Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers

Lingyang Zhu,1,4 Jill Wilken,2 Nelson B. Phillips,3 Umadevi Narendra,3 Ging Chan,1 Stephen M. Stratton,2 Stephen B. Kent,2 and Michael A. Weiss1,3–5

1Center for Molecular Oncology, Departments of Biochemistry & Molecular Biology and Chemistry, The University of Chicago, Chicago, Illinois 60637-5419 USA; 2Gryphon Sciences, South San Francisco, California 94080 USA; 3Department of Biochemistry, Case Western Reserve School of Medicine, Cleveland, Ohio 44106-4935 USA

Sex determination is regulated by diverse pathways. Although upstream signals vary, a cysteine-rich DNA-binding domain (the DM motif) is conserved within downstream transcription factors of Drosophila melanogaster (Doublesex) and Caenorhabditis elegans (MAB-3). Vertebrate DM genes have likewise been identified and, remarkably, are associated with human sex reversal (46, XY gonadal dysgenesis). Here we demonstrate that the structure of the Doublesex domain contains a novel zinc module and disordered tail. The module consists of intertwined CCHC and HCCC Zn2+-binding sites; the tail functions as a nascent recognition α-helix. Mutations in either Zn2+-binding site or tail can lead to an intersex phenotype. The motif binds in the DNA minor groove without sharp DNA bending. These molecular features, unusual among zinc fingers and zinc modules, underlie the organization of a Drosophila enhancer that integrates sex- and tissue-specific signals. The structure provides a foundation for analysis of DM mutations affecting sexual dimorphism and courtship behavior.

[Key Words: Drosophila melanogaster; DNA-binding protein; gene regulation; transcription factor; NMR]

Received April 14, 2000; accepted in revised form May 22, 2000.

Sexual dimorphism exhibits a diversity of molecular mechanisms. In mammals formation of the testis and subsequent male development is initiated by a gene on the Y chromosome (Page et al. 1987). Studies of human sex reversal led to the identification of SRY (Gubbay et al. 1990; Sinclair et al. 1990), the testis-determining factor (Berta et al. 1990; Koopman et al. 1990). SRY contains a high-mobility group (HMG) box, a minor groove DNA-binding domain (DBD) shared by a family of architectural transcription factors. Regulation of organogenesis requires sharp DNA bending (Giese et al. 1992; van Houte et al. 1993; Pontiggia et al. 1994; Bewley et al. 1998). Alternative genetic mechanisms operate in Drosophila melanogaster (Nagoshi et al. 1989; Burtis et al. 1991; An et al. 1996; Cho and Wensink 1997, 1998; Li and Baker 1998; Marin and Baker 1998) and Caenorhabditis elegans (Villeneuve and Meyer 1990, Hodgkin 1997) wherein sex is determined by the X:autosome ratio, a process linked to X dosage compensation (Cline and Meyer 1996; Dawes et al. 1999). The pathways of the fly and worm (Fig. 1A) seem otherwise unrelated. Whereas somatic sexual differentiation in C. elegans requires cell–cell signaling, analogous decisions in D. melanogaster are cell autonomous (Cline and Meyer 1996). D. melanogaster [but not C. elegans] uses a hierarchy of sex-specific RNA splicing events.

Despite diverse upstream mechanisms, downstream sex-determining genes may be conserved in part (Raymond et al. 1998). An example is provided by homologs doublesex (dsx; Hildreth 1965) and mab-3 [Shen and Hodgkin 1988]. In C. elegans MAB-3 is expressed in the male and repressed in the hermaphrodite, whereas the RNA-splicing pathway of D. melanogaster regulates male- and female-specific DSX isoforms [DSXm and DSXf; Baker and Wollner 1988; Burtis and Baker 1989]. Despite these differences, dsx and mab-3 each function downstream of “master” regulatory factors to define one branch of a ramifying pathway (Fig. 1A). Furthermore, dsx and mab-3 each encode transcription factors that in part regulate analogous dimorphic tissues [Bowens et al. 1983; Baker and Wollner 1988; Burtis and Baker 1989; Coschigano and Wensinck 1993]. Functional analogies in downstream pathways exceed the extent of molecular or histologic conservation: for example, DSX and MAB-3 each regulate expression of yolk–protein genes (yp and vit families, respectively), although these yolk proteins are unrelated in sequence [Bowens et al. 1983; Coschigano and Wensinck 1993; An and Wensinck 1995a,b; Li and Baker 1998; Yi and Zarkower 1999]. Similarly, DSX-
regulated differentiation of male sensory bristles in *D. melanogaster* and MAB-3-regulated formation of male sensory rays in *C. elegans* seem analogous even as the organs themselves differ in histology. Strikingly, DSXM functions in a *mab-3*− XO worm (ordinarily a chromosomal male with an intersex phenotype; Shen and Hodgkin 1988) to rescue such male features (Raymond et al. 1998). Complementation across phyla demonstrates that downstream mechanisms of gene regulation are in part conserved.

DSX and MAB-3 domains exhibit zinc-dependent DNA binding (Erdman and Burtis 1993) with distinct but
Cysteines and histidines that coordinate Zn\(^{2+}\) are aligned as two intertwined binding sites (boxes): site I (red) and site II (green).

