
Universiteit
Leiden
The Netherlands

Histone-DNA assemblies in archaea : shaping the genome on the edge of life

Henneman, B.

Citation
Henneman, B. (2019, December 4). *Histone-DNA assemblies in archaea : shaping the genome on the edge of life*. Retrieved from https://hdl.handle.net/1887/81191

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/81191

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle http://hdl.handle.net/1887/81191 holds various files of this Leiden University dissertation.

Author: Henneman, B.
Title: Histone-DNA assemblies in archaea : shaping the genome on the edge of life
Issue Date: 2019-12-05

HISTONE-DNA ASSEMBLIES IN ARCHAEA

SHAPING THE GENOME ON THE EDGE OF LIFE

HISTONE-DNA ASSEMBLIES IN ARCHAEA

BRAM HENNEMAN

BRAM HENNEMAN

HISTONE-DNA ASSEMBLIES IN ARCHAEA

Shaping the genome on the edge of life

Bram Henneman

Histone-DNA assemblies in archaea - Shaping the genome on the edge of life

Bram Henneman

Doctoral Thesis, Leiden University, 2019

ISBN: 978-94-6332-577-6

© Bram Henneman, 2019. All rights reserved.

Cover: RAUW Grafisch Design | Sem Henneman

Printed by: GVO drukkers & vormgevers

HISTONE-DNA ASSEMBLIES IN ARCHAEA

Shaping the genome on the edge of life

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 5 december 2019
klokke 10:00 uur

door

Bram Henneman

geboren te Tilburg
in 1990

Promotors:	Dr. R. T. Dame
	Prof. Dr. M. Ubbink
Promotiecommissie:	Prof. Dr. H. S. Overkleeft (voorzitter)
	Prof. Dr. A. Briegel (secretaris)
Overige commissieleden:	Prof. Dr. Ir. E. Peeters (Vrije Universiteit Brussel)
	Prof. Dr. Ir. T. J. G. Ettema (Wageningen University)
	Prof. Dr. Ir. S. J. T. van Noort
	Dr. F. Mattiroli (Hubrecht Institute)

This work was supported by the Human Frontier Science Program (HFSP) (RGP0014/2014)

TABLE OF CONTENTS

List of abbreviations 8

CHAPTERS

1 Introduction 11

2 Structure and function of archaeal histones 31

3 Quantitation of DNAbinding affinity using Tethered Particle Motion 73

4 Mechanical and structural properties of archaeal hypernucleosomes 97

5 DNA binding by histones from *Ca.* Nanohaloarchaeota 125

6 Specific DNA binding of archaeal histones 141

7 General discussion 151

References 159

Summary & Nederlandse samenvatting 180

Curriculum Vitae 186

List of publications 187

Acknowledgments 188

LIST OF ABBREVIATIONS

AFM	atomic force microscopy
AFS	acoustic force spectroscopy
Alba	acetylation lowers binding affinity
ANME	anaerobic methanotrophs
BB	bead buffer
BRE	B recognition element
Ca.	*Candidatus*
CARD-FISH	catalyzed reporter deposition fluorescence *in situ* hybridization
DPANN	superphylum originally named after Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota
Dps	DNA-binding protein from starved cells
EB	experimental buffer
EMSA	electrophoretic mobility shift assay
FECA	first eukaryotic common ancestor
FJC	freely-jointed chain
F_{norm}	normalized fluorescence
G1P	*sn*-glycerol-1-phosphate
G3P	*sn*-glycerol-3-phosphate
GRN	gene regulatory network
GTF	general transcription factor
H-NS	histone-like nucleoid-structuring protein
HADDOCK	high-ambiguity driven docking
HU	histone-like protein from *Escherichia coli* strain U93
IHF	integration host factor
Inr	initiator element
ITC	isothermal titration calorimetry
K_D	binding constant
k_{off}	off-rate constant
k_{on}	on-rate constant
LECA	last eukaryotic common ancestor
L_p	persistence length
LUCA	last universal common ancestor
MC1	methanogen chromosomal protein 1
MNase	micrococcal nuclease
MST	microscale thermophoresis
NAP	nucleoid-associated protein
PCR	polymerase chain reaction

pI	isoelectric point
PIC	preinitiation complex
PIFE	protein induced fluorescence enhancement
PPE	promoter proximal DNA element
PTM	post-translational modification
RMS	root mean square displacement
RNAP	RNA polymerase
SAMP	small archaeal modifier protein
SD	standard deviation
SELEX	systematic evolution of ligands by exponential enrichment
smFRET	single molecule Förster resonance energy transfer
SPR	surface plasmon resonance
sRNA	small RNA
TACK	superphylum originally named after Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota
TBP	TATA-binding protein
TF	transcription factor
TFB	transcription factor B
TFEα	transcription factor E
TPM	tethered particle motion
TSS	transcription start site
WLC	worm-like chain
ΔF_{norm}	difference in normalized fluorescence

CHAPTER 1

Introduction

This chapter is based on:

Bram Henneman & Remus T. Dame, (2015) *Archaeal histones: dynamic and versatile genome architects*. AIMS Microbiology. doi: 10.3934/microbiol.2015.1.72

and

Bram Henneman, Clara van Emmerik, Hugo van Ingen & Remus T. Dame, (2018) *Structure and function of archaeal histones*. PLoS Genetics. doi: 10.1371/journal.pgen.1007582

Organisms from the domain archaea - one of the three domains of life - share similarities with both bacteria and eukaryotes (1). Genomic comparisons demonstrate that bacteria and archaea share a common ancestor; eukaryotes are to date classified as being part of the archaeal branch (2-4). The archaeal domain comprises single-cellular organisms found in diverse habitats. Although archaea and bacteria have common features, such as a circular genome and the absence of a nucleus, at the genetic level archaea seem to be more related to eukaryotes. Amongst others, archaeal RNA polymerase, a key component of cellular life in all domains, is more similar to RNA polymerase from eukaryotes than bacterial RNA polymerase (5, 6). Archaeal ribosomes share their size and structural core with bacterial ribosomes, but are more similar to eukaryotic ribosomes when it comes to protein and rRNA sequence and some specific domains (7-9). Also, some cellular processes thought to be unique to eukaryotes, such as endosomal sorting and the ubiquitin system, have been identified in some archaea (10).

Chromatin proteins are found in every domain of life. Bacteria express DNA bending and DNA bridging proteins, such as HU and H-NS, to structure and functionally organize the genome and to regulate genome activity (11, 12). In eukaryotes and most archaeal lineages, histones are responsible for packaging

and compaction of the DNA. Histones act together with other chromatin proteins, such as linker histones and histone variants in eukaryotes, Alba, MC1 and Sso10a in archaea, and SMC in both domains (13). These proteins jointly shape the genome, which has an effect on transcription regulation and has a role in response to environmental cues.

This thesis focuses on the histones of archaea and the roles that they have in chromatin architecture and transcription regulation. In this chapter, I will extensively introduce archaea as a domain of life, discussing their habitats and their metabolism. Next, I will discuss their genomic features and evolutionary relations to each other, as well as their mechanisms of transcription regulation. In CHAPTERS 2-6 I will further describe the evolution of histones and molecular mechanisms of genome compaction by histones.

Archaea as a domain of life

Evolution and the tree of life

Archaea were identified as a domain that is separate from bacteria by Carl Woese and George Fox (14). They proposed a three-domain model, in which bacteria, archaea and eukaryotes are monophyletic groups sharing a common ancestor (15). As the availability of information on phylogenetic and genomic features of archaea grew, the model was changed to a two-domain model, in which eukaryotes are a sister group of a phylum that is now known as Crenarchaeota (16). Together with Euryarchaeota, these phyla were the only known archaeal phyla for a long time. This changed with the discovery of Thaumarchaeota, *Candidatus* (*Ca.*) Aigarchaeota and *Ca.* Korarchaeota (17-19), which together form the TACK superphylum (an acronym of Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota). At that point, eukaryotes were hypothesized to branch from the TACK superphylum. Later, metagenomic sequencing boosted the discovery of new archaeal species and phyla. This resulted in the discovery of the DPANN superphylum (an acronym of Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota), and most recently the Asgard superphylum, which revealed that eukaryotes are most closely related to extant members of Asgard archaea. Recent additions to the archaeal branch of the tree of life include the plyla *Ca.* Heimdallarchaeota, *Ca.* Odinarchaeota, *Ca.* Thorarchaeota and *Ca.* Lokiarchaeota (belonging to Asgard archaea) (20, 21), *Ca.* Huberarchaeota, *Ca.* Micrarchaeota, *Ca.* Pacearchaeota, *Ca.* Woesearchaeota and *Ca.* Mamarchaeota (belonging to DPANN) (22-27), *Ca.* Bathyarchaeota, *Ca.* Geothermarchaeota (also

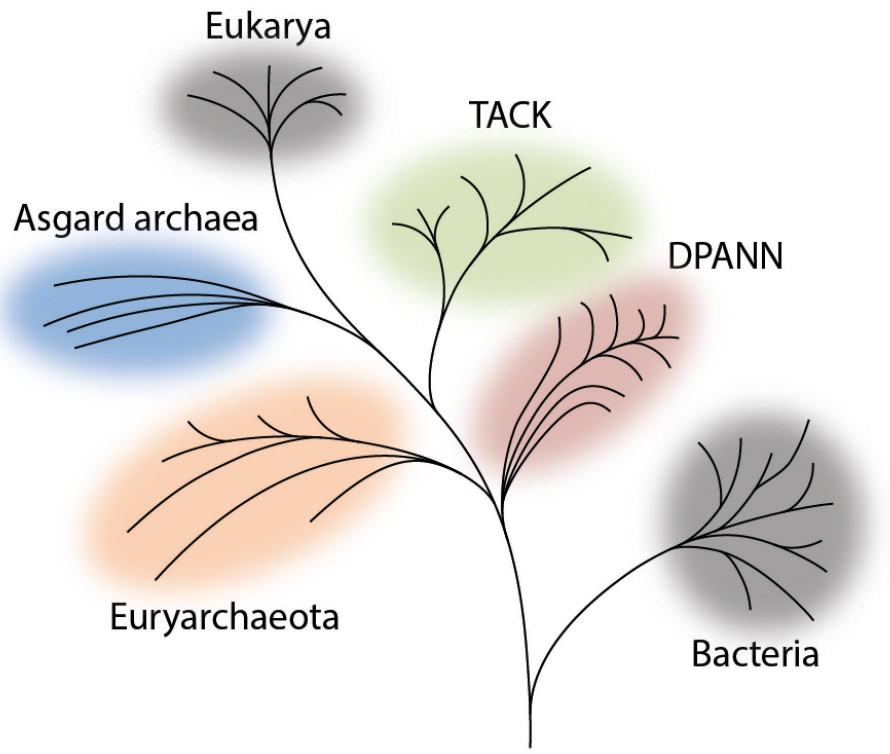


Figure 1.1: Cartoon of the tree of life, with emphasis on archaea. The tree is rooted at the last universal common ancestor (LUCA), and shows the two domains: bacteria and archaea/eukarya. The archaeal superphyla Asgard archaea, TACK and DPANN, and the phylum Euryarchaeota, together represent all archaeal genomes that are known to date. Modified from: Eme *et al.* Nat. Rev. Microbiol. 2017 (47).

referred to as *Ca.* Geoarchaeota), *Ca.* Verstraetearchaeota and *Ca.* Marsarchaeota (belonging to TACK) (28-31), and the classes Methanonatronarchaeia, Theoinarchaea and Hadesarchaea (belonging to Euryarchaeota) (32-34) (**FIGURE 1.1**). Most of the more recently discovered archaeal phyla have the *Candidatus* status, since their existence is thus far solely based on metagenomic sequencing, and they are categorized based on their 16S ribosomal RNA sequence. For now, only phyla with at least one member strain maintained in microbiological culture get a definitive phylum name (35-37). Also, the categorization of archaeal phyla is not set in stone, as new sequencing data and insights into archaeal evolution may change the layout of the archaeal tree of life. Especially the categorization of DPANN is subject to discussion (38). While most studies regard DPANN as a separate and deep-branching superphylum within archaea (39-44), others suggest that some or all DPANN lineages are part of Euryarchaeota, and that their categorization as part of the DPANN superphylum is an artifact (45). Also the position of *Ca.* Lokiarchaeota has been challenged (46).

The study of archaeal evolution is the key to understanding eukaryogenesis, the evolutionary events that have led to the emergence of eukaryotes from prokaryotes. Eukaryogenesis is best understood by identifying the first eukaryotic common ancestor (FECA) and the last eukaryotic common ancestor (LECA) (47). FECA can be defined as the archaeon that is the oldest ancestor of all eukaryotes that are living today. LECA is the organism of which all now-living eukaryotes are descendants. Although all present-day eukaryotes are descendants of both FECA and LECA, not all descendants of FECA and LECA are living today, and FECA may have had archaeal descendants that have gone extinct. FECA and LECA are phylogenetic entities, identified as nodes in the tree of life (47). The discovery of more archaea that are closely related to eukaryotes and deep-branching eukaryotes can add branches to the tree of life, and as a result, FECA and LECA can move closer together. The narrowing of the gap between FECA and LECA provides us with information about when and how the first eukaryote emerged. Also, it sheds light on the order in which eukaryotic characteristics, such as nucleus and mitochondria, first appeared in the evolutionary predecessors of eukaryotes. It is therefore important to identify common eukaryotic features, in order to determine the cellular processes of LECA. It has been established that LECA lived 1 to 1.9 billion years ago and had a nucleus containing linear chromosomes, encoding genes with introns (48-50). Histones played a role in the transcription regulation system, together with an RNAi system and small non-coding RNAs (48, 51). Cell motility machinery involving an actin- and tubulin-based cytoskeleton and motor proteins was present in LECA, just like mitochondria that allow for aerobic respiration (48). The mitochondria were descendants from alphaproteobacteria, which have once been engulfed by an ancestor of LECA, as described by the endosymbiotic theory (52, 53). The membrane of LECA was composed of phospholipids made of fatty acids and *sn*-glycerol-3-phosphate (G3P), just like present-day eukaryotes and bacteria (54). In contrast, archaeal membranes consist of phospholipids based on isoprenoids and *sn*-glycerol-1-phosphate (G1P). This 'lipid divide' was initially used as an argument for the three-domain theory, but it does not rule out the existence of archaea with a G3P-based membrane. It has been suggested that some archaea, including the Asgard archaea member phylum *Ca.* Lokiarchaeota, may have the ability to produce eukaryotic-like membranes, based on the presence of genes coding for enzymes involved in the metabolism of G3P and fatty acids (55-57). However, this has not been confirmed experimentally.

Like the composition of the membrane, other features of LECA may have been inherited from FECA and its ancestors. Of the estimated 4000 genes of LECA, 550-1100 genes are hypothesized to be of prokaryotic descendance; 18-37% of those genes have their origin in archaea. This does not seem like a large propor-

tion, although it should be noted that also some archaeal genes are of bacterial origin. Furthermore, the recent discovery of new phyla belonging to TACK and Asgard archaea has revealed that some components of processes thought to occur exclusively in eukaryotes, are also present in some archaea (47). Most Asgard archaea contain some, if not all, components of the eukaryotic trafficking and membrane remodeling machinery (including ESCRT-machinery), cytoskeleton, oligosaccharyl-transferase system and ubiquitin system, as well as some informational proteins and eukaryotic-like ribosomal proteins (21). Fluorescence *in situ* hybridization and catalyzed reporter deposition (CARD-FISH) experiments have shown that *Ca.* Lokiarchaeota cells occur in a wide spectrum of shapes and sizes, which probably reflects the expression of cytoskeletal and membrane remodeling proteins (58). To a lesser extent, the ribosomal, ubiquitin-related, trafficking-related and informational proteins were also identified in *Ca.* Verstraetearchaeota, Crenarchaeota, *Ca.* Korarchaeota, *Ca.* Bathyarchaeota and Thaumarchaeota (47, 59-62). This suggests that FECA likely contained the before-mentioned 'eukaryotic' components, and may have passed these on to its eukaryotic descendants.

Habitats and metabolism

Archaea were initially known as organisms that live at - from a human perspective - extreme conditions, such as hot springs, hydrothermal vents, hypersaline lakes and highly acidic water. Later, archaea were also found in oceans, soil and intestines of humans and cattle. Now, archaea are considered to inhabit nearly every possible habitat on earth, from permafrost to sewage and from ocean floors to goldmines (63). Furthermore, archaea occupy a large array of niches on or in the human body (64). The role of archaea in the human microbiome has largely been undiscovered, although knowledge of archaeal presence on or in humans is rapidly expanding. Recently, *Ca.* Woesearchaeota species were found in human lungs, Thaumarchaeota members were found on skin, and Crenarchaetoa and Euryarchaeota were found in intestines (65, 66). Surprisingly, to date, no real archaeal pathogens have been identified, but it was suggested that *Methanobrevibacter oralis* may be involved in periodontitis and brain abscesses (67, 68). Archaea are also suspected to be involved in inflammatory bowel diseases (69). Additionally, in their capacity of improving the microbiome, the role of archaea as probiotics is investigated, and archaea may help against trimethylaminuria and cardiovascular diseases (70). Other animals also live in symbiosis with archaea, as has been shown in studies on farmhouse animals, nonhuman primates, macropodidae, birds, reptiles, arthropods and sponges (66). Methanogens account for the vast majority of archaea in animal gastrointernal tracts, with *Methanobrevibacter*, *Methanobacterium* and *Methanosphaera* (all Euryarchaeota)

as main contributing families.

Archaea not only live in symbiosis with eukaryotes, some are co-dependent on other archaea. A well-studied example is *Nanoarchaeum equitans* (Nanoarchaeota), which can only be grown in co-culture with *Ignicoccus hospitalis* (Crenarchaeota) (71, 72). These organisms may exchange membrane lipids, amino acids and enzymes involved in several metabolic pathways (71, 72).

Archaea play an important role in biogeochemical cycles together with other life forms. Via their unique and highly variable metabolic pathways, they recycle carbon, nitrogen and sulfur, as well as other elements (73). Euryarchaeota, Crenarchaeota and Thaumarchaeota are relatively well characterized in terms of metabolism, as they can be cultured. In other phyla, some information about their metabolic pathways can be deduced from the enzymes that their genomes encode. Archaea are an important contributor to production and consumption of greenhouse gases such as carbon dioxide and methane (73). Significant amounts of methane are known to be produced by oceanic archaea (including archaea living inside oceanic plankton) and archaea living inside cattle. Most Euryarchaeota are anaerobic and autotrophic, producing organic compounds from carbon dioxide and bicarbonate (74). Also, some Crenarchaeota and Thaumarchaeota are autotrophs. Mixotrophic organisms can be found in the phyla Crenarchaeota and Euryarchaeota (73). Chemoorganotrophic archaea reduce organic compounds to carbon dioxide or partially oxidized compounds such as organic acids and alcohols. These types of archaea are identified as Euryarchaeota and Crenarchaeota. Nanohaloarchaeota, *Ca.* Korarchaeota, *Ca.* Geoarchaeota, *Ca.* Aigarchaeota, *Ca.* Micrarchaeota, *Ca.* Parvarchaeota and *Ca.* Marsarchaeota can likely also be regarded as chemoautotrophs, and can be anaerobic or (facultatively) aerobic (30, 73, 75). Phototrophic organotrophs, which get their energy from sunlight, have also been identified (76).

A large group of euryarchaeal species are methanogenic, producing methane with hydrogen gas as electron donor and carbon dioxide as electron acceptor (77). The reverse process, methane oxidation, is done by the anaerobic methanotrophs (ANME) group of Euryarchaeota (78, 79). ANME usually use sulfate as electron acceptor, but some use nitrite or iron- or manganese-containing compounds instead (80). Some archaea are thereby also coupled to biogeochemical cycles of metals.

Archaea are mainly involved in the nitrogen cycle via the oxidation of nitrate, ammonia or nitrogen-containing organic compounds to nitrite (81). However,

in the absence of those compounds, some anaerobic methane-oxidizing Euryarchaeota can use nitrogen gas as an electron donor (82). Denitrification, the reverse process (nitrite or nitrate to nitrogen gas), can also be done by some archaea (83). Few examples of archaea capable of producing nitrous oxide are currently known (84).

Some Crenarchaeota and Euryarchaeota contribute to the sulfur cycle via sulfidogenesis: the anaerobic production of hydrogen sulfide (85). A number of sulfur-containing compounds are used as electron acceptors, most notably elementary sulfur, sulfite, thiosulfate and sulfate. Sulfur oxidation, in which the sulfur compound is the electron donor, is observed in some crenarchaeal and euryarchaeal species, which use the iron-containing compounds pyrite, marcasite and chalcopyrite (86).

Transcription regulation in archaea

Like in any organism, archaea regulate transcription in order to adapt to their environment. The archaeal transcription machinery contains some elements that resemble eukaryotic transcription machinery, while other elements are more similar to their bacterial counterparts. Organisms from all domains express RNA polymerase (RNAP), a well-conserved and multi-subunit enzyme responsible for transcribing DNA into mRNA (87). The structure of archaeal RNAP resembles that of eukaryotic RNAPII, with most of the 13 subunits matching. Also, the initiation- and elongation-related stalk domain, absent in bacterial RNAP, is present in archaeal RNAP and resembles that of eukaryotes (88, 89). Transcription occurs in three stages: initiation, elongation and termination, together forming the transcription cycle. All three stages are regulated, although to date, the most information regarding transcription regulation is available on transcription initiation (88, 90). It should be noted that for historic reasons, almost all research regarding archaeal transcription regulation has been done on organisms belonging to Euryarchaeota and Crenarchaeota, which are also the easiest classes to cultivate. Although homologs of transcription machinery elements are present in genomes belonging to DPANN and Asgard archaea, the exact transcription mechanisms in these superphyla have yet to be unveiled.

Basal transcription regulation

On a basal level, archaeal regulation of transcription initiation is similar to its eukaryotic equivalent (90-92). Two DNA elements are required for transcription

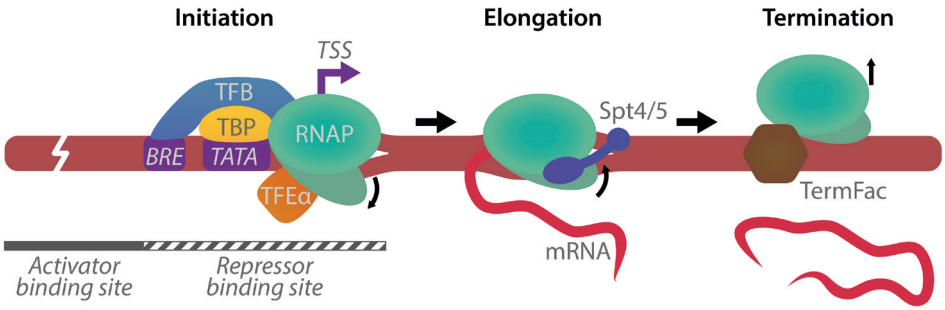


Figure 1.2: The archaeal transcription cycle. At the initiation stage, TATA-binding protein (TBP) binds the TATA-box (TATA), and is stabilized by transcription factor B (TFB) that binds to the B recognition element (BRE) and to TBP. This complex, situated just upstream of the transcription start site (TSS), recruits RNA polymerase (RNAP), which melts the DNA, forming the open complex. Open complex formation is sometimes assisted by transcription factor E (TFEα), which leaves during early elongation. Transcription factors with a repressive or activatory role can bind the promotor region at or (far) upstream of the initiation-related DNA elements, respectively. During elongation, elongation factors Spt4 and Spt5 stabilize the clamp domain of RNAP, thereby preventing dissociation of the elongation complex. Transcription is likely terminated by termination factors (TermFac) or by intrinsic termination due to weak base pairing at dedicated DNA sequences (not shown). RNAP dissociates from the DNA and can start a new transcription cycle. DNA elements are written in italics.

initiation: the TATA-box and the B recognition element (BRE). Both are located within the first 50 bp upstream of the transcription start site (TSS), in the promoter region (FIGURE 1.2). These elements require binding of general transcription factors (GTFs) in order to start transcription (90, 91). TATA-binding protein (TBP) binds to the TATA-box, and Transcription factor B (TFB) binds to the BRE and to TBP, thereby stabilizing the TBP-TATA complex (93). The assembly of TBP and TFB on DNA in the preinitiation complex (PIC) is a prerequisite for RNAP recruitment. Unlike eukaryotic transcription initiation, no energy is required by the PIC to separate DNA strands and form a transcription bubble (94). However, the non-essential GTF transcription factor E (TFEα) assists promoter opening by inducing the opening of the RNAP clamp (95, 96). Additionally, the promoter proximal DNA element (PPE), situated just downstream of the TSS, stabilizes TFB binding at weak promoters, and is able to provide additional promoter selectivity. Also, the initiator element (Inr) can increase the strength of the promoter, and in some cases it is bound by transcription activators. Both PPE and Inr are not essential for transcription initiation.

Not much is known about elongation regulation in archaea. The universally conserved protein Spt5 (also called Spt5 in eukaryotes; NusG in bacteria) with its conserved binding partner Spt4, and transcription factor S (TFS) are the only elongation factors that have been identified (97). TFS was hypothesized to be

the main elongation factor, with Spt4/5 as a secondary factor (98). TFS enhances RNAP's endonucleolytic cleavage activity, which allows removal of stalled elongation complexes, thereby making way for other elongation complexes. Spt4/5 replaces the GTFs at promoter escape or during early elongation and enhances processivity (99, 100). Also, Spt4/5 closes the clamp domain of RNAP, presumably resulting in a more stable elongation complex, which reduces dissociation from the DNA (95). In addition, both TFS and Spt4/5 allow for faster processivity of RNAP through histone-induced barriers (98). The interplay between elongation factors and histones provides a possible mechanism of transcription regulation.

Besides early transcription termination facilitated by elongation factors, regular transcription termination is regulated via DNA elements and likely also via termination factors. Intrinsic termination occurs when base pairing between the DNA template and the mRNA becomes too weak. Similar to eukaryotes, the main cue for archaeal intrinsic termination is a series of adenines, which results in a poly-U tract in the mRNA (101, 102). Subsequent destabilization of the RNA-DNA hybrid results in transcription termination. In some cases however, a stem-loop structure helps termination, like in bacteria (103). Not much is known about factor-mediated termination, because homologs of bacterial and eukaryotic termination factors have never been identified (90). Eta (euryarchaeal termination activity) from *Thermococcus kodakarensis* is the only known termination factor to date in archaea (104). Eta is conserved in most archaeal lineages, and therefore it may be functional as a transcription terminator throughout the domain.

Advanced transcription regulation

On a more advanced level, transcription is regulated by transcription factors (TFs), which, in addition to GTFs, are necessary for transcription initiation at specific genes and operons. These TFs resemble their bacterial counterparts (105). TFs function as activators or repressors, and together with the GTFs, they promote or block transcription initiation (106). The promotive or repressive function of a TF can be deduced from the position of their binding site: if it overlaps the BRE, TATA-box or TSS, it likely prevents PIC formation (91, 93). In case a TF's binding position is located more upstream of the BRE, it may recruit RNAP, TBP or TFB, resulting in transcription activation (**FIGURE 1.2**). However, more complex mechanisms of transcription regulation have also been found, which involve for example autoactivation and autorepression by the same TF, complex formation by repressors in order to block RNAP recruitment, and activation of alternative TSSs within the coding region (107-109). Many TFs are able to respond to environmental cues by binding ligands, such as sugars or amino acids (105). Ligand

binding subsequently enhances or decreases DNA binding affinity of TFs, allowing them to bind to or dissociate from repressor- or activator binding sequences. Some TFs are able to bind a large set of ligands, thereby gaining or losing affinity to defined promoters, while others bind only one ligand and operate more binary (110, 111). When groups of genes related to one physiological trait need to be activated or silenced, this is in general done by one TF, controlling the so-called regulon. Regulon-related TFs are almost always bifunctional, which allows them to promote one set of genes while repressing another set with an opposite physiological effect (91).

Like in bacteria, there are a few master regulators that control a large group of genes. All TFs and genes that have a transcriptional relation, operate in a gene regulatory network (GRN). In these networks, paralogs of GTFs play an important role. More than 70% of all archaeal genomes contain at least two genes encoding the same GTF: TBP or TFB (112). Specific combinations of GTF paralogs promote transcription of defined genes (113). Combining TBP and TFB paralogs results in a number of possible combinations that is equal to the product of the number of paralogs. This allows archaea to use some combinations for standard growth conditions, whereas other combinations may specifically be used during stress. Also, some promoters have unique GFT paralog combinations. In addition, also TFE has different paralogs in some species. Especially the genomes of the phylum Thaumarchaeota and the euryarchaeal class Halobacteria encode a large number of GTF paralogs (91).

Transcription regulation via chromatin architecture

Archaea have more options when it comes to gene regulation. Small RNAs (sRNA) are highly abundant and play a role in GRNs (114). sRNAs have been studied in Euryarchaeota, Crenarchaeota and Nanoarchaeota. Also, the covalent modification of proteins by small archaeal modifier proteins (SAMPs), a process known as SAMPylation, can be regarded as regulatory mechanism (115). Like ubiquitination in eukaryotes, SAMPylation targets proteins for degradation. Furthermore, post-translational modification (PTM) of proteins by adding or removing functional group can alter a protein's function. Another important regulatory factor is the modulation of chromatin architecture. Archaea express nucleoid-associated proteins (NAPs) and histones (together called chromatin proteins) that are able to change chromatin formation by bending, wrapping or bridging DNA (13, 116). Modulating genome architecture can result in activation or repression of genes and operons (117). For example, changes in chromatin conformation may result in blockage of the promoter region or inhibition of the transcription elongation

complex. In contrast, certain chromatin protein-induced DNA conformations may allow DNA-bound activators to facilitate RNAP recruitment, leading to transcription initiation. Some chromatin proteins are highly conserved among archaea, while others are unique to one or a few phyla or classes. Chromatin proteins are highly abundant and usually do not bind at specific DNA sequences (116). Therefore, they may affect transcription regulation globally, which would resemble transcription regulation via chromatin conformation in bacteria. However, structural data on archaeal genome architecture *in vivo* and its dynamics is mostly lacking.

A widespread NAP, encoded in a subset of genomes from all archaeal superphyla, is Alba (acetylation lowers binding affinity). Alba is able to bridge two DNA duplexes. Depending on the presence of a phenylalanine at position 60 (present in the variant Alba1, but absent in Alba2), Alba is able to form filaments along DNA via dimer-dimer interactions along the DNA (118). DNA binding by Alba can be modulated by trimethylation of the lysine at position 16, which lowers DNA-binding affinity (119, 120). In bridging mode, Alba may facilitate loop formation. In bacteria, loop formation by H-NS functionally organizes the genome, and in archaea, Alba may have a similar role (121). Its lateral filament mode may repress transcription via binding across promoter regions. Also, bridging by H-NS is able to cause repression of transcription, which is arguably the favored model for the involvement of H-NS in transcription regulation, as bridging can be modulated (122, 123).

In the TACK-phylum Crenarchaeota, many NAPs are expressed, such as Sul7, Cren7, CC1, Sso10a and Sso7c (124). Cren7 is best conserved within Crenarchaeota, whereas Sso10a can also be found in some classes of Euryarchaeota. The abundance of different NAPs in Crenarchaeota may be related to the absence of genes encoding histones in most genomes of the phylum. Cren7, Sul7, and likely also Sso7c, organize chromatin by DNA bending, whereas Sso10a has bending, bridging and lateral filament modes (125-127). The role of CC1 in chromatin architecture has yet to be elucidated (128).

Methanogen chromosomal protein 1 (MC1) is another NAP that is presumably involved in functional organization of the archaeal genome. MC1 induces sharp kinks into the DNA (129). It is encoded by genomes in the euryarchaeal classes Halobacteria, Methanomicrobia and Hadesarchaea, and by the genome of the DPANN-phylum Altiarchaeota.

Homologues of bacterial NAPs were identified in some archaea. Dps (DNA-

binding protein from starved cells), which protects DNA during oxidative stress by condensing chromatin in rosette-like structures, can be found in some species belonging to Euryarchaeota and Thaumarchaeota (130-132). Also, some genomes spread across the archaeal tree of life encode homologs of the bacterial DNA bender HU (133). It seems likely that these proteins have been acquired via horizontal gene transfer.

The most conserved protein related to chromatin architecture, however, is the histone. It has been found in almost every phylum of archaea (*see also* **CHAPTER 2** of this thesis), and is believed to be a major contributor to DNA conformation on a local and global level (134). The archaeal histone resembles the well-studied eukaryotic histone in both sequence and in structure of the histone fold, the core of the histone monomer. However, there are also some clear differences between histones from eukaryotes and archaea. For a better understanding of these differences, I introduce the eukaryotic histone first, before comprehensively discussing the archaeal histone.

The eukaryotic histone

In eukaryotes, octameric histone cores compact DNA by wrapping an approximately 150 bp unit twice around its surface, forming a nucleosome (135, 136). Nucleosomes interact with each other, yielding an additional level of DNA organization in the form of a fiber. Besides a role in compaction, histones also play roles in genome organization, DNA replication, DNA repair and gene expres-

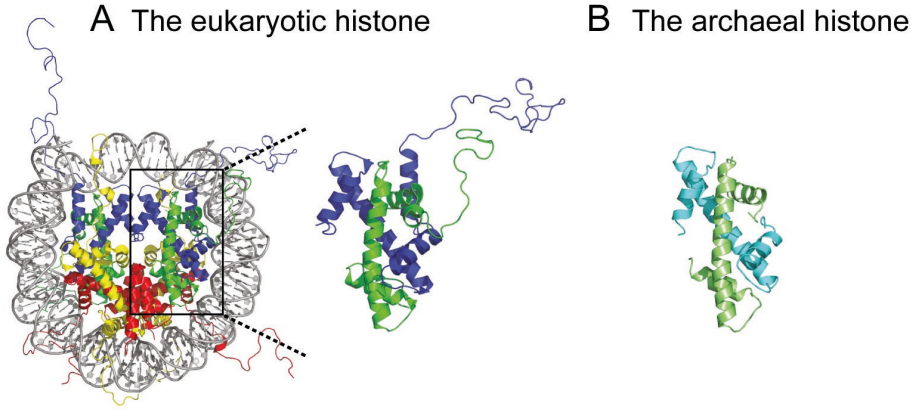


Figure 1.3: Eukaryotic and archaeal histones. A) Eukaryotic nucleosome consisting of DNA wrapped around a core of a (H3-H4)2 tetramer and two H2A-H2B dimers. Zoom-in of the black box shows a H3-H4 histone dimer. Yellow: H2A; red: H2B; blue: H3; green: H4. B) Archaeal histone homodimer of HMfB.

sion, which highlights the nucleosome as a very important complex affecting a vast array of cellular processes. Characteristic of core histone proteins is a common 'histone fold': two short and one long α-helix, separated by loops (137-141). In eukaryotes, the histone core consists of two H2A-H2B dimers and a H3-H4 tetramer, around which ~146 bp of DNA is wrapped 1.65 times (**FIGURE 1.3A**). It has been suggested that smaller histone assemblies, such as tetrasomes (H3-H4 tetramers), hexasomes (H3-H4 tetramers plus one H2A/H2B dimer) and hemisomes (a H3-H4 dimer plus one H2A/H2B dimer), have functional roles as intermediate structures during for example transcription elongation (142-145). The linker histone H1 (which lacks the characteristic histone fold) binds at the entry and exit points of the DNA wrapped around the octameric histone core (146, 147). The association of histone H1 constrains an additional 20 bp of DNA and helps formation of the 30-nm fiber, which results in tighter compaction (148, 149). Also, flexible N-terminal tails that protrude from eukaryotic histones contribute to a tighter DNA packaging. These tails may interact with either the DNA or the histone surface of another nucleosome, which facilitates close association of nucleosomes (150-152). Furthermore, post-translational modifications of amino acid residues in the N-terminal tails, such as acetylation, methylation, phosphorylation, ubiquitination and biotinylation, are a key instrument for the cell to regulate gene expression, the DNA damage response and many other processes (153-155). For instance, while heterochromatin (tightly packed DNA) is typically devoid of acetylated lysines, euchromatic (lightly packed) regions typically contain histones with acetylated lysines. Histone acetylation is believed to cause a locally less condensed chromatin structure *in vivo*, which is permissive to transcription. Thus, euchromatin contains actively transcribed genes. In particular the lysine-rich histone H4 tail seems to be crucial in the modulation of chromatin structure (150). *In vitro*, H4 tails are required for higher order chromatin folding (156-158), which can be disrupted by acetylation of K16 (150). Nucleosome function and level of genome compaction can be altered in a multitude of ways, providing flexible and versatile mechanisms for tuning the cell's chromatin structure and transcription regulation.

Archaeal histones: structure and alignment

Histones are the only archaeal chromatin proteins that resemble the eukaryotic histones H3 and H4 in structure and sequence. These proteins occur in almost every archaeal phylum, although histone-coding genes are not conserved throughout all orders of every phylum (159). The best-studied histones are HMfA and HMfB, two histones from the euryarchaeon *Methanothermus fervidus*, which are closely related in terms of sequence (160). These histones occur in solution

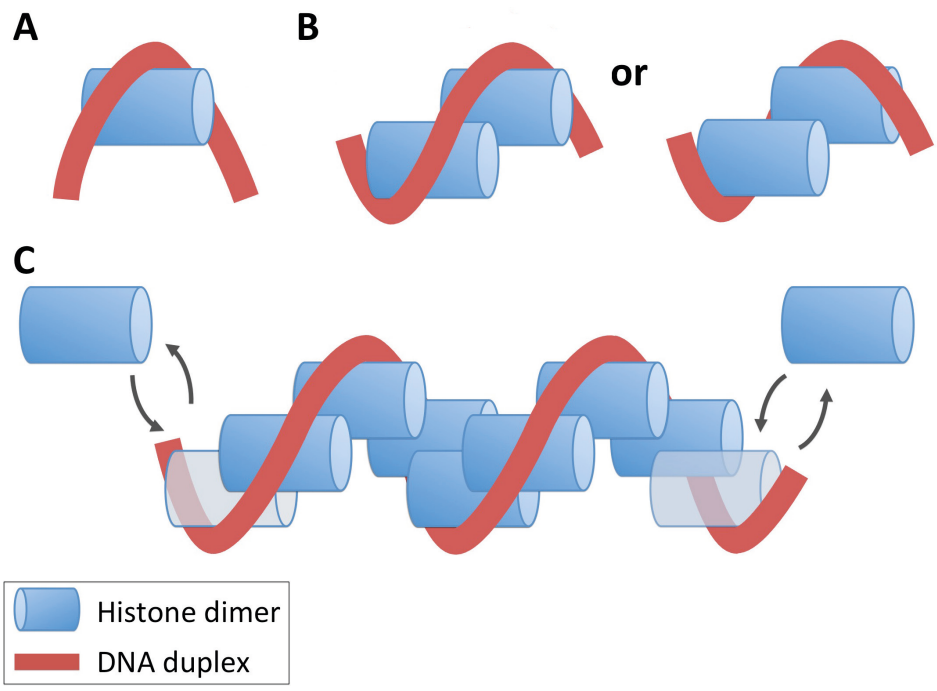


Figure 1.4: Model for multimerization of histones in *Thermococcus kodakarensis* and *Methanothermobacter thermautotrophicus*. Blue: histone dimers; red: DNA duplex. A) DNA is bent by a histone dimer, which covers 30 bp of DNA. B) A histone tetramer wraps the DNA in a left- or right-handed manner and is able to switch between these configurations, for example in response to changes in salt concentrations (172). C) Histone dimers can further associate to form a larger multimer. Every additional dimer wraps another 30 bp of DNA. At both extremities of the multimer, dimers can be added or removed. The left-handed wrapping depicted here is arbitrary, since also this large histone-multimer-DNA complex might accommodate both left-handed and right-handed wrapping.

as homo- or heterodimers (FIGURE 1.3B), but have been shown to bind to specific DNA sequences as tetramers, possibly resembling the (H3-H4)2 tetrasomes formed as intermediates during histone octamer assembly in eukaryotes (161). Archaeal histone tetramers were shown already in early studies to protect ~60 bp of DNA from micrococcal nuclease (MNase) digestion (162) and to assemble at specific high affinity sequences (163, 164). Conversely, MNase digestion of chromatin from the euryarchaeal species *Thermococcus kodakarensis* yields digests with fragments 60 bp in size, representing a tetrameric histone, but also fragments of 30 bp and multiples of 30 bp up to ~500 bp. This observation suggests that 'nucleosomes', which likely protected DNA from MNase digestion, formed by histone HTkB are not of defined size and that dimers, tetramers and larger multimers are functionally relevant (either in chromatin organization or in relation to transcription regulation) (165). Our lab proposed a model for multimerization

of histone proteins, in which histone dimers can be added and removed from both ends to assemble into a DNA-coiling multimer, similar in structure to the eukaryotic nucleosome, that consists of dimers that each cover 30 bp of DNA (140) (**FIGURE 1.4**). Studies on *Methanothermobacter thermautotrophicus* chromatin are in accordance with this observation, which suggests that the basic functional unit of archaeal histones might not be a tetramer, as generally assumed, but a dimer (165, 166). The observed differences between species might be due to differences in genomic sequence and/or protein sequence, altering dimer-dimer interactions. Although sequence homology between histones is sometimes limited, the histone fold is well conserved between species (167). Histones consist of three α-helices separated by short β-loops. Residues R10, K13, R19, K53, T54 and K56, located in the N-terminal α-helix, N-terminal β-loop and the C-terminal β-loop, were, based on crystal structure and mutagenesis studies, predicted to be mainly responsible for interacting with the DNA backbone. Dimer-dimer interaction, tetramerization, has been proposed to be mediated by residues E42, L46, H49, D59, L62 and R66 in the C-terminal and middle α-helix and the C-terminal β-loop (137, 160) Euryarchaeal histones have a high degree of sequence identity and are very similar in length, although some have an extended N-terminal loop and/or helix, or an extended C-terminus of ~30 amino acids. It has been proposed that these unconventional histone proteins are modulators of DNA binding by interaction with other histones (167, 168). Also, some histones from halophilic archaea consist of two histone domains fused together, whereas other histones assemble into dimers following synthesis (169). In these naturally occurring histone-histone fusion proteins, the histone fold at the C-terminus is more similar to the histone fold of other euryarchaeal histones than the one at the N-terminus (167, 170). HTkB and HMtA2, which likely bind along the genome as a dimer, tetramer or larger multimer in *T. kodakarensis* and *M. thermautotrophicus*, respectively, contain all residues identified as responsible for tetramerization in HMfB from *M. fervidus*. Therefore the difference in minimal size of the functional histone unit cannot be attributed to residues known to be involved in tetramerization.

Binding motif

Archaeal histones have a preference for binding GC-rich sequences with alternating (G/C)2/3 and (A/T)2/3 motifs, which are separated by half a helical turn. This compresses the minor and major groove on one side of the helix with A:T facing towards the histones, and G:C facing outwards, causing the DNA to bend. This mode of binding is very similar to that of eukaryotic histones. Studies on HMfB from *M. fervidus* reveal that binding motifs are more complex: AT and GC are not equal to TA and CG, respectively, in terms of facilitating archaeal

nucleosome assembly (164). Also, when analyzing substitution patterns in the DNA of nucleosomes in sister lineages of *Haloferax volcanii*, it was found that there is a preference for G:C near the dyad, whereas changes towards A:T are more common further away from the dyad, near the ends of the nucleosome DNA (172, 173). These sequence patterns are again very similar to human nucleosome positioning sequences. (174, 175).

In eukaryotes, a histone wraps 147 bp of DNA 1.65 times around its core, by means of which the DNA is compacted and supercoils are constrained (176). In archaea, tetrameric histones have been reported to wrap DNA without making a full turn, which results in a horseshoe-like conformation (177). However, interactions with the DNA may determine the extent of wrapping, which means DNA sequence motifs may affect the exact number of turns that is made by the DNA. HMf proteins from *M. fervidus* can wrap DNA around its tetramer core at high affinity sequences, but also are able to bend DNA as a dimer *in vitro* (178). Wrapping can be considered an advanced form of bending, rather than a separate mechanism, and may be an evolutionary consequence of recruitment of dimers at adjacent sites due to high affinity sequences or protein-protein interactions between dimers. *In vivo* bending and wrapping by histones is expected to occur in parallel at different locations along the genome. Not only wrapping by histones, but also loop formation by trans-acting elements such as Alba contributes to DNA organization and compaction. Loop formation may stabilize topologically isolated domains in which supercoiling is preserved, and in which a subset of genes can be co-regulated. (121, 179, 180).

Evolution of the histone protein

It has been suggested that eukaryotic histones evolved from archaeal histones (181). This hypothesis is supported by the high similarity at the amino acid sequence level and in secondary structure (182, 183). Suggestive of an archaeal origin of eukaryotic histones is also the dimeric nature of archaeal histones; archaeal histone complexes are built from dimers, but members of the archaeal class halobacteria express a 'tandem histone'. In these tandem histones, the histone folds are linked end-to-end (184-186). This implies that the histone folds always occupy the same position and role in the naturally linked dimer. This leads to the relaxation of evolutionary constraints in parts of the histone, an example of subfunctionalization (187, 188). According to this hypothesis, the histone folds further evolved in a divergent way, leading to an asymmetric dimer. This may have been an ancestor of H3-H4, which later separated to become two individual proteins and corresponding genes (182). It has been suggested that H2A and H2B have arisen

from H3 and H4 later on in histone evolution (182). Indeed, H3 and H4 are more similar to archaeal histones than H2A and H2B, supporting this hypothesis. From this point in time, eukaryotic histones have further evolved into histone variants: highly homologous substitutes of canonical eukaryotic histones, which often play a specialist role in a wide variety of cellular processes (189). Unlike canonical histones, which are mainly expressed during DNA replication, histone variants are expressed in a replication-independent manner (190, 191). Histone variants of H2A and H3 are widely known and studied, whereas only few examples have been found of diversified H2B and H4 (192). The evolutionary pressure for the evolution of dimer-based histones to octameric histones and their subsequent variants is believed to be DNA compaction (182). The fact that eukaryotic cells undergo mitosis, in which chromosomes are highly compacted, together with the abundance of gene-poor regions may have favored a histone conformation that wraps DNA twice (eukaryotic octamer) instead of once (archaeal tetramer), and that via its N-terminal tails has the ability to compact DNA at a higher order. By understanding the structural properties and functionality of archaeal histones, we can learn how evolution led to the eukaryotic histone.

