Functional characterisation of POLYCOMBLIKE and a novel, chromosomal protein interactor from *Drosophila melanogaster*

A thesis submitted for the degree of Doctor of Philosophy

by
Stanley Robert, B.Sc. (Hons).

Department of Genetics
University of Adelaide
South Australia 5005

3rd of March 1997

Table of Contents

Abstract

The Polycomb-Group of genes is responsible for the maintenance of repression of many genes during *Drosophila* development. Sequence analyses of those members of the Polycomb-Group that have been cloned provide few clues about the mechanism of this repression. However, some genetic and molecular data suggest that the mechanism is similar to that operating during the repression of genes in Position Effect Variegation.

In order to determine the mechanism of Polycomb-Group repression, the identification of proteins that interact with a key member of the Polycomb-Group, POLYCOMBLIKE, was undertaken. Two proteins were found to interact with POLYCOMBLIKE. One of these, ENHANCER of ZESTE, is a previously characterised Polycomb-Group member that is known to be essential for the attachment of Polycomb-Group proteins to chromosomes. The other, temporarily named 2.1, is a novel protein. A region of POLYCOMBLIKE containing two PHD fingers, a recently identified, putative protein-interaction motif, appears to be responsible for binding ENHANCER of ZESTE but not 2.1.

Antibodies were raised against 2.1 for use as a molecular probe to characterise its nature and distribution. It was found that 2.1 is present ubiquitously during embryogenesis, as has been found for members of the Polycomb-Group, however, it was distributed differently on polytene chromosomes. Its interbanded deposition and close physical abutment to sites of deposition of POLYCOMBLIKE suggest a role for 2.1 in boundary element structures that abate the spread of Polycomb-Group repression, as a general chromosomal factor, or as a general transcriptional activator.

A previously identified mutant, *l(3)SG23*, was defined as being a novel Polycomb-Group gene in the course of attempting to identify a mutation within the *2.1* gene.

The mechanism by which the Polycomb-Group protein complex attaches to DNA is presently unkown. During the course of this study, significant sequence similarity was detected between POLYCOMLIKE and a murine protein, M96, that was reported to have DNA binding activity, thus potentially providing a mechanism for the DNA attachment of the Polycomb-Group protein complex. However, it appears that POLYCOMBLIKE is not able to bind to DNA *in vitro*.

Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give full consent for this copy of my thesis to be made available for loan and photocopying.

Stanley Robert

3rd of March 1997

Acknowledgments

Many people have contributed to the last 4 years. Thanks go to:

Rob, for giving me permission to join his lab, for his enthusiasm and for allowing me to work independently

Steve, Andy and Dan, for providing plenty of stimulating and cherished conversation

Michelle, for being an ever-helpful Pc-G co-worker, and for having a good, sarcastic sense of humour

Helena and Ulrik, for their advice on many experiments

Hugh Brock, for allowing me to work in his lab and teaching me the knack of polytene immuno-stainings (and the maple syrup)

Jacob Hodgson, for delivering my protein chemist side

Julie, for telling me about third chromosome balancers

Sinead, for being a good sport (mostly)

Peter Kolesik, for his excellent confocal microscope assistance and knowledge of eucalypt galls

David Hayman, for his diamond cutter

Heidi, for writing well on the first attempt

The rest of the lab for providing a generally industrious and happy atmosphere

Alice, for making me smile everyday

Briony, for everything as well as some of the above

Chapter 1 - Introduction

The concept that the differentiation of a cell is caused by expression of a particular suite of transcriptional regulatory proteins has become firmly entrenched. However, we are just beginning to understand that the expression of a specific set of transcriptional regulatory proteins must persist for the cell to retain its identity and that this persistence often occurs in the absence of the original signals that established the transcriptional state. How a cell 'remembers' its transcriptional or epigenetic state is the question which the work presented in this thesis attempts to address.

1.1 - Transcriptional state maintenance systems

There are several well characterised systems or components of systems responsible for maintaining established transcriptional states within a cell. These include mammalian X - chromosome inactivation, mating-type maintenance in *Saccharomyces cerevisiae*, DNA methylation and histone modification. These systems will not be reviewed in detail independently here but will be considered within the more general discussion of epigenetic regulatory mechanisms that follows.

1.2 - Transcriptional state maintenance during *Drosophila* development

The question of maintenance of epigenetic states is especially intriguing with respect to development. During the development of any multicellular organism, individual cells or cell groups modify their gene expression (epigenetic state) in order to differentiate and produce diverse tissues in highly ordered patterns. The breakdown of epigenetic state maintenance during development would be especially conspicuous. Cell groups would lose their differentiated identity, adopt alternate, spurious fates and the process of development would degenerate. *Drosophila* development is exceptionally well described. Perturbations in the developmental program, therefore, are relatively easy to identify and characterise. This, combined with the ease of genetic manipulation, has led to the identification of two groups of genes that are responsible for epigenetic state maintenance during *Drosophila* development. The trithorax Group (trx-G) is responsible for maintenance of active gene states while the Polycomb Group (Pc-G) is responsible for the maintenance of repressed gene states. The canonical members of these groups, *trithorax* (*trx*) and *Polycomb* (*Pc*), were identified by their homeotic phenotypes.

1.2.1 - Homeotic gene expression during *Drosophila* embryogenesis

In *Drosophila* there are two clusters of homeotic genes ; the bithorax complex (BX-C) (Lewis, 1978) and the Antennapedia complex (ANT-C) (Kaufman *et al.*, 1980). Each of the genes in these clusters is expressed in a different, contiguous block of the embryo along the antero-posterior axis. These complex patterns of expression are effected by the products of the segmentation genes

(reviewed in Lawrence and Morata, 1994). That is, in every parasegment of the embryo a different suite of segmentation gene products specifically activates or represses all of the homeotic genes. Loss-of-function or gain-of-function mutations in homeotic genes result in homeotic phenotypes - segments develop with the identity of other segments (Lewis, 1978). This genetic evidence strongly favoured the idea that the homeotic genes are master genes, whose specific combination of expression regulates specific batteries of structural genes to produce segment specific structures. The homeotic genes were shown to encode homeodomain containing proteins (McGinnis *et al.*, 1984; Scott and Weiner, 1984) supporting the idea that they directly regulated the expression of structural genes.

trx mutants were shown to have phenotypes resembling those seen in loss-of-function homeotic mutants (Ingham and Whittle, 1980), while *Pc* mutants were shown to resemble gain-of-function homeotic mutants (Lewis, 1978). In both of these cases the presence of wild type homeotic complexes was necessary for the visualisation of the phenotypic effect. These data suggested that *trx* and *Pc* are responsible for the general activation and repression, respectively, of homeotic gene expression.

1.3 - The trx-G and maintenance of activation

Subsequent to the identification of *trx*, numerous other genes were found to exhibit similar homeotic phenotypes. These genes were first identified on the basis of their suppression of the *Pc* phenotype (Kennison and Tamkun, 1988). The mechanism of action of the proteins encoded by these genes was a mystery until it was discovered that one member of the trx-G, *brahma* (*brm*), encoded a protein with significant sequence similarity to the yeast general transcriptional activator Swi2/Snf2 (Tamkun *et al.*, 1992).

1.3.1 - The Swi/Snf complex and transcriptional promotion through nucleosome disruption

A group of genes in yeast, the SWI/SNF genes, appear to be responsible for the activation of transcription of a range of genes including those involved in mating type switching (SWI) and sucrose fermentation (SNF) (Peterson and Herskowitz, 1992). These genes encode proteins that form a large (~2 MDa) complex consisting of 11 subunits (reviewed in Peterson and Tamkun, 1995; Peterson, 1996). An insight into the mechanism of Swi/Snf complex action came from experiments which showed that second site suppressors of swi/snf$^-$ mutations include mutations in genes encoding histones and chromatin proteins. In order to test the possibility that the Swi/Snf complex mediates transcriptional activation by disrupting the repressive effects of chromatin, Hirschhorn *et al.*, (1992) showed that reduced levels of the core histones H2A and H2B were able to relieve the repressive effect of mutations in SWI2/SNF2 and SNF5 on the transcription of the SUC2 gene. Based on the assumption that reduced histone H2A and B levels would result in less compacted chromatin, one interpretation of these data is that Swi2/Snf2 and Snf5 normally act to de-compact chromatin. To test

this further these researchers analysed chromatin structure around the SUC2 promoter, in different mutant strains, in conditions where SUC2 transcription should be derepressed. Their results showed that in snf5$^-$ mutants, the region near the TATA box, which was susceptible to micrococcal nuclease cleavage in chromatin extracted from wild-type yeast, was resistant to cleavage, indicating protection by a nucleosome. In contrast, in the snf5$^-$/H2A and B mutant (reduced H2A and B levels), this site became susceptible to nuclease cleavage, indicating the absence of a nucleosome. These data correlate with the absence of SUC2 transcription in a snf5$^-$ mutant and presence of transcription in a snf/H2A, H2B mutant. Finally to show that these fluctuations in chromatin structure were not merely a consequence of the state of transcription of the SUC2 gene rather than having a causative effect, they showed that in a strain where there was no SUC2 transcription (due to a mutated TATA box), a snf5 mutation still rendered the region near the TATA box inaccessible to nuclease digestion, whereas in the wild-type case this region was accessible (Hirschhorn *et al.*, 1992). Further evidence for the idea that the Swi/Snf complex acts by disrupting chromatin structure came from the observation that suppressors of swi1, 2 and 3$^-$ mutants, SIN1 and 2, encode an HMG1 like protein (SIN1) and histone H3 (SIN2) (Kruger *et al.*, 1995), a core histone. Different mutations in SIN2 and another subsequently identified SIN mutant encoding histone H4, were mapped to regions which are responsible for DNA contact. It was predicted that DNA/histone octamer contacts in these mutants would probably be compromised (Kruger *et al.*, 1995). The result of this would be an unpackaging of chromatin. That unpackaged chromatin relieves the requirement for the Swi/Snf complex supports the idea that the normal function of the Swi/Snf complex is to disrupt nucleosome structure to allow transcription factor access and transcription.

1.3.2 - What is the mechanism of Swi/Snf nucleosome disruption?

How the Swi/Snf complex mediates this nucleosomal disruption is still largely unknown. Recently, it was shown that the purified Swi/Snf complex was able to bind to DNA with high affinity *in vitro* (Quinn *et al.*, 1996). Further, it was shown that an ATPase activity of the complex was stimulated when the complex bound to four-way junction DNA *in vitro* (Quinn *et al.*, 1996). It has been shown that the ATPase activity of the Swi/Snf complex is essential for its function of disrupting nucleosomes *in vivo* (Laurent *et al.*, 1993)and *in vitro* (Côté *et al.*, 1994). Four - way junction DNA has been postulated to resemble either the appearance of DNA as it enters and exits a nucleosome or as it kinks when it folds around the octamer (Peterson, 1996). Therefore, it is possible that the Swi/Snf complex recognises DNA packaged into nucleosomes and disrupts this packaging in an ATP dependent manner. How the Swi/Snf complex is targeted to specific promoters to disrupt nucleosomes is a question that remains to be answered, although some evidence pertaining to this topic is discussed below.

1.3.3 - A *Drosophila* Swi/Snf complex?

Given that Swi2/Snf2 shares significant sequence similarity with BRM and that Swi2/Snf2 is part of a large protein complex in yeast, is it possible that there is an homologous protein complex in *Drosophila*? The identification and characterisation of a *Drosophila* homologue of the yeast gene SNF5, named *snr1* (Dingwall *et al.*, 1995), and its subsequent characterisation suggested that there is. *snr1* was isolated by homology to the human homologue of SNF5. Antibodies against SNR1 were raised and used to show that SNR1 and BRM could be co-immunoprecipitated from *Drosophila* embryonic extracts. Furthermore, it was shown that both of these proteins were present in a very large, ~2 MDa, protein complex (Dingwall *et al.*, 1995). This corresponds well with the size of the yeast Swi/Snf complex and strongly suggests that there is indeed a *Drosophila* equivalent.

It is possible that the proteins encoded by the trx-G are the components of such a complex. However, the presence of TRX or any other of the trx-G proteins in the same complex as BRM and SNR1 has not been demonstrated. Further cloning and characterisation of the trx-G genes would clarify the question of whether or not the trx-G corresponds to the Swi/Snf complex and therefore maintains transcriptional activation in *Drosophila* by a similar mechanism. Even if there is an equivalent Swi/Snf complex in *Drosophila* it is quite possible that not all the members of the trx-G are contained within it. The phenotypes of trx-G mutations with respect to homeotic gene expression, although similar, are not exclusively suggestive of an effect on chromatin structure. It is formally possible that some members of the trx-G are responsible for transcriptional control independently, or that some members are involved at other levels of regulation (i.e. message stability). One member of the trx-G, *Trithorax-like* (*Trl*), encodes the *Drosophila* homologue of GAGA factor (Farkas *et al.*, 1994). Although *Trl* phenotypes are similar to other trx-G members, it seems possible that TRL can act independently of other trx-G proteins to stimulate transcription. This is based on experiments which showed that purified human GAGA factor is able to disrupt nucleosomes near a heat shock promoter in an ATP dependent manner *in vitro* (Tsukiyama *et al.*, 1994).

1.3.4 - Targeting the Swi/Snf complex to stimulate, not activate transcription

Although at least some members of the trx-G can disrupt nucleosome organisation to promote transcription, they do not directly activate transcription themselves. Rather, the current thinking is that the Swi/Snf complex (and the trx-G) 'open up' a promoter region for specific transcription factor access, and help to keep this region open for transcription to continue, if necessary. A significant question is how the Swi/Snf complex is targeted to promoters to have its effect? Recently it was shown that some components of the yeast Swi/Snf complex - Swi2/Snf2, Swi3, Snf5 and Snf11 - are constituents of the RNA polymerase II holoenzyme complex in yeast (Wilson *et al.*, 1996). Although these data have been questioned (Wang *et al.*, 1996), this would suggest these components are targeted to promoter regions by whatever mechanism targets the transcription initiation complex (i.e. TATA binding proteins), allowing a localised effect on nucleosomes. It appears that the Swi/Snf

complex in yeast (which can be purified separately) is part of a larger protein complex - the RNA polymerase II holoenzyme complex. Perhaps then the ~2 mDA complex containing the *Drosophila* homologues of Swi2/Snf2 (BRM) and Snf5 (SNR1) as observed by Dingwall *et al.*, (1995) may be contained within the *Drosophila* RNA polymerase II holoenzyme complex. Following from this and the observations that RNA polymerase II mutants behave in a similar ways to swi/snf mutants in yeast (Wilson *et al.*, 1996), it is possible that some of the yet to be characterised trx-G genes may encode other members of the RNA polymerase II holoenzyme complex, for example the SRB proteins.

In summary then, although it appears that the trx-G of genes may encode diverse proteins, at least some of them are related to proteins which in yeast appear to be constituents of a large protein complex, the RNA polymerase II holoenzyme complex. These and other data strongly suggest that at least some members of the trx-G encode proteins that bind indirectly at the promoters of many genes, including the homeotic genes in *Drosophila*, and promote transcription by disrupting nucleosome structure. Several important questions remain. The Swi/Snf complex requires gene specific activators to directly activate transcription of genes under their control. The same is probably true of *Drosophila*. For example, the segmentation gene products probably co-operate with some of the trx-G gene products to initiate transcription of homeotic genes. What mechanism maintains homeotic gene transcription in the absence of the segmentation gene products later in development? Are some trx-G encoded proteins able to do this even in the absence of gene specific activators, unlike in yeast? Alternatively, do other transcription factors, including the homeotic genes which are known to auto-regulate and cross-regulate (Chouinard and Kaufman, 1991; Gonzalez-Reyes *et al.*, 1990), maintain transcriptional activation with the help trx-G proteins?

As a separate question, it is clear from experiments including those on the *proboscipedia* (*pb*) gene from the ANT-C, that the levels of homeotic gene expression are crucial to segmental identity (Cribbs *et al.*, 1995). If and how the trx-G proteins regulate the levels of homeotic gene expression is unknown.

1.4 - The need for a stable repression system

The need for a stable repression system is very similar to the need for a stable activation system like that encoded by the trx-G. As mentioned above the homeotic gene products determine segmental identity. Their specific patterns of expression are established early in embryogenesis by the products of the segmentation genes. During gastrulation (~ 5hrs) these gene products are no longer present. As homeotic function is required throughout development (Morata and Kerridge, 1981; Struhl, 1981b), the specific patterns of homeotic gene expression must be maintained, throughout the rest of development. This is no mean feat, as repression is required to be maintained through many cellular divisions, and therefore through DNA replication.

1.4.1 - The Pc-G is responsible for maintenance, not initiation, of gene repression

As was described above, *Pc* was identified because the mutant phenotype resembles gain-of-function homeotic mutations. Strong *Pc* mutations are embryonic lethal showing transformation of all segments towards the eighth abdominal segment (A8) (Lewis, 1978). This is consistent with a ubiquitous and ectopic expression of the gene responsible for A8 identity, *Abd-B*. Weaker (haploinsufficient) *Pc* mutations are viable and show mild homeotic transformations in adults including the characteristic extra sex combs phenotype. Normally, male flies have an appendage on the first leg only, known as the sex comb. The extra sex combs phenotype is characterised by the presence of sex combs on the second and third legs also. This phenotype is indicative of ectopic expression of the homeotic gene *Sex combs reduced* (*Scr*). SCR is normally only present in the first thoracic segment (T1), but in *Pc* mutants it is also present in T2 and T3 (Pattatucci and Kaufman, 1991). Given that homeotic expression patterns are correctly initiated in Pc mutant embryos (Simon *et al.*, 1992), these data show that *Pc* function is required for maintenance of repression of homeotic gene expression.

By screening for mutations that resulted in enhancement of the homeotic phenotypes of *Pc* (and other subsequently identified Pc-G mutants), twelve other genes were identified that are functionally similar to *Pc* and hence constitute the Pc-G. These are *extra sex combs* (*esc*) (Struhl, 1981a), *Polycomblike* (*Pcl*) (Duncan, 1982), *super sex combs* (*sxc*) (Ingham, 1984), *polyhomeotic* (*ph*) (Dura *et al.*, 1987), *Additional sex combs* (*Asx*) (Jürgens, 1985), *Posterior sex combs* (*Psc*) (Jürgens, 1985), *Sex combs on midleg* (*Scm*) (Jürgens, 1985), *Sex combs extra* (*Sce*) (Breen and Duncan, 1986), *Enhancer of zeste/polycombeotic* (*E(z)/pco*) (Jones and Gelbart, 1990; Phillips and Shearn, 1990), *pleiohomeotic* (*pho*) (Girton and Jeon, 1994), *multi sex combs* (*mxc*) (Santamaria and Randsholt, 1995) and *l(3)SG23* (this study). Two genes that did not fall strictly into the Pc-G category because although they enhance Pc-G phenotypes, they have no detectable homeotic phenotypes of their own, are *Suppressor of zeste 2* (*Su(z)2*) (Wu and Howe, 1995) and *Enhancer of Polycomb* (*E(Pc)*) (Sato *et al.*, 1984). The latter has subsequently been shown to be a suppressor of position effect variegation (*Su(var)*) (H. Brock, personal communication).

1.4.2 - Pc-G proteins are novel

Sequence analysis of Pc-G genes has shown that all of these genes encode novel proteins. Curiously, none of the proteins contain known DNA binding motifs. In some cases similarity to non-Pc-G proteins identifying conserved domains provides clues as to the mechanism of action of the Pc-G.

chromodomain

Most notably, PC contains a 37 amino acid sequence motif towards its amino-terminus with direct similarity to the protein HP1, termed the chromodomain (Paro and Hogness, 1991). HP1 is a non-histone, heterochromatin associated protein encoded by the gene *Su(var)205*. Since

heterochromatin can lead to the stochastic repression of an artifactually juxtaposed region of euchromatin, an effect referred to as Position Effect Variegation (PEV), it was postulated that PC, and other Pc-G proteins, might repress their target genes by a similar mechanism. Specifically the mechanism proposed involved local modification of the chromatin structure to a compacted (silent) form, of a type that had been observed in heterochromatin regions of the genome (Paro, 1990; see section 1.5.1) . The chromodomain has been shown to be necessary to localise PC to specific sites on polytene chromosomes (Messmer *et al.*, 1992) but not for repression (Müller, 1995). Further data showed that the differences between the chromodomains of PC and HP1 were sufficient to account for the specificity of these two proteins for regions of euchromatin or heterochromatin, respectively, as assayed by polytene chromosome immunostaining (Platero *et al.*, 1995). Given that PC probably does not bind directly to DNA, it seems likely that the chromodomain is involved in specific protein-protein interactions.

SET domain

E(Z) contains a region of similarity, at its carboxy terminal end, to TRX (Jones and Gelbart, 1993) and the product of the gene *Su(var)3-9* (Tschiersch *et al.*, 1994), termed the SET domain. Firstly, the similarity to TRX, a protein involved in an apparently opposite role, suggested that there may be some level of antagonism or competition between these two proteins for a common target and that therefore E(Z) and TRX may represent components of a system which 'decides' whether a gene should be active or silenced (Jones and Gelbart, 1993). Secondly, the similarity between SU(VAR)3-9 and E(Z) once again infers that the mechanism of silencing by the Pc-G and that seen in PEV is similar (see later). Although it was suggested that the SET domain was either a nucleic acid or protein binding domain (Jones and Gelbart, 1993), subsequent experiments argue strongly against the former possibility (R. Jones, personal communication).

RING and PHD finger

A wide variety of proteins from many species, including the Pc-G protein PSC, contain a novel type of zinc chelating structure known as the RING finger. This is characterised as Cys3 - His - Cys4 and has been shown to co-ordinate two zinc ions (Arnim and Deng, 1993). Although there is some primary sequence resemblance of the RING finger to zinc finger DNA binding motifs, RING finger proteins are not found exclusively in the nucleus of cells (Song and Donner, 1995), and have been shown to be responsible for protein-protein interaction (Borden *et al.*, 1995). At the primary sequence level, PHD fingers - Cys4 - His - Cys3 (Aasland *et al.*, 1995; Stassen *et al.*, 1995) - appear very similar to RING fingers. PHD fingers have been found in numerous nuclear regulatory proteins including POLYCOMBLIKE (PCL) and TRX, which each contain two of these motifs. Given that TRX contains another domain which appears to be responsible for its DNA binding activity (Stassen *et al.*, 1995), it seems unlikely that the PHD fingers are responsible for DNA binding. Perhaps, as for the RING fingers, the PHD fingers are responsible for protein-protein interaction. This or any other function of the PHD fingers has yet to be determined. The presence of a conserved motif in TRX and a Pc-G protein, PCL, once again raises the possibility of competition or antagonism between these two

classes of proteins. The PHD fingers will be discussed specifically with respect to PCL in section 1.6.1.

WD40 repeats

ESC contains either 5 copies (Simon *et al.*, 1995) or 6 copies (Gutjahr *et al.*, 1995) (depending on the method used for alignment) of a motif known as WD40 repeat. This motif is characterised by a very loose consensus including 15 conserved residues over an approximately 40 residue stretch (van der Voorn and Ploegh, 1992). Although this repeat has been found in functionally diverse proteins, significantly, it has been found in the yeast transcriptional repressor protein Tup1 (7 repeats), where it is responsible for a direct protein interaction with the α2 homeodomain mating type transcriptional repressor protein (Komachi *et al.*, 1994) . It is possible therefore that ESC might also use this domain to interact with other repressor proteins.

The presence of putative protein-protein interaction domains in several Pc-G members supports the possibility that at least some members of the Pc-G form a protein complex.

1.4.3 - Pc-G proteins form a multimeric complex

As mentioned above, Pc-G mutants interact genetically with each other; double heterozygotes have more severe homeotic phenotypes. This interaction is often synergistic; double heterozygous mutants have stronger phenotypes than the sum of the phenotypic effects in each single heterozygous mutant (Campbell *et al.*, 1995). This suggests that either Pc-G proteins are components of a pathway that leads to repression, or that these proteins form a multimeric protein complex that represses genes. More specifically, the reduction in dose of a gene product, caused by heterozygosity for a mutation in that gene, could reduce the rate at which a pathway would process information or a complex could assemble. Reducing the dosage of a second protein would further disrupt either of these processes resulting in increased phenotype severity. It has been proposed that Pc-G proteins may assemble and disassemble by the laws of mass action into a protein complex (Locke *et al.*, 1988). This model implies that Pc-G proteins assemble into a complex that is very sensitive to levels of the individual components of the complex. A small reduction in dose of one component pushes the equilibrium towards complex disassembly. Conversely then, an increase in dosage of one component might increase the rate of complex assembly. It is difficult to envisage what effect the over-formation of Pc-G complexes would have. It is possible however that this may lead to spurious repression of some genes that are not supposed to be repressed by the Pc-G. There is some evidence that this may occur. Overexpression of *Psc* and *Su(z)2* results in bristle abnormalities that are reminiscent of reduced levels of NOTCH (Sharp *et al.*, 1994). Furthermore, overexpression of *Su(z)2* resulted in the reduced expression of a lacZ enhancer trap inserted in the gene *neuralised*. These data show spurious repression, by Pc-G proteins, of genes which are not normally under their control. This is consistent with the ectopic localisation of these two proteins at hundreds of polytene chromosome loci (these two proteins are normally only found at 60-80 chromosome sites - see next

section; Sharp *et al.*, 1994). Overexpression of *esc* from an *esc* containing transgene results in anteriorly directed transformations in adult *Drosophila*. These phenotypes are similar to those observed in trx-G mutants and indicate, therefore, that in this case extra levels of ESC may be ectopically repressing homeotic gene transcription (Campbell *et al.*, 1995). These transformations were not enhanced in flies heterozygous for mutations in *brm* or *trx*. However, given that flies heterozygous for *brm* or *trx* mutations show no transformations themselves, these mutations may not represent an adequate reduction in BRM or TRX levels for the visualisation of an interaction (Campbell *et al.*, 1995).

In cases where the assembly of the Pc-G complex is reduced by a diminished dose of a single constituent of the complex, it may be expected that increased dosage of a different component might drive the equilibrium back towards complex assembly, thereby suppressing the phenotypes associated with the initial mutation. This has been observed in several cases. Increased dose of PCL suppresses the phenotypes associated with reduced levels of PC (Kennison and Russell, 1987). Similarly increased dosage of PH suppresses the posteriorly directed transformations associated with *Pc* and *Pcl* mutations (Cheng *et al.*, 1994). Interestingly, duplications containing *Asx* and *Pc* weakly enhanced the extra sex combs phenotype associated with flies heterozygous for *Pcl* mutations (Campbell *et al.*, 1995). These data and that which shows that most duplications including Pc-G genes have no effect on the phenotypes associated with *Pc* mutations (Kennison and Russell, 1987), indicate that if the mass action model is true there must be some caveats placed upon it. For example, it may be that there is not just one type of Pc-G complex. There may be several types of complexes with different components that have different roles (for evidence of this see later). This may explain why some Pc-G mutants or duplications do not interact (Campbell *et al.*, 1995). Alternatively, to explain this observation, it may be that not all Pc-G proteins are equal. Although all Pc-G members are required for at least some function (hence their mutant phenotypes), it may be that there are some Pc-G proteins that are the key to the initiation of protein complex formation (see next section).

Although all the genetic evidence presented above strongly suggests that Pc-G proteins form a complex, it is the combination of these data with a range of molecular data that provides compelling evidence for this hypothesis. Most notable are the data which show that PC and PH are found in the same large (2-5 mDa) multimeric protein complex in *Drosophila* embryonic nuclear extracts (Franke *et al.*, 1992). Immunoprecipitation with anti-PC antibodies showed that 10-15 proteins are co-immunoprecipitated with PC. One of these proteins was shown to be PH. Furthermore, by size exclusion chromatography, most of the PC and PH immunoreactivity in these nuclear extracts was found to reside in a protein complex that was between 2 and 5 mDa in size. These data are suggestive of a few things. Firstly, they show good evidence that there is indeed a large protein complex that may correspond to the Pc-G complex. Secondly, given that the protein extracts were of soluble nuclear proteins, this protein complex either forms prior to chromosome binding or is weakly associated with chromosomes. Furthermore, there is some evidence that PCL is co-immunoprecipitated with anti-PC antibodies (Lonie, 1994), PH is co-immunoprecipitated with anti-E(Z) antibodies (R. Jones, personal communication) and that PCL interacts directly with E(Z) (chapter 4).

polytene chromosome immunostaining

Polytene chromosome immunostaining data add weight to the hypothesis that Pc-G proteins form a protein complex. PC, PH and PCL have been shown to localise to exactly the same ~100 sites on polytene chromosomes (DeCamillis *et al.*, 1992; Franke *et al.*, 1992; Lonie *et al.*, 1994; Zink and Paro, 1989). Although cytological resolution in these assays has been estimated to be in the order of 100-200 kb (Rastelli *et al.*, 1993), these data are supportive of the co-immunoprecipitation data discussed above. In addition PSC is found at 83 sites, 63 of which appear to coincide with PC/PH/PCL sites and SU(Z)2 is found at 89 sites of which 52 overlap with the PSC sites and 48 with the PC/PH/PCL sites (Rastelli *et al.*, 1993). Although, in this paper double immunofluorescence was not carried out, so it is possible that there is more or less overlap between the sites. Aside from supporting the 'complex' model these polytene data provide other important information. Firstly, although the proteins are found at the cytological sites of the ANT-C and BX-C as expected, they are also present at many other sites. This indicates that Pc-G proteins may have a more general role than repression maintenance of homeotic gene transcription. In fact there is ample data to support this assertion. Pc-G mutants have pleiotropic phenotypes including non-homeotic phenotypes. For example *E(z)* mutants have imaginal disc defects (Phillips and Shearn, 1990), *pho* and *ph* mutants show nervous system defects (Girton and Jeon, 1994; Smouse *et al.*, 1988), *mxc* mutants show germline proliferative defects (Docquier *et al.*, 1996) and phenotypes associated with tumour suppressors (Santamaria and Randsholt, 1995) and several Pc-G mutants show segmentation defects (Breen and Duncan, 1986; Ingham, 1984). At a molecular level it has been shown that *engrailed* (*en*) is ectopically expressed in various Pc-G mutants (Moazed and O'Farrell, 1992). Note that PC/PH/PCL/PSC proteins are present at 48A (DeCamillis *et al.*, 1992; Franke *et al.*, 1992; Lonie *et al.*, 1994; Rastelli *et al.*, 1993) which contains the *en* locus (Lindsley and Zimm, 1992). Secondly, the lack of complete overlap in the localisation of PSC and SU(Z)2 with PC/PH/PCL, suggests that any model of Pc-G function must allow for different Pc-G protein complexes with different components at different loci.

1.4.4 - Assembly of Pc-G protein complexes

An assimilation of a large amount of data gives a skeletal picture of how the Pc-G protein complex may assemble.

Firstly, it appears that ESC may have a unique role. It has been shown that *esc* function is required only early in development (6-8 hrs) unlike other Pc-G gene function which is required throughout development (Struhl and Brower, 1982). This early role is supported by the evidence that ESC is present highly only early in embryogenesis (Gutjahr *et al.*, 1995), and that rescue of *esc* null mutant associated embryonic lethality can only occur when *esc* is expressed, under heat shock control, at or before 4 hrs of embryogenesis (Simon *et al.*, 1995). Interestingly, this early window of *esc* function coincides with the time in embryogenesis when the initial repressors of homeotic gene transcription, the gap gene products, are lost. The yeast protein Tup1 is involved in transcriptional repression of several genes in yeast by interacting with several protein partners via specific WD40 contained within it (see above). Similarly, it was shown that individual WD40 repeats are required for

ESC function *in vivo*, indicating that specific WD40 repeats may interact with specific protein partners (Simon *et al.*, 1995). Therefore, it has been proposed that the role of ESC is to interact specifically with the homeotic gene transcriptional repression initiation proteins (e.g. the gap gene products HUNCHBACK and KRUPPEL) and recruit the Pc-G protein complex to establish stable silencing of the locus (Simon *et al.*, 1995).

Secondly, it does appear that a large protein complex (2 - 5 mDa) containing PC, PH and also presumably at least PCL and E(Z) is present in extracts of soluble nuclear proteins (see above). This is strongly suggestive that a Pc-G protein complex, or at least part of the final complex, assembles in solution and then binds to loci that are to be silenced. Some of the members of the complex may bind directly to ESC.

Thirdly, it appears that there is some kind of hierarchy in complex assembly. Although all of the Pc-G proteins appear to be required for wild - type function, it seems that boosting the concentrations of only some Pc-G members can provide the impetus for additional complex assembly (see above). Overexpression of *Psc* , *Su(z)2* and *esc* causes ectopic repression of transcription (see above) indicative of an increased rate of complex formation. Similarly only duplications in *Pcl* and *ph* were able to suppress the phenotypes associated with mutations at other Pc-G genes, indicating that increased levels of these proteins were able to 'rescue' reduced complex assembly. Alternatively, PSC, SU(Z)2 and ESC may be able to act as repressors independently when overexpressed, and there may be some redundancy which allows PCL and PH to substitute for the function of other Pc-G proteins. Analysis of the distribution of other Pc-G proteins on polytene chromosomes of larvae overexpressing PSC and SU(Z)2 would help to address the first of these alternative hypotheses. In order to address the second alternative, it may be necessary to attempt rescue of embryos that are completely null for one Pc-G member with a duplication in *Pcl* or *ph*. If rescue was possible it would suggest that PCL or PH could indeed functionally substitute for that absent member.

Lastly, by artificially tethering PC by the use of a GAL4 DNA binding domain, upstream of a reporter gene, it was shown that PC was able to initiate and maintain silencing of the reporter gene throughout *Drosophila* embryogenesis (Müller, 1995; see section 1.4.5). This silencing was dependent on wild type *Pc*, *Asx*, *Pcl*, *Scm*, *Sce*, *E(z)* and *esc* function. These data suggest a few things. Firstly and most importantly, it appears that PC can recruit a silencing complex to chromatin when it is artificially tethered. Secondly, it may suggest that ESC is not only required for attaching a Pc-G complex to chromatin, but also for silencing. More likely is the possibility that GAL-PC can interact with ESC which then recruits the endogenous Pc-G complex (including endogenous PC) to the reporter gene to effect silencing. This is consistent with the idea that the endogenous Pc-G complex exists already assembled in solution.

In summary, it appears that ESC may bind to chromosomes first and then recruit at least a partially pre-formed Pc-G complex. This complex probably assembles according to the laws of mass action, although some members, e.g. PCL and PH, may be more critical to initiate assembly possibly because they contain many protein interacting faces, for example. It has been suggested that a complex that forms by the laws of mass action would be too unstable to stably repress transcription

(Henikoff, 1996). As an alternative model Henikoff (1996) proposed that there are sites in chromatin that bind certain proteins. These sites act as anchorage sites which interact cooperatively with other proteins causing a compaction of the local chromatin structure. This model need not be mutually exclusive to the one presented above. It is possible that these anchorage sites occur, and partially pre-formed Pc-G complexes interact with these sites and bring them together resulting in repression. There is no molecular evidence for this however.

1.4.5 - How does the Pc-G recognise loci to be silenced?

All the presently characterised Pc-G genes are expressed ubiquitously, spatially and temporally, during embryogenesis - their expression pattern contains no positional information (Bornemann *et al.*, 1996; Franke *et al.*, 1992; Jones and Gelbart, 1993; Lonie *et al.*, 1994; Martin and Adler, 1993). In contrast the homeotic genes which are regulated by the Pc-G are differentially transcribed along the antero-posterior axis. This means that in any one parasegment, the Pc-G must only repress some homeotic genes and not others. This decision is obviously not mediated by the presence or absence of Pc-G proteins in that parasegment, so, how does it occur? Various groups have recently found, while analysing control regions of homeotic genes, that as well as DNA elements that are responsible for the initiation of homeotic gene transcription (initiation elements), there are elements that are responsible for transcriptional repression maintenance (maintenance elements). The transcriptional repression maintenance conferred by these elements is dependent on and therefore mediated by the Pc-G.

By fusing regulatory fragments from the BX-C (Simon *et al.*, 1993), *Antennapedia* (*Antp*) (Zink *et al.*, 1991), *Ubx* (Chan *et al.*, 1994; Müller and Bienz, 1991), *Abdominal - B* (*Abd-B*) (Busturia and Bienz, 1993) and *Scr* (Gindhart and Kaufman, 1995) upstream of a lacZ reporter gene it has been shown that there are two types of transcriptional control element. Different control fragments can initiate transcription of the reporter gene in embryos in patterns that mimic subsets of the entire pattern corresponding to the homeotic gene from which the fragment came. These initiation elements presumably contain binding sites for gap and pair-rule gene products which repress and activate transcription in a parasegmental specific manner. In some cases these initiated patterns degenerate later in embryogenesis - there is stochastic derepression of lacZ transcription. This derepression is reminiscent of the effect of mutations in Pc-G genes. Some of the control fragments, however, as well as initiating specific expression patterns were able to maintain repression throughout embryogenesis. This repression was shown to be dependent on the wild-type function of several Pc-G genes, and therefore, these maintenance elements have been termed Pc-G Response Elements (PRE) (after Simon *et al.*, 1993). Since PC (Chiang *et al.*, 1995; Zink *et al.*, 1991) , PCL (Lonie *et al.*, 1994) and PSC (Chan *et al.*, 1994) have been shown be present at the insertion sites of transgenes containing these fragments, as assayed by polytene chromosome immunostaining, it seems clear that Pc-G proteins act directly through these DNA elements. Furthermore, these PRE fragments appear to be modular in nature, as they are able to confer stable repression maintenance when linked to heterologous initiation elements (Chiang *et al.*, 1995). Also note that at present PREs have not been

divisible into small restriction fragments. Smaller fragments lose their PRE activity (unpublished results cited in Pirrotta, 1995). Therefore, consensus sequences required for PRE activity have yet to be determined.

PREs can confer repression maintenance when linked in *cis* to initiation elements. Note that this is only repression maintenance - essentially only the transcriptional repression which is initiated by the initiation element is faithfully preserved. Therefore, it appears that a PRE acts as a 'docking site' for attachment of the Pc-G complex. If no PRE is present then a Pc-G complex cannot 'dock' to maintain the established repression state. However, the presence of a PRE alone is not sufficient for Pc-G repression to establish - a gene must also have been repressed initially by a transient mechanism.

There are two main models which attempt to explain how the Pc-G is recruited to a transiently repressed gene that contains a PRE. The first model is that some components of the Pc-G complex interact specifically with the transient repression proteins (Zhang and Bienz, 1992). This protein interaction recruits the complex to the repressed gene where it is stabilised by an interaction with the PRE. Presumably this interaction would be stable in the absence of the original transient repressor protein, as would occur late in embryogenesis. The second model suggests that either directly or indirectly, the Pc-G recognises some as yet undetermined aspect of a target gene's transcriptionally repressed state (e.g. silent chromatin conformation). This recognition process, however it occurs, results in a stable interaction between the Pc-G complex and the PRE (Chan *et al.*, 1994).

model 1

Three separable control regions are responsible for the sub-patterns of the normal *Ubx* expression pattern. These control regions are called ABX, PBX and BXD (Müller and Bienz, 1991; figure 1.4.5.1). Whereas the normal expression domain of *Ubx* is from parasegment 5 to parasegment 13 inclusively, ABX, fused to a lacZ reporter, directed expression of lacZ in parasegments 5, 7, 9, 11 and 13, PBX directed expression in parasegments 6, 8, 10 and 12 and BXD directed expression from 'head to tail' (i.e. parasegments 1 - 14) (Müller and Bienz, 1991). Note that the boundaries of lacZ expression in strains carrying ABX or PBX constructs are not maintained in late embryogenesis. It was further shown that the linking of the three control regions resulted in the normal *Ubx* expression pattern, with the boundaries of expression maintained. This indicated that the ABX and PBX expression patterns are additive, and that the repression initiation conferred by these two control fragments (i.e. anterior to parasegment 5 and posterior to parasegment 13) is dominant to BXD, with the maintenance of repression conferred by the BXD fragment. In order to test the role of *Pc* in the repression from these fragments, lacZ expression was analysed in embryos that contained different combinations of the fragments in a Pc^- mutant background. It was found that in the absence of *Pc* function, repression maintenance conferred by the BXD fragment was lost. This indicates that there is a PRE within the BXD fragment. Given that *hb* is expressed in the anterior of the embryo early, it was postulated that HB may be responsible for the repression initiation seen with the ABX and PBX fragments. In fact it was found that there were numerous HB binding sites within the PBX fragment (Zhang *et al.*, 1991) and that joining synthetic HB DNA binding sites next to the BXD fragment could initiate lacZ expression in the normal *Ubx* domain (except with no repression of lacZ in parasegment

14), with maintenance of this anterior *hb* mediated repression (Zhang and Bienz, 1992). These and other data have suggested that HB may directly recruit the Pc-G complex to make the transition to stable silencing of *Ubx* transcription in the anterior of the embryo (Zhang and Bienz, 1992)

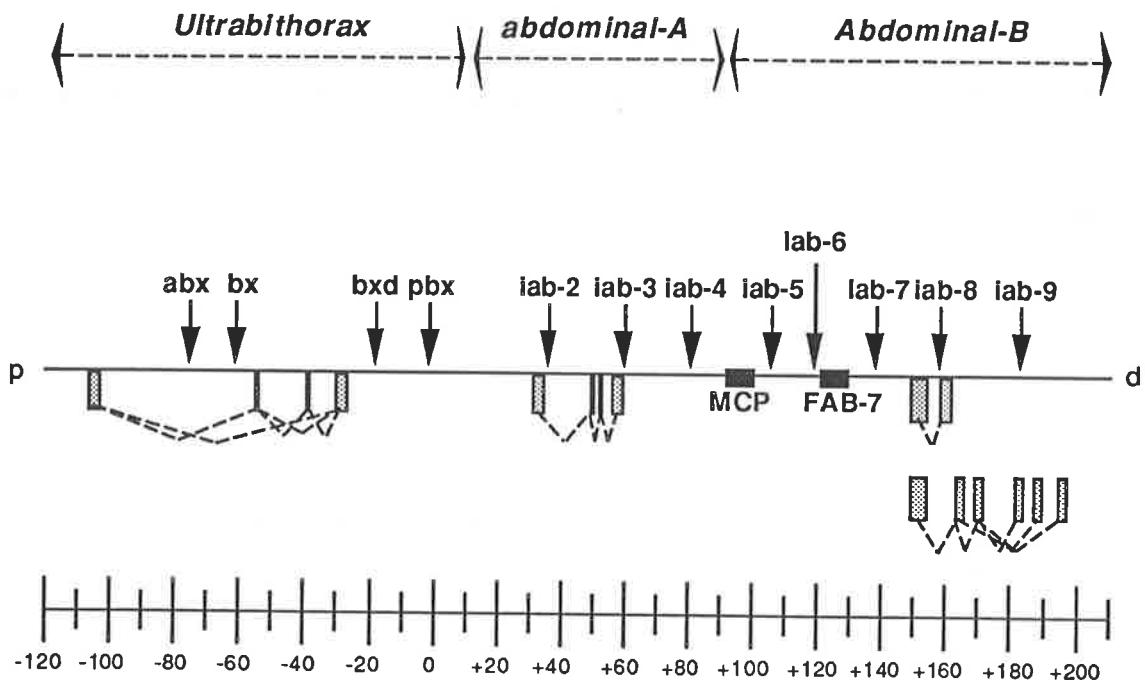


figure 1.4.5.1 - The Bithorax Complex region (taken from Zink and Paro, 1995)
This is a schematic representation of the BX-C. The top line shows the limits of the *Ubx*, *abd-A* and *Abd-B* genes. The middle line shows the positions of the *cis* - regulatory elements, MCP and FAB - 7 fragments and the transcription units (p - proximal, d - distal). The bottom line is a scale in kb.

Experiments by Müller (1995) (as described briefly in chapter 1.4.4) are consistent with this model. PC fused to the DNA binding domain of GAL4 (GAL-PC) was expressed in embryos under the control of the *hb* promoter such that GAL-PC was produced at the same time and in the same anterior region of the embryo as endogenous HB. When GAL4 DNA binding sites (UAS) were fused with the BXD, repression of lacZ transcription in the anterior of the embryo was initiated and maintained throughout embryogenesis. The maintenance was dependent on wild-type Pc-G function (Müller, 1995). Given that this could occur in the absence of endogenous HB, it was concluded that the physical tethering of PC next to a PRE bypassed the requirement for normal HB tethering of the Pc-G.

model 2

The second model suggests that the Pc-G complex can bind to a PRE, and repress transcription, at a locus by virtue of some generic aspect of its silent transcriptional state. In support of this it was shown that a fragment from *Ubx* containing a PRE and no patterning activity (including no detectable HB binding sites) was able to create a novel PSC binding site on polytene chromosomes (Chan *et al.*, 1994). Whether this fragment is able to confer Pc-G mediated repression is unknown because, if the reporter gene is not initially patterned, repression maintenance cannot be assayed. In

accord with these data is the observation that transgenes with two PRE containing fragments from the *Abd-B* control region with no enhancer activity, MCP and FAB - 7 (Busturia and Bienz, 1993), create novel PC binding sites in polytene chromosomes and effect repression upon a heterologous GAL-4 - UAS activated reporter gene (Zink and Paro, 1995). In order to test further whether a PRE can establish Pc-G mediated silencing, one would need to fuse a PRE with an enhancer fragment that conferred activation on a reporter gene at a late stage of development, well after the Pc-G proteins become active. If the Pc-G can bind to the PRE early in development while the promoter of the reporter gene is silent, then presumably the reporter would not be activated at its normal stage later in development. A series of experiments by Poux *et al.*, (1996), show exactly this. Previously it has been shown that as well as the enhancer activity that establishes the expression pattern of *Ubx* in the early (blastoderm) stage embryo, there are separate enhancers that activate *Ubx* during later embryonic and larval stages in the imaginal discs. One of these sets of enhancers resides in the ABX region (Simon *et al.*, 1990), and the other resides between the PBX region and the main *Ubx* PRE (Pirrotta *et al.*, 1995). Although the expression of *Ubx* during larval stages respects the parasegment 5 and 13 boundaries that were established during the blastoderm stage, HB (which initiates the PS 5 boundary) and TAILLESS (TLL) (which initiates the PS 13 boundary) are no longer present. When tested with a reporter gene in isolation these enhancers drive expression in imaginal discs in a parasegmental non-specific way (Poux *et al.*, 1996). Presumably, the parasegmental specific boundaries of Ubx expression late in development are maintained from the time that they were established during embryogenesis by the Pc-G. As well as the imaginal enhancers, Poux *et al.*, (1996) characterised an enhancer fragment (2218R6) that is active just after the blastoderm stage (when most of the other enhancers are active) at germ band extension. When this enhancer (2218R6) or the imaginal enhancers are coupled with a PRE in isolation, activity from these enhancers is strongly repressed in a *Pc* dependent manner throughout the animal. These enhancers were shown not to contain any *in vitro* detectable HB binding sites, but this is irrelevant, as they are silenced in regions of the embryo that do not contain HB. In contrast, when these enhancers are coupled with a PRE and early (blastoderm) enhancers, the late enhancers are Pc-G silenced only in the regions of the animal that were transcriptionally inactive during the blastoderm stage (due to the transient repressive effects of HB and TLL) (Poux *et al.*, 1996). These data strongly suggest that the Pc-G complex is active around the blastoderm stage and binds to PRE fragments that are in transcriptionally inactive chromatin.