D. melanogaster cDmrt1 (chicken, accession no. AF123456). TERRA: DM domain of zebrafish (accession no. AF130729). Other homologs: mDmrt1 (murine, accession no. AF202778), pDmrt1 (porcine, accession no. AF216651), and rDmrt1 (rat, accession no. AF101315); C27C12 (accession no. Z69883); and Y43F8C (accession no. AL032637.1). Vertebrate DM sequences: DMRT1 and DMRT2, human homologs on the short arm of chromosome 9 (DMRT1: accession no. AF130728 and DMRT2: accession no. AF130729). Other homologs: mDmrt1 (murine, accession no. AF202778), pDmrt1 (porcine, accession no. AF216651), and cDmrt1 (chicken, accession no. AF130729). TERRA: DM domain of zebrafish (accession no. AF080622). Sequences shown include genes such as TERRA, Meng et al. 1999) known to not be involved in sex determination. Arrows without parentheses indicate sites of point mutations in dsx or mab-3 associated with intersex development; substitutions in parenthesis indicate variants characterized only by biochemical assays (Erdman and Burtis 1993). Three additional putative DM genes of unknown function have recently been found in the genome of D. melanogaster (accession nos. AA56619, AA55843, and AA48261, sequences not shown).

overlapping specificities [Erdman et al. 1996; Yi and Zarkower 1999]. The proteins contain homologous DBDs (the DM motif) remarkable for a novel pattern of cysteines and histidines [Fig. 1B; Erdman and Burtis 1993; Raymond et al. 1998]. Mutations in the motif are associated with intersex development (arrowheads in Fig. 1B; Hildreth 1965; Hodgkin 1983; Erdman et al. 1996). Phenotypes correlate with loss of DNA-binding activity [Erdman and Burtis 1993; Raymond et al. 1998]. DSX mutations in the invariant cysteines or histidines impair both Zn\(^{2+}\) coordination and DNA recognition, an intersex mutation in the motif’s carboxy-terminal segment disrupts DNA binding without affecting Zn\(^{2+}\) coordination [R91Q, Erdman and Burtis 1993]. DSX\(^{M}\) and DSX\(^{F}\) share the same amino-terminal DBD but distinct carboxy-terminal extensions [Fig. 2A; Burtis and Baker 1989]. Unlike DSX, MAB-3 contains two DM domains [Fig. 2A], each required in vivo [Raymond et al. 1998] and proposed to function coordinately in DNA binding [Yi and Zarkower 1999]. Multiple homologs of unknown function are encoded in the C. elegans genome [see Fig. 1B; C. elegans Sequence Consortium 1998].

The DM motif is conserved within vertebrates [Fig. 1B, bottom]. These include DMRT1 and DMRT2, adjacent genes on human chromosome 9 (9p24; Raymond et al. 1999a). Remarkably, deletions spanning DMRT1 and DMRT2 are associated with XY sex reversal in the absence of abnormalities in SRY [Fryns et al. 1986; Crocker et al. 1988; Magenis et al. 1990; Bennett et al. 1993; McDonald et al. 1997; Flejter et al. 1998; Guirao et al. 1998; Veitia et al. 1998]. This clinical correlation suggests broad conservation of DM genes in metazoan sex determination, a hypothesis supported by the embryologic expression of murine Dmrt1 in the differentiating gonadal ridge [Raymond et al. 1999b]. Point mutations in DMRT1 or DMRT2 (unlike deletions) are apparently rare in XY humans with unexplained sex reversal or intersex phenotypes, presumably due to functional redundancy [Raymond et al. 1999a]. Zebras and mammals, like C. elegans, contain additional DM genes, one of which (TERRA) appears to function in mesodermal patterning in both sexes [Meng et al. 1999].

We present here the solution structure of the DSX DM domain and define its overall mode of DNA recognition. The structure reveals a novel zinc module containing intertwined CCHC and HCCC Zn\(^{2+}\)-binding sites. A carboxy-terminal tail, disordered in the free domain, is stabilized on DNA binding as a recognition a-helix. Unlike classic zinc fingers and zinc modules [Klug and Schwebe 1995], the DM domain binds in DNA’s minor groove. Unlike SRY and other architectural motifs targeted to the minor groove [Giese et al. 1992; van Houte et al. 1993; Haq et al. 1994], the DM motif does not induce sharp DNA bending. These unusual features rationalize the architecture of a DSX response element active in a dimorphic lineage [the female fat body, Burtis et al. 1991; Coschigano and Wensink 1993]. The fat body enhancer (f) [Fig. 2B] contains tightly overlapping minor and major groove DNA control sites, which function coordinately to integrate sex- and tissue-specific signals [Fig. 2C; An and Wensink 1995a,b]. The structure of the DSX DM domain provides a foundation for analysis of intersex mutations in diverse metazoans [Raymond et al. 1998]. Because in the fly such mutations lead to allele-specific changes in sexual preference [Villella and Hall...
1996], their characterization may provide insight into the genetic encoding of a complex behavior.

Results

Seventy-one-residue DSX- and DMRT1-derived DM polypeptides were obtained by total synthesis. The synthetic strategy exploited the native chemical ligation reaction (Dawson et al. 1994) and used an invariant cysteine residue as a central ligation site. Studies focused on DSX, which was found to be more stable to air oxidation than DMRT1. The DSX polypeptide was shown by mass spectrometry to bind two Zn$^{2+}$ ions; the (Zn$^{2+}$)$_2$–polypeptide complex was found by analytical ultracentrifugation to be monomeric in the concentration range 60–140 µM. The synthetic domain’s affinity for the fbe \( K_d \) 8 nM as inferred by gel mobility-shift assay; GMSA) is similar to published values (An et al. 1996; Erdman et al. 1996). Zinc-dependent protein folding was demonstrated by circular dichroism (CD) and proton nuclear magnetic resonance (\(^1\)H-NMR). The visible absorption spectrum of a (Co$^{2+}$)$_2$–peptide complex is consistent with tetrahedral coordination. Addition of two equivalents of Zn$^{2+}$ leads to attenuation of Co$^{2+}$ d–d transitions, indicating (as in the zinc finger; Frankel et al. 1987; Lee et al. 1989) preferential binding of Zn$^{2+}$.

