Structure determination of *Mycoplasma pneumoniae* genome

Trussart Marie

TESI DOCTORAL UPF / ANY 2015

THESIS DIRECTORS

Dr. Luis Serrano and Dr. Marc A. Marti-Renom

DEPARTMENT

EMBL-CRG System Biology Unit

Center for Genomic Regulation (CRG)

Acknowledgements

I would like to express my gratitude to all of the people whose various contributions have made my PhD an enriching experience and have supported me along this challenging journey.

First of all, I would like to thank my thesis supervisor Dr. Luis Serrano for offering me the opportunity to do my doctoral thesis in his laboratory, whose expertise and suggestions have added considerably to my graduate experience. With his knowledge, he gave me many ideas and intellectual freedom in my work and supported me to start doing experiments in the lab, as well as suggested several fruitful collaborations. I would like to thank my other thesis supervisor Dr. Marc Marti-Renom for supporting and guiding my work throughout the years. I am especially grateful for your encouragement and patience, for your advices and for keeping a sense of humour when I had lost mine.

I have been very fortunate to work with Maria Lluch over the last years. Without her enthusiastic support and her confidence in me, I would not have been able to jump into the dark side of wet lab and microscopy.
Many more people in both the Serrano lab and marcius lab deserve special acknowledgments: Jae-Seong Yang for your unconditional help on a daily basis, not only for sharing your broad knowledge but also for being like a brother to me; Davide Bau for introducing and helping me understand 3D genome modeling; Javier Delgado for your time and support whenever I experienced any problem; Besray Ünal for your support and positive attitude in the bad days; Hannah and Tony for your friendship and for making my phd experience an enjoyable one; Thomas Pengo for introducing and helping me with the microscopy; Sira and Eva for your efforts and time with the experiments, which was essential to this work; Francois Serra for your help, motivation and support with TADbit; Ivan Junier for sharing your knowledge on polymer physics and passion for chromatin structure; all my colleagues: Kiana, Erik, Anne, Martin, Carolina, Vero, Dina, Marc; Violeta and Christina for your support; in short I would like to thanks all the past and current members of the Serrano lab and marcius lab for contributing to the nice atmosphere in the lab.

Quiero agradecer también a todos mis amigos en Barcelona, especialmente a Silvia que estuvo siempre presente durante todos estos años y a Marina por compartir juntas los altibajos de un doctorado. Gracias a Django por inspirarme.

Finalement je voudrais remercier mes amis et ma famille, en particulier mes parents pour être toujours la, pour m'avoir toujours soutenu dans mes décisions, pour leur confiance en moi tout au long de mes études et pour leurs précieux conseils.

Quiero agradecer a Federico por ser una persona tan especial y por haber estado a mi lado, tu apoyo me ha dado fuerzas para seguir adelante en los momentos mas críticos.

Abstract

Recent progress in imaging and chromosome conformation capture (3C) experiments has enabled the determination of the chromosome structure of different organisms, from bacteria to humans. Through applying computational approaches, researchers have used such experimental data to model the three-dimensional (3D) structure of genomes and genomic domains. However, despite these numerous studies, no systematic analysis of the accuracy of the 3D modeling of genomes has been performed. Moreover it has been shown that chromatin structure plays an essential role in the regulation of gene expression. This correlative observation, however, leads us to the question of how does the structural organization of genomes interplay with the transcriptional regulation of the resident genes.

In this thesis, we aim at addressing these two issues: first we developed a methodology to evaluate the accuracy of 3D modeling approaches; second, we applied this approach to explore the chromosome structure of the genome-reduced bacterium *Mycoplasma pneumoniae*. By combining super-resolution microscopy and Hi-C, we determined the 3D chromosome structure of this reduced-genome bacteria and established fundamental principles of its organization. For example, we studied the impact of chromosomal interacting domains (CIDs) on transcriptional regulation. Finally, our work suggests that a defined chromosomal structure could be a universal feature of all living systems, including those with minimal genomes.

Resumen

Los últimos progresos en microscopia y el desarrollo de las técnicas de captura de la conformación del cromosoma (3C) han permito determinar la estructura del cromosoma de diferentes organismos, desde bacterias a humanos. Investigadores han desarrollado metodologías para modelar la estructura del cromosoma en tres-dimensiones (3D). A pesar del gran número de estudios, no se ha evaluado aun la precisión y la metodología de la modelización de la conformación en 3D de los cromosomas. Además, se ha demostrado que la estructura de la cromatina tiene un papel esencial en la regulación de la expresión genética. ¿Cual es el papel de esa interacción entre la organización de la cromatina y la regulación de la transcripción en moldear la estructura del núcleo?

En esa tesis, hemos abordado estos dos problemas: hemos desarrollado una metodología para evaluar la precisión de los modelos reconstruidos; segundo hemos aplicado este método de modelización para explorar la estructura del cromosoma de la bacteria *Mycoplasma pneumoniae* que tiene un genoma reducido y pocas proteínas que unen DNA. Combinando microscopia de alta-resolución con Hi-C, hemos determinado la estructura tridimensional de su genoma y hemos establecido principios fundamentales de la organización de un cromosoma. Por ejemplo hemos estudiado el impacto de los dominios en la regulación de la transcripción. En conclusión, sugerimos que la estructura del cromosoma podría ser una característica de todos seres vivos, incluyendo los que tienen un genoma mínimo.

Objectives

The two main objectives of this thesis are: 1) to evaluate mean-field restraint-based reconstruction of genomes by considering diverse chromosome architectures and different levels of data noise and structural variability; 2) determine the 3D structure and the possible impact of chromatin organization in transcriptional regulation in *M. pneumoniae.*

The first chapter of the thesis is an introduction of genome organization and the different modeling and experimental approaches to unveil the 3D conformation of genomes. We presented the different factors that compact DNA and the role of chromatin structure in regulation. We also discussed the application of such methods to determine the chromosome structure of *Mycoplasma pneumoniae* and we end the chapter presenting the characteristics of this genome-reduced bacterium.

Over the second chapter of the thesis, we designed a pipeline to evaluate the restraint-based modeling approach that consists in simulating 'toy genome' structures, deriving interaction matrices from them, reconstructing their 3D structure, assessing their quality and predicting their accuracy using the Matrix Modeling Potential (MMP) score. Next, we described the results of assessing the predictive power for determining the 'real' assembly structure of 'toy genome' structures as well as *a priori* evaluate the input interaction matrices modeling potential. Finally, we summarized our conclusions on the limits of mean-field restraint-based approaches and how a measure such as the MMP can be used to *a priori* evaluate the reconstructed models.

The goal of chapter three was to uncover the 3D genome structure of *M. pneumoniae* with restraint-based modeling by combining electron microscopy, high-resolution light microscopy (STORM) and Hi-C. We selected *M. pneumoniae* as model organism because of its small genome size and its expected simplified regulatory network, compared to other bacteria that have several TFs and alternative sigma factors to reprogram RNA polymerase and coordinate gene transcription. By analyzing the resulting 3D models, we identified fundamental principles of genome organization and their impact in gene expression and provide evidence that the chromosome structure is exploited to control transcription. Moreover we detected that the chromosome is organized into CIDs and we provided the first evidence that genes inside CIDs tend to be co-regulated. We then studied the effect of inhibiting supercoiling on genome structure and observed that it significantly reduced the sharpness and positions of CIDs, suggesting that supercoiling is regulating those domains formation in bacteria. In conclusion, this study expands the current understanding of bacterial genome.

Table of contents

Pag.

Abstract..... v
Objectives..... vii

1. Introduction..... 1
1.1 Genome organization and their role in regulation..... 1
a) Supercoiling..... 2
b) Molecular crowding..... 4
c) Polyamines..... 5
d) Nucleoid-associated proteins..... 5
1.2 Modeling approaches of 3D genomes and their limitations..... 13
a) Experimental methods..... 13
b) Modeling methods..... 17
1.3 Application on *Mycoplasma pneumoniae* to understand the impact of structural organization in transcriptional regulation..... 20
a) *Mycoplasma*..... 20
b) *M. pneumoniae* morphology..... 21
c) *M. pneumoniae* ultrastructure..... 23
d) *M. pneumoniae* gliding motility..... 24
e) *M. pneumoniae* cell division..... 25
f) *M. pneumoniae* transcriptional regulation..... 28
2. Assessing the limits of restraint-based 3D modeling of genomes and genomic domains..... 31
2.1. Abstract..... 32
2.2. Introduction..... 33
2.3. Material and Methods..... 35
a) Overall pipeline..... 35
b) Matrix generation from toy genome architectures..... 36
c) Model building by TADbit..... 39
d) Model accuracy..... 40
e) Matrix Modeling Potential..... 41
2.4. Results..... 43
a) Toy genome structures and derived matrices..... 43
b) Accuracy of the generated models..... 45
c) Genome architecture and model accuracy..... 47
d) The accuracy of the models is sensitive to structural variability but robust to noise..... 48
e) The TADbit-SCC is an accurate scoring function for modeling..... 49
f) Reconstructed models capture part of the structural variability in the matrices..... 49
g) Statistics of the input matrices correlate with the accuracy of the models..... 50
h) The Matrix Modeling Potential (MMP) score..... 51
2.5. Discussion..... 54

2.6. Availability..... 57
3. Defined chromosome structure in a minimal cell..... 60
3.1. Abstract..... 60
3.2. Introduction..... 61
3.3. Material and Methods..... 65
a) Overview of Methodology..... 65
b) Chromosome conformation capture with next generation sequencing..... 35
c) Generation of contact matrix..... 68
d) Reproducibility of Hi-C data..... 69
e) Matrix Modeling Potential using MMP score..... 69
f) Integrative 3D Modeling with TADbit..... 69
g) TEM imaging..... 70
h) 3D reconstruction and cell volume..... 71
i) Estimation of chromosome dimensions and volume..... 71
j) Fluorescence In Situ Hybridization (FISH) combined with Immunofluorescence..... 72
k) FISH imaging acquisition and processing..... 74
l) Domain detection on Hi-C contact map..... 74
m) Co-expression levels analysis (RNA-seq)..... 74
n) HpaII sites number on domains borders..... 74
o) High co-expression levels within domains..... 75
p) Low co-expression levels surrounding domains borders..... 75
3.4. Results..... 76
a) Generation of the Hi-C map of the *M. pneumoniae* genome..... 76
b) 3D modeling reveals a chromosome structure with Ori and Ter localized at the two opposite poles..... 77
c) Chromosome occupancy is about two-thirds of the total cell volume..... 78
d) Validation of 3D models with fluorescent imaging..... 79
e) Genes are co-expressed within chromosome interaction domains..... 81
f) Inhibiting supercoiling decreases domain sizes and interaction frequencies..... 84
3.5. Discussion..... 86
3.6. Acknowledgments..... 94
3.7. Author contributions..... 94

Discussion..... 97
References..... 107

1. INTRODUCTION

1.1 Genome organization and their role in regulation

Identifying how genomes and chromosomes are spatially organized, and how their organization changes during physiological processes could help unveiling the complexity of regulation of gene-transcriptional networks determining all aspects of life. Indeed, the identification of some groups of transcription units in eukaryotes, called "transcription factories" [1, 2], as well as "replication factories" [3], have shed new light on the importance of genome architecture and chromatin looping in the coordination and regulation of biological processes, such as gene expression and replication. In prokaryotes, by analogy with the organization of transcription and replication in eukaryotes [4], the first imaging of bacterial RNA polymerase (RNAP) in *Bacillus subtilis* [5] revealed an organization in the RNAP distribution with the concentration of RNAP into transcription foci or "factories" at high growth rate. Similarly, imaging of RNAP in *Escherichia coli* provided more evidence into the role of RNAP and transcription in the organization of the bacterial chromosome as changes in the distribution of RNAP accompany changes in nucleoid structure [6-8]. Such interplay suggests that chromatin organization have a role in regulating gene expression at both global and gene-specific level [9-15].

