The *Drosophila* doublesex proteins share a novel zinc finger related DNA binding domain

Scott E. Erdman and Kenneth C. Burtis

Department of Genetics, University of California, Davis, CA 95616, USA

Communicated by R. Nöthiger

The *doublesex* gene of *Drosophila melanogaster* is the final member of a well characterized hierarchy of genes that controls somatic sex determination and differentiation. The male-specific and female-specific doublesex polypeptides occupy a terminal position in the hierarchy, and thus regulate those genes responsible for the development of sexually dimorphic characteristics of the fly. To investigate the molecular mechanism by which these two related proteins interact with specific target genes, we have identified and characterized their DNA binding domains. Using gel mobility shift experiments with sequentially deleted polypeptides, site-directed mutagenesis and spectrophotometric assays, we have shown that the two doublesex proteins share a common and novel zinc finger-related DNA binding domain distinct from any reported class of zinc binding proteins. We have further shown that of 10 null *dsx* alleles, six encode proteins deficient in DNA binding activity, and that three of these alleles are the result of mutations that alter cysteine and histidine residues in the metal binding domain. Our results provide evidence that both the male-specific and female-specific doublesex proteins share and depend upon the same DNA binding domain for function *in vivo*, suggesting that both proteins bind to, but differentially regulate, a common set of genes in both sexes.

Key words: *doublesex/Drosophila melanogaster*/sex determination/zinc finger

Introduction

Like many other organisms, adult *Drosophila* are sexually dimorphic with respect to a wide variety of morphological and physiological characteristics. Sex-specific differences include a number of externally visible characteristics, such as the morphology of the genitalia and analia, differences in pigmentation of the abdominal cuticular segments, and morphology of certain bristles on the foreleg. Examples of sexually dimorphic gene expression have also been characterized, such as the female-specific transcription of the *yolk protein* genes (Belote *et al.*, 1985). A particularly interesting developmental problem lies in how the multiple genetic functions responsible for generating a wide variety of sex-specific differences are coordinately controlled in a sex-specific manner.

In *Drosophila* a well characterized hierarchy of genes is known to control somatic sex determination and differentiation (Baker, 1989; Cline, 1989; Slee and Bownes,

1990; Steinmann-Zwicky *et al.*, 1990). The *doublesex* (*dsx*) gene occupies a unique position at the end of this hierarchy of genes. While the genes preceding *dsx* in the hierarchy serve to read and interpret the X chromosome to autosome ratio, the primary determinant of sex in *Drosophila*, and subsequently to transmit this information down the hierarchy, *dsx* acts as the immediate regulator of the genes whose products lead to terminal sexual differentiation. Absence of *dsx* function leads to the unregulated differentiation of somatic sexual tissues appropriate to both sexes, resulting in an intersexual phenotype (Hildreth, 1965), rather than the transformations from one sex to another observed with mutations in genes acting above *dsx* in the regulatory hierarchy. Genetic studies have suggested that the cause of the intersexual *dsx* phenotype may be the simultaneous expression of both the male and female sets of somatic sexual differentiation genes, thereby causing mutant cells to acquire sexually intermediate fates (Baker and Ridge, 1980). Thus, the role of the *dsx* gene product in males appears to be prevention of the expression of female-specific differentiation functions, while the role of the *dsx* gene product in females is conversely to prevent the expression of male-specific differentiation functions. Control of these downstream or target genes by *dsx* can be either continuous or transitory (reviewed by Wolfner, 1988; Burtis and Wolfner, 1992). The specific molecular mechanisms of this control remain unknown in most cases.

A direct molecular interaction between the *dsx* gene products and a set of target genes known to be regulated by *dsx* has been recently reported (Burtis *et al.*, 1991). Both the male-specific *dsx* protein (DSX^M) and female-specific *dsx* protein (DSX^F) produced in *E. coli* are able to bind *in vitro* to specific DNA sequences that direct sex-specific transcription of the *yolk protein 1* (*Yp1*) and *yolk protein 2* (*Yp2*) genes in the adult fat body. This finding established that the *dsx* proteins act as DNA binding transcription factors, and raised two key questions: how do the *dsx* proteins bind DNA, and how does this binding regulate gene expression? One essential component of the answer to these questions is a structural and functional characterization of the two proteins.

Previous studies characterizing the *dsx* gene products have shown that the male-specific and female-specific *dsx* proteins share common amino acid sequences (Burtis and Baker, 1989). Sex-specific transcripts produced by alternative splicing encode *dsx* proteins with identical N-termini comprising 397 amino acids, and sex-specific C-termini of either 152 (DSX^M) or 30 (DSX^F) amino acids. Inspection of the primary sequence of these proteins failed to reveal any convincing similarities to previously characterized DNA binding proteins, leaving the questions: where in the *dsx* polypeptides is the DNA binding domain, and what is the biochemical nature of this domain?

In this study we report a characterization of the DNA binding domain of the DSX^M and DSX^F polypeptides. We

Fig. 1. (A) Gel mobility shifts of progressively larger C-terminal deletions of dsx proteins. Lanes are labeled by the name of the expression construct used to prepare protein extracts. Details of plasmid construction are in Materials and methods. Lanes 1 and 2 depict shifts using extracts produced from cells expressing undeleted full length male and female dsx coding sequences, respectively. Extract in lane 3 was from cells bearing only the original T7-7 vector without a dsx insert. Lanes 4–11 contain deletion protein extracts as labeled (refer to Figure 2 for details). (B) Gel mobility shifts of increasingly more extensive N-terminal deletions. Lanes 1, 2 and 3 are as in (A). Lanes 4–10 contain deletion extracts as labeled (see Figure 2). The probe is a 33 bp oligonucleotide containing a high affinity binding site for the DSX proteins found in the fat body enhancer of the *Yp1* and *Yp2* genes (Materials and methods).

demonstrate that the domain is located in the common region found in both polypeptides. Evidence is presented that the dsx DNA binding domain represents a novel zinc finger related DNA binding domain, distinct from the three major classes of zinc finger DNA binding domains (Harrison, 1991). By site-directed mutagenesis of specific cysteine residues in this domain, we establish that these residues are in fact critical to DNA binding. We demonstrate *in vitro* that a metal cation is associated in a sulfhydryl-dependent manner with the functional dsx DNA binding domain, and find that this association is disrupted by replacement of specific cysteine and histidine residues. Characterization of the DNA binding activities and protein sequence alterations of several null dsx alleles indicates the importance *in vivo* of key cysteine and histidine residues within the dsx DNA binding domain. These results provide evidence that both DSX^M and DSX^F share and depend on the same DNA binding domain for function *in vivo*. Models are presented to explain how two proteins with identical DNA binding domains might mediate distinct regulatory effects on individual target genes that must underlie sex-specific differentiation.

