The dynamic organization of prokaryotic genomes: DNA bridging and wrapping proteins across the tree of life

Erkelens, A.M.

Chapter 4

Specific DNA binding of archaeal histones HMfA and HMfB

This chapter is based on the following article: Erkelens, A.M., Henneman, B., van der Valk, R.A., Kirolos, N.C.S., Dame, R.T. (2023), Specific DNA binding of archaeal histones HMfA and HMfB. *Frontiers in Microbiology*, 14, 1166608

Abstract

In archaea, histones play a role in genome compaction and are involved in transcription regulation. Whereas archaeal histones bind DNA without sequence specificity, they bind preferentially to DNA containing repeats of alternating A/T and G/C motifs. These motifs are also present on the artificial sequence “Clone20”, a high-affinity model sequence for binding of the histones from *Methanothermus fervidus*. Here, we investigate the binding of HMfA and HMfB to Clone20 DNA. We show that specific binding at low protein concentrations (<30 nM) yields modest DNA compaction, attributed to tetrameric nucleosome formation, whereas nonspecific binding strongly compacts DNA. We also demonstrate that histones impaired in hypernucleosome formation are still able to recognize the Clone20 sequence. Histone tetramers indeed exhibit a higher binding affinity for Clone20 than nonspecific DNA. Our results indicate that a high-affinity DNA sequence does not act as a nucleation site, but is bound by a tetramer which we propose is geometrically different from the hypernucleosome. Such a mode of histone binding might permit sequence-driven modulation of hypernucleosome size. These findings might be extrapolated to histone variants that do not form hypernucleosomes. Versatile binding modes of histones could provide a platform for functional interplay between genome compaction and transcription.

Introduction

Every organism needs to compact its genome dynamically. Eukaryotes express histone proteins that form a defined octameric core with ~ 147 bp DNA wrapped around it, called the nucleosome (1). Archaea express histone homologues, which are involved in genome compaction and transcription regulation (2, 3). Together with other architectural proteins, such as Alba and MC1, archaeal histones have been hypothesized to function as transcription regulators (4, 5). Expression of model histones HMfA and HMfB from *Methanothermus fervidus* in *Escherichia coli* resulted in a mild generic repressive effect on transcription (6). Also, in their native environment, the histones of *Thermococcus kodakarensis* were shown to repress transcription, which was dependent on their multimerization state (7). Archaeal histones are dimers in solution, although micrococcal nuclease (MNase) digestion studies in *M. fervidus, Haloferax volcanii* and *Methanobacterium thermoautotrophicum* point to a tetramer as the smallest relevant unit on DNA, showing protection of ~ 60 bp (8, 9). Similar studies in *T. kodakarensis*, however, show protection of DNA increases with ~ 30 bp steps up to 450 bp, suggesting multimerization by adding dimers (10).

This multimer of archaeal histone dimers, called the hypernucleosome, is a rod-like structure with DNA wrapped around it (11, 12). The formation of a hypernucleosome coats and compacts the DNA and could potentially play an important role in transcription regulation. Assembly of histone dimers into a hypernucleosome is dependent on stacking interactions between a dimer and its second and third neighbor (12, 13). Most histones throughout the archaeal domain are predicted to be able to form hypernucleosomes, but some archaea encode histones that lack some or all stacking interactions (13). As archaea encode up to 11 histone variants within a single genome, many different combinations of dimers, tetramers and multimers are possible. Depending on different expression levels during the growth cycle and environmental cues, heteromerization could play an essential role in modulating (hyper)nucleosome size and structure, potentially affecting transcription (14–16). Histone variants lacking stacking interactions could act as ‘capstones’ and limit the size of hypernucleosome (17).

Archaeal histones, like their eukaryotic counterparts, bind DNA without sequence specificity, but with a preference for more GC-rich sequences (8, 18). Transcription start sites (TSSs) are often AT-rich and depleted from histones, both in archaea and eukaryotes (19, 20). HMfB preferentially binds GC-rich sequences with alternating GC and AT motifs (21, 22). Such a sequence motif also positions histone tetramers on genomic DNA in *H. volcanii* (8, 18). Using systematic evolution of ligands by exponential enrichment (SELEX), sequences with high affinity for HMfB were identified (22). One of the resulting sequences, “Clone20”, consists of alternating A/T-

and G/C-rich regions (see Materials & Methods) and has a high binding affinity for HMfA and HMfB tetramers (23). However, it is unclear whether such a high-affinity site functions as a nucleation site for hypernucleosome formation.

Here we show that HMfA and HMfB modestly compact Clone20 DNA by forming a tetrameric complex before hypernucleosome formation and that histone derivatives with impaired stacking interactions are still able to recognize the Clone20 sequence. High-affinity sites are likely bound by a geometrically different, more closed, tetramer, which is incompatible with hypernucleosome formation. This might indicate a previously unknown ability of histone variants that lack stacking interactions as tetrameric roadblocks halting hypernucleosome progression.

Materials and Methods

Protein expression and purification

HMfA and HMfB were kindly provided by John Reeve and Kathleen Sandman. $HMfA_{K31A\ E35A}$ and $HMfB_{D14A\ K30A\ E34A}$ were purified as previously described (12). Identity of the proteins was confirmed with mass spectrometry. Plasmids pRD323 ($HMfA_{K31A\ E35A}$) and pRD324 ($HMfA_{D14A\ K30A\ E34A}$) for expression of mutated HMfA and HMfB derivatives were deposited at Addgene with ID 198044 and 198045 respectively.

DNA substrate preparation

For the Tethered Particle Motion (TPM) DNA substrate, the Clone20 sequence (GCACAGTTGAGCGATCAAAAACGCCGTAGAACGCTTTAATTGATAATCAAAGGCC GCAGA, (22)) was cloned into pBR322 using restriction digestion with EcoRI and HindIII (Thermo Scientific), resulting in plasmid pRD120. The same approach was used to create pRD123 containing Clone20R. Gibson assembly was used to create pRD196 containing Clone20L (24). We used PCR to generate and amplify a 685 bp linear substrate containing the cloned sequence, using digoxygenin- and biotin-labeled oligonucleotides and DreamTaq DNA polymerase (Thermo Scientific) (25) or Phusion® High-fidelity DNA polymerase (Thermo Scientific). The products were purified with the GenElute PCR Clean-up kit (Sigma Aldrich). The nonspecific DNA substrate was prepared as previously described (12).

For microscale thermophoresis, 78 bp complementary oligonucleotides were designed using the Nonspecific and Clone20 sequence (table S4.1). The top strand was labeled with Cy5 and the complementary oligonucleotides were mixed 1:1 to a final concentration of 40 µM. Subsequently, they were heated to 95°C and slowly cooled to room temperature to anneal the strands.

Tethered particle motion

The tethered particle motion experiments, data analysis and representation of results were performed as previously described (26). To select single-tethered beads, we used a standard deviation cut-off of 8% and an anisotropic ratio cut-off of 1.3. As measurement buffer 50 mM Tris-HCl pH 7, 75 mM KCl was used.