In both eukaryotes and prokaryotes, the genome must be compacted to fit into the nucleus or nucleoid, respectively, while maintaining accessibility for efficient transcription, replication and segregation. Eukaryotic genomes are localized into the cell nucleus, where each chromosome is confined into a discrete region, referred to as chromosome territories [16], which in turn are spaced into two compartment types [17]. Co-regulated genes of the same compartment will be brought into physical proximity to coordinate their activities leading to a more efficient gene expression. In prokaryotes, a recent study has shown in fast-growing *E. coli* cells that the transcription machinery is spatially organized into functional compartments, which suggest that functional compartmentalization is also present in bacterial chromosome organization [18]. Besides, mammalian genomes are further partitioned into domains, the so-called

Topological Associated Domains (TADs) ranging from 200-kilobases (kb) to 1 megabase (Mb) and conserved across different species and cell types [19, 20].

Although technical limitations for chromosome visualization have hampered the characterization of detailed organization of bacterial chromosome, several levels of regulation have been identified. At the molecular level, bacteria have evolved mechanisms that condense their chromosome like DNA supercoiling, macromolecular crowding forces exerted by the cytoplasm, polyamines and nucleoid-associated proteins (NAPs). Paradoxically, the coupled transcription and translation of transmembrane proteins, called transertion expand the nucleoid towards the cell membrane [21-23]. In the following sections, we will discuss the mechanisms mentioned that compact the chromosome.

a) Supercoiling

The first factor that condense DNA is the supercoiling that is mediated by the action of topoisomerases and gyrases [24, 25]. In bacteria the DNA is negatively supercoiled, which means twisted in the opposite direction to the Watson-Crick helix. DNA gyrase introduces negative supercoiling, while DNA topoisomerases I such as topA gene relaxes negative supercoiling and DNA topoisomerases IV relaxes both positive and negative supercoiling [26, 27]. Structurally, gyrases act as tetramers with two monomers encoded gyrA and gyrB genes. The introduction of supercoils occurs at the expense of ATP hydrolysis, gyrB subunit forms the site of ATP binding and hydrolysis while gyrA subunit is involved in catalyzing the breakage and ligation of the cut DNA. Negative supercoiling forms plectonemic loops (Fig. 1) or micro-domains from 2 kb to 65 kb with about 10 kb average size, depending on the studies. Such plectonemic loops are maintained by gyrases and topoisomerases to prevent relaxation of the entire chromosome [28-30].

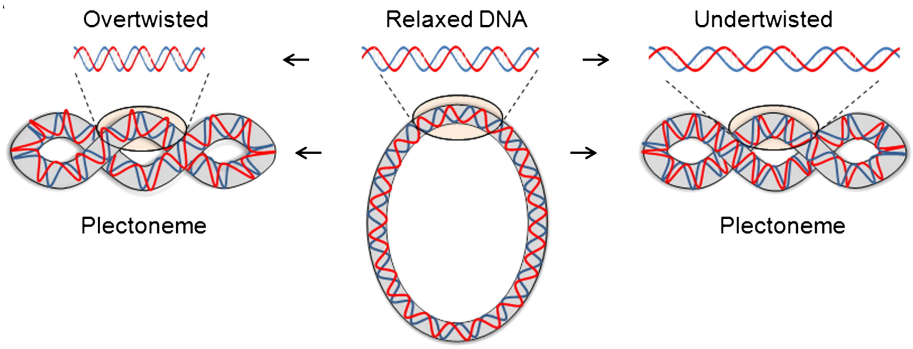

Figure 1: DNA topology. The DNA topology is described quantitatively by the twists of double helix and by the number of times the helix crosses over on itself (plectoneme). The figure was extracted from [31].

The best investigated example of source of DNA supercoiling is the transcription generated supercoiling. Indeed transcription and translation processes dynamically change DNA topology inducing local and temporal supercoiling of the DNA template. As the RNAP is immobilized in transcription factories, the DNA template being transcribed is forced to rotate around its axis as the double helix threaded through the transcriptional machinery [32]. In bacteria, as transcription proceeds, DNA in front of the transcription ensemble becomes positively supercoiled, and DNA behind the ensemble becomes negatively supercoiled [21, 33] (Fig. 2). A decrease in the degree of negative supercoiling elevates the transcription of the gyrA and gyrB genes and reduces the transcription of the topA genes whereas an increase in the degree of negative supercoiling has the opposite effects on the expression of those genes [34, 35]. The level of supercoiling induced by transcription depends on the rate of transcriptional elongation as well as the rate of transcriptional initiation and the possible RNA polymerase pausing [21]. Low levels of transcription thus produces torsional stress followed by DNA relaxation while high level of transcription with repetitive initiation may establish stable dynamic supercoiling upstream of the transcription start sites [36].

Figure 2: Illustration of the mechanism of transcription and supercoiling. A transcription ensemble R is illustrated including the polymerase, the nascent RNA and proteins bound to the RNA. If R is moving from left to right as indicated by the arrow, the DNA is front of the polymerase becomes overwound, or positively supercoiled while the DNA behind the polymerase becomes underwound, or negatively supercoiled. The figure was extracted from [21].

Additionally the distribution of promoters in divergent orientation could reinforce DNA supercoiling upstream of the transcription start sites by untwisting the double helix as well as by inducing directly plectonemes [37].

Another source of supercoiling is provided by the reorganization of eukaryotic chromatin with the assembly and disassembly of nucleosomes. Eukaryotic organisms lack enzymes such as DNA gyrase that directly introduce supercoils into DNA, but have nucleosomes and statically their genome is supercoiled to a similar degree of bacterial genome [38]. Each nucleosome of the chromosome is wrapped by DNA 1.8 times which constrains the chromosome until released by nucleosome removal [39]. Special proteins called chromatin remodelers complexes are indeed able to remove or slide nucleosomes in an ATP-dependent fashion [40, 41].

b) Molecular crowding

The second factor that condenses the genome is the molecular crowding that causes strong depletion and attraction forces [42, 43]. It has been proposed that genome organization is mainly entropy-driven [44] as nucleoid is expanded in low-crowding conditions and compacted under high-crowding conditions [45]. However, crowding forces are non-specifically driving compaction and therefore cannot regulate specific DNA interactions. Still this crowded environment could contribute to DNA-DNA

interactions and NAP interactions with genomic DNA [42] that also have an effect on genome folding.

c) Polyamines

Additionally, some studies revealed the essential role of polyamines in chromosome condensation in mammalian cells [46] and bacterial cells [47, 48]. In prokaryotes, the most abundant polyamines are putrescine and spermidine that account for the majority of intracellular cationic charge [49]. The interaction of polyamines with DNA induces conformational changes such as transitions from B to A and Z DNA forms [50] or DNA bending [51] as well as modify the interactions of DNA with sequence-specific DNA-binding proteins [52]. Moreover at high cellular abundance, spermidine stabilizes the condensed bacterial chromosome in isolated nucleoid, suggesting an important role of polyamines in the compaction of DNA in bacterial cells [53, 54].

d) Nucleoid-associated proteins

The fourth important factor in DNA compaction is the action of nucleoid-associated proteins NAPs, which also play a role in chromosome segregation and DNA repair [55-57]. NAPs were initially referred to as ‘histone-like’ proteins [58, 59], by analogy to the histone eukaryotic proteins that alter the shape of DNA, but they are now referred to as NAPs, reflecting their cellular location. Interestingly, some NAPs seem to have a clear role in compaction, whereas others seem to act both as compacting agents and as antagonists of compaction. The best characterized ones, due to their high intracellular abundance, which depends on growth phase [60, 61], are heat unstable (HU) [62], factor for inversion stimulation (Fis) [63], integration host factor (IHF) [64] and histone-like nucleoid structuring (H-NS) [65]. A non-exhaustive list of NAPs found in gram-positive and gram-negative bacteria and their different functional interactions with DNA has been published [33], where they were classified into three categories: DNA-wrappers, DNA-bridgers, and DNA-benders (Table 1).

