Development 114, 921-929 (1992)
Printed in Great Britain © The Company of Biologists Limited 1992

The mouse has a *Polycomb*-like chromobox gene

JONATHAN J. H. PEARCE, PRIM B. SINGH and STEPHEN J. GAUNT

Department of Molecular Embryology, AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge, CB2 4AT, UK

Summary

The *Drosophila* gene *Polycomb* (*Pc*) has been implicated in the clonal inheritance of determined states and is a *trans*-regulator of the *Antennapedia*-like homeobox genes. *Pc* shares a region of homology (the chromobox) with the *Drosophila* gene Heterochromatin Protein 1 (HP1), a component of heterochromatin. The *Pc* chromobox has been used to isolate a mouse chromobox gene, M33, which encodes a predicted 519 amino acid protein. The M33 chromodomain is more similar to that in the Pc protein, than that in the HP1 protein. In addition to the chromodomain, the M33 and Pc proteins also share a region of homology at their C termini. The temporal and spatial expression patterns of M33 have been studied by in situ hybridization and northern analysis. During the final 10 days of embryonic development, M33 expression mirrors that of the cell-cycle-specific cyclin B gene. It is therefore suggested that the rate of cellular proliferation controls M33 expression. From comparisons of the characteristics of M33 with those of *Pc* it is proposed that M33 is a *Pc*-like chromobox gene. The roles of M33 and *Pc* in models of cellular memory are examined and implications of the memory models addressed.

Key words: determination, mouse, chromobox, M33, cyclin B, *Drosophila*, *Polycomb*, HP1.

Introduction

During the early stages of regulative development, embryos assign positional values to their parts. Wolpert has suggested, in his influential theory of positional information (Wolpert, 1969), that this specification is achieved by a morphogen concentration gradient. Mathematical studies have concluded that such gradients can only lay down a pattern over a distance of about one millimetre, or about 100 cell widths (Slack, 1991). This estimate corresponds very well with the actual size of embryos undergoing positional specification: the anteroposterior (A/P) axis of the *Drosophila* embryo at cellular blastoderm stage, the mouse embryo at primitive streak stage, and the *Xenopus* embryo at gastrulation each cover some 100 cells (Rugh, 1968; Nelsen, 1953).

The morphogen needs a cellular mechanism that can convert its concentration into fate. The *Antennapedia (Antp)*-like homeobox gene family can fulfil the requirements of this mechanism. In situ hybridization on mice and *Drosophila* embryos has shown that, at the time of positional specification, the members of this family form a series of different but overlapping domains of expression along the A/P axis (Gaunt, 1991). Each area of the embryo therefore contains a unique blend of *Antp*-like homeobox proteins. Experimentally altering the blend causes the area to adopt an abnormal fate (Kuziora and McGinnis, 1988). This suggests that normal cellular fates are specified by normal levels of the *Antp*-like homeobox proteins.

With a morphogen and a set of *Antp*-like homeobox genes, a small embryo can assign positional values and fates to its parts. However, it faces a problem: it needs to grow. Due to their physical properties, the concentration gradients that provide the initial positional information cannot continue to supply this to the expanding embryo. It is therefore necessary for the parts of the growing embryo to commit their newly specified positional values to memory. A consequence of this operation is that, when part of the embryo is moved to an ectopic site, it differentiates as it would have done if left undisturbed. Classically a piece of an embryo that behaves in this manner is called determined, and the process of committing positional value to memory is known as determination.

Determination has been shown to occur in many organisms. In chickens, when segmented or unsegmented somitic mesoderm is moved from cervical regions to thoracic regions, it differentiates to form vertebrae of cervical form (Kieny et al., 1972). Transplantation experiments on amphibian neurulas have revealed patches of determined mesoderm, the so called secondary fields (De Robertis et al., 1991). These patches form such features as the ears, the balancers, the gills, the heart and the limbs. Insects also undergo determination. When *Drosophila* blastoderm stage cells or imaginal discs are transplanted, they differentiate in a manner appropriate for their donor sites and not their host sites (Simcox and Sang, 1983; Gehring, 1967).

A group of *Drosophila* mutants, the *Polycomb* (*Pc*)

group, has problems in fixing and maintaining the determined state (Jürgens, 1985). These mutants can specify their initial positional values but are unable to remember them (Dura and Ingham, 1988; Struhl, 1981), a forgetfulness that leads to larvae and flies made up of inappropriate parts. Strong mutants of the groups namesake, *Pc*, have initially correct patterns of *Antp*-like homeobox gene expression, but at the extended-germ band stage (stage 11, ~6 hours) these become indiscriminate (Kuziora and McGinnis, 1988). This results in all the segments of the larva adopting the fate of what appears to be the systems ground state, the abdominal 8th segment (Wedeen et al., 1986). The switch from correct to aberrant patterns of *Antp*-like homeobox gene expression in *Pc* mutants suggests that *Drosophila*, having established its positional values, commits these to memory at the extended-germ band stage by tying down the states of activity of its *Antp*-like homeobox genes.

A clue as to how the *Drosophila* memory mechanism may operate has emerged from the cloning and sequencing of the *Pc* gene (Paro and Hogness, 1991). The Pc protein shares a short region of homology with *Drosophila* Heterochromatin Protein 1 (HP1), a component of heterochromatin that plays a role in the position effect variegation (PEV) phenomenon (James and Elgin, 1986). This finding has led to the proposal that the *Antp*-like homeobox gene expression patterns are fixed by preserving the active genes in an open and competent chromatin state, whilst encapsulating the inactive genes within heterochromatin-like complexes, thereby rendering them inexpressible (Gaunt and Singh, 1990; Paro, 1990). Clonal inheritance of these different chromatin states maintains the expression patterns of the *Antp*-like homeobox genes through time, so maintaining the determined state.

