The bacterial chromatin protein HupA can remodel DNA and associates with the nucleoid in *Clostridium difficile*

Ana M. Oliveira Paiva[1,3], Annemieke H. Friggen[1,3], Liang Qin[2,3], Roxanne Douwes[1], Remus T. Dame[2,3], Wiep Klaas Smits[1,3]*

1) Department of Medical Microbiology, Section Experimental Bacteriology, Leiden University Medical Center, Leiden, The Netherlands
2) Faculty of Science, Leiden Institute of Chemistry, Leiden University, Leiden, Netherlands
3) Center for Microbial Cell Biology, Leiden, The Netherlands

* Corresponding author: Department of Medical Microbiology, Section Experimental Bacteriology, Postzone E4-P, Leiden University Medical Center, PO Box 9600, 2300 RC, Leiden, The Netherlands, phone +31-71-5261229, e-mail w.k.smits@lumc.nl

Keywords: HU, bacterial chromatin protein, HupA, tethered particle motion, chromosome, fluorescence microscopy, electrophoretic mobility shift assay, bending

ORCID: AMOP: 0000-0002-6122-832X; AHF: 0000-0001-5780-2053; LQ: 0000-0002-5419-8089; RD: 0000-0003-2038-0998; RTD: 0000-0001-9863-1692; WKS: 0000-0002-7409-2847

Abstract

The maintenance and organization of the chromosome plays an important role in the development and survival of bacteria. Bacterial chromatin proteins are architectural proteins that bind DNA, modulate its conformation and by doing so affect a variety of cellular processes. No bacterial chromatin proteins of *C. difficile* have been characterized to date.

Here, we investigate aspects of the *C. difficile* HupA protein, a homologue of the histone-like HU proteins of *Escherichia coli*. HupA is a 10 kDa protein that is present as a homodimer *in vitro* and self-interacts *in vivo*. HupA co-localizes with the nucleoid of *C. difficile*. It binds to the DNA without a preference for the DNA G+C content. Upon DNA binding, HupA induces a conformational change in the substrate DNA *in vitro* and leads to compaction of the chromosome *in vivo*.

The present study is the first to characterize a bacterial chromatin protein in *C. difficile* and opens the way to study the role of chromosomal organization in DNA metabolism and on other cellular processes in this organism.

Introduction

Clostridium difficile (also known as *Clostridioides difficile*)[1] is a gram-positive anaerobic bacterium that can be found in the environment like the soil, water, and even meat products [2, 3]. It is an opportunistic pathogen and the leading cause of antibiotic associated diarrhea in nosocomial infections [4]. *Clostridium difficile* infection (CDI) can present symptoms that range from mild diarrhea to more severe disease, such as pseudomembranous colitis, and can even result in death [4]. Over the past two decades the incidence of CDI worldwide, in a healthcare setting as well as in the community has increased [4-6]. *C. difficile* is resistant to a broad range of antibiotics and recent studies have reported cases of decreased susceptibility of *C. difficile* to some of the available antimicrobial therapies [7, 8]. Consequently, the interest in the physiology of the bacterium has increased in order to explore new potential targets for intervention.

The maintenance and organization of the chromosome plays an important role in the development and survival of bacteria. Several proteins involved in the maintenance and organization of the chromosome have been explored as potential drug targets [9-11]. The bacterial nucleoid is a highly dynamic structure organized by factors such as the DNA supercoiling induced by the action of topoisomerases [12], macromolecular crowding [13, 14] and interactions with nucleoid-associated proteins (NAPs) [15, 16]. Bacterial NAPs have been implicated in efficiently compacting the nucleoid while supporting the regulation of specific genes for the proliferation and maintenance of the cell [16].

NAPs are present across all bacteria and several major families have been identified [16, 17]. One of the most abundant NAPs in the bacterial cell are bacterial chromatin proteins like the histone-like HU/IHF protein family [18, 19]. *Escherichia coli* contains three HU/IHF family proteins (αHU, βHU, IHF) that have been extensively characterized [19-22]. By contrast, *Bacillus subtilis* and several other gram-positive organisms only contain one protein of the HU/IHF protein family [17, 19, 23]. In *E. coli* disruption of αHU and/or βHU function leads to a variety of growth defects or sensitivity to adverse conditions, but HU is not essential for cell survival [24, 25]. However, in *B. subtilis* the HU protein HBsu is

essential for cell viability, likely due to the lack of functional redundancy of the HU proteins such as in *E. coli* [17, 23].

In solution, most HU proteins are found as homodimers or heterodimers and are able to bind DNA through a flexible DNA binding domain. The crystal structure of the *E. coli* αHU-βHU heterodimer suggests the formation of higher order complexes at higher protein concentrations [22]. Modelling of these complexes suggest HU proteins have the ability to form higher-order complexes through dimer-dimer interaction and make nucleoprotein filaments [22, 26, 27]. However, the physiological relevance of these is still unclear [18, 22, 27].

The flexible nature of the DNA-binding domain in HU proteins confers the ability to accommodate diverse substrates. Most proteins bind with variable affinity and without strong sequence specificity to both DNA and RNA [28]. Some bacterial chromatin proteins have a clear preference for AT-rich regions [29-31] or for the presence of different structures on the DNA [28, 32].

HU proteins can modulate DNA topology in various ways. They can stabilize negatively supercoiled DNA or constrain negative supercoils in the presence of topoisomerase [22, 33]. HU proteins are involved in modulation of the chromosome conformation and have been shown to compact DNA [16, 26, 34]. This compaction of DNA is possible through the ability of HU proteins to introduce flexible hinges and/or bend the DNA [16, 26, 34, 35].

The ability to induce conformational changes in the DNA influences a variety of cellular processes due to an indirect effect on global gene expression [36-40]. In *E. coli* HU proteins are differentially expressed during the cell cycle. The αHU-βHU heterodimer is prevalent in exponential growth phase, while when in stationary phase HU is predominantly present as homodimers [41]. Several studies suggest an active role of HU proteins in the transcription and translation of other proteins and even on DNA replication and segregation of the nucleoids [42-44].

The diverse roles of HU proteins are underscored by their importance for metabolism and virulence in bacterial pathogens. Disruption of both HU homologues (HUα and HUβ) in *Salmonella typhimurium*, for example, results in the down-regulation of the pathogenicity island SPI2 and consequently a reduced ability to survive during macrophage invasion [45]. Other studies have shown the importance of HU proteins for the adaptation to stress conditions, such as low pH or antibiotic treatment [46-48]. For instance, in *M. smegmatis* deletion of *hupB* leads to increased sensitivity to antimicrobial compound [47].

Despite the wealth of information from other organisms, no bacterial chromatin protein has been characterized to date in the gram-positive enteropathogen *Clostridium difficile*. In this study, we show that *C. difficile* HupA (CD3496) is a legitimate homologue of the bacterial HU proteins. We show that HupA exists as a homodimer, binds to DNA and co-localizes with the nucleoid. HupA binding induces a conformational change of the substrate DNA and leads to compaction of the chromosome. This study is the first to characterize a bacterial chromatin protein in *C. difficile*.

Results and Discussion

C. difficile *encodes a single HU protein, HupA*

To identify bacterial chromatin proteins in *C. difficile*, we searched the genome sequence of *C. difficile* for homologs of characterized HU proteins from other organisms. Using BLASTP (https://blast.ncbi.nlm.nih.gov/), we identified a single homologue of the HU proteins in the genome of the reference strain 630 [49]; GenBank: AM180355.1), encoded by the *hupA* gene (CD3496)(e-value: 1e-22). This is similar to other gram-positive organisms, where also a single member of this family is found [17, 19, 23] and imply an essential role of this protein on the genome organization in *C. difficile*. Moreover, lack of *hupA* mutants during random transposon mutagenesis of the epidemic *C. difficile* strain R20291 supports that the *hupA* gene (CDR20291_3333) is essential [50].

Alignment of HupA amino acid sequence with selected homologs from other organisms shows a sequence identity varying between 58% to 38% (Fig. 1a). HupA displays the highest sequence identity with *Staphylococcus aureus* HU (58%). When compared to the *E. coli* HU proteins, HupA has a higher sequence identity with βHU (47%) than with αHU (43%).

The overall structure of HU proteins is conserved has previously described by the analysis of several nucleoid-associated proteins [19, 51]. To confirm the structural similarity of the *C. difficile* HupA protein to other HU proteins, we performed a PHYRE2 structure prediction [52]. All top-scoring models are based on structures from the HU family. The model with the highest confidence (99.9) and largest % identity (60%) is based on a structure of the *S. aureus* HU protein (PDB: 4QJU). Next, we generated a structural model of HupA using SWISS-MODEL [53] and *S. aureus* HU protein (Uniprot ID: Q99U17)[54] as a template. As expected, the predicted structure (Fig. 1b) is a homodimer, in which each monomer contains two domains as is common for HU proteins [51, 54]. The α-helical dimerization domain contains a helix-turn-helix (HTH) and the DNA-binding domain consists of a protruding arm composed of 3 β-sheets (Fig. 1b). In the dimer the two β-arms form a conserved pocket that can extensively interact with the DNA [54](Fig 1a).

Crystal structures of HU-DNA complexes have shed light on the mode of interaction of HU proteins with DNA and an overall mechanism for DNA binding has been proposed [35, 54-56]. In the co-crystal structure of *S. aureus* HU the arms embrace the minor groove of the dsDNA [54]. Proline residues at the terminus of the arms cause distortion of the DNA helix, by creating or stabilizing kinks [35, 54]. Further electrostatic interactions between the sides of HU dimers and the phosphate backbone facilitate DNA bending [57]. In *Borrelia burgdorferi* direct interactions between the DNA backbone and the helices of the Hbb protein dimerization domain were observed [56]. The overall similarity of *C. difficile* HupA to other HU family proteins (Fig. 1a) and a similar predicted electrostatic surface potential (Fig. 1c) suggest a conserved mechanism on HupA DNA binding in *C. difficile*.

Mutating arginine residues in the beta-arm of HupA eliminates DNA binding

Based on the alignment and structural model of HupA (Fig.1) we predict that several aminoacid residues in *C. difficile* HupA could be involved in the interaction with DNA. Specifically, the positively charged arginine residues R55, R58 and R61 on the β-arms of HupA (Fig. 1a and b) were of interest. In *B. stearothermophilus* arginine 55 of BstHU (residue reference to *C. difficile*) is essential for the interaction with DNA, while residues R58 and R61 have a minor effect [58]. In contrast, R58 and R61 play an important role in DNA binding of *E. coli* βHU [59]. In *S. aureus* substitutions of the residue R58 reduced the affinity of HU for DNA while R55 and R61 were crucial for proper DNA binding [54].

As it has been shown that disruption of a single residue may not be sufficient to abolish DNA binding [32, 58, 59], we substituted the residues R55, R58 and R61 (Fig. 1b, blue sticks) in *C. difficile* HupA based on the published mutations in HU from other organisms [54, 58, 59]. Residue R55 was changed to glutamine (Q), a neutral residue with long side chain. R58 and R61 were replaced by glutamic acid (E) and aspartic acid (D), respectively, both negatively charged residues. The resulting protein is referred to as $HupA^{QED}$. Evaluation of the effect of these mutations on the electrostatic surface potential of the structural model of HupA reveals that compared to the wild type protein (Fig. 1c), $HupA^{QED}$ exhibits a reduced positively charged surface of the DNA binding pocket (Fig. S1), which is expected to prevent the interaction with DNA.

To test the DNA binding of HupA and $HupA^{QED}$ we performed gel mobility shift assays. *C. difficile* HupA and $HupA^{QED}$ were heterologous produced and purified as 6x histidine-tagged fusion proteins ($HupA_{6xHis}$ and $HupA^{QED}{}_{6xHis;}$ see Materials and Methods). We incubated increasing concentrations of protein with different [γ-^{32}P]-labelled 38 bp double stranded DNA (dsDNA) fragments with different G+C content. When $HupA_{6xHis}$ was incubated with the DNA fragment a shift in mobility is evident, dependent on the protein concentration (Fig. 2a). At 2 µM of protein approximately 70% of DNA is present as a

DNA:protein complex (Fig. 2b). This clearly demonstrates that $HupA_{6xHis}$ is capable of interacting with DNA.

Some nucleoid-associated proteins demonstrate a clear preference for AT-rich regions [29, 30, 60]. We considered that binding of HupA could show preference for low G+C content DNA, since *C. difficile* has a low genomic G+C content (29.1% G+C). We tested DNA binding to dsDNA with 71.1%; 52.6% and 28.9% G+C content and observed no notable difference in the affinity (Fig. 2b). Our analyses does not exclude possible sequence preference or differential affinity for DNA with different structures (e.g. bent, looped, or otherwise deformed)[28, 54]. To investigate possible sequence specificity in more detail, chromatin immune-precipitation experiments (ChIP-chip or ChIP-seq) should be performed on live cells.

Having established DNA binding by $HupA_{6xHis}$, we examined the effect of replacing the arginine residues in the β-arm in the same assay. When $HupA^{QED}{}_{6xHis}$ was incubated with all three tested DNA fragments, no shift was observed (Fig. 2a and b). This suggests that the introduction of the R55Q, R58E and R61D mutations successfully abolished binding of HupA to short dsDNA probes. We conclude that the arginine residues are crucial for the interaction with DNA and that the mode of DNA-binding by HupA is consistent with HU homologs from other organisms [35, 54, 58].

