Characterization of the Extended Myb-like DNA-binding Domain of Trithorax Group Protein Zeste

Lisette Mohrmann, Arnoud J. Kal, and C. Peter Verrijzer

From the Department of Molecular and Cell Biology, Center for Biomedical Genetics, Leiden University Medical Center, P.O. Box 9503, 2300 RA Leiden, The Netherlands

Received for publication, March 11, 2002, and in revised form, September 26, 2002
Published, JBC Papers in Press, September 26, 2002, DOI 10.1074/jbc.M202341200

Zeste is a Drosophila sequence-specific DNA-binding protein that performs a variety of functions during chromatin-directed gene regulation. Its DNA-binding domain (DBD) was previously identified, but no similarities to established DNA-binding structures are known. Here we present sequence comparisons suggesting that the Zeste-DBD is a novel variant of the tri-helical Myb-DBD. Using band shift assays, we mapped the Zeste-DBD to 76 residues, corresponding to a single Myb repeat of only 50 residues. All residues involved in formation of the hydrophobic core of the Myb domain are conserved in Zeste, suggesting it forms an extended Myb domain. Mutagenesis studies determined (T/C/g)GAGTG(A/G/c) as the consensus Zeste recognition sequence. Reconstituted transcription experiments established that deviations from this optimal consensus compromise transcriptional activation by Zeste. In addition, flanking DNA is critical because Zeste-DBD binding requires a DNA sequence of minimally 16 base pairs, which is much longer than the consensus site. The DNA flanking the consensus is contacted by Zeste through sequence-independent backbone contacts. Interestingly, hydroxyl radical footprinting revealed that the Zeste-DNA backbone contacts all map to one face of the DNA. We compare the DNA-binding properties of Zeste with those of classical tri-helical DBDs harboring a helix-turn-helix motif and suggest a model for Zeste-DNA recognition.

Recognition of specific DNA sequences is the most fundamental mechanism by which transcriptional regulators are able to function in a gene-specific manner. The Drosophila Zeste protein is a sequence-specific DNA-binding transcription factor that belongs to the trithorax group (trxG) of regulators. The trxG comprises diverse activators that act together with the polycomb group (PcG) of repressors to maintain the correct expression of key developmental genes including the homeotic genes (1–6). Zeste is a transcriptional activator of homoeotic genes, including Ultrabithorax (Ubx), as well as other genes such as white and decapentaplegic (dpp) (7–10). Consistent with a role in the regulation of multiple genes, Zeste is found associated with over 60 sites on polytene chromosomes of the larval salivary gland (11, 12). Zeste recognition elements are found in the promoter areas of its target genes as well as in some polycomb response elements that mediate the maintenance of either gene activity or silencing by trxG and PcG proteins. So far, only three members of the trxG/PcG proteins have been shown to possess sequence-specific DNA-binding activity. In addition to Zeste, this group comprises GAGA, a zinc-finger trxG protein, and the PcG protein Pleiohomeotic, which contains a Krüppel-like zinc-finger DBD (5, 6, 13). These factors are attractive candidate candidate tethering factors that may direct non-DNA binding trxG/PcG proteins to polycomb response elements and promoters of homoeotic genes.

A combination of genetic and biochemical studies has suggested that Zeste activates gene expression by alleviation of chromatin-mediated repression. Zeste loss-of-function mutations are enhancers of position effect variegation, indicating a role in counteracting heterochromatin-mediated silencing (14). Biochemical experiments revealed that Zeste recruits the SWI/SNF-related Drosophila Brahma complex to mediate transcriptional activation on a chromatin template (15). In addition, Zeste shows positive as well as negative genetic interactions with a number of PcG genes (16, 17). More recently, Zeste was identified as a component of a large Pc-containing complex, PRC1 (18). Thus, Zeste may perform dual functions and act as a tethering factor for trxG activators as well as PcG repressors. Moreover, Zeste is involved in chromosome pairing-dependent gene control phenomena. Firstly, Zeste bound to an enhancer on one chromosome is able to activate an allelic promoter on a homologously paired chromosome. This process is referred to as transvection and has been described for loci such as Ubx, dpp, and yellow (19–22). Secondly, particular neomorphic Zeste mutations result in pairing-dependent repression of transcription (23). Taken together, these studies show that Zeste is a multifunctional regulator of chromatin-directed gene regulation.