The region of homology between the Pc protein and HP1 shares 65% identity over 37 amino acids and is called the chromodomain (*chr*omatin *o*rganization *mo*difier). Zooblot analysis has shown it to be conserved across the animal and plant kingdoms, and a number of genes from the mouse and man have been cloned and shown to contain chromodomains (Singh et al., 1991). The probe used to isolate these initial clones was derived from an HP1 clone. It is perhaps, therefore, not surprising that the genes characterized so far appear to be more similar to HP1 than *Pc*. To see if genes of a more *Pc* type also occur in the mouse, a finding that might suggest a broadening of the applicability of the *Drosophila* memory models to other species, we have conducted a search using a probe derived from a *Pc* clone. It is the results of this screen that are presented here.

Materials and methods

Preparation of Pc *chromobox probe, isolation of clones and sequencing*

A 5′ primer (5′-GAA-TTC-TAC-GCG-GCT-GAG-AAA-ATC-3′) and a 3′ primer (5′-GGA-TCC-GTT-TAC-CTC-CGG-TTC-CCA-3′), which demarcate the *Pc* chromobox, were used to generate a PCR probe from a *Pc* cDNA clone (Paro and Hogness, 1991). The probe was used in a low-stringency screen of an 8.5-day mouse embryo library (Fahrner et al., 1987). Nylon filters (NEN; NEF-978) were hybridized overnight at 58°C in NEN's alternative mix [1 M NaCl, 50 mM Tris-HCl (pH 7.5), 1% SDS, 10% PEG-8000, 5× Denhardt's Solution, 0.1% sodium pyrophosphate], with denatured salmon sperm DNA at a final concentration of 10 μg ml^{-1}, and a probe concentration of 2×10^5 cts minute^{-1} ml^{-1}. After being washed twice for 20 minutes at 50°C in 2× SSC and 1% SDS, they were autoradiographed with intensifying screens for 10 days at −70°C. Positives were isolated as pure clones and their inserts subcloned into Bluescript KS+ (Stratagene). The inserts were sequenced, using the double-stranded dideoxy method (Sanger et al., 1977) described for Sequenase Version 2.0 (USB), by a combination of primer walks and directed exonuclease III deletions (Henikoff, 1984).

Screening of ES cell genomic library

Because the M33 cDNA lacks an initiation codon, a genomic clone was isolated. An ES cell genomic library (a kind gift from A. J. H. Smith) was screened with a radioactively labelled 280 base-pairs (bp) *Eco*RI-*Pst*I probe from the 5′ end of the M33 cDNA. The library filters (NEN; NEF-978) were hybridized overnight at 65°C in the NEN alternative mix, with a final denatured salmon sperm DNA concentration of 100 μg ml^{-1}, and a probe concentration of 1×10^5 cts minute^{-1} ml^{-1}. They were washed twice for 30 minutes at 65°C in 1× SSC and 1% SDS. After autoradiography at −70°C with intensifying screens, positives were picked and purified. The inserts were subcloned into Bluescript KS+ (Stratagene) ready for sequencing. A primer homologous to the 5′ end of the cDNA was used to extend the M33 sequence.

Northern blot analysis

RNA was prepared by a guanidinium thiocyanate method (Chomczynski and Sacchi, 1987) and dissolved in a 40 unit ml^{-1} RNasin (Pharmacia) solution. Total RNA (15 μg) was separated by electrophoresis through a 1% agarose/2.2 M formaldehyde denaturing gel in MOPS buffer and transferred to Gene Screen Plus (NEN; NEF-976) by a capillary method (Lehrach et al., 1977). The filters were heated at 80°C for 2 hours then UV cross-linked (Church and Gilbert, 1984). The filters were prehybridized overnight at 60°C in the NEN alternative mix with a final denatured salmon sperm DNA concentration of 100 μg ml^{-1} and then hybridized, with radioactively labelled probes, in the same mix overnight at 60°C. After washing twice for 30 minutes at 60°C in 2× SSC and 1% SDS, the filters were autoradiographed at −70°C with intensifying screens. The 280 bp M33 *Eco*RI-*Pst*I fragment was used at 4×10^5 cts minute^{-1} ml^{-1}, and the filter was autoradiographed for 3 days. The β-actin control was used at 1×10^5 cts minute^{-1} ml^{-1}, and the filter was autoradiographed overnight. The cyclin B probe is a 156 bp *Sau*3a fragment from a genomic clone of murine cyclin B (Hamers and Singh, unpublished result). This fragment, which codes for an amino acid sequence that has been conserved between the cyclins of many species (MQNSCVPKK VLQLVGVMAM FIAS-KYEEMY PPETGDFAFV TNNTYTKH), was used at 3×10^5 cts minute^{-1} ml^{-1}, and the filter was autoradiographed for 3 days.

In situ hybridization

A ^{35}S-labelled antisense probe was generated from the 280 bp M33 *Eco*RI-*Pst*I fragment using a previously described

method (Gaunt, 1987). Methods for embryo sectioning, alkaline hydrolysis of labelled probes, in situ hybridization and autoradiography were performed as described elsewhere (Gaunt, 1987).

Results

The M33 sequence

The 8.5-day mouse embryo cDNA library yielded 12 positives from 3×10^5 plaques screened with the PCR generated *Pc* chromobox probe. Sequencing one of these, M33, revealed a 1.5 kb open reading frame. This clone does not contain an initiation codon, so its genomic locus was cloned, from an ES cell genomic library, and sequenced. An initiation codon was found immediately upstream of the cDNA-derived sequence. Some 50 bp upstream of this putative start site, there is an in-frame stop codon, suggesting that this is indeed the M33 protein initiation codon. The full-length M33 sequence encodes a 519 amino acid (aa) protein (Fig. 1). A 44 aa stretch of the N-terminal region of this protein shares 61% identity, and a further 16% similarity, with the chromodomain containing N-terminal region of the Pc protein (Paro and Hogness, 1991).

