DNA-sequence dependent positioning of the linker histone in a chromatosome: a single-pair FRET study

Madhura De[1,2,3], Mehmet Ali Öztürk[2,4], Sebastian Isbaner[1,7], Katalin Tóth[1*], Rebecca C. Wade[2,3,5,6*]

[1]Department of Biophysics of Macromolecules, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

[2] Molecular and Cellular Modelling Group, Heidelberg Institute for Theoretical Studies (HITS), 69118 Heidelberg, Germany

[3] Faculty of Biosciences, Heidelberg University, 69120 Heidelberg, Germany

[4]Centre for Biological Signalling Studies (BIOSS) and Centre for Integrative Biological Signalling Studies (CIBSS), University of Freiburg, 79104 Freiburg, Germany

[5]Center for Molecular Biology (ZMBH), DKFZ-ZMBH Alliance, Heidelberg University, 69120 Heidelberg, Germany

[6]Interdisciplinary Center for Scientific Computing (IWR), 69120 Heidelberg, Germany

[7]Present address: Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Campus-Vienna-Biocenter 1, 1030 Vienna, Austria

*To whom correspondence should be addressed. Emails: Katalin Tóth: dr.toth.katalin@med.unideb.hu , Rebecca C. Wade: rebecca.wade@h-its.org

Abstract:

The linker histone (LH) associates with the nucleosome with its globular domain (gH) binding in an on or off-dyad binding mode. The positioning of the LH may play a role in the compaction of higher-order structures of chromatin. Preference for different binding modes has been attributed to the LH's amino acid sequence. We here study the effect of the linker DNA (L-DNA) sequence on the positioning of a full-length LH, *Xenopus laevis* H1.0b, by employing single-molecule FRET spectroscopy. Chromatosomes were fluorescently labelled on one of the two 40bp long L-DNA arms, and on the gH. We varied 11bp of DNA flanking the core (non-palindromic Widom 601) of each chromatosome construct, making them either A-tract, purely GC, or mixed, with 64% AT. The gH consistently exhibited higher FRET efficiency with the L-DNA containing the A-tract, than that with the pure-GC stretch, even when the stretches were swapped. However, it did not exhibit higher FRET efficiency with the L-DNA containing 64% AT-rich mixed DNA, compared to the pure-GC stretch. We explain our observations with a FRET-distance restrained model that shows that the gH binds on-dyad and that two arginines mediate recognition of the A-tract via its characteristically narrow minor groove.

Keywords: chromatosome, linker histone, nucleosome, single molecule-FRET, linker DNA sequence, A-tract, minor groove recognition

Introduction:

The genetic material of cells consists of nucleoprotein polymers which constitute chromatin, the smallest repeat unit of which is the chromatosome. The chromatosome comprises of a nucleosome – a core histone octamer wrapped by 1.65 turns (or ca. 150 bp) of the DNA, flanked by 20-60 bp long linker-DNA (L-DNA) on either side, and an additional protein, the linker histone (LH). Numerous cell-type and cell-cycle stage specific LHs exist (1), (2). They all have a length of about 200 amino acid residues and are highly positively charged. The LH has a tripartite structure. They have a short, flexible, N-terminal domain, a highly conserved and structured globular domain (gH) of about 70 to 80 residues, and a long, disordered C-terminal domain (CTD or 'tail'). Being highly positively charged, the LH, along with the core histones, is well-suited to neutralizing the negatively charged, mutually repulsive DNA strands and aiding in the packaging of the chromatin. Indeed, the major function of the LH is chromatin condensation (3) and it has been observed to play a crucial role in the formation of higher-order chromatin structures (4). LH mediated compaction of chromatin mediates silencing of genes and repetitive sequences (5), (6).

The location of the LH on the nucleosome has been widely studied both experimentally and computationally (7), (8). X-ray crystallographic studies of the *Gallus gallus* H5 (9),(10) and on *Xenopus laevis* H1.0b (11) in association with nucleosomes have shown an 'on-dyad' binding mode in which the gH binds at the centre of the stretch of core or ncp DNA (stands for nucleosome core particle or the part of DNA wrapped around the octamer) bounded on either side by the entering and exiting linker-DNA (L-DNA). A second possible binding mode, referred to as 'off-dyad', has been observed by NMR (12) and cryo-EM (13) where the gH binds in the region bounded by the dyad axis (the axis passing through the nucleosome, equidistant from both entry-exit points) and the entering or exiting L-DNA. Apart from these two observed locations, the gH has also been observed to adopt different orientations on the nucleosome (9, 11–13). As pointed out from Brownian dynamics simulations by Öztürk *et. al.* (14), the gH can associate with nucleosomes in a range of configurations, that depend upon the sequences of the gH and the DNA and the presence of various post-translational modifications.

It is important to note that these binding modes and orientations of the LH refer specifically to the gH and that the CTD is highly flexible. Using FRET, Caterino and Hayes (15) demonstrated that the highly disordered CTD develops local structuring in the presence of nucleosomes. Fang and colleagues (16) further demonstrated that the ordering of the CTD in the presence of mononucleosomes is distinct from that observed in the presence of nucleosomal arrays. Cryo-EM structural studies (11, 53) showed that while the gH was on-dyad, the CTD was associated with one of the L-DNA arms. The local ordering of the CTD in the presence of nucleosomes could be a result of electrostatic interactions of the negatively charged linker-DNA and the

positively charged CTD, as observed by Luque *et. al.* (17), using mesoscale modeling at different salt concentrations. However, using nanosecond FCS measurements, Heidarsson *et. al.* (18) demonstrated that the CTD remains highly disordered even on binding mononucleosomes.
In-depth experimental and computational studies have pin-pointed the amino acid residues in the gH contributing to DNA binding (8, 19–21) and to the two different binding modes, on- and off-dyad (22). But the possible contributions of the linker DNA (L-DNA) sequence to the LH binding mode and orientation have largely remained unclear.
Although the LH is not reported to have any consensus DNA binding sequence, preferences for certain sequences have been reported. A number of in-vitro studies on the binding of LH to naked DNA (23, 24) showed an affinity of the LH to AT-rich sequences. Evidence of the LH having an affinity for AT- and A-tract rich scaffold attachment regions (SAR) (25) and to satellite sequences (26) led to the proposition that the LH has an affinity for not just AT-rich sequences but specifically for A-tract regions (sequences with a stretch of 4 to 6 or more adenine bases present in SARs and satellite sequences (27)). This was further proven using binding assays by Roque *et.al.* (28) who showed that LH was selectively bound to A-tract rich SARs in the presence of competing DNA sequences having 75% AT but lacking in A-tracts. Recent TEM images and sedimentation studies on A-tract containing nucleosomal arrays also showed strong compaction in the presence of LH (29).
The affinity of LH for AT-rich sequences has been attributed to both the gH, and also solely to the CTD alone (28).
The biological relevance of the LH's preference for AT-rich or A-tract sequences was pointed out in the review by Zlatanova *et. al.* (23) who stated that a preference of the LH for AT-rich sequences (including A-tracts) might aid in the LH-mediated compaction of AT-rich heterochromatin. On the other hand, while the LH has shown an affinity for A-tract sequences, the opposite has been observed for GC-rich sequences (26, 30, 53).
The aim of the current study is to observe the possible effect of the linker-DNA sequence on the positioning and orientation of the LH in the chromatosome, and to pin-point the underlying reason for the affinity of the LH towards specific DNA sequences such as A-tracts. We used single-molecule FRET to probe the location and orientation of the full-length *Xenopus laevis* LH, H1.0b, on mononucleosomes reconstituted from non-palindromic Widom 601 sequences. We modified a stretch of 11 bp on each linker-DNA, adjacent to the points where the DNA enters and exits the nucleosome (the 'entry-exit' sites). Since these stretches flank the nucleosome core, we refer to them as 'flanks'. We study the effects of a pure GC stretch, a mixed sequence (64% AT-rich), and A-tracts pairing with thymine as well as deoxy-uridine, on the position and orientation of the LH. Based on FRET-derived distances, we come to a possible suggestion as to how A-tracts are recognized by the LH. Our model shows that the gH of the LH is oriented towards A-tracts in such a way as to position arginines 42 and 94 close to the narrow minor groove of A-tracts. Recognition of A-tracts via their characteristic narrow minor grooves by proteins containing positively charged amino acid residues such as arginines, has been well

documented (27, 31, 32). We suggest that this very same mechanism could be responsible for the specific recognition of A-tracts by the LH.

