

Molecular characterisation of the *Polycomblike* gene of *Drosophila melanogaster*, a *trans*-acting negative regulator of homeotic gene expression

Andrew Lonie¹, Richard D'Andrea^{1,*}, Renato Paro² and Robert Saint^{1,†}

¹Department of Biochemistry, University of Adelaide, South Australia 5005, Australia

²University of Heidelberg, Heidelberg, Germany

*Present address: Hanson Centre for Cancer Research, PO Box 14, Rundle Mall, Adelaide SA 5000, Australia

†Author for correspondence at present address: Department of Genetics, University of Adelaide, Adelaide, South Australia 5005, Australia

SUMMARY

The *Polycomblike* gene of *Drosophila melanogaster*, a member of the Polycomb Group of genes, is required for the correct spatial expression of the homeotic genes of the Antennapedia and Bithorax Complexes. Mutations in Polycomb Group genes result in ectopic homeotic gene expression, indicating that Polycomb Group proteins maintain the transcriptional repression of specific homeotic genes in specific tissues during development. We report here the isolation and molecular characterisation of the *Polycomblike* gene. The *Polycomblike* transcript encodes an 857 amino acid protein with no significant homology to other proteins. Antibodies raised against the product of this open reading frame were used to show that the Polycomb-

like protein is found in all nuclei during embryonic development. Antibody staining also revealed that the Polycomblike protein is found on larval salivary gland polytene chromosomes at about 100 specific loci, the same loci to which the Polycomb and polyhomeotic proteins, two other Polycomb Group proteins, are found. These data add further support for a model in which Polycomb Group proteins form multimeric protein complexes at specific chromosomal loci to repress transcription at those loci.

Key words: Polycomb group, *Polycomblike*, repression, homeotic regulation, polytene chromosomes, *Drosophila*

INTRODUCTION

In the insect *Drosophila melanogaster*, morphological diversity of the body segments arises during embryogenesis by selective expression of homeotic genes of the Antennapedia (ANT-C) and Bithorax (BX-C) Complexes (see Ingham, 1988 for a review). Spatial misexpression of the homeotic selector genes can result in inappropriate development of body segments, leading to the formation of structures characteristic of other segments. Clonal analysis has demonstrated that correct homeotic gene expression patterns must be maintained throughout most of *Drosophila* development to yield the normal adult segmental structures (Morata and Garcia-Bellido, 1976).

A number of *trans*-regulatory factors are involved in establishment and maintenance of the pattern of homeotic gene expression. One group of unlinked genes, known as the Polycomb group (PcG), appear to regulate homeotic expression by acting as transcriptional repressors of the BX-C and ANT-C. Mutations of PcG genes result in ectopic expression of homeotic genes within the BX-C and ANT-C, generating homeotic transformations of body segments (Jürgens, 1985; Wedeen et al., 1986; McKeon and Brock, 1991).

The PcG is composed of at least twelve genes. While all known members exhibit mutant phenotypes consistent with

misexpression of the ANT-C and BX-C, double and triple mutants exhibit, in general, much stronger transformations than single mutants (Jürgens, 1985). It is noteworthy that lesions in one member of the PcG, *Polycomb* (*Pc*), cause transformations that are as strong as those normally observed in double mutants. Differences in PcG mutant phenotypes may reflect differing maternal contributions within the PcG (Breen and Duncan, 1986). Various loci in the group exhibit pleiotropic effects such as segmentation defects and interactions with the *zeste* locus, indicating that genes of the PcG have more general roles than regulation of the ANT-C and BX-C. (Ingham, 1984; Breen and Duncan, 1986; Dura and Ingham, 1988; Smouse et al., 1988; Adler et al., 1991; Moazed and O'Farrell, 1992).

Several lines of evidence have led to suggestions that the products of the PcG genes form multimeric complexes at specific chromosomal locations. Two of the PcG proteins, Polycomb (*Pc*) and polyhomeotic (*ph*), have been shown to be localized to the same set of polytene chromosome loci, including those of the ANT-C and BX-C (Zink and Paro, 1989; DeCamillis et al., 1992). The protein encoded by another member of the group, *Posterior sex combs* (*Psc*), appears to localize to a set of sites on polytene chromosomes that overlaps the set of *Pc/ph* sites (Martin and Adler, 1993; Rastelli et al., 1993), although co-immunolocalisation has yet to be reported. Anti-*Pc* antibodies immunoprecipitate formaldehyde cross-

linked chromatin from within the BX-C (Orlando and Paro, 1993). Localisation of binding to *Antennapedia* sequences was confirmed by showing that the Pc protein binds to insertion sites of transgenes containing PcG-responsive *Antennapedia* regulatory regions (Zink et al., 1991). Significantly, this binding is sensitive to the position of insertion of the transgene and binding is not observed in cases where ectopic expression of the transgene is observed. Finally, the Pc and ph proteins were shown to coprecipitate from embryonic nuclear extracts (Franke et al., 1992).

Circumstantial evidence suggests that the transcriptional repression is mediated through regional changes in chromatin structure. The Pc protein has significant homology to the HP-1 heterochromatin-specific protein in a region termed the 'chromodomain'. HP-1 is encoded by the *Su(var)205* locus (Paro and Hogness, 1991). Mutations at this locus suppress an independent mutant phenotype termed Position Effect Variegation (PEV). PEV occurs as a result of the translocation of a euchromatic gene next to heterochromatic sequences. PEV is the heritable inactivation of the expression of such a translocated gene in particular somatic cells during development. It is thus likely to be caused by a property of the heterochromatin. According to one model, for example, the heterochromatin in a subset of cells spreads to include the translocated gene, inactivating it (see Henikoff, 1990 for a review). Various loci of the PcG are able to enhance or repress this effect in a gene-dosage-dependent manner (Sinclair, Clegg, Grigliatti and Brock, personal communication), further implicating the PcG in chromatin organisation.

Despite the similarities between PcG mutant phenotypes, the molecular and functional properties of PcG proteins are diverse. Of the four PcG genes whose molecular characterisation has been reported, no two share significant homology, although several have identifiable domains. As described above, the $44 \times 10^3 M_r$ Pc protein contains a conserved 37 amino-acid region, termed the 'chromodomain'. The *Psc* gene encodes a putative $175 \times 10^3 M_r$ protein containing a 200 amino-acid domain found also in the murine oncogene *bmi-1* (Brunk et al., 1991). This domain is postulated to bind DNA, although the DNA-binding properties of the *Psc* protein are yet to be reported. The *ph* gene encodes a $169 \times 10^3 M_r$ protein containing a single putative zinc finger and glutamine-rich repeats (DeCamillis et al., 1992). Finally, the *Enhancer of zeste* (*E(z)*), also known as *polycombeotic*, *pco* gene encodes an $87 \times 10^3 M_r$ protein with sequence similarity to the *Drosophila* trithorax and human ALL-1/Hrx proteins (Jones and Gelbart, 1993).