On aligning the M33 protein sequence with the previously described chromodomains (Fig. 2) two subgroups become apparent. The HP1 class is characterized by a block of negatively charged glutamic acid residues immediately upstream of the chromodomain and an amino acid sequence of the form LDCpdLI immediately downstream. Together, these features extend the region of homology between members of the HP1 class from 37 to 50 aa. In comparison, the members of the Pc class do not have the upstream glutamic acid stretch, and they possess a downstream identity, ILDpRLi, different to the HP1 class. HP1, M31 and M32 are members of the HP1 class; *Pc* and M33 are members of the Pc class. This classification extends to the sizes of the members of the HP1 and Pc classes: the HP1 class are all around 190 aa long, while both the Pc protein and M33 are considerably bigger, 390 and 519 aa, respectively.

Outside the chromodomain, comparisons between the predicted amino acid sequences of the chromobox genes (Fig. 3) have revealed a region of homology at the C terminus which is conserved within, but not between, classes. So, although M33 does not contain the Pc protein's very noticeable polyhistidine blocks, it does share, in addition to the chromodomain, a block of 30 aa at its C terminus which has 53% identity and 20% similarity with the C terminus of the Pc protein (Fig. 4). At the whole protein level, M33's high proportion of charged residues [basic (H, K, R) 17%, acidic (D, E) 9%] is comparable to the Pc protein's charged residues [basic 20%, acidic 15%]. Immunohistochemisty has shown that the Pc protein is a nuclear protein (Zink and Paro, 1989) and three putative nuclear localization signals (NLSs) have been reported in its sequence (Paro and Hogness, 1991). Three potential NLSs also emerge in the M33 sequence when it is compared with the general features of an NLS, as defined by studies of the SV40 T antigen minimal NLS (Fig. 1) (Garcia-Bustos et al., 1991). M33 is therefore probably a nuclear protein, and a member of the Pc class of the chromodomain family. This assignment is strengthened by the M33 and Pc protein C-terminal homology.

The M33 promoter

The 147 bp genomic region upstream of the initiation codon has a very high G+C content, 84%. A visual examination reveals a large number of CpGs, and indeed the observed-to-expected ratio as calculated by the formula given below, from Gardiner-Garden and Frommer (1987), equals 1.1.

$$\frac{O}{E}(CpG) = \frac{\text{No. of CpG}}{\text{No. of C} \times \text{No. of G}} \times \text{No. of nucleotides}$$

A genomic digest with the methyl sensitive restriction enzyme *Bss*HII (recognition sequence GCGCGC) has demonstrated that at least some of the Cs of the CpGs surrounding the initiation codon are unmethylated (data not shown). The M33 5′ genomic region therefore has the main characteristics of a CpG island: unmethylated Cs and an observed-to-expected ratio greater than 0.6 (Gardiner-Garden and Frommer, 1987).

The M33 expression pattern

The spatial and temporal expression patterns of M33 have been investigated using a combination of in situ hybridizations and northern blots. Embryos of 7.25, 8.25 and 12.5 days were sectioned and examined by in situ hybridization. Total RNA from 10.5-, 12.5-, 14.5-, 16.5-, and 18.5-day embryos, and newborn mice was separated by electrophoresis and examined by northern blot analysis.

In situ hybridizations to 7.25-day embryos sectioned within their deciduae showed labelling in both maternal and embryonic tissues. Maternal tissue labelling was confined to the outer layer of the decidua (Fig. 5A). Labelling of embryonic tissue was abundant within parts destined to form the foetus (ectoderm and mesoderm germ layers) but was not detected above background in the amnion, chorion or ectoplacental cone (Fig. 5B). Embryos of 8.25 days (Fig. 5C) maintained the distinction between labelled embryonic tissue and apparently unlabelled extraembryonic regions. At this stage, there was little or no specific labelling in the allantois. Embryos of 12.5 days showed labelling in all tissues (Fig. 6). None of the developmental stages examined showed anterior-to-posterior differences in the abundance of transcripts.

A northern blot of RNA samples from the final 10 days of murine embryonic development has been used to investigate the temporal nature of M33 expression. The blot clearly shows (Fig. 7A) that the level of expression gradually falls away from a high point at 10.5 days. One interpretation of this result would be that M33 expression is under developmental stage control. However, since M33 is a potential chromatin protein, the change in its expression may instead be due to a change in the rate of cellular division during this