Materials and Methods:

Preparation of labelled DNA

226 bp unlabelled and singly-labelled DNA was used for reconstituting mononucleosomes. Labelled DNA was produced using PCR, from either pGEM3z vector (construct MG, see Supplementary Information S2 for sequences), or from 226 bp templates (Biolegio, Nijmegen, The Netherlands), containing the strongly positioning sequence, Widom 601 (33). Primers used to amplify the DNA (IBA Lifesciences, Goettingen, Germany) contained the fluorophore Alexa 594, covalently attached via an amino-dT C6 linker. For the TU construct, the stretch of 11 deoxy-uridines was incorporated in the reverse primer (Biolegio, Nijmegen, The Netherlands), (see Supplementary Information S2 for sequences). PCR was performed using Taq-polymerase containing master mix (Thermo Scientific) and products were purified on Gen-Pak FAX HPLC (Waters, USA). The final 226 bp product consisted of the highly positioning, non-palindromic Widom 601 sequence at the centre, bounded by two equally long (~40 bp) L-DNA. The fluorophores were attached 20 bp from the ends of the DNA, on one or the other L-DNA arm (Fig. 1a). The DNAs were purified from PCR products using Gen-Pak FAX HPLC (Waters) and gel electrophoresis was performed to check for the correct size of the product.
The central base pair of the DNA was numbered 0. Bases to the left were assigned negative numbers and bases to the right were assigned positive numbers. The labelling positions on the L-DNA arms were either +94 or -93. Based on the flank sequences (i.e. a stretch of 11 bp flanking the central non-palindromic Widom 601), the 226 bp DNAs were named AG (A-tract minus flank, purely GC plus flank), GA (purely GC minus flank, A-tract plus flank), MG (mixed sequence minus flank, purely GC plus flank), GM (purely GC minus flank, mixed sequence plus flank), and TU (A(T)-tract minus flank, A(U)-tract plus flank) (for sequences, see Supplementary Information S2). Chromatosomes reconstituted from these DNA were named accordingly.

Preparation and purification of LH:

Full-length, recombinant LH from *Xenopus laevis*, H1.0b, was used for the study. Unlabeled, wild-type LH was used with either unlabeled chromatosomes or chromatosomes labelled only on the DNA (acceptor-only chromatosomes used to calculate spectral overlap, as described in Supplementary Information S1). We labelled the LH with Alexa 488, the donor fluorophore, either on the C-terminal 'tail' domain (CTD), or on the globular domain (gH). For labelling, cysteine residues needed to be incorporated by site-directed mutagenesis, in the cysteine-free,

wild-type LH. The plasmid for the gH label (threonine 77 mutated to cysteine, or T77C), and the protein for the CTD label (G101C) were kindly provided by Prof. Jeffrey Hayes. The plasmid containing the T77C gene was transformed into and amplified in *E.coli* XL-1 Blue competent cells. Amplified plasmids were purified and retransformed to *E.coli* BL21(DE3) cells for protein expression, with IPTG induction as described by (34). The bacterial pellet was sonicated (Branson Sonifier 250) in a lysis buffer (50mM Tris-HCl, pH 7.5, 100mM KCl, 1mM EDTA, 1mM phenylmethanesulphonylfluoride, 0.1% volume/volume Nonidet P-40). After multiple rounds of washing the bacterial pellet, the supernatant was concentrated and simultaneously buffer exchanged to an unfolding buffer (7M guanidium hydrochloride, 20mM Tris.HCl, pH 7.5), using Vivaspin 2 columns (10 kDa molecular weight cutoff, Sartorius). The resulting, concentrated protein solution was passed through a size exclusion chromatography column, Sephacryl S-200 (pre-equilibrated by SAU-1000 buffer: 7M deionized urea, 20mM sodium acetate, pH 5.2, 1mM EDTA, 1M KCl), at a flow rate of 3ml/min. The eluted protein fractions were concentrated and simultaneously buffer exchanged to SAU-50 buffer (7M deionized urea, 20mM sodium acetate, pH 5.2, 1mM EDTA, 50mM KCl) using Vivaspin 20 columns (10 kDa molecular weight cutoff, Sartorius). This was followed by cation exchange chromatography using Mono-S-HR 10/10 FPLC column pre-equilibrated with SAU-50 and SAU-1000 buffers. The protein was eluted with SAU-1000 buffer. Protein eluate was concentrated and refolded by overnight dialysis at 4°C into 2M NaCl, 10mM Tris-HCl, 0.1mM EDTA, pH 7.5.
Wild type, recombinant, *X. laevis* core histones were purchased from Planet Protein (Colorado State University, USA), assembled into octamers and purified as described (35, 36).

LH labelling

The LH was labelled with Alexa 488 C5 maleimide on the cysteine residue either on the gH or CTD (see Supplementary Information S3 for labelled protein gels). Unlabelled, cysteine-containing protein was unfolded in a denaturing buffer containing 7M guanidine hydrochloride, 10mM dithiothreitol, 20mM Tris-HCl, at pH 7.5. Tris (2-carboxyethyl) phosphine was added as a reducing agent, at a concentration ten times higher than the protein concentration, to prevent the formation of protein dimers through a disulfide bridge. Alexa 488 C5 maleimide dye (Thermo Fischer Scientific) was dissolved in N'N'-dimethyl formamide, and added in three steps every hour, to the protein shaking in the dark at room temperature, to achieve a final dye concentration of ten times the number of moles of protein. This mix was stored at 4ºC overnight. The maleimide reaction was stopped on the following day, by adding L-cysteine at the same concentration as the dye molecules. The labelled protein was dialysed into an unfolding buffer (urea 7M, sodium acetate 20mM, pH 5.2, EDTA 1mM, KCl 1M). To remove excess, unbound dyes, the protein was passed through a sephacryl S200 column. Finally, the protein was refolded by multiple dialysis steps in 2M NaCl 1xTE (10mM Tris-HCl, 0.1 mM EDTA) buffer (pH 7.5).

Chromatosome reconstitution

Chromatosomes were reconstituted from recombinant *Xenopus laevis* core histones, 226 bp labelled or unlabelled DNA, and full length, recombinant, *Xenopus laevis* LH that were either labelled or unlabelled. Reconstitution involved reducing the salt concentration by a two-step dialysis method, using Slide-A-Lyzer MINI Dialysis unit (Thermo Fischer) having a cut-off of 7 kDa. DNA and core histone octamers were mixed in a molar ratio of 1:1.65 (36) in 2M NaCl 1xTE, pH 7.5. The first step of the dialysis, performed for four hours at 4ºC, was against 0.6M NaCl 1xTE. This lowered the salt concentration of the solution containing DNA and core histones from 2M to 0.6M, enabling reconstitution of mononucleosomes. The LH was added at this stage (37), at a molar ratio of 1.6:1 mole DNA (38), taking care to have a dye stoichiometric ratio (measured in ALEX, see section Single-molecule FRET and Supplementary Information S1) of 1:1 (S=0.5) donor (on LH) and acceptor (on DNA), while avoiding aggregation due to excess of LH. Dialysis was performed against 1xTE without NaCl, at 4ºC for overnight, until the salt concentration of the DNA and protein mix was lowered to 25mM. All single-molecule and bulk spectroscopic measurements were performed at 25mM NaCl in TE buffer. Chromatosome formation was checked by 6% polyacrylamide gel electrophoresis (acrylamide/bisacrylamide 60:1) (see Supplementary Information S4). Single labeled chromatosomes such as donor-only (Alexa 488 label on the LH) and acceptor-only (Alexa 594 label on the DNA, the LH remains unlabeled) samples were reconstituted in the same way as double labelled samples.

Single-molecule FRET

Single-molecule FRET with Alternating Laser Excitation (ALEX) (39, 40) experiments were performed on freely-diffusing, doubly labelled chromatosomes, using our in-house instrument described in (41) that was further modified (https://doi.org/10.11588/heidok.00022135) by the inclusion of the acceptor laser (561 nm, Cobolt Jive, Hübner Photonics) and an acousto-optic tunable filter (AOTFnC-VIS-TN, AA Optoelectronics) to switch between the two wavelengths (491nm, Cobolt Calypso, Hübner Photonics, and 561nm) on a microsecond timescale. The laser intensity at the objective, for measuring at single-molecule levels, was 40µW. The light collected from the sample comprises a mix of fluorescence emitted from the donor and acceptor fluorophores. The emitted beam is split using a dichroic mirror (600DCXR, Omega Optical, Brattleboro, USA) to donor and acceptor signals which are sent through filters (donor channel: 520df40, acceptor channel: 610ALP, Omega Optical) to two avalanche photodiodes (SPAD-AQ-14, PerkinElmer Optoelectronics, Boston, Massachusetts, USA), see Supplementary Information S1 for a schematic of the setup adapted from (https://doi.org/10.11588/heidok.00022135). The instrumentation is described in detail in (https://doi.org/10.11588/heidok.00022135). Although we obtained both the stoichiometry (proportion of the donor and acceptor dyes) and the proximity ratio by this procedure (see Supplementary Information S1 and (40)), we extracted the proximity ratios of only the double labelled samples, i.e., having a stoichiometry range of 0.25 to 0.75.