Protein	Bacteria	Binding motif	Molecular Mass	Function	Native protomer	Refs
HU	Gram-negative ; Gram-positive	DNA structural motif in dsDNA to either dsDNA or ssDNA with preference for AT-rich or curved DNA ; ND	~9 kDa ; ~10 kDa	DNA wrapping and bending ; DNA bending	Heterodimer (for example HUα-HUβ); Homodimer	[62, 66-70]
Fis	Gram-negative	A-6-tracts and AT tracts	~11 kDa	DNA wrapping bridging and bending	Homodimer	[71, 72]
IHF	Gram-negative	(A/T)ATCAANNNNT T(A/G)	~11 kDa	DNA bending	Heterodimer (IHFα-IHFβ)	[73, 74]
H-NS	Gram-negative	AT-rich DNA and TCGATAAATT	~15 kDa	DNA bridging	Homodimer or heterodimer (H-NS–StpA)	[75]
Lrp	Gram-negative ; Gram-positive	(T/C)AG(A/T/C)A(A/T)ATT(A/T)T(A/T/G) CT(A/G) ; ND	~18 kDa ; ~17 kDa	DNA wrapping and bridging	Homodimer	[76, 77]
MukB	Gram-negative ; Gram-positive	ND ; Preference for ssDNA	~175 kDa ; ~130 kDa	DNA bridging	Homodimer	[78, 79]
Dps	Gram-negative	ND	~19 kDa	ND	Monomer or dodecamer	[80]
StpA	Gram-negative	AT-rich DNA	~15 kDa ;	DNA bridging	Homodimer or heterodimer (StpA–H-NS)	[81]

CbpA	Gram-negative	Curved DNA	~33 kDa ;	ND	Homodimer or heterodimer (CbpA–CbpM)	[82]
CbpB	Gram-negative	Curved DNA	~33 kDa ;	ND	Monomer	[83]
EbfC	Gram-negative	GTNAC	~11 kDa ;	DNA bridging suggested	Homodimer	[84]
MvaT	Gram-negative	AT-rich DNA	ND	DNA bridging	Homodimer	[85]
Lsr2	Gram-positive	AT-rich DNA	~12 kDa ;	DNA bridging	Homodimer	[86]
Hlp	Gram-positive	ND	~21 kDa ;	ND	Monomer	[87]
MrgA	Gram-positive	ND	~17 kDa ;	ND	Monomer or dodecamer	[88]

Table 1: Nucleoid-associated proteins in bacteria. Legend for abbreviations: Curved-DNA-binding protein A (CbpA); Curved-DNA-binding protein B (CbpB), also known as Rob; DNA protection from starvation (Dps); Double-stranded DNA (dsDNA); Factor for inversion stimulation (Fis); Histone-like protein (Hlp); Histone-like nucleoid-structuring (H-NS); Integration host factor (IHF); Leucine-responsive regulatory protein (Lrp); Metalloregulation DNA-binding stress protein (MrgA); Not determined (ND); Single-stranded DNA (ssDNA). This table was adapted from [57].

Similarly as histone proteins in eukaryotes, which were thought to be the dominant mode of DNA packaging, the first family of DNA wrappers induces a considerable volume reduction of the DNA. Indeed, since bacteria mostly lack histone proteins, there are additional NAPs that help in DNA compaction (Table 1). The second family of DNA bending NAPs can relieve repression or counteract the effect of the third family of DNA bridging NAPs, which are thought to repress transcription. Indeed, DNA bridges have the potential to trap RNA polymerase (RNAP) and exclude it from the promoters of the genes concerned. Moreover, an evidence of co-localization of such proteins with RNAP is consistent with this trapping mechanism [89, 90].

From the first family, leucine-responsive regulatory protein (Lrp) forms disc-shaped octameric structures with multiple binding sites that wrap DNA around themselves in a right-handed superhelix [77]. This wrapping constrains positive supercoiling and compacts DNA.

From the second family, HU, which might be the most universally conserved and abundant NAP, can induce DNA bends, condense DNA in a fibre and can also interact with single stranded DNA (ssDNA) [28-33]. HU has similar overall structure and shares several conserved regions with IHF [91]. IHF can bend DNA strongly to specific DNA sites [64] and participates in forming higher-order DNA structures required for replication, site-specific recombination, phage packaging or regulation of transcription initiation [73, 74, 92]. It can also influence global transcription in *E. coli* [93]. IHF can also bind to DNA non-specifically and therefore can be substituted by HU. In addition, Fis, which is abundant in the early exponential phase of growth, can bend DNA and has considerable affinity for non-specific DNA with high AT-content [94].

From the third family of DNA bridging NAPs, H-NS can form bridges between adjacent tracts of double-stranded DNA, preferentially in high AT-content regions [75]. The distribution of its binding sites is found in the same location of domain loop boundaries [95]. H-NS is widely found in bacteria and have been studied in pathogens such as *Mycobacterium tuberculosis* [96] and *Vibrio cholerae* [97].

All those proteins act in concert either with others members of the same family, either with NAPs from another family leading to synergistic or antagonistic effects [98, 99]. Those antagonistic effects also play a role at the level of global nucleoid organization and could be responsible for the considerable heterogeneity within the nucleoid [100]. Indeed the action of HU, IHF or Fis could result in regions that are less compact than others and this local reorganization of the nucleoid could affect the expression of genes that are hundreds of bases away [100].

All the proteins mentioned above have relatively small size of about 30 kDa for most of them. There are also larger proteins called structural maintenance of chromosome (SMC) that are able to act on larger distances and might be the most conserved of all architectural proteins. They are able to build DNA bridging and can promote long-range interactions between DNA segments and condense the chromosome to facilitate its

segregation [101, 102]. They were found within or near the origin of replication [103, 104]. For example, MukB is a SMC homolog protein from *E. coli* that can also form bridges and have a role in efficient segregation of chromosomes during cell division [78].

NAPs are also altering the level of DNA supercoiling [105]. For example, HU interacts with topoisomerase I leading to alterations in the superhelicity of DNA, nucleoid structure and gene expression [106]. MukB has been implicated in the formation of independent topological domains in the *E. coli* chromosome probably in association with DNA gyrase [107]. Also H-NS and Fis are thought to assist directly the supercoiling of domains by forming topological barriers on the *E. coli* chromosome [108]. However, their depletion do not affect the global nucleoid structure [109]. This result suggests that additional factors could be important to the maintenance of nucleoid structure. In fact, about half the amount of supercoiling is not constrained by proteins and is thought to be present in the form of plectonemes. Therefore, it has been proposed that transcription from highly active promoters can also introduce new domain boundaries [110, 111]. The fact that transcription of ribosomal RNA (rRNA) operons takes place in the transcription factories in the cell, which have been linked to a role in DNA compaction, might be related to this hypothesis [112]. Rather than being important for the stabilization or formation of large DNA loops, NAPs could be essential for the local structure of smaller loops within these larger ones [100]. Following this hypothesis, a recent study in *Caulobacter crescentus* revealed that its genome is divided into 23 chromosome interacting domains (CIDs) or highly self-interacting regions, equivalent to the TADs found in eukaryotes, but with a ranging size of 30 to 400 kilobases (kb) [113]. Their formation seems to be related to the presence of highly expressed genes where the DNA is kept free of plectonemic loops by active transcription [113], rather than the deletion of NAP or SMC proteins as it was previously suggested [114, 115]. An alternative hypothesis is that domain boundaries arise through attachment of DNA to the membrane that might occur during transertion and expression of transmembrane proteins [23, 116].

In addition to their role in condensing the chromosome, changes in DNA supercoiling could control transcription in bacteria [117-120], as promoter activities are sensitive to the local level of supercoiling and can be increased or decreased in response to super-

helical variation [119]. This could be more important in small genomes such as *Mycoplasma pneumoniae* or *Mycoplasma genitalium* [121] where many structural DNA-binding proteins are absent [122], and code for few transcription factors (TFs) and only two sigma factors (Table 2). Therefore, in such genomes gyrases and topoisomerases might be controlling gene expression through changes in the DNA local structure [121]. Indeed *M. pneumoniae* only has few TFs compared to other bacteria, with MPN529: IHF-HU possibly affecting DNA topology [123], MPN426: SMC family, MPN 229: SSB binding ssDNA [124], MPN 554: binding ssDNA [125] and possible evidence for a homolog of CbpA MPN002: xdj1.

Gene number	Gene name	Protein name
MPN002	*cbpA*	Curved DNA-binding protein CbpA
MPN003	*gyrB*	DNA gyrase subunit B
MPN004	*gyrA*	DNA gyrase subunit A
MPN089	*hsdS*	Putative type-1 restriction enzyme specificity protein
MPN122	*parB*	DNA topoisomerase 4 subunit B
MPN123	*parC*	DNA topoisomerase 4 subunit A
MPN124	*hrcA*	Heat-inducible transcription repressor hrcA
MPN229	*ssbA*	SSB binding single stranded DNA (ssDNA)
MPN239	*gntR*	Probable HTH-type transcriptional regulator gntR
MPN241	*whiA*	Transcription factor with WhiA C-terminal domain
MPN266	*spxA*	Transcriptional regulator Spx
MPN275	*ybaB*	DNA-binding protein, YbaB/EbfC family
MPN289	*HsdS1B*	Putative type-1 restriction enzyme specificity protein
MPN294	*araC*	AraC-like transcriptional regulator
MPN332	*lon*	ATP-dependent protease La (EC 3.4.21.53)
MPN352	*sigA*	RNA polymerase sigma factor rpoD (Sigma-A) (EC 2.7.7.6)
MPN424	*ylxM*	Putative helix-turn-helix protein, YlxM/p13-like protein
MPN426	*smc*	SMC family, chromosome/DNA binding/protecting functions
MPN478	*yrbC*	YebC family protein (transcription factor of the tetR family)
MPN529	*ihf*	Histone-like bacterial DNA-binding protein
MPN554	*ssbB*	Putative single-stranded DNA-binding protein
MPN572	*pepA*	Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase) (LAP) (Leucyl aminopeptidase)
MPN608	*phoU*	Transcriptional regulator involved in phosphate transport system
MPN615	*hsdS*	Putative type-1 restriction enzyme specificity protein
MPN626	*mpn626*	Alternative sigma factor
MPN638	*hsdS*	Putative type-1 restriction enzyme specificity protein
MPN686	*dnaA*	Chromosomal replication initiator protein dnaA
MPN688	*SojA/ParA*	Member of the ParA family of ATPases involved in plasmid and chromosomal segregation

Table 2: List of *Mycoplasma pneumoniae* transcription factors, possible DNA binding structural proteins and sigma factors.

In summary, the role of individual proteins in chromatin organization has been widely studied [56, 61, 79, 100]. However individual and in concert contributions of supercoiling, molecular crowding and NAPs to genome compaction are yet to be fully elucidated. Indeed as mentioned before, macromolecular crowding increases the association of NAPs [42] and moreover the binding of NAPs is enhanced by supercoiling. To progress in the understanding of chromatin organization, studying on a

global level, rather than understanding in detail the role of specific proteins, has enabled elucidating genome structure. For example, fluorescence imaging of living *C. crescentus* cells revealed that its chromosome is orderly arranged [126]. It was therefore suggested a chromosome model in which the two arms of the chromosomes are arranged linearly as a series of loops perpendicular to the cell axis with the origin of replication (Ori) found at the flagellated pole of the bacterium and the terminal of replication (Ter) at the opposite end of the cell [95]. Similarly *E. coli* and *B. subtilis* genomes were also found to fold in the linear order of genes [127-129], although in *E. coli*, Ori and Ter were found in the cell centre. Additionally, it was shown that *E. coli* genome consists of four macro-domains of about 1Mb each and two less-constrained regions [130] that influence the segregation and mobility of the chromosome [131]. Those studies have shown evidence that chromosomal DNA is not distributed randomly within the cell and has a highly conserved organization. More recent approaches such as Chromosome Conformation Capture (3C) [132] helped in elucidating the three-dimensional genome organization of genomes and the processes that shape the chromatin structure.