The DM domain binds cooperatively in the minor groove without DNA bending

Binding of the DSX DM domain to dsxA (Fig. 3A,B) leads to discrete complexes (c1 and c2). Protein:DNA stoichi-
Figure 3. DNA-binding studies of DSX domain demonstrate minor groove binding mode. (A, top) Consensus sequence of DSX-binding sites, an imperfect palindrome containing an odd number of base pairs. (Second from top) DSX target site in fbe1–13. (Third from top) Analog containing multiple AT → IC substitutions (boldface type). (Bottom two sequences) Variants containing additional AT or TA base pairs in center. Labels b, c, d, e, f (circled) refer to lanes of gels in B. [8] 32P-detected GMSA at 4°C of DSX domain binding to sites defined in A. Bands c1 and c2 designate 1:1 and 2:1 protein–DNA complexes. Protein binds equally well to native and IC-containing sites. Insertion of central AT or TA base pairs reduces protein binding by about 100-fold. Protein concentration was 8 nM and DNA concentration was 0.5 nM. (C) Cooperative low-affinity binding of the DSX DM domain to a palindromic (sym) DNA site (5’-ACTACAATTGTTCGCA-3’; central base pair in boldface type) using fluorescein (at 5’ of top strand) as probe. DNA concentration was 500 nM in lanes a–i, with respective protein concentrations 0, 250, 500, 600, 700, 800, 900, and 1000 nM. Lane j contains an equimolar solution in which concentrations of DNA and protein were each 5 µM; predominance of 2:1 complex and free DNA indicates cooperativity is retained. (D, left) Structures of AT and variant nebularine NT base pair. (D, right) Structures of GC and inosine IC base pair. Positions of major and minor grooves are indicated. (E, right gel) Major groove base modifications do not perturb binding of DSX domain to 15 base-pair DNA probe (5’-CACTACAATGTTCGCA-3’ and complement): (lanes e–g) single nebularine substitutions at positions 5, 7, and 8 of upper strand and (lanes h–j) single 5-methylcytosine or uridine substitutions at positions 6, 9, and 11 of upper strand. (E, left gel) DMRT1 DM domain (25 nM) recognizes DSX target site with weaker apparent affinity; cooperative 2:1 binding is maintained. [F] 32P-detected GMSA of fbe analog in which the 8 AT bp in the DSX-binding site (boxed in Fig. 3A, sequence c,d) are substituted by 2,6-diaminopurine (structure at right). No binding is observed to λ operator site O1, (lanes e,d) or to the diaminopurine fbe analog (lanes e–d) at successive protein concentrations of 0, 1, 2, 4, 8, 16, 32, 64, and 128 nM. DNA concentration was –1 nM. Protein concentrations were 8 nM for lanes b and d. Sequence of λ probe is 5’-TACCTACTGCGCTACATA-3’ and complement.
omeries were verified to be 1:1 and 2:1 by GMSA studies of equimolar and 2:1 solutions 500-fold higher than the apparent dissociation constant. The predominance of the 2:1 complex under conditions containing appreciable free DNA demonstrates its cooperative assembly. The domain’s sequence specificity is highlighted by its 100-fold decreased affinity for palindromic dsxA analogs containing an additional central AT or TA base pair [Fig. 3A–C]; cooperativity is nevertheless retained [Fig. 3C]. The human DMRT1 domain also recognizes dsxA [Fig. 3E, lane l]. Minor groove binding was established by GMSA studies of DNA sites containing base analogs. Systematic AT → IC substitutions (boldface type in b in Fig. 3A; Lee et al. 1991; Starr and Hawley 1991) alter the global pattern of functional groups in the DNA major groove but preserve a native minor groove [Fig. 3D]. Multiple IC substitutions have no effect on specific DNA binding (Fig. 3B, lane b). Base minor groove (Fig. 3D). Multiple IC substitutions have no effect on specific DNA binding (Fig. 3B, lane b). Base substitutions by uridine, 5-methylcytosine, or nebulamine—complementary probes in the major groove—likewise have no effect [Fig. 3E]. Robustness of DM–dsxA recognition to major groove chemical modification stands in contrast to its disruption by insertion of a palindromic central base pair [discussed above]. An additional negative control is provided by systematic substitution of adenine by diaminopurine, which introduce exocyclic amino groups in the minor groove (right-hand panel of Fig. 3F). As expected, such bulky modifications prevent protein binding [Fig. 3F, lanes c–n]. Methylphosphonate interference (Labeots and Weiss 1997) was used to obtain a footprint of DM–DNA contacts [red circles in Fig. 2D]. Nine sites of interference were observed. These sites, potential contacts between the DM domain and the DNA backbone, are almost symmetric in distribution and restricted to the core consensus sequence [5′-NACAATGTN-3′ and complement; Erdman et al. 1996].

The staggered pattern of half-site contacts is consistent with protein binding to the minor groove of B-DNA. Despite minor groove binding, DSX-induced DNA bending is negligible as inferred from permutation gel electrophoresis: relative mobilities of either /fbe complex c1 or c2 exhibit only slight dependence on site position [Fig. 4A,B]. Such results are in contrast to the marked electrophoretic anomalies induced by binding of the SRY HMG box [Fig. 4B, control lanes 6,7; Giese et al. 1992]. Insufficient DNA bending was corroborated by analysis of protein-induced changes in fluorescence resonance energy transfer [FRET] between probes [fluorescein and tetramethylrhodamine] linked to opposite 5′ ends of a 15-bp duplex [Parkhurst et al. 1996]. Whereas binding of the SRY HMG box significantly enhances FRET efficiency, consistent with the decreased end-to-end distance of a sharply bent DNA, the effect of DM binding is minor (supplemental information).