Many members of the PcG remain to be characterized. One such member, the *Polycomblike* (*Pcl*) gene, was assigned and named on the basis of the similarity of its mutant phenotype to the *Polycomb* gene (Duncan, 1982). We describe here the isolation and characterisation of the *Pcl* gene and show that it encodes a putative 857 amino acid protein that is localized to polytene chromosomes in an identical pattern to that found for both the Pc and ph proteins, further supporting the proposal that the PcG proteins form a multimeric complex that represses the transcription of homeotic genes during development.

MATERIALS AND METHODS

General techniques

Routine methods including screening of libraries, cloning, northern

and Southern analysis are described in Sambrook et al. (1989). DNA sequencing was performed using the Pharmacia Nested Deletion and Sequenase kits.

Preparation of antisera and immunolocalisation of proteins on polytene chromosomes

Fusion proteins were prepared using the pGEX glutathione-s-transferase bacterial expression system (Amrad). A 1.3 k bp *XmnI* fragment of the *Pcl* cDNA (see Fig. 3A) was cloned into pGEX1 and transformed into DH5 α bacteria. Expression of the glutathione-s-transferase-Pcl fusion protein was induced by the addition of IPTG. The bacteria were harvested, sonicated until lysis was complete and the insoluble fraction was collected by centrifugation. (The majority of the fusion protein was found to be in the insoluble fraction). This fraction was resuspended in 1% Triton X-100, 1% Tween-20 and subjected to SDS-PAGE. The fusion protein was excised from the gel, mixed with an adjuvant and used to immunise rabbits. After three boost injections, blood was collected and serum was tested for Pcl-binding activity on western blots of Pcl fusion proteins. The antiserum was further purified by passage successively over two affinity columns, the first containing whole bacterial lysate (covalently bound to CNBr-Sepharose) from which the eluate was collected and the second containing Pcl fusion protein from which the bound, Pcl-specific antibodies were eluted and used for immunohistochemistry. Immunostaining of chromosomes followed the procedure of Franke et al. (1992).

In situ hybridisation

Whole-mount in situ hybridisations were performed with staged Canton-S embryos according to the protocol of Tautz and Pfeifle (1989), using digoxigenin probes (Boehringer Mannheim) prepared by random priming with $10 \times$ the normal concentration of oligonucleotides (C. Oh and B. Edgar, unpublished modifications of Feinberg and Vogelstein, 1983).

RESULTS

Isolation of the *Polycomblike* gene

The *Pcl* gene had previously been localised to cytological region 55A on the right arm of chromosome 2. Characterisation of deficiencies within this region revealed the *Pcl* gene to be genetically close to the maternal effect gene, *staufen* (*stau*) (R. Tearle, personal communication). A cosmid (cos8.1), isolated as part of the cloning and characterisation of *stau* was found to rescue the *Pcl* mutant phenotype (D. St. Johnston, personal communication).

Characterisation of the region contained within the cos8.1 genomic clone revealed the presence of three transcription units in addition to the *stau* transcription unit (summarized in Fig. 1). Two of these correspond to previously characterized genes. The one closest to *stau* was found by Southern blot hybridisation to correspond to the extensively characterized HSF gene (Clos et al., 1990) while the one furthest from *stau* was found by sequence analysis to correspond to the Poly(A)-binding Protein (*PABP*) gene (Lefrere et al., 1990). Both of these transcripts were considered unlikely to correspond to *Pcl*. A third previously undescribed transcription unit, represented by cDNA clone AL15, lay between the HSF and PABP genes.

To determine which transcription unit corresponds to the *Pcl* gene, P-element-induced insertion alleles of *Pcl* were generated by mobilising a P-element inserted in *stau* (St. Johnston et al., 1991). Two such insertion alleles, termed *Pcl^{P1}*

