

# An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the *Drosophila engrailed* gene

J. Lesley Brown, Daniel J. Grau, Sarah K. DeVido and Judith A. Kassis\*

Laboratory of Molecular Genetics, National Institutes of Child Health and Human Development,  
National Institutes of Health, Bethesda, MD, 20892, USA

Received June 13, 2005; Revised and Accepted August 23, 2005

## ABSTRACT

**Polycomb-group response elements (PREs) are DNA elements through which the Polycomb-group (PcG) of transcriptional repressors act. Many of the PcG proteins are associated with two protein complexes that repress gene expression by modifying chromatin. Both of these protein complexes specifically associate with PREs *in vivo*, however, it is not known how they are recruited or held at the PRE. PREs are complex elements, made up of binding sites for many proteins. Our laboratory has been working to define all the sequences and DNA binding proteins required for the activity of a 181 bp PRE from the *Drosophila engrailed* gene. Here we show that one of the sites necessary for PRE activity, Site 2, can be bound by members of the Sp1/KLF family of zinc finger proteins. There are 10 Sp1/KLF family members in *Drosophila*, and nine of them bind to Site 2. We derive a consensus binding site for the Sp1/KLF *Drosophila* family members and show that this consensus sequence is present in most of the molecularly characterized PREs. These data suggest that one or more Sp1/KLF family members play a role in PRE function in *Drosophila*.**

## INTRODUCTION

Polycomb-group response elements (PREs) and Trithorax-group response elements (TREs) are DNA elements that are the targets of the *Drosophila* Polycomb-group (PcG) and Trithorax-group (TrxG) of genes, respectively. PcG genes encode proteins that act as transcriptional repressors and TrxG genes encode transcriptional activators. The best-studied

targets of the PcG and TrxG genes are the homeotic genes in *Drosophila* where PcG and TrxG genes do not initiate the pattern of gene expression, but rather serve as the molecular memory to keep genes in either the off or on state. At the homeotic genes, PREs and TREs are often closely associated or interspersed [for reviews see (1,2,3)].

Many PcG and TrxG genes encode proteins that work in complexes to modify chromatin. For example, there are two well characterized PcG complexes, PRC2, or the E(z)/Esc complex (4–7) and PRC1, a complex that contains the PcG proteins Polycomb (Pc), Polyhomeotic (Ph), dRing/Sex combs extra (Sce), (8) and Posterior sex combs (Psc) (9). Chromatin-immunoprecipitation experiments have shown that PcG proteins are specifically bound to PREs *in vivo* [e.g. (10,11)], however, it is not known how they get recruited to the DNA, since these complexes do not specifically associate with PRE sequences *in vitro*.

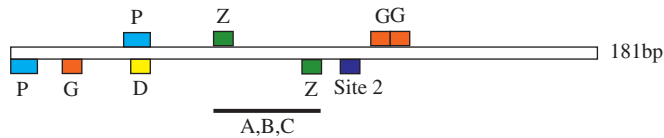
Much effort has gone into trying to understand what constitutes a PRE in order to understand how PcG protein complexes are recruited to the DNA and to predict the occurrence of PREs by sequence analysis (12). PREs are complex elements—no single DNA binding site can act as a PRE. Instead, PREs are made up of binding sites for many different proteins. To date most PREs studied at the molecular level have been shown to contain DNA binding sites (usually multiple copies) for the proteins Pleiohomeotic (Pho) and its partially redundant homolog Pleiohomeotic-like (Phol), Zeste, GAGA Factor (GAF)/Pipsqueak (Psq) [reviewed in (1)] and Dorsal Switch Protein 1 (Dsp1) (13). Ringrose *et al.* (12) have reported that clustered pairs of GAF/Psq, Zeste and Pho/Phol sites can predict the location of many PREs. However it has recently been shown that a combination of Zeste, Dsp1, GAF/Psq and Pho/Phol sites in the same number, orientation and spacing as in the native PRE is insufficient to restore full PRE activity (13). This suggests that all of the DNA binding sites necessary for PRE activity are still not known.

\*To whom correspondence should be addressed. Tel: +1 301 496 7879; Fax: +1 301 496 0243; Email: jkassis@mail.nih.gov  
Present address:

Daniel J. Grau, Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org



**Figure 1.** Summary of factor binding sites identified within the 181 bp *engrailed* PRE. The 181 bp fragment is shown by an unfilled box going from 5' to 3' (left to right) with respect to the *engrailed* transcription start site. The locations of the binding sites are shown by colored boxes above (for sites on the + strand) or below (for sites on the – strand) the line. Binding site abbreviations are as follows; P, Pho/Phol; G, GAF/Psq; D, Dsp1; Z, Zeste; A,B,C are currently unidentified factor binding sites described in Americo *et al.* (16).

With the aim of identifying the binding sites and factors necessary for PRE function, our lab has been trying to define the components that constitute a minimal 181 bp PRE at –576 to –395 upstream of the *Drosophila engrailed* gene. This element was originally identified in a pairing-sensitive silencing assay (14), an assay used to detect the function of many PREs. When PREs are included in the vector pCaSpeR, they have an unusual effect on the eye color marker *mini-white*. Normally, when transgenic flies are made with pCaSpeR, flies homozygous for the transgene have a darker eye color than flies heterozygous for the transgene. However, when a PRE is included in the pCaSpeR vector, the eye color of homozygous flies is often lighter than that of heterozygotes. This phenomenon is dependent on the chromosomes being able to pair. Fragments of DNA that mediate pairing-sensitive silencing are called pairing-sensitive elements (PSEs). Not all PSEs have been shown to act as PREs and vice versa [for review see (15)]. The *engrailed* 181 bp element behaves both as a PRE and a PSE (16).

The 181 bp *engrailed* PRE contains 3 GAF/Psq sites, 2 Pho/Phol sites, 2 potential Zeste sites and 1 Dsp1 site that almost entirely overlaps the Pho/Phol site that we have studied by mutational analysis (Figure 1). In addition to these known sites, we have identified a number of other protein-binding sites required for pairing-sensitive silencing [Figure 1, ref (16)]. Here we investigate the role of one of these protein-binding sites, Site 2, in PRE function and show that the Sp1/KLF family of proteins bind to this site. We find that Sp1/KLF binding sites are present in most well characterized PREs.

