

Physical interaction between MSL2 and CLAMP assures direct cooperativity and prevents competition at composite binding sites

Nikolas Eggert , Fotios Gkoutromichos, Silke Krause, Aline Campos-Sparr and Peter B. Becker *

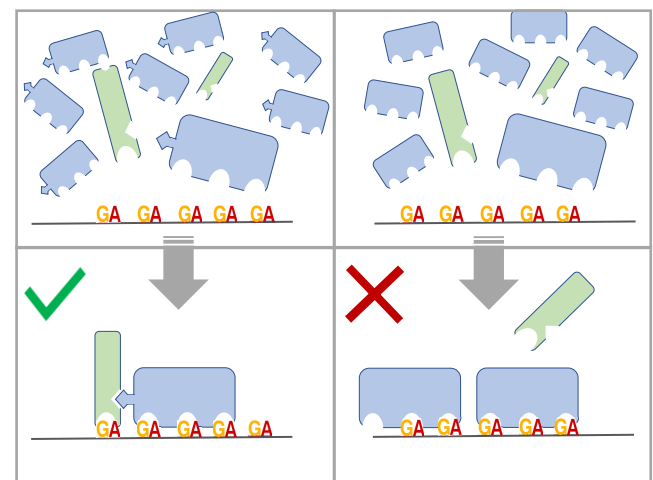
Biomedical Center, Molecular Biology Division, LMU, Munich, Germany

Received April 05, 2023; Revised July 13, 2023; Editorial Decision July 26, 2023; Accepted August 09, 2023

ABSTRACT

MSL2, the DNA-binding subunit of the *Drosophila* dosage compensation complex, cooperates with the ubiquitous protein CLAMP to bind MSL recognition elements (MREs) on the X chromosome. We explore the nature of the cooperative binding to these GA-rich, composite sequence elements in reconstituted naïve embryonic chromatin. We found that the cooperativity requires physical interaction between both proteins. Remarkably, disruption of this interaction does not lead to indirect, nucleosome-mediated cooperativity as expected, but to competition. The protein interaction apparently not only increases the affinity for composite binding sites, but also locks both proteins in a defined dimeric state that prevents competition. High Affinity Sites of MSL2 on the X chromosome contain variable numbers of MREs. We find that the cooperation between MSL2/CLAMP is not influenced by MRE clustering or arrangement, but happens largely at the level of individual MREs. The sites where MSL2/CLAMP bind strongly *in vitro* locate to all chromosomes and show little overlap to an expanded set of X-chromosomal MSL2 *in vivo* binding sites generated by CUT&RUN. Apparently, the intrinsic MSL2/CLAMP cooperativity is limited to a small selection of potential sites *in vivo*. This restriction must be due to components missing in our reconstitution, such as *roX2* lncRNA.

GRAPHICAL ABSTRACT



INTRODUCTION

Transcriptional regulation is a complex process involving multiple levels of control. At the first level of the ‘cis-regulatory code’ (1) transcription factors (TFs) bind to specific DNA sequence motifs (2,3). However, in the dynamic and competitive nuclear environment, the affinity of isolated binding sites is insufficient to ensure stable binding. Instead, productive interactions occur at regulatory regions where clusters of binding sites promote TF cooperation and synergistic DNA binding (1). This cooperativity underlies much of the combinatorial regulation of transcription, integrating multiple signaling cues.

To understand the fundamental principles of TF cooperativity and faithful binding site selection, we are studying the *Drosophila* dosage compensation system. The male-specific-lethal (MSL) dosage compensation complex (DCC) binds exclusively to the X chromosome in male cells. It boosts the transcription of genes to match the output of the two female X chromosomes (4). The DCC subunit MSL2,

*To whom correspondence should be addressed. Tel: +49 89 2180 75427; Email: pbecker@bmc.med.lmu.de

the main determinant of X chromosome-selective binding, has only limited intrinsic ability to identify X-specific DNA motifs, the so-called PionX sites (5,6). PionX sites contain a 5'-extended variant of the MRE. Depending on the presence of a distinctive DNA shape signature these sequences are recognized by MSL2 alone (PionX sites) or are strong MRE sequences where MSL2 still profits somewhat from CLAMP cooperativity (PionX motifs) (5).

To overcome this limitation, MSL2 cooperates with the ubiquitous zinc finger protein CLAMP ('chromatin-linked adaptor of MSL proteins'). The two TFs stabilize each other's binding at some 300 'high affinity sites' (HAS) on the X chromosome (7,8). These sites contain one or more so-called 'MSL recognition elements' (MREs), GA-rich motifs that constitute composite TF binding sites (5,9,10).

To study the intrinsic and cooperative DNA binding specificity of MSL2 and CLAMP, we have developed an *in vitro* approach that allows the analysis of TF binding in complex, physiological chromatin at a genome-wide level (11). The cell-free system relies on extracts from pre-MBT (mid blastula transition) *Drosophila* embryos [DREX, (12)], which assemble maternal stockpiles of chromatin components on genomic DNA. The assembled chromatin shows physiological nucleosome spacing, a rich complement of non-histone proteins and bound insulators surrounded by phased nucleosomes (11,13). In line with its pre-MBT origin, it is transcriptionally inactive and shows dynamic nucleosome mobility catalyzed by abundant ISWI-type nucleosome remodeling complexes. Because endogenous TFs are mostly absent, DREX-reconstituted chromatin provides a naïve substrate to test for genome-wide interactions of purified TFs with thousands of potential binding sites. The experimental strategy mimics the *in vivo* situation at the zygotic genome activation (ZGA), when the first TFs encounter a dynamic chromatin (14,15), but additionally allows to manipulate the quantity and quality (mutants) of added TFs at will.

When we explored the ability of MSL2 and CLAMP to bind individual MREs in chromatin we observed strong cooperativity between the two proteins, which was, however, not restricted to functional X chromosomal sites, but included many autosomal sites with similar sequences. We also found that the GAGA factor (GAF) competed with MSL2/CLAMP at non-functional sites, thereby improving the X-chromosomal enrichment (11).

Our current work extends this earlier study in several important ways. First, we investigate the mechanisms that underlie the cooperativity between MSL2 and CLAMP. This cooperativity may involve protein-protein interactions that result in increased residence times of the TFs at composite *cis*-elements and lead to a synergistic effect that exceeds the sum of each individual contribution [direct cooperativity, (2)]. Weak protein-protein interactions may only become relevant when two proteins are brought into close proximity through their binding to adjacent DNA sequence motifs, which is referred to as 'DNA-mediated' interaction (16). Interestingly, high-throughput assays that monitor the binding of multiple TFs to a diverse library of sequences have shown that, with the exception of obligatory factor dimerization, the binding of two TFs to neighboring motifs typically leads to additive rather than synergistic effects. This

suggests that direct physical cooperation between TFs may not be a common mechanism (17).

In addition to the direct cooperativity of interacting TFs, indirect cooperativity can also be observed in a chromatin context. Nucleosome assembly maximizes the nucleosome organization of genomic DNA (18). TFs often cannot bind to nucleosomal DNA and, therefore, compete with nucleosomes to access their binding sites. At clustered *cis* elements, the successful binding of one TF can hinder nucleosome formation thus facilitating the interaction of other TFs to neighboring DNA. This nucleosome-mediated cooperativity or 'assisted loading' of TFs, does not necessarily involve protein interactions and can broaden the combinatorial space (19,20). Nucleosome-mediated cooperativity appears to be the prevalent form of TF cooperativity (21). Some TFs may be better competitors against nucleosomes due to their abundance or mode of DNA interaction, while others may benefit from the action of good competitors (22).

The cooperativity between MSL2 and CLAMP might involve direct protein interactions (8,23). Indeed, we now find that disrupting the physical contact between MSL2 and CLAMP completely abolishes their cooperativity. Surprisingly, rather than the expected nucleosome-mediated cooperativity between non-interacting factors, we find that the stronger GA binder, CLAMP, inhibits the association of MSL2 at shared sites. This suggests that the physical contact between MSL2 and CLAMP not only increases the affinity of MSL2 for target sites (23), but also ensures a defined geometry or registry between the proteins, which not only underlies their cooperation, but also prevents competition. We find that the binding strength of MSL2/CLAMP depends on cooperativity at individual composite binding sites and not on clustering or arrangement of such motifs in the genome.

Finally, we explore to which extent the observed TF cooperativity explains the X chromosome-specific binding of MSL2 *in vivo*. Taking advantage of the superior signal-to-noise ratio and resolution of the CUT&RUN (C&R) technique, we generated the first C&R profiles for MSL2. The new profiles reveal more than twice as many MSL2 binding sites than corresponding chromatin immunoprecipitation (ChIP) profiles, which are, remarkably, all on the X chromosome. The improved profiling method not only visualizes HAS, but also degenerate sites of lower affinity that are still restricted to the X (24). Interestingly, we observed only a minimal overlap between the *in vitro* and *in vivo* binding sites. This suggests that the extensive cooperativity between MSL2 and CLAMP observed *in vitro* on all chromosomes is restricted to a relatively small number of functional sites *in vivo*. We discuss a potential role for non-coding *roX* RNA in this process.

MATERIALS AND METHODS

Co-immunoprecipitation

The protocol was adapted from (8). Briefly, we mixed 50 nM of recombinant MSL2-FLAG and CLAMP-FLAG or CLAMP- Δ ZF1-FLAG in 100 μ l DREX or EX50 buffer (10 mM HEPES, 50 mM KCl, 1.5 mM MgCl₂, 10% (v/v) glycerol, 50 μ M ZnCl₂, 1 mM DTT and 1 \times cComplete

EDTA-free Proteinase inhibitor cocktail), and incubated the mixture with end-over-end rotation for 30 min at room temperature (RT). Next, we added the samples to 15 μ l of antibody-coupled magnetic ProteinG Dynabeads, which had been pre-coupled with the respective antibody. For CLAMP immunoprecipitations only 5 μ l were used. For pre-coupling, beads were washed thrice with 1 ml EX50 and incubated with end-over-end rotation for over-night at 4°C with 1 μ l rabbit anti-MSL2 antibody per 10 μ l beads and with 2 μ l rabbit anti-CLAMP antibody (SDIX). The beads were then washed thrice with 1 ml of EX50 buffer, followed by incubation with the IP reaction for 15 min at RT. After incubation, the beads were washed thrice with 1 ml of EX50 buffer, resuspended in SDS-sample buffer, and processed for western blotting.

Western blots

Samples were denatured with Laemmli buffer at 95°C for 5 min. Then samples were electrophoresed on 4–20% SDS ServaGel TGPrime (Serva) for 1 h at 100 V. Proteins were transferred to Amersham™ Protran™ 0.45 μ M Nitrocellulose Blotting Membrane for 1 h at 400 mA. All following incubations were for 1 h at RT. Membranes were blocked with 5% BSA. The membrane was incubated with primary antibody in TBS-T, washed thrice with TBS-T and incubated with secondary antibody. Images were taken using the LICOR Odyssey CLx.

