

Cell-free genomics reveal intrinsic, cooperative and competitive determinants of chromatin interactions

Nikolas Eggert and Peter B. Becker *

Biomedical Center, Molecular Biology Division, Ludwig-Maximilians-Universität, 82152 Planegg, Germany

Received May 03, 2021; Revised June 08, 2021; Editorial Decision June 09, 2021; Accepted June 21, 2021

ABSTRACT

Metazoan transcription factors distinguish their response elements from a large excess of similar sequences. We explored underlying principles of DNA shape read-out and factor cooperativity in chromatin using a unique experimental system. We reconstituted chromatin on *Drosophila* genomes in extracts of preblastoderm embryos, mimicking the naïve state of the zygotic genome prior to developmental transcription activation. We then compared the intrinsic binding specificities of three recombinant transcription factors, alone and in combination, with GA-rich recognition sequences genome-wide. For MSL2, all functional elements reside on the X chromosome, allowing to distinguish physiological elements from non-functional ‘decoy’ sites. The physiological binding profile of MSL2 is approximated through interaction with other factors: cooperativity with CLAMP and competition with GAF, which sculpts the profile by occluding non-functional sites. An extended DNA shape signature is differentially read out in chromatin. Our results reveal novel aspects of target selection in a complex chromatin environment.

INTRODUCTION

Upon import into the nucleus, a transcription factor (TF) scans the entire genome for potential binding sites to finally associate with a small subset of physiologically relevant loci. This exploration of the genome through 3D diffusion and sliding along the chromatin fiber is surprisingly fast, happening within minutes (1,2). The affinity of TFs to their binding sites is modulated by variations in short DNA consensus sequence motifs. Typically, complex metazoan genomes contain thousands of sequences that match TF consensus recognition sequences, but for reasons that are not immediately obvious only a small fraction are actually bound. Increasingly, we appreciate the subtle discriminators of stable protein interactions. Well-known parameters on the DNA side are the precise shape of the DNA polymer (3,4), the properties of DNA flanking the target

sequence that slow down the search diffusion (1) and chromatin organization (5). On the protein side, the intrinsic specificity of DNA-binding domains may be allosterically modulated by small molecules or by cooperative interactions with other TFs. Transcription factors must compete with nucleosome assembly or cooperate with nucleosome remodeling factors to integrate themselves into the chromatin landscape (6). Despite significant progress, it is still difficult to predict genomic binding profiles for most TFs. Clearly, our understanding of the underlying rules is rudimentary.

We address the question of binding site discrimination in a chromatin context combining unique experimental and biological systems. First, we employ a cell-free system for reconstitution of *Drosophila* genomes into physiological, embryonic chromatin representing the preblastoderm stage. Extracts from preblastoderm embryos (about 1.5 h old) efficiently assemble DNA into complex chromatin (7) containing hundreds of proteins (8) ISWI-type nucleosome sliding factors (9) and bound insulator complexes (10). Chromatin-reconstituted genomes also exhibit physiological condensation (this work).

In this system, binding of TFs to the genome *in vitro* is measured by chromatin immunoprecipitation and sequencing of associated DNA (ChIP-seq) along with mapping of the corresponding nucleosome positions by Micrococcal Nuclease sequencing (MNase-seq). Addition of recombinant TFs to chromatin-assembled genomes mimics the process of the ‘zygotic genome activation’ (ZGA) to some extent, when the first wave of transcription leads to functional diversification of the naïve preblastoderm chromatin (11). Our approach allows manipulations that cannot be easily achieved *in vivo*. For example, we can freely adjust the concentrations of TFs in absolute terms and relative to each other. We can also assess the effect of linker histone H1 on TF affinity and specificity by reconstituting genomes with and without H1 – a setting difficult to achieve in cells.

We explore the requirements for sequence-specific DNA recognition of a protein, MSL2, for which all functional binding sites are localized on the X chromosome. MSL2 is the DNA-binding subunit of the Dosage Compensation Complex (DCC). The DCC activates the transcription of the single X chromosome in male flies with re-

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

markable selectivity, a vital requirement for balancing genome expression (12). Autosomal binding therefore immediately exposes non-functional binding. Enrichment of X-chromosomal binding provides a convenient readout for the effects of experimental manipulation in our cell-free genomics experiments.

MSL2 recognizes GA-rich ‘MSL recognition elements’ [MRE, (13,14)] at some 300 X-chromosomal loci, the so-called high affinity sites (HAS). However, the consensus MRE motif (represented by the position weight matrix, PWM) occurs thousands of times throughout the genome. The small number of functional MRE sequences within X-chromosomal HAS cannot easily be discriminated from a large excess of similar sequences that do not bind MSL2 and lack functional significance. We previously identified a subset of the MREs termed ‘PionX sites’, which are specifically recognized by MSL2 through its C-terminal CXC domain (15). These sites are characterized by an extension of the MRE motif and a distinct DNA shape signature. Only these are highly enriched on the X chromosome.

Simpler GA-rich sequences are frequent in the genome and are bound by several other TFs. Two particularly abundant ones are the CLAMP protein (‘chromatin linked adaptor for MSL proteins’) (16) and the GAGA factor GAF (17). Both proteins bind to thousands of GA-rich sites across all chromosomes in both sexes. CLAMP is an essential protein that, in cooperation with MSL2, localizes to the majority of HAS, but only in male cells. MSL2 and CLAMP physically interact and together compete with nucleosomes for binding at HAS (18–20). Apparently, MSL2 coopted a very general GA-binding protein for the specific task of stabilizing its association at GA-rich MREs.

The transcription factor GAF is known to facilitate the chromatin association of other TFs in promoters, enhancers or polycomb response elements (21,22). GAF and CLAMP rarely co-localize in chromatin suggesting discriminating features in GA-rich sequences that provide exclusive selectivity (23). However, the minimal recognition sequence for GAF should allow binding to many CLAMP sites (17). In summary, the literature suggests that TF occupancy of GA-rich sequences is negotiated dynamically between several proteins with GA-binding potential. The rules that define this regulatory system are very difficult to uncover *in vivo*.

Our cell-free genomics approach allowed us to determine – for the first time – the effect of chromatin on TF binding in the context of a metazoan genome *in vitro*. We observe direct cooperativity of physically interacting TFs in competition with nucleosomes as well as indirect, nucleosome-mediated cooperativity between TFs. The cooperation of MSL2 with CLAMP enhances binding but tends to deroute MSL2 to non-functional sites. These ‘decoy’ sites resemble binding sites but are not used *in vivo*. Remarkably, GAF outcompetes MSL2/CLAMP at many decoy sites and so indirectly increases the X-chromosomal targeting of MSL2. In hindsight this observation provides a mechanistic explanation to the old observation that flies bearing a hypomorphic allele of the *trl* gene (encoding GAF) show elevated levels of male-specific lethality and inappropriate binding of MSL2 to autosomes (24).

We found that DNA shape is differentially read out by GA-binders in chromatin. Our finding that the chromoso-

mal interaction profile of a given TF does not only depend on its intrinsic properties (including its interactions with a cooperation partner), but is sculpted by an unrelated DNA binding protein with overlapping specificity, establishes a principle of broad relevance.

MATERIALS AND METHODS

DNA purification

Genomic DNA (gDNA) was obtained from male BG3-c2 cells (Drosophila Genomics Resource Center) for best male ploidy (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 and regularly tested for mycoplasma. The DNA of 10⁷ cells was purified using the Blood and Cell Culture DNA Midi Kit (Qiagen) following the supplier’s protocol. The resulting 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)

DREX was prepared from preblastoderm embryos within 90 min after egg laying (7). About 50 ml of settled embryos were dechorionated in 200 ml embryo wash buffer (EW: 0.7% NaCl, 0.04% Triton X-100) and 60 ml 13% sodium hypochlorite (VWR) for 3 min at room temperature (RT) while stirring. Embryos were rinsed for 5 min on a sieve with cold water and transferred into a glass cylinder with EW. Settled embryos were washed first in 0.7% NaCl and then in extract buffer [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)]. Embryos were settled in a homogenplus homogenizer (Schuett-Biotec), the supernatant was decanted and the embryos were homogenized with one stroke at 3000 rpm and 10 strokes at 1500 rpm. The MgCl₂ concentration of the homogenate was adjusted to 5 mM (final concentration) and centrifuged for 15 min at 27 000 g at 4°C. The white lipid layer was discarded and the supernatant was centrifuged for 2 h at 245 000 g at 4°C. The clear extract was collected with a syringe, leaving the lipid layer and pellet behind. Extracts were stored in 200 µl aliquots -80°C after shock frosting in liquid N₂. Extracts were only thawed once before use. EDTA was excluded from all steps of the purification to avoid chelation of Zn.

Chromatin assembly

About 1 µg of genomic DNA was assembled into chromatin by adding 15 µl 10× McNAP buffer (0.3 M creatine phosphate, 30 mM ATP, 3 mM MgCl₂, 1 mM DTT, 10 ng/µl creatine phosphokinase), 100 µl DREX extract and up to 150 µl total amount EX50 buffer (10 mM Hepes/KOH pH 7.6, 50 mM KCl, 1.5 mM MgCl₂, 50 µM ZnCl₂, 10% glycerol, 1 mM DTT, 1× PIC). Exact amounts of extract necessary were determined empirically for each batch. Assembly took place at 26°C for 4 h at 300 rpm on a shaking heat block.