## MATERIALS AND METHODS

### Analysis of PRE activity

The construction of the 181-*bx-d-Ubx-LacZ* construct was described in Americo *et al.* (16). The MutSite2 181 PRE was amplified from the pCasper/MutSite2 construct as described in Americo *et al.* (16) using 5' and 3' primers containing Xho1 sites. The amplified product was cloned as an Xho1 fragment into the Xho1 site of the *bx-d-Ubx-lacZ* construct. The base changes in MutSite2 are shown in Figure 3. The pCasper/MutPho construct described in Americo *et al.* (16) was used as the starting point to introduce mutations in the second potential Pho/Phol binding site in the 181 PRE. A 5' primer carried an Xho1 site and the relevant Pho/Phol site mutations from (GAGATGGC to GAGCGTGC, the match to the Pho/Phol consensus binding site is on the lower strand). The 3' primer carried an Xho1 site. The amplified product was

cloned into the Xho1 site of the *bx-d-Ubx-lacZ* construct. All plasmids were sequenced over the 181 bp PRE to confirm the presence of the described mutations. The *bx-d-Ubx-lacZ* 181 wild-type and mutated derivatives were injected into a ry<sup>506</sup> strain by Genetic Services Inc.

### Yeast one-hybrid screen

A yeast one-hybrid screen was performed using the Clontech yeast one-hybrid kit. Standard yeast techniques were used as recommended by Clontech. A double-stranded oligonucleotide comprised of Site 2 (ACTCGGTAACGCCCGTGA) repeated five times bounded by EcoR1 and Xho1 restriction sites was cloned into the EcoR1, Xho1 cleaved pHisi-1 vector (Clontech) to create pHisi-1:Site2. The pHisi-1:Site2 plasmid was linearized and integrated into the yeast genome. A *Drosophila* 0–18 h cDNA library (Clontech) was transformed into the yeast strain with the integrated pHisi-1:Site2 plasmid. Positive colonies were isolated after 4–6 days of incubation at 30°C on selective media. Plasmid DNA was isolated and sequenced. In addition to the clones reported here, we sequenced 31 other potential positives, only two others had potential DNA binding domains. These proteins did not bind to Site 2 after *in vitro* transcription/translation.

### Cloning of the zinc finger domains into the *pT7link* expression vector

The zinc finger domains of the Sp1/KLF proteins were generated either by PCR directly from the cDNA or by RT-PCR from genomic DNA. Primers were designed that would PCR amplify the entire zinc finger region of each protein with an additional 8 N-terminal amino acids. The primers contained BamHI and XhoI ends for cloning the zinc finger regions downstream of the  $\beta$ -globin 5'-UTR and ATG in the *pT7link* expression vector (provided by R. Treisman). Each of the clones was confirmed by sequencing and used for *in vitro* transcription/translation.

### Gel mobility shift assay

Recombinant proteins were synthesized *in vitro* using the TNT-coupled transcription/translation system (Promega). Gel mobility shift assays were performed as described previously (16) using 3  $\mu$ l of the *in vitro* translation reaction.

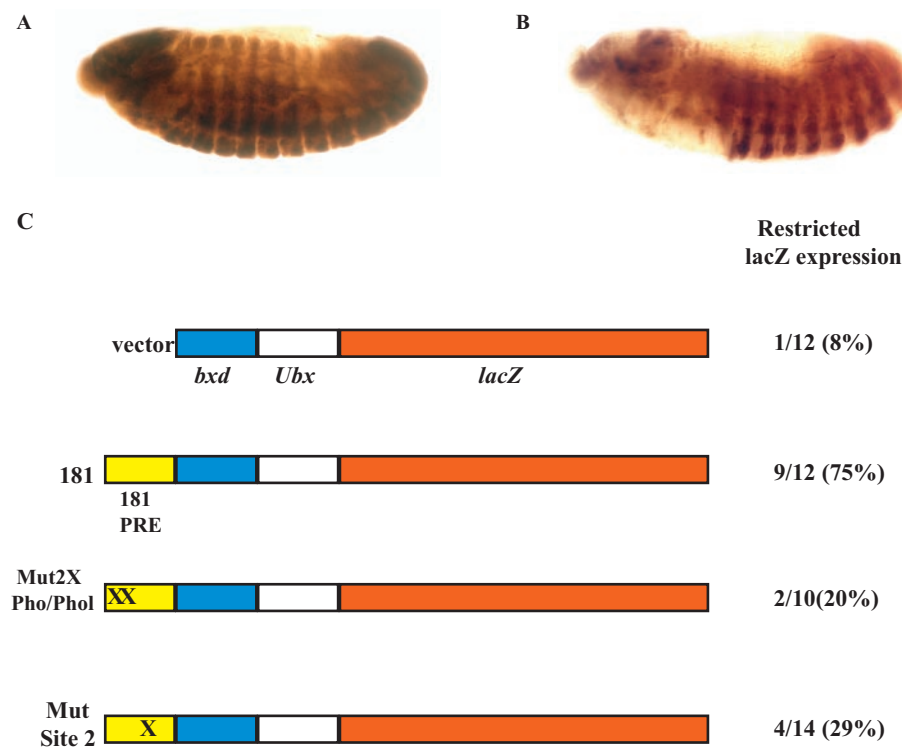
### Searching known PRE and PSE sequences for Site 2 and other known protein-binding sites

The sequences of known PREs and PSEs were searched for the presence of consensus binding sites for Sp1/KLF family members, Pho/Phol, GAF/Psq, Zeste and DSP1 proteins. The exact parameters of the consensus sequences are given in the legend to Table 1. Consensus sequences were localized using a sequence visualization software program named Gene Palette ([www.genepalette.org](http://www.genepalette.org)), (17). Pho/Phol, GAF/Psq and one of the Zeste consensus sites were chosen based on Ringrose *et al.* (12). The Sp1/KLF consensus site is as described in the text. We used an additional version of the Zeste consensus site (BGAGTGV) as described by Mohrmann *et al.* (18). The DSP1 consensus site came from Dejardin *et al.* (13).