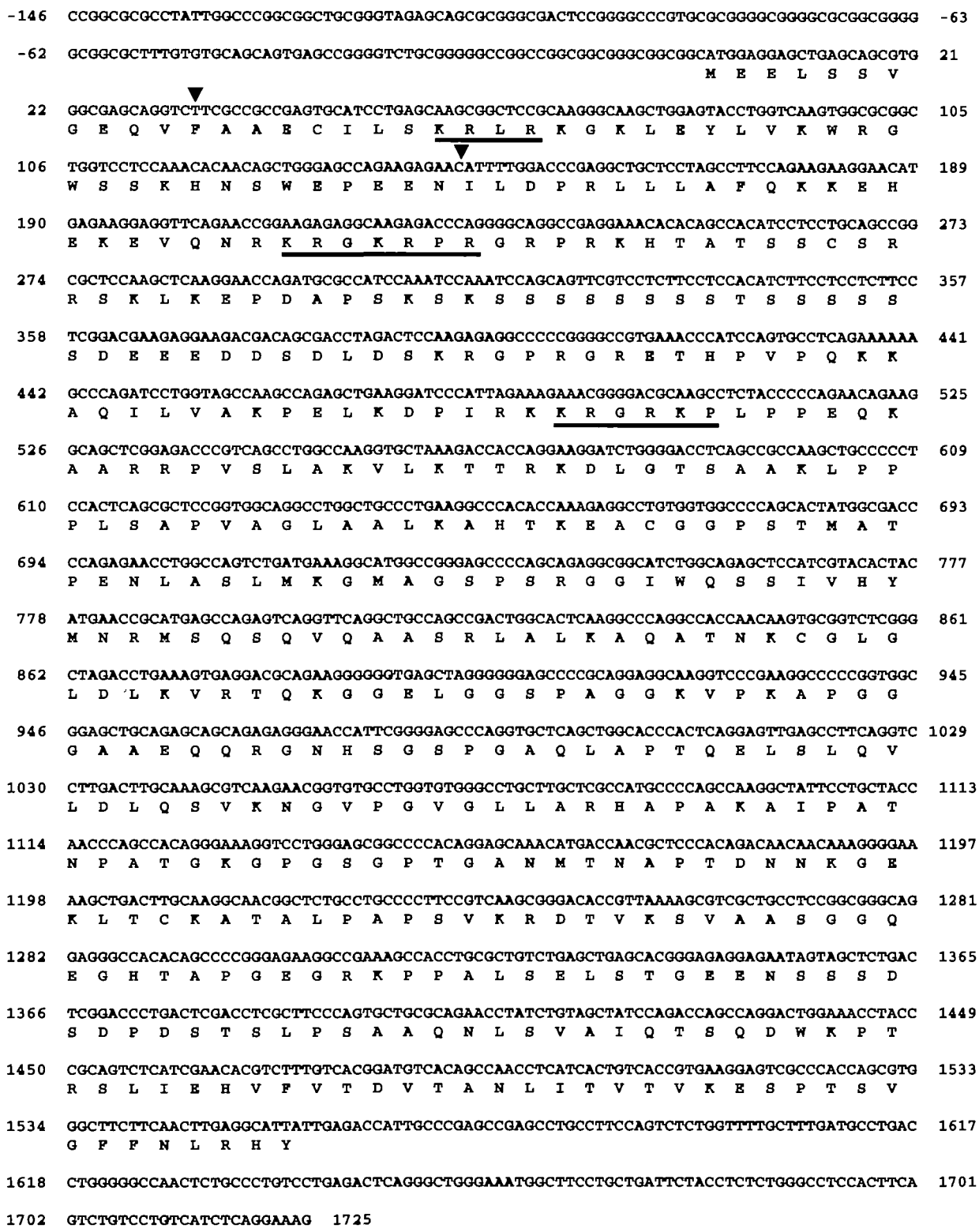

Fig. 1. Nucleotide sequence of M33 and its predicted amino acid sequence. The sequence upstream of base no. 3 is from the genomic clone, that downstream is from the cDNA. The chromobox is demarcated by triangles and the three putative nuclear localization signals are underlined.

embryonic phase. This hypothesis has been tested by comparing the pattern of M33 expression with that of the murine cyclin B gene, a conserved cell-cycle-specific gene whose expression level reflects the degree of cellular proliferation (Lehner and O'Farrell, 1990). The pattern of cyclin B expression (Fig. 7B) closely mirrors

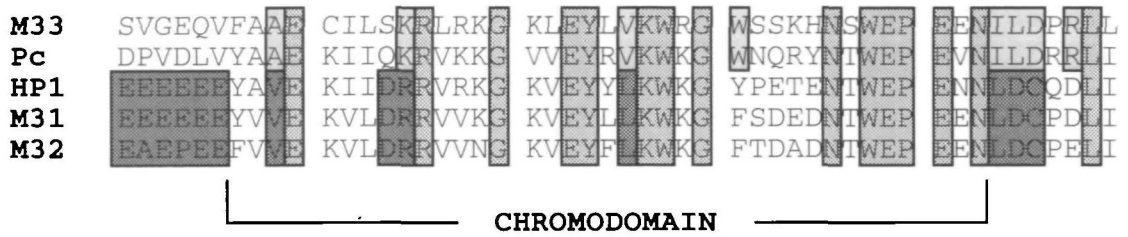

Fig. 2. Comparison of five chromodomain sequences: M31,M32 and M33 are murine genes; *Pc* and HP1 are *Drosophila* genes. The shaded boxes highlight regions that characterize the chromodomain family, the HP1 class, and the Pc class.

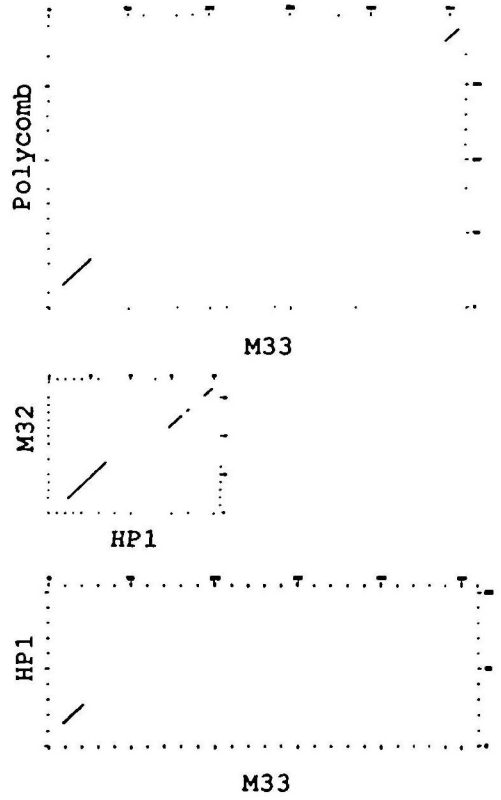

Fig. 3. A graphic matrix comparison of the predicted amino acid sequences of four of the chromobox genes; M33, M32, HP1 and Pc. All were made using the UWGCG Compare program, with a window of 18 and a stringency of 16 (Devereux et al., 1984). The chromodomains are in the N-terminal regions, bottom left. The C-terminal region, class-specific homologies can be seen in the upper right hand corners of the M33/Pc and M32/HP1 comparisons.