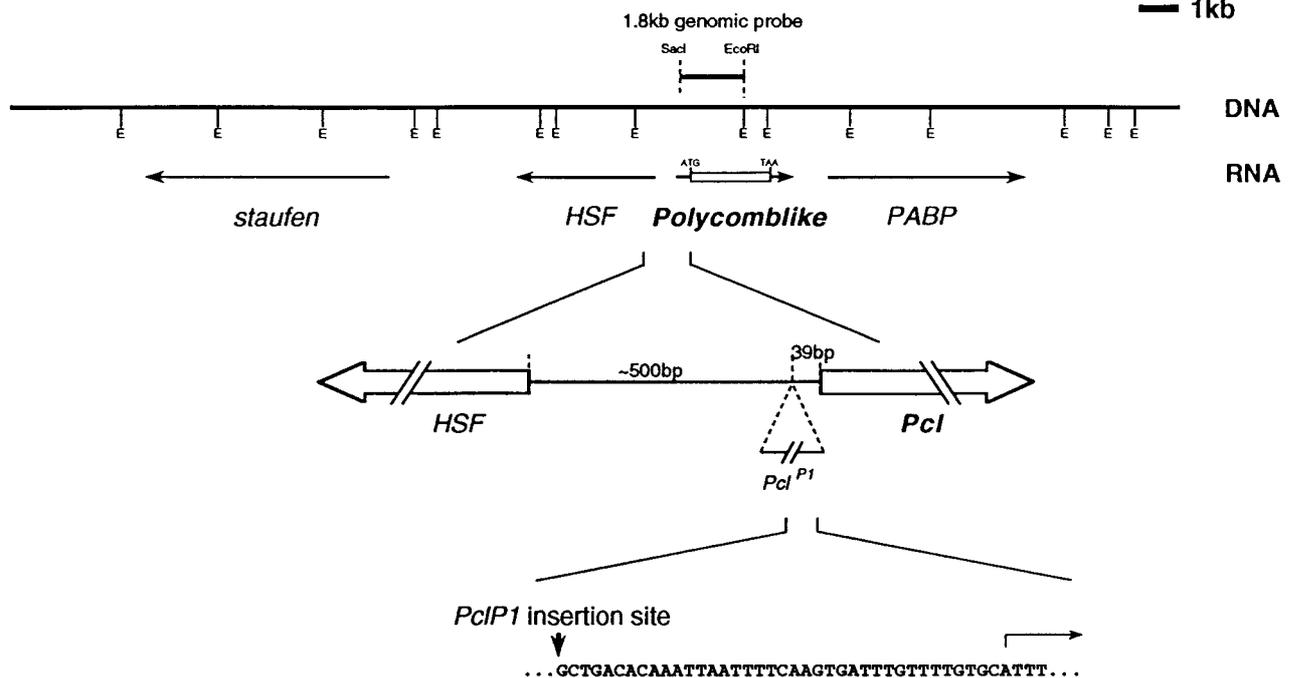


Fig. 1. (A) Positions of *stau*, *Pcl*, *HSF* and *PABP* genes in genomic DNA contained within the *cos1* genomic clone. The 1.8 kb *EcoRI-SacI* genomic fragment was used to screen a 0-4 hour cDNA library, from which the *Pcl* cDNA clones were isolated. Note the position of the *Pcl*^{P1} insertion site: the P-element causing this allele has inserted 39 bp upstream of the genomic equivalent of the 5' end of the longest *Pcl* cDNA clones. A possible TATA box is located 24 bp upstream of the 5' end of the cDNA, suggesting that the cDNA may be full length.

and *Pcl*^{P2}, were generated. Southern blot analysis showed both insertions to lie adjacent to the 5' end of the AL15 cDNA (results not shown). DNA covering the site of insertion of the P-element in the *Pcl*^{P1} allele was isolated and its sequence determined. The P-element was found to have inserted 39 nucleotides upstream of the 5' end of the 3.9 kb AL15 cDNA (Fig. 1 and see below) and approximately 500 base pairs from the *HSF* gene. A possible TATA box exists 12 bp downstream of the P-element insertion site. P-element-induced mutations frequently involve insertion very close to the transcription start site of the mutated gene (Tsubota et al., 1985; Chia et al., 1986; Searles and Volker, 1986; Kelley et al., 1987; Roiha et al., 1988). We conclude that the transcription unit defined by cDNA clone AL15 corresponds to the *Pcl* locus.

Characterisation of the *Pcl* transcript

The longest *Pcl* cDNA clone isolated was 3.9 kb in length (Fig. 1). Shorter clones appeared to correspond to incomplete cDNAs, as they formed a nested set with a common origin at the poly(A) tail. The entire cDNA was used as a probe to determine the size of the *Pcl* transcript on a northern blot (Fig. 2). Only one species of transcript was detected, corresponding in size to the largest cDNA that was isolated, which suggests that AL15 is a full-length cDNA. Sequence analysis revealed the presence of one major open reading frame (ORF), capable of encoding an 857 amino acid protein (Fig. 3). Two overlapping consensus start sequences (Cavener, 1987) were associated with two consecutive, in frame methionine codons at the 5' end of this ORF. We have assumed that the first ATG is that used in vivo, although the second ATG also forms a reasonable consensus start site. The sequence of the *Pcl* protein was compared to sequences in

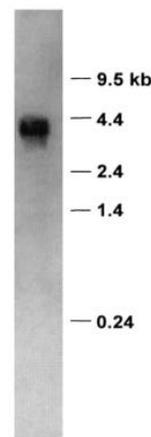


Fig. 2. Northern analysis of the *Pcl* message using adult mRNA. A single transcript of approximately 3.9 kb, corresponding to the size of the longest species of cDNA isolated, was detected.

all available databases, but no significant homologies were identified. Thus, like all other known members of the PcG, the *Pcl* protein shares no homology with any other PcG protein. A consensus polyadenylation site is present 21 nucleotides 5' to the poly(A) tail in the cDNA (Fig. 3).

Digoxigenin-labelled DNA probes were prepared from *Pcl* cDNA clone AL15 and hybridised in situ to *Drosophila* whole-mount embryos. *Pcl* mRNA was found to be evenly distributed throughout all tissues at all stages of embryogenesis (data not shown).

Characterisation of the *Pcl* protein

To characterise the *Pcl* protein, rabbit antibodies were raised against the 857 amino acid product of the ORF encoded by the *Pcl* transcript. A 1.3 kb *XmnI* fragment of *Pcl* cDNA clone

