Cumulative contributions of weak DNA determinants to targeting the Drosophila dosage compensation complex

Gregor D. Gilfillan, Cornelia König, Ina K. Dahlsveen, Nicky Prakoura, Tobias Straub, Rosemarie Lamm, Torsten Fauth and Peter B. Becker*

Adolf-Butenandt Institut, Molekularbiologie, Ludwig Maximilian Universität, Schillerstrasse 44, 80336 Munich, Germany

Received February 20, 2007; Revised and Accepted April 11, 2007

ABSTRACT

Fine-tuning of X chromosomal gene expression in Drosophila melanogaster involves the selective interaction of the Dosage Compensation Complex (DCC) with the male X chromosome, in order to increase the transcription of many genes. However, the X chromosomal DNA sequences determining DCC binding remain elusive. By adapting a ‘one-hybrid’ assay, we identified minimal DNA elements that direct the interaction of the key DCC subunit, MSL2, in cells. Strikingly, several such novel MSL2 recruitment modules have very different DNA sequences. The assay revealed a novel, 40 bp DNA element that is necessary for recruitment of DCC to an autosomal binding site in flies in the context of a longer sequence and sufficient by itself to direct recruitment if trimerized. Accordingly, recruitment of MSL2 to the single 40 bp element in cells was weak, but as a trimer approached the power of the strongest DCC recruitment site known to date, the roX1 DH site. This element is the shortest MSL2 recruitment sequence known to date. The results support a model for MSL2 recruitment according to which several different, degenerate sequence motifs of variable affinity cluster and synergise to form a high affinity site.

INTRODUCTION

In Drosophila melanogaster, as in many animals, the female state is defined by the presence of two X chromosomes, but males contain only a single X chromosome in addition to the gene-poor Y chromosome. A priori, the genome of male fruit flies appears unbalanced due to the halved dosage of X chromosomal genes. Re-establishment of proper balance requires a compensatory mechanism that is able to distinguish the X chromosome from the autosomes and to bind the X selectively. The DCC consists of five core proteins (MSL1, -2, -3, MOF and MLE) and two non-coding RNAs, roX1 and roX2, which is able to distinguish the X chromosome from the autosomes and to bind the X selectively. The DCC consists of five core proteins (MSL1, -2, -3, MOF and MLE) and two non-coding RNAs, roX1 and roX2, which is able to distinguish the X chromosome from the autosomes and to bind the X selectively. The DCC consists of five core proteins (MSL1, -2, -3, MOF and MLE) and two non-coding RNAs, roX1 and roX2 (for review see (8) and references therein). Gene activation involves modification of X chromosomal chromatin by the DCC-associated acetyltransferase MOF, which acetylates histone H4 at lysine 16 (9–11), but contributions from more general factors, such as the H3 serine 10 kinase Jil1 (12,13), the supercoiling factor (14) and nuclear pore components (15) have been suggested. The targeting of these effects to the X chromosome relies primarily on MSL1 and MSL2, the two DCC proteins that are able to recognize a subset of sites on the X chromosome even in the absence of all other factors (16–18). However, faithful occupancy of all sites on the X chromosome requires additional factors, such as the activities of the acetyltransferase MOF and the RNA helicase MLE, as well as the presence of the roX RNAs (19,20).

How the DCC recognizes a single chromosome is a question of great interest. Ultimately, recognition must involve X-specific DNA sequences. Combining chromatin immunoprecipitation (ChIP) with probing of high-density
oligonucleotide arrays ‘tiling’ the entire X chromosome, 
the interaction of the DCC with the X chromosome has 
recently been mapped (21,22). These studies revealed 
that about 25% of the X chromosomal DNA is bound by 
DCC in tissue culture cells or Drosophila embryos and, 
notably, the majority of DCC binding is found within 
coding sequences, reaffirming an earlier suggestion that 
the DCC may act to facilitate transcription elongation 
rather than initiation (23). However, despite this 
wealth of interaction data it has not been possible to 
distil a set of ‘consensus’ DNA sequences that define DCC 
binding (21,22). It therefore remains possible that the 
observed X chromosomal binding pattern of the DCC is 
governed by more than just DNA sequence. DNA 
sequence may just define a subset of ‘primary’ targeting 
sites from which the DCC is distributed to secondary 
sites in neighbouring chromatin. Because the DCC 
interacts preferentially with active genes it is possible 
that the process of transcription itself or a transcription-
associated epigenetic modification of chromatin generates 
secondary sites (21,22).

While the nature of those presumed secondary sites is 
entirely unclear, the existence of primary sites, defined by 
DNA sequences with autonomous DCC recruitment 
activity, has been inferred from P-element-mediated 
insertion of X-derived sequences into autosomes, where 
association of the DCC with these ectopic sites can 
be monitored on polytene chromosomes by immuno-
histochemistry. The first DCC binding sites characterized 
in this way correspond to ~200 bp sequences found in the 
coding regions of the roX1 and roX2 genes (24,25). 
A prominent feature in these sequences is an abundance 
of GA sequences, and mutation of GAGAG tracts 
significantly reduced the recruiting power of the roX2 
element (25). Unfortunately, bioinformatic efforts failed 
to detect related DCC targeting sequences on the X 
chromosome (25) and the third binding site character-
ized in some detail lacked GAGAG sequences altogether 
(26). Primary DCC targeting elements thus differ in DNA 
sequence and may therefore belong to different classes.

