INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Effects of Repetitiveness, Pairing and Linkage on Position-Effect Variegation in *Drosophila*

by

Joy F. Sabl

A dissertation submitted in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
University of Washington
1996

Approved by

Charles D. Laird

Professor Charles D. Laird
(Chairperson of Supervisory Committee)

Steven Henikoff

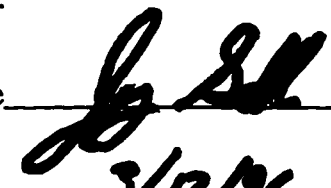
Barbara Wakimoto

Program Authorized
to Offer Degree Zoology

Date March 12, 1996

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Request for copying or reproduction of this dissertation may be referred to University Microfilm, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 3/12/96

University of Washington

Abstract

Effects of Repetitiveness, Pairing and Linkage on Position-Effect Variegation in *Drosophila*

by Joy F. Sabl

Chairperson of the Supervisory Committee: Professor Charles D. Laird
Department of Zoology and
Fred Hutchinson Cancer
Research Center

Position-effect variegation (PEV) is a long distance silencing effect that results in variable inactivation of one or many genes. PEV is caused by the abnormal juxtaposition of heterochromatin and euchromatin following chromosome breakage. By using artificially constructed transgene arrays, I have found that the effects of PEV can be modified by changes in the local repetitiveness of a genetic region. Interestingly, this modification takes place at a distance from the breakpoint, suggesting that the control of variegation does not occur only at the point of propagation. I here propose a new model, emphasizing the role of paired repetitive elements, to account for heterochromatization and for the propagation of PEV. In addition, by using the phenomenon of dominant PEV, I compare the relative contribution of intrachromosomal and allelic pairing to *trans*-silencing in *Drosophila*, look at the effects of chromosomal location on dominant PEV, and examine a large array that does not seem to be prone to variegation, in spite of its repetitive nature. Finally, I review the history of chromosomal pairing phenomena and suggest a mechanism that might unify many known pairing and homology-sensing effects in diverse organisms.

Table of Contents

	Page
List of Figures	iii
List of Tables	iv
Foreword	vi
Chapter I: An Introduction to Position-Effect Variegation, Part 1: Historical Observations and Models	1
An Introduction to Position-Effect Variegation, Part 2: Experimental Background	9
Sequence composition and PEV	9
PEV and the heterochromatin-distance effect	10
Pairing and PEV	15
Heterochromatin and euchromatin: packing the genome	19
Chapter II: Copy Number and Orientation Determine the Susceptibility of a Gene to Silencing by Nearby Heterochromatin in *Drosophila*	34
Introduction	34
Materials and Methods	37
Results: An increase in local repetitiveness enhances PEV	41
PEV is sensitive to gene orientation	43
Discussion: Repeat arrays and orientation effects	48
A pairing-looping model for PEV	52
Chapter III: Linkage modification of PEV in two related *P[bw+]* lines	63
Introduction	63
Materials and Methods	64
Results: Linkage enhancers	67
cis-and *trans*-inactivation	70
Discussion	72
Chapter IV: Orientation-Based Disruption of Local Homologous Pairing and Dominant PEV	85
Introduction	85
Materials and Methods	88
Results: Orientation determines responsiveness to dominant PEV	90
Discussion	91
Chapter V: *P[PRAT-bw+]* Repeat Expansions	98
Introduction	98
Materials and Methods	103
Results: Large repeat arrays	108
Discussion	113
Chapter VI: Conclusions	120

Appendix A: A Partial History of Somatic Pairing Phenomena 124
Before the genetic code: early cytological concepts of homology 125
The rise of genetics: connecting homology and allelism 128
PEV and the idea of the genetic unit 132
The importance of pairing 138
The biochemistry of repetitiveness in genomes: from C_0t to CATG 143

Appendix B: Somatic Pairing and Homology Sensing--Mechanisms 152
RIP and MIP: interactions of repeated sequences in fungi 152
Mechanisms 159
DNA-DNA interactions... heteroduplex 165

Appendix C: The Suitability of *16X98* and *98v1* Derivatives for Interphase Cytological Analysis of *trans*-inactivation 174

Appendix D: PEV As a Model for the Genetics of Human Disease--Huntington Disease, Trinucleotide Expansions and Heterochromatin Formation 178
Epigene Conversion: A Proposal with Implications for Gene Mapping in Humans (a model from 1992) 178
An update of the epigene conversion model; its relevance 185

References 188

LIST OF FIGURES

Number		*Page*
I-1.	Position-effect variegation (PEV) and Dominance	26
I-2.	The "spreading" of PEV	27
I-3.	The heterochromatin-distance effect	29
I-4.	Some *brown*^{Dominant} pairing effects	30
I-5.	Different meanings of "pairing"	31
I-6.	Pairing mechanisms	32
I-7.	Pairing and dominant PEV	33
II-1.	Screens for enhancement of *V21*	57
II-2.	Phenotypes and Southern analysis of duplication-bearing lines	58
II-3.	Generation and interconversion of *V21*-derived alleles	59
II-4.	Orientation of *P[bw+]* affects PEV	60
II-5.	Heterochromatic proximity affects PEV	61
II-6.	A model for PEV	62
III-1.	Comparison of *16X98* and *98v1* lines	79
III-2.	A model for linkage enhancement of *16X98*	80
III-3.	*98v1 trans*-inactivation phenotypes	81
III-4.	*16X98* lines: *trans*-inactivation	82
III-5.	Modeling the relative contributions of pairing regions	83
III-6.	Can *trans* chromosomes enhance PEV?	84
IV-1.	Inverting *P[bw+]* at 92C	94
IV-2.	Flipping a transgene	95
IV-3.	Orientation controls susceptibility to *trans* inactivation	96
IV-4.	Models for local pairing	97
V-1.	*P[PRAT-bw+]* vs. *P[bw+]*	118
V-2.	*P[PRAT-bw+]* arrays	119
A-1.	Cytogenetic and genetic pairing	151
B-1.	Strand invasion in unbroken DNA	172
B-2.	Close packing of helices leads to non-parallel allignments	173

LIST OF TABLES

Number *Page*

II-I Transposase screens of *V21, 92C* and derived lines 56
III-I Linkage Enhancers 77
IV-I Orientation and *trans*-inactivation 93
V-I Expanding *P[PRAT-bw+]* arrays 116
V-II Irradiation of *P[PRAT-bw+]* arrays 117

Acknowledgements

Before thanking my committee for their contributions to my academic and scientific development, I would like to thank the people, friends or otherwise, who first provided me with some clue of how lab work should progress; without Denise Clark and (said with some amazement) Deb Wines, I would probably not be writing this document. Thanks are always due to Debbie Birnby for support, Nick Kosuk for enlightened cynicism, and Doug Dorer for the oddest little things. Doreen Gillespie and Celeste Berg provided memorable hours in my first days in Seattle, and memorable minutes thereafter. I have been lucky in labs; and so I also thank Jim Birchler and all the Birchlerites; long may they flourish!

And, for my reading committee, Barbara, Charles and Steve: thanks for the tolerance, as well as the good ideas that you have sent my way. I couldn't be more delighted; and that's saying a great deal.

Foreword

From the early days of genetics, certain forms of gene inactivation aroused particular interest for the insight they might provide into the nature of genes and of chromosomes. When genes were conceived of as indivisible structures of unknown composition, many observations were mysterious. In time, these mysteries have yielded to our growing understanding of the nature of DNA and of the proteins that package it, open it, shape it and interpret it. For example, until the 1960's, geneticists hotly defended the unitary identity of genes, laboring to explain how mutations could share a phenotype, map to the same locus, and fail to complement each other, and yet (occasionally) recombine. Knowing more, we identify the mutations as alleles, and accept the existence of intragenic recombination. When genes were believed to be essentially autonomous, it was debated whether different phenotypes could arise from different disruptions of a single gene or whether the chromosome contained clusters of related genes with slightly diverged functions. Now, we think of genes as essentially regulatable, and are comfortable with the idea of tissue-specific enhancers located several kilobases from the gene and from each other.

The variegating position effects caused by chromosomal disruptions have always been central to debates on chromosomal organization, gene regulation, and gene identity. Unlike the phenomena of intragenic recombination and allelic difference, position-effect-variegation has been described, but not fully explained. That is, explanations have abounded, with most (after the 1940's) being based on chromatin compaction. As a group, recent models for PEV-associated chromatin compaction have relied on the idea that PEV is mediated by protein conglomerates possessing very specific and entirely unprecedented physical characteristics. Recently, the

Henikoff lab has been involved in a concerted effort to model PEV-dependent compaction as a set of homologous and ectopic physical associations caused by the somatic pairing of like sequences. The following work is part of that effort.

Chapter I presents some features of PEV that require explanation. In chapter II, I show that changes in local sequence repetitiveness and gene orientation at a variegating locus can produce the entire detectable range of variegating phenotypes. From this observation I conclude that the only sequences needed to account for PEV are the heterochromatic breakpoint and the affected locus, and that these two regions communicate with each other. Chapter II concludes with the presentation of a "looping" model which proposes that variegation may be mediated by direct contact between the affected locus and centric heterochromatin. Next (Chapter III), I address the effects of linkage changes on the ability of two, related heterochromatic blocks to induce PEV on adjacent euchromatic genes. Linkage changes demonstrate that PEV can be induced *de novo* when an "isolated" block of heterochromatin is brought closer to centric heterochromatin. However, the intrinsic ability of the block to cause variegation strongly affects both the recovery rate and the phenotype caused by variegating rearrangements. The role of chromosomal pairing in dominant PEV is addressed in these "linkage enhancer" lines as well. Chapter IV considers the importance of local pairing in mediating dominant PEV. Inversion of the *trans*-inactivation target demonstrates, perhaps surprisingly, that the absolute orientation of the *trans*-inactivated gene is more important than the relative orientations of the *cis* and *trans* alleles. Chapter V presents an allelic series of progressively larger transgene amplification arrays, and discusses the implications for repeat-based PEV and also their potential use

for cytological studies of chromatin compaction. Finally, chapter VI is a short summary linking the conclusions of the above chapters.

The appendices to this dissertation are intended to add depth, breadth and mental stimulation, and should be read! Appendix A is a short history of the idea of "homology," a term which is used in many senses, each with its own set of assumptions. Appendix B is a discussion of models for the sensing of homology, in *Drosophila* and (more particularly) other systems. In it, I argue that somatic heteroduplex formation (i.e. standing Holliday junctions) may occur in the absence of strand breakage, and provide a sensitive mechanism for both chromosomal and ectopic pairing. Appendix C details the possible usefulness of the linkage enhancers (described in chapter III) for FISH studies of nuclear organization. Appendix D presents a model for an epigenetic gene-conversion-like mechanism that I proposed in 1992 (Am. J. Hum. Genet. **50**: 1171-1177) to account for some of the difficulties encountered in mapping the human gene responsible for Huntington's Disease, and concludes by discussing how work on repeat-based silencing in *Drosophila* may apply to some of the other trinucleotide-repeat-expansion diseases in humans.

CHAPTER I

An Introduction to Position-Effect Variegation Part 1: Historical Observations and Models

In higher eukaryotes, chromosomes possess two distinct fractions, heterochromatin and euchromatin. Heterochromatin differs from euchromatin in both cytological appearance and sequence organization. Heterochromatin appears condensed during interphase, stains differently from euchromatin at metaphase and participates in non-homologous associations (Heitz 1929). The sequences that are found in heterochromatin typically consist of blocks of highly or moderately repetitive DNA with a low abundance of genes (Willard 1990; Lohe and Hilliker 1995; Weiler and Wakimoto 1995).

When chromosomes are rearranged, linked genes can be variably inactivated by the unnatural juxtaposition of chromosomal regions, even if they lie at a considerable distance from the breakpoint itself. This type of mosaic expression, first noticed by Muller (1930) in *Drosophila melanogaster*, is called position-effect variegation (PEV). More specifically, PEV occurs following the apposition heterochromatin and euchromatin (Schultz 1936). Comparable gene silencing phenomena occur in mammals (Cattanach 1974), plants (Cocciolone and Cone 1993; Meyer, Heidmann and Niedenhoff 1993) and yeast (Allshire *et al.* 1994; Allshire *et al.* 1995) but *Drosophila* has remained the organism of choice for studying PEV.

Drosophila boasts not only a vast selection of variegating alleles of perhaps 80 genes, but also an impressive collection of genic modifiers, the *E(var)* and *Su(var)* loci, which participate directly or indirectly in the establishment of chromatin domains (Reuter and Spierer 1992).

In PEV, heterochromatin appears to spread into euchromatin in a polar manner (Demerec and Slizynska 1937). In addition, the effect of a variegating breakpoint can extend through dozens of genes and span megabases of DNA (Weiler and Wakimoto 1995). Investigation of this remarkable action at a distance has driven research on PEV.

Over the course of 60 years' study, "rules" have been devised to systematize observations made on different variegating genes, stressing commonalities while omitting exceptions. Some of these so-called exceptions, however, have a history that is almost as long and detailed as the history of PEV itself. Models based on these rules, therefore, are only partially successful in explaining known features of PEV. Examination of rules and exceptions such as the following three examples has encouraged us to come up with alternative models for PEV that we hope are more inclusive.

Rule #1: genes variegate only when linked to the disruption breakpoint. Reversion by recombination has served as the definitive test for PEV (reviewed in Lewis 1950). However, for some genes, unlinked alleles can also variegate, if the chromosomal disruption occurs near an allelic position on the homologous chromosome (Schultz 1936). Normally, such sensitivity is noticed as the apparent dominance of a variegating allele to its wild-type homologue; thus, the effect is called dominant PEV (Figure

I-1a). Evidence suggests that heterochromatin inactivates both the *cis* and the *trans* copy of the gene directly: a structurally intact allele in *cis* to the heterochromatic breakpoint is unnecessary for dominant PEV (Slatis 1955a; Dreesen, Henikoff and Loughney 1991; Figure I-1b). Slatis (1955a) found that variegated inactivation of endogenous *brown*$^{+}$ alleles in *trans* to a heterochromatic breakpoint was stronger when a few bands' worth of chromatin remained between the breakpoint and the *brown* locus. The extra chromatin presumably allowed stronger pairing in the *brown* region (reviewed in Spofford 1976). Similarly, local deletions in a variegating *brown*$^{+}$ transgene or in its normal homologue partially rescue the normal allele from dominant PEV (Dreesen, Henikoff and Loughney 1991; Martin-Morris *et al.* 1993), perhaps by disrupting pairing (Figure I-1c; though see also chapter IV). There is no evidence for specific chromosomal sites that mediate the passage of inactivation from one homologue to the other. If such sites did exist, they would have to be common; when copies of *brown*$^{+}$ introduced elsewhere in the genome have been induced to variegate, they have all proven to be dominant (Dreesen, Henikoff and Loughney 1991; Martin-Morris and Henikoff 1995).

Rule #2: When two variegating genes can be assayed in the same tissue, inactivation of the gene farther from the breakpoint guarantees inactivation of the closer gene (Baker 1968). This is a rule based on minimal evidence. Joint inactivation has only been tested for a tiny subset of variegating genes (and of rearrangements affecting these genes): namely, *white* (*w*) *split* (*spl*) and *roughest* (*r*) using the rearrangements $w^{258\text{-}18}$ and $w^{a\text{-}De3}$ (Demerec and Slizynska 1937; Schultz 1941b). Though there is no *a priori* reason to expect that levels of PEV will be the same throughout a fly, joint inactivation of genes that are active in different tissues has been

successfully assayed in some cases by comparing individual flies. For example, Muller (1930) could correlate variegation of white and Notch in w^{m1} (the first described variegating line); this correlates with the compaction of the loci in polytene chromosomes (Hartmann-Goldstein 1967). Generally, genes are more frequently silenced when they are close to a variegating breakpoint than when they are far away (Demerec 1940). PEV in diploid tissues has a cytological correlate in polytene tissues. In salivary gland chromosomes, progressive compaction may extend from a heterochromatic breakpoint into euchromatin, accompanied by loss of distinct chromosome banding (Prokofyeva-Belgovskaya 1939; Schultz 1941a). The extent of compaction responds appropriately to known modifiers of PEV (see below) showing that the two effects may be causally linked.

On the basis of these few examples of genetic polarity, it has been argued that the passage of variegation along the chromosome must involve a physically continuous change in chromatin composition or conformation reaching from the chromosomal discontinuity to the affected gene(s). Compositional changes have been stressed by Tartof *et al.* (1984), Locke *et al.* (1988) and others, who propose that linear arrays of heterochromatin-specific proteins can propagate from heterochromatic breakpoints (Fig. I-2a). Purely conformational changes ("propagative stress") have also been suggested as a mechanism of chromatin regulation (Ephrussi and Sutton 1944; Fig. I-2b). The same idea may underlie the expectations of modern researchers that changes in nucleosome regularity or nuclease accessibility will accompany variegation (Hayashi *et al.* 1990; Schlossherr *et al.* 1994; Locke 1993; Wines *et al.* 1995; Wallrath and Elgin 1995).

Zuckerkandl (1974) earlier proposed that heterochromatin formation could be modeled as a combination of three factors. These factors are

1. a local sequence's intrinsic tendency to compact (e.g., due to the regular curvature induced by the repetition of simple sequences);
2. the ability of nearby regions to pack with the same periodicity; and
3. the ability of repeated and non-repeated sequences to bind the same hypothetical "locking molecule" that would "creep" along the chromosome, stabilizing packing (see Fig. I-2c).

All of these models assume and strive to explain the continuity of heterochromatic propagation. Yet, continuity of PEV has only been shown by cytology; discontinuity can be inferred on similar grounds. For example, the compaction associated with PEV is not always continuous: Belyaeva and Zhimulev (1991) have reported discontinuous compaction in large regions. There is also indirect evidence for some sort of communication between heterochromatic and euchromatic sequences. Demerec (1940) recognized that not all heterochromatic sequences are equally effective in causing PEV. More recently it has been recognized that the "effectiveness" of heterochromatin can depend strongly on the euchromatic context of the reporter gene used (discussed by Spofford 1976).

In general, the likelihood that a chromosomal aberration will induce variegation on a euchromatic gene is not a simple function of the "strength" of the heterochromatin and the "sensitivity" of the euchromatin involved. Instead, the two regions must share some further "specificity." This property may result in the distinctive, non-random distribution of variegating heterochromatic breakpoints seen for certain genes (recently discussed by Henikoff 1996). Continuous-spread or

"oozing" models for PEV (e.g., Tartof *et al., op. cit.*) do not convincingly account for discontinuity of spreading or for the specificity requirement seen in some cases of PEV. The Zuckerkandl "locking molecule" model described above does account for specificity by predicting that different heterochromatic regions will exhibit a variety of mutually exclusive packing periodicities or be stabilized by different locking molecules. Nevertheless, like other "oozing" models, the Zuckerkandl model has difficulty accounting for disparities in heterochromatin propagation through a given, long region (Henikoff, 1995). For example, the degree of inactivation of the *white* gene close to a variegating breakpoint does not correlate with the likelihood that inactivation will spread to more distant genes. If distinctive packing is required to inactivate the region surrounding the *white* locus (as suggested by Zuckerkandl) then that packing should spread consistently into the region beyond the *white* locus, contrary to observations. Furthermore, the range of spreading in and beyond the *white* region is not simply bimodal; instead, it varies widely between alleles. By the Zuckerkandl model, each distinct allele type would imply that the *white* gene was being inactivated by a different form of heterochromatin. This would, in turn, indicate that the *white* locus has specificity for a wide range of hypothetical "locking molecules." In the final analysis, such wide "specificity" is not specific at all, and the Zuckerkandl model begins to resemble the simpler spreading model discussed above.

Rule #3: The clonality observed in some variegating alleles is evidence for the inheritance of a heterochromatic state instated early in development (reviewed in Baker 1968), though the actual likelihood of expression can be modified by later effects of temperature (Becker 1961, quoted in Becker

1966). Responding to suggestions by Schultz that polarity is caused by a delay in the inactivation of more distal genes, Becker (*op. cit.*), Janning (1961), and Baker (1963) all demonstrated that the boundaries of variegating patches follow clonal patterns. Neither the size of the clone nor the location of its boundaries, however, shows much developmental plasticity. Instead, variegating patches follow boundaries similar to those of somatic clones induced in first-instar larvae.

Summing up these observations, Baker (1968) writes that the clonal nature of determination is most simply explained by the assumption that "a polarized suppression of potential gene activity occurs at the pattern-determining stage and spreads from the breakpoint." Early establishment of heterochromatin-induced inactivation followed by clonal maintenance has been well documented in other organisms, notably for the mammalian Barr body. In contrast, a host of environmental factors can modify PEV even late in development (reviewed in Spofford 1976). The state of a variegating allele, therefore, may not necessarily be maintained in the period between the time of determination and the time of eye-disk differentiation (Baker 1968). One possible explanation for developmental plasticity is that the determinant of inactivation is not, in itself, sufficient for inactivation. In a spreading model, however, the "signal" that targets a gene for inactivation is the inactivation complex itself.

The looping model for PEV: an alternative model for PEV is based on idea that nuclear structure may help to coordinate gene expression. Heterochromatin's condensed nature (Heitz 1928; 1929) has long allowed us to see that heterochromatin is clumped within the nucleus, creating or revealing nuclear asymmetry. Nuclei may contain functionally distinct

"compartments." One or more of these compartments may recruit factors that optimize the expression of heterochromatic genes, while minimizing the expression of euchromatic genes. By this hypothesis, variegating position effects arise from the variable mislocalization of a gene to the inappropriate compartment (Devlin, Bingham and Wakimoto 1990; Wakimoto and Hearn 1990; Eberl, Duyf and Hilliker 1993). As described below, there is good cytological evidence for direct association between displaced blocks of heterochromatin. Such associations are predicted to bring displaced blocks back to the heterochromatic compartment, allowing normal activity of heterochromatic genes in the block (see Figure I-4b).

Talbert *et al.* (1994) and Henikoff *et al.* (1995) have suggested that ectopic association of heterochromatic blocks can also lead to inactivation of closely linked euchromatic genes when these encounter the heterochromatin "compartment." Furthermore, regions on cytologically-normal chromosomes can also be mislocalized to the heterochromatin compartment if drawn in by a homologous pairing with a rearranged chromosome. (The terms "ectopic association" and "ectopic pairing" will be used to refer to associations that are reminiscent of allelic pairing, but occur between distinct loci. The first term will be used preferentially when no sequence similarity is implied; the second term will be used generally. Sequence-specific association will be described as such.) In the following chapters, I will argue that a looping model can explain the above-mentioned "rules" and "exceptions," and will present a specific looping model for the polar spreading of heterochromatin observed in PEV of euchromatic genes.

An Introduction to Position-Effect Variegation Part 2: Experimental Background

We study PEV because it provides us with a window on the functional organization of the genome and the nucleus. What, then, can PEV tell us about the role of chromosome composition, chromosome linkage, chromosome pairing and chromatin packing in organizing and regulating chromosomal domains?

Sequence Composition and PEV

PEV was first described by Muller (1930). By 1936, it was clear that the variegating position effects were caused by the juxtaposition of dissimilar domains (Schultz 1936), one of which, heterochromatin, was proposed to contain a preponderance of repeated regions (Bridges 1935). In the modern, molecular era, it has been shown that repetitive DNA is associated with breakpoints which cause variegation, e.g., of *white* (w^{m4} and derivatives, Reuter, Wolff and Friede 1985), of *Acph-1*, a gene coding for the acid-phosphatase-1 enzyme (Shaffer and MacIntyre 1990), or of *brown* (bw^{D}, Csink and Henikoff, in press). Of these, the bw^{D} allele clearly owes its phenotype to an isolated block of insertional heterochromatin which is largely, if not entirely, composed of simple-sequence repeats. Based on the paucity of restriction sites, the heterochromatic break of the variegating *Acph-1* chromosome also contains simple-sequence repeats, while derivatives of w^{m4} have breakpoints that are rich in middle-repetitive elements. As far as assigning causality, these results are largely uninterpretable. Perhaps they show that local repetitiveness affects variegation directly; alternatively, they might just reflect the sequence composition of heterochromatin. Recently, we have recognized our

ability to manipulate local repetitiveness by tandem amplification of P-transposon vectors (Dorer and Henikoff 1994), and have begun to test the role of local repetitiveness in the propagation of PEV (Chapter II).

PEV and the heterochromatin distance effect

The study of position-effect variegation of heterochromatic genes in *Drosophila* has provided evidence for communication between distant chromosomal regions (Baker, 1968; Eberl, Duyf and Hilliker 1993; Figure I-3a). Heterochromatic genes variegate when removed to a point about two-thirds of the distance along a chromosomal arm (Wakimoto and Hearn 1990, Figure I-3a, middle). Variegation of a heterochromatic gene in an isolated heterochromatic block is correspondingly relieved when the block is brought closer to other heterochromatin, apparently responding to a "heterochromatin-distance effect," rather than a "centromere-distance effect" (Eberl, Duyf and Hilliker 1993; Figure I-3a, bottom).

Because isolated heterochromatic blocks can be rearranged without further (detectable) heterochromatin breakage or loss, they provide a particularly precise detection system for heterochromatin-distance effects. Using this system, Eberle *et al.* (*op. cit.*) show that the distance effect can be quantitative: for a given block of heterochromatin, there is a rough linear correlation of the cytological distance between heterochromatic blocks and the expression level of the variegating gene. This suggests that the heterochromatic state of the isolated block is affected, at a distance, by other blocks of heterochromatin -- a form of long-distance, ectopic communication. An isolated block of heterochromatin presumably must retain its identity to induce PEV on adjacent genes. We therefore predict that the heterochromatin-distance effect could be recognized indirectly by

looking for distance-dependent modification of a heterochromatic block's ability to cause euchromatic PEV (Figure I-3b).

To test this prediction, the right genetic material is needed. Above all, a consistent block of heterochromatin should be maintained. Reversion of PEV has often been accomplished by removing a variegating euchromatic gene and adjacent heterochromatin from a proximal site to a more distal location (Hannah 1951; Lewis 1950; Baker, 1968). Unfortunately, these reversion events involved breaks in the PEV-inducing heterochromatin. Heterochromatic breaks change both sequence composition and location, so the relative contribution of the two effects cannot be separated.

Separation of the effects has been achieved by using an isolated block of heterochromatin which confers variegation on the endogenous *brown* gene (Dubinin 1936). "Linkage modification" of *brown*^{Dominant} (*bw*^{D}), another, dominant *brown* allele, was recently described (Talbert, LeCiel and Henikoff 1994; Henikoff, Jackson and Talbert 1995). Modification of *bw*^{D} variegation more closely resembled the "threshold effect" seen by Wakimoto and Hearn (1990) than the quantitative distance effect seen by Eberl *et al.* (1993). Genetic peculiarities of the *bw*^{D} allele might have prevented recognition of a strict heterochromatin-distance effect. Because a heterochromatic insert interrupts coding sequences, *bw*^{D} is a null allele, and its *cis*-inactivation cannot be assayed directly. Instead, variegation is assayed by looking at inactivation of the paired *brown*$^{+}$ allele (*trans*-silencing, Figure I-4a,b).

Talbert *et al.* (1994) imply that in spite of its null phenotype, *bw*^{D} is a normal variegating allele, and furthermore, that *trans*-inactivation of the

bw^{+} allele serves as a simple detector system which reliably mirrors the state of its *cis* inactivation at *bw*^{D}. Five points support this argument. First, though dominance is a common property of variegating *brown* alleles, their *cis*-acting phenotype is indistinguishable from PEV of genes such as *white;* that is, dominant-variegating *brown* alleles are not, as a class, unusual. There is no reason to expect *bw*^{D} to be more unusual than any other variegating *brown* allele. Secondly, the *bw*^{D} *trans-* inactivation system is highly effective in screening for mutations in the classical *Su(var)* loci, implying shared components with *cis*-acting PEV. Thirdly, no second-site modifiers specific for *trans*-inactivation have been found, despite several attempts to find such genes (including the screens of Talbert *et al.* (1994) which instead produced the "linkage modifier" derivatives of *bw*^{D}). Next, linkage-enhanced derivatives of *bw*^{D} (*R: bw*^{D}) can drag an unrearranged *bw*^{D} chromosome into the polytene chromocenter, showing that productive pairing interactions are occurring in these particular cells (Talbert, LeCiel and Henikoff 1994). Finally, recent results have shown that nuclear localization patterns of *bw*^{D} homozygotes, *bw*$^{+}$ homozygotes and *bw*^{D} /*bw*$^{+}$ heterozygotes in interphase somatic cells correlate strongly with levels of *brown* variegation, and that the localization patterns of *bw*^{D} /*bw*$^{+}$ heterozygotes and of *bw*^{D} homozygotes are essentially identical (Csink and Henikoff, in press). Nevertheless, it is possible that *trans-* inactivation and *cis*-inactivation do not respond identically to changes in linkage, so that instead of looking at a single effect, we are again looking at two, confounding effects.

Henikoff *et al.* (1995) looked at a series of constructs, the pigmented *Byron* derivatives of *bw*^{D}, with the intent of assaying linkage-based changes more directly. (*Byron* lines contain a duplication of the *brown* locus, inserting a

brown^{+} allele in tandem with the disrupted *brown* allele of the *brown*^{Dominant} chromosome; they are named after Byron Kadel, who first reported the isolation of such alleles.) The *brown*$^{+}$ half of the *Byron* duplication variegates as if affected directly by PEV. However, the authors argue, based on a comparison of alleles, that the variegation is not due to direct *cis*-inactivation of the *brown*$^{+}$ gene by the heterochromatic block. Rather, they believe that variegation is a secondary effect caused by (and dependent upon) pairing of the active allele with the adjacent, heterochromatized null allele (*para*-inactivation, Figure I-4c). Accepting this argument, we are again left without an assay for linkage-modification of *cis*-inactivation.

Our inability to differentiate between direct and indirect changes in *trans*-inactivation is particularly onerous because Henikoff *et al.* (1995) also found that a subset of chromosome disruptions in *trans* to a heterochromatic block can affect the phenotype of bw^{D} derivatives. We therefore must consider the possibility that enhancement of PEV may occur whenever pairing forces draw "sensitive" genes towards centromeric heterochromatin (Henikoff, Jackson and Talbert 1995; Figure I-4d).

None of the above-described mutageneses has produced data in support of a quantitative heterochromatin-distance-effect on euchromatic genes: the distance of the heterochromatic blocks from each other does not correlate with the degree of phenotypic enhancement (or suppression). Talbert *et al.* (1994) and Henikoff *et al.* (1995) suggest that a subset of the breakpoints may interfere with efficient local chromosomal pairing by causing steric hindrance. I suggest, in addition, that *trans*- inactivation may not consistently mimic *cis*-inactivation. Therefore, even if bw^{D} is subject to a

strict heterochromatin-distance effect, the effect may be undetectable due to the imprecision of the detection system.

Alternatively, the lack of a strict distance effect may not be an artifact. Instead, it may reflect different requirements for nuclear localization between heterochromatin and euchromatin. Howe *et al.* (1995) have recently produced a series of deletions near a breakpoint that induces both heterochromatic and euchromatic PEV. Presumably, these deletions should have a similar effect on the localization of the heterochromatic and euchromatic genes that flank them. If so, they should have reciprocal effects on heterochromatic and euchromatic gene expression.

Unexpectedly, the effects on the two regions were not reciprocal: variegation of the euchromatic transgene (*P[white$^+$]*) did not correlate with variegation of the nearby heterochromatic *light* gene. One possible explanation is that heterochromatic genes have a strict requirement for heterochromatic surroundings (both in terms of nuclear localization and sequence). Euchromatic genes, in contrast, might be very sensitive to nuclear localization. As long as the euchromatic sequences retain a minimal heterochromatic "tag"-- a sequence common to heterochromatin that could mediate further contact with this nuclear compartment--their variegation would be maintained.

A third alternative is that the heterochromatin-distance-effect does not depend only on the linear proximity of blocks along the chromosome. Instead, it may be a measure of the likelihood of spatial proximity between any two blocks in the nucleus, whether tethered in *cis-* by linkage or in *trans-* by pairing interactions. In chapter III, I attempt to distinguish

between these possibilities by producing linkage enhanced lines of a *P[bw+]* transposon array.

Pairing and PEV

Dominant PEV is the silencing of a normal allele by its variegating counterpart. As described above, joint inactivation of homologous genes in dominant PEV is triggered by the change in the chromosomal environment of one of the homologues. Dominant PEV may be regarded as an elaboration of *cis*-acting PEV: first, the gene on the disrupted chromosome is inactivated by heterochromatin; then the normal allele is inactivated because it, too, is re-localized by the global effects of somatic pairing along the chromosome (Csink and Henikoff, in press). Somatic pairing may play other roles in the establishment of PEV as well, perhaps by mediating ectopic contacts near the variegating gene (Chapter II).

To investigate the role of somatic pairing, two types of lesions can be superimposed on lines that show dominant PEV: chromosomal rearrangements and local changes near affected genes. Chromosomal disruptions such as translocations and inversions of either homologue can influence gene expression. Often located far from the region of interest, these rearrangements generally preserve the local environment of affected gene(s), yet alter chromosomal structure and pairing on a cytologically visible scale. Chromosome-modification experiments have been carried out for decades (Dubinin 1936; Slatis 1955a). Precise manipulation of known, short sequences, possible only in recent years, now allows us to alter the local pairing properties of specific regions and sequences. We can now start to dissect the local pairing requirements for effective dominant PEV.

Somatic pairing and association of ectopic sequences

To discuss *trans*-inactivation in detail, we must first consider what we mean by an oft-used term: somatic pairing. This term has been applied loosely to cover any situation where sequences or chromosomal regions associate with each other in somatic nuclei (Figure I-5 a-d). What does this term imply; just as importantly, what does it not imply?

The term "somatic pairing" does not imply a mechanism: visible association of allelic and non-allelic genomic regions is common, yet we do not know the basis or bases of these interactions. For regions sharing sequence homology, several sorts of associations are possible, and are not mutually exclusive (Figure I-6 a-c). Similar irregularities in DNA twist and writhe might be adequate, over long regions, to encourage direct DNA-DNA interactions of a non-Watson-Crick nature (Timsit and Moras 1995). Alternatively, transient single-strandedness might provide a mechanism for genome scanning, and result in transient Watson-Crick interactions between the most nearly identical regions of homologous chromosomes or genes. It is perhaps more likely that homologous associations are caused indirectly, by protein-protein interactions. These would not need to be specific proteins (like recombination complexes) or even distinctive structural proteins (like the synaptonemal complex). Instead, histones and/or regulatory proteins could provide the requisite irregularities in phasing; two similarly deformed chromatin segments would then preferentially pack together (Henikoff 1996; Figure I-6b; see also appendix B). Even in the absence of homology, regions may co-associate, perhaps by binding similar proteins. Candidate proteins might be common transcription factors with fairly broad binding specificities, or proteins that completely lack binding specificity, such as

single-strand-binding proteins (Lamb and Laird 1987).

While many models are possible, one point should be emphasized: all of these models allow for multiple contacts between two or more sequences, occurring concurrently or sequentially. A single locus, therefore, can theoretically participate in both ectopic and allelic pairing, or make multiple, allelic contacts in polytene or polyploid tissues.

Pairing is not a static and exclusionary relationship between two partners. What is the role of somatic pairing in dominant PEV? One easy assumption is that dominant PEV reflected a "partnering" process: a normal allele pairs with its variegating homologue. Responding to its partner's inactivation, perhaps by coupling itself to the homologue's transcription pattern, it is silenced (Figure I-7a). This model can account for various phenotypes seen when more than two homologues are present in the genome (e.g., through segmental aneuploidy); if somatic pairing were a largely competitive process, similar sequences would preferentially associate with each other, to the exclusion of less-similar sequences (Figure I-7b). However, aneuploidy experiments have shown that a single variegating allele can inactivate two normal homologues. Thus, "pairing" can involve more than two sequences, either concurrently or sequentially (Henikoff and Dreesen 1989; Figure I-7c). On a scale where we can observe its consequences, then, "pairing" is a misnomer; nevertheless, I will continue to use it (rather than the more correct "association") in recognition of the fact that most models for the self-recognition of homologous sequences require strict pairing on a finer scale (Appendix B). Pairing has been studied in many ways. Genetically, the phenomena of transvection and pairing-dependent *zeste* inactivation have served both as

a motivation for the study of pairing and as useful tools (e.g., Wu 1989; also see Appendix A). Cytologically, the very existence of polytene chromosomes shows that related sequences can form stable, persistent associations. Chromosome pairing is a common phenomenon in diploid cells as shown by interphase cytology (Hiraoka *et al.* 1993). In sum, evidence suggests that cytological association is common. The forces that underlie chromosome pairing may affect genetic phenomena directly.

Genetic and cytological pairing interactions correlate in considerable detail. For example, the relative orientation of transposable element (TE) duplications determines both their (cytologically visible) pairing properties and their genetically defined pairing properties (Gubb *et al.* 1990; see Appendix A for details). Dorer and Henikoff (1994) have proposed that similar interactions occur on a cytologically invisible scale, and have used this assumption to infer the pairing state of alleles by virtue of their phenotypes. Similarly, large disruptions which cause a cytologically-visible disruption of pairing in polytene chromosomes can also interfere with genetic phenomena that depend on pairing (Henikoff, Loughney and Dreesen 1993). Henikoff and Dreesen (1989), Dreesen *et al.* (1991) and Martin-Morris *et al.* (1993) have extrapolated from this observation to explain how small deletions might interfere with the local pairing of homologous *brown* alleles.

The results of Dorer and Henikoff (1994) suggest that repeat arrays participate in exquisitely sensitive *cis*-pairing interactions, while the results of Dreesen *et al.* (1991) and Martin-Morris *et al.* (1993) suggest that the sensitivity of *trans*-inactivation to local homologous pairing is equally extreme. How do these two sensitive systems affect each other? By

modifying the orientation and the repetitiveness of transgenes in *cis* and in *trans* to a variegating breakpoint, I address this question in chapter IV.

Heterochromatin and euchromatin: packing the genome

Heterochromatin was first named and described by virtue of its distinctive cytological compaction (Heitz 1929). This compaction was proposed to be directly responsible for the inactivating effects of PEV (Prokovyeva-Belgovskaya 1947). Recent work on genic modifiers of PEV has focused on their potential role in initiating or maintaining a high level of chromatin compaction (Tartof *et al.* 1989; Reuter and Spierer 1992). Heterochromatin formation is not the only process believed to inactivate sequences by changing their packing. Certain developmentally-regulated genes are inappropriately activated if any of a set of proteins, collectively known as the Polycomb-group (PcG) proteins, is mutant or absent (Moehrle and Paro 1994). Local aggregations of Polycomb, polyhomeotic, and other Polycomb-group proteins are found at sites containing genes affected by *PcG* mutations, supporting the idea that PcG proteins coordinately form a packaging complex (Franke *et al.* 1992; Lonie *et al.* 1994).

Some of the *Polycomb-group* genes also act as modifiers of PEV (e.g. Farkas *et al.* 1994). In addition, the best-characterized *su(var)*-encoded protein, HP1, shares a putative chromatin-binding domain with Polycomb (Paro and Hogness 1991), suggesting that they could have partially analogous functions in their respective silencing complexes.

If PcG proteins form a protein complex involved in silencing (Paro 1990; Franke *et al.* 1992; Kennison 1993), and the *E(var)* and *Su(var)* gene products are similarly involved in the aberrant silencing of euchro-matic

genes placed near or in heterochromatin (Locke, Kotarski and Tartof 1988; Dorn *et al.* 1993), are these packing domains truly different? Genetic and cytological analysis both allow us to distinguish between the two complexes. Considering that both complexes can cause stable (or clonally heritable) repression of euchromatic gene activity, it is not surprising that a few mutations affect both systems; rather, it is surprising how little genetic overlap has been found. The low degree of overlap between the groups of genes probably implies parallel, rather than cooperative, function of their respective products (Tartof and Henikoff 1991). Exclusivity can also be visualized cytologically using the localization patterns of antibodies on polytene chromosomes. Binding patterns differ strikingly for anti-HP1 (James and Elgin 1986) and anti-Polycomb-domain antibodies (Messmer, Franke and Paro 1992; see also Platero *et al.* 1995). Nor are these the only two regional-silencing mechanisms in *Drosophila*: yet another silencing system apparently exists at *Drosophila* telomeres. Telomeric silencing does not respond to *e(var)*s, *su(var)*s or the majority of *PcG* genes tested (Talbert, LeCiel and Henikoff 1994; Wallrath and Elgin 1995); the search for a novel class of telomere-binding-proteins (TbP's) is on-going.

Given these tantalizing hints about the similarities and differences of the proposed silencing complexes, one would like the answers to the following questions:

1) What mediates the spatial separation of different complexes, and what does this say about the control of heterochromatin formation?

2) How similar is chromatin packing in heterochromatin and euchromatin, and what might the latter tell us about the former?

How are complexes spatially segregated? Polycomb-group proteins can be recruited to a site by short DNA sequences, termed "polycomb response elements" (PRE's) which function autonomously and additively in transgene constructs (Gindhart and Kaufman 1995). In contrast, no sequence-specific recruitment or binding activity has (yet) been found for *E(var)'s* and *Su(var)'s*. Another factor may be the (proposed) preferential recruitment of heterochromatin-binding proteins to repetitive sequences (Dorer and Henikoff 1994; this work, chapter II); this idea is supported by the recent observation that transgene duplications can bind the HP1 protein (D. Dorer, personal communication). Intuitively, this is a pleasing explanation; repeats, or high reiteration, are a common feature of constitutive heterochromatin.

However, it now seems that increased repetitiveness may also initiate silencing by providing multiple binding sites for a sequence-specific binding protein. Gindhart and Kaufman (1995) propose the following scenario for transvection, an allele-specific genetic phenomenon affecting some of the genes regulated by PcG-based silencing: PRE's may cluster in certain developmentally-regulated genes. Stable gene repression might require cross-talk between nearby PRE's (in *cis*). Alleles retaining two (or more) copies of a PRE in tandem are not dependent on a homologous allele for pairing of PRE's; however, when an allele retains only one PRE site, *cis*-pairing of PRE's is abolished, and silencing can only be rescued by *trans*-pairing of the lone PRE with an allelic PRE. This predicts an effect very like that of transvection, where the effects of certain mutations are exacerbated by chromosomal rearrangements that interfere with allelic pairing (see Appendix A).

Arguing along similar lines, Hopmann, Duncan and Duncan (1995) suggest that juxtaposition of regulatory regions that happen to contain single (ineffective) PRE's may create silencers *de novo*. Extending this idea, it is possible that specific cases of repeat-enhanced silencing could be mediated by PcG complex formation, if the sequences involved happen to have PRE's. Similarly, the local repetitiveness of retrotransposons (i.e. *TART, HeT-A*) and other sequences could, in theory, be responsible for telomeric silencing in *Drosophila* (Biessmann *et al.* 1992; Levis *et al.* 1993; Wallrath and Elgin 1995), yet not involve the same repeat recognition system as heterochromatin or PcG-based compaction.

How similar can packaging be in heterochromatin and euchromatin?
At their extremes, heterochromatin and euchromatin differ in compaction, density of expressed genes, representation levels in polytene tissues, and sequence composition. Nevertheless, euchromatic packaging has often been compared to the more invariant compaction of constitutive heterochromatin. Facultative heterochromatin, such as that formed in PEV, is interesting precisely because it bridges the conceptual and physical gap between the grand-scale, invariant compaction of simple-sequence repeats at the chromocenter and the local and highly regulated compaction of inactivated euchromatic genes. We study PEV because its boundaries define the physical border where heterochromatin and euchromatin blend together. The study of chromatin packing, and especially the study of chromatin packing of repeated sequences, also brings us to a different sort of border between euchromatin and heterochromatin. To quote from Nagl (1978):

> "Although heterochromatin is well characterized by its genetic inertness and its content of simple-sequence,

> highly repetitive DNA (e.g. satellite DNA) its definition is rather of a quantitative nature. For instance, the chromomeres (or bands) are defined as euchromatin, although they are in a condensed state, contain repetitive DNA sequences, may replicate asynchronously, and their DNA may be underreplicated or amplified to some extent... Actually, the euchromatin of species with many repetitive DNA sequences (e.g. that of the Allium species; see (Flavell *et al.* 1974) shows exactly the same electron density in ultrathin sections as does the heterochromatin (Nagl 1977). Evidently, the proportion of repetitive DNA within a genome determines the degree of chromatin condensation. A euchromatic band in *Chironomus* was found to become heterochromatic (i.e. Giemsa C-banding positive) as a result of duplication (Hagele 1977). Thus it seems that a certain extent of reiteration causes a chromosome segment to condense into the heteropycnotic state."

In terms of the previous section, components of chromatin and regulators of their packing may differ between chromosomal regions, yet the final compaction state that they produce may be similar (Pirrotta and Rastelli 1994). The degree of similarity may reflect analogous early steps in chromatin compaction. Alternatively, it may reflect the uniform contribution of proteins that are more generally involved in chromatin packaging: for example, there may only be a few relatively stable nucleosome packing states. Chromatin condensation in different domains is presently being observed and compared on a molecular level. Nuclease and restriction enzyme digestion are used *in vitro*, i.e. on isolated nuclei (Hayashi *et al.* 1990; Wallrath and Elgin 1995). Induction of methylase activity can be used to probe and mark chromatin accessibility *in vivo* (Wines *et al.* 1995). To date -- and with the exception of euchromatic genes actually inserted into heterochromatin (Wallrath and Elgin 1995) -- neither PEV nor PcG-based silencing seems to produce distinctive changes in

chromatin structure that can be distinguished at this level (reviewed and discussed in Henikoff 1996). In summary, there is little evidence for the induction of an impenetrable nucleoprotein coat, nor for regular, compact, higher-order packaging of DNA in these domains. The possibility that nucleosome phasing is affected at the exact point of heterochromatin-euchromatin junctures is still being investigated; however, other models for chromatin condensation should also be considered.

Condensation (of heterochromatin in interphase cells or of bands in polytene chromosomes) need not be regulated through active modification of chromatin complexes at all. For example, consider the differential condensation seen in polytenes bands and interbands. This distinctive pattern could result from the interspersion of nucleation sites for a variety of chromatin-packing complexes; but alternatively, it could simply reflect local differences in the strength or complexity of *cis*- and *trans*-pairing interactions between homologous regions (Henikoff 1996, also see Appendix B). Using such a model, the "crisp" nature of the band boundary can be explained by nothing more than the cooperativity of lateral packing (and thus, stabilization) of the 30 nm fiber. This idea can be seen as a variant of the Ephrussi and Sutton pairing model (Ephrussi and Sutton 1944):

> "it will be noted that the intimate association of chromosomes in somatic pairing...may reasonably be assumed to lead to a more precise alignment of the submicroscopic structures than that which prevails in unpaired segments. That synaptic association may actually lead to a stretching out of the whole structure of the chromosome is conceivable and moreover suggested by certain cytological features of the salivary gland chromosomes."

[The cytological features to which they refer are presumably those described later by Beermann (1963), namely, that each round of endoreduplication lengthens the chromosomes, as if pairing stability is being increased.]

Similarly, interphase cells could regulate condensation, at least in part, by the cooperative co-localization of like sequences, both in *cis* and in *trans* (see Henikoff 1996; Appendix B).

In chapter V, I describe another set of repeat arrays, produced to study PEV. This array was never successfully induced to variegate. Nevertheless, due to the size and organization of the arrays, we suspect that they might have some features of both heterochromatin and euchromatin, though they appear to be minimally compacted. In the same way that genetic analysis has allowed us to distinguish both differences and similarities in the *trans*-acting factors that mediate heterochromatic compaction and euchromatic compaction, this (and other) repeat arrays may eventually allow us to dissect the *cis*-acting differences that trigger or resist regional compaction and/or silencing in the two chromatin types.

Figure I-1. Position-effect variegation (PEV) and dominance. a) PEV is usually recessive: the combination of a wild-type and a variegating allele produces a wild-type eye (left). Variegating alleles of a few genes show dominance (right), inactivating their wild-type homologue, which would otherwise be sufficient for full expression. This is called *trans*-inactivation. b) Dominance can be explained by the pairing of the wild-type copy of the gene with heterochromatin on the homologous chromosome (top). Lesions on either chromosome that disrupt pairing also interfere with *trans*-inactivation(middle). *Trans-inactivation* of an allele on the structurally normal chromosome does not require the presence of a homlogous allele on the inactivating chromosome (bottom).

