MOLECULAR AND CELLULAR BIOLOGY, Feb. 1988, p. 615–623
0270-7306/88/020615-09$02.00/0

Vol. 8, No. 2

DNA-Binding Properties of the *Drosophila melanogaster* zeste Gene Product

ALKA MANSUKHANI, ANJA CRICKMORE, PETER W. SHERWOOD, AND MICHAEL L. GOLDBERG*
Section of Genetics and Development, Cornell University, Ithaca, New York 14853

Received 24 August 1987/Accepted 4 November 1987

The ability of the zeste moiety of β-galactosidase-zeste fusion proteins synthesized in *Escherichia coli* to bind specific DNA sequences was examined. Such fusion proteins recognize a region of the white locus upstream of the start of transcription; this region has previously been shown to be required for genetic interaction between the zeste and white loci. Another strong binding site was localized to a region between 50 and 205 nucleotides before the start of the *Ubx* transcriptional unit; expression of the bithorax complex is also known to be influenced by the zeste locus. Weaker binding sites were also seen in the vicinity of the *bxd* and *Sgs-4* genes, but it is currently unclear whether these binding sites play a role in transvection effects. The DNA-binding activity of the zeste protein is restricted to a domain of approximately 90 amino acids near the N terminus. This domain does not appear to contain homeobox or zinc finger motifs found in other DNA-binding proteins. The DNA-binding domain is not disrupted by any currently characterized zeste mutations.

The zeste locus (*z*; 1 - 1.0) of *Drosophila melanogaster* appears to mediate a class of genetic phenomena termed transvection effects. Transvection refers to alterations in the phenotypic expression of a variety of genes when chromosomal rearrangements are introduced which disrupt synapsis between the two alleles of these genes in the diploid genome. For example, certain heterozygous combinations of mutant alleles at the bithorax complex (*Bx-C*; 3 - 58.8) display partial phenotypic complementation. This complementation is abolished by either of two conditions: (i) the introduction of particular rearrangements which disrupt pairing of these bithorax complex mutations, or (ii) the presence of certain mutant alleles of the zeste locus (10, 25, 30). Similar effects are seen at the gene decapentaplegic (*dpp*; 2 - 4.0); zeste locus mutations and rearrangements preventing pairing of *dpp* alleles both influence observed phenotypes (10, 11). The interaction of the z^1 allele with the white locus (*w*; 1 - 1.5) provides a third example of transvection phenomena. The z^1 mutation produces abnormal yellow eye color only in the presence of two or more copies of the wild-type white locus, and only when all w^+ genes are juxtaposed in the nucleus by virtue of their tandem duplication or synapsis (8, 11, 13, 14, 21, 23).

Several recent observations suggest that transvection effects are associated with the transcriptional regulation of target genes. First, as originally noted by Jack and Judd (21), the interplay between zeste and white is likely to occur in the nucleus, as it is difficult to envision how chromosomal synapsis could modulate gene expression in the cytoplasm. Next, it has recently been found that the DNA sequences at the white locus required for zeste interaction are located between 1.1 and 1.8 kilobases (kb) upstream of the apparent transcriptional initiation site (29, 40). The activity of the z^1 allele must therefore influence events before or coincident with white locus transcription. This view is substantiated by the finding that steady-state levels of white mRNA are markedly decreased in the heads of flies displaying the zeste phenotype (3; unpublished data).

The zeste gene product can thus be thought to act as a tissue-specific regulator of the transcription of white and other target loci such as the bithorax complex and decapentaplegic. However, the above experiments do not indicate whether the observed effects of zeste mutations on gene expression are the result of a direct interaction between the zeste product and DNA sequences in the vicinity of target loci or an interaction mediated by other, as yet undefined molecules. In this study we attempted to distinguish between these alternatives by examining the hypothesis that the protein encoded by the zeste locus shares one functional property with well-characterized transcriptional factors in procaryotic and eucaryotic organisms: the ability to bind specific DNA sequences at target genes. We demonstrated that a β-galactosidase-zeste fusion protein synthesized in *Escherichia coli* cells has specific affinity for those DNA sequences at the white locus that have been shown genetically to be required for interaction with zeste. Binding sites in the vicinity of sequences encoding the 5′ end of the *Ubx* and *bxd* transcripts in the bithorax complex were also observed. Finally, we showed that a limited domain of the zeste protein consisting of approximately 90 amino acids near the N terminus is sufficient for this DNA-binding activity.

MATERIALS AND METHODS

Plasmids and DNA techniques. *lacZ*-zeste fusion plasmids were constructed by standard techniques (33). We have previously described the isolation of a 2.4-kb *Eco*RI fragment containing a near full-length zeste cDNA sequence (15; A. Mansukhani, P. H. Gunaratne, P. W. Sherwood, B. J. Sneath, and M. L. Goldberg, Mol. Gen. Genet., in press). This fragment was subcloned into the *Eco*RI site of the vector Bluescribe BBM− (Stratagene). Deletions progressively removing sequences encoding the 5′-untranslated leader of the zeste mRNA and amino acids from the N terminus of the zeste protein were generated as described by Henikoff (18), utilizing the *Bam*HI and *Pst*I sites in the polylinker of the vector. The location of the deletion endpoint was determined by direct DNA sequencing of the supercoiled DNA of each deletion construct (5, 43). For each N-terminal deletion employed, the *Hin*dIII site remaining in the polylinker was converted into an *Eco*RI site utilizing adaptor oligonucleotides (New England Biolabs, Inc., Bev-

* Corresponding author.

erly, Mass.). The resultant *Eco*RI fragments were cloned into either vector pWR590 (for the 36C and 178C constructs) or pWR590-1 (for the NC, 47C, and 327C constructs) (16), such that the zeste sequences would be in the same frame as the truncated *lacZ* gene contained on these vectors. Direct DNA sequencing verified the identity of nucleotides at the β-galactosidase-zeste junction.

Successive C-terminal deletions of zeste sequences were synthesized starting from the NC construct. The NC plasmid was cleaved in its polylinker with *Bam*HI and *Pst*I, preliminary to exonuclease III and S1 nuclease digestion as described by Henikoff (18). The endpoints of all resultant deletions were determined by direct DNA sequencing with a synthetic oligonucleotide primer derived from the pWR590-1 vector.

Plasmids used as targets in the DNA binding assays were obtained from several sources. A 2.2-kb *Eco*RI-*Bam*HI fragment containing sequences between 0.7 and 2.9 kb upstream of the start of white transcription was originally provided as a subclone in the vector M13mp8 by R. Levis. To simplify interpretation of the binding analysis, this fragment was subcloned into the Bluescript vector KS− (Stratagene), yielding clone pBE2. An M13mp8 *Bam*HI-*Hin*dIII subclone containing sequences between 0.7 kb upstream of the start of white transcription and 0.6 kb into transcribed sequences was also the kind gift of R. Levis. W. Bender and M. Peifer generously provided several cloned DNAs from the bithorax complex. These included bacteriophages L2206, L2212, L2218, and L2229, which contain the region between coordinates −44 and +16 on the map of Bender et al. (1). More precise analysis in the regions encoding the 5′ ends of the *Ubx* and *bxd* transcriptional units was performed on subclones 3101 (coordinates −3 to +2 at *bxd*) and 3107 (coordinates −31 to −27 at *Ubx*) (1). pbx2-10, a bacteriophage lambda clone containing genomic DNA from the pbx^2 mutation, was originally isolated by M. Peifer. A 3.2-kb *Eco*RI subclone of this phage (pbx2-3.2) including the deletion breakpoints was prepared in our laboratory by R. Colvin. L. Prestidge and D. Hogness provided subclones 1553 and 1543 (35), which contain *Hin*dIII fragments including almost the entire *Sgs-4* transcriptional unit as well as approximately 3 kb of upstream sequences from the Canton S and Hikone R strains, respectively.

