© 1991 Oxford University Press

A sequence motif found in a *Drosophila* heterochromatin protein is conserved in animals and plants

Prim B.Singh, J.Ross Miller, Jonathan Pearce, Rashmi Kothary*, Robert D.Burton,
Renato Paro[1], Tharappel C.James[2] and Stephen J.Gaunt
Department of Molecular Embryology, Institute of Animal Physiology and Genetics Research,
Cambridge Research Station, Babraham, Cambridge CB2 4AT, UK, [1]Zentrum für Molekulare
Biologie, Universität Heidelberg, D-6900 Heidelberg 1, FRG and [2]Molecular Biology and Biochemistry
Department, Wesleyan University, Middletown, CT 06457-6032, USA

Received November 20, 1990; Revised and Accepted January 21, 1991 EMBL accession nos X56683 and X56690

ABSTRACT

Modifiers of position-effect-variegation in *Drosophila* encode proteins that are thought to modify chromatin, rendering it heritably changed in its expressibility. In an attempt to identify similar modifier genes in other species we have utilized a known sequence homology, termed chromo box, between a suppressor of position-effect-variegation, Heterochromatin protein 1 (HP1), and a repressor of homeotic genes, *Polycomb* (*Pc*). A PCR generated probe encompassing the HP1 chromo box was used to clone full-length murine cDNAs that contain conserved chromo box motifs. Sequence comparisons, *in situ* hybridization experiments, and RNA Northern blot analysis suggest that the murine and human sequences presented in this report are homologues of the *Drosophila* HP1 gene. Chromo box sequences can also be detected in other animal species, and in plants, predicting a strongly conserved structural role for the peptide encoded by this sequence. We propose that epigenetic (yet heritable) changes in gene expressibility, characteristic of chromosomal imprinting phenomena, can largely be explained by the action of such modifier genes. The evolutionary conservation of the chromo box motif now enables the isolation and study of putative modifier genes in those animal and plant species where chromosomal imprinting has been described.

INTRODUCTION

Heterochromatin represents a cytologically visible state of heritable gene repression (1). When genes that lie within euchromatin are brought close to heterochromatin by rearrangement they are subject to an inactivating influence by the spreading of the heterochromatic condition, across the breakpoint, into the euchromatin. The extent of this spreading usually varies between individual cells, resulting in mosaic or variegated patterns of gene expression within a tissue (position-effect-variegation) (2). Studies on position-effect-variegation in *Drosophila* have shown that the inactivation is at the level of transcription (3-5), and examination of the rearranged chromosomes in polytene nuclei has revealed that the variegating locus assumes the condensed morphology of heterochromatin (4-7). This latter observation has pointed to changes in chromatin structure as being responsible for the repression (8,9).

Over several decades, a number of unlinked genes known as modifiers of variegation have been described (2). Mutations in these genes lead either to an enhancement or suppression of the variegating phenotype (6,10-13). Cytological analysis has also shown that mutation in modifier genes can directly affect the degree of heterochromatin formation at the variegating breakpoint (6). These studies have led to the proposal that modifier genes encode proteins which are either structural components of heterochromatin or enzymes that modify these components (reviewed in Refs. 8 and 33). Considerable support for this proposal has come from the recent cloning and characterization of two *Drosophila* modifier genes whose mutations suppress variegation (16-18). One of them, HP1, is allelic to *Suvar(2)5* (16,17) and its product has been shown to be a structural component of heterochromatin (14,15,17), the other is identified by the *Suvar(3)7* mutation (18).

Inactivation of relatively large chromosomal domains has also been invoked as a mechanism of genetic control in the regulation of the bithorax complex of *Drosophila* (18,19). In this system, heritable repression of the homeotic genes is achieved by a family of dominant *trans*-repressor genes known as the *Polycomb*-group (Pc-G) (20). Interestingly, the cloning and sequencing of the central member of this group, *Pc*, has revealed a region of homology with HP1 (R.P. and D.S. Hogness, manuscript in preparation, 21). Thus, position-effect-variegation and control of homeotic genes most probably represent analogous mechanisms of gene repression (18,19). Furthermore, the finding presented here that the region of homology between HP1 and *Pc*, termed chromo box, is conserved suggests that similar mechanisms of heritable gene repression also function in a wide range of animal and plant species.