**Table 1.** Sites for PRE-binding proteins within known PREs and PSEs

Regulatory DNA	Length (bp)	Pho	GAF/Psq	Sp1/KLF	Zeste	Dsp1	Order of sites within regulatory DNA <sup>a</sup>	Accession number (basepairs)
<i>Engrailed</i> PRE (16)	181	2	3	1	2	1	P,G,P,D,Z,Z,S,G,G	AE003825.3 (62974–63154)
<i>lab-7</i> PRE (53)	219	3	2	1	3	1	P,D,S,P,G,G,Z,Z,P,Z	AE003715.4 (100059–100278)
<i>lab-8</i> PRE (54)	1413	5	1	2	1	16	G,P,D,P,Z,D,D,D,D,D,P,P,S,D,D, D,D,D,D,D,S,P,D	AE003715.4 (119929–121341)
PRE <sub>D</sub> <i>Ubx</i> PRE <sub>D</sub> (24)	567	6	5	1	2	3	D,D,Z,D,P,G,S,P,G,P,G,Z,G,G,P,P,P	AE003714.2 (195965–195399)
<i>Engrailed</i> PSE (14)	450	2	3	3	4	2	S,Z,D,S,G,P,G,Z,G,Z,D,P,S,Z	AE003825.3 (64078–64527)
<i>Escargot</i> PSE (14)	985	1	6	2	4	3	S,D,G,G,G,G,P,S,Z,G,Z,Z,Z,G,D,D	AE003647.3 (67517–68504)
<i>Even-skipped</i> PSE (55)	601	1	2	2	1	4	D,D,D,G,D,G,Z,S,S,P	AE003831.3 (139977–140578)
<i>Proboscipedia</i> PSE (56)	580	3	1	3	1	6	S,D,S,Z,G,D,D,D,D,D,P,S,P,P	AE003674.3 (262509–263088)
Minimal MCP (28)	142	4	1	0	0	2	G,P,P,P,D,P,D	AE003715.4 (69732–69885)
MCP 823 bp (57)	823	5	1	1	1	9	D,D,P,D,D,S,D,Z,G,P,P,P,D,P,D,D,D	AE003715.4 (69199–70021)

<sup>a</sup>Pho/Phol (P), GAGA Factor/Pipsqueak (G), Sp1/KLF family members (S), Zeste (Z), Dsp1 (D). Consensus sites used: P-GCCAT, G-GAGAG, S-RRGGYG, Z-BGAGTGV, YGAGYG, Dsp1 GAAAA.



**Figure 2.** PRE activities of a wild-type and mutated 181 bp PRE from the *Drosophila engrailed* gene. (A and B) Representative embryos showing unrestricted (A) and restricted  $\beta$ -galactosidase expression (B) as visualized by immunoperoxidase staining using anti- $\beta$ -galactosidase antibodies. Embryos are oriented with anterior left, dorsal up. Lateral views of embryos at ~10 h after egg laying are shown. (C) Schematic representation of the constructs and a summary of the number of lines for each construct that show restricted  $\beta$ -galactosidase expression in late embryos/total number of lines. Percentages are also shown. (X) represents mutations in the binding sites as indicated in the name of the constructs. The sequence changes of the two mutated Pho/Phol sites and the mutated Site 2 sequence are given in the Materials and Methods and Figure 3, respectively.

## RESULTS

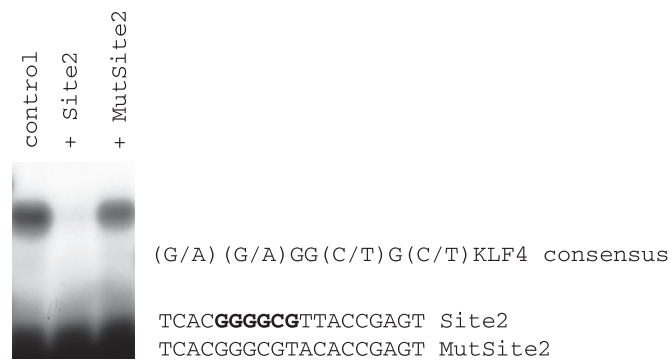
### Site 2 is required for PRE activity of the 181 bp *engrailed* fragment

We have previously shown that Site 2 is required for pairing-sensitive silencing of the 181 bp *engrailed* PRE (16). Here we ask whether Site 2 is also important for the PRE activity of that fragment. We used the *bxd-Ubx-lacZ* reporter construct (19) to test the effect of mutations in Site 2 on PRE activity. In the absence of a PRE, the *bxd-Ubx-lacZ* reporter construct expresses *lacZ* throughout the embryo, in both the ectoderm

and the nervous system, late in development (Figure 2A). When the 181 bp PRE from the *Drosophila engrailed* gene is included in this vector, *lacZ* expression is restricted to parasegment six and posterior segments [Figure 2B and ref (16)]. We saw restricted expression patterns in 75% of our lines, a number consistent with what has been seen by other investigators as PREs are not active in all chromosomal insertion sites.

We next introduced mutations into either Site 2 or the two Pho/Phol binding sites in the 181 bp fragment in the context of the *bxd-Ubx-lacZ* vector. The effect on PRE activity, as

assayed by embryonic  $\beta$ -galactosidase expression patterns, was analyzed. Mutation of either the Pho/Phol sites or Site 2 caused a reduction in the percentage of lines that had PRE activity. Both mutated constructs gave results that were significantly different from the expected frequency of 75% PRE activity for the unmutated construct ( $P < 0.001$  by a chi-squared test). There were still a small percentage of lines with restricted expression in the Pho/Phol (20%), and the Site 2 mutant transgene (29%) as well as in the vector only control (8%). This probably reflects the fact that the *bxd* enhancer and the *Ubx* promoter are poised to work with flanking genomic PREs and may contain weak PRE activity on their own (19). When we scanned the first 350 bp of the *bxd* sequence that borders the 181 bp element in our constructs for the known PRE DNA binding sites, we found 3 Pho/Phol sites, 3 Dsp1 sites, 1 Zeste site and 2 potential Sp1 sites. This may explain why mutation of Pho/Phol or Site 2 did not completely eliminate the PRE activity of the 181 bp *engrailed* fragment in this vector. Nevertheless, our results suggest that both Pho/Phol and Site 2 binding sites contribute to the activity of the *engrailed* PRE, and that neither alone is sufficient for PRE activity.