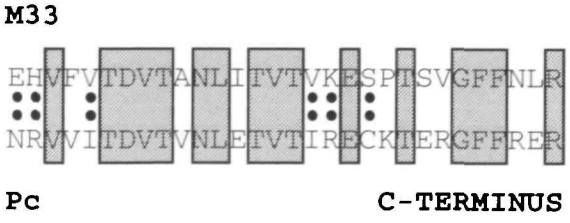

Fig. 4. The region of homology between the C termini of the M33 and Pc proteins. The double dots denote conservative changes and the boxes identities.

that of M33, both decline from a 10.5-day peak. It therefore appears that M33 expression is influenced by the embryo's rate of cellular division, and not by its developmental stage. M33 has been seen to be expressed, at levels lower than in embryos, in adult tissues (data not shown), a finding in accordance with these tissues lower mitotic indexes (Altman and Dittmer, 1972).

Discussion

The roles of M33 and Pc

The sequencing of M33 has led to the classification of two subgroups of chromodomain, the HP1 and Pc classes, and to the assignment of M33 to the Pc class. The intraclass homologies discovered in the C-terminal regions of the chromodomain proteins support this subgrouping, and may point to domains of functional significance. Although it has yet to be shown that M33 is a functional murine homolog of *Pc*, the data presented are not inconsistent with the hypothesis that M33 plays a role similar to *Pc*. The in situ hybridization study has revealed that M33, like *Pc*, shows ungraded expression. The *Pc* expression pattern is thought to be a reflection of the simplicity of its promoter (Paro and Hogness, 1991). Similarly, the embryonic M33 expression pattern may reflect its association with a CpG island, a simple vertebrate promoter.

Pc is a repressor of the *Antp*-like homeobox genes, and it maintains its repression through successive cell generations (Struhl, 1981). The molecular basis of this repression is unknown. However, two memory models propose that the Pc protein may participate in the formation of a heterochromatin-like complex that cloaks inactive *Antp*-like homeobox genes, rendering them inexpressible (Gaunt and Singh, 1990; Paro, 1990). The Pc protein complex may, like the heterochromatin complex of PEV (Henikoff, 1990; Tartof et al., 1984), be able to spread along the chromosome. The models cast *Pc* in a simple role: it is a ubiquitously expressed building block. The model's smart players are other genes that initiate spread of the heterochromatin-like complex at specific initiator sites.

In the context of the memory models, the conserved motifs at both ends of the M33 and Pc proteins may, we suggest, allow these proteins to act like interlocking building blocks, each addition spreading the complex further down the chromosome. The finding that the