---

* Present address: Institut du Cancer de Montréal, 1560 rue Sherbrooke est, Montréal, Québec, Canada H2L 4M1

MATERIALS AND METHODS

Generation of PCR probe, isolation of clones and sequencing

A 5′ primer (5′-TAC-GCC-GTG-GAA-AAG-ATC-3′) and a 3′ primer (5′-ATT-GTT-CTC-CGG-CTC-CCA-3′) to the extremities of the region homologous between HP1 and *Pc* were used to generate a PCR probe from the HP1 cDNA clone (14). The annealing temperature was 50°C. The probe was used in a low stringency screening of a 8½ day mouse embryo library (22). Nylon filters (NEN; NEF-978) were hybridised at 58°C in 1% SDS, 1M NaCl and 10% PEG 8000 (Sigma). Filters were washed twice for 20 minutes at 50°C in 2×SSC/1% SDS and put down for autoradiography for 3 days at −70°C with intensifying screens. Inserts were subcloned into Bluescripts KS$^+$ (Stratagene). A human library (23) was screened with radiolabelled (24) M31 probe under conditions of high-stringency (25). Plasmids were sequenced double-stranded using the dideoxy method (26) and a combination of primer-walking and deletions.

Northern blot analysis

RNA was prepared from tissue homogenised in 10 volumes of guanidinium isothiocyanate buffer (28). The homogenate was extracted with an equal volume of 1:1 phenol:chloroform and then precipitated with ethanol. The resulting nucleic-acid pellet was washed with 70% ethanol and resuspended in 0.5% SDS. Total RNA was electrophoresed through a formaldehyde/agarose denaturing gel in MOPS buffer and transferred to Gene Screen plus (NEN; NEF-976). The filter was next U.V. cross-linked and hybridized at 65°C to ^{32}P-labelled (24) M31 probe in a modification of Church's buffer (29) (0.5 M $NaHPO_4$ pH7.2, 7% $NaDodSO_4$ and 1 mM EDTA) and later (16 hrs), washed in a modification of Church's wash buffer (40 mM $NaHPO_4$, 1% SDS) again at 65°C. After washing, the filter was exposed at −70°C to X-ray film (Fuji) for 2 days with intensifying screens.

In situ hybridization

For use in *in situ* hybridization, ^{35}S-labelled antisense M31 probe was synthesised in a direction opposite to that of normal transcription, as previously described (27). The probe included the entire coding sequence plus 665 b.p. of 3′ untranslated sequence. Control (sense) M31 probe was synthesised in the direction of normal transcription. Embryos were dissected from deciduae prior to fixation in 3% paraformaldehyde in PBS (w/v) and were then embedded in paraffin wax. Methods for embryo sectioning, alkaline hydrolysis of labelled probes, *in situ* hybridization and autoradiography were all as previously described (27).

Southern blot analysis

All DNAs were digested with EcoR1 and electrophoresed through a 0.8% agarose gel before transferring to Gene Screen plus (NEN; NEF-976). The lanes contained genomic DNAs from vertebrates (*Mus musculus*, 12 μg; *Homo sapiens*, 12μg; *Xenopus laevis*, 13 μg), insects (*Drosophila melanogaster*, 4 μg; *Planococcus citri*, 4 μg), a nematode (*Caenorhabditis elegans*, 3 μg) and plants (*Zea mays*, 13 μg; *Antirrhinum majus*, 13 μg). The filters were prehybridized at 58°C for 6−8 hrs in 50 mM Tris-HCl pH7.5 containing 1 M NaCl, 1% SDS, 100 mg/ml PEG 8000, 5×Denhardts solution, 1 mg/ml Sodium pyrophosphate and 10 μg/ml sonicated boiled salmon sperm DNA. The prehybridization buffer was removed from the bag and replaced with the same buffer containing HP1 modifier box probe (at 3×10^5 cpm/ml). Hybridization was allowed to proceed overnight and the filters were washed as for the low stringency screening of the mouse embryo cDNA library. After washing, filters were exposed to X-ray film (Fuji) for 10−14 days with intensifying screens.

```
           1                                                              63
                   |------------- Chromo Domain -----------|
HP1        DAEEEEEEYA VEKIIDRRVR KGKVEYYLKW KGYPETENTW EPENNLDCQD LIQQYEASRK DEE

Polycomb   TDDPVDLVYA AEKIIQKRVK KGVVEYRVKW KGWNQRYNTW EPEVNILDRR LIDIYEQTNK SSG

M31        VLEEEEEEYV VEKVLDRRVV KGKVEYLLKW KGFSDEDNTW EPEENLDCPD LIAEFLLSQK TAH

HSM1       VLEEEEEEYV VEKVLDRRVV KGKVEYLLKW KGFSDEDNTW EPEENLDCPD LIAEFLLSQK TAH

M32        VEEAEPEEFV VEKVLDRRVV NGKVEYFLKW KGFTDADNTW EPEENLDCPE LIEDFLNSQK AGK

Consensus  vleeeeeeyv vEKvldrRVv kGkVEYllKW KGfsdedNTW EPEeNldcpd LIaefllsqK tah
```

Figure 1. Comparison of chromo domain sequences from *Homo Sapiens* (HSM1), *Mus musculus* (M31 and M32) and *Drosophila melanogaster* (HP1 and *Polycomb*). The two lines above the *Drosophila* HP1 amino acid sequence delimits a region of 37 amino acids of which 24 are identical to the sequence of the *Polycomb* gene product. This region is termed chromo domain (chromatin organization modifier), and the 111 b.p. encoding it, the chromo box.

