Copyright © 1985 by the Genetics Society of America

GENETIC AND PHYSICAL STUDIES OF A PORTION OF THE WHITE LOCUS PARTICIPATING IN TRANSCRIPTIONAL REGULATION AND IN SYNAPSIS-DEPENDENT INTERACTIONS IN DROSOPHILA ADULT TISSUES

DAN DAVISON, CAROLYN H. CHAPMAN, CATHY WEDEEN AND PAUL M. BINGHAM

Department of Biochemistry, State University of New York, Stony Brook, New York 11794

Manuscript received December 26, 1984
Revised copy accepted March 25, 1985

ABSTRACT

We have identified and sequenced the portion of the white locus affected by an idiosyncratic set of white mutant alleles (the w^{sp} alleles). The affected white locus portion (w^{sp} region) extends from ca. 590 base pairs (bp) to ca. 1270 bp 5′ to the apparent start site for the major white transcription unit. Based on the properties of these mutant alleles, we infer the existence of two distinct *cis*-acting regulatory elements in the w^{sp} region and a third element mapping 3′ to this region (3′ to position ca. −670). Our analysis allows us to define the apparent position of one of the two w^{sp} region elements with substantial precision. Examination of the DNA sequences in this region suggests that it is functionally similar to the enhancers identified in vertebrates. This same element participates in synapsis-dependent genetic interactions, suggesting a largely unexpected relationship between enhancer-like, *cis*-acting genetic elements and the genetic elements responsible for the synapsis-dependent genetic interactions in *trans* revealed by the existence of transvection effects. Our results further suggest that a presumptive regulatory locus (suppressor-of-white-spotted) regulates white transcription in adult tissues and is not involved in regulating white expression in larvae. We discuss the regulation of white expression in light of our studies. We also demonstrate unusual structures for an X-ray-induced deletion and a spontaneous deletion.

THE w^{sp} mutant alleles at the white locus of Drosophila have long been recognized to have quite special properties as assessed by genetic and cytological studies (LINDSLEY and GRELL 1968 and references therein): These alleles produce a characteristic, drastic reduction in eye pigment deposition. Moreover, the w^{sp} mutations disrupt the synapsis-dependent genetic interactions between synapsed white alleles revealed by the existence of transvection effects (GANS 1953; LEWIS 1954; GREEN 1959; KAUFMAN, TASAKA and SUZUKI 1973; JACK and JUDD 1979; BABU and BHAT 1980, 1981; BINGHAM 1980). Furthermore, in spite of their quite extreme effects on adult eye color phenotype, w^{sp} mutations have no effect on pigmentation of larval malpighian tubules. This is in contrast to essentially all comparably extreme white mutant alleles which exert similar effects on pigmentation in both of these tissues

(LINDSLEY and GRELL 1968). This last observation strongly suggests that w^{sp} mutations have tissue-specific effects on white locus expression.

Molecular studies (ZACHAR and BINGHAM 1982) revealed several additional novel features of these mutations: First, the four w^{sp} mutations affect the same portion of the locus. Second, the detailed structure of these mutations suggested that they affected a genetic element outside of the white locus transcription unit and, thus, were good candidates for regulatory mutations.

Several recent studies (O'HARE, LEVIS and RUBIN 1983; LEVIS, O'HARE and RUBIN 1984; PIRROTTA and BROCKL 1984; CHAPMAN and BINGHAM 1985) support the original proposal of ZACHAR and BINGHAM (1982) that the w^{sp} mutations map outside of the white transcription unit. Furthermore, CHAPMAN, ZACHAR and BINGHAM (1985) demonstrated the existence of a novel type of suppressor locus (suppessor-of-white-spotted) that interacts specifically with the w^{sp} mutant alleles. Mutant alleles at suppressor-of-white-spotted largely restore a wild-type eye color phenotype to w^{sp} individuals.

Thus, the w^{sp} region represents a well-defined opportunity to study presumptive regulatory elements. Moreover, the effects of these mutations on synapsis-dependent genetic interactions at white suggest that these mutations may reveal unexpected properties of such elements. We report here the results of physical, genetic and transcriptional studies of mutations affecting the w^{sp} region of the white locus.

MATERIALS AND METHODS

Fly strains and culture: Description of fly strains can be found in LINDSLEY and GRELL (1968), in ZACHAR and BINGHAM (1982) and in CHAPMAN and BINGHAM (1985). All files were reared at 21.5–22° throghout the developmental stages examined. Control of culture temperature is extremely important in visualizing the effects of $su(w^{sp})$ on white transcript levels in mature adult tissues. At higher temperatures (greater than 24°) the adult effects described are much less extreme (CHAPMAN and BINGHAM 1985; D. DAVISON, C. CHAPMAN and P. BINGHAM, unpublished observations). Adult eye color phenotypes described were assessed within 24–48 h after eclosion.

DNA sequence determination: DNA sequences were determined essentially as described by SANGER, NICKLEN and COULSON (1977), SANGER and COULSON (1978) and BARNES, BEVAN and SON (1983). A description of the fragments sequenced is in Figure 9.

Analysis of DNA sequence similarities: Analysis of nucleotide sequence similarities was performed in several ways. The dot matrix-plotting programs of CONRAD and MOUNT (1982) and PUSTELL and KAFATOS (1984) were used for initial examination of the sequence and for comparison to the BPV enhancer region (WEIHER and BOTCHAN 1984) and the SV40 enhancer core sequence (WEIHER, KONIG and GRUSS 1983). The GOAD and KANEHISA (1982) implementation of the NEEDLEMAN and WUNSCH (1970) and SELLERS (1974) (NWS) metric search procedures were then used to further refine the similarities noted in the dot matrix procedures. Analysis of the direct repeats in the $su(w^{sp})$p region (see RESULTS) was performed using the global alignment implementation of the NWS procedures (FITCH and SMITH 1983).

Statistical analysis was performed by the method of GOAD and KANEHISA (1982); in particular, equations 4 though 7 were used. For all of these calculations the composition of the entire sequence in Figure 2 was used.

The expectation (E) value is the average number of times one would expect to find a match of comparable quality in two randomly chosen sequences of the same length and composition (GOAD and KANEHISA 1982; KANEHISA 1984). For the range of E values calculated below, E values are good approximations of the frequency with which the matches in question should occur at random and we will refer to E values as expected frequencies in the text. The expectation values for the alignments in Figure 8 are 0.012 (SV40) and 0.048 (BPV). The relevant parameters for all

calculations are given below. The direct repeat in the *su(w^{sp})* interval (see DISCUSSION) has an E of 0.0002.

Although the GOAD and KANEHISA (1982) method has been criticized (LIPMAN *et al.* 1984), we note that the E resulting from the method is an upper bound and, therefore, similarities are likely to be more statistically significant than this calculation indicates. Furthermore, the problem pointed out by LIPMAN *et al.* (1984) occurs at particular, unusual boundary conditions which are not applicable in the cases presented here.

The data for the calculations of Figure 8 are: top, $p = 0.228$, $I = 896$ (in this specific case $J = I$; W. GOAD, personal communication); bottom, $p = 0.252$, $I = 896$, $J = 64$. For the direct repeats near the *su(w^{sp})p* region the data are: $p = 0.255$, $l = 68$, $m = 46$, $r = 18$, $g_1 = \phi$, $g_2 = 2$, $I = J = 896$.