In this thesis, I will further discuss archaeal histones in terms of structure, function, and their relation to DNA. **Chapter 2** highlights the interactions that histones have in a complex with other histones and DNA. Also, in **chapter 2**, all currently known (proposed) superphyla and phyla are discussed in terms of encoding and possibly expressing histones and histone variants. Furthermore, based on the protein sequences of these histones, I predict what kind of complexes they may form together with DNA, and what consequences this may have for the organism's genome architecture. In **chapter 3**, I describe a protocol for Tethered Particle Motion (TPM), a single molecule technique which is able to provide a read out of DNA conformation. Addition of histones and other DNA-binding proteins may change the DNA conformation. TPM is the main technique used in this thesis. Results of TPM experiments are described in **chapters 4**, **5** and **6**. In **chapter 4**, I focus on the largest histone-DNA assembly known so far: the hypernucleosome. Using TPM, as well as magnetic tweezers, I show at what concentrations hypernucleosomes of HMfA and of HMfB, the best-studied archaeal histones, form, and how they obtain their stability. **Chapter 5** describes histones from the candidate phylum Nanohaloarchaeota, which have never been reported to have been purified. I use TPM and microscale thermophoresis to characterize these histones. Finally, in **chapter 6**, I show how specific high affinity sites on the DNA alter the binding of histones to DNA using TPM. This chapter also provides an extensive overview of the possible functions of histones and histone variants in archaea. **Chapter 7** is a general discussion.

CHAPTER 2

Structure and function of archaeal histones

Part of this chapter is based on:

Bram Henneman, Clara van Emmerik, Hugo van Ingen & Remus T. Dame, (2018) *Structure and function of archaeal histones*. PLoS Genetics. doi: 10.1371/journal.pgen.1007582

Abstract

The genomes of all organisms throughout the tree of life are compacted and organized in chromatin by association of chromatin proteins. Eukaryotic genomes encode histones, which are assembled on the genome into octamers, yielding nucleosomes. Post-translational modifications of the histones, which occur mostly on their N-terminal tails, define the functional state of chromatin. Like eukaryotes, most archaeal genomes encode histones, which are believed to be involved in the compaction and organization of their genomes. Instead of discrete multimers, *in vivo* data suggest assembly of "nucleosomes" of variable size, consisting of multiples of dimers, which are able to induce repression of transcription. Based on these data and a model derived from X-ray crystallography, it was recently proposed that archaeal histones assemble on DNA into "endless" hypernucleosomes. In this chapter, we analyse archaeal histones from all archaeal domains and phyla. We identify archaeal histones differing from the consensus, which are expected to be unable to assemble into hypernucleosomes. Finally, we identify atypical archaeal histones with short N- or C-terminal extensions and C-terminal tails similar to the tails of eukaryotic histones, which are subject to post-translational modification. Based on the expected characteristics of these archaeal histones, we discuss possibilities of involvement of histones in archaeal transcription regulation.

Introduction

Architectural chromatin proteins are found in every domain of life. Bacteria express DNA-bending and DNA-bridging proteins, such as histone-like protein from *Escherichia coli* strain U93 (HU) and histone-like nucleoid-structuring protein (H-NS), to structure and functionally organize the genome and to regulate genome activity (11, 12). In eukaryotes and most archaeal lineages, histones are responsible for packaging and compaction of the DNA (**TABLE 2.1**). The histones found in archaea are widespread throughout the domain but are absent in some lineages, notably *Candidatus* (*Ca.*) Parvarchaeota and *Ca.* Marsarchaeota, and most Crenarchaeota and Thermoplasmata. They have the same histone fold as eukaryotic histones, but N-terminal histone tails have not been identified. Linker histones, homologous to eukaryotic H1, have not been found. Archaeal histones exist as dimers in solution, which have been shown to bend DNA (160, 193). These histone dimers can be homodimeric or heterodimeric (194), as many archaeal species express, or at least encode, more than one histone variant. In *Methanothermus fervidus* (class Methanobacteria), two histone variants are expressed at different levels and ratios at different growth phases, suggesting a distinct function for both proteins (193). In addition to binding as dimers, archaeal histones have been reported *in vivo* and *in vitro* to bind DNA as tetramers (160, 162, 195), wrapping the DNA once. However, micrococcal nuclease (MNase) digestion patterns of *Thermococcus kodakarensis* (class Thermococci) chromatin suggest that histone–DNA complexes consist of discrete multiples of a dimeric histone subunit (i.e., not limited to dimers and tetramers) *in vivo* without obvious dependence on the DNA sequence (196). Based on the latter observations, it was proposed that histone dimers multimerize and wrap DNA into a filament of variable length (140, 196). The crystallography study of Luger and coworkers on histone HMfB from *M. fervidus* indicates that these histones assemble into an endless left-handed rod *in vitro*, which we propose to call a "hypernucleosome" (**FIGURE 2.1**). Note that these complexes were assembled on SELEX-optimized DNA previously shown to favor tetrameric nucleosome assembly (197). The number of wraps in the hypernucleosome, which is the number of times that the DNA is bent by 360° around the histone multimer, scales linearly with the number of histone subunits, resulting in a tight packaging of DNA. The authors also provide evidence that mutation-directed perturbation of hypernucleosome function *in vivo* alters the response to nutrient change in *T. kodakarensis*, suggesting a role in transcription. Both eukaryotes and archaea encode histone proteins, which seem to be involved in response to environmental cues by their involvement in transcription regulation.

Table 2.1: phylogenetic subdivision of the archaeal domain

Superphylum	Phylum	Class	Histones
-	**Euryarchaeota**	Archaeoglobi	Yes
		Hadesarchaea	Yes
		Halobacteria	Yes
		Hydrothermarchaeota	Yes
		Methanobacteria	Yes
		Methanococci	Yes
		Methanomicrobia	Yes
		Methanonatronarchaeia	Yes
		Methanopyri*	Yes
		Theionarchaea	Yes
		Thermococci	Yes
		Thermoplasmata*	Yes
DPANN	*Ca.* Aenigmarchaeota		Yes
	Ca. Altiarchaeota		Yes
	Ca. Diapherotrites*		Yes
	Ca. Huberarchaeota		Yes
	Ca. Micrarchaeota		Yes
	Nanoarchaeota		Yes
	Ca. Nanohaloarchaeota		Yes
	Ca. Pacearchaeota*		Yes
	Ca. Parvarchaeota		No
	Ca. Woesearchaeota		Yes
TACK	*Ca.* Bathyarchaeota		Yes
	Crenarchaeota*		Yes
	Ca. Geothermarchaeota		Yes
	Ca. Korarchaeota		Yes
	Ca. Marsarchaeota		No
	Thaumarchaeota		Yes
	Ca. Verstraetearchaeota		Yes
Asgard Archaea	*Ca.* Heimdallarchaeota		Yes
	Ca. Lokiarchaeota		Yes
	Ca. Odinarchaeota		Yes
	Ca. Thorarchaeota		Yes

Division of archaea in superphyla and phyla, including the euryarchaeal classes. Presence or absence of histone-coding genes on the genome of at least one of the members of the phyla and classes have been indicated. *) of these groups, histones are found in a minority of sequenced genomes.

Histones are found in most newly discovered archaea

With the widespread use of metagenomic sequencing, entire new branches within the archaeal domain have been discovered. Genomes of the recently discovered archaeal superphylum Asgard archaea and phyla *Ca.* Bathyarchaeota, *Ca.* Woesearchaeota, *Ca.* Pacearchaeota, *Ca.* Aenigmarchaeota, *Ca.* Diapherotrites, *Ca.* Huberarchaeota, *Ca.* Verstraetearchaeota, *Ca.* Geothermarchaeota and *Ca.* Micrarchaeota encode histones (20-24, 26, 28, 29, 31, 39) (TABLE 2.1). With the publication of the genome sequences of these organisms, we were able to scrutinize the sequence divergence of histones by comparing sequences of histones from archaea throughout the domain (FIGURE 2.2). The selection of histones shown here is based on the presence of histone-coding genes in different phyla. Since many of those phyla were only discovered in the last five years, our selection includes a relatively large number of histones that have not yet been studied *in vivo* or *in vitro*.

We found that in *Methanosphaera* sp. SHI613 (class Methanobacteria), 11 different histones are encoded, which is the highest number of histones found in

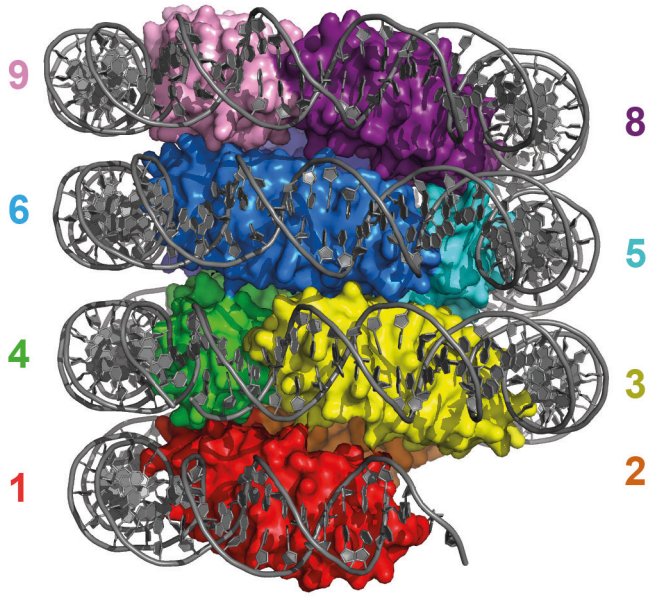


Figure 2.1: Overview of the hypernucleosome structure. HMfB dimers stack to form a continuous, central protein core that wraps the DNA in a left-handed superhelix. Nine HMfB dimers are shown, each dimer in surface mode and in rainbow colors. Numbering indicates position of the nine histone dimers; note that dimer 5 and 6 occlude the view of dimer 7. DNA is in gray and shown as cartoon. Image generated using Protein Data Bank (PDB) entry 5T5K (197).

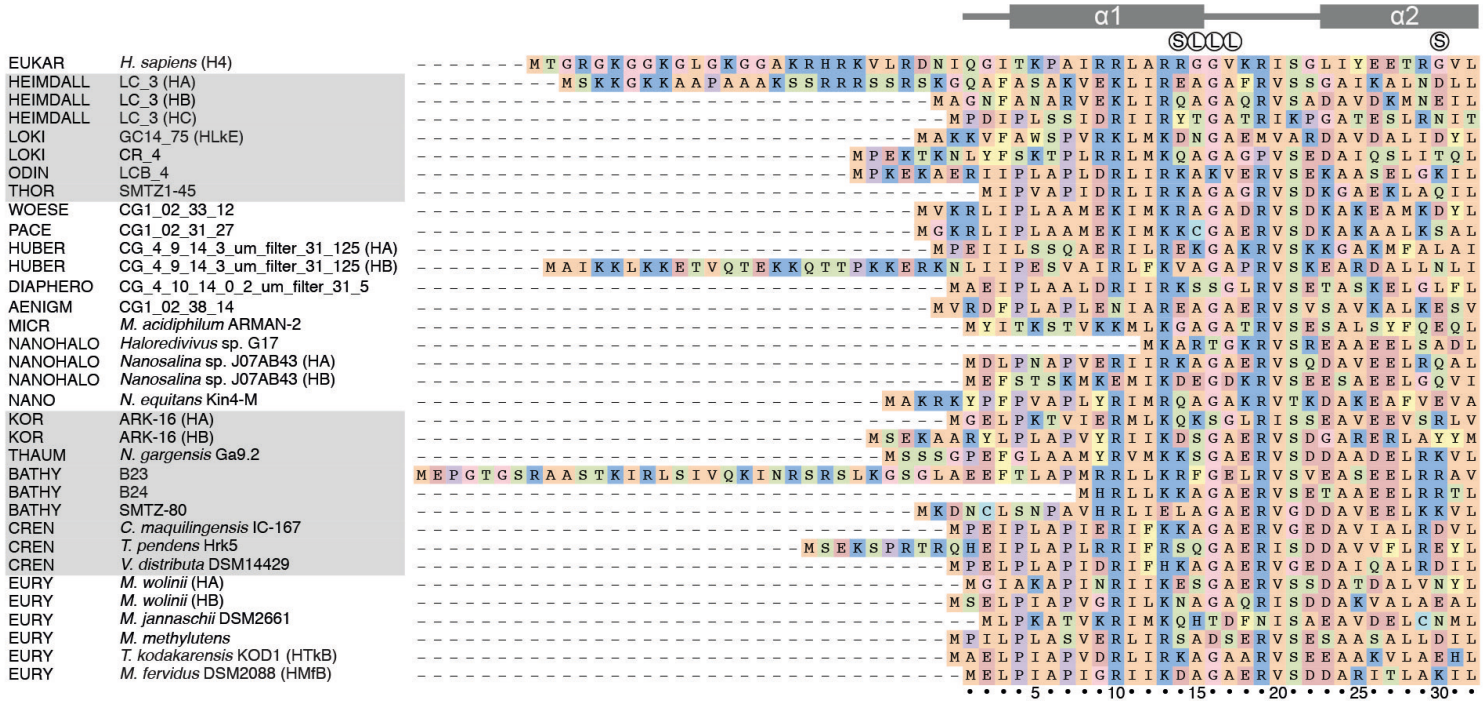
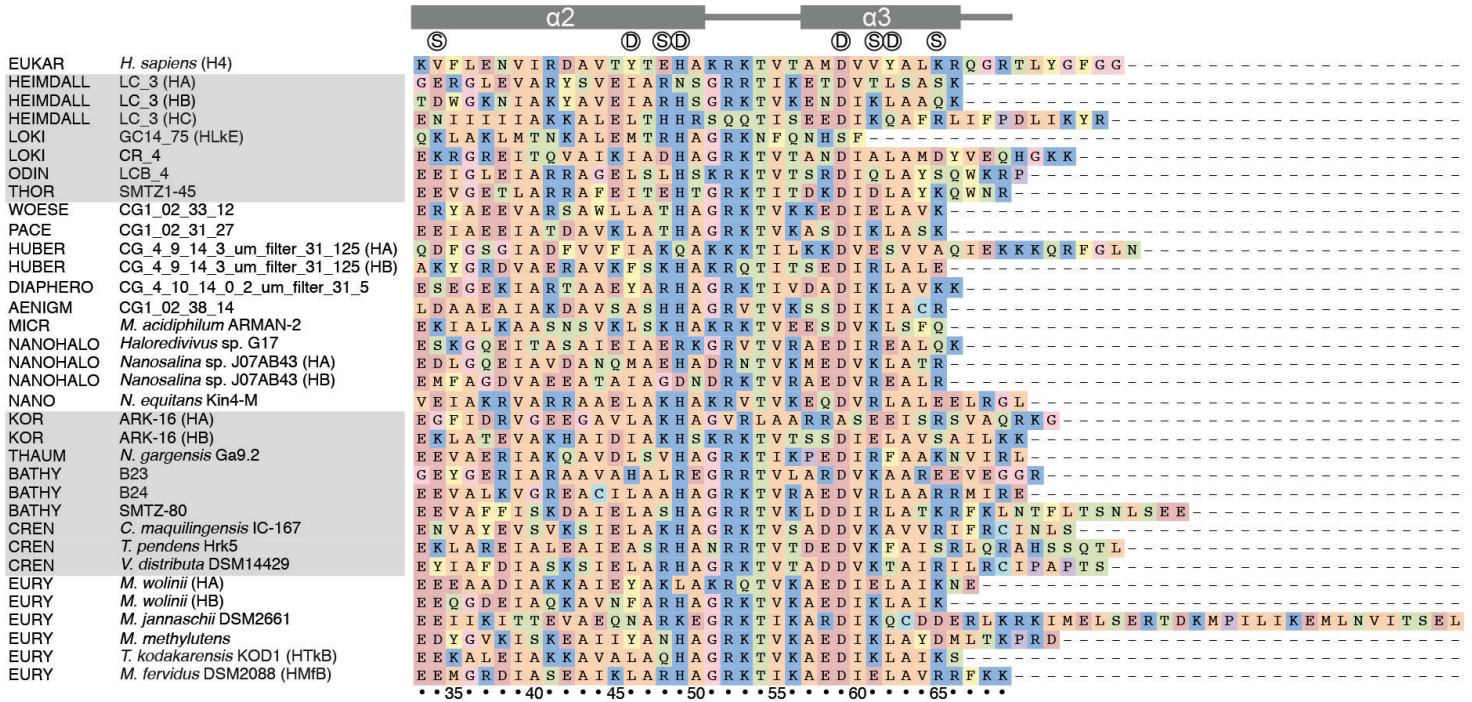


Figure 2.2. Alignment of histones from different archaeal species and human histone H4. Colors indicate the side chain group: R, H, K: blue; D, E: red; A, V, I, L, M: orange; F, Y, W: yellow; S, T, N, Q: green; C: turquoise; G: pink; P: purple. Symbols above the alignment indicate the dimer–dimer interface (D), the loop of the stacking interface (L), and the putative stacking interactions (S) (based on HMfB). Secondary structure and numbering of HMfB is used for reference. EUKAR H4, eukaryotic (human) histone H4 NP_724344.1; HEIMDALL LC_3 HA, HB, and HC, *Ca.* Heimdallarchaeota OLS22332.1, OLS24873.1, and OLS21974.1, respectively; LOKI GC14_75 HLkE and CR_4, *Ca.* Lokiarchaeota KKK41979.1 and OLS16336.1, respectively; ODIN, *Ca.* Odinarchaeota OLS18261.1; THOR, *Ca.* Thorarchaeota KXH71038.1; WOESE, *Ca.* Woesearchaeota OIO61677.1; PACE, *Ca.* Pacearchaeota OIO41945.1; HUBER, *Ca.* Huberarchaeota CG_4_9_14_3_um_filter_31_125 HA and HB, PJB03565.1, and PJB04497.1, respectively; DIAPHERO, *Ca.* Diapherotrites PJA17623.1;



(Figure 2.2, continued) AENIGM, *Ca.* Aenigmarchaeota OIN88081.1; MICR, *Ca.* Micrarchaeota *Micrarchaeum acidiphilum* ARMAN-2 EET90461.1; NANOHALO, *Ca.* Nanohaloarchaeota *Haloredivivus* sp. G17 and *Nanosalina* sp. J07AB43 HA and HB, EHK01841.1, EGQ42849.1, and EGQ43804.1, respectively; NANO, Nanoarchaeota *Nanoarchaeum equitans* Kin4-M AAR39197.1; THAUM, Thaumarchaeota *Nitrososphaera gargensis* Ga9.2 AFU59009.1; BATHY B23, B24, and SMTZ-80, *Ca.* Bathyarchaeota KYH36356.1, KYH37304.1, and KON27866.1, respectively; CREN, Crenarchaeota *Caldivirga maquilingensis* IC-167, *Thermofilum pendens* Hrk5, and *Vulcanisaeta distribute* DSM14429, ABW02527.1, ABL77757.1, and ADN51226.1, respectively; EURY, Euryarchaeota *Methanobrevibacter wolinii* HA and HB, *Methanocaldococcus jannaschii* DSM2661, *Methanococcoides methylutens*, *Thermococcus kodakarensis* KOD1, and *Methanothermus fervidus* DSM2066, WP_42707783.1, WP_42706862.1, AAB99668.1, KGK98166.1, BAD86478.1, and ADP77985.1, respectively.

one archaeal genome (198, 199). Also notable are genomes LC_3 of the phylum *Ca.* Heimdallarchaeota, and *Methanosphaera* sp. BMS, which both encode 10 histone paralogs (21). We have not found any histones in the genomes of phyla *Ca.* Parvarchaeota and *Ca.* Marsarchaeota, although it should be noted that abundance of available genomes and completeness of the genomes differs. The majority of available genomes from the phyla that do not seem to encode histones have an estimated completeness of between 40% and 99% (26, 30). This means that we cannot rule out the possibility that any of those genomes does contain one or more genes coding for histones. The absence of histones suggests that other NAPs may be involved in genome compaction. In that light, it is notable that in genomes from *Ca.* Parvarchaeota and *Ca.* Marsarchaeota, as well as in genomes from the phyla from Asgard Archaea, *Ca.* Woesearchaeota, *Ca.* Bathyarchaeota, *Ca.* Pacearchaeota, *Ca.* Verstraetearchaeota, *Ca.* Geothermarchaeota, *Ca.* Aenigmarchaeota and *Ca.* Micrarchaeota, genes coding for the DNA-bridging protein Alba1 (and in some cases, Alba2) are present. Like histones, Alba (or Sso10b) proteins are likely involved in transcription repression. They are highly abundant in nonhistone-coding Crenarchaeota (200), possibly taking the functional role of histones as found in other archaea. Some *Ca.* Parvarchaeota genomes encode Alba but not histones, and their genomes may therefore be shaped or regulated in a similar way as in Crenarchaeota. Furthermore, we found that only genomes of *Ca.* Thorarchaeota and Thermoplasmata contain an HU gene (a DNA-bending protein generally found in bacteria and archaea without histones (183)). Also, some genomes from Euryarchaeota and TACK encode Dps, a DNA protection protein that is found in bacteria. Dps has several mechanisms to protect the DNA in bacteria, such as strong DNA compaction and protection from hydrogen peroxide toxicity, but it is not known if archaeal Dps is able to function in a similar way. Additionally, the genomes of *Ca.* Huberarchaeota, *Ca.* Altiarchaeota and some Euryarchaeota encode an MC1 homologue, which is a monomeric DNA-bending protein often found in organisms from the euryarchaeal class Halobacteria. Genes coding for other known archaeal NAPs (201-203) were not found.

Some archaeal histones have eukaryote-like N-terminal tails

A striking finding based on the amino acid sequence comparison reveals that two histones from *Ca.* Heimdallarchaeota archaeon LC_3 (only Histone A [HA] shown in **FIGURE 2.2**), one from *Ca.* Huberarchaeota archaeon CG_4_9_14_3_um_filter_31_125 and one from *Ca.* Bathyarchaeota archaeon B23, contain an N-terminal tail, which was previously thought to exist only in eukaryotic histones and only recently reported for *Ca.* Heimdallarchaeota (197). In eukaryotes, these tails stabilize a higher order of compaction by interacting with either the DNA or

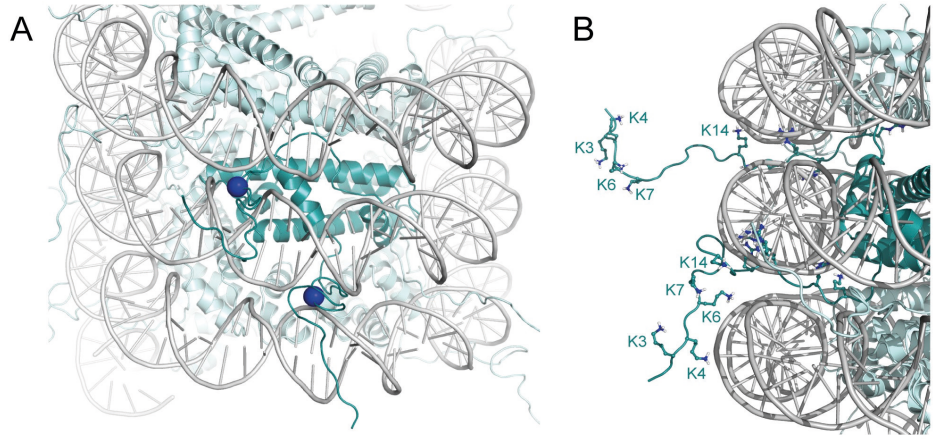


Figure 2.3: Model of a *Ca.* Heimdallarchaeota archaeon LC_3 hypernucleosome with N-terminal tails. A) View showing histone tails protruding through the DNA minor grooves. The R17 Cα-atom is shown as a blue sphere to mark the exit point of the tail. B) Close up of the histone tails with lysine and arginine residues shown as sticks, and N-terminal lysines are labeled. Homodimers of Heimdall LC_3 histone HA are shown in teal; one dimer is highlighted in darker colors. Models are based on the structure of HMfB (PDB entry 5T5K); the tail in the top (bottom) of panel B is modeled in the H3 (H4) tail conformation (PDB entry 1KX5).

another nucleosome. The tails of the two histones from *Ca.* Heimdallarchaeota and *Ca.* Huberarchaeota are of roughly the same length and sequence composition as eukaryotic H4 tails (*see* **FIGURE 2.2**). Prompted by the importance of the eukaryotic histone tails in modulating chromatin structure and function (150, 155), collaborator Clara van Emmerik constructed a molecular model of a hypernucleosome formed by Histone A (HA) from *Ca.* Heimdallarchaeota LC_3 to investigate its potential function (*see* **METHODS SECTION**).

The model illustrates how three subsequent arginines (R17–R19) could facilitate passing of the tails through the DNA gyres (**FIGURE 2.3**). The tails exit the hypernucleosome through DNA minor grooves, similar to eukaryotic histone tails, and might position their lysine side chains to bind to the hypernucleosomal DNA or to other DNA close by, facilitating (long-range) genomic interactions in trans. Like the H4 tail that is subject to acetylation of lysines K5, K8, K12, and K16 (204), lysines in the heimdallarchaeal histone tail may well be subject to acetylation. Archaeal genomes are known to have several candidate lysine acetyltransferase and deacetylase enzymes, including proteins belonging to the ELP3 superfamily, to which transcription elongation factor and histone acetyltransferase ELP3 belongs (205-207). Searches using the ProSite database* (208) and Protein Information Resource** (209) further reveal that the *Ca.* Heimdallarchaeota LC_3 genome

* http://prosite.expasy.org
** http://pir.georgetown.edu

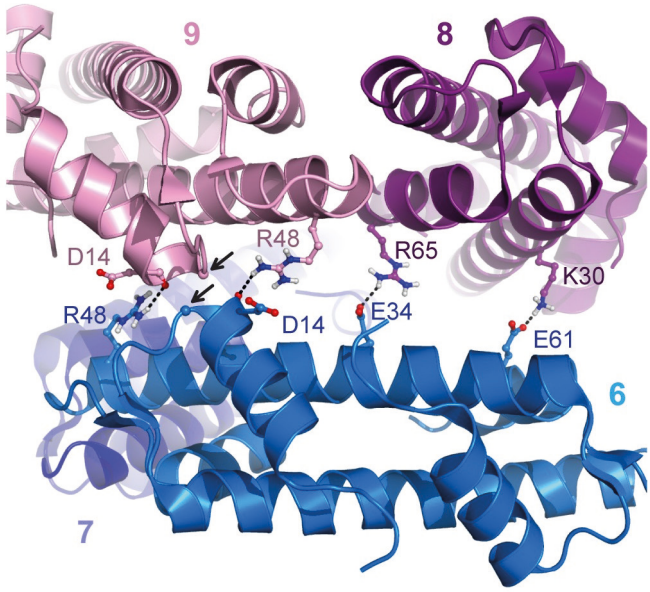


Figure 2.4: Stacking interface between HMfB dimers in the hypernucleosome. Each dimer *i* forms stacking interactions with dimer *i*+2 and *i*+3, shown here for dimer 6. Residues deemed important are shown in ball-and-sticks and labeled; close stacking of G16 in dimer 6 and 9 is indicated by arrows (Cα shown as spheres); hydrogen bonds are indicated with dashes. Image generated using PDB entry 5T5K (197)

contains multiple gene products containing the Gcn5-related N-acetyltransferase domain, which is present in many histone acetyltransferases (210). Interestingly, a potential 'reader' protein that binds modified lysines can also be identified. This protein, HeimC3_47440, contains a YEATS-domain, which has recently been shown to bind histone tails that carry acetylated or crotonylated lysines (211-214). Comparison with the closest homolog of known 3D structure, YEATS2 (35% identity, PDB-id 5IQL, (215)), shows that the binding site for the modified lysine side chain is strictly conserved in the archaeal protein. Notably, only *Ca.* Bathyarchaeota, which also features tailed histones, contains a detectable homolog of HeimC3_47440. The presence of lysine-containing N-terminal tails in combination with histone modification writers and readers suggests that archaea use post-translational modifications in a similar way to Eukaryotes as modulators of genome compaction and gene activity. The tail of the *Ca.* Huberarchaeota histone also contains lysine residues that are found at the same position as some of the lysines of the H4 tail. However, we have not identified any proteins resembling post-translational modification-related proteins from other organisms in terms of sequence in this phylum.

Other histones, for example from *Ca.* Lokiarchaeota CR_4, *Ca.* Odinarchaeota

LBC_4, *Nanoarchaeum equitans* (Nanoarchaeota), and *Thermofilum pendens* (Crenarchaeota), contain a short N-terminal tail of 5–10 residues. Also, histones with a C-terminal tail have been found. The histone from the euryarchaeal species *Methanocaldococcus jannaschii* (class Methanococci) has a 28-residue C-terminal tail, which seems to be unique among archaeal histones. Other C-terminal tails are up to 11 residues long (as compared to *Methanothermus fervidus* HMfB) and appear in *Caldiarchaeum subterraneum* (Thaumarchaeota), *Ca.* Bathyarchaeota SMTZ-80, *Ca.* Heimdallarchaeota LC_3, *Ca.* Lokiarchaeota CR_4, and all histones found in Crenarchaeota. These short C-terminal tails are similar in length to the H4 C-terminal tail, that is reported to play a role in the promotion of histone octamer formation in eukaryotes (216). The genomes of some archaeal species contain genes for histone truncates. The histone from *Haloredivivus* sp. G17, member of *Ca.* Nanohaloarchaeota, and the histone from *Ca.* Bathyarchaeota archaeon B24 both lack part of the N-terminal α-helix (α1), and one histone from *Ca.* Lokiarchaeota GC14-75 is reduced in length at the C-terminus. The remainder of the C-terminal amino acids likely does not form a C-terminal helix (α3) in this histone from *Ca.* Lokiarchaeota. Although histones of reduced length or containing tails lack part of the histone fold, they likely still possess DNA-binding properties. Therefore, they possibly have functional roles in the regulation of genes.

Multimerization of histones

Both eukaryotic histones and HMfB form dimers, a process that is driven by a hydrophobic core (involving residues A24, L28, L32, I39, and A43 in HMfB) as well as a crucial salt bridge for a stable histone fold (R52-D59 in HMfB) (217). These hydrophobic residues and the salt bridge are conserved among archaea. This indicates that archaeal histones have very similar tertiary structures (217, 218). Also, residues that play an important role in DNA binding are present in all examined histones, including the arginines that anchor archaeal histone dimers to the DNA minor grooves (R10 and R19 in HMfB) (217). Both eukaryotic H3-H4-dimers and HMfB dimers can form tetramers by hydrogen bonding of H49 and D59 (HMfB) and additional hydrophobic interactions in the interface (L46 and L62 in HMfB) (171), pairs of residues that, too, are generally conserved among archaeal histones (**FIGURE 2.2**).

Recently, a crystallography study on HMfB shed new light on the compaction of DNA by archaeal histones. The HMfB–DNA cocrystal structure reveals left-handed wrapping of DNA around a histone-multimer core (197) (**FIGURE 2.1**). This structure supports the model in which HMfB dimers multimerize along

DNA into an 'infinite' hypernucleosome, thereby linearly compacting the DNA approximately ten-fold. It is likely that hypernucleosomes grow or shrink by association or dissociation of dimers at both ends. The resolution of the crystal structure allowed us to identify several interacting residues between layers of dimers that may be important for stabilizing the complex (**FIGURE 2.4**). Based on this structural information, the propensity of different archaeal histones to multimerize can be predicted.

In **TABLE 2.2**, we set out three criteria for hypernucleosome formation by archaeal histones. Firstly, conservation of residues in the dimer–dimer interface (E42, L46, H49, D59, L62 and R66 in HMfB) is required, as forming a tetramer is the first step in multimerization. Secondly, residue G16, which is positioned at the stacking interface of the hypernucleosome (**FIGURE 2.4**), is crucial in permitting formation of the hypernucleosome (197). Bulkier residues at this position interfere with multimerization (197). Lastly, favorable interactions between histone dimers *i* and *i*+2 and *i*+3, here termed stacking interactions, will contribute to stability of the compacted hypernucleosome. The HMfB hypernucleosome crystal structure shows three stacking interactions, hydrogen bonds from K30 to E61, E34 to R65, and R48 to D14 (**FIGURES 2.2** and **2.4**).

Scrutiny of a selection of histone sequences reveals that most archaeal histones meet these criteria and are thus likely to form hypernucleosomes, (**TABLE 2.2**, marked +). We identified two to seven potential stacking interactions for this group of histones, which may affect hypernucleosome stability and compactness. Fewer interactions may allow for more 'breathing' of the hypernucleosome structure, yielding hypernucleosomes that are more flexible or 'floppy'. We predict such structures to be formed also by a number of archaeal histones that do not fully meet our criteria (**TABLE 2.2**, marked ±). For example, *Ca.* Heimdallarchaeota LC_3 HA and *Ca.* Lokiarchaeota GC14_75 HLkE have H49N and D59S substitutions, respectively, which likely weakens the crucial hydrogen bonding interaction at the dimer–dimer interface (171). Similarly, substitution of the hydrophobic residues 46 and 62 for more hydrophilic or bulkier ones would lead to a less stable dimer–

Caption table 2.2 (page 43): Dimer-dimer interactions in the tetrameric interface were assumed to be essential for hypernucleosomes formation. Absence of bulky residues in the first loop and a high number of potential hydrogen bonds in the stacking interface will enhance the compactness and stability of the hypernucleosome. Likely, uncertain and unlikely stacking ability is indicated with +, ± and -, respectively.
[a] Dimer-dimer interface includes residues at position 42, 46, 49, 59, 62 and 66.
[b] Stacking interface includes residues at positions 15-17.
[c] For all potential stacking interactions, residue numbering of HMfB was used according to the alignment in **FIGURE 2.2**.

Table 2.2: Assessment of possible hypernucleosome formation by archaeal histones.

		Dimer-dimer interface[a]	Stacking interface[b]	Potential stacking interactions[c]	Hypernucleosome formation	Histone features
Heimdall	LC_3 HA	±	+	3 (E14-R48, K26-E57, R41-E45)	±	N-terminal tail
Heimdall	LC_3 HB	+	+	5 (E30-K61, Q14-R48, R13-Q18, K27-E57, K37-E45)	+	
Heimdall	LC_3 HC	±	+	3 (N34-R65, T15-K41, Y14-Q53)	±	
Loki	GC14_75 (HLkE)	-	+	2 (D14-R48, K34-E45)	-	Truncated C-term.
Loki	CR_4	+	+	2 (Q14-D48, Q41-Q41)	+	
Odin	LCB_4	+	-	3 (K30-Q61, K14-E18, E38-R41)	±	
Thor	SMTZ1-45	+	+	5 (Q30-D61, E34-K65, K14-E48, E37-R41, E26-K58)	+	
Woese	CG1_02_33_12	+	+	4 (R14-T48, R34-E61, E26-K57, E37-R41)	+	
Pace	CG1_02_31_27	+	+	4 (S30-K61, E34-K65, K14-T48, E37-K45)	+	
Aenigm	CG1_02_38_14	+	+	5 (E30-K61, D34-R65, E14-H48, A15-K41, E37-K41)	+	
Micr	*M. acidiphilum* ARMAN-2	+	+	3 (E30-K61, K34-Q65, Y2-K48)	+	
Nanohalo	*Haloredivivus* sp. G17	±	-	2 (E27-R61, Q37-E45)	-	Truncated N-term.
Nanohalo	*Nanosalina* sp. J07AB43 (HA)	+	+	5 (Q30-K61, D34-R65, K14-E48, K14-E18, Q37-Q45)	+	
Nanohalo	*Nanosalina* sp. J07AB43 (HB)	-	±	2 (Q30-R61, D14-K18)	-	
Nano	*N. equitans* Kin4-M	+	+	4 (E30-R61, Q14-K48, Q14(bb)-R41, K37-E45)	+	
Aig	*C. subterraneum*	±	+	4 (K30-E61, K14-Q48, E27-K57, K41-E45)	±	
Thaum	*N. gargensis* Ga9.2	+	+	4 (E34-K65, K14-E18, E27-R61, E37-K41)	+	
Bathy	B23	±	±	4 (R14-V44, E34-K61, E37-R41, E26-R58)	±	N-terminal tail
Bathy	B24	+	+	3 (E34-R65, K14-E18, E27-R61)	+	Truncated N-term.
Bathy	SMTZ-80	+	+	3 (E34-K65, K41-E45, E27-R61)	+	
Cren	*C. maquilingensis* IC-167	+	+	4 (D30-K61, N34-R65, K14-E18, Y37-K48)	+	
Cren	*T. pendens* Hrk5	+	+	4 (E30-K61, S14-R48, R37-E45, R13-E18)	+	
Cren	*V. distributa* DSM14429	+	+	4 (D30-K61, Y34-R65, K14(bb)-R48, K14-E18)	+	
Eury	*M. wolinii* (HA)	±	+	4 (N30-E61, E34-K65, E14-K48, K41-E45)	±	
Eury	*M. wolinii* (HB)	+	+	5 (E30-K61, E34-K65, N14-R48, N14-Q18, Q41-Q41)	+	
Eury	*M. jannaschii* DSM2661	+	-	4 (N30-K61, Q14-R48, K37-Q45, D26-R58)	±	C-terminal tail
Eury	*M. methylutens*	±	-	2 (D30-K61, S14-E18)	-	
Eury	*T. kodakarensis* KOD1 (HTkB)	+	+	4 (E30-K61, E34-K65, K14-Q48, K26-E58)	+	
Eury	*M. fervidus* DSM2088 (HMfB)	+	+	3 (K30-E61, E34-R65, D14-R48)	+	

dimer interface, as for *Ca.* Heimdallarchaeota LC_3 HC and *Ca.* Bathyarchaeota B23. In the presence of the canonical dimer–dimer interface, bulky substitutions at position 16 likely also result in a more open hypernucleosome structure, as for *Ca.* Odinarchaeota LCB_4.

Three archaeal histone species fail multiple criteria in our analysis, indicating that these cannot form hypernucleosomes. These histone species are *Haloredivivus* sp G17, *Nanosalina* J07AB43 HB, and Euryarchaeal *Methanococcoides methylutens* (class Methanomicrobia) that all combine defects in the dimer interface with a bulky substitution at position 16 and few potential stacking interactions (**TABLE 2.2**, marked –). In particular, *Nanosalina* J07AB43 Histone B (HB) shows a H49D substitution and a glutamic acid at position 62, making the dimer surface highly negatively charged and thus very unlikely to interact with another dimer.

It is remarkable that most of the histones having N- or C-terminal tails or N- or C-terminal truncations additionally have substitutions in the dimer–dimer and/or stacking interface that will affect hypernucleosome formation. Histones with reduced ability to form compact hypernucleosomes are expected to exhibit different roles in shaping the genome, like simple DNA bending or site-specific interference with histone multimerization. Interestingly, the genomes of several organisms encode histones that we predict are able to multimerize as well as histones that probably do not multimerize. Histones have been hypothesized to bind to promoters, thereby making the promoter inaccessible to transcription factors that onset gene expression. The ability to multimerize into hupernucleosomes suggests that histones may also have an effect on gene expression via multimerization, thereby potentially silencing the gene by binding into the coding region.

Sequence analysis of archaeal histones

Although a sequence alignment of a selection of histones from throughout the archaeal domain as shown in **FIGURE 2.2** and a structural analysis as discussed in **TABLE 2.2** give a good overview of the multeity of structural and sequential elements, a more systematic sequence analysis of all known archaeal histones allows for a better characterization of phylum-specific features of archaeal histones. Here, we will discuss the histones of all archaeal phyla and all euryarchaeal classes. Sequences of these histones can be found in the supplementary Excel file (accessible via the Leiden Repository). Main points of discussion are length and isoelectric point (pI), the residues involved in DNA binding and dimerization, the well-conserved 'HAGRKT-motif' at the C-terminal linker, the stacking interface or AGA-loop situated at the N-terminal linker, and the stacking interactions (as

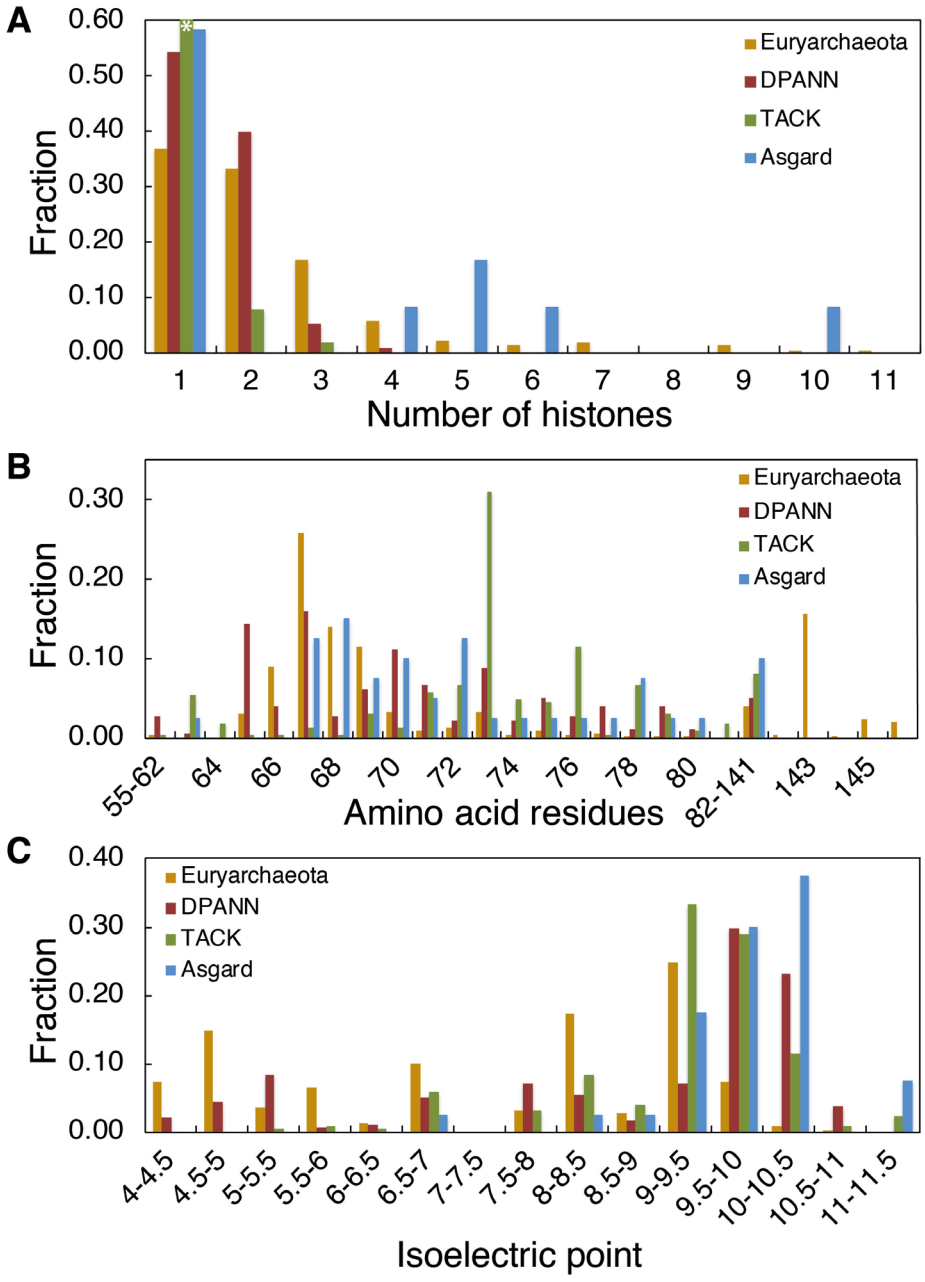


Figure 2.5: Characteristics of archaeal histones categorized by superphylum. A) Histones per genome, as fraction of the total number of genomes per category. The double-histone-fold histones of Halobacteria are not included in this analysis. *) The value of TACK at one histone is 0.90. Total number of genomes included: Euryarchaeota, 228; DPANN, 118; TACK, 203; Asgard archaea, 12. B) Division of histone lengths, as fraction of the total number of histones per category. Total number of histones: Euryarchaeota, 662; DPANN, 181; TACK, 225; Asgard archaea, 40. C) Division of theoretical isoelectric point, as fraction of the total number of histones per category. Total number of histones: *see* B.

found in HMfB). Histones from genomes that were not assigned to a phylum or class, are disregarded. Some histones have been identified as being part of the same histone type within a taxon. These histones are more similar to histones within the same type than histones from the same organism, and are identified using sequence alignment with Clustal Omega (*see also* **Methods**). Histone types are assigned independently of other types, and so histones of a certain type in one taxon do not have any relation with histones from another type with the same name in another taxon.

Euryarchaeota

In all superphyla and Euryarchaeota, the mode of number of histones encoded per genome is one. However, in Euryarchaeota, 63% of the genomes evaluated in this study encode two or more histones (**figure 2.5A**). Most euryarchaeal histones are between 65 and 70 amino acid residues long, with the double-histone-fold histones from Halobacteria being the main exception, at 142-150 amino acid residues (**figure 2.5B**). The theoretical isoelectric point (determined using ProtParam (219)) of euryarchaeal histones varies between 4 and 11 (**figure 2.5C**). The pI of histones from halophilic organisms is usually between 4 and 6, whereas the pI of histones from other organisms is mostly between 7.5 and 10.

Archaeoglobi

The majority of Archaeoglobi genomes encodes two histones, although some encode three histones. There is little variance between the histones, and unlike histones from for example Thermococci or *Ca.* Nanohaloarchaeota, no types can be identified. Therefore, it remains unclear how the two or three different histones in Archaeoglobi are functionally different. Likely, Archaeoglobi histones are able to bind DNA. The number of tetramerization interactions differs. Some histones can probably form five tetramerization interactions, while others can form only four. The AGA-loop is conserved among all histones. Stacking interactions can likely be formed by most histones in this class, which may result in formation of hypernucleosomes.

