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

Erkelens, A.M.

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

Proefschrift
ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 27 september 2023
klokke 11:15 uur

door

Amanda Monique Erkelens
geboren te Rotterdam
in 1995

Promotores:	Prof dr. R.T. Dame
	Prof. dr. ir. S.J.T. van Noort
Promotiecommissie:	Prof. dr. M. Ubbink
	Prof. dr. A. Briegel
	Prof. dr. ir. T.J.G. Ettema (Wageningen University)
	Dr. W.K. Smits (Leids Universitair Medisch Centrum)
	Dr. W.J.J. Meijer (Universidad Autónoma Madrid)

ISBN: 978-90-8333-308-3

Cover design: The microscope was made using the crochet pattern that can be found in Knotmonsters: Biology edition by Michael Cao, independently published in 2022.

Printed by: Print Service Ede – www.proefschriftenprinten.nl

Most of us tend to know a little bit about everything, but seldom do we know a great deal about something, and only very exceptionally do we know more than everyone else

The Art of Flight - Fredrik Sjöberg

Table of contents

Chapter 1

Introduction

Every organism in the tree of life faces the same challenge: the effective volume of its genome far exceeds the volume of the cell or cellular compartment in which it is contained. Therefore, strategies have evolved to ensure proper genome compaction and organization. At the same time, the DNA must be accessible for genomic transactions such as transcription and replication. The main factors contributing to these processes are DNA supercoiling, macromolecular crowding and binding of chromatin proteins with architectural properties (1–4). These chromatin-associated proteins can be classified based on their architectural properties: DNA wrapping, bending, bridging, or formation of a nucleoprotein filament. Proteins from the different classes can exhibit structural interplay in either a synergistic or antagonistic manner (2–4).

The transcription process is, to a large extent, conserved throughout the tree of life. Bacteria, archaea and eukaryotes all express RNA polymerase (RNAP) to transcribe DNA into mRNA. Transcription initiation in bacteria occurs when a σ factor binds to the promoter -10 element and guides the RNAP to the DNA (5). In archaea and eukaryotes, the TATA-box and the B recognition element (BRE) are necessary for transcription to initiate (6, 7). The TATA-binding protein (TBP) and Transcription factor B (TFB) bind to the respective DNA elements and recruit RNAP (8). Transcription elongation and, therefore, correct, uninterrupted mRNA synthesis is aided by the transcription elongation factor called Spt5 (in archaea and eukaryotes) or NusG (in bacteria) (9). Chromatin proteins, such as histones and other nucleoid-associated proteins (NAPs), can interfere with these processes in several ways and thereby regulate gene expression. They can bind to the transcription initiation sites and exclude RNAP from binding, they can trap RNAP by loop formation, or interfere with transcription elongation by binding across the coding region (10, 11). In general, chromatin proteins are relatively small and basic proteins. They can bend, wrap, stiffen or bridge DNA duplexes upon binding to the DNA. In eukaryotes, histones are the main chromatin proteins, while bacteria encode a plethora of NAPs. For instance, at least 12 proteins have been classified as NAPs in *Escherichia coli* (12). They are mainly defined as proteins that combine a function in transcriptional regulation and genome structure with limited sequence specificity, in contrast to transcription factors with a sequence-specific regulatory function. Archaea encode both histone proteins and NAPs (table 1.1). At the time of discovery, archaea were defined as a third branch of the tree of life, next to bacteria and eukaryotes. However, with advances in the phylogenetic description of archaea, a two-domain model of the tree of life was adopted, in which eukaryotes are a sister group of the archaea. This implies that archaea were at the basis of eukaryogenesis, and study of archaeal evolution could provide insight into this important evolutionary event.

In this thesis, I will focus on chromatin proteins throughout the tree of life and their DNA binding properties. In this chapter, I will introduce the different chromatin proteins based on their architectural effect on DNA and explore their distribution among species. In chapter 2, I will describe the known characteristics and functions of bacterial DNA bridging proteins in more depth. In chapters 3-6, I will describe my experimental studies on bacterial and archaeal DNA bridging proteins and archaeal histones.

DNA wrappers

Eukaryotic histones

The best-known architectural proteins, prototypical DNA wrappers, are eukaryotic histones. Conserved throughout the eukaryotic domain of life, they share the characteristic histone fold of three α-helices connected by two loops (13). They form obligatory H3-H4 and H2A-H2B heterodimers. Two H3-H4 dimers interact to form a tetramer with which two H2A-H2B dimers can associate. The result is an octameric protein core with around 147 bp DNA wrapped around it (figure 1.1A). Depending on the length of DNA between nucleosomes and other factors such as the linker histone H1, nucleosomes can be arranged in higher-order structures such as the 30 nm-fiber (14, 15). Nucleosomes are generally associated with repression of transcription by excluding other factors from binding to the DNA (16).

The N-terminal tails of histones can be post-translationally modified, which affects the accessibility of the chromatin and, as a consequence, gene transcription (17). Another mechanism of modulating nucleosome (and therewith chromatin) structure is exchange of histones with histone variants. The H2A-H2B dimers are generally more exchangeable than the H3-H4 tetramer (18). Histone variants, each with their own specific role in genome organization and gene regulation, exist of all core histones with (in number) a bias towards H2A and H3 (19).

Archaeal histones

Most archaeal genomes encode histones that resemble their eukaryotic counterparts (table 1.1). Histones are present at least in a minority of available genomes in every superphylum, phylum and class. Only in *Ca. Marsarchaeota* no histone genes have been found to date. Crenarchaeota, which have been considered histone-free for a long time, encode histones in a minority of genomes (20, 21). Due to the relatively low amount of available genomes for the three histone-free phyla, histone genes may be discovered later. Also, as they all still have the *Candidatus* status, advances in culturing

Archaea could provide more insight into their genome organization and the presence of histones.

Archaeal histones share the characteristic histone fold with eukaryotic histones, but generally lack the N- and C-terminal tails (21) (figure 1.1B). Also, a preference for GC-rich sequences is shared between archaea and eukaryotes (22, 23). Due to the resemblance between the archaeal histone and the eukaryotic H3/H4 tetramer, it was suggested that eukaryotic histones evolved from the archaeal ones (24). The H2A and H2B histones likely later evolved from the H3/H4 tetramer (25). The expanding archaeal branch of the tree of life supports this evolutionary relationship. *Ca.* Asgardarchaeota are currently considered the closest living relatives to eukaryotes, and strikingly, some of their histones have an N-terminal tail (21, 26–28). As this tail also includes several lysine residues, a similar function and acetylation pattern to eukaryotic tails might be very well possible.

The archaeal histones HMfA and HMfB from *Methanothermus fervidus* and HTkA and HTkB from *Thermococcus kodakarensis* are the best studied. These model histones form homo- and heterodimers in solution. Micrococcal nuclease (MNase) digestion of *M. fervidus, Haloferax volcanii* and *Methanobacterium thermoautotrophicum* showed protection of ~60 bp of DNA, suggesting binding as tetramers (22, 29). High-affinity sites for HMfB found by systematic evolution of ligand by exponential enrichment (SELEX) were suggested to be bound by and wrapped around a tetrameric protein core, highlighting the importance of the tetrameric structure (30, 31). However, MNase digestion studies on chromatin of *T. kodakarensis* showed that the dimer is the basic unit when binding to the DNA. The size of protected DNA increased as multiples of 30 bp up until ~ 500 bp, suggesting that structures larger than the eukaryotic octamer are also relevant (32). Indeed, X-ray structures and single-molecule experiments on HMf show that a so-called hypernucleosome can be formed, which is stabilized by stacking interactions between dimers (33, 34) (figure 1.1B). Although these interactions are predicted to be widespread throughout the archaeal domain, several species encode multiple histone variants with different stacking propensities (21). The incorporation of different variants might be key to modulating the size and stability of hypernucleosomes by acting as capstones (35).

Little is known about the effects of archaeal histones and variants of these on transcription. A mildly repressive effect was found for HMf in *E. coli* cells without severe growth defects. In *T. kodakarensis,* gene expression patterns change depending on the presence of histones (un)able to form hypernucleosomes (36). This suggests that naturally occurring histone variants that are less likely to oligomerize have a function in modulating hypernucleosome size and structure, and therewith gene expression.

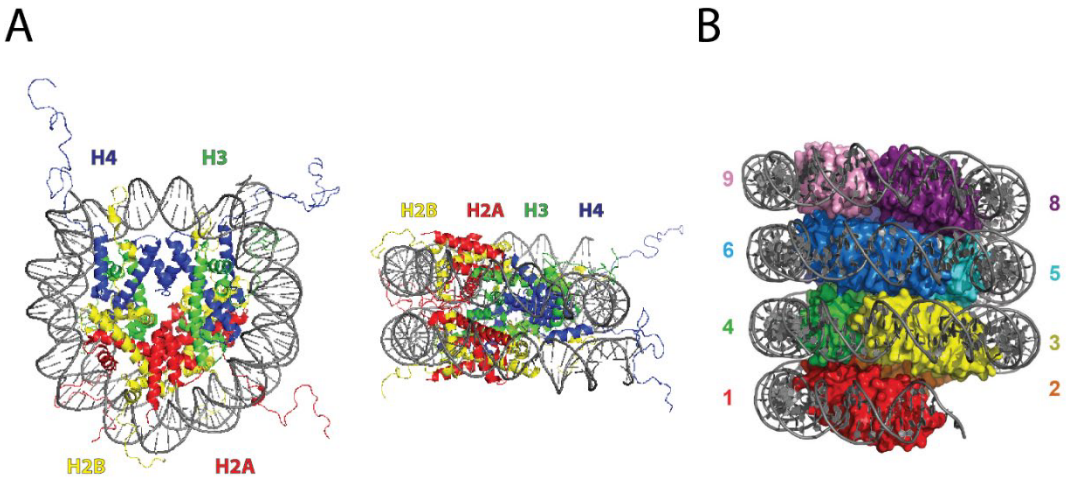


Figure 1.1 Eukaryotic and archaeal histones are examples of DNA wrappers. A) Two views of a eukaryotic nucleosome (PBD: 1KX5 (37)) consisting of a H3/H4 tetramer (H3: green, H4: blue) and two H2A/H2B dimers (H2A: red, H2B: yellow) that wrap 147 bp of DNA. B) A model of an archaeal hypernucleosome (PBD: 5T5K (33)) consisting of nine HMfB dimers that wrap around 270 bp of DNA. This image was reproduced from Henneman et al. 2018 (21).

Table 1.1 Distribution of chromatin proteins across the archaeal domain of life. For the archaeal histones, Alba and MC1, the entries from the NCBI protein database that were annotated as such and assigned to a superphylum and phylum were included. For Sso10a (CAH69222.2), Cren7 (P0C835.1), Sul7 (AAK42679.1) and CC1 (WP_ 053240420.1), BLAST with the reference sequences was performed. For SMC-proteins, the organisms found in Yoshinaga et al. are indicated (38). Presence of a protein in the genome of organisms of the phylum/class are indicated with Y and absence with N. Proteins indicated with Y* means that it was found only in the minority of the genomes of that specific class. Due to the incomplete picture of SMC-diversity in archaea, phyla without known SMC-proteins are left open instead of indicated with N. DPANN is an acronym for Diapherotrites, Parvarchaeaota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota and TACK for Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota.

Superphylum	Phylum	Class	Histones	SMC	Alba	MC1	Sso10a	Cren7/Sul7/CC1
	Euryarchaeota	Archaeoglobi	Y	Y*	Y	N	Y*	N
		Hadesarchaea	Y		Y	N	N	N
		Halobacteria	Y	Y*	N	Y	N	N
		Hydrothermarchaeota	Y		Y	N	N	N
		Methanoatronarchaeia	Y		Y	N	N	N
		Methanobacteria	Y		Y	N	N	N
		Methanococci	Y		Y	N	Y*	N
		Ca. Methanofastidiosa	Y*		Y	N	N	N
		Ca. Methanoliparia	Y		Y	N	N	N
		Methanomicrobia	Y	Y*	Y	Y	Y*	N
		Methanopyri	Y*		Y*	N	N	N
		Nanohaloarchaeota	Y		Y	N	N	N
		Theionarchaea	Y	Y*	Y	N	N	N
		Thermococci	Y	Y*	Y	N	N	N
		Thermoplasmata	Y*	Y*	Y	Y	Y*	N
DPANN	Ca. Aenigmarchaeota		Y		Y	Y	N	N
	Ca. Altiarchaeota		Y		Y	Y	N	N
	Ca. Diapherotrites		Y*		Y	N	N	N
	Ca. Huberarchaeota		Y		N	Y	N	N
	Ca. Micrarchaeota		Y		Y	Y*	N	N
	Nanoarchaeota		Y		Y	N	N	N
	Ca. Pacearchaeota		Y*		Y	N	N	Y*
	Ca. Parvarchaeota		Y		N	N	N	N
	Ca. Woesearchaeota		Y		Y	Y*	N	Y*
TACK	Ca. Bathyarchaeota		Y	Y*	Y	N	Y*	N
	Crenarchaeota		Y*	Y*	Y	N	Y*	Y*
	Ca. Geothermarchaeota		Y		Y	N	N	N
	Ca. Korarchaeota		Y		Y	N	N	N
	Ca. Marsarchaeota		N		Y	N	N	N
	Ca. Nezhaararchaeota		Y		Y	N	N	N
	Thaumarchaeota		Y		Y	N	Y*	N
	Ca. Verstraetearchaeota		Y		Y	N	N	N
Asgard Archaea	Ca. Heimdallarchaeota		Y	Y*	Y	N	N	N
	Ca. Lokiarchaeota		Y		Y	N	N	N
	Ca. Odinarchaeota		Y		Y	N	N	N
	Ca. Thorarchaeota		Y		Y	N	N	N

DNA bridgers

SMC proteins – in all domains of life

The only architectural chromatin protein family that is conserved throughout the tree of life, is the structural maintenance of chromosomes (SMC) family of proteins (39). These proteins structure the chromosome by bridging two DNA strands followed by loop extrusion (40). This is an active process involving the hydrolysis of ATP (41, 42) (figure 1.2A). SMC proteins consist of a hinge dimerization domain, an ATPase head domain and an anti-parallel coiled-coil arm between the two domains (43, 44). The ATPase function was also found in the universally conserved Rad50, which has a function in DNA repair. Together they form an 'SMC-like' superfamily (39, 45). Eukaryotes encode for six SMC subfamilies (SMC1-6). Proteins belonging to these subfamilies form obligatory heterodimers called cohesin (SMC1/3), condensin (SMC2/4) and the SMC5/6 complex (46–49). Each heterodimeric complex is associated with distinct accessory proteins and functions in chromosome condensation during replication and DNA repair (50). In bacteria, several SMC-like proteins have been identified. For example, in *E. coli* the MukBEF complex has been shown to fulfill the SMC function (51) and in *B. subtilis* and *Caulobacter crescentus* this function is carried out by SMC-ScpAB (52–55). More widespread throughout the bacterial domain is the MksBEF complex (56).

In archaea, until recently, only a few SMC-like proteins were identified. Two SMC-like proteins in *Halobacterium salinarum* (Sph) were found (57) and Archadin-4 was identified in Thermoproteales archaea (58). More recently, coalescin (ClsN) in *Sulfolobus acidocaldarius* and *Sulfolobus islandicus* was shown to be involved in chromosome compartmentalization (59). SMC-like proteins, Sph, Archadin-4 and ClsN seemed to be restricted to specific lineages only. Considering the wide distribution of SMC-like proteins in bacteria and eukaryotes, this raised the question whether other SMC-like proteins are present in the archaeal domain. Yoshinaga et al. discovered a new, widespread group which they called Archaea-specific SMC-related proteins (ASRPs) (38). Although experimental validation is still lacking, this increased the potential diversity of SMC-like proteins in archaea. Because the diversity and distribution in the archaeal domain is a topic of ongoing investigation, SMC proteins could only partially be included in table 1.1.

SMC-like proteins are the only class of DNA bridging proteins in bacteria, archaea and eukaryotes. Below, the domain-specific DNA bridging proteins are discussed.

H-NS-like proteins – in bacteria

The histone-like nucleoid structuring protein (H-NS) is a key actor in both genome organization and transcription regulation in *E. coli.* H-NS regulates around 5-10% of the *E. coli* genes, especially genes acquired by horizontal gene transfer (60, 61). In contrast to the active, ATP-driven SMC proteins, H-NS is an example of a passive DNA bridger (figure 1.2B). Structurally, H-NS consists of an N-terminal oligomerization domain containing a dimerization and an oligomerization site, followed by a flexible linker and a C-terminal DNA binding domain (10, 62, 63). This last domain harbors an AT-hook-like motif to recognize the minor groove of the DNA. H-NS has two modes of DNA binding, resulting in DNA bridging or nucleofilament formation (64–66). The switch between these two modes is dependent on environmental conditions and interaction with protein partners (10, 67–69) (see also Chapter 2). Several functional and structural homologs of H-NS have been found in other bacteria, which are discussed in more depth in Chapter 2.

Alba – in archaea and some eukaryotes

The most widespread NAP in archaea is Alba (Acetylation lowers binding affinity). At least one copy of Alba is present in nearly every phylum and class (table 1.1). There are a few exceptions, namely Halobacteria (Euryarchaeota), *Ca.* Huberarchaeota and *Ca.* Parvarchaeota (both DPANN). In eukaryotes, RNA-binding proteins such as subunits from ribonucleoprotein complexes (RNase P/MRP) and ciliate macronuclear development protein 2 (Mdp2) are related to archaeal Alba (70). These proteins were mainly studied in protozoan parasites, such as the malaria parasite *Plasmodium falciparum* (71, 72), but similar proteins have been found in *Caenorhabditis elegans, Homo sapiens* and *Arabidopsis* (73). Most of the research on archaeal Alba was done in non-histone containing *Sulfolobus spp.* where Alba is the main chromatin protein. Alba is a 10 kDa protein and forms homodimers in solution (74). It contains a β-sheet arrangement and two α-helices and the dimer has a highly basic surface which functions as DNA binding interface (75, 76). Alba constitutes about 4% of the cellular protein of *S. shibatae* and binds DNA without apparent sequence selectivity (77). The DNA binding affinity of Alba was found to be dependent on its modification status. Originally it was thought that Alba was subject to lysine acetylation (hence its name), but further research identified trimethylation of Lys16 as the factor that lowers the DNA binding affinity (78, 79).

Alba binds to ssDNA and RNA with a similar affinity as to dsDNA *in vitro* (80). In many cases, such as the eukaryotic Alba proteins mentioned above, Alba domains are found in proteins related to RNA metabolism (70). Recently it was found that Alba

catalyzes RNA unwinding and unfolding, especially at elevated temperature, which is Alba's natural environment in hyperthermophilic *Sulfolobus spp.* Therefore, a role as RNA chaperone was proposed (81). This second function of Alba was shown to be dependent on Lys17, which is involved in RNA binding (82).

Two distinct DNA binding modes have been identified: at low protein:DNA ratio (about 1 dimer per 15 bp) Alba bridges two DNA duplexes (figure 1.2B), while at higher protein:DNA ratios (one dimer per 5 bp), it binds cooperatively along the DNA (75, 83, 84). The phenylalanine on position 60 has been shown to be important for dimer-dimer interactions (76), along with hydrophobic interactions between the two α1-helices (76, 85). Phe60 is responsible for side-by-side interactions and is, therefore necessary for cooperative binding along the DNA.

S. solfataricus, among many archaea, encode two Alba proteins: Alba1 and Alba2 (86), where Alba1 does have Phe60 while it is absent in Alba2. Alba1 is expressed at a higher level than Alba2. When Alba2 is present, obligate heterodimers are formed, resulting in a mixture of Alba1 homodimers and Alba1:Alba2 heterodimers (86). Alba1 homodimers can both bridge two DNA duplexes and form a stiffening filament along the DNA in a concentration-dependent manner (87). However, as Alba2 lacks the Phe60 for effective dimer-dimer interactions, Alba1:Alba2 heterodimers only bridge DNA due to a loss of cooperativity. Therefore, tuning the relative concentrations of Alba1 and Alba2 was proposed to be an effective way to regulate DNA bridging behavior and nucleoprotein filament formation, akin to H-NS and its interaction partners (87). Besides, Alba and H-NS share high cellular expression levels and a lack of sequence specificity. It is currently unknown, but well possible that Alba has a similar function in gene regulation and/or genome organization as H-NS-like proteins.

Sso10a – in Archaea

Most Crenarchaeota do not express histone proteins, but encode several NAPs instead. Sequences of the Sso10a family of proteins have been found in the TACK superphylum (Ca. Bathyarchaeota, Crenarchaeota and Thaumarchaeota) and Euryarchaeota (Methanomicrobia, Methanococci, Archaeoglobi and Thermoplasmata) (88) (table 1.1). The best-studied members of this family are from *Sulfolobus spp* (74, 89). Sso10a proteins are small 10 kDa proteins with a winged helix-turn-helix (wHTH) DNA binding domain and an anti-parallel coiled-coil structure as dimerization site (89–92). *S. solfataricus* expresses three Sso10a homologues: Sso10a1, Sso10a2 and Sso10a3. Both Sso10a1 and Sso10a2 were shown to bend DNA at low protein concentrations (92). At higher concentrations, however, the behaviour of the two proteins is different. Where Sso10a1 is able to bridge two DNA duplexes, Sso10a2 forms a

stiffening filament along the DNA. This difference in DNA binding properties might have a structural basis as Sso10a2 contains four extra residues in the wHTH domain, creating an extended loop with two additional charged residues compared to Sso10a1.This could lead to the formation of extra electrostatic interactions between Sso10a2 dimers, resulting in the formation of a nucleofilament (92). Therefore, the absence of DNA stiffening behavior for Sso10a1 is most likely the result of differences in the dimer-dimer interface.

The multiple effects on DNA conformations of the Sso10a proteins resemble the multiple architectural properties of bacterial NAPs. For instance, the bacterial chromatin protein HU is a DNA bender (see below), but also stiffens DNA at higher concentrations (93, 94). The DNA bridging behaviour of Sso10a1 resembles that of H-NS-like proteins (64, 65, 95–98). It is likely, based on sequence similarity, that Sso10a1 and Sso10a2 can heterodimerize, and possibly also form heterodimeric complexes with Sso10a3. By regulating the relative expression levels of the Sso10a proteins, the cell can potentially regulate genome architecture. Although it is currently unclear whether Sso10a proteins affect transcription, this feature might be relevant in gene expression.

H1 and BAF – in eukaryotes

Histone H1 and the barrier-to-autointegration factor (BAF) are two examples of passive DNA bridging proteins in eukaryotes. H1 binds at the entry/exit site of the nucleosome and influences the nucleosome repeat length (99). Structurally, H1 consists of a winged helix domain with an unstructured N-terminal tail and a highly basic, unstructured C-terminal domain, which is necessary for DNA binding *in vivo* (100) and nucleosome condensation in higher-order structures (101). The structured globular domain of H1 was found to bridge the nucleosome complex with linker DNA, thereby compacting the chromatin structure (102). There are multiple subtypes of H1 in eukaryotic cells with subtype-specific PTMs, and they play particular roles in the formation of chromatin structure (nicely reviewed by Hergeth and Schneider in 2015 (103)).

BAF was originally identified as a protein to prevent autointegration of retroviral DNA (104, 105), but it also has a function in repair of nuclei ruptures (106) and is involved in various diseases (107). It is a dimeric protein in solution and uses a helix-hairpin-helix DNA binding domain (108, 109). When bound to the DNA, BAF can bridge two DNA strands (110–112) by either forming a higher-order complex (a dodecamer) or binding as a dimer depending on the length of the DNA (113). Because BAF binds DNA without sequence specificity, a role in chromatin organization rather than transcription regulation was proposed (111). The binding of BAF to the DNA and its interaction partners depends on its phosphorylation status (114, 115). Interestingly, BAF interacts with histone H3 and the histone variant H1.1 and affects the modification status of histones, but the *in vivo* function of this interplay between BAF and histones is still unclear (116, 117)

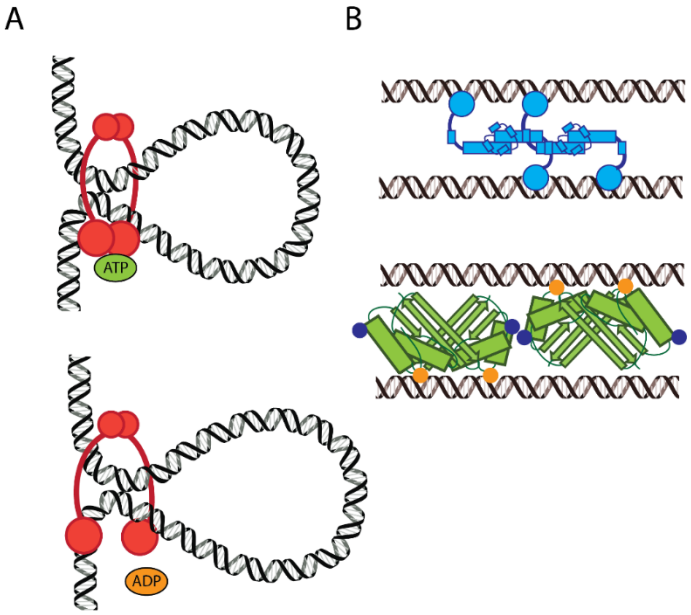


Figure 1.2 Active and passive DNA bridgers A) SMC proteins are examples of active DNA bridgers. When ATP is bound (top) the head domains dimerize and a ring structure is formed. Upon hydrolysis of ATP to ADP (bottom), the head domains release and DNA can be pulled through the ring. B) Bacterial H-NS (top) and archaeal Alba (bottom) are examples of passive DNA bridgers. H-NS dimerizes via a hand-shake topology and multimerizes using a helix-turn-helix interface. Alba forms dimers via a β-sheet arrangement and residue Phe60 (in blue) is used for multimerization. Residue Lys16 important for DNA binding is indicated in orange. Rectangles represent α-helices and arrows indicate β-sheets.

DNA benders

DNA wrapping, with histones as prototypical examples, can be considered an extended form of DNA bending. Bacteria encode several DNA bending proteins, but mostly lack histones and other DNA wrappers. Recently, a first indication that bacteria also encode histones was published, but their DNA binding mode might be distinct from eukaryotic and archaeal histones (118). Clustered binding of DNA bending proteins would result in a structure comparable to a DNA wrap (12, 119–121). Several archaeal phyla, most notably the histone-lacking Crenarchaeota, also encode unique DNA bending proteins (122–125). In other archaeal phyla that lack Alba proteins, another DNA bender is encoded called MC1 (table 1.1).

HU, IHF and Fis

The histone-like protein from *E. coli* strain U93 (HU) is a widely conserved NAP among bacteria (126). Most bacteria encode one HU protein, while *E. coli* expresses HUα and HUβ that can heterodimerize (126). HU functions in many cellular processes such as DNA organization, gene expression and protection of DNA against various stresses (127–130). Due to being an abundant protein that binds DNA without sequence specificity, HU binds throughout the bacterial genome. However, HU does have a higher affinity for already bent and/or distorted DNA (12, 127, 131).

HU consists of an α-helical body with two β-ribbon arms (121) (figure 1.3A). Proline residues at the end of the arms intercalate in the minor groove of the DNA and three lysine residues facilitate DNA binding (120, 127, 132). The bending angle caused by HU binding is flexible between 105° and 140°, suggesting that HU acts as a flexible hinge (93, 121). Next to the DNA bending mode of HU, a second DNA binding mode has been described, where HU forms nucleofilaments along the DNA (93). The switch between the two binding modes is dependent on local HU concentration.

Next to bacteria, the archaeon *Thermoplasma acidophilum* encodes a HU homolog called HTa (133). Despite sharing its primary, and predicted secondary to quaternary structure with bacterial HU, HTa evolved to behave like an archaeal histone protein in terms of DNA binding preferences and oligomerization behavior.

With 40% sequence identity, integration host factor (IHF) is similar to HU in many respects (126). However, for IHF a consensus sequence has been found (134) and IHF binding induces a DNA bend of 160° (120). IHF is only present in Gram-negative bacteria in contrast to HU (126). IHF fulfills a more specific role than the general HU protein in transcriptional regulation, replication and integration of phage DNA (135–137).

Fis is another DNA bending protein across the bacterial domain of life. It binds DNA as a dimer using a helix-turn-helix motif (138). The binding of Fis induces a bend in

the DNA between 50-90°(139). It recognizes a 15 bp palindromic sequence, mainly by the sequence-dependent width of the minor groove (138). Fis is mainly present in intergenic regions of the genome, but it also acts as a transcription regulator (140, 141). As Fis can often be found at overlapping and branched DNA strands, an architectural role was proposed next to its regulatory function (142). The binding profiles of H-NS, IHF and Fis partly overlap and they work together to repress certain genes (140, 143). This highlights the structural and functional interplay between different NAPs.

MC1

Some archaea lack genes encoding for Alba proteins, but express Methanogen Chromosomal protein 1 (MC1), a small monomeric protein of 93 amino acids (144), instead. These are mainly Halobacteria (Euryarchaeota) and *Ca.* Huberarchaeota (DPANN). However, some phyla encode both Alba and MC1, which are Methanomicrobia (Euryarchaeota), *Ca*. Aenigmarchaeota and *Ca.* Altiarchaeota (both DPANN). So far, no MC1 sequences have been found in the TACK and Asgard superphyla. Most organisms have only one copy of the gene encoding MC1, but two or more have also been found. For instance, *Halococcus thailandensis JCM 13552* harbors nine copies of this gene. To what extent these MC1 paralogues are all expressed in the cell and have similar or different DNA binding characteristics remains to be determined.

Structurally, MC1 consists of five β-sheets and one α-helix leading to the formation of a pseudo-barrel connected to a long flexible arm (145, 146) (figure 1.3B). The residues identified as important for DNA binding are Arg25, Trp74 and Lys86, which are conserved among the MC1 containing species (147). MC1 binds to the DNA in a noncooperative manner as monomer with a binding site size of around 11 bp (148). MC1s affinity for double-stranded DNA is high (K_D<100 nM). It preferentially binds to the following DNA sequence only consisting of adenines and cytosines: [AAAAACACAC(A/C)CCC(C/A)] (149). Furthermore, MC1 binds strongly to bent DNA, such as four-way junctions (150).

Upon binding of MC1, the DNA undergoes a significant conformational change caused by two bends of 55° and 75° resulting in a V-turn (151, 152). On longer DNA, MC1 can stabilize multiple V-turn conformations, leading to a structure which resembles a DNA wrap. This observation lead to the hypothesis that MC1 has two different DNA binding modes: DNA bending at lower concentration and DNA wrapping at higher concentrations (152). This might relate to the two observed effects MC1 has on transcription *in vitro*, which is activated at low MC1/DNA ratio (DNA bending), but repressed at higher ratios (DNA wrapping), but this is still an open question (144). Next to regulation by tuning the expression level, methylation on Lys37 of MC1 was found

(153), but what effect this has on the DNA binding properties of MC1 is currently unknown.

Cren7, Sul7 and CC1

Next to DNA bridging proteins Alba and Sso10a, several Crenarchaeota-specific DNA benders were identified in the search for proteins with a different architectural effect on the DNA in these organisms. Cren7 and Sul7 are both small, monomeric proteins, which are not related in terms of amino acid sequence, but their tertiary structure and biochemical properties are similar (122, 123, 125, 154, 155). They contain two antiparallel beta-sheets with either an extended loop in between (Cren7, figure 1.3C) or an additional C-terminal α-helix (Sul7) (125, 155). Crystal structures with DNA show binding to the minor groove of the DNA and an induced bend of 50-60° (156–159) . The binding site size for Cren7 is 6-7 bp, binding as a monomer to the DNA in a head-to-tail manner. Cren7 has a slight preference for AT-rich DNA, while no sequence preference was found for Sul7 (157). Lysine methylation was found on five positions for Cren7 and on seven and nine positions for the two Sul7 proteins, respectively in *S. islandicus*, but their relative occurrence and function *in vivo* remain to be investigated (160, 161).

In the search for a second chromatin protein next to Alba in the crenarchaea *Pyrobaculum aerophilum* and *Thermoproteus tenax,* another small (6 kDa) protein was identified: Crenarchaeal Chromatin Protein 1 (CC1). This protein has further homologues in *Aeropyrum pernix* (124) and a few sequences were found in *Ca.* Pacearchaeota and *Ca*. Woesearchaeota (both DPANN) (table 1.1). Secondary structure prediction suggested mainly β-sheet organization, similar to Cren7 and Sul7. However, CC1 binds both ss and dsDNA, while Cren7 and Sul7 do not bind ssDNA (162).

HMG-box proteins

Known DNA-bending proteins in eukaryotes are the high motility group (HMG)-box proteins. The global fold of a HMG-box consists of three α-helices in an L-shape (163). They bind the DNA using a hydrophobic interface and introduce a substantial bend in the DNA (164–166). Often, HMG-box proteins have an N- or C-terminal extension that stabilizes their binding to the DNA and the introduced bend (167, 168). They recognize DNA independently of DNA sequence but mainly on the basis of structure, for instance distorted DNA or four-way junctions (169, 170). Functionally, they have been associated with both activation and repression of transcription *in vitro*. This could be related to interactions of HMG-box proteins with transcription factors and the basal transcription machinery (171–178). Also, a more architectural role has been proposed where HMG-box proteins ‘pre-bend’ the DNA for other chromatin proteins to form nucleoprotein

complexes (179, 180). In *Saccharomyces cerevisiae,* loss of a HMG-box protein resulted in higher susceptibility for micrococcal nuclease (MNase) of the DNA, suggesting a structural and protecting function (163).

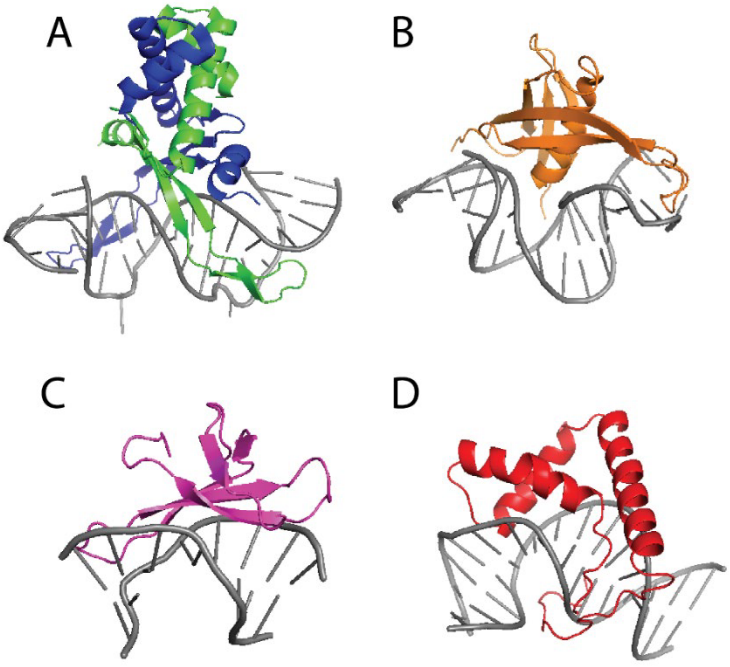


Figure 1.3 DNA bending proteins A) Bacterial HU protein dimer (PDB: 1P78 (121)) using a β-sheet "clamp" to bind the DNA and α-helices to dimerize. B-C) The archaeal monomeric DNA bending proteins MC1 (B) (PDB: 2NBJ (146)) and Cren7 (C) (PDB: 5K17 (181)). D) Eukaryotic HMG-box domain from the LEF-1 protein from mouse (PDB: 2LEF (165)). Note that HU and MC1 bend the DNA in the opposite direction compared to Cren7 and the HMG-box domain.

Nucleoprotein filament formation

Several proteins mentioned above exhibit a second DNA binding mode, where they form a filament along the DNA, effectively stiffening it. Such binding is often observed for passive DNA bridgers, such as H-NS-like proteins, Alba and Sso10a (66, 87, 92, 182–184). Also, the DNA bender HU exhibits DNA stiffening behavior (93). For H-NS-like proteins, it has been a topic of discussion which DNA binding mode is relevant *in vivo*. It is possible that the change between DNA bridging and the formation of nucleofilaments functions as a switch between a repressive state and a state permissive of transcription (69, 185). Multimerization of proteins is necessary for nucleoprotein filament formation, which can result from either high local protein concentration, as for HU, or cooperative interactions, as for Alba and Sso10a, or both (87, 92, 93). However, as *in vivo* data is lacking for these proteins, the likelihood and functional relevance of nucleoprotein filament formation remains an open question.

Discussion

Across the tree of life, architectural chromatin proteins act in concert to organize the genome and regulate gene expression. Conserved structural effects of these proteins on DNA, such as wrapping or bridging, can be found in every domain of life. It has been hypothesized that archaea need at least two chromatin proteins (most likely two that execute a different structural effect) (186), but there is very little data about the interplay between archaeal chromatin proteins. In *S. solfataricus,* the interplay between the chromatin proteins Alba and Cren7 was investigated *in vitro*. It was found that Cren7 disrupts larger structures formed by Alba (187), which could result in more open chromatin. The effects of this interplay might be relevant *in vivo*.

The lack of standard laboratory strains of most phyla currently limits research in the archaeal domain of life. Most data is either from Euryarchaeaota, such as *M. fervidus* and *T. kodakarensis,* or Crenarchaeota, such as *Sulfolobus spp.* Recent advances in culturing two Lokiarchaeota from the Asgard archaea could help to establish more standard laboratory strains across different phyla to study chromatin proteins *in vivo* by, for instance, Chromosome Conformation Capture (3C)-based techniques and Chromatin Immunoprecipitation Sequencing (ChIP-seq). Close monitoring of genomic changes in newly cultured organisms is necessary as studies of a laboratory strain of *Methanothermobacter thermautotrophicus* revealed that two chromatin proteins (the histone HMtB and MtAlba) had lost their ability to bind DNA by specific arginine to isoleucine mutations after several passages. Histone HMtB could likely still interact with the DNA-binding histones HMtA1 and HMtA2 and fusion proteins were able to compact

DNA (188). What evolutionary pressure drove these mutations is unknown, but *M. thermautotrophicus* was able to alter the DNA binding properties of chromatin proteins without effects on growth rate under laboratory conditions.

In contrast to the archaea they evolved from, eukaryotes use histones and SMC proteins as their main, nearly exclusive chromatin proteins. As all Asgard archaea found so far contain genes for histones and Alba (table 1.1), it is likely that the first eukaryotic cells also had both. Several RNA-binding proteins in eukaryotes do contain an Alba domain (see above), but Alba, as a separate protein, is mostly lost. The hypothesis that archaea need at least two chromatin proteins was based on extended polymerization on DNA observed for reconstituted DNA-protein complexes *in vitro* (186). A second protein would be necessary to prevent unlimited polymerization of the first protein. An outstanding difference between archaeal and eukaryotic histones is the size of the 'nucleosomal complex'. Association of archaeal histones yields a theoretically endless hypernucleosome, while eukaryotic histones do not yield nucleosomes with more than eight histone subunits, although larger structures have been recently observed at telomers (189). The loss of 'endless' polymerizing histones might have resulted in a partial loss of Alba's cellular function. Instead, other regulatory mechanisms, such as PTMs on the histone tails, might have taken over as main regulatory mechanism in eukaryotes.

Thesis outline

In this thesis I describe studies on the structural properties of chromatin proteins from bacteria and archaea. I study several DNA bridging proteins across the tree of life and discuss the DNA binding properties of some archaeal histones. In Chapter 2, I discuss the structural and functional characteristics of a family of bacterial DNA bridging proteins, H-NS-like proteins. Also, I propose in this chapter that the protein charge distribution is an good predictor of the responsiveness of a protein to physico-chemical cues. In Chapter 3, I demonstrate that Rok is an atypical H-NS-like protein. I investigate the DNA structuring properties of Rok and demonstrate that the binding of the protein is not affected by environmental conditions. Also, I investigated its, naturally occurring, truncated derivative sRok and the interplay between the two proteins. We identify differences in the DNA binding characteristics of the two proteins and demonstrate that this translates in different regulons in B. subtilis. In Chapter 4, I describe studies on the binding of archaeal histones HMfA and HMfB from *M. fervidus* to a specific DNA sequence. This is relevant in the context of nucleosome-positioning *in vivo*, with possible impact on transcription. In Chapter 5 we investigate histones from *M. jannaschii*. Specifically, we show that one of these histones, MJ1647, is a novel atypical histone capable of DNA bridging. The ability to bridge is attributed to the presence of a C-terminal which promotes tetramerization. In Chapter 6, I discuss a toolbox for the biological expression and synthesis of archaeal histones with HA and HB from Heimdallarchaeota as example. Chapter 7 is a general discussion.

References

1. Stuger,R., Woldringh,C.L., van der Weijden,C.C., Vischer,N.O., Bakker,B.M., van Spanning,R.J., Snoep,J.L. and Westerhoff,H. V (2002) DNA supercoiling by gyrase is linked to nucleoid compaction. *Mol Biol Rep*, **29**, 79–82.

2. Dame,R.T. (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol Microbiol*, **56**, 858–870.

3. Luijsterburg,M.S., White,M.F., Van Driel,R. and Th. Dame,R. (2008) The major architects of chromatin: Architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol*, **43**, 393–418.

4. Luijsterburg,M.S., Noom,M.C., Wuite,G.J.L. and Dame,R.Th. (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *J Struct Biol*, **156**, 262–272.

5. Feklistov,A., Bae,B., Hauver,J., Lass-Napiorkowska,A., Kalesse,M., Glaus,F., Altmann,K.H., Heyduk,T., Landick,R. and Darst,S.A. (2017) RNA polymerase motions during promoter melting. *Science (1979)*, **356**, 863–866.

6. Karr,E.A. (2014) Transcription regulation in the third domain. *Adv Appl Microbiol*, **89**, 101–133.

7. Gehring,A.M., Walker,J.E. and Santangelo,T.J. (2016) Transcription regulation in archaea. *J Bacteriol*, **198**, 1906–1917.

8. Peeters,E., Peixeiro,N. and Sezonov,G. (2013) Cis-regulatory logic in archaeal transcription. In *Biochemical Society Transactions*. Biochem Soc Trans, Vol. 41, pp. 326–331.

9. Martinez-Rucobo,F.W., Sainsbury,S., Cheung,A.C.M. and Cramer,P. (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO Journal*, **30**, 1302–1310.

10. Qin,L., Erkelens,A.M., Ben Bdira,F. and Dame,R.T. (2019) The architects of bacterial DNA bridges: A structurally and functionally conserved family of proteins. *Open Biol*, **9**, 190223.

11. Dame,R.T., Rashid,F.-Z.M. and Grainger,D.C. (2020) Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat Rev Genet*, **21**, 227–242.

12. Azam,T.A. and Ishihama,A. (1999) Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *Journal of Biological Chemistry*, **274**, 33105–33113.

13. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251–260.

14. White,A.E., Hieb,A.R. and Luger,K. (2016) A quantitative investigation of linker histone interactions with nucleosomes and chromatin. *Sci Rep*, **6**.

15. Brouwer,T., Pham,C., Kaczmarczyk,A., De Voogd,W.J., Botto,M., Vizjak,P., Mueller-Planitz,F. and Van Noort,J. (2021) A critical role for linker DNA in higher-order folding of chromatin fibers. *Nucleic Acids Res*, **49**, 2537–2551.

16. Shilatifard,A. (2006) Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem*, **75**, 243–269.

17. Rossetto,D., Avvakumov,N. and Côté,J. (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics*, **7**, 1098–1108.

18. Louters,L. and Chalkley,R. (1985) Exchange of histones H1, H2A, and H2B in vivo. *Biochemistry*, **24**, 3080–3085.

19. Martire,S. and Banaszynski,L.A. (2020) The roles of histone variants in fine-tuning chromatin organization and function. *Nature Reviews Molecular Cell Biology 2020 21:9*, **21**, 522–541.

20. Čuboňová,L., Sandman,K., Hallam,S.J., DeLong,E.F. and Reeve,J.N. (2005) Histones in Crenarchaea. *J Bacteriol*, **187**, 5482–5485.

21. Henneman,B., van Emmerik,C., van Ingen,H. and Dame,R.T. (2018) Structure and function of archaeal histones. *PLoS Genet*, **14**, e1007582.

22. Ammar,R., Torti,D., Tsui,K., Gebbia,M., Durbic,T., Bader,G.D., Giaever,G. and Nislow,C. (2012) Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife*, **1**, e00078.

23. Warnecke,T., Becker,E.A., Facciotti,M.T., Nislow,C. and Lehner,B. (2013) Conserved Substitution Patterns around Nucleosome Footprints in Eukaryotes and Archaea Derive from Frequent Nucleosome Repositioning through Evolution. *PLoS Comput Biol*, **9**.

24. Arents,G. and Moudrianakis,E.N. (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A*, **92**, 11170.

25. Malik,H.S. and Henikoff,S. (2003) Phylogenomics of the nucleosome. *Nat Struct Biol*, **10**, 882–891.

26. Eme,L., Spang,A., Lombard,J., Stairs,C.W. and Ettema,T.J.G. (2017) Archaea and the origin of eukaryotes. 10.1038/nrmicro.2017.133.

27. Spang,A., Saw,J.H., Jørgensen,S.L., Zaremba-Niedzwiedzka,K., Martijn,J., Lind,A.E., Van Eijk,R., Schleper,C., Guy,L. and Ettema,T.J.G. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, **521**, 173–179.

28. Bulzu,P.A., Andrei,A.Ş., Salcher,M.M., Mehrshad,M., Inoue,K., Kandori,H., Beja,O., Ghai,R. and Banciu,H.L. (2019) Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. *Nature Microbiology 2019 4:7*, **4**, 1129–1137.

29. Pereira,S.L., Grayling,R.A., Lurz,R. and Reeve,J.N. (1997) Archaeal nucleosomes. *Proc Natl Acad Sci U S A*, **94**, 12633–12637.

30. Bailey,K.A., Pereira,S.L., Widom,J. and Reeve,J.N. (2000) Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. *J Mol Biol*, **303**, 25–34.

31. Bailey,K.A., Marc,F., Sandman,K. and Reeve,J.N. (2002) Both DNA and histone fold sequences contribute to archaeal nucleosome stability. *Journal of Biological Chemistry*, **277**, 9293–9301.

32. Maruyama,H., Harwood,J.C., Moore,K.M., Paszkiewicz,K., Durley,S.C., Fukushima,H., Atomi,H., Takeyasu,K. and Kent,N.A. (2013) An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. *EMBO Rep*, **14**, 711–7.

33. Mattiroli,F., Bhattacharyya,S., Dyer,P.N., White,A.E., Sandman,K., Burkhart,B.W., Byrne,K.R., Lee,T., Ahn,N.G., Santangelo,T.J., *et al.* (2017) Structure of histone-based chromatin in Archaea. *Science (1979)*, **357**, 609–612.

34. Henneman,B., Brouwer,T.B., Erkelens,A.M., Kuijntjes,G.-J., van Emmerik,C., van der Valk,R.A., Timmer,M., Kirolos,N.C.S., van Ingen,H., van Noort,J., *et al.* (2021) Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res*, **49**, 4338–4349.

35. Stevens,K.M., Swadling,J.B., Hocher,A., Bang,C., Gribaldo,S., Schmitz,R.A. and Warnecke,T. (2020) Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proc Natl Acad Sci U S A*, **117**, 33384–33395.

36. Sanders,T.J., Ullah,F., Gehring,A.M., Burkhart,B.W., Vickerman,R.L., Fernando,S., Gardner,A.F., Ben-Hur,A. and Santangelo,T.J. (2021) Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in Thermococcus kodakarensis. *Front Microbiol*, **12**, 1071.

37. Davey,C.A., Sargent,D.F., Luger,K., Maeder,A.W. and Richmond,T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol*, **319**, 1097–1113.

38. Yoshinaga,M., Nakayama,T. and Inagaki,Y. (2022) A novel structural maintenance of chromosomes (SMC)-related protein family specific to Archaea. *Front Microbiol*, **13**, 913088.

39. Cobbe,N. and Heck,M.M.S. (2004) The Evolution of SMC Proteins: Phylogenetic Analysis and Structural Implications. *Mol Biol Evol*, **21**, 332–347.

40. Gruber,S., Haering,C.H. and Nasmyth,K. (2003) Chromosomal cohesin forms a ring. *Cell*, **112**, 765–777.

41. Haering,C.H., Schoffnegger,D., Nishino,T., Helmhart,W., Nasmyth,K. and Löwe,J. (2004) Structure and Stability of Cohesin's Smc1-Kleisin Interaction. *Mol Cell*, **15**, 951–964.

42. Lammens,A., Schele,A. and Hopfner,K.P. (2004) Structural Biochemistry of ATP-Driven Dimerization and DNA-Stimulated Activation of SMC ATPases. *Current Biology*, **14**, 1778–1782.

43. Melby,T.E., Ciampaglio,C.N., Briscoe,G. and Erickson,H.P. (1998) The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. *J Cell Biol*, **142**, 1595–1604.

44. Anderson,D.E., Losada,A., Erickson,H.P. and Hirano,T. (2002) Condensin and cohesin display different arm conformations with characteristic hinge angles. *J Cell Biol*, **156**, 419–424.

45. Cobbe,N. and Heck,M.M.S. (2000) Review: SMCs in the world of chromosome biology - From prokaryotes to higher eukaryotes. *J Struct Biol*, **129**, 123–143.

46. Losada,A., Hirano,M. and Hirano,T. (1998) Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev*, **12**, 1986–1997.

47. Hirano,T. and Mitchison,T.J. (1994) A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell*, **79**, 449–458.

48. Lehmann,A.R., Walicka,M., Griffiths,D.J., Murray,J.M., Watts,F.Z., McCready,S. and Carr,A.M. (1995) The rad18 gene of Schizosaccharomyces pombe defines a new subgroup of the SMC superfamily involved in DNA repair. *Mol Cell Biol*, **15**, 7067–7080.

49. Fousteri,M.I. and Lehmann,A.R. (2000) A novel SMC protein complex in Schizosaccharomyces pombe contains the Rad18 DNA repair protein. *EMBO J*, **19**, 1691–1702.

50. Yoshinaga,M. and Inagaki,Y. (2021) Ubiquity and Origins of Structural Maintenance of Chromosomes (SMC) Proteins in Eukaryotes. *Genome Biol Evol*, **13**.

51. Niki,H., Jaffe,A., Imamura,R., Ogura,T. and Hiraga,S. (1991) The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of E. coli. *EMBO J*, **10**, 183–193.

52. Schwartz,M.A. and Shapiro,L. (2011) An SMC ATPase mutant disrupts chromosome segregation in Caulobacter. *Mol Microbiol*, **82**, 1359–1374.

53. Jensen,R.B. and Shapiro,L. (1999) The Caulobacter crescentus smc gene is required for cell cycle progression and chromosome segregation. *Proc Natl Acad Sci U S A*, **96**, 10661–10666.

54. Moriya,S., Tsujikawa,E., Hassan,A.K.M., Asai,K., Kodama,T. and Ogasawara,N. (1998) A Bacillus subtilis gene-encoding protein homologous to eukaryotic SMC motor protein is necessary for chromosome partition. *Mol Microbiol*, **29**, 179–187.

55. Wang,X., Tang,O.W., Riley,E.P. and Rudner,D.Z. (2014) The SMC condensin complex is required for origin segregation in Bacillus subtilis. *Curr Biol*, **24**, 287–292.

56. Petrushenko,Z.M., She,W. and Rybenkov,V. V. (2011) A new family of bacterial condensins. *Mol Microbiol*, **81**, 881–896.

57. Herrmann,U. and Soppa,J. (2002) Cell cycle-dependent expression of an essential SMC-like protein and dynamic chromosome localization in the archaeon Halobacterium salinarum. *Mol Microbiol*, **46**, 395–409.

58. Ettema,T.J.G., Lindås,A.C. and Bernander,R. (2011) An actin-based cytoskeleton in archaea. *Mol Microbiol*, **80**, 1052–1061.

59. Takemata,N., Samson,R.Y. and Bell,S.D. (2019) Physical and Functional Compartmentalization of Archaeal Chromosomes. *Cell*, **179**, 165-179.e18.

60. Hommais,F., Krin,E., Laurent-Winter,C., Soutourina,O., Malpertuy,A., Le Caer,J.P., Danchin,A. and Bertin,P. (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol Microbiol*, **40**, 20–36.

61. Navarre,W.W., Porwollik,S., Wang,Y., McClelland,M., Rosen,H., Libby,S.J. and Fang,F.C. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science (1979)*, **313**, 236–238.

62. Arold,S.T., Leonard,P.G., Parkinson,G.N. and Ladbury,J.E. (2010) H-NS forms a superhelical protein scaffold for DNA condensation. *Proceedings of the National Academy of Sciences*, **107**, 15728–15732.

63. Gordon,B.R., Li,Y., Cote,A., Weirauch,M.T., Ding,P., Hughes,T.R., Navarre,W.W., Xia,B. and Liu,J. (2011) Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc Natl Acad Sci U S A*, **108**, 10690–10695.

64. Dame,R.T., Wyman,C. and Goosen,N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res*, **28**, 3504–3510.

65. Dame,R.T., Noom,M.C. and Wuite,G.J.L. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature*, **444**, 387–390.

66. Amit,R., Oppenheim,A.B. and Stavans,J. (2003) Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophys J*, **84**, 2467–2473.

67. Liu,Y., Chen,H., Kenney,L.J. and Yan,J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev*, **24**, 339–344.

68. van der Valk,R.A., Vreede,J., Qin,L., Moolenaar,G.F., Hofmann,A., Goosen,N. and Dame,R.T. (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-Bridging activity. *Elife*, **6**, e27369.

69. Boudreau,B.A., Hron,D.R., Qin,L., van der Valk,R.A., Kotlajich,M. V, Dame,R.T. and Landick,R. (2018) StpA and Hha stimulate pausing by RNA polymerase by promoting DNA–DNA bridging of H-NS filaments. *Nucleic Acids Res*, **46**, 5525–5546.

70. Aravind,L., Iyer,L.M. and Anantharaman,V. (2003) The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biol*, **4**, R64.

71. Chêne,A., Vembar,S.S., Rivière,L., Lopez-Rubio,J.J., Claes,A., Siegel,T.N., Sakamoto,H., Scheidig-Benatar,C., Hernandez-Rivas,R. and Scherf,A. (2012) PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. *Nucleic Acids Res*, **40**, 3066–3077.

72. Goyal,M., Alam,A., Iqbal,M.S., Dey,S., Bindu,S., Pal,C., Banerjee,A., Chakrabarti,S. and Bandyopadhyay,U. (2012) Identification and molecular characterization of an Alba-family protein from human malaria parasite Plasmodium falciparum. *Nucleic Acids Res*, **40**, 1174–1190.

73. Goyal,M., Banerjee,C., Nag,S. and Bandyopadhyay,U. (2016) The Alba protein family: Structure and function. *Biochim Biophys Acta Proteins Proteom*, **1864**, 570–583.

74. Grote,M., Dijk,J. and Reinhardt,R. (1986) Ribosomal and DNA binding proteins of the thermoacidophilic archaebacterium Sulfolobus acidocaldarius. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, **873**, 405–413.

75. Wardleworth,B.N., Russell,R.J.M., Bell,S.D., Taylor,G.L. and White,M.F. (2002) Structure of Alba: An archaeal chromatin protein modulated by acetylation. *EMBO Journal*, **21**, 4654–4662.

76. Tanaka,T., Padavattan,S. and Kumarevel,T. (2012) Crystal structure of archaeal chromatin protein alba2-double-stranded DNA complex from Aeropyrum pernix K1. *Journal of Biological Chemistry*, **287**, 10394–10402.

77. Xue,H., Guo,R., Wen,Y., Liu,D. and 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**, 3929–3933.

78. Bell,S.D., Botting,C.H., Wardleworth,B.N., Jackson,S.P. and White,M.F. (2002) The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science (1979)*, **296**, 148–151.

79. Cao,J., Wang,Q., Liu,T., Peng,N. and 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**, 192–208.

80. Guo,R., Xue,H. and Huang,L. (2003) Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo. *Mol Microbiol*, **50**, 1605–1615.

81. Zhang,N., Guo,L. and Huang,L. (2020) The Sac10b homolog from Sulfolobus islandicus is an RNA chaperone. *Nucleic Acids Res*, **48**, 9273–9284.

82. Guo,L., Ding,J., Guo,R., Hou,Y., Wang,D.C. and Huang,L. (2014) Biochemical and structural insights into RNA binding by Ssh10b, a member of the highly conserved Sac10b protein family in Archaea. *J Biol Chem*, **289**, 1478–1490.

83. Lurz,R., Grote,M., Dijk,J., Reinhardt,R. and Dobrinski,B. (1986) Electron microscopic study of DNA complexes with proteins from the Archaebacterium Sulfolobus acidocaldarius. *EMBO J*, **5**, 3715–3721.

84. Jelinska,C., Petrovic-Stojanovska,B., Ingledew,W.J. and White,M.F. (2010) Dimer-dimer stacking interactions are important for nucleic acid binding by the archaeal chromatin protein Alba. *Biochem J*, **427**, 49–55.

85. Wardleworth,B.N., Russell,R.J.M., Bell,S.D., Taylor,G.L. and White,M.F. (2002) Structure of Alba: An archaeal chromatin protein modulated by acetylation. *EMBO Journal*, **21**, 4654–4662.

86. Jelinska,C., Conroy,M.J., Craven,C.J., Hounslow,A.M., Bullough,P.A., Waltho,J.P., Taylor,G.L. and White,M.F. (2005) Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure*, **13**, 963–971.

87. Laurens,N., Driessen,R.P.C., Heller,I., Vorselen,D., Noom,M.C., Hol,F.J.H., White,M.F., Dame,R.T. and Wuite,G.J.L. (2012) Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat Commun*, **3**, 2330.

88. Chen,L., Chen,L.R., Zhou,X.E., Wang,Y., Kahsai,M.A., Clark,A.T., Edmondson,S.P., Liu,Z.J., Rose,J.P., Wang,B.C., *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**, 73–91.

89. Edmondson,S.P., Kahsai,M.A., Gupta,R. and Shriver,J.W. (2004) Characterization of Sac10a, a hyperthermophile DNA-binding protein from Sulfolobus acidocaldarius. *Biochemistry*, **43**, 13026–13036.

90. Chen,L., Chen,L.R., Zhou,X.E., Wang,Y., Kahsai,M.A., Clark,A.T., Edmondson,S.P., Liu,Z.J., Rose,J.P., Wang,B.C., *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**, 73–91.

91. Biyani,K., Kahsai,M.A., Clark,A.T., Armstrong,T.L., Edmondson,S.P. and Shriver,J.W. (2005) Solution structure, stability, and nucleic acid binding of the hyperthermophile protein Sso10b2. *Biochemistry*, **44**, 14217–14230.

92. Driessen,R.P.C., Lin,S.N., Waterreus,W.J., Van Der Meulen,A.L.H., Van Der Valk,R.A., Laurens,N., Moolenaar,G.F., Pannu,N.S., Wuite,G.J.L., Goosen,N., *et al.* (2016) Diverse architectural properties of Sso10a proteins: Evidence for a role in chromatin compaction and organization. *Sci Rep*, **6**, 29422.

93. Van Noort,J., Verbrugge,S., Goosen,N., Dekker,C. and Dame,R.T. (2004) Dual architectural roles of HU: Formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci U S A*, **101**, 6969–6974.

94. Dame,R.T. and Goosen,N. (2002) HU: Promoting or counteracting DNA compaction? *FEBS Lett*, **529**, 151–156.

95. Dame,R.T., Luijsterburg,M.S., Krin,E., Bertin,P.N., Wagner,R. and Wuite,G.J.L. (2005) DNA Bridging: a Property Shared among H-NS-Like Proteins. *J Bacteriol*, **187**, 1845–1848.

96. Chen,J.M., Ren,H., Shaw,J.E., Wang,Y.J., Li,M., Leung,A.S., Tran,V., Berbenetz,N.M., Kocíncová,D., Yip,C.M., *et al.* (2008) Lsr2 of Mycobacterium tuberculosis is a DNA-bridging protein. *Nucleic Acids Res*, **36**, 2123–2135.

97. Qin,L., Bdira,F. Ben, Sterckx,Y.G.J., Volkov,A.N., Vreede,J., Giachin,G., van Schaik,P., Ubbink,M. and Dame,R.T. (2020) Structural basis for osmotic regulation of the DNA binding properties of H-NS proteins. *Nucleic Acids Res*, **48**, 2156–2172.

98. Erkelens,A.M., Qin,L., van Erp,B., Miguel-Arribas,A., Abia,D., Keek,H.G.J., Markus,D., Cajili,M.K.M., Schwab,S., Meijer,W.J.J., *et al.* (2022) The B. subtilis Rok protein is an atypical H-NS-like protein irresponsive to physico-chemical cues. *Nucleic Acids Res*, 10.1093/nar/gkac1064.

99. Woodcock,C.L., Skoultchi,A.I. and Fan,Y. (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res*, **14**, 17–25.

100. Hendzel,M.J., Lever,M.A., Crawford,E. and Th'Ng,J.P.H. (2004) The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. *J Biol Chem*, **279**, 20028–20034.

101. Srinivas Bharath,M.M., Ramesh,S., Chandra,N.R. and Rao,M.R.S. (2002) Identification of a 34 amino acid stretch within the C-terminus of histone H1 as the DNA-condensing domain by site-directed mutagenesis. *Biochemistry*, **41**, 7617–7627.

102. Zhou,B.R., Feng,H., Kato,H., Dai,L., Yang,Y., Zhou,Y. and Bai,Y. (2013) Structural insights into the histone H1-nucleosome complex. *Proc Natl Acad Sci U S A*, **110**, 19390–19395.

103. Hergeth,S.P. and Schneider,R. (2015) The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep*, **16**, 1439–1453.

104. Lee,M.S. and Craigie,R. (1994) Protection of retroviral DNA from autointegration: involvement of a cellular factor. *Proc Natl Acad Sci U S A*, **91**, 9823–9827.

105. Lee,M.S. and Craigie,R. (1998) A previously unidentified host protein protects retroviral DNA from autointegration. *Proc Natl Acad Sci U S A*, **95**, 1528.

106. Halfmann,C.T., Sears,R.M., Katiyar,A., Busselman,B.W., Aman,L.K., Zhang,Q., O'Bryan,C.S., Angelini,T.E., Lele,T.P. and Roux,K.J. (2019) Repair of nuclear ruptures requires barrier-to-autointegration factor. *J Cell Biol*, **218**, 2136–2149.

107. Sears,R.M. and Roux,K.J. (2020) Diverse cellular functions of barrier-to-autointegration factor and its roles in disease. *J Cell Sci*, **133**.

108. Cai,M., Huang,Y., Zheng,R., Wei,S.-Q., Ghirlando,R., Lee,M.S., Craigie,R., Gronenborn,A.M. and Clore,G.M. (1998) Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration. *Nature Structural Biology 1998 5:10*, **5**, 903.

109. Umland,T.C., Wei,S.Q., Craigie,R. and Davies,D.R. (2000) Structural basis of DNA bridging by Barrier-to-autointegration factor. *Biochemistry*, **39**, 9130–9138.

110. Lee,M.S. and Craigie,R. (1998) A previously unidentified host protein protects retroviral DNA from autointegration. *Proc Natl Acad Sci U S A*, **95**, 1528–1533.

111. Zheng,R., Ghirlando,R., Lee,M.S., Mizuuchi,K., Krause,M. and Craigie,R. (2000) Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *Proc Natl Acad Sci U S A*, **97**, 8997–9002.

112. Skoko,D., Li,M., Huang,Y., Mizuuchi,M., Cai,M., Bradley,C.M., Pease,P.J., Xiao,B., Marko,J.F., Craigie,R., *et al.* (2009) Barrier-to-autointegration factor (BAF) condenses DNA by looping. *Proc Natl Acad Sci U S A*, **106**, 16610–16615.

113. Bradley,C.M., Ronning,D.R., Ghirlando,R., Craigie,R. and Dyda,F. (2005) Structural basis for DNA bridging by barrier-to-autointegration factor. *Nat Struct Mol Biol*, **12**, 935–936.

114. Nichols,R.J., Wiebe,M.S. and Traktman,P. (2006) The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. *Mol Biol Cell*, **17**, 2451–2464.

115. Bengtsson,L. and Wilson,K.L. (2006) Barrier-to-autointegration factor phosphorylation on Ser-4 regulates emerin binding to lamin A in vitro and emerin localization in vivo. *Mol Biol Cell*, **17**, 1154–1163.

116. Montes De Oca,R., Lee,K.K. and Wilson,K.L. (2005) Binding of barrier to autointegration factor (BAF) to histone H3 and selected linker histones including H1.1. *Journal of Biological Chemistry*, **280**, 42252–42262.