2632 A. Lonie and others

1		AT	1893	<u>AAGATCACAAAACCAGCGGGAAATTTGTCGCCAGTTGCCCTACCATGCGGATAAG</u>
3	TTTCGCTGGTGTAAAGATGAACTGGGCAACACGCTTCAAAGCGGATAGCCG		482	K I T K P A G I C R Q L P Y H A D K
57	AATCAATAAGAGCAAAATGCTAATTTAATGACAGTTCGAAACACACTTAAAGG		1947	<u>CTCAGCTGGGATGAGAAGCATCGGGTCAACGAGAGCAGATCTACTGCTACTGC</u>
111	TGGGTAATTCGGCAACCGAACTAGAAAGTGCCTGGTGGCTGTGGAAAAGTAC		500	L S W D E K H R V N E E Q I Y C Y C
165	GAATGAGGAAGACACAGAGAATACAAAAGGGCCGCTCCAGTCCGTATGTGT		2001	<u>GGCAAACCGGGAAAATTCGATCACAATATGCTGCAGTCTGCAAAATGTCGGAAC</u>
219	AGCAAAGTGTGCATGTATGTTCGGCGGGCAACAACAATTCGACGAAAAAGG		518	G K P G K F D H N M L Q C C K C R N
273	CGCCAGATTAATGAGCTCGTGTAAATCAGCGGAAACAATCGCAAAATCAATGA		2055	<u>TGGTTCACACCCAATGCATGCAGAACTTCAAGAAGAAGCTGCTGCGCGCGAT</u>
327	AAAAACGGCCCGGAAAAGCGGCAAAATCGAAGTGAAGATCCATATTTTTCACA		536	W F H T Q C M Q N F K K K L L R G D
381	AAAGACTGCGCACTGAAAAAGACTGCGCACTGAAAAATCAACCCATCTAGCAA		2109	ATGTTCTTTGTTTTCTGCTGCAGGCTGCAACAATGGCATAGAAATCGTTCGT
435	GCACACAGCCGCAATGATGAACAACCATTTTCACTTGCACACGATCATCCG		554	M F F V F C C T V C N N G I E F V R
1	M M N N H F H L Q H D H P		2163	CGTATGCAAAATCGAAATGGTGGAGCTGCTCCACATGCATATACAATTCGCA
489	CCGAGAAAGTGGACATCCGTTTATGCAACAGCCGAGTACCGCTGTCCCATCA		572	R M Q I E W V D V L H I A L L Y N L R
14	P Q N V A H P F M Q Q P S T A V P S		2217	AAGCATCAGCACCAGAAGTACCATCACTGCTAAATGACATTTGGCCCTTCATA
543	GCTCCTCCGGCCACCTACGGCTACTTAGCAGCCAGCTGGCCAGCAGCCGCG		590	K H Q H Q K Y H H L L N D I W P F I
32	A P P A T Y G Y L A Q P A G Q P A G Q P Q		2271	CTCAGCAGCGCCACCAGCTGCCTATCTGCGAGAAGTGGCGCACACTGCCGAA
597	TGGATGACGACCACCTACCAGTCTGCCGCCAGTGTAGGACCTGCCACGGTG		608	L E Q R H Q L P I C E K W R T L P E
50	W M T T T Y Q I L P P S V G P A T V		2325	ACGGCACTTATGGAGCTCTCAAGCAAACCCCTTAAGGATTACTCTGACAGATT
651	GCCAAAGCATACTATGCCACCACTGGGCGCGAGCCAGCATCCACACATCCC		626	T A L M E R L K Q T L K D Y S D R F
68	A K R Y Y A T T G P Q T T H P T H P		2379	GTTTGGCGAGGAGTTTAAAGGGCTCCGGCCCTCTATGCAGTGCAGACAGT
705	AGCACCATCCAGATCAGCAACAATTTCTGCTCAGCAATCAACGCCACCGAAGCAA		644	V C G R E F K R A P A F Y A L R H S
86	S T I Q I T N N F A Q Q S T P P K Q		2433	GGCCGCCACACATTCGCAAGGATTCCTTGCAGCCCAAGGAGGAGCTGTCTGAT
759	CAGGGCGCAACAAGCTGCAGCCCATTCAGGCCAACACATAGCAATCATCTCG		662	G P P H I P K V F L E P H E E L S D
104	Q A A T S C S P F K A N N I R I I S		2487	GAGCTGCTCGAAAAAGATTTAACTAATGCTCATGCCAGGAGCCTGATGAA
813	<u>ACGGCGCAAGTGTATACAGTCTGAATTAAGCCGCCAGGAAGCCGATCCACA</u>		680	E L L E K R F K L M L M P E E P D E
122	T A P S V Y S L N K P P Q E A H S T		2541	GGTGCCAACGAGCTCCGAAAAGAGTTCGCAAGATGTGTATGAGTTAATACA
867	<u>TATGCTCCCGTTCAGTCTACTACTTGGCCAGTGGCGGTGGCCAACTGCGGGA</u>		698	G A N E L P K R V P K D V Y E F N T
140	Y A P V Q S Y Y L P S G G G Q T A G		2595	GATGAAGATGACCCAGTGCAGACGAGGACGAAATACCCATCAACAGATT
921	<u>CBAATCAATCTACTGGCGCCAGCGGAACTGGCAAGCAACTGCACCGCCGCGG</u>		716	D E D D P V E T S E D E I P I K Q I
158	Q I N L L A A S G T G K Q L Q P P P		2649	ATTGAGAGGCCAAAAAGCAGCCGCTCAGAAGGCCGATAAGCATGATGAGTTG
975	<u>CTGGTGCCTGTGACCAATTCACCTCACCGCCCTCCACCGTCTGTGTAGACCGC</u>		734	I E K A K A K Q A A G A K A D K H D E L
176	L V P V T N S T S P P S T V V L D R		2703	CCACTGAAGCCGACCTCGCCGACGACAATGCCAACGATGGCGATCCCGGCAA
1029	<u>ATAACATTTGCATCAATAACCACTATACGGGAGCGCCCACTTCGCTAAGTTCA</u>		752	P L K P D L A D D N A N D G D P G K
194	I N I C I N N H Y T E T P T S L S S		2757	CTACCAGCTCCCATTCACCTCTGCTGGACGCAAAATGACAGTCCGAAAGCGAAG
1083	<u>TCCGCTTACCAGCGCCCAACCACTTCGGCCATCATCCGGCCATCCAGCACAAG</u>		770	L P A P I P P L L D A N S S R K R K
212	S L T T A Q Q P S P I I P A I Q H K		2811	GCTTTCGATGTCCAACGCTATGACAACAGCCGCAATCACTGTGATCTATCA
1137	<u>GCCATCCTGCCCTCATCGACAGCAGCAGCGGGATAGCAGCAGCTGCAGCAGC</u>		788	A F R L S K R Y D N S R N H C D L S
230	A I L P L I D S S T A D S S S C S S		2865	TCCGAGAGAATTCAAGCAGCAGTCCGGGAACCAAGTTCGTTGGACCTCATATAC
1191	<u>AGTTCTGTATCCAGCAGCAGTATATAGTGTACCGGACCACTCAGCAGCAGTT</u>		806	S E R I Q A A V G E P V R W T S S Y
248	S S V S S S S Y S G T A T T S A A V		2919	CACCCGGTCAACTTTTTGGGACGCAACAATCCATTCCTTATGGCCATCCCA
1245	<u>GTGATGTGGATGAGCCGGATTTACGACCAACAAGCCACAGACACCGCAACT</u>		824	H R R S T F W D A T I H S L W P L R
266	V I V D E P D S T T T P Q T P P T		2973	AAAAGCGTCCGAGGACGAAGTATTCTGTGGAACTGGAGTGGAGTAAATG
1299	<u>ACGGCCGAGGCGCATGAGTTACCCGGTAAAGTTCCGCCCTCACCCTCCGCTACTA</u>		842	K R R R R D E V F L W E L E S E *
284	T P E A M S S P G K S P S P S P P P L L		3027	GCATCATAAACAGCATTTTCAAGCTTAAAGGCACTCCAAAGGAGCAGCCGCGCA
1353	<u>GCCACGCAATCGCTCTCABAGCGCTCAACTCAATGAAGCAAGCTTCAAGAGC</u>		3081	TGGTGGCCACCATCAAGCGAAGACTGAGTGCAAAGACATTACGATAGGGCCCA
302	A T Q S L L K G V N S M K P S F K T		3135	ATCAGGAAGTGGCGAGCAGCAGCACTCCGGCCCTGACCAAGGCAATGAGGTCA
1407	<u>GTGGAAGCAGCACCAGCAAGCCGCAACCACTCCATCACCACCAACCCAGCCG</u>		3189	TCAGCACCACCACCAATCCCATACCCAGCCACTACCTTCCTATCTATGCCA
320	V E A A P P T P P T P P S P P P P P		3243	AAGATCTGCAGCCACCCGCGCCGCAATGGAAAGCCGACACATGGGGCCCTGT
1461	<u>CCGGCACCACAGTTCGCGCTCCCTGGCAGCGGTACAGTATGCGCTTCAGGAG</u>		3297	TGCGCCAGCGACCAAGAAGCAGTACCAGCCAAAGCCGCGGAACTCCACCA
338	P A P P V A A P S P A V T Y A L Q E		3351	GCTCGACGGCAACCAGCAGCAGCAACGGCATTGGAGCCCAAGGACATTCCA
1515	<u>GAGCTTCTCATCAATGCAACGATGGCAGATTTTACCTGGGACCAATATTGAC</u>		3405	TGCTGGACTTGAAGCAGTCCGTTGAACAAGTATTGGTGGCCTATGAATCGCA
356	D V F I K C N D G R F Y L G T I I D		3459	TAGATGCTGGTGGCCCTTGGCATTCCGGCCAAAGCCGCGATGGAAATGGAC
1569	<u>CAGACAAGCGATCAATATCTTATCCGCTTTGATGACCAGTCCGAGCAATGGTGT</u>		3513	AGGTCCAGTACCTGGTGGAGTGGGCGGAGATACAGGCAACGGCGATGGAT
374	Q T S D Q Y L I R F D D Q S E Q W C		3567	TGCTTGGAAATGAACCCCTTGGCATACCGACTCATTATGAATTTTATCTT
1623	<u>GAGCCGATAAGCTGCGCAAGCTGGGTGGCGGTAGCTCGATTACCCTGGCGGA</u>		3621	CATTTCCATTATCATTTCTTTTCGTTATATGTATTAATGAACGCTGCCAACTA
392	E P D K L R K L G G G S S I T A G G		3675	CGGGCCTCTTATTTGTTAATGTTTAAAGTATTATTACCAACCCACTGCCAAAA
1677	<u>GGCGGTGCCTCTACACCGAATCCACGAATACCTCCCCAGTGGACCCATGTGC</u>		3729	TAGAAAAAGATTGTTCAAATATTGGTTCAGCATTAAGCGATAAAATGTTAAA
410	G G A S T T E S T N T S P S G P M C		3783	TTGTAATAAAAAAAAAA
1731	<u>GTGGCTGCAAGCGATCGGATATCGAAGACGTGGTGGAAATTTGCGAGCGCTGC</u>			
428	V A C K R S D I E D V V E I C E R C			
1785	<u>GGACCTGGCTATCATCTGTTGACCCCTTGAATTTGTTACTGGCAGCGGAATT</u>			
446	G R G Y H R G C T V E I V T G S G I			
1839	<u>TGGAGTTGCAACCGCTGCGCTAAACCCATGAAAATGCAACAGCCTGTCAGCCAT</u>			
464	W S C K R C A K P M K M Q Q P V S H			