Proximity ratio is the ratio of the number of photons detected in the acceptor channel divided by the total number of photons detected in both the donor and acceptor channels (https://doi.org/10.11588/heidok.00008342). Proximity ratio histograms were created using our in-house AlexEval software (https://doi.org/10.11588/heidok.00022135) having 50 bins, with a bin width of 0.02, and ranging from 0 to 1. Multiple Gaussian peaks were fitted to the proximity ratio histograms (see Supplementary Information S5). The proximity ratio is related to the FRET efficiency E, via an instrumentation detection factor γ (Proximity ratio = FRET efficiency when γ = 1), that we measured prior to sample measurements, using two FRET standards: 30 bp oligos labelled with Alexa 488 and Alexa 594 at a separation of 10 and 21 bp (Supplementary Table T2, γ factor calculation is described in (40)).
Using bulk fluorescence and absorption measurements, we measured a Förster distance, R_0, of 52 Å, assuming a refractive index of 1.4 (see Supplementary Information S1, Supplementary Table T1).
From the FRET efficiency, using R_0, we obtained inter-fluorophore distances (Supplementary Information S1), which were used to model the chromatosomes AG, GA, and TU.
For measurements at single-molecule levels, we used 50 to 100 pM of doubly-labelled samples, and added unlabelled samples to raise the total concentration of the samples to 300pM. This procedure avoided spontaneous dissociation of the reconstituted chromatosomes (41). Measurements were done for 30 minutes in 25 mM NaCl 1xTE buffer, pH 7.5. The sample solution also contained 0.8 mM ascorbic acid for scavenging free radicals to prevent photobleaching of the dyes as described previously (35), and 0.01 mM nonidet-P40 (43) to stabilize the mononucleosomes.

Modelling of the AG, GA and TU chromatosome structures

Distance information between the dyes on the gH and the L-DNA arms was used to generate chromatosome models. Although we used a full-length LH for our experiments, we modelled only the gH. The N-terminal domain does not have any label and thus its position cannot be determined. Moreover, the N and C-terminal domains are both intrinsically disordered. Thus, the CTD, despite having a label, cannot be accurately modelled. Similarly, despite using 226 bp DNA for our experiments, we modelled only 193 bp of DNA, omitting 16-17 bp from both the ends, but taking care to include the two labelling positions on the DNA, 3 to 4 bp from the new ends. This was done because the distance information obtained from the DNA labels cannot be used to define the positions of the parts of the L-DNA arms extending beyond the labels.
Chromatosomes AG, GA, and TU were modelled using the 193 bp chromatosome structure with PDB id: 5nl0 (11) as the starting structure. We removed the associated gH and mutated the DNA sequence to that of our construct using Chimera (44). The gH was separately mutated to include the cysteine residue at position 77. Using Chimera, we repositioned the gH on the nucleosome, generating a range of conformations based on varying orientations of the gH, for each of the chromatosome constructs. The Fluorescence Positioning and Screening (FPS) (42, 45) software

was used to screen the structures using experimentally derived inter-dye distances as the input. Screened structures that satisfied the FRET-derived distances, were protonated in Chimera, considering hydrogen bonds, and energy minimisation of the protonated structures was performed. Keeping the core histone octamers fixed, we minimised the DNA and the gH, in Chimera, using the ff99bsc0 (46) force field, and in subsequent steps, minimised only the gH using the ff14SB force field (47). 100 steps of steepest descent energy minimization were performed at each stage to remove clashes. This was followed by 10 conjugate gradient steps. After energy minimisation, the FRET distances were recalculated by accessible volume simulation of the fluorophores, using FPS (for dye parameters, see Supplementary Table T5). For the TU constructs, the AG nucleosome was used as a starting structure. The GC flank was mutated in Chimera (44) to an A-tract complementary to deoxy-uridine. The gH was positioned and fluorophores were built using FPS (42, 45), in the same way as for the AG and GA constructs. Using a R_0 value of 52Å, the theoretical FRET efficiency was calculated between the gH dye and the L-DNA dyes with the FPS software. To compare with experimental results, we converted the proximity ratio data to FRET efficiencies, assuming a γ factor of 1.

Results:

The *Xenopus laevis*, H1.0b LH globular domain show a preference for A-tracts

In this study, we asked how the L-DNA sequence may affect the location and/or orientation of the LH. Does the LH prefer an AT-rich L-DNA arm over a GC-rich arm? Does the LH specifically prefer A-tracts (one strand purely adenines and the complementary strand purely thymines) or simply any sequence having a higher AT content compared to GC?
To address these questions, we constructed two types of chromatosomes: in the AG construct, the minus flank had an A-tract and the plus flank was purely GC (Table 1). In the MG construct, the minus flank A-tract was replaced by a mixed DNA sequence that had an AT content of 64%, while the plus flank was purely GC, and the same as the plus flank of AG (Table 1). In the GA and the GM constructs, the minus and the plus flanks of AG and MG, respectively, were swapped.
The gH of the full-length LH was labelled with the donor (Alexa 488) dye at residue 77 (threonine mutated to cysteine prior to labelling), at the C-terminal end of helix α3, see Fig1B. The DNA was labelled with the acceptor (Alexa 594) dye on either the minus arm (-93 T) or on the plus arm (+94 T) (Fig 1A). The gH dye was observed to show a higher proximity ratio with the minus dye in the AG construct (and with the plus dye in the GA construct (Fig 2A and B). Accordingly, the plus dye of AG and the minus dye of GA showed a lower proximity ratio (see Supplementary Information S5 i-iv for the individual peaks). We also observed a high proximity ratio peak (mean proximity ratio = 0.9) for the minus dye in the GA construct. However, the intermediate peak (mean proximity ratio = 0.58, Supplementary Table T3) was the dominant population, comprising ~75% of the total area under the curve.

The distances of the major populations were derived from the proximity ratio histograms by a multi-step procedure based on both bulk and single-molecule spectroscopic measurements as described in Materials and Methods and Supplementary Information S1. We used a Förster distance of 52 Å to calculate the distances (see Supplementary S1 and T1).
We found that in the AG construct, the gH dye is 45 ± 1.4 Å away from the -93 dye and 50 ± 0.2 Å away from the +94 dye (Fig 2A and Supplementary Information S6 and Table T4 for replicates). In the GA construct, the gH is 41 ± 0.5 Å away from +94 dye and 48.4 ± 0.4 Å away from -93 dye (Fig 2B). These data show that the labelled site on the gH of the full-length LH, is in both systems closer to the A-tract (minus flank of AG and plus flank of GA chromatosome) than the purely GC flanks (plus flank of AG, and minus flank of GA).
For the MG construct, the gH dye is observed to show higher FRET for the +94 dye (Fig. 2D), than for the -93 dye, suggesting that the gH dye is closer to the 100% GC flank plus L-DNA arm than the 36% GC/64% AT minus flank. This observation is at odds with previous works (26, 30, 53) reporting that the LH has less affinity for GC-rich regions as compared to AT-rich regions, and led us to test this observation further. We thus built the GM construct, with the 100% GC flank and the mixed 36% GC/64% AT flank swapped. However, in the GM construct, the FRET profile of the gH dye is the same with the 100% GC flank minus arm and the 36% GC / 64% AT flank containing plus arm (Fig. 2D). The high FRET efficiency peak shown for the GC-flank in the MG construct could be due to the orientation of the gH.
The observations for the MG and GM constructs suggest that (a) the gH does not have a preference for pure GC containing regions over a mixed DNA sequence, and (b) the gH does not have any preference for mixed DNA sequences even if they have a higher AT content than GC. However, as shown by the AG and GA constructs, when the mixed, 36% GC/64% AT containing sequence of MG and GM is replaced by an A-tract, the gH consistently exhibits a higher proximity ratio with the A-tract flank. We next wanted to investigate the possible reason behind the preference of the LH for A-tract regions.