The DM motif encodes a novel intertwined zinc module and disordered tail

The 1H-NMR spectrum of the DSX DM domain provides evidence of an ordered moiety and disordered tail. The ordered moiety [DSX residues 36–82, including invariant cysteines and histidines; see Fig. 1B] exhibits marked...
dispersion of chemical shifts, whereas the tail is manifest by an envelope of poorly resolved spin systems at near-random coil chemical shifts. Line widths and chemical shifts are similar in the protein concentration range 60 μM–1.4 mM. Sequential assignment of the ordered moiety, obtained by homonuclear two-dimensional NMR, was facilitated by comparative study of a fragment lacking most of the tail [DSX residues 35–86; DM\textsubscript{A}].

The specific DNA-binding activity of DM\textsubscript{A} is reduced by at least 100-fold (not shown). With the exception of tail resonances, \textsuperscript{1}H-NMR spectra of the intact and truncated domains are essentially identical (Fig. 5A,B). The fragment’s structure thus corresponds to the ordered moiety of the intact domain, demonstrating that folding of the zinc module is independent of the tail. The envelope of carboxy-terminal resonances in the spectrum of the intact domain (boxed in Fig. 5A) is similar to that of an isolated tail peptide (19 residues; Fig. 5C).

A network of 346 interproton nuclear Overhauser effects (NOEs) was observed in the ordered moiety. Of particular interest are 92 long-range NOEs, which include contacts within putative metal-binding sites. An example is provided by NOEs linking the side chains of residues C44, C47, H59, and C63 and in turn linking the side chain of C44 to aliphatic protons of I54 and L56. Contacts likewise occur from C47 to L52 and I54. Analogous NOEs are observed among H50, C68, C70, and C73. Notably, the aromatic ring of F65 contacts both sets of ligands (side chains of C47, H60, and C68).

Preliminary distance geometry/simulated annealing (DG/SA) calculations were performed in the absence of explicit Zn\textsuperscript{2+} coordination. Two potential zinc-binding sites were well defined; each exhibited dimensions consistent with tetrahedral coordination. Such interim models were refined by imposition of zinc-related distance restraints (see Materials and Methods; Qian et al. 1993). An extensive series of control DG/SA calculations was performed imposing alternative ligation schemes; each was excluded based on significant NOE restraint violations. The native model predicted a specific pattern of seven amide-sulfur hydrogen bonds involving four of the six cysteate Sy atoms. These bonds rationalize a corresponding pattern of protected amide protons in D\textsubscript{2}O, not otherwise accounted for by secondary structure. The final ensemble was refined by inclusion of amide–sulfur hydrogen bond restraints (Qian et al. 1993). A summary of restraints, violations, and empirical energy terms is provided in Table 1. Two glycines (G51 and G58) exhibit positive \( \phi \) values. No restraints were obtained for the carboxy-terminal tail [residues 87–105]; therefore, they were omitted from the model.

The structure of the DM domain is remarkable for intertwined CCHC and HCCC zinc-binding sites (sites I and II in Fig. 6A; outlined in schematic form in Fig. 1B). These sites nucleate a unified globular domain containing two \( \alpha \)-helices [residues 46–50 and 71–79; Fig. 6B]. The first helix, containing C47 and H50, connects Zn\textsuperscript{2+}-binding sites; the second and longer helix extends from site II into the tail. Ordered side chains adjoin each site (Fig. 6C). R46, a site of intersex mutation in \textit{C. elegans} (motif MAB-3b), is conserved in DMRT1, DMRT2, and other vertebrate homologs [Fig. 1B].

Figure 5. NOEsY spectra demonstrate “peptide dissection” of intact DM motif (A) into major and minor fragments: Zn-binding module (DM\textsubscript{A}, B) and carboxy-terminal peptide (peptide DM-p, C). Shown are respective spectra in H\textsubscript{2}O at 25°C and 600 MHz; mixing times were in each case 175 msec. Cross peaks from amide and aromatic protons (vertical scale, \( \omega_1 \)) and aliphatic protons (horizontal scale, \( \omega_2 \)) are shown. Boxed regions highlight unresolved envelope in intact DM domain (A) assigned to nascent carboxy-terminal helix (C); this feature is absent in spectrum of DM\textsubscript{B}. Peptides were made 1.5 mM in 50 mm deuterated Tris-HCl (\( \text{pH} \) 6.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O. Spectra A and B also were obtained in presence of 4 mM ZnCl\textsubscript{2}.
The tail contains a nascent DNA recognition helix

Although the carboxy-terminal tail is largely disordered, a strong helical propensity was predicted [Erdman and Burris 1993]. To resolve this apparent discrepancy, the tail’s mean conformation was investigated by CD. Comparison of the intact domain and DMΔ [Fig. 7A, a,b] demonstrates that the tail forms a nascent α-helix at low temperature. The inferred helix content of DMΔ (20% or 11 residues) is in accord with its structure [Fig. 7A, b]. The intact domain at 4°C [solid line in Fig. 7A, a] exhibits a greater helix content (30% or 21 residues). Their difference spectrum at 4°C exhibits an overall helical signature (Δ in Fig. 7A, c) whose normalization ([Δ] in Fig. 7A, d) predicts that the tail contains at least 10 additional residues of helix [in a static model] or subpopulation of helical structures in an ensemble [dynamic model] as implied by NMR features. The helical character of the tethered tail at 4°C is greater than that of an isolated peptide containing the carboxy-terminal 19 residues (DM-p in Fig. 7A, d), presumably due to nucleation by helix 2.