Results

DNA binding activity resides in the common N-terminal domain of the male-specific and female-specific dsx polypeptides

To identify the N-terminal and C-terminal boundaries of the minimal functional dsx DNA binding domain, we generated a group of expression vector constructs encoding dsx polypeptides serially truncated from either end. Overexpression of these truncated polypeptides in *Escherichia coli* was accomplished using the pT7-7 vector (Tabor and

Richardson, 1985). Extracts prepared from cells overexpressing each truncated polypeptide were tested for the presence of sequence-specific DNA binding activity. Binding was detected by a gel mobility shift assay using a radioactively labeled 33 bp oligonucleotide containing the high affinity dsx binding site from the fat body enhancer of the *Yp1* and *Yp2* genes as a target (Burtis *et al.*, 1991).

Figure 1A displays the mapping of the C-terminal border of the dsx DNA binding domain. Deletion clones C3-1 and C6-45 (which was subsequently determined by sequence analysis to be identical to deletion C6-27) define the C-terminal boundary of the functional dsx DNA binding domain to a region of 25 amino acids. These 25 residues and the 10 residues located immediately N-terminal to them comprise a segment of the polypeptide predicted by protein structure computer algorithms to adopt an extended α -helix (Chou and Fasman, 1978; Devereux *et al.*, 1984). The elimination of binding activity in clones C6-45 and C6-27 which prematurely terminate this potential helix suggests the importance of this region for DNA binding by dsx.

Figure 1B shows the effect of progressive N-terminal deletions of DSX^M. Deletion points N3C and N4A thus define the N-terminal boundary of the dsx DNA binding domain, and indicate that no additional DNA binding domains capable of binding the target sequence used lie C-terminal to the boundary defined by deletion clone C3-1 in DSX^M. A combined summary of our deletion analyses of the dsx proteins to localize their DNA binding domain are presented in Figure 2. The minimal DNA binding domain is located entirely within the shared N-terminal region of the male- and female-specific proteins. We discuss here features of this domain pertinent to our further characterization of the DNA binding activity.

Fig. 2. Summary of the DNA binding activities of DSX wild type and deletion proteins as determined in Figure 1A and B. The stippled regions indicate the sex-specific C-termini of the DSX^M and DSX^F proteins, while the solid region denotes the minimal DNA binding domain as defined by deletions that retain activity. The relative lengths of the various protein domains are depicted to scale. The endpoints of the critical deletions defining the minimal binding domain are as follows (identified by the last amino acid residue remaining as numbered by Burtis and Baker, 1989): ΔC3-1 (Glu105), ΔCC6-45 (Gln80), ΔN2A (Arg39), ΔN3C (Pro41) and ΔN4A (Cys44).

Fig. 3. Amino acid sequence of the *dsx* DNA binding domain. The 66 amino acid sequence of the minimal DNA binding domain is shown. Missense mutations that abolish DNA binding are indicated by arrows and corresponding amino acid replacements. Site-directed mutations (SDM) are indicated above the sequence. Protein and DNA sequence alterations characterizing mutant *dsx* alleles are illustrated below the sequence. The striped bar identifies a predicted α -helical region (see text). Black bars indicate the position of amino acids deleted by two *dsx* alleles (*dsx*^{M+R40} introduces a frame shift as well). Numbers indicate the positions of amino acids with respect to the common sequence shared by the *dsx* proteins (Burtis and Baker, 1989).

Figure 3 presents the amino acid sequence of the 66 amino acid minimal DNA binding domain identified by the deletion experiments. Inspection of the sequence yields no exact homologies to any known DNA binding motifs. However, the location and large number of cysteine and histidine residues found in the N-proximal half of the domain suggest the possibility of a zinc-coordinating structure (Berg, 1986). Additionally, the defined region contains a high proportion

of basic residues, a property common to many DNA binding motifs. No precise alignments of conserved residues in this portion of the domain to any known classes of zinc-containing DNA binding proteins (Harrison, 1991 and references therein) were found. Nevertheless, the occurrence of paired cysteine and histidine residues separated by a region containing hydrophobic residues (although in non-canonical positions relative to known zinc finger proteins), led us to

Fig. 4. Mutational effects on activity of the *dsx* DNA binding domain. DNA binding activities of the GST–DSX fusion proteins were assayed by gel mobility shift analyses. The fastest migrating DNA fragment corresponds to the fat body enhancer (FBE) of the *Yolk protein 1* and 2 genes and contains several binding sites for the DSX proteins. The wild type *dsx* DNA binding domain interacts specifically with this fragment, thereby shifting its mobility in the gel [compare lane 1 (Canton S wild type) with lane 16 (no protein added)]. Lanes 2–5 demonstrate the elimination of sequence-specific DNA binding by site-directed mutations located at specific cysteine residues within the *dsx* polypeptide. Lanes 6–15 show the activities of the DNA binding domains of a series of *dsx* mutant alleles. Lanes are labeled with the names of each mutation. Site-directed mutations (SDM) are labeled according to the name and position of the residue affected followed by the new residue introduced at that position; previously existing *dsx* mutations are identified according to their original names. The location and nature of alterations resulting in inactive DNA binding domains are shown in Figure 3.

the hypothesis that the *dsx* DNA binding domain might represent a novel zinc finger related DNA binding structure. Initial results showing that the DNA binding activity was sensitive to EDTA (K.C.Burtis, unpublished observations) were supportive of the presence of a metal-dependent domain, and led us to study this domain of the protein in greater detail. The predicted α -helical region (amino acids 71–104) adjacent to this apparent metal binding domain was also determined to be essential for DNA binding activity. This region may serve either a structural role in the folding of the DNA binding domain, or perhaps be an integral part of the DNA binding domain itself.