117. Montes de Oca,R., Andreassen,P.R. and Wilson,K.L. (2011) Barrier-to-autointegration factor influences specific histone modifications. *Nucleus*, **2**, 580–590.

118. Hocher,A., Laursen,S.P., Radford,P., Tyson,J., Lambert,C., Stevens,K.M., Picardeau,M., Elizabeth Sockett,R. and Luger,K. (2023) Histone-organized chromatin in bacteria. *bioXrviv*, 10.1101/2023.01.26.525422.

119. Pan,C.Q., Finkel,S.E., Cramton,S.E., Feng,J.A., Sigman,D.S. and Johnson,R.C. (1996) Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J Mol Biol*, **264**, 675–695.

120. Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–306.

121. Swinger,K.K., Lemberg,K.M., Zhang,Y. and Rice,P.A. (2003) Flexible DNA bending in HU-DNA cocrystal structures. *EMBO Journal*, **22**, 3749–3760.

122. Mai,V.Q., Chen,X., Hong,R. and Huang,L. (1998) Small abundant DNA binding proteins from the thermoacidophilic archaeon Sulfolobus shibatae constrain negative DNA supercoils. *J Bacteriol*, **180**, 2560–2563.

123. Edmondson,S.P. and Shriver,J.W. (2001) DNA-binding proteins Sac7d and Sso7d from Sulfolobus. *Methods Enzymol*, **334**, 129–145.

124. Luo,X., Schwarz-Linek,U., Botting,C.H., Hensel,R., Siebers,B. and White,M.F. (2007) CC1, a novel crenarchaeal DNA binding protein. *J Bacteriol*, **189**, 403–409.

125. Guo,L., Feng,Y., Zhang,Z., Yao,H., Luo,Y., Wang,J. and Huang,L. (2008) Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea. *Nucleic Acids Res*, **36**, 1129–1137.

126. Grove,A. (2011) Functional evolution of bacterial histone-like HU proteins. *Curr Issues Mol Biol*, **13**, 1–12.

127. Azam,T.A., Iwata,A., Nishimura,A., Ueda,S. and Ishihama,A. (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J Bacteriol*, **181**, 6361–6370.

128. Stojkova,P., Spidlova,P. and Stulik,J. (2019) Nucleoid-associated protein Hu: A lilliputian in gene regulation of bacterial virulence. *Front Cell Infect Microbiol*, **9**, 159.

129. Boubrik,F. and Rouviere-Yaniv,J. (1995) Increased sensitivity to gamma irradiation in bacteria lacking protein HU. *Proc Natl Acad Sci U S A*, **92**, 3958.

130. Li,S. and Waters,R. (1998) Escherichia coli strains lacking protein HU are UV sensitive due to a role for HU in homologous recombination. *J Bacteriol*, **180**, 3750–3756.

131. Kamashev,D. and Rouviere-Yaniv,J. (2000) The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J*, **19**, 6527.

132. Bettridge,K., Verma,S., Weng,X., Adhya,S. and Xiao,J. (2021) Single-molecule tracking reveals that the nucleoid-associated protein HU plays a dual role in maintaining proper nucleoid volume through differential interactions with chromosomal DNA. *Mol Microbiol*, **115**, 12–27.

133. Hocher,A., Rojec,M., Swadling,J.B., Esin,A. and Warnecke,T. (2019) The DNA-binding protein HTa from thermoplasma acidophilum is an archaeal histone analog. *Elife*, **8**, e52542.

134. Hales,L.M., Gumport,R.I. and Gardner,J.F. (1994) Determining the DNA sequence elements required for binding integration host factor to two different target sites. *J Bacteriol*, **176**, 2999–3006.

135. Huo,Y.X., Zhang,Y.T., Xiao,Y., Zhang,X., Buck,M., Kolb,A. and Wang,Y.P. (2009) IHF-binding sites inhibit DNA loop formation and transcription initiation. *Nucleic Acids Res*, **37**, 3878–3886.

136. Deog Su Hwang and Kornberg,A. (1992) Opening of the replication origin of Escherichia coli by DnaA protein with protein HU or IHF. *Journal of Biological Chemistry*, **267**, 23083–23086.

137. Laxmikanthan,G., Xu,C., Brilot,A.F., Warren,D., Steele,L., Seah,N., Tong,W., Grigorieff,N., Landy,A. and Van Duyne,G.D. (2016) Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *Elife*, **5**.

138. Stella,S., Cascio,D. and Johnson,R.C. (2010) The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev*, **24**, 814–826.

139. Hancock,S.P., Stella,S., Cascio,D. and Johnson,R.C. (2016) DNA sequence determinants controlling affinity, stability and shape of DNA complexes bound by the nucleoid protein Fis. *PLoS One*, **11**, e0150189.

140. Grainger,D.C., Hurd,D., Goldberg,M.D. and Busby,S.J.W. (2006) Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic Acids Res*, **34**, 4642–4652.

141. Kelly,A., Goldberg,M.D., Carroll,R.K., Danino,V., Hinton,J.C.D. and Dorman,C.J. (2004) A global role for Fis in the transcriptional control of metabolism and type III secretion in Salmonella enterica serovar Typhimurium. *Microbiology (N Y)*, **150**, 2037–2053.

142. Schneider,R., Lurz,R., Lüder,G., Tolksdorf,C., Travers,A. and Muskhelishvili,G. (2001) An architectural role of the Escherichia coli chromatin protein FIS in organising DNA. *Nucleic Acids Res*, **29**, 5107–5114.

143. Browning,D.F., Cole,J.A. and Busby,S.J.W. (2000) Suppression of FNR-dependent transcription activation at the Escherichia coli nir promoter by Fis, IHF and H-NS: Modulation of transcription by a complex nucleo-protein assembly. *Mol Microbiol*, **37**, 1258–1269.

144. Chartier,F., Laine,B., Belaïche,D., Touzel,J.P. and Sautière,P. (1989) Primary structure of the chromosomal protein MC1 from the archaebacterium Methanosarcina sp. CHTI 55. *BBA - Gene Structure and Expression*, **1008**, 309–314.

145. Paquet,F., Culard,F., Barbault,F., Maurizot,J.C. and Lancelot,G. (2004) NMR solution structure of the archaebacterial chromosomal protein MC1 reveals a new protein fold. *Biochemistry*, **43**, 14971–14978.

146. Paquet,F., Loth,K., Meudal,H., Culard,F., Genest,D. and Lancelot,G. (2010) Refined solution structure and backbone dynamics of the archaeal MC1 protein. *FEBS Journal*, **277**, 5133–5145.

147. Paquet,F., Delalande,O., Goffinont,S., Culard,F., Loth,K., Asseline,U., Castaing,B. and Landon,C. (2014) Model of a DNA-protein complex of the architectural monomeric protein MC1 from Euryarchaea. *PLoS One*, **9**.

148. Culard,F., Laine,B., Sautiére,P. and Maurizot,J.C. (1993) Stoichiometry of the binding of chromosomal protein MCl from the archaebacterium, Methanosarcina spp. CHTI55, to DNA. *FEBS Lett*, **315**, 335–339.

149. De Vuyst,G., Aci,S., Genest,D. and Culard,F. (2005) Atypical recognition of particular DNA sequences by the archaeal chromosomal MC1 protein. *Biochemistry*, **44**, 10369–10377.

150. Paradinas,C., Gervais,A., Maurizot,J.C. and Culard,F. (1998) Structure-specific binding recognition of a methanogen chromosomal protein. *Eur J Biochem*, **257**, 372–379.

151. Cam,E. Le, Culard,F., Larquet,E., Delain,E. and Cognet,J.A.H. (1999) DNA bending induced by the archaebacterial histone-like protein MC1. *J Mol Biol*, **285**, 1011–1021.

152. Loth,K., Largillière,J., Coste,F., Culard,F., Landon,C., Castaing,B., Delmas,A.F. and Paquet,F. (2019) New protein-DNA complexes in archaea: a small monomeric protein induces a sharp V-turn DNA structure. *Sci Rep*, **9**, 14253.

153. Manzur,K.L. and Zhou,M.M. (2005) An archaeal SET domain protein exhibits distinct lysine methyltransferase activity towards DNA-associated protein MC1-alpha. *FEBS Lett*, **579**, 3859–3865.

154. Napoli,A., Zivanovic,Y., Bocs,C., Buhler,C., Rossi,M., Forterre,P. and Ciaramella,M. (2002) DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of Sulfolobus solfataricus. *Nucleic Acids Res*, **30**, 2656–2662.

155. Baumann,H., Knapp,S., Lundback,T., Ladenstein,R. and Hard,T. (1994) Solution structure and DNA-binding properties of a thermostable protein from the archaeon Sulfolobus solfataricus. *Nat Struct Biol*, **1**, 808–819.

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

157. Zhang,Z., Gong,Y., Guo,L., Jiang,T. and Huang,L. (2010) Structural insights into the interaction of the crenarchaeal chromatin protein Cren7 with DNA. *Mol Microbiol*, **76**, 749–759.

158. Gao,Y.G., Su,S.Y., Robinson,H., Padmanabhan,S., Lim,L., McCrary,B.S., Edmondson,S.P., Shriver,J.W. and Wang,A.H.J. (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat Struct Biol*, **5**, 782–786.

159. Robinson,H., Gao,Y.G., McCrary,B.S., Edmondson,S.P., Shriver,J.W. and Wang,A.H.J. (1998) The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature*, **392**, 202–205.

160. Vorontsov,E.A., Rensen,E., Prangishvili,D., Krupovic,M. and Chamot-Rooke,J. (2016) Abundant Lysine Methylation and N-Terminal Acetylation in Sulfolobus islandicus Revealed by Bottom-Up and Top-Down Proteomics. *Mol Cell Proteomics*, **15**, 3388–3404.

161. Niu,Y., Xia,Y., Wang,S., Li,J., Niu,C., Li,X., Zhao,Y., Xiong,H., Li,Z., Lou,H., *et al.* (2013) A prototypic lysine methyltransferase 4 from archaea with degenerate sequence specificity methylates chromatin proteins Sul7d and Cren7 in different patterns. *J Biol Chem*, **288**, 13728–13740.

162. Hardy,C.D. and Martin,P.K. (2008) Biochemical characterization of DNA-binding proteins from Pyrobaculum aerophilum and Aeropyrum pernix. *Extremophiles*, **12**, 235–246.

163. Lu,J., Kobayashi,R. and Brill,S.J. (1996) Characterization of a high mobility group 1/2 homolog in yeast. *J Biol Chem*, **271**, 33678–33685.

164. Teo,S.-H., Grasser,K.D. and Thomas,J.O. (1995) Differences in the DNA-binding properties of the HMG-box domains of HMG1 and the sex-determining factor SRY. *Eur J Biochem*, **230**, 943–950.

165. Love,J.J., Li,X., Case,D.A., Giese,K., Crosschedl,R. and Wright,P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791–795.

166. Werner,M.H., Huth,J.R., Gronenborn,A.M. and Marius Clore,G. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, **81**, 705–714.

167. Lnenicek-Allen,M., Read,C.M. and Crane-Robinson,C. (1996) The DNA bend angle and binding affinity of an HMG box increased by the presence of short terminal arms. *Nucleic Acids Res*, **24**, 1047.

168. Payet,D. and Travers,A. (1997) The acidic tail of the high mobility group protein HMG-D modulates the structural selectivity of DNA binding. *J Mol Biol*, **266**, 66–75.

169. Webb,M. and Thomas,J.O. (1999) Structure-specific binding of the two tandem HMG boxes of HMG1 to four-way junction DNA is mediated by the A domain. *J Mol Biol*, **294**, 373–387.

170. Travers,A. (2000) Recognition of distorted DNA structures by HMG domains. *Curr Opin Struct Biol*, **10**, 102–109.

171. Zwilling,S., König,H. and Wirth,T. (1995) High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. *EMBO J*, **14**, 1198.

172. Zappavigna,V., Falciola,L., Citterich,M.H., Mavilio,F. and Bianchi,M.E. (1996) HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. *EMBO J*, **15**, 4981–4991.

173. Jayaraman,L., Moorthy,N.C., Murthy,K.G.K., Manley,J.L., Bustin,M. and Prives,C. (1998) High mobility group protein-1 (HMG-1) is a unique activator of p53. *Genes Dev*, **12**, 462.

174. Decoville,M., Giraud-Panis,M.J., Mosrin-Huaman,C., Leng,M. and Locker,D. (2000) HMG boxes of DSP1 protein interact with the Rel homology domain of transcription factors. *Nucleic Acids Res*, **28**, 454.

175. Boonyaratanakornkit,V., Melvin,V., Prendergast,P., Altmann,M., Ronfani,L., Bianchi,M.E., Taraseviciene,L., Nordeen,S.K., Allegretto,E.A. and Edwards,D.P. (1998) High-Mobility Group Chromatin Proteins 1 and 2 Functionally Interact with Steroid Hormone Receptors To Enhance Their DNA Binding In Vitro and Transcriptional Activity in Mammalian Cells. *Mol Cell Biol*, **18**, 4471.

176. Giese,K., Kingsley,C., Kirshner,J.R. and Grosschedl,R. (1995) Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev*, **9**, 995–1008.

177. Ge,H. and Roeder,R.G. (1994) The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. *Journal of Biological Chemistry*, **269**, 17136–17140.

178. Sutrias-Grau,M., Bianchi,M.E. and Bernués,J. (1999) High mobility group protein 1 interacts specifically with the core domain of human TATA box-binding protein and interferes with transcription factor IIB within the pre-initiation complex. *J Biol Chem*, **274**, 1628–1634.

179. Bustin,M. and Reeves,R. (1996) High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog Nucleic Acid Res Mol Biol*, **54**, 35–100.

180. Thomas,J.O. and Travers,A.A. (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci*, **26**, 167–174.

181. Tian,L., Zhang,Z., Wang,H., Zhao,M., Dong,Y. and Gong,Y. (2016) Sequence-Dependent T:G Base Pair Opening in DNA Double Helix Bound by Cren7, a Chromatin Protein Conserved among Crenarchaea. *PLoS One*, **11**.

182. Winardhi,R.S., Fu,W., Castang,S., Li,Y., Dove,S.L. and Yan,J. (2012) Higher order oligomerization is required for H-NS family member MvaT to form gene-silencing nucleoprotein filament. *Nucleic Acids Res*, **40**, 8942–8952.

183. Liu,Y., Chen,H., Kenney,L.J. and Yan,J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev*, **24**, 339–44.

184. Qu,Y., Lim,C.J., Whang,Y.R., Liu,J. and Yan,J. (2013) Mechanism of DNA organization by Mycobacterium tuberculosis protein Lsr2. *Nucleic Acids Res*, **41**, 5263–5272.

185. Kotlajich,M. V, Hron,D.R., Boudreau,B.A., Sun,Z., Lyubchenko,Y.L. and Landick,R. (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife*, **2015**, 4970.

186. Sandman,K. and Reeve,J.N. (2005) Archaeal chromatin proteins: Different structures but common function? *Curr Opin Microbiol*, **8**, 656–661.

187. Cajili,M.K.M. and Prieto,E.I. (2022) Interplay between Alba and Cren7 Regulates Chromatin Compaction in Sulfolobus solfataricus. *Biomolecules*, **12**.

188. Sandman,K., Louvel,H., Samson,R.Y., Pereira,S.L. and Reeve,J.N. (2008) Archaeal chromatin proteins histone HMtB and Alba have lost DNA-binding ability in laboratory strains of Methanothermobacter thermautotrophicus. *Extremophiles*, **12**, 811–817.

189. Soman,A., Wong,S.Y., Korolev,N., Surya,W., Lattmann,S., Vogirala,V.K., Chen,Q., Berezhnoy,N. V., van Noort,J., Rhodes,D., *et al.* (2022) Columnar structure of human telomeric chromatin. *Nature 2022 609:7929*, **609**, 1048–1055.

Chapter 2

The architects of bacterial DNA bridges: a structurally and functionally conserved family of proteins

This chapter has been adapted from: Qin, L., Erkelens, A.M., Ben Bdira, F. and Dame, R.T., The architects of bacterial DNA bridges: a structurally and functionally conserved family of proteins, *Open Biology*, **9**:, 190223

Abstract

Every organism across the tree of life compacts and organizes its genome with architectural chromatin proteins. While eukaryotes and archaea express histone proteins, the organization of bacterial chromosomes is dependent on nucleoid-associated proteins (NAPs). In *Escherichia coli* and other proteobacteria, the histone-like nucleoid structuring protein (H-NS) acts as a global genome organizer and gene regulator. Functional analogs of H-NS have been found in other bacterial species: MvaT in *Pseudomonas* species, Lsr2 in actinomycetes and Rok in *Bacillus* species. These proteins complement *hns*$^-$ phenotypes and have similar DNA-binding properties, despite their lack of sequence homology. In this review, we focus on the structural and functional characteristics of these four architectural proteins. They can bridge DNA duplexes, which is key to genome compaction and gene regulation and respond to changing environmental conditions. Structurally the domain organization and charge distribution of these proteins is conserved, which we suggest is the basis of their conserved environment-responsive behavior. These observations could be used to find and validate new members of this protein family and to predict their response to environmental changes.

Introduction

All organisms compact and organize their genomic DNA. Structuring of the genome is achieved by the action of small, basic architectural proteins that interact with DNA. These proteins wrap, bend and bridge DNA duplexes. Despite the lack of both sequence and structural homology between architectural proteins in species across the tree of life, the basic concepts appear conserved, with all organisms harboring functional analogs(1). An essential feature of genome organization is its intrinsic coupling to genome transactions, such that a process like gene expression is both dependent upon chromatin structure and a driving factor in chromatin (re)organization(2, 3). The structure of chromatin is affected by environmental signals, which can be translated into altered gene expression(4).

No doubt, the best-known architectural proteins are the histones expressed by eukaryotes. The binding of these proteins to DNA yields nucleosomes in which DNA is wrapped around an octameric histone protein core. Aided by other architectural proteins, these nucleosomal fibers are further organized into higher-order structures(1, 5–7). Histone H1 and BAF (barrier-to-autointegration factor) are examples of eukaryotic architectural proteins capable of bridging DNA(8, 9). In addition, Structural Maintenance of Chromosome (SMC) proteins (e.g. cohesin and condensin) act upon chromatin, forming large chromatin loops by bridging, at the expense of ATP(10). SMC proteins are the only chromatin proteins that are universally conserved(11, 12). Finally, eukaryotes express small proteins that bend DNA, such as HMG-box proteins(13). Archaea also express histones. Unlike their eukaryotic counterparts, archaeal histones assemble into oligomeric filaments along DNA, yielding hypernucleosomes(14–17). In addition, they express DNA bridging proteins such as Alba (acetylation lowers binding affinity), which can both form nucleofilaments and bridge DNA, depending on the protein:DNA stoichiometry(18, 19). Some archaeal species lack histones and these express DNA bending proteins instead, like Cren7 and Sul7(20).

Bacteria lack homologs of the histone proteins expressed by eukaryotes and archaea. The organization of bacterial genomes is dependent on a group of architectural proteins collectively referred to as nucleoid-associated proteins (NAPs). At least 12 NAPs have been described for *Escherichia coli* and closely related species(21–23). A shared feature of many of these proteins is their ability to bend DNA. Examples include the histone-like protein from strain U93 (HU), integration host factor (IHF) and the factor for inversion stimulation (Fis)(24–26). The histone-like nucleoid structuring protein (H-NS) has an overarching role in the organization of the *E. coli* genome. It acts as a global regulator of gene expression: 5-10% of *E. coli* genes are affected, primarily repressed,

by H-NS(27). Due to its preference for A/T-rich DNA, it specifically targets and silences horizontally acquired genes, a process referred to as xenogeneic silencing(28). Key to the role of H-NS in both processes is the formation of nucleofilaments along the DNA and protein-mediated DNA-DNA bridges(29–31). H-NS-like proteins are passive DNA bridgers in contrast to SMC proteins which are active, ATP-driven DNA bridgers (figure 2.1).

Over the last two decades, functional homologs of H-NS have been identified in other bacterial species. Despite low sequence similarity, these proteins have similar DNA binding properties, resulting in the formation of structurally and functionally similar protein-DNA complexes. This ability is elegantly demonstrated by the genetic complementation of *hns*$^{-}$ phenotypes (like mucoidy, motility and β-glucoside utilization) in *E. coli* by MvaT from *Pseudomonas* species and Lsr2 from *Mycobacterium* and related actinomycetes(32, 33). *In vitro* both proteins can also bridge DNA in a manner similar to H-NS(29, 34, 35) (figure 2.1). MvaT regulates hundreds of genes in *P. aeruginosa* and Lsr2 binds to one fifth of the *M. tuberculosis* genome, especially to horizontally acquired genes(36–39). These properties endow them with functions as global gene regulators and spatial chromatin organizers. A newly proposed functional homolog of H-NS is the repressor of *comK* protein (Rok) of *Bacillus subtilis.* This classification is primarily based on the observation that Rok binds extended regions of the *B. subtilis* genome and especially A/T rich regions acquired by horizontal gene transfer, which it aids to repress (40). Also, Rok decreases intra- and interspecies chromosomal transformation (41). This specific property of silencing foreign genes makes Rok, like H-NS, MvaT and Lsr2, a xenogeneic silencer. It is also associated with a large subset of chromosomal domain boundaries identified in *B. subtilis* by Hi-C(42). As Rok forms interacting clusters creating chromosomal loops, this implies a role as DNA bridging architectural protein.

In this review we focus on the properties of DNA bridging proteins in bacteria with a proposed role in genome architecture and gene regulation: H-NS, MvaT, Lsr2 and Rok. We describe and compare their structure and function to define conserved features. Also, we discuss the mechanisms by which the architectural and regulatory properties of these proteins are modulated.

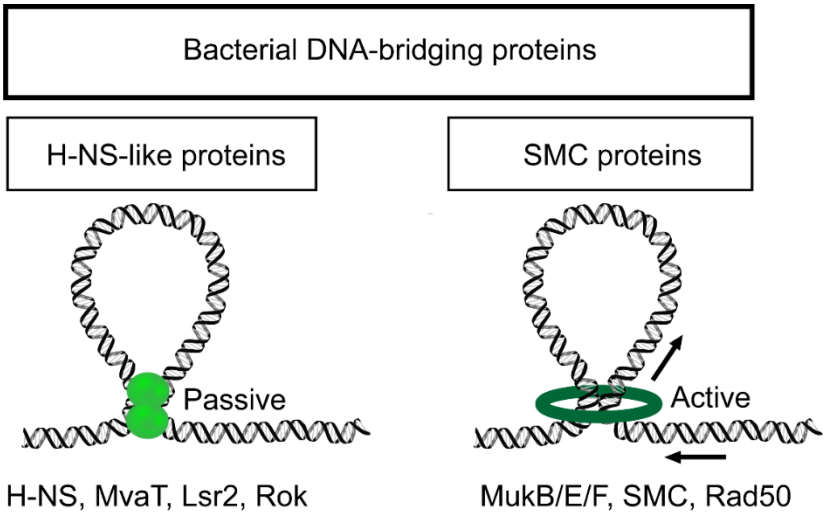


Figure 2.1 Bacterial DNA-bridging proteins. Two types can be distinguished: passive DNA bridgers, such as H-NS-like proteins (light green), which bind distant segments of DNA duplexes and bring them together, and active DNA bridgers, such as SMC proteins (dark green), that are able to connect two double-stranded DNA segments, translocating along the DNA molecule with motor activity resulting from ATP hydrolysis. Note that the exact molecular mechanisms by which SMC proteins operate and are involved in loop formation only start to be defined and are a topic of much discussion.

Fold topology of H-NS-like proteins

Structural studies have revealed that H-NS, Lsr2 and MvaT harbor similar functional modules: 1) an N-terminal oligomerization domain consisting of two dimerization sites, 2) a C-terminal DNA binding domain and 3) an unstructured linker region (figure 2.2 a,b and c)(43–47). For Rok, a similar overall domain architecture has been found: the C-terminal domain is capable of DNA binding and the N-terminal domain is responsible for oligomerization(48).

The N-terminal domain

The N-terminal domain of H-NS and MvaT is involved in the formation of oligomers, which is a property essential for gene repression(38, 49). Both Lsr2 and Rok are capable of oligomerization, but it is currently unknown whether oligomerization is required for gene regulation(43, 48). As the N-terminal structure of most of these DNA bridging proteins is known, differences and similarities in the mechanism of forming high-order complexes have become evident. H-NS of *Salmonella typhimurium* has two dimerization sites in the N-terminal domain (1-83)(44). The N-terminal dimerization domain (site 1, 1-40) is formed by a “hand-shake” topology between α1 and α2 and part of α3. The central dimerization domain (site 2, 57–83), has two α-helices α3 and α4 that form a helix-turn-helix dimerization interface. H-NS dimers are formed via site 1 in a tail-to-tail manner, which can oligomerize via site 2 via head-to-head association (figure 2.2a). The resulting crystal structure is superhelical. Therefore, it was proposed that DNA-H-NS-DNA filaments involve superhelical wrapping of DNA around an oligomeric protein core. However, apart from the X-ray crystal structure(44) and molecular simulations (50), there is no evidence for this type of H-NS nucleofilaments organization.

The crystal structure of the MvaT homolog, TurB from *Pseudomonas putida,* revealed a similar fold topology of the N-terminal dimerization site 2 as that of H-NS (45). In contrast, site 1 exhibits a standard “coiled-coil” architecture in MvaT/TurB, whereas H-NS due to the presence of two additional N-terminal helices (α1 and α2 in H-NS) compared to TurB/MvaT exhibits a “hand-shake” topology (figure 2.2). Despite this difference in site 1, both proteins form a head-head and tail-tail dimers organization in their protein filaments.

The N-terminal structure of Lsr2 from *Mycobacterium tuberculosis*, however, is completely different from that of H-NS and MvaT(43). The flexible N-terminus is followed by a β-sheet formed by two anti-parallel β-strands and a kinked α-helix. When forming dimers, the two β-sheets of the monomers align to form a four-stranded antiparallel β-sheet with an antiparallel arrangement of the α-helices on the opposite sides of the sheet

(figure 2.2c). Notably, oligomerization does not occur with the first four amino acids of Lsr2 of *M. tuberculosis* present, but is triggered by trypsin cleavage, removing these residues (43). The oligomerization between Lsr2 dimers occurs through an antiparallel association between two N-terminal β-strands from adjacent monomers (figure 2.2c). The triggering of Lsr2 oligomerization by proteolysis indicates that this process is possibly controlled via protease activity *in vivo*, offering a mechanism for genome protection by Lsr2 under stress conditions(51). Note, however, that these four amino acids are not highly conserved and, for example, are lacking in *Mycobacterium sinensis* Lsr2 (Genbank: AEF37887.1).

Although the 3D structure of Rok's N-terminal domain has not been experimentally determined, it is predicted to contain two α-helices (α1 and α2) and a part of the first α-helix (residues 1-43) is predicted to form a "coiled coil" dimerization motif (52, 53) (figure 2.2d), similar to MvaT. Based on these secondary structure predictions, it is plausible that Rok exhibits higher-order oligomerization with a structural organization resembling that of TurB/MvaT.

The C-terminal domain

The C-terminal domain of all four proteins recognizes and binds to DNA. *In vivo* the DNA segments bound are generally AT-rich compared to other parts of the genome. Although H-NS and Lsr2 differ in the overall structure of the C-terminal domain, both proteins recognize the minor groove of DNA with a similar 'AT-hook-like' motif composed of three consecutive residues 'Q/RGR'(46) (figure 2.2a,c). H-NS and Lsr2(-like proteins) generally favor similar AT-rich DNA target sequences, both with a preference for TpA steps over A tracts(46, 54, 55). This can be related to the width of the minor groove. A-tracts narrow the minor groove compared to TpA steps, while GC-rich sequences result in a wider minor groove. TpA steps result in a favorable width for H-NS and Lsr2 binding to the DNA(46).

The C-terminal domain of MvaT exhibits a similar overall fold as H-NS but has a different DNA binding mechanism(56) (figure 2.2b). The C-terminus of MvaT recognizes AT-rich DNA via both the "AT-pincer" motif consisting of three non-continuous residues 'R-G-N' targeting minor groove DNA and a "lysine network" interacting with the DNA backbone by multiple positive charges. MvaT has similar preferences in binding DNA sequences as H-NS and Lsr2, preferring TpA steps over A-tracts(54, 56). MvaT is, however, more tolerant to G/C interruptions in the DNA sequence than H-NS and Lsr2.

Binding of the C-terminal domain of H-NS and Lsr2 to DNA causes no notable changes in DNA conformation(37, 46), whereas the C-terminus of MvaT triggers significant distortions in the DNA molecule(56). It is likely that the 'AT-hook motif' of H-

NS and Lsr2 forms a narrow crescent-shaped structure that inserts into the minor groove without significantly interrupting the DNA helical trajectory. When bound by the C-terminal domain of MvaT, the minor groove of DNA is expanded, leading to a significant rearrangement in base-stacking(47). Therefore, the binding of full-length MvaT dimers to DNA results in DNA bending (57).

The structure of the C-terminal DNA-binding domain of Rok reveals that it employs a winged helix domain fold utilizing a unique DNA recognition mechanism different from the other three proteins(48) (figure 2.2d). Rather than using an 'AT-hook-like' motif or 'AT-pincer' motif, the C-terminus of Rok targets the DNA minor groove via the three non-continuous residues N-T-R. As in the case of MvaT, DNA binding is stabilized by a hydrogen bond network between several lysine residues and the phosphate groups of the DNA backbone(48). Like H-NS, Lsr2 and MvaT, Rok interacts with AT-rich DNA sequences with a preference for TpA steps(48, 54). Rok has selectivity towards some specific DNA sites, comparable with the affinity of H-NS for its high-affinity sites, where the highest affinity is noticed for AACTA and TACTA sequences (58). Compared to MvaT, Rok induces a more pronounced conformational change in its target DNA substrate. Rok binding leads to bending of DNA by ~25 degrees(48).

The function of the conformational changes in target DNA induced by MvaT and Rok binding is unknown, but the changes may be of importance in gene regulation, where often multiple architectural proteins operate in concert. An example is the reversion of Rok repression by ComK at the *comK* promotor(59). It has been suggested that the DNA-bending by Comk reverses the conformational changes in the DNA induced by Rok.

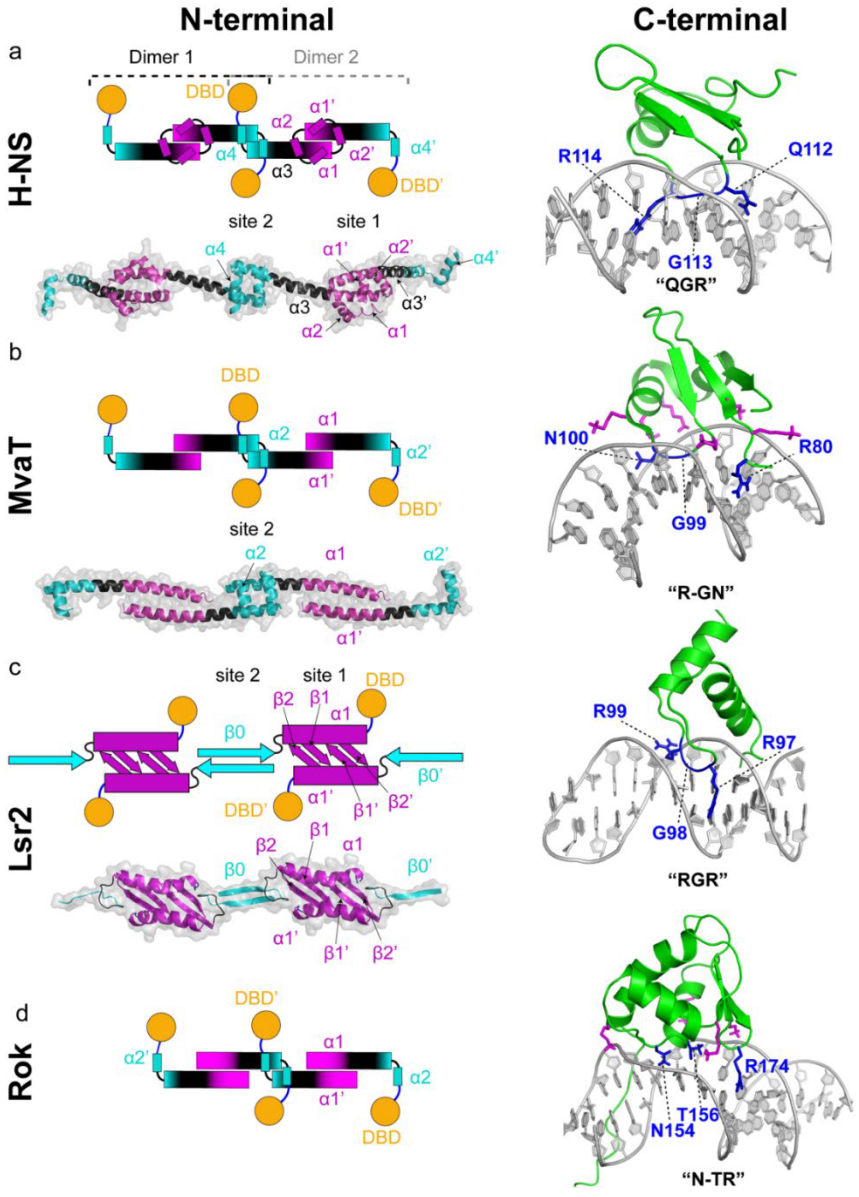


Figure 2.2 Fold topology and oligomerization states of H-NS-like proteins. Left panels show models of the structures of the N-terminal oligomerization domains of (a) H-NS(44), (b) MvaT(45) and (c) Lsr2(43) as determined by crystallography. A schematic representation of the higher-order oligomerization states of H-NS-like proteins is shown above the crystal structures of the N-terminal domains. For Rok (d) the schematic representation of its higher-

order oligomerization state is based on secondary structure prediction. The dimerization sites (site 1) are shown in magenta and the oligomerization sites (site 2) in cyan. The DNA binding domains are shown in orange spheres and the linker regions in blue lines. Right panels show the NMR structures of the DNA binding domains of H-NS-like proteins and their DNA recognition mechanisms(48). The loops of the DNA binding motif are shown in blue and the residues involved in the direct interactions with the DNA minor groove and in complex stabilization are shown in sticks.

Protein-DNA complexes formed by H-NS-like proteins

Two types of protein-DNA complexes can be formed by H-NS-like proteins: 1) nucleoprotein filaments and 2) bridged complexes. Assembly of protein-DNA complexes by H-NS, MvaT and Lsr2 is believed to proceed via a multi-step process (figure 2.3). First, the C-terminal DNA-binding domain directs the protein to a high-affinity site (nucleation) (58). This step is likely assisted by the positively charged amino acid residues of the linker region, which interact with the DNA and recruit H-NS to bind non-specifically (60–62). This then allows H-NS to scan on DNA to search for the specific site where the C-terminal domain can engage with a higher affinity. Next, the proteins spread cooperatively along the DNA, forming a nucleoprotein filament by oligomerization through their N-terminal domains. If the surrounding conditions are favorable, these nucleoprotein filaments can interact with another DNA duplex to form a bridge. This occurs without changing the positions of the C-terminal domains, which are spaced ~10 bp (63). Both types of protein-DNA complexes are thought to play important roles in genome structuring and gene silencing. Evidence in support of the formation of nucleoprotein filaments comes from atomic force microscopy (AFM) and single-molecule studies which revealed that H-NS, Lsr2 and MvaT all form rigid protein-DNA filaments, suggestive of protein oligomerization along DNA(31, 64–67). DNA-DNA bridging has been visualized *in vitro* using microscopy(29, 34, 35) and corroborated using solution-based assays(30, 68). The ability to oligomerize is important for the function of these proteins in chromosome organization and gene regulation(38, 66, 67). To date, there are no indications that Rok rigidifies DNA suggesting that this protein might not be able to oligomerize along DNA. Nevertheless, the protein induces DNA compaction and is capable of DNA-DNA bridging ((69), Chapter 3).

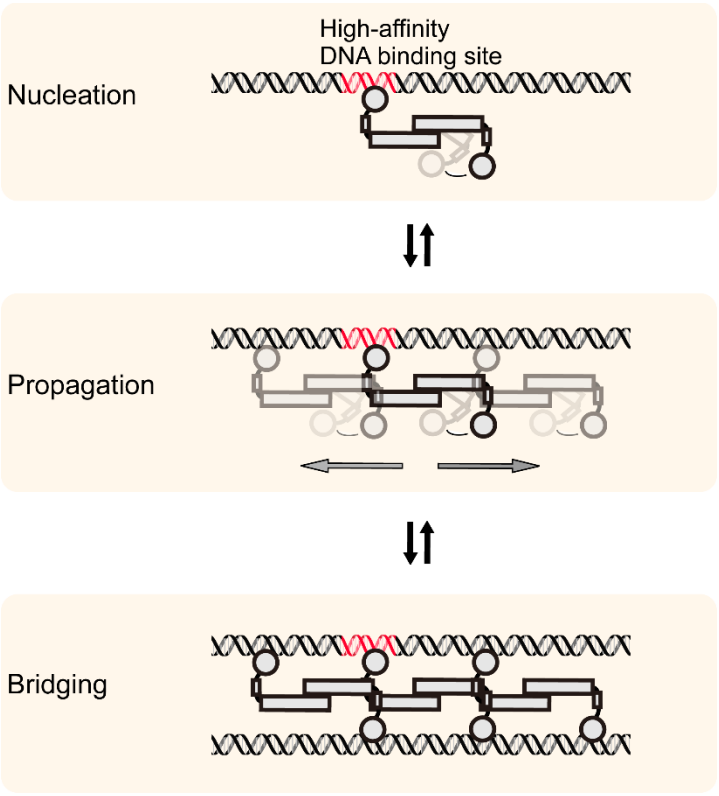


Figure 2.3: Assembly of functional protein-DNA complexes by H-NS-like proteins. DNA binding of H-NS-like proteins initiates at a nucleation site (high-affinity binding site, red) in the genome; H-NS-like proteins cooperatively propagate along DNA due to protein-protein interactions; DNA-protein-DNA bridging complex can be formed under favorable bridging conditions, bringing distant DNA duplexes together. Further propagation (not indicated in the figure) may occur in the bridged complex due to both protein-protein interactions and high effective local DNA concentration. Note that all steps are reversible, which is important for the modulation of the function of these proteins (section 5).

Following the initial observations that two types of complexes can be formed(35, 38, 65, 67, 70), the mechanism that drives the switch between these two DNA binding modes has long been elusive. Two recent studies on H-NS have provided the first mechanistic insights into this process (68, 71). The switch between the two DNA binding modes(65) involves a conformational change of the H-NS dimers from a half-open to an open conformation driven by Mg^{2+}(68). These conformational changes are modulated by the interactions between the N-terminal domain of H-NS and its C-terminal DNA binding

domain. Mutagenesis at the interface of these domains generated an H-NS variant no longer sensitive to Mg^{2+}, which can form filaments and bridge DNA(68). Recently, these interactions were confirmed by Arold and co-workers using H-NS truncated domains(71). The linker of H-NS was shown to be essential for the interdomain interaction between the N-terminal and C-terminal domain(71). Studies on MvaT in our lab further support a model in which both an increase in ionic strength and the DNA substrate cooperatively destabilize these interdomain interactions, inducing the dimers to release their second DNA binding domain to bind and bridge a second DNA molecule in trans (figure 2.4a) (57).

The interdomain interactions described above for H-NS and MvaT might be driven by the asymmetrical charge distribution within the protein sequence: the N-terminal domain is mainly negatively charged, while the linker and the DNA-binding motif are positively charged (figure 2.4b, c). Analysis of the average charge of the primary sequences of H-NS-like proteins revealed that this characteristic is a conserved feature among H-NS/MvaT proteins across species and extends to Lsr2 (figure 2.4b, c, d and supplementary figure)(57). The conserved asymmetrical charge distribution might provide an explanation for how H-NS, MvaT and Lsr2 act as sensors of environmental changes. Indeed, several salt bridges in H-NS were shown to have an important function in environmental sensing (72). For Rok, this asymmetrical charge distribution between its folded domains is less pronounced (figure 2.4e). In addition to that, Rok contains a neutral Q linker instead of the basic linker integrated in H-NS, MvaT and Lsr2 polypeptide chains. Previously, the Q linker was defined as a widespread structural element connecting distinct functional domains in bacterial regulatory proteins(73). Thus, the difference in charge distribution and the linker region between Rok and the other proteins could have functional implications (see below).

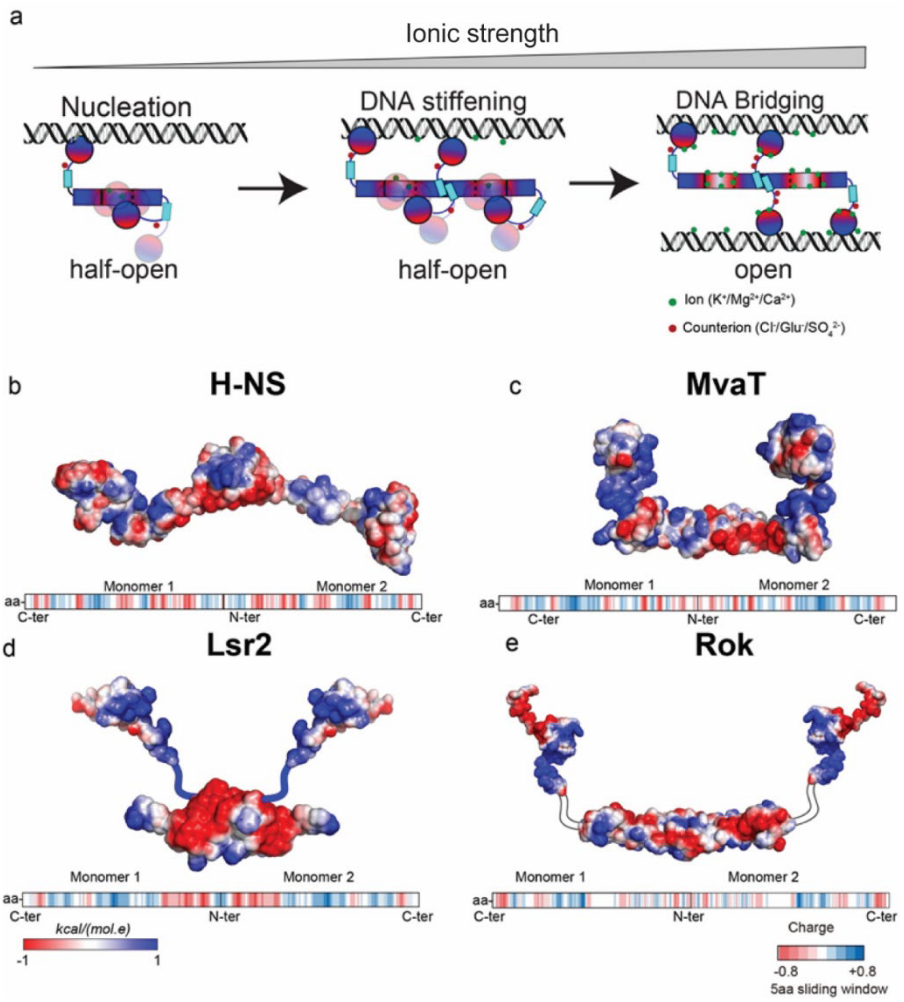


Figure 2.4 H-NS-like proteins DNA binding modes and Electrostatics: (a) Schematic representation of the switching mechanism between H-NS-like proteins DNA binding modes. The red/blue gradient represents the electrostatics of H-NS-like protein surfaces(57, 68). The red and blue are for negatively and positively charged surface regions, respectively. The electrostatic potential surface of (b) H-NS, (c) MvaT, (d) Lsr2 and (e) Rok are depicted on full-length protomer structural models of the proteins using a red/white/blue color gradient. The missing linkers in the structural models of Lrs2 and Rok are shown in blue (positively charged) and white (neutral), respectively. In the lower panel, the five amino acid sliding window averaged charge of the protein's protomers primary sequences generated by EMBOSS charge is shown. Positive, negative and neutral charged amino acid fragments are shown in blue, red and white bars, respectively.

Functional properties

Genome organization

The chromosomal DNA of *E.coli* is structured into domains of various sizes(3). The first layer of organization involves division into four macrodomains (Ori, Ter, Right and Left) of about 1 Mb in size(74–77). Although it is not completely clear how the borders of these domains are formed, several *E. coli* DNA-binding proteins (e.g. SeqA, SlmA and MatP) are associated with certain macrodomains only (77–80). One scale smaller, microdomains have been described of roughly 10 kb in size, which are attributed to loop formation in *E.coli* (81–84). H-NS is important for domain formation *in vivo* and the distribution of H-NS along the chromosome is suggestive of a role in establishing microdomains(82, 85). Recently, it was shown with Hi-C that H-NS mediates short-range contacts along the chromosome(86). The DNA bridging ability of H-NS matches well with the structural properties detected *in vivo*. Genome-wide 3C-based studies reveal chromosomal interaction domains (CIDs) along the chromosome in *Caulobacter crescentus, B. subtilis* and *Vibrio cholerae*(42, 87–89). These domains are tens to hundreds of kb in size and the boundaries of CIDs are often formed by highly transcribed genes(87).

The other DNA bridging proteins likely organize the bacterial genome in similar ways. The genome of *B. subtilis* consists of three global domains and smaller local domains(42). A subset of the barriers between local domains corresponds to genomic positions bound by Rok(40, 42), and the observed Rok clusters also interact with each other forming large chromosomal loops (89). This suggests a role for Rok in the genome organization of *B. subtilis.* For *Pseudomonas* and *Mycobacterium* species, such studies have not been done yet. But ChIP-on-chip data shows that MvaT and Lsr2 bind to defined regions throughout the whole genome(36, 37), supporting a similar role in genome organization as H-NS and Rok.

Gene regulation by H-NS-like proteins

There has been a lot of discussion in the field as to which DNA binding mode is relevant *in vivo* and which of the two modes is needed for gene silencing. The short answer is that both modes of binding could explain gene repression. H-NS-DNA filament formation at or across a promotor region potentially occludes RNA polymerase (RNAP), preventing the initiation of gene transcription. H-NS mutants incapable of gene silencing were indeed found *in vitro* to be defective for nucleofilament formation(66). Note, however, that oligomerization is essential to both filament formation and bridging and that also transcription can be affected by failed assembly of both types of complexes if

oligomerization is perturbed. Data from many recent *in vitro* studies indeed favors models in which DNA bridging plays a key role. H-NS might inactivate promotors that are sensitive to local DNA topology due to its ability to constrain supercoils(90). Generally, these promotors are affected at a distance and H-NS-mediated bridging could also constrain supercoils by generating a diffusion barrier. Bridge formation in promotor regions not only has the potential to occlude RNAP, but also to physically trap RNAP, thereby preventing promoter escape and silencing the associated gene(91–93). Evidence from genome-wide binding profiles and a protein-protein interactions study indicate that H-NS and RNAP colocalize at some promotors, which is compatible with a trapping mechanism(94, 95). On the other hand, H-NS also binds extensively across coding regions to prevent spurious intragenic transcription (96–98). *In vitro* transcription studies show that whilst bridged H-NS-DNA complexes can inhibit progression of the elongating RNAP, the H-NS-DNA filament cannot, highlighting the importance of H-NS' DNA bridging activity(99, 100). Further mechanistic dissection of the repressive role of H-NS in transcription awaits single molecule *in vitro* transcription experiments and the application of novel approaches permitting structural investigation of the process *in vivo.* Much less mechanistic information is available for the other architectural DNA bridging proteins. MvaT and MvaU prevent spurious intragenic transcription similar to H-NS (101). Lsr2 can inhibit transcription *in vitro* and inhibit topoisomerase I, thereby introducing supercoils into the DNA(102). Rok reduces the binding of RNAP at the promotor of *comK*(59). Rok is antagonized at this promotor by ComK itself and although their binding sites partially overlap(103–105), this occurs without preventing Rok binding(59). In contrast to H-NS, Rok binds nearly exclusively around promotor regions which could be explained by differences in specificity of RNAP and its σ factors between *E. coli* and *B. subtilis* (98). It was postulated that anti-repression is achieved through modulation of DNA topology, which would imply that Rok itself also impacts DNA topology(59). In *P. aeruginosa,* MvaT is known to repress the *cupA* gene, which is important in biofilm formation. Mutants unable to silence *cupA* could not form a nucleofilament(106). For these three proteins, it remains to be investigated if DNA bridging is as important for gene silencing as in the case of H-NS.

Functional regulation

Environmental conditions

Bacteria need to respond fast to environmental changes to be able to adapt to diverse living conditions. The bacterial genome operates as an information processing machine, translating environmental cues into altered transcription of specific genes

required for adaptation and survival(4). Key to this process is the dynamic organization of the bacterial genome driven by such cues(2).

The role of temperature

Temperature change is one of the many environmental changes encountered by bacteria that need altered gene expression for adaptation. H-NS regulates the expression of many genes responsive to changes in environment. For example, in *E. coli,* H-NS controls more than 60% of the temperature-regulated genes: they have higher expression at 37°C than at 23°C(107–109). These genes are involved in the nutrient, carbohydrate, and iron utilization systems, and their changes in expression at 37°C may be of advantage for host colonization(110–112). The involvement of H-NS in temperature regulation is corroborated by *in vitro* transcription studies, which indicate that H-NS can no longer pause elongating RNAP at 37°C (see section 4)(100). The underlying mechanism may be an increased off-rate of H-NS or a reduced propensity to oligomerize at high temperature, as observed *in vitro*(30, 31, 99, 107).

Analogous to the situation for H-NS, Lsr2-DNA filaments are also sensitive to temperature changes *in vitro*: the rigidity of Lsr2–DNA complexes is lower at 37°C than at 23°C(67). This can be due to either a change in the ability of Lsr2 to oligomerize along DNA or a change in Lsr2 DNA-binding affinity at 37°C compared to 23°C. Qualitatively, Lsr2-induced DNA folding seems insensitive to changes in temperature, suggesting but not conclusively proving that DNA bridging by Lsr2 is not affected by temperature(67). However, if the general binding scheme depicted in figure 2.3 applies, the effects of temperature on protein-DNA filament may translate into effects on DNA-protein-DNA bridging activity. Thus, taken together, it might be that Lsr2 is involved in regulating thermosensitive genes in *Mycobacteria*.

Unlike the protein-DNA filaments formed by H-NS and Lsr2, the stiffness of MvaT-DNA filaments is not altered in the range of 23°C to 37°C(106). However, in the same temperature range an increase in DNA compaction was observed with increasing temperature. Such a switch between bridging and non-bridging modes is suggestive of temperature-controlled gene regulation at MvaT-bridged genes in *Pseudomonas*.

DNA bridging by Rok is not sensitive to changes in temperature from 25°C to 37°C, suggesting that genes in *B. subtilis* that are repressed by Rok are not regulated by temperature ((69), Chapter 3) .

The role of salt

H-NS has also been shown to regulate genes sensitive to salt stress. The *proU* (*proVWX*) operon is one of the best-studied operons that is osmoregulated by H-NS. The

expression of *proU* is significantly upregulated by high osmolarity(113, 114). *In vitro*, the stiffness of the nucleoprotein filament formed by H-NS and Lsr2 is sensitive to changes in salt concentration from 50 mM to 300 mM KCl(65, 70): the rigidity of the protein-DNA complexes decreased as salt concentration increased. However, the stiffness of the MvaT-DNA filament is not affected by salt over the same concentration range(106). The formation of DNA-DNA bridges by H-NS and MvaT is sensitive to both $MgCl_2$ and KCl(57, 65, 115). Changes in $MgCl_2$ (0-10 mM) or KCl concentration (50-300 mM) drive a switch between the DNA stiffening mode and bridging mode. This structural switch could be the mechanism underlying the regulation of osmoregulated genes by H-NS and MvaT. Qualitatively, Lsr2-induced DNA folding seems insensitive to changes in salt, suggesting that the formation of the Lsr2-DNA bridged complex is not affected by salt(67). However, the salt effects on the structure of the Lsr2-DNA filament could alter the activity of DNA-Lsr2-DNA bridging. DNA bridging by Rok is independent of and not sensitive to both $MgCl_2$ (0-60 mM) and KCl (35-300 mM) concentration ((69), Chapter 3). This insensitivity to changes in salt concentration may be related to the different charge distribution of Rok compared to the other H-NS-like proteins (figure 2.4), where the lack of charges may lead to less interdomain interactions. In this way, Rok could always be in an open conformation, suitable for DNA bridging.

The role of pH

H-NS has been reported to be involved in pH-dependent gene regulation(116). *In vitro*, the rigidity of H-NS-DNA nucleofilaments is sensitive to changes in pH(65). A reduction in stiffness of the protein-DNA complex was observed with increasing pH from 6.5 to 8. However, the formation of the bridged DNA-H-NS-DNA complex might be insensitive to pH changes as H-NS-induced DNA folding was unaffected over the same range of pH values(65).

Unlike the pH sensitivity observed for H-NS-DNA filaments, Lsr2-DNA nucleofilaments were insensitive to pH changes from 6.8 to 8.8. Similar to H-NS, DNA-folding by Lsr2 is not sensitive to pH changes in the same range(67).

Also, the MvaT-DNA nucleoprotein filament is not sensitive to pH changes from 6.5 to 8.5, but DNA compaction is affected(106): MvaT induced stronger folding at pH 6.5 than pH 8.5. Note that the DNA folding induced by H-NS, MvaT and Lsr2 was detected by a qualitative 'folding assay', in which observed DNA folding does not necessarily result from DNA bridging. A quantitative assay such as the 'bridging assay', developed by van der Valk et al.(117) , is essential to determine better the sensitivity of DNA-protein-DNA bridging to changes in environmental conditions. With this bridging assay, DNA bridging by Rok was shown to be insensitive to pH changes from 6.0 to 10.0. Strikingly, even

crossing the pI of Rok (9.31) did not affect its bridging capacities, indicating that charge interactions are unlikely to play a role in DNA bridging by Rok ((69), Chapter 3).

For H-NS, Lsr2 and MvaT, the sensitivity of protein-DNA nucleofilaments or bridged complexes to environmental changes *in vitro* agrees with the proposed role in the regulation of genes sensitive to changes in the environment. However, the involvement in the physiological response to changes remains to be established for Lsr2 and MvaT.

Binding-partners / antagonists

Not only physico-chemical conditions regulate the DNA binding and gene regulation properties of the four proteins. To date, such binding partners (paralogs, truncated derivatives, and non-related modulators and inhibitors) have been primarily identified for H-NS; some modulators were also identified for MvaT, Lsr2 and Rok. They are summarized in table 2.1, together with the characteristics of the four proteins.

Paralogs

The H-NS paralog StpA shares 58% sequence identity. Expression of StpA partially complements an *hns*$^-$ phenotype of *E.coli*(118, 119). H-NS and StpA show negative autoregulation and repress transcription of each other's genes(120, 121). StpA is upregulated during growth at elevated temperature and high osmolarity(122, 123). Like H-NS, StpA can form dimers and higher-order oligomers *in vitro*(124). *In vivo,* it is believed to exist only as a heteromer with H-NS as it is otherwise susceptible to Lon degradation(125, 126). The structural effects of StpA binding to DNA *in vitro* are rather similar to H-NS: StpA can bridge DNA and forms protein filaments along DNA(34, 127). Despite the large similarities between the two proteins, H-NS, StpA and H-NS-StpA heteromers exhibit functionally distinct behavior. At 20°C H-NS mediated DNA-DNA bridges induce transcriptional pausing, whereas they do not at 37°C. (see section 5a)(99). The upregulation of StpA during growth at higher temperature might be explained by H-NS's inability to repress StpA at this temperature. At both 20°C and 37°C, StpA does increase the pausing of RNAP, thereby repressing transcription(100). StpA filaments on DNA are mostly present in a bridged conformation(100). Also, bridged DNA-DNA complexes built using StpA-H-NS heteromers can induce RNAP pausing at 37°C(100). This robustness of StpA could contribute to gene silencing under stress conditions, forming an extra layer of gene regulation by H-NS. A third H-NS paralogue has been reported for several strains: the uropathogenic *E. coli* strain 536 expresses Hfp (also called H-NSB(128)), which is primarily expressed at 25°C and could thus be specifically implicated in regulating gene expression outside the host(123).

Heteromerization of Hfp with H-NS occurs, but, different from StpA, this is not required for the stability of the protein(123). The enteroaggregative *E. coli* strain 042 and several other *Enterobacteriaceae* also carry a second H-NS gene (or third in e.g. strain 536 besides H-NS and Hfp) which can partially complement the *hns*$^-$ phenotype(129). This protein, called H-NS2, is expressed at significantly lower levels than H-NS during exponential growth, but the expression increases in the stationary phase(123). Also, expression of H-NS2 is higher at 37°C than at 25°C, which could relate to the pathogenic nature of the investigated *E. coli* strains(123). A C>T mutation in the *hns2* promoter has been identified which results in the upregulation of H-NS2, and cells readily select for this mutation in the absence of a functional *hns* gene (130). When comparing the amino acid sequences of all mentioned H-NS paralogs, it becomes apparent that the N- and C-terminal domains are quite conserved, particularly the DNA binding domain (figure 2.5a). Because of this, the charge distribution of these proteins is also conserved, which makes it likely that they are similarly regulated by environmental conditions (supplementary figure). The differences in expression levels and thermal stability are most likely responsible for their different regulons.

Several plasmids have also been reported to encode H-NS paralogs. Often these proteins have distinct functional properties. For example, the conjugative IncHI1 plasmid pSfR27 encodes the H-NS paralogue Sfh(131, 132). It was proposed that the plasmid moves from one host to another without causing a large change in host gene expression because the binding of Sfh to pSfR27 prevents H-NS from being titrated away from the chromosome (133). Sfh can form homodimers and heterodimers with H-NS and StpA *in vivo*, but the DNA binding properties of these complexes are still unknown(132). The IncHI plasmid R27 encodes an H-NS variant that can partially complement the *hns*$^-$ phenotype(134). This H-NS$_{R27}$ binds to horizontally acquired DNA, but not to the core genes regulated by chromosomal H-NS(134). The difference is encoded in the linker region, which is hypothesized to be more rigid in H-NS$_{R27}$(135). Both Sfh and H-NS$_{R27}$ are quite conserved in N- and C-terminal domain compared to H-NS, and also the charge distribution is conserved (figure 2.5a and supplementary figure), which likely accounts for their “stealth” function. These plasmid-borne H-NS homologs help the plasmid to stay unnoticed by the host cell upon entry. By silencing horizontally acquired DNA (including the plasmid itself), they prevent H-NS from relieving its chromosomal targets and binding to the plasmid. This allows the plasmid to be transmitted without fitness costs for the host. Several other paralogs of H-NS have been identified via phylogenetic analysis: HlpC (H-NS-like proteins on chromosomes), HlpP (H-NS-like proteins in *Pantoea*) and several clades of Hpp (H-NS plasmid proteins) (136). While the DNA binding motif and site 1 of the N-terminal domain are conserved, diversification was found in site 2, which suggests

different tendencies to heteromerize with H-NS or StpA. How this would affect gene transcription and if the regulons of these paralogues are comparable to H-NS is, however, still unknown.

In *Pseudomonas aeruginosa,* MvaU was identified as a paralogue of MvaT. The two proteins occupy the same chromosomal regions and work coordinately(36). They can form heteromers, but this is not necessary for gene regulation(36, 137). Under conditions of fast growth, they are functionally redundant, becoming only essential when the other partner is deleted from the chromosome(36). Uncharacterized paralogues of MvaT/U are present in several *Pseudomonas* genomes. For example, in *P. alcaligenes* strain RU36E, four MvaT/U proteins are present (figure 2.5b). They share residues in the N-terminal domain and the AT-pincer motif. One of these is closely related to MvaT (Mva2: 81.7% sequence identity). The homology for the other three proteins is not that clear as the sequence identities to MvaT and MvaU are comparable (1: 54.7% and 52.1% 3: 54.0% and 53.7% 4: 53.2% and 54.1%). Another paralogue was found on the IncP-7 plasmid pCAR1 in *P. putida*(138). The protein Pmr (plasmid-encoded MvaT-like regulator) shares 58% sequence identity with MvaT and *in vitro* forms homodimers and heteromers with MvaT-like proteins(139). This is not surprising as many residues and the charge distribution are conserved between Pmr, MvaT and MvaU (figure 2.5b and supplementary figure). Pmr is able to regulate genes both on the plasmid and on the host chromosome(140). Whereas the expression levels of the chromosomally encoded MvaT-like proteins alter between the log and stationary phase, the level of Pmr is constant during different growth phases(141). The regions on the chromosomes bound by Pmr and MvaT are identical. Still, their regulons differ(140) and the regulons of the individual proteins also change upon acquisition of the plasmid and Pmr expression (142). This may be attributed to the formation of different heteromers between Pmr and the two MvaT-like proteins with slightly different functions. Therefore, it remains unsure if Pmr can complement a *mvaT*$^{-}$ phenotype in *P. putida*.

For Lsr2, several candidate paralogs have been identified so far. Several genomes of *Mycobacterium* species harbor two genes encoding for Lsr2 (e.g. *Mycobacterium smegmatis*) or carry a plasmid with an Lsr2 gene (e.g. *Mycobacterium gilvum*)(43). In the case of *M. smegmatis* MKD8, some features of Lsr2, like its DNA binding motif, are conserved between the two proteins, but the second Lsr2 has a much longer linker domain (figure 2.5c). It enlarges the positive patch already present in the Lsr2 linker, so it may provide the protein with extra charged surface for interdomain interactions (supplementary figure). This second Lsr2 also binds DNA and colocalizes with the entire nucleoid (143, 144). Because of this widespread binding and its role in helping *M. smegmatis* to cope with various stresses, it was suggested to be a StpA-like

NAP (144). The fact that its expression is downregulated by Lsr2 itself further supports this hypothesis. It was noticed that all *Streptomycetes* carry a second Lsr2(145), although a recent study shows that Lsr2 is more important than the second, Lsr2-like protein(146). They appear to be similar in charge distribution (supplementary figure). The DNA binding modes of these candidate paralogs or the formation of heteromers with Lsr2 have not been investigated.

Truncated derivatives

Some pathogenic *E. coli* strains, like uropathogenic *E. coli* strain CFT073, encode a truncated version of H-NS: H-NST(128). It lacks the DNA binding domain and the oligomerization site (figure 2.5a). H-NST is able to counteract gene silencing by H-NS, which was proposed to occur by interfering with its oligomerization(128). This model is supported by truncated H-NS products, mimicking H-NST, that are also able to perturb DNA bridging(68).

V. cholerae expresses a virulence regulator with a weak similarity to the N-terminal domain of H-NS: type VI secretion system regulator A (TsrA) (147) (21% identity and 43% identity for TsrA 38-76 to *E. coli* H-NS). The regulons of H-NS and TsrA partially overlap and TsrA has a preference for AT-rich DNA like H-NS (148). As it is not clear if TsrA can bind DNA on its own because of the lack of the DNA binding domain, an interaction between H-NS' N-terminal domain and TsrA seems likely. However, TsrA acts as a repressor of gene expression like H-NS; therefore, it does not seem only a regulator of H-NS' function like H-NST.

A small Rok variant (sRok) was identified on the *B. subtilis* plasmid pLS20(149). sRok can complement the *rok* phenotype in the competence pathway and binds genome-wide across the host chromosome. The DNA binding motif in the C-terminal domain and parts of the N-terminal domain are conserved, suggesting that the function of these domains is also conserved (figure 2.5d). The difference between the two proteins is mostly the length of the linker. The neutral Q-linker of Rok is absent in sRok, which could lead to differences in DNA binding and responsiveness to changes in environmental conditions (figure 2.5d and supplementary figure). In contrast to *B. subtilis,* which carries the *srok* gene nearly always on a plasmid, the *srok* gene is present on the chromosome of *B. licheniformis* and *B. paralicheniformis*. Possibly, the gene was transferred to the chromosome from the pLS20 plasmid. The *rok gene* is not present in all *Bacillus* species, and its introduction has been attributed to a horizontal gene transfer event (105). The absence of *rok* in several *Bacillus* species means that its proposed genome-organizing function is redundant and can be compensated for by other proteins. Recently, it was shown that Rok and sRok can form heteromers and that this changes the susceptibility

to osmolarity ((69), Chapter 3). Although the global pattern of affected genes by either Rok or sRok alone or together is different, further research is needed to understand their function in gene silencing and genome organization.