A unique feature of Zeste is that it binds cooperatively to multiple Zeste elements as a higher-order homo-oligomer (24, 25). Zeste oligomerization is the result of coiled-coil interactions mediated by carboxyl-terminal leucine zipper motifs (Fig. 2A). Interestingly, a Zeste oligomer cannot bind efficiently to a single Zeste site; rather, DNA binding requires at least 4 or more recognition elements. Indeed, natural Zeste-responsive elements such as the Ubx promoter (8) or the Ubx polycomb response elements typically contain multiple Zeste-binding elements. Although the Zeste-DBD (Fig. 2) was identified over a decade ago (24, 26, 27), no similarities with well characterized DNA-binding motifs have been identified so far.

Many DNA-binding proteins can be classified into distinct groups that use a common structural motif for DNA recognition. The first DNA recognition motif discovered, and one of the best characterized, is the helix-turn-helix (HTH) structure (28).
The HTH motif is part of the DBDs of a wide range of distinct gene regulatory proteins found in organisms ranging from bacteria and archaea to man. The second of the two helices that form the HTH motif inserts into the major groove of the DNA where it makes base-specific as well as backbone contacts. Consequently, this helix is referred to as the recognition helix. There is substantial variation in the DNA-binding protein structures harboring an HTH and in the way the recognition helix inserts into the major groove. One of the most common examples of an eukaryotic HTH DBD is the homeodomain. Within the ~60-amino acid homeodomain, the HTH motif is part of a stable tri-helical protein fold with helix 2 and 3 forming the HTH motif. A unique feature of homeodomain-DNA interactions is that, in addition to the insertion of helix 3 into the major groove, DNA contacts are made by a short unstructured amino-terminal arm, which reaches into the minor groove (29).

A distinct class of tri-helical DBDs containing an HTH motif is formed by the Myb domain (30). As indicated by its name, Myb repeats were first recognized in the transcriptional regulator c-Myb, which is involved in the control of cell proliferation and differentiation (31). The Myb DNA-binding domain comprises three imperfect tandem repeats (R1, R2, and R3), each forming a tri-helical protein fold in which an HTH motif is formed by helix 2 and 3 (32, 33). Together, R2 and R3 are sufficient for recognition of specific DNA sequences, whereas R1 enhances the stability of the Myb-DNA complex (34, 35). The three helices in each repeat are maintained by a hydrophobic core that includes three strictly conserved tryptophan residues (Fig. 1) (36, 37). The turn between the second and third helices in each repeat is one amino acid longer than in the classical HTH motif. Another related domain is found in the yeast regulator Rap1p (repressor activator protein 1), which contains a DBD with properties that are somewhat in between those of a homeodomain and a Myb domain. Rap1p is a bifunctional regulator, which can either activate or repress transcription; its binding sites are found in the promoters of a large number of genes as well as at telomeric DNA (38). The Rap1p DBD contains two very similar subdomains that each bind the sequence GGTGT arranged in a tandem orientation. The structure of each of these is very similar to that of the Myb domain, but, like homeodomains, the Rap1p subdomains contain an amino-terminal arm that contacts the minor groove (39). Another protein sequence motif that shows similarity to the Myb domain is the SANT domain, which was first recognized in several proteins involved in transcriptional control including SWI3, ADA2, N-CoR, and TFIIIB (40). Although no DNA-binding activity has been reported for the SANT domain, it has been suggested, based on sequence similarities, that the Myb and SANT domains are folded into a related conformation. Finally, the DBD of the vertebrate telomere repeat-binding factor TRF1 comprises a single Myb domain that suffices for sequence-specific DNA binding (41). Helix 3 of the TRF1 DBD makes base-specific contacts in the major groove, whereas its amino-terminal arm binds within the minor groove (42). Hence, the Myb domain and homeodomain are related tri-helical structural arrangements harboring an HTH motif that binds DNA in a related but distinct fashion.