DNA purification

Genomic DNA (gDNA) was obtained from male BG3-c2 cells (Drosophila Genomics Resource Center) (25). They were cultured at 26°C in Schneider's Drosophila Medium (GIBCO) with 10% fetal calf serum (FCS), Penicillin–Streptomycin and 10 mg/ml human insulin. The DNA of 10⁷ cells was purified using the Blood & Cell Culture DNA Midi Kit (Qiagen) following the supplier's protocol. The DNA was dissolved in EDTA-free 10 mM Tris–NaCl pH 8. Concentrations were determined using Qubit (Thermo Fisher).

Preparation of preblastoderm embryo chromatin assembly extract (DREX)

Preblastoderm embryos were collected within 90 min after egg laying, as previously described (12). Briefly, settled embryos (50 ml) were dechorionated in 200 ml of embryo wash buffer (EW) (0.7% NaCl and 0.04% Triton X-100), followed by treatment with 13% sodium hypochlorite (VWR) for 3 min at RT with constant stirring. After rinsing the embryos for 5 min with cold water, they were transferred to a glass cylinder containing EW and allowed to settle. The settled embryos were then washed successively with 0.7% NaCl and extract buffer, consisting of 10 mM HEPES/KOH pH 7.6, 10 mM KCl, 1.5 mM MgCl₂, 10% glycerol, 1 mM DTT and 1× cComplete EDTA-free Proteinase inhibitor cocktail (PIC, Roche).

The embryos were homogenized using a homogenplus homogenizer (Schuett-Biotec) with one stroke at 3000 rpm

and 10 strokes at 1500 rpm after which the MgCl₂ concentration of the homogenate was adjusted to 5 mM (final concentration) and centrifuged at 27 000 g for 15 min at 4°C. The lipid layer was discarded, and the supernatant was further centrifuged at 245 000 g for 2 h at 4°C. The clear extract was collected using a syringe, while the lipid layer and pellet were left behind. To avoid chelation of Zn, EDTA was excluded from all steps of the purification. Extracts were stored in 200 μ l aliquots at –80°C after shock freezing in liquid nitrogen. Extracts were only thawed once before use.

Chromatin assembly

One μ g of genomic DNA was assembled into chromatin by adding 15 μ l 10× McNAP buffer (0.3 M creatine phosphate, 30 mM ATP, 3 mM MgCl₂, 1 mM DTT, 10 ng/ μ l creatine phosphokinase), 100 μ l DREX extract and EX50 buffer (10 mM HEPES/KOH pH 7.6, 50 mM KCl, 1.5 mM MgCl₂, 50 μ M ZnCl₂, 10% glycerol, 1 mM DTT, 1× PIC) up to a final volume of 150 μ l. The amounts of extract necessary were determined empirically for each batch. Assembly took place at 26°C for 4 h.

MNase digestions

One μ g of DNA assembled into chromatin in 150 μ l was digested with MNase by adding 200 μ l MNase digestion solution (186 μ l EX50, 10 μ l 1 M CaCl₂ and 4 μ l MNase solution 333 U/ μ l). At times 15, 30 and 120 sec, 110 μ l were transferred to tubes containing to 40 μ l 100 mM EDTA solution each to stop the digest. Two μ l glycogen (10 mg/ml) and 150 μ l 7.5 M ammonium acetate were added and samples were mixed. Then, 880 μ l of 100% ethanol was added and samples were vortexed vigorously and cooled at –20°C for 10 min. After centrifugation at 21 000 g for 15 min at 4°C, the supernatant was removed and pellets were washed with ice-cold 70% ethanol. After pelleting the DNA again at 21 000 g for 5 min at 4°C, it was dissolved in 8 μ l 10 mM TE buffer and 2 μ l Orange-G loading dye. Samples were separated on a 2% agarose gel, stained with ethidium bromide and imaged using the Quantum ST-4 from PeqLab to control for proper chromatin assembly. Samples used for later ChIP-seq experiments were always digested for 2 min.

Baculovirus infections

Sf21 cell cultures at 10⁶ cells/ml (2.5 × 10⁸ cells) were infected 1:1000 (v/v) with baculovirus, expressing the respective FLAG tagged proteins as described in (26). After 72 h, cells were harvested and washed once in PBS, frozen in liquid nitrogen and stored at –80°C.

Protein purification

EDTA was excluded from all steps of the purification to avoid chelation of Zn in later experiments.

MSL2-FLAG: Sf21 cell pellets were rapidly thawed and resuspended in ice-cold Lysis buffer (300 mM KCl, 50 mM HEPES/KOH pH 7.6, 5% glycerol, 0.05% NP-40, 1 mM MgCl₂, 50 μ M ZnCl₂, 1 mM DTT, 1× PIC). 25 ml buffer was added to the cell pellet (2.5 × 10⁸ cells). After 15 min

incubation on ice, the suspension was sonicated (5×10 sec pulses, 20 s break, 20% amplitude, Branson sonifier model 250-D) and centrifuged for 45 min at 30 000 g at 4°C. The soluble protein fraction was incubated with Lysis buffer-equilibrated FLAG beads (Anti-FLAG M2 Agarose, Sigma) for 3 h at 4°C on a rotating wheel. 0.5 ml beads were used per 2.5×10^8 cells. The beads were washed twice with 10 ml ice-cold Lysis buffer, twice with 10 ml Wash buffer (Lysis buffer, but with 1 M KCl and 1% NP-40) and twice with 10 ml Elution buffer (Lysis buffer, but with 100 mM KCl). The FLAG-tagged MSL proteins were eluted for 3 h at 4°C on a rotating wheel in the presence of 0.5 mg/ml FLAG-Peptide (Sigma) in 1 ml Elution buffer. Purified proteins were then rapidly frozen in liquid nitrogen and finally stored at -80°C. Protein concentrations were determined via SDS-PAGE and Coomassie staining using BSA (New England Biolabs) as a standard. Cloning for the MSL2 expression construct is described by (26).

CLAMP-FLAG and CLAMP-ΔZF1-FLAG: Sf21 cell pellets were rapidly thawed and resuspended in 1 ml Buffer C per 10 mL of culture (50 mM HEPES pH7.6, 1 M KCl, 1 mM MgCl₂, 5% (v/v) glycerol, 0.05% NP-40, 50 μM ZnCl₂, 375 mM L-arginine (according to (27)) supplemented with 0.5 mM TCEP and 1× PIC. After 15 min incubation on ice, the suspension was sonicated (5×10 s pulses, 20 s break, 20% amplitude, Branson sonifier model 250-D). The extract was adjusted with Buffer C containing 1× PIC to 2 ml per 10 ml of culture and supplemented with 0.1% (v/v) polyethyleneimine by adding 2% (v/v) polyethyleneimine (neutralized with HCl to pH 7.0) drop-by-drop while string in an ice bath [according to (28)] and then centrifuged for 45 min at 30 000 g at 4°C. The soluble protein fraction was incubated with Buffer C-equilibrated FLAG beads (Anti-FLAG M2 Agarose, Sigma) for 3 h at 4°C on a rotating wheel. 0.5 ml beads were used per 2.5×10^8 cells. Beads were pelleted at 4°C for 5 min at 500 g and supernatant was removed. Beads were washed 5 times with 20 bed volumes of Buffer C. The FLAG-tagged CLAMP proteins were eluted for 3 h at 4°C on a rotating wheel in the presence of 0.5 mg/ml FLAG-Peptide (Sigma) in 1 ml Buffer C containing 1× PIC. Purified proteins were then rapidly frozen in liquid nitrogen and finally stored at -80°C. Protein concentrations were determined via SDS-PAGE and Coomassie staining using BSA (New England Biolabs) as a standard. Cloning for the CLAMP constructs is described by (8).

Antibodies

The following antibodies were used: rabbit α-MSL2 (26); guinea-pig α-MSL2 (8); rabbit α-CLAMP (8); IRDye 680RD goat α-rabbit (Licor 925-68071); IRDye 800CW goat α-guinea-pig (Licor 926-32211).

ChIP-seq

Recombinant proteins were added to 1 μg of assembled chromatin as described in 'chromatin assembly' and were allowed to bind for 1 h. Samples were crosslinked by adding formaldehyde (Thermo Fisher Scientific, Ref 28908) up to 0.1% final concentration for 10 min and then quenched by 125 mM glycine for 5 min. Samples were digested by MNase

as described above for 2 min. After adding 1× RIPA buffer up to 500 μl, samples were precleared on a rotating wheel with 20 μl buffer-equilibrated protein-AG beads per 1 μg chromatin for 1 h at 4°C. Supernatant was collected and 2 μl of purified antibody was added to the precleared sample and let to bind overnight at 4°C on a rotating wheel. Then samples were bound to freshly washed protein AG beads for 3 h. Samples were washed 4 times for 5 min with 1 ml of 1× RIPA buffer per sample (1 μg chromatin on 20 μl beads). Then the beads were suspended in 100 μl 1× TE buffer and de-crosslinked overnight at 65°C while shaking. Samples were then digested with 10 μg RNase A for 30 min at 37°C and 100 μg proteinase K at 56°C for 1h. beads were pelleted at 1000 g for 1 min and supernatant was transferred to a fresh tube for purification. DNA was purified by two extractions with phenol:chloroform:isoamyl-alcohol (25:24:1, Sigma Aldrich) and precipitated by adding it to 2 μl of glycogen, 0.1× volume 3 M sodium acetate. Then 2.5× volume 100% ethanol was added and samples cooled at -20°C for 15 min and pelleted in a tabletop centrifuge. The DNA was washed once with 70% ethanol and dissolved in EDTA-free 10 mM Tris/NaCl, pH 8. Concentrations were determined using Qubit (Thermo Fisher).

For correlation analyses between replicates, see Supplementary Figure S2.

Cut&run

CUT&RUN was performed as described in (29). Briefly, 10⁶ S2 cells were harvested, washed and bound to pre-activated Concanavalin A-coated magnetic beads, in wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine and 1× cComplete EDTA-free PIC). The cells were then permeabilized in Antibody buffer containing 0.05% digitonin (Dig-Wash) and 2 mM EDTA and incubated with rabbit α-MSL2 serum or pre-immune serum, diluted 1/500, at 4°C overnight on a nutator. After antibody incubation, the beads were washed thrice with Dig-Wash buffer and incubated with pAG/MNase at a concentration of 700 ng/ml, for 1 h at 4°C. After 2 washes in Dig-Wash buffer the beads were resuspended in 100 μl of Dig-Wash buffer and chromatin was digested, by addition of 100 mM CaCl₂, for 30 min at 0°C. The digestion reaction was stopped by addition of 100 μl 2× STOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.05% Digitonin, 50 μg/ml glycogen, 100 μg/ml RNase A). The digested chromatin fragments were released by incubating the beads for 30 min at 37°C. The beads were then placed on a magnet stand and the liquid containing the digested chromatin was transferred to a fresh tube. Total DNA was extracted using the phenol/chloroform protocol. Concentrations were determined using Qubit (Thermo Fisher).