Mus musculus modifier 1 (M31)

```
               .         .         .         .         .         .
   1  ATGGGGAAAAAGCAAAACAAGAAGAAAGTGGAGGAGGTACTAGAAGAAGAGGAAGAGGAA  60
      M  G  K  K  Q  N  K  K  K  V  E  E  V  L  E  E  E  E  E  E
      ▼        .         .         .         .         .         .
  61  TATGTGGTGGAAAAAGTTCTTGATCGGCGAGTTGTCAAGGGCAAGGTGGAATATCTTCTA  120
      Y  V  V  E  K  V  L  D  R  R  V  V  K  G  K  V  E  Y  L  L
                              -------------
               .         .         .         .        ▼.         .
 121  AAGTGGAAGGGTTTCTCAGATGAGGACAACACTTGGGAGCCAGAAGAGAATCTGGATTGC  180
      K  W  K  G  F  S  D  E  D  N  T  W  E  P  E  E  N  L  D  C
               .         .         .         .         .         .
 181  CCTGACCTTATTGCTGAGTTTCTACAGTCACAGAAAACAGCTCATGAGACAGATAAGTCA  240
      P  D  L  I  A  E  F  L  Q  S  Q  K  T  A  H  E  T  D  K  S
               .         .         .         .         .         .
 241  GAGGGAGGCAAGCGCAAAGCTGATTCTGATTCTGAAGATAAAGGAGAGGAAAGCAAACCA  300
      E  G  G  K  R  K  A  D  S  D  S  E  D  K  G  E  E  S  K  P
               .         .         .         .         .         .
 301  AAGAAGAAGAAAGAAGAGTCAGAAAAGCCACGAGGCTTTGCCCGGGGTTTGGAGCCAGAG  360
      K  K  K  K  E  E  S  E  K  P  R  G  F  A  R  G  L  E  P  E
               .         .         .         .         .         .
 361  CGGATTATTGGAGCTACTGACTCCAGTGGAGAGCTCATGTTCCTGATGAAATGGAAAAAC  420
      R  I  I  G  A  T  D  S  S  G  E  L  M  F  L  M  K  W  K  N
               .         .         .         .         .         .
 421  TCTGATGAGGCTGACCTGGTCCCTGCCAAGGAAGCCAATGTCAAGTGCCCACAGGTTGTC  480
      S  D  E  A  D  L  V  P  A  K  E  A  N  V  K  C  P  Q  V  V
               .         .         .         .         .         .
 481  ATATCCTTCTATGAGGAAAGGCTAACGTGGCATTCCTACCCCTCAGAGGATGATGACAAA  540
      I  S  F  Y  E  E  R  L  T  W  H  S  Y  P  S  E  D  D  D  K
               .
 541  AAAGACGACAAGAATTAG                                            558
      K  D  D  K  N  *
```

Mus musculus modifier 2 (M32)

```
               .         .         .         .         .       ▼ .
   1  ATGGGAAAGAAACAAAATGGAAAGAGTAAAAAAGTTGAAGAGGCAGAGCCTGAAGAATTT  60
      M  G  K  K  Q  N  G  K  S  K  K  V  E  E  A  E  P  E  E  F
               .         .         .         .         .         .
  61  GTGGTAGAAAAAGTACTGGACCGTCGTGTAGTGAATGGGAAGGTGGAGTATTTCCTGAAG  120
      V  V  E  K  V  L  D  R  R  V  V  N  G  K  V  E  Y  F  L  K
                           -------------
               .         .         .         .     ▼   .         .
 121  TGGAAGGGGTTCACAGATGCTGATAATACTTGGGAACCAGAAGAAAATTTAGATTGTCCA  180
      W  K  G  F  T  D  A  D  N  T  W  E  P  E  E  N  L  D  C  P
               .         .         .         .         .         .
 181  GAATTAATTGAAGACTTTCTTAATTCTCAAAAAGCTGGTAAAGAAAAAGATGGTACAAAA  240
      E  L  I  E  D  F  L  N  S  Q  K  A  G  K  E  K  D  G  T  K
               .         .         .         .         .         .
 241  AGGAAATCTTTATCTGACAGTGAATCTGATGATAGCAAATCGAAGAAGAAGAGAGATGCT  300
      R  K  S  L  S  D  S  E  S  D  D  S  K  S  K  K  K  R  D  A
               .         .         .         .         .         .
 301  GCTGACAAACCAAGGGGCTTTGCCAGAGGTCTCGACCCTGAACGAATAATCGGCGCCACA  360
      A  D  K  P  R  G  F  A  R  G  L  D  P  E  R  I  I  G  A  T
               .         .         .         .         .         .
 361  GACAGCAGCGGAGAGTTAATGTTTCTCATGAAGTGGAAGGACTCGGACGAGGCCGACTTG  420
      D  S  S  G  E  L  M  F  L  M  K  W  K  D  S  D  E  A  D  L
               .         .         .         .         .         .
 421  GTGCTGGCAAAGGAGGCGAACATGAAGTGTCCTCAGATTGTCATTGCCTTCTACGAGGAG  480
      V  L  A  K  E  A  N  M  K  C  P  Q  I  V  I  A  F  Y  E  E
               .         .         .         .
 481  CGGCTGACTTGGCATTCTTGTCCTGAAGATGAAGCACAATAA                    522
      R  L  T  W  H  S  C  P  E  D  E  A  Q  *
```

