The EMBO Journal vol.7 no.12 pp.3907 – 3915, 1988

The *Drosophila zeste* protein binds cooperatively to sites in many gene regulatory regions: implications for transvection and gene regulation

Mark Benson and Vincenzo Pirrotta

Department of Cell Biology, Baylor College of Medicine, Texas Medical Center, Houston, TX 77030, USA

Communicated by V.Pirrotta

The *Drosophila zeste* protein binds *in vitro* to several sites in the *white*, *Ultrabithorax*, *decapentaplegic*, *Antennapedia*, and *engrailed* genes and to at least one site in the *zeste* gene itself. The distribution of these sites corresponds often with that of regulatory elements in these genes as defined by mutations or, in the case of *white*, by molecular analysis. A *zeste* binding site is frequently found in the immediate vicinity of the promoter. *zeste* binding sites are composed of two or more *zeste* recognition sequences T/CGAGT/CG. Isolated consensus sequences do not bind or footprint. Cooperative interactions are involved both in binding to a given site and between proteins bound at independent sites. *zeste* bound to one DNA molecule can in fact bind simultaneously to another DNA molecule. These results suggest a general role for *zeste* in bringing together distant regulatory elements controlling the activity of a target gene. In this model, transvection effects are a by-product of normal intragenic *zeste* action.

Key words: cooperative binding/enhancers/DNA looping/DNA cross-linking

Introduction

The *zeste* gene of *Drosophila* is required for transvection and related phenomena that imply the possibility for interaction between two copies of a gene present on two homologously paired chromosomes. *zeste* dependent transvection effects have been detected at the *Ultrabithorax* (*Ubx*), *decapentaplegic* (*dpp*) and *white* genes (Lewis, 1954; Gans, 1953; Kaufman *et al.*, 1973; Babu and Bhat, 1980; Gelbart and Wu, 1982). The genetic and molecular results imply that *zeste* allows control elements on one copy of the gene to regulate the activity of the second copy if it is held in physical proximity by homologous pairing. In addition, in the presence of the z^1 allele of *zeste*, the expression of the *white* gene in the eye is specifically repressed when two copies of the *white* gene are physically close together due either to chromosome pairing or tandem duplication (Jack and Judd, 1979; Bingham and Zachar, 1985). The ability of the z^1 mutant product to alter the expression of the *white* gene is dependent on a sequence element located about 1.2 kb upstream of the transcription start site that has some of the features typical of eukaryotic enhancer elements (Pirrotta *et al.*, 1985; Levis *et al.*, 1985; Davison *et al.*, 1985). It has been suggested that elements such as this might be responsible for transvection effects by a mechanism similar to that which allows enhancers to control their own promoters in *cis* except that in transvection the enhancer effect would occur in *trans* between paired chromosomes (Zachar *et al.*, 1985; Pirrotta *et al.*, 1985). Alternative models to explain proximity-dependent transvection effects have invoked the existence of regulatory RNA species that are spatially restricted to the vicinity of the gene from which they are transcribed and require the interaction with *zeste* product for their function (Jack and Judd, 1979; Micol and Garcia-Bellido, 1988).

The molecular cloning of the *zeste* gene (Mariani *et al.*, 1985; Gunaratne *et al.*, 1986) has allowed the molecular study of the *zeste* gene product and its interactions. *zeste* protein (Benson and Pirrotta, 1987) or a *zeste*-β-galactosidase hybrid expressed in *Escherichia coli* (Mansukhani *et al.*, 1988) have been shown to bind specifically to DNA from the *white* and *Ubx* genes. Moreover, Biggin *et al.* (1988) have shown that *zeste* purified from *E.coli* or from *Drosophila* acts *in vitro* as a sequence-specific transcription factor, stimulating the activity of the *Ubx* promoter. These results strongly suggest that at least one of the functions of *zeste* is similar to that of other proximal promoter factors such as SP1 or AP1 (Briggs *et al.*, 1986; Lee *et al.*, 1987). However, such a role would not immediately explain the pairing dependence of the z^1 effect on *white* or the interchromosomal nature of transvection effects. Other difficulties arise from the structure of the regulatory regions of complex loci such as *dpp* or *Ubx*. In these cases, control regions defined by mutations and involved in transvection may be as much as 50 kb distant from the promoter they control in *cis*. It is possible that the *zeste*-dependent mechanisms that allow the interaction between paired chromosomes may also explain the long range interactions between distant regulatory elements and the promoter they control on the same chromosome.

Here we report the structure and sequence of a number of *zeste* binding sites, some features of their interaction with *zeste* protein and their distribution in a number of genes, some of which were not previously known to interact with *zeste*. We find that *zeste* bound at one site can facilitate binding to a second, distant site on the same DNA molecule and that, in fact, *zeste* can hold together two separate DNA molecules. These observations suggest a role of *zeste* in mediating the action at a distance by regulatory elements that may be located many tens of kilobases away from the promoter.

Results

Zeste binding sites in the white gene

We used the *white* gene as the most convenient substrate in which to identify *zeste* binding sites because its entire sequence has been determined (O'Hare *et al.*, 1984) and some of its regulatory elements have been approximately located (Pirrotta *et al.*, 1985; Levis *et al.*, 1985). We had previously shown that two contiguous *Hin*fI restriction

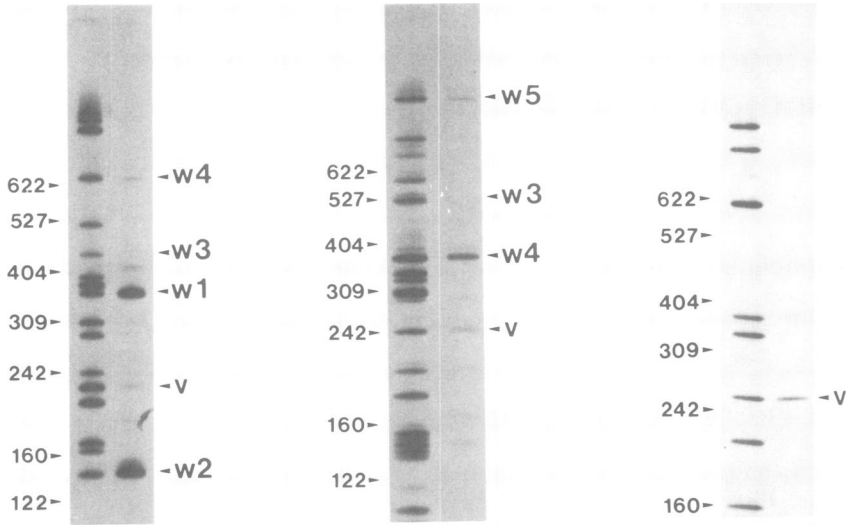

Fig. 1. Immunoprecipitation of *white* DNA. For each pair of lanes, one shows the total digestion products, the other the fragments selected by *zeste* in an immunoprecipitation experiment. The left panel shows a clone containing *white* sequences from position −6.3 to +0.5 kb, using the numbering shown in Figure 2, cut with *Hin*dIII and *Eco*RV and *Hin*fI. The clone used in the centre panel contained *white* DNA from −0.7 to +4.3 kb, and was cut with *Eco*RI and *Sal*I and *Sau*3A. The clone in the right panel extended from +4.1 to +6.7 kb and was cut with *Hin*dIII and *Sau*3A. For each pair of lanes, the positions of mol. wt markers are on the left and the binding fragments are identified on the right. W1, W2, W4, W5 represent the binding sites whose positions are shown in Figure 2. V: Vector fragments.