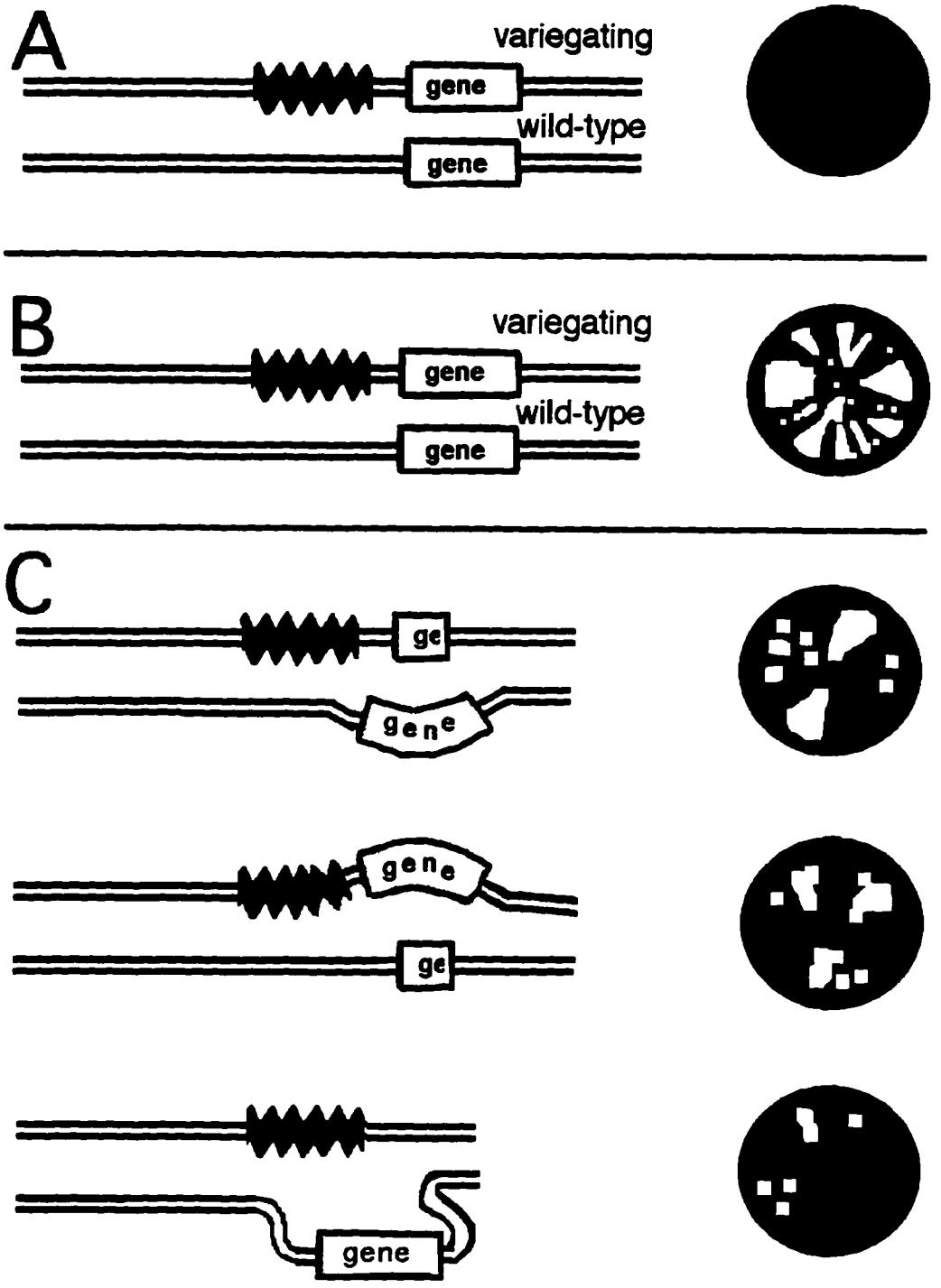
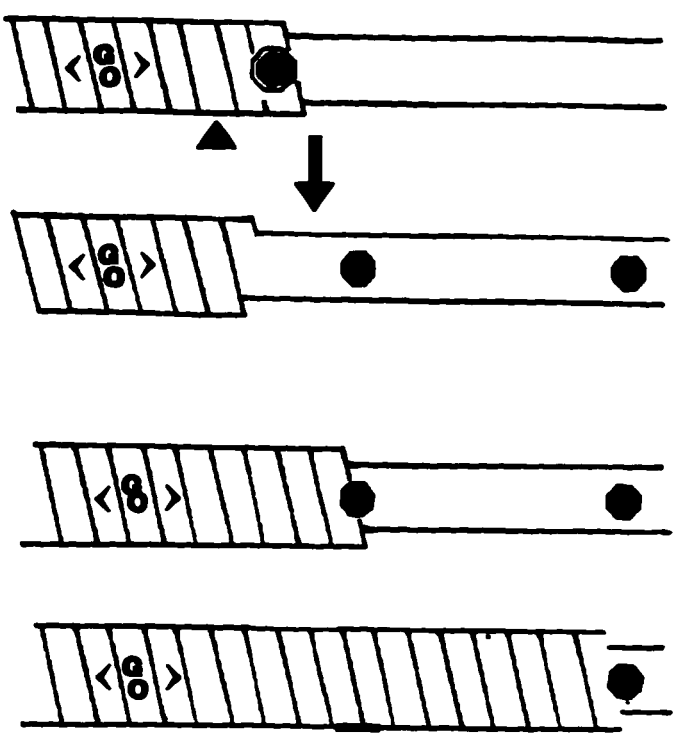

A
variegating
gene
wild-type
gene
B
variegating
gene
wild-type
gene
C
gene
gene
gene
gene
gene

Figure I-2. The "spreading" of PEV. Many models for the propagation of PEV invoke linear spreading of a component or effect along the chromosome. a) In the Tartof model for PEV, assembly of a chromatin complex (//////) is initiated by sequences common in heterochromatin (<GO>) and then propagates until it reaches a termination element (stop sign) at the normal heterochromatin boundary point (top). If the chromosome is broken between these points (arrowhead), and joined to a euchromatic region where potential terminators are rarer or less effective (middle) then propagation will continue variably to these terminators (bottom) causing variable expression of genes in the intervening region. b) Other models (see text) do not imply that special chromatin components are being recruited to form heterochromatin; instead, components already present undergo an allosteric-like change (shown as a shift in the angle of (| | | |). Stress caused by the disruption of homologous pairing might initiate such a shift. The presence of heterochromatin at the breakpoint junction is proposed to disrupt pairing particularly strongly.

A



B

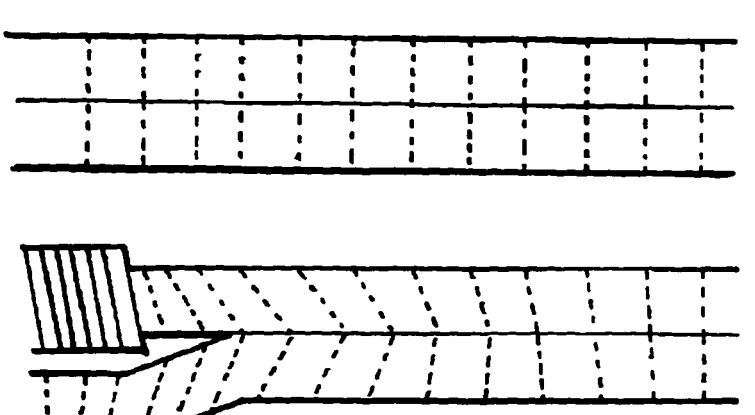


Figure 1-2 continued

c) The Zuckerkandl model for PEV has many components. Repetition of any given sequence (black bar or striped bar) may cause an overall regularity in the chromosome which is conducive to compaction (zig-zag), forming distinct domains (striped, flat zig-zag; dark, tall zig-zag). Heterochromatin, then, would form in regions where the concentration of repeated sequences is high, and would be limited by the presence of adjacent, "unpackable" sequences (irregular lines). Additional sequences juxtaposed to heterochromatin following chromosome breakage (various arrowheads) would have different fates depending on their ability to pack with the form of heterochromatin at the breakpoint. Some will be packable (left; result of breaks at arrows) and others will not be packable (right, result of breaks at left arrow and either arrowhead). To explain the spread of packing through heterologous sequences, Zuckerkandl proposes "locking molecules" (balls) which bind to the repeat and then propagate linearly and cooperatively. This stabilizes compaction, allowing it to pass beyond the boundaries of repeat domains (bottom left; balls not shown above for clarity). If the locking molecule cannot bind or the sequences cannot pack (right, middle and bottom respectively) variegation will not occur. Otherwise, this model is much like the Tartof spreading model, which it pre-dates.

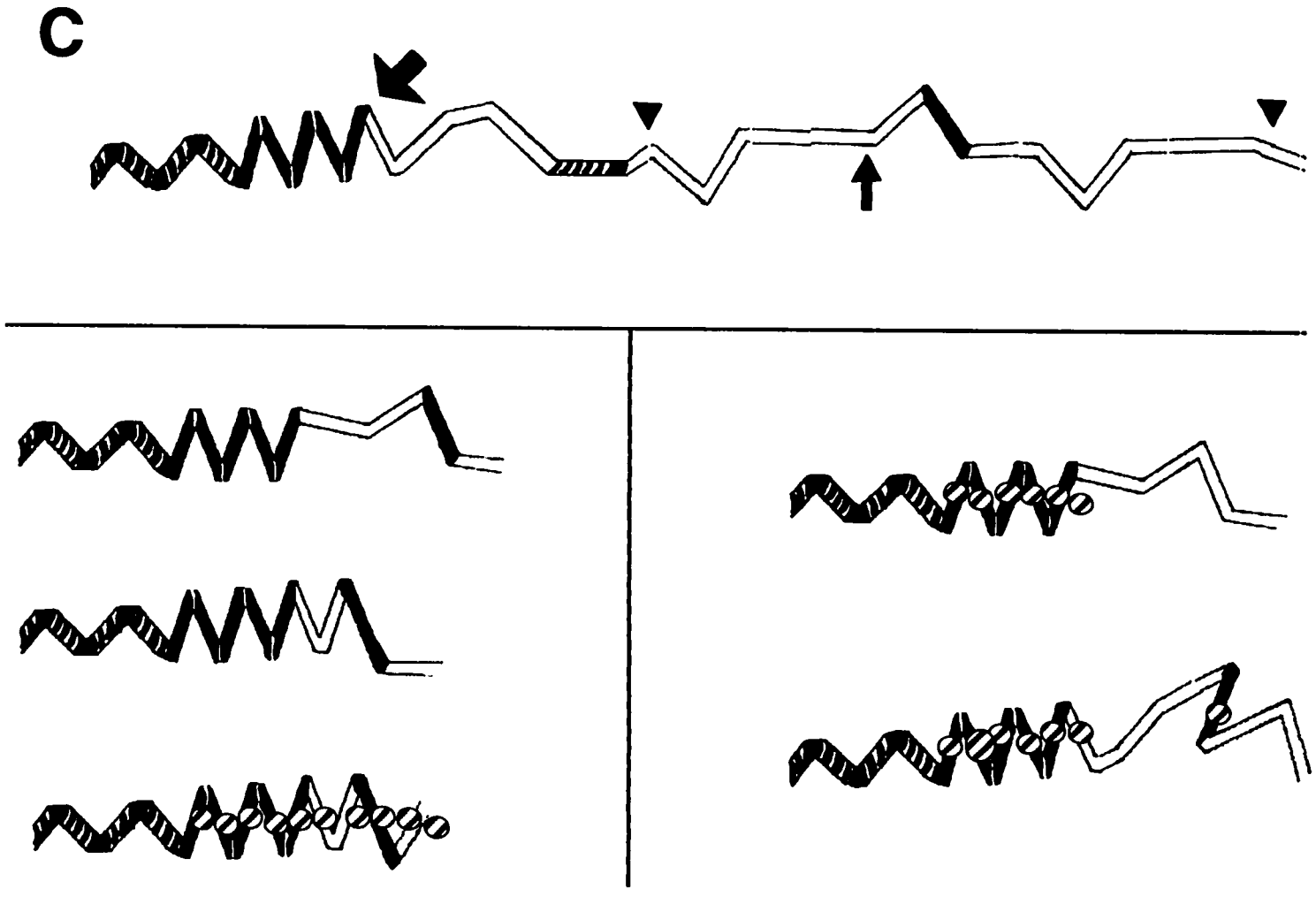

C

Figure I-3. The heterochromatin distance effect. a) When a heterochromatic gene (gray zig-zag rectangle) is displaced from heterochromatin (zig-zag line) into euchromatin (straight line), it may be inactivated (dark X, middle). If the distance to another block of heterochromatin is decreased (bottom), inactivation can be partially or fully suppressed (dotted X). b) A euchromatic gene (rectangle) is partially inactivated (dotted X) by a block of heterochromatin (zig-zag line). If the distance to another block of heterochromatin is decreased (bottom), is inactivation of the euchromatic gene enhanced (dark X)?

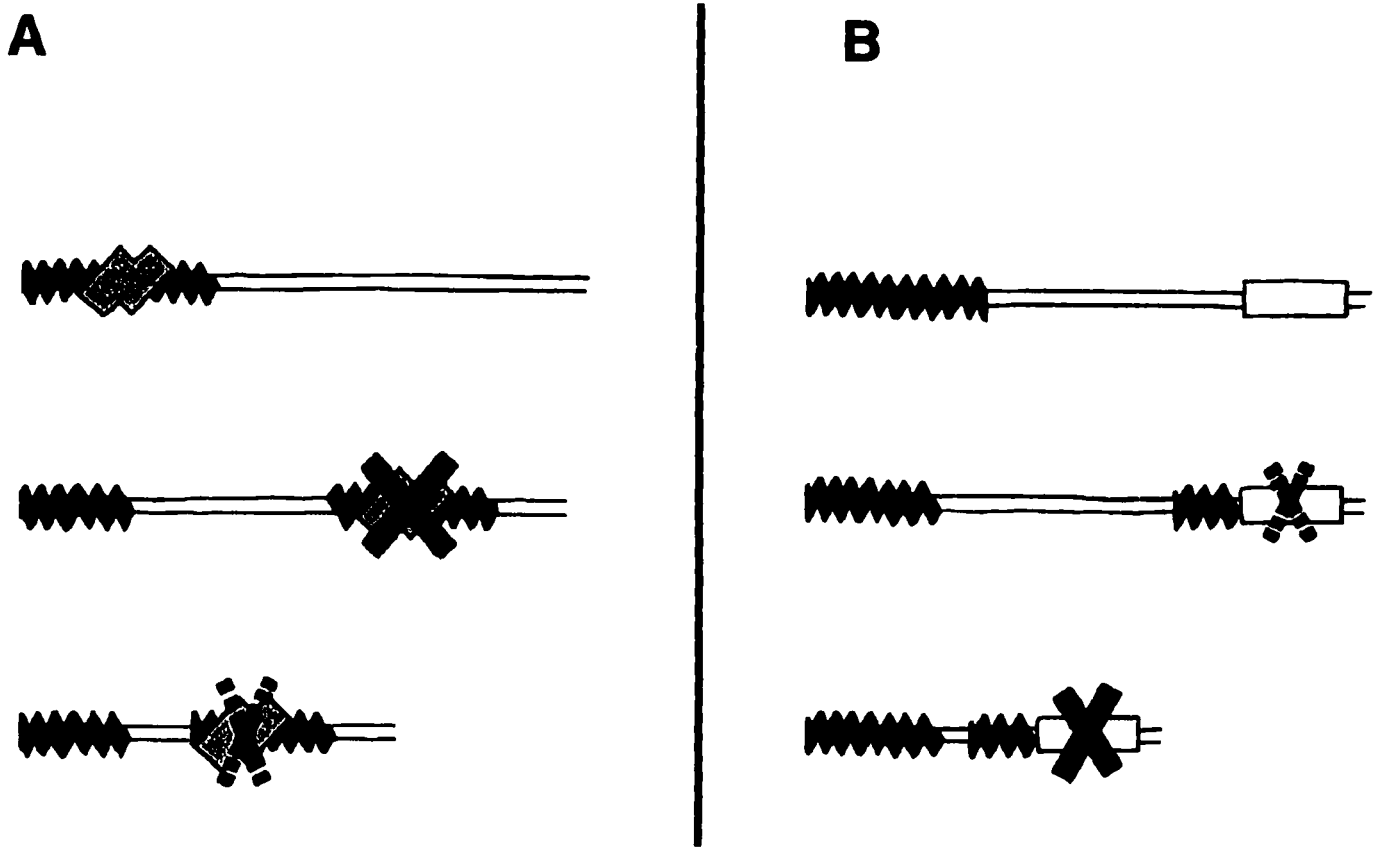

A
B

Figure I-4 Some *bw*^{D} pairing effects. In the gene disruptions *bw*^{D} and *bw*^{D} *bw*$^{+}$ (also called *Byron*), inactivation of *brown*$^{+}$ does not result directly from the effect of heterochromatin on an adjacent gene, as shown in figure I-3. Instead, inactivation is seen following pairing of *bw*^{D} and *bw*$^{+}$ (allelically or intrachromosomally). a) *bw*^{D} is a null allele of *brown* (rectangle) caused by insertion of ~2 Mb of simple-sequence repeats (zig-zag). It's phenotype can only be assayed by its effect on a paired *bw*$^{+}$ gene (dots indicate pairing of homologous chromosomes). b) Cytological evidence supports the idea that variegation depends on the location of sequences within the nucleus. Euchromatic genes may be inactivated when rearrangements reposition them in a heterochromatic compartment (HET) instead of a euchromatic compartment (EU). Variable localization from cell to cell of a tissue leads to variegated expression of the gene. The *bw*^{D} chromosome can reposition the *bw*$^{+}$ chromosome in the nucleus so that it, too, is in the inappropriate compartment. c) based on comparison of different *bw*^{D} and *bw*$^{+}$ derivatives, the *bw*$^{+}$ gene in the *Byron* construct appears to be inactivated indirectly by the adjacent *bw*^{D} gene (middle) not directly by interactions with centric heterochromatin (bottom). d) If a gene is moderately inactivated by a nearby block of heterochromatin (zig-zag), then a rearrangement in the homologous chromosome (arrows) may produce an allelic combination (bottom) where chromosomal pairing (dots) enhances PEV (dark X) by bringing the region closer to heterochromatin.

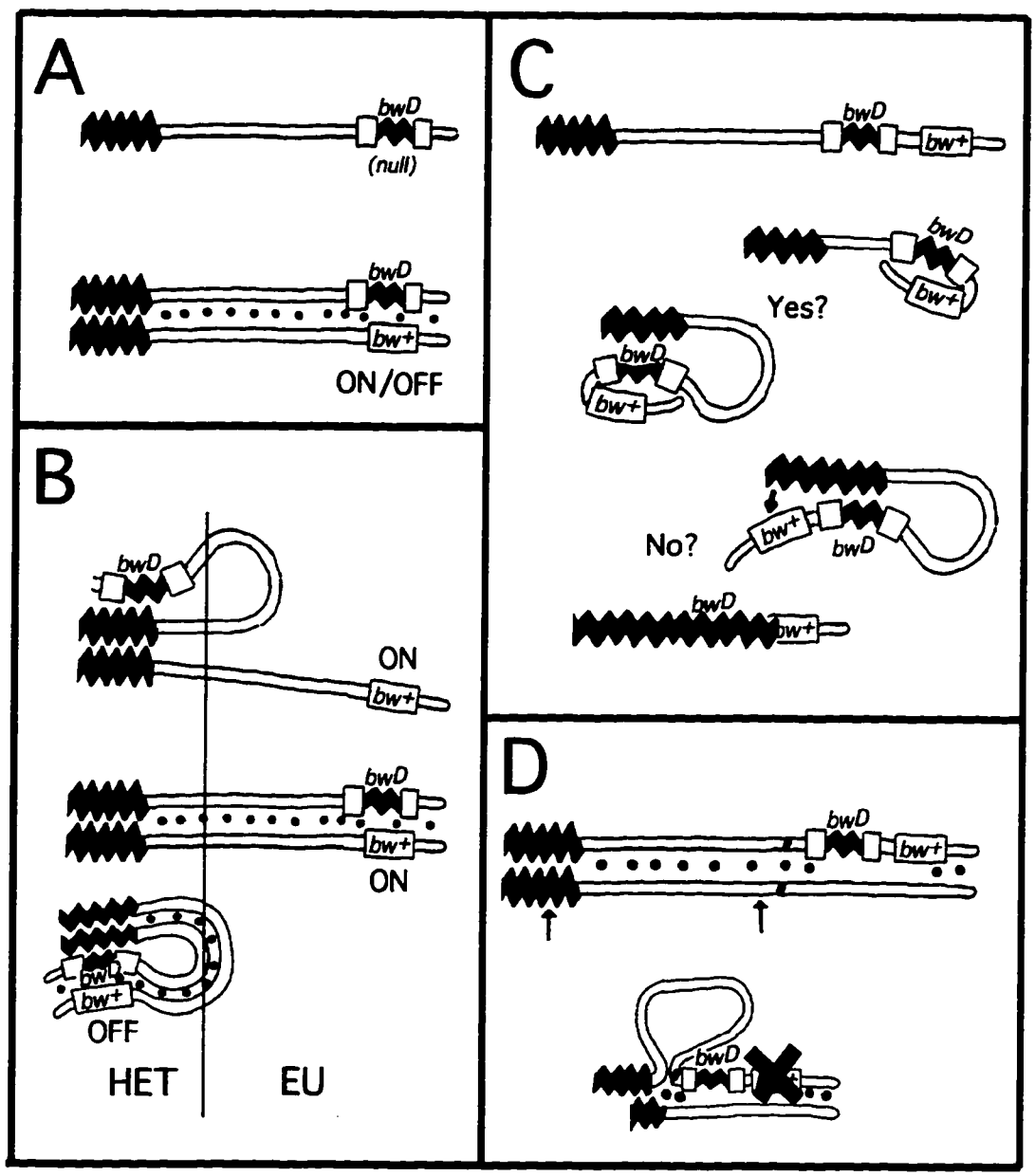
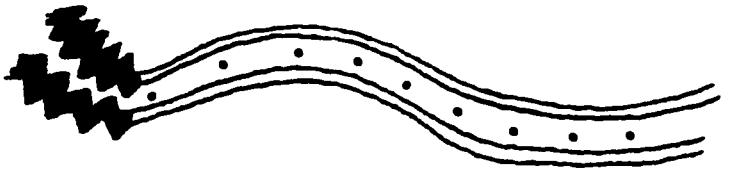
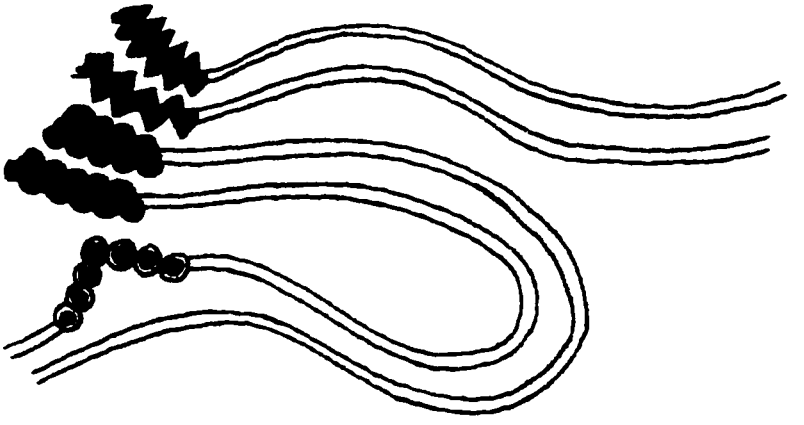

A
bwD
(null)
bwD
bw+
ON/OFF
B
bwD
ON
bw+
bwD
bw+
ON
bw+
bw+
OFF
HET
EU
C
bwD
bw+
bwD
bw+
Yes?
bwD
bw+
No?
bw+
bwD
bwD
bw+
D
bwD
bw+
bwD

Figure I-5. Different meanings of "pairing." We refer to many forms of contact as pairing. a) Pairing of chromosome arms in somatic cells is dependent on sequence similarity. b) In polytene cells, different repetitive sequences (represented by zig-zags, squiggles, etc.) co-segregate into a chromocenter. c) Repeated sequences, both inverted (left) and direct (right) may pair intrachromosomally. d) Ectopic intrachromosomal contacts may also occur between distant sequences of high, moderate, dubious, or no sequence homology.

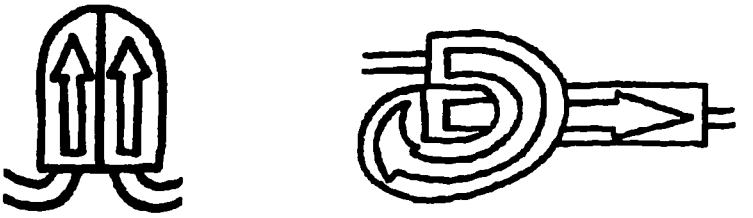
A



B



C



D

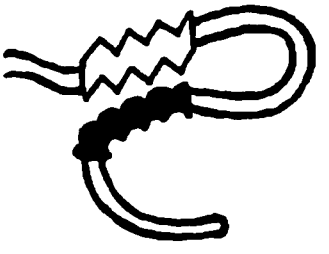


Figure I-6. Pairing mechanisms. There are many levels on which recognition of similar genetic regions may occur. a) Watson-Crick base pairing (hatch marks) may form heteroduplex structures, even in the absence of chromosome breakage; these may be short and "naked," i.e. non-nucleosomal (top) or they may be packaged and stabilized by nucleosomes (bottom; circles=nucleosomes). b) B-DNA regions differ considerably in twist and writhe; similar strands may stabilize each other's conformational preferences. c) One model for 30 nm fiber topology suggests that it is no more than the conformationally- preferred state of nucleosome-bound DNA. If so, then these open "zig-zag" structures would also be stabilized by alignment of like regions, albeit on a larger scale (circles=nucleosomes).

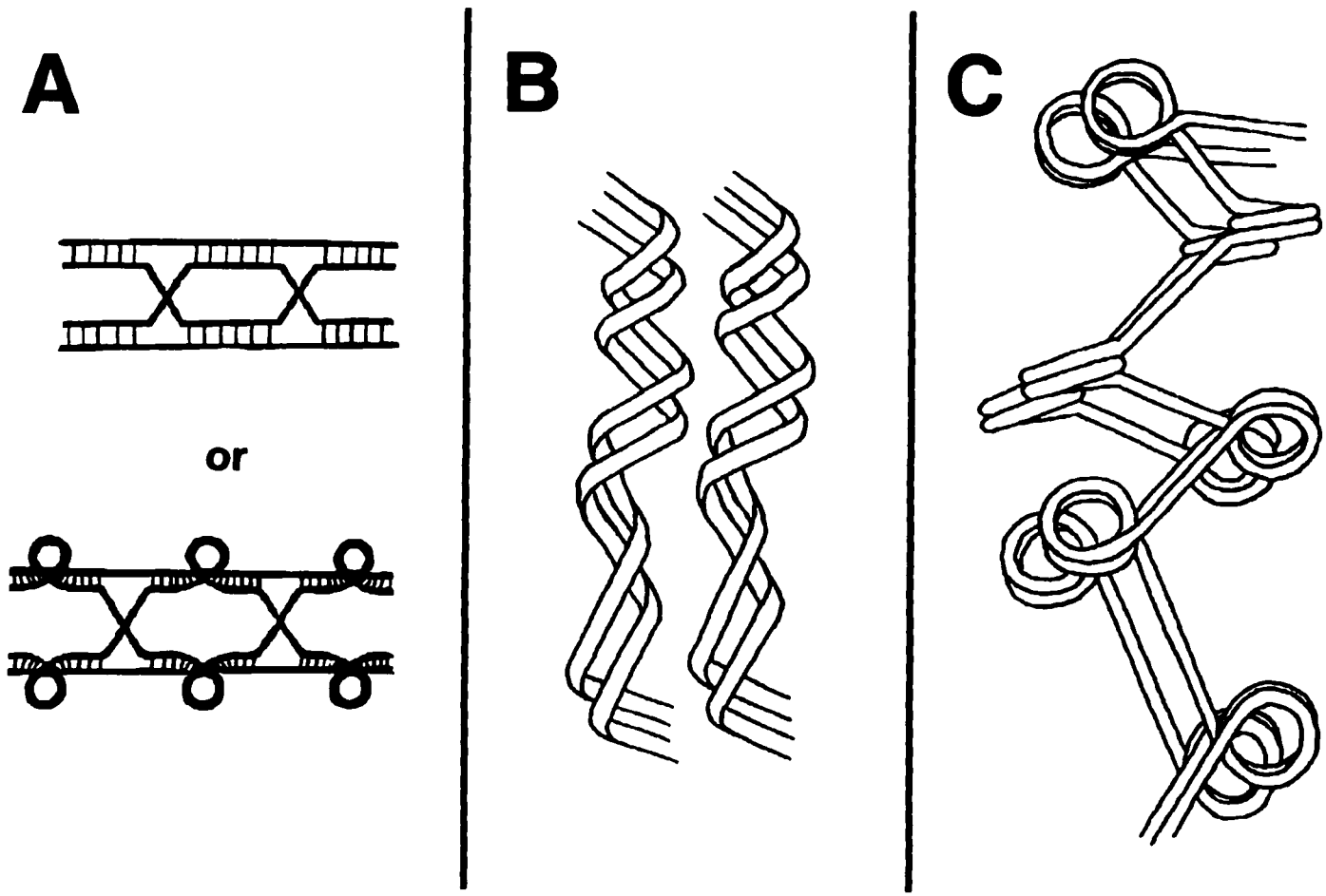
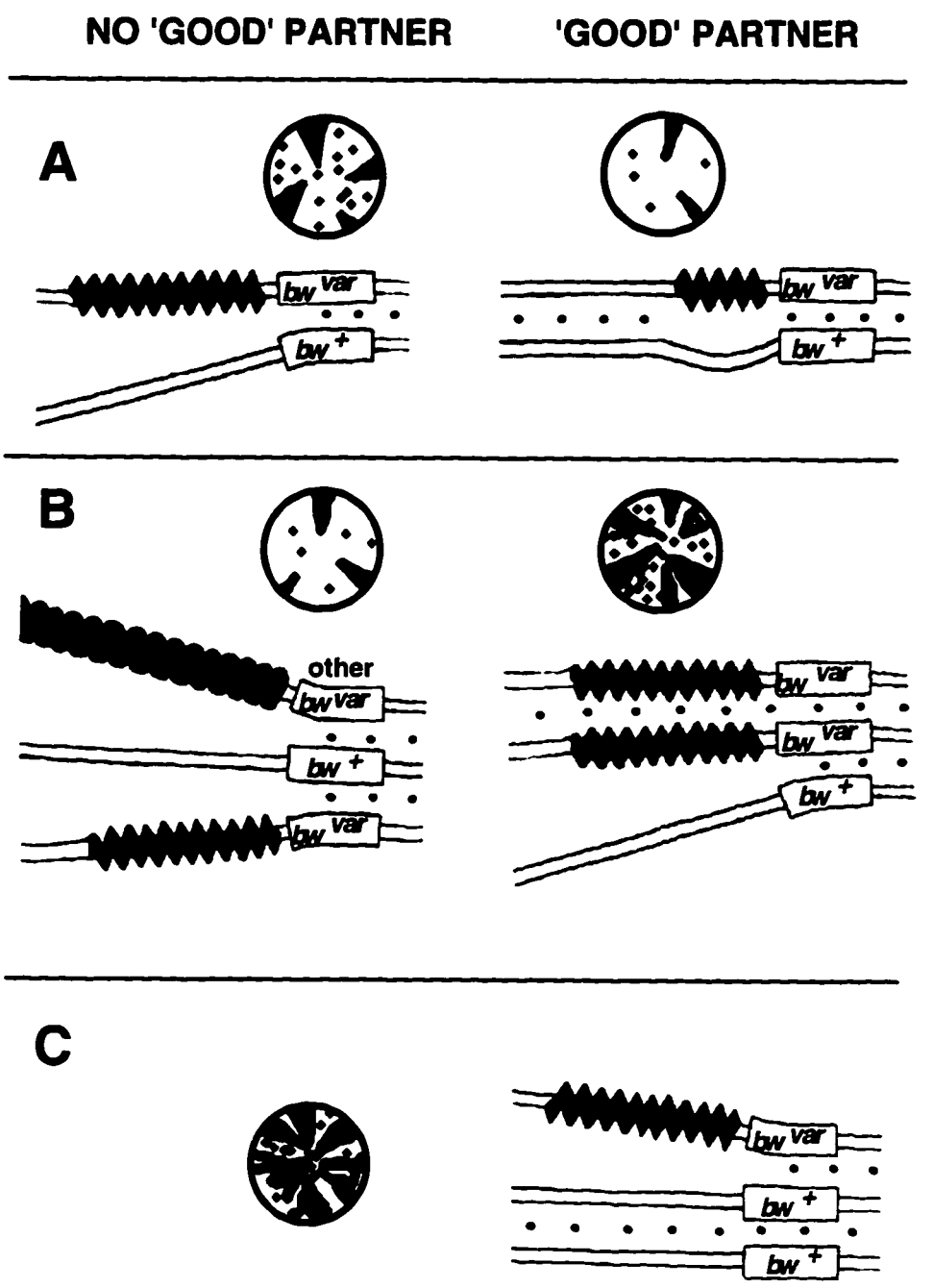

A
or
B
C

Figure I-7. Pairing and dominant PEV. Lines represent chromosomes, circles represent the resulting eye phenotype. a) Some of the variability in *trans*-inactivation can be explained by the relative degree of pairing achievable by the "homologous" chromosomes. PEV is induced by the introduction of heterochromatin (zig-zag lines) near a gene (rectangle) in euchromatin (straight line). When the disruption is large, homologous pairing (shown by dots) is impeded, and the homologue sometimes escapes inactivation (left); when the disruption is smaller, the homologue is more consistently paired, and thus more often inactivated (right). b) Attempts have been made to explain the phenotypes seen for multiple alleles (e.g. in triploid flies) using arguments based on competitive pairing. At first, this seems to work. Left: two different variegating alleles *trans*-inactivate a wild-type allele strongly (two chances to pair in every cell). Two copies of the same variegating allele (left) are less effective than two different variegating alleles (left) and less well than a single copy of the variegating allele (above), as if the two identical alleles tend to pair and sequester each other, allowing the unpaired wild-type allele to be expressed. c) However, competition cannot be absolute, because a single variegating allele can effectively *trans*-inactivate two wild-type alleles in the same cell.


NO 'GOOD' PARTNER
'GOOD' PARTNER
A
bw var
bw +
bw var
bw +
B
other
bw var
bw +
bw var
bw var
bw var
bw +
C
bw var
bw +
bw +

CHAPTER II

Copy Number and Orientation Determine the Susceptibility of a Gene to Silencing by Nearby Heterochromatin in *Drosophila*

INTRODUCTION

The organization and spatial regulation of the diploid interphase nucleus is largely a mystery. One of the few obvious aspects of large-scale genomic organization is the segregation and condensation of certain genomic regions into a dense nuclear aggregate termed heterochromatin (Heitz 1928; 1929). Most eukaryotic cells contain heterochromatin-rich regions near their centromeres, and, often, near their telomeres (John 1988). Sex-chromosomes may be constitutively or facultatively heterochromatic (Lyon 1992; Speed *et al.* 1993) and supernumerary B-chromosomes in plants and animals are also largely heterochromatic (Jones and Rees 1982). An enigmatic feature of heterochromatin is that when euchromatic genes are moved nearby, they can show mosaic expression, or position-effect variegation (PEV) as described by Muller (1930; 1932) and reviewed by Lewis (1950), Spofford (1976) and Weiler and Wakimoto (1995).

Because PEV can affect multiple genes over considerable distances, it has become a paradigm for long-distance genomic interactions, and has also served as a genetically tractable model for that most extreme example of

facultative heterochromatin formation, the condensation of mammalian X chromosome.

Following the prescient proposal that chromosomes might contain internally repeated domains (Bridges 1935; see appendix A), Ephrussi and Sutton (1944) proposed that spreading occurs because homologous pairing interactions between heterochromatic repeats disrupt the structure of the chromosome, affecting nearby genes (reviewed by Henikoff 1994). Later, based on the (probably false) idea that variegating breakpoints exert a polar silencing effect by encouraging the continuous propagation of a gene-silencing complex, it was argued that spreading is instead controlled by discrete *cis*-acting elements, including sequences in constitutive heterochromatin that initiate spreading and possibly terminators that limit spreading (Tartof *et al.* 1989; Eissenberg 1989; Grigliatti 1991; Moehrle and Paro 1994). However, there is no evidence that such elements exist. Moreover, exceptions to the continuity of compacted chromatin (Belyaeva *et al.* 1993; Belyaeva *et al.* 1993) challenge the assumption that spreading occurs by linear propagation of an altered chromatin conformation. A further challenge to this model springs from observations of cooperativity between discontinuous blocks of heterochromatin, as assayed by effects on variegating heterochromatic genes (Wakimoto and Hearn 1990; Eberl, Duyf and Hilliker 1993) and variegating euchromatic genes (Talbert, LeCiel and Henikoff 1994; Henikoff, Jackson and Talbert 1995; see also chapter III).

PEV is the preferred system for studying heterochromatin in part because *trans*-acting modifiers of PEV encode proteins involved in heterochromatin formation (Henikoff 1979; Reuter and Wolff 1981; Moore, Sinclair and Grigliatti 1983; James and Elgin 1986; Locke, Kotarski

and Tartof 1988; Reuter *et al.* 1990; Garzino *et al.* 1992). These include *Suppressor-of-variegation* loci such as *Su(var)205* (Grigliatti 1991), which encodes HP1, a protein component of heterochromatin (James and Elgin 1986; Powers and Eissenberg 1993). Genic modifiers of PEV provide insight into the proteins responsible for initiating and maintaining heterochromatin and also provide simple tools for assaying and manipulating PEV. We know less about genic and non-genic *cis*-acting modifiers of PEV, especially those located in heterochromatin. A complementary genetic approach might focus on the isolation of these modifiers. However, the large size of heterochromatic blocks and their repetitive sequence composition compromise standard genetic mapping.

Recently, a new genetic approach to the study of heterochromatin was introduced (Dorer and Henikoff 1994). Repetitive sequences were generated as closely linked duplications of a transposon carrying a mini-*white* eye pigment gene. Repeat arrays of three or more copies showed PEV mutant phenotypes that strengthened with increasing copy number and that were responsive to *trans*-acting modifiers of PEV. This result implies that heterochromatin forms because of the repetitive nature of the sequences found there, not because of the presence of any specific sequences.

Although artificial repeat arrays provide a model for heterochromatin formation, they do not address the mechanism of spreading associated with PEV. In part, this is because the same repeat arrays include both the heterochromatin initiation point and the reporters at which heterochromatin acts, so that spreading along the chromosome cannot be assessed. In addition, there are differences between heterochromatin

formation in mini-*white* repeat arrays and in classical PEV. For example, classical PEV involves chromosome breaks that juxtapose euchromatin and natural heterochromatin. In the case of mini-*white* repeat arrays, neither chromosome breaks nor natural heterochromatin was involved. Moreover, the mini-*white* reporter gene lacks the normal *white* regulatory elements and therefore is especially sensitive to position effects (Kellum and Schedl 1991; Roseman, Pirrotta and Geyer 1993; Chung, Whiteley and Felsenfeld 1993).

Here we report an example of classical PEV in which an X-ray induced chromosome rearrangement has placed natural heterochromatin adjacent to the genomic sequence of a *brown* eye pigment gene with its normal regulatory elements (Dreesen, Henikoff and Loughney 1991; Martin-Morris *et al.* 1993). Manipulation of this reporter gene using P transposase caused striking effects on PEV. These phenotypic changes resulted from local alterations in transposon copy number and orientation. Our findings favor a somatic pairing model over current linear propagation models to explain PEV.

MATERIALS AND METHODS

Fly strains and culture conditions: All *Drosophila melanogaster* stocks were kept at room temperature on standard cornmeal molasses agar medium. Crosses were performed at 25°C. Mutants not described below are described by Lindsley (1992). Unless otherwise noted, all lines carry $brown^{Dominant}$ (bw^{D}) and *scarlet* (*st*) giving a white background for easy scoring of changes in expression from bw^{+} transposons. *P[Δ2,3(99B)]* is a

third chromosome insertion that encodes P transposase (Robertson 1988). The line *P[bw+]* 92C is one of 32 previously reported non-variegating lines (Dreesen, Henikoff and Loughney 1991) produced by injection and subsequent mobilization of P-element constructs containing a wild-type genomic copy of the *brown* gene in the pDm24 vector (Mismer and Rubin 1987); a moderately variegating derivative, *V21*, was produced by X-irradiation (Dreesen, Henikoff and Loughney 1991). *V21* is a translocation that places 2R heterochromatin ~55-70 kb from the *P[bw+]* transposon (Figure II-1a; J. Sabl, unpublished results). *V21* derivatives produced by P transposase mutagenesis are labeled sequentially by phenotype, *i.e.* (E)nhanced, (R)everted to *V21*-like or (S)uppressed, and isolate number. For example, *V21-E2E1R1* is the product of three rounds of transposase mutagenesis with selection for enhancement (*V21-E2*), further enhancement (*V21-E2E1*), and finally reversion to a *V21*-like phenotype. Transposon orientation is indicated in parentheses after the allele name, *e.g. V21(L), V21-E1(LR).* Orientations are listed beginning with the most proximal transposon; "L" indicates that *brown* transcription is directed towards the translocation breakpoint. Transposase-generated derivatives of *V21* are denoted as a group by *V21**, and are all marked with *Sb*. Transposase derivatives of *92C* are denoted as a group by *92C**, and are unmarked except by *brown*$^{+}$ expression.

Mutagenesis: Transposase screens were as follows. Males hemizygous for *V21* or *V21** or homozygous for *92C* were crossed to *P[Δ2,3(99B)]* virgin females (*Δ2,3*). The *92C/Δ2,3* and *V21*/Δ2,3* males were crossed individually or in pairs to 5-7 *bw*^{D}*; st* virgin females, and transferred at 4-7 day intervals for 2 weeks. For *92C* crosses, male progeny were aged at least 5 days and scored for increased pigment intensity; for *V21** crosses, male

progeny were scored for degree of variegation. Putative mutants were crossed to *bw^D; st* females and maintained by selection. Several concerns influenced the scoring and stocking of putative events. Unwanted events (transposon jumps and mutations in second-site modifiers of PEV) were recovered frequently. Many putative lines were kept for 2-4 generations to determine linkage before Southern analysis. In the process, some linked lines may have been lost. Additionally, no more than four phenotypically similar events were kept from any one vial. Because of this, frequency estimates are all lower limits. Multiple events from the same vial were scored as independent only if different by Southern analysis. The screen that produced *V21-E2(LL)* and *V21-E3(LR)* used a *bw$^+$; st$^+$* strain of *Δ2,3*. All other screens were done in a *bw^D; st* background.

The translocation screen used 2-10 day old males hemizygous for the *V21-E2E1(multi)* translocation chromosome marked with *Sb*. These were X-irradiated (3,000 rads) and groups of 6 were immediately crossed to 30 *bw^D; st* virgin females. After 3-1/2 days males were removed to insure that only independent events were recovered. Females were transferred after 7 days. Recovery rates of linked suppressors were calculated relative to the total number of *Sb* flies. These lines were designated *V21-(multi)RX*.

Testing with enhancers of PEV: To see if non-variegating lines could be induced to variegate, females from several *92C** and *V21-(multi)RX* lines were crossed to males carrying second-site enhancers of variegation *E(var)39A, E(var)8* and *E(var)66* (Locke, Kotarski and Tartof 1988). For second chromosome enhancers, the test class produced was *w$^+$/w^{m4}* or *w$^+$/Y; Enhancer, bw$^+$/bw^D; 92C** or *V21*, st/st$^+$*. For third chromosome

enhancers, the enhancer was on the st^+ chromosome, but the test class was otherwise identical. The endogenous *brown*$^+$ gene was inactivated by bw^D. Weak *brown* variegation is generally easier to detect in a st^+ background.

Cytogenetic analysis: Neuroblast chromosomes were prepared essentially as in Gatti, Pimpinelli and Santini (1976) without colchicine, and were then stained in 1 μg/ml 4',6-diamidino-2-phenylindole (DAPI) 4-8 min, and mounted under 90% glycerol, phosphate-buffered saline, 1 mg/ml phenylene-diamine (Johnson and Nogueria-Araujo 1981) and photographed with TechPan film (Kodak) through a 100X planapochromatic oil immersion lens under ultraviolet epifluorescent illumination (Nikon UV-1A filter).

Hybridization to polytene chromosomes (Simon *et al.* 1985) with plasmid, lambda or cosmid DNA was done using randomly primed templates for incorporation of biotinylated dUTP (Bethesda Research Laboratories), and detected with the Detek-1-HRP kit (ENZO Biochemical).

DNA preparation and analysis: *D. melanogaster* genomic DNA was purified by the method of Bender *et al.* (1983) as modified by J. Hirsh. Restriction digestion and electrophoresis (0.28%-1.5% agarose in 1 X TAE) were carried out by standard methods (Sambrook, Fritsch and Maniatis 1989). For standard pulsed field gel electrophoresis, ovary DNA plugs were prepared and digested essentially as described by Karpen and Spradling (1990) for larval disks and brains. Blotting [without neutralization to Biodyne-B (GIBCO)] and radiolabeling with random primers were done by standard methods.

RESULTS

An increase in local repetitiveness enhances PEV

Selection for enhanced variegation yields transgene duplications: This work began with the fortuitous recovery of an extreme allele derived from a *brown*-variegating translocation line. The original line, referred to as *V21* (Figure II-1a), shows patchy pigmentation over about half of the ommatidia (Figure II-1b, upper left). This line had been generated by X-ray mutagenesis of flies carrying a *P[bw+]* transposon inserted into 92C (*P[bw+]92C*). The extreme allele, *V21-E1* (previously called "*V21^{e}*"), shows only scattered spots and patches (Figure II-1b, lower left). Both *V21* and *V21-E1* carry the same translocation between chromosome 2R heterochromatin and chromosome 3 broken within 92B (Dreesen, Henikoff and Loughney 1991, data not shown). The juxtaposition of 2R heterochromatin and the transposon causes classical PEV of the *brown+* gene present on the transposon. Both variegating lines had transposable, *brown+*-expressing, and therefore presumably intact P-elements, so that the *V21-E1* phenotype appeared to result from *cis*-enhancement of the variegating position effect in *V21* (Henikoff, Loughney and Dreesen 1993). Structurally, *V21-E1* (like *V21*) is an example of classical PEV, because it was induced by chromosome rearrangement; functionally, its classical nature was confirmed in a screen for *Su(var)*s: 90% of the modifiers that suppressed *V21-E1* also suppressed *white-mottled-4* (Talbert, LeCiel and Henikoff 1994), a prototypic PEV allele (Reuter and Spierer 1992).

To investigate the basis for this unexpected *cis*-enhancement of the *V21* phenotype, we compared the extreme allele both to its parent allele and to

the original non-translocated line, *P[bw+]92C*. Southern analysis revealed no differences among the three chromosomes for ~15 kb to either side of the original insertion (data not shown). This suggests that enhancement was not caused by a lesion in nearby sequences. Rather, an alteration involving the *P[bw+]* transposon was seen in the extreme line. The entire transposon was duplicated at the original insertion site.

Transposase mutagenesis of *V21* produced three additional linked enhancers from 1260 progeny. All were similar in phenotype to *V21-E1* (Figure II-1b), and all were associated with closely linked duplications of *P[bw+]* in tandem or reverse orientation as demonstrated by Southern analysis (Figure II-2 b-c and Figure II-4c). This perfect correlation between enhancement and duplication (Figure II-3) is evidence for a cause-effect relationship.

Enhancement is revertable: Direct proof that duplications are responsible for enhancement requires that removal of the extra copy of *P[bw+]* restore the original *V21* phenotype. To this end, we re-mutagenized two of the duplication lines by exposure to P transposase, screened for any phenotypic change, and analyzed the resulting lines for transposon number and structure.

Revertants, which are flies with a moderate-to-dark variegating phenotype approximating that of *V21*, were found at high frequency (see Table II-I and Figure II-3). For one line, *V21-E2(LL)*, 10 revertants from 690 flies were characterized. For the other line, *V21-E3(LR)*, three revertants from 70 flies were characterized. All revertant lines had a single copy of the *brown+* transgene in the same location and orientation as in *V21*. We

conclude that for both (LL) and (LR) alleles, duplications of *P[bw+]* cause enhancement of the variegating phenotype.

In addition to revertants, other phenotypic alterations were found. A more extreme derivative of *V21-E2(LL)* had two additional copies of the *brown* transgene at the site (Figure II-2). This line, *V21-E2E1(multi)*, showed pigment spots only on occasion. *V21-E2E1(multi)* was subjected to transposase mutagenesis; it, too, reverted to a *V21*-like phenotype or to an intermediate phenotype at high frequency. Of the five derivatives of *V21-E2E1(multi)* that were characterized, all three *V21*-like revertant lines had single intact copies of the *brown* gene, whereas both of the intermediate lines had 2-3 approximately intact copies of *P[bw+]*. These results with *V21-E2E1(multi)* further demonstrate that increased gene copy number leads to decreased frequency of expressing cells.

PEV is sensitive to gene orientation

Orientation changes susceptibility to PEV: The re-mutagenesis of *V21-E3(LR)* provided the opportunity to examine the effect of PEV on the same transgene inserted in either orientation at the same location. Precise excision of the (R) element from *V21-E3(LR)* results in reversion to a *V21*-like (L) transposon, whereas precise excision of the (L) element in effect reverses the orientation of the transposon at the site. We had expected that the two orientations would lead to the same phenotype. Surprisingly, this was not so. All revertants to the *V21(L)*-like phenotype had lost the (R) copy of the element, reverting to the original (L) orientation. Additionally, flies with nearly wild-type eyes were also

retrieved (Figure II-4). Upon analysis, these were found to be examples of the complementary event: they contained a single transposon in the (R) orientation. Repetition of this screen with intentional selection for the two classes confirmed that phenotype is a perfect predictor of transposon orientation, with 6 of 6 lines correctly classified for each orientation (Table II-I).

Orientation dependent phenotypes were also seen for tandem double insert lines. Transposase mutagenesis of *V21-E3S1(R)* produced lines with a weakly enhanced phenotype not previously observed. These were found by Southern analysis to carry two tandem elements in the (RR) orientation (Figure II-4 a,c). Although these (RR) lines are enhanced relative to (R) lines, they are strongly suppressed relative to (LL) lines (Figure II-4b).

Mutagenesis of *V21-E3S1(R)* also yielded lines with strongly enhanced phenotypes characteristic of (LL) and (LR) lines and lines with moderate phenotypes characteristic of (L) lines (like *V21*). Southern analysis revealed that for 10 lines with unambiguous structures, copy number and orientation were consistent with results of previous mutageneses (Figs. II-2c, II-4c and data not shown). This ability to generate lines with the same phenotypes from different starting points further confirms that the phenotypic effects depend solely on copy number and orientation of *P[bw$^+$]*.

Variegation requires adjacent heterochromatin: Our observations that PEV strengthens with increasing copy number parallel those reported previously for arrays of mini-*white* transgenes (Dorer and Henikoff 1994). An important question is whether arrays of *P[bw$^+$]*, like arrays of

mini-*white*, autonomously form heterochromatin at sites that are distant from heterochromatic regions, so that silencing is intrinsic to the array. In the case of *P[bw+]*, arrays were produced at the site of a classical PEV allele that juxtaposes euchromatin and heterochromatin via a chromosomal rearrangement. Therefore, our interpretation is that these *P[bw+]* arrays primarily enhance classical PEV rather than create new blocks of heterochromatin as in the case of mini-*white* repeat arrays. To address this, we asked whether *P[bw+]* duplications could form heterochromatin autonomously at the original 92C site of insertion, which is distant from heterochromatin.

We followed the same protocol used for mini-*white* to expand repeat arrays by transposase mutagenesis of the original (un-rearranged) *P[bw+]* 92C chromosome. A single *P[bw+]* copy at 92C (and most other sites) gives rise to dull red, rather than full red eyes (Dreesen, Henikoff and Loughney 1991). This mild hypomorphic effect allowed us to screen for duplications at the site by selecting flies with full red eyes. Of these, mutants that showed linkage to 92C were subjected to Southern analysis. This procedure led to the recovery of direct (LL) and reversed (LR) *P[bw+]* duplications at a rate of approximately 1%. These duplication lines did not variegate; nor were any variegating lines recovered. No (RL) derivatives of *92C* (or *V21*) were ever characterized; 3'-to-3' P-transposon duplications are known to occur only rarely (Zhang and Spradling 1993).

Three of the 92C *P[bw+]* duplication lines were tested with the *Enhancer-of-variegation* mutations *E(var)66* and *E(var)8* (Locke, Kotarski and Tartof 1988); even then, no variegation was observed. Additional rounds of transposase mutagenesis on the tandem lines also produced no

variegators, even though more than 1230 flies were scored. We conclude that *P[bw+]* duplications and small repeat arrays at 92C do not have intrinsic silencing ability. Instead, *P[bw+]* repeats seem to increase the effect of heterochromatin from the PEV-inducing rearrangement ~55-70 kb away.