For correlation analyses between replicates, see Supplementary Figure S2.

Library preparation and sequencing

For ChIP-seq samples, libraries were prepared using NEB-Next Ultra II DNA Library (New England Biolabs) according to manufacturer's instructions. For CUT&RUN, sequencing libraries were prepared as described in (30,31). All

libraries were sequenced on an Illumina NextSeq1000 sequencer. About 20 million paired-end reads were sequenced per sample for each of the ChIP replicates. Replicates were performed using a separate batch of purified proteins and DREX extracts. For CUT&RUN about 5 million paired-end reads were sequenced per sample for each of the biological replicates. Base calling was performed by Illumina's RTA software, version 1.18.66.3.

DATA ANALYSIS

Read processing

Sequence reads were Demultiplexed by JE demultiplexer (32) using the barcodes from the Illumina Index read files. Demultiplexed files were aligned to the *Drosophila melanogaster* release 6 reference genome (BDGP6) using Bowtie2 (33) version 2.2.9. (parameter '-end-to-end -very-sensitive -no-unal -no-mixed -no-discordant -X 400') and filtered for quality using samtools 1.6 (34) with a MAPQ score cutoff of -q 2.

Replicate correlation

Replicate correlation was determined by first searching the dm6 genome for 5000 best hits of the GAGAGA MRE-motif by FIMO (MSL2-ChIP $n = 3$, CLAMP-ChIP $n = 2$). Then each replicate was down-sampled to receive the same number of reads per replicate, and reads per motif were counted and plotted against each other. If replicates were sufficiently similar, the sampled reads were merged and used for further analysis. This avoids normalization against an input and retains individual read information. For correlation analyses between replicates, see Supplementary Figure S2.

Peak calling

Peaks were called using Homer (35) version 4.9.1 calling the functions makeTagDirectory (parameters -single -fragLength 150) and findPeaks (parameters -style factor -size 150 -F 6) using the corresponding negative samples in which the IP was done without adding the respective protein as control. We called peaks against a negative control (IP in the absence of added protein) where possible, as this allows us to account for antibody cross reactivity in immunoprecipitations. Peak calling was done with the summarized replicates for each sample and the control, resulting in more robust peaks through the additional coverage used. HAS and PionX regions and motifs were used as defined by (5) with 309 HAS and 56 PionX in total.

De novo motif discovery

Enriched motifs in peak region were discovered using MEME (36) (version 4.11.4 (parameters -mod anr -dna -revcomp -nmotifs 2)). Before analysis the peaks were resized to 200 bp to include 25 bp of sequences directly bordering the peaks.

Find motif occurrences

Motif search using position weight matrixes from MEME in peak regions on the genome was performed with FIMO (37) version 5.0.2.

Browser profiles

Browser profiles were generated using UCSCutils (<http://genome.ucsc.edu>.) version 3.4.1. calling the function makeUCSCfile using the summarized sample replicates Tag Directories, also used for the peak calling, and were normalized against the control. Values are fold change over control. Profiles were visualized using the IGV software (38).

Data analysis and plotting

Data Analysis was conducted in R (39), using the tidyverse libraries (40).

Heatmaps and cumulative plots

Heatmaps were made using the R library 'Complex-heatmaps' (41) by cutting windows of 2000 bp around sites of interest of the calculated coverages normalized against a control if applicable and aligning them. The cumulative plots are made by calculating the mean of each column. Window identities are retained in the data and used for the annotation by overlapping them with the known HAS or the X chromosome.

Fragment size analysis

Plots depicting individual fragments at each peak and annotation of motifs within peaks was done by custom code using the libraries zoo (42), rtracklayer (43) and ggpmisc.

Alphafold

Protein structures were predicted using ColabFold (44), which uses the AlphaFold2 algorithm (45).

RESULTS

CLAMP-MSL2 cooperativity is dependent on CLAMP's N-terminal zinc finger

The interaction surfaces between MSL2 and CLAMP were previously mapped (8,23). CLAMP binds MSL2 with its N-terminal zinc finger (ZF1, Figure 1A). For MSL2 the interactions required an evolutionary conserved sequence just downstream of the CXC domain, which we termed the CLAMP binding domain (CBD, Figure 1A). Tikhonova *et al.* concluded from NMR studies of isolated domains that the CBD was largely disordered (46). Because unstructured regions are usually not conserved, we used the AlphaFold2 protein structure database to predict the CBD structure. The query of an MSL2 fragment between amino acids 510–690 including both CXC and CBD domains revealed a high-confidence structural prediction for both domains connected by an unstructured linker (Figure 1B).

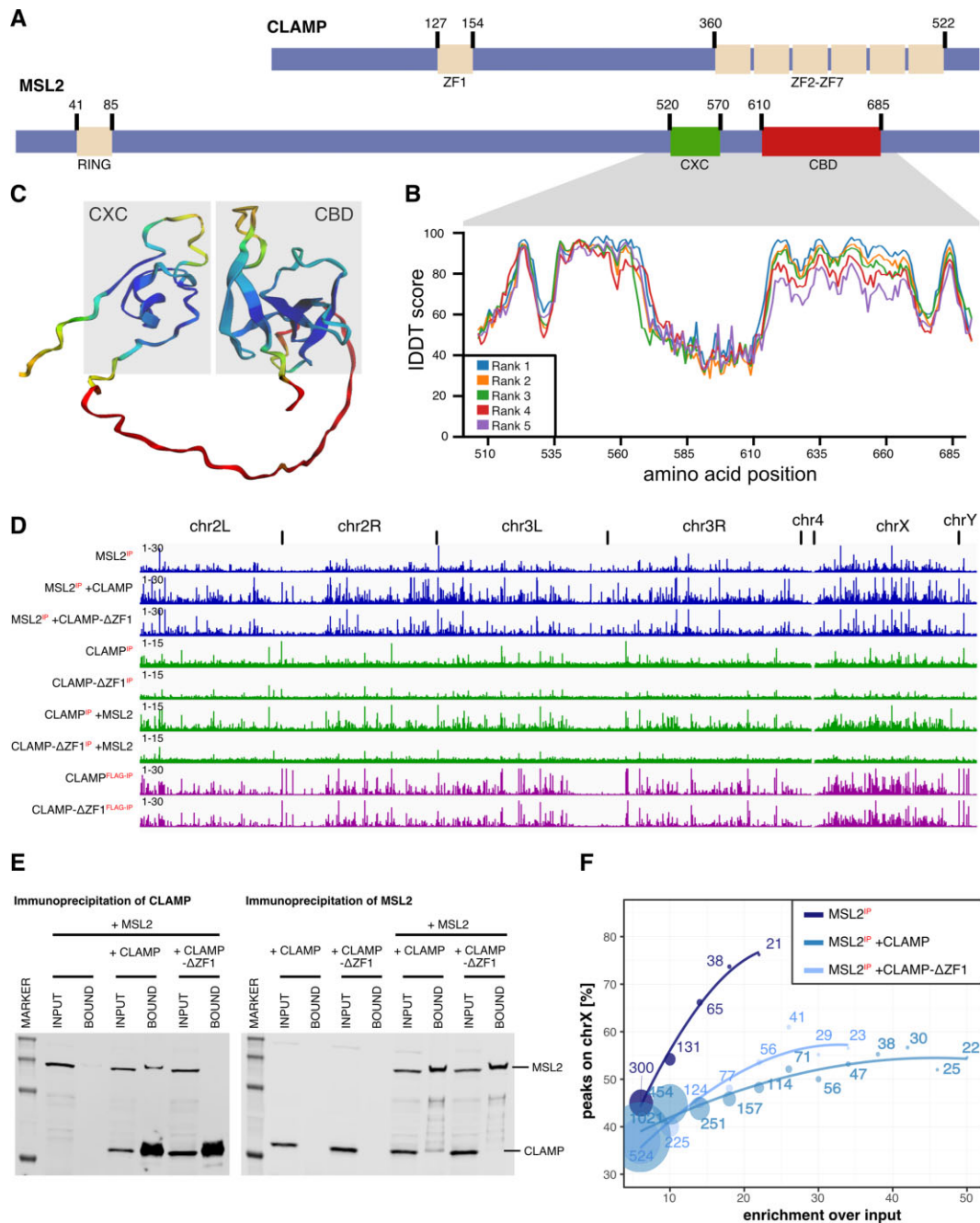


Figure 1. Deletion of CLAMP N-terminal zinc finger (ZF1) reduces binding cooperativity with MSL2. (A) The domain organization for CLAMP and MSL2. CLAMP contains seven zinc finger (ZF) domains, of which ZF1 interacts with MSL2 and a subset of ZF2-ZF7 is involved in MRE recognition. The known functional domains of MSL2 RING, CXC and CLAMP binding domain (CBD) are indicated. Amino acid positions within the proteins are labeled. (B) Relative IDDT prediction scores from AlphaFold2 for MSL2 calculated for amino acids 510–690. High scores indicate high prediction confidence. (C) Predicted protein structure of MSL2 amino acids 510–690, generated by AlphaFold2. The IDDT scores from Figure 1B are color-coded (high = blue, low = red). The CXC and CBD domains are indicated by shading. (D) Genome-wide chromatin binding profiles. Summarized genome browser profile for all biological replicates shows binding of the protein marked with 'IP' determined by ChIP-seq along the *Drosophila* genome (MSL2-ChIP $n = 3$, CLAMP-ChIP $n = 2$). The chromosomes are indicated on the top of the panels. Maxi-mum values for each window are shown after normalization against the control. For exemplary close-up views, see Supplementary figures S1C and S1D. (E) Western blot analysis of co-immunoprecipitation of MSL2 and CLAMP. 50 nM of respective proteins were added to DREX and immunoprecipitated with anti-MSL2 or anti-CLAMP antibody as annotated. The input (10%) before IP and the bound fraction (30%) were resolved. Blots were probed with anti-FLAG antibody to avoid bias and imaged using a Licor Odyssey DLx. The amount of co-immunoprecipitated protein was quantified, and the bound fraction was normalized against input amounts (see also Supplementary Figure S1A). (F) X-chromosomal enrichment of MSL2 peaks in absence and presence of CLAMP or CLAMP- Δ ZF1. Enrichment is related to the respective peak cut-off, defined as the enrichment over background of the peaks. Enrichment is only calculated for datasets with at least 20 remaining peaks.

The prediction further showed that CXC and CBD contain surfaces that are compatible with mutual interactions (Figure 1C).

Attempting to disrupt the interaction between CLAMP and MSL2 we generated protein variants with respective ZF1 and CBD deletions. We found that deleting the CBD yielded a protein that lost most DNA binding activity (not shown), in agreement with earlier findings that partial deletion of the CBD compromises DNA binding (6). In light of the AlphaFold2 prediction it is tempting to speculate that CBD and CXC domains are part of a composite DNA reader module. In order to assay cooperativity, it is of importance to assure that the DNA binding of the mutant and wild type TFs alone is similar. This requirement was not satisfied in the MSL2 CBD deletion mutant.