fragments beginning 1118 bp and 1260 bp upstream of the transcription start site are able to bind *zeste* protein independently in an immunoprecipitation assay (Benson and Pirrotta, 1987). For a more careful analysis, we used a series of subclones covering the entire *white* gene, including the transcription unit plus 6360 bp of upstream and 830 bp of downstream flanking sequences. The DNA of these clones was cut with suitable restriction enzymes, end-labelled and the fragments tested for binding to *zeste* protein by the immunoprecipitation assay (Figure 1). The results of such experiments, summarized in Figure 2, reveal the existence of two weaker binding sites in addition to the two strong sites previously characterized. Using combinations of restriction enzymes, the position of these additional binding sites could be narrowed down to the intervals −150 to −110 in the immediate vicinity of the promoter and +670 to +1330 in the major intron, measured from the transcription start site (Steller and Pirrotta, 1985). Efficient binding to these sites requires a 5-fold higher concentration of *zeste* protein than that required for the two upstream sites. We conclude that they have a lower affinity for *zeste* protein, the site near the promoter being of intermediate strength and the intron binding site the weakest. In the course of these experiments we found that weak binding sites are apparently also contained in the pUC8 vector, for example in a 191-bp *Hin*fI−*Eco*RI fragment. We have used this very weak binding as an internal standard against which to compare the strength of other binding sites. Binding weaker or equivalent to that of pUC8 DNA was considered not significant.

Footprinting analysis

Our previous results with one of the two strong upstream binding sites in the *white* regulatory region implicated a *Dde*I recognition sequence at position −1218 in the interaction with *zeste* (Benson and Pirrotta, 1987). The resulting *Dde* 234-bp fragment failed to bind to *zeste* but bound strongly when the *Dde* site was reconstructed. These observations suggested that part of the sequence required for *zeste* binding was at or very near the *Dde* site. To determine the sequences physically occupied by the *zeste* protein on the DNA of the *white* gene, we undertook a footprinting analysis of the regions indicated by the immunoprecipitation experiments. Figure 3 shows some of these results for the upstream binding site. The pattern of protection indicates that these are compound sites, in each case comprising two or more protected regions of 16−18 nucleotides each, containing a common or related hexanucleotide sequence: C/TGAGC/TG or its complement. Altogether, this upstream region contains five footprinting sites. The two most upstream sites are 42 bp apart, centre to centre, and in the same orientation. Together they constitute *zeste* binding site *white*-1, able to bind to *zeste* protein independently of the other sites in the immunoprecipitation assay. The next set of three footprinting sites begins 69 nucleotides further downstream (position −1254 to −1181). These three sites are at intervals of 37 and 22 bp respectively, two of them in one orientation and one inverted. The two more downstream sites are able to bind independently of the third when the *Dde* fragment containing them is cloned in the *Hin*dII site of pUC8 (thereby reconstructing the *Dde* site). The most upstream of these three, when separated from the others as a 21 bp *Taq*I fragment, fails to show independent binding in the immunoprecipitation assay. These three footprinting sites, constituting binding site *white*-2, are almost precisely excised by the w^{sp2} deletion which removes 111 nucleotides from position −1182 to −1293. It will be necessary in the remainder of this paper to distinguish between a footprinting site, i.e. the sequence protected by a *zeste* molecule bound to DNA, and a *zeste* binding site, which we define for the present as one or more footprinting sites which together allow a DNA fragment to bind to *zeste* in the immunoprecipitation assay. The two, as we shall show, are not always equivalent.

Armed with the information from these footprinting experiments we were then able to identify consensus sequences in other fragments selected in the immunoprecipitation assay. Two consensus sequences spaced 17 bp apart (from centre to centre) are present about 120 nucleo-

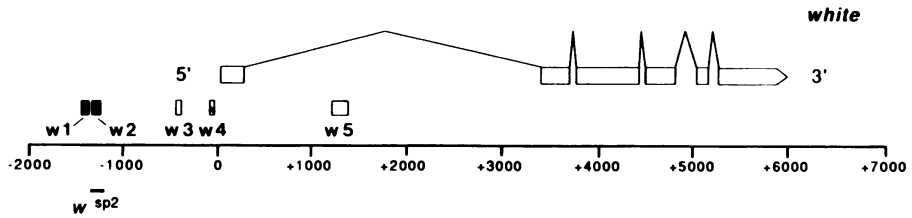

Fig. 2. *zeste* binding sites in the *white* locus. A map of the *white* transcription unit is shown with the scale in base pairs according to O'Hare *et al.* (1984) with minor corrections (Hazelrigg, 1987). The scale is in base pairs numbered from the transcription start site. Intervals within which *zeste* binding sites were mapped are shown as black boxes (strong sites), hatched boxes (intermediate sites) or white boxes (weak sites) labelled W1 to W5. The extent of the 111 bp w^{sp2} is shown below the scale.

tides upstream of the *white* transcription start site. The binding of *zeste* to these sites was confirmed by footprinting analysis. Another cluster of consensus sequences is present in the middle of the major intron, in a fragment identified by immunoprecipitation. Here we find four consensus sequences within 100 bp, two of them immediately adjacent to one another, forming a palindrome (see Figure 5).

An examination of the *white* gene sequence reveals the presence of many consensus sequences, generally isolated rather than in clusters of two or more within a 100 bp interval, and not associated with fragments showing detectable *zeste* affinity in the immunoprecipitation assay. Two exceptions to this statement are notable. One of these is the *white* upstream region, about 420 bp from the transcription start site. We find here two consensus sequences in the same orientation 31 nucleotides apart. The restriction fragment containing these sequences binds extremely poorly in the immunoprecipitation assay. When assayed by footprinting, however, these two sites are protected by *zeste*. The other cluster of consensus sequences revealed by sequence inspection is at the 3′ end of the gene, about 300 bp beyond the polyadenylation site. Here we find two consensus sequences in opposite orientation and 14 nucleotides apart. No detectable immunoprecipitation of restriction fragments containing these sequences was observed.