According to the currently accepted model, many 
targeting elements of varying affinity are spread across 
the X chromosome. Strong targeting elements are able to 
automonomously recruit the DCC to an ectopic integration 
site on an autosome even at reduced concentrations of 
MSL proteins. In contrast, weaker sites are only bound in 
the X chromosomal context, presumably because the 
density of targeting elements leads to an increased local 
DCC concentration (26–30). This model was 
substantiated by the analysis of 11 X-chromosomal 
fragments of varying affinity for DCC isolated by ChiP 
(29). An attempt to identify sequence motifs responsible 
for DCC recruitment from a subset of these ‘DCC Binding 
Fragments’ (DBFs) with high- and moderate-affinity 
highlighted a number of clustered motifs (29). 
Accordingly, a DCC binding site may be composed of 
clusters of variable combinations of several degenerate 
sequence motifs. However, perhaps due to the relatively 
large size of the fragments analysed, it was not possible to 
predict further DCC binding based on the clustering 
of these motifs (29).

Thus, several previous studies indicate that the 
DNA sequences comprising high affinity DCC binding 
sites are diverse, possess varying affinities for the DCC, 
and can be dispersed over several kb (21,22,24,26,29). 
Given the degeneracy of the DNA motifs seen in 
association with DCC interaction, one has to assume 
that the contributions of individual elements to overall 
DCC targeting may be small and hence difficult to 
document by established methodology. In order to 
monitor the effect of weak targeting determinants, we 
developed a sensitive transfection-based ‘one-hybrid’ 
assay that amplifies weak DNA interaction events into a 
strong transcriptional read-out. The assay allows rapid 
identification of sequence elements able to recruit the 
MSL2 protein. We applied the assay to localize the 
DCC binding determinants within several previously 
described high affinity DBFs (29). We describe several 
sequence motifs that contribute to MSL binding and show 
that high affinity MSL binding sites can be generated by 
oligomerization of weaker elements.

MATERIALS AND METHODS

Drosophila genetics

Fly genetic manipulation and crossing was performed as 
in ref. (29). A more detailed description is provided in the 
Supplementary Data online.

FISH and immunostaining of polytene chromosomes

FISH and Immunostaining were performed exactly as 
described in (29). For immunostaining, one of two affinity 
purified rabbit anti-MSL1 antibodies kindly provided by 
M. Kuraku (31) and E. Schulze (32) were used at a 
dilution of 1:200 and 1:400, respectively.

DNase-I hypersensitive site mapping

DNase-I hypersensitive (DH) sites in adult flies were 
mapped as described (24), except that nuclei from 2.5 g 
adult flies, sorted according to sex, were divided into 
seven portions and digested with a titration of up to 
60 units DNase-I (Roche, Penzberg, Germany). Ten µg 
DNase-I-digested DNA per lane was digested with 
restriction enzymes as described in figure legends, run on 
0.8% agarose gels, then blotted and hybridized as 
described (33).

Northern hybridization

Total RNA was recovered by grinding adult male or 
female flies under liquid nitrogen, then extracting 
the powder with Qiagen (Qiagen, Hilden, Germany) 
reagent according to manufacturer’s instructions. Twenty µg of RNA from male and females was run on 
a 1.2% agarose gel containing formaldehyde, then 
transferred and hybridized according to standard labora-
tory protocols (34).

Plasmid constructs

Plasmids made for use in this publication are detailed in 
Supplementary Table 2.
Transient transfection and luciferase assay

SL2 cells were maintained in culture at 26°C and split every 3–4 days. For transfection, cells were diluted to 2.5 × 10^5 cells/ml at 1 ml per well of a 12-well tissue culture plate. The following day, the cells were transfected using the Effectene reagent (Qiagen) according to manufacturer’s instructions, with the single exception that 20 μl Effectene reagent was used per μg of transfected DNA. Per well of cells, the following amounts of plasmid were transfected: pRL-TK, 15 ng; pGL3-TK and derivatives, 315 ng; pVP16 and other activator constructs, 160 ng. Each tested firefly reporter plasmid was transfected into duplicate wells, and on at least two occasions using DNA prepared from two different maxipreps. DNA was prepared using Qiagen or Promega maxiprep kits. Two days following transfection, cells were harvested by centrifugation and washed in 1 ml PBS. Cells were then lysed and luciferase activities determined using the Dual Luciferase Kit (Promega, Mannheim, Germany) according to manufacturer’s instructions. Light emission was measured using a Lumat 9501 Luminometer (Berthold, Bad Wildbad, Germany).

Western blotting

Western analysis was performed according to standard laboratory protocols (34) using antibodies directed against MSL1 (32), MSL2 and HSV VP16 (Santa Cruz Biotech, SC7545, Santa Cruz, CA, USA).

RESULTS

DNase-I hypersensitive site mapping is of limited value for DCC binding site localization

The previously isolated DBF5, DBF6, DBF7, DBF9 and DBF12 are considered to contain high affinity binding sites for the DCC (29). However, given that these fragments are between 2.5 and 6.7 kb long, localizing the targeting elements requires further mapping. So far, DBFs have been characterized by integration of candidate fragments into autosomes through P-element-mediated gene transfer and monitoring of DCC recruiting power by MSL1 immunostaining on polytene chromosomes. In order to guide the construction of further clones for this analysis we first explored whether DNase-I hypersensitivity (DH) would highlight regions of interest. DNase-I hypersensitivity analysis indicates chromosomal loci where chromatin is disrupted due to the interaction of non-histone proteins (35). The three known high-affinity DCC binding sites within the roX genes and at cytological position 18D10 (24,25,36) all contain regions of male-specific DNase-I hypersensitivity, and in the case of the roX genes these isolated sites are alone able to recruit the DCC.