In contrast, upon deletion of ZF1 CLAMP retained DNA binding activity (see below). When we compared the binding of the wildtype and mutant CLAMP by ChIP, we found the CLAMP- Δ ZF1 profile to be systematically weaker (Figure 1D). We speculated that this may be because the epitope, against which the CLAMP antibody was raised, lies close to the ZF1 deletion site. We therefore generated ChIP-seq profiles for CLAMP and CLAMP- Δ ZF1 using an antibody detecting the C-terminal FLAG-tag shared by both proteins. The α FLAG ChIP yielded profiles of roughly similar intensities, indicating that deletion of ZF1 generally does not impair the ability of CLAMP to bind chromatin as seen when comparing the binding to known *in vitro* CLAMP site (S1B) and shown in coverage profiles (Figure 1D, Supplementary Figure S1B–D).

To test whether the deletion of ZF1 impairs the interaction with MSL2 in the complex chromatin reconstitution extract, we performed reciprocal co-immunoprecipitation (IP) assays. MSL2 was incubated with either the wildtype or mutant CLAMP in a chromatin assembly reaction lacking DNA. To control for non-specific adsorption of protein to antibody-coated paramagnetic beads, mock-IPs were performed, in which one protein was omitted. In the presence of DREX, this nonspecific adsorption was minimal (Supplementary Figure S1A and data not shown). The Western blots were probed with α FLAG antibody that visualize MSL2 and CLAMP simultaneously. We found that a significant co-IP of MSL2 with CLAMP (20%) or CLAMP with MSL2 (7%) was reduced to background levels (<1%) upon deletion of CLAMP ZF1 (Figure 1E).

We then tested the cooperativity of DNA binding of CLAMP and MSL2 in reconstituted chromatin as before (11). Genomic DNA was assembled into chromatin, the TFs were added and crosslinked (Supplementary Figure S1E). The chromatin was digested with MNase before ChIP and paired-end sequencing of the associated DNA. Representative genome browser coverages are shown in Figure 1D for a whole-genome view and Supplementary Figures S1C, D for regions of the X chromosome.

In the presence of CLAMP, the binding of MSL2 to chromatin was generally increased, in line with our earlier observations (11). The fact that MSL2 is recruited to strong CLAMP binding sites on autosomes is illustrated in Figure 1F, which relates the strength of binding events ('signal over input') to the enrichment on the X. Of the 21 strongest MSL2 peaks in the absence of CLAMP (25-fold over in-

put) more than 75% are located on the X. This illustrates the intrinsic ability of MSL2 to select strong X chromosome determinants. Sites of lower affinity are less enriched on the X (e.g. 131 binding events with 10-fold signal over input reside only 55% on the X). In the presence of CLAMP, MSL2 binds strongly to many more sites, but about half of the strongest sites with 30–50-fold enrichment over input are located on autosomes.

In presence of CLAMP- Δ ZF1 the number of MSL2 peaks is reduced and the X-specificity remains low (Figure 1F). This result suggests that ZF1 is required for CLAMP-MSL2 cooperativity and that this cooperativity is direct. Nevertheless, this cooperativity does not explain the exclusive MSL2 binding to the X chromosome *in vivo*, a notion that will be further developed, below.

CLAMP- Δ ZF1 competes against MSL2 for MRE binding

To further explore the effect of ZF1 deletion on cooperativity we picked 264 autosomal sites where CLAMP and CLAMP- Δ ZF1, when assayed alone, bind strongly as judged by α FLAG IP (Figure 2A, purple tracks). Addition of MSL2 enhanced the binding of CLAMP to these sites, but not of the mutant CLAMP- Δ ZF1 (Figure 2A, green tracks). The binding of MSL2 alone was strongly improved by CLAMP. To our surprise, addition of CLAMP- Δ ZF1 had the opposite effect: MSL2 binding was reduced (Figure 2A, blue tracks). We had expected that MSL2 would either not be affected by the interaction-defective CLAMP or perhaps even benefit somewhat from indirect, nucleosome-mediated cooperativity. On the contrary, the data suggest that CLAMP- Δ ZF1 competes for MSL2 binding. The phenomenon of cooperativity by CLAMP and competition by CLAMP- Δ ZF1 was also seen at 309 HAS (5) (Figure 2B).

To determine if there was a sequence dependency to the effects seen in the heatmaps we called peaks for MSL2 and divided them into groups according to whether FIMO annotates an MRE, PionX sequence (Figure 2D) or no clear motif. For quantitative assessment we determined the normalized read counts at each peak (Figure 2C). This shows that MSL2 alone binds best at sites with a PionX motif and worst at sites without a clear motif, as expected. Addition of CLAMP increases binding to all sites. CLAMP- Δ ZF1 competes with MSL2 only at sites with a clear composite motif. This competition may be due to the fact that MSL2 and CLAMP both recognize GA-dinucleotides in MREs (7,9). At PionX sites, MSL2 may bind the 5' extension with its CXC domain and more 3' GA-dinucleotides with parts of the CBD (5). CLAMP- Δ ZF1 may compete with this GA-binding. So, why does wild-type CLAMP not compete for MSL2 binding? We propose that the direct physical interaction between the two proteins establishes a defined distance between the cooperating proteins, ensuring that they bind to composite sites together and side by side, rather than alone and in competition (Figure 2E).

Exploring MSL2-CLAMP cooperativity beyond single MREs

The compendium of X-chromosomal binding sites for MSL2 has been redefined several times in recent years. We

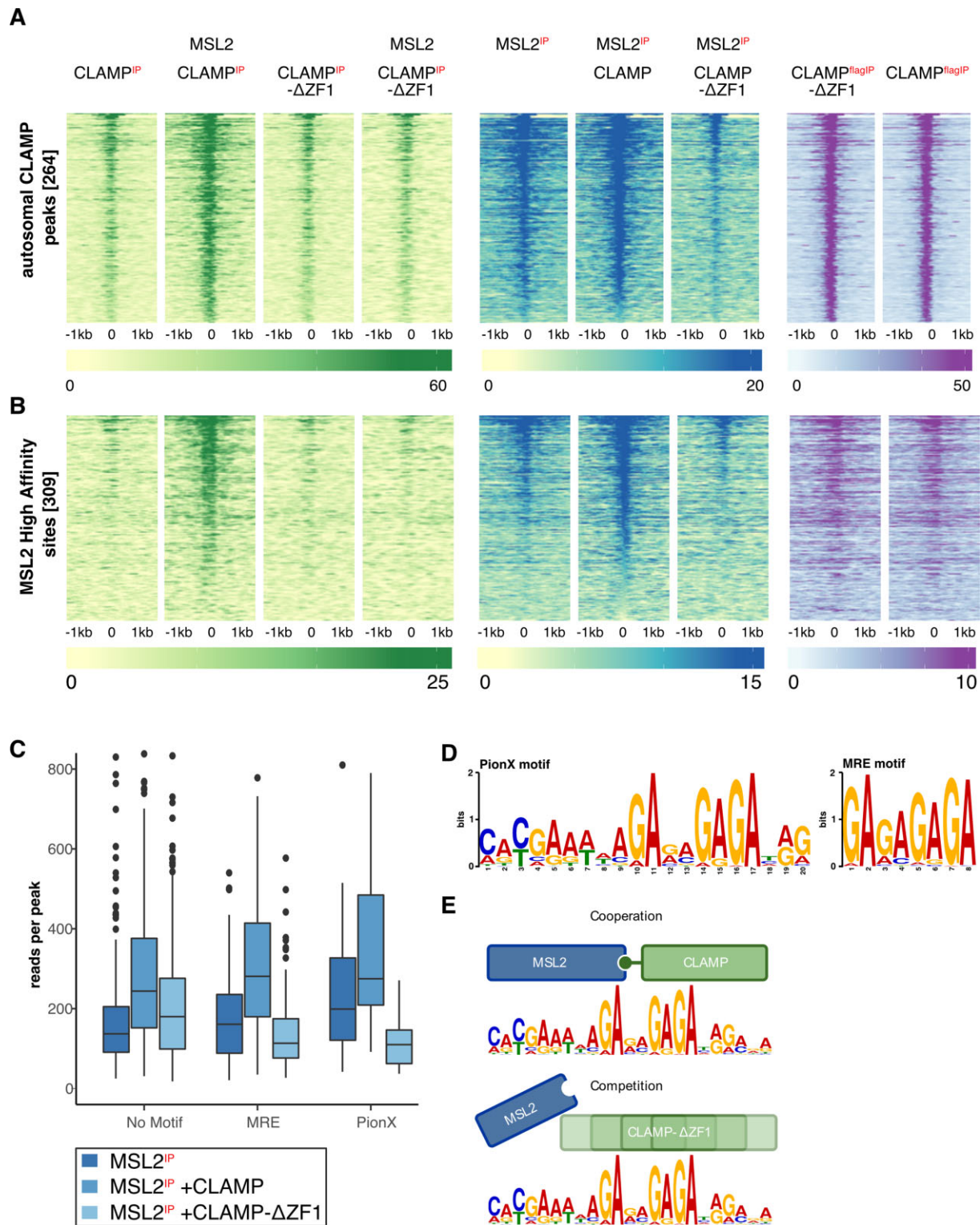


Figure 2. In absence of ZF1, CLAMP competes with MSL2 at composite binding sites. (A) The indicated proteins were allowed to bind to chromatin-assembled genomic DNA and MSL2 or CLAMP binding determined by ChIP-seq (see Supplementary Figure S1C). Enrichment of the targeted factor marked with ‘IP’ was plotted to 264 autosomal sites that were bound by both CLAMP and CLAMP-ΔZF1. Heatmaps of individual regions are shown. Coverage windows of 2000 bp around the sites were cut out and aligned. (B) As in A, but plotted at the 309 MSL2 High-Affinity sites (5). (C) Boxplot of normalized read counts per peak. Peaks were determined by Homer for each sample and then separated into groups depending on whether they harbor no strong motif, an MRE motif or a PionX motif, respectively. (D) Position Weight Matrix of the ‘MSL recognition element’ (MRE) and the ‘Pioneering on the X’ (PionX) element used to determine motif-containing sites. (E) Model explaining the binding behavior of MSL2 and CLAMP. Intact proteins interact cooperatively and lock into a defined geometry on the large, composite binding motif. In absence of the ZF1 interaction domain CLAMP binds ‘out of phase’ to the motif, leading to competition and displacement of MSL2 from the binding site.

use a list of 309 HAS defined as overlapping peak regions in two high-quality MSL2 profiles (5). HAS loci often contain multiple MRE and PionX motifs. MSL2 binding affinity may therefore correlate with the number of MREs in a given chromosomal region or with a defined spatial architecture at HAS. This ‘higher order’ level of MSL2 binding as well as MSL2-CLAMP cooperativity has not been systematically investigated so far.