Zeste binding sites at selected loci

We next examined a number of other loci known to have complex regulatory regions, beginning with the two other cases known to interact with *zeste* in transvection phenomena, *Ubx* and *dpp*. We previously reported the existence of a strong *zeste* binding site in the immediate vicinity of the *Ubx* promoter. As has recently been shown (Biggin *et al.*, 1988), this site consists of five consensus sequences of the form described in the previous section, each protected by *zeste* protein in a footprinting assay. This multiple site is located between 40 and 140 nucleotides upstream of the *Ubx* transcription start site. To look for additional *zeste* binding sites, we then scanned the entire *Ubx* locus using a series of large genomic clones spanning the transcriptional unit and including 50 kb of upstream flanking sequences. We cut the DNA of these lambda clones with *Eco*RI, *Hin*dIII or *Bam*HI, alone or in suitable combinations to yield several restriction fragments and tested them for *zeste* binding with the immunoprecipitation assay. Several binding sites could be identified and localized with the aid of a restriction map of the *Ubx* chromosomal walk (Bender *et al.*, 1983). The map in Figure 4 summarizes these results and shows that the binding sites are distributed not only in the 40 kb preceding the transcription start, but also along

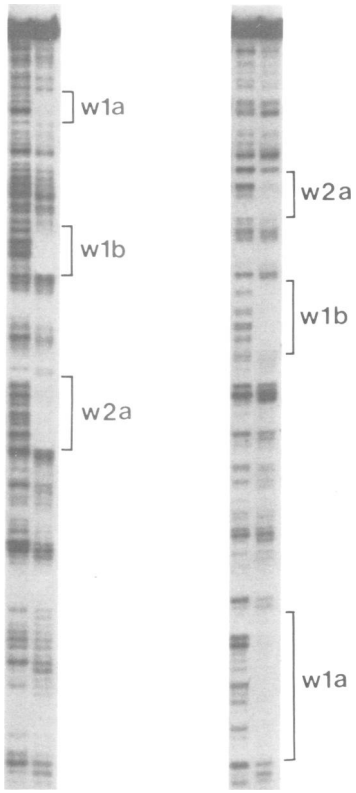

Fig. 3. Footprinting of upstream *zeste* binding sites in *white*. A *Dde* fragment (position −1220 to −1444 from the transcription start site) was used as probe. Footprinting sites on both strands are identified using the nomenclature of Figure 5.

the transcription unit. The position of these binding sites bears a rough similarity to positions of *Ubx* regulatory elements defined by mutations. However, we do not find a *zeste* binding site near the promoter of the early *bxd* transcripts as reported by Mansukhani *et al.* (1988). We have isolated and determined the sequence of two of these sites: one is the quintuple site at the *Ubx* promoter, the other is the site about 500 bp beyond the 3′ end of the longest *Ubx* transcript. This is a strong binding site whose unusual properties are discussed below.

Using a similar approach, we found binding sites at *dpp* as well as in the *Antennapedia* (*Antp*) and *engrailed* (*en*) loci. Maps of the *zeste* binding fragments in these loci are shown in Figure 4. In each case, the immunoprecipitation assay showed several binding fragments that have different affinities for *zeste* and are scattered both upstream and downstream of the transcription start sites. Frequently, as

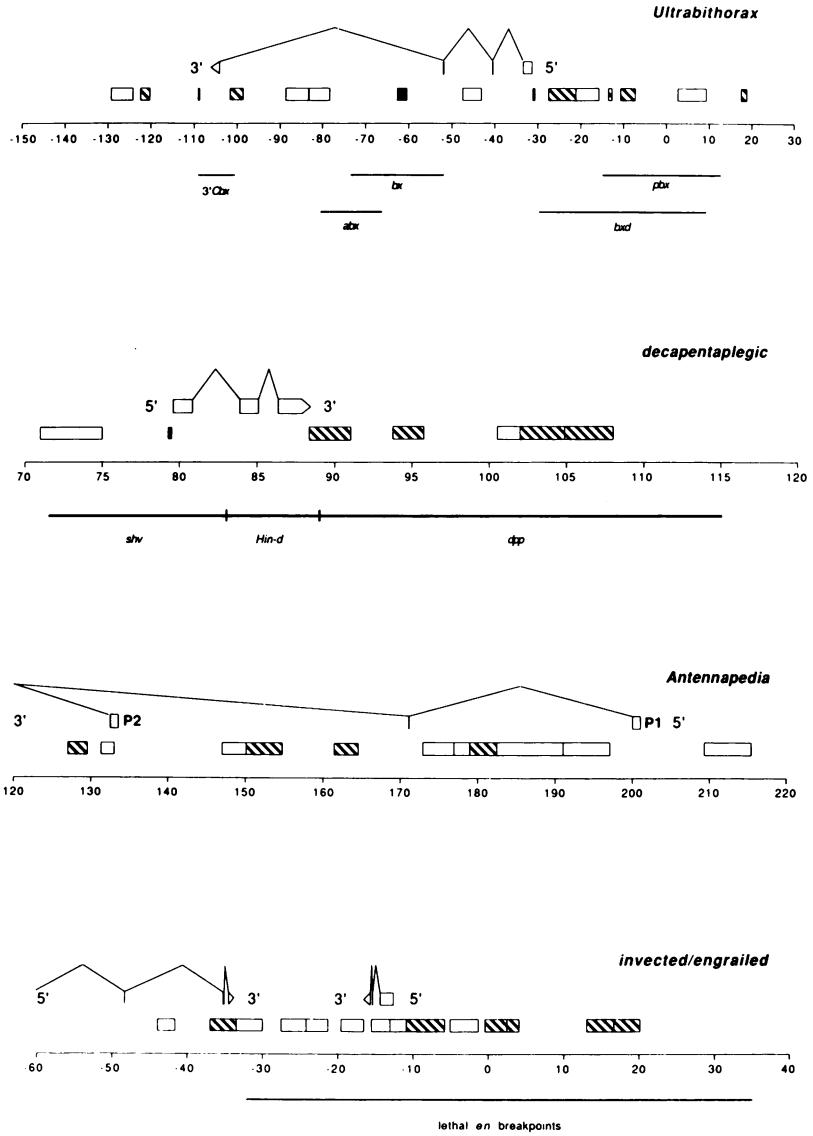

Fig. 4. Maps of *zeste* binding sites at several loci. The principal transcripts are shown for the *Ubx* (O'Connor *et al.*, 1988) and *dpp* loci (Gelbart *et al.*, 1985; St. Johnston *et al.*, in preparation). For the *Antp* gene, only the first half of the gene, containing the two promoters, P1 and P2, is shown (Laughon *et al.*, 1986; Stroeher *et al.*, 1986; Schneuwly *et al.*, 1986). The *engrailed* and *invected* maps are according to Poole *et al.* (1985), Kuner *et al.* (1985) and Coleman *et al.* (1987). Intervals containing *zeste* binding sites are symbolized as in Figure 2 and genetically defined regions are indicated below the maps of *Ubx*, *dpp* and *engrailed*.

at *white* and *Ubx*, one of the binding sites is found in the vicinity of the promoter. At *dpp*, where transcriptional analysis indicates the existence of multiple promoters (Gelbart *et al.*, 1985; St. Johnston *et al.*, manuscript in preparation), we find one binding site close to the start site of the 4.5 kb transcript. The sequence of this binding site (R.W.Padgett, S.Findley and W.M.Gelbart, personal communication), as expected, reveals the presence of multiple *zeste* consensus sequences that are protected by *zeste* protein in footprinting experiments (not shown). *Zeste* binding sites are also found near the *Antp* P2 promoter. In this case, however, we have no precise sequence information and the published data suggest that the binding site may be as far as 2 kb from the transcription start site. At *engrailed*, the strong binding site nearest the promoter is probably 2–5 kb upstream of the transcription start site. The sequence of the region immediately preceding the transcriptional start (Soeller *et al.*, 1987) contains a very striking nest of five overlapping *zeste* consensus sequences, beginning at position −133 (see Figure 5). However, in immunoprecipitation experiments, the restriction fragment containing this sequence fails to bind to *zeste* and footprinting analysis shows no occupancy. Two consensus sequences, 20 nucleotides apart, are found in the *zeste* gene itself 20 nucleotides downstream of the presumed transcription start site (Pirrotta *et al.*, 1987). Restriction fragments containing this region bind very weakly in the immunoprecipitation assay but the two consensus sequences are well protected by *zeste* protein in footprinting experiments.