Hadesarchaea

The four genomes of Hadesarchaea included in this study encode one, two (twice) or three histones. The AGA-loop is found in only two histones, but four others have SGA instead, which may lead to a more flexible hypernucleosome. Two histone-coding genes do not encode an AGA-loop. The hadesarchaeal histones

without AGA or SGA are predicted to be able to bind DNA and tetramerize, like the others, but may do this with fewer interactions. Affinity for DNA might therefore be lower than for the other histones. The histones with AGA or SGA loop are able to form at least two stacking interactions, and may therefore (at least, if SGA does not sterically hinder it) assemble into hypernucleosomes. For the other two, this is unlikely, not only because the amino acid residues in their N-terminal loop probably cause sterical hindrance and thereby prevent dimerization, but also because these histones do not have enough stacking interactions.

Halobacteria

Halobacteria express a unqiue kind of histones, with two histone folds. These histones can be regarded as naturally linked dimers. Because of the double histone fold, they are much longer than other histones. Their length is 140-150 amino acid residues, with a mode of 143 residues. Within Halobacteria, histone sequence is highly conserved (FIGURE 2.6). The pI of halobacterial histones ranges from 4.24 to 5.37, which is very low compared to other archaeal histones, although these values are common for salinophilic archaea. Because of its two histone folds, one histone fold of the halobacterial histone is able to 'specialize' in certain interactions (182). Therefore, not all residues involved in DNA-

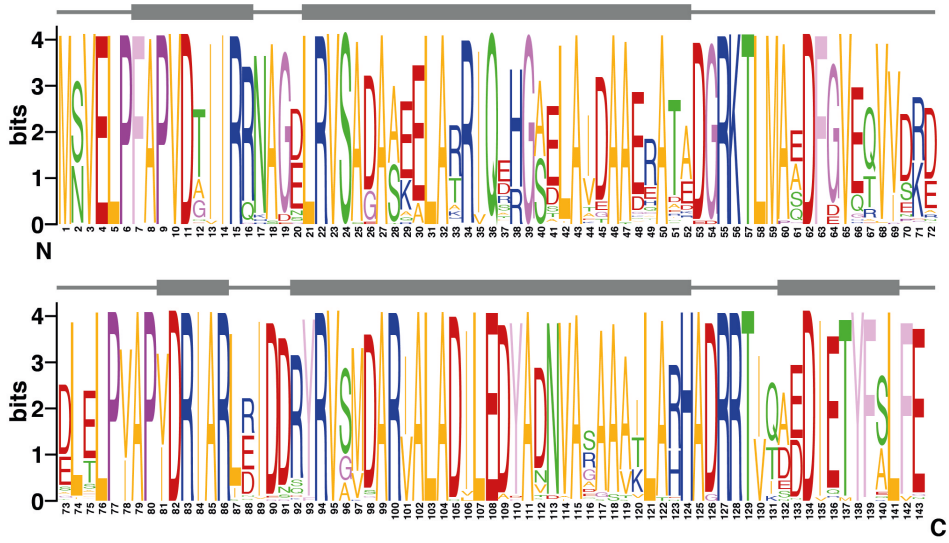


Figure 2.6: Sequence logo of histones from Halobacteria. The top panel represents the N-terminal histone fold, the bottom panel represents the C-terminal histone fold. 104 histones with a length of 143 residues were included in this figure. Grey bars represent predicted α-helices. Figure made using WebLogo (224).

binding- and tetramerization interactions are present in one of the histone folds, which makes it difficult to discuss those interactions here. From *in vivo* experiments, it is known that halobacterial histones bind DNA as tetramers, and probably do not form hypernucleosomes (220, 221). Histones from Halobacteria do not have an AGA-loop, but in the N-terminal histone fold, XGA is often present at positions 15-17. Only the second half of the HAGRKT-motif is found in the N-terminal histone fold, whereas the C-terminal histone fold usually contains a full (variant of the) HAGRKT-motif. The least conserved part of halobacterial histones is the central α-helix and C-terminus of the N-terminal histone fold. Positively charged and hydrophobic residues are generally well-conserved (**FIGURE 2.6**). *In vitro* studies suggest that histone HpyA from *Halobacterium salinarum* plays a minor role in DNA compaction, but is important for growth phase-dependent transcription regulation (185). MC1 may be the main DNA compacting protein in Halobacteria (222, 223).

Hydrothermarchaeota

Two histone-encoding genomes of Hydrothermarchaeota were identified, both encoding one histone. These histones are identical. Remarkably, this histone does not contain an AGA loop. Also, no potential stacking interactions could be identified, which would make hypernucleosome formation highly unlikely. Hydrothermarchaeal histones are likely able to bind DNA and to tetramerize, as residues generally involved in these processes are present. Most hydrothermarchaeal genomes encode Alba proteins (225). Based on the hypothesized inability of histones to form hypernucleosomes, Alba may be the main chromatin protein responsible for DNA compaction in Hydrothermarchaeota.

Methanobacteria

The histone distribution of histone-coding genes in Methanobacteria is highly diverse. Of the 42 genomes that we have examined, half encoded three histones. However, two genomes encoded a single histone, whereas one genome encoded ten histones and one genome encoded eleven histones. Methanobacterial histones as a group cannot be categorized into types, since histones are better conserved within a certain genus than within subgroups. The class Methanobacteria contains only one order, Methanobacteriales, which is subsequently subdivided into two families: Methanobacteriaceae and Methanothermaceae. The former is further subdivided into the histone-encoding genera *Methanobacterium*, *Methanobrevibacter*, *Methanothermobacter* and *Methanosphaera*, whereas the latter is made up of a single genus: *Methanothermus*.

Genomes from the genus *Methanobacterium* mostly encode three histones, with the main exception being *Methanobacterium bryantii* M.o.H., encoding seven histones. The histones of *Methanobacterium* seem to be equal in terms of DNA binding and stacking interactions, but based on tetramerization interactions, three types can be discerned, which form four, five or six stacking interactions. Presence of the AGA-loop and putative tetramerization and stacking interactions suggest that these histones can assemble into a hypernucleosome.

Most of the genomes from *Methanobrevibacter* encode three or four histones. The histones can be categorized into three types, A, B and C. Histones from type A likely form more stacking interactions than the others, which in theory could result in more stable hypernucleosome formation. However, the histones of type A lack H49, which is very important in tetramerization. Thus, whether type A histones are able to form hypernucleosome remains unclear. This is supported by the structural analysis of this chapter, in which we concluded that histone HA of *Methanobrevibacter wolinii* may or may not form hypernucleosomes. Histones of type B and C are likely able to multimerize into hypernucleosomes, and do not differ in residues that are important for DNA binding, tetramerization or stacking. Within *Methanobrevibacter*, large differences between the pIs of histones are observed (5.10-9.52). This may have an effect on transcription regulation in response to environmental cues.

All genomes of the genus *Methanothermobacter* included in this study encode three histones. These histones form three types, D, E, and F. Type D is probably slightly impaired in tetramerization, as it forms five tetramerization interactions while the histones from the other types form six tetramerization interactions. Unlike *Methanobrevibacter* histones, all types contain the important and well-conserved H49, and therefore tetramerization is likely not affected in *Methanothermobacter* histones. Interstingly, the histones from type E exhibit more potential stacking interactions than the others. Histones from type D or F do not differ at residues involved in any known interactions. The AGA-loop is also present, which together with the likely presence of tetramerization and stacking interaction leads to the conclusion that *Methanothermobacter* histones of any type probably form hypernucleosomes. The histones from *Methanothermobacter thermautotrophicus*, HMtA1 (type D), HMtA2 (type E) and HMtB (type F) were studied *in vitro*. Substitution R19I in HMtB, found in a laboratory strain, resulted in the loss of DNA binding of the HMtB homodimer. However, $HMtB_{R19I}$-HMtA2 heterodimers were able to bind DNA (226).

Methanosphaera is the genus of which the genomes encode on average the largest

number of histones: between four and eleven. These histones form seven loosely-defined types, and histones from these types are not equally distributed among the *Methanosphaera* genomes. Histones from most types exhibit six potential DNA-binding interactions, although some types are likely able to form just five DNA-binding interactions. Histones from most types form five tetramerization interactions, but this number varies from three to six. The histones of one type, expected to form just 3 tetramerization interactions, also lack H49, which means the histones belonging to this type likely cannot tetramerize or form hypernucleosomes. The number of potential stacking interactions varies even within types, and therefore it is impossible to make statements about the hypernucleosome formation abilities of the types. Similar to *Methanobrevibacter*, within *Methanosphaera*, large differences between the pIs of histones are observed (4.64-9.77), which may be of importance for transcription regulation in response to environmental cues.

Within the genus *Methanothermus*, only three species have been identified: *Methanothermus fervidus*, *Methanothermus sociabilis* and *Methanothermus jannaschii*. Histones are exclusively encoded by *M. fervidus*. These histones, HMfA and HMfB, are regarded as archaeal 'model histones', and are very well studied (*see also* **CHAPTER 4** and **6** of this thesis). Hypernucleosome formation was first shown for *M. fervidus* histones (197), and sequence and structural analysis in this chapter is based on experiments on HMfA and HMfB.

Methanococci

Of the 17 genomes included in this study, eight encode three histones, while the others encode two to five histones. Some histone-coding genes encode C-terminal tails with a length of ~32 residues, which are relatively well-conserved. Most genomes encode at least one histone with C-terminal tail, possibly forms an additional α-helix (identified with Jpred4 (227)). The function of this tail remains unclear. Remarkably, the histones with C-terminal tail likely have very different properties compared to Methanococci histones without a tail. The histones with tail have a poorly conserved HAGRKT-motif and lack an intact AGA-loop. They almost completely lack residues involved in tetramerization, and are probably unable to form stacking interactions. This means that it is very unlikely that histones with a C-terminal tail form hypernucleosomes. However, these histones can likely bind to DNA. In contrast, Methanococci histones without tail have an AGA-loop and can probably form DNA-binding-, tetramerization- and stacking interactions, which leads to the conclusion that they likely form hypernucleosomes. Most Methanococci genomes encode two histones without tail that are very similar. One of the two often contains a C-terminal methionine of which the function is unknown, whereas the other does not. The clear difference between

Methanococci histones with and without tail indicates that the histones without tail form hypernucleosomes, while histones with tail regulate hypernucleosome length or position hypernucleosomes onto specific loci.

Methanomicrobia

Of the 120 Methanomicrobia histones included in this study, more than half is the only histone encoded by its genome. ~75% of histones have an intact AGA-loop. Most histones are likely able to form DNA-binding- and tetramerization interactions, although a large minority of histones has one or more tetramerization defects. The histidine at position 49 is conserved in 99% of histones. Stacking interactions are found in Methanomicrobia histones, although the number varies from zero to three. Therefore, some histones likely form hypernucleosomes, some likely do not, and for some the analysis is inconclusive. The histone of *Methanococcoides methylutens* DSM 2657, the only histone encoded by this genome, are likely unable to form hypernucleosomes, as shown by the structural analysis (**table 2.2**). Thus, hypernucleosome formation is likely nonessential in Methanomicrobia. Some histones in this class have a highly charged C-terminus, which may be used for transcripton factor recruitment. *Methanosarcina*, a genus within Methanomicrobia of which the genome encodes one histone, also expresses MC1 (formerly known as HMb). In this genus, MC1 was identified as the major component of extracted nucleoprotein, which may indicate that this NAP is the most important protein involved in genome compaction (228-230).

Methanonatronarchaeia

One genome from Methanonatronarchaeia encoding one histone was included in this analysis. This histone is likely able to form DNA-binding-, tetramerization- and stacking interactions. Its AGA-loop is not fully conserved, but instead it has GGA at positions 15-17. The A15G substitution probably does not interfere with hypernucleosome formation, since glycine does not possess a side chain and therefore it likely does not sterically hinder multimerization of histones.

Methanopyri

Of the six Methanopyri genomes that have been identified, only *Methanopyrus kandleri* AV19 encodes histones. The three gene products annotated as histones are vastly different from each other. One gene product is 77 amino acid residues long, has a pI of 9.50 and lacks features that are associated with histones. The gene product is predicted to contain a histone fold (identified with Jpred4 (227)). It is

therefore questionable if this gene product really is a histone, or if it was wrongly annotated. Another histone is 91 residues long, has a pI of 9.15 and has a HAGRKT-motif. It exhibits some similarities the histones with tail from Methanococci, and also unlikely to tetramerize or form hypernucleosomes. The third histone has 154 residues, a pI of 4.91 and has been shown to contain two histone folds (231). This histone, referred to in literature as MkaH, thus somewhat resembles halobacterial histones, although they lack the AGA-loop and HAGRKT-motif. Although there is little sequence similarity to other archaeal histones, *in vitro* experiments have shown that MkaH is able to dimerize into structures that resemble eukaryotic (H3-H4)2 tetramers, and to compact DNA (232).

Theionarchaea

Although only two genomes of Theionarchaea are included in this study, combined they encode twelve histones, five from archaeon DG-70 and seven from archaeon DG-70-1. The AGA-loop is present, although it is substituted with GGA in some cases. This likely does not interfere with hypernucleosome formation. Most of the histones are expected to multimerize into hypernucleosomes. However, one histone of each genome probably cannot tetramerize and therefore cannot form hypernucleosomes. Four out of twelve histones are probably able to form sufficient stacking interactions in order to form stable hypernucleosomes. The other histones either cannot form hypernucleosomes because of the before-mentioned defects in tetramerization interactions, or cannot be predicted to form hypernucleosomes based on this analysis.

Thermococci

Over 75% of Thermococci genomes encode two histones. Of these two histones, one usually has a lower theoretical pI than the other, the difference being approximately 1. We categorized thermococcal histones into type A, histones with lower pI, and type B, histones with higher pI. Although length and sequence of the histones of both types are relatively similar when compared to histones from other classes or phyla, there are four clear differences. E19, well conserved in type A, is poorly conserved in type B and can be alanine, glutamine or proline (**FIGURE 2.7**). The aromatic tyrosines in type A at position 32 and 36, are replaced by the positively charged residues histidine and lysine in type B, respectively. Finally, the aromatic phenylalanine or tyrosine at position 47 is not present in type B, which instead has leucine at position 47. The first three residues do not have a known effect on DNA binding, tetramerization, or hypernucleosome formation, however, L47 is involved in tetramerization, but this substitution was shown to

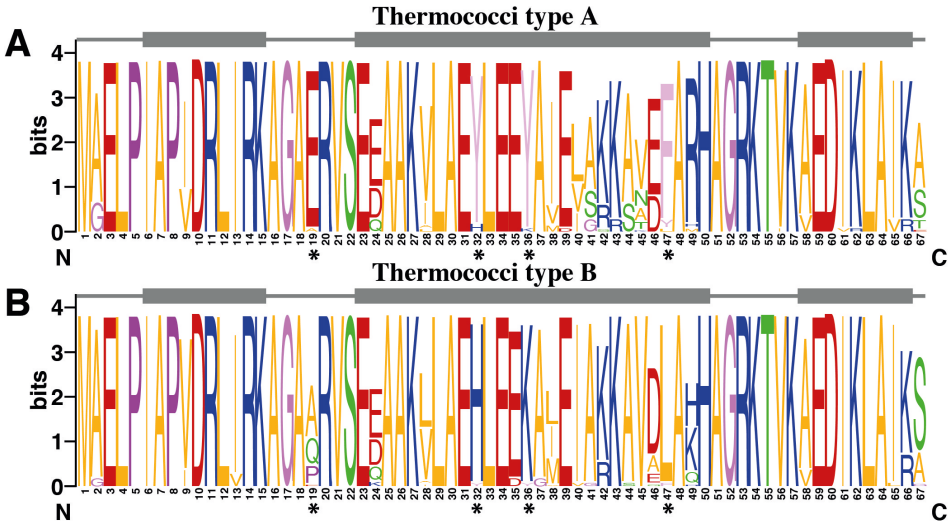


Figure 2.7: Sequence logo of the two types of thermococcal histones. A) Type A and B) type B. Type C was not included because of the heterogeneity in histone length. Grey bars represent predicted α-helices. Main differences between type A and type B are represented by asterisks. Figure made using WebLogo (224).

have no effect in electrophoretic mobility shift assays (EMSAs) (233). Despite these differences, histones from both types can likely multimerize into hypernucleosomes and bind DNA. This is supported by the finding that type A histone HTkA from *Thermococcus kodakarensis* forms hypernucleosomes *in vivo* (197). Most, but not all Thermococci histones can be placed in either types A or B. Type C, which represents the remaining histones, is probably less conserved than types A or B. The length of some type C histones does not exactly match the lengths of histones from type A or B, and the number of potential tetramerization interactions varies within type C. Additionally, some type C histones likely form weaker stacking interactions, which may have its effect on hypernucleosome formation. The genomes to which thermococcal type C histones belong, were all identified recently using metagenomic sequencing. The habitat from which they were retrieved does not differ from that of other Thermococci genomes (225).

Thermoplasmata

Most Thermoplasmata genomes do not encode histones. Currently, only five, all identified via metagenomic sequencing, encode a single histone. One of these, although annotated as histone, is missing all signature motifs and conserved elements, and is therefore probably not a real histone. It is therefore disregarded in this analysis. The remaining histones all have an AGA loop, but the number of potential stacking interactions differs, with some histones likely able to form

a hypernucleosome, while others are likely unable to do so. Also the expected number of DNA interactions and tetramerization interactions varies. It is therefore difficult to predict the function of Thermoplasmata histones. However, *Thermoplasma acidophilum* expresses HTa, a NAP that is a member of the bacterial HU protein family (234, 235). It is one of the most highly expressed proteins in *T. acidophilum*, and is a functional analog of histones (235, 236). Also, Alba is encoded by most Thermoplasmata genomes. HTa and Alba may together compact and regulate the genomes of Thermoplasmata.

DPANN

In DPANN, genomes that encode more than one histone, make up 56% of the data set. In contrast to Euryarchaeota, genomes encoding more than four histones were never found (**figure 2.5A**). Histones from DPANN come in a wide range of lengths, although most histones are between 65 and 70 amino acid residues long (**figure 2.5B**). Also, a broad range of theoretical pIs of histones can be found in DPANN, although the average pI is significantly higher than in Euryarchaeota (8.72 compared to 7.31) (**figure 2.5C**). Similar to Euryarchaeota, genomes found in high-salinity habitats, such as Nanoarchaeota and *Ca.* Nanohaloarchaeota, encode histones with low pIs (pI 4-6).

Ca. Aenigmarchaeota

Most of the twenty genomes of *Ca.* Aenigmarchaeota that encode histones, have one or two histone-coding genes. The AGA loop is usually intact, although some variations occur, such as YGA, LGA, NGA, VGA, CGA and SGA. It is not known what the effect of these AGA-loop variants is with respect to hypernucleosome formation. However, sequences of histones with an AGA-loop variant are likely less conserved in general. Potential stacking interactions are exhibited by a majority of aenigmarchaeal histones, which, together with the presence of the AGA-loop, makes it likely that those histones are able to form a hypernucleosome. The six residues involved in DNA binding are found in aenigmarchaeal histones, but none of the histones contains all six residues. This may result in distinct DNA binding affinities within this group of histones, and even within the same organism. Residues involved in tetramerization are usually found in aenigmarchaeal histones, although also here, some defects are found.

Ca. Altiarchaeota

Of the eleven genomes of *Ca.* Altiarchaeota included in this study, eight genomes

encode one histone. Of the AGA-loop, only the glycine is conserved in all histones, and the majority of histones has a motif that is different from AGA. A subset of histones may be able to form two stacking interactions and is therefore possibly able to form hypernucleosomes, whereas the remaining histones likely do not stack and thus cannot form hypernucleosomes. Also, the histones without stacking interface are partially defective in tetramerization. These histones are somewhat longer than the others, with an extension at the C-terminus. Since DNA interaction is likely possible for all histones, the non-stacking histones may function as a regulator for hypernucleosome extension. All genomes with more than one histone-coding gene encode one non-stacking histone.

Ca. Diapherotrites

The majority of *Ca.* Diapherotrites genomes does not encode histones. Only four histones from *Ca.* Diapherotrites are included, encoded by four different genomes. Remarkably, the AGA-loop is absent in all histones, although the variant GGA was found in one histone. Two histones likely form stacking interactions. All histones are likely to be able to tetramerize and bind DNA. Since no histone contains amino acid residues at positions 15-17 that cause no steric hindrance when multimerizing, and also is predicted to form stacking interactions, it is likely that none of the histones from *Ca.* Diapherotrites assemble into hypernucleosomes.

Ca. Huberarchaeota

Although we included twelve histones from seven huberarchaeal genomes, effectively there are only two histones included in this analysis, since all huberarchaeal histones are identical to either of those two histones. The histone of 91 amino acid residues contain an AGA-loop, whereas the 79-residue histone has a variant of which it is unknown if it allows for hypernucleosome formation. Stacking interactions are identified in both histones, albeit between different residues. The long histone is likely able to form a hypernucleosome, but the analysis is inconclusive for the short histone. Residues of the DNA-binding interface are best conserved in the short histone, whereas tetramerization interactions are best conserved in the long histone. The long histone has an N-terminal tail that is highly positively charged because of lysines, as discussed elsewhere in this chapter, whereas the short histone has a short C-terminal extension with positive charges due to the presence of lysines and arginines. We hypothesize that the long histone is involved in hypernucleosome formation, whereas the short histone is able to regulate the hypernucleosome by binding to its extremities and possibly positioning the hypernucleosome on the DNA.

Ca. Micrarchaeota

Most *Ca.* Micrarchaeota genomes encode two histones, while one-third of the micrarchaeal genomes encodes only one histone. We found a broad spectrum of micrarchaeal histones, with lengths ranging from 59 to 113 residues and theoretical pIs from 5.02 to 10.53. The histones can roughly be divided in two types: type A that is likely able to form a hypernucleosome via stacking interactions, with likely strong DNA-binding- and tetramerization interactions, and type B that lacks stacking interactions and likely has weaker DNA-binding- and tetramerization interactions. Types A and B are not correlated to histone length and pI. Also, the histones of the two types are not equally divided between the genomes, as some genomes encode two histones from the same type (FIGURE 2.8). This suggests contradiction with the hypothesis that the histones of a single organism have different functions, with one being able to form a hypernucleosome and one involved in a different process, for example regulation of the hypernucleosome. It should be noted though that differences on a sequence level between two histones of the same type may result in a yet unidentified from of differentiation. Also, encoding of multiple histones of the same type may be a mechanism for increasing histone expression levels. Still, the majority of genomes encoding two histones encode a histone from either type. Four type B histones have a shorter N-terminal α-helix, which may also contribute to a differential function in genome compaction or transcription regulation.

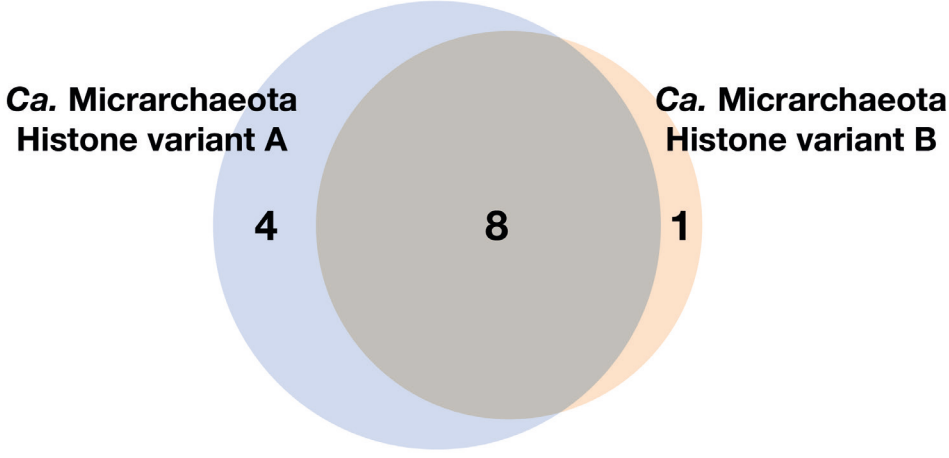


Figure 2.8: Venn diagram of micrarchaeal genomes encoding two histones, and the presence of histone type A (likely forming a hypernucleosome) and type B (likely unable to form a hypernucleosome).

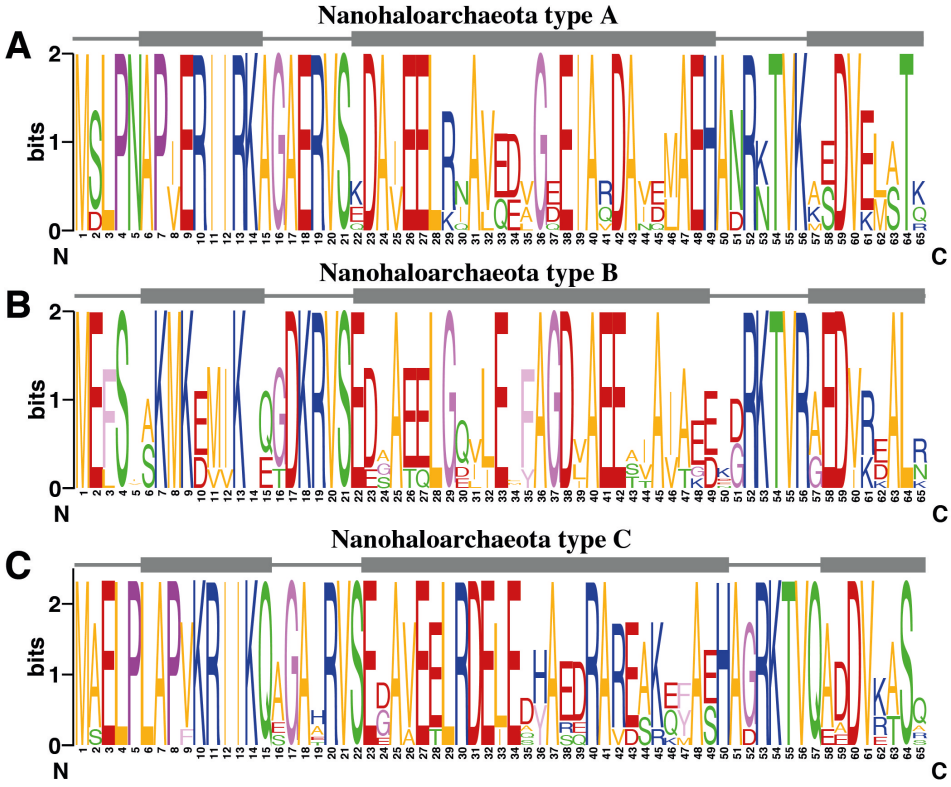


Figure 2.9: Sequence logo of the three types of nanohaloarchaeal histones. A) Type A, which is likely able to form hypernucleosomes. B) Type B, which is unlikely to form hypernucleosomes. C) Type C, which may form hypernucleosomes. Grey bars represent predicted α-helices. Figure made using WebLogo (224).

Nanoarchaeota

Like *Ca.* Micrarchaeota, within the five genomes of Nanoarchaeota that were included in this study, two types can be distinguished. Here, every genome encodes one histone of each type (with Nanoarchaeota archaeon B36_G17, encoding only one histone, being the exception). Histones of type A have a fixed length of 75 residues, whereas histones of type B are always longer. The types are not related to pI, DNA interactions or tetramerization interactions, but they do seem to be related to the presence of potential stacking interactions, with histones of type A generally being expected to form stacking interactions while histones of type B are not. Type B histones also have AGV instead of the AGA-loop, which may prevent hypernucleosome formation for the members of this type. Here, again, histones of type A are potentially involved in hypernucleosome formation, and

histones of type B likely have different structural properties. The type B histone of *Ca. Nanobsidianus stetteri* has a length of 103 residues and is predicted to contain an N-terminal tail. However, an internal methionine at exactly the location where other histones of the same type have the N-terminal methionine, suggests that this gene was misannotated. The internal methionine is in reality likely the N-terminal methionine, and the N-terminal tail is probably never translated.

Ca. Nanohaloarchaeota

Histones of *Ca.* Nanohaloarchaeota are, like histones from other (presumably) salinophilic species, characterized by their low (<6) theoretical pI. Most histones have a length of 65 amino acid residues. Within the nanohaloarchaeal histones, three types can be distinguished. Type A has an intact AGA loop, six DNA-binding interactions and six tetramerization interactions (**FIGURE 2.9A**). Also, it has several tetramerization interactions, although the exact number varies per histone. Therefore, histones of type A can likely form a hypernucleosome (*see also Nanosalina* sp. J07AB43 HA in **FIGURE 2.2** and **TABLE 2.2**). Histones of type B completely lack the AGA loop, and also most histones of type B cannot facilitate stacking (**FIGURE 2.9B**). Additionally, H49, important for tetramerization, is lacking. Therefore, it is highly unlikely that histones of this type are involved in hypernucleosome formation. Amino acids involved in DNA binding are present, although with five interactions, they are slightly smaller in number compared to type A. This may cause slightly weaker DNA binding affinities. Lastly, histones belonging to type C have an AGA loop in which at least the glycine and C-terminal alanine are well conserved (**FIGURE 2.9C**). Also, they are able to form stacking interactions, but less in number than histones of type A. Instead of six DNA-binding and tetramerization interactions, histones of type C probably form five of both. It is unclear if these histones are able to form a hypernucleosome, but if they do, type C hypernucleosomes are likely less stable than those of type A. In this phylum, the stability of the hypernucleosome may therefore be regulated by expressing histones of type A or C. Genomes that encode three histones, encode all three variants. Genomes encoding two histones encode either A and B, A and C or two histones of type B. Five nanohaloarchaeal histones were difficult to categorize. These histones are either shorter or longer than 65 residues, and in some cases their pIs are much different from the other histones.

Ca. Pacearchaeota

A minority (15%) of pacearchaeal genomes encodes histones, while Alba is encoded by most genomes belonging to *Ca.* Pacearchaeota. All pacearchaeal genomes investigated here encode a single histone. The AGA-loop is only present

in one histone, other histones have variants, such as LGA and CGA. Four out of six histones that were investigated here, are likely able to form stacking interactions. However, the histone with the AGA-loop is unlikely to form stacking interactions. This suggests that none of the pacearchaeal histones are able to assemble into a hypernucleosome. In contrast, we concluded from the structural analysis that at least the histone from genome Pacearchaeota CG1_02_31_27 is able to form a hypernucleosome, despite having CGA instead of AGA. We do not expect the other histones to form hypernucleosomes as well: only two out of six histones are likely able to form tetramers; the other four histones likely cannot form structures larger than dimers. Pacearchaeal histones lack potential DNA-binding interactions and are therefore probably unable to bind DNA.

Ca. Parvarchaeota

In the genomes of *Ca.* Parvarchaeota, no genes coding for histones have been identified to date.

Ca. Woesearchaeota

The genomes of *Ca.* Woesearchaeota encode one, or in some cases two histones. These histones can likely bind DNA and form tetramers. Some woesearchaeal histones are probably able to form stacking interactions, although the AGA-loop is conserved in less than half of the histones. Variants of the AGA motif similar to those in *Ca.* Pacearchaeota are often found, which means we cannot rule out the possibility of hypernucleosome formation for the majority of the histones of *Ca.* Woesearchaeota. The histones from this phylum cannot be grouped into types, as seen for some other DPANN histones. There are relatively few conserved residues as compared to other phyla, although length and pI are very constant, considering the variation in sequence.

TACK

Over 90% of TACK genomes encode a single histone; the maximum is three histones encoded by one genome (**figure 2.5A**). Histone length is variable, although almost a third of the TACK histones in this study consist of 73 amino acid residues (**figure 2.5B**). The pI of TACK histones is relatively high, mostly between 9 and 10.5, with an average of 9.19 (**figure 2.5C**).

Ca. Bathyarchaeota

50 out of 60 genomes of *Ca.* Bathyarchaeota in this study encode one histone. The remaining ten genomes encode two or three histone homologs. Bathyarchaeal histones range from 62 to 105 amino acid residues in length, although almost half of the histones in this study consist of exactly 73 residues. Also the most extreme theoretical pIs are far apart (from 5.06 to 11.30), but most pIs lie between 9 and 10.5. More than 85% of the histones have an AGA-loop. Those histones seem to share common features of archaeal histones: DNA binding, tetramerization, and multimerization into hypernucleosomes. The analysis suggests that the remaining group of histones form less DNA-binding interactions and less tetramerization interactions, and are not able to form stacking interactions. Remarkably, most of the histones without conserved AGA-loop are identified as the only histone encoded by the genome it belongs to. This suggests that these histones do not regulate transcription via hypernucleosome modulation, as hypothesized for *Ca.* Huberarchaeota and *Ca.* Nanohaloarchaeota, i.a.. A small subset of bathyarchaeal histones contains an N-terminal tail. The N-terminal histone tails within *Ca.* Bathyarchaeota do not resemble each other, nor do they resemble the tails of eukaryotic histones. Also, two histones have a short C-terminal tail, which also does not resemble any other known histone tail.

Crenarchaeota

Crenarchaeota were long known as the archaeal phylum in which histones do not occur. The chromosomes were thought to be compacted and regulated by Crenarchaeota-specific chromatin proteins such as Cren7 and Sul7 (*see also* **Chapter 1**). Metagenomic sequencing resulted in the identification of 48 genomes that do encode histones. Except for one, all genomes encode a single histone; only Thermoprotei archaeon B132_G9 encodes two. These histones likely bind to DNA and are able to tetramerize. Also, the AGA-loop is present in nearly all crenarchaeal histones, and most are able to form stacking interactions. This means that most histones may be able to assemble into a hypernucleosome. Of all histones with a single histone fold, the ones from Crenarchaeota are on average the longest, which is probably related to their short C-terminal tail. This tail is poorly conserved within the phylum. Additionally, some histone-coding genes encode short N-terminal extensions of unknown function.

Ca. Geothermarchaeota

In *Ca.* Geothermarchaeota, two genomes were identified to encode histones. One genome encodes a long, 83-residue histone, whereas the other encodes two

shorter histones. All geothermarchaeal histones are likely able to tetramerize. The AGA-loop is conserved only in the two short histones. One of these histones is probably also able to assemble into a hypernucleosome, whereas the other cannot do this because it lacks stacking interactions. The histone that likely cannot form a hypernucleosome has six DNA-binding interactions, which possibly results in a high DNA binding affinity. The hypernucleosome-forming histone encoded by the same genome likely has a lower binding affinity, as a result of having only five residues involved in DNA binding. The non-hypernucleosome-forming histone may therefore function as an inhibitor against extension of the hypernucleosome. The long histone is probably unable to multimerize into a hypernucleosome. However, its positively charged N-terminal tail and C-terminal tail may play a role in regulation of DNA compaction and transcription. The N-terminal tail does not resemble its eukaryotic counterpart.

Ca. Korarchaeota

Although most TACK histones do not have homologs in the same genome, *Ca.* Korarchaeota is the only phylum within TACK that has more genomes that encode two histones than those that encode one histone (six and five, respectively). No histone types could be identified. The AGA-loop is conserved in most korarchaeal histones. Stacking interactions are also likely to be formed by these histones, although the combination of residues forming a stacking interaction widely varies

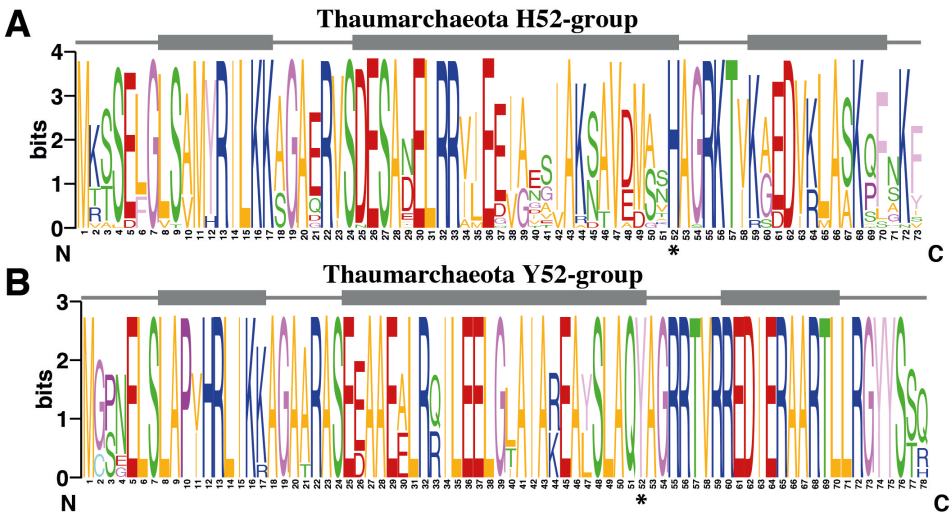


Figure 2.10: Sequence logo of the two groups of thaumarchaeal histones. A) The H52-group, which is possibly able to form hypernucleosomes. B) The Y52-group, which is unlikely to form hypernucleosomes. Asterisks indicate the residue at position 52. Grey bars represent predicted α-helices. Figure made using WebLogo (224).

in this phylum. Based on our analysis, DNA binding is likely conserved in most korarchaeal histones. There are generally less tetramerization interactions than in histones from other phyla, which may make tetramerization weaker in some histones. Therefore, it is not possible to predict hypernucleosome formation based on this analysis. Also, structural analysis proved to be inconclusive in terms of hypernucleosome formation (**TABLE 2.2**).

Ca. Marsarchaeota

In the genomes of *Ca.* Marsarchaeota, no genes coding for histones have been identified to date.

Thaumarchaeota

Of the 78 thaumarchaeal genomes encoding histones that were analysed in this study, 77 genomes encode one histone. Two histones are encoded by the remaining genome, and these histones are almost identical. The AGA-loop was found in most histones. The tetramerization-related histidine at position 52, which is well conserved in most histones throughout the domain, is in 13% of thaumarchaeal histones substituted by tyrosine (**FIGURE 2.10**). Here, remarkably, the histones with Y52 form a group that possibly are able to form stacking interactions, but their weakened tetramerization interactions may affect their ability to form hypernucleosomes. Also, histones of this group possess a short C-terminal extension. Examination of the metadata of the genomes to which the 'Y52-group' histones belong, showed that these genomes were found at deep sea hydrothermal vent sediments, whereas the genomes of the 'H52-group' histones are found in sea water and soil samples. Since this is the only clear difference between the two histone groups, and the Y52-group consists of a much smaller number of histones, we will not assign them to a type, as for other phyla. The vastly different habitat in which the organisms of the Y52-histones live, may demand for other histone properties than the H52-histones. Since the H52Y substitution in the Y52-group is the only element that possibly prevents hypernucleosome formation while all other elements required for forming hypernucleosomes are in place, it is not possible to be conclusive about the hypernucleosome formation abilities based on this sequence analysis. Structural analysis has shown that at least one member of the H52-group likely can assemble into hypernucleosomes (**TABLE 2.2**).

Ca. Verstraetearchaetoa

Of *Ca.* Verstraetearchaeota, only three histones from three different genomes are known. Two of those histones are identical, which effectively means that this

analysis includes only two verstraetearchaeal histones. In these histones, the AGA loop is conserved, and so are probably also its DNA-binding and tetramerization properties. Prediction of the ability of verstraetearchaeal histones to form hypernucleosomes is not possible, since identification of the stacking interactions via sequence analysis is challenging for these histones.

Asgard archaea

A small number of available Asgard archaea genomes leads to a small number of histones available for investigation. However, within this small population of genomes we found a relatively broad range of histones per genome and histone lengths (**FIGURE 2.5A** and **B**). The variety of lengths is partially caused by histone truncates and histones with N-terminal tails. With a value of 9.84, Asgard archaeal histones account for the highest average pI of all superphyla (**FIGURE 2.5C**). Some histones have a pI that comes close to the pI of eukaryotic histones (>11).

Ca. Heimdallarchaeota

Of three genomes of *Ca.* Heimdallarchaeota in this study, one genome encodes one histone, one genome encodes five histones and one genome encodes ten histones. Although few genomes are included, the number of histones per genome varies a lot. Additionally, three genomes were identified, which all encode one gene product annotated as histone. However, these gene products lack all main conserved amino acid residues, such as the HAGRKT-motif and the AGA-loop. Other parts do show resemblance to heimdallarchaeal histones, especially to the central α-helix, which is usually the least-conserved part of the histone. It is highly uncertain if these gene products are really histones, and therefore they were not included in the analysis. The pI of most heimdallarchaeal histones is very high, with an average of 10.05. The N-terminal tail of two of the histones encoded by *Ca.* Heimdallarchaeota archaeon LC_3 has been extensively discussed in this chapter. The histones with tail have a length of 91 residues, whereas the other heimdallarchaeal histones are between 67 and 76 residues long. Heimdallarchaeal histones exhibit the potential interactions that are involved in DNA binding and and tetramerization, except for the histones with N-terminal tail. They lack the histidine of the HAGRKT-motif, which is crucial in tetramerization. The HAGRKT-motif of other histones is conserved but in most cases the alanine is substituted by a serine. The consequenes are not clear, but DNA binding and tetramerization are likely not affected. In general, most heimdallarchaeal histones likely form hypernucleosomes, since they exhibit a stacking interface and potential stacking interactions. However, H49N substitution in histones of archaeon

LC_3 that likely express tails, may interfere with hypernucleosome formation. Also, two histones from archaeon LC_3 (which encodes ten histones) and one histone from archaeon J3-BM-08 (which encodes five histones) do not have an intact AGA-loop and cannot form stacking interactions. These histones likely cannot multimerize into hypernucleosomes.

Ca. Lokiarchaeota

Ca. Lokiarchaeota histones are possibly the group that least resembles histones from other phyla. These histones have mostly variants of the AGA-loop and HAGRKT-motif, and also they possess some but not all possible DNA-binding- and tetramerization interactions that are found in most other histones. Three lokiarchaeal genomes are included in this analysis, archaeons CR_4, GC14_75 and B53_G9, which encode four, five and six histones, respectively. Archaeon B53_G9 encodes three histones that are likely able to form hypernucleosomes and archaeon CR_4 encodes one histone that probably does the same, which is supported by the structural analysis discussed in this chapter. Histones from archaeon GC14_75 likely do not multimerize into hypernucleosomes (140), although these histones are very dissimilar to other histones, which makes it difficult to draw conclusions. Archaeon GC14_75 also encodes a truncated histone, which lacks part of the C-terminal α-helix. The N-termini of histones from archaeon B53_G9 are highly positively charged, and one histone from the same genome has a short N-terminal extension that contains three more positive charges. Also, one histone from archaeon GC14_75 and one from CR_4 have positively charged N-termini. The function of these charged regions remains unclear, but they may be involved in recruitment of other proteins or may play a role in DNA binding.

Ca. Odinarchaeota

Only one histone from *Ca.* Odinarchaeota has been identified to date. The AGA-loop is absent, its HAGRKT-motif is different from other histones (although the residues involved in tetramerization and DNA binding are not affected), and its stacking interface likely disables hypernucleosome formation. The only feature that likely is present, is DNA binding.

Ca. Thorarchaeota

Five genomes of *Ca.* Thorarchaeota encoding one histone have been found. These histones are quite different from each other. Two histones have an N-terminal tail that does not resemble the heimdallarchaeal or eukaryotic N-terminal tails. These histones have AKA instead of AGA at the N-terminal loop, which likely interferes

with hypernucleosome formation. Stacking interactions were identified, as well as DNA-binding interactions and tetramerization interactions. Two other histones do possess an intact AGA-loop, but are missing a histidine at position 49, which makes tetramerization and hypernucleosome formation unlikely. These histones are able to form stacking interactions and probably can bind to DNA. The last histone of five somewhat resembles the thorarchaeal histones with AGA-loop, but does have a conserved H49. This histone is the most likely candidate to form hypernucleosomes, while this is unlikely or uncertain for the other thorarchaeal histones.

Histones in genome regulation

MNase-seq experiments have shown that histones position upstream and downstream of a promoter region (221). This, in combination with knock-out studies showing both up- and down-regulation of transcription levels, leads to the hypothesis that histones are important for transcription regulation in the relatively well-studied phylum Euryarchaeota (117, 185, 237, 238) and may play a similar role in other histone-coding phyla. The exact mechanisms by which histones act in regulation are at this moment largely unknown. What is the mechanistic role of histones in the regulation of gene expression? Is the hypernucleosome, with a mechanism analogous to that in bacterial gene repression, able to block promoter regions and other regulatory elements, thereby making them inaccessible to the transcription machinery (239-242)? In Bacteria, such a mechanism exists for H-NS and partition protein B (ParB) proteins, in which filaments laterally spread from a nucleation site, often a high affinity DNA sequence (123, 243-245). Specific high affinity sites have been identified both *in vivo* and *in vitro* in archaea (164, 220, 221, 246). The role of such high affinity sites may be to position the hypernucleosome on the genome and could be a key feature in archaeal genome regulation. In archaea, cooperative lateral spreading of filaments has been reported for Alba proteins (116, 118, 247, 248). Also, promoter occlusion mechanisms and competitive binding of archaeal NAPs and transcription factors have been reported (117, 249, 250).

In addition, how dynamic are hypernucleosomes, and how does the cell control the size of the hypernucleosome in order for it to be functional? Is up- and down-regulation of histone expression important in fine tuning this process? Another option for control of hypernucleosome size is heteromerization of histone variants with different stacking propensity. Heteromerization of such histone variants, for instance a histone from type A and a histone from type B from *Ca.* Nanohaloarchaeota (**TABLE 2.2** and **FIGURE 2.9**), could restrict hypernucleosome size to fewer subunits. Distinct expression patterns of histone variants or histones

from different types at different growth phases or as a result of environmental cues such as osmolarity (237, 251), may alter the composition and size of the hypernucleosome. However, so far, histone variants and types of histones within a phylum or class have been poorly studied in archaea. The results of our predictions on hypernucleosome formation clearly point out the need for *in vitro* and *in vivo* studies explicitly addressing all of these questions.