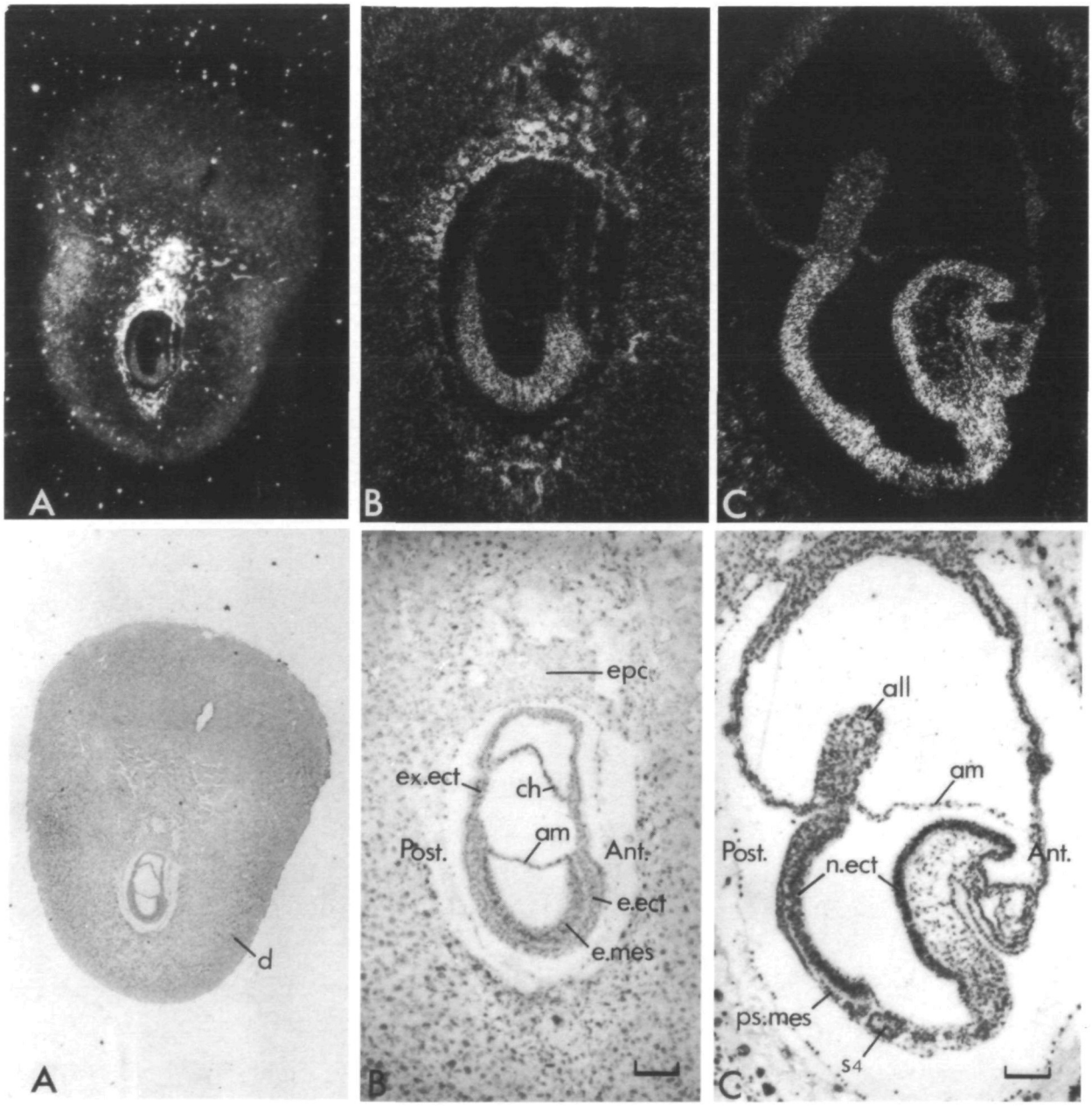

Fig. 5. (A,B) 7.25-day and (C) 8.25-day embryos, sectioned within their deciduae and showing the distribution of M33 transcripts detected by in situ hybridization. B shows a high power view of the embryo shown in A. Upper panels, dark-field; lower panels, bright-field illumination. The intensely white material shown in A is not due to silver grains but to light scattering from maternal blood. In C, the reduced labelling seen over somites 2 and 3 is probably a plane-of-section artefact, since parallel sections (not shown) showed all somites to be equally labelled. Ant, anterior; Post, posterior; d, decidua; e.ect, e.mes, embryonic ectoderm and mesoderm; ex.ect, extraembryonic ectoderm; am, amnion; ch, chorion; epc, ectoplacental cone; n.ect, neural ectoderm; S4, somite 4; ps.mes, presomitic mesoderm; all, allantois. Bars, 0.1 mm.

level of M33 is dependent upon the degree of cellular proliferation suggests that its gene product is required during the process of cellular division. This fits well with the proposal that M33 is a *Pc*-like gene since, according to the models, an *Antp*-like homeobox gene's chromatin state remains static between divisions, but is clonally inherited at division in a process that requires a supply of the Pc protein and other heterochromatin-like building blocks.

There are hints in the literature that the chromodomain is not the only link between PEV and the memory mechanism. The memory models suggest that barriers to the spreading of a heterochromatin-like complex must form within the *Antp*-like homeobox gene clusters. The existence of such barriers would appear to be confirmed by the finding that the homeobox genes of the bithorax complex (BX-C) are resistant to PEV repression (Henikoff, 1990).

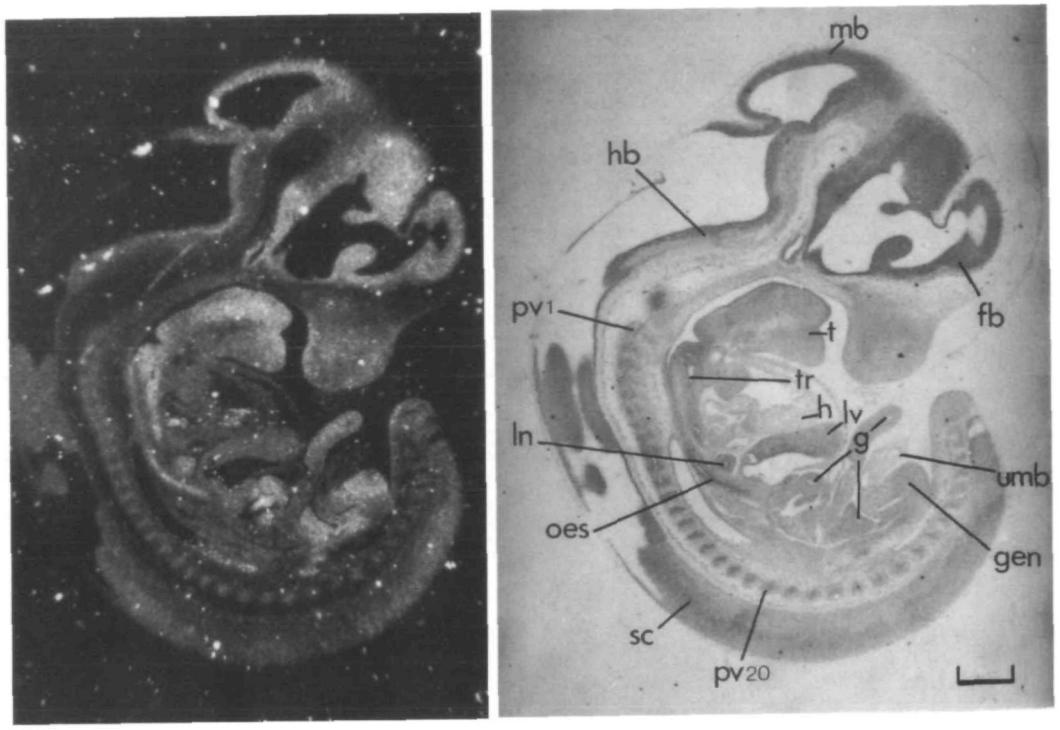

Fig. 6. 12.5-day embryo, sectioned sagittally and showing the distribution of M33 transcripts within all tissues. Left, dark-field; right, bright-field illumination; fb, forebrain; mb, midbrain; hb, hindbrain; pv1,20, prevertebrae 1 and 20; ln, lung; oes, oesophagus; sc, spinal cord; gen, genital eminence; umb, umbilicus; g, gut; lv, liver; h, heart; tr, trachea; t, tongue. Bar, 0.5 mm.

Ramifications of the memory models

The memory models predict that an *Antp*-like homeobox gene control element removed from its chromosomal context will be able to set up an initially correct pattern of expression, but, if the disruption separates the control element from its initiator site, this pattern will not be correctly maintained. Dissection of the regulatory regions of the *Ultrabithorax* (*Ubx*) homeobox gene, a member of the *Drosophila* BX-C, has provided evidence which supports this prediction. The *Ubx* gene product is spatially restricted during embryogenesis to parasegments 5 to 13. Restriction of the Ubx protein to parasegment 5 is the role of the *abx/bx* domain. The *abx* subdomain has been incorporated into promoter-LacZ constructs to study its control elements (Simon et al., 1990). On its own, this subdomain is able to initiate LacZ expression with an anterior boundary coincident with the anterior edge of parasegment 5. However, it is unable to maintain this pattern. LacZ expression is seen to creep forward of parasegment 5 from about 9 hours onwards, until by larval stages all the imaginal discs are expressing LacZ. The inability of *abx* to maintain its expression pattern therefore suggests that *abx* may lack an initiator site. A similar defect in the memory mechanism might, we suggest, be the basis of the Antennapedia transformation. This homeotic transformation is due to ectopic head expression of the *Antp* gene from its P2 promoter (Jorgensen and Garber, 1987). Most of the mutants that give this phenotype have chromosomal inversions that map between P1 and P2 (Scott et al., 1983). In terms of the memory model, these inversions may separate P2 from its initiator site and thereby disrupt the correct maintenance of the P2 expression pattern, resulting in ectopic P2-driven expression.