Disruption of DNA binding does not affect oligomerization

HU proteins from various organisms have been found to form homo- or heterodimers [18, 19, 22, 54]. To determine the oligomeric state of *C. difficile* HupA protein, we performed size exclusion chromatography (SEC). The elution profile of the purified protein was compared to molecular weight standards on a Superdex 75 HR 10/30 column. Wild type $HupA_{6xHis}$ protein exhibited a single clear peak with a partition coefficient (Kav) of 0.19 (Fig. 2c). These values correspond to an estimated molecular weight of a 38 kDa, suggesting a multimeric assembly of $HupA_{6xHis}$ (theoretical molecular weight of monomer is 11 kDa). Similar to $HupA_{6xHis}$, $HupA^{QED}{}_{6xHis}$ exhibits only one peak with a Kav of 0.20 and a calculated

molecular weight of 37 kDa (Fig. 2c). Thus, mutation of the residues in the DNA-binding pocket of HupA did not interfere with the ability of HupA to form multimers in solution.

The calculated molecular weight for both proteins is higher than we would expect for a dimer (22 kDa), by analogy with HU proteins from other organisms. However, we cannot exclude the possibility the conformation of the proteins affects the mobility in the size exclusion experiments. Therefore, to further understand the oligomeric state of HupA, we performed glutaraldehyde crosslinking experiments. HupA monomers cross-linked with glutaraldehyde were analyzed by western-blot analysis using anti-his antibodies. Upon addition of glutaraldehyde (0.0006% and 0.006%) we observed an additional signal around 23 kDa (Fig. 2d), consistent with a HupA dimer. No higher order oligomers were observed under the conditions tested. A similar picture was obtained for $HupA^{QED}_{6xHis}$ (Fig. 2d). Together, these experiments support the conclusion that HupA of *C. difficile* is a dimer in solution, similar to other described HU homologs, and that the ability to form dimers is independent of DNA-binding activity.

***HupA self-interacts* in vivo**

Above, we have shown that HupA of *C. difficile* forms dimers *in vitro.* We wanted to confirm that the protein also self-interacts *in vivo*. We developed a split-luciferase system to allow the assessment of protein-protein interactions in *C. difficile*. Our system is based on NanoBiT (Promega)[61] and our previously published codon-optimized variant of Nanoluc, $sLuc^{opt}$ [62]. The system allows to study protein-protein interactions *in vivo* in the native host, and thus present an advantage over heterologous systems. The large (LgBit) and small (SmBit) subunits of this system have been optimized for stability and minimal self-association by substitution of several amino acid residues [61]. When two proteins are tagged with these subunits and interact, the subunits come close enough to form an active luciferase enzyme that is able to generate a bright luminescent signal once substrate is added. We stepwise adapted our $sLuc^{opt}$ reporter [62] by 1) removing the signal sequence (resulting in an intracellular luciferase, Luc^{opt}), 2) introducing the mutations corresponding to

the amino acid substitutions in NanoBiT (resulting in a full length luciferase in which SmBiT and LgBiT are fused, bitLuc^{opt}) and finally, 3) the construction of a modular vector containing a polycistronic construct under the control of the anhydrotetracycline (ATc) -inducible promoter P_{tet} [63] (see Supplemental Methods).

To assess the ability of HupA to form multimers *in vivo*, we genetically fused HupA to the C-terminus of both SmBit and LgBit subunits and expressed them in *C. difficile* under the control of the ATc-inducible promoter. As controls, we assessed luciferase activity in strains that express full length luciferase (bitLuc^{opt}) and combinations of HupA-fusions with or without the individual complementary subunit of the split luciferase (Fig. 3). Expression of the positive control bitLuc^{opt} results in a 2-log increase in luminescence signal after 1 hour of induction (1954024 ± 351395 LU/OD, Fig. 3). When both HupA-fusions are expressed from the same operon a similar increase in the luminescence signal is detected (264646 ± 122518 LU/OD at T1, Fig. 3). This signal is dependent on HupA being fused to both SmBit and LgBiT, as all negative controls demonstrate low levels of luminescence that do not significantly change upon induction (Fig. 3). We conclude that HupA self-interacts *in vivo.*

HupA overexpression leads to a condensed nucleoid

To determine if inducible expression of HupA leads to the condensation of the chromosome in *C. difficile*, we introduced a plasmid carrying *hupA* under the ATc-inducible promoter P_{tet} into strain *630Δerm* [64]. This strain (AP106) also encodes the native *hupA* and induction of the plasmid-borne copy of the gene is expected to result in overproduction of HupA. AP106 cells were induced in exponential growth phase and imaged 1 hour after induction. In wild type or non-induced AP106 cells nucleoids can be seen, after staining with DAPI stain, with a signal spread throughout most of the cytoplasm (Fig. 4a). In some cells a defined nucleoid is observed localized near the cell center (Fig. 4a). This heterogeneity in nucleoid morphology is likely a reflection of the asynchronous growth.

When HupA expression is not induced, the average nucleoid size is 3.10 ± 0.93 µm, similar to wild type *C. difficile* 630Δ*erm* cells (3.32 ± 1.16 µm). Upon induction of HupA

expression a significant decrease in size of the nucleoid is observed (Fig. 4a and b, white arrow). When cells are induced with 50, 100 or 200 ng/mL ATc the average chromosome size was 1.91 ± 0.80 µm; 1.90 ± 0.82 µm and 2.02 ± 0.94 µm, respectively (Fig. 4b). No significant difference was detected between the strains induced with different ATc concentrations (Fig. 4b). The decrease in the nucleoid size when HupA is overexpressed suggests that HupA can compact DNA *in vivo*. This observation is reminiscent of HU overexpression effects reported for other organisms, like *B. subtilis* and *Mycobacterium tuberculosis* [10, 23].

HupA co-localizes with the nucleoid

If HupA indeed exerts an influence on the nucleoid, as suggested by our experiments above, it is expected that the protein co-localizes with the DNA. To test this, we imaged HupA protein and the nucleoid in live *C. difficile*. Here, we use the HaloTag protein (Promega)[65] for imaging the subcellular localization of HupA. Tags that become fluorescent after covalently labeling by small compounds, such as HaloTag, are proven to be useful for studies in bacteria and yeast [66-68]. In contrast to GFP, does not require the presence of oxygen for maturation and should allow live-cell imaging in anaerobic bacteria.

We introduced a modular plasmid expressing HupA-HaloTag from the ATc-inducible promoter P_{tet} [63] into strain 630Δ*erm* [64], yielding strain RD16. Repeated attempts to create a construct that would allow us to integrate the fusion construct on the chromosome of *C. difficile* using allelic exchange failed, likely due to toxicity of the *hupA* upstream region in *E. coli* (cloning intermediate). For the visualization of HupA-HaloTag we used the Oregon green substrate, that emits at Em_{max} 520 nm. Although autofluorescence of *C. difficile* has been observed at wavelengths of 500-550 nm [69, 70] we observed limited to no green signal in the absence of the HaloTag (our unpublished observations and Fig. 5a, -ATc).

HupA-Halotag expression was induced in RD16 cells during exponential growth phase with 200 ng/mL ATc and cells were imaged after 1 hour of induction. In the absence of ATc, no green fluorescent signal is visible, and the nucleoid (stained with DAPI) appears

extended (Fig. 5a). Upon HupA-HaloTag overexpression, the nucleoids are more defined and appear bilobed (Fig. 5a and b), similar to the previous observations (Fig. 4a). The Oregon Green co-localizes with the nucleoid, located in the center of the cells, with a bilobed profile that mirrors the profile of the DAPI stain (Fig. 5a and b). This co-localization is observed for individual cells at different stages of the cell cycle and is independent of the number of nucleoids present (data not shown). The localization pattern of the *C. difficile* HupA resembles that of HU proteins described in other organisms [23, 71, 72] (Fig 5a). Expression levels of HupA-Halotag were confirmed by SDS-PAGE in-gel fluorescence of whole cell extracts, after incubation with Oregon Green. Accumulation of HupA-HaloTag is observed over time (Fig. 5c).

ATc-induced RD16 cells exhibit a heterogeneous Oregon Green fluorescent signal. This has previously been observed with other fluorescent reporters in *C. difficile* [68-70, 73] and can likely be explained by both heterogeneous expression from inducible systems [74] and different stages of the cell cycle. For instance, the localization of cell division proteins, such as MldA or FtsZ is dependent on septum formation and thus dependent on cells undergoing cell division [69, 73].

We found that $HupA^{QED}_{6xHis}$ does not bind dsDNA in the electrophoretic mobility shift assay (Fig. 3b). We introduced the triple substitution in the HupA-HaloTag expression plasmid to determine its effect on localization of the protein in *C. difficile.* We found that the $HupA^{QED}$-HaloTag protein was broadly distributed throughout the cell and no compaction of the nucleoid is observed (Fig. S2). Moreover, unlike observed for ATc-induced RD16 cells (HupA-Halotag), we did not observe clear compaction of the nucleoid.

Together, these results indicate that HupA co-localizes with the nucleoid, and that nucleoid compaction upon HupA overexpression is possibly dependent on its DNA-binding activity. We cannot exclude that the nucleoid compaction observed could be an indirect outcome of HupA overexpression by influencing possible interaction with the RNA and/or other proteins, or by altering transcription/translation [40, 75].

***HupA compacts DNA* in vitro**

To substantiate that the decrease in nucleoid size is directly attributable to the action of HupA, we sought to demonstrate a remodeling effect of HupA on DNA *in vitro*. We performed a ligase-mediated DNA cyclization assay. Previous work has established that dsDNA fragments with a length smaller than 150 bp greatly reduces the possibility of the extremities to meet and therefore less probability to ligate into closed rings [76]. However, in the presence of DNA bending proteins exonuclease III (ExoIII)-resistant (thus closed) rings can be obtained [57, 76].

We tested the ability of $HupA_{6xHis}$ to stimulate cyclization of a [γ-^{32}P]-labeled 123-bp DNA fragment (Fig. 6a). The addition of T4 DNA-ligase alone results in multiple species, corresponding to ExoIII-sensitive linear multimers (Fig. 6a, lane 2 and 3). In the presence of $HupA_{6xHis}$, however, an ExoIII-resistant band is visible (Fig. 6a, lanes 4 to 6). In the absence of ExoIII, the linear dimer is still clearly visible in the HupA-containing samples (Fig. 6a, last lane). We conclude that *C. difficile* HupA is able to bend the DNA, or otherwise stimulate cyclization by increasing flexibility and reducing the distance between the DNA fragment extremities, allowing the ring closure in the presence of ligase.

To more directly demonstrate remodeling of DNA by HupA, we performed tethered particle motion (TPM) experiments. TPM is a single molecule technique that provides a readout of the length and flexibility of a DNA tether (Supplemental Fig. S3)[77]. The binding of proteins to DNA alters its conformation, resulting in a change in RMS (Root Mean Square). If a protein bends DNA, makes DNA more flexible or more compact, the RMS is reduced compared to that of bare DNA, as represented in Supplemental Fig. S3 [77]. If a protein stiffens DNA, the RMS is expected to be larger than that of bare DNA [78].

We performed TPM experiments according to established methods [78] to determine the effects of HupA on DNA conformation at protein concentrations from 0 – 1600 nM (Fig. 6b). For this assay a non-tagged HupA was purified from *C. difficile* cells overexpressing HupA and compared to $HupA_{6xHis}$ to assess potential subtle effects of the 6xhistidine-tag on the protein functionality. The experiments show that binding of both native HupA and

$HupA_{6xHis}$ to DNA reduces the RMS. The RMS of bare DNA is 148 ± 1.9 nm. In the presence of HupA at different concentrations (100, 200, 400 nM) the RMS decreases (113 ± 0.1 nm; 103 ± 0.7 nm and 97 ± 1.5 nm respectively). Even at higher concentrations of HupA (800, 1600 nM) the RMS is 97-100 nm (Fig. 6b). The strongly reduced RMS of DNA bound by non-tagged HupA at 1600 nM suggests a more compacted conformation of DNA compared to that of bare DNA. The curves are overall highly similar for HupA and $HupA_{6xHis}$ proteins; the small difference in the observed effects is attributed to interference of the tag and/or protein stability.

The effects of *C. difficile* HupA of *C. difficile* on DNA topology observed by TPM indicates similar structural properties to those of *E. coli* HU, which was shown to compact DNA by bending at low protein coverage [26, 79, 80]. However, in contrast to *E. coli* [26], there is no clear stiffening of the DNA tether at high concentrations of protein in our assay, suggesting that there is lower or reduced dimer-dimer interaction in our experimental condition. Bending of DNA by HU proteins has also been shown for other organisms. Interestingly, in *B. burgdorferi* [56] and *Anabaena* [35] it was shown that bending is influenced by interaction of the DNA with a positively charged lateral surface, although the main interaction region with the DNA is through the β-arms. *C. difficile* HupA demonstrates an electrostatic surface potential compatible with such a mechanism (Fig 1C). It will be of interest to determine if and which residues in this region contribute to the bending of the DNA.

Conclusions

In this work, we present the first characterization of a bacterial chromatin protein in *C. difficile*. HupA is a member of the HU family of proteins and is capable of binding DNA and does so without an obvious difference in affinity as a result of the G+C content. DNA binding is dependent on the residues R55, R58 and R61 that are located in the predicted β-arm of the protein. These observations in combination with the predicted structure suggest a

conserved mode of DNA binding, although the role of other regions of the protein in DNA binding is still poorly understood. HupA is present as a dimer in solution and disruption of the residues of the DNA binding domain did not affect the oligomeric state of HupA.