The *Xenopus laevis*, H1.0b LH globular domain shows similar FRET for A-tracts with thymines and with deoxy-uridines

It was proposed by Cui *et.al.* (48) that the affinity of gH to sequences with high AT content could be a result of hydrophobic interactions between the thymine methyl groups on the DNA and a hydrophobic patch on the gH, the GVGA motif in the 'wing' domain. This interaction was observed in MD simulations of a chromatosome with a gH in the off-dyad binding mode (49), showing that the GVGA motif interacted with the thymine methyl groups exposed on the major groove of the DNA. Our LH, subtype H1.0b, contains the GVGA motif on the wing domain (see Supplementary Information S2 for sequence). Why does it interact only with the A-tract, but not with the 64% AT containing, mixed sequence? To investigate this phenomenon experimentally, we synthesized the TU construct, where both the plus and minus flanks have A-tracts. However, the minus flank was base paired to dT, and the plus flank was base paired to deoxy-uridine. The

reason for using deoxy-uridine was that dU lacks the C5 methyl group present in thymine. Thus, the TU minus flank is decorated with methyl groups, that are lacking in the plus flank.
We observed that the proximity ratio distributions for the gH dye with the plus and minus dyes closely overlap (Fig. 2C). The lowest peak (mean Proximity ratio = 0.1) could be a result of the LH being non-specifically bound to the nucleosome (Fig. 2C and Supplementary Info S5 v and vi). The chromatosome peak (proximity ratio 0.4 to 0.9) could be fitted by two Gaussians, the intermediate peak with a mean proximity ratio of 0.54 and a high proximity ratio peak having a mean proximity ratio of 0.76. This FRET profile could be explained by the existence of two populations of chromatosomes, one with the gH dye closer to the AT flank and the other, with the gH dye closer to the AU flank. In each plot, the high proximity ratio peak corresponds to the chromatosomes where the gH is proximal to that particular labelled arm. The intermediate peak corresponds to the chromatosomes where the gH is oriented away from that labelled arm and is oriented towards the other, unlabeled arm.
These observations show that the gH, despite having the GVGA motif, has an equal preference for A-tracts paired to methyl group lacking deoxy-uridines and to A-tracts paired to methyl group containing thymines. This similar preference suggests that the hydrophobic interaction between thymine methyl groups and the GVGA motif (48, 49), may not be the only driving force behind the affinity of the gH for A-tracts. Our earlier observation that the gH does not show any preference towards random sequences having a higher AT-content, further suggests that it is not just the hydrophobic interactions with the thymine methyl group, but possibly some other parameters of the DNA that come into play. To investigate this, we modelled the AG, GA and the TU constructs, as will be presented in the section "Results: Arginines 42 and 94 in the globular domain of the LH mediate A-tract recognition by LH".

The *Xenopus laevis*, H1.0b LH C-terminal domain is not affected by the sequences of the L-DNA arms

In order to explore whether the CTD is affected by the flank sequences, we labelled the LH at residue 101. Fig. 3 shows the proximity ratio distribution between the donor dye on the CTD and the acceptor fluorophores on the plus and minus arms. We observed that for each of the chromatosomes AG, GA (Fig. 3A), MG and GM (Fig. 3B), the proximity ratio distributions between the gH and the plus and minus arms nearly overlap with each other. The mean proximity ratios differ significantly between the plus and minus arms in the AG construct only ($\Delta P = 0.1$) (Table 2), translating to a distance difference (assuming Förster distance = 52 Å) of 4 Å. The tail dye in the AG construct is observed to mirror the gH dye, giving a higher proximity ratio for the GC-flank plus arm with respect to the A-tract flank minus arm. For GA and GM, the mean proximity ratio of the tail dye differs very slightly between the minus and plus arms and we suggest that the tail dye is nearly equidistant from both the acceptor fluorophores. For the MG construct, the tail dye exhibits higher mean proximity ratio for the mixed DNA arm,

suggesting that the dye is oriented away from the flank towards which the gH is oriented, just as in AG construct. But the distance difference is again small.
From our data for the tail dye, we cannot conclude whether the CTD is locally structured on the mononucleosome, or it is disordered. But the good signal to noise ratio of the proximity ratio peak suggests that the dye is present on a more ordered part of the CTD. This could be because of the labelling position at the 101st residue, is on the upstream end of the tail, just six residues away from the gH. A dye attached further towards the end of the tail might show higher noise due to the disordered nature of the tail.
There is no effect of the flank sequences on the tail dye, contrary to what is seen for the gH dye. The tail dye is either equidistant from both the arms (GM and GA constructs), or shifted slightly away from the flank that the gH is proximal to, as in the MG and the AG construct.

Arginines 42 and 94 in the gH mediate A-tract recognition by the *Xenopus laevis*, H1.0b LH

Previous studies (9–11, 19, 21, 50), and structures (PDB ids: 5nl0 (11), 4qlc (9), 5wcu (10)) of chromatosomes with the gH bound in an on-dyad position, show that there are three DNA contacting surfaces on the gH (Fig. 1B). In the chromatosome structure from Bednar *et. al.* (11), where the gH is bound on-dyad on the nucleosome, Zone C contacts the ncp or core DNA. Zones A and B contact the two L-DNA arms (Fig. 1B). We modelled the AG, GA and TU chromatosomes, using the PDB 5nl0 as the starting structure, and placed the gH in different orientations on the nucleosome, generating a range of configurations. We then screened these structures using FPS (see Materials and Methods for details), giving our FRET-derived inter-dye distance as an input. By comparing the residues proximal to DNA in prior structural studies and the selected conformers of our AG and GA models, we find that our 'FRET-restrained' gH is located on the DNA in a similar fashion to the previously observed on-dyad binding mode (Supplementary Table T6 for a comparison of DNA proximal residues in X-ray crystallographic structures to our model**)**. Although for the GA construct, the 'wing' domain is not close enough to the dyad DNA, on aligning the 'open' conformation of gH (PDB 1hst chain A (51)), we observed that the interaction between the dyad DNA and the wing domain of the gH could become possible (Supplementary Fig. T6). The closed to open transition of the gH in a chromatosome has been observed *in silico* (49, 52), suggesting that this transition in the GA constructs could recover the wing domain/ncp DNA interactions.
In our AG and GA (Fig. 2A and B) models, we observe that Zone B is proximal to the A-tract. Specifically, arginines 42 (loop 1) and 94 (beta sheet 2) are strong candidates for interactions with the minor groove of the A-tract. The dye position (Cys 77) is observed to be oriented away from the L-DNAs, enabling free rotation of the donor dye. A concern could be that the highly negatively charged Alexa 488 dye tends to turn away from the DNA and affect the orientation of the gH, placing Zone B closer to the A-tract minor groove. We therefore also considered an alternate model (Supplementary Information S9) where we placed the gH rotated approximately 180 degrees with respect to the dyad axis. However, screening by FPS (42) clearly shows us that

only the first model, and hence only one gH orientation is possible, in which Zone B, with its two arginine residues (42 and 94) is positioned to interact with the very narrow minor groove of the A-tract (see Supplementary Information S9 for alternate model, Fig. 2A and B and table 3 for model satisfying our FRET data). The fact that the gH interacts with the minor and not the major groove of the L-DNA flank sequences in our model and also in previous studies (11, 19) makes it clear why the gH (or rather gH-dye in LH) shows equal proximity ratio distributions for the A(T)-tract flank and for the A(U)-tract flank. This is because the difference between dT and dU lies in the methyl group which is exposed on the major groove side of an A:T base pair and is not in its minor groove. Thus, if A-tract recognition is mediated by minor groove interactions, the LH cannot be able to distinguish between dU and dT containing A-tracts. Our models of the TU chromatosome (Fig. 2C) have the gH Zone B (R42/R94) oriented towards either the A(T)-tract minor groove (A and T shown in orange and blue colours, Fig. 2C) or the A(U)-tract minor groove (dU shown in sea-green, Fig. 2C extreme right). For each case, i.e. Zone B towards either A(U)-tract or A(T)-tract, the simulated FRET efficiencies between the gH dye and the two L-DNA arms are comparable to the FRET efficiencies of the two FRET populations observed (Fig. 2C and table 3), assuming a γ value of 1 (see Materials and Methods).

In all our models, the gH is positioned on-dyad. In this position, and with the modelled L-DNA separation, Zone B of the gH points towards the minor groove of the A-tracts. To investigate whether off-dyad binding might be consistent with our FRET data, we built off-dyad models for both AG and GA construct (See Supplementary Information S8), while keeping the GVGA motif on the gH wing domain proximal to the major groove of the A-tracts. Using FPS, we calculated the possible inter-fluorophore distances. We found that an off-dyad positioning of the gH would not result in inter-dye distances consistent with our experimental observations. Further support for an on-dyad positioning of the gH comes from the observations that LH subtype H1.0 associates with the nucleosome on-dyad ((11, 52, 53)).

It should be considered whether the observed FRET could be a result of sequence dependent bending of the DNA. It is well known that A-tracts confer an intrinsic bend in the DNA (29, 54, 55). To resolve this question, we modelled the L-DNA arms of the AG and GA constructs using the cgdna webserver (56) (Supplementary Information S7), and kept the gH position unchanged, on the nucleosome. Although we observe a sequence induced deviation of about 3.5 to 5.3 Å between the C7 atom of the thymines at the points of attachment of the dye-linker with the L-DNA, the simulated inter-fluorophore distance differs by only 0.3 to 0.6Å. Therefore, this bending of the DNA does not account for our observed differences in FRET. Furthermore, proximity ratio profiles of the two arms of the TU construct showed two populations corresponding to two different orientations of the gH (Fig. 2C). If the position of the gH on the nucleosome had remained constant and the two L-DNA arms of the TU construct bent in the same way, then we would have obtained only one population. Thus, our modelling shows that the sequences of the L-DNA affect the orientation of a gH bound on-dyad on the nucleosome. In the presence of A-tracts, the gH of *X. laevis* H1.0b orients itself in a way that two conserved

arginines can mediate electrostatic interactions with the narrow minor groove (thereby high negative charge density) of the A-tracts.