Protein gel electrophoresis and Western blots (immunoblots). Expression clones were grown in *E. coli* HB101, induced with 1 mM isopropyl-β-D-thiogalactopyranoside at A_{600} = 0.5, and collected 3 h later by centrifugation. Cell pellets from 1 ml of culture were suspended and boiled for 5 min in 50 μl of sodium dodecyl sulfate (SDS) sample buffer (5% 2-mercaptoethanol, 3% SDS, 10% glycerol, 62.5 mM Tris hydrochloride [pH 6.8], 0.1% bromophenol blue). Samples were loaded on 10% polyacrylamide-SDS gels, and proteins were visualized by staining with Coomassie brilliant blue.

Gels utilized for Western blotting (immunoblotting) were transferred immediately after electrophoresis to nitrocellulose filters with a Bio-Rad TransBlot cell according to the instructions of the manufacturer. All subsequent treatments were performed at room temperature. The filter was blocked in TBS buffer (20 mM Tris hydrochloride [pH 7.5], 500 mM NaCl) with 3% gelatin for 2 h. The filter was then transferred into TBS buffer containing 1% gelatin and a 1:2,000 dilution of affinity-purified rabbit anti-β-galactosidase immunoglobulin G (Organon Teknika, Malvern, Pa.). Incubation proceeded for a minimum of 5 h. The filter was then washed twice for 10 min in TBS buffer and then transferred into a solution of TBS buffer containing 1% gelatin and a 1:2,000 dilution of peroxidase-labeled goat anti-rabbit immunoglobulin G (Organon Teknika) for 1 h. The filter was rinsed twice for 10 min in TBS buffer and then immersed in a development solution (20 ml of 0.3% 4-chloro-1-naphthol in methanol plus 100 ml of 0.015% hydrogen peroxide in TBS) for 10 to 30 min.

Preparation of bacterial extracts. Overnight cultures of expression clones in *E. coli* HB101 were inoculated into 100 ml of Luria broth and induced with 1 mM isopropyl-β-D-thiogalactopyranoside at A_{600} = 0.5. Cells were collected 3 h later by centrifugation, and the cell pellet was suspended in 1 ml of sonication buffer (25 mM sodium HEPES [*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid] [pH 7.5], 5 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 10% glycerol, 0.1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride). The suspension was sonicated on ice for 30 s and then centrifuged at 30,000 × *g* for 10 min. The supernatant was stored at −20°C after the addition of glycerol to a final concentration of 30% (vol/vol). The extract could be stored in this manner for several weeks without obvious loss of DNA-binding activity.

Total protein concentration in these extracts was measured by spotting twofold dilutions of the extracts and of a standard solution of bovine serum albumin on Whatman 3MM paper. The filter was dipped first in ice-cold 10% trichloroacetic acid and then in acetone and air dried at room temperature. The filter was stained for 10 min in 50% methanol–10% acetic acid–2.5% Coomassie brilliant blue and destained for 20 min in the same solution without the dye. The average protein concentration in these extracts was 10 to 15 mg/ml. SDS-polyacrylamide gel electrophoresis and Western blotting indicates that approximately 5 to 10% of the fusion protein synthesized in these cells is retained in these soluble extracts (data not shown).

DNA binding assay. DNA-binding reactions were performed essentially by the method of Desplan et al. (7). Target DNAs were cleaved with suitable restriction endonucleases and end labeled with [α-^{32}P]deoxynucleotide triphosphates by using the Klenow fragment of DNA polymerase I. Between 10 and 20 ng of the labeled DNA was incubated with bacterial extracts containing 10 μg of protein in the presence of 2 μg of poly(dI-dC)-(dI-dC) carrier DNA (Pharmacia, Inc., Piscataway, N.J.) in a total volume of 40 μl of binding buffer (100 mM NaCl, 20 mM Tris hydrochloride [pH 7.6], 0.25 mM EDTA, 1 mM dithiothreitol, 10% glycerol). Incubation was for 30 min at 0°C, after which 2 μl of undiluted rabbit antibodies was added. Incubation was continued for an additional 30 min at 0°C. Next, 8 μl of a 10% suspension of fixed *Staphylococcus aureus* cells (Sigma Chemical Co., St. Louis, Mo.) was added, and the reaction was pelleted by centrifugation in a microcentrifuge after an additional 30 min of incubation at 0°C. The pellet was washed three times in 100 μl each time of binding buffer on ice and resuspended in 20 μl of binding buffer. The solution was extracted with phenol-chloroform (1:1, vol/vol), and the aqueous phase containing the DNA was analyzed on polyacrylamide or agarose gels. Approximately 5 to 10% of the input radioactivity of those fragments recognized by the β-galactosidase-zeste fusion protein was recovered in the immunoprecipitate.

All binding reactions were performed with affinity-purified rabbit anti-β-galactosidase immunoglobulin G with the exception of those displayed on lanes g and h on Fig. 2, which employed antisera against the zeste moiety of the fusion protein. These antisera were obtained by immunizing Flem-

ish giant chinchilla rabbits with 75 μg of NC fusion protein prepared by electroelution from a polyacrylamide gel slice. Anti-zeste moiety antibodies were purified by adsorption of the serum with a protein extract from *E. coli* cells harboring the pWR590-1 vector (16); Western blots exposed to this purified antibody showed reaction only with β-galactosidase-zeste fusion proteins, but not with β-galactosidase or other *E. coli* proteins (data not shown).

RESULTS

Binding of a β-galactosidase-zeste fusion protein to specific DNA fragments. At the outset of these experiments, we reasoned that synthesis in *E. coli* cells of a β-galactosidase-zeste fusion protein should provide a convenient material to test the hypothesis that the zeste protein recognizes specific DNA sequences. Johnson and Herskowitz (22) and Desplan et al. (7) have shown that the β-galactosidase moiety of such a fusion protein can be used as an immunological tag to enable identification and characterization of complexes formed between the fusion protein and nucleic acids. To construct such a fusion, we prepared a 2.2-kb *Eco*RI fragment derived from a near full-length zeste cDNA clone as described in Materials and Methods; this fragment includes 16 nucleotides from the zeste mRNA immediately preceding the ATG initiation codon, information encoding the entire zeste protein, and approximately 350 nucleotides from the untranslated 3′ end of the zeste mRNA. This fragment was subcloned into the expression vector pWR590-1 (18), yielding the recombinant plasmid NC, described more fully in Table 1. Plasmid NC directed the synthesis in *E. coli* cells of substantial amounts of a fusion protein of approximately the predicted size (132 kilodaltons) (Fig. 1).