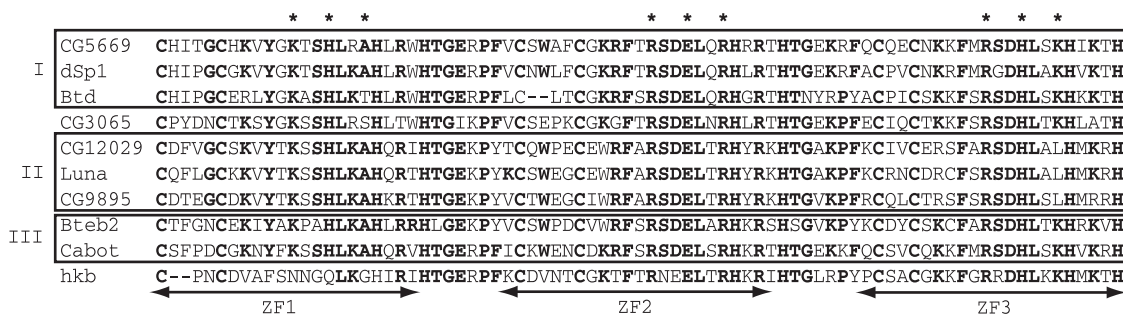
**Figure 3.** The Sp1/KLF family member CG5669 binds specifically to Site 2. The Site 2 sequence was radioactively labeled and used in a gel shift experiment with the *in vitro* transcribed/translated zinc finger domain of CG5669. A single specific complex was detected (control) that was competed by 100-fold excess of the Site 2 oligo (Lane 2) but not by a Site 2 oligo with five base substitutions (MutSite2 oligo, Lane 3). The sequence of the Site 2 and MutSite2 oligonucleotides are shown. The base substitutions in MutSite2 are underlined. The KLF4 derived consensus sequence is shown. The bases within Site 2 that are a match to the Sp1/KLF consensus sequence are shown in bold.

### Members of the Sp1/KLF family bind to Site 2

In order to identify the protein(s) that interact with Site 2 we carried out a yeast one-hybrid screen using multimerized Site 2 as bait. We isolated four cDNA clones that showed specific binding to Site 2 but not to a mutated Site 2 (Figure 3). Each of these four clones encoded a member of the Sp1/KLF family of zinc finger proteins [CG5669, CG12029 (two different length clones) and *luna*]. The consensus binding sequence for the Sp1/KLF family based on mutational analysis of a KLF family member (KLF4) binding site is (G/A)(G/A)GG(C/T)G(C/T), (20). The Site 2 sequence contains a perfect match to this consensus sequence (Figure 3).

The Sp1/KLF family is an important group of proteins that in mammals have been shown to be involved in cell morphogenesis, differentiation and cancer. These proteins share a high degree of homology over 3 Cys2/His2 zinc fingers (>65% sequence identity with each other), located at or close to the C-terminal end of the protein. The N-terminal regions are generally unique [for reviews see (21,22)].

We identified three members of the Sp1/KLF family in our yeast one-hybrid screen, however our screen was not saturating. When we searched the *Drosophila* genome sequence for homology to the zinc finger region of these proteins we identified a total of 10 members of this class in *Drosophila*. The zinc finger regions of these members are highly conserved as summarized in Figure 4. The mammalian factors can be subdivided into three classes (Sp1/class I family, class II family and class III family) based on closer homology between members of the subfamily relative to members outside the subfamily [reviewed in (21)]. This homology sometimes extends to functional protein domains that lie outside the DNA binding domain. The *Drosophila* proteins can also be placed into different groups based on homology of the zinc fingers with the zinc fingers of the human Sp1/KLF proteins. CG5669, dSp1 and Btd are more closely related to the Sp1/Class I proteins than to the KLF proteins. In fact, dSp1 has 97% amino acid identity with hSp8 zinc fingers and also has homology to hSp8 in the region N-terminal to the zinc fingers. For CG5669 and Btd it is hard to assign a homolog since they have significant identity with a number of Class I proteins. Sequence identities of CG5669 with the zinc fingers of Class I range from 76 to 87%, for Btd the range is 65–74%. CG12029, Luna and CG9895 have closer amino acid identity to the Class II human proteins. Luna is most closely related to



**Figure 4.** Amino acid sequence alignment of the zinc finger domains of the *Drosophila* Sp1/KLF family of proteins. Residues that are identical in the majority of proteins are shown in bold. An asterisk denotes amino acids that are predicted from mammalian studies to interact with the DNA binding site. The zinc fingers are delimited under the sequence. The boxed areas denote proteins that can be grouped into different subclasses based on homology with the human Sp1/KLF family of proteins, refer to text.

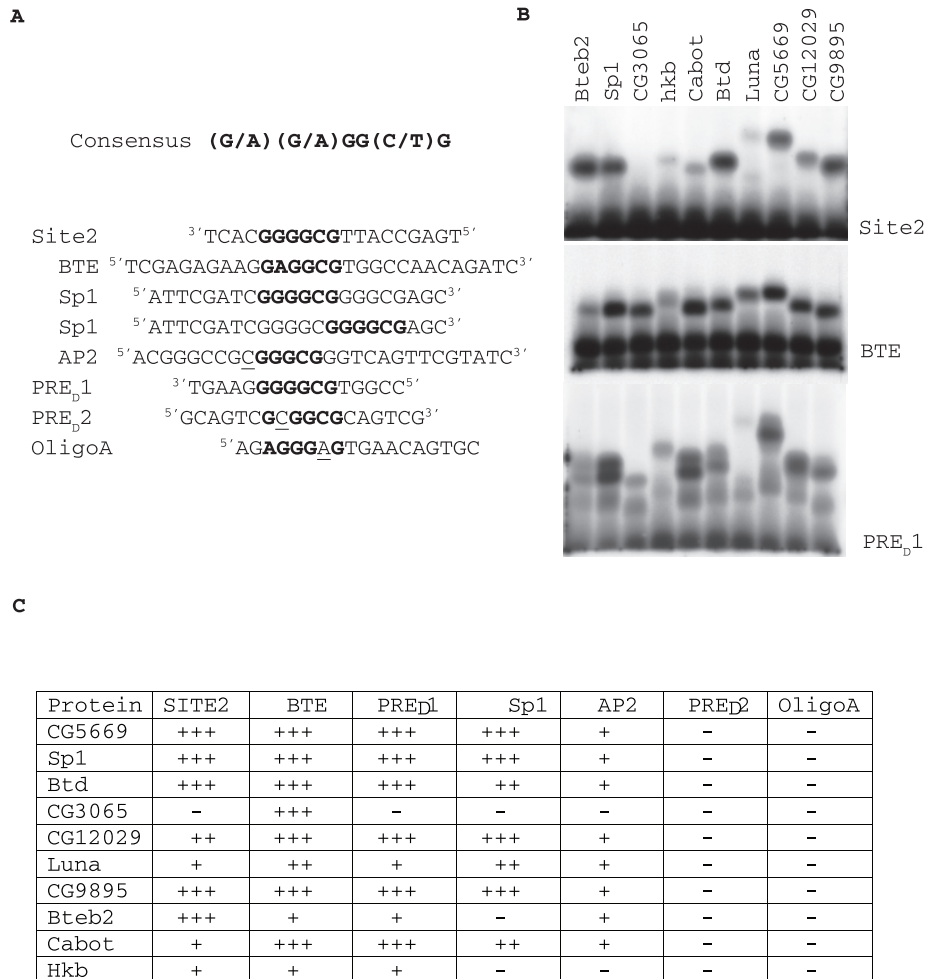


KLF6 and KLF7 and shows conservation of a putative activation domain at the N-terminus (23). Cabot (also known as CG4427) and Bteb2 seem to belong to the class III subclass showing 72–85% identity with human members of this class. Hkb and CG3065 are harder to place. CG3065 has 62–65% identity to members of the Sp1 family and 62–64% identity with KLF9 and KLF14 members of class III. Hkb is the most diverged member of this class of proteins. The highest degree of homology for Hkb is 53% identity with hSp5 and it is difficult to say whether it should be placed in this family at all. We note that Hkb binds only weakly to sequences that match the Sp1/KLF consensus (Figure 5).