Although mounting data strongly suggest that the Pc-G recognises transcriptionally silent, PRE - containing genes early in development, the mechanism for such a recognition is unknown. It presumably occurs via protein interaction and given its early role (see above) probably involves ESC. One way to identify the proteins involved in a presumptive interaction with ESC, would be to screen for proteins that interact with individual WD 40 repeats within ESC, given that it appears that each of the 5 WD 40 repeats are essential (Simon *et al.*, 1995).

1.5 - Mechanism of Pc-G silencing

The mechanism by which the Pc-G silences gene transcription is largely unknown. However, experimental data from a wide range of sources have shed light on this mechanism. Firstly, it has been shown that mutations in some Pc-G genes (*Asx*, *Pcl*, *Psc* and *E(Pc)*) consistently modify PEV, while mutations in some others (*Pc*, *Scm* and *E(z)*) modify PEV variably (H. Brock, personal communication). Secondly, there are molecular similarities between some Pc-G members (PC and E(Z)) to proteins encoded by *Su(var)* genes. This is consistent with the fact that mutations in these genes modify PEV. These data strongly suggest that the mechanisms of gene silencing by the Pc-G and in PEV are overlapping.

1.5.1 - Position Effect Variegation (PEV) and gene silencing

In general, eukaryotic genomes contain two types of chromatin that appear cytologically different. Euchromatin contains most of the single copy DNA and mutable (expressed) genes. Euchromatin replicates right through S - phase and decondenses during interphase. In contrast, heterochromatin contains most of the repetitive DNA, is always condensed and replicates at the end of S - phase. PEV is the phenomenon of variable but heritable gene silencing that occurs to a euchromatin gene when juxtaposed with heterochromatin by chromosome rearrangement (reviewed in Karpen, 1994). Right from the beginning, the observation that heterochromatin appears cytologically more condensed led to the concept that heterochromatin is more compacted than euchromatin. Further, the observation on *Drosophila* polytene chromosomes that euchromatin artificially juxtaposed to heterochromatin is banded diffusely and apparently highly compacted has led to the idea that heterochromatin 'spreads' into the juxtaposed euchromatin causing euchromatin gene inactivation by compaction. This model has been elaborated by recent molecular data which suggest that heterochromatin is composed of large, multimeric protein complexes that are responsible for the packaging of heterochromatin into its compacted state (Grigliatti, 1991). By modifying the doses of these proteins (encoded by *Suppressors of variegation* (*Su(var)*)) or the regulators of complex assembly (encoded by *Enhancers of variegation* (*E(var)*)) the extent of spread of heterochromatin into juxtaposed euchromatin can be altered, either decreasing or increasing, respectively, the frequency at which the euchromatin gene becomes silenced (Locke *et al.*, 1988). The presence of the heterochromatin protein HP1 in PEV repressed euchromatin was demonstrated by immunostaining heterochromatin-induced condensed euchromatin using anti-HP1 antibodies (Belyaeva *et al.*, 1993).

There are however, significant problems with this simplistic view of both heterochromatin and PEV. Firstly, heterochromatin is not transcriptionally inert. There are at least 21 mutable and hence expressing genes within or near the heterochromatin of chromosomes 2 and 3 in *Drosophila* (cited in Devlin *et al.*, 1990) . Secondly, the discontinuous compaction of some variegating euchromatin is inconsistent with the spreading model for PEV (Belyaeva and Zhimulev, 1991).

1.5.2 - Heterochromatin is not merely an inert chromatin

As mentioned above *Drosophila* heterochromatin contains at least 21 genes that are expressed at some stage during development. This number may be significantly higher as a recent study showed that a P-element mutagenesis carried out in a *Su(var)* mutant background returned large numbers of new mutations at diverse heterochromatic sites (Zhang and Spradling, 1994). This screen was carried out on the basis that in a *Su(var)* background, heterochromatin may be less compacted and therefore more accessible to P-element insertion. Additionally, it has been shown that there are 'islands' of complex DNA within the repetitive DNA of heterochromatin in *Drosophila* (Le *et al.*, 1995).

The study of one particular heterochromatic gene in *Drosophila*, *light* (*lt*) which is necessary for eye pigmentation, is intriguing. PEV which involves the juxtaposition of a euchromatic gene into heterochromatin is enhanced by the absence of the Y chromosome and suppressed by the presence of an additional copy of the Y chromosome. This, it has been speculated, is due to the fact that the Y chromosome is predominantly heterochromatic and therefore, less or more Y chromosomes make available or consume components of heterochromatin thereby enhancing or suppressing the formation of heterochromatin elsewhere (Karpen, 1994). Chromosomal rearrangements that move *lt* to euchromatic regions also result in PEV, which , as for inversions that affect the *white* (*w*) gene, can be monitored as pigment level changes in the cells of the adult eye (Wakimoto and Hearn, 1990). Interestingly, modifiers of euchromatic PEV act in the opposite way on these *lt* rearrangements. Mutations that normally suppress PEV (*Su(var)*) now enhance PEV and vice versa (unpublished results as cited in Devlin *et al.*, 1990). Also, the copy number of the Y chromosome has the opposite effect (Baker and Rein, 1962). These data indicate that although heterochromatin consists of proteins that are encoded by the *Su(var)* genes, and these proteins repress the transcription of variegating euchromatin genes, their normal role in heterochromatin is probably not in gene repression. It appears that heterochromatin may be a completely 'functional' and different type of chromatin to euchromatin.

1.5.3 - Silencing by PEV is not mediated by inaccessible, compacted chromatin

There is conflicting, published molecular evidence regarding a compaction model for PEV mediated gene silencing. It was shown that there are only very minor changes in DNase I sensitivity and nucleosome organisation associated with the euchromatic variegating allele of *w*, $In(1)w^{m4}$ in *Su(var)* and *E(var)* genetic backgrounds (Hayashi *et al.*, 1990). In a similar way it was shown that access to restriction endonuclease sites at the *w* gene in chromatin extracted from $In(1)w^{m4}$ flies was not altered in *Su(var)* or *E(var)* genetic backgrounds (Schlossherr *et al.*, 1994). These researchers hand dissected eyes from adult, female flies that carried the mutation $Su\text{-}var(2)1^{01}$ (very strong suppression of variegation - eyes almost wild - type) and the mutation $En\text{-}var(2)1^{01}$ (very strong enhancement - eyes almost white) and extracted the nuclei. The whole nuclei were subjected to restriction enzyme digestion with three different enzymes, one of which cuts within an exon of the *w*

gene and the other two of which cut in the 5' flanking region of the *w* gene. They showed that although no restriction site was accessed 100% of the time, there was no difference in the accessibility to a particular site between genetic backgrounds. This strongly indicates that the mechanism that is involved in the repression of the *w* gene in the *En-var(2)1*^{01} background does not involve changes to the chromatin accessibility of an exogenous protein (Schlossherr *et al.*, 1994). A different group show apparently contradictory results. In this study, however, a naturally variegating euchromatic allele was not used. Rather, a P-element containing the *w* gene and a heat shock inducible *hsp26* gene was mobilised from a euchromatic site on the X chromosome and lines with variegated eyes were recovered. Of these lines, several had the P-elements within pericentric heterochromatin or on the largely heterochromatic Y (fourth) chromosome, as might be expected. These variegating lines responded to extragenic modifiers of PEV in a similar way to 'normal' variegating alleles and were thus considered analogous to these systems. It was then shown that access to a single Xba I site, in nuclei extracted from whole third instar larvae, in the *hsp26* gene (i.e. not the same site as was used by Schlossherr *et al.*, 1994) was reduced in the 7 variegating lines. The absolute access to the Xba I site in the original strain was 68%. This was then set at 100% and the values obtained for the other lines corrected against this figure. It was shown that the only other euchromatic line tested had a relative Xba I site accessibility of 82%. The 7 heterochromatic lines tested had relative values ranging from 9% to 47%. It was concluded that this reduced accessibility correlated with the repression seen for the *white* gene. However, the ranking of strains in order of their Xba I site accessibility does not align with the ranking of strains based on the apparent amount of eye pigmentation. Note that this ranking of eye pigmentation was done by visual inspection and really should be quantitated from eye extracts. This apparent lack of correlation was not mentioned or discussed by the authors. It may be that different sites within a nucleus have different accessibilities to an exogenous protein (restriction endonuclease), but that this accessibility has no correlation to the relative transcriptional states of those sites. In this sense, the experiment by Schlossherr *et al.*, (1994) is much more informative as it compares the same locus in an activated and repressed state. Therefore, at this stage it must be concluded that there is no significant difference in the accessibility of PEV repressed or unrepressed chromatin to a foreign protein. Whether this relates to the accessibility of transcriptional machinery to the chromatin is a question that is yet to be answered.

These data and those which show that a heterochromatic gene is not expressed correctly when moved to a euchromatic region, have led some to propose a different model for PEV gene silencing. In this model it is proposed that there are two compartments within a nucleus: a heterochromatic and euchromatic compartment. There are specific transcription factors for each compartment that cannot move between compartments. If a gene from one compartment is transferred to another compartment (i.e. by a chromosomal inversion), then it cannot access the appropriate transcription factors for its expression (Devlin *et al.*, 1990; Wakimoto and Hearn, 1990). This model is appealing but it has some significant deficiencies. Firstly, it has been shown that two variegating loci can have different gene expression, compaction and HP1 binding states in the same cell (Belyaeva *et al.*, 1993). This indicates that a simplistic model of a heterochromatin compartment which contains a homogeneous pool of

heterochromatin specific proteins cannot be true. Another deficiency in the model is that it does not explain how exogenous proteins (e.g. restriction enzymes) can access all nuclear compartments. This will be discussed further at the end of the next section, when the comparison is made to Pc-G silencing.

1.5.4 - Pc-G silencing is not mediated by compacted, inaccessible chromatin

Schlossherr *et al.*, (1994), extended their study on chromatin accessibility at PEV silenced sites to accessibility at Pc-G silenced sites. They analysed the *Abd-B* and *Antp* genes. *Abd-B* is normally only expressed in parasegments 10 - 14 of the embryo, however in Pc^3 homozygous embryos, it is derepressed throughout all of the embryo. Three different restriction enzyme sites around the *Abd-B* gene were assayed in single wild - type and Pc^3 homozygous embryos. It was found that there is no difference in accessibility to these sites in these two genetic backgrounds. Similarly, two restriction sites were assayed in the *Antp* gene. In this case chromatin from dissected eye-antennal imaginal discs (where *Antp* is not expressed) and wing discs (where *Antp* is expressed) was compared. Here, it was shown that there was a two - fold increase in accessibility in the wing discs. However, this difference in accessibility was concluded as not being sufficient to account for the expression difference of the *Antp* gene (Schlossherr *et al.*, 1994). Therefore, these experiments show once again that chromatin accessibility to an exogenous protein is not a significant factor in the transcriptional state of genes. The weakness with these experiments is that they bear no relationship to whether a locus is accessible to transcriptional machinery. This question has been addressed by two recent elegant studies.

In the first study, a P-element containing UAS (the GAL4 DNA binding site) driving lacZ was used to replace a characterised P-element within the *bx* region of the regulatory region of *Ubx* (bxUASlacZ). This characterised P-element was known to be under the control of the Pc-G, maintaining its repression anterior to PS 5 and posterior to PS 12. A control reporter P-element (UASlacZ) that was not inserted in the *Ubx* gene was also used. By crossing both of these reporter lines to flies expressing GAL4 in different tissues of the embryo, it was shown that lacZ transcription was repressed anterior to PS 5 and posterior to PS 12 in a Pc-G dependent manner in the embryos carrying bxUASlacZ but not in embryos carrying UASlacZ (McCall and Bender, 1996). Interestingly, when these lines were crossed to flies expressing GAL4 in the pattern of the *paired* gene, although anterior and posterior repression was not initially observed, it became apparent, only in bxUASlacZ embryos, at the time of embryogenesis when the Pc-G begins its repression maintenance of homeotic genes (at the beginning of germ band retraction). *paired* is a segmentation gene that is expressed from very early in embryogenesis. This result suggests that the Pc-G can 'shut down' the expression of a gene if it is inserted into a region of the genome that is under normal Pc-G repression. In contrast to these results with GAL4 transcriptional activation are the results observed with phage T7 polymerase mediated transcription. In a similar way, a P-element containing a T7 promoter driving a lacZ reporter gene was used to replace the same P-element in the *bx* region of *Ubx*. T7 RNA

polymerase directs autonomous transcription when it binds to its 23 bp promoter. It was shown that transcription of the reporter gene, mediated by ubiquitously expressed T7 RNA polymerase, in this case did not respect the parasegmental boundaries (McCall and Bender, 1996). The difference between transcription from bxUASlacZ and the T7lacZ is that in the former case GAL4 uses the endogenous RNA polymerase II transcriptional machinery, whereas, as mentioned above, in the latter case transcription is mediated independently by the T7 RNA polymerase. It was concluded then, that although Pc-G mediated repression did not physically exclude proteins from DNA, or even prevent transcription (usually considered disruptive of nucleosomes), transcription mediated by the *Drosophila* RNA polymerase II was blocked, indicating a more specific mechanism of repression than chromatin compaction (McCall and Bender, 1996). Experiments carried out by Zink and Paro (1995) confirm and elaborate the effect of Pc-G silencing on GAL4 - RNA polymerase II transcription. As mentioned above (section 1.4.5) the MCP and FAB - 7 fragments containing PREs and not patterning information, were fused to a lacZ reporter gene under the control of a UAS. P-elements containing various permutations of these fragments were inserted randomly into the genome (i.e. not into the *Ubx* gene as for the experiments by McCall and Bender, 1996). Flies containing these reporters were crossed to different GAL4 expressing lines, the salivary glands dissected from third larval instar progeny and assayed for β - galactosidase activity. The results showed that reporter genes carrying PREs had significantly lower activity than control reporter constructs without any PRE (Zink and Paro, 1995). Further, it was shown that overexpression of the GAL4 was able to alleviate Pc-G silencing. This interplay between Pc-G silencing and GAL4 activation correlated well with the complementary presence or absence of PC or GAL4 at the site of transgene insertion on polytene chromosomes (Zink and Paro, 1995). Therefore, as well as showing that Pc-G mediated silencing can block RNA polymerase II transcription, these experimenters showed that this silencing could be alleviated by high levels of the GAL4 protein, indicating some kind of competition between repression and activation mechanisms.

In summary, it appears that Pc-G mediated silencing is not caused by a simple compaction of chromatin resulting in its inaccessibility to proteins that activate transcription. Although it must be noted that the possibility exists that the chromatin is compacted but only to a degree and single, autonomous, high affinity proteins such as restriction endonuclease and T7 polymerase are not inhibited by this compaction, but endogenous transcription complexes are. The differences between the chromatin of Pc-G silenced loci as opposed to unsilenced loci are still largely unknown. In fact it is unclear whether there is any feature of the chromatin of Pc-G silenced loci that is in itself responsible for the silencing observed.

As has been proposed for PEV gene silencing, Pc-G silenced loci may occupy a certain compartment of the nucleus that transcriptional activators cannot access. Given the molecular and genetic nexus between Pc-G silencing and PEV silencing, it may be that this compartment is the heterochromatin compartment as described in the previous section. One problem with the compartment model was described above. A further problem is that inhibition of access of factors to a nuclear compartment or to compacted chromatin is essentially a similar proposal. The data presented above which show that exogenous proteins have no difficulty in accessing Pc-G silenced loci as well

as suggesting that the chromatin fibre is not inaccessible, suggest that the chromatin is not stored in an inaccessible compartment within the nucleus. Experiments which lead to the proposal of an alternative model are discussed below.

1.5.5 - PEV silencing of bw^D is correlated with a heterochromatic association

brown (*bw*) is a gene responsible for eye pigmentation. It is located at a very distal portion of chromosome 2 far away from heterochromatin (Lindsley and Zimm, 1992). A variegating allele of *bw*, bw^D, was shown not to be associated with a chromosomal inversion but with an ~2 Mb insertion of repetitive heterochromatic sequence into the *bw* gene. Interestingly, this PEV is dominant: a wild type *bw* allele is silenced by the bw^D allele in heterozygotes. This dominant silencing is dependent on the physical pairing of the homologues (Dreesen *et al.*, 1991). Based on the observation that the variegation was enhanced when bw^D was positioned closer to the centromere, it has been proposed that repression of the *bw* gene in bw^D occurs by an association with centromeric heterochromatin. Using fluorescent *in situ* hybridisation (FISH) and 3-D microscopy, it has been shown that in a proportion of interphase nuclei from eye imaginal discs and brains, bw^D and its wild - type homologue associate non - randomly with a block of heterochromatin near the centromere of chromosome 2 (Csink and Henikoff, 1996; Dernburg *et al.*, 1996). This proportion closely reflects the rate at which the *bw* gene is repressed in bw^D flies. In a similar way heterochromatin that has been translocated distally in an inverted X chromosome, FM7, is also stochastically, non-randomly associated with centromeric regions of the X chromosome (Dernburg *et al.*, 1996). Another important point to come from this work is that although some heterochromatin appears to associate, not all heterochromatin is grouped within interphase nuclei (Dernburg *et al.*, 1996). These striking data strongly suggest that there are a number regions within the nucleus that are enriched for heterochromatin. It seems likely that either DNA homology directly, or protein interactions cause the interactions between heterochromatin, as it is non-random (Dernburg *et al.*, 1996).

Previous to these experiments a serendipitous discovery was made which, in the light of these recent experiments, now can be explained. A line was constructed that had a P-element with a *w* gene inserted near the heterochromatin/euchromatin breakpoint of an inverted chromosome. The P-element was mobilised in an attempt to generate small deletions and bring the P-element closer to this boundary, so that the *w* gene would variegate. Indeed flies exhibiting *white* variegation were recovered, but these were not associated with the predicted small deletions rather, they were associated with tandem repeated arrays of the P-element. In fact, subsequent selection for stronger variegation of the eye pigmentation resulted in the recovery of longer P-element arrays or arrays in which neighbouring P-elements were in the opposite orientation to each other (5' to 5'). This phenomenon could occur in medial euchromatin but variegation was stronger when the transgene array was closer to heterochromatin. Furthermore, that this variegation is similar mechanistically to classic PEV was demonstrated by the observation that extragenic modifiers of PEV have similar effects (Dorer and Henikoff, 1994). What appears to have happened in these experiments is that a

small block of 'heterochromatin' type DNA has been created which then behaves in the same way as the ~ 2 Mb block of heterochromatin inserted in the *bw* gene in bw^D. It appears that construction of DNA repeats, in this case transgene arrays, can mimic heterochromatin DNA. It is important to note that this effect increases with the size of the heterochromatin block. In support of this, it has been shown that a ~20 kb block of heterochromatin cannot induce variegation of a neighbouring euchromatin gene (Pokholkova *et al.*, 1993).

In the case of bw^D, it appears that the spurious association of a euchromatic gene (*bw*) with centromeric heterochromatin may cause the repression of transcription of the *bw* gene. An analogous mechanism appears to be operating for the transgene repeats. It is possible however that heterochromatin proteins may associate with the bw^D heterochromatic insertion, repressing transcription and then translocating it to the centromere. This seems unlikely, as it does not explain why *bw* in bw^D is not always repressed and not just repressed at the same rate at which the locus is associated with the centromere. The idea that is emerging here is that there are regions of the nucleus that are enriched for heterochromatin and heterochromatin proteins. Some people might refer to these regions as compartments. If a euchromatic gene is mispositioned to one of these regions, it may not be able to access the pool of euchromatin transcription factors that would normally activate its transcription, thereby leading to the repression seen in PEV. This repression would also be associated with the presence of heterochromatic proteins on the euchromatic fragment (as has been observed for HP1 (Belyaeva *et al.*, 1993) and the characteristic appearance of heterochromatin (compacted).

What evidence is there for heterogeneity of regulatory proteins and chromatin within the nucleus?

1.5.6 - Silenced compartments within a yeast nucleus

In yeast, silencing at the silent mating type loci, HML and HMR, is mediated by two *cis* silencer elements (**E** and **I**) and the proteins Rap 1 (Shore and Nasmyth, 1987), Abf1 (Buchman *et al.*, 1988) and the ORC complex (Bell *et al.*, 1993), which initiate repression, and the Sir proteins which maintain repression. RNA polymerase II transcribed genes placed near yeast telomeres were found to be silenced in a stochastic way reminiscent of PEV in *Drosophila* (Gottschling *et al.*, 1990). The telomeres of yeast contain poly (TG_{1-3}) tracts which bind Rap 1 protein (Gottschling *et al.*, 1990) and effect silencing by recruiting Sir 3 and Sir 4 proteins (Moretti *et al.*, 1994). The silencing by Sir proteins is mediated by an interaction with histones H3 and H4 (Hecht *et al.*, 1995) and is therefore postulated to involve changes in chromatin structure. The HML locus is ~ 13 kb and the HMR locus is ~ 25 kb from the telomere of chromosome III. Recently the requirement for proximity to the telomere for silencing at the HMR and HML loci was tested. Reporter gene constructs were made that contained E and I elements flanking a LEU2-lacZ fusion gene. These constructs were integrated at various positions along yeast chromosomes and the effect on silencing was assayed. It was found that reporter genes closest to the telomeres silenced the best, whereas, moving away from the telomeres abolished silencing. Furthermore it was shown that truncating chromosomes to produce novel telomeres

rescued silencing at more centromere proximal reporter genes. Lastly in was shown that overexpression of one or more of the Sir proteins also rescued silencing at centromere proximal reporters. Immunodetection of Sir proteins in yeast nuclei was particularly informative. In wild type yeast Sir 3 and 4 are found localised to a few discrete foci coincident with the telomeres. However, when these proteins are overexpressed, their distribution in the nucleus is uniform possibly enabling them to silence non-telomere associated loci (Maillet *et al.*, 1996). These data fit squarely with the idea that silencing proteins are not normally heterogeneous within a nucleus, and that repressed loci may be clustered to regions of high concentrations of these proteins.

1.5.7 - Clustering of Pc-G silenced loci?

As was described above, Pc-G proteins are found at many sites (100 for PC/PCL/PH) on squashed polytene chromosomes. Polytene nuclei are interphase nuclei, so it would be of interest to determine if there was a clustering of these Pc-G binding sites in these nuclei when analysed in the un-squashed 3-D state. It has been shown that polytene chromosomes are not arranged randomly within the nucleus, rather, there is a specific reproducible organisation (Mathog *et al.*, 1984). Therefore, there is at least a possibility that Pc-G sites might cluster in these nuclei. Further, it would be of interest to determine if there was any clustering of Pc-G sites in non-polytene (diploid) nuclei. It has been shown that PC is present at several foci of staining in SL - 2 cells. Mutant PC proteins, that cause homeotic phenotypes *in vivo*, showed, in contrast, uniform nuclear staining (Messmer *et al.*, 1992). It is therefore possible that PC, and hence Pc-G, bound sites are present in clusters to effect silencing in these nuclei. In section 1.4.3 it was discussed that overexpression of *Psc* and *Su(z)2* resulted in the ectopic repression of gene transcription. In the light of the data which showed that overexpression of Sir 3 and 4 resulted in ectopic silencing of sites that were not close to the normal clustered Sir 3 and 4 repressed sites, probably as a result of ectopic nuclear distribution of Sir 3 and Sir 4, it seems reasonable to suppose that overexpression of *Psc* and *Su(z)2* results in the ubiquitous nuclear distribution of these proteins and ectopic silencing of non-clustered sites. The presence of these proteins at hundreds of polytene chromosome sites supports this (see section 1.4.3). In a similar but converse way, overexpression of GAL4 to a high enough concentration may have enabled the recruitment of the RNA polymerase II complex to a Pc-G clustered site to activate transcription from the transgene that contained a PRE and UAS enhancer binding site (see section 1.5.4) (Zink and Paro, 1995).

Is there any molecular evidence for interaction between Pc-G bound sites that may result in their nuclear clustering? Many researchers have observed a phenomenon termed 'homing' (Chan *et al.*, 1994; Fauvarque and Dura, 1993; Kassis *et al.*, 1992). When a P-element transgene containing a PRE fragment is transformed into *Drosophila*, the transgene inserts into the genome at a Pc-G bound site (as assayed by polytene *in situ* and immunostaining) at a high rate. Although there is no direct molecular evidence, it seems likely that protein-protein interactions between Pc-G proteins bound at a genomic site and those bound to the incoming PRE containing transgene result in the site specific

insertion of the transgene. That PREs can interact in this manner suggests that genomic PREs may be able to interact by virtue of protein-protein interactions.

Based on the evidence discussed above, a model which describes the association and incidental segregation of Pc-G silenced sites is presented below (figure 1.5.7.1).

Initially, ESC recognises, by protein-protein interaction, some aspect of a non-transcribed gene, for example a nucleosome in a particular configuration. Then by further protein interactions, ESC interacts with a pre-formed Pc-G complex in solution which then either binds directly or indirectly (via an already bound 'anchor' protein) to a PRE. Then, by further protein-protein interactions, multiple nearby PREs are brought into close proximity. In this dynamic way a cluster of Pc-G bound loci are brought together, possibly interacting with heterochromatin clusters, given the molecular and genetic similarities between Pc-G and PEV silencing. In a converse way, active loci are separated from the repressed loci. A critical aspect of this model is the recognition that transcriptional proteins (activators, repressors, heterochromatin proteins) are not present homogeneously in the nucleus, because of their limiting amounts and association with the loci with which they are having an effect. In this way, active (by putative interactions between activator proteins) and repressed loci are segregated into multiple clusters within the nucleus and then are effectively separated. Note, that no actual physical barrier or compartment is postulated to form. This would be consistent with the observation that exogenous restriction enzyme or T7 RNA polymerase would have access to all DNA in the nucleus (see section 1.5.4), but that RNA polymerase II (endogenous) transcription would be limited to clustered active loci where RNA polymerase II is present at highest concentration. Explaining the relationship to artifactual GAL4 mediated RNA polymerase II transcription is more complicated. Zink and Paro (1995) observed that GAL4 and PC could essentially each bind partially (remember that polytene chromosomes contain thousands of DNA strands bundled together, and partial binding indicates binding to some of these DNA strands) at the site of insertion of a transgene that contains a UAS and a PRE. At a site that contains a PRE and a UAS, both PC and GAL4 would be able to bind to the transgene insertion site, maybe even at the same time. Then it would be expected that this locus would stochastically be either associated with other repressed or active loci, respectively. In turn this would explain the lower level of both PC and GAL 4 protein at this site, as there would be a much lower level of PC and GAL4 at active and repressed sites, respectively, and the locus would only be associated with each cluster for part of the time. This stochastic, partial association with clusters of active loci would also explain the low expression levels of the reporter gene that were observed (Zink and Paro, 1995). In these experiments, when GAL4 was expressed at very high levels, PC binding at the site of transgene insertion was completely abolished, GAL4 binding was complete, and there was a high level of transcription from the reporter gene. This observation can be explained by this model also. If GAL4 is present at very high levels, it would probably be ubiquitously present in the nucleus (as observed for Sir 3 and Sir 4 - see section 1.5.6) and not just at the clusters of active loci due to its interaction with the transcriptional initiation complex. This ubiquitous distribution would allow GAL4 to be bound to the transgene even when it was associated with a repressed locus cluster. This binding would then

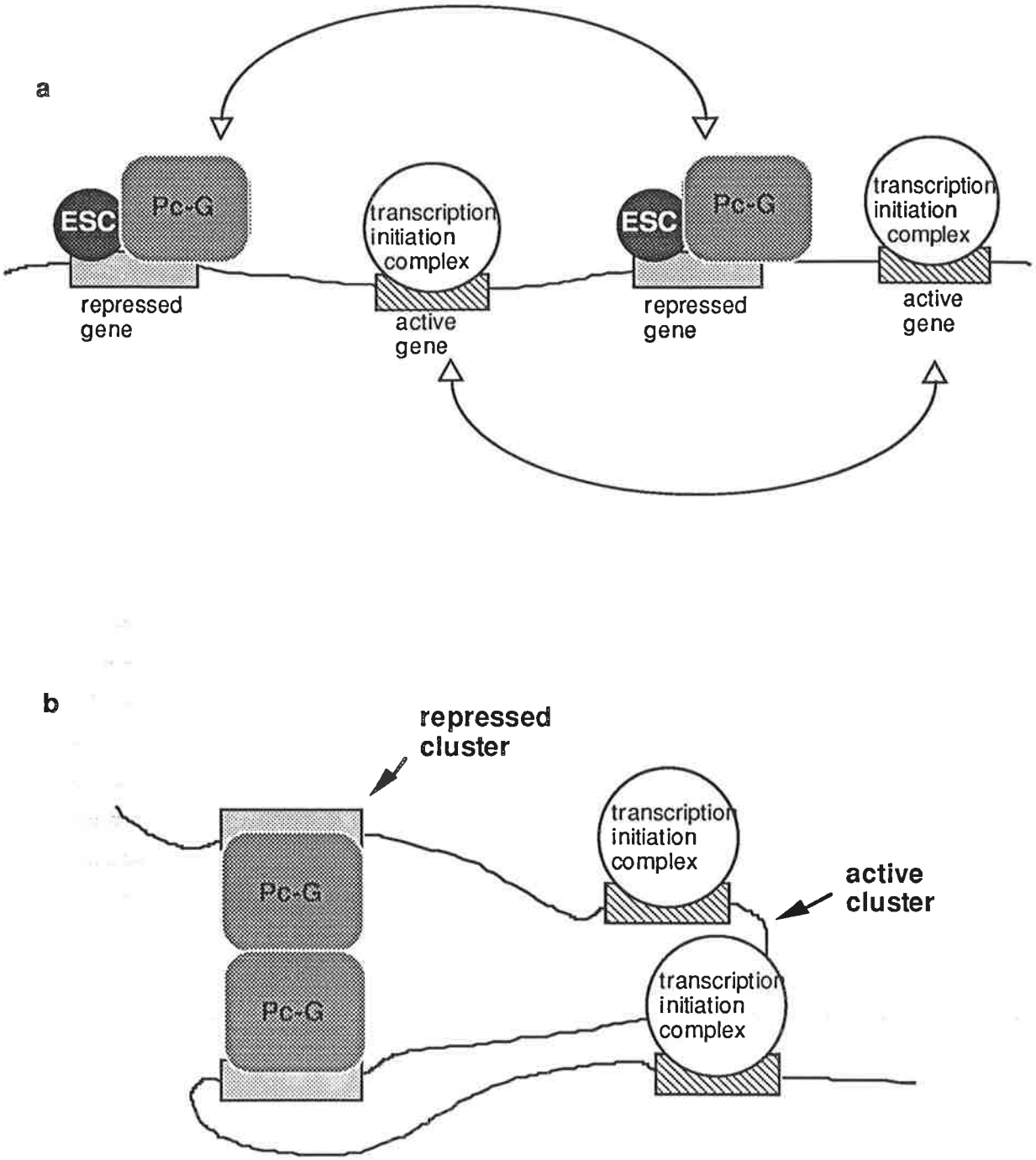


figure 1.5.7.1 - Model for generation and segregation of silent gene clusters

In the initiation phase (**a**), ESC, which is bound to repressed loci, recruits the Pc-G. The active genes have the transcription initiation complex bound. There is attraction between Pc-G complexes and possibly also between proteins that are within the transcription initiation complex (arrows). This leads to the association of the repressed genes and possibly, the separate association of active genes (**b**). Due to the limiting nature of repressive and activating proteins within a nucleus, repressed genes are effectively separated from transcriptional activators.

displace the transgene from this cluster, bringing it into an active cluster (by virtue of the GAL4 protein interaction with the transcriptional initiation complex) where there is no PC protein, and stimulate transcription from it. It is important to note that the clustering mechanism may be quite different in polytene nuclei. Given that any locus already represents a cluster of many individual gene copies (aligned in register), it may be that inter - gene clustering would not need to occur. If this were the case then the partial binding of GAL4 and PC seen at the insertion site of the transgene may represent the presence of active and repressed clusters at this site, and that stochastic association with other active or repressed loci would not occur.

This model can also explain what is seen in PEV. To illustrate this in simple form, the *w*, *lt* and *bw* genes, although not syntenic have been depicted on the same chromosome arm (figure 1.5.7.2a). An inversion is shown which positions *w* near centric heterochromatin and *lt* within euchromatin. bw^D stays in medial euchromatin (figure 1.5.7.2b). Note the proteins encoded by *Su(var)* genes bound (presumably to repeat DNA) to heterochromatin. In the inversion case (1.5.7.2b), *w* is present within heterochromatin and cannot access euchromatin transcriptional activators in a fraction of cells. *lt* and bw^D interact stochastically with centric heterochromatin, based on interaction between *Su(var)* encoded proteins (see section 1.5.5). When the interaction occurs, *lt* is expressed because it has access to heterochromatin specific transcription factors, and bw^D is repressed, because it no longer has access to euchromatic transcription factors. Mutation of a *Su(var)* leading to a reduced functional dosage of that protein would mean that the centric heterochromatin would interact less often with other heterochromatin (figure 1.5.7.2c), enabling *w* incidentally to have more access to euchromatic regions of the nucleus and euchromatic transcription factors, resulting in a stochastically increased transcription of *w*. The distal heterochromatin associated with *lt* and bw^D would interact less frequently with the centric heterochromatin, resulting in less access of *lt* to heterochromatic factors (i.e. less trancriptional activation) and more access to *bw* of euchromatic specific factors (more transcriptional activation) due to *bw* being associated with euchromatin more often. This is consistent with the observation that mutations in *Su(var)* genes have opposite effects on variegating euchromatic and heterochromatic genes (see section 1.5.2).

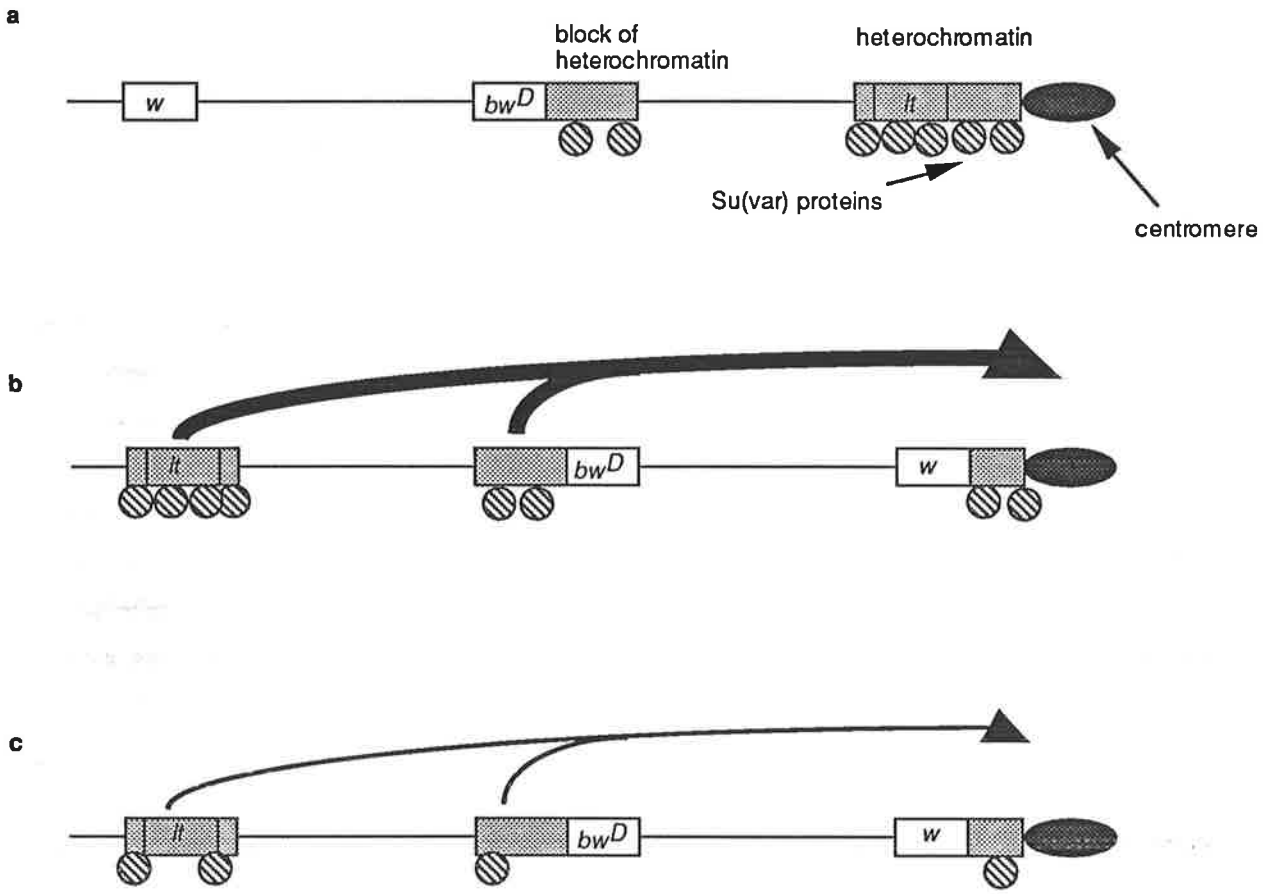


figure 1.5.7.2 - Associations between heterochromatic regions

A fictitious chromosome containing *w* near the telomere, bw^D in a medial region and *lt* within centric heterochromatin, is shown (**a**). **b** shows the result of a chromosomal inversion with breakpoints outside the *w* and *lt* genes. The result of this is that *w* is translocated to centric heterochromatin and *lt* is translocated to distal euchromatin. The position of bw^D does not change. Due to interactions between *Su(var)* encoded heterochromatin proteins, *lt* and bw^D stochastically interact with centric heterochromatin (thick arrows in **b**). When the dose of heterochromatin proteins is reduced, these stochastic interactions occur less frequently (thin arrows in **c**).

1.5.8 - What is the nature of Pc-G complex association with repressed genes?

Given that the Pc-G does not appear to silence by simply engulfing a gene in compacted heterochromatin, it would be of interest to know how the Pc-G complex associates with silenced genes. Is it only associated with the PRE, for example? Orlando and Paro (1993) used anti - PC antibodies to immunoprecipitate formaldehyde cross-linked chromatin from *Drosophila* SL - 2 cells to answer this question. By extracting the DNA from such chromatin for use as a probe onto Southern blots containing DNA fragments of *Antp*, they showed that PC is associated with approximately 1.5 kb of DNA overlapping one transcriptional start site of *Antp*. By using the probe onto a Southern blot containing a walk covering the BX-C they found that PC was associated with most of the DNA with peaks of PC present at known PREs (e.g. MCP, FAB - 7 and BXD). Strikingly PC was not associated with the *Abd-B* gene, and subsequent analysis showed that this was most likely due to the fact that *Abd-B* is not repressed in SL - 2 cells (Orlando and Paro, 1993). Therefore, although Pc-G silencing is not due to a general compaction of chromatin (see above) PC does appear to be associated with extensive regions (including coding regions) of repressed but not unrepressed genes. In light of this it is possible that the Pc-G complex is somehow partially involved in either a steric blocking or specific blocking of the transcription initiation complex. If there is some aspect of steric blocking, an explanation must be given as to how restriction enzymes and T7 RNA polymerase can access Pc-G silenced chromatin. It is possible that steric blocking only applies to large protein complexes, such as the transcription initiation complex and not small autonomously acting proteins like restriction endonuclease or T7 RNA polymerase. Crosslinking a large protein (e.g. myosin or β - galactosidase) to a restriction enzyme or expressing a T7 RNA polymerase fusion with a large protein and assaying for their access to Pc-G repressed chromatin would test whether or not steric blocking occurs. Although some kind of steric blocking may occur, it seems clear from the evidence presented above that segregation of active and repressed clusters occurs and is part of the repression maintenance mechanism. In fact a two - tiered repression mechanism model - segregation and steric blocking - may be required to explain the stability of Pc-G repression that is observed.

1.5.9 - Pc-G silencing 'spreads' in an analogous way to PEV silencing

The observation that PC is associated with large tracts of DNA but that PREs can be refined to relatively smaller regions implies that Pc-G repression can 'spread'. It has been observed by many workers that flies containing transgenes with PREs and a *w* reporter gene (used to detect the presence of the P-element) show variegated distribution of the eye pigmentation in a manner reminiscent of classic PEV (Chan *et al.*, 1994; Fauvarque and Dura, 1993; Gindhart and Kaufman, 1995; Zink and Paro, 1995) . This variegated pigmentation has therefore been ascribed to a stochastic spreading of Pc-G repression from the PRE to the neighbouring *w* reporter gene. This phenomenon

was termed DREV (developmental regulator effect variegation) and characterised in comparison to PEV. It was shown that unlike PEV, DREV was not modified by the absence of the Y chromosome. Also, almost all of 22 *Su(var)* and 9 *E(var)* mutations tested for their ability to modify DREV had no effect. However, interestingly, one suppressor of PEV weakly enhanced DREV and two enhancers of PEV weakly suppressed DREV. Most strikingly, temperature had the opposite effect on DREV as it does on PEV. Normally, a temperature increase suppresses PEV and a decrease enhances PEV. In all lines tested temperature increase strongly enhanced DREV (Fauvarque and Dura, 1993).

In the light of the model proposed above for the clustering of Pc-G repressed loci, the spreading of Pc-G repression into neighbouring genes may be caused by the stochastic 'dragging' of the neighbouring gene into a Pc-G repressed gene cluster. This spurious shutting down of neighbouring genes poses a potential problem. This problem is particularly obvious within a homeotic gene complex. Somehow, Pc-G repression must be prevented from spreading from one repressed homeotic gene into a very close active gene. An example of this is described above in the previous section. Orlando and Paro (1993) showed that in SL - 2 cells PC is associated with large tracts of DNA in the BX - C, but abruptly, this association is no longer present in the *Abd-B* gene, which is active in these cells. How is PC and presumably the rest of the Pc-G blocked from spreading into and repressing the *Abd-B* gene in this case?

1.5.10 - Pc-G repression is blocked by boundary elements

Within the BX-C there are 3 transcription units: *Ubx*, *abd-A* and *Abd-B*. Transcription of these, as alluded to earlier, is specific to a block of parasegments. It turns out that transcription within each parasegment is defined by a separable *cis*-regulatory element. Within the BX-C these elements are, in order from the most anterior to posterior parasegment they specify (which also happens to be their order on the chromosome from proximal to distal), *abx/bx*, *bxd/pbx*, *iab-2*, *iab-3*, *iab-4*, *iab-5*, *iab-6*, *iab-7*, *iab-8* and *iab-9* (reviewed in Peifer *et al.*, 1987; figure 1.4.5.1) .

These elements correspond to parasegments 5 - 14. Individuals homozygous for mutations in any of these elements exhibit transformation of a parasegment to the identity of the adjacent, anterior parasegment. In addition to these *cis* - regulatory enhancers, two other regions with interesting function have been identified within the BX-C; MCP (in between *iab-4* and *iab-5*) and FAB - 7 (in between *iab-6* and *iab-7*). When FAB - 7 is deleted, for example, the parasegment identity that is specified by *iab-7* is imposed on the adjacent, anterior parasegment (i.e. parasegment 11 becomes parasegment 12). It appears therefore that FAB - 7 acts as a boundary element to separate the effects of the *iab-6* and *iab-7 cis*-regulatory regions. This was shown by the insertion of an enhancer trap P-element (bluetail element) into FAB - 7. Lac Z expression from the bluetail element was initiated in PS 12 by *iab-7* but not in PS 11 by *iab-6*. Although the story is somewhat more complicated (*iab-6* is normally able to activate *Abd-B* in PS 11 and *Abd-B* is located distally to *iab - 7*, i.e. the bluetail element is between *iab-6* and *Abd-B*) it clearly indicates that FAB - 7 acts as a boundary element or insulator to

block the effects of *iab-6* (and more proximal *cis* - regulatory regions) on the bluetail element. What is of further interest is that the anterior boundary of Lac Z expression from the bluetail element (PS 12) is maintained by the Pc-G. It was shown that in *Pc* mutant animals repression of Lac Z anterior to PS 12 is lost - there is lac Z transcription in the anterior of the embryo. These researchers concluded that *iab - 7*, which is normally repressed in anterior parasegments was in this case able to activate transcription of Lac Z in more anterior parasegments. In wild - type embryos, the distribution of ABD-B is different in PS 11 and PS 12. PS 11 staining is weaker than that in PS 12. It appears, therefore, that in wild - type embryos *iab - 7* is repressed in PS11. However in a FAB - 7 deletion, ABD-B becomes PS 12 like in PS 11, indicating that *iab - 7* has become derepressed in PS 11 (Galloni *et al.*, 1993). These data indicate that Pc-G repression acts on *iab - 7* anterior to PS 12. In a *Pc* mutant or in a FAB - 7 deletion, this repression is no longer present, causing *iab - 7* to be derepressed anteriorly (only up to PS 11 in the FAB - 7 deletion as there would still be a presumed boundary element between *iab - 6* and *iab - 5*). Therefore, there is strong suggestive evidence that FAB - 7 acts as a boundary element to prevent parasegment non - specific activation (in this case *iab - 6* activation of bluetail in PS 12), and also to prevent Pc-G repression from spreading into an adjacent *cis* - regulatory region. MCP, and other as yet undiscovered boundary elements presumably act in the same way. It was shown that the boundary between PC - associated DNA and the PC - free *Abd-B* region in SL - 2 cells lies within the FAB - 7 region (Orlando and Paro, 1993). It is no surprise then that FAB - 7 and MCP contain PREs (see section 1.5.4; Zink and Paro, 1995). It may be that there is interaction between Pc-G proteins and boundary element associated proteins to prevent the inappropriate spreading of Pc-G repression.

It would also be expected that boundary elements would prevent Pc-G repression from spreading into neighbouring genes that are normally never under Pc-G control. In the case where transgenes containing PREs and the *w* gene show Pc-G repression of *w* (see section 1.5.9), there is most likely no intervening boundary element present. It is difficult to imagine how a boundary element would prevent the spurious 'dragging' of a gene (or *cis* - regulatory region) into a Pc-G repressed locus cluster. There is some speculation that boundary elements may provide attachment sites to the nuclear matrix (reviewed in Eissenberg and Elgin, 1991), thus providing an intervening anchorage site between active and repressed chromatin. However, there is no good molecular data pertaining to the structure of boundary elements.

1.6 - Functional characterisation of the Pc-G by screening for Pc-G interactors

The information described above, although describing the characteristics of the Pc-G and its function, highlight the vast lack of understanding that persists. In particular, several key questions remain to be answered.

- By what mechanism does the Pc-G complex attach to the PRE? Is there a member of the Pc-G that interacts with DNA directly or with an as yet unidentified DNA binding protein?