The end-to-end distance was calculated by selecting the 25 beads closest to the fitted RMS at the respective protein concentration. Next, the 2.5% most distant positions of each bead were collected. The end-to-end distance was calculated for each point using triangular calculations and the diameter of the beads (0.44 µm). Next, the data was represented as histograms and fitted with a skewed Gaussian fit. The difference between the two populations was obtained by taking a pairwise distance distribution and fitting the resulting histogram with a Gaussian distribution.

Microscale thermophoresis

The DNA substrates described above with a concentration of 40 nM were diluted 1:1 with the HMf proteins. The final experimental buffer consisted of 50 mM Tris-HCl pH 8, 75 mM KCl. In MST experiments with $HMfB_{D14A\ K30A\ E34A}$, 0.2% Tween20 was added for optimal solubility of the protein. The samples were incubated for 5 minutes at room temperature and transferred to MST capillaries (Monolith NT.115 Premium Capillaries, NanoTemper, Germany). The measurement was done at 40% LED power and medium MST laser power using the NanoTemper Monolith NT.115. Total measurement time was 40 seconds, with 5 seconds laser off, 30 seconds laser on and 5 seconds laser off. F_{norm} values were evaluated after 20 seconds of laser on. ΔF_{norm} values were calculated by subtracting F_{norm} of DNA only. Occupancy values were calculated and fitted with a Hill binding model.

Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS)

The molecular weight of HMf complexes in solution was measured using a SEC-MALS system comprising a miniDAWN® TREOS®, NanoStar DLS, Optilab differential refractometer (Wyatt technology) and 1260 Infinity II multiple wavelength absorbance detector (Agilent). The samples containing at least 1 mg/ml HMfA or HMfB were run on a Superdex75 10/300 Increase GL column (Cytiva) with phosphate-buffered saline (12 mM $NaPO_4$ pH 7.4, 137 mM NaCl) as running buffer. The ASTRA 8 software package was used to select the peaks and report the molecular weight.

Results

HMfA and HMfB bind as tetramers to the Clone20 sequence before hypernucleosome formation

To determine the effect of specific DNA sequences, we carried out TPM experiments with a 685 bp DNA substrate with the Clone20 sequence at its center. The reduction of the root mean square displacement (RMS) of the DNA tether in TPM indicates that both HMfA and HMfB compact the Clone20 substrate (figure 4.1A and B). Compaction as a function of protein concentration occurs in two steps. The compaction step at high protein concentrations (at > ~30 nM for both HMfA and HMfB) resembles the strong cooperative compaction of nonspecific DNA into a hypernucleosome (figure 4.1A and B). This step occurred at slightly higher protein concentrations on Clone20 DNA than on nonspecific DNA. The first compaction step, occurring at low protein concentrations (at 1-30 nM for HMfA and 20-30 nM for HMfB), was not observed for nonspecific DNA, and is therefore due to specific binding of HMfA and HMfB to the Clone20 sequence. At this step, the RMS is reduced to ~125 nm. For HMfB, this state is unpopulated up to 20 nM, partially populated at 20-22 nM and completely populated at 23-25 nM. The ratio of both populations is expressed as occupancy for HMfB, to which the Hill equation was fit (figure 4.1C). This resulted in a binding constant of (K_D) of 21 ± 0.2 nM and a Hill binding coefficient (n) of 32 ± 8. HMfA directly fully populates this intermediate state at 1-30 nM. (figure 4.1A). Therefore, no exact binding constant could be calculated as the intermediate state is already fully populated at 1 nM, which means that the binding constant of HMfA for Clone20 is in the sub-nanomolar concentration range. The Clone20 site consists of 60 bp, theoretically permitting binding of a tetramer to this sequence. To determine whether this is indeed the case, we calculated the end-to-end distance of the DNA molecule without protein and with 5 nM HMfA (figure 4.1D). This resulted in an end-to-end distance of 101 ± 11 nm and 78.9 ± 11 nm for 0 and 5 nM respectively. The pairwise distribution gives a difference of 22.8 ± 10 nm, corresponding to 67 ± 30 bp (where each bp is 0.34 nm). The same analysis was done for HMfB at a concentration of 21 nM, where two populations were observed (figure S4.1), and this yielded a difference of 23.0 ± 9 nm or 68 ± 27 bp. These observations suggest that both HMfA and HMfB form a structurally identical tetrameric histone-DNA complex at the Clone20 site. However, this site is unable to act as a nucleation site as it does not promote hypernucleosome formation.

The finding that HMfA exhibits a higher binding affinity for Clone20 than HMfB contradicts results from EMSA experiments (23). The difference may be caused by a different pH (7.0 in our experiments vs 8.0 in the studies of Bailey et al.) as the isoelectric points of HMfA and HMfB are different (8.06 and 9.59, respectively). Another possibility

is that a difference in measured affinity is a result of using different methods, with EMSA involving a gel matrix and TPM using DNA in solution attached to a glass surface. Also, the DNA substrate length is different; our 685 bp substrate is much longer than the 110 bp used by Bailey et al., which could have effects on apparent binding affinity and cooperativity.