In a complementary approach, we asked whether removal of the bulk of centric heterochromatin from the vicinity of the largest *P[bw+]* array would fully revert the variegating phenotype. X-ray mutagenesis of *V21-E2E1(multi)* gave suppressed F1 progeny at a rate of approximately 0.5%; of these, almost 20% (10 flies) were fully wild-type (Figure II-5b). Southern analysis revealed that the transgene array was undisturbed in all five full and all four partial revertant lines examined (data not shown). All three of the full revertant lines tested displayed uniform eye pigmentation in the presence of *E(var)66*. Polytene and mitotic cytological analyses on one of these lines, *V21-E2E1X31*, revealed that an inversion had occurred with breaks at 92B and 98D, removing the *P[bw+]* array at 92C from the vicinity of heterochromatin. Therefore complete removal of heterochromatin from an extreme variegating array results in full reversion of the phenotype, further suggesting that *P[bw+]* repeat arrays act to enhance PEV without forming new heterochromatin.

Another full revertant line insensitive to *E(var)66*, *V21-E2E1X98*, proved to be especially interesting. Linkage analysis showed that the *P[bw+]* array was no longer present on a translocation chromosome, confirmed by examination of chromosome 3 in neuroblasts, which showed no visible abnormalities (data not shown). Examination in polytene salivary gland nuclei revealed that this chromosome had "healed," restoring a nearly

normal third chromosome but leaving behind a visible block of heterochromatin at 92B5-10 (Figure II-5a). In other words, full pigmentation is produced in the *X98* line that retains the full multi-copy array, the cytologically visible heterochromatin and the intervening euchromatin of *V21-E2E1*. This result is reminiscent of the many complete revertants of the w^{m4} PEV mutation that were found to involve one break in heterochromatin and another break in euchromatin (Tartof, Hobbs and Jones 1984; Reuter, Wolff and Friede 1985). Similarly, suppression of a PEV phenotype by movement of a sensitive reporter gene away from pericentric heterochromatin has been seen for *trans*-inactivation of the *brown* gene by the $brown^{Dominant}$ (bw^{D}) heterochromatic element (Talbert, LeCiel and Henikoff 1994; Henikoff, Jackson and Talbert 1995). There, suppression was found to correlate with a decreased frequency of looping of bw^{D} into the chromocenter of polytene chromosomes (and *vice versa* for enhancement), lending support to the notion that associations with the heterochromatic compartment of the nucleus underlie PEV.

In addition to the gross rearrangements that modify PEV, more precise changes can have a striking effect on variegation. Recently, it has been shown that PEV is triggered by P transposase-induced deletions that bring a single $white^{+}$ transgene close to an isolated block of 2L heterochromatin (Howe *et al.* 1995). We wondered whether the trapped heterochromatin in the healed *X98* line could similarly induce PEV. Accordingly, we mutagenized *X98* with P transposase and recovered a weakly variegating mutant from ~550 red-eyed progeny (Figure II-5b). Polytene chromosome analysis revealed no gross alterations resulting from transposase treatment; however, Southern analysis showed that a deletion had removed intervening sequences, evidently fusing the transgene array to

heterochromatin (data not shown). Therefore, proximity of the array to a heterochromatic block influences PEV sensitivity. This inference made on a molecular scale extends previous conclusions based on large-scale cytology (Wakimoto and Hearn 1990; Eberl, Duyf and Hilliker 1993; Talbert, LeCiel and Henikoff 1994; Henikoff, Jackson and Talbert 1995).

DISCUSSION

Repeat arrays and orientation effects

Natural and artificial repeat arrays can interact in forming heterochromatin: We have found that classical PEV, in which a rearrangement causes silencing of a euchromatic gene by juxtaposing it to heterochromatin, can be strikingly altered by local changes at the gene. Tandem and inverted duplications of a *brown*$^+$ (*bw*$^+$) transgene strongly enhance PEV, and longer transgene arrays enhance PEV still more strongly. To an extent, our results with the full *bw*$^+$ genomic sequence parallel previous findings that transposon repeat arrays become susceptible to heterochromatic inactivation (Dorer and Henikoff 1994). These effects are unlikely to involve any feature specific to either the sensitive mini-*white* or the fully functional *bw*$^+$ reporter gene, indicating that heterochromatic silencing of reporter genes is a general property of repeat arrays. Fittingly, *white* and *brown* have been the most intensively studied genes in PEV research, beginning with Muller's discovery of variegation (1930).

Repeat arrays of the *bw*$^+$ transgene are insufficient to induce heterochromatic effects autonomously, but rather require the presence of

nearby natural heterochromatin, as is the case for classical PEV. This finding seems at odds with the autonomy demonstrated by mini-*white* repeat expansions. However, mini-*white* repeats do not appear to be completely autonomous, because an array closer to heterochromatin was more strongly affected than an array farther away (Dorer and Henikoff 1994), a correlation that has been extended to rearranged derivatives of array-bearing chromosomes (D. Dorer, personal communication). Therefore, the difference between *bw*$^+$ and mini-*white* transgene arrays might be that *bw*$^+$ arrays are inherently less sensitive to PEV silencing by nearby heterochromatin. Reduced sensitivity would result if insulator elements are associated with *P[bw*$^+$*]* but not with mini-*white*. The presence of insulator elements tightly associated with the *brown* gene is implied by the insensitivity of all *P[bw*$^+$*]* transposons to position effects caused by insertions at various euchromatic sites (Dreesen, Henikoff and Loughney 1991; Martin-Morris *et al.* 1993). In contrast, mini-*white* lacks insulator elements, as evident from its extreme sensitivity to euchromatic position effects (Kellum and Schedl 1991; Roseman, Pirrotta and Geyer 1993; Chung, Whiteley and Felsenfeld 1993). Addition of insulator elements protects mini-*white* from euchromatic position effects and provides partial protection from heterochromatic inactivation at sites very close to heterochromatin (Roseman, Pirrotta and Geyer 1993). That is, mini-*white* with insulator elements resembles the 8.4 Kb *bw*$^+$ genomic segment both in euchromatin and heterochromatin. One known insulator element, scs', corresponds to two DNase I hypersensitive sites (Udvardy, Maine and Schedl 1985) that map exactly to the shared promoter region of two divergently transcribed genes (Glover *et al.* 1995). Insulation may be a common feature of *Drosophila* promoter regions.

At mini-*white* repeat arrays, strong pairing interactions based on sequence identity have been proposed to nucleate heterochromatin formation (Dorer and Henikoff 1994). Similarly, *P[bw+]* arrays would be subject to the same forces and participate in the same kinds of pairing interactions. The *brown* insulator sequences would merely inhibit later steps in heterochromatin formation, such as the assembly of *Su(var)*-encoded proteins.

Orientation dependence suggests sequence-specific interactions with heterochromatin: An unexpected finding was that reversing the orientation of the transposon carrying a *bw+* gene modified the variegating phenotype. Differences were seen both for single copies and for tandem duplications. These differences cannot be explained by the proximity of heterochromatin to the promoter, because the promoter is slightly farther from heterochromatin in the more strongly inactivated lines. Similarly, they cannot be explained by imagining an insulating block specific to either end of the element. For example, a partial block 5' to the transcription unit could indeed account for the observed differences between (L) and (R): such a block would protect (R) but not (L) from PEV. However, in such a case, both (LL) and (LR) two-copy lines would have the more distal copy protected, and therefore be dark in phenotype, rather than light.

Pairing properties might contribute to some of the phenotypes, explaining why (LR) lines resemble (LL) lines, whereas (RR) lines are darker. For single transposons, flipping from (L) to (R) gives a darker phenotype. This leads to the expectation that (LR) lines should be darker than (LL) lines. However, a counter expectation, based on the phenotypes of mini-*white* repeat arrays, is that pairing within (LR) lines should be stronger than

pairing within (LL) lines, so that (LR) lines should be lighter than (LL) lines. We suppose that these conflicting effects cancel one another.

Weaker pairing properties could account for the orientation effect, in which (R) lines are darker than (L) lines, and (RR) lines are darker than (LL) lines. It is possible that limited stretches of sequence similarity exist between the transposon and nearby heterochromatic sequences, and occasional pairing occurs. Another possibility is that stretches of chromatin similarity exist, such as similar patterns of nucleosome phasing or similarly ordered DNA-binding proteins, leading to pairing of heterologous sequences. These possibilities cannot be distinguished given our lack of understanding of the mechanism that underlies somatic pairing. In either case, the frequency of pairing would be greater for anti-parallel than for parallel orientation, resulting in a higher frequency of *brown* gene inactivation.

Another possible explanation for orientation dependence arises from an unusual feature of the *brown* gene, that heterochromatin can inactivate the gene in *trans*, causing dominant position-effect variegation. This dominant effect might be mediated by a heterochromatin-sensitive transcription factor which makes direct contact with heterochromatic proteins (Dreesen, Henikoff and Loughney 1991; Martin-Morris *et al.* 1993; Martin-Morris and Henikoff 1995). This hypothetical transcription factor would then interact with heterochromatin in a particular topological orientation. However, recent results show that mini-*white*, which lacks transcription factor binding sites, can be *trans*-inactivated (D. Dorer, personal communication), undermining this explanation.
Although other models might be proposed to explain orientation

dependence, this effect is not explained by sequence-independent properties of chromatin. That is, simply reversing the orientation of a DNA segment without changing the distance to heterochromatin should not affect the distribution of chromatin proteins that lack sequence specificity. Rather, we suggest that orientation dependence of single and tandem copies is understandable as a consequence of the inherent polarity of DNA sequences. Relevant to this suggestion, bacterial site-specific recombination can show an orientation dependence over distances of more than 10 kb (Howe and Schumm 1981). Recombination studies *in vitro* suggest that orientation effects do not involve linear propagation or tracking; rather, topology constrains looping interactions between similar sequences (Gellert and Nash 1987).

A pairing-looping model for PEV

Current models for PEV are motivated by the well-established polarity of the effect, in which genes close to a heterochromatic break are more strongly affected than genes farther away (reviewed by(Spofford 1976). These models propose that the heterochromatic state propagates linearly along the chromosome from sites within constitutive heterochromatin (Zuckerkandl 1974; Tartof *et al.* 1989; Eissenberg 1989; Grigliatti 1991; Moehrle and Paro 1994). Chromosome breaks that join heterochromatin to euchromatin might allow propagation into euchromatin. To account for mosaicism of gene expression, linear propagation must terminate variably either before or after that gene in different cells. In a single cell, the gene is active only if propagation has stopped in the 55-70 kb stretch proximal to the gene. Changes distal to the stopping point should not have any effect, unless it is imagined that propagation can be attracted or repelled from a distance. Yet we find that such distal changes are

extremely effective in modifying inactivation of the *brown* gene: for example, *V21-E3S1(R)* has eyes that are nearly wild-type with rare mutant patches, whereas *V21-E2E1(multi)* has eyes that are fully mutant except for an occasional eye with one or more wild-type spots. That is, changes distal to the hypothetical propagation stopping points produce phenotypes spanning the full observable range. This contradicts models that require continuous linear propagation of heterochromatin.

Rather than invoking linear propagation or action at a distance to explain our results, we suggest that silencing results from direct contact between heterochromatin and the affected gene. This contact could be mediated by the forces of somatic pairing, an interaction seen most vividly in the salivary gland polytene chromosomes. Polytene chromosomes display both homologous association of chromatids and non-homologous association of heterochromatic blocks into a chromocenter. We propose that sequences moved close to heterochromatin can associate directly with the chromocenter or heterochromatic compartment of the nucleus. Orientation dependence seen for single or tandem *P[bw+]* transposons suggests that such associations reflect pairing interactions between sequences on either side of the euchromatin/heterochromatin boundary. Increased silencing with increased size of the transposon array would reflect the tendency for repetitive sequences to pair, form recognizable structures (Dorer and Henikoff 1994), and co-localize with heterochromatin.

Our model can account for the bias in breakpoint distributions and frequencies seen for different variegating genes (Spofford 1976). This bias suggests that the large majority of breaks that place a reporter gene near

heterochromatin do not produce a variegating phenotype, and so are not selected in a screen. The middle repetitive elements found clustered in heterochromatin and scattered throughout euchromatin could provide the compatibility needed for mediating PEV (Figure II-6). When a chromosomal rearrangement moves a gene close to heterochromatin, then the net effect of pairing between any element near the gene and similar sequences clustered in heterochromatin will be to drag the gene close to the heterochromatic compartment, causing inactivation. This possibility also explains the surprising presence of middle repetitive heterochromatic sequences at PEV breakpoints (Tartof, Hobbs and Jones 1984), rather than the more abundant simple sequence DNA in heterochromatin (Lohe, Hilliker and Roberts 1993). Unlike middle repetitive elements, simple sequence satellites are rare or absent in *Drosophila* euchromatin and so cannot mediate contact between euchromatic sequences and the chromocenter. The model is also consistent with the lack of correspondence between the quantity of heterochromatin at a single breakpoint location and the severity of PEV (Howe *et al.* 1995), because specific repetitive sequences in heterochromatin that mediate looping presumably lie at varying distances from compatible euchromatic sequences in different lines. The model is economical, in that it proposes a single underlying mechanism, somatic pairing, for both heterochromatin formation and PEV.

Our model also can rationalize the binding pattern of antibody to the *Su(var)*-encoded HP1 protein. Anti-HP1 binds not only to the polytene chromocenter, but also to euchromatic sites, such as the banded regions of the fourth chromosome (James and Elgin 1986; Powers and Eissenberg 1993). Interestingly, this pattern is also seen for known middle repetitive

elements (Miklos *et al.* 1988). An HP1-containing protein complex might bind specifically to somatically paired structures that form at locally repetitive sequences (Dorer and Henikoff 1994), and this would favor looping associations with similar sequences nearby. Comparable local looping of DNA-protein complexes has been proposed to explain the spread of silencing by *Polycomb*-group proteins (Pirrotta and Rastelli 1994). In this way, unrelated gene silencing phenomena mediated by different *trans*-acting factors with different specificities might nevertheless involve analogous protein-protein interactions (Paro 1990).

In conclusion, we have shown that PEV is strikingly modified by purely local effects that can be understood in terms of somatic pairing forces acting on repeats to form heterochromatin (Ephrussi and Sutton 1944). Our results support the view that pairing associations organize the nuclear compartment and regulate the effects of chromosomal rearrangements on gene activity (Devlin, Bingham and Wakimoto 1990).

TABLE II-I
Transposase screens of *V21, 92C* and derived lines

Parent line	Mutant phenotype[a]	# independent events[b]	Total flies scored (vials)	Number of *P[bw+]* copies[c] (names)
V21	enhanced var.	4	1260 (31)	2 *(V21-E2, E3, E4)*
V21-E2	medium var.	>10	690 (12)	1
	super-en. var.	1		4[c] *(V21-E2E1)*
V21-E2E1	medium var.	>3	570 (16)	1
	enhanced var.	>2		2-3
V21-E3				
Screen 1	medium var.	>3[f]	70 (10)	1
	dark var.	2		1[e] *(V21-E3S1, S2)*
Screen 2	medium var.	>6[f]	192 (11)	1
	dark var.	>6[f]		1[e] *(V21-E3S1, S2)*
	en. dark var.	>2		2[e]
V21-E3S1	enhanced var.	>3	900 (22)	2
	en. dark var.	2		2[e]
	medium var.	2		1
V21-E2E1X98	weak var.	1	~550 (8)	4[d]
92C --->	full *brown+*	5	>510 (24)	2

[a]Relative to *V21*

[b]Number subjected to Southern analysis; phenotypically similar mutants not tested are indicated by ">". For all screens, *brown-* flies were recovered at a rate of 4-30%.

[c]At least one example with complete transposons; other examples have insertions or deletions no larger than 3 kb within transposon copies.

[d]Based on array size for this complex expansion.

[e]Inverted relative to V21.

[f]Numbers are for characterized lines; total frequencies were ~10X higher.

Figure II-1. Screens for enhancement of *V21*. (**A**) The *P[bw+]* transposon consists of the pDm24 vector (indicated by 3' and 5' transposon ends) containing 8.4 kb of *brown*+ genomic DNA (E-E) including a 3.8 Kb fragment (G-G) sufficient for rescue of *brown* (MARTIN-MORRIS *et al.* 1993). Restriction sites are indicated (N=*Nru*I, E=*Eco*RI, G=*Bgl*II, H=*Hin*dIII, B=*Bst*EII, K=*Kpn*I, S=*Sma*I). The non-cutters *Eco*RV and *Nhe*I were also used. The region used to probe Southern blots is indicated by shading. The *T(2;3)* breakpoint joining 92BC to 2R heterochromatin (zig-zag) was localized using overlapping P1 clones from the Berkeley Drosophila Genome Project, 2 lambda clones from Kate Loughney and a cosmid from a library generated by Mark Champe. Based on Southern analysis (with standard and pulsed-field electrophoretic gels) and *in situ* hybridizations, the heterochromatic break maps to P1 clones DS08505 and DS04082, and proximal to DS01965; the transposon insert site maps to the distal end of DS01965 and within DS00599. Arrowheads indicate the 3' and 5' transposon ends, and the wavy arrow represents the extent and 5'-to-3' orientation of the *brown* transcript. (**B**) Transposase mutagenesis produces *cis*-acting modifiers of PEV. Here and below, we refer to the various alleles by phenotypic effect, where "E" refers to enhancers, and (L) and (R) for left and right transgene orientation. *V21(L)* is at upper left; clockwise follow *V21-E2(LL)*, *V21-E3(LR)* and *V21-E1(LR)*. Curved arrows represent *Δ2,3* transposase mutagenesis; small arrows represent copy number and orientation (3' --> 5') of the transposon.

A

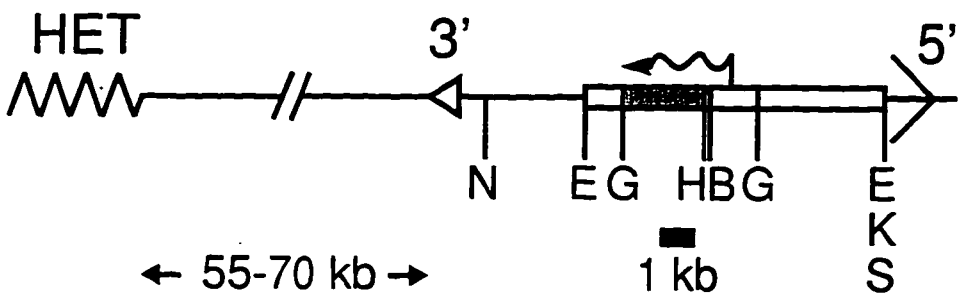


Figure II-3. Generation and interconversion of *V21*-derived alleles. Each dashed arrow represents the result of one round of transposase mutagenesis. Typical phenotypes are shown. In addition, flies with fully pigmented eyes (presumed mobilizations) were recovered at every step, demonstrating that at least one transposon and its *brown* gene always remain fully functional. Symbols and nomenclature are as in Figure 1, with "S" standing for suppressed and "DE" for dark-enhanced. (L) elements have 3' transposon ends oriented towards heterochromatin, and (R) elements are flipped.

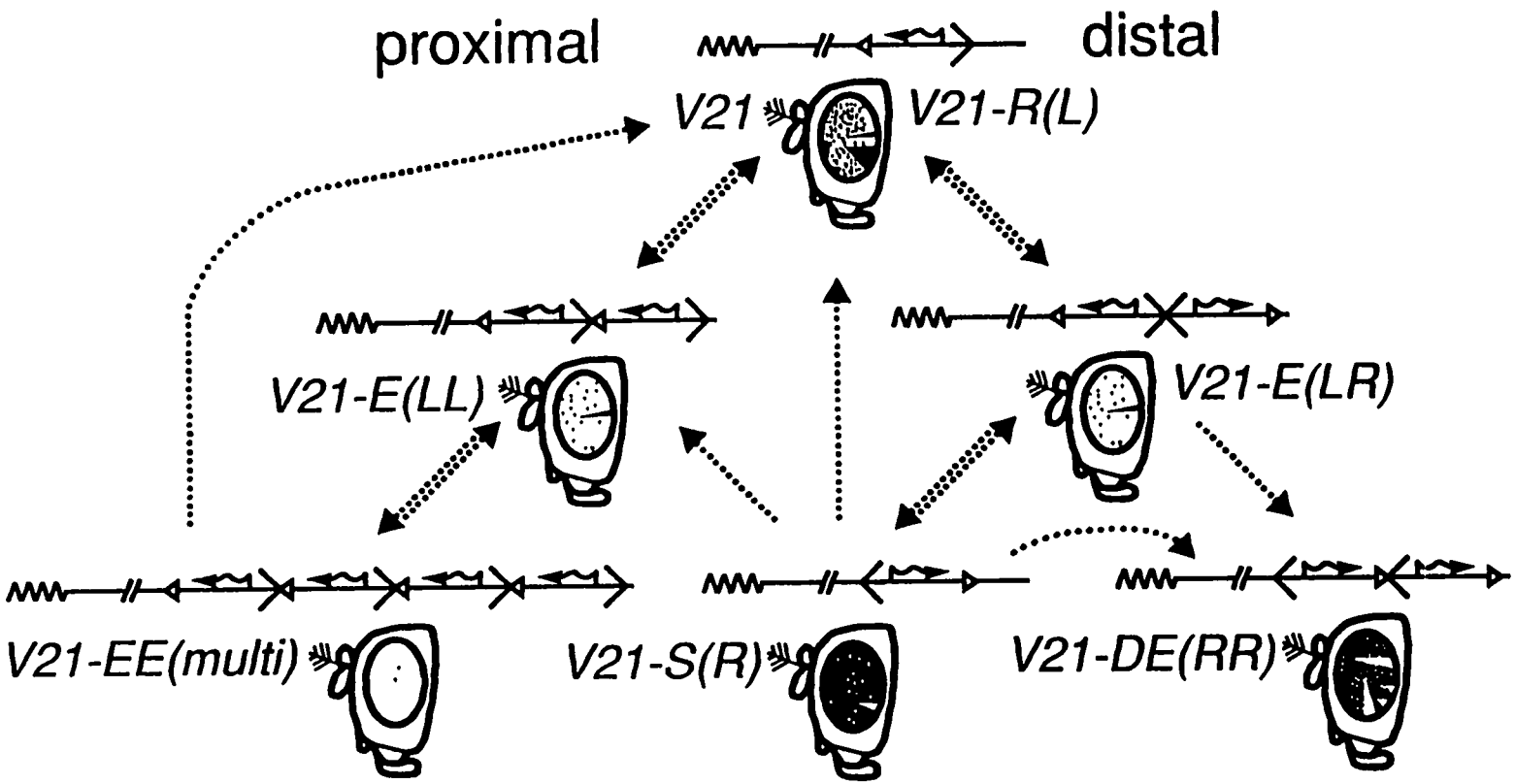
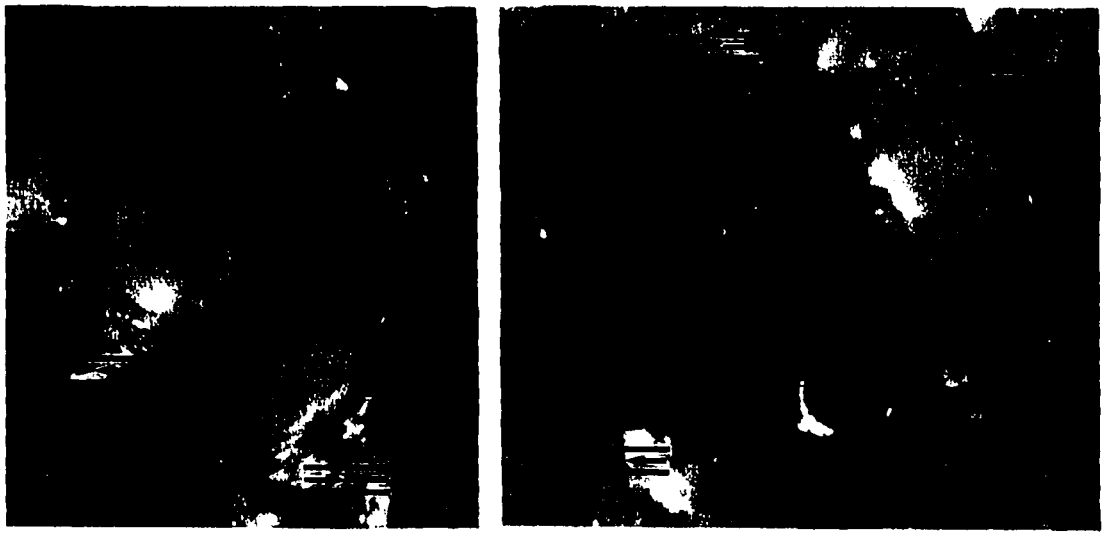

proximal
distal
V21
V21-R(L)
V21-E(LL)
V21-E(LR)
V21-EE(multi)
V21-S(R)
V21-DE(RR)

Figure II-4. Orientation of *P[bw+]* affects PEV. (**A**) Dark-suppressed lines generated from an (LR) duplication have transposons that are flipped relative to that of *V21(L)*. The dark line *V21E3S1(R)* can give rise to enhanced duplications identical to those produced by *V21(L)*. Clockwise from upper left: *V21-E3(LR); V21-E3S1(R); V21-E3S1E1(LR)*. (**B**) Tandem duplication of *P[bw+]* leads to enhancement of variegation (right side versus left side) and the variegation for the two orientations is strikingly different (top versus bottom). Top: *V21(L), V21-E2(LL)*. Bottom: *V21-E3S1(R), V21-E3S1DE1(RR)*. (**C**) *Bst*EII digests with a probe from the 3' end of the *brown* gene (shaded fragment in Figure 1A) showing the characteristic 3' flanking fragment for single and double transposons in two orientations: L=left, R=right. Tandem (T) duplications are also shown.



C

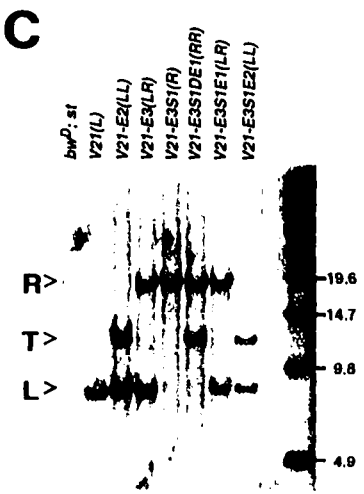

bwD: st
V21(L)
V21-E2(LL)
V21-E3(LR)
V21-E3S1(R)
V21-E3S1DE1(RR)
V21-E3S1E1(LR)
V21-E3S1E2(LL)
R>
T>
L>
19.6
14.7
9.8
4.9

Figure II-5. Heterochromatic proximity affects PEV. (**A**) X-ray induced healing of *V21-E2E1(multi)* rejoins the tip and the base of the third chromosome, trapping 2R heterochromatin (shown in hemizygote, arrow) between 92B and 92C. The absence of *Eco*RV and *Nhe*I restriction sites associated with this heterochromatic block extends for at least 250 kb. (**B**) Clockwise from left, phenotypes of *V21-E2E1(multi)*; non-variegating X-ray revertant, *V21-E2E1X98*; weakly variegating transposase-induced derivative. Based on Southern hybridization analysis (with standard and pulsed-field electrophoretic gels), the weakly variegating line contains a deletion that starts within the array and extends into 92BC euchromatin, ending near or within heterochromatin (data not shown). Zig-zags indicate 2R heterochromatin, dashes 92BC euchromatin and arrowheads transposons. Open curved arrow with X represents X-ray mutagenesis.

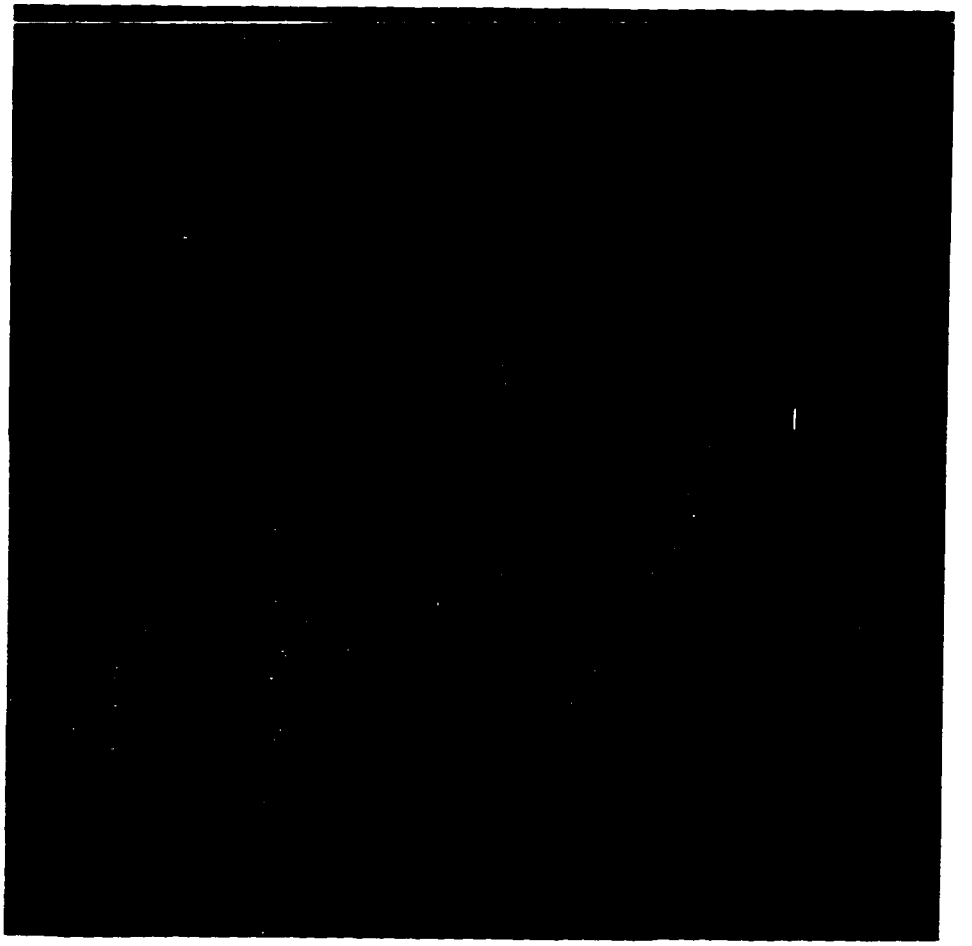
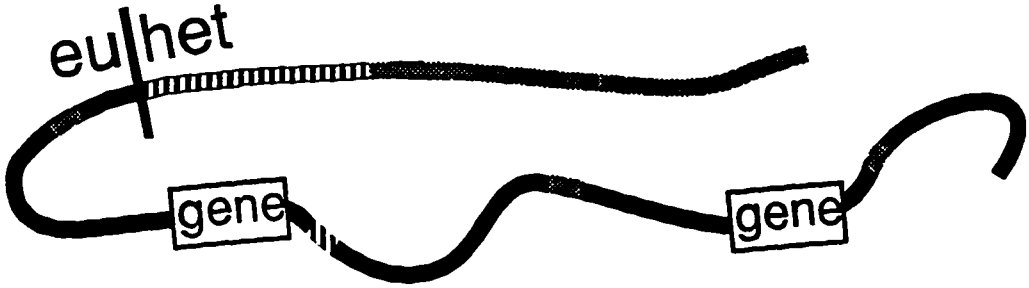
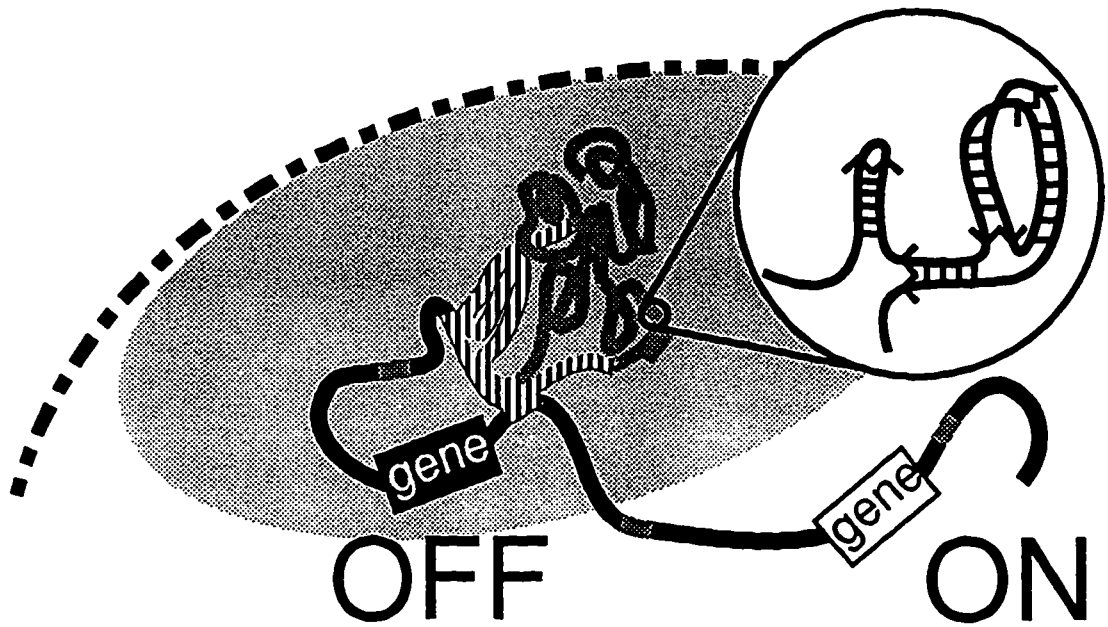


FIGURE II-6. A model for PEV. The repeat organization of a chromosome is shown in the vicinity of a classical variegating rearrangement breakpoint. The solid line represents single-copy DNA in euchromatin (eu) and the striped and shaded lines represent two different families of repetitive sequences, present both in large blocks in heterochromatin (het) and as dispersed middle repetitive elements in euchromatin. Heterochromatin forms as a result of local pairing between homologous double stranded DNA sequences forming hairpin, loop and more complex structures (magnified), which participate in the formation of a chromocenter (grey oval) on the nuclear envelope (dotted line). Silencing of a gene by PEV can occur when a dispersed repeat nearby pairs with a homologous sequence in a block, sequestering the gene into a heterochromatic environment.

Repeat organization



Heterochromatin formation



CHAPTER III

Linkage Modification of PEV in Two Related *P[bw+]* Lines

INTRODUCTION

As discussed in chapter I, the bw^D system has allowed us to see that changes in linkage can modify euchromatic PEV, a situation that is reminiscent of the heterochromatin-distance effect seen for heterochromatic PEV. However, a correlation between linear distance and variegation strength is not observed for linkage modifiers of bw^D (Henikoff, Jackson and Talbert 1995). Indeed, no correlation is seen even when consideration is restricted to rearrangements of similar complexity. Perhaps the "heterochromatin-distance effect" (in the strictest sense of the term) does not exist for euchromatic genes, or of secondary importance relative to other factors, such as changes in the relative proximity of sequences that mediate or modify the spread of variegation. Alternatively, the heterochromatin-distance effect might consistently modify *cis*-acting PEV, yet have no consistent effect on *trans*-inactivation or *para*-inactivation, thus escaping detection in the above screens. After all, in cases of *trans*-inactivation and *para*-inactivation the phenotype relies on local pairing of bw^D and bw^+ alleles and might therefore be particularly sensitive to variability in rearrangement-induced disruption of local pairing.

I have produced linkage modifiers of two closely related lines containing an isolated block of heterochromatin near a reporter transgene array at a distal location on chromosome arm 3R (92B/C). These lines are potentially superior to bw^D because they allow us to examine both *cis*- and *trans*-inactivation, and compare the effects of linkage changes in each circumstance. Furthermore, the two parent lines have different baseline phenotypes due to local differences in the proximity (and perhaps size) of the repeat array and the heterochromatic block. Thus, we can ask how local changes in a gene's environment influence the its susceptibility to linkage enhancement. Finally (Appendix C), the alleles generated in these screens may be useful for further cytological studies on the "looping" interactions and *trans*-inactivation.

MATERIALS AND METHODS

Fly strains and culture conditions: Unless otherwise specified, standard conditions (cornmeal-molasses medium, temperature of 22-25 degrees) and a standard genetic background (second chromosomes carry $brown^D$, third chromosomes carry *scarlet*) were used. Two lines derived from the *V21-E2E1* strain were used. The first line, *16X98*, was produced by the imperfect "healing" of a variegating T(2;3) translocation (see chapter III). That is, re-translocation reverted the right arm of the third chromosome to wild-type morphology (as seen in mitotic metaphase preparations), except for the trapping and isolation of a block of heterochromatin (derived from the base of the right arm of chromosome two) at 92B (Figure II-5). The *P[bw+]* transgene array from *V21-E2E1* remains at a location about 60 Kb distal to the heterochromatic block. Even though the

transgene array is exquisitely sensitive to PEV when in the *V21* configuration, it is entirely wild-type (*bw*$^{+}$) in the *16X98* line.

The second line, *16X98var1* (also described in Chapter II; sometimes called *98v1*), is derived from *16X98* by a transposase-induced event that removes or rearranges the proximal part of the transgene array fuses the array to the heterochromatic block, and deletes intervening sequences to (or into) the heterochromatin block (by CHEF and Southern analysis, Figure III-1a). The deletion induces weak variegation (20-30 spots on a red background). The phenotypic change is probably due primarily to the increased proximity of heterochromatin, and to changes in the heterochromatic block; similar rearrangements of the transgene array have minimal effects on phenotype when the chromosome is in the *V21* configuration (not shown). Due to the re-translocation event, which reverts local crossover-suppression caused by the original translocation, *16X98* and is no longer as strongly linked to *Sb* (which is otherwise present on the *V21*-derivative series). The *Sb* marker has been maintained by selection or balancing in one stock of *16X98*, and lost in others; both types were used in the following screens. When present, *Sb* theoretically allows distinction of extremely enhanced variegating lines from their *bw*^{D}; *st* (white-eyed) siblings.

Mutagenesis: One- to four-day-old males from *16X98* were irradiated (3,000 rads) and mated in groups of 4-6 to 25-35 *bw*^{D}; *st* virgin females (instant medium, Carolina Biological Supply). To aid calculation of independent events, males were removed and mated to additional *bw*^{D}; *st* virgin females after 4-5 days, and separate tallies were kept. The *16X98* flies were marked with *Sb* in 6 of 11 bottles, and were unmarked (except for

transgene pigment) in the remaining 5 bottles. The presence of *Sb* made no difference in the number or phenotype of mutants recovered. Males and females were scored. Variegating progeny were crossed to bw^D; *st* flies and kept as selection stocks. For the *16X98var1* screen, *16X98var1/Payne* males were collected, irradiated, crossed in pairs to 12 bw^D; *st* virgin females (instant food, vials), transferred, scored and recovered in the same way, except as noted.

Use of recombination to assess distance and pairing: Recombination rates provide a rough estimate of the relative distance of enhancing breaks from the 92BC region. Proximity to heterochromatin and proximity to inversion and translocation breakpoints all suppress recombination (see, for example, Offerman and Muller 1932; Dobzhansky and Sturtevant 1931; Glass 1933). The translocation effect is particularly high near the middle of chromosome arms, peaking, for 3R, in the 93-94 region (Roberts 1972). If significant levels of recombination are seen in a linkage enhanced line of *16X98* or *98v1*, it implies that the enhancing breakpoint is a considerable distance away from the 92BC region. Recombination can be recognized by the recovery of the (darker) parental phenotype in stocks of the linkage-enhanced progeny. Crosses were not set up specifically to test for recombination frequencies. Instead, all non-male-sterile lines were propagated by crossing males to bw^D; *st* virgin females, and then kept as selection stocks. A rough estimate of recombination frequencies was calculated by counting dark F2 progeny (~80 flies/line). In lines with no F2 revertants, the count was continued in succeeding generations, up to ~200-300 flies. No statistical analysis was done because common events and uncommon events were not counted in an equivalent way.

Meiotic recombination rates also provide a simple assay for the pairing of homologous chromosomes; this is important because pairing of homologous chromosomes in the interval between a heterochromatic block and an enhancing breakpoint may influence *trans*-inactivation phenotypes.

RESULTS

Linkage enhancers

Linkage enhancement can cause variegation of the 16X98 line. Irradiation of *16X98* produced 8 flies with clearly variegating eyes from a test class of 1,174 pigmented progeny (0.7%). Two males were sterile; one male had a Y-linked translocation(line *98X6.2*); and two females and three males contained stable, autosomal rearrangements (lines *98X2, 98X3, 98X5, 98X6.3* and *98X8* respectively). One additional sterile variegating male and two fertile variegators (one male, one female) were recovered from the transfer bottles. As described in chapter II, the *16X98* line is insensitive to enhancers of variegation. The variegation of the *16X98v1* deletion derivative implies that the 92B heterochromatic block in *16X98* retains some vestige of its heterochromatic character (and thus, perhaps, its ability to communicate with centric heterochromatin). Nevertheless, in the context of the *16X98* chromosome, the block is unable to pass this effect on to the transgene array at 92BC.

Following translocation or inversion, however, variegation begins anew. Looked at in terms of a looping-pairing model, perhaps the 92B block is

contacting the centric heterochromatin more efficiently, so that the inefficiency of pairing between it and the array is mitigated (Figure III-2a). Alternatively, when more proximally located, the block may be directly "revitalized"-- returned to its heterochromatic state, perhaps by recruitment of heterochromatin components -- and thus pair more actively with the repeated sequences of the array (Figure III-2b).

Linkage enhancers of *16X98var1* are recovered at almost four times the frequency of *16X98* linkage enhancers. Irradiation of *16X98var1* produced 27 clearly variegating flies from a test class of ~1010 pigmented progeny (2.7%). Three males and two females were sterile; one female was semi-fertile and produced sterile males; one male had a Y-linked translocation (*98v1XNm1*); and one of the two most extreme variegators either did not breed true or were too extreme to detect in the next generation, and was lost. The remaining 17 *98v1X* lines will be described below or in Table III-I. Additionally, one sterile female and four fertile flies were recovered from the transfer vials. One line did not breed true or was too extreme for detection; the remaining 3 were kept (*98v1XCDe2; 98v1XEFe1; 98v1XIJm1*).

Irradiation and recovery protocols were essentially the same for *16X98* derivatives and *16X98var1* derivatives. The only known difference between the parent lines is a local deletion running from the heterochromatic block at 92B/C into the transgene array at 92C, in what will be called the "het-*bw*$^+$" region (Figure III-1b). Unexpectedly, these local changes in the het-*bw*$^+$ region make *16X98var1* flies sensitive to heterochromatin at almost four times as many sites as *16X98* flies.

Extreme linkage enhancers are often produced from *16X98var1* but never from *16X98*. Mutagenesis of *16X98* produces enhanced lines with a characteristic set of phenotypes; mutagenesis of *16X98var1* produces lines with a very different set of phenotypes (Figures III-3, III-4). 18 of 21 *98v1x* lines are moderate-to-extreme variegators. Two additional lines were lost before characterization because they were so extreme that they could not be distinguished from their bw^D; *st* siblings. In contrast, linkage enhancement of *16X98* produces weak-to-moderate PEV only (Table III-I). The bias in recovery of mutants is not caused by the screening procedure: pigment level in the non-variegating and variegating lines *16X98* and *16X98var1* are similar, so the ability to score weakly, moderately or strongly enhanced derivatives should also be similar for the two lines. Both groups produce a few moderately variegating lines. Comparison of the moderate variegators suggests that the quality of variegation also differs between the lines: *16X98* derivatives are strongly clonal (or patchy), while *16X98var1* derivatives are salt-and-pepper alleles (i.e. spotted). Therefore, it seems that the difference in variegation level and variegation character is controlled primarily at the level of the het-bw^+ region.

Reversion by recombination: Why do we get more linkage enhancers from *98v1* than from *16X98*? This change could reflect an increase in the number of productive breaks within a set euchromatic region; and increase in the number of productive breaks within a set euchromatic region; or an increase in the total size of the productive region. The latter case can be distinguished because it predicts the many of the "excess" lines derived from *98v1* will owe their enhancement to euchromatic breaks that are far from the 92BC region. Greater distance, in turn, should produce higher recombination rates in the interval, which can be recognized as

reversion to the parental phenotype. Among the *16X98var1* derivatives, reversion was seen for 7 of the 14 lines analyzed. In comparison, reversion has not been noticed in *16X98* derivatives. This result is consistent with the bias in recovery rates for the two lines, and is supported by preliminary cytological analysis. By all criteria, the het-bw^+ region from *16X98var1* really is more sensitive to PEV than the het-bw^+ region from *16X98*, allowing more distal cytological breakpoints to cause PEV in *98v1X* lines.

Cis- and *trans-*inactivation

Equivalent levels of *cis*-inactivation of the same repeat array differentially *trans*-inactivate a non-variegating transgene. The *98v1* linkage enhancers produce a fairly full range of variegating phenotypes, with a fairly large cluster of comparable phenotypes at the extreme end of the spectrum (Table III-I). Seven lines were tested for *trans*-iactivation phenotypes. Strikingly, little or no correlation can be seen between the *cis*-inactivation and *trans*-inactivation phenotypes (Figure III-3, Table III-I), even though most of the lines tested showed a reasonably good (and in some cases, similar) degree of pairing by the meiotic-reversion-rate test.

Pairing effects do not account for the discordance between *cis*- and *trans*-inactivation phenotypes: Several of the *98v1* derivatives were rated for both *trans* inactivation of *P[bw^+] 92C* and for their ability to pair well in the region between their enhancing breakpoint and the 92BC region (as assayed by recombination). Interestingly, pairing strength alone cannot account for the relative differences between the *cis*- and the *trans*-

inactivation phenotypes (Table III-I, Figure III-3). For example, of two alleles with 2-4% rates of cross-over reversion, the *trans*-inactivation strength of *98v1Ce1* is much greater than that of *98V1Je1*, though the *cis*-inactivation phenotypes are similar (Figure III-3 b,d). Nevertheless, efficient *trans*-inactivation may require some local pairing: *98v1Ee3*, an allele with no detectable crossover reversion, is a particularly poor *trans*-inactivator (Figure III-3 f).

The difference between *cis* and *trans* inactivation is inexplicable at this point. In theory, the additional hetero-chromatin could have differential effects on *cis*- and *trans*- inactivation by interacting directly with one allele or the other; for example, enhancement of *cis*-inactivation could be entirely mediated through the closely apposed 92B block of heterochromatin, while enhancement of *trans*-inactivation could include direct effects of the additional heterochromatin on the *trans*- copy of the gene. However, the maintenance of distinctive phenotypes from *cis*- to *trans*-inactivation weakens this argument: sectored alleles remain sectored, while salt-and-pepper alleles remain speckled (compare Figures III-3, III-4).

Imprinting-like effects can be seen for *trans*-inactivation: While discussing trans-inactivation phenotypes, a peculiar effect should be noted: one linkage enhanced line, *16X98X3*, shows a strong parental effect, specific for *trans*-inactivation(Figure III-4c). No such effect was seen in other linkage-enhanced lines tested. The line has not been tested for Y-chromosomal effects or maternal effects.

DISCUSSION

Recovery of linkage enhancers---*cis*-inactivation vs. *trans*- (or *para*-) inactivation: Linkage enhancement screens were done on *16X98* and *98v1* to see whether scoring phenotypes by *cis*-inactivation alone influence the correlation between the phenotype and the cytology of recovered lines. In addition, we hoped to ask whether screening by *cis*-inactivation might change the number of lines recovered, or their breakpoint distribution. To date, none of the lines scored by *cis*-inactivation would have been missed if scored by *trans*-inactivation. In addition, we have no reason to think that we have changed the frequency or the type of events recovered. For example, linkage-enhancers of *16X98var1* and linkage enhancers of the (slightly variegating) *Byron* derivative of bw^D are recovered at similar frequencies: 2.7% vs. 2% (Henikoff, Jackson and Talbert 1995). Furthermore, recombination rates suggest that the enhanced lines share a similar range and degree of linkage to their enhancing breakpoints (data not shown).

Comparing *98v1* derivatives to *V21-E2E1*: *V21-E2E1(multi)* is phenotypically similar to several of the *16X98var1* X-ray derivatives (Figure II-2, III-3). Furthermore, it is structurally similar in the het-bw^+ region, and is therefore predicted to have similar local pairing properties. Any strong differences in their *trans*-inactivation ability, then, should be due either to the small changes that distinguish the parent alleles from each other, or to the specific rearrangement that has produced a given *16X98var1* X-ray derivative (Figure III-5). Three of four extreme *98V1* derivatives tested *trans*-inactivate far more strongly than *V21-E2E1*. First, this result supports our finding that the degree of *cis*-inactivation is not a

straightforward predictor of the degree of *trans*-inactivation. Secondly, this result allows us to speculate on the relative importance of local vs. chromosomal pairing in mediating *trans* inactivation (an idea that will be examined further in the next chapter). If allelic pairing of sequences between the transposon and the heterochromatic block (deleted in the *98v1* line) were relatively important, we would expect the *98v1* lines to be weaker, *trans*-inactivators than *V21-E2E1*. Instead, the *98v1* lines are stronger *trans*-inactivators. Therefore, the loss of the het-*bw*$^+$ region apparently plays only a small roll in local pairing. How can we explain the phenotype of the *98v1* lines relative to that of *V21-E2E1*? Enhancement may be due to some combination of three factors: first, there may be increased pairing of the homologues outside of the het-*bw*$^+$ region (minimal in *V21-E2E1*, in which the 92B region is juxtaposed to the chromocenter). Next, the *trans*-inactivated allele might be responding directly to the increased proximity of heterochromatin. Finally, the drop in stearic hinderance caused by moving the 92BC region away from the massive chromocenter may allow better contact between alleles.

"Distance effects": what matters besides chromosomal distance? Linkage effects can modify *cis*-acting euchromatic PEV caused by isolated heterochromatic blocks. The relative ability of an isolated heterochromatic block to function as heterochromatin therefore extends to its effect on adjacent euchromatic genes. However, preliminary results on the cytology of *16X98* and *98v1* linkage-enhanced derivatives give no indication that the strength of variegation correlates with the distance between heterochromatic blocks. First, the cytology of these lines is often complex (data not shown). For *16X98*, one simple two-break rearrangement was recovered; the other derivatives all appear to involve at least three (and

possibly as many as five) breaks, with one or more in heterochromatin. Break complexity may influence heterochromatic PEV (B. Wakimoto, personal communication); these results suggest that the complexity effect may also be felt, indirectly, by euchromatic genes subjected to PEV.

Interestingly, one event, producing the moderately-enhanced line *98v1XAm1*, does not interrupt the third chromosome anywhere between the base of the arm and 92BC; that is, there is no change in the intra-chromosomal distance between heterochromatin blocks. However, in this case, it is clear that chromosome pairing could cause re-localization of the isolated block towards the chromocenter (Figure III-6a). Perhaps similar effects occur in the other complex rearrangements, masking or modifying the strictly intrachromosomal effect of linkage distance. Given the complexity of the rearrangements and the (relatively) small sample size, further cytological analysis seemed unwarranted.

As to our main question, we have shown that the *trans*-inactivation phenotype of an allele is not necessarily a good predictor of its *cis*-inactivation phenotype. However, we still see no suggestion of a strict heterochromatin distance effect even when using *cis*-inactivation to assay changes. That is, the intrachromosomal distance between heterochromatic blocks does not correlate with the strength of variegation either in *cis* or in *trans*. Presumably, previous "failures" in observing this effect were not, therefore, due to such problems as the disruption of local pairing, as suggested, for example, by Henikoff, Jackson and Talbert (1995).