- What specific protein associations occur between Pc-G members? Do these associations give some insight into how Pc-G repressed may sites cluster?

- What non - Pc-G proteins do Pc-G members interact with physically?

This last question is very broad and accordingly, answers to it may provide the most information for the mechanism of Pc-G action. If Pc-G proteins interact with heterochromatin (*Su(var)* encoded) proteins, the hypothesis of similar mechanisms of Pc-G and PEV silencing would be re-inforced. It is apparent that clustering of Pc-G repressed loci is probably not sufficient to account for the specificity and stability of Pc-G silencing. Pc-G proteins may interact with some components of the transcription initiation complex specifically interfering or competing to effect Pc-G silencing. If Pc-G proteins interact with histones (in an analogous manner to the Sir 3,4/Histone H3,H4 interaction), then it could be postulated that Pc-G silencing involves nucleosome modification (however this would not be predicted to result in complete inaccessibility to the DNA, as described above). If Pc-G proteins interact with boundary element factors, there would be a molecular handle to characterise the postulated interaction between Pc-G silencing and boundary elements.

1.6.1 - POLYCOMBLIKE and the aims of this study

Polycomblike (*Pcl*) was identified in a mutagenesis screen for enhancers of the partial homeotic transformations exhibited by Pc^3 heterozygote adults (Duncan, 1984). *Pcl* is an essential gene, as loss of *Pcl* function results in embryonic lethality. It was further shown by Duncan (1984) that *Pcl* mutant phenotypes closely resemble gain-of-function homeotic mutations, indicating that *Pcl* has a role in repressing homeotic gene function. Using clonal analysis experiments, it was shown that *Pcl* exerts this repressive function, at least for the BX-C, until at least 24-48hrs pre-pupation (Duncan, 1984).

Consistent with the predicted role for PCL in the maintenance of homeotic gene repression, it has been shown that *Ubx* is ectopically expressed in late homozygous mutant*Pcl* embryos (Lonie, 1994). Interestingly, the requirement for PCL may be different in different tissues. Soto *et al.*, (1995) showed that in homozygous mutant*Pcl* embryos, ABD-B is ectopically expressed throughout all parasegments of the central nervous system (CNS) but only as far anterior as parasegment 6/7 in the epidermis (ABD-B is normally present between parasegments 10-15 in the epidermis). This contrasts with the observation that *Asx* mutant embryos have a converse misexpression pattern in the CNS with respect to the epidermis, and other Pc-G mutant embryos show no tissue specificity of homeotic gene misexpression. The mechanistic basis for the apparent tissue specificity of PCL action has not been determined.

The *Polycomblike* gene has been cloned and partially characterised (Lonie, 1994; Lonie *et al.*, 1994). *Pcl* encodes an 857 amino acid polypeptide that, at the time of commencement of this study, showed no extensive sequence similarity to any other known protein. Although the predicted

size of PCL is approximately 95kD, it was observed that PCL is present on western blots of *Drosophila* extracts as a doublet band migrating at approximately 140kD (Lonie, 1994). It was proposed that PCL is post-translationally modified *in vivo*, but the nature and function of this modification is unknown (Lonie, 1994). PCL was shown to encode a nuclear protein that is present in nuclei throughout development. This is also true of the other Pc-G proteins examined to date. The ubiquitous distribution of PCL is consistent with its role as a general repressor of gene transcription throughout the anterior-posterior axis of *Drosophila* (see section 1.4.5).

PCL is present at approximately 100 polytene chromosome binding sites that directly overlap with the binding sites of PC (Lonie, 1994) and therefore PH (DeCamillis *et al.*, 1992). The binding sites of PCL are overlapped partially by the binding sites of PSC, SU(Z)2 and E(Z) (see section 1.4.3). These data and the observation that PCL may be co-immunoprecipitated with PC (Lonie, 1994) suggest that PCL may be a component of the proposed Pc-G multimeric protein complex. It must be noted however, that the co-immunoprecipitation of PCL and PC as described by Lonie (1994), may have been due to background precipitation of PCL. The appropriate controls (e.g. the use of pre-immune serum, or probing the blots with an antibody to an unrelated nuclear protein) were not performed. As such this result should be interpreted with caution.

Although PCL has no sequence similarity to other Pc-G proteins, it appears to behave similarly at a molecular level. That is, PCL is found at similar chromosomal sites and appears to be part of the Pc-G protein complex. Genetic evidence, however, suggests that *Pcl* plays a critical role in Pc-G function. Pcl- mutants genetically interact most consistently with other Pc-G mutants (Campbell *et al.*, 1995). It could be that PCL is responsible for key protein interactions to establish and maintain Pc-G mediated repression. Furthermore, Pcl mutations modify PEV more consistently than any other Pc-G mutations, indicating that PCL may have a unique role in Pc-G function.

Recently, I identified that PCL contains two copies of a motif known as the PHD finger. This motif is descibed as Cys4-His-Cys3 (see figure 7.1.1) and could potentially chelate two metal ions. The PHD finger is thought to be responsible for mediating protein-protein interactions (see section 1.4.2). It is possible, given that the genetic evidence suggests that PCL is a central Pc-G component, that PCL interacts with proteins critical to Pc-G function via its PHD fingers.

AIM 1

Given that PCL appears to be crucial to Pc-G function and that it contains putative protein-protein interaction domains, the identification of PCL protein interactors could provide a large amount of information about the way PCL and the Pc-G complex function. The major aim of the work presented in this thesis was the identification and characterisation of PCL protein interactors.

Toward the latter part of this study, I discovered that PCL shares significant sequence similarity to a murine protein, M96, with reported DNA binding ability. This sequence similarity includes the PHD fingers as well as sequence outside of the PHD fingers (see figure 7.1.1). Given this, and that PCL

does appear to have a critical role in Pc-G function, I hypothesised that PCL may bind to DNA directly thereby anchoring the Pc-G complex to the genes which they repress.

AIM 2

The second aim of this thesis was to test for the ability of PCL to bind directly to DNA. These experiments are described in chapter 7.

Chapter 2 - Materials and Methods

2.1 - Materials

2.1.1 - Chemical reagents

All reagents were of analytical grade, or the highest grade obtainable.

2.1.2 - Enzymes

Restriction endonucleases	Boehringer Mannheim, New England Biolabs, Pharmacia
Alkaline calf intestinal phosphatase	Boehringer Mannheim
Proteinase K	
T4 DNA ligase	Promega
Klenow	Bresatec
T4 Polynucleotide Kinase	
DNase	Sigma
RNase	

2.1.3 - Radio-labelled compounds

α-^{32}P-dATP (3000 Ci/mmole)	Bresatec
α-^{35}S-dATP (1500 Ci/mmole)	
γ-^{32}P-ATP (4000 Ci/mmole)	
^{35}S-L-methionine	Amersham

2.1.4 - Strains

LE392: F^-,*hsd*R574, (r_K^-, m_K^+), *sup*E44, *sup*F58, *lac*Y1, *gal*K2, *gal*T22, *met*B1, *trp*R55(Murray *et al.*, 1977)

DH5α: F^-, f80, *lac*ZΔM15, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17, (r_K^-, m_K^+),
*sup*E44, *rel*A1, *deo*R,Δ(*lac*ZYA-*arg*F) U169 (Hanahan, 1983)

BL21(DE3):*hsdS*, *gal* (lcI*ts*857, *ind*1, *S*am7, *nin*5, *lac*UV5-T7 gene *1*)
(Studier and Moffatt, 1986)

BL21(DE3) p*lys*S (Studier *et al.*, 1990)

Unless otherwise indicated, *Drosophila* strains are as described (Lindsley and Zimm, 1992) and obtained from the Indiana Stock Centre, Bloomington, IA.

- *D. melanogaster*

Wild-type strain: Canton-S or w^{1118}

2.1.5 - Media and buffers

a) Media

All buffers and media were prepared with distilled and deionised water and sterilised by autoclaving, except heat labile reagents, which were filter sterilised.

All bacterial strains were propagated in LB-broth or on LB-agar plates, except LE392 which was grown in T-broth or on T-broth plates.

LB-broth:
1% (w/v) tryptone
0.5% yeast extract
1% NaCl, pH 7.0

T-broth:
0.5% NaCl
1% tryptone

SOC:
2% tryptone
0.5% yeast extract
100 mM NaCl
25 mM KCl
100 mM $MgCl_2$
100 mM $MgSO_4$
0.2% glucose

Yeast minimal medium:
0.17% yeast nitrogen base
0.5% ammonium sulphate
2% carbon source
amino acids as required

Plates: liquid broth with 1.5% bacto-agar.

Where required for selection, ampicillin was added to a final concentration of 100 µg/ml, and chloramphenicol to 30 µg/ml.

All *Drosophila* strains were grown on *Drosophila* culture media:

10% treacle
20% yeast
1% agar
10% polenta
2.5% tegosept
1.5% propionic acid

b) Buffers

Commonly used buffers were:

Protein gel running buffer :
1.5% tris base
7.2% glycine
0.5% SDS

Protein 'sample' buffer:
62.5 mM Tris-HCl pH 6.8
10% glycerol
2% SDS
5% 2-b-mercaptoethanol
0.00125% bromophenol blue

PBS: 7.5 mM Na_2HPO_4
2.5 mM NaH_2PO_4
145 mM NaCl

PSB: 10 mM Tris-HCl pH 7.4
10 mM NaCl
100 mM $MgCl_2$

TBE: 50 mM Tris-borate pH 8.3
1 mM EDTA

TBS: 20 mM Tris-HCl pH 7.5
150 mM NaCl

TE: 10 mM Tris-HCl pH 7.4

0.1 mM EDTA

TAE: 40 mM Tris-acetate pH 8.2
1 mM EDTA

10 x agarose gel load buffer: 80% glycerol
50 mM EDTA
0.1% bromophenol blue

Sequencing gel load buffer: 98% deionised formamide
10 mM EDTA pH 8.0
0.025% xylene cyanol
0.025% bromophenol blue

2.1.6 - Libraries

pNB40 0-4, 4-8 h embryonic cDNA libraries (Brown and Kafatos, 1988) λ gt11 0-18 h embryonic cDNA library (Clontech).

2.1.7 - Plasmids

a) Cloning and expression vectors

pBluescript (Stratagene)
pGEX - 3X and 1 (Smith and Johnson, 1988)
pET15b (Novagen)
pMALc2 (NEB)
pUCBM20 (Boehringer Mannheim)
pEG202 (Gyuris *et al.*, 1993)
JG4 - 5 (Gyuris *et al.*, 1993)
pAS - CYH2

b) recombinant plasmids (most plasmid constructions are described in the text)

pBS-*Pcl* Sac/Not
pGEX-*2.1* Eco
pBS-*2.1* Eco
pMAL-*2.1* Eco
pGEX-*Pcl* Bam - high fidelity PCR using the primers *Pcl* 5' Bam and *Pcl* 3' Bam and AL 15 as template was performed. The PCR product was

digested with BamHI and cloned into the BamHI site of pGEX-3X in the orientation that would produce a GST-PCL fusion.

pET15b-*Pcl* Nde/Sal - the same BamHI ended PCR product above was cloned into the BamHI site of pAS1-CYH2 in the orientation such that the NdeI site in the polylinker was at the 5' end of the *Pcl* ORF. The *Pcl* ORF was then excised with SalI at the 3' end and NdeI (partial digest) at the 5' end and then sub-cloned into NdeI/SalI digested pET15b.

pGEX-cDOM

pGEX-ΔEH

pGEX-C430S- two high fidelity (Pfu) PCRs were done using the primer pairs *Pcl*5' Bam/Nhe bottom and Nhe top/*Pcl*3' Bam. The products were restricted with NheI and then ligated to each other. The ligated product was digested with NdeI/XhoI and cloned into NdeI/XhoI digested pGEX-*Pcl* Bam (i.e replacement of the NdeI/XhoI fragment of *Pcl*).

2.1.8 - Oligonucleotides

a) Sequencing Primers

T3:	5'-d(ATTAACCCTCACTAAAGGGA)-3'
T7:	5'-d(TAATACGACTCACTATAGGG)-3'
SP6:	5'-d(GAATTTAGGTGACACTATAG)-3'

b) other primers

cDOM 5':	5'-d(AAGGATCCGTAGCTCGATTACCGCTGGCGGA)-3'
cDOM 3':	5'-d(AAGGATCCGGGCCAAATGTCATTTAGCAG)-3'
PCL5' Bam:	5'-d(GGAGGATCCTGATGAACAACCATT)-3'
PCL3' Bam:	5'-d(GGAGGATCCTTATGATGCCATTTAC)-3'
05197:	5'-d(CCATCCATGGATGATGAACAACCATT)-3'
11787:	5'-d(AAGCCCATGGTTATGATGCCATTTAC)-3'
Nhe top:	5'-d(AAAGCTAGCAAGCGATCGGATATCGAA)-3'
Nhe bottom:	5'-d(AAAGCTAGCCACGCACATGGGTCCACT)-3'
2.1bb:	5'-d(GGTCGGCTGAAAATGTTGC)-3'
2.15'seq.:	5'-d(GAATTCCGGATGAAGCCC)-3'

2.1.9 - Molecular weight markers

a) DNA

λ DNA digested with BstEII and SalI produces fragments of sizes (in kb): 14.14, 7.24, 4.82, 4.32, 3.68, 3.13, 2.74, 2.32, 1.93, 1.37, 1.26, 0.70, 0.45, 0.22 and 0.11

c) Protein

Prestained high molecular weight markers (GIBCO BRL), High and low molecular weight markers (SIGMA)

2.2 - Methods

Miscellaneous, well established molecular biological techniques were carried out according to the protocols published previously (Ausubel *et al.*, 1987)

2.2.1 - λ bacteriophage propagation

LE392 plating cells were prepared by resuspending a mid-log phase culture grown in T-broth, 10 mM $MgCl_2$, in 0.5 volume 10 mM $MgCl_2$. Appropriate dilutions of λ phage were added to 100 µl LE392 plating cells, mixed with 3 ml molten 0.7% agar at 42°C, poured onto 85 mm T-broth plates and incubated at 37°C for 8-16 h. For 140 mm plates volumes were scaled up three fold.

Agar plugs containing plaques of interest were removed with the narrow or wide end of a Pasteur pipette and phage eluted in 1 ml PSB at rt for 4 h. Serial dilutions for plating were made in PSB. Phage stocks were stored at 4°C with addition of 50 µl $CHCl_3$.

2.2.3 - Isolation of λ bacteriophage DNA

Plugs of single plaques were transferred into 0.5 ml of stationary phase LE392 culture + 16 µl of 1M $MgSO_4$ and incubated at 37°C for 30 min. This was transferred to 10 ml T-broth and incubated at 37°C with shaking until lysis (usually 4 h) whereupon 0.5 ml $CHCl_3$ was added and incubation proceeded for a further 10 min. DNase and RNase to 10 µg/ml were added for 30 min at 37°C and debris pelleted by spinning for 5 min at 1500g. 0.5 g PEG 6000 and 0.44 g NaCl were gently dissolved in the supernatant and left at 4°C for 16 h.

Phage were pelleted at 10,000 g, 4°C, for 10 min and resuspended in 500 µl PSB, 20 mM EDTA, 50 µg/ml proteinase K, 0.5% SDS at 65°C for 1 h. λ DNA was extracted with phenol/$CHCl_3$ twice and ethanol precipitated at rt, pelleted and resuspended in 100 µl water. Typical yield was 10 µg.

2.2.4 - Radio-labelling of DNA fragments

DNA fragments were labelled by Klenow catalysed α-^{32}P-dATP incorporation in random oligonucleotide primed synthesis products (Feinberg and Vogelstein, 1983) using a "Megaprime" kit (Amersham). Unincorporated nucleotides were removed by size exclusion chromatography on Sepharose CL-6B mini column (Section 2.2.13).

2.2.5 - Hybridisation of radio-labelled probes to membrane immobilised nucleic acids

Filters were pre-hybridised with hybridisation mix (50% formamide, 5 x SSC, 0.5% blotto, 100 µg/ml sonicated and denatured salmon sperm DNA) at 42°C for at least 2 h. If nylon based filters (Plaquescreen) were used hybridisation mix was supplemented with 1% SDS. Heat denatured and snap cooled radio-labelled probe was added to the membranes with fresh hybridisation mix and incubated at 42° for 4-16 h.

Membranes were washed typically (at high stringency) with two 10 min washes each of 2 x SSC, 0.1% SDS at rt, then 0.1 x SSC, 0.1% SDS at 65°C.

2.2.6 - Autoradiography

Membranes or dried gels were exposed for variable periods to X-Omat AR X-ray film (Kodak) in an autoradiography cassette (Ilford) at rt, or at -80°C in the presence of a calcium tungstate intensifying screen. Alternatively, exposure was to a pre-erased phosphorimager capture screen followed by laser scanning and image analysis (Fujix BAS1000 scanner and MacBas version 2 software)

2.2.7 - "Mini prep" isolation of plasmid DNA

A single colony was used to inoculate 2 ml of LB-broth plus ampicillin and incubated for 5-16 h at 37°C with shaking. DNA was isolated by the "boiled lysis" method (Murphy and Kavanagh, 1988) to the stage of isopropanol precipitation, where the pellet (typical yield 10 µg) was resuspended in 20µl of water.

2.2.8 - Restriction analysis of DNA

DNA was digested with restriction endonucleases under conditions recommended by the suppliers. 1/10th volume of agarose gel load buffer was added and samples were run on a 1.0% agarose horizontal minigel (Hoefer HE 33) in TAE buffer at 5-10 V/cm. DNA was visualised by staining the gel with 10 µg/ml ethidium bromide and viewing under UV light.

2.2.9 - Polymerase Chain Reaction (PCR)

PCR was generally done in a volume of 20µL in solution containing 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl2, 0.001% w/v gelatine, 0.2 mM dNTPs, 20 µM each primer, 0.025 U/µL Taq DNA polymerase (Perkin Elmer Cetus) and approximately 10,000 target DNA molecules. Reactions were subjected to various numbers of a temperature cycle consisting of 95°C for denaturation, 50-65°C for annealing and 72°C for extension. The template amount, times for each temperature step and

annealing temperature were optimised empirically for each type of PCR. If high fidelity PCR was required, Pfu DNA polymerase (Stratagene) was used instead of Taq using the manufacturers supplied buffer. An optimised annealing temperature determined using Taq was dropped by 2-3^{o}C and glycerol was also added to a final concentration of 5%.

2.2.10 - DNA fragment purification

DNA was isolated from agarose gel slices by application of the frozen gel slice to a syringe plugged with glass wool. The liquid from the gel was squeezed out, phenol/chloroform extracted and then ethanol precipitated prior to resuspension in a suitable amount of water.

2.2.11 - Creation of recombinant plasmids

Plasmid vector DNA was prepared by digestion with the appropriate restriction endonuclease in the presence of 1 U CIP to remove 5' terminal phosphates. Linear vector molecules were then phenol/chloroform extracted and purified on Sepharose CL-6B mini-columns (Section 2.2.13). Ligations of 100 ng total DNA were performed with insert:vector of 3:1 in 10-20µL 30 mM Tris - HCl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP and 1 U T4 DNA ligase at room temperature for 4 - 16 h.

2.2.12 - Transformation with recombinant molecules

a) Heat shock method

A 50 mL mid-log phase culture of DH5α (or BL21(DE3)) was harvested, resuspended in 20 mL 50 mM $CaCl_2$ and left on ice for 20 min. The cells were harvested and carefully resuspended in 2 mL 50 mM $CaCl_2$. 100 µL cell suspension was typically mixed with 1 µL ligation mix and left on ice for 30 min before heat shock at 37^{o}C for 3 min. The mixture was incubated at 37^{o}C for 30 min following addition of 0.5 mL LB-broth, plated on LB-broth plates with ampicillin (and with chloramphenicol for BL21(DE3)p*lys*S), and grown at 37^{o}C for 16 h. For "blue/white" selection of pBluescript recombinant clones 10 µL each of 20% IPTG and 10% BCIG was added to plating mixture.

b) electroporation method

500 mL of LB-broth was inoculated with 100th volume of a fresh overnight culture (usually of XL1-Blue) and grown at 37^{o}C until the $O.D._{600}$ was ~0.7-8. The flask was chilled on ice and then the bacteria pelleted. After a wash with 500 mL then 250 mL of ice cold water and a wash with 10 mL of ice cold 10% glycerol, the cell pellet was resuspended in 1.5 mL of ice cold 10% glycerol, dispensed as 45 µL aliquots, snap frozen and then stored at -80^{o}C. One aliquot of cells was thawed, mixed with 5 µl of salt free DNA and then electroporated at 2.5 kV (Bio-Rad). The cells were rescued in 1 mL of SOC, incubated at 37^{o}C for 0.5 hr and then plated.

2.2.13 - Colony cracking for analysis of putative recombinant clones

Lids were removed from microfuge tubes and 15µL of cracking solution was dispensed into each one. A colony was picked up with a yellow tip, gently touched to a fresh LB plate (a masterplate) then transferred to an microfuge tube containing the solution. The tip was 'swirled' until the solution rose up into it by capillary action. The tubes were incubated at 65°C with an additional 'swirl' for 15 minutes. The samples were loaded into a non-submerged agarose gel and were 'run in' at 30V. After this time the gel was submerged with the running buffer and subjected to electrophoresis at 90V until the bromophenol blue dye reached the bottom of the gel. Any clones that migrated slower in the gel than the control parental colony were selected for further analysis.

Cracking solution-	2M NaOH	25µL
	10% SDS	50µL
	0.5 M EDTA	10µL
	80% Glycerol	125µL
	water	785µL
	bromophenol blue	a pinch

2.2.14 - Nucleotide sequence analysis

a) Sequencing template preparation

9 µg of plasmid DNA was RNase treated, alkali denatured and purified on a Sepharose CL-6B mini column as described (Murphy and Kavanagh, 1988). 3 µg of this was annealed with 10 ng of primer at 37°C for 1 h.

b) Sequencing reactions

DNA was sequenced by the dideoxy method (Sanger, 1977) using α-^{35}S-dATP and a Sequenase Sequencing Kit (United States Biochemical) using the protocol recommended by the manufacturer.

c) Electrophoresis

Products of sequencing reactions were resolved on 0.4 mm 6% acrylamide (acrylamide:bis-acrylamide, 20:1), 8 M urea, TBE gels. Gels were dried on a vacuum gel drier at 80°C for 30 min onto 3MM paper (Whatman), and autoradiographed for 16 h.

d) Sequence Analysis

Sequence analysis and comparisons were carried out using computer programs of (Staden, 1980) and the Genetics Computer Group (University of Wisconsin, WI). Searches of GenBank and EMBL and other databases used the FASTA (Pearson and Lipman, 1988) and BLAST (Altschul *et al.*, 1990)programs.

2.2.15 - Medium efficiency yeast transformations

A yeast culture was established at an $O.D._{600}$ of 0.2 from a fresh overnight culture. This was incubated with shaking at 30°C for 3 hours (YPD medium) or 5 hours (minimal medium). The culture was transferred to a 10 mL sterile tube and centrifuged at 2000 X g for 5 minutes at room temperature in a benchtop centrifuge. The yeast pellet was resuspended in 1 ml of 900 mM LiOAc/TE and transferred to a microfuge tube. The yeast were pelleted at 8000 rpm for 30 seconds, the supernatant discarded and then resuspended in 100µL/0.2 OD units of 900 mM LiOAC/TE. For each transformation 8µL of a standard plasmid mini-prep that had not been RNase treated (the bacterial RNA acts as a carrier), 12µL of yeast mixture and 45µL of sterile 50% PEG 4000 were mixed in a glass test tube. The tube was incubated at 30°C with agitation for 1 hour followed by a 5 minute heat shock at 42°C and plating onto selective plates.

2.2.16 - Analytical total yeast protein preparation

Yeast were grown overnight at 30°C in selective liquid medium (usually a 50 mL volume). After pelleting in a benchtop centrifuge, the yeast were resuspended in 6µL of water and 3µL of sample buffer per $O.D._{600}$ unit/mL of yeast. This suspension was transferred to a microfuge tube containing 50µL of acid washed glass beads. The tube was then vortexed for 2 minutes, followed by boiling for 2 minutes. The vortexing and boiling were repeated then a hot 18 gauge needle was used to pierce the bottom of the tube (to form a 'smile' shaped hole), and then the protein extract was allowed to flow out into a fresh tube. The sample was centrifuged at high speed to pellet any remaining debris prior to loading on an SDS gel.

2.2.17 - Maintenance of *Drosophila* stocks

Stocks were routinely cultured at 25°C in plastic vials or half-pint glass bottles containing *Drosophila* culture medium. Stocks needed for collections of large numbers of eggs were maintained in a population cage.

2.2.18 - Isolation of genomic DNA from adult *Drosophila*

20 adult flies were collected in a microfuge tube and placed on ice. They were macerated with glass pestle in 100 µl of ice cold 0.1 M Tris-HCl pH 9, 0 0.1 M EDTA followed by addition of 100 µl of 2% SDS and incubation at 65°C for 30 min. 42 µl of 5 M CH_3COOK was added, the mixture left on ice for 30 min and cell debris pelleted at 12,000g in a microfuge. The spin was repeated on the supernatant after another 10 min on ice. DNA was pelleted by addition of 120 µl isopropanol and spinning at 12,000g for 10 min in a microfuge. DNA (typical yield 20 µg) was resuspended in 20 µl H_2O.

Routinely, 10 µg of DNA was restricted and electrophoresed in each track of a 5 x 13.5 x 0.5 cm horizontal 1.2% agarose gel for Southern analysis.

2.2.19 - Southern blot analysis

After visualisation with ethidium bromide, agarose gels were soaked in 0.25 M HCl, then in 0.5 M NaOH, 1.5 M NaCl until bromophenol blue in the gel turned yellow, then blue respectively. Following neutralisation in 1.0 M Tris-HCl pH 7.4, 1.5 M NaCl for 30 mins the gel was placed in a sandwich that consisted of (from bottom): 5 sheets of 3MM paper drawing on a reservoir of 20 x SSC, the gel, a sheet of nitrocellulose (prewetted in H_2O), 3 sheets of 3MM paper, 5 cm of paper towels, and a glass plate with a 0.5 kg weight upon it. Sandwich was left for 1-16 hours and then dismantled, with orientation marks made on the nitrocellulose. The membrane was allowed to air dry then UV cross - linked. Radio-labelled probe was hybridised to the DNA affixed to the nitrocellulose as in 2.2.5.

2.2.20 - Polytene *in situ* hybridisation

Digoxigenin labelling of probes

200 ng of linearised pBS-*2.1* was added to 100 µg of random hexamers in a total volume of 13 µL. The DNA was denatured at 95°C and then snap cooled on ice for 10 min. 2 µL of Vogel's buffer (0.95 M Pipes pH 6.6, 50 mM $MgCl_2$, 92.5 mM β - mercaptoethanol), 4 µL of 5 X dNTP + DIG - 11 - dUTP mix (1 mM each of dATP, dCTP, dGTP, 0.65 mM dTTP, 0.35 mM Digoxigenin - 11 - dUTP) and 1 µL of 10 U/µL Klenow fragment were added . The mixture was incubated at 18°C ON. The mixture was heated to 95°C for 10 min to denature the DNA again and then snap cooled on ice for 10 min. A further 1 µL of Klenow was added and incubation was at rt for 4 hrs. The reaction was stopped by adding 21 µL of 100 mM EDTA followed by a 10 min incubation at 65°C. 82 µL of H_20, 2 µL (20 mg/mL) glycogen, 11 µL (4 M) LiCl and 300 µL of 95% ethanol were added and mixed in. Precipitation of the labelled DNA was at -80°C for 30 min. The DNA was pelleted at maximum speed for 30 min, washed with 70% ethanol, air dried and resuspended in 50 µL of H_20.

Preparation of chromosome squashes and hybridisation

Salivary glands were dissected in 0.7% NaCl and as much fat body as possible was removed. The glands were transferred to a siliconised coverslip with a drop of fresh 45% acetic acid for 1.5 - 2 min. After this time the liquid was drawn off carefully with a small piece of tissue while viewing under the dissecting scope so as not to pick up the glands. 15 µL of freshly prepared fixing solution (1:2:3 lactic acid:H_20:acetic acid) was added to the glands and left for 4 - 5 min. The coverslip was picked up with a clean slide, and the chromosomes were spread by tapping the coverslip with the 'lead' end of a pencil. The chromosomes were viewed under phase - contrast to determine if adequate spreading had occurred. The slide/coverslip was then inverted (coverslip down) onto a piece of paper towel and squashed ever so firmly with a thumb. Chromosomes were viewed under phase for flatness (absence of refractile appearance). The location of the chromosomes was marked on the slide with a diamond

pencil. The slides were incubated (stored flat) at 4°C ON. The slides had their coverslips popped off as in section 2.2.34 and were then placed in -20°C 95% ethanol for 30 min. The slides were then air dried. Slides were washed in 2 X SSC at 65°C for 30 min, then dehydrated through 2 X washes in 70% ethanol and 1 wash in 95% ethanol at rt. The slides were then air dried. To denature the chromosomes, the slides were placed in freshly prepared 0.07 M NaOH for 2 min at rt, followed by 3 X washes in 2 X SSC (5 min each at rt). The dehydration steps were repeated and the slides were air dried again. 9.5 µL of hybridisation mix (5 µL of probe, 7.5 µL of 20 X SSC, 12.5 µL H_2O and 25 µL deionised formamide - heat denatured at 65°C for 15 min, chilled on ice for 5 min) was applied per slide over the chromosomes. The slides were covered with an 18 X 18 mm plastic coverslip and the edges of the coverslip sealed with 2 layers of bike tube repair cement. Incubated in a humid chamber at 37°C ON. The glue was peeled off and the slides were washed twice in 3 X SSC/ 50% formamide at 37°C (the coverslips fall off). The slides were then washed twice in 2 X SSC at rt. For colour detection, the slides were washed in PBST at rt for 20 min, followed by an incubation for 1 hr with a 1:1000 dilution of anti - dig - AP conjugated antibody diluted in PBST, at rt. The slides were washed 4 X in PBST (10 min each at rt), then incubated for 5 min in staining buffer (100 mM NaCl, 50 mM $MgCl_2$, 100 mM tris - HCl pH 9.0, 0.1% Tween - 20), followed by an incubation in staining buffer with 4.5 µL of (100 mg/mL in 70% DMF) NBT and 3.5 µL of (50 mg/mL in 70% DMF) BCIP. The colour was developed with occasional checking under the microscope. When done the slides were washed in PBST, then counterstained with Giemsa before being mounted in water for photography.

2.2.21 - Phage expression library screen using radio-labelled protein

Expressing the protein onto nitrocellulose filters and pre-treating them

A single colony of y1090r- was inoculated into 50 mL of LB + 0.2% maltose and 50 µg/µL ampicillin and grown overnight at 37°C with shaking. At the same time 6 large (13.7 cm) LB plates were dried overnight with the lids on at 37°C. The next morning the plates were pre-warmed with the lids off and the plates turned upside down at 42°C. The y1090r- culture was pelleted in a 50 mL tube in a bench centrifuge, and resuspended in 20 mL of 10 mM $MgSO_4$. 200 µL of cell solution was mixed with 30,000 pfu of the library and allowed to adsorb at 37°C for 15 min. 7-8 mL of LB 0.7% agarose/10 mM MgSO4 pre-cooled to 47°C was added, mixed in gently and poured onto one LB plate, swirling to cover the plate evenly. Once all the plates were layered identically, and the layer was set, they were incubated at 42°C for 3.5 hrs to achieve lytic growth.

The plates were then transferred to a 37°C room. Nitrocellulose filters pre-treated by soaking in 10 mM IPTG then air drying on cling wrap for 30 minutes were layered onto the plates (being careful not to introduce air bubbles between the filter and the agarose). Distinctive orientation marks were applied using India ink and a 26 gauge needle at this time. The plates were then incubated for 5 hrs at 37°C. The filters were gently peeled off and placed protein side up on 3MM paper to air dry for at least 30 min. The second round of filters (for the duplicate lift) were applied to the plates in the same way and incubated at 37°C for a further 14 hrs. During this time the first set of filters was stored interleaved

with the nitrocellulose packaging paper in a container at 4°C while waiting for the second set of filters. The second set of filters were removed and air dried as for the first set of filters. All the filters were then rinsed thoroughly in TBST. The filters were pre-hybridised in TBST containing 3% BSA - into each of 4 large petri dishes 20 mL of the pre-hybridisation fluid was dispensed, 5 filters were layered into each (protein side down), one by one, using a glass spreader each time to dislodge air bubbles that became trapped under the filters. The filters were placed on a gently orbiting platform to pre-hybridise at RT while preparing the labelled protein.

Probing the prepared filters

10 mL of 3% BSA in TBST was dispensed into each of 2 petri dishes. Probe solution (pre-hybe + purified probe - approx $7x10^5$ cpm/filter) was added to each and the dishes were swirled to mix. The pre-hybridised filters (protein side down) were transferred, one by one, to the petri dishes with the hybridisation solution, once again using a spreader to dislodge trapped air bubbles. Once all the filters had been transferred, a sheet of plastic cut to the shape of a filter was layered on top of the stacks of filters, once again dislodging any trapped bubbles (the plastic sheet helps to keep the top filter wet during the hybridisation). The dishes were sealed in parafilm to make them airtight, and incubated them at 4°C overnight on a gently orbiting platform. The next day the dishes were transferred to an orbiting platform at room temperature and allowed to equilibrate to room temperature. The filters were removed and washed 3 times for 5 min each in large volumes of TBST. The filters were then placed protein side up on 3MM paper and dried thoroughly. Using the 3MM as a backing paper, the filters were stuck down at their very edges with magic tape, and exposed directly to fast X-ray film for one week.

Note : the hybridisation solution was saved for the second round screen, and supplemented with freshly labelled protein.

TBST- 50 mM Tris-HCl pH 7.5
150 mM NaCl
0.05% Tween-20

Preparing the radio-labelled protein

^{35}S-PCL was prepared using the TnT coupled *in vitro* transcription/ translation kit (Promega) according to the manufacturer's instructions. ^{35}S-Methionine was incorporated as the label. After the reaction was completed, the protein was separated from the unincorporated label by spin column chromatography using Sephadex G-25 (Pharmacia) equilibrated in TEN. An aliquot of the protein was analysed using SDS-PAGE and scintillation counting.

TEN- 10 mM Tris-HCl pH 7.5
0.1 mM EDTA
150 mM NaCl

2.2.22 - Direct protein binding to proteins immobilised on nitrocellulose membranes ("far-western blot analysis")

Nitrocellulose blots of the desired proteins were prepared as described elsewhere in this section. The filter was washed thoroughly in TBST. The filter was treated with HBB (20 mM HEPES-KOH pH 7.6, 1 mM KCl, 5 mM MgCl2, 1 mM ZnSO4, 10 mM beta-mercaptoethanol) containing 6M guanidine-HCl for 30 minutes. Then, to renature the proteins with HBB containing half the previous concentration of guanidine-HCl until this reached 0.75M. The filter was then washed extensively in HBB alone, followed by HBB (with variable concentrations of KCl)+ 5% Blotto to block. Approximately $5X10^3$ cpm of *in vitro* translated protein (see above) per square centimetre of membrane was added in approximately 3 mL of HBB (with variable KCl) + 1% Blotto and allowed to hybridise at 4^oC overnight. The membrane was washed at room temperature 3 times for 10 minutes each in PBS + 0.2% Triton X-100, dried completely and then exposed to film or a phosphorimaging plate.

2.2.23 - Expression of bacterial fusion proteins

Clones in pMAL c-2 and the pGEX plasmids were transformed into bacterial strains DH5α or XLI-blue. Clones in pET15b were transformed into BL21 or BL21/pLysS. A single colony was transferred into a flask containing LB-broth and 100µg/mL ampicillin and grown overnight at 37^oC as a pre-culture. A 1:10 dilution of this culture was made into a flask containing LB-broth and 150-250µg/mL of ampicillin (the higher concentration of antibiotic serves to maintain the plasmid under the potentially stressful conditions of induction), incubated at 37^oC until the $O.D._{600}$ reached 0.6-0.8. IPTG was added to a final concentration of 0.1-0.3 mM (for the pMAL and pGEX plasmids) or 1 mM (for the pET15b clones). The culture was then incubated at 37^oC for a further 3-6 hrs to allow the accumulation of expressed protein. pMAL-*2.1* clones were incubated at 25^oC overnight to enhance the expression of soluble protein. The bacteria were then pelleted at low speed and the medium discarded.

2.2.24 - Harvesting expressed protein

The bacterial pellet was resuspended in 1/50th the original culture volume using a buffer that was appropriate for the particular use of the crude protein. The suspension was sonicated to lyse the bacteria using a MSE sonicator at medium power level and amplitude setting number 4, using 3 cycles of 30 seconds of sonication followed by 30 seconds rest, whilst the sample was on ice. At this point the suspension was either directly centrifuged at 10000 X g in a Beckman JA-20 rotor for 10 minutes or was treated as described in a later section. Lysis of BL21/pLysS bacteria was carried by simply freezing the bacterial pellet at -70^oC for 1 hour prior to resuspension in an appropriate buffer and centrifugation as described just above. The supernatant was transferred to a fresh tube and the pellet was resuspended in an appropriate buffer (usually PBS + 1% Triton X-100). Then, 10µL of each was mixed with sample buffer, boiled for 2 minutes and analysed by SDS-PAGE in comparison to

identically prepared samples from bacteria expressing the parental plasmid alone, to determine both the levels of induction and the solubility of the fusion protein.

2.2.25 - Purification of 2.1-MBP or 6Xhis-PCL fusion proteins

Purification of 2.1-MBP and 6Xhis-PCL (both were solubly expressed) fusion proteins was done using amylose resin (New England Biolabs) and Ni-NTA silica resin (Qiagen) respectively using the manufacturers' specifications.

2.2.26 - Solubilising GST-PCL fusion protein

The insoluble pellet from a 500 mL induction of GST-PCL was resuspended in 25 mL of 20 mM Tris-HCl pH 7.5, 20% sucrose, 1 mM EDTA followed by a 10 minute incubation on ice. The sample was pelleted at 4000 X g in a Beckman JA-20 rotor for 10 minutes. The pellet was resuspended in 25 mL of ice cold water and incubated on ice for 10 minutes. The bacterial spheroplasts were then pelleted at 8000 X g for 10 minutes. The pellet was then resuspended in 5 mL of PBS, 5 mM EDTA, 0.5 mM PMSF, and sonicated as described above. RNase and DNase was added to a final concentration of 0.1 mg/mL followed by incubation at RT for 15 minutes. The suspension was diluted to 25 mL by the addition of PBS, 5 mM EDTA, and then centrifuged at 13000 X g for 30 minutes to pellet the crude inclusion bodies. The inclusion bodies were washed 3 times for 10 minutes each on ice by resuspension in PBS, 25% sucrose, 5 mM EDTA, 1% Triton X-100, with intervening centrifugation at 25000 X g to pellet inclusion bodies. The washed inclusion body pellet was dissolved in 5 mL of 20 mM Tris-HCl or HEPES-KOH pH 8.0, 5M guanidine-HCl, by incubation on ice for 60 minutes and occasional gentle mixing. This solution was clarified by centrifugation at 12000 X g for 30 minutes, and then added to 50 mL of 20 mM Tris-HCl or HEPES-KOH pH 8.0, 2 mM b-mercaptoethanol, 0.05% Triton X-100, 10μM $ZnSO_4$, 0.5 mM PMSF in a 100 mL Schott bottle under magnetic stirring at 4°C overnight. The substance from the bottle was transferred to a centrifuge tube an spun at 13500 X g for 30 minutes. The supernatant and the pellet were analysed using SDS-PAGE and Coomassie staining to determine the degree of solubilisation of the GST-PCL protein. At this point the GST-PCL protein was in buffer containing 454 mM guanidine-HCL. The protein was then dialysed against the buffer containing gradually decreasing amounts of the denaturant, followed by dialysis against the desired buffer containing 10% glycerol. Samples were then aliquoted, snap-frozen and stored at -70°C. This protocol resulted in a partial purification of the GST-PCL protein.

2.2.27 - Solubilising the GST-cDOM and GST-ΔEH fusion proteins

Both of these fusions were insoluble when expressed using the method described above for full length GST-PCL. The insoluble pellet from a 500 mL culture was dissolved in 10 mL of Tris-HCl, 10μM $ZnSO_4$, 8M urea. After centrifugation at 13000 X g to clarify the solution, it was dialysed serially against

the buffer containing 4M, 2M, 1M, 0.5M and then no urea. This was followed by dialysis against the desired buffer containing 10% glycerol, snap-freezing and storage at -70^{o}C.

2.2.28 - Protein gel electrophoresis and western blotting

All SDS-PAGE of protein samples and subsequent western transfer to nitrocellulose membrane, Coomassie blue staining of gels and Ponceau S or India ink staining of blots was performed exactly as described by (Harlow and Lane, 1988).

Nitrocellulose blots were washed thoroughly with PBS + 0.1% Tween-20 and then blocked for 1 hour in PBS + 0.1% Tween-20, 5% Blotto. Primary and secondary antibody incubations were carried out overnight at 4^{o}C and for 45 minutes at room temperature respectively, with the appropriate dilutions of antibody in the described blocking solution. The secondary antibodies were almost always horseradish peroxidase conjugated (Jackson) and therefore detection was either by the Enhanced Chemiluminescence method (Amersham) or by colorimetric detection using nickel enhanced DAB staining (Harlow and Lane, 1988).

2.2.29 - Preparation of antigen for immunisation of rats

GST-2.1 protein was electrophoresed on SDS containing acrylamide gels. The protein was visualised by staining the gel in Coomassie prepared in water (Harlow and Lane, 1988), and the appropriate gel slice excised. The slice was homogenised in an equal volume of PBS by passage through gradually narrower needles. The gel slurry was given to Joe Wrin for appropriate mixing with adjuvant and administering to rats. Approximately 10µg of antigen was given to each rat at each injection. Unfortunately, no pre-immune serum was harvested. As a pre-immune control, serum from rats of the same strain and from the same animal house was used. Serum was tested on nitrocellulose strips containing the GST-2.1 antigen, for levels of antibody activity.

2.2.30 - Preparation and use of acetone powders for pre-absorbing anti-serum

A 500 mL culture of bacteria containing a parental pGEX plasmid was grown to an O.D.600 of approximately 1.0. Expression was induced with IPTG as described above. The bacterial pellet was weighed and resuspended in 0.9% NaCl to 50% w/v. The suspension was chilled on ice for 5 minutes. 4 volumes of acetone pre-chilled to -20^{o}C was added to the suspension and mixed vigorously with a vortex. The tube was incubated on ice for 30 minutes with occasional vortexing. The precipitate was harvested by pelleting at 10000 X g for 10 minutes, resuspended in fresh acetone and re-pelleted at 12000 X g for 10 minutes. The supernatant was poured off and the pellet was transferred with a spatula to a piece of glass. The pellet was dispersed and pulverised as it was air drying using a spatula. The dry, fine, ivory coloured powder was stored at 4^{o}C in an airtight tube. To pre-absorb anti-serum

against bacterial proteins and GST, 1% of powder was added to the serum and incubated at 4°C for 1 hour. The tube was centrifuged at high speed and the supernatant used for immunoreaction.

2.2.31 - Electrophoretic mobility shift assays (EMSA)

Soluble protein samples that were stored in 20 mM Tris-HCl, 60 mM KCl (occasionally) and 10-20% glycerol were used in EMSAs. Up to 10µL of protein was mixed with approximately 0.1 nM end labelled target DNA, 50 ng/µL of herring sperm DNA or poly-dIdC (as a non-specific DNA competitor) in a buffer that contained 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM KCl, 0.1 mM DTT, 5% glycerol, 50µg/mL of acetylated BSA (as a non-specific protein competitor) in a total volume of 20µL. The sample was incubated at room temperature for 30-45 minutes. In cases where specific DNA competitor was used, it was added prior to the addition of the probe with a pre-incubation for 20 minutes followed by incubation with the probe for a further 30 minutes. 2µL of 64% glycerol buffered as the EMSA reactions was added to each reaction prior to resolution of protein/DNA complexes on native 5%-12% (depending on the size of the DNA probe) 30:1 polyacrylamide gels containing 10% glycerol and 0.5 X TBE. The running buffer was 0.5 X TBE and the gels were left to electrophorese at 30V at 4°C until the bromophenol blue (loaded in an adjacent track as a marker) migrated to the bottom of the gel.

2.2.32 - Purification of antiserum using Protein-G affinity

Purification of antiserum using Protein-G affinity was performed using the MAb-Trap kit (Pharmacia) using the manufacturers protocols.

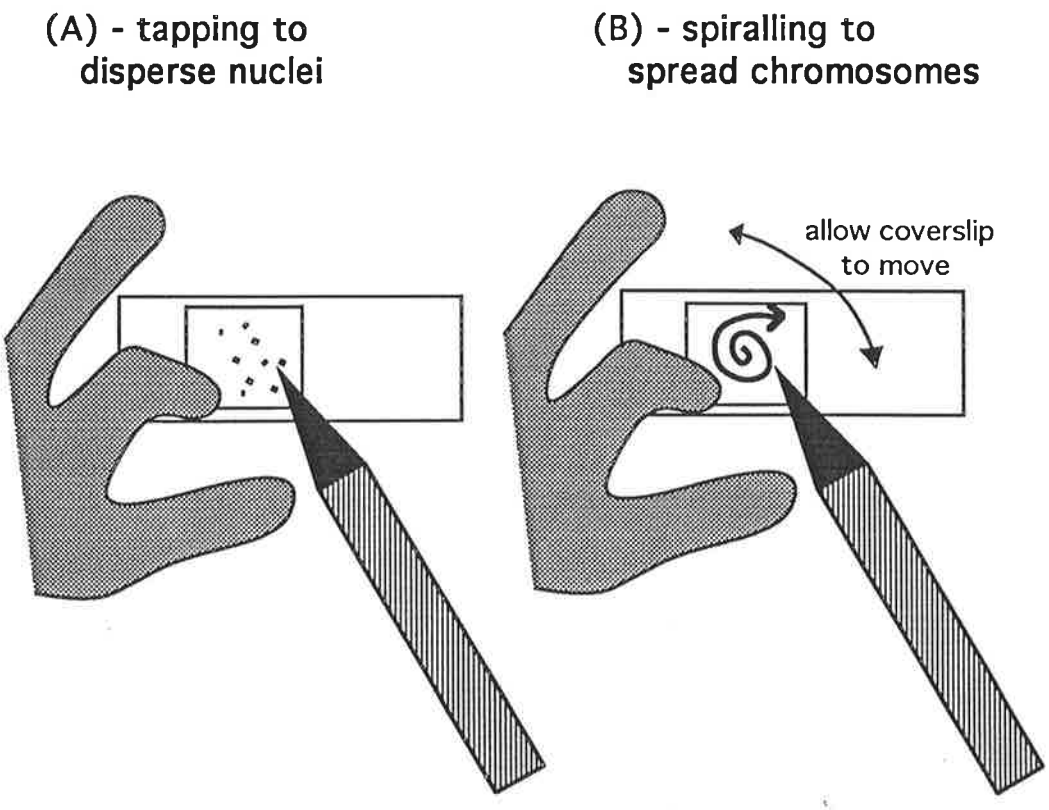
2.2.33 - Affinity purification of polyclonal antiserum

Affinity purification was performed using the nitrocellulose strip method. The antigen used was always bacterially expressed 2.1 or PCL fusion protein. Approximately 5-10 µG of the unpurified fusion protein was resolved electrophoretically on 2 mm thick preparative SDS gels using the Bio-Rad mini protean II gel apparatus. Protein in the whole gel was transferred to nitrocellulose and the presence of the antigen on the blot was determined by staining the blot with Ponceau S. The smallest strip of nitrocellulose containing all the antigen was excised from the membrane, rinsed in PBS + 0.1% Tween-20 and then blocked for 1 hour in PBS + 5% Blotto. The strips were then washed in PBS for 15 minutes to remove any excess Blotto. The strips were layered, protein side up, onto a piece of parafilm in a humid chamber (e.g. a petridish with a damp tissue inside), and 250-300µL of crude antiserum was gently pipetted on top. Incubation occurred for 2 hours with agitation of the chamber on an orbiting platform. After this time the serum was removed from the strip and saved for analysis as a depleted fraction. The strips were washed in PBS for at least 30 minutes with several changes of buffer. Antibodies were eluted in different stages, using different elution conditions to account for any differences in the affinities of the IgG molecules present. Two elutions of 100-150µL each using 200 mM glycine, 1 mM EGTA, pH 2.5 layered onto the strip and incubated with agitation for 15 minutes

were done. The eluates were neutralised by the addition of 5M KOH (the pH was tested by spotting some solution onto pH test strips). This was followed by two elutions of 100-150µL each of 4.5M $MgCl_2$. The eluates were tested for antibody activity by western blot on bacterially expressed fusion protein or directly on fixed *Drosophila* embryos. Fractions containing satisfactory levels of antibody activity were dialysed against large volumes of PBS + 20% glycerol. Sodium azide was added to the samples to a final concentration of 0.02% prior to storage at 4°C.