We have detected no appreciable binding of *zeste* to DNA fragments from a *fushi tarazu* clone containing 8.5 kb of 5′ flanking sequence and 4.5 kb of 3′ sequence. We also tested the *sgs*-4 gene, for which transvection effects have been reported (Korge, 1981; Kornher and Brutlag, 1986)

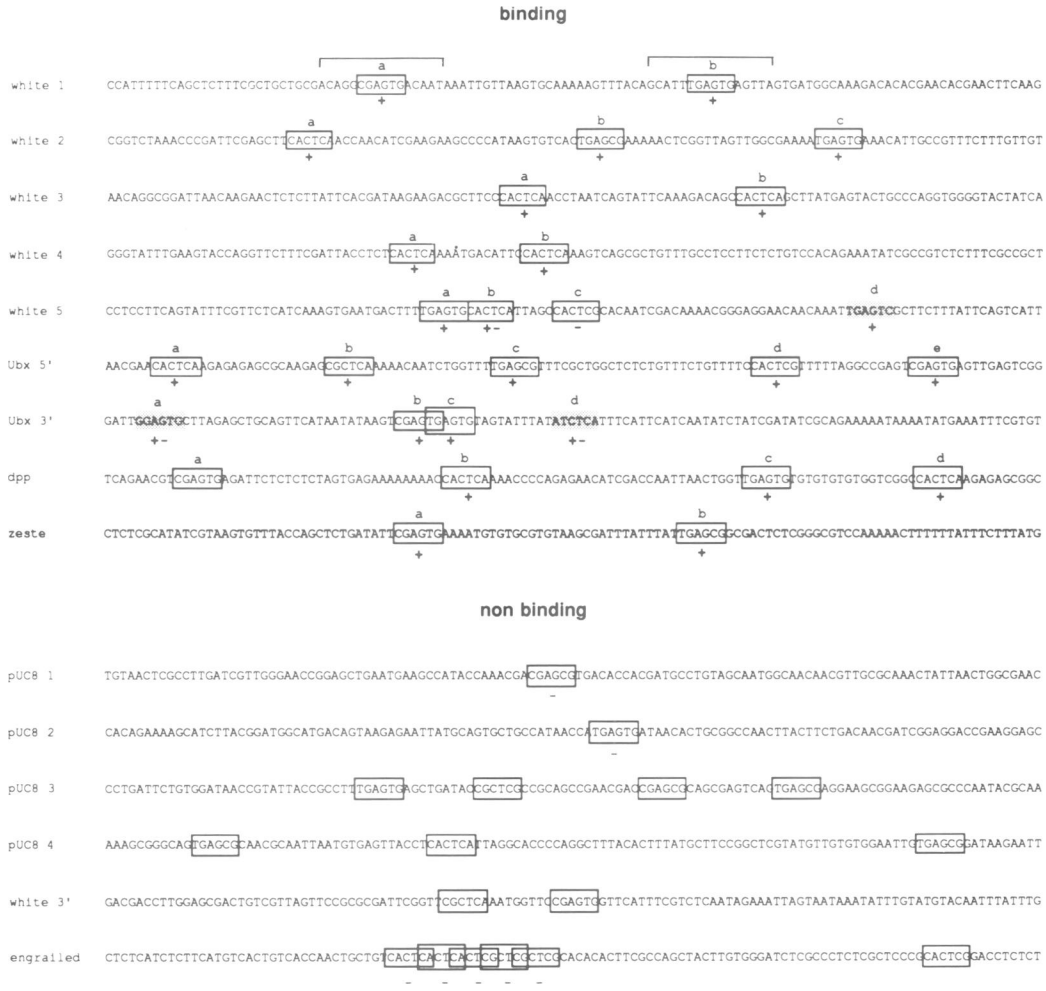

Fig. 5. *zeste* binding site sequences. The sequences of nine *zeste* binding and six non-binding regions are shown. The solid boxes indicate *zeste* consensus sequences T/CGAGT/CG, the shaded boxes near-consensus sequences that appear at least partly protected in footprinting experiments. The extent of the protected region is shown at the *white*-1 sequence. The + and − under each box indicate whether this sequence is protected against DNase while ± means weak protection. Unmarked boxes were not tested. Sites *white*-1 to 5 are indicated in the map shown in Figure 2. *White* 3′ lies 300 bp beyond the *white* polyadenylation site (O'Hare *et al.*, 1984). *Ubx* 5′ and 3′ are the *Ubx* promoter site and the site near the *Ubx* polyadenylation sequence (Saari and Bienz, 1987; K.Kornfeld and D.Hogness, personal communication). *dpp* refers to the site near the promoter of the *dpp* 4.5 kb RNA (R.W.Padgett, S.Findley and W.M.Gelbart, personal communication). The *zeste* site is 20 bp downstream of the presumed *zeste* transcription start site (Pirrotta *et al.*, 1987). The *engrailed* overlapping multiple consensus sequence lies 132 bp upstream of the transcription start site (Soeller *et al.*, 1988). Except for the pUC8 sites, the sequences are given 5′ to 3′ in the direction of the corresponding transcription unit.

although they have not been shown to be *zeste*-dependent. No strong *zeste* binding site was found in a genomic region containing the *sgs*-4 gene and extending from 12 kb upstream to 8 kb downstream of the coding sequences. A DNA fragment within the coding region of the gene gave, however, a very weak binding response. Although the *sgs*-4 gene contains an isolated *zeste* consensus sequence 70 bp upstream of the transcription start site, no good consensus hexanucleotides are found in the coding region. However, a near-consensus sequence TGAGCC is present in the repetitive block that constitutes much of the glue protein coding sequence. As a result, this motif is found ten times within 200 nucleotides and is very likely responsible for the weak binding we observe.