In *C. difficile* we co-localized HupA with the nucleoid and demonstrated that overexpression of HupA leads to nucleoid compaction. In line with these observations, HupA stimulates the cyclization of a short dsDNA fragment and compacts DNA *in vitro*.

We also developed a new complementation assay for the detection of protein-protein interactions in *C. difficile,* complementing the available tools for this organism, and confirmed that HupA self-interacts *in vivo*. Additionally, to our knowledge, our study is the first to describe the use of the fluorescent tag HaloTag for imaging the subcellular localization of proteins in live *C. difficile* cells.

In sum, HupA of *C. difficile* is an essential bacterial chromatin protein required for nucleoid (re)modelling. HupA binding induce bending or increase the flexibility of the DNA, resulting in compaction. Further research is required to explore the roles of HupA in the physiology of *C. difficile*.

Methods

Sequence Alignments and Structural Modelling

Multiple sequence alignment of amino acid sequences was performed with Clustal Omega [81]. The sequences of HU proteins identified in *C. difficile* 630Δ*erm* (Q180Z4), *E. coli* (P0ACF0 and P0ACF4), *Bacillus subtilis* (A3F3E2), *Geobacillus stearothermophilus* (P0A3H0), *Bacillus anthracis* (Q81WV7), *Staphylococcus aureus* (Q99U17), *Salmonella typhimurium* (P0A1R8), *Streptococcus pneumoniae* (AAK75224), *S. mutans* (Q9XB21), *M. tuberculosis* (P9WMK7), *Thermotoga maritima* (P36206) and *Anabaena sp.* (P05514), were selected for alignment. Amino acid sequences were retrieved from the Uniprot database.

Homology modelling was performed using PHYRE2 (http://www.sbg.bio.ic.ac.uk/phyre2, [52] and SWISS-MODEL [53] using default settings. For SWISS-MODEL, PDB 4QJN was used as a template. Selection of the template was based on PHYRE2 results, sequence identity (59,55%) and best QSQE (0,80) and GMQE (0,81). Graphical representations and mutation analysis were performed with the PyMOL Molecular Graphics System, Version 1.76.6. Schrödinger, LLC. For electrostatics calculations APBS (Adaptive Poisson-Boltzmann Solver) and PDB2PQR software packages were used [82]. Default settings were used.

Strains and growth conditions

E. coli strains were cultured in Luria Bertani broth (LB, Affymetrix) supplemented with chloramphenicol at 15 µg/mL or 50 µg/mL kanamycin when appropriate, grown aerobically at 37°C. Plasmids (Table 1) were maintained in *E. coli* strain DH5α. Plasmids were transformed using standard procedures [83]. *E. coli* strain Rosetta (DE3) (Novagen) was used for protein expression and *E. coli* CA434 for plasmid conjugation [84] with *C. difficile* strain 630Δ*erm* [64, 85].

C. difficile strains were cultured in Brain Heart Infusion broth (BHI, Oxoid), with 0,5 % w/v yeast extract (Sigma-Aldrich), supplemented with 15 µg/mL thiamphenicol and

Clostridium difficile Selective Supplement (CDSS; Oxoid) when necessary. *C. difficile* strains were grown anaerobically in a Don Whitley VA-1000 workstation or a Baker Ruskinn Concept 1000 workstation with an atmosphere of 10% H_2, 10% CO_2 and 80% N_2.

The growth was followed by optical density reading at 600 nm. All the *C. difficile* strains are described in Table 2.

Construction of the E. coli expression vectors

All oligonucleotides and plasmids from this study are listed in Tables 1 and 3.

To construct an expression vector for $HupA_{6xHis}$, the *hupA* gene (CD3496 from *C. difficile* 630 GenBank accession no. NC_009089.1) was amplified by PCR from *C. difficile* 630Δ*erm* genomic DNA using primers oAF57 and oAF58 (Table 3). The product was inserted into the *Nco*I-*Xho*I digested pET28b vector (Table 1) placing it under control of the T7 promoter, yielding plasmid pAF226.

To generate the HupA triple mutant ($HupA^{QED}_{6xHis}$) site-directed mutagenesis was used according to the QuikChange protocol (Stratagene). Initially the arginine at position 55 and at position 58 were simultaneously substituted for glutamine (R55Q) and glutamic acid (R58E) respectively, using primers oAF73/oAF74 (Table 3), resulting in pAF232 (Table 1). The arginine at position 61 was subsequently substituted for aspartic acid (R61D) using primer pair oAF75/oAF76 (Table 3) and pAF232 as a template, yielding pAF234 (Table 1). All the constructs were confirmed by Sanger sequencing.

Construction of the C. difficile expression vectors

To overexpress non-tagged HupA the *hupA* gene was amplified by PCR from *C. difficile 630*Δ*erm* genomic DNA using primers oWKS-1519 and oAP47 (Table 3) and cloned into SacI-BamHI digested pRPF185 vector [63], placing it under control of the ATc-inducible promoter P_{tet}, yielding vector pAP103 (Table 1).

For microscopy experiments HaloTagged protein (Promega) was used. The *halotag* gene was amplified from vector pH6HTC (Promega, GenBank Accession no. **JN874647**) with primers oWKS-1511/oWKS-1512 and inserted into pCR2.1-TOPO according to the instructions of the manufacturer (ThermoFisher), yielding vector pWKS1746 (Table 1). This primer combination also introduces a 6xHis-tag at the C-terminus of the HaloTag. The *hupA* gene was amplified with primers oWKS-1519/oWKS-1520 (Table 3) and inserted into vector pCR2.1-TOPO according to the instructions of the manufacturer (ThermoFisher), generating vector pWKS1744 (Table 1). The primers introduces the *cwp2* ribosomal binding site upstream and a short DNA sequence encoding a GS-linker downstream (SGSGSGS) of the *hupA* open reading frame. To generate the expression construct for HupA-Halotag the open reading frame encoding the $HaloTag_{6xHis}$ protein was amplified from pWKS1746 using primers oRD5/oWKS-1512 (Table 3). The *hupA* gene was amplified from pWKS1744 with primers oWKS-1519/oWKS-1520 (Table 3). Gene fusions were made by overlapping PCR using the PCR amplified fragments encoding HupA and Halotag proteins as templates with primers oWKS-1519 and oWKS-1512 (Table 3). The fragment was cloned into SacI-BamHI digested pRPF185 [63], placing it under control of the ATc inducible promoter P_{tet}, yielding vector pRD4 (Table 1).

To generate the HupA triple mutant fused to the Halotag ($HupA^{QED}$-Halotag) site-directed mutagenesis was used, according to the QuikChange protocol (Stratagene). The arginines at position 55 and at position 58 were substituted to glutamine (R55Q) and glutamic acid (R58E), using primers oAF73/oAF74 (Table 3) and pRD4 as template, resulting in pAF235 (Table 1). The arginine at position 61 was subsequently substituted to aspartic acid (R61D), using pAF235 as template and primers oAF75/oAF76 (Table 3), yielding pAF237 (Table 1). All the constructs were confirmed by Sanger sequencing.

Construction of the bitLuc^{opt} expression vectors

The bitLuc^{opt} complementation assay for *C. difficile* described in this study is based on NanoBiT (Promega) [61] and the codon optimized sequence of sLuc^{opt} [62]. Details of its construction can be found in Supplemental Material.

Gene synthesis was performed by Integrated DNA Technologies, Inc. (IDT). Fragments were amplified by PCR from synthesized dsDNA, assembled by Gibson assembly [86] and cloned into *Sac*I/*BamH*I digested pRPF185 [63], placing them under control of the ATc-inducible promoter P_{tet}. As controls a non-secreted luciferase (Luc^{opt}; pAF254) and a luciferase with the NanoBiT aminoacid substitutions (Promega)[61](bitLuc^{opt}; pAF259) were constructed. We also constructed vectors expressing only the SmBiT and LgBiT domains, alone (pAF260 and pAF255) or in combination (pAF262), as controls.

To assay for a possible interaction between HupA monomers, vectors were constructed that encode HupA-SmBiT/HupA-LgBiT (pAP118), HupA-SmBiT/LgBiT (pAF256), SmBiT/HupA-LgBiT (pAF257). DNA sequences of the cloned DNA fragments in all recombinant plasmids were verified by Sanger sequencing.

Note that all our constructs use the HupA start codon (GTG) rather than ATG; a minimal set of vectors necessary to perform the *C. difficile* complementation assay (pAP118, pAF256, pAF257 and pAF258) is available from Addgene (105494-105497) for the *C. difficile* research community.

Overproduction and purification of $HupA_{6xHis}$ *and* $HupA^{QED}{}_{6xHis}$

Overexpression of $HupA_{6xHis}$ and $HupA^{QED}{}_{6xHis}$ was carried out in *E. coli* strains Rosetta (DE3)(Novagen) harbouring the *E. coli* expression plasmids pAF226 and pAF234, respectively. Cells were grown in LB and induced with 1mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at an optical density (OD_{600}) of 0,6 for 3 hours. The cells were collected by centrifugation at 4°C and stored at -80°C.

Pellets were suspended in lysis buffer (50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM imidazole, 5 mM β mercaptoethanol, 0.1% NP40 and Complete protease inhibitor cocktail (CPIC, Roche Applied Science). Cells were lysed by the addition of 1 mg/ml

lysozyme and sonication. The crude lysate was clarified by centrifugation at 13000 g at 4°C for 20 min. The supernatant containing recombinant proteins was collected and purification was performed with TALON Superflow resin (GE Healthcare) according to manufacturer instructions. Proteins were stored at -80°C in 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl and 12% glycerol.

Overproduction and purification of HupA

Overexpression of HupA was carried out in *C. difficile* strain AP106 that carries the plasmid encoding HupA under the ATc-inducible promoter P_{tet}. Cells were grown until OD_{600} 0.4-0.5 and induced with 200 ng/mL ATc for 3 hours. Cells were collected by centrifugation at 4°C.

Pellets were resuspended in HB buffer (25 mM Tris (pH 8.0), 0.1 mM EDTA, 5 mM β mercaptoethanol, 10% glycerol and CPIC). Cells were lysed by French Press and phenylmethylsulfonyl fluoride was added to 0.1 mM. Separation of the soluble fraction was performed by centrifugation at 13000g at 4°C for 20 min. Purification of the protein from the soluble fraction was done on a 1 mL HiTrap SP (GE Healthcare) according to manufacturer instructions. The protein was collected in HB buffer supplemented with 300 mM NaCl. Fractions containing the HupA protein were pooled together and applied to a 1 mL Heparin Column (GE Healthcare) according to manufacturer's instructions. Column washes were performed with a 500 mM – 800 mM NaCl gradient in HB buffer. Proteins were eluted in HB buffer supplemented with 1 M NaCl and stored in 10% glycerol at -80°C.

DNA labelling and electrophoretic mobility shift Assay (EMSA)

For the gel shift assays double stranded oligonucleotides with different GC contents were used. Oligonucleotides oAF61/oAF62 have a 71,1% G+C-content, oAF63/oAF64 have a 52,6% G+C-content and oAF65/oAF66 have a 28,9% G+C-content. The oligonucleotides were labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase (PNK) (Invitrogen) according to the PNK-manufacturer's instructions. The fragments were purified with a Biospin P-30 Tris

column (BioRad). Oligonucleotides with same G+C content were annealed by incubating them at 95˚C for 10 min, followed by ramping to room temperature.

Gel shift assays were performed with increasing concentrations (0,25 -2 µM) of $HupA_{6xHis}$ or $HupA^{QED}_{6xHis}$ in a buffer containing 20 mM Tris pH 8.0; 50 mM NaCl; 12 mM $MgCl_2$; 2.5 mM ATP; 2 mM DTT; 10% glycerol and 2.4 nM [γ-^{32}P]ATP-labelled oligonucleotides. Proteins were incubated with the oligonucleotide substrate for 20 min at room temperature prior to separation. Reactions were analyzed in 8% native polyacrylamide gels in cold 0,5X TBE buffer supplemented with 10 mM $MgCl_2$. After electrophoresis gels were dried under vacuum and protein-DNA complexes were visualized by phosphorimaging (Typhoon 9410 scanner; GE Healthcare). Analysis was performed with Quantity-One software (BioRad).

Size-exclusion chromatography (SEC)

Size-exclusion experiments were performed on an Äkta pure 25L1 instrument (GE Healthcare). 200 µL of $HupA_{6xHis}$ and $HupA^{QED}_{6xHis}$ was applied at a concentration of ~100 µM, to a Superdex 75 HR 10/30 column (GE Healthcare), in buffer containing 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl and 12% glycerol. UV detection was done at 280 nm. Lower concentrations of HupA were not possible to analyses due to the lack of signal. HupA protein only contains 3 aromatic residues, and lacks His, Trp, Tyr or Cys to allow detection by absorbance at 280 nm. The column was calibrated with a mixture of proteins of known molecular weights (Mw): Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic anhydrase (29 kDa), Ribonuclease A (13,7 kDa), and Aprotinin (6,5 kDa). Molecular weight of the HupA proteins was estimated according to the equation MW=10(Kav-b)/m where m and b correspond to the slope and the linear coefficient of the plot of the logarithm of the MW as a function of the Kav. The Kav is given by the equation $Kav=(V_e-V_0)/(V_t-V_0)$ [87], where V_e is the elution volume for a given concentration of protein, V_0 is the void volume (corresponding to the elution volume of thyroglobulin), and V_t is the total column volume (estimated from the elution volume of a 4% acetone solution).