Extracts prepared from cells harboring the NC plasmid were added to mixtures of radiolabeled DNA fragments from a plasmid (pBE2) containing a region of DNA upstream of the transcription start of the *Drosophila* white locus, which had previously been shown to be required for the zeste-white interaction. After incubation, antibody directed against the β-galactosidase moiety of the fusion protein was added, followed by the addition of fixed *S. aureus* cells. The DNA content of pellets obtained after centrifugation was analyzed by gel electrophoresis. Specific fragments of the input DNA were preferentially found in immunoprecipitates obtained in this manner (Fig. 2, lanes a and b). The presence of these bands was dependent on the addition of an extract containing the fusion protein (Fig. 2, lanes c and d). The immunoprecipitation of these DNA fragments required the addition of antibody directed against either the β-galactosidase or zeste region of the NC polypeptide (Fig. 2, lanes f to h). Furthermore, addition of unlabeled pBE2 DNA substantially reduced the amount of radiolabeled DNA detected (Fig. 2, lane e); pBE2 is considerably more efficient as a competitor than unrelated DNA (data not shown). These results demonstrate the occurrence of a specific interaction between certain DNA fragments and the zeste moiety of the fusion protein in the extract. The protein-DNA complex appears to be quite stable, being resistant to salt concentrations of 300 mM, at least under conditions of fusion protein excess (Fig. 2, lanes i to l).

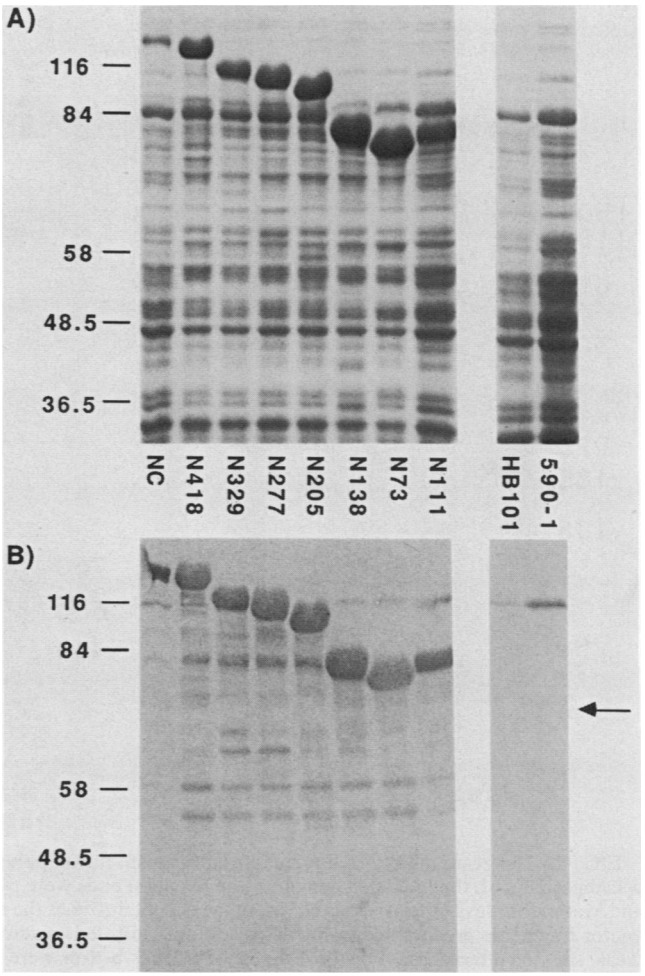

FIG. 1. Fusion proteins synthesized in *E. coli* cells. Expression clones encoding fusion polypeptides containing the first 583 amino acids of β-galactosidase and either a complete zeste protein (NC) or derivatives progressively deleted from the zeste C terminus (N418, N329, N277, N205, N138, and N73, Table 1) were grown in *E. coli* HB101 and induced as described in Materials and Methods. Cell pellets were boiled in SDS sample buffer and divided into two aliquots run on separate 10% SDS-polyacrylamide gels. (A) Staining with Coomassie brilliant blue to reflect total protein content. (B) Western blot of gel run in parallel, reacted with anti-β-galactosidase immunoglobulin G, and stained with peroxidase-labeled second antibody. Fusion proteins are seen as prominent bands in both panels. Endogenous β-galactosidase is observed as the only band on the Western blot of HB101 cells, demonstrating the specificity of the antibody employed. In this particular experiment, the vector pWR590-1 produced limited quantities of truncated β-galactosidase (seen as a faint band marked with an arrow at 66 kilodaltons), while in other experiments, such as those depicted in Fig. 2 and 4, a much stronger signal was observed (not shown). This variability has been ascribed by R. Wu and colleagues (16) to the fact that the copy number of pWR590-derived plasmids sharply decreases as the cells are passaged subsequent to transformation. Numbers on left show size in kilodaltons.

Zeste-binding sites at the white locus. We localized the sites on the pBE2 plasmid recognized by the NC fusion protein by assaying DNA fragments produced by cleavage with a variety of restriction enzymes (Fig. 2). In each case, a discrete subset of input fragments appeared in the immunoprecipitates. Since the DNA sequence of this region of the white locus is known (6, 36) these fragments could be mapped with some precision (Fig. 3). Zeste-binding sites were found within or near molecular lesions associated with the mutations w^{sp1} and w^{sp2}, alleles which disrupt the zeste-white interaction (6, 13). At least two separable elements in this region were recognized by the zeste fusion