Conclusion

Histones from archaea and eukaryotes are similar in tertiary but not in quaternary structure when bound to DNA. While eukaryotic histones form octamers on the DNA, archaeal histones form filaments of variable size: hypernucleosomes. Important residues responsible for DNA binding, dimer–dimer interactions, and stacking interactions are mostly conserved among archaea, including Asgard archaea, Bathyarchaeota, and other newly discovered archaea. In these recently discovered Archaeal phyla, histone tails and truncated histone variants were also found. In terms of evolution, it appears that, based on fragmentary data derived from extant lineages, the hypernucleosome has progressively become more flexible as histones with N-terminal and C-terminal tails and additional terminal helices (like in H2A and H2B in the nucleosome) developed. Furthermore, the appearance of additional DNA-binding residues and positively charged N-terminal tails may have increased the affinity of histones for DNA (252). These changes in dimer structure and DNA affinity may have stabilized octameric nucleosomes and disfavored multimerization. Specifically, the emergence of the eukaryotic H2A-H2B heterodimer blocked hypernucleosome formation since H2A lacks the dimer–dimer interface, and H2B contains an additional helix at its C-terminus that blocks the stacking interface.

The histone tails from *Ca.* Heimdallarchaeota are likely to function in similar ways as those of eukaryotic histones. They are lysine rich and potentially subject to post-translational modification, thereby possibly affecting the histone's interactions with other actors. Alternatively, they may provide stabilization of the hypernucleosome via interactions with DNA in cis or in trans. Since it is believed that eukaryotes share their latest common ancestor with *Ca.* Heimdallarchaeota, eukaryotic histones may have evolved from the predecessors of the tail-containing Heimdallarchaeal histones. As some histone proteins that have an N-terminal tail (*Ca.* Heimdallarchaeota LC_3 HA and *Ca.* Bathyarchaeota archaeon B23) seem to form less stable hypernucleosomes, these histones may represent an evolutionary transition towards a different mechanism of gene regulation, switching from regulation by multimerization and compaction toward regulation by histone tail modifications.

Although the hypernucleosome structure is suggestive of stacking interactions between dimers in adjacent turns, experimental evidence for such interactions has been lacking. However, we provide evidence for such interactions in chapter 4. Also, the functional role of tails, as well as truncates, has yet to be proven experimentally. *In vitro* hypernucleosome reconstitution experiments and *in vivo*

foot-printing assays of species expressing nonstandard histones combined with mutation of the residues proposed to be involved in stacking interactions could answer these questions. Lastly, the existence of post-translational modifications of residues in archaeal histone tails, as well as their effect on transcription regulation, remains to be discovered and would give an important insight into the evolution of transcription regulation and genome folding from archaea to eukaryotes.

Methods

Alignment of archaeal histone sequences

We have included histones from every histone-encoding (candidate) phylum within the archaeal domain in our analysis. We show different histones from the same organism if the predicted stacking properties are very dissimilar. Sequences were aligned with Clustal Omega (253) using default parameters, removing gaps.

Sequence analysis of archaeal histones

All sequences of archaeal histones that were available on 01-01-2019 via the NCBI protein database*. Only sequences of histones were included that had been annotated as histones in the database, and that originate from genomes that were assigned to a superphylum or phylum. A total of 1107 sequences were included, which are listed in the supplemental Excel file. Histone types were identified with the help of Clustal Omega (253) using default parameters.

Analysis of potential hypernucleosome formation

Structural analysis of the selected archaeal histones and assessment of potential hypernucleosome formation was done by inspecting the conservation of residues that are important for multimerization in the published HMfB hypernucleosome structure (197). Comparative multichain modeling was performed in MODELLER (254) using default parameters to construct dimer models of the archaeal histones. These models were superimposed onto HMfB dimers in the hypernucleosome crystal structure to assess whether alternative or additional interactions were possible in the different archaeal histone complexes.

Model of Heimdall HA tails in hypernucleosome

The molecular model of the histone HA dimer from the Heimdallarchaeota LC_3 genome was constructed by multitemplate modeling in MODELLER (254) using otherwise default parameters. The HMfB dimer in the hypernucleosome (197) was used as a structural template for the histone fold and eukaryotic histone H3 and H4 as structural templates for the N-terminal tails. An initial model for the Heimdall HA hypernucleosome was obtained by superimposing the HA dimer model onto HMfB in the hypernucleosome crystal structure, with either an H3-like or an H4-like tail conformation. To optimize the path of the tails through

* https://www.ncbi.nlm.nih.gov/protein

the DNA gyres and remove major steric clashes, the HA dimer model and surrounding DNA was excised from the initial model and water refined separately using High-Ambiguity Driven Docking (HADDOCK) (255), imposing ambiguous interaction restraints between HA residues 14–19 and the surrounding 3-bp section of DNA, using otherwise default parameters.

CHAPTER 3

Quantitation of DNA binding affinity using Tethered Particle Motion

This chapter was reprinted by permission from Springer Nature and based on:

Bram Henneman, Joost Heinsman, Julius Battjes & Remus T. Dame, (2018) *Quantitation of DNA affinity using Tethered Particle Motion*. In: Dame, R. (eds) Bacterial Chromatin. Methods in Molecular Biology vol 1837. Humana Press, New York, NY. doi: 10.1007/978-1-4939-8675-0_14

Abstract

The binding constant is an important characteristic of a DNA-binding protein. A large number of methods exist to measure the binding constant, but many of those methods have intrinsic flaws that influence the outcome of the characterization. Tethered Particle Motion (TPM) is a simple, cheap and high-throughput single molecule method that can be used to reliably measure binding constants of proteins binding to DNA, provided that they distort DNA. In TPM, the motion of a bead tethered to a surface by DNA is tracked using light microscopy. A protein binding to the DNA will alter bead motion. This makes it possible to measure binding properties. We use the bacterial protein Integration Host Factor (IHF) as an example to show how specific binding to DNA can be measured. Moreover, we show a new intuitive quantitative approach to displaying data obtained via TPM.

Introduction

Determination of binding affinity is an important part of the characterization of DNA-binding proteins. Typically, the binding affinity is expressed in terms of the binding constant (K_D). The K_D is the ratio of the off-rate constant, k_{off}, and the on-rate constant, k_{on}. Operationally, the K_D is defined as the concentration at which half of the substrate is bound by ligand (here: DNA and protein, respectively). For protein-DNA binding, the K_D is conventionally determined using Electrophoretic Mobility Shift Assays (EMSA), Filter Binding Assays, and Fluorescence Anisotropy (256-258). Additionally, Isothermal Titration Calorimetry (ITC), Surface Plasmon Resonance (SPR) and Microscale Thermophoresis (MST) were used (259-261). These techniques measure protein-DNA affinity in bulk, and therefore cannot distinguish different populations of bound and unbound molecules. Some of these techniques separate the protein-DNA complexes from solution or use small sample volumes, thus introducing large surface areas that may influence effective compound concentrations. Single molecule techniques such as Optical Tweezers, Magnetic Tweezers, Acoustic Force Spectroscopy (AFS) and Atomic Force Spectrometry (AFM) can be used to estimate the K_D (262, 263), but require a high level of expertise and often expensive experimental setups. Single molecule Förster Resonance Energy Transfer (smFRET) allows for measuring of DNA that is not bound to a surface and/or a bead (264), but requires fluorescently labeled DNA and protein and/or a sufficiently large and predictable conformational change of the DNA.

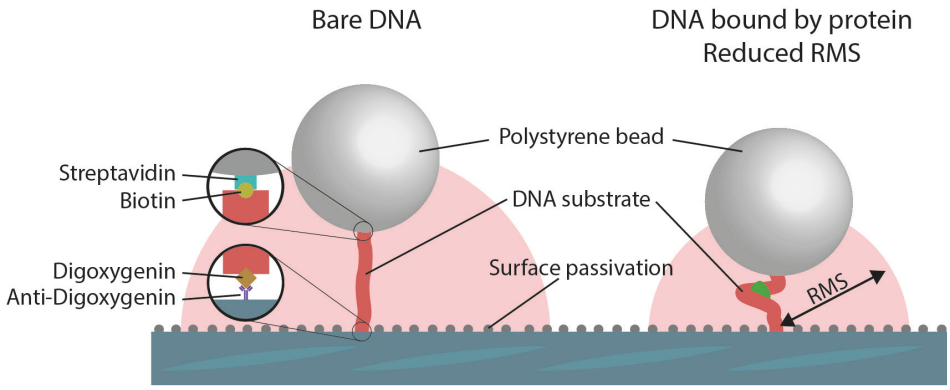


Figure 3.1: Principle of Tethered Particle Motion (TPM). A bead is tethered to a glass surface by DNA. The position of the bead is tracked over time. Proteins that bind to the DNA and thereby deform it, change its apparent persistence length, which results in a change in the Root Mean Square displacement (RMS) of the bead.

Tethered particle motion (TPM) is a simple, label-free single molecule technique that can be used to characterize DNA-binding proteins in a high-throughput manner. The technique relies on the tracking of a bead in solution that has been tethered to a glass surface by a DNA molecule or other macromolecule (**FIGURE 3.1**) (265). The bead exhibits Brownian motion, which is restricted by the tether. In contrast to some of the above-mentioned techniques, no additional external force is applied to the bead. The xy-positions of the bead are recorded over time and are used to calculate the root mean square displacement (RMS) of the bead (**EQUATION 3.1**). TPM has a high throughput (*see* **NOTE 1**), which makes it a powerful tool to study the characteristics of DNA-binding proteins. DNA-binding proteins that deform the DNA upon binding, change the apparent persistence length (L_p) of the DNA, a measure for rigidity, when bound, thereby altering the RMS (266). Although the L_p cannot be directly calculated with TPM, the change in RMS is a reliable measure for protein binding. TPM has been used to measure DNA binding by architectural proteins, proteins involved in DNA replication, transcription, DNA repair and recombination (123, 127, 267-274).

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[(x_i - \bar{x})^2 + (y_i - \bar{y})^2]} \qquad (3.1)$$

Equation 3.1. Calculation of the Root Mean Square displacement. x and y are the coordinates of the two-dimensional position of the bead. The summation runs over all measurements taken in the 45 second time window.

In this chapter we describe the use of TPM to measure the K_D of sequence-specific DNA-binding proteins. As an example we characterized the bacterial Integration Host Factor (IHF) (275). IHF is a nucleoid-associated protein (NAP) that is involved in shaping genome architecture in bacteria. IHF also plays a role in up- and downregulation of genes (276). This function is attributed to the ability of IHF to bend DNA and bring regulatory elements together, as well as due to direct interactions with RNA polymerase (277, 278). IHF is known to bend DNA when bound at its high affinity recognition sequence (279) and also bends DNA when bound to DNA non-specifically (280). Bound at its specific site, binding angles of 120–160° have been reported (281-283).

Materials & Methods

All materials are listed in the SUPPLEMENTARY INFORMATION section.

Parafilm semi-circles

1. Cut a strip of parafilm at a length fitting the grid of the laser cutter.
2. Remove the protection layer from the parafilm and place the parafilm onto the grid of the laser cutter.
3. Load the template of the semi-circles in the laser cutter software of choice. Design has to be such that two parafilm semicircles on a 28 mm coverglass create a 6 mm wide channel. Position the parafilm so that the template virtually overlays the parafilm.
4. Fix the parafilm on the grid with office tape.
5. Cut the parafilm (*see* NOTE 9).
6. Remove parafilm strip. The semi-circles will stick to the grid of the laser cutter. Take the semi-circles off the grid individually with tweezers.
7. Store the parafilm semi-circles at 4°C to prevent sticking to each other.

Sample chambers

The following steps should be carried out with gloves, both to protect the hands from exposure to chemicals and to keep the sample chambers free of contaminants.

1. Place the cover glasses in the slide holder using the tweezers with pointed tips (*see* NOTE 10).
2. Place the holder in a beaker and submerge in acetone. Place the beaker in a sonication bath and sonicate for one cleaning cycle (*see* NOTE 11).
3. Transfer the slide holder into a new beaker and submerge the slide holder in ethanol. Place the beaker in a sonication bath and sonicate for one cleaning cycle.
4. Remove the holder from the beaker. Place the cover glasses on a precision wipe and dry the cover glasses by gently wiping them with precision wipes. Turn the cover glasses over and leave exposed to air for 10 minutes to dry.
5. Using tweezers place two parafilm semi-circles in parallel on opposing sides of a 28 mm cover glass, so that a straight channel is created (FIGURE 3.2). The parafilm should align with the glass on the round edges. Push the parafilm firmly but carefully onto the cover glass with the tweezers so that the parafilm adheres to the cover glass (*see* NOTE

12).

6. Place the 35 mm cover glass on top of the parafilm-covered side of the 28 mm cover glass using tweezers, carefully centered onto the 28 mm cover glass. Push firmly but carefully with the tweezers so that the slides stick together.
7. Place the sample chamber on the hot plate (pre-heated to 100°C) using tweezers. Cover the sample chamber with the lid and press firmly and evenly on the lid for 10 seconds. Remove the lid from the sample chamber (*see* NOTE 13).
8. Lift the sample chamber off the hot plate using tweezers and let it cool down on a precision wipe. Store the sample chamber in a small plastic storage box at room temperature.

Bead suspension

1. Dilute the stock of streptavidin-coated beads to 0.01% w/v in 5 ml BB (*see* NOTE 16).
2. Vortex thoroughly for 5 minutes.
3. Sonicate the bead suspension on ice for 3 minutes (active), 2.5 s on and 9.9 s off, at 25%of the maximum power of the immersion sonicator. The total run time will be around 15 minutes. The tip of the immersion sonicator should be submerged by roughly 3 cm. Never sonicate the bead

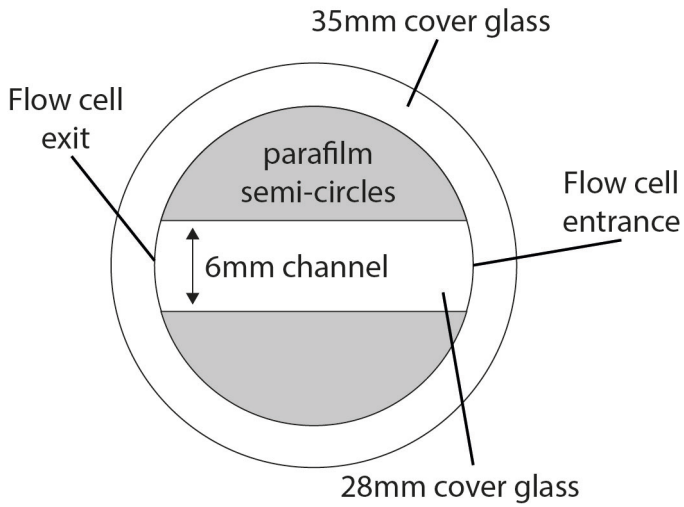


Figure 3.2: Sample chamber for TPM. The sample chamber consists of two parafilm semi-circles sandwiched between two cover glasses. The sample chamber is agglutinated by heating to 100°C.

suspension for more than 5 minutes of active time, as this may damage the beads and its coating.

4. Flow 30 µl of bead suspension into an empty sample chamber. If less than 1% of the beads is clustered, it is considered a good bead suspension.
5. Store the bead suspension at 4°C.

DNA substrate

1. Add the following components to a PCR reaction tube on ice:

Component	*Quantity*
DNA template in 10 mM Tris pH 7.5	10 ng
Biotinylated forward primer	25 pmol
Digoxygenin-labeled reverse primer	25 pmol
Phusion master mix	25 µl
DMSO	1.5 µl
Milli-Q	To total volume of 50 µl

2. Place the PCR reaction tube in the thermal cycler and use the following touchdown protocol (*see* **NOTE 14**):

Temperature	*Duration*	*Cycles*
98°C	30 s	
98°C	10 s	
72°C (-1°C per cycle)	20 s	15x
72°C	60 s	
98°C	10 s	
57°C	20 s	25x
72°C	60 s	
72°C	600 s	
12°C	∞	

3. Purify the PCR product using a PCR clean-up kit (*see* **NOTE 15**).
4. Run 2 µl of PCR product on a 1% agarose gel, pre-stained with GelRed, alongside a suitable DNA marker. Verify the length of the DNA substrate and estimate the concentration using a spectrophotometer. Store the DNA substrate at -20°C at a concentration of 100-500 pM.

Anti-digoxygenin solution

1. Dissolve 200 µg of anti-digoxygenin in 1 ml of BB.
2. Aliquot per 20 µl and store the anti-digoxygenin at -20°C.
3. When using the anti-digoxygenin solution, thaw an aliquot and dilute 10 times with BB.

DNA-tethered bead in sample chamber

1. Take a clean sample chamber and flow in 20 µg/ml anti-dig solution by holding the tip of a pipette against the opening on the right side of the sample chamber at a 45° angle and carefully pressing the plunger button (*see* NOTE 17). Incubate 10 minutes at room temperature.
2. Wash the sample chamber with 100 µl passivation solution and incubate for 10 minutes at room temperature. The excess liquid flowing out of the sample chamber is collected using Whatman paper. This promotes an undisturbed and uninterrupted flow in the channel.
3. Wash the sample chamber with 100 µl BB (*see* NOTE 18).
4. Flow in 100 µl 100 pM DNA substrate (*see* NOTE 19). Incubate for 10 minutes at room temperature.
5. Wash the sample chamber with 100 µl BB.
6. Pipette the bead suspension stock up and down and flow in 100 µl bead suspension. Incubate for 10 minutes at room temperature.
7. Wash the sample chamber with 100 µl EB.
8. Flow in 100 µl protein solution, diluted to the desired concentration in EB. Incubate for 10 minutes at room temperature.
9. Wash the sample chamber with 100 µl protein solution of the same concentration as in step 8.
10. Seal the sample chamber on both sides by applying nail polish to the opening of the sample chamber. When the nail polish is dry, the sample chamber is ready to be used.

Dilution series

1. Calculate a dilution series with a range from 0 to 1000 nM using twofold dilution steps (*see* NOTE 20). Determine the concentration range in which DNA binding is to be expected. Literature values can be used if available; if not, preliminary experiments may be required.
2. Dilute a protein stock aliquot to a concentration that can be used to make further protein dilutions with EB (*see* NOTE 21). Store this stock

on ice (*see* NOTE 22).

3. Using the working stock, make a dilution series by diluting part of the working stock with EB to the desired concentration. Dilute part of this sample with EB in order to create the next sample (*see* NOTE 21). Store samples on ice (*see* NOTE 22).
4. Measure at least two, but preferably three dilution series (*see* NOTE 23).

Imaging

1. An hour before imaging, turn on the lamp and the heating stage.
2. Place immersion oil on the objective and start imaging software.
3. Place a sample chamber in the sample chamber holder.
4. Raise the objective until the beads are in focus (*see* NOTE 24).
5. Close the isolation chamber. The temperature of the chamber and the objective will now rise to a set temperature (*see* NOTE 25).
6. Re-focus on the beads and ensure that no focal drift is observed (*see* NOTE 26).
7. Select a field of view by moving the stage of the microscope. Select the beads and track them for at least 1500 frames. Images are taken at 25 Hz, with an exposure time of 20 ms. At least 100 'good' bead-tether combinations are needed for quantification after quality check (explained in 'Data analysis'; *see* NOTE 27).
8. Use the tracking software to determine x- and y-positions of the beads (*see* NOTE 8).

Data analysis

1. Calculate the anisotropic ratio (EQUATION 3.2), the root mean square displacement (RMS) and the standard deviation of the RMS for all measurements at a given concentration combined (*see* NOTE 28). Discard tethers with an anisotropic ratio of above 1.3. Discard measurements with a standard deviation of the RMS larger than 6% of the RMS. We thereby exclude beads sticking to the surface during the measurement (*see* NOTE 29).
2. Plot all RMS values that meet the requirements as explained in step 1 in a histogram.
3. Fit a Gaussian curve to the histogram to determine the average RMS and standard deviation for each condition.
4. Calculate the standard error of the mean for each condition.

$$\alpha = \frac{l_{major}}{l_{minor}} \tag{3.2}$$

Equation 3.2: Calculation of the anisotropic ratio (α). l_{major} and l_{minor} represent the length of the major and minor axis of the xy-scatter plot, respectively.

Analysis of specific protein-DNA binding

1. Determine the RMS of the DNA-tethered bead in the absence of protein. This is the position of the peak in the Gaussian fit at 0 nM (**FIGURE 3.3A**).
2. Determine the RMS at saturation levels of specific binding. This is the concentration at which the peak seen at 0 nM has completely disappeared and a peak at a different RMS has appeared (**FIGURE 3.3F**).
3. Concentrations at which two peaks are observed, represent a situation in which a subpopulation of the DNA is bound by protein (**FIGURE 3.3B-E**).
4. For the concentrations at which subpopulations of bound and unbound states can be observed, calculate the occupancy (**EQUATION 3.3**). Occupancy is the ratio between bound and unbound DNA substrate.
5. Find the K_D and Hill coefficient (n) by solving the Hill binding model (**EQUATION 3.4**) for K_D and n (*see* **NOTE 30**).
6. Fit the occupancies using the Hill binding model (**FIGURE 3.4**; *see* **NOTE 31**). Use graphing and data analysis software to display the occupancies per concentration and the Hill fit in a graph.

$$\theta = \frac{h_{peak,\,bound}}{h_{peak,\,bound} + h_{peak,\,bound}} \tag{3.3}$$

Equation 3.3: Calculation of occupancy. The occupancy is the fraction of substrate bound by protein and can be calculated from the area of the peaks of the fitted Gaussian distributions.

$$\theta = \frac{1}{\left(\frac{K_D}{[L]}\right)^n + 1} \tag{3.4}$$

Equation 3.4: Hill binding model. Occupancy θ, binding constant K_D, ligand concentration [L], Hill coefficient n.

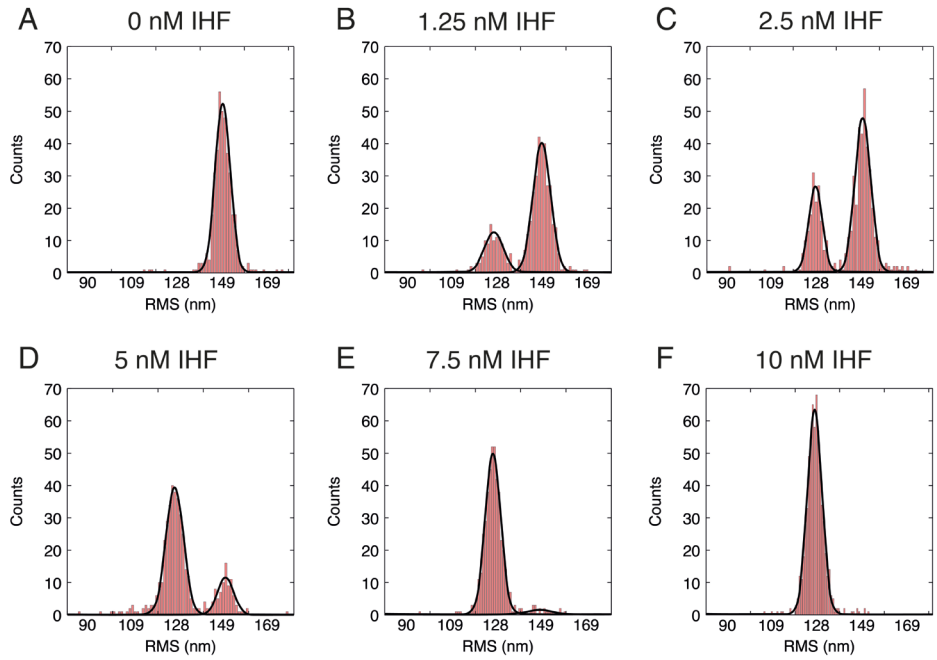


Figure 3.3: Specific binding of IHF to DNA at 0-10 nM. A-F: histograms of RMS values obtained for DNA-tethered beads at IHF concentrations of 0, 1.25, 2.5, 5, 7.5 and 10 nM. The histograms have been fitted with a Gaussian function in KaleidaGraph (black, semi-opaque line).

Example of KD estimation for DNA-binding protein IHF

We used IHF as an example to illustrate how to characterize specific DNA-binding proteins. In our example, we observe a reduction in RMS at low (1.25 nM) IHF concentrations from 149 nm to 127 nm (FIGURE 3.3). This reduction in RMS is caused by DNA bending as a consequence of IHF binding at its specific binding site (*see* NOTE 32). Saturation of specific binding is achieved at 10 nM, the concentration at which only the population with reduced RMS is observed. Fitting the data using the Hill binding model (χ^2: 4.6) yields a binding affinity (K_D) of 3.0 ± 0.3 nM, which is in agreement with values reported using EMSA on different variants of the binding consensus sequence (284). A Hill coefficient (n) of 2.5 ± 0.5 was found. A Hill coefficient >1 indicates positively cooperative binding; n<1 indicates negatively cooperative binding (*see* NOTE 33).

Data representation

TPM is a single molecule method with a high throughput. As a consequence data are often reported in terms of an average RMS value. However, displaying the full population of RMS values can be insightful to illustrate distribution shape and

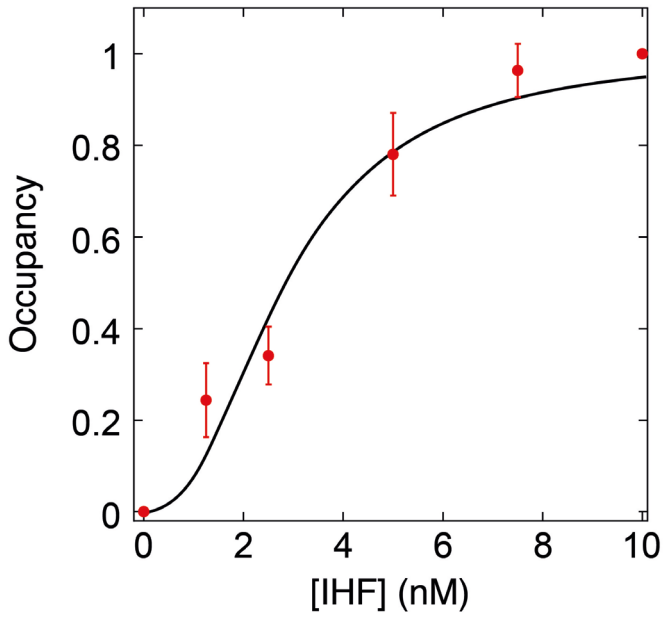


Figure 3.4: Binding curve for specific binding of IHF to DNA. The data points were fit using the Hill binding model. The Hill coefficient (n) is 2.5 ± 0.5, and the binding affinity (K_D) is 3.0 ± 0.3 nM. Error bars represent the standard error of the mean.

population heterogeneity. The new representation approach presented here accommodates both average values, as well as unprocessed population information.

1. Plot all data points as black squares using the graphing and analysis software.
2. In the same plot, include the values of the average RMS at all concentrations and display in a color different from that of the data points.
3. Export the plot as a vector image.
4. Open the plot in Adobe Illustrator and select all black squares. Change the opacity from 100% to 5% (FIGURE 3.5; *see* NOTE 34).

In our example, specific and non-specific binding by IHF can be observed. Specific binding of IHF is observed between 0 and 10 nM, where the RMS decreases from 149 ± 5 nm to 127 ± 5 nm and is characterized by the two populations. Intensity profiles of the populations are equivalent to the histograms of FIGURE 3.3. In the non-specific binding regime, at IHF concentrations >30 nM, we find a single population of which the RMS continues to reduce as a function of IHF concentration up to 400 nM. This further reduction in RMS is reduced to sequence unspecific binding of the protein. At concentrations >400 nM the RMS increases,

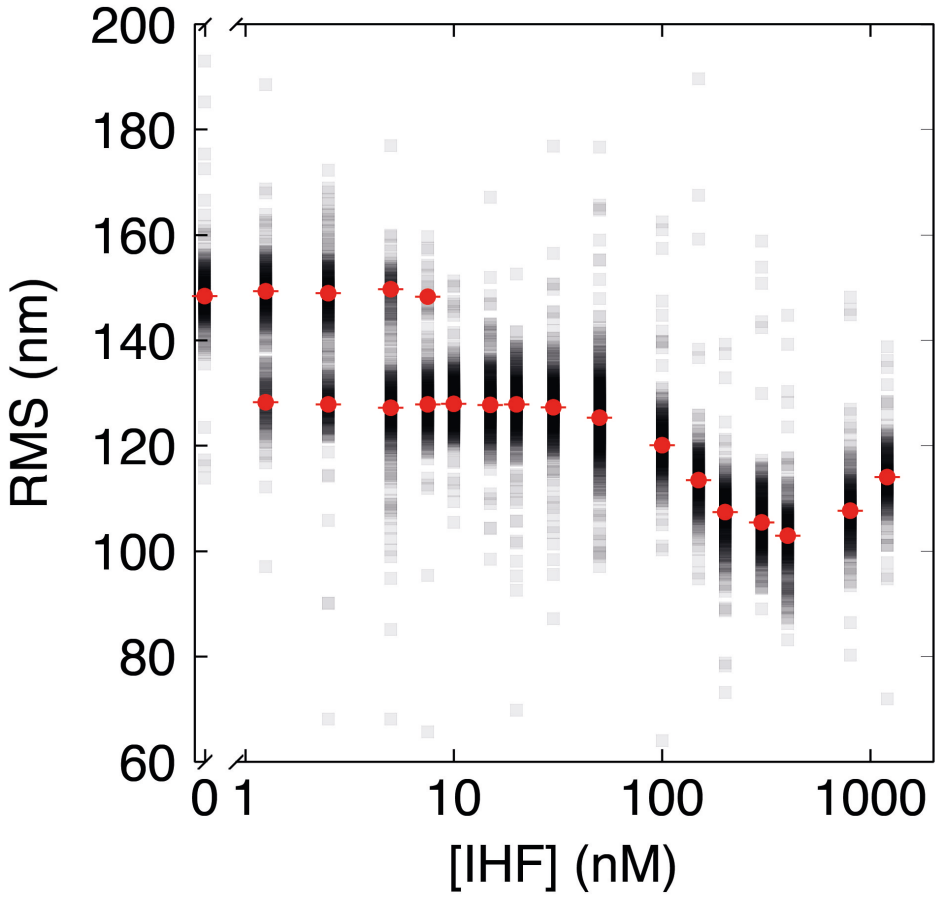


Figure 3.5: Compaction of DNA by IHF. Black, partially transparent squares represent the individual RMS values of DNA-bead complexes after application of anisotropic ratio and standard deviation of the RMS as selection criteria. Red dots and bars represent the average RMS that resulted from fitting with a Gaussian function.

which is attributed to stiffening of DNA by IHF bound at high density along the DNA, similar to the effects of observed for the HU protein (285).

Notes

1) Up to 100 beads were measured for 1 minute.

2) A custom made hot plate was used to heat the sample chambers. The hot plate is made from a circular massive metal block in which a shallow cavity is carved (diameter 35 mm). A sample chamber fits into the cavity. A metal lid (diameter 35 mm) with a plastic handle covers the sample chamber. The hot plate is placed on a heating block and heats the sample chamber to 100°C.

3) Nitrile gloves were used, as the powder that is often found in rubber gloves may contaminate the sample chamber.

4) As a template the plasmid pGp1 was used (286), which contains part of the bacteriophage mu genome. A 685bp fragment that contains 1 IHF binding site exactly at the center was amplified, which is expected to result in the highest effect on end-to-end distance (283). Our experience is that a DNA substrate in the order of 700 bp is ideal, as protein binding results in a clearly detectable change in RMS.

5) The forward and reverse primers that we used are 5' BIO-TGATT AAGGTTGTGGTTAATTTGTTTATCAGTTCCG 3' and 5' DIG-ATTACTTCCGGTTTAGTTTCTAAGGCG 3'.

6) The experimental buffer containined 10 mM Tris-HCl pH 8.0 and 100 mM KCl, which is suitable for specific binding of IHF (287). The osmolarity of the buffer should be sufficiently high to prevent non-specific binding of specific binding proteins. A different protein of interest may require buffer optimization.

7) Recombinant IHF at a concentration of 120 μM was used here. This stock was diluted 100-fold at minimum. The number of freeze-thaw cycles was minimized by aliquoting the protein stock, to guarantee protein activity.

8) Software written for Acoustic Force Spectrometry was used (288), which is freely available online*. Other particle tracing software has been described and compared by Chenouard *et al.* (289). Our lab also

* http://figshare.com/articles/AFS_software/1195874

has positive experience with PolyParticleTracker (290).

9) Here, the computer-aided design software AutoCAD was used to design the cutting template in combination with the VersaLaser tabletop Laser cutter. We cut the parafilm at 55% laser power, 30% of the maximum speed, 1000 pulse and 1.00 mm Z-axis. Different laser cutters may require different settings, allowing the parafilm to be cut straight and without melting or burning.

10) Ensure that each slot is occupied by only one cover glass. The cover glasses tend to stick together, which prevents them from being cleaned properly.

11) For sonication of the glass slides, the Emag EMMI 4 sonicator was used, which is pre-programmed in cycles of 9 minutes at 42.5 kHz.

12) Note that the parafilm leaves grease marks on the cover glass, so avoid sliding after placing them.

13) Heating the sample chamber for too long may result in excessive melting of the parafilm, which causes a non-straight sample chamber. This prevents a good flow in the sample chamber. As a result, coating of the surface with anti-digoxygenin and/or passivation agent may be heterogenous, which reduces throughput and results in lower effective protein concentrations. Especially experiments at low concentration may then suffer from reduced reproducibility.

14) Conventional protocols with fixed annealing temperatures may be used as well, but need to be individually optimized for different primer-template combinations and synthesized fragment lengths.

15) Purification using the GenElute PCR clean-up kit separates the PCR product from other components that were added to the reaction tube to facilitate the PCR reaction, such as excess primers, nucleotides and salts. The result of this purification is the labeled DNA-substrate in 10 mM Tris-HCl pH 8.5 (provided with the kit).

16) The bead stock is vortexed and sonicated to homogenize the suspension. Beads tend to stick together, which will distort measurements.

17) Check if the sample chamber has no cracks in either of the cover glasses, as defects may create a flow inside the sample chamber while measuring. This will affect the RMS by giving it a directional bias. Also, check if the channel is straight; *see* **FIGURE 3.3**. Left-handed people may prefer using the opening on the left side of the sample chamber for sample loading.

18) The volumes used for washing depend on the volume of the sample chamber. As a rule of thumb, we wash with approximately three times the volume of the sample chamber.

19) The DNA substrate concentration is 100 pM, but when we find that this concentration is not sufficient for a particular experiment, we raise the concentration up to 500 pM.

20) A dilution series with concentrations from 0 to 1000 nM is used, with twofold dilution steps. Based on results and literature values, an additional range and/or different dilution steps may be used, which may result in a more reliable calculation of the K_D.

21) When measuring many samples, consider breaking up the dilution series in two or more overlapping ranges. Here, we diluted a 120 μM IHF stock to a working stock of 1.2 μM. From that, we made a serial dilution which included IHF concentrations of 800, 400, 300, 200, 150, 100 and 50 nM: the high range concentrations. For the low range concentrations, we diluted the 1.2 μM working stock directly to 50 nM, from which we subsequently made a serial dilution which included IHF concentrations of 30, 20, 15, 10, 7.5, 5, 2.5 and 1.25 nM. We verified that the RMS values at 50 nM from the high range dilution series were identical to the RMS values at 50 nM from the low range dilution series. We also measured the RMS in the absence of IHF, indicated with 0 nM.

22) Do not freeze the protein solutions, since the activity of the protein may be affected.

23) For reliable results and to minimize the effect of pipetting errors, experiments should be carried out at least in duplicate, but preferably in triplicate, and independently from each other. Especially when the concentration range over which binding takes place is narrow, small changes in sample concentrations can be of significant influence on the

outcome of the binding curve.

24) Due to a sedimentation effect, there may be a slight difference in RMS between beads tethered to the bottom surface and beads tethered to the top surface. Therefore, always choose the same surface for measurements. Also, the focus for top and bottom surface may differ.

25) It may take 5 to 20 minutes, depending on the temperature, before the isolation chamber has completely reached the desired temperature.

26) While the temperature of the immersion oil on the objective rises, the refractive index of the oil changes slightly. As a result, focal drift is observed. Wait until the immersion oil has reached the desired temperature and then refocus. When no focal drift is observed, the immersion oil has the desired temperature.

27) At least 100 good bead-tether combinations are needed, but the actual amount of measured bead-tether combinations to reach this limit may differ. For IHF, 300 measured bead-tether combinations always resulted in more than 100 good combinations. The amount of bead-tether combinations that need to be measured, depends on the protein that is used.

28) Verify that individual experiments done at the same concentration give similar outcomes. If so, RMS values can be combined and can be fitted together. If one measurement is a clear outlier due to a known cause, discard the measurement. In other cases, re-measure the sample and check if the results align with any previous measurements.

29) Discard any bead with an anisotropic ratio >1.3, which is considered non-symmetrical. Non-symmetrical beads indicate for example that a bead is attached to two DNA tethers. An extensive study that links motion patterns to setup defects has been done by Visser *et al.* (291) Also discard any bead with a standard deviation of the RMS of >6%, since this indicates rapid protein dissociation and sticking and/or releasing of the bead to or from the sample chamber surface.

30) The Hill equation can be solved using IGOR Pro, which was done here, but MatLab and other data analysis programs are also suitable for this.

31) As error bars, the standard error of the mean of the individual measure-

ment series per concentration has been used.

32) If desired, the effective persistence length of the DNA can be determined for a single protein bound to DNA, which could be used to calculate the DNA bending angle as described by Kulić and coworkers (266). However, this requires an extensive simulation series, which reveals the relation between RMS and L_p. The simulations are highly dependent on the length and sequence of the DNA substrate, buffer conditions and temperature of the experiment. This means that for every experiment in which one or more of these parameters is changed, a new simulation series is required. The simulations have been described by Driessen *et al.* (292).

33) Here, the Hill coefficient is 2.5, which indicates positively cooperative binding. However, in this example, only one IHF dimer binds to DNA. The apparent positively cooperative binding is possibly a result of the dimeric nature of the protein. Both components of the dimer have to bind to the DNA in order to bend the DNA.

34) For representation of all data points, squares are most suitable, because squares are easy distinguishable when data points overlap. Circles may make it harder to distinguish individual data points in the plot. The opacity that suits your data best depends on the number of data points. When using more data points than shown in the example, it may be useful to use a lower opacity. The Gaussian distribution of the data points should be visible from the intensity of the combination of data points.

Supplementary information

Parafilm semi-circles

1. Parafilm
2. Office tape
3. Laser cutter
4. Computer-aided design (CAD) software (for example AutoCAD or CorelDRAW)
5. Tweezers

Sample chambers

1. Round cover glass, diameter 35 mm (VWR)
2. Round cover glass, diameter 28 mm (Thermo Scientific)
3. Parafilm semi-circles (*see* above)
4. Cover slip rack (Diversified Biotech)
5. Ethanol
6. Acetone
7. Beaker
8. Sonication bath
9. Heating block, capable of heating up to 100°C, with hot plate for 35 mm cover glasses (*see* **NOTE 2**)
10. Tweezers with pointed tips
11. KIMTECH SCIENCE* precision wipes tissue wipers (Kimberly-Clark)
12. Nitrile gloves (*see* **NOTE 3**)
13. Whatman paper

Bead suspension

1. Streptavidin-coated polystyrene beads, diameter 0.44 µm (Kisker Biotech)
2. Immersion sonicator with fine tip
3. Bead buffer (BB) (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA pH 8.0, 1 mM DTT, 3% glycerol, 100 µg/mL BSA-acetylated [Ambion])
4. Vortex

Anti-digoxigenin solution

1. Anti-digoxigenin (anti-Dig) polyclonal antibodies 200 μg (Roche)
2. Bead buffer (BB) (*see* above)

DNA substrate

1. Template for DNA substrate containing the binding site of interest. PCR amplification with compatible primers should yield a substrate of desired length and sequence (*see* NOTE 4).
2. Biotinylated forward primer, 10 μM in 10 mM Tris, pH 7.5, designed to be compatible with the template (*see* NOTE 5)
3. Digoxygenin-labeled reverse primer, 10 μM in 10 mM Tris, pH 7.5, compatible with the template (*see* NOTE 5)
4. Phusion master mix (New England Biolabs)
5. DMSO
6. Milli-Q
7. GenElute PCR clean-up kit (Sigma Aldrich)
8. PCR reaction tubes
9. Biorad C1000 Thermal Cycler

Substrate quality control

1. Agarose gel, 1% in Tris/Borate/EDTA (TBE) buffer
2. GelRed nucleic acid gel stain (Biotium)
3. GeneRuler DNA marker
4. Spectrophotometer to measure DNA concentrations (Nanodrop, Qubit, SimpliNano or comparable device)

Sample chamber reagents

1. Bead buffer (BB) (*see* above)
2. Anti-digoxigenin solution
3. Passivation solution (4 mg/ml Blotting Grade Blocker [Bio-Rad] dissolved in SB)
4. Experimental buffer (EB) suitable for protein binding (*see* NOTE 6)
5. Protein of interest, preferrably at high concentrations to avoid as much as possible components of the protein storage buffer on the EB (*see* NOTE 7)

Imaging

1. Isolation table
2. Inverted microscope (Nikon Diaphot 300) (**FIGURE S3.1**)
3. 100x oil-immersion objective (NA 1.25)
4. Immersion oil for microscopy (Merck)
5. Heat chamber (Bioscience tools TC-HLS-025 heating element for heating of the objective and a custom sample chamber heating element consisting of 40 parallel resistors); Pt100 sensors attached to the sample chamber holder and the objective, and two SA200 PID digital temperature controllers)
6. Climatized room
7. Isolation box
8. Thorlabs CMOS camera DCC1545M
9. Lens paper
10. Nacre-free nail polish
11. Particle-tracking software (software available online (288); *see* **NOTE 8**)

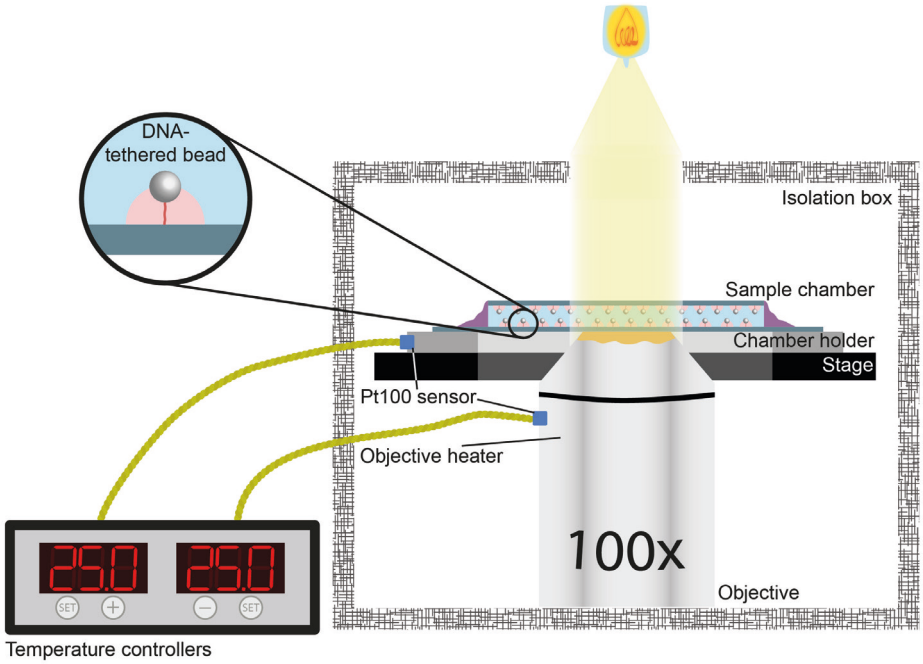


Figure S3.1: Schematic of the Tethered Particle Motion setup. The sample chamber is heated by two heating elements, one placed in the sample chamber holder and one on the objective. An isolation box covers the stage, sample chamber and heating elements. Schematic is not to scale.

Data analysis and representation

1. Computer
2. Custom MatLab routine for calculating RMS, standard deviation of the RMS and anisotropic ratio. The routine also selects the beads that meet the requirements for reliable beads (*see* 'Data analysis' in the **Methods** section)
3. Graphing and data analysis software, such as KaleidaGraph, Origin or IGOR Pro
4. Adobe Illustrator

CHAPTER 4

Mechanical and structural properties of archaeal hypernucleosomes

This chapter is based on:

Bram Henneman*, Thomas Brouwer*, Clara van Emmerik, Ramon A. van der Valk, Mandy Erkelens, Monika Timmer, Nancy Kirolos, Hugo van Ingen, John van Noort & Remus T. Dame. *Mechanical and structural properties of archaeal hypernucleosomes*. Submitted.

* These authors contributed equally

Abstract

Many archaeal species express homologues of eukaryotic histones H3, H4, H2A and H2B, which organize the genome and play a key role in gene regulation. The structure and function of archaeal histone-DNA complexes remain largely unclear. Recent X-ray crystallography studies of *Methanothermus fervidus* histone HMfB bound to DNA fragments shows formation of hypernucleosomes consisting of DNA wrapped around an 'endless' histone-protein core. Such a structure is in line with observations suggesting that histone-DNA complexes consisting of an integer number of dimeric units of archaeal histones are formed *in vivo*. However, whether such a hypernucleosome structure assembles on a long DNA substrate and which interactions provide for its stability, remained unclear.

Here we describe micromanipulation studies of complexes of the *M. fervidus* histones HMfA and HMfB in solution. Our experiments show that hypernucleosome assembly results from cooperative binding of histones to DNA, facilitated by stacking interactions between neighboring histone dimers. Furthermore, rotational force spectroscopy demonstrates that the HMfB-DNA complex is left-handed, but that torque can drive it in a right-handed conformation. The structure of the hypernucleosome is thus dependent on stacking interactions and force. Electrostatic interactions facilitate ionic modulation of structural plasticity,

and torsional plasticity allows for absorption of supercoiling. Therefore, stacking interactions, differential behavior of different histones, and force-dependent handedness imply an important role for archaeal histones as transcription regulators, responsive to environmental changes.