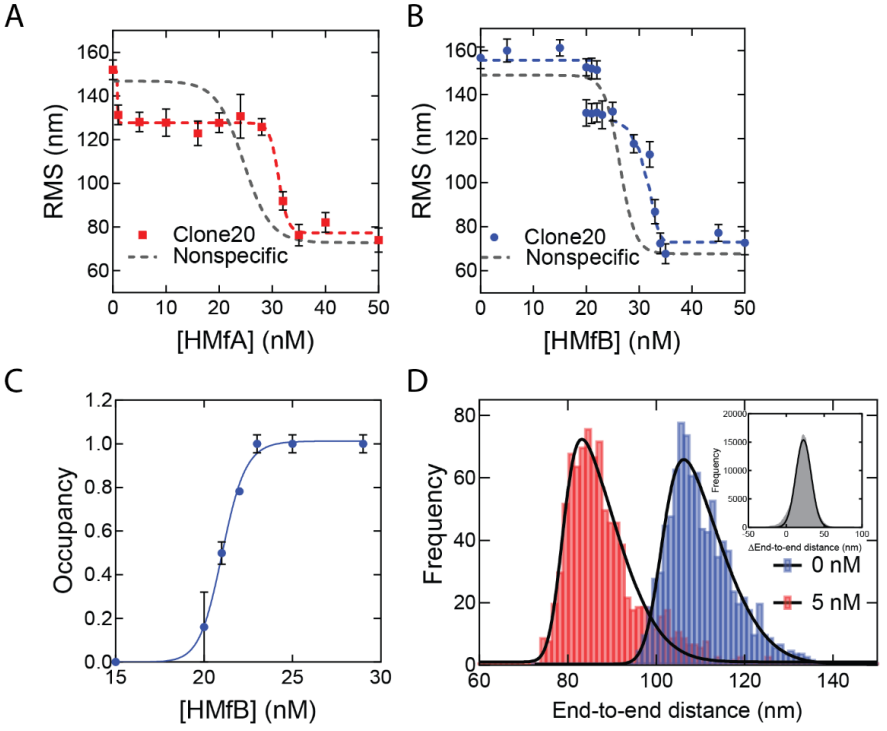

Figure 4.1 HMfA and HMfB bind as tetramers to the Clone20 site preceding hypernucleosome formation A) Root mean square displacement (RMS) values of Nonspecific and Clone20 DNA tethers incubated with HMfA and B) with HMfB measured by TPM in 50 mM Tris-HCl pH 7, 75 mM KCl. Histograms were fitted with a Gaussian function and the mean values are represented by red and blue dots, respectively. Data for Nonspecific DNA was reproduced from Henneman et al. (12) and depicted as a line to guide the eye. Error bars represent the propagated standard deviation of at least two replicates C) Binding curve for specific binding of HMfB to the Clone20 substrate. The data points were fitted using the Hill binding model. Error bars represent the standard deviation of the replicates and propagated error for data points at saturation. D) Calculated end-to-end distances for unbound Clone20 DNA and with 5 nM HMfA. Histograms were fitted with a skewed normal distribution. Insert: pairwise distribution plot of the difference between the two end-to-end distance populations. Histogram was fitted with a Gaussian distribution.

The Clone20 DNA sequence is recognized by HMfA/B derivatives impaired in hypernucleosome formation

Previously, we found that the HMfA and HMfB derivatives $HMfA_{K31A\ E35A}$ and $HMfB_{D14A\ K30A\ E34A}$ require higher concentrations to fully compact nonspecific DNA and that the resulting hypernucleosome is less stable compared to the wildtype, especially for HMfB (12). These observations underscore the importance of the mutated residues in stabilizing hypernucleosome structure via electrostatic interactions between hypernucleosomal stacks. These HMfA and HMfB derivatives have additional relevance as mimics of histone variants from other species that lack stacking interactions (13). We examined if these proteins still exhibit specific binding to the Clone20 sequence. Both derivatives compact the Clone20 DNA into a tetramer at comparable protein concentrations as the wildtype proteins (figure 4.2A and B). This result indicates that HMfA and HMfB recognize the Clone20 site independent of their stacking interactions, as expected. Nonspecific binding, leading to hypernucleosome formation occurs at >125 nM for $HMfA_{K31A\ E35A}$ and >80 nM for $HMfB_{D14A\ K30A\ E34A}$. These concentrations are higher than observed for the wildtype proteins, which indicates delayed hypernucleosome formation attributed to the missing stacking interactions. Also the transition from tetramer to hypernucleosome is more gradual for the histone derivatives than for the wildtype proteins. The distinct binding at a specific DNA sequence by archaeal histones at concentrations below the effective K_D for nonspecific compaction implies that specific sites may have a functional role in archaea. Also, the difference in affinity for the Clone20 sequence between HMfA and HMfB (and their mutated derivatives) supports the hypothesis that histone variants have distinct functional roles, potentially in transcription regulation (13, 17, 27).

Histone tetramers have increased affinity for Clone20 and can bind in different conformations

To further investigate the properties and affinities of the respective tetramers formed on the different DNA sequences, we used microscale thermophoresis (MST) with short (78 bp) DNA substrates designed to accommodate maximally two HMf dimers (figure 4.3 and figure S4.2) and fitted the binding curves with the Hill binding model (figure S4.3). For HMfA, the affinity for Clone20 DNA is higher than for nonspecific DNA, while cooperativity stayed the same (table 4.1). Judged by the in general higher ΔF_{norm}, for HMfA compared to HMfB, the protein-DNA complexes formed by HMfB are more compact than those formed by HMfA (figure 4.3A and B). This agrees with earlier observations where the hypernucleosome formed by HMfB is more compact and has a higher stacking energy than that formed by HMfA (12). Also the ΔF_{norm} at the highest protein concentration of

nonspecific DNA is higher than that of Clone20 for both proteins, indicating a more compact structure formed on the specific site.

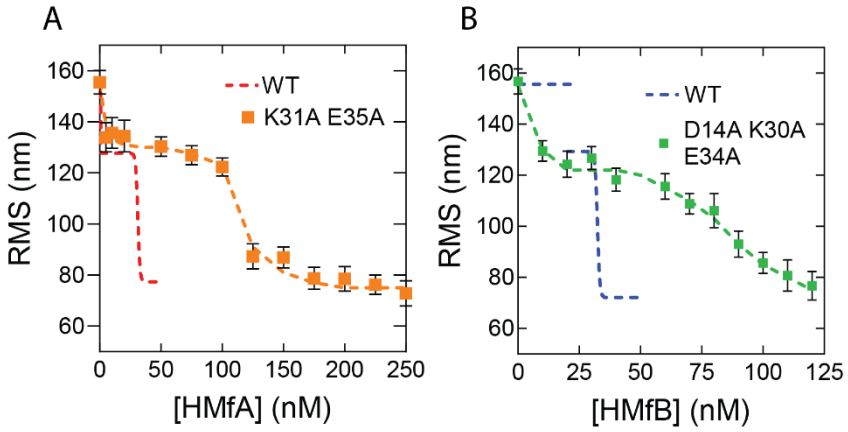

Figure 4.2 Histone derivatives HMfA$_{K31A\ E35A}$ and HMfB$_{D14A\ K30A\ E34A}$ recognize the Clone20 sequence A) Root mean square displacement (RMS) values of Clone20 DNA tethers with HMfA$_{K31A\ E35A}$ or B) HMfB$_{D14A\ K30A\ E34A}$ as measured by TPM in 50 mM Tris-HCl pH 7, 75 mM KCl. Wildtype data was reproduced from figure 4.1. Histograms were fitted to a Gaussian distribution. Error bars represent the propagated standard deviation of at least two replicates. Dashed lines are to guide the eye.

HMfB exhibited two-step behavior on Clone20 DNA (figure 4.3B). The first state, attributed to specific binding to the Clone20 site, resulted in a negative ΔF_{norm}, so a more compact structure compared to unbound DNA. While a slight decrease in ΔF_{norm} was observed for HMfA as well (figure 4.3A), it was less pronounced than for HMfB and we were unable to fit any binding constant. The second state showed increasing ΔF_{norm} and corresponds to nonspecific binding. There are multiple possibilities to explain this two-step behavior. An HMfB tetramer could bind first, forming a compact bent structure. At higher protein concentrations, a hexamer with suboptimal protein-DNA interaction interface might assemble on the DNA. This would be a metastable structure as the DNA substrate is shorter than expected for hexamer binding (78 bp compared to 90 bp theoretically). Another option might be binding of an HMfB dimer, which bends the DNA resulting in the observed compact structure. The second binding regime would then represent tetramer (or even hexamer) formation on the DNA substrate.