Sequence requirements for binding

Figure 5 shows the sequence of nine *zeste* binding regions and points out the consensus recognition sequences occupied by *zeste* in the footprinting analysis. In all cases but one, the *zeste* binding site consists of two or more consensus sequences of the form T/CGAGT/CG with a pronounced preference for T rather than C at the two pyrimidine positions. The footprinting results show that these consensus sequences are in each case contained within the region protected by *zeste* protein. The distance between consensus sequences in strong binding sites, measured from centre to centre, varies from 20 to 42 bp and their relative orientation can be direct or inverted or, in the case of sites with multiple consensus sequences, mixed, without obvious correlation with the binding affinity. Not surprisingly, sites containing multiple consensus sequences bind better in general than sites containing only two but even the latter differ greatly from one another in their affinity for *zeste*. However, in some cases, even sites containing multiple consensus sequences fail to bind or bind extremely poorly. This low affinity might be explained in some cases by the very short distance between consensus sequences (e.g. the *white* 3′ site). If *zeste* protects 16−18 bp against DNase

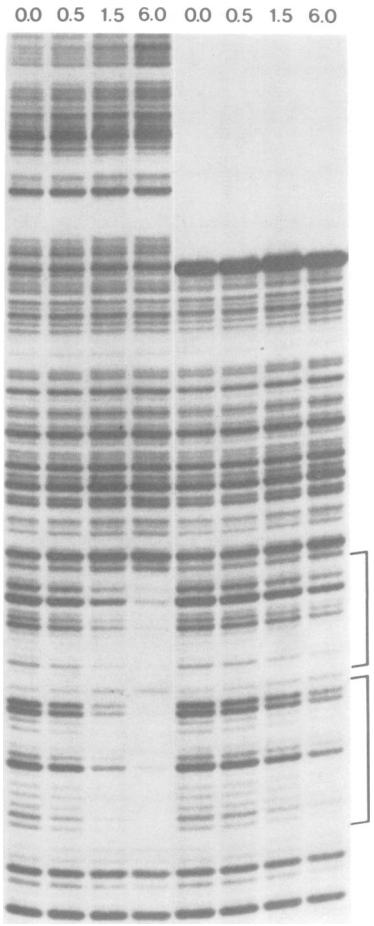

Fig. 6. Cooperative binding between distant sites. A DNA fragment containing all the *zeste* binding sites in the *white* 5′ flanking region (from −110 to −1850 from the transcription start) was labelled near the promoter site and used for footprinting experiments either uncut (left four lanes) or cut with *Eco*RV at position −223 (right four lanes). Amounts of protein in microlitres are indicated above each lane. Binding to the intact fragment is more efficient than to the cut fragment.

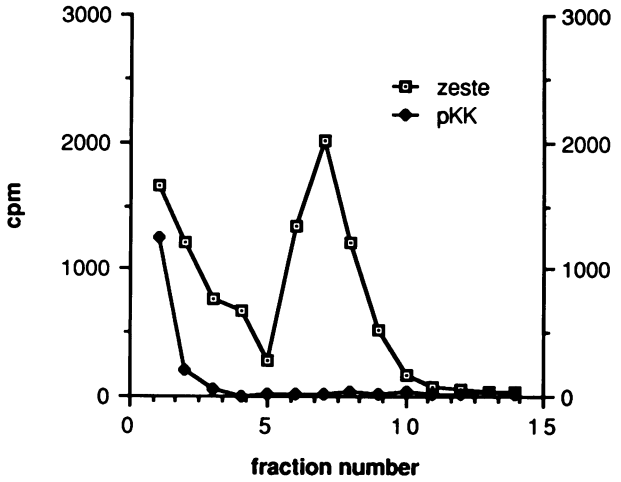

Fig. 7. A *zeste*−DNA complex is retained on a DNA affinity column. Extracts from bacteria containing the pKK223-*zeste* expression vector (Benson and Pirrotta, 1987) or the pKK223 vector alone were incubated with equal amounts of ^{32}P-labelled DNA containing the *Ubx* 5′ binding site, then applied to a DNA affinity column bearing the *zeste* consensus sequence. Fractions 1−4 represent 200 μl washed, fractions 5−14 are of 50 μl each and were eluted with 0.6 M NaCl.

attack, it might be expected that consensus sequences separated by less than 16−18 bp could not be simultaneously occupied. Another feature correlated with weak or no binding is the presence of Cs in the consensus sequence. Although several good binding sites contain one C at the first or fifth position in the sequence, no strong binding site contains Cs at both positions as does, for example, the extremely weak binding site pUC8-3. These considerations however do not explain all the instances of very weak binding, suggesting that the surrounding sequences may also play a role in the ability of a given pair of consensus sequences to form a stable complex with *zeste* protein. It may be relevant for example that strong binding sites are frequently flanked by sequences rich in runs of As or Ts and that the consensus sequence is almost always (17/24) preceded by a T.

Isolated consensus sequences do not bind to *zeste* by the immunoprecipitation assay nor do they footprint. An apparent exception is the *Ubx* 3′ site. This is a very strong binding site located just a few hundred nucleotides beyond the 3′-most polyadenylation site of the *Ubx* transcription unit. This site is unusual because instead of two separate consensus sequences, it contains two partially overlapping sequences: CGAGTGAGTG. Footprinting experiments show that this sequence is rapidly occupied and protected by *zeste*. However, the protection is never complete, even with very high amounts of *zeste* protein, probably because either one or the other consensus sequence can be occupied but not both simultaneously. The strong binding detected at this site is probably explained by the presence within 300 bp of several near-consensus sequences which are partly protected in footprinting experiments. Two of these sequences are indicated by shaded boxes in Figure 5. The strength of the binding, assayed by immunoprecipitation or footprinting, in fact decreases drastically if some of these near-consensus sequences are removed by cutting the fragment with *Eco*RV, *Spe*I or *Sal*I which cleave at 37, 83 and 118 bp from the double consensus site.

Cooperative interactions

The finding that isolated consensus sequences neither bind nor footprint and that stable binding requires the presence of two or more footprinting sites suggests that cooperative protein−protein interactions are involved. Binding of *zeste* at the quintuple *Ubx* 5′ binding sites in immunoprecipitation experiments is greatly reduced by the progressive deletion of consensus sequences, becoming virtually undetectable by the time three of the five footprinting sites are removed (Biggin *et al.*, 1988). Footprinting experiments show that when two sites are deleted the affinity of the remaining three sites is reduced four-fold (data not shown), indicating a degree of cooperativity between *zeste* bound at neighbouring sites.

We then considered the question of cooperativity between *zeste* bound to independent binding sites. We used a DNA fragment from the *white* regulatory region, containing two independent *zeste* binding sites in the region −1100 to −1300, one very weak binding site at −400 to −450 and one intermediate binding site near the promoter at −119 to −143. Using the footprinting assay, we compared the binding of *zeste* to the promoter site using the intact fragment as opposed to the fragment cut with *Eco*RV at position −223

(hence separating the promoter site from the other sites). The results shown in Figure 6 indicate that the presence of the distant binding sites facilitates the binding of *zeste* to the promoter site, increasing the affinity by approximately a factor of five. Using other restriction enzymes, we found that cleavage at position −1085 (*Bgl*II) or −538 (*Sph*I) has little effect but the 5-fold drop in affinity occurs when the fragment is cut at −408 (*Sca*I). No additional effect resulted from cutting at −223 (*Eco*RV). These results indicate that the complex is stabilized by the interaction of *zeste* bound at the *white*-3 site (−400 to −450) with *zeste* bound at the promoter site (−119 to −143).