2.2.34 - Polytene immuno-stainings

Third instar larvae were raised in uncrowded conditions in bottles of standard cornmeal medium containing fresh live yeast. Larvae that had crawled out of the medium and had slowed down in preparation for pupariation were selected. The larvae were rinsed in PBS + 0.1% Triton X-100 and then their salivary glands were dissected in a fresh lot of the same solution. Once all the glands had been dissected (usually 10 pairs), 2 pairs of glands at a time were transferred to a petri dish containing PBXF and allowed to incubate for 30 seconds (this time was determined empirically for 2.1 immunostaining). Each set of glands (2 pairs) was transferred into a droplet of FA on a 22 mm X 22 mm glass coverslip. After 2 minutes incubation, the coverslip was picked up with a poly-L-lysine (Sigma) treated glass microscope slide and then inverted so that the coverslip was facing up. The bottom left hand corner (I am right handed) of the coverslip was held down gently with a gloved (to prevent finger fixation) index finger. Using a pencil (with a not too sharp end) the coverslip was tapped gently about 10 times to disperse the tissue (A). Then, using the pencil once again, a spiral was traced on the coverslip starting at its centre using gentle pressure. The coverslip should move during this procedure (B).



The slide/coverslip sandwich was then inverted onto a paper towel and squashed really firmly with the thumb. The position of the coverslip was marked onto the back of the slide using a diamond pencil. The slide was then dipped into liquid nitrogen until the sizzling stopped, removed and the coverslip 'popped' off using a scalpel blade. The slide was left at room temperature until the frozen condensation on it was thawed, and then it was placed into a Coplin jar containing PBS. When all the slides were in the Coplin jar, it was placed on an orbiting platform for 15 minutes. After this time the PBS was poured off and replaced with PBN and incubated for at least 30 minutes at room temperature with agitation. The slides were drained but not to complete dryness and then used to pick up 22 mm X 22 mm coverslips with 16μL of antibody diluted (the dilution was determined empirically - it was 1:12 for affinity purified rat anti-2.1 and 1:10 for affinity purified rabbit anti-PCL) in PBN spotted on them. The slides were then incubated overnight at $4^{o}C$ in a humid chamber. The coverslips were then removed by jiggling each slide in a large beaker of PBS and then the slides were placed into a Coplin jar containing PBS + 0.1% NP-40 and washed with agitation for 10 minutes. This wash step was repeated twice more. The wash solution was replaced with PBN and the slides were incubated with agitation for 30 minutes. The slides were incubated with fluorescently labelled secondary antibodies (Jackson) in the identical manner as for the primary antibody incubation at a dilution of 1:200 in PBN in a dark humid chamber for 3 hours at room temperature. Occasionally this solution also contained Chromomycin A3 at 20μg/mL and 10 mM MgCl2 as a DNA specific counter-stain. The slides were then washed as before, drained and then mounted under a 22 mm X 22 mm coverslip using 8μL of antifade mountant, occasionally containing 10μg/mL of bis-benzimide (Hoechst 33258) as a DNA specific counter stain. The chromosomes were viewed using a Zeiss axiophot epifluorescence microscope, and then images were captured using a Bio-Rad MRC 1000 laser scanning confocal microscope connected to a Nikon microscope using a 60 X oil - immersion objective. The images were adjusted but not internally modified using Adobe Photoshop v3.0 on a Macintosh Quadra 840av and printed on a Kodak XLT 7720 digital continuous tone printer.

PBXF- 1% Triton X-100
3.7% formaldehyde*
in PBS

FA- 50% acetic acid
3.7% formaldehyde*

PBN- 2% BSA
0.5% NP-40
in PBS

antifade mountant-
30 mM Tris-HCl pH 9.0
70% Glycerol

2.5% n-propyl gallate

*this was diluted from 37% paraformaldehyde dissolved by boiling in 15 mM KOH.

2.2.35 - Harvesting and 'fixing' *Drosophila* embryos

Embryos were collected on grape juice agar plates smeared with yeast. They were then harvested and washed thoroughly in a sieve using copious amounts 'embryo wash buffer'. The sieve was then transferred into a container with 50% commercially available bleach (2% sodium hypochlorite) for 2 minutes to de-chorionate the embryos. The embryos were once again washed in the sieve thoroughly using 'embryo wash buffer'. They were then transferred to a glass scintillation vial containing a two-phase mix of 4 mL of 4% formaldehyde in PBS (made fresh by boiling paraformaldehyde in PBS) and 4 mL of heptane. The vial was then shaken on an orbiting platform such that the interface between the liquid phases was disrupted and the embryos were bathing in an emulsion, for between 15 and 30 minutes to 'fix' the embryos. The bottom phase (aqueous) was drawn off and replaced with 4 mL of methanol and the vial was shaken vigorously for 1 minute to de-vitellinise the embryos. De-vitellinised embryos sink from the interface and were collected from the bottom phase (methanol). Embryos were rinsed thoroughly in methanol at which point they were either processed for whole mount in situ hybridisation, immuno-staining or storage at -20^{o}C in methanol.

embryo wash buffer- 0.7% NaCl
0.15% Triton X-100

2.2.36 - Whole mount immuno-staining of *Drosophila* embryos

The methanol was removed from embryos in a microfuge tube and replaced with PBST. Several rinses were done using PBST followed by a single wash for 30 minutes. The embryos were then 'blocked' in 1 ml of PBST containing 5% Blotto for at least 1 hour. The blocking solution was removed and primary antibody diluted in fresh blocking solution was added (usually 200µL). The embryos were routinely incubated with gentle agitation at 4^{o}C overnight. The next day, the antibody solution was removed and the embryos were washed extensively in PBST (several changes of buffer over a 2 hour time period). The embryos were then incubated with secondary antibody diluted in fresh blocking solution for at least 2 hours at room temperature with gentle agitation. Following a period of washing as for the primary antibody, the embryos were mounted in PBS/80% Glycerol (if the secondary antibody was fluorescent tag conjugated). If the secondary antibody was conjugated with horseradish peroxidase, the antibody localisation was detected colorimetrically. The embryos were incubated in a solution of 0.5 mg/mL DAB, 0.045% H_2O_2, 0.064% $NiCl_2$ until the staining had developed (as assayed on a dissecting microscope), and then rinsed thoroughly using PBST prior to mounting in PBS/80% Glycerol. Occasionally, DAB reacted embryos were cleared in methyl salicylate

to give better visualisation of the staining pattern. In this case, embryos washed in PBST were dehydrated in 95% ethanol and then transferred to and mounted in methyl salicylate.

2.2.37 - Analytical preparation of protein extract from *Drosophila*

Approximately 100 de chorionated embryos were homogenised then boiled in 100 µL of sample buffer. A single third instar larva was homogenised in 100 µL of 50 mM HEPES-KOH pH 7.6, 50 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.05% NP-40 and protease inhibitors. 100 µL of sample buffer was added prior to boiling.

2.2.38 - Regulatory considerations

All manipulations involving recombinant DNA were carried out in accordance with the regulations and approval of the Genetic Manipulation Advisory Committee and the University Council of the University of Adelaide.

All manipulations involving animals were carried out in accordance with the regulations and approval of the Animal Ethics Committee and the University Council of the University of Adelaide.

2.3 - Abbreviations

Abbreviations are as described in "Instructions to authors", *Biochem. J.* (1978) **169**, 1-27. In addition:

aa	amino acid residues
APS	ammonium persulphate
BCIG	5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside
bisacrylamide	N,N'-methylene-bisacrylamide
bp	base pairs
BSA	bovine serum albumin
blotto	skim milk powder
CIP	alkaline calf intestinal phosphatase
Da	Dalton
DTT	dithiothreitol
GST	glutathione-s-transferase
HEPES	N-2-Hydroxyethyl piperazine-N-2-ethane sulphonic acid
HRP	horse radish peroxidase
IPTG	isopropyl-b-D-thiogalactopyranoside
kb	kilobase
MBP	maltose binding protein
nt	nucleotide
ON	overnight

ORF	open reading frame
PEG	polyethylene glycol
pfu	plaque forming unit
PMSF	phenylmethanesulfonylfluoride
rpm	revolutions per minute
rt	room temperature
SDS	lauryl sulphate sodium salt (sodium dodecyl sulphate)
TEMED	N,N,N',N'-tetramethylethylenediamine
U	unit

Chapter 3 - Screening for PCL interactors

3.1 - Background

As outlined in chapter 1, a rational strategy for determining the mechanism of repression of the Pc-G proteins would include the identification of proteins that interact physically with any of the individual members. Determination of the sequence of these interactors could reveal that proteins with known function or characterised functional domains had been isolated. For example, if histone proteins were found to interact with a Pc-G member, it would be reasonable to establish the working hypothesis that the Pc-G exerts its repressive effect by modifying the superstructure of nucleosomes.

The main aim of the work that this thesis describes was to identify protein/s that interact with one member of the Pc-G, PCL. Although PCL was chosen for the practical reason that it was isolated and characterised in this laboratory, there are very good reasons why if given the choice of any Pc-G proteins with which to isolate interactors, PCL would be selected. As outlined in chapter 1, *Pcl* mutants show amongst the most consistent and strong genetic interaction with modifiers of PEV (H. Brock, personal communication) as well as interacting more consistently with Pc-G mutants than any other individual Pc-G mutant (Campbell *et al.*, 1995). These data strongly suggest that PCL is a central component of the Pc-G complex and that it has a critical role in its function.

Although many different ways were available to screen for protein interactors, it was decided that a good method would be to probe a λ GT11 expression library with radio-labelled PCL. This screen was chosen for two main reasons. Firstly, the identification of an interactor would be concomitant with the cloning of the corresponding cDNA (unlike co-immunoprecipitation for example). Secondly, detection of an interaction would not be reliant on transcription of a reporter gene (unlike the yeast two-hybrid system) which may be obscured by the repressive effects of PCL. As disadvantages, the sensitivity of the system chosen may be lower and post-translational modifications potentially required for interactions *in vivo* would not be present.

3.2 - Cloning *2.1* by screening a λ GT11 library

^{35}S-methionine labelled PCL was produced by coupled *in vitro* transcription and translation using the plasmid pBS-*Pcl* Sac/Not (SacI/NotI fragment from AL15 (Lonie *et al.*, 1994) cloned into pBS KS+) as the template and T7 RNA polymerase for transcription. ^{35}S-methionine was chosen as the labelled amino acid as the predicted amino acid sequence of PCL contained 13 methionine residues and therefore a probe of high specific activity could be synthesised. The labelled protein was assayed for the level of radioactivity by scintillation counting as well as SDS-PAGE followed by autoradiography. Figure 3.2.1a shows a western blot, probed with anti-PCL antibodies, of a test transcription/translation reaction run along side larval protein extract. The *in vitro* generated protein migrated at approximately 110 kD which did not equate with the predicted size from the conceptually translated sequence (95 kD) but was the same as had been previously observed (Lonie, 1994), and differed significantly from

the protein *in vivo* (140 kD) as previously noted (Lonie, 1994). That the translated protein was specifically recognised by anti-PCL antibodies from amongst the proteins present in the reticulocyte lysate indicated that PCL was indeed produced in the transcription/translation reaction. The transcription/translation was then repeated on a larger scale utilising ^{35}S-methionine label to generate a probe. Figure 3.2.1b shows the autoradiograph of a small aliquot of the reaction resolved by SDS-PAGE. Once again, an approximately 110 kD protein was present which presumably corresponded with PCL. Given that the majority of the radioactive label had incorporated into the full length PCL protein and not into other background proteins, it was decided that it was unnecessary to purify the synthesised PCL protein as any positive in the screen would almost certainly be due to PCL binding.

Approximately 4X10^{6}cpm of labelled PCL was used to screen 180,000 pfu of a 0-18 hr embryonic cDNA λ GT11 library. Autoradiography of the resultant filters was for one week. From this preliminary screen, one duplicate positive (clone λ *2.1*) was recovered (figure 3.2.1c and d). The size (approximately 2.1 kb) and orientation of the insert in the λ GT11 vector were determined by restriction digestion. The EcoRI-ended insert was then sub-cloned into the bacterial expression plasmid pGEX1 in the orientation that would result in expression of the same protein except fused to GST rather than beta-galactosidase (as is the case for λ GT11 clones). Figure 3.2.1e shows a simple restriction map of λ *2.1* and of the resultant sub-clone named pGEX-*2.1* Eco.

In parallel, the same 2.1 kb insert was sub-cloned into pBS KS+ (clone pBS-*2.1* Eco) for sequence determination. Full sequencing of the *2.1* cDNA followed by database searching showed that 2.1 did indeed correspond to a novel protein (M. Coulson, personal communication). Analysis of the reading frame that was the continuation of the reading frame 1 from the EcoRI site (i.e the ORF corresponding to the protein that would have been expressed from the λGT11 clone) showed that this was the longest reading frame (638 amino acids) and that the codon usage was comparable to most *Drosophila* open reading frames. (M. Coulson, personal communication).

3.3 - Cloning the 5'-end of *2.1*

The sequence of the 5' most end of the 2.1 kb cDNA did not appear to contain the 5' end of the mRNA. The reading frame of the cDNA was expected to be the continuation of reading frame 1 from the EcoR1 site, the site of fusion with the lacZ coding region in λ GT11 (figure 3.3.1a). Although there is an in frame methionine at position 4, there is not the consensus A or C at position -1 expected of a *Drosophila* translation start site (Cavener, 1987). This suggested that this methionine may not be the 'natural' start of translation within *2.1*. Extensive re-screening of the 0-18 hr embryonic λ GT11 library from which the original λ *2.1* clone was isolated did not provide any cDNA clones that extended further 5' than the original clone (M. Coulson, personal communication). In an attempt to isolate the 5' end of the *2.1* transcript, a PCR approach was carried out utilising the directionally cloned embryonic cDNA libraries of Nick Brown (Brown and Kafatos, 1988) as template material. A *2.1* specific non-coding strand complementary oligonucleotide as the 'reverse' primer (primer 2.1bb) and the SP6 primer as the forward primer were used to amplify any 5' extended sequences present in these libraries. An initial round of PCR was performed using a combination of a 0-4 and 4-8 hour embryonic

Figure 3.2.1 - The expression screen and identification of the clone λ *2.1*.

a - Western blot probed with anti-PCL antibodies of *in vitro* translated PCL (~110 kD, lane 1) and larval protein extract (~140 kD)(lane 2). **b** - autoradiography of ^{35}S-labelled, *in vitro* translated PCL (~110 kD). Marker sizes in kD for both **a** and **b** are shown to the left.

c and **d** - sectors of autoradiographs of the resultant primary and duplicate lifts onto nitrocellulose respectively, probed with labelled PCL, showing the duplicate positive corresponding to the clone λ *2.1* (arrows).

e - bar diagram showing the relationship between the clone λ *2.1* (top) and the constructed expression clone pGEX-*2.1* Eco (bottom). RI = EcoRI, K = KpnI

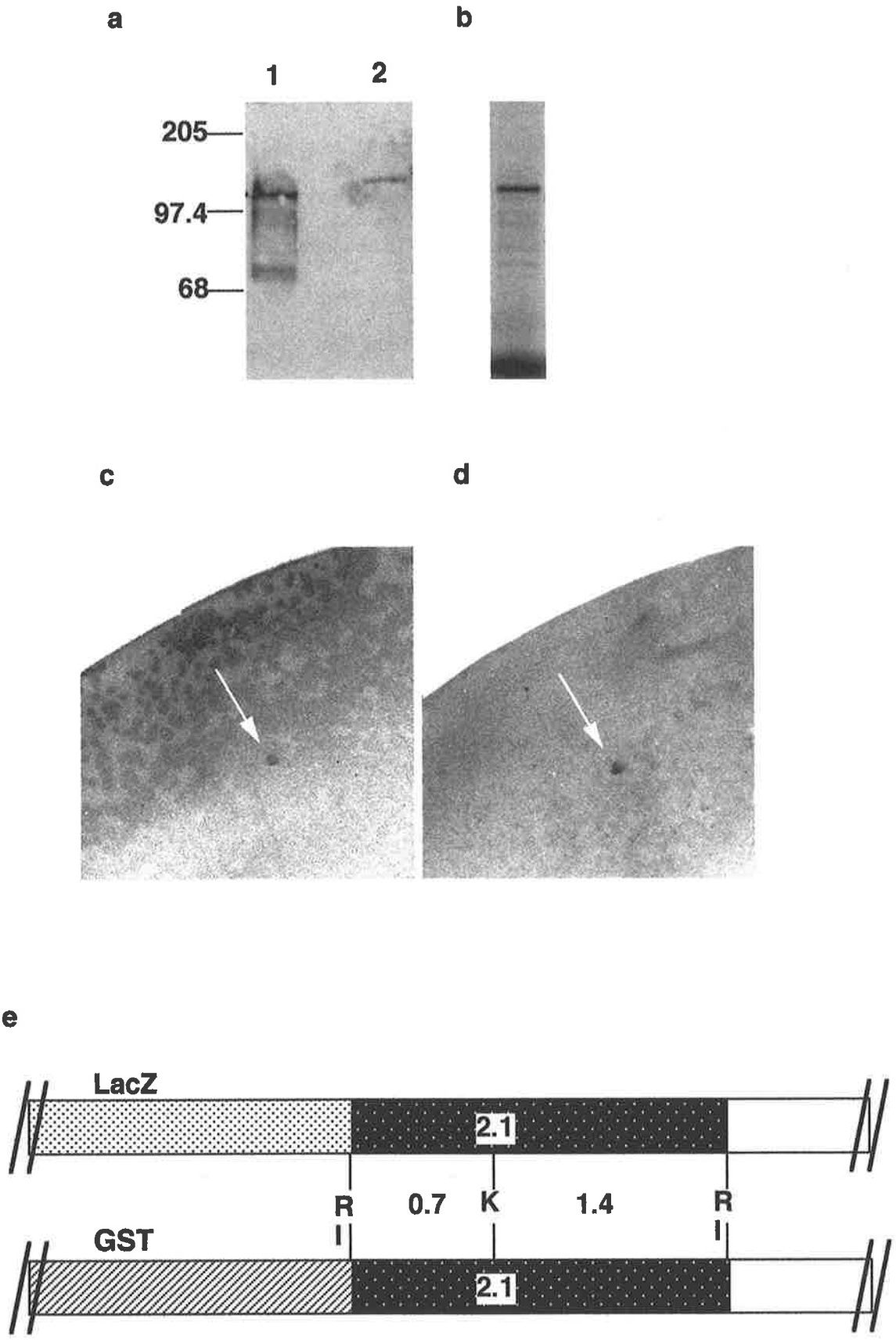

a
b
1
2
205
97.4
68
c
d
e
LacZ
2.1
R
I
0.7
K
1.4
R
I
GST
2.1

Figure 3.3.1 - Cloning and sequence analysis of the 5'-end of the *2.1* cDNA

a - shows the sequence alignment of the 5' most portion of the clone λ *2.1* and the 5'-end PCR clone. The bottom strand is the λ *2.1* clone. Note the engineered EcoRI site at the very 5' end (underlined). The three artifactually added amino acid residues encoded by the EcoRI linker are shown underlined beneath the DNA sequence alignment. The predicted amino acid sequence of the NH2-terminus of 2.1 is shown above the alignment. Two potential initiation codon encoded methionines are shown in bold. The nucleotide residues at positions -1 and -4 from these codons are shown outlined. An in-frame amber stop codon is marked with **AMB**. The poly-dC,G tract at the very 5' end is shown underlined.

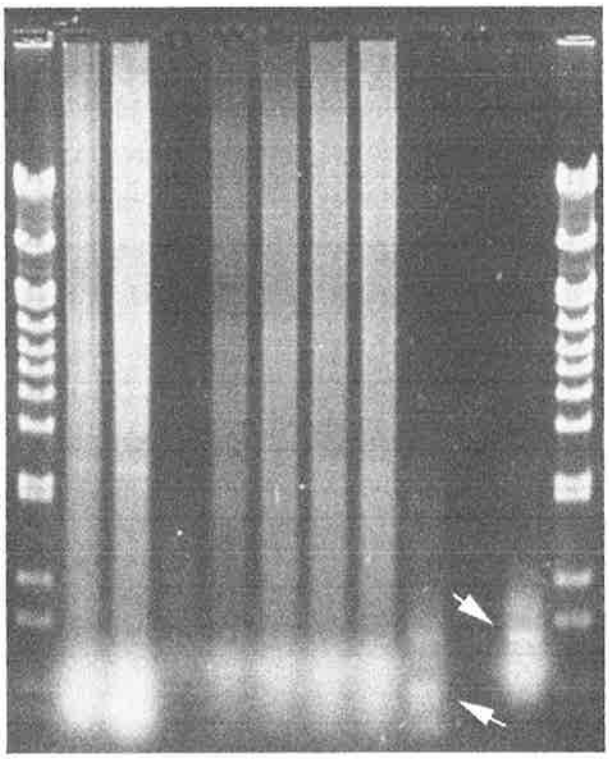
b - The first round of PCR using SP6 primer only (lane 1) and both SP6 and *2.1*bb primer (lane 2). Lanes 3-7 show the second round of PCR using 1, 3, 5, 7, and 9 μL respectively of round 1 PCR product as template. Lanes 8 and 9 show the third round products amplified when the primer sets 5'sequencing/*2.1*bb and SP6/*2.1*bb, respectively, were used. The arrows indicate the 250 bp product (lane 8) and the 450 bp product (lane 9).

c - Autoradiograph of a Southern blot probed with radiolabelled *2.1* cDNA, of lanes 8 and 9 in **b** (lanes 8 and 9 respectively)

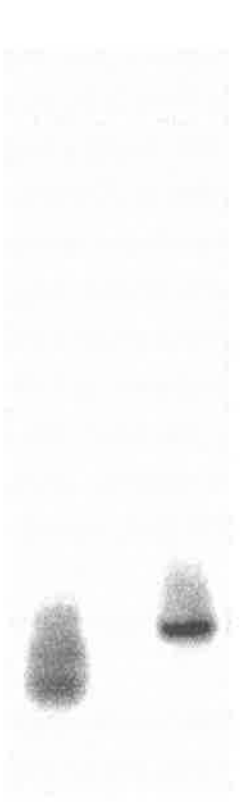
a

```
                          AMB
CCCCCCCCCCCGATTCGATAGCGTTTAGGCCAGCGGTACGAGCGGTACGGA


TACGGACACGATTGCGATTACGGTTACGGTCTTTTCCTGGCCAGCGAATC
                                                ::::
                                               GAATT
                                                E  F

     M  K  P  T  V  M  A  A  N  Q  A  Q  S  N  P
ATCGATGAAGCCCACTGTGATGGCGGCCAATCAGGCGCAAAGCAATCCCG
   :::::::::::::::::::::::::::::::::::::::::::::::::
CCGGATGAAGCCCACTGTGATGGCGGCCAATCAGGCGCAAAGCAATCCCG
  R

A  T  A  A  G  N  N  V  S
CCACCGCCGCCGGTAATAATGTCTCA
:::::::::::::::::::::::::::
CCACCGCCGCCGGTAATAATGTCTCA......
```

b



c



cDNA library as template. Approximately 150 ng/μL of library plasmid DNA, 25 ng/μL of SP6 primer and 5 ng/μL of 2.1bb primer was used. Figure 3.3.1b tracks 1 and 2 show the result of the first round of amplification using only SP6 primer (a negative control) and SP6 and 2.1bb primer respectively. The large 'smear' of DNA present in both tracks is presumably due to the large amount of template DNA alone. In the second round of PCR, five different reactions were performed with the same reaction conditions as the first round of PCR, except that as template, the products generated in the first round of amplification were used. Tracks 3-7 show the amplification products of PCR utilising 1, 3, 5, 7, and 9μL respectively, of first round product diluted into a 20 μL reaction mix. Given that there was some DNA present in track 3, this reaction was used as a template for a third round of PCR, in the hope of diminishing the amount of starting template material, using the same reaction conditions as before. 1μL of the reaction in track 3 was added to a 20μL PCR mix containing the SP6 and 2.1bb primers (track 9) or a specific *2.1* 5' primer complementary to the coding strand at the very 5' end of the λ *2.1* clone (5' sequencing primer), and 2.1bb (track 8). This last reaction was carried out to determine if any *2.1* cDNA that went as far 5' as the existing clone had been amplified in the first 2 rounds of PCR. The expected size of this control PCR product is approximately 250 bp. As can be seen in track 8, an approximately 250 bp product was amplified. Track 9 shows that an approximately 450 bp was amplified in the SP6/2.1bb PCR. This portion of the gel was Southern-blotted and probed with the existing 2.1 kb *2.1* probe. As can be seen in figure 3.3.1c both the 250 bp product (and not the lower molecular weight products) and the 450 bp product hybridised, indicating that they contain *2.1* sequence. The 450 bp product was then gel purified and sub-cloned into the vector pUCBM20 for DNA sequence analysis. Sequence was generated using the SP6 and 2.1bb primers and is shown aligned with the previously generated *2.1* sequence (figure 3.3.1a). As can be seen, there is no other in frame methionine upstream of the methionine that was previously designated to be at position 4. Furthermore, there is an in frame stop codon (TAG) at nucleotide position 18 of the newly generated sequence. This strongly suggested that the methionine previously designated to be at position 4 is the natural start of translation of the 2.1 protein. However given that this methionine and the one 5 residues downstream of it do not have the consensus nucleotide residues at positions -1 and -4, either of these methionines may be the true start codon. It must be one of these two residues, however, to account for the observed size of bacterially expressed 2.1 (see next section). The newly generated sequence contains at its 5' end a poly-C tract followed by a G. Given the way the cDNA library was synthesised, this strongly suggested that this cDNA extends to the 5' end of the *2.1* mRNA (Brown and Kafatos, 1988). Sequencing of genomic clones covering this region, and primer extension or RNAse protection would need to be done to confirm this.

3.4 - Generation and preliminary analysis of 2.1 antibodies

In order to characterise the distribution and function of the 2.1 protein it was necessary to generate specific anti-2.1 antibodies.

Expression of GST-full-length 2.1 (GST-2.1) fusion protein was engineered in *E.coli* using the pGEX-*2.1* plasmid. This fusion protein was approximately 100 kD in size (figure 3.4.1a), which

corresponded with the predicted polypeptide size (~69 kD) with the addition of GST (25 kD). This fusion protein was purified, processed and used to inject rats to raise antibodies. Serum from the terminal bleed of 3 rats was tested for antibody activity on western blots of the same GST-2.1 protein that was used to inoculate the rats. All three serum samples had high levels of antibody activity in this assay (figure 3.4.1c). The sera were pooled and an aliquot was affinity purified against GST-2.1 fusion protein immobilised on nitrocellulose membrane. To test that the affinity purified antibodies recognised the 2.1 moiety of the fusion protein, a second bacterial expression construct pMAL-*2.1* Eco was made by sub-cloning the EcoRI-ended 2.1 kb *2.1* fragment into the EcoRI site of pMAL-c2 (unfortunately, this clone was not made in time to express protein for affinity purification of 2.1 specific antibodies). This clone produced a 2.1-MBP fusion protein of size approximately 120 kD (figure 3.4.1b track 1). This fusion protein and the MBP alone that was expressed from the parental pMAL-c2 plasmid were partially purified on an amylose-agarose column (figure 3.4.1b). 2 identical western blots were prepared that contained the MBP and the 2.1-MBP protein. One blot was stained with India ink to show the total protein on the blot (figure 3.4.1d), prior to being reacted with the affinity purified anti-2.1 antibodies diluted to 1:500 (figure 3.4.1e). The second blot was processed and probed with ^{35}S-methionine labelled PCL protein in a far-western assay (figure 3.4.1f). The data presented in figure 3.4.1d and e show that the affinity purified anti-2.1 antibodies recognise only the 2.1-MBP and not the MBP protein alone. This suggests that the antiserum contained antibodies that were specific for the 2.1 moiety. Figure 3.4.1f shows that *in vitro* synthesised PCL interacts directly with the 2.1-MBP protein specifically. These data, as well as confirming the interaction visualised in the original library screen, showed that the antibodies were indeed generated to the same protein that interacts specifically *in vitro* with PCL.

In order to test the specificity of the generated and purified antibodies for *in vivo* proteins, western blots containing *Drosophila* embryonic and larval proteins were made. As a control, 2.1 was *in vitro* transcribed and translated using the clone pBS *2.1*Eco as a template and T7 RNA polymerase. As is shown in figure 3.4.2, the antibodies recognise 3 electrophoretic species strongly as well as at least 3 others weakly in both embryonic and larval extracts. Of the strongly detected bands one corresponds well with the predicted size of 2.1 at approximately 69 kD. The two other strongly detected bands are present as a doublet of approximately 97 kD. None of these, or the weakly detected bands, are present in identical blots reacted with rat serum or affinity purified PCL anti-serum, suggesting that these proteins are specifically detected by the purified anti-2.1 anti-serum. Surprisingly, the *in vitro* translated 2.1 appears to migrate at approximately 97 kD rather than at the predicted 69 kD. Given that the predicted 2.1 polypeptide size of approximately 69 kD was confirmed by expression of 2.1 in bacteria (see above) and that correspondingly there is a ~69 kD immunoreactive protein in *Drosophila* extracts, it must be assumed that 2.1 is post-translationally modified in the *in vitro* translation system. This apparent modification may also occur in *Drosophila* given that the size of the *in vitro* translated product corresponds well with the larger protein in the ~97 kD doublet found in *Drosophila* extracts. The lower band of the 97 kD doublet may correspond to an alternatively modified protein. Two weakly detected specific products in the *in vitro* translated track present below the strongly detected band appear to correspond in size with two weakly detected

Figure 3.4.1 - Production and testing of anti-2.1 antisera

a - Coomassie stained gel showing induction of an ~100 kD GST-2.1 protein (lane 2). Lane 1 shows protein from an un-induced culture.

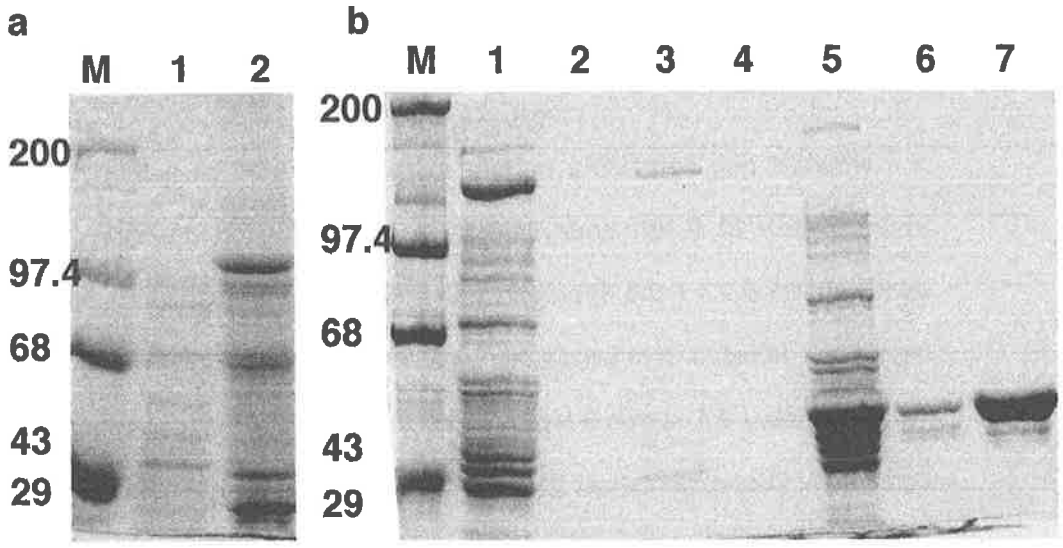
b - Coomassie stained gel showing the purification of MBP-2.1 (~140kD) and MBP alone bacterially expressed fusion proteins. Lanes 1 and 5 show a sample of the MBP-2.1 and MBP protein respectively loaded onto the column. Lanes 2 and 6 show the protein remaining on the column after elution. Lanes 3 and 7 show the pooled elution fractions 1-5. Lane 4 shows the pooled elution fractions 6-11 of the 2.1-MBP column.

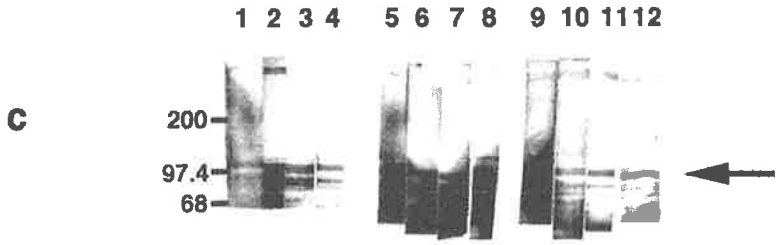
c - GST-2.1 on western blot strips probed with anti-2.1 sera. Bacterial extract containing approximately 50 ng of GST-2.1 fusion protein was loaded into each track. Tracks 1, 5 and 9 were probed with serum diluted 1:100 from rats 1, 2 and 3 respectively. Similarly, the next three strips in each set were probed with serum diluted 1:1000, 1:10000, 1:100000. Each rat's serum contained high levels of antibody activity against the GST-2.1 protein. The arrow indicates the presence of the GST-2.1 protein.

d - western blot containing MBP protein (lane 1) and MBP-2.1 protein (lane 2) stained with India ink to show the total protein on the blot.

e - the same blot as in **d** probed with affinity purified 2.1 antibodies, showing that the antibodies recognise the 2.1 moiety of the fusion protein.

f - a duplicate blot as in **d** and **e** probed with ^{35}S-labelled PCL. This shows that PCL interacts with the same protein that the antibodies recognise.


a
M 1 2
200
97.4
68
43
29
b
M 1 2 3 4 5 6 7
200
97.4
68
43
29


c
1 2 3 4 5 6 7 8 9 10 11 12
200
97.4
68

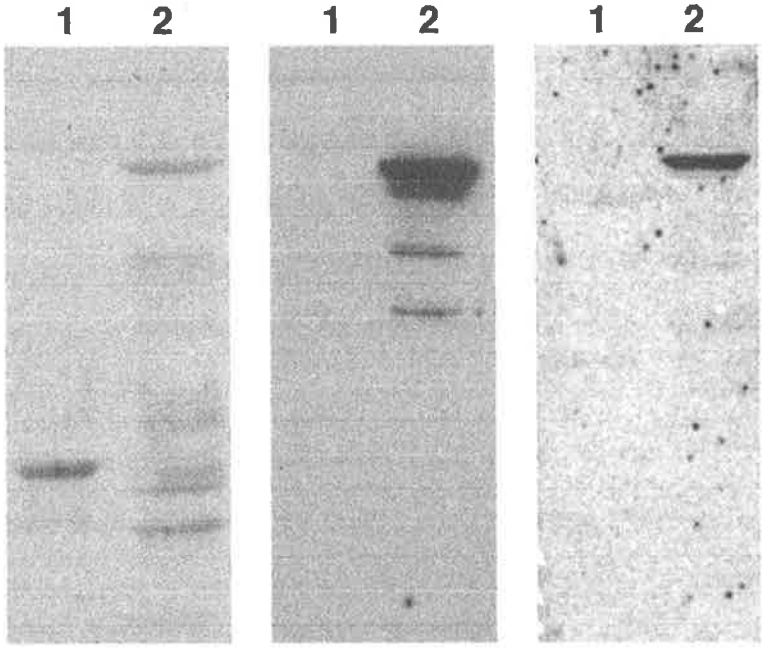

1 2
1 2
1 2
d
e
f

Figure 3.4.2 - Immunodetection of 2.1 protein *in vitro* translated and from *Drosophila* extracts

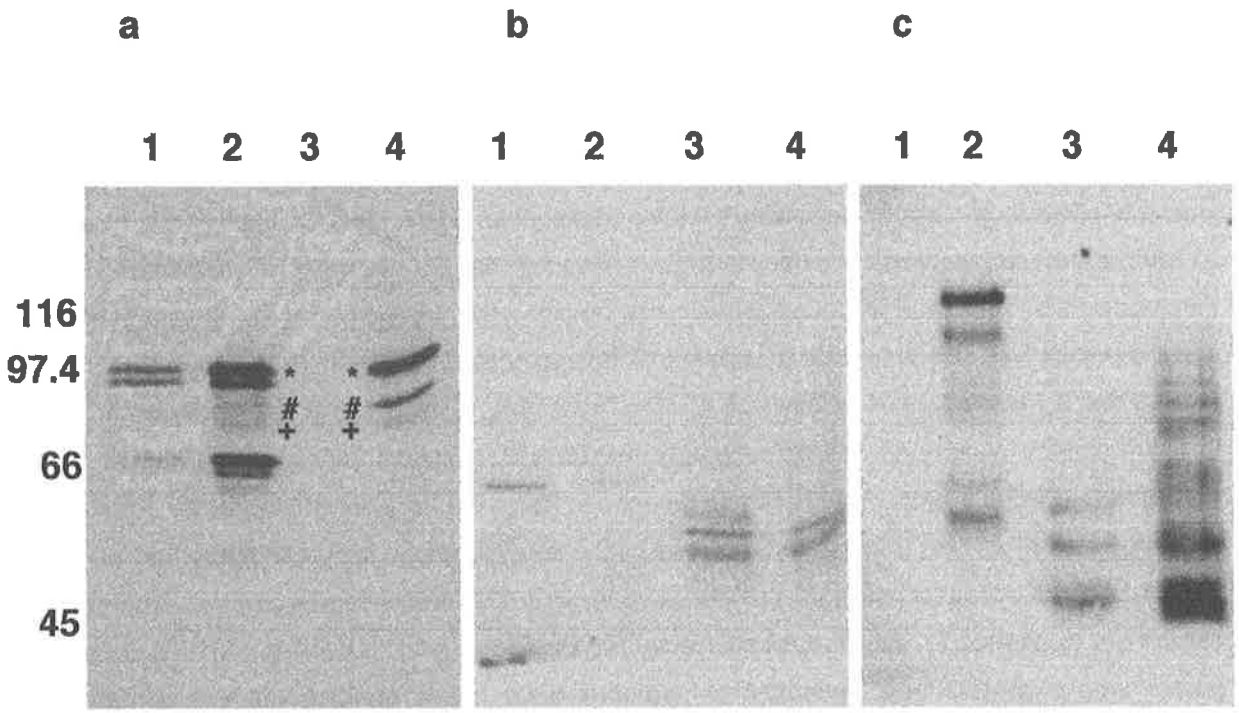
a - western blot containing protein extract from 1/10 of a third instar larva (lane 1), approximately 20, 0-18hr embryos (lane 2), unprogrammed *in vitro* translation (lane 3) and *in vitro* translated 2.1 (lane 4) probed with affinity purified anti-2.1 antibodies. At least 3 main bands are detected from Drosophila extracts (approximately 69 kD and a 97 kD doublet). *In vitro* translated 2.1 shows 3 bands of immunoreactivity that correspond in size with bands detected in *Drosophila* extract (*, #, +).

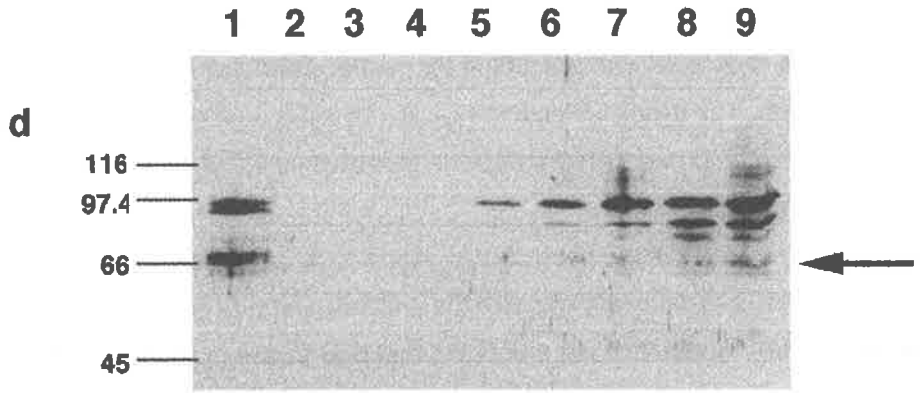
b - a duplicate blot of that in **a** probed with pre-immune serum. This blot was exposed to film for 60 seconds compared to that in **a** which was exposed for 5 seconds. This shows that the 2.1 antiserum specifically detects the bands seen in **a**

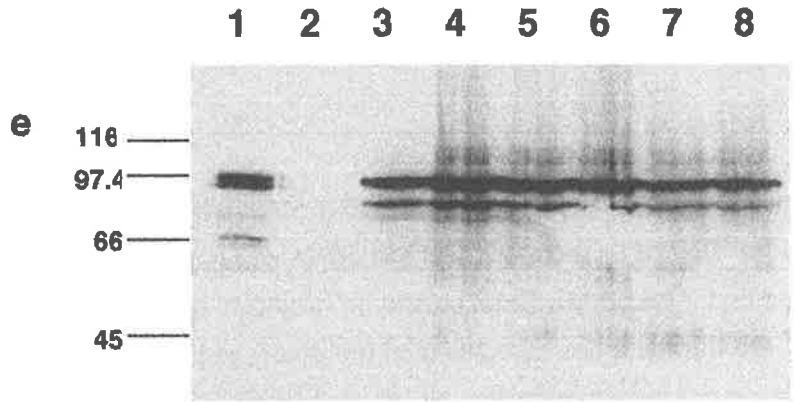
c - a triplicate blot probed with affinity purified anti-PCL antibodies to show further that the bands detected when probed with anti-2.1 antibodies, are specific. The bands seen in lanes 3 and 4 are non - specific as they are only seen with this serum.

d - western blot probed with anti-2.1 antibodies containing a time course of *in vitro* translation of 2.1. Lane 1 contains larval extract (as in **a**, **b** and **c**) as a size control. The other tracks contain un-programmed *in vitro* translation (lane 2) translation for 15, 30, 45, 60, 75 and 90 minutes (lanes 3-8). Lane 9 contains *in vitro* translation of 2.1 for 90 minutes followed by treatment with calf intestinal phosphatase for 30 minutes. In all cases the 97 kD band is the most obvious. The 69 kD band is detected weakly (arrow). Given that there is no difference between lane 8 and 9, the post-translational modification which occurs *in vitro* is probably not phosphorylation.

e - western blot probed with anti-2.1 antibodies containing larval extract (lane 1), un-programmed *in vitro* translation (lane 2) and 90 minute *in vitro* translations of 2.1 in the presence of 1, 2, 3, 4, 5, and 5.6 mM of benzamide (an ADP-ribosyl polymerase inhibitor) in lanes 3-8 respectively. Given that the predominantly detected band is at 97 kD, it appears that the modification that occurs to 2.1 *in vitro* is not poly-ADP ribosylation.


a
b
c
1 2 3 4
1 2 3 4
1 2 3 4
116
97.4
66
45
*
#
+


d
1 2 3 4 5 6 7 8 9
116
97.4
66
45


e
1 2 3 4 5 6 7 8
116
97.4
66
45

bands in *Drosophila* extracts. These products may represent alternatively modified 2.1 proteins or modification intermediates. The type of modification that was observed is of an unknown nature. The only post-translational modification that is known to occur to nuclear proteins (see chapter 6) that could account for the size difference observed between the predicted polypeptide and the observed *in vitro* translated peptide (~28 kD) is poly-ADP ribosylation (Naegeli *et al.*, 1989). This modification is known to occur to chromatin associated proteins (Baksi *et al.*, 1987; Durkacz *et al.*, 1992; Naegeli *et al.*, 1989) and was therefore considered a good candidate for the observed modification. However, this modification occurs in the nucleus, whereas reticulocytes, from whence the *in vitro* translation system was derived, are anucleate (K. Huston, Promega, personal communication), and so it seemed unlikely that 2.1 was poly-ADP ribosylated *in vitro*. To test this, *in vitro* translation of 2.1 was carried out in the presence of varying concentrations of benzamide, a known inhibitor of poly-ADP ribosyl polymerase (Durkacz *et al.*, 1992). As can be seen in figure 3.4.2e, benzamide has no effect on the observed size of *in vitro* translated 2.1 at any of the concentrations tested showing that poly-ADP ribosylation was not responsible for the modification observed. A time course *in vitro* translation was carried out to attempt to visualise the nascently translated 2.1 prior to it being modified. At all time points where protein was detected on western blots using anti-2.1 antibodies, the main electrophoretic species corresponded to 97 kD. This suggested that the modification was occurring rapidly after translation had occurred. Calf intestinal phosphatase treatment of the *in vitro* translated protein did not change its mobility indicating that the modification that occurs is not phosphorylation (figure 3.4.2d). There is at least one precedent for post-translational modification occurring in an *in vitro* translation system. The predicted size of PCL is ~95 kD (Lonie *et al.*, 1994) and that is the size that is observed when PCL is expressed in bacteria (see chapter 7). *In vitro* translated PCL is ~110 kD (Lonie, 1994; also see section 3.2) .