In this study, we have analyzed the structural and functional properties of the Zeste-DBD. Although previous studies failed to uncover a relationship between the Zeste-DBD and known DNA recognizing structures, our results indicate that the Zeste-DBD is related both to the Myb- and to the homeodomain. We found that a 76-amino acid domain of Zeste, corresponding to a single Myb repeat, mediates sequence-specific DNA binding. Interestingly, the Zeste-DBD is significantly larger than the 60-amino acid homeodomain or the ~50-amino acid Myb domain. However, the strict conservation of key structural residues and structure predictions suggested that the Zeste-DBD might form an extended tri-helical DBD related to the Myb domain and the homeodomain. Using a combination of band shift assays, DNase I, and hydroxyl radical footprinting, we characterized the DNA-binding properties of the Zeste-DBD. The optimal Zeste recognition sequence was identified, and functional assays revealed that changes in DNA-protein interactions affect transcriptional activation by Zeste. In addition to base-specific contacts, sequence-independent backbone contacts were shown to be critical for efficient Zeste binding.
Characterization of the Zeste DNA-binding Domain

EXPERIMENTAL PROCEDURES

Expression and Purification of Recombinant Proteins—To investigate whether the DBD of Zeste based on the homology to Myb/SANT domains (between amino acid 42 and 143) were still able to bind DNA, six different expression constructs were created (indicated in Fig. 2A). DNA fragments encoding the different DBDs were generated using a PCR-based strategy. Primers were designed to generate a BamHI restriction site directly in front of the indicated amino-terminal Zeste amino acid and a stop codon and EcoRI site following the indicated thrombin cleavage site (right, R). These proteins were expressed in Escherichia coli BL21-CodonPlus(DE3)-RIL cells and, after extract preparation, affinity purified on glutathione-Sepharose 4B. All protein procedures were carried out at 4 °C. All protein procedures were carried out at 4 °C using HEMG buffer (25 mM HEPES-KOH, pH 7.6, 0.1 mM EDTA, 12.5 mM MgCl₂, and 10% glycerol) and varying amounts of NaCl. To isolate the recombinant protein the lysis was carried out in HEMG buffer containing 1 mM dithiothreitol, 0.5 mM AEBSF ((2-aminoethyl)benzenesulfonyl fluoride), 1 µM pepstatin, 0.1 mM leupeptin, 0.5 mM sodium metabisulphite, 0.2 mM phenylmethylsulfonyl fluoride, 0.1% Nonidet P-40, and 0.5 mg/ml lysozyme. After incubation for 45 min, a freeze-thaw step was performed, and the cells were sonicated several times. After centrifugation at 15,000 rpm in a SS-34 rotor for 25 min, the tagged proteins were affinity-purified from the supernatant on glutathione-Sepharose 4B. After extensive washes, the recombinant proteins were eluted by thrombin cleavage at room temperature for 45 min. The elution was done in HEMG buffer containing 150 mM NaCl and 1 mM dithiothreitol, giving two extra amino acids (glycine-serine) at the amino terminus of the recombinant proteins. The protein concentrations of the preparations were determined by UV-measurements (all around 0.4 mM) and verified by Coomassie staining. The fractions were aliquoted and stored at −20 °C. Analyses of the affinity-purified DBD proteins by denaturing SDS-PAGE showed a single band with the expected molecular mass, ranging from ~7 to 17 kDa (Fig. 2B, upper panel).

DNA-binding Assays—For band shift assays, double-stranded oligonucleotides (Figs. 3 and 5) were end-labeled with T4 polynucleotide kinase. DNA-binding reactions were performed in 25 µl of binding buffer (0.5 x HEMG, 50 mM NaCl, 0.05% Nonidet P-40) in the presence of 50 nM poly(dIdC)-poly(dIdC), ~150 fmol of labeled probe, and 1.15, 0.58, or 0.29 µM protein. All binding reactions were carried out at room temperature for 25 min and were analyzed on an 8% polyacrylamide gel, run in 0.5 x Tris-glycine/0.1% Nonidet P-40 buffer at room temper-
Characterization of the Zeste DNA-binding Domain

FIG. 4. Determination of the $K_d$ of Zeste-DBD. The double-stranded oligonucleotide 17L (see Fig. 3) was incubated with increasing amounts of Zeste-DBD. Next, the DNA occupancy was determined in a mobility shift assay. The percentage of bound DNA was determined by quantitation of the gel using a Amersham Biosciences PhosphorImager and plotted against the DBD concentration (see "Experimental Procedures"). Under conditions when the total DNA concentration is well below the $K_d$, the $K_d$ approximates the protein concentration at which 50% of the DNA is bound. The concentration of 17L in these reactions was about 6 nM, which allowed us to estimate a $K_d$ of $1.3 \times 10^{-7}$ M, the protein concentration at which 50% of the DNA is bound.