1.2 Modeling approaches of 3D genomes and their limitations

The 3D organization of chromosomes has traditionally been studied by imaging methods such as fluorescent *in situ* hybridization (FISH), that uses florescent probes to bind to the genomic regions of interest, and then measures the spatial distances between pairs of florescent probes. But such techniques are limited by low throughput, low resolution and probe sequence specificity, resulting in an analysis of a few hundred cells [133]. Complementary to those imaging techniques, the latest developments of 3C-based techniques [132] enabled the determination of global chromosome organization of genomes or genomic domains. Indeed such techniques give information on the physical position of genomic regions in the 3D organization of the genome and measure the frequency at which those regions interact in the 3D space. In 3C-based technologies chromatin is cross-linked with formaldehyde, in such a way that only DNA regions that are covalently linked together form ligation products [134]. In this respect, an important difference arises between eukaryotes and prokaryotes. While eukaryotic DNA is almost fully covered by protein complexes (nucleosomes), prokaryotic DNA, have DNA regions not occupied by proteins [135]. This raises the question of how formaldehyde crosslinking works since it ligates primary amines and bases. One explanation could be the fact that DNA is covered by polyamines that bind unspecifically and neutralize the DNA charge. Thus polyamines could act as crosslinking bridges in prokaryotes.

a) Experimental methods

The principle of 3C-based experiments is based on cells that are first cross-linked by formaldehyde, then digested by restriction enzymes, as for example using HindIII, and finally the digested fragments that were originally close in the 3D space, are ligated. For 3C, ligation products are quantified by PCR using locus specific primers to measure the relative frequency of interactions of the regions assessed and identify chromatin loops formed in regions of several hundreds of kilobases distance. Further development of 3C-based methods allowed to identify interactions between regions of larger distances, with 4C [136, 137], 5C [138] and Hi-C [139]. For 4C, inverse PCR is assessing the

interaction frequency of single loci against the whole genome, generating one-to-many interaction profiles [136, 137]. 5C generates many-to-many interaction profiles using annealing and ligating of oligonucleotides in a multiplex setting [138]. Finally Hi-C generates genome-wide interaction profiles using biotin to isolate the ligation products [139] (Fig. 2).

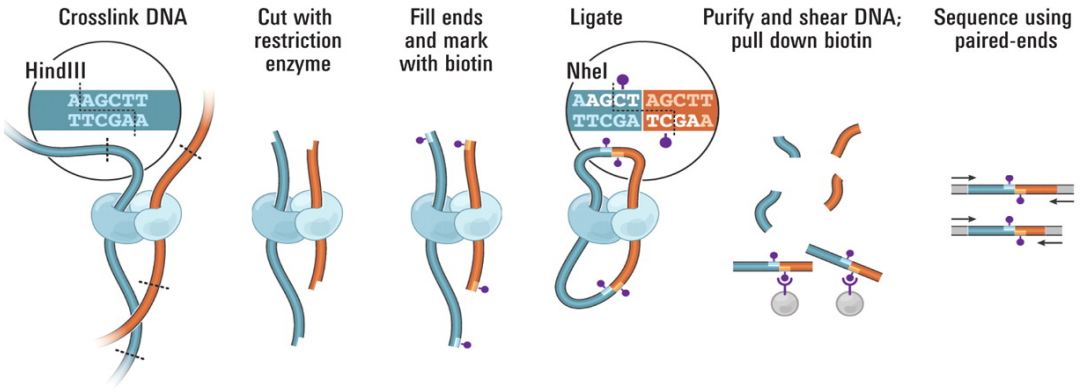

Figure 3: Hi-C overview. Cells are cross-linked with formaldehyde, resulting in covalent links between spatially adjacent chromatin segments (DNA fragments: dark blue, red; Proteins, which can mediate such interactions, are shown in light blue and cyan). Chromatin is digested with a restriction enzyme and the resulting sticky ends are filled in with nucleotides, one of which is biotinylated (purple dot). Ligation is performed under extremely dilute conditions favoring intramolecular ligation events. DNA is then purified and sheared, and biotinylated junctions are isolated using streptavidin beads. Finally, paired-end sequencing identifies interacting fragments. The figure was extracted from [139].

More recent high-throughput derivation techniques, which improve the ligation step performance, have also been published with Tethered Conformation Capture (TTC) on a solid phase support [140] or inside the nucleus with in situ-Hi-C [141]. Additionally, chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [142] can detect chromatin interactions bound by specific proteins at base-pair resolution. And the latest Capture-C [143] and T2C [144] techniques that use oligonucleotide capture technology with 3C and high-throughput sequencing allow to interrogate selected genomic regions at high-resolution.

Compared to the microscopic techniques and other existing 3C-based methods, Hi-C technology provides high throughput genome-wide chromatin interaction maps over a large population of cells. The preprocessing of paired-end reads obtained from Hi-C experiments is an essential step of the Hi-C analysis that allows removing and

correcting experimental artifacts. First paired-end reads are mapped to the reference genome, then reads are filtered, followed by fragment-level filtering and finally reads are pooled into bins to create a Hi-C matrix. The mapping step can be done using softwares such as Bowtie [145], GEM [146] or MAQ [147]. Originally, the read length was fixed and a pair of reads was conserved only if both reads of the pair map uniquely with this fixed length. Otherwise if one of the two reads did not map in a unique site with the selected length, the pair of reads was discarded. An alternative was proposed to avoid discarding so many pairs of reads, with the iterative mapping strategy [148] that supports different read lengths for each read, resulting in a larger number of mapped reads. Then, mapped reads are subjected to filtering to discard self-circle and dangling ends products, given the directions of the two reads (Fig. 4). Additionally, redundant reads due to PCR amplification also need to be removed from the analysis. The remaining reads will then be subjected to fragment-level filtering to remove fragments whose mappability score is too low, as well as reads from restriction fragments that are too short (<100 bp) or too long (>100 kb) in length and to remove the top 0.5% fragments with the greatest number of reads which are likely arising from PCR amplification artifacts [148]. Finally reads are then pooled into bins or equal sized genomic loci to create a contact matrix at a resolution that depends on the total number of reads or sequencing depth and the restriction enzyme used.

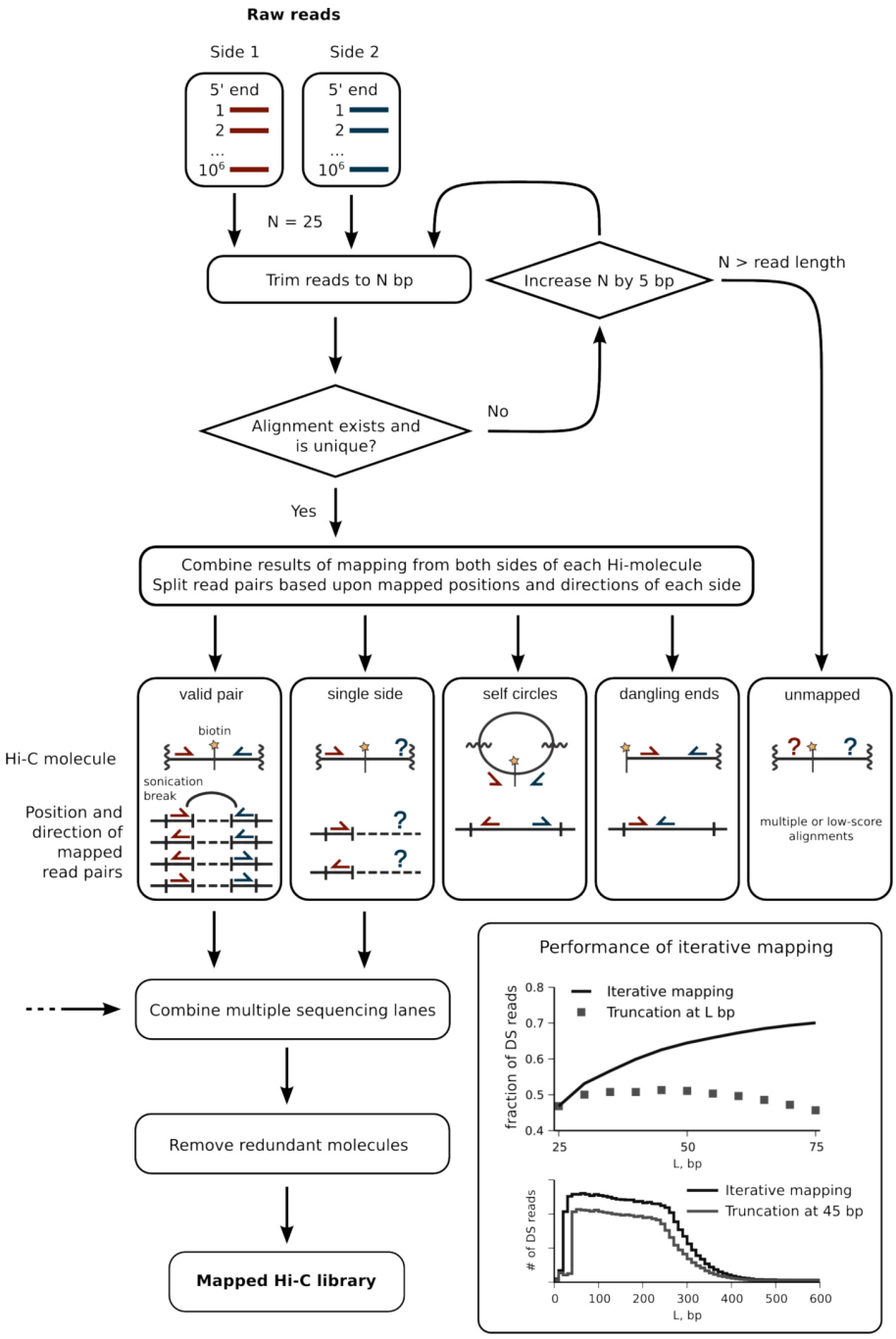

Figure 4: Iterative mapping and filtering pipeline. This flowchart illustrates how read pairs are separately mapped to the genome at increasing truncation lengths and collected if uniquely mapped. Later reads are classified based on the relative positions and directions of both side, only valid pairs and single-side are conserved for further analysis and self-circles, dangling ends and unmapped are discarded. This image was extracted from [148]

The main sources of biases of Hi-C experiments are restrictions sites that are not uniformly distributed along the genome, digestion efficiency, GC content, sequence uniqueness and fragments lengths that leads to variable ligation efficiency [149].