We first explored this question computationally by applying the ‘enhanced chromatin occupancy’ (EChO) strategy (47) to binding data from DREX-assembled genomes. Although this method was developed for CUT&RUN, it can also be applied to ChIP data, which involve MNase digestion and paired-end sequencing. The method exploits the observation that subnucleosome-sized fragments arise from MNase cleavage on either side of a DNA-bound transcription factor. The shortest fragments are centered around a bound TF. When the fragment lengths are plotted across a region of interest, the minima will point to TF contacts on DNA. In absence of binding factors at the sites the profile consists of few reads with lengths of about 150 bp on average, which reflects the nucleosomal background (not shown).

We randomly selected 10 from the 264 autosomal sites where MSL2 and CLAMP bind cooperatively *in vitro* and visualized an area of 401 bp centered on the peaks (Figure 3A). The vertical green annotation indicates the positions of the MREs and PionX motifs, revealing a great diversity in the number and arrangement of such sequences. We then plotted the fragment sizes obtained from experiments, in which either CLAMP was assayed alone or MSL2 was mapped in the presence of CLAMP (Figure 3A). The number of reads (individual dots) provides a measure of binding intensity. The minima of the fitted curve points to sites where proteins bind DNA. Several conclusions can be drawn from the plots. (1) The predicted TF binding sites generally match the annotated MREs and PionX sequences. (2) Strong binding events often coincide with MRE clustering, but there are also cases where binding occurs in the absence of annotated MRE motifs. (3) The curves for CLAMP alone and MSL2 in the presence of CLAMP generally, but not always, follow a similar path.

The data suggest that CLAMP and MSL2 do in fact bind together to the composite sequence elements. If this was the case, an increased MNase ‘footprint’ (indicated by increased fragment sizes) would be detectable at sites where more than one TF binds. We computed the fragment lengths across all peak areas in all *in vitro* binding experiments, determined minima at each site by loess regression and summarized all the reads found at these locations (Figure 3B). The ‘fragment size at minima’ gives a relative measure of the TF footprint. MSL2 binding alone correlated with an average fragment length of 120 bp. In the presence of CLAMP, the average length of fragments was increased by 10 bp. In contrast, addition of CLAMP- Δ ZF1 had no effect. These findings support the notion of cooperative binding of MSL2 and CLAMP at individual composite binding sites, which are largely defined by the presence of CLAMP.

The anecdotal fragment size profiles (Figure 3A) illustrate a great diversity of binding site architecture, but they

do not allow general conclusions about binding cooperativity between clustered MRE sequences. To address this question more systematically, we determined an ‘MRE score’ for each peak area (see Materials and Methods) by summing up all nucleotides contained in an 8 bp MRE PWM (note that PionX sequences conform with the general MRE definition). The higher the MRE score, the more MRE motifs cluster at a given site. We then related this MRE score to the corresponding binding strength, indicated by the normalized read coverage (Figure 3C). For the binding of MSL2 alone, the correlation between MRE score and binding strength was poor. This is explained by the fact that MSL2 alone binds best to single PionX sequences and not so much to extended GA-rich regions (5,11). In contrast, CLAMP alone is known to bind extended GA sequences (48), which is reflected by a good correlation between MRE score and binding strength. If MSL2 binding strength is assessed in the presence of CLAMP, the correlation to MRE score remains poor (although slightly better than for MSL2 alone). Interestingly, there are many sites with a single MRE (MRE score of 8) with very different binding strengths. Finally, if CLAMP binding is assessed in the presence of MSL2 the correlation to MRE score is reduced, presumably because MSL2 can stabilize CLAMP at otherwise low affinity binding sites (Figure 2B). We conclude that the cooperativity in chromatin binding between MSL2 and CLAMP occurs predominantly at the level of individual, composite binding sites and less so between TFs bound to clustered elements within a HAS.

CUT&RUN detects X chromosomal MSL2 binding sites beyond the known HAS

These findings, along with our previous *in vitro* genome-wide binding assays (8,11), reveal fundamental principles of cooperativity and competition at GA-rich MRE sequences on all chromosomes. However, *in vivo* the cooperativity between MSL2 and CLAMP is strictly limited to X chromosomal HAS (8). This discrepancy may be resolved *in vivo*, if the cooperativity at the level of MRE clustering was more important, due to chromosome organization (49).

To obtain binding profiles suitable for EChO analyses, we generated the first C&R profiles for MSL2 in the male S2 cell line. The C&R methodology avoids immunoprecipitation and thus yields a better signal-to-noise ratio and improved resolution (29). It also avoids formaldehyde crosslinking, a major source of inconsistency between protocols. If combined with paired-end sequencing, C&R is ideal for EChO plot analysis.

Our C&R analysis impressively documented the exclusive binding of MSL2 to the X chromosome *in vivo*, and in this respect was superior to previous ChIP-seq profiles [Figure 4A; (5)]. C&R yielded twice as many MSL2 binding sites (709) than corresponding ChIP profiles. Apparently, the improved profiling method not only visualizes HAS, but also degenerate sites of lower affinity that are nevertheless restricted to the X (24,50). As such sites have not been previously characterized, we divided the total set of peaks into thirds according to their binding strength and performed motif searches using MEME. The ‘top’ peaks, where MSL2 binds most strongly, contain sequences that resemble the

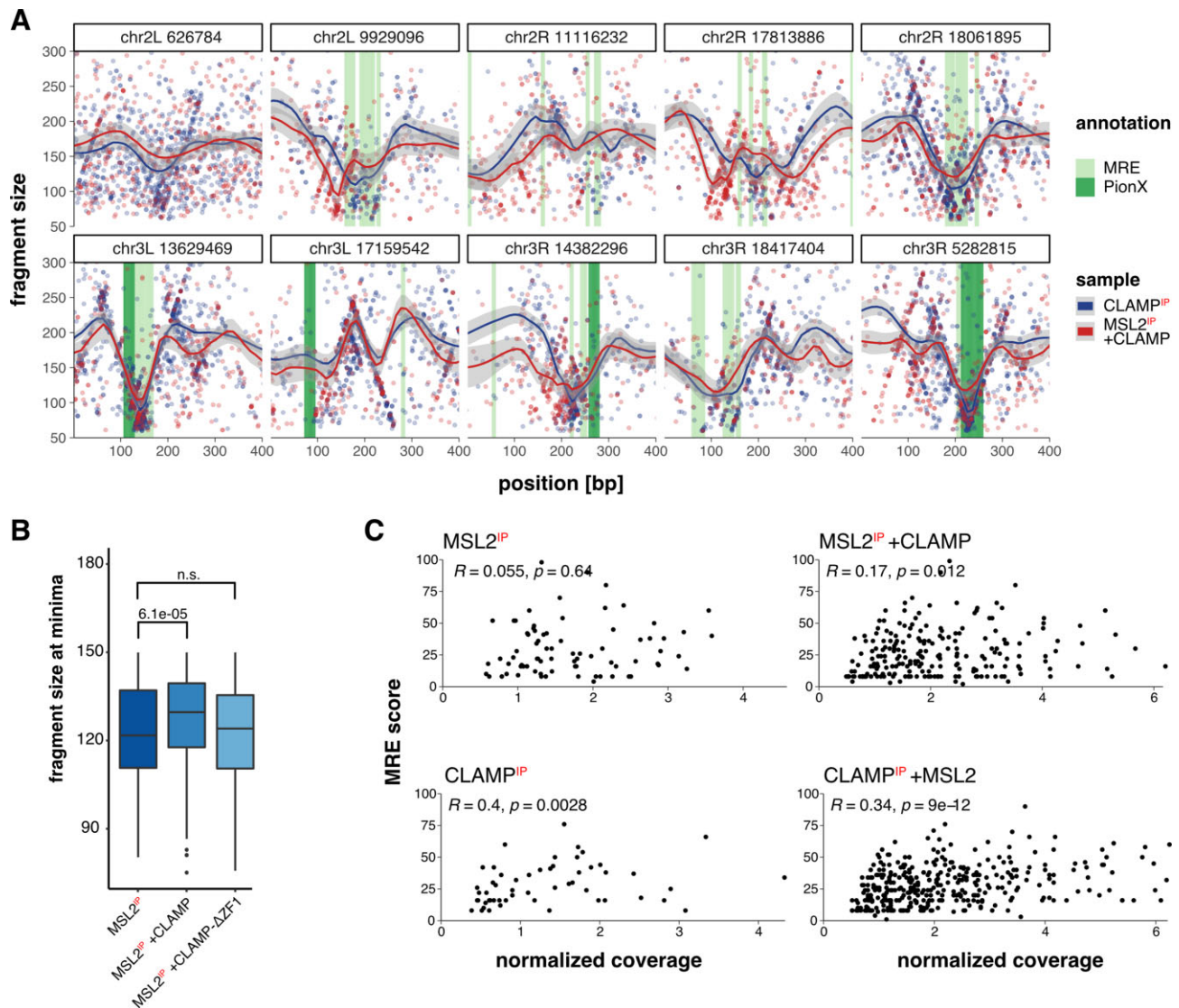


Figure 3. MSL2 binding sites are heterogeneous with respect to motif number, arrangement and MSL2/CLAMP interactions. (A) MSL2 and CLAMP binding to chromatinized genome was determined by ChIP and paired-end sequencing of fragments. Ten binding sites bound by both proteins were sampled from the sites shown in Figure 2A. The size of individual reads within the binding sites was plotted against their relative position. MRE and PionX sites were determined using the PWMs from Figure 2D and are annotated with green shading. A regression curve is fitted using loess and the shaded area represents the 5–95% confidence interval. As TFs protect a smaller area of DNA from cleavage than nucleosomes, the precise binding site of a TF can be found by determining the minima of fragment sizes within a given peak. (B) Using the regression curves the minima within each peak are determined and the size of all minima <150 bp are summarized. Boxplots are drawn and the correlation between groups measured using Wilcoxon signed-rank test. *P* values are annotated above the boxes. (C) The binding strength of the indicated protein was derived from normalizing the number of reads per peak to total read count and plotted against the MRE score. MRE score was determined as the number of GA dinucleotides that fall within MREs, as defined by a FIMO search using the PWM shown in Figure 2D.

ideal PionX motif (Figure 4B). Binding sites of intermediate affinity yield the familiar MRE motif in a longer GA-rich context. The ‘bottom’ group of weaker MSL2 interaction contains more degenerate motifs that are still GA-rich. Even these low-affinity motifs tend to avoid the minimal binding site of GAF, GAGAG. We previously showed that GAF can outcompete MSL2/CLAMP at sites containing GAGAG sequences, thereby excluding non-functional sites from MSL2 interaction (11). TF binding motifs are typically short, often less than 10 bp. The fact that MEME detects PWM signatures that are 20 bp and longer (Figure 4B) supports the notion that cooperation and competition be-

tween MSL2 and CLAMP occur at single, long composite binding sites.