Discussion

The location of the LH on the nucleosome has garnered much attention because of its possible biological role: the two predominant modes by which the LH associates with the nucleosome (on and off-dyad) has been hypothesized to have major implications in the compaction of higher-order structures of the chromatin (9, 57). The positioning of the gH has been observed to be dependent on the sequence variations of the gH (22) and on post-translational modifications (14) of the gH. However, despite numerous studies showing the preference of the LH for AT-rich regions, the contribution of the DNA sequence to the positioning of the LH is yet unclear. The preference of LH for AT-rich regions, and A-tracts has been suggested to play a role in the compaction of AT-rich heterochromatin (23).

In the present work, we revisited the possible effect of the DNA sequence on the location and the orientation of the LH. Employing single-molecule FRET and distance-restrained modelling, we found that certain L-DNA sequences flanking the core or ncp DNA have an effect on the orientation of the gH, but do not affect the CTD. Three types of sequences were studied: A-tracts, purely GC, and a mixed 64% AT-rich sequence.

We tested the preference of the LH for the 64% AT-rich mixed sequence, vs. a purely GC-tract, in the MG and GM constructs. Surprisingly, in the MG construct, the gH showed higher FRET efficiency with the L-DNA arm containing purely GC sequence, but did not do so in the GM construct. These apparently contradicting observations led us to question whether the ncp DNA (non-palindromic Widom 601 sequence), owing to being asymmetric, induced the gH to associate in a specific orientation. We rule out this possibility because, in all our constructs, the ncp DNA remains the same. We would have had similar observations in all constructs if the ncp DNA was responsible for the orientation of the gH. A second possibility, that the gH has a preference for GC-tracts over a mixed sequence, is ruled out, not just by the similar proximity ratio profiles of the two arms in the GM construct, but also by reports that suggested that the LH has no preference for purely GC tracts (26, 30, 53).

When we substituted the 64% AT-rich mixed sequence with an A-tract (AG and GA constructs), we consistently observed that the gH dye showed a high proximity ratio for the A-tract containing L-DNA. Our distance-restrained models showed that the gH is located on-dyad, and is associated in a canonical fashion. Out of the three DNA proximal zones, zone B was observed to be close to the minor groove of the A-tract (minus arm of AG, plus arm of GA, and both plus and minus arms of TU). In this zone, two amino acid residues likely mediate the A-tract recognition. A highly conserved arginine is present in the loop 1 region of zone B (Arg42). A second arginine (Arg 94), conserved in H1.0 subtypes (see Supplementary Table T7), is present on the beta sheet at the C-terminal end of the gH. These two arginines protrude into the characteristically narrow minor groove of the A-tract.

A-tract regions, which are abundantly present in S/MAR (scaffold/matrix attachment regions) (58) and satellite sequences (26), are known to have the narrowest minor groove, and hence have a high density of negative charges. Moreover, A-tracts complementary to deoxy-uracils would have very similar minor groove widths to A-tracts complementary to thymines. Replacing thymines with uracils has been previously observed not to confer any significant change in the minor groove width (59). Arginines, being highly positively charged, interact electrostatically with the minor grooves of A-tracts (27, 31, 32), and are preferred over lysines in this particular interaction (60) because of the guanidium group that strongly hydrogen bonds with the DNA (14), and because of the lower electrostatic desolvation penalty compared to lysines (61). Electrophoretic mobility shift assays between lysine-rich somatic LH and arginine-rich sperm LH associated with DNA (62) showed a greater DNA binding capability of arginine-rich sperm LH. This was also reflected in binding assays (28) that showed that arginine-rich protamine nuclear proteins were selectively bound to A-tract rich DNA. On the basis of these findings and our observations, we propose that, specifically in H1.0 subtypes, the zone B region, with its two arginines in a 'pincer'-like motif, is a key contributor to A-tract recognition by LH.

Although the LH has been widely observed to have a preference for AT-rich DNA, our data for the MG and GM constructs contradicts this. This could be because of the mode of association of the LH isoform (*Xenopus laevis* H1.0b) that we studied with the nucleosome. The H1.0 subtype has been experimentally (11, 53) reported to associate in an on-dyad manner. Woods and Wereszczynski (52) showed computationally that the H1.0 subtype thermodynamically favours an on-dyad binding mode. In this binding mode, the gH is proximal to and can interact with the minor grooves of the flank regions of the L-DNA arms. To access the major groove, and thereby interact with the thymine methyl groups of the 64% AT-rich region through hydrophobic interactions, the gH has to associate in an off-dyad binding mode (48, 49). Despite containing the hydrophobic GVGA motif (Zone C), the H1.0b LH, associating on-dyad, is able to recognise the A-tract solely by electrostatic interactions with its minor groove.

Considering previous studies on AT-rich DNA recognition by the LH (48,49) and the results reported here, we suggest that there is more than one mechanism by which LHs can recognize AT-rich sequences. The recognition can be mediated by hydrophobic (zone C) or electrostatic (zone B) interactions. The mode of recognition largely depends upon whether the LH is able to access the major or the minor groove of the AT-rich sequence or A-tract. This in turn depends upon the binding mode of the gH on the nucleosome. The existence of multiple recognition mechanisms has biological implications, with the binding preferences of LHs for AT-rich and A-tract regions enabling compaction of AT-rich heterochromatin.

Data and software availability: Data for this work is available at the link https://doi.org/10.5281/zenodo.4312185. The software used to obtain proximity ratio and stoichiometry histograms, AlexEval is available at the link: https://github.com/sisbaner/AlexEval.

Funding: This work was supported by the Helmholtz International Graduate School for Cancer Research (DKFZ) and the Klaus Tschira Foundation.

Acknowledgements:

We dedicate this work to the memory of late Prof. Dr. Jörg Langowski, a remarkable pioneer in the field of chromatin biophysics. We are very grateful to Prof. Jeffrey Hayes (University of Rochester Medical Center) and Dr. Amber Cutter for kindly providing us with the plasmids and protein for unlabelled and labelled LH. We are thankful to Nathalie Schwarz for DNA purification and alongside Maria Mildenberger and Gabriele Mueller, for protein purification. M.D. is thankful to Angga Prawira for measuring some technical replicates of AG, and MG chromatosomes. We are grateful to Dr. Kathrin Lehmann for helping in designing the MG chromatosome and also for very helpful feedback on this manuscript.

References

1. Kinkade JM Jr, Cole RD. (1966) The resolution of four lysine-rich histones derived from calf thymus. *J Biol Chem*. 241(24):5790-7. PMID: 5954358.

2. Hohmann P. (1980) Species- and cell-specific expression of H1 histones in tissue culture cells. *Arch Biochem Biophys*. 205(1):198-209. doi: 10.1016/0003-9861(80)90099-5. PMID: 7447476.

3. Billett,M.A. and Barry,J.M. (1974) Role of Histones in Chromatin Condensation. *Eur J Biochem*, 49, 477–484.

4. Fan,L. and Roberts,V.A. (2006) Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. *Proceedings of the National Academy of Sciences*, 103, 8384–8389.

5. Fyodorov,D.V., Zhou,B.-R., Skoultchi,A.I. and Bai,Y. (2018) Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol*, 19, 192–206.

6. Healton,S.E., Pinto,H.D., Mishra,L.N., Hamilton,G.A., Wheat,J.C., Swist-Rosowska,K., Shukeir,N., Dou,Y., Steidl,U., Jenuwein,T., *et al.* (2020) H1 linker histones silence repetitive elements by promoting both histone H3K9 methylation and chromatin compaction. *Proc Natl Acad Sci USA*, 10.1073/pnas.1920725117.

7. Öztürk,M.A., De,M., Cojocaru,V. and Wade,R.C. (2020) Chromatosome Structure and Dynamics from Molecular Simulations. *Annu. Rev. Phys. Chem.*, 71, 101–119.

8. Öztürk,M.A., Cojocaru V, Wade RC. (2018) Toward an Ensemble View of Chromatosome Structure: A Paradigm Shift from One to Many. *Structure*. 26(8):1050-1057.

9. Zhou,B.-R., Jiang,J., Feng,H., Ghirlando,R., Xiao,T.S. and Bai,Y. (2015) Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell*, 59, 628–638.

10. Zhou,B.-R., Jiang,J., Ghirlando,R., Norouzi,D., Yadav,K.N.S., Feng,H., Wang,R., Zhang,P., Zhurkin,V. and Bai,Y. (2018) Revisit of reconstituted 30-nm nucleosome arrays reveals an ensemble of dynamic structures. *J Mol Biol*. 430(18 Pt B):3093-3110.

11. Bednar,J., Garcia-Saez,I., Boopathi,R., Cutter,A.R., Papai,G., Reymer,A., Syed,S.H., Lone,I.N., Tonchev,O., Crucifix,C., *et al.* (2017) Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell*, 66, 384-397.e8.