Regardless of the type of modification that occurs to 2.1, it is clear from the above experiments that 2.1 is modified post-translationally and that the antibody that was generated was able to detect all the forms of the protein. Whether or not any of the detected bands (e.g. the lower band of the 97 kD doublet) do not correspond to 2.1 protein can only be tested by performing western blots on extracts of *Drosophila* that are *2.1* null mutants. Unfortunately no such mutants, including any deficiencies that uncover *2.1*, were available at the time of this study.

3.5 - Discussion

The work in this chapter describes the isolation of a cDNA whose encoded protein interacts directly with PCL *in vitro*. Complete sequencing of the cDNA clone containing the complete open reading frame revealed that it encoded a novel protein (M. Coulson, personal communication). This finding, unfortunately, did not meet the expectation at the outset of the work, that PCL interactors would have known function or functional domains. However, it remained likely that the characterisation of the 2.1 protein in terms of its distribution during embryogenesis and within cells would shed light on the subject of Pc-G repression. To this end anti-2.1 antibodies were generated in rats. These antibodies were shown to be specific for the same polypeptide that was able to interact specifically

with PCL. The antibodies detected a 69 kD protein in *Drosophila* extracts. Given that the predicted size of the 2.1 polypeptide is 69 kD this result suggested that *in vivo* 2.1 was detected with the antibodies that were generated. A 97 kD doublet was also detected in the *Drosophila* extracts. Initially it was thought that this may be non-specific detection, however the observation that *in vitro* translated 2.1 is also 97 kD strongly suggests that the 97 kD proteins in *Drosophila* extracts also correspond to 2.1 and that 2.1 is post-translationally modified by an unknown mechanism. The data shown here indicate that the purified anti-2.1 antibodies were specific for multiple forms of the 2.1 protein *in vivo* and therefore immunohistochemistry on *Drosophila* tissues using these antibodies would specifically detect 2.1 protein. However formal and definitive proof that the antibodies are specific to 2.1 await the generation of specific *2.1* null mutants.

Chapter 4 - Confirming and testing for interactions between PCL, 2.1 and other Pc-G members

4.1 - Background

As discussed in chapter 1, the Pc-G proteins most likely form a multimeric protein complex. A rationale for screening for PCL interactors was to place PCL conceptually within such a complex, as a better knowledge of the structure of this complex would aid in the understanding of its function. Another rationale was to identify other non Pc-G proteins, which would aid in the elucidation of the repression mechanism of the Pc-G (see above). Given that no interactions with known Pc-G members were forthcoming from the initial screen, it was decided that assaying for interactions between PCL, 2.1 and Pc-G proteins should be undertaken. At this point the opportunity was kindly provided by Dr Hugh Brock to visit his laboratory, where a yeast two-hybrid 'matrix of interaction' was being constructed. Essentially all the cloned Pc-G members had been placed into plasmids which generated LEXA DNA binding domains fusions (EG clones) or activation domain fusions (JG clones) in yeast (Gyuris *et al.*, 1993). Every EG clone was tested against every JG clone to assay for specific interactions between Pc-G proteins.

4.2 - PCL, 2.1 and E(Z) all co-interact

Full length *Pcl* and the M96 conserved region of *Pcl* (see chapter 7) were subcloned into the EG plasmid. Full length *2.1* was subcloned into the EG and JG plasmids. The full length *Pcl* was generated by high fidelity PCR using Pfu polymerase, the cDNA clone AL15 being used as a template (Lonie, 1994) and the primers 05197 (overlapping the start codon) and 11787 (overlapping the termination codon) as forward and reverse primers respectively. These primers incorporated an NcoI restriction site at each end of the PCR product, so that the product could be cloned into the NcoI site of the plasmid pEG202 to produce clone pEG-*Pcl*. High fidelity PCR was also used to produce the M96 conserved region clone. Primers cDOM 5' and cDOM 3' were used as the forward and reverse primers respectively. These primers incorporated a BamHI site at each end of the PCR product which was cloned into the BamHI site of pEG202 to produce the clone pEG-cDOM. This clone was expected to express residues 403-605 of PCL. The full *2.1* clone (2.1 kb EcoRI insert from λ *2.1*) was subcloned into the EcoRI sites of pEG202 and JG4-5 to produce the clones pEG-*2.1* and JG-*2.1* respectively. Note that each of these clones was produced prior to the isolation of the 5' end of *2.1* (see section 3.3) and therefore contain an extra 9 nucleotides which would result in the addition of the residues E F R to the NH_2- terminus of 2.1. Each of these four clones was transformed into the yeast assay strain EGy48 (Gyuris *et al.*, 1993). Transformants were selected and then co-transformed with the entire panel of JG clones for each of the 3 EG clones and the entire panel of EG clones for the JG-*2.1* clone. These resultant yeast strains were then assayed for growth on leucine deficient media in the absence or presence of galactose which stimulates transcription from the JG plasmids, on the

basis that an interaction would 'marry' the LEXA DNA binding domain and the activation domain in the proximity of the LEU2 reporter gene, activating it's transcription and causing prototrophy for leucine. Table 4.2.1 shows all the co-transformations and the interactions detected (i.e growth on leucine deficient medium containing galactose as the sole carbon source).

Table 4.2.1 - yeast two-hybrid interaction matrix

test clones	pEG-*Pcl*	pEG-cDOM	pEG-*2.1*	JG-*2.1*
matrix clones				
(either as EG or JG clones)				
JG4-5	-	-	-	NA
pEG 202	NA	NA	NA	-
ph	-	-	-	-
phDN	-	-	-	-
phDS	-	-	-	-
Pc	-	-	-	-
Psc	-	-	-	-
PscDB	-	-	-	-
PscDS	-	-	-	-
PscDN	-	-	-	-
Su(z)2	-	-	-	-
Su(z)2DB	-	-	-	-
Su(z)2DNco	-	-	-	-
Su(z)2DNot	-	-	-	-
ph D	-	-	-	-
D chr Pc	-	-	-	-
chr Pc	-	-	-	-
Psc box	-	-	-	-
E(z)	++	++	++	++
Pcl	NA	NA	NA	-
cDOM	NA	NA	NA	-
2.1	-	-	-	-

- : no growth on leucine deficient medium

++ : growth on leucine deficient medium

NA : indicates an incompatible co-transformation i.e. two JG or two EG plasmids

Pc - Polycomb, ph - polyhomeotic, Psc - Posterior Sex Combs, *E(z)* - Enhancer of zeste, Su(z)2 - Suppressor of Zeste 2, chr - chromodomain, D(n) - refers to a deletion or truncation construct, ph D - is ph with the rae-28 homology region deleted, Psc box - the Psc/bmi-1 homology region.

Note: all of these test clones other than the *Pcl* and *2.1* clones were made by Michael Kyba (UBC)

As shown in Table 4.2.1, interactions were detected between PCL and E(Z), cDOM and E(Z) and 2.1 and E(Z). This latter interaction was observed when the reciprocal co-transformation was performed and thus acts as a verification for the interaction. That the former interaction was observed using full length PCL and a sub-region of PCL suggests that it may not be artifactual. Note also that all the interactions were dependent on galactose and hence expression of the JG clones and therefore probably represent interaction-dependent reporter gene activation. Interestingly, PCL and 2.1 do not appear to interact in this yeast two-hybrid assay. It is possible therefore that either the λ GT11 screen and subsequent far-western assay confirmation of the interaction was spurious, or that PCL and 2.1 do interact in yeast but that their interaction causes a repression of the LEU2 reporter gene that would normally have been activated.

4.3 - *In vitro* confirmation of the two-hybrid interactions

It is possible that interactions detected in yeast two-hybrid assays may be artifactual so it is desirable to obtain independent confirmation of the apparent interactions. It was decided therefore to utilise the far-western blot assay to confirm the PCL/E(Z) and 2.1/E(Z) interactions.

Figure 4.3.1a shows a western blot that contains GST, GST-PCL, GST-cDOM and GST-2.1 crude bacterial extracts in tracks 1-4 respectively. A duplicate blot was processed for far-western assays and then probed with *in vitro* transcribed/translated ^{35}S-methionine labelled E(Z). The E(Z) probe was produced from the clone of *E(z)* in pBS (pBS-KS-*E(z)*e32 provided by Rick Jones). Figure 4.3.1b (track 4) shows that a band corresponding to GST-2.1, but not GST, GST-PCL or GST-cDOM, was specifically bound by the E(Z) protein. Therefore, the 2.1/E(Z) interaction was confirmed by independent means whereas the PCL/E(Z) interaction was not. There are several ways in which the PCL/E(Z) negative result could be interpreted. Firstly, it is possible that the yeast two-hybrid interaction between PCL and E(Z) was artifactual. Secondly, it is possible that the interaction is too weak or transient to be detected by far western assay, compared to the exquisitely sensitive two-hybrid assay. Thirdly, either or both of PCL and E(Z) may require post-translational modifications prior to interacting. There is circumstantial evidence that PCL is post-translationally modified in *Drosophila* (Lonie, 1994). Figure 4.3.1c shows a western blot of total protein extract from yeast carrying the pEG-*Pcl* plasmid probed with anti-PCL antibodies. The expected size of a PCL-LEXA DNA binding domain fusion is approximately 119 kD (95 kD for PCL and 24 kD for the LEXA moiety). As can be seen in the figure, a doublet band (as has previously been noted for PCL in *Drosophila*, Lonie, 1994) of immunoreactivity for PCL of size approximately 140-150 kD is detected. This strongly suggested that PCL is also modified in yeast, although the relative change in mobility between predicted and observed proteins is different in *Drosophila* and yeast. (Drosophila: 95 kD predicted - 140 kD observed; yeast: 119 kD predicted 140-150 kD observed). Given that the yeast protein is a fusion with LEXA, thereby having a greater predicted size than the *Drosophila* protein, it is not possible to compare the post-translational modifications that occur to the yeast protein and the *Drosophila* protein.

To test the second possibility, that the far-western blot assay is not sensitive enough to detect the interaction, a co-immunoprecipitation method was employed. In this assay, two soluble proteins are mixed in a suitable buffered solution. Antibodies to each protein are then added and immobilised onto protein-G sepharose beads. After washing, the proteins bound to the beads are resolved using SDS-PAGE, and then the presence of each protein is assayed. As well as mixing ^{35}S-methionine labelled PCL and E(Z) proteins, unlabelled MBP-2.1 protein was added on the basis that it may form a 'bridge' between PCL and E(Z), given that it can interact independently with both of these proteins. Figure 4.3.1d shows the results of such an experiment. Track 1 contains labelled PCL and E(Z) proteins as a size standard. Tracks 2 and 4 shows that neither PCL nor E(Z) were immunoprecipitated when PCL pre-immune serum is used as an immunoprecipitation control. Tracks 3, 5 and 6 show that only PCL and not E(Z) was immunoprecipitated either without any accessory protein, with MBP-2.1 or with MBP alone added respectively. Therefore no interaction between PCL and E(Z), either direct or indirect, was detected in this assay.

4.4 - Refining the PCL and E(Z) interaction domains within 2.1

As was previously noted, 2.1 showed no common domains with any protein of known or unknown function. The mapping of the regions of 2.1 that interact with PCL and E(Z) would define functional regions of the protein. This knowledge would be useful in several ways. Firstly, more rigorous database searching could be performed using a small amino acid sequence to identify more distantly related domains. Secondly, these regions could eventually be specifically mutated, the resultant 2.1 cDNA be transformed into *Drosophila* and then assayed for modified function. This would partition the entire function of 2.1 into those individual functions that require PCL or E(Z) and aid in describing the mechanism of Pc-G repression as it relates to 2.1

In order to define the PCL and E(Z) interacting regions in 2.1 a far-western blot assay was once again employed. The clone pMAL-*2.1* was digested with the restriction enzymes PstI and SalI to 'drop-out' different portions of the 3' end of the *2.1* open reading frame (figure 4.4.1a). These religated clones were used to express COOH-terminal truncated MBP fusion proteins. Figure 4.4.1d shows a western blot reacted with anti-MBP antibodies (NEB) containing the truncated fusion proteins, the full length MBP-2.1 fusion protein and MBP protein alone. Identical blots were probed with either ^{35}S-PCL or ^{35}S-E(Z). Figures 4.4.1b and c show that both PCL and E(Z) respectively will only interact with the full length 2.1 (lane 3 in each blot) and not either of the truncations or MBP alone (as seen before). This indicated that the region/s of 2.1 required to interact with PCL and E(Z) lie towards the COOH-terminus from residue 315 (figure 4.4.1a). It is possible that either of these truncations disrupt the domain/s of interaction, or that 2.1 has a bipartite domain.

4.5 - 2.1 does not interact with the M96 conserved domain of PCL

In chapter 1 the possibility that the PHD fingers of PCL, that are contained within the region of similarity to M96, may be responsible for protein-protein interactions was discussed. The data

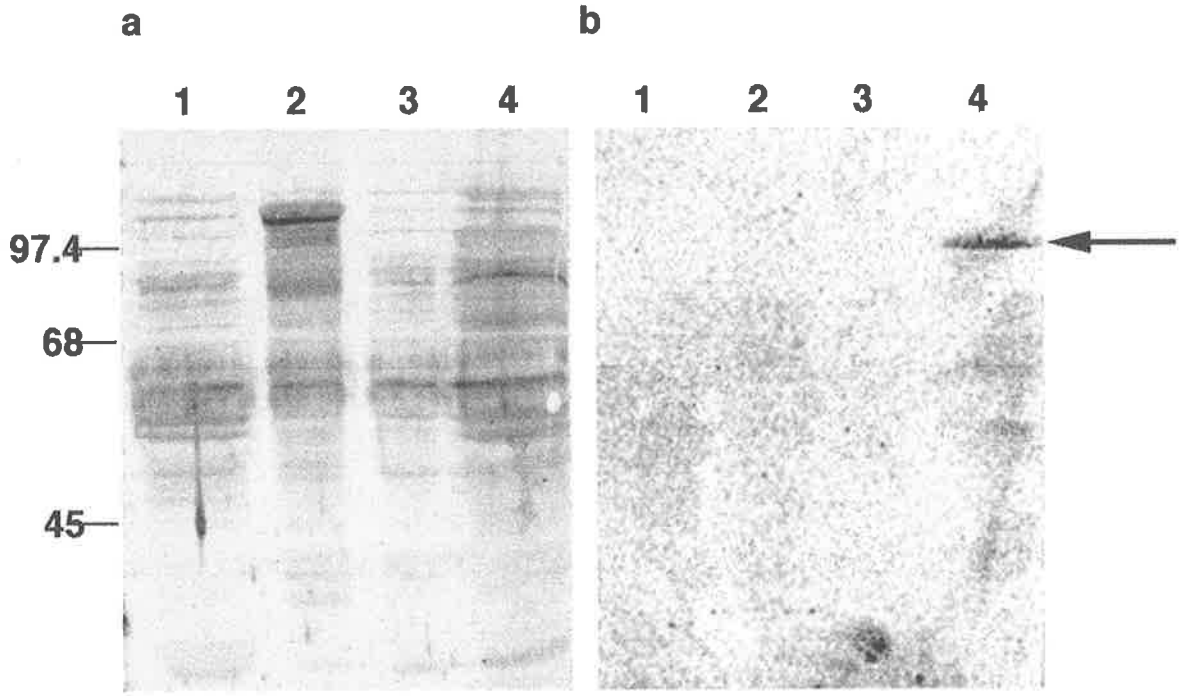
Figure 4.3.1 - *In vitro* Interactions between PCL, 2.1 and E(Z)

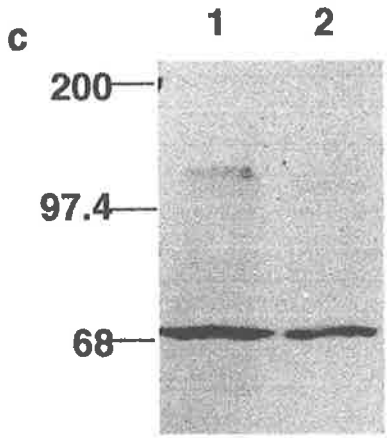
a - ink stained western blot containing bacterial extracts with GST (lane 1), GST-PCL (lane2), GST-cDOM (lane 3) and GST-2.1 (lane 4). Approximately equal amounts of protein were in each lane.

b - a duplicate blot of **a** probed, as a far-western, with ^{35}S-labelled E(Z). Only 2.1 appears to interact with E(Z) in this assay.

c - western blot of total protein extract from yeast containing the plasmid pEG-Pcl (lane 1) or the plasmid pEG with Pcl in the opposite orientation (lane 2) probed with anti-PCL antibodies. Note the doublet of approximately 140 kD in lane 1. As the band of ~ 69 kD was detected in all tracks including the marker track (not shown), it was deemed non-specific.

d - Co-immunoprecipitation of *in vitro* translated PCL and E(Z). Lane 1 contains radio - labelled PCL and E(Z) as a size control. Lane 2 - PCL and E(Z) mixed and precipitated with pre-immune serum. Lane 3 - PCL and E(Z) mixed and precipitated with anti-PCL. Lane 4 - PCL, E(Z) and MBP-2.1 mixed and precipitated with pre-immune serum. Lane 5 - as lane 4 except precipitated with anti-PCL. Lane 6 - as lane 5 except with MBP and not MBP-2.1. Lane 7 - as lane 5 except with no PCL added. E(Z) was never co-immunoprecipitated with anti-PCL antibodies.


a
b
1
2
3
4
1
2
3
4
97.4
68
45


c
1
2
200
97.4
68

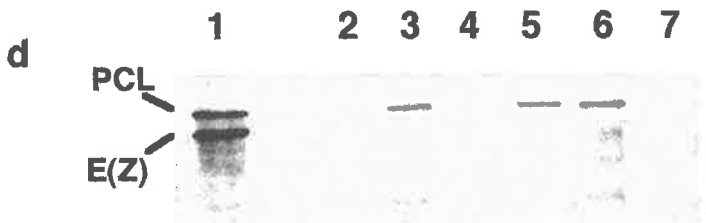

d
1
2
3
4
5
6
7
PCL
E(Z)

Figure 4.4.1 - Refining the interaction domains of 2.1

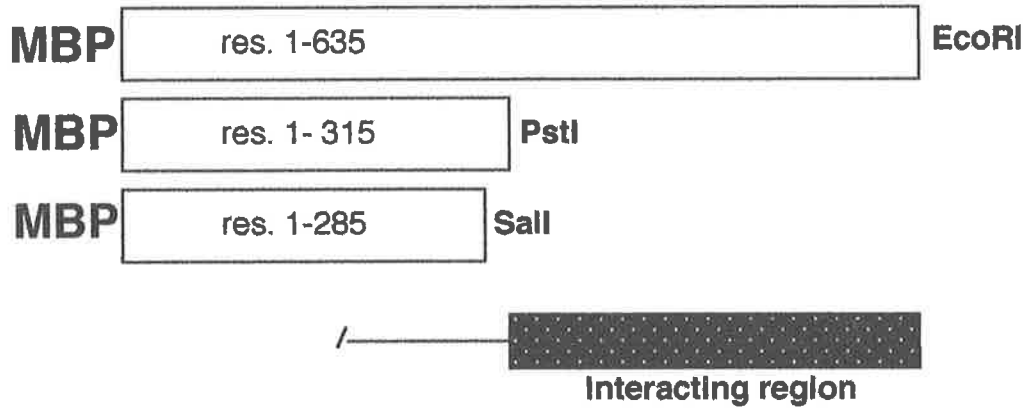
a - diagram showing a representation of the fusion proteins that would be produced from the pMAL-*2.1* deletion constructs that were generated.

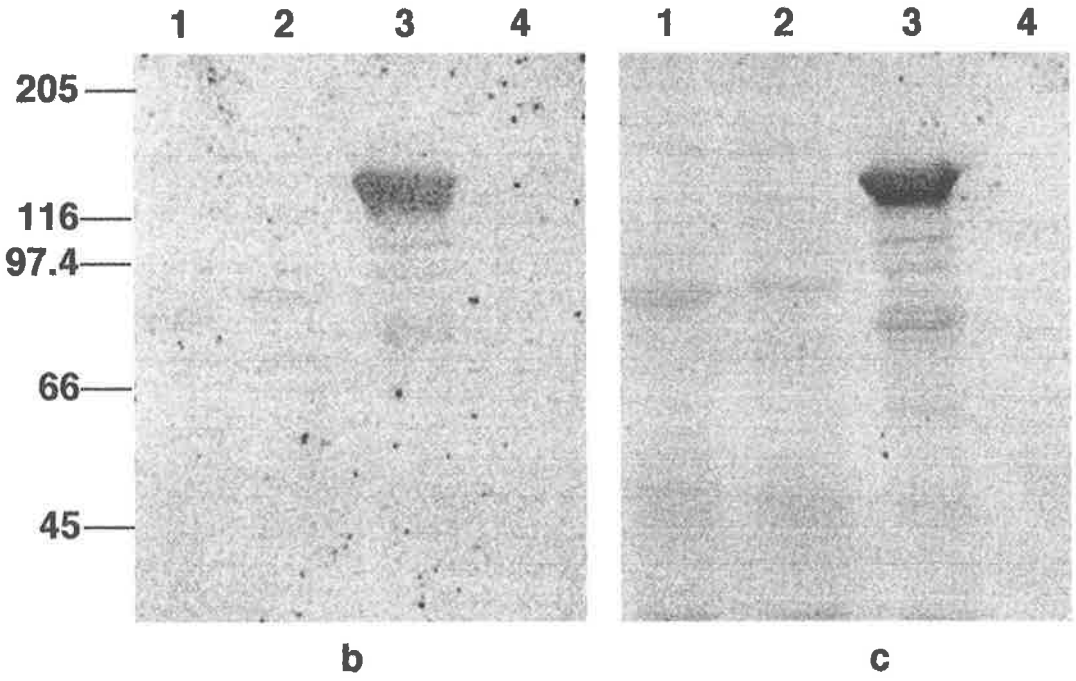
b - Western blot containing bacterial extracts with MBP-2.1Sal (lane 1), MBP-2.1Pst (lane 2), MBP-2.1 (lane 3) and MBP alone (lane 4) probed with ^{35}S-labelled PCL. Only the full length 2.1 interacts with PCL in this assay.

c - duplicate blot probed with ^{35}S-labelled E(Z). Only the full length 2.1 interacts with E(Z).

d - triplicate blot, probed with anti-MBP antibodies.

a


2.1MBP deletion constructs
MBP
res. 1-635
EcoRI
MBP
res. 1- 315
PstI
MBP
res. 1-285
SalI
Interacting region


1 2 3 4
205
116
97.4
66
45
1 2 3 4

b c

d

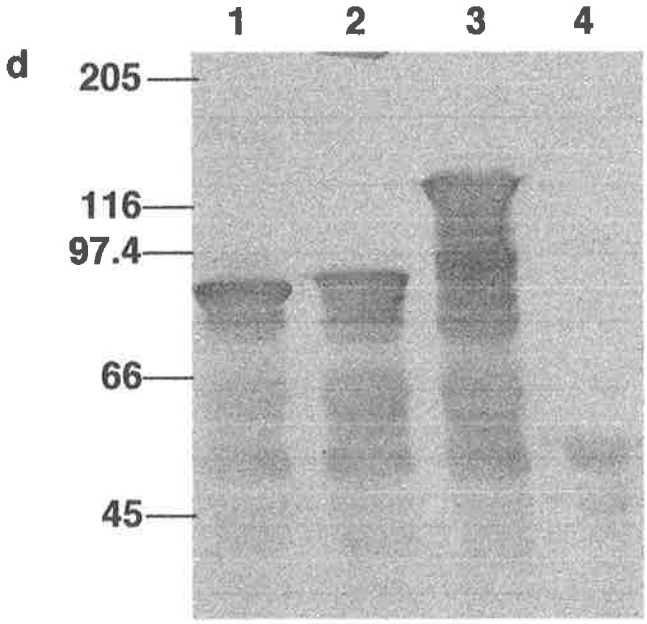
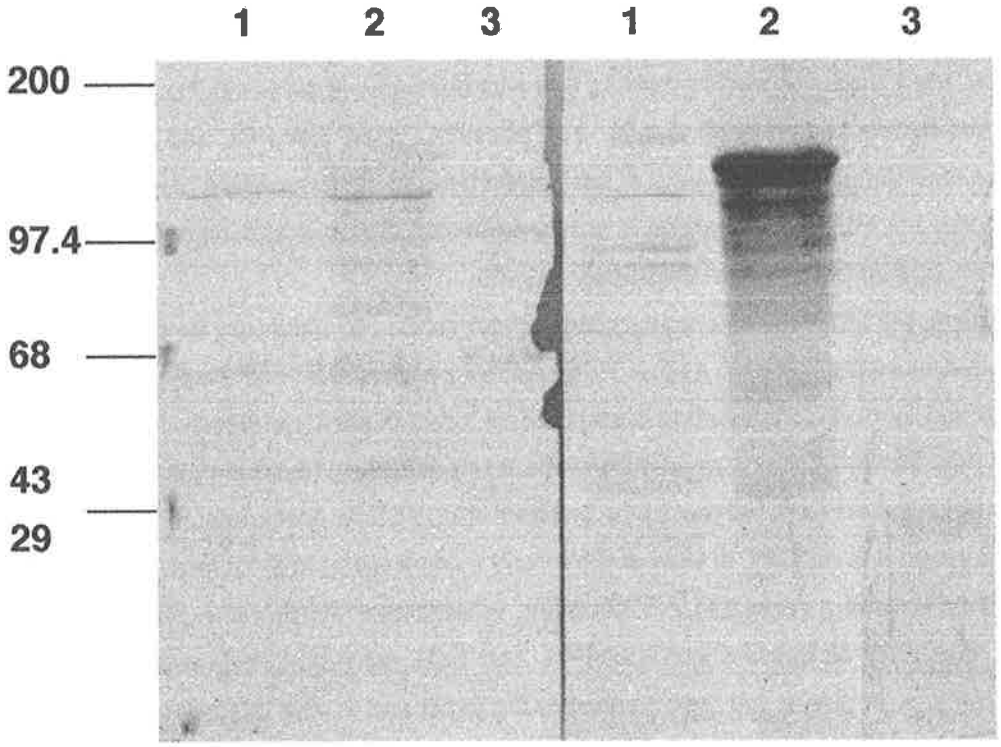

1 2 3 4
205
116
97.4
66
45

Figure 4.5.1 - 2.1 does not interact with the M96 similarity region of PCL

a - far western blot containing bacterial extract with GST (lane 1), GST-PCL (lane 2) and GST-cDOM (lane 3) probed with bacterial extract containing MBP protein alone, followed by probing with anti-MBP antibodies.

b - a duplicate blot probed with bacterial extract containing MBP-2.1 protein followed by probing with anti-2.1 antibodies. This shows that 2.1 specifically interacts with PCL but not with the M96 similarity region of PCL (cDOM).


1
2
3
1
2
3
200
97.4
68
43
29
a
b

presented in section 4.2 indicated that indeed this region of PCL is responsible for the interaction observed with E(Z). To test whether 2.1 interacts with PCL in this same region, a far western blot assay was employed, given that PCL and 2.1 did not appear to interact with each other in the yeast two-hybrid assay. Identical western blots containing bacterial extract with GST alone, GST-PCL or GST-cDOM were prepared for far western assay. One blot was probed with partially purified MBP-2.1 and a control blot was probed with MBP alone (see chapter 3), prior to being probed with anti-MBP antibodies (NEB). The results of this experiment (figure 4.5.1) confirm once again that 2.1 can interact with PCL, this time in the converse way (i.e PCL immobilised on the blot) and also shows that 2.1 does not interact solely with the cDOM portion of PCL.

4.6 - Discussion

PCL and 2.1 were each shown to interact with E(Z) independently in the yeast two-hybrid assay. However, the PCL/2.1 *in vitro* interaction that was described in section 3.2 could not be verified by the yeast two-hybrid assay (section 4.2). It is possible, given that PCL and 2.1 appear to be modified *in vivo*, that the unmodified forms of the protein used in the far-western blot analyses interact spuriously, i.e. that the PCL/2.1 interaction is purely an *in vitro* artifact that does not reflect the *in vivo* (in yeast and *Drosophila*) behaviour of these two proteins.

Conversely, the PCL/E(Z) interaction could not be verified by either the far-western blot assay or an *in vitro* co-immunoprecipitation assay. This is either because the yeast two-hybrid result in this case was artifactual, or that the interaction is far too weak or transient to be detected by the two *in vitro* means, or that PCL and/or E(Z) require post-translational modifications to interact. The last possibility is raised by the observation from western blot analysis that PCL is post-translationally modified in *Drosophila* and yeast. It is unlikely, in view of the strong evidence that E(Z) interacts with 2.1, that the interaction with PCL, itself a candidate 2.1 interactor, is artifactual. Additionally, the partial evidence that PCL, 2.1 and E(Z) co-interact and that PCL and E(Z) are functionally related but have no sequence similarity, strengthens the hypothesis that the yeast and *in vitro* interactions between these three proteins represents *in vivo* interactions. In support of this, E(Z) did not interact with any of the other Pc-G members in the two-hybrid 'interaction matrix' (M. Kyba, personal communication), indicating that E(Z) was not prone to spurious interactions. As further evidence that PCL and E(Z) can and do interact *in vivo*, *Pcl* mutants show a stronger genetic interaction with *E(z)* mutants than any other Pc-G mutants (Campbell *et al.*, 1995). In conclusion, it seems more likely that the PCL/E(Z) interaction is real but required a sensitive assay system, or post-translational modifications which were not present in the *in vitro* assay system used, to detect their interaction.

Although the compilation of far-western blot and yeast two-hybrid interaction data strongly suggest that the interactions between PCL, 2.1 and E(Z) are biologically significant, in order to verify that the PCL/2.1, PCL/E(Z) and 2.1/E(Z) interactions occur *in vivo*, co-immunoprecipitation assays would need to be employed using extracts from *Drosophila*. These kinds of experiments are currently being undertaken.

That MBP-2.1 could not bridge the interaction between PCL and E(Z), even though 2.1 can interact with both PCL and E(Z) *in vitro*, may only point to the possibility that the complex of three proteins was too unstable to be co-immunoprecipitated, that 2.1 can only interact with either PCL or E(Z) (i.e. PCL and E(Z) compete for interaction with 2.1) or that there was a technical problem with the co-immunoprecipitation, such as antibody interference with complex stability.

Data which showed requirement for amino acids 315-650 of 2.1 for the PCL and E(Z) interaction was presented. At present there is no evidence to suggest that PCL and E(Z) interact with the same domain within 2.1 (see below). Furthermore it is possible that the domain or domains are bipartite and are only formed in the native protein, in which case only part of the domain/s may reside towards the NH_2-terminus from residue 315. To test this the region from the PstI site to the 3' end of the ORF of 2.1 would need to be used to produce protein as a target for PCL and E(Z) interaction. An interaction with this portion of the protein would suggest that the entire interaction domain is present. Conversely no interaction would suggest that one of the two possibilities listed above was true.

From the data presented in the last two chapters a crude model for the interaction of these three proteins, PCL, 2.1 and E(Z), can be postulated. In section 4.5 it was shown that 2.1 did not interact with PCL in its M96 conserved region. This is in contrast to the yeast two-hybrid data that showed that E(Z) interacted with PCL within the M96 conserved region. None of these three genes share any sequence similarity that might indicate the presence of a common interaction domain. If this is the case, then there appears to be no reason why the three proteins could not interact at the same time.

Chapter 5 - In search of a *2.1* mutant: *l(3)SG23* is a Pc-G mutant but does not correspond to 2.1

5.1 - Background

The work presented in the last two chapters, in particular the evidence that the newly isolated 2.1 interacts with PCL and E(Z), both known members of the Pc-G, strongly suggested the possibility that 2.1 is a novel member of the Pc-G. This hypothesis could be tested partially at a molecular level, for example, by analysing the distribution of 2.1 protein on polytene chromosomes in contrast to known Pc-G members (see chapter 6). However, the most definitive way in which to characterise the role of the 2.1 protein would be to generate *2.1* mutant alleles and characterise the 2.1 mutant phenotype. Pc-G genes are defined by two criteria. Firstly, a Pc-G mutation must cause gain of function homeotic transformations. Secondly, these mutations must interact synergistically with other Pc-G mutations to produce more extreme phenotypes (see chapter 1). Once a *2.1* mutation was generated it would be relatively simple to construct these two genetic assays. At a more detailed level, the possibility that homeotic genes are ectopically expressed in homozygous *2.1* mutants could be tested directly by immunohistochemical staining of such mutants.

If *2.1* could not be classified as a member of the Pc-G by these assays, the existence of a *2.1* mutant would be even more valuable. As was described in chapter 1, a protein that interacts with PCL may be expected to fall within a few functional classes. As examples, 2.1 may be a component of heterochromatin, be involved in long distance chromosomal interactions or be involved in restricting the spread of Pc-G components. Each of these hypotheses could be tested genetically. *2.1* mutants could be assayed for modification of PEV, for modification of the zeste-white effect and transvection or for modification of enhancer blocking.

Establishing the importance of the availability of *2.1* mutants is trivial compared to generating and identifying them. Fortunately the *Drosophila* Genome Project now provides valuable resources for mutant generation. For example, as part of this project a large scale P-element insertional mutagenesis is being undertaken. Once the cytological location of a cloned gene is determined, P-elements within that vicinity, it they exist, can be obtained and analysed. In some cases, a P-element may occur within the gene of interest. If one is not so fortunate, a nearby P-element can be mobilised to generate insertional mutations within the gene of interest.

5.2 - Cytological location of *2.1*

The clone pBS-*2.1* Eco was used to make a digoxygenin labelled probe to map the position of the *2.1* gene by *in situ* hybridisation to polytene chromosomes. Analysis of many chromosomal spreads showed that only one signal was detected and that this was always at the same position. Figure 5.2.1a shows a segment from a polytene chromosome showing the hybridisation signal which was mapped to the interval 70B. None of the existing mutants within this region, except those

discussed below, (Lindsley and Zimm, 1992) had the characteristics that suggested they were good candidates for *2.1* mutants.

5.3 - Candidate *2.1* EMS mutants

As reported in Lindsley and Zimm (1992), several groups generated EMS-induced mutant lines that are late third instar larval lethal and have imaginal disc phenotypes. Four of these lethals, *l(3)SG22* (Shearn and Garen, 1974), *l(3)SG23*, *l(3)SG24* and *l(3)SG25* (Shearn *et al.*, 1971) mapped at the genetic positions 3-37.8, 3-39.4, 3-41 and 3-41.4 respectively (Lindsley and Zimm, 1992). Given that some of these mutants (in particular *l(3)SG23*) were reported to have abnormally differentiated imaginal discs and their genetic map position corresponded roughly with the cytological location of *2.1* (70B), they were obtained for further analysis as putative *2.1* mutants. Flies heterozygous mutant for *l(3)SG22*, *23*, *24* and *25* were crossed to flies heterozygous for *Pcl* alleles T1, W6 or 1. Progeny from these crosses that were heterozygous for both *l(3)SG23* and *Pcl* showed a moderate extra sex combs phenotype (see figure 5.4.1). This result corresponded well with the fact that *l(3)SG23* had the strongest abnormal disc differentiation phenotype of the 4 lethals tested which was suggestive of abnormal homeotic gene expression in this line. *l(3)SG23* was then chosen for further analysis as a strong candidate for a *2.1* mutant.

At the same time Landecker *et al.*, (1994) reported the identification of 4 cytological regions that contained novel Pc-G genes. One of these locations was defined by the overlap of the two deficiencies *Df(3L)fzGF3b* (70B;70C6) and *Df(3L)fzCAL5* (70C2-7;70E1-3). That is, the novel Pc-G gene was within 70C2-7. Based on the possibility that the cytological mapping of *2.1* or the deficiency strains was inaccurate, these deficiency strains were obtained to assess whether or not *2.1* was uncovered by them. As can be seen in the southern blot of genomic DNA prepared from these strains and wild type flies, *2.1* is not removed by either deficiency (figure 5.2.1b and c). These data are not surprising in the light of more recent information that two other groups independently mapped the left hand breakpoint of *Df(3L)fzGF3b* to 70C2 and not 70B (Lindsley and Zimm, 1992). From this it is clear that *2.1* is not the same as the novel Pc-G gene identified by Landecker *et al.*, (1994). Formally it is possible that *Df(3L)fzGF3b* removes part of the *2.1* gene that is not within the probe fragment used for the Southern blot analyses. If this were the case the Pc-G enhancer identified by Landecker *et al.*, (1994) may correspond to *2.1*. Complementation analyses between *2.1* mutants and the abovementioned deficiencies would need to be carried out to test this.

5.4 - *l(3)SG23* enhances *Pcl* and *Pc* mutant phenotypes but is not a 2.1 mutant

As described above, 4 EMS induced mutant lines that mapped genetically in the approximate location of the *2.1* gene, *l(3)SG22*, *23*, *24* and *25* , were crossed to lines bearing *Pcl* mutations and the resultant progeny analysed for Pc-G type homeotic transformations. It was discovered that

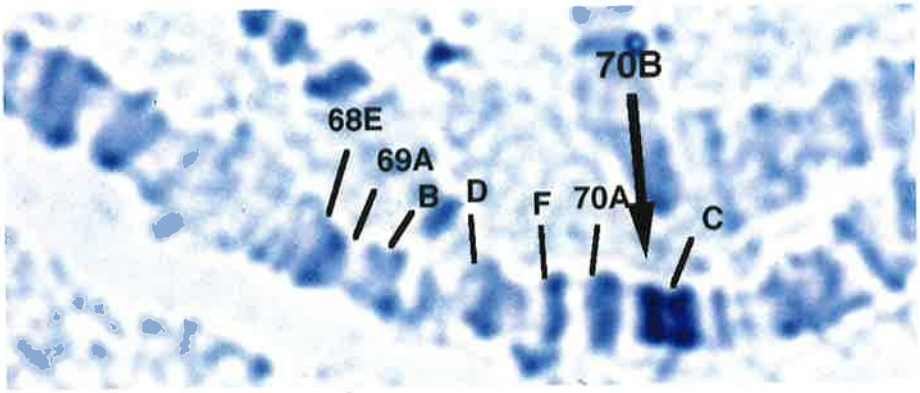
Figure 5.2.1 - Cytological localisation of the *2.1* gene

a - segment of polytene chromosome 3L probed with a *2.1* cDNA probe and counter-stained with giemsa. The large puff acted as a distinctive reference marker. The intervals have been numbered. The signal generated by the *2.1* probe is arrowed and is within 70B.

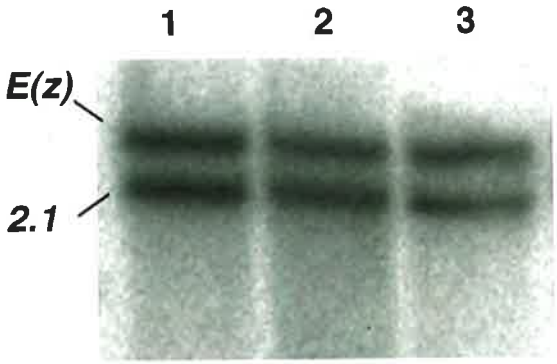
b - Southern blot of Cla I digested genomic DNA prepared from wild type flies (lane 3) or flies heterozygous for *Df(3L)fzGF3b*, the deficiency that uncovered the Pc-G enhancer of interest (lane 1) or *Df(3L)GS1a*, a nearby unrelated deficiency as a control (lane 2). The blot was separately probed with a *2.1* cDNA probe, and an *E(z)* cDNA probe as a control (*E(z)* is at a distant part of the genome (67E3-4)).

c - the ratio of arbitrary densitometric units found in a rectangle encompassing the signal from the *2.1* probe and the same sized rectangle encompassing the signal from the *E(z)* probe in each of the three tracks. This shows that there is an approximately equal amount of *2.1* DNA compared to *E(z)* DNA in all three fly strains, indicating that *Df(3L)fzGF3b* does not uncover the *2.1* gene.

a



b



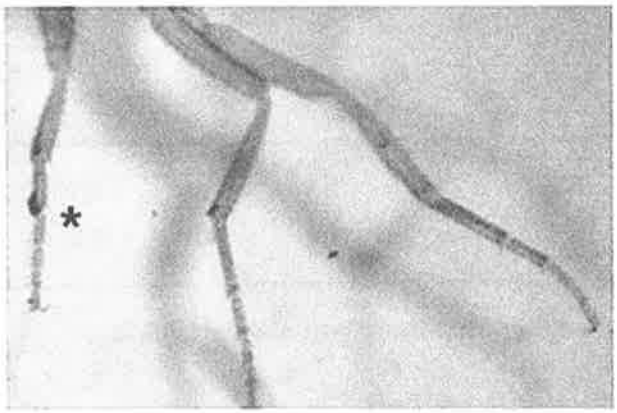
c

1.06 0.86 0.87

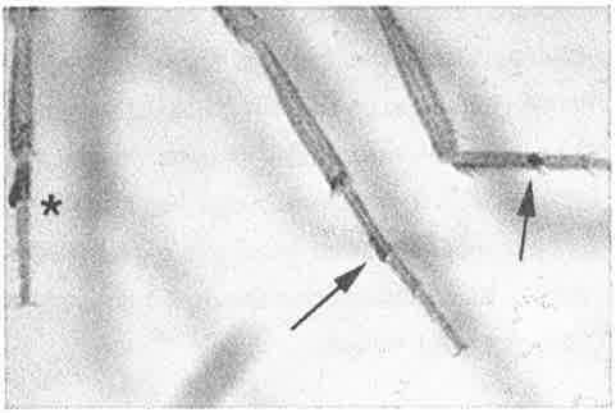
Figure 5.4.1 - Enhanced extra sex combs phenotypes in *l(3)SG23;Pcl* and *l(3)SG23/Pc* trans-heterozygote male flies

Legs from longitudinally dissected thoraces of adult male flies, with anterior (first leg) to the left and posterior (third leg) to the right. **a** *CyO;TMI* (*l(3)SG23*$^{+}$, *Pc*$^{+}$, *Pcl*$^{+}$), **b** *l(3)SG23/Pc*^{1} trans-heterozygote, **c** *l(3)SG23;Pcl*^{T1} trans-heterozygote. Normal sex combs on first legs are indicated with asterices. The additional sex combs on second and third legs, never complete ones, are indicated by arrows.

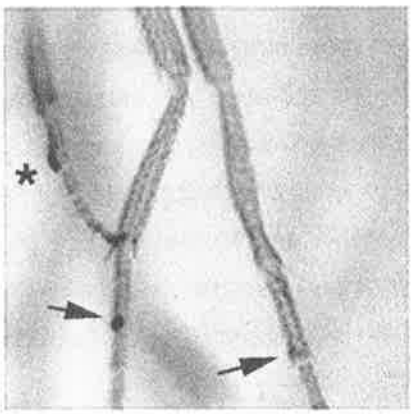
a



b



c



l(3)SG23 was able to enhance dominantly the extra sex combs phenotype of flies heterozygous for *Pcl* or *Pc* mutations (table and figure 5.4.1).

Table 5.4.1

Numbers of legs with sex comb bristles in trans-heterozygotes

	CyO	Pcl^{T1}	Pcl^{W6}	Pcl^{1}	TMI	Pc^{1}
TMI	NS	2.2	2.2	2.1	NS	2.0*
SG23	2.0	4.8	4.6	3.9	2.0*	5.2

*** - these classes were phenotypically indistinguishable**

NS - not scored

note - at least 30 male flies of each class were scored

Flies homozygous for the *l(3)SG23* mutation die at the late third instar/pupal transition stage probably as a result of inappropriate differentiation of their imaginal discs (Lindsley and Zimm, 1992). These data are suggestive of ectopic homeotic gene expression in *l(3)SG23* animals, and satisfy the criteria for defining *l(3)SG23* as a Pc-G mutant. Complementation analysis showed that *l(3)SG23* was completely lethal over the deficiency *Df(3L)fzGF3b* (not shown), suggesting that *l(3)SG23* could correspond to a specific mutation of the Pc-G enhancer identified within this region (70C2-D5) (Lindsley and Zimm, 1992) by Landecker *et al.*, (1994). As was mentioned above, *2.1* does not lie within *Df(3L)fzGF3b*. By deduction therefore, *l(3)SG23* is not a *2.1* mutant.

5.5 - A P-element insertional mutation of *2.1*

A 'Cyto-search' of the flybase database (Gelbart *et al.*, ongoing) identified two P-element mutations in the 70B interval. Both of these strains *l(3)00543* (cytological location 70B1-3) and *l(3)03699* (cytological location 70B4-5) were obtained from the Bloomington stock centre for further analysis. Plasmid rescue of the P-elements and flanking genomic DNA followed by probing with this DNA on Southern blots containing *2.1* genomic DNA indicated that the P-element associated with the mutation *l(3)00543* was inserted within an intron of the *2.1* gene (M. Coulson, personal communication). Phenotypic analysis of this insertional mutant as well as the generation of small deletions by imprecise P-element excision is being carried out (M. Coulson, personal communication).

5.6 - Discussion

The *2.1* gene was cloned by "reverse genetics" and therefore no genetic mutant was available initially. For a complete characterisation of *2.1* function, *2.1* mutants needed to be generated and analysed. To this end the cytological location of *2.1*, 70B, was determined by *in situ* hybridisation to

polytene chromosomes. This information led to the identification of *l(3)SG23* as a putative *2.1* mutant. However, *l(3)SG23* turned out not to be a *2.1* mutant, but a novel Pc-G mutant. *l(3)00543* was identified as a P-element insertional mutation within *2.1* and therefore most likely a *2.1* mutant. The large amount of work to characterise this mutation as well as the generation of other mutations is underway.

Chapter 6 - Analysis of the distribution of 2.1

6.1 - Background

As described in chapter 3, sequence analysis of *2.1* revealed that it encodes a novel protein (M. Coulson, personal communication). It interacts *in vitro* and in yeast two-hybrid assays with two Pc-G members, PCL and E(Z). Many questions are posed about the role of 2.1. Is it a member of the Pc-G? Is it a protein involved in chromosome structure or some nuclear function which requires or is required by the Pc-G? Analysis of the distribution of 2.1 within a cell may help to address these issues. Firstly, 2.1 would be expected to be localised to the nucleus of a cell. Resolution of the sub-nuclear distribution of 2.1 may help to answer some of the specific questions about its role. If 2.1 was localised to the same polytene chromosomal sites as PCL, PC and PH for example, it would be reasonable to establish the working hypothesis that 2.1 is a Pc-G member. If 2.1 was distributed in a different way chromosomally, analysis of this distribution may cast some light on the function of 2.1 and, given the interaction with PCL and E(Z), on the mechanism of Pc-G mediated repression. To this end antibodies against 2.1 were raised (see section 3.4).