S_1 protection analysis of deletion breakpoints: Approximately 10 μg of Drosophila genomic DNA were mixed with approximately 50 ng of an M13-cloned fragment (virion DNA) in 100 μl of 10 mM Tris, 1 mM EDTA (pH 7.4). The mixture was vortexed vigorously to shear genomic DNA slightly and then heated to 95° for 5 min. The mixture was immediately chilled on ice, NaCl was added to 0.2 M and the mixture was incubated at 65° for 1 h. After chilling, an equal volume of a solution containing 50 mM sodium acetate and 6 mM zinc sulfate was added to the mixture and followed by the appropriate amount of S_1. Digestion was carried out at 37° for 15 min. S_1 protection products were analyzed by Southern transfer after fractionation on 1% formaldehyde agarose gels (MANIATIS, FRITSCH and SAMBROOK 1982).

Northern gel analysis: RNAs were analyzed after Northern transfer (MANIATIS, FRITSCH and SAMBROOK 1982) using the hybridization procedure of HU and MESSING (1982) modified by BINGHAM and ZACHAR (1985). Purified head tissues and polyadenylated RNAs were prepared as previously described (BINGHAM and ZACHAR 1985). Larvae were washed from food, and polyadenylated RNAs were prepared as from adult tissues.

white locus DNA sequence probes: white locus DNA sequences were originally cloned by BINGHAM, LEVIS and RUBIN (1981). All white cloned segments used herein were derived from these original segments or were retrieved from clone libraries using these sequences as probes (ZACHAR and BINGHAM 1982; LEVIS, BINGHAM and RUBIN 1982; D. DAVISON, C. CHAPMAN and P. BINGHAM, unpublished observations).

RESULTS

The w^{sp} region mutations all affect the same very small region of the white locus: The structures of the mutant alleles we have analyzed are diagramed in Figure 1. It is convenient to describe the sequences affected by these mutations in relation to the apparent start site for the major w^+ transcript (PIRROTTA and BROCKL 1984; D. DAVISON, C. CHAPMAN and P. BINGHAM, unpublished observations). We will assume this start site to be at coordinate ca. +3.7 kb on the standard white locus map (ZACHAR and BINGHAM 1982; LEVIS, BINGHAM and RUBIN 1982). By convention we designate points to the right (5′ to the major white transcription unit) of this position as having negative values and points to left as having positive values.

[It is important to note that two sources of uncertainty currently exist in placement of the white transcription start site. First, it is inherently difficult to establish that no exons exist 5′ to those currently identified (PIRROTTA and BROCKL 1984; Figure 1). We will assume for simplicity that no such undetected exons exist. Moreover, our results (Figures 6 and 7) strongly suggest that the white transcription unit does not extend into the *w^{sp}* region. Second, we have placed the 5′-most known white exon based on our sequence analysis of the region (identification of potential splice donors and peptide-coding regions, D. DAVISON, unpublished results) and S_1 protection results (PIRROTTA and BROCKL

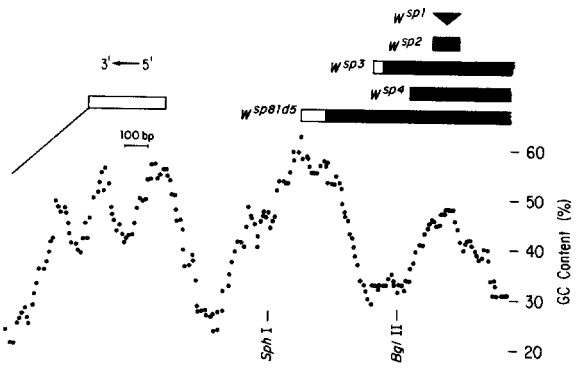

FIGURE 1.—Structures of w^{sp} region mutations and their relation to the major white transcription unit. Top, The five w^{sp} region mutations discussed in the text are diagramed. The triangle represents the point of insertion of the B104 transposon responsible for the w^{sp1} mutation (ZACHAR and BINGHAM 1982; RESULTS). The black bars represent deletions apparently responsible for the other four mutations. The uncertainty in placement of the w^{sp3} and w^{sp81d5} breakpoints is indicated by the white extensions of the black bars. The w^{sp3} deletion extends ca. 17 kb rightward from the *Bgl*II site shown (ZACHAR and BINGHAM 1982); the w^{sp4} deletion extends ca. 1.0 kb rightward from the *Bgl*II site (ZACHAR and BINGHAM 1982; Figure 4), and the w^{sp81d5} deletion extends for an undetermined distance beyong a point 5 kb rightward of the *Bgl*II site as assessed by Southern gel analysis (results not shown). The single white bar represents our placement of the presumptive first white exon (see text) and the direction of the transcription is shown. This figure is oriented according to the conventional genetic map (left to right is telomere to centromere). All subsequent figures showing sequence data are oriented so that left to right corresponds to the 5′ to 3′ direction of the white transcription unit. The sequences shown extend from coordinates +588 to −1470 using the coordinate system described in Figure 2 and in the text. Shown in the bottom portion of the figure is the GC composition of the region. Each point represents the GC composition of the 100 bp leftward of the point and there is a point every 10 bp. The GC composition is determined from sequence analysis (Figure 2; D. DAVISON, C. CHAPMAN and P. BINGHAM, unpublished observations).

1984; D. DAVISON, C. CHAPMAN and P. BINGHAM, unpublished observations) rather than on the basis of the sequence of a full-length cDNA clone of the white message. However, although our placement of the presumptive 5′-most exon is inferential, we believe that it is unlikely to be in error by more than 150 bases and is, thus, usefully accurate in this context.]

We have sequenced the white locus interval extending from position −473 through −1369 and that sequence is shown in Figure 2. We have analyzed the mutational lesions responsible for the w^{sp1}, w^{sp2} and w^{sp4} mutations by DNA sequence analysis of cloned portions of these alleles (Figures 2–4). We have analyzed the deletion breakpoints for the w^{sp3} and w^{sp81d5} deficiency mutations by S_1 protection of cloned DNA segments with genomic DNA (Figure 5).

w^{sp1} results from the insertion of the B104 transposon (SCHERER *et al.* 1982) at position −1170 (Figure 2). w^{sp2} is a spontaneous deficiency mutation whose deleted segment extends from coordinates −1231 to −1115 (Figures 2 and 3). w^{sp3} is an X-ray-induced deficiency mutation whose deleted segment begins between coordinates −925 and −870 and extends further 5′ for at least 17 kb (Figures 2 and 5). w^{sp4} is an X-ray-induced deficiency mutation whose deleted segment extends from coordinates ca. −2200 to −1029 (Figures 2 and 4).