The basal transcription machinery was provided by addition of 0.2 l of RNase inhibitor, and 0.8 mM DDT. Transcription reactions were performed either in the absence or presence of increasing amounts of purified full-length Zeste. Transcription reactions were performed in a total volume of 25 l and 10-fold reduced. These results suggest (T/C/g)GAGTG(A/G/c) as the Zeste consensus recognition sequence (shown at the bottom of the table). The lines flanking the consensus sequence indicate sequence-independent bases, critical for DNA binding by the Zeste-DBD. We used the double-stranded oligonucleotide 17L and various 17L-derived mutants in a band shift assay in the presence of decreasing amounts of Zeste-DBD. A selection of representative autoradiographs are shown. The mutants were grouped into either wild-type (+) affinity, less than 2-fold reduced (+), over 4-fold reduced (−), or more than 10-fold (−) reduced affinity. The lines flanking the consensus sequence indicate sequence-independent bases, critical for DNA binding by the Zeste-DBD. The double-stranded oligonucleotide 17L and various 17L-derived mutants were used to compare the level of Zeste-dependent transcription on a template harboring 4 consensus Zeste elements (wild-type, mut 11 and 23) or templates containing suboptimal Zeste sites (mut 15, 18, 19, 20, 22, and 24). The mutations present in the various templates (indicated above the lanes) are shown in Fig. 5A. Transcription reactions were performed either in the absence or presence of increasing amounts of purified full-length Zeste. Transcription products were detected by primer extension, separated on an 8% denaturing polyacrylamide gel, and subjected to autoradiography. The gels were quantified by phosphorimager analysis, which revealed an approximately 20-fold activation of transcription by Zeste on the templates with optimal binding sites, whereas the level of activation on the mutant templates was between 5–10-fold reduced.
% denaturing polyacrylamide gel. Transcription gels were quantified using an Amersham Biosciences PhosphorImager.

RESULTS

The Zeste-DBD Is Related to the Myb Domain—Like most transcription factors, Zeste has a modular organization, and distinct structural and functional domains can be recognized (Fig. 2A). Among its prominent features are an amino-terminal DBD and a carboxyl-terminal leucine zipper oligomerization domain. The leucine zipper domain mediates coiled-coil interactions that lead to the formation of higher-order Zeste multimers (26). As a result, Zeste binds DNA cooperatively as an oligomer, and a typical natural response element for Zeste contains 4 or more Zeste-binding sites (24). Although the DBD of Zeste was identified and narrowed down to amino acids 47–138 well over a decade ago (27), its relationship with other DBDs remained unresolved. In an attempt to identify potential structural motifs in Zeste, we first identified protein sequences that are highly conserved between Zeste from two distinct fly species, Drosophila virilis and Drosophila melanogaster, and used these regions to search protein databases. A BLAST search of the SWISS-PROT data base revealed a clear relationship between Zeste and the Myb-DBD as well as with the Myb-related SANT domain. Fig. 1 illustrates the primary sequence similarity between the Zeste-DBD and the Myb-DBD obtained following a clustal W alignment and further optimization by manual editing. In addition to an overall sequence similarity, there is strict conservation of critical hydrophobic residues that are involved in the formation of the hydrophobic core of the Myb-DBD. Furthermore, the conserved glycines and prolines at the end of helices 2 and 3 in the Zeste-DBD are conserved, whereas the glycine/proline after helix 1 is not conserved in Zeste. Moreover, a structure prediction using the PredictProtein program (Swiss Institute of Bioinformatics) suggested that the Zeste-DBD may also have a tri-helical structure. Taken together, these observations are consistent with the idea that the Zeste-DBD may form a protein fold similar to that found in the Myb domain. However, all three putative α helices within the Zeste-DBD are extended relative to those present in the Myb domain. Compared with the Myb domain, the putative helices 1, 2, and 3 are respectively 12, 4, and 10 residues longer. It should be noted that helix 1 is extended in the area where the conserved glycine is lacking. Although this hypothesis awaits a three-dimensional structure determination, these findings suggest that the Zeste-DBD forms an extended Myb-like domain.