Significance statement

Histones play an important role in genome compaction and gene regulation of Eukaryotes. Archaea also express histones, but they lack a key feature of the eukaryotic histone: the N-terminal tail, important in eukaryotic transcription regulation. Archaea are highly relevant as they are one of the main pillars of life, and the role of archaea as a link between the last universal common ancestor in life on earth (LUCA) and Eukaryotes becomes more unequivocal. With no N-terminal tail, archaea must have an alternative way of modulating DNA accessibility. We investigated how archaeal histones compact DNA. Its mechanisms of compaction are suggestive of primordial bacterial-like transcription regulatory mechanisms preceding eukaryotic regulatory mechanisms.

Introduction

Dynamic genome organization is a prerequisite for a compact yet active genome throughout all domains of life. Eukaryotes use histones, which wrap DNA into nucleosomes to compact and functionally organize their genomes. These histones have N-terminal tails that can be post-translationally modified, changing their physico-chemical properties, which is key in defining the functional chromatin state (293). Current views on evolution agree that eukaryotes are part of the branch of archaea, which are single-cellular organisms that share many cellular mechanisms with Eukaryotes (3, 10, 47, 294). Most archaeal species express rudimentary homologues of eukaryotic histones with the characteristic histone fold (13, 177, 295), but lack the N-terminal tail (140, 186, 296-298). In addition to histones, archaea express other small architectural proteins, nucleoid-associated proteins (NAPs). These NAPs are also involved in genome organization and may complement and compete with histones to regulate genes (299, 300).

The prototypical archaeal histones are HMfA and HMfB from *Methanothermus fervidus* (193). HMfA and HMfB are homologues of eukaryotic histones in terms of sequence and structure (140). HMfA and HMfB share 84% sequence identity. These proteins are expressed at varying ratios as a function of growth phase (251). Although estimates of absolute expression levels of individual histones are lacking, HMfA is prevalent during exponential growth (1.5x the amount of HMfB), and present in equal amounts as HMfB during stationary growth (251). This might imply that these proteins have evolved different DNA-binding properties – and correspondingly distinct functions – in the cell. HMfA and HMfB are capable of forming homo- and heterodimers, which subsequently can form larger structures (171, 251, 301, 302). The binding of HMf dimers to DNA results in bends in DNA at low protein concentrations, and has been reported to yield "beads on a string" nucleosome-like structures at higher protein to DNA ratios (193). HMf proteins bind preferentially to intrinsically curved DNA sequences *in vitro* (303). There is currently no evidence of the existence of specific DNA sequence signatures enhancing histone binding *in vivo*, although studies in *Methanothermobacter thermautotrophicus* and *Thermococcus kodakarensis* indicate that histone positioning favors repetitive motifs of AT and GC base pair steps (221). These motifs facilitate the distortion needed for DNA wrapping. Systematic evolution of ligands by exponential enrichment (SELEX) experiments have yielded two high affinity DNA-binding sequences for HMfB *in vitro* (164). Binding of HMf proteins to this sequence yields nucleosome-like structures, in which 60 bp of DNA is proposed to be wrapped around a tetrameric protein core composed of two interacting HMf dimers (160). Here, we will refer to the DNA compaction mode

of HMf proteins as wrapping, even when the DNA is wrapped less than one turn. The histone tetramer model is supported by micrococcal nuclease (MNase) digestion studies of chromatin from *Haloferax volcanii*, which yield undigested DNA fragments 60 bp in size (220, 221). However, the histone tetramer model has been challenged by the results of similar studies in *T. kodakarensis*, which yield DNA fragments ranging from 30 – 450 bp in 30 bp increments (165). The sizes of these DNA fragments were interpreted as reflecting the existence of histone multimers of varying sizes bound along the genome. A notable recent crystallography study of HMfB on 90 base pair DNA fragments supports the formation of a multimeric histone filament established by interactions between adjacent histone dimers and of wrapping of DNA fragments around an endless protein core, creating a

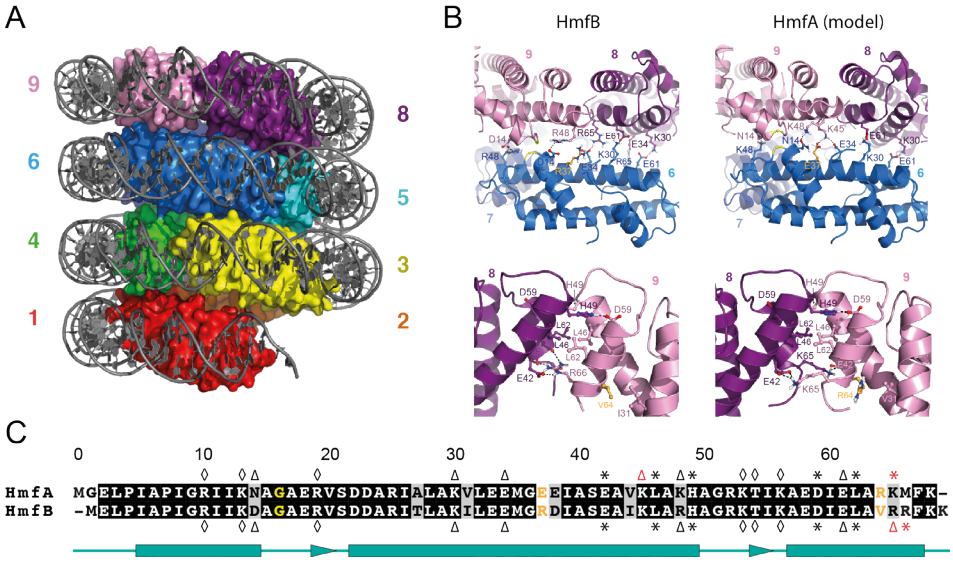


Figure 4.1: The hypernucleosome and *Methanothermus fervidus* histones HMfB and HMfA. A) The HMfB hypernucleosome structure as described by Mattiroli *et al*. (PDB: 5T5K (27)). Figure taken from Henneman *et al*. 2018 with permission (28). B) Hypernucleosome interfaces of HMfB (left; water-refined and starting from the 5T5K structure) and HMfA (right; water-refined and starting from the homology model). Top panels show the stacking interface, bottom panels show the tetramer interface. Interacting residues are labeled; hydrogen bonds are shown by dashed lines. Important residue differences between HMfA and HMfB are colored in orange. The closely packed residues G16 in the stacking interface are shown in yellow spheres (Cα only). C) Sequence alignment of HMfA and HMfB. Conserved residues are shown in black boxes, conservative changes in grey and non-conservative changes in white boxes. Residues are numbered according to HMfB sequence. The two sequences are 84% identical and have 94% similarity. Residues directly contacting the DNA are indicated with diamonds (◊), residues forming the tetramerization interface are marked with asterisks (*) and stacking residues are indicated with triangles (Δ) for HMfA (top) and HMfB (bottom). Differences in interacting residues between HMfA and HMfB are indicated with red symbols. Important residue differences between HMfA and HMfB are colored in orange and G16 residues are shown in yellow, matching panel B.

quasi-continuous superhelix (**FIGURE 4.1A**) (197). This structure is referred to as the hypernucleosome (134). The hypernucleosome is left-handed like the eukaryotic nucleosome, although it has been reported that archaeal histones can accommodate both left- and right-handed wrapping configurations (171, 304, 305). It was proposed that hypernucleosomes are not just held together by interactions between adjacent dimers, but also by stacking interactions between the layers of dimers within the hypernucleosome (**FIGURE 4.1B, C**) (134). The disparate findings from *H. volcanii* and *T. kodakarensis*, as well as the crystal structure, indicate that molecular insights into the role of histones in archaeal chromatin organization are important for understanding archaeal genome compaction.

In this study, we provide direct evidence of hypernucleosome formation in solution and of its capability of transitioning in handedness. The hypernucleosomes formed by HMfA and HMfB homodimers exhibit subtle differences in amino acid composition, resulting in different stacking energies. This underlies different structures of the hypernucleosome, which may be the key to a mechanism for transcription regulation in archaea.

Results

Both HMfA and HMfB cooperatively compact DNA

To determine the degree of DNA compaction by HMfA and HMfB, we performed Tethered Particle Motion experiments (TPM) (195). Here, the Root Mean Squared displacement (RMS) of a bead connected to a surface-attached DNA tether provides a quantitative readout of DNA conformation (FIGURE S4.1A). We investigated the binding of HMfA and HMfB to a 685 bp DNA molecule containing a random, naturally occurring sequence (*see* MATERIALS AND METHODS). We observed an abrupt and drastic reduction in RMS of the DNA tether over a narrow concentration range during titration of both proteins, which indicates that DNA compaction by both HMfA and HMfB is highly cooperative. Such cooperativity is due to either direct protein-protein interactions or facilitated protein-DNA binding through structural effects of adjacently bound HMfA or HMfB proteins (171, 306) (FIGURE 4.2).

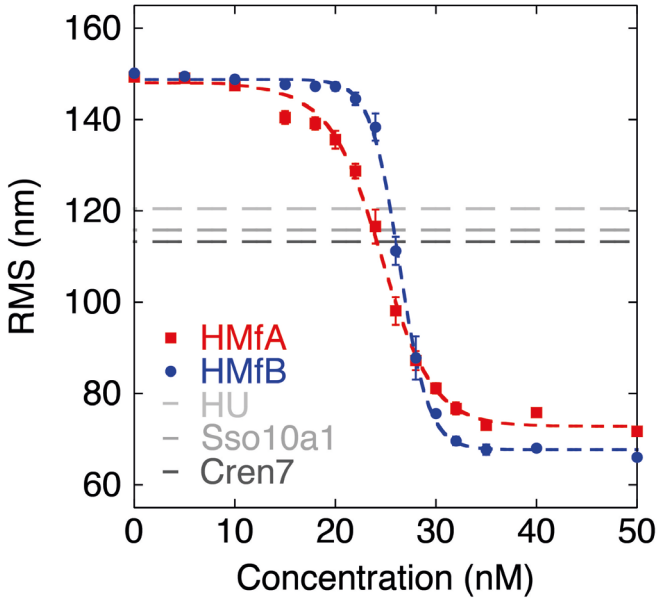


Figure 4.2. Both *Methanothermus fervidus* histone proteins HMfA and HMfB drastically and cooperatively compact DNA in Tethered Particle Motion experiments. Root mean square displacement (RMS) excursion of the bead for HMfA and HMfB on a DNA substrate is shown as a function of protein concentration. Dashed lines are to guide the eye. Error bars indicating the standard error of the data points (N≥100) are mostly hidden behind the data points. In grey, saturation levels of the bacterial NAP HU and archaeal NAPs Sso10a1 and Cren7 – which are known to organize and compact DNA – are indicated, as reported by Driessen *et al.* 2014, Driessen *et al.* 2016 and Driessen *et al.* 2013 (33-35).

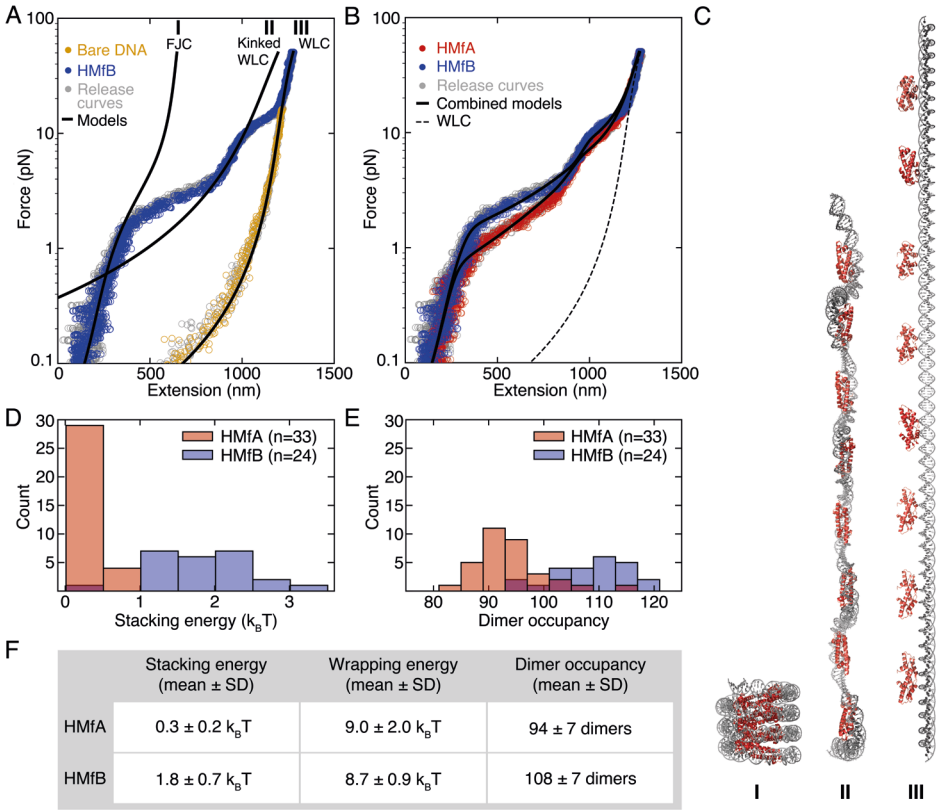
As a dimer, HMf proteins bend DNA by partial wrapping to an angle that is similar to those induced by DNA-bending proteins such as the bacterial NAP HU and the archaeal NAPs Cren7 and Sso10a1 (125, 127, 193, 292, 307, 308). These proteins reduce the RMS on a substrate of equal length from 150 to 110-120 nm at saturating protein concentrations (127, 292). A much stronger reduction in RMS down to ~70 nm was observed for HMfB and ~75 nM for HMfA. The extent of this change cannot be explained by DNA bending induced by individual HMfB dimers. The high degree of DNA compaction together with the pronounced binding cooperativity suggests adjacent packing of HMfB (and HMfA) proteins that together wrap DNA. By determining the maximum deflection of the bead in TPM, we found an end-to-end distance of 24 ± 4.7 nm for the HMfB-DNA complex at saturating protein concentrations for this specific DNA substrate (**figure S4.1**). This number agrees with the theoretical end-to-end distance of the hypernucleosomal structure observed in the crystallography studies (197). Both histones bind to DNA in a cooperative manner, with HMfB showing a more pronounced cooperative binding than HMfA. Differences in the stacking interface or tetramerization interface of HMfA and HMfB may underlie the difference in DNA compaction (**figure 4.1B** and **C**).

HMfB organizes long DNA tethers into endless hypernucleosomes

The mechanical stability of the hypernucleosome was investigated by force spectroscopy. HMfB hypernucleosomes were reconstituted *in vitro* on a torsionally unconstrained DNA construct, at 100 nM HMfB, well above the K_D. Hence, we expect full coverage of the DNA. Force spectroscopy revealed a characteristic unfolding response to low-high-low force cycling (**figure 4.3A** and **B**). Binding of HMfB resulted in a large compaction of the DNA. The absence of hysteresis at the low force regime (<50 pN) indicates that the manipulation of the HMfB-DNA complex occurs in thermodynamic equilibrium: stretch-release cycles could be repeated multiple times without qualitative change to the data. The force extension curve of the HMfB-DNA complex shows three regimes of extension (**figure 4.3A**). In regime I, at forces below 1 pN, the force-extension curve features a small extension and high stiffness. In regime II, between 2 and 10 pN, the tether is longer and appears to be less stiff than in regime I. In regime III, at forces above 20 pN, the extension of the complex follows that of bare DNA, which is stiffer than the complex in regime II. Most of the force extension curve of the HMfB-coated DNA could be captured in these three regimes, and we interpret the more horizontal parts as transitions between these regimes.

We modeled the two most extended structures, in force regime II and III, with

the kinked and regular worm-like chains (WLCs), respectively. In regime III the complex followed the force-extension behavior of bare DNA with a persistence length of 50 nm. For regime II, we fitted a persistence length of approximately 4 nm, suggesting a much more flexible tether. The most compacted structure, regime I, could not be modeled by a WLC and shows a linear extension between 0.1 and 2 pN. We used the linear part of a freely-jointed chain (FJC) to describe



F

	Stacking energy (mean ± SD)	Wrapping energy (mean ± SD)	Dimer occupancy (mean ± SD)
HMfA	0.3 ± 0.2 k_BT	9.0 ± 2.0 k_BT	94 ± 7 dimers
HMfB	1.8 ± 0.7 k_BT	8.7 ± 0.9 k_BT	108 ± 7 dimers

Figure 4.3: Force spectroscopy experiments on the hypernucleosome reveal stronger stacking in HMfB tethers than in HMfA tethers. Red: HMfA; blue: HMfB; yellow: bare DNA. A) Force spectroscopy on a HMfB-DNA complex at 100 nM of HMfB and a bare DNA molecule reveals three levels of compaction. We fitted these to a freely-jointed chain (FJC) (I), kinked worm-like chain (WLC) (II), and WLC (III). The release curves overlapped with the stretch curves, which indicates that HMfB-DNA the stretch-release cycle was in equilibrium. B) Comparison between HMfA hypernucleosomes and HMfB hypernucleosomes. The WLC model for bare DNA was added as a reference (dashed black line). The free energy for dimer stacking and DNA wrapping was extracted by fitting the unfolding model (black lines). C) Cartoons of three states of the HMfB-DNA complex corresponding to the model fit in A, with representative compaction, illustrating the 7-fold compaction of the hypernucleosome compared to bare DNA. D) Histogram of stacking energies of HMfA and HMfB. E) Histogram of occupancy of DNA by HMfA and HMfB dimers. F) Mean stacking energy, wrapping energy and dimer occupancy, with standard deviation (SD).

this regime of the force-extension curve.

The mechanical properties of the complex in these three force regimes suggest three different structures of the tether. Stacking of HMfB dimers into a hypernucleosome would create a relatively stiff structure that can only be stretched to a limited extent without breaking the protein-protein stacking interactions. Based on the crystal structure of the HMfB hypernucleosome (197), we expect a maximal extension of 4 nm per dimer for the most condensed structure (**figure 4.3C, I**). When the stacking interactions are broken due to excessive force, individual dimers may remain bound to the DNA. Such a structure would have a much larger extension per HMfB dimer than the stacked hypernucleosome. If all DNA-protein contacts remain intact, the structure follows a highly curved DNA trajectory (**figure 4.3C, II**). Such curved DNA corresponds to a decrease of the apparent persistence length (285, 309), which is consistent with the observed force-extension relation in regime II. Increasing the force further breaks protein-DNA interactions to yield an unperturbed DNA trajectory (**figure 4.3C, III**), which would have similar mechanical properties as bare DNA. Indeed, at forces larger than 20 pN the force-extension curve overlaps with that of DNA. Thus, the experimental force-extension curves suggest a two-step transition from a hypernucleosome into a fully stretched DNA tether.

For a more quantitative analysis we developed a statistical mechanics model that includes the transitions between the three structures, as described in materials and methods. In analogy with the forced unfolding of eukaryotic chromatin (310), we modeled the entire force-extension curve as a linear combination of the extension of individual dimer-DNA complexes in either of the three states and in addition to a small fraction of bare DNA (*see* **Materials and Methods** for model details). For every dimer that binds to the DNA tether, the extension changes and the free energy is reduced by a binding and/or stacking energy. For each dimer in state II, we imposed a deflection angle of 10 degrees resulting in a reduction of the persistence length as described by Kulić and Schiessel (309). Release curves fully overlap with the stretch curves. The absence of hysteresis in all curves indicates that all transitions are in thermodynamic equilibrium at low forces, allowing an equilibrium model to describe the transitions between the states.

The force-induced transitions between the three states of an HMfB-coated DNA tether are captured by the model (**figure 4.3B**). The first transition (I to II) at 2 pN indicates a low stacking energy of $1.8 \pm 0.7\ k_BT$. The increase in extension is large with ~500 nm, which means that the stacked hypernucleosome is easily disrupted by force (**figure 4.3D**). Unwrapping of the DNA from the dimers at 10

pN takes more energy (8.7 ± 0.9 k_BT), and increases the extension by only ~200 nm. These transitions appear as gradual changes in extension, likely because of the large number of dimers, the small changes in extension and the fast kinetics. We found an average of 108 ± 7 dimers per tether. A footprint of 30 bp per HMfB dimer would allow 121 dimers to bind this DNA molecule. It therefore appears that the DNA molecule is not fully saturated, implicating defects in stacking of the hypernucleosome (**figure 4.3E**). Interestingly, the fitted number of dimers did not change in multiple stretch and release curves of the same hypernucleosome, suggesting that the dimers remain bound to and positioned on the DNA, even at high forces.

TPM experiments showed that HMfB compacts DNA slightly more than HMfA (**figure 4.2**). In force spectroscopy, the force-extension curves for HMfA-DNA complexes were similar to those for HMfB-DNA complexes (**figure 4.3B**), although the transitions between stacked and unstacked states occurred at lower forces, 0.5 rather than 2 pN. Correspondingly, we found a higher stacking energy and dimer occupancy for the HMfB-DNA complex compared to the HMfA-DNA complex, but a similar wrapping energy (**figure 4.3D-F**). The high similarity in stiffness and extension of the complexes at forces below the unstacking transition (I to II) and the similar wrapping energy after unstacking, highlights the universality of the hypernucleosome structure, whereas the difference in stacking energy and occupancy clearly distinguishes the mechanical properties of the protein-DNA filaments formed by the two homologues. This may be biologically relevant as differences in mechanical stability may modulate the balance between gene compaction and accessibility *in vivo*.

Stacking interactions mediate hypernucleosome formation

The basis for stacking of HMfA and HMfB homodimers lies in their primary structure. We analyzed both histones in terms of interaction residues (**figure 4.1C**). Using water refinement with HADDOCK (311) on the HMfB hypernucleosome crystal structure and modeling the HMfA hypernucleosome structure based on that same crystal structure, we were able to identify the possible stacking interactions within the hypernucleosome. For HMfB, we determined interactions between R48-D14, K30-E61 and E34-R65. For HMfA, we found interactions between K31-E62 and E35-K62. In order to investigate whether these interactions indeed mediate hypernucleosome formation, we expressed and purified HMfA stacking mutant $HMfA_{K31A\ E35A}$. We performed TPM experiments similar to those described above. We found that the mutants reduced the RMS to ~78 nm in TPM experiments (**figure 4.4**), which is similar to the values found for $HMfA_{wt}$ on

DNA. However, the concentrations at which this reduction in RMS takes place, is much higher for $HMfA_{K31A\ E35A}$ than for $HMfA_{wt}$. For $HMfA_{wt}$, a K_D of 25 nM is estimated, whereas for $HMfA_{K31A\ E35A}$ the estimated K_D is 750 nM. This indicates that cooperative hypernucleosome formation is strongly dependent on stacking interactions mediated by residues K31 and E35.

Although HMfB has 1 stacking interaction more than HMfA, the stacking interactions of both proteins have a similar number and type of hydrogen bonds. When we calculate the total free energy of all interface residues (stacking interfaces and dimer-dimer interfaces including hydrophobic interactions), we find that HMfA contacts are energetically more stable than HMfB contacts (**FIGURE 4.3F**). Reduced repulsive forces between the HMfA dimers because of the charge swap at position 37 and the lack of positive charge at position 69 also favor a more stable HMfA complex. However, at HMfB position 64, there is a valine that features a hydrophobic interaction with I31, whereas HMfA has an arginine at this position (**FIGURE 4.1C**). The large size of this residue disrupts the hydrophobic interaction. This may make the C-terminal α-helix of HMfA more flexible, leading to a less compact and less stable hypernucleosome. Earlier studies support this interpretation, as a V64R mutation in HMfB leads to reduced electrophoretic mobility of the histone-DNA complex and a slightly impaired nucleosome formation (137, 312, 313).

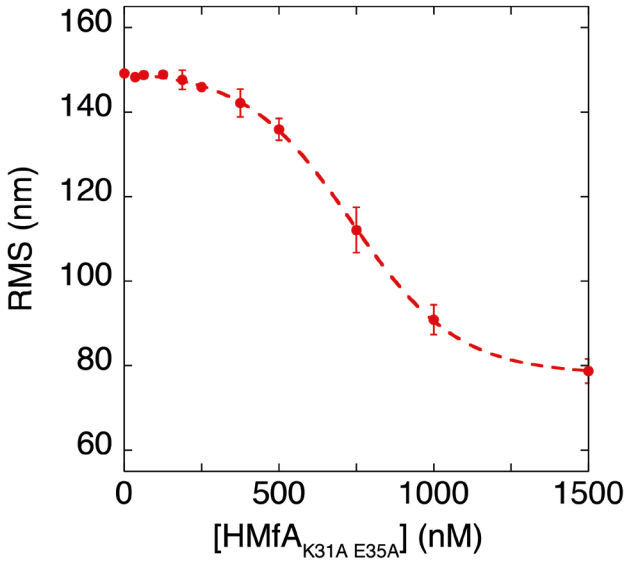


Figure 4.4: Characterization of $HMfA_{K31A\ E35A}$. Root mean square displacement (RMS) excursion of the bead for $HMfA_{K31A\ E35A}$ on a DNA substrate is shown as a function of protein concentration. Dashed lines are to guide the eye. Error bars indicate the standard error of the data points (N≥100). Note that the x-axis is very different compared to **FIGURE 4.2**.

The hypernucleosome forms a left-handed structure on DNA

Wrapping DNA into a hypernucleosome imposes a distinct chirality. To reveal handedness and torsional stiffness of the HMfB hypernucleosome, we used rotational force spectroscopy on a torsionally constrained hypernucleosome. Therefore, the hypernucleosome is unable to release torsional stress by swiveling around the attachment point. The resulting torque would affect the mechanical properties of the hypernucleosome, which we refer to as the HMfB-DNA complex when stretched. We monitored the extension of the hypernucleosome at a constant force, after applying negative (FIGURE 4.5A) and positive (FIGURE 4.5B) twist to a HMfB hypernucleosome.

The unstacking transition of negatively twisted complexes was shifted to higher forces. The force plateau shifted up to 10 pN. Thus, the energy for the unstacking transition is increased: more force is required to unstack the hypernucleosome.

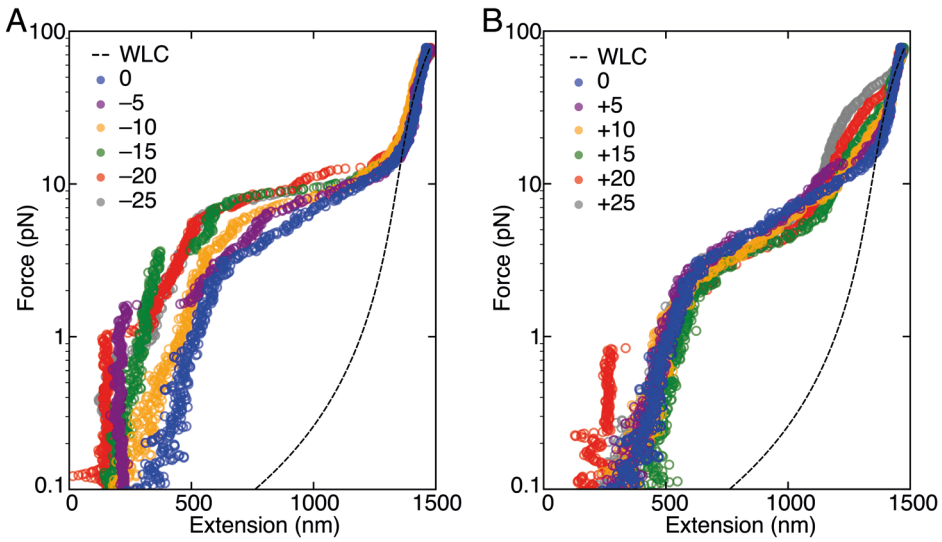


Figure 4.5: Stretching a torsionally constrained HMfB-DNA complex reveals left-handedness of the hypernucleosome, and shows increased affinity of HMfB dimers for positively twisted DNA. A) Negative twist as a result of 5-25 rotations increases the unstacking force of the hypernucleosome. In addition, excess negative twist causes the HMfB-DNA complex to buckle, resulting in a dramatic decrease of its extension. Both observations point to the left-handedness of the hypernucleosome. The unwrapping transition appears largely unaffected, resulting in a continuous decondensation of the tether around 10 pN. B) Positive twist as a result of 5-25 rotations recovers the unstacking plateau. Strikingly, positive twist increases the rupture force of the unwrapping transition of HMfB dimers. An increasing force is necessary to reverse the bending effect of HMfB dimers on DNA. At very low forces the beads appear to stick to the cover slip for certain curves (-5, -20, and +20 rotations, especially).

Large amounts of negative twist further compact the HMfB-DNA complex by buckling the tether complex at forces below 1 pN. The second transition to unbent DNA appeared unaffected by negative twist.

Positive twist on the other hand somewhat extended the range of the unstacking transition to 2-10 pN. Such behavior is a feature characteristic of a left-handed structure, and in agreement with the HMfB-DNA co-crystal structure (197). More strikingly, the second transition between bent and straight DNA is disfavored by positive twist, but not by negative twist. It appears that unstacked HMfB dimers can better accommodate positive twist than bare DNA. The affinity for HMfB dimers increases with the number of twists.

An unstacked HMfB-DNA complex imposes a right-handed structure on DNA

Unfolding a torsionally constrained HMfB-DNA complex builds up torque due to the untwisting of the hypernucleosome (**figure S4.2**). The unstacking transition (between I and II in **figure 4.3A**) of a torsionally constrained hypernucleosome features an increased and more gradual force plateau compared to a torsionally unconstrained hypernucleosome. As a result, the unstacking transition merges with the transition that we attribute to unwrapping DNA from the HMfB dimers (regime II to III in **figure 4.3A**). Since HMfB dimers are likely coupled on the DNA, the unwrapping of one dimer creates torsional stress, which impedes unwrapping of the next dimer, resulting in an anti-cooperative transition.

By maintaining a fixed force larger than the unstacking force of the hypernucleosome, and monitoring the extension under twist, we were able to determine the handedness of the unstacked HMfB-DNA complex (**figure 4.6**). For low forces (f=2.0 pN) either negative or positive twist induced a reduction in extension due to buckling of the complex superstructure. Note that the presence of HMfB-induced bends favors plectonemes over DNA melting, which would occur in bare DNA for negative twist at this force. The asymmetry in twist response in HMfB-decorated tethers only occurs at force above 3 pN. Alternatively, negative twist may unwrap the HMfB dimers, increasing the length of the tether. Similar to bare DNA, positive twist only supercoils the tether. This implies that the unstacked HMfB-DNA complex forms a right-handed superhelix and that the HMfB dimers are stabilized on the DNA by positive torque. The force required to completely unwrap the DNA was correlated with the applied twist. HMfB dimers appeared to have a higher affinity for positively twisted DNA than for negatively twisted or relaxed DNA. This observation is supported by the release curves of the force spectros-

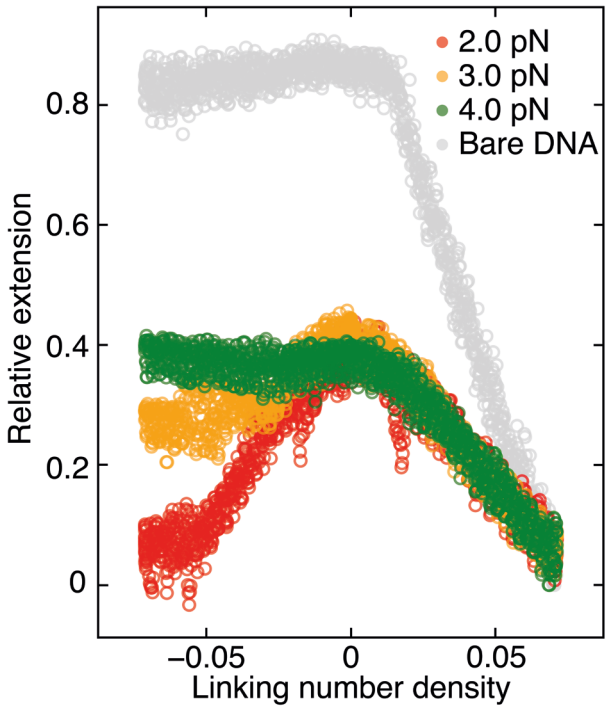


Figure 4.6: Torque stabilizes the hypernucleosome and reveals a right-handed chirality of unstacked HMfB-DNA complexes. A torsionally constrained HMfB hypernucleosome was compacted by twisting the molecule under constant force. At 2.0 pN, both positive and negative twist reduce the extension presumably by buckling. At higher forces, negative twist caused the HMfB dimers to unwrap from the DNA, which induced stretching of the HMfB-DNA complex. These measurements suggest that the unstacked HMfB-DNA complex is right-handed. As a reference, we included a twist-extension curve of bare DNA at 1.0 pN. Note that the slope for the HMfB-DNA complex is much smaller than for bare DNA.

copy experiments (FIGURE S4.3), which show that positively twisted complexes more readily restack into hypernucleosomes than relaxed or negatively twisted DNA.

The slope of the curve beyond the buckling transitions reveals more details of the twisted complex. Bare plectonemic DNA reduces its extension by approximately 40-100 nm per rotation, which depends on force, as the size of the loop at the tip of the plectoneme decreases with force (314). For unstacked HMfB-DNA complexes we measured a reduction of the extension of 16 nm per rotation, much lower than bare DNA, independent of force. This is consistent with a fixed HMfB-induced bend at the tip of the plectoneme, which defines the tip curvature, and thus the geometry of the plectoneme.

Discussion

The hypernucleosome structure may be a key feature of archaeal chromosome compaction. Here we characterized the structural and mechanical properties of the hypernucleosome in solution. We found that hypernucleosomes composed of stacked *M. fervidus* histone homodimers HMfA and HMfB compact DNA much stronger than similar DNA-bending proteins. Both the HMfA- and the HMfB-DNA complexes featured a two-step unfolding mechanism that could be captured quantitatively in the transitions between a hypernucleosome, an array of wrapped dimers that kink the DNA trajectory, and a stretched DNA molecule to which dimers remain bound. HMfB featured a higher stacking energy than HMfA, resulting in a larger occupancy as well as a higher stability of the hypernucleosome against force.

The stiffness of hypernucleosomes is larger than that of stacked eukaryotic nucleosomes (315). This may be explained by the different nature of the stacking interactions. Interactions between eukaryotic nucleosomes are known to be mediated by flexible histone tails. Our data suggests that the archaeal hypernucleosome on the other hand is mediated by stacking interactions between the globular histones that form the hypernucleosome. Stacking energies of HMfA and HMfB dimers (~2 k_BT) are much lower than the 17 k_BT reported for eukaryotic nucleosome stacking (316). The unwrapping of DNA from the HMfA and HMfB dimers also starts at similar forces as the unwrapping of eukaryotic nucleosomes. However, in hypernucleosomes this happens reversibly, whereas nucleosome unwrapping is generally not in equilibrium.

The equilibrium conditions made it possible to capture the entire force-extension curve in a statistical physics model. This worked better for the unstacking transition than for the unwrapping transition. It should be noted though that modelling of the wrapped dimers as kinks in the DNA trajectory might be naive, as the high occupancy of the DNA yields a dimer-dimer distances that are much smaller than the typical deflection length for which the Kulić and Schiessel model is valid (309). Moreover, this model assumes a flat one-dimensional kink, while the super-helical structure of the unfolded hypernucleosome may be more complex. A deflection angle (10°), which matched the curves best, is rather small for the 30 base pair footprint of a HMfA or HMfB dimer. Nevertheless, the resulting force dependent reduction of the end-to-end distance matched quantitatively with the force-extension curve. We resolved a clear transition from a flexible (regime II) to a stiffer WLC (regime III), using the same number of dimers as in the low force transition, indicating that this unfolding intermediate can be effectively modeled

as an array of wrapped dimers.

In the crystal structure, the hypernucleosome compacts DNA in a left-handed manner, which resembles DNA wrapping by eukaryotic histones. In accordance, we found that negative supercoiling stabilizes the most compact regime of the HMfB-DNA complex. However, after breaking the stacking by force a right-handed HMfB-DNA structure was observed. The right-handed HMfB-DNA complex has been described previously as a functional form, responding to environmental cues *in vitro* (251, 305). Depending on conditions, force and torque, a right-handed HMfB-DNA complex may have a biological function in archaea.

The archaeal genome is not only organized by histones but also by other NAPs. The components of the organizational machinery however differ per phylum, class and order. Histones are found in almost every branch of the archaeal domain. Currently, evidence in support of hypernucleosome formation has only been obtained for *M. fervidus* using x-ray crystallography, and for *T. kodakarensis* using MNase digestion of isolated chromatin and analysis of effects of histone binding on transcription *in vivo* (165, 197). It is likely that also histones from other species are able to assemble into hypernucleosomes (134). Formation of a hypernucleosome is likely stochastic, and initiation at multiple sites may lead to lateral frustration. Nevertheless, long stretches of hypernucleosome will probably be rare in the presence of processivity factors driving transcription elongation (317). Also, as nucleosomes are occluded at transcription start sites based on MNase digestion studies, the length of hypernucleosomes may in most cases be limited to a length on the order of genes or operons (172).

The majority of archaeal chromosomes contains genes coding for two or more histones. This may provide archaea with a means to regulate gene expression via changing the length and stability of the hypernucleosome in response to growth phase or environmental cues (251). *M. fervidus* histones form homo- and heterodimers, and, based on sequence similarities, histones from other species may be able to do so as well. This would provide archaea with a wide range of tools for transcription regulation (318). For instance, heterodimers may bind to DNA and interact with other heterodimers via one multimerization site, while other sites remain unbound. Homodimers may form hypernucleosomes, which can be interrupted by other homodimers incapable of certain interactions. Furthermore, sequence specificity in certain histone variants may direct histones to specific genomic targets. The emergence of the N-terminal tail may have diverted the eukaryotic histone from the archaeal histone, eventually resulting in alternative abilities to regulate genes than in archaea (134). The current methods of biological

analysis at the single molecule level allow for detailed, mechanistic measurements that have the potential to shed light on the evolutionary path of genome compaction and transcription regulation.

Materials and Methods

DNA substrate preparation

For the TPM DNA subtrate, a random sequence of 50% GC content (CGGCGCAAATTCGTGACCAGTTGCATCAGCTGCGTGAGCTGT-TTATCGCAGCATCGTAACAGGATAGTGAAGAAGACT) was cloned into pBR322, as described previously (127), resulting in plasmid pRD121. We used PCR to generate and amplify a 685 bp linear substrate (sequence in **SUPPLEMENTARY INFORMATION**) containing the cloned sequence, using digoxygenin- and biotin-labeled oligonucleotides (195).

For magnetic tweezers a 3646 bp DNA fragment based on plasmid pFW01 (pBP-CYC1(wt)/3 derivative (319)) was digested with BsaI and BseYI (New England Biolabs). The fragment was separated using agarose gel electrophoresis and purified using a Promega Wizard® SV Gel and PCR Clean-Up System. It was subsequently labeled with digoxigenin and biotin at either end by a Klenow reaction.

For the torsionally constrained construct, pUC18-based plasmid pTB01, (containing 12 repeats of the Widom 601 sequence (320)), was digested with BsaI and BseYI. The 4092 bp digested product was separated using agarose gel electrophoresis and purified using a Promega Wizard® SV Gel and PCR Clean-Up System. Subsequently, a 650 bp handle containing multiple digoxigenin- or biotin-modified bases was ligated on each end. Details on both approaches are provided elsewhere (321). The 4092 bp substrate based on plasmid pTB01 qualitatively behaved the same as the 3646 bp pFW01-based substrate.

HMfA and HMfB

HMfA and HMfB were kindly provided by John Reeve and Kathleen Sandman (16).

Expression and purification of HMfA mutant

The histone mutant gene was synthesized by GeneArt (Thermo Fisher Scientific, MA) and cloned into pET30b vectors using Gibson assembly. This resulted in plasmids pRD323, expressing $HMfA_{K31A\ E35A}$. pRD323 was transformed to *E. coli* BL21 pLysS (DE3). 1L LB medium with 25 mg/ml chloramphenicol and 50 mg/ml kanamycin was inoculated with 1% v/v culture grown to saturation o/n. The culture was grown at 37°C with shaking at 250 rpm. At $OD_{600} \approx 0.6$, expression of

the mutant was induced with 1 mM IPTG. The culture was grown for 1 h at 37°C with shaking at 250 rpm. Cells were harvested after 1 h and spun down for 30 minutes at 11,325 xg and at 4°C. Cells were resuspended in resuspension buffer (50 mM Tris-HCl pH8, 100 mM NaCl and 5 mM $MgCl_2$ in water), then frozen at -80°C and thawed the next day. DNase I, PMSF and lysozyme were added and cells were lysed using the pressure homogenizer unit 5 (Homogenising Systems Ltd, UK) at 350 MPa in resuspension buffer. The lysate was then centrifuged for 30 minutes at 11,325 xg and at 4°C. Pellets were resuspended in 3M NaCl. The resulting mixture was incubated at 95°C for 10 minutes, then spun down for 20 minutes at 10,967 xg and at 4°C. Treated lysate was desalted using a 5 ml HiTrap desalting column (GE Healthcare, IL) using the Äkta start protein purification system (GE Healthcare, IL) at RT. Flow-through was analyzed using Coomassie-stained SDS-PAGE gels. The fractions containing the proteins of interest were loaded onto a 5 ml HiTrap DEAE Sepharose FF column (GE Healthcare, IL) and eluted with 150 ml 50 mM Tris-HCl pH 8 and a sodium chloride gradient from 0 to 1 M using the Äkta pure chromatography system (GE Healthcare, IL) at RT. Peak fractions of 0.5 ml were collected and checked on Coomassie-stained SDS-PAGE gels. Fractions containing the protein of interest were combined and concentrated to 500 µl using Amicon ultrafiltration membranes (Merck, Germany; cut-off: 3 kDa) at 3700 rpm. The combined fractions were loaded on a Superdex 75 Increase 10/300 GL gel filtration column (GE Healthcare, IL) and eluted with a degased solution of 50 mM Tris-HCl pH 8 and 100 mM NaCl in water, using the Äkta pure chromatography system at RT. Peak fractions of 0.5 ml were collected, checked on Coomassie-stained SDS-PAGE gels and purity was examined using mass spectrometry. Fractions with less than ~5% impurities were considered pure. Concentration of $HMfA_{K31A\ E35A}$ was determined using a Pierce BCA Protein Assay kit (Thermo Fisher Scientific, MA) and was 151.3 µM. Protein solutions were plunge frozen and stored at -80°C.

Tethered Particle Motion

The tethered particle motion experiments were carried out as described previously (322). The measurement buffer contained 50 mM Tris-HCl, pH 7 and 75 mM KCl. We used an anisotropic ratio cut-off of 1.3 and a standard deviation cut-off of 8% to select single-tethered beads. The end-to-end distance of the DNA substrate was determined by taking the average of the extremities (2.5% largest values) of the deflection of 25 beads in the xy plane at 50 nM of HMfB and subtracting the radius of the bead (**FIGURE S4.1**). The end-to-end distance was obtained by triangular calculation.

Force spectroscopy and rotational spectroscopy

HMfA and HMfB were diluted to 100 nM in measurement buffer. 1 pM DNA substrate was introduced in the flow cell and incubated with histones, following the TPM protocol, but using paramagnetic M270 beads (Dynabeads; Thermo Fisher Scientific, USA) instead of polystyrene beads. Details of the force spectroscopy experiments and the multiplexed magnetic tweezers setup have been described elsewhere (321). In each measurement, the force was increased exponentially from 0.01 to 40 pN in 70 seconds, and decreased at the same speed.

Quantitative modeling of stretching the hypernucleosome

To interpret the force-extension behavior of the hypernucleosomes, we developed a statistical mechanics model in collaboration with Thomas Brouwer. HMfA and HMfB dimers bound to the DNA were distributed over three conformations, representing different levels of compaction. In the lowest force regime, HMfA and HMfB homodimers assembled into a hypernucleosome. The extension of the hypernucleosome was modeled by a freely-jointed chain (FJC) (323, 324) (**EQUATION 4.1**),

$$z_{FJC}(f) = L_{dimer}\left(coth\left(\frac{f_b}{k_B T}\right) - \frac{k_B T}{f_b}\right) \tag{4.1}$$

where $z_{FJC}(f)$ is the extension per dimer, b the Kuhn length, f the force, k_B Boltzmann's constant, T the temperature and L_{dimer}, which was fixed at 4 nm, which is the length of the DNA segment that is wrapped around a single dimer. The FJC was chosen for its initial linear increase of extension with force, as observed experimentally. In the final model only the linear part of the FJC i.e. an extension up to 2 nm per dimer mattered, as the dimers unstack at larger forces. For low forces ($f_b \ll k_B T$), the FJC describes a Hookean spring with linear extension, and Kuhn length can be rewritten to include stiffness k of the stacked fiber (**EQUATION 4.2**) (323).

$$b = \frac{3k_B T}{kL_{dimer}} \tag{4.2}$$

The second force regime is characterized as a beads-on-a-string structure. In this structure, DNA wraps onto non-interacting HMfA and HMfB homodimers resulting in moderate compaction of the fiber. The beads-on-a-string structure was modeled as an extensible worm-like chain (WLC) (310, 325) (**EQUATION 4.3**).

$$z_{WLC}(f) = L\left(1 - \frac{1}{2}\sqrt{\frac{k_B T}{fP}} + \frac{f}{S}\right) \tag{4.3}$$

in which $z_{WLC}(f)$ is the extension of protein-DNA complex, L is the contour length of the DNA substrate, S the stretch modulus and P the persistence length. Since the footprint of an HMfB homodimer is 30 bp (326), each transition of a dimer from the hypernucleosome into a beads-on-a-string conformation reduces the contour length of the FJC by L_{dimer} and increases the contour length of the WLC with 30 bp. Moreover, each HMf dimer that bends the DNA reduces the persistence length of DNA to an apparent persistence length P_{app} (309) (**EQUATION 4.4**).

$$P_{app} = \frac{P}{\left(1 + PN^2 8\left[1 - \cos\left(\frac{\alpha}{4}\right)\right]/L\right)^2} \tag{4.4}$$

in which N is the number of dimers in the beads-on-a-string conformation and α is the DNA deflection angle induced by the dimer. In addition we added a small free energy contribution per kink g_{kink} that is proportional to the square root of force and does not exceed 0.01 $k_B T$ (327). The model matched the data best using a deflection angle of 10 degrees, which corresponds roughly to the structure depicted in **FIGURE 4.5B**.