Several approaches have been recently proposed to correct and reduce those biases [148-150]. The first method published uses a probabilistic framework to remove systematic biases and significantly increases the reproducibility between biological replicates [149]. An alternative method was proposed, based on Poisson regression model, that has a reduced number of parameters and a shorter computing time [150]. More recently the iterative correction and eigenvector decomposition (ICE) method, which is based on the equal visibility assumption, resulted in good reproducibility between replicates with different restriction enzymes [148]. Compared with previous methods, which required the specification of known biases, this method can remove any type of known or unknown biases. However this normalization method is only valid for equal sized genomic loci, or bins. Although those methods are effectively reducing biases, additional controls are necessary for further validation such as the reproducibility of biological replicates.

b) Modeling methods

The resulting filtered interaction matrices have been extensively used for computationally analyzing the organization of genomes and genomic domains [151]. In particular, a significant number of approaches for modeling the 3D organization of genomes have recently flourished [140, 152-156]. The main goal of such approaches is to provide an accurate 3D representation of the bi-dimensional interaction matrices, which can then be more easily explored to extract biological insights. Chromatin modeling is performed using two main complementary methodologies to simulate an ensemble of 3D conformations of the chromosomes, compatible with input contact maps [151]. The first, known as restraint-based (RB) modeling assumes a relationship between the frequencies of interactions of genomic regions and their distances in 3D space, transforming those frequencies into spatial restraints and aiming to satisfy as many distance constraints as possible. The second, called thermodynamics-based (TB) modeling, simulate polymer fibers applying physical principles of the chromatin fiber, with the aim to identify genome conformations that, as an ensemble, reproduce the observed experimental frequencies.

Within RB approaches, there are two main categories, those that aim to find a single

solution of genome conformation and those that explore the variability of the experimental data over a population of cells and simulate a large number of solutions of genome conformation. The first category that converts the interaction matrix into a 3D object [154, 157], are more appropriate for single-cell 3C-based studies [158]. In the second category, frequencies of interactions between genomic regions are converted into a set of spatial restraints to build 3D models of the genome by satisfying as many input restraints as possible using Monte Carlo sampling [140, 156, 159] or Bayesian approaches [153, 155, 160]. Some of the approaches are doing several independent simulations, based on restraints on the entire interaction map, to reproduce the experimental variability, while others approaches are simulating sub-populations of genomes, where each subpopulation restraints are based on a selection of interactions, and in this case, the sub-populations are representing the experimental variability [140].

RB approaches are able to simulate genome conformation faster than TB approaches, as they do not fully take into considerations the physical properties of chromatin fiber. Unfortunately, TB methods simulations are time-consuming and not suitable for simulations of full eukaryotic chromosome. Additionally, TB approaches require *a priori* characterization of the chromatin fiber, which could vary across the population of cells, or not always be experimentally defined. Whereas RB approaches do not take chromatin fiber properties as input in the simulations and therefore might be more suitable to reproduce the heterogeneity of population of cells.

RB methods have been already applied to define chromosomal organization features as for example for the identification of chromatin globules in the alpha-globin domain of the human genome [161], the ellipsoidal conformation of *C. crescentus* genome [162], the genomic organization of the yeast [159], the spatial organization of the X inactivation center in the human genome [156] and the effect of hormones on TADs structure [163]. However, no internal and systematic analysis of the accuracy of the resulting models has been performed and only an assessment of the reproducibility of these 3D reconstruction methods has been addressed [164]. The goal of the first chapter of the thesis is to address the lack of such analysis by assessing the limits of 3D reconstruction using restraint-based modeling approaches. We focused on a recently developed RB method for modeling 3D structures of genomes and genomic domains called TADbit [152], that was developed around the Integrative Modeling Platform (IMP, http://integrativemodeing.org), a general framework for restraint-based modeling

of 3D bio-molecular structures [165]. Although our analysis was based solely on models generated by TADbit, the conclusions are likely to hold for alternative mean-field restraint-based approaches.

1.3 Application on *Mycoplasma pneumoniae* to understand the impact of structural organization in transcriptional regulation

As previously mentioned, 3C-based approaches helped in elucidating the bacterial chromosome organization and its regulation. Such studies have been carried out in *B. subtilis* with a Hi-C map at 30kb resolution [166], *E. coli* at 20kb [167] and *C. crescentus* at 13kb [162] showing that genome structure is globally related to the process of chromosome segregation in *C. crescentus* and DNA replication and transcription in *E. coli*. More recently, a high-resolution Hi-C map of *C. crescentus* at 10kb [113] revealed that its genome is divided into 23 CIDs that are not affected by the deletion of HU or SMC proteins. No such domains were described in the lower resolution Hi-C maps of *B. subtilis* and *E. coli*. Moreover, it was shown that histone-like proteins in *E. coli* do not contribute to the global organization of the genome [167]. These bacteria have large and complex genomes with sizes above 4 Mbp, more than 300 regulatory TFs in *E. Coli* [168] and 120 TFs in *B. subtilis* [169], multiple DNA structural proteins, and several sigma factors that play key roles in responding to physiological and environmental signals [170]. The goal of the chapter two of this thesis is to understand how a bacterial chromosome structure organization is achieved and identify its role in transcriptional regulation. To do so, we have looked for a genome reduced bacterium, *M. pneumoniae*, a human pathogenic bacterium causing atypical pneumonia [171].

a) *Mycoplasmas*

The infection caused by *M. pneumoniae* is slow to develop and includes symptoms such as fever, cough, sore throat and hoarseness. It is commonly treated with antibiotic [172]. Like retroviruses, *Mycoplasmas* have the capacity for cellular invasion, self-replication [173] and the initiation of a variety of immune responses. However, unlike viruses, they are viable in body fluids and do not require living cell hosts for DNA replication and growth. They are distinguished phenotypically by their minute size and a lack of a cell wall [174], bounded by a plasma membrane only, which taxonomically separate them

from other bacteria in a class named *Mollicutes* (*mollis*, soft; *cutis*, skin, in Latin). The term mycoplasma (*mykes*: fungus and *plasma*, formed, in Greek) replaced the term pleuropneumonia-like organisms (PPLO) that was found in various diseases of animals and human in early 1900 [175]. The allusion to a fungus in their name was originally describing the growth of *Mycoplasma mycoides*, but it was conserved afterwards. Within the class, there are 102 species of mycoplasmas and they are phylogenetically related to the eubacterial subgroup of gram-positive that included the bacilli, streptococci and lactobacilli [176]. Additionally, they are characterized by their small genomes consisting of a single circular chromosome containing 0.58 to 1.35 kilobases (kb) with a low GC content [177]. In 1989, the small subunit of the 16S rRNA was used to develop a classification system over 50 mycoplasmas species and their walled relatives [178]. The class *Mollicutes* was therefore divided into five groups: the pneumoniae group, the hominis group, the spiroplasma group, the anaeroplasma group and a fifth group containing only the *Asteroleplasma anaerobiu*. *Mollicutes* are thought to have evolved from a common ancestor with *Firmicutes* through successive genome losses [179]. The result of a gradual reduction in genome size from a common ancestor is known as degenerative evolution [176].

Mycoplasmas usually exhibit tissue preference with *M. pneumoniae* preferentially found in respiratory tract. The closest relatives of *M. pneumoniae* are the human urogenital pathogen *M. genitalium* [180] and the avian pathogen *Mycoplasma gallisepticum* [181]. In contrast to *M. pneumoniae*, *M. genitalium* infections are typically found in the urogenital tract [180] and *M. gallisepticum* infections are in the respiratory tract and conjunctiva of avian species [181]. With the lack of a cell wall, *Mycoplasmas* are pleomorphic [182] and have different morphologies ranging from rod-shaped with *Mycoplasma insons* [183], coccus-shaped with *Mycoplasma hyopneumoniae* [184], to flask-shaped with *M. pneumoniae*.

b) *M. pneumoniae* morphology

M. pneumoniae grows forming dense networks on the surface and in colonies composed mainly of rounded and elongated forms [185]. Morphological changes from spherical to filamentous have been observed with scanning-beam electron microscope (SEM) in cultures of cells grown and fixed in liquid suspension, varying during its life cycle

[186]. In early growth phase, from 8h to 2 days of growth, the predominant morphology is symmetric round forms with aggregates and clusters of tightly packed spherical cells. While after 2 to 6 days of growth, cells have filamentous both branched and straight forms, as well as flask-shaped in microcolony, to finally turn with asymmetrical and larger round forms in later growth phases. The first images of transmission electron microscopy (TEM) with carbon replicas [187] showed three-dimensional shape of *Mycoplasmas* attached to surfaced, comparable to those acquired by SEM across the growth cycle [188].

Mycoplasmas represent the smallest self-replicating organisms, in both genome size and cellular dimensions [189]. According to our estimation, *M. pneumoniae* cell has an average length of 1.38 μm, which can reach up to 2.5 μm and an average width of 365 nm (Fig. 5), compared to a length of 1 to 4 μm in a typical bacillus and 0.5 to 1 μm in width. Regarding the cell volume, we estimated it to be of 0.08 μm^3, compared to 1 μm^3 in *E. coli*.

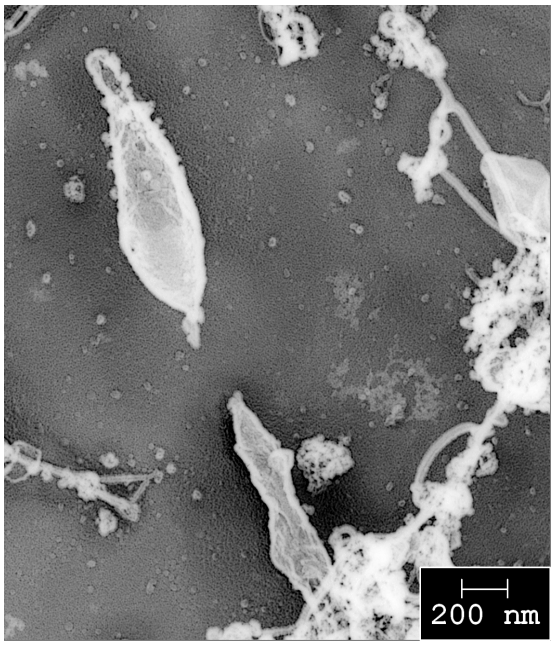

Figure 5 TEM image of *M. pneumoniae* cell with quick-freeze deep-etch replica

c) *M. pneumoniae* ultrastructure

The ultrastructure of *Mycoplasmas* was studied by electron microscopy, and a cystoskeleton-like structure was detected in *M. pneumoniae* with a rod-like structure in the end of the cell, the tip or attachment organelle (AO), as well as thin fibrous structures, cytoskeletal filaments, extending in the cell body [190-192] (Fig. 6). Indeed the flask shape of *M. pneumoniae* is conferred by the presence of this polarized structure, the AO that has a length between 220-300 nm long and about 50-80 nm width with a terminal button and a basal node [190].