In order to be able to distinguish between the two possible models described above, we performed MST experiments with derivatives of the Clone20 DNA substrate, where only either the left (Clone20L) or the right (Clone20R) site of the sequence is present (table S4.1). The other half was replaced with the nonspecific DNA sequence. For HMfA, this leads to a generally lower affinity than for the entire Clone20 sequence but higher than for nonspecific DNA (figure 4.3C and table 4.1). HMfB still shows the two-step binding behavior mainly on Clone20R. (figure 4.3D and table 4.1). This suggests that either a dimer is binding, and therefore half of the Clone20 sequence is sufficient, or that half the site is enough to position a tetramer on the DNA. Strikingly, the ΔF_{norm} at the highest HMfB concentration increased compared to the fully nonspecific and Clone20 substrates, especially for Clone20R. This means that the resulting structure is less compact or the DNA is more permissive to HMfB multimerization. TPM experiments with only Clone20L or Clone20R present were in agreement with the MST experiments (figure S4.4). For HMfA, tetramer binding cannot be observed for both half sites; instead, HMfA shows similar binding behavior as on nonspecific DNA (figure S4.4A). Tetrameric complex formation by HMfB, as observed by having two populations (figure 4.1B), was only found on Clone20R (figure S4.4B), but with a slightly reduced affinity compared to the full Clone20 site (Kd of 28.8 ± 1.1 nM versus 21 ± 0.2 nM) (figure S4.4C). We calculated the end-to-end distance of the two observed populations and found a pairwise distance of 27.6 ± 11 nm or 81 ± 32 bp, confirming that a tetramer is most likely bound to the Clone20R site (figure S4.4D). The RMS of Clone20L for 10-30 nM HMfB is slightly lower than unbound Clone20R DNA, but higher than the second population corresponding to the tetrameric complex (figure S4.4B). This could be suggestive of binding of a dimer, but the resolution of TPM experiments is not high enough to confirm this.

MST experiments with the HMf derivatives showed that $HMfA_{K31A\ E35A}$ had too low an affinity for both DNA substrates to be reliably fitted (figure S4.2A and S4.3). $HMfB_{D14A\ K30A\ E34A}$ showed increased aggregation in MST experiments; therefore, 0.2% Tween20 had to be added (figure S4.2B). Most likely, this is an artefact of using protein concentrations in the micromolar range for MST experiments in comparison to nanomolar for TPM. To be able to compare, also an HMfB wildtype titration with nonspecific DNA was done in the presence of 0.2% Tween20. The affinities of $HMfB_{D14A\ K30A\ E34A}$ for both DNA substrates are similar (table 4.1) and qualitatively the curves are also comparable. No two-step behavior was observed on the Clone20 DNA substrate.

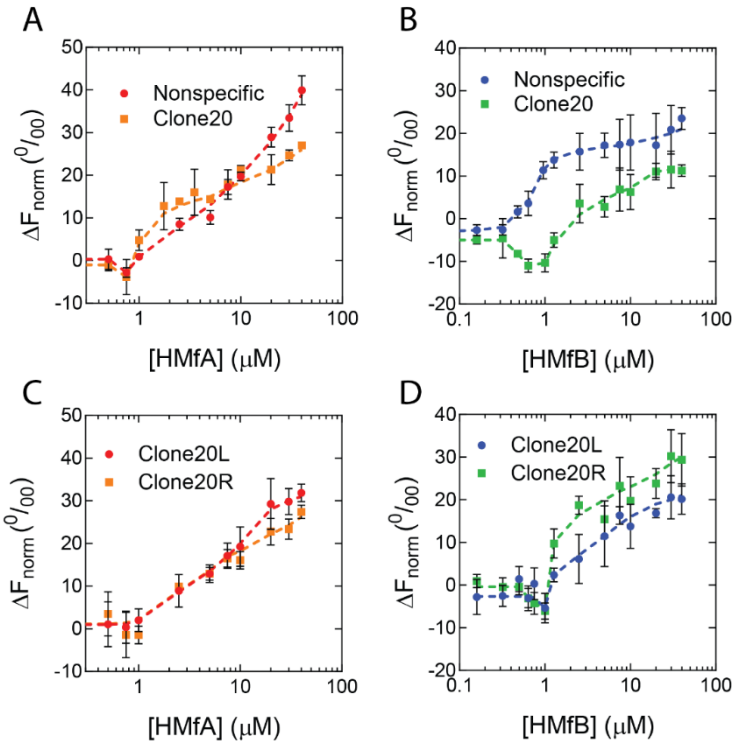

Figure 4.3 Binding of HMf proteins to short DNA substrates using Microscale Thermophoresis. Normalized thermophoresis curves of Nonspecific or Clone20 DNA as a function of A) HMfA or B) HMfB or of Clone20L or Clone20R as a function of C) HMfA or D) HMfB. Error bars indicate the standard deviation of three independent measurements. Dashed lines are lines to guide the eye.

Table 4.1: Binding affinities (K_D) and Hill binding coefficients (h) of HMf to 78 bp DNA substrates. The values were determined by fitting of the MST data to the Hill binding model.

Protein	DNA substrate	K_D (μM)	h
HMfA	Nonspecific	16.5 ± 6.7	1.08 ± 0.19
	Clone20	3.75 ± 0.10	1.08 ± 0.21
	Clone20L	8.09 ± 1.4	1.23 ± 0.16
	Clone20R	7.34 ± 2.8	0.992 ± 0.21
HMfB	Nonspecific	0.915 ± 0.089	3.30 ± 0.97
	Nonspecific + 0.2% Tween20	19.6 ± 5.4	1.22 ± 0.15
	Clone20 specific	0.243 ± 0.16	1.24 ± 0.91
	Clone20 nonspecific	2.99 ± 0.81	1.44 ± 0.47
	Clone20L specific	n.a.	n.a.
	Clone20L nonspecific	4.11 ± 1.0	1.30 ± 0.38
	Clone20R specific	0.660 ± 0.047	7.45 ± 3.5
	Clone20R nonspecific	12.3 ± 27	0.564 ± 0.30
HMfA $_{K31A\ E35A}$	Nonspecific	n.a.	n.a.
	Clone20	n.a.	n.a
HMfB $_{D14A\ K30A\ E34A}$	Nonspecific + 0.2% Tween 20	22.0 ± 0.71	4.09 ± 0.62
	Clone20 + 0.2% Tween 20	19.6 ± 1.194	3.51 ± 0.77

Discussion

A DNA substrate containing the artificial high-affinity sequence Clone20 is compacted by *M. fervidus* histones in two distinct steps, representing two distinct types of complexes. HMf is a dimer in solution, even at high concentrations above 1 mg/mg (figure S4.5). We propose a model where the first step is binding of a dimer to the DNA, directly followed by recruitment of the second dimer to form a stable tetrameric complex. Recruitment of the second dimer is cooperative due to interactions with both DNA and the dimer already bound to the DNA. We found that the tetramer on the Clone20 site exists in a distinct structural, possibly more closed, state incompatible with hypernucleosome formation. Therefore the high affinity sequence is unable to act as a nucleation site. This closed state is in equilibrium with the more open state, which is geometrically permissive to multimerization (figure 4.4). On nonspecific DNA, most likely only open tetramers can bind, which explains why such dynamics at the dimer-dimer interface were not observed with molecular dynamics simulations of HMfB (28).