Fig. 3. Nucleotide sequence of the *Pcl* cDNA and predicted amino acid sequence of the protein product. A consensus polyadenylation sequence (bold) lies 21 bp 5' to the poly(A) tail of the cDNA. A 1.3 kb *XmnI* fragment, used to generate a fusion protein for antibody production, is underlined.

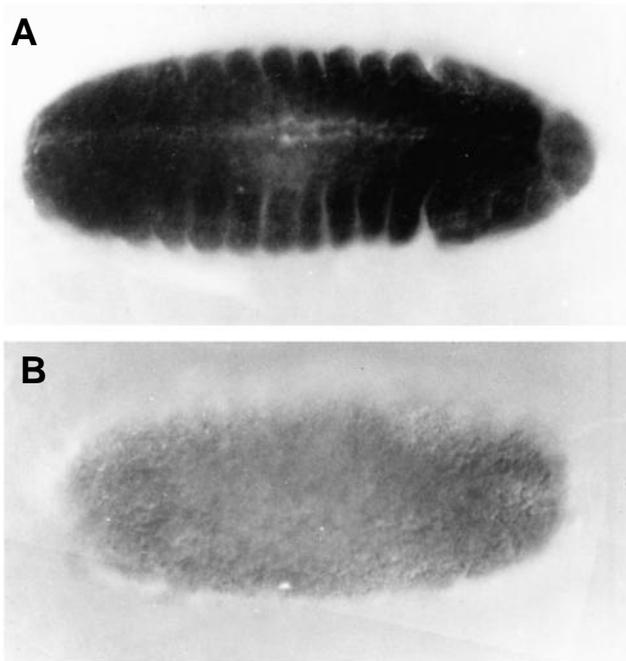


Fig. 4. Staining of wild-type and homozygous *Df(2R)PC4* (*Pcl* deficient) germ band retracted embryos. Wild-type embryos (A) were discriminated from homozygous mutant embryos (B) by the homozygous *three rows* cell proliferation phenotype also associated with *Df(2R)PC4* (see text). The poor morphology of the embryo shown in B is a consequence of the *three rows* mutation. No staining is seen in embryos deficient for *Pcl* at this developmental stage.

AL15 (see Fig. 3) was ligated to the 3' end of the glutathione-S-transferase (GST) gene in the vector pGEX2. The GST-*Pcl* fusion protein was produced in a bacterial expression system, isolated on SDS-polyacrylamide gels and used to immunise rabbits. Sera collected from the animals was affinity purified on CNBr-Sepharose columns to which fusion protein had been bound. Probing of western blots of the GST-*Pcl* fusion protein demonstrated that the antibodies were specific for the *Pcl* portion of the fusion protein (data not shown). Subsequently, we used the purified antisera to probe embryos that were homozygous for the *Df(2R)PC4* deletion. The homozygous deficient embryos were also missing the closely linked *three rows* gene, a gene required for chromosome segregation and cytokinesis, but not for any other aspect of cell cycle progression (D'Andrea et al., 1993), so that 12-15 hour homozygous *Pcl* mutant embryos were identified by virtue of their *three rows* phenotype. In contrast to wild-type and *Df(2R)PC4* heterozygous embryos, the homozygous mutant embryos showed no staining with the affinity purified anti-*Pcl* antibodies, indicating that the antibodies are specific for *Pcl* in vivo (Fig. 4). The staining of all cleavage, blastoderm and gastrulating embryos derived from heterozygous *Df(2R)PC4* parents indicated the presence of a maternal *Pcl* component in the homozygous deficient embryos, consistent with the genetic analysis of *Pcl* mutants (Breen and Duncan, 1986).