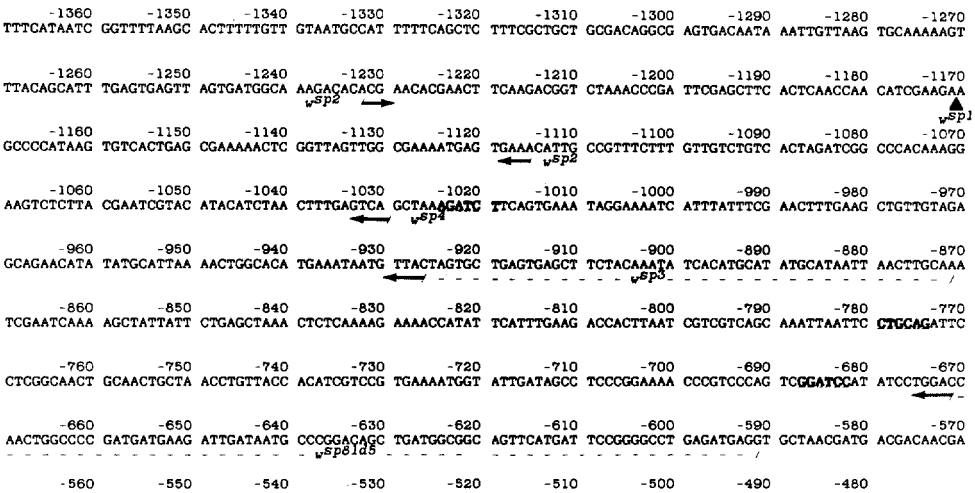

FIGURE 2.—DNA sequence of the w^{sp} region. Shown is the sequence of the portion of the white locus affected by the w^{sp} region mutations. Shown on the sequence is the insertion point of the B104 transposon responsible for w^{sp1} (as assessed by sequencing from right to left on the sequence shown). The w^{sp1} allele is indistinguishable in sequence from the Canton-S w^+ allele on the interval from the *Bgl*II site at coordinate −1019 to the B104 insertion point. We have not sequenced across the w^{sp1} B104 insertion from left to right, and thus, a more complex structure than a simple insertion is not excluded: however, no such structure is indicated by Southern gel analysis (ZACHAR and BINGHAM 1982). O'HARE, LEVIS and RUBIN (1983) reportedly sequenced across the w^{sp1} insertion in the clone of ZACHAR and BINGHAM (1982) and found the insertion to be simple, although no sequence data were presented. Shown as well are the end points of deleted segments in w^{sp} region deficiency mutations. Our estimate of the uncertainty in placement of the w^{sp3} and w^{sp81d5} deletion breaks is indicated by the dashed portion of the line. Also shown (highlighted) are the *Bgl*II (−1019), *Pst*I (−774), *Bam*HI (−682) and *Sph*I (−473) cleavage sites. Strategy for generation of this sequence shown in Figure 9 and additional details of the w^{sp2} and w^{sp4} alleles are shown in Figures 3 and 4.

TGATGGCAAAGA	Canton-S 5′
GTAATGGCAAACACACA---CGAACAGAAC	Oregon-R 5′
GTAATGGCAAACACACA---**CATTGCCGTTTCTT**	w^{sp2}
GAAAATGAGTGAAA---**CATTGCCGTTTCTT**	Oregon-R 3′

FIGURE 3.—The structure of the w^{sp2} spontaneous deletion mutation. Shown are the sequences immediately surrounding the w^{sp2} deletion in the Canton-S and Oregon-R w^+ alleles and in the w^{sp2} allele. The w^{sp2} deficiency removes bases between coordinates −1231 and −1115 (Figure 2). The dashed lines indicate the position of the deletion break, and the highlighted Oregon-R sequences are those not removed by the deletion. The italicized bases denote strain-specific base substitution differences between w^{sp2} and the Canton-S w^+ allele. These differences are shared by w^{sp2} and the Oregon-R w^+ allele and are thus unlikely to be responsible for the w^{sp2} mutant phenotype. On the interval for which fully reliable sequence data are available for all alleles (between coordinates −1269 and −1080; Figure 2) the w^{sp2} allele differs from the Canton-S allele at one other point; between coordinates −1092 and −1085 the w^{sp2} sequence is GTC*GG*TAG and the Canton-S and Oregon-R w^+ sequences are GTC*AC*TAG. We note that these last base substitution differences between w^{sp2} and the two wild-type alleles map 83 bases 3′ to the w^{sp1} insertion, whereas the w^{sp2} deletion covers the w^{sp1} insertion site. On the basis of these observations, we suggest that the base substitution differences between w^{sp2} and the Canton-S w^+ allele represent strain-specific differences present in the wild-type allele parental to w^{sp2} and that the w^{sp2} deletion is solely responsible for the mutant phenotype. The strategy for generating the sequence is described in Figure 9.

```
CCCCCCCACCACCACCACCGACCATTT                                                    Canton-S 5
CCCCCCCTCCACCACCACCGACCATTTACCACTTTTGCCACTTTT--TCCATGGATCGTACTTTTCGGGGCATG     Oregon-R 5
CCCCCCCT  CCACCACCACCGATTTACCACTTTTGCCACTTTT--GCCACTTTTACCACTTTTGCCATTTTT--GCTAAAGATCT   w^sp4
                                              ATCGCACATACAGCTAACTTTGAGTCA--GCTAAAGATCT   Oregon-R 3
```

FIGURE 4.—The structure of the w^{sp4} X-ray-induced deletion mutation. Shown are the sequences immediately surrounding w^{sp4} deletion in the Canton-S and Oregon-R w^+ alleles and in the w^{sp4} allele. The w^{sp4} deficiency removes sequences beginning at coordinate −1029 and extending 5′ approximately 1000 bp (ZACHAR and BINGHAM 1982). The dashed lines represent the positions of the deletions breakpoint(s), and the highlighted Oregon-R sequences are those not removed by the deletion. The italicized bases denote strain-specific base substitutions. Notice that the w^{sp4} has an additonal 3-bp deletion 34 bases 5′ to the large deleted segment. Furthermore, the w^{sp4} allele contains an additional 27 bases (three copies of sequence PuCCACPyTTTT) found only in w^{sp4}. This sequence feature might have been generated during repair of X-ray-induced damage. Other than the changes shown in the figure, w^{sp4} is indistinguishable from the Canton-S w^+ sequence (Figure 2) on the interval for which reliable sequence is available (extending from the *Bgl*II cleavage site at approximately coordinate −2220 through coordinate −992; Figures 1, 2 and 9). The strategy for generating the sequence is described in Figure 9.