Globally, archaeal histone variants can be divided into three functional groups. The first group consists of histones that contain the amino acid residues involved in both dimer-dimer interactions (tetramer formation) and stacking interactions (hypernucleosome formation). Members of this group include the archaeal model histones HMfA and HMfB, and HTkB from *Thermococcus kodakarensis* (13). Generally,

they show cooperative extension on DNA, resulting in hypernucleosome formation once the first tetramer is in the right position (figure 4.4). However, differences in DNA binding properties between members of this group do exist, and environmental or growth phase related response may bias the expression of one histone variant over another, resulting in changes in (local) chromosome organization, potentially translating into an altered expression of genes (14). Hypernucleosome formation by HMfB is more cooperative than for HMfA, and the level of DNA compaction achieved is slightly higher (12). HMfA, on the other hand, has a higher affinity for the Clone20 sequence in the context of longer DNA (figure 4.1C). This finding was unexpected as the Clone20 sequence was obtained via SELEX optimization with HMfB. Nevertheless, this finding may be indicative of distinct functions in chromosome organization. HMfA may more effectively position tetramers at specific locations on the genome, setting boundaries for hypernucleosome formation and the action of other chromatin proteins, whereas HMfB forms predominantly hypernucleosomes. However, this is contradicted by experiments on shorter DNA, such as in Bailey et al. (23) and in our MST experiments (figure 4.3). Bailey et al. found that the difference in affinity between HMfA and HMfB was at least partially dependent on the C-terminal residues of helix α3, which does not make direct contact with the DNA, but is important in dimer-dimer interactions (29). Also, it has been proposed before that changes in the dimer-dimer interface might result in tetramers that bend the DNA with either a negative or positive supercoil akin to the eukaryotic $(H3\text{-}H4)_2$ tetramer (30–32). Potentially, this interface is involved in forming the closed and open conformation of the HMf tetramer, proposed here (figure 4.4). This would require extensive structural follow-up studies on the different protein-DNA complexes. Also, the genomic context and amount of other proteins bound to the DNA might be of importance. Synergistic or antagonistic interplay between histones and other architectural proteins could be expected, but has not been studied in detail so far.

The second group of histone variants consists of histones that are able to form dimers and tetramers, but lack the stacking interactions implied in the stabilization of hypernucleosomes. Examples are the histone derivatives $HMfA_{K31A\ E35A}$ and $HMfB_{D14A\ K30A\ E34A}$ and the *Haloredivivus* sp. G17 and *Methanococcoides methylutens* histones (13). They are able to recognize a specific DNA sequence in a similar concentration range as histones from the first group (figure 4.2), but hypernucleosome formation will occur at higher concentration and less cooperatively due to the absence of stabilizing stacking interactions. The presence of a tetramer formed by these histones could act as a roadblock for hypernucleosome progression or act as a capstone by preventing further multimerization on one side of the hypernucleosome (figure 4.4). In this way, changing

expression levels of histone variants might affect DNA compaction and potentially transcriptional regulation.

The last group of histone variants lacks the residues implied in dimer-dimer interactions. Therefore, these histones are likely bound as dimers only or, when incorporated in a heterodimer, prevent a hypernucleosome from further multimerization and thus act as capstones (17). They may have intact stacking interactions, potentially permitting the formation of hypernucleosomes (of reduced stability compared to the model histones HMfA and HMfB). Some predicted members of this group are *Ca. Lokiarchaeota* GC14_75 HLkE and *Nanosalina* J07AB43 HB (13).

Clone20 can be regarded as the archaeal counterpart of the 601 nucleosome positioning sequence, a sequence that energetically favors nucleosome formation. The 601 sequence is often used in studies on eukaryotic nucleosomes (33–36). However, sequences with high similarity to Clone20 and 601 sequences have thus far not been identified in genomes, and affinity for the 601 sequence was found to be much higher than for natural sequences (37). Based on our results of HMfB binding to the right site of Clone20 (figure 4.3D and figure S4.4), it might be possible that a smaller site is sufficient to act as a high-affinity sequence. This would increase the possibility of encountering such a sequence in genomes.

Taken together, the interplay between archaeal histone variants and specific genomic sequences can result in the formation of structurally different protein-DNA complexes. Positioning of these complexes along the genome might have a potential to act in archaeal transcription regulation.

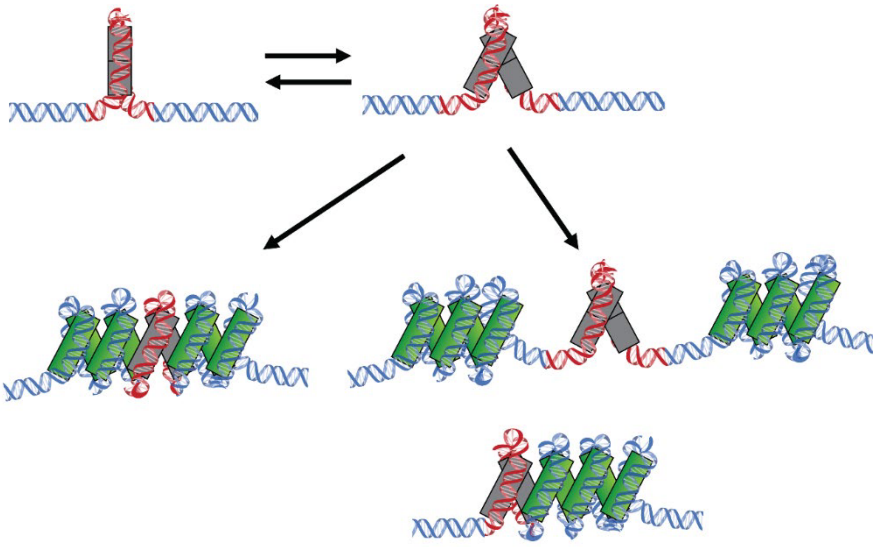

Figure 4.4 Mechanisms of HMf tetramers binding to specific DNA sequences followed by hypernucleosome formation. HMf tetramers bind to the Clone20 sequence and form a closed complex incompatible with further multimerization (top left). This structure can dynamically open and close (top right). The open structure can facilitate hypernucleosome formation (bottom left). If histone variants are bound that lack either stacking interactions or dimer-dimer interactions are bound, this tetramer could potentially act as a barrier of hypernucleosome progression or act as a 'capstone' (bottom right). Including different homo- and heterodimers into one structure could also result in limited extension of the hypernucleosome.