Each of the zinc finger regions of these 10 proteins were cloned in frame into an *in vitro* transcription/translation vector and expressed *in vitro*. The products were tested for binding to Site 2. Nine of the 10 *Drosophila* Sp1/KLF family member zinc fingers show binding to Site 2 and only CG3065 shows no binding (Figure 5B). Binding to each of these factors is

specific based on competition experiments (Figure 3 and data not shown). No binding was seen with product from the vector alone (data not shown). Binding of Luna, Cabot and Hkb to Site 2 is very weak. This result is not due to the production of inactive protein since, with the exception of Hkb, we see strong binding of these factors to other binding site sequences (see below Figure 5).

In order to get a better idea of the binding specificity of the *Drosophila* Sp1/KLF family members, we tested their binding to six additional oligonucleotides (Figure 5). BTE, Sp1 and AP2 are sequences that have previously been used to test binding specificity of human KLF4 (20). Note that the oligonucleotide named Sp1 contains two matches to the consensus sequence. PRE<sub>D</sub>1 and PRE<sub>D</sub>2 are sequences within an *Ubx* PRE, PRE<sub>D</sub> (24) that most closely matched the Sp1/KLF binding site. Oligo A (16) is a sequence from the *engrailed* 181 PRE with only one mismatch with the Sp1/Klf consensus. Not surprisingly, like the mammalian Sp1/KLF family



**Figure 5.** DNA binding characteristics of the zinc finger regions of *Drosophila* Sp1/KLF family members. (A) Shows the sequence of the DNA probes used in the bandshift experiments. The Sp1/KLF consensus sequence is shown. Bases that match the Sp1/KLF consensus sequence are denoted in bold type. Nucleotides that lie within the consensus but do not match it are underlined. The Sp1 oligo contains two potential Sp1/KLF binding sites. (B) Gel mobility shift assay using radioactively labeled binding site probes incubated with *in vitro* transcribed/translated zinc finger regions of the *Drosophila* Sp1/KLF family of proteins. The probe used in the bandshift is denoted beside each panel. The zinc finger regions used for the bandshift are shown at the top of the lanes. These experiments were performed simultaneously, and were repeated three times. (C) Binding of each Sp1/KLF family member to seven oligonucleotide probes. (+++) Strong binding (++) Moderate binding (+) Weak binding.

members, our data indicate that the *Drosophila* Sp1/KLF proteins bind with different affinities to the different sites. Both sequences within the consensus and bases that flank the consensus were important for binding. For example although Site 2 and PRE<sub>D</sub>1 perfectly match each other in the Sp1/KLF core consensus sequence, the relative binding of the Sp1/KLF proteins differ between the two sites. Bteb2 binds strongly to Site 2 but weakly to PRE<sub>D</sub>1 and conversely, Cabot binds strongly to PRE<sub>D</sub>1 but weakly to Site 2. This result shows the influence of flanking sequences on DNA binding affinity. The exact base composition of the core consensus sequence is also important. The BTE and PRE<sub>D</sub>1 core differ in only one base, a G versus an A in the second position and sequences flanking the consensus are identical for four bases on each side. Despite this, CG3065 binds more strongly to the BTE sequence than to PRE<sub>D</sub>1.

We next tested sequences with mismatches within the consensus to get a better idea of the binding specificity of the *Drosophila* family members. PRE<sub>D</sub>2, has one mismatch to the consensus, it has a C rather than a G or A in position 2. PRE<sub>D</sub>2 did not bind any of the *Drosophila* Sp1/KLF family members. An AP2 binding site, that contained a C instead of a G or A at the first position did not bind any of the Sp1/KLF zinc fingers with high affinity. The importance of the first base is also seen with the Mutsite2 oligonucleotide (Figure 3). Mutsite2 contains the sequence CGGGCGT, a mismatch in the first position of the consensus, and does not compete for binding with the Site 2 oligonucleotide. Oligo A, which differs from the consensus sequence at position 5 having an A instead of a C or T, did not bind any of the 10 Sp1/KLF zinc fingers. This data set tells us that the Sp1/KLF consensus derived by Shields and Yang (20) is largely predictive of which sites will be bound by the *Drosophila* factors. Shields and Yang (20) report a C/T for the seventh base of the KLF consensus binding site. However, their experiments showed that an A in the seventh position was also able to compete for binding of the KLF4 factor although not as well as a C or T in this position. This suggests that A can also be added to the seventh position. In addition, binding sites for the mammalian Sp1 class of factors would predict that a G in the last position of the consensus sequence is also a valid binding site (25). Note that the Sp1 oligonucleotide we used contained both the sequences GGGGCGG and GGGGCGA. This oligonucleotide was bound quite well by several factors. Based on our analysis here, the data of Shields and Yang (20), and what is known of the Sp1 binding site we propose that any base may function in the seventh position of the consensus making it (G/A)(G/A) GG(T/C)G.