The strict heterochromatin-distance effect, if it really exists, appears to be highly subject to modification. The phenotype of one line, *98v1XAm1*,

suggests that in the absence of any intrachromosomal change in linkage distance, variegating phenotypes can still be modified by changes in linkage which could draw the variegating region towards heterochromatin through the mechanism of chromosomal pairing -- an interchromosomal "heterochromatin distance effect." These effects may make up a considerable sub-set of the (previously mentioned) "complexity effects." Presumably, these effects have influenced the recovery and classification of previous linkage-modifiers. In addition, it is possible that euchromatic genes, embedded in euchromatin, respond to their chromosomal or nuclear position in ways that are more complex, sequence-specific and idiosyncratic than those of heterochromatic genes embedded in blocks of heterochromatin. Thus, even if we had a multitude of linkage enhanced lines, analyzed only simple rearrangements, and scored strictly by the *cis*-enhancement phenotype, we might still not see a continuous correlation between heterochromatin distances and phenotypes.

Summary: In this chapter, I have tried to emphasize the idea that inactivation of genes in *cis* and in *trans* to variegation-inducing breakpoints both involve some level of somatic pairing. In energetic terms, sequences lying in *cis* to each other are already tethered together; keeping these sequences in relatively close proximity to each other does not strongly suppress the overall entropy of the chromosomal association. Under these conditions, local pairing of short homologous regions is a relatively favorable process. By comparison, *trans*-inactivation requires that associations be maintained between two separate chromosomes. In fact, if we accept a looping model for PEV, this association is strongly increasing entropy by anchoring a normal chromosome to the chromocenter. Under these conditions, the overall pairing of chromosomes (and

the relative position of chromosomal breaks) is more important than the local alignment of short sequences. Further analysis using modern cytological methods may be able to demonstrate this point directly (Appendix C).

TABLE III-I
Linkage enhancers

allele name	*cis*[1]	*trans*[1] L(92C)	*trans*[1] R(any)	*trans* imprint?	cross-over	comments	cytology (polytene)	southern
98v1XHe1	8	nd	nd	nd	NO	from f., see only in m.	T(2;3) prox. to 91F; 41?	nd
98v1XIe2	8	nd	nd	nd	NO	-----	nd	nd
98v1XCDe2	7	nd	nd	nd	2-4%	-----	nd	nd
98v1XIe1	7	7	1	NO	nd	-----	In(3R)83D,94D and T(2;3) 32?;94D	nd
98v1XEe1	7	nd	nd	nd	NO	-----	nd	nd
98v1XCe1	6	8	nd	NO	2-4%	-----	T(2;3) 41E;96EF	nd
98v1XDe1	6	nd	nd	nd	NO	-----	92A-??	Y
98v1XLe2	6	nd	nd	nd	~1%	-----	nd	nd
98v1XEFe1	6	3	<1	NO	nd	darker with age	have	nd
98v1XAe1	5	nd	nd	nd	nd	m(s); a/p gradient	nd	nd
98v1XJe1	5	3	nd	nd	2-4%	-----	nd	nd
98v1XLe1	5	nd	nd	nd	<1%	a/p gradient	nd	nd
98v1XIe3	4/6	nd	nd	nd	nd (1%?)	strong a/p gradient	nd	nd
98v1XDe2	4	2	nd	nd	1-3%	-----	nd	Y

cont.

TABLE III-I
Linkage enhancers (con't.)

98v1XAm1	4	nd	nd	nd	NO	a/p gradient	In(2R)59AB,53A and T(2;3)96C;59E	Y	cont..
98v1XIJm1	4	2	<1	nd	5-7%?	-----	nd	Y	
98v1XJm1	4	nd	nd	nd	NO?	-----	nd	nd	
98v1XEe3	4	?	nd	nd	NO	-----	nd	nd	
98v1XJm2	3	nd	nd	nd	nd(NO?)	-----	nd	nd	
98v1XEm4	3	nd	nd	nd	nd	-----	nd	nd	
98v1XNm1	2	nd	nd	nd	na	Y-linked	nd	nd	
98X3	3	1/3	nd	strong	NO[2]	patchy	T(2;3) (mitotic)	nd	
98XC	3	2	nd	NO	NO[2]	patchy	nd	nd	
98X5	3	nd	nd	nd	NO[2]	-----	T(2:3) 90D-2L base? plus "other"	nd	
98X8	2	nd	nd	nd	NO[2]	-----	T(2;3) 91A;41?	nd	
98XA	2	nd	nd	nd	NO[2]	-----	nd	nd	
98X6.3	1	nd	nd	nd	NO[2]	-----	nd; disruption near 92A?	nd	
98X6.2	1	nd	nd	nd	NO[2]	-----	complex; 30A, 67E, 81Fhet, ~90?	nd	
98X2	1	nd	nd	nd	NO[2]	-----	nd	nd	

nd= not determined; NO=not observed; na=not applicable

1 scale based on visual comparison to *V21* alleles; *V21-EE*=8; *V21-E3*=7; *V21-E2*=5; *V21*=3/4;*V21-S*=2; *98v1*=1.

2 difficult to detect revertants in dark lines. No stable revertant lines recovered.

Figure III-1. Comparison of *16X98* and *98v1* lines. a) Using a probe for *brown* (described in chapter II) to probe genomic DNA cut with enzymes whose sites flank the transgene insertion site at 92C, we can see that *16X98v1* line is derived from the *16x98* by an event which fuses the transgene array to a block of DNA lacking restriction sites (arrows). For comparison, the large bw^D band lies at the limiting mobility of the CHEF gel under the conditions used. b) *16X98* and *16X98v1* are related alleles containing a heterochromatic block derived from chromosome 2R (zig-zag) at 92B separated by ~60 kb from a *P[bw+]* transgene array (triangles) inserted at 92C. *98v1* differs from *16x98* by a local deletion which removes and/or rearranges a defective (*brown-*) *P[bw+]* transposon at the proximal end of the transgene array, removes all intervening sequences, and may remove some of the heterochromatin in the block. The effect of the change in the transgene array itself is assumed to be small, based on transposase-derived alleles of *V21-E2E1* which are still quite extreme when they contain three copies of the *P[bw+]* transposon in a variety of relative orientations (data not shown). In addition, reduction of transgene copy number tends to suppress PEV; therefore, a reduction in the transgene array of *16X98var1* relative to *16X98* is presumably not responsible for enhancement of PEV. c) The *16X98* line and some (average) derivatives are shown clockwise from top left:*16X98* , *16X98X3*, *16X98XA*, *16X98X8*. d) The *16X98v1* line and some (average) derivatives are shown clockwise from top left: and (average) derivatives *98v1XIe1*, *98v1XDe2* and *98v1XCE1*.

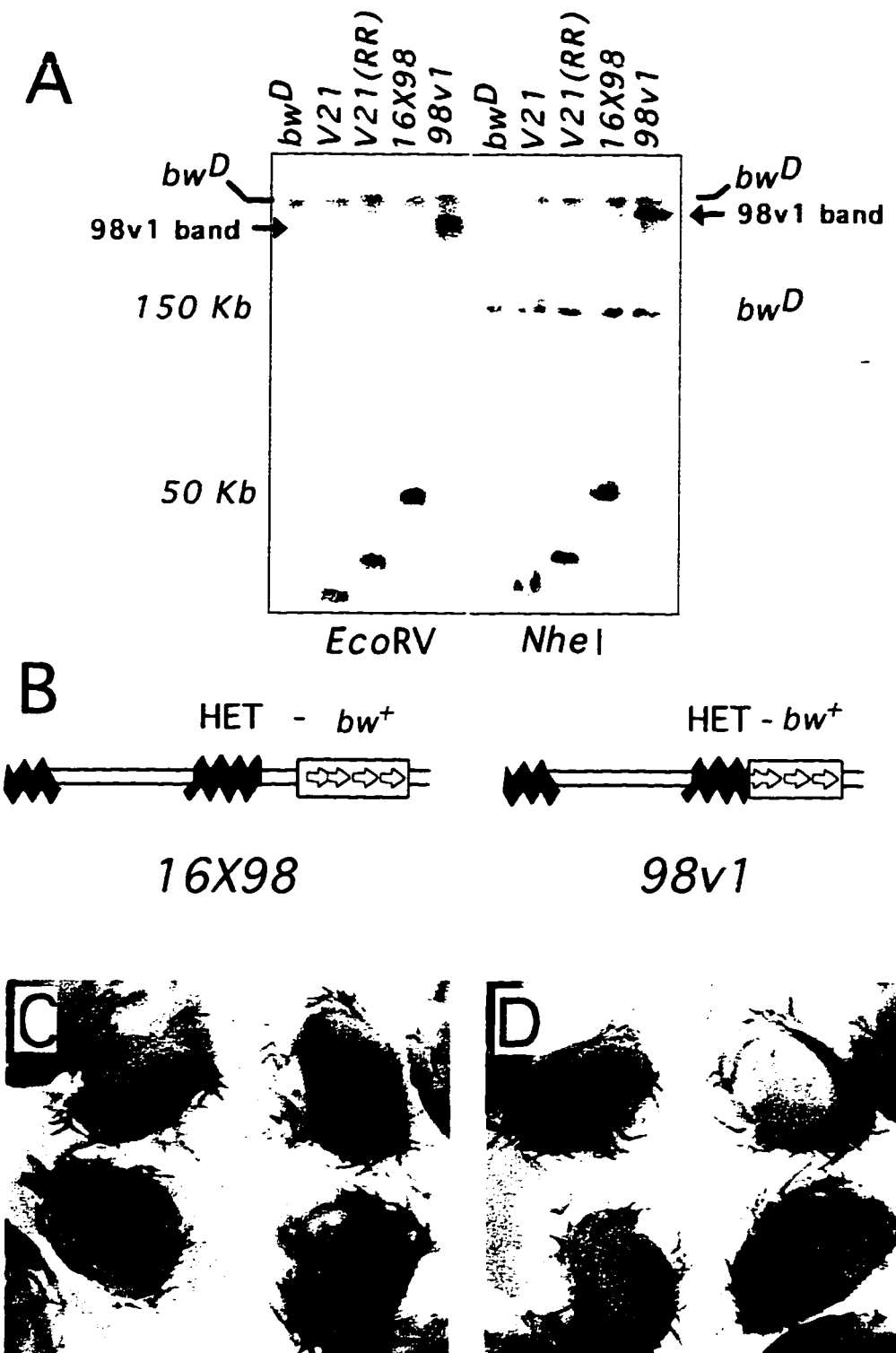

A
bwD
V21
V21(RR)
16X98
98v1
bwD
V21
V21(RR)
16X98
98v1
bwD
98v1 band
150 Kb
50 Kb
bwD
98v1 band
EcoRV
NheI
B
HET - bw+
HET - bw+
16X98
98v1
C
D

Figure III-2. A model for linkage enhancement of *16X98*. Variegation is never observed for *16X98* in the absence of linkage enhancement. By a looping argument, this means that the frequency of contact between the heterochromatic block at 92B and the chromocenter (dark zig-zags) is low (left) and/or the frequency of contact between the reporter array (arrows) and the heterochromatic block (zig-zag) is low (right). Presumably, linkage enhancement improves the chances of contact between the heterochromatic block at 92B and other heterochromatin; it is also possible that proximity to heterochromatin changes the pairing properties of the 92B block, enhancing it's "stickiness" and stabilizing contacts between the block and the array, changing both terms of the equation.

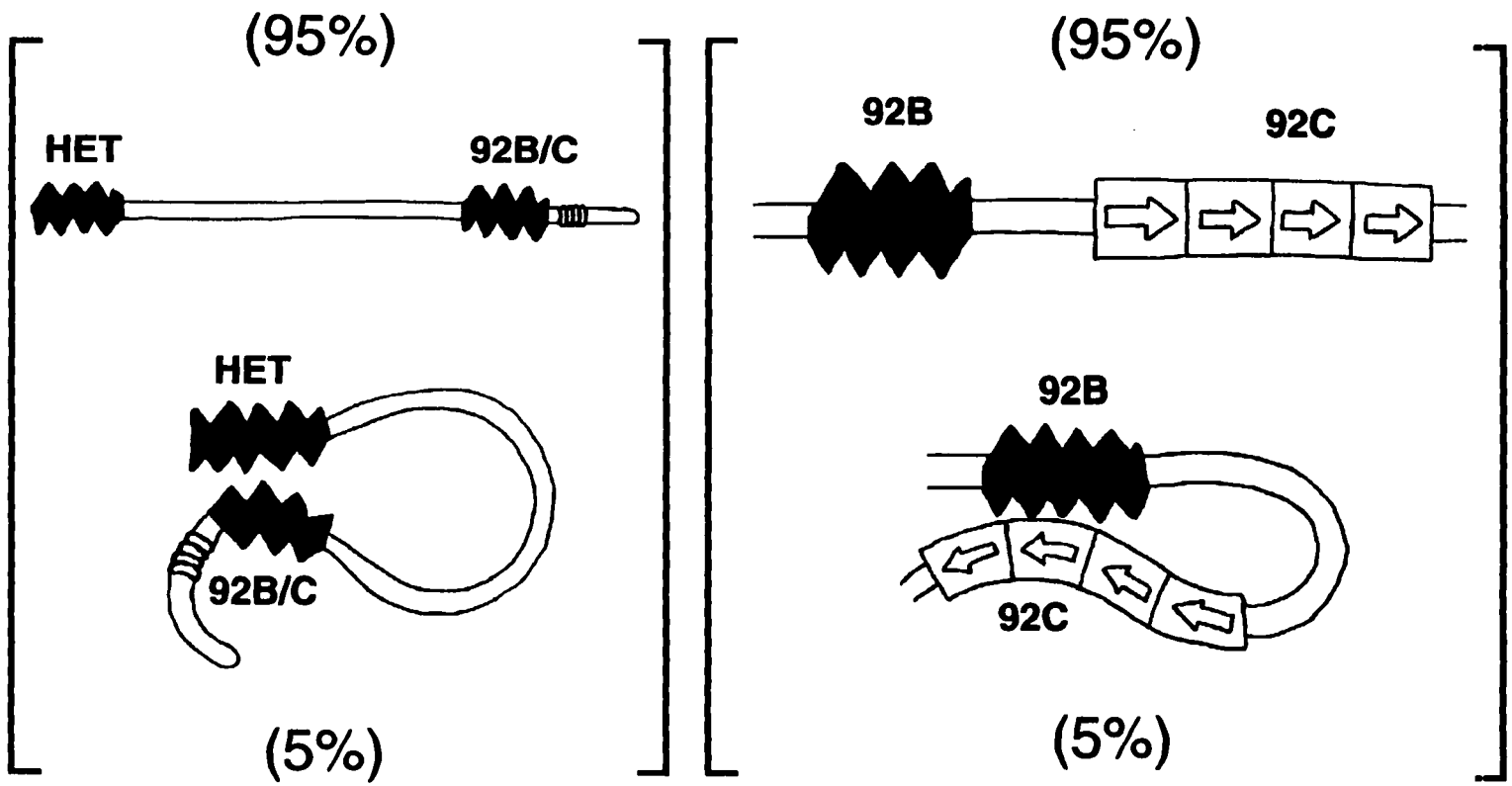

(95%)
HET
92B/C
HET
92B/C
(5%)
(95%)
92B
92C
92B
92C
(5%)

Figure III-3. *98v1 trans-* inactivation phenotypes. In each panel, males are on top, females on the bottom; flies on the left of the panel are *98v1Xderivative/-*, while flies on the right are *98v1Xderivative / P[bw+]92C* (unless otherwise noted). The lines are arranged in the order of their *cis-* PEV phenotypes (upper left to lower right). (The conditions used for the *trans* inactivation experiments produce a slightly different order from that in Table III-I). The *trans*-inactivation phenotypes clearly do not correlate well with this order. a) *98v1XEFe1*, b) *98v1XJe1*, c) *98v1XIe1*, d) *98v1XCe1*, e) *98v1XIJm1*, f) *98v1XEe3* (females only, males were darker), g) *98v1XDe2* females only; *98v1XDe2 / -* left, *92C / -* lower right, *98v1XDe2 / 92C* upper right.

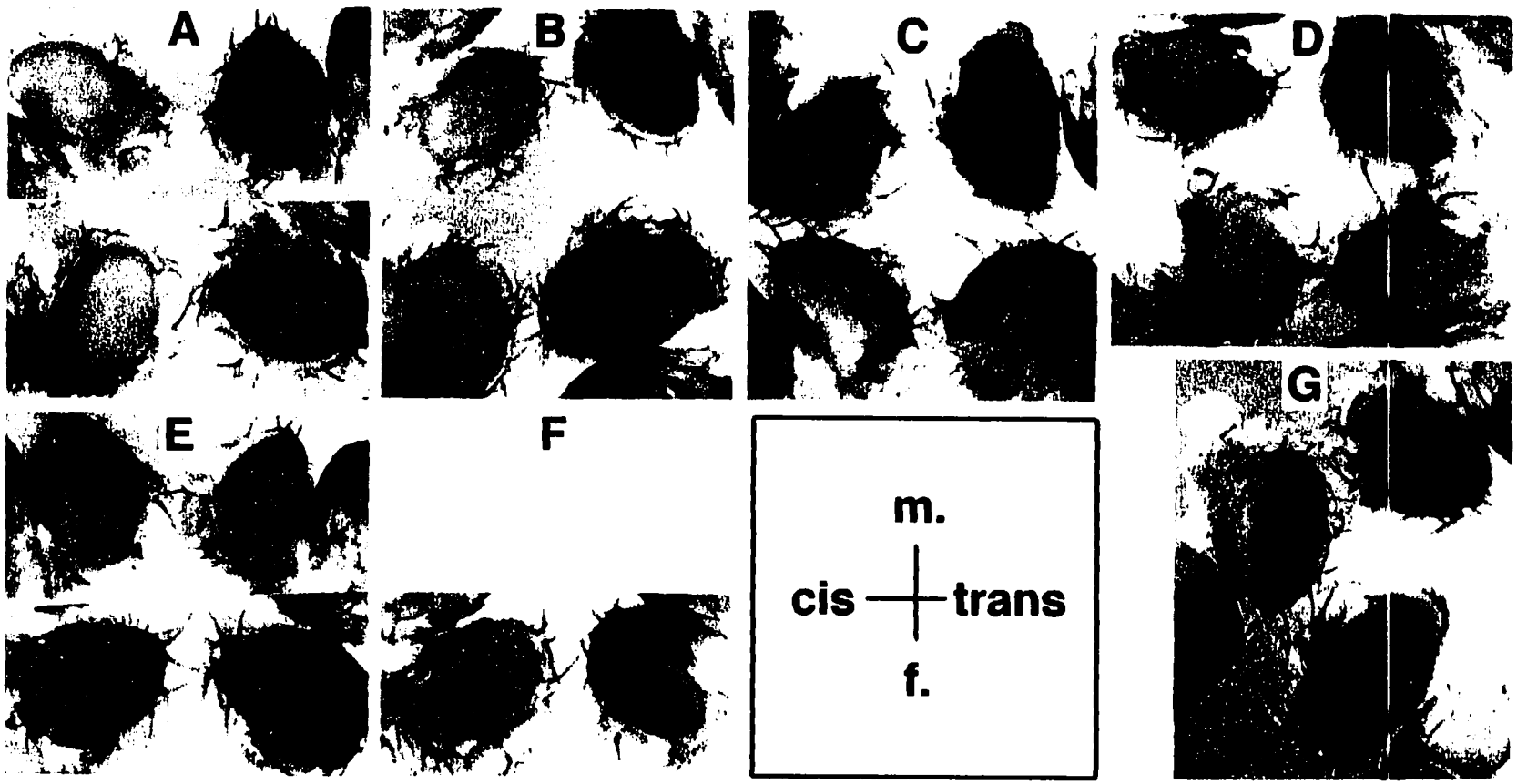

A
B
C
D
E
F
G
m.
cis
trans
f.

Figure III-4. *16X98* lines: *trans-* inactivation. a)*16X98* lines *trans-*inactivate; clockwise from top, *16X98X3 / -- ; -- / 92C;* and *16X98X3 / 92C.* b,c) Parental effects of *trans* inactivation. b) For most *16X98* derivative lines, there is no difference in the *trans*-inactivation phenotype depending on the parent providing the variegating allele; here, *16X98XC / 92C* is shown (maternal, left; paternal, right; males, top; females, bottom). c) One line, *16X98X3,* shows a strong parental effect; the paternally-derived array trans-inactivates more strongly *16X98X3 / 92C* is shown (maternal, left; paternal, right; males, top; females, bottom).


A
B
C

Figure III-5. Modeling the relative contributions of pairing regions. *trans*-inactivation can be looked at in terms of the competition between pairing forces (productive contact) and stearic hindrance (unproductive contact). a) The repeat array (arrows) in *V21-E2E1(multi)* is separated from heterochromatin (zig-zag) by over 50 Kb; the heterochromatin is present as a single, unbroken block. In *V21-E2E1*, the opportunity for both local pairing (in the ~60 kb het-*bw*$^+$ interval) and for steric hindrance (between the rest of the homologous chromosome and the chromocenter) is high. b) The intermediate *98var1* demonstrates the removal of sequences in the 92BC region, and the restoration of homologous pairing in the region proximal to 92B. A block of heterochromatin is isolated at 92BC. c) In the (idealized) *98v1* derivative, local pairing and stearic hindrance are both lower than for *V21-E2E1 (multi)* ; however, the potential for direct contact between the array and the block has increased. Thus, the drop in steric hindrance and/or the increased contact between the array and the heterochromatic block are more important determinants of the *trans*-inactivation phenotype than local pairing interactions in the 60 kb het-*bw*$^+$ region. For ease of comparison, potentially helpful paring is indicated by "yes," unpaired regions caused by insertion or deletion are labeled "no," and regions of potential (but unproductive) pairing are labeled "(no)."

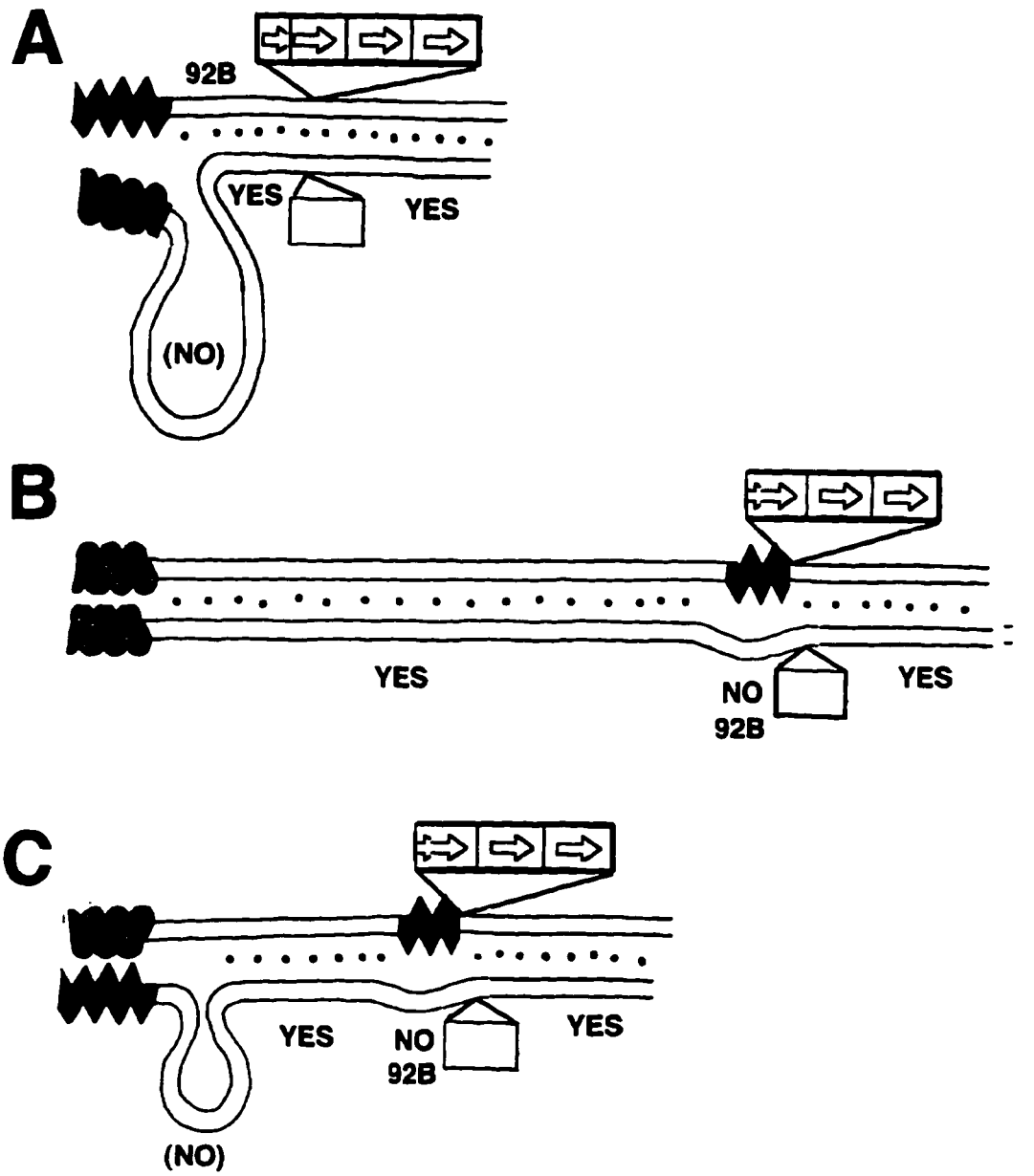
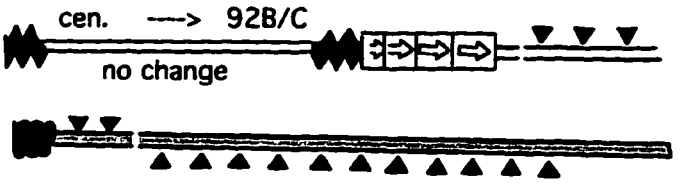
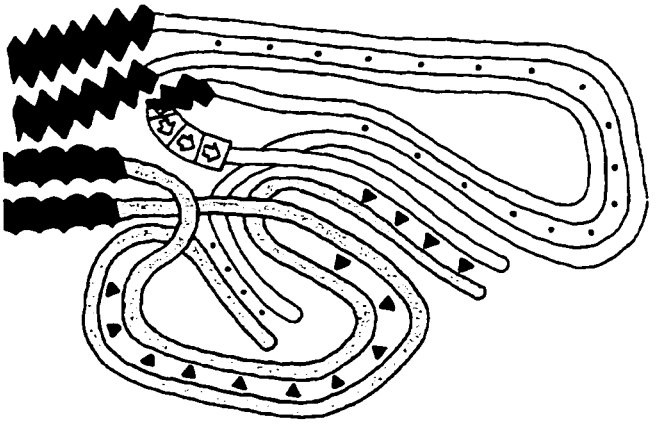

A
92B
YES
YES
(NO)
B
YES
NO
92B
YES
C
YES
NO
92B
YES
(NO)

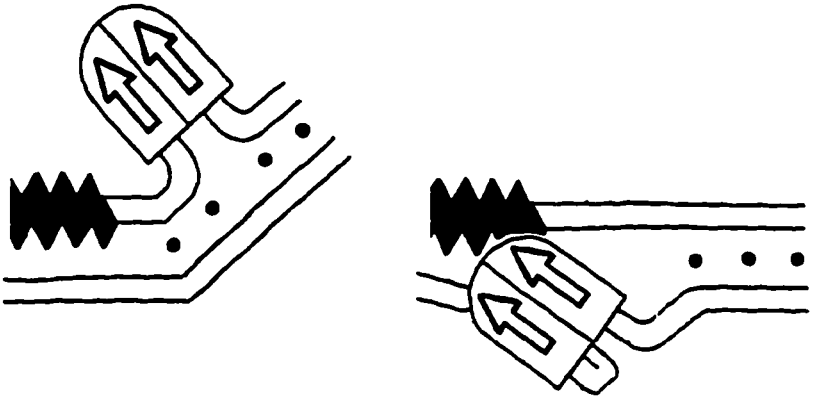
Figure III-6. Can *trans* chromosomes enhance PEV? a) linkage enhancement is sometimes observed in cases where the 92C region is no closer to heterochromatin on the same chromosome. For example, an inversion on another chromosome (triangles) coupled with a distal translocation (dotted/undotted switch, shown on right) can lead to re-localization of the 92C region to the chromocenter (zig-zag and squiggle = heterochromatin) if all homologous regions pair (below). b) If pairing properties of inverted duplications (thumb-like projection) facilitate contact with heterochromatin in *cis*, it is possible that such duplications could actively help to enhance PEV, even when present on the "wrong" chromosome (right), perhaps by binding the heterochromatin-associated protein HP1. The *98v1X* derivatives might provide a sensitive system for detecting such properties at 92C.

A





B



CHAPTER IV

Orientation-Based Disruption of Local Homologous Pairing: Effects on Dominant PEV

INTRODUCTION

Before the discovery of repeat-based enhancement of PEV, and thus before our proposal that transgene arrays could participate in intrachromosomal pairing interactions, the pairing properties of *brown* alleles were already a subject of interest and study. In an attempt to dissect the mechanism of *trans*-inactivation, a transposase-induced deletion series of the *P[bw+]* transposon of *V21* was constructed (Dreesen, Henikoff and Loughney 1991). The *V21* translocation breakpoint induced *trans*-inactivation of the allelic *P[bw+]* element, even when the *P[bw+]* element on the *V21* translocation chromosome had been fully deleted; however, the strength of the *trans*-inactivation was reduced. In a second set of experiments, partial deletions of the *P[bw+]* transposon on the un-rearranged chromosome also suppressed, but did not eradicate, *trans*-inactivation, while retaining *brown* expression (Martin-Morris *et al.* 1993; see Figure I-1c). Together, these results suggested that the strength of *trans*-inactivation might correlate only with the amount of local homology near the affected alleles. On this basis, the authors suggested that the efficiency of *trans*-inactivation could be a direct manifestation of the relative ability of the different alleles to pair with each other.

In theory, these lesions demonstrate the importance of local interactions in promoting chromosomal pairing and co-localization. However, the deletion experiments were not designed to look at pairing requirements. Instead, they were originally intended to find specific sequences mediating *trans*-inactivation. The transposase-induced deletions did not merely remove inert DNA; all of the *V21*-related deletions and several of the *P[bw+]92C* deletions also removed sequences from the *brown+* transgene. Now that we recognize the potential role of intrachromosomal contacts in mediating PEV (the pairing looping model for PEV, chapter II), we have to consider an alternative interpretation, namely, that the *P[bw+]* deletions suppress PEV by removing sequences that mediate intrachromosomal contacts. If so, then it is possible that local lesions do not, in fact, interfere directly with allelic pairing or with *trans*-inactivation *per se.*

Do other examples support the idea that *trans*-inactivation depends on the effectiveness of local, allelic pairing? A series of experiments carried out using the *brown+* gene, the *brown^{Dominant}* (*bw^{D}*) rearrangement, and more complex *brown* derivatives may serve, after the fact, to rank the relative strength of *cis*-pairing and *trans*-pairing of different alleles (see Henikoff and Dreesen 1989; Henikoff, Jackson and Talbert 1995; see also Figure I-7). The *bw^{D}* chromosome consists of a two-megabase, heterochromatic, simple-sequence repeat inserted into the (disrupted) *brown* gene. In derivatives that combine tandem copies of *bw+* and *bw^{D}*, the heterochromatin of *bw^{D}* can inactivate the linked *bw+* gene. Introducing an extra copy of *bw^{D}*, however, can partially suppresses *trans*-inactivation of linked or unlinked copies of *bw+*. This suggests that both sorts of inactivation are mediated by a pairing mechanism, and that pairing of two identical copies of *bw^{D}* precludes pairing of *bw^{D}* with the less-similar *bw+* sequences.

Considerable evidence of this type supports the idea that preferential pairing of identical sequences in *trans* can partially exclude pairing between non-identical partners either in *cis* or in *trans*, though the exclusion is not absolute (Henikoff and Dreesen 1989; Henikoff, Loughney and Dreesen 1993; Henikoff, Jackson and Talbert 1995, also see Figure I-7). Given the complexity and size of the loci involved, however, it is hard to call this effect a purely local phenomenon, especially in terms of adding extra, megabase-sized heterochromatic elements to the system. Furthermore, the additional chromatin in bw^D is distinctive: it is composed of simple-sequence repeats, has a strong strand-bias in base composition, and is under-represented in polytene tissues. If this chromatin interferes with pairing, it could do so in a variety of ways that may differ in different tissues.

To summarize previous experiments:

1. Large disruptions of chromosomal pairing can strongly modify PEV

2. local changes that disrupt pairing alter the susceptibility of nearby genes to dominant PEV,

3. *cis*- inactivation ("*para*-inactivation") and *trans*-inactivation seem to be competitive processes that share some mechanistic aspects,

and, drawing on the information offered in chapter II, we can add that

4. repeat-based enhancement of PEV also involves pairing, but its interaction with the above phenomena is not yet clear.

If we hope to tie together the very local effect of repeat-based enhancement of PEV and the local effects involved in *trans*-sensing, we must have a more precise system than any of the ones detailed above. Ideally, the heterochromatin, the breakpoint, and the euchromatin would all remain constant so the only variable would be simple change in the pairing properties of the reporter gene itself. Altering the orientation of a *P[bw+]* transgene at 92C fulfills the above conditions. Combining alleles of *V21* and of *92C* allows us to ask whether *cis*- or *trans*- pairing forces dominate, or whether all copies (*cis*- and *trans*-) associate indiscriminately.

MATERIALS AND METHODS

Fly strains and culture conditions: All conditions are described in chapter II. Unless noted, all second chromosomes carry bw^D and all third chromosomes carry *scarlet* (*st*), providing a white background for easy scoring of expression from *P[bw+]* transposons.

Mutagenesis: As described in Chapter III, derivatives of *P[bw+]92C* were produced from *P[bw+]92C* by transposase mutagenesis as follows: flies were screened for bright red eyes to retrieve any duplications of the *P[bw+]* transgene and linkage was checked in the F3 and F4 generations. Next, two lines with perfect, divergent duplications, as determined by Southern analysis, were chosen. These lines are *92C*4D* and *92C*2B*. We know from transposase-mutagenesis of an equivalent duplication on the *V21* translocation chromosome that more than 50% of progeny from the dysgenic cross have a full or partial reversion event. This very high reversion frequency includes single transposons in either orientation,

making this type of duplication a highly efficient way to generate transgene alleles that differ only in their chromosomal orientation. Accordingly, two males from homozygous *92C*4D* and *92C*2B* lines were crossed to ten *st Sb P[Δ2,3(99B)] / Payne* virgin females; seven of the resulting *Sb*-marked F1 males were mated individually to ten bw^D; *st* virgin females; and resulting non-*Sb* progeny were scored for reversion to single-transgene pigment levels. Two or three putative revertant males were chosen from each vial and were mated to bw^D; *st* virgins females. The resulting lines were made homozygous by selecting F2 individuals with bright red eyes and were characterized by Southern analysis: *Eco*RV (which does not cut within the transposon) and the single-cutter *Bst*EII were used to winnow prospective lines for insert size and orientation, then orientation was confirmed using *Kpn*I, *Avr*II and *Nru*I, with *V21-E3S1* as a standard (Figure IV-I). Three individual lines with inverted transposons were recovered from *92C*4D*, and called *4D-2B*, *4D-3B* and *4D-5A* (in addition, eight separate lines were recovered with a single transposon in the original *92C* (L) orientation, and two lines had damaged elements). No inverted alleles were recovered from *92C*2B*. Ideally, a new (L) line could have been derived from one of the (R) lines by again selecting for (LR) duplications, and again reducing the duplication to a single (L) copy (two additional rounds of mutagenesis) to show that no lesion had been introduced in *92C*4D*; in the interests of efficiency, this was not done.

Test crosses: Six related *trans*-inactivating alleles were used: *V21* (L); *V21-E3*, a (LR) duplication derivative of *V21*; *V21Δ116* (L), a null derivative of *V21* containing a small internal deletion; *V21-E3S1*, an (R) allele; and two tandem duplication alleles, one in the (LL) and one in the

(RR) orientation. The *V21-E3* inverted duplication should pair equally well with both (L) and (R) derivatives (figure IV-2a), while *V21*, a (L) transposon allele, is expected to pair well with (L) derivatives of *92C*, but not with (R) derivatives, in spite of the fact that they share full sequence homology (Figure IV-2b,c). By a similar argument, the R alleles of *V21* and of *92C* are expected to pair well. In each case, males from the *V21* derivative lines (*st, V21-E3, Sb / st* or *st, V21 / st Sb H*) were crossed to homozygous virgin females from the *92C* derivative lines, and progeny were scored and compared using the presence or absence of *Sb* to distinguish the test class. Heterozygous *92C* derivatives were occasionally used when pilot experiments suggested that the resulting classes would be unambiguous.

RESULTS

Orientation predicts the sensitivity of *P[bw+]* at 92C to *trans*-inactivation. For all combinations tested, (R) alleles were far less *trans*-inactivated than the original (L) allele (Figure IV-3, Table IV-I). This was true when the allele on the inactivating chromosome was single or double; when it lay parallel or antiparallel to the allelic copy; and when it was either an (L) or (R) allele. This surprising degree of autonomy in the *trans*-inactivated allele is so strong as to overwhelm the effects of local allelic pairing requirements (if any). Interestingly, the suppression of variegation by a (R) orientation was also seen for *cis*-inactivation of *V21* derivatives (Chapter II); though, as there are only two possible outcomes for this comparison, this point should not be over-stressed. Nevertheless, the consistency in the orientation preference between *cis*- and *trans*-

inactivated alleles may mean that the same physical process is involved in both the *cis-* and the *trans-* orientation effects. The concordance of orientation-based phenotypes for separately-derived *cis-* and *trans-* alleles of *P[bw+]* is also comforting, in that it suggests that the dark phenotype of variegating (R) alleles is not caused by some unrelated, undetected lesion in the (R) chromosome. If such a lesion exists, it must have arisen twice; in addition, it must have no detectable effect in boosting baseline gene expression in non-variegating genotypes.

DISCUSSION

Preferential maintenance of *cis*-pairing? It has been proposed (Dorer and Henikoff 1994) that intrachromosomal pairing of inverted repeats may create "hairpin" structures that are a) stable and b) recognizable as a chromosomal disruptions. Supporting this idea, duplications are able to localize the heterochromatin-associated protein HP1 (D. Dorer, personal communication). I have suggested that in the case of *V21*-derived alleles, such structures enhance PEV by serving as a "tag" for looping of repeated sequences into the chromocenter (chapter II; Figure IV-4a). If *trans*-pairing were to compete efficiently with *cis*-pairing, we would expect *V21*(LR) / *92C* (L) heterozygotes to be very dark: loss of *cis*-pairing would abolish the structure needed for enhanced localization, while pairing of the two (L) copies of the transgene would leave only the inefficient (R) copy exposed for ectopic interactions with heterochromatin (Figure IV-4b). Clearly, this is not so; *V21*(LR) / *92C* (L) has the strongest mutant phenotype of any allelic combination tested. Alternatively, pairing of multiple transgenes both in *cis* and in *trans* might be a fully cooperative process (Figure IV-4c).

If this were so, we would expect that both (L) and (R) alleles of 92C would pair well with V21(LR), but that 92C (L) alleles would pair better with V21 (L) alleles, while 92C (R) alleles would pair better with V21 (R) alleles (Figure IV-3). Though this model does not predict whether pairing of single alleles in *trans* would enhance or suppress variegation, it does predict that increased pairing would affect both pairs of alleles in the same way. In other words, if *V21*(L) / *92C* (L) is more variegated than *V21*(L) / *92C* (R), then *V21*(R) / *92C* (R) should also be more variegated than V21(R) / *92C* (L). However, this is not so. Instead, the L/L combination is the lightest, the R/R combination is the darkest, and the other two combinations fall between these extremes. This result is consistent with a model whereby single elements on either or both chromosomes contribute independently to the inactivation potential of the region, with (L) orientations enhancing inactivation. At the same time, orientation preferences do not seem to interfere with the stronger, enhancing effect of inverted duplications on the V21 chromosome. Thus, results of *cis-* and *trans-* pairing taken together are consistent with a model where local intrachromosomal pairing of inverted duplications is strong, while equally local interchromosomal pairing is weaker (Figure IV-4d). Finally, it should be pointed out that linkage enhancer lines (*98v1* derivatives) described in chapter III also support this model, in that they *trans-*inactivate 92C (L) lines strongly, but only weakly inactivate 92C (R) lines.

TABLE IV-I
Orientation and *trans*-inactivation

Allele name (orientation)[1]	*92C*(L)	*4D-3B* (R)	*4D-5A* (R)	*4D-2B* (R)
V21 (L)	Medium[2]	Weak	Weak	Weak
V21Δ116 (L)	Med.-Strong	Very weak	n.d.	n.d.
V21-E3 (LR)	Strong	Medium	Medium	n.d.
V21-E3R1 (R)	Weak	Almost none	n.d.	None(?)
98V1XEFe1(M)	Strong	Almost none	n.d.	n.d.
98V1XIe1(M)	Strong	n.d.	Very weak	n.d.
98V1XJm1(M)	Medium	n.d.	n.d.	Almost none

[1] (L) = Left, (R) = Right, (LR) = Left+Right (Divergent), (M) = Multiple (Complex array, at least 2 L copies). Column lists *cis*-alleles, row lists *trans*-alleles.

[2] Terms are relative, and specific to this table. "Very weak" and "Almost none" differ in that the former consistently showed one or two *bw*⁻ facets per eye, while the later often had non-variegating eyes, plus occasional eyes with several *bw*⁻ facets.

Figure IV-1. Inverting *P[bw+]* at 92C. Potential single-copy lines were tested for orientation by looking at the size of a flanking fragment (top) and promising lines were characterized with additional enzymes (bottom); they are indistinguishable from each other and from *V21* (R) alleles by these criteria.

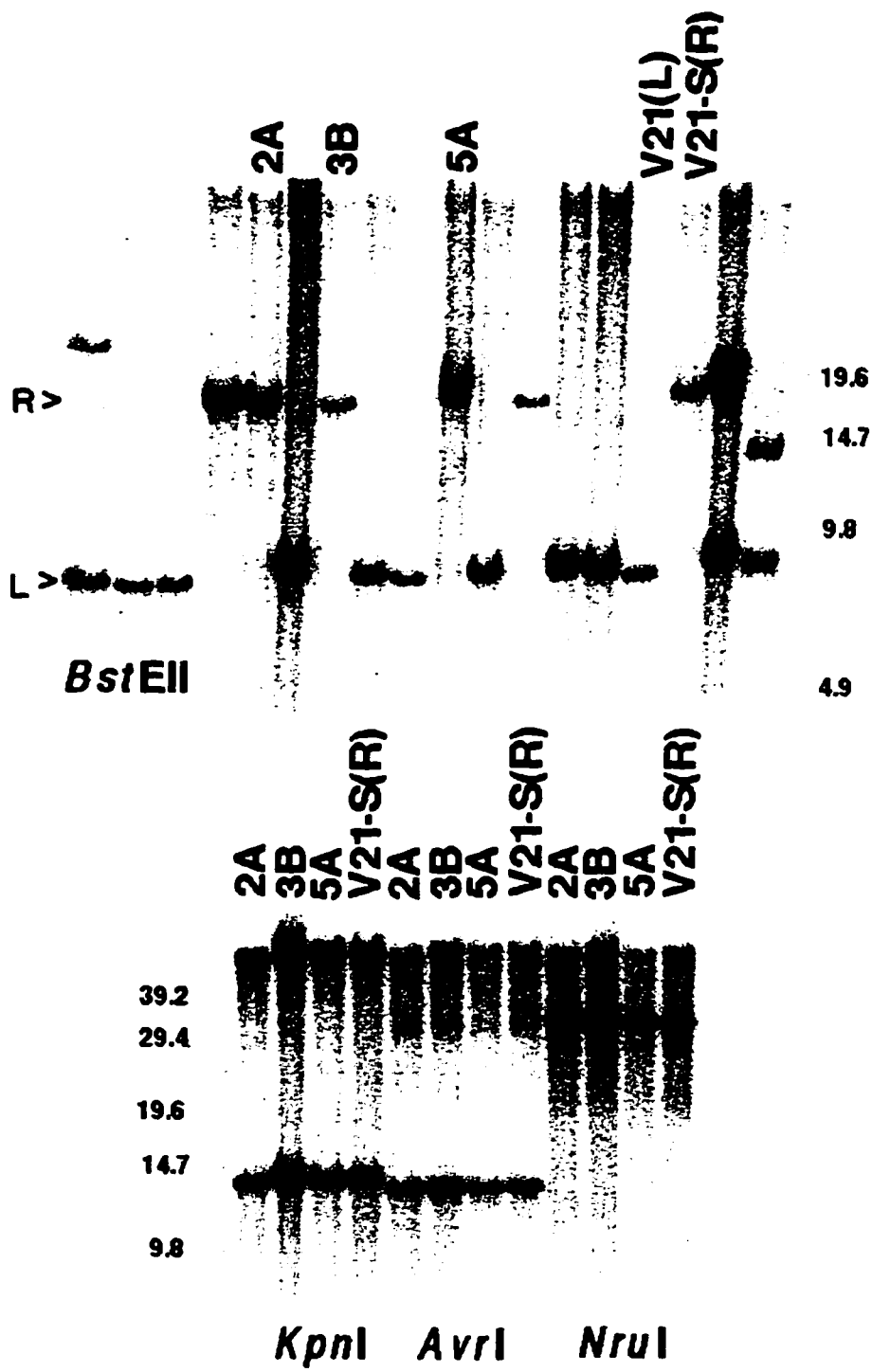

2A
3B
5A
V21(L)
V21-S(R)
R>
L>
19.6
14.7
9.8
4.9
BstEII
2A
3B
5A
V21-S(R)
2A
3B
5A
V21-S(R)
2A
3B
5A
V21-S(R)
39.2
29.4
19.6
14.7
9.8
KpnI
AvrI
NruI

Figure IV-2. Flipping a transgene. Dots indicate unhindered pairing. a) Pairing interactions should be roughly identical for L or R alleles pairing with LR alleles. However, local pairing should be very different for combinations of unduplicated alleles. b) L / L and R / R are predicted to pair well. c) L / R and R / L are predicted to pair less well.

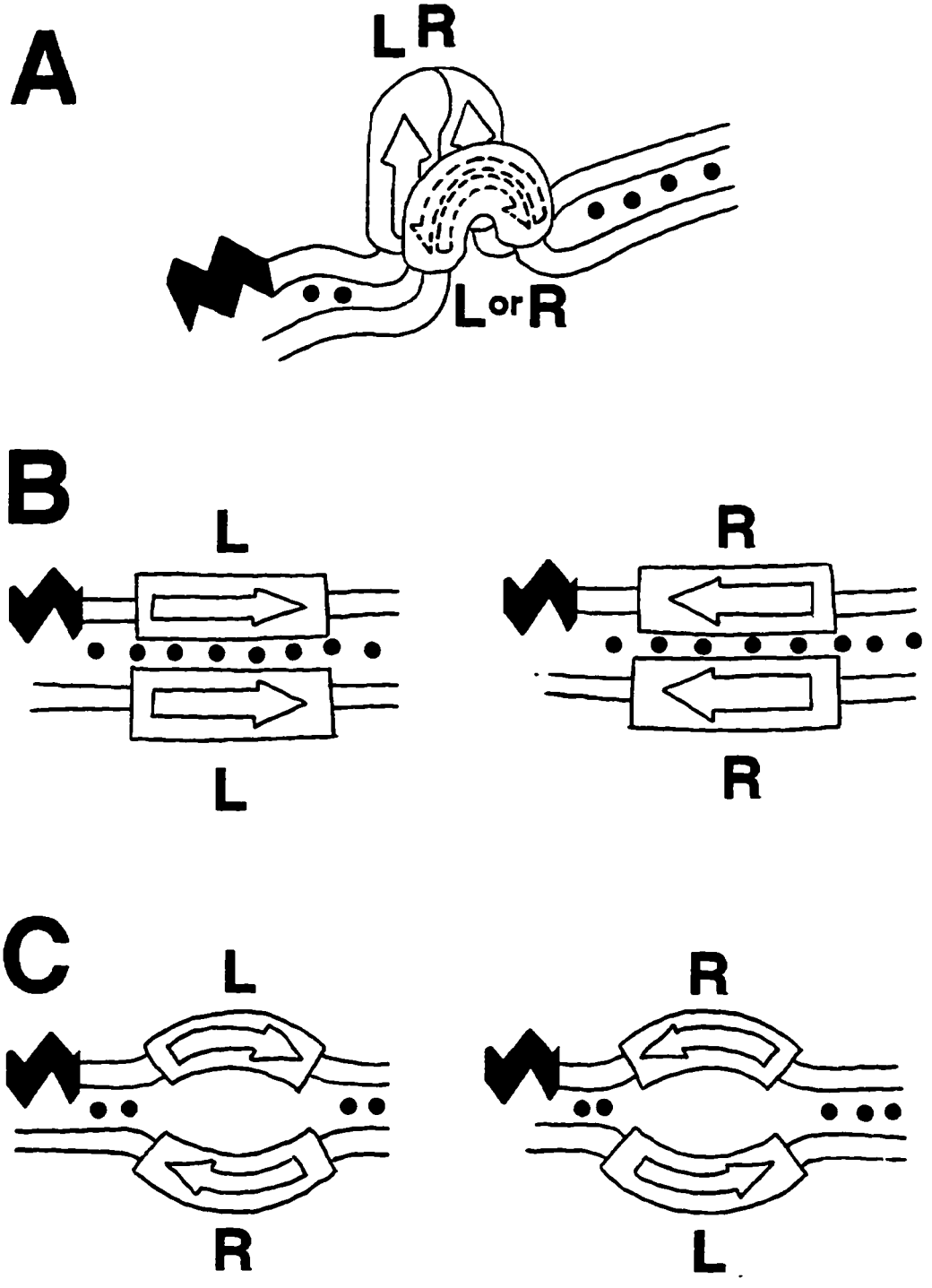

A
L R
L or R
B
L
L
R
R
C
L
R
R
L

Figure IV-3. Orientation controls susceptibility to *trans*-inactivation. a) orientation differences between single elements are most easily seen when the *V21* copy of *P[bw+]* is rendered null by a small deletion (*V21-116*). At left, V21-116 (L) / 92C-3B (R); at right, V21-116 (L) / 92C (L). This result would be predicted by a pairing model. b) Unexpectedly, an even more striking difference is seen for *trans* inactivation by a duplicated transgene. Clockwise from top left: V21-E3 (LR) / --; V21-E3 (LR) / 92C-3B (R); V21-E3 (LR) / 92C-5A (R); V21-E3 (LR) / 92C (L). c) 92C (R) alleles were less trans-inactivated in all tests, even with 98v1 derivatives: clockwise from top left, 98v1XIJm1 / --; -- / 92C (L); 98v1XIJm1 / 92C (R); 98v1XIJm1 / 92C (L).