References

1. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251–260.
2. Sandman,K. and Reeve,J.N. (2005) Archaeal chromatin proteins: Different structures but common function? *Curr. Opin. Microbiol.*, **8**, 656–661.
3. Wilkinson,S.P., Ouhammouch,M. and Geiduschek,E.P. (2010) Transcriptional activation in the context of repression mediated by archaeal histones. *Proc. Natl. Acad. Sci.*, **107**, 6777–6781.
4. Peeters,E., Driessen,R.P.C., Werner,F. and Dame,R.T. (2015) The interplay between nucleoid organization and transcription in archaeal genomes. *Nat. Rev. Microbiol.*, **13**, 333–341.
5. Henneman,B. and Dame,R.T. (2015) Archaeal histones: dynamic and versatile genome architects. *AIMS Microbiol.*, **1**, 72–81.
6. Rojec,M., Hocher,A., Stevens,K.M., Merkenschlager,M. and Warnecke,T. (2019) Chromatinization of Escherichia coli with archaeal histones. *Elife*, **8**.
7. Sanders,T.J., Ullah,F., Gehring,A.M., Burkhart,B.W., Vickerman,R.L., Fernando,S., Gardner,A.F., Ben-Hur,A. and Santangelo,T.J. (2021) Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in Thermococcus kodakarensis. *Front. Microbiol.*, **12**, 1071.
8. Ammar,R., Torti,D., Tsui,K., Gebbia,M., Durbic,T., Bader,G.D., Giaever,G. and Nislow,C. (2012) Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife*, **1**, e00078.
9. Pereira,S.L., Grayling,R.A., Lurz,R. and Reeve,J.N. (1997) Archaeal nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 12633–12637.
10. Maruyama,H., Harwood,J.C., Moore,K.M., Paszkiewicz,K., Durley,S.C., Fukushima,H., Atomi,H., Takeyasu,K. and Kent,N.A. (2013) An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. *EMBO Rep.*, **14**, 711–7.
11. Mattiroli,F., Bhattacharyya,S., Dyer,P.N., White,A.E., Sandman,K., Burkhart,B.W., Byrne,K.R., Lee,T., Ahn,N.G., Santangelo,T.J., *et al.* (2017) Structure of histone-based chromatin in Archaea. *Science (80-.).*, **357**, 609–612.
12. Henneman,B., Brouwer,T.B., Erkelens,A.M., Kuijntjes,G.-J., van Emmerik,C., van der Valk,R.A., Timmer,M., Kirolos,N.C.S., van Ingen,H., van Noort,J., *et al.* (2021) Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res.*, **49**, 4338–4349.
13. Henneman,B., van Emmerik,C., van Ingen,H. and Dame,R.T. (2018) Structure and function of archaeal histones. *PLoS Genet.*, **14**, e1007582.
14. Sandman,K., Grayling,R.A., Dobrinski,B., Lurz,R. and Reeve,J.N. (1994) Growth-phase-dependent synthesis of histones in the archaeon Methanothermus fervidus. *Proc. Natl. Acad. Sci.*, **91**, 12624–12628.

15. Marc,F., Sandman,K., Lurz,R. and Reeve,J.N. (2002) Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J. Biol. Chem.*, **277**, 30879–30886.
16. Musgrave,D., Forterre,P. and Slesarev,A. (2000) Negative constrained DNA supercoiling in archaeal nucleosomes. *Mol. Microbiol.*, **35**, 341–349.
17. Stevens,K.M., Swadling,J.B., Hocher,A., Bang,C., Gribaldo,S., Schmitz,R.A. and Warnecke,T. (2020) Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 33384–33395.
18. Warnecke,T., Becker,E.A., Facciotti,M.T., Nislow,C. and Lehner,B. (2013) Conserved Substitution Patterns around Nucleosome Footprints in Eukaryotes and Archaea Derive from Frequent Nucleosome Repositioning through Evolution. *PLoS Comput. Biol.*, **9**.
19. Nalabothula,N., Xi,L., Bhattacharyya,S., Widom,J., Wang,J.P., Reeve,J.N., Santangelo,T.J. and Fondufe-Mittendorf,Y.N. (2013) Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics*, **14**, 391.
20. Segal,E. and Widom,J. (2009) Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.*, **19**, 65–71.
21. Pereira,S.L. and Reeve,J.N. (1999) Archaeal nucleosome positioning sequence from Methanothermus fervidus. *J. Mol. Biol.*, **289**, 675–681.
22. Bailey,K.A., Pereira,S.L., Widom,J. and Reeve,J.N. (2000) Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. *J. Mol. Biol.*, **303**, 25–34.
23. Bailey,K.A., Marc,F., Sandman,K. and Reeve,J.N. (2002) Both DNA and Histone Fold Sequences Contribute to Archaeal Nucleosome Stability. *J. Biol. Chem.*, **277**, 9293–9301.
24. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison 3rd,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343–345.
25. van der Valk,R.A., Laurens,N. and Dame,R.T. (2017) Tethered particle motion analysis of the DNA binding properties of architectural proteins. In *Methods in Molecular Biology*. Humana Press Inc., Vol. 1624, pp. 127–143.
26. Henneman,B., Heinsman,J., Battjes,J. and Dame,R.T. (2018) Quantitation of DNA-binding affinity using tethered particle motion. In *Methods in Molecular Biology*. Humana Press Inc., Vol. 1837, pp. 257–275.
27. Sandman,K. and Reeve,J.N. (2006) Archaeal histones and the origin of the histone fold. *Curr. Opin. Microbiol.*, **9**, 520–525.
28. Bowerman,S., Wereszczynski,J. and Luger,K. (2021) Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways. *Elife*, **10**, e65587.

29. Decanniere,K., Babu,A.M., Sandman,K., Reeve,J.N. and Heinemann,U. (2000) Crystal structures of recombinant histones HMfA and HMfB, from the hyperthermophilic archaeon Methanothermus fervidus. *J. Mol. Biol.*, **303**, 35–47.
30. Hamiche,A., Carot,V., Alilat,M., De Lucia,F., O'Donohue,M.F., Révet,B. and Prunell,A. (1996) Interaction of the histone (H3-H4)2 tetramer of the nucleosome with positively supercoiled DNA minicircles: Potential flipping of the protein from a left- to a right-handed superhelical form. *Proc. Natl. Acad. Sci.*, **93**, 7588–7593.
31. Hamiche,A. and Richard-Foy,H. (1998) The Switch in the Helical Handedness of the Histone (H3-H4)2 Tetramer within a Nucleoprotein Particle Requires a Reorientation of the H3-H3 Interface. *J. Biol. Chem.*, **273**, 9261–9269.
32. Sandman,K. and Reeve,J.N. (2000) Structure and functional relationships of archaeal and eukaryal histones and nucleosomes. *Arch. Microbiol.*, **173**, 165–169.
33. Lowary,P.T. and Widom,J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.*, **276**, 19–42.
34. Thåström,A., Lowary,P.T., Widlund,H.R., Cao,H., Kubista,M. and Widom,J. (1999) Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.*, **288**, 213–229.
35. Vasudevan,D., Chua,E.Y.D. and Davey,C.A. (2010) Crystal Structures of Nucleosome Core Particles Containing the '601' Strong Positioning Sequence. *J. Mol. Biol.*, **403**, 1–10.
36. Eslami-Mossallam,B., Schiessel,H. and van Noort,J. (2016) Nucleosome dynamics: Sequence matters. *Adv. Colloid Interface Sci.*, **232**, 101–113.
37. van der Heijden,T., van Vugt,J.J.F.A., Logie,C. and Van Noort,J. (2012) Sequence-based prediction of single nucleosome positioning and genome-wide nucleosome occupancy. *Proc. Natl. Acad. Sci. U. S. A.*, **109**.