Non-related modulators and inhibitors

Hha and YdgT are members of the Hha/YmoA family acting as modulators of H-NS activity and function(150). Unlike H-NS truncated derivatives, these proteins have very limited sequence identity with H-NS. Hha is involved in co-regulating a subset of known H-NS-regulated genes, especially in silencing horizontally acquired genes(134, 151, 152). It is, therefore, not obvious *a priori* how these proteins would modulate the DNA binding properties of H-NS. Hha was found to interact with the N-terminal domain of H-NS, specifically with the first two helices(153, 154). The H-NS-Hha co-crystal structure shows two Hha monomers binding to either site of the H-NS dimer, exposing two positively charged Hha surfaces per H-NS dimer(155). Based on this structure, it was proposed that Hha affects H-NS-mediated DNA bridging and thus co-regulates specific genes with H-NS(155). Indeed, it was shown that both Hha and YdgT enhance DNA bridging by H-NS(68). Mechanistically, this could be explained in two ways: 1) Hha provides additional electrostatic interactions with DNA(155) or 2) the H-NS dimer 'opens' upon Hha binding, resulting in a conformation capable of DNA bridging(68). In the latter scenario, Hha could additionally stabilize the complex by the interactions implied in the first scenario. Hha has also been shown to enhance the pausing of RNAP by H-NS and H-NS:Hha complexes preferentially bridge DNA(100). In this manner, Hha could help silence a subset of H-NS-regulated genes. Genomes of a wide-range of pathogenic *E. coli* strains also contain two extra *hha* genes: *hha2* and *hha3*(156) in addition to the H-NS paralogues as described above. The presence of the extra *hha* genes is correlated to the duplication of gene clusters regulated by H-NS and Hha and may be important for virulence(157). Also, the *hha2* gene is present in one of the duplicated regions, meaning that it originates from a duplication event. This shows a relation between duplicated virulence regions and extra genes encoding for regulators like H-NS and Hha. A gene encoding for Hha is also present on several plasmids, including the R27 plasmid described above(158, 159).

Several phage-encoded proteins can inhibit H-NS to counteract gene silencing. Gp5.5 from phage T7 interacts, for example, with the oligomerization domain of H-NS(160, 161), thereby inhibiting oligomerization and gene silencing by H-NS. Another strategy used for counteracting gene silencing by H-NS is mimicking DNA, thereby competing with H-NS' genomic targets.(162–164). This strategy is used by Ocr from phage T7 and Arn from phage T4(162–164). Both strategies result in relief of gene

silencing by H-NS and making H-NS unable to bind to the phage. Lack of repression could lead to replication of the phage and entering of the lytic cycle to kill the host cell. Also MvaT can be inhibited by a protein encoded by a phage. The phage LUZ24 in *P. aeruginosa* expresses a protein called gp4, which was shown to perturb the DNA binding of MvaT and proposed to inhibit the silencing of virus genes by MvaT(165). It was found that gp4 specifically interferes with the formation and stability of a bridged MvaT-DNA complex, again highlighting the importance of the bridging mode in gene regulation (166).

Lsr2 of *M. tuberculosis* binds to the architectural protein HU (167). This interaction involves the N-terminal domain of Lsr2 and the C-terminal tail of HU, which has (P)AKKA repeat motifs and thereby resembles histone tails. This tail is absent in HU of other bacteria discussed in this review (*E. coli, B. subtilis* and *Pseudomonas* species). The Lsr2-HU complex binds DNA, creating thick linear filaments instead of DNA bridges as seen for Lsr2 alone or DNA compaction as seen for HU(167).

Rok interacts with bacterial replication initiator and transcription factor DnaA; jointly, the two proteins interact with a subset of Rok-bound genes(168). DnaA enhances gene repression by Rok for this subset of Rok-bound genes. DnaA binding to these genes is dependent on Rok and the DNA binding domain of DnaA is neither necessary for this interaction nor does it bind DNA in the Rok:DnaA complex. This is consistent with a model in which DnaA modulates the function of Rok. Because DnaA is an ATPase and its activity levels change during the cell cycle(169, 170), it may be an indirect way to modulate the activity of Rok according to the cell cycle or energy status of the cell.

Table 2.1: Characteristics of bacterial DNA-bridging proteins

	H-NS	MvaT	Lsr2	Rok
Bacteria	*Enterobacteriaceae* Gram-negative	*Pseudomonas sp.* Gram-negative	*Actinomycetes* Gram-positive	*Bacillus sp.* Gram-positive
Size	15.5 kDa	14.2 kDa	12.0 kDa	21.8 kDa
Protomer size	Dimer	Dimer	Dimer	Unknown
Oligomerization	Yes	Yes	Yes	Yes
Nucleofilament	Yes	Yes	Yes	N.D.
DNA-bridging	Yes	Yes	Yes	Yes
DNA-bending	No	Yes	No	Yes
Heteromers	Yes	Yes	Predicted	Yes
Modulators				
Paralogues	StpA, Hfp, H-NS2, $H\text{-}NS_{R27}$, Sfh	MvaU, Pmr	Predicted	Unidentified
Truncated derivatives	H-NST, TsrA	Unidentified	Unidentified	sRok
Non-related interaction partners	Hha, YdgT, gp5.5, Ocr, Arn	gp4	HU	DnaA

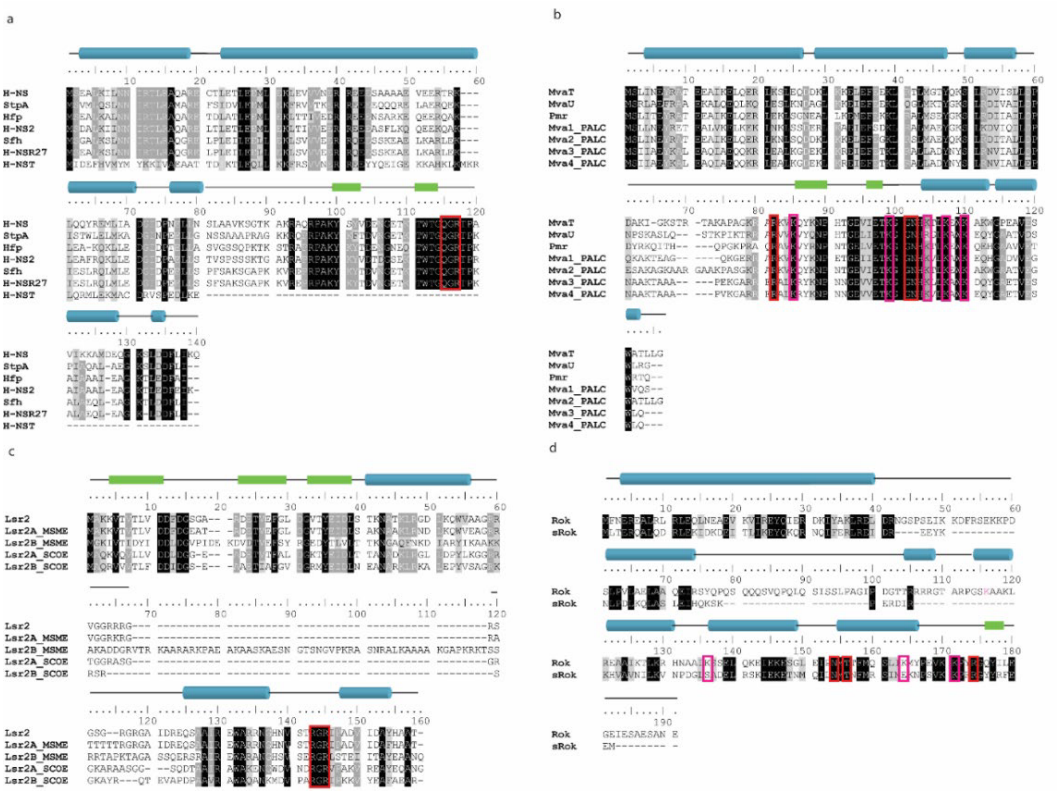


Figure 2.5 Comparison of H-NS-like architectural proteins and their structurally related protein-protein interaction partners. Alignment of identified paralogues and truncated derivatives of H-NS (a), MvaT (b), Lsr2 (c) and Rok (d). Identical residues in all sequences are highlighted in black, identical residues in >75% of the sequences are highlighted in dark grey and conserved residues are highlighted in light grey. The DNA binding motif is highlighted in red and the lysine network of MvaT and Rok in magenta. α-Helices are indicated with blue cylinders and β-sheets with green boxes. The indicated structure corresponds to the most upper sequence. H-NS, *E. coli* strain K-12, NP_415753.1; StpA, *E. coli* strain K-12, NP_417155.1; Hfp, *E. coli* strain 536, ABG69928.1; H-NS2, *E. coli* strain 042, CBG35667.1; Sfh, *Shigella flexneri* 2a, AAN38840.1; H-NSR27, *Salmonella enterica subsp. enterica serovar Typhi* strain CT18, NP_569380.1; H-NST, *Escherichia coli* CFT073, NP_754305; MvaT, *P. aeruginosa* PAO1, NP_253005.1; MvaU, *P. aeruginosa* CLJ1, PTC37345.1; Pmr, *P. resinovorans*, NP_758612.1; Mva1/2/3/4_PALC, *P. alcaligenes* RU36E SIQ98833.1, SIQ72658.1, SIP93681.1 and SIP94365.1; Lsr2_MTUB, *M. tuberculosis* H37Rv, NP_218114.1; Lsr2A/B_MSMEG, *M. smegmatis* MKD8, AWT56911.1 and AWT52048.1; Lsr2A/B_SCOEL, *S. coelicolor* A(3)2 , CAB40875.1 and CAB56356.1; Rok, *B. subtilis* strain 168, NP_389307.1; sRok, *B. subtilis subsp. natto*, YP_004243533.

Post-translational modifications

The post-translational modification (PTMs) of eukaryotic histones has been studied for years. The functional significance of these modifications in biological processes, including DNA repair, gene regulation, and cell division, is well established(171). Although PTMs have been identified in recent years on NAPs in bacteria, their functional importance remains unclear.

H-NS has been found to undergo many PTMs, which potentially add an extra layer to its function in chromatin organization and regulation (172). PTMs discovered for H-NS include acetylation, succinylation, and 2-hydroxyisobutyrylation (hib) of lysines, methylation of arginines, phosphorylation of serines, tyrosines and threonines, deamidation of asparagines and oxidation of methionines(172). These modifications may influence diverse functional properties of H-NS, such as DNA binding, oligomerization, and interaction with other proteins. Acetylation or succinylation occurs on Lys96 and Lys121 near the H-NS DNA-binding motif(173). These modifications could reduce the DNA binding affinity due to the change from positive to neutral or negative charge(174, 175). Acetylation at Lys83 and Lys87 located in the linker region has been identified(174, 175). This could decrease the H-NS DNA binding affinity since the positive charge residues of the linker region are important for DNA binding(61). The addition of the hib modification on Lys121 neutralizes a positive charge and was shown to reduce the DNA binding activity of H-NS (176). Also, H-NS K121hib was shown to improve the acid resistance of *E. coli* as H-NS regulates several genes related to acid tolerance. The identified phosphorylation on Tyr61 induces a negative charge and could interfere with H-NS oligomerization(177). Indeed, the phosphorylation-mimicking mutation Y61D was shown to be important in H-NS dimer-dimer interaction. A similar reasoning was used for the phosphorylation of Thr13, where the resulting neutralization of Arg14 would lead to a defect in the dimerization of H-NS (178). Phosphorylation at Ser42 was shown to be temperature-dependent, where H-NS is phosphorylated at higher temperatures, and this is necessary for silencing prophage CP4So (179). Lys6 has been shown to be involved in the interaction between Hha and H-NS(155). Succinylation of Lys6 may reduce the strength of Hha binding by inducing steric hindrance, which could act in modulating the silencing of genes by H-NS at which Hha is involved as a co-partner(175).

Lsr2 of *M. tuberculosis* has been found to be phosphorylated at Thr112, located near the DNA binding motif(180). A recent study shows that phosphorylation of this residue decreases DNA binding by Lsr2, thus resulting in altered expression of genes important for *M. tuberculosis* growth and survival(181). Some residues (Thr8, Thr22 and Thr31) in the N-terminal domain of Lsr2 were also found to be phosphorylated *in vitro*(181). All these residues are located at β-sheets which are important for the

formation of dimers or oligomers(43). The PTMs on these residues could influence the dimerization or oligomerization of Lsr2 by the addition of negative charges. Also the interaction with HU, which involves the N-terminal domain of Lsr2(167), could be regulated by phosphorylation.

In MvaT from *P. aeruginosa PA01,* several residues located at the N-terminal domain and linker region are acetylated and succinylated and at the C-terminal domain succinylated(182). MvaT Lys86 is directly adjacent to Lys85, which is involved in DNA binding(47). Lys86 is succinylated, leading to charge inversion and, thereby, a decrease in DNA binding affinity(182). Acetylation occurs on Lys22 and Lys31, located in the dimerization domain. On Lys39, shown to be important for oligomerization by forming hydrogen bonds(45), both acetylation and succinylation can occur. By altering positive charge to neutral or negative charge, these PTMs could affect the dimerization or oligomerization of MvaT. In addition, phosphorylation occurs at S2 in *P. putida PNL-MK25* MvaT*,* which may also affect its dimerization properties(183). The linker region has been found to be important for H-NS interdomain interaction between C-terminal and N-terminal domain, which plays a role in H-NS controlling genes sensitive to temperature(71). For the MvaT linker region, acetylation of Lys63 and succinylation of Lys72 was discovered. By changing positive charge to neutral or negative charge, these PTMs could interfere with the intradomain interaction and modulate the function of MvaT in gene regulation.

Proteome analyses of *B. amyloliquefaciens* and *B. subtilis* show that Rok can be acetylated on residues K51 and K142(184, 185). K51 is present in an (predicted) unstructured part of the N-terminal domain. Because it is currently unknown which residues are important for dimer- and oligomerization of Rok, we cannot predict the effect of acetylation on K51. K142 is close to the DNA binding motif of Rok and, while it was not identified as being directly involved, could affect the strength of the ‘lysine network’(48). Nevertheless, it could have an effect on the DNA binding affinity of Rok by changing the charge of K142 from positive to neutral.

PTMs that lead to changes in charge could affect the charge distribution of H-NS-like proteins. This may alter the proteins’ response to environmental changes, which is important for their role in gene regulation. Unfortunately, functional studies on PTMs on these proteins are currently lacking, but they will prove essential in better understanding the function of PTMs in global gene regulation and physiological adaptation.

New H-NS-like proteins?

In recent years, additional proteins have been suggested to be H-NS-like proteins. One is MucR from the MucR/Ros family of proteins in α-proteobacteria such as *Brucella spp.* MucR has a functional role in regulation of genes important for host infection and virulence and autoregulates expression by binding at the *mucR* promotor (186–188). Interestingly, no H-NS homologs were previously identified in α-proteobacteria (23). The DNA binding preferences of MucR resemble H-NS as MucR favors AT-rich DNA with a TpA steps and binds in the minor groove (189). The C-terminal part of MucR consists of a zinc-finger domain that is responsible for DNA binding (190, 191) unlike any other H-NS-like protein (figure 2.2) (46–48). However, not all MucR proteins need zinc to form a functional DNA binding domain (192). MucR can form higher-order oligomers via a conserved hydrophobic region in its N-terminal part (189). Specifically, Leu36, Leu39 and Ile40 were shown to be responsible for the formation of MucR oligomers (193). In contrast to the undefined size of H-NS oligomers made up of an integer number of dimers, MucR forms defined decamers, and mutating the three hydrophobic residues results in MucR monomers. This suggests that MucR does not have a separate dimer- and oligomerization site like most H-NS-like proteins (figure 2.2) (43–45). The decamer is most likely the functional unit *in vivo* as MucR L36A L39A I40A cannot complement for MucR wildtype in *B. abortus* (193). The N-terminal domain was also shown to be necessary for MucR to form DNA bridges (194). *In vivo* complementation experiments with chimeric proteins where the N- or C-terminal domain of MucR was replaced with that of H-NS or Lsr2 show that all the N-terminal domains are functionally similar . The C-terminal domain of H-NS was able to replace MucR's, while this was not possible with the DNA binding domain of Lsr2. This is unexpected as H-NS and Lsr2 use the same AT-hook motif Q/RGR to bind DNA (46). Taken together, the MucR/Ros family of proteins might be a class of H-NS-like proteins based on structural and functional similarities. Open questions that remain to be answered are if MucR can form a nucleoprotein filament next to DNA bridges and whether MucR is regulated *in vivo* by, for example, physico-chemical conditions, interaction partners, and PTMs.

Another NAP found in α-proteobacteria is GapR from *Caulobacter crescentus* (195). The absence of GapR affects global gene expression and likely chromosome organization (196). However, GapR binding does not seem to correlate with changes in gene expression, suggesting mainly an indirect effect (197). GapR cannot complement

for H-NS in *E. coli*, suggesting different DNA binding preferences or DNA-protein complexes *in vivo* (195).

GapR has a binding preference for AT-rich DNA but also associates specifically with overtwisted DNA in accordance with its function in chromosome segregation during DNA replication (197). Recently, this property of GapR was used to map positive supercoiling along chromosomes of different organisms (198). Structurally, a GapR monomer is composed of one larger α-helix followed by two shorter ones. The first helix is responsible for dimerization, and the second and third form the tetrameric interface (197, 199). This GapR tetramer was found to be the protomer size of the protein in solution (200). It encircles the overtwisted DNA as a dimer-of-dimers like a clamp. The inside of the clamp contains six lysine residues responsible for DNA binding (199, 201). Using AFM, GapR was found to stimulate the formation of junctions in DNA, suggestive of DNA bridging similarly to H-NS-like proteins (201). It is, however, still unclear if this is mediated by individual tetramers or higher-order structures possibly mediated by a salt bridge (197, 201). Interestingly, the N-terminal region of GapR shares 61% sequence similarity with site 1 of H-NS, suggesting a similar function for both coiled-coil regions. Also the second α-helix of GapR shares 67% similarity with the C-terminal DNA binding domain of H-NS. Complementation experiments with chimeric proteins show that those domains indeed exhibit a similar function (201). This is surprising, considering the completely different tertiary and quaternary structures of GapR and H-NS. The sequence similarity would imply a shared evolutionary history between the two proteins.

As described above, especially pathogenic bacteria seem to express paralogues, truncated derivatives, and other modulators of H-NS-like proteins, mainly to regulate the expression of (plasmid-based) virulence genes. In *Edwardsiella* and *Salmonella* a new NAP was discovered recently: EnrR (202). It is encoded in a horizontally transferred genomic island and EnrR is essential for virulence in pathogenic bacteria. It targets AT-rich regions using a winged helix-turn-helix domain and binds to both the minor and major groove of DNA (202). EnrR competes with H-NS at horizontally acquired genes, replacing H-NS at their promotors to activate virulence without direct interaction between the proteins. Incubation of EnrR with plasmid DNA resulted in DNA compaction and possibly DNA bridging (202). The crystal structure of EnrR, however, shows a monomer with one DNA binding interface, contradicting the possibility of binding and bridging two DNA duplexes. Therefore, more research is necessary to confirm whether EnrR is a DNA bridging protein and if it can form dimers and higher order oligomers.

Conclusion and perspectives

The architectural chromatin proteins H-NS, MvaT, Lsr2, and the newly proposed functional homolog Rok play important roles in the organization and regulation of the bacterial genome. Although their sequence similarity is low, their domain organization is the same: the N-terminal domain functions in dimerization and oligomerization, the C-terminal domain binds DNA, and a flexible linker connects the two. Moreover, except for Rok, the charge distribution of these proteins along the sequence is highly conserved.

All four proteins are capable of bridging DNA duplexes, which is important for spatial genome organization and gene regulation (figure 2.6). Although a vast amount of research is done on gene regulation by H-NS, it remains to be investigated if the other three proteins regulate genes in similar ways. To mechanistically decipher how H-NS represses genes, more advanced single-molecule and *in vivo* experiments are needed.

H-NS, MvaT and Lsr2 are functionally modulated by changes in environmental conditions, protein partners, and PTMs (figure 2.6). PTMs are, to date, poorly explored and could be an entirely new field of research. The first functional study for phosphorylation of Lsr2 shows that PTMs can indeed be important for the DNA binding properties of these chromatin-organizing proteins and could be a new, uncharacterized way of regulation.

We propose that the shared domain organization and asymmetric charge distribution of the H-NS-like proteins is key to their response to changes in environmental conditions. This information could be used to predict a protein's behavior and may be employed in fighting pathogenic strains by either activating or repressing specific gene transcription.

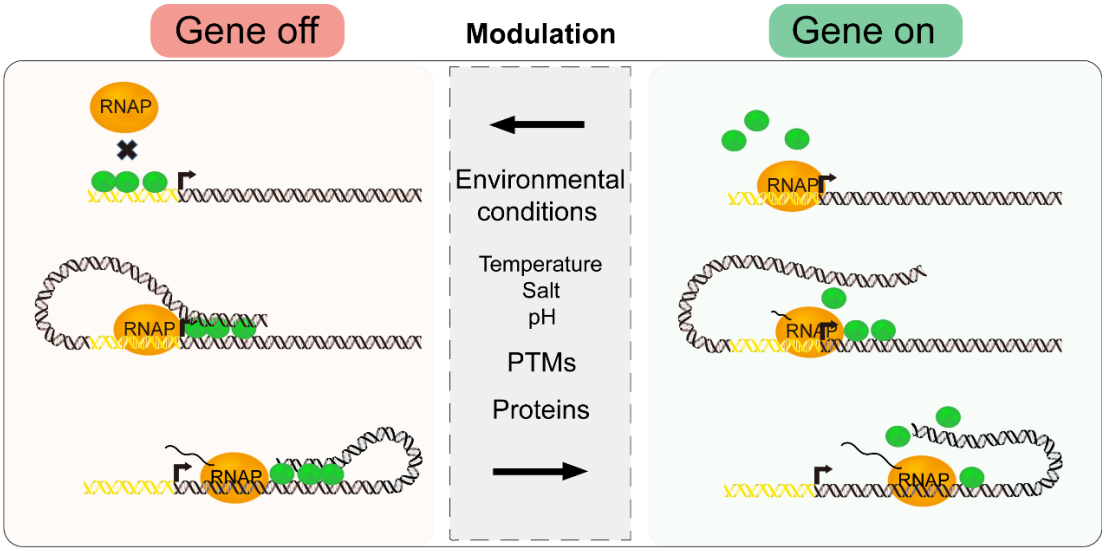


Figure 2.6: H-NS-like proteins functional regulation in gene silencing by different factors. H-NS-like proteins may inhibit the activity of RNA polymerase (RNAP) in three ways: 1) RNAP binding to the promotor region can be inhibited by protein-DNA filaments or bridged complexes at the promoter; 2) RNAP can be trapped in the open initation complex by DNA-protein-DNA bridge complexes. 3) Actively transcribing RNAP can be paused or blocked by DNA-protein-DNA complexes. The inhibition of RNAP by H-NS-like proteins may be modulated by factors such as environmental conditions (temperature, salt and pH), proteins and post-translational modifications (PTMs), allowing genes to be expressed.

References

1. Luijsterburg,M.S., White,M.F., Van Driel,R. and Th. Dame,R. (2008) The major architects of chromatin: Architectural proteins in bacteria, archaea and eukaryotes. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 393–418.
2. Felsenfeld,G., Boyes,J., Chung,J., Clark,D. and Studitsky,V. (1996) Chromatin structure and gene expression. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 9384.
3. Dame,R.T., Rashid,F.-Z.M. and Grainger,D.C. (2020) Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.*, **21**, 227–242.
4. Mekalanos,J.J. (1992) Environmental signals controlling expression of virulence determinants in bacteria. *J. Bacteriol.*, **174**, 1–7.
5. Dorigo,B., Schalch,T., Bystricky,K. and Richmond,T.J. (2003) Chromatin fiber folding: Requirement for the histone H4 N-terminal tail. *J. Mol. Biol.*, **327**, 85–96.
6. Dorigo,B., Schalch,T., Kulangara,A., Duda,S., Schroeder,R.R. and Richmond,T.J. (2004) Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science (80-.).*, **306**, 1571–1573.
7. Luger,K. and Hansen,J.C. (2005) Nucleosome and chromatin fiber dynamics. *Curr. Opin. Struct. Biol.*, **15**, 188–196.
8. Clark,D.J. and Thomas,J.O. (1988) Differences in the binding of H1 variants to DNA: Cooperativity and linker-length related distribution. *Eur. J. Biochem.*, **178**, 225–233.
9. Zheng,R., Ghirlando,R., Lee,M.S., Mizuuchi,K., Krause,M. and Craigie,R. (2000) Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 8997–9002.
10. Hirano,M., Anderson,D.E., Erickson,H.P. and Hirano,T. (2001) Bimodal activation of SMC ATPase by intra- and inter-molecular interactions. *EMBO J.*, **20**, 3238–3250.
11. Hirano,T. (2016) Condensin-Based Chromosome Organization from Bacteria to Vertebrates. *Cell*, **164**, 847–857.
12. Cobbe,N. and Heck,M.M.S. (2004) The Evolution of SMC Proteins: Phylogenetic Analysis and Structural Implications. *Mol. Biol. Evol.*, **21**, 332–347.
13. Thomas,J.O. and Travers,A.A. (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem. Sci.*, **26**, 167–174.
14. Henneman,B. and T. Dame,R. (2015) Archaeal histones: dynamic and versatile genome architects. *AIMS Microbiol.*, **1**, 72–81.
15. Mattiroli,F., Bhattacharyya,S., Dyer,P.N., White,A.E., Sandman,K., Burkhart,B.W., Byrne,K.R., Lee,T., Ahn,N.G., Santangelo,T.J., *et al.* (2017) Structure of histone-based chromatin in Archaea. *Science (80-.).*, **357**, 609–612.
16. Henneman,B., van Emmerik,C., van Ingen,H. and Dame,R.T. (2018) Structure and function of archaeal histones. *PLoS Genet.*, **14**, e1007582.
17. Henneman,B., Brouwer,T.B., Erkelens,A.M., Kuijntjes,G.-J., van Emmerik,C., van der

Valk,R.A., Timmer,M., Kirolos,N.C.S., van Ingen,H., van Noort,J., *et al.* (2021) Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res.*, **49**, 4338–4349.

18. Jelinska,C., Conroy,M.J., Craven,C.J., Hounslow,A.M., Bullough,P.A., Waltho,J.P., Taylor,G.L. and White,M.F. (2005) Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA Packaging. *Structure*, **13**, 963–971.
19. Laurens,N., Driessen,R.P.C., Heller,I., Vorselen,D., Noom,M.C., Hol,F.J.H., White,M.F., Dame,R.T. and Wuite,G.J.L. (2012) Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat. Commun.*, **3**, 2330.
20. Driessen,R.P.C. and Dame,R.T. (2013) Structure and dynamics of the crenarchaeal nucleoid. *Biochem. Soc. Trans.*, **41**, 321–5.
21. Azam,T.A. and Ishihama,A. (1999) Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.*, **274**, 33105–13.
22. Dillon,S.C. and Dorman,C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, **8**, 185–195.
23. Tendeng,C. and Bertin,P.N. (2003) H-NS in Gram-negative bacteria: a family of multifaceted proteins. *Trends Microbiol.*, **11**, 511–518.
24. Pan,C.Q., Finkel,S.E., Cramton,S.E., Feng,J.-A., Sigman,D.S. and Johnson,R.C. (1996) Variable Structures of Fis-DNA Complexes Determined by Flanking DNA – Protein Contacts. *J. Mol. Biol.*, **264**, 675–695.
25. Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–306.
26. Swinger,K.K., Lemberg,K.M., Zhang,Y. and Rice,P.A. (2003) Flexible DNA bending in HU-DNA cocrystal structures. *EMBO J.*, **22**, 3749–60.
27. Hommais,F., Krin,E., Laurent-Winter,C., Soutourina,O., Malpertuy,A., Le Caer,J.P., Danchin,A. and Bertin,P. (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol. Microbiol.*, **40**, 20–36.
28. Navarre,W.W., Porwollik,S., Wang,Y., McClelland,M., Rosen,H., Libby,S.J. and Fang,F.C. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science (80-.).*, **313**, 236–238.
29. Dame,R.T., Wyman,C. and Goosen,N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.*, **28**, 3504–3510.
30. Dame,R.T., Noom,M.C. and Wuite,G.J.L. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature*, **444**, 387–390.
31. Amit,R., Oppenheim,A.B. and Stavans,J. (2003) Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophys. J.*, **84**, 2467–

2473.

32. Tendeng,C., Soutourina,O.A., Danchin,A. and Bertin,P.N. (2003) MvaT proteins in Pseudomonas spp.: A novel class of H-NS-like proteins. *Microbiology*, **149**, 3047–3050.
33. Gordon,B.R.G., Imperial,R., Wang,L., Navarre,W.W. and Liu,J. (2008) Lsr2 of Mycobacterium represents a novel class of H-NS-like proteins. *J. Bacteriol.*, **190**, 7052–7059.
34. Dame,R.T., Luijsterburg,M.S., Krin,E., Bertin,P.N., Wagner,R. and Wuite,G.J.L. (2005) DNA Bridging: a Property Shared among H-NS-Like Proteins. *J. Bacteriol.*, **187**, 1845–1848.
35. Chen,J.M., Ren,H., Shaw,J.E., Wang,Y.J., Li,M., Leung,A.S., Tran,V., Berbenetz,N.M., Kocíncová,D., Yip,C.M., *et al.* (2008) Lsr2 of Mycobacterium tuberculosis is a DNA-bridging protein. *Nucleic Acids Res.*, **36**, 2123–2135.
36. Castang,S., McManus,H.R., Turner,K.H. and Dove,S.L. (2008) H-NS family members function coordinately in an opportunistic pathogen. *Proc. Natl. Acad. Sci.*, **105**, 18947–18952.
37. Gordon,B.R.G., Li,Y., Wang,L., Sintsova,A., van Bakel,H., Tian,S., Navarre,W.W., Xia,B. and Liu,J. (2010) Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci.*, **107**, 5154–5159.
38. Castang,S. and Dove,S.L. (2010) High-order oligomerization is required for the function of the H-NS family member MvaT in Pseudomonas aeruginosa. *Mol. Microbiol.*, **78**, 916–931.
39. Colangeli,R., Helb,D., Vilchèze,C., Hazbón,M.H., Lee,C.-G., Safi,H., Sayers,B., Sardone,I., Jones,M.B., Fleischmann,R.D., *et al.* (2007) Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in M. tuberculosis. *PLoS Pathog.*, **3**, e87.
40. Smits,W.K. and Grossman,A.D. (2010) The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in Bacillus subtilis. *PLoS Genet.*, **6**, 1001207.
41. Serrano,E., Torres,R. and Alonso,J.C. (2021) Nucleoid-associated Rok differentially affects chromosomal transformation on Bacillus subtilis recombination-deficient cells. *Environ. Microbiol.*, **23**, 3318–3331.
42. Marbouty,M., Le Gall,A., Cattoni,D.I., Cournac,A., Koh,A., Fiche,J.-B., Mozziconacci,J., Murray,H., Koszul,R. and Nollmann,M. (2015) Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Mol. Cell*, **59**, 588–602.
43. Summers,E.L., Meindl,K., Usó,I., Mitra,A.K., Radjainia,M., Colangeli,R., Alland,D. and Arcus,V.L. (2012) The Structure of the Oligomerization Domain of Lsr2 from

Mycobacterium tuberculosis Reveals a Mechanism for Chromosome Organization and Protection. *PLoS One*, **7**, 38542.

44. Arold,S.T., Leonard,P.G., Parkinson,G.N. and Ladbury,J.E. (2010) H-NS forms a superhelical protein scaffold for DNA condensation. *Proc. Natl. Acad. Sci.*, **107**, 15728–15732.
45. Suzuki-Minakuchi,C., Kawazuma,K., Matsuzawa,J., Vasileva,D., Fujimoto,Z., Terada,T., Okada,K. and Nojiri,H. (2016) Structural similarities and differences in H-NS family proteins revealed by the N-terminal structure of TurB in Pseudomonas putida KT2440. *FEBS Lett.*, **590**, 3583–3594.
46. Gordon,B.R.G., Li,Y., Cote,A., Weirauch,M.T., Ding,P., Hughes,T.R., Navarre,W.W., Xia,B. and Liu,J. (2011) Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc. Natl. Acad. Sci.*, **108**, 10690–10695.
47. Ding,P., McFarland,K.A., Jin,S., Tong,G., Duan,B., Yang,A., Hughes,T.R., Liu,J., Dove,S.L., Navarre,W.W., *et al.* (2015) A Novel AT-Rich DNA Recognition Mechanism for Bacterial Xenogeneic Silencer MvaT. *PLoS Pathog.*, **11**, 1004967.
48. Duan,B., Ding,P., Hughes,T.R., Navarre,W.W., Liu,J. and Xia,B. (2018) How bacterial xenogeneic silencer rok distinguishes foreign from self DNA in its resident genome. *Nucleic Acids Res.*, **46**, 10514–10529.
49. Ueguchi,C., Suzuki,T., Yoshida,T., Tanaka,K.I. and Mizuno,T. (1996) Systematic mutational analysis revealing the functional domain organization of Escherichia coli nucleoid protein H-NS. *J. Mol. Biol.*, **263**, 149–162.
50. Zhao,X., Remington,J.M., Schneebeli,S.T., Arold,S.T. and Li,J. (2021) Molecular Basis for Environment Sensing by a Nucleoid-Structuring Bacterial Protein Filament. *J. Phys. Chem. Lett.*, **12**, 7878–7884.
51. Ribeiro-Guimarães,M.L. and Pessolani,M.C.V. (2007) Comparative genomics of mycobacterial proteases. *Microb. Pathog.*, **43**, 173–178.
52. Drozdetskiy,A., Cole,C., Procter,J. and Barton,G.J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.*, **43**, W389–W394.
53. Lupas,A., Dyke,M. Van and Stock,J. (1991) Predicting coiled coils from protein sequences. *Science (80-.).*, **252**, 1162–1164.
54. Duan,B., Ding,P., Navarre,W.W., Liu,J. and Xia,B. (2021) Xenogeneic Silencing and Bacterial Genome Evolution: Mechanisms for DNA Recognition Imply Multifaceted Roles of Xenogeneic Silencers. *Mol. Biol. Evol.*, **38**, 4135.
55. Wiechert,J., Filipchyk,A., Hünnefeld,M., Gätgens,C., Brehm,J., Heermann,R. and Frunzke,J. (2020) Deciphering the rules underlying xenogeneic silencing and counter-silencing of lsr2-like proteins using cgps of corynebacterium glutamicum as a model. *MBio*, **11**, e02273-19.
56. Ding,P., McFarland,K.A., Jin,S., Tong,G., Duan,B., Yang,A., Hughes,T.R., Liu,J., Dove,S.L., Navarre,W.W., *et al.* (2015) A Novel AT-Rich DNA Recognition Mechanism

for Bacterial Xenogeneic Silencer MvaT. *PLoS Pathog.*, **11**, 1004967.

57. Qin,L., Bdira,F. Ben, Sterckx,Y.G.J., Volkov,A.N., Vreede,J., Giachin,G., van Schaik,P., Ubbink,M. and Dame,R.T. (2020) Structural basis for osmotic regulation of the DNA binding properties of H-NS proteins. *Nucleic Acids Res.*, **48**, 2156–2172.
58. Bouffartigues,E., Buckle,M., Badaut,C., Travers,A. and Rimsky,S. (2007) H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat. Struct. Mol. Biol.*, **14**, 441–448.
59. Smits,W.K., Hoa,T.T., Hamoen,L.W., Kuipers,O.P. and Dubnau,D. (2007) Antirepression as a second mechanism of transcriptional activation by a minor groove binding protein. *Mol. Microbiol.*, **64**, 368–381.
60. Halford,S.E. and Marko,J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
61. Gao,Y., Foo,Y.H., Winardhi,R.S., Tang,Q., Yan,J. and Kenney,L.J. (2017) Charged residues in the H-NS linker drive DNA binding and gene silencing in single cells. *Proc. Natl. Acad. Sci.*, **114**, 12560–12565.
62. Gulvady,R., Gao,Y., Kenney,L.J. and Yan,J. (2018) A single molecule analysis of H-NS uncouples DNA binding affinity from DNA specificity. *Nucleic Acids Res.*, **46**, 10216–10224.
63. Shen,B.A., Hustmyer,C.M., Roston,D., Wolfe,M.B. and Landick,R. (2022) Bacterial H-NS contacts DNA at the same irregularly spaced sites in both bridged and hemi-sequestered linear filaments. *iScience*, **25**, 104429.
64. Winardhi,R.S., Fu,W., Castang,S., Li,Y., Dove,S.L. and Yan,J. (2012) Higher order oligomerization is required for H-NS family member MvaT to form gene-silencing nucleoprotein filament. *Nucleic Acids Res.*, **40**, 8942–8952.
65. Liu,Y., Chen,H., Kenney,L.J. and Yan,J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev.*, **24**, 339–44.
66. Lim,C.J., Lee,S.Y., Kenney,L.J. and Yan,J. (2012) Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. *Sci. Rep.*, **2**.
67. Qu,Y., Lim,C.J., Whang,Y.R., Liu,J. and Yan,J. (2013) Mechanism of DNA organization by Mycobacterium tuberculosis protein Lsr2. *Nucleic Acids Res.*, **41**, 5263–5272.
68. van der Valk,R.A., Vreede,J., Qin,L., Moolenaar,G.F., Hofmann,A., Goosen,N. and Dame,R.T. (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-Bridging activity. *Elife*, **6**, e27369.
69. Erkelens,A.M., Qin,L., van Erp,B., Miguel-Arribas,A., Abia,D., Keek,H.G.J., Markus,D., Cajili,M.K.M., Schwab,S., Meijer,W.J.J., *et al.* (2022) The B. subtilis Rok protein is an atypical H-NS-like protein irresponsive to physico-chemical cues. *Nucleic Acids Res.*, 10.1093/nar/gkac1064.
70. Dame,R.T. (2005) The role of nucleoid-associated proteins in the organization and

compaction of bacterial chromatin. *Mol. Microbiol.*, **56**, 858–870.

71. Shahul Hameed,U.F., Liao,C., Radhakrishnan,A.K., Huser,F., Aljedani,S.S., Zhao,X., Momin,A.A., Melo,F.A., Guo,X., Brooks,C., *et al.* (2018) H-NS uses an autoinhibitory conformational switch for environment-controlled gene silencing. *Nucleic Acids Res.*, **1**, 23955–6900.
72. Zhao,X., Hameed,U.F.S., Kharchenko,V., Liao,C., Huser,F., Remington,J.M., Radhakrishnan,A.K., Jaremko,M., Jaremko,Ł., Arold,S.T., *et al.* (2021) Molecular basis for the adaptive evolution of environment-sensing by H-NS proteins. *Elife*, **10**, 1–18.
73. Wootton,J.C. and Drummond,M.H. (1989) The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng.*, **2**, 535–43.
74. Valens,M., Penaud,S., Rossignol,M., Cornet,F. and Boccard,F. (2004) Macrodomain organization of the Escherichia coli chromosome. *EMBO J.*, **23**, 4330–4341.
75. Valens,M., Thiel,A. and Boccard,F. (2016) The MaoP/maoS Site-Specific System Organizes the Ori Region of the E. coli Chromosome into a Macrodomain. *PLOS Genet.*, **12**, e1006309.
76. Duigou,S. and Boccard,F. (2017) Long range chromosome organization in Escherichia coli: The position of the replication origin defines the non-structured regions and the Right and Left macrodomains. *PLoS Genet.*, **13**, e1006758.
77. Mercier,R., Petit,M.-A., Schbath,S., Robin,S., El Karoui,M., Boccard,F. and Espéli,O. (2008) The MatP/matS site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. *Cell*, **135**, 475–85.
78. Sánchez-Romero,M.A., Busby,S.J.W., Dyer,N.P., Ott,S., Millard,A.D. and Grainger,D.C. (2010) Dynamic Distribution of SeqA Protein across the Chromosome of Escherichia coli K-12. *MBio*, **1**.
79. Tonthat,N.K., Arold,S.T., Pickering,B.F., Van Dyke,M.W., Liang,S., Lu,Y., Beuria,T.K., Margolin,W. and Schumacher,M.A. (2011) Molecular mechanism by which the nucleoid occlusion factor, SlmA, keeps cytokinesis in check. *EMBO J.*, **30**, 154–64.
80. Dame,R.T., Kalmykowa,O.J. and Grainger,D.C. (2011) Chromosomal macrodomains and associated proteins: Implications for DNA organization and replication in gram negative bacteria. *PLoS Genet.*, **7**.
81. Deng,S., Stein,R.A. and Higgins,N.P. (2005) Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.*, **57**, 1511–21.
82. Hardy,C.D. and Cozzarelli,N.R. (2005) A genetic selection for supercoiling mutants of Escherichia coli reveals proteins implicated in chromosome structure. *Mol. Microbiol.*, **57**, 1636–52.
83. Postow,L., Hardy,C.D., Arsuaga,J. and Cozzarelli,N.R. (2004) Topological domain structure of the Escherichia coli chromosome. *Genes Dev.*, **18**, 1766–1779.

84. Worcel,A. and Burgi,E. (1972) On the structure of the folded chromosome of Escherichia coli. *J. Mol. Biol.*, **71**, 127–47.
85. Noom,M.C., Navarre,W.W., Oshima,T., Wuite,G.J.L. and Dame,R.T. (2007) H-NS promotes looped domain formation in the bacterial chromosome. *Curr. Biol.*, **17**, R913-4.
86. Lioy,V.S., Cournac,A., Marbouty,M., Phane Duigou,S., Mozziconacci,J., Espéo., Dé Ric Boccard,F. and Koszul,R. (2018) Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins Escherichia coli Article Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell*, **172**, 771-783.e18.
87. Le,T.B.K., Imakaev,M. V., Mirny,L.A. and Laub,M.T. (2013) High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science (80-.).*, **342**, 731–734.
88. Val,M.-E., Marbouty,M., de Lemos Martins,F., Kennedy,S.P., Kemble,H., Bland,M.J., Possoz,C., Koszul,R., Skovgaard,O. and Mazel,D. (2016) A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. *Sci. Adv.*, **2**, e1501914.
89. Dugar,G., Hofmann,A., Heermann,D.W. and Hamoen,L.W. (2022) A chromosomal loop anchor mediates bacterial genome organization. *Nat. Genet. 2022 542*, **54**, 194–201.
90. Blot,N., Mavathur,R., Geertz,M., Travers,A. and Muskhelishvili,G. (2006) Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep.*, **7**, 710–5.
91. Shin,M., Song,M., Joon,H.R., Hong,Y., Kim,Y.J., Seok,Y.J., Ha,K.S., Jung,S.H. and Choy,H.E. (2005) DNA looping-mediated repression by histone-like protein H-NS: Specific requirement of Eσ70 as a cofactor for looping. *Genes Dev.*, **19**, 2388–2398.
92. Dame,R.T., Wyman,C., Wurm,R., Wagner,R. and Goosen,N. (2001) Structural Basis for H-NS-mediated Trapping of RNA Polymerase in the Open Initiation Complex at the rrnB P1. *J. Biol. Chem.*, **277**, 2146–2150.
93. Schröder,O. and Wagner,R. (2000) The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. *J. Mol. Biol.*, **298**, 737–748.
94. Grainger,D.C., Hurd,D., Goldberg,M.D. and Busby,S.J.W. (2006) Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic Acids Res.*, **34**, 4642–4652.
95. Butland,G., Peregrín-Alvarez,J.M., Li,J., Yang,W., Yang,X., Canadien,V., Starostine,A., Richards,D., Beattie,B., Krogan,N., *et al.* (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature*, **433**, 531–537.
96. Singh,S.S., Singh,N., Bonocora,R.P., Fitzgerald,D.M., Wade,J.T. and Grainger,D.C. (2014) Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev.*, **28**, 214–219.

97. Lamberte,L.E., Baniulyte,G., Singh,S.S., Stringer,A.M., Bonocora,R.P., Stracy,M., Kapanidis,A.N., Wade,J.T. and Grainger,D.C. (2017) Horizontally acquired AT-rich genes in Escherichia coli cause toxicity by sequestering RNA polymerase. *Nat. Microbiol.*, **2**.
98. Forrest,D., Warman,E.A., Erkelens,A.M., Dame,R.T. and Grainger,D.C. (2022) Xenogeneic silencing strategies in bacteria are dictated by RNA polymerase promiscuity. *Nat. Commun.*, **13**, 1149.
99. Kotlajich,M. V, Hron,D.R., Boudreau,B.A., Sun,Z., Lyubchenko,Y.L. and Landick,R. (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife*, **2015**, 4970.
100. Boudreau,B.A., Hron,D.R., Qin,L., van der Valk,R.A., Kotlajich,M. V, Dame,R.T. and Landick,R. (2018) StpA and Hha stimulate pausing by RNA polymerase by promoting DNA–DNA bridging of H-NS filaments. *Nucleic Acids Res.*, **46**, 5525–5546.
101. Lippa,A.M., Gebhardt,M.J. and Dove,S.L. (2021) H-NS-like proteins in Pseudomonas aeruginosa coordinately silence intragenic transcription. *Mol. Microbiol.*, **115**, 1138–1151.
102. Colangeli,R., Helb,D., Vilchèze,C., Hazbón,M.H., Lee,C.-G., Safi,H., Sayers,B., Sardone,I., Jones,M.B., Fleischmann,R.D., *et al.* (2007) Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in M. tuberculosis. *PLoS Pathog.*, **3**, e87.
103. Hamoen,L.W., Van Werkhoven,A.F., Bijlsma,J.J.E., Dubnau,D. and Venema,G. (1998) The competence transcription factor of Bacillus subtilis recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev.*, **12**, 1539–1550.
104. Hamoen,L.W., Van Werkhoven,A.F., Venema,G. and Dubnau,D. (2002) The pleiotropic response regulator DegU functions as a priming protein in competence development in Bacillus subtilis. *Proc. Natl. Acad. Sci.*, **97**, 9246–9251.
105. Albano,M., Smits,W.K., Ho,L.T.Y., Kraigher,B., Mandic-Mulec,I., Kuipers,O.P. and Dubnau,D. (2005) The Rok protein of Bacillus subtilis represses genes for cell surface and extracellular functions. *J. Bacteriol.*, **187**, 2010–9.
106. Winardhi,R.S., Fu,W., Castang,S., Li,Y., Dove,S.L. and Yan,J. (2012) Higher order oligomerization is required for H-NS family member MvaT to form gene-silencing nucleoprotein filament. *Nucleic Acids Res.*, **40**, 8942–8952.
107. Ono,S., Goldberg,M.D., Olsson,T., Esposito,D., Hinton,J.C.D. and Ladbury,J.E. (2005) H-NS is a part of a thermally controlled mechanism for bacterial gene regulation. *Biochem. J.*, **391**, 203–13.
108. Göransson,M., Sondén,B., Nilsson,P., Dagberg,B., Foreman,K., Emanuelsson,K. and Uhlin,B.E. (1990) Transcriptional silencing and thermoregulation of gene expression in Escherichia coli. *Nature*, **344**, 682–685.

109. White-Ziegler,C.A. and Davis,T.R. (2009) Genome-wide identification of H-NS-controlled, temperature-regulated genes in Escherichia coli K-12. *J. Bacteriol.*, **191**, 1106–10.
110. Lease,R.A. and Belfort,M. (2000) A trans-acting RNA as a control switch in Escherichia coli: DsrA modulates function by forming alternative structures. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 9919–24.
111. Jordi,B.J., Dagberg,B., de Haan,L.A., Hamers,A.M., van der Zeijst,B.A., Gaastra,W. and Uhlin,B.E. (1992) The positive regulator CfaD overcomes the repression mediated by histone-like protein H-NS (H1) in the CFA/I fimbrial operon of Escherichia coli. *EMBO J.*, **11**, 2627–32.
112. Han,Y., Zhou,D., Pang,X., Song,Y., Zhang,L., Bao,J., Tong,Z., Wang,J., Guo,Z., Zhai,J., *et al.* (2004) Microarray Analysis of Temperature-Induced Transcriptome of *Yersinia pestis*. *Microbiol. Immunol.*, **48**, 791–805.
113. Ueguchi,C. and Mizuno,T. (1993) The Escherichia coli nucleoid protein H-NS functions directly as a transcriptional repressor. *EMBO J.*, **12**, 1039.
114. Nagarajavel,V., Madhusudan,S., Dole,S., Rahmouni,A.R. and Schnetz,K. (2007) Repression by binding of H-NS within the transcription unit. *J. Biol. Chem.*, **282**, 23622–23630.
115. Winardhi,R.S., Yan,J. and Kenney,L.J. (2015) Biophysical Perspective H-NS Regulates Gene Expression and Compacts the Nucleoid: Insights from Single-Molecule Experiments. *Biophysj*, **109**, 1321–1329.
116. Atlung,T., Knudsen,K., Heerfordt,L. and BrØndsted,L. (1997) Effects of σ(s) and the transcriptional activator AppY on induction of the Escherichia coli hya and cbdAB-appA operons in response to carbon and phosphate starvation. *J. Bacteriol.*, **179**, 2141–2146.
117. van der Valk,R.A., Qin,L., Moolenaar,G.F. and Dame,R.T. (2018) Quantitative determination of DNA bridging efficiency of chromatin proteins. In Dame R.T. (ed), *Methods in Molecular Biology*. Humana Press, New York, NY, Vol. 1837, pp. 199–209.
118. Zhang,A. and Belfort,M. (1992) Nucleotide sequence of a newly-identified Escherichia coli gene, stpA, encoding an H-NS-like protein.
119. Shi,X. and Bennett,G.N. (1994) Plasmids bearing hfq and the hns-like gene stpA complement hns mutants in modulating arginine decarboxylase gene expression in Escherichia coli. *J. Bacteriol.*, **176**, 6769–75.
120. Zhang,A., Rimsky,S., Reaban,M.E., Buc,H. and Belfort,M. (1996) Escherichia coli protein analogs StpA and H-NS: Regulatory loops, similar and disparate effects on nucleic acid dynamics. *EMBO J.*, **15**, 1340–1349.
121. Sonden,B. and Uhlin,B.E. (1996) Coordinated and differential expression of histone-like proteins in Escherichia coli: regulation and function of the H-NS analog StpA. *EMBO J.*, **15**, 4970–80.

122. Free,A. and Dorman,C.J. (1997) The Escherichia coli stpA gene is transiently expressed during growth in rich medium and is induced in minimal medium and by stress conditions. *J. Bacteriol.*, **179**, 909–18.
123. Müller,C.M., Schneider,G., Dobrindt,U., Emödy,L., Hacker,J. and Uhlin,B.E. (2010) Differential effects and interactions of endogenous and horizontally acquired H-NS-like proteins in pathogenic *Escherichia coli*. *Mol. Microbiol.*, **75**, 280–293.
124. Leonard,P.G., Ono,S., Gor,J., Perkins,S.J. and Ladbury,J.E. (2009) Investigation of the self-association and hetero-association interactions of H-NS and StpA from *Enterobacteria*. *Mol. Microbiol.*, **73**, 165–179.
125. Johansson,J. and Uhlin,B.E. (1999) Differential protease-mediated turnover of H-NS and StpA revealed by a mutation altering protein stability and stationary-phase survival of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 10776–81.
126. Johansson,J., Eriksson,S., Sondén,B., Sun Nyunt Wai and Uhlin,B.E. (2001) Heteromeric interactions among nucleoid-associated bacterial proteins: Localization of StpA-stabilizing regions in H-NS of Escherichia coli. *J. Bacteriol.*, **183**, 2343–2347.
127. Lim,C.J., Whang,Y.R., Kenney,L.J. and Yan,J. (2012) Gene silencing H-NS paralogue StpA forms a rigid protein filament along DNA that blocks DNA accessibility. *Nucleic Acids Res.*, **40**, 3316–28.
128. Williamson,H.S. and Free,A. (2005) A truncated H-NS-like protein from enteropathogenic Escherichia coli acts as an H-NS antagonist. *Mol. Microbiol.*, **55**, 808–827.
129. Prieto,A., Bernabeu,M., Aznar,S., Ruiz-Cruz,S., Bravo,A., Queiroz,M.H. and Juárez,A. (2018) Evolution of Bacterial Global Modulators: Role of a Novel H-NS Paralogue in the Enteroaggregative Escherichia coli Strain 042. *mSystems*, **3**.
130. Prieto,A., Bernabeu,M., Falgenhauer,L., Chakraborty,T., Hüttener,M. and Juárez,A. (2020) Overexpression of the third H-NS paralogue H-NS2 compensates fitness loss in hns mutants of the enteroaggregative Escherichia coli strain 042. *Sci. Rep.*, **10**, 18131.
131. Beloin,C., Deighan,P., Doyle,M. and Dorman,C.J. (2003) Shigella flexneri 2a strain 2457T expresses three members of the H-NS-like protein family: characterization of the Sfh protein. *Mol. Genet. Genomics*, **270**, 66–77.
132. Deighan,P., Beloin,C. and Dorman,C.J. (2003) Three-way interactions among the Sfh, StpA and H-NS nucleoid-structuring proteins of Shigella flexneri 2a strain 2457T. *Mol. Microbiol.*, **48**, 1401–1416.
133. Doyle,M., Fookes,M., Ivens,A., Mangan,M.W., Wain,J. and Dorman,C.J. (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science*, **315**, 251–2.
134. Baños,R.C., Vivero,A., Aznar,S., García,J., Pons,M., Madrid,C. and Juárez,A. (2009) Differential Regulation of Horizontally Acquired and Core Genome Genes by the

Bacterial Modulator H-NS. *PLoS Genet.*, **5**, e1000513.

135. Fernández-De-Alba,C., Berrow,N.S., Garcia-Castellanos,R., García,J. and Pons,M. (2013) On the Origin of the Selectivity of Plasmidic H-NS towards Horizontally Acquired DNA: Linking H-NS Oligomerization and Cooperative DNA Binding. *J. Mol. Biol.*, **425**, 2347–2358.
136. Fitzgerald,S., Kary,S.C., Alshabib,E.Y., MacKenzie,K.D., Stoebel,D.M., Chao,T.C. and Cameron,A.D.S. (2020) Redefining the H-NS protein family: A diversity of specialized core and accessory forms exhibit hierarchical transcriptional network integration. *Nucleic Acids Res.*, **48**, 10184–10198.
137. Vallet-Gely,I., Donovan,K.E., Fang,R., Joung,J.K. and Dove,S.L. (2005) Repression of phase-variable cup gene expression by H-NS-like proteins in Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 11082–7.
138. Suzuki,C., Yun,C.-S., Umeda,T., Terabayashi,T., Watanabe,K., Yamane,H. and Nojiri,H. (2011) Oligomerization and DNA-Binding Capacity of Pmr, a Histone-Like Protein H1 (H-NS) Family Protein Encoded on IncP-7 Carbazole-Degradative Plasmid pCAR1. *Biosci. Biotechnol. Biochem.*, **75**, 711–717.
139. Suzuki,C., Kawazuma,K., Horita,S., Terada,T., Tanokura,M., Okada,K., Yamane,H. and Nojiri,H. (2014) Oligomerization Mechanisms of an H-NS Family Protein, Pmr, Encoded on the Plasmid pCAR1 Provide a Molecular Basis for Functions of H-NS Family Members. *PLoS One*, **9**, e105656.
140. Yun,C.-S., Takahashi,Y., Shintani,M., Takeda,T., Suzuki-Minakuchi,C., Okada,K., Yamane,H. and Nojiri,H. (2016) MvaT Family Proteins Encoded on IncP-7 Plasmid pCAR1 and the Host Chromosome Regulate the Host Transcriptome Cooperatively but Differently. *Appl. Environ. Microbiol.*, **82**, 832–42.
141. Sun,Z., Vasileva,D., Suzuki-Minakuchi,C., Okada,K., Luo,F., Igarashi,Y. and Nojiri,H. (2017) Growth phase-dependent expression profiles of three vital H-NS family proteins encoded on the chromosome of Pseudomonas putida KT2440 and on the pCAR1 plasmid. *BMC Microbiol.*, **17**.
142. Nakamura,T., Suzuki-Minakuchi,C., Kawano,H., Kanesaki,Y., Kawasaki,S., Okada,K. and Nojiri,H. (2020) H-NS Family Proteins Drastically Change Their Targets in Response to the Horizontal Transfer of the Catabolic Plasmid pCAR1. *Front. Microbiol.*, **11**.
143. Kriel,N.L., Gallant,J., van Wyk,N., van Helden,P., Sampson,S.L., Warren,R.M. and Williams,M.J. (2018) Mycobacterial nucleoid associated proteins: An added dimension in gene regulation. *Tuberculosis*, **108**, 169–177.
144. Kołodziej,M., Łebkowski,T., Płociński,P., Hołówka,J., Paściak,M., Wojtaś,B., Bury,K., Konieczny,I., Dziadek,J. and Zakrzewska-Czerwińska,J. (2021) Lsr2 and Its Novel Paralogue Mediate the Adjustment of Mycobacterium smegmatis to Unfavorable Environmental Conditions. *mSphere*, **6**, e00290-21.

145. Chandra,G. and Chater,K.F. (2014) Developmental biology of Streptomyces from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.*, **38**, 345–379.
146. Gehrke,E.J., Zhang,X., Pimentel-Elardo,S.M., Johnson,A.R., Rees,C.A., Jones,S.E., Hindra, Gehrke,S.S., Turvey,S., Boursalie,S., *et al.* (2019) Silencing cryptic specialized metabolism in Streptomyces by the nucleoid-associated protein Lsr2. *Elife*, **8**.
147. Zheng,J., Shin,O.S., Cameron,D.E. and Mekalanos,J.J. (2010) Quorum sensing and a global regulator TsrA control expression of type VI secretion and virulence in Vibrio cholerae. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 21128–21133.
148. Caro,F., Caro,J.A., Place,N.M. and Mekalanos,J.J. (2020) Transcriptional Silencing by TsrA in the Evolution of Pathogenic Vibrio cholerae Biotypes. *MBio*, **11**, e02901-20.
149. Singh,P.K., Ramachandran,G., Durán-Alcalde,L., Alonso,C., Wu,L.J. and Meijer,W.J.J. (2012) Inhibition of Bacillus subtilis natural competence by a native, conjugative plasmid-encoded comK repressor protein. *Environ. Microbiol.*, **14**, 2812–2825.
150. Madrid,C., Balsalobre,C., García,J. and Juárez,A. (2007) The novel Hha/YmoA family of nucleoid-associated proteins: use of structural mimicry to modulate the activity of the H-NS family of proteins. *Mol. Microbiol.*, **63**, 7–14.
151. Nieto,J.M., Madrid,C., Prenafeta,A., Miquelay,E., Balsalobre,C., Carrascal,M. and Juárez,A. (2000) Expression of the hemolysin operon in Escherichia coli is modulated by a nucleoid-protein complex that includes the proteins Hha and H-NS. *Mol. Gen. Genet.*, **263**, 349–58.
152. Ueda,T., Takahashi,H., Uyar,E., Ishikawa,S., Ogasawara,N. and Oshima,T. (2013) Functions of the Hha and YdgT Proteins in Transcriptional Silencing by the Nucleoid Proteins, H-NS and StpA, in Escherichia coli. *DNA Res.*, **20**, 263–271.
153. García,J., Cordeiro,T.N., Nieto,J.M., Pons,I., Juárez,A. and Pons,M. (2005) Interaction between the bacterial nucleoid associated proteins Hha and H-NS involves a conformational change of Hha. *Biochem. J.*, **388**, 755–762.
154. García,J., Madrid,C., Juárez,A. and Pons,M. (2006) New Roles for Key Residues in Helices H1 and H2 of the Escherichia coli H-NS N-terminal Domain: H-NS Dimer Stabilization and Hha Binding. *J. Mol. Biol.*, **359**, 679–689.
155. Ali,S.S., Whitney,J.C., Stevenson,J., Robinson,H., Howell,P.L. and Navarre,W.W. (2013) Structural Insights into the Regulation of Foreign Genes in Salmonella by the Hha/H-NS Complex. *J Biol Chem*, **288**, 13356–13369.
156. Prieto,A., Urcola,I., Blanco,J., Dahbi,G., Muniesa,M., Quiros,P., Falgenhauer,L., Chakraborty,T., Huttener,M. and Juarez,A. (2016) Tracking bacterial virulence: Global modulators as indicators. *Sci. Rep.*, **6**, 25973.
157. Bernabeu,M., Sánchez-Herrero,J.F., Huedo,P., Prieto,A., Hüttener,M., Rozas,J. and Juárez,A. (2019) Gene duplications in the E. coli genome: Common themes among

pathotypes. *BMC Genomics*, **20**.

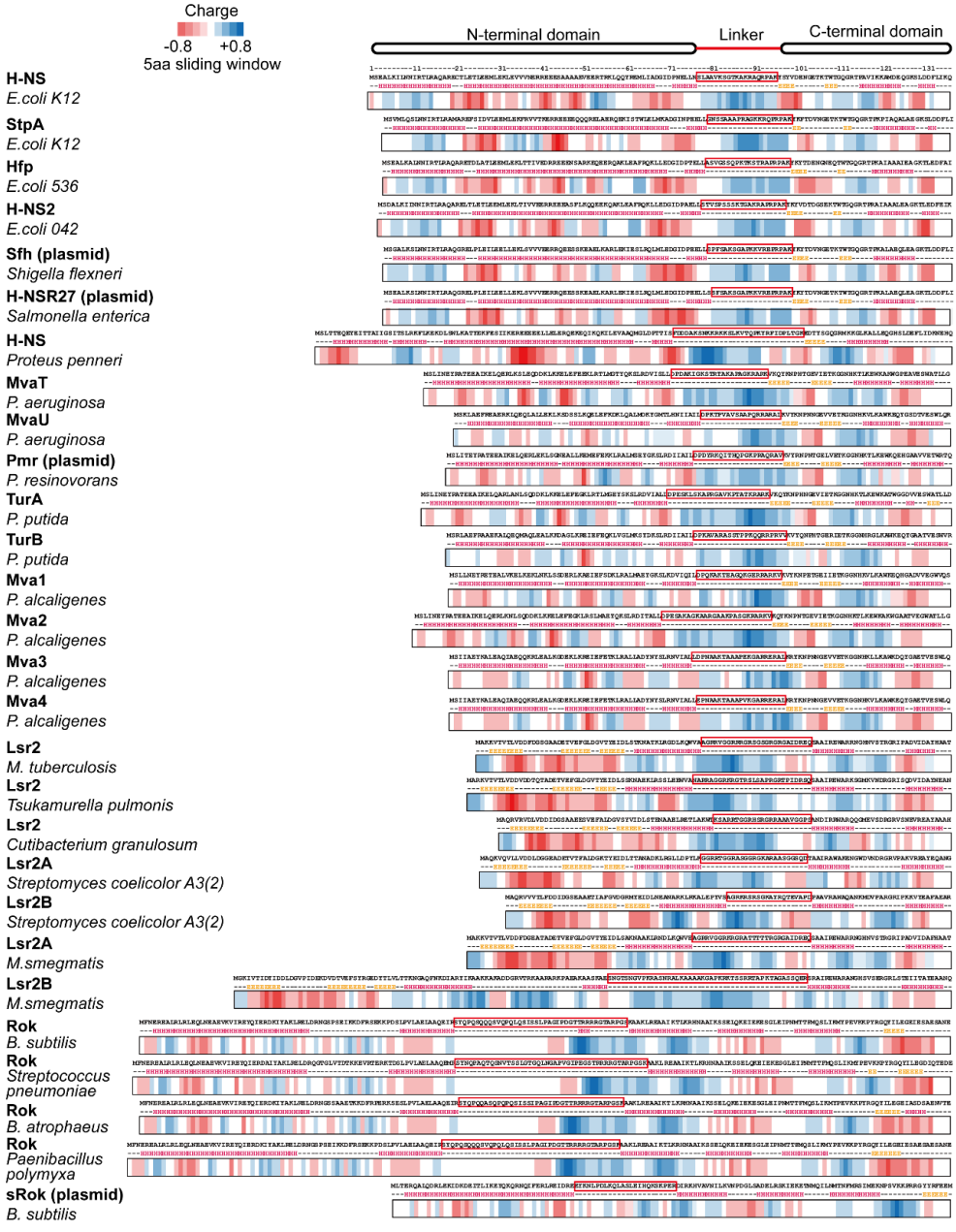
158. Forns,N., Baños,R.C., Balsalobre,C., Juárez,A. and Madrid,C. (2005) Temperature-dependent conjugative transfer of R27: Role of chromosome- and plasmid-encoded Hha and H-NS proteins. *J. Bacteriol.*, **187**, 3950–3959.
159. Shintani,M., Suzuki-Minakuchi,C. and Nojiri,H. (2015) Nucleoid-associated proteins encoded on plasmids: Occurrence and mode of function. *Plasmid*, **80**, 32–44.
160. Liu,Q. and Richardson,C.C. (1993) Gene 5.5 protein of bacteriophage T7 inhibits the nucleoid protein H-NS of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 1761–5.
161. Ali,S.S., Beckett,E., Bae,S.J. and Navarre,W.W. (2011) The 5.5 protein of phage T7 inhibits H-NS through interactions with the central oligomerization domain. *J. Bacteriol.*, **193**, 4881–92.
162. Walkinshaw,M.D., Taylor,P., Sturrock,S.S., Atanasiu,C., Berge,T., Henderson,R.M., Edwardson,J.M. and Dryden,D.T.F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell*, **9**, 187–94.
163. Ho,C.-H., Wang,H.-C., Ko,T.-P., Chang,Y.-C. and Wang,A.H.-J. (2014) The T4 phage DNA mimic protein Arn inhibits the DNA binding activity of the bacterial histone-like protein H-NS. *J. Biol. Chem.*, **289**, 27046–54.
164. Melkina,O.E., Goryanin,I.I. and Zavilgelsky,G.B. (2016) The DNA–mimic antirestriction proteins ArdA ColIB-P9, Arn T4, and Ocr T7 as activators of H-NS-dependent gene transcription. *Microbiol. Res.*, **192**, 283–291.
165. Wagemans,J., Delattre,A.-S., Uytterhoeven,B., De Smet,J., Cenens,W., Aertsen,A., Ceyssens,P.-J. and Lavigne,R. (2015) Antibacterial phage ORFans of Pseudomonas aeruginosa phage LUZ24 reveal a novel MvaT inhibiting protein. *Front. Microbiol.*, **6**, 1242.
166. Bdira,F. Ben, Erkelens,A.M., Qin,L., Volkov,A.N., Lippa,A.M., Bowring,N., Boyle,A.L., Ubbink,M., Dove,S.L. and Dame,R.T. (2021) Novel anti-repression mechanism of H-NS proteins by a phage protein. *Nucleic Acids Res.*, **49**, 10770–10784.
167. Datta,C., Jha,R.K., Ahmed,W., Ganguly,S., Ghosh,S. and Nagaraja,V. (2019) Physical and functional interaction between nucleoid-associated proteins HU and Lsr2 of *Mycobacterium tuberculosis* : altered DNA binding and gene regulation. *Mol. Microbiol.*, 10.1111/mmi.14202.
168. Seid,C.A., Smith,J.L. and Grossman,A.D. (2017) Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in Bacillus subtilis. *Mol. Microbiol.*, **103**, 798–817.
169. Kurokawa,K., Nishida,S., Emoto,A., Sekimizu,K. and Katayama,T. (1999) Replication cycle-coordinated change of the adenine nucleotide-bound forms of DnaA protein in Escherichia coli. *EMBO J.*, **18**, 6642–6652.
170. Burkholder,W.F., Kurtser,I. and Grossman,A.D. (2001) Replication Initiation Proteins Regulate a Developmental Checkpoint in Bacillus subtilis. *Cell*, **104**, 269–279.