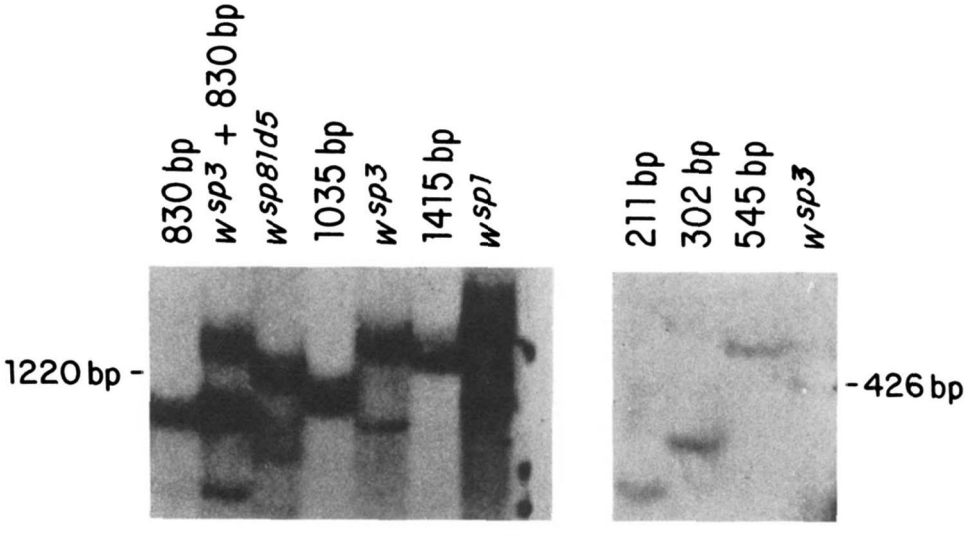

FIGURE 5.—S_1 protection analysis of w^{sp3} and w^{sp81d5} breakpoints. The approximate positions of the left-hand extremes of these deletions were determined previously by Southern gel analysis (ZACHAR and BINGHAM 1982; D. DAVISON, C. CHAPMAN and P. BINGHAM, unpublished observations). Shown here are the results of genomic S_1 protection experiments (MATERIALS AND METHODS) to refine the positioning of these breakpoints. The first seven channels show protection products corresponding to the distance from the white *Hin*dIII cleavage site at coordinate +588 (see Figure 2 for coordinate system; this *Hin*dIII site is precisely at the left-hand extreme of the sequence interval diagramed in Figure 1) to several synthetic deletion breakpoints (sizes of fragments indicated) or to the w^{sp3} and w^{sp81d5} breakpoints. The estimated distance to the w^{sp81d5} deletion break is indicated as 1220 bp, corresponding to its position indicated in Figures 1 and 2. The right-hand four panels show protection products measuring the distance from the *Sph*I cleavage site (at coordinate −473; Figure 2) to the *Bam*HI (211 bp), *Pst*I (302 bp) and *Bgl*II (545 bp) cleavage sites and to the w^{sp3} breakpoint. The estimated distance from *Sph*I to the w^{sp3} break is indicated at 426 bp, corresponding to its position diagramed in Figures 1 and 2. The M13 DNA protector for the left-hand panel was the *Hin*dIII + *Eco*RI fragment, extending from coordinate +588 to coordinate ca. −2900, and for the right-hand panel was the *Sph*I + *Eco*RI fragment, extending from coordinate −473 to coordinate ca. −2900 (see Figure 2 and text for coordinate system).

The w^{sp81d5} mutation arose after EMS mutagenesis and produces an eye color phenotype very much more darkly pigmented than the w^{sp} phenotype (M. M. GREEN, personal communication). We find that this mutation behaves as a spotted allele by conventional genetic tests (GREEN 1959): It produces a spotted-like eye color when heterozygous with w^{sp1} and it is a dominant suppressor of the z^1 mutation allele at zeste.

We find that w^{sp81d5} is a deficiency mutation whose deleted segment begins between coordinates −670 and −590 and extends for an uncharacterized distance beyond a point 5 kb 5′ to the white transcription unit (Figures 2 and 5 and results not shown).

It is convenient to define two generic allele designations. The first is w^{sp} mutations or w^{sp} alleles referring exclusively to w^{sp1}, w^{sp2}, w^{sp3} and w^{sp4}. (We will refer to the presumptive genetic element inactivated by the w^{sp} mutations as the w^{sp} genetic element.) The second is w^{sp} region mutations referring, collectively, to the w^{sp} mutations and the w^{sp81d5} mutation.

The sequence of the w^{sp} *genetic element suggests that it is enhancer-like*: The apparent stage or tissue specificity of w^{sp} mutant effects on white expression (LINDSLEY and GRELL 1968 and references therein; introduction) and the position of the w^{sp} genetic element outside of the major white transcription unit (Figure 1; introduction) suggests that the element exerts positive effects on white expression from ca. 1000 bp outside the transcription unit. These properties are suggestive of elements similar to the enhancers identified in vertebrate systems (WEIHER, KONIG and GRUSS 1983; WEIHER and BOTCHAN 1984 and references therein).

We have examined the sequenced interval containing the w^{sp} genetic element (−1369 to −473) for similarities to the following portions of well-characterized vertebrate enhancers: base pairs 5021 through 5291 on the polyoma map (TOOZE 1981), base pairs 179 through 270 on the SV40 map (TOOZE, 1981) and base pairs 4391 through 4454 on the BPV map (WEIHER and BOTCHAN 1984). These studies reveal statistically significant similarities to the functional portions of the BPV and SV40 enhancers in the w^{sp} region (Figure 8). Moreover, these regions of similarity are partially coextensive with one another and map to the small white locus region containing the w^{sp} genetic element (−1215 to −1240; Figures 1 and 8). These results are suggestive and we propose that the w^{sp} element is functionally related to these papovavirus enhancers.

The w^{sp81d5} *deletion defines one of the sites of suppressor-of-white-spotted locus action on the white locus*: white transcript levels in mature adult tissues are substantially reduced in the presence of the wild-type allele at a second locus, suppressor-of-white-spotted [$su(w^{sp})$] (CHAPMAN and BINGHAM 1985; Figures 6 and 7). On the basis of these results, these authors speculated that the wild-type $su(w^{sp})$ locus causes the production of a repressor of white transcription. [For simplicity, we will refer to this hypothetical repressor as the $su(w^{sp})$ repressor.] On this hypothesis, the suppressor mutation inactivates this repressor leading to the observed elevated white transcript levels.

If this hypothesis is correct, it should be possible to mimic the effects of mutant alleles at $su(w^{sp})$ on expression of w^{sp} alleles by removing the white locus site on which the $su(w^{sp})$ repressor acts. With this prediction in mind we