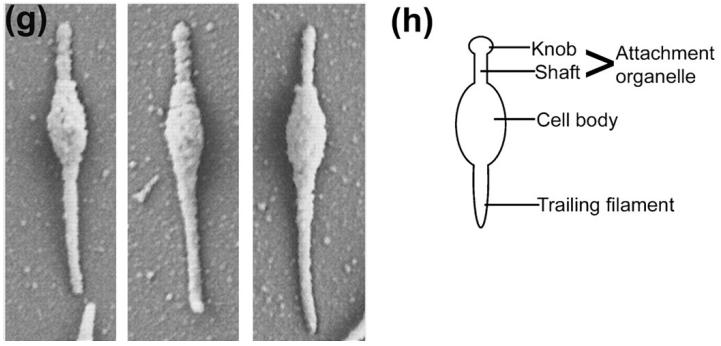

Figure 6: Scanning electron micrograph of Mycoplasma Pneumoniae cells grown on glass coverslips and schematic of mycoplasma cell. The figure was adapted from [192].

Images from cryo-ultrathin sectioning of the tip show a rod-like structure that striped perpendicular to its long axis [193] (Fig. 7).

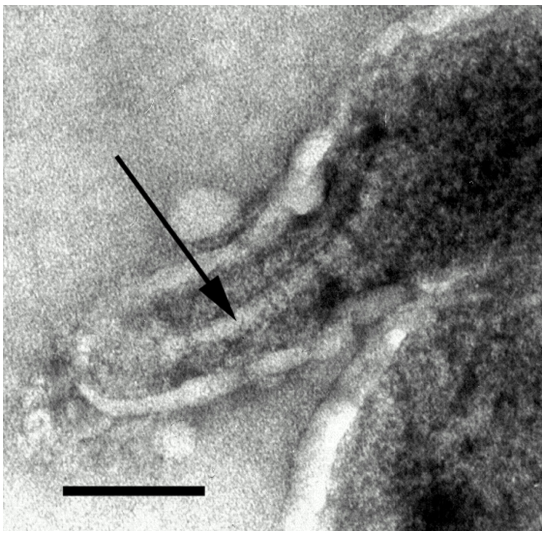

Figure 7: Ultrathin cryosection of M. pneumoniae. The arrow marks a structure in the tip of attachment organelle, which is striped perpendicular to its long axis. Bar =100 nm. The image was extracted from [193].

Later studies using cryo-electron tomography were able to visualize the cells three-dimensionally and resolved the structure of the tip [194] that is surrounded by an electron-dense complex and a structure at the proximal end of the rod that attaches the rod to the cell membrane. High-resolution images of the electron-dense core detail the structure of the core, which can be divided into three regions: a terminal button, a rod made up of two parallel subunits and a bowl-shaped base [195].

d) *M. pneumoniae* gliding motility

This complex terminal structure AO, that is not found in model bacteria like *E. coli* enables *M. pneumoniae* to adhere to the host cell surface [196] and give them ability to glide on surface like glass [197]. Similarly, *M. genitalium, Mycoplasma imitans, M. gallisepticum* and *Mycoplasma pirum* have an attachment organelle structure [198-200]. This membrane-bound extension of the cell, which is supported by cytoskeleton-like structure, is characterized by a dense cluster of the adhesin protein P1 [196, 201, 202] that enables the cell to interact productively with host cells by cytoadherence. Additionally P30 proteins are only found at the AO [203, 204]. Surface proteins P1 and P30 are thought to function as adhesins to allow successful colonization of the respiratory tract with mycoplasmas penetrance and adherence to the respiratory epithelium [205]. P1 is essential for virulence [196] and is considered to be the primary

adhesion protein. Indeed the addition of antibodies against P1 to a population of attached and motile cells resulted in a decrease in motility and a final release from the substrate [206]. Similarly, antibodies to P30 interfere with the attachment to surface [207]. Moreover P30 also plays a role in gliding motility, as mutants missing the protein are unable to glide [208] and mutants that are missing some of the C-terminal repeats of the P30 are capable of gliding but at a reduced speed [209]. Other essential proteins for attachment and gliding were found in the electron-dense core of the AO in addition to the adhesins such as HMW1 and HMW2, P24, P28, P65 and P41 [204, 210-213]. Most of these proteins are helping in stabilizing or localizing other proteins during the assembly of the electron-dense core, ensuring the proper formation of the AO [214]. The core proteins are therefore allowing the cytadherence of *M. pneumoniae*. Indeed without a core, the AO does not form properly because the adhesins do not localized in the tip and therefore the cell cannot adhere to host cells, making them avirulent [204]. A significant number of species of mycoplasmas are motile and the movement is in the direction of the AO [215]. Gliding motility is dependent on adherence via the adhesins found in the AO, leading the cell in a continuous unidirectional pattern. *Mycoplasma mobile* is the fastest of the species, as indicated by his name, with a speed of gliding motility of 2.0-4.5 μm/s [216] without rest periods, compared to an average speed of 0.3-0.4 μm/s in *M. pneumoniae* [197]. Their ability to move is related to the pathogenicity, allowing the cells to spread out during an infection, which is therefore an essential function of *M. pneumoniae* [217].

e) *M. pneumoniae* cell division

The cell division of the *Mollicutes* has been less studied than model bacteria such as *E. coli*, *B. subtilis* and *C. crescentus*. In *M. pneumoniae*, the cell division is linked to the AO [212], that is the leading end of the cell during gliding motility [182] and the location of the gliding motor [218]. AO has been suggested to have an essential role in the cell division process. It was shown that during cell division, the AO duplicates itself and move towards the opposite pole of the cell [211] (Fig. 8). Imaging of cells stained for P1, that marks the number and position of terminal organelles, and for DNA, using 4',6-diamidino-2-phenylindole (DAPI) allow to identify that duplication of the AO is also accompanied by an increase in the amount of DNA within the cell, that increases as

the distance between the two AO increases [211]. These data strongly suggest that AO duplication is linked not only to the cell division but also to DNA replication and might also be related to nucleoid segregation. In addition, it was shown that the nucleoid of *M. pneumoniae* appears to occupy nearly the entire volume of the cell [211]. Furthermore, in *Mycoplasma gallisepticum,* attachment organelles were found enriched for newly synthesized DNA, in a subcellular fraction of cells [219]. Altogether these results suggest a possible interaction between the chromosome and the AO during cell division. The segregation of chromosome could be driven by interactions between the chromosome and the migrating AO during cell division, further supported by the absence of partitioning machinery in *M. pneumoniae*, compared to model bacteria. We have thus investigated whether we observed such interactions between a specific region of the chromosome and AO that could give insights into this possible role of AO in chromosome segregation.

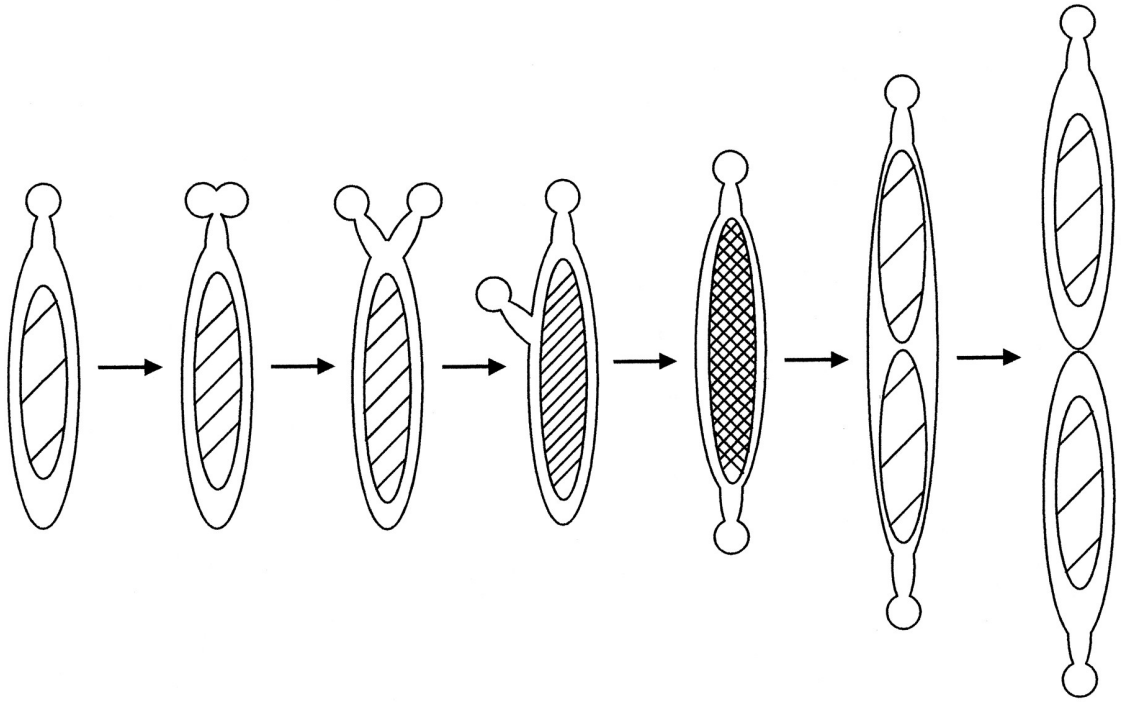

Figure 8: Model for cell division scheme in *M. pneumoniae* in relation to the formation and migration of attachment organelles. *M. pneumoniae* cell is represented with the AO at one cell pole and its nucleoid is shown in the center with oblique lines. The cell division process is depicted in the serie of images. First the cell has a single AO, then the AO is duplicated and migrate to the opposite cell pole, also accompanied by an increase in the amount of DNA. Finally the two cells segregates. The image was extracted from [211].

Later studies revealed that the new AO remains attached to the surface while the old AO pulls the cell away [220], allowing the migration of the old AO to the opposite pole before cytokinesis occurs. On occasion they observed that the duplication and separation of AO is followed directly by cytokinesis, as previously described [211], but for most dividing cells examined, additional AO developed before new daughter cells emerged [220] (Fig. 9).