Fluorescent microscopy of condensates

A standard chromatin assembly reaction was allowed to proceed for 4 h. The DNA was stained with 1 μ M final concentration SiR-DNA (Spiro-Chrome) for 15 min at RT. Omitting either DNA (extract only) or DREX (DNA only) served as references. Samples were placed in sealed sample chambers made by punching a hole into a double-sided sticky tape, which was in turn taped onto a glass slide and sealed with a coverslip. Widefield fluorescence microscopy was performed at the Core Facility Bioimaging of the Biomedical Center with an inverted Leica DMI8 microscope, equipped with a SPECTRA X light engine from Lumencor and a Leica DFC365 FX CCD camera. Images were acquired with a 63x/1.4 NA oil immersion objective; image pixel size was 102 nm. SiR-DNA was excited with 13% power of the SPECTRA X light engine red LED with an effective excitation range of 625–650 nm. The emitted signal was detected with a quad band filter cube with the relevant emission band 670–770 nm. The exposure time was set to 200 ms.

MNase digestions

One microgram of DNA assembled into chromatin in 150 μ l solution 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 s, 110 μ l were transferred to tubes containing to 40 μ l of 100 mM EDTA solution each to stop the digest. About 2 μ l glycogen (10 mg/ml) and 150 μ l of 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 of 10 mM TE buffer and 2 μ l of Orange-G loading dye. Samples were separated on a 2% agarose gel pre-stained with ethidium bromide and imaged using the Quantum ST-4 from PeqLab.

Baculovirus infections

Sf21 cell cultures at 10⁶ cells/ml (2.5 \times 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 \times PIC). About 25 ml buffer was added to the cell pellet (2.5 \times 10⁸ cells). After 15-min incubation on ice, the suspension was sonicated

(5 \times 10 s pulses, 20 s break, 20% amplitude, Branson digital 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. About 0.5 ml beads were used per 2.5 \times 10⁸ cells. The beads were washed twice with 10 ml ice-cold Lysis buffer, twice with 10 ml of Wash buffer (Lysis buffer, but 1 M KCl and 1% NP-40) and twice with 10 ml of Elution buffer (Lysis buffer, but 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 of 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 (NewEngland Biolabs) as a standard. Cloning for the MSL2 expression construct is described by (26).

GAF-FLAG was expressed and purified analogously to MSL2. Cloning for the GAF expression construct is described in (27) and was re-cloned for this work.

CLAMP-FLAG. Sf21 cell pellets were rapidly thawed and resuspended in 1 ml of 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 μ M ZnCl₂, 375 mM L-Arginine (according to (28)) supplemented with 0.5 mM TCEP and 1 \times PIC. After 15 min incubation on ice, the suspension was sonicated (5 \times 10 s pulses, 20 s break, 20% amplitude, Branson digital sonifier model 250-D). The extract was adjusted with Buffer C containing PI 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 (29)] 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. About 0.5 ml beads were used per 2.5 \times 10⁸ cells. Beads were pelleted at 4°C for 5 min at 500 g and supernatant was removed. Beads were washed five 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 \times 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 (NewEngland Biolabs) as a standard. Cloning for the CLAMP construct is described by (18).

Antibodies

Antibodies	source	Identifier
Anti-MSL2 monoclonal	(18)	
Anti-CLAMP monoclonal	(18)	
Rabbit Anti-GAF polyclonal	(30)	
Rabbit anti-NURF301 polyclonal:	Dr Paul Badenhorst	
IRDye 680RD Goat anti-Rabbit	Licor	925-68071
IRDye 800CW Goat anti-Rabbit	Licor	926-32211

Western blots

DREX and purified protein samples were denatured with 1× Laemmli buffer at 95°C for 5 min. Then, samples were electrophoresed on 8% SDS ServaGel TGPrime (Serva) for 1 h at 100 V. Proteins were transferred to Amersham™ Protran™ 0.45 μM Nitrocellulose Blotting Membrane using the BIO-RAD Trans-Blot® Turbo™ semi-dry transfer system for 8 min at 25 V. Membranes were then blocked with 5% milk in TBS and washed three times with 1× TBS-T. Primary antibodies were bound overnight at 4°C in TBS-T. After washing the blots three times with 1× TBS-T secondary antibodies were bound for 1 h. Images were taken using the LICOR Odyssey CLx.

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 0.1% formaldehyde for 5 min and then quenched by 125 mM glycine for 10 min. Samples were partially digested by MNase as described under ‘MNase digestions’ for 2 min. After adding 1× RIPA buffer up to 500 μl samples were precleared on a rotating wheel with 20 μl protein AG beads per 1 μg of chromatin for 1 h at 4°C.

For immunoprecipitations with monoclonal antibody 20 μl beads per sample were bound to monoclonal antibodies by adding 1 ml of culture supernatant and rotating at 4°C for 3 h. Beads were pelleted at 1000 g for 1 min and the supernatant discarded. Antibody coated beads were washed once with 1× RIPA buffer. The precleared samples were pelleted at 1000 g for 1 min and supernatant was transferred to the antibody bound beads. Binding was done overnight at 4°C on a rotating wheel.

For immunoprecipitations with polyclonal antibody IPs 1 μl of purified antibody was added to the precleared sample supernatant 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.

Then, both kinds of antibody bound samples were washed 4 times for 5 min with 1 ml of 1× RIPA buffer per sample (1 μg of chromatin on 20 μl of beads). Then, the beads were suspended in 100 μl of 1× TE buffer and de-crosslinked overnight at 65°C while shaking. Samples were then digested with 10 μg of RNaseA for 30 min at 37°C and 100 μg of proteinase K at 56°C for 1 h. beads were pelleted at 1000 g for 1 min, and supernatant was transferred to a fresh tube for purification.

DNA 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 and 2.5× volume 100% ethanol, cooling at -20°C for 15 min and pelleting 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).

Library preparation and sequencing

Next-generation sequencing libraries were prepared using NEBNext Ultra II DNA Library (New England Biolabs) according to manufacturer’s instructions and sequenced by the Laboratory for Functional Genome Analysis (LAFUGA), Gene Center Munich, Germany using an Illumina HiSeq1500 sequencer. About 25 million paired-end reads were sequenced per sample for ChIP samples and 60 million paired-end reads for MNase-sequencing samples. 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 (31) using the barcodes from the Illumina Index read files. Demultiplexed files were aligned to the *Drosophila melanogaster* release 6 reference genome (BDGP6) using Bowtie2 (32) version 2.2.9. (parameter ‘-end-to-end -very-sensitive -no-unal -no-mixed -no-discordant -I 10 -X 220’) and filtered for quality using samtools 1.6 (33) with a MAPQ score cutoff of -q 10, allowing only for high quality reads of sizes between 10 and 220 bp.

Replicate correlation

To summarize the replicates within the same experiment reads for each replicate were formatted to ‘.bed’ format using bedtools2 (34) by calling the function bamToBed and sampled to the same read count. ‘.bed’ files were imported to R and coverages were calculated. For ChIP-seq samples cumulative coverages of a 100 bp window around sites of interest were correlated between the samples to affirm similarity between them. For MNase samples average dyad densities around sites of interest were plotted and compared for each replicate to affirm similarity. If they were sufficiently similar the sampled reads were added and used for further analysis. This allowed us to avoid normalization against an input with possible zero values and to use the larger combined dataset for peak calling, thus improving the robustness of the resulting peaks.

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 8 -L 2) using the corresponding negative samples in which the IP was done without adding the respective protein as control (for NURF301 IPs the input was used). We called peaks against a negative control (IP in the absence of added protein) where possible, as this allows us to account for antibody bias 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 were used as defined by (15) 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 1). Before analysis the peaks were resized to 200 bp to include 25 bp of sequences directly bordering the peaks.

Motif search

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).

Venn diagrams

Venn diagrams were made using the peaks resized as for the *de novo* motif discovery and allowing for a maximal gap between overlapping sites of 100 bp, effectively scoring sites as overlapping if their centers are separated by <1 peak width. Plots were drawn in R using the library Vennerable (<https://github.com/js229/Vennerable>).

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.

Shape analysis

To find and align all bound motifs within the peaks we extended the peaks by 25 bp on each side and used the ‘Find Individual Motif Occurrences’ (FIMO) (37) provided by the MEME suit version 5.0.2. using the parameters -qv-thresh and -thresh 0.01. If multiple hits were recorded for one peak only the best hit was considered, to allow unbiased selection of weak motifs if no stronger motifs are nearby, while ignoring false positive overlapping motifs. False positives can happen otherwise as the motif is degenerate and repetitive allowing for multiple hits in a GA rich region. Shapes were then calculated using DNASHapeR (42) that is based on Monte-Carlo simulations to estimate approximate values for DNA Roll and Propeller Twist. Plots were drawn using ggplot2.

RESULTS

Reconstituting drosophila genomes into complex chromatin

To evaluate the cooperative DNA binding of MSL2 and CLAMP in a chromatin context we reconstituted chromatin in extracts of preblastoderm *Drosophila* embryos (7,10). This extract (DREX) assembles physiological chromatin with phased nucleosomes at chromatin boundaries (10) and a complex non-histone proteome (8). The reconstituted chromatin is in a dynamic state due to the abundance of ATP-dependent nucleosome remodeling factors (43–46). Nucleosome sliding generates ‘windows of opportunity’ for DNA-binding proteins in search for their binding sites (47). As DREX is obtained from preblastoderm embryos before the onset of transcription or dosage compensation, it is devoid of the relevant factors we wish to test, allowing us to control their concentrations by supplementing recombinant proteins [Figure 1A and Supplementary Figure S1B,C (8)].

Purifying genomic DNA from *Drosophila* cells or embryos by common kits yields fragment sizes ranging between 50 and 150 kb. Chromatin assembly of this DNA results in nucleosomes with physiological spacing, measured by digestion with Micrococcal Nuclease (MNase, e.g. Figure 2B). Other hallmarks of this physiological chromatin are phased nucleosomal arrays (PNAs) flanking sites of tightly bound protein, such as the ‘suppressor of hairy wing’ Su(Hw) insulator complex or Phaser (10). Nucleosome phasing at these sites is seen upon mapping nucleosome positions by sequencing the mononucleosomal MNase fragments (MNase-seq) following the workflow depicted in Figure 1A and aligning their dyad densities relative to genomic Su(Hw) motifs in cumulative plots. Nucleosome phasing (Figures 1B and 2C) serves as important quality control for efficient chromatin assembly and as internal reference for binding of recombinant factors.