Figure 2. Nucleotide sequences of M31 and M32 with their predicted amino acid sequences for the ORFs. Inverted triangles denote the limits of the chromo domain and possible nuclear localization signals (52) found within the chromo domain are underlined.

RESULTS

Cloning of cDNAs containing the chromo box motif

Using an HP1 cDNA clone (14) as template, we prepared a 111 b.p. PCR-generated probe to the region of homology with *Pc* (Figure 1). This probe was then used to screen an 8½ day mouse embryo cDNA library. The clones isolated fell into several different classes that represent transcripts from different genes (PBS, RDB and SJG, unpublished result). Sequencing the largest cDNA of one class (M31) revealed a stretch of 37 amino acids that shares 70% identity with the chromo domain of HP1 (Figure 1). An adjacent stretch of six glutamic acid residues is also shared in both M31 and HP1. We have proceeded to screen a human cDNA library with M31, and have cloned its human equivalent (HSM1). The amino-acid sequence of the HSM1 chromo domain (Figure 1) is identical to that of M31, although sharing 94% identity at the nucleotide level (not shown). Again, adjacent to the HSM1 domain are the six glutamic acid residues characteristic of the canonical modifier protein, HP1.

Sequencing a cDNA of another class of transcripts (M32; Figure 1) isolated from the mouse cDNA library screen has revealed a chromo domain that is different but clearly related to those found in M31 and HSM1. The amino acid sequence of the M32 chromo domain shares 86% identity with that of M31. Furthermore, the presence of a group of negatively charged amino acids next to the chromo domain in M32 suggests that it may also possess a function similar to that of the HP1 homologues. Sequencing cDNAs from the other classes isolated in the screen has revealed at least two other genes containing chromo boxes (unpublished result).

Comparisons of the protein coding sequences of M31 and M32 with HP1

The chromo domains of the murine proteins lie close to the N-terminal portion of the molecules (Figure 2), as does the chromo domain of the human modifier HSM1 (Figure 1), whose amino acid sequence (data not shown) is identical to that of M31. The chromo domains of the *Drosophila* proteins, HP1 (14,17) and *Pc* (21), also lie close to the N-termini. M31 and M32 are both of similar size to HP1, the predicted molecular weights being 24,700 and 22,850 respectively. The predicted molecular weight of HP1, as calculated from the corrected amino acid sequence (17, Genbank accession No. M33749), is 24,300. Both murine sequences are also very hydrophillic, approximately 40% of the residues being highly charged (D,E, K or R), with the ratio of acidic to basic amino acids being 1.3:1 for M31 and 1.25:1 for M32.

When the amino-acid sequences of M31 or M32 are compared to that of HP1 (17) over their entire length, they share 51% identity. However, dot matrix comparisons (Figures 3a and 3b), at high stringency, of the nucleotide sequences of M31 or M32 versus HP1 show that the most highly conserved region is that which encompasses the chromo box. This result attests to the high degree of evolutionary conservation placed on this motif and suggests an important role for the chromo domain in the functioning of this class of modifier proteins. Dot matrix comparison of M31 with M32 (Figure 3c) shows that they share considerable homology on two sides of a central region (nucleotides 200–300 of M31) where the sequences diverge. From this analysis it appears that there are possibly two functional parts to these proteins that are constrained, by selection, in the degree to which they can diverge by mutation. These two parts appear to be joined by a 'hinge' which has been allowed to accumulate mutational changes.