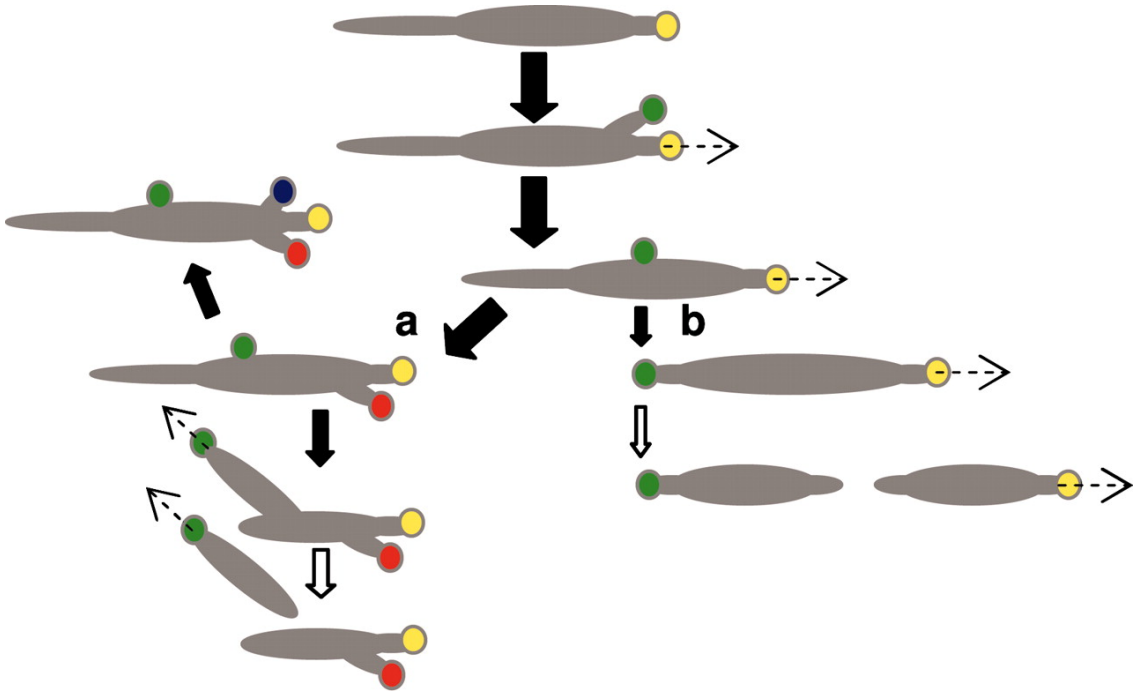

Figure 9: Model for *M. pneumoniae* terminal organelle duplication and growth cycle. The yellow circle represents the initial terminal organelle, and green, red, and blue circles represent subsequent organelles, appearing in that order. The dashed arrows reflect movement of the indicated terminal organelle, and open arrows indicate cytokinesis. Solid arrows indicate steps in the cell cycle, with arrow size reflecting relative frequency. In most cases, multiple duplications of the terminal organelle occur before daughter cells emerge (a), although rarely some cells do undergo a single duplication of the terminal organelle followed by cytokinesis (b), according to the previous model for cell division in *M. pneumoniae* [211]. The image was extracted from [220]

These data indicate that AO duplication and cytokinesis are not tightly coordinated. Whether the AO duplication is coordinated with DNA replication remains unclear, as the previous study was limited to *M. pneumoniae* cells having only one or two AO

[211]. Further characterization of the possible interaction between the AO and the chromosome would give insights into the dynamics of DNA replication in *M. pneumoniae* and the possible role of AO in DNA segregation.

f) *M. pneumoniae* transcriptional regulation

Mycoplasmas fast evolution has been marked by genome reduction and therefore a reduced coding ability, as well as a limited number of metabolic pathways [179]. *M. pneumoniae* has a single circular double-stranded genome of only 816,394 bp [221], which is about five smaller than the *E. coli* genome comprising 4,639,221 bp [222]. It codes for 694 ORFs (32 of which are smORFs having less than 100aa), 311 ncRNAs and 43 conventional RNAs [223], compared with about 4,300 genes in *E. coli* [222] and more than 20,000 in human [224]. The related species *M. genitalium* have even fewer genes with only 482 protein-coding genes [225]. The small genome of *M. pneumoniae* and its limited biosynthetic capabilities requires a complex medium for their in vitro culturing. This organism is particularly sensitive to osmotic stability, due to the lack of a rigid cell wall, and desiccation. As its reduced genome is less complex than those of other bacteria, *M. pneumoniae* offers a unique model to understand an entire organism that can be grown axenically in a laboratory.

In 1984, Morowitz was the first researcher to aim at defining the comprehensive machine of mycoplasmas [226]. Indeed, the organism is ideal not only to study genomes of a minimal cell, but also to understand the evolution of "reduced genomes" and give insights into the minimal cellular requirements. In recent years, *M. pneumoniae* has been systematically characterized in a quantitative manner, revealing an unexpected complexity of its transcriptome, the main components of its proteome organization as well as metabolism organization [223, 227-233]. Altogether, it has been proposed that the minimal essential genome for living mycoplasmas is comprised of 33% of the total genome (269,410 bp) [223]. Despite its genomic reduction, efficient antisense detection and the precise mapping of transcription start sites and untranslated regions, as well as transcriptional responses to perturbations, revealed that *M. pneumoniae* transcriptional machinery is much more complex than previously thought [227].

With few DNA binding protein, transcription factors (Table 2) and only two sigma factors: *MPN352* and *MPN626* that recognizes a slightly altered version of a standard

sigma 70 promoter region [229], this organism is an ideal biological model to study the impact of genome architecture organization on transcriptional regulation. Indeed compared to the most studied bacteria *E. coli, C. crescentus and B. subtilis* with several TFs and DNA bindings proteins, *M. pneumoniae* is ideal to uncover the specific role of chromatin structure in transcriptional regulation, in the absence of many NAP, which are one of the most important factors of DNA compaction.

Marie Trussart, François Serra, Davide Baù, Ivan Junier, Luís Serrano, and Marc A. Marti-Renom*: Assessing the limits of restraint-based 3D modeling of genomes and genomic domains, Nucleid Acids Research 2015, 43 (7): 3465-3477 doi:10.1093/nar/gkv221

2. ASSESSING THE LIMITS OF RESTRAINT-BASED 3D MODELING OF GENOMES AND GENOMIC DOMAINS

2.1 Abstract

Restraint-based modeling of genomes has been recently explored with the advent of Chromosome Conformation Capture (3C-based) experiments. We previously developed a reconstruction method to resolve the 3D architecture of both prokaryotic and eukaryotic genomes using 3C-based data. These models were congruent with fluorescent imaging validation. However, the limits of such methods have not systematically been assessed. Here we propose the first evaluation of a mean field restraint-based reconstruction of genomes by considering diverse chromosome architectures and different levels of data noise and structural variability. The results show that: first, current scoring functions for 3D reconstruction correlate with the accuracy of the models; second, reconstructed models are robust to noise but sensitive to structural variability; third, the local structure organization of genomes, such as Topologically Associating Domains, results in more accurate models; fourth, to a certain extent, the models capture the intrinsic structural variability in the input matrices; and fifth, the accuracy of the models can be *a priori* predicted by analyzing the properties of the interaction matrices. In summary, our work provides a systematic analysis of the limitations of a mean-field restrain-based method, which could be taken into consideration in further development of methods as well as their applications.

2.2 Introduction

Recent studies of the three-dimensional conformation of genomes are revealing insights into the organization and the regulation of biological processes, such as gene expression regulation and replication [15, 16, 141, 151, 234, 235]. The advent of the so-called Chromosome Conformation Capture (3C) assays [132], which allowed identifying chromatin-looping interactions between pairs of loci, helped deciphering some of the key elements organizing the genomes. High-throughput derivations of genome-wide 3C-based assays were established with Hi-C technologies [139] for an unbiased identification of chromatin interactions. The resulting genome interaction matrices from Hi-C experiments have been extensively used for computationally analyzing the organization of genomes and genomic domains [151]. In particular, a significant number of new approaches for modeling the three-dimensional organization of genomes have recently flourished [152-156, 236]. The main goal of such approaches is to provide an accurate 3D representation of the bi-dimensional interaction matrices, which can then be more easily explored to extract biological insights. One type of methods for building 3D models from interaction matrices relies on the existence of a limited number of conformational states in the cell. Such methods are regarded as mean-field approaches and are able to capture, to a certain degree, the structural variability around these mean structures [237].

We recently developed a mean-field method for modeling 3D structures of genomes and genomic domains based on 3C interaction data [152]. Our approach, called TADbit, was developed around the Integrative Modeling Platform (IMP, http://integrativemodeing.org), a general framework for restraint-based modeling of 3D bio-molecular structures [238]. Briefly, our method uses chromatin interaction frequencies derived from experiments as a proxy of spatial proximity between the ligation products of the 3C libraries. Two fragments of DNA that interact with high frequency are dynamically placed close in space in our models while two fragments that do not interact as often will be kept apart. Our method has been successfully applied to model the structures of genomes and genomic domains in eukaryote and prokaryote organisms [161, 162, 239]. In all of our studies, the final models were partially validated by assessing their accuracy using Fluorescence *in situ* hybridization (FISH) imaging. However, no internal and systematic analysis of the accuracy of the resulting models

has been performed and only an assessment of the reproducibility of these 3D reconstruction methods has been addressed [164].

Here, our main objective is to address the lack of such analysis by assessing the limits of 3D reconstruction based on mean-field restraint-based modeling. Although our analysis is based solely on models generated by TADbit, the conclusions are likely to hold for alternative mean-field restraint-based approaches. Over the next sections of the manuscript, we detail the methods for simulating "toy genome" structures, deriving interaction matrices from them, reconstructing their 3D structure, assessing their quality and evaluating their accuracy using the MMP score (Materials and Methods). Next, we describe the results of assessing the predictive power for determining the "real" assembly structure of "toy genome" structures as well as *a priori* evaluate the input interaction matrices modeling potential (Results). Finally, we summarize our conclusions on the limits of mean-field restraint-based approaches and how a measure such the Matrix Modeling Potential (MMP) can be used to *a priori* evaluate the reconstructed models (Discussion).

2.3 Material and Methods

a) Overall pipeline

With the aim of assessing the accuracy of restraint-based modeling of genomes and genomic domains by TADbit [152, 240] we devised a computational pipeline consisting of the following three steps (Fig. 1A). First, using polymer modeling we simulated six artificially generated genomes (here called "toy genomes") of a single chromosome with different architectures, from which we extracted 168 simulated interaction matrices with increasing noise levels and structural diversity. Second, we reconstructed with TADbit three-dimensional (3D) models of the toy genomes based on their simulated "Hi-C" interaction matrices. And third, we analyzed the reconstructed models for each simulation to assess their structural similarity to the original simulated toy genomes.