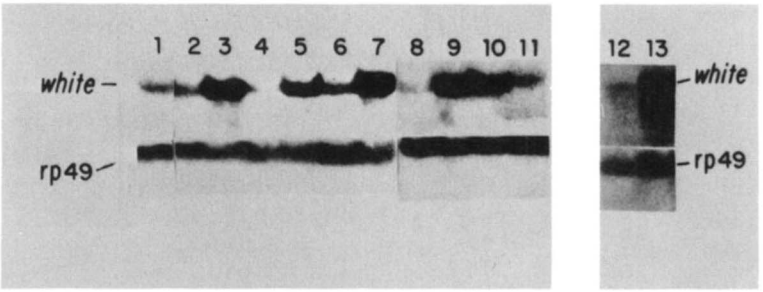

FIGURE 6.—Effects of *su*(w^{sp}) mutation on white transcript levels in mature adult head tissues. Shown are the results of Northern gel analysis of polyadenylated RNAs (10 μg/channel) isolated from purified adult head tisues. The genotypes analyzed are as follows by channel number: 1, w^+; 2, w^{sp1}; 3, *su*(w^{sp}) w^{sp1}; 4, w^{sp2}; 5, *su*(w^{sp}) w^{sp2}; 6 and 8, w^{sp3}; 7 and 9, *su*(w^{sp}) w^{sp3}; 10, *su*(w^{sp}) w^{a2}; 11, w^{a2}; 12, w^{sp81d5}; 13, *su*(w^{sp}) w^{sp81d5}. The figure shown was generated by cutting a Northern transfer in half, probing the top half with white sequences and the bottom half with *rp49* sequences. The white locus probe is a mixture of fragments extending from coordinates −1.2 to −0.4 kg and from +3.2 to +6.5 kb on the standard white locus map. The *rp49* probe is the 0.6-kb *Eco*RI + *Hin*dIII fragment containing most of the *rp49* transcription unit (O'CONNELL and ROSBASH 1984). The ratio of *rp49* to white transcripts is generally reproducible for any tissue and genotype, and *rp49* RNA constitutes an internal control for the amounts of polyadenylated RNA loaded into each channel. The w^{a2} allele produces a transcript indistinguishable in size, hybridization pattern and levels from w^+ alleles. white transcript levels in whole adults and in adult body tissues of w^{sp1-3} and w^{a2} show an essentially identical pattern of response to *su*(w^{sp}); white transcript levels in these other tissues are likewise unaffected by the w^{sp} mutations (results not shown). We note that the small depression in white transcript levels seen in channels 4 and 8 are within the fluctuations normally seen and are not reproducible (for example, compare channels 6 and 8 which both contain w^{sp3} RNA). Also, note that the w^{sp} alleles produce a wild-type transcript in normal amounts in mature adult tissues and w^{a2} produces a wild-type white regulatory region; we presume, therefore, that mature adult transcript levels produced by w^+ alleles will respond to the allelic state of *su*(w^{sp}) in the same way as do w^{sp} and w^{a2} alleles.

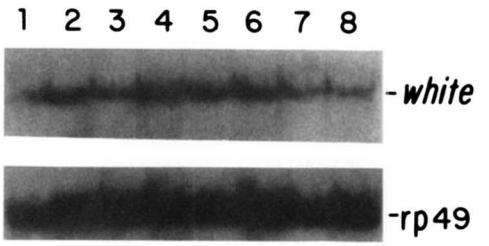

FIGURE 7.—w^{sp} and *su*(w^{sp}) mutations do not affect white transcript levels in larval tissues. Shown is a Northern gel analysis of white transcript levels in combined late second through midthird instar larvae. The analysis was done as described in the legend to Figure 6. The genotypes analyzed are as follows by channel: 1, w^{sp81d5}; 2, *su*(w^{sp}) w^{sp81d5}; 3, w^{sp2}; 4, *su*(w^{sp}) w^{sp2}; 5, w^{sp3}; 6, *su*(w^{sp}) w^{sp3}; 7, w^{a2}; 8, *su*(w^{sp}) w^{a2}. The transcript levels produced by all of the alleles shown are indistinguishable from those produced by w^+ alleles (PIRROTTA and BROCKL 1984; Z. ZACHAR and P. BINGHAM, unpublished observations) and the transcripts produced by the w^{sp} alleles are structurally wild type; on this basis we presume that larval transcript levels produced by the w^+ allele will be likewise nonresponsive to the allelic state of *su*(w^{sp}).

noted that the eye color phenotypes produced by the $su(w^{sp})^1w^{sp1\text{-}3}$ double mutant combinations were quite strikingly similar to those produced by the w^{sp81d5} mutation alone. These eye color phenotypes are both readily distinguishable from either the wild-type phenotype or any white mutant phenotype we have examined. This observation and the structures of the w^{sp} region mutations suggested to us that the w^{sp81d5} deficiency (which deletes the w^{sp} genetic element) might also delete the site of action of the presumptive $su(w^{sp})$ repressor.

This hypothesis, in turn, is directly testable. First, it predicts that the eye color phenotype produced by w^{sp81d5} should not be altered by replacing the wild-type allele at $su(w^{sp})$ with a mutant allele, in spite of the fact that the w^{sp81d5} deletion removes the w^{sp} genetic element. This is shown to be the case by the following results: w^{sp81d5}/w^{def} (w^{def} is a white deficiency allele deleting sequences from coordinates +10 to −15 kb on the standard white locus map; Z. ZACHAR and P. BINGHAM, unpublished observations) heterozygous females have much less eye pigment than w^{sp81d5}/w^{sp81d5} homozygous females or w^{sp81d5}/Y hemizygous males. This demonstrates that the w^{sp81d5} eye color phenotype is sensitive to two-fold differences in dosage of the w^{sp81d5} allele or, equivalently, that a two-fold increase in w^{sp81d5} expression in eye pigment cells of w^{sp81d5}/w^{def} females would be readily detected. $su(w^{sp})$ has been shown previously to produce an increase of at least two-fold on eye pigment produced by the w^{sp} alleles (CHAPMAN, ZACHAR and BINGHAM 1985), and w^{sp81d5}/w^{def} individuals show an eye color phenotype quite similar to w^{sp}/w^{sp} individuals. Collectively, these results demonstrate that effects of $su(w^{sp})$ on w^{sp81d5} similar in magnitude to its effects on $w^{spl\text{-}3}$ should be readily discernible in w^{sp81d5}/w^{def} individuals. Such an effect is demonstrated not to occur by the fact that w^{sp81d5}/w^{def} of all of the following genotypes are indistinguishable from one another: $su(w^{sp})^+/su(w^{sp})^+$, $su(w^{sp})^+/su(w^{sp})$ and $su(w^{sp})/su(w^{sp})$. (Notive also that these experiments demonstrate that the w^{sp81d5} allele dosage compensates.)

The question arises as to whether $su(w^{sp})$ might increase w^{sp81d5} eye pigmentation by the same absolute amount (rather than by the same factor) as it does in the case of the w^{sp} alleles without this increase being discernible in the experiment described in the preceding paragraph. This possibility must be considered because w^{sp81d5} produces a higher level of eye pigmentation than do $w^{sp1\text{-}3}$. The following experiment demonstrates that this is unlikely: $su(w^{sp})$ $w^{sp81d5}/su(w^{sp})$ w^{sp1} females have noticeably more pigmented eyes than do the corresponding homozygous $su(w^{sp})^+$ females. This increased pigmentation presumably results from $su(w^{sp})$ action on w^{sp1} in this heterozygote.

The nonresponsiveness of w^{sp81d5} to $su(w^{sp})$ is further corroborated by our observation that the eye color phenotype of w^{sp81d5} homozygous females is unaffected by the $su(w^{sp})$ mutation in either heterozygous or homozygous condition.

These experiments demonstrate that the eye color phenotype produced by w^{sp81d5} does not respond to $su(w^{sp})$, in contrast to all of the other w^{sp} region mutations. This and the eye color phenotype of w^{sp81d5} are strong evidence that it deletes the white locus element necessary to allow response to the $su(w^{sp})$ locus during the deposition of adult eye pigment in pupal development.

On the one hand, deposition of eye pigment occurs during pupal develop-

ment. On the other hand, white transcripts detected in adult heads may or may not be synthesized during those times or in those tissues where eye pigment is being deposited. We, thus, wished to ask whether the mature adult transcript levels produced by the w^{sp81d5} allele would be nonresponsive to $su(w^{sp})$ as is eye pigment deposition. The results in Figure 6 show that w^{sp81d5} transcript levels respond to $su(w^{sp})$ precisely as do other tested white alleles.