171. Bannister,A.J. and Kouzarides,T. (2011) Regulation of chromatin by histone modifications. *Cell Res.*, **21**, 381–395.
172. Dilweg,I.W. and Dame,R.T. (2018) Post-translational modification of nucleoid-associated proteins: an extra layer of functional modulation in bacteria? *Biochem. Soc. Trans.*, **46**, 1381–1392.
173. Schmidt,A., Kochanowski,K., Vedelaar,S., Ahrné,E., Volkmer,B., Callipo,L., Knoops,K., Bauer,M., Aebersold,R. and Heinemann,M. (2016) The quantitative and condition-dependent Escherichia coli proteome. *Nat. Biotechnol.*, **34**, 104–10.
174. Kuhn,M.L., Zemaitaitis,B., Hu,L.I., Sahu,A., Sorensen,D., Minasov,G., Lima,B.P., Scholle,M., Mrksich,M., Anderson,W.F., *et al.* (2014) Structural, Kinetic and Proteomic Characterization of Acetyl Phosphate-Dependent Bacterial Protein Acetylation. *PLoS One*, **9**, e94816.
175. Weinert,B.T., Schölz,C., Wagner,S.A., Iesmantavicius,V., Su,D., Daniel,J.A. and Choudhary,C. (2013) Lysine Succinylation Is a Frequently Occurring Modification in Prokaryotes and Eukaryotes and Extensively Overlaps with Acetylation. *Cell Rep.*, **4**, 842–851.
176. Dong,H., Zhao,Y., Bi,C., Han,Y., Zhang,J., Bai,X., Zhai,G., Zhang,H., Tian,S., Hu,D., *et al.* (2022) TmcA functions as a lysine 2-hydroxyisobutyryltransferase to regulate transcription. *Nat. Chem. Biol.*, **18**, 142–151.
177. Hansen,A.-M., Chaerkady,R., Sharma,J., Díaz-Mejía,J.J., Tyagi,N., Renuse,S., Jacob,H.K.C., Pinto,S.M., Sahasrabuddhe,N.A., Kim,M.-S., *et al.* (2013) The Escherichia coli Phosphotyrosine Proteome Relates to Core Pathways and Virulence. *PLoS Pathog.*, **9**, e1003403.
178. Hu,L., Kong,W., Yang,D., Han,Q., Guo,L. and Shi,Y. (2019) Threonine phosphorylation fine-tunes the regulatory activity of histone-like nucleoid structuring protein in Salmonella transcription. *Front. Microbiol.*, **10**, 1515.
179. Liu,X., Lin,S., Liu,T., Zhou,Y., Wang,W., Yao,J., Guo,Y., Tang,K., Chen,R., Benedik,M.J., *et al.* (2021) Xenogeneic silencing relies on temperature-dependent phosphorylation of the host H-NS protein in Shewanella. *Nucleic Acids Res.*, **49**, 3427–3440.
180. Turapov,O., Forti,F., Kadhim,B., Ghisotti,D., Sassine,J., Straatman-Iwanowska,A., Bottrill,A.R., Moynihan,P.J., Wallis,R., Barthe,P., *et al.* (2018) Two Faces of CwlM, an Essential PknB Substrate, in Mycobacterium tuberculosis. *Cell Rep.*, **25**, 57-67.e5.
181. Alqaseer,K., Turapov,O., Barthe,P., Jagatia,H., De Visch,A., Roumestand,C., Wegrzyn,M., Bartek,I.L., Voskuil,M.I., O'hare,H., *et al.* Protein kinase B controls Mycobacterium tuberculosis growth via phosphorylation of the global transcriptional regulator Lsr2. 10.1101/571406.
182. Gaviard,C., Jouenne,T. and Hardouin,J. (2018) Proteomics of *Pseudomonas aeruginosa* : the increasing role of post-translational modifications. *Expert Rev.*

Proteomics, **15**, 757–772.

183. Ravichandran,A., Sugiyama,N., Tomita,M., Swarup,S. and Ishihama,Y. (2009) Ser/Thr/Tyr phosphoproteome analysis of pathogenic and non-pathogenic Pseudomonas species. *Proteomics*, **9**, 2764–75.
184. Liu,L., Wang,G., Song,L., Lv,B. and Liang,W. (2016) Acetylome analysis reveals the involvement of lysine acetylation in biosynthesis of antibiotics in Bacillus amyloliquefaciens. *Sci. Rep.*, **6**, 20108.
185. Reverdy,A., Chen,Y., Hunter,E., Gozzi,K. and Chai,Y. (2018) Protein lysine acetylation plays a regulatory role in Bacillus subtilis multicellularity. *PLoS One*, **13**, e0204687.
186. Caswell,C.C., Elhassanny,A.E.M., Planchin,E.E., Roux,C.M., Weeks-Gorospe,J.N., Ficht,T.A., Dunman,P.M. and Martin Roop,R. (2013) Diverse genetic regulon of the virulence-associated transcriptional regulator MucR in Brucella abortus 2308. *Infect. Immun.*, **81**, 1040–1051.
187. Mirabella,A., Terwagne,M., Zygmunt,M.S., Cloeckaert,A., De Bolle,X. and Letesson,J.J. (2013) Brucella melitensis MucR, an orthologue of Sinorhizobium meliloti MucR, is involved in resistance to oxidative, detergent, and saline stresses and cell envelope modifications. *J. Bacteriol.*, **195**, 453–465.
188. Jiao,J., Zhang,B., Li,M.L., Zhang,Z. and Tian,C.F. (2022) The zinc-finger bearing xenogeneic silencer MucR in α-proteobacteria balances adaptation and regulatory integrity. *ISME J.*, **16**, 738–749.
189. Baglivo,I., Pirone,L., Pedone,E.M., Pitzer,J.E., Muscariello,L., Marino,M.M., Malgieri,G., Freschi,A., Chambery,A., Roop,R.M., *et al.* (2017) Ml proteins from Mesorhizobium loti and MucR from Brucella abortus: an AT-rich core DNA-target site and oligomerization ability. *Sci. Rep.*, **7**.
190. Russo,L., Palmieri,M., Baglivo,I., Esposito,S., Isernia,C., Malgieri,G., Pedone,P. V. and Fattorusso,R. (2010) NMR assignments of the DNA binding domain of Ml4 protein from Mesorhizobium loti. *Biomol. NMR Assign.*, **4**, 55–57.
191. Palmieri,M., Malgieri,G., Russo,L., Baglivo,I., Esposito,S., Netti,F., Del Gatto,A., De Paola,I., Zaccaro,L., Pedone,P. V., *et al.* (2013) Structural Zn(II) implies a switch from fully cooperative to partly downhill folding in highly homologous proteins. *J. Am. Chem. Soc.*, **135**, 5220–5228.
192. Baglivo,I., Russo,L., Esposito,S., Malgieri,G., Renda,M., Salluzzo,A., Di Blasio,B., Isernia,C., Fattorusso,R. and Pedone,P. V. (2009) The structural role of the zinc ion can be dispensable in prokaryotic zinc-finger domains. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 6933.
193. Pirone,L., Pitzer,J.E., D'Abrosca,G., Fattorusso,R., Malgieri,G., Pedone,E.M., Pedone,P.V., Roop,R.M. and Baglivo,I. (2018) Identifying the region responsible for Brucella abortus MucR higher-order oligomer formation and examining its role in gene regulation. *Sci. Rep.*, **8**, 17238.

194. Shi,W.T., Zhang,B., Li,M.L., Liu,K.H., Jiao,J. and Tian,C.F. (2022) The convergent xenogeneic silencer MucR predisposes α-proteobacteria to integrate AT-rich symbiosis genes. *Nucleic Acids Res.*, **50**, 8580–8598.
195. Ricci,D.P., Melfi,M.D., Lasker,K., Dill,D.L., McAdams,H.H. and Shapiro,L. (2016) Cell cycle progression in Caulobacter requires a nucleoid-associated protein with high AT sequence recognition. *Proc. Natl. Acad. Sci. U. S. A.*, **113**, E5952–E5961.
196. Arias-Cartin,R., Dobihal,G.S., Campos,M., Surovtsev,I. V, Parry,B. and Jacobs-Wagner,C. (2017) Replication fork passage drives asymmetric dynamics of a critical nucleoid-associated protein in Caulobacter. *EMBO J.*, **36**, 301–318.
197. Guo,M.S., Haakonsen,D.L., Zeng,W., Schumacher,M.A. and Laub,M.T. (2018) A Bacterial Chromosome Structuring Protein Binds Overtwisted DNA to Stimulate Type II Topoisomerases and Enable DNA Replication. *Cell*, **175**, 583–597.
198. Guo,M.S., Kawamura,R., Littlehale,M., Marko,J.F. and Laub,M.T. (2021) High-resolution, genome-wide mapping of positive supercoiling in chromosomes. *Elife*, **10**.
199. Tarry,M.J., Harmel,C., Taylor,J.A., Marczynski,G.T. and Schmeing,T.M. (2019) Structures of GapR reveal a central channel which could accommodate B-DNA. *Sci. Rep.*, **9**.
200. Huang,Q., Duan,B., Dong,X., Fan,S. and Xia,B. (2020) GapR binds DNA through dynamic opening of its tetrameric interface. *Nucleic Acids Res.*, **48**, 9372–9386.
201. Lourenço,R.F., Saurabh,S., Herrmann,J., Wakatsuki,S. and Shapiro,L. (2020) The Nucleoid-Associated Protein GapR Uses Conserved Structural Elements To Oligomerize and Bind DNA. *MBio*, **11**.
202. Ma,R., Liu,Y., Gan,J., Qiao,H., Ma,J., Zhang,Y., Bu,Y., Shao,S., Zhang,Y. and Wang,Q. (2022) Xenogeneic nucleoid-associated EnrR thwarts H-NS silencing of bacterial virulence with unique DNA binding. *Nucleic Acids Res.*, **50**, 3777–3798.


Charge
-0.8 +0.8
5aa sliding window
N-terminal domain
Linker
C-terminal domain
H-NS
E.coli K12
StpA
E.coli K12
Hfp
E.coli 536
H-NS2
E.coli 042
Sfh (plasmid)
Shigella flexneri
H-NSR27 (plasmid)
Salmonella enterica
H-NS
Proteus penneri
MvaT
P. aeruginosa
MvaU
P. aeruginosa
Pmr (plasmid)
P. resinovorans
TurA
P. putida
TurB
P. putida
Mva1
P. alcaligenes
Mva2
P. alcaligenes
Mva3
P. alcaligenes
Mva4
P. alcaligenes
Lsr2
M. tuberculosis
Lsr2
Tsukamurella pulmonis
Lsr2
Cutibacterium granulosum
Lsr2A
Streptomyces coelicolor A3(2)
Lsr2B
Streptomyces coelicolor A3(2)
Lsr2A
M.smegmatis
Lsr2B
M.smegmatis
Rok
B. subtilis
Rok
Streptococcus pneumoniae
Rok
B. atrophaeus
Rok
Paenibacillus polymyxa
sRok (plasmid)
B. subtilis

Supplementary figure 2.1: Electrostatics and fold topology of H-NS-like proteins. Comparison of the charge distribution of the four families of H-NS-like protein across a range of species. The average charge of a five-amino acid window as analyzed with EMBOSS charge. Positively and negatively charged regions are colored blue and red respectively. Unknown secondary structures were predicted using JPRED(52). α-Helices are represented with (H) in pink, β-sheets with (E) in yellow and random coils with (-) in black. The linker domain is indicated with a red box. H-NS, *E. coli* strain K-12, NP_415753.1; StpA*, E. coli* strain K-12, NP_417155.1; Hfp, *E. coli* strain 536, ABG69928.1; H-NS2, *E. coli* strain 042, CBG35667.1; Sfh, *Shigella flexneri 2a*, AAN38840.1; H-NSR27, *Salmonella enterica subsp. enterica serovar Typhi* strain CT18, NP_569380.1; H-NS, *Proteus penneri*, SUB98598.1; MvaT, *P. aeruginosa* PAO1, NP_253005.1; MvaU, *P. aeruginosa* CLJ1, PTC37345.1; Pmr, *P. resinovorans*, NP_758612.1; TurA, *P. putida*, SUD72464.1; TurB, *P. putida*, VEE40761.1; Mva1/2/3/4_PALC, *P. alcaligenes* RU36E SIQ98833.1, SIQ72658.1, SIP93681.1 and SIP94365.1; Lsr2_MTUB, *M. tuberculosis* H37Rv, NP_218114.1; Lsr2, *Tsukamurella pulmonis*, SUP14481.1; Lsr2, *Cutibacterium granulosum*, SNV28945.1; Lsr2A/B_SCOEL, *S. coelicolor* A(3)2 , CAB40875.1 and CAB56356.1; Lsr2A/B_MSMEG, *M. smegmatis* MKD8, AWT56911.1 and AWT52048.1; Rok, *B. subtilis* strain 168, NP_389307.1; Rok, *Streptococcus pneumonia*, CVM76913.1; Rok, *B. atrophaeus*, KFK83781.1; Rok, *Paenibacillus polymyxa*, SPY12450.1; sRok, *B. subtilis subsp. natto*, YP_004243533.1.

Chapter 3

The *B. subtilis* Rok protein is an atypical H-NS-like protein irresponsive to physico-chemical cues

This chapter is based on the following article: Erkelens, A.M., Qin, L., van Erp, B., Miguel-Arribas, A., Abia, D., Keek, H.G.J., Markus, D., Cajili, M.K.M., Schwab, S., Meijer, W.J.J. and Dame, R.T. (2022) The *B. subtilis* Rok protein is an atypical H-NS-like protein irresponsive to physico-chemical cues. *Nucleic Acids Research,* 50, 12166-12185

Abstract

Nucleoid-associated proteins (NAPs) play a central role in chromosome organization and environment-responsive transcription regulation. The *Bacillus subtilis*-encoded NAP Rok binds preferentially AT-rich regions of the genome, which often contain genes of foreign origin that are silenced by Rok binding. Additionally, Rok plays a role in chromosome architecture by binding in genomic clusters and promoting chromosomal loop formation. Based on this, Rok was proposed to be a functional homolog of *E. coli* H-NS. However, it is largely unclear how Rok binds DNA, how it represses transcription and whether Rok mediates environment-responsive gene regulation. Here, we investigated Rok's DNA binding properties and the effects of physico-chemical conditions thereon. We demonstrate that Rok is a DNA bridging protein similar to prototypical H-NS-like proteins. However, unlike these proteins, the DNA bridging ability of Rok is not affected by changes in physico-chemical conditions. The DNA binding properties of the Rok interaction partner sRok are affected by salt concentration. This suggests that in a minority of *Bacillus* strains Rok activity can be modulated by sRok, and thus respond indirectly to environmental stimuli. Despite several functional similarities, the absence of a direct response to physico-chemical changes establishes Rok as disparate member of the H-NS family.

Introduction

The bacterial chromosome, like the chromosomes of eukaryotic cells, is highly organized and compactly folded. At the same time, housekeeping genes need to be accessible for the transcription machinery while other (sets of) genes must become available for transcription upon changing environmental conditions. This requires a flexible and dynamic organization of the bacterial chromosome in response to environmental cues (1–3). Many factors contribute to the organization and compaction of the nucleoid, including DNA supercoiling, macromolecular crowding and nucleoid-associated proteins (NAPs) (4–6). The histone-like nucleoid structuring protein (H-NS), one of the main NAPs in *Escherichia coli*, plays important roles in both chromosome organization and gene regulation (7, 8). H-NS can non-specifically bind DNA across the genome, but has a preference for AT-rich DNA. DNA acquired via horizontal gene transfer (HGT) is often AT-rich and is therefore recognized as xenogeneic DNA by H-NS (9, 10). Although genes acquired via HGT are key to the evolution of bacteria by conferring new genetic traits, inappropriate transcription of acquired genes can lead to loss of competitive fitness. H-NS family proteins, which include H-NS of *E. coli*, MvaT of *Pseudomonas sp.* and Lsr2 of *Mycobacteria*, function as silencers of xenogeneic genes, until repression is relieved by particular environmental signals (11–13).

Rok of *Bacillus subtilis,* which was originally identified as the repressor of competence regulator *comK* (14), plays a role in the repression of genes for cell surface and extracellular functions (15). Rok has been proposed to be a functional homolog of H-NS, based on the observation that it binds preferentially to AT-rich regions of the chromosome acquired via horizontal gene transfer (16). Rok contributes to silencing of the genes within such regions (16), including those encoding antimicrobial compounds (17). Rok also decreases chromosomal transformation (14), possibly by altering nucleoid architecture (18). This behavior classifies Rok as a xenogeneic silencer like H-NS, MvaT and Lsr2. Rok primarily silences expression by interfering with the initial steps of transcription (19). In contrast, H-NS also binds intragenic regions to repress spurious transcription. Such repression by Rok is less needed in *B. subtilis* due to enhanced specificity of its RNA polymerase (19). Lastly, Rok was found to be associated with a large subset of chromosomal domain boundaries in *B. subtilis* (20), suggesting that it contributes to chromosome organization similarly as was proposed previously for H-NS (21, 22). Recently, it was demonstrated that Rok indeed establishes long-range chromosomal loops in *B. subtilis* (23). This is suggestive of Rok acting as a DNA bridging protein. However, the molecular basis underlying the precise roles of Rok in transcriptional regulation and chromosome organization and their possible interplay

remains unknown. In addition, it is unknown whether gene silencing by Rok can be modulated by changes in physico-chemical growth conditions such as temperature, pH and salt. A distinct group of *rok* genes has been identified on several *Bacillus* plasmids belonging to the pLS20 family and on some *Bacillus* chromosomes (24, 25). Because these *rok* genes are smaller, we refer to them as small Rok (sRok) (8). The most outstanding difference between the two types of Rok is that sRok lacks a part of Rok's linker domain. sRok can associate with the host chromosome and it can replace Rok in the competence pathway (24). Therefore, we hypothesized that sRok has similar DNA binding features as Rok.

In this study, we used a combined *in vitro* (single-molecule) and *in vivo* approach to study the DNA binding properties of Rok, sRok and the artificial Rok Δ75-96 variant. We found that Rok and sRok are both DNA bridging proteins, but that they respond differently to changes in environmental conditions. We show that the DNA binding properties of Rok are only very mildly modulated by physico-chemical changes, while the DNA bridging activity of sRok is osmo-sensitive. We also demonstrate that Rok and sRok can form heterodimers and show that these interactions alter DNA bridging properties, which in turn, most probably, affect the regulatory role of Rok and sRok *in vivo*. Therefore, interactions with sRok and possibly other interactions with proteins might be key to the regulation of Rok-mediated gene repression.

Materials and methods

Cloning and mutagenesis

Oligonucleotides used for cloning and mutagenesis procedures are listed in table S3.1. The Rok coding sequence from *B. subtilis* 168 (Ref seq: NC_000964.31493787-1494362, Uniprot O34857, NCBI protein database: NP_389307.1) was cloned into pET30b using Gibson Assembly (26) resulting in plasmid pRD231. This plasmid was used as template to create the expression vector for Rok Δ75-96 (pRD415) using Gibson Assembly, where amino acids 75-96 were deleted from the sequence. pRD231 was also used to create plasmid pRD461 containing Rok with a C-terminal His-tag (Rok 6xHis) using Gibson assembly. The coding sequence for sRok was created and codon optimized with GeneArt (Thermo Fisher Scientific) (Uniprot: E9RJ31 NCBI protein database: BAJ76946.1) and cloned into pET30b using Gibson assembly creating plasmid pRD411. The plasmids for *in vivo* complementation use the pUC19 vector as backbone. All inserts into pUC19 were cloned into the vector using Gibson Assembly. pRD408 was created by taking the *E. coli* MG1655 *hns* promoter including the upstream regulatory region up to position -150 (genomic location 1292358-1292508)

followed by the H-NS coding sequence (genomic location 129509-129923). The Rok, sRok and Rok Δ75-96 plasmids (pRD424, pRD410 and pRD412) were created by replacing the H-NS coding sequence of pRD408 with the Rok, sRok or Rok Δ75-96 sequence from pRD231, pRD411 or pRD415 respectively. In a previous publication, pRD424 was named pRok (19).The sequence of all constructs was verified by DNA Sanger sequencing (BaseClear). All plasmids were deposited at Addgene and their information and identification numbers are summarized in table S3.2.

DNA substrates

Tethered particle motion and bridging assay experiments were performed using an AT-rich (32% GC) 685 bp DNA substrate described earlier (27, 28) unless otherwise stated. The DNA substrate was generated by PCR using Thermo Scientific® Phusion® High-Fidelity DNA Polymerase and the products were purified using the GenElute PCR Clean-up kit (Sigma-Aldrich). As single-stranded DNA substrates for the bridging assay, parts with comparable GC-content (around 32%) of the 685 bp DNA substrate were ordered as oligonucleotides (Sigma-Aldrich) and turned into double-stranded DNA using PCR and complementary oligonucleotides (table S3.3). The poly(A) single-stranded DNA with an average length between 250-500 bp was ordered from Sigma-Aldrich and turned into double-stranded DNA using PCR and poly(T) oligonucleotides. For use in the DNA bridging assay (see below), DNA was ^{32}P-labeled (29). For microscale thermophoresis, complementary oligonucleotides of 78 bp were designed (table S3.4) and the top strand was 5' labelled with Cy5. The oligonucleotides were mixed 1:1 to a final concentration of 40 μM, heated to 95°C and slowly cooled down to room temperature for annealing. For Atomic Force Microscopy, pUC19 plasmid was incubated with nicking endonuclease Nb.BsrDI (New England Biolabs) for one hour at 65°C followed by heat inactivation at 80°C for 20 minutes and cooling to room temperature. The nicked plasmid was then purified by phenol chloroform extraction and the buffer was exchanged with HPLC water (Sigma-Aldrich) through overnight dialysis at RT using a Slide-A-Lyzer cassette with a 3.5 kDa cut-off (Thermo Scientific).

Protein purification

E. coli BL21(DE3) pLysS cells, transformed with pRD231, pRD411 or pRD415 were grown at 37°C, 250 rpm until an OD600 (optical density at 600 nm) of 0.6. Expression was induced with 1 mM IPTG and cell growth was continued at 16°C, 180 rpm overnight. Cells were pelleted at 6354 g, 4°C and resuspended in 20 mM Tris HCl pH 8.0, 130 mM NaCl, 10% glycerol with 100 μM PMSF and 20.5 μg/mL DNase I. The cells were lysed using a French press and the lysate was centrifuged with an

ultracentrifuge (Beckman Coulter) for 30 min at 100736 g. The supernatant was filtered with a 0.22 µm Millex-GP Syringe Filter and loaded on a HiTrap Heparin HP 1 mL affinity column (GE Healthcare). The Rok protein was eluted using an NaCl gradient from 130 mM to 1.5 M. The eluted fractions were checked for the presence of Rok with SDS-PAGE and the relevant fractions were pooled, concentrated with an Amicon 10 kDa cut-off filter and buffer exchanged with a PD10 column (GE Healthcare) to a buffer with 130 mM NaCl. Next, the protein was loaded on and eluted from a HiTrap SP HP 1 mL column (GE Healthcare) using a NaCl gradient from 130 mM to 1.5 M. The fractions were again checked with SDS-PAGE and concentrated to 500 µl with an Amicon 10 kDa cut-off filter. The protein was then loaded on a GE Superdex 200 10/300 Increase GL column pre-equilibrated with storage buffer (20 mM Tris HCl pH 8.0, 300 mM KCl, 10% glycerol). Rok Δ75-96 and sRok were purified according to the same general protocol with minor modifications. For both proteins, the pH of the buffers (except the storage buffer) was changed to 7.5 to increase affinity for the columns. For sRok, 3 µM benzamidine was added to the lysis buffer to prevent cleavage and a P11 column using a gradient from 100 mM to 1.5 M NH_4Cl was used first. The eluate containing sRok was dialysed overnight against 20 mM Tris HCl pH 8.0, 130 mM NaCl, 10% glycerol and the purification continued with a Heparin column as described. Cells that expressed Rok with a C-terminal his-tag (Rok-6xHis, pRD461) were lysed in 20 mM Tris HCl pH 8, 300 mM NaCl, 10 mM imidazole, 3.5% glycerol. Rok-6xHis was purified using a 5 mL HisTrap HP column (GE Healthcare) with a gradient from 10 mM to 1 M imidazole, followed by gel filtration as described above. Protein concentrations were determined using a Pierce™ BCA Protein Assay Kit (Thermo Scientific) or a Qubit™ Protein assay kit (Invitrogen). Purity (>95%) and identity of the proteins was verified with SDS-PAGE and mass spectrometry.

Tethered particle motion

The DNA used for Tethered Particle Motion (TPM) experiments was an AT-rich (32% GC) 685 bp DNA substrate (30, 31). Measurements were performed as previously described (30, 32) with minor modifications. Briefly, the flow cell was washed with 100 µL experimental buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA, 5% glycerol, 50 mM KCl) to remove excess beads and 100 µL protein diluted in experimental buffer was flowed in and incubated for 5 minutes. Next, the flow cell was washed with protein solution one more time, sealed with nail polish and incubated for 5 minutes. After incubation, the flow cell was directly transferred to the holder and incubated for 5 more minutes in the instrument to stabilize the temperature at 25 °C for the measurement. For each flow cell,

more than 200 beads were measured and measurements for each concentration were performed at least in duplicate.

Data analysis including the calculation of occupancy was done as described previously (30). The occupancy (θ) is the fraction of DNA bound by Rok (equation 1) and can be calculated from the area under the peaks of the fitted Gaussian distributions.

$$\theta = \frac{A_{peak,bound}}{A_{peak,bound}+A_{peak,unbound}} \quad \text{(Equation 1)}$$

In which, θ is the occupancy and A is the area under the peaks of the fitted Gaussian distribution. Fitting the occupancies as a function of the concentration of Rok using the Hill-binding model (Equation 2) yields the apparent binding affinity K_D and Hill coefficient (n).

$$\theta(c) = \frac{1}{\left(\frac{K_D}{[c]}\right)^n + 1} \quad \text{(Equation 2)}$$

In which, $\theta(c)$ is the occupancy, K_D is the apparent binding affinity, [c] is the ligand concentration, and n is the Hill coefficient.

Bridging assay

The DNA used for the bridging assay is the same as that used for TPM and was ^{32}P-labeled (29). The DNA bridging assay was performed as described previously (27, 33) with minor modifications. Streptavidin-coated Magnetic M-280 Dynabeads (Invitrogen) were resuspended in buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 2 M NaCl, 2 mg/mL BSA (ac), 0.04% Tween 20) containing 100 fmol biotinylated 32% GC DNA (685 bp) and incubated at 1000 rpm for 20 min at 25 °C in an Eppendorf Thermomixer with an Eppendorf Smartblock™ 1.5 mL. The beads with associated DNA were washed twice before resuspension in buffer (10 mM Tris-HCl, pH 8.0, 5% v/v glycerol, 1 mM spermidine, 0.02% Tween20, 1 mg/ml acetylated BSA). Radioactive ^{32}P-labeled DNA and unlabeled DNA were combined to maintain a constant (2 fmol/µl) concentration and a radioactive signal around 8000 cpm, and then added to each sample. Next, protein was added to initiate formation of bridged protein-DNA complexes. Salt concentration (KCl or $MgCl_2$), protein concentration, temperature and pH were individually varied in line with the experiments. For pH 6 and 6.5 10 mM MES (2-morpholinoethanesulfonic acid) was used instead of Tris-HCl and for pH 9, 9.5 and 10 10 mM CHES (*N*-Cyclohexyl-2-aminoethanesulfonic acid) was used. The samples were incubated for 20 min at 1000 rpm at 25°C in an Eppendorf Thermomixer with an Eppendorf Smartblock™ 1.5 mL. After the incubation the beads were washed with the same experimental buffers once and then resuspended in counting buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 200 mM NaCl,

0.2% SDS). The radioactive signal of DNA was quantified by liquid scintillation and was used for the calculation of protein DNA recovery (%) based on a reference sample containing the same amount of labeled ^{32}P 685 bp DNA used in each sample. All DNA bridging experiments were performed at least in triplicate.

Atomic force microscopy

Complexes of DNA and Rok protein were formed by incubating Rok with 50 ng of nicked pUC19 in AFM Buffer A (40 mM HEPES pH 7.9, 10 mM $MgCl_2$, 60 mM KCl) at 37°C for 30 minutes. This mixture was then diluted 20-fold to a final buffer of 5 mM HEPES pH 7.9, 5.5 mM $MgCl_2$, 3 mM KCl. 10 uL of the mixture was deposited onto a freshly cleaved mica disk. After 30 seconds, the mica disk was gently rinsed with 10 mL HPLC water, excess water on the surface was absorbed with lint-free tissue paper, and dried with filtered N_2 gas. Images were acquired on a JPK Instruments NanoWizard 3 system in AC mode using a TESPA probe (Nanoworld) at a resonance frequency of 320 kHz. All images were captured at 512x512 square pixel resolution at a line rate of 1.5 Hz. The images were then processed using JPK Data Processing software. Images were flattened, and contrast was adjusted for clarity.

Microscale thermophoresis

A serial dilution of the Rok or sRok protein was made from 32 to 0.250 µM using dilution buffer (20 mM Tris HCl pH 8, 300 mM KCl, 10% glycerol, 0.1% Tween20 and 0.16 mg/ml acetylated BSA). Then the samples were diluted 1:1 with 80 nM DNA substrate in MilliQ water. This resulted in samples with a constant DNA substrate concentration of 40 nM with (s)Rok concentrations between 16 and 0.125 µM in 10 mM Tris HCl pH 8, 150 mM KCl, 5% glycerol, 0.05% Tween20 and 0.08 mg/ml acetylated BSA. $MgCl_2$ was added in dilution buffer when required. The samples were incubated for 5 minutes at room temperature and transferred to MST capillaries (Monolith NT.115 Premium Capillaries, NanoTemper, Germany). The measurement was done at 40% LED power and medium MST laser power using the NanoTemper Monolith NT.115. Total measurement time was 40 seconds, with 5 seconds laser off, 30 seconds laser on and 5 seconds laser off. F_{norm} values were evaluated after 1.5 seconds of laser on. ΔF_{norm} values were calculated by subtracting F_{norm} of DNA only. Occupancy values were calculated and fitted with a McGhee-von Hippel fitting algorithm assuming a binding site size (n) of 30 bp (27, 34).

Alphafold2 protein structure prediction

For the AlphaFold predictions MMseqs2 and LocalColabFold were run on the high performance computing facility ALICE at Leiden University (35–37). Multiple sequence alignments (MSAs) for Rok, sRok, and Rok:sRok heterodimer were generated with MMseqs2 (38). Target databases used for these MSAs were constructed by the ColabFold team (https://colabfold.mmseqs.com/) and include UniRef30, BFD, Mgnify, MetaEuk, SMAG, TOPAZ, MGV, GPD, and MetaClust2. The search sensitive parameter -s was set to 8. The constructed MSA was used as an input for LocalColabFold to predict dimer structures for Rok, sRok, and Rok:sRok. No templates, 12 recycles, and AlphaFold-Multimer-v2 were used for these predictions. The structures were relaxed by AlphaFold's AMBER forcefield.

His-tag pull down assay

40 µl of HisPur™ Ni-NTA Magnetic beads (Thermo Scientific) per sample were pipetted into an Eppendorf tube. The buffer was exchanged for binding buffer (20 mM Tris-HCl pH 8.0, 500 mM KCl, 50 mM imidazole, 3.5% glycerol, 0.05% Tween, 0.8 mg/ml acetylated BSA) using a magnetic stand. 400 µl of the desired combination of Rok-6xHis and sRok was added to the bead suspension (final concentration in assay 8.75 µM) and incubated 30 minutes while mixing on an end-over-end rotator. The beads were washed once with binding buffer using a magnetic stand. The proteins were eluted with 25 µl of elution buffer (20 mM Tris-HCl pH 8.0, 500 mM imidazole, 3.5% glycerol) and incubated for 10 minutes with end-over-end rotation. The eluate was collected with a magnetic stand and 10 µl was mixed with 5 µl cracking buffer (50 mM Tris-HCl pH 6.8, 1% SDS, 25% glycerol, 1% β-mercaptoethanol, 0.05% bromophenol blue). 10 µl was loaded on a 4-15% mini-PROTEAN® TGX™ Precast protein gel (Bio-Rad) and run at constant voltage (200V) for 35 minutes. After the run, the gel was stained with 0.1% Coomassie Brilliant Blue in destain solution containing 40% ethanol and 10% acetic acid.

In vivo complementation

E. coli hns::kan (strain NT135) cells were created by λ Red recombination (39). The cells were made chemically competent and transformed via heat shock with plasmid pRD408, pRD410, pRD412, pRD424 or an empty pUC19 vector. The transformed cells were plated on lysogeny broth (LB) agar (40) for growth curves, or on MacConkey agar (Sigma-Aldrich) with 0.4% salicin (Sigma-Aldrich) or bromothymol blue (Sigma-Aldrich) indicator plates with 0.5% salicin, all with 50 µg/mL kanamycin and 75 µg/mL ampicillin and incubated overnight at 37 °C.

For growth curve measurements, liquid starter cultures were prepared by inoculating LB medium with the appropriate antibiotics with single colonies from LB agar plates. Cultures were grown until an OD_{600} of 0.8-1.0. Next day, all cultures were diluted to an OD_{600} of 0.01. Next, the OD_{600} value of each culture was measured at regular time intervals during growth at 37°C, 200 rpm. After finishing the growth curve, the plasmids were isolated using the Thermo Scientific GeneJET Plasmid Miniprep Kit and next checked with DNA Sanger sequencing (BaseClear) to rule out the occurrence of mutations.

Construction of B. subtilis strains

All constructed *B. subtilis* strains are derivatives of strain 168. Cassettes allowing the controlled expression of *rok*, *srok* or *srok+rok* were placed at the *amyE* locus of the *rok* deletion strain PKS21 (table S3.1) as follows. Genes *srok* and *rok* were amplified by PCR using as template total DNA isolated from strain PKS11 harboring pLS20 in combination with primer sets [oAND314/oAND315] and [oAND316/oAND317], respectively. The PCR products were digested with HindIII and SalI, or SalI and NheI (New Engeland Biolabs, USA) and then cloned in vector pDR110 (a gift from D. Rudner) digested with the same enzymes to produce pAND520 and pAND521. Plasmid pAND522 was generated cloning the fragment of *rok* behind the srok gene on plasmid pAND520. Plasmid pDR110 is a *B. subtilis amy*E integration vector that contains a multiple cloning site located behind the IPTG-inducible P_{spank} promoter. Next, plasmids pAND520, pAND521 and pAND522 were used to transfer competent PKS21 cells and selecting for spectinomycin resistant transformants. *B. subtilis* competent cells were prepared as described (41). Double cross over events of the resulting strains AND520, AND521 and AND522 were confirmed by the loss of a functional amylase gene.

RNA-seq

B. subtilis strains were grown in Lysogeny Broth (LB) medium (40) or on 1.5% LB agar plates supplemented with spectinomycin (100 μg/ml). For total RNA extraction, the bacteria were grown in liquid media with shaking overnight (ON) at 37 ºC and diluted 1/100 times in fresh LB media with 1mM of IPTG. At OD_{600}=0.8-1, 1.5ml of the culture was harvested by centrifugation and stored at -80ºC. Pelleted cells were thawed on ice and then lysed using a bead mill. Total RNA was isolated using Monarch® Total RNA Miniprep Kit (New England Biolabs, USA) and stored at -80ºC.

Concentrations of RNA samples were measured using the Nanodrop, and 500 ng was loaded on a 1% agarose bleach gel to verify the quantity and quality. The RiboCop rRNA depletion kit (Lexogen Vienna, Austria) was used to remove ribosomal

RNA from 500 ng total RNA. Subsequently, the CORALL Total RNA-SeqTotal RNA Library Kit (Lexogen Vienna, Austria) was used to prepare the library preps for Illumina sequencing. Samples were sequenced on the Illumina NextSeq 1000 to generate 100 bases single end reads (100SE) with an average read-depth of 8-12M reads per sample. The quality of the resulting fastq reads we checked using FastQC v0.11.9 (Babraham Bioinformatics, Cambridge) and mapped on the reference genome (*Bacillus subtilis subsp. subtilis str. 168* (GenBank identifier AL009126.3)) using Bowtie2 v2.4.2 using default settings (42). Resulting SAM files were converted to BAM using SAMtools 1.11 and featuresCounts 2.0.1 was used to obtain the gene counts for each gene annotated in the reference genome taking into account the orientation of genes and reads (43, 44). Duplicate samples were summarized by calculating the mean counts for each gene.

RNA-seq data processing and analysis

To calculate the effect of ectopic induction of *rok*, *srok* or *rok*+*srok* on expression on the *B. subtilis* genes, the average of normalized counts obtained for the Δ*rok* strain PKS21 were subtracted from the normalized counts derived from wild type, AND520, AND521 and AND522 strains. Note that all strains were grown in the presence of IPTG, which avoids differential gene profile effects due to IPTG. These differences were transformed by taking their square root. Plots showing differential gene expression profiles as a function of genome position were generated using the gnuplot tool (www.gnuplot.info). For comparison, such plots were also generated using previously published data of DNA binding of Rok determined by ChIP and DNA coverage techniques (16, 23). We next fitted several distributions to the aggregated values from differences, which revealed that our data were best modelled by the logistic distribution, and therefore we used this distribution to identify differentially expressed genes (DEGs). These exercises were done using R (Foundation for statistical computing (https://www.R-project.org/)) in combination with the package "fitdistrplus" (45) obtaining the following parameters: location=0.0715, scale=0.66927. From this distribution, the associated p-value for every gene in each condition was transformed into a q-value that incorporates a false discovery rate-based multiple testing correction, using the “qvalue” package for R (https://bioconductor.org/packages/release/bioc/html/qvalue.html). Those genes that exhibit an associated q-value < 1E-3 were selected as DEGs. The list of DEGs from each strain was uploaded to the FUNAGE-Pro (FUNctional Analyis and Gene set Enrichment for Prokaryotes) server (46), to detect enrichment in operons and KEGG pathways (47).

Analysis of Rok and sRok sequences

The NCBI protein database was used to search for all sequences annotated as Rok. The resulting sequences were divided based on sequence length: ≥190 aa for Rok and ≤189 aa for sRok as previously suggested (24). Also, sequences NP_389307.1 (Rok) and BAJ76946.1 (sRok) were used for BLAST searches, but no additional candidate Rok proteins were identified. Rok sequences from species with or without sRok present on the chromosome were gathered with Batch Entrez and aligned using EMBOSS Clustal Omega using default parameters(48). Sequence logos were generated using Skylign (https:/skylign.org) using default parameters (49).

Results

Rok compacts DNA

To determine the architectural properties of Rok, we investigated the effect of Rok binding on the conformation of DNA using Tethered Particle Motion (TPM) (30, 32). In TPM experiments the Root Mean Square displacement (RMS) of a bead (exhibiting thermal motion) at the extremity of a DNA substrate attached to a glass surface, provides a readout of DNA conformation. Whereas an increase in RMS following the binding of proteins is indicative of DNA stiffening, an RMS reduction reflects DNA softening, binding or bridging upon protein binding. We investigated the interaction between Rok and an AT-rich (32% GC) DNA substrate, which we used earlier to study the DNA-binding properties of H-NS (27) and MvaT (28, 50). We determined the effect of Rok on DNA conformation by titration from 0-10 nM (figure 3.1A). Bare DNA had an RMS of 159 ± 2 nm. Upon addition of 3 nM Rok, a second population at an RMS of ~105 nm appeared. Saturation of Rok binding was achieved at 10 nM; at this concentration only the population with reduced RMS was observed (figures 3.1A and B). At several concentrations two extra minor populations can be observed at an RMS of ~125 and ~80 nm. As these populations were not consistently present (and therefore do not represent a main Rok-DNA complex) and could not be fitted due to low occupancy, they were not taken into account for occupancy calculations and further analysis (see below). The observed reduction of RMS implies that Rok does not form DNA stiffening filaments along DNA as observed for other H-NS-family proteins under similar conditions (51–53), but instead indicates that binding of Rok compacts the DNA (figure 3.1C). However, it cannot be ruled out that a Rok filament is formed that effectively shortens the DNA, as has been found for HMf proteins (figure 3.1C) (54). The compaction of the DNA would then be due to DNA bending induced by binding of individual Rok proteins in a filament. The reduction in RMS might also be attributed to DNA bending without filament formation as observed

for HU (figure 3.1C) (55) or to DNA bridging. The fact that compaction occurred at low protein concentration and that the structural transition was abrupt suggests cooperative behavior.

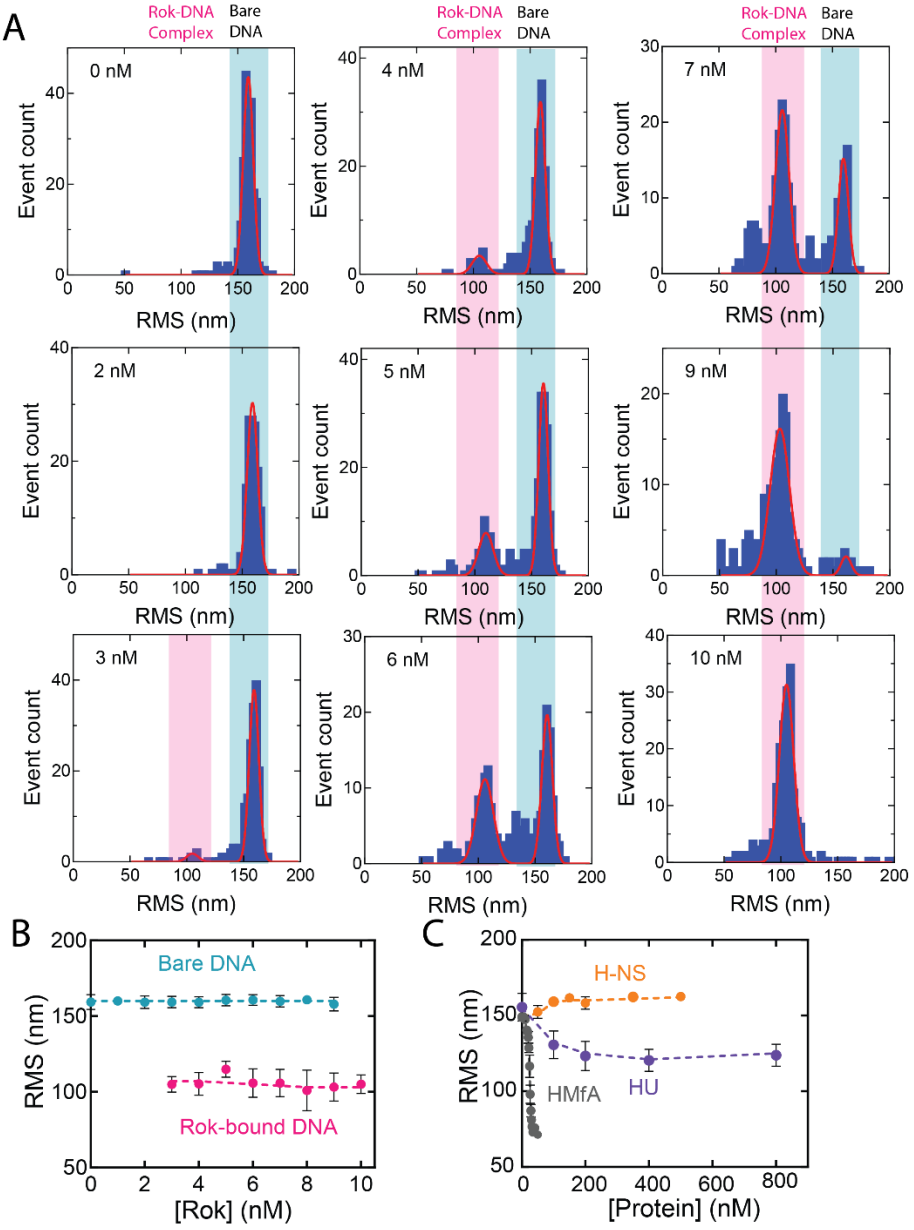


Figure 3.1 *B. subtilis* Rok compacts DNA. A) Histograms of Root Mean Square displacement (RMS) obtained for 32%GC DNA as a function of Rok at concentrations of 0, 2, 3, 4, 5, 6, 7, 9 and 10 nM as measured by TPM in the presence of 50 mM KCl. The histograms were fitted to Gaussian distributions, in which the RMS value at ~150 nm represents bare DNA and the population with an RMS at ~100 nm represents DNA bound by Rok. The bare DNA and Rok-DNA complex populations are highlighted with a light blue and magenta box, respectively. The data for each concentration originates from at least two independent measurements. B) RMS values obtained for 32%GC DNA as a function of Rok at concentrations from 0 nM to 10 nM. Blue and magenta dots represent the average RMS resulting from fitting with a Gaussian distribution, where blue represents bare DNA and magenta Rok-DNA complexes, respectively. Error bars represent the propagated standard deviation from at least two independent measurements. Due to their small size, some error bars are hidden behind the data points. How the DNA tethers are distributed between the two populations is not taken into account in this representation. Dashed lines are lines to guide the eye. C) RMS as a function of protein concentration of *E. coli* H-NS and HU and *M. fervidus* HMfA. Data were taken from van der Valk et al. 2017 (H-NS) (27), Driessen et al. 2014 (HU) (56) and Henneman et al. 2021 (HMfA) (54). Error bars represent the standard deviation; due to their small size they are hidden behind the data points. Dashed lines serve as lines to guide the eye.

Rok is able to bridge DNA

The DNA compaction observed in the TPM experiments described above could have its structural basis in DNA bridging, which is supported by the fact that Rok can induce chromosomal loops (23). This led us to hypothesize that Rok, like H-NS, is able to bridge DNA. We therefore investigated the ability of Rok to bridge DNA in a quantitative biochemical DNA bridging assay, which we used earlier to evaluate the effect of altering physico-chemical conditions on the DNA bridging efficiency of H-NS and MvaT (27, 28). In this assay, biotinylated DNA bound to streptavidin-coated magnetic beads is used as bait in combination with ^{32}P-labeled prey DNA offered *in trans* that can be recovered by magnetic pull-down of beads when bridged by protein. The radioactive signal of the DNA pulled down is a proxy of DNA bridging efficiency. The DNA used in the bridging assay was the same as that used in TPM experiments. The TPM and bridging experiments are fundamentally different in the sense that they interrogate DNA at single molecule and bulk DNA level respectively, which is why higher protein concentrations are needed in the bridging assay (57). To determine whether Rok bridges DNA we carried out a titration with Rok from 0 - 0.5 μM. In the absence of Rok, no radioactive DNA was recovered. DNA recovery increased with increasing Rok concentrations. Saturation of DNA recovery occurred at a Rok concentration of 0.3 μM (figure 3.2A). These data unambiguously show

that Rok is a DNA bridging protein, which we confirmed by AFM imaging (figure S3.1). Rok has a similar DNA bridging efficiency as H-NS at 25ºC, yet reaches this efficiency at 10 times lower concentration (0.3 μM vs 3 μM (27)), which we attribute to the high DNA binding cooperativity of Rok. The minimal length of DNA that Rok could bridge was 100 bp (figure S3.2), which suggests that multiple Rok dimers are needed to form a stable bridge. During natural transformation, a process in which Rok has been implicated (18), dsDNA is processed into ssDNA and only a single DNA strand (ssDNA) is absorbed. Therefore, we tested if Rok might be able to bridge incoming ssDNA with genomic dsDNA and therewith play a possible role in recombination. However, Rok was unable to bridge prey ssDNA with dsDNA bait (figure S3.2). Some bridging was observed with a 685 bp ssDNA substrate, but this was most likely due to binding of Rok to dsDNA structures generated by internal hybridization as we were unable to reproduce these observations using poly(A) ssDNA. As internal hybridization of ssDNA can also occur inside the cell, Rok might still play a role in recombination by bridging small DNA structures to genomic DNA.

DNA bridging activity of Rok is only mildly sensitive to environmental conditions

Bacteria adapt to environmental changes and environmental cues are known to have a direct effect on the function of H-NS-like proteins (51, 58–61). *B. subtilis*, found in soil and the gastrointestinal tract of livestock and humans, is exposed to rapid changing conditions, which requires an ability to adapt to different environmental conditions via changes in transcription of specific genes. To study if Rok plays a direct role in this response by responding to environmental changes through altering its binding (mode) at Rok-regulated genes as seen for H-NS (58–60, 62, 63), we tested the effect of various physico-chemical conditions on Rok's DNA bridging efficiency.

First, we investigated the effect of temperature and pH. An increase in temperature from 25 ºC to 37 ºC caused a slight drop in DNA recovery from 80 to 60% in DNA bridging assays (figure 3.2A). DNA was also rather efficiently recovered over a pH range from 6 to 10 (figure 3.2B). Strikingly, even crossing the pI of Rok (9.31) did not interfere with its capacity to bridge DNA.

Besides temperature and pH, *B. subtilis* is frequently challenged to adapt to osmotic up- and downshifts in its natural habitat (64, 65). To determine whether Rok's DNA bridging activity is osmo-sensitive, we investigated the effect of changing concentrations of monovalent and divalent cations. An increase in concentration of KCl from 50 mM to 300 mM had no significant effect on the DNA bridging activity of Rok (figure 3.2C). A mild decrease in DNA recovery from 80 to 60% was observed when increasing the $MgCl_2$ concentration from 0 to 20 mM and this remained constant until 60

mM (figure 3.2D). This result might indicate that the Mg^{2+} concentration modulates the affinity of Rok for DNA as has been previously suggested for H-NS (66). To check if this is the case, we performed microscale thermophoresis (MST) experiments with Rok and a 78 bp DNA substrate. Without $MgCl_2$ an apparent binding affinity (K_D) of 79.4 ± 20 µM was obtained (figure S3.3). The affinity improved slightly in the presence of 10 mM $MgCl_2$ (K_D of 34.6 ± 8.0 µM), but it decreased substantially with $MgCl_2$ concentrations above 25 mM (K_D of around 800 µM). We concluded that Rok exhibits the highest DNA binding affinity at 10 mM $MgCl_2$ in contrast to what has been previously observed for H-NS. Thus, instead of decreased DNA binding observed for H-NS, Rok displayed increasing DNA binding activities at 5-20 mM $MgCl_2$. We also tested a specific DNA substrate, which differs from the aspecific substrate by having a specific Rok binding site (TACTA) present in the middle, which was found previously to be one of the most favorable for Rok binding (67). We observed similar behavior as for the aspecific DNA substrate, but Rok-binding remained in the high affinity regime of K_D ~ 100 µM over a wider range of $MgCl_2$ concentration (up to 35 mM $MgCl_2$ instead of 25 mM) before transitioning to the low-affinity regime (K_D ~ 800 µM) (figure S3.3).

The intracellular concentrations of K^+ and Mg^{2+} in *B. subtilis* have been determined to be 27 ± 10 mM and 1-2 mM respectively (68). Figures 3.2C and D show that DNA recoveries are at maximum levels at these cation concentrations; the decrease in DNA recoveries observed above 400 and 50 mM concentrations of KCl and $MgCl_2$ respectively are well above the physiological concentrations and hence are unlikely to be relevant under natural conditions. We attribute the reduction in DNA recovery at high cation concentration to complete disintegration of DNA-Rok-DNA bridged complexes, associated with the reduction in DNA binding affinity, rather than a switch from a Rok-DNA bridge to a Rok-DNA nucleofilament. Unlike H-NS and MvaT (27, 28), the formation of bridged Rok-DNA complexes also does not require a particular concentration of monovalent (K^+) or divalent (Mg^{2+}) cations. These observations highlight that DNA bridging activity of Rok neither requires KCl or $MgCl_2$ for bridging, nor is strongly inhibited by these ions at biologically relevant concentrations. To examine if Rok might be able to induce liquid-liquid phase separation (LLPS), 1,6-hexanediol, a commonly used alcohol to dissolve LLPS assemblies (68-70 and references therein), was added to the bridging assay before Rok-DNA bridge formation. DNA was efficiently recovered up to 5% 1,6-hexanediol (figure S3.4), suggesting that Rok does not form LLPS assemblies.

Taken together, these observations indicate that the DNA bridging activity of Rok is only mildly affected by changes in physico-chemical conditions. This is unexpectedly different from H-NS and MvaT where much larger effects were observed (27, 28, 53, 58).

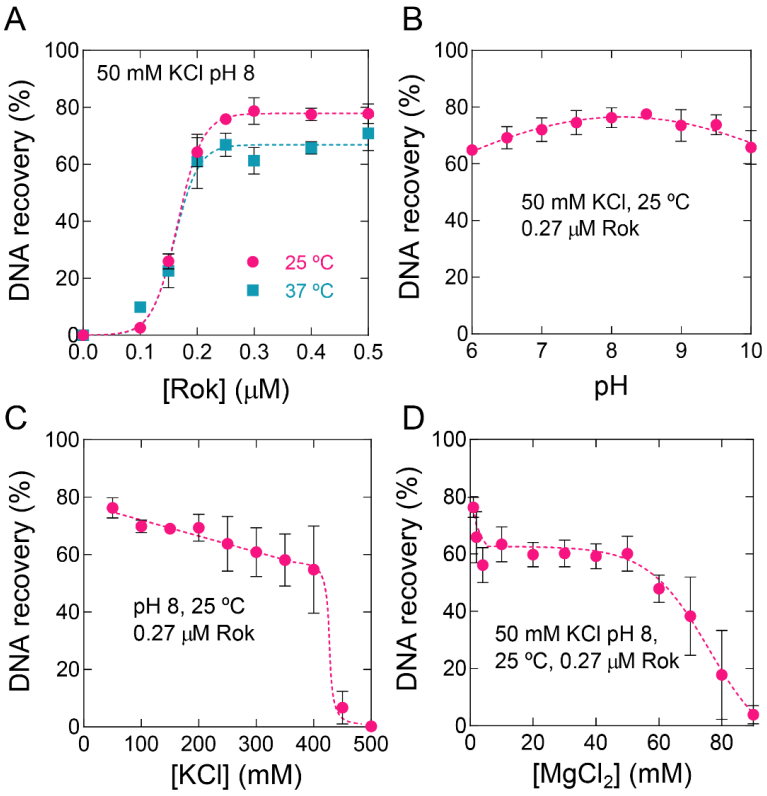


Figure 3.2 Rok exhibits DNA bridging activity, which is only mildly affected by temperature, pH and salt concentration A) DNA recovery (as a percentage of the input DNA) as a function of Rok concentration from 0 to 0.5 μM as measured using the DNA bridging assay in the presence of 50 mM KCl at 25 ºC (red) and 37ºC (blue), respectively. B) DNA recovery as a function of pH from 6 to 10 in the presence of 0.27 μM Rok at 25 ºC. C) DNA recovery as a function of KCl concentration from 35 to 500 mM in the presence of 0.27 μM Rok at 25 ºC. D) DNA recovery as a function of $MgCl_2$ concentration from 1 to 90 mM in the presence of 0.27 μM Rok at 25 ºC. Data are plotted as mean values of three independent measurements and the error bars represent the standard deviation. Dashed lines serve as lines to guide the eye.

The neutral linker of Rok has a role in DNA binding cooperativity

Previously, we found that H-NS-like proteins have a conserved asymmetrical charge distribution, with the N-terminal domain mainly negatively charged and the linker and C-terminal domain positively charged (8). This asymmetrical charge distribution is needed for interdomain interactions which are characteristic for nucleoprotein filament formation (28, 72). Rok is a notable exception with a less pronounced charge distribution and a neutral linker (8). Therefore, we proposed that Rok cannot form a nucleoprotein filament because of weaker interdomain interactions. We attempted to test if the introduction of charges could result in DNA stiffening behavior as observed for H-NS (figure 3.1C). Unfortunately, a recombinant Rok variant in which the neutral part of the linker was replaced with the (charged) H-NS linker was present in the insoluble fraction after cell lysis (data not shown). Next, we investigated whether removal of Rok's neutral amino acids (residues 75-96) would allow the recombinant protein Rok Δ75-96 to form nucleofilaments (figure 3.3A). Although, this Rok variant could successfully be expressed and purified, TPM experiments with the Rok Δ75-96 variant revealed the same level of DNA compaction as observed for wild type Rok (Rok WT figure 3.3B and S3.5A). This strongly indicates that removal of the neutral linker in Rok does not promote DNA stiffening and that hence this Rok variant does not form filaments upon DNA binding.

Removal of the linker however affected DNA binding cooperativity, as evident from the less steep increase of DNA occupancy for the Rok Δ75-96 variant compared to the wild type protein (figure 3.3C). When fitted to the Hill equation, affinities for Rok WT and the Rok Δ75-96 variant were 6.1 ± 0.07 and 8.8 ± 0.3 nM respectively and the Hill coefficients (*n*) were 6.5 ± 0.4 for Rok wildtype and 5.7 ± 0.9 for the Rok Δ75-96 variant. Similar behavior was observed for the Rok Δ75-96 variant in the DNA bridging assay: it can bridge DNA with similar efficiency as Rok wildtype, but with the transition from low to high DNA recovery occurring over a wider concentration range: between 0.1-0.3 μM for Rok wildtype and between 0.1-0.5 for Rok Δ75-96. (figure S3.5B). Together, these results show that the deletion of the neutral linker in Rok does not affect DNA filament formation and bridging, but that it affects the cooperativity of DNA binding.

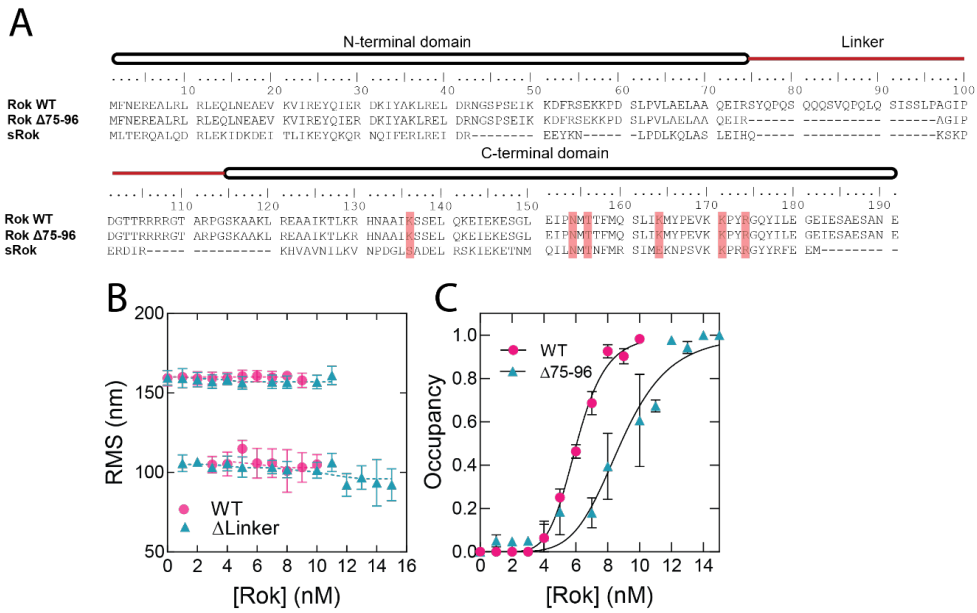


Figure 3.3 The linker domain of Rok is important for its cooperative DNA binding. A) Sequence alignment of Rok wildtype (NCBI accession number: NP_389307.1), Rok Δ75-96 and sRok (NCBI accession number: YP_004243533.1). The N- and C-terminal domain are indicated with a black box and the linker domain with a red line. The residues important for DNA binding are highlighted in red. B) RMS values obtained for 32%GC DNA as a function of Rok Δ75-96 concentration as measured by TPM in the presence of 50 mM KCl. The RMS values were determined from fitting with a Gaussian distribution. For reference, Rok WT data is shown (reproduced from figure 3.1B). Error bars represent the propagated standard deviation from at least two independent measurements. C) DNA occupancy (fraction of Rok-bound tethers of the total amount of DNA tethers) as function of protein concentration in nM. The data points were fitted using the Hill binding model.

sRok exhibits nucleoprotein filament formation and can modulate the DNA bridging activity of Rok

Besides the artificial Rok Δ75-96 variant, a number of *Bacillus* species and strains encode a variant of Rok that lacks the linker region. We refer to this variant as small Rok (sRok) (8) (figure 3.3A). sRok was first identified on the large conjugative *B. subtilis* plasmid pLS20 (24). It was shown to associate with the host chromosome and to be able to replace Rok as regulator in the competence pathway (24). Therefore, we expected that Rok and sRok would have similar DNA binding properties, while the sensitivity to the environment may differ due to the lack of a linker in sRok.

To test this, we investigated the DNA binding properties of purified sRok in TPM assays using the same DNA fragment as for Rok in figure 3.1. The experiments revealed that the RMS of the DNA tether was only mildly – if at all – affected by protein concentration (figure 3.4A). While this could point at low DNA binding affinity or non-functional protein, DNA bridging experiments (see below) demonstrate that the protein binds DNA. These observations suggest that sRok binds DNA without affecting DNA conformation, which is in sharp contrast to the DNA compaction observed upon Rok binding (figure 3.1A). In the bridging assay, sRok reached its maximal DNA recovery in the same concentration range as Rok (figure 3.4B).