12. Zhou,B.-R., Feng,H., Kato,H., Dai,L., Yang,Y., Zhou,Y. and Bai,Y. (2013) Structural insights into the histone H1-nucleosome complex. *Proc Natl Acad Sci USA*, 110, 19390.

13. Song,F., Chen,P., Sun,D., Wang,M., Dong,L., Liang,D., Xu,R.-M., Zhu,P. and Li,G. (2014) Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science*, 344, 376.

14. Öztürk,M.A., Cojocaru,V. and Wade,R.C. (2018) Dependence of Chromatosome Structure on Linker Histone Sequence and Posttranslational Modification. *Biophysical Journal*, 114, 2363–2375.

15. Caterino,T.L. and Hayes,J.J. (2011) Structure of the H1 C-terminal domain and function in chromatin condensationThis paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Journal's usual peer review process. *Biochem. Cell Biol.*, 89, 35–44.

16. Fang,H., Wei,S., Lee,T.-H. and Hayes,J.J. (2016) Chromatin structure-dependent conformations of the H1 CTD. *Nucleic Acids Res*, Volume 44, Issue 19, 2 November 2016, Pages 9131–9141.

17. Luque,A., Collepardo-Guevara,R., Grigoryev,S. and Schlick,T. (2014) Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. *Nucleic Acids Res*, 42, 7553–7560.

18. Heidarsson,P.O., Mercadante,D., Sottini,A., Nettels,D., Borgia,M.B., Borgia,A., Kilic,S., Fierz,B., Best,R.B. and Schuler,B. (2020) Disordered Proteins Enable Histone Chaperoning on the Nucleosome.
bioRxiv 2020.04.17.046243; doi: https://doi.org/10.1101/2020.04.17.046243

19. Brown,D.T., Izard,T. and Misteli,T. (2006) Mapping the interaction surface of linker histone H10 with the nucleosome of native chromatin in vivo. *Nature Structural & Molecular Biology*, 13, 250–255.

20. Pachov,G.V., Gabdoulline,R.R. and Wade,R.C. (2011) On the structure and dynamics of the complex of the nucleosome and the linker histone. *Nucleic Acids Research*, 39, 5255–5263.

21. Öztürk,M.A. and Wade,R.C. (2020) Computation of FRAP recovery times for linker histone – chromatin binding on the basis of Brownian dynamics simulations. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1864, 129653.

22. Zhou,B.-R., Feng,H., Ghirlando,R., Li,S., Schwieters,C.D. and Bai,Y. (2016) A Small Number of Residues Can Determine if Linker Histones Are Bound On or Off Dyad in the Chromatosome. *Journal of Molecular Biology*, 428, 3948–3959.

23. Zlatanova,J. and Yaneva,J. (1991) Histone H1-DNA Interactions and Their Relation to Chromatin Structure and Function. *DNA and Cell Biology*, 10, 239–248.

24. An,W., Leuba,S.H., van Holde,K. and Zlatanova,J. (1998) Linker histone protects linker DNA on only one side of the core particle and in a sequence-dependent manner. *Proc Natl Acad Sci USA*, 95, 3396.

25. Käs, E., Izaurralde,E. and Laemmli,U.K. (1989) Specific Inhibition of DNA Binding to Nuclear Scaffolds and Histone Hl by Distamycin: The Role of Oligo(dA) -Oligo(dT) Tracts.

26. Cao,K., Lailler,N., Zhang,Y., Kumar,A., Uppal,K., Liu,Z., Lee,E.K., Wu,H., Medrzycki,M., Pan,C., *et al.* (2013) High-Resolution Mapping of H1 Linker Histone Variants in Embryonic Stem Cells. *PLoS Genet*, 9, e1003417.

27. Haran,T.E. and Mohanty,U. (2009) The unique structure of A-tracts and intrinsic DNA bending. *Quart. Rev. Biophys.*, 42, 41–81.

28. Roque,A. (2004) The preferential binding of histone H1 to DNA scaffold-associated regions is determined by its C-terminal domain. *Nucleic Acids Research*, 32, 6111–6119.

29. Buckwalter,J.M., Norouzi,D., Harutyunyan,A., Zhurkin,V.B. and Grigoryev,S.A. (2017) Regulation of chromatin folding by conformational variations of nucleosome linker DNA. *Nucleic Acids Research*, 45, 9372–9387.

30. Tomaszewski,R. (1997) The AT-rich flanks of the oocyte-type 5S RNA gene of Xenopus

laevis act as a strong local signal for histone H1-mediated chromatin reorganization in vitro. *Nucleic Acids Research*, 25, 458–466.

31. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein–DNA recognition. *Nature*, 461, 1248–1253.

32. West,S.M., Rohs,R., Mann,R.S. and Honig,B. (2010) Electrostatic Interactions between Arginines and the Minor Groove in the Nucleosome. *Journal of Biomolecular Structure and Dynamics*, 27, 861–866.

33. Lowary,P.T. and Widom,J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, 276, 19–42.

34. Luger,K., Rechsteiner,T.J. and Richmond,T.J. (1999) Preparation of nucleosome core particle from recombinant histones. In *Methods in Enzymology*. Elsevier, Vol. 304, pp. 3–19.

35. Gansen,A., Hauger,F., Tóth,K. and Langowski,J. (2007) Single-pair fluorescence resonance energy transfer of nucleosomes in free diffusion: Optimizing stability and resolution of subpopulations. *Analytical Biochemistry*, 368, 193–204.

36. Lehmann,K., Zhang,R., Schwarz,N., Gansen,A., Mücke,N., Langowski,J. and Toth,K. (2017) Effects of charge-modifying mutations in histone H2A α3-domain on nucleosome stability assessed by single-pair FRET and MD simulations. *Sci Rep*, 7, 13303.

37. Tóth,K., Brun,N. and Langowski,J. (2006) Chromatin Compaction at the Mononucleosome Level. *Biochemistry*, 45, 1591–1598.

38. Würtz,M., Aumiller,D., Gundelwein,L., Jung,P., Schütz,C., Lehmann,K., Tóth,K. and Rohr,K. (2019) DNA accessibility of chromatosomes quantified by automated image analysis of AFM data. *Sci Rep*, 9, 12788.

39. Kapanidis,A.N., Lee,N.K., Laurence,T.A., Doose,S., Margeat,E. and Weiss,S. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *PNAS*, 101, 8936–8941.

40. Lee,N.K., Kapanidis,A.N., Wang,Y., Michalet,X., Mukhopadhyay,J., Ebright,R.H. and Weiss,S. (2005) Accurate FRET Measurements within Single Diffusing Biomolecules Using Alternating-Laser Excitation. *Biophysical Journal*, 88, 2939–2953.

41. Gansen,A., Tóth,K., Schwarz,N. and Langowski,J. (2009) Structural Variability of Nucleosomes Detected by Single-Pair Förster Resonance Energy Transfer: Histone Acetylation, Sequence Variation, and Salt Effects. *J. Phys. Chem. B*, 113, 2604–2613.

42. Kalinin,S., Peulen,T., Sindbert,S., Rothwell,P.J., Berger,S., Restle,T., Goody,R.S., Gohlke,H. and Seidel,C.A.M. (2012) A toolkit and benchmark study for FRET-restrained high-precision

structural modeling. *Nature Methods*, 9, 1218–1225.

43. Koopmans,W.J.A., Brehm,A., Logie,C., Schmidt,T. and van Noort,J. (2007) Single-Pair FRET Microscopy Reveals Mononucleosome Dynamics. *J Fluoresc*, 17, 785–795.

44. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25, 1605–1612.

45. Dimura,M., Peulen,T.O., Hanke,C.A., Prakash,A., Gohlke,H. and Seidel,C.A. (2016) Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Current Opinion in Structural Biology*, 40, 163–185.

46. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E.,3rd, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, 92, 3817–3829.

47. Maier,J.A., Martinez,C., Kasavajhala,K., Wickstrom,L., Hauser,K.E. and Simmerling,C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, 11, 3696–3713.

48. Cui,F. and Zhurkin,V.B. (2009) Distinctive sequence patterns in metazoan and yeast nucleosomes: Implications for linker histone binding to AT-rich and methylated DNA. *Nucleic Acids Research*, 37, 2818–2829.

49. Öztürk,M.A., Pachov,G.V., Wade,R.C. and Cojocaru,V. (2016) Conformational selection and dynamic adaptation upon linker histone binding to the nucleosome. *Nucleic Acids Res*, 44, 6599–6613.

50. George,E.M., Izard,T., Anderson,S.D. and Brown,D.T. (2010) Nucleosome Interaction Surface of Linker Histone H1c Is Distinct from That of H1 ^{0}. *J. Biol. Chem.*, 285, 20891–20896.

51. Ramakrishnan,V., Finch,J.T., Graziano,V., Lee,P.L. and Sweet,R.M. (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, 362, 219–223.

52. Woods,D.C. and Wereszczynski,J. (2020) Elucidating the influence of linker histone variants on chromatosome dynamics and energetics. *Nucleic Acids Research*, 48, 3591–3604.