Observing the reconstituted chromatin by fluorescence microscopy, we found it condensed in aggregates of variable sizes ranging between 1 and 10 μm (Supplementary Figure S1A), as expected from polymer theory and previous reports (48–50). We assume that local chromatin concentration in these condensates approaches values of preblastoderm nuclei. We conclude that the reconstituted chromatin provides a *bona fide* physiological substrate to test the binding of recombinant transcription factors on a genome-wide scale. To our knowledge, this is the first study of this kind using a chromatin-reconstituted metazoan genome.

Genome-wide assessment of in vitro CLAMP-chromatin interactions

As a proof of principle for the approach we first assessed the binding of CLAMP. We previously established the intrinsic DNA-binding specificity of CLAMP in genome-wide DIP experiments (18), providing excellent reference profiles. CLAMP binds >4000 sites in the genome that are characterized by strings of GA dinucleotides. Roughly 30% of these sites overlap the *in vivo* binding sites for CLAMP, where CLAMP may cooperate with other TFs (18). ‘False positive’ binding only observed *in vitro* may be due to lack of nucleosome competition. This hypothesis can now be tested by chromatin reconstitution.

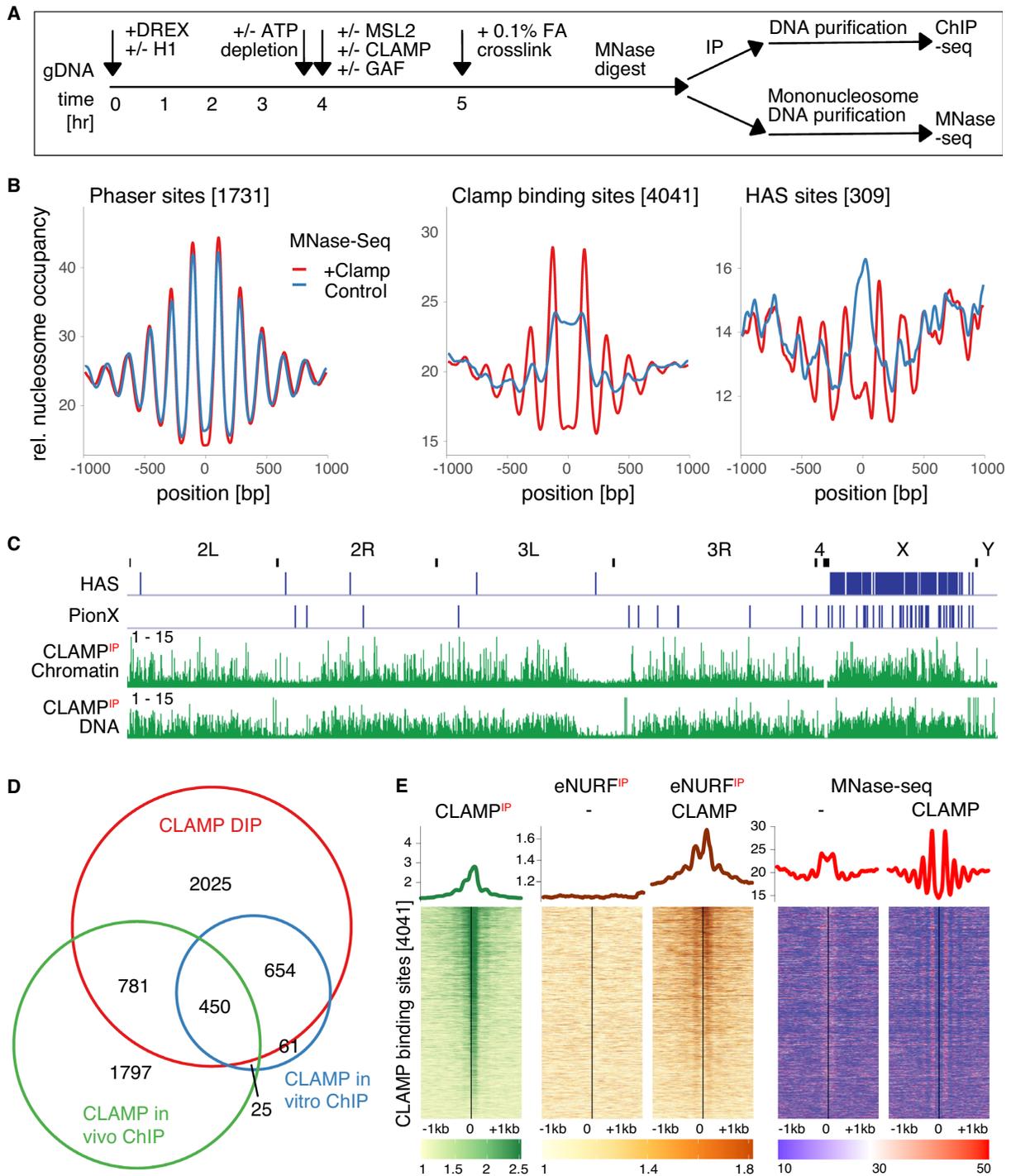


Figure 1. Genome-wide binding profile of CLAMP in reconstituted chromatin. **(A)** Timeline of a ChIP-seq and MNase-seq experiment. **(B)** Nucleosome occupancy at selected sites. Chromatin was assembled in the presence and absence of CLAMP and averaged nucleosome dyad densities were determined by MNase-seq of the mononucleosomal bands. Summarized nucleosomal occupancy of three biological replicates are shown relative to the motif position at the sites of interest ($n = 3$). For separate analysis of each replicate at CLAMP-binding sites, see Supplementary Figure S1D. **(C)** Genome-wide browser overview of ChIP-seq and DIP-seq profiles for CLAMP. Chromatin was assembled and CLAMP added as depicted in (A). Summarized genome browser profile from two biological replicates showing CLAMP binding over the whole genome determined by ChIP-seq. Coverage was normalized against the control (ChIP in absence of factor addition), the maximum values for each window are shown. The *Drosophila* chromosomes (2L, 2R, 3L, 3R, 4, X, Y) are indicated and HAS, PionX sites are annotated. CLAMP DIP sequencing profile from (18) was added for comparison. For replicate correlation analysis see Supplementary Figure S6A. **(D)** Venn Diagram relating the genomic binding events (peaks) of CLAMP binding, as determined from ChIP profile in (B), from *in vitro* DIP and from *in vivo* conditions (18). Numbers shown in Venn diagrams might differ from numbers shown elsewhere, as peaks from one set may overlap multiple peaks in another. **(E)** Average profiles and heatmaps of factor enrichment (ChIP-seq) or dyad density (MNase-seq) around known CLAMP-binding sites. Green: CLAMP ChIP; brown: ChIP of endogenous NURF301 in the absence (left) and presence (right) of CLAMP. Blue: nucleosome positioning in the absence (left) and presence (right) of CLAMP. For replicate correlation analysis see Supplementary Figure S6B.