According to the models, a cell's positional address is preserved in the static chromatin state of its *Antp*-like homeobox genes. Having committed its address to memory in this way, the models predict that the state of activity of the *Antp*-like homeobox genes is cell autonomous. Consistent with this prediction, experiments with chick wing buds, where material from anterior regions was transplanted to posterior regions and vice versa, have shown that expression of the XlHbox 1 antigen is appropriate for the donor site and not the host site (Oliver et al., 1990). Interestingly, similar experiments looking at the expression of two non-*Antp*-like homeobox genes, *Hox-7.1* and *Hox-8.1*, neither of which is a member of the *Hox* clusters, have shown that post-transplantation material adopts the host site's expression pattern (Davidson et al., 1991). The finding that a *Hox* gene from within the *Hox* clusters preserves its state of activity on transplantation, but that *Hox* genes from outside the clusters do not, supports the notion that clustering and memory are associated mechanisms.

Errors in the transmission of the chromatin states at cellular division have been proposed to account for shifts in the fate of *Drosophila* imaginal disc cells (Gaunt and Singh, 1990). Shifts in fate can have dire

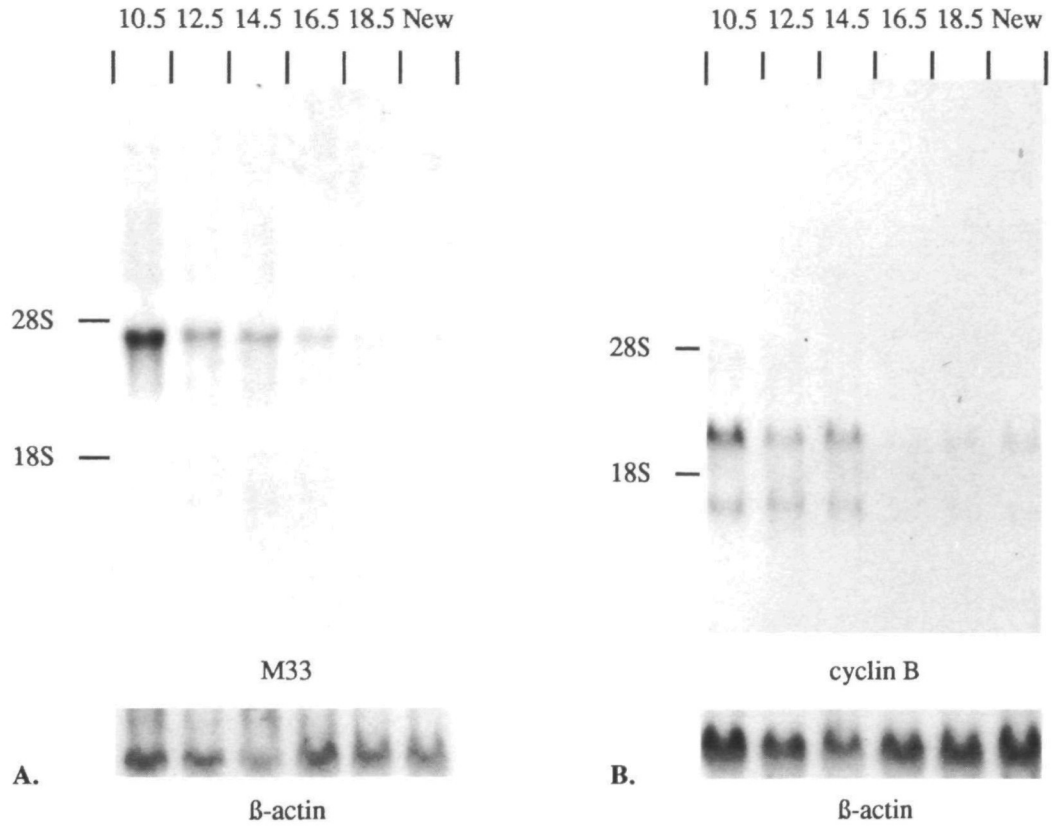

Fig. 7. Northern blot analysis of the M33 and cyclin B genes. Total RNA (15 μg per track) from embryos spanning the final ten days of murine embryonic development was probed with (A) the M33 gene and (B) the cyclin B gene. The blots were also probed with the β-actin gene as a control for loading and transfer.

consequences, as a cell in a new state may be unable to respond to its surroundings. A lack of responsiveness is thought to be a primary cause of cells becoming cancerous. It is intriguing therefore that a recently characterized member of the *Pc*-group, *Posterior Sex Combs* (*Psc*), has been found to have homology with the murine oncogene *bmi-1* (Brunk et al., 1991; van Lohuizen et al., 1991).

The characterization of a *Pc*-like gene in mice extends the credible range of the memory models to include vertebrates. As the members of the *Pc*-group are characterized, it will be interesting to see if they have vertebrate counterparts. Support for the models, in either *Drosophila* or vertebrates, could come from an analysis of the chromatin state of the homeobox genes in different embryonic regions. The M33 protein product will be examined to clarify the degree of functional homology between it and the Pc protein, and a targeted mutagenesis of M33 is planned as it would clearly be of use in defining function.