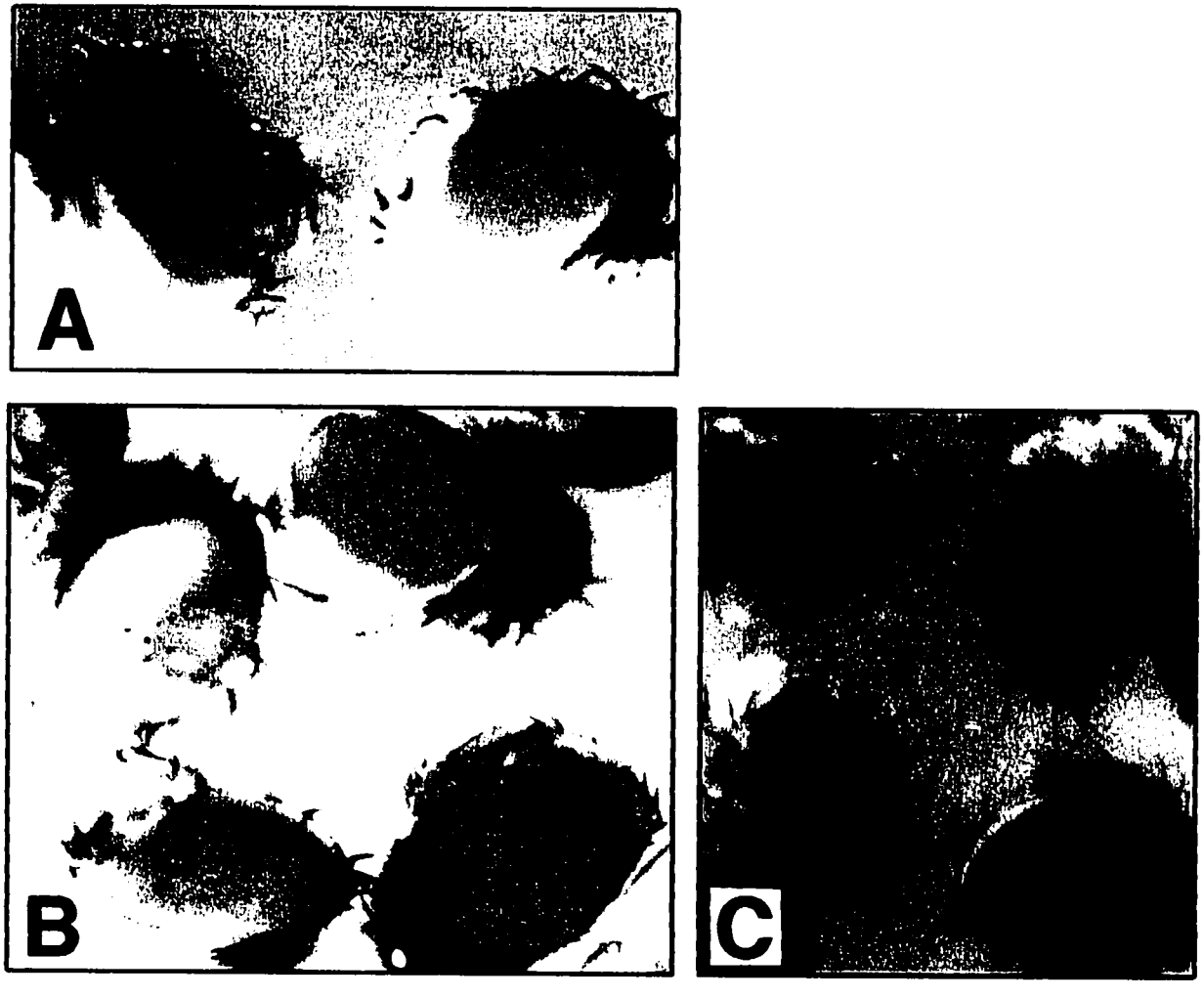
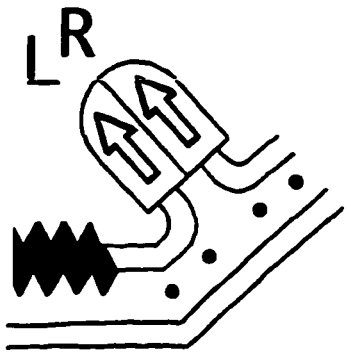
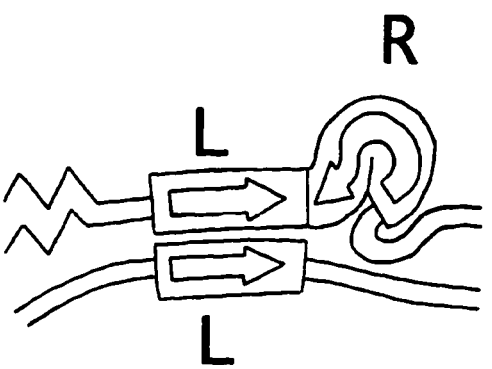

A
B
C

Figure IV-4. Models for local pairing. Arrows indicate transgenes; dots indicate potential pairing. a) Pairing of transgene duplications is proposed to have a critical effect in enhancing PEV. b) If *cis* and *trans* pairing were competitive processes, and it *trans*-pairing were stronger, then the disruption of *cis*-pairing would be expected to suppress PEV in two ways. First, by disrupting the "hairpin" which is proposed to act as a signal for co-localization with heterochromatin, the *trans*-copy would strongly reduce the heterochromatic nature of its homologue. Secondly, a disruption of cis-pairing would be predicted to "release" an R copy of the transgene for ectopic contacts with heterochromatin; however, we already know (chapter II) that R copies are not effective in this role. b) If *cis* and *trans* pairing were fully cooperative processes, then we might expect that L and R transgenes would be inactivated equally efficiently by LR alleles of *V21*--after all, with full pairing, there is neither the need nor the opportunity for the *trans*-inactivated allele to interact with heterochromatin directly. d) instead, we find that the orientation of the 92C allele has a very strong effect on the level of trans-inactivation, most strikingly in the presence of the *V21-E3* (LR) allele. A model where strong *cis*-pairing of the *V21* LR transgenes excludes the L or R allelic transgene is consistent with this unexpected autonomy. In turn, this finding suggests that earlier experiments involving partial transgene deletions may need to be reinterpreted.

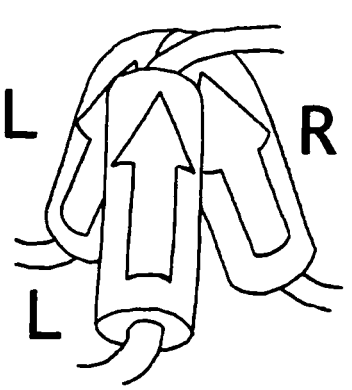
A



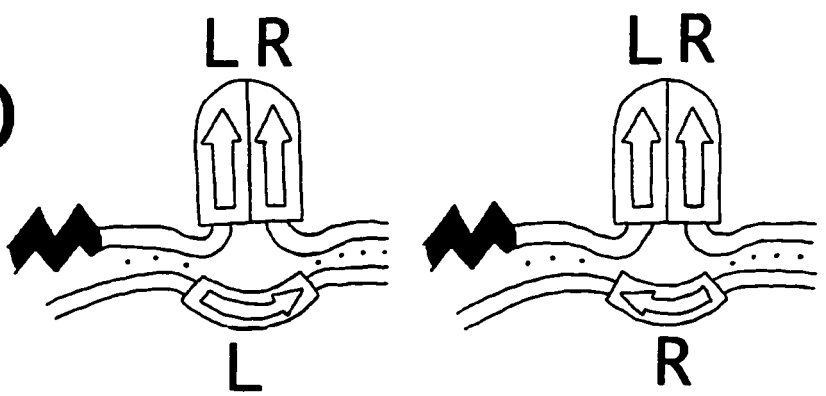
B



C



D



CHAPTER V

P[PRAT-bw+] Repeat Expansions

INTRODUCTION

Historically, certain types of genes in a subset of *Drosophila* tissues have been disproportionately used to study PEV. For example, of the intensively studied genes *white, roughest, brown, yellow, Stubble* and *Notch*, three are eye specific. Variegating alleles of essential genes can be recognized through Y-chromosome-dosage dependent lethality; on this basis, Lindsley *et al.* (1960) and Eddington *et al.* (1962) found that ~20% of sex-linked, X-ray induced recessive lethals were PEV alleles. Baker (1968) concluded that "in fact, there is no evidence against the assumption that every gene is subject to [PEV]." This may be so; however, the fact remains that a majority of the work on PEV has been done on a small fraction of the known variegating genes.

The study of repeat-induced heterochromatin formation (Dorer and Henikoff 1994) and repeat-induced enhancement of PEV (Chapter II), still in its infancy, shows more bias than the field of PEV research as a whole. So far, all examples of repeat-induced or repeat-enhanced PEV known to us (personal communication, A. Csink, D. Dorer, L. Martin-Morris) use eye-color gene constructs as a reporter system (*mini-white*, a second *white*$^+$ transgene, *brown*$^+$ as described above, and a 3.8 Kb mini-*brown*$^+$). We

originally felt that two genes (in different vectors) should serve as an adequate control for gene-specific or allele-specific effects. On the basis of evidence from these two genes we concluded that increased local repetitiveness increases the susceptibility of a region to PEV, and argued that enhancement of local heterochromatin formation was likely to be a natural property of many, or even most, repeated sequences. However, while there is no detectable sequence conservation in their regulatory regions, it has been pointed out that the *brown*$^+$ and *white*$^+$ genes may be more similar than their seemingly unrelated sequences would suggest: First, they each code for one of a pair of structurally-related membrane transport proteins that interact, probably as heteromeric complexes, in eye pigment deposition (Dreesen, Johnson and Henikoff 1988). Besides coding for functionally related proteins, *brown*$^+$ and *white*$^+$ are also coordinately regulated in the eye, for example by the *Inr-a* modifier gene (Rabinow, Nguyen-Huynh and Birchler 1991). Though the *mini-white* reporter construct used by Dorer and Henikoff (1994) lacks most *white*-derived regulatory sequences and the first intron of the *white*$^+$ structural gene, the possibility remained that repetition of eye-specific regulatory sequences, rather than repetition *per se*, was the real cause of silencing. In a similar vein, part of the *pUC*-derived vector sequences could theoretically be responsible for enhanced variegation, as it, too, is shared between *P[lac-w]* and *P[bw$^+$]*. I therefore planned to amplify a construct that would differ in several ways from both of the two previous constructs. The *P[PRAT-bw$^+$]* construct (Figure V-1) contains the *brown* cDNA (providing an eye pigment phenotype) under the control of 5′ and 3′ regulatory regions from a housekeeping gene, *PRAT*, which controls the rate-limiting step in *de-novo* purine biosynthesis (Clark 1994). *P[PRAT-bw$^+$]* loses all *pUC*-derived sequences upon integration. *P[PRAT-bw$^+$]* does contain the

rosy$^+$ gene (*ry*$^+$), which is named for its effect on eye color. However, the *rosy*$^+$ gene codes for xanthine dehydrogenase (Xdh) which is not functionally related to the brown and white proteins. Furthermore, the *ry*$^+$ gene is neither cell-autonomous nor transcribed in the eye, and would not, therefore, be expected to contain related tissue-specific regulatory regions.

The experimental design of the *P[PRAT-bw*$^+$*]* amplification screen was intended to counter another criticism frequently encountered in the early stages of the *P[bw*$^+$*] / V21-E* repeat array experiments. While we could show that enhanced lines contained extra repeats of the transgene, we could not show that all transgene repeat arrays enhanced variegation. Because one copy of *P[PRAT-bw*$^+$*]* produces a very pale eye color (Figure V-IIb), I reasoned that it would be possible to select for increased gene dosage by phenotype, to define close linkage by the absence of recombination over several generations, and to distinguish arrays by Southern analysis, all in the absence of heterochromatin or PEV. At this point, I hoped to induce variegation by X-irradiation, and then demonstrate that variegation could be suppressed by reduction of *P[PRAT-bw*$^+$*]* copy number.

We eventually recognized that transgene arrays, differing in size by multiples of a single repeat unit, and located at a single euchromatic site were a useful tool for other, more general inquiries into chromosome structure and chromatin packing. Thus, instead of focusing solely on (unsuccessful) attempts to induce PEV, I also began the characterization of the arrays, to provide a system for cytological observation of DNA compaction and morphology.

Banding, compaction, and gene activity

In *Drosophila*, it has long been recognized that puffing -- a disruption of banding in polytene chromosomes -- might correlate with local gene activation (Beermann 1963; Callan 1992, for review). Conversely, it has been suggested that increased definition in the familiar polytene banding patterns of salivary nuclei might be due, in part, to the inactivation of other, less-strongly transcribed genes. Banding patterns in polytene nuclei (and condensed regions in diploid nuclei as well) might, therefore, mirror genetic regulation: "The chromomeric pattern...may reflect]... a subdivision of the chromosome which primarily is of operational importance and which is only realized when the independent operation of functional units becomes a necessity, e.g. in cellular differentiation (Beermann 1963)."

On the other hand, there is evidence that overall polytene morphology need not follow local gene activity. First, the banding pattern of polytene chromosomes in other tissues (e.g. in nuclei of nurse cells from *otu* mutant lines) is fairly invariant from that of salivary glands in its major details, though the two tissues have distinct functions. At minimum, this means that banding does not generally reflect tissue-specific suppression of genes. However, as many as 95% of active puffs (including transcribed interband regions) might represent housekeeping genes (Berendes 1966), this argument is inconclusive.

The idea that a morphologically distinct region would equate with a structural gene seemed very appealing; it was suggested that gene transcription, translation, morphology and identity would all be controlled at the same, few, local points (Nagl 1978, reviewed p. 4-7). For decades, the

size and number of polytene chromosomes bands seemed more than adequate in to contain all the known and predicted *Drosophila* genes (Bridges 1938; Beermann 1967). However, before the discovery of introns and of mRNA splicing, genes seemed to be much to small to span even the smallest of bands, suggesting that something else must be present (regulatory regions, repetitive sequences, etc.), and necessarily clouding the issue of how a gene and a band might align. As the size and number of eukaryotic genes was recognized, the problem quickly changed -- the number of bands (~5000), which had originally seemed like a generous allotment, was considerably smaller than the number of predicted genes (Nagl 1978, p. 5).

With the advent of easy cloning of chromosomal regions, and precise DNA and RNA *in situ* hybridization techniques, it has been shown that many genes occupied only part of a band; that bands can partially decondense; that some genes seemed to extend into interband sequences; and that several genes can "share" a band (Semeshin 1986, 1989; Rykowski *et al.* 1988).

Micro dissection, cloning, electron microscopy and fine-scale genetic analysis have been used to determine packing ratios, theoretically with great precision. Kozlova *et al.*, for example, report that the DNA packing ratio may be as little as 17-77 for band 10A4-5 (a "light" band) or as much as 151-161 for band 10A1-2 (a "heavy" band). This corresponds to a packing ratio of roughly 2-20 times that of simple nucleosomal packaging.

When done with natural chromatin, the above methods have technical drawbacks. The precise edge of a band or interband can be hard to define;

furthermore, it is difficult to map genes (or gene-free regions) within large bands due to loss of optical contrast in otherwise informative rearrangements (Zhimulev, Semeshin and Belyaeva 1981). In sum, cytological measurement of small bands is proportionally less precise because the band "edge" makes up a larger proportion a small band, while determination of total DNA length is difficult for large bands.

Historically, naturally occurring tandem amplification events could be easily recognized at specific sites such as *Bar* and *Beadex* (Ashburner 1989, p. 607 for review). Because of their internal consistency, they were useful for analysis of sequence-specific chromatin packing. More recently, transgene integration events have been particularly prized as well-defined molecular yardsticks (Semeshin *et al.* 1986; Semeshin *et al.* 1989). The *P[PRAT-bw+]* array series has promise as a yet-more-powerful tool for the fine-scale cytological analysis needed to explain the easily-observed but still-unexplained packing phenomena that produce the band/interband pattern characteristic of polytene chromosomes.

MATERIALS AND METHODS

Fly strains and culture conditions: All conditions were as described in previous chapters, except as noted below.

Construction and introduction of *P[PRAT-bw+]* : The plasmid construction, transformation and first amplification step were done by Denise Clark to provide a reporter system for screening for suppressors and enhancers of *PRAT* (unpublished). When integrated, the

P[PRAT-bw+] construct contains a *brown+* cDNA attached to 5′ and 3′ *PRAT* flanking sequences, followed by the *rosy+* gene (7 Kb) in the same orientation.

Three independent *P[PRAT-bw+]* transformants, *D18-2*, *E1-2*, and *D9-1*, were made by co-injection of the construct and a "wings clipped" transposase source (Karess and Rubin 1984) into a ry^{506} mutant line, followed by screening for ry^{+} expression (D. Clark, personal communication). After introduction of the bw^{D} and *st* mutations, and of ry^{+} in place of ry^{506}, a transposon duplication scheme was carried out by Denise Clark on line *D18-2* (like that for *92C*, as described in chapter III). I received five of the resulting lines.

Transposase Mutagenesis: Several, similar rounds of mutagenesis were performed to produce transgene insertion arrays. In the first round, males from *E1-2*, *D9-1* and each of the five *D18-2*-derived "duplication" lines were crossed to bw^{D}; *st* Δ*2,3* virgin females. In all seven cases sectoring was seen in the F1 generation, showing that each line contained a mobilizable transposon. *P[PRAT-bw+]* /Δ*2,3* males were crossed to bw^{D}; *st* virgin females, and male progeny were scored (Table V-I, lines 1-7). Flies with suspected expansions (i.e. darker eyes) and no transposase source (eyes not sectored) made up 0.6-3.4% of the progeny. These males were crossed to bw^{D}; *st* virgin females; if the pigment in their progeny was uniform, females were crossed to bw^{D}; *st* males for one or more generations to see if the additional copies of the transposon would segregate. At this low level of expansion, lines were discarded if any pale segregants were seen. Southern and pigment analyses were carried out on a subset of the lines during this period.

For the next round of mutagenesis, five of the seven lines were represented by one or two apparent expansion derivatives. It was hoped that further array expansion might begin to produce variegating derivatives, therefore, males were crossed to *bw^{D}; st, Sb, Δ2,3 /Payne* virgin females, so as not to mistake the sectored phenotype of the transposase-containing flies for true variegation. Otherwise, the second mutagenesis was identical to the first. There were sectored F1 flies in all lines except the two derived from *E1-2* (which was therefore abandoned as being insensitive to transposase). Dark-eyed flies were produced at a rate of 0.9-4%, and a large class of flies with intermediate pigment levels was also produced (3.9-13% of total progeny; Table V-I, lines 8-11). Based on a higher-than expected number of pale pigmented progeny, one of the two *D9-1* derivatives was reclassified as a transposable mobilization event rather than a local duplication, and the line was abandoned (not shown).

As in the previous screen, males were recovered and then females were used to check for linkage. However, at this stage, a few of the darker lines were kept even if very pale-eyed flies were seen among the progeny. On the other hand, lines that showed a progressive loss of very dark-eyed flies or new classes of flies with moderately-pigmented eyes were thrown out. This counter-intuitive selection scheme was based on the assumption that lines could develop several transposons' worth of additional pigment in two ways: either an array could expand significantly, or numerous transposons could move to new sites (together or separately). Of course, it was also possible to have both an expanded array and one or two scattered transposons. Transposons in "clean" expansions would never segregate, thus lines containing them would not produce any pale-eyed flies. In contrast, simple mobilizations would segregate, producing flies with

moderately-pigmented eyes as well as flies with light eyes and flies with dark eyes. Finally, "contaminated" expansions -- that is, lines with both expanded arrays and a stray transposon insertion -- would produce occasional pale-eyed flies (with the segregation of the stray transposon away from the array) but never flies with moderate pigment, because the array would remain intact. By passaging "dirty" lines through one or two hemizygous males, the array could be recovered without "contaminating" transposons.

Pigment Analysis: Only preliminary analysis of pigment levels was carried out; thus, only one or two samples were taken from each genotype. Pigment was extracted from one individual head per genotype in ~0.25 ml of acidified ethanol (30% EtOH, pH adjusted to 2.0 with HCl) by gentle inversion for three days, and measured at OD_{480} in a 0.2 ml quartz cuvette.

Irradiation Mutagenesis: Several lines were irradiated (see Table V-II). Males were collected and irradiated as described in chapters II and III, and mated in groups of 12-15 to 50-75 bw^D; *st* virgin females (cornmeal-molasses medium, bottles, some 25 degrees, some 18 degrees). At first, males were removed after 4 days. As the low recovery rate and lack of clustering became evident, males were left in. Females were transferred to new food after 7-8 days.

DNA preparation and Southern analysis: The 5′ and 3′ regions of *PRAT* are not yet fully sequenced, and restriction mapping has focused on common enzymes. Because *P[PRAT-bw+]* is a large transposon, it contains restriction sites for most available enzymes. Enzymes that cut at only one

site (as determined using the plasmid form of *P[PRAT-bw+]* , which contains an extra ~3 Kb of vector sequences) include *Kpn*I, *Mlu*I, *Stu*I, *Nhe*I, *Spe*I and *Rsr*II. Enzymes having no restriction sites within *P[PRAT-bw+]* include *Avr*II, *Eag*I, and *Not*I. *Avr*II and *Eag*I were both used to size transgene arrays. At 65A10-11, the endogenous *Eag*I fragment that is expanded by the array is only about 5 Kb; the endogenous *Avr*II fragment is about 60 Kb. The small size of the *Eag*I fragment allows considerable precision in sizing the arrays, while comparable expansion of the large *Avr*II fragment insures that deletions of the *Eag*I site are not occurring, and being mistaken for expansion events. The *Stu*I site lies within *ry+*; by cutting with *Stu*I and probing sequentially with *brown* and 3′ *rosy* sequences, tandem arrays can be recognized. Furthermore, sizes can be predicted for both 5′-5′ and 3′-3′ inverted duplications, which are uniquely recognized by *brown-* and *rosy*-derived probes, respectively. The *brown+* probe was prepared by in-gel random priming (BRL) as in chapter III; the 3′ *ry+* probe was prepared from a PCR fragment of approximately 200 bp amplified by Adrian Quintanilla.

Gels, blotting and hybridization were done by standard methods, and as described in chapter II, except that plugs for CHEF and FIGE pulsed-field analysis were made from a simple nuclei mini-preparation, as adapted from Bingham *et al.* (1981): 50-100 flies are homogenized, filtered through nitex and pelleted three times (8.5K in a microcentrifuge for 7 min., with resuspension by gentle pipetting). In a slight variant on the protocol of Karpen and Spradling (1990) nuclei are then resuspended in 50-100 microliters of Buffer A mixture and processed at 25-50 flies/200 microliter plug. Restriction digests (20-40 units of enzyme/digest; 6-16 hours, 37°) are done on 20-25 microliters worth of the plug (1/10-1/8 plug) after perfusing

in enzyme (4°C for 6-16 hours) in the appropriate buffer.

Polytene cytology: *in situ* analysis was carried out as described in chapter II; *brown* + probe was made by in-gel labeling with the Bioprime kit (BRL), using double the recommended amount of polymerase, and incubating for 4 hours.

RESULTS

Large transgene repeat arrays can be produced by phenotypic selection: At three (presumably) different sites of transgene integration, putative transposase-induced duplications were retrieved at a rate of ~1.2-3%. Fortuitously, expression of the *P[PRAT-bw+]* construct varied by a factor of at least two at different genomic locations, which turned out to be of use in differentiating jumps from duplication events. As listed in Table VIII-II, most duplication events share a fairly consistent ("dark") phenotype, while many jumps vary slightly in their pigment levels ("med" class). At two locations, further attempts at amplification were unsuccessful. For one integrant, *E1-2*, this resistance to amplification was prefigured by a total absence of transgene-induced somatic sectoring in the two duplication alleles used. The other line, *D9-1*, produced 3 strong putative amplification alleles, but these were all mobilization events as determined by recombination (two lines) or Southern analysis (line *D9D6*, a very dark line). In general, we assume that transgene inactivity stems from damage to transposon ends, though the formation of locally inaccessible chromatin which excludes transposase could also theoretically occur. The *E1-2* line was not studied further. The *D9-1* line was also

abandoned. It is not known why amplification events were not detected in this line. It is possible that the baseline pigmentation of a *P[PRAT-bw+]* transposon at this site is too similar to the "average" expression level, so that amplification and mobilization events are harder to distinguish.

In contrast, the third site supported a high degree of transgene amplification. Amplification was aided by two phenomena: first, the primary amplification lines tested retained active elements, allowing a second and third round of amplification {lines *1d(2)*, *13d(2)* and *25d3(4)*}. Next, the combination of transposon, site, and screening procedure favors recovery of large expansions, including expansions that double, or more-than-double the size of the transgene array. Such expansions occurred independently at least four times (1 copy --> 4 copies; 3 copies --> 6 and 7 copies; 4 copies --> 13 copies), that being 0.07% of events at that site. Other increases of three or more copies at a time made up an additional 0.05% of all events. In each case, mutagenized flies hemizygous for the *P[PRAT-bw+]* -array were compared to homozygous adults from the parent line to provide a benchmark for a meaningful increase in pigment. It was later determined that homozygous flies have more than twice the pigment level of hemizygous flies; therefore, the selection criterion for "dark" flies was very strict.

Large *P[PRAT-bw+]* arrays at 65A10-11 and induction of PEV

The original intent of the *P[PRAT-bw+]* -array scheme was to provide arrays on which variegation could be induced, to demonstrate three things. First, I wanted to test whether array size and variegation would correlate when the array in question was not amplified on the basis of enhancement of variegation. Next, I hoped to see whether the *PRAT*

housekeeping gene could variegate, and whether it could participate in repeat-based enhancement of this variegation. Finally, I wanted to test whether *pUC*-derived sequences were necessary for repeat-enhancement of PEV.

Because some genomic sites may not support PEV of adjacent genes, I intended to mutagenize *P[PRAT-bw+]* arrays at two (or more) locations. However, the irradiation screen was carried out before *in situ* localization of the *P[PRAT-bw+]* arrays. At that time, it was assumed that *25d(4)A(13)* [called *A(13)*] represented a mobilization of one or more *P[PRAT-bw+]* transposons to a new location. In turn, we assumed that *25A1*3(16)*, *25A3*1(19)*, and *25A5* were moderate array expansions at the new location. Finally, it seemed likely that the stable, dark allele *D9D6* represented an expansion at the *D9-1* site (location unknown). These assumptions were incorrect. Instead, all expansion lines derived from D18-2 remain at the original insertion site; furthermore, the dark derivative of D9-1 does not involve amplification of the element. Therefore, instead of irradiating arrays at three sites as intended, I only irradiated many related arrays at the same site, and also what appears to be a single element at some other (unknown) site.

A few spotted (as well as sectored) flies were produced at a rate of 0.05%, but these either were sterile or the phenotype was not heritable. Do these flies demonstrate that the *P[PRAT-bw+]* array can variegate? Irradiation can produce non-uniform phenotypes in several ways. Thus, it is possible that the sectored "variegating" flies (see Table V-II) represented delayed strand repair or improper chromosome segregation following radiation damage. The two flies with classic variegation phenotypes (0.008% of

progeny, see Table V-II) are harder to account for, except as examples of PEV. One fly might have been an XO male, with a variegating mutation of *white* on the male-derived X chromosome; alternatively, he could represent spontaneous mutation at the *white* locus in the female parent, or a novel case of dominant PEV for another eye-color gene. The other, a female with one variegating eye and one normal eye, may actually have been a gynandromorph, with white variegation uncovered (and enhanced) in the XO sector; "she" only produced normally-pigmented progeny. Therefore, we still do not know whether *PRAT* can support variegation, and if so, whether this variegation would be subject to repeat-enhancement.

Inserts into the 65A10-11 band have distinct morphologies:
Hybridization with *bw*$^+$ sequences showed that the original *D18-2* insert, the *1d2E6(5)* insert, and the *25A1*3(16),25A3*1(19)*, and *25A5* inserts were all at the same site: 65A10-11, a thin, sharp band known to contain the laminin A gene (Flybase). On closer observation, unhybridized polytene chromosomes showed striking changes in the banding pattern in 65A10-11 that were dependent on the size of the transgene array. Hemizygotes were examined first, so that pairing could be used to determine the endogenous 65A10-11 band, then homozygotes were also used to eliminate distortion, which was fairly severe for hemizygotes involving the larger arrays. When only one copy of the transgene is present, the endogenous band is split by a thin interband into two visibly thinner bands, both of which pair with the unsplit band on the homologous chromosome. A single copy of *P[PRAT-bw*$^+$*]* appears as an interband which splits the pale endogenous band; that is, it is fairly highly decompacted. At the other extreme, the *25A1*3(16)* and *25A3*1(19)*arrays produces a very wide, fuzzy band that is

distinctive enough to serve as a landmark band. While not as dark as other large bands (and presumably not as compacted), the new band is much darker (and presumably more compact) than the interband seen upon insertion of a single *P[PRAT-bw+]* transposon (Figure V-2c). Duplications of the transgene *1d(2)* and *13d(2)* show a faint pattern of bands and interbands comparable to that seen for other transgene arrays (Semeshin *et al.* 1986; Semeshin *et al.* 1989). Hints of this pattern are visible in slightly larger arrays; but at an intermediate level of amplification (e.g. *1d2E6(5)*, a broad puff is seen that presumably includes the split endogenous band at each end, and similarly compacted (and thus indistinguishable) chromatin from the transposon in the middle. In essence, it appears that even in this non-variegating transgene array, increased repetitiveness of the region may correlate with some change in the chromatin state.

Infertility may reflect local effects of array expansion: In the largest array, *25A3*1(19)*, female sterility and low male fertility are evident. Some depression in breeding efficiency and viability were also seen for several generations of *25A1*3(16)* and *25A5*, however this was never quantified and now seems to have been suppressed. Loss of fecundity may be due to repeat-induced PEV on the *laminin A* gene, whose disruption is known to have pleiotropic effects including infertility (Flybase), or on another nearby gene. This could be tested by looking for increased fertility following reduction in transgene copy number. If PEV can be shown to affect a nearby gene, this implies that the *PRAT-bw+* construct is probably insensitive to PEV. Should this be the case, Northern analysis could be used to check whether the *ry+* gene in the *P[PRAT-bw+]* construct is being affected by PEV as well. Alternatively, pleiotropic effects on fertility and viability could be due to the titration of a *PRAT*-specific regulatory factor.

As the limiting step in the *de novo* purine biosynthesis pathway, *PRAT* must have one (or more) regulatory proteins that are responsible for its tight regulation. The addition of 16 or more extra copies of the *PRAT* regulatory region (32 in homozygotes) could have a significant impact on this pathway).

DISCUSSION

Array packing: Semeshin *et al.* (*op. cit.*) have described compaction, decompaction, and banding of several short transgene arrays. They find that arrays show a regular banding pattern that is essentially uniform from one *P[ry$^+$]* copy to the next. The formation of distinct bands demonstrates that DNA compaction varies along the length of a single transposon. In contrast, the regularity of the banding pattern probably indicates that packing is consistent between different transposons within the array. Upon activation of genes, decompaction and puffing is seen; in these puffs, physical distances correlate with the known repeat length, though for complex transposons, puffing may be restricted to that portion of a band where the activatable gene resides. In this sense, the *P[PRAT-bw$^+$]* array seems to have properties of a constitutively active puff. Furthermore, the cytology of the array is consistent with this idea.

Insensitivity to variegation: *P[PRAT-bw$^+$]* arrays at 65A10-11 may not support variegation. This could be due to a dominant lethal effect of genes in nearby euchromatin; to a lack of "fit" between heterochromatin and nearby euchromatin; or to some property of *P[PRAT-bw$^+$]* itself. At this time, there is no way to distinguish between these possibilities. The first

two suggestions have previously been invoked when variegation screens have failed, and do not demand further discussion. The last possibility is more interesting. It is possible, for example, that the *brown* and the *white* genes really do share some factor that makes them unusually sensitive to repeat-induced gene silencing; if so, this "element" is retained in the *mini-white* construct, which retains a minimal capacity for autonomous regulation. Similarly, it is possible that pUC effectively mediates packing, perhaps by virtue of its intrinsic packing properties, or by protein binding properties (i.e. in the mode of Zuckerkandl -- see Figure I-2c). We have not favored these explanations for repeat-based PEV. Nevertheless, as we generate internal transposon deletions in the course of mobilizing variegating transgenes, we routinely recover and discard lines that could answer this question definitively. If repeat-based inactivation turns out to be unexpectedly uncommon or elusive in *Drosophila*, these might bear closer examination.

Following a path of greater interest, I would like to put forth an alternative explanation for the behavior of *P[PRAT-bw⁺]*. I feel it is possible that the *P[PRAT-bw⁺]* array's resistance to PEV and its distinctive, puffy appearance share an etiology. The *PRAT* gene is a housekeeping gene; as such, it seems to be active at all stages (Clark 1994, and personal communication). It is possible that activity before cycle 14, when heterochromatic compaction is fully established (Hiraoka *et al.* 1993) prevents the efficient inactivation of *P[PRAT-bw⁺]* even in a heterochromatic milieu; constant, continued activity might also have a protective value. Further work with the *P[PRAT-bw⁺]* construct at different sites, or work with modifiers of PRAT expression (when such are discovered) will be needed to support this idea.

The *P[PRAT-bw+]* array does not variegate, but it does have a distinctive morphology in comparison to other large puffs. It is possible that its own increased repetitiveness is just beginning to cause *PRAT-bw+*to respond to some higher-order packing phenomenon: that is, more copies of the transgene seem to create a band that has increased in visual density (but that is still far less dense than the majority of polytene bands). However, when measurements are made of the distance between flanking bands, a regular progression is seen paralleling the molecularly-determined length of the array (Figure V-2c; S Henikoff, personal communication). Perhaps this implies that a qualitative change occurs before (or in addition to) an increase in the packing density of repeated sequences. Alternatively, it could reflect the high levels of transcription believed to be occurring in this distinctive puff. Experiments are ongoing to determine and compare the interphase cytological arrangement and the transcriptional activity of the *PRAT-bw+*, *rosy+* and (flanking) laminin sequences. It is hoped that the large (up to 325 Kb), precisely sized arrays will eventually allow comparison between packing in polytene and mitotic interphase chromatin.

TABLE V-I
Expanding *P[PRAT-bw+]* arrays

Parent line	Flies (vials)	Dark (%)	Med. (%)	Sterile	Jumps	Name(#)	(Pigment)
Previous screen							
D*18-2*[1]	1423[2] (51)	42 (3%)	na na	nd	nd		(1.0)
						1d(2)	(1.7)
						3d3	(5.3)
						13d(2)	(2.1)
						25d3(4)	(3.1)
						29d2	(3.0)
Screen 1							
E1-2	593[4] (10)	6 (2.0%)	14[5] (4.7%)	1	6		(2.2)
						E1B(2)	nd
						E1E2(2-3)	nd
						E1D(2)	(2.7)
D9-1	662[4] (10)	4 (1.2%)	9[5] (2.7%)	2	2		(1.5)
						D9D(1-2)	nd
						D9A(1-2)	nd
						D9L(2)	(3.6)
						D9M(2)	(3.0)
1d(2)	516[4] (10)	3 (1.2%)	0	2	0		
						1d(2)E(3)	(2.7)
3d3	597[4] (10)	10 (3.4%)	4[5] (1.3%)	0	7		
13d(2)	723[4] (10)	0	24[5] (6.6%)	0	>6		
						13d(2)B(3)	(3.3)
25d3(4)	764[4] (10)	5 (1.3%)	4[5] (1.0%)	0	6		
						25d3(4)A(13)	(11.0)
29d2	562[4] (10)	14 (5.0%)	1[5] (0.4%)	1	>5		
						29d2d3(nd)	(3.9)
Screen 2							
D9L(2)	323 (7)	3 (0.9%)	20[5] (6.2%)	0	9		
						D9D6(<2)	nd
1d2E(3)	407 (9)	4 (1.0%)	16[5] (3.9%)	0	11		
						1d(2)E6(5)	nd
						1d(2)E6'(5)	nd
13d(2)B(3)	421 (9)	17 (4.0%)	39[5] (9.2%)	0	9		
						13d(2)B2(7)	nd
						13d(2)B4(4)	nd
						13d(2)B7(6)	nd
25d3(4)A(13)	401 (12)	8 (2.0%)	52[5] (13%)	5	>4		
						*25A1*3(16)*	nd
						*25A3*1(19)*	nd
						25A5(16-17)	nd

[1] Denise Clark, unpublished
[2] scored for only 3-4 days
[3] reference transposon.

bold face marks copy-number of tandem arrays at 65A10-11
nd=not determined; na= not applicable;
Pigment values =raw OD_{480} values x 10^3 (no replicates)

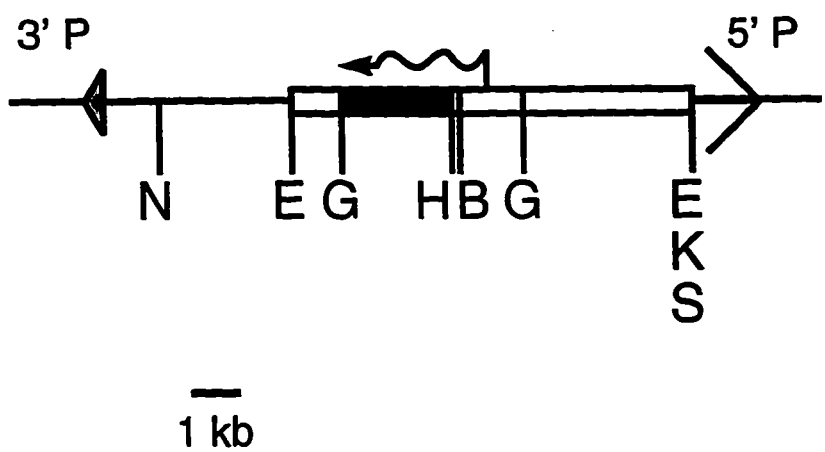
[4] half of total males; test class not determined by color or by absence of a marked balancer
[5] most "medium" putatives not crossed or not kept; dark = progeny is 1.5-2X darker than the parent

TABLE V-II
Irradiation of *P[PRAT-bw+]* arrays

Name(copies)	Temp.	Flies(bottles)	Putatives(%)		Comments[1]
D9D6	25°	683 (2)	2	(0.29%)	1 sectored, 1 var., no var. in progeny
1d(2)E6(5)	25°	874 (4)	0	(<0.1%)	
1d(2)E6'(5)	25°	3103 (9)	3	(0.1%)	*e(var)*-like m., sterile; m. small *bw*⁻ patch; f. *bw*⁻ eye and *bw*⁺ eye with small *bw*⁻ patch, not heritable. Also 3 rough-eye, 1 pale.
	18°	3050 (12)	0	(<.03%)	
13d(2)B2(7)	25°	2951 (9)	1	(0.03%)	small *bw*⁻ patch, not heritable
	18°	884 (4)	0	(<0.1%)	
13d(2)B4(4)	25°	1411 (3)	0	(<0.7%)	two pale flies
13d(2)B7(6)	25°	6483 (19)	3	(.05%)	m. with small *bw*⁻ patch; m. with a red eye and an *e(var)*-like eye; weak var. (f). None heritable.
*25A1*3(16)*	25°	1746 (5)	1	(0.06%)	small *bw*⁻ patch, not heritable.
	18°	1235 (7)	0	(<.08%)	also 1 pale, 1 rough-eye from 18°.
*25A3*1(19)*	25°	1208 (7)	3	(0.2%)	cluster of 3 very weak var. females
	18°	363 (2)	0	(<0.3%)	1 sterile, 2 non-heritable or dying
Total		23991	13	(0.05%)	nothing heritable

[1] Details on variable eyes are given to assess the likelihood that PEV has actually been seen in an array line. m=male. All viable lines were kept for two generations before being discarded as "variegation not heritable."

Figure V-1. *P[PRAT-bw+]* vs. *P[bw+]*. Sequence similarity is limited to the brown coding regions (present as a cDNA in *P[PRAT-bw+]* and as genomic sequence in *P[bw+]*) and the P-element ends.


3' P
5' P
N
E G
H B G
E
K
S
1 kb

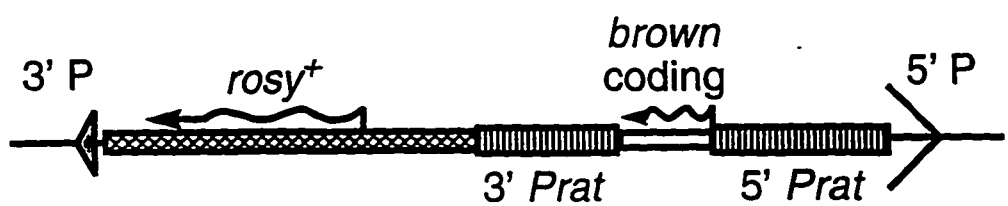

3' P
rosy+
brown
coding
5' P
3' Prat
5' Prat

Figure V-2. *P[PRAT-bw+]* arrays. a) Amplification of a transgene array was confirmed by Southern analysis of CHEF electrophoretic gels; copy number was determined on the basis of size alone. b) darkening of eye color was used to screen for transgene amplifications. c) The polytene cytology of *P[PRAT-bw+]* transgene arrays correlates well with both the molecularly-predicted size of the arrays and the increase in eye color. The original insertion divides a faint band (1); amplification changes the size and optical properties of the band.

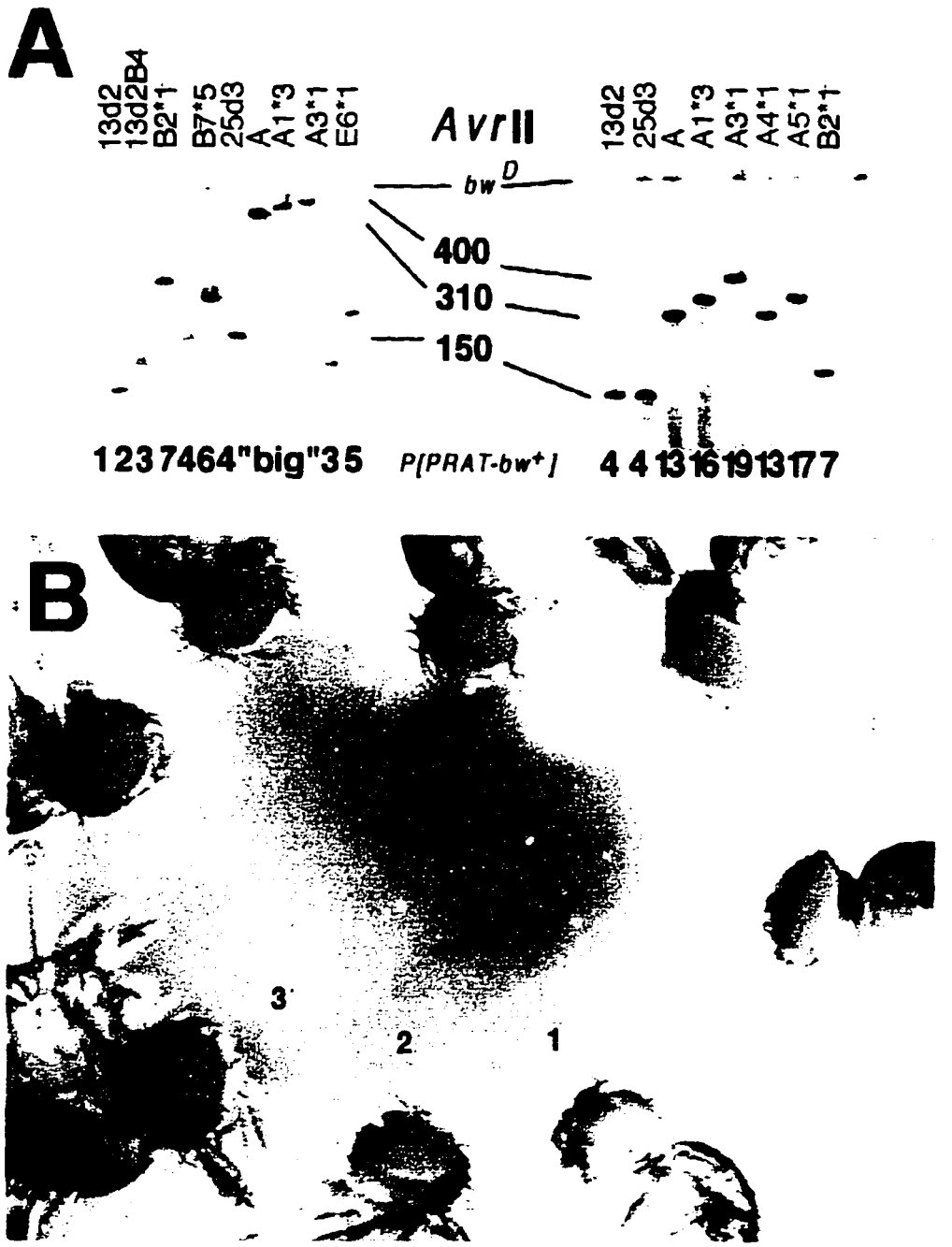
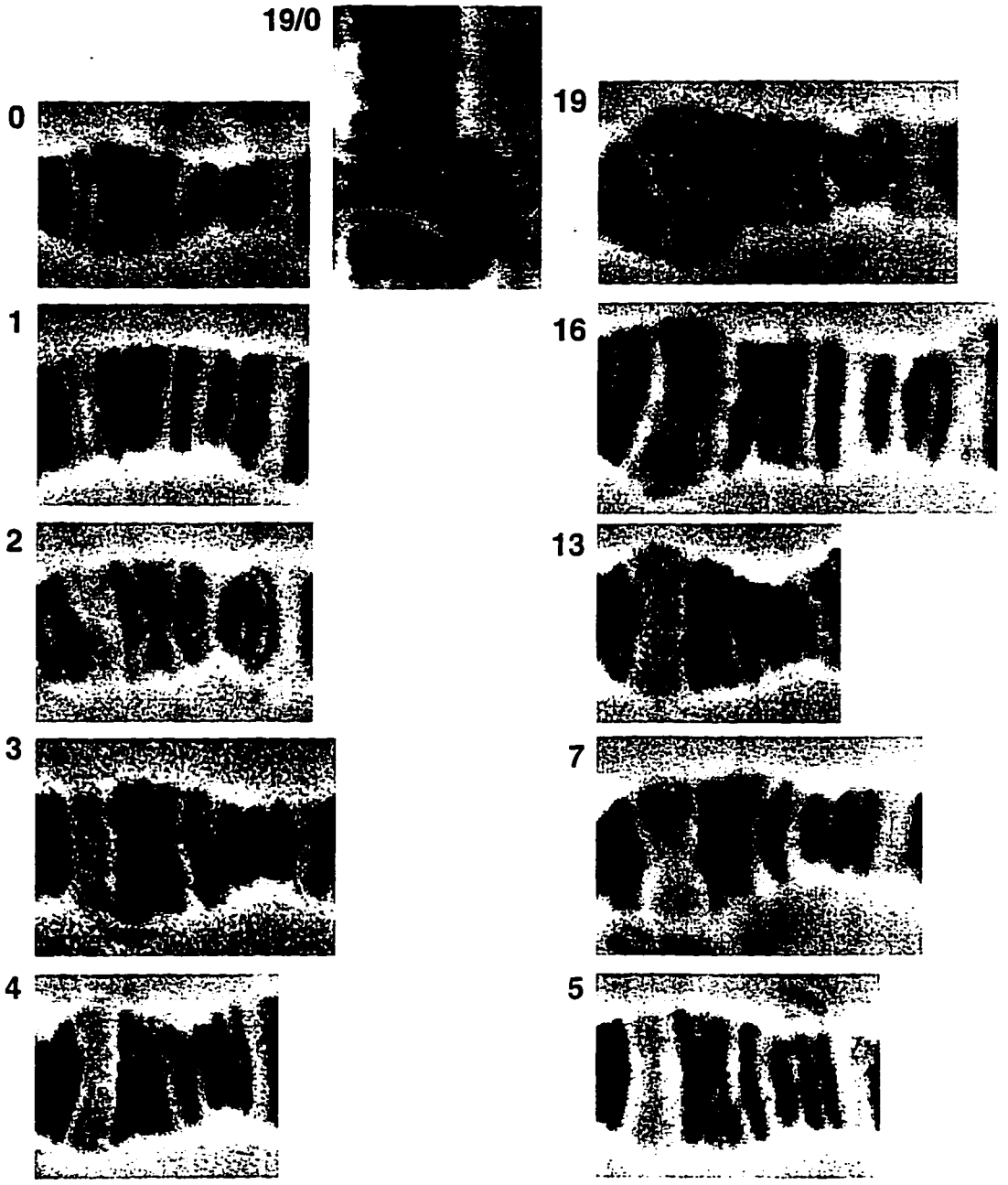

A
13d2
13d2B4
B2*1
B7*5
25d3
A
A1*3
A3*1
E6*1
AvrII
13d2
25d3
A
A1*3
A3*1
A4*1
A5*1
B2*1
bw D
400
310
150
1237464"big"35
P[PRAT-bw+]
4413161913177
B
3
2
1

Figure V-II cont. (part C)


19/0
0
19
1
16
2
13
3
7
4
5

VI

Conclusions

In the above chapters, I have shown that changing the local repetitiveness near a variegating breakpoint changes the level of variegation, and suggested that this occurs through *cis*-acting homologous association of sequences. Next, I have argued that, in the presence of a chromosomal discontinuity (such as a variegating breakpoint), the direct (ectopic) pairing of transgene sequences with heterochromatin is a strong determinant of PEV in *cis* to the breakpoint. Furthermore, I have extended this observation to the *trans* copy of a gene: the phenotype of dominant PEV is more strongly controlled by the relative proximity and orientation of the *trans*-inactivated gene to heterochromatin than by the ability of local sequences to pair homologously. However, the "local" heterochromatin still controls the character of the variegation in *trans*-inactivated genes, suggesting that pairing forces do more than simply move the homologous region into the general vicinity of the chromocenter. In addition, the heterochromatin distance effect (which can also be seen, in cytological terms, as enhancement of the ectopic association of heterochromatic blocks) is shown to effectively enhance euchromatic PEV in *cis* to the heterochromatic block. However, the correlation between phenotype and linkage distance (intrachromosomal distance) is weak; instead, physical interchromosomal distances within the interphase nucleus may control variegation levels directly.

Taken together, my results suggest that homologous pairing can be loose, or permissive enough to allow complex and diverse intrachromosomal associations to occur. At the same time, a large body of evidence (e.g. transvection) suggests that homologous pairing in *Drosophila* is strong

enough to suppress many ectopic interactions between (and even within) chromosomes. This view is consistent with a mode of genome organization that is both compartmentalized and flexible.