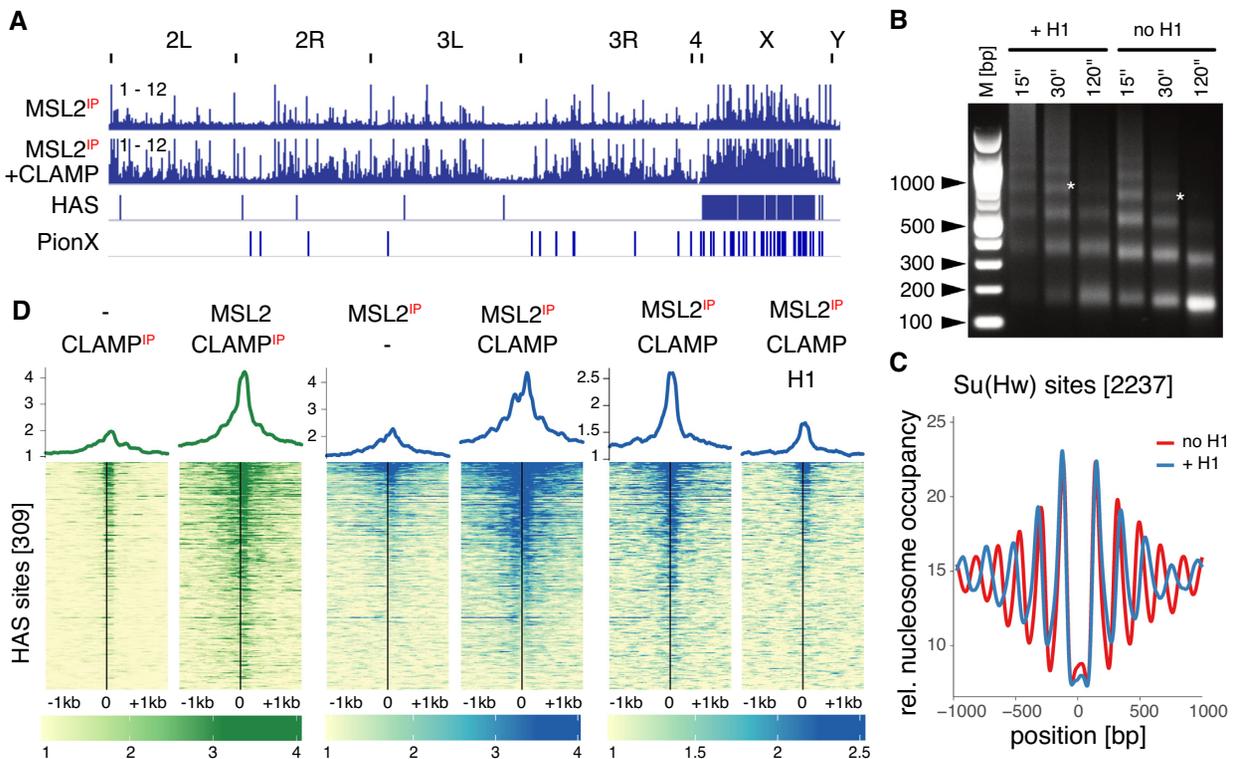


Figure 2. Cooperativity of MSL2 and CLAMP in chromatin interactions. (A) Samples were prepared as depicted in Figure 1A. Browser Overview along the genome of MSL2 binding in absence and presence of CLAMP determined by ChIP-seq ($n = 3$). Coverage was normalized against the control, maximum values for each window are shown and HAS, PionX sites and chromosomes are annotated. For replicate correlation analysis see Supplementary Figure S6C. (B) Reconstitution of H1-containing chromatin. Chromatin was assembled in the presence and absence of 100 nM histone H1 and then digested with MNase for the indicated times. The resulting fragments were visualized after agarose gel electrophoresis by ethidium bromide staining. M: DNA size marker in bp. Asterisks: tetranucleosomal fragments. (C) Nucleosomes in reconstituted chromatin (containing H1 or not) were mapped by MNase-seq. Nucleosome dyad densities were aligned to 2237 genomic Su(Hw) sites. Coverage windows of 2000 bp around the sites were cut out, aligned and the mean for each column calculated. The average profiles summarize the results of two biological replicates ($n = 2$). For separate analysis of each replicate for Su(Hw) and Phaser binding sites see Supplementary Figure S2B. (D) Binding of MSL2 and CLAMP to HAS. Samples were prepared as depicted in Figure 1A and added factors are indicated. Enrichment of the targeted factor marked by red 'IP' normalized to the control at 309 High Affinity sites (HAS) is illustrated by average profiles (top) and heatmaps of individual regions. Coverage windows of 2000 bp around the sites were cut out, aligned and the mean for each column calculated. The loci are sorted by MSL2 ChIP peak height in presence of CLAMP. For replicate correlation analysis see Supplementary Figure S6C,D.

DREX does not contain CLAMP (Supplementary Figure S1B), which enables assessing the chromatin interactions of recombinant protein. Chromatin was assembled for 4 h, CLAMP was added, allowed to bind for 1 h and crosslinked with formaldehyde (following *in vivo* protocols). CLAMP-binding sites were mapped by ChIP-seq using a specific monoclonal antibody (Figure 1A). The minimal amount of recombinant protein to yield high-quality ChIP-seq libraries was determined empirically to be 25 nM and used throughout, unless stated otherwise.

CLAMP binding was detected at numerous sites in reconstituted chromatin (Figure 1C). The CLAMP ChIP profile resembled the profile on naked genomic DNA (18), with 1004 of 1090 ChIP peaks overlapping in the DIP (93%, Figure 1D). Stringent peak calling highlighted about four-fold less CLAMP binding events in chromatin versus protein-free genome (Figure 1D), but heat maps show that most of the DIP sites are bound to some extent (Figure 1E, left). The dominant motif of CLAMP-binding sites in chromatin

or free DNA are stretches of GA (Figure 4, row 13, note that Figure 4 summarizes salient features of all binding profiles of this study). Chromatin apparently suppresses the binding to very long, nonphysiological dinucleotide repeats. The mild enrichment of CLAMP binding to the X chromosome was also observed *in vivo* and is explained by a corresponding enrichment of GA repeats (51).

Because CLAMP *in vitro* ChIP and DIP are so similar, we considered that CLAMP binding sites may be devoid of nucleosomes. Nucleosome mapping clearly showed that this was not the case: aligning cumulative nucleosome dyad density profiles to the 4041 CLAMP *in vitro* binding sites and corresponding heat maps showed that in the absence of CLAMP nucleosomes were even enriched on the GA recognition sequences (Figure 1B, center). Remarkably, upon addition of CLAMP the nucleosomes realigned to form phased arrays flanking the GA elements. Nucleosome phasing at 1731 Phaser binding sites (10) confirmed that the addition of CLAMP did not change nucleosome po-

sitions globally (Figure 1B, left). The dynamic repositioning of nucleosomes is explained by the presence of abundant ATP-dependent nucleosome sliding factors, such as CHRAC, ACF and NURF (9).

CHRAC and ACF are thought to slide nucleosomes in an untargeted manner and can facilitate the chromatin binding of ectopic proteins, like restriction enzymes (43,44). By contrast, NURF is recruited by a range of transcription factors, including CLAMP (52). Since DREX contains abundant NURF we monitored its distribution in genomic chromatin. NURF binding was detected at the CLAMP-binding sites only in presence of CLAMP demonstrating a direct recruitment. NURF was crosslinked predominantly to the nucleosomes flanking bound CLAMP, suggesting that CLAMP directs the remodeler to specific nucleosome substrates (Figure 1E). We conclude that the dynamic nature of chromatin assembled in DREX provides an ideal system to explore the mechanisms of transcription factor interactions with the *Drosophila* genome.

Cooperative binding with CLAMP promotes binding of MSL2 to GA rich sites

Physiological CLAMP binding events that are not observed in genome-wide DIP suggest that CLAMP cooperates with other factors *in vivo*. Indeed, a small subset of CLAMP-binding sites coincide with the X-chromosomal HAS for the DCC, where CLAMP cooperates with MSL2 to promote stable association (18,19). Mapping the nucleosome dyad densities at the 309 HAS reveals them to be largely occupied by nucleosomes in the absence of CLAMP, but CLAMP can clear these sites of nucleosomes and establish phasing (Figure 1B, right panel).

Nucleosome-free regions with positioned flanking nucleosomes are a hallmark of HAS *in vivo* (18,53,54). Does this CLAMP-dependent reorganization of chromatin affect the binding of MSL2 *in vitro*? Comparing the chromatin interactions of recombinant MSL2 in the absence of CLAMP (Figure 2A) to previous DIP profiles (Supplementary Figure S2A), we found that the profile changed. Many smaller peaks were repressed but others remained robust. Requiring the replicated signals to be 8-fold over background (see Materials and Methods section), MSL2 was mapped to 131 prominent sites in the chromatinized genome (Figure 2A; representative profiles of all *in vitro* ChIP experiments are shown in Supplementary Figure S3A,B). Although of these 131 binding sites, 74% are located on the X chromosome, only 28 overlap with HAS (Figure 2D). A search for shared motifs among the MSL2-binding sites yielded a consensus sequence consisting of a long string of GA repeats (Figure 4, row 4). This is reminiscent of the dominant feature of functional MREs (51), but longer. Apparently, MSL2 alone can interact well with long stretches of GA in chromatin.

Next, we added equimolar amounts of CLAMP and MSL2 to reconstituted genomic chromatin and scored their mutual effects on chromatin interactions by ChIP using specific antibodies. In the presence of CLAMP, MSL2 was able to bind to the majority of HAS, although with a gradient of intensities (Figure 2D). Remarkably, the reverse was also true: in the presence of MSL2, CLAMP bound to

most HAS (Figure 2D, Figure 4, row 14). This cooperativity between MSL2 and CLAMP to bind HAS in a chromatin context is an important aspect of HAS identification in cells (18). These chromatin interactions relied on ATP-dependent nucleosome remodeling, since depletion of ATP after chromatin assembly by adding hexokinase and glucose (55), abolished all binding (Supplementary Figure S2C).

In the presence of CLAMP, MSL2 not only bound HAS but also to 321 other sites in the genome, including many shorter GA strings (Figure 4, row 5). Many of these sites are on autosomes, leading to a decrease in X chromosomal specificity from 74% to 64.5% (Figure 4, rows 4, 5; a more visual illustration of X chromosome enrichments of sequence features and binding events is displayed in Supplementary Figure S3C). CLAMP promotes the binding of MSL2 to HAS (20% versus 9% HAS bound by MSL2 alone), but due to many more binding events on autosomes, the percentage of peaks on HAS drops from 21% to 16% (Figure 4, rows 4, 5). This is reminiscent of the situation seen in genome-wide DIP experiments, where CLAMP stabilizes the binding of MSL2, but also de-routes MSL2 to non-functional decoy sites. This contrasts the *in vivo* situation where the synergistic interaction of CLAMP and MSL2 is only observed at functional HAS (18).

Histone H1 generally reduces MSL2 binding without influencing X specificity

The inability of nucleosomes to occlude decoy sites might be due to the lack of canonical linker histone H1 in preblastoderm embryos, which instead contain mostly the linker binding proteins HMG-D and bigH1 (56,57). This chromatin may be particularly open for the transcription factors that orchestrate the first wave of zygotic transcription. The linker histone is essential for proper chromosome organization and genome function (58,59). Conceivably, H1 may modulate nucleosome positioning and fiber folding to render functional and nonfunctional binding sites differentially accessible to MSL2/CLAMP.

Purified H1 can be faithfully incorporated into plasmid chromatin if added to DREX at the onset of the assembly (7,60,61). We scaled up this experiment to the genome-wide level, where we can, for the first time, compare nucleosome positions in the absence and presence of the linker histone. Titrating in physiological levels of H1 leads to the expected increase in nucleosome repeat length (NRL) in bulk chromatin [(60,62), Figure 2B]. The lengthening of the NRL by 25 bp is best observed at sites of nucleosome phasing, e.g. by aligning the nucleosome dyad density profiles at the 2237 binding sites for the Su(Hw) insulator (Figure 2C).

In the presence of H1 (and CLAMP) MSL2 binding to chromatin was reduced by about 75%. Binding was equally reduced at all sites (Figure 2D) and, accordingly, the linker histone did not affect the enrichment of peaks to the X chromosome (Figure 4, row 6). A MEME analysis of the respective peak sets showed an increase of motif strength after addition of H1 suggesting that peaks with a more concise MRE motif tend to be bound better (Figure 4, row 6). We conclude that H1 uniformly reduces the accessibility of all binding sites throughout the genome and hence does not contribute to X chromosomal specificity.