Adult flies were sorted according to sex, nuclei prepared and treated mildly with DNase-I to digest only the most exposed sites in chromatin. DH sites within the DBFs were identified by indirect end-labelling (Supplementary Figure S1). DH sites were present in every DBF clone, but notably, only two contained male-specific sites. A strong and a weak general DH site (i.e. common to both sexes) were found in DBF12 within intronic sequences of the Smr gene. A single, weak male-specific site was found in DBF6, spanning an intron and coding sequences of the nej gene. The remaining DH sites revealed in clones DBF-5, -7 and -9 (Supplementary Figure S1) are summarized as follows: Two general sites were found at the 5’ ends of two neighbouring genes (CG15892 and CG3815) in DBF5. In DBF7, a male-specific site was seen in the vicinity of two small introns of CG2025, and a general site was present between the CG2025 and CG1847 genes. Lastly, in DBF9, three DH sites present in both sexes were found in an intron of the Tao-1 gene.

To examine whether DH sites found in both sexes could direct DCC targeting, the strong (DHS-S) and weak (DHS-W) general DH sites from DBF12 were chosen for rigorous testing of DCC recruitment potential. P-elements containing these sequences were generated and integrated into autosomes of transgenic flies. We determined the insertion sites by DNA FISH on polytene chromosomes (data not shown) and then monitored the recruitment of DCC to these ectopic sites by staining with MSL1 antibody. The 480 bp DHS-S was capable of recruiting MSL1 in both wild type and msl-3^I male-specific mutant flies (Figure 1A and B), demonstrating that this sequence contains a high affinity DCC binding site. In contrast, a 700 bp fragment containing DHS-W failed to recruit MSL1 proteins in three transgenic lines under wild type DCC expression (data not shown).

The single, weak male-specific site found in DBF6, spanning an intron and coding sequences of the nej gene, was also tested for DCC recruitment in transgenic flies. A 500 bp fragment containing the DH site and flanking sequences recruited MSL1 to an autosomal insertion site in wild type males (Figure 1C), but was relatively weak compared to the DBF12 DHS-S. No MSL1 binding could be detected on the DBF6 DH insertion in flies carrying the msl-3^I mutation (data not shown) or at the reduced DCC levels obtained when MSL2 is provided only by the expression mutant SXB-1 or NOPU in females (data not shown) (27,29,37). Only under conditions of MSL1 and MSL2 overexpression was robust binding to the DBF6 DH site seen (Figure 1D). A hexamer of this sequence showed improved binding in wild type flies (Figure 1E), but was still not capable of recruiting the DCC at lower DCC concentrations that characterize the SXB-1 and NOPU genetic backgrounds. Therefore in this instance, a male-specific DHS was not sufficient to define a high affinity binding site for the DCC, and the presence of a male specific site is a poor indicator of DCC recruiting ability.

A ‘one-hybrid’ transfection assay to examine MSL2 recruitment to candidate DNA

It had become clear that DHS mapping was of only limited diagnostic value for the identification of DCC interaction sites. Also, P-element-mediated transgenesis was considered too time consumping to attempt narrowing down the DBFs to minimal DNA elements. We therefore...
developed a novel strategy for the characterization of candidate DCC targeting sequences employing co-transfection of three plasmids in male Drosophila SL2 cells. The assay is based on the activation of a reporter gene after transient transfection into SL2 cells (Figure 2A). Candidate DCC binding sites are cloned in front of the firefly luciferase gene, which is driven by a minimal thymidine kinase (tk) promoter. To convert recruitment of a key DCC component (MSL2) into a robust signal, an MSL2-VP16 fusion protein is expressed from a second, co-transfected plasmid. This hybrid protein consists of the entire MSL2 protein, to which the C-terminus (38,39), is fused. The assay therefore does not measure dosage compensation, but recruitment of a key DCC component (MSL2) into a recruitment site is shown in Figure 3. From the first two clones splitting the DHS in two (DBF12-L2 and -L4), recruited element lies from the promoter may influence the expression of firefly luciferase. Within the roX1 cDNA, the DHS is the primary element responsible for DCC recruitment (24), and in the cDNA lies ~2.3 kb from the promoter. The effect of distance from the promoter was examined in greater detail with subclones derived from DBF12 (see below).

Isolation of a novel MSL2 binding element from DBF-12

Having determined that the one-hybrid assay could be used to isolate small sequences crucial to DCC binding, we applied it to isolate minimal MSL2 binding elements on three clones (DBF-6, -9 and -12) and performed less detailed analysis on DBF-5 and DBF-7. We first applied the one-hybrid assay to the ~500 bp DBF12 DHS-S. A summary of the clones analysed to map an MSL2 recruitment site is shown in Figure 3. From the first two clones splitting the DHS in two (DBF12-L2 and -L4), sequence, either directly or indirectly through incorporation into a binding complex, the VP16 activator will boost the expression of firefly luciferase. The resulting increase in luciferase activity can be expressed as activation over the light emission seen with the control activator, which consists of a mutant version deleted for most of MSL2, leaving the VP16 activation domain alone. A second, obligatory normalization control for non-specific effects, such as transfection efficiency, involves co-transfection of a Renilla luciferase expression vector.