6.2 - 2.1 is localised to the nucleus and ubiquitous during embryogenesis

As is evident from the data presented in figure 6.2.1, 2.1 is distributed ubiquitously in embryos throughout embryogenesis. Additionally, there appear to be higher levels of the 2.1 protein in the mesoderm of germ band extended embryos (stage 9) and in a series of segmentally reiterated, single cell width stripes in later (stage 13) embryos. 2.1 is localised to the nucleus, as was expected of a protein that interacts with PCL and E(Z). The 'grainy' nuclear localisation, indicative of chromosomal localisation (Kuzin *et al.*, 1994), was observed at all stages of development that were observed.

6.3 - The 2.1 stripes are posterior to the ENGRAILED stripes

The position of the 2.1 stripes relative to the stripes of expression of a key segment polarity gene, *engrailed* (*en*), was investigated. Embryos were subjected to immunotreatment as before using both the rat anti-2.1 and the mouse monoclonal 4D9 (anti-EN) antibodies (Patel *et al.*, 1989). This time the location of these antibodies and therefore the 2.1 and EN antigens was visualised by the use of rhodamine conjugated anti-rat and FITC conjugated anti-mouse secondary antibodies respectively, followed by confocal microscope imaging. Negative control immunofluorescences showed that neither rat serum nor the two secondary antibodies used produced any significant background fluorescence (figure 6.3.1a,b). Figures 6.3.1c-e show that the 2.1 stripes are directly posterior to the EN stripes in each parasegment. A potential role for 2.1 in the generation of segment polarity or segmental structures was not investigated further.

6.4 - 2.1 is a chromosomal protein

In order to assess whether or not 2.1 is likely to be stereotypical Pc-G member immunostaining of polytene chromosomes was carried out. This required that 2.1 be present in larval salivary gland polytene nuclei, the source of polytene chromosomes. Salivary glands were dissected from larvae, fixed and immunostained with anti-2.1 antibodies. 2.1 was present in salivary gland nuclei as visualised by a punctate nuclear staining characteristic of chromosomal localisation (not shown).

To examine more precisely the chromosomal distribution of 2.1, polytene squashes and immunofluorescence were carried out as described in section 2.2.34. The concentration of antibody for use in these assays was determined empirically, but was close to ten times the concentration required for whole mount immuno-staining of embryos. As is evident from figures 6.4.1.1-2, and 6.4.2-3, 2.1 is not distributed in the same way as stereotypical Pc-G proteins. It is quite clear from the double fluorescence detection of DNA and 2.1 that 2.1 is present predominantly in the interband regions throughout the chromosomes (figure 6.4.1.1f and 6.4.2a). Strikingly, there is a very strong 2.1 containing region within the chromocentre (figure 6.4.1.1c and 6.4.1.2a). That the level of 2.1 protein here is higher than elsewhere can be visualised in figure 6.5.1 where the intensity of the 2.1 signal was reduced until none of the signal from the chromosome arms was detectable. The signal in the chromocentre was still visible at this level. It must be noted that we have no good idea about the sensitivity of 2.1 detection in these assays. It is formally possible that 2.1 is present at other locations on chromosomes at lower, undetectable levels.

6.5 - 2.1 does not appear to colocalise with PCL on polytene chromosomes

Double immuno-fluorescence of polytene chromosomes using anti-2.1 and anti-PCL antibodies revealed that there is little colocalisation of 2.1 with PCL (figure 6.4.1.2g and 6.4.2b) and, given the total number of sites, the small amount of overlap may be due to chance alone. It must be noted that not all of the PCL sites could be visualised in these experiments. This was probably due to the optimisation of the fixation conditions for the 2.1 immunofluorescence. However, the approximately seventy PCL sites that were present in the best spread obtained (figure 6.4.1.2) appeared to correspond with sites that were previously mapped for PCL/PC/PH (DeCamillis *et al.*, 1992; Lonie *et al.*, 1994). It is striking that almost all of the visualised PCL sites overlap completely with polytene bands (as visualised by the light blue staining - see figure 6.4.2c). This is not surprising given that the bands are considered to be regions of transcriptionally silent chromatin (Rykowski *et al.*, 1988). This is in stark contrast to the observation that 2.1 is present predominantly in interbands. That this is the case was validated by the observation made above that there appears to be almost no obvious overlap between PCL and 2.1 sites. This observation was difficult to explain given the *in vitro* interaction between PCL and 2.1. It was only when the 2.1 protein distribution was considered in isolation that a possible explanation was forthcoming. As discussed below, the distribution of 2.1 (i.e.

Figure 6.2.1 - 2.1 immunostaining of whole mount *Drosophila* embryos

Lateral views of embryos with anterior to the left and dorsal up unless stated otherwise

a - an early blastoderm (stage 5) embryo showing ubiquitous nuclear staining

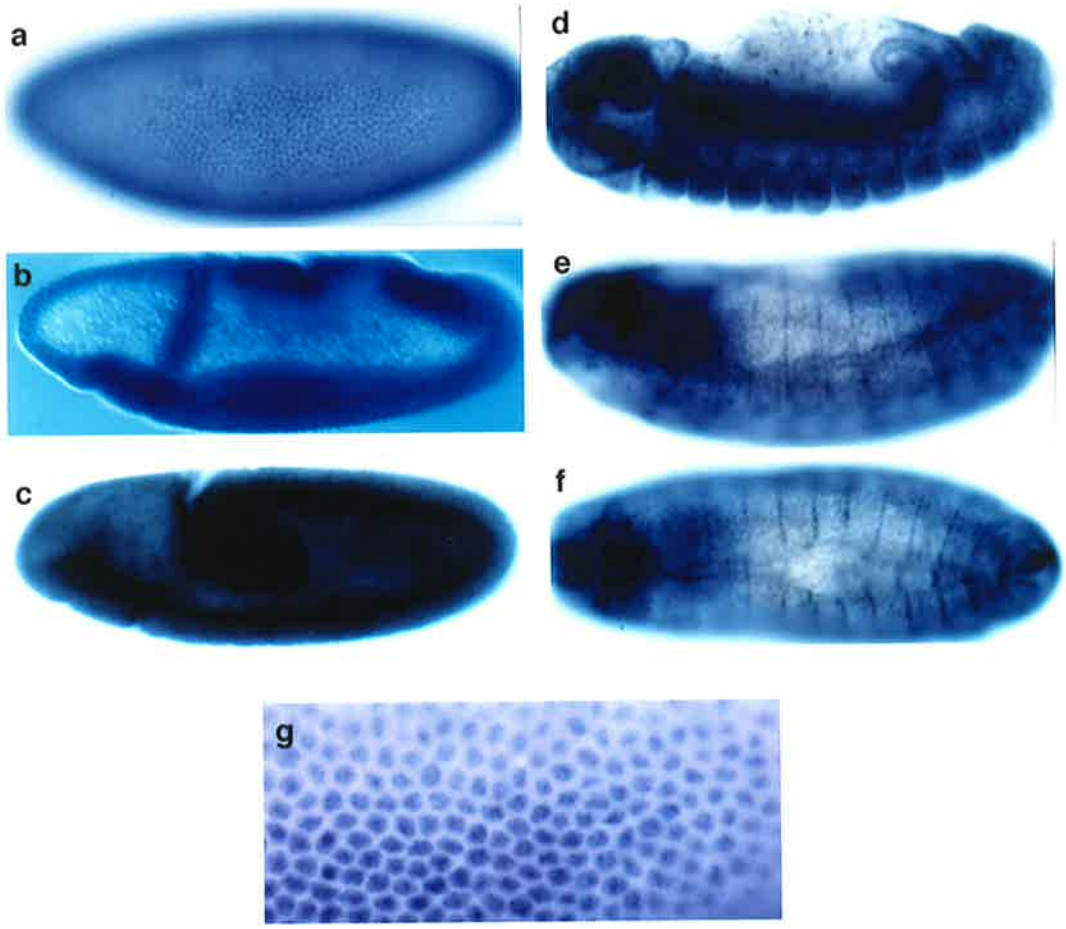
b - an early germ band extended (stage 8) embryo showing ubiquitous staining

c - a fully germ band extended (stage 9) embryo showing ubiquitous staining as well as a higher level of staining in the mesoderm.

d - an early germ band retracted (late stage 12) embryo showing ubiquitous nuclear staining .

e and **f** - late stage 13 embryo showing ubiquitous 2.1 distribution as well as a higher level of staining in a series of segmentally re-iterated, single nucleus width stripes. **f** shows a dorsal view of the same embryo in **e** showing that the stripes extend to the dorsal edge of the epidermis (note that dorsal closure has not completed)

g - high magnification of the embryo in **a** showing the nuclear localisation of 2.1. The granular staining is indicative of chromosomal localisation of the protein


a
d
b
e
c
f
g

widespread predominantly interband localisation) is similar to that reported for two boundary element associated proteins, BEAF-32 (Zhao *et al.*, 1995) and BANGDOO (Gaszner *et al.*, 1996). Bearing in mind the hypothesised necessity for a mechanism to limit the spreading of Pc-G repression (see section 1.5.10), it was postulated that 2.1 is a boundary element/insulator associated protein and the interaction between PCL and 2.1 may occur at the margins of a site of PCL localisation in order to halt the further spread of Pc-G mediated repression. To examine this possibility, high magnification images of PCL polytene sites were analysed with respect to their interactions with 2.1 sites. After analysing numerous PCL sites at random (all the PCL sites shown in the spread in figure 6.4.1.1-2 and many in other spreads) it was found that in nearly all cases PCL sites were intimately flanked on one or both sides by a 2.1 site (figure 6.5.2). In many cases the PCL and 2.1 sites show a small amount of overlap at their margins as indicated by the presence of the purple colour. More obviously in some cases than others the contours of the interacting margins of the PCL and 2.1 sites appear perfectly complementary, indicating that these 'interactions' may be significant (figure 6.5.2).

6.6 - Discussion

As may have been expected for a protein that apparently interacts with PCL and E(Z), 2.1 is present ubiquitously throughout embryogenesis. As was discussed in chapter 1, the ubiquitous distribution of Pc-G proteins enables them to act as transcriptional regulators in any tissue with their specificity of action being determined at a different and as yet unknown level. It is possible therefore that 2.1 could be acting in concert with PCL in any tissue at any stage of embryogenesis. Given, however, that 2.1 is not a stereotypical Pc-G member it cannot be assumed that its direct role would be in transcriptional repression.

As well as the ubiquitous distribution of 2.1 there are increased levels of 2.1 protein in the mesoderm of germ band extended embryos (stage 9) and in segmentally reiterated stripes in late germ band retracted (stage 13) embryos. It is possible that these increased levels have no functional significance but merely reflect the differential strengths of certain enhancers that are utilised by the *2.1* gene to effect ubiquitous expression. Alternatively, it is possible that there is a requirement for an increased level of 2.1 protein in these tissues at these stages. Difficulty in assigning specific roles for 2.1 in these tissues stems from an inadequacy in the understanding of the general role of 2.1. However, analysis of any disruption of mesodermal derivatives and segment polarity defects in *2.1* mutant flies would aid in answering whether or not 2.1 had a specific role in these tissues. Although the 2.1 stripes only appear late in embryogenesis, after segment polarity has been established, preliminary data from detection of β - galactosidase in embryos carrying the enhancer trap P-element within *2.1* (*l(3)00543* - see chapter 5) suggest that the 2.1 stripes are visible much earlier during embryogenesis. It appears that lac Z expression from this P-element is not driven in the ubiquitous pattern (M Coulson and S. O'Connell, personal communication). Therefore the ubiquitous staining seen when detecting 2.1 protein in embryos may obscure the early presence of the 2.1 stripes.

Immunostaining at various stages during *Drosophila* development showed clearly that 2.1 is localised to the nucleus, consistent with the *in vitro* interactions with PCL and E(Z).

Figure 6.3.1 - Double Immunofluorescence of ENGRAILED and 2.1

Embryos are shown laterally with anterior to the left and ventral to the bottom.

a and **b** - stage 13 embryo reacted with rat serum, anti-rat rhodamine conjugated secondary antibodies and anti-mouse FITC conjugated secondary antibodies as a negative control for fluorescence. **a** - detection of 'red' channel fluorescence. **b** - detection of green channel fluorescence. There is no appreciable background fluorescence caused by the rat serum or either of the two secondary antibodies. Notice that although the plane of focus is at the surface of the embryo, the auto-fluorescent yolk present at the centre of the embryo is visible (and out of focus) in both channels. This is inherent to the epi-fluorescence microscopy used to capture these images. This yolk auto-fluorescence is not visualised in the embryo shown in **c-e**, as confocal microscopy was utilised to visualise only the epidermis (surface) of the embryo.

c-e - the posterior half of a stage 13 embryo reacted with anti-2.1 antibodies and anti-ENGRAILED monoclonal antibodies. 2.1 is detected with rhodamine conjugated secondary antibodies and is shown in red (**c** and **e**). ENGRAILED is detected with FITC conjugated secondary antibodies and is shown in green (**d** and **e**). Any overlap between the two signals is shown in yellow (**e**).

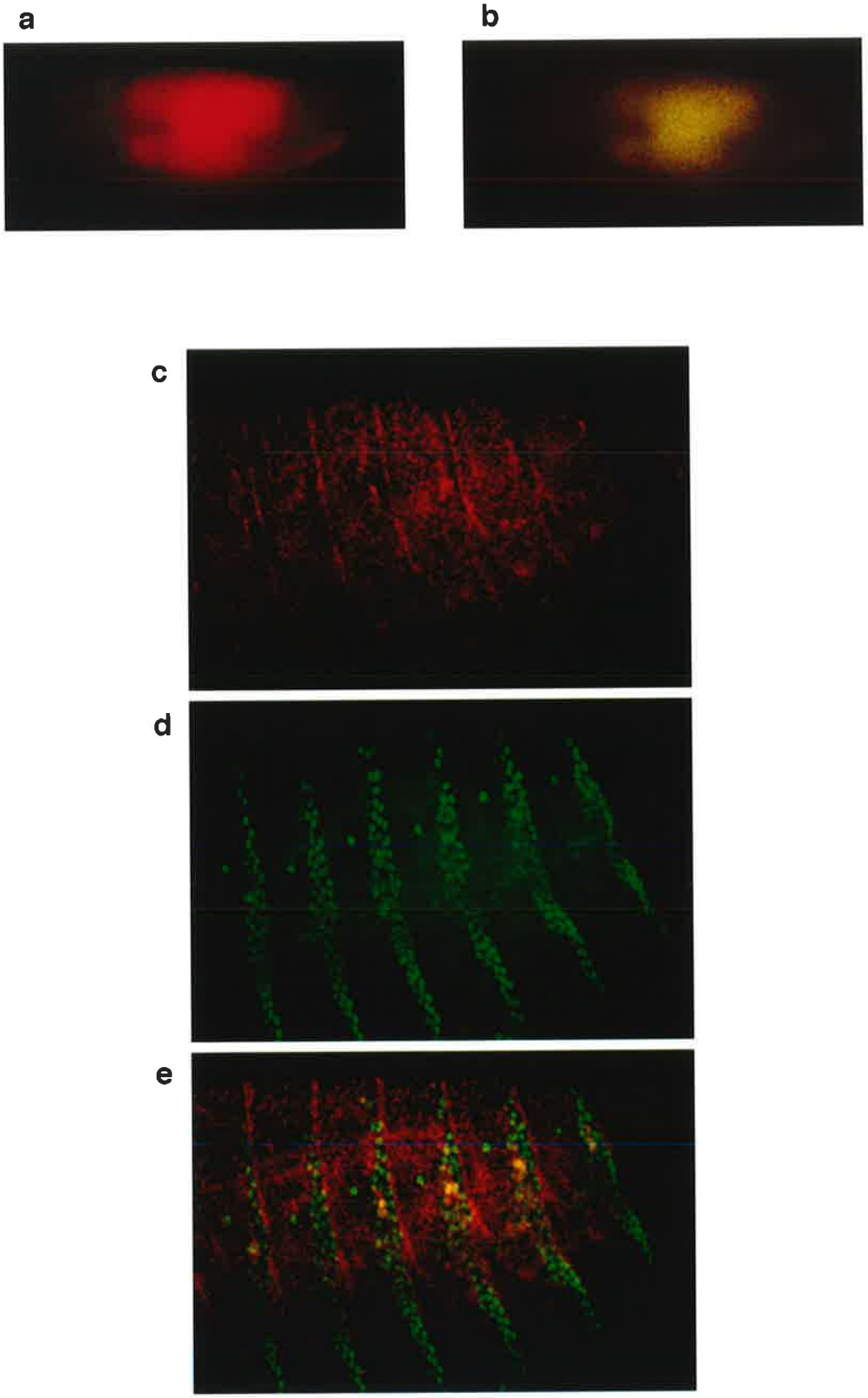

a
b
c
d
e

Figure 6.4.1.1 - Immunofluorescence of 2.1 and PCL on polytene chromosomes

a-b - negative control immunofluorescence of polytene chromosomes. The same chromosome spread, representative of all that were visualised, is shown in **a** and **b**. The chromosomes were treated with rat serum and anti-rat rhodamine conjugated secondary antibodies. The DNA, detected with Hoechst 33258, is shown in **a**. No fluorescence is detected in the red channel in **b** (the exposure time was 20 times that in **a**). This indicates that there is no background staining caused by the rat serum secondary antibodies. Therefore, any fluorescence visualised in the red channel in the experimental spreads shown in figures 6.4.1.1-2, 6.4.2-3 and 6.5.1-2 must be due to the specific detection of 2.1 antigen.

c-h - show the same chromosome spread immunotreated to detect 2.1 (red), PCL (blue) and DNA (green).**c**, **d** and **e** are the single immunofluorescences of 2.1, PCL and DNA, respectively. Overlap between 2.1 and DNA (shown in **f**) is visualised as yellow. Overlap between 2.1 and PCL (shown in **g**) is visualised as purple. Overlap between PCL and DNA (shown in **h**) is visualised as pale blue.

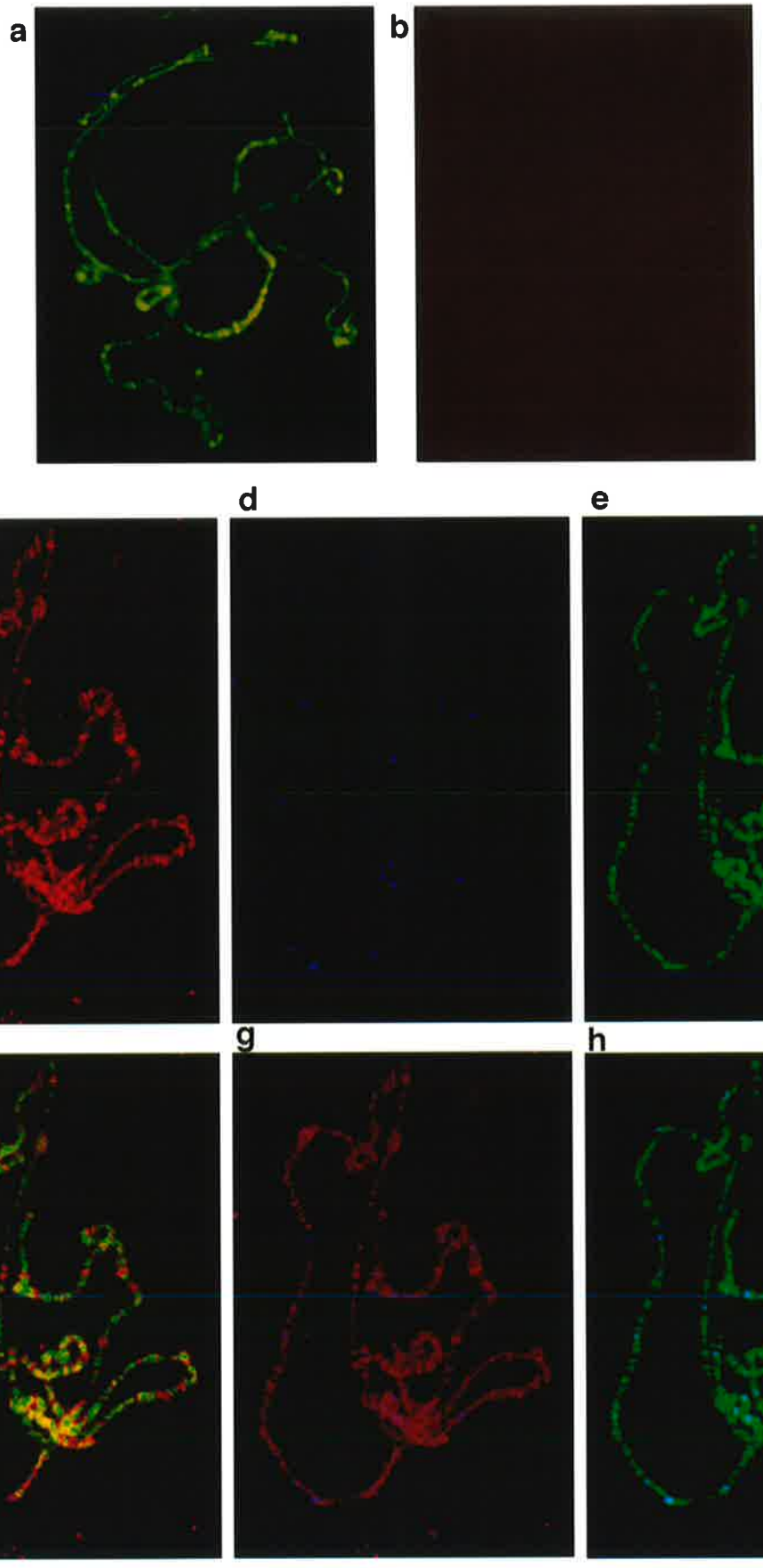

a
b
c
d
e
f
g
h

Figure 6.4.1.2 - Immunofluorescence of 2.1 and PCL on polytene chromosomes

The same chromosome spread is shown in **a**, **b** and **c**. It appears to lack chromosome arm 2L. 2.1 detection is shown in red (**a**), PCL is shown in blue - (**b**) and DNA (stained with chromomycin) is shown in green (**c**).

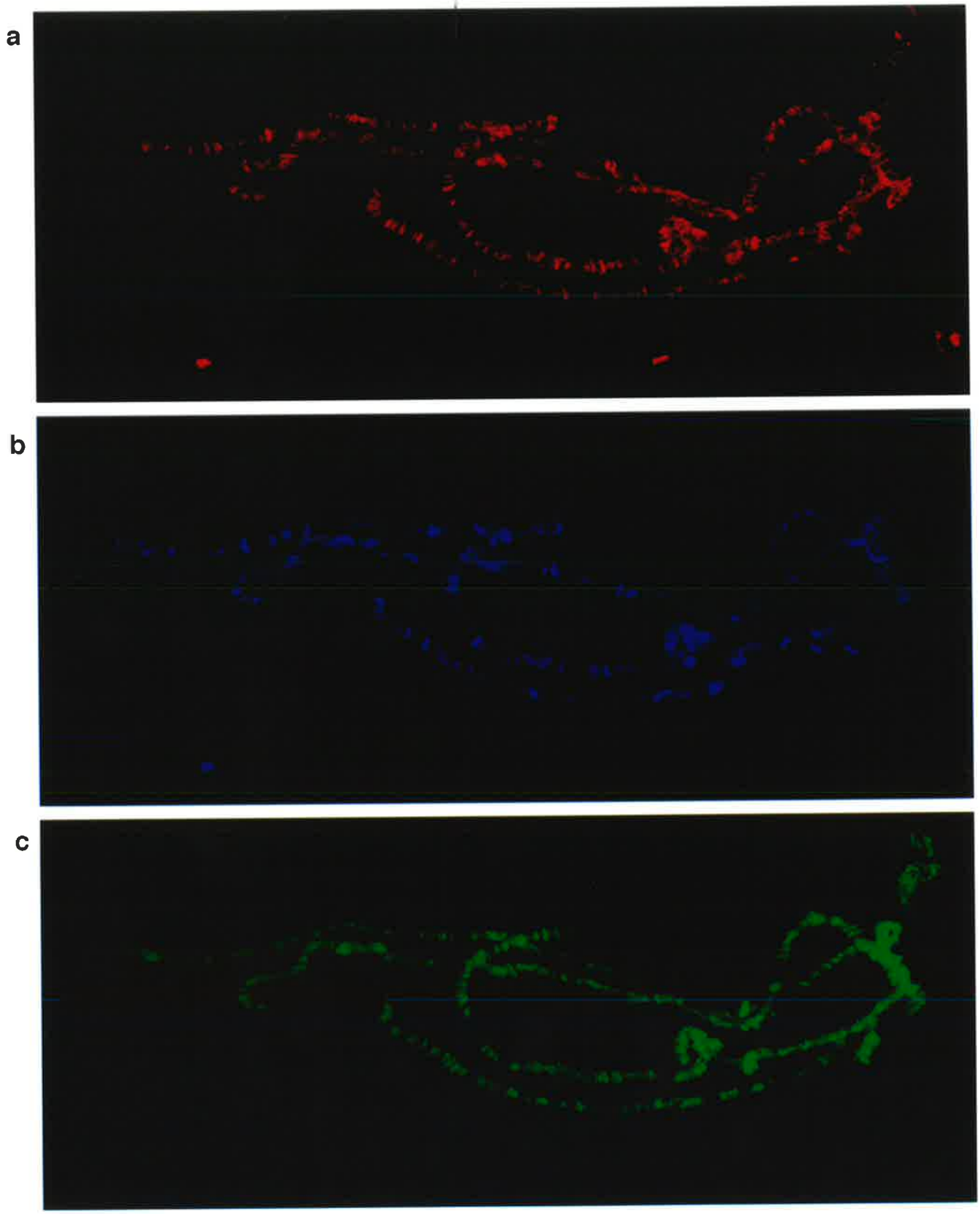

a
b
c

Figure 6.4.2 - Double fluorescence of polytene chromosomes

The same chromosome spread as shown in figure 6.4.1.2 is shown in these three panels. 2.1 is shown in red, PCL is shown in blue and DNA is shown in green. Overlap between red and green appears yellow, between red and blue appears purple and between blue and green appears pale blue.

a - 2.1 (red)/DNA (green) double fluorescence

b - 2.1 (red)/PCL (blue) double fluorescence

c - PCL (blue)/DNA (green) double fluorescence

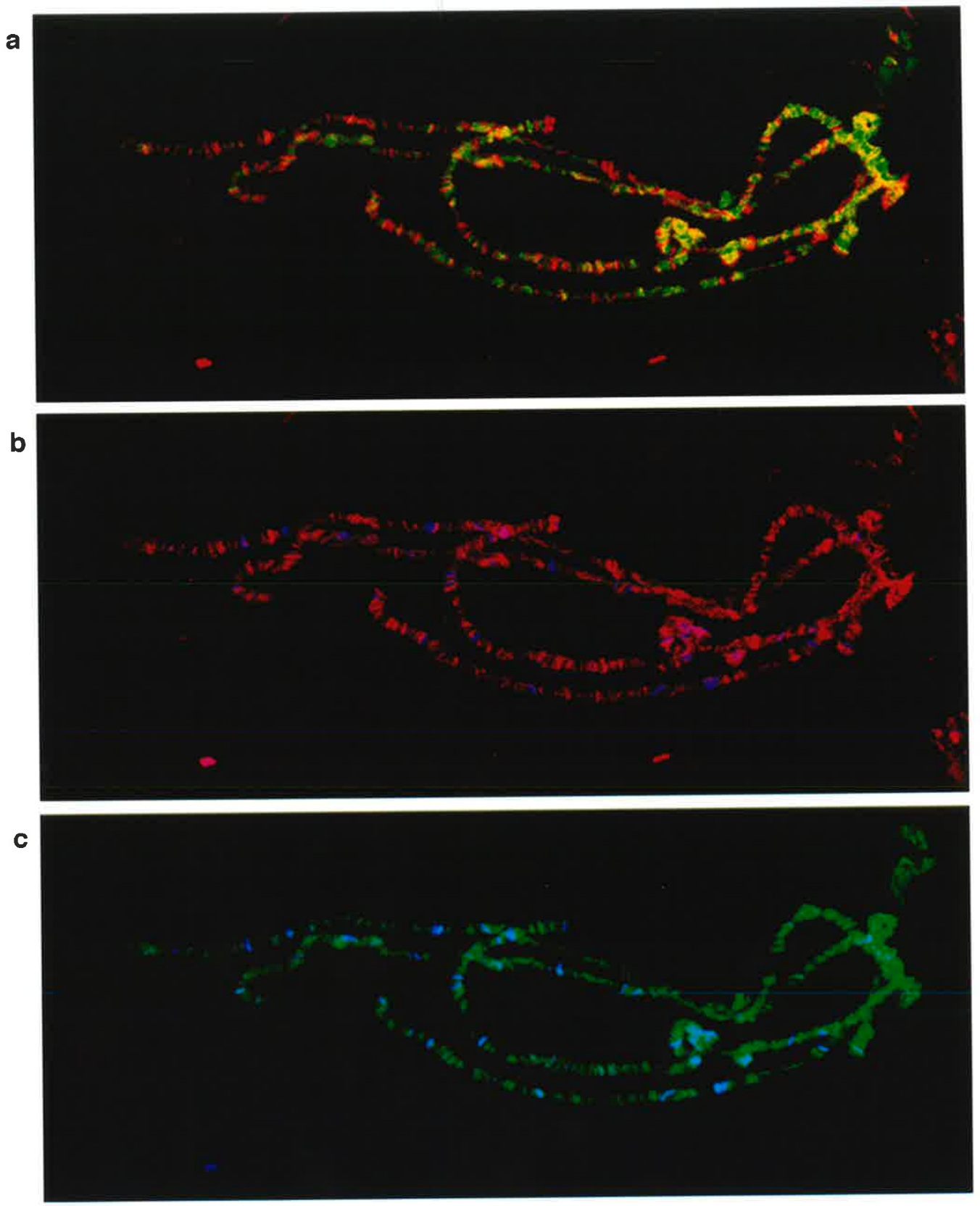

a
b
c

Figure 6.4.3 - High magnification analysis showing the interbanded deposition of 2.1

2.1 is shown in red and DNA is shown in green. The overlap between 2.1 and DNA is yellow. These are high magnification images taken from the same chromosome spread as shown in figures 6.4.1.2 and 6.4.2.

a and **d** - 2.1/DNA

b and **e** - 2.1

c and **f** - DNA

A high magnification image which clearly shows the predominantly interbanded deposition of 2.1 taken from the region in **d**, **e** and **f**. **g** - 2.1/DNA, **h** - 2.1 and **i** - DNA.

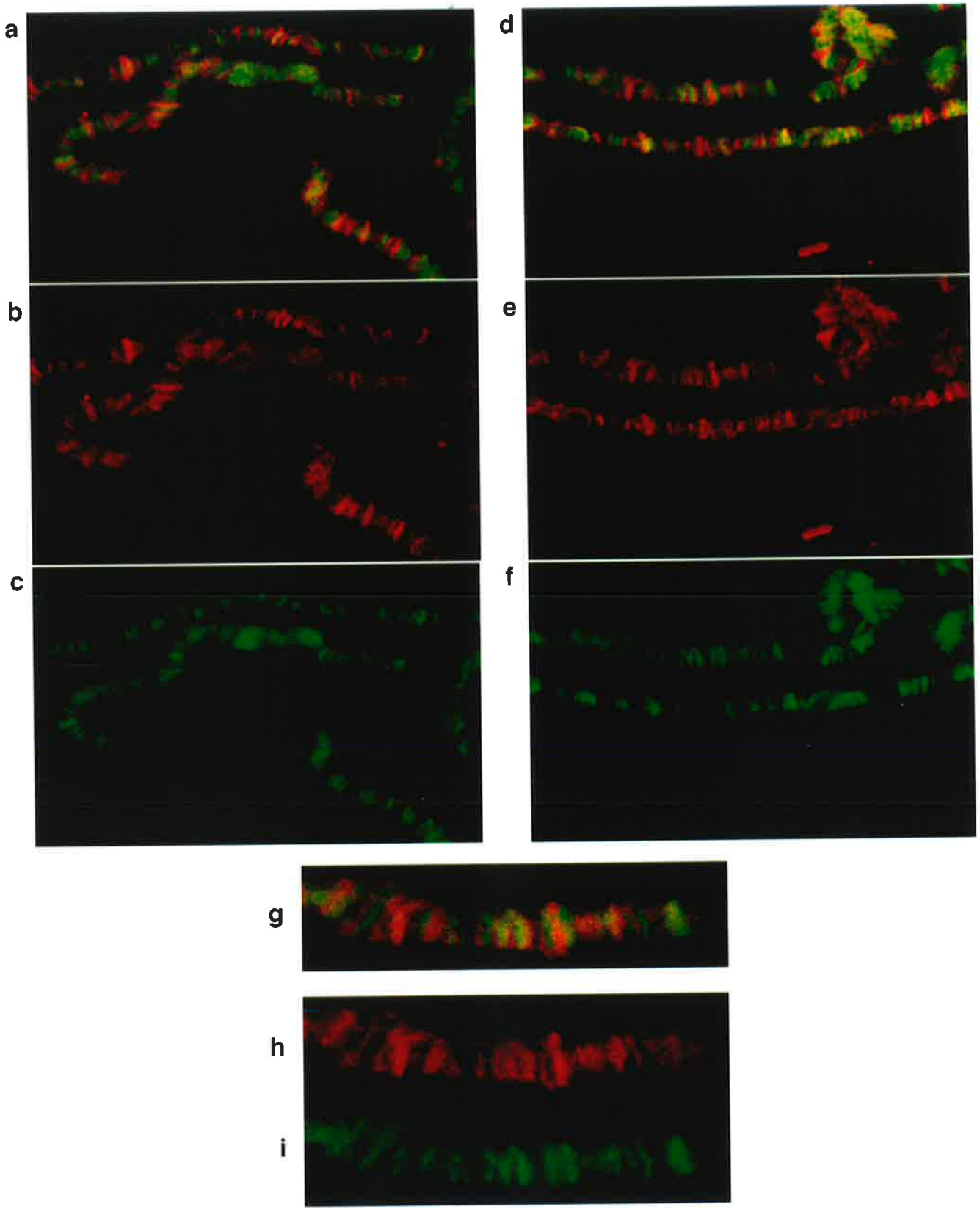

a
b
c
d
e
f
g
h
i

Figure 6.5.1 - High magnification analysis of 2.1 and PCL double immunofluorescence and 2.1 at the chromocentre

The same two chromosomal regions shown in figure 7.4.3 are depicted here. 2.1 is shown in red, PCL is shown in blue and DNA is shown in green.

a, **b** and **c** are 2.1, 2.1/PCL and PCL fluorescence respectively of one region

d, **e** and **f** are 2.1, 2.1/PCL and PCL fluorescence respectively of the other region

g - a high magnification image showing the chromocentre of a chromosome spread. The 2.1 fluorescence was reduced until none of the euchromatin signal was observable, indicating that the staining at the chromocentre is stronger than the staining in other regions

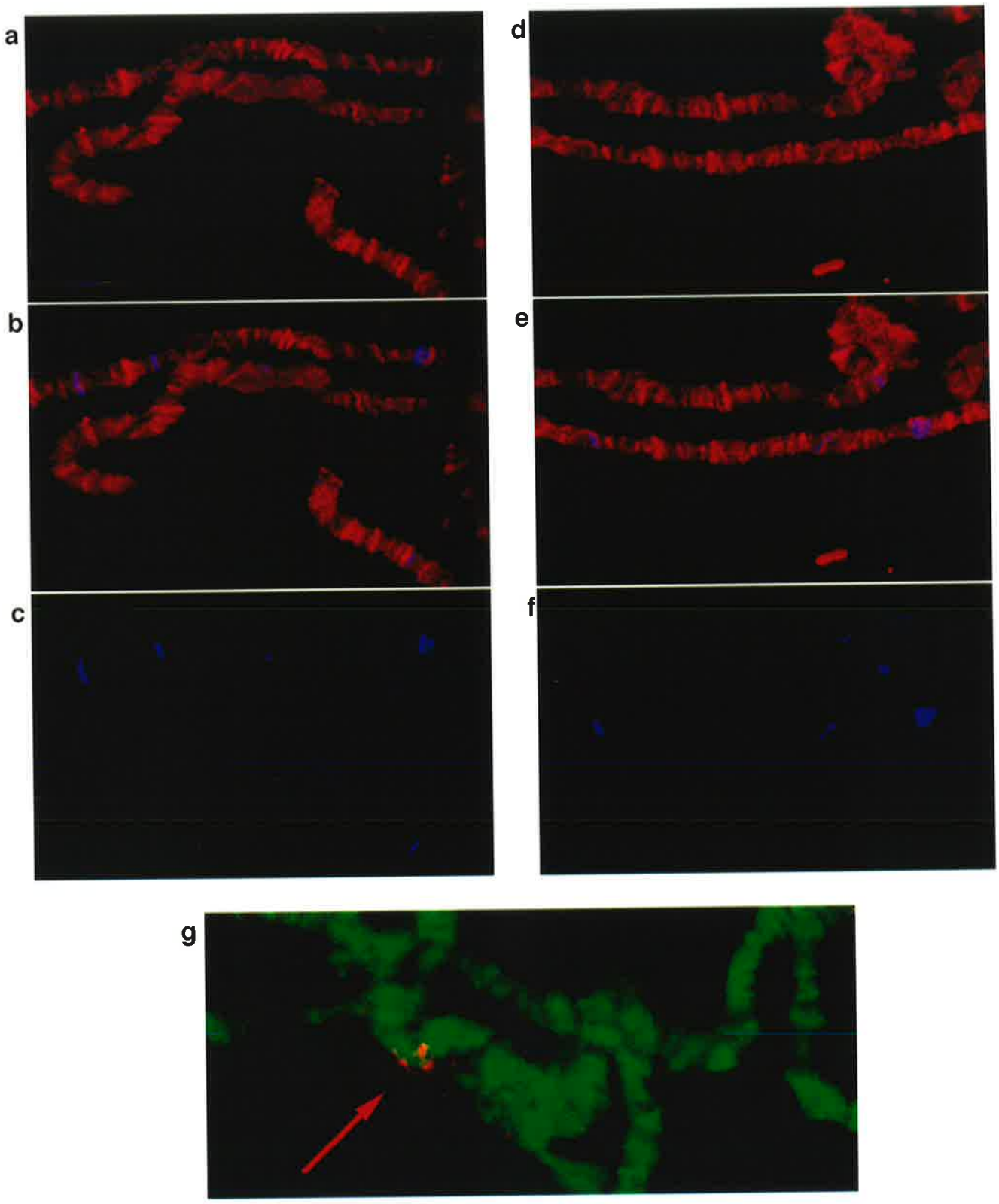
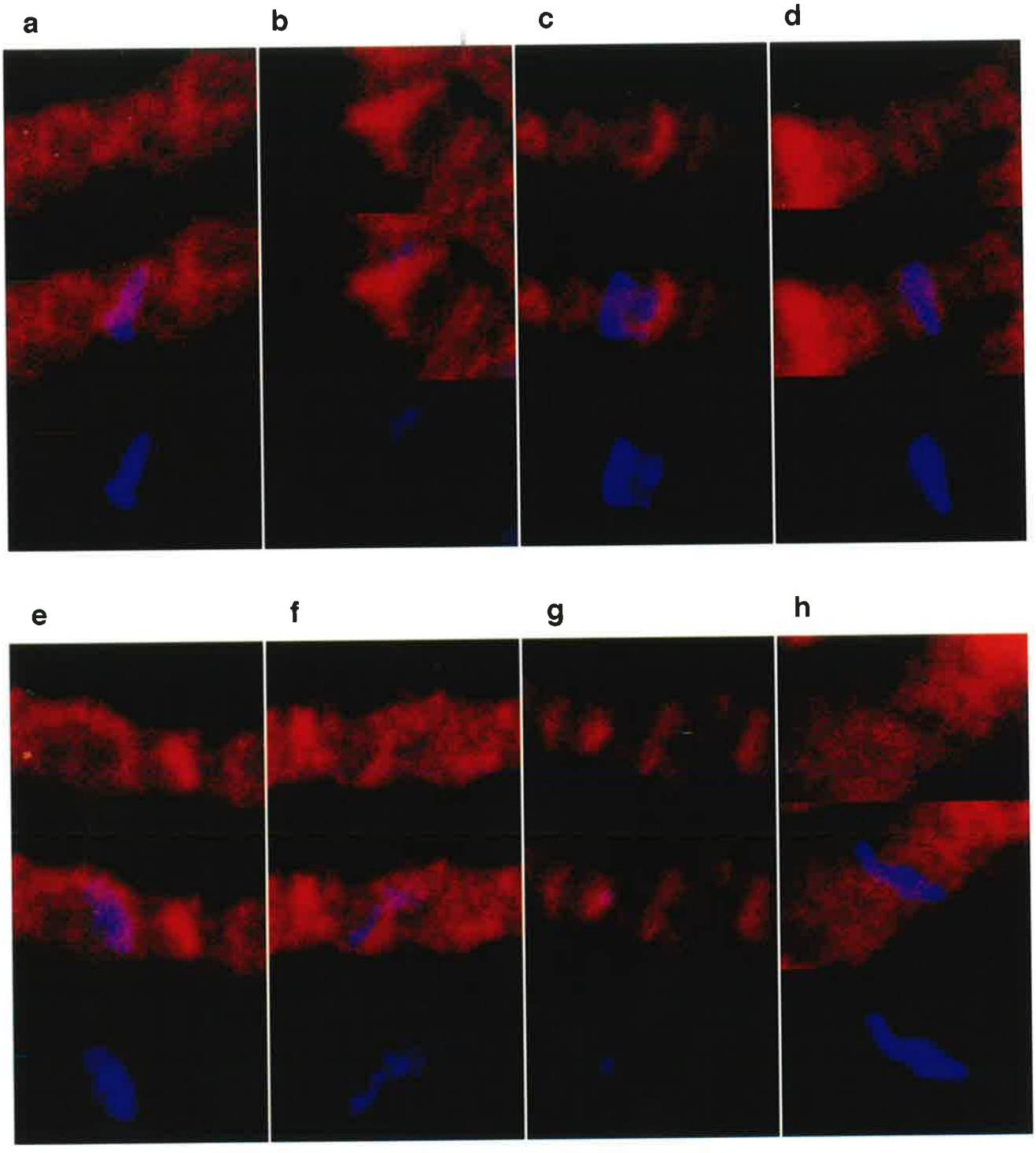

a
b
c
d
e
f
g

Figure 6.5.2 - Double immunofluorescence of 2.1 and PCL showing that 2.1 is present at the margins of PCL bands.

a-h - show high magnification images of polytene chromosomes. 2.1 is shown in red on top, PCL is shown in blue at the bottom and the overlayed images are shown in the middle. Any overlap of 2.1 and PCL is shown in purple.

In some cases PCL bands are flanked on both sides by 2.1 (**d** and **h**). Note how the margin of the 2.1 signal and PCL signal appear to complement each other (especially in **c**, **e**, **f** and **h**) suggesting abutment of the proteins, although the resolution is too low to be certain.


a
b
c
d
e
f
g
h

The data generated by immunofluorescence of polytene chromosomes raised more questions than they answered. Firstly and most strikingly, 2.1 does not have the same distribution as PCL and, hence, most of the Pc-G proteins. It was not found at about 100 discrete sites but was found at many sites throughout the chromosomes, generally in between the polytene bands (interbands).

Double immunofluorescence of 2.1 and PCL superficially appeared to give the unexpected result of no significant co-localisation of these two proteins. Except for a few sites, none of the strong bands of 2.1 fluorescence overlapped with those of PCL. How could this be explained given the strong *in vitro* association of PCL and 2.1? Firstly, it is possible that PCL and 2.1 do not interact *in vivo* at all and that the observed *in vitro* interaction is artifactual. Secondly, it is possible that PCL and 2.1 interact in some tissues but not in the tissue (salivary gland) that was used in these immunofluorescence experiments. This could be tested by performing co-immunoprecipitations on salivary gland protein extract. However, given that most of the PCL and 2.1 protein present on polytene chromosomes does not appear to interact, this low level of putative interaction may not be detected by this method.

Two striking features of the immunofluorescence were noted. Almost all the PCL sites detected corresponded exactly with chromosomal bands. This was not surprising given that the bands are transcriptionally inactive chromatin regions (Rykowski *et al.*, 1988). It is only the few PCL bands that are present in interband regions that appear to co-localise obviously with 2.1 (as seen in figure 6.4.2). This seems to support the idea that this apparent co-localisation is merely coincidental. Alternatively this observation may support the idea that 2.1 may have the very specific role of interacting with PCL at un-repressed (interband) loci.

From these data, what could be said about the nature of 2.1 and the PCL interaction? Given that interband chromatin is considered to be transcriptionally active (Rykowski *et al.*, 1988), it is possible that 2.1 has a role in generating and/or maintaining transcriptionally active chromatin. If the absence of 2.1/PCL interaction was observed by co-immunoprecipitation experiments using *Drosophila* extracts, it may have to be concluded that the *in vitro* interaction detected between 2.1 and PCL is indeed artifactual. However, for the reasons mentioned above, co-immunoprecipitation experiments may not be a sensitive enough assay. It is possible that 2.1 contains spurious similarity with a veritable PCL interactor and that this causes a biologically insignificant *in vitro* interaction with PCL. If, however, co-immunoprecipitation experiments provide evidence for an *in vivo* interaction between these two proteins, their apparent lack of co-localisation on polytene chromosomes would have to be explored further. One possibility, based on the assumptions that PCL is exclusively associated with repressed chromatin and 2.1 is exclusively associated with active chromatin (see above), is as follows. In chapter 1 it was described that during early *Drosophila* embryogenesis genes such as the homeotic genes have their transcriptional states determined and then set for the remainder of development. It is possible that PCL and 2.1 interact at this stage of development only, as part of the process initiating the maintenance of repressed or active transcriptional states, respectively. Co-immunoprecipitation experiments should be carried out to determine if 2.1 and PCL interact at this early stage of embryogenesis.

In contrast to the possibility that PCL and 2.1 do not colocalise on polytene chromosomes, it was also considered possible that 2.1 and PCL interact only where the PCL and 2.1 sites of fluorescence abut each other. Further investigation of this idea demonstrated that this indeed may be the case given that in almost all cases PCL sites were intimately flanked by 2.1 sites and that the margins of interaction were complementary and appeared to contain a narrow region of overlap. It is possible however that this merely reflects PCL association with bands and 2.1 association with interbands and their apparent colocalisation at these band-interband borders is a by-product of the complementary nature of bands and interbands. If it is assumed that the apparent colocalisation of PCL and 2.1 at the margins of their bands of localisation is biologically relevant, how could these data be explained and further explored? The interband localisation found for 2.1 and several other *Drosophila* proteins has been reported for two boundary element associated proteins, BEAF-32 (Zhao *et al.*, 1995) and BANGDOO (Gaszner *et al.*, 1996). In chapter 1 it was described that Pc-G proteins apparently respect chromatin boundaries by not 'spreading' to repress genes distal to boundary elements. The mechanism by which this occurs is unknown. If it is assumed that 2.1 is present at chromatin boundaries, it is possible that the PCL/2.1 interaction is responsible for the generation and/or stabilisation of the interface between Pc-G repressed domains and chromatin boundaries. It should be noted that although 2.1 is interband deposited, it does not occupy the entire extent of each interband space. Rather it appears that 2.1 is present right next to chromosome bands (see figure 6.4.3). This suggests that 2.1 may not just be associated with all active chromatin and supports the idea that 2.1 may be a boundary element associated factor. To analyse whether 2.1 is localised on polytene chromosomes in the same way as boundary element factors double immunofluorescence with antibodies to 2.1 and the known boundary element proteins (BEAF-32 and BANGDOO) should be carried out.