Zeste can cross-link two DNA molecules

One possible explanation for the apparent long-range cooperativity described above is that the *zeste* molecules bound to two independent sites can interact or even form a single multivalent aggregate able to hold together two DNA regions. This kind of behaviour would provide an attractive basis for the understanding of transvection phenomena. In transvection, *zeste* mediates in some way the interaction between two genes on homologously paired chromosomes. In the case of the *zeste*−*white* effect, *zeste* can also mediate the interaction between two tandem copies of *white*. The simplest model for the role of *zeste* would be one in which the *zeste* protein can bind simultaneously to both copies of the gene.

To determine whether *zeste* protein can interact simultaneously with two DNA molecules, we first mixed extracts from *E.coli* expressing *zeste* with radiolabelled DNA fragments containing the *Ubx* promoter binding site. We then applied the binding mixture to an affinity column made with a synthetic oligonucleotide containing the consensus binding sequence (Kadonaga and Tjian, 1986). After the binding reaction was loaded, the column was washed with ten column volumes of binding buffer containing cold calf thymus DNA and then eluted with 0.6 M NaCl. As Figure 7 shows, the profile of the radioactivity eluting from the affinity column indicates that a substantial fraction of the counts is retarded and elutes slowly in the column wash while ~10% of the counts are bound more stably and elute only with high salt. As a control we performed a parallel experiment using equivalent amounts of protein extracts from bacteria containing the expression vector but no *zeste* coding sequences. In this case, the labelled DNA is neither bound nor retarded and elutes from the column in the flow-through fractions. These results demonstrate directly that *zeste* protein bound to the radiolabelled DNA is able to bind simultaneously to the affinity column and therefore cross-link two DNA molecules.

Discussion

The DNA sequence of many different *zeste* binding sites indicates that *zeste* protein recognizes a consensus hexanucleotide with the sequence T/CGAGT/CG with a strong preference for T at the pyrimidine positions. Such a hexanucleotide with two two-fold degeneracies would be a relatively frequent sequence, statistically expected on one strand or the other every 512 base pairs. The much lower frequency of *zeste* binding sites actually found is due to the requirement for two or more such consensus sequences with a spacing of 17−50 nucleotides. Multiple binding sites for DNA binding proteins are well known in cases ranging from the lambda operators (Maniatis *et al.*, 1975) to the cluster of SP1 sites in the SV40 early promoter (Dynan and Tjian, 1985). What is unusual in the case of *zeste* is that the affinity for a single consensus sequence is undetectable in our experiments (hence more than an order of magnitude lower than an average site) while the cooperating consensus sequences can be separated by a variable distance without any clear evidence of helical periodicity or orientation specificity. These results suggest not only that stable binding requires cooperativity but that the interaction between *zeste* proteins bound at two consensus sequences is remarkably flexible.

The *zeste* protein used for most of the experiments reported here was in the form of extracts from *E.coli* expressing the *zeste* gene. Affinity-purified *zeste* protein made in *E.coli* or extracted from *Drosophila* appears to behave in an entirely similar way (Biggin *et al.*, 1988) at least with respect to binding and footprinting at the *Ubx* promoter site. Mansukhani *et al.* (1988) have reported binding to what are undoubtedly some of the same sites at *white* and *Ubx* using a *zeste*-β-galactosidase hybrid protein containing as little as 205 of the 575 amino acid residues of the *zeste* protein. It is possible that the β galactosidase moiety, which normally forms tetramers, could contribute to the protein−protein interactions. Another possibility is that some part of the *zeste* polypeptide required for cooperative interactions may lie very close to the DNA binding domain.

Distribution of binding sites

The results presented in this paper and other recent reports on the *zeste* gene and its product (Biggin *et al.*, 1988; Pirrotta *et al.*, submitted) have greatly expanded our conception of what *zeste* might do. It clearly binds to DNA at a large number of sites at many more loci than those predicted by the genetic evidence. In the *white* gene, the binding sites are located primarily in the 5′ regulatory region and are in rough correspondence with known control elements responsible for the expression of the *white* gene in testes, eyes and malpighian tubules (Pirrotta *et al.*, 1985; Levis *et al.*, 1985). The excision of one of these binding sites, *white*-2, in the w^{sp2} mutation results in greatly decreased eye pigmentation and failure to interact with z^1 but no impairment of testis or malpighian tubule pigmentation. The implication that the *white*-2 binding site is associated with an eye-specific determinant has been directly confirmed using a *lacz* gene as a reporter (V.P., unpublished results). The *white* gene then appears to have two eye specific determinants, one very close to the promoter and one more than one kilobase upstream; both of these are associated with *zeste* binding sites.

The other loci we have examined so far are much more complex than *white*. At *Ubx*, mutations with altered regulatory effects are found in the 40 kb preceding the gene as well as within the 75 kb transcription unit and even downstream of the polyadenylation sites. *zeste* binding sites at *Ubx* are similarly spread over a 130 kb region. As at *white*, one binding site is in the immediate vicinity of the transcription start site. Other sites lie in the *Ubx* regulatory regions defined by the *abx*, *bx*, *bxd*, *pbx* and *Cbx* mutations. At *dpp*, the 30 kb following the 3′ end of the transcription unit contain a number of regulatory elements affecting the function of the gene in imaginal discs (Gelbart *et al.*, 1985) and participating in transvection effects (Gelbart and Wu, 1982). We find several *zeste* binding sites in this region as

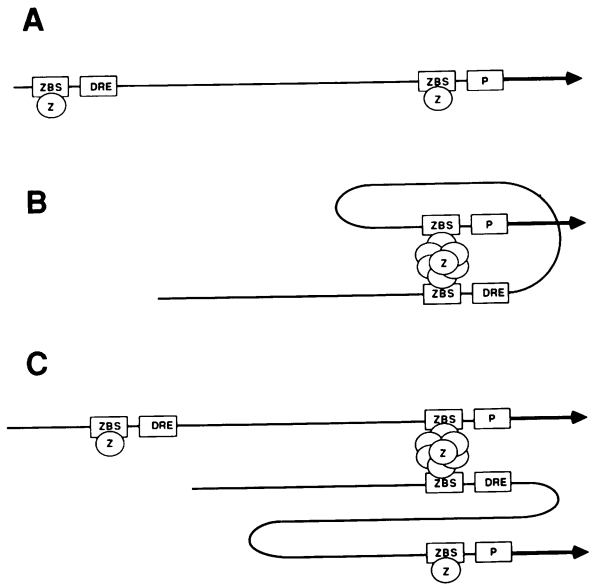

Fig. 8. A model of *zeste* action. (**A**) The promoter P of a transcriptional unit is associated with a *zeste* binding site (ZBS). A second *zeste* binding site is associated with a distant regulatory element (DRE) lying tens of kilobases away from the promoter. (**B**) A multivalent *zeste* protein complex can bind to both the proximal and distal *zeste* binding sites, bringing the DRE to the vicinity of the promoter. (**C**) A similar interaction between the proximal ZBS on one chromosome and the distant ZBS on the synapsed chromosome permits transvection effects.