Site-directed mutagenesis identifies specific amino acids critical for DNA binding activity

To investigate the functional importance of the putative zinc finger related domain to DNA binding, two specific residues of the *dsx* polypeptide, cysteines 47 and 68, were altered by site-directed mutagenesis [see Figure 3 for the locations of these residues within the DNA binding domain; Burtis and Baker (1989) for the complete amino acid sequence of the *dsx* polypeptides]. These residues were chosen on the basis of their locations relative to histidines and other cysteines. If the *dsx* zinc finger was to follow even the least stringent criteria ($C-X_2-4-C-X_2-15-a-X_2-4-a$, where 'a'

represents either cysteine or histidine) for potential cysteine/histidine metal binding domains (Berg, 1986), then these residues should participate in any group of four residues tetrahedrally coordinating metal ions.

Site-directed mutagenesis was accomplished by PCR with degenerate oligonucleotides (Materials and methods). Mutant *dsx* DNA binding domains (including *dsx* amino acid residues 1–106) were subcloned in-frame into the *E.coli* expression vector pGEX-KG (Guan and Dixon, 1991) and glutathione-S-transferase–DSX fusion proteins prepared (Materials and methods). A gel mobility shift assay was used to determine the effect of specific mutations. As Figure 4 (lanes 2–5) indicates, amino acid replacements at both Cys47 and Cys68 eliminated sequence-specific DNA binding activity of the corresponding mutant proteins. Thus, both cysteine residues 47 and 68 are essential for this function.

The inability of the polypeptides with either alanine or histidine substituted for Cys47 to bind DNA in a sequence-specific manner is particularly informative. Substitution of alanine for Cys47 is a relatively conservative structural change, as both amino acids are of similar size, charge and hydrophobicity. Significantly, however, alanine is unable to participate in the coordination of metal ions, supporting a role for Cys47 in this process. The inability of histidine, which is capable of metal ion coordination, to substitute functionally for Cys47 suggests additional structural constraints on the domain. This substitution results in the apposition of two histidine residues at positions 47 and 50. It has been previously noted that paired histidines involved in zinc coordination are consistently separated by at least three amino acids in the primary sequence, and suggested that this spacing is necessary to accommodate the relatively bulky histidine residues (Severne *et al.*, 1988). This may explain the failure of the C47H polypeptide to display DNA binding activity.

Four *dsx* null alleles have altered amino acid sequences in the DNA binding domain

The importance of the zinc finger-related DNA binding domain *in vivo* is reflected in the nucleotide sequence of several mutant *dsx* alleles. Reasoning that one way in which null *dsx* mutations might inactivate both the male-specific and female-specific proteins would be to alter the shared DNA binding domain, we chose 10 null alleles from existing *dsx* mutations (described in Lindsley and Zimm, 1992; A.Garen, unpublished) for further characterization. The alleles used included both X-ray and EMS induced mutations that were cytologically normal (Baker *et al.*, 1991).

We obtained by PCR amplification the regions encoding the DNA binding domains (residues 1–106 of the *dsx* polypeptide) from flies bearing individual *dsx* alleles in either homozygous condition or heterozygous over *Df(3R)dsx^{D+R2}* (a chromosomal deficiency that removes the entire *dsx* locus). Following amplification the DNA was subcloned in-frame directly into the pGEX-KG expression vector and fusion proteins were affinity purified. Glutathione-S-transferase–DSX fusion proteins were then assayed for sequence-specific DNA binding activities using a gel mobility shift assay.

The DNA binding domain encoded by each mutant allele was tested for the presence of sequence-specific DNA binding activity. Figure 4 shows the results of this screen. Four alleles (lanes 6–9) in addition to the Canton S wild

type control (lane 1) proved to encode functional DNA binding domains, while six alleles were identified whose sequences failed to encode polypeptides with sequence-specific DNA binding activity (lanes 10–15). The amplified fragments from the latter alleles were subsequently sequenced. Of the six null alleles, four possessed missense mutations in the DNA binding domain, while two (dsx^{M+R40} and dsx^{127-1}) were deletions encompassing portions of the DNA binding domain. Figure 3 illustrates the nature and locations of alterations encoded by the DNA binding null alleles.

Strikingly, three of the four missense mutations ($dsx^{100.41}$, dsx^{128-1} , dsx^{122-1}) altered either cysteine or histidine residues, suggesting that these amino acids play a critical role in the structure of the dsx DNA binding domain. While the protein sequence alterations encoded by these mutant alleles do not prove direct functions of these residues in DNA binding *per se*, their locations are intriguing. It seems unlikely that the occurrence of three mutations at cysteine and histidine residues within the zinc finger related region is a coincidence, since mutations generally disrupting the folding of the domain might be expected to be distributed more uniformly within the region. We therefore interpret these results as further evidence of a zinc finger-related structure contained within the dsx DNA binding domain. Similar patterns of mutations targeting residues demonstrated to be involved in metal ion coordination have been observed for other zinc finger DNA binding domains (Blumberg *et al.*, 1987; Schena *et al.*, 1989).

The remaining mutation (dsx^{EFK43}) alters a residue in the putative α -helical region discussed above. Although the resolution of our deletion mapping data was insufficient to demonstrate the necessity of the entire α -helical region for DNA binding activity, the existence of a point mutation that abolishes DNA binding activity provides strong evidence for a central role for this region in DNA binding. This mutation may identify a residue involved in the maintenance of local α -helical structure. Alternatively, it may alter a residue involved in making specific DNA contacts, perhaps in the context of a recognition helix.