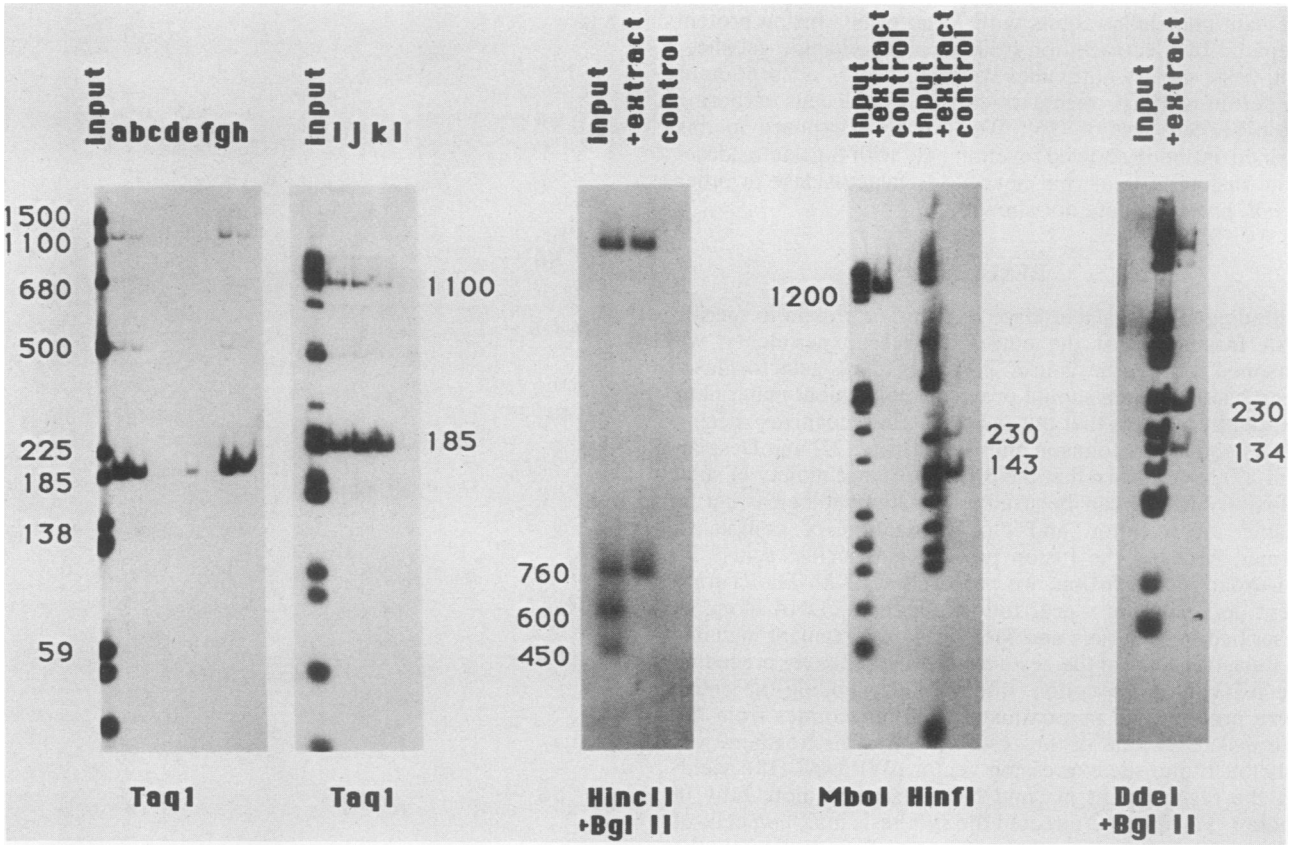

FIG. 2. The zeste moiety of a β-galactosidase-zeste fusion polypeptide binds to specific DNA fragments of the white locus. pBE2 DNA was digested with the indicated enzymes, the resultant ends were radiolabeled, and DNA binding assays were done as described in Materials and Methods unless otherwise noted. Input lanes contain 5% of the radioactive restriction fragments added to the reactions. The sizes of large vector fragments precipitated in the assay are not indicated; reactivity of these fragments presumably results from the presence of multiple weak sites of interaction. For *Taq*I digestions, lanes a to h were run on a 7% polyacrylamide gel, while lanes i to l were run on a 10% polyacrylamide gel. Lanes: a, 10 μg of total extract protein from cells harboring the NC plasmid; b, 20 μg of the same extract; the reaction appears to be saturated for fusion protein at this level; c, 10 μg of extract from cells containing vector 590-1; d, no extract; e, 10 μg of NC extract, with 4 μg of pBE2 competitor DNA; f, no antibody; g and h, purified antibody against the zeste moiety of the NC polypeptide instead of anti-β-galactosidase; i, 50 mM NaCl; j, 100 mM NaCl; k, 300 mM NaCl; l, 500 mM NaCl. In the remaining panels, +extract indicates the presence of 10 μg of NC extract, while control denotes the addition of 10 μg of vector 590-1 extract. *Hin*cII-*Bgl*II digests were run on 1% agarose gels, while the remaining reactions were analyzed on 7% acrylamide gels. Numbers indicate sizes in base pairs.

protein. In most digests, the promoter proximal of the two sites bound with relatively greater efficiency. However, approximately equivalent amounts of two adjacent *Dde*I fragments were found in the immunoprecipitate. The 23 nucleotides between the *Dde*I site separating these two fragments and the next most promoter-distal *Taq*I site thus appear to be of some importance in determining the strength of the recognition; the B104 transposon found in w^{sp1} DNA is inserted into this 23-nucleotide stretch (6, 36). Analysis of sequences closer to the initiation point of white gene transcription (between coordinates −0.7 and +0.6 kb) shows no additional strong zeste-binding sites in this region of the white locus (data not shown).

Zeste-binding sites in the bithorax complex. Zeste-mediated transvection effects involving combinations of a variety of bithorax complex alleles have been noted (30, 31). As these mutations map throughout the bithorax complex, which encompasses over 100 kb of DNA (1), it is possible that a number of zeste-binding sites might be found through the region, each of which might influence *Bx-C* expression. We thus initiated a search for such potential binding sites at a low degree of resolution by analyzing digests of four bacteriophage lambda clones, each containing approximately 20 kb of *Drosophila* genomic DNA. These clones collectively span the region between coordinates −44 and +16 on the map of Bender et al. (1). These clones were digested with the enzymes *Eco*RI, *Bam*HI, and *Hin*dIII, which in general yielded fragments greater than 1 kb in length. Although such large fragments often precipitated adventitiously in the binding reactions (see Discussion), only two regions of this part of *Bx-C* were found to interact significantly with the β-galactosidase-zeste fusion protein.

The stronger of these two sites was found on a DNA fragment including sequences encoding the 5′ end of the *Ubx* transcript. To localize the binding site more precisely, we performed reactions using restriction digests of a 900-base-pair *Eco*RI-*Hin*dIII fragment containing this site. These results are shown in Fig. 4 and summarized diagrammatically in Fig. 5. A region between 50 and 205 nucleotides upstream of the *Ubx* transcription start was strongly bound by the fusion protein. No mutations are currently known which alter the sequences in this binding region, so its significance to *Bx-C* expression remains undefined though suggestive.

The second fragment of the bithorax complex that contains a zeste-binding site was found very near the 5′ end of

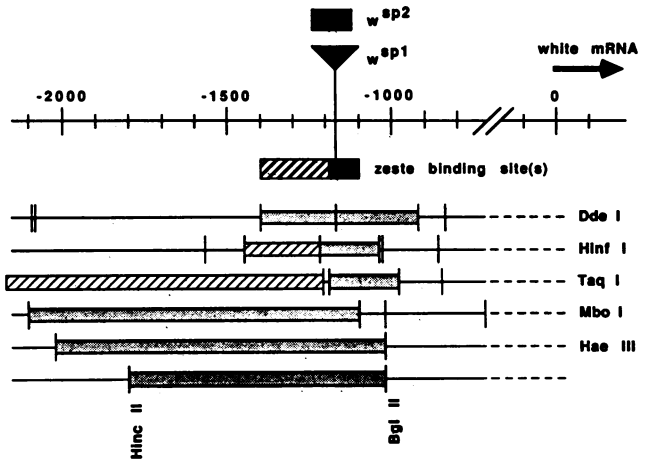

FIG. 3. Schematic representation of zeste-binding sites in the vicinity of the w^{sp1} mutation. The coordinates employed are in nucleotide pairs, with coordinate +1 representing the 5′ end of the white mRNA according to O'Hare et al. (36). The w^{sp1} mutation is associated with the insertion of a B104 transposable element at position −1171 (6, 36); while the w^{sp2} lesion is associated with a 119-nucleotide-pair deletion from −1115 to −1234 (6). DNA fragments reacting strongly with the β-galactosidase-zeste fusion protein are indicated by shaded bars on the restriction map below, while fragments less strongly though reproducibly bound by the fusion polypeptide are indicated by hatched bars. Two adjacent *Dde*I fragments are precipitated with approximately equal efficiencies, as observed on gels electrophoresed for a long time to resolve these two fragments (data not shown). On the *Dde*I-*Bgl*II digest in Fig. 2, only the *Dde*I end of the smaller bound fragment is labeled, so the specific activity of this fragment is one-half that of the other precipitated fragment. The solid bar summarizes the position of sequences strongly bound by the zeste fusion protein, while the bar hatched with dark lines indicates a region with somewhat lower affinity.