We considered that Rok might be functionally modulated by sRok, analogous to the modulation of DNA bridging efficiency of H-NS and other H-NS-like proteins by their protein-partners (8). Rok has been shown to dimerize via its N-terminal domain (67). The possibility that Rok and sRok hetero(di)merize is realistic considering the high level of conservation between the N-terminal multimerization domain of Rok and sRok (8). Using Alphafold2 we predicted the structures of Rok and sRok homodimers and of Rok-sRok heterodimers (figure 3.4C) with high confidence levels (figure S3.6A-C). Both proteins exhibit a similar tertiary structure and dimerize using their N-terminal helix (residues 1-47). At the monomer-monomer interface, the two α-helices interact in a coiled-coil-like manner and form similar salt bridges and hydrophobic interactions in all three predicted structures (figure S3.6D). We also attempted to predict higher order structures to gain insight in the oligomerization interface, but Alphafold2 was unable to generate a confident model. To test whether the predicted heteromerization indeed can occur we generated a Rok variant with a 6xHis-tag at the C-terminus that we exploited for pulldown using HisPur™ Ni-NTA Magnetic beads (see Materials and Methods). Figure 3.4D shows that sRok lacking a His-tag was pulled down together with Rok-6xHis confirming heteromerization of the two proteins. Next, we performed bridging assays using different ratios of Rok and sRok (figure 3.4E). Interestingly, the presence of only 20% of sRok in a Rok:sRok mixture was sufficient to lower the DNA recovery to about 40%, similar to the DNA recovery of sRok alone. This latter result not only supports that Rok and sRok can form heterodimers, but also shows that sRok can modulate the DNA bridging efficiency of Rok. To investigate further the interplay between Rok and sRok we studied the effects of salt concentration on the DNA bridging efficiency of sRok and the Rok:sRok complex. For this, bridging experiments were performed using different KCl concentrations for both sRok and a 1:1 Rok:sRok complex. While Rok was only mildly sensitive to these changes (figure 3.2 and 3.4F), DNA bridging by sRok peaked around 130 mM KCl, but became strongly inhibited at higher KCl concentrations (figure 3.4F). MST also showed that sRok is less tolerant than Rok to increasing $MgCl_2$ concentrations (figure S3.7). At 20 mM

$MgCl_2$ or higher the binding was too weak to fit a K_D value. Also, the presence of a specific high-affinity DNA sequence for Rok yielded no improvement in sRok binding. The Rok:sRok complex showed intermediate behavior with a constant DNA bridging efficiency up to 200 mM KCl (figure 3.4F). These results show that – in contrast to Rok - sRok-mediated DNA bridging is osmo-sensitive and that to a lesser extent this also translates to the Rok:sRok complex. These results open up the possibility of controlling Rok-mediated gene repression via protein partners such as sRok.

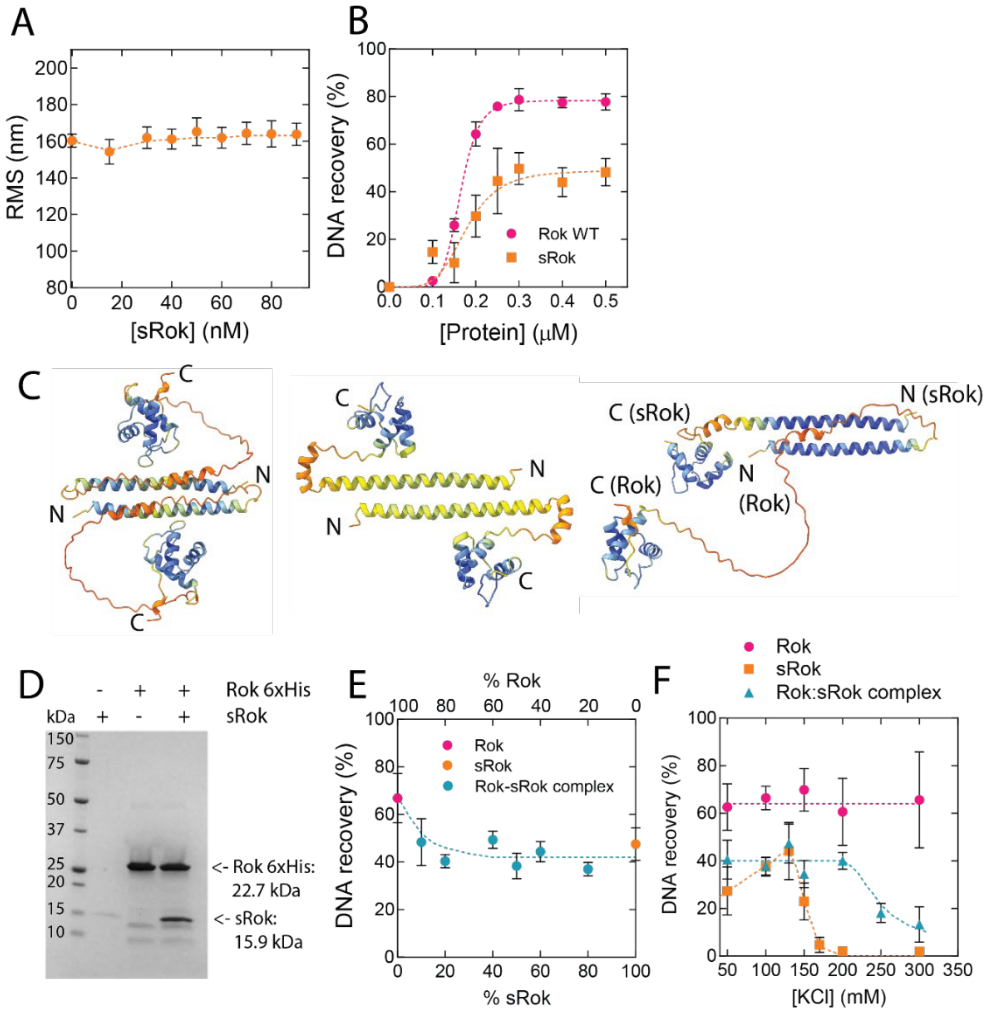


Figure 3.4 The bridging activity of sRok can be modulated by salt concentration A) RMS values obtained for 32%GC DNA as a function of sRok concentration as measured by TPM in the presence of 50 mM KCl. The RMS values were determined from fitting with a Gaussian distribution. Error bars represent the propagated standard deviation from at least two independent measurements. Some error bars are hidden behind the data points. B) DNA recovery (%) as function of (s)Rok concentration in the presence of 50 mM KCl at 25°C. For reference, Rok WT is shown (reproduced from figure 3.2A). Data are plotted as mean values from three independent measurements and the error bars represent the standard deviation. Dashed lines serve as lines to guide the eye. C) Structural predictions using Alphafold of Rok homodimer, sRok homodimer and Rok:sRok heterodimer (left to right). The protein structures are colored by the predicted DT-Cα (PLDDT) values with the following color scheme: blue (> 90), light blue (90 to 70), yellow (70 to 50), orange (< 50). The PLDDT indicates the local confidence in the predicted structures from 0-100, with 100 corresponding to highest confidence. D) SDS-PAGE analysis of His-tag pull down assay. Rok 6xhis was captured on HisPur™ Ni-NTA Magnetic beads and, when applicable, sRok was added in 1:1 molar ratio. E) DNA recovery (%) measured in the presence of 50 mM KCl + 1 mM $MgCl_2$ at 25°C with different ratios Rok:sRok. The total amount of protein used was constant at 0.5 μM. F) DNA recovery (%) measured in the presence of 1 mM $MgCl_2$ at 25°C and 0.5 μM protein with different KCl concentrations. Dashed lines serve as lines to guide the eye.

Rok and sRok cannot complement the absence of hns in E. coli

The results in figure 3.4 might suggest that sRok is functionally more similar to H-NS than Rok due to its responsiveness to changes in osmolarity. To test this hypothesis, we performed *in vivo* complementation experiments in *E. coli.* Previously, this approach showed that MvaT and Lsr2 can complement the absence of H-NS *in vivo* (73, 74). We used *E. coli hns::kanR* (NT135) and transformed it either with an empty pUC19 vector or with a pUC19-derived vector containing the *hns* promoter followed by the respective protein coding sequence. One of the best-characterized operons that is repressed by H-NS is the *bgl* operon (75). Expression of this operon, which results in the uptake and fermentation of aryl-β-D-glucosides (76), has been used to test whether H-NS variants or potential H-NS like proteins can complement the *hns* knockout phenotype. The pH difference caused by either the ability (acidic pH) or the inability (basic pH) to ferment an aryl-β-D-glucoside can be visualized on MacConkey agar or bromothymol blue indicator (BTB) plates supplemented with salicin (74, 77, 78). As expected, *E. coli* NT135 harboring the empty plasmid was able to use salicin as carbon source resulting in red and yellow colonies on MacConkey (figure S3.8A) and BTB agar plates (figure S3.8B) respectively. *E. coli* NT135 ectopically expressing *hns* from the plasmid (containing the *hns* gene) cannot use salicin and grow on peptone. They formed yellow

and blue colonies on MacConkey (figure S3.8A) and BTB plates (figure S3.8B), respectively. When transformed with pRD424 (Rok), the colonies were red on MacConkey agar, sometimes with a yellow halo around them and yellow on BTB plates, indicating that Rok represses the *bgl* operon at most only partly. Different results were obtained for sRok; the ectopic expression of *srok* resulted in formation of yellow colonies on MacConkey plates, indicating that the *bgl* operon is repressed in these cells. However, on BTB plates also yellow colonies were obtained indicating that the *bgl* operon is expressed. This discrepancy might be explained by partial repression of the *bgl* operon and different sensitivities of the pH indicator used in the respective plates. Finally, cells expressing *rok Δ75-96* formed yellow and brown/green colonies on MacConkey and BTB plates, respectively. This indicates that Rok Δ75-96 can repress the *bgl* operon to a better extent than Rok and sRok, but cannot achieve the complete repression caused by H-NS.

The (dis)ability of Rok, sRok and Rok Δ75-96 to complement H-NS *in vivo* was further analyzed by studying possible reversion of growth defects observed for the *Δhns* strain NT135. Growth curves of *E. coli* NT135 expressing *hns, rok, srok* or *rok Δ75-96* are shown in figure S3.8C. Expression of either Rok or sRok did not restore the growth defect caused by the lack of H-NS; in fact, their expression aggravated the growth defect caused by the lack of H-NS. This indicates that both Rok and sRok cannot bind to and repress most of H-NS-regulated genes but instead may bind to non-H-NS-regulated genes, which could explain the detrimental effect of (s)Rok on cell growth. Another, not mutually exclusive, possibility is that (s)Rok binds (most) H-NS-regulated genes, but exhibits a different mode of DNA binding, causing aberrant gene expression. Interestingly, cells expressing *rok Δ75-96* grew at the same rate as the positive H-NS control, indicating complementation of the *Δhns* growth defect by Rok Δ75-96. This suggests substantial occupation by the Rok Δ75-96 variant of genomic sites otherwise bound by H-NS with regulatory function or key function in transcriptional regulation or structural organization.

Rok and sRok exert different effects on transcription in B. subtilis

The *in vitro* results presented above demonstrate that while Rok and sRok are both able to bridge DNA, they exhibit different physico-chemical behaviors. Due to these differences, the two proteins may exert distinct effects on transcription. Upon heterologous expression in *E. coli* we indeed observed different abilities of the proteins to complement H-NS in a *Δhns* strain. We next directly investigated the possibility that Rok and sRok have distinct effects on transcription in *B. subtilis* using an RNAseq approach. First, we generated a *rok* null mutant (Δ*rok*) and used this strain to construct derivatives containing a cassette at the *amyE* locus having a copy of s*rok* or *rok* under

the control of the IPTG-inducible promoter P_{spank} (strains AND520 and AND521, respectively). In addition, we constructed a strain (AND522) in which P_{spank} drives the expression of both *srok* and *rok*. Total RNA was isolated from these and control strains grown under the same conditions in the presence of 1 mM IPTG. After processing, the RNA samples were used to generate cDNA libraries using a ‘‘directional RNAseq” procedure that preserved information about a transcript’s direction. The generated libraries were subjected to Illumina sequencing to generate 100-nt fragments, and those that passed quality controls (see Material and Methods) were used to calculate the apparent expression level of individual genes.

Several genes known to be repressed by Rok are given in figure 3.5A and table S3.5 that also lists the effect of ectopic expression of Rok in a Δ*rok* background on these genes observed in our studies. The table shows that under these conditions *rok* was expressed 2-fold higher from the P_{spank} promoter as compared to its native promoter. Importantly, the observation that most of these reported Rok-regulated genes were also repressed in our experiments in which *rok* was expressed form its native or the P_{spank} promoter validates our approach (figure 3.5A, table S3.5). The exceptions for which the expression was not or only slightly affected in our experiments can be explained by the differences in growth conditions in the different studies. For instance, our samples were taken at late exponential growth phase from cultures growing in rich LB medium when *comK* and the *sdp* operon are expressed at only very low levels. Remarkably, ectopic expression of *srok* or *srok*+*rok* affected several of these genes differently compared to *rok*. For example, *sboA* was down- and upregulated by Rok and sRok, respectively, and *htpX* was upregulated in the presence of both Rok+sRok, but downregulated by each of the two Rok variants individually. Thus, the two highly related transcriptional regulators (48% sequence similarity for the total proteins, 55% similarity for the DNA binding domains) do not seem to target an identical set of genes. Moreover, the Rok variants seem to affect each other’s role in transcription regulation. To study the possibility that besides the small set of selected genes mentioned above also other genes are differentially regulated by Rok, sRok and Rok+sRok, we plotted the expression levels along the entire genome for each of these three overexpressing strains with respect to the Δ*rok* strain (figure S3.9). These plots show that Rok, sRok and Rok+sRok affect the expression at multiple different loci along the entire *B. subtilis* chromosome and confirm therefore that their regulatory effect is not identical. In line with the results presented in figure 3.5A, these plots also show that the expression profile observed for simultaneous expression of Rok and sRok is distinct to those observed for the individual Rok variants. The genome wide expression data were then used to select statistically differentially expressed genes (DEG) (qvalue <1e-3). This resulted in 175, 252 and 259 DEG for

conditions in which sRok, Rok and Rok+sRok were overexpressed, respectively (tables S3.6, S3.7 and S3.8). These numbers correspond to about 5% of all protein-encoding *B. subtilis* genes. To gain insight into (dis)similarities between the DEG of the three experimental conditions they were presented as Venn diagram (figure 3.5B) that provided the following information. First, different subsets of DEG were regulated by both Rok and sRok (132 genes [38.4%]), or by Rok+sRok with either Rok (193 genes [56.1%]) or sRok (129 genes [37.5%]). Second, about one third of the DEG was regulated under each of the three different conditions (112 genes [32.6%]). Thus, as might have been expected for two highly related proteins, substantial overlap in DEG were observed for Rok, sRok and Rok+sRok. However, as already hinted at by the analysis of a small number of genes (figure 3.5A), the sets of genes regulated by Rok, sRok and Rok+sRok are not identical. Besides considerable subsets of DEG being affected by only Rok (120 genes, [34.8%]) or sRok (43 genes, [12.5%]), a substantial subset of DEG was affected by only Rok+sRok (49 genes, [14.2%]). Particularly this latter observation strongly suggests that the regulatory capacity of Rok is influenced by sRok and *vice versa*. Importantly, these results are in line with the *in vitro* data presented above which show that the presence of sRok affects the DNA binding properties of Rok (figure 3.4).

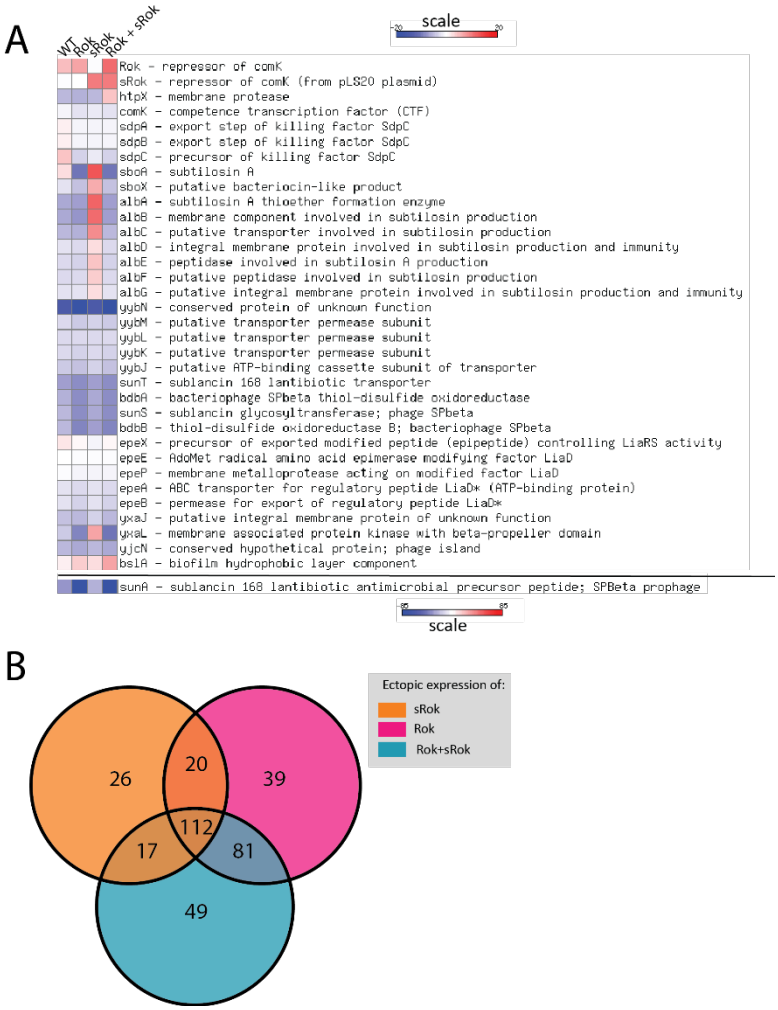


Figure 3.5 Effects of Rok, sRok and Rok+sRok on the expression of *B. subtilis* genes
A) Heat map representations of a set of known Rok-regulated genes in response to *rok*, *srok* and *rok+srok* expressed from an ectopic promoter. The heat map as a response to *rok* expressed from its native promoter is included as a control. Changes in expression observed for *sunA* in response to the different *rok* variants exceeded the -20 to +20 range. Therefore, for clarity, this gene was plotted separately using a larger range. Blue and red color reflect lower and higher expression levels with respect to the *Δrok* strain. B) Venn diagram of differentially expressed *B. subtilis* genes when *rok*, *srok* or *rok+srok* are ectopically expressed in an otherwise isogenic background. The number of DEG are indicated for each of the three different conditions: ectopic expression of *rok*, *srok* or *rok+srok*. Numbers in the intersections correspond to DEG shared by the corresponding different conditions.

The results above show that expression of Rok, sRok and Rok+sRok affect non-identical sets of DEGs. Possibly, the observed differences in DEGs may underlie alterations of distinct pathways or regulons. To study this possibility the differential RNAseq data for strains expressing *rok*, *srok* or *rok*+*srok* with respect to the Δ*rok* strain were uploaded to the Funage Pro server that allows automatic analysis of gene set enrichment (46). Evidence that Rok, sRok and Rok+sRok indeed affect certain operons or pathways differently was obtained. For instance, whereas the presence of Rok or sRok alone did not affect expression of any of the six genes in the *dhb* operon that encodes the biosynthesis of the siderophore bacillibactin (79), all six genes of the operon were highly expressed in the presence of both Rok and sRok (figure 3.6A). Other effects were observed for the *dlt* operon that encodes proteins required for incorporation of d-alanine in teichoic acids (80) (figure 3.6B). In this case, except for *dlt*C, the presence of sRok hardly affected expression of the *dlt* genes, while Rok caused more than 6-fold decrease in expression of all *dlt* genes. Yet other effects were observed for the genes involved in the non-ribosomally synthesized lipopeptide antibiotics surfactin and lichenysin D (figure 3.6C and D). Thus, whereas Rok strongly stimulated expression of both the *srf*A-C and *lic*A-C/H genes, sRok repressed the *srf*A-C genes but hardly affected expression of *lic*A-C/H. Interestingly, the presence of Rok+sRok acted differently on these different operons. As mentioned, whereas Rok and sRok alone did not affect much the bacillibactin genes, these genes were upregulated in the presence of Rok+sRok. Expression of the *dlt* genes in the presence of Rok+sRok was similar to that observed for Rok alone, suggesting that sRok is unable to alter Rok-mediated expression of these genes. The opposite though was observed for the surfactin genes, in which very similar expression profiles were observed for Rok+sRok and sRok alone. Altogether, the RNAseq data analysis demonstrate that Rok and sRok alter the expression of non-identical sets of genes and that Rok and sRok affect each other´s regulatory activity.

Presence of Rok and sRok sequences suggest multiple horizontal gene transfer events

sRok was first identified on the *B. subtilis* plasmid pLS20 (24). Its presence on a plasmid implies that sRok is not always present in all *Bacillus* species and strains therein. Also, Rok itself was found to be present in only a subset of *Bacillus* species (15). An extensive search in the NCBI protein database extracted 1085 sequences annotated as (s)Rok protein. The *Bacillus* species that contain a *rok* gene form a cluster within group 1 of *Bacillus* species (figure 3.7A) (81). While in most species only *rok* was found, some *Bacillus* species have both the *rok* and *srok* gene on their chromosome, most notably *B. licheniformis, B. sonorensis* and *B. subtilis subsp. spizizenii*.

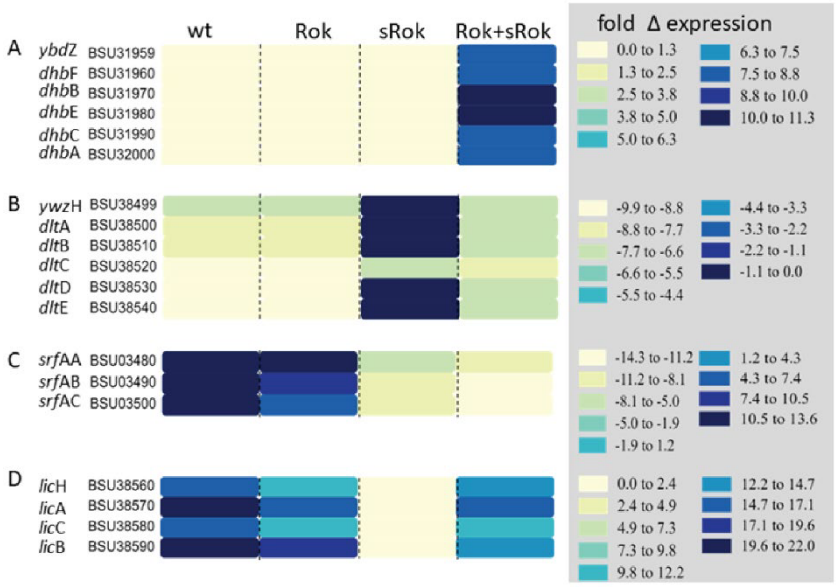


Figure 3.6 Differential effects of either Rok, sRok and Rok+sRok on sets of functionally related *B. subtilis genes.* Heat map representations of genes involved in the synthesis of bacillibactin (A), incorporation of D-alanine in teichoic acids (B) or the non-ribosomally synthesized lipopeptide antibiotic surfactin (C) and lichenysin D (D). Differential expression profiles were generated using the Funage Pro webserver (46). Note that the range of the color scale indicating the expression levels is different for each panel.

Most likely, *srok* was introduced to these lineages via separate horizontal gene transfer events as its location on the chromosome is not conserved in contrast to *rok* (24). A single horizontal gene transfer event was previously proposed for *rok,* where it was introduced in the common ancestor of the *B. subtilis* - *B. licheniformis* - *B. amyloliquefaciens* group (15, 24). We investigated whether the Rok protein in *Bacillus* species containing also an *srok* gene was adapted to the presence of this binding partner. When comparing the sequence logos of Rok from *Bacillus* species encoding sRok or not, only minor differences in amino acid occurrences were observed, for example at position 75 where both isoleucine and methionine can be found (figure 3.7B). Most likely, these minor differences arise from general differences between the *Bacillus* species and the number of Rok sequences used per sequence logo (620 Rok sequences without sRok, 110 sequences with sRok) rather than adaptation of the Rok protein.

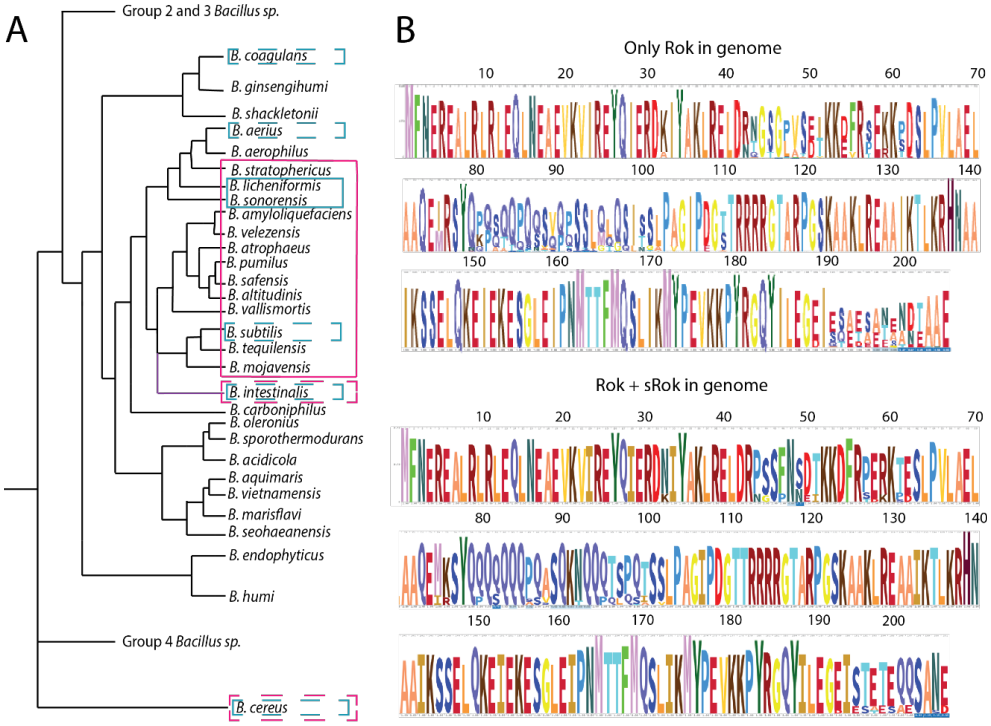


Figure 3.7 Rok and sRok are only present in a subset of *Bacillus sp.* A) Phylogenetic tree of *Bacillus sp.* Group 1 adapted from Wang and Sun 2009 (81) and extended with *B. intestinalis* according to Tetz and Tetz 2017 (82). The distances between species are not on scale. A purple line was used for *B. intestinalis* as its exact position is currently unknown. Magenta indicates the presence of a *rok* gene in the genome and blue an *srok* gene. The presence of a gene in only the minority of the genomes available is indicated with dashed lines. (B) Sequence logos of the Rok sequences of genomes without (top) or with (bottom) an *srok* gene present.

Discussion

In this work we have discovered that Rok is a DNA bridging protein. This feature is in line with the recent finding that, besides being a transcriptional regulator, Rok plays an important role in chromosome organization by formation of long-ranged loops (23). Formation of such loops is most likely a direct consequence of the DNA bridging capacity of Rok that we established here. Detailed analysis showed that DNA bridging by Rok is only mildly sensitive to changes in physico-chemical conditions (figure 3.2). H-NS, MvaT and Lsr2 can use these conditions to switch between DNA bridging and formation of a

nucleoprotein filament along DNA, causing DNA stiffening (51–53, 83). The bridging activity of H-NS and MvaT can be modulated by both monovalent (Na^+, K^+) and divalent (Mg^{2+}, Ca^{2+}) cations (27, 28, 53, 84). For Lsr2, the effect of changes in ionic strength on the protein's DNA binding properties has not been investigated in detail (51). The formation of either nucleoprotein filaments or DNA-protein-DNA bridges by H-NS family proteins is sensitive to temperature, pH and salt, which facilitates cellular response to environment. We did not observe such a switch for Rok, instead only DNA compaction was observed (figure 3.1A), which most likely can be attributed by DNA bridging. It has been reported that binding of Rok's isolated DNA binding domain to DNA causes a bend of around 25° (67). Therefore, it cannot be fully excluded that DNA bending contributes to the DNA compaction behavior of Rok. We propose that Rok bridges DNA by employing dimeric Rok as bridging units (figure 3.8). This is different from H-NS-mediated DNA bridging which requires the formation of an H-NS multimer (27). At least 100 bp were needed to recover DNA (figure S3.2), which indicates that multiple dimers are needed to form a stable Rok-DNA bridge. Therefore, in our model, Rok dimers cluster cooperatively due to the high local DNA concentration adjacent to existing Rok mediated bridges (figure 3.8). We cannot rule out the formation of oligomers once Rok is bound to the DNA. Earlier studies suggest that Rok is capable of forming oligomers in solution (67), but we were not able to predict oligomeric structures with Alphafold2. Possibly, oligomers are only formed in solution at very high protein concentrations. Similarly, oligomer formation may be favored while bridging DNA due to the intrinsic cooperative nature of this process. This might be comparable to HU switching from its DNA bending to DNA stiffening mode which involves oligomerization along the DNA at high protein concentrations (85). But, as we did not observe DNA stiffening upon binding of Rok, it remains unclear if Rok forms oligomers when bound to DNA and whether this is needed for stable bridge formation.

Although the sequence similarity of Rok with H-NS, MvaT and Lsr2 is low, these proteins share a similar domain organization. Structural studies have revealed that H-NS, Lsr2 and MvaT have an N-terminal oligomerization domain consisting of two dimerization sites, a C-terminal DNA binding domain and a flexible linker region (9, 86–89). Rok has a C-terminal DNA binding domain as well and the N-terminal domain was suggested to be responsible for oligomerization (67). A main difference we identified previously, is the presence of a neutral linker for Rok, while the other proteins have a clear asymmetrical charge distribution (8). Here we showed that this neutral linker is not responsible for the lack of a strong response of Rok to changes in biologically-relevant conditions as we suggested previously (8). However, we cannot rule out that the subtle changes in Rok's DNA bridging capacity under several physico-chemical conditions might in fact affect gene expression. Removal of the neutral linker decreased the DNA binding cooperativity

of Rok and we showed that Rok Δ75-96 can (partially) complement the absence of *hns* in *E. coli*. Therefore, the detrimental effect of ectopic Rok and sRok expression on *E. coli* growth may be due to the high cooperativity in DNA binding of Rok and sRok.

Contrary to Rok, sRok is responsive to different KCl concentrations and is able to form heterodimers with Rok, making the Rok:sRok complex also osmo-sensitive (figure 3.4). In *Bacillus sp.* where both Rok and sRok are present, homo- and heterodimers exist in an unknown ratio, making part of the complexes osmo-sensitive (figure 3.8). Therefore, we propose that protein-protein interactions with Rok are a more important, primary mechanism to regulate genes repressed by Rok than changes in environmental conditions. Because *srok* is only present in a subset of *Bacillus sp.*, either on a plasmid or encoded on the chromosome (figure 3.7), this hints at the existence of other convergently evolved sRok-like NAPs in sRok deficient *Bacillus sp*. Also the fact that *srok* is nearly exclusively present together with *rok* – in contrast to *rok* itself, which is frequently found alone – suggests that sRok is rather a modulator of Rok activity than a main NAP itself. Our RNA seq data and Funage Pro analysis show that both Rok and sRok on their own and their combination have a unique, only partially overlapping, regulon. Particularly the observation that the combination of Rok and sRok is not merely adding the individual genes/regulons together, suggests an extensive interplay between the two proteins. However, we cannot exclude indirect effects on transcription mediated by the effect of Rok and sRok on the expression level of other transcriptional regulators and/or proteins involved in chromatin organization (figure S3.10A). HBsu (generally referred to as HU) was downregulated upon ectopic expression of (s)Rok, while its level of transcription is enhanced in the wildtype compared to the *Δrok* strain. Second, the combined expression of Rok and sRok activated gyrase and topoisomerase genes. Which parts of the different expression patterns observed in figure 3.5 are direct consequences of (s)Rok and which are indirect effects, remains therefore unanswered.

The distinct regulons might also reflect the different osmo-sensitivity of the two proteins and their combined complex (figure 3.4). The known regulon of Rok so far mainly contains genes with a function in membrane maintenance and antimicrobial activity (15, 17, 90, 91). This suggests that the Rok regulon is not directly involved in the response to environmental cues, in contrast to for example the *proU* operon in *E. coli*, which is regulated by H-NS (62, 92). This might explain why Rok has not evolved a strong response to changes in physico-chemical conditions. We compared the previously determined Rok regulon (19) with the genes that change significantly upon salt shock

(93) and indeed very few genes overlap (figure S3.10B). Also the effects on transcription of these genes are either very minor or do not significantly change between the overexpression of Rok, sRok or Rok+sRok.

The stronger responsiveness of sRok to salt *in vitro* suggested that sRok regulates genes with a function in salt response and the formation of a complex between Rok and sRok, binding to (environmentally sensitive) operons not directly bound by Rok (or sRok) might be observed. It is a possibility that this happens at other osmo-sensitive genes or that indirect effects as described above might play a role. It makes it tempting to speculate that the habitat of the specific *Bacillus* species having both a *rok* and *srok* gene is different from that of other *Bacillus sp* containing only *rok*. It has been found that *B. licheniformis* and *B. sonorensis* have a slightly higher salt tolerance than *B. subtilis* (94), which might explain the need for a NAP (sRok) that is stimulated by higher salt concentrations.

It has been noted previously that two strains with both Rok and sRok on their chromosome showed a lack of competence development (24, 95). This might be due to *comK* repression by both Rok and sRok, however this was not visible in our data due to different growth conditions which do not induce *comK* expression. Further research into sRok's function *in vivo* is needed to see to what extent the direct regulon of sRok is different from that of Rok and what differences might be due to indirect effects via other regulators of transcription. Also, whether sRok is sufficiently transcribed to play a significant role in the respective *Bacillus* strains remains an open question.

Another transcriptional regulatory partner of Rok is DnaA. These two proteins cooperate in transcriptional repression of various genes (96), but it was shown that DnaA is not needed for chromosomal loop formation by Rok (23). Although it is unknown if DnaA affects the ability of Rok to bridge DNA, these observations strengthen the hypothesis that interaction with other proteins is one of the main ways to modulate Rok-mediated gene regulation. Future research is needed to determine if besides DnaA and sRok other proteins can influence the role of Rok in regulation. An example of a Rok antagonist is ComK that can relieve gene repression mediated by Rok at the *comK* promoter (97). There is no direct evidence that this involves physical interaction between the two proteins, but binding to DNA is not mutually exclusive. Activation of *comK* transcription is believed to be induced by alteration of local nucleoprotein structure, permitting transcription initiation (97). Based on the robustness of Rok in binding to DNA we expect the existence of similar antagonistic factors operating at other sites across the

chromosome. This bears similarity to regulation of transcription through local chromatin remodeling at complex promoters in *E. coli* (98). Studying the protein partners of Rok might shed light on how the modulation of Rok works in *B. subtilis* and provide new insights into gene regulation by NAPs in general.

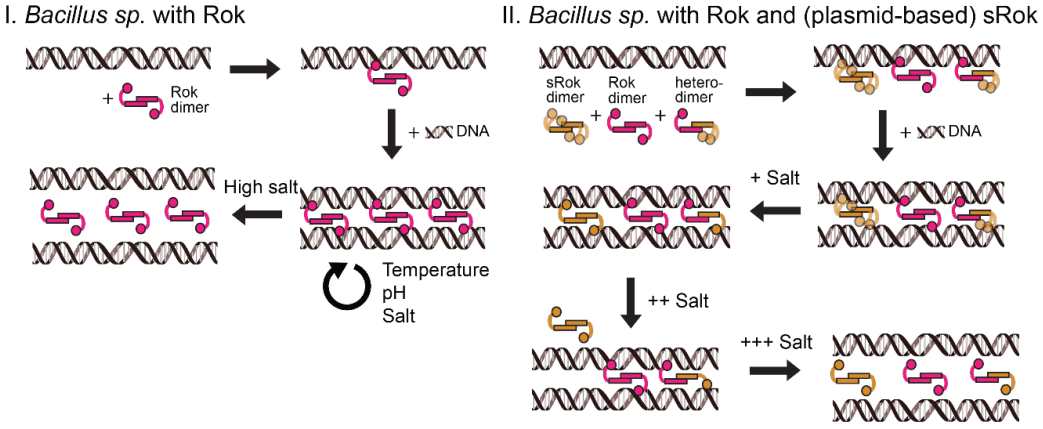


Figure 3.8 The proposed mechanisms of DNA bridging by (s)Rok. In *Bacillus sp.* with Rok on their genome (I), Rok binds DNA as a dimer without associating into nucleoprotein filaments. Dimeric Rok acts as bridging unit. Rok dimers cluster cooperatively in between two DNA duplexes, not due to dimer-dimer interactions, but due to high local DNA concentration which drives association and bridging by additional dimers. The DNA bridging activity of Rok is not sensitive to changes in physico-chemical conditions (temperature, pH and salt). When sRok is present, DNA bridging activity can be modulated by salt concentration. In *Bacillus sp.* where both Rok and sRok are present either on the genome or on a plasmid (II), a mix of homo- and heterodimers exist. Rok homodimers behave as in (I), while sRok homodimer DNA binding is first stimulated and then inhibited by salt. Heterodimers exhibit average behaviour of the two proteins.

References

1. Badrinarayanan,A., Le,T.B.K. and Laub,M.T. (2015) Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.*, **31**, 171–199.

2. Dorman,C.J. (2004) H-NS: A universal regulator for a dynamic genome. *Nat. Rev. Microbiol.*, **2**, 391–400.

3. Dame,R.T., Rashid,F.-Z.M. and Grainger,D.C. (2020) Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.*, **21**, 227–242.

4. Dame,R.T. (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.*, **56**, 858–870.

5. Luijsterburg,M.S., Noom,M.C., Wuite,G.J.L. and Dame,R.T. (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *J. Struct. Biol.*, **156**, 262–272.

6. Stuger,R., Woldringh,C.L., van der Weijden,C.C., Vischer,N.O., Bakker,B.M., van Spanning,R.J., Snoep,J.L. and Westerhoff,H. V (2002) DNA supercoiling by gyrase is linked to nucleoid compaction. *Mol Biol Rep*, **29**, 79–82.

7. Grainger,D.C. (2016) Structure and function of bacterial H-NS protein. *Biochem. Soc. Trans.*, **44**, 1561–1569.

8. Qin,L., Erkelens,A.M., Ben Bdira,F. and Dame,R.T. (2019) The architects of bacterial DNA bridges: A structurally and functionally conserved family of proteins. *Open Biol.*, **9**, 190223.

9. Gordon,B.R.G., Li,Y., Cote,A., Weirauch,M.T., Ding,P., Hughes,T.R., Navarre,W.W., Xia,B. and Liu,J. (2011) Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc. Natl. Acad. Sci.*, **108**, 10690–10695.

10. Dorman,C.J. (2014) H-NS-like nucleoid-associated proteins, mobile genetic elements and horizontal gene transfer in bacteria. *Plasmid*, **75**, 1–11.

11. Diggle,S.P., Winzer,K., Lazdunski,A., Williams,P. and Cámara,M. (2002) Advancing the quorum in Pseudomonas aeruginosa: MvaT and the regulation of N-acylhomoserine lactone production and virulence gene expression. *J. Bacteriol.*, **184**, 2576–86.

12. Stoebel,D.M., Free,A. and Dorman,C.J. (2008) Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology*, **154**, 2533–2545.

13. Bartek,I.L., Woolhiser,L.K., Baughn,A.D., Basaraba,R.J., Jacobs,W.R., Lenaerts,A.J. and Voskuil,M.I. (2014) Mycobacterium tuberculosis Lsr2 is a global transcriptional regulator required for adaptation to changing oxygen levels and virulence. *MBio*, **5**.

14. Hoa,T.T., Tortosa,P., Albano,M. and Dubnau,D. (2002) Rok (YkuW) regulates genetic competence in Bacillus subtilis by directly repressing comK. *Mol. Microbiol.*, **43**, 15–26.

15. Albano,M., Smits,W.K., Ho,L.T.Y., Kraigher,B., Mandic-Mulec,I., Kuipers,O.P. and Dubnau,D. (2005) The Rok protein of Bacillus subtilis represses genes for cell surface and extracellular functions. *J. Bacteriol.*, **187**, 2010–9.

16. Smits,W.K. and Grossman,A.D. (2010) The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in Bacillus subtilis. *PLoS Genet.*, **6**, 1001207.

17. Denham,E.L., Piersma,S., Rinket,M., Reilman,E., de Goffau,M.C. and van Dijl,J.M. (2019) Differential expression of a prophage-encoded glycocin and its immunity protein suggests a mutualistic strategy of a phage and its host. *Sci. Rep.*, **9**, 2845.

18. Serrano,E., Torres,R. and Alonso,J.C. (2021) Nucleoid-associated Rok differentially affects chromosomal transformation on Bacillus subtilis recombination-deficient cells. *Environ. Microbiol.*, **23**, 3318–3331.

19. Forrest,D., Warman,E.A., Erkelens,A.M., Dame,R.T. and Grainger,D.C. (2022) Xenogeneic silencing strategies in bacteria are dictated by RNA polymerase promiscuity. *Nat. Commun.*, **13**, 1149.

20. Marbouty,M., Le Gall,A., Cattoni,D.I., Cournac,A., Koh,A., Fiche,J.-B., Mozziconacci,J., Murray,H., Koszul,R. and Nollmann,M. (2015) Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Mol. Cell*, **59**, 588–602.

21. Hardy,C.D. and Cozzarelli,N.R. (2005) A genetic selection for supercoiling mutants of Escherichia coli reveals proteins implicated in chromosome structure. *Mol. Microbiol.*, **57**, 1636–52.

22. Noom,M.C., Navarre,W.W., Oshima,T., Wuite,G.J.L. and Dame,R.T. (2007) H-NS promotes looped domain formation in the bacterial chromosome. *Curr. Biol.*, **17**, R913-4.

23. Dugar,G., Hofmann,A., Heermann,D.W. and Hamoen,L.W. (2022) A chromosomal loop anchor mediates bacterial genome organization. *Nat. Genet. 2022 542*, **54**, 194–201.

24. Singh,P.K., Ramachandran,G., Durán-Alcalde,L., Alonso,C., Wu,L.J. and Meijer,W.J.J. (2012) Inhibition of Bacillus subtilis natural competence by a native, conjugative plasmid-encoded comK repressor protein. *Environ. Microbiol.*, **14**, 2812–2825.

25. Val-Calvo,J., Miguel-Arribas,A., Abia,D., Wu,L.J. and Meijer,W.J.J. (2021) pLS20 is the archetype of a new family of conjugative plasmids harboured by Bacillus species. *NAR Genomics Bioinforma.*, **3**.

26. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison 3rd,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343–345.

27. van der Valk,R.A., Vreede,J., Qin,L., Moolenaar,G.F., Hofmann,A., Goosen,N. and Dame,R.T. (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-Bridging activity. *Elife*, **6**, e27369.

28. Qin,L., Bdira,F. Ben, Sterckx,Y.G.J., Volkov,A.N., Vreede,J., Giachin,G., van Schaik,P., Ubbink,M. and Dame,R.T. (2020) Structural basis for osmotic regulation of the DNA binding properties of H-NS proteins. *Nucleic Acids Res.*, **48**, 2156–2172.

29. Wagner,K., Moolenaar,G.F. and Goosen,N. (2011) Role of the insertion domain and the zinc-finger motif of Escherichia coli UvrA in damage recognition and ATP hydrolysis. *DNA Repair (Amst).*, **10**, 483–496.

30. Henneman,B., Heinsman,J., Battjes,J. and Dame,R.T. (2018) Quantitation of DNA-binding affinity using tethered particle motion. In *Methods in Molecular Biology*. Humana Press Inc., Vol. 1837, pp. 257–275.

31. Laurens,N., Driessen,R.P.C., Heller,I., Vorselen,D., Noom,M.C., Hol,F.J.H., White,M.F., Dame,R.T. and Wuite,G.J.L. (2012) Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat. Commun.*, **3**, 2330.

32. van der Valk,R.A., Laurens,N. and Dame,R.T. (2017) Tethered particle motion analysis of the DNA binding properties of architectural proteins. In *Methods in Molecular Biology*. Humana Press Inc., Vol. 1624, pp. 127–143.

33. van der Valk,R.A., Qin,L., Moolenaar,G.F. and Dame,R.T. (2018) Quantitative determination of DNA bridging efficiency of chromatin proteins. In Dame R.T. (ed), *Methods in Molecular Biology*. Humana Press, New York, NY, Vol. 1837, pp. 199–209.

34. McGhee,J.D. and von Hippel,P.H. (1974) Theoretical aspects of DNA-protein interactions: Co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.*, **86**, 469–489.

35. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nat. 2021 5967873*, **596**, 583–589.

36. Evans,R., O'Neill,M., Pritzel,A., Antropova,N., Senior,A., Green,T., Žídek,A., Bates,R., Blackwell,S., Yim,J., *et al.* (2021) Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 10.1101/2021.10.04.463034.

37. Mirdita,M., Schütze,K., Moriwaki,Y., Heo,L., Ovchinnikov,S. and Steinegger,M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods 2022 196*, **19**, 679–682.

38. Steinegger,M. and Söding,J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol. 2017 3511*, **35**, 1026–1028.

39. Yu,D., Ellis,H.M., Lee,E.C., Jenkins,N.A., Copeland,N.G. and Court,D.L. (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 5978.

40. Bertani,G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J. Bacteriol.*, **62**, 293–300.

41. Singh,P.K., Ramachandran,G., Ramos-Ruiz,R., Peiró-Pastor,R., Abia,D., Wu,L.J. and Meijer,W.J.J. (2013) Mobility of the native Bacillus subtilis conjugative plasmid pLS20 is regulated by intercellular signaling. *PLoS Genet.*, **9**.

42. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods 2012 94*, **9**, 357–359.

43. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

44. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

45. Delignette-Muller,M.L. and Dutang,C. (2015) fitdistrplus: An R Package for Fitting Distributions. *J. Stat. Softw.*, **64**, 1–34.

46. De Jong,A., Kuipers,O.P. and Kok,J. (2022) FUNAGE-Pro: comprehensive web server for gene set enrichment analysis of prokaryotes. *Nucleic Acids Res.*, **50**, W330–W336.

47. Kanehisa,M., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.

48. Madeira,F., Park,Y.M., Lee,J., Buso,N., Gur,T., Madhusoodanan,N., Basutkar,P., Tivey,A.R.N., Potter,S.C., Finn,R.D., *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.*, **47**, W636–W641.

49. Wheeler,T.J., Clements,J. and Finn,R.D. (2014) Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics*, **15**, 7.

50. Bdira,F. Ben, Erkelens,A.M., Qin,L., Volkov,A.N., Lippa,A.M., Bowring,N., Boyle,A.L., Ubbink,M., Dove,S.L. and Dame,R.T. (2021) Novel anti-repression mechanism of H-NS proteins by a phage protein. *Nucleic Acids Res.*, **49**, 10770–10784.

51. Qu,Y., Lim,C.J., Whang,Y.R., Liu,J. and Yan,J. (2013) Mechanism of DNA organization by Mycobacterium tuberculosis protein Lsr2. *Nucleic Acids Res.*, **41**, 5263–5272.

52. Winardhi,R.S., Fu,W., Castang,S., Li,Y., Dove,S.L. and Yan,J. (2012) Higher order oligomerization is required for H-NS family member MvaT to form gene-silencing nucleoprotein filament. *Nucleic Acids Res.*, **40**, 8942–8952.

53. Liu,Y., Chen,H., Kenney,L.J. and Yan,J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev.*, **24**, 339–44.

54. Henneman,B., Brouwer,T.B., Erkelens,A.M., Kuijntjes,G.-J., van Emmerik,C., van der Valk,R.A., Timmer,M., Kirolos,N.C.S., van Ingen,H., van Noort,J., *et al.* (2021) Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res.*, **49**, 4338–4349.

55. Driessen,R.P., Sitters,G., Laurens,N., Moolenaar,G.F., Wuite,G.J., Goosen,N. and Dame,R.T. (2014) Effect of temperature on the intrinsic flexibility of DNA and its interaction with architectural proteins. *Biochemistry*, **53**, 6430–6438.

56. Driessen,R.P.C., Sitters,G., Laurens,N., Moolenaar,G.F., Wuite,G.J.L., Goosen,N. and Dame,R.T. (2014) Effect of temperature on the intrinsic flexibility of DNA and its interaction with architectural proteins. *Biochemistry*, **53**, 6430–6438.

57. Dame,R.T. and Wuite,G.J.L. (2003) On the Role of H-NS in the Organization of Bacterial Chromatin: From Bulk to Single Molecules and Back... *Biophys. J.*, **85**, 4146–4148.

58. Boudreau,B.A., Hron,D.R., Qin,L., van der Valk,R.A., Kotlajich,M. V, Dame,R.T. and

Landick,R. (2018) StpA and Hha stimulate pausing by RNA polymerase by promoting DNA–DNA bridging of H-NS filaments. *Nucleic Acids Res.*, **46**, 5525–5546.

59. Liu,Y., Chen,H., Kenney,L.J. and Yan,J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev.*, **24**, 339–344.

60. Ono,S., Goldberg,M.D., Olsson,T., Esposito,D., Hinton,J.C.D. and Ladbury,J.E. (2005) H-NS is a part of a thermally controlled mechanism for bacterial gene regulation. *Biochem. J.*, **391**, 203–13.

61. White-Ziegler,C.A. and Davis,T.R. (2009) Genome-wide identification of H-NS-controlled, temperature-regulated genes in Escherichia coli K-12. *J. Bacteriol.*, **191**, 1106–10.

62. Ueguchi,C. and Mizuno,T. (1993) The Escherichia coli nucleoid protein H-NS functions directly as a transcriptional repressor. *EMBO J.*, **12**, 1039.

63. Kotlajich,M. V, Hron,D.R., Boudreau,B.A., Sun,Z., Lyubchenko,Y.L. and Landick,R. (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife*, **2015**, 4970.

64. Hoper,D., Bernhardt,J. and Hecker,M. (2006) Salt stress adaptation of Bacillus subtilis: a physiological proteomics approach. *Proteomics*, **6**, 1550–1562.

65. Wendel,B.M., Pi,H., Krüger,L., Herzberg,C., Stülke,J. and Helmann,J.D. (2022) A Central Role for Magnesium Homeostasis during Adaptation to Osmotic Stress. *MBio*, **13**.

66. Ryan Will,W., Whitham,P.J., Reid,P.J. and Fang,F.C. (2018) Modulation of H-NS transcriptional silencing by magnesium. *Nucleic Acids Res.*, **46**, 5717–5725.

67. Duan,B., Ding,P., Hughes,T.R., Navarre,W.W., Liu,J. and Xia,B. (2018) How bacterial xenogeneic silencer rok distinguishes foreign from self DNA in its resident genome. *Nucleic Acids Res.*, **46**, 10514–10529.

68. Szatmári,D., Sárkány,P., Kocsis,B., Nagy,T., Miseta,A., Barkó,S., Longauer,B., Robinson,R.C. and Nyitrai,M. (2020) Intracellular ion concentrations and cation-dependent remodelling of bacterial MreB assemblies. *Sci. Reports 2020 101*, **10**, 1–13.

69. Sabari,B.R., Dall'Agnese,A., Boija,A., Klein,I.A., Coffey,E.L., Shrinivas,K., Abraham,B.J., Hannett,N.M., Zamudio,A. V., Manteiga,J.C., *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science (80-.).*, **361**.

70. Düster,R., Kaltheuner,I.H., Schmitz,M. and Geyer,M. (2021) 1,6-Hexanediol, commonly used to dissolve liquid–liquid phase separated condensates, directly impairs kinase and phosphatase activities. *J. Biol. Chem.*, **296**, 100260–100261.

71. Elbaum-Garfinkle,S. (2019) Matter over mind: Liquid phase separation and neurodegeneration. *J. Biol. Chem.*, **294**, 7160–7168.

72. Shahul Hameed,U.F., Liao,C., Radhakrishnan,A.K., Huser,F., Aljedani,S.S., Zhao,X., Momin,A.A., Melo,F.A., Guo,X., Brooks,C., *et al.* (2018) H-NS uses an autoinhibitory conformational switch for environment-controlled gene silencing. *Nucleic Acids Res.*, **1**, 23955–6900.

73. Tendeng,C., Soutourina,O.A., Danchin,A. and Bertin,P.N. (2003) MvaT proteins in Pseudomonas spp.: A novel class of H-NS-like proteins. *Microbiology*, **149**, 3047–3050.

74. Gordon,B.R.G., Imperial,R., Wang,L., Navarre,W.W. and Liu,J. (2008) Lsr2 of Mycobacterium represents a novel class of H-NS-like proteins. *J. Bacteriol.*, **190**, 7052–7059.

75. Dole,S., Nagarajavel,V. and Schnetz,K. (2004) The histone-like nucleoid structuring protein H-NS represses the Escherichia coli bgl operon downstream of the promoter. *Mol. Microbiol.*, **52**, 589–600.

76. Kharat,A.S. and Mahadevan,S. (2000) Analysis of the β-glucoside utilization (bgl) genes of Shigella sonnei: Evolutionary implications for their maintenance in a cryptic state. *Microbiology*, **146**, 2039–2049.

77. Schnetz,K., Toloczyki,C. and Rak,B. (1987) Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two Bacillus subtilis genes. *J. Bacteriol.*, **169**, 2579–2590.

78. Sankar,T.S., Neelakanta,G., Sangal,V., Plum,G., Achtman,M. and Schnetz,K. (2009) Fate of the H-NS–Repressed bgl Operon in Evolution of Escherichia coli. *PLoS Genet.*, **5**, e1000405.

79. May,J.J., Wendrich,T.M. and Marahiel,M.A. (2001) The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J. Biol. Chem.*, **276**, 7209–7217.

80. Kovács,M., Halfmann,A., Fedtke,I., Heintz,M., Peschel,A., Vollmer,W., Hakenbeck,R.

and Brückner,R. (2006) A functional dlt operon, encoding proteins required for incorporation of d-alanine in teichoic acids in gram-positive bacteria, confers resistance to cationic antimicrobial peptides in Streptococcus pneumoniae. *J. Bacteriol.*, **188**, 5797–5805.

81. Wang,W. and Sun,M. (2009) Phylogenetic relationships between Bacillus species and related genera inferred from 16S rDNA sequences. *Brazilian J. Microbiol.*, **40**, 505–521.

82. Tetz,V. and Tetz,G. (2017) Draft Genome Sequence of a Strain of Bacillus intestinalis sp. nov., a New Member of Sporobiota Isolated from the Small Intestine of a Single Patient with Intestinal Cancer. *Genome Announc.*, **5**.

83. Lim,C.J., Lee,S.Y., Kenney,L.J. and Yan,J. (2012) Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. *Sci Rep*, **2**, 509.

84. Winardhi,R.S., Yan,J. and Kenney,L.J. (2015) H-NS Regulates Gene Expression and Compacts the Nucleoid: Insights from Single-Molecule Experiments. *Biophys. J.*, **109**, 1321–1329.

85. Van Noort,J., Verbrugge,S., Goosen,N., Dekker,C. and Dame,R.T. (2004) Dual architectural roles of HU: Formation of flexible hinges and rigid filaments. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 6969–6974.

86. Summers,E.L., Meindl,K., Usó,I., Mitra,A.K., Radjainia,M., Colangeli,R., Alland,D. and Arcus,V.L. (2012) The Structure of the Oligomerization Domain of Lsr2 from Mycobacterium tuberculosis Reveals a Mechanism for Chromosome Organization and Protection. *PLoS One*, **7**, 38542.

87. Arold,S.T., Leonard,P.G., Parkinson,G.N. and Ladbury,J.E. (2010) H-NS forms a superhelical protein scaffold for DNA condensation. *Proc. Natl. Acad. Sci.*, **107**, 15728–15732.

88. Suzuki-Minakuchi,C., Kawazuma,K., Matsuzawa,J., Vasileva,D., Fujimoto,Z., Terada,T., Okada,K. and Nojiri,H. (2016) Structural similarities and differences in H-NS family proteins revealed by the N-terminal structure of TurB in Pseudomonas putida KT2440. *FEBS Lett.*, **590**, 3583–3594.

89. Ding,P., McFarland,K.A., Jin,S., Tong,G., Duan,B., Yang,A., Hughes,T.R., Liu,J., Dove,S.L., Navarre,W.W., *et al.* (2015) A Novel AT-Rich DNA Recognition Mechanism for Bacterial Xenogeneic Silencer MvaT. *PLoS Pathog.*, **11**, 1004967.

90. Kovács,Á.T. and Kuipers,O.P. (2011) Rok regulates yuaB expression during

architecturally complex colony development of Bacillus subtilis 168. *J. Bacteriol.*, **193**, 998–1002.

91. Zhu,B. and Stülke,J. (2018) SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. *Nucleic Acids Res.*, **46**, D743–D748.

92. Nagarajavel,V., Madhusudan,S., Dole,S., Rahmouni,A.R. and Schnetz,K. (2007) Repression by binding of H-NS within the transcription unit. *J. Biol. Chem.*, **282**, 23622–23630.

93. Hahne,H., Mäder,U., Otto,A., Bonn,F., Steil,L., Bremer,E., Hecker,M. and Becher,D. (2010) A comprehensive proteomics and transcriptomics analysis of Bacillus subtilis salt stress adaptation. *J. Bacteriol.*, **192**, 870–882.

94. Palmisano,M.M., Nakamura,L.K., Duncan,K.E., Istock,C.A. and Cohan,F.M. (2001) Bacillus sonorensis sp. nov., a close relative of Bacillus licheniformis, isolated from soil in the Sonoran Desert Arizona. *Int. J. Syst. Evol. Microbiol.*, **51**, 1671–1679.

95. Duitman,E.H., Wyczawski,D., Boven,L.G., Venema,G., Kuipers,O.P. and Hamoen,L.W. (2007) Novel methods for genetic transformation of natural Bacillus subtilis isolates used to study the regulation of the mycosubtilin and surfactin synthetases. *Appl. Environ. Microbiol.*, **73**, 3490–3496.

96. Seid,C.A., Smith,J.L. and Grossman,A.D. (2017) Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in Bacillus subtilis. *Mol. Microbiol.*, **103**, 798–817.

97. Smits,W.K., Hoa,T.T., Hamoen,L.W., Kuipers,O.P. and Dubnau,D. (2007) Antirepression as a second mechanism of transcriptional activation by a minor groove binding protein. *Mol. Microbiol.*, **64**, 368–381.

98. Browning,D.F., Cole,J.A. and Busby,S.J.W. (2000) Suppression of FNR-dependent transcription activation at the Escherichia coli nir promoter by Fis, IHF and H-NS: Modulation of transcription by a complex nucleo-protein assembly. *Mol. Microbiol.*, **37**, 1258–1269.

99. Marciniak,B.C., Trip,H., Fusetti,F. and Kuipers,O.P. (2012) Regulation of ykrL (htpX) by Rok and YkrK, a novel type of regulator in Bacillus subtilis. *J. Bacteriol.*, **194**, 2837–2845.

Supplementary figures

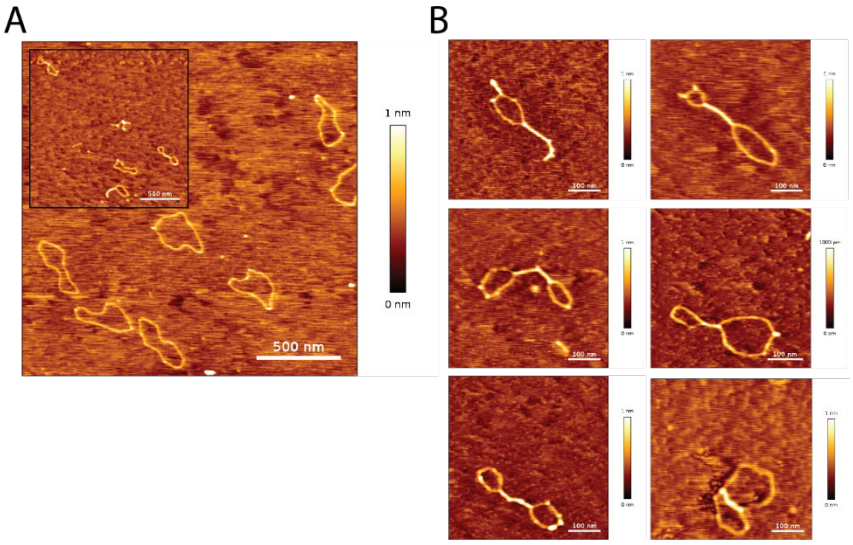


Figure S3.1 Rok bridges DNA as observed by AFM imaging A) Nicked pUC19 molecules incubated without Rok or with Rok at a concentration of 200 nM (insert). B) Close-ups of representative Rok-DNA complexes.

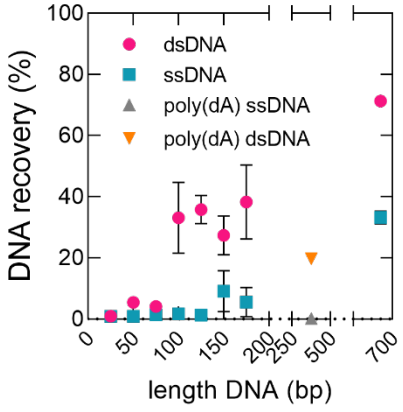


Figure S3.2 Rok bridges dsDNA but not ssDNA DNA recovery (%) as a function of the length of prey DNA in bp in the presence of 0.27 μM Rok with 50 mM KCl at 25°C. The bait DNA was the 685 bp dsDNA as used for the other bridging experiments. Data are plotted as mean values and the error bars represent the standard deviation from three independent measurements.

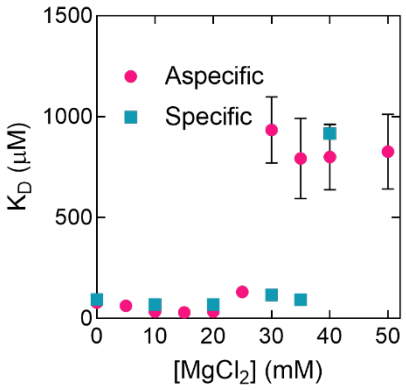


Figure S3.3 $MgCl_2$ does not affect the DNA binding affinity of Rok Kd-values (μM) obtained from fitting protein titration data to the McGhee-von Hippel equation as a function of $MgCl_2$ concentration. The DNA used was 78 bp in length with or without a specific Rok binding site and the Rok concentration was varied between 0.125 and 16 μM. The final measurement buffer consisted of 10 mM Tris HCl pH 8, 150 mM KCl, 5% glycerol, 0.05% Tween20 and 0.08 mg/ml acetylated BSA. $MgCl_2$ was added accordingly. Each data point was measured at least in triplicate and error bars represent the standard deviation. Some error bars are hidden behind the data points due to their small size.

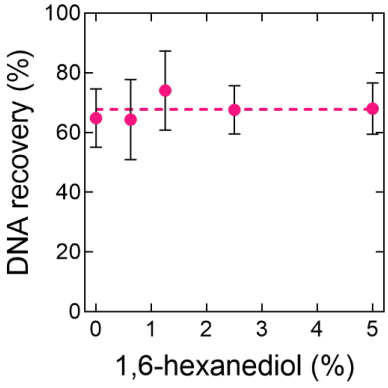


Figure S3.4 1,6-hexanediol cannot prevent Rok-DNA bridges being formed DNA recovery (%) as a function of added 1,6-hexanediol (%) in the presence of 0.27 μM Rok with 50 mM KCl at 25°C. Data are plotted as mean values and the error bars represent the standard deviation from three independent measurements. Dashed line serves as line to guide the eye.