Supplementary figures

Table S4.1: Sequences of DNA substrates used for MST experiments. The Cy5-label is attached on the '5-end of the top strand. For Clone20L and Clone20R, the nonspecific part is indicated in black and part of the Clone20 substrate in red.

Name DNA substrate	Sequence (5'-3')
Nonspecific	CGGCGCAAATTCGTGACCAGTTGCATCAGCTGCGTGAGCTGTTTAT CGCAGCATCGTAACAGGATAGTGAAGAAGACT
Clone20	GGATCCCTGTCGGCACAGTTGAGCGATCAAAAACGCCGTAGAACG CTTTAATTGATAATCAAAGGCCGCAGAGAGCTC
Clone20L	GGATCCCTGTCGGCACAGTTGAGCGATCAAAAACGCCGTAGATTA TCGCAGCATCGTAACAGGATAGTGAAGAAGACT
Clone20R	CGGCGCAAATTCGTGACCAGTTGCATCAGCTGCGTGCGTAGAACG CTTTAATTGATAATCAAAGGCCGCAGAGAGCTC

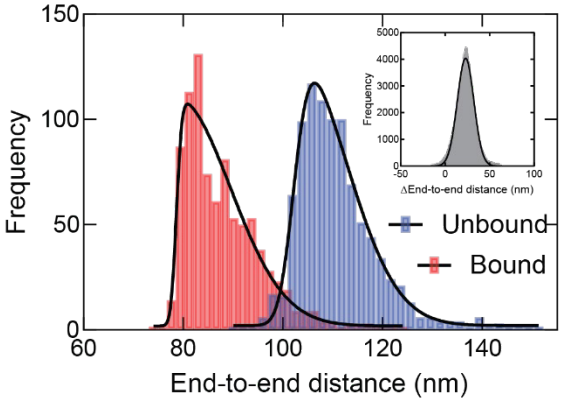

Figure S4.1 Calculated end-to-end distances for the unbound and bound population of 21 nM HMfB on Clone20 DNA. Histograms were fitted with a skewed normal distribution, resulting in end-to-end distances of 102 ± 10 nm and 78.6 ± 11 nm for unbound and bound DNA respectively. Insert: pairwise distribution plot of the differences between the two end-to-end distance peaks. Histogram was fitted with a Gaussian distribution resulting in a difference of 23.0 ± 9.3 nm.

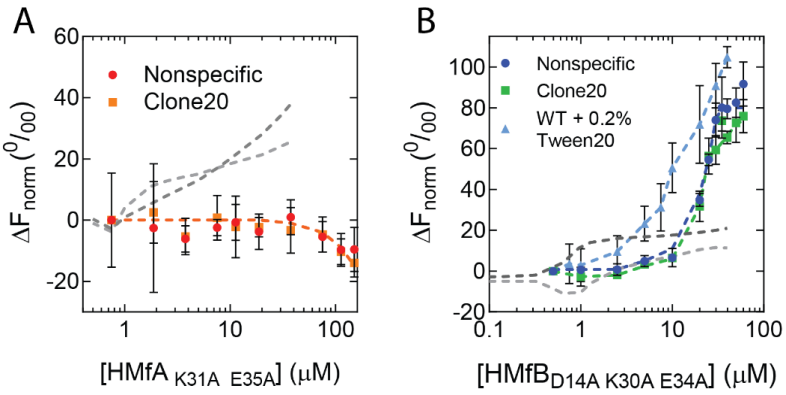

Figure S4.2: Binding of HMf derivatives to Nonspecific and Clone20 DNA substrates. Normalized thermophoresis curves of Nonspecifc or Clone20 DNA as a function of A) $HMfA_{K31A\ E35A}$ or B) $HMfB_{D14A\ K30A\ E34A}$. For HMfB WT, an extra curve on Nonspecific DNA was measured with 0.2% Tween20 in the buffer. Error bars indicate the standard deviation of three independent measurements. Dashed lines are lines to guide the eye. Wildtype curves from WT proteins (figure 4.3) are included in dark (Nonspecific) and light (Clone20) grey for easy comparison.