We first tested the functionality of the assay using the entire roX1 cDNA, and the 217 bp roX1 DHS, which are known to contain high affinity DCC binding sites (see above; Figure 2B). As can be seen when comparing the empty firefly vector pGL3 to those containing the entire roX1 cDNA or the isolated DHS site, the basal level of luciferase activity (in the presence of VP16 only) is heavily influenced by sequences inserted into the vector. The roX1 cDNA produced a drop in luciferase expression relative to empty vector, whilst the DHS sequence alone caused an increase. This phenomenon was seen with all sequences tested, but the roX1 sequences shown represent the most extreme cases of repression and activation observed. It most likely reflects the juxtaposition of both real and cryptic transcription factor binding sites to the promoter driving firefly luciferase expression, or of nucleosome exclusion sequences, and must be considered non-specific.

The effect of recruiting MSL2-VP16 is therefore best expressed as the fold activation in luciferase activity in the presence of pMSL2-VP16 over that seen with pVP16 (Figures 2B and 4). The assay yielded a reproducible, ~5-fold activation of transcription in the presence of roX1 cDNA, but the roX1 DHS led to a robust 24-fold activation. An additional MSL2 construct, fused to GFP instead of VP16 and previously shown to be capable of substituting the native MSL2 protein (41), showed negligible activation over pVP16 in the assay, demonstrating the need for the artificial VP16 activator fusion to observe MSL2 binding (Figure 2B).

The stronger activation of the roX1-DHS compared to the entire cDNA suggested that the distance an MSL2 recruiting element lies from the promoter may influence the level of luciferase activation. Within the roX1 cDNA, the DHS is the primary element responsible for DCC recruitment (24), and in the cDNA lies ~2.3 kb from the promoter. The effect of distance from the promoter was examined in greater detail with subclones derived from DBF12 (see below).
it was clear that the MSL2 recruiting sequences resided in the 5' half of the DHS, within DBF12-L2.

Proximity of the MSL2 recruiting element to the tk promoter influences luciferase expression. In several instances, cloning the DBF12 fragments in the reverse orientation resulted in a higher luciferase activity, suggesting the MSL2 binding site was closer to one end of the fragment, which was indeed the case (compare clones L13, L14 and L15). In an effort to systematically evaluate this distance effect we modified L2 by inserting varying lengths of spacer DNA from a fragment of DBF12 shown to be devoid of targeting elements and found that the assay would only give robust activation when the MSL2 binding site was within 350 bp of the promoter (Supplementary Table 1). This is however a conservative estimate, because the MSL2 binding site within the DBF12-L2 clone itself already lies more than 100 bp from the promoter (see below).

Restricting clone DBF12-L2 into ever-smaller fragments led to the identification of a 40 bp MSL2 binding element (DBF12-L15, see Figure 3A), which retained the ~4-fold activation potential of DBF12-L2. Trimerization of L15 led to an MSL2-dependent 24-fold enhanced luciferase activity (Figure 3A), rendering this oligomer almost as potent as the native roX1 DHS (Figure 4).

Clearly, the interaction of MSL2-VP16 can be boosted by clustering of interaction modules.

We used the one-hybrid assay to further define and mutate the minimal MSL2 binding sequence within DBF12. The sequence of DBF12-L15 contains runs of adenosines and a GAGA sequence (Figure 3B). A still shorter 16 bp element retaining these sequences (DBF12-L22) still promoted MSL2-dependent luciferase activation (Figure 3B). Mutating blocks of 4–5 bp of this 16 bp ‘core’ sequence in the context of the original 40 bp L15 fragment (clones L18, L19 and L23) caused a complete loss of MSL2 recruitment, confirming the importance of this element for MSL2 binding. Conversely, mutation of a 5 bp sequence outwith this core did not abolish activity (clone L20). To explore whether the ‘GAGA motif’ in L22 was important we systematically mutated the last ‘GA’ dinucleotide. The adenine was shown to be dispensable (clones L30–L32). In contrast, a transversion of the adjacent guanine to a cytosine (L33) or a thymine (L35) resulted in a complete loss of MSL2 recruitment, whilst a transition to an adenine (L34) had little or no effect. These experiments therefore allow the requirements for MSL2 recruitment to be dissected at the single nucleotide level.

Figure 2. A ‘one-hybrid’ transfection assay for isolation of DCC binding sites. (A) Summary of the transient transfection assay (inset) and transfected plasmids (not to scale). pGL3-TK is modified by the insertion of candidate DCC binding sites (indicated by question marks) in front of a minimal Herpes Simplex Virus tk promoter, and is transfected together with pRL-TK and either pMSL2-VP16 or pVP16 into Drosophila SL2 cells. (B) An example transfection experiment performed with modified pGL3 vectors containing the roX1 c3 DNA and the roX1 DHS. Numbers above the bars are fold-activation of MSL2-VP16 over VP16 luciferase expression for replicate experiments for the roX1 constructs. Western blots above graphs show expression of activator constructs pVP16, pMSL2-VP16 and pMSL2-eGFP. Transfection of pVP16 resulted in the expression of two proteins migrating at ~27 kDa. Transfected VP16 and eGFP fusions of MSL2 appear as a single additional band migrating above the endogenous protein.
Diverse sequences may direct MSL2 binding