At high forces, the HMfA and HMfB homodimers appeared to remain bound, but do not bend the DNA. Each transition therefore reduces the number of kinks in the DNA, resulting in an increase of the apparent persistence length to that of bare DNA, for which we used P = 50 nm and S = 1000 pN (324, 325, 328).

Many different states j of the fiber can be defined containing different distributions of each of the three dimer conformations i, which can be the hypernucleosome, beads-on-a-string, or straight conformation. The total extension $z_j(f)$ of the tether is a linear combination of the extension of each dimer conformation $z_i(f)$ multiplied by the number of dimers n_i in conformation i (**EQUATION 4.5**).

$$z_j(f) = \sum_i n_i z_i(f) \tag{4.5}$$

The free energy per dimer in the hypernucleosome is given by:

$$g_{FJC} = \int z_{FJC}(f)\mathrm{d}f = L_0 \frac{k_B T}{b}\left(\ln\left(\sinh\left(\frac{fb}{k_B T}\right)\right) - \ln\left(\frac{fb}{k_B T}\right)\right) + g_{stack} + g_{wrap} \tag{4.6}$$

using a protein-protein interaction energy g_{stack} and a protein-DNA interaction energy g_{wrap}. The free energy per dimer in the beads-on-a-string and the straight structure is given by:

$$g_{WLC} = \int z_{WLC}(f)\mathrm{d}f = L\left(-\sqrt{\frac{fk_B T}{P}} + \frac{f^2}{2S}\right) \tag{4.7}$$

optionally complemented by g_{wrap} for the wrapped conformation. In addition, the work W done by the bead corresponds to:

$$W = \int f(z_j)\mathrm{d}z \tag{4.8}$$

The total free energy $G_j(f)$ of a particular tether is the work by the bead minus the sum of the free energy contributions of each dimer with an additional force-dependent free energy contribution for the number of kinks $Ng_{kink}(f)$ (327):

$$G_j(f) = W - \sum_i n_i g_i(f) + Ng_{kink}(f) \tag{4.9}$$

Finally, the force dependent extension of the hypernucleosome is calculated as the Boltzmann-weighted mean extension:

$$< z_{tot}(f) >= \frac{\sum_j z_j(f)\exp(-G_j(f)/k_B T)}{Z} \tag{4.10}$$

with partition function $Z = \sum_j \exp(-G_j(f)/k_B T)$.

Prior to fitting, the z-offset z_0 was determined by fitting the WLC for bare DNA at forces >30 pN. Subsequently, the number of dimers on the DNA was determined based on the maximum compaction of state I and state II (*see also* **FIGURE 4.3C**). The stiffness k of the hypernucleosome was fixed at 1.2 pN/nm which yielded a good fit to the data, and since the hypernucleosome unstacks at a relatively low force fitting this value was problematic. **EQUATION 4.10** was fitted to the data between 0.5 and 55 pN to extract the composition, stacking and energy g_{stack} and wrapping energy g_{wrap} from each curve.

Supplementary information

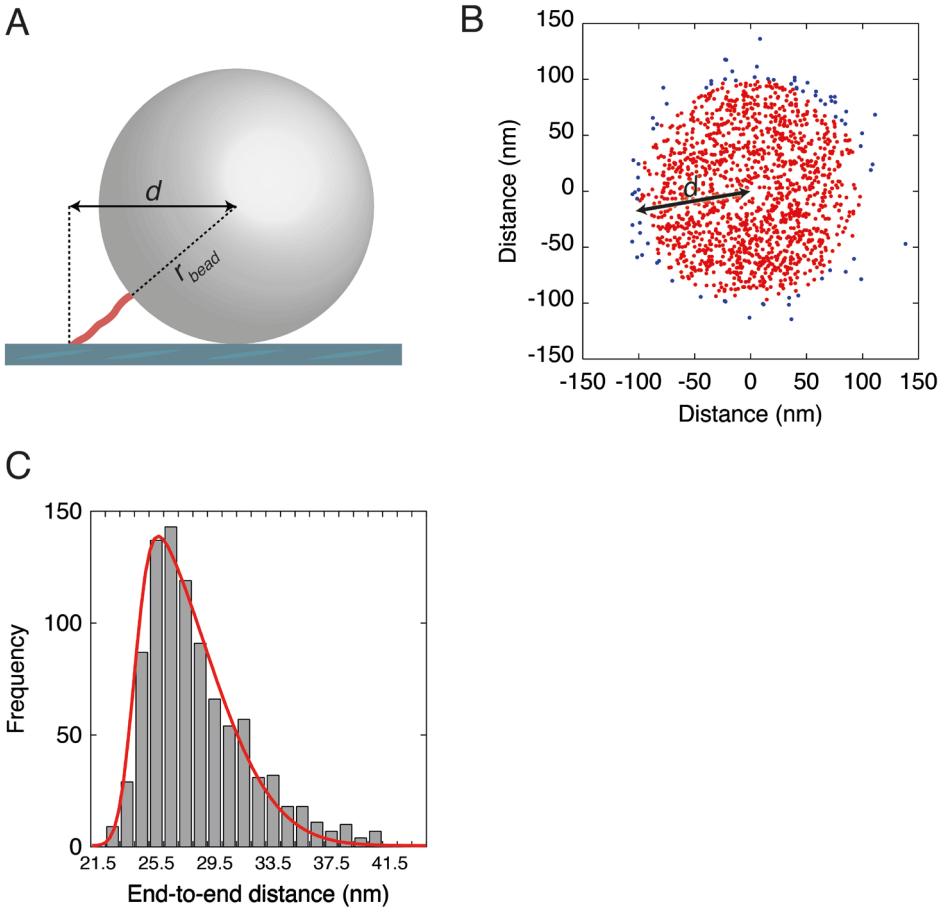


Figure S4.1: End-to-end distance of the HMfB-DNA complex using TPM. A) Schematic of the parameters used in the calculation of the end-to-end distance. *d* is the horizontal distance between the center of the bead and the place where the DNA is attached to the flow cell; r_{bead} is the radius of the bead. B) Positions of TPM bead over 60 seconds at 50 nM HMfB (red and blue). The 2.5% positions at the extremities of the plot (blue) represent the end-to-end distance of the HMfB-DNA complex. C) Skewed Gaussian fit (red) of the end-to-end distance of the 2.5% most distant positions with respect to the center of 25 beads.

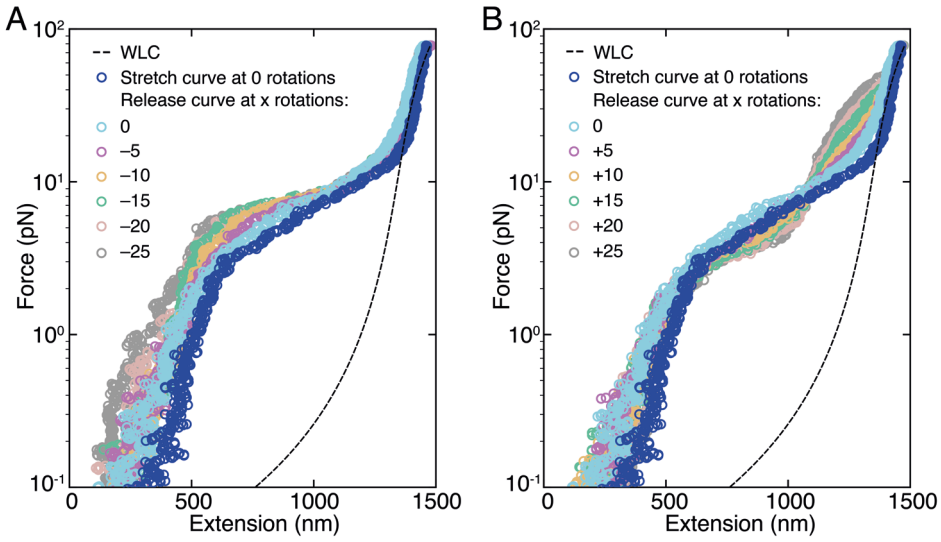


Figure S4.2: Force spectroscopy on torsionally constrained and torsionally free HMfB-DNA complex. At low force (<2 pN) both complexes are identical in length, and the curve overlaps with the freely-jointed chain model (FJC). At high force (>20 pN) the curves overlap with the worm-like chain model (WLC) of bare DNA. In between, the hypernucleosome undergoes an unstacking transition and subsequently an unwrapping transition. Due to the chiral structure of the hypernucleosome, unstacking increases torque, which stabilizes the structure. The torsionally constrained hypernucleosome unstacks at ~10 pN, compared to ~1.5 pN for the torsionally unconstrained hypernucleosome. The unwrapping transition appears unaffected. Fitting the torsionally constrained hypernucleosome yields a higher stacking energy (4.2 ± 0.1 k_BT vs 1.6 ± 0.0 k_BT per dimer), reflecting an increased energy cost for unstacking the hypernucleosome. The wrapping energy is largely unaffected (9.9 ± 0.1 k_BT vs 10.0 ± 0.1 k_BT per dimer). The fit, however deviates somewhat from the data in the transition region. Both molecules are associated with an identical number of HMfB dimers: 103 dimers, from which 95 are stacked in a hypernucleosome. The force extension curves of both molecules converge to a WLC with a contour length of 3646 bp at forces >30 pN.

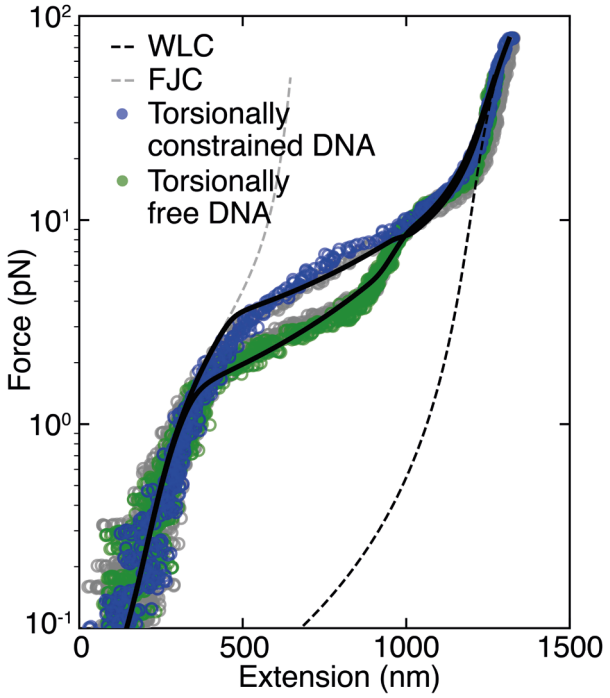


Figure S4.3: Release curves of torsionally constrained HMfB-DNA complexes. A) Force spectroscopy on negatively twisted HMfB-DNA complexes. The release curves suggest that overstretched DNA has higher affinity for HMfB dimers compared to relaxed DNA, shown by the reduced extension of the release curve. As a reference, the stretch curve for 0 twist has been added in bright blue. Upon negative twist, this hysteresis becomes less. B) Force spectroscopy experiments on positively twisted HMfB-DNA complexes. At forces between 10 and 40 pN the extension is reduced with increasing twist, which was also observed in the pulling curve.

DNA substrates

TPM substrate

5’ TTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGG-
GAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCAT-
TTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAG-
GGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGA-
CATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCCGGCGCAAAT-
TCGTGACCAGTTGCATCAGCTGCGTGAGCTGTTTATCGCAGCATCGTAACAGGATAGTGAA-
GAAGACTAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGTG-
TATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCATA-
GGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGT-
CACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAG-
CACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGG 3’

CHAPTER 5

DNA binding by histones from *Ca.* Nanohaloarchaeota

Abstract

Members of the archaeal phylum *Candidatus* Nanohaloarchaeota have recently been found in hypersaline lakes and salt crusts. For functionality, the macromolecules of these organisms, including histones, must be adapted to high cytoplasmic salt concentrations. Based on our sequence analysis and classification, three histone types have been identified in *Candidatus* Nanohaloarchaeota. Histones of type A are predicted to form tetramers and hypernucleosomes, whereas histones of type B can probably form no structure larger than a dimer, and may have a lower binding affinity than histones of type A. Prediction of characteristics of type C histones was less conclusive. Here, we investigate two histones encoded by *Nanosalina* sp. J07AB43. *Nanosalina* HA and HB, type A and type B histones, respectively, were recombinantly expressed and their interactions with DNA were characterized using microscale thermophoresis and tethered particle motion. We find that at 75 mM KCl, HA binds and probably compacts DNA, whereas HB does not. At 1.5 M KCl, no direct binding is observed, but both proteins do affect fluorescence when added, suggesting binding to DNA. HA does this at lower concentrations than HB, indicating that HA indeed is able to better bind DNA than its type B counterpart.

Introduction

Archaea, the domain of prokaryotic organisms from which eukaryotes have emerged, have recently seen a sharp increase in the availability of genetic information. Whereas initially archaea were found in extreme environments such as hydrothermal vents and hot springs, scrutiny of the archaeal domain revealed that they exist in nearly every habitat on earth. The archaeal tree of life dramatically changed shape over the last ten years as a result of the discovery of novel archaeal superphyla, phyla and classes (47). One of these more recent additions is the candidate phylum Nanohaloarchaeota (329). After its discovery at Lake Tyrell, Australia, it was initially classified as a new euryarchaeal class, but was later categorized as a phylum part of the DPANN superphylum (39). Three classes of *Candidatus* (*Ca.*) Nanohaloarchaeota have been identified, Haloredivivus, Nanosalina and Nanosalinarum, all resulting from metagenomic sequencing of samples from hypersaline lakes (329, 330). Additionally, fifteen unclassified genomes of *Ca.* Nanohaloarchaea were discovered on salt crusts in the Chilean Atacama desert (331, 332). None of the *Ca.* Nanohaloarchaeota have been reported to have grown in culture. The cells of *Ca.* Nanohaloarchaeota are small with an estimated cell size of 0.6 μm (329). Their genomes are also small, ranging from 0.40 to 1.48 Mbp. Most of these genomes have a GC content of 40-50%, which is low compared to other archaeal genomes that were found in the same habitat, but similar to Methanobacteria and Nanoarchaeota (329).

The proteins predicted to exist in archaeon SG9 (genome length 1.1 Mbp) have a median theoretical isoelectric point (pI) of 4.7. This would place archaeon SG9 in the first percentile of average pIs of more than 5000 proteomes from throughout the tree of life (333). The average pIs of some halophilic archaea from the euryarchaeal class Halobacteria are comparable to those of Nanohaloarchaeota. The median pI of the average of all archaeal proteomes is 6.4, with half of all archaeal proteomes having an average pI between 5.8 and 7.1 (331, 333). Scrutiny of the proteins encoded by the genome of *Nanosalina* sp. J07AB43 also indicated pIs between 4 and 6, similar to archaeon SG9. Living in conditions with high salinity is challenging, as high salt concentrations in the environment lead to an increase of the osmotic pressure inside the cell. A low average pI of the proteome is indicative of the 'salt-in' osmotic strategy, which microorganisms inhabiting hypersaline environments use to balance the osmotic pressure of their cellular contents with the pressure of the environment (334). In this strategy, ions, mainly Cl^- and K^+, are accumulated inside the cell, which means that the proteome needs to be adapted for functionality at very high intracellular salt concentrations (335, 336). For species of Halobacteria, intracellular KCl concentrations up to 7 M have

been recorded (335). For many proteins that are expressed by organisms that use the salt-in strategy, a high intracellular salt concentration contributes to protein folding. Exposure to low salt environments can therefore lead to denaturation. Only some halobacterial archaea, as well as some bacteria, use the salt-in strategy for haloadaptation. Most other salt loving or salt tolerant organisms exclude rather than let in salt ions, using high concentrations of sugars, alcohols, glycerol or other organic compounds without net charge to counterbalance osmotic pressure (337). This places members of *Ca.* Nanohaloarchaeota in a small subgroup of extremophiles.

Genomes of *Ca.* Nanohaloarchaeota encode the nucleoid-associated protein Alba, as well as histones. Alba-encoding genes were identified on approximately half of the annotated nanohaloarchaeal genomes, whereas genes coding for histones can be found in all annotated genomes. Both Alba and histones have, like most other proteins in this phylum, exceptionally low theoretical pIs, on average 4.2 and 5.3, respectively. All nanohaloarchaeal Alba proteins contain the residue Lys16, which is subject to acetylation or methylation (119, 120). Post-translational modification alters Alba's binding affinity to DNA, and is therefore important for modulation of its activity. Most Alba proteins encoded by *Ca.* Nanohaloarchaeota are of type Alba1, and only *Haloredivivus* sp. G17 encodes a protein of type Alba2 (338). Histones encoded by the genomes of *Ca.* Nanohaloarchaeota are predicted to contain a single histone fold, unlike the 'tandem histones' from Halobacteria, which are found in the same habitat (185, 339). They likely form dimers in solution, like other archaeal histones (160, 193). Nanohaloarchaeal histones have been categorized in three types (*see also* **CHAPTER 2**). Although histones of all nanohaloarchaeal types are predicted to bind DNA, histones of type A are predicted to be able to tetramerize and assemble into hypernucleosomes, which strongly compact DNA (197, 339). Histones of type B lack the residues providing interactions necessary for tetramerization, which likely makes them unable to form structures larger than dimers (**FIGURE 5.1**). Especially the absence of H49 likely makes tetramerization impossible. Furthermore, they lack AGA at position 15-17, situated at the N-terminal loop region. Bulkier amino acid residues at this

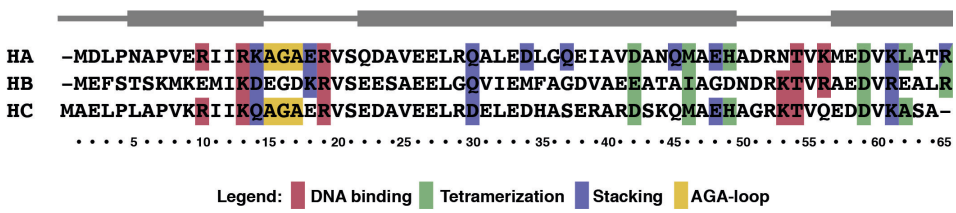


Figure 5.1: Alignment of histones HA, HB and HC encoded in the genome of *Nanosalina* sp. J07AB43. Grey bars represent α-helices.

position generally cause steric hindrance between histone dimers when multimerizing, which means that substitution of AGA for amino acids with larger side chains prevents hypernucleosome formation. Also, histones of type B have 5 sites of potential DNA-binding interactions, whereas histones of type A have 6. This may result in weaker DNA binding of type B histones compared to histones of type A. Histones of type C may be able to tetramerize, although their tetramers are likely less stable than those of type A, due to a lower number of tetramerization interactions. It is unknown if histones of type C form hypernucleosomes, but if they do, these hypernucleosomes are likely less stable than hypernucleosomes composed of histones of type A, as a result of less stacking interactions (*see also* CHAPTER 2).

Here, we investigate DNA binding and compaction by histones from the genome *Nanosalina* sp. J07AB43, HA and HB. These histones belong to nanohaloarchaeal histone types A and B, respectively. Histone types A and B are hypothesized to exhibit clear differences in terms of multimerization, which can be empirically tested by *in vitro* experiments. We use microscale thermophoresis (MST), which measures difference in the sum of molecular size, charge and hydration shell, and tethered particle motion (TPM), which gives a read-out of DNA organization, to qualitatively probe compaction of DNA by these histones, permitting assessment of the protein's ability to form tetramers or hypernucleosomes.

Results and discussion

In order to investigate interaction between *Nanosalina* sp. J07AB43 histones HA and HB, we expressed and purified recombinant HA and HB. To determine whether HA or HB binding affects DNA conformation, we used TPM. We tethered polystyrene beads to a glass surface using 685 bp long digoxigenylated and biotinylated nonspecific DNA substrates. The root mean square displacement of the tethered beads gives a quantitative readout of changes in DNA conformation. DNA tethers were incubated with HA or HB at 75 mM KCl, a concentration at which other histones have been reported to bind DNA (*see also* **CHAPTERS 4** and **6**). Although low pH is physiologically more relevant considering the mode of life of members of *Ca.* Nanohaloarchaeota, in TPM, buffers with pH below 6.0 cannot be used following routine protocols: lower pH causes sticking of beads to the surface. Therefore, a pH of 6.0 was chosen for TPM experiments. Addition of up to 1000 nM HA or HB did not affect the root mean square displacement (RMS) of DNA (**FIGURE 5.2**). This suggests that in these conditions, HA and HB do not change DNA conformation.

We further investigated HA and HB interacting with DNA using microscale thermophoresis, which also responds to proteins binding to DNA without DNA

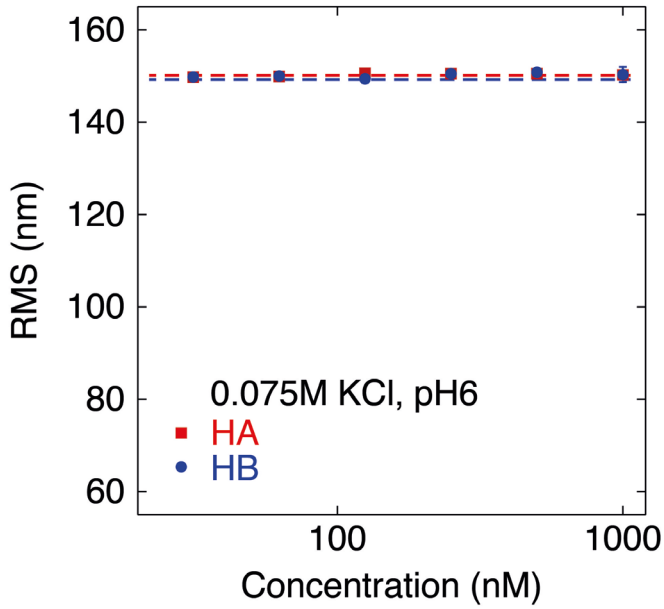


Figure 5.2: Tethered particle motion experiments show that HA and HB do not deform DNA at pH6 and 75 mM KCl. Error bars represent the standard deviation of the different experiments, and are mostly hidden behind the data points. Dashed lines are to guide the eye.

deformation. We used linear, Cy5-labeled, nonspecific DNA with a length of 78 bp and a GC content of 50%, close to the average GC content of nanohaloarchaeal genomes. The length of the DNA substrate is sufficient to harbor a histone tetramer, but is too short to accommodate larger histone assemblies (340, 341). To mimic physiological pH values, we incubated Nanosalina histones with the DNA substrate in a solution containing 75 mM KCl at pH 4.0. Next, we measured thermophoresis at histone concentrations between 0 and 1000 nM (**FIGURE 5.3A** and **B**). We quantify thermophoresis by taking the ratio of two regimes of the

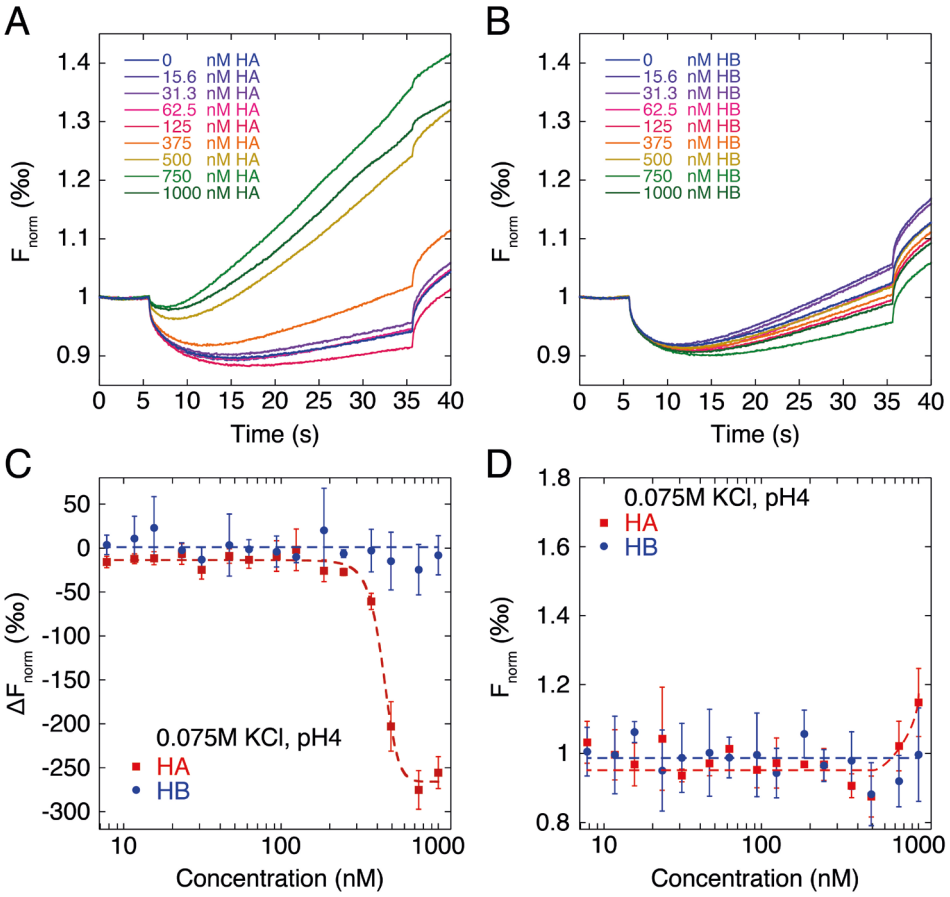


Figure 5.3: Microscale thermophoresis experiments show reduction in thermophoresis by HA, but not by HB at pH4 and 75 mM KCl. Representative normalized time traces of thermophoresis experiments on 78 bp nonspecific DNA substrates for Nanosalina histones HA (A) and HB (B). Normalized fluorescence (F_{norm}) is expressed as function of time. C) Thermophoresis of Nanosalina histones HA and HB on 78 bp nonspecific DNA substrates at pH4, expressed as the difference in normalized fluorescence compared to the value at 0 nM (ΔF_{norm}) as function of histone concentration. Error bars represent standard deviation of the experiments. D) Initial normalized fluorescence (F_{norm}) of Nanosalina histones HA and HB on 78 bp nonspecific DNA substrates, as function of concentration. Error bars represent standard deviation of the experiments.

thermophoresis time traces, by which we compare fluorescence before heating and fluorescence just before the end of heating (*see also* **Materials & Methods**). The ratio gives a read-out of how fast the DNA substrates diffuse out of the heated spot, which in turn indicates the change in size, charge and hydration shell when compared to a situation where only labeled DNA is present. Thermophoresis experiments showed that at concentrations of HA >250 nM, the difference in normalized fluorescence (ΔF_{norm}) decreases (**figure 5.3C**). This means that the combination of change in molecular assembly, charge and/or hydration shell changes as a result of histone binding and possibly DNA deformation (342). The response to HA is similar to the results of MST experiments on HMfA and HMfB, although that study also showed that ΔF_{norm} increases at protein concentrations >1000 nM (341). Due to limited availability of protein, we were not able to reach these concentrations. In contrast to the response to thermophoresis caused by HA, ΔF_{norm} did not change in the presence of HB, which suggests that HB does not bind and compact the DNA substrate. Since the footprint of archaeal histones is 30 bp per dimer, the 78 bp substrate can accommodate a tetramer. Based on structural and sequential analyses, we predicted that HA is able to bind DNA as tetramers, whereas HB also binds DNA but probably cannot form tetramers. The results of the experiments presented here are in line with this hypothesis. In addition to thermophoresis, protein binding to DNA can also have an effect on the initial fluorescence before heating of the sample. Therefore we normalized and plotted initial fluorescence as function of concentration. The initial fluorescence did not significantly change in response to protein concentration (**figure 5.3D**).

Based on the notion that the average pI of nanoarchaeal proteins is very low, we expect that members of *Ca.* Nanohaloarchaeota use the salt-in strategy for regulation of osmotic pressure. Reduction of salt concentrations to below 0.5 M (as used in the experiments reported above) leads to a shortage of cations shielding the protein surface, resulting in repulsive forces that weaken the conformation of some proteins (337). As it may lead to more stable histones, it is thus relevant to measure nanoarchaeal histones at high salt concentrations. However, when we carried out HA protein titrations using MST at 1.5 M of KCl, we no longer detect an effect of HA on thermophoresis (**figure 5.4A-C**). This suggests that HA is no longer binding and compacting the DNA substrate, or binds with low affinity. A similar behavior was observed for HB. However, when we look at initial fluorescence, we observe a positive correlation between histone concentration and fluorescence for both HA and HB (**figure 5.4D**). Fluorescence increases starting at a concentration of 125 nM for HA, whereas for HB, we observe an increase in fluorescence from 375 nM and onwards. One possible explanation for the increase in fluorescence signal is protein induced fluorescence enhancement (PIFE). This effect entails that fluorophore intensity increases upon binding of

a protein in its vicinity, more precisely within 0 and 3 nm of the dye (343, 344). Cy5, the dye attached to our DNA substrates, has been shown to exhibit PIFE (343). This means that if PIFE is a proxy of histone binding to DNA, initial fluorescence data does not correspond with thermophoresis data. It should be noted that fluorescence is proportional to the viscosity of the experimental buffer and that protein concentration contributes to viscosity (345, 346). However, we do not observe a proportional relation between protein concentration and initial fluorescence. Also, HA and HB are very similar in terms of size, charge and structure, which would suggest that their contribution to buffer viscosity is roughly equal. As our data indicate that HA and HB affect initial fluorescence at different con-

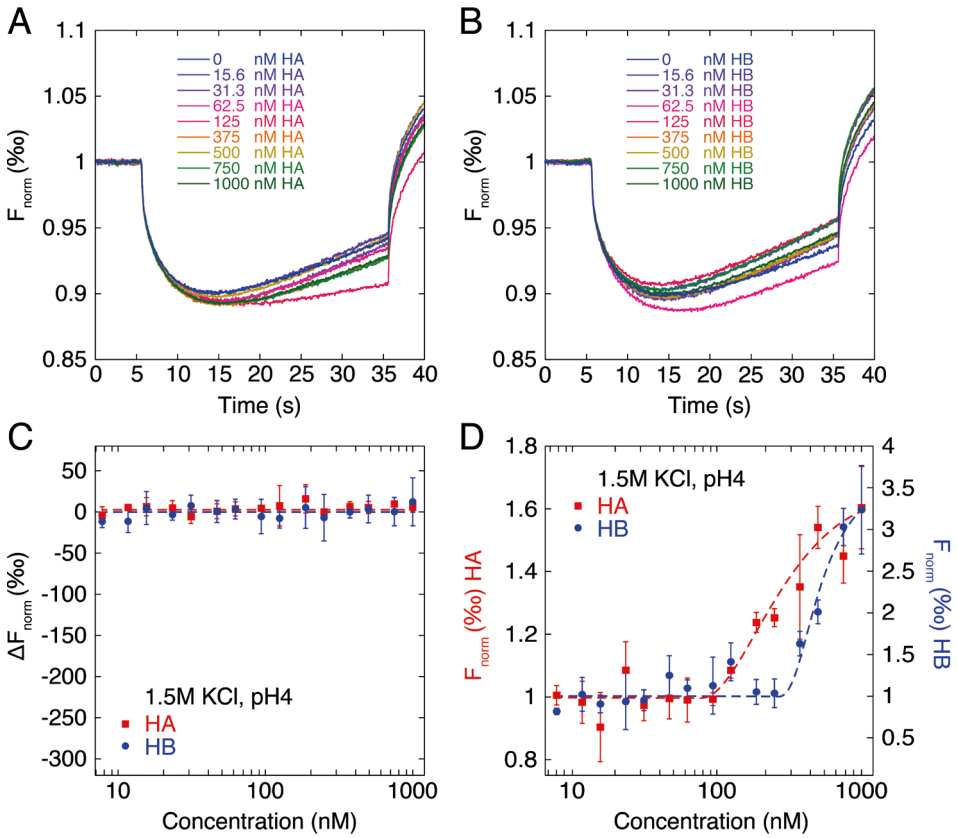


Figure 5.4: Microscale thermophoresis experiments indicate binding of DNA by HA and HB at pH4 and 1.5 M KCl. Representative normalized time traces of thermophoresis experiments on 78 bp nonspecific DNA substrates for Nanosalina histones HA (A) and HB (B). Normalized fluorescence (F_{norm}) is expressed as function of time. C) Thermophoresis of Nanosalina histones HA and HB on 78 bp nonspecific DNA substrates, expressed as the difference in normalized fluorescence compared to the value at 0 nM (ΔF_{norm}) as function of histone concentration. Error bars represent standard deviation of the experiments. D) Initial normalized fluorescence (F_{norm}) of Nanosalina histones HA and HB on 78 bp nonspecific DNA substrates, as function of concentration. Error bars represent standard deviation of the experiments.

centrations (**FIGURE 5.4D**), a viscosity-induced increase in fluorescence unlikely. The discrepancy between thermophoresis and initial fluorescence experiments may be explained by binding dynamics of HA and HB. High dissociation rates may result in a PIFE effect but may not be visible as an effect on thermophoresis. We did not observe an increase in initial fluorescence as function of protein concentration for HA and HB at 75 mM KCl. Possibly, higher protein concentrations would result in an effect on initial fluorescence, as the initial fluorescence seems to slightly increase at the highest concentrations of the titration used in this study. This may indicate a different form of binding at 1.5 M KCl compared to 75 mM. The observation of DNA binding at high salt concentrations might be indicative of the involvement of hydrophobic interactions. It should be noted that initial fluorescence cannot be quantitatively compared between conditions with different KCl concentrations, as fluorescence is inversely related to KCl concentration (**FIGURE S5.1**). We did not use concentrations of 1.5M KCl in TPM, because high salt concentrations are not compatible with our routine protocols. In order to establish whether HA and HB form multimers, we recommend further optimization to establish conditions at which histone-DNA complexes are stable. Establishing such conditions could also allow us to quantify DNA deformation by HA and HB. These proposed experiments may provide further insight into the function of multimerizing and non-multimerizing histones.

Conclusion

Nanohaloarchaeal histone HA from *Nanosalina* sp. J07AB43 binds DNA at 75 mM KCl and pH4. Although we have not been able to confirm DNA compaction by HA using TPM, the strong effect of the protein on fluorescence is, in accordance with previous studies, indicative of a large conformational change. This conformational change is probably caused by wrapping around a histone tetramer. HA was predicted to better tetramerize and bind to DNA than HB, based on sequential analysis. This study indirectly confirms the prediction, although optimization of pH and KCl concentrations may reveal the conditions at which HB also binds and compacts DNA. Experiments at KCl concentrations intended to mimic physiological conditions of *Ca.* Nanohaloarchaeota did not directly show binding, but increase in initial fluorescence in MST experiments suggests that HA and HB interact with DNA. Further optimization of experimental conditions may shed more light on nanohaloarchaeal histones and their involvement in DNA compaction. Insight into the DNA-binding characteristics of nanohaloarchaeal histones may reveal how mechanisms for DNA compaction and transcription regulation were established in the archaeal domain.

Materials & Methods

Purification of Nanosalina HA and HB

Genes coding for *Nanosalina* sp. J07AB43 histones HA (accession number EGQ42849) and HB (accession number EGQ43804) were codon-optimized for *Escherichia coli* and synthesized at GeneArt (Thermo Fisher Scientific, MA). Genes were cloned into pET30b vectors using Gibson assembly, resulting in plasmids pRD321 expressing HA and pRD322 expressing HB. pRD321 was transformed to *E. coli* BL21 star (DE3), and pRD322 was transformed to *E. coli* BL21 Rosetta (DE3), resulting in strains NT494 and NT495. For protein overexpression 500 ml LB medium with 25 mg/ml chloramphenicol and 50 mg/ml kanamycin was inoculated with 1% v/v culture grown to saturation o/n. Cultures were grown at 37°C with shaking at 250 rpm. At $OD_{600} \approx 0.5$, expression of HA was induced with 1 mM IPTG. Cultures were grown for 1 h at 37°C with shaking at 250 rpm. Cultures were harvested by centrifugation for 30 minutes at 11,325 xg and at 4°C. Cells were frozen at -80°C and thawed the next day. DNase I, PMSF and lysozyme were added and cells were lysed using the pressure homogenizer unit 5 (Homogenising Systems Ltd, UK) at 350 MPa in 50 mM Tris-HCl pH 8.0 and 100 mM NaCl in water. Cell lysate was separated from the pellet by centrifugation for 30 min at 10,967 xg and at 4°C. Lysates were incubated at 95°C for 10 minutes, then spun down for 15 min at 10,967 xg and at 4°C. The supernatant was diluted 3x in 16.7 mM Tris-HCl pH 7.0. The treated lysates were loaded onto a 5 ml HiTrap DEAE Sepharose FF column (GE Healthcare, IL) and eluted with 150 ml 50 mM Tris-HCl pH 8 and a NaCl gradient from 0 to 250 mM (for lysates of NT494) or 0 to 1 M (for lysates of NT495) using the Äkta start protein purification system (GE Healthcare, IL) at RT. Peak fractions were analysed on Coomassie-stained SDS-PAGE gels and combined. Combined fractions were concentrated to 400 µl using Amicon ultrafiltration membranes (Merck, Germany; cut-off: 3 kDa) at 3700 rpm. The combined fractions were loaded on a Superdex 75 Increase 10/300 GL gel filtration column (GE Healthcare, IL) and eluted with a degased solution of 50 mM Tris-HCl pH 8 and 100 mM NaCl in water, using the Äkta pure chromatography system (GE Healthcare, IL) at RT (flow rate: 0,5 ml/min). Peak fractions of 0.5 ml were collected, checked on Coomassie-stained SDS-PAGE gels and protein purity was examined using mass spectrometry. Fractions with less than ~5% protein impurities were considered pure. Coomassie-staining showed bands for HA and HB corresponding to 1x and 2x the expected molecular, indicating that a significant fraction of both proteins is present as dimer. Since the proteins are able to interact, we expect that the proteins are properly folded. Concentrations of HA and HB were determined using a Pierce BCA Protein Assay kit (Thermo Fisher

Scientific, MA) and were 100 µM and 561 µM, respectively. Protein solutions were plunge frozen and stored at -80°C.

DNA substrate preparation

For the MST substrate, two complementary oligos (Merck, Germany) of 78 nucleotides containing the 50% GC sequence as described in chapter 4 were mixed at 100 µM each. The oligos were annealed by heating to 95°C for 5 minutes, then decreasing the temperature in steps of 5°C, followed by 5 minutes incubation, and so on until a temperature of 20°C was reached. Substrates were subsequently diluted to 100 nM in 10 mM Tris-HCl pH 8 and 20 mM NaCl.

For the TPM DNA substrate, plasmid pRD121 was used, as described previously (*see* CHAPTER 4). We used PCR to generate and amplify a 685 bp linear substrate, using digoxygenin- and biotin-labeled oligonucleotides (195).

Microscale Thermophoresis

A dilution series of HA and HB starting at 2000 nM down to 15.6 nM was made using measurement buffer (50 mM acetate buffer pH 4.0 [11.3 mM sodium acetate and 38.7 mM acetic acid], KCl 75 mM or 1.5 M in water). Protein dilutions were mixed 1:1 with 100 nM Cy5-labeled MST substrate, resulting in protein concentrations from 1000 nM downwards and a constant substrate concentration of 50 nM in a total volume of 40 µl. Sample mixtures were transferred to MST capillaries (Monolith NT.115 Premium Capillaries, NanoTemper, Germany). Capillaries were measured using the Monolith NT.115 (NanoTemper, Germany) at 20% LED power and 20% MST power. Total measuring time was 40.4 s. NanoTemper Analysis software was used for data analysis of MST experiments. Time traces were normalized by alignment of the linear regime before onset of the MST laser (0-5.6 s). Thermophoresis was calculated by the ratio of normalized fluorescence between time windows 6.2-6.7 s and 34.7-36.6 s as described by Corbeski *et al.* (340). For the initial fluorescence curves, the average of initial fluorescence values at 0-62.5 nM was normalized to 1. The initial fluorescence values of the other data points are then shown as relative values to the values at 0-62.5 nM. MST experiments were carried out in triplicate.

Tethered particle motion

The tethered particle motion experiments, data analysis and representation of results were performed as previously described (322). 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 we used 50 mM phosphate buffer pH 6.0 (6.8 mM dipotassium phosphate and 43.2 mM monopotassium phosphate) and 75 mM KCl in water. TPM experiments were carried out in duplicate.

Supplemental figure

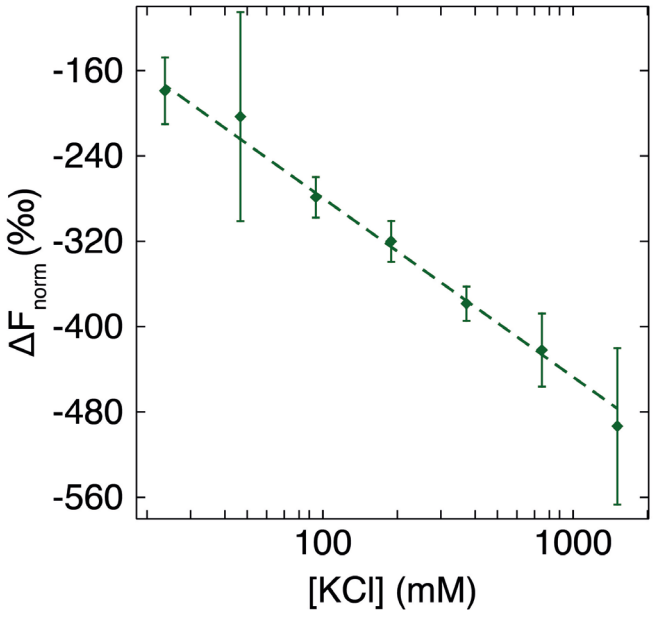


Figure S5.1: Initial fluorescence as a function of KCl concentration. Fluorescence decreases at larger KCl concentrations. Error bars represent standard deviation of the experiments. Dashed line is to guide the eye.

CHAPTER 6

Specific DNA binding of archaeal histones

Abstract

In archaea, histones play a role in genome compaction and are involved in transcription regulation. Although no specific sequence is required for DNA binding, archaeal histones prefer binding to DNA containing repeats of alternating A/T and G/C motifs. These motifs are also present on an artificial sequence Clone20, a high affinity model sequence for binding of the histones from *Methanothermus fervidus*. Here, we investigate binding of HMfA and HMfB to Clone20 DNA. We show that specific binding at low protein concentrations (<22 nM) modestly compacts DNA, whereas nonspecific binding strongly compacts DNA. We propose mechanisms by which high affinity sequences on the genome may interplay with other components that are likely involved in archaeal transcription regulation, such as histone variants and the hypernucleosome.

Archaea and eukaryotes express histones, which are involved in genome compaction and transcription regulation (177, 237). Together with other architectural proteins, such as Alba and MC1, archaeal histones are hypothesized to function as transcription regulators, although the mechanisms employed are currently unknown (117). The hypernucleosome, a rod-like structure consisting of a protein core of an archaeal histone multimer (multiple associated dimers), with DNA wrapped around compacts DNA, and may play a key role in archaeal transcription regulation (134, 197) (*see also* CHAPTER 4). Histones from species throughout the archaeal domain of life are predicted to be able to assemble into hypernucleosomes, although some genomes also (sometimes exclusively) encode histones that likely cannot multimerize (134). Archaeal histones have also been proposed to have defined roles as dimers or tetramers (196, 238). The dimer is the smallest functional unit of archaeal histones, and may be formed by monomers of the same or of different histone variants. Homo- and heterodimerization of histones could be key to modulating DNA compaction and transcription (251). Archaeal genomes harbor up to 11 histone variants, which allows archaea to vary the composition of histone dimers, tetramers and hypernucleosomes. Each form may have different functionality, with its formation dependent on growth phase or environmental cues (171, 251, 305).

Histones, like most archaeal chromatin proteins, bind DNA in a nonspecific way. However, histones have been suggested to be positioned at specific locations on the genome (196, 347). Also, histones are depleted from transcription start sites (TSSs), which is common in most unicellular organisms (221, 246). These TSSs are AT-rich, which disfavors nucleosome formation in both archaea and eukaryotes (221, 348). HMfB, one of the two histones from the archaeon *Methanothermus fervidus*, was reported to preferentially bind GC-rich sequences with alternating GC and AT motifs (164, 349). Specifically, (T)3(A)3GCCG repeat sequences exhibit a very high affinity towards *M. fervidus* histones (350). In *Haloferax volcanii*, patterns of GC and AT were shown to position histone tetramers in a similar way as reported for eukaryotes (220, 351). Sequences to which HMfB binds with high affinity were selected in early studies by Reeve and co-workers using a systematic evolution of ligands by exponential enrichment (SELEX) approach (164). Especially Clone20, a sequence resulting from this SELEX experiment, was shown to have a very high binding affinity for HMfB tetramers (an increase of two orders of magnitude compared to control DNA for HMfB, while the increase in affinity of HMfA was less than twofold, as measured via electrophoretic mobility shift assay [EMSA])(160). Clone20 contains repeats of alternating A/T- and

G/C-rich regions (*see* **Materials & Methods**). Clone20 can be regarded as the archaeal counterpart of the nucleosome positioning sequence 601, a sequence that energetically favors nucleosome formation, which is used in numerous studies on eukaryotic histones (320, 352-354). DNA sequences with high similarity to Clone20 and 601 have thus far not been identified in genomes, but high affinity sequences may be present on archaeal genomes (163, 221). High affinity sequences may be involved in positioning histones at or close to regulatory DNA elements, by means of which they could have an effect on transcription regulation.

Nonspecific DNA is cooperatively bound and compacted by HMfA and HMfB. Although affinity of HMfA and HMfB to Clone20 is high, it is unclear whether high affinity sequences play specific roles in DNA compaction, for instance as nucleation sites. Here, we use tethered particle motion (TPM) to investigate the role of the Clone20 sequence in DNA compaction by *M. fervidus* histones. The experiments were carried out as described elsewhere (195), using a linear DNA substrate with the Clone20 sequence at its center (*see* **Materials and Methods**).