GAF competes with CLAMP to refine MSL2-MRE specificity

The dominant mechanism through which MSL2 is re-routed to decoy sites *in vitro* is the physical interaction with CLAMP, which binds GA-rich sequences in general. *In vivo*, many of these sites are occupied by the GAGA factor [GAF, (22,23)]. GAF recognizes GAGAG sequences, but oligomerizes to select longer GA repeats (17,63,64). Since GAF is absent from DREX [Supplementary Figure S1C, (27,65)], we hypothesized that CLAMP may bind native GAF sites *in vitro*. To test this hypothesis, we included recombinant GAF as a potential competitor in MSL2/CLAMP binding reactions (Figure 3A). Monitoring GAF binding by ChIP in the absence of CLAMP and MSL2 shows that at a concentration of 25 nM GAF binds to most of the 4041 *in vitro* CLAMP sites (Figure 3A, panel 1). We found that GAF on average binds best to its own physiological binding sites followed by known CLAMP-binding sites, but does not bind HAS as tightly (Supplementary Figure S5A). Next, we titrated GAF into MSL2/CLAMP binding reactions at GAF/CLAMP ratios of 1/3 (8 nM, low), 1/1 (25 nM, equimolar) and 3/1 (75 nM, high). The binding reaction was split and used partly to monitor the increase in GAF binding (Figure 3A, panels 2–4) and partly to test for MSL2 binding (Figure 3, panels 6–8). In Figure 3A, the mean peak heights can be derived from the cumulative plots, while the heatmaps show the individual binding events. For an overview of the peaks bound by each factor see Figure 4.

Surprisingly, adding GAF at low concentration increased the MSL2 association, particularly with decoy sites (Figure 3, panels 5, 6). Apparently, GAF promoted binding of MSL2 *in vitro*, which may well be due to nucleosome-mediated cooperativity (5). GAF can keep regulatory sequences clear of nucleosomes and cooperate with a variety of DNA-binding proteins at composite regulatory sequences (21). However, the quality of cooperation between GAF and MSL2, which are not known to interact, is different from the cooperativity between CLAMP and MSL2 (compare Figure 3, panel 5 versus 9). CLAMP and MSL2 interact physically (18) and it is likely that the defined geometry of the complex selects a different subset of sites.

High amounts of GAF effectively competed for MSL2/CLAMP binding and led to a reduction in binding and specificity (Figure 3A, panel 5 versus 8). Equimolar concentrations of GAF resulted in intermediate competition with significantly reduced numbers of MSL2 peaks (panel 7). Remarkably, and in contrast to the global repression by H1, 76% (84/111) of the MSL2 binding events in the presence of CLAMP and GAF were now on the X chromosome (Figure 4, row 9). GAF binding to decoy sites on the autosome prevented MSL2/CLAMP interaction, leading to substantial X chromosome enrichment. Indeed, GAF did not affect the association of MSL2/CLAMP with HAS much, as GAF binds HAS less well than most CLAMP sites (Figure 3A, panels 1–4 versus 10–13). MSL2/CLAMP binding at HAS appears relatively resistant against GAF competition (panels 14–16).

The differential competition by GAF is best illustrated by ChIP profiles at representative individual loci. Strong

GAF binding to non-HAS sites on the X (Figure 3B) or autosomal GA-sequences (Supplementary Figure S4D) effectively competes for MSL2/CLAMP interaction. By contrast, GAF tends not to bind to HAS on the X, and therefore does not compete with MSL2/CLAMP. This leads to a substantial enrichment of MSL2/CLAMP at HAS. In the absence of GAF, only 16% of MSL2 binding sites overlapped with HAS, while in the presence of GAF 35% overlap (Figure 4, rows 5 and 9). Remarkably, a MEME analysis among the MSL2 peaks gained through the addition of GAF yielded a PWM that strongly resembles the PionX motif, with a 5' extension of the GA-rich MRE. The binding sites that are lost upon competition are mostly characterized by long GA stretches of low complexity (Figure 4, row 11). These findings establish binding site competition as an important principle that contributes to the refinement of binding site selection.

DNA shape distinguishes functional and decoy sites in complex chromatin

We wondered why GAF competes with MSL2/CLAMP only for binding at decoy sites but less so at functional sites. We know that the functional PionX signature is defined by a DNA signature that combines sequence and shape. Accordingly, many DNA sequences that conform only with the PionX sequence are non-functional (15). What are the DNA shape features in chromatin that enables MSL2 to distinguish functional sites from PionX decoys?

To address this question, we performed a FIMO search with the PionX PWM (Supplementary Figure S4A) to obtain an unbiased pool of functional and non-functional sites, which, due to the asymmetric nature of the motif, can be well aligned for DNA shape analysis. Of the 500 sites with best PWM-hit-score, roughly 40% are located on the X chromosome, including 71 HAS, and 60% are located on autosomes and, therefore, correspond to decoy sites (Supplementary Figure S4A,B).

The MSL2 ChIP signals at these 500 sites in the presence of CLAMP and GAF were displayed in a heat map (Supplementary Figure S4C). Sorting this heatmap by MSL2 binding strength revealed an enrichment of HAS in the upper third of the sites, but only in the presence of GAF. In the absence of GAF, the HAS sites were distributed throughout the map. Evidently, GAF occludes decoy sites and leaves functional PionX sites for MSL2 to bind. This points to a difference in these sites that is not explained by the PionX PWM. To understand how decoy sites differ from functional PionX sites, a look at extreme cases is instructive. We first determined the most common motif within the top or bottom 50 MSL2-binding sites in the heat map, in presence or absence of GAF. In the presence of GAF, the bottom sites (not bound by MSL2) share an extended GA dinucleotide repeat. The top MSL2-binding sites, however, are characterized by a shorter and discontinuous sequence more closely resembling the original PWM (Supplementary Figure S4C). This discrimination is only observed in the presence of GAF.

We then considered the DNA shape features of these sites. We generated peak sets of all binding reactions and precisely aligned the bound motifs within each set (see Materials and

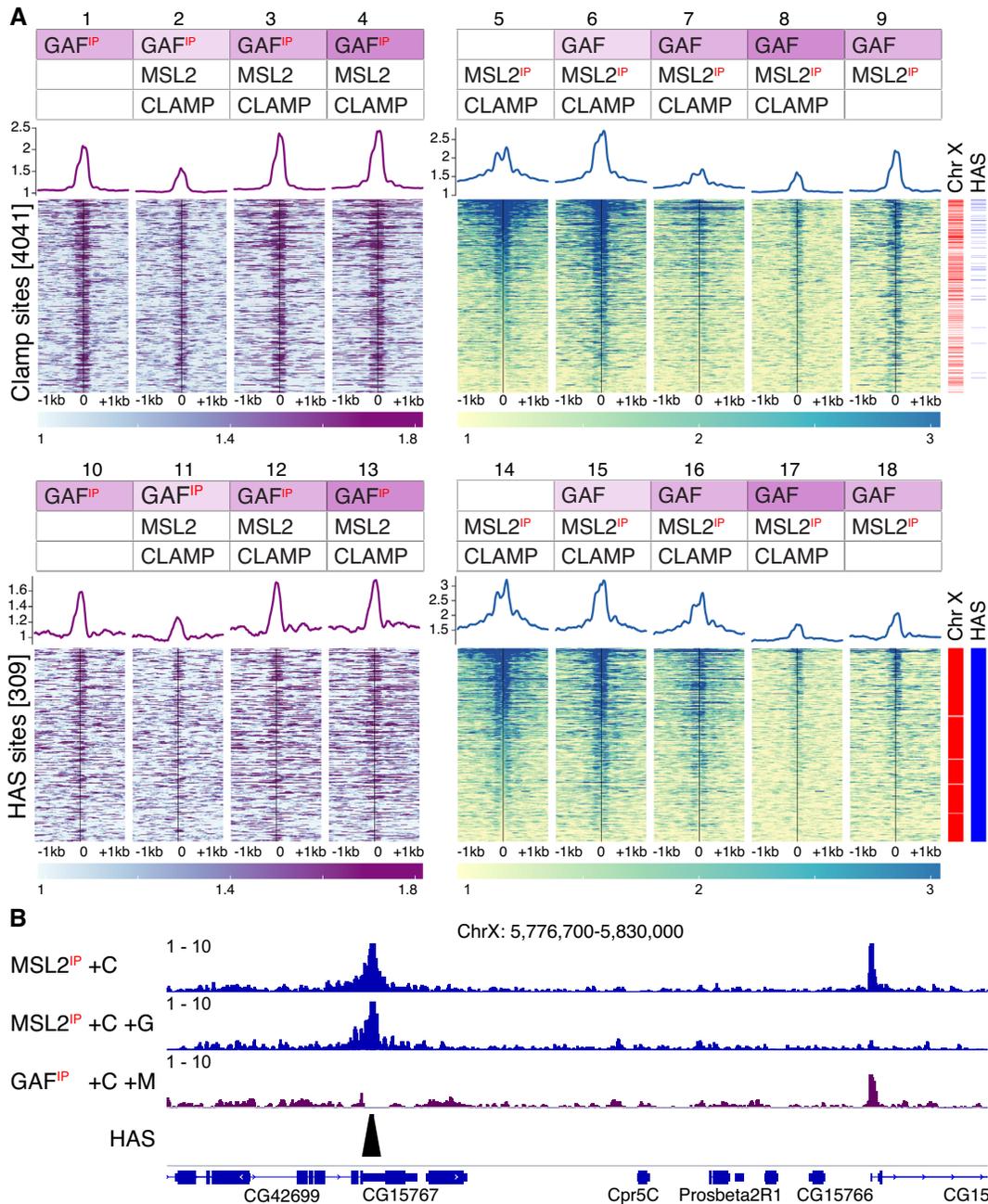


Figure 3. GAF competes with MSL2/CLAMP for binding at decoy sites. **(A)** After chromatin assembly the annotated proteins were added and MSL2 or GAF binding determined by ChIP-seq. Enrichment of the targeted factor marked in red by 'IP' normalized to the control at 4041 CLAMP-binding sites (1–9) or HAS (10–18) illustrated by average profiles (top) and heatmaps of individual regions. CLAMP sites represent potential but mostly nonfunctional MSL2-binding sites (decoy sites). Coverage windows of 2000 bp around the sites were cut out, aligned and the mean for each column calculated. Two biological replicates were normalized and summarized values are shown. Concentration of MSL2 and CLAMP are 25 nM where present. The color gradient indicates GAF concentrations (8.3, 25 and 75 nM). Heatmaps are sorted by the signal strength of MSL2 IP in presence of CLAMP, representing our positive control, in a 100 bp window around the center (panels 5 and 14). For replicate correlation analysis of GAF IPs see Supplementary Figure S6E, for MSL2 IPs see Supplementary Figure S6F. **(B)** Browser profile from two biological replicates showing MSL2 binding in presence of CLAMP (C) and/or GAF (G) or GAF binding under the same conditions at a representative locus determined by ChIP-seq ($n = 2$). Coverage was normalized against the respective controls, the maximum values for each window are shown scaled as indicated and HAS are annotated.