Establishing general rules for MSL2 recruitment requires the isolation of a panel of binding elements. We therefore applied the one-hybrid approach to the DH sites of DBF6 and DBF9. The DBF6-L4 fragment, which contains the male-specific DH site, recruited MSL2-VP16 to a similar extent as DBF12 (Figure 5A). Sequentially removing sequences from the 3’ end of the DH site first caused an increase in luciferase activity, most likely due to moving MSL2-binding sequences closer to the promoter, followed by complete loss when the intron in the centre of the DH site was removed (compare clones DBF6-L4, L5, L6 and L7). A clone containing 68 bp of only intronic sequences (L8) recruited MSL2, although the fact that clone L9 showed double this activity argued that additional 5’ nef coding sequences, incapable of attracting MSL2 on their own, contribute to recruitment (Figure 5A). The presence of similar contributing elements in coding sequences 3’ to the intron is also suggested by the higher activity of clone L5 than L6. Removing 17 bp from the 5’ end of L8 did affect its activity somewhat (L10), but further removal of a CGAGAAA sequence (L11) almost abolished activity. Further deletion into a GA repeat resulted in complete loss of MSL2 recruitment (L12). Restriction from the 3’ end (L13 and L15) suggested that the central GA repeat may form a core element, flanked by other sequences that contribute to MSL2 recruitment.

Fine mapping of DBF6 was therefore carried out on the DBF6-L15 fragment, which consists of a GA repeat flanked by two short sequences (a 5’ CGAGAAA and a 3’ TATA motif, Figure 5B). Clones L18, -L19 and -L23 were constructed with dinucleotide substitutions within the 5’ CGAGAAA sequence. None of the dinucleotide mutants caused complete loss of activation compared to the parent clone, suggesting that, similar to the mutants of DBF12 discussed above, mutation of individual bases (or pairs in this case) can be tolerated within an essential motif. Likewise, mutation of the entire 3’ TAT motif (clone L17) revealed its importance, but single base substitutions (L20–L24) had little effect. Deletions and substitutions within the central dinucleotide repeat region highlighted the role of uninterrupted GA repeats (L26–L31). In order to explore whether GA repeats alone were able to recruit MSL2 we mutated three bases within L15 to construct a fragment consisting entirely of GA repeats (L16). This sequence did not support any MSL2 recruitment, confirming the importance of the CGAGAAA and/or TATA motifs flanking the GA repeat (Figure 5B). However, a trimer of the 26 bp L16 GA repeat sequence demonstrated robust, 7.8-fold MSL2 recruitment, providing a second example for the earlier notion that weak elements, which by themselves are unable to recruit MSL2, may gain affinity by oligomerization and clustering (see also Figure 4).

A similar approach was also applied to DBF9. Three DH sites, named A, B and C had been identified in intronic sequences at the 3’ end of this clone (Supplementary Figure S1C). Each of the DH sites was analysed, but only DHS-C exhibited modest MSL2 recruitment (Figure 6A). Dividing DHS-C into clones