***Pcl* is a nuclear protein that binds to specific loci on polytene chromosomes**

The distribution of *Pcl* in embryos was examined by incubating fixed whole-mount embryos with the anti-*Pcl* primary and

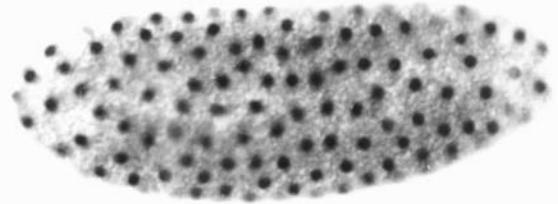


Fig. 5. Localisation of the *Pcl* protein. Embryos aged 0-24 hours were fixed, stained with anti-*Pcl* antibodies and the bound antibodies detected using horseradish peroxidase-conjugated secondary antibodies. *Pcl* protein was found in nuclei throughout the developing embryo. The example shown here is of a cleavage stage embryo.

horseradish peroxidase linked secondary antibodies (Fig. 5). By comparing the *Pcl* staining pattern with that of the DNA-specific fluorochrome, DAPI, *Pcl* was found to be localised to the nucleus (data not shown). The protein was found in all nuclei throughout development (Fig. 5 and data not shown), consistent with the distribution of *Pcl* mRNA.

Two other PcG-encoded proteins, *Pc* and *ph*, are known to bind to an identical set of approximately 100 sites on salivary gland polytene chromosomes (Franke et al., 1992). We used double-labelled antibody stainings to demonstrate that the *Pcl* protein binds to an identical set of sites as those bound by the *Pc* protein (Fig. 6 and results not shown). As found for the *Pc* protein, two of the most intensely stained bands corresponded to the ANT-C and BX-C at cytological locations 84A-B and 89E, respectively (results not shown).

Segments of DNA derived from the *Antp* regulatory region and transposed to new chromosomal locations are capable of binding *Pc* (Zink et al., 1991), as demonstrated by antibody staining polytene chromosomes carrying such transposed fragments. We used this system to show that *Pcl* also binds to the same transposed DNA segments derived from the *Antp* regulatory regions. For example, a new *Pcl*-binding site in polytene chromosomes was found to be present on the third chromosome at the site of insertion of the pAPT2.3 construct (Fig. 7). This construct contains approximately 12.6 kb of DNA from the region of the *Antp* P2 promoter.

DISCUSSION

The Polycomb Group (PcG) of genes are defined on the basis of a common homeotic mutant phenotype that results from derepression of the genes of the ANT-C and BX-C. As such, they may represent a unique mechanism for the negative regulation of eukaryotic genes. The extraordinary conservation of the homeotic gene clusters (reviewed in McGinnis and Krumlauf, 1992) and, possibly, of their regulation (Awgulewitsch and Jacobs, 1992; Malicki et al., 1992) makes it likely that the PcG-mediated repression mechanism will be found to operate throughout the metazoa. The finding that proteins encoded by the *Pc*, *Psc* and *E(z)* genes share sequence similarity with mammalian proteins supports this speculation (Singh et al., 1991; Brunk et al., 1991; Jones and Gelbart, 1993).

We report here the molecular isolation and characterisation of the *Polycomblike* (*Pcl*) gene, a member of the PcG. The gene

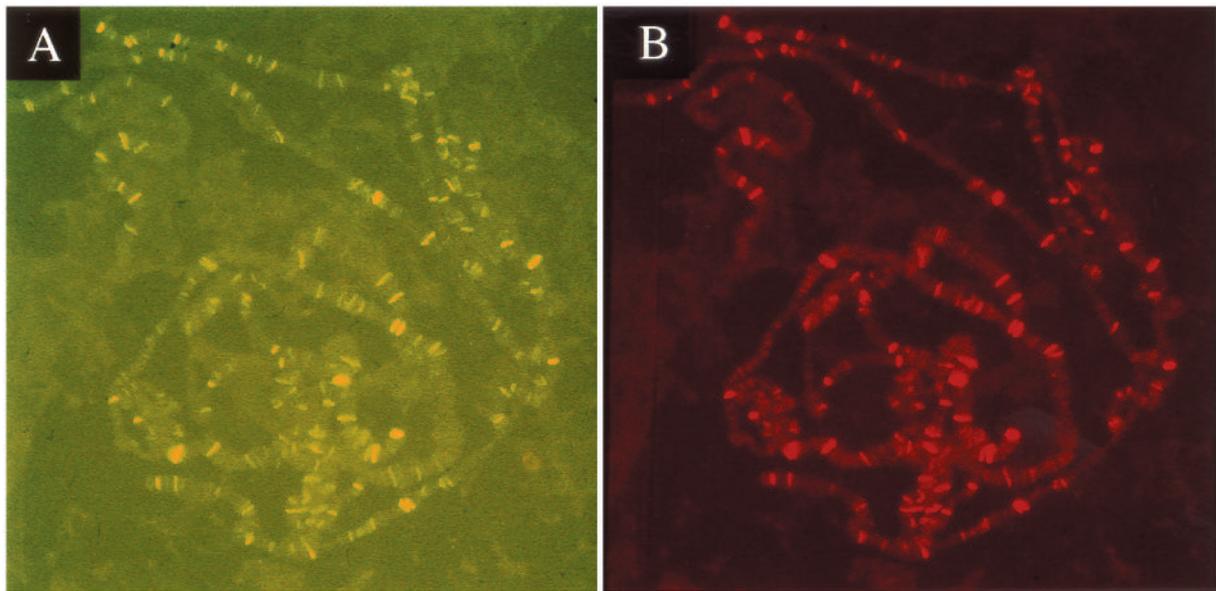


Fig. 6. Localisation of Pcl and Pc proteins on polytene chromosomes using double immunofluorescence staining. (A) Staining of chromosomes with polyclonal rabbit anti-Pcl antibodies and DTAF-conjugated anti-rabbit antibodies. (B) Staining of the same chromosomes with polyclonal mouse anti-Pc antibodies and rhodamine-conjugated anti-mouse antibodies. Distribution of the two proteins appears to be identical.

was identified by the molecular characterisation of transcripts from a cosmid that rescued the *Pcl* mutant phenotype and from the molecular analysis of a P insertion *Pcl* allele.

Sequence analysis of *Pcl* cDNA clones revealed the presence of a single large open reading frame that encodes an 857 amino acid protein. Exhaustive searching of databases yielded no clues to the function of this protein. It seems likely that at least one member of the PcG will encode a protein that exhibits sequence-specific DNA-binding activity, but Pcl does not possess any known DNA-binding motifs. Furthermore, the relatively weak phenotype of *Pcl* mutants in comparison to, for instance, *Pc* mutants, argues that Pcl is unlikely to provide sequence specificity for the PcG complex.