To determine whether the Myb-related domain in Zeste was sufficient to mediate sequence-specific DNA binding, we expressed the series of polypeptides indicated in Fig. 2A. The various constructs were expressed as GST-fusion proteins in E. coli, and following purification on a Gluthathione-Sepharose matrix the Zeste polypeptides were released from the GST portion by thrombin cleavage. These polypeptides were resolved by SDS-PAGE and visualized by Coomassie staining, revealing comparable concentrations and purity (Fig. 2B, top panel). Next, we compared the ability of the distinct Zeste-DBD polypeptides to recognize and bind a 22 base pair DNA oligomer harboring a single Zeste-binding site derived from the Drosophila Ubx promoter (9). As shown in Fig. 2B (bottom panel), all Zeste polypeptides tested efficiently bound the Zeste site. Thus, the minimal 76-residue Zeste polypeptide Z (51–126) that is homologous to a single Myb repeat is sufficient for DNA binding. Hope-Struhl bandshift assays demonstrated that the Zeste-DBD binds DNA as a monomer (our data not shown) (24). In the DNA-binding experiments described below we used this polypeptide and refer to it as DBD.

Determination of the Minimal DNA Sequence Required for

DBD Binding—To determine the outer borders of the DNA sequence required for Zeste-DBD binding, we generated a series of oligonucleotides that were progressively shortened toward a Zeste site derived from the Ubx promoter (core sequence: CGAGTTG, Fig. 3). We chose an oligonucleotide of 22 base pairs as our longest DNA fragment because it encompassed the Zeste DNase I footprinting borders (data not shown, see also Fig. 6). For reference we numbered the top strand bases from 1 to 22, as indicated in Fig. 3. A series of double-
Characterization of the Zeste DNA-binding Domain

stranded oligonucleotides, which were either 22 base pairs or shortened upstream (left, L) or downstream of the Zeste site (right, R), were labeled and tested for binding in a band shift assay using decreasing DBD concentrations. The autoradiographs were examined, and the mutants were grouped into either wild-type (+ + ) affinity, less than 2-fold reduced (+), over 4-fold reduced (−), and more than 10-fold (− − ) reduced affinity. The shortest oligonucleotide recognized with apparent wild-type affinity (16RL) was 16 base pairs long. Interestingly, the core Zeste recognition sequence CGAGTG is located asymmetrically within this minimal DNA oligomer. Efficient binding by the Zeste-DBD required 6 additional base pairs downstream of the core recognition motif (20R) but only 3 base pairs upstream of the core (17L). Moreover, an oligonucleotide with only 1 flanking base pair upstream of the core (15L) was still recognized by Zeste. In contrast, binding to an oligonucleotide with 4 flanking base pairs downstream of the core was no longer detectable (18R). These results indicated that DNA binding by the Zeste-DBD required extensive contacts with the DNA downstream of the core recognition sequence (Fig. 3).

To determine the equilibrium dissociation constant \( K_d \) of the Zeste-DBD DNA binding, we used the band shift assay. The amount of protein was increased in the presence of a fixed amount of radiolabeled, double-stranded oligonucleotide DNA at a concentration of about 6 nm. The fraction of DNA bound by Zeste-DBD was obtained by computer quantitation of the gel using an Amer sham Biosciences PhosphorImager and plotted against the DBD concentration (Fig. 4). Under conditions when the total DNA concentration is well below the \( K_d \), the protein concentration at which 50% of the DNA is bound approximates the \( K_d \). Using this approach, we obtained a \( K_d \) value of 1.8 ± 10^{-7} M for the longest oligonucleotide (22 base pairs) and a \( K_d \) of 1.3 ± 10^{-7} M for 17L, whereas oligonucleotide 16RL was bound with an estimated \( K_d \) of 2.4 ± 10^{-7} M. These results indicated that the isolated Zeste-DBD bound its recognition sequence with a modest binding affinity of about 1.8 ± 10^{-7} M. As discussed below, oligomerization of full-length Zeste results in highly cooperative binding to multiple binding sites present in its natural response elements (24).