Glutaraldehyde cross-linking Assay

100 ng HupA protein was incubated with different concentrations of glutaraldehyde (0 0,006%) for 30 min at room temperature. Reactions were quenched with 10 mM Tris. The samples were loaded on a 6.5% SDS-PAGE gel and analysed by western-blotting. The membrane was probed with a mouse anti-His antibody (Thermo Fisher) 1:3000 in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4) with 0.05% Tween-20 and 5% w/v Milk (Campina), a secondary anti-mouse HRP antibody 1:3000 and Pierce ECL2 Western blotting substrate (Thermo Scientific). A Typhoon 9410 scanner (GE Healthcare) was used to record the chemiluminescent signal.

bitLuc^{opt} Assay

For the *C. difficile* complementation assay, cells were grown until OD_{600} 0.3-0.4 and induced with 200 ng/mL anhydrotetracline for 60 min. To measure luciferase activity 20 µL NanoGlo Luciferase (Promega N1110) was added to 100 µL of culture sample. Measurements were performed in triplicate in a 96-well white F-bottom plate according to manufacturer's instructions. Luciferase activity was determined using a GloMax instrument (Promega) for 0.1 s. Data was normalized to culture optical density measured at 600 nm (OD_{600}). Statistical analysis was performed with Prism 7 (GraphPad, Inc, La Jolla, CA) by two-way ANOVA.

Ligase-mediated cyclization assay

A 119 bp DNA fragment was amplified by PCR amplification with primers oAF81/oAF82, using pRPF185 plasmid as a template. The PCR fragment was digested with *EcoR*I and 5'end labelled with [ɣ3^{2}P] ATP using T4 polynucloetide kinase (Invitrogen) according to the manufacturer's instructions. Free ATP was removed with a Biospin P-30 Tris column (BioRad).

The labelled DNA fragment (~ 0.5 nM) waspAF235 incubated with different concentrations of HupA for 30 min on 30°C in 50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, and 0.5 mM ATP in a total volume of 10 µl. 1 Unit of T4 ligase was added and incubated for 1 h at 30°C followed by inactivation for 15 minutes at 65°C. When appropriate, samples were treated with 100 U of Exonuclease III (Promega) at 37°C for 30 minutes. Enzyme inactivation was performed by incubating the samples for 15 minutes at 65°C. Before electrophoresis the samples were digested with 2 µg proteinase K and 0.2% SDS at 37°C for 30 minutes. Samples were applied to a pre-run 7% polyacrylamide gel in 0.5X TBE buffer with 2% glycerol and run at 100V for 85 min. After electrophoresis the gel was vacuum-dried and analysed by phosphor imaging. Analysis was performed with Quantity-One software (BioRad).

Fluorescence microscopy

The sample preparation for fluorescence microscopy was carried out under anaerobic conditions. *C. difficile* strains were cultured in BHI/YE, and when appropriate induced with different ATc concentrations (50, 100 and 200 ng/mL) for 1 hour at an OD_{600} of 0.3-0.4. When required, cells where incubated with 150 nM Oregon Green substrate for HaloTag (Promega) for 30 min. 1 mL culture was collected and washed with pre-reduced PBS. Cells were incubated with 1 µM DAPI (Roth) when necessary. Cells were spotted on 1.5% agarose patches with 1 µL of ProLong Gold antifading mountant (Invitrogen). Slides were sealed with nail polish.

Samples were imaged with a Leica DM6000 DM6B fluorescence microscope (Leica) equipped with DFC9000 GT sCMOS camera using a HC PLAN APO 100x/1.4 OIL PH3 objective, using the LAS X software. The filter set for imaging DAPI is the DAPI ET filter (n. 11504203, Leica), with excitation filter 350/50 (band pass), long pass dichroic mirror 400 and emission filter 460/50 (band pass). For imaging of Oregon Green the filter L5 ET was used (n. 11504166, Leica), with excitation filter 480/40, dichroic mirror 505 and emission filter 527/30.

Data was analysed with MicrobeJ package version 5.12d [88] with ImageJ 1.52d software [89]. Recognition of cells was limited to 2 - 16 µm length. For the nucleoid and Halotag detection the nucleoid feature was used for the nucleoid length and fluorescent analysis. Cells with more than 2 identified nucleoids and defective detection were excluded from analysis. Statistical analysis was performed with MicrobeJ package version 5.12d [88].

In-gel fluorescence

C. difficile strains were cultured in BHI/YE, and when appropriate induced at an OD_{600} of 0.3-0.4 with 200 ng/mL ATc concentrations for up to 3 hours. Samples were collected and centrifuged at 4°C. Pellets were resuspended in PBS and lysed by French Press. Samples were incubated with 150 nM Oregon Green substrate for HaloTag (Promega) for 30 minutes at 37°C. Loading buffer (250 mM Tris-Cl pH 6.8, 10 % SDS, 10% β-mercaptoethanol, 50% glycerol, 0.1% bromophenol blue) was added to the samples without boiling and samples were run on 12% SDS-PAGE gels. Gels were imaged with Uvitec Alliance Q9 Advanced machine (Uvitec) with F-535 filter (460 nm).

Tethered Particle Motion measurements

A dsDNA fragment of 685bp with 32% G+C content (sso685) was used for Tethered Particle Motion experiments. This substrate was generated by PCR using the forward biotin-labelled primer Sso10a-2Nde and the reverse digoxygenin (DIG) labelled primer Sso10a-2Bam685 from pRD118 as previously described [90]. The PCR product was purified using the GenElute PCR Clean-up kit (Sigma-Aldrich).

Tethered Particle Motion (TPM) measurements were done as described previously [77, 78] with minor modifications. In short, anti-digoxygenin (20 µg/mL) was flushed into the flow cell and incubated for 10 minutes to allow the anti-digoxygenin to attach to the glass surface. To block unspecific binding to the glass surface, the flow cell was incubated with BSA and BGB (Blotting grade Blocker) in buffer A (10 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 3% glycerol, and 100 µg/mL acetylated BSA, 0.4% BGB) for 10

minutes. To tether DNA to the surface, DNA (labeled with Biotin and DIG) diluted in buffer B (10 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 3% glycerol, and 100 µg/mL acetylated BSA) was flushed into the flow cell and incubated for 10 minutes. Streptavidin-coated polystyrene beads (0,44 µm in diameter) diluted in buffer B were introduced into the sample chamber and incubated for at least 10 minutes to allow binding to the biotin-labelled DNA ends. Before flushing in the protein in buffer C (20 mM HEPES (pH 7.9), 60 mM KCl, and 0.2% (w/v) BGB), the flow cell was washed twice with buffer C to remove free beads. Finally the flow cell was sealed, followed by incubation with protein or experimental buffer for 10 minutes. The measurements were started after 6 minutes of further incubation of the flow cell at a constant temperature of 25 °C. More than 300 beads were measured for each individual experiment. All experiments were performed at least in duplicate.

The analysis of the TPM data was performed as previously described [78]. Equation 1 was used to calculate the RMS of the individual beads.

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

where x and y are the coordinates of the beads, $\overline{x}$ and $\overline{y}$ are averaged over the full-time trace. The RMS value of each measured condition was acquired by fitting a Gaussian to the histogram of the RMS values of individual beads.

All the pictures were prepared for publication in CorelDRAW X8 (Corel).

Acknowledgements

Work in the group of WKS is supported by a Vidi Fellowship (864.10.003) of the Netherlands Organization for Scientific Research (NWO) and a Gisela Their Fellowship from the Leiden University Medical Center. Work in the group of RTD is supported by a VICI grant (016.160.613) of the Netherlands Organization for Scientific Research (NWO), Human Frontier Science Program [RGP0014/2014] and a China Scholarship Council to L.Q. [201506880001].

We thank Dr. Jeroen Corver for helpful discussions.

Conflict of interest

WKS has performed research for Cubist. The company had no role in the design or interpretation of these experiments or the decision to publish. Others: none to declare.

References

[1] Lawson PA, Citron DM, Tyrrell KL, Finegold SM. Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prevot 1938. Anaerobe. 2016;40:95-9.
[2] Hensgens MP, Keessen EC, Squire MM, Riley TV, Koene MG, de Boer E, et al. Clostridium difficile infection in the community: a zoonotic disease? Clin Microbiol Infect. 2012;18:635-45.
[3] Warriner K, Xu C, Habash M, Sultan S, Weese SJ. Dissemination of Clostridium difficile in food and the environment: Significant sources of C. difficile community-acquired infection? J Appl Microbiol. 2017;122:542-53.
[4] Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. Clostridium difficile infection. Nature Reviews Disease Primers. 2016;2:16020.
[5] Crobach MJT, Vernon JJ, Loo VG, Kong LY, Pechine S, Wilcox MH, et al. Understanding Clostridium difficile Colonization. Clin Microbiol Rev. 2018;31.
[6] Pituch H. Clostridium difficile is no longer just a nosocomial infection or an infection of adults. Int J Antimicrob Agents. 2009;33 Suppl 1:S42-5.
[7] Spigaglia P. Recent advances in the understanding of antibiotic resistance in Clostridium difficile infection. Ther Adv Infect Dis. 2016;3:23-42.
[8] Peng Z, Jin D, Kim HB, Stratton CW, Wu B, Tang YW, et al. Update on Antimicrobial Resistance in Clostridium difficile: Resistance Mechanisms and Antimicrobial Susceptibility Testing. J Clin Microbiol. 2017;55:1998-2008.
[9] Ehmann DE, Lahiri SD. Novel compounds targeting bacterial DNA topoisomerase/DNA gyrase. Curr Opin Pharmacol. 2014;18:76-83.
[10] Bhowmick T, Ghosh S, Dixit K, Ganesan V, Ramagopal UA, Dey D, et al. Targeting Mycobacterium tuberculosis nucleoid-associated protein HU with structure-based inhibitors. Nat Commun. 2014;5:4124.
[11] van Eijk E, Wittekoek B, Kuijper EJ, Smits WK. DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens. J Antimicrob Chemother. 2017;72:1275-84.
[12] Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. Nat Rev Mol Cell Biol. 2002;3:430-40.
[13] Pelletier J, Halvorsen K, Ha BY, Paparcone R, Sandler SJ, Woldringh CL, et al. Physical manipulation of the Escherichia coli chromosome reveals its soft nature. Proc Natl Acad Sci U S A. 2012;109:E2649-56.
[14] Zimmerman SB, Murphy LD. Macromolecular crowding and the mandatory condensation of DNA in bacteria. FEBS Lett. 1996;390:245-8.
[15] Dame RT, Tark-Dame M. Bacterial chromatin: converging views at different scales. Current Opinion in Cell Biology. 2016;40:60-5.
[16] Dillon SC, Dorman CJ. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nat Rev Microbiol. 2010;8:185-95.

[17] Ohniwa RL, Ushijima Y, Saito S, Morikawa K. Proteomic analyses of nucleoid-associated proteins in Escherichia coli, Pseudomonas aeruginosa, Bacillus subtilis, and Staphylococcus aureus. PloS one. 2011;6:e19172.
[18] Grove A. Functional evolution of bacterial histone-like HU proteins. Curr Issues Mol Biol. 2011;13:1-12.
[19] Kamashev D, Agapova Y, Rastorguev S, Talyzina AA, Boyko KM, Korzhenevskiy DA, et al. Comparison of histone-like HU protein DNA-binding properties and HU/IHF protein sequence alignment. PloS one. 2017;12:e0188037.
[20] Boubrik F, Bonnefoy E, Rouviere-Yaniv J. HU and IHF: similarities and differences. In Escherichia coli, the lack of HU is not compensated for by IHF. Res Microbiol. 1991;142:239-47.
[21] Claret L, Rouviere-Yaniv J. Variation in HU composition during growth of Escherichia coli: the heterodimer is required for long term survival. Journal of molecular biology. 1997;273:93-104.
[22] Guo F, Adhya S. Spiral structure of Escherichia coli HUalphabeta provides foundation for DNA supercoiling. Proc Natl Acad Sci U S A. 2007;104:4309-14.
[23] Kohler P, Marahiel MA. Association of the histone-like protein HBsu with the nucleoid of Bacillus subtilis. Journal of bacteriology. 1997;179:2060-4.
[24] Giangrossi M, Giuliodori AM, Gualerzi CO, Pon CL. Selective expression of the beta-subunit of nucleoid-associated protein HU during cold shock in Escherichia coli. Mol Microbiol. 2002;44:205-16.
[25] Painbeni E, Caroff M, Rouviere-Yaniv J. Alterations of the outer membrane composition in Escherichia coli lacking the histone-like protein HU. Proc Natl Acad Sci U S A. 1997;94:6712-7.
[26] van Noort J, Verbrugge S, Goosen N, Dekker C, Dame RT. Dual architectural roles of HU: formation of flexible hinges and rigid filaments. Proc Natl Acad Sci U S A. 2004;101:6969-74.
[27] Dame RT, Hall MA, Wang MD. Single-molecule unzipping force analysis of HU-DNA complexes. Chembiochem. 2013;14:1954-7.
[28] Balandina A, Kamashev D, Rouviere-Yaniv J. The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. J Biol Chem. 2002;277:27622-8.
[29] Grainger DC, Hurd D, Goldberg MD, Busby SJ. Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. Nucleic Acids Res. 2006;34:4642-52.
[30] Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, et al. Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. Science. 2006;313:236-8.
[31] Gordon BR, Li Y, Cote A, Weirauch MT, Ding P, Hughes TR, et al. Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. Proc Natl Acad Sci U S A. 2011;108:10690-5.
[32] Kohler P, Marahiel MA. Mutational analysis of the nucleoid-associated protein HBsu of Bacillus subtilis. Mol Gen Genet. 1998;260:487-91.
[33] Bensaid A, Almeida A, Drlica K, Rouviere-Yaniv J. Cross-talk between topoisomerase I and HU in Escherichia coli. Journal of molecular biology. 1996;256:292-300.
[34] Dame RT, Goosen N. HU: promoting or counteracting DNA compaction? FEBS Lett. 2002;529:151-6.
[35] Swinger KK, Lemberg KM, Zhang Y, Rice PA. Flexible DNA bending in HU-DNA cocrystal structures. EMBO J. 2003;22:3749-60.
[36] Aki T, Choy HE, Adhya S. Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by GAL repressor. Genes Cells. 1996;1:179-88.
[37] Morales P, Rouviere-Yaniv J, Dreyfus M. The histone-like protein HU does not obstruct movement of T7 RNA polymerase in Escherichia coli cells but stimulates its activity. Journal of bacteriology. 2002;184:1565-70.
[38] Oberto J, Nabti S, Jooste V, Mignot H, Rouviere-Yaniv J. The HU regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. PloS one. 2009;4:e4367.