Note

The M33 sequence described in this paper has been submitted to the EMBL Data Library, and has been assigned the accession number X62537.

We thank R. D. Burton, Dr A. Ferguson-Smith and Dr W. Reik for technical assistance and Dr R. Paro for the kind gift of the *Pc* clone. J. J. Pearce is a recipient of an AFRC Ph.D. studentship and P. B. Singh is a Babraham Research Fellow.

References

Altman, P. L. and Dittmer, D. S. (1972). *Biology Data Book*, vol 1, pp. 119-126. Bethesda, Maryland: FASEB.

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 murine *bmi-1* oncogene. *Nature* **353**, 351-353.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analyt. Biochem.* **162**, 156-159.

Church, G. M. and Gilbert, W. (1984). Genomic Sequencing. *Proc. natl. Acad. Sci. U.S.A.* **81**, 1991-1995.

Davidson, D. R., Crawley, A., Hill, R. E. and Tickle, C. (1991). Position-dependent expression of two related homeobox genes in developing vertebrate limbs. *Nature* **352**, 429-431.

De Robertis, E. M., Morita, A. and Cho, K. W. Y. (1991). Gradient fields and homeobox genes. *Development* **112**, 669-678.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

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.

Fahrner, K., Hogan, B. L. M. and Flavell, R. (1987). Transcription of *H-2* and *Qa* genes in embryonic and adult mice. *EMBO J.* **6**, 1265-1271.

Garcia-Bustos, J., Heitman, J. and Hall, M. N. (1991). Nuclear protein localization. *Biochem. Biophys. Acta* **1071**, 83-101.

Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261-282.

Gaunt, S. J. (1987). Homeobox gene *Hox-1.5* expression in mouse embryos; earliest detection by *in situ* hybridization is during gastrulation. *Development* **101**, 51-60.

Gaunt, S. J. (1991). Expression patterns of mouse Hox genes: clues to an understanding of developmental and evolutionary strategies. *BioEssays* **13**, 505-513.

Gaunt, S. J. and Singh, P. B. (1990). Homeogene expression patterns and chromosomal imprinting. *Trends Genet.* **6**, 208-212.

Gehring, W. (1967). Clonal analysis of determination dynamics in cultures of imaginal discs in *Drosophila melanogaster*. *Dev. Biol.* **16**, 438-456.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351-359.

Henikoff, S. (1990). Position effect variegation after 60 years. *Trends Genet.* **6**, 422-426.

James, T. C. and Elgin, S. C. R. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. *Mol. Cell. Biol.* **6**, 3862-3872.

Jorgensen, E. M. and Garber, R. L. (1987). Function and misfunction of the two promoters of the *Drosophila Antennapedia* gene. *Genes Dev.* **2**, 1600-1614.

Jürgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature* **316**, 153-155.

Kieny, M., Mauger, A. and Sengel, P. (1972). Early regionalization of the somitic mesoderm as studied by development of the axial skeleton of the chick embryo. *Dev. Biol.* **28**, 142-161.

Kuziora, M. A. and McGinnis, W. (1988). Different transcripts of the *Drosophila Abd-B* gene correlate with distinct genetic subfunctions. *EMBO J.* **7**, 3233-3244.

Lehner, C. F. and O'Farrell, P. H. (1990). The roles of Drosophila cyclins A and B in mitotic control. *Cell* **61**, 535-547.

Lehrach, H., Diamond, D., Wozney, J. M. and Boedtker, H. (1977). RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochem.* **16**, 4743-4751.

Nelsen, O. E. (1953). *Comparative Embryology of the Vertebrates*. New York: The Blakiston Company.

Oliver, G., De Robertis, E. M., Wolpert, L. and Tickle, C. (1990). Expression of a homeobox gene in the chick wing bud following application of retinoic acid and grafts of polarizing region tissue. *EMBO J.* **9**, 3093-3099.

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. U.S.A.* **88**, 263-267.

Rugh, R. (1968). *The Mouse*. Minneapolis: Burgess.

Sanger, F., Nicken, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. natl. Acad. Sci. U.S.A.* **74**, 5463-5467.

Scott, M. P., Weiner, A. J., Hazelrigg, T. I., Polisky, B. A., Pirrotta, V., Scalenghe, F. and Kaufman, T. C. (1983). The molecular organization of the *Antennapedia* locus of Drosophila. *Cell* **35**, 763-776.

Simcox, A. A. and Sang, J. H. (1983). When does determination occur in *Drosophila* embryos. *Dev. Biol.* **97**, 212-221.

Simon, J., Peifer, M., Bender, W. and O'Connor, M. (1990). Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis. *EMBO J.* **9**, 3945-3956.

Singh, P. B., Ross Miller, J., 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. *Nucl. Acids Res.* **19**, 789-794.

Slack, J. M. W. (1991). *From Egg to Embryo: Regional Specification in Early Development*. Cambridge: Cambridge University Press.

Struhl, G. (1981). A gene product required for correct initiation of segmental determination in *Drosophila*. *Nature* **293**, 36-41.

Tartof, K. D., Hobbs, C. and Jones, M. (1984). A structural basis for variegating position effects. *Cell* **37**, 869-878.

van Lohuizen, M., Frasch, M., Wientjens, E. and Berns, A. (1991). Sequence similarity between the mammalian *bmi-1* proto-oncogene and the *Drosophila* regulatory gene *Psc* and *Su(z)2*. *Nature* **353**, 353-355.

Wedeen, C., Harding, K. and Levine, M. (1986). Spatial regulation of Antennapedia and bithorax gene expression by the *Polycomb* locus in Drosophila. *Cell* **44**, 739-748.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. *J. theor. Biol.* **25**, 1-47.

Zink, B. and Paro, R. (1989). In vivo binding pattern of the *trans*-regulator of homeotic genes in *Drosophila melanogaster*. *Nature* **337**, 468-471.

(Accepted 17 December 1991)