---

**Figure 3.** Identification of a 40 bp MSL2 binding sequence within DBF12. (A) DBF12 is illustrated with the identified DH sites shown as grey bars above a scale in kb. Exons (boxes) and introns (lines) of the Smr gene are indicated, with the arrow showing direction of transcription. Subsequences containing DHS-S are magnified, beneath which the sequences cloned into the firefly luciferase reporter vector pGL3 are indicated. Arrows under the heading ‘orientation’ refer to the direction the fragment was cloned into the pGL3 vector; a forward pointing arrow represents clones with the 3’ end of the sequence closest to the tk promoter. Activation of firefly activity with co-transfected pMSL2-VP16 over firefly luciferase activity with pVP16 is indicated, ±standard deviation. (B) Fine mapping deletion and mutant clones derived from DBF12-L15. Mutated bases are underlined. Sequences shown are the reverse complement to their actual insertion in pGL3, to allow easier comparison to GA sequences in DBF6 in Figure 5. Activation of firefly activity with co-transfected pMSL2-VP16 compared to firefly luciferase activity with pVP16 is indicated, ±standard deviation.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Fold activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBF12-L10</td>
<td>CACATGGCGTAGCCAATCACCAAATCAAAGAAATGGCCGCA</td>
<td>3.2 ± 0.8</td>
</tr>
<tr>
<td>DBF12-L11</td>
<td>AAAAATCAAAATCAAAATCAAAGAAATGGCCGCA</td>
<td>3.0 ± 0.3</td>
</tr>
<tr>
<td>DBF12-L12</td>
<td>CACATGGCGTAGCAAAATCAAAGAAATGGCCGCA</td>
<td>1.0 ± 0.1</td>
</tr>
<tr>
<td>DBF12-L13</td>
<td>AAAAATCAAAATCAAAGAAATGGCCGCA</td>
<td>0.9 ± 0.1</td>
</tr>
<tr>
<td>DBF12-L14</td>
<td>CACATGGCGTAGGAAATCAAAGAAATGGCCGCA</td>
<td>0.9 ± 0.2</td>
</tr>
<tr>
<td>DBF12-L15</td>
<td>AAAAATCAAAATCAAAGAAATGGCCGCA</td>
<td>0.8 ± 0.1</td>
</tr>
<tr>
<td>DBF12-L16</td>
<td>CACATGGCGTAGGAATCAAAGAAATGGCCGCA</td>
<td>2.8 ± 0.1</td>
</tr>
<tr>
<td>DBF12-L17</td>
<td>AAAAATCAAAATCAAAGAAATGGCCGCA</td>
<td>2.3 ± 0.6</td>
</tr>
<tr>
<td>DBF12-L18</td>
<td>CACATGGCGTAGGAAATCAAAGAAATGGCCGCA</td>
<td>3.3 ± 0.5</td>
</tr>
<tr>
<td>DBF12-L19</td>
<td>AAAAATCAAAATCAAAGAAATGGCCGCA</td>
<td>0.9 ± 0.0</td>
</tr>
<tr>
<td>DBF12-L20</td>
<td>CACATGGCGTAGGAAATCAAAGAAATGGCCGCA</td>
<td>0.9 ± 0.0</td>
</tr>
<tr>
<td>DBF12-L21</td>
<td>AAAAATCAAAATCAAAGAAATGGCCGCA</td>
<td>2.8 ± 0.1</td>
</tr>
<tr>
<td>DBF12-L22</td>
<td>CACATGGCGTAGGAAATCAAAGAAATGGCCGCA</td>
<td>3.3 ± 0.2</td>
</tr>
<tr>
<td>DBF12-L23</td>
<td>AAAAATCAAAATCAAAGAAATGGCCGCA</td>
<td>1.0 ± 0.1</td>
</tr>
<tr>
<td>DBF12-L24</td>
<td>CACATGGCGTAGGAAATCAAAGAAATGGCCGCA</td>
<td>1.0 ± 0.1</td>
</tr>
</tbody>
</table>
L6 (or L9) and L10, revealed that most MSL2 recruitment partitioned with clone L10. Restriction of L10 from the 3' end resulted in a modest increase in activity (L7), suggesting that MSL2 recruiting elements had been moved closer to the promoter. Further efforts therefore focused on restricting the L7 clone to a minimal activation element. Clones L11 and L12 defined a 5' end for the MSL2 recruiting sequences. Trimming sequences from the 3' end of L11 resulted in a sequential loss of MSL2 recruitment (L13–L15; see also Figure 6B). These observations suggested that the 5' end of L11 contains an essential 'core' MSL2 recruiting element, followed at the 3' end by weaker elements, which are not essential but contribute in a cumulative fashion to MSL2 recruitment.

Fine mapping of the DBF9 fragments therefore focused on mutating bases in the 5' 'core' or 3' accessory elements to confirm this hypothesis (Figure 6B). Mutating two 5 bp blocks in the core of L13 abolished recruitment activity (clones L16 and L17), whilst disrupting two CACA elements in the 3' end resulted only in reduced activity (clones L18 and L19). Therefore, these experiments confirmed that DBF9 contained an important element of ~25 nucleotides, flanked by accessory elements that contribute to MSL2 binding. The analysis pointed to the CA dinucleotide as one such accessory motif.

In summary, examination of the smallest sequences isolated capable of recruiting MSL2 in the transfection assay revealed a striking lack of sequence similarity. Whilst both DBF12-L15 and DBF6-L15 share the AGAGA motif, and both these clones and DBF9-L13 are generally A-rich, the only truly conserved motifs shared by all three are AGA, GAG and AAA. These motifs are also found in two further MSL2 recruitment sites we identified in a less systematic search in two introns within the DBF7 and DBF5 sequences (Supplementary Figure S2).

Extensive attempts to find larger motifs conserved between the sequences by allowing degenerate positions, or including flanking sequences that might contain motifs missed in this analysis, failed to return significant results.

None of the DCC recruiting regions span male-specific RNAs

Both roX genes encode non-coding RNAs that span also the DHS sequences responsible for DCC recruitment, although in the case of roX2 very few, if any, transcripts read through the DHS (25). At the 18D site, however, no RNA could be detected in the region of the male-specific DH site (26). We therefore examined all the DBF sequences seen to recruit the DCC for transcripts in male and female adult flies by Northern blotting. In contrast to both roX DH sites, but similar to the 18D site, no male-specific transcripts could be detected (Supplementary Figure S3), although we cannot exclude that we have missed rare, large (>8 kb), developmentally regulated transcripts, or micro-RNAs.

The one-hybrid assay identifies bona fide DCC binding sites

Comparison of the results presented here to high resolution ChIP-chip data for MSL1 (22) demonstrates that most MSL2-recruiting sequences are found within peaks of MSL1 binding (Supplementary Figure S4). These peaks are embedded in broader regions of MSL1 binding and it is currently unclear whether these interactions are entirely defined by DNA sequence or whether secondary targeting determinants, such as histone modification marks, contribute to the observed profile. In order to ascertain that the one-hybrid assay indeed selects high affinity DCC binding sites, we tested whether the DBF12-L15 fragment was able to recruit MSL1 to ectopic, autosomal locations in transgenic flies. An autosomal insertion of the 40 bp DBF12-L15 was not able to recruit the DCC in wild type flies (Figure 7A). However, deletion of the L15 fragment from the DBF12 DHS-S clone caused loss of DCC recruitment, even in a wild type male background, confirming the essential nature of the
40 bp L15 sequence (Figure 7B). Notably, a trimer of L15 recruited DCC to a similar or greater extent than the parent DHS-S construct in wild type males (Figure 7C, compare to Figure 1A), and also demonstrated robust recruitment at low DCC concentration (SXB-1 background; Figure 7D). Thus, similar to the high affinity site at 18D (26), multimerizing an essential element not sufficient to be classed as a high affinity site on its own created an artificial high affinity DCC binding site. Therefore, DBF12-L15 may well be the shortest DCC targeting element identified to date.