Sequence Determinants for Recognition by Zeste-DBD—Because we previously showed that Zeste-binding elements suggested (T/C)GAG(T/C)G as its consensus recognition sequence (8), because our determination of the minimal DNA sequence required for DBD binding suggested the need for extended DNA contacts, we wondered whether there might be sequence constraints for the positions flanking the consensus. We were particularly intrigued by this possibility because we noted that natural Zeste sites are frequently flanked by a sequence-independent contacts with the DNA backbone, in particular downstream of the consensus site.

To ascertain that changes in Zeste-DNA interactions influence transcriptional activity, we performed reconstituted transcription experiments. We compared the ability of Zeste to activate transcription on a template harboring 4 consensus Zeste elements with its ability to activate a series of templates containing either optimal or suboptimal Zeste sites. For this experiment we made a selection of the mutants analyzed in Fig. 5A, which involve distinct positions within the Zeste consensus sequence. Other than the point mutations in the Zeste recognition sequences, all templates used in the transcription experiments were identical. The general transcription factors were provided by a partially purified Drosophila embryo extract, and full-length Zeste was isolated from SF9 cells infected with recombinant baculoviruses (15). Transcription reactions with each of the 9 distinct templates were performed either in the absence or presence of increasing amounts of purified full-length Zeste, and transcription was monitored by primer extension. As shown in Fig. 5B, Zeste strongly stimulated transcription of the templates with optimal Zeste-binding sites (wild-type, mut 11 and 23). In contrast, activation of transcription by Zeste on the various mutant templates (mut 15, 18, 19, 20, 22, and 24) was clearly decreased (about 5–10-fold reduced) compared with the templates with optimal Zeste-binding sites. It should be noted that the transcription templates contain multiple Zeste sites, allowing for cooperative DNA binding. Therefore, mutations that impair Zeste binding could still support weak transcriptional activation (e.g. mut 15 and 20). Nevertheless, all mutations in the recognition sequence that affect Zeste-DNA interactions compromise the ability of Zeste to activate transcription. Thus, we conclude that transcriptional activation by Zeste is critically dependent on its recognition sequence.

Identification of Zeste-DBD DNA Backbone Contacts—Our analyses suggested that, in addition to essential sequence-specific contacts with the core recognition element, sequence-independent contacts with the DNA backbone play an important role during Zeste DNA binding. To determine the DNA sequence contacted by the Zeste-DBD, we first performed a DNase I footprinting analysis using a 51-base pair oligomer containing the Zeste recognition site and flanking sequences used in previous experiments (Fig. 3). As shown in Fig. 6A, Zeste-DBD protects 11 bases on the top strand (A7–G10) on the top strand is protected, −T7) and a larger, strongly positive footprint (G20), respectively, from DNase I digestion. Because DNase I is a bulky, asymmetrical enzyme, its use in footprinting reactions provides only a low-resolution probe for protein-DNA contacts. To examine Zeste-DBD DNA contacts with a higher resolution we used hydroxyl radical footprinting analysis (Fig. 6B). Hydroxyl radicals, generated by iron(II)-promoted reduction of hydrogen peroxide, attack the deoxyribose sugars in the DNA backbone. Sugars that are in close contact with protein are completely or partially protected against cutting. Thus, this technique identifies contacts with the sugar/phosphate backbone of DNA. The hydroxyl radical footprints of Zeste-DBD revealed only a small protected region on the top strand and two distinct regions, separated by about 5 base pairs, on the bottom strand (Fig. 6B). The bands were quantified by phosphorimaging analysis, and the relative amounts of label present at the various positions were plotted in a graph (Fig. 6C). The quantification confirmed the striking difference between the top and bottom strand with respect to the extent of backbone contacts made by Zeste-DBD. Although a single small region (A3–G10) on the top strand is protected, there are two separate footprints on the bottom strand: a small, partially protected region (T9–T7) and a larger, strongly pro-
Zeste-DBD are discussed below. Recognition by Zeste. Models to explain DNA binding by the lap with the positions that are critical for sequence-specific radical footprinting cover 13 base pairs and only partially overlap with the positions that are critical for sequence-specific recognition by Zeste. Models to explain DNA binding by the Zeste-DBD are discussed below.