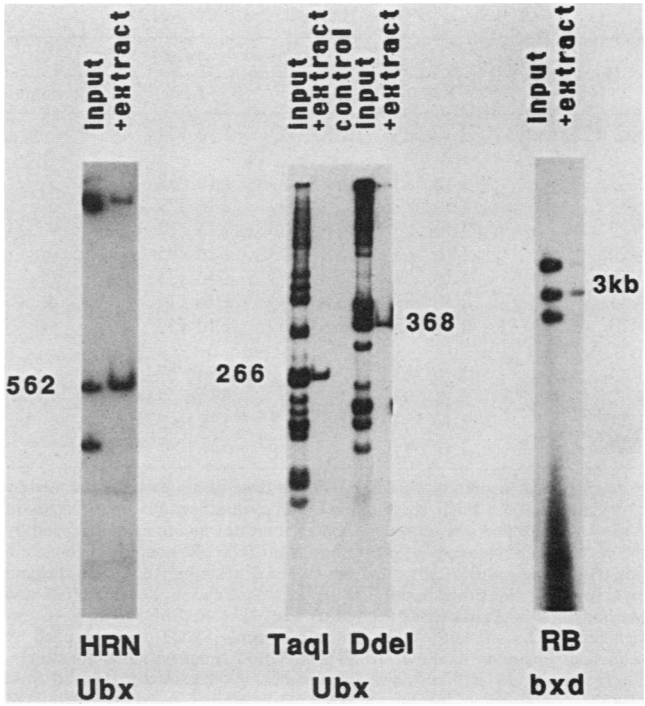

FIG. 4. Binding of fusion protein to sequences in the bithorax complex. DNA binding assays were performed with extracts from cells containing the NC construct as described in Materials and Methods. Targets, which were cut with the indicated restriction enzymes, include a 900-base-pair *Hin*dIII-*Eco*RI fragment containing the start of the *Ubx* transcriptional unit cloned into the Bluescript vector and a 5.1-kb *Eco*RI fragment containing the *bxd* transcription start cloned into pBR322. Sizes in base pairs of fragments preferentially precipitated by the β-galactosidase-zeste fusion protein are indicated.

the *bxd* transcriptional unit (Fig. 4, *bxd* panel). Analysis of a subclone containing this region showed that the binding site is mainly localized to a 600-base-pair *Taq*I fragment which lies between an *Eco*RI site at coordinate −3 and a *Bam*HI site at coordinate +0.2 on the map of Bender et al. (1) (data not shown). As the sequence of this region of DNA is not currently available, the exact position of the binding site relative to the transcript cannot yet be determined. However, we further attempted to analyze whether this binding site might be correlated with differences between the pbx^1 and pbx^2 mutations with respect to transvection: pbx^2 shows partial phenotypic complementation with *Ubx* and *bxd* alleles, whereas pbx^1 does not (31). These mutations are associated with deletions which partially overlap each other. This weak zeste-binding site is clearly absent from the pbx^1 chromosome, but is very close to the breakpoint of the pbx^2 deletion. A fragment of pbx^2 DNA containing the breakpoint does not appear to be specifically precipitated by the fusion protein (data not shown). However, because the binding observed even with wild-type DNA is quite weak, it cannot be stated unequivocably that the pbx^2 mutation completely deletes this site. The conclusion that this weaker site is removed by both *pbx* deletions and is not correlated with the differences in their transvection behavior must therefore be regarded as provisional.

Localization of DNA-binding domain of zeste protein. Several recent studies in yeasts have shown that limited domains of transcription factors are sufficient to bind specific DNA sequences at target genes (19, 20, 26). Similarly, the homeobox motif, restricted to approximately the C-terminal 60 amino acids of several presumed regulatory proteins in *D. melanogaster* and other organisms, appears to have specific DNA-binding properties (7, 9). In the cases studied, it is furthermore clear that DNA binding and transcriptional activation are separable functions of these proteins (4, 20,

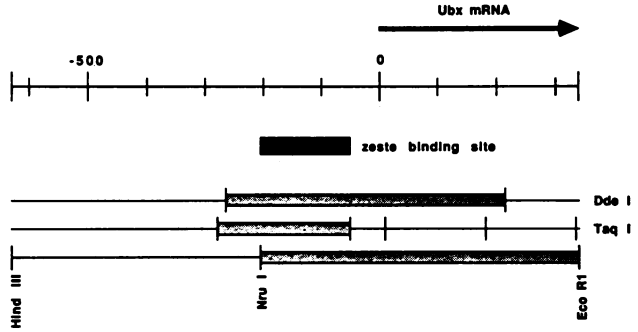

FIG. 5. Schematic representation of the strong binding site at the start of the *Ubx* transcription unit. Representation of bound fragments is as described in the legend to Fig. 3; distances are given in base pairs. The location of restriction sites relative to the 5′ end of the *Ubx* transcript is from Saari and Bienz (42), although a *Dde*I site predicted by their sequence at position +95 does not appear to be present in the clone used as a target in our experiments, presumably as the result of a DNA polymorphism.

TABLE 1. β-Galactosidase-*zeste* derivatives[a]

Derivative	Amino acids present			
	β-Gal	Spacer	Zeste	C terminus
NC (complete)	1 to 583	1SKQTQ	1 to 574	
N418	1 to 583	1SKQTQ	1 to 418	3
N329	1 to 583	1SKQTQ	1 to 329	3
N277	1 to 583	1SKQTQ	1 to 277	4
N205	1 to 583	1SKQTQ	1 to 205	4
N138	1 to 583	1SKQTQ	1 to 138	R3
N111	1 to 583	1SKQTQ	1 to 111	4
N73	1 to 583	1SKQTQ	1 to 73	4
36C	1 to 583	1	36 to 574	
47C	1 to 583	2H	47 to 574	
178C	1 to 583	2P	178 to 574	
327C	1 to 583	1	327 to 574	

[a] Amino acid sequence of β-galactosidase-zeste fusion proteins encoded by expression clones. Each polypeptide contains the first 583 amino acids of β-galactosidase at its N terminus. Spacer indicates amino acids encoded by the vector pWR590 or pWR590-1 (16), a *Hin*dIII-*Eco*RI adaptor oligonucleotide (New England BioLabs), and the vector Bluescribe BBM$^-$ (Stratagene) as a result of the procedures used in the construction (see Materials and Methods). 1 = PGELEFEPLRACMP and 2 = RASSNSNPFELAC, as signified by the one-letter code. The amino acids SKQTQ are encoded by sequences upstream of the *zeste* ATG initiation codon in some constructs. C-terminal amino acid residues are encoded by nucleotides derived from pBR322 contained in the vector pWR590-1, with 3 = AQACARSFGQR YQLTQRR and 4 = PKLALGRSAAASGISSLKGGNTVIHRIRG.

32). We thus initiated studies to define functional domains of the zeste protein by analyzing the ability of β-galactosidase-zeste fusion proteins serially truncated at the zeste N or C terminus to bind specific DNA sequences.

A set of constructs successively deleting C-terminal-coding information from the complete fusion gene was constructed. When these plasmids were introduced into *E. coli* cells, prominent polypeptides containing β-galactosidase antigenic determinants were synthesized (Fig. 1). The mobility of these fusion proteins on denaturing polyacrylamide gels approximated the sizes predicted by the nucleotide sequence (Table 1), though several of these polypeptides migrated slightly slower in these gels than expected. Similarly, a series of derivatives of the β-galactosidase-zeste fusion proteins was synthesized in which DNA from codon 1 of the zeste gene to various positions throughout the coding region was deleted, while ensuring that the correct reading frame with respect to β-galactosidase was maintained (Table 1). These constructs again produced appropriately sized fusion proteins in *E. coli* cells (data not shown).