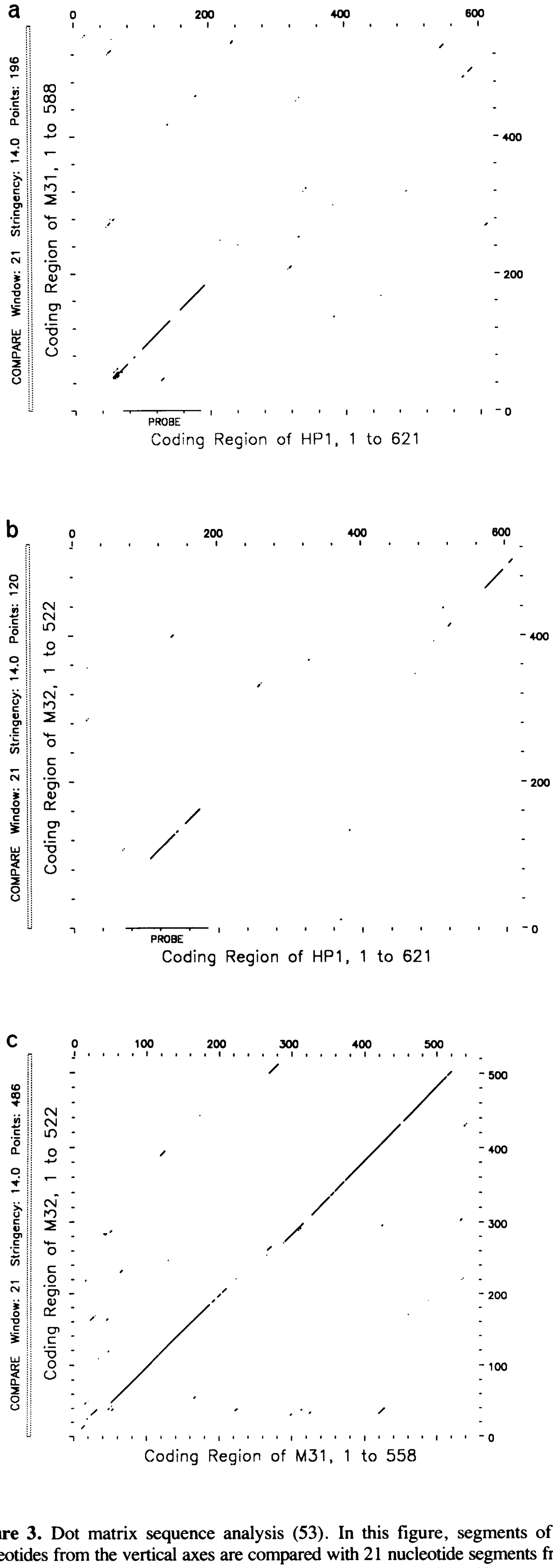

Figure 3. Dot matrix sequence analysis (53). In this figure, segments of 21 nucleotides from the vertical axes are compared with 21 nucleotide segments from the horizontal axes, and a dot is placed in the appropriate position if 14 or more nucleotides are shared between the segments compared. In this manner homologies between sequences can be detected visually. a, Comparison of M31 with HP1. b, Comparison of M32 with HP1. c, Comparison of M32 with M31. The PCR probe (see **MATERIALS AND METHODS**) encompassing the HP1 chromo box is underlined in 3a and 3b.

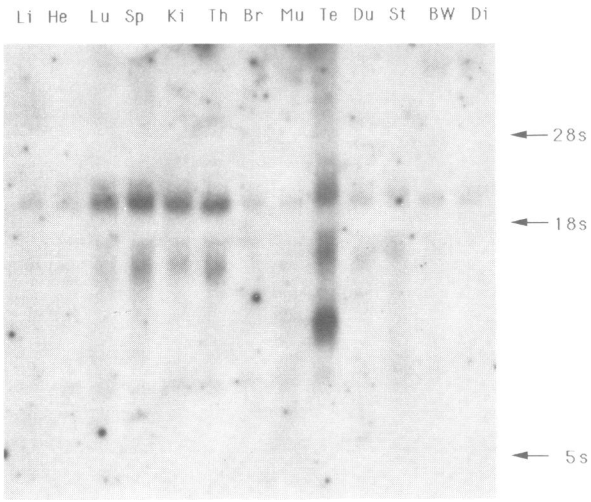

Figure 4. RNA Northern blot analysis of M31 expression. Shown is a blot of adult total RNA (~ 10 to 20 μg per lane) prepared from different tissues of a male mouse, and hybridised with the M31 probe. The lanes are: Li, liver; He, heart; Lu, lung; Sp, spleen; Ki, kidney; Th, thymus; Br, brain; Mu, skeletal muscle; Te, testes; Du, duodenum; St, stomach; Bw, bodywall; Di, diaphragm.