DISCUSSION

We have identified a structural relationship between the DBDs of Zeste and Myb. Sequence alignments revealed a clear similarity between the Zeste-DBD and the Myb repeats and SANT domains. In particular, all the key residues involved in the formation of the hydrophobic core of Myb are strictly conserved in the DBD of Zeste (Fig. 1). However, whereas Myb requires minimally two of its three repeats for sequence-specific DNA binding (44, 45), the minimal DBD of Zeste corresponds to a single, extended Myb repeat. Structure predictions suggest that, like Myb, the DBD of Zeste forms a tri-helical fold. However, all three helices in Zeste appear to be significantly extended compared with Myb and Myb-related DBDs, such as those of Rap1p, TRF1, and homeodomains (Fig. 1 and data not shown). All these proteins comprise a tri-helical structure, with helix 2 and 3 forming an HTH motif in which helix 3 mediates specific base sequence recognition through major groove contacts.

The manner of disposition of the recognition helix within the major groove is distinct among Myb-related DBDs. For instance, the recognition helices of the scRap1p DBD domains are aligned at almost right angles to the DNA axis (39), whereas those of the Myb-protein essentially follow the path of the major groove (32). Moreover, the recognition helix in homeodomain DBDs is tilted toward the direction of the major groove so that the helix lies along the floor of the groove (46). It is unclear how Zeste docks on the DNA; however, both sequence conservation and structure predictions suggest that Zeste forms an extended Myb-like DBD with helix 2 and 3, forming an elongated HTH motif that interacts with the major groove. The elongated helix 1 of Zeste might reach further than the amino-terminal arm of Myb, TRF1, or Rap1p, which would be consistent with the remarkably extended DNA contacts observed in this study.

Our results also suggest further differences in the way that Zeste and the other Myb domain proteins interact with their recognition sites. DNA mutagenesis studies revealed (T/C/G)GAGTG(A/G/c) as the consensus Zeste recognition sequence. This sequence does not appear to be related to the Myb (47), Rap1p (39), and TRF1 (41, 42) recognition sites. Moreover, in addition to the sequence-specific base recognition, critical non-sequence-dependent contacts are made with the flanking sequences. Surprisingly, Zeste binding requires particularly extended non-sequence-specific contacts downstream of the consensus sequence, whereas only a few positions directly upstream are important. Supporting asymmetrical docking of the Zeste-DBD to DNA, DNase I footprinting showed a relatively small footprint on the top strand (11 base pairs) compared with a large (17 base pairs) protected region on the bottom strand. Moreover, the high-resolution hydroxyl radical footprinting revealed one small footprint on the top strand but two separate footprints on the bottom strand. Taken together, these results suggest that Zeste docks asymmetrically on its recognition element.

The protein contacts with the backbone sugars that we identified using hydroxyl radical footprinting (Fig. 6B) as well as the bases critical for sequence recognition by Zeste (Fig. 5) are indicated on a DNA double helix structure (Fig. 7). For comparison, we have included the backbone (phosphate) and base contacts made by c-Myb repeat 2, as identified by Ogata et al. (32). The base-specific contacts are indicated in blue, and both the sugar and phosphate contacts are indicated in red. It should be noted that the Zeste DNA contacts map to one face of the DNA and are more spread out than the c-Myb repeat 2 contacts. The Zeste DNA contacts cover 13 base pairs of DNA compared with 9 base pairs contacted by Myb repeat 2 (32).

In summary, our results indicate that the Zeste-DBD is a special variant of the Myb domain DNA recognition fold. Not only is it extended compared with other tri-helical DBDs such as the homeodomain and Myb domain, but there also appears to be significant differences in the way the DNA is contacted. It will be of great interest to determine the three-dimensional
structure of the Zeste-DBD DNA recognition site to reveal the structural basis of this interaction.

The striking positioning of the Zeste DNA contacts on one side of the DNA helix may have important functional consequences for the ability of Zeste to interact with the chromatin template. Previously, we obtained data suggesting that Zeste was able to efficiently bind to a chromatinized site in the template. Our future studies will be aimed at testing this possibility, to allow Zeste binding to nucleosomal DNA, would be that Zeste contacts the DNA double helix facing away from the histones. Our future studies will be aimed at testing this hypothesis and determining whether Zeste may first bind chromatin and subsequently recruits chromatin-remodeling factors, such as the Brahma complex, to mediate further opening of the chromatin structure. This, in turn, may allow binding of other transcription factors. Such studies may provide further insight in the molecular mechanism by which Zeste regulates gene expression.