53. Zhou,B.-R., Feng,H., Kale,S., Fox,T., Khant,H., de Val,N., Ghirlando,R., Panchenko,A.R. and Bai,Y. (2020) Distinct Structures and Dynamics of Chromatosomes with Different Human Linker Histone Isoforms. *Molecular Cell*, ISSN 1097-2765

54. Marin-Gonzalez,A., Vilhena,J.G., Moreno-Herrero,F. and Perez,R. (2019) DNA Crookedness Regulates DNA Mechanical Properties at Short Length Scales. *Phys. Rev. Lett.*, 122, 048102.

55. Marin-Gonzalez,A., Pastrana,C.L., Bocanegra,R., Martín-González,A., Vilhena,J.G., Pérez,R., Ibarra,B., Aicart-Ramos,C. and Moreno-Herrero,F. (2020) Understanding the paradoxical mechanical response of in-phase A-tracts at different force regimes. *Nucleic Acids Research*, 48, 5024–5036.

56. De Bruin,L. and Maddocks,J.H. (2018) cgDNAweb: a web interface to the cgDNA sequence-dependent coarse-grain model of double-stranded DNA. *Nucleic Acids Research*, 46, W5–W10.

57. Perišić,O., Portillo-Ledesma,S. and Schlick,T. (2019) Sensitive effect of linker histone binding mode and subtype on chromatin condensation. *Nucleic Acids Research*, 47, 4948–4957.

58. Narwade,N., Patel,S., Alam,A., Chattopadhyay,S., Mittal,S. and Kulkarni,A. (2019) Mapping of scaffold/matrix attachment regions in human genome: a data mining exercise. *Nucleic Acids Research*, 47, 7247–7261.

59. Fadda,E. and Pomès,R. (2011) On the molecular basis of uracil recognition in DNA: comparative study of T-A versus U-A structure, dynamics and open base pair kinetics. *Nucleic Acids Research*, 39, 767–780.

60. Lewis,J.D., Saperas,N., Song,Y., Zamora,M.J., Chiva,M. and Ausio,J. (2004) Histone H1 and the origin of protamines. *Proceedings of the National Academy of Sciences*, 101, 4148–4152.

61. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem*, 79, 233–269.

62. Piscopo,M., Conte,M., Di Paola,F., Conforti,S., Rana,G., De Petrocellis,L., Fucci,L. and Geraci,G. (2010) Relevance of Arginines in the Mode of Binding of H1 Histones to DNA. *DNA and Cell Biology*, 29, 339–347.

Tables:

Name of construct	Minus L-DNA WIDOM 601	Plus L-DNA
AG	5'-AAAAAAAAAAA-3' 3'- TTTTTTTTTTT -5'	5' – GGGCGGCCGCG – 3' 3' - CCCGCCGGCGC – 5'
GA	5' – CGCGGCCGCCC – 3' 3' - GCGCCGGCGGG – 5'	5' – TTTTTTTTTTT – 3' 3' - AAAAAAAAAAA – 5'
MG	5' – ATACATGCACA – 3' 3' – TATGTACGTGT – 5'	5' - GGGCGGCCGCG – 3' 3' - CCCGCCGGCGC – 5'
GM	5' – CGCGGCCGCC – 3' 3' - GCGCCGGCGG – 5'	5' –AGGCATGTAT -3' 3' - TCCGTACATA -5'
TU	5'-AAAAAAAAAAA-3' 3'- TTTTTTTTTTT -5'	5' – AAAAAAAAAAA -3' 3'- UUUUUUUUUUU- 5'

Table 1: The chromatosome constructs with the different 'flank' sequences on either side of the core or ncp DNA that were investigated. 'U' denotes deoxyuridine (dU) in the plus L-DNA of the TU construct. See Supplementary Information S2 for the full DNA sequences.

	gH		CTD	
	-93	+94	-93	+94
AG	0.70 ± 0.04	0.53 ± 0.01	0.65 ± 0.02	0.75 ± 0.02
GA	**0.58 ± 0.01** 0.89 ± 0.02	0.80 ± 0.02	0.66 ± 0.01	0.62 ± 0.002
TU	0.54 ± 0.01 0.76 ± 0.02	0.54 ± 0.04 0.76 ± 0.02		
MG	0.56 ± 0.01	0.73 ± 0.02	0.70 ± 0.02	0.65 ± 0.004
GM	0.60 ± 0.002	0.62 ± 0.01	0.68 ± 0.02	0.72 ± 0.003

Table 2: Proximity ratio between the fluorophores on the LH (gH and CTD) and the minus or plus L-DNA arms. If more than one population are consistently present, both are indicated. Major population indicated in bold.

	Experiment		Model	
	-93	+94	-93	+94
AG (R in Å)	45 ± 1.4	50 ± 0.2	45	51
GA (R in Å)	48.4 ± 0.4	41 ± 0.5	47.4	42.5
TU (P=E, $\gamma = 1$)	0.54, 0.76	0.54, 0.76	0.55, 0.77	0.53, 0.72

Table 3: Comparison between experimental and computed values: inter-fluorophore distances (between gH and plus and minus L-DNA arms) in AG and GA constructs and FRET efficiency values in TU construct, assuming $\gamma = 1$.

Figure:

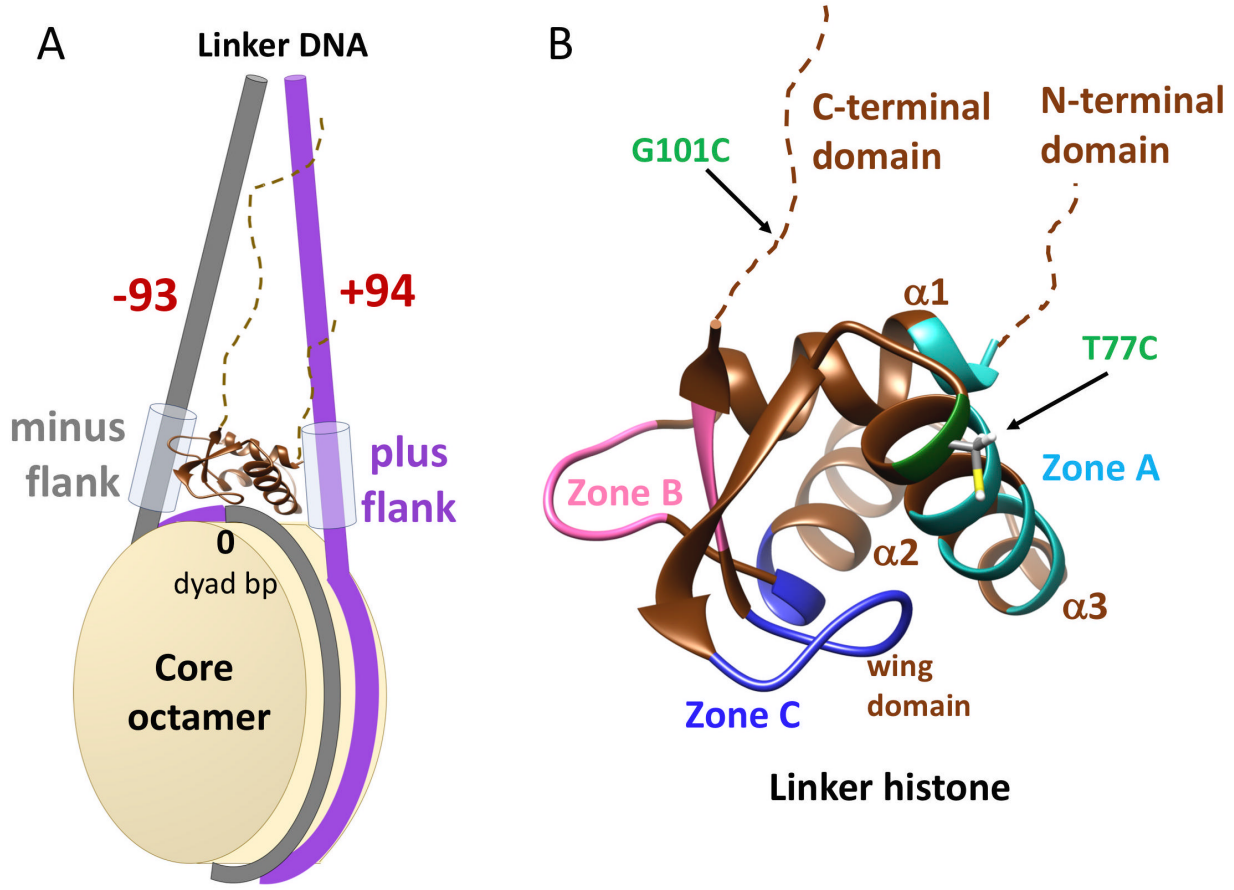

Figure 1: A. Schematic diagram of the chromatosome reconstituted from 226 bp non-palindromic Widom 601 DNA. The dyad base pair is numbered 0. The negative arm of the L-DNA is coloured grey and the plus arm is coloured purple. The Alexa 594 labelling positions on the DNA are on thymines -93 and +94 (in red), 20 bp from the ends of the L-DNA.
B. Linker histone with the gH in cartoon representation showing the two Alexa 488 labelling positions on the gH (T77C) and the CTD (G101C). The three DNA-proximal zones are shown: A (cyan) and B (pink) interacting with the two L-DNA arms, and C (blue) interacting with the core or ncp DNA.