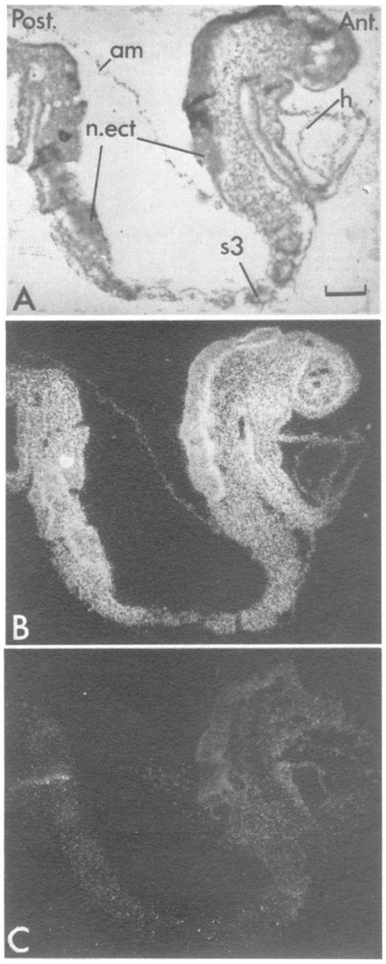

Figure 5. M31 transcripts detected by *in-situ* hybridization within complete parasagittal sections of the 8½ day mouse embryo. A, Bright field. B, Dark field view of same section probed with anti-sense M31 RNA probe. C, Nearby section probed with the control M31 sense probe. Ant, anterior; Post, posterior; n.ect, neural ectoderm; s3 somite 3; am, amnion; h, heart. Bar, 0.1 mm.

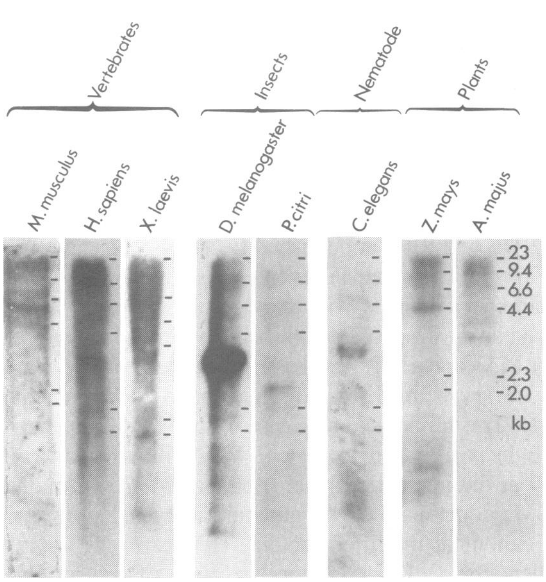

Figure 6. Detection of the chromo box homology in plant and animal genomic DNAs. More than one band can be seen in all of the DNA tracks suggesting that each genome contains genes possessing the chromo box motif. The lines next to the DNA tracks denote the positions of λ Hind III marker fragments.

M31 Expression Analysis

A study of the expression of M31 has shown that it is transcribed in all adult tissues (Figure 4) and gives rise to three different transcript sizes. A 2.2 Kb transcript is found in all tissues and is at elevated levels in lung, spleen, kidney, thymus and testis. In each of these tissues an additional transcript of 1.6 Kb is also present. The 1.6 Kb transcript can also be detected in the duodenum and stomach, but at lower levels. The testis, in addition to these two transcript sizes, contains another highly abundant transcript of 1 Kb that is not found in any other tissue. *In situ* hybridization experiments on 8½ day mouse embryo sections (Figure 5) have similarly indicated that M31 transcripts are abundant in all embryonic tissues. This ubiquitous expression of M31 parallels results obtained for HP1 in *Drosophila* embryos (15).

Conservation of the chromo box motif

The zooblot of Figure 6 shows that the chromo box motif is conserved in a range of organisms taken from both animal and plant kingdoms. More than one band can be seen in each of the DNA tracks when the blots are probed with the *Drosophila* HP1 chromo box, suggesting multiple genes containing the motif. As a common ancestor of plants and animals is thought to have lived approximately 10^9 yrs ago (30), Figure 6 illustrates the strict conservation placed on this motif through evolutionary time.

DISCUSSION

Although we cannot, at present, be certain of the function of M31, M32 and HSM1 proteins, several lines of evidence lead us to suggest that they may function as gene repressors, perhaps by serving as components of heterochromatin. Thus, by analogy, the two chromo box genes so far isolated in *Drosophila*, HP1 and *Pc*, are both known repressors of gene activity. HP1 is a structural component of heterochromatin (14,15,17). *Pc* is also a component of chromatin (31) and might form part of a heterochromatin-like complex (reviewed in Refs. 19 and 21). We suggest that the protein domain encoded by the chromo box might therefore be essential for packaging of proteins within heterochromatin or heterochromatin-like complexes. Since neither *Pc* (21) nor HP1 proteins (TCJ, unpublished) bind directly to DNA, the chromo domain is more likely involved in protein-protein that in protein-DNA interaction. The function of the group of negatively charged glutamic acid residues next to the chromo domains (Figures 1 and 2) in HP1, M31, HSM1 and M32 is unclear. However, experiments on the silencing of the HML and HMR loci in yeast (32) have shown that the positively charged residues 16−18 on histone H4 are critical in the repression of the mating-type loci. It is tempting to speculate that this sequence may be one site with which the negatively charged glutamic acid stretch interacts during the formation of compact heterochromatin.