Collectively, our results suggest that the $su(w^{sp})$ locus acts on one white locus site during eye pigment deposition and a second during deposition of white transcripts in mature adult tissues. We note that studies of the w^{DZL} allele at white also suggest that the eye pigment and head transcript phenotypes are separately controlled (BINGHAM and ZACHAR 1985). We will refer to the white locus element allowing $su(w^{sp})$ responsiveness during pigment deposition as the $su(w^{sp})$ genetic element. This element resides partially or entirely between the left-hand breakpoints of the w^{sp3} and w^{sp81d5} deletions. The observation that both w^{sp3} and w^{sp81d5} deletions extend for substantial distances rightward (5′) suggests that the white locus element allowing responsiveness of mature adult white transcript levels to $su(w^{sp})$ maps leftward (3′) to the w^{sp81d5} deletion.

We note that our analysis of the sequence from −1024 to −473 revealed the existence of two small regions with extensive homology to one another (MATERIALS AND METHODS). These regions extend from −895 to −815 and from −964 to −896 and are ca. 62% homologous to one another. By the measure described in MATERIALS AND METHODS a match of this quality is expected to occur in approximately 0.02% of randomly chosen comparable sequences. Moreover, 24 Monte Carlo scramblings of the sequence preserving nearest neighbor frequencies did not result in comparably extensive sequence repetitions. These observations allow the hypothesis that this imperfect, direct repetition is maintained by selection, although alternatives to this supposition are not excluded.

w^{sp} *region mutations and the* su(w^{sp}) *mutation have no effects on white transcript levels in larvae*: Neither the w^{sp1} insertion mutation (PIRROTTA and BROCKL 1984; LEVIS, O'HARE and RUBIN, 1984; CHAPMAN, ZACHAR and BINGHAM 1985; Figure 6) nor the w^{sp} region deletion mutations (CHAPMAN, ZACHAR and BINGHAM 1985; Figure 6) have a discernable effect on white transcript levels in mature adult tissues. Moreover, it has also been observed previously that tne $su(w^{sp})$ mutation causes a substantial elevation in white transcript levels in mature adult tissues (CHAPMAN and BINGHAM 1985). We, therefore, wished to ask whether these mutations affected larval white transcript levels. Figure 7 shows that w^{sp2}, w^{sp3} and w^{sp81d5} all fail to have detectable effects on accumulation of white transcripts at the larval stage examined. Moreover, the $su(w^{sp})$ mutation has no effect on larval white transcript levels in any cases that we have examined. This last observation suggests that the $su(w^{sp})$ locus may act exclusively in regulating white expression in adults.

On the nature of spontaneous and X-ray-induced deletion formation: The w^{sp2} deficiency probably arose spontaneously in the recombination-proficient female germline (MOHLER 1956). Our sequence data show that the segment deleted in this event is bounded by a TGTGTG segment on one side and a single TG

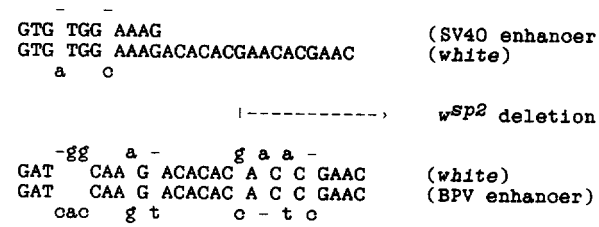

FIGURE 8.—Similarities between the w^{sp} region and the SV40 and BPV enhancer regions. Sequence matches shown are between the w^{sp} region of the Canton-S w^+ allele (coordinates −1248 to −1221; Figure 2) and the SV40 enhancer (top portion) or the BPV enhancer (bottom portion). (See text for detailed discussion of these sequence matches). These regions of similarity overlap. This region of overlap abuts the small segment deleted in the w^{sp2} mutation and is deleted in the w^{sp3} and w^{sp4} alleles (Figures 1 and 2). Moreover, this region is 5′ to the w^{sp1} insertion and is thus separated from the white transcription unit by this insertion (Figures 1 and 2). One expects to encounter matches of comparable quality between randomly chosen, comparable sequences at frequencies of approximately 0.012 (SV40) and 0.048 (BPV), respectively (MATERIALS AND METHODS). Note that the similarities of these enhancer sequences to the Oregon-R w^+ allele are less extensive than those to the Canton-S w^+ allele (compare Figures 3 and 8).

on the other. These are combined to form a TGTGTGTG segment in the deletion allele (Figure 2). Evidence exists that such poly-TG segments participate in some recombination events (PROUDFOOT and MANIATIS 1980; SLIGHTOM, BLECHL and SMITHIES 1980; STRINGER 1982; D. TRECO and N. ARNHEIM, personal communication).

w^{sp4} is an X-ray-induced deficiency, and our sequence data show the presence of 27 bp of additional sequence (not present in any other white allele we have examined) at the deletion breakpoint (Figure 4). This additional 27 bp consists of three copies of the sequence PCCAYTTT, where P is a purine and Y is pyrimidine. Two copies of this sequence are present in the wild-type allele and precisely abut one of the deletion breakpoints (Figure 4). This unusual structure suggests that the repair of X-ray-induced damage may be unexpectedly complex.

GC content of the w^{sp} *region*: Our DNA sequence analysis of the w^{sp} region shows the existence of very large variation in GC content (Figures 1, 2, 4, 8 and 9). Sequence analysis of several other Drosophila genes (O'CONNELL and ROSBASH 1984; SNYDER *et al.* 1982; BENYAJATI *et al.* 1983; MUSKAVITCH and HOGNESS 1982; MCGINNIS *et al.* 1983) suggests that relatively high GC content correlates with regions whose sequences are under selective constraint.

DISCUSSION

A refined model for regulation of white transcription in adult tissues: Our results suggest that the w^{sp} region is internally complex. It apparently contains at least two distinct genetic elements: First, mapping partially or entirely in the interval between positions −1231 to −1115 is the w^{sp} genetic element. All analyzed mutants eliminating only this region drastically reduce white-dependent eye pigmentation. Second, mapping partially or entirely between positions −590 and −925 is a presumptive genetic element [$su(w^{sp})$p element] necessary to