Sample	total peaks	on chrX [%]	% peaks on HAS	% HAS bound	% peaks on PionX	% PionX bound	Position Weight Matrix	found in n peaks
1 Genome		18						
2 HAS	309	98.4	100	100	12	66		301
3 PionX	56	78.6	66	12	100	100		56
4 MSL2	131	74	21	9	6	14		93
5 MSL2 + C	383	64.5	16	20	5	34		264
6 MSL2 + C + H1	91	61.5	22	6	8	12		65
7 MSL2 + C + 8 nM G	797	55	9	23	3	41		484
8 MSL2 + C + 75 nM G	93	34.4	11	3	9	14		30
9 MSL2 + C + 25 nM G	111	75.7	35	13	23	43		84
10 MSL2 gained by GAF*	36	77.8	28	3	22	14		17
11 MSL2 lost by GAF*	306	62.1	11	11	1	4		217
12 MSL2 + 25 nM G	363	49.3	3	4	2	14		242
13 CLAMP	1259	36.5	5	22	1	16		1238
14 CLAMP + M	1584	42.2	7	33	1	27		1448
15 25 nM GAF	155	42.6	3	2	1	2		88
16 25 nM GAF + C + M	913	35.2	2	6	0	0		697
17 25 nM GAF + M	363	49.3	3	4	2	14		242
18 75 nM GAF + C + M	2148	30.4	1	10	0	2		1603

Figure 4. Summary table comparing transcription factor binding profiles. Summary Supplementary Table Showing characteristics of genome-wide binding profiles of transcription factor binding profiles under the conditions listed under ‘sample’ (for details, see text). ‘Genome’ indicates the fraction of DNA on the X chromosome, ‘HAS’ and ‘PionX’ represent the collection of MSL2 High Affinity Sites (53) and PionX sites (15) as references. The table displays the absolute number of peaks called (total peaks), their X chromosomal enrichment, the fraction of peaks that overlap with HAS or PionX sites (% peaks on HAS or PionX, respectively) and the fraction of HAS and PionX sites bound by MSL2 (% peaks of HAS or PionX bound, respectively). Position Weight Matrices are derived from the corresponding peak sets as determined by MEME. The shading refers to important shape features derived from Figure 5A,B. Lastly, the number of motif hits within each peak set (that contributed to the PWM) is shown. For description of subsample marked with * see Supplementary Figure S5B.

methods section). We then plotted the ‘Roll’ and ‘Propeller Twist’ parameters for each base pair along the motif for each condition and connected the median values to provide a visual aid to better appreciate the signature (Figure 5A,B and Supplementary Figure S5C,D). Shape profiles of HAS serve as reference for functionally relevant sites, while the profile of motifs found within GAF peaks represents decoy sites. The motifs bound by MSL2 alone and in presence of CLAMP resemble sequences bound by GAF and differ from the functional HAS profile (Figure 5A,B, left and middle panel). Only in the presence of CLAMP and GAF, MSL2 selected a DNA shape that resembles the HAS

signature. Specifically, these show a high median Roll at position +3 and +18 and a region of low Roll and Propeller Twist between positions +6 to +8 (Figure 5A,B). Additionally, we see a high roll at position +1, the previously known signature for PionX motifs. All HAS together show a lower median roll at this position, but at a high variability, as PionX sites are only a subset of HAS (Supplementary Figure S4B).

Interestingly, the functional signature includes particularly high variability of DNA shapes at some positions, e.g. undefined DNA Roll at positions +8, +13 and +18, directly adjacent to the GA sequence in the ‘classical’ MRE mo-

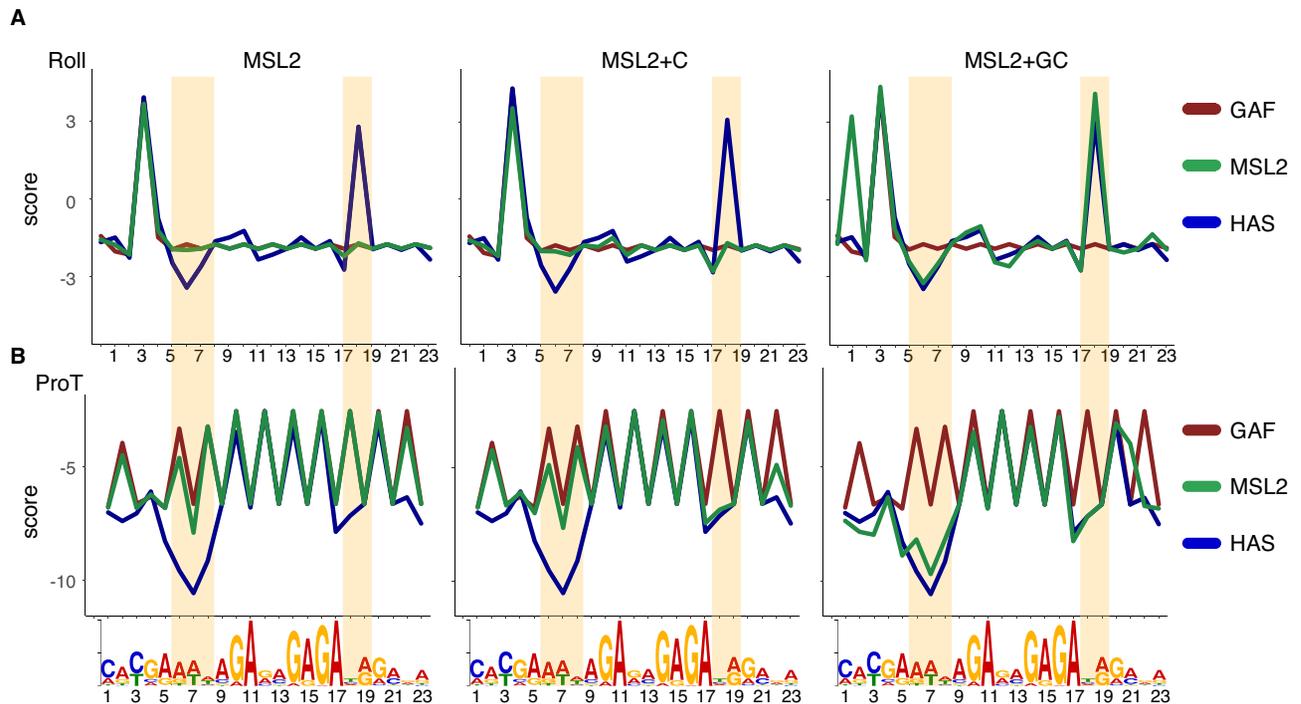


Figure 5. Sculpting of the functional MSL2 binding profile through GAF competition. **(A)** DNA shape features of factor binding sites. A large collection of sites containing PionX PWM sequences were interrogated for MSL2 binding by FIMO ($-\text{qv-thresh} -\text{thresh} 0.01$). If an MSL2 peak had multiple PionX motifs only the best-scoring one was considered. For the subset of MSL2-bound motifs the DNA shape feature ‘Roll’ was determined along the 23 base pairs of the motif and plotted relative to the consensus sequence (Roll is an inter-base feature, for simplicity +1 Roll reflects the Roll from base +1 onto +2). This was done for MSL2 alone ($n = 131$), or MSL2 in presence of CLAMP (+C, $n = 383$), or MSL2 in presence of CLAMP and GAF (+C+G, $n = 111$), as indicated in the panels. The data plotted is the median of the full dataset represented in Supplementary Figure S5C,D. The median values have then been connected with line plots for better visualization. The shape signatures of HAS and of sites bound by GAF serve as references for functional and non-functional binding sites, respectively. **(B)** As in (A) but for Propeller Twist. Propeller Twist is an intra-base feature therefore +1 reflects the value at base +1.

tif. The large variability in Roll at locations +9 and +18 in MSL2 binding sites reveals that MSL2 itself does not distinguish Roll there. Apparently, the precise shape at these sites is less important, provided the GA dinucleotide repeat is broken and ‘GAGAG’, a strong GAF-binding motif, is avoided (Supplementary Figure S5C,D). MSL2 thus binds to the sites left unoccupied by GAF, explaining how after addition of GAF the variability of Roll and Propeller Twist of the sites bound by MSL2 increases.

Previously, we had found a local Roll feature as a hallmark of PionX sites that are recognized by MSL2 via its CXC domain (15). The more profound reconstitution now yields a more complex shape signature with diagnostic Propeller Twist and Roll features. We conclude that the discrimination of functional HAS versus decoy sites that is a hallmark of MSL2 binding *in vivo*, is only recapitulated *in vitro* in the presence of CLAMP and GAF, indicating that MSL2 relies on extrinsic factors to achieve proper X/autosome discrimination.