The sensitivity of variegating phenotypes to changes in the dosage of modifiers has led to the proposal that heterochromatin domains consist of a repeating multimeric complex whose subunits are encoded by modifier genes such as HP1 (13,33). As proposed by Tartof and co-workers, the assembly and dissolution of heterochromatin domains obey the chemical law of mass action. Accordingly, small changes in the concentration of constituent modifiers have dramatic and opposite (antipodal) effects on heterochromatin formation (and therefore of variegation). Although put forward with regard to heterochromatin formation, similar repressor complexes may be involved in other instances of heritable gene repression that must occur elsewhere in the genome. Such a view is supported by the observation that the chromo box motif can be detected in *C.elegans* DNA (Figure 6), an organism that may not possess cytologically visible heterochromatin (34). Indeed, the observations that the cytologically euchromatic short arm of the fourth chromosome in *Drosophila* binds HP1 (15), and may be variegation inducing (35,36), that *Pc* binds to at least 50 different sites on polytene chromosomes (31), and that transgenes in mice can exhibit mosaic expression (37−39), reminiscent of position-effect-variegation (40), suggests to us that extensive regions of the genome might be subject to repression by heterochromatin-like complexes.

Modifier proteins provide a possible molecular basis for chromosomal imprinting (40). Chromosomal imprinting is the epigenetic process by which a chromosome or gene is changed in its ability to function in subsequent cell generations. So far, studies on imprinting have been largely concerned with heritable differences in parental chromosomes (41,42). As first documented by Metz (43), for example, paternally inherited X chromosomes in males of the species *Sciara coprophilia* are never expressed due to their elimination. Interestingly, the controlling region on the X chromosome that governs this behaviour maps to the heterochromatin adjacent to the X centromere (44,45). Heterochromatization has also been invoked as the mechanism for chromosomal imprinting in *Coccids* (41). For example, in male members of the species *Planococcus citri* (46) the entire paternal chromosome set is inactivated by heterochromatization. In mammals, X-chromosome inactivation is another case of chromosomal imprinting known to be due to heterochromatization (47). Differences in the degree of assembly of repressor complexes along homologous chromosomes as seen in the above examples might also, we believe, provide an explanation for the phenotypes of embryos produced in mammalian pronuclear transplantation experiments (48). Thus, gynogenones have an antipodal phenotype when compared to androgenones. Similarly, when paternal disomies are compared to their reciprocal maternal disomies their phenotypes are also antipodal (49).

If, as proposed (13,33), repressor complexes consist of interacting subunits encoded by modifier genes, then the alleles that an individual bears at modifier loci could determine the stability of the complex. Phenotypically, therefore, the alleles at these loci may affect the penetrance and expressivity of the repressed trait. Mutant alleles at such loci could then be the source of some inherited human diseases (50). Indeed, it has been proposed that the alleles present at an X-linked modifier locus determine the age-of-onset of Huntington's Chorea (51). Consistent with this model, we have now mapped two homologues of HSM1 (Figure 1) to the human X chromosome (PBS, SJG and W. Reik, unpublished result).

ACKNOWLEDGEMENTS

We thank Drs. S.C.R.Elgin, K.D.Tartof, U.Nur, R.Blackman, M.E.N.Majerus and D.Albertson for helpful discussions regarding this work. For help with PCR and encouragement we thank Drs. A.J.H.Smith and F.Hill. Drs. S.Munro and R.Krumlauf provided the human and mouse cDNA libraries respectively. *C.elegans* and *D.melanogaster* DNA was provided by Dr. T.Barnes; plant DNA by Dr. R.B.Flavell. PBS is indebted to Dr. M.A.H.Surani for generous provision of laboratory space and facilities. JP is a recipient of an AFRC Ph.D. studentship. RK was supported by the MRC of Canada. PBS is a Babraham Research Fellow.