Looking at the clustering of MRE motifs at MSL2 binding sites, we again observe a large diversity in the number and arrangement of MRE, similar to the *in vitro* sites. Figure 4C shows 10 randomly sampled sites for which the motif positions were determined by FIMO, as before. For EChO analysis, we plotted the lengths of individual reads within each binding site. The minima of the fitted curve indicate protein binding sites. Unlike the *in vitro* experiment, where we can control the input of TFs, we do not know which other proteins bind along with MSL2 at these sites. The

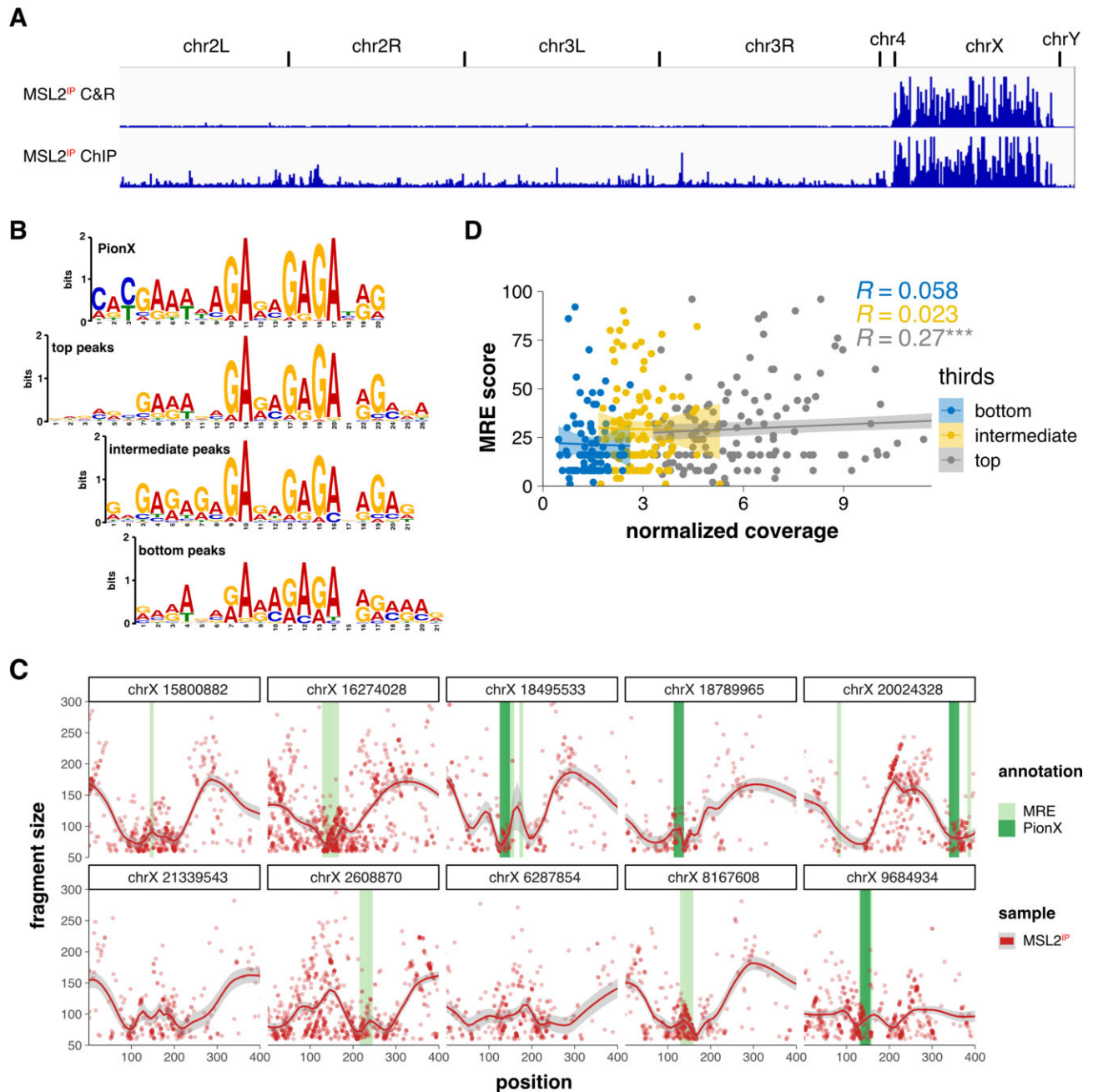


Figure 4. Mapping MSL2 binding sites *in vivo* by CUT&RUN (C&R) allows insights into motif and binding architecture. (A) Genome browser profile over the *Drosophila* genome summed up from three biological replicates of C&R profiles for MSL2 in S2 cells (‘C&R’) or ChIP-seq (‘ChIP’)(5). Maximum values for each window are shown and chromosomes are annotated. (B) Position Weight Matrix (PWM) of the PionX motif compared to the PWMs found by MSL2 C&R peaks as determined by MEME. The 709 peaks were split into thirds depending on their enrichment over input and analyzed separately. (C) Characterization of MSL2 binding sites with respect to motif number and arrangement and footprinting by MNase. The MSL2 binding sites were determined by C&R *in vivo* and peaks were called using Homer (34). For each peak the fragment length of individual reads was plotted against the position of the respective read center within the peak. Ten binding sites were sampled from the top third of binding sites and are used for illustration. MRE and PionX sites were determined using the PWMs from Figure 2D and are annotated. A regression curve is fitted using loess and the shaded area represents the 5–95% confidence interval. As TFs protect a smaller area of DNA from MNase cleavage than nucleosomes, the precise binding site of a TF can be found by determining the minima of fragment sizes within a given peak. (D) The binding strength of the indicated protein was derived from normalizing the number of reads per peak to total read count and plotted against the MRE score. MRE score was determined as the amount of GAs that fall within an MRE. MREs were found using a FIMO search using the PWM shown in Figure 2D. The 709 peaks found by MSL2 C&R were split into thirds depending on their enrichment over input and analyzed separately.

fact that we sometimes observe more than one minimum and that some minima are broad suggests that other factors, such as CLAMP, bind in the vicinity of MSL2. To approach the question of inter-MRE cooperativity more quantitatively, we calculated the MRE score within each site and related it to the binding strength, as before. For this evaluation of *in vivo* sites, the ‘top’, intermediate and ‘bottom’ thirds of the peak areas were analyzed separately (Figure 4D). Linear regression shows no correlation for the two lower thirds of MSL2 peaks *in vivo* (0.058 and 0.023 respectively) and a very weak correlation for the upper third of MSL2 peaks (0.27). This finding suggests that, similar to the *in vitro* analysis, the binding strength is mainly determined by the interaction of MSL2 with individual MRE motifs and not dependent on the number or arrangement of MRE-type sequences.

The intrinsic DNA binding cooperativity between MSL2 and CLAMP is largely overridden *in vivo*

So far, the *in vitro* and *in vivo* analyses have yielded MSL2 binding sites that appear similarly diverse with respect to number, MRE PWM and arrangement. We therefore expected the *in vitro* sites to overlap considerably with the expanded set of X chromosome sites obtained by C&R. To our surprise, the two sets of binding sites were very distinct. Heat maps sorted according to the strength of the 709 binding sites identified by C&R reveals that only a small fraction of these sites is actually bound *in vitro* (Figure 5). We conclude that the *in vivo* binding profile cannot be explained by intrinsic DNA binding cooperativity of MSL2 and CLAMP, but that this cooperativity is largely offset *in vivo* and only allowed in a context-specific manner.

DISCUSSION

Our experimental strategy involves the assembly of physiological chromatin on genomic DNA to provide a complex substrate for TF interaction assays. Since the assembly system is derived from *Drosophila* preblastoderm embryos, the chromatin that arises has all hallmarks of pre-MBT embryos: it does not contain significant concentrations of endogenous TFs or components of the transcription machinery. The addition of defined TFs somewhat mimics aspects of zygotic genome activation, when newly translated proteins initiate the complex genome expression program. The analysis of large compendia of potential TF binding sites in different genomic contexts allows statistical generalization, in addition to anecdotal illustration.

Our experimental system is well suited to discovering intrinsic protein-DNA interactions of well-characterized components. These may even include heterologous TFs of other species that bind short recognition motifs statistically distributed in *Drosophila* genomes (51). Characterizing the intrinsic binding properties of TFs in a complex chromatin environment is important, even if TF binding specificities commonly observed *in vitro* do not explain physiological chromosome interactions (52).

Our focus has been on the DNA binding proteins that target the ‘male-specific-lethal’ (MSL) dosage compensation complex (DCC) exclusively to X chromosome-specific

DNA elements to deploy the vital activation of transcription. *In vivo*, the MSL proteins do not bind autosomal DNA. We found that (i) MSL2, the DNA-binding subunit of the DCC, bears intrinsic sequence selectivity for a complex X chromosome-specific DNA signature, which we termed PionX sites (5); (ii) MSL2 cooperates with the abundant GA-binding protein CLAMP to associate with *bona fide* HAS on the X chromosome *in vivo* (8). *In vitro* this cooperation results in the recruitment of MSL2 to hundreds of strong CLAMP binding sites on all chromosomes (8,11); (iii) the GA-binding GAF competes with CLAMP/MSL2, preventing binding to a class of unphysiological sites (11). We now investigated the mechanism underlying the CLAMP-MSL2 cooperativity.

Cooperativity turned into competition

We found that the cooperativity between CLAMP and MSL2 depends on the physical interaction between the two proteins. Although a protein complex is formed in solution, it is possible or even likely that the interaction is promoted by DNA (16). We expected that upon disruption of the TF interaction the direct cooperativity would turn into indirect, nucleosome-mediated cooperativity, the dominant mechanism of synergism between TFs (21,22,53). To our surprise, we found that under those conditions, the stronger GA-binder, CLAMP, competes with MSL2 for binding to composite sites. These sites are unusually long: MEME detects PWMs of 20 bp and more in peaks, so there should be space for two or more proteins to bind side by side. This is in line with the comprehensive study of Taipale and colleagues, who showed that often cooperating TFs bind composite sequence elements where individual TF recognition motifs overlap (16). These findings argue that the physical interaction not just serves to provide additional surfaces to reduce the off-rate of TFs, but in this case assures that the two cooperating factors bind DNA in a defined arrangement that assures that the long, composite binding sites are fully used and to prevent dominance by the more avid GA-binder CLAMP, which outcompetes MSL2 if not physically connected to it.

Since CLAMP binds degenerate GA repeats (48), it is possible that two CLAMP molecules can bind to these long sequences, or that single CLAMP molecules bind with variable translational positions, thus occluding GA sequences required for MSL2 interaction (see model in Figure 2E). Accordingly, MSL2 can resist competition at sites that contain the PionX signature, but will be competed away from sites of lower affinity that contain degenerate GA repeats (Figure 4B). In such a situation, the physical interaction between CLAMP ZF1 and the CBD of MSL2 may ensure that both proteins lock into a defined geometry that prevents competition.