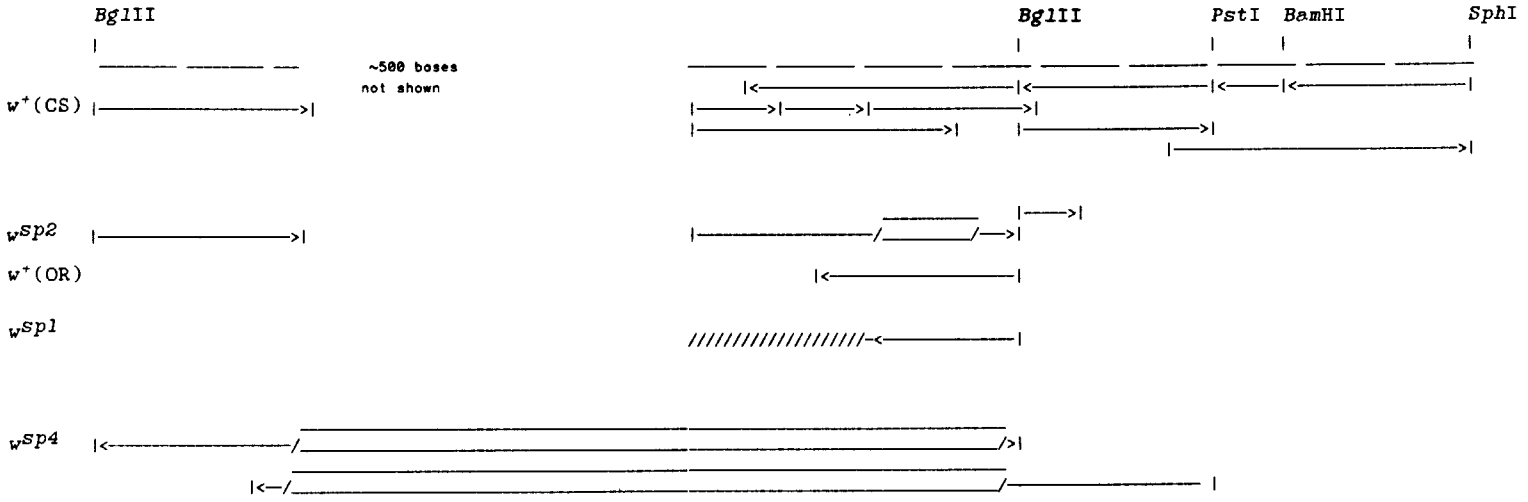

FIGURE 9.—Sequencing strategies for various white alleles. The top line represents w^{sp} region DNA sequences oriented so that left to right corresponds to the 5′ to 3′ direction for the sequence shown in Figures 2–4 and 8. The *Sph*I, *Bam*HI and *Pst*I sites occur at coordinated −473, −682, and −774, respectively (Figure 2). The rightmost *Bgl*II site occurs at coordinate −1019 and the leftmost *Bgl*II site at approximately coordinate −2200. [Additional white locus restriction map information can be found in ZACHAR and BINGHAM (1982) and LEVIS, BINGHAM and RUBIN (1982).] The line is broken every 100 bp and is broken at a segment of ca. 500 bp which we have not analyzed in detail. The arrows below this top line show the direction and extent of individual sequencing runs (MATERIALS AND METHODS) yielding the sequence in Figures 2–4. The open boxes are the sequences deleted in the w^{sp2} (middle panel) and w^{sp4} (lower panel) mutations. In the case of the w^{sp4} allele the *Bgl*II fragment containing the deleted segment was sequenced in both orientations and an additional sequencing run was done extending from the *Pst*I site (−774) across the w^{sp4} deletion. The hashed box is the B104 transposon insertion responsible for the w^{sp1} mutation. All portions of the wild-type sequence in Figure 2 were determined in both directions, and all restriction sites internal to the wild-type sequence were crossed by at least one sequencing run.

allow responsiveness of w^{sp} allele-dependent eye pigment deposition to the allelic state of the $su(w^{sp})$ locus.

Third, our studies suggest a genetic element allowing the accumulation of white transcripts in mature adult tissues to be responsive to the allelic state of the $su(w^{sp})$ locus. This element is not deleted by any of the mutations analyzed, including w^{sp81d5}, and, thus, presumably maps 3′ to position −670.

We propose the following model for the regulation of white expression in adult tissues on the basis of our studies: A repressor of white transcription is produced in mature and immature adult tissues. Production of this repressor is dependent on a wild-type allele at the suppressor-of-white-spotted locus. This repressor acts at a white locus site during eye pigment deposition (developing, immature adult tissue) different from during deposition of white transcripts in mature adult tissues. In inmature eye pigment cells, the effect of this repressor is entirely antagonized by the action of the enhancer-like w^{sp} genetic element. Genetic evidence (GREEN 1959; BABU and BHAT 1980) suggests that the product of the zeste locus participates in this action of the w^{sp} genetic element. The enhancer-like w^{sp} element has no discernible effect on white transcript levels in mature adult tissues.

We also note that the w^{sp81d5} allele dosage compensates. This observation, together with earlier observations (GREEN 1959; JUDD 1974; LEVIS, BINGHAM and RUBIN 1982; ZACHAR and BINGHAM 1982; HAZELRIGG, LEVIS and RUBIN 1984), demonstrates that the element necessary to allow dosage compensation in adult eye pigment cells maps between coordinates −3.0 and +4.1 kb on the white locus map.

Synapsis-dependent interactions and cis*-acting regulatory elements*: We wish to emphasize an important implication of the proposal that the w^{sp} element is an enhancer (RESULTS and DISCUSSIONS above): On the one hand, the w^{sp} element participates in the special, genetically defined interactions that occur between two white alleles when the two alleles pair or synapse (GREEN 1959; JACK and JUDD 1979; BINGHAM 1980; BABU and BHAT 1980). The mechanistic basis of this class of interactions is still obscure, but such interactions have been clearly recognized at several loci in Drosophila and may be relatively common (LEWIS 1954; GELBART and WU 1982; BINGHAM and ZACHAR 1985). On the other hand, the wide occurrence of enhancer-like elements in vertebrates is well established, but the mechanisms whereby these elements act from distances of hundreds to thousands of base pairs to stimulate transcript accumulation is likewise not established. The proposal that the w^{sp} element is an enhancer suggests an insight into both of these unresolved problems.

This insight results from the fact that one mechanism previously proposed for synapsis-dependent interactions in Drosophila is quite similar to a mechanism previously proposed for the action of enhancers. On the one hand, it has been proposed (BINGHAM and ZACHAR 1985) that synapsis-dependent interactions reflect the propensity of synapsed alleles to behave as a unit in choosing nuclear positions or compartments relevant to the expression of those alleles. Although the evidence for this hypothesis is circumstantial, it is suggestive. On the other hand, it has been proposed that enhancer-like *cis*-acting regulatory

elements might act by controlling nuclear placement of associated genes [see MIRKOVITCH, MIRAULT and LAEMMLI (1984) for a recent discussion], although the evidence for this proposal is likewise circumstantial. We suggest that our observation of possible identity between an element participating in synapsis-dependent interactions and a presumptive enhancer supports each of these very similar mechanistic hypotheses.

We are grateful to W. GELBART for making *su*(w^{sp}) available to us, to M. M. GREEN for making w^{sp81d5} available to us and to Z. ZACHAR for criticisms of a draft of this manuscript. We would like to thank W. FITCH for supplying a copy of his metric alignment program and WALTER GOAD and MINORU KANEHISA for discussions and for providing some sequences for the Los Alamos sequence data bank. This work was supported by a United States Public Health Service grant to P.M.B.

LITERATURE CITED

BABU, P. and S. G. BHAT, 1980 Effect of *zeste* on *white* complementation. pp. 35–44. In: *Development and Neurobiology of Drosophila*, Edited by O. SIDDIQI, P. BABU, L. M. HALL, and J. C. HALL. Plenum Press, New York.

BABU, P. and S. G. BHAT, 1981 Role of *zeste* in transvection at the *bithorax* locus of Drosophila. Mol. Gen. Genet. **183:** 400–402.

BARNES, W. M., M. BEVAN and P. H. SON, 1983 Kilo-sequencing: creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13. Methods Enzymol. **110:** 343–380.

BENYAJATI, C., N. SPOEREL, H. HAYMERLE and M. ASHBURNER, 1983 The mRNA for alcohol dehydrogenase in Drosophila melanogaster differs in its 5′ end in different development stages. Cell **33:** 125–133.