### Sp1/KLF binding sites are present in many PREs and PSEs

We have found sequences that are bound by the Sp1/KLF family of proteins in both the *engrailed* 181 bp PRE and in PRE<sub>D</sub> of *Ubx*. Are Sp1/KLF binding sites a general feature of PREs as has been found for Pho/Pho1, Dsp1, GAF/Psq and Zeste binding sites? To address this question we have searched the molecularly well-defined PRE (fragments under 1.5 kb) and PSEs sequences for the presence of the Sp1/KLF consensus sequence using the Gene Palatte sequence analysis program (see Materials and Methods). We searched

using the consensus sequence described above [(G/A) (G/A) GG (T/C) G]. We note that this consensus is based on a small data sample and may not detect all potential Sp1/KLF binding sites. For example, some of the mammalian family members bind to a GT box [GGTGTGGGG, (26)] a sequence that is not contained in the Sp1/KLF consensus we used. We also included in our search the distribution of potential Pho/Pho1, GAF/Psq, Zeste and Dsp1 sites in these PREs/PSEs. Of the fragments we examined, the 181 *engrailed* PRE, *iab-7* PRE and the *iab-8* PRE have all been shown to act as both PREs and to mediate pairing-sensitive silencing (PSS). PRE<sub>D</sub> has not been tested for PSS, but multimerized copies of sub-fragments of this PRE have been shown to mediate PSS (27). A 450 bp *engrailed* fragment located from -1944 to -1503 upstream of the *engrailed* transcription start site and fragments of DNA from *escargot*, *even-skipped* and *proboscipedia* mediate PSS. The minimal MCP element, a 142 bp fragment from the *iab-5* regulatory region of the *Abdominal-B* gene, acts as a PRE in vectors designed to test for PRE activity but does not mediate PSS (28,29).

The distributions of the various binding sites in the PRE/PSEs are summarized in Table 1. Of the PREs and PSEs we examined, all except the minimal MCP element have potential Sp1/KLF binding sites and about one-half of them contain more than one potential Sp1/KLF consensus binding site. In addition all but the minimal MCP element contain potential Pho/Pho1, GAF/Psq, Zeste and Dsp1 sites, often in multiple copies (see Table 1). The minimal MCP1 element seems to be atypical by not having Zeste or Sp1/KLF consensus binding sites. This MCP element was tested for PRE activity in a vector that, like the vector we used to test the PRE activity of the *engrailed* 181 bp PRE, contains *Ubx* sequences containing many potential Sp1/KLF binding sites as well as binding sites for other PRE-binding factors. When we searched the larger 823 bp MCP sequence that has been shown to have insulator and pairing activity, we detect a potential Sp1/KLF binding site and a match to the Zeste consensus binding site. It is interesting to note that the minimal MCP element lacks Zeste sites and does not mediate PSS. Zeste has long been known to be involved in transvection, mediating *trans*-interactions between chromosomes and it has been speculated that it may play a role in mediating interactions between the PRE and the promoter in PcG silencing (30).

The presence of potential Sp1/KLF sites in so many different PREs and PSEs suggests that Sp1/KLF binding sites may play a general role in PRE function. We note that the Sp1/KLF consensus sequence was not one of the sequence motifs found to be statistically enriched in PREs over random sequence by Ringrose *et al.* (12). However, this may be because it is so short and degenerate (RRGGYG). The Zeste binding motif is also very short and degenerate (YGAGYG) and was also not statistically enriched in PREs over random sequence (Ringrose *et al.* (12). Clearly the function of Sp1/KLF binding sites in other PREs will need to be tested, preferably in vectors that do not contain nearby Sp1/KLF consensus sequences.

### Function of the *Drosophila* Sp1/KLF family members

The Sp1/KLF family of proteins in mammals is very complex. These proteins can be activators or repressors; some can do

both depending on cellular and binding site context. Some proteins function ubiquitously; whilst others function only in specific cells, and some do both at different stages of development. For example, Sp1 and Sp3 proteins function redundantly early in embryogenesis although they have different functions later in development. There is also competition for binding sites. Post transcriptional modifications can also affect the function of these proteins [for reviews see (21,22)].

Of the 10 *Drosophila* Sp1/KLF members, four are well characterized genetically (*btd*, *dSp1*, *hkb* and *cabot*). Since these four genes are involved in the development of particular structures in the fly and are not ubiquitously expressed, we believe it unlikely that they play general roles in PcG repression. *buttonhead* (*btd*) is important for the development of head structures and mechano-sensory organs (31). *dSp1* is located 50 kb from *btd* and may act with *btd* in mechano-sensory organ development (32). *hkb* mutations have been studied extensively and are found to affect the specification of endoderm (33), salivary gland development (34), germ cell migration (35), as well as many other developmental processes. Overexpression of *cabot* affects sensory organ development (36). The function of the *luna* gene during embryogenesis has been studied using RNAi (23), and the phenotypes obtained do not mimic PcG phenotypes. However, analysis of *luna* mutants will be necessary to assess its role in PcG function.

We obtained a mutant allele of the *Bteb2* gene from the *Drosophila* genome project. These flies contain an insertion of a piggyback transposon into the coding region of the *Bteb2* gene and should produce a null mutation in this gene. These flies are homozygous viable and fertile and have no phenotypic defects. Thus, if the *Bteb2* gene plays a role in PcG repression, its role is likely redundant with that of another member of the family. Mutations in the other Sp1/KLF family members have not yet been isolated.

## DISCUSSION

Here we describe another binding site important for function of the *engrailed* 181 bp PRE in *Drosophila* and present evidence that Sp1/KLF family members can bind to that site. Consensus binding sites for this family of proteins have been found in most of the well characterized PREs implying that this binding site may play a general role in PRE activity.

Identifying which Sp1/KLF factor acts through Site 2 is not a easy task. Not only are a number of members of this class genetically uncharacterized there is also the possibility that there may be functional redundancy as is seen with Pho and Phol. The existence of a viable and fertile *Bteb2* mutant suggests that functional redundancy will be observed with the Sp1/KLF family in *Drosophila*. Experiments using family member-specific antibodies in chromatin-immunoprecipitation experiments on PREs will help elucidate which Sp1/KLF family members play a role in PRE function in *Drosophila*.

What role the Sp1/KLF family of proteins play in recruiting the PcG complexes to the PRE remains to be elucidated. In fact, for most of the other proteins required for PRE function, their roles are not yet clear. Both GAF and Psq bind the

sequence GAGAG (37,38), a sequence shown to be important for PRE function (28,30). Psq has been shown to be in a complex with PcG proteins isolated from the *Drosophila* cell line SL2 (39,40) and *psq* mutations enhance the mutant phenotypes of the PcG genes polyhomeotic (Ph) and Polycomb (Pc) in larval and adult tissues (39,41). This suggests that Psq may be important for PRE function. GAF [for review see ref (42)] has also been reported to co-purify with some PcG proteins (27) and has been shown by chromatin-immunoprecipitation experiments to be present at PREs (10). GAF is a member of the TrxG of genes but may also play a role in PcG repression (27,30,43). The DNA binding protein Pho has been shown to bind *in vitro* to a chromatinized PRE template only if GAF is present (44). GAF and Psq can interact through their BTB protein-protein interaction domains and it has been proposed that they may function together *in vivo* (45).