Structural independence of the zinc module and tail suggests uncoupled thermal unfolding transitions. Indeed, whereas the CD spectrum of DMΔ changes little between 4° and 37°C, the spectrum of the intact domain undergoes marked thermal attenuation [Fig. 7A, a]. The domain’s thermal unfolding exhibits an exaggerated and noncooperative overall transition [Fig. 7B, c]. The difference spectrum of the intact domain at 4° and 37°C (solid line in Fig. 7A, c) is α-helical in character and similar in magnitude to that between DM and DMΔ at 4°C [dashed dotted line in Fig. 7A, c]. Because the zinc module’s NMR resonances do not undergo significant changes in this temperature range, we ascribe the domain’s attenuation in helix content to thermal fraying of the tail. Similar CD patterns have previously been seen in studies of DNA-binding motifs containing nascent α-helical arms motifs [Weiss et al. 1990; Ferre-D’Amare et al. 1994; Carroll et al. 1997].

To investigate the function of the tail, binding of the DM domain to dsxA was monitored by CD [Fig. 7B]. Experiments at 4° and 37°C are shown in Figure 7B, a and b. Difference spectra at 37°C [Fig. 7B, d] indicate a change in structure of both DNA [250–300 nm; arrow in Fig. 7B, d] and protein [200–250 nm; induced α-helix]. The α-helical difference feature is more prominent at 37°C than 4°C due to thermal fraying of the free tail; the DNA difference feature is not significantly affected by temperature. DNA-stabilized α-helical structure is maintained in the complex at 4–40°C as indicated by its melting curve [Fig. 7B, c]. Negligible CD difference features were observed in control studies using λ operator site O1,1 (not shown). Similar temperature- and DNA-dependent α-helical CD difference features accompany the induced fit of nascent α-helical arms in otherwise unrelated DNA-binding motifs [Weiss 1990; Ferre-D’Amare et al. 1994; Carroll et al. 1997]. These results demonstrate that the DM tail folds on DNA binding but do not resolve whether the DNA-stabilized structure consists of a single continuous helix or discrete segments. The DM domain’s DNA-specific difference feature, unrelated to those of previous CD studies, may reflect the structural accommodation of an α-helix in the minor groove.

The DM tail may be stabilized on DNA binding by either tail–DNA contacts or DNA-induced dimerization [or both]. To test whether specific binding of a DM monomer to DNA is sufficient to stabilize the tail, CD studies were repeated using an isolated dsxA half site (5’-AGTACATTG-3’ and complement). The domain forms a low affinity 1:1 complex and on binding at 37°C yields a similar α-helical CD difference feature. Furthermore, formation of a 1:1 half-site complex is sufficient to confer thermal stability between 4° and 40°C. DNA-dependent folding can thus occur without dimerization. Direct tail–DNA contacts are also suggested by GMSA studies of an intersex variant (R91Q; see Fig. 1B); DNA recognition is impaired [as previously reported, Erdman and Burris 1993] but DNA-dependent dimerization is retained [see Fig. 4C]. These data strongly suggest that this
conserved basic side chain—and hence the tail itself—contributes to the protein–DNA interface; participation of other tail residues in dimerization cannot be excluded. Mutational analyses (Erdman and Burtis 1993; U. Narendra and M.A. Weiss, unpubl.) suggest that additional DNA contacts are made by the Zn module.

Discussion

The present study has defined the structure of the DM motif and its overall mode of DNA recognition. Whereas the genetic regulation of mammalian DM proteins [and in particular their relationship to SRY] is not understood, dsx and mab-3 have been characterized extensively [Fig. 1B; Cline and Meyer 1996]. Intersex phenotypes demonstrate that only a subset of dimorphic lineages in either D. melanogaster or C. elegans are affected by DM mutations [Hildreth 1965; Shen and Hodgkin 1988], reflecting the function of dsx and mab-3 in one branch of a ramifying pathway. Downstream pathways are not well defined. A model DSX responsive enhancer (the fbe; Burtis et al. 1991; Coschigano and Wensink 1993) has been isolated in yolk-protein (yp) promoters, active in the fat body of the female adult fly and inactive in the male [Fig. 2B,C; An and Wensink 1995a,b]. Although an analogous MAB-3 responsive element has not been characterized, positive regulatory sites have been defined in yolk protein-related vitellogenin (vit) promoters. Mutations that block binding of MAB-3 in vitro also attenuate sex-specific transcriptional regulation in vivo (Yi and Zarkower 1999). Whether the ubiquity of the DM motif indicates conservation of other downstream genes is unknown.

A nascent tail is required for DNA recognition and biological function

The structure of the DM zinc module, like the C_4–C_4 DBD of the nuclear hormone-receptor superfamily, con-
Figure 7. The DM tail is a nascent $\alpha$-helix that folds on DNA binding. (A,B) CD studies of free [A] and bound (B) DSX domain. (A,a) Spectra of intact domain at 4°C [solid line] and 37°C [dotted line]; (A,b) corresponding spectra of DM, $\Delta$CD. (A,c) Difference spectra: (solid line a), between DM spectra at 4° and 37°C; (dashed dotted line $\Delta$), between DM and DM, at 4°C; (dotted line b), baseline attenuation of stable helices in DM between 4° and 37°C. That difference spectra a and $\Delta$ are similar suggests that the labile segment in DM corresponds to the missing element in DM$\Delta$, i.e., a nascent carboxy-terminal helix. (A,d). Difference spectrum $\Delta$ [dashed dotted line], obtained from $\Delta$ normalized relative to 19 deleted residues, exhibits a helical signature. For comparison, spectra of isolated carboxy-terminal peptide DM-p 19 residues, underlined in Fig. 1B) are shown in $\Delta$ at 4° and 37°C [solid and dotted lines]. Magnitude of $\Delta$ at 222 nm [P$\Delta$/$\Delta$C] is similar to that of difference spectra $\Delta$ and a in A,C. (B) Comparison of spectra of protein–DNA complexes [solid line], free domain [dashed dotted line], and free DNA [dotted or dashed line] at 4°C [B,a] and 37°C [B,b]. (B,c) CD melting curves of free DM domain [dashed dotted line] and specific 2:1 complex [solid line] as monitored at 222 nm from 4° to 95°C. Whereas free domain exhibits steep loss in mean residue ellipticity with increasing temperature, the complex exhibits an attenuation rate $\Delta$ at 222 nm [P$\Delta$/$\Delta$C]. (B,d) Comparison of CD difference spectra at 37°C: [solid line $\Delta$] between complex and the free DNA [solid line $\Delta$], between the spectrum of a complex and the sum of the spectra of its isolated parts. Induced helical feature in far UV is labeled. Arrow in near-UV indicates difference feature arising from change in DNA structure on protein binding. Peptide and DNA concentrations were 10 and 5 µM, respectively. For reference, the spectrum of free DM domain is also shown [dashed dotted line].