Extracts from bacteria harboring both sets of truncated β-galactosidase-zeste fusion proteins were assayed for their ability to bind specifically to the w^{sp} region upstream of the sequences transcribed into white mRNA. All the C terminally deleted proteins bound specifically to the same DNA fragments recognized by the complete fusion product, except for the two smallest derivatives, N111 and N73 (Fig. 6A). The amount of the labeled DNA fragments precipitated in the presence of the N418, N329, N277, and N205 deletion proteins was approximately the same as that observed with the complete fusion extract. However, the polypeptide containing only the first 138 amino acids of zeste reproducibly appeared to react with these DNA fragments less efficiently in these assays. In a similar fashion, fusions deleting the first 36 or 47 amino acids of zeste retained specific DNA-binding activity, although the specific activity of binding in these extracts appeared reduced relative to that of extracts containing the complete β-galactosidase-zeste fusion protein. Deletion of the first 178 or 327 N-terminal amino acids of zeste abolished the ability of the resultant fusion proteins to bind specific DNA fragments from the w^{sp} region of the white locus (Fig. 6B).

These results are summarized schematically in Fig. 7. At the current resolution of these experiments, the domain of the zeste protein absolutely required for specific DNA binding was localized between amino acids 47 and 138.

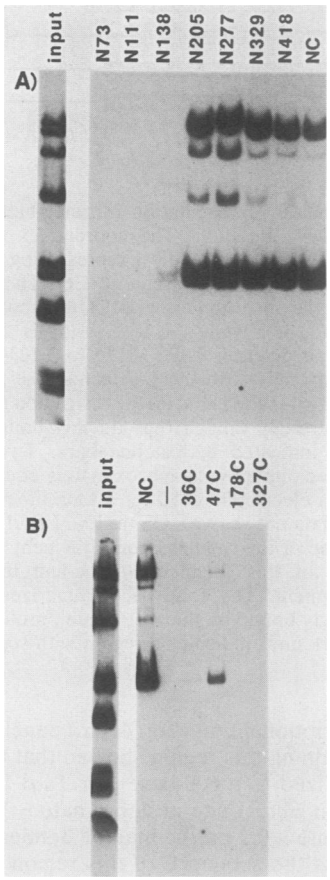

FIG. 6. DNA binding assays performed with extracts containing truncated β-galactosidase-zeste fusion polypeptides. Plasmid pBE2 DNA, with an insert from the w^{sp} region of the white locus, was digested with *Taq*I, labeled, and subjected to DNA binding assays as described in Materials and Methods. Input lanes show 5% of the mixture of *Taq*I fragments added to each reaction; other lanes show assays performed with 10 μg of extracts from cells harboring the indicated plasmids. The two most prominently precipitated bands are the same fragments of 185 and 1,100 base pairs seen in Fig. 2. (A) Extracts prepared from constructs which delete the C terminal of zeste. Gel electrophoresis and Western blot analysis of the extracts used in this experiment are shown in Fig. 1. Lower exposures of the autoradiograms shown in this figure indicate that the specificity of the zeste-deleted fusion proteins synthesized by plasmids N205 and N277 is similar to that of the full-length NC fusion. Extracts from cells harboring plasmids N205 and N277 do appear to have increased specific activity of binding, perhaps owing to decreased concentrations of degradation products separating the zeste and β-galactosidase moieties of the fusion proteins. (B) Extracts prepared from constructs which delete the N terminal of zeste. Western blot analysis indicates that the amount of β-galactosidase-zeste fusion protein in each extract is approximately equal (not shown). The bands faintly observed in the 36C lane are readily visible upon overexposure of the autoradiogram and are indistinguishable from those obtained with the NC extract.

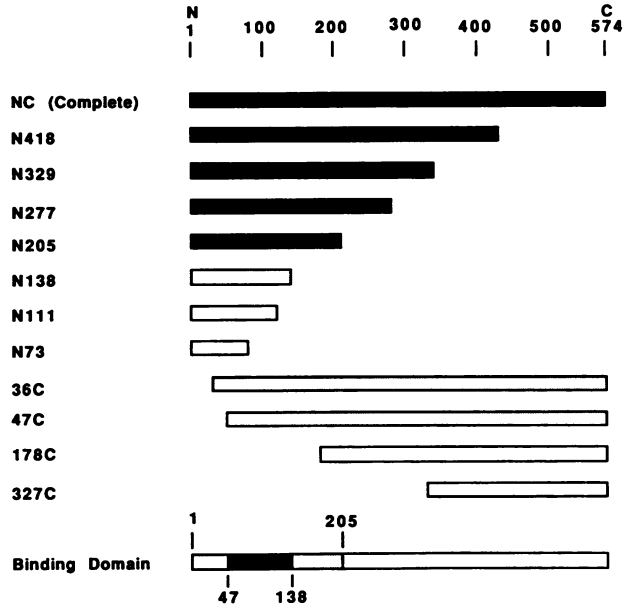

FIG. 7. Schematic representation of DNA binding assays performed with extracts containing truncated β-galactosidase-zeste fusion polypeptides. Bars indicate amino acids from the zeste protein encoded by each fusion construct; complete information on the structure of these fusion proteins is given in Table 1. Solid bars indicate that the extract in question binds specific DNA sequences from the white locus with approximately the same affinity as extracts containing fusion proteins containing the complete zeste polypeptide, while shaded bars represent extracts with reproducibly reduced specific binding activity. Fusion proteins denoted by open bars show no detectable recognition of these DNA fragments. These data are summarized at the bottom of the figure. The region between amino acids 47 and 138 of the zeste protein is absolutely required for binding (solid bar), while hatched segments denote regions which, when deleted, result in reduced binding activity.

Deletion of amino acids 1 to 46 or 139 to 205 appeared to reduce the affinity at which the fusion proteins recognize these same DNA fragments. However, the effect of these latter alterations could also be ascribed to other factors, such as decreased accessibility of the binding domain to DNA sequences owing to peculiarities in the three-dimensional structure of these particular fusion proteins.

DISCUSSION

Sequences recognized by zeste fusion proteins. We demonstrated here that β-galactosidase-zeste fusion proteins exhibit specific DNA-binding properties in vitro. For the white locus, the sites recognized by the zeste fusion protein are correlated with sequences known from genetic experiments to be required for interaction with zeste (13, 29, 40), indicating that the observed binding is of biological significance. Although no evidence to date directly implicates the strong zeste-binding site observed in the upstream region of the *Ubx* transcriptional unit in the regulation of the bithorax complex, phenotypes of heterozygous combinations of certain *Bx-C* mutant alleles are clearly influenced by the allelic state of zeste.

The presence of the β-galactosidase moiety on the fusion polypeptides used in these studies is unlikely to influence the observed binding activity. β-Galactosidase itself does not bind to DNA fragments (7, 22) (Fig. 2). Moreover, Benson and Pirrotta (2) have recently performed experiments analogous to those reported here, utilizing zeste protein synthesized in *E. coli* cells, but not fused to additional amino acids. The results obtained in both laboratories, at least with respect to binding sites in the vicinities of the w^{sp} mutations and the *Ubx* transcription start, appear to be indistinguishable. It is, however, not clear whether posttranslational modifications of the zeste protein might occur in *D. melanogaster* cells, but not in bacterial cells, which could alter the specificity or affinity of the zeste protein for various DNA sequences.