The dsx proteins contain a metal binding domain

To provide additional evidence that the cysteine/histidine rich region of the dsx polypeptides functions as a zinc binding domain, we used a spectrophotometric detection system to assay the presence of sulfhydryl-coordinated heavy metal ions (Jezorek and Freiser, 1979; Hunt *et al.*, 1985; see Materials and methods). This assay, based upon changes in absorbance of the metallochromic dye 4-(2-pyridylazo) resorcinol (PAR), is a sensitive indicator of several heavy metal ions, including Zn^{2+} , Ni^{2+} , Hg^{2+} and Co^{2+} . In the cases where atomic absorption experiments with DNA binding proteins have identified bound heavy metals, the cation present is almost always zinc, and we will assume that the same is true for dsx. We have used this assay to compare the amounts of sulfhydryl-coordinated zinc present in wild type and various mutant *doublesex* polypeptides (Figure 5).

Comparison of the amounts of PMPS-releasable zinc associated with the dsx DNA binding domain fusion protein (WT) and the protein encoded by the vector alone (GST) indicates the presence of cysteine-coordinated zinc specifically associated with the dsx fusion protein. Using excess

Fig. 5. Relative concentrations of sulfhydryl-coordinated divalent metal cation in wild type and mutant dsx DNA binding domains. Solid bars depict the absorbance change at 500 nm resulting from the PMPS-induced release of divalent metal cation from dsx polypeptides in the presence of the metallochromic indicator dye PAR (Jezorek and Freiser, 1979; Hunt *et al.*, 1985). The identity of each protein tested is indicated below the abscissa. The first two bars depict the values for glutathione-S-transferase (GST) and the wild type GST-dsx DNA binding domain fusion protein (WT). Mutant polypeptides are labeled with the name and position of the residue affected followed by the new residue at that position. Mutant polypeptides represent the result of either site-directed mutagenesis (C47A and C68Y), or isolation of sequences from existing dsx mutations $dsx^{100.41}$ (H50Y), dsx^{128-1} (H59Y), dsx^{122-1} (C70Y), and dsx^{EFK43} (R91Q).

PAR and PMPS, and assuming $\Delta\epsilon = 6.6 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$ at 500 nm (Hunt *et al.*, 1985), we have calculated for one preparation of dsx fusion protein a value of $\sim 1\text{--}2$ mol of zinc per mole of protein (K. Burtis, unpublished data). However, this value must be considered a preliminary estimate due to uncertainties in the protein concentration resulting from the presence of proteolytic breakdown products in the sample. Nevertheless, this estimate is consistent with the potential number of coordination ligands present within the domain.

Several mutations that eliminate DNA binding activity substantially reduce the amounts of zinc associated with these mutant proteins relative to the wild type protein (Figure 5), suggesting participation of the altered residues in zinc coordination. However, these mutant proteins in some cases retain intermediate quantities of PMPS-releasable zinc, greater than the background level of the GST polypeptide encoded by the vector alone (compare wild type, C47A, H50Y, C68Y, C70Y and GST; see figure legend for allele identities). We have noted some variability in the relative concentration of zinc found in different preparations of the proteins displaying intermediate values, but values are always substantially reduced from wild type levels (S. Erdman, unpublished data). One explanation for the residual zinc in these proteins is that the remaining unaltered ligand groups, perhaps with the assistance of other cysteine or histidine residues present in the domain, may still bind zinc with low affinity. The resulting domains, however, may be improperly folded and in any event are no longer able to bind DNA. A precedent for this hypothesis exists in experiments involving the adenovirus E1A protein (Webster *et al.*, 1991), in which a directed mutation of one of four zinc coordinating cysteine residues resulted in a protein that was completely non-functional with respect to transactivation activity, but which still bound high levels of zinc. Zinc binding was

maintained by recruitment of two histidine residues into a novel coordination domain with two of the remaining unaltered cysteine residues.

Interestingly, two mutant proteins (H59Y, R91Q) contain near wild type levels of zinc (Figure 5). One of these mutations, R91Q, is located in the putative α -helical region adjacent to the finger related domain. The presence of nearly wild type levels of zinc in this mutant protein suggests that organization of the finger domain is not dependent on this arginine, which may alternatively serve an essential role in a recognition helix. The H59Y mutation alters a histidine residue located in the center of the cysteine/histidine rich region. The nearly wild type levels of zinc present in this protein could be explained if this histidine residue were located within a loop of the zinc finger structure, rather than acting as one of the coordinating residues. Detailed studies of the three dimensional structure of the DNA binding domain will be essential for clarification of the role of specific amino acid residues.

Discussion

In this paper we have demonstrated that the male-specific and female-specific polypeptides encoded by the *dsx* locus have in common a non-canonical zinc finger-related DNA binding domain that binds *in vitro* to sequences regulating yolk protein gene expression, and have presented evidence that this binding domain is essential *in vivo* for the genetic functions of *dsx* in both sexes. The products of the *dsx* locus play reciprocal roles in the control of somatic sexual differentiation in *Drosophila melanogaster*. These two related polypeptides, produced as a result of sex-specific alternative splicing of the *dsx* primary transcript, are required in their respective sexes to repress the differentiation of sexually dimorphic characteristics of the opposite sex, and perhaps in some cases to activate the differentiation of characteristics of their cognate sex (Jursnich and Burtis, 1992; Taylor and Truman, 1992).

The recent discovery that bacterially produced proteins encoded by both the male and female *dsx* cDNAs bound with equal affinity *in vitro* to identical sites in enhancer sequences able to direct female-specific expression of the *Yp1* and *Yp2* genes implied that the two polypeptides shared a common DNA binding domain (Burtis *et al.*, 1991). This observation raised two principal questions. First, since computer-assisted searches did not identify any regions of the *dsx* polypeptides with high degrees of similarity to other polypeptides, what was the nature of the *dsx* DNA binding domain? Second, was this the only DNA binding domain in the male and female polypeptides, and if so, how could a single binding domain be reconciled with the very different effects of the two polypeptides on gene expression *in vivo*?