**DISCUSSION**

A novel assay to map DCC targeting elements

How the DCC of *Drosophila* recognizes the X chromosome for selective interaction is an unsolved question. Although there is ample evidence that DNA sequences are involved, defining consensus sequence elements that may serve as binding sites for DCC components has been difficult. The available evidence points to the existence of different sequence motifs, clustering in regions covering several kb, which form the highest affinity binding sites. Such a definition necessitates testing numerous candidate binding sites and extensive mutagenesis. So far, the established method to evaluate X chromosomal sequences for DCC recruitment is time consuming since it involves generating stable fly lines containing candidate sequences integrated into an autosome. The ‘one-hybrid’ strategy we introduced abbreviates this process dramatically. Fusing the VP16 transactivation domain to MSL2 leads to a robust activation of a reporter gene provided that MSL2 is targeted to the candidate DNA upstream of a minimal promoter. This strategy has several important features. First, the assay solely measures chromosome binding of MSL2 without constraints imposed by a requirement for normal function in dosage compensation. This allows mutating MSL2 regardless of potential consequences on functions other than recruitment. Second, for the assay to work it does not matter whether MSL2 binds the chromosome directly or indirectly via an adaptor molecule or even the entire DCC. Third, the assay appears to be more sensitive for the identification of minimal targeting determinants than the polytene chromosome recruitment assay. The DBF12-L15 fragment would have been missed in the chromosome recruitment assay because its affinity for MSL2 is too weak if present as a monomer. The element is nonetheless essential for DCC binding. However, the dramatic increase in MSL2-responsiveness upon trimerization led to uncovering its autonomous recruitment potential in flies. This enhanced sensitivity may be due to the fact that MSL2 is overexpressed in SL2 cells and hence present in artificially high concentrations that allow recognition of weak elements. Native high affinity sites may be composites of several weak elements (see below) that individually are unable to attract DCC autonomously, but which can be detected in the one-hybrid assay. Finally, the assay may be adapted to a high-throughput format, which should allow screening many DNA sequences in parallel.

![Figure 5. Restriction of DBF6 with the one-hybrid assay. (A) DBF6 is schematized, as in Figure 3, with two magnified sections. Activation of firefly activity with co-transfected pMSL2-VP16 compared to firefly luciferase activity with pVP16 is indicated, ± standard deviation. (B) Fine mapping and deletion clones derived from DBF6-L15.](image-url)
Figure 6. Restriction of DBF9 with the one-hybrid assay. (A) DBF9 is schematized, as in Figure 3, with two magnified sections. For the illustrated transcripts, filled boxes indicate coding sequence and empty boxes contain untranslated regions. Activation of firefly activity with co-transfected pMSL2-VP16 compared to firefly luciferase activity with pVP16 is indicated, ± standard deviation. (B) Fine mapping and deletion clones derived from DBF9-L13.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Fold activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBF9-L3</td>
<td>ACGCAACCGAAAAAGAACATGACATACACACACGTACACGAGCATACGCATT</td>
<td>0.9 ± 0.1</td>
</tr>
<tr>
<td>DBF9-L4</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>0.7 ± 0.0</td>
</tr>
<tr>
<td>DBF9-L5</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>2.1 ± 0.1</td>
</tr>
<tr>
<td>DBF9-L10</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>2.3 ± 0.0</td>
</tr>
<tr>
<td>DBF9-L6</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>1.5 ± 0.1</td>
</tr>
<tr>
<td>DBF9-L9</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>1.5 ± 0.2</td>
</tr>
<tr>
<td>DBF9-L7</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>2.6 ± 0.2</td>
</tr>
<tr>
<td>DBF9-L11</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>5.2 ± 1.3</td>
</tr>
<tr>
<td>DBF9-L12</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>0.7 ± 0.3</td>
</tr>
<tr>
<td>DBF9-L13</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>3.2 ± 0.4</td>
</tr>
<tr>
<td>DBF9-L14</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>1.4 ± 0.1</td>
</tr>
<tr>
<td>DBF9-L15</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>1.0 ± 0.2</td>
</tr>
<tr>
<td>DBF9-L16</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>0.8 ± 0.0</td>
</tr>
<tr>
<td>DBF9-L17</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>1.0 ± 0.4</td>
</tr>
<tr>
<td>DBF9-L18</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>1.9 ± 0.1</td>
</tr>
<tr>
<td>DBF9-L19</td>
<td>ACGCAACCGAAAAAGAACATGACATGCACATACACACACGTACACGAGCATACGCATT</td>
<td>1.9 ± 0.3</td>
</tr>
</tbody>
</table>
Notably, the DNA sequence elements identified as targeting determinants resemble those found earlier with the more established assay (see below (22,25,29)). However, at this point we cannot exclude that the assay only detects a subset of DCC binding sites with special characteristics. For example, all the core sequences identified to recruit MSL2 in this analysis lie in non-coding regions of the genome, whereas the majority of DCC binding is seen in coding regions (21,22). According to a recent model (8), the DCC may interact with chromosomes in two (or more) distinct binding modes: a primary mode, determined largely by DNA sequences, and a secondary mode employing transcription-associated epigenetic features. A similar model has recently been proposed for dosage compensation in C. elegans (42). Accordingly, distribution of the DCC over the X chromosome may involve primary recruitment to a subset of sites (possibly including those identified in this analysis) from which DCC is distributed to the majority of secondary sites (8).