Relevance and limitation of intrinsic sequence specificity

MSL2 and CLAMP mutually cooperate to bind stably to X-chromosomal HAS *in vivo* (7,8). However, the chromosomal MSL2 binding profile only overlaps to a low extent with the pattern of MSL2-CLAMP binding *in vitro*. In some ways, this is not surprising given that MSL2 functions as a subunit

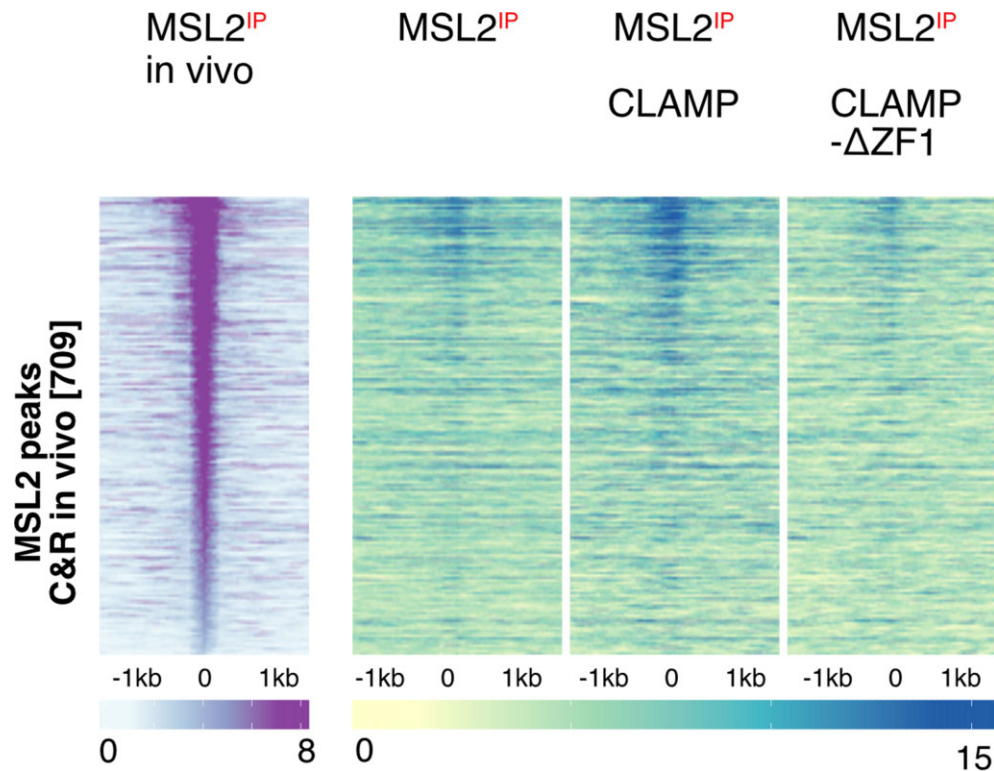


Figure 5. The intrinsic MSL2-CLAMP cooperativity found *in vitro*, is restricted *in vivo*. MSL2 binding was determined by C&R *in vivo* and by ChIP-seq *in vitro*. Enrichment was plotted to 709 MSL2 C&R sites and illustrated by heatmaps of individual regions. Coverage windows of 2000 bp around the sites were cut out, aligned and the mean for each column calculated. For details, see text.

of the large DCC. It is possible that MSL1-induced dimerization improves the selectivity of DNA recognition (54). The DCC also contains the long, non-coding *roX* RNA. Indeed, several studies concluded that the presence of *roX* RNA is required for faithful localization of MSL2 on the X chromosome (55,56). We found MSL2 delocalized to strong CLAMP binding sites on all chromosomes in the absence of *roX* RNA *in vivo* (6). Although the underlying mechanism is unknown, we note that *roX2* interacts with the C-terminus of MSL2, which also harbors the intrinsic GA-binding specificity of MSL2 as well as the CBD (57,58). We speculated that *roX* may modulate the DNA and protein interactions of MSL2, perhaps by restricting the cooperativity with CLAMP (6).

Remarkably, Larschan and colleagues recently observed that at nuclear division cycle 13, prior to the expression of *roX2* and the assembly of a mature DCC, MSL2 and CLAMP localize to CLAMP binding sites on all chromosomes (59). We hypothesized that the binding profile in reconstituted preblastoderm chromatin might reflect some of these earliest MSL2 interactions with chromosomes prior to *roX2* expression. However, a comparison of the profiles did not support this hypothesis (Supplementary Figure S3).

Zygotic gene activation is also accompanied by the appearance of the linker histone H1, which may modulate TF cooperativity (60,61). We explored this possibility in our earlier work (11), but found that genome-wide incorporation of H1 did not affect the cooperativity of CLAMP and MSL2 qualitatively, but led to a general dampening of all TF binding to chromatin.

Identification of an expanded set of X-chromosomal MSL2 binding sites

Taking advantage of the superior resolution and signal/noise output of C&R analyses we determined MSL2 binding sites in S2 cells. We obtained more than 700 distinct peaks of MSL2 interaction, all of which were localized on the X chromosome. Stratifying the peaks according to their signal strength reveals a hierarchy of sequences. Motifs with 5' extension, resembling the PionX motif ranged among the sites with highest affinity, followed by sites of intermediate affinity that display the MRE consensus sequence. Weaker binding sites contain degenerate MRE motifs dominated by GA repeats. These data support earlier conclusions from low-resolution studies on polytene chromosomes about the existence of a hierarchy of binding sites, where sites of higher affinity are primary attractants for MSL proteins and sites of lower affinity profit from local TF enrichment (24,50).

Is there a defined anatomy of high affinity sites for the MSL2?

The X-chromosomal HAS or 'chromosome entry sites, CES' were initially defined at low resolution on polytene chromosomes. With the advent of the ChIP-seq methodology it became clear that the peaks of DCC binding are several hundred base pairs long and often contain two or more MRE sequences (62,63). MEME analysis often focuses only on the strongest MRE within a larger ChIP-seq peak. Intuitively, loci with multiple MREs may attract

more MSL2 than sites with a single MRE. It is also conceivable that higher affinity sites have a defined arrangement of sites, which may facilitate cooperative interactions between MSL complexes bound to different elements. However, to our knowledge these questions have not been systematically investigated.

With respect to the diversity of MRE number and arrangement, which may or may not include PionX motifs, the intrinsic binding sites obtained *in vitro* and the physiological X chromosomal binding sites appear very similar. In both cases, we found no strong correlation between the number, length or arrangement of MRE sequences with ChIP-seq peak intensity. We conclude that the binding strength is mainly determined by the affinity of MSL2 for individual MRE motifs, where the cooperativity with CLAMP plays an important role. The predominant binding event can be inferred from our EChO analyses. MSL2 appears particularly dependent on this cooperation, since embryonic nuclei only contain very limited amounts of MSL2, barely enough to distribute one copy for each binding site mapped by C&R (64).

Quinn *et al.* suggested that MREs originally evolved from pyrimidine-rich splicing enhancers in introns and were not under selection for precise location, but rather for proximity to genes (65). Our finding that MSL2/CLAMP binding does not require a defined architecture of binding sites resonates with the observation that defined spacing and orientation of multiple binding sites are commonly not needed to assure integration of different TF input (17).

In conclusion, our analyses show that the direct contact between MSL2 and CLAMP ensures cooperativity and avoids competition between the two proteins at individual MREs. Remarkably, *in vivo* this intrinsic mechanism only operates at X chromosomal sites where the DCC is functional. Future research will focus on finding the molecular mechanism that restricts this cooperative binding to the X chromosome.

DATA AVAILABILITY

For the *in vitro* reconstitution assays the raw sequencing files in fastq format and the summarized genome browser tracks in bigwig format are available in the GEO database at GSE224981. For the CUT&RUN experiments the raw sequencing files in fastq format and the summarized genome browser tracks in bigwig format are available in the GEO database at GSE228935. To review GEO accession GSE228935. Custom code can be found as Zenodo repository and number 7816989. DOI: 10.5281/zenodo.7816989.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank S. Krebs from the LAFUGA Genomics Facility for next-generation sequencing. We thank T. Straub and T. Schauer (BMC Bioinformatic Core Facility) for scripts and advice, M. Müller for advice on protein handling and all members of the Becker lab for discussions throughout the

project. We thank E. Larschan for the generous gift of α -CLAMP antibody.

Author contributions: N.E. and P.B.B. conceived experiments; N.E. and F.G. performed experiments; S.K. cloned baculovirus expression constructs and carried out viral infections; A.C.S. prepared the sequencing libraries and optimized the CUT&RUN protocol; N.E. analyzed data; P.B.B. provided feedback and supervision; N.E. and P.B.B. wrote the manuscript; P.B.B. secured funding.

FUNDING

German Research Council (DFG) [Be1140/8-1 to P.B.B.]. The open access publication charge for this paper has been waived by Oxford University Press – NAR Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.

Conflict of interest statement. None declared.

REFERENCES

- Kim, S. and Wysocka, J. (2023) Deciphering the multi-scale, quantitative cis-regulatory code. *Mol. Cell*, **83**, 373–392.
- Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R. and Weirauch, M.T. (2018) The Human transcription factors. *Cell*, **172**, 650–665.
- Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y. *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.*, **22**, 1798–1812.
- Samata, M. and Akhtar, A. (2018) Dosage compensation of the X chromosome: a complex epigenetic assignment involving chromatin regulators and long noncoding rnas. *Annu. Rev. Biochem.*, **87**, 323–350.
- Villa, R., Schauer, T., Smialowski, P., Straub, T. and Becker, P.B. (2016) PionX sites mark the X chromosome for dosage compensation. *Nature*, **537**, 244–248.
- Villa, R., Jagtap, P.K.A., Thomae, A.W., Sparr, A.C., Forné, I., Hennig, J., Straub, T. and Becker, P.B. (2021) Divergent evolution towards sex chromosome-specific gene regulation in *Drosophila*. *Genes Dev.*, **35**, 1055–1070.
- Soruco, M.M.L., Chery, J., Bishop, E.P., Siggers, T., Tolstorukov, M.Y., Leydon, A.R., Sugden, A.U., Goebel, K., Feng, J., Xia, P. *et al.* (2013) The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes Dev.*, **27**, 1551–1556.
- Albig, C., Tikhonova, E., Krause, S., Maksimenko, O., Regnard, C. and Becker, P.B. (2019) Factor cooperation for chromosome discrimination in *Drosophila*. *Nucleic Acids Res.*, **47**, 1706–1724.
- Straub, T., Grimaud, C., Gilfillan, G.D., Mitterweger, A. and Becker, P.B. (2008) The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS Genet.*, **4**, e1000302.
- Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J. *et al.* (2008) A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell*, **134**, 599–609.
- Eggers, N. and Becker, P.B. (2021) Cell-free genomics reveal intrinsic, cooperative and competitive determinants of chromatin interactions. *Nucleic Acids Res.*, **49**, 7602–7617.
- Becker, P.B. and Wu, C. (1992) Cell-free system for assembly of transcriptionally repressed chromatin from *Drosophila* embryos. *Mol. Cell. Biol.*, **12**, 2241–2249.
- Baldi, S., Jain, D.S., Harpprecht, L., Zabel, A., Scheibe, M., Butter, F., Straub, T. and Becker, P.B. (2018) Genome-wide rules of nucleosome phasing in *Drosophila*. *Mol. Cell*, **72**, 661–672.
- Ahmad, K. and Henikoff, S. (2022) The *Drosophila* embryo as a tabula rasa for the epigenome. *Fac. Rev.*, **11**, 40.
- Duan, J., Rieder, L., Colonna, M.M., Huang, A., McKenney, M., Watters, S., Deshpande, G., Jordan, W., Fawzi, N. and Larschan, E.