Although the PcG genes isolated to date exhibit no molecular homology with each other, the similarities in mutant phenotypes, i.e. derepression of the homeotic genes, argue that they are involved in the same process. The studies of the *Pcl* gene and its protein product reported here show that it shares certain characteristics with the other PcG genes reported to date. Pcl, like Pc, Psc and ph, is a nuclear protein that is expressed throughout embryonic development. This result is not surprising, as genetic analysis shows that *Pcl* function is required throughout development (Duncan, 1982). *Pcl* mutants exhibit both embryonic and adult phenotypes: *Pcl* null mutant homozygotes die during embryogenesis, while heterozygotes exhibit an adult phenotype in which the second and third legs of males are transformed to resemble the first leg. This generalised expression of PcG members suggests that the transcriptional repression of particular homeotic genes in a subset of tissues cannot be due to specific expression of PcG members in those tissues.

Another functional similarity between the PcG proteins analysed to date is that the Pcl, Pc and ph proteins bind to the same large number of loci on polytene chromosomes. The observation that Pcl represents a third member of the PcG family of proteins that co-localises to specific sites on polytene

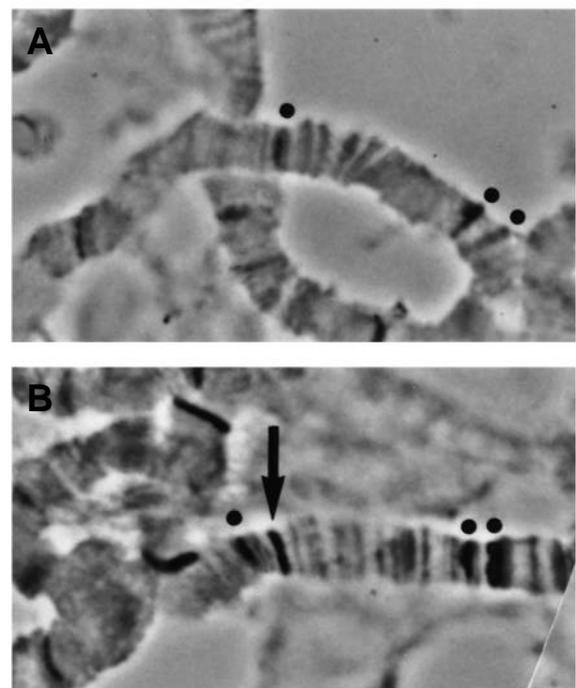


Fig. 7. In vivo binding of Pcl to a transgene containing *Antp* promoter sequences. (A) Immunohistochemical detection of Pcl protein on the tip of a wild-type 3rd chromosome. (B) Corresponding section of a chromosome containing a 12.6 kb portion of the *Antp* P2 promoter which has inserted at position 99B on chromosome 3 (Zink et al., 1991). Wild-type binding sites are marked by dots; a new Pcl-binding site, indicated by an arrow, coincides with the site of the transgene insertion.

chromosomes further strengthens the proposal that a heteromeric complex of PcG proteins binds to specific sites in chromosomes to repress particular genes (Paro, 1990).

While Pcl exhibits certain functional similarities with other PcG genes, it is also similar in having no molecular homology to any other member of the group. Thus while the different PcG genes are functionally related, the known PcG proteins share no molecular similarities. Each gene, therefore, appears to play a unique role in the repressing mechanism. Loss of any part of the PcG complex would be expected to lead to loss of function, so mutations in any one PcG gene would give a similar phenotype, despite its unique function within the complex. Phenotypes of PcG mutants are not, however, identical. This suggests that there are differences in the function of individual members, despite the overall functional similarity. Some of the differences in the zygotic null phenotypes very likely reflect differences in the provision and/or stability of the different maternal components. It is also possible, however, that particular PcG proteins exhibit subtle differences in function at different loci and in different cell types.

Our studies have not addressed the basis of specificity of binding of PcG proteins to specific sites on chromosomes. In contrast to the tissue specificity of repression of particular homeotic genes, Pcl, like the previously described Pc and ph proteins (Franke et al., 1992), is present in all nuclei examined. This suggests that the PcG-mediated repression of particular genes in some, but not all, tissues is not a function of the confinement of PcG expression to those tissues. Thus, repression of ANT-C and BX-C genes only in segments that do not actively express those genes during the establishment phase of homeotic gene expression, suggests that the postulated heteromeric PcG complex is able specifically to recognise a transcriptionally inactive gene. This recognition may occur at the level of chromatin structure, or may be the result of a PcG complex interacting with transcription factors that regulate ANT-C and BX-C expression. Understanding the basis of this recognition is likely to require further characterisation of the nature of the PcG genes and the molecular properties of the proteins they encode.

The authors would like to thank Carl Wu for unpublished information and for providing the HSF cDNA clone and Rick Tearle for providing unpublished information. A.L. was supported by an Australian Postgraduate Research Award and R.D'A by an Australian Research Council QEII Fellowship. This work was supported by the Australian Research Council.