Identification of novel targeting determinants

The one-hybrid assay allowed the fast mapping of minimal MSL2 targeting elements within larger DBFs identified by conventional means. These are the smallest known binding sites for the DCC. Deleting sequences from the DBFs in our quest for minimal elements we noticed the existence of ‘accessory’ elements, which by themselves are not sufficient to recruit the DCC in the transfection assay, but in the vicinity of a ‘core’ element contribute to the overall affinity. One such accessory motif consists of short CA dinucleotide repeats. The core elements appear purine-rich on one strand, and although purine-pyrimidine transversion affected activity of one nucleotide position, it could not account for all observed changes in activity, nor was the length of purine tract required for activity consistent between the different clones. The results therefore suggested that despite a general tolerance of mutation, some nucleotide positions within the core may be more important than others. However, the only sequence motifs shared between the three shortest elements isolated are AGA, GAG and AAA, and attempts to build longer, more flexible motifs, did not produce convincing results. The importance of GA-rich sequences for DCC recruitment has already been established for the roX loci (25), and they are also common in the DBF clones (29). However, the removal of the GAG motif from the shortest clones isolated from DBF9 (clones DBF9-L13 and -L14) does not lead to complete loss of MSL2 recruitment, confirming that not even this is an essential motif. On the other hand, oligomerization of an element that essentially only consists of GA repeats can recruit MSL2 in our assay.

Together, these observations strengthen the earlier hypothesis that high affinity DBFs are composites of several distinct sequence motifs with variable DCC recruitment potential that synergize to generate a high affinity site. These motifs may be dispersed, but cluster to form high affinity DBFs. However, diversity in sequence appears not to be a fundamental requirement since high affinity sites can also be generated from homotypic elements by oligomerization, as shown here for two examples and as was also previously observed for a larger element (36). The relative tolerance towards point mutations emphasizes that these elements are degenerate. The binding specificity of the DCC therefore seems surprisingly plastic, which may explain the failure of genome-wide binding analyses to define a single consensus. The observed degeneracy implies that each single targeting determinant has a relatively low affinity for DCC, and we have to assume that the sum of many weak interactions effectively generate high-affinity DBFs (43). Ultimately, the future identification of a greater number of targeting elements should allow a better definition of the motifs recognized by the DCC. Our conclusions are in broad agreement with recent results from C. elegans, where isolated or clustered motifs can render a high affinity DBF, but are not sufficient to explain all observed DCC binding (44,42).

The nature of the binding sites

Emulating previous studies (24–26), DH sites were found in all of the five high affinity DCC binding sites studied in this analysis. However, we found for the first time DH sites common to both sexes that overlapped with DCC recruiting elements, and only two of the five DBFs contained a male-specific DH site. The single example of a male-specific DH site tested in the transgenic fly for DCC recruitment was not capable of recruiting the DCC with high affinity, suggesting that male-specific DNase-I hypersensitivity is merely a crude indicator of chromatin

Figure 7. The DBF12-L15 fragment is essential for DCC recruitment. (A) Polytenic immunofluorescence with anti MSL1 antibody (red) and DNA (blue), of an autosomal insertion of the 40 bp L15 sequence in wild type male background. Inset shows FISH staining (green) for location of the autosomal insertion. Arrows show the position of the wild type male background. Inset shows FISH staining (green) for DNA (blue), of an autosomal insertion of the 40 bp L15 sequence in wild type male background, (B) (L15) in wild type male background, (C) (L15) in wild type male background, and (D) (L15) in female SXB-1 background (w+/SXB1–2 D).
accessibility and does not reflect the strength of the underlying sequence to attract the DCC. This is similar to the findings of the DBF at 18D, where a male-specific DH site was necessary for high affinity of a larger fragment, but was alone not sufficient (26). Conceivably, regulatory elements attracting the DCC in males may have co-opted sequence elements and interacting factors that facilitate chromatin opening, and so perform this function also in females. Accessory elements may therefore not contribute to defining a DBF, but rather facilitate the interaction of the MSLs with DNA in chromatin. For example, runs of poly A/T tend not to be assembled into nucleosomes, which could aid interaction of the DCC (45). Furthermore, GAGAG sequences are found in many regulatory elements, where the interacting proteins, such as GAF, recruit nucleosome remodelling factors that render nucleosomal DNA accessible (46,47).

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank R. Aasland for kindly making remote bio-computing services available. We would also like to thank Angelika Mitterweger and Heike Mitlöchner for excellent technical assistance, E. Schulze, M. Kuroda and M. Meisterernt for providing antibodies, and M. Bartkuhn and R. Renkawitz for gifts of plasmids. This work was supported by the Deutsche Forschungsgemeinschaft through Transregio 5 and the Leibniz Programme and R. Renkawitz for gifts of plasmids. This work was technical assistance, E. Schulze, M. Kuroda and M. computing services available. We would also like to thank R. Aasland for kindly making remote bio-

REFERENCES

Downloaded from https://academic.oup.com/nar/article-abstract/35/11/3561/2402493 by guest on 30 December 2018