758 [39] Priyadarshini R, Cugini C, Arndt A, Chen T, Tjokro NO, Goodman SD, et al. The nucleoid-
759 associated protein HUbeta affects global gene expression in Porphyromonas gingivalis. Microbiology.
760 2013;159:219-29.
761 [40] Berger M, Farcas A, Geertz M, Zhelyazkova P, Brix K, Travers A, et al. Coordination of genomic
762 structure and transcription by the main bacterial nucleoid-associated protein HU. EMBO reports.
763 2010;11:59-64.
764 [41] Abebe AH, Aranovich A, Fishov I. HU content and dynamics in Escherichia coli during the cell
765 cycle and at different growth rates. FEMS Microbiol Lett. 2017;364.
766 [42] Lewis DEA, Geanacopoulos M, Adhya S. Role of HU and DNA supercoiling in transcription
767 repression: specialized nucleoprotein repression complex at gal promoters in Escherichia coli.
768 Molecular Microbiology. 1999;31:451-61.
769 [43] Chodavarapu S, Felczak MM, Yaniv JR, Kaguni JM. Escherichia coli DnaA interacts with HU in
770 initiation at the E. coli replication origin. Mol Microbiol. 2008;67:781-92.
771 [44] Jaffe A, Vinella D, D'Ari R. The Escherichia coli histone-like protein HU affects DNA initiation,
772 chromosome partitioning via MukB, and cell division via MinCDE. Journal of bacteriology.
773 1997;179:3494-9.
774 [45] Mangan MW, Lucchini S, T OC, Fitzgerald S, Hinton JC, Dorman CJ. Nucleoid-associated protein
775 HU controls three regulons that coordinate virulence, response to stress and general physiology in
776 Salmonella enterica serovar Typhimurium. Microbiology. 2011;157:1075-87.
777 [46] Alvarez A, Toledo H. The histone-like protein HU has a role in gene expression during the acid
778 adaptation response in Helicobacter pylori. Helicobacter. 2017;22.
779 [47] Sakatos A, Babunovic GH, Chase MR, Dills A, Leszyk J, Rosebrock T, et al. Posttranslational
780 modification of a histone-like protein regulates phenotypic resistance to isoniazid in mycobacteria.
781 Sci Adv. 2018;4:eaao1478.
782 [48] Ferrandiz MJ, Carreno D, Ayora S, de la Campa AG. HU of Streptococcus pneumoniae Is Essential
783 for the Preservation of DNA Supercoiling. Front Microbiol. 2018;9:493.
784 [49] Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, et al. The multidrug-
785 resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. Nature genetics.
786 2006;38:779-86.
787 [50] Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, et al. High-throughput analysis
788 of gene essentiality and sporulation in Clostridium difficile. MBio. 2015;6:e02383.
789 [51] Swinger KK, Rice PA. Structure-based analysis of HU-DNA binding. Journal of molecular biology.
790 2007;365:1005-16.
791 [52] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein
792 modeling, prediction and analysis. Nat Protoc. 2015;10:845-58.
793 [53] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL:
794 homology modelling of protein structures and complexes. Nucleic Acids Res. 2018.
795 [54] Kim DH, Im H, Jee JG, Jang SB, Yoon HJ, Kwon AR, et al. beta-Arm flexibility of HU from
796 Staphylococcus aureus dictates the DNA-binding and recognition mechanism. Acta Crystallogr D Biol
797 Crystallogr. 2014;70:3273-89.
798 [55] Hammel M, Amlanjyoti D, Reyes FE, Chen JH, Parpana R, Tang HY, et al. HU multimerization shift
799 controls nucleoid compaction. Sci Adv. 2016;2:e1600650.
800 [56] Mouw KW, Rice PA. Shaping the Borrelia burgdorferi genome: crystal structure and binding
801 properties of the DNA-bending protein Hbb. Mol Microbiol. 2007;63:1319-30.
802 [57] Kamau E, Tsihlis ND, Simmons LA, Grove A. Surface salt bridges modulate the DNA site size of
803 bacterial histone-like HU proteins. Biochem J. 2005;390:49-55.
804 [58] Saitoh F, Kawamura S, Yamasaki N, Tanaka I, Kimura M. Arginine-55 in the beta-arm is essential
805 for the activity of DNA-binding protein HU from Bacillus stearothermophilus. Biosci Biotechnol
806 Biochem. 1999;63:2232-5.
807 [59] Goshima N, Kohno K, Imamoto F, Kano Y. HU-1 mutants of Escherichia coli deficient in DNA
808 binding. Gene. 1990;96:141-5.

809 [60] Smits WK, Grossman AD. The transcriptional regulator Rok binds A+T-rich DNA and is involved in
810 repression of a mobile genetic element in Bacillus subtilis. PLoS Genet. 2010;6:e1001207.
811 [61] Dixon AS, Schwinn MK, Hall MP, Zimmerman K, Otto P, Lubben TH, et al. NanoLuc
812 Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells.
813 ACS Chem Biol. 2016;11:400-8.
814 [62] Oliveira Paiva AM, Friggen AH, Hossein-Javaheri S, Smits WK. The Signal Sequence of the
815 Abundant Extracellular Metalloprotease PPEP-1 Can Be Used to Secrete Synthetic Reporter Proteins
816 in Clostridium difficile. ACS Synth Biol. 2016.
817 [63] Fagan RP, Fairweather NF. Clostridium difficile has two parallel and essential Sec secretion
818 systems. Journal of Biological Chemistry. 2011;286:27483-93.
819 [64] van Eijk E, Anvar SY, Browne HP, Leung WY, Frank J, Schmitz AM, et al. Complete genome
820 sequence of the Clostridium difficile laboratory strain 630Deltaerm reveals differences from strain
821 630, including translocation of the mobile element CTn5. BMC genomics. 2015;16:31.
822 [65] Los GV, Encell LP, McDougal MG, Hartzell DD, Karassina N, Zimprich C, et al. HaloTag: A Novel
823 Protein Labeling Technology for Cell Imaging and Protein Analysis. American Chemical Society. 2010.
824 [66] Stagge F, Mitronova GY, Belov VN, Wurm CA, Jakobs S. SNAP-, CLIP- and Halo-tag labelling of
825 budding yeast cells. PloS one. 2013;8:e78745.
826 [67] Ke N, Landgraf D, Paulsson J, Berkmen M. Visualization of Periplasmic and Cytoplasmic Proteins
827 with a Self-Labeling Protein Tag. Journal of bacteriology. 2016;198:1035-43.
828 [68] Cassona CP, Pereira F, Serrano M, Henriques AO. A Fluorescent Reporter for Single Cell Analysis
829 of Gene Expression in Clostridium difficile. Methods Mol Biol. 2016;1476:69-90.
830 [69] Buckley AM, Jukes C, Candlish D, Irvine JJ, Spencer J, Fagan RP, et al. Lighting Up Clostridium
831 Difficile: Reporting Gene Expression Using Fluorescent Lov Domains. Sci Rep. 2016;6:23463.
832 [70] Ransom EM, Ellermeier CD, Weiss DS. Use of mCherry Red fluorescent protein for studies of
833 protein localization and gene expression in Clostridium difficile. Applied and environmental
834 microbiology. 2015;81:1652-60.
835 [71] Wery M, Woldringh CL, Rouviere-Yaniv J. HU-GFP and DAPI co-localize on the Escherichia coli
836 nucleoid. Biochimie. 2001;83:193-200.
837 [72] Holowka J, Trojanowski D, Ginda K, Wojtas B, Gielniewski B, Jakimowicz D, et al. HupB Is a
838 Bacterial Nucleoid-Associated Protein with an Indispensable Eukaryotic-Like Tail. MBio. 2017;8.
839 [73] Ransom EM, Williams KB, Weiss DS, Ellermeier CD. Identification and characterization of a gene
840 cluster required for proper rod shape, cell division, and pathogenesis in Clostridium difficile. Journal
841 of bacteriology. 2014;196:2290-300.
842 [74] Binder D, Probst C, Grunberger A, Hilgers F, Loeschcke A, Jaeger KE, et al. Comparative Single-
843 Cell Analysis of Different E. coli Expression Systems during Microfluidic Cultivation. PloS one.
844 2016;11:e0160711.
845 [75] Jin DJ, Cagliero C, Zhou YN. Role of RNA polymerase and transcription in the organization of the
846 bacterial nucleoid. Chem Rev. 2013;113:8662-82.
847 [76] Hodges-Garcia Y, Hagerman PJ, Pettijohn DE. DNA ring closure mediated by protein HU. J Biol
848 Chem. 1989;264:14621-3.
849 [77] Driessen RP, Sitters G, Laurens N, Moolenaar GF, Wuite GJ, Goosen N, et al. Effect of
850 temperature on the intrinsic flexibility of DNA and its interaction with architectural proteins.
851 Biochemistry. 2014;53:6430-8.
852 [78] van der Valk RA, Laurens N, Dame RT. Tethered Particle Motion Analysis of the DNA Binding
853 Properties of Architectural Proteins. Methods Mol Biol. 2017;1624:127-43.
854 [79] Skoko D, Wong B, Johnson RC, Marko JF. Micromechanical analysis of the binding of DNA-
855 bending proteins HMGB1, NHP6A, and HU reveals their ability to form highly stable DNA-protein
856 complexes. Biochemistry. 2004;43:13867-74.
857 [80] Xiao B, Johnson RC, Marko JF. Modulation of HU-DNA interactions by salt concentration and
858 applied force. Nucleic Acids Res. 2010;38:6176-85.

[81] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.
[82] Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018;27:112-28.
[83] Sambrook J, Fritsch EF, Maniatis T. Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 1989.
[84] Purdy D, O'Keeffe TAT, Elmore M, Herbert M, McLeod A, Bokori-Brown M, et al. Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier. Molecular Microbiology. 2002;46:439-52.
[85] Hussain HA, Roberts AP, Mullany P. Generation of an erythromycin-sensitive derivative of Clostridium difficile strain 630 (630Deltaerm) and demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. J Med Microbiol. 2005;54:137-41.
[86] Gibson DG. Enzymatic assembly of overlapping DNA fragments. Methods Enzymol. 2011;498:349-61.
[87] Laurent TC, Killander J. A theory of gel filtration and its exeperimental verification. Journal of Chromatography A. 1964;14:317-30.
[88] Ducret A, Quardokus EM, Brun YV. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat Microbiol. 2016;1:16077.
[89] Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, et al. ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics. 2017;18:529.
[90] Driessen RP, Lin SN, Waterreus WJ, van der Meulen AL, van der Valk RA, Laurens N, et al. Diverse architectural properties of Sso10a proteins: Evidence for a role in chromatin compaction and organization. Sci Rep. 2016;6:29422.

Table 1. Plasmids used in this study.

Name	Relevant features *	Source/Reference
pH6HTC	$P_{T7,}$ HaloTag-His6, *amp*	Promega
pCR2.1-TOPO	TA vector; pMB1 oriR; *km amp*	ThermoFisher
pET28b	$lacI^{q}$, P_{T7} expression vector, *km*	Novagen
pRPF185	*tetR* P_{tet}-*gusA*; *catP*	[64]
pAP24	*tetR* P_{tet}-$sLuc^{opt}$; *catP*	[63]
pRD118	P_{T7}-s*so685*	[91]
pAF226	P_{T7}-$hupA_{6xHis}$; *km*	This study
pAF232	P_{T7}-$hupA^{QE}{}_{6xHis}$; *km*	This study
pAF234	P_{T7}-$hupA^{QED}{}_{6xHis}$; *km*	This study
pAF235	*tetR* P_{tet}-$hupA^{QE}$-$HaloTag_{6xHis}$; *catP*	This study
pAF237	*tetR* P_{tet}-$hupA^{QED}$-$HaloTag_{6xHis}$; *catP*	This study
pAF254	*tetR* P_{tet}-luc^{opt}; *catP*	This study
pAF255	*tetR* P_{tet}-*lgbit*; *catP*	This study
pAF256	*tetR* P_{tet}-*hupA-smbit*/*lgbit*; *catP*	This study
pAF257	*tetR* P_{tet}-*smBit*/*hupA-lgbit*; *catP*	This study
pAF259	*tetR* P_{tet}-$bitluc^{opt}$; *catP*	This study
pAF260	*tetR* P_{tet}-*smbit*; *catP*	This study
pAF262	*tetR* P_{tet}-*smbit-lgbit*; *catP*	This study
pAP103	*tetR* P_{tet}-*hupA*; *catP*	This study
pAP118	*tetR* P_{tet}-*hupA-sbit-hupA-lgbit*; *catP*	This study
pAP134	*tetR* P_{tet}-*hupA-lgbit*; *catP*	This study
pAP135	*tetR* P_{tet}-*hupA-sbit*; *catP*	This study
pAP159	*tetR* P_{tet}-*sbit-lgbit* (GTT); *catP*	This study
pRD4	*tetR* P_{tet}-$hupA$-$HaloTag_{6xHis}$; *catP*	This study
pWKS1744	pCR2.1-TOPO with *hupA; km amp*	This study
pWKS1746	pCR2.1-TOPO with $HaloTag_{6xHis}$*; km amp*	This study

* *amp* – ampicillin resistance cassette *catP* – chloramphenicol resistance cassette, *km* – kanamycin resistance cassette

Table 2. *C. difficile* strains used in this study.