At the white locus, two adjacent stretches of DNA are independently recognized by zeste fusion proteins in these assays. These sites are separated by the insertion of a B104 transposon in the w^{sp1} mutation (6, 36), which disrupts the zeste-white interaction (13). It is thus possible that the physical juxtaposition of the two sites is important for the in vivo effects of zeste. Alternatively, transcription of the B104 element may in some manner interfere with the ability of the zeste protein to recognize one or both of the binding sites in *Drosophila* tissues. Clustered binding sites may be a characteristic of many regions of DNA which interact with the zeste protein; it is at least possible that the strong binding fragment at *Ubx* could be resolved into subsites by appropriate cleavage. Comparison of the sequence of the DNA fragments at white and *Ubx* precipitated in our experiments shows no obvious relationships. Zeste binding may thus be dependent on relatively short sequence motifs, or alternatively, several distinct sequences may be employed for zeste binding, as has been observed for the yeast HAP1 transcriptional activator protein (38). Resolution of this question must await results of experiments, currently in progress, to refine the localization of these zeste-binding sites.

Caution should be employed in the interpretation of weaker zeste-binding sites identified by these assays. It appears that the presence of several such sites of weak interaction on a single DNA fragment allows the efficient recovery of that fragment in the immunoprecipitate. As a case in point, at least three *Taq*I fragments of the Bluescript vector were weakly recognized by the β-galactosidase-zeste fusion protein, while the entire vector was precipitated almost as readily as the *Hin*cII-*Bgl*II fragment containing the w^{sp} region (Fig. 2). This caveat is of particular relevance for larger restriction fragments, which are more likely to contain multiple weak, presumably adventitious, sites.

On the other hand, the possibility that at least certain weak sites, such as those seen within or near the *bxd* transcriptional unit, are of biological significance cannot be ignored. Lower affinity protein-DNA interactions might nonetheless modulate gene expression; alternatively, binding to these sites in vivo may be enhanced by additional factors present in *Drosophila* nuclei. In this light, we have found that the zeste fusion protein reproducibly shows weak affinity for certain DNA fragments from two wild-type variants of the *Sgs-4* locus (Canton S and Hikone R [data not shown]). In a region covering the entire *Sgs-4* transcriptional unit and an additional 3 kb of upstream sequences, these weak interactions appear localized to the protein-coding regions of both genes. This result is potentially of interest because the levels of transcripts from certain weakly expressed alleles of the *Sgs-4* locus increase when paired with more strongly expressed alleles, a phenomenon similar in many ways to transvection (27, 28). It is not currently known whether the state of the zeste locus influences these effects; the weak interaction observed may or may not be of relevance to proximity-induced enhancement of *Sgs-4* expression.

Functional domains of zeste locus product. Results summarized in Fig. 7 of this paper show that the region of the zeste protein between amino acids 47 and 138 contains information sufficient to interact with the binding sites at w^{sp}. This domain, as currently defined, includes amino acids encoded by the first two exons of the zeste locus. The region of the zeste protein required for binding is highly enriched for charged residues, with a preponderance of basic amino acids (22% arginine plus lysine; 11% glutamic acid plus aspartic acid) (34, 40). No obvious similarities to homeobox (9) or zinc finger (17) motifs characteristic of other known DNA-binding proteins are observed. Two parts of the protein adjacent to the binding domain (amino acids 1 to 46 and amino acids 139 to 205) are not required for the DNA-binding function but influence the affinity of the zeste protein for recognized DNA fragments.

The molecular lesions associated with a variety of zeste locus alleles have recently been characterized (39; Mansukhani et al., in press). It is of interest that none of these mutations [z^1, z^{OP6}, z^{v77h}, z^{11G3}, z^{π}, and *In(1)e(bx)*] map to or are predicted to affect the DNA-binding domain of the zeste protein as defined in these experiments. Thus, DNA binding cannot be the sole function of the zeste locus product. We anticipate that many of the above mutations might alter the ability of the zeste protein to interact with other components of the transcriptional machinery, defining other functional domains. This idea is also in accordance with the suggestion of Pirrotta et al. (39) that the zeste locus is essential for *Drosophila* viability. Thus, homozygous viable z^a-like alleles such as *In(1)e(bx)*, previously regarded as null alleles (25), may not in fact represent the completely inactive state of zeste. Even though the zeste product encoded by the *In(1)e(bx)* rearrangement lacks the entire C-terminal half of the normal polypeptide, the truncated product should retain at least the ability to bind specific DNA sequences.

Zeste locus product and transvection phenomena. Genetic data strongly suggest that at least certain mutant derivatives of the zeste protein can interact with unpaired or unduplicated targets at w^{sp}. Thus, altered eye color phenotypes are seen in z^{OP6} and z^{v77h} males carrying only a single copy of the w^+ locus. As mentioned above, neither of these two mutations alters the DNA-binding domain of the zeste protein as defined by these in vitro studies. Transvection, that is, pairing-dependent modulation of gene expression, is therefore likely to involve properties of the zeste protein separate from the intrinsic affinity of a zeste polypeptide for individual binding sites at target genes.

One attractive model consistent with this limitation posits that the zeste protein is able to associate into dimeric or multimeric aggregates in *Drosophila* nuclei. If this was the case, cooperative binding to two or more *cis*-acting target sites brought into proximity by synapsis or tandem duplication might result. This postulated cooperativity (which may also be considered as an increased local concentration of the zeste product) would then be the physical basis of transvection effects. Similar mechanisms involving the looping of DNA between two regions simultaneously bound by a dimeric regulatory protein have recently been described in procaryotic systems (34, 41). The zeste mutations mentioned above could then be imagined to affect the ability of the protein to form multimers. This model requires that a zeste aggregate could simultaneously bind sites on homologous chromosomes or sites separated by up to 80 kb of DNA (12). Approximately one-half of the zeste protein is composed of stretches of glutamine or alanine residues or of the alternating copolymer (39; Mansukhani et al., in press); it is conceivable that this region might form an arm allowing zeste aggregates to span rather large distances within the nucleus.

The molecular nature of the events preceding and subsequent to the binding of zeste protein to its target sites remains unclear. A variety of unlinked suppressors and enhancers of the z^1 eye color phenotype have been characterized (24, 37). Participation of the factors encoded by these loci might be essential for in vivo binding of the zeste protein, formation of a zeste aggregate as suggested by the hypothesis above, or modulation of the transcription of target genes, or alternatively, the effects of these modifiers may be more indirect. Because of the absence of indisputable zeste null alleles, it is difficult to define even basic parameters concerning the role of the wild-type zeste protein in transcriptional regulation. For example, though various target genes interact quite differently with the spectrum of zeste alleles (11), it is conceivable that binding of the wild-type zeste protein may enhance transcription at some loci while repressing gene expression elsewhere. The demonstration here that the zeste gene product interacts directly with specific DNA sequences should provide a focus for future experiments and hypotheses designed to probe the nature of gene regulation by zeste and the molecular nature of transvection phenomena.

ACKNOWLEDGMENTS

We thank R. Levis, W. Bender, M. Peifer, R. Colvin, and L. Prestidge for providing plasmids containing target sequences and J. Lis, R. MacIntyre, and M. Wolfner for critical evaluation of the manuscript.