REFERENCES

- Adler, P. N., Martin, E. C., Charlton, J. and Jones, K. (1991). Phenotypic consequences and genetic interactions of a null mutation in the *Drosophila* *Posterior sex combs* gene. *Dev. Genet.* **12**, 349-361.
- Awgulewitsch, A. and Jacobs, D. (1992). *Deformed* autoregulatory element from *Drosophila* functions in a conserved manner in transgenic mice. *Nature* **358**, 341-345.
- Breen, T. and Duncan, I. (1986). Maternal expression of genes that regulate the bithorax complex of *Drosophila melanogaster*. *Dev. Biol.* **118**, 442-56.
- Brunk, B. P., Martin, E. C. and Adler, P. N. (1991). *Drosophila* genes *Posterior Sex Combs* and *Suppressor two of zeste* encode proteins with homology to the murine *bmi-1* oncogene. *Nature* **353**, 351-353.
- Cavener, D. (1987). Comparison of the consensus flanking translational start sites in *Drosophila* and vertebrates. Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucleic Acids Res.* **15**, 1353-1361.
- Chia, W.G., Howes, M., Martin, Y.B., Meng, K., Moses, K. and Tsubota, S. (1986). Molecular analysis of the *yellow* locus of *Drosophila* *EMBO J.* **5**, 3597-3605.
- Clos, J., Westwood, J. T., Becker, P. B., Wilson, S., Lambert, K. and Wu, C. (1990). Molecular cloning and expression of a hexameric *Drosophila* Heat Shock Factor subject to negative regulation. *Cell* **63**, 1085-1097.
- D'Andrea, R., Lehner, C., John, U., Stratmann, R. and Saint, R. B. (1993). The *three rows* gene of *Drosophila melanogaster* encodes a novel protein required for chromosome separation during mitosis. *Molec. Biol. Cell* **4**, 1161-1174.
- DeCamillis, M., Cheng, N. S., Pierre, D. and Brock, H. W. (1992). The *polyhomeotic* gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome-binding sites with Polycomb. *Genes Dev.* **6**, 223-232.
- Duncan, I. M. (1982). *Polycomblike*: A gene that appears to be required for the normal expression of the *bithorax* and *Antennapedia* gene complexes of *Drosophila melanogaster*. *Genetics* **102**, 49-70.
- Dura, J.-M. and Ingham, P. (1988). Tissue- and stage-specific control of homeotic and segmentation gene expression in *Drosophila* embryos by the *polyhomeotic* gene. *Development* **103**, 733-741.
- Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6-13.
- Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H. W. and Paro, R. (1992). *Polycomb* and *polyhomeotic* are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* **11**, 2941-2950.
- Henikoff, S. (1990). Position-effect variegation after 60 years. *Trends Genet.* **6**, 422-426.
- Ingham, P. (1984). A gene that regulates the *bithorax* complex differentially in larval and adult cells of *Drosophila*. *Cell* **37**, 815-823.
- Ingham, P. W. (1988). The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* **335**, 25-34.
- Jones, R.S. and Gelbart, W.M. (1993). The *Drosophila* Polycomb-Group gene *Enhancer of zeste* contains a region with sequence similarity to *trithorax*. *Mol. Cell. Biol.* **13**, 6357-6366.
- Jurgens, G. (1985). A group of genes controlling the spatial expression of the *bithorax* complex in *Drosophila*. *Nature* **316**, 153-155.
- Kelley, M.R., Kidd, S., Berg, R.L. and Young, M.W. (1987). Restriction of P element insertions at the *Notch* locus of *Drosophila melanogaster*. *Mol. Cell. Biol.* **7**, 1545-1548.
- Lefrere, V. A., Vincent, A. and Amalric, F. (1990). *Drosophila melanogaster* poly(A)-binding protein: cDNA cloning reveals an unusually long 3'-untranslated region of the mRNA, also present in other eukaryotic species. *Gene* **96**, 219-225.
- Malicki, J., Cianetti, L., Peschle, C. and McGinnis, W. (1992). A human HOX4B regulatory element provides head specific expression in *Drosophila* embryos. *Nature* **358**, 345-347.
- Martin, E.C. and Adler, P.N. (1993). The *Polycomb* group gene *Posterior sex combs* encodes a chromosomal protein. *Development* **117**, 641-655.
- McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and Axial patterning. *Cell* **68**, 283-302.
- McKeon, J. and Brock, H. (1991). Interactions of the Polycomb group of genes with homeotic loci of *Drosophila*. *Roux's Arch. Dev. Biol.* **199**, 387-396.
- Moazed, D. and O'Farrell, P. H. (1992). Maintenance of the *engrailed* expression pattern by *Polycomb* group genes in *Drosophila*. *Development* **116**, 805-810.
- Morata, G. and Garcia-Bellido, A. (1976). Developmental analysis of some mutants of the *bithorax* system of *Drosophila*. *Wilhelm Roux's Arch. Dev. Biol.* **179**, 125-143.
- Orlando, V. and Paro, R. (1993). Mapping Polycomb-repressed domains in the Bithorax Complex using in vivo formaldehyde cross-linked chromatin. *Cell* **75**, 1187-1196.
- Paro, R. (1990). Imprinting a determined state into the chromatin of *Drosophila*. *Trends Genet.* **6**, 416-421.
- Paro, R. and Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **88**, 263-267.
- Rastelli, L., Chan, C.S. and Pirotta, V. (1993). Related chromosomal binding sites for *zeste*, suppressors of *zeste* and *Polycomb* group proteins in *Drosophila* and their dependence on *Enhancer of zeste* function. *EMBO J.* **12**, 1513-1522.
- Roiha, H., Rubin, G.M. and O'Hare, K. (1988). P element insertions and rearrangements at the *singed* locus of *Drosophila melanogaster*. *Genetics* **119**, 75-83.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Plainview, NY

- Searles, L.L. and Voelker, R.A.** (1986). Molecular characterization of the *Drosophila vermilion* locus and its suppressible alleles. *Proc. Natl. Acad. Sci. USA* **83**, 404-408.
- Singh, P.B., Miller, J.R., Pearce, J., Kothary, R., Burton, R.D., Paro, R., James, T.C. and Gaunt, S.J.** (1991). A sequence motif found in a *Drosophila* heterochromatin protein is conserved in animals and plants. *Nuc. Acids Res.* **19**, 789-794.
- Smouse, D., Goodman, C., Mahowald, A. and Perrimon, N.** (1988). *Polyhomeotic*: a gene required for the embryonic development of axon pathways in the central nervous system of *Drosophila*. *Genes Dev.* **2**, 830-842.
- St. Johnston, D., Beuchle, D. and Nusslein-Volhard, C.** (1991). *staufer*, a gene required to localise maternal RNAs in the *Drosophila* egg. *Cell* **66**, 51-63.
- Tautz, D. and Pfeifle, C.** (1989). A non-radioactive *in situ* hybridisation method for the localisation of specific RNAs in the *Drosophila* embryo reveals translational control of the segmentation gene *hunchback*. *Chromosoma* **94**, 81-85.
- Tsubota, S., Ashburner, M. and Schedl, P.** (1985) P element induced control mutations at the *r* gene of *Drosophila melanogaster*. *Mol. Cell. Biol.* **5**, 2567-2574.
- Wedeen, C., Harding, K. and Levine, M.** (1986). Spatial regulation of Antennapedia and bithorax gene expression by the Polycomb locus in *Drosophila*. *Cell* **44**, 739-48.
- Zink, B., Engstrom, Y., Gehring, W. J. and Paro, R.** (1991). Direct interaction of the Polycomb protein with *Antennapedia* regulatory sequences in polytene chromosomes of *Drosophila melanogaster*. *EMBO J.* **10**, 153-162.
- Zink, B. and Paro, R.** (1989). *In vivo* binding pattern of a trans-regulator of homeotic gene in *Drosophila melanogaster*. *Nature* **337**, 468-471.

(Accepted 20 June 1994)