Acknowledgments—We thank Karin Langenberg and Tokameh Mahmoudi for assistance with the transcription experiments and Eric Kalkhoven, Jan van der Knaap, Jennifer Roha, and Lee Fradkin for critical reading of the manuscript.

REFERENCES
1. Kennison, J. A. (1995) *Annu. Rev. Genet.* 29, 289–303
2. Faro, R. (1995) *Trends Genet.* 11, 295–297
3. Pirrotta, V. (1998) *Cell* 93, 333–336
4. Lyko, F., and Paro, R. (1999) *Bioessays* 21, 824–832
5. Francis, N. J., and Kingston, R. E. (2001) *Nat. Rev. Mol. Cell. Biol.* 2, 409–421
6. Mahmoudi, T., and Verrijzer, C. P. (2001) *Oncogene* 20, 3055–3066
7. Benson, M., and Pirrotta, V. (1987) *EMBO J.* 6, 1387–1392
8. Biggin, M. D., Bickel, S., Benson, M., Pirrotta, V., and Tjian, R. (1988) *Cell* 53, 713–722
9. Biggin, M. D., and Tjian, R. (1988) *Cell* 53, 699–711
10. Laney, J. D., and Biggin, M. D. (1992) *Genes Dev.* 6, 1531–1541
11. Pirrotta, V., Bickel, S., and Mariani, C. (1988) *Genes Dev.* 2, 1839–1850
12. Rastelli, L., Chan, C. S., and Pirrotta, V. (1993) *EMBO J.* 12, 1513–1522
13. Brown, J. L., Mucci, D., Whiteley, M., Dirksen, M. L., and Kassotis, J. A. (1998) *Mol. Cell* 1, 1057–1064
14. Judd, B. H. (1995) *Genetics* 141, 245–253
15. Kal, A. J., Mahmoudi, T., Zak, N. B., and Verrijzer, C. P. (2000) *Genes Dev.* 14, 1058–1071
16. Phillips, M. D., and Shearn, A. (1990) *Genetics* 125, 91–101
17. Pelegri, F., and Lehmann, R. (1994) *Genetics* 136, 1541–1553
18. Saura, A. J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R. E. (2001) *Nature* 412, 655–660
19. Lewis, E. B. (1954) *Am. Nat.* 88, 225–239
20. Benson, M., and Pirrotta, V. (1988) *EMBO J.* 7, 3907–3915
21. Dorsett, D. (1999) *Curr. Opin. Genet. Dev.* 9, 505–514
22. Pirrotta, V. (1999) *Biochim. Biophys. Acta* 1424, M1-M8
23. Rosen, C., Dorsett, D., and Jack, J. (1998) *Genetics* 148, 1865–1874
24. Chen, J. D., and Pirrotta, V. (1993) *EMBO J.* 12, 2075–2083
25. Chen, J. D., and Pirrotta, V. (1993) *EMBO J.* 12, 2061–2073
26. Chen, J. D., Chan, C. S., and Pirrotta, V. (1992) *Mol. Cell. Biol.* 12, 598–608
27. Mansukhani, A., Crickmore, A., Sherwood, P. W., and Goldberg, M. L. (1988) *Mol. Cell. Biol.* 8, 615–623
28. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.* 61, 1053–1095
29. Wolberger, C. (1996) *Curr. Opin. Struct. Biol.* 6, 62–68
30. Gabrielsen, O. S., Sentenac, A., and Fromageot, P. (1991) *Science* 253, 1140–1143
Characterization of the Extended Myb-like DNA-binding Domain of Trithorax Group Protein Zeste
Lisette Mohrmann, Arnoud J. Kal and C. Peter Verrijzer

J. Biol. Chem. 2002, 277:47385-47392. doi: 10.1074/jbc.M202341200 originally published online September 26, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M202341200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 51 references, 14 of which can be accessed free at http://www.jbc.org/content/277/49/47385.full.html#ref-list-1