If 2.1 was indeed responsible for binding to PCL and abating the spread of Pc-G repression, it would be reasonable to question why some PCL sites are only flanked on one side by a 2.1 binding site. It is possible that there are other boundary element proteins that also function in the same way thereby resulting in some PCL sites being flanked by 2.1 on one side and another protein on the other. In fact there is a precedent for polarity in a chromosomal domain with respect to its boundary elements. The heat shock puff at 87A7 is bounded on one side by the scs' element and therefore BEAF-32 and on the other side by the scs element and therefore some scs associated protein (Zhao *et al.*, 1995). The reason for this polarity is unclear. Experiments which explore the role of 2.1 as a boundary element factor are proposed in detail in chapter 8.

The presence of a high level of 2.1 in a discrete position within the chromocentre of polytene chromosomes is very curious. No Pc-G proteins localise to this heterochromatin region. BEAF-32 also does not localise to the chromocentre (Zhao *et al.*, 1995). In chapter 1 the hypothesis that Pc-G repressed loci may be localised to regions of the nucleus that contain heterochromatin was discussed. Part of this hypothesis was that protein interactions might mediate these long distance chromosomal interactions. If this were the case it may be possible that 2.1 is responsible for localising some Pc-G repressed loci to the particular region of heterochromatin with which it appears to be associated, via

interactions with PCL and E(Z). Experiments to test this are proposed and discussed in detail in chapter 8.

The work presented in this chapter demonstrates that 2.1 is a chromosomal protein with a novel distribution and supports the idea that it is a protein of novel function.

Chapter 7 - Is PCL a DNA binding protein?

7.1 - Background

To understand the mechanism of repression by the Pc-G proteins, an understanding of the mechanism of DNA association by the complex is necessary. Somehow the Pc-G must come into contact with DNA either directly or indirectly before it can stably repress a gene. None of the characterised Pc-G members encode any known DNA binding domain, although an as yet uncloned member of the Pc-G may encode a DNA binding domain or one of the presently cloned members may contain a novel DNA binding domain. An alternative possibility is that a 'super-domain' is formed by the association of several of the members or that the complex does not even contact DNA but recognises some aspect of chromatin. The first of these possibilities is being addressed incidentally whenever a group identifies, clones and sequences a new Pc-G gene. The 'super-domain' hypothesis could be tested by reconstituting the complex *in vitro* in a step-wise manner while assaying for DNA binding activity. Presumably the most likely target DNA to use in such assays would be a PRE, given the circumstantial genetic and molecular data that suggest a Pc-G/PRE interaction. That the Pc-G recognises some aspect of chromatin rather than naked DNA is more difficult to test, but it remains an important possibility to bear in mind. The second possibility, that one of the cloned Pc-G genes encodes a novel DNA binding domain has been tested or is in the process of being tested in various laboratories. None of the characterised Pc-G members have been reported to encode an assayable DNA binding activity, except for PH which may have some non-specific DNA binding activity (H.Brock, personal communication).

Using a BLAST search algorithm (Altschul *et al.*, 1990) and the PCL amino acid sequence as a search probe, a murine cDNA termed M96 was identified as encoding a protein with significant sequence similarity to PCL (figure 7.1.1). M96 is the first protein that has significant similarity to PCL outside the PHD finger domain.

M96 is a murine gene that encodes a protein that was isolated by its ability to bind to a human metallothionein response element (MRE) sequence in a yeast one-hybrid assay(Inouye *et al.*, 1994). The possibility was considered that PCL may also bind to DNA in a similar fashion to M96 given this high level of sequence similarity. In an effort to address this question, bacterially expressed PCL was synthesised and used in DNA binding assays. The experiments described below showed that PCL appeared to exhibit a sequence specific DNA binding activity which upon closer scrutiny was shown to be artifactual.

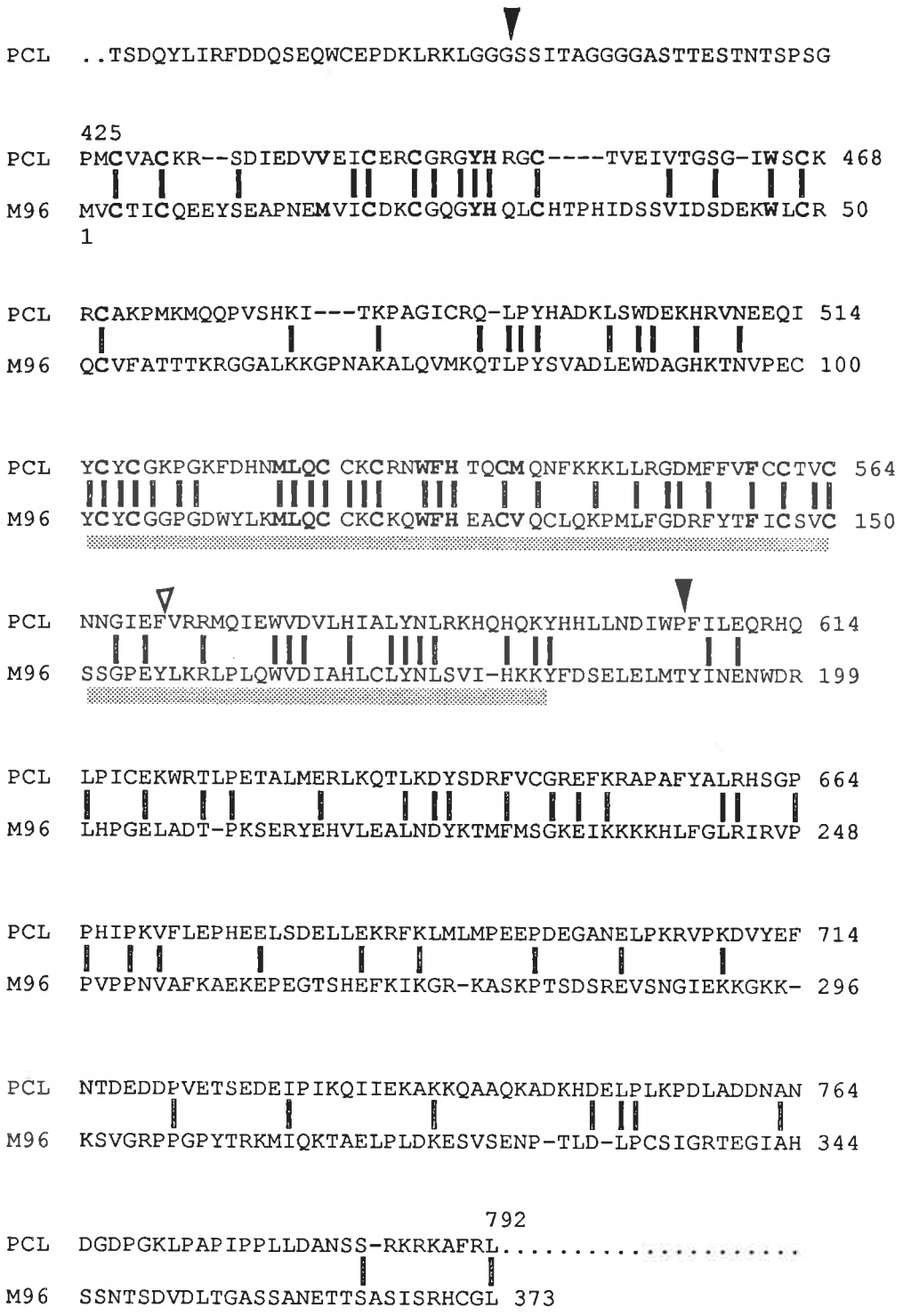


Figure 7.1.1 - alignment of PCL and M96

Only the identical residues are indicated (with bold vertical lines). PHD finger conserved residues (Aasland *et al.*, 1995) are in bold type. A block of higher sequence identity that overlaps the second PHD finger and then extends further towards the carboxy-terminus is highlighted with a grey line below the alignment. The black arrow heads denote the limits of the PCL protein expressed in cDOM constructs (see section 7.4). The open arrow marks the COOH-terminal limit of the PCL peptide expressed in ΔEH constructs (see section 7.4). Note the colinear alignment between PCL and M96 for the entire extent of the M96 protein. The proteins have 42/83 (50.6 % identity) residues identical in the high identity region and 104/373 residues identical (27.9 % identity) over the whole protein.

7.2 - PCL has an apparent DNA binding activity

GST-PCL was expressed from the plasmid pGEX-*Pcl* Bam and refolded from the insoluble fraction. In order to repeat the experiments as were done with M96, the exact DNA probe containing two MRE binding sites (MRE d and MRE c) (Inouye *et al.*, 1994) was synthesised. This probe was radioactively end-labelled, using polynucleotide kinase, to high specific activity and then used in retardations. As a negative control for the binding GST alone was expressed and purified on GSH-agarose according to the manufacturer's instructions (Pharmacia). As shown in figure 7.2.1a, PCL appeared to bind to the MRE probe in a sequence specific fashion. Lanes 1 and 2 contain probe alone (2 ng in each track) and probe with 800 ng of GST protein respectively. Lane 3 contains probe and GST-PCL extract with approximately 800 ng of the GST-PCL protein. This track shows two retarded complexes. Lanes 4, 5 and 6 are the same as lane 3 except they contain 1, 2 and 3 µg of poly dIdC (Pharmacia) as a non-specific DNA competitor. Even 3 µg of this competitor (1500 fold mass excess) failed to compete the binding significantly of the higher mobility complex. The lower mobility complex was competed and therefore is non-specific. Lanes 7 and 8 are identical to lane 3 except that they contain 0.5 µg and 1 µg of cold MRE oligonucleotide as a specific competitor respectively. 0.5 µg of specific competitor (250 fold excess) was sufficient to compete the binding. This strongly indicated that the observed binding was sequence specific. Now that the DNA binding activity of PCL appeared to have been established, it was decided that the binding of PCL to a PRE, a biologically relevant target DNA, should be attempted.

The PRE from the regulatory region of *Ubx* at map position of approximately -8.2 - -6.7 (Chan *et al.*, 1994) was used. Note however that only one consensus MRE was found within the sequence of this 1.6 kb PRE (the sequence of the PRE was extracted from GenBank). This DNA was obtained as a plasmid clone which contained this region as well as flanking DNA. The 1.6 kb PRE was excised as an EcoRI/StyI fragment, digested with HinfI and end-labelled with α-^{32}P-dATP and DNA polymerase Klenow fragment. The collection of 5 fragments generated were used in gel mobility shift assays using the same reaction conditions as used for the above retardation. Figure 7.2.1b shows the results of such an experiment. Lanes 1 and 2 show the probe alone and the probe with approximately 800 ng GST protein added respectively. No mobility shift is detected with the GST protein once again. Lane 3 shows the result of adding approximately 800 ng of GST-PCL protein. Here PCL specifically and completely retards the mobility of the 825 bp fragment and possibly shifts some of the other fragments only partially. The 825 bp fragment contains the consensus MRE but presumably contains other sequences that can be bound by PCL thereby explaining the increased efficiency of the DNA binding as compared to that shown for the MRE oligonucleotide.

7.3 - PCL appears to be a low affinity DNA binding protein

In order to assess the DNA binding affinity of PCL (dissociation constant, K_d) a mobility shift experiment using a fixed amount of DNA probe and increasing amounts of GST-PCL was employed. Figure 7.3.1a shows a retardation in which 80, 240, 400, 560 and 720 ng of GST-PCL protein were

used in tracks 2-5 respectively. Lane 1 contained the 825 bp fragment alone. Quantitation of the unretarded 825 bp band relative to lane 1 where no protein was added showed that in lane 4 (400 ng of GST-PCL) approximately half the probe was unbound and therefore half was bound.

$$Kd = \frac{[protein].[DNA]}{[protein/DNA]}$$

Assuming that when the [protein] is at >100 fold excess over the [DNA], that the K_d approximates the concentration of the protein required for binding half the DNA and that the molecular weight of the GST-PCL is 120000 (25 kD for GST and 95 kD for PCL), the K_d for GST-PCL is approximately $1.6X10^{-7}$ M. This is on the low side for sequence specific DNA binding proteins in which the observed range is 10^{-8} - 10^{10} M (Ausubel *et al.*, 1987). 6Xhis tagged PCL (6H-PCL), using the clone pET15b-*Pcl* Nde/Sal, was expressed and purified from the soluble fraction (figure 7.4.1b). Use of this protein in identical retardations, side by side with the GST-PCL fusion protein, showed that this protein apparently had a significantly lower binding affinity (not shown). Although this was initially a concern, it was subsequently postulated that the GST-PCL may bind with higher affinity because of the fact that GST dimerises in solution (Henderson *et al.*, 1996; Riley *et al.*, 1996). This dimerising activity may cause the GST-PCL to bind to the DNA in a co-operative manner.

7.4 - Defining the DNA binding domain of PCL

In order to define the region of PCL which contained the DNA binding activity, three mutant GST-PCL constructs were synthesised. These mutations were designed on the assumption that the DNA binding domain was contained within the region of similarity with M96, given the similarity of DNA binding specificity between the two proteins.

The BamHI insert from the clone pEG-cDOM (see above) was excised and subcloned into pGEX-3X to create the clone pGEX-cDOM. This construct was expected to produce a fusion protein containing GST and the M96 conserved region (residues 403-605) of PCL. Analysis of the M96 conserved region suggested that this region could be separated into two parts. At the NH_2-terminal end are the two PHD fingers, comprising the first part. Th second part, overlaps with the second PHD finger and extends further towards the COOH-terminal end of this region. This region is unique to PCL and M96 and is termed the extended homology domain (EH) (chapter 1). In order to test whether or not it is the EH domain that is responsible for the DNA binding a clone was produced which would express both PHD fingers and a disrupted EH domain. pGEX-cDOM was restricted with EcoRI to release an approximately 80 bp fragment from the EcoRI site at position 1632 in the *Pcl* ORF to the EcoRI site in the polylinker of pGEX-3X at the 3' end of the clone. This 'drop-out' was religated to form the clone pGEX-ΔEH. This clone was expected to produce a GST fusion with residues 403-581 of PCL, thereby removing most of the EH domain but leaving both PHD fingers intact. The third mutant clone was full length PCL with a single residue change in the first PHD finger. The Cys at position 430 was changed to a Ser using an overlapping PCR mutagenesis approach (see chapter 2). Codons 429 and 430 were changed from GCC TGC (encoding Ala and Cys) to GCT AGC (encoding Ala and Ser), thereby also incorporating a novel, unique NheI restriction site which was used to assay for the

Figure 7.2.1 - Apparent retardation of an MRE and a PRE by PCL

a - shows the apparent sequence-specific retardation of the MRE d oligonucleotide by PCL. Lane 1 contains probe alone. Lane 2 contains GST extract. Lane 3 is the same as lane 2 except that 1 µg of poly-dIdC was added as a non-specific competitor. Lane 4 contains GST-PCL extract. Note the presence of a high mobility complex that is competed by non-specific DNA competitor (lane 5-7). Lanes 5-7 are the same as lane 4 except with 1, 2, and 3 µg of poly-dIdC added. Lanes 8 and 9 are the same as lane 4 except with 250 and 500 fold excess respectively, of the cold MRE d oligonucleotide was added as a sequence specific competitor. Lane 10 is the same as lane 9 except with 1 µg of poly-dIdC added.

b - retardation by GST-PCL extract of a PRE from *Ubx* digested with HinfI and end-labelled. Lane 1 shows the probe with no added test protein. Lane 2 and 3 are the same as lane 1 except that GST extract and GST-PCL extract respectively, were added. All lanes contain 1 µg of poly-dIdC as a non-specific competitor. Notice the complete, specific retardation of the 825 bp fragment by GST-PCL extract only.

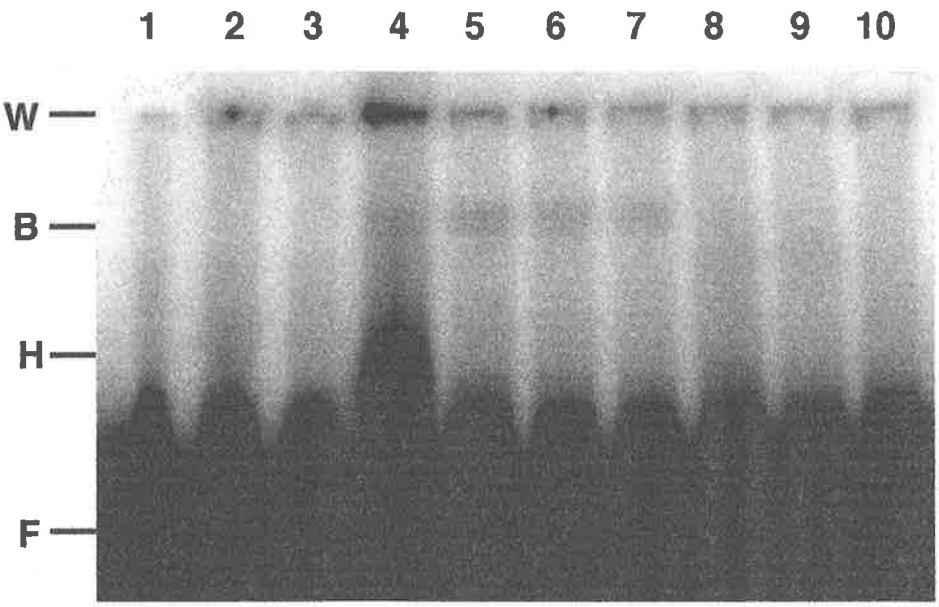
W = non-specific probe adherence to the wells

B = bound probe

H = high mobility complex

F = free probe

a



b

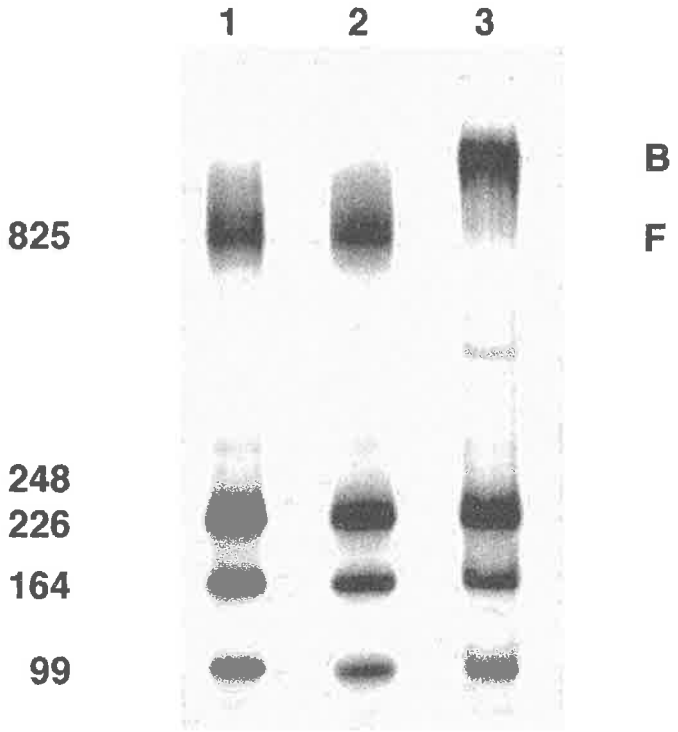
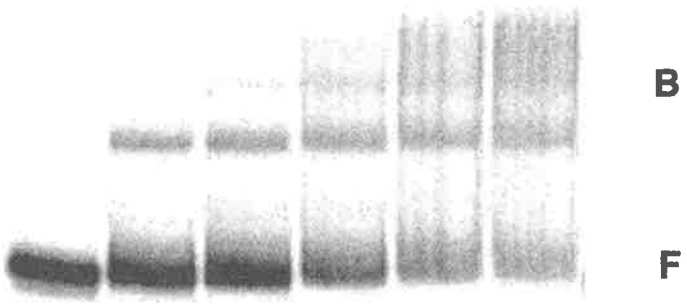


Figure 7.3.1 - GST-PCL extract binds to the 825 bp PRE fragment with low affinity

a - retardation of the 825 bp PRE fragment with 0, 80, 240, 400, 560 and 720 ng of GST-PCL (as GST-PCL extract) in lanes 1-6 respectively. The identical amount of probe is added in each reaction. B = bound probe, F = free probe.

b - ratio of free probe to the free probe present in lane 1. This figure is shown in bold in lane 4 as it corresponds to approximately half of the probe bound.

a



1 2 3 4 5 6

b

1	0.87	0.81	**0.52**	0.36	0.26

Figure 7.4.1 - A bacterial protein is responsible for the DNA binding activity

a - Coomassie stained gel of inductions of the proteins used in the study. Bacterial cultures were pelleted, resuspended, sonicated, pelleted and then the supernatant and the resuspended pellet (insoluble fraction) were run on the gel. Lane 1 - supernatant from an induction of GST alone. Lanes 2 and 3, 4 and 5 - pellet and supernatant from inductions of GST-PCL and GST-cDOM respectively. Lane 6 - the pellet from an induction of GST-ΔEH.

b - protein preparations (10 µL each) used in the retardations in **c** and **d**. Lane 1 - GST extract ; soluble GST purified and reconstituted with resolubilised insoluble proteins. Lane 2, 3, 4, and 5 - resolubilised GST-PCL, GST-cDOM, GST-C430S and GST-ΔEH extracts respectively. Lane 6 - 6H-PCL purified on Ni-NTA silica.

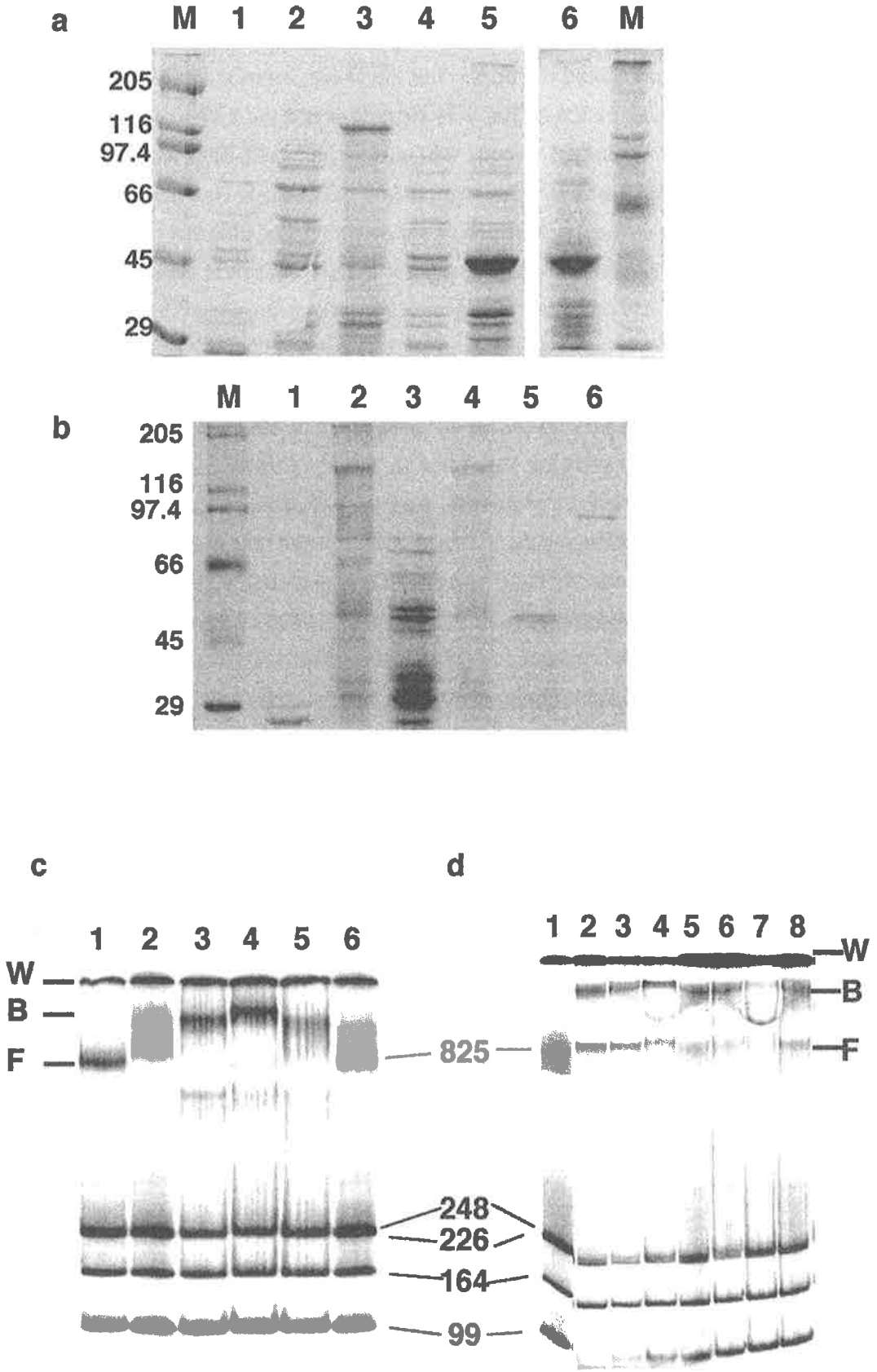
c - retardation of the PRE digested with HinfI. Lane 1 - probe alone. Lanes 2-6 - as lane 1 with GST, GST-PCL, GST-cDOM, GST-C430S and GST-ΔEH extracts added respectively. The amounts of protein added are shown in the Coomassie stained gel in **b**.

d - similar to the retardation in **c**. Lane 1 - is probe alone. Lanes 2 and 3 - with 5 and 10 µL of GST extract added respectively. Lane 4 - is the same as lane 3 except with anti-PCL antibodies added. Lanes 5, 6 and 7 are similar to lanes 2, 3 and 4 except that GST-ΔEH extract was added instead of GST extract. Lane 8 - contains 10 µL of GST-PCL extract.

W = non-specific probe adherence in the wells

B = bound probe

F = free probe


a
M 1 2 3 4 5 6 M
205
116
97.4
66
45
29
b
M 1 2 3 4 5 6
205
116
97.4
66
45
29
c
1 2 3 4 5 6
W
B
F
825
248
226
164
99
d
1 2 3 4 5 6 7 8
W
B
F

presence of the mutation (not shown). Ser was chosen as a replacement for Cys as it was not expected to be able to chelate a metal ion, but is approximately the same size and charge as Cys. In this way the folding of the PHD finger should be disrupted but primary disruptive effects on general folding of the protein should not occur. This clone was inserted into the BamHI site of pGEX-3x and named pGEX-C430S. pGEX-C430S was expressed and the fusion protein was purified side by side with GST-PCL as described above. GST-cDOM and GST-ΔEH were purified from the insoluble fraction. At this point it was decided that purified GST protein was not a suitable negative control for the binding assays as it did not include all the contaminating bacterial proteins that were re-solubilised from the insoluble fraction that were present in the preparation of the other fusions. In order to rectify this situation, insoluble proteins from bacteria that were expressing GST alone were re-solubilised and then added back to the purified GST protein (GST-extract) (figure 7.4.1b). This mixture was then used as the new negative protein control.

A gel mobility shift was carried out once again on the HinfI digested PRE DNA using all of the newly prepared protein samples. Figure 7.4.1b shows a Coomassie stained gel of the protein samples used. The same amount of protein as was loaded in each track of the gel was used in each corresponding retardation reaction. Figure 7.4.1c shows the results of this experiment. Lane 1 is probe alone. Lane 2 contains GST-extract and shows significant retardation of the 825 bp fragment. GST-PCL (lane 3) once again showed retardation of the 825 bp fragment. This was similar to what was observed for GST-cDOM (lane 4) and GST-C430S (lane 5). GST-ΔEH (lane 6) showed a small amount of retardation of this fragment. Given that the negative control showed retardation, the retardation observed in all of the experimental tracks was viewed with suspicion. Closer inspection of the Coomassie stained gel of the protein samples used in the experiment (figure 7.4.1b) revealed an interesting correlation. A ranking of the levels of retardation observed from lowest to highest was GST-ΔEH, GST-extract, GST-PCL and GST-C430S followed by GST-cDOM. This order was the same for the levels of contaminating bacterial proteins present in each sample from lowest to highest. The possibility that a protein or proteins in the bacterial extract were responsible for the DNA binding activity was therefore considered. To test this another retardation using GST-extract, GST-ΔEH and GST-PCL was performed this time with approximately equivalent amounts of bacterial proteins in each reaction. Note that in the previous reaction the samples were made equivalent for the level of the expressed fusion protein and the levels of bacterial proteins reflected the relative differences in expression efficiency from each of the clones (figure 7.4.1b). Furthermore, to test for a specific PCL involvement in the binding anti PCL-antibodies were added in order to either supershift or disrupt the binding. Figure 7.4.1d shows the results of this experiment. Lane 1 is probe alone. Lane 2 and 3 have 5 and 10 μL respectively of GST-extract. Notice the significant retardation of the 825 bp product to equivalent levels as that observed with 10 μL of GST-PCL (lane 8). Similarly, lanes 5 and 6 show reactions that contained 5 and 10 μL of GST-ΔEH. Once again the levels of retardation are similar to that seen in lane 8. Lanes 4 and 7 show that no significant supershifting or disruption of the complexes is observed upon the addition of 1 μL of affinity purified anti-PCL antibodies to the reactions containing GST-extract and GST-ΔEH respectively (note the strange effect on the electrophoresis in these tracks presumably due to the addition of a large amount of extra protein

(antibody)). This result confirmed that the retardation observed was due to a bacterial protein present in the extracts and not GST-PCL or its derivatives.

7.5 - Discussion

This chapter describes a series of experiments to test whether or not PCL exhibits any DNA binding activity. The possibility that PCL could bind to DNA in a sequence specific manner was raised by the observation of sequence similarity between PCL and a novel murine protein of unknown function, M96. M96 was identified in a yeast one-hybrid screen for proteins that could bind to a human MRE-containing oligonucleotide. M96 was subsequently shown to have *in vitro* DNA binding activity in a series of mobility shift assays. GST-PCL was expressed in bacteria and re-solubilised but not purified and used in mobility shift assays. This protein extract was shown to contain an MRE binding activity *in vitro* using a mobility shift assay as described above. As with M96 the ability of PCL to bind to the MRE oligonucleotide was shown to be sequence specific. Subsequently, the ability of GST-PCL to bind to a PRE from the Ubx gene was tested. GST-PCL extract was able to retard one region of the PRE in a sequence specific manner. It was in the course of attempting to define the DNA binding domain of PCL that it was decided that purified GST protein did not represent an adequate negative control. In order to rectify this problem insoluble bacterial proteins were re-solubilised and added to purified GST protein to re-constitute a suitable negative control. It was then shown that this control extract (GST-extract) as well as extracts containing any of the fusion proteins used above were able to retard the mobility of the PRE in the same way as the extract containing GST-PCL. Furthermore it was shown that this retardation, as would be expected, was not modified by the addition of anti PCL-antibodies. Together these data strongly suggested that a bacterial protein or proteins were responsible for the DNA binding activity observed. This assessment is consistent with the observation that the addition of more bacterial extract caused greater retardation. The observation that purified 6H-PCL extract could bind to DNA but not as well as GST-PCL extract could be explained by this hypothesis also. In the purified 6H-PCL sample bacterial proteins were present at much lower levels than in the GST-PCL extracts (figure 7.4.1b). A few very faint bands representing contaminating bacterial proteins could just be visualised. Given the sensitivity of Coomassie staining (Harlow and Lane, 1988), almost 50 ng of the bacterial protein/s could have been present without being visualised in the gel shown in figure 7.4.1b. Given that the DNA binding activity was shown to be sequence specific it is not at all difficult to account for all the DNA binding activity by the presence of this small amount of bacterial protein. Sequence specific prokaryotic DNA binding proteins have very high affinities (in the order of 10^{-10}M) for their target DNAs.

What of the DNA binding activity of M96? Inouye *et al.*, (1994), used 250 ng of "partially purified GST-M96". The 250 ng apparently referred to the amount of the GST-M96 and not to the total protein used. In their mobility shift assays using this amount of protein, only a small proportion of the probe was bound. An estimate by eye was that in the order of 1/10-50 th of the probe was bound. This estimate is probably on the generous side and it is possible that an even smaller fraction of the probe is bound. In order for half the probe to be bound, therefore, approximately 2.5-7.5 µg of GST-M96

protein may have been required. Given this and the molecular weight of GST-M96 (~76 kD), the Kd for GST-M96 is between $1.5X^{10-5}$ and $3.3X^{10-6}$ M. This is extraordinarily low and, similarly for the Kd estimated for the 'PCL DNA binding activity', does not fit within the observed range of binding affinities for sequence specific DNA binding proteins (see above). Therefore it seems possible that the *in vitro* DNA binding activity that was ascribed to GST-M96 was also due to contaminating bacterial protein.

Questions are raised by this assertion. Firstly, why was no DNA binding seen with purified GST protein either in this study or by Inouye *et al.*, (1994), while binding was observed with purified 6H-PCL? It has been observed that in attempts to purify GST protein alongside GST fusions that the fusions will not purify as well (unpublished observations; Ausubel *et al.*, 1987). This is probably because of non-specific hydrophobic interactions between the non-GST moiety of the fusion protein and bacterial proteins. It seems plausible therefore that purified GST samples used in the experiments above and by Inouye *et al.*, (1994) were significantly more pure than the GST-M96 fusion or the 6H-PCL fusion. Inouye *et al.*, (1994) went one step further in that they showed an abolition of DNA binding activity when their fusion protein was expressed in the absence of zinc. They ascribed this to the inability of the 'zinc fingers' in M96 to form. There are a few ways to explain this observation. Firstly, the zinc fingers in M96 are known as PHD fingers (Cys4-His-Cys3, chapter 1). These are found in many nuclear proteins and have not as yet been shown to have DNA binding activity. In fact, given the similarity at the primary sequence level to RING fingers, it has been postulated that PHD fingers constitute protein-protein interaction domains. It is possible that M96 does bind to DNA and PCL does not because the DNA binding domain is not inclusive of the region of similarity between the two proteins. The folding of the whole M96 may be have been altered in the absence of zinc therefore disrupting this putative DNA binding domain. Alternatively, it is possible that the contaminating bacterial protein that is responsible for the DNA binding seen in this study and postulated to be responsible for the binding seen in the study of Inouye et al. is a zinc dependent protein.

How is it then that M96 was apparently able to bind to the MRE in the yeast one-hybrid system? GAL4-M96 was able to activate transcription of a HIS3 reporter gene downstream of 4 MRE d oligonucleotides, as well as a separate reporter construct that consisted of 4 MRE d oligonucleotides upstream of the LacZ gene. This activation was dependent on the presence of the MRE oligonucleotides. Unfortunately, none of the yeast data were shown. Given these controls, the most likely assumption is that GAL4-M96 was indeed able to activate transcription. It is possible that M96 was not able to bind to DNA independently, however, but required the assistance of a yeast nuclear protein. It must be noted that the GAL4-M96 clone as well as the GEX-M96 clone contained "an additional 99 amino acids that are not present in the authentic protein". This is because both the GAL4 and GST fusions occurred with M96 in its 5' untranslated region. Therefore, although highly unlikely, this 'artifactual' 99 amino acid peptide could have been responsible for the low affinity DNA binding activity that was seen. Formally, it is possible that GST-M96 does indeed bind to the MRE *in vitro* but it does so with a region that it does not have in common with PCL, for example the 'artifactual' 99 amino acid peptide. That a bacterial protein could also bind to the same MRE (this study) must then be viewed as a freak coincidence.

There are several experiments which would be required to test the suggestions that have been made.

-PCL fused to the GAL4 activation domain should be used identically in a yeast one-hybrid assay to determine if it can activate the identical MRE-reporter system employed by Inouye et al.

-GST-extract (see above) should be used in retardations on MRE containing oligonucleotides to test whether this protein extract does bind to the MRE oligonucleotide.

-Super-shift assays should be employed using anti M96-antibodies to determine if M96 is really responsible for the binding activity seen by Inouye *et al.*, (1994)

Why has the bacterial DNA binding activity reported here not been reported previously? Presumably any DNA binding assays using bona fide DNA binding proteins that were purified from bacterial extracts would only show activity of the protein of interest. In these experiments only small amounts of protein would be required to observe sequence specific DNA binding. This would result in the inadvertent use of less contaminating bacterial protein and therefore the bacterial protein binding activity described above would not be observable. rae-28, the murine ph homologue, has been observed to encode a low affinity sequence specific DNA binding activity (H. Brock, personal communication). This would not be expected given that ph does not have such an activity, and therefore may be explained by the above hypotheses.

In summary then, it appears that PCL does not bind to DNA and the DNA binding of M96 requires further investigation. This further assessment would be of great interest as the level of similarity between PCL and M96 suggests that M96 may perform a role similar to PCL.

Chapter 8 - Discussion

In chapter 1 the role of the Pc-G in the maintenance of repression of a wide range of genes was discussed. It was mentioned that the mechanism by which the Pc-G perform their silencing activity is still largely unknown, although, it appears that three non-exclusive mechanisms may be operating.

1) The Pc-G proteins may sterically block access of the transcriptional machinery
2) The Pc-G proteins may interfere specifically with transcriptional activators or with the basic transcriptional machinery
3) Pc-G repressed loci may associate by protein-protein interactions to segregate them from regions of the nucleus that contain transcriptional activators

Pcl was cloned in our laboratory, and its role in Pc-G function was unknown at the beginning of this study. *Pcl* is a critical member of the Pc-G and as such the identification of PCL protein interactors may provide powerful insights into the mechanism of PCL and therefore Pc-G function.

In this thesis it was shown that PCL interacts *in vitro* with a novel protein, termed 2.1, and a known Pc-G member, E(Z).

8.1 - Isolation of *2.1* and the biological significance of its interaction with PCL

In order to investigate the mechanism of Pc-G repression, a screen for interactors of PCL was undertaken. A single interactor was identified in a non-exhaustive *in vitro* screen. This protein, temporarily named 2.1, was found to be a chromosomal associated protein but different to Pc-G proteins in its chromosomal distribution. The distribution of 2.1 on polytene chromosomes was analysed, and its interaction with other Pc-G proteins was tested. Using far-western or yeast two-hybrid interaction assays, it was shown that PCL, 2.1 and E(Z) (see section 8.2) all co-interact. However, given inconsistencies in detecting these interactions using these two systems, the biological significance of these interactions is unknown until *in vivo* interactions can be demonstrated.

8.2 - Interactions between PCL, 2.1 and E(Z)

In chapter 4, data which showed that PCL and 2.1 interact independently with E(Z) were described. As the name suggests, mutations in *E(z)* enhance the *zeste-white* effect. The *zeste-white* effect is a pairing dependent repression of a wild-type *w* gene mediated by a ZESTE protein encoded by the neomorphic *zeste* (*z*) allele, z^1 (reviewed in Pirrotta, 1991). That E(Z) is involved in a general chromosomal maintenance role such as chromosome pairing is supported by the observation that E(Z) may be associated almost ubiquitously with polytene chromosomes as well as at a higher level with approximately 44 sites that overlap with PC/PH/PCL sites (R. Jones, personal communication).

However, recent published data suggest that the ubiquitous E(Z) immunostaining on polytene chromosomes may be background staining (Carrington and Jones, 1996). Other evidence for a general chromosomal role is that larvae homozygous for a temperature sensitive allele of *E(z)* grown at the restrictive temperature, have decondensed polytene chromosomes with almost complete abolition of PSC and SU(Z)2 chromosomal localisation (Rastelli *et al.*, 1993). This apparent dependence on E(Z) for Pc-G proteins to be bound to chromosomes could be mediated by the interaction between PCL and E(Z). Overexpression of the domain that interacts with E(Z) may titrate binding of endogenous PCL to E(Z). Loss of binding of other Pc-G proteins to chromosomes in this experiment would strongly support the above hypothesis. PCL and TRX each have PHD fingers and may compete for protein partners. The results of chapter 4 suggest that the PHD fingers of PCL may be responsible for the interaction with E(Z). This raises the possibility that TRX and PCL may compete for binding to E(Z). Given the role for E(Z) mentioned above, it may be that this competition determines whether or not the Pc-G or trx-G becomes associated with a gene. This model may be too simplistic - TRX and PC (and therefore PCL) have been found at multiple overlapping chromosomal sites (Chinwalla *et al.*, 1995). However, the resolution of polytene immuno - fluorescence is too low to say whether these proteins actually occupy the same DNA site.

The interaction between 2.1 and E(Z) is interesting in the light of the data presented in this thesis that 2.1 appears not to be a general Pc-G member. This interaction, given the widespread distribution of 2.1, could further support the idea that E(Z) has a general chromosomal structural role. Conversely, given that E(Z) is proposed to play a general chromosomal structural role, it could be that 2.1, with its distribution pattern on polytene chromosomes that may be described as widespread, has a similar general chromosomal structural role. Such a general role could be tested once *2.1* mutant animals were generated. In particular, if conditional *2.1* mutants could be generated, polytene chromosomes from larvae raised in restrictive conditions could be analysed to determine if they had altered morphology and/or altered binding of Pc-G and trx-G proteins. Further to this, effects of *2.1* mutations on transvection and the zeste-white interaction, both sensitive assays of chromosomal pairing upon which *E(z)* mutations have an effect, could be studied.

8.3 - Possible roles for 2.1

Although the data presented in this thesis show that 2.1 is a novel chromosomal protein, its function is yet to be determined. Some equally plausible hypotheses pertaining to its function are described below.

8.3.1 - 2.1 as a boundary element associated factor

In chapter 6, the distribution of 2.1 protein in *Drosophila* was described. Most interesting were the data generated on the distribution of 2.1 on polytene chromosomes. Firstly, 2.1 was found predominantly in interbands intimately flanking sites of PCL localisation. Given that two other *Drosophila* boundary element associated proteins, BEAF - 32 (Zhao *et al.*, 1995) and BANGDOO (Gaszner *et al.*, 1996), are

associated with interbands and that Pc-G repression is probably prevented from spreading ectopically by boundary elements (see section 1.5.10), the possibility arises that 2.1 is a boundary element associated factor. This possibility is particularly exciting as it may provide the first tangible link between Pc-G silencing and the prevention of repression spreading by boundary elements. There are several ways in which this proposed role for 2.1 could be tested. It would be expected that the loss of *2.1* function would result in the spurious shutdown of genes or enhancers that neighbour a gene or enhancer that is normally under Pc-G repression. Once a *2.1* mutant is generated, analysis of homeotic gene expression in *2.1* mutants should be carried out. It would be expected that homeotic expression may be completely repressed late in embryogenesis after the perdurance of any maternal 2.1. For example, repression of *Ubx* transcription outside of PS 5 and 13 may spread to include repression of the *cis* - regulatory regions that normally activate *Ubx* between PS 5 and 13. Similarly, other homeotic genes might be stochastically, ectopically repressed, in a manner similar to the repression of the *w* gene in transgenes that contain a PRE. In this situation, *2.1* mutant phenotypes would be similar to loss of function homeotic phenotypes (late, due to perdurance) of the BX - C and ANT - C.

It would also be interesting to examine if repression could spread to shutdown genes that normally are never susceptible to Pc-G repression. For this it would be necessary to identify genes that flank a Pc-G repressed locus, and analyse their expression patterns in *2.1* mutants. If a gene that is normally never susceptible to Pc-G silencing, but flanks a Pc-G silenced locus, is silenced in a *2.1* mutant, then it would be likely that this silencing is due to ectopic Pc-G repression. This kind of experiment should be coupled with polytene immuno-staining to determine if there is any visible spread of PC/PCL/PH binding in *2.1* mutants. Of course, it would be necessary to show that in *2.1*/Pc-G double mutants that the ectopic silencing is abolished.

Another test for a boundary element effect would be to perform formaldehyde crosslinking of chromatin followed by immunoprecipitation with either 2.1 or PCL antibodies in a manner similar to Orlando and Paro (1993). It would be expected that while PCL is found associated with large tracts of DNA of repressed BX - C genes, 2.1 is associated with the boundary elements flanking an active gene (e.g. *Abd-B* in SL - 2 cells).

A further way to test for a role for 2.1 in boundary element function is to analyse the effect *2.1* mutations have in enhancer - blocking assays (Kellum and Schedl, 1992). In these assays, putative boundary element DNA fragments are interposed between enhancers and reporter genes and their effect on the enhancer's ability to activate reporter gene transcription assayed. Given that 2.1 appears to interact with PCL and E(Z), it seems likely that if 2.1 is a boundary element associated protein that it would be associated with boundary elements within the homeotic gene complexes such as MCP and FAB-7. The function of these boundary elements as well other known elements- scs, scs' and gypsy transposon (reviewed in Eissenberg and Elgin, 1991) could be assayed in *2.1* mutants. If *2.1* mutations have no inhibitory effect on the insulators in these assays, it may be for a couple of reasons. Firstly, 2.1 may not act through any of these particular boundary elements. Secondly, 2.1 may only be involved in blocking repression and not activation. To test this idea, repressor blocking assays could be constructed. In this case, boundary elements could be interposed between PREs and a reporter

gene (for example *w*). In the wild - type case, it would be expected that *w* transcription is not repressed. However, in *2.1* mutants, it would be expected that the *w* gene would be stochastically repressed as if there were no boundary element present (as seen for the *w* gene adjacent adjacent to PRE-containing fragments). Zink and Paro (1995) used FAB - 7 and MCP fragments as PREs in their transgenes (see chapter 1). In these transgenes, *w* is not completely insulated from repression even though there is a boundary element within the MCP and FAB - 7 fragments (Galloni *et al.*, 1993). However, it was noted that FAB - 7 was able to effect Pc-G repression more strongly when it was placed with it's proximal end towards the lac Z gene (Zink and Paro, 1995). Perhaps this indicates that the boundary element exists towards the distal end of FAB - 7 fragment, but that this element is not sufficient to completely block silencing in this artifactual situation. In this experiment the FAB - 7 fragment used is a 3.7 kb Hind III fragment from position +122 - + 126. When the nuclease sensitive sites within FAB - 7 were mapped, they were positioned from ~ 3 kb - 3.4 kb distal to a Sac II site at +121 (Galloni *et al.*, 1993). This indicates that these sites are 2 - 2.4 kb from the proximal end of the 3.7 kb FAB - 7 fragment used by Zink and Paro (1995), and therefore towards the distal end of this fragment as predicted. What this analysis also suggests is that the PRE is within the proximal 2 - 2.4 kb of the FAB - 7 fragment. The loss of polarity of repression blocking of this FAB - 7 fragment could be assayed in *2.1* mutants.

A model for 2.1 prevention of Pc-G repression spreading in the BX - C, at the parasegment level, is shown below.