This work was supported by Public Health Service grant GM-31935 from the National Institutes of Health.

LITERATURE CITED

1. **Bender, W., M. Akam, F. Karch, P. A. Beachy, M. Pfeifer, P. Spierer, E. B. Lewis, and D. S. Hogness.** 1983. Molecular genetics of the bithorax complex in *Drosophila melanogaster*. Science **221**:23–29.
2. **Benson, M., and V. Pirrotta.** 1987. The product of the *Drosophila zeste* gene binds to specific DNA sequences in *white* and *Ubx*. EMBO J. **6**:1387–1392.
3. **Bingham, P. M., and Z. Zachar.** 1985. Evidence that two mutations w^{DZL} and z^1 affecting synapsis-dependent genetic behavior of *white* are transcriptional regulatory mutations. Cell **40**:819–825.
4. **Brent, R., and M. Ptashne.** 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43**:729–736.
5. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**: 165–170.
6. **Davison, D., C. H. Chapman, C. Wedeen, and P. M. Bingham.** 1985. Genetic and physical studies of a portion of the *white* locus participating in transcriptional regulation and in synapsis-dependent interactions in *Drosophila* adult tissues. Genetics **110**:479–494.
7. **Desplan, C., J. Theis, and P. H. O'Farrell.** 1985. The *Drosophila* developmental gene *engrailed* encodes a sequence-specific DNA binding activity. Nature (London) **318**:630–634.
8. **Gans, M.** 1953. Etude génétique et physiologique du mutant *z* de *Drosophila melanogaster*. Bull. Biol. Fr. Belg. Suppl. **38**:1–90.
9. **Gehring, W. J.** 1987. Homeo boxes in the study of development. Science **236**:1245–1252.
10. **Gelbart, W. M.** 1982. Synapsis-dependent allelic complementation at the *decapentaplegic* gene complex in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **79**:2636–2640.
11. **Gelbart, W. M., and C.-T. Wu.** 1982. Interactions of *zeste*

mutations with loci exhibiting transvection effects in *Drosophila melanogaster*. Genetics **102**:179–189.

12. **Goldberg, M. L., J.-Y. Sheen, W. J. Gehring, and M. M. Green.** 1983. Unequal crossing-over associated with asymmetrical synapsis between nomadic elements in the *Drosophila melanogaster* genome. Proc. Natl. Acad. Sci. USA **80**:5017–5021.
13. **Green, M. M.** 1959. Spatial and functional properties of pseudoalleles at the *white* locus in *Drosophila melanogaster*. Heredity **13**:303–315.
14. **Green, M. M.** 1963. Unequal crossing over and the genetical organization of the *white* locus of *Drosophila melanogaster*. Z. Vererb. **94**:200–214.
15. **Gunaratne, P. H., A. Mansukhani, S. E. Lipari, H.-C. Liou, D. W. Martindale, and M. L. Goldberg.** 1986. Molecular cloning, germline transformation, and transcriptional analysis of the *zeste* locus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **83**:701–705.
16. **Guo, L.-H., P. P. Stepien, J. Y. Tso, R. Brousseau, S. Narang, D. Y. Thomas, and R. Wu.** 1984. Synthesis of human insulin gene. VIII. Construction of expression vectors for fused proinsulin production in *Escherichia coli*. Gene **29**:251–254.
17. **Harrison, S. C.** 1986. Fingers and DNA half turns. Nature (London) **322**:597–598.
18. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
19. **Hope, I. A., and K. Struhl.** 1985. GCN4 protein, synthesized *in vitro*, binds to HIS3 regulatory sequences: implications for the general control of amino acid biosynthetic genes in yeast. Cell **43**:177–188.
20. **Hope, I. A., and K. Struhl.** 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell **46**:885–894.
21. **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.
22. **Johnson, A. D., and I. Herskowitz.** 1985. A repressor (MAT 2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell **42**:237–247.
23. **Judd, B. H.** 1959. Studies on some position pseudoalleles at the *white* region in *Drosophila melanogaster*. Genetics **44**:34–42.
24. **Kalisch, W.-E., and B. Rasmuson.** 1974. Changes of *zeste* phenotype induced by autosomal mutations in *Drosophila melanogaster*. Hereditas **78**:97–104.
25. **Kaufman, T. C., S. E. Tasaka, and D. T. Suzuki.** 1973. The interaction of two complex loci *zeste* and *Bithorax* in *Drosophila melanogaster*. Genetics **75**:299–321.
26. **Keegan, L., G. Gill, and M. Ptashne.** 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science **231**:699–704.
27. **Korge, G.** 1981. Genetic analysis of the larval secretion gene *Sgs-4* and its regulatory chromosome sites in *Drosophila melanogaster*. Chromosoma **84**:373–390.
28. **Kornher, J. S., and D. Brutlag.** 1986. Proximity-dependent enhancement of *Sgs-4* gene expression in *D. melanogaster*. Cell **44**:879–883.
29. **Levis, R., T. Hazelrigg, and G. M. Rubin.** 1985. Separable *cis*-acting control elements for expression of the *white* gene of *Drosophila*. EMBO J. **4**:3489–3500.
30. **Lewis, E. B.** 1954. The theory and application of a new method of detecting chromosomal rearrangements in *Drosophila melanogaster*. Am. Natur. **88**:225–239.
31. **Lewis, E. B.** 1982. Control of body segment differentiation in *Drosophila* by the *bithorax* gene complex, p. 269–288. *In* M. Burger (ed.), Embryonic development: genes and cells. Alan R. Liss, Inc., New York.
32. **Ma, J., and M. Ptashne.** 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell **48**:847–853.
33. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Martin, K., L. Huo, and R. F. Schleif.** 1986. The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie in these same sites. Proc. Natl. Acad. Sci. USA **83**:3654–3658.
35. **Muskavitch, M. A. T., and D. S. Hogness.** 1982. An expandable gene that encodes a *Drosophila* glue protein is not expressed in variants lacking remote upstream sequences. Cell **29**:1041–1051.
36. **O'Hare, K., C. Murphy, R. Levis, and G. M. Rubin.** 1984. DNA sequence of the *white* locus of *Drosophila melanogaster*. J. Mol. Biol. **180**:437–455.
37. **Persson, K.** 1976. Modification of the eye color mutant *zeste* by suppressor, enhancer, and minute genes in *Drosophila melanogaster*. Hereditas **82**:111–120.
38. **Pfeifer, K., T. Prezant, and L. Guarente.** 1987. Yeast HAP1 activator binds to two upstream activation sites of different sequence. Cell **49**:19–27.
39. **Pirrotta, V., E. Manet, E. Hardon, S. E. Bickel, and M. Benson.** 1987. Structure and sequence of the *Drosophila zeste* gene. EMBO J. **6**:791–799.
40. **Pirrotta, V., H. Stellar, and M. P. Bozetti.** 1985. Multiple upstream regulatory elements control the expression of the *Drosophila white* gene. EMBO J. **4**:3501–3508.
41. **Ptashne, M.** 1986. Gene regulation by proteins acting nearby and at a distance. Nature (London) **322**:697–701.
42. **Saari, G., and M. Bienz.** 1987. The structure of the *Ultrabithorax* promoter of *Drosophila melanogaster*. EMBO J **6**:1775–1779.