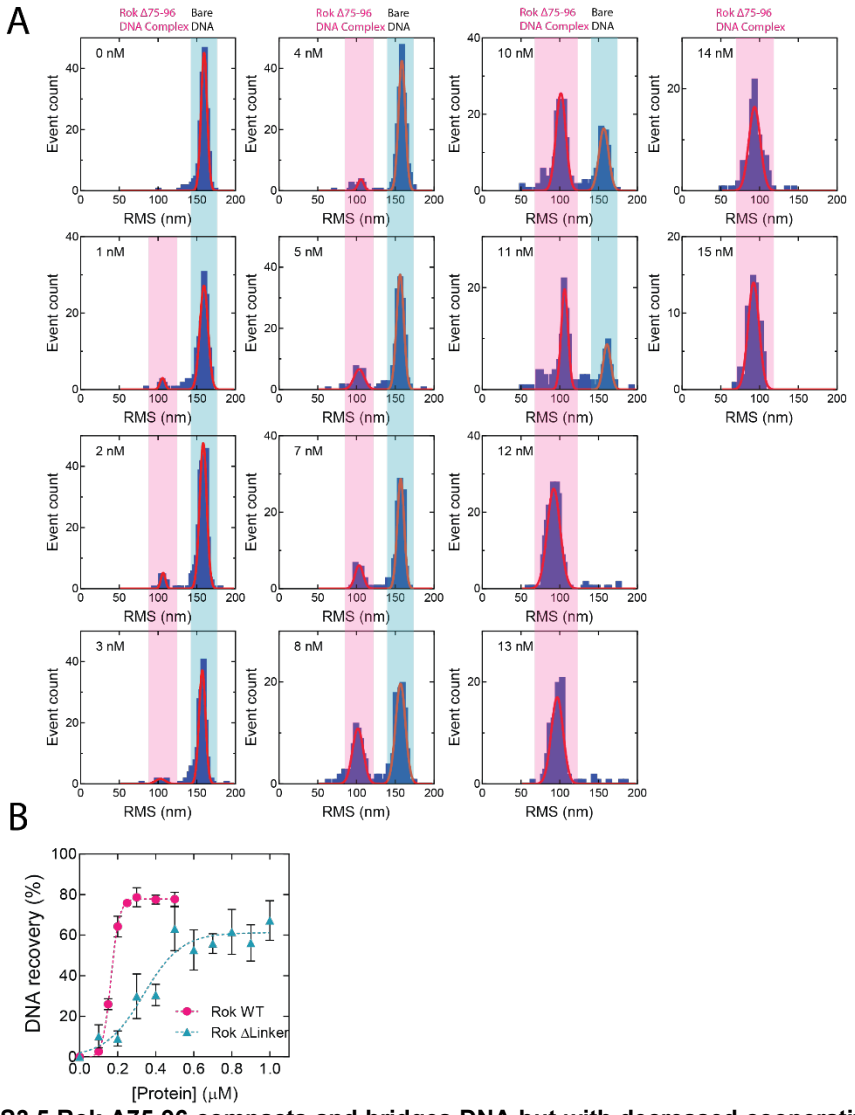


Figure S3.5 Rok Δ75-96 compacts and bridges DNA but with decreased cooperativity. A) Histograms of Root Mean Square displacement (RMS) obtained for 32%GC DNA as a function of Rok Δ75-96 at concentrations of 0, 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14 and 15 nM as measured by TPM in the presence of 50 mM KCl. The histograms were fitted to Gaussian distributions, in which the RMS value at ~150 nm represents bare DNA and the population with an RMS at ~100 nm represents DNA bound by Rok Δ75-96. The bare DNA and Rok-DNA complex populations are highlighted with a light blue and magenta box, respectively. The data for each concentration originates from at least two independent measurements. B) DNA recovery (%) as a function of protein concentration in μM in the presence of 50 mM KCl at 25°C. For reference, Rok WT is shown (reproduced from figure

3.1A). Data are plotted as mean values and the error bars represent the standard deviation from three independent measurements. Dashed lines serve as lines to guide the eye.

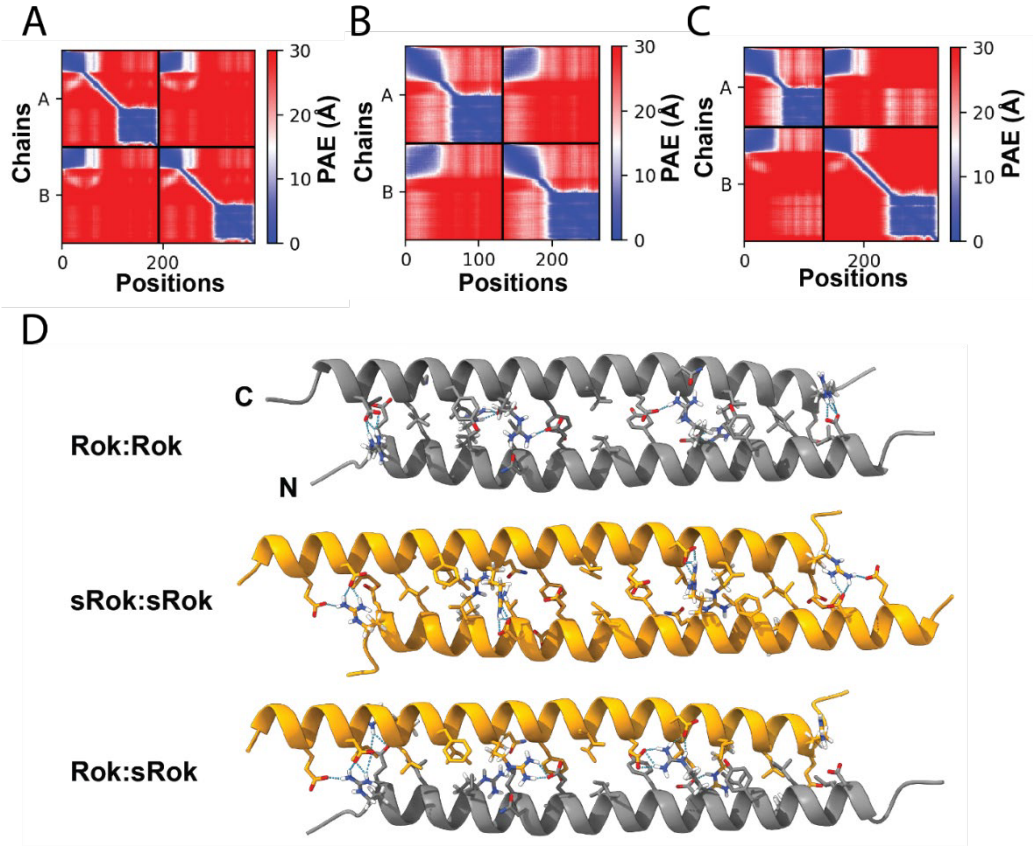


Figure S3.6 Alphafold2 predicts (s)Rok homo- and heterodimers with high confidence A) Predicted aligned error (PAE) plot of the Rok homodimer, B) sRok homodimer and C) Rok:sRok heterodimer structure. D) The dimerization domains of homodimers Rok (grey, amino acids 1 to 46) and sRok (orange, amino acids 1 to 48) and heterodimer Rok:sRok as predicted by AlphaFold2. Amino acids that are located within the coiled-coil dimerization interface are drawn as sticks.

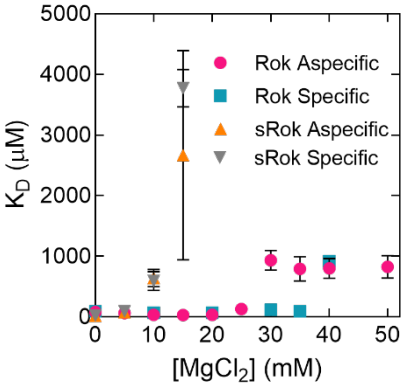


Figure S3.7 $MgCl_2$ does affect the DNA binding affinity of sRok Kd-values (µM) obtained from fitting protein titration data to the McGhee-von Hippel equation as a function of the $MgCl_2$ concentration. The DNA used was 78 bp in length with or without a specific Rok binding site and the sRok concentration was varied between 0.125 and 16 µM. The Rok data was reproduced from figure S3.2 for comparison. The final measurement buffer consisted of 10 mM Tris HCl pH 8, 150 mM KCl, 5% glycerol, 0.05% Tween20 and 0.08 mg/ml acetylated BSA. $MgCl_2$ was added accordingly. Each data point was measured at least in triplicate and error bars represent the standard deviation. Some error bars are hidden behind the data points due to their small size.

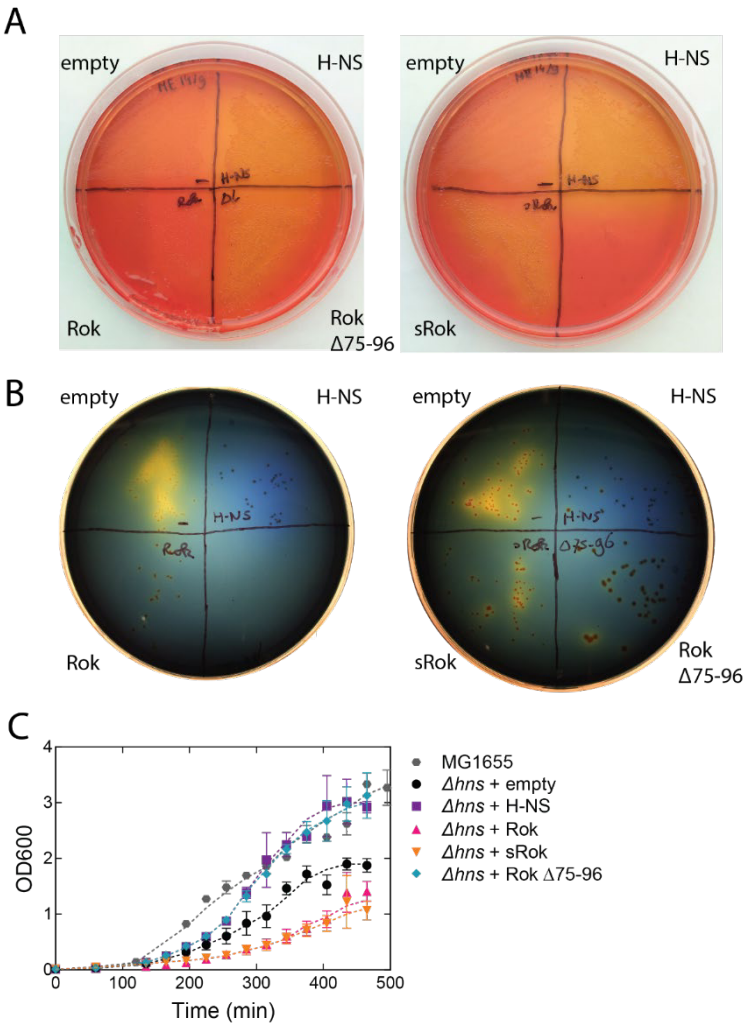


Figure S3.8. Rok and sRok cannot complement a *Δhns* phenotype, but Rok Δ75-96 can A) Complementation of *bgl* operon repression tested on MacConkey agar plates supplemented with 0.4% salicin. *E. coli* NT135 cells were transformed with either an empty pUC19 plasmid or with an insert containing the *hns* promoter followed by the *hns, rok, srok* or *rokΔ75-96* coding sequence. B) Complementation of *bgl* operon repression tested on BTB indicator plates supplemented with 0.5% salicin, using the same cells as in A. C) Growth curves in LB medium over time. Data are plotted as mean values and the error bars represent the standard deviation of three independent growth curves. Dashed lines serve as lines to guide the eye.

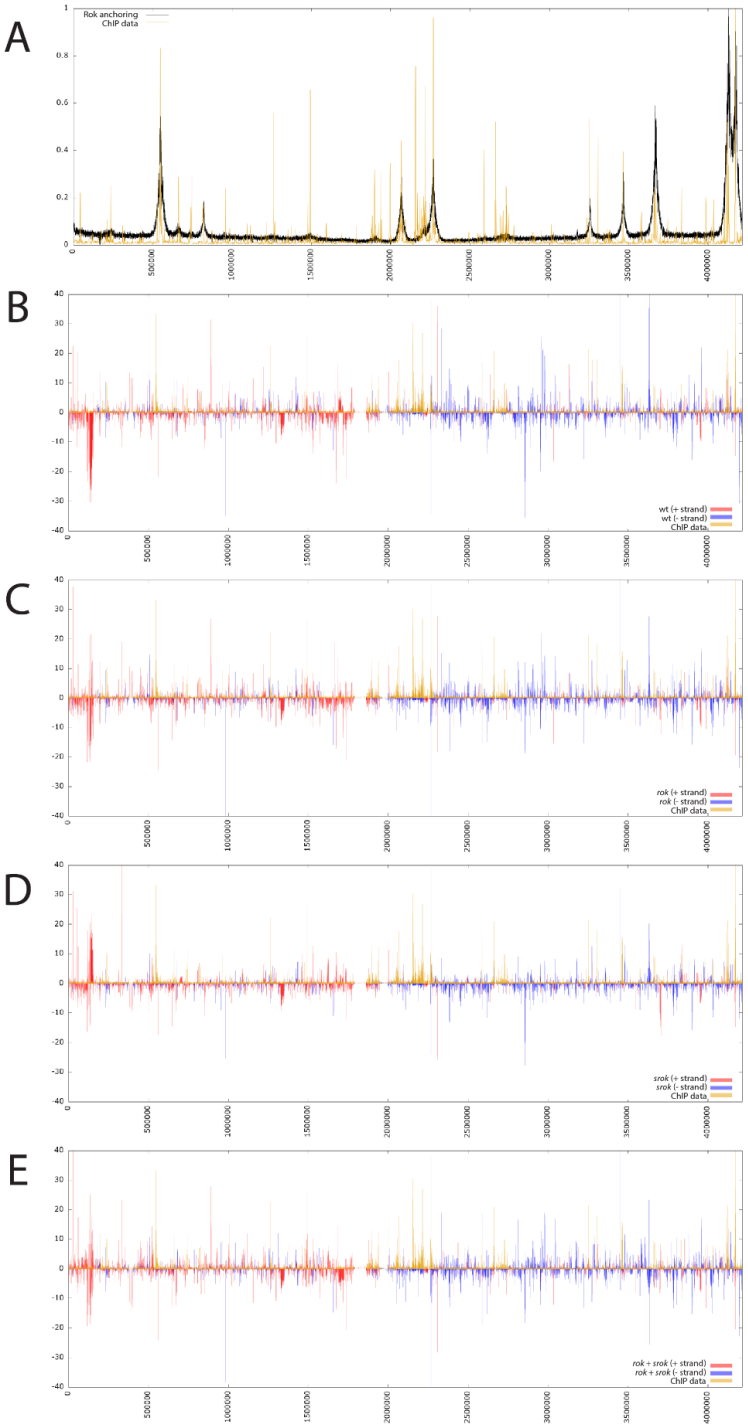

A
Rok anchoring
ChIP data
B
wt (+ strand)
wt (- strand)
ChIP data
C
rok (+ strand)
rok (- strand)
ChIP data
D
srok (+ strand)
srok (- strand)
ChIP data
E
rok + srok (+ strand)
rok + srok (- strand)
ChIP data

Figure S3.9. Rok, sRok and Rok+sRok affect multiple loci across the *B. subtilis* genome. A) Overlay of previously published ChIP data (16) and DNA coverage data (23) for Rok. B-E) Overlay of the expression levels on the + strand, - strand for wt (B), Rok (C), sRok (D) and Rok+sRok (E). For reference, the ChIP data for Rok was included (16).

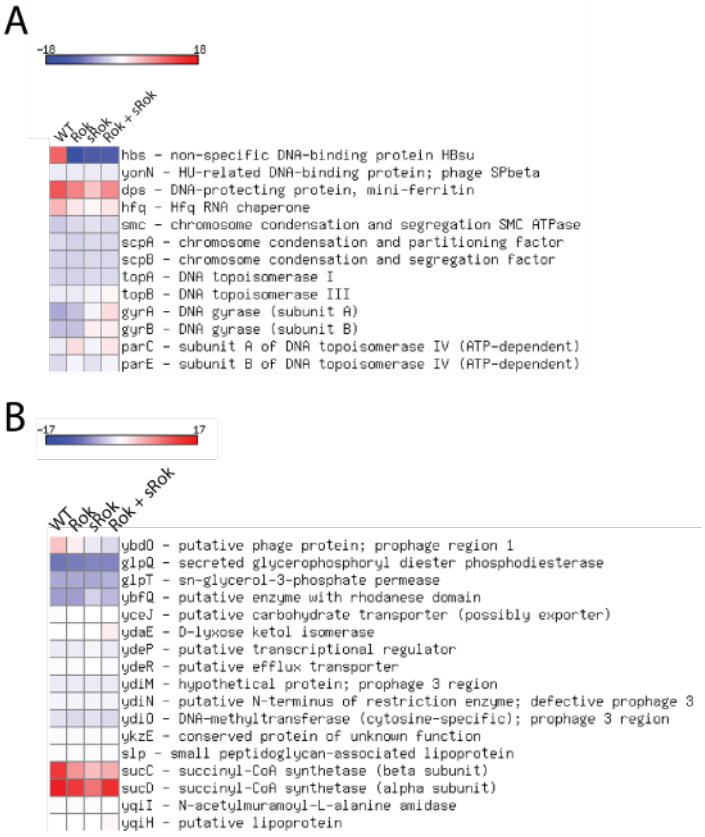


Figure S3.10. Effect of Rok, sRok and Rok+sRok on osmo-regulated genes and proteins involved in chromatin organization. Heat map representations of a set of proteins involved in chromatin organization (A) or osmo-regulated genes (B) in response to *rok*, *srok* and *rok+srok* expressed from an ectopic promoter.

Table S3.1. Oligonucleotides used for cloning and mutagenesis. The name and sequence of the primer are given and the plasmid(s) in which construction they were used are indicated.

Name primer	Resulting plasmid	Sequence (5'-3')
Rok_fragment_F	pRD231/pRD415	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTTTAATGAAAGAGAAGCTTTGCGC
Rok_fragment_R	pRD231/pRD415	CCTTTCGGGCTTTGTTAGCAGTTATTCGTTTGCTGATTCTGCAGATTCGA
Rok_vector_F	pRD231/pRD415	GCGCAAAGCTTCTCTTTCATTAAACATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC
Rok_vector_R	pRD231/pRD415	TCGAATCTGCAGAATCAGCAAACGAATAACTGCTAACAAAGCCCGAAAGG
Rok_delta_linker F	pRD415	CGGCTCAGGAAATCAGAGCCGGTATACCAGACGG
Rok_delta_linker R	pRD415	CCGTCTGGTATACCGGCTCTGATTTCCTGAGCCG
sRok_pET30b F	pRD411	AACTTTAAGAAGGAGATATACATATGCTGACCGAACGTCAGGC
sRok_pET30b R	pRD411	CCTTTCGGGCTTTGTTAGCAGTTACATTTCTTCAAAGCGAT
sRok_pET30b_vector F	pRD411	ATCGCTTTGAAGAAATGTAACTGCTAACAAAGCCCGAAAGG
sRok_pET30b_vector R	pRD411	GCCTGACGTTCGGTCAGCATATGTATATCTCCTTCTTAAAGTT
ivc_H-NS_pUC19_F	pRD408	GAGCTCGGTACCCGGGGATCTTCTGGCTAATTTTATGAAA
ivc_H-NS_pUC19_R	pRD408	CCTGCAGGTCGACTCTAGAGCAAGTGCAATCTACAAAAGA
ivc_pUC19_H-NS_F	pRD408	TTTCATAAAATTAGCCAGAAGATCCCCGGGTACCGAGCTC
ivc_pUC19_H-NS_R	pRD408	TCTTTTGTAGATTGCACTTGCTCTAGAGTCGACCTGCAGG
sRok_pUC19 F	pRD410	GATAGGGGGTACATTGAGGAATGCTGACCGAACGTCAGGC
sRok_pUC19 R	pRD410	CAAGCAGTTTTTCTTTATATTTACATTTCTTCAAAGCGAT

sRok_pUC19_vector F	pRD410	ATCGCTTTGAAGAAATGTAAATATAAAGAAAAA CTGCTTG
sRok_pUC19_vector R	pRD410	GCCTGACGTTCGGTCAGCATTCCTCAATGTAC CCCCTATC
Rok_pH-NS F	pRD424/pRD412	ATAAGTTTGAGATTACTACAATGTTTAATGAAA GAGAAGC
Rok_pH-NS R	pRD424/pRD412	CAAGTGCAATCTACAAAAGATTATTCGTTTGCT GATTCTG
pH-NS_Rok F	pRD424/pRD412	GCTTCTCTTTCATTAAACATTGTAGTAATCTCA AACTTAT
pH-NS_Rok R	pRD424/pRD412	CAGAATCAGCAAACGAATAATCTTTTGTAGATT GCACTTG
Forward pET30b	pRD461	CCGCTATCGCTACGTGACTGGGTCATGGCTG CGCCCCGACACCCG
Reverse pET30b	pRD461	AGCCATGACCCAGTCACGTAGCGATAGCGGA GTGTATACTGGCTT
Rok N-term F	pRD461	CATCATCATCATCATCATtaaCTGCTAACAAAGC CCGAAAG
Rok 46-191 his R	pRD461	TTAATGATGATGATGATGATGTTCGTTTGCTGA TTCTGCAGATTCGA
oAND314	pAND520	TTTTaagcttAAAGGAGAGATATAAATGCTTACAG AAAGAC
oAND315	pAND520	TTTTgtcgacGTGTTACATTTCTTCGAATCTATAA TAACC
oAND316	pAND521/pAND522	TTTTgtcgacAAAGGAGAGATATAAATGTTTAATG AAAGAG
oAND317	pAND521/pAND522	TTTTgctagcGAAAAAGAAAACAAACCTTCACAG AAAAAACCTC

Table S3.2. Plasmids created for this study. For each plasmid, the backbone, insert, antibiotics resistance are given. All plasmids are deposited to Addgene and the respective identification numbers are given.

Name	Backbone	Insert	Resistance	Addgene number
pRD231	pET30b	Rok	Kanamycin	178195
pRD411	pET30b	sRok	Kanamycin	178196
pRD415	pET30b	Rok Δ75-96	Kanamycin	178197
pRD461	pET30b	Rok 6xhis	Kanamycin	178198
pRD408	pUC19	p*hns* + H-NS	Ampicillin	178199
pRD410	pUC19	p*hns* + sRok	Ampicillin	178200
pRD412	pUC19	p*hns* + Rok Δ75-96	Ampicillin	178201
pRD424	pUC19	p*hns* + Rok	Ampicillin	178202

Table S3.3. Oligonucleotides used for single-stranded DNA substrates. Highlighted sequences were found to be favorable for Rok binding by Duan et al. 2018 (67).

Length	Sequence
25 bp	GGAGTAGTATGGTAATAACTATTTT
50 bp	GCAAATATATAATGTATAAGTTCTGGAGTAGTATGGTAATAACTATTTTA
75 bp	CTTAGTGGCAAATATATAATGTATAAGTTCTGGAGTAGTATGGTAATAACTATTTTATTT TCGATAGCTTGAATG
100 bp	GATATATGGACTCTTAGTGGCAAATATATAATGTATAAGTTCTGGAGTAGTATGGTAAT AACTATTTTATTTTCGATAGCTTGAATGTTTATTTTCCCAG
125 bp	ATTTAGTCTATTCGATATATGGACTCTTAGTGGCAAATATATAATGTATAAGTTCTGGAG TAGTATGGTAATAACTATTTTATTTTCGATAGCTTGAATGTTTATTTTCCCAGAGACATT AGGTG
150 bp	TTTCTCCAATTGATTTAGTCTATTCGATATATGGACTCTTAGTGGCAAATATATAATGTA TAAGTTCTGGAGTAGTATGGTAATAACTATTTTATTTTCGATAGCTTGAATGTTTATTTT CCCAGAGACATTAGGTGGTCTTTCAAACTC
175 bp	TCTCTTCAAATGTTTTCTCCAATTGATTTAGTCTATTCGATATATGGACTCTTAGTGGCA AATATATAATGTATAAGTTCTGGAGTAGTATGGTAATAACTATTTTATTTTCGATAGCTT GAATGTTTATTTTCCCAGAGACATTAGGTGGTCTTTCAAACTCAATAATGTGGGT

Table S3.4. Oligonucleotides used for MST DNA substrates. Highlighted sequence was found to be favorable for Rok binding by Duan et al. 2018 (67).

Aspecific top strand	Cy5- CGGCGCAAATTCGTGACCAGTTGCATCAGCTGCGTGAGCTGTTTATCGCAGC ATCGTAACAGGATAGTGAAGAAGACT
Aspecific bottom strand	AGTCTTCTTCACTATCCTGTTACGATGCTGCGATAAACAGCTCACGCAGCTGA TGCAACTGGTCACGAATTTGCGCCG
Specific top strand	Cy5- CGGCGCAAATTCGTGACCAGTTGCATCAGCTACTAGAGCTGTTTATCGCAGC ATCGTAACAGGATAGTGAAGAAGACT
Specific bottom strand	AGTCTTCTTCACTATCCTGTTACGATGCTGCGATAAACAGCTCTAGTAGCTGA TGCAACTGGTCACGAATTTGCGCCG

Table S3.5. Effects of ectopic expression of Rok as fold change of selected set of genes reported before to be regulated by Rok. The differences in expression between each of the strains and the *Δrok* strain are calculated by taking the square root of the difference between the normalized mean counts of each of the strains with respect to those of the *Δrok* strain, for each of the genes.

operon	Gene(s)	Reference	reported effect Rok	These studies (differential expression with respect to Δ*rok* strain)			
				wt strain	ectopic expression		
					Rok	sRok	sRok+Rok
	*htp*X	(99)	Min	-5.38	-6.05	-5.20	4.43
	*com*K	(14)	Min	-1.24	-1.92	-1.82	-1.92
sdp	*sdp*A	(15)	Min	1.29	-1.05	-0.92	-1.05
	*sdp*B		Min	1.27	-0.90	-0.72	-0.91
	*sdp*C		min	4.61	-3.23	-1.38	-3.39
alb	*sboA*		Min	2.76	-10.51	13.36	-10.13
	sboX		Min	-2.06	-4.96	6.28	-4.89
	albA		Min	-6.41	-7.29	12.44	-7.16
	albB		Min	-6.32	-7.56	11.28	-7.27
	albC		Min	-5.01	-5.72	8.78	-5.53
	albD		Min	-2.24	-2.56	2.83	-2.55

	albE	(15)	Min	-2.94	-3.31	4.55	-3.27
	albF		Min	-2.66	-2.98	4.02	-2.92
	albG		Min	-1.97	-2.37	-3.05	-2.29
	rok	(14)	Min	5.33	7.5	-0.18	11.85
ybb	*yybN*	(15, 23)	Min	-14.70	-19.36	-14.76	-20.37
	yybM		Min	-2.96	-3.91	-3.22	-4.03
	yybL		Min	-2.52	-2.92	-2.64	-2.94
	yybK		Min	-2.93	-3.23	-2.84	-3.26
	yybJ		Min	-4.08	-4.78	-3.70	-4.85
sun	*sunA*	(15, 17)	Min	-34.10	-82.33	-24.08	-82.88
	sunT		Min	-7.43	-8.45	-7.16	-8.47
	bdbA		Min	-6.12	-8.16	-6.58	-8.18
	sunS		Min	-5.35	-8.39	-6.53	-8.43
	bdbB		Min	-5.36	-9.24	-6.98	-9.27
epe	*epeX*	(15)	Min	2.03	1.10	-0.77	1.07
	epeE		Min	0.13	-0.6	-0.57	-0.62
	epeP		Min	-0.37	-1.00	-0.75	-1.05
	epeA		Min	-2.01	-2.92	-1.91	-3.00
	epeB		Min	-2.27	-3.23	-2.13	-3.31
	yxaJ	(15)	Min	-4.79	-5.25	-3.63	-5.29
	yxaL	(15)	Min	-3.96	-9.34	7.38	-10.46
	yjcN	(15)	Min	-5.20	-6.54	-5.13	-6.73
	yuaB (=*bsl*A) (indirect)	(90)	min	1.76	4.31	2.98	7.41

Chapter 4

Specific DNA binding of archaeal histones HMfA and HMfB

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

Abstract

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

Introduction

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

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

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

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

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

Materials and Methods

Protein expression and purification

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

DNA substrate preparation

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

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

Tethered particle motion

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

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

Microscale thermophoresis

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

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

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

Results

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

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

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

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

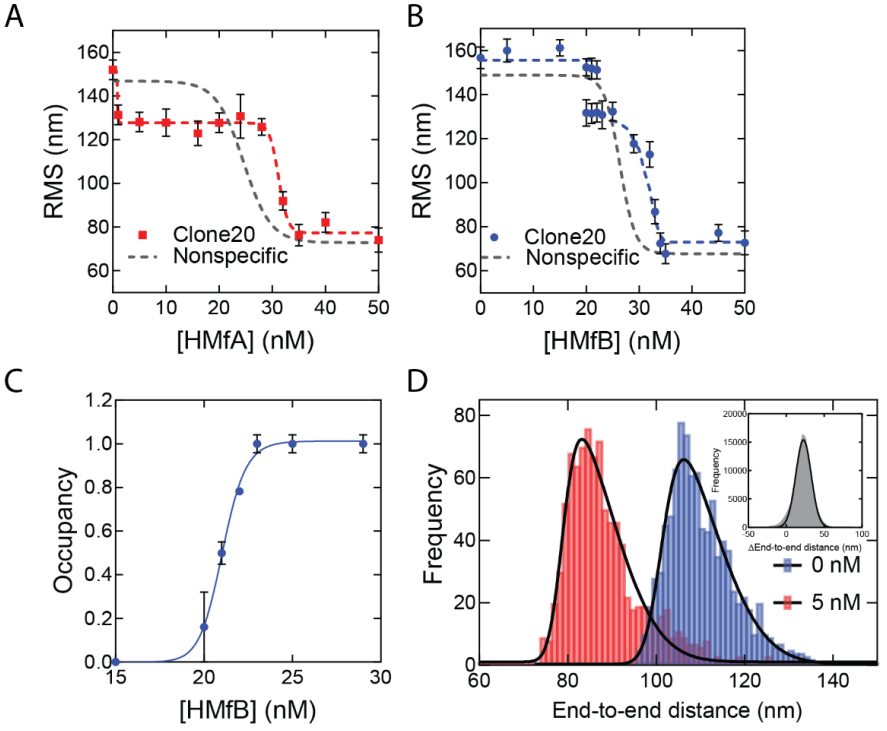


Figure 4.1 HMfA and HMfB bind as tetramers to the Clone20 site preceding hypernucleosome formation A) Root mean square displacement (RMS) values of Nonspecific and Clone20 DNA tethers incubated with HMfA and B) with HMfB measured by TPM in 50 mM Tris-HCl pH 7, 75 mM KCl. Histograms were fitted with a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

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

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

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

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

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

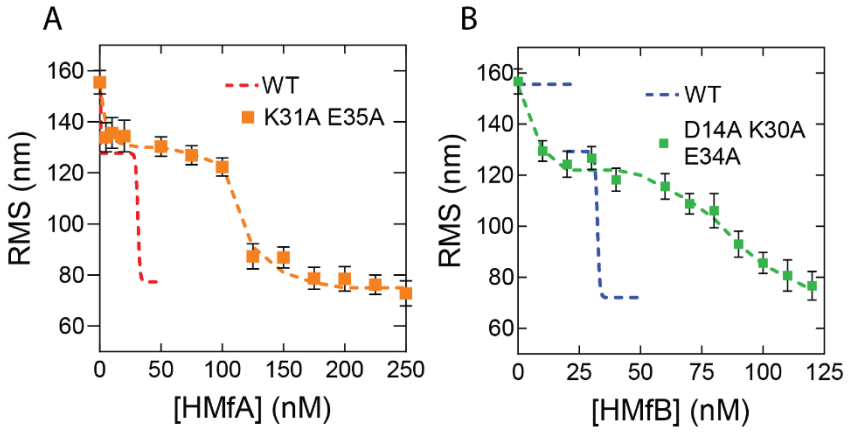


Figure 4.2 Histone derivatives HMfA$_{K31A\ E35A}$ and HMfB$_{D14A\ K30A\ E34A}$ recognize the Clone20 sequence A) Root mean square displacement (RMS) values of Clone20 DNA tethers with HMfA$_{K31A\ E35A}$ or B) HMfB$_{D14A\ K30A\ E34A}$ as measured by TPM in 50 mM Tris-HCl pH 7, 75 mM KCl. Wildtype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

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

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

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

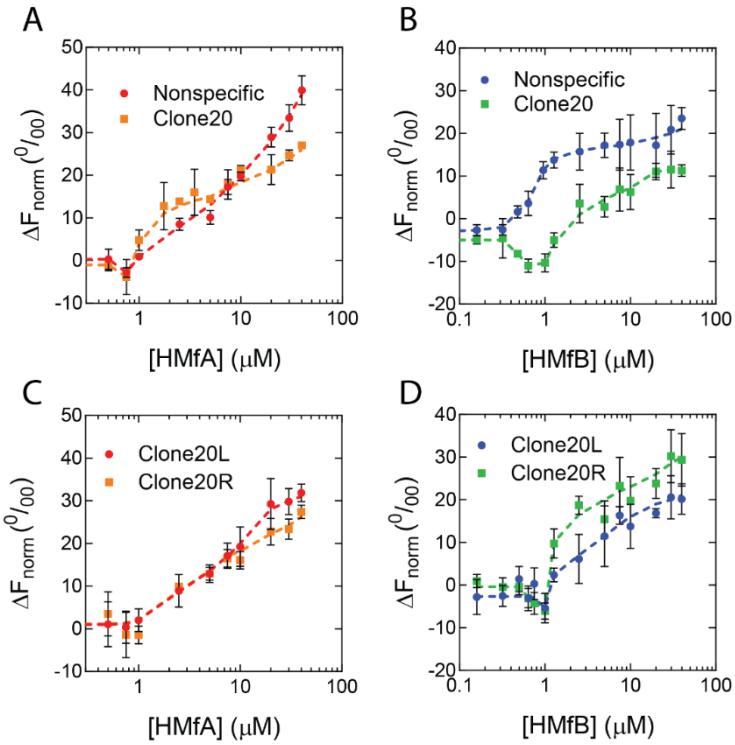


Figure 4.3 Binding of HMf proteins to short DNA substrates using Microscale Thermophoresis. Normalized thermophoresis curves of Nonspecific or Clone20 DNA as a function of A) HMfA or B) HMfB or of Clone20L or Clone20R as a function of C) HMfA or D) HMfB. Error bars indicate the standard deviation of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

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

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

Discussion

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

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

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

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

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

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

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

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

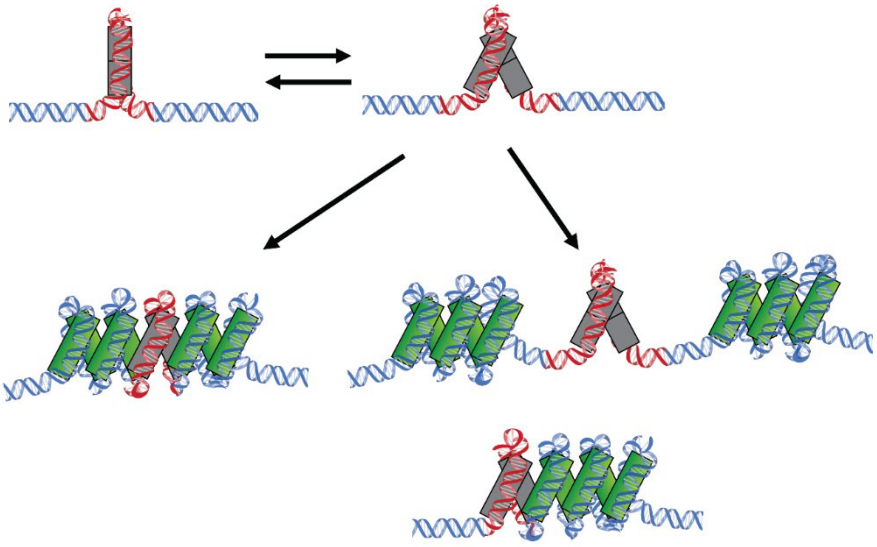


Figure 4.4 Mechanisms of HMf tetramers binding to specific DNA sequences followed by hypernucleosome formation. HMf tetramers bind to the Clone20 sequence and form a closed complex incompatible with further multimerization (top left). This structure can dynamically open and close (top right). The open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

References

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

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

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

Supplementary figures

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

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

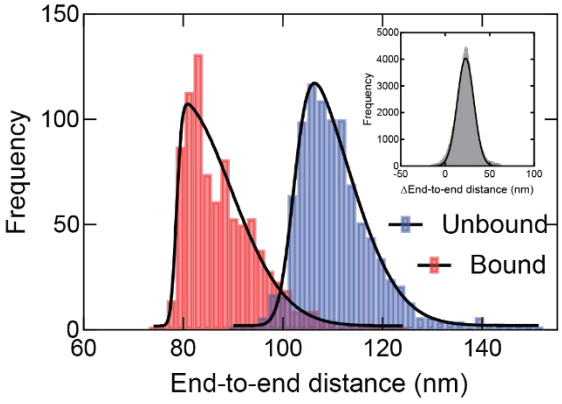


Figure S4.1 Calculated end-to-end distances for the unbound and bound population of 21 nM HMfB on Clone20 DNA. Histograms were fitted with a skewed normal distribution, resulting in end-to-end distances of 102 ± 10 nm and 78.6 ± 11 nm for unbound and bound DNA respectively. Insert: pairwise distribution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

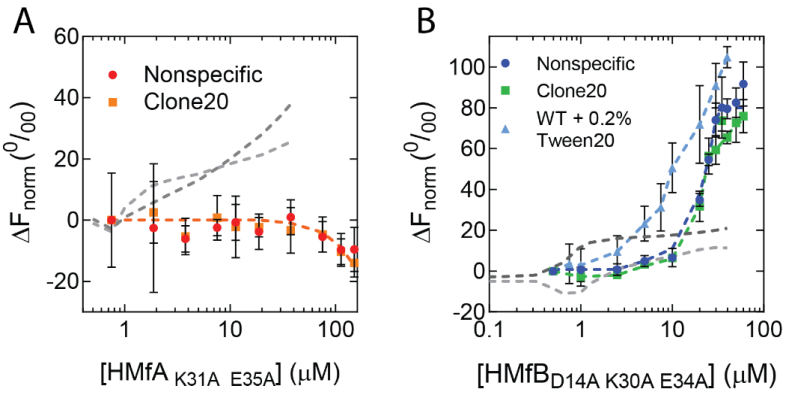


Figure S4.2: Binding of HMf derivatives to Nonspecific and Clone20 DNA substrates. Normalized thermophoresis curves of Nonspecifc or Clone20 DNA as a function of A) $HMfA_{K31A\ E35A}$ or B) $HMfB_{D14A\ K30A\ E34A}$. For HMfB WT, an extra curve on Nonspecific DNA was measured with 0.2% Tween20 in the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

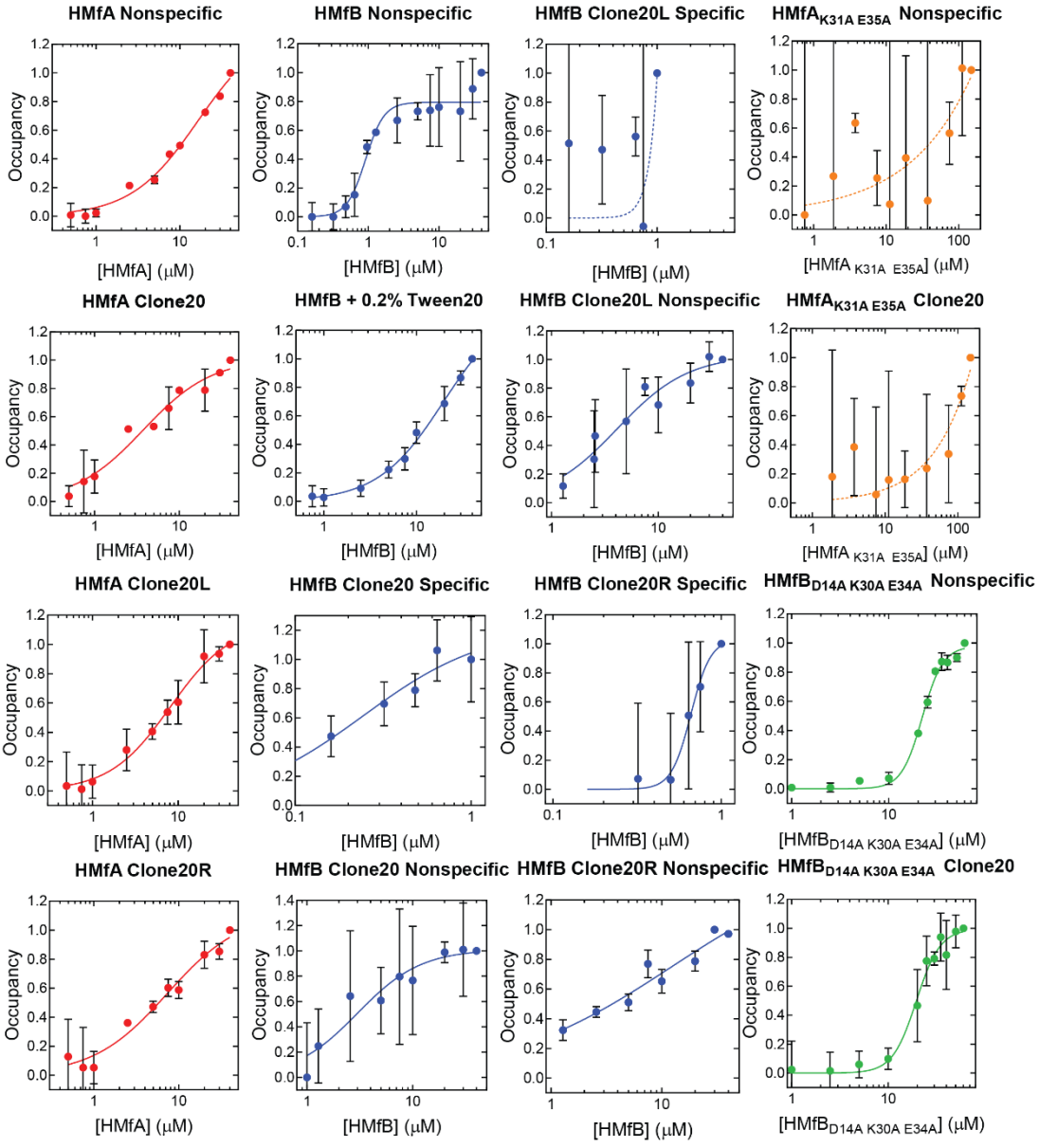


Figure S4.3: Hill fits of MST curves Occupancy was fitted against protein concentration where the occupancy at the highest concentration was set as 1.0. The dotted lines indicate fits that did not result in any reliable results.

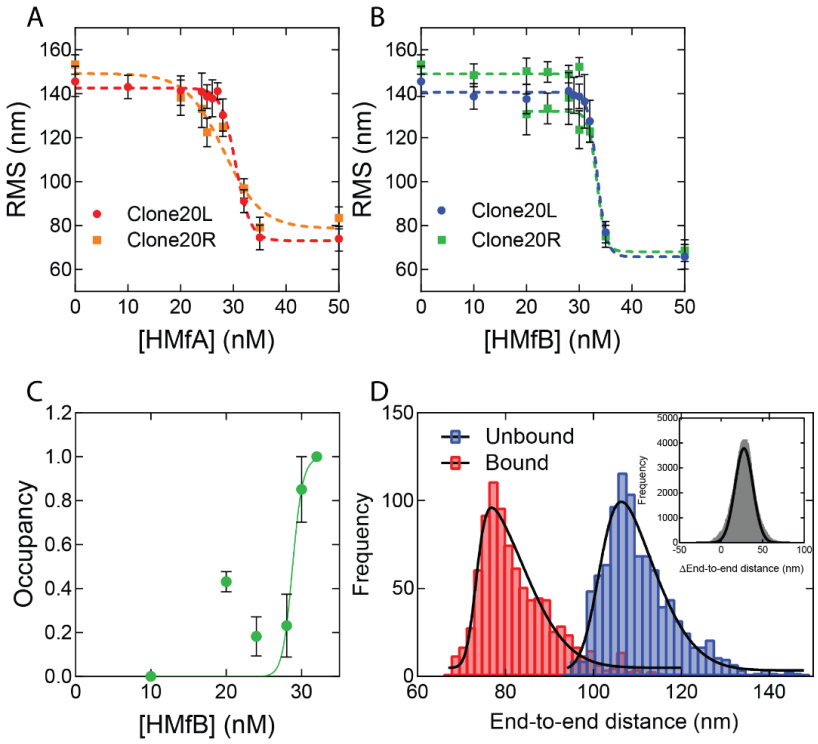


Figure S4.4 Binding of HMfA and HMfB to Clone20L or Clone20R in TPM experiments Root mean square displacement (RMS) of Clone20L and Clone20R DNA incubated with A) HMfA or B) HMfB in 50 mM Tris-HCl pH 7, 75 mM KCl. Histograms were fitted to a Gaussian distribution. Error bars represent the propagated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

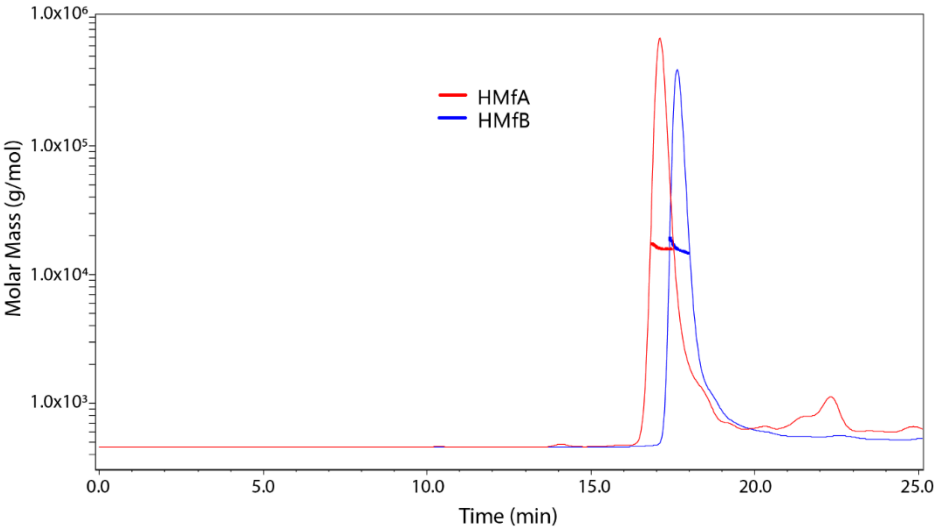


Figure S4.5 HMfA and HMfB are dimers in solution. SEC-MALS result of at least 1 mg/ml HMfA and HMfB. The determined molecular weight for HMfA was 16.3 ± 0.3 kDa and for HMfB 16.0 ± 0.1 kDa. The theoretical monomer mass of HMfA is 7.5 kDa and for HMfB 7.7 kDa.

Chapter 5

An atypical archaeal histone from *M. jannaschii* can form DNA bridges

This chapter is published as part of the following preprint: Ofer, S., Blombach, F., Erkelens, A.M., Barker, D., Schwab, S., Smollett, K., Matelska, D., Fouqueau, T., van der Vis, N., Kent, N.A., Dame, R.T. and Werner, F., DNA-bridging by an archaeal histone variant via a unique tetramerization interface, *Res Sq*, 10.21203/rs.3.rs-2183355/v1

Abstract

In eukaryotes, histone paralogues form obligate heterodimers such as H3/H4 and H2A/H2B that assemble into octameric nucleosome particles. Archaeal histones are dimeric and assemble on DNA into hypernucleosome particles of varying sizes, with each dimer wrapping 30 bp of DNA. These particles are likely composed of canonical and variant histone paralogues, but the function of these variants is poorly understood. Here, we characterize the structure and function of the histone paralogue MJ1647 from *Methanocaldococcus jannaschii,* which has a unique C-terminal extension that enables MJ1647 homo-tetramerization. Structural modelling and site-directed mutagenesis reveal that the C-terminal extension forms a tetramerization module consisting of two alpha-helices in a handshake arrangement. Unlike canonical histones, including *M. jannaschii* A3, the MJ1647 tetramer can bridge DNA duplexes *in vitro*. Using single-molecule tethered particle motion analyses, we show that MJ1647 tetramers bind ~60 bp and compact DNA in a highly cooperative manner. Our results reveal MJ1647 as the first histone with DNA bridging properties, which has important implications for genome structure and gene expression in archaea.

Introduction

Histone-based chromatin originated in Archaea (1, 2). Archaeal histones bear a close structural resemblance to their eukaryotic counterparts but differ in several important aspects (3). The archaeal histone encompasses the classical histone fold but lacks N- or C-terminal extensions, and there is no evidence of the extensive posttranslational modifications that are characteristic of eukaryotic histones and which regulate transcription (4). The tertiary structure of archaeal histones is near-identical to their eukaryotic counterparts, but the quaternary structure of the chromatin formed through interactions between histones and DNA is fundamentally different (5). While eukaryotic histones H2A, H2B, H3 and H4 form well-characterized octameric nucleosomes that wrap ~147 bp of DNA, canonical archaeal histones form dimers wrapping 30 bp of DNA (3, 5). The X-ray structure encompasses three *M. fervidus* histone homodimers assembled into a rod-like protein core stabilized by electrostatic interactions with a 90 bp DNA fragment wrapped around it in a left-handed solenoid (5). Notably, the dimensions of this structure, including the diameter and step size for each solenoid turn, are identical to those of the eukaryotic nucleosome, which reflects the highly conserved structure of the histone protein. Both analytical ultracentrifugation and cryo-electron microscopy (cryo-EM) experiments (6) and tethered particle motion (TPM) and magnetic tweezer (MT) experiments (7) measuring DNA compaction for two archaeal model histones, HMfA and HMfB, indicate extended hypernucleosome formation only limited in size by the length of the DNA used. Limited digestion of archaeal chromatin with Micrococcal Nuclease (MNase) produces a nucleosome ladder with step increments of 30 bp up to >400 bp length (8). The MNase data are thus consistent with histone homodimers binding 30 bp DNA as minimal chromatin subunit and forming longer hypernucleosomes *in vivo* (8, 9).

Eukaryotic nucleosomes self-interact to form phase-separated aggregates or structured 30 nm fibers (10), which are further stabilized by the linker histone H1 and heterochromatin protein 1 (HP1), which in turn can regulate the expression of specific genes (11–13). Analytical ultracentrifugation experiments, in contrast, do not provide evidence that archaeal hypernucleosome fibers associate with each other (6). These findings gave rise to the idea that the observed breathing of hypernucleosomal stacks is required in the absence of eukaryotic-like chromatin remodeling factors to facilitate transcription of chromatinised DNA (6). This inherent flexibility gives rise to the possibility

of binding of histone variants at the end of the hypernucleosome. Most archaea encode several histone paralogues that are closely related on the sequence level with little variation in length, amino acid composition and domain organization. However, hypernucleosome size may be limited *in vivo* by the incorporation of 'capstone' histone variants with weakened dimer-dimer interfaces in several archaea (14). Moreover, hypernucleosome stability is likely affected by differences in the number of stacking interactions available to different histone variants ((3, 7) and Chapter 4) and posttranslational modifications (acetylation) of the lysines involved in these interactions (15). Besides histone paralogues, different archaea encode a plethora of nucleoid-associated proteins (NAPs) (Chapter 1), including small basic proteins such as the highly abundant Alba, MC1 and Dps. These proteins can potentially affect the length of the hypernucleosome, and thus chromatin structure, as a function of growth conditions.

Importantly, interplay occurs between chromatin structure and genomic DNA transactions, such as transcription. Chromatin has the potential to deny access of the transcription machinery to gene promoters and thereby regulate gene expression. Supporting this notion, the deletion or mutation of one or multiple histones in *H. salinarum* and *T. kodakarensis* results in aberrant gene expression patterns (5, 16, 17) and severe impairment of DNA recombination (18). Several studies have characterized the impact of archaeal histones on transcription under rigorously defined conditions *in vitro* (19–22). These studies reveal that archaeal histones have an overall attenuating or inhibitory effect on transcription. Transcription elongation factors, including the transcript cleavage factor TFS and the processivity factor Spt4/5, enhance the transcription of archaeal RNA polymerase (RNAP) through histone-based chromatin (19). However, it remains unknown how transcription, as an extra layer of complexity, affects chromatin structure and *vice versa*. Also, it is unclear how the combinations of histone variants in the hypernucleosome and the binding of other NAPs to DNA would affect transcription.

M. jannaschii encodes four canonical histone paralogues, three on the main chromosome termed histone A1 to A3 that are highly abundant (4, 23), a less abundant paralogue on the extra-chromosomal elements (MJECL29), and an unusual histone variant MJ1647, which is encoded by most members of the Methanococcales (24, 25). MJ1647 homologues harbor a divergent histone fold that is followed by a C-terminal extension of ~27 amino acids. To investigate archaeal histones and the effect of their variants on hypernucleosome formation *in vitro*, we have characterized the structure and DNA binding properties of two representative histone variants from *M. jannaschii*, A3 and MJ1647. A combination of (single-molecule) biochemical experiments and structural

modeling demonstrate that both A3 and MJ1647 bind and compact DNA cooperatively. Moreover, we show that MJ1647 is a tetramer formed by two histone dimers that can bridge DNA duplexes and compact DNA in 60 bp steps.

Materials and methods

Protein purification

All *M. jannaschii* histones used in this study were kindly provided by Declan Barker and Finn Werner (26).

DNA substrate preparation

Tethered particle motion and bridging assay experiments were performed using a 47% GC 685 bp DNA substrate described earlier (7) unless otherwise stated. The DNA substrates were generated by PCR using Thermo Scientific® Phusion® High-Fidelity DNA Polymerase and the products were purified using the GenElute PCR Clean-up kit (Sigma-Aldrich).

For atomic force microscopy experiments, pUC19 plasmid (2686 bp) was digested with Nb.BsrDI (New England Biolabs) for one hour at 65 °C followed by inactivation at 85°C for 20 minutes and purified using phenol-chloroform extraction. The buffer was replaced with HPLC water (Sigma-Aldrich) through overnight dialysis using Slide-A-Lyzer 3,500 MWCO dialysis cassette (Thermo Scientific). Nicking of the plasmid was verified by agarose gel electrophoresis.

Tethered particle motion

Measurements were performed as previously described (27, 28) with minor modifications. Briefly, the flow cell was washed with 100 µL experimental buffer (50 mM Tris-HCl, pH 7 and 75 mM KCl) to remove excess beads and 100 µL protein diluted in experimental buffer was flowed in and incubated for 10 minutes. Next, the flow cell was washed with protein solution one more time and sealed with nail polish. After incubation, the flow cell was directly transferred to the holder and incubated for 5 more minutes in the instrument to stabilize the temperature at 25 °C for the measurement. For each flow cell, more than 200 beads were measured. Measurements were performed at least in duplicate at each concentration. Data analysis was done as described previously (28).

For the calculation of the end-to-end distance, 25 beads closest to the fitted RMS of a population were selected and the 2.5% most distant positions of each bead were collected. The end-to-end distance was determined by triangular calculation for each

point. Next, the data were plotted as histograms and the resulting distributions were fitted with a skewed Gaussian fit. Finally, a pairwise distance distribution was obtained by taking the difference in distance between each data point at one concentration to all data points at a second concentration. The resulting populations were fit with a Gaussian distribution.

DNA bridging assay

The DNA used for the bridging assay is the same as that used for TPM and was ^{32}P-labelled (29). The DNA bridging assay was performed as described previously (30, 31) with minor modifications. First, streptavidin-coated Magnetic M-280 Dynabeads (Invitrogen) were resuspended in buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 2 M NaCl, 2 mg/mL BSA (ac), 0.04% Tween 20) containing 100 fmol biotinylated 47% GC DNA (685 bp) and incubated at 1000 rpm for 20 min at 25 °C in an Eppendorf Thermomixer with an Eppendorf Smartblock™ 1.5 mL. The beads with associated DNA were washed twice before resuspension in incubation buffer. Radioactive ^{32}P-labeled DNA and unlabeled DNA were combined to maintain a constant (2 fmol/µl) concentration and a radioactive signal around 8000 cpm and then added to each tube. Next, protein was added to initiate the formation of bridged protein-DNA complexes. The composition of the experimental buffer after combining all components in 20 µl was 10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM KCl, 5% v/v glycerol, 0,016% Tween20, 0,8 mg/ml acetylated BSA, 1 mM $MgCl_2$, 1 mM spermidine, 1 mM DTT, 5 mM EDTA. The samples were incubated for 20 min at 1000 rpm at 25°C in an Eppendorf Thermomixer with an Eppendorf Smartblock™ 1.5 mL. After the incubation the beads were washed with 20 µl of the same experimental buffer and then resuspended in 12 µl counting buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 200 mM NaCl, 0.2% SDS). The radioactive signal of DNA was quantified by liquid scintillation. It was used to calculate protein DNA recovery (%) based on a reference sample containing the same amount of labelled ^{32}P 685 bp DNA used in each sample. All DNA bridging experiments were performed at least in triplicate.

Atomic Force Microscopy

MJ1647 was incubated with 50 ng nicked pUC19 in AFM buffer A (10 mM HEPES pH 7.5, 100 mM NaCl) for 30 minutes at 37°C to form complexes. The resulting mixture was then diluted to a final buffer composition of 4 mM HEPES, 3 mM $MgCl_2$, 5 mM NaCl. Next, 10 µl of the solution was applied onto a freshly cleaved mica disk for 10 minutes. Subsequently, the mica disk was rinsed with 10 ml HPLC water, with the excess water being removed using lint-free tissue and dried with filtered N_2 gas.

Imaging was done in a JPK instruments Nanowizard 3 system in AC mode using a TESPA probe (320 kHz resonance frequency) (Nanoworld). Images were captured in a 512x512 resolution at a line rate of 1.5 kHz. Image flattening and contrast adjustments were made using JPK Data Processing software for optimal viewing. Measurement of contour length was done using WsXM 5.0 Develop 10.2 (32).

Structural modelling of MJ1647 dimers and tetramers

For the AlphaFold predictions, MMseqs2 and LocalColabFold were run on the high-performance computing facility ALICE at Leiden University (33–36). A multiple sequence alignment (MSA) for MJ1647 was generated with MMseqs2 (commit bfc6f85 from December 5, 2021). Target databases used for this MSA were constructed by the ColabFold team (https://colabfold.mmseqs.com/) and included UniRef30, BFD, Mgnify, MetaEuk, SMAG, TOPAZ, MGV, GPD, and MetaClust2. The search-sensitive parameter was set to 8. The constructed MSA was used as an input for LocalColabFold (LocalColabFold: commit 6b76904 from December 4, 2021, ColabFold: commit 33fcb9a from December 7, 2021) to predict the dimer and tetramer structures of MJ1647. No templates and 3 recycles were used for these predictions. The structures were relaxed by AlphaFold's AMBER forcefield.

Results

Single-molecule experiments reveal DNA compaction

To test the DNA compaction properties of A3 and MJ1647, we carried out Tethered Particle Motion (TPM) experiments. This single-molecule technique reports on the length and conformation of dsDNA duplexes tethered to a surface at one end by reporting the Root Mean Square motion (RMS) of a bead attached to the other end. The RMS is reduced upon DNA compaction (27). First, we investigated the effect of the binding of the A3 histone on DNA conformation. Experiments with A3 demonstrate a gradual progressive DNA compaction that is evident from a reduction in RMS upon titration of A3 and indicative of hypernucleosome formation (figure 5.1A). The RMS value under saturated conditions is ~80 nm, a value that is comparable to that obtained with HMfB (7). The formation of hypernucleosomes occurs with lower cooperativity compared to HMfB, which can be attributed to A3 having only one predicted stacking interaction (E35-K66) instead of three in HMfB (D14-R48, K30-E61 and E34-R65).

Besides having a C-terminal extension, MJ1647 is also divergent from canonical archaeal histones in terms of amino acid sequence, including a substitution of residue

R19 (HMfB numbering) of the so-called 'RT-pair'. This motif is highly conserved in canonical histones and mediates DNA binding, suggesting a different DNA binding mode for MJ1647 or reduced DNA affinity (5, 25, 37, 38). Titration of MJ1647, like A3, leads to increasing DNA compaction, but strikingly distinct populations corresponding to distinct successive binding events of MJ1647 were observed (figure 5.1B). Overall, MJ1647 compacted DNA in a cooperative manner, but the distinct populations make it challenging to quantify this effect. To extract quantitative structural information on MJ1647 binding from these data, we transformed the RMS into end-to-end distance ((7) and Materials & Methods; see figure 5.1C). Notably, the pairwise distance between the peaks is ~ 20 nm. Previous experiments with a SELEX-optimized high-affinity site at which HMfB tetramers bind and wrap 60 bp of DNA (39), also yielded a reduction in end-to-end distance of ~ 20 nm (Chapter 4). Our data would thus be compatible with MJ1647 wrapping ~60 bp of DNA in a manner similar to that of canonical histones. This result is in agreement with MJ1647 forming tetramers in Size Exclusion Chromatography experiments combined with Multi-Angle Light Scattering (SEC-MALS) and binding of MJ1647 to 60 bp, but not to 30 bp DNA fragments in Electromobility Shift Assays (EMSAs) (26). The tetramerization of MJ1647 was shown to be dependent on the 27 amino acid C-terminal extension, called the tetramerization module (TM, (26)). TPM experiments with MJ1647ΔTM revealed gradual DNA compaction as a function of protein concentration without visible steps, confirming that the deletion of the TM leads to a DNA binding mode similar to that of canonical histone A3 and HMfB (figure 5.1D).

MJ1647 but not A3 mediates DNA bridging

As MJ1647 forms tetramers in solution and interacts with 60 bp of DNA, we speculated whether MJ1647 could interact with DNA in trans by bridging two 30 bp DNA sites in addition, or alternatively, to binding to two adjacent 30 bp DNA binding sites (i.e. 60 bp) in cis. We tested the bridging properties of histones using a radiolabeled DNA pulldown assay where a biotin-labelled DNA molecule is immobilized on magnetic streptavidin-coated beads (30, 31, 40). Histone variants are added together with a second 32P-labelled DNA, and if bridging between the two DNA species occurs, the 32P-labelled DNA is immobilized on the beads and quantified as % recovery. The results show that MJ1647 bridges DNA duplexes with a half-maximal recovery at 2.5 μM (figure 5.2A). This property depends on the MJ1647 TM as neither A3 nor MJ1647ΔTM can bridge DNA (figure 5.2A).

To confirm the DNA bridging ability of MJ1647 and examine the conformations of DNA induced by bridging, we performed Atomic Force Microscopy (AFM) on MJ1647-DNA complexes. Next, MJ1647-DNA complexes were classified as 'circular' or 'entangled' (figure 5.2B). The relative amount of entangled complexes increases with increasing MJ1647 concentration (figure 5.2C). This result confirms that MJ1647 is responsible for the observed DNA crossovers and is most likely the result of MJ1647's DNA bridging ability.

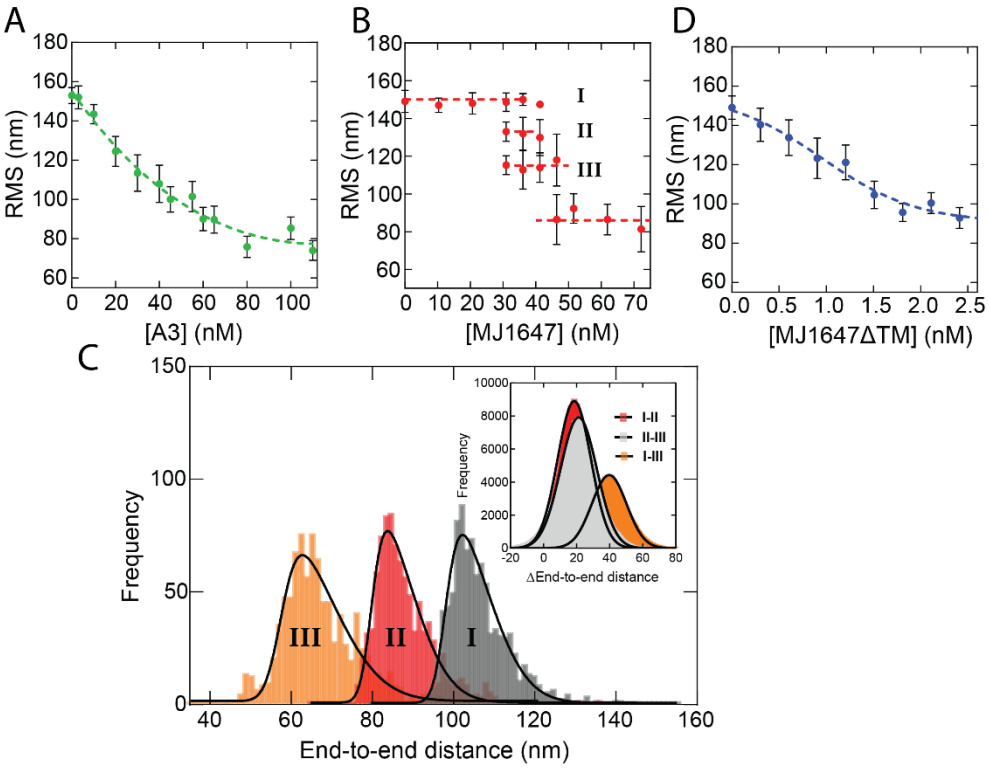


Figure 5.1 MJ1647 tetramers compact DNA. Tethered particle motion (TPM) experiments were used to investigate DNA compaction by A3 (A) or MJ1647 (B). Root mean square displacement (RMS) values are plotted as a function of histone concentration. The RMS values were obtained by fitting the data with a Gaussian distribution. Error bars represent the propagated standard deviation from at least two individual measurements. Dashed lines serve as lines to guide the eye. C) Histograms of calculated end-to-end distances of the three populations observed at 36 nM MJ1647. Of each population, the 25 beads closest to the fitted RMS value were selected and the end-to-end distance was calculated for the 2.5% most distant positions with respect to the center of the beads. Histograms were fit with a skewed

normal distribution. Insert: pairwise distribution plot of the differences between the end-to-end distance peaks I, II and III. Histograms were fitted with a Gaussian distribution. D) TPM experiments testing DNA compaction by MJ1647ΔTM.