Our experiments show that both HMfA and HMfB compact the Clone20 substrate, indicated by reduction of the root mean square displacement (RMS) of the DNA tether in TPM (**figure 6.1A** and **B**). Compaction occurs in two steps. The strong decrease in RMS at high protein concentrations (at >18 nM for HMfA and >22 nM for HMfB) is similar to that observed for strong cooperative compaction of nonspecific DNA into a hypernucleosome, as has been shown in chapter 4 of this thesis. (**figure S6.1**). Therefore we conclude that specific binding to the Clone20 sequence at concentrations <18 nM does not increase the effective binding affinity of the hypernucleosome, disfavoring a mechanism of site-specific nucleation, followed by propagation. Compaction is both somewhat more cooperative (reflected in a steeper slope) and stronger for HMfB than HMfA on the Clone20 substrate (an RMS of ~77 vs ~72 nm, **figure 6.1C**). Similar behavior was observed for nonspecific compaction on a DNA substrate without high affinity sequence (**figure S6.1**). The compaction at low protein concentrations (5-10 nM for HMfA and 10-22 nM for HMfB) was not observed for nonspecific DNA, and is attributed to specific binding of HMfA and HMfB to the Clone20 sequence. In this concentration regime, two populations of tethers with RMS values of ~155 nm and ~134 nm can be observed (**figure 6.1A** and **B**). The population sizes were used to calculate Clone20 sequence occupancy as a function of protein concentration, and a Hill equation was fit (**figure 6.1D**). This yielded a significantly lower binding constant (K_D) for binding to this sequence for HMfA than HMfB (5.7 ± 1.6 nM and 14.6 ± 1.6 nM, respectively). The higher affinity of HMfA for the Clone20 sequence compared to HMfB is unexpected as the sequence was

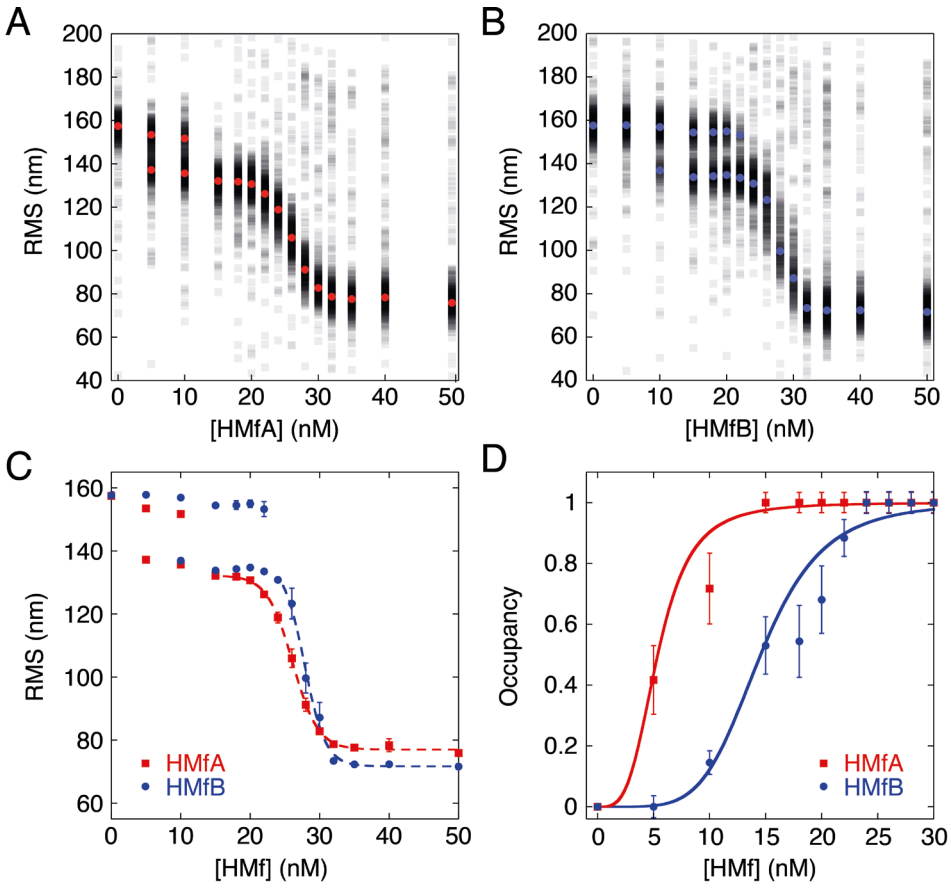


Figure 6.1: Compaction of a DNA substrate containing the Clone20 sequence by HMfA and HMfB using tethered particle motion (TPM). A) Root mean square displacement (RMS) values of all DNA tethers incubated with HMfA and B) with HMfB. Fitted mean RMS values are represented by red and blue dots, respectively. Grey squares represent the individual RMS values. C) Mean RMS values of DNA tethers with HMfA and HMfB. Error bars represent standard error of the mean of the replicates. Dashed lines are to guide the eye. D) Binding curve for specific binding of HMfA and HMfB to the Clone20 substrate. The data points were fit using the Hill binding model. The binding affinity (K_D) resulting from the fit is 5.7 ± 1.6 nM and 14.6 ± 1.6 nM and the Hill binding coefficient (n) is 3.6 ± 0.8 and 5.2 ± 2.0 for HMfA and HMfB, respectively. Error bars represent the standard error of the mean of the replicates, and propagated error for data points at saturation.

SELEX optimized for HMfB and is in disagreement with results from EMSA experiments (160). The difference may be caused by a different pH (7.0 in our experiments, 8.0 in the experiments of Bailey *et al.*). The difference in measured affinity could also be attributed to differences in methodology, with EMSA measuring mobility of protein-DNA complexes in a gel matrix and TPM measuring the conformation of protein-DNA complexes in solution attached to a glass surface. The observed reduction in RMS at concentrations <22 nM is attributed to for-

mation of a tetrameric histone-DNA complex (160, 171). This complex does not function as a nucleation site, as the concentration at which the hypernucleosome is formed, was not affected by this complex.

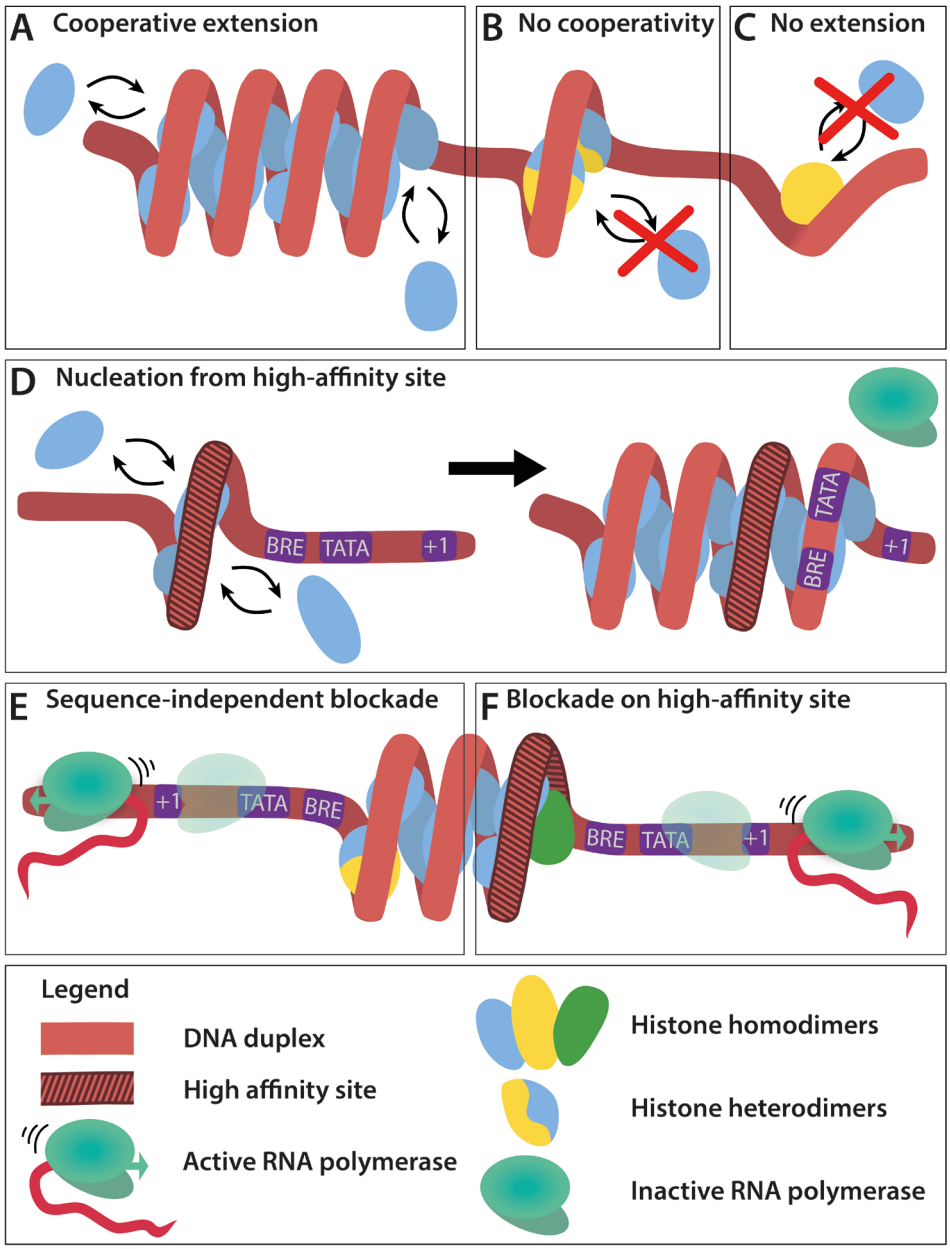


Figure 6.2: Possible mechanisms for involvement of archaeal histones in transcription regulation. Histones cooperatively bind and compact DNA in hypernucleosomes, which can be modulated via extension or dissociation from both ends (A). As tetramers (B) or dimers (C),

Conclusion

A DNA substrate containing the artificial high affinity sequence Clone20 is compacted by *M. fervidus* histones in two distinct steps, indicating two distinct complexes. At low concentrations, a histone assembly that is likely a tetramer, modestly compacts DNA. Although no sequences with a binding affinity similar to that of the Clone20 sequence have been identified on the *M. fervidus* genome thus far, it suggests a functional role for high affinity sequences. HMfA has a lower K_D than HMfB on Clone20. However, HMfB multimerizes more cooperatively into a hypernucleosome, and also nonspecifically compacts DNA more than HMfA. This may reflect a distinct cellular function, with HMfA being able to better position tetramers and/or hypernucleosomes at specific locations on the genome, and HMfB forming hypernucleosomes that are more stable than those formed by HMfA. Differential expression of HMfA and HMfB is in accordance with the hypothesis that histone variants play a role in transcription regulation, and these variants may concur with high affinity sequences and different forms of histone multimers in archaeal transcription regulation.

(Figure 6.2, continued) histones are also able to non-cooperatively bind DNA, thereby inducing less compaction than a hypernucleosome. Histones can function as transcription factors, possibly binding to specific sequences on the genome (D). These specific sequences may function as a positioning sequence for the hypernucleosome. Histone variant assemblies that are not able to multimerize may be able to nonspecifically (E) or specifically (F) prevent hypernucleosome formation in the promoter region, thereby leaving the promoter accessible to RNA polymerase.

Materials and Methods

DNA substrate preparation

For the TPM DNA subtrate, the Clone20 sequence (GCACAGTTGAGCGATCAAAAACGCCGTAGAACGCTTTAATTGATAATCAAAGGCCGCAGA, (164)) was cloned into pBR322, as described previously (127), resulting in plasmid pRD120. We used PCR to generate and amplify a 685 bp linear substrate containing the cloned sequence, using digoxygenin- and biotin-labeled oligonucleotides (195).

Tethered particle motion

The tethered particle motion experiments, data analysis and representation of results were performed as previously described (322). 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 we used 50 mM Tris-HCl pH 7 and 75 mM KCl in water. HMfA and HMfB were kindly provided by John Reeve and Kathleen Sandman (20). We verified the identity of both proteins using mass spectrometry.

Supplemental figure

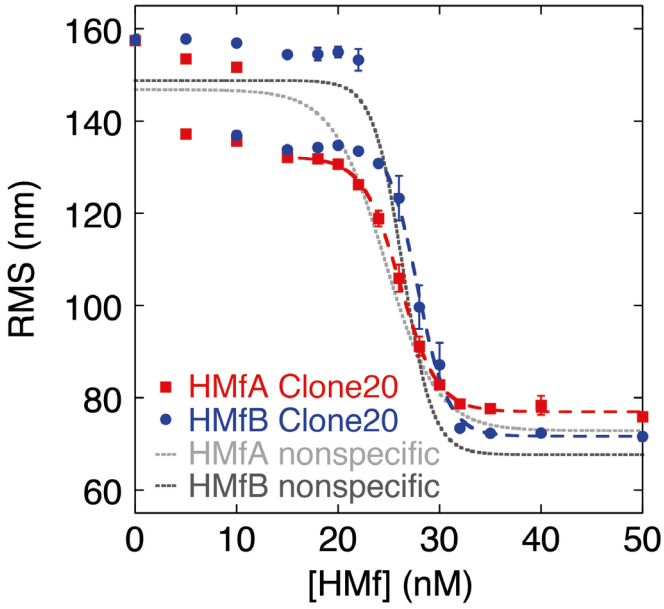


Figure S6.1: Compaction of HMfA and HMfB on a specific (Clone20) and nonspecific DNA substrate. Data of HMfA and HMfB on the Clone20 substrate is from this chapter (dashed lines are to guide the eye, error bars represent standard error of the mean), data of HMfA and HMfB on the nonspecific substrate is reproduced from CHAPTER 4 and depicted as a line to guide the eye.

CHAPTER 7

General discussion

Archaea appeal to the imagination of many. As inhabitants of hydrothermal vents, salt lakes, deserts, polar ice, geysers and mines, they raise questions about the limits of terrestrial life. The prefix 'hyper' is connected to archaea, as some are described to live in hypersaline lakes and hyperarid polar deserts, or are hyperthermophilic (355-357). Although extremities draw attention, it should be noted that many archaea have been identified in more common habitats, such as soil, oceans and intestines of animals, places that are also occupied by bacteria and eukaryotes. It seems that many extremophilic and mesophilic archaea are connected via another 'hyper' term: the hypernucleosome. This mechanism of DNA compaction via a rod-shaped histone core around which DNA is wrapped, has thus far only been found in archaea. The existence of the hypernucleosome was hypothesized based on micrococcal nuclease (MNase) digestion experiments, and crystallography studies revealed that archaeal histones really can form these structures (165, 197). We showed that hypernucleosomes are formed in solution. Also, we have made predictions on the ability of histones to form hypernucleosomes, based on the sequences encoded by archaeal genomes.

Histone variants and high affinity sequences

Histones, of which variants are encoded in most archaea, form dimers, tetramers and larger multimers. The hypernucleosome is potentially a very versatile tool for genome compaction and transcription regulation. Histone variants can modulate the structure of the hypernucleosome, which may contribute to a dynamic system that regulates transcription. Expression of histone variants in response to environmental cues could alter the length, stability or position of the hypernucleosome, which in turn changes compaction or availability of promoters. Also, specific high affinity sequences in DNA may position hypernucleosomes or tetramer complexes. Like many unicellular organisms, histones are generally depleted from the promoter region (246). However, on an individual basis, promoters may be prone to histone binding, if histones or histone variants have a higher affinity to certain sequences.

Eukaryotes use expression of histone variants and histone positioning by high affinity sequences for transcription regulation (191, 358, 359). Histone variants are (often essential) elements in eukaryotic gene expression, stress response, development, DNA damage repair and chromatin organization (360-365). Defects may result in disease. High affinity sites directly influence transcription initiation and histone remodeling (366). Although functionality of histone variants and the existence and relevance of high affinity sequences has yet to be established in archaea, it is not unlikely that these factors have a significant effect on vital cellular processes in archaea in a similar way as they do in eukaryotes. As the eukaryotic domain branches off from the domain of archaea, eukaryotes may have inherited this system from the first eukaryotic common ancestor (FECA). Now that research on archaeal histones and evolution of archaea are gaining momentum, the functionality of histone variants and high affinity sequences in archaea becomes an urgent question to be addressed.

Interplay with other NAPs

Archaea express a number of nucleoid-associated proteins (NAPs), of which histones, the main focus of this thesis, are arguably the best-studied examples. However, many other NAPs are encoded by the genomes of archaea. Alba, MC1 and Dps are the most widespread examples. In this thesis we focus mainly on interactions between histones, but we should not ignore the interplay that histones may have with other NAPs, both directly and indirectly. Alba-coding genes are found in the majority of archaeal genomes, and Alba is able to differentially regulate gene expression (117, 367). Like histones, Alba is hypothesized

to block the binding sites of the general transcription factors (GTFs), which are key elements in transcription initiation (117). Therefore it is important to know what factors influence the competition between Alba, histones and transcription factors. Post-translational modification of Alba alters its binding affinity, which allows for regulation of transcription initiation (119, 368). Alba is able to induce and constrain twist strain (200). In this way it may indirectly influence the binding of histones. Although this effect has not been studied for Alba together with histones, it has for the combination of the local twist-inducing transcription factor-like architectural protein TrmBL2 and one of the histones from *Thermococcus kodakarensis*. At a moderate level of twist strain, TrmBL2 is the dominant protein in terms of DNA binding, whereas at low and high superhelical densities, histones become the main DNA occupant (304). This effect may be explained by the fact that histones can both left- and right-handedly wrap DNA, which relieves twist strain. In contrast, TrmBL2 laterally binds DNA, a form of binding that benefits from a neutral superhelical density. TrmBL2 has been found in a very limited number of organisms, but Alba, as well as other proteins capable of altering or stabilizing twist, may influence histone occupancy and positioning in a similar way as TrmBL2. Furthermore, Alba exhibits DNA-binding properties similar to those of linker histones found in eukaryotes, which provides potential structural synergy in binding.

Another notable NAP is the poorly studied MC1, which, like histone dimers, induces kinks in the DNA. MC1 was suggested to be the main NAP in the euryarchaeal class Halobacteria and in the genus *Methanosarcina*, part of the euryarchaeal class Methanomicrobia (223, 228, 230). Although MC1 has not exclusively been found in above-mentioned class and genus, it is striking that the histones of these taxa are either predicted to not form hypernucleosomes, or to form hypernucleosomes that are likely non-essential for genome compaction and DNA organization. The hypothesis that hypernucleosome formation is in some cases non-essential, has been based on the notion that in some phyla, histones predicted to form hypernucleosomes are not equally distributed among its species, leaving some species without multimerizing histones. This suggests that MC1 plays a role in the compaction of the chromosome when hypernucleosomes are not formed. The function of MC1 in Halobacteria and Methanosarcina may be similar to that of the HU-family protein HTa, which also induces kinks in the DNA. HTa, expressed by *Thermoplasma acidophilum*, is the main NAP in a species of which the genome does not encode histones. It was hypothesized that HTa fills the architectural role that histones play in other archaea (235). Immunoprecipitation studies found that HTa is mainly located at compacted DNA (369). HTa has a histone-like nucleotide preference and is able to cover both short and long stretches of DNA, the abundance of which is growth phase dependent (235). Although not

a sequential homolog of HTa, MC1 may play a similar role as HTa. This functional conservation has been widespread through all domains, as archaea, bacteria and eukaryotes express NAPs that can bend, bridge and wrap DNA in order to alter genome architecture on a global and local level (13).

Discovery and characterization of archaeal histones

The recent emergence of metagenomic sequencing has made available an enormous amount of genomic information, and with it a large number of histone sequences. Thanks to this information, the archaeal tree of life is rapidly expanding, and as a result, the gap between archaea and eukaryotes is narrowing. This gives insight in how processes such as transcription regulation and genome compaction have evolved. The identification of the Asgard archaea has not only resulted in the discovery of some thus far eukaryote-exclusive features in archaea, but also revealed new types of archaeal histones that ask for urgent characterization. The fraction of archaeal histones that have been experimentally investigated remains small to date, but high throughput single molecule technologies and systematic sequence analysis may rapidly lead to a better understanding of how histones interact with DNA, each other, histone variants and other NAPs, and it may shed light on what structures they form. The discovery of the hypernucleosome is an important lead in terms of characterization of histones. Questions about its length, its response to environmental cues and its role in transcription regulation remain to be answered. Chromosome conformation capture techniques, which have recently revealed the architecture of the chromosome of many eukaryotes and some bacteria, will soon gain a foothold in the archaeal field. It likely is able to give new insights in how the archaeal genome is organized, and may demonstrate what the role of histones is in archaeal genome organization.

The hypernucleosome formed by archaeal histones may not only give us a better understanding of how archaea organize their genomes, but it may also make more clear what the importance of genome organization is in all domains of life. As NAPs are functionally conserved, we may find similar mechanisms and structures in other parts of life. Thus far, hypernucleosomes have not been found in eukaryotes, but as all eukaryotes express histones, they may not be exclusive to archaea. This especially applies to the organisms that are most closely related to archaea. Also, the hypernucleosome likely is an important link that explains how NAPs have evolved over time. Until now, archaea have not been proven to cause human disease, but as more information on the archaeome of humans and animals becomes rapidly available, they may be identified as the cause of health problems in the near future. An accurate understanding of archaeal transcription regulation

and genome architecture may make it a target for treatment. Also, archaea have been an important source for biocatalysts, and applications for 'extremozymes' range from flavor enhancing in the food and feed industry to ink production and the synthesis of pharmaceuticals (370). Not only extracted enzymes, but also whole cells are being used for biocatalysis (371). A better understanding of the archaeal transcription machinery may help in the design of future whole cell biocatalysts (372). Furthermore, the role of archaea on the global carbon, nitrogen and sulfur cycles is only emerging, while greenhouse gases, nitrate pollution and emission of sulfur oxides are more topical than ever. Insight into the archaeal DNA organization machinery could potentially also lead to a better comprehension of the contribution of archaea to geochemical cycles, which affect all life on earth.

7

REFERENCES

1. Zillig W (1991) Comparative biochemistry of Archaea and Bacteria. *Curr Opin Gen Dev* 1(4):544-551.
2. Eme L, Spang A, Lombard J, Stairs CW, & Ettema TJG (2017) Archaea and the origin of eukaryotes. *Nat Rev Microbiol* 15(12):711-723.
3. Williams TA, Foster PG, Cox CJ, & Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504(7479):231-236.
4. Cox CJ, Foster PG, Hirt RP, Harris SR, & Embley TM (2008) The archaebacterial origin of eukaryotes. *Proc Natl Acad Sci USA* 105(51):20356-20361.
5. Huet J, Schnabel R, Sentenac A, & Zillig W (1983) Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *EMBO J* 2(8):1291-1294.
6. Lane WJ & Darst SA (2010) Molecular evolution of multisubunit RNA polymerases: structural analysis. *J Mol Biol* 395(4):686-704.
7. Armache JP, *et al.* (2013) Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. *Nucleic Acids Res* 41(2):1284-1293.
8. Petrov AS, *et al.* (2014) Evolution of the ribosome at atomic resolution. *Proc Natl Acad Sci USA* 111(28):10251-10256.
9. Yutin N, Puigbo P, Koonin EV, & Wolf YI (2012) Phylogenomics of prokaryotic ribosomal proteins. *PLoS One* 7(5):e36972.
10. Zaremba-Niedzwiedzka K, *et al.* (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541(7637):353-358.
11. Dorman CJ (2013) Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat Rev Microbiol* 11(5):349-355.

12. Dame RT & Tark-Dame M (2016) Bacterial chromatin: converging views at different scales. *Curr Opin Cell Biol* 40:60-65.

13. Luijsterburg MS, White MF, van Driel R, & Dame RT (2008) The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* 43(6):393-418.

14. Woese CR & Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 74(11):5088-5090.

15. Woese CR, Kandler O, & Wheelis ML (1990) Towards a Natural System of Organisms - Proposal for the Domains Archaea, Bacteria, and Eucarya. *P Natl Acad Sci USA* 87(12):4576-4579.

16. Fox GE, *et al.* (1980) The Phylogeny of Prokaryotes. *Science* 209(4455):457-463.

17. Elkins JG, *et al.* (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *P Natl Acad Sci USA* 105(23):8102-8107.

18. Brochier-Armanet C, Boussau B, Gribaldo S, & Forterre P (2008) Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology* 6(3):245-252.

19. Nunoura T, *et al.* (2011) Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Research* 39(8):3204-3223.

20. Spang A, *et al.* (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551):173-179.

21. Zaremba-Niedzwiedzka K, *et al.* (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541(7637):353.

22. Probst AJ, *et al.* (2018) Differential depth distribution of microbial function and putative symbionts through sediment- hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol* 3(3):328-336.

23. Durbin AM & Teske A (2012) Archaea in organic-lean and organic-rich marine subsurface sediments: an environmental gradient reflected in distinct phylogenetic lineages. *Front Microbiol* 3.

24. Takai K & Horikoshi K (1999) Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* 152(4):1285-1297.

25. Castelle CJ, *et al.* (2015) Genomic Expansion of Domain Archaea Highlights Roles for Organisms from New Phyla in Anaerobic Carbon Cycling. *Curr Biol* 25(6):690-701.

26. Baker BJ, *et al.* (2010) Enigmatic, ultrasmall, uncultivated Archaea. *P Natl Acad Sci USA* 107(19):8806-8811.

27. Castelle CJ & Banfield JF (2018) Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Cell* 172(6):1181-1197.

28. Meng J, *et al.* (2014) Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *Isme J* 8(3):650-659.

29. Vanwonterghem I, *et al.* (2016) Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat Microbiol* 1:16170.

30. Jay ZJ, *et al.* (2018) Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat Microbiol* 3(6):732-740.

31. Kozubal MA, *et al.* (2013) Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *Isme J* 7(3):622-634.

32. Sorokin DY, *et al.* (2017) Discovery of extremely halophilic, methyl-reducing euryarchaea provides insights into the evolutionary origin of methanogenesis. *Nat Microbiol* 2:17081.

33. Lazar CS, Baker BJ, Seitz KW, & Teske AP (2017) Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. *Isme J* 11(5):1118-1129.

34. Baker BJ, *et al.* (2016) Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea (vol 1, 16002, 2016). *Nat Microbiol* 1(3).

35. Murray RG & Schleifer KH (1994) Taxonomic notes: a proposal for recording the properties of putative taxa of procaryotes. *Int J Syst Bacteriol* 44(1):174-176.

36. Murray RG & Stackebrandt E (1995) Taxonomic note: implementation of the provisional status Candidatus for incompletely described procaryotes. *Int J Syst Bacteriol* 45(1):186-187.

37. Euzeby JP, Tindall BJ, & International Committee on Systematics of P (2004) Status of strains that contravene Rules 27(3) and 30 of the Bacteriological Code. Request for an opinion. *Int J Syst Evol Microbiol* 54(Pt 1):293-301.

38. Dombrowski N, Lee JH, Williams TA, Offre P, & Spang A (2019) Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol Lett* 366(2).

39. Rinke C, *et al.* (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459):431-437.

40. Spang A, *et al.* (2013) Close encounters of the third domain: the emerging genomic view of archaeal diversity and evolution. *Archaea* 2013:202358.

41. Castelle CJ, *et al.* (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol* 25(6):690-701.

42. Saw JH, *et al.* (2015) Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. *Philos Trans R Soc Lond B Biol Sci* 370(1678):20140328.

43. Spang A, Caceres EF, & Ettema TJG (2017) Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* 357(6351).

44. Williams TA, *et al.* (2017) Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc Natl Acad Sci USA* 114(23):E4602-E4611.

45. Aouad M, *et al.* (2018) Extreme halophilic archaea derive from two distinct methanogen Class II lineages. *Mol Phylogenet Evol* 127:46-54.

46. Da Cunha V, Gaia M, Gadelle D, Nasir A, & Forterre P (2017) Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *Plos Genet* 13(6).

47. Eme L, Spang A, Lombard J, Stairs CW, & Ettema TJG (2017) Archaea and the origin of eukaryotes. *Nat Rev Microbiol* 15(12):711-723.

48. Koumandou VL, *et al.* (2013) Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit Rev Biochem Mol* 48(4):373-396.

49. Eme L, Sharpe SC, Brown MW, & Roger AJ (2014) On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks. *Csh Perspect Biol* 6(8).

50. Martin W & Koonin EV (2006) Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440(7080):41-45.

51. Shabalina SA & Koonin EV (2008) Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* 23(10):578-587.

52. Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, & Woese CR (1985) Mitochondrial Origins. *P Natl Acad Sci USA* 82(13):4443-4447.

53. Gray MW (2012) Mitochondrial Evolution. *Csh Perspect Biol* 4(9).

54. Lykidis A (2007) Comparative genomics and evolution of eukaryotic phospholipid biosynthesis. *Prog Lipid Res* 46(3-4):171-199.

55. Villanueva L, Schouten S, & Damste JSS (2017) Phylogenomic analysis of lipid biosynthetic

genes of Archaea shed light on the "lipid divide'. *Environ Microbiol* 19(1):54-69.

56. Lombard J, Lopez-Garcia P, & Moreira D (2012) An ACP-Independent Fatty Acid Synthesis Pathway in Archaea: Implications for the Origin of Phospholipids. *Mol Biol Evol* 29(11):3261-3265.
57. Dibrova DV, Galperin MY, & Mulkidjanian AY (2014) Phylogenomic reconstruction of archaeal fatty acid metabolism. *Environ Microbiol* 16(4):907-918.
58. Salcher MM, *et al.* (2019) Visualization of Loki- and Heimdallarchaeia (Asgardarchaeota) by fluorescence in situ hybridization and catalyzed reporter deposition (CARD-FISH). bioRxiv Preprint.
59. Makarova KS, Yutin N, Bell SD, & Koonin EV (2010) Evolution of diverse cell division and vesicle formation systems in Archaea. *Nature Reviews Microbiology* 8(10):731-741.
60. Yutin N & Koonin EV (2012) Archaeal origin of tubulin. *Biol Direct* 7.
61. Ettema TJG, Lindas AC, & Bernander R (2011) An actin-based cytoskeleton in archaea. *Mol Microbiol* 80(4):1052-1061.
62. Makarova KS & Koonin EV (2010) Archaeal Ubiquitin-Like Proteins: Functional Versatility and Putative Ancestral Involvement in tRNA Modification Revealed by Comparative Genomic Analysis. *Archaea-an International Microbiological Journal.*
63. Adam PS, Borrel G, Brochier-Armanet C, & Gribaldo S (2017) The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *Isme J* 11(11):2407-2425.
64. Dridi B, Raoult D, & Drancourt M (2011) Archaea as emerging organisms in complex human microbiomes. *Anaerobe* 17(2):56-63.
65. Koskinen K, *et al.* (2017) First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin. *mBio* 8(6).
66. Moissl-Eichinger C, *et al.* (2018) Archaea Are Interactive Components of Complex Microbiomes. *Trends Microbiol* 26(1):70-85.
67. Perez-Chaparro PJ, *et al.* (2014) Newly identified pathogens associated with periodontitis: a systematic review. *J Dent Res* 93(9):846-858.
68. Drancourt M, *et al.* (2017) Evidence of Archaeal Methanogens in Brain Abscess. *Clin Infect Dis* 65(1):1-5.
69. Lecours PB, *et al.* (2014) Increased Prevalence of *Methanosphaera stadtmanae* in Inflammatory Bowel Diseases. *Plos One* 9(2).
70. Brugere JF, *et al.* (2014) Archaebiotics Proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes* 5(1):5-10.
71. Huber H, *et al.* (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417(6884):63-67.
72. Waters E, *et al.* (2003) The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. *P Natl Acad Sci USA* 100(22):12984-12988.
73. Offre P, Spang A, & Schleper C (2013) Archaea in Biogeochemical Cycles. *Annu Rev Microbiol* 67:437-457.
74. Berg IA, *et al.* (2010) Autotrophic carbon fixation in archaea. *Nature Reviews Microbiology* 8(6):447-460.
75. Chen LX, *et al.* (2018) Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. *Isme J* 12(3):756-775.
76. Iverson V, *et al.* (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335(6068):587-590.
77. Liu Y & Whitman WB (2008) Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann N Y Acad Sci* 1125:171-189.

78. Knittel K & Boetius A (2009) Anaerobic oxidation of methane: progress with an unknown process. *Annu Rev Microbiol* 63:311-334.

79. Milucka J, *et al.* (2012) Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* 491(7425):541-546.

80. Beal EJ, House CH, & Orphan VJ (2009) Manganese- and iron-dependent marine methane oxidation. *Science* 325(5937):184-187.

81. Cabello P, Roldan MD, & Moreno-Vivian C (2004) Nitrate reduction and the nitrogen cycle in archaea. *Microbiology* 150(Pt 11):3527-3546.

82. Dekas AE, Poretsky RS, & Orphan VJ (2009) Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science* 326(5951):422-426.

83. Mancinelli RL & Hochstein LI (1986) The occurrence of denitrification in extremely halophilic bacteria. *FEMS Microbiol Lett* 35:55-58.

84. Duan PP, Fan CH, Zhang QQ, & Xiong ZQ (2019) Overdose fertilization induced ammonia-oxidizing archaea producing nitrous oxide in intensive vegetable fields. *Sci Total Environ* 650:1787-1794.

85. Sorokin DY, *et al.* (2012) Sulfidogenesis in hypersaline chloride-sulfate lakes of Kulunda Steppe (Altai, Russia). *FEMS Microbiol Ecol* 79(2):445-453.

86. Rawlings DE (2002) Heavy metal mining using microbes. *Annu Rev Microbiol* 56:65-91.

87. Werner F & Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology* 9(2):85-98.

88. Werner F (2013) Molecular Mechanisms of Transcription Elongation in Archaea. *Chem Rev* 113(11):8331-8349.

89. Grohmann D & Werner F (2011) Cycling through transcription with the RNA polymerase F/E (RPB4/7) complex: structure, function and evolution of archaeal RNA polymerase. *Res Microbiol* 162(1):10-18.

90. Gehring AM, Walker JE, & Santangelo TJ (2016) Transcription Regulation in Archaea. *J Bacteriol* 198(14):1906-1917.

91. Martinez-Pastor M, Tonner PD, Darnell CL, & Schmid AK (2017) Transcriptional Regulation in Archaea: From Individual Genes to Global Regulatory Networks. *Annu Rev Genet* 51:143.

92. Bell SD (2005) Archaeal transcriptional regulation - variation on a bacterial theme? *Trends Microbiol* 13(6):262-265.

93. Peeters E, Peixeiro N, & Sezonov G (2013) Cis-regulatory logic in archaeal transcription. *Biochem Soc T* 41:326-331.

94. Nagy J, *et al.* (2015) Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nat Commun* 6.

95. Schulz S, *et al.* (2016) TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *P Natl Acad Sci USA* 113(13):E1816-E1825.

96. Blombach F, Smollett KL, Grohmann D, & Werner F (2016) Molecular Mechanisms of Transcription Initiation-Structure, Function, and Evolution of TFE/TFIIE-Like Factors and Open Complex Formation. *J Mol Biol* 428(12):2592-2606.

97. Martinez-Rucobo FW, Sainsbury S, Cheung ACM, & Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO Journal* 30(7):1302-1310.

98. Sanders TJ, *et al.* (2019) TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Mol Microbiol* 111(3):784-797.

99. Grohmann D, *et al.* (2011) The Initiation Factor TFE and the Elongation Factor Spt4/5

Compete for the RNAP Clamp during Transcription Initiation and Elongation. *Mol Cell* 43(2):263-274.

100. Smollett K, Blombach F, Reichelt R, Thomm M, & Werner F (2017) A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat Microbiol* 2(5).
101. Santangelo TJ & Reeve JN (2006) Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J Mol Biol* 355(2):196-210.
102. Santangelo TJ, Cubonova L, Skinner KM, & Reeve JN (2009) Archaeal intrinsic transcription termination *in vivo*. *J Bacteriol* 191(22):7102-7108.
103. Maier LK & Marchfelder A (2019) It's all about the T: transcription termination in archaea. *Biochem Soc Trans* 47(1):461-468.
104. Walker JE, Luyties O, & Santangelo TJ (2017) Factor-dependent archaeal transcription termination. *P Natl Acad Sci USA* 114(33):E6767-E6773.
105. Karr EA (2014) Transcription regulation in the third domain. *Adv Appl Microbiol* 89:101-133.
106. Ouhammouch M & Geiduschek EP (2005) An expanding family of archaeal transcriptional activators. *Proc Natl Acad Sci USA* 102(43):15423-15428.
107. Peeters E, Thia-Toong TL, Gigot D, Maes D, & Charlier D (2004) Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Mol Microbiol* 54(2):321-336.
108. Weidenbach K, Ehlers C, Kock J, & Schmitz RA (2010) NrpRII mediates contacts between NrpRI and general transcription factors in the archaeon *Methanosarcina mazei* Go1. *FEBS J* 277(21):4398-4411.
109. Koide T, *et al.* (2009) Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol Syst Biol* 5.
110. Lee SJ, *et al.* (2005) TrmB, a sugar sensing regulator of ABC transporter genes in *Pyrococcus furiosus* exhibits dual promoter specificity and is controlled by different inducers. *Mol Microbiol* 57(6):1797-1807.
111. Todor H, Sharma K, Pittman AMC, & Schmid AK (2013) Protein-DNA binding dynamics predict transcriptional response to nutrients in archaea. *Nucleic Acids Research* 41(18):8546-8558.
112. Perez-Rueda E & Janga SC (2010) Identification and Genomic Analysis of Transcription Factors in Archaeal Genomes Exemplifies Their Functional Architecture and Evolutionary Origin. *Mol Biol Evol* 27(6):1449-1459.
113. Baliga NS, *et al.* (2000) Is gene expression in Halobacterium NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol Microbiol* 36(5):1184-1185.
114. Babski J, *et al.* (2014) Small regulatory RNAs in Archaea. *RNA Biol* 11(5):484-493.
115. Maupin-Furlow JA (2013) Archaeal proteasomes and sampylation. *Subcell Biochem* 66:297-327.
116. Visone V, *et al.* (2014) Chromatin Structure and Dynamics in Hot Environments: Architectural Proteins and DNA Topoisomerases of Thermophilic Archaea. *Int J Mol Sci* 15(9):17162-17187.
117. Peeters E, Driessen RPC, Werner F, & Dame RT (2015) The interplay between nucleoid organization and transcription in archaeal genomes. *Nature Reviews Microbiology* 13(6):333-341.
118. Laurens N, *et al.* (2012) Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat Commun* 3.
119. Bell SD, Botting CH, Wardleworth BN, Jackson SP, & White MF (2002) The interaction of

Alba, a conserved archaeal, chromatin protein, with Sir2 and its regulation by acetylation. *Science* 296(5565):148-151.

120. Cao JJ, Wang Q, Liu T, Peng N, & Huang L (2018) Insights into the post-translational modifications of archaeal Sis10b (Alba): lysine-16 is methylated, not acetylated, and this does not regulate transcription or growth. *Mol Microbiol* 109(2):192-208.
121. van der Valk RA, Vreede J, Cremazy F, & Dame RT (2014) Genomic looping: a key principle of chromatin organization. *J Mol Microbiol Biotechnol* 24(5-6):344-359.
122. Liu Y, Chen H, Kenney LJ, & Yan J (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev* 24(4):339-344.
123. van der Valk RA, *et al.* (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-bridging activity. *eLife* 6.
124. Kalichuk V, *et al.* (2016) The archaeal "7 kDa DNA-binding" proteins: extended characterization of an old gifted family. *Sci Rep-Uk* 6.
125. Driessen RP, *et al.* (2013) Crenarchaeal chromatin proteins Cren7 and Sul7 compact DNA by inducing rigid bends. *Nucleic Acids Res* 41(1):196-205.
126. Hsu CH & Wang AH (2011) The DNA-recognition fold of Sso7c4 suggests a new member of SpoVT-AbrB superfamily from archaea. *Nucleic Acids Res* 39(15):6764-6774.
127. Driessen RP, *et al.* (2016) Diverse architectural properties of Sso10a proteins: Evidence for a role in chromatin compaction and organization. *Sci Rep* 6:29422.
128. Luo X, *et al.* (2007) CC1, a novel crenarchaeal DNA binding protein. *J Bacteriol* 189(2):403-409.
129. Le Cam E, Culard F, Larquet E, Delain E, & Cognet JAH (1999) DNA bending induced by the archaebacterial histone-like protein MC1. *Journal of Molecular Biology* 285(3):1011-1021.
130. Bayer B, *et al.* (2016) Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *Isme J* 10(5):1051-1063.
131. Bridger SL, Lancaster WA, Poole FL, Schut GJ, & Adams MWW (2012) Genome Sequencing of a Genetically Tractable *Pyrococcus furiosus* Strain Reveals a Highly Dynamic Genome. *J Bacteriol* 194(15):4097-4106.
132. Cai SF, *et al.* (2012) Identification of the Haloarchaeal Phasin (PhaP) That Functions in Polyhydroxyalkanoate Accumulation and Granule Formation in *Haloferax mediterranei*. *Appl Environ Microb* 78(6):1946-1952.
133. Kamashev D, *et al.* (2017) Comparison of histone-like HU protein DNA-binding properties and HU/IHF protein sequence alignment. *Plos One* 12(11).
134. Henneman B, van Emmerik C, van Ingen H, & Dame RT (2018) Structure and function of archaeal histones. *Plos Genet* 14(9):e1007582.
135. Luger K, Mader AW, Richmond RK, Sargent DF, & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389(6648):251-260.
136. Kornberg RD (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* 184(4139):868-871.
137. Decanniere K, Babu AM, Sandman K, Reeve JN, & Heinemann U (2000) Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J Mol Biol* 303(1):35-47.
138. Sandman K & Reeve JN (2006) Archaeal histones and the origin of the histone fold. *Curr Opin Microbiol* 9(5):520-525.
139. Arents G, Burlingame RW, Wang BC, Love WE, & Moudrianakis EN (1991) The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci USA* 88(22):10148-10152.

140. Henneman B & Dame RT (2015) Archaeal histones: dynamic and versatile genome architects. *Aims Microbiol* 1(1):72-81.

141. Malik HS & Henikoff S (2003) Phylogenomics of the nucleosome. *Nat Struct Biol* 10(11):882-891.

142. Katan AJ, Vlijm R, Lusser A, & Dekker C (2015) Dynamics of nucleosomal structures measured by high-speed atomic force microscopy. *Small* 11(8):976-984.

143. Levchenko V, Jackson B, & Jackson V (2005) Histone release during transcription: displacement of the two H2A-H2B dimers in the nucleosome is dependent on different levels of transcription-induced positive stress. *Biochemistry* 44(14):5357-5372.

144. Hamiche A, *et al.* (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 USA* 93(15):7588-7593.

145. Arimura Y, Tachiwana H, Oda T, Sato M, & Kurumizaka H (2012) Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. *Biochemistry* 51(15):3302-3309.

146. Bednar J, *et al.* (2017) Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. *Mol Cell* 66(3):384-397 e388.

147. Zhou BR, *et al.* (2013) Structural insights into the histone H1-nucleosome complex. *P Natl Acad Sci USA* 110(48):19390-19395.

148. Cutter AR & Hayes JJ (2017) Linker histones: novel insights into structure-specific recognition of the nucleosome. *Biochem Cell Biol* 95(2):171-178.

149. Robinson PJ & Rhodes D (2006) Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* 16(3):336-343.

150. Shogren-Knaak M, *et al.* (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311(5762):844-847.

151. Pepenella S, Murphy KJ, & Hayes JJ (2014) A distinct switch in interactions of the histone H4 tail domain upon salt-dependent folding of nucleosome arrays. *J Biol Chem* 289(39):27342-27351.

152. Zhang R, Erler J, & Langowski J (2017) Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction. *Biophys J* 112(3):450-459.

153. Zhao Y & Garcia BA (2015) Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* 7(9):a025064.

154. Jenuwein T & Allis CD (2001) Translating the histone code. *Science* 293(5532):1074-1080.

155. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128(4):693-705.

156. Kalashnikova AA, Porter-Goff ME, Muthurajan UM, Luger K, & Hansen JC (2013) The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J R Soc Interface* 10(82):20121022.

157. Zhou J, Fan JY, Rangasamy D, & Tremethick DJ (2007) The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression. *Nat Struct Mol Biol* 14(11):1070-1076.

158. Dorigo B, Schalch T, Bystricky K, & Richmond TJ (2003) Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J Mol Biol* 327(1):85-96.

159. Peeters E, Driessen RPC, Werner F, & Dame RT (2015) The interplay between nucleoid organization and transcription in archaeal genomes. *Nature Rev Microbiol* 13(6):333-341.

160. Bailey KA, Marc F, Sandman K, & Reeve JN (2002) Both DNA and histone fold sequences contribute to archaeal nucleosome stability. *J Biol Chem* 277(11):9293-9301.

161. Luger K, Dechassa ML, & Tremethick DJ (2012) New insights into nucleosome and chroma-

tin structure: an ordered state or a disordered affair? Nature reviews. *Molecular cell biology* 13(7):436-447.

162. Pereira SL, Grayling RA, Lurz R, & Reeve JN (1997) Archaeal nucleosomes. *Proc Natl Acad Sci USA* 94(23):12633-12637.

163. Pereira SL & Reeve JN (1999) Archaeal nucleosome positioning sequence from *Methanothermus fervidus*. *J Mol Biol* 289(4):675-681.

164. Bailey KA, Pereira SL, Widom J, & Reeve JN (2000) Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. *J Mol Biol* 303(1):25-34.

165. Maruyama H, *et al.* (2013) An alternative beads-on-a-string chromatin architecture in *Thermococcus kodakarensis*. *EMBO Rep* 14(8):711-717.

166. Xie Y & Reeve JN (2004) Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J Bact* 186(11):3492-3498.

167. Sandman K & Reeve JN (2006) Archaeal histones and the origin of the histone fold. *Curr Op Microbiol* 9(5):520-525.

168. Li WT, Sandman K, Pereira SL, & Reeve JN (2000) MJ1647, an open reading frame in the genome of the hyperthermophile *Methanococcus jannaschii*, encodes a very thermostable archaeal histone with a C-terminal extension. *Extremophiles* 4(1):43-51.