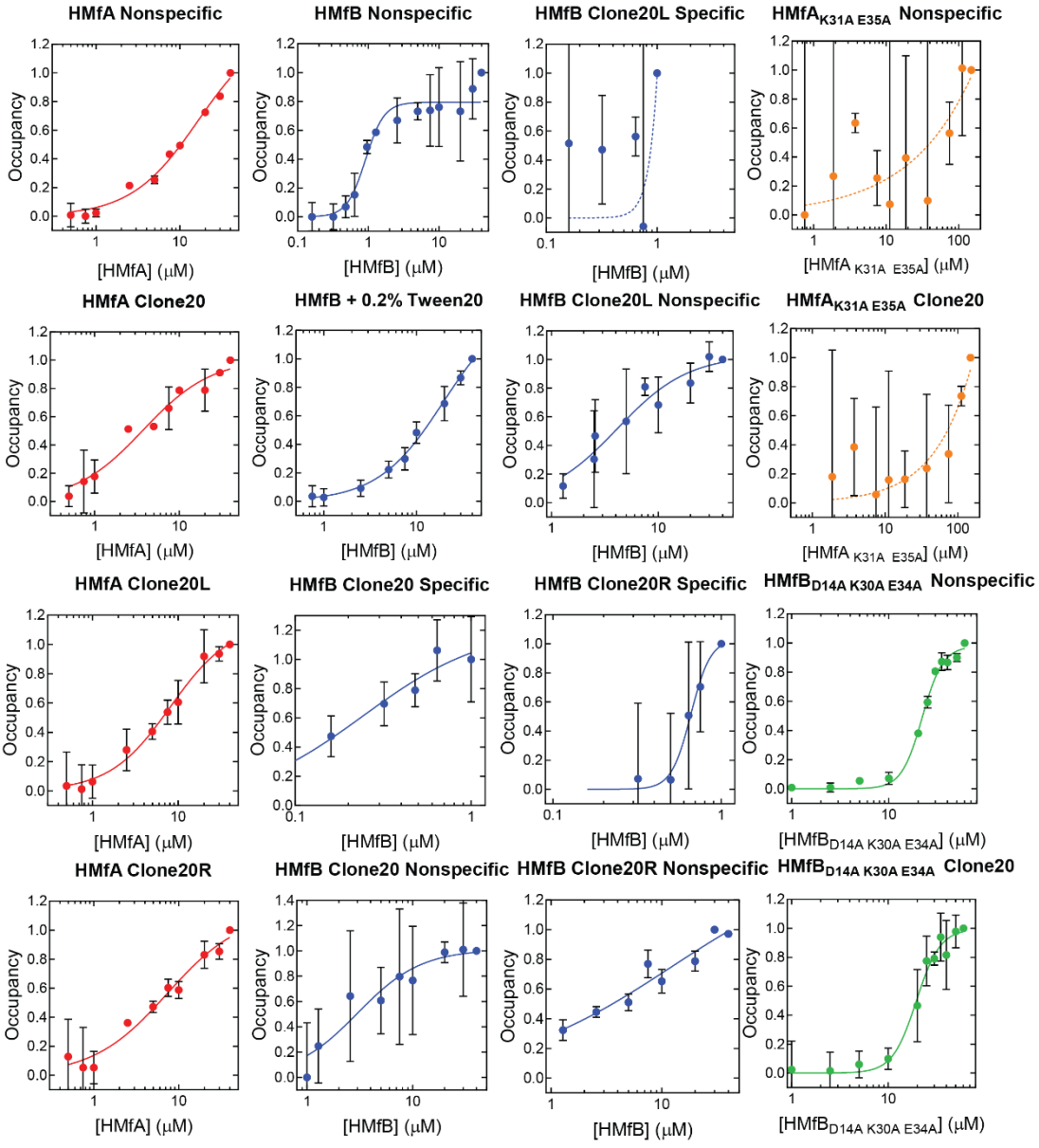

Figure S4.3: Hill fits of MST curves Occupancy was fitted against protein concentration where the occupancy at the highest concentration was set as 1.0. The dotted lines indicate fits that did not result in any reliable results.

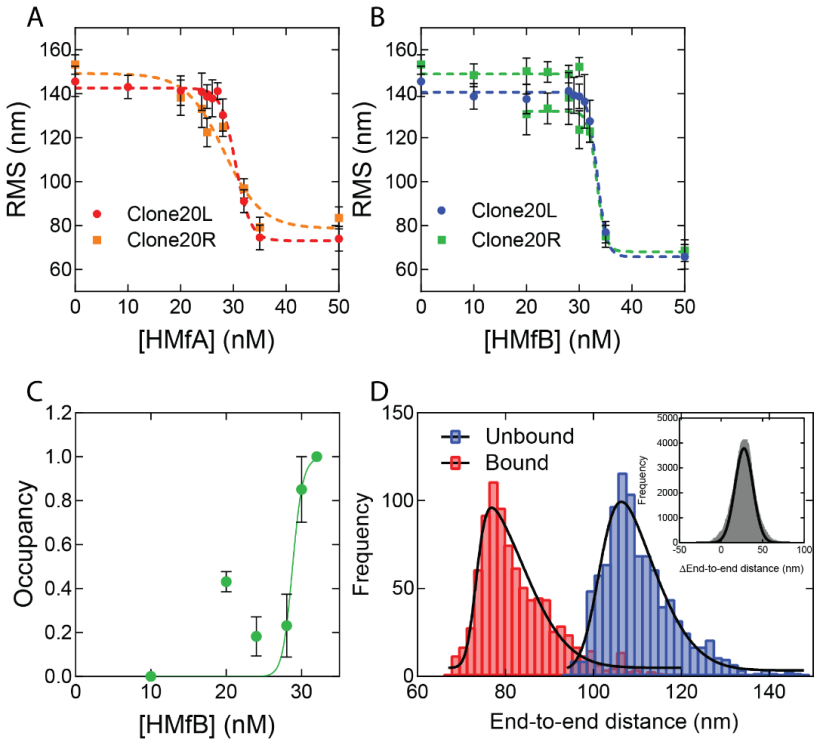

Figure S4.4 Binding of HMfA and HMfB to Clone20L or Clone20R in TPM experiments Root mean square displacement (RMS) of Clone20L and Clone20R DNA incubated with A) HMfA or B) HMfB in 50 mM Tris-HCl pH 7, 75 mM KCl. Histograms were fitted to a Gaussian distribution. Error bars represent the propagated standard deviation of two replicates. Dashed lines are lines to guide the eye. C) Binding curve of HMfB on Clone20R DNA. Data point were fitted using the Hill binding model. D) Calculated end-to-end distance for bound and unbound Clone20R DNA incubated with 30 nM HMfB. Histograms were fitted with a skewed normal distribution. Insert: pairwise distribution plot of the difference between the two populations. Histogram was fitted with a Gaussian distribution.

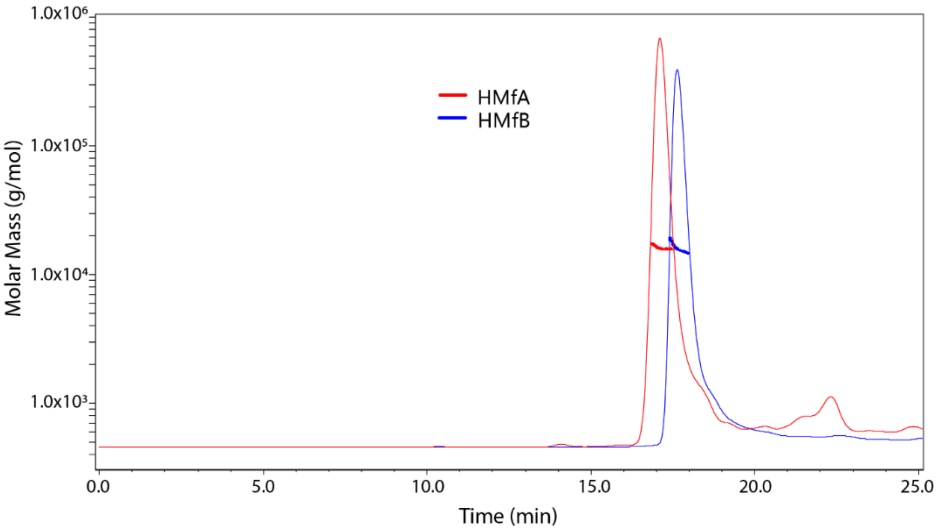

Figure S4.5 HMfA and HMfB are dimers in solution. SEC-MALS result of at least 1 mg/ml HMfA and HMfB. The determined molecular weight for HMfA was 16.3 ± 0.3 kDa and for HMfB 16.0 ± 0.1 kDa. The theoretical monomer mass of HMfA is 7.5 kDa and for HMfB 7.7 kDa.