Zeste has been shown to be important for both PRE and TRE activity. Zeste is a stoichiometric component of the biochemically purified PcG complex, PRC1 (9) suggesting a role in PcG repression and Hur *et al.* (30) report that Zeste is required for the PcG-mediated repression of an *Ubx* transgene. In contrast, experiments with the *iab-7* PRE have shown that Zeste binding sites are important for the ability of this DNA to act as a TRE, not as a PRE (46).

Pho and Phol (47,48) have recently been shown to be required to recruit an Esc-E(z) complex to a PRE (49). *In vitro*, Pho interacts directly with E(z) and Esc whereas Phol interacts with Esc. Recruitment of the Esc-E(z) complex leads to methylation of lysine 27 of histone H3 by the SET domain of E(z). The methylated K27 recruits a Pc-containing complex through interaction of the chromo-domain of Pc with the methylated histone tails. Pho has also been shown to interact with Pc *in vitro* (50). Pho/Phol double mutants have a very strong PcG phenotype, much stronger than mutations in the genes encoding the other PRE-binding factors suggesting that Pho/Phol play a central role in PRE function. It has been proposed that Dsp1 facilitates the binding of Pho/Phol to the PRE (13).

The role that the Sp1/KLF family may play remains to be elucidated but it is intriguing to note that mammalian Sp1 has been reported to interact directly with YY1 (the mammalian homolog of Pho). This interaction requires the first one and a half zinc fingers of YY1, a region that is 96% identical between the *Drosophila* and mammalian proteins. The 158 amino acid C-terminal region of Sp1 (includes the three zinc fingers and one of the activation domains, domain D), can mediate the interaction leading to an increase in the level of correctly initiated transcripts (51,52). These data raise the possibility that Pho or Phol may interact with Sp1/KLF proteins at PREs.

## ACKNOWLEDGEMENTS

We thank Mark Mortin and Karl Pfeifer for critical reading of the manuscript and Jim Kennison for discussions. This research was supported by the Intramural Research Program of the NIH, NICHD. Funding to pay the Open Access publication charges for this article was provided by NIH, NICHD.

*Conflict of interest statement.* None declared.



## REFERENCES

- Ringrose, L. and Paro, R. (2004) Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. *Annu. Rev. Genet.*, **38**, 413–443.
- Levine, S.S., King, I.F. and Kingston, R.E. (2004) Division of labor in polycomb group repression. *Trends Biochem. Sci.*, **29**, 478–485.
- Lund, A.H. and Lohuizen, M. (2004) Polycomb complexes and silencing mechanisms. *Curr. Opin. Cell Biol.*, **16**, 239–246.
- 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*, **125**, 275–286.
- Tie, F., Furuyama, T., Prasad-Sinha, J., Birve, A., Rasmuson-Lestander, A. and Harte, P.J. (2003) A 1-megadalton ESC/E(Z) complex from *Drosophila* that contains polycomblike and RPD3. *Mol. Cell Biol.*, **23**, 3352–3362.
- Czermin, B., Melfi, R., McCabe, D., Steitz, V., Imhof, A. and Pirrotta, V. (2002) *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, **111**, 185–196.
- Mueller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E. and Simon, J.A. (2002) Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell*, **111**, 197–208.
- Fritsch, C., Beuchle, D. and Müller, J. (2003) Molecular and genetic analysis of the Polycomb group gene sex combs extra/ring in *Drosophila*. *Mech. Dev.*, **120**, 949–954.
- Saurin, A.J., Shao, Z., Erdjument-Bromage, H., Tempst, P. and Kingston, R. (2001) A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. *Nature*, **412**, 655–660.
- Strutt, H., Cavalli, G. and Paro, R. (1997) Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.*, **16**, 3621–3632.
- 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.
- Ringrose, L., Rehmsmeier, M., Dura, J.M. and Paro, R. (2003) Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev. Cell*, **5**, 759–771.
- Dejardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decouvill, M., Locker, D. and Cavalli, G. (2005) Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1. *Nature*, **434**, 533–538.
- Kassis, J.A. (1994) Unusual properties of regulatory DNA from the *Drosophila engrailed* gene: three 'pairing-sensitive' sites within a 1.6 kb region. *Genetics*, **136**, 1025–1038.
- Kassis, J. (2003) Pairing-sensitive silencing, polycomb group response elements and transposon homing in *Drosophila*. *Adv. Genet.*, **46**, 421–438.
- Americo, J., Whiteley, M., Brown, J.L., Fujioka, M., Jaynes, J.B. and Kassis, J.A. (2002) A complex array of DNA-binding proteins required for pairing-sensitive silencing by a polycomb group response element from the *Drosophila engrailed* gene. *Genetics*, **160**, 1561–1571.
- Rebeiz, M. and Posakony, J.W. (2004) GenePalatte: a universal software tool for genome sequence visualization and analysis. *Dev. Biol.*, **15**, 431–438.
- Mohrmann, L., Kal, A.J. and Verrijzer, C.P. (2002) Characterization of the extended Myb-like DNA-binding domain of trithorax group protein Zeste. *J. Biol. Chem.*, **277**, 47385–47392.
- Müller, J. and Bienz, M. (1991) Long range repression conferring boundaries of Ultrabithorax expression in the *Drosophila* embryo. *EMBO J.*, **10**, 3147–3155.
- Shields, J.M. and Yang, V.W. (1998) Identification of the DNA sequence that interacts with the gut-enriched Kruppel-like factor. *Nucleic Acids Res.*, **26**, 796–802.
- Kaczynski, J., Cook, T. and Urrutia, R. (2003) Sp1- and Kruppel-like transcription factors. *Genome Biol.*, **4**, 206.
- Li, L., He, S., Sun, J.M. and Davie, J.R. (2004) Gene regulation by Sp1 and Sp3. *Biochem. Cell Biol.*, **82**, 460–471.
- De Graeve, F., Smaldone, S., Laub, F., Mlodzik, M., Bhat, M. and Ramirez, F. (2003) Identification of the *Drosophila* progenitor of mammalian Kruppel-like factors 6 and 7 and a determinant of fly development. *Gene*, **314**, 55–62.
- 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.
- Kriwacki, R.W., Schultz, S.C., Steitz, T.A. and Caradonna, J.P. (1992) Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor Sp1. *Proc. Natl Acad. Sci. USA*, **89**, 9759–9763.
- Kingsley, C. and Winoto, A. (1992) Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. *Mol. Cell Biol.*, **12**, 4251–4261.
- Horard, B., Tatout, C., Poux, S. and Pirrotta, V. (2000) Structure of a polycomb response element and *in vitro* binding of polycomb group complexes containing GAGA factor. *Mol. Cell Biol.*, **20**, 3187–3197.
- Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H. and Sakonju, S. (2001) The MCP silencer of the *Drosophila Abd-B* gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development*, **128**, 2163–2173.
- Gruzdeva, N., Kyrchanova, O., Parshikov, A., Kullyev, A. and Georgiev, P. (2005) The MCP element from the bithorax complex contains an insulator that is capable of pairwise interactions and can facilitate enhancer-promoter communication. *Mol. Cell Biol.*, **25**, 3682–3689.
- Hur, M.W., Laney, J.D., Jeon, S.H. and Biggin, M.D. (2002) Zeste maintains repression of Ubx transgenes: support for a new model of Polycomb repression. *Development*, **129**, 1339–1343.
- Wimmer, E.A., Jackle, H., Pfeifle, C. and Cohen, S.M. (1993) A *Drosophila* homologue of human Sp1 is a head-specific segmentation gene. *Nature*, **366**, 690–694.
- Schock, F., Purnell, B.A., Wimmer, E.A. and Jackle, H. (1999) Common and diverged functions of the *Drosophila* gene pair D-Sp1 and buttonhead. *Mech. Dev.*, **89**, 125–132.
- Bronner, G., Chu-LaGriff, Q., Doe, C.Q., Cohen, B., Weigel, D., Taubert, H. and Jaekle, H. (1994) Sp1/egr-like zinc-finger protein required for endoderm specification and germ-layer formation in *Drosophila*. *Nature*, **369**, 664–668.
- Myat, M.M. and Andrew, D.J. (2002) Epithelial tube morphology is determined by the polarized growth and delivery of apical membrane. *Cell*, **111**, 879–891.
- Moore, L.A., Broihier, H.T., Van Doren, M., Lunsford, L.B. and Lehmann, R. (1998) Identification of genes controlling germ cell migration and embryonic gonad formation in *Drosophila*. *Development*, **125**, 667–678.
- Abdelilah-Seyfried, S., Chan, Y.M., Zeng, C., Justice, N.J., Younger-Shepherd, S., Sharp, L.E., Barbel, S., Meadows, S.A., Jan, L.Y. and Jan, Y.N. (2000) A gain-of-function screen for genes that affect the development of the *Drosophila* adult external sensory organ. *Genetics*, **155**, 733–752.
- Biggin, M.D. and Tjian, R. (1988) Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. *Cell*, **53**, 699–711.
- Lehman, M., Siegmund, T., Lintermann, K.G. and Korge, G. (1998) The pipsqueak protein of *Drosophila melanogaster* binds to GAGA sequences through a novel DNA-binding domain. *J. Biol. Chem.*, **273**, 28504–28509.
- Huang, D.H., Chang, Y.L., Yang, C.C., Pan, I.C. and King, B. (2002) Pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. *Mol. Cell Biol.*, **22**, 6261–6271.
- Huang, D.H. and Chang, Y.L. (2004) Isolation and characterization of CHRASCH, a polycomb-containing silencing complex. *Methods Enzymol.*, **377**, 267–282.
- Hodgson, J.W., Argiropoulos, B. and Brock, H.W. (2001) Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Mol. Cell Biol.*, **21**, 4528–4543.
- Lehmann, M. (2004) Anything else but GAGA: a nonhistone protein complex reshapes chromatin structure. *Trends Genet.*, **20**, 15–22.
- Hagstrom, K., Muller, M. and Schedl, P. (1997) A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the *Drosophila* bithorax complex. *Genetics*, **146**, 1365–1380.
- Mahmoudi, T., Zuijderduijn, L.M., Mohd-Sarip, A. and Verrijzer, C.P. (2003) GAGA facilitates binding of Pleiohomeotic to a chromatinized Polycomb response element. *Nucleic Acids Res.*, **31**, 4147–4156.
- Schwendemann, A. and Lehmann, M. (2002) Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. *Proc. Natl Acad. Sci. USA*, **99**, 12883–12888.