DISCUSSION

We employed a unique system for the reconstitution of entire *Drosophila* genomes into complex chromatin to explore the mechanisms of binding site selection by transcription factors under most physiological conditions. The system recapitulates several known principles: the readout of

DNA sequence and shape by DNA-binding domains, the direct cooperativity between TFs and indirect, nucleosome-mediated cooperativity. Remarkably, the study also revealed the importance of a less appreciated principle (66): we found that the selection of genomic binding sites did not only depend on the intrinsic properties of the TF, but on competition with a much more abundant factor for similar target sequences.

Competition by abundant GAF refines the MSL2 binding profile

The dosage compensation regulator MSL2 is ideally suited for an in-depth analysis of TF selectivity. All functionally relevant binding sites reside on the X chromosome and, consequently, association with autosomal sequences reveals erroneous binding site selection. To a first approximation, the X/autosome enrichment of binding events provides an objective measure of ‘faithfulness’ in binding site selection.

MSL2 is not abundant in nuclei. The experimentally determined number of about 600 copies MSL2 per nucleus (67) must be related to about 300 high affinity sites (HAS) of the complex. This latter value underestimates the number of *bona fide* MSL binding sites since many HAS contain several MREs and sites of lower affinity may still be functionally relevant. The limiting amounts of MSL2 are critical for the selectivity of X-chromosome binding. If MSL2 lev-

els are experimentally elevated *in vivo*, the excess protein will bind to nonfunctional, autosomal sites (68,69).

Perhaps due to its low concentration, MSL2 cooperates with the abundant CLAMP protein for robust MRE association (18,19,70). Because CLAMP has thousands of physiological binding sites throughout the genome, there is a risk of ‘de-routing’ MSL2 to these inappropriate sites. Here, competition by the GAGA factor comes into play. CLAMP and GAF bind GA-rich sequences but curiously never colocalize *in vivo* (18,23). GAF is 30–100-fold more abundant than MSL2 (17,67). Our observation that the abundant GAF occludes many GA-rich sequences and prevents inappropriate MSL2/CLAMP binding to decoy sites provides a mechanistic explanation to the old observation that flies bearing reduced levels of GAF (due to a hypomorph allele of the *trl* gene that encodes GAF) show elevated levels of male-specific lethality and inappropriate binding of MSL2 to autosomes (24).

Our work documents the contribution of MSL2 and of the GA-binding proteins it interacts with, to X chromosome targeting through recognition of X-specific DNA sequence elements. However, the principles discussed here cannot explain the exclusive localization of the DCC to the X. The DCC lncRNA subunit roX2 has most recently been shown to play an important role in X chromosome targeting (71,72). RoX2 interacts with MSL2 close to the CLAMP binding domain (71,73) and thus may modulate the cooperative interactions and de-routing activity of CLAMP (72). MSL2 and roX can form phase-separated condensates (71) and nascent roX RNA can concentrate MSL proteins around its gene (74,75). Because roX RNAs are transcribed from the X chromosome, it is plausible that the multivalent interactions that underlie such condensation contribute to X chromosome targeting.

Cell-free chromatin genomics

The situation we reconstitute *in vitro* is reminiscent of the chromatin transition upon zygotic genome activation (ZGA) in early fly embryogenesis, when H1 is expressed, TFs are induced and dosage compensation is established (11). CLAMP and GAF are translated from maternal RNAs to orchestrate early zygotic transcription. They may be considered pioneer factors due to their interaction with NURF (52,76). The binding sites for CLAMP and MSL2 were occupied by nucleosomes in the absence of the interacting transcription factors, as is the case *in vivo* (18).

Nucleosomes, either randomly positioned or specifically placed on functional sites, may serve as ‘gate-keepers’, occluding irrelevant sequences and leave functional sites available for interaction (77). Strikingly, all TFs added to the reconstituted, physiologically spaced and compacted chromatin were able to efficiently access their recognition sequences in an ATP-dependent manner. GAF was shown earlier to rely on ATP-dependent sliding factors for chromatin interactions *in vitro* (27,65). The stoichiometric incorporation of linker histone H1 dampened the interaction efficiency somewhat, but allowed qualitatively similar TF interactions. This is in line with biochemical observations that ISWI can slide chromatosomes (78,79) This situation

may well reflect the *in vivo* situation, since ISWI remodelers are very abundant (80).

Blending the biochemistry of TF interactions on chromatin with ‘genomics’ yields the statistical summation of many events, yet allows zooming in on individual binding at chromosomal loci. The complexity of the genomic DNA matches the *in vivo* situation. This includes competition by repetitive sequences that are usually not considered in routine ChIP analyses, but are nevertheless relevant. Of note, GAF was shown to bind to centromeric heterochromatin at the GAGA satellite repeat during mitosis (81,82) and, likewise, MSL2 may also be trapped there, if in the absence of roX RNA the DCC is not assembled properly (83).

Distinct types of cooperativity between GA-binding proteins

Monitoring the influence of CLAMP and GAF on the interaction of MSL2 with the chromatinized genome revealed two types of cooperativity (Figure 3A, panels 5, 9). At low concentrations, GAF promoted MSL2 binding to long GA stretches. We assume that this is explained by nucleosome-mediated cooperativity, or ‘assisted loading’ if one TF profits from action of the other, keeping nucleosomes off the shared binding site (5). GAF, perhaps due to its ability to interact with NURF (76) can keep regulatory sequences sites clear of nucleosomes. Any TF with a response element close to the GAGAG binding site will profit from this increased accessibility (84). This indirect cooperativity does not rely on direct, physical interaction between TFs and, accordingly, is widely used *in vivo* (85).

By contrast, the cooperativity between MSL2 and CLAMP is reciprocal and direct, i.e., mediated by physical contact between both factors (18). Both proteins may contribute to a composite DNA binding surface explaining the different MSL2 profiles in complex with CLAMP or in presence of GAF. A further effect of a dynamic, direct interaction is that interaction partners tend to enrich around their shared chromatin binding site. This increased local concentration would implement the next level of cooperativity, namely between two or more MREs within HAS.

Competition and cooperativity explained by an extended DNA shape signature

DNA shape parameters are defined for B-form DNA and *a priori* it cannot be assumed that the DNA adopts the same conformation in a chromatin context. Nevertheless, computing DNA shape parameters improves the prediction of physiological chromatin binding profiles for many TFs (3). We found the PionX signature prevailed in chromatin and is recognized by MSL2. For functional MREs in general, we observed a more elaborate shape signature. Remarkably, MSL2 only bound this signature in the presence of GAF and CLAMP. GAF led to a refinement of the MSL2/CLAMP profile by occluding decoy sites. To our knowledge this is the first description of an extrinsic, sculpturing effect on global binding site selectivity.

GAF only competes for sites that contain its minimal binding sequence GAGAG (23,64). A good MSL2 binding site should, therefore, not contain this sequence. The PWM of PionX sites formally allows sequence combinations that constitute GAGAG (Supplementary Figure

S4A). Our shape analysis of sites selected under various conditions revealed that good PionX sites and strong MREs do not show 'GA' at positions 12/13 or 17–19 of the motif. This information cannot be derived from the PWM, which is blind to nucleotide succession. The analysis of Propeller Twist nicely illustrates how 'GAG' is avoided between positions 17–19 of the PionX signature. Furthermore, monitoring Roll reveals a localized variability of this parameter at motif position 12/13 that is matched by MSL2 only in the presence of GAF. Conceivably, at this position it does not matter which nucleotide occurs, provided it is not GA.

Outlook

Our approach of cell-free genomics bears significant potential to study the self-assembly of chromosome folding and nuclear structures. Immediate applications relate to targeting and spreading phenomena of transcription factors in chromatin, of which dosage compensation is only one example. It will be interesting to explore to which extent local structural chromosomal heterogeneity and the process of chromatin programming at the maternal-to-zygotic transition can be reconstructed in a cell-free system lacking the confines of a nucleus. For the specific question of dosage compensation, we anticipate that reconstituting the complete 6-subunit DCC, including roX RNA, will yield further insights into the targeting of MSL2.

DATA AVAILABILITY

The raw sequencing files in fastq format and the summarized genome browser tracks in bigwig format are available in the GEO database at GSE169222 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE169222>). Custom code was deposited on Github: <https://github.com/nikolas848/egggers2021>

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

ACKNOWLEDGEMENTS

We thank S. Krause for baculovirus expression and her help in cloning and protein purifications, A. Campos-Sparr and A. Zabel for help in preparing the sequencing libraries. We thank A. Thomae from the Core Facility Bioimaging of the Biomedical Center and A. Scacchetti for help with microscopy and S. Krebs from the LAFUGA Genomics Facility for next-generation sequencing. We thank T. Straub and T. Schauer from the Core Facility Bioinformatics of the Biomedical Center for scripts and advice, P. Korber, M. Müller, R. Villa and T. Straub for comments on the manuscript and all members of the Becker lab (particularly S. Baldi) for discussions throughout the project. We are grateful to Paul Badenhorst (Institute of Cancer and Genomic Sciences, Birmingham) for a kind gift of NURF-301 antibodies.