BINGHAM, P. M., 1980 The regulation of white locus expression: a dominant mutant allele at the white locus of *Drosophila melanogaster*. Genetics **95:** 341–353.

BINGHAM, P. M., R. LEVIS, and G. M. RUBIN, 1981 Cloning of DNA sequences from the *white* locus of D. melanogaster by a novel and general method. Cell **25:** 693–704.

BINGHAM, P. M. and Z. ZACHAR, 1985 Evidence that two mutations, w^{DZL} and z^1, affecting the synapsis-dependent genetic behavior of *white* are transcriptional regulatory mutations. Cell **40:** 819–825.

CHAPMAN, C. H. and P. M. BINGHAM, 1985 Evidence that the locus of a novel type of suppressor mutation regulates *white* locus transcription. Drosophila Inform. Serv. In press.

CONRAD, B. and D. W. MOUNT, 1982 Microcomputer programs for DNA sequence analysis. Nucleic Acids Res. **10:** 31–38.

FITCH, W. and T. F. SMITH, 1983 Optimal sequence alignments. Proc. Natl. Acad. Sci. USA **80:** 1382–1386.

GANS, M., 1983 Etude genetique et physiologique du mutant z de Drosophila melanogaster. Bull Biol. F. Belg. **38U:** (Suppl.) 1–19.

GELBART, W. M. and C.-T. WU, 1982 Interactions of zeste mutations with loci exhibiting transvection effects in *Drosophila melanogaster*. Genetics **102:** 179–189.

GOAD, W. and M. KANEHISA, 1982 Pattern recognition in nucleic acid sequencies. I. A general method for finding local homologies and symmetries. Nucleic Acids Res. **10:** 247–263.

GREEN, M. M., 1959 Spatial and functional properties of pseudoalleles at the *white* locus in Drosophila melanogaster. Heredity **13:** 302–315.

HAZELRIGG, T., R. LEVIS and G. M. RUBIN, 1984 Transformation of *white* locus DNA in Drosophila: dosage compensation, zeste interaction and position effects. Cell **36:** 469–481.

HU, N. and J. MESSING, 1982 The making of strand-specific M13 probes. Gene **17:** 271–277.

JACK, J. W., and B. H. JUDD, 1979 Allelic pairing and gene regulation: a model for the zeste-white interaction in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **76:** 1368–1372.

JUDD, B. H., 1974 Genes and chromosomes of Drosophila. pp. 169–184. In: *The Eukaryotic Chromosome*, Edited by W. J. PEACOCK and R. D. BROCK. Australian National University Press, Canberra.

KANEHISA, M., 1984 The use of statistical criteria for screening potential homologies in nucleic acid sequences. Nucleic Acids Res. **12:** 203–213.

KAUFMAN, T. C., S. E. TASAKA and D. T. SUZUKI, 1973 The interaction two complex loci, zeste and bithorax, in *Drosophila melanogaster*. Genetics **75:** 299–321.

LEVIS, R., P. M. BINGHAM and G. M. RUBIN, 1982 Physical map of the *white* locus of Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **79:** 564–568.

LEVIS, R., K. O'HARE and G. M. RUBIN, 1984 Effects of transposable element insertions on RNA encoded by the *white* gene of Drosophila. Cell **38:** 471–481.

LEWIS, E. B., 1954 The theory and applications of a new method of detecting chromosomal rearrangement in Drosophila melanogaster. Am. Nat. **88:** 225–239.

LINDSLEY, D. L. and E. H. GRELL, 1968 Genetic variations of *Drosophila melanogaster*. Carnegie Inst. Wash. Publ. 627.

LIPMAN, D. J., W. J. WILBUR, T. F. SMITH and M. S. WATERMAN, 1984 On the statistical significance of nucleic acid similarities. Nucleic Acids Res. **12:** 215–266.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

MCGINNIS, W., A. SHERMOEN, J. HAEMSKERK and S. BECKENDORF, 1983 DNA sequence changes in an upstream DNase-I hypersensitive region correlates with reduced gene expression. Proc. Natl. Acad. Sci. USA **80:** 1063–1067.

MIROKOVITCH, J., M.-E. MIRAULT and U. K. LAEMMLI, 1984 Organization of the higher-order chromatin loop: specific DNA attachment sites one nuclear scaffold. Cell **39:** 223–232.

MOHLER, J. D., 1956 New mutants: white-spotted^2. Drosophila Inform. Serv. **30:** 78.

MUSKAVITCH, M. A. T. and D. S. HOGNESS, 1982 An expandable glue gene that encodes a Drosophila glue protein is not expressed in variants lacking remote upstream sequences. Cell **29:** 1041–1051.

NEEDLEMAN, S. B. and C. D. WUNSCH, 1970 A general method applicable to the search for similarities in the amino-acid sequence of two proteins. J. Mol. Biol. **42:** 245–261.

O'CONNELL, P. and M. ROSBASH, 1984 Sequence, structure and codon preference of the Drosophila ribosomal protein 49 gene. Nucleic Acids Res. **12:** 5495–5513.

O'HARE, K., R. LEVIS and G. M. RUBIN, 1983 Transcription of the *white* locus in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **80:** 6917–6921.

PIRROTTA, V. and CH. BROCKL, 1984 Transcription of the Drosophila *white* locus and some of its mutants. EMBO J. **3:** 563–568.

PROUDFOOT, N. J. and T. MANIATIS, 1980 The structure of a human alpha-globin pseudogene and its relationship to alpha-globin gene duplication. Cell **21:** 537–544.

PUSTELL, J. and F. KAFATOS, 1984 A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis, and homology determination. Nucleic Acids Res. **12:** 643–655.

SANGER, F. and A. R. COULSON, 1978 The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87:** 107–110.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHERER, G., C. TSCHUDI, J. PERERA, H. DELIUS and V. PIRROTTA, 1982 *B104*, a new dispersed repeated gene family in Drosophila melanogaster and its analogies with retroviruses. J. Mol. Biol. **157:** 435–451.

SELLERS, P., 1974 The theory and computation of evolutionary distances: pattern recognition. J. Algorithms **1:** 359–373.

SLIGHTOM, J. L., A. E. BLECHL and O. SMITHIES, 1980 Human fetal G-gamma and A-gamma globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21:** 627–638.

SNYDER, M., M. HUNKAPILLER, D. YUEN, D. SILVER, J. FISTROM and N. DAVIDSON, 1982 Cuticle protein genes of Drosophila: structure, organization and evolution of four clustered genes. Cell **29:** 1027–1040.

STRINGER, J. R., 1982 DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. Nature **296:** 363–366.

TOOZE, J., 1981 *DNA Tumor Viruses, Ed. 2.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

WEIHER, H. and M. R. BOTCHAN, 1984 An enhancer sequence from bovine papilloma virus DNA consists of two essential regions. Nucleic Acids Res. **12:** 2901–2916.

WEIHER, H., M. KONIG and P. GRUSS, 1983 Multiple point mutations affecting the simian virus 40 enhancer. Science **219:** 626–631.

ZACHAR, Z. and P. M. BINGHAM, 1982 Regulation of *white* locus expression: the structure of mutant alleles at the *white* locus of Drosophila melanogaster. Cell **30:** 529–541.

Communicating editor: W. M. GELBART