The *dsx* DNA binding domain is a novel zinc finger-related polypeptide

On the basis of clustered cysteine and histidine residues, PMPS-releasable bound metal ions, and EDTA-sensitive DNA binding activity, we have included the *dsx* DNA binding domain in the class of zinc finger-related motifs. However, the primary sequence of the *dsx* motif does not match at defined positions the canonical residues of the three well characterized classes of DNA binding zinc proteins, alternatively termed Class I (e.g. Zif 268), Class II (e.g.

glucocorticoid receptor) and Class III (e.g. GAL4) (Harrison, 1991) or zinc finger, zinc twist and zinc cluster, respectively (Vallee *et al.*, 1991). The arrangement of cysteine and histidine residues in *dsx* (C-X₂-C-X₂-H-X₈-H-X₃-C-X₄-C-X₁-C-X₂-C), although not consistent with any of the three major classes, does barely fall within the boundaries of the sequence (C-X₂₋₄-C-X₂₋₁₅-a-X₂₋₄-a, where a is either Cys or His) used by Berg (1986) to carry out a computer search for potential metal binding domains. The *dsx* polypeptide thus joins a rather small class of proteins for which site-specific DNA binding and metal ion coordination have been demonstrated, but which do not fall into one of the major classes of zinc finger-related proteins.

Our proposal that the *dsx* DNA binding domain represents a novel zinc finger-related domain adds yet another member to the group of non-standard zinc finger-related polypeptides, and as such has implications for the functional constraints underlying the structural organization of zinc binding domains. X-ray crystal structures exist for only one representative of each class of zinc binding domain (Harrison, 1991; Vallee *et al.*, 1991). However, identification of some common structural features has been possible (Luisi, 1992). Understanding the rules and mechanisms by which zinc finger-related domains are folded will require the crystallization of both additional examples from each major class as well as examples of domains such as the *dsx* DNA binding domain that do not fall into one of the major classes. Crystallography will be necessary to determine whether the *dsx* protein, despite its extreme sequence divergence, is a distantly related member of a known class of zinc finger domain, or if it represents a new structural motif. As more zinc finger-related polypeptides are identified it is possible that other examples of a *dsx*-like domain will be found.

Only one DNA binding domain is active *in vitro* and *in vivo*

A critical factor in understanding how DSX^M and DSX^F control the expression of the terminal genes that function in somatic sexual differentiation is to know whether the two proteins have similar or different DNA binding activities. We have pursued this question using both *in vitro* and *in vivo* approaches.

Combined N-terminal and C-terminal deletion experiments confirmed that for DSX^M there exists only one domain capable of binding *in vitro* to the fat body enhancer of the *Yp1*–*Yp2* intergenic region, and that this domain is indeed located in the common region shared between the male and female *dsx* polypeptides. Because the fat body enhancer is the only available sequence known to be a direct target of *dsx* regulation in males, and no gene known to be a target of *dsx* repression in females has been identified, it is not yet possible to determine *in vitro* if DSX^F (and perhaps DSX^M) have a different binding domain that recognizes sequences adjacent to other male-specific or female-specific genes. However, evidence discussed below argues against the existence of a second domain.

Our results indicate that the same binding domain identified *in vitro* is essential for the activities of the *dsx* polypeptides in both males and females *in vivo*. Four of the 10 null alleles of *dsx* examined have missense mutations that result in single amino acid substitutions in the common zinc finger-related DNA binding domain. This finding indicates that the

identified DNA binding domain is essential for *dsx* function in both sexes. Furthermore, it is unlikely that a second functional DNA binding domain exists in the *dsx* polypeptides. If recognition of target genes relied on two independent DNA binding domains, then one would predict the existence of three types of *dsx* mutations: those that affected both domains, and those that uniquely affected either one of the two domains. It is reasonable to assume that flies bearing domain-specific mutations would display only a subset of the phenotypic characteristics seen in flies lacking the function of both domains. However, the phenotypes in both sexes of flies carrying mutations we have shown to be in the zinc finger-related domain are indistinguishable from those of flies bearing *dsx* alleles demonstrated at the molecular level to be completely non-functional, suggesting that the DNA binding domain identified is essential for all *dsx* functions.

If both male and female *dsx* functions are mediated through the same DNA binding domain, then what is the nature of the *dsx* mutations (Baker and Ridge, 1980; Nöthiger *et al.*, 1987) that principally affect only one sex? We have determined that two such alleles, *dsx*¹³⁶ (male-specific allele, A.Garen, unpublished) and *dsx*^{EFH55} (a predominantly female-specific allele, Baker *et al.*, 1991) have functional DNA binding domains (S.Erdman, unpublished). A partial sequence of the *dsx*¹³⁶ allele has been obtained (K.Burtis, unpublished) and reveals a nonsense mutation in the 11th codon of the male-specific domain. Other sex-specific alleles may similarly be due to missense and nonsense mutations outside of the DNA binding domain, with the exception of the female-specific dominant mutations, which are known to be caused by incorrect splicing of the *dsx* mRNA in females (Nagoshi and Baker, 1990).

Differential regulation of gene expression by proteins sharing a common DNA binding domain

Since the male and female *dsx* polypeptides share a common DNA binding domain, how and why do the two polypeptides impose such different effects on somatic sexual differentiation? Two fundamentally different models can be considered; one in which the two proteins bind to the same target sequences *in vivo*, but cause different effects when bound, or an alternative model in which the identical DNA binding domains of the two proteins are somehow influenced *in vivo* to recognize different target nucleotide sequences. We shall consider the two models separately, and discuss the extent to which each is supported by existing data.

In the first model, dissimilar effects result from the binding of DSX^M or DSX^F at the same sites. This model requires that each of the two proteins display unique regulatory properties in their interactions with specific targets. These unique properties may be inherent in the primary structure of the sex-specific domains, or may result from domain-specific interactions with other regulatory proteins. Regulation of transcription in higher eukaryotes commonly involves interactions between different proteins bound to enhancer sequences (reviewed by Thompson and McKnight, 1992). These proteins are often modular in nature, with separate domains responsible for different functions such as DNA binding and activation or repression (reviewed by Frankel and Kim, 1991). If DSX^M and DSX^F are considered as identical DNA binding domains fused to sex-specific repression (or activation) domains, then the activity at any

particular binding site will depend on how the sex-specific domain interacts with the other factors bound adjacent to that site.