Solution structure of the DM domain

contains two Zn$^{2+}$ atoms and adjoining $\alpha$-helices in a unified core. Although unrelated in detail, these motifs are substantially but not completely folded in the absence of DNA [Hard et al. 1990, Schwabe et al. 1990, Holmbeck et al. 1998]. Like the glucocorticoid and estrogen receptors (GR and ER) DBDs in particular, the DM domain recognizes pseudopalindromic DNA sites as an induced homodimer [Luiski et al. 1991, Schwabe et al. 1993]. Insertion of a symmetrical central base pair to make a palindromic DNA element in each case leads to a 100-fold reduction in affinity with retention of cooperativity among the basic arms of major groove-binding motifs (Weiss et al. 1990, Ferre-D’Amare et al. 1994, Carroll et al. 1997). Whereas the dimensions of the major groove of B-DNA are commensurate with those of an $\alpha$-helix, the minor groove is ordinarily too narrow. DM–DNA recognition thus provides an unusual example of the minor groove acting as a template for protein folding [Yang and Steitz 1995]. The R91Q intersex mutation in DSX demonstrates that a native tail is necessary for DNA recognition and sex-specific gene regulation in vivo. Tail sequences showed an overall helical propensity. Conservation of tail sequences is more stringent than within the zinc module [exclusive of the invariant cysteines and histidines; Fig. 1B]. The DSX tail is similar, for example, to those of MAB-3b and the vertebrate DMRT and TERRA families: the motif L[V/T][X[D/E/R][Q/RVMA][A/L]Q[V/T]-ALRR[Q/A]QA (DSX residues 75–94 including DSX intersex site R91Q, underlined) is invariant. We propose that this segment functions as a DNA recognition helix and therefore, defines a subgroup of DM proteins of related specificity. Such tail–DNA contacts are presumably extended by interactions between the DNA and the Zn module itself.

The tail of MAB-3a is strikingly divergent: nonconser-
orative substitutions are highlighted by arrows in Figure 1B. In particular, MAB-3a contains Ile rather than Arg at DSX intersex site R91Q. Such divergence, retained in C. briggsae [V91; Yi and Zarkower 1999], suggests that MAB-3a recognizes a different DNA sequence or is positioned differently relative to a DNA half site. Indeed, although DSX and MAB-3 exhibit overlapping specificities, random binding-site selection has revealed important differences [Yi and Zarkower 1999]. In particular, MAB-3’s consensus sequence exhibits marked asymmetry ([5′-AATGTTGCGA[T/A]NT-3′ and complement], which contrasts with near-palindromic dsxA half-sites. Because respective DSX and MAB-3 target sites are similar in length, MAB-3 is proposed to bind DNA as an “internal dimer” of DM domains. We imagine that the divergent tails of MAB-3a and MAB-3b contact asymmetric half-sites with different intrinsic specificity. Whereas the genetic function of MAB-3 requires both domains [Raymond et al. 1998], other known DM proteins contain a single domain. We speculate that such proteins undergo DNA-dependent homo- or heterodimerization as an example of combinatorial gene regulation.

DNA-binding properties of the DM motif rationalize the organization of a DSX response element

The fbe, a model DSX response element in a dimorphic tissue, contains three overlapping-factor-binding sites [An and Wensink 1995a,b; Fig. 2B]. Binding of DSX to the central site (dsxA; red box in Fig. 2B) is proposed to regulate binding of tissue-specific factors to flanking sites (aef1 and bzip1; Fig. 2C). Critical bases in dsxA recognized by DSX are palindromic about a central base pair [Fig. 2B, bottom; An and Wensink 1995b; Erdman et al. 1996]. The associated tissue-specific factors [AEF1 and an uncharacterized bZIP transcription factor] contain conserved DNA-binding motifs [classic zinc fingers and basic region-leucine zipper, respectively]. In the fat body coordinate binding of DSXf and bZIP1 to the fbe displaces AEF1 and recruits an unknown activator to proximal site ref1, resulting in fat body-specific activation [Fig. 2C, top; other regulatory features are summarized in the caption]. Overlapping DNA target sites are thus proposed to provide a mechanism by which sex- and tissue-specific signals are integrated to regulate gene expression [An and Wensink 1995a,b]. Although the biochemical basis of fbe regulation remains speculative, functional integration of its discrete factor-binding sites has been demonstrated in transgenic flies. Such features are likely to reflect general principles of transcriptional activation in eukaryotes [Tjian and Maniatis 1994].