well as a particularly strong site associated with the promoter of the 4.5 kb transcript. The distribution of *zeste* binding sites at *Antennapedia* and *engrailed* is similarly complex although we have not found in these cases a binding site in close proximity of the promoter as at *white*, *Ubx* and *dpp*. In some cases, the apparent affinity of a DNA fragment for *zeste* might be due to the presence of several, individually very weak sites. These may however collaborate to achieve stable binding, as our results at the *white* promoter site indicate. At present, the evidence for the significance of these *zeste* binding sites is purely circumstantial (except for the *white*-2 site), based on the fact that their distribution at *Ubx* or *dpp* resembles that of regulatory mutations that are known to be involved in transvection phenomena.

Although *zeste* binding sites are abundant in the loci discussed above, they are not ubiquitous and many of the genomic clones we have examined contained no sites with appreciable affinity for *zeste in vitro* (for example, clones containing the *fushi tarazu* gene and its flanking sequences). Using indirect immunofluorescence on salivary gland chromosomes, *zeste* has been localized at at least 60 sites on the euchromatic arms of the chromosomes in late third instar larvae and the evidence suggests that the occupation of potential binding sites might be developmentally regulated (Pirrotta *et al.*, submitted). We suspect therefore that a *zeste* binding site detected *in vitro* might be occupied *in vivo* only in a particular stage or tissue. Some might in fact not be physiologically significant. Conversely, it is possible that sites with low affinity *in vitro* might become significant *in vivo* or *in vitro* in the presence of suitable cofactors.

A possible model of zeste function

Our results suggest the following conclusions: (i) *zeste* binding sites are associated with promoter regions and/or regulatory elements, particularly in genes with complex and extensive regulatory regions; (ii) *zeste* binding requires cooperative interactions; (iii) cooperativity can also be detected between *zeste* bound at distant sites; (iv) *zeste* protein or aggregates of *zeste* can bind simultaneously to two different DNA molecules or two regions of the same molecule. *In vitro* studies show in fact that the *zeste* protein tends to form very large aggregates and that the DNA binding activity is found preferentially in such higher mol. wt forms (S.Bickel and V.Pirrotta, manuscript in preparation). We propose then that the role of *zeste* is to bind in the vicinity of different regulatory elements that may be tens of kilobases distant from one another and to bring them together by means of *zeste*–*zeste* interactions. Alternatively, a large aggregate of *zeste* protein bound at one site might capture a second *zeste* binding site on a loop of DNA from the same gene or from the homologous chromosome as illustrated in Figure 8. In complex regulatory regions such as those of *Ubx* or *dpp*, the binding of *zeste* to a given site might be itself regulated by other factors that would block a potential *zeste* binding site or facilitate *zeste* binding at an otherwise weak site. A suitable series of such factors appearing, for example, at different developmental stages, would allow *zeste* to bind successively to a number of different sites. This would generate the possibility for a programmed looping and folding of long DNA regulatory domains, successively juxtaposing different parts and allowing them in turn to affect RNA polymerase poised at the transcription start site. Thus, in addition to having a direct effect as a promoter-proximal transcription factor (Biggin *et al.*, 1988), *zeste* could also modulate the activity of a gene by allowing distant regulatory elements to control the promoter. Interactions of this sort, as proposed by Biggin *et al.* (1988), could easily accommodate explanations of transvection phenomena simply by allowing the looping to occur not only between regulatory elements within a gene but also between two copies of a gene brought together by chromosome pairing.

Looping mechanisms have been frequently invoked to explain the action of enhancer elements (Dunn *et al.*, 1984; Ptashne, 1986). The looping is generally attributed to protein–protein contacts between a *trans*-acting factor bound at the enhancer and RNA polymerase or transcription factors at the promoter. While these are very likely mechanisms in many cases in which the enhancer is found at a distance of one to several hundred nucleotides from the promoter, more explanation is required in the case of complex genes such as *Ubx* in which the presumed enhancer-like elements may lie as far away as 50 kb from the transcription start. What is there about *zeste* that would justify such long-range looping models? We can adduce three arguments. One is that transvection phenomena clearly indicate that such long-range interactions do occur and that they involve *zeste*. A second argument is that, as we have shown here, *zeste* proteins can bind simultaneously to two DNA molecules or two regions of the same molecule. Thirdly, the *zeste* protein, whether produced in bacteria or extracted from *Drosophila* cells, has a strong tendency to associate, forming large aggregates which are the molecular species responsible for most of the DNA binding activity (S.Bickel and V.Pirrotta, in preparation). Finally, it can be argued that some form of looping mediated by *zeste* or other proteins with similar properties is necessary to explain the specificity of action of distant

regulatory elements. We know that such elements can act on promoters many tens of kilobases distant. We also know of cases in which two genes with different tissue or developmental specificity reside in close proximity without interfering with one another's specific pattern of expression. A protein such as *zeste*, binding at sites made accessible according to a specific program (for example, by means of other factors that block or facilitate the occupancy of a given site), would allow a specific loop to be established between a control element and its corresponding promoter to the exclusion of other promoters in the vicinity.

Needless to say, these models are, at present, purely conjectural. A functional role for *zeste* binding sites remains to be proved. Most critical for any model that proposes an important role for *zeste* in the expression of its target genes is the lack of genetic evidence to indicate that the *zeste* function is essential. Mutations in *zeste* affect its ability to mediate transvection but do not impair the viability of the fly. We have presented arguments for the possibility that none of the available *zeste* alleles are in fact null (Pirrotta *et al.*, 1987) but no understanding of the role of *zeste* will be possible without a study of true *zeste* null mutants.

Materials and methods

DNA binding assays

The procedure was identical to that previously described (Benson and Pirrotta, 1987).

Footprinting

white-1 and *white*-2 footprints. DNA was end-labelled using Klenow DNA polymerase and 2 femtomoles were incubated in ice for 10 min with *E.coli* extract (35 μg protein) containing *zeste* in 50 μl of 40 mM KCl, 10 mM Tris pH 7.6, 0.04 mM EDTA, 5 mM $MgCl_2$, 0.4 mM dithiothreitol and 2% polyvinyl alcohol. After the addition of 50 μl of 10 mM $MgCl_2$, 5 mM $CaCl_2$ at 25°C, 10 μl of 2.5 μg/ml DNase I were rapidly mixed in. Digestion continued for 1 min at 25°C and was terminated with 100 μl of 40 μg/ml tRNA, 0.2 M NaCl, 20 mM EDTA and 1% SDS. The reaction volume was extracted once with phenol–chloroform, ethanol precipitated and analysed on a 7% acrylamide 8 M urea gel.