Name	Relevant Genotype/Phenotype*	Source/Reference
AP6	*C. difficile 630*Δ*erm*; Erm^S	[64, 85]
WKS1588	*630*Δ*erm* pRPF185; $Thia^R$	This study
RD16	*630*Δ*erm* pRD4; $Thia^R$	This study
AF239	*630*Δ*erm* pAF237; $Thia^R$	This study
AP34	*630*Δ*erm* pAP24; $Thia^R$	[62]
AP106	*630*Δ*erm* pAP103; $Thia^R$	This study
AP122	*630*Δ*erm* pAP118; $Thia^R$	This study
AP152	*630*Δ*erm* pAP134; $Thia^R$	This study
AP153	*630*Δ*erm* pAP135; $Thia^R$	This study
AP181	*630*Δ*erm* pAF254; $Thia^R$	This study
AP182	*630*Δ*erm* pAF259; $Thia^R$	This study
AP183	*630*Δ*erm* pAF256; $Thia^R$	This study
AP184	*630*Δ*erm* pAF257; $Thia^R$	This study
AP199	*630*Δ*erm* pAF255; $Thia^R$	This study
AP201	*630*Δ*erm* pAF260; $Thia^R$	This study
AP202	*630*Δ*erm* pAF262; $Thia^R$	This study

* Erm^S – Erythromycin sensitive, $Thia^R$ – Thiamphenicol resistant

Table 3. Oligonucleotides used in this study.

Name	Sequence (5'>3') *
oAF57	GTCGCCATGGATGAATAAAGCTGAATTAGTATCAAAG
oAF58	GACGCTCGAGTCCATTTATTATATCCTTTAATCC
oAF61	CGCCAGGCCAGGGCTGTCACTGTGCAGCTCGTGGACGC
oAF62	GCGTCCACGAGCTGCACAGTGACAGCCCTGGCCTGGCG
oAF63	CATCAGGCAAGAGTAGTCACTGTGTAGCTCGTGGATGC
oAF64	GCATCCACGAGCTACACAGTGACTACTCTTGCCTGATG
oAF65	CATTAAGTATGAGTATTCTATGTATAGATCATTGATGC
oAF66	GCATCAATGATCTATACATAGAATACTCATACTTAATG
oAF73	CATTTGAGACAAGAGAACAGGCTGCTGAACAAGGAAGAAATCCAAGAG
oAF74	CTTGGATTTCTTCCTTGTTCAGCAGCCTGTTCTCTTGTCTCAAATGTTC
oAF75	GGCTGCTGAACAAGGAGATAATCCAAGAGATCCAGAGC
oAF76	CTGGATCTCTTGGATTATCTCCTTGTTCAGCAGCCTG
oAF81	GCTAGAATTCGCCACTGGCAGCAGCCAC
oAF82	CCTAGAATTCCTGTCCTTCTAGTGTAGCCG
oAP47	TAGGATCCTTATCCATTTATTATATCCTTTAATCC
oAP48	CT GAGCTCCTGCAGTAAAGGAGAAAATTTTGTTTTTACACTTGAAGATTTTGTGG
oAP49	TAGGATCCCTATGCTAGAATACGTTCAC
oAP54	CTGAGCTCCTGCAGTAAAGGAGAAAATTTTGTTTTTACACTTGAAGATTTTGTG
oAP55	TAGGATCCCTATAGAATTTCTTCAAAAAGTCTATAACCTGTAACACTGTTTATAGTTAC
oAP58	GGATCCTATAAGTTTTAATAAAACTTTAAATAG
oAP59	AGCTCAGATCTGTTAACGCTACGATCAAGC
oAP60	GCTTGATCGTAGCGTTAACAGATCTGAGC
oAP61	CTCCTTTACTGCAGCGATCGAGCTATAG
oAP62	GAAGAAATTCTATAGCTCGATCGCTGCAG
oAP63	GTTTTATTAAAACTTATAGGATCCCTAACTGTTTATAG
oAP64	GATCTGAGCTCCTGCAGTAAAGGAGAAAATTTTGTGAATAAAGC
oAP65	CTTATAGGATCCAGCTATAGAATTTCTTC
oAP66	GATCTGAGCTCCTGCAGTAAAGGAGAAAATTTTGTTACAGGTTATAGAC
oAP67	GCTCGATCGCTGCAGTAAAGGAGAAAATTTTGTTTTTACACTTGAAGATTTTGTG
oAP96	GCAGTAAAGGAGAAAATTTTGTGTTTACACTTGAAGATTTTG
oAP97	CACAAAATCTTCAAGTGTAAACACAAAATTTTCTCCTTTAC
oAP98	GCAGTAAAGGAGAAAATTTTGTGACAGGTTATAGACTTTTTG
oAP99	CTTCAAAAAGTCTATAACCTGTCACAAAATTTTCTCCTTTAC
oRD5	CAGGATCTGGTTCAGGAAGTCTCGAGGGTTCCGAAATCGGTACTGG
Sso10a-2Nde	ATACATATGCAACTTGAACGGCGTAAAAGAGGAACAATGG
Sso10a-2Bam685	GGTGGATCCTTTTCATCCCTTTAGTTCTTCCAG
oWKS-1511	CTCGAGTCAGGATCTGGTTCAGGAAGTGGTTCCGAAATCGGTACTGGCTTTCC
oWKS-1512	GGATCCTTAGTGGTGATGGTGATGATGACC
oWKS-1519	GAGCTCAAATTTGAATTTTTTAGGGGGAAAATACCGTGAATAAAGCTGAATTAGTATCAAAG
oWKS-1520	CTCGAGACTTCCTGAACCAGATCCTGATCCATTTATTATATCCTTTAATCCTTTTC

* Restriction enzyme sites used underlined

Legends

Fig. 1. *Clostridium difficile* HupA is a homologue of bacterial HU proteins. (a) Multiple sequence alignment (ClustalOmega) of *C. difficile* HupA with homologous proteins from the Uniprot database. The protein sequences from *C. difficile* 630Δ*erm* (Q180Z4), *Escherichia coli* αHU (P0ACF0), *E. coli* βHU (P0ACF4), *Bacillus subtilis* (A3F3E2), *Geobacillus stearothermophilus* (P0A3H0), *Bacillus anthracis* (Q81WV7), *Staphylococcus aureus* (Q99U17), *Salmonella typhimurium* (P0A1R8), *Streptococcus pneumonia* (AAK75224), *Streptococcus mutans* (Q9XB21), *Mycobacterium tuberculosis* (P9WMK7), *Thermotoga. maritima* (P36206) and *Anabaena sp.* (P05514) were selected for alignment. Conserved residues (*) are also indicated in grey. (b) Structural model of the *C. difficile* HupA dimer based on homology with the crystal structure of DNA-bound nucleoid associated protein SAV1473 (SWISS-MODEL, PDB: 4QJN, 58.43% identity). α-helixes are represented in red, β-sheets in orange and unstructured regions in grey. Both the N-terminus and the C-terminus are indicated in the figure. A DNA binding pocket is formed by the arm regions of the dimer, composed by 4 β-sheets in each monomer. The localization of the substituted residues (R55, R58 and R61) are indicated (blue, sticks). (c) Electrostatic surface potential of *C. difficile* HupA. The electrostatic potential is in eV with the range shown in the corresponding color bar.

Fig. 2 – Dimerization of HupA is independent of DNA binding. (a) Electrophoretic mobility shift assays with increasing concentrations (0.25 -2 µM) of $HupA_{6xHis}$ and $HupA^{QED}{}_{6xHis}$. Gel shift assays were performed with 2.4 nM radio-labeled ([ɣ-^{32}P] ATP) 29% G+C dsDNA oligonucleotide incubated with HupA for 20 min at room temperature prior to separation. Protein-DNA complexes were analysed on native 8% polyacrylamide gels, vacuum-dried and visualized by phosphorimaging. ssDNA and dsDNA (without protein added) were used as controls. (b) Quantification of the gel-shift DNA-protein complex by densitometry. Gel shift

assays were performed with 2.4 nM radio labelled ([ɣ-^{32}P] ATP) dsDNA oligonucleotides with different 29-71% G+C content and the indicated concentration of $HupA_{6xHis}$ (red) and $HupA^{QED}_{6xHis}$ (blue). (c) Elution profiles of $HupA_{6xHis}$ (red) and $HupA^{QED}_{6xHis}$ (blue) from size exclusion chromatography. The experiments were performed with purified protein (100 µM) on a Superdex HR 75 10/30 column. The elution position of protein standards of the indicated MW (in kDa) are indicated by vertical grey dashed lines. The elution profiles show a single peak, corresponding to a ~38 kDa multimer, when compared to the predicted molecular weight of the monomer (11 kDa). No significant difference in the elution profile of the $HupA^{QED}_{6xHis}$ compared to $HupA_{6xHis}$ was observed. (d) Western-blot analysis of glutaraldehyde cross-linking of $HupA_{6xHis}$ and $HupA^{QED}_{6xHis}$. 100 ng HupA was incubated with 0%, 0.0006% and 0.006% glutaraldehyde for 30 min at room temperature. The samples were resolved by SDS-PAGE and analysed by immunoblotted with anti-his antibody. Crosslinking between the HupA monomers is observed with the approximate molecular weight of a homo-dimer (~22 kDa). Additional bands of lower molecular weight HupA are observed (*) that likely represent breakdown products.

Fig. 3 – HupA demonstrates self-interaction in *C. difficile*. A split luciferase complementation assay was used to demonstrate interactions between HupA monomers *in vivo*. Cells were induced with 200 ng/mL ATc for 60 min. Optical density-normalized luciferase activity (LU/OD) is shown right before induction (T0, blue bars) and after 1 hour of induction (T1, red bars). The averages of biological triplicate measurements are shown, with error bars indicating the standard deviation from the mean. Luciferase activity of strains AP182 (P_{tet}-*bitluc^{opt}*), AP122 (P_{tet}-*hupA-smbit/hupA-lgbit*), AP152 (P_{tet}-*hupA-lgbit*), AP153 (P_{tet}-*hupA-smbit*), AP183 (P_{tet}-*hupA-smbit-lgbit*) and AP184 (P_{tet}-*smbit-hupA-lgbit*). A positive interaction was defined on the basis of the negative controls as a luciferase activity of >1000 LU/OD. No significant difference was detected at T0. AP122 and AP182 were significantly higher with $p<0.0001$(*) by two-way ANOVA.

Fig. 4 – HupA overexpression leads to compaction of the nucleoid *in vivo*. (a) Fluorescence microscopy analysis of *C. difficile* 630Δ*erm* harboring the vector for ATc-dependent overexpression of HupA (AP106). For HupA overexpression, cells were induced at mid-exponential growth in liquid medium with different ATc concentrations (50, 100 and 200 ng/mL) for 1h. *C. difficile* 630Δ*erm* and non-induced AP106 were used as controls. The cells were stained with DAPI for DNA visualization (nucleoid). Phase contrast (PC) and an overlay of both channels is shown. Because growth is asynchronous in these conditions, cells representing different cell cycles stages can be found. In the presence of ATc, the chromosome appears more compacted. White arrow indicates the cells with mid-cell nucleoid. Scale bar = 2 µm. (b) Boxplots of mean nucleoid length. Whiskers represent minimum and maximum nucleoid length observed. Black dots represent the mean and the blue lines represent the median values. Quantifications were performed using MicrobeJ from at least two biological replicates for each condition. n is the number of cells analyzed per condition. * $p<0,05$ ** $p<0,0001$ by one-way ANOVA.

Fig. 5 – HupA co-localizes with the nucleoid. (a) Fluorescence microscopy analysis of *C. difficile* 630Δ*erm* harboring a vector for expression of HupA-HaloTag (RD16). For visualization of HupA-HaloTag cells were induced at mid-exponential growth phase with 200 ng/mL ATc for 1h and incubated with Oregon Green HaloTag substrate for 30 min. The cells were stained with DAPI stain to visualize DNA (nucleoid). As control non-induced RD16 was used. Phase contrast (PC) and an overlay of the channels is shown. Because growth is asynchronous under these conditions, cells representing different cell cycles stages can be found. In the presence of ATc the chromosome appears more compacted and HupA-HaloTag co-localizes with the nucleoid. Scale bar = 2 µm. (b) Average intensity profile scans for the nucleoid (DAPI, blue line) and HupA-HaloTag (Oregon Green, green line) obtained

from a MicrobeJ analysis from at least two biological replicates in each condition. 289 cells were analyzed. Standard deviation of the mean is represented by the respective color shade. (c) In-gel fluorescent analysis of RD16 samples before induction (T0), 1 and 3 hours after induction (T1, T3). Samples were incubated with Oregon Green substrate for 30 min and run on 12%SDS-PAGE.