A prediction of the first model is that both the male and female *dsx* polypeptides should bind *in vivo* to a known target, the fat body enhancer of the *Yp1* and *Yp2* genes, with different regulatory consequences. Both genetic experiments (Postlethwait *et al.*, 1980; Bownes and Nöthiger, 1981; Ota *et al.*, 1981; Belote *et al.*, 1985) and *in vitro* results (Burtis *et al.*, 1991) predict that the male protein can bind to and repress the fat body enhancer of the yolk protein genes. Recent studies have confirmed that the interaction of DSX^M with the fat body enhancer *in vivo* is dependent on the presence of the DSX binding sites identified *in vitro*, and have furthermore shown that transcription of the *Yp1* and *Yp2* genes directed by the fat body enhancer is actually increased *in vivo* by the presence of DSX^F in females (K.Coschigano and P.Wensink, personal communication). Thus, it is clear from the example provided by this case that the two polypeptides can indeed bind to the same sites *in vivo* with very different regulatory effects.

The second model proposes that the male and female *dsx* polypeptides bind different target sequences *in vivo*, although as argued above, this binding must depend on the non-sex-specific DNA binding domain we have identified. There are several possible mechanisms by which interactions with additional proteins might modify the recognition site of the two polypeptides. For example, if the *dsx* proteins bind DNA as homodimers, interactions with an additional protein might alter the spacing and orientation of the DNA binding domains. A well characterized example of this type of mechanism is found in *Saccharomyces cerevisiae*, where interaction with the MCM1 protein alters the binding site specificity of dimers of the $\alpha 2$ homeodomain protein (Smith and Johnson, 1992). Alternatively, additional proteins might form heterodimers with the *dsx* polypeptides to create novel chimeric DNA binding domains, as occurs in the case of the bZIP and bHLH families of transcription factors (reviewed by Jones, 1990). Yet a third possibility would be an alteration of binding site specificity due to post-translational modifications of either of the *dsx* polypeptides.

Which of the above models best describes the mechanism by which the *dsx* polypeptides regulate somatic sexual differentiation? In the case of the target gene for which there is molecular data, results support the first model, in which *dsx* proteins bind to common sites, but with different effects on expression. There is no direct evidence for or against the second model in which the proteins bind to different target sequences. However, it is clear from the results of the *in vivo* experiments with the yolk protein fat body enhancer that not all target site recognition can be sex specific. A final determination of which model is most representative of the actual mechanism by which regulatory specificity is achieved will require the isolation and analysis of additional target genes.

Any complete model of *dsx* function must account for the role of the product of the *intersex* (*ix*) gene. The *ix* gene is required only in females, and loss of *ix* function results in an intersexual phenotype indistinguishable from that caused by loss of *dsx* function. The epistatic interactions between *ix* and other sex determination mutations (Baker and Ridge, 1980; Nöthiger *et al.*, 1987; Nagoshi *et al.*, 1988) suggest that *ix* protein is required along with DSX^F to

regulate target gene expression in females. It is possible to incorporate the role of *ix* into both of the models considered. In the first model, *ix* protein may interact with DSX^F and provide an essential domain that mediates interactions with other regulatory proteins. If all functions of DSX^F are mediated through this domain, then complete loss of either *ix* or *dsx* function should have identical phenotypic consequences. Alternatively, *ix* might represent the additional protein factor hypothesized in the second model to alter the binding specificity of DSX^F. While this is possible, the evidence that both DSX^M and DSX^F can bind the fat body enhancer of the yolk protein genes means that not all molecules of DSX^F can have altered specificity. If two classes of target exist, those recognized by the DSX^F-*ix* complex and those recognized by DSX^F alone, then it might be expected that loss of *ix* function would lead to misregulation of only a fraction of female target genes. The identical phenotypes resulting from *dsx* and *ix* mutations do not support this expectation, although the visible phenotype may reveal the influence of only a subset of target genes.

To understand fully at the molecular level how DSX^F and DSX^M can cause such different effects on gene expression while using a common DNA binding domain, two key issues need to be addressed. Additional target genes regulated directly by *dsx* must be identified, and the functional domains other than the DNA binding domain that contribute to the activities of the *dsx* polypeptides must be characterized. By understanding how these functional domains interact with the other regulatory sequences and proteins controlling expression of specific target genes, it should be possible to understand the mechanisms that allow DSX^M and DSX^F to impose different developmental fates on sexually dimorphic tissues.

Materials and methods

DNA manipulations

Genomic, plasmid and PCR derived DNAs were routinely handled as described (Sambrook *et al.*, 1989). C-terminal protein truncations were generated by insertion of an oligonucleotide linker with translation termination codons in all three reading frames (5'-CTAGTCTAGACTAG). This linker was ligated into blunt-ended expression plasmids derived either from digestion of pT7-7-DSX^M at restriction sites unique to *dsx* (*Bam*HI and *Eag*I) or following linearization at the *Bam*HI site and treatment with exonuclease III and S1 nuclease to generate a deletion series. Non-phosphorylated linker ligations were performed in 30 mM Tris-HCl pH 7.5, 30 mM NaCl, 7.6 mM MgCl₂, 0.25 mM ATP, 2 mM DTT, 0.2 mM EDTA, 1.0 mM spermidine, 0.1 mg/ml BSA buffer at 4°C with 200 units of T4 DNA ligase in a 20 µl reaction. N-terminal protein deletions were created using a modified pT7-7 vector. Partially complementary oligonucleotides (5'-TATGAGCTCGGTACCG, 5'-AATTCGGTACCGAGCTCA) were annealed and inserted into pT7-7 (Tabor and Richardson, 1985) that had been digested with *Nde*I and *Eco*RI, thus creating *Sac*I and *Kpn*I sites upstream of the *Eco*RI site and downstream of the vector-encoded ATG initiation codon. DSX protein coding sequences were then recloned into the deletion vector in an identical manner to the construction of the original pT7-7-DSX^M expression plasmid (Burtis *et al.*, 1991). The resulting clone was digested with *Sac*I and *Eco*RI and treated with exonuclease III and S1 nuclease to generate a deletion series. Following ligation, clones corresponding to various time points were chosen and sequenced to verify (i) the presence of an intact initiation codon, and (ii) in-frame fusions of the initiation codon to the DSX coding sequence. Computer algorithms used for protein structure predictions and database searching were part of the UWGCG Package (Devereux *et al.*, 1984).