In contrast to the above results, I have also demonstrated that a high level of repetitiveness does not guarantee autonomous variegation of a transgene. We cannot yet say whether this is due to factors at the insertion site, to a lack of pairing factors within the transgene, to physical constraints on the topology of the repeat ends, or to some unusual robustness of the promoter involved. On a technical note, this work has demonstrated that manipulation of local repetitiveness is possible both in the presence and the absence of a variegation-inducing breakpoint. It should, in theory, be possible to distinguish in the future between some of the above possibilities.

Drawing on my results; on the looping model of Wakimoto and Hearn (1990) as adapted by Talbert *et al.* (1994); and on the pairing models of Gubb *et al.* (1990) and Dorer and Henikoff (1994), I have proposed a new model for the propagation of PEV. The model proposes that inactivation of euchromatic genes may result from the formation of multiple small loops that preferentially re-localize the euchromatin near a heterochromatic breakpoint into a heterochromatic compartment. Looping might be initiated or stabilized by contact between repeated sequences interspersed through the genome. A subset of these sequences may preferentially associate with heterochromatin due to structural properties common to tandemly repeated sequences (Henikoff 1996, see also chapter II and Appendix B). This model can account for the commonly-observed polarity and the occasionally-observed discontinuity that are characteristic of PEV;

for the compaction and loss of chromatin banding that are cytological earmarks of PEV in polytene nuclei; and for the existence of dominant PEV.

Perhaps provocatively, this model implies that heterochromatization of euchromatic sequences is directly dependent on spatial proximity to heterochromatin, and only indirectly dependent on genetic linkage (which nevertheless performs a strong tethering function). In other words, dominant PEV might be the most basic form of variegation, rather than an elaboration of the cis-limited form of PEV seen for most alleles. This contention is supported by recent evidence that ectopic copies of the *white*$^{+}$ gene, stripped of their flanking DNA, can exhibit dominant PEV, even though the endogenous *white* gene has never been observed to do so.

The looping model for PEV may even allow us to explain the unexpectedly unstable developmental behavior of some clonally variegated alleles (as described in chapter I). If inactivation relies on the association of non-homologous sequences, contacts made early in development may help to organize the nucleus, so that the same contacts will be made preferentially following each cell division. Nevertheless, stable inactivation might only be realized in nuclei when the duration of interphase has reached a certain level, allowing a full set of contacts to occur.

Looking beyond *Drosophila*, the looping model for PEV links two fields of inquiry--long-distance gene regulation and chromosomal pairing -- that might benefit from sharing some of their mechanistic models (appendix B). Finally, this model also supports the suggestion of Dorer and Henikoff

(1994) that repeat-based induction of PEV in *Drosophila* may serve as a model system for Repeat-Induced Gene Silencing (RIGS; (Assaad, Tucker and Signer 1993) in plants and animals.

Appendix A: a Partial History of Somatic Pairing Phenomena

The word "homology" has carried several different meanings -- often concurrently -- over the last century. When chromosomes were assumed to be either homogeneous or entirely heterogeneous along their length, it was easy to conclude that perfect homologous pairing of chromosomes could result from self-association of like parts. That is, local homology could be inferred on the basis of global homology. Over time, awareness of cytologically-visible "non-homologous" associations within and between chromosomes increased. This realization, when augmented by genetic evidence for "pairing" of like genes in tandemly-duplicated chromosomal regions and bolstered by biochemical evidence for high levels of genomic repetitiveness, led to the idea that similar but non-allelic sequences tend to associate physically; similarity of sequences was also called "homology." As most genomes contain dispersed sequences that share "homology," the simple link between local paring of similar chromosomal regions and global pairing of similar chromosomes became more complex. How were local pairing forces integrated to produce pairing of homologous chromosomes?

In the specific case of meiotic pairing, cytological evidence has long suggested that relatively large, dissimilar regions can be forced to align (following the proper pairing of flanking sequences) through the intervention of the proteinaceous structure known as the synaptonemal complex. Perhaps, more generally, all pairing might be a partially cooperative, dynamic processes that encouraged the continued association

of longer regions of homology relative to short ones. But how, then, could detectable ectopic contacts occur? In a dynamic nucleus, linked sequences might contact each other sooner or more often than allelic (unlinked) sequences. Under some conditions, the mere proximity of sequences might encourage their association, without regard to sequence similarity. Thus, the organization of contacts between and within the chromosomes of the somatic, diploid, interphase nucleus might represent the sum of many pairing forces. The study of homologous associations, then, is the study of these forces, both individually and in combination.

Before the genetic code: early cytological concepts of homology

The term "homology" has been used to describe everything from the similarity of chromosomes to the percent-identity of DNA sequences in a database. We make many assumptions in using the term so broadly; what follows is an attempt to pin down some of these assumptions by looking at the early history of pairing phenomena.

Chromosomal homology: Use of the term "homology" was originally motivated by the then-revolutionary idea that individual chromosomes had different identities, and that two chromosomes could share an identity. The recognition of homology in mitotic and meiotic chromosomes slightly predated -- and strongly supported -- the idea that chromosomes carry the hereditary material of the cell. Observing chromatids of the salamander in mitotic prophase, Walter Flemming wrote

> "...their longitudinal splitting is a typical occurrence... beyond all doubt...The halves of the threads lie almost precisely parallel in epithelial nuclei and red blood cell nuclei; in endothelial cells and often connective tissue

> cells they lie slightly diverging...Later the threads push apart from each other along their whole length and thus a fine rayed star arises...one longitudinal half of each thread could move into one half of the nuclear figure, the other longitudinal half of thread into the other half of the nuclear figure..." (Flemming, 1879, trans.(Voeller 1968).

Because the "threads" in question were derived from the apparent "splitting" of a single thread, the homology of the "half threads" was evident. However, Flemming clearly distinguishes the half-threads as independent units: he notes that there are differences in the degree of alignment in different cell types, and writes of threads "push[ing] apart," rather than threads "being released," implying that two independent units are being formed, rather than one unit disintegrating.

In 1883, Wilhelm Roux argued that separation of arrayed homologous units is the only efficient way to partition a large number of distinctive "nuclear qualities" whose presence and ratio must be maintained; this lies in contrast to a "free mixing" method of nuclear partitioning, which becomes very inefficient as the number of "qualities" rises and the quantity of each "quality" diminishes. (Roux, 1883, trans. (Voeller 1968). Weismann, in 1887, extended this argument, writing

> "The vast number together with the minute quantity of the ancestral germ-plasms permit us to conclude that they are, upon the whole, arranged in a linear manner in the thin thread-like loops: in fact, the longitudinal splitting of these loops appears to me to be almost a proof of the existence of such an arrangement...(Weismann 1887)."

As part of what we now see as a mis-analysis of the meiotic process,

Weissman reinterpreted earlier work on oocyte formation in ascarids. Because he did not accept that actual loss of information occurs ("the first of these two [meiotic] divisions would have to be looked upon as a reduction...but this assumption is impossible; although on the other hand it cannot be directly disproved..."), Weissman argued that the number of "loops" had been doubled previous to meiosis; the loops which find each other and pair in the first division cycle are actually identical. Implicit in this argument is the assumption that identical determinants retain the property of self-association when separated.

With the recognition by Montgomery (1901) and Sutton (1903), that the chromosome "pairs" in the first meiotic division were made up of one maternally-derived and one paternally derived chromosome (rather than duplicates of the same chromosome), a shift occurred in the meaning of "homology." Previously, it had indicated *a priori* sameness-by-descent. Now, the "similarity" of the pairing partners was assumed on the basis of their morphological similarity (Sutton 1903; Baumgartner 1904), and by the semi-tautological argument that they were "homologously" paired: because identical chromosomes pair, all paired chromosomes were assumed to be identical in their pairing properties, though dissimilar in their coding properties; for, as Sutton (1903) noted, "it seems highly probable that homologous chromatin-entities are not usually of strictly uniform constitution, but present minor variations corresponding to the various expressions of the characters they represent." Interestingly, morphological dimorphism was occasionally observed between homologues, and was eventually used to demonstrate the independent assortment of chromosomes (Carothers 1913); the acceptance of these clearly dissimilar chromosomes as homologues stresses that pairing, not

morphology or associated "expression of characters," was accepted as the touchstone for chromosomal identity.

Ectopic "homology": The recognition of homology within chromosomes, or between "non-homologous" chromosomes, has a related but separable intellectual history. Most obvious were those associations that involved distinctive cytological co-segregation of "non-homologous" chromosomal regions, especially associations of many (or all) of the chromosomes. Promiscuous association of telomeres (with each other and with the nuclear periphery) was recognized (for *Drosophila*, (Hinton and Atwood 1941). Because of their distinctive location (on the chromosome and in the nucleus), telomeres could be expected to share some structure or function; therefore they might be "similar" in the same way that homologous chromosomes were "similar." Even more obvious was the functional significance of the primary constrictions, or centromeric regions in mitosis and meiosis; these also may co-localize within nuclei to form chromocenters (Heitz 1929, and as described above). If regions known to share functional properties could co-localize, it was easy to infer that co-localization of other regions also implied shared function, or even shared identity.

The rise of cytogenetics: connecting homology and allelism through recombination

Genetic identity and chromosomal homology: The rise of genetics as a discipline in the 'teen's and '20's made possible the recognition that groups of genes had a set, linear relationship to each other (Morgan 1911; Sturtevant 1913); furthermore, the genetic maps produced to describe these

relationships had a one-on-one correlation with the cytologically identifiable chromosomes (Bridges 1916). Genetic maps were produced using meiotic recombination frequencies. Of course, the idea that different alleles are essentially interchangeable is intrinsic to the concept of a standard map. As pointed out by Kohler (1994), however, the genetic map and the lab *Drosophila* were both modified to fit a conceptual framework:

> "[Measuring genetic distance] assumes that the rate of crossing over depends only on distance, when in fact it is quite variable. Indeed, almost none of the early *Drosophila* stocks acted in this ideal fashion(p. 65)...genes that modified or suppressed crossing-over in nearby loci...could be eliminated only by rebuilding the chromosomes. That was done by selecting a dozen or so stocks in which the rate of crossing over was least variable and selectively breeding from them a standard stock in which the rate of crossing over was uniform...Wild *Drosophila* were thus physically reconstructed to conform with the fundamental principles of genetic mapping: namely, that 'distance' is proportional to the rate of crossing over and that the rate of crossing over is constant from one region of a chromosome to another (p. 78)."

Specific gross chromosomal heterogeneities were crucial in connecting the previously distinct fields of genetics and cytology. With the recognition of translocations in *Drosophila* (Bridges, 1923, cited in Dobzhansky, 1929), and maize (McClintock 1930), the co-linearity of the genetic map and the physical (mitotic) chromosomes could be established (Dobzhansky 1929; Creighton and McClintock 1931). the cytological evidence of region-specific chromosomal morphology and the genetic evidence of predictable recombination frequencies supported the idea of gross and fine scale similarities between homologous chromosomes and chromosomal

regions. Interestingly, the recognition of chromosomal aberrations -- that is, of heterogeneity between "homologous" chromosomes -- did not challenge the idea of "homology" *per se*. So long as genes maintained a unitary identity, homologous chromosomes were also essentially identical, and homologous pairing of chromosomes could still be derived simply from the local association of similar sequences (Van Atta 1932).

(Van Atta 1932) used metaphase cytology to demonstrate that a displaced chromosomal region would either pair strongly or not pair at all with its homologous partner. The pairing decision could be explained by the physical constraints imposed by pairing of flanking sequences, given the following assumptions:

> "Since the oogonial metaphase plates of normal *Drosophila* females almost always show close pairing of the homologous chromosomes, it has been assumed that the corresponding parts (perhaps the genes) have some property causing them to exert an attractive force on each other...[maintenance of pairing in translocated chromosomes] can be explained only by assuming attraction of homologous parts of chromosomes regardless of their relative positions in the chromosomal configurations (Van Atta 1932)."

The adoption of polytene cytology by *Drosophila* geneticists (Painter 1934b; Painter 1934a) provided a more detailed look at the ability of translocated and inverted material to remain synapsed in salivary gland somatic cells, more strongly supporting a similar conclusion:

> "if the attraction is chemical in nature...it is particulate and not a general mass reaction, for not only do like chromosomes lie together, but like parts are usually

> apposed...[even when] the form of the entire chromosome must be distorted in order to bring this about (Painter 1934b)."

By suggesting that the pairing of chromosomes is a global manifestation of nothing more than the local pairing of attractive factors, this idea hearkens back to Roux and Weismann. However, this thesis is partially undercut by another of Painter's observations (Painter, 1934a):

> "In triploid larvae, the three homologues unite in the same intimate way as in diploid cells...If one of the chromosomes carries an inversion, it has been observed that all three elements unite, beginning at the point of spindle fiber attachment and continuing up to the point of inversion, and from this point the aberrant element remains single while the other two homologues are paired in the usual way [Painter 1934b]."

Triploid pairing is in accordance with the observations of (Metz 1925), who described somatic pairing in the prophase of triploid females; competition is consistent with the observations of Newton and Darlington {1929, quoted in (Ashburner 1989)} that in triploid meiosis, partners pair "two-by-two," but that a homologue may switch pairing partners along its length.

In other words, "pairing," in a global sense, can involve more than two partners (as suggested by earlier observations of telomere aggregation and chromocenter formation). However, the pairing potential of chromosomes is not unlimited. When a rearrangement is encountered, tripartite "pairing" is reduced, and partner-choice preference is seen. Why? Any two of the following assumptions can explain the result:

1. The pairing process shows some level of cooperativity (as well as autonomy) in *cis* (i.e. along the chromosome);

2. paired chromosomes stabilize each other, reducing their ability to distort so as to match a third, inverted homologue;

3. pairing at any site occurs between only two homologues, but individual sites can interact with different homologues;

4. only a subset of sites can mediate pairing.

Or, more succinctly, either the divisibility of the chromosome into pairing units or the autonomy of such units is not infinite. The trade-offs seen in cases of tripartite pairing provided the clearest evidence to date that chromosomal pairing was not merely a simple extension of some ideal property of the minimal genetic unit. Rather, pairing was a relative property of similar chromosomal regions, and was subject to competition, to interchromosomal modification and to experimental manipulation. "The primary assumption of attraction of homologous parts" (Van Atta, 1932) would need some refinements or qualifications.

PEV and the idea of the genetic unit: Most of the chromosomes examined as metaphase plates by Van Atta (*op. cit.*) were recovered because their aberrations produced dominant, mottled eye color phenotypes. First recognized by Muller (1930), these "eversporting displacements" (dominant or recessive) shared features that set them apart from other genetic aberrations, and made them a topic of keen interest. First, they were "ever-sporting"-- that is, a single allele located on a variegating chromosome produced two phenotypes within a single tissue. This implied instability of a chromosome (Van Atta 1932); Painter, 1934; allelic instability (Gowen and Gay 1934; Demerec 1940); or instability of the gene's activity (allelic state), with no corresponding change in the allelic compo-

sition (Raffel and Muller 1940; Demerec 1940). Furthermore, more than one gene could be affected at a time.

The key to approaching variegation (and all that it implied) was the equally perplexing role of the "displacement." Variegating alleles were generally correlated with large chromosomal aberrations. Why was this so? Van Atta (1932) argued that translocations and inversions caused complex pairing, and that this pairing in turn caused mitotic stresses that led to chromosome breakage and somatic loss. Many lines of evidence argued against such chromosome loss (and against individual gene disruptions). Gowen and Gay (1934) pointed out that a cell-autonomous lethal gene lying between two affected genes was not, itself, affected by variegation, demonstrating that the phenotype was not caused by loss of the whole region (though individual lesions might be involved). Schultz (1934) noted that the existence of dominant variegation argued against the gene-loss hypothesis. In addition, only certain combinations of breaks produced variegating effects (Raffel and Muller 1940); such specificity was inconsistent with the idea of chromosome breakage and loss. Finally, (Demerec 1940) showed that insertional translocations into heterochromatin could produce variegation for several genes at both ends of an insert, while genes in the middle of the insert were unaffected suggesting that loss of the insert was not occurring.

As a group, these results strongly implied that gene expression was modifiable, not absolute; that a gene's neighbors (or neighborhood) could serve as modifiers of expression; and that the regulatory region of a gene (in the widest sense) could encompass and affect several other genes. In 1903, Sutton (*op. cit.*) had suggested that alleles of a gene must differ, but in

ways that only modify their phenotype, without affecting their essential identity or their pairing properties. By challenging the idea of the gene as a consistent and indivisible unit, the work of the 1940's collapsed this artificial dichotomy. Gene boundaries, gene identities and gene pairing properties might all be essentially mutable and relative:

> "In genetic theory, genes have been considered as (1) crossover units -- hypothetical segments over which crossing over does not occur; (2) breakage units -- again, hypothetical segments within which chromosome breakage and reattachment do not occur (at any rate, not without destruction of one or both fragments); (3) mutational and functional units -- those minute regions of chromosomes, changes within one part of which may be so connected with changes in the functioning of the rest of that region as to give rise to the phenomenon of (multiple) allelism; or (4) reproductive units -- the smallest blocks into which, theoretically, the gene-string could be divided without loss of the power of self-reproduction of any part. A category of auto-attractive units might also be added. Although it seems often to have been assumed, there is as yet no empiric evidence, and only doubtful theoretical ground, for assuming that the lines of demarcation between genes, as defined on any one of these systems, would coincide with those on any of the others, or even...that such lines are necessarily invariable, non-overlapping, well-defined and absolute (Raffel and Muller 1940)."

This willingness to consider the gene as an ill-defined entity was, by Raffel and Muller' s own admission, new: "[in 1935]...we did not think of genes as being further genetically divisible...if smaller, linearly arranged parts of genes...may become rearranged with respect to each other...sufficient plasticity of result is introduced to make plausible the conception that most gene mutations might be of this kind."

And what is the nature of the "neighbor" that affects expression in heterochromatin-induced PEV? In one model, Raffel and Muller suggest that heterochromatin serves to hide strong but redundant genes whose activity can affect neighboring genes:

> "...not all genes are indispensable for life...[or] cause readily visible abnormalities. And yet, such 'invisible genes' might, theoretically, be as active as any others in exerting a 'position effect' upon neighboring genes...the so-called 'inert' or chromocentral region of the proximal end of the X chromosome, as before noted, is known to be homologous, at least in part, with the Y chromosome, and probably with the chromocentral regions of other chromosomes as well. Hence deficiencies of genes in this region would more often be 'covered' by the presence of normal alleles of them in the other chromocentral regions (Raffel and Muller 1940)."

As Ephrussi and Sutton (1944) put it, this line of inquiry directly questions "the degree of discreetness of the hereditary material." In more current terms, this model suggests that PEV is caused by regulatory fusion; the role of heterochromatin is to provide an unusual level of genetic redundancy, so that fusions will be non-lethal, and thus can be recovered.

Muller's willingness to challenge the orthodox, genetically-based model of chromosome structure stands in contrast to the equally ambitious but slightly more orthodox ideas of Calvin Bridges. Bridges was the most determined interpreter of the new science of polytene cytology. He believed that chromosome structure would bear the traces of evolutionary events that had been responsible for increasing complexity in the genome. Evolutionary arguments then (as now) often invoked the concept of gene duplication followed by gene divergence. This in turn implied that

chromosomes might be riddled with large and small regions of similarity. Additional support for this idea came from Bridges' discovery that the unstable *Bar* phenotype and its more extreme *double- Bar* derivative were caused by cytologically-visible gene duplications (Bridges 1936).

Each of Bridges' successive polytene chromosome maps was accompanied by an analysis of probable duplications within the genome (Bridges 1935; 1938; 1942). He diagnosed repeats "of all lengths, from doublets to whole sections, both direct and symmetrically reversed in tandem" by "general appearance, the folding back into synapsis of reversed repeats, and the synapsis of bands from one chromosome with those of other sections of the same or different chromosomes (Bridges 1942)." In euchromatin, synapsis was monitored by the "weak and occasional" (but band-specific) presence of ectopic fibers or by band alignment; this, Bridges termed "real" synapsis. By comparison, Bridges was less taken with the co-aggregation of condensed chromatin, which he termed "apparent synapsis": "this latter type is actually only the general manifestation of the non-specific attraction of bits of heterochromatin for each other...Wherever such synapsis occurs, the dark bands involve synapsis only as region to region, rather than as band to band. [Note: Centromere-associated hetero-chromatin does not form clear bands, but identity and orientation of heterochromatic regions can nevertheless be inferred from connections to euchromatin. Heterochromatic regions appear to associate in a variety of orientations, unlike euchromatic bands.]"

Fueling Bridges determination to see gene duplications, ever-finer mapping studies were suggesting that some genes had phenotypically-similar, partially redundant copies located extremely nearby. Once the

divisibility of the gene was recognized, these would be reinterpreted as "cistrons" of the gene, and later yet, as "complementing alleles." For the next two or three decades, however, the idea that many genes had near (and distant) neighbors with whom they shared a common line of descent and whom they could contact encouraged many geneticists to consider the role played by genetic neighbors in determining the action of genes (e.g. Lewis 1950; Green 1963).

Bridges analysis also helped to change the common understanding of the pairing of "like" sequences. By invoking the pairing of diverged genes to explain ectopic connections, and by treating heterochromatin as a co-aggregation of non-identical regions, Bridges assumes that it is possible for non-identical entities to associate, just as if they were identical.
What does this mean in terms of homology and pairing? First, the idea of regions possessing "adequate" -- as opposed to "absolute" -- similarity is essentially a foreshadowing of our concept of sequence divergence, and of per-cent sequence homology; it would not be dealt with explicitly until the 60's and 70's, and even then, only rarely, e.g.:

> "It is usually inferred that the physical basis of homology is the same as that which underlies genetic specificity...but ...direct evidence on this point is lacking. The assumption would seem, indeed, to lead to a paradox, since if both homology and genetic specificity reside in DNA, and mutations involve alterations in DNA, then increased heterozygosity for "point mutations" should be accompanied by decreasing chromosomal homology (Menzel 1964, discussing the preferential pairing of homologues over homeologues)."

In appendix B, I will discuss questions related to sequence inhomogeneity, such as "how much sequence divergence is required to make a sequence

'non-homologous' in terms of biological sensing mechanisms?" For now, I only want to point out that pairing of large regions of chromosomal arms can occur even in hybrids of different drosophilid species, where the average divergence of coding -- and especially non-coding -- regions is presumably comparable to the levels of divergence between repeated regions within a species.

Bridges' belief in the pairing of non-identical genes had another, even more general, effect on the concept of chromosomal homology. If "homology" between regions can be strong or weak, and if many regions can share homology, then "homologous pairing" becomes an explicitly conditional term: "homologous sequences" are those that can pair under a given set of circumstances. It follows that if pairing is a) relative and b) modifiable, then it is also potentially a) regulatable and b) manipulable. Two questions suggest themselves. First, is there any evidence that pairing can affect gene regulation? Next, how strong are the relative pairing forces that might contribute to such regulation?

The importance of pairing -- pairing dependent genetic phenomena

In the past 40 years, several genetic interactions have been described that depend largely or entirely on the proper pairing of chromosomes or sequences for their effect. The best-described of these are the transvection effects seen for certain mutant alleles of the Bithorax and Decapentaplegic gene complexes, and the *zeste* -mediated silencing of paired target sequences.

Transvection (*Ubx/bx* and *dpp*): Lewis (1954) coined the term "transvection" for his observation that certain alleles of genes in the *Ultrabithorax* (*Ubx*) complex were sensitive, as heterozygotes, to a wide range of genomic rearrangements. Specifically, breaks proximal to one allele (e.g., Ubx^{1}) which placed the gene in an unusually distal location, caused a mutant phenotype in the otherwise-complementing "*trans*-heterozygous" combination Ubx^{1}/bx^{34e}. The same effect could be seen following rearrangement of only the bx^{34e}-bearing chromosome. In combination, however, similar rearrangements again allowed the *bx34e and Ubx1 alleles* to complement each other. Though the pairing requirements for this effect are actually quite lax, transvection at *Ubx*, and later, at *decapentaplegic* (*dpp*) and other loci (Gelbart 1982; Benson and Pirrotta 1988), have been taken as the premier example for the regulation of gene function by chromosomally-mediated pairing effects (see also Chapter I).

Pairing requirements of the *zeste-white* interaction: Originally described by Gans (1953), the *zeste-white* interaction is now understood to be the interaction of closely apposed white sequences by multimerization of $zeste^{1}$, a complex-forming neomorph of the $zeste^{+}$ regulatory protein(Bickel and Pirrotta 1990). Gans noticed the effect as a yellow eye color affecting females (rather than males), and demonstrated pairing dependence by showing that females hemizygous for $white^{+}$ were unaffected, while males duplicated for $white^{+}$ (in *cis*) were affected. Chromosomal rearrangements near the $white^{+}$ region (or removal of one $white^{+}$ allele to an ectopic location) were shown to prevent $white^{+}$ from participating in $zeste^{1}$ -induced silencing; in contrast, tandem duplications $white^{+}$ at a variety of chromosomal locations were sensitive to $zeste^{1}$ (Green 1967; Jack and Judd 1979). Thus, the phenotype was strictly dependent on pairing -- much

more so than in the case of transvection at *Ubx* (as noted by Ashburner 1989) -- with the interesting proviso that tandem duplications and allelic copies of the *white*$^{+}$ gene had esentially indistinguishable responses to *zeste*^{1}.

Other, phenotypically similar *trans*-sensitive effects have also been described. Several are now known to occur due o the increased proximity of regulatory proteins that bind within the duplicated regions. Pairing-sensitive effects at the *engrailed* gene, for example, fall into this class of duplication-based phenomena (see Kassis 1994).

Pairing as a regulatory component, or force: The idea that pairing could influence genetic phenotypes was proposed in 1944 by Ephrussi and Sutton, who invoked both classical, allelic pairing and intrachromosomal pairing of repeated regions to account for disparate phenomena associated with PEV. As mentioned in chapter I, the phenomenon of dominant PEV spawned many models. Among these was a model based entirely on the forces of homologous attraction (Ephrussi and Sutton 1944). In fact, this was not so much a single model as two models. The first of these suggested that disruption caused by the pairing of rearranged chromosomes caused local regions of chromosomal stress; these could then be propagated along the chromosomes, affecting both homologues equally. A prediction from this model, namely, that disruptions of a homologous chromosome should generally affect the phenotype of dominant, variegating genes was soon disproven (Slatis 1955a; Slatis 1955b). This prediction, however, has been partially validated for a set of dominant, variegating *brown* alleles (Henikoff, Jackson and Talbert 1995, and Figure I-4d). Nevertheless, the basis for the effect is probably not pairing-induced

stress, *per se*; otherwise, as pointed out by Lewis (1950) homozygosis of variegating rearrangements should suppress variegation, because they restore pairing and reduce stress; instead, homozygosis of a variegating gene generally strengthens inactivation (Demerec and Slizynska 1937; Schultz 1941b).

As an extension or variant of their first model, Ephrussi and Sutton proposed a second model of more-lasting interest. This model proposes that stress of a different kind (caused, presumably, by differences in packaging) allows for efficient intrachromosomal contact and pairing of similar regions within heterochromatin and within the euchromatin near the affected alleles, leading to variegation. This model is attractive, but was untestable at the time it was proposed, because there was no way to manipulate the local repetitiveness of the sequences involved, and no way to gauge the relative strength of intrachromosomal pairing.

The *zeste-white* interaction, described above, has helped to remedy the second problem, by demonstrating how the strength of local pairing can affect gene expression. Because *cis*-pairing of a tandemly-duplicated *white* allele and normal (*trans*)pairing of *white* alleles were equally effective in promoting the phenotype, *zeste*^{1} interactions provided a very sensitive test for the relative strength of ectopic and homologous pairing in a region. Gubb *et al.*(1990) examined this effect using two examples from a series of Transposable Element (TE) duplications. These duplications arose from a spontaneous inclusion of several polytene chromosome bands from the *white* region of the X between naturally- occurring transposable ends (Ising and Block 1981; Ising and Block 1984). Spontaneous mobilization of this region brought it (or parts of it) to a few autosomal locations, and produced

both tandem and inverted duplications (Ashburner 1989, p. 643-649).

Using a tandem duplication in the 35B region, Gubb *et al.* irradiated *white*$^-$, *zeste*^{1} flies to produce chromosomal breaks, and crossed them to marked *white*$^-$, *zeste*^{1}, TE-duplication females, and scored for *zeste*- and *zeste*^{variegated} flies. As expected, many breaks allelic to the duplication region partially or entirely suppressed the duplication-based phenotype by "ironing out" the (cytologically-visible) pairing of the (TE) *white*$^+$ duplications (Figure A-1). Interestingly, pairing of direct duplications was strongly suppressed, but an inverted duplication at the same site was unaffected; that is, its pairing was resistant to "stress" induced by chromosomal pairing forces. How does allelic pairing compare to the pairing of direct or inverted regions of a similar size? To compare the forces involved, it helps to note that the pairing dependence of the *zeste-white* intereaction was first recognized because local allelic pairing was entirely ineffective in promoting pairing when *white*-containing regions were displaced to other chromosomal locations. Thus, we can provisionally rank the relative efficacy of pairing interactions as follows: allelic -- weakest; direct tandem -- intermediate; inverted tandem -- strongest. (Presumably, this reflects a difference in the tethering, or overall physical proximity of the sequences involved.) On this basis, it is reasonable to suggest that the local, physical pairing of like sequences could be a stable component in the structuring of intrachromosomal domains.

Pairing-dependent genetic interactions might seem to be obscure or unnatural forms of genetic regulation. For example, both transvection and the *zeste-white* interaction generally require primary mutations to "sensitize" the system; otherwise, pairing effects are not seen.

Furthermore, as detailed above, different pairing-dependent interactions show different degrees of pairing sensitivity, or have different pairing requirements. Nevertheless, each pairing-dependent phenomenon demonstrates that chromosomal regions can, in fact, pair strongly enough to allow for cross-regulation of similar sequences. Different phenomena, by having different requirements, provide a window through which we can peer at the different requirements for effective searching and productive contact of like sequences.

The biochemistry of repetitiveness in genomes: from Cot to CATG

Once we know that "like" regions can potentially interact, two questions remain. First, how similar (and how long) do regions have to be to interact? Next, how many regions of similarity exist in the genome? I shall address the second question here, and the first question in Appendix B.

Repetitiveness suspected: Homologous pairing of chromosomes is not only complicated by the inhomogeneity of allelic sequences, but also by the similarity of nonallelic sequences. Before sequence-based biochemical analysis of DNA began in the 60's, various regions of the genome were suspected of having unusual properties. Based not on their morphology, but rather on perceived functional similarities to centromeric heterochromatin -- involvement in non-homologous contacts especially with proximal heterochromatin or terminal adhesions and/or a high rate of induced breakage relative to (polytene) map length -- these regions were referred to as "intercalary heterochromatin" (Hannah 1951, reviewing the work of Kaufmann, Slyzinski and Hinton). Because of their high rate of ectopic associations, these regions were suspected to contain duplicated segments of the genome.

Hannah (1951) pointed out that the some explanation needed to be made of the functional link between repetitiveness, a tendency towards ectopic associations, and the other properties of this so-called intercalary heterochromatin:

> "There is no a priori reason why intercalary heterochromatin should be associated with repeats. If ...the heterochromatic property is not an expression of degeneration of euchromatin [to inertness], the relation between repeats and heterochromatin must have another explanation. If the heterochromatic property is an expression of the linear order of chromomeres of similar function, as suggested by Pontecorvo (1944) then a heterochromatic segment should arise every time a euchromatic region undergoes duplication and the replicas remain adjacent to each other. Or the heterochromatic regions may [predate]...the duplications ...[making them] less subject to rigid selection."

If repetitiveness could mediate local chromatin conformation and also mediate contacts between widely dispersed regions of the genome, did this imply that it could serve a master regulatory function throughout the nucleus? While questioning the statistical significance, Kaufmann (1963) points out that of 71 regions recognized by Becker to form salivary puffs, only 12 are more than 10 bands away from ectopically-paired regions, and suggests that "puffing depends on the coordinated action of an operator-receptor system, the intercalary heterochromatin serving as a controlling element in the system."

Repetitiveness demonstrated: In the last two decades, proof for the existence of distinctive heterochromatin-like regions in euchromatin has been ephemeral, and the term "intercalary heterochromatin" has largely fallen into disuse, and with it, the idea that ectopic contacts are regulators

of gene expression. Repeated sequences interspersed throughout the genome may, however, play a role in gene expression very similar to that once proposed for intercalary heterochromatin. The relative repetitiveness of genomes was first assayed by ultracentrifugation (which isolated and defined the "satellite" sequences -- that is, DNA fragments representing specific, uniform and highly repetitive DNA sub-species). while the time-normalized DNA reassociation (C_0t) curve provided an overall view of sequence repetitiveness in the genome (Britten and Kohne 1968; Laird and McCarthy 1968). Because fast renaturation of repetitive DNA can anchor attached single-copy DNA in a renaturation complex, C_0t analysis of size-fractionated, sheared DNAs made it possible to estimate not only the relative representation of sequences, but also their relative length and interspersion patterns (reviewed by Nagl 1978). By various estimates, ~12% of the *melanogaster* genome is moderately repetitive. Other results suggest that perhaps 15Mb of middle-repetitive DNA is dispersed throughout the genome in an average pattern of 5.5 Kb (repetitive) | 13 Kb unique sequence (Manning, Schmid and Davidson 1975; Crain *et al.* 1976). Another 6% of the genome re-associates with zero-order kinetics, indicating that it is composed of inverted repeats (Schmid, Manning and Davidson 1975); if interspersed, these would have an average pattern of 1.35 Kb inverted | 38 Kb non-inverted sequence. (All reviewed in Ashburner 1989, p. 91-98).

Some of the dispersed, repetitive sequences were found to be highly transcribed, and therefore presumably in a (locally) open chromatin conformation; others, preferentially localized in compacted regions of the genome, were assumed to be silent. In terms of functional localization, Mayfield and Ellison (1975) demonstrated that segregation of different

satellite repeats was maintained even within the more-generally co-localized mass of the chromocenter. Repetitive sequences, though common and ubiquitous, were nevertheless non-uniform in their chromosomal and nuclear distribution. The tantalizing question remained: did these sequences play a role in gene regulation or chromosome organization?

Repetitiveness incorporated in theories: Drosophilists studying repetitive sequences tended to focus on their possible role in gene regulation or in the homologous association of polytene chromosomes (discussed below). In contrast, recognition of genome repetitiveness stimulated plant geneticists to reconsider the role of homology in meiotic and pre-meiotic pairing (driven, largely, by their appreciation of the paradox of homologous *vs.* non-homologous pairing (or homeologous pairing -- see above). Riley and Flavell (1977) list the qualities of chromosomes or sub-chromosomal regions "that may permit them to be described as 'homologous.'" These conditions are 1) evolutionary relatedness (presumably, and "usually"); 2) overall morphology (arm lengths, constrictions, distribution of heterochromatin, and banding patterns); and 3) gene colinearity. They also point out the limits of these conditions: gross morphology is commonly modified by structural rearrangement; gene identity is compromised by mutation and selection. Nevertheless, "lengths of very similar nucleotide sequences [within genes] would remain [and] these sequences would probably result in a similar array of histone and nonhistone protein being attached...[Giving] similar coiling or condensation patterns along their length [so that] chromosome homology...[would] crucially depend upon chromatin configuration." Non-homologous pairing, then, results from occasional contact between

repetitive sequences at ectopic locations or in homeologous chromosomes. Because it is based on sequence similarity, this pairing is "legitimate" in an immediate sense; nevertheless, it will generally be displaced by the preponderance of correct pairing between two homologues. (As pointed out by Kleckner (1992) in her review of McClintock's work (1933), the displacement may be due to additional forces produced by the synaptonemal complex which favor "2-by-2" pairing, rather than by direct competition of competing regions of homologous and homeologous pairing.)

While maize work was concentrated on the role of repetitive sequences in mediating interchromosomal contacts, *Drosophila* work was often concerned with the role of repetitive DNA in intrachromosomal contacts as well, especially in the light of its possible contribution to gene regulation. Bonner and Wu (1973), for example, suggested that repetitiveness could control packing, or transcription, or processing, or crossing over between pseudo-alleles. Mayfield and Ellison (1975), arguing from their observation of co-localization or related sequences, suggested that each band might have its own repetitive sequence family, and that a separate, repeat family-specific binding protein would mediate the packing of each band through small-scale ectopic interactions.

Repetitiveness supplanted: These theories of genome regulation suffered from extreme over-specificity (of numerical values, models, and thus predictions), and were not so much proven wrong as rendered obsolete. The cloning of genes and the resulting elucidation of promoters, enhancers, and splicing made other levels of control seem superfluous. As far as the link between centromeric heterochromatin and "intercalary

heterochromatin," the potentially simple correlation between specific sequences, local repetitiveness, chromosome fragility, and late replication turned out to be more complex and less universal than predicted (though, for example, sequences were found that hybridized not only to telomeres and centric heterochromatin, but also to the ectopic fibers that occasionally connected them; Ashburner 1989, p. 88). Finally, the sheer number and variety of repetitive sequences (simple and complex) did not suggest a simple or tractable unifying factor that could profitably be dissected; and the relative interest in repetitive sequences waned.

Repetitiveness redux: More recently, we have found a new appreciation for the intermingling of chromatin types in the genome: the simple-sequence rich centric heterochromatin has been found to be riddled with islands of unique and middle-repetitive sequences (Le, Duricka and Karpen 1995) while the patterns of specific middle-repetitive element families through the euchromatin (and heterochromatin) have been charted in detail. In addition, (Dorer and Henikoff 1994) demonstrated that repetition of an arbitrary euchromatic gene could produce a region with heterochromatic properties, as suggested 50 years previously (Pontecorvo 1944). Drawing on these results and on the observations in the previous chapters, I have suggested (chapter II) that middle-repetitive sequences could mediate the direct pairing interactions between juxtaposed heterochromatic and euchromatic regions in rearranged chromosomes. What does this imply for normal chromosomes? While it is true that PEV is an unnatural phenomenon caused by mis-regulation of genes, the very fact that such regulation can occur suggests that the "threat" of such mis-regulation may have helped to constrain the evolution of genomes, and may have subtly sculpted the characteristics of chromosomal regions.

In addition, the chromosome complement of wild Drosophilids, containing a high background of chromosomal and genetic diversity, might produce genomes with a high potential for ectopic contacts. As argued by Riley and Flavell in the case of maize (*op. cit.*.), such contacts might actually stabilize, rather than destabilize, chromosomal identity and localization due to the chromosome-specific and clustered nature of many ectopic sequences (Figure A-2). In light of this idea, perhaps some of the outmoded concepts of gene regulation through ectopic association of repeated sequences may bear reexamination.

Summary:

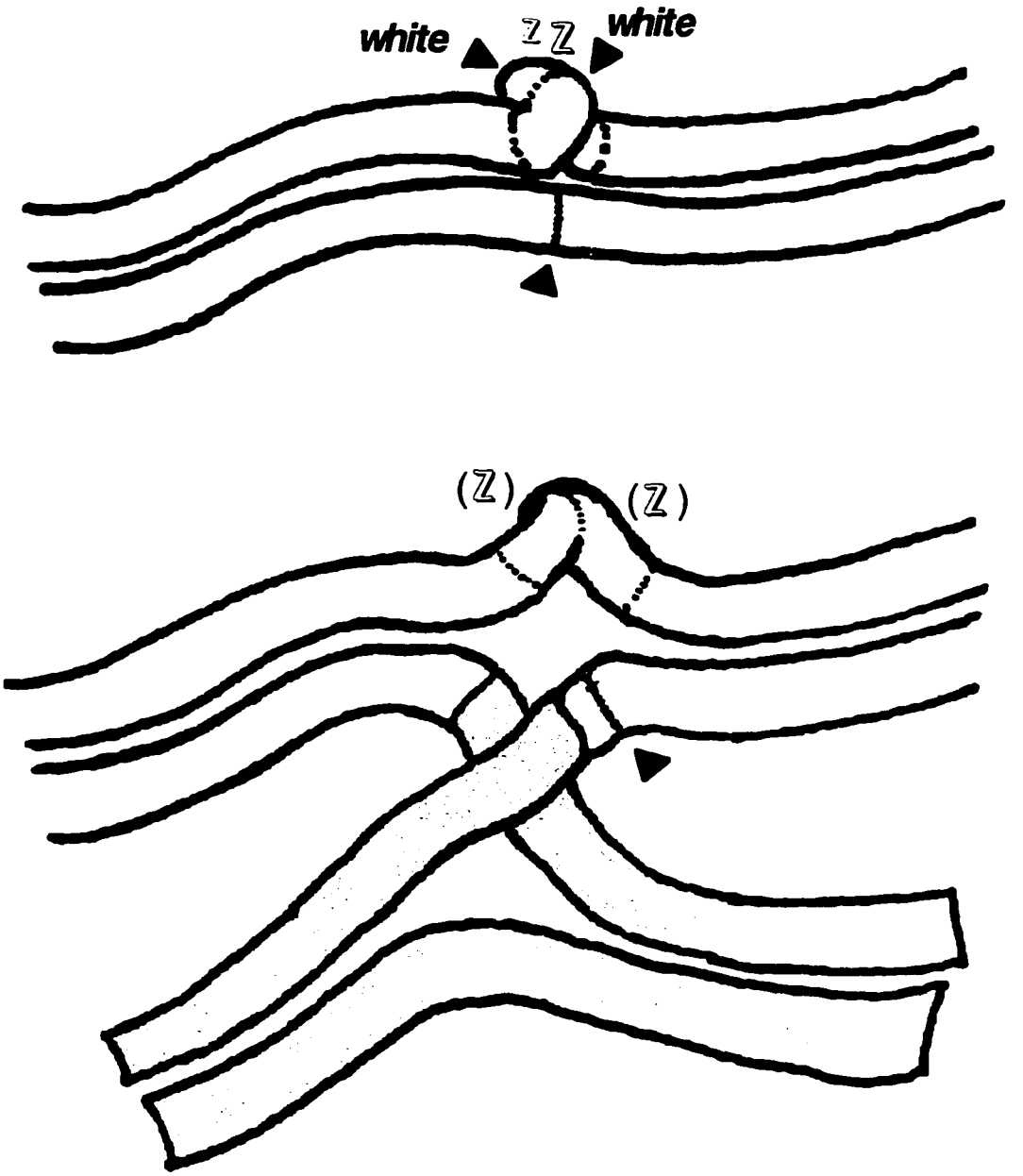
In this appendix chapter, I have traced the evolution of the idea of homologous pairing, and used this history to point out some of the problems that still face us in understanding the pairing of similar sequences, and the pairing of homologous chromosomes. Once, homology was considered to be an absolute property of chromosomes, created by the uniqueness of chromosomal constituents. Soon, it was shown that pairing was not an absolute property of chromosomes. Next, it was suggested, and later shown, that chromosomal constituents were neither unitary nor unique, suggesting that the pairing which we observe represents a balance between cooperative and competitive pairing forces within the nucleus.

When "pairing" was a property of whole chromosomes, its functional importance was of limited, primarily, to a structural role in chromosome mechanics. After all, if all regions pair equally and uniquely, then paring can have no individual regulatory effect. As we have come to recognize the relativistic nature of pairing, we have also begun to ask whether

differential pairing could also serve as a mechanism of intrachromosomal organization and genetic control.

To date, the mechanism(s) of homologous pairing remain mysterious. Manipulation of duplicated sequences and sequence arrays may allow us to look at the conditions for, and constraints on homologous pairing in a variety of organisms (appendix B).

Figure A-1. Cytogenetic and genetic pairing. The *zeste*^{1} neomorphic product inactivates *white*$^{+}$ genes when they are paired, either as alleles or as tandem duplications. Disruptions that interfere with pairing interfere with *zeste-white*$^{+}$ interactions as well. Top: the two halves of an insertional tandem duplications pair with each other; *white*$^{+}$ genes in the duplication pair and respond to *zeste*^{1} (Z). The point on the homologous chromosome allelic to the duplication insertion site is also marked. Bottom: chromosomal disruptions near the duplication insertion site "tug" on the duplication, interfering with its pairing and suppressing the genetic interaction.


white
ZZ
white
(Z)
(Z)

Appendix B: Somatic Pairing and Homology Sensing -- a Discussion of Mechanisms and Models

RIP and MIP; interactions of repeated sequences in fungi

Examination of repetitiveness in eukaryotic genomes has produced many descriptive studies which contain a wealth of sequence and positional information (for *D. melanogaster*, see Gatti and Pimpinelli, 1992; see also John and Miklos 1979). Repetitive sequences have not, however, been particularly amenable to direct experimentation. Speculation has attributed a wide variety of functions to repetitive sequences (Yunis and Yasmineh 1971), some of which have been refuted by looking at chromosomes with large heterochromatic deletions (John and Miklos 1979). In *Drosophila*, recent descriptive work has located specific gene families in chromosomal domains (Lohe, Hilliker and Roberts 1993; Carmena and Gonzalez 1995; Le, Duricka and Karpen 1995; Pimpinelli *et al.* 1995).

Recent experimental manipulation of repeated sequences (naturally-occurring or created) has included the work of Wakimoto and Hearn (1990); Eberl, Duyf and Hilliker (1993); Talbert, LeCiel and Henikoff (1994); Henikoff, Jackson and Talbert (1995); Howe *et al.* (1995); and most directly, Dorer and Henikoff (1994) and the above-described work. Still, our ability to precisely manipulate repetitiveness in the *Drosophila* genome is new. Also, our detection abilities are not (yet) well-defined. We hope that variegation can serve as a general detection system for changes in local repetitiveness. We do not, however, know which genes, transgenes or

alleles will, in fact, detect such changes. We do not know the range of their sensitivity, either in terms of distance, or in terms of the types of repeats that they will detect. In addition, *Drosophila* contains a considerable amount of naturally-occurring, interspersed repetitive DNA. These repeats may influence results in unpredictable ways.

There are other organisms that may be better-suited to the study of repeat-recognition systems. These organisms possess manipulable genomes; furthermore, the natural repetitiveness of these genomes is low, while the ability of the organism to recognize repetitiveness -- and respond to it -- is high. The two obvious candidate organisms are *Neurospora crassa* and *Ascobolus imersus*, two distantly-related haploid fungi. These fungi both exhibit inactivation of repeated genes by the pre-meiotic silencing phenomena of (respectively) RIP and MIP (described below). The phenomenology of the silencing differs, as do the chromosomal properties and life cycles of the organisms involved. However, each of these systems can tell us something about the ability of biological systems to sense the presence of homology. While the same can be said for silencing phenomena in plants, and for homologous recombination in a variety of organisms, RIP and MIP provide us with an unusually complete and compact body of data that can be applied to the following questions:

1. What is the minimum length of homology needed for recognition?

2. What is the minimum sequence identity (per-cent homology) needed for recognition? Does it vary?

3. Does chromosomal location (relative or absolute) affect gene silencing?

4. Is recognition a pair-wise process, or are multiple partners silenced together?

5. Does silencing depend on the duplication of regulatory regions?

At the same time, it should be pointed out that the silencing process (which may or may not be mediated by the same molecules as the homology-searching process) is almost certainly different among all of these systems (and *Drosophila*).

RIP in *Neurospora*: RIP, or "Repeat-Induced Point Mutation," is a premeiotic mutation mechanism affecting repeated sequences in *Neurospora crassa* (reviewed in Selker 1990). Because a single round of RIP commonly produces many changes in the DNA sequence, the products of RIP are easy to recognize and are rich in information. (RIP causes GC-->AT transversions with a strong dinucleotide bias, and thus even evolutionary RIP artifacts can be recognized.)

The smallest effective repeat length reported is probably less than 450 bp but greater than 380 bp for linked genes (Stadler, Macleod and Dillon 1991); for unlinked genes, a repeat of 800 bp allows efficient RIP in a variety of natural and introduced duplications. As a lower limit, the 5s rDNA genes survive at multiple sites, due to a combination of their small size (~150 bp) and their relatively high level of non-RIP-related divergence (see Selker 1990).

Naturally occurring repeats appear to have been RIPped to a divergence level of ~15% (with some dependence on linkage). For example, the 800 bp linked duplication known as the *zeta-eta* region has 267 GC-->AT

transversions (and one extraneous mutation) for a probable RIP-based divergence of less than 17%. (Though the RIP process acts randomly with respect to individual sites, some sites may be RIPped identically in the two halves of the duplication, purely by chance.) It can be shown that cessation of RIPping is due to loss of homology rather than to loss of RIP-able sites by re-duplicating a paired region (such as *zeta-eta*) that has been RIPped to completion: RIP is indeed re-initiated at the appropriate frequency (though the number of transversions is lower due to site saturation). So far, all sequences >1kb RIP at equivalent frequencies (Selker, 1990): ~100% of linked duplications suffer some degree of RIP, as do ~50% of unlinked duplications.

The idea that one-on-one pairing is involved in every round of RIP is supported by the observation that RIP can affect two (of three) genes severely, while leaving the third largely untouched (Fincham *et al.* 1989); this can be explained (Faugeron *et al.* 1990) by the pairing of two genes randomly in the first stages of RIP, followed by a preferential association of these genes as RIP continues, due to a greater sequence similarity between the two RIPped genes relative to the un-RIPped gene.

It is possible that the mechanism responsible for the actual RIP mutagenesis is actually more sensitive to local pairing than is the homology-searching mechanism (if, in fact, the two are separable). Thus, the estimates for minimum length and minimum homology may be high.

MIP in *Ascobolus*: MIP, or "Methylation Induced Premeiotically," is a premeiotic epimutation mechanism affecting repeated sequences in *Ascobolus immersus* (Rossignol and Faugeron 1994). MIP was named for

its phenomenological similarity to RIP. Instead of resulting in point mutations, MIP causes high degrees of methylation, thereby silencing duplicated promoter regions.

The smallest effective repeat length reported is less than 450 bp but greater than 320 bp for linked genes(Rossignol and Faugeron 1994); for unlinked genes, a repeat of 600 bp allows MIP, but at a lowered frequency. Naturally occurring duplications in *Ascobolus* are generally heavily methylated, as if MIPed.

The rate of MIP is essentially 100% for closely linked genes, and is 50% for unlinked genes, as for RIP. In both systems, this pattern may indicate that the stability or persistence of the search mechanism falls off at the same rate as the rate of crossing over, or it may demonstrate that chromosomal regions are somehow compartmentally segregated, so that there is only a 50% chance of encounter between unlinked sequences. Because no changes in sequence occur, pairing can occur through multiple rounds (and multiple generations) of RIP. Less work has been done with RIP to examine the requirements for per-cent sequence homology. Given full homology of all repeated sequences, there is no basis for preferential pairing through multiple rounds of MIP. Thus, though it is argued (by analogy to RIP) that MIP inactivates multiple gene copies in a pair-wise fashion, this is manifested by inactivation of two or three (of three) competing sequences.

Proposed pairing mechanisms of RIP and MIP -- are they applicable to other organisms? At first glance, the pair wise inactivation argued to occur in RIP and MIP contrasts with the results seen for dominant PEV in

Drosophila, where two (or even three) wild-type alleles of the *brown* gene can be inactivated by a variegating (dominant) *brown* allele, implying that the two phenomena are not mechanistically similar. However, the same mechanism could produce different effects in the two systems, which allow only for very different assays: in *Drosophila*, PEV allows us to look (indirectly) at the mechanics of entire chromosomes; while RIP and MIP focus attention on very short regions of homology in otherwise haploid fungal genomes. The length of the diploid *Drosophila* chromosome provides a vastly greater number of potential pairing sites. Even if any one "site" (of whatever minimum size) pairs exclusively with a single partner, this wealth of sites might allow the chromosomal region as a whole to pair promiscuously with two or more partners: if, for example, each site randomly chooses its pairing partner, a loose, interlocked mesh-work of 1:1 contacts could be formed between several regions of homology.