Fig. 6 – HupA alters the topology of DNA *in vitro*. (a) Ligase-mediated cyclization assay. A 119 bp [ɣ-^{32}P]ATP-labelled dsDNA fragment was incubated in the presence of increasing concentrations of $HupA_{6xHis}$ (1, 10 µM), exonuclease III and ligase, as indicated above the panel. The presence of ExoIII-resistant (i.e. circular) DNA fragments is observed when samples are incubated with $HupA_{6xHis}$ ("circle"). (b) The effect of increasing concentrations of HupA and $HupA_{6xHis}$ on the persistence length of DNA in tethered particle motion experiments. Root mean square (RMS; see Equation 1/Materials and methods) values as a function of protein concentration obtained is shown. Increasing concentrations of HupA and $HupA_{6xHis}$ lead to a decreased RMS, suggesting compaction of the DNA.

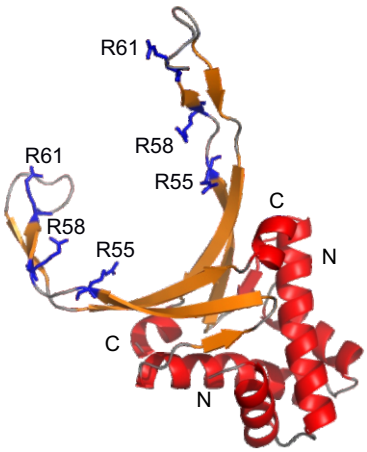
Figure 1

(a)

Organism	Nt	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
C. difficile 630		-	V	N	**K**	A	E	**L**	V	S	K	M	A	E	K	S	G	L	T	K	K	E	A	E	A	A	L	N	A	F	M	S	S	V	Q	D	A	L	V	N	N	E	K	V	Q	L	V	**G**	**F**	**G**	T	**F**
E. coli aHU		-	M	N	**K**	T	Q	**L**	I	D	V	I	A	E	K	A	E	L	S	K	T	Q	A	K	A	A	L	E	S	T	L	A	A	I	T	E	S	L	K	E	G	D	A	V	Q	L	V	**G**	**F**	**G**	T	**F**
E. coli ßHU		-	M	N	**K**	S	Q	**L**	I	D	K	I	A	A	G	A	D	I	S	K	A	A	A	G	R	A	L	D	A	I	I	A	S	V	T	E	S	L	K	E	G	D	D	V	A	L	V	**G**	**F**	**G**	T	**F**
B. subtilis		-	M	N	**K**	T	E	**L**	I	N	A	V	A	E	A	S	E	L	S	K	K	D	A	T	K	A	V	D	S	V	F	D	T	I	L	D	A	L	K	N	G	D	K	I	Q	L	I	**G**	**F**	**G**	N	**F**
B. stearothermophilus		-	M	N	**K**	T	E	**L**	I	N	A	V	A	E	T	S	G	L	S	K	K	D	A	T	K	A	V	D	A	V	F	D	S	I	T	E	A	L	R	K	G	D	K	V	Q	L	I	**G**	**F**	**G**	N	**F**
B. anthracis		-	M	N	**K**	T	E	**L**	I	K	N	V	A	Q	N	A	E	I	S	Q	K	E	A	T	V	V	V	Q	T	V	V	E	S	I	T	N	T	L	A	A	G	E	K	V	Q	L	I	**G**	**F**	**G**	T	**F**
S. aureus		-	M	N	**K**	T	D	**L**	I	N	A	V	A	E	Q	A	D	L	T	K	K	E	A	G	S	A	V	D	A	V	F	E	S	I	Q	N	S	L	A	K	G	E	K	V	Q	L	I	**G**	**F**	**G**	N	**F**
S. typhimurium		-	M	N	**K**	S	Q	**L**	I	E	K	I	A	A	G	A	D	I	S	K	A	A	A	G	R	A	L	D	A	I	I	A	S	V	T	E	S	L	K	E	G	D	D	V	A	L	V	**G**	**F**	**G**	T	**F**
S. pneumonia		M	A	N	**K**	Q	D	**L**	I	A	K	V	A	E	A	T	E	L	T	K	K	D	S	A	A	A	V	E	A	V	F	A	A	V	A	D	Y	L	A	A	G	E	K	V	Q	L	I	**G**	**F**	**G**	N	**F**
S. mutans		M	A	N	**K**	Q	D	**L**	I	A	K	V	A	E	A	T	E	L	T	K	K	D	S	A	A	A	V	D	A	V	F	S	A	V	S	S	Y	L	A	K	G	E	K	V	Q	L	I	**G**	**F**	**G**	N	**F**
M. tuberculosis		-	M	N	**K**	A	E	**L**	I	D	V	L	T	Q	K	L	G	S	D	R	R	O	A	T	A	A	V	E	N	V	V	D	T	I	V	R	A	V	H	K	G	D	S	V	T	I	T	**G**	**F**	**G**	V	**F**
T. maritima		-	M	T	**K**	K	E	**L**	I	D	R	V	A	K	K	A	G	A	K	K	K	D	V	K	L	I	L	D	T	I	L	E	T	I	T	E	A	L	A	K	G	E	K	V	Q	I	V	**G**	**F**	**G**	S	**F**
Anabaena		-	M	N	**K**	G	E	**L**	V	D	A	V	A	E	K	A	S	V	T	K	K	Q	A	D	A	V	L	T	A	A	L	E	T	I	I	E	A	V	S	S	G	D	K	V	T	L	V	**G**	**F**	**G**	S	**F**
			.		*		:	*	:			:	:							:							:				.		:	:				:			.	:		:		:		*	*	*		*

Organism	Nt	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
C. difficile 630		E	T	R	E	R	A	A	**R**	Q	G	R	**N**	**P**	R	D	P	E	Q	V	I	D	I	P	A	S	K	A	**P**	V	**F**	K	A	**G**	K	G	L	K	D	I	I	N	G	-	-	-	-	-	-	-	-	-
E. coli aHU		K	V	N	H	R	A	E	**R**	T	G	R	**N**	**P**	Q	T	G	K	-	E	I	K	I	A	A	A	N	V	**P**	A	**F**	V	S	**G**	K	A	L	K	D	A	V	K	-	-	-	-	-	-	-	-	-	-
E. coli ßHU		A	V	K	E	R	A	A	**R**	T	G	R	**N**	**P**	Q	T	G	K	-	E	I	T	I	A	A	A	K	V	**P**	S	**F**	R	A	**G**	K	A	L	K	D	A	V	N	-	-	-	-	-	-	-	-	-	-
B. subtilis		E	V	R	E	R	S	A	**R**	K	G	R	**N**	**P**	Q	T	G	E	-	E	I	E	I	P	A	S	K	V	**P**	A	**F**	K	P	**G**	K	A	L	K	D	A	V	A	G	K	-	-	-	-	-	-	-	-	-
B. stearothermophilus		E	V	R	E	R	A	A	**R**	K	G	R	**N**	**P**	Q	T	G	E	-	E	M	E	I	P	A	S	K	V	**P**	A	**F**	K	P	**G**	K	A	L	K	D	A	V	K	-	-	-	-	-	-	-	-	-	-
B. anthracis		E	V	R	E	R	A	A	**R**	T	G	R	**N**	**P**	Q	T	G	E	-	E	M	Q	I	A	A	S	K	V	**P**	A	**F**	K	A	**G**	K	E	L	K	E	A	V	K	-	-	-	-	-	-	-	-	-	-
S. aureus		E	V	R	E	R	A	A	**R**	K	G	R	**N**	**P**	Q	T	G	K	-	E	I	D	I	P	A	S	K	V	**P**	A	**F**	K	A	**G**	K	A	L	K	D	A	V	K	-	-	-	-	-	-	-	-	-	-
S. typhimurium		A	V	K	E	R	A	A	**R**	T	G	R	**N**	**P**	Q	T	G	K	-	E	I	T	I	A	A	A	K	V	**P**	S	**F**	R	A	**G**	K	A	L	K	D	A	V	N	-	-	-	-	-	-	-	-	-	-
S. pneumonia		E	V	R	E	R	A	E	**R**	K	G	R	**N**	**P**	Q	T	G	K	-	E	M	T	I	A	A	S	K	V	**P**	A	**F**	K	A	**G**	K	A	L	K	D	A	V	K	-	-	-	-	-	-	-	-	-	-
S. mutans		E	V	R	E	R	A	A	**R**	K	G	R	**N**	**P**	Q	T	G	E	-	E	I	K	I	K	A	S	K	V	**P**	A	**F**	K	A	**G**	K	A	L	K	D	A	V	K	-	-	-	-	-	-	-	-	-	-
M. tuberculosis		E	Q	R	R	R	A	A	**R**	V	A	R	**N**	**P**	R	T	G	E	-	T	V	K	V	K	P	T	S	V	**P**	A	**F**	R	P	**G**	A	Q	F	K	A	V	V	S	G	A	Q	R	L	P	A	E	G
T. maritima		E	V	R	K	A	A	A	**R**	K	G	V	**N**	**P**	Q	T	R	K	-	P	I	T	I	P	E	R	K	V	**P**	K	**F**	P	K	**G**	K	A	L	K	E	K	V	K	-	-	-	-	-	-	-	-	-	-
Anabaena		E	S	R	E	R	K	A	**R**	E	G	R	**N**	**P**	K	T	N	E	-	K	M	E	I	P	A	T	R	V	**P**	A	**F**	S	A	**G**	K	L	F	R	E	K	V	A	P	P	K	A	-	-	-	-	-	-
			.	.					*		.		*	*	:			:			:		:					.	*		*			*			:	:			:									

(b)



(c)

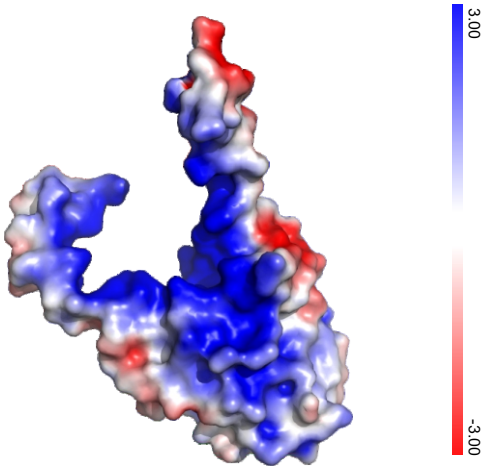


Figure 2

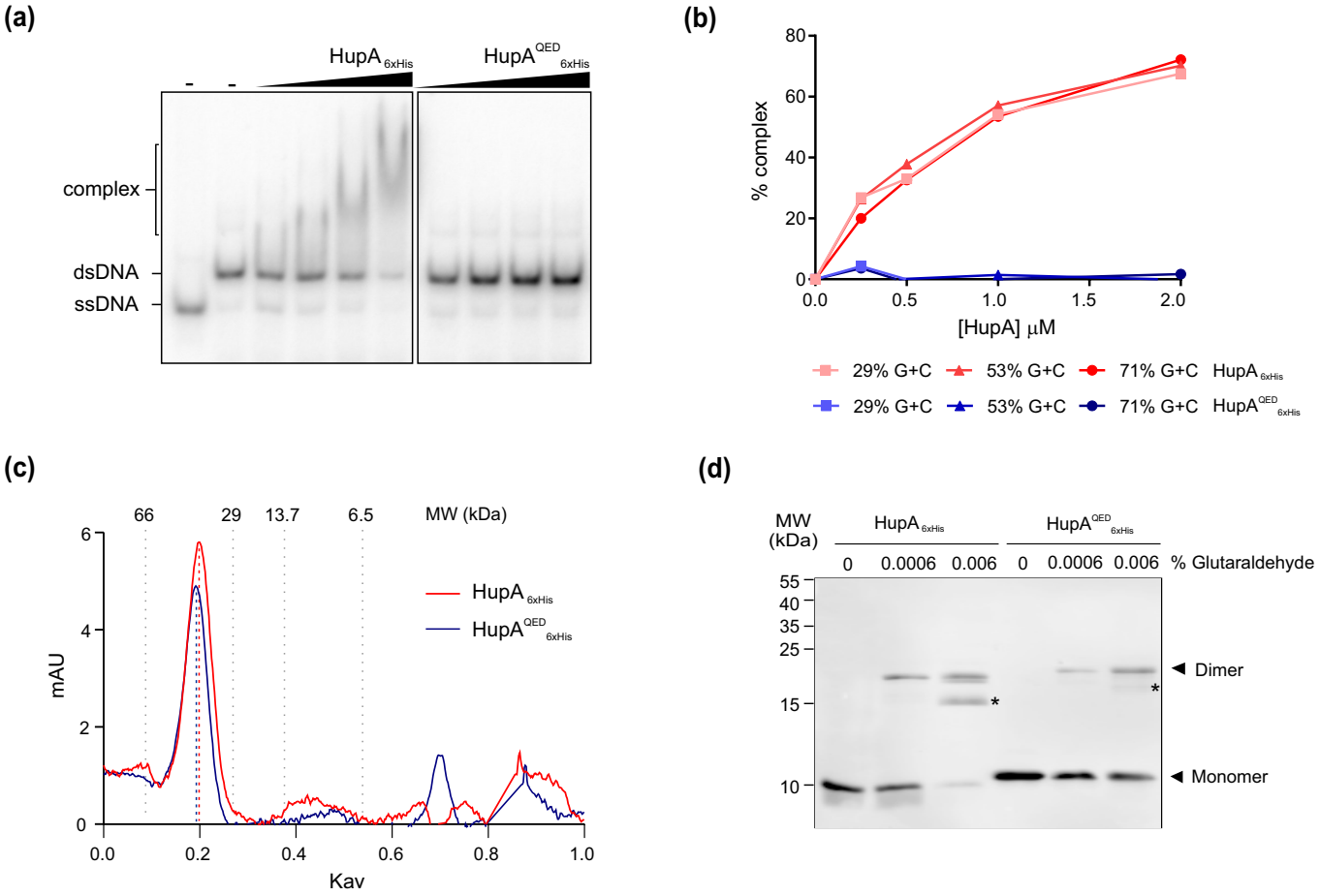

(a)
HupA 6xHis
HupA QED 6xHis
complex
dsDNA
ssDNA
(b)
% complex
[HupA] µM
29% G+C
53% G+C
71% G+C
HupA 6xHis
HupA QED 6xHis
(c)
MW (kDa)
66
29
13.7
6.5
mAU
Kav
HupA 6xHis
HupA QED 6xHis
(d)
MW (kDa)
HupA 6xHis
HupA QED 6xHis
% Glutaraldehyde
Dimer
Monomer