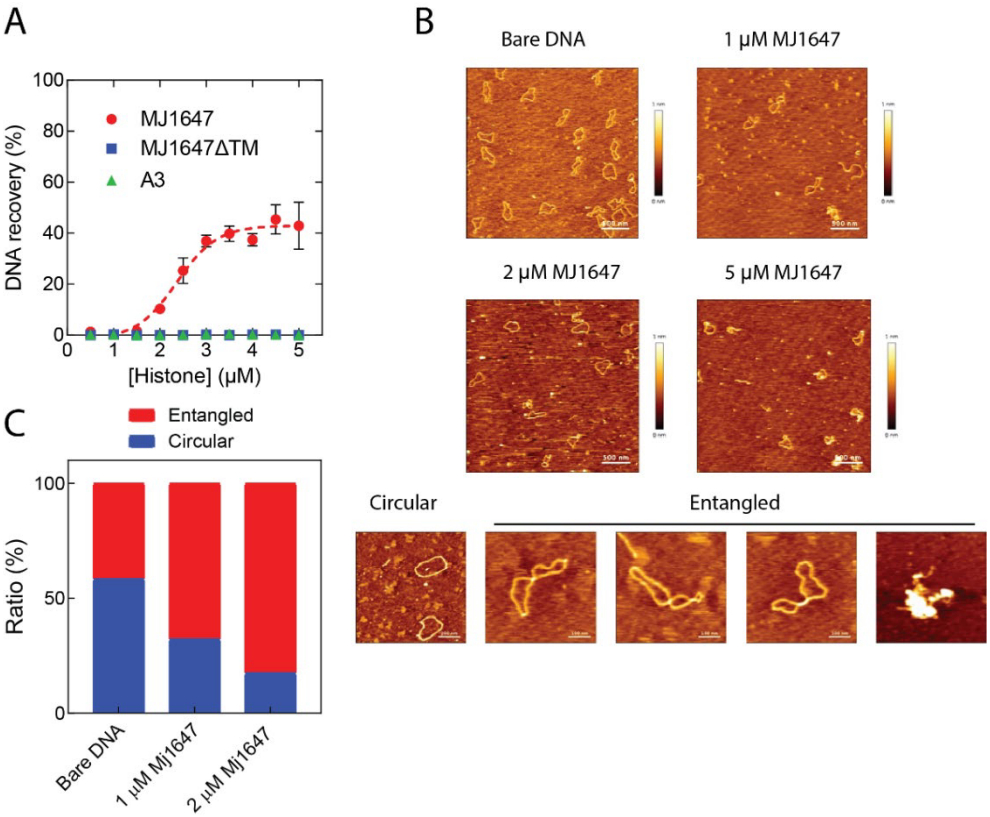


Figure 5.2: MJ1647 bridges DNA. A) DNA recovery (%) as a function of protein concentration. Data are plotted as mean values of three independent measurements and error bars represent the standard deviation. The dashed line is a line to guide the eye. B) Nicked pUC19 was incubated with an increasing amount of MJ1647, as indicated. Bridges and DNA-protein aggregates were formed with the addition of protein and were classified as 'entangled'. C) Bar graph summarizing the population of circular and entangled plasmids incubated with the indicated amount of MJ1647 (N=130 for Bare DNA, N=93 for 1 μM MJ1647, N=80 for 2 μM MJ1647). 'Circular DNA' refers to DNA duplexes without structural features, such as crossovers. 'Entangled DNA' refers to DNA containing single or multiple crossovers, overlapping sections and aggregation.

Structural basis of MJ1647 dimer and tetramer formation

The molecular structure of an MJ1647 dimer has been solved at a resolution of 1.9 Å (26). The model revealed the classical histone fold of MJ1647 followed by two alpha-helices separated by a loop (corresponding to the TM). The two helices of each monomer are packed against each other to create an interleaved or 'handshake' arrangement. Thus the TM of MJ1647 is structured in contrast to the tails of eukaryotic histones.

Unable to obtain crystals of the MJ1647 tetramer, we generated structural models for MJ1647 tetramers using the structure prediction algorithm AlphaFold2 (AF2, (33–36)) (figure 5.3A and figure S5.1). In the MJ1647 tetramer model, the two DNA-binding histone folds are located opposite each other and connected by the four MJ1647 TMs. While TM helices 4 and 5 form monomer-monomer interactions within MJ1647 dimers in the crystal structure (26), TM helices 4 and 5 'open up' in the tetrameric model and make new interactions with opposing monomers that enable dimer-dimer interactions within the tetramer (figure 5.3A and B). The predicted aligned error (PAE) indicates high confidence (PAEs < 10 Å) in the relative positions of the residues in the tetramerization module (figure S5.1B). The interface between the two dimers consists of the tetramerization module's hydrophobic core that is stabilized by four salt bridges between residues K80 and E95, both of which are strictly conserved in all MJ1647 homologues (figure S5.2). In addition, K68 makes polar contacts with the backbone of T93 and L96 of the tetramer model, albeit with low confidence for side chain orientation for residues 68 to 80.

To experimentally validate the tetrameric model of MJ1647, two charge-reversal mutants (K80E and E95K) and the K80E E95K double mutant were generated. While wildtype MJ1647 eluded as a tetramer in SEC-MALS experiments, the K80E substitution resulted in a delayed and broader elution profile indicating interference with tetramerization (26). We tested the salt bridge mutants for their capability to bind and bridge DNA. All three mutants retained their ability to bridge DNA, although the concentration range in which bridging occurred was shifted (figure 5.3C and figure S5.3). This suggests that the salt bridge mutations alone do not disrupt tetramerization entirely, and a lower affinity to form tetramers might still be sufficient for the observed DNA bridging behavior of MJ1647.

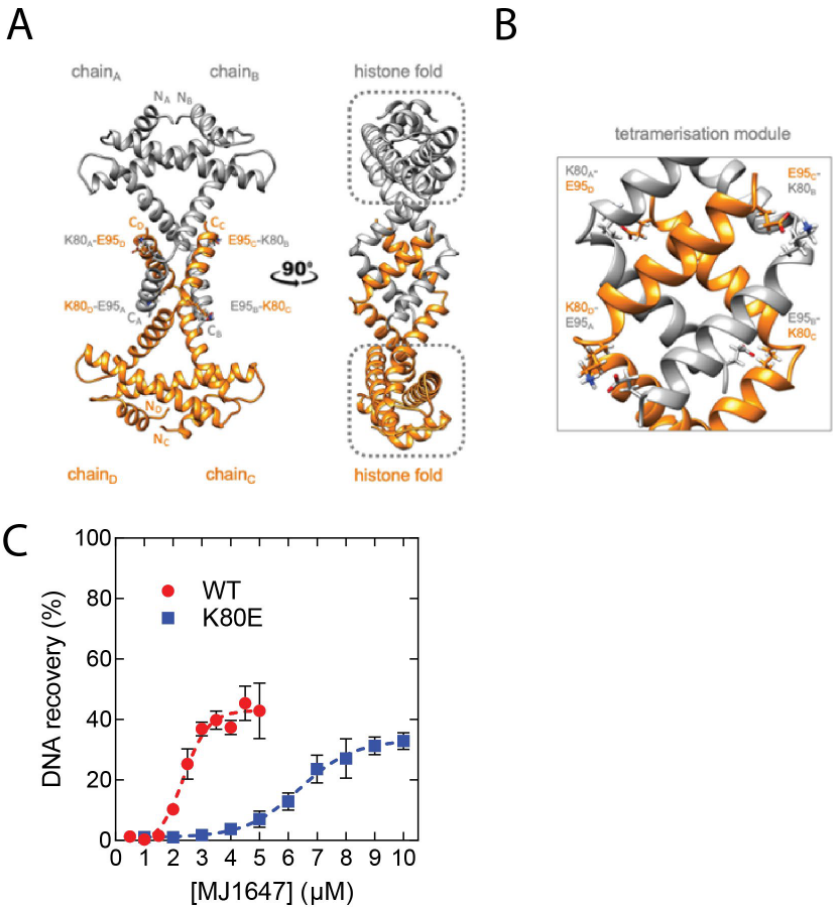


Figure 5.3: AlphaFold2 predicts tetramerization interface. A) AlphaFold2 model of the tetrameric 'dimer of dimer' MJ1647 species with the two dimers (chain A/B and chain C/D) highlighted in dark grey and orange, respectively. The tetramer is organized into two histone folds and a tetramerization center formed by the TM. B) Four salt bridges contribute to tetramer stability. These are formed between residues K80 and E95 between chains A and D and B and C, respectively. C) DNA recovery (%) as a function of MJ1647 WT or K80E concentration. MJ1647 WT data was reproduced from figure 5.2 for easy comparison. Data are plotted as mean values of three independent measurements and error bars represent the standard deviation. Dashed lines serve as lines to guide the eye.

Discussion

In contrast to eukaryotes that utilize the ubiquitous histone octamer as functional chromatin building block, archaea employ a range of different chromatin proteins to enable genome compaction and gene regulation (1). In most archaeal phyla, histones, often a combination of well-characterized canonical histones and variants with largely unknown structure and function, are believed to be critical players in chromatin organization. Here we have set out to systematically compare the canonical A3 histone from *M. jannaschii* and the MJ1647 variant that piqued our interest due to its unusual C-terminal extension.

At the single-molecule level, TPM results demonstrate that A3 forms hypernucleosomes *in vitro* despite providing fewer stacking interactions than other histones, including HMfB. When comparing the TPM curves of A3 and HMfB (7) it becomes clear that A3 binds DNA with lower cooperativity, which could be the result of having fewer stacking interactions. In contrast, MJ1647 shortens the DNA in experimentally discernible larger steps consistent with 60 bp DNA wraps. This more extensive DNA wrapping behavior depends on the TM (figure 5.1), and it suggests that the two histones exhibit two distinct DNA binding modes. DNA bridging assays demonstrate that MJ1647, but not A3, can bridge two DNA duplexes, likely enabling the formation of DNA loops (figure 5.2). This highlights MJ1647, to the best of our knowledge, as the first histone able to form inter-duplex DNA connections.

MJ1647 forms tetramers in solution and this requires the TM (26). In our high-confidence model of tetrameric MJ1647 generated by AlphaFold2, two MJ1647 dimers interact via the TM, where the two TM helices hinge open and form contacts with the TM helices of the opposite dimer. Four salt bridges stabilize this arrangement. We confirmed this homology model by introducing charge reversal and double charge reversal substitutions that impaired and restored tetramers, respectively (figure 5.3 and figure S5.3).

Combining this structural information with our (single-molecule) experimental data results in a model where MJ1647 has two distinct conformations (figure 5.4). A “standard” histone-like DNA wrapping conformation explains the observed tetramer binding of MJ1647 in TPM experiments. The tetramer can be stabilized either by interactions in TM or by dimer-dimer interactions, which are conserved between MJ1647 and canonical archaeal histones like HMfB (3). This conformation is also conserved in the eukaryotic $(H3\text{-}H4)_2$ tetramer (41). The second conformation, predicted by AlphaFold2, is a DNA-bridging conformation where two dimers individually bind to a DNA

duplex and the TM connects the two dimers to bridge the DNA duplexes. Changing the equilibrium between these two states could play a role in gene regulation by MJ1647. It is, however, currently unknown and hard to predict which state could be repressive and which could be permissive of gene transcription. As MJ1647 does not heterodimerize with A3 (26), *M. jannaschii* may partition its chromatin into canonical histone and MJ1647 filaments, similar to the partitioning between histones and the NAP TrmBL2 found in *T. kodakarensis* (42).

What drove the evolution of MJ1647 homologues in Methanococcales? Growth temperature has been proposed as a major driving force for the evolution of additional NAPs and histones in archaea (43). In line with this idea, the ancestor of Methanococcales, the lineage in which MJ1647 evolved, is predicted to have been a thermophile (45), just as *M. jannaschii*. Notably, MJ1647 homologues were retained not only in thermophilic Methanococcales species but also those Methanococcales species adapting to mesophilic growth conditions such as *Methanococcus maripaludis*. Other, eukaryotic, histone variants with a C-terminal extension have also been implied in adaptation to environmental conditions. Deletion of the C-terminal tail of canonical H2A also increases stress sensitivity (45). Organisms belonging to the class Bdelloid rotifers are extremely resistant to radiation and undergo frequent dehydration (46). They encode an H2A variant with an unusual C-terminal tail, which is suggested to aid the organism in dealing with DNA damage caused by these environmental stresses (47). Another driving force of histone evolution and diversification might have been the inherent sequence symmetry of histones (48). Alignment of H2A, HMfB and MJ1647 with their reversed amino acid sequences reveals a symmetric distribution of the hydrophobic interactions that promote dimer formation (figure S5.4). This could lead to the formation of a stable inverted histone dimer where helix 1 of the first monomer is close to helix 3 of the second monomer instead of both helices 1 (and 3) being on the same side of the dimer complex (48). This conformation could give extra possibilities for sequence diversity and additional domains at the N- or C-terminus of the histone fold. However, the inverted conformation has only been predicted by computational methods and its existence *in vitro* or *in vivo* remains to be established.

Archaea lack linker histones and hypernucleosomes formed by model archaeal histones HMf and HTk do not self-associate, unlike eukaryotic nucleosomes (6). Our finding that MJ1647 bridges DNA, opens the possibility that MJ1647 plays a role in the higher-order organization of archaeal chromatin. The best characterized archaeal NAP

mediating DNA bridging is Alba (49–52). Alba constitutes a large fraction of chromatin in many archaeal species, such as *Sulfolobus shibatae,* in which it constitutes 4% of all cellular protein (53). In contrast, shotgun proteomics and quantitative immunodetection data suggest much lower expression for Alba in *M. jannaschii* (23, 54). Expression levels of Alba in the mesophile *M. maripaludis* are even lower at about 0.01% of total cellular protein, coinciding with the evolution of Alba into a protein with DNA sequence specificity (54). It is tempting to speculate that MJ1647 homologues functionally replaced Alba in its role of higher-order DNA compaction.

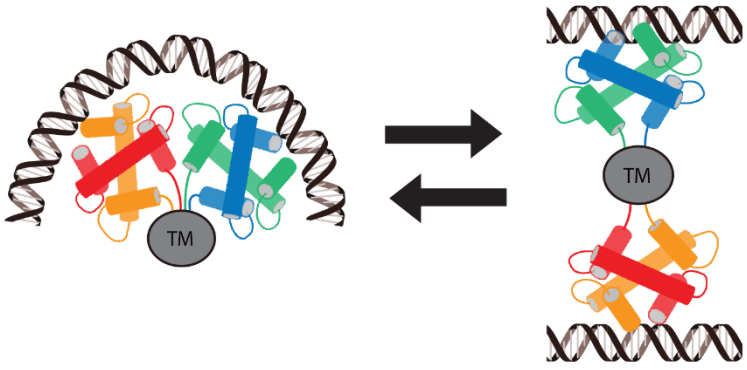


Figure 5.4 MJ1647 has two conformations. A “standard” histone-like DNA wrapping conformation is shown on the left, where MJ1647 binds as a tetramer to a single DNA duplex. The tetramer can be formed by either the TM or the dimer-dimer interactions as for other archaeal histones, or both. The DNA-wrapping conformation is in equilibrium with the DNA-bridging conformation, where both dimers bind a DNA duplex, and the dimers are connected via the TM to form a DNA-bridging tetramer, as shown on the right.

References

1. Peeters,E., Driessen,R.P.C., Werner,F. and Dame,R.T. (2015) The interplay between nucleoid organization and transcription in archaeal genomes. *Nature Reviews Microbiology*, **13**, 333–341.

2. Ammar,R., Torti,D., Tsui,K., Gebbia,M., Durbic,T., Bader,G.D., Giaever,G. and Nislow,C. (2012) Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife*, **1**, e00078.

3. Henneman,B., van Emmerik,C., van Ingen,H. and Dame,R.T. (2018) Structure and function of archaeal histones. *PLoS Genet*, **14**, e1007582.

4. Forbes,A.J., Patrie,S.M., Taylor,G.K., Kim,Y. Bin, Jiang,L. and Kelleher,N.L. (2004) Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 2678–2683.

5. Mattiroli,F., Bhattacharyya,S., Dyer,P.N., White,A.E., Sandman,K., Burkhart,B.W., Byrne,K.R., Lee,T., Ahn,N.G., Santangelo,T.J., *et al.* (2017) Structure of histone-based chromatin in Archaea. *Science (1979)*, **357**, 609–612.

6. Bowerman,S., Wereszczynski,J. and Luger,K. (2021) Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways. *eLife*, **10**, e65587.

7. Henneman,B., Brouwer,T.B., Erkelens,A.M., Kuijntjes,G.-J., van Emmerik,C., van der Valk,R.A., Timmer,M., Kirolos,N.C.S., van Ingen,H., van Noort,J., *et al.* (2021) Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res*, **49**, 4338–4349.

8. Maruyama,H., Harwood,J.C., Moore,K.M., Paszkiewicz,K., Durley,S.C., Fukushima,H., Atomi,H., Takeyasu,K. and Kent,N.A. (2013) An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis. *EMBO Rep*, **14**, 711–7.

9. Nalabothula,N., Xi,L., Bhattacharyya,S., Widom,J., Wang,J.P., Reeve,J.N., Santangelo,T.J. and Fondufe-Mittendorf,Y.N. (2013) Archaeal nucleosome positioning in vivo and in vitro is directed by primary sequence motifs. *BMC Genomics*, **14**, 391.

10. Hansen,J.C., Maeshima,K. and Hendzel,M.J. (2021) The solid and liquid states of chromatin. *Epigenetics and Chromatin*, **14**, 1–33.

11. Cutter,A.R. and Hayes,J.J. (2017) Linker histones: novel insights into structure-specific recognition of the nucleosome. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, **95**, 171–178.

12. Robinson,P.J. and Rhodes,D. (2006) Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Current opinion in structural biology*, **16**, 336–343.

13. Jedrusik-Bode,M. (2013) Histone H1 and heterochromatin protein 1 (HP1) regulate specific gene expression and not global transcription. *Worm*, **2**.

14. Stevens,K.M., Swadling,J.B., Hocher,A., Bang,C., Gribaldo,S., Schmitz,R.A. and Warnecke,T. (2020) Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proc Natl Acad Sci U S A*, **117**, 33384–33395.

15. Alpha-Bazin,B., Gorlas,A., Lagorce,A., Joulié,D., Boyer,J.B., Dutertre,M., Gaillard,J.C., Lopes,A., Zivanovic,Y., Dedieu,A., *et al.* (2021) Lysine-specific acetylated proteome from the archaeon Thermococcus gammatolerans reveals the presence of acetylated histones. *Journal of Proteomics*, **232**, 104044.

16. Dulmage,K.A., Todor,H. and Schmid,A.K. (2015) Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein. *mBio*, **6**, e00649-15.

17. Sanders,T.J., Ullah,F., Gehring,A.M., Burkhart,B.W., Vickerman,R.L., Fernando,S., Gardner,A.F., Ben-Hur,A. and Santangelo,T.J. (2021) Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in Thermococcus kodakarensis. *Front Microbiol*, **12**, 1071.

18. Čuboňová,L., Katano,M., Kanai,T., Atomi,H., Reeve,J.N. and Santangelo,T.J. (2012) An Archaeal Histone Is Required for Transformation of Thermococcus kodakarensis. *Journal of Bacteriology*, **194**, 6864.

19. Sanders,T.J., Lammers,M., Marshall,C.J., Walker,J.E., Lynch,E.R. and Santangelo,T.J. (2019) TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Molecular Microbiology*, **111**, 784–797.

20. Wilkinson,S.P., Ouhammouch,M. and Geiduschek,E.P. (2010) Transcriptional activation in the context of repression mediated by archaeal histones. *Proceedings of the National Academy of Sciences*, **107**, 6777–6781.

21. Xie,Y. and Reeve,J.N. (2004) Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *Journal of Bacteriology*, **186**, 3492–3498.

22. Soares,D., Dahlke,I., Li,W.T., Sandman,K., Hethke,C., Thomm,M. and Reeve,J.N. (1998) Archaeal histone stability, DNA binding, and transcription inhibition above 90°C. *Extremophiles 1998 2:2*, **2**, 75–81.

23. Zhu,W., Reich,C.I., Olsen,G.J., Giometti,C.S. and Yates,J.R. (2004) Shotgun proteomics of Methanococcus jannaschii and insights into methanogenesis. *Journal of Proteome Research*, **3**, 538–548.

24. Agha-amiri,K. and Klein,A. (1993) Nucleotide sequence of a gene encoding a histone-like protein in the archaeon Methanococcus voltae. *Nucleic Acids Research*, **21**, 1491.

25. Li,W.T., Sandman,K., Pereira,S.L. and Reeve,J.N. (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**, 43–51.

26. Ofer,S., Blombach,F., Erkelens,A.M., Barker,D., Schwab,S., Smollett,K., Matelska,D., Fouqueau,T., van der Vis,N., Kent,N.A., *et al.* (2023) DNA-bridging by an archaeal histone variant via a unique tetramerisation interface. *Res Sq*, 10.21203/rs.3.rs-2183355/v1.

27. van der Valk,R.A., Laurens,N. and Dame,R.T. (2017) Tethered particle motion analysis of the DNA binding properties of architectural proteins. In *Methods in Molecular Biology*. Humana Press Inc., Vol. 1624, pp. 127–143.

28. Henneman,B., Heinsman,J., Battjes,J. and Dame,R.T. (2018) Quantitation of DNA-binding affinity using tethered particle motion. In *Methods in Molecular Biology*. Humana Press Inc., Vol. 1837, pp. 257–275.

29. Wagner,K., Moolenaar,G.F. and Goosen,N. (2011) Role of the insertion domain and the zinc-finger motif of Escherichia coli UvrA in damage recognition and ATP hydrolysis. *DNA repair*, **10**, 483–496.

30. van der Valk,R.A., Vreede,J., Qin,L., Moolenaar,G.F., Hofmann,A., Goosen,N. and Dame,R.T. (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-Bridging activity. *Elife*, **6**, e27369.

31. van der Valk,R.A., Qin,L., Moolenaar,G.F. and Dame,R.T. (2018) Quantitative determination of DNA bridging efficiency of chromatin proteins. In Dame R.T. (ed), *Methods in Molecular Biology*. Humana Press, New York, NY, Vol. 1837, pp. 199–209.

32. Horcas,I., Fernández,R., Gómez-Rodríguez,J.M., Colchero,J., Gómez-Herrero,J. and Baro,A.M. (2007) WSXM: A software for scanning probe microscopy and a tool for nanotechnology. *Review of Scientific Instruments*, **78**, 013705.

33. Steinegger,M. and Söding,J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology 2017 35:11*, **35**, 1026–1028.

34. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, **596**, 583–589.

35. Evans,R., O'Neill,M., Pritzel,A., Antropova,N., Senior,A., Green,T., Žídek,A., Bates,R., Blackwell,S., Yim,J., *et al.* (2021) Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 10.1101/2021.10.04.463034.

36. Mirdita,M., Schütze,K., Moriwaki,Y., Heo,L., Ovchinnikov,S. and Steinegger,M. (2022) ColabFold: making protein folding accessible to all. *Nature Methods 2022 19:6*, **19**, 679–682.

37. Soares,D.J., Sandman,K. and Reeve,J.N. (2000) Mutational analysis of archaeal histone-DNA interactions. *Journal of Molecular Biology*, **297**, 39–47.

38. Decanniere,K., Babu,A.M., Sandman,K., Reeve,J.N. and Heinemann,U. (2000) Crystal structures of recombinant histones HMfA and HMfB, from the hyperthermophilic archaeon Methanothermus fervidus. *Journal of Molecular Biology*, **303**, 35–47.

39. Bailey,K.A., Pereira,S.L., Widom,J. and Reeve,J.N. (2000) Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution (vol 303, pg 25, 2000). *Journal of Molecular Biology*, **304**, 493.

40. Qin,L., Bdira,F. Ben, Sterckx,Y.G.J., Volkov,A.N., Vreede,J., Giachin,G., van Schaik,P., Ubbink,M. and Dame,R.T. (2020) Structural basis for osmotic regulation of the DNA binding properties of H-NS proteins. *Nucleic Acids Res*, **48**, 2156–2172.

41. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251–260.

42. Maruyama,H., Shin,M., Oda,T., Matsumi,R., Ohniwa,R.L., Itoh,T., Shirahige,K., Imanaka,T., Atomi,H., Yoshimura,S.H., *et al.* (2011) Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus kodakarensis. *Molecular Biology of the Cell*, **22**, 386.

43. Hocher,A., Borrel,G., Fadhlaoui,K., Brugère,J.F., Gribaldo,S. and Warnecke,T. (2022) Growth temperature and chromatinization in archaea. *Nature Microbiology 2022 7:11*, **7**, 1932–1942.

44. Lecocq,M., Groussin,M., Gouy,M. and Brochier-Armanet,C. (2021) The Molecular Determinants of Thermoadaptation: Methanococcales as a Case Study. *Molecular Biology and Evolution*, **38**, 1761.

45. Vogler,C., Huber,C., Waldmann,T., Ettig,R., Braun,L., Izzo,A., Daujat,S., Chassignet,I., Lopez-Contreras,A.J., Fernandez-Capetillo,O., *et al.* (2010) Histone H2A C-Terminus Regulates Chromatin Dynamics, Remodeling, and Histone H1 Binding. *PLoS Genetics*, **6**, 1–12.

46. Gladyshev,E. and Meselson,M. (2008) Extreme resistance of bdelloid rotifers to ionizing radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 5139–5144.

47. Van Doninck,K., Mandigo,M.L., Hur,J.H., Wang,P., Guglielmini,J., C.Milinkovitch,M., Lane,W.S. and Meselson,M. (2009) Phylogenomics of Unusual Histone H2A Variants in Bdelloid Rotifers. *PLOS Genetics*, **5**, e1000401.

48. Zhao,H., Wu,H., Guseman,A., Abeykoon,D., Camara,C.M., Dalal,Y., Fushman,D. and Papoian,G.A. (2022) The Molecular and Evolutionary Principles of Histone Folding in Eukarya and Archaea. *bioRxiv*, 10.1101/2022.10.15.512373.

49. Laurens,N., Driessen,R.P.C., Heller,I., Vorselen,D., Noom,M.C., Hol,F.J.H., White,M.F., Dame,R.T. and Wuite,G.J.L. (2012) Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat Commun*, **3**, 2330.

50. Jelinska,C., Conroy,M.J., Craven,C.J., Hounslow,A.M., Bullough,P.A., Waltho,J.P., Taylor,G.L. and White,M.F. (2005) Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA Packaging. *Structure*, **13**, 963–971.

51. Tanaka,T., Padavattan,S. and Kumarevel,T. (2012) Crystal structure of archaeal chromatin protein alba2-double-stranded DNA complex from Aeropyrum pernix K1. *Journal of Biological Chemistry*, **287**, 10394–10402.

52. Wardleworth,B.N., Russell,R.J.M., Bell,S.D., Taylor,G.L. and White,M.F. (2002) Structure of Alba: An archaeal chromatin protein modulated by acetylation. *EMBO Journal*, **21**, 4654–4662.

53. Xue,H., Guo,R., Wen,Y., Liu,D. and 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**, 3929–3933.

54. Liu,Y., Guo,L., Guo,R., Wong,R.L., Hernandez,H., Hu,J., Chu,Y., Amster,I.J., Whitman,W.B. and Huang,L. (2009) The Sac10b homolog in Methanococcus maripaludis binds DNA at specific sites. *Journal of bacteriology*, **191**, 2315–2329.

Supplementary figures

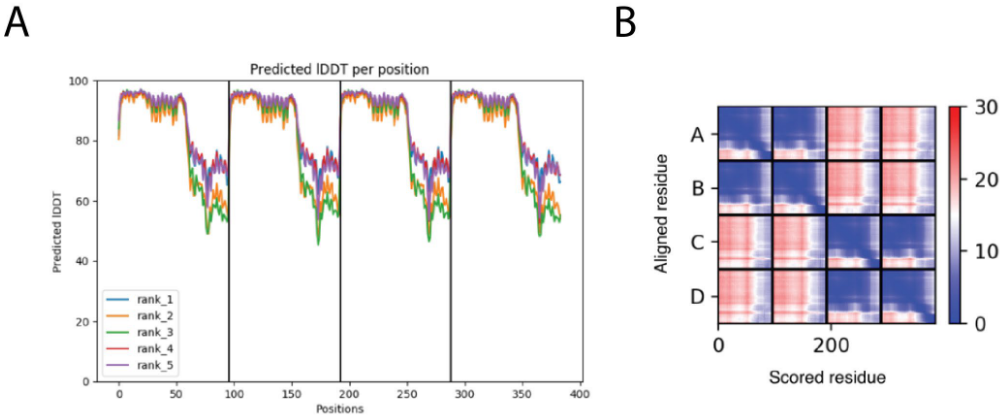


Figure S5.1 A) Predicted local Distance Difference Test (pIDDT) for the five MJ1647 tetramer models produced by AlphaFold2**.** B) Predicted alignment error (PAE) for the rank 1 MJ1647 tetramer model produced by AlphaFold2. A to D denote the four amino acid chains.

Figure S5.2 K80 and E95 are conserved residues in MJ1647 homologues. Multiple sequence alignment was performed with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) using Uniprot entries that have 50% sequence identity to MJ1647 (Uniprot Reference Cluster: UniRef50_Q59041). Sequences are listed by reviewed (sp) or unreviewed (tr) status on Uniprot, followed by the Uniprot accession code and entry name. Residues K80 and E95 in MJ1647 (Q59041) are highlighted in red.

```
sp|Q03576|HMVA_METVO            ------------MIPKGTVKRIMKDNTEMYVSTESVVALVDILQEMIVTTTKIAEENAA
sp|Q59041|HMVA_METJA            ------------MLPKATVKRIMKQHTDFNISAEAVDELCNMLEEIIKITTEVAEQNAR
tr|A0A076LDI7|A0A076LDI7_9EURY  ----------------------MKQHTNFNISAEAVDELCNMLEEIIKITTEVAEQNAR
tr|A0A2L1CD04|A0A2L1CD04_METMI  ------------MIPKGTVKRIMKENTDMNVSAESVVALVEILQEMVVTTTKIAEENAA
tr|A0A2Z5PDL5|A0A2Z5PDL5_METMI  ------------MIPKGTVKRIMKENTDMNVSAESVVALVEILQEMVVTTTKIAEENAA
tr|A0A2Z5PL52|A0A2Z5PL52_METMI  ------------MIPKGTVKRIMKENTDMNVSAESVVALVEILQEMVVTTTKIAEENAA
tr|A0A7J9PFZ2|A0A7J9PFZ2_METMI  ------------MIPKGTIKRIMKENTDMNVSAESVAALVEILQEMVVTTTKIAEENAE
tr|A0A7J9S218|A0A7J9S218_METMI  ------------MIPKGTVKRIMKENTDLNISSEAVDELCNMLEEIIKITTEVAEQNAA
tr|A0A832SV22|A0A832SV22_9EURY  ------------MLPKATVKRIMKQHTDFNISAEAVDELCNMLEEIIKITTEVAEQNAR
tr|A0A8D6SUC3|A0A8D6SUC3_9EURY  ------------MLPKTTIKRIMKQYTDFNISSEAVDELNNLLMEIIKITTEVAEQNAK
tr|A0A8D6T0P6|A0A8D6T0P6_9EURY  ------------MLPKTTIKRIMKQYTDFNISSEAVDELSNLLMEIIKITTEVAEQNAK
tr|A0A8J7S442|A0A8J7S442_METVO  ------------MIPKGTVKRIMKDNTEMYVSTESVVALVDILQEMIVTTTKIAEENAA
tr|A0A8J8C7G0|A0A8J8C7G0_9EURY  ------------MIPKGTIKRIMKENTNLNISAESVEKLVDILQEYIVTTTKLAEENAI
tr|A4FXG4|A4FXG4_METM5          ------------MIPKGTVKRIMKENTEMNVSAESVAALVEILQEMVVTTTKIAEENAA
tr|A6UP30|A6UP30_METVS          ------------MIPKGTVKRIMKQHTEMNVSAESVEKLVELLQEIIVTTTQIAEQNAG
tr|A6UUU1|A6UUU1_META3          ------------MIPKGTVKRIMKQNTDMNVSAESVVKIVEILQEYIVTTTRLAEENAA
tr|A6VFV1|A6VFV1_METM7          ------------MIPKGTIKRIMKENTDMNVSAESVAALVEILQEMVVTTTKIAEENAE
tr|A9AAT3|A9AAT3_METM6          ------------MIPKGTVKRIMKENTDMNVSAESVAALVEILQEMVVTTTKIAEENAE
tr|C7P6C0|C7P6C0_METFA          ------------MLPKATVKRLMKQYTDFNISAEAVDELCNMLEEIIKITTEVAEQNAR
tr|C9REV3|C9REV3_METVM          ------------MLPKATIKRIMKEHTDFNISSEAVDELCNMLEEIIKITTEVAEQNAR
tr|D3S432|D3S432_METSF          ------------MLPKATVKRIMKQYTNFNISAEAVDELCNMLEEIIKITTEVAEQNAR
tr|D5VTW4|D5VTW4_METIM          ------------MLPKTTIKRVMKNYTDLNISSEAVDELINLLEEMIKVTTEVAEKNAK
tr|D7DTQ7|D7DTQ7_METV3          ------------MIPKGTVKRIMKENTDMYVSTESVVALVDILQEMIITTTRIAEENAA
tr|F6BAU5|F6BAU5_METIK          ------------MIPKGTIKRIMKKHTDMNISSEAVEELSNILEEIIVITTKTAEENAR
tr|F8AK59|F8AK59_METOI          ------------MIPKGTVKRLMKENTDMNISAESVEKLVEILQEYIVTTTKMAEDNAR
tr|G0H0G5|G0H0G5_METMI          ------------MIPKGTVKRIMKENTDMNVSAESVVALVEILQEMVVTTTKIAEENAA
tr|H1KZH1|H1KZH1_9EURY          ------------MIPKGTIKRIMKKHTDLNISSEAVEELANILEEIIAITTKTAEENAK
tr|N6VQ57|N6VQ57_9EURY          ------------MLPKTRIKKIMKQYTKLNISSEAVDELNNILLEIIKIITETAEKNAK
tr|Q6LYH6|Q6LYH6_METMP          MPYDSKDKIRGLIMIPKGTVKRIMKENTDMNVSAESVVALVEILQEMVVTTTKIAEENAA
                                                **. *.: :*:*:*  : ::* * :   *. **.**

sp|Q03576|HMVA_METVO            KDKRKTIKARDIEECDAERLKEKILQVSERTEKVNMLANEILHVIASELERY-      99
sp|Q59041|HMVA_METJA            KEGRKTIKARDIKQCDDERLKRKIMELSERTDKMPILIKEMLNVITSEL----      96
tr|A0A076LDI7|A0A076LDI7_9EURY  KEGRKTIKARDIKQCDDERLKRKIIELSERTDKMPILIKEMLNVITSELK---      87
tr|A0A2L1CD04|A0A2L1CD04_METMI  KDKRKTLKARDIEQCDAERLRKKVIEVSERTEKVNMLTNEILNVIANELERY-      99
tr|A0A2Z5PDL5|A0A2Z5PDL5_METMI  KDKRKTLKARDIEQCDAERLRKKVIEVSERTEKVNMLTNEILNVIANELERY-      99
tr|A0A2Z5PL52|A0A2Z5PL52_METMI  KDKRKTLKARDIEQCDAERLRKKVIEVSERTEKVNMLTNEILNVIANELERY-      99
tr|A0A7J9PFZ2|A0A7J9PFZ2_METMI  KDKRKTLKARDIEQCDAERLRKKVVEVSERTEKVNMLTNEILNVIANELERY-      99
tr|A0A7J9S218|A0A7J9S218_METMI  NDKRKTLKARDIEQCDAERLRKKVIEVSERTEKVNMLTNEILNVIANELERY-      99
tr|A0A832SV22|A0A832SV22_9EURY  KEGRKTIKARDIKQCDDERLKRKIMELSERTDKMPILIKEMLNVITSEL----      96
tr|A0A8D6SUC3|A0A8D6SUC3_9EURY  KDGRKTIKARDIRNCDDERLKRKIIELSERTDKMPILIKEMLNVITSELE---      97
tr|A0A8D6T0P6|A0A8D6T0P6_9EURY  KDGRKTIKAKDIRNCDDERLKRKIIELSERTDKMPILIKEMLNVITSELE---      97
tr|A0A8J7S442|A0A8J7S442_METVO  KDKRKTIKARDIEECDAERLKEKILQVSERTEKVNMLANEILHVIASELERY-      99
tr|A0A8J8C7G0|A0A8J8C7G0_9EURY  KDKRKTIKGRDIENCDEERLRSKIIEISDRTEKVNILTREFLKVLSSELKRH-      99
tr|A4FXG4|A4FXG4_METM5          NDKRKTLKARDIEQCDAERLRKKVVEVSERTEKVNMLTNEILNVIANELERY-      99
tr|A6UP30|A6UP30_METVS          KDKRKTLKARDIEQCDAERLRRKIVEVSERTEKVNILTNEILNVVANELERY-      99
tr|A6UUU1|A6UUU1_META3          KDKRKTIKARDVENCDGERVRQKILEVADRTEKVQILTKEFLKVLSSELTREE      100
tr|A6VFV1|A6VFV1_METM7          KDKRKTLKARDIEQCDAERLRKKVVEVSERTEKVNMLTNEILNVIANELERY-      99
tr|A9AAT3|A9AAT3_METM6          KDKRKTLKARDIEQCDAERLRKKVVEVSERTEKVNMLTNEILNVIANELERY-      99
tr|C7P6C0|C7P6C0_METFA          KDGRKTIKARDIKMCDDERLKRKIMELSERTDKMPILVKEMLNVITSELE---      97
tr|C9REV3|C9REV3_METVM          KEGRKTIKARDIKNCDDERLKRRIMELSERTDKMPILIKEMLNVITSELK---      97
tr|D3S432|D3S432_METSF          KEGRKTIKARDIKQCDDERLKRRIMELSERTDKMPILIKEMLNVITSELE---      97
tr|D5VTW4|D5VTW4_METIM          REGRKTILRRDIKNCDEERLKRKILELSERTDKMPIIVKEILAIITSELE---      97
tr|D7DTQ7|D7DTQ7_METV3          KDKRKTIKARDIEECDAERLKEKILQVSERTEKVNMLANEILHVIASELERY-      99
tr|F6BAU5|F6BAU5_METIK          ADNRKTIKARDIKKCDKERIREKIIELANRTEKMNILTREFLNVISSELE---      97
tr|F8AK59|F8AK59_METOI          KDKRKTIKARDIENCDEERLRVKIMEIADRTEKVNILTKEFLKVLASELLRE-      99
tr|G0H0G5|G0H0G5_METMI          KDKRKTLKARDIEQCDAERLRKKVIEVSERTEKVNMLTNEILNVIANELERY-      99
tr|H1KZH1|H1KZH1_9EURY          AENRKTIKARDIKKCDEERLREKIIELANRTEKMNILTREFLNVISSELE---      97
tr|N6VQ57|N6VQ57_9EURY          REGRKTIKGRDIKECDDERLKRKIIELSKRTDKMPILIKEILNVITSELE---      97
tr|Q6LYH6|Q6LYH6_METMP          KDKRKTLKARDIEQCDAERLRKKVIEVSERTEKVNMLTNEILNVIANELERY-      112
                                 : ***:  :*:. ** **:: ::::::.**:*: :: .*:* :::.**
```

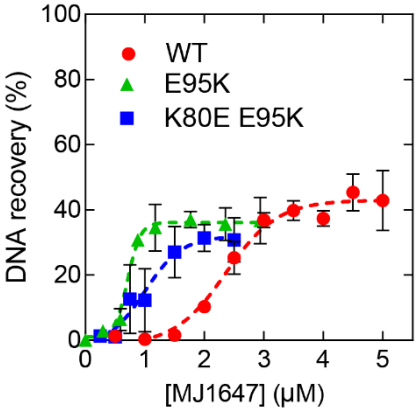


Figure S5.3 MJ1647 E95K and K80E E95K maintain DNA bridging behavior. DNA recovery (%) as a function of MJ1647 WT, E95K or K80E E95K concentration. MJ1647 WT data was reproduced from figure 5.2 for easy comparison. Data are plotted as mean values of three independent measurements and error bars represent the standard deviation. Dashed lines serve as lines to guide the eye.

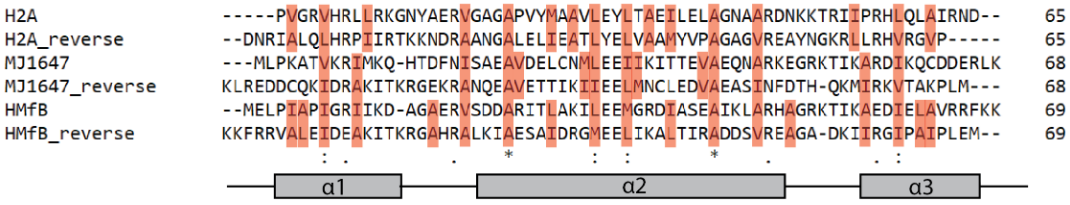


Figure S5.4 Sequence alignment reveals a symmetrical distribution of hydrophobic interactions. Sequences of the histone folds of H2A (NCBI code: NP_003507.1 residues 26-91), MJ1647 (AAB99668.1, residues 1-68) and HMfB (ADP77985.1, full seuqence) and their reversed sequences were aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) using default parameters discouraging gaps (Gap open penalty = 100). Aligned hydrophobic residues between the forward and reversed sequences are highlighted in red. The three alpha helices of the histone fold are indicated as grey bars.

Chapter 6

Histones from Heimdallarchaeota: on the way to eukaryotic nucleosomes

Part of the experiments described in this chapter were performed by Willem Marulanda Valencia, Samuel Schwab and Shane R. Dulfer under supervision of Amanda M. Erkelens, Aimee L. Boyle and Remus T. Dame.

Abstract

Eukaryogenesis is considered one of the main evolutionary events in the history of life. Asgard archaea are considered the closest relatives to eukaryotes today and are likely candidates to have provided the host cell that merged with an α-proteobacterium. Archaeal histones share the histone fold with their eukaryotic counterparts but generally lack the tails. Also, archaeal histones are not limited to forming octameric nucleosomes, as found in eukaryotes, but they wrap DNA into hypernucleosomes of various lengths. *Ca. Heimdallarchaeota* LC_3 encodes multiple histones of which one has an N-terminal tail similar to eukaryotic histones. Expression of recombinant histones HA and HB from *Ca. Heimdallarchaeota* LC_3 is challenging. Therefore, we here examined alternative methods to obtain these proteins. We were able to synthesize HB. HB folds correctly and can bind DNA. However, unlike earlier predictions, HB most likely does not form a hypernucleosome in the prototypical HMfB-like fashion. Instead, we propose that HB mediates formation of higher-order structures only by dimer-dimer interactions without additional interactions between the stacks, resulting in a more open kind of hypernucleosome.

Introduction

The origin of the eukaryotic cell has been described as a series of endosymbiotic events where a host cell engulfed bacterial partners. Several eukaryotic organelles, such as mitochondria and chloroplasts, have a bacterial origin in the α-proteobacteria and cyanobacteria, respectively (1–3). The nature of the host cell has so far been less clear. According to several eukaryogenesis models, the host cell that acquired mitochondria was an archaeon (4). The discovery of Asgard archaea near Loki's castle supported this as phylogenetic analyses show that eukaryotes are a sister group of the Asgard archaea (5–8). Most members of the Asgard archaea encode multiple eukaryotic signature proteins (ESPs), such as the ubiquitin-coupled ESCRT system and homologues to eukaryotic actin to form a cytoskeleton (5–7, 9–11). Only two Asgard archaea have successfully been cultured in a laboratory so far (*Ca. Prometheoarchaeum syntrophicum* and *Ca. Lokiarchaeum ossiferum*, both Lokiarchaeota) (12, 13). They are anaerobic, slow-growing organisms that metabolize amino acids and grow in syntropy with hydrogen-consuming and sulfate-reducing bacteria. Their morphology showed long-branched protrusions created by their cytoskeleton. This, together with their requirement of co-culture with several bacteria, strengthens the hypothesis that an Asgard archaeon was able to entangle, engulf and endogenize (known as the E3-model) one of its syntrophic partners to create the first eukaryotic common ancestor (FECA).

Eukaryotes organize their genome using histone proteins that assemble on DNA into nucleosomes. Nucleosomes consist of a histone $(H3\text{-}H4)_2$ tetramer together with two histone H2A-H2B dimers forming an octameric protein core with ~147 bp DNA around it (14). Archaea express homologues of eukaryotic histones that share the histone fold, but lack the tails of eukaryotic histones (15). Also, archaeal histones are not limited to forming an octameric structure; they can form hypernucleosomes, rod-like structures in which histone dimers can, in theory endlessly, stack upon each other with DNA wrapped around it (16, 17). There are several elements in the formation of a hypernucleosome (15) (figure 6.1). First, dimer-dimer interactions promote side-by-side association of dimers along DNA. Second, stacking interactions to form stable interactions between dimer *i* and dimers *i*+2 and *i*+3 can be used to further stabilize the hypernucleosome. Additionally, absence of bulky residues at certain positions and the possibility of hydrogen bonds increases the likelihood of hypernucleosome formation. Asgard archaea mainly encode histones that, based on homology modelling, fulfill these requirements, but clear exceptions have been found. For example, histone HLkE from *Ca.*

Lokiarchaeum sp. GC14_75 lacks the dimer-dimer interface and has a truncated C-terminus (15). *Ca. Heimdallarchaeota* LC_3 encodes 10 different histone proteins, of which histone A (HA) contains an N-terminal tail akin to those found in eukaryotic histones (figure 6.1). A structural model of HA shows that the tail protrudes similarly through the DNA and is a likely candidate for post-translational modifications (PTMs) due to the presence of several lysine residues (15). Next to these eukaryotic features, HA does contain residues that could mediate stacking interactions yielding hypernucleosomes similar to those formed by HMfB (figure 6.1). *Ca. Heimdallarchaeota* LC_3 also encodes prototypical archaeal-like histones, such as histone B (HB) and histone C (HC), which lack tails and consist of a histone-fold only, that are predicted to form hypernucleosomes.

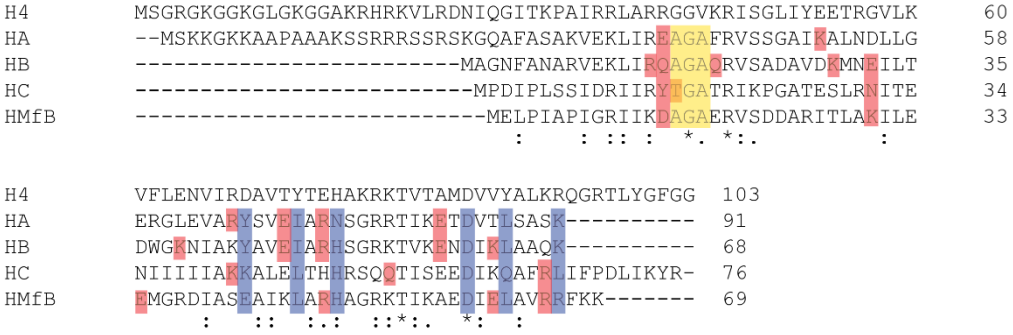


Figure 6.1 Alignment of human histone H4 with archaeal histones HA, HB, HC (*Ca. Heimdallarchaeota* LC_3) and HMfB (*Methanothermus fervidus*) For the archaeal histones, the dimer-dimer interface is indicated in blue, the first loop in yellow and the predicted stacking interactions in red (15). NCBI accession codes: NP_ 003486.1 (H4), OLS22332.1 (HA), OLS24873.1 (HB), OLS21974.1 (HC) and ADP77985.1 (HMfB).

Here, we aimed to investigate the DNA binding characteristics of HA and HB from *Ca. Heimdallarchaeota* LC_3. As attempts for recombinant expression failed, we turned to Solid Phase Peptide Synthesis (SPSS). We were able to synthesize HB successfully and show that, although it is folded and binds DNA, HB does not form a hypernucleosome as tightly packed as HMfB, but forms a more open hypernucleosome only via its dimer-dimer interactions not stabilized by stacking interactions.

Materials and methods

Plasmid construction

Plasmids for protein expression were constructed using the coding sequences for HA (NCBI protein database: OLS22332.1) and HB (NCBI protein database: OLS24873.1). These sequences were separately codon optimized for expression in *Escherichia coli* and *Pichia pastoris* using GeneArt (Thermo Fischer Scientific). The optimized sequences for *E. coli* were cloned into pET30b using Gibson assembly (18) resulting in plasmids pRD317 (HA) and pRD318 (HB). For *P. pastoris*, the optimized sequences were cloned into pPICZ A from the EasySelect™ *Pichia* Expression kit (Invitrogen) using restriction digestion with EcoRI and ApaI (Thermo Fisher Scientific) resulting in plasmids pRD469 (HA) and pRD470 (HB). The sequence of all constructs was verified by DNA Sanger sequencing (BaseClear).

Protein expression in E. coli

pRD317 and pRD318 were used to transform several *E. coli* expression strains: BL21(DE3), BL21(DE3) pLysS, BL21(DE3) Rosetta, BL21(DE3) Star and OverExpress™ C43(DE3) pLysS (Sigma-Aldrich). The cells were grown at 37°C, 200 rpm in Lysogeny Broth (LB) medium until an OD600 of 0.6. When appropriate, 50 µg/ml kanamycin and 25 µg/ml chloramphenicol were present in the medium. Expression was induced with 1 mM IPTG and samples were taken for SDS-PAGE analysis at 0, 1, 3 and 20 hours after induction.

Protein expression in P. pastoris

Details regarding the used protocols and medium recipes can be found in the EasySelect™ *Pichia* Expression kit manual (Invitrogen). In short, plasmids pRD469 and pRD470 were linearized using SacI (Thermo Fisher Scientific) digestion at 37°C overnight, followed by heat inactivation at 65°C for 20 minutes. The linearized plasmids were used to transform chemically competent *P. pastoris* GS115 cells. The cells were plated on Super Optimal Broth with Catabolite repression (SOC) medium with 1% bacteriological agar (Oxoid) and 75 µg/ml Zeocin™ (Invitrogen) and incubated at 30°C. Transformants were checked for their methanol utilization (Mut) phenotype by comparing growth on minimal dextrose medium + 0.004% histidine and minimal methanol medium + 0.004% histidine. Genomic inserts were amplified using PCR for direct screening of *Pichia* clones and the sequences were verified by DNA Sanger sequencing (BaseClear).

Small-scale expression tests were performed as described in the manual. All minimal media were supplemented with 0.004% histidine. In short, cells were grown in

either minimal glycerol medium (MGH), buffered minimal glycerol medium (BMGH) or buffered glycerol-complex medium (BMGY) at 30°C 250 rpm overnight. The cells were harvested and resuspended in either minimal methanol medium (MMH), buffered minimal methanol medium (BMMH) or buffered methanol-complex medium (BMMY) to induce expression. Every 24 hours, expression was maintained by the addition of 0.5% methanol. Samples for SDS-PAGE analysis were collected at 0, 1, 2, 3 and 4 days after induction.

SDS-PAGE analysis of protein expression

The samples taken during protein expression were normalized to equal OD600 and the cells were spun down for 1 minute at maximum speed. For *P. pastoris,* the pellet was resuspended in 200 µl 0.1M NaOH and spun down for 5 minutes at maximum speed. The pellet was resuspended in 100 µl cracking buffer (50 mM Tris-HCl pH 6.8, 1% SDS, 25% glycerol, 1% β-mercaptoethanol, 0.05% bromophenol blue). For *E. coli,* the cell pellet was directly resuspended in 50 µl cracking buffer. 5 µl was loaded on a 4-15% mini-PROTEAN® TGX™ Precast protein gel (Bio-Rad) and run at constant voltage (200V) for 30 minutes. After the run, the gel was stained with 0.1% Coomassie Brilliant Blue in destain solution containing 40% ethanol and 10% acetic acid.

In vitro translation

Plasmids pRD317, pRD318 and the NEBExpress Control DHFR-His Plasmid were used with the NEBExpress® Cell-free *E. coli* Protein Synthesis System (NEB) according to the manufacturer's protocol and the reactions were analyzed using SDS-PAGE.

Solid Phase Peptide Synthesis

Peptide synthesis was performed using the microwave-assisted peptide synthesizer Liberty Blue (CEM) at a 0.1 mmol scale using 5-fold excess of reagents. These reagents include 0.2 M amino acids, 1M N,N-diispropylcarbodiimide (DIC), 1M oxyma and 20% piperidine, all in dimethylformamide (DMF). Syntheses were done on an Fmoc-Lys-Wang resin (Sigma-Aldrich). During synthesis, all arginines were double-coupled. After synthesis, the peptides were cleaved from the resin and the protecting groups were removed from the amino acid side chains with a trifluoracetic acid (TFA)-H_2O- triisopropyl silane (TIPS) (95:2.5:2.5) mix for 3 hours on a rocker. The peptides were precipitated in ice-cold diethyl ether and redissolved in water and acetonitrile or 50 mM Tris-HCl pH 7, 75mM KCl buffer. To increase solubility of the peptides, they were incubated in a 37°C-45°C water bath for 30-60 minutes. Peptides were purified using

Reverse-Phase HPLC (RP-HPLC), ion exchange chromatography (HiTrap SP HP column, Cytiva) or affinity chromatography (HiTrap Heparin HP column, Cytiva). RP-HPLC purified protein was lyophilized and stored at -20°C. Proteins purified with ion exchange or affinity chromatography were stored at -80°C.

CD spectroscopy

CD spectra were taken with a Jasco J-815 CD spectrometer with 10 μM protein in 50 mM Tris-HCl pH 7, 75 mM KCl at room temperature. Scans were taken from 260 to 200 nm with a bandwidth of 1 nm and the reported spectra are an average of 5 accumulations. Secondary structure content was determined using BeStSel (19).

DNA substrate preparation

Tethered particle motion experiments were performed using a 47% GC 685 bp DNA substrate described earlier (17) unless otherwise stated. The same DNA substrate was used for electromobility shift assays. The DNA substrate was generated by PCR using Thermo Scientific® Phusion® High-Fidelity DNA Polymerase and the products were purified using the GenElute PCR Clean-up kit (Sigma-Aldrich).

Electromobility shift assay

The 685 bp DNA substrate was incubated with protein for 45 minutes at room temperature in 50 mM Tris HCl pH 7, 75 mM KCl. The final concentration of DNA was kept constant at 42 ng/μl. After incubation, 6X DNA loading dye (Thermo Scientific) was added and the samples were loaded on a 4-15% mini-PROTEAN® TGX™ Precast protein gel (Bio-Rad). The gels were run in native PAGE buffer (25 mM Tris HCl pH 8.3, 192 mM glycerine) for 20-24 hours at 4°C and 25-45V. The gels were stained with GelRed to detect the DNA, followed by 0.1% Coomassie Brilliant Blue staining to detect the protein.

Tethered particle motion

Tethered particle motion experiments were performed as previously described (20). Flow cells were washed with experimental buffer (50 mM Tris-HCl, pH 7, 75 mM KCl) once and 100 μl protein in experimental buffer was flowed in. After 10 minutes incubation at RT, the cells were washed once with protein dilution and sealed with nail polish. The cells were transferred to the instrument and left to stabilize at 25°C for 5 minutes. For each flow cell, more than 200 beads were measured, and experiments were done at least in duplicate for each protein concentration. To select single-tethered beads,

a standard deviation cut-off of 8% and an anisotropic ratio cut-off of 1.3 were used. The resulting histograms were fitted with a Gaussian distribution.

For end-to-end distance analysis, 25 beads close to the fitted RMS were chosen and the 2.5% most distant points of each bead were collected. Using triangular calculations and the bead diameter of 0.44 µm, the end-to-end distance was calculated for each point. The data points were plotted as histograms and fitted with a skewed Gaussian distribution. The difference between the two populations was calculated using a pairwise distance distribution and fitted with a Gaussian distribution.

Results

Recombinant and in vitro production of HA and HB was unsuccessful

To be able to study the properties of the *Ca. Heimdallarchaeota* LC_3 histones, *E. coli* expression plasmids were created with a codon-optimized sequence of HA or HB behind the high-expressing T7 promotor. Protein expression was attempted in various *E. coli* expression strains in the presence of 1 mM IPTG, but no protein expression could be detected (figure S6.1).

Next, we investigated whether HA and HB could be recombinantly produced in the eukaryotic expression system *P. pastoris*. Strains with the yeast codon-optimized HA or HB inserted in the genome behind the methanol-inducible AOX-1 promoter were created and protein expression was attempted in both minimal and complex growth medium. However, also *P. pastoris* was unable to express the histone proteins (figure S6.2).

One possible explanation for the difficult recombinant expression of proteins is cell toxicity. This can be circumvented by the use of an *in vitro* translation system. An *E. coli*-extract for coupled transcription/translation was used with the *E. coli* expression plasmids described above. This approach also yielded no detectable protein (figure S6.3).

Solid-phase peptide synthesis results in well-folded histone proteins

Generally, histones are small: HA is 91 and HB is 68 amino acids. Also, the histone fold is a relatively easy fold, consisting of three α-helices connected by two loops. Therefore, solid phase peptide synthesis (SPPS) is a potentially suitable method to create these proteins. SPSS was used before to obtain modified eukaryotic histones (reviewed in (21)). For HB this approach was successful and protein of the right mass was synthesized (figure S6.4A). However, HA turned out to be too long to reliably synthesize (figure S6.4B). We also synthesized HMfB from *Methanothermus fervidus* to

be able to compare a synthesized protein with its recombinantly produced version. While this was overall successful, some shorter products were present after purification (figure S6.4C). The folding of synthesized HB and HMfB was analyzed using CD spectroscopy and the spectra obtained were compared with that of recombinant HMfB (figure 6.2). All spectra are typical of a protein with high α-helical content with recombinant HMfB having the highest percentage α-helix (~80%), while this is slightly lower for HB and synthesized HMfB (~75%). This means that the synthesized proteins were folded correctly.

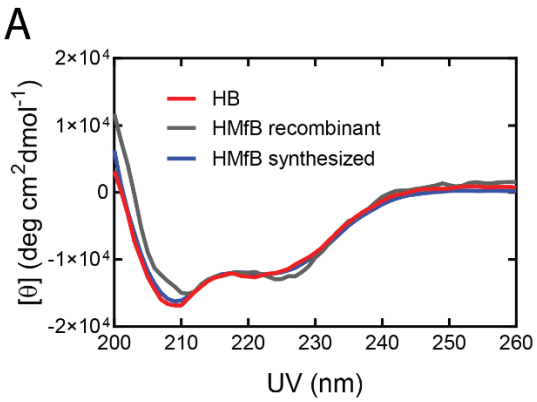


Figure 6.2 Synthesized HB and HMfB are folded CD spectra of synthesized HB and both recombinant and synthesized HMfB. The spectra were normalized at 222 nm.

HB compacts DNA by forming a loosely packed kind of hypernucleosome

To investigate the DNA binding properties of synthesized HB, we performed Tethered Particle Motion (TPM) experiments. We observed a gradual decrease in Root Mean Square deviation (RMS) upon titration with HB (figure 6.3A), indicating DNA compaction. The RMS value under saturating conditions (600 nM) is ~97 nm, which is higher than the ~70 nm found for HMfB (17) and the ~80 nm for A3 from *Methanococcus jannaschii* ((22), Chapter 5). DNA compaction occurred more gradually and over a larger concentration range for HB than for HMfB and A3, where hypernucleosome formation was observed at 50 and 110 nM respectively. This suggests lower cooperativity, or no cooperativity at all, for HB DNA binding. We calculated the end-to-end distance for bare DNA and DNA bound by 600 nM HB and the pairwise distribution between the two states (figure 6.3B). The difference between bound and unbound DNA is 67 ± 8 nm, corresponding to 197 ± 24 bp with each bp being 0.34 nm. The end-to-end distance under saturating conditions is also much higher than previously found for the HMfB

hypernucleosome (47 ± 7 nm versus 24 ± 5 nm, (17)). It is, therefore, not likely that HB forms a hypernucleosome in the prototypical sense as HMfB. HB dimers could bind evenly spaced along the DNA without any interactions between each other effectively being a DNA-bending protein. But, as dimer-dimer interactions were predicted to be present, it is likely that HB dimers associate side-by-side along the DNA and thus form higher order structures only via those interactions. Without stacking interactions to keep the stacks tightly packed, HB would then yield a more open hypernucleosome structure.

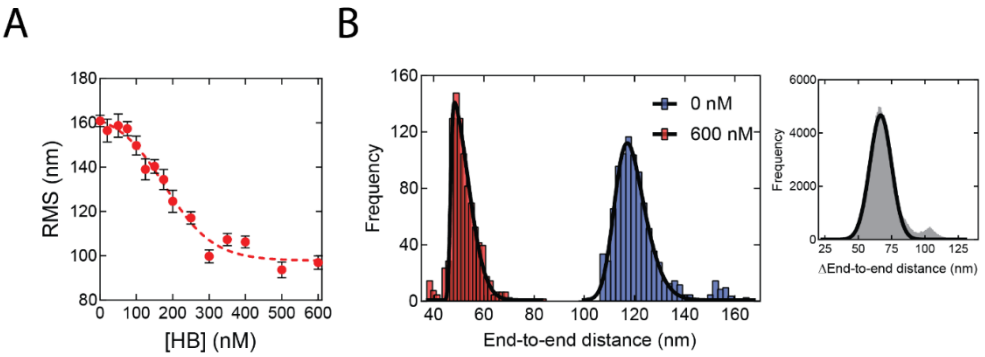


Figure 6.3 Synthesized HB compacts DNA A) Fitted root mean square displacement (RMS) values of DNA tethers incubated with HB in Tethered particle motion (TPM) experiments. Error bars represent the propagated standard deviation of at least two replicates. The dashed line serves as line to guide the eye. B) Calculated end-to-end distances for bare (blue) DNA and bound DNA incubated with 600 nM HB (blue). Histograms were fitted with a skewed normal distribution. Right: pairwise distribution of the difference between unbound and bound DNA. The histogram was fitted with a Gaussian distribution.

Discussion

We were unable to obtain HA and HB via recombinant expression in multiple cellular systems. As we were also unable to obtain protein from *in vitro* translation, cell toxicity is most likely not the main issue (figure S6.3). It could be that the codon-optimized sequences result in a folded mRNA too stable for proper translation. Another possibility is that initial histone production is enough to repress further transcription, resulting in protein amounts that are too low to detect. As archaea are often extremophiles and *Heimdallarchaeota* were discovered near a hydrothermal vent (7), an expression system growing in closer to native conditions might be more suitable. For example, protein expression could be done in the histone-free thermophilic *Sulfolobus sp.* (23).

SPPS was used to successfully synthesize HB from *Ca. Heimdallarchaeota* LC_3 and we showed that it is correctly folded. Therefore, we can conclude that SPPS is suitable for creating functional archaeal histones. An advantage of SPPS is the possibility of incorporating PTMs at specific residues. This could be used in the future to study the effects of modifications on the DNA binding properties of archaeal histones. HA was too long to synthesize completely, and an attempt to synthesize it in two parts was unsuccessful as well, mostly due to impurities in the N-terminal part of the protein (data not shown). Another option to obtain a synthesized HA variant is semi-synthesis, where the N-terminal part is expressed in *E. coli* and the C-terminal part is synthesized, followed by native chemical ligation (24).

Despite it being correctly folded, HB is unlikely to form a hypernucleosome in the sense of its original definition based on HMfB. Instead, our experimental observations favor a model in which HB multimerizes via its dimer-dimer interactions, but the stacks do not interact with each other, yielding a toroidal superhelix with a larger pitch (figure 6.4). This suggests that the stacking interactions predicted for HB in earlier studies do not occur, under the conditions of our experiments (figure 6.1). It shows that stacking interactions provide additional stability to the hypernucleosome, but are not a requirement to form such a higher order structure. Five stacking interactions (E30-K61, Q14-R48, R13-Q18, K27-E57, K37-E45) were originally predicted for HB, based on the homology modelling on the crystal structure of HMfB, for which three stacking interactions were identified (15, 16). It might be, however, that the residues predicted to mediate stacking interactions at the interface between HB dimers are slightly differently positioned in three-dimensional space, not permitting interaction. As HB and HMfB have only 38% sequence identity (and 65% similarity), such differences are not unlikely. Molecular dynamics simulations and cryo-EM structures, as done previously for HTkA from *Thermococcus kodakarensis* (25), might give insight into the (dynamic) structure of HB-DNA complexes. Such an open structure of the hypernucleosome might also be relevant in the context of osmolarity. HTkA showed DNA wrapping of the outer stack perpendicular to the other part of the hypernucleosome with increasing Mg^{2+} concentration (25). These dynamics could therefore be part of controlling hypernucleosome structure by changes in environmental conditions.