Site-directed mutagenesis of the DNA binding domain by PCR

Site-directed mutagenesis was performed by a two step PCR process. Four priming sites for PCR were chosen along the DNA sequences encoding

a portion of the *dsx* polypeptide. The physical order of these primers from 5' to 3' was A-C-B-D, with primers A and C in the sense orientation and primers B and D in antisense orientation. Two degenerate versions of primer C were used; C1 to introduce mutations at cysteine 47 and C2 to introduce mutations at cysteine 68. Primer A contained sequences including the vector-encoded initiation codon in the modified pT7-7-DSX^M expression construct mentioned above. Primer B is complementary to nucleotides designated 1597-1611 by Burtis and Baker (1989); primer D is complementary to nucleotides 2059-2073. Amplification of DNA fragments between priming sites A and C produced the primary product AC, while amplification between sites B and D yielded primary product BD. The primary products AC and BD share a region of overlap between priming sites B and C that mediates their subsequent rejoining in a second round of PCR. Following treatment with T4 DNA polymerase to remove non-templated thymidines present on 3' ends of the primary PCR products caused by the terminal transferase activity associated with Taq polymerase (Clark, 1988), and gel isolation to remove primers, ~10 ng of each primary product were mixed, using a slight molar excess of the mutation-encoding product. The mixed products were then denatured, annealed and extended in the absence of primers. Three cycles were used to ensure that an adequate number of extended template strands were produced for the subsequent step. Addition of primer A, followed by several linear rounds of synthesis and denaturation, caused overrepresentation of the mutant template strand relative to the wild type strand. The second flanking primer, D, was then added to begin a 25 cycle amplification of the final product. The final products were subcloned into pBluescript SK(+) (Stratagene) and screened by DNA sequencing for the presence of desired mutations.

Oligonucleotides used in this procedure were:

A 5'-TATGAGCTCGGTACCG

B 5'-TGACTCAGGAGGGTC

C1 5'-CTGCGCCCGCYCCGCAATCAT X = 33.3% A, G, T

C2 5'-GTTCCGCTACXYCACGTGCGAG Y = 50% A, C

D 5'-CCGCTGGATCCGTTG

Amplification and subcloning of *dsx* sequences from mutant alleles

DNA sequences corresponding to individual mutant *dsx* alleles were PCR amplified from genomic DNAs of flies of appropriate genotypes (see text) using the primers 5'-CCGAATTCATATGGTTTCGGAGGAG and 5'-GCTGGATCCGTTGCT; the former has an *Eco*RI linker followed by sequences spanning the DSX initiation codon (indicated by the ATG), and the latter is complementary to nucleotides 2057-2071 of the *dsx* gene, as designated by Burtis and Baker (1989). PCR products were digested by *Eco*RI and *Nco*I (an internal *Nco*I site appears in the amplified *dsx* sequences), and subcloned into the vector pGEX-KG (Guan and Dixon, 1991) between the *Eco*RI and *Nco*I sites. All nucleic acid sequences generated by PCR were determined and verified to contain only the sequence alterations indicated in the text.

Protein purifications and gel mobility shifts

Proteins produced by T7 expression were purified in Z-100 buffer as previously described (Burtis *et al.*, 1991). Glutathione-S-transferase fusion proteins were prepared and affinity purified by batch in Z-100 buffer as follows. Cultures grown overnight in 2XYT with 100 µg/ml carbenicillin were induced by 10-fold dilution into 2XYT with 50 µg/ml carbenicillin and 0.44 mM IPTG. Induced cultures were grown 3 h prior to harvesting cells. Cleared cell lysates were prepared as previously described (Burtis *et al.*, 1991). Glutathione agarose beads (Sigma) were swelled in distilled water, followed by equilibration in Z-100 buffer. Beads were added to 1 vol per 400 vol of culture (typically 50 µl/20 ml culture) and allowed to bind fusion proteins at 4°C for 30 min. After washing, proteins were eluted from the beads twice in two bead volumes of Z-100, 5 mM glutathione, 5 mM NaOH, then stored at 4°C.

Gel mobility shift analyses of DNA binding activities of proteins and extracts were performed essentially as previously described (Burtis *et al.*, 1991). ³²P-end-labeled DNA fragments used as probes were either a set of restriction fragments from sequences upstream of the *Yp1* gene, the smallest of which contains the 127 bp FBE controlling expression of the *Yp1* and *Yp2* genes; or alternatively a 33 bp oligonucleotide containing the highest affinity binding site for the DSX proteins ('A' site of Burtis *et al.*, 1991) present in the FBE. The sequences of the complementary oligonucleotides that were annealed to form A sites with synthetic *Eco*RI sticky ends for end-filling with radiolabeled nucleotides by Klenow polymerase were:

5'-AATTCACAACACTACAATGTTGCAATCAGCTAGCC

5'-AATTGGCTAGCTGATTGCAACATTGTAGTTGTG

Metal divalent cation quantitation using PAR

GST fusion proteins were prepared as described above with the following modifications. Prior to elution of fusion proteins, affinity beads were washed with an additional 3×30 bead volumes of Z-100 lacking NP-40. Proteins were eluted from the beads in 2×2 bead volumes in Z-100 (without NP40), 5 mM glutathione, 5 mM NaOH. Proteins were then dialyzed with several buffer changes against 8 l of HGK buffer (25 mM HEPES pH 7.6, 10% v/v glycerol, 100 mM KCl) at 4°C. Protein concentrations were determined by A_{280} , and normalized by appropriate dilutions with HGK dialysate. The relative concentration of sulfhydryl-coordinated heavy metal ions in each sample was determined by measuring the change in absorbance at 500 nm after addition of saturating quantities (50 μ l of 4 mM solution) of *p*-hydroxymercuriphenyl sulfonic acid (PMPS) to plastic cuvettes with 1 ml protein and 20 μ l of a 5 mM solution of the metalochromic indicator 4-(2-pyridylazo)resorcinol (PAR) at pH 8.8 (Jezorek and Freiser, 1979; Hunt *et al.*, 1985). PMPS reacts covalently with the sulfhydryl residues of cysteines. One ml of HGK dialysate with 20 μ l 5 mM PAR served as a blank. Upon addition of PAR to the protein samples, an initial increase in A_{500} was seen, presumably due to loosely coordinated ions. PMPS was added after A_{500} values were stable for 60 s, and the change in absorbance recorded.