Author Contributions: N.E. and P.B.B. conceived experiments. N.E. performed experiments. 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

- Suter, D.M. (2020) Transcription factors and DNA play hide and seek. *Trends Cell Biol.*, **30**, 491–500.
- Lionnet, T. and Wu, C. (2021) Single-molecule tracking of transcription protein dynamics in living cells: seeing is believing, but what are we seeing? *Curr. Opin. Genet. Dev.*, **67**, 94–102.
- Mathelier, A., Xin, B., Chiu, T.P., Yang, L., Rohs, R. and Wasserman, W.W. (2016) DNA shape features improve transcription factor binding site predictions in vivo. *Cell Syst.*, **3**, 278–286.
- Yang, L., Orenstein, Y., Jolma, A., Yin, Y., Taipale, J., Shamir, R. and Rohs, R. (2017) Transcription factor family-specific DNA shape readout revealed by quantitative specificity models. *Mol. Syst. Biol.*, **13**, 910.
- Mirny, L.A. (2010) Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. USA*, **107**, 22534–22539.
- Morgunova, E. and Taipale, J. (2017) Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.*, **47**, 1–8.
- 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.
- Voelker-Albert, M.C., Pusch, M.C., Fedisch, A., Schilcher, P., Schmidt, A. and Imhof, A. (2016) A quantitative proteomic analysis of in vitro assembled chromatin. *Mol. Cell. Proteomics*, **15**, 945–959.
- Längst, G. and Becker, P.B. (2001) Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J. Cell Sci.*, **114**, 2561–2568.
- 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.
- Hamm, D.C. and Harrison, M.M. (2018) Regulatory principles governing the maternal-to-zygotic transition: insights from *Drosophila melanogaster*. *Open Biol.*, **8**, 180183.
- 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.
- 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.
- 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.
- 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.
- Larschan, E., Soruco, M.M., Lee, O.K., Peng, S., Bishop, E., Chery, J., Goebel, K., Feng, J., Park, P.J. and Kuroda, M.I. (2012) Identification of chromatin-associated regulators of MSL complex targeting in *Drosophila* dosage compensation. *PLoS Genet.*, **8**, e1002830.
- Wilkins, R.C. and Lis, J.T. (1998) GAGA factor binding to DNA via a single trinucleotide sequence element. *Nucleic Acids Res.*, **26**, 2672–2678.
- 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.
- Soruco, M.M., 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.

20. Urban, J.A., Urban, J.M., Kuzu, G. and Larschan, E.N. (2017) The Drosophila CLAMP protein associates with diverse proteins on chromatin. *PLoS One*, **12**, e0189772.
21. Fuda, N.J., Guertin, M.J., Sharma, S., Danko, C.G., Martins, A.L., Siepel, A. and Lis, J.T. (2015) GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. *PLoS Genet.*, **11**, e1005108.
22. Adkins, N.L., Hagerman, T.A. and Georgel, P. (2006) GAGA protein: a multi-faceted transcription factor. *Biochem. Cell. Biol.*, **84**, 559–567.
23. 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.
24. Greenberg, A.J., Yanowitz, J.L. and Schedl, P. (2004) The Drosophila GAGA factor is required for dosage compensation in males and for the formation of the male-specific-lethal complex chromatin entry site at 12DE. *Genetics*, **166**, 279–289.
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 of the Dosage Compensation Complex to the X chromosome. *Nucleic Acids Res.*, **38**, 3209–3221.
27. Wall, G., Varga-Weisz, P.D., Sandaltzopoulos, R. and Becker, P.B. (1995) Chromatin remodeling by GAGA factor and heat shock factor at the hypersensitive Drosophila hsp26 promoter in vitro. *EMBO J.*, **14**, 1727–1736.
28. 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**, e52482.
29. 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.
30. Strutt, H., Cavalli, G. and Paro, R. (1997) Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.*, **16**, 3621–3632.
31. 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 Bioinform.*, **17**, 419.
32. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
33. 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.
34. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
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.
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. Chiu, T.P., Comoglio, F., Zhou, T., Yang, L., Paro, R. and Rohs, R. (2016) DNashapeR: an R/Bioconductor package for DNA shape prediction and feature encoding. *Bioinformatics*, **32**, 1211–1213.
43. Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M. and Becker, P.B. (1997) Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature*, **388**, 598–602.
44. Längst, G., Bonte, E.J., Corona, D.F. and Becker, P.B. (1999) Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell*, **97**, 843–852.
45. Tsukiyama, T. and Wu, C. (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell*, **83**, 1011–1020.
46. Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R. and Kadonaga, J.T. (1997) ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell*, **90**, 145–155.
47. Becker, P.B. (2002) Nucleosome sliding: facts and fiction. *EMBO J.*, **21**, 4749–4753.
48. Erdel, F. and Rippe, K. (2018) Formation of chromatin subcompartments by phase separation. *Biophys. J.*, **114**, 2262–2270.
49. Gibson, B.A., Doolittle, L.K., Schneider, M.W.G., Jensen, L.E., Gamarra, N., Henry, L., Gerlich, D.W., Redding, S. and Rosen, M.K. (2019) Organization of chromatin by intrinsic and regulated phase separation. *Cell*, **179**, 470–484.
50. Maeshima, K., Rogge, R., Tamura, S., Joti, Y., Hikima, T., Szerlong, H., Krause, C., Herman, J., Seidel, E., DeLuca, J. *et al.* (2016) Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. *EMBO J.*, **35**, 1115–1132.
51. Kuzu, G., Kaye, E.G., Chery, J., Siggers, T., Yang, L., Dobson, J.R., Boor, S., Bliss, J., Liu, W., Jögl, G. *et al.* (2016) Expansion of GA dinucleotide repeats increases the density of CLAMP binding sites on the X-chromosome to promote Drosophila dosage compensation. *PLoS Genet.*, **12**, e1006120.
52. Urban, J., Kuzu, G., Bowman, S., Scruggs, B., Henriques, T., Kingston, R., Adelman, K., Tolstorukov, M. and Larschan, E. (2017) Enhanced chromatin accessibility of the dosage compensated Drosophila male X-chromosome requires the CLAMP zinc finger protein. *PLoS One*, **12**, e0186855.
53. 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.
54. Ramirez, F., Lingg, T., Toscano, S., Lam, K.C., Georgiev, P., Chung, H.R., Lajoie, B.R., de Wit, E., Zhan, Y., de Laat, W. *et al.* (2015) High-affinity sites form an interaction network to facilitate spreading of the MSL complex across the X chromosome in Drosophila. *Mol. Cell*, **60**, 146–162.
55. Tatei, K., Kimura, K. and Ohshima, Y. (1989) New methods to investigate ATP requirement for pre-mRNA splicing: inhibition by hexokinase/glucose or an ATP-binding site blocker. *J. Biochem.*, **106**, 372–375.
56. Ner, S.S. and Travers, A.A. (1994) HMG-D, the Drosophila melanogaster homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histone H1. *EMBO J.*, **13**, 1817–1822.
57. Climent-Canto, P., Carbonell, A., Tatarski, M., Reina, O., Bujosa, P., Font-Mateu, J., Bernues, J., Beato, M. and Azorin, F. (2020) The embryonic linker histone dBigH1 alters the functional state of active chromatin. *Nucleic Acids Res.*, **48**, 4147–4160.
58. Lu, X., Wontakal, S.N., Emelyanov, A.V., Morcillo, P., Konev, A.Y., Fyodorov, D.V. and Skoultchi, A.I. (2009) Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. *Genes Dev.*, **23**, 452–465.
59. Zhou, B.R. and Bai, Y. (2019) Chromatin structures condensed by linker histones. *Essays Biochem.*, **63**, 75–87.
60. Sandaltzopoulos, R., Blank, T. and Becker, P.B. (1994) Transcriptional repression by nucleosomes but not H1 in reconstituted preblastoderm Drosophila chromatin. *EMBO J.*, **13**, 373–379.
61. 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.
62. Blank, T.A. and Becker, P.B. (1995) Electrostatic mechanism of nucleosome spacing. *J. Mol. Biol.*, **252**, 305–313.

63. Katsani, K.R., Hajibagheri, M.A. and Verrijzer, C.P. (1999) Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *EMBO J.*, **18**, 698–708.
64. Omichinski, J.G., Pedone, P.V., Felsenfeld, G., Gronenborn, A.M. and Clore, G.M. (1997) The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat. Struct. Biol.*, **4**, 122–132.
65. Tsukiyama, T., Becker, P.B. and Wu, C. (1994) ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature*, **367**, 525–532.
66. Zhou, X. and O'Shea, E.K. (2011) Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. *Mol. Cell*, **42**, 826–836.
67. 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.
68. 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.
69. Villa, R., Forne, I., Muller, 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.
70. 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.
71. 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.
72. 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.*, <http://www.genesdev.org/cgi/doi/10.1101/gad.348411.121>.
73. 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.
74. Kelley, R.L., Meller, V.H., Gordadze, P.R., Roman, G., Davis, R.L. and Kuroda, M.I. (1999) Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell*, **98**, 513–522.
75. Oh, H., Park, Y. and Kuroda, M.I. (2003) Local spreading of MSL complexes from roX genes on the *Drosophila* X chromosome. *Genes Dev.*, **17**, 1334–1339.
76. Judd, J., Duarte, F.M. and Lis, J.T. (2021) Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. *Genes Dev.*, **35**, 147–156.
77. Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S. and Natoli, G. (2014) Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol. Cell*, **54**, 844–857.
78. Maier, V.K., Chioda, M., Rhodes, D. and Becker, P.B. (2008) ACF catalyses chromatosome movements in chromatin fibres. *EMBO J.*, **27**, 817–826.
79. Clausell, J., Happel, N., Hale, T.K., Doenecke, D. and Beato, M. (2009) Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF. *PLoS One*, **4**, e0007243.
80. Tsukiyama, T., Daniel, C., Tamkun, J. and Wu, C. (1995) ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell*, **83**, 1021–1026.
81. Wei, K.H., Grenier, J.K., Barbash, D.A. and Clark, A.G. (2014) Correlated variation and population differentiation in satellite DNA abundance among lines of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*, **111**, 18793–18798.
82. Raff, J.W., Kellum, R. and Alberts, B. (1994) The *Drosophila* GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. *EMBO J.*, **13**, 5977–5983.
83. 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.
84. Granok, H., Leibovitch, B.A., Shaffer, C.D. and Elgin, S.C. (1995) Chromatin. Ga-ga over GAGA factor. *Curr. Biol.*, **5**, 238–241.
85. Sönmezer, 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.