Genetic dissection of the fbe has novel and previously unrecognized structural consequences. Because bZIP proteins bind within the major groove [Ellenberger et al. 1992; Glover and Harrison 1995; Keller et al. 1995; Chen et al. 1998], the An-Wensink model [Fig. 2C] predicts that the DM domain—like SRY but unlike classic zinc fingers and other zinc modules—would bind in the minor groove. The model further requires that the DM domain—unlike SRY and other HMG boxes—would not induce sharp DNA bending lest associated compression of the major groove displaces bZIP1. The present studies have tested and confirmed these implicit predictions. Minor groove binding without bending is unusual and may be analogous to that of the T domain–DNA complex of the Brachyury transcription factor [Muller and Herrmann 1997]. The putative ternary DSXf–bZIP1–DNA complex would be remarkable for adaptive binding of one nascent α-helix in the DNA major groove (bzip1) and of another nascent α-helix in the overlapping DNA minor groove (dsxA). Methylphosphonate interference experiments suggest intimate juxtaposition of DSXf and bZIP1 complexes. bZIP1–phosphate contacts [blue circles in Fig. 2D], inferred from crystal structures of the Gcn4 bZIP–DNA complex [Ellenberger et al. 1992; Keller et al. 1995], overlap sites of DSX–methylphosphonate interference (red circles in Fig. 2D). Such interdigitation of protein–DNA phosphate contacts is consistent with a previous “missing base” interference assay suggesting corecognized bases [An and Wensink 1995b]. Simultaneous occupancy of major and minor grooves is likely to be a general feature of enhanceosome assembly [Tjian and Maniatis 1994].

The mechanism of synergistic transcriptional activation between DSXf and bZIP1 [Fig. 2C; An and Wensink 1995a,b] is not well understood. Studies of mammalian liver-specific gene regulation by C/EBPα and NF-Y have demonstrated that synergy can occur in the absence of cooperative promoter binding, binding of C/EBP in fact impairs binding of NF-Y and yet their simultaneous binding leads to synergistic activation of a minimal promoter [Milos and Zaret 1992]. Nonetheless, synergy between by CCAAT enhancer-binding factor α (C/EBPα) and nuclear factor Y (NF-Y) results in the formation of a preinitiation complex that is stable through multiple rounds of transcription. Additional biochemical studies will be required to distinguish whether [1] DNA binding by DSXf and bZIP1 is cooperative [a thermodynamic mechanism], [2] the ternary DSXf–DNA–bZIP1 complex dissociates more slowly than either binary complex [a kinetic mechanism], or [3] binding of DSX induces a conformational change in bZIP1 enhancing its potency in transcriptional activation [an allosteric mechanism]. The mechanisms by which DSXM represses bzip1 are also not well understood.

Mutations in DSX cause allele-specific changes in courtship behavior

doublesex is representative of a class of genes that not only specify aspects of body plan but also influence behavior [for review, see Yamamoto et al. 1996]. Chromosomal female [XX] flies expressing DSXM, although male in external appearance, do not court wild-type females [Villella and Hall 1996]. Chromosomal male [haplo-X] flies bearing a deletion in one dsx allele and various point mutations in the other exhibit specific and quantifiable changes in courtship behavior [Hall 1994; Villella and Hall 1996]. The haplo-X dsx phenotype differs from
that associated with *fruitless* [male chaining, Ito et al. 1992, Villella et al. 1997]. DSX-associated anomalies include an altered courtship song, selective lack of vigor in the pursuit of wild-type females, increased elicitation of courtship from wild-type males, and decreased rejection of subsequent attempted copulation [Villella and Hall 1996]. The mechanics of courtship and individual elements of mating behavior are normal with the exception of the extent of abdominal bending during attempted copulation and altered song [Villella and Hall 1996]. The latter is of particular interest: the humming component of subsequent attempted copulation (Villella and Hall 1996). Whether and how *fruitless, dsx* does not regulate the male-specific induction or female-specific suppression of the abdominal muscle of Lawrence [Gailey et al. 1991], which thus represents a dsx-independent branch of the ramified sex-determining pathway (Fig. 1A).

The extent and spectrum of behavioral change associated with the various point mutations in *dsx* are allele specific. The biochemical bases of these phenotypic differences have not been elucidated. Courtship behavior reflects the complex integration of multiple systems, including pheromone production, elaboration, and sensing, development of the peripheral nervous system and its central ramifications; and function of a putative courtship command center [Villella and Hall 1996] is not well characterized. In light of the allele-specific effects of mutations on extent of courtship derepression, the DSX pathway promises to provide an attractive model for study of the biochemical basis of a complex behavior.

Materials and methods

Protein synthesis

Solid phase peptide synthesis used an Applied Biosystems 430A synthesizer. t-Boc amino acids were used with the following protection: Arg[Tosyl], Asn[Xanthyl], Asp(OcHxl), Cys[4MeBzl], Glu(OcHxl), His[DNP], Lys[2ClZ], Ser[Bzl], Thr[Bzl], and Tyr[BrZ]. Peptide DSX(35–67) was synthesized on a carbboxy-terminal thioester-containing resin. The peptides DSX(68–105) and acetyl-DSX(36–105) were synthesized on a Boc-Glu(OcHxl)–OCH2-Pam resin, and DSX(68–86) on a Boc-Gln–OCH2–Pam resin. Full-length polypeptides were prepared by native chemical ligation [Dawson et al. 1994]. The amino-terminal peptide consisted of DSX residues 35–67; the carboxy-terminal peptide consisted of residues 68–105 [DM domain] or 68–86 [DMD fragment]. The R91Q variant domain was prepared similarly. Ligation products were purified by reverse-phase HPLC and characterized by electrospray mass spectrometry. Residue numbers refer to native sequence.

Analytical ultracentrifugation

Self-association was characterized in 50 mM Tris-HCl [pH 6.5] at 25°C using a 60Tvi rotor and Beckman model XL-A [20,000 rpm] at a concentration of 70 and 140 µM. Although data were inconsistent with a monomer–dimer model, the domain does not behave as an ideal species; fitting yielded an apparent molecular mass 9350 ± 590 Da larger than that of the monomer [8481 Da], presumably due to hydrodynamic consequences of the disordered tail. The sedimentation equilibrium of DMΔ (DSX residues 35–86) was consistent with an ideal monomer of expected mass.