46. Dejardin, J. and Cavalli, G. (2004) Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. *EMBO J.*, **23**, 857–868.
47. Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.L. and Kassis, J.A. (1998) The *Drosophila* Polycomb group gene *pleiohomeotic* encodes a sequence-specific DNA binding protein with homology to the multifunctional transcription factor YY1. *Mol. Cell*, **1**, 1057–1064.
48. Brown, J.L., Fritsch, C., Mueller, J. and Kassis, J.A. (2003) The *Drosophila* *pho-like* gene encodes a YY1-related DNA binding protein that is redundant with *pleiohomeotic* in homeotic gene silencing. *Development*, **130**, 285–294.
49. Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A. and Jones, R.S. (2004) Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell*, **14**, 637–646.
50. Mohd-Sarip, A., Venturini, F., Chalkley, G.E. and Verrijzer, C.P. (2002) Pleiohomeotic can link Polycomb to DNA and mediate transcriptional repression. *Mol. Cell. Biol.*, **22**, 7473–7483.
51. Lee, J.S., Galvin, K.M. and Shi, Y. (1993) Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl Acad. Sci. USA*, **90**, 6145–6149.
52. Seto, E., Lewis, B. and Shenk, T. (1993) Interaction between transcription factors Sp1 and YY1. *Nature*, **365**, 462–464.
53. Mishra, R.K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., Schweinsberg, S.E. and Schedl, P. (2001) The *iab-7* polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. *Mol. Cell Biol.*, **21**, 1311–1318.
54. Barges, S., Mihaly, J., Galloni, M., Hagstrom, K., Muller, M., Shanower, G., Schedl, P., Gyurkovics, H. and Karch, F. (2000) The Fab-8 boundary defines the distal limit of the bithorax complex *iab-7* domain and insulates *iab-7* from initiation elements and a PRE in the adjacent *iab-8* domain. *Development*, **127**, 779–790.
55. Fujioka, M., Emi-Sarker, Y., Yusibova, G.L., Goto, T. and Jaynes, J.B. (1999) Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development*, **126**, 2527–2538.
56. Kapoun, A.M. and Kaufman, T.C. (1995) Regulatory regions of the homeotic gene *proboscipedia* are sensitive to chromosomal pairing. *Genetics*, **140**, 643–658.
57. Muller, M., Hagstrom, K., Gyurkovics, H., Pirrotta, V. and Schedl, P. (1999) The *mcp* element from the *Drosophila melanogaster* bithorax complex mediates long-distance regulatory interactions. *Genetics*, **153**, 1333–1356.