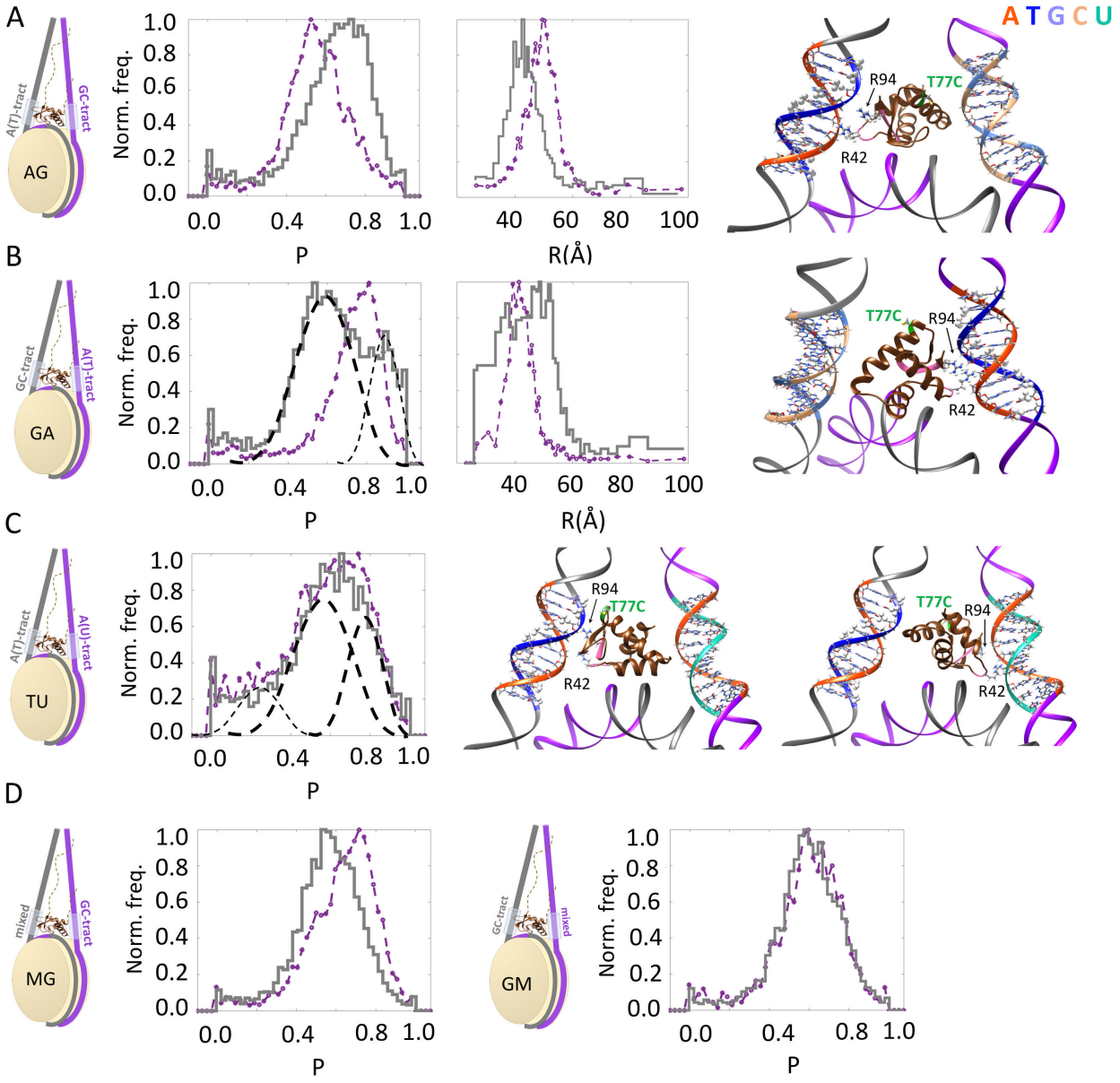

Figure 2: FRET data for the dyes on the L-DNA arms and on the gH and the distance-restrained models derived for the five chromatosomes studied with gH binding on-dyad. The constructs are indicated by schematic cartoons. For the AG (A) and GA (B) constructs, proximity ratio (P) and distance (R) histograms are shown with the colouring corresponding to the cartoons (purple for plus arm, grey for minus arm). When more than one population exists, major and minor populations of the proximity ratio histogram are indicated by thick and thin dashed black lines, respectively. The derived distance-restrained model (for the major population) is shown on the right with the gH binding with arginines 42 and 94 of Zone B suitably positioned to interact with the A-tract (dA in orange base paired to dT in blue) via the minor groove. The C7 methyl groups of the thymines, shown as spheres, are not accessible in the minor groove. The gH labelling position, T77C, is coloured green.

C. TU construct. Proximity ratio histograms of the two L-DNA arms with the gH overlap, and can each be subdivided into three populations, of which two are major (thick dashed lines). Both the L-DNA dyes show two populations, suggesting that the gH can be oriented both towards the minor groove of A(T)-tract (grey, minus) (left) as well as the minor groove of A(U)-tract (purple, plus) (right). The computed FRET efficiencies, E (assuming a γ value of 1, i.e. E = P), between the dyes on the L-DNA arms and the gH from the two models are very similar to the experimentally determined proximity ratios of the two major peaks.
D. MG (left) and GM (right) constructs. For the MG construct, the gH dye has a higher proximity ratio with the GC-tract flank (purple) than the mixed flank (grey). This is not observed when the flanks are swapped in the GM construct.

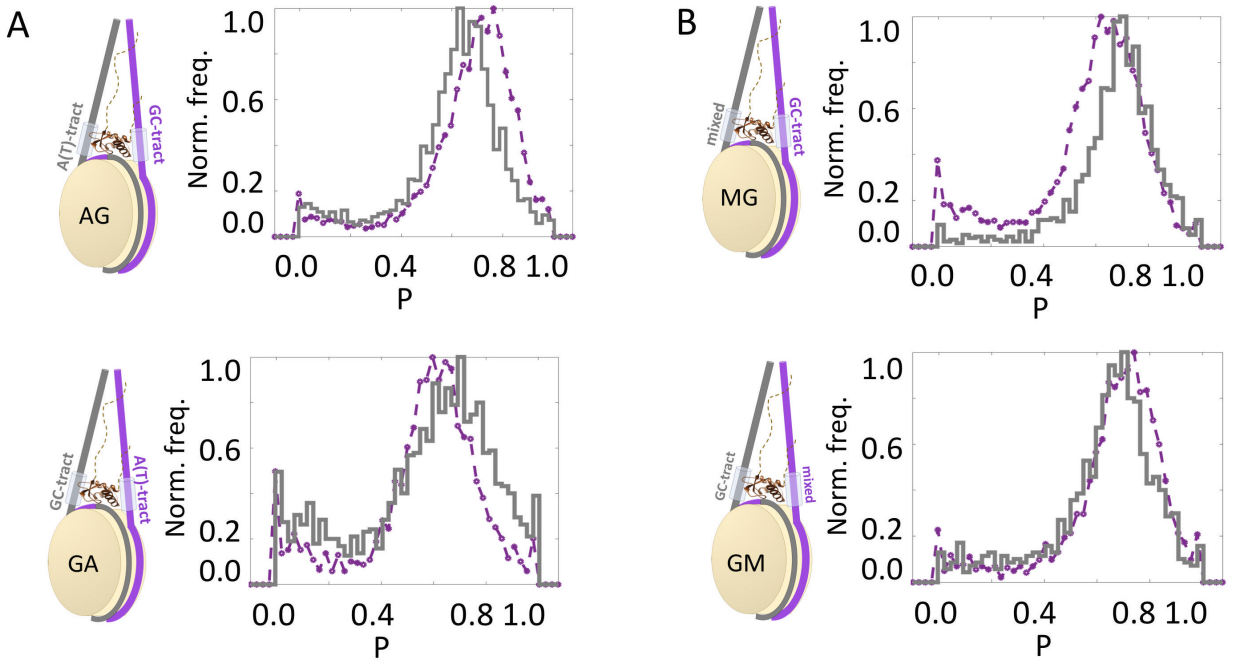

Figure 3: FRET data show the LH CTD location is not affected by the flank sequences. Proximity ratio (P) histograms between the dyes on the L-DNA arms and CTD of the LH are shown for the four chromatosome constructs shown as schematic cartoons: (A) AG and GA and (B) MG and GM.