Are there pairing-dependent phenomena in other organisms that bridge the gap between *Drosophila* and fungal homology-sensing mechanisms? Plants, like *Drosophila*, but unlike *Neurospora* and *Ascobolus*, have a diploid chromosome complement and often form polyploid tissues (Nagl 1978). Plants exhibit several subtle variants of repeat-induced transcriptional silencing. In some systems, silencing preferentially affects nearby repeated sequences, or allelic sequences, while in others, there are minimal positional preferences. Interestingly, tandem repetition of sequences may predispose a region towards allelic silencing, implying that multiple "pairing" may also occur in plants. Due to the existence of other (post-transcriptional) silencing phenomena in plants, the minimum requirements for repeat-based transcriptional silencing are still open to interpretation. However, it is clear that the length requirement for

recognition is less than several kilobase-pairs, and may be as low as 100 base-pairs (reviewed, e.g., by Matzke, Matzke and Mittelsten-Scheid 1994).

What might account for the differences in repeat-based silencing between plants and Drosophilids? Fly chromosomes may be unusually predisposed to forming strong homologous contacts. Dipteran chromosomes, somatic and polytene, exhibit considerably tighter pairing than is seen in most other diploid or polyploid tissues (Metz 1916). By a variety of cytological criteria (reviewed in Nagl 1978), the somatic pairing in plants appears to be considerably looser than that of *Drosophila*, implying that non-allelic sites might be more exposed to each other, or less constrained within the nucleus, and thus be more available for productive ectopic pairing. In terms of ectopic pairing, a loosely-paired diploid genome is not unlike two adjacent haploid genomes -- allelic sites are not strictly bound to each other, but rather, are free to scan ectopic sites for homology. In general, it is possible that in diploid organisms the strength of chromosomal pairing and the ability to sense homology at non-allelic sites are inversely correlated.

How might all of these sensing mechanisms operate? As pointed out by Rossignol and Faugeron (1994), RIP and MIP have been seen for all sequences tested, at all positions tested, and therefore "must imply a mechanism of search for homology that involves any sequence, or most of the sequences in the genome." Furthermore, this search must be sensitive to islands of homology that are as little as 400 bp long, and that diverge by perhaps 10% of their sequence. These requirements are not unlike those for homologous recombination in mammals (~300 bp minimum, Liskay *et al.* 1987) or for ectopic recombination in yeast (Hawley and Arbel 1993).

Perhaps we may profitably combine information from such disparate pairing-dependent phenomena as recombination, premeiotic gene inactivation and pairing-dependent gene silencing to ask how similar sequences might recognize each other.

MECHANISMS

DNA and histones are perhaps the most highly conserved structures in eukaryotic evolution. Many chromatin-associated proteins involved in gene silencing may share conserved domains and functions: the chromodomain region shared by the *Drosophila* HP1 and Polycomb proteins has also been found in the product of the yeast *swi-6* gene (Lorentz *et al.* 1994) and in mouse and human HP1-like proteins (Singh *et al.* 1991; Wreggett *et al.* 1994). In addition, all cells are at least periodically faced with a certain level of homology and/or repetitiveness in their genomes (though in the case of asexual organisms and prokaryotic cells, this is, of course, limited to the special case of identical sister strands present after replication). Thus, it is quite possible that homology-sensing mechanisms in diverse organisms could be ancient and (largely) conserved. In this appendix, I will try to demonstrate that one simple pairing mechanism might serve as an adequate basis for a diverse multitude of pairing phenomena.

In chapter III and appendix A, I touched briefly on possible mechanisms for pairing of similar sequences, both allelic and arrayed. Several different types of pairing models have been proposed. These can be divided into three categories: DNA-DNA models, chromatin-chromatin models, and

protein-protein models (Table B-I). As discussed in chapter I, protein-protein interactions play a necessary role in our observation of pairing effects. The neomorphic *zeste*^{1} protein is necessary to observe the *zeste-white* interaction, while the *Polycomb*-group and *thrithorax* group genes are required to see the effect of transvection at the *Bithorax* and *Decapentaplegic* loci. It is also possible that sequential complexes of these proteins (which can be immunolocalized to >100 sites along the polytene chromosome arms) could play a primary role in actively mediating chromosome pairing. Alternatively, special "pairing proteins" might exist for this purpose (Comings and Riggs 1971). Protein-protein models for pairing will not, however, be discussed here, for three reasons. First (with the possible exception of the aberrant *zeste*^{1} protein, which forms large aggregates) it is unclear whether there exist appropriate proteins with strong enough DNA-binding and self-association properties to mediate pairing. Next, protein-mediated pairing is not, strictly speaking homologous pairing; it presumably requires only binding-site homology, and therefore, is likely to be indistinguishable from other regulatory interactions such as the cooperativity of enhancer binding proteins or promoter-enhancer looping interactions (Bellomy and Record 1990). Finally, we should not expect that all protein-protein interactions can be modeled on any one example of protein-based chromatin pairing. Rather, each protein may have distinctive dimerization characteristics, different binding site distributions, etc. Therefore, I have chosen to look elsewhere for general mechanisms of homologous pairing.

The other two classes of models share the promise of being more generally applicable. Chromatin-path-based models are, essentially, packing models. By this, I mean that they rely primarily on the weak but cooperative

interactions (Van der Waals, hydrophilic) that might stabilize any two chemical entities having the same molecular composition and preferred spatial conformation (e.g. Riley and Flavell, 1977 quoted above; Henikoff 1996; Figure I-6). Some DNA-based models are proposed to work in a similar manner, albeit on a smaller scale: aligned or interdigitated helices are proposed to stabilize each other by weak interactions. One popular model is the paranemic pairing model of Kleckner *et al.* (Kleckner and Weiner 1993; Weiner and Kleckner 1994; Storlazzi *et al.* 1995). Depicted in two-dimensions, this model makes intuitive sense; the helices follow each others' conformations, and pack neatly (Figure I-6b). However, if we extrapolate from the 3-dimensional crystalographic studies of (Timsit and Moras 1995), we see that the favored packing of identical DNA helices involves a partial interdigitation of 3-4 bases-worth of the major groove. Groove-groove contact then positions the helices so that they jut away from each other at a considerable angle. (Timsit and Moras 1995) point out that this sort of packing could stabilize nearby protein-protein interactions; in contrast, it will not support continuous pairing along the helices, even the presence of supercoiling. In particular, it is hard to see how such packing would produce a requirement for continuous sequence homology.

Other DNA-based models, in contrast, involve major perturbations of the double helix. These include multi-strand (triplex) models and hetero-duplex models. Almost all of the DNA-based models and some of the chromatin-path-based models were originally proposed to account specifically for premeiotic or meiotic pairing, and are evolving in tandem with our understanding of the pre-meiotic events leading to meiotic recombination and gene conversion. Because double-strand breaks occur early in the meiotic recombination process, and seem to set the sites (or the

distribution) of productive recombination events, many of the models for homology sensing have incorporated the idea that strand invasion only occurs in the presence of prior DNA strand breakage. Smithies and Powers (1986) have suggested, for example, that single-stranded DNA "feelers" or "whiskers" emanating from nicks could be responsible for polytene pairing in *Drosophila*, in spite of the fact that such a model would require a large number of standing nicks to be maintained through the endoreduplication process. On the other end of the scale, Kleckner and co-authors (Kleckner and Weiner 1993; Weiner and Kleckner 1994; Storlazzi *et al.* 1995) maintain that strand invasion cannot be involved in early premeiotic events largely because pairing must occur before breakage is initiated (reviewed in Roeder 1995).

How good is the evidence that effective strand invasion can only be mediated by a broken chromosomal end? Chromosome breakage is clearly involved in initiating productive recombination nodules *in vivo*. Furthermore, pairing intermediates can be detected biochemically if they contain partially-resolved chromosomal aberrations (Schwacha and Kleckner 1995, and references therein). However, evidence that all heteroduplex formation must be preceded by chromosome breakage is lacking.

On the other hand, evidence for local homology-sensing by heteroduplex formation in the absence of breaks has been similarly sparse. Notably, in the case of *Neurospora*, there is minimal correlation between RIP tracts and recombination tracts (Irelan, Hagemann and Selker 1994). Extrapolating from yeast, it is believed that the number of strand-breakage events correlates roughly with the number of productive recombination

events. RIP, however, is many times more frequent than recombination, though it requires a similar homologous tract length. Thus, RIP cannot be dependent on chromosomal breakage unless it both encourages chromosome breakage and excludes recombination complexes from the breaks. RIP and recombination may share similar initial requirements that are easily met, but recombination may involve additional requirements that are only occasionally met, such as chromosome breakage.

Henikoff (1996) has recently proposed a model for chromosome packing at the level of the 30 nm fiber. Tomography work (Horowitz *et al.* 1994) has recently supported a model for the 30 nm fiber that is unusual in its degree of accessibility and in its degree of irregularity (Woodcock *et al.* 1993). By this model, much of the chromatin in a cell is organized in a three-dimensional, zig-zag structure bounded at each apex by nucleosomes, and possessing a preferred spatial path from one apex to the next. Side-by-side (paranemic-like) contacts between two homologous fibers could be stabilized both by the packing of like DNA helices and by the packing of chromatin zig-zags. Contact would be favored in several ways. Packing would reduce the entropic effects of ordering molecules around two, separate chromatin tracks. Face-to-face nucleosome packing could occur; other protein contacts could stabilize the interaction; paranemic contacts between the DNA helices could also occur more frequently.

This model is attractive because it suggests how the formation of polytene chromosomes might lead to the appearance of strong boundaries between polytene bands and interbands, without invoking domain-control regions at the borders. Instead, the cooperativity of packing would strengthen with each round of endoreduplication, emphasizing slight differences in

packing preferences. In a similar vein, ectopic contacts between repeated sequences in diploid cells might be actively promoted by the resulting cooperativity of packing. Henikoff (1996) suggests that such contacts may structure and order the nucleus.

While the 30 nm fiber model has many good points, other models also bear consideration. This chapter examines an alternative model in greater depth, and considers the possibility that all homologous pairing is mediated directly by DNA heteroduplex formation. The reason for looking at this model in such detail is that heteroduplex formation can accommodate semi-stable recognition of very short DNA segments; if such a system exists, it could account equally well for the recognition of whole chromosome arms and the recognition of 400 base stretches of homology (such as are detected by RIP and MIP). Furthermore, it may allow us to draw parallels between gene inactivation by PEV in *Drosophila* and gene inactivation by trinucleotide-repeat expansions in humans (see Appendix C). The attractiveness of this model, however, rests precariously on its ability to tolerate mismatches, and an absence of evidence for heteroduplex formation in undamaged somatic cells.

Because it relies on chromatin packing, not DNA contacts, the 30 nm chromatin fiber model of Henikoff (1996) is predicted to show far greater stability in the event of mismatches. However, Henikoff's 30 nm fiber pairing model also has at least one weakness: it relies on greater lengths of DNA homology than are available in cases of RIP or MIP. In other words, 30 nm-level pairing may operate in *Drosophila*, but if so, it is hard to see how the homology-sensing mechanisms of *Ascobolus* and *Neurospora* can be functionally related to it.

DNA-DNA interactions: is non-recombinogenic, interphase heteroduplex formation possible? It may be instructive to consider somatic pairing in the light of a much-studied pairing process, RecA-mediated recombination. What follows is a brief introduction to recombination as observed *in vitro* (Stasiak and Egelman 1994) with reference to a model from Howard-Flanders *et al.* (1984, Figure B-1a-c).

In RecA-mediated recombination between linear single-stranded (ss) DNA molecule and a homologous circular, double-stranded (ds) DNA molecule, strand "invasion" generally starts at sites internal to the linear strand, proceeding first in the 5' direction, then the 3' direction of the invading strand. In the absence of excess ss binding protein (ssbp), which would otherwise stabilize the strand being displaced, branch migration within the RecA-based "linear complex" can produce extended stretches of three-strand, DNA structures, especially in the proximal direction, suggesting that coaxial alignment precedes actual changes in Watson-Crick base pairing. (Based on nuclease experiments, this alignment is equally protective of the "invading" strand and the "displaced" strand, while the "exchange" strand which switches partners is less solidly bound by RecA (Chow *et al.* 1986, 1988). In the presence of RecA, homology searching requires as little as 8 bp; following synapsis, pairing is stable even in the absence of RecA with as little as 26 bp of homology (Hsieh, Camerini-Otero and Camerini-Otero 1992). For fully homologous sequences, this process is essentially isoenergetic, provided that the invading strand is originally single-stranded, and thus has a very high affinity for RecA. In other words, bound RecA seems to provide an adequate energetic sink to power both homologous recognition and strand exchange (Stasiak and Egelman 1994).

Equivalent interactions between two duplex (ds) DNA molecules have been investigated (Kim *et al.*, 1992, reviewed in Stasiak and Egelman 1994), and have been found to require ATP (that is, they are NOT isoenergetic). It has been pointed out (e.g. Camerini-Otero and Hsieh 1993)-- and is intuitively obvious -- that a genome-wide homology sensing mechanism can not afford to be an energetically costly process. Thus, scanning by strand invasion of one ds DNA molecule by another would seen to be an energetically costly mechanism for homology sensing.

Two factors, however, should be taken into account when discussing energetics. First, homology sensing may begin as a locally restricted process involving only nearby regions of perfect homology (and thus have relatively low energetic costs). Pairing -- and sensing -- might then be extended or augmented once "buttons" of homology had been stably formed. These "buttons" would tether like regions, lowering the scope of the search.

Secondly, much of the measured energetic cost of ds-ds interactions can be explained by the energetic requirements (as studies in a purified system) for opening up otherwise-inert DNA. If some other "purposeful" process leads to local DNA melting -- e.g. transcription (Thuriaux, 1977) or replication -- then the overall energetic requirements for ds-ds strand invasion and migration could well be lowered considerably *in vivo*. In fact, work with non-hydrolysable ATP-analogs suggests that most of the observed cost of RecA-mediated strand exchange may actually occur only when recombination tracts "bridge" heterologous regions (Rosselli and Stasiak 1991) or during the final, non-reversible steps which complete the recombination/conversion process (Rosselli and Stasiak 1990). A somatic

homology sensing system would, by definition, not involve irreversible changes, forestalling those energetic requirements. Furthermore, if somatic homology sensing operated in a patch-wise manner, and if the boundaries of those patches and set by heterologous sequences, then the mechanism would involve neither of these steps, and would be far more energetically favorable. (It should be noted here that Stasiak and Egelman {1994} draw the opposite conclusion, suggesting that the binding of proteins to DNA *in vivo* would make it less prone to melting; I feel that their argument refers to an average state of DNA, while mine refers to transient, favorable conditions in active DNA.) In addition, if strand displacement is not required either to proceed or to resolve by productive recombination, then heterologous regions will remain un-melted, serving as energetic "anchors," and helping to correctly resolve heteroduplexed regions before or during mitosis. Thus, it seems reasonable to suggest that Watson-Crick base-pairing may be directly involved in the sensing of homology and recognition of repetitive sequences within complex genomes.

An argument against DNA-based pairing

In 1975, Mayfield and Ellison examined homology sensing in polytene chromosome compaction (Mayfield and Ellison 1975, see also chapter V). While focused on the role of pairing in the linear (as opposed to lateral) structure of polytene chromosome bands, their analysis also showed the following: first, polytene morphology is essentially unchanged under conditions that lead to DNA denaturation; next, in contrast, structure is highly disrupted under conditions that denature and/or strip chromosomal proteins. While it is not strictly necessary that polytene pairing and mitotic pairing must follow the same rules, we will assume that this is so.

Does loss of morphology in polytene chromosomes imply that somatic pairing mechanisms are lost? Loss of morphology and loss of pairing need not be equivalent. In plant chromosomes, for example, diffuse morphology is natural to polytene chromosomes; only after chilling or chemical treatment (which suppress gene activity) do *Phaseolus* "polytene chromosomes" assume a regularly-banded structure (Nagl 1978), p. 62). Nevertheless, these "loose" polytenes are certainly paired, even though they are not tightly packed.

Finally, it is possible that proteins are needed for the stable maintenance of pairing, but that the specificity of the pairing is provided entirely by DNA-DNA interactions (Figure I-3c); that is, structures such as histones could package paired DNA cooperatively, yet not interfere directly with the pairing process.

A stronger challenge to the idea of DNA-based pairing mechanisms comes from the observation that physical assays do not detect heteroduplex DNA until late meiotic prophase (Roeder 1995, review). Roeder notes that this seems to invalidate a number of recombination models, including the much-favored double-strand-break(DSB)-repair model. She suggests that present methods for detection of heteroduplex DNA cannot detect unstable heteroduplexes, which might "self-destruct" during purification because they are rectified by branch migration or are denatured. In addition, short heteroduplex structures might simply be so variable in conformation that they defy electrophoretic analysis. For example, unbroken, double Holliday junction intermediates are recovered only after the period of DSB activity in meiosis (Schwacha and Kleckner 1995); but this might just indicate an increase in their stability following strand

breakage. This suggestion is energetically reasonable: non-coordinated movement of two Holliday junctions produces a large increase in local torsional stress and helical disruption (Fu and Seeman 1993), destabilizing the junction structure; DSB's (or nicks) can reduce local tension, thus stabilizing the structures allowing and allowing their recovery. In other words, it pairing in somatic cells is dependent on short heteroduplexes, one would probably expect these heteroduplexes to be too unstable for recovery by normal means. If Watson-Crick base-pairs form between allelic or ectopic regions of homology, the challenge lies in finding a method to show that they exist.

A proposal: As discussed above, recovery of heteroduplex regions from native chromatin may be technologically infeasible. However, such structures might be detectable in other ways, both cytological and biochemical. If stable heteroduplex formation lies behind the pairing effects suggested to cause heterochromatic compaction, the ability to form and maintain somatic heteroduplexes is likely to be necessary for PEV. Though I have argued that heteroduplex formation might be a largely isoenergetic process, driven directly by DNA homology, it is quite possible that stable somatic structures requires proteins (passive or active) that are localized to the point of strand exchange. Furthermore, if such proteins exist, then their loss or reduction should cause a *su(var)*-like effect. Following this argument, one can make some predictions and suggest a tentative course of experimental inquiry. First, as more *su(var)*s are cloned and sequenced, several will no doubt have predicted products which contain DNA binding motifs. Instead of just producing adequate protein for antibody production, enough protein should be generated to look directly at DNA binding preferences, comparing the relative binding

efficiency for ds, ss, and "crossed oligo" DNA substrates (that is, stable oligomeric substrates which mimic the conformation and allow the binding of Holliday-junction-specific binding proteins). To date, no direct DNA binding activity has been described for a *su(var)* -encoded protein, leading to the idea that they might associate more generally with histones or other chromatin proteins. Instead, perhaps binding activity does exist, but is restricted to unusual DNA structures. If a su(var) product were found to bind preferentially to a Holliday-junction-like structure, it would support the idea that heterochromatin formation is linked to heteroduplex formation.

As another experimental approach to this question, common "crossover-specific" binding proteins such as RecA could be used to probe polytene structure. It might be possible to bind the protein directly to polytene squashes. Alternatively, agarose embedded salivary glands (with a minimum of sheared DNA and maximally preserved structure) could be restricted into small fragments, and the resulting DNA exposed to RecA, followed by antibody purification for RecA. This would provide a DNA sample enriched for regions of high local heteroduplex formation; random PCR could then be used to make probes for polytene and mitotic chromosomes, to localize these regions. If heterochromatin formation is linked to heteroduplex formation at repeated sequences,

Conclusions: Gene manipulation and genome analysis have advanced to the point where we can use genetics to probe the finer details of the biology of repeated sequences. So far, detailed manipulation of local repetitiveness has primarily been done in simple, minimally-repetitive, and (mostly) haploid genomes. Someday, we would like to bring a similar level of

analysis to bear on the richly complex, diploid genome of *Drosophila*. We can then use terms like "repetitiveness," "homology" and "pairing" in an meaningful and analytical way when talking about the organization and regulation of entire chromosomes, and even entire genomes.

Figure B-1. Strand invasion in unbroken DNA. From top to bottom, the model for strand invasion and exchange *in vitro*. The center of a single stranded region, rather than its end, leads the invasion process. (Tripartite coiling is favored by some experimental conditions, and is shown here for visualization, but is not a stable state.) A) invasion (of the dotted strand) is followed by the processes of co-linearization (B), strand migration and (C) strand displacement (which may actually occur in concert). If two duplexes each contributed an invading strand, they could form a structure like that shown in figure I-6A. As for many nuclear processes, rotational stress would have to be accommodated by supercoiling or relieved by the action of topoisomerases if both partners were anchored (rather than free as shown here).

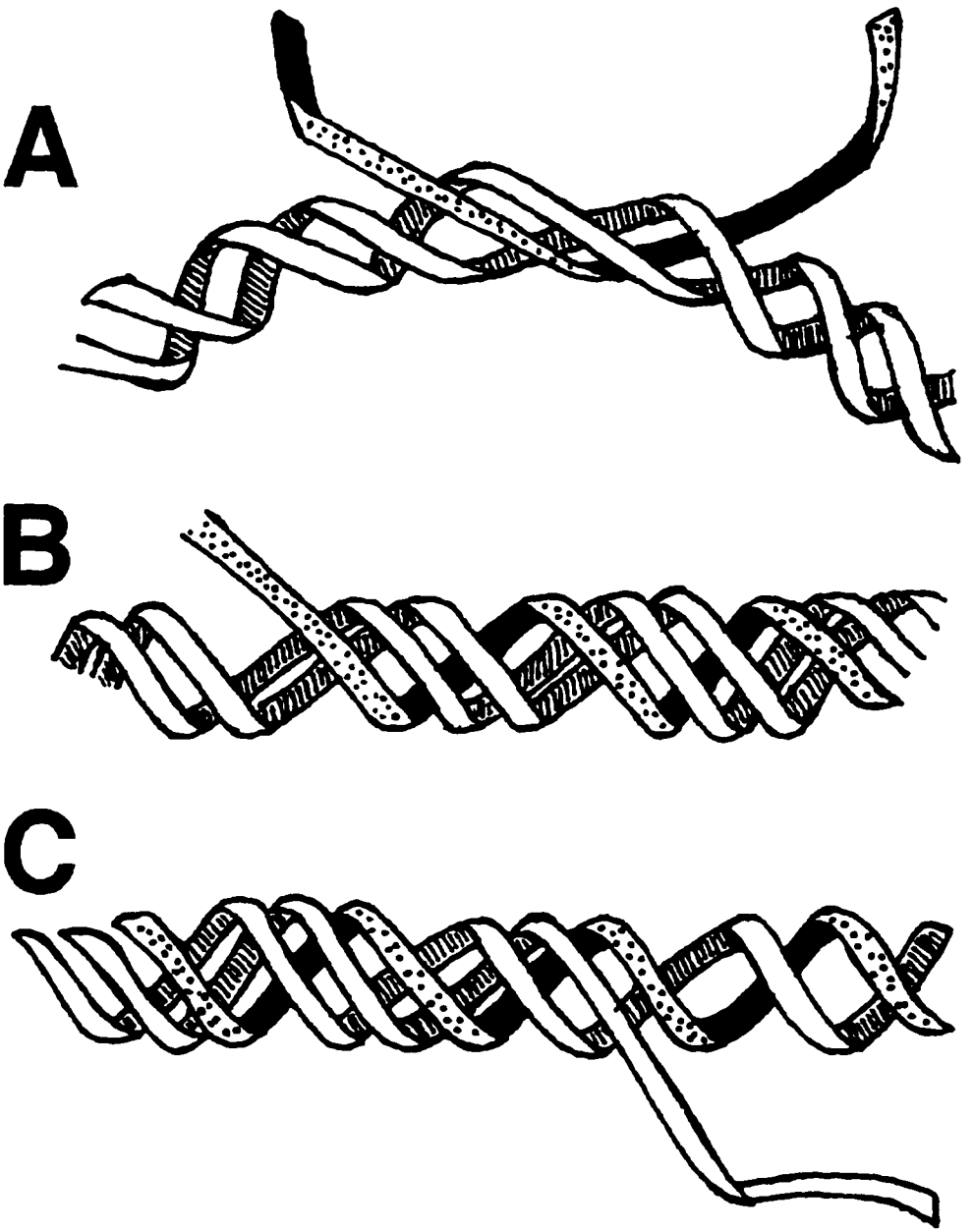
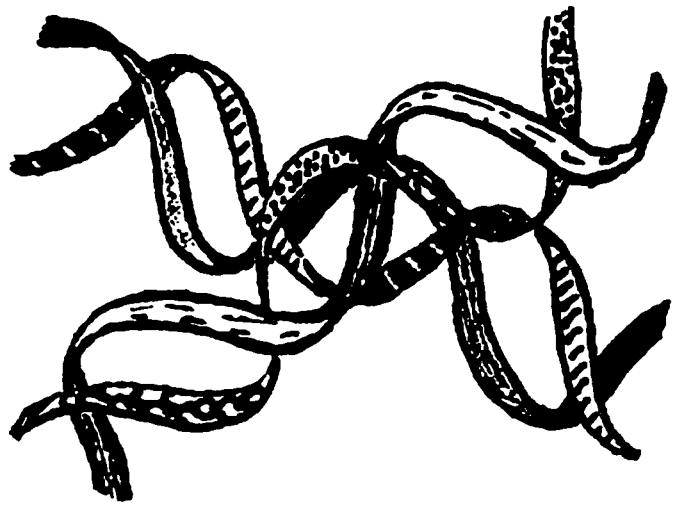

A
B
C

Figure B-2. Close packing of helices leads to non-parallel allignments (modified from Timsit and Moras, 1995). When examined crystalographically, short DNA fragments interdigitate in their major grooves and pack at an angle, not paranemically (see Figure I-6).



Appendix C: Other uses of *16X98* and *98v1* derivatives

Part 1: interphase cytological analysis of *trans*-inactivation

In Chapter I, I described recent cytological results from our lab that show re-localization of bw^D and its homologue to the heterochromatic "compartment" of the nucleus in larval neuroblast cells. However, not all of the relocalization is co-localization; that is, it is possible for one copy of the *brown* gene (presumably that of $brown^{Dominant}$) to be associated with the heterochromatic compartment, while the other copy remains in the euchromatic compartment. Because bw^D is itself a null allele, we do not have a phenotype that corresponds uniquely to this " in / out" cytological pattern. Expression from the $brown^+$ allele could come from either "out / out" or from "in / out" cells in the eye primordium. Because neuroblast tissue may differ from eye imaginal disk tissue, it would be reassuring to have as strong a correlation as possible between types of inactivation and types of localization. With this in mind, I suggest that certain *V21*-derivative lines could be useful in extending the bw^D results.

98X3 and *98XC* are variegating X-ray derivatives of the non-variegating "healed break" line *V21-16X98*. They are moderately inactivated, and are occasionally quite strikingly clonal. These large sectors may allow us to look directly at *trans*-inactivation of *92C* by *98X3* and *98XC*. The *92C* allele is itself a hypomorph. Thus, inactivation of the array, but not of *92C* produces regions of intermediate pigmentation, while joint inactivation produces null (white) patches, and joint expression produces bright red areas. Thus, it is potentially possible to assay *cis-* and *trans-* inactivation in the same cells (just as it is possible to assay localization of either homologue in neuroblast cells). As an added bonus, the imprinting-like

effect observed for *trans*-inactivation of *92C* by *98X3* might prove useful (see chapter III); potentially, enhancement of localization or of pairing would also be seen in neuroblast nuclei.

For a chromosome to be useful in interphase cytology, there must be "free play" between its centromeric region, which will be localized on the nuclear periphery, and the variegating region of interest, with its smaller, isolated heterochromatic block. That is, for purposes of statistical analysis, it should be equally possible to find the reporter gene in any part of the nucleus, except for the "looping" and "pairing" forces being studied. In this particular respect, $brown^{Dominant}$ is ideal. At 59E, it is very distally located. Also, $brown^{Dominant}$ exists on an otherwise wild-type chromosome; therefore, any other wild-type chromosome can be used as a negative control for $brown^{Dominant}$. By comparison, the *16X98* derivatives are all closely linked to their respective centromeres (as evidenced by low rates of crossing over, and by cytology on a subset of the derivatives). Additionally, the majority of the *16X98* lines involve complex rearrangements. On the other hand, work in diploid yeast cells have shown that separations of as little as 150 bp are adequate to randomize the placement of "adjacent" signals within yeast nuclei -- that is, the range and spread of the distance distribution for closely linked loci and for unlinked loci are identical at a 5 micrometer level of resolution (Weiner and Kleckner 1994). For any cell, the distance (in kb) needed for randomization will depend on the local level of chromatin packing, which remains an undefined variable. Finally, though it is unclear what to use for a control (to produce a "no association" distance distribution, transposase mutagenesis could probably produce such a line by removing or reducing the heterochromatic block.

In response to the first two concerns, attention was diverted to the extreme X-ray derivatives of *V21-16X98var1* (the *98v1x* lines). Tests for *trans*-inactivation of *P[bw+]* 92C were done on seven lines, chosen for a relatively large distance between the 92BC block and the enhancing breakpoint (by cytology or reversion; see chapter III). As shown in Figure III-3, some of these alleles produce strong *trans*-inactivation, suggesting that these lines might be very amenable to cytological analysis. As a tradeoff, phenotypic assessment of *cis* and *trans* inactivation cannot be assayed in individual regions of 98v1x eyes (as might be possible with the clonal *16X98* derivatives); instead, overall *cis*- and *trans*-inactivationwould have to be assayed in siblings, perhaps by pigment assays (which should be very consistent, due to the even, salt-and-pepper phenotype).

Part 2: generation of lines to examine the contribution of duplications on the unrearranged chromosome to PEV Clearly, the number of possible combinations of inactivating and inactivatable transgene arrays at 92BC is already large; yet with further mutagenesis, local sequence repetitiveness (and orientation) can be varied further on both chromosomes. One question of continuing interest is whether inverted transgene arrays on a structurally-normal chromosome can participate in enhancing the PEV of a rearranged chromosome (Figure III-6b), perhaps by binding HP1, as has been seen for another transgene duplication (D. Dorer, personal communication). A previous attempt to see such cooperativity was based on mutagenizing duplications of the *P[bw+] 92C* transposon with transposase, crossing them to *V21-E1* females, and screening for enhancement of PEV based on expansion of the *92C* array; this screen was unsuccessful due to the moderate and variable phenotype of *P[bw+] 92C/V21-E1* flies. By comparison, some of the *98v1X* derivatives have strong and consistent phenotypes that would allow such a screen to be successful.

Another path of inquiry would be further modification of the size of the transgene array on the *98v1X* derivative chromosomes, to see whether reduction of gene copy number suppresses PEV even when the gene in question is placed directly against heterochromatin. Preliminary results involving a *P[bw+]* transgene inserted at the base of chromosome 2 suggest that repeat-enhancement may not play a strong role in PEV when the local milieu is already highly repetitive (data not shown). Deletion of the transgene array in a strong *98v1X* derivative would be useful in its own right, as it would provide a very strong, dominant null allele with which to study *trans*-inactivation at the 92BC site.

Appendix D: PEV As a Model for the Genetics of Human Disease--Huntington Disease, Trinucleotide Expansions and Heterochromatin Formation

Epigene Conversion: A Proposal with Implications for Gene Mapping in Humans

Joy F. Sabl* and Charles D. Laird*†‡§

Departments of *Zoology and †Genetics, and ‡Child Development and Mental Retardation Center, University of Washington; and §Fred Hutchinson Cancer Research Center, Seattle

Summary

Epigenetic modification of DNA is now recognized as a potentially important factor in the inheritance and expression of some mutations; its ability to complicate human genetic analysis is concurrently becoming apparent. One unusual form of epigenetic modification, dominant position-effect variegation (PEV), has been used as a model for Huntington disease. In dominant PEV, a fully dominant mutant phenotype results from stable epigenetic inactivation of an allele adjacent to the structural alteration (*cis*-inactivation) combined with a complementary inactivation of the homologous normal allele (*trans*-inactivation). We now propose that *trans*-inactivation of the normal allele may occasionally persist through meiosis. Such "epigene conversion" occurring at the Huntington disease locus in a few percent of meioses would largely account for the published anomalies in that region's genetic map. This concept could also explain anomalous linkage map data for other disease-causing alleles in humans.

Introduction

The use of DNA sequence polymorphisms for linkage analysis has brought the power of gene mapping to the genetics of human disease. Such analysis, however, depends on the constant nature of alleles. Unusual genetic and epigenetic events that alter allelic structure or function merely complicate linkage analysis in non-human organisms. Their effects are potentially far more disruptive, however, in the assessment of human linkage data, because of restricted sample size, rarity of informative meioses, and recovery of only a single product of each meiosis. Occasional inconstancy of normal and mutant alleles could have important but unrecognized consequences in the search for human disease genes.

The *HD* (Huntington disease) locus was the first human locus to be approximately mapped by RFLP analysis (Botstein et al., 1980; Gusella et al. 1983). Precise mapping of the locus and identification of a candidate gene, however, have thus far been intractable (Pritchard et al. 1991). Inconsistent RFLP mapping data have targeted two mutually exclusive regions on the short arm of chromosome 4 as likely gene sites, on the basis of apparent crossover events in four families; additional unpublished data also exist (MacDonald et al. 1989, 1991; Robbins et al. 1989; Bates et al. 1991; M. Hayden, personal communication). A realistic explanation is still needed for these apparent inconsistencies, in spite of recent analysis that redefines the nature of the event in one of the original families (Bates et al. 1991; MacDonald et al. 1991; Pritchard et al. 1991; M. Hayden, personal communication). We propose that events similar to gene conversion, but of epigenetic origin, can account for the apparent inconsistencies. This proposal is an extension of our previous model, in which HD results from dominant position-effect variegation (PEV) (Laird 1990).

Dominant PEV and its Application to Huntington Disease

PEV is the variable but clonally stable inactivation of a euchromatic gene that has been placed adjacent to heterochromatic sequences (Muller 1930; Henikoff 1990). In an example of dominant PEV found in *Drosophila melanogaster* (Muller 1932), the juxtaposi-

Received December 2, 1991.

Address for correspondence and reprints: Joy Sabl, Department of Zoology, NJ-15, University of Washington, Seattle, WA 98195.

0002-9297/92/5006-0003$02.00

tion of heterochromatin causes the adjoining *brown* locus on the disrupted chromosome to be epigenetically inactivated (*cis*-inactivation) and also creates a secondary epigenetic inactivation of the allele on the homologous chromosome (*trans*-inactivation), probably through the binding of heterochromatic proteins (Henikoff and Dreesen 1989). Dominant PEV has been proposed to be the basis of Huntington disease (Laird 1990) because it provides an explanation for two unusual aspects of the disease's inheritance and expression. First, the *HD* mutation is fully dominant (Wexler et al. 1987; Myers et al. 1989); that is, the range of phenotypes in apparently homozygous individuals is identical to that in heterozygotes. Second, the severe, early-onset variant of Huntington disease shows an odd pattern of gender-biased transmission: patients with an early onset of Huntington disease are more likely to have inherited the disease from their fathers (Merritt et al. 1969; Ridley et al. 1988). The dominant PEV model for Huntington disease has two main components. The first component is a proposed chromosomal alteration that results in *cis*-inactivation of an essential, but otherwise normal, coding region, thereby creating the *HD* allele (fig. 1*a–c*). This *cis*-inactivation leads to *trans*-inactivation of the homologous HD^+ allele (fig. 1*d*), mediated by somatic pairing at the *HD* locus. *Trans*-inactivation of the normal (HD^+) allele presumably results directly from the *cis*-inactivation rather than from the proximity of the original chromosomal alteration. In accordance with this argument, any inactivated allele would potentially retain the ability to *trans*-inactivate its properly paired homologue, a point to which we shall return. Recent observations demonstrate somatic pairing in a chromosome-specific and a cell type–specific manner in humans (Arnoldus et al. 1989).

The second component of the model is a rare, recessive allele of an X-linked locus—the *enhancer of HD*—that stably increases the strength of the inactivation of the *HD* locus within the germ line of the parent, resulting in affected progeny with early onset of disease symptoms. Two potential candidate genes for the proposed X-linked *enhancer of HD* have been identified through sequence similarity to a gene encoding a *Drosophila* heterochromatic protein (Singh et al. 1991).

Implicit in the dominant PEV model for Huntington disease is the assumption that the probability of *cis*-inactivation can be stably modified in the parental germ line; that is, the proposed *enhancer of HD* has an effect that is realized one generation after the enhancer has encountered the mutant *HD* allele. In contrast, the

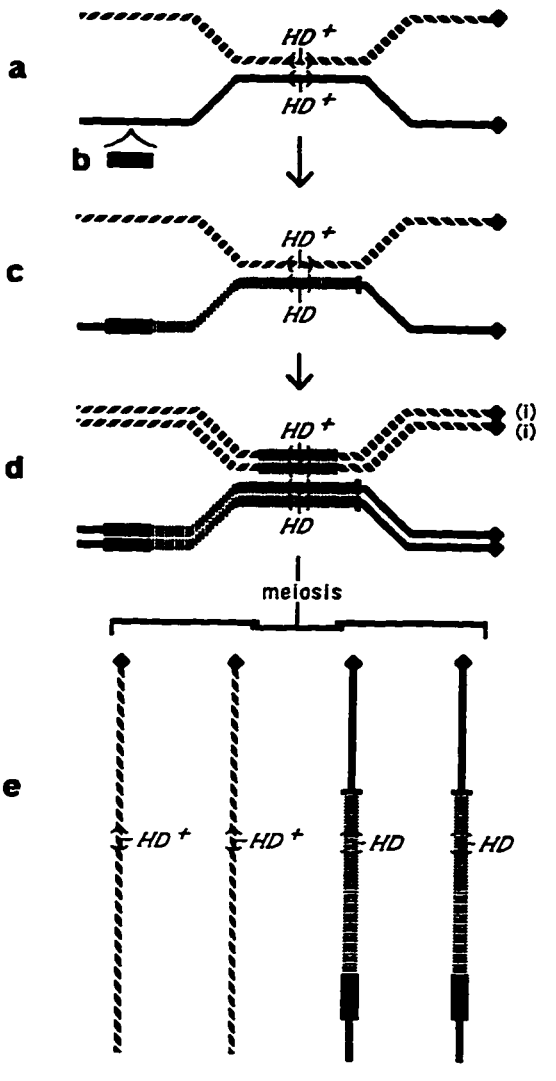


Figure 1 PEV and Huntington disease. The dominant PEV model (*a*) proposes that pairing occurs at or near the *HD* locus. Telomeres are represented as blackened diamonds (◆) to orient the chromosome. A chromosome alteration (*b*; white-spotted black bar) leads to inactivation (*c*; hatch marks [▩]) of the adjacent *HD* allele. This *cis*-inactivation (*d*) then causes *trans*-inactivation of the homologous wild-type allele—inactivation is indicated by (i). The *trans*-inactivation is usually erased prior to meiosis, leading to gametes (*e*) that will produce approximately Mendelian ratios of the Huntington disease phenotype in the next generation.

trans-inactivation effect is usually lost as the parental germ-line cells pass through meiosis, producing gametes that are normal at the *HD* locus (fig. 1*e*). This

latter inference is based on the observation of approximately normal meiotic segregation of the mutant and normal alleles at the *HD* locus and on the generally effective use of RFLP mapping at the *HD* locus.

Epigene Conversion: Persistence of Inactivation through Meiosis

We now propose that, in rare cases, the *trans*-inactivation of the HD^+ allele is retained through meiosis, allowing the inactivated epigenetic state to persist into the next generation (fig. 2). This effectively turns the previously normal allele into an *HD* allele for that generation. As argued above, it can then act on its homologous locus through the same pairing-based mechanism used by the "true" *HD* allele, producing an individual affected with Huntington disease. *Trans*-inactivation that persists through meiosis will be termed "epigene conversion."

The change of a normal allele to a phenotypically mutant allele may have precedence in two phenomena described elsewhere. "Paramutation," as defined for maize (Brink 1973), is a set of processes whereby the presence of an unusual mutant allele causes the heritable alteration of a normal allele to the mutant state. For some loci studied, this change is correlated with changes in DNA methylation patterns, although causality has not been demonstrated (V. Chandler and J. Kermicle, personal communications). Paramutation differs from epigene conversion in that the former occurs with very high frequency. In the ascomycete *Ascobolus*, the presence of multiple copies of any one of several genes leads to the transient inactivation of all copies of that gene; the inactivation occurs by methylation, is reversible, and appears to be regulated through a pairing-dependent mechanism (Faugeron et al. 1990).

How does epigene conversion compare with the recognized process of gene conversion? The latter process is a nonreciprocal *genetic* change in which the nucleotide sequence of one allele becomes entirely or partially like that of its homologous allele, as seen in *Neurospora* (Mitchell 1955) and yeast (reviewed in Roman 1986). By comparison, *epigene* conversion, as postulated here, nonreciprocally changes the functional (*epigenetic*) state of one allele to that of its homologous allele, without a change in nucleotide sequence (fig. 2*a*). Because RFLP mapping is based on the assumption that only standard crossover events occur, both gene conversion and epigene conversion could confound RFLP maps if their phenotypic effects were recorded but their molecular basis went unrecognized.

If gene conversion spans a polymorphic DNA marker, the conversion event can be documented directly by sequence analysis. In contrast, epigene conversion cannot be localized by sequence analysis. If DNA-methylation changes are involved in epigene conversion, RFLP analysis with methyl-sensitive restriction enzymes potentially could be used to identify the crucial region.

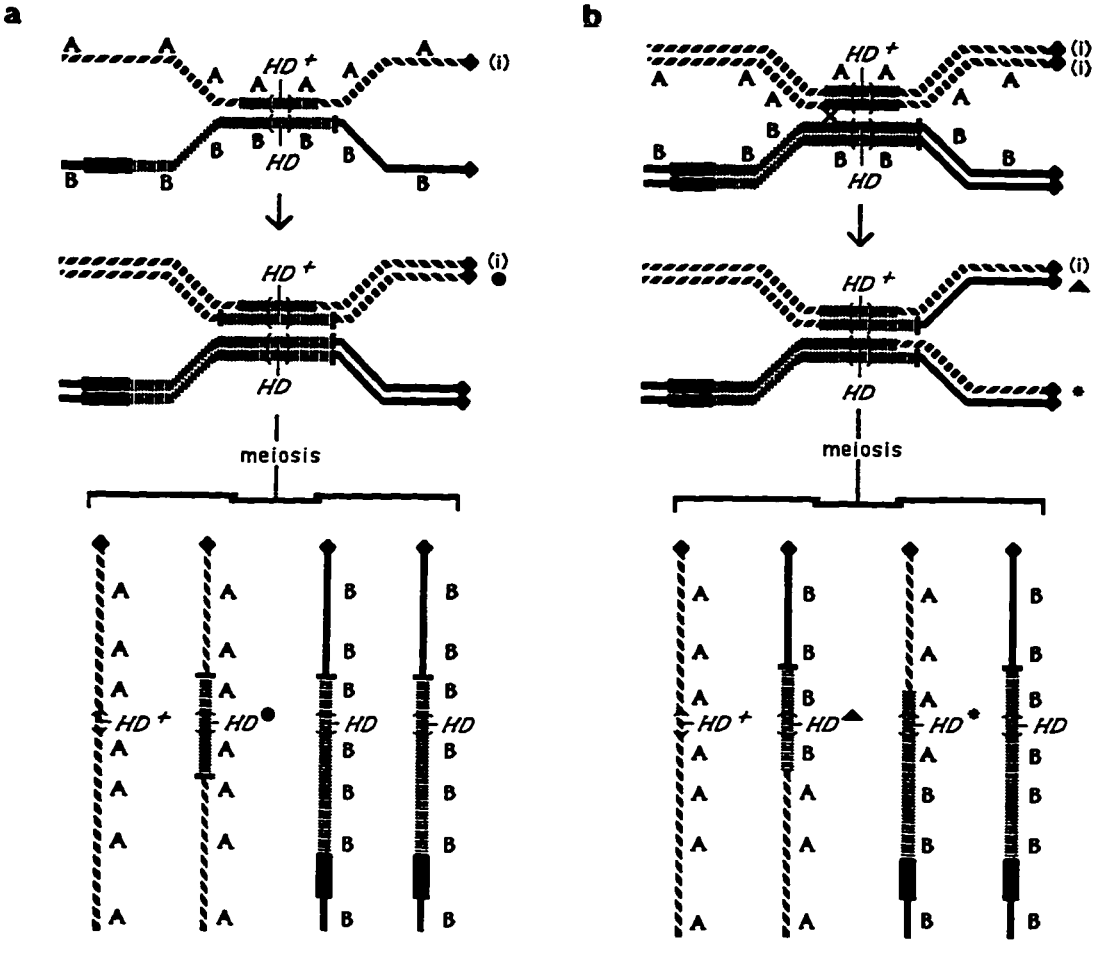
Another important distinction between the two processes is that high levels of gene conversion will disrupt linkage disequilibrium. If epigene conversion is unstable through a second generation, as we predict, it would not reduce any preexisting linkage disequilibrium in a population. Epigene conversion is therefore fully compatible with the demonstrated linkage disequilibrium at the *HD* locus (Theilmann et al., 1989).

Possible Mechanisms for Epigene Conversion

In the simplest scenario, epigene conversion results directly from stabilized *trans*-inactivation of the HD^+ allele and is not linked to or mediated by any genetic event (fig. 2*a*). However, just as gene conversion may occur in conjunction with crossing-over and yet be recognized as a distinctive event in its own right, so too may epigene conversion benefit from and be associated with certain crossover events while retaining its inherent characteristics. To see how the two processes may be associated, we consider in greater detail the PEV model of Huntington disease.

According to the PEV model, the element responsible for the Huntington disease phenotype is not a single entity. It includes (*a*) the original chromosome alteration postulated to act by initiating inactivation, (*b*) the transcribed *HD* locus encoding a crucial gene product, and (*c*) the inactivation complex spreading from the alteration thereby prohibiting transcription. When one of these three elements is separated from the other two by crossing-over in the parental germ line (fig. 2), three of the four potential gametes contain part of the inactivation system. Because the effect of the inactivation system is carried out epigenetically rather than genetically, it is possible for both of the crossover products to maintain inactivation of the structural gene through gametogenesis and beyond (fig. 2*b* and *c*). Either crossover product might inactivate a homologue in the resulting progeny and thereby cause Huntington disease.

Different apparent contradictions in RFLP mapping data would arise from each of two types of epigene conversion. In the absence of crossing-over, epigene conversion would give the appearance that the locus being mapped lay distal to any of the informative



RFLP markers (fig. 2*a*). This happens because the most parsimonious explanation of the data would invoke a single, far-distal crossover event to explain the segregation pattern.

Subtler inconsistencies would result from epigene conversion with adjacent crossing-over. If the crossover lay between the coding region and the chromosomal alteration, the apparent site of the *HD* "gene" would be shifted proximally (fig. 2*b*). Because the phenotype would be determined primarily by the segregation of the chromosome alteration, the crossover point could be significantly proximal to the "actual" gene—the *HD* coding locus. (This argument is based on the assumption that the linear order is centromere–alteration–structural locus; if the chromosome alteration is actually the most distal element, the discussion can be appropriately modified.)

If the crossover lay distal to the *HD* region, the existing *trans*-inactivation may be stabilized (fig. 2*c*), contributing to the frequency of epigene conversion and causing a slight distal bias in mapping: the crossover point could lie recognizably distal to the structural locus, leading to the belief that the gene was more distally located. In this way, each class of epigene conversion would lead to aberrant RFLP mapping data; data from each class considered separately would suggest three mutually exclusive locations for the gene, one of which is greatly distal relative to the

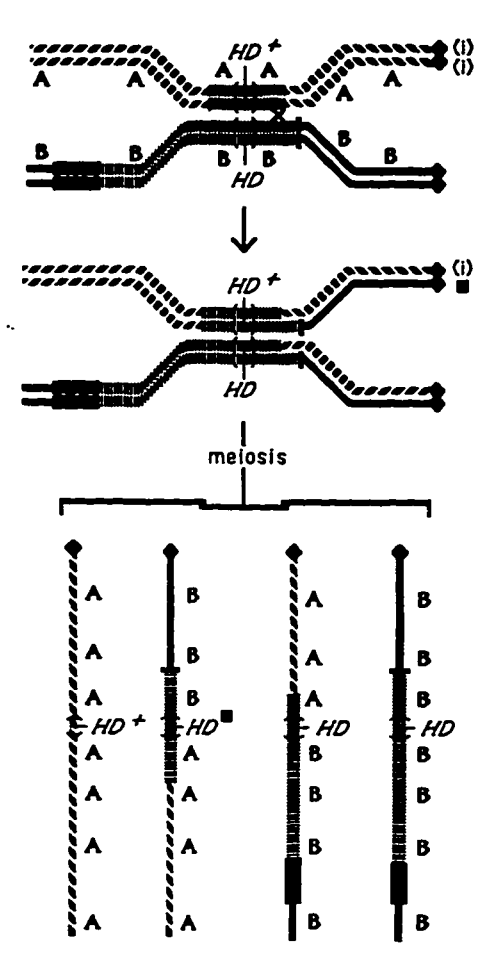


Figure 2 Epigene conversion mechanisms. Epigene conversion is expected to produce rare, non-Mendelian segregation of phenotypes. It can result from either crossover or noncrossover events. *a*, Epigene conversion without cross-over. When *trans*-inactivation is stabilized in a germ-line cell (■), the inactivation of the HD^+ allele can persist beyond meiosis. In the next generation, the Huntington disease phenotype can result from this epigenetically converted allele (●), as well as from the original *HD* allele (*HD*). *b*, Epigene conversion with proximal crossover. If crossing-over occurs within the inactivation domain proximal to the *HD* locus, both crossover products receive part of the inactivation complex. The HD^+ allele will be converted into a stable *HD* allele (*); the original *HD* allele will eventually be reactivated but may retain its inactivation in the generation following the crossover event (▲). *c*, Epigene conversion with distal crossover. If crossing-over occurs within the inactivation domain immediately distal to the *HD* locus, the original *HD* allele will be retained. Juxtaposition of *cis*-inactivated sequences may stabilize the *trans*-inactivation effect on the HD^+ allele, causing it to function as an *HD* allele in the next generation (■). Symbols are as in Fig. 1. Markers (AAA/BBB) are included

other two. It might be possible to distinguish the classes by the persistence of the crossover allele in subsequent generations: for one class it is the more stable crossover product that will be misinformative (fig. 2*b*), while for the other classes it is the presumably unstable, converted allele that directs attention away from the true location of the structural gene (fig. 2*a* and *c*).