Figure 3

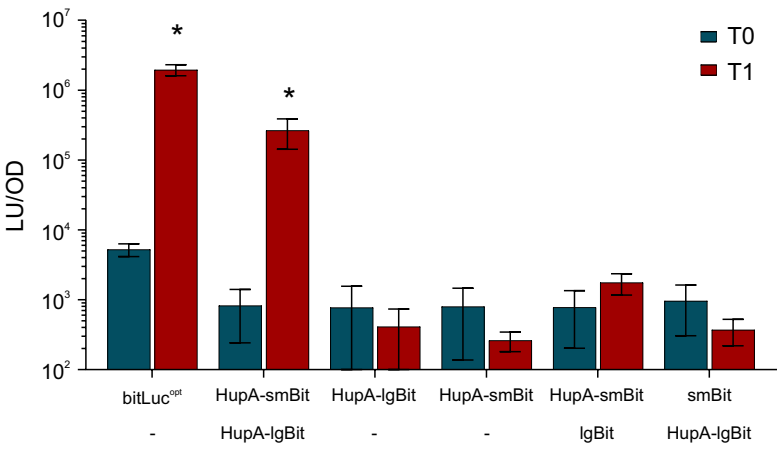
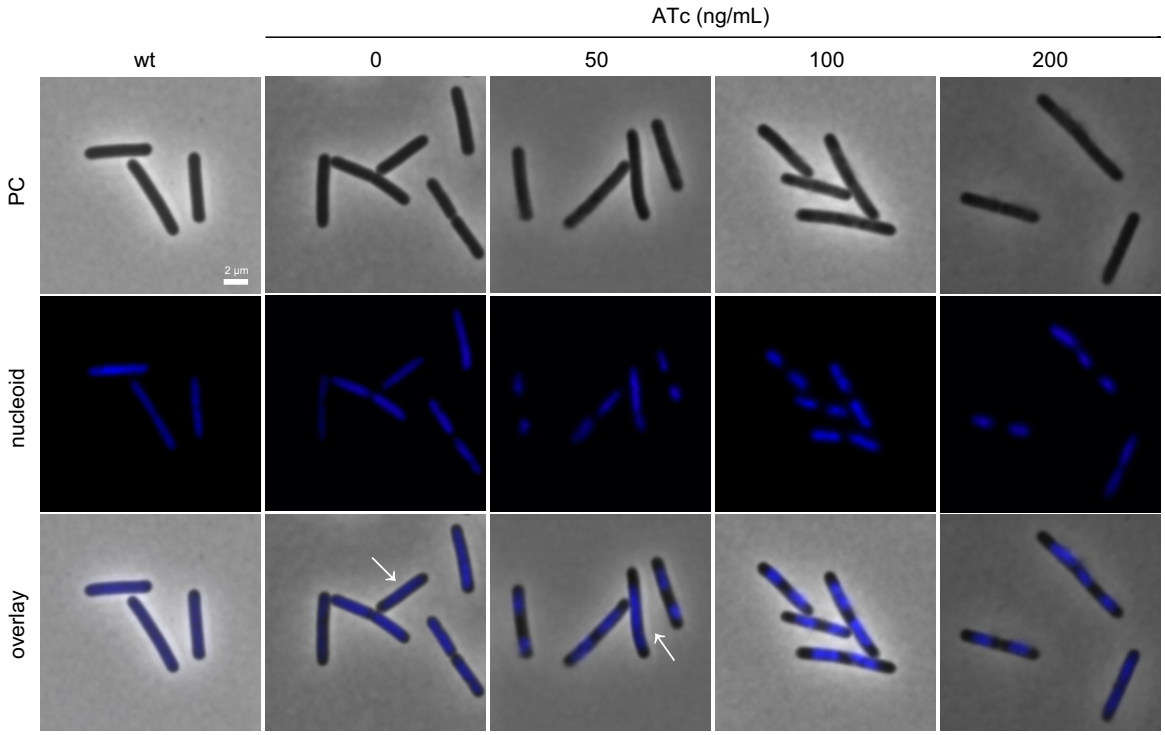

LU/OD
10^7
10^6
10^5
10^4
10^3
10^2
*
*
T0
T1
bitLuc^{opt}
HupA-smBit
HupA-lgBit
HupA-smBit
HupA-smBit
smBit
-
HupA-lgBit
-
-
lgBit
HupA-lgBit

Figure 4

(a)


ATc (ng/mL)
wt
0
50
100
200
PC
nucleoid
overlay
2 µm

(b)

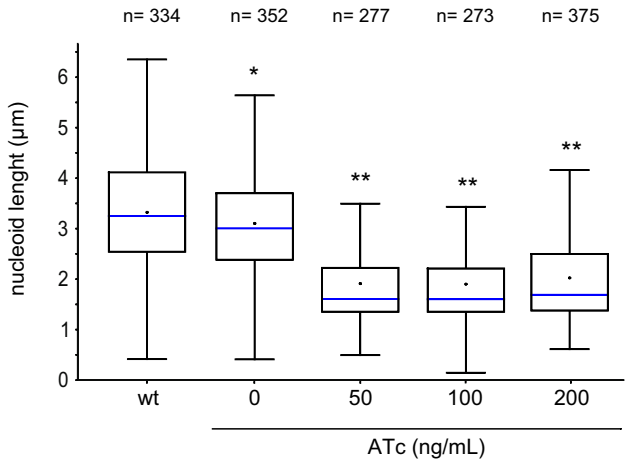

n= 334
n= 352
n= 277
n= 273
n= 375
nucleoid lenght (µm)
0
1
2
3
4
5
6
*
**
**
**
wt
0
50
100
200
ATc (ng/mL)

Figure 5

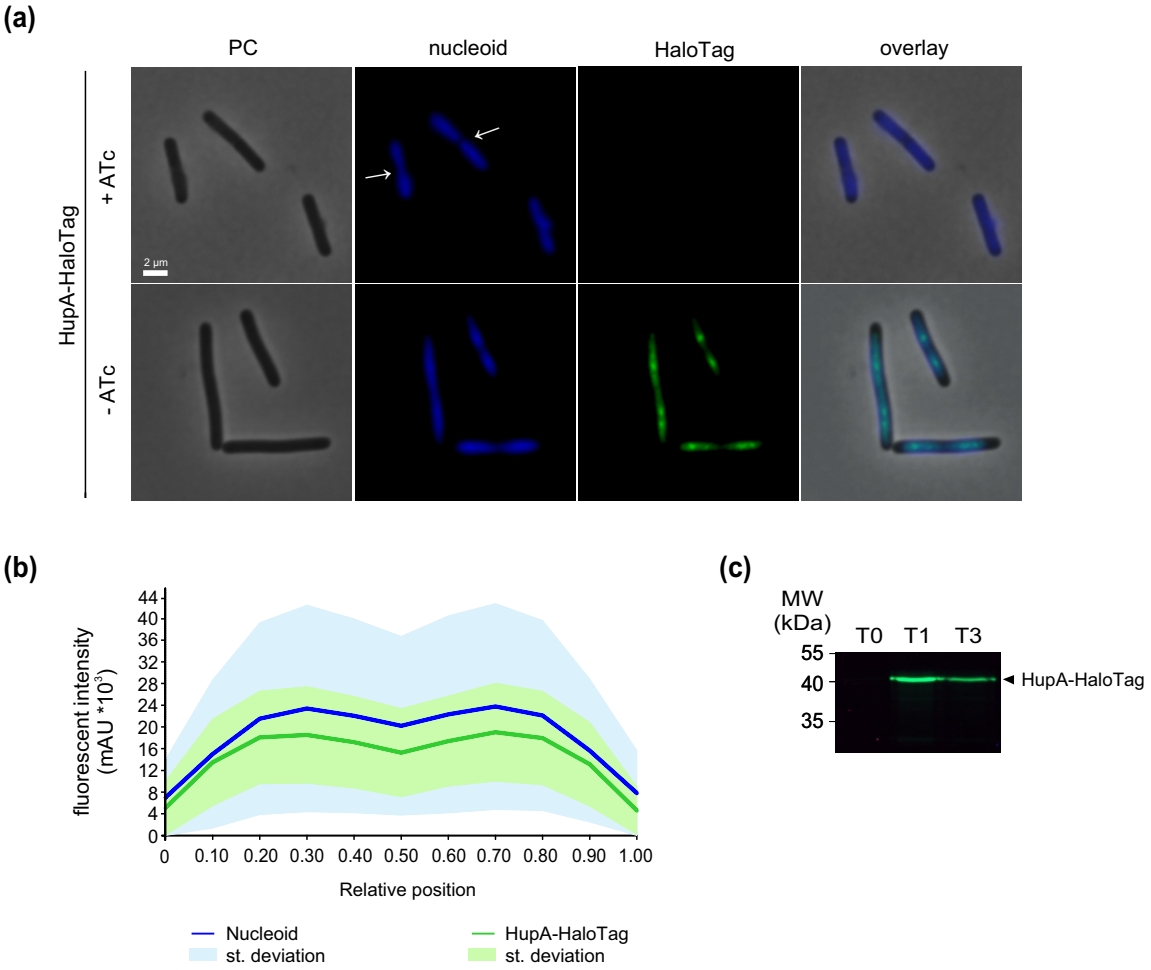
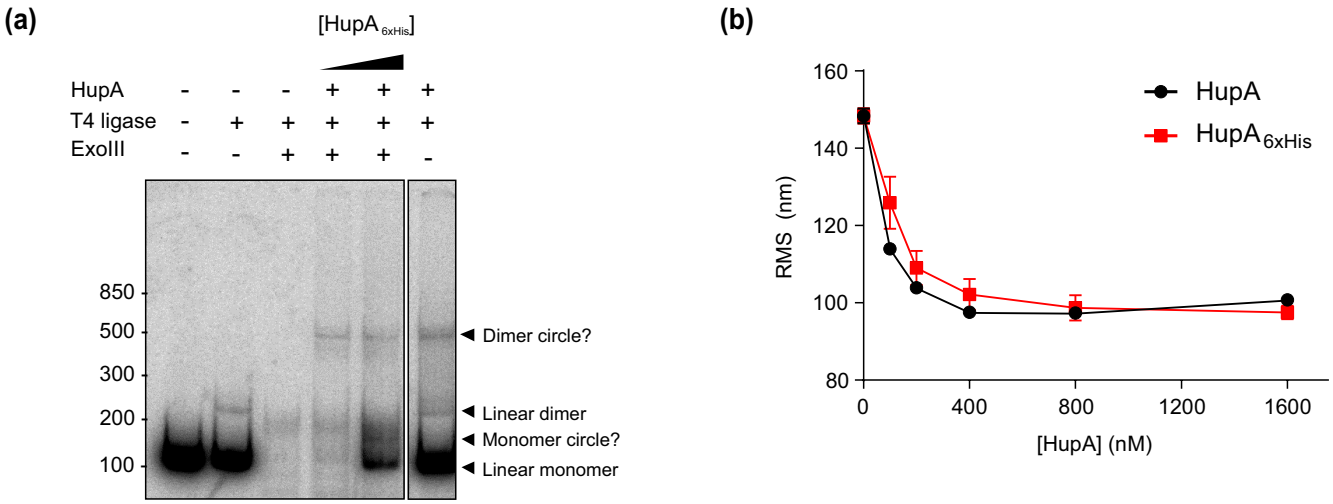

(a)
PC
nucleoid
HaloTag
overlay
HupA-HaloTag
+ ATc
- ATc
2 µm
(b)
fluorescent intensity (mAU *10^3)
44
40
36
32
28
24
20
16
12
8
4
0
0
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Relative position
Nucleoid
st. deviation
HupA-HaloTag
st. deviation
(c)
MW (kDa)
T0
T1
T3
55
40
35
HupA-HaloTag

Figure 6


(a)
[HupA 6xHis]
HupA - - - + + +
T4 ligase - + + + + +
ExoIII - - + + + -
850
500
300
200
100
Dimer circle?
Linear dimer
Monomer circle?
Linear monomer
(b)
HupA
HupA 6xHis
RMS (nm)
160
140
120
100
80
0
400
800
1200
1600
[HupA] (nM)