Also, as *Ca. Heimdallarchaeota* LC_3 encodes for 10 histones, the effect of heteromerization is most likely substantial and heteromerization might affect the formation of hypernucleosomes. So far, the effects of heteromerization have mainly been investigated *in silico* (26). Models of homo- and heterotetramers showed a wide range of

DNA binding affinities and tetramer stability across archaeal species. Several possible tetramers of *Ca. Heimdallarchaeota* LC_3 were included in these in silico studies, but HA, HB and HC were not. Therefore, it remains to be investigated what their contributions are to DNA binding and how stability of tetrameric complexes and possibly hypernucleosomes is affected.

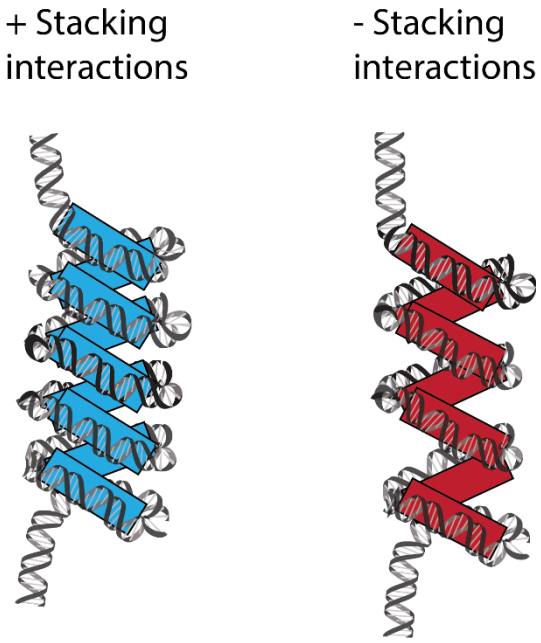


Figure 6.4 Structural modulation of the hypernucleosome via stacking interactions Left: A hypernucleosome with additional stabilizing stacking interactions as formed by HMfB. Right: A more open hypernucleosome where the histone dimers only interact via dimer-dimer interactions creating a larger pitch between the stacks. This could be due to the absence of stacking interactions, as for HB, or to disruption of the stacking interactions by high osmolarity, as was shown for HTkA (25).

References

1. Martin,W. and Kowallik,K. (1999) Annotated english translation of mereschkowsky's 1905 paper 'Über natur und ursprung der chromatophoren impflanzenreiche'. *European Journal of Phycology*, **34**, 287–295.
2. Wallin,I.E. (1922) On the nature of mitochondria. I. Observations on mitochondria staining methods applied to bacteria. II. Reactions of bacteria to chemical treatment. *American Journal of Anatomy*, **30**, 203–229.
3. Gray,M.W., Burger,G. and Lang,B.F. (1999) Mitochondrial evolution. *Science*, **283**, 1476–1481.
4. Mills,D.B., Boyle,R.A., Daines,S.J., Sperling,E.A., Pisani,D., Donoghue,P.C.J. and Lenton,T.M. (2022) Eukaryogenesis and oxygen in Earth history. *Nature Ecology & Evolution 2022 6:5*, **6**, 520–532.
5. Spang,A., Saw,J.H., Jørgensen,S.L., Zaremba-Niedzwiedzka,K., Martijn,J., Lind,A.E., Van Eijk,R., Schleper,C., Guy,L. and Ettema,T.J.G. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, **521**, 173–179.
6. Eme,L., Spang,A., Lombard,J., Stairs,C.W. and Ettema,T.J.G. (2017) Archaea and the origin of eukaryotes. 10.1038/nrmicro.2017.133.
7. Zaremba-Niedzwiedzka,K., Caceres,E.F., Saw,J.H., Bäckström,Di., Juzokaite,L., Vancaester,E., Seitz,K.W., Anantharaman,K., Starnawski,P., Kjeldsen,K.U., *et al.* (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, **541**, 353–358.
8. Williams,T.A., Cox,C.J., Foster,P.G., Szöllősi,G.J. and Embley,T.M. (2019) Phylogenomics provides robust support for a two-domains tree of life. *Nature Ecology & Evolution 2019 4:1*, **4**, 138–147.
9. Akıl,C., Tran,L.T., Orhant-Prioux,M., Baskaran,Y., Manser,E., Blanchoin,L. and Robinson,R.C. (2020) Insights into the evolution of regulated actin dynamics via characterization of primitive gelsolin/cofilin proteins from Asgard archaea. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 19904–19913.
10. Liu,Y., Makarova,K.S., Huang,W.-C., Wolf,Y.I., Nikolskaya,A.N., Zhang,X., Cai,M., Zhang,C.-J., Xu,W., Luo,Z., *et al.* (2021) Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature*, **593**, 553–557.
11. Hatano,T., Palani,S., Papatziamou,D., Salzer,R., Souza,D.P., Tamarit,D., Makwana,M., Potter,A., Haig,A., Xu,W., *et al.* (2022) Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nature Communications 2022 13:1*, **13**, 1–16.

12. Imachi,H., Nobu,M.K., Nakahara,N., Morono,Y., Ogawara,M., Takaki,Y., Takano,Y., Uematsu,K., Ikuta,T., Ito,M., *et al.* (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature*, **577**, 519–525.
13. Rodrigues-Oliveira,T., Wollweber,F., Ponce-Toledo,R.I., Xu,J., Rittmann,S.K.M.R., Klingl,A., Pilhofer,M. and Schleper,C. (2023) Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature*, **613**, 332–339.
14. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251–260.
15. Henneman,B., van Emmerik,C., van Ingen,H. and Dame,R.T. (2018) Structure and function of archaeal histones. *PLoS Genet*, **14**, e1007582.
16. Mattiroli,F., Bhattacharyya,S., Dyer,P.N., White,A.E., Sandman,K., Burkhart,B.W., Byrne,K.R., Lee,T., Ahn,N.G., Santangelo,T.J., *et al.* (2017) Structure of histone-based chromatin in Archaea. *Science (1979)*, **357**, 609–612.
17. Henneman,B., Brouwer,T.B., Erkelens,A.M., Kuijntjes,G.-J., van Emmerik,C., van der Valk,R.A., Timmer,M., Kirolos,N.C.S., van Ingen,H., van Noort,J., *et al.* (2021) Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res*, **49**, 4338–4349.
18. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison 3rd,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343–345.
19. Micsonai,A., Wien,F., Bulyáki,É., Kun,J., Moussong,É., Lee,Y.H., Goto,Y., Réfrégiers,M. and Kardos,J. (2018) BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Research*, **46**, W315–W322.
20. Henneman,B., Heinsman,J., Battjes,J. and Dame,R.T. (2018) Quantitation of DNA-binding affinity using tethered particle motion. In *Methods in Molecular Biology*. Humana Press Inc., Vol. 1837, pp. 257–275.
21. Qi,Y.K., Ai,H.S., Li,Y.M. and Yan,B. (2018) Total Chemical Synthesis of Modified Histones. *Frontiers in Chemistry*, **6**, 19.
22. Ofer,S., Blombach,F., Erkelens,A.M., Barker,D., Schwab,S., Smollett,K., Matelska,D., Fouqueau,T., van der Vis,N., Kent,N.A., *et al.* (2023) DNA-bridging by an archaeal histone variant via a unique tetramerisation interface. *Res Sq*, 10.21203/rs.3.rs-2183355/v1.
23. Berkner,S., Grogan,D., Albers,S.V. and Lipps,G. (2007) Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for Sulfolobus acidocaldarius and Sulfolobus solfataricus, model organisms of the (cren-)archaea. *Nucleic acids research*, **35**.
24. Jing,Y., Ding,D., Tian,G., Jonathan Kwan,K.C., Liu,Z., Ishibashi,T. and Li,X.D. (2020) Semisynthesis of site-specifically succinylated histone reveals that succinylation

regulates nucleosome unwrapping rate and DNA accessibility. *Nucleic Acids Research*, **48**, 9538–9549.

25. Bowerman,S., Wereszczynski,J. and Luger,K. (2021) Archaeal chromatin 'slinkies' are inherently dynamic complexes with deflected DNA wrapping pathways. *Elife*, **10**, e65587.
26. Stevens,K.M., Swadling,J.B., Hocher,A., Bang,C., Gribaldo,S., Schmitz,R.A. and Warnecke,T. (2020) Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proc Natl Acad Sci U S A*, **117**, 33384–33395.

Supplementary figures

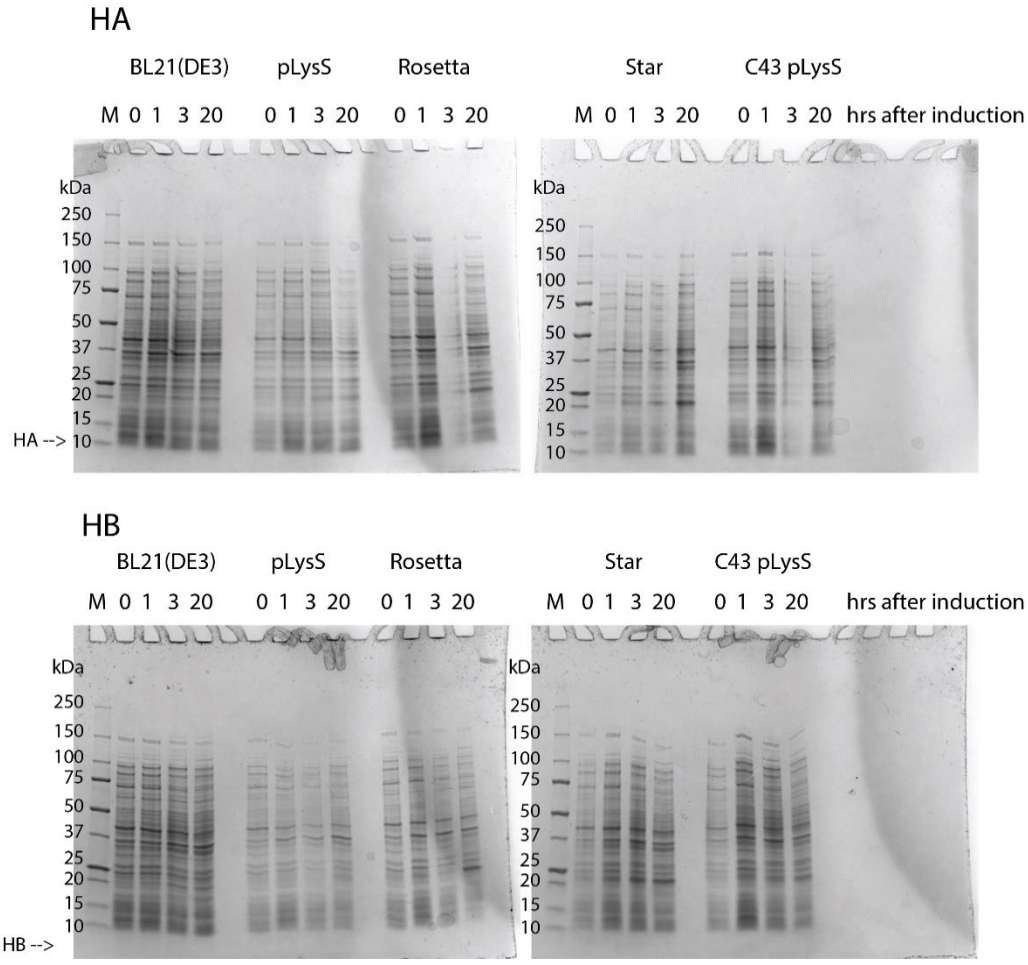


Figure S6.1 *E. coli* is not able to express HA and HB SDS-PAGE analysis of *E. coli* cell lysates collected during expression tests for HA (top) and HB (bottom). The strains used are BL21(DE3), BL21(DE3) pLysS, BL21(DE3) Rosetta, BL21(DE3) Star and OverExpress™ C43(DE3) pLysS. Samples were taken directly upon addition of 1 mM IPTG to the culture (0h) and after 1, 3 and 20 hours. The theoretical mass of HA is 9.7 kDa and HB 7.5 kDa; the expected positions are indicated.

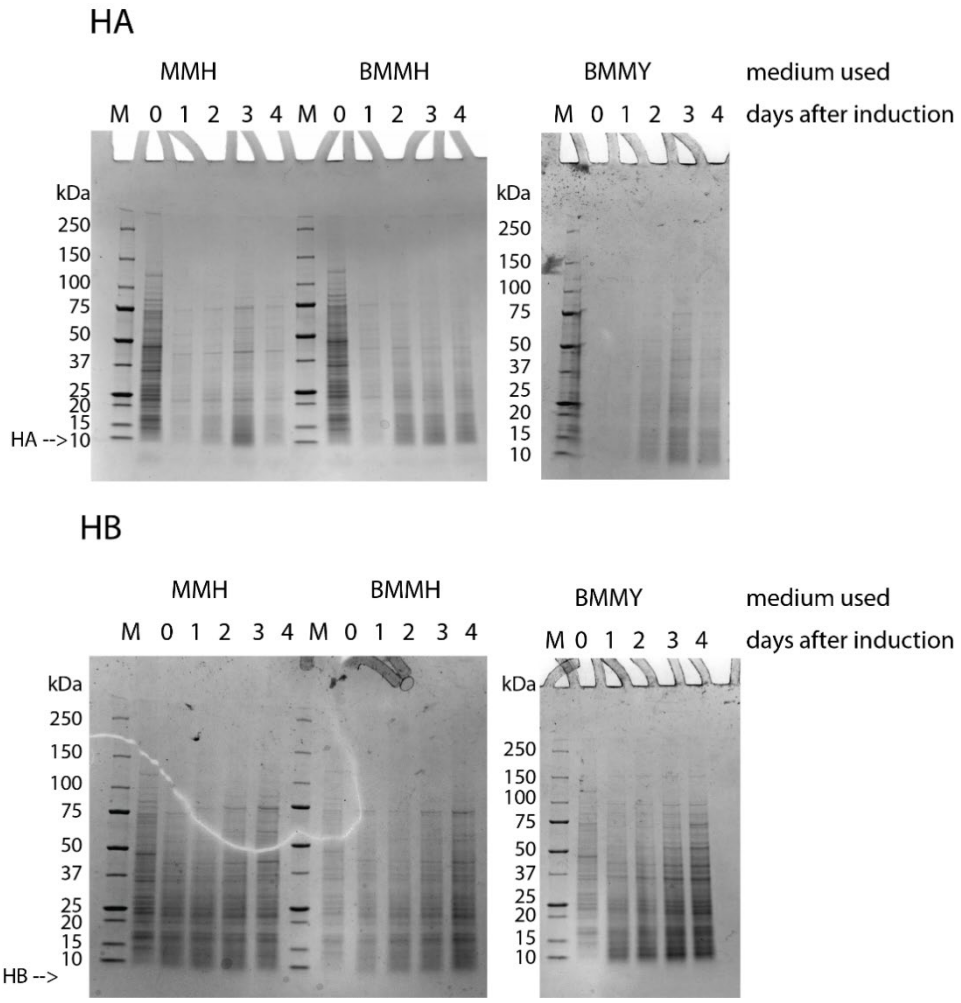


Figure S6.2 *P. pastoris* does not express HA and HB SDS-PAGE analysis of *P. pastoris* GS115 cell lysates collected during expression tests for HA (top) and HB (bottom). *P. pastoris* G115 with HA or HB were grown in MGH, BMGH or BMGY and expression was induced by replacing glycerol with methanol in the medium (MMH, BMMH and BMMY respectively). Cell lysates were collected directly after induction (0 days) after 1, 2, 3 and 4 days. The theoretical mass of HA is 9.7 kDa and HB 7.5 kDa.

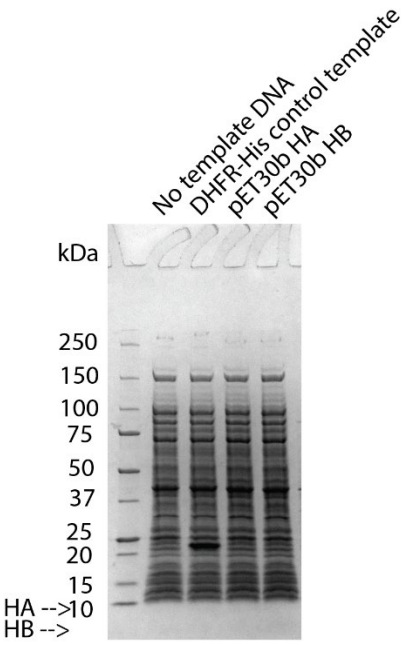
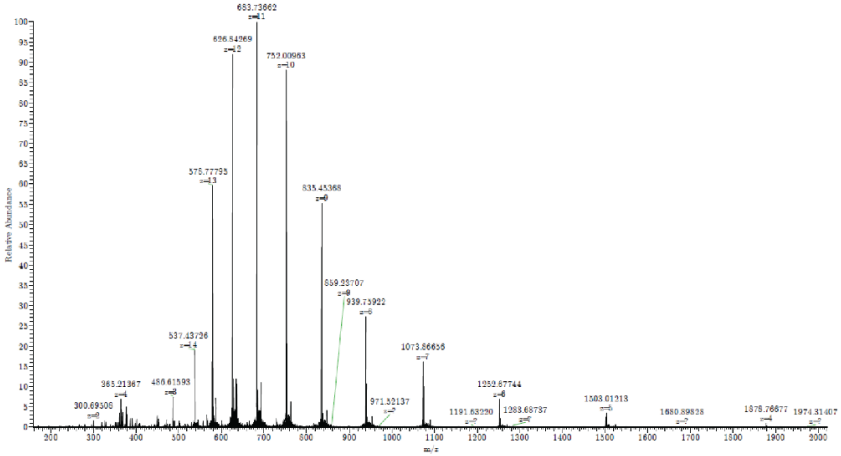
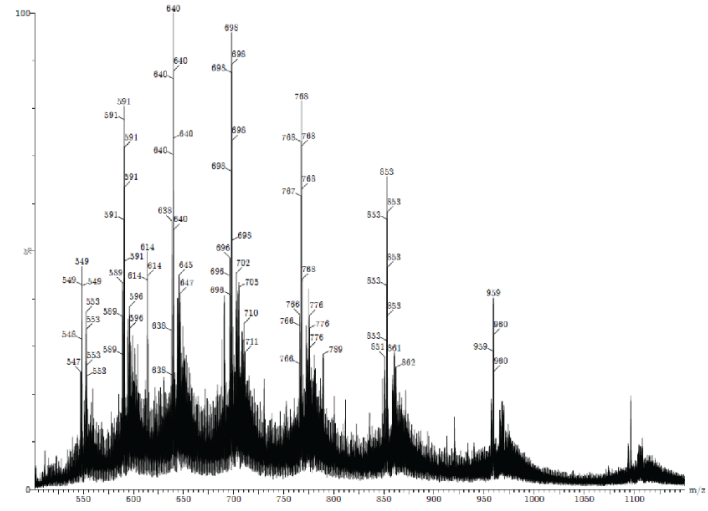


Figure S6.3 Cell-free transcription and translation yielded no protein The NEBExpress® Cell-free *E. coli* Protein Synthesis System was used with plasmids pRD317 (HA) and pRD318 (HB). A reaction without DNA was used as negative control and the included DHFR-His control plasmid was used as positive control.

A


683.73662
z=11
636.54769
z=12
753.00983
z=10
575.77793
z=13
835.45368
z=9
859.25707
z=8
939.73922
z=6
537.43736
z=14
1073.86656
z=7
365.21367
z=4
486.61593
z=8
300.69306
z=3
971.52137
z=5
1252.67744
z=6
1191.63230
z=7
1293.68737
z=6
1503.01313
z=5
1680.89838
z=7
1878.76677
z=4
1974.31407
z=5
Relative Abundance
m/z

B


m/z

C

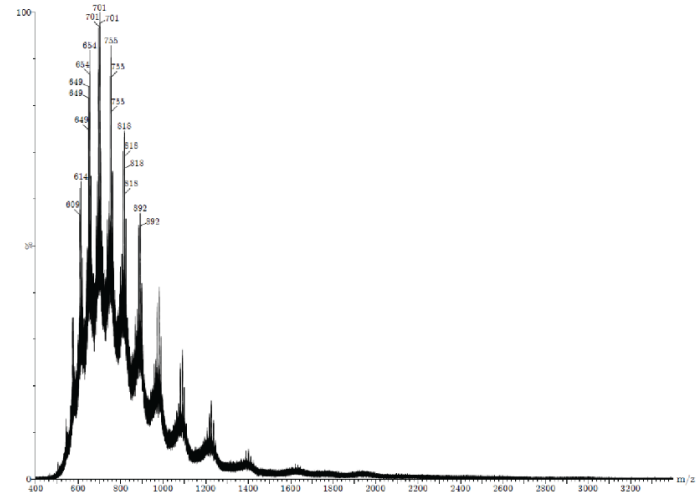

m/z

Figure S6.4 Mass spectrometry results of synthesized histones A) High-resolution mass spectrometry of synthesised HB, puried by ion exchange. Peaks 537.4, 578.8, 626.8, 683.7, 752.0, 835.5, 939.8, 1074, 1253, and 1503 correspond to HB (7511 Da). B) Full protein mass spectrometry of synthesised HMfB, puried by RP-HPLC. Peaks 549, 591, 640, 698, 768, 853, and 959 correspond to HMfB (7667 Da). C) Full protein mass spectrometry of synthesised HA (crude). Peaks 609, 649, 694, 748, 811, and 884 correspond to HA (9721 Da). Peaks 614, 654, 701, 755, 818, and 892 correspond to a delta mass of +80 compared to HA (9801 Da).

Chapter 7

General discussion

Across the tree of life, organisms have evolved ways of organizing their chromosomes and regulating gene expression. Chromatin is dynamically organized and structured as a function of growth cycle and responds to environmental factors and specific physico-chemical cues. The tree of life consists of two major branches: the bacterial and the archaeal/eukaryotic branch (1). Eukaryotes structure their chromosomes locally using nucleosomes consisting of an octameric protein core with DNA wrapped around it (2). Most flexibility in this structure comes from either post-translational modifications (PTMs) on the histone tails or the inclusion of histone variants (3–6). These variants are often involved in specific processes such as DNA damage repair, stress response and spermatogenesis. Defects in these variants or differences in their relative expression levels can result in disease (7). A common mechanism of higher-order DNA organization in eukaryotes is DNA looping by proteins like cohesin and the CCCTC-binding factor (CTCF) (8, 9). Looking back in evolutionary history, eukaryotes are a sister group of the Archaea. Many archaeal enzymes and other proteins are quite similar to their eukaryotic counterparts, but often they contain fewer subunits, or they are a "simplified" version as in the case of RNA polymerase (10). Archaeal histones, for example, lack the tails for PTMs present in eukaryotic histones. Therefore, studying archaeal chromatin and transcription is partly looking back in evolutionary history from a eukaryotic perspective. On the other side of the tree of life we find the domain of Bacteria where nucleoid-associated proteins (NAPs) are the main chromatin organizers. Despite the major differences between and within the branches of the tree of life, architectural chromatin proteins use common mechanisms to organize DNA. The structural and functional properties of the main proteins are introduced in Chapter 1.

NAPs in bacteria

A major global challenge is the rise in antibiotic resistance of pathogenic bacteria such as *Mycobacterium tuberculosis* and methicillin-resistant *Staphylococcus aureus* (MRSA) (11). In this "race of arms" between humans and pathogenic bacteria, it is important to understand the evolutionary mechanisms underlying acquisition of resistance. Horizontally-transferred genes (HGTs) often contain traits that are advantageous in (for the bacteria) challenging situations (12). NAPs, such as the histone-like nucleoid structuring protein (H-NS), contribute to bacterial evolution by binding to and regulating these HGTs. Such binding can be used as a defence mechanism against bacteriophages, but can also help bacteria to build a new, advantageous, gene into a

pre-existing transcriptional network. In Chapter 2, we review the H-NS-like family of proteins and discuss their shared structural and functional characteristics. We also highlight the open questions in this field such as the influence of post-translational modifications (PTMs). Many PTMs of H-NS and other important NAPs in *E. coli* have been identified in proteomics studies (13), but their functional implications are thus far unknown. Another possible layer of regulation of H-NS-like proteins is the expression of paralogs, of which we are only beginning to understand the effect they have on the DNA-binding properties of H-NS-like proteins.

An unusual member of the H-NS-like protein family is the Rok protein from *Bacillus subtilis* which we studied in Chapter 3. We found no indications for nucleoprotein filament formation besides DNA bridging and Rok is unable to complement for the loss of H-NS in *E. coli,* which is different from other H-NS-like proteins such as MvaT and Lsr2. Also its smaller paralog sRok cannot replace H-NS, but can modulate Rok's function. Interestingly, the combined expression of Rok and sRok resulted in a different transcriptional profile than for expression of the individual proteins. *Bacillus spp.* can produce a plethora of antimicrobial compounds relevant to the food industry and as therapeutic agents. As Rok is one of the regulators of these genes (14), understanding the role of these two proteins in transcriptional regulation is an important topic. To be able to benefit from the fundamental knowledge of (s)Rok in a biotechnology setting, we first need to understand how it is regulated. So far, environmental cues in the physiological range have little effect. On the other hand, the fact that Rok shows the same behavior (DNA bridging) over a wide range of environmental conditions, could make it a reliable tool for future applications.

The difference of Rok and sRok compared to other H-NS-like proteins, raises the question whether the initial inclusion of (s)Rok into the H-NS protein family was correct. Also, with the discovery of new potential H-NS-like proteins, such as MucR, GapR and EnrR (15–17), reaching a well-defined description of when a protein is an H-NS-like protein becomes very important. Until now, we have mainly used fold topology, functional properties and the types of protein-DNA complexes formed as common denominators of the protein family (Chapter 2). A protein family is a set of proteins that is evolutionarily related, reflected by related functions or their similarities in sequence or structure. H-NS-like proteins are not related by the strict definition of shared origin but rather via convergent evolution: finding a similar solution to the same evolutionary challenge using proteins unrelated by sequence and evolutionary history. One of these challenges could be horizontally transferred genes and their unregulated transcription. The solution offered by H-NS-like proteins would be xenogeneic silencing. However, this

definition ignores other gene regulatory functions and the genome-organizing properties of H-NS. From a more structural perspective, the presence of an N-terminal oligomerization domain and a C-terminal DNA-binding domain are shared features of the current members of the H-NS-like protein family. Also, they all form DNA-protein-DNA bridges. With the rise of structure prediction programmes like AlphaFold, categorizing proteins based on (predicted) structure becomes easier, although experiments should be performed to verify the predictions. Previously hidden underlying patterns might in this way become visible.

One clear example is the histone protein family. The presence of histone proteins in bacteria has long been a topic of debate. In the 1970's and 1980's, proteins like HU were considered bacterial histones based on similar amino acid content and the ability to form 'nucleosome-like' structures plasmid DNA (18). Later, bacterial NAPs were instead considered functional homologues of histone proteins (such as H-NS being 'Histone-like Nucleoid Structuring protein') and until recently, histones themselves were considered absent in bacteria. However, using fold homology, people have started to note histone sequences in bacterial genomes (19). They lack the tails of eukaryotic histones, just like archaeal histones. Despite some structural difference in their histone fold compared to eukaryotic and archaeal histones, they assemble as dimers. Crystal structures show, however, that these bacterial histones do not wrap the DNA but coat the DNA as a nucleofilament. This is an unexpected deviation from the DNA binding properties of eukaryotic and archaeal histones, but it remains to be investigated if these structures are formed *in vivo* and whether they have any regulatory function. Irrespective of the answers to these questions, it is an important message that histones are more widespread in bacteria than commonly believed, and their (re)discovery might lead to new interesting insights into bacterial chromatin organization.

Hypernucleosome formation – and beyond

In the other prokaryotic branch of the tree of life, the Archaea, histone proteins are widespread (20). Most archaea encode for histone proteins that are able or predicted to form a hypernucleosome, defined as a histone multimer that wraps the DNA, creating a quasi-continuous superhelix (20, 21). Although eukaryotic histones mainly form the characteristic octameric nucleosome core, recently a more compact extended structure was found at the telomers of chromosomes, which strikingly resembles a hypernucleosome (22).

In Chapter 4 we discuss a SELEX-optimized DNA sequence, Clone20, that is bound by histone tetramers with a higher (effective) affinity than observed for hypernucleosome formation. We hypothesized that this sequence could function as a nucleation site for hypernucleosome assembly, but this is not the case. Sequences identical to the artificially optimized DNA sequences have not been identified in archaeal genomes, although one would expect that similar sequences do exist. More general characteristics for high-affinity DNA sequences have been identified, such as a TA or AA dinucleotide every ~ 10 bp, following the average helical repeat around a nucleosome (23, 24). It is unclear what the characteristics of specific positioning DNA sequences are in archaeal genomes and what is their potential function in transcriptional regulation. So far, only an indirect relationship between archaeal histones and repressive effects on transcription have been found in *Thermococcus kodakarensis* and *E. coli* (25, 26). Combining RNA-seq with a technique like ChIP-seq that can detect histone binding at a specific location, could give us a hint of which genes might be directly regulated by histones. Such model genes could help us studying the gene regulatory function of histones.

Another important aspect of hypernucleosome formation of which our understanding is very limited, is the incorporation of histone variants. Based on their (predicted) properties, speculative models have been proposed. *In silico*, different combinations of histone variants can alter the DNA binding affinity and tetramer stability (27). Thus far, most experimental studies on archaeal histones have been performed on homodimers, where slight differences between histone paralogs have been observed. For example, the hypernucleosome formed by HMfA has weaker stacking interactions and is slightly less compact than an HMfB hypernucleosome (28). Due to their high sequence similarity (84%), heterodimers are likely formed *in vivo.* Sometimes, heteromerization seems even necessary for DNA binding. For example, one of the histones of *Methanothermobacter thermoautotrophicus* OSU, HMtB, has lost its DNA binding ability in laboratory culture and can only bind DNA when forming heteromeric complexes with HMtA2 (29). Similar behaviour was observed for the two histones of *Nanoarchaeum equitans* (30). Archaea of the class Halobacteria encode double histones, where two histone folds are connected by a short linker peptide (20, 31, 32). This results in intramolecular heterodimerization. So far, no examples of such obligate heterodimerization have been observed for archaeal histones that are not physically attached as found for their eukaryotic counterparts.

In Chapter 5, we show that histone MJ1647, a variant that contains a C-terminal extension, can both bridge DNA and bind the DNA as a wrapping tetramer. Comparable to eukaryotic cohesin and other SMC proteins, the DNA bridging function of MJ1647 could potentially result in DNA looping. MJ1467 might, therewith, have a role in higher-order genome organization. Despite having a histone fold, MJ1647 does not form heteromeric complexes with the canonical histone A3 (33). This suggests that MJ1647 does not have a direct effect on the hypernucleosome formed by A3, but may act as a roadblock for hypernucleosome extension and might compete for similar DNA sequences. In this respect, MJ1647 might be more comparable to archaeal NAPs like Alba and MC1. It has been shown *in vitro* that Alba can indeed compete with histones on DNA (34), but details of the interplay between NAPs and the hypernucleosome, and the effects on gene regulation and genome organisation are still open questions.

Limitations in studying the role of archaeal histones *in vivo* mainly come from the lack of laboratory strains and accompanying genetic tools which might need to be optimized for the respective archaeal species used. Most cultured strains belong to the Euryarchaeota and TACK-superphylum (mainly Crenarchaeota), which gives us a limited view of the archaeal domain of life. Two recent articles demonstrate the culturing of two Lokiarchaeota, but also show that they are slow-growing, syntrophic organisms (35, 36). In Chapter 6, we attempted to study the DNA binding properties of histone A and B (HA and HB) from *Ca. Heimdallarchaeota* LC_3. However, obtaining the proteins turned out to be non-trivial. The successful synthesis of HB opens up opportunities of studying histones that are difficult to obtain otherwise. This technique also makes it possible to selectively incorporate PTMs at specific sites. The finding of acetylated histones in *T. kodakarensis* and *Thermococcus gammatolerans* (37) and the presence of an N-terminal tail containing several lysine resides in HA, highlights the importance of studying the effects of these modifications. Advances in culturing archaea from different phyla could provide us with more ways to study (modified) histones. Also, it could provide us with more information about abundance of different histone variants across the growth cycle and allow for studying the 3D organization of their chromosomes by chromosome conformation capture techniques like 3C and Hi-C.

Although no pathogenic archaea have been discovered yet, they might play a role in some human diseases such as periodontal disease and inflammatory bowel disease (38, 39). Most archaea, however, are extremophiles and live in the most remote places on earth, including hydrothermal vents and saline lakes. Their habitat makes it necessary for their proteins to be resistant to these extreme conditions. Also the hypernucleosome formed by archaeal histones can be regarded as an ‘extreme’ version

of the eukaryotic nucleosome. Considering the theoretically endless nature of this complex, evident follow-up studies would focus on the factors that have an influence on modulating the size and stability of the hypernucleosome. Also, the interplay between hypernucleosome and the transcription machinery, and a possible specific role in transcription regulation needs to be further investigated.

Studying archaeal transcription regulation is important in the bigger scheme of nutrient cycles. As archaea play a role in methane metabolism, but also nitrogen and sulphur cycles (40–42), they might be a key player in addressing today's challenges such as global warming. On the other side, they could also provide solutions to sustainability issues. Many industrial processes involve the use of chemical compounds that are (potentially) harmful to the environment. Due to their high resistance, archaeal enzymes are good candidates to catalyse chemical reactions and thereby make industrial processes more sustainable (43).

Increasing our knowledge about genome organization of prokaryotes and the proteins responsible for this, might show us functional and structural conservation patterns previously unknown. In turn, this might explain how NAPs and histones evolved. The recent discoveries of histones in bacteria and hypernucleosome-like structures in eukaryotic telomers are good examples of evolutionary patterns that still need to be fully elucidated. Insights into the diversity of DNA organization and transcription regulation across the tree of life could help to understand the role that microorganisms play in major challenges the world faces today.

References

1. Williams,T.A., Foster,P.G., Cox,C.J. and Embley,T.M. (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, **504**, 231–236.
2. Luger,K., Mäder,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251–260.
3. Shilatifard,A. (2006) Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem*, **75**, 243–269.
4. Rossetto,D., Avvakumov,N. and Côté,J. (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics*, **7**, 1098–1108.
5. Louters,L. and Chalkley,R. (1985) Exchange of histones H1, H2A, and H2B in vivo. *Biochemistry*, **24**, 3080–3085.
6. Martire,S. and Banaszynski,L.A. (2020) The roles of histone variants in fine-tuning chromatin organization and function. *Nature Reviews Molecular Cell Biology 2020 21:9*, **21**, 522–541.
7. Quénet,D. (2018) Histone Variants and Disease. *International Review of Cell and Molecular Biology*, **335**, 1–39.
8. Losada,A., Hirano,M. and Hirano,T. (1998) Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev*, **12**, 1986–1997.
9. Kim,S., Yu,N.K. and Kaang,B.K. (2015) CTCF as a multifunctional protein in genome regulation and gene expression. *Experimental & Molecular Medicine 2015 47:6*, **47**, e166–e166.
10. Werner,F. (2007) Structure and function of archaeal RNA polymerases. *Molecular Microbiology*, **65**, 1395–1404.
11. Vestergaard,M., Frees,D. and Ingmer,H. (2019) Antibiotic Resistance and the MRSA Problem. *Microbiology Spectrum*, **7**.
12. Soucy,S.M., Huang,J. and Gogarten,J.P. (2015) Horizontal gene transfer: building the web of life. *Nature Reviews Genetics 2015 16:8*, **16**, 472–482.
13. Dilweg,I.W. and Dame,R.T. (2018) Post-translational modification of nucleoid-associated proteins: An extra layer of functional modulation in bacteria? *Biochemical Society Transactions*, **46**, 1381–1392.
14. Denham,E.L., Piersma,S., Rinket,M., Reilman,E., de Goffau,M.C. and van Dijl,J.M. (2019) Differential expression of a prophage-encoded glycocin and its immunity protein suggests a mutualistic strategy of a phage and its host. *Scientific Reports*, **9**, 2845.
15. Baglivo,I., Pirone,L., Pedone,E.M., Pitzer,J.E., Muscariello,L., Marino,M.M., Malgieri,G., Freschi,A., Chambery,A., Roop,R.M., *et al.* (2017) Ml proteins from Mesorhizobium loti and MucR from Brucella abortus: an AT-rich core DNA-target site and oligomerization ability. *Scientific reports*, **7**.

16. Ricci,D.P., Melfi,M.D., Lasker,K., Dill,D.L., McAdams,H.H. and Shapiro,L. (2016) Cell cycle progression in Caulobacter requires a nucleoid-associated protein with high AT sequence recognition. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, E5952–E5961.
17. Ma,R., Liu,Y., Gan,J., Qiao,H., Ma,J., Zhang,Y., Bu,Y., Shao,S., Zhang,Y. and Wang,Q. (2022) Xenogeneic nucleoid-associated EnrR thwarts H-NS silencing of bacterial virulence with unique DNA binding. *Nucleic Acids Research*, **50**, 3777–3798.
18. Rouvière-Yaniv,J., Yaniv,M. and Germond,J.E. (1979) E. coli DNA binding protein HU forms nucleosome-like structure with circular double-stranded DNA. *Cell*, **17**, 265–274.
19. Hocher,A., Laursen,S.P., Radford,P., Tyson,J., Lambert,C., Stevens,K.M., Picardeau,M., Elizabeth Sockett,R. and Luger,K. (2023) Histone-organized chromatin in bacteria. *bioXrviv*, 10.1101/2023.01.26.525422.
20. Henneman,B., van Emmerik,C., van Ingen,H. and Dame,R.T. (2018) Structure and function of archaeal histones. *PLoS Genet*, **14**, e1007582.
21. Mattiroli,F., Bhattacharyya,S., Dyer,P.N., White,A.E., Sandman,K., Burkhart,B.W., Byrne,K.R., Lee,T., Ahn,N.G., Santangelo,T.J., *et al.* (2017) Structure of histone-based chromatin in Archaea. *Science (1979)*, **357**, 609–612.
22. Soman,A., Wong,S.Y., Korolev,N., Surya,W., Lattmann,S., Vogirala,V.K., Chen,Q., Berezhnoy,N. V., van Noort,J., Rhodes,D., *et al.* (2022) Columnar structure of human telomeric chromatin. *Nature 2022 609:7929*, **609**, 1048–1055.
23. Bailey,K.A., Pereira,S.L., Widom,J. and Reeve,J.N. (2000) Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. *J Mol Biol*, **303**, 25–34.
24. Bailey,K.A. and Reeve,J.N. (1999) DNA repeats and archaeal nucleosome positioning. *Research in Microbiology*, **150**, 701–709.
25. Rojec,M., Hocher,A., Stevens,K.M., Merkenschlager,M. and Warnecke,T. (2019) Chromatinization of Escherichia coli with archaeal histones. *eLife*, **8**.
26. Sanders,T.J., Ullah,F., Gehring,A.M., Burkhart,B.W., Vickerman,R.L., Fernando,S., Gardner,A.F., Ben-Hur,A. and Santangelo,T.J. (2021) Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in Thermococcus kodakarensis. *Front Microbiol*, **12**, 1071.
27. Stevens,K.M., Swadling,J.B., Hocher,A., Bang,C., Gribaldo,S., Schmitz,R.A. and Warnecke,T. (2020) Histone variants in archaea and the evolution of combinatorial chromatin complexity. *Proc Natl Acad Sci U S A*, **117**, 33384–33395.
28. Henneman,B., Brouwer,T.B., Erkelens,A.M., Kuijntjes,G.-J., van Emmerik,C., van der Valk,R.A., Timmer,M., Kirolos,N.C.S., van Ingen,H., van Noort,J., *et al.* (2021) Mechanical and structural properties of archaeal hypernucleosomes. *Nucleic Acids Res*, **49**, 4338–4349.

29. Sandman,K., Louvel,H., Samson,R.Y., Pereira,S.L. and Reeve,J.N. (2008) Archaeal chromatin proteins histone HMtB and Alba have lost DNA-binding ability in laboratory strains of Methanothermobacter thermautotrophicus. *Extremophiles*, **12**, 811–817.
30. Friedrich-Jahn,U., Aigner,J., Längst,G., Reeve,J.N. and Huber,H. (2009) Nanoarchaeal origin of histone H3? *Journal of Bacteriology*, **191**, 1092–1096.
31. Slesarev,A.I., Belova,G.I., Kozyavkin,S.A. and Lake,J.A. (1998) Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. *Nucleic Acids Research*, **26**, 427–430.
32. Fahrner,R.L., Cascio,D., Lake,J.A., Slesarev,A., Shemyakin,M.M. and Ovchinnikov,Y.A. (2001) An ancestral nuclear protein assembly: Crystal structure of the Methanopyrus kandleri histone. *Protein Science*, **10**, 2002–2007.
33. Ofer,S., Blombach,F., Erkelens,A.M., Barker,D., Schwab,S., Smollett,K., Matelska,D., Fouqueau,T., van der Vis,N., Kent,N.A., *et al.* (2023) DNA-bridging by an archaeal histone variant via a unique tetramerisation interface. *Res Sq*, 10.21203/rs.3.rs-2183355/v1.
34. Maruyama,H., Prieto,E.I., Nambu,T., Mashimo,C., Kashiwagi,K., Okinaga,T., Atomi,H. and Takeyasu,K. (2020) Different Proteins Mediate Step-Wise Chromosome Architectures in Thermoplasma acidophilum and Pyrobaculum calidifontis. *Frontiers in Microbiology*, **11**, 1247.
35. Imachi,H., Nobu,M.K., Nakahara,N., Morono,Y., Ogawara,M., Takaki,Y., Takano,Y., Uematsu,K., Ikuta,T., Ito,M., *et al.* (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature*, **577**, 519–525.
36. Rodrigues-Oliveira,T., Wollweber,F., Ponce-Toledo,R.I., Xu,J., Rittmann,S.K.M.R., Klingl,A., Pilhofer,M. and Schleper,C. (2023) Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature*, **613**, 332–339.
37. Alpha-Bazin,B., Gorlas,A., Lagorce,A., Joulié,D., Boyer,J.B., Dutertre,M., Gaillard,J.C., Lopes,A., Zivanovic,Y., Dedieu,A., *et al.* (2021) Lysine-specific acetylated proteome from the archaeon Thermococcus gammatolerans reveals the presence of acetylated histones. *J Proteomics*, **232**, 104044.
38. Lepp,P.W., Brinig,M.M., Ouverney,C.C., Palm,K., Armitage,G.C. and Relman,D.A. (2004) Methanogenic Archaea and human periodontal disease. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 6176–6181.
39. Lecours,P.B., Marsolais,D., Cormier,Y., Berberi,M., Haché,C., Bourdages,R. and Duchaine,C. (2014) Increased prevalence of Methanosphaera stadtmanae in inflammatory bowel diseases. *PloS one*, **9**.
40. Cabello,P., Roldán,M.D. and Moreno-Vivián,C. (2004) Nitrate reduction and the nitrogen cycle in archaea. *Microbiology (Reading, England)*, **150**, 3527–3546.
41. Liu,Y., Beer,L.L. and Whitman,W.B. (2012) Sulfur metabolism in archaea reveals novel processes. *Environmental Microbiology*, **14**, 2632–2644.

42. Evans,P.N., Boyd,J.A., Leu,A.O., Woodcroft,B.J., Parks,D.H., Hugenholtz,P. and Tyson,G.W. (2019) An evolving view of methane metabolism in the Archaea. *Nature Reviews Microbiology 2018 17:4*, **17**, 219–232.
43. Cabrera,M.Á. and Blamey,J.M. (2018) Biotechnological applications of archaeal enzymes from extreme environments. *Biological Research 2018 51:1*, **51**, 1–15.

Summary

A skein of cotton yarn with dimensions of approximately 10 by 6 centimeters has a length of 85 meters when completely stretched out. Yarn can be organized in a skein, hank, ball or cake to remain untangled and ready to use for the purpose at hand. In a similar way, the length of prokaryotic genomic DNA (in the order of 2 millimeters) far exceeds the volume of a prokaryotic cell (1-2 by 5-10 micrometers), requiring an even larger degree of compaction while keeping the biological information stored in DNA accessible. Several strategies have evolved to achieve this, one of them being the expression of proteins that bind and organize the DNA. The best-known examples are the histone proteins of eukaryotes, but many other proteins have the same function. They are collectively referred to as nucleoid-associated proteins (NAPs). There are multiple ways in which histones and NAPs can organize DNA. In Chapter 1 we discuss the different possibilities: DNA wrapping, DNA bridging, DNA bending and nucleofilament formation. In each domain of life examples of (nearly) all categories can be found, although the proteins involved are often unrelated. This could be an example of convergent evolution: independent evolution of different proteins that fulfill the same function.

In the model bacterium *Escherichia coli* at least 12 NAPs have been described. The histone-like nucleoid structuring protein (H-NS) organizes the genome and has an effect on the expression of 5-10% of *E. coli*'s genes. H-NS is an example of both a DNA bridging and a nucleofilament forming protein. Several factors play a role in determining which structure is formed by H-NS, such as environmental conditions, protein interaction partners and post translational modifications. In Chapter 2, we review the functional and structural properties of H-NS and its (functional) homologues in other bacteria. We propose that charge distribution in an important characteristic to predict if an H-NS-like protein will react to changes in the environment. If the charge distribution is asymmetrical, and the protein has a certain level of interna flexibility, this is an indication that osmotic strength can be the switch between nucleofilament formation and DNA bridging.

The NAP Rok from *Bacillus subtilis* is one of the proposed H-NS-like proteins, however its charge distribution is different. Instead of an asymmetrical charge distribution, the charges are more spread out and the middle part of the protein is neutral. In Chapter 3, we show that Rok only forms DNA bridges and does only mildly react to environmental changes. This suggests that these conditions are not as important in the regulation of Rok as for other H-NS-like proteins. In search of a mechanism to perturb Rok-DNA bridges, we investigated a smaller variant of Rok, called sRok. This NAP can form both a nucleofilament along the DNA and DNA-DNA bridges and is sensitive to salt

concentration. In a situation where Rok and sRok are together, they form heterodimers and influence each other's behaviour. These results show us the importance of protein partners in the regulation of Rok and other DNA organizing proteins as well.

The other prokaryotic domain of life, the archaea, has its own, sometimes lineage specific, DNA organizing proteins. Archaea were long regarded the third domain of life, but new models show two domains: the bacterial domain and the archaea-eukaryotes domain where eukaryotes evolved from archaea. This is partly reflected in their DNA binding proteins. In eukaryotes, the well-studied histone proteins are the main DNA organizing proteins. Most archaea also express histones, but they generally lack the tails of eukaryotic histones used for gene regulation. Previously, it was found that archaeal histones can wrap DNA in a continuous, rod-like manner. This 'endless' structure is called a hypernucleosome. Practically, the hypernucleosome can be limited in size by several factors such as histone variants, posttranslational modifications, environmental conditions and specific DNA sequences. In Chapter 4, we examined the effect of an artificial high-affinity DNA sequence on the formation of a hypernucleosome by model archaeal histones HMfA and HMfB. We found that the specific DNA sequence is first bound by a tetramer at low protein concentrations, but that this tetramer does not promote hypernucleosome formation. We propose that this is due to a more closed tetrameric conformation of the histones which is not compatible with hypernucleosome formation. Combined with histone variants that are less likely to form a hypernucleosome but do recognize this sequence, this might be a way to mark the start and end of the hypernucleosome.

Some histone variants do not only consist of a histone fold, variants exist with either an N- or C-terminal tail. In Chapter 5, we investigated MJ1647, a histone with a C-terminal tail from *Methanocaldococcus jannaschii*. We found that it has two DNA binding modes: it can wrap DNA like other archaeal histones, but it can also bridge two DNA strands. In both cases, MJ1647 forms tetramers and the behavior is dependent on the C-terminal tail, which we called the tetramerization domain. Due to steric hindrance caused by the C-terminal tail, MJ1647 cannot form a hypernucleosome. It might however act as a roadblock to stop progression of a hypernucleosome.

Following evolution from archaea towards the eukaryotes in the tree of life brings us at the Asgard archaea as the closest relatives to eukaryotes today. This is not only reflected by their high amount of eukaryotic signature proteins, but also the presence of histones with an N-terminal tail. *Heimdallarchaeota* LC_3 encodes 10 histone proteins of which one has such a tail. In Chapter 6, we attempted to produce the eukaryotic-like histone HA and the archaeal-like histone HB, however we were unsuccessful in our

attempts both in *E. coli* and *Pichia pastoris.* We were able to chemically synthesize HB and study its DNA binding behaviour. In contrast to predictions, HB does not seem to form a hypernucleosome.

The results presented in this thesis show that there is great variation in DNA organizing proteins and that proper DNA organization can be achieved in different ways. In Chapter 7, the broader impact of this research is discussed. Prokaryotes play an important part in today's major challenges such as the rise of antibiotic resistant bacteria and the role of methanogenic archaea in climate change. Therefore, it is important to study how prokaryotes regulate their gene expression and how they change its DNA organization according to environmental changes.

Nederlandse samenvatting

Een bol katoengaren met een afmetingen van ongeveer 10 bij 6 centimeter heeft een draadlengte van 85 meter wanneer het helemaal is afgewikkeld. Garen kan worden gesponnen in de vorm van een bal, een meer langwerpige ovaal of losjes om zichzelf gewikkeld. In iedere vorm is het essentieel dat het garen niet in de knoop raakt en makkelijk is in gebruik. Hetzelfde principe geldt voor het DNA. Het genoom van een prokaryoot is ongeveer 2 millimeter lang terwijl de afmetingen van de cel maar 1-2 bij 5-10 micrometer zijn. Er is dus nog meer vouwing en compactie nodig om dit te laten passen en tegelijkertijd moet de cel bij de informatie kunnen die in het DNA is opgeslagen. De cel heeft verschillende mogelijkheden om dit te doen: één daarvan is het maken van eiwitten die aan het DNA binden en het organiseren. De meeste bekende categorie van deze eiwitten zijn histonen van eukaryotische cellen (bijvoorbeeld menselijke cellen). In prokaryoten kun je eiwitten vinden met dezelfde functie, die noemen we nucleoid-associated proteins (NAPs). Er zijn verschillende manieren waarop deze eiwitten het DNA op kunnen vouwen. In Hoofdstuk 1 bespreken we deze mogelijkheden: het opwinden en buigen van DNA, bruggen vormen tussen twee stukken DNA en het vormen van een filament langs het DNA om het stijver te maken. In ieder domein van het leven worden (bijna) alle mogelijkheden gebruikt al gebeurt dit vaak met behulp van verschillende eiwitten. Dit is waarschijnlijk een voorbeeld van convergente evolutie: onafhankelijke evolutie van verschillende eiwitten met dezelfde functie.

Voor de modelbacterie *Escherichia coli* zijn minstens 12 NAPs bekend. Het H-NS-eiwit (histone-like nucleoid structuring protein) maakt het genoom compact en heeft een effect op het alfezen van ongeveer 5-10% of de genen van *E. coli*. H-NS kan zowel bruggen tussen DNA vormen als een filament langs het DNA vormen. Welke structuur H-NS gebonden aan het DNA vormt hangt af van verschillende factoren zoals omgevingsfactoren (temperatuur, pH en zout), andere eiwitten waarmee het een interactie aan kan gaan en posttranslationele modificaties. In Hoofdstuk 2 vatten we de functionele en structurele eigenschappen van H-NS en homologe eiwitten in andere bacteriën samen. We stellen dat de ladingsverdeling van het eiwit kan voorspellen of een H-NS-achtig eiwit kan reageren op zijn omgeving. Als de ladingsverdeling asymmetrisch is, en het eiwit genoeg flexibel is, is dat een indicatie dat de zoutconcentratie het H-NS-achtige eiwit kan laten schakelen tussen de vorming van een filament en van DNA-bruggen.

De NAP Rok in *Bacillus subtilis* wordt gezien als een mogelijk H-NS-achtig eiwit. De ladingsverdeling is echter anders dan de asymmetrische verdeling van de meeste H-NS-achtige eiwitten. De ladingen zijn bij Rok meer uitgespreid en het middelste gedeelte

van Rok is neutraal. In Hoofdstuk 3 laten we zien dat Rok alleen DNA-bruggen kan vormen en maar minimaal reageert op omgevingsfactoren. Dit suggereert dat Rok veel minder wordt gereguleerd door zijn omgeving dan andere H-NS-achtige eiwitten. De zoektocht naar andere factoren om de Rok-DNA bruggen te reguleren leidde naar een kleinere variant van Rok genaamd sRok. Deze NAP kan wel zowel DNA-bruggen als een filament op het DNA vormen. Ook reageert sRok meer dan Rok op verschillende zoutconcentraties. Wanneer Rok en sRok gemengd worden, vormen ze heterodimeren en hebben ze een invloed op elkaars gedrag. Deze resultaten laten zien dat eiwitpartners belangrijk zijn voor het reguleren van Rok en daarmee waarschijnlijk ook voor andere DNA-organiserende eiwitten.

In het andere prokaryotische domein, de archaea, zijn andere DNA-organiserende eiwitten te vinden, die soms alleen in een bepaalde tak organismen voorkomen. Archaea werden lange tijd beschouwd als het derde domein van het level, maar nieuwe modellen laten zien dat de boom van het leven waarschijnlijk maar uit twee takken bestaat: de tak met bacteriën en de tak met archaea en eukaryoten. In dit model zijn de eukaryoten geëvolueerd vanuit de archaea. Deze relatie is zichtbaar in hun DNA-organiserende eiwitten. In eukaryoten zijn histonen de voornaamste DNA-organiserende eiwitten. De meeste archaea coderen ook voor histonen in hun genoom, maar deze histonen missen de flexibele staarten van de eukaryotische varianten die voor regulatie worden gebruikt. In eerder onderzoek is gevonden dat archaeale histonen het DNA kunnen opwinden rond een soort cilinder gevormd door histonen. Deze, in theorie eindeloze, structuur noemen we een hypernucleosoom. In de praktijk is deze structuur niet eindeloos, maar wordt gelimiteerd door verschillende factoren zoals histonvarianten, post-translationele modificaties, omgevingsfactoren en specifieke DNA-sequenties. In Hoofdstuk 4 hebben we het effect onderzocht van een artificiële DNA-sequentie met een hoge affiniteit voor histonen en wat dit betekent voor het vormen van een hypernucleosoom door de archaeale model histonen HMfA en HMfB. Deze specifieke DNA-sequentie wordt eerst gebonden door een tetrameer bij lage eiwitconcentraties, maar dit vormt geen voordeel voor het vormen van een hypernucleosoom ten opzichte van aspecifiek DNA. Ons voorstel is dat dit komt doordat de tetrameer op de specifieke sequentie meer gesloten is en daarom niet geschikt om een hypernucleosoom mee te starten. Als je dit combineert met varianten van histonen die minder geschikt zijn om een hypernucleosoom te vormen, maar wel deze specifieke sequentie herkennen, kan dit een manier zijn om het begin en het einde van een hypernucleosoom mee aan te geven.

Sommige varianten van histonen zijn wat groter dan de standaard histonen door staarten aan het begin of einde van het eiwit. In Hoofdstuk 5 onderzoeken we de

eigenschappen van MJ1647, een histon met een staart aan het einde van *Methanocaldococcus jannaschii*. Dit histon kan op twee manieren het DNA organiseren: het kan zowel het DNA opwinden zoals andere archaeale histonen, maar het kan ook DNA bruggen vormen. Voor allebei de toestanden is het nodig dat MJ1647 tetrameren vormt via de staart. Omdat deze staart teveel ruimte inneemt, kan MJ1647 geen hypernucleosoom vormen. Het zou echter wel een rol kunnen hebben als versperring op het DNA om verdere groei van een hypernucleosoom te stoppen.

In het model waarin de boom van het leven twee takken heeft, komen we uit bij de Asgard archaea als de archaea die het meest verwant zijn aan eukaryoten. Dit kunnen we zien aan de grote hoeveelheid 'eukaryotic signature proteins', maar ook aan de aanwezigheid van histonen die beginnen met een flexibele staart. *Heimdallarchaeota* LC_3 codeert voor 10 histonen waarvan er één zo'n staart heeft, namelijk HA. In Hoofdstuk 6 hebben we tevergeefs geprobeerd om HA en het meer archaeale histon HB te produceren in *E. coli* en de gist *Picha pastoris*. Het was mogelijk om HB chemisch te synthetiseren en op deze manier de DNA bindende eigenschappen te onderzoeken. In tegenstelling tot de voorspellingen lijkt HB geen hypernucleosoom te kunnen vormen.

De resultaten in dit proefschrift laten zien dat er grote variatie te vinden is in DNA-organiserende eiwitten en dat DNA op verschillende manieren georganiseerd kan worden. De bredere impact van dit onderzoek wordt beschreven in Hoofdstuk 7. Prokaryoten spelen een belangrijke rol in een aantal globale uitdagingen van deze tijd, zoals de toenemende resistentie van bacteriën tegen antibiotica en het effect van methaan-producerende archaea op klimaatverandering. Het is daarom belangrijk om te bestuderen hoe prokaryoten hun genen reguleren en hoe ze op DNA-niveau reageren op een veranderende omgeving.

Curriculum Vitae

Mandy Erkelens was born on August 16th, 1995 in Rotterdam. In 2013, she started her her bachelor's degree Molecular Life Sciences in Wageningen which was followed by a master's degree Molecular and Cellular Life Sciences in Utrecht. During her master's, she did internships in the group of prof. dr. Ineke Braakman at Utrecht University on posttranslational modifications of the cystic fibrosis protein CFTR and in the group of prof. dr. Per Jemth at Uppsala University in Sweden on the evolution of a protein-protein interaction. In 2018, she joined the group of prof. dr. Remus Dame in Leiden as a PhD student. Her research focused on the properties of DNA bridging and wrapping proteins in bacteria and archaea. Mandy is now a Postdoctoral researcher in the Dame group.

List of publications

Described in this thesis:

A.M. Erkelens, B. Henneman, R.A. van der Valk, N.C.S. Kirolos, R.T. Dame, (2023) Specific DNA binding of archaeal histones HMfA and HMfB, *Frontiers in Microbiology*, 14, 1166608

S. Ofir, F. Blombach, A.M. Erkelens, D. Barker, S. Schwab, K. Smollett, D. Matelska, T. Fouqueau, N. van der Vis, N.A. Kent, R.T. Dame, F. Werner, (2023) DNA-bridging by an archaeal histone variant via a unique tetramerization interface, *Res Sq*, 10.21203/rs.3.rs-2183355/v1

A.M. Erkelens, L. Qin, B. van Erp, A. Miguel-Arribas, D. Abia, H.G.J. Keek, D. Markus, M.K.M. Cajili, S. Schwab, W.J.J. Meijer, R.T. Dame, (2022) The *B. subtilis* Rok protein is an atypical H-NS-like protein irresponsive to physico-chemical cues, *Nucleic Acids Research*, 50, 12166-12185

L. Qin, A.M. Erkelens, F. Ben Bdira, R.T. Dame, (2019) The architects of bacterial DNA bridges: a structurally and functionally conserved family of proteins, *Open Biology*, 9, 190223

Other publications:

C. Du, J.J. Willemse, A.M. Erkelens, V.J. Carrion Bravo, R.T. Dame, G.P. van Wezel, (2022) System-wide analysis of the GATC-binding nucleoid-associated protein Gbn and its impact on Streptomyces development, *mSystems*, 7, e00061-22

D. Forrest, E.A. Warman, A.M. Erkelens, R.T. Dame, D.C. Grainger, (2022) Xenogeneic silencing strategies in bacteria are dictated by RNA polymerase promiscuity, *Nature Communications*, 13, 1149

F. Ben Bdira, A.M. Erkelens, L. Qin, A.N. Volkov, A.M. Lippa, N. Bowring, A.L. Boyle, M. Ubbink, S.L. Dove, R.T. Dame, (2021) Novel anti-repression mechanism of H-NS by a phage protein, *Nucleic Acids Research,* 49 10770-10784

B. Henneman, T.B. Brouwer, A.M. Erkelens, G.J. Kuijntjes, C. van Emmerik, R.A. van der Valk, M. Timmer, N.C.S. Kirolos, H. van Ingen, J. van Noort, R.T. Dame (2021), Mechanical and structural properties of archaeal hypernucleosomes, *Nucleic Acids Research,* 49, 4338-4349

E. Karlsson, C. Paissoni, A.M. Erkelens, Z.A. Tehranizadeh,, F.A. Sorgenfrei, E. Andersson, W. Ye, C. Camilloni, P. Jemth, (2020), Mapping the transition state for a binding reaction between ancient intrinsically disordered proteins, *The Journal of Biological Chemistry,* 295, 17698-17712

Acknowledgements

After 5 years of working at MacBio, it is time to look back and thank all the wonderful people I met along the way.

Allereerst, Remus bedankt dat je me de kans gaf om mijn PhD in jouw groep te doen. Ik heb veel van je geleerd over onderzoek doen en het opschrijven ervan. We zullen vast nog af en toe contact hebben over de lopende projecten en verder onderzoek naar archaeale chromatine. John, bedankt voor je input als tweede supervisor en de nuttige discussies over histonen tijdens de DaNo meetings.

Bert, ik denk dat niemand verbaasd is dat jij een van mijn paranimfen bent. Je ging van master student naar collega en nu ben je een van mijn beste vrienden. Je staat altijd voor me klaar en je hebt de gave om me altijd weer aan het lachen te krijgen, bedankt daarvoor. Marc, you are a nice addition to the group and to my office. Although your proteins turned out to be complicated, you never give up and keep on trying. Now, you will continue with the HMf project and I wish you the best of luck.

The general opinion is that Remus' group consists of interesting people and I agree with that statement. Fatema Zahra, thank you for the nice chats and the advice you've given me over the years. Good luck with the next step of your career! Ilias, veel succes met de laatste loodjes en het afronden van je eigen thesis. Samuel, veel succes met het in kaart brengen van alle nieuwe interessante histonen. Ik ben benieuwd naar de resultaten! Pingzhuang, good luck with your *in vivo* experiments to expand our knowledge about the famous H-NS protein. I would also like to thank the past members of Remus' group who have helped me during the earlier stages of my research.

Anneloes, Monika en Patrick, bedankt voor alle technische ondersteuning en adviezen wanneer de experimenten niet wilden lukken. Ilona, als het goed is verdedig je net iets eerder dan ik. Heel veel succes! We liepen ongeveer gelijk qua planning en daardoor liepen we vaak tegen hetzelfde aan. Waar je hierna ook gaat werken, ik weet zeker dat het lab altijd netjes zal zijn. Laura, op het oog leek je vooral chaos, maar eigenlijk had je alles altijd precies onder controle. Ik hoop dat je ook snel aan je verdediging toe bent!

The EE2 wing has seen many people come and go over the years. To thank everyone personally would require a separate thesis. So, to all my EE2 colleagues: thank you for creating a nice environment, good luck with your research and don't forget to enjoy it too!

Tijdens mijn PhD onderzoek heb ik met veel plezier studenten begeleid. Willem, Erik, Helena, Nico, Bert, Samuel, Shane en Lars, bedankt voor alle gezelligheid en het verder helpen van mijn onderzoek. Jullie kunnen allemaal wel iets van jullie stages terugvinden in dit boekje.

Of course, one never does science completely on its own. I had the chance to work on several articles with many nice collaborators: Finn and Fabian in London, David and Dave in Birmingham and David and Wilfried in Madrid. The Rok project will be continued by Xam and Gijs in Amsterdam and Fredrik in Göteborg, good luck!

Many thanks to my colleagues of the Green Team, Anneloes, Laura, Antonella, Imogen, Cristina, Nicole and Thomas. I am very proud of what we have achieved the past years and I hope you do continue to make the lab more sustainable. Thomas, bedankt voor je inzichten in de wetenschappelijke wereld en voor mij een zetje in de goede richting geven toen ik dat nodig had.

Marijke, dankjewel voor alle gezellige middagen met thee en chocola. Je gaf me altijd je kijk van buiten de academische wereld op dingen, wat alles een beetje in perspectief hield.

Natuurlijk mijn lieve ouders, broertje en andere (schoon)familie, bedankt voor alle steun tijdens mijn studie en PhD onderzoek. Ook al hadden jullie vaak geen idee wat ik precies aan het doen was, jullie waren er altijd voor me als het even moeilijk was. Jullie waren ook onmisbaar tijdens onze verhuizingen.

Mijn lieve man Fatih, zonder jou had ik dit nooit kunnen doen. Je was er voor me op de goede en slechte dagen met een lach, een knuffel en altijd heerlijk zelfgemaakt eten. Nu wonen we in ons mooie huisje in Alphen aan den Rijn met onze thuiswerk collega en emotional support animal Neena en ik hoop dat dat nog lang zo mag blijven.