An alternative protocol was used for some experiments including that shown in Figure 6. DNA was end-labelled with T_4 polynucleotide kinase and [^{32}P-γ]ATP, the amount of protein added was 3.5–42 μg and the binding reaction took place in ice in 20 mM Tris pH 7.6, 0.25 mM EDTA, 1 mM dithiothreitol and 150 mM NaCl. Digestion was done by adding 50 μl of ice cold binding buffer containing 12.5 μg/ml DNase I and either 10 mM $MgCl_2$ and 5 mM $CaCl_2$ or 1 mM $MnCl_2$.

Affinity column

Approximately 25 ng of kinase-labelled DNA was incubated with crude *E.coli* extract (480 μg protein) in 200 μl of 20 mM Tris pH 7.6, 0.25 mM EDTA, 1 mM dithiothreitol, 150 mM NaCl and 50 μg/ml sonicated calf thymus DNA. After binding at 0°C, the reaction volume was passed three times through a 100 μl bed volume affinity column made with oligonucleotides containing the *zeste* consensus binding sequence (Biggin *et al.*, 1988). The columns were washed with a total of 1 ml binding buffer containing 10 μg/ml calf thymus DNA, then eluted with ten 50 μl aliquots of 0.6 M NaCl, 10% glycerol, 1 M dithiothreitol, 10 mM Tris pH 7.6, 1 mM EDTA. All column procedures were done at 4°C.

Acknowledgements

We are grateful to Welcome Bender, Bill Gelbart, Pete Harte, Tom Kornberg, Kerry Kornfeld, Mark Krasnow, Matthew Scott and Robert Weinzierl for providing us with genomic clones, information and sequences. We thank Mark Biggin and members of the Tsai labs for advice on footprinting and Sharon Bickel for help with the oligonucleotide affinity columns. M.B. is an MSTP trainee and the research was supported by a grant to V.P. from the NIH.

References

Babu,P. and Bhat,S.G. (1980) In Siddiqi,O., Babu,P., Hall,L.M. and Hall,J.C. (eds), *Development and Neurobiology of Drosophila*. Plenum Press, New York, pp. 35–38.

Bender,W., Akam,M., Karch,F., Beachy,P.A., Peifer,M., Spierer,P., Lewis,E.B. and Hogness,D.S. (1983) *Science*, **221**, 23–29.

Benson,M. and Pirrotta,V. (1987) *EMBO J.*, **6**, 1387–1392.

Biggin,M.D., Bickel,S., Benson,M., Pirrotta,V. and Tjian,R. (1988) *Cell*, **53**, 713–722.

Bingham,P.M. and Zachar,Z. (1985) *Cell*, **40**, 819–825.

Briggs,M.R., Kadonaga,J.T., Bell,S.P. and Tjian,R. (1986) *Science*, **234**, 47–52.

Coleman,K.G., Poole,S.J., Weir,M.P., Soeller,W.C. and Kornberg,T. (1987) *Genes Dev.*, **1**, 19–28.

Davison,D., Chapman,C.H., Wedeen,C. and Bingham,P.M. (1985) *Genetics*, **110**, 479–494.

Dunn,T.M., Hahn,S., Ogden,S. and Schleif,R.F. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5017–5020.

Dynan,W.S. and Tjian,R. (1985) *Nature*, **316**, 774–778.

Gans,M. (1953) *Bull. Biol. Fr. Belg. Suppl.*, **38**, 1–90.

Gelbart,W.M. and Wu,C.-T. (1982) *Genetics*, **102**, 179–189.

Gelbart,W.M., Irish,V.F., St.Johnston,R.D., Hoffmann,F.M., Blackman,R., Segal,D., Posakony,L.M. and Grimaila,R. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 119–125.

Gunaratne,P.H., Mansukhani,A., Lipari,S.E., Liou,H.-C., Martindale, D.W. and Goldberg,M.L. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 701–705.

Hazelrigg,T. (1987) *TIG*, **3**, 43–47.

Jack,J.W. and Judd,B.H. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1368–1372.

Kadonaga,J.T. and Tjian,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5889–5893.

Kaufman,T.C., Tasaka,S.E. and Suzuki,D.T. (1973) *Genetics*, **75**, 299–321.

Korge,G. (1981) *Chromosoma*, **84**, 373–390.

Kornher,J.S. and Brutlag,D. (1986) *Cell*, **44**, 879–883.

Kuner,J.M., Nakanishi,M., Ali,Z., Drees,B., Gustavson,E., Theis,J., Kauvar,L., Kornberg,T. and O'Farrell,P.H. (1985) *Cell*, **42**, 309–316.

Laughon,A., Boulet,A.M., Bermingham,J.R., Laymon,R.A. and Scott,M.P. (1986) *Mol. Cell. Biol.*, **6**, 6476–6481.

Lee,W., Haslinger,M.K. and Tjian,R. (1987) *Nature*, **325**, 368–372.

Levis,R., Hazelrigg,T. and Rubin,G.M. (1985) *EMBO J.*, **4**, 3489–3500.

Lewis,E.B. (1954) *Am. Natur.*, **88**, 225–239.

Maniatis,T., Ptashne,M., Backman,K., Kleid,D., Flashman,S., Jeffrey,A. and Maurer,R. (1975) *Cell*, **5**, 109–113.

Mansukhani,A., Crickmore,A., Sherwood,P.W. and Goldberg,M.L. (1988) *Mol. Cell. Biol.*, **8**, 615–623.

Mariani,C., Pirrotta,V. and Manet,E. (1985) *EMBO J.*, **4**, 2045–2052.

Micol,J.L. and Garcia-Bellido,A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1146–1150.

O'Connor,M.B., Binari,R., Perkins,L.A. and Bender,W. (1988) *EMBO J.*, **7**, 435–445.

O'Hare,K., Murphy,C., Levis,R. and Rubin,G.M. (1984) *J. Mol. Biol.*, **180**, 437–455.

Pirrotta,V., Steller,H. and Bozzetti,M.P. (1985) *EMBO J.*, **4**, 3501–3508.

Pirrotta,V., Manet,E., Hardon,E., Bickel,S.E. and Benson,M. (1987) *EMBO J.*, **6**, 791–799.

Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.

Ptashne,M. (1986) *Nature*, **322**, 697–701.

Saari,G. and Bienz,M. (1987) *EMBO J.*, **6**, 1775–1779.

Schneuwly,S., Kuroiwa,A., Baumgartner,P. and Gehring,W.J. (1986) *EMBO J.*, **5**, 733–739.

Soeller,W.C., Poole,S.J. and Kornberg,T. (1987) *Genes Dev.*, **2**, 68–81.

Steller,H. and Pirrotta,V. (1985) *EMBO J.*, **4**, 3765–3772.

Stroeher,V.L., Jorgensen,R.M. and Garber,R.L. (1986) *Mol. Cell Biol.*, **6**, 4709–4716.

Zachar,Z., Chapman,C.H. and Bingham,P.M. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 337–346.

Received on June 30, 1988; revised on August 22, 1988