DNA-binding assays

DNA-binding was monitored by GMSA [10% acryl-amide] in 0.5x TBE [1x, 89 mM Tris-borate (pH 8.0), and 2 mM ethylene-diamine-tetraacetic acid] at 175 V and 4°C. Sequences are given in Figure 3; binding was studied in 10 mM Tris-HCl [pH 7.4], 50 mM KCl, 0.1 mM ZnCl2, 34 µg/ml bovine serum albumin, and 0.6 ng/µl poly[dI–dC] DNA [unlabeled]. The concentration of 32P-labeled DNA was 1 nM; binding in Figure 4 was at higher salt [150 mM KCl] and nonspecific competitor DNA (0.8 ng/µl poly[dI–dC]) to limit nonspecific binding to the fbe probe. GMSA studies of DNA analogs containing systematic AT→IC substitutions used sequences 5’-TGGACACAC-CICICCGCCGCIICACGCTAGC-3’ annealed to 3’-AGCTG-TGTCGCGGCICICCGCCTGTCGTACCGG-5’. Methylphosphonate interference was evaluated in a 15-bp dsxA duplex [Labets and Weiss 1997]. To obtain predicted hZIP contacts [Fig. 2D], bzip1 sequence was aligned with ATF/CREB site of yeast activator GCN4 [5’-GAGATGACGTCATCC-3’; Keller et al. 1995]. Underlined bases indicate homology with bzip1 [5’-ACAAT-GTGCAGAACGCCG-3’; An and Wensink 1995a,b].

Solution structure of the DM domain

An fbe duplex [5’-GTGCAACACTAACAGTTGCAATCGAG-3’ and complement] containing dsxA was cloned in unique EcoRI site of pBluescript-M13 [Stratagene Inc., La Jolla, CA]. Probes of equal length, with the binding site at varying distance from ends were generated by PCR and 5’-labeled with 32P[ATP and T4 polynucleotide kinase. Binding reactions [25 µl] contained 150 mM KCl, 20 mM Tris-HCl [pH 7.4], 5 mM MgCl2, 20 ng poly[dI–dC], 0.8 ng bovine serum albumin, 5% glycerol, approximately 1 nM 32P-labeled DNA probe, and 24 nM purified DSX DM domain. Control binding reactions contained the SRY HMG box [King and Weiss 1993] at a concentration of 4 nM in 150 mM KCl, 60 mM Tris-HCl [pH 7.4], 5 mM MgCl2, 33 ng poly[dI–dC], 0.8 ng of bovine serum albumin, and 5% glycerol.

Circular dichroism

Spectra were obtained using an Aviv spectropolarimeter. Helix contents were estimated by magnitude of mean residue ellipticity at 222 nm, assuming that 100% α-helix corresponds to a value of −33,000 at 4°C. DNA-binding studies used site 5’-CAC-TACAAATTGTTGCA-3’ and complement [target site underlined]. The protein concentration was 10 µM in a 50 mM Tris-HCl [pH 6.5] buffer containing 50 mM KCl and a 2.5-fold molar excess of ZnCl2. The DNA concentration was 5 µM.

NMR spectroscopy

Samples were prepared as in CD studies using deuterated Tris-HCl buffer and protein concentrations of 60 µM, 460 µM, and 1.4 mM (intact domain) and 2 mM [DMΔ] and DM-p. Solutions were purged with N2 before use. Experiments were performed at 25°C on Varian Inova 400, 500, and 600 MHz spectrometers: NOESY
Binding sites were restrained to be 2.2 Å, Zn = 3.4 Å. Protected amide resonances in D₂O were observed at 25 °C as described (Qian et al. 1993). Root mean square deviation (RMSD) was calculated using program Superpose (Diamond 1992). RMSD for each residue was calculated using X-PLOR. The CA ensemble exhibits a main-chain RMSD of 0.43 Å. The eight Zn²⁺-binding residues are characterized with a precision of 0.13 Å (main chain RMSD) and 0.15 Å (side chain RMSD). In refined models respective Zn-S=O and S=O-S distances in metal-binding sites were restrained to be 2.2–2.4 Å and 3.6–3.9 Å. Zn–N and N–S distances were restrained to be 1.9–2.1 Å and 3.4–3.8 Å.

Assignment of hydrogen bond

Protected amide resonances in D₂O were observed at 25 °C in one-dimensional ¹H-NMR and TOCSY spectra obtained immediately after dissolution of Zn–DM powder. Protected amide protons were assigned as putative hydrogen bond donors; acceptors were identified based on preliminary DG/SA models. Two distance restraints were imposed per hydrogen bond. Nine protons were assigned as putative hydrogen bond donors; acceptors were identified. The former constraints were NH–O (1.8–2.3 Å) and NH–N (1.6–2.4 Å) or (side chain RMSD). In refined models respective Zn–S=O and S=O–S distances in metal-binding sites were restrained to be 2.2–2.4 Å and 3.6–3.9 Å. Zn–N and N–S distances were restrained to be 1.9–2.1 Å and 3.4–3.8 Å.

Acknowledgments

We thank P.K. Donahoe, E. Haas, A. Mahowald, and J. Rowley for discussion, K. Hallenga, Q.X. Hua, V. Ittah, D. Jones, L. Milos, M. ter Horst, and B. Wang for advice regarding methods, B. Li for database searches, M. Jeu for assistance with gel shifts, and W. Jia for assistance with figures. M.A.W. thanks A. Kosiakoff for encouraging the continuation of these studies in Cleveland. This work was supported by a grant from the NIH to M.A.W. and by Gryphon Sciences, Inc. L.-Y.Z. was an NIH Post-Doctoral Fellow in the Cardiovascular Pathophysiology and Biochemistry Training Program at The University of Chicago (5T32HL07237).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References


1762 GENES & DEVELOPMENT