D.melanogaster stocks and culturing

Stocks were maintained and crosses were carried out on synthetic media at 25°C. All *dsx* alleles are described in Lindsley and Zimm (1992) except *Df(3R)dsx^{D+R2}* (Baker *et al.*, 1991) and *dsx¹²²⁻¹* (A. Garen, unpublished).

Acknowledgements

The authors wish to acknowledge the technical assistance of Brian Trenchak and especially Molly Cadle and Steve Reinl for synthesis of oligonucleotides. We also thank Scott Hawley, Holly Irick and Vicki Jursnich for useful discussions throughout the course of this work, and Scott Hawley for critical comments on the manuscript. This work was supported by grants from the National Science Foundation and the Searle Family Trust/Chicago Community Trust to K.C.B.

References

- Baker, B.S. (1989) *Nature*, **340**, 521–524.
 Baker, B.S. and Ridge, K. (1980) *Genetics*, **94**, 383–423.
 Baker, B.S., Hoff, G., Kaufman, T.C., Wolfner, M.F. and Hazelrigg, T. (1991) *Genetics*, **127**, 125–138.
 Belote, J.M., Handler, A.M., Wolfner, M.F., Livak, K.J. and Baker, B.S. (1985) *Cell*, **40**, 339–348.
 Berg, J.M. (1986) *Science*, **232**, 485–487.
 Blumberg, H., Eisen, A., Sledziewski, A., Bader, D. and Young, E.T. (1987) *Nature*, **328**, 443–445.
 Bownes, M. and Nöthiger, R. (1981) *Mol. Gen. Genet.*, **182**, 222–228.
 Burtis, K.C. and Baker, B.S. (1989) *Cell*, **50**, 997–1010.
 Burtis, K.C. and Wolfner, M.F. (1992) *Sem. Dev. Biol.*, **3**, 331–340.
 Burtis, K.C., Coschigano, K.T., Baker, B.S. and Wensink, P.C. (1991) *EMBO J.*, **10**, 2577–2582.
 Chou, P.Y. and Fasman, G.D. (1978) *Adv. Enzymol.*, **47**, 45–148.
 Clark, J.M. (1988) *Nucleic Acids Res.*, **16**, 9677–9686.
 Cline, T.W. (1989) *Cell*, **59**, 231–234.
 Devereux, J., Haeblerli, P. and Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
 Frankel, A.D. and Kim, P.S. (1991) *Cell*, **65**, 717–719.
 Guan, K.L. and Dixon, J.E. (1991) *Anal. Biochem.*, **192**, 262–267.
 Harrison, S.C. (1991) *Nature*, **353**, 715–719.
 Hildreth, P.E. (1965) *Genetics*, **51**, 659–678.
 Hunt, J.B., Neece, S.H. and Ginsburg, A. (1985) *Anal. Biochem.*, **146**, 150–157.
 Jezorek, J. R. and Freiser, H. (1979) *Anal. Chem.*, **51**, 373–376.
 Jones, N. (1990) *Cell*, **61**, 9–11.
 Jursnich, V.A. and Burtis, K.C. (1992) *Dev. Biol.*, **154**, in press.
 Lindsley, D.L. and Zimm, G.G. (1992) *The Genome of Drosophila melanogaster*. Academic Press, San Diego, CA.
 Luisi, B. (1992) *Nature*, **356**, 379–380.
 Nagoshi, R.N. and Baker, B.S. (1990) *Genes Dev.*, **4**, 89–97.
 Nagoshi, R.N., McKeown, M., Burtis, K.C., Belote, J.M. and Baker, B.S. (1988) *Cell*, **53**, 229–236.
 Nöthiger, R., Leuthold, M., Andersen, N., Gerschwiler, P., Grüter, A., Keller, W., Leist, C., Roost, M. and Schmid, H. (1987) *Genet. Res. Camb.*, **50**, 113–123.
 Ota, T., Fukunaga, A., Kawabe, M. and Oishi, K. (1981) *Genetics*, **99**, 429–441.
 Postlethwait, J., Bownes, M. and Jowett, T. (1980) *Dev. Biol.*, **79**, 379–387.
 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 Schena, M., Freedman, L.P. and Yamamoto, K.R. (1989) *Genes Dev.*, **3**, 1590–1601.
 Severne, Y., Wieland, S., Schaffner, W. and Rusconi, S. (1988) *EMBO J.*, **7**, 2503–2508.
 Slee, R. and Bownes, M. (1990) *Q. Rev. Biol.*, **65**, 175–204.
 Smith, D.L. and Johnson, A.D. (1992) *Cell*, **68**, 133–142.
 Steinmann-Zwicky, M., Amrein, H. and Nöthiger, R. (1990) *Adv. Genet.*, **27**, 189–237.
 Tabor, S. and Richardson, C.C. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1074–1078.
 Taylor, B.J. and Truman, J.W. (1992) *Development*, **114**, 625–642.
 Thompson, C.C. and McKnight, S.L. (1992) *Trends Genet.*, **8**, 232–236.
 Vallee, B.L., Coleman, J.E. and Auld, D.S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 999–1003.
 Webster, L.C., Zhang, K., Chance, B., Ayene, I., Culp, J.S., Huang, W.-J., Wu, F.Y.-H. and Ricciardi, R.P. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9989–9993.
 Wolfner, M.F. (1988) *Trends Genet.*, **4**, 333–337.

Received on July 27, 1992; revised on October 29, 1992