Testing the Epigene-Conversion Theory

How might rare cases of epigene conversion be recognized in Huntington disease families? Particular attention should be given to progeny of probands for whom RFLP data appear to be aberrant; decreased stability of a "converted" *HD* allele might be detected in these progeny as apparent cases of reversion to HD^+. Attention should also be given to a possible effect of the postulated X-linked *enhancer of HD* (Laird 1990; Singh et al. 1991) on the recovery of seemingly revertant progeny. Familial correlation (positive or negative) of epigene conversion with early-onset Huntington disease may provide insight into a possible mechanistic relationship between strengthening of somatic *trans*-inactivation and stability through meiosis of epigenetically converted *HD* alleles. Autosomal enhancers of epigenetic stability may contribute to general familial variation in age at onset (beyond the extreme early-onset cases discussed in the Introduction). One might find an elevated level of epigene conversion in families with a young overall age at onset. As discussed above, molecular tests of epigene conversion may also be possible.

A Complementary Proposal: "Incomplete Penetrance" in Huntington Disease

Our proposal that epigene conversion can lead to expression of the Huntington disease phenotype in the absence of a true *HD* allele has a complementary prediction: the Huntington disease phenotype may be absent in spite of the presence of a true mutant *HD* allele (incomplete penetrance). This unusual prediction stems from the proposed structure of the *HD* locus. If crossing-over occurs immediately adjacent to the original site of chromosome alteration, the unstable *trans*-inactivation could be lost before *cis*-inactiva-

to aid in the visualization of RFLP mapping results. Given the combination of a potentially large inactivation domain (based on the reach of PEV effects in *Drosophila*) and a high crossover rate due to pairing, the potential for epigene conversion based on crossing-over may be as high as that for simple epigene conversion.

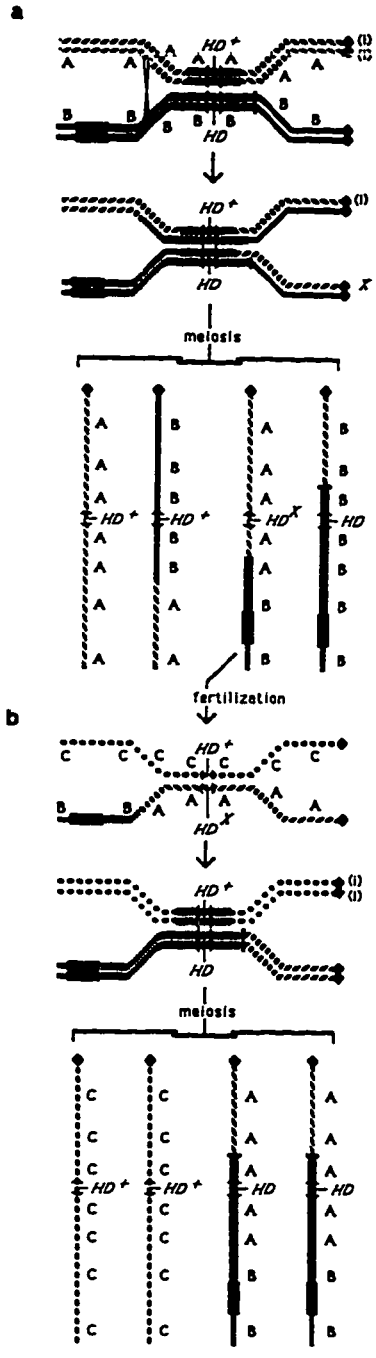


Figure 3 Predicted incomplete penetrance. As shown in panel *a*, if *cis*-inactivation propagates slowly along the chromosome after crossover at the same time as the previously present *trans*-inactivation complex recedes at meiosis, then the newly created *HD* allele (HD^X) may function as an HD^+ allele during the development of the resulting individual. This effect will be dependent on the relative rates of gamete formation and of postmeiotic epigenetic inactivation and on the timing of the activity of the *HD* gene. As can be seen in panel *b*, once inactivated in the course of that generation, the allele should be stable, causing Huntington disease in succeeding generations. Rare pedigrees where apparent transmission through an unaffected individual follows a crossover event could be used to delimit the region of chromosome alteration. Because the absence of Huntington disease is not definitively scored as a phenotype until very late in life, the implications of a "silent" *HD* allele for RFLP mapping are lessened. Symbols are as in figs. 1 and 2.

tion has propagated along the chromosome to reach the *HD* structural locus (fig. 3). RFLP data for the family would show crossing-over in the generation prior to the nonpenetrant individual, with possible reappearance of the Huntington disease phenotype in the individual's progeny. Detection of this rare phenomenon would be consistent with the epigene-conversion effects proposed here.

Concluding Remarks

The significance of epigene conversion for Huntington disease is that it provides a new theoretical framework wherein linkage data that currently seem inconsistent are not only tolerated, but expected. In addition, candidate *HD* genes may reside in chromosome regions that previously were thought to have been ruled out.

The general significance that epigene conversion has for human genetics will depend largely on the number and importance of mutant phenotypes that are influenced by epigenetic modifications, of mutations that affect normal alleles in *trans*, and of the frequency of epigene-conversion events. More broadly, recessive as well as dominant genes could be affected through the mechanism of crossover-based epigene conversion. In the absence of molecular data, these events would appear at least transiently as gene conversion; but they would more properly be described as epigene conversion due to the lack of change in the DNA sequence of the transcription unit and its control regions. The occasional occurrence of epigene conversion could explain anomalous RFLP data for Huntington disease and other human diseases.

Acknowledgments

We thank T. Bird, V. Chandler, G. Côté, R. Esposito, J. Evans, L. Farrer, P. Harper, M. Hayden, M. Howe,

N. Kosuk, M. Lamb, R. Myers, R. M. Myers, and D. Stadler for comments on the manuscript or on ideas therein. This work was supported by grants from NIH (General Medical Sciences), NSF (Eukaryotic Genetics), and the Joseph P. Kennedy, Jr., Foundation.

References

Arnoldus EPJ, Peters ACB, Bots GTAM, Raap AK, van der Ploeg M (1989) Somatic pairing of chromosome 1 centromeres in interphase nuclei of human cerebellum. Hum Genet 83:231–234

Bates GP, MacDonald ME, Baxendale S, Youngman S, Lin C, Whaley WL, Wasmuth JJ, et al (1991) Defined physical limits of the Huntington disease gene candidate region. Am J Hum Genet 49:7–16

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314–331

Brink R (1973) Paramutation. Annu Rev Genet 7:129–152

Faugeron G, Rhounim L, Rossignol J-L (1990) How does the cell count the number of ectopic copies of a gene in the premeiotic inactivation process acting in *Ascobolus immersus*? Genetics 124:585–591

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234–238

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

Henikoff S, Dreesen TD (1989) Trans-inactivation of the *Drosophila* brown gene: evidence for transcriptional repression and somatic pairing dependence. Proc Natl Acad Sci USA 86:6704–6708

Laird CD (1990) Proposed genetic basis of Huntington's disease. Trends Genet 6:242–247

MacDonald ME, Haines JL, Zimmer M, Cheng SV, Youngman S, Whaley WL, Wexler N, et al (1989) Recombination events suggest potential sites for the Huntington's disease gene. Neuron 3:183–190

MacDonald ME, Lin C, Srinidhi L, Bates G, Altherr M, Whaley WL, Lehrach H, et al (1991) Complex patterns of linkage disequilibrium in the Huntington disease region. Am J Hum Genet 49:723–734

Merritt AD, Conneally PM, Rahman NF, Drew AL (1969) Juvenile Huntington's chorea. In: Barbeau A, Brunette JR (eds) Progress in neuro-genetics. Excerpta Medica, Amsterdam, pp 645–650

Mitchell MB (1955) Aberrant recombination of pyroxidine mutants of *Neurospora*. Proc Natl Acad Sci USA 41:215–220

Muller HJ (1930) Types of visible variation induced by X-rays in *Drosophila*. J Genet 22:299–334

——— (1932) Further studies on the nature and causes of gene mutation. Proceedings of the Sixth International Congress of Genetics (Ithaca) 1:213–255

Myers RH, Leavitt J, Farrer LA, Jagadeesh J, McFarlane H, Mastromauro CA, Mark RJ, et al (1989) Homozygote for Huntington disease. Am J Hum Genet 45:615–618

Pritchard C, Cox DR, Myers RM (1991) The end in sight for Huntington disease? Am J Hum Genet 49:1–6

Ridley RM, Frith CD, Crow TJ, Conneally PM (1988) Anticipation in Huntington's disease is inherited through the male line but may originate in the female. J Med Genet 25:589–595

Robbins C, Theilmann J, Youngman S, Haines J, Altherr MJ, Harper PS, Payne C, et al (1989) Evidence from family studies that the gene causing Huntington disease is telomeric to D4S95 and D4S90. Am J Hum Genet 44: 422–425

Roman H (1986) The early days of yeast genetics: a personal narrative. Annu Rev Genet 20:1–12

Singh PB, Miller JR, Pearce J, Kothary R, Burton RD, Paro R, James TC, et al (1991) A sequence motif found in a *Drosophila* heterochromatin protein is conserved in animals and plants. Nucleic Acids Res 19:789–794

Theilmann J, Kanani S, Shiang R, Robbins C, Quarrell O, Huggins M, Hedrick A, et al (1989) Non-random association between alleles detected at D4S95 and D4S98 and the Huntington's disease gene. J Med Genet 26:676–681

Wexler NS, Young AB, Tanzi RE, Travers H, Starosta-Rubinstein S, Penney JB, Snodgrass SR, et al (1987) Homozygotes for Huntington's disease. Nature 326:194–197

An update of the epigene conversion model; its relevance to trinucleotide expansion diseases:

When it was proposed that Huntington's Disease (HD) might result from a process analogous to dominant PEV, nothing was known about the structure of the HD gene, or of the mutations affecting it (Laird, 1990). In 1992, this proposal was expanded, with the hope that incorporation of results from other epigenetic phenomena might help to sort out the confusing linkage data that had made the genetic locus so elusive (above). Shortly thereafter, HD was localized and the HD mutation was shown to result from an expansion of a short trinucleotide repeat (Group THDCR, 1993). Since then, the mutant Huntingtin protein has been identified, localized, and shown to be abundant (Gutekunst *et al.* 1995; Trottier *et al.* 1995), demonstrating that our model is incorrect for Huntington's disease.

Huntington's, however, is but one of a range of human diseases associated with expanded trinucleotide arrays (reviewed in McMurray 1995, Ashley and Warren 1995). While the (CAG/CTG) repeats involved in HD and other diseases tend to be in the coding region of genes, (CGG/CCG) expansions need not be in coding regions (reviewed in Ashley and Warren 1995) suggesting that some repeat expansions interfere with transcription, instead of modifying the final gene product (Mitas *et al.* 1995b). In addition, repeat-based inactivation of these genes may be quite variable, as assayed by phenotype and/or by methylation of the repeat tract (Hansen *et al.* 1993). This has helped to focus attention on two questions about chromatin structure which are familiar to students of *Drosophila* PEV. First, how might the repeated nature of sequences potentially contribute to repeat instability? Secondly, how do changes in the local repetitiveness of DNA influence gene expression? The first question was recently reviewed

by McMurray (*op. cit.*) and will not be discussed further here. The second question is a source of lively conjecture. One possible answer (among several) is that trinucleotide repeats encourage local chromatin condensation (Otten and Tapscott 1995). Condensation may be mediated by proteins that bind preferentially to specific trinucleotide repeats; alternatively, the repeats may form distinctive structures.

With the recent demonstration that disease-related trinucleotide repeats will engage in non-Watson-Crick base pairing to form hairpin DNA structures *in vitro* (Gacy *et al.* 1995; Mitas *et al.* 1995a), attention has been focused on the role of semi-stable hairpins in mediating repeat expansion (McMurray 1995). I would like to suggest that the formation of hairpins may once again link human trinucleotide expansion diseases and PEV in *Drosophila*. I have proposed that repeat-sensing in *Drosophila* and other organisms may be mediated by the formation of semi-stable, recombinationally inert somatic heteroduplex regions, and that such structures could then be recognized by proteins with a binding specificity for Holliday-junction-like structures. Similarly, if hairpins form at trinucleotide repeat expansions, the resulting DNA structures could be recognized and sequestered, leading to the inactivation, compaction, and methylation of the gene involved. As before, I wish to stress that pairing of a normal allele with a mutant homologue could lead to dominant inactivation. For example, formation of hairpin structures could be induced in the short trinucleotide array of the normal allele, once this array finds itself unpaired, due to hairpin formation in the abnormal allele. Alternatively, dominance could occur through simple co-localization, as predicted previously.

The link between repetitiveness and gene silencing has recently received another boost from the field of mammalian genetics. Neumann et al. (1995) have found that imprinted genes and transgenes contain an unexpectedly high concentration of tandemly-repeated sequences in comparison to randomly chosen, non-imprinted genes. Furthermore, in at least one gene, *Igf2/Mpr*, the repeats are clustered in the putative "imprinting box" -- the control region for parent-specific methylation and parent-specific inactivation. Perhaps repeats directly encourage silencing; or perhaps structures formed by the repeats serve as a memory system for imprinting in early development, a time when methylation is largely abolished.

REFERENCES

Allshire, R. C., J. P. Javerzat, N. J. Redhead and G. Cranston, 1994 Position effect variegation at fission yeast centromeres. Cell **76:** 157-169.

Allshire, R. C., E. R. Nimmo, K. Ekwall, J.-P. Javerzat and G. Cranston, 1995 Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. Genes Dev **9:** 218-33.

Ashburner, M., 1989 Drosophila*: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Ashley, C.T. and S.T. Warren, 1995 Trinucleotide repeat expansion and human disease. Ann Rev Genet **29:** 703-728.

Assaad, F. F., K. L. Tucker and E. R. Signer, 1993 Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. Plant Mol Biol **22:** 1067-1085.

Baker, W., 1963 Genetic control of pigment differentiation in somatic cells. Am Zool **3:** 57-69.

Baker, W., 1968 Position-effect variegation. Adv Genet **14:** 133-169.

Baumgartner, W. J., 1904 Some new evidences for the individuality of the chromosomes. Biol Bull **8:** 1-23.

Becker, H. J., 1961 Verhandl. Deut. Zool. Ges., Bonn, 1960 : 283.

Becker, H. J., 1966 Genetic and variegation mosaics in the eye of *Drosophila*. Curr Top Dev Biol **1:** 155-171.

Beermann, W., 1963 Structure and function of interphase chromosomes. Genet Today **1:** 375-384.

Beermann, W., 1967 Gene action at the level of the chromosome. In *Heritage from Mendel*, edited by R. A. Brink. The University of Wisconsin Press, Madison.

Bellomy, G. R., and M. T. Record, 1990 Stable DNA loops in vivo and in vitro: roles in gene regulation at a distance and in biophysical characterization of DNA. Prog Nuc Aci Res **39:** 81-128.

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:** 583-590.

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:** 453-466.

Bender, W., P. Spierre and D. S. Hogness, 1983 Chromosomal Walking and Jumping to Isolate DNA from the *Ace* and *Rosy* Loci and the *Bithorax* Complex in *Drosophila melanogaster*. J. Mol. Biol. **168:** 17-33.

Benson, M., and V. Pirrotta, 1988 The *Drosophila zeste* protein binds cooperatively to sites in many gene regulatory regions: Implications for transvection and gene regulation. EMBO J **7:** 3907-3915.

Berendes, H. D., 1966 Differentiation at the level of the chromosomes. In *Cell Differentiation and Morphogenesis*, pp. 24-54. North-Holland, Amsterdam.

Bickel, S., and V. Pirrotta, 1990 Self-association of the *Drosophila* zeste protein is responsible for transvection effects. EMBO J **9:** 2959-2967.

Biessmann, H., K. Valgeirsdottir, A. Lofsky, C. Chin, B. Ginther, R. W. Levis and M. L. Pardue, 1992 *HeT-A*, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. Mol Cell Biol **12:** 3910-3918.

Bingham, P., R. Levis and G. Rubin, 1981 Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. Cell **25:** 693-704.

Bonner, J., and J. R. Wu, 1973 A proposal for the structure of the *Drosophila* genome. PNAS USA **70:** 535-537.

Bridges, C., 1916 Nondisjunction as proof of the chromosome theory of heredity. Genetics : 107-163.

Bridges, C., 1936 The Bar "gene," a duplication. Science **83:** 210-211.

Bridges, C. B., 1935 Salivary chromosome maps with a key to the banding of the chromosomes of *Drosophila melanogaster*. J Hered **26:** 60-64.

Bridges, C. B., 1938 A revised map of the salivary gland X-chromosome of *Drosophila melanogaster*. J Hered **29:** 11-13.

Bridges, C. B., 1942 A new map of the salivary gland 2L chromosome of *Drosophila melanogaster*. J Hered **33:** 403-407.

Britten, R. J., and D. E. Kohne, 1968 Repeated sequences in DNA. Science **161:** 529-540.

Callan, H. G., 1992 A survey of the first hundred volumes of Chromosoma. Chromosoma **101:** 527-537.

Camerini-Otero, R. D., and P. Hsieh, 1993 Parallel DNA triplexes, homologous recombination, and other homology-dependent DNA interactions Review. Cell **73:** 217-223.

Carmena, M., and C. Gonzalez, 1995 Transposable elements map in a conserved pattern of distribution extending from *beta*-heterochromatin to centromeres in *Drosophila melanogaster*. Chromosoma **103:** 676-684.

Carothers, E. E., 1913 The Mendelian ratio in relation to certain Orthopteran chromosomes. J Morph **24:** 487-509.

Cattanach, B. M., 1974 Position effect variegation in the mouse. Genet Res **23:** 291-306.

Chow, S. A., S. M. Honigberg R. J. Bainton and C. M. Radding 1986 Patterns of nuclease protection during strand exchange. J Biol Chem **261:** 6961-6971.

Chow, S. A., S. M. Honigberg and C. M. Radding 1988 Dnase protection by RecA protein during strand exchange; assymetric protection of the Holliday structure. J Biol Chem **263:** 3335-3347.

Chung, J. H., M. Whiteley and G. Felsenfeld, 1993 A 5' element of the chicken *beta-globin* domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. Cell **74:** 505-514.

Clark, D. V., 1994 Molecular and genetic analyses of *Drosophila Prat*, which encodes the first enzyme of *de Novo* purine biosynthesis. Genetics **136:** 547-557.

Cocciolone, S. M., and K. C. Cone, 1993 *Pl-Bh*, an anthocyanin regulatory gene of maize that leads to variegated pigmentation. Genetics **135:** 575-588.

Commings, D. E., A. D. Riggs 1971 Molecular mechanisms of chromosome pairing, folding and function. Nature **233:** 48-50.

Crain, W. R., F. C. Eden, W. R. Pearson, E. H. Davidson and R. J. Britten, 1976 Absence of short period interspersion of repetitive and non-repetitive sequences in the DNA of *Drosophila melanogaster*. Chromosoma **56:** 309-326.

Creighton, H. B., and B. McClintock, 1931 A correlation of cytological and genetical crossing-over in *Zea mays*. PNAS USA **17:** 492-497.

Demerec, M., 1940 Genetic behavior of euchromatic segments inserted into heterochromatin. Genetics **25:** 618-627.

Demerec, M., and H. Slizynska, 1937 *Mottled white* $^{258\text{-}18}$ of *Drosophila melanogaster*. Genetics **22:** 641-649.

Devlin, R. H., B. Bingham and B. T. Wakimoto, 1990 The organization and expression of the *light* gene, a heterochromatic gene of *Drosophila melanogaster*. Genetics **125:** 129-140.

Dobzhansky, T., 1929 Genetical and cytological proof of translocations involving the third and the fourth chromosomes of *Drosophila melanogaster*. Biolog Zent **49:** 408-419.

Dobzhansky, T., and A. H. Sturtevant, 1931 Translocations between the second and third chromosomes of *Drosophila* and their bearing on *Oenothera* problems. Carneg Inst Wash Publ **421:** 29-59.

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

Dorn, R., J. Szidonya, G. Korge, M. Sehnert, H. Taubert, E. Archoukieh, B. Tschiersch, H. Morawietz, G. Wustmann and A. L. Et, 1993 P transposon-

induced dominant enhancer mutations of position-effect variegation in *Drosophila melanogaster*. Genetics **133:** 279-290.

Dreesen, T. D., S. Henikoff and K. Loughney, 1991 A pairing-sensitive element that mediates *trans*-inactivation is associated with the *Drosophila brown* gene. Gen Dev **5:** 331-340.

Dreesen, T. D., D. H. Johnson and S. Henikoff, 1988 The brown protein of *Drosophila melanogaster* is similar to the white protein and to components of active transport complexes. Mol Cell Biol **8:** 5206--5215.

Dubinin, N. P., 1936 A new type of position effect. Biolog Zh **5:** 851-874.

Eberl, D. F., B. J. Duyf and A. J. Hilliker, 1993 The role of heterochromatin in the expression of a heterochromatic gene, the *rolled* locus of *Drosophila melanogaster*. Genetics **134:** 277-292.

Eddington, C. W., J. L. Epler and J. D. Regan, 1962 The frequency-dose relation of X-ray induced Y-suppressed lethals in *Drosophila*. Genetics **47:** 397-406.

Eissenberg, J. C., 1989 Position effect variegation in *Drosophila*: Towards a genetics of chromatin assembly. Bioessays **11:** 14-17.

Ephrussi, B., and J. L. Herold, 1944 Studies of eye pigments of *Drosophila*. I. Methods of extractions and quantitative estimation of the pigment components. Genetics **29:** 148-175.

Ephrussi, B., and E. Sutton, 1944 A reconsideration of the mechanism of position effect. PNAS USA **30:** 183-197.

Farkas, G., J. Gausz, M. Galloni, G. Reuter, H. Gyurkovics and G. Karch, 1994 The *Trithorax-like* gene encodes the *Drosophila* GAGA factor. Nature **371:** 806-808.

Faugeron, G. ,L. Rhounim and J.-L. Rossignol 1990 How does the cell count the number of ectopic copies of a gene in the premeiotic inactivation process acting in *Ascobolus immersus*? Genetics **124:** 585-591.

Fincham, J. R. S., I. F. Connerton, E. Notarianni, and K. Harrington 1989 Premeiotic disruption of duplicated and triplicated copies of the

Neurospora crassa am (glutamate dehydrogenase) gene. Curr Genet **15:** 327-334.

Flavell, R. B., M. D. Bennett, J. B. Smith and D. B. Smith, 1974 Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet **12:** 257-269.

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:** 2941-2950.

Fu, T. J., and N. C. Seeman, 1993 DNA double-crossover molecules. Biochem **32:** 3211-3220.

Gacy, A. M., G. Goellner, N. Juranic, S. Macura and C. T. McMurray, 1995 Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell **81:** 533-540.

Gans, M., 1953 Étudé génétique et physiologique du mutant z de *Drosphila melanogaster*. Bull Biol Fr Belg Suppl **38:** 1-90.

Garzino, V., A. Pereira, P. Laurenti, Y. Graba, R. W. Levis, Y. Le Parco and J. Pradel, 1992 Cell lineage-specific expression of *modulo*, a dose-dependent modifier of variegation in *Drosophila*. EMBO J **11:** 4471-4479.

Gatti, M., and S. Pimpinelli, 1992 Functional elements in *Drosophila melanogaster* heterochromatin Ann Rev Genet **26:** 239-275.

Gatti, M., S. Pimpinelli and G. Santini, 1976 Characterization of *Drosophila* heterochromatin. I. Staining and decondensation with Hoechst 33258 and quinacrine. Chromosoma **75:** 351-375.

Gelbart, W. M., 1982 Synapsis-dependent allelic complementation at the decapentaplegic gene complex in *Drosophila melanogaster*. PNAS, USA **79:** 2636-2640.

Gellert, M., and H. Nash, 1987 Communication between segments of DNA during site-specific recombination. Nature **325:** 401-404.

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

Glass, H. B., 1933 A study of dominant mosaic eye colour mutants in *Drosophila melanogaster*. II. Tests involving crossing-over and non-disjunction. J Genet **28:** 69-112.

Glover, D. M., M. H. Leibowitz, D. A. McLean and H. Parry, 1995 Mutations in *aurora* prevent centrosome separation leading to the formation of monopolar spindles. Cell **81:** 95-105.

Gowen, J., and E. H. Gay, 1934 Chromosome constitution and behavior in ever-sporting and mottling in *Drosophila melanogaster*. Genetics **19:** 189-208.

Green, M. M., 1963 Unequal crossing over and the genetical organization of the *white* locus of *Drosophila melanogaster*. Hered **13:** 302-315.

Green, M. M., 1967 Variegation of the eye color mutant *zeste* as a function of rearrangements at the *white* locus in *Drosophila melanogaster*. Biol Zent (suppl) **86:** 211-220.

Grigliatti, T., 1991 Position-effect variegation-An assay for nonhistone chromosomal proteins and chromatin assembly and modifying factors. In *Functional organization of the nucleus: A laboratory guide*, pp. 587-627. Academic Press, San Diego.

Group THDCR, 1993 A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell **72:** 971-983

Gubb, D., M. Ashburner, J. Roote and T. Davis, 1990 A novel transvection phenomenon affecting the *white* gene of *Drosophila melanogaster*. Genetics **126:** 167-176.

Gutekunst, C. A., A. I. Levey, C. J. Heilman, W. L. Whaley, H. Yi, N. R. Nash, H. D. Rees, J. J. Madden and S. M. Hersch, 1995 Identification and localization of huntingtin in brain and human lymphoblastoid cell lines with anti-fusion protein antibodies.PNAS USA **92:** 8710-8714.

Hagele, K., 1977 Differential staining of polytene chromosome bands in *Chironomus* by Giemsa banding methods. Chromosoma **59:** 207-216.

Hannah, A., 1951 Localization and function of heterochromatin in *Drosophila melanogaster*. Adv Genet **4:** 87-125.

Hansen, R. S., T. K. Canfield, M. M. Lamb, S. M. Gartler and C. D. Laird, 1993 Association of fragile X syndrome with delayed replication of the FMR1 gene. Cell **73:** 1403-9.

Hartmann-Goldstein, I. J., 1967 On the relationship between heterochromatization and variegation in *Drosophila* with special reference to temperature-sensitive periods. Genet Res **10:** 143-159.

Hawley, R. S., and T. Arbel, 1993 Yeast genetics and the fall of the classical view of meiosis Review. Cell **72:** 301-303.

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.

Heitz, E., 1928 Das Heterochromatin der Moose. I. Jahrb wiss Bot **69:** 762-818.

Heitz, E., 1929 Heterochromatin, Chromocentren, Chromomeren. Ber Dtsch bot Ges **47:** 274-84.

Henikoff, S., 1979 Position effects and variegation enhancers in an autosomal region of *Drosophila melanogaster*. Genetics **93:** 105-115.

Henikoff, S., 1994 A reconsideration of the mechanism of position effect. Genetics **138:** 1-5.

Henikoff, S., 1995 A pairing-looping model for position-effect variegation in *Drosophila*. In *Genomes*, edited by J. P. Gustafson, and R. B. Flavell. Plenum Press.

Henikoff, S., 1996 Position-effect variegation in *Drosophila*: recent progress. In *Epigenetics*, edited by V. E. A. Russo, V. R. Martienssen and A. Riggs. Cold Spring Harbor Press, Cold Spring Harbor NY.

Henikoff, S., and T. Dreesen, 1989 Trans-inactivation of the *Drosophila brown* gene: evidence for transcriptional repression and somatic pairing dependence.PNAS USA **86:** 6704-6708.

Henikoff, S., J. M. Jackson and P. B. Talbert, 1995 Distance and pairing effects on the *brown*^{Dominant} heterochromatic element in *Drosophila*. Genetics **140:** 1007-1017.

Henikoff, S., K. Loughney and T. D. Dreesen, 1993 The enigma of dominant position-effect variegation in *Drosophila*. In *The Chromosome,*

edited by J. S. Helsop-Harrison, and R. B. Flavell, pp. 193-206. BIOS Scientific Publishers, Oxford.

Hinton, T., and K. C. Atwood, 1941 Terminal adhesions of salivary gland chromosomes in *Drosophila*. PNAS USA **27:** 491-496.

Hiraoka, Y., A. F. Dernburg, S. J. Parmelee, M. C. Rykowski, D. A. Agard and J. W. Sedat, 1993 The onset of homologous chromosome pairing during *Drosophila melanogaster* embryogenesis. J Cell Biol **120:** 591-600.

Hopmann, R., D. Duncan and I. Duncan, 1995 Transvection in the *iab-5,6,7* region of the bithorax complex of *Drosophila*: homology independent interactions in *trans*. Genetics **139:** 815-833.

Horowitz, R. A., D. A. Agard, J. W. Sedat and C. L. Woodcock, 1994 The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. J Cell Biol **125:** 1-10.

Howard-Flanders, P., S. C. West and A. Stasiak, 1984 Role of RecA protein spiral filaments in genetic recombination. Nature **309:** 215-220.

Howe, M., P. Dimitri, M. Berloco and B. T. Wakimoto, 1995 Cis-effects of heterochromatin on heterochromatic and euchromatic gene activity in *Drosophila melanogaster*. Genetics **140:** 1033-1045.

Howe, M. M., and J. W. Schumm, 1981 Transposition of bacteriophage *mu*: Properties of *lambda* containing both ends of *mu*. C S H S Q B **45:** 337-346.

Hsieh, P., C. S. Camerini-Otero and R. D. Camerini-Otero, 1992 The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. PNAS USA **89:** 6492-6496.

Irelan, J. T., A. T. Hagemann and E. U. Selker, 1994 High frequency repeat-induced point mutation (RIP) is not associated with efficient recombination in *Neurospora*. Genetics **138:** 1093-1103.

Ising, G., and K. Block, 1981 derivation-dependent distribution of insertion sites for a *Drosophila* transposon. C S H S Q B **45:** 527-554.

Ising, G., and K. Block, 1984 A transposon as a cytogenetic marker in *Drosophila melanogaster*. Mol Gen Genet **196:** 6-16.

Jack, J. W., and B. H. Judd, 1979 Allelic pairing and gene regulation: A model for the *zeste-white* interaction in *Drosophila melanogaster*. PNAS USA **76:** 1368-1372.

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

Janning, W., 1961 [Timing of Heterochromatization in *white*-variegated *Drosophila melanogaster* by X-ray induced mitotic recombination in the eye anlage]. Molecular and General Genetics **107:** 128-149.

John, B., 1988 The biology of heterochromatin. In *Heterochromatin molecular and structural aspects*, edited by R. S. Verma, pp. 1-147. Cambridge University Press, Cambridge, UK.

John, B., and G. L. Miklos, 1979 Functional aspects of satellite DNA and heterochromatin Review. Int Rev Cyt **58:** 1-114.

Johnson, G. D., and G. M. D. C. Nogueria-Araujo, 1981 A simple method of reducing the fading of immunofluorescence during microscopy. J Immunol Meth **43:** 349-350.

Jones, R. N., and H. Rees, 1982 *B chromosomes*. Academic Press, New York.

Karess, R. E., and G. M. Rubin, 1984 Analysis of P transposable element functions in *Drosophila*. Cell **38:** 135-146.

Karpen, G. H., and A. C. Spradling, 1990 Reduced DNA polytenization of a minichromosome region undergoing position-effect variegation in *Drosophila*. Cell **63:** 97-107.

Kassis, J. A., 1994 Unusual properties of regulatory DNA from the *Drosophila engrailed* gene: three "pairing-sensitive" sites within a 1.6-kb region. Genetics **136:** 1025-1038.

Kellum, R., and P. Schedl, 1991 A position-effect assay for boundaries of higher order chromosomal domains. Cell **64:** 941-950.

Kennison, J. A., 1993 Transcriptional activation of *Drosophila* homeotic genes from distant regulatory elements. Trends Genet **9:** 75-79.

Kleckner, N., 1992 McClintock (1933): implications for meiotic chromosome pairing. In *The Dynamic Genome.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Kleckner, N., and B. M. Weiner, 1993 Potential advantages of unstable interactions for pairing of chromosomes in meiotic, somatic, and premeiotic cells. C S H S Q B **58:** 553-565.

Kohler, R. E., 1994 *Lords of the Fly.* The University of Chicago Press, Chicago and London.

Laird, C. D., and B. J. McCarthy, 1968 Nucleotide sequence homology within the genome of *Drosophila melanogaster.* Genetics **60:** 323-334.

Lamb, M. M., and C. D. Laird, 1987 Three euchromatic DNA sequences under-replicated in polytene chromosomes of *Drosophila* are localized in constrictions and ectopic fibers. Chromosoma **95:** 227-235.

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

Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar and F.-M. Sheen, 1993 Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell **75:** 1083-1093.

Lewis, E. B., 1950 The phenomenon of position effect. Adv Genet **3:** 73-115.

Lewis, E. B., 1954 The theory and application of a new method of detecting chromosomal rearrangements in *Drosophila melanogaster.* Amer Nat **88:** 225-239.

Lindsley, D. L., C. W. Edington and E. S. von Halle, 1960 Sex-linked recessive lethals in *Drosophila* whose expression is suppressed by the Y chromosome. Genetics **45:** 1649-1670.

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

Liskay, R. M., A. Letsou and J. L. Stacheleck, 1987 homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics **115**: 161-167

Locke, J., 1993 Examination of DNA sequences undergoing chromatin conformation changes at a variegating breakpoint in *Drosophila melanogaster*. Genetica **92:** 33-41.

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:** 181-198.

Lohe, A. R., and A. J. Hilliker, 1995 Return of the H-word (heterochromatin). Curr Op Gen Dev **5:** 746-755.

Lohe, A. R., A. J. Hilliker and P. A. Roberts, 1993 Mapping simple repeated DNA sequences in heterochromatin of *Drosophila* melanogaster. Genetics **134:** 1149-1174.

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 homeotic gene expression. Development **120:** 2629-2636.

Lorentz, A., K. Ostermann, O. Fleck and H. Schmidt, 1994 Switching gene swi6, involved in repression of silent mating-type loci in fission yeast, encodes a homologue of chromatin-associated proteins from *Drosophila* and mammals. Gene **143:** 139-143.

Lyon, M. F., 1992 Some milestones in the history of X-chromosome inactivation. Ann Rev Genet **26:** 16-28.

Manning, J. E., C. W. Schmid and N. Davidson, 1975 Interspersion of repetitive and nonrepetitive DNA sequences in the *Drosophila melanogaster* genome. Cell **4:** 141-155.

Martin-Morris, L. E., and S. Henikoff, 1995 Conservation of *brown* gene *trans*-inactivation in *Drosophila*. Genetics **140:** 193-199.

Martin-Morris, L. E., K. Loughney, E. O. Kershisnik, G. Poortinga and S. Henikoff, 1993 Characterization of sequences responsible for *trans*-inactivation of the *Drosophila brown* gene. C S H S Q B **28:** 577-584.

Matzke, M. A., A. J. M. Matzke and O. Mittelsten-Scheid, 1994 Inactivation of repeated genes --DNA-DNA interaction? In *Homologous recombination in plants*, pp. 271-307. Kluwer, Amsterdam.

Mayfield, J. E., and J. R. Ellison, 1975 The organization of interphase chromatin in Drosophilidae. The self adhesion of chromatin containing the same DNA sequences. Chromosoma **52:** 37-48.

McClintock, B., 1930 A cytological demonstration of the location of an interchange between two non-homologous chromosomes of *Zea mays*. PNAS USA **16:** 791-796.

McMurray, C. T. 1995 Mechanisms of DNA expansion. Chromosoma **104:** 2-13.

Menzel, M., 1964 Preferential chromosome pairing in allotetraploid *Lycopersicon esculentum--Solanum lycopersicoides*. Genetics **50:** 855-862.

Messmer, S., A. Franke and R. Paro, 1992 Analysis of the functional role of the *Polycomb* chromo domain in *Drosophila melanogaster*. Gen Dev **6:** 1241-1254.

Metz, C. W., 1916 Chromosome studies on the Diptera II. The paired association of chromosomes in the diptera, and its significance. J Exp Zool **21:** 213-280.

Metz, C. W., 1925 Prophase chromosome behavior in triploid individuals of *Drosophila melanogaster*. Genetics **10:** 345-350.

Meyer, P., I. Heidmann and I. Niedenhoff, 1993 Differences in DNA methylation are assoicated with a paramutation phenomenon in transgenic plants. Plant J **4:** 89-100.

Miklos, G. L. G., M.-T. Yamamoto, J. Davies and V. Pirrotta, 1988 Microcloning reveals a high frequency of repetitive sequences characteristic of chromosome 4 and the *beta*-heterochromatin of *Drosophila melanogaster*. PNAS USA **85:** 2051-2055.

Mismer, D., and G. R. Rubin, 1987 Analysis of the promoter of the *ninaE* opsin gene in *Drosophila melanogaster*. Genetics **116:** 565-578.

Mitas, M., A. Yu, J. Dill and I. S. Haworth, 1995a The trinucleotide repeat sequence d(CGG)$_{15}$ forms a heat-stable hairpin containing *Gsyn.Ganti* base pairs. Biochem **34:** 12803-12811.

Mitas, M., A. Yu, J. Dill, T. J. Kamp, E. J. Chambers and I. S. Haworth, 1995b Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)$_{15}$. Nuc Acids Res **23:** 1050-1059.

Moehrle, A., and R. Paro, 1994 Spreading the silence: epigenetic transcriptional regulation during *Drosophila* development. Dev Genet **15:** 478-484.

Montgomery, T. H. J., 1901 A study of the chromosomes of the germ cells of metazoa. Trans Am Phil Soc **20:** 154-230.

Moore, G. D., D. A. Sinclair and T. A. Grigliatti, 1983 Histone gene multiplicity and position effect variegation in *Drosophila melanogaster*. Genetics **105:** 327-344.

Morgan, T. H., 1911 An attempt to analyze the constitution of the chromosomes on the basis of sex-limited inheritance in *Drosophila*. J Exp Zool **11:** 365-411.

Muller, H. J., 1930 Types of visible variations induced by X-rays in *Drosophila*. J Genet **22:** 299-334.

Muller, H. J., 1932 Further studies on the nature and causes of gene mutations. Proc Intl Congr of Genet **1:** 213-225.

Nagl, W., 1977 Early and late DNA replication in respectively condensed and decondensed heterochromatin of *Allium carinatum*. Protoplasma **91:** 389-407.

Nagl, W., 1978 *Endopolyploidy and Polyteny in Differentiation and Evolution*. Elsevier/North-Holland Biomedical Press, Amsterdam, The Netherlands.

Neumann, B., P. Kubicka and D. P. Barlow, 1995 Characteristics of imprinted genes. Nat Genet **9:** 12-13.

Offerman, C. A., and H. J. Muller, 1932 Regional differences in crossing over as a function of chromosome structure. Proc Int Con Genet **2:** 143-145.

Orr, H. T., M. Y. Chung, S. Banfi, T. J. Kwiatkowski, Jr., A. Servadio, A. L. Beaudet, A. E. McCall, L. A. Duvick, L. P. Ranum and H. Y. Zoghbi, 1993 Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet **4:** 221-226.

Otten, A. D., and S. J. Tapscott, 1995 Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. PNAS USA **92:** 5465-5469.

Painter, T. S., 1934a The morphology of the X chromosome in salivary glands of *Drosophila melanogaster* and a new type of chromosome map for the element. Genetics **19:** 448-469.

Painter, T. S., 1934b A new method for the study of chromosome aberrations and the plotting of chromosome maps in *Drosophila melanogaster*. Genetics **19:** 175-188.

Paro, R., 1990 Imprinting a determined state into the chromatin of *Drosophila*. Trends Genet **6:** 416-421.

Paro, R., and D. S. Hogness, 1991 The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. PNAS USA **88:** 263-267.

Pimpinelli, S., M. Berloco, L. Fanti, P. Dimitri, S. Bonaccorsi, E. Marchetti, R. Caizzi, C. Caggese and M. Gatti, 1995 Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. PNAS USA **92:** 3804-3808.

Pirrotta, V., and L. Rastelli, 1994 *white* gene expression, repressive chromatin domains and homeotic gene regulation in *Drosophila*. Bioessays **16:** 549-556.

Platero, J. S., H. T and J. C. Eissenberg, 1995 Functional analysis of the chromodomain of HP1. EMBO J **14:** 3977-3986.

Pontecorvo, G., 1944 Structure of heterochromatin. Nature **153:** 365-367.

Powers, J. A., and J. C. Eissenberg, 1993 Overlapping domains of the

heterochromatin-associated protein HP1 mediate nuclear localization and heterochromatin binding. J Cell Biol **120:** 291-299.

Prokofyeva-Belgovskaya, A., 1939 C. R. (Dokl.) Acad. Sci. U.R.S.S. N. S. **22:** 270-273.

Prokovyeva-Belgovskaya, A., 1947 Heterochromatization as a change of the chromosome cycle. J Genet **48:** 80-98.

Rabinow, L., A. T. Nguyen-Huynh and J. A. Birchler, 1991 A *trans*-acting regulatory gene that inversely affects the expression of the *white, brown* and *scarlet* loci in *Drosophila*. Genetics **129:** 463-480.

Raffel, D., and H. J. Muller, 1940 Position effect and gene divisibility considered in connection with three strikingly similar *scute* mutations. Genetics **25:** 541-583.

Reuter, G., M. Giarre, J. Farah, J. Gausz, A. Spierer and P. Spierer, 1990 Dependence of position-effect variegation in *Drosophila* on dose of a gene encoding a unusual zinc-finger protein. Nature **344:** 219-223.

Reuter, G., and P. Spierer, 1992 Position effect variegation and chromatin proteins. Bioessays **14:** 605-612.

Reuter, G., and I. Wolff, 1981 Isolation of dominant suppressor mutations for position-effect variegation in *Drosophila melanogaster*. Mol Gen Genet **182:** 516-519.

Reuter, G., I. Wolff and B. Friede, 1985 Functional properties of the heterochromatic sequences inducing w^{m4} position-effect variegation in *Drosophila melanogaster*. Chromosoma **93:** 132-139.

Riley, R., and R. B. Flavell, 1977 A first view of the meiotic process. Phil Trans Royal Soc Lond B: Biol Sci**277:** 191-199.

Roberts, P. A., 1972 Differences in synaptic affinity of chromosome arms of *Drosophila melanogaster* revealed by differential sensitivity to translocation heterozygosity. Genetics **71:** 401-415.

Robertson, M., 1988 Homeo boxes, POU proteins and the limits to promiscuity. Nature **336:** 522-523.

Roeder, S., 1995 Sex and the single cell: meiosis in yeast. PNAS USA **92:** 10450-10456.

Roseman, R. R., V. Pirrotta and P. K. Geyer, 1993 The su(Hw) protein insulates expression of the *Drosophila melanogaster white* gene from chromosomal position-effects. EMBO J **12:** 435-442.

Rosselli, W., and A. Stasiak, 1990 Energetics of RecA-mediated recombination reactions: without ATP hydrolysis RecA can mediate polar strand exchange but is unable to recycle. J Mol Biol **216:** 333-350.

Rosselli, W., and A. Stasiak, 1991 The ATP-ase activity of RecA is needed to push the DNA strand through heterologous regions. EMBO J **10:** 4391-4396.

Rossignol, J.-L., and G. Faugeron, 1994 Gene inactivation triggered by recognition between DNA repeats. Experientia **50:** 307-317.

Rykowski, M. C., S. J. Parmelee, D. A. Agard and J. W. 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.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecualr Cloning: A Laboratory Manual (Second Edition)* Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

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 accessiblity of the chromatin fibre. Mol Gen Genet **243:** 453-462.

Schmid, C. W., J. E. Manning and N. Davidson, 1975 Inverted repeat sequences in the *Drosophila* genome. Cell **5:** 159-172.

Schultz, J., 1936 Variegation in *Drosophila* and the inert heterochromatic regions.PNAS USA **22:** 27-33.

Schultz, J., 1941a The evidence of the nucleoprotein nature of the gene. C S H S Q B **9:** 55-65.

Schultz, J., 1941b The function of heterochromatin. Proc 7th Int Cong Genet, 1939 : 257-262.

Schultz, J., and T. Dobzhansky, 1934 The relation of a dominant eye color in *Drosophila melanogaster* to the associated chromosome rearrangement. Genetics **19:** 344-364.

Schwacha, A., and N. Kleckner, 1995 Identification of double Holliday junctions as intermediates in meiotic recombination. Cell **83:** 783-791.

Semeshin, V. F., E. S. Belyaeva, I. F. Zhimulev, J. T. Lis, G. Richards and M. Bourouis, 1986 Electron microscopical analysis of *Drosophila* polytene chromosomes IV. Mapping of morphological structures appearing as a result of the transformation of DNA sequences into chromosomes. Chromosoma **93:** 461-468.

Semeshin, V. F., S. A. Demakov, M. P. Alonso, E. S. Belyaeva, J. J. Bonner and I. F. Zhimulev, 1989 Electron microscopical analysis of *Drosophila* polytene chromosomes V. Characteristics of structures formed by transposed DNA segments of mobile elements. Chromosoma **97:** 396-412.

Shaffer, C. D., and R. J. MacIntyre, 1990 The isolation of the acid-phosphatase -1 gene of *Drosophila melanogaster* and a chromosomal breakpoint inducing its position effect variegation. Mol Gen Genet **224:** 49-56.

Simon, J. A., C. A. Sutton, R. B. Lobel, R. L. Glaser and J. T. Lis, 1985 Determinants of heat shock-induced chromosome puffing. Cell **40:** 805-817.

Singh, P. B., J. R. Miller, J. Pearce, R. Kothary, R. D. Burton, R. Paro, T. C. James and S. J. Gaunt, 1991 A sequence motif found in a *Drosophila* heterochromatin protein is conserved in animals and plants. Nuc Acids Res **19:** 789-794.

Slatis, H. M., 1955a Position effects at the *brown* locus in *Drosophila melanogaster*. Genetics **40:** 5-23.

Slatis, H. M., 1955b A reconsideration of the *brown-dominant* position effect. Genetics **40:** 246-251.

Smithies, O. And P. A. Powers1986 Gene conversions and their relation to homologous chromosome pairing. Phil. Trans R Soc B **312**: 291-302

Speed, R. M., P. Vogt, M. R. Kohler, T. B. Hargreave and A. C. Chandley, 1993 Chromatin condensation behaviour of the Y chromosome in the human testis. Chromosoma **102:** 421-427.

Spofford, J. B., 1976 Position-effect variegation in *Drosophila*. In *The Genetics and Biology of Drosophila*, edited by M. Ashburner, and E. Novitiski. Academic Press, New York.

Stadler, D., H. Macleod and D. Dillon, 1991 Spontaneous mutation at the *mtr* locus of *Neurospora*: the spectrum of mutant types. Genetics **129:** 39-45.

Stasiak, A., and E. H. Egelman, 1994 Structure and function of RecA-DNA complexes. Experientia **50:** 192-203.

Storlazzi, A., L. Xu, L. Cao and N. Kleckner, 1995 Crossover and noncrossover recombination during meiosis: timing and pathway relationships. PNAS USA **92:** 8512-8516.

Sturtevant, A. H., 1913 The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. J Exp Zool **14:** 43-59.

Sutton, W., 1903 The chromosomes in heredity. Biol Bull **4:** 231-251.

Talbert, P. B., C. D. LeCiel and S. Henikoff, 1994 Modification of the *Drosophila* heterochromatic mutation *brown*^{Dominant} by linkage alterations. Genetics **136:** 559-571.

Tartof, K. D., C. Bishop, M. Jones, C. A. Hobbs and J. Locke, 1989 Towards an understanding of position effect variegation. Dev Genet **10:** 162-176.

Tartof, K. D., and S. Henikoff, 1991 Trans-sensing effects from *Drosophila* to humans. Cell **65:** 201-203.

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

Timsit, Y., and D. Moras, 1995 Self-fitting and self-modifying properties of the B-DNA molecule. J Mol Biol **251:** 629-647.

Trottier, Y., D. Devys, G. Imbert, F. Saudou, I. An, Y. Lutz, C. Weber, Y. Agid, E. C. Hirsch and J. L. Mandel, 1995 Cellular localization of the Huntington's disease protein and discrimination of the normal and mutated form. Nat Genet **10:** 104-110.

Udvardy, A., E. Maine and P. Schedl, 1985 The 87A7 chromomere: identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. J Mol Biol **185:** 341-358.

Van Atta, E. A., 1932 Genetic and cytological studies on X-radiation induced eye colors of *Drosophila*. Genetics **17:** 637-659.

Voeller, B. R. (Editor), 1968 *The chromosome theory of inheritance.* Appleton-Century-Crofts, New York.

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*. **125:** 141-154.

Wallrath, L. L., and S. C. Elgin, 1995 Position effect variegation in *Drosophila* is associated with an altered chromatin structure. Genes Devel **9:** 1263-1277.

Weiler, K. S., and B. T. Wakimoto, 1995 Heterochromatin and gene expression in *Drosophila*. Ann Rev Genet **29:** 577-605.

Weiner, B. M., and N. Kleckner, 1994 Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast. Cell **77:** 977-991.

Weismann, A., 1887 On the number of polar bodies and their significance in heredity. In *Essays Upon Heredity and Kindred Biological Problems,* edited by E. B. Poulton, S. Schonland and A. E. Shipley. Clarendon Press, Oxford.

Willard, H. F., 1990 Centromeres of mammalian chromosomes. Trends Genet **6:** 410-416.

Wines, D. R., P. B. Talbert, D. V. Clark and S. Henikoff, 1995 Introduction of a DNA methyltransferase in *Drosophila* to probe chromatin structure *in vivo*. Chromosoma **104:** 332-340.

Woodcock, C. L., S. A. Grigoryev, R. A. Horowitz and N. Whitaker, 1993 A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. PNAS USA **90:** 9021-9025.

Wreggett, K. A., F. Hill, P. S. James, A. Hutchings, G. W. Butcher and P. B. Singh, 1994 A mammalian homologue of *Drosophila* heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. Cytogenet Cell Genet **66:** 99-103.

Wu,C.-t, M. L. Goldberg, 1989 The *Drosophila zeste* gene and transvection. Trends Genet **5:** 189-194

Yunis, J. J., and W. G. Yasmineh, 1971 Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. Science **174:** 1200-1209.

Zhang, P., and A. C. Spradling, 1993 Efficient and dispersed local P element transposition from *Drosophila* females. Genetics **133:** 361-373.

Zhimulev, I. F., V. F. Semeshin and E. S. Belyaeva, 1981 Fine cytogenetical analysis of the band 10A1-2 and the adjoining regions in the *Drosophila melanogaster* X chromosome II. Cytology of the region and mapping of chromosomal rearrangements. Chromosoma **82:** 9-23.

Zuckerkandl, E., 1974 A possible role of "inert" heterochromatin in cell differentiation. Action of and competition for "locking" molecules. Biochimie **56:** 937-954.

VITA

Joy Sabl was born May 27, 1966, two months premature, and has never been early for anything else in her life since then. Nevertheless, she managed to graduate from Harvard University in January, 1989. She experienced the joys of *Drosophila* genetics in the laboratories of Jim Birchler at Harvard and Celeste Berg at the University of Washington before beginning her Doctoral studies.