169. Hartman AL, *et al.* (2010) The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* 5(3):e9605.

170. Fahrner RL, Cascio D, Lake JA, & Slesarev A (2001) An ancestral nuclear protein assembly: crystal structure of the *Methanopyrus kandleri* histone. *Prot Science* 10(10):2002-2007.

171. Marc F, Sandman K, Lurz R, & Reeve JN (2002) Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J Biol Chem* 277(34):30879-30886.

172. Ammar R, *et al.* (2012) Chromatin is an ancient innovation conserved between Archaea and Eukarya. *eLife* 2012(1):1-11.

173. Warnecke T, Becker EA, Facciotti MT, Nislow C, & Lehner B (2013) Conserved Substitution Patterns around Nucleosome Footprints in Eukaryotes and Archaea Derive from Frequent Nucleosome Repositioning through Evolution. *PLoS Comp Biol* 9(11):9-14.

174. Kaplan N, *et al.* (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458(7236):362-366.

175. Struhl K & Segal E (2013) Determinants of nucleosome positioning. *Nat Struct Mol Biol* 20(3):267-273.

176. Das C & Tyler JK (2013) Histone exchange and histone modifications during transcription and aging. *Biochim Biophys Acta* 1819:332-342.

177. Sandman K & Reeve JN (2005) Archaeal chromatin proteins: different structures but common function? *Curr Opin Microbiol* 8(6):656-661.

178. Sandman K, Krzycki JA, Dobrinski B, Lurz R, & Reeve JN (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc Natl Acad Sci USA* 87(15):5788-5791.

179. Dame RT (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol Microbiol* 56(4):858-870.

180. Hofmann A & Heermann DW (2015) The role of loops on the order of eukaryotes and prokaryotes. *FEBS letters* 589(20 Pt A):2958-2965.

181. Arents G & Moudrianakis EN (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci USA* 92(24):11170-11174.

182. Malik HS & Henikoff S (2003) Phylogenomics of the nucleosome. *Nat Struct Biol* 10(11):882-891.

183. Sandman K, Pereira SL, & Reeve JN (1998) Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell Mol Life Sci* 54(12):1350-1364.

184. Ng WV, *et al.* (2000) Genome sequence of Halobacterium species NRC-1. *Proc Natl Acad Sci USA* 97(22):12176-12181.

185. Dulmage KA, Todor H, & Schmid AK (2015) Growth-Phase-Specific Modulation of Cell Morphology and Gene Expression by an Archaeal Histone Protein. *mBio* 6(5):e00649-00615.

186. Becker EA, *et al.* (2014) Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. *Plos Genet* 10(11):e1004784.

187. Stoltzfus A (1999) On the possibility of constructive neutral evolution. *J Mol Evol* 49(2):169-181.

188. Lynch M & Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154(1):459-473.

189. Ausio J (2006) Histone variants--the structure behind the function. *Brief Funct Genomic Proteomic* 5(3):228-243.

190. Marzluff WF & Duronio RJ (2002) Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr Opin Cell Biol* 14(6):692-699.

191. Weber CM & Henikoff S (2014) Histone variants: dynamic punctuation in transcription. *Genes Dev* 28(7):672-682.

192. Henikoff S & Smith MM (2015) Histone variants and epigenetics. *Cold Spring Harb Perspect Biol* 7(1):a019364.

193. Sandman K, Krzycki JA, Dobrinski B, Lurz R, & Reeve JN (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc Natl Acad Sci USA* 87(15):5788-5791.

194. Nishida H & Oshima T (2017) Archaeal histone distribution is associated with archaeal genome base composition. *J Gen Appl Microbiol* 63(1):28-35.

195. van der Valk RA, Laurens N, & Dame RT (2017) Tethered Particle Motion Analysis of the DNA Binding Properties of Architectural Proteins. *Methods Mol Biol* 1624:127-143.

196. Maruyama H, *et al.* (2013) An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. *EMBO Reports* 14(8):711-717.

197. Mattiroli F, *et al.* (2017) Structure of histone-based chromatin in Archaea. *Science* 357(6351):609-611.

198. Fricke WF, *et al.* (2006) The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H2 for methane formation and ATP synthesis. *J Bacteriol* 188(2):642-658.

199. Hoedt EC, *et al.* (2018) Culture- and metagenomics-enabled analyses of the Methanosphaera genus reveals their monophyletic origin and differentiation according to genome size. *Isme J* 12(12):2942-2953.

200. Xue H, Guo R, Wen YF, Liu DX, & Huang L (2000) An abundant DNA binding protein from the hyperthermophilic archaeon *Sulfolobus shibatae* affects DNA supercoiling in a temperature-dependent fashion. *J Bacteriol* 182(14):3929-3933.

201. Driessen RPC & Dame RT (2011) Nucleoid-associated proteins in Crenarchaea. *Biochem Soc T* 39:116-121.

202. Driessen RP & Dame RT (2013) Structure and dynamics of the crenarchaeal nucleoid. *Biochem Soc Trans* 41(1):321-325.

203. White MF & Bell SD (2002) Holding it together: chromatin in the Archaea. *Trends Genet*

18(12):621-626.

204. O'Neill LP & Turner BM (1995) Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *EMBO J* 14(16):3946-3957.

205. Reeve JN (2003) Archaeal chromatin and transcription. *Mol Microbiol* 48(3):587-598.

206. Beltrao P, Bork P, Krogan NJ, & van Noort V (2013) Evolution and functional cross-talk of protein post-translational modifications. *Mol Syst Biol* 9.

207. Eichler J & Adams MWW (2005) Posttranslational protein modification in Archaea (vol 69, 393, 2005). *Microbiol Mol Biol R* 69(4):696-696.

208. Sigrist CJA, *et al.* (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research* 38:D161-D166.

209. Wu CH, *et al.* (2003) The Protein Information Resource. *Nucleic Acids Research* 31(1):345-347.

210. Berger SL (1999) Gene activation by histone and factor acetyltransferases. *Current Opinion in Cell Biology* 11(3):336-341.

211. Li YY, *et al.* (2014) AF9 YEATS Domain Links Histone Acetylation to DOT1L-Mediated H3K79 Methylation. *Cell* 159(3):558-571.

212. Shanle EK, *et al.* (2015) Association of Taf14 with acetylated histone H3 directs gene transcription and the DNA damage response. *Gene Dev* 29(17):1795-1800.

213. Andrews FH, *et al.* (2016) The Taf14 YEATS domain is a reader of histone crotonylation. *Nat Chem Biol* 12(6):396-U333.

214. Li YY, *et al.* (2016) Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol Cell* 62(2):181-193.

215. Zhao D, *et al.* (2016) YEATS2 is a selective histone crotonylation reader. *Cell Res* 26(5):629-632.

216. Chavez MS, *et al.* (2012) The conformational flexibility of the C-terminus of histone H4 promotes histone octamer and nucleosome stability and yeast viability. *Epigenetics Chromatin* 5(1):5.

217. Decanniere K, Babu AM, Sandman K, Reeve JN, & Heinemann U (2000) Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *Journal of Molecular Biology* 303(1):35-47.

218. Fahrner RL, Cascio D, Lake JA, & Slesarev A (2001) An ancestral nuclear protein assembly: Crystal structure of the *Methanopyrus kandleri* histone. *Protein Science* 10(10):2002-2007.

219. Wilkins MR, *et al.* (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531-552.

220. Ammar R, *et al.* (2012) Chromatin is an ancient innovation conserved between Archaea and Eukarya. eLife 1:e00078.

221. Nalabothula N, *et al.* (2013) Archaeal nucleosome positioning *in vivo* and *in vitro* is directed by primary sequence motifs. *Bmc Genomics* 14:391.

222. Toulme F, *et al.* (1995) Conformational changes of DNA minicircles upon the binding of the archaebacterial histone-like protein MC1. *J Biol Chem* 270(11):6286-6291.

223. De Vuyst G, Aci S, Genest D, & Culard F (2005) Atypical recognition of particular DNA sequences by the archaeal chromosomal MC1 protein. *Biochemistry* 44(30):10369-10377.

224. Crooks GE, Hon G, Chandonia JM, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188-1190.

225. Dombrowski N, Teske AP, & Baker BJ (2018) Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat Commun* 9(1):4999.

226. Sandman K, Louvel H, Samson RY, Pereira SL, & Reeve JN (2008) Archaeal chromatin proteins histone HMtB and Alba have lost DNA-binding ability in laboratory strains of *Methanothermobacter thermautotrophicus*. *Extremophiles* 12(6):811-817.

227. Drozdetskiy A, Cole C, Procter J, & Barton GJ (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43(W1):W389-394.

228. Weidenbach K, *et al.* (2008) Deletion of the archaeal histone in *Methanosarcina mazei* Go1 results in reduced growth and genomic transcription. *Mol Microbiol* 67(3):662-671.

229. Chartier F, Laine B, & Sautiere P (1988) Characterization of the chromosomal protein MC1 from the thermophilic archaebacterium *Methanosarcina* sp. CHTI 55 and its effect on the thermal stability of DNA. *Biochim Biophys Acta* 951(1):149-156.

230. Chartier F, Laine B, Sautiere P, Touzel JP, & Albagnac G (1985) Characterization of the Chromosomal Protein Hmb Isolated from *Methanosarcina-Barkeri*. *FEBS letters* 183(1):119-123.

231. Fahrner RL, Cascio D, Lake JA, & Slesarev A (2001) An ancestral nuclear protein assembly: crystal structure of the *Methanopyrus kandleri* histone. *Protein Sci* 10(10):2002-2007.

232. Pavlov NA, Cherny DI, Jovin TM, & Slesarev AI (2002) Nucleosome-like complex of the histone from the hyperthermophile *Methanopyrus kandleri* (MkaH) with linear DNA. *J Biomol Struct Dyn* 20(2):207-214.

233. Soares DJ, Sandman K, & Reeve JN (2000) Mutational analysis of archaeal histone-DNA interactions. *Journal of Molecular Biology* 297(1):39-47.

234. DeLange RJ, Williams LC, & Searcy DG (1981) A histone-like protein (HTa) from *Thermoplasma acidophilum*. II. Complete amino acid sequence. *J Biol Chem* 256(2):905-911.

235. Hocher A, Rojec M, Swadling JB, Esin A, & Warnecke T (2019) The DNA-binding protein HTa from *Thermoplasma acidophilum* is an archaeal histone analog. bioRxiv Preprint.

236. Sun N, *et al.* (2010) Quantitative proteome and transcriptome analysis of the archaeon *Thermoplasma acidophilum* cultured under aerobic and anaerobic conditions. *J Proteome Res* 9(9):4839-4850.

237. Wilkinson SP, Ouhammouch M, & Geiduschek EP (2010) Transcriptional activation in the context of repression mediated by archaeal histones. *Proc Natl Acad Sci USA* 107(15):6777-6781.

238. Xie Y & Reeve JN (2004) Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J Bacteriol* 186(11):3492-3498.

239. Ueguchi C, Kakeda M, & Mizuno T (1993) Autoregulatory expression of the *Escherichia coli* hns gene encoding a nucleoid protein: H-NS functions as a repressor of its own transcription. *Mol Gen Genet* 236(2-3):171-178.

240. Dorman CJ (2004) H-NS: A universal regulator for a dynamic genome. *Nature Reviews Microbiology* 2(5):391-400.

241. Baek JH, Rajagopala SV, & Chattoraj DK (2014) Chromosome Segregation Proteins of *Vibrio cholerae* as Transcription Regulators. *mBio* 5(3).

242. Ringgaard S, Ebersbach G, Borch J, & Gerdes K (2007) Regulatory cross-talk in the double par locus of plasmid pB171E. *Journal of Biological Chemistry* 282(5):3134-3145.

243. Breier AM & Grossman AD (2007) Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the *Bacillus subtilis* chromosome. *Mol Microbiol* 64(3):703-718.

244. Pratto F, *et al.* (2008) Streptococcus pyogenes pSM19035 requires dynamic assembly of ATP-bound ParA and ParB on parS DNA during plasmid segregation. *Nucleic Acids Research* 36(11):3676-3689.

245. Lang B, *et al.* (2007) High-affinity DNA binding sites for H-NS provide a molecular basis for

selective silencing within proteobacterial genomes. *Nucleic Acids Res* 35(18):6330-6337.

246. Tompitak M, Vaillant C, & Schiessel H (2017) Genomes of Multicellular Organisms Have Evolved to Attract Nucleosomes to Promoter Regions. *Biophys J* 112(3):505-511.
247. Jelinska C, Petrovic-Stojanovska B, Ingledew WJ, & White MF (2010) Dimer-dimer stacking interactions are important for nucleic acid binding by the archaeal chromatin protein Alba. *Biochem J* 427(1):49-55.
248. Zhao KH, Chai XM, & Marmorstein R (2003) Structure of a Sir2 substrate, Alba, reveals a mechanism for deacetylation-induced enhancement of DNA binding. *Journal of Biological Chemistry* 278(28):26071-26077.
249. Grohmann D & Werner F (2011) Recent advances in the understanding of archaeal transcription. *Current Opinion in Microbiology* 14(3):328-334.
250. Sheppard C & Werner F (2017) Structure and mechanisms of viral transcription factors in archaea. *Extremophiles* 21(5):829-838.
251. Sandman K, Grayling RA, Dobrinski B, Lurz R, & Reeve JN (1994) Growth-phase-dependent synthesis of histones in the archaeon *Methanothermus fervidus*. *Proc Natl Acad Sci USA* 91(26):12624-12628.
252. Soares DJ, Marc F, & Reeve JN (2003) Conserved eukaryotic histone-fold residues substituted into an archaeal histone increase DNA affinity but reduce complex flexibility. *J Bacteriol* 185(11):3453-3457.
253. Sievers F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7.
254. Webb B & Sali A (2014) Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* 47:5 6 1-32.
255. Dominguez C, Boelens R, & Bonvin AMJJ (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125(7):1731-1737.
256. Hellman LM & Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* 2(8):1849-1861.
257. Gilman AG (1970) A protein binding assay for adenosine 3':5'-cyclic monophosphate. *Proc Natl Acad Sci USA* 67(1):305-312.
258. Heyduk T, Ma Y, Tang H, & Ebright RH (1996) Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. *Methods Enzymol* 274:492-503.
259. Ziegler A & Seelig J (2007) High affinity of the cell-penetrating peptide HIV-1 Tat-PTD for DNA. *Biochemistry* 46(27):8138-8145.
260. Hart DJ, Speight RE, Cooper MA, Sutherland JD, & Blackburn JM (1999) The salt dependence of DNA recognition by NF-kappaB p50: a detailed kinetic analysis of the effects on affinityand specificity. *Nucleic Acids Res* 27(4):1063-1069.
261. Jerabek-Willemsen M, Wienken CJ, Braun D, Baaske P, & Duhr S (2011) Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev Technol* 9(4):342-353.
262. Yang Y, Sass LE, Du C, Hsieh P, & Erie DA (2005) Determination of protein-DNA binding constants and specificities from statistical analyses of single molecules: MutS-DNA interactions. *Nucleic Acids Res* 33(13):4322-4334.
263. McCauley MJ & Williams MC (2011) Measuring DNA-protein binding affinity on a single molecule using optical tweezers. *Methods Mol Biol* 749:305-315.
264. Coats JE, Lin Y, Rueter E, Maher LJ, 3rd, & Rasnik I (2013) Single-molecule FRET analysis of DNA binding and bending by yeast HMGB protein Nhp6A. *Nucleic Acids Res* 41(2):1372-1381.

265. Schafer DA, Gelles J, Sheetz MP, & Landick R (1991) Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature* 352(6334):444-448.

266. Kulic IM, Mohrbach H, Thaokar R, & Schiessel H (2007) Equation of state of looped DNA. *Phys Rev E Stat Nonlin Soft Matter Phys* 75(1 Pt 1):011913.

267. Chintakayala K, *et al.* (2015) DNA recognition by *Escherichia coli* CbpA protein requires a conserved arginine-minor-groove interaction. *Nucleic Acids Res* 43(4):2282-2292.

268. Wang H, Yehoshua S, Ali SS, Navarre WW, & Milstein JN (2014) A biomechanical mechanism for initiating DNA packaging. *Nucleic Acids Res* 42(19):11921-11927.

269. Mack AH, Schlingman DJ, Salinas RD, Regan L, & Mochrie SG (2015) Condensation transition and forced unravelling of DNA-histone H1 toroids: a multi-state free energy landscape. *J Phys Condens Matter* 27(6):064106.

270. Wu HY, Lu CH, & Li HW (2017) RecA-SSB Interaction Modulates RecA Nucleoprotein Filament Formation on SSB-Wrapped DNA. *Sci Rep* 7(1):11876.

271. Lu CH & Li HW (2017) DNA with Different Local Torsional States Affects RecA-Mediated Recombination Progression. *Chemphyschem* 18(6):584-590.

272. Fan HF, Cheng YS, Ma CH, & Jayaram M (2015) Single molecule TPM analysis of the catalytic pentad mutants of Cre and Flp site-specific recombinases: contributions of the pentad residues to the pre-chemical steps of recombination. *Nucleic Acids Res* 43(6):3237-3255.

273. Fan HF, Hsieh TS, Ma CH, & Jayaram M (2016) Single-molecule analysis of varphiC31 integrase-mediated site-specific recombination by tethered particle motion. *Nucleic Acids Res* 44(22):10804-10823.

274. Chen YF, *et al.* (2016) Modulation of yeast telomerase activity by Cdc13 and Est1 *in vitro*. *Sci Rep* 6:34104.

275. Mangan MW, *et al.* (2006) The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium. *Mol Microbiol* 59(6):1831-1847.

276. Peacock S, Weissbach H, & Nash HA (1984) *In vitro* regulation of phage lambda cII gene expression by *Escherichia coli* integration host factor. *Proc Natl Acad Sci USA* 81(19):6009-6013.

277. Goosen N & van de Putte P (1995) The regulation of transcription initiation by integration host factor. *Mol Microbiol* 16(1):1-7.

278. Khodr A, Fairweather V, Bouffartigues E, & Rimsky S (2015) IHF is a trans-acting factor implicated in the regulation of the proU P2 promoter. *FEMS Microbiol Lett* 362(3):1-6.

279. Wang S, Cosstick R, Gardner JF, & Gumport RI (1995) The specific binding of *Escherichia coli* integration host factor involves both major and minor grooves of DNA. *Biochemistry* 34(40):13082-13090.

280. Swinger KK & Rice PA (2004) IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* 14(1):28-35.

281. Rice PA, Yang S, Mizuuchi K, & Nash HA (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87(7):1295-1306.

282. Travers A (1997) DNA-protein interactions: IHF--the master bender. *Curr Biol* 7(4):R252-254.

283. Dame RT, *et al.* (2005) Analysis of scanning force microscopy images of protein-induced DNA bending using simulations. *Nucleic Acids Res* 33(7):e68.

284. Yang SW & Nash HA (1995) Comparison of protein binding to DNA *in vivo* and *in vitro*: defining an effective intracellular target. *EMBO J* 14(24):6292-6300.

285. van Noort J, Verbrugge S, Goosen N, Dekker C, & Dame RT (2004) Dual architectural

roles of HU: formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci USA* 101(18):6969-6974.

286. Giphart-Gassler M, Goosen T, van Meeteren A, Wijffelman C, & van de Putte P (1979) Properties of the recombinant plasmid pGP1 containing part of the early region of bacteriophage Mu. *Cold Spring Harb Symp Quant Biol* 43 Pt 2:1179-1185.
287. Holbrook JA, Tsodikov OV, Saecker RM, & Record MT, Jr. (2001) Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. *J Mol Biol* 310(2):379-401.
288. Sitters G, *et al.* (2015) Acoustic force spectroscopy. Nat Methods 12(1):47-50.
289. Chenouard N, *et al.* (2014) Objective comparison of particle tracking methods. *Nat Methods* 11(3):281-289.
290. Rogers SS, Waigh TA, Zhao X, & Lu JR (2007) Precise particle tracking against a complicated background: polynomial fitting with Gaussian weight. *Phys Biol* 4(3):220-227.
291. Visser EW, van ILJ, & Prins MW (2016) Particle Motion Analysis Reveals Nanoscale Bond Characteristics and Enhances Dynamic Range for Biosensing. *ACS Nano* 10(3):3093-3101.
292. Driessen RP, *et al.* (2014) Effect of temperature on the intrinsic flexibility of DNA and its interaction with architectural proteins. *Biochemistry* 53(41):6430-6438.
293. Bowman GD & Poirier MG (2015) Post-translational modifications of histones that influence nucleosome dynamics. *Chem Rev* 115(6):2274-2295.
294. Cox CJ, Foster PG, Hirt RP, Harris SR, & Embley TM (2008) The archaebacterial origin of eukaryotes. *P Natl Acad Sci USA* 105(51):20356-20361.
295. Peeters E, Driessen RP, Werner F, & Dame RT (2015) The interplay between nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol* 13(6):333-341.
296. Bult CJ, *et al.* (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* 273(5278):1058-1073.
297. Darcy TJ, Sandman K, & Reeve JN (1995) *Methanobacterium formicicum*, a mesophilic methanogen, contains three HFo histones. *J Bacteriol* 177(3):858-860.
298. Tabassum R, Sandman KM, & Reeve JN (1992) HMt, a histone-related protein from *Methanobacterium thermoautotrophicum* delta H. *J Bacteriol* 174(24):7890-7895.
299. Fukui T, *et al.* (2005) Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with Pyrococcus genomes. *Genome Res* 15(3):352-363.
300. Anderson I, *et al.* (2010) Complete genome sequence of *Methanothermus fervidus* type strain (V24S). *Stand Genomic Sci* 3(3):315-324.
301. Musgrave DR, Sandman KM, & Reeve JN (1991) DNA binding by the archaeal histone HMf results in positive supercoiling. *Proc Natl Acad Sci USA* 88(23):10397-10401.
302. Grayling RA, Sandman K, & Reeve JN (1996) DNA stability and DNA binding proteins. *Adv Protein Chem* 48:437-467.
303. Howard MT, Sandman K, Reeve JN, & Griffith JD (1992) HMf, a histone-related protein from the hyperthermophilic archaeon *Methanothermus fervidus*, binds preferentially to DNA containing phased tracts of adenines. *J Bacteriol* 174(23):7864-7867.
304. Efremov AK, *et al.* (2015) Transcriptional Repressor TrmBL2 from *Thermococcus kodakarensis* Forms Filamentous Nucleoprotein Structures and Competes with Histones for DNA Binding in a Salt- and DNA Supercoiling-dependent Manner. *J Biol Chem* 290(25):15770-15784.
305. Musgrave D, Forterre P, & Slesarev A (2000) Negative constrained DNA supercoiling in archaeal nucleosomes. *Mol Microbiol* 35(2):341-349.

306. Tomschik M, Karymov MA, Zlatanova J, & Leuba SH (2001) The archaeal histone-fold protein HMf organizes DNA into bona fide chromatin fibers. *Structure* 9(12):1201-1211.

307. Feng Y, Yao H, & Wang J (2010) Crystal structure of the crenarchaeal conserved chromatin protein Cren7 and double-stranded DNA complex. *Protein Sci* 19(6):1253-1257.

308. Chen L, *et al.* (2004) The hyperthermophile protein Sso10a is a dimer of winged helix DNA-binding domains linked by an antiparallel coiled coil rod. *J Mol Biol* 341(1):73-91.

309. Kulic IM, Mohrbach H, Thaokar R, & Schiessel H (2007) Equation of state of looped DNA. *Phys Rev E* 75(1).

310. Marko JF & Siggia ED (1995) Stretching DNA. *Macromolecules* 28(26):8759-8770.

311. van Zundert GCP, *et al.* (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* 428(4):720-725.

312. Soares DJ, Sandman K, & Reeve JN (2000) Mutational analysis of archaeal histone-DNA interactions. *J Mol Biol* 297(1):39-47.

313. Sandman K & Reeve JN (2001) Chromosome packaging by archaeal histones. *Adv Appl Microbiol* 50:75-99.

314. Lipfert J, Klijnhout S, & Dekker NH (2010) Torsional sensing of small-molecule binding using magnetic tweezers. *Nucleic Acids Res* 38(20):7122-7132.

315. Meng H, Andresen K, & van Noort J (2015) Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibers. *Nucleic Acids Res* 43(7):3578-3590.

316. Kaczmarczyk A, *et al.* (2017) Single-molecule force spectroscopy on histone H4 tail-cross-linked chromatin reveals fiber folding. *Journal of Biological Chemistry* 292(42):17506-17513.

317. Sanders TJ, Marshall CJ, & Santangelo TJ (2019) The Role of Archaeal Chromatin in Transcription. *J Mol Biol*.

318. Nishida H & Oshima T (2017) Archaeal histone distribution is associated with archaeal genome base composition. *Journal of General and Applied Microbiology* 63(1):28-35.

319. Pollock WB, Rosell FI, Twitchett MB, Dumont ME, & Mauk AG (1998) Bacterial expression of a mitochondrial cytochrome c. Trimethylation of lys72 in yeast iso-1-cytochrome c and the alkaline conformational transition. *Biochemistry* 37(17):6124-6131.

320. Lowary PT & Widom J (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* 276(1):19-42.

321. Brouwer TB, Kaczmarczyk A, Pham C, & van Noort J (2018) Unraveling DNA Organization with Single-Molecule Force Spectroscopy Using Magnetic Tweezers. *Methods Mol Biol* 1837:317-349.

322. Henneman B, Heinsman J, Battjes J, & Dame RT (2018) Quantitation of DNA-Binding Affinity Using Tethered Particle Motion. *Methods Mol Biol* 1837:257-275.

323. Strick TR, Allemand JF, Croquette V, & Bensimon D (1998) Physical approaches to the study of DNA. *J Stat Phys* 93(3-4):647-672.

324. Strick T, Allemand JF, Croquette V, & Bensimon D (2000) Twisting and stretching single DNA molecules. *Prog Biophys Mol Bio* 74(1-2):115-140.

325. Meng H, Andresen K, & van Noort J (2015) Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibers. *Nucleic Acids Research* 43(7):3578-3590.

326. Pereira SL, Grayling RA, Lurz R, & Reeve JN (1997) Archaeal nucleosomes. *P Natl Acad Sci USA* 94(23):12633-12637.

327. Kulic IM & Schiessel H (2004) DNA spools under tension. *Phys Rev Lett* 92(22):228101.

328. Bustamante C, Bryant Z, & Smith SB (2003) Ten years of tension: single-molecule DNA mechanics. *Nature* 421(6921):423-427.

329. Narasingarao P, *et al.* (2012) De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *Isme J* 6(1):81-93.

330. Ghai R, *et al.* (2011) New abundant microbial groups in aquatic hypersaline environments. *Sci Rep* 1:135.

331. Crits-Christoph A, *et al.* (2016) Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. *Environ Microbiol* 18(6):2064-2077.

332. Finstad KM, *et al.* (2017) Microbial Community Structure and the Persistence of Cyanobacterial Populations in Salt Crusts of the Hyperarid Atacama Desert from Genome-Resolved Metagenomics. *Front Microbiol* 8:1435.

333. Kozlowski LP (2017) Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res* 45(D1):D1112-D1116.

334. Sleator RD & Hill C (2002) Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol Rev* 26(1):49-71.

335. Lanyi JK (1974) Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol Rev* 38(3):272-290.

336. Oren A (2008) Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems* 4:2.

337. Roberts MF (2006) Characterization of organic comvatible solutes of halotolerant and lialophilic microorganisms. *Extremophiles* 35:615-647.

338. Jelinska C, *et al.* (2005) Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure* 13(7):963-971.

339. Henneman B, van Emmerik C, van Ingen H, & Dame RT (2018) Structure and function of archaeal histones. *Plos Genetics* 14(9).

340. Corbeski I, *et al.* (2018) Microscale Thermophoresis Analysis of Chromatin Interactions. *Methods Mol Biol* 1837:177-197.

341. Van Der Valk RA (2016) Responding to environmental cues: the adaptive qualities of chromatin compaction proteins. PhD thesis (Leiden University, Leiden).

342. Duhr S & Braun D (2006) Why molecules move along a temperature gradient. *P Natl Acad Sci USA* 103(52):19678-19682.

343. Hwang H, Kim H, & Myong S (2011) Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proc Natl Acad Sci USA* 108(18):7414-7418.

344. Hwang H & Myong S (2014) Protein induced fluorescence enhancement (PIFE) for probing protein-nucleic acid interactions. *Chem Soc Rev* 43(4):1221-1229.

345. Aramendia PF, Negri RM, & Sanroman E (1994) Temperature-Dependence of Fluorescence and Photoisomerization in Symmetrical Carbocyanines - Influence of Medium Viscosity and Molecular-Structure. *J Phys Chem*-Us 98(12):3165-3173.

346. Levitus M & Ranjit S (2011) Cyanine dyes in biophysical research: the photophysics of polymethine fluorescent dyes in biomolecular environments. *Q Rev Biophys* 44(1):123-151.

347. Sosa D, *et al.* (2013) Periodic distribution of a putative nucleosome positioning motif in human, nonhuman primates, and archaea: mutual information analysis. *Int J Genomics* 2013:963956.

348. Segal E & Widom J (2009) Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struc Biol* 19(1):65-71.

349. Pereira SL & Reeve JN (1999) Archaeal nucleosome positioning sequence from *Methanothermus fervidus*. *Journal of Molecular Biology* 289(4):675-681.

350. Bailey KA & Reeve JN (1999) DNA repeats and archaeal nucleosome positioning. *Res Microbiol* 150(9-10):701-709.

351. Warnecke T, Becker EA, Facciotti MT, Nislow C, & Lehner B (2013) Conserved substitution patterns around nucleosome footprints in eukaryotes and Archaea derive from frequent nucleosome repositioning through evolution. *PLoS Comput Biol* 9(11):e1003373.

352. Thastrom A, *et al.* (1999) Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J Mol Biol* 288(2):213-229.

353. Vasudevan D, Chua EYD, & Davey CA (2010) Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. *J Mol Biol* 403(1):1-10.

354. Eslami-Mossallam B, Schiessel H, & van Noort J (2016) Nucleosome dynamics: Sequence matters. *Adv Colloid Interfac* 232:101-113.

355. Roessler M & Muller V (2001) Osmoadaptation in bacteria and archaea: common principles and differences. *Environ Microbiol* 3(12):743-754.

356. Singh BK (2010) Archaea in a hyper-arid polar desert. *P Natl Acad Sci USA* 107(2):E1-E1.

357. Stetter KO (2013) A brief history of the discovery of hyperthermophilic life. *Biochem Soc T* 41:416-420.

358. Jiang C & Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10(3):161-172.

359. Shi L, Wen H, & Shi X (2017) The Histone Variant H3.3 in Transcriptional Regulation and Human Disease. *J Mol Biol* 429(13):1934-1945.

360. Ahmad K & Henikoff S (2002) The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* 9(6):1191-1200.

361. Draker R, *et al.* (2012) A combination of H2A.Z and H4 acetylation recruits Brd2 to chromatin during transcriptional activation. *Plos Genet* 8(11):e1003047.

362. Wiedemann SM, *et al.* (2010) Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *J Cell Biol* 190(5):777-791.

363. Hoghoughi N, Barral S, Vargas A, Rousseaux S, & Khochbin S (2018) Histone variants: essential actors in male genome programming. *J Biochem* 163(2):97-103.

364. Campisi J & d'Adda di Fagagna F (2007) Cellular senescence: when bad things happen to good cells. Nature reviews. *Molecular cell biology* 8(9):729-740.

365. Quenet D (2018) Histone Variants and Disease. *Int Rev Cell Mol Biol* 335:1-39.

366. Segal E, *et al.* (2006) A genomic code for nucleosome positioning. Nature 442(7104):772-778.

367. Liu Y, *et al.* (2009) The Sac10b homolog in *Methanococcus maripaludis* binds DNA at specific sites. J Bacteriol 191(7):2315-2329.

368. Marsh VL, Peak-Chew SY, & Bell SD (2005) Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba. *J Biol Chem* 280(22):21122-21128.

369. Bohrmann B, Arnold-Schultz-Gahmen B, & Kellenberger E (1990) Ultrastructural localization of the histone-like protein HTa from the archaeon *Thermoplasma acidophilum*. *Journal of Structural Biology* 104(1-3):112-119.

370. Cabrera MA & Blamey JM (2018) Biotechnological applications of archaeal enzymes from extreme environments. *Biol Res* 51(1):37.

371. de Carvalho CC (2011) Enzymatic and whole cell catalysis: finding new strategies for old processes. *Biotechnol Adv* 29(1):75-83.

372. Haque RU, Paradisi F, & Allers T (2019) *Haloferax volcanii* as immobilised whole cell biocatalyst: new applications for halophilic systems. *Appl Microbiol Biotechnol* 103(9):3807-3817.

Summary
Nederlandse samenvatting
Curriculum Vitae
List of publications
Acknowledgments

SUMMARY

All life on earth contains DNA, which is used to store biological information. Organisms compact their DNA in order for it to fit inside their cell(s). Specialized proteins, which are referred to as nucleoid-associated proteins (NAPs), help organize the DNA to keep it compact but accessible for transcription. These NAPs often function as DNA benders, bridgers and wrappers. In most eukaryotic and archaeal species, histones are important NAPs that are able to bend or wrap the DNA duplex around its core. In this thesis, the histones from archaea are described in terms of their interaction with DNA.

Archaea, sometimes referred to as the third domain of life besides eukaryotes and bacteria, are a group of single-cellular organisms that are found in nearly every habitat on earth. Although optically somewhat resembling bacteria, genetically they are more related to eukaryotes. Archaea have only been recognized as a separate domain since the 1970s. Their role in the global geochemical cycles is significant, and some components of archaea have found their way to technological applications. Although archaea are a prominent inhabitant of the human gut, their role in health and disease has largely to be established. In CHAPTER 1, we give a broad introduction of archaea as a domain. Also, this chapter gives an over-

view of transcription regulation in archaea on a molecular level. Furthermore, we introduce the eukaryotic and archaeal histone, and we highlight their main differences and similarities.

Recently, it was discovered that some archaeal histones assemble onto DNA as continuous, rod-like structures, which we named hypernucleosomes. In **CHAPTER 2**, we describe on a molecular level which components of the archaeal histone contribute to the formation of the hypernucleosome. We also analyze the primary structure of all archaeal histones known to date, and we predict if these histones together with DNA can potentially assemble into hypernucleosomes.

For the characterization of NAPs and their interaction with DNA, single molecule techniques can be a powerful tool. To study the effect of NAPs on DNA conformation, tethered particle motion (TPM) can be a very useful high-throughput method. In TPM, the root mean square displacement of a bead tethered to a glass slide via a DNA duplex gives a read-out of DNA conformation. NAP concentration and physico-chemical conditions can be varied in order to study their effect on NAP-DNA assemblies. In **CHAPTER 3**, we show how TPM can be used when studying NAPs. Also, we provide new ways of TPM data representation, which allow insight into how individual measurements contribute to the population average.

In **CHAPTER 4**, TPM is used to study how histones HMfA and HMfB from the archaeal organism *Methanothermus fervidus* compact DNA. In addition, we use magnetic tweezers to perform force spectroscopy experiments, from which we conclude that the hypernucleosome formed by HMfB is more robust than the one formed by HMfA. We furthermore investigate if the components that were predicted to contribute to hypernucleosome formation as described in chapter 2, indeed play a role in that process *in vitro*. Also, we perform twisting experiments, which show that HMfB hypernucleosomes are left-handed, but torque can drive their conformation to change to a right-handed structure.

Like *M. fervidus*, the genomes of some archaeal species encode multiple histones. An interesting example are the histones from the salt-loving species *Nanosalina* sp. J07AB43, belonging to the recently discovered phylum *Candidatus* Nanohaloarchaeota. In chapter 2, we predict that some of these histones may be able to form a hypernucleosome, whereas one particular Nanosalina histone is very unlikely to do so. In **CHAPTER 5**, we perform microscale thermophoresis (MST) experiments of which the results suggest that histone HA together with DNA assembles into hypernucleosomes, whereas HB does not. We also measure

at very high salt concentrations, which better resembles the conditions in which species from *Ca.* Nanohaloarchaeota are found.

Since archaeal histones can compact parts of the genome, they are likely a part of the transcription regulation system. It is hypothesized that specific DNA sequences play a role in the positioning of archaeal histones onto the genome. In **CHAPTER 6**, the effect of a known specific sequence called Clone20 on *M. fervidus* histones is investigated. We find that these histones modestly compact Clone20 DNA at histone concentrations below 22 nM, whereas this does not happen on a random DNA sequence. Also, we propose mechanisms by which sequences like Clone20 may interplay with other elements of the archaeal transcription regulation system.

A broad discussion of the work presented in this thesis is given in **CHAPTER 7**.

NEDERLANDSE SAMENVATTING

Alle leven op aarde bevat DNA, dat gebruikt wordt voor de opslag van biologische informatie. Organismen maken hun DNA compact zodat het in de cel(len) past. Speciale eiwitten, die worden aangeduid als nucleoïde-geassocieerde eiwitten (NAP's), helpen met het compact doch toegankelijk maken van DNA. NAP's zijn vaak eiwitten die DNA buigen, oprollen, of een brug vormen tussen twee DNA duplexen. In de meeste eukaryote en archaeale soorten worden histonen gebruikt waar DNA omheen gerold of gebogen wordt. In dit proefschrift wordt beschreven hoe archaeale histonen interacteren met DNA.

Archaea, soms het derde domein van het leven genoemd naast eukaryoten en bacteriën, zijn een groep eencellige organismen die in bijna elke habitat op aarde worden aangetroffen. Hoewel ze uiterlijk op bacteriën lijken, zijn ze genetisch meer verwant aan eukaryoten. Archaea worden pas sinds de jaren '70 van de twintigste eeuw gezien als een afzonderlijk domein. Ze spelen een grote rol in geochemische cycli, en sommige componenten van archaea worden gebruikt voor technische toepassingen. Alhoewel archaea een groot aandeel hebben als bewoners van de menselijke darm, is hun rol in gezondheid en ziekte nog niet duidelijk. In **HOOFDSTUK 1** geven wij een brede introductie van het domein archaea. Ook

geeft dit hoofdstuk een overzicht van archaeale transcriptieregulatie op moleculair niveau. Daarnaast worden de eukaryote en archaeale histonen geïntroduceerd, en gaan we in op hun belangrijkste overeenkomsten en verschillen.

Recent is ontdekt dat sommige archaeale histonen samen met DNA een aaneengesloten staafvormige structuur vormen, die wij tot hypernucleosoom hebben gedoopt. In **HOOFDSTUK 2** beschrijven we op moleculair niveau welke componenten van het histon bijdragen aan de formatie van het hypernucleosoom. Wij hebben ook de primaire structuur van alle histonen die vandaag de dag bekend zijn, bestudeerd, en we voorspellen of het mogelijk is dat deze histonen samen met DNA een hypernucleosoom vormen.

Om NAP's en hun interactie met DNA te kunnen karakteriseren, kunnen single molecule technieken uitkomst bieden. Bij het bestuderen van het effect van NAP's op de structuur van DNA, is tethered particle motion (TPM) een handige methode met een hoge verwerkingscapaciteit. Bij TPM drukt de beweging van een balletje dat via DNA aan een glasplaatje vast zit, de structuur van het DNA nauwkeurig in cijfers uit. De concentratie van de NAP en fysisch-chemische condities kunnen worden gevarieerd om het effect daarvan op complexen van NAP's en DNA te bestuderen. In **HOOFDSTUK 3** laten we zien hoe TPM gebruikt kan worden in een studie van NAP's. Ook bieden we nieuwe mogelijkheden waarop data verkregen via TPM kan worden getoond, die inzicht verschaffen in hoe individuele metingen bijdragen aan het populatiegemiddelde.

In **HOOFDSTUK 4** gebruiken we TPM om te bestuderen hoe de histonen HMfA en HMfB, afkomstig van het archaeale organisme *Methanothermus fervidus*, DNA compacter maken. Daarnaast gebruiken we magnetic tweezers om kracht te kunnen uitoefenen op eiwit-DNA-complexen, waaruit we afleiden dat het hypernucleosoom gevormd door HMfB robuuster is dan het hypernucleosoom gevormd door HMfA. Ook onderzoeken we of de onderdelen waarvan we in hoofdstuk 2 voorspelden dat ze bij zouden dragen aan de vorming van de hypernucleosoom, inderdaad die rol spelen *in vitro*. Vervolgens laten de we resultaten zien van experimenten waarbij het DNA gedraaid wordt. Die tonen aan dat hypernucleosomen opgebouwd rondom HMfB , in principe linkshandig zijn, maar onder invloed van een krachtmoment kunnen veranderen in een rechtshandige structuur.

Zoals bij *M. fervidus* bevatten de genomen van sommige archaeale soorten meerdere genen die coderen voor histonen. Een interessant voorbeeld zijn de histonen van het zoutminnende organisme *Nanosalina* sp. J07AB43, dat behoort tot de recent ontdekte stam *Candidatus* Nanohaloarchaeota. In hoofdstuk 2

voorspelden we dat sommige van deze histonen een hypernucleosoom kunnen vormen, terwijl een ander histon van Nanosalina dat hoogstwaarschijnlijk niet kan. In **HOOFDSTUK 5** tonen we de resultaten van microscale thermophoresis (MST) experimenten, die suggereren dat histon HA samen met DNA inderdaad hypernucleosomen vormt, terwijl HB dat niet doet. We meten ook bij zeer hoge zoutconcentraties, die beter overeenkomen met de condities waarin soorten behorende tot *Ca.* Nanohaloarchaeota gevonden zijn.

Omdat archaeale histonen delen van het genoom compact kunnen maken, zijn ze waarschijnlijk ook betrokken bij transcriptieregulatie. Er wordt vermoed dat specifieke DNA-sequenties een rol spelen in het positioneren van histonen op het genoom. In **HOOFDSTUK 6** wordt onderzocht wat het effect is van een bekende specifieke DNA-sequentie genaamd Clone20 op de histonen van *M. fervidus*. We zien dat deze histonen zorgen voor gedeeltelijke compactie van het Clone20-DNA bij concentraties beneden 22 nM, terwijl dit bij willekeurig DNA niet gebeurt. Ook stellen wij mechanismen voor waarmee sequenties zoals Clone20 mogelijk interacteren met andere elementen van het archaeale transcriptieregulatiesysteem.

Een brede discussie van het werk dat in dit proefschrift wordt gepresenteerd, wordt gegeven in **HOOFDSTUK** 7.

CURRICULUM VITAE

Bram Henneman was born in Tilburg, the Netherlands on the 25th of August 1990. In 2008 he completed pre-university education (atheneum) at the Theresialyceum in Tilburg. In that same year, he started to attend the Bachelor's program Life Science & Technology at Delft University of Technology and Leiden University. As part of this program, he did an internship in the group of dr. Wiep Klaas Smits at the department of Medical Microbiology at the Leiden University Medical Center (LUMC). He received his Bachelor's degree in 2012. Between 2010 and 2011, Bram was a member of the 14th board of Study Association LIFE. He also represented Delft University of Technology and Leiden University as student ambassador between 2010 and 2013. In 2012, he enrolled in the Master's program Life Science & Technology at Leiden University, which he successfully completed in 2014. He went to the lab of dr. Remus Dame at the Leiden Institute of Chemistry at Leiden University and the lab of prof. dr. David Grainger at the University of Birmingham, United Kingdom, for internships. In 2015, Bram returned to the Dame lab for his PhD research, the results of which are presented in this thesis. In his spare time, Bram is a homebrewer and co-founder of brewery 'De Bierjuwelier'.

LIST OF PUBLICATIONS

Bram Henneman & Remus T. Dame, (2015) *Archaeal histones: dynamic and versatile genome architects*. AIMS Microbiology. doi: 10.3934/microbiol.2015.1.72

Bram Henneman, Clara van Emmerik, Hugo van Ingen & Remus T. Dame, (2018) *Structure and function of archaeal histones.* PLoS Genetics. doi: 10.1371/journal.pgen.1007582

Bram Henneman, Joost Heinsman, Julius Battjes & Remus T. Dame, (2018) *Quantitation of DNA affinity using Tethered Particle Motion.* In: Dame, R. (eds) Bacterial Chromatin. Methods in Molecular Biology vol 1837. Humana Press, New York, NY. doi: 10.1007/978-1-4939-8675-0_14

Bram Henneman, Thomas Brouwer, Clara van Emmerik, Ramon A. van der Valk, Amanda M. Erkelens, Monika Timmer, Nancy Kirolos, Hugo van Ingen, John van Noort & Remus T. Dame, *Mechanical and structural properties of archaeal hypernucleosomes*, submitted

Bram Henneman, Ramon A. van der Valk, Nancy Kirolos & Remus T. Dame, *Specific DNA binding of archaeal histones*, in preparation

ACKNOWLEDGMENTS

Apart from myself and my promotor, the following people have contributed to parts of this thesis: Clara van Emmerik and dr. Hugo van Ingen (**CHAPTERS 1, 2** and **4**), Julius Battjes and Joost Heinsman (**CHAPTER 3**), dr. Ramon van der Valk (**CHAPTERS 4** and **6**) and prof. dr. ir. John van Noort, Thomas Brouwer and dr. Monika Timmer (**CHAPTER 4**). I would like to thank my promotors dr. Remus Dame and prof. dr. Marcellus Ubbink for giving me the opportunity to work in their labs and for their contribution to this thesis. I also thank Anneloes Cramer-Blok, dr. Monika Timmer and Patrick Voskamp for their support in and around the lab. I thank the people who have shared an office with me, Daan Brocken, dr. Ramon van der Valk, dr. Kathy Chaurasiya, dr. Frédéric Crémazy and Ivar Dilweg, for making the days a little brighter. I also thank all members of the Macromolecular Biochemistry group for their help in the past couple of years, especially Folkert Kan and Anja Herdtfelder, the students who I supervised. Special thanks go to my friends and family, as well as to dr. Elleke Peterse, without whom I could not have done this.