REFERENCES

1. Brown, S.W. (1966) *Science,* **151,** 417−425.
2. Spofford, J. (1976) In Ashburner, M. and Novitski, E. (eds.), The Genetics and Biology of *Drosophila.* Academic Press. New York, Vol. 1c, pp 955−1018.
3. Nix, C.E. (1973) *Biochem. Genet.,* **10,** 1−12.
4. Henikoff, S. (1981) *Chromosoma (Berl.),* **83,** 381−393.
5. Kornher, J.S. and Kauffman, S.A. (1986) *Chromosoma (Berl.),* **94,** 205−216.
6. Reuter, G., Werner, W. and Hoffman, H.J. (1982) *Chromosoma (Berl.),* **85,** 539−551.
7. Zhimuler, I.F., Beyaeva, E.S., Formina, O.V., Protopopov, M.O. and Bolshakov, V.N. (1986) *Chromosoma (Berl.),* **94,** 492−504.
8. Eissenberg, J.C. (1989) *BioEssays,* **11,** 14−17.
9. Tartof, K.D., Hobbs, C. and Jones, M. (1984) *Cell,* **37,** 869−878.
10. Sinclair, D.A.R., Mottus, R.C. and Grigliatti, T.A. (1983) *Mol. Gen. Genet.,* **191,** 326−33.
11. Reuter, G. and Wolff, I. (1981) *Mol. Gen. Genet.,* **182,** 516−519.
12. Reuter, G., Dorn, R., Wustmann, G., Friede, B. and Rauh, G. (1986) *Mol. Gen. Genet.,* **202,** 481−487.
13. Locke, J., Kotarski, M.A. and Tartof, K.D. (1988) *Genetics,* **120,** 181−198.
14. James, T.C. and Elgin, S.C.R. (1986) *Mol. Cell Biol.* **6,** 3862−3872.
15. James, T.C., Eissenberg, J.C., Craig, C., Dietrich, V., Hobson, A. and Elgin, S.C.R. (1989) *Eur. J. Cell Biol.* **50,** 179−189.
16. Eissenberg, J.C., Elgin, S.C.R. and James, T.C. (1987) *Genetics,* **116,** s4.
17. Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V. and Elgin, S.C.R. (1990) *Proc. Natl. Acad Sci USA.* In Press.
18. Reuter, G., Giarre, M., Farah, J., Gausz, J., Spierer, A. and Spierer, P. (1990) *Nature,* **344,** 219−223.
19. Gaunt, S.J. and Singh, P.B. (1990) *TIG,* **6,** 208−212.
20. Jürgens, G. (1985) *Nature,* **316,** 153−155.
21. Paro, R. (1990) *TIG.,* **16,** 416−421.
22. Fahrner, K., Hogan, B.L.M. and Flavell, R. (1987) *EMBO J.* **6,** 1265−1271.
23. Aruffo, A. and Seed, B. (1987) *Proc. Natl. Acad. Sci. USA,* **84,** 8573−8577.
24. Feinberg, A. and Vogelstein, B. (1983) *Anal. Biochem.,* **132,** 6−13.
25. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
26. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74,** 5463−5467.

27. Gaunt, S.J. (1987) *Development*, **101**, 51–60.
28. Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156–159.
29. Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
30. Gilbert, W., Marchionni, M. and McKnight, G. (1986) *Cell*, **46**, 151–154.
31. Zink, B. and Paro, R. (1989) *Nature*, **337**, 468–471.
32. Magee, P.C., Morgan, B.A. and Smith, M.M. (1990) *Science*, **247**, 841–845.
33. Tartof, K.D., Bishop, C., Jones, M., Hobbs, C.A. and Locke, J. (1989) *Devel. Genet.*, **10**, 162–176.
34. *The Nematode* Caenorhabditis elegans (1988). Wood, W. (ed.) Cold Spring Harbor University Press, Cold Spring Harbor.
35. Panshin, I.B. (1938) *Biol. Zh. (Mosk.)*, **7**, 295–302.
36. Griffen, A.B. and Stone, W.S. (1940) *Univ. Texas Publ.*, **4032**, 201–207.
37. McGowan, R., Campbell, R., Peterson, A. and Sapienza, C. (1989) *Genes Dev.*, **3**, 1669–1676.
38. Katsuki, M., Sato, M., Kimura, M., Yokoyama, M., Kobayashi, K. and Nomura, T. (1988) *Science*, **241**, 593–595.
39. Sweetser, D., Hauft, S., Hoppe, P., Birkenmeir, E. and Gordon, J.I. (1988) *Proc. Natl. Acad. Sci USA*, **85**, 9611–9615.
40. Sapienza, C. (1990) *Scientific American*, **263**, 52–60.
41. Chandra, H.S. and Brown, S.W. (1975) *Nature*, **253**, 165–168.
42. Lyon, M.F. and Rastan, S. (1984) *Differentiation*, **26**, 63–67.
43. Metz, C.W. (1938) *Am. Naturalist*, **72**, 485–520.
44. Crouse, H.V. (1960) *Genetics*, **45**, 1429–1443.
45. Crouse, H.V. (1979) *Chromosoma*, **74**, 324–339.
46. Brown, S.W. and Nur, U. (1964) *Science*, **145**, 130–136.
47. Takagi, N. and Sasaki, M. (1975) *Nature*, **256**, 640–642.
48. Surani, M.A.H., Barton, S.C. and Norris, M.L. (1984) *Nature*, **308**, 548–550.
49. Cattanach, B.M. and Kirk, M. (1985) *Nature*, **315**, 496–498.
50. Hall, J.G. (1990) *Am. J. Hum. Genet.*, **46**, 857–873.
51. Laird, C.D. (1990) *TIG*, **6**, 242–247.
52. Goldfarb, D.S. (1989) *Curr. opinion in Cell Biol.* **1**, 441–446.
53. Devereux, J., Haeberli, P. and Smithies, O. (1984) *NAR*, **12**, 387–395.