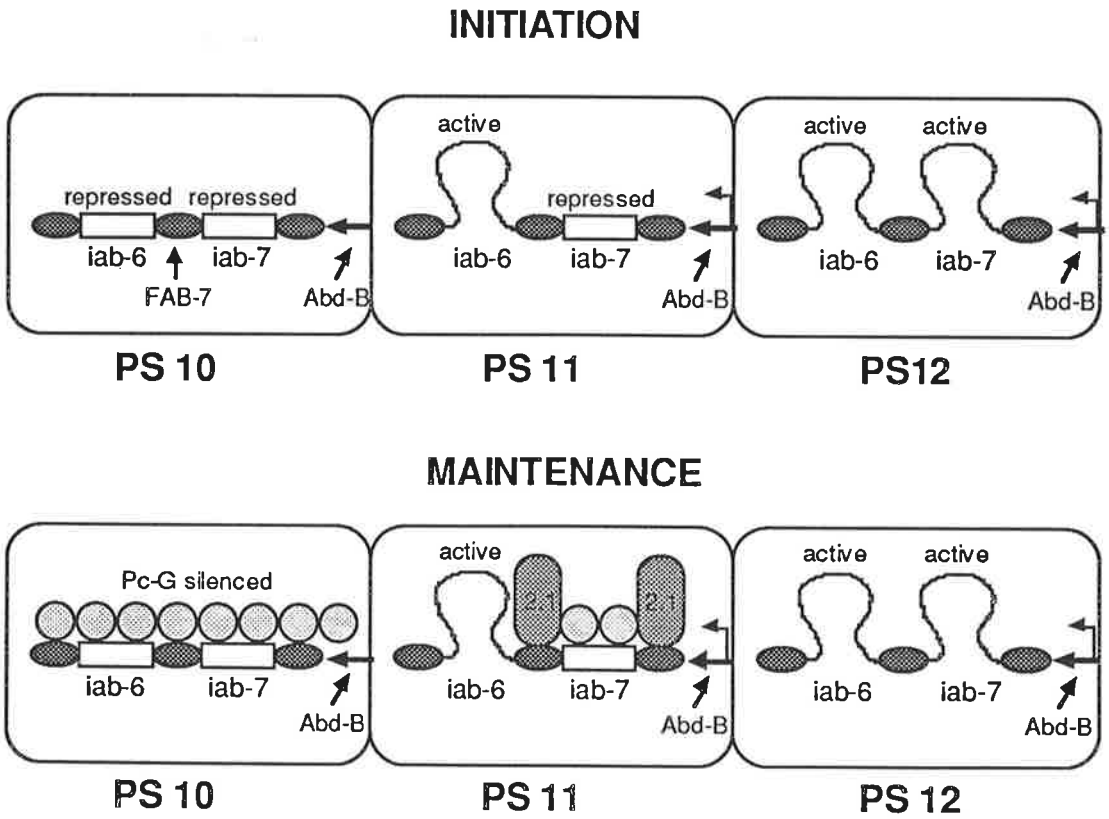


figure 8.3.1.1 - model for 2.1 prevention of Pc-G spreading in the BX - C

The figure (adapted from one in Galloni *et al.*, 1993) shows the state of activity of the *cis* - regulatory regions iab - 6 and 7 during the initiation and maintenance phases, in parasegments 10, 11 and 12. In PS 10, neither element is active therefore they become repressed in the maintenance phase. Pc-G proteins (grey discs) are associated with these elements and the *Abd-B* coding region. In PS 11, iab - 6 is active, but iab - 7 is not. Pc-G proteins become associated with iab - 7 but not iab - 6 and the *Abd-B* coding region. 2.1 (dark grey oval) which binds to the boundary elements (black ovals), like FAB - 7, prevents Pc-G repression from spreading. Note that iab - 6 must be able to access the *Abd-B* coding region even in the presence of 2.1 at FAB - 7 and a putative FAB - 8. This indicates that a 2.1 mediated boundary element may be specific to containing Pc-G repressed regions. It may be that when a silenced iab - 7 is recruited to a Pc-G silenced nuclear region the regions outside of the 2.1 binding sites (i.e. iab - 6 and *Abd-B*) can 'loop' out to access each other and transcriptional machinery.

8.3.2 - 2.1 as a component of transcriptionally active chromatin

The evidence that 2.1 is a boundary element associated protein is only circumstantial. Given that interbands are regions of transcriptionally active chromatin (Rykowski *et al.*, 1988), it may be that 2.1 is associated with active chromatin, rather than just the boundaries between active and inactive chromatin, performing some general structural role (e.g. keeping the chromatin in its open configuration). In this model the Pc-G, via PCL, would be proposed merely to exclude 2.1 from band regions by interacting with it at the margins between bands and interbands. One could invoke a model of dynamic antagonism between PCL and 2.1 in their competition to either repress or activate a locus. Just as PCL is not associated with all repressed chromatin (bands) (see figure 6.4.2c), 2.1 does not appear to associated with all active chromatin (interbands). There are interband regions and fractions of interbands that do not appear to contain 2.1 protein (figure 6.4.3). Experiments which test the possibility that 2.1 is an essential general structural component of active chromatin would be essentially the same as those described in section 8.3.1. Polytene chromosomes from 2.1 mutant larvae could be analysed to determine, at the cytological level, if interband regions had degenerated (i.e. that band regions had expanded). At the molecular level, experiments which analyse the levels of transcription of various genes in 2.1 mutants would be similar to those described below.

It is also possible, given the interband deposition of 2.1, that 2.1 is a general transcriptional activator protein that could be categorised, for example, with trx-G proteins. As examples, 2.1 may act to disrupt nucleosomal structure (as alluded to above), may be a component of the RNA polymerase II holoenzyme complex, or be some other type of global transcriptional activator protein and therefore act to promote and maintain transcriptional activation of a large range of genes including the homeotic genes. It must be noted then, that an interaction between PCL and a 2.1, a proposed activator protein, appears incongruous. It is possible that the PCL/2.1 interaction *in vitro* is purely artifactual and there is no *in vivo* interaction between these proteins. Alternatively, there are at least two other possibilities. PCL and 2.1 may interact and compete for loci to repress or activate respectively and incidentally form an interface as visualised on polytene chromosomes, or it is possible that PCL and

2.1 do not interact on polytene chromosomes but interact in other tissues at other developmental stages. One possibility here is that they may interact earlier in development, possibly in an antagonistic way to determine whether a gene will enter a repressed or active maintenance phase. To test the hypothesis that 2.1 is a transcriptional activator, *2.1* mutant flies would need to be generated. It would then be possible to construct expression assays to determine whether or not expression of homeotic genes, for example, was reduced in *2.1* mutants.

It must be noted that reduced or abolished transcription of genes such as the homeotics is the predicted outcome if 2.1 is a boundary element factor that abates the spread of Pc-G repression or if 2.1 is a general structural or transcriptional factor in active chromatin. In fact, experimentally, it would be very difficult to distinguish between these possibilities. One experiment that may distinguish between these two models is as follows. From genetic studies (see chapter 1) and given that TRX and PC appear to bind to many of the same sites on polytene chromosomes, it seems that genes that are susceptible to Pc-G silencing and trx-G activation are at least an overlapping set. Therefore, observing transcriptional loss of homeotic genes, for example, in 2.1 mutant animals would not distinguish between whether there was a loss of transcriptional activation (i.e. 2.1 is a trx-G-type transcriptional activator) or a gain of ectopic Pc-G silencing (i.e. 2.1 is a boundary element associated factor that normally prevents the spread of Pc-G repression). One way to overcome this would be to analyse the transcriptional states of a gene that is not normally under the control of Pc-G and trx-G and is either close by or distant to a gene that is normally under Pc-G/trx-G control in a 2.1 mutant versus wild type genetic background. If a gene that is not normally under Pc-G/trx-G control and flanks a Pc-G/trx-G controlled gene is inactivated in a 2.1 mutant background, it would be possible to say that 2.1 is most likely not a trx-G-type activator and is either a boundary element factor (i.e. Pc-G repression has spread from the flanking gene) or some other type (non trx-G) of general transcriptional activator (i.e. loss of transcriptional activating function). If a gene that is not normally under Pc-G/trx-G control and is not close to a gene that is normally under Pc-G/trx-G control is inactivated in a 2.1 mutant background, then it would be possible to exclude the possibility that Pc-G repression had 'spread ' to repress this gene. It may then be possible to conclude that 2.1 is most likely a general transcriptional activator. Of course these experiments would need to be combined with immunostaining experiments to determine if there was any cytologically visible spreading of Pc-G proteins on polytene chromosomes.

Another possibility that has not been discussed is what would happen if 2.1 were a boundary element protein that prevents Pc-G repression from 'spreading' into an active region and concomitantly prevents activation from 'spreading' into a Pc-G repressed region? In this case if 2.1 function was compromised, a metastable state would be generated. In some cells repression would spread, in some cells activation would spread and in some cells there may be no effect. This possibility must be borne in mind as a potential difficulty in screening phenotypically for *2.1* mutants.

8.3.3 - 2.1 at the chromocentre

The two hypotheses described above pertain to the interband deposition of 2.1. However, 2.1, unlike Pc-G proteins and BEAF-32 (Zhao *et al.*, 1995), is present at a discrete position within the chromocentre of polytene chromosomes.

As was described in chapter 1, the similarities between Pc-G silencing and silencing in PEV suggest that there is overlap in the mechanisms involved. It was suggested that Pc-G repressed loci may cluster to isolate these silenced genes. Further, given the relationship to PEV silencing, it was suggested that Pc-G clustered sites may also interact with some heterochromatin regions. Given the observation that PCL and 2.1 can interact *in vitro*, it is possible that the presence of 2.1 at the chromocentre (heterochromatin) (see chapter 6) facilitates such a proposed interaction. This would be very difficult to test. 3 - D microscopy of nuclei stained with anti - 2.1 antibodies would most likely not work given the number of 2.1 binding sites (Dernburg *et al.*, 1996). Given that the region of heterochromatin with which 2.1 is interacting is unknown and is difficult to identify (due to the homogenous appearance of the chromocentre - see chapter 6), at present it would not be possible to mark this part of the chromatin by *in situ* hybridisation to attempt 3 - D microscopy. It may be possible to perform triple immunofluorescence on whole nuclei. Antibodies could be used to simultaneously mark the chromocentre (e.g. HP1), Pc-G binding sites (e.g. PCL) and 2.1 sites. The level of detection of 2.1 could be reduced so that only the chromocentre signal is visible (see chapter 6). Then, analysis of the frequency with which Pc-G sites come into association with the 2.1 site at the chromocentre could be undertaken.

8.3.4 - tissue specific functions of 2.1

If 2.1 is involved in transcriptional activation, it is possible that the higher concentration of 2.1 in the mesoderm early in embryogenesis and in stripes late in embryogenesis is required for a higher level of transcriptional activation in these tissues at these stages of embryogenesis.

It is possible that the striped distribution of 2.1 has a specific role in generating sub-parasegment specific boundaries of repression within the homeotic gene complexes. For example, it may be that a higher level of 2.1 protein is required to establish a functional boundary element and interact with PCL within a portion of a parasegment at the appropriate inter *cis* - regulatory region boundary element site. Although each *cis* - regulatory element is responsible for transcription in a parasegment, there is evidence that transcription of homeotic genes is non-uniform within a parasegment For example *Ubx* is expressed only in the abdominal part of PS 6 at stage 9 of embryogenesis, whereas in stage 10 expression extends more anteriorly (Castelli-Gair and Akam, 1995). Coincidentally, the stripes of 2.1 are present towards the anterior of parasegments (one stripe posterior to *engrailed* which defines the absolute anterior).

8.4 - 2.1 further in the future

Given that the role of 2.1 is unclear, screening for 2.1- protein partners may be particularly informative about the function of 2.1. If 2.1 was shown to interact with known boundary element proteins or proteins involved in transcriptional activation, then it would be possible to predict more confidently that 2.1 was either a boundary element or transcriptional activating protein respectively.

8.5 - General Pc-G issues

Many questions still remain to be asked about the nature of Pc-G repression.

- How does the Pc-G complex assemble and what are its characteristics?

- How does the Pc-G interact with a PRE?

- How does the Pc-G become associated with larger tracts of DNA outside of the PRE?

- What prevents the Pc-G complexes from 'spreading' into neighbouring genes or *cis* - regulatory elements inappropriately?

- Do Pc-G repressed loci cluster and if so what interactions are involved?

- How does the Pc-G sterically or specifically block access of transcriptional machinery?

The possibility that the second question could be answered was raised by the discovery that PCL shares significant sequence similarity to a reported DNA binding protein.

8.5.1 - Does PCL bind to DNA?

In chapter 7 the possibility that PCL may bind directly to DNA because of its similarity to an apparent DNA binding protein, M96, was raised. This possibility was very exciting as it potentially provided an answer to the second question above. It was shown, however, that PCL in fact was not able to bind to PRE sequences independently *in vitro*. Furthermore, the possibility that M96 also did not bind independently to DNA *in vitro* was put forward. Although the published data regarding the DNA binding activity of M96 *in vitro* were questioned, it seems clear that M96 is able to bind to DNA in a yeast one - hybrid system. Testing whether PCL can bind to DNA in this system is underway (H. Weckert, personal communication). Given the high degree of sequence similarity between PCL and M96, it seems quite possible that PCL will also bind to DNA in the one-hybrid system. It may be that M96 is able to interact with endogenous yeast proteins to enable it to bind directly or indirectly to DNA. It is possible therefore that similar interactions will occur with PCL. Although Pc-G homologues have not been identified in yeast by sequence similarity alone, it is possible that there are structural homologues that can interact with PCL to allow it to interact with DNA. Alternatively, interactions with

M96, and potentially PCL, resulting in DNA binding may be completely spurious and not reflect any functional similarities between the proteins involved. The *in vitro* data presented here make it unlikely that PCL would be able to bind to DNA independently. In *Drosophila*, Pc-G binding to DNA must be modulated; Pc-G is present in all cells but not at all PREs. If PCL could bind to DNA independently, this selective Pc-G repression may not be able to occur. If PCL can bind to DNA at all, it probably requires interactions with other proteins in the Pc-G and possibly ESC. UV-cross-linking of *Drosophila* chromatin could be used to determine if PCL is actually associated with DNA directly *in vivo*.

Given the high degree of similarity between PCL and M96 outside of the PHD fingers, it is quite possible that M96 represents a murine homologue of PCL. Experiments which attempt to rescue *Pcl* mutant phenotypes with an M96 expressing transgene would need to be attempted to test this. This kind of experiment has been successful with *Pc* and its murine homologue M33 (Müller *et al.*, 1995). Furthermore it would be of interest to determine if any genetic diseases are associated with mutations in the human homologue of M96. Two different human genes with high similarity to M96 have been isolated for this purpose (M. Coulson, personal communication).

8.5.2 - Other Pc-G questions

Some of the questions posed above could be answered with relatively straightforward experimental approaches. For example, it would be informative to screen for interactors of ESC. Given that ESC potentially forms the bridge between transiently repressed and Pc-G silenced genes, ESC interactors might be proteins that are associated with repressed chromatin and Pc-G proteins respectively.

3 - D microscopy on Pc-G immunofluorescing nuclei should be carried out to determine if there is any clustering of Pc-G loci, a possibility discussed in this thesis. If there is such clustering, it would be important to determine a) which protein interactions are responsible for the clustering and b) is clustering sufficient for heritable silencing. To answer the first question, it would be necessary to look at the effect on clustering of different Pc-G mutations. These data would have to be combined with information on which Pc-G proteins interact with each other. The second question would be answered if any Pc-G mutants were found that did not affect clustering (as all Pc-G mutants, by definition affect silencing).

Another interesting problem that has not been raised previously in this thesis is how does the Pc-G maintain silencing through the presumably disruptive process of DNA replication. It has been shown that the Pc-G is required for homeotic gene silencing even in the absence cell division or DNA replication (Gould *et al.*, 1990), suggesting that the role of the Pc-G is not merely to stabilise the silencing of genes during replication. Given that the transcriptionally active chromatin is 're-nucleosomed' immediately after replication, requiring further nucleosome disruption to re-stimulate transcription (Lucchini and Sogo, 1995), it seems likely that Pc-G repression may have to be reset after each replication event, prior to transcriptional activators spuriously activating those genes. That this occurs and then how it occurs is completely unknown.

Answers to these questions and those presented throughout this thesis, will enable a full understanding of one epigenetic state maintenance system.

References

Aasland, R., T.J. Gibson and A.F. Stewart. (1995) "The PHD finger: implications for chromatin-mediated transcriptional regulation." *Trends in Biochemical Science* **20**: 56-59.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman. (1990) "Basic Local Alignment Search Tool." *J. Mol. Biol.* **215**: 403-410.

Arnim, A.G.v. and X.-W. Deng. (1993) "Ring Finger Motif of *Arabidopsis thaliana* COP1 Defines a New Class of Zinc-binding Domain." *The Journal of Biological Chemistry* 268.26: 19626-19631.

Ausubel, S.F., R. Brent, R.E. Kingston, D. Moore, J.A. Smith, J.G. Seidman and K. Struhl. Current Protocols in Molecular Biology. New York: Wiley, 1987.

Baker, W.K. and A. Rein. (1962) "The dichotomous action of Y chromosomes on the expression of position-effect variegation." *Genetics* 47: 1399-1407.

Baksi, K., H. Alkhatib and M.E. Smulson. (1987) "In vivo characterization of the poly(ADP-ribosylation) of SV40 chromatin and large T antigen by immunofractionation." *Exp Cell Res* 172.1: 110-23.

Bell, S., R. Kobayashi and B. Stillman. (1993) "Yeast Origin Replication Complex functions in transcription silencing and DNA replication." *Science* 262: 1844-1848.

Belyaeva, E.S., O.V. Demakova, G.H. Umbetova and I.F. Zhimulev. (1993) "Cytogenetic and molecular aspects of position-effect variegation in *Drosophila melanogaster*. V. Heterochromatin-associated protein HP1 appears in euchromatic chromosomal regions that are inactivated as a result of position-effect variegation." *Chromosoma* 102.8: 583-90.

Belyaeva, E.S. and I.F. Zhimulev. (1991) "Cytogenetic and molecular aspects of position effect variegation in *Drosophila*. III. Continuous and discontinuous compaction of chromosomal material as a result of position effect variegation." *Chromosoma* 100.7: 453-66.

Borden, K.L., M.N. Boddy, J. Lally, N.J. O'Reilly, S. Martin, K. Howe, E. Solomon and P.S. Freemont. (1995) "The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML." *Embo J* 14.7: 1532-41.

Bornemann, D., E. Miller and J. Simon. (1996) "The *Drosophila Polycomb* group gene *Sex comb on midleg* (*Scm*) encodes a zinc finger protein with similarity to polyhomeotic protein." *Development* 122: 1621-1630.

Breen, T.R. and I.M. Duncan. (1986) "Maternal expression of genes that regulate the bithorax complex of *Drosophila melanogaster*." *Dev Biol* 118.2: 442-56.

Brown, N.H. and F.C. Kafatos. (1988) "Functional cDNA Libraries from *Drosophila* Embryos." *J. Mol. Biol.* 203: 425-437.

Buchman, A., W. Kimmerly, J. Rine and R. Kornberg. (1988) "Two DNA-binding factors recognise specific sequences at silencers, upstream activating sequences, autonomously replicating sequneces and telomeres in *S. cerevisiae*." *Molecular and Cellular Biology* 8: 210-225.

Busturia, A. and M. Bienz. (1993) "Silencers in *Abdominal-B*, a homeotic *Drosophila* gene." *The EMBO Journal* 12.4: 1415-1423.

Campbell, R.B., D.A.R. Sinclair, M. Couling and H.W. Brock. (1995) "Genetic interactions and dosage effects of *Polycomb* group genes of *Drosophila*." *Molecular and General Genetics* 246: 291-300.

Carrington, E.A. and R.S. Jones. (1996) "The drosophila enhancer of zeste gene encodes a chromosomal protein - examination of wild-type and mutant protein distribution." *Development* 122.12: 4073-4083.

Castelli-Gair, J. and M. Akam. (1995) "How the Hox gene Ultrabithorax specifies two different segments: the significance of spatial and temporal regulation within metameres." *Development* 121.9: 2973-2982.

Cavener, D.R. (1987) "Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates." *Nucleic Acids Research* 15: 1353-1361.

Chan, C.-S., L. Rastelli and V. Pirrotta. (1994) "A *Polycomb* response element in the *Ubx* gene that determines an epigenetically inherited state of repression." *EMBO J.* 13.11: 2553-2565.

Cheng, N.N., D.A.R. Sinclair, R.B. Campbell and H.W. Brock. (1994) "Interactions of *polyhomeotic* with *Polycomb* group genes of *Drosophila melanogaster*." *Genetics* 138: 1151-1162.

Chiang, A., M.B. O'Connor, R. Paro, J. Simon and W. Bender. (1995) "Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex." *Development* 121: 1681-1689.

Chinwalla, V., E.P. Jane and P.J. Harte. (1995) "The *Drosophila trithorax* protein binds to specific chromsomal sites and is co-localized with *Polycomb* at many sites." *EMBO J.* 14.9: 2056-2065.

Chouinard, S. and T.C. Kaufman. (1991) "Control of expression of the homeotic labial (lab) locus of *Drosophila melanogaster*: evidence for both positive and negative autogenous regulation." *Development* 113.4: 1267-80.

Côté, J., J. Quinn, J.L. Workman and C.L. Peterson. (1994) "Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex." *Science* 265: 53-60.

Cribbs, D.L., C. Benassayag, F.M. Randazzo and T.C. Kaufman. (1995) "Levels of homeotic protein function can determine developmental identity: evidence from low-level expression of the *Drosophila* homeotic gene proboscipedia under Hsp70 control." *Embo J* 14.4: 767-78.

Csink, A.K. and S. Henikoff. (1996) "Genetic modification of heterochromatic association and nuclear organization in *Drosophila*." *Nature* 381.6582: 529-31.

DeCamillis, M., N. Cheng, D. Pierre and H.W. Brock. (1992) "The *polyhomeotic* gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome-binding sites with *Polycomb*." *Genes and Development* 6: 223-232.

Dernburg, A.F., K.W. Broman, J.C. Fung, W.F. Marshall, J. Philips, D.A. Agard and J.W. Sedat. (1996) "Perturbation of nuclear architecture by long-distance chromosome interactions." *Cell* 85.5: 745-59.

Devlin, R.H., B. Bingham and B.T. Wakimoto. (1990) "The Organisation and Expression of the *light* Gene, a heterochromatic gene of *Drosophila melanogaster*." *Genetics* 125: 129-140.

Dingwall, A.K., S.J. Beek, C.M. McCallum, J.W. Tamkun, G.J. Kalpana, S.P. Goff and M.P. Scott. (1995) "The *Drosophila* snr1 and brm proteins are related to Yeast SWI/SNF proteins and are components of a large protein complex." *Molecular Biology of the Cell* 6: 777-791.

Docquier, F., O. Saget, F. Forquignon, N.B. Randscholt and P. Santamaria. (1996) "The *multi sex combs* gene of *Drosophila melanogaster* is required for proliferation of the germline." *Roux's Archives of Developmental Biology* 205: 203-214.

Dorer, D.R. and S. Henikoff. (1994) "Expansions of transgene repeats cause heterochromatin formation and gene silencing." *Cell* 77: 993-1002.

Dreesen, T.D., S. Henikoff and K. Loughney. (1991) "A pairing-sensitive element that mediates trans-inactivation is associated with the *Drosophila* brown gene." *Genes Dev* 5.3: 331-40.

Duncan, I.M. (1982) "Polycomblike: a gene that appears to be required for the normal expression of the bithorax and Antennapedua gene complexes of *Drosophila melanogaster*." *Genetics* 102: 49-70.

Dura, J.-M., N.B. Randsholt, J. Deatrick, I. Erk, P. Santamaria, J.D. Freeman, S.J. Freeman, D. Weddel and H.W. Brock. (1987) "A complex locus, *polyhomeotic*, is required for segmental specification and epidermal development in D. *melanogaster*." *Cell* .51: 829-839.

Durkacz, B.W., J. Lunec, H. Grindley, S. Griffin, O. Horner and A. Simm. (1992) "Murine melanoma cell differentiation and melanogenesis induced by poly(ADP-ribose) polymerase inhibitors." *Exp Cell Res* 202.2: 287-91.

Eissenberg, J.C. and S.C.R. Elgin. (1991) "Boundary functions in the control of gene expression." *Trends in Genetics* 7.10: 335-340.

Farkas, G., J. Gausz, M. Galloni, G. Reuter, H. Gyurkovics and F. Karch. (1994) "The Trithorax-like gene encodes the *Drosophila* GAGA factor." *Nature* 371.6500: 806-8.

Fauvarque, M.-O. and J.-M. Dura. (1993) "*polyhomeotic* regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in *Drosophila*." *Genes and Development* 7: 1508-1520.

Feinberg, A.P. and B. Vogelstein. (1983) "A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity." *Anal. Biochem.* 132: 6-13.

Franke, A., M. DeCamillis, D. Zink, N. Cheng, H.W. Brock and R. Paro. (1992) "*Polycomb* and *polyhomeotic* are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*." *EMBO J.* 11.8: 2941-2950.

Galloni, M., H. Gyurkovics, P. Schedl and F. Karch. (1993) "The bluetail transposon: evidence for independent cis-regulatory domains and domain boundaries in the bithorax complex." *Embo J* 12.3: 1087-97.

Gaszner, M., J. Vazquez and P. Schedl. (1996) "Characterisation of the scs element" In *37th Annual Drosophila Research Conference* San Diego, California

Gelbart, W., W. Rindone, J. Chillemi, S. Russo, J.P. Lenart, D. Emmert, B. Matthews, L. Crosby, M. Ashburner, R. Drysdale, E. Whitfield, G. Millburn, A. de Grey, T. Kaufman, K. Matthews, D. Gilbert, V. Strelets and C. Tolstoshev. (Ongoing) "Flybase."

Gindhart, J.G. and T.C. Kaufman. (1995) "Identification of *Polycomb* and *trithorax* group repsonsive elements in the regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*." *Genetics* 139: 797-814.

Girton, J.R. and S.H. Jeon. (1994) "Novel embryonic and adult homeotic phenotypes are produced by pleiohomeotic mutations in *Drosophila*." *Dev Biol* 161.2: 393-407.

Gonzalez-Reyes, A., N. Urquia, W. Gehring, G. Struhl and G. Morata. (1990) "Are cross regulatory interactions between homeotic genes functionally significant?" *Nature* 344: 78-80.

Gottschling, D.E., O.M. Aparicio, B.L. Billington and V.A. Zakian. (1990) "Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription." *Cell* 63.4: 751-62.

Gould, A.P., R.Y. Lai, M.J. Green and R.A. White. (1990) "Blocking cell division does not remove the requirement for Polycomb function in *Drosophila* embryogenesis." *Development* 110.4: 1319-25.

Grigliatti, T. (1991) "Position-effect variegation--an assay for nonhistone chromosomal proteins and chromatin assembly and modifying factors." *Methods Cell Biol* 35.587: 587-627.

Gutjahr, T., E. Frei, C. Spicer, S. Baumgartner, R.A.H. White and M. Noll. (1995) "The Polycomb-group gene, *extra sex combs*, encodes a nuclear member of the WD-40 repeat family." *EMBO J.* 14.17: 4296-4306.

Gyuris, J., E. Golemis, H. Chertkov and R. Brent. (1993) "Cdi1, a Human G1 and S Phase Protein Phosphatase That Associates with Cdk2." *Cell* 75.4: 791-803.

Hanahan, D. (1983) "Studies on transformation of E. coli with plasmids." *Journal of Molecular Biology* 166: 577-580.

Harlow, E. and D. Lane. 1st ed. New York: Cold Spring Harbor, 1988.

Hayashi, S., A. Ruddell, D. Sinclair and T. Grigliatti. (1990) "Chromosomal structure is altered by mutations that suppress or enhance position effect variegation." *Chromosoma* 99: 391-400.

Hecht, A., T. Laroche, S. Strahl-Bolsinger, S.M. Gasser and M. Grunstein. (1995) "Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast." *Cell* 80: 583-592.

Henderson, R.M., S. Schneider, Q.A. Li, D. Hornby, S.J. White and H. Oberleithner. (1996) "Imaging romk1 inwardly rectifying atp-sensitive k+ channel protein using atomic force microscopy." *Proceedings of the National Academy of Sciences of the United States of America* 93.16: 8756-8760.

Henikoff, S. (1996) "Dosage-dependent modification of position-effect variegation in drosophila." *Bioessays* 18.5: 401-409.

Hirschhorn, J.N., S.A. Brown, C.D. Clark and F. Winston. (1992) "Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure." *Genes and Development* 6: 2288-2298.

Ingham, P. and R. Whittle. (1980) "*trithorax* a new homeotic mutation of *Drosophila melanogaster* causing transformations of abdominal and thoracic imaginal segments. I. Putative role during embryogenesis." *Molecular and General Genetics* 179: 607-614.

Ingham, P.W. (1984) "A gene that regulates the bithorax complex differentially in larval and adult cells of *Drosophila*." *Cell* 37.3: 815-23.

Inouye, C., P. Remondelli, M. Karin and S. Elledge. (1994) "Isolation of a cDNA encoding a metal response element binding protein using a novel expression cloning procedure: the one hybrid system." *DNA and Cell Biology* 13.7: 731-742.

Jones, R.S. and W.M. Gelbart. (1990) "Genetic analysis of the *Enhancer of zeste* locus and its role in gene regulation in *Drosophila melanogaster*." *Genetics* 126: 185-199.

Jones, R.S. and W.M. Gelbart. (1993) "The *Drosophila* Polycomb-Group gene *Enhancer of zeste* Contains a Region of Sequence similarity to *trithorax*." *Molecular and Cellular Biology* 13.10: 6357-6366.

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

Jurgens, G., E. Wieschaus, C. Nusslein-Volhard and H. Kluding. (1984) "Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the third chromosome." *Willhelm Roux's Arch. Dev. Biol.* 193: 283-295.

Karpen, G.H. (1994) "Position-effect variegation and the new biology of heterochromatin." *Current Opinion in Genetics and Development* 4: 281-291.

Kassis, J.A., E. Noll, E.P. Van Sickle, W.F. Odenwald and N. Perrimon. (1992) "Altering the insertional specificity of a *Drosophila* transposable element." *Proceedings National Academy of Sciences (USA)* 89: 1919-1923.

Kaufman, T., R. Lewis and B. Wakimoto. (1980) "Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homeotic gene complex in polytene chromosome interval 84A-B." *Genetics* 94: 115-133.

Kellum, R. and P. Schedl. (1992) "A group of scs elements function as domain boundaries in an enhancer-blocking assay." *Mol Cell Biol* 12.5: 2424-31.

Kennison, J.A. and M.A. Russell. (1987) "Dosage-dependent modifiers of homeotic mutations in *Drosophila melanogaster*." *Genetics* 116: 75-86.

Kennison, J.A. and J.W. Tamkun. (1988) "Dosage-dependent modifiers of Polycomb and Antennapedia mutations in *Drosophila*." *Proceedings National Academy Science USA* 85: 8136-8140.

Komachi, K., M.J. Redd and A.D. Johnson. (1994) "The WD repeats of Tup1 interact with the homeo domain protein a2." *Genes and Development* 8: 2857-2867.

Kruger, W., C.L. Peterson, A. Sil, C. Coburn, G. Arents, E.N. Moudrianakis and I. Herskowitz. (1995) "Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription." *Genes and Development* 9: 2770-2779.

Kuzin, B., S. Tillib, Y. Sedkov, L. Mizrokhi and A. Mazo. (1994) "The *Drosophila trithorax* gene encodes a chromosomal protein and directly regulates the region-specific homeotic gene *fork head*." *Genes and Development* 8: 2478-2490.

Landecker, H.L., D.A.R. Sinclair and H.W. Brock. (1994) "Screen for Enhancers of *Polycomb* and *Polycomblike* in *Drosophila melanogaster*." *Developmental Genetics* 15: 425-434.

Laurent, B.C., I. Treich and M. Carlson. (1993) "Role of yeast SNF and SWI proteins in transcriptional activation." *Cold Spring Harb Symp Quant Biol* 58.257: 257-63.

Lawrence, P.A. and G. Morata. (1994) "Homeobox genes: their function in *Drosophila* segmentation and pattern formation." *Cell* 78: 181-189.

Le, M.-H., D. Duricka and G.H. Karpen. (1995) "Islands of complex DNA are widespread in *Drosophila* centric heterochromatin." *Genetics* 141: 283-303.

Lewis, E.B. (1978) "A gene complex controlling segmentation in *Drosophila*." *Nature* 276: 565-570.

Lindsley, D.L. and G.G. Zimm. The Genome of *Drosophila melanogaster*. San Diego: Academic Press, 1992.

Locke, J., M.A. Kotarski and K.D. Tartof. (1988) "Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effect." *Genetics* 120.1: 181-98.

Lonie, A. "Cloning and Characterisation of *Polycomblike*." Ph.D. University of Adelaide, 1994.

Lonie, A., R. D'Andrea, R. Paro and R. Saint. (1994) "Molecular characterisation of the *Polycomblike* gene of *Drosophila melanogaster*, a *trans*-acting negative regulator of homeotic gene expression." *Development* 120: 2629-2636.

Lucchini, R. and J.M. Sogo. (1995) "Replication of transcriptionally active chromatin." *Nature* 374.6519: 276-80.

Maillet, L., C. Boscheron, M. Gotta, S. Marcand, E. Gilson and S. Gasser. (1996) "Evidence for silencing compartments within a yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression." *Genes and Development* 10: 1796-1811.

Martin, E.C. and P.N. Adler. (1993) "The *Polycomb* group gene *Posterior Sex Combs* encodes a chromosomal protein." *Development* 117: 641-655.

Mathog, D., M. Hochstrasser, Y. Gruenbaum, H. Saumweber and J. Sedat. (1984) "Characteristic folding pattern of polytene chromosomes in *Drosophila* salivary gland nuclei." *Nature* 308: 414-421.

McCall, K. and W. Bender. (1996) "Probes for chromatin accessibility in the *Drosophila* bithorax comples respond differently to *Polycomb*-mediated repression." *EMBO J.* 15.3: 569-580.

McGinnis, W., R.L. Garber, J. Wirz, A. Kuroiwa and W.J. Gehring. (1984) "A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans." *Cell* 37.2: 403-8.

Messmer, S., A. Franke and R. Paro. (1992) "Analysis of the functional role of the *Polycomb* chromo domain in *Drosophila melanogaster*." *Genes and Development* 6: 1241-1254.

Moazed, D. and P.H. O'Farrell. (1992) "Maintenance of the *engrailed* expression pattern by *Polycomb* group genes in *Drosophila*." *Development* 116: 805-810.

Morata, G. and S. Kerridge. (1981) "Sequential functions of the bithorax complex of *Drosophila*." *Nature* 290.5809: 778-81.

Moretti, P., K. Freeman, L. Coodly and D. Shore. (1994) "Evidence that a complex of SIR proteins interacts with the silencer and telomere binding protein RAP1." *Genes and Development* 8: 2257-2269.

Müller, J. (1995) "Transcriptional silencing by the Polycomb protein in *Drosophila* embryos." *The EMBO Journal* 14.6: 1209-1220.

Müller, J. and M. Bienz. (1991) "Long range repression conferring boundaries of *Ultrabithorax* expression in the *Drosophila* embryo." *EMBO J.* 10.11: 3147-3155.

Müller, J., S. Gaunt and P.A. Lawrence. (1995) "Function of the Polycomb protein is conserved in mice and flies." *Development* 121: 2847-2852.

Murphy, G. and T. Kavanagh. (1988) "Speeding-up the sequencing of double-stranded DNA." *Nucleic Acids Res.* 16: 5198.

Murray, N.E., W.J. Brammar and K. Murray. (1977) "Lambdoid phages that simplify the recovery of in vitro recombinants." *Mol. Gen. Genet.* 150: 53-58.

Naegeli, H., P. Loetscher and F.R. Althaus. (1989) "Poly ADP-ribosylation of proteins. Processivity of a post-translational modification." *J Biol Chem* 264.24: 14382-5.

Orlando, V. and R. Paro. (1993) "Mapping Polycomb-repressed domains in the Bithorax complex using in vivo formaldehyde cross-linked chromatin." *Cell* 75: 1187-1198.

Paro, R. (1990) "Imprinting a determined state into the chromatin of *Drosophila*." *Trends in Genetics* 6.12: 416-421.

Paro, R. and D.S. Hogness. (1991) "The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*." *Proc Natl Acad Sci U S A* 88.1: 263-7.

Patel, N.H., E. Martin-Blanco, K.G. Coleman, S.J. Poole, M.C. Ellis, T.B. Kornberg and C.S. Goodman. (1989) "Expression of *engrailed* proteins in arthropods, annelids, and chordates." *Cell* 58: 955-968.

Pattatucci, A.M. and T.C. Kaufman. (1991) "The homeotic gene Sex combs reduced of *Drosophila melanogaster* is differentially regulated in the embryonic and imaginal stages of development." *Genetics* 129.2: 443-61.

Pearson, W.R. and D.J. Lipman. (1988) "Improved tools for biological sequence comparison." *Proc. Natl Acad. Sci. USA* 85: 2444-2448.

Peifer, M., F. Karch and W. Bender. (1987) "The bithorax complex : control of segmental identity." *Genes and Development* 1: 891-898.

Peterson, C. and J.W. Tamkun. (1995) "The SWI-SNF complex: a chromatin remodeling machine?" *Trends in Biochemical Science* 20: 143-146.

Peterson, C.L. (1996) "Multiple SWItches to turn on chromatin?" *Current Opinion in Genetics and Development* 6: 171-175.

Peterson, C.L. and I. Herskowitz. (1992) "Characterization of the yeast *SWI1, SWI2* and *SWI3* genes, which encode a global activator of transcription." *Cell* 68: 573-583.

Phillips, M.D. and A. Shearn. (1990) "Mutations in *polycombeotic*, a *Drosophila* Polycomb-Group gene, cause a wide range of maternal and zygotic phenotypes." *Genetics* 125: 91-101.

Pirrotta, V. (1991) "The genetics and molecular biology of *zeste* in *Drosophila melangaster*." *Advances in Genetics* 29: 301-348.

Pirrotta, V. (1995) "Chromatin complexes regulating gene expression in *Drosophila*." *Current Opinion in Genetics and Development* 5: 466-472.

Pirrotta, V., C.S. Chan, D. McCabe and S. Qian. (1995) "Distinct parasegmental and imaginal enhancers and the establishment of the expression pattern of the Ubx gene." *Genetics* 141.4: 1439-50.

Platero, J.S., T. Hartnett and J.C. Eissenberg. (1995) "Functional analysis of the chromo domain of HP1." *EMBO J.* 14.16: 3977-3986.

Pokholkova, G.V., I.V. Makunin, E.S. Belyaeva and I.F. Zhimulev. (1993) "Observations on the induction of position effect variegation of euchromatic genes in *Drosophila melanogaster*." *Genetics* 134.1: 231-42.

Poux, S., C. Kostic and V. Pirrotta. (1996) "Hunchback-independent silencing of late *Ubx* enhancers by a Polycomb Group Response Element." *EMBO Journal* 15.17: 4713-4722.

Quinn, J., A.M. Fryberg, R. Ganster, M.C. Schmidt and C.L. Peterson. (1996) "DNA-binding properties of the yeast SWI/SNF complex." *Nature* 379: 844-847.

Rastelli, L., C.S. Chan and V. Pirrotta. (1993) "Related chromosome binding sites for *zeste*, suppressors of *zeste* and *Polycomb* group proteins in *Drosophila* and their dependence on *Enhancer of zeste* function." *EMBO J.* 12.4: 1513-1522.

Riley, L.G., G.B. Ralston and A.S. Weiss. (1996) "Multimer formation as a consequence of separate homodimerization domains - the human c-jun leucine zipper is a transplantable dimerization module." *Protein Engineering* 9.2: 223-230.

Rykowski, M., S. Parmelee, D. Agard and J. Sedat. (1988) "Precise Determination of the Molecular Limits of a Polytene Chromosome Band : Regulatory Sequences for the *Notch* Gene Are in the Interband." *Cell* 54: 461-472.

Santamaria, P. and N.B. Randsholt. (1995) "Characterization of a region of the X chromosome of *Drosophila* including *multi sex combs (mxc)*, a *Polycomb* group gene which also functions as a tumour suppressor." *Molecular and General Genetics* 246: 282-290.

Sato, T., H.P. H and D.R. E. (1984) "Homoeosis in *Drosophila*:Maternal effects of the enhancer of Polycomb locus and its interaction with Polycomb and related loci." *Developmental Genetics* 4: 185-198.

Schlossherr, J., H. Eggert, R. Paro, S. Cremer and R.S. Jack. (1994) "Gene inactivation in *Drosophila* mediated by the Polycomb gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fibre." *Mol Gen Genet* 243.4: 453-62.

Scott, M.P. and A.J. Weiner. (1984) "Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*." *Proc Natl Acad Sci U S A* 81.13: 4115-9.

Sharp, E.J., E.C. Martin and P.N. Adler. (1994) "Directed Overexpression of *Suppressor 2 of zeste* and *Posterior Sex Combs* Results in Bristle Abnormalities in *Drosophila melanogaster*." *Developmental Biology* 161: 379-392.

Shearn, A. and A. Garen. (1974) "Genetic control of imaginal disc development in *Drosophila*." *Proc Natl Acad Sci U S A* 71.4: 1393-7.

Shearn, A., T. Rice, A. Garen and W. Gehring. (1971) "Imaginal disc abnormalities in lethal mutants of *Drosophila*." *Proc Natl Acad Sci U S A* 68.10: 2594-8.

Shore, D. and K. Nasmyth. (1987) "Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements." *Cell* 51: 727-732.

Simon, J., D. Bornemann, K. Lunde and C. Schwarz. (1995) "The *extra sex combs* product contains WD40 repeats and its time of action implies a role distinct from other *Polycomb* group products." *Mechanisms of Development* 53: 197-208.

Simon, J., A. Chang, W. Bender, M.J. Shimell and M. O'Connor. (1993) "Elements of the *Drosophila* bithorax Complex that mediate repression by *Polycomb* Group products." *Developmental Biology* 158: 131-144.

Simon, J., A. Chiang and W. Bender. (1992) "Ten different *Polycomb* group genes are required for spatial control of the *abdA* and *AbdB* homeotic products." *Development* 114: 493-505.

Simon, J., M. Peifer, W. Bender and M. O'Connor. (1990) "Regulatory elements of the bithorax complex that control expression along the anterior-posterior axis." *EMBO J.* 9.12: 3945-3956.

Smith, D.B. and K.S. Johnson. (1988) "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase." *Gene* 67: 31-40.

Smouse, D., C.S. Goodman, A.P. Mahowald and N. Perrimon. (1988) "*polyhomeotic*: a gene required for the embryonic development of axon pathways in the central nervous system of *Drosophila*." *Genes and Development* 2: 830-842.

Song, H.Y. and D.B. Donner. (1995) "Association of a RING finger protein with the cytoplasmic domain of the human type-2 tumour necrosis factor receptor." *Biochem J* :

Staden, R. (1980) "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing." *Nucleic Acids Res.* 8: 3673-3694.

Stassen, M.J., D. Bailey, S. Nelson, V. Chinwalla and P.J. Harte. (1995) "The *Drosophila trithorax* proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins." *Mechanisms of Development* 52: 209-223.

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

Struhl, G. (1981b) "A homoeotic mutation transforming leg to antenna in *Drosophila*." *Nature* 292: 36-41.

Struhl, G. and D. Brower. (1982) "Early role of the esc+ gene product in the determination of segments in *Drosophila*." *Cell* 31.1: 285-92.

Studier, F.W. and B.A. Moffatt. (1986) "Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes." *J. Mol. Biol.* 189: 113-130.

Studier, F.W., A.H. Rosenberg, J.J. Dunn and J.W. Dubendorff. (1990) "Use of T7 RNA Polymerase to Direct Expression of Cloned Genes." *Methods Enzymol.* 185: 60-89.

Tamkun, J.W., R. Deuring, M.P. Scott, M. Kissinger, A.M. Pattatucci, T.C. Kaufman and J.A. Kennison. (1992) "*brahma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2." *Cell* 68: 561-572.

Tschiersch, B., A. Hofmann, V. Krauss, R. Dorn, G. Korge and G. Reuter. (1994) "The protein encoded by the *Drosophila* position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes." *Embo J* 13.16: 3822-31.

Tsukiyama, T., P.B. Becker and C. Wu. (1994) "ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor." *Nature* 367: 525-532.

van der Voorn, L. and H.L. Ploegh. (1992) "The WD-40 repeat." *Febs Lett* 307.2: 131-4.

Wakimoto, B.T. and M.G. Hearn. (1990) "The Effects of Chromosome Rearrangements on the Expression of Heterochromatic Genes in Chromosome 2L of *Drosophila melanogaster*." *Genetics* 125: 141-154.

Wang, W., Y. Xue, S. Zhou, A. Kuo, B. Cairns and G. Crabtree. (1996) "Diversity and specialisation of mammalian Swi/Snf complexes." *Genes and Development* 10.2117-2130:

Wilson, C.J., D.M. Chao, A.N. Imbalzano, G.R. Schitzler, R.E. Kingston and R.A. Young. (1996) "RNA Polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling." *Cell* 84: 235-244.

Wu, C.T. and M. Howe. (1995) "A genetic analysis of the Suppressor 2 of zeste complex of *Drosophila melanogaster*." *Genetics* 140.1: 139-81.

Zhang, C.-C. and M. Bienz. (1992) "Segmental determination in *Drosophila* conferred by hunchback (*hb*), a repressor of the homeotic gene Ultrabithorax (*Ubx*)." *Proceedings National Academy Science USA* 89: 7511-7515.

Zhang, C.-C., J. Müller, M. Hoch, H. Jäckle and M. Bienz. (1991) "Target sequences for *hunchback* in a control region conferring *Ultrabithorax* expression boundaries." *Development* 113: 1171-1179.

Zhang, P. and A.C. Spradling. (1994) "Insertional mutagenesis of *Drosophila* heterochromatin with single P elements [see comments]." *Proc Natl Acad Sci U S A* 91.9: 3539-43.

Zhao, K., C.M. Hart and U.K. Laemmli. (1995) "Visualization of chromosomal domains with boundary element-associated factor BEAF-32." *Cell* 81.6: 879-89.

Zink, B., Y. Engström, W.J. Gehring and R. Paro. (1991) "Direct interaction of the *Polycomb* protein with *Antennapedia* regulatory sequences in polytene chromosomes of *Drosophila melanogaster*." *EMBO J.* 10.1: 153-162.

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

Zink, D. and R. Paro. (1995) "*Drosophila* Polycomb-group regulated chromatin inhibits accessibility of a *trans*-activator to its target DNA." *EMBO J.* 14.22: 5660-5671.