- (2021) CLAMP and Zelda function together to promote drosophila zygotic genome activation. *eLife*, **10**, e69937.
16. Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E. and Taipale, J. (2015) DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, **527**, 384–388.
 17. Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C.O. *et al.* (2022) Sequence determinants of human gene regulatory elements. *Nat. Genet.*, **54**, 283–294.
 18. Oberbeckmann, E., Wolff, M., Krietenstein, N., Heron, M., Ellins, J.L., Schmid, A., Krebs, S., Blum, H., Gerland, U. and Korber, P. (2019) Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.*, **29**, 1996–2009.
 19. Voss, T.C., Schiltz, R.L., Sung, M.H., Yen, P.M., Stamatoiyannopoulos, J.A., Biddie, S.C., Johnson, T.A., Miranda, T.B., John, S. and Hager, G.L. (2011) Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell*, **146**, 544–554.
 20. Mirny, L.A. (2010) Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 22534–22539.
 21. Sönmez, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schubeler, D., Benes, V., Molina, N. and Krebs, A.R. (2020) Molecular Co-occupancy identifies transcription factor binding cooperativity In Vivo. *Mol. Cell*, **81**, 255–267.
 22. Zhao, Y., Vartak, S.V., Conte, A., Wang, X., Garcia, D.A., Stevens, E., Kyoung Jung, S., Kieffer-Kwon, K.R., Vian, L., Stodola, T. *et al.* (2022) “Stripe” transcription factors provide accessibility to co-binding partners in mammalian genomes. *Mol. Cell*, **82**, 3398–3411.
 23. Tikhonova, E., Fedotova, A., Bonchuk, A., Mogila, V., Larschan, E.N., Georgiev, P. and Maksimenko, O. (2019) The simultaneous interaction of MSL2 with CLAMP and DNA provides redundancy in the initiation of dosage compensation in *Drosophila* males. *Development*, **146**, dev179663.
 24. Dahlsveen, I.K., Gilfillan, G.D., Shelest, V.I., Lamm, R. and Becker, P.B. (2006) Targeting determinants of dosage compensation in *Drosophila*. *PLoS Genet.*, **2**, e5.
 25. Lee, H., McManus, C.J., Cho, D.Y., Eaton, M., Renda, F., Somma, M.P., Cherbas, L., May, G., Powell, S., Zhang, D. *et al.* (2014) DNA copy number evolution in *Drosophila* cell lines. *Genome Biol.*, **15**, R70.
 26. Fauth, T., Muller-Planitz, F., König, C., Straub, T. and Becker, P.B. (2010) The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. *Nucleic Acids Res.*, **38**, 3209–3221.
 27. Leibly, D.J., Nguyen, T.N., Kao, L.T., Hewitt, S.N., Barrett, L.K. and Van Voorhis, W.C. (2012) Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. *PLoS One*, **7**, e2482.
 28. Patel, A., Hashimoto, H., Zhang, X. and Cheng, X. (2016) Characterization of how DNA modifications affect DNA binding by C2H2 zinc finger proteins. *Methods Enzymol.*, **573**, 387–401.
 29. Meers, M.P., Bryson, T.D., Henikoff, J.G. and Henikoff, S. (2019) Improved CUT&RUN chromatin profiling tools. *eLife*, **8**, e46314.
 30. Liu, N., Hargreaves, V.V., Zhu, Q., Kurland, J.V., Hong, J., Kim, W., Sher, F., Macias-Trevino, C., Rogers, J.M., Kurita, R. *et al.* (2018) Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. *Cell*, **173**, 430–442.
 31. Liu, N. (2021) Library prep for CUT&RUN with NEBNext® Ultra™ II DNA Library prep kit for Illumina® (E7645). *Protocols.io*, <https://doi.org/10.17504/protocols.io.bagaibse>.
 32. Girardot, C., Scholtalbers, J., Sauer, S., Su, S.Y. and Furlong, E.E. (2016) Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinf.*, **17**, 419.
 33. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
 34. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G.P.D.P. (2009) The sequence alignment/map format and samtools. *Bioinformatics*, **25**, 2078–2079.
 35. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
 36. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
 37. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
 38. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
 39. R Core Team (2014) In: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
 40. Wickham, H., Averick, M., Bryan, J., Chang, W., D’Agostino McGowan, L., Romain, F., Garrett, G., Hayes, A., Henry, L., Hester, J. *et al.* (2019) Welcome to the tidyverse. *J. Open Source Software*, **4**, 1686.
 41. Gu, Z., Eils, R. and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, **32**, 2847–2849.
 42. Zeileis, A. and Grothendieck, G. (2005) zoo: S3 infrastructure for regular and irregular time series. *J. Stat. Softw.*, **14**, 1–27.
 43. Lawrence, M., Gentleman, R. and Carey, V. (2009) rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, **25**, 1841–1842.
 44. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*, **19**, 679–682.
 45. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
 46. Tikhonova, E., Mariasina, S., Efimov, S., Polshakov, V., Maksimenko, O., Georgiev, P. and Bonchuk, A. (2022) Structural basis for interaction between CLAMP and MSL2 proteins involved in the specific recruitment of the dosage compensation complex in *Drosophila*. *Nucleic Acids Res.*, **50**, 6521–6531.
 47. Meers, M.P., Tenenbaum, D. and Henikoff, S. (2019) Peak calling by Sparse Enrichment analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin*, **12**, 42.
 48. Kaye, E.G., Booker, M., Kurland, J.V., Conicella, A.E., Fawzi, N.L., Bulyk, M.L., Tolstorukov, M.Y. and Larschan, E. (2018) Differential occupancy of two GA-binding proteins promotes targeting of the *Drosophila* dosage compensation complex to the male X chromosome. *Cell Rep.*, **22**, 3227–3239.
 49. Villa, R., Forne, I., Müller, M., Imhof, A., Straub, T. and Becker, P.B. (2012) MSL2 combines sensor and effector functions in homeostatic control of the *Drosophila* dosage compensation machinery. *Mol. Cell*, **48**, 647–654.
 50. Demakova, O.V., Kotlikova, I.V., Gordadze, P.R., Alekseyenko, A.A., Kuroda, M.I. and Zhimulev, I.F. (2003) The MSL complex levels are critical for its correct targeting to the chromosomes in *Drosophila melanogaster*. *Chromosoma*, **112**, 103–115.
 51. Michael, A.K., Stoos, L., Crosby, P., Eggers, N., Nie, X.Y., Makasheva, K., Minnich, M., Healy, K.L., Weiss, J., Kempf, G. *et al.* (2023) Cooperation between bHLH transcription factors and histones for DNA access. *Nature*, **619**, 385–393.
 52. De Kumar, B. and Darland, D.C. (2021) The Hox protein conundrum: the “specifics” of DNA binding for Hox proteins and their partners. *Dev. Biol.*, **477**, 284–292.
 53. Moyle-Heyman, G., Tims, H.S. and Widom, J. (2011) Structural constraints in collaborative competition of transcription factors against the nucleosome. *J. Mol. Biol.*, **412**, 634–646.
 54. Hallaceli, E., Lipp, M., Georgiev, P., Spielman, C., Cusack, S., Akhtar, A. and Kadlec, J. (2012) Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. *Mol. Cell*, **48**, 587–600.
 55. Li, F., Schiemann, A.H. and Scott, M.J. (2008) Incorporation of the noncoding roX rnas alters the chromatin-binding specificity of the *Drosophila* MSL1/MSL2 complex. *Mol. Cell Biol.*, **28**, 1252–1264.
 56. Figueiredo, M.L., Kim, M., Philip, P., Allgardsson, A., Stenberg, P. and Larsson, J. (2014) Non-coding roX rnas prevent the binding of the MSL-complex to heterochromatic regions. *PLoS Genet.*, **10**, e1004865.

57. Valsecchi, C.I.K., Basilicata, M.F., Georgiev, P., Gaub, A., Seyffarth, J., Kulkarni, T., Panhale, A., Semplicio, G., Manjunath, V., Holz, H. *et al.* (2021) RNA nucleation by MSL2 induces selective X chromosome compartmentalization. *Nature*, **589**, 137–142.
58. Muller, M., Schauer, T., Krause, S., Villa, R., Thomae, A.W. and Becker, P.B. (2020) Two-step mechanism for selective incorporation of lncRNA into a chromatin modifier. *Nucleic Acids Res.*, **48**, 7483–7501.
59. Rieder, L.E., Jordan, W.T. 3rd and Larschan, E.N. (2019) Targeting of the dosage-compensated male X-chromosome during early drosophila development. *Cell Rep.*, **29**, 4268–4275.
60. Ner, S.S., Blank, T., Perez-Paralle, M.L., Grigliatti, T.A., Becker, P.B. and Travers, A.A. (2001) HMG-D and histone H1 interplay during chromatin assembly and early embryogenesis. *J. Biol. Chem.*, **276**, 37569–37576.
61. Bayona-Feliu, A., Casas-Lamesa, A., Carbonell, A., Climent-Canto, P., Tatarski, M., Perez-Montero, S., Azorin, F. and Bernues, J. (2016) Histone H1: lessons from *Drosophila*. *Biochim. Biophys. Acta*, **1859**, 526–532.
62. Straub, T., Zabel, A., Gilfillan, G.D., Feller, C. and Becker, P.B. (2013) Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. *Genome Res.*, **23**, 473–485.
63. Alekseyenko, A.A., Ho, J.W., Peng, S., Gelbart, M., Tolstorukov, M.Y., Plachetka, A., Kharchenko, P.V., Jung, Y.L., Gorchakov, A.A., Larschan, E. *et al.* (2012) Sequence-specific targeting of dosage compensation in *Drosophila* favors an active chromatin context. *PLoS Genet.*, **8**, e1002646.
64. Bonnet, J., Lindeboom, R.G.H., Pokrovsky, D., Stricker, G., Celik, M.H., Rupp, R.A.W., Gagneur, J., Vermeulen, M., Imhof, A. and Muller, J. (2019) Quantification of proteins and histone marks in *drosophila* embryos reveals stoichiometric relationships impacting chromatin regulation. *Dev. Cell*, **51**, 632–644.
65. Quinn, J.J., Zhang, Q.C., Georgiev, P., Ilik, I.A., Akhtar, A. and Chang, H.Y. (2016) Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. *Genes Dev.*, **30**, 191–207.