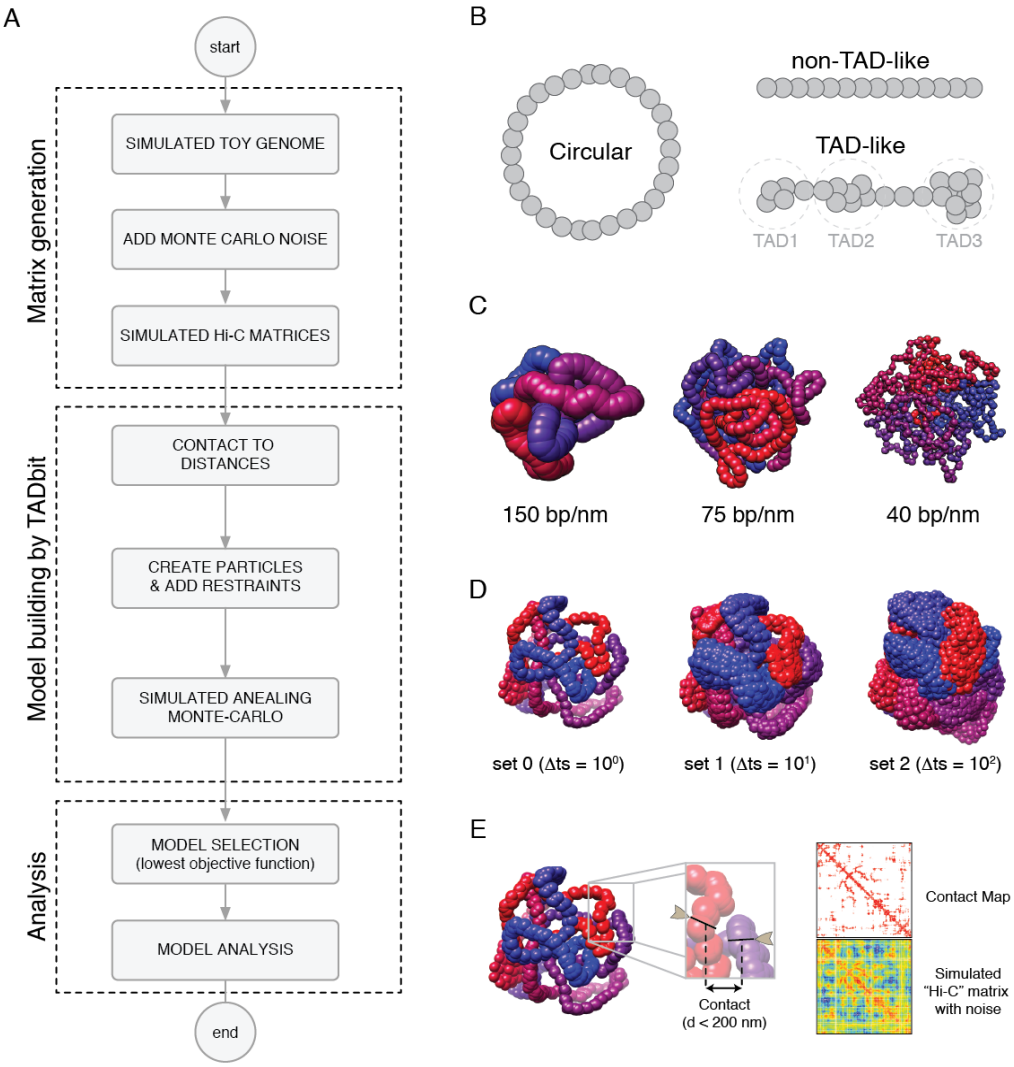

Figure 1. Matrix generation and model building. (A) Flowchart from toy genome generation to reconstructed 3D models. (B) Types of simulated genomic architectures. (C) Genomic density of simulated genomes. (D) Structural variability depending on the selection of conformations between distant times steps in the simulated genomes. (E) Derivation of interaction matrices from toy genome structures based on simulated restriction sites (grey arrows) and distance cut-off. Noise was added by a Monte-Carlo procedure with a probability proportional to the distance between the simulated restriction sites.

b) Matrix generation from toy genome architectures

The toy genomes were generated using a worm-like chain (WLC) model, which provides a coarse-grained description of protein-coated DNA (*e.g.*, the eukaryotic chromatin). At the "microscopic level", a WLC is characterized by three parameters: the diameter (nm), the persistence length (nm) and the linear density (bp/nm), which respectively account for the physical thickness, the stiffness and the level of DNA compaction of the chain. Here, we considered a "chromatin fiber" structure with a

diameter of 30 nm and a persistence length of 100 nm, and investigated three densities: 40, 75 and 150 bp/nm. The toy genomes consisted of a single circular chromosome of approximately 1Mb long (Fig. 1B) with a circular architecture to prevent the formation of knots during the WLC simulation. For half of the simulations, we forced into the toy-genomes the formation of a Topologically Associated Domain (TAD)-like architecture by defining a limited number of locally interacting regions in the chromosome. To this end, we added a harmonic potential between all pairs of loci within the region considered as TAD so that they were constrained to remain close-by in space [241]. Altogether, considering the combination of the three linear densities and the architectural properties (TAD or non-TAD), we investigated six types of genome architectures. Using a Monte-Carlo algorithm [242], we then simulated the equilibrium folding of these chromosomes in a cube of side 400 nm, which leads to the typical DNA density that is found in eukaryotic nucleus (0.015 $bp*nm^{-3}$).

Each of the six simulations generated many successive conformations of the chromosomes, whose likelihood is dictated by thermodynamic laws [243]. Using the outcome of these simulations, we generated simulated Hi-C matrices as explained below. To this end, each spatial conformation of the toy genome was segmented into *N* spherical bins of equal lengths, which determined the resolution of the Hi-C matrix. Given the ~1Mb length of our simulated chromosomes, we respectively considered bins of length 1.6 kb (626 bins), 2.5 kb (402 bins) and 5 Kb (202 bins) for the bp densities 40, 75 and 150 bp/nm, respectively (Fig. 1C).

To assess the impact of cell-to-cell variability on our reconstruction method [158], we examined the effect of increasing the level of structural variability by selecting conformations of the toy genomes at different times of the simulations. For each of the six simulations (corresponding to the six chromosome architectures), we created a total of seven sets of 100 models, each differing in the number of simulation steps that separated them (Δt) from 1 to 1,000,000 steps. The corresponding sets of toy genomes were named set 0 to 6. The larger the Δt between two selected models, the larger their structural variability (Fig. 1D).

Finally, for each set of toy genome structures we derived an interaction matrix to obtain a "simulated Hi-C matrix" by computationally mimicking the published Hi-C protocol [139] (Fig. 1E). We set a restriction enzyme cutting frequency and defined all restriction site positions that would be tested for interactions (*i.e.*, contact in the models). We

considered about 2,000 restriction sites over the 1Mb toy genome, which resulted in an average cutting frequency of 500 bp. We selected this frequency to consider it a middle range value of the restriction site frequencies used in the Hi-C experiments [134]. Restriction enzymes recognizing a 6-base pair sequence (*e.g.*, HindIII), have an approximate cutting frequency of 4 Kb in the Human genome, while restriction enzymes recognizing a 4-base pair sequence cut on average every 256 base pairs. The genomic position of each restriction site was determined randomly, maintaining the defined cutting frequency of 500 bp per genome. Once the restriction sites were assigned, we interpolated its 3D coordinates in the simulated toy genomes to obtain Euclidian distances between all the restriction sites. Next, we applied a 200 nm distance cut-off to generate a contact map between all the restriction sites in a set of structures (Fig. 1E); this cut-off can be viewed as a maximum size of protein macro-complexes that can lead to Hi-C interactions through formaldehyde cross-linking. In addition, since several steps of the Hi-C protocol may affect the detection of interacting fragments (*e.g.*, inefficient formaldehyde cross-linking or inefficient digestion and/or re-ligation) [134], we simulated the experimental noise by selecting pairwise interactions with a probability defined by a Gaussian procedure with an α value varying from 50 to 200 in steps of 50. The α parameter is related to the decay of the Gaussian function between the probability of interactions and the Euclidian distance between the restriction sites. A large α of 200 will increase the total probability of interactions, while a smaller α of 50 will decrease the total probability of interactions. The selection of the Gaussian procedure allowed for a large dynamical range of maps across the tested structural variability. The resulting interaction matrices, that is, our "simulated Hi-C matrices", thus contain a varying proportion of noise compared to a direct contact map generated from the models (Fig. 1E). Finally, the total number of interactions between restriction sites was then pooled into bins according to the linear density of the genome (see above). The simulated Hi-C matrices contain thus a varying degree of experimental noise (α from 50 to 200), which are then complemented by an increasing degree of structural variability (sets 0 to 6) representing cell-to-cell variability in a population of millions of cells of a typical Hi-C experiment (Fig. 2).

Before building models using TADbit, the input matrices were normalized by first calculating the weight ($W_{i,j}$) for each pair of interactions:

$$w_{i,j} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} M_{i,j}}{\sum_{i=1}^{N} M_{i,j} \times \sum_{j=1}^{N} M_{i,j}}, i,j \in 1..N$$

where $M_{i,j}$ is the raw counts in the simulated interaction matrix between bins *i* and *j*. The normalized matrix resulted from the multiplication of $M_{i,j}$ by its weight $W_{i,j}$, which corresponds to a single iteration of the ICE normalization procedure [244]. Next, a decimal logarithm transformation was applied to the normalized interactions and its $Zscore_{i,j}$ was computed for non-zero interaction cells in the matrix as:

$$Zscore_{i,j} = \frac{log_{10}(M_{i,j} \times w_{i,j}) - \mu}{\sigma}$$

where the average μ and the standard deviation σ from the entire matrix were obtained as:

$$\mu = log_{10}\left(\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} M_{i,j} \times w_{i,j}}{N \times N}\right) \text{ and } \sigma = \sqrt[2]{\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}((M_{i,j} \times w_{i,j})^2 - \mu)}{N}}$$

The resulting Z-scored matrices were used as input for modeling with TADbit.

c) Model building by TADbit

To build the 3D models of the genomes, we used the TADbit python library developed around the Integrative Modeling Platform (IMP), which involves the translation of the data into particles; the assignment of spatial restraints between them and the search for optimal solutions maximizing the satisfaction of the imposed restraints. Next, we describe the used of our modeling protocol, which has been previously detailed [240].

Briefly, 3D models in TADbit are defined by N particles determined by the resolution of the input interaction matrix. Each particle has an excluded volume defined as a sphere with a radius proportional to the number of base-pairs in each particle. Here, we consider an inverse relationship between spatial distances and the corresponding frequencies of interactions. Given this assumption, TADbit transforms the frequencies of interactions into spatial restraints differently for consecutive and non-consecutive particles. Two consecutive particles are spatially restrained (that is, kept at an equilibrium distance) according to their occupancy, which corresponds to the sum of their radii. Non-consecutive particles are restrained based on empirically identified parameters that define a set of restraints, their distances, and the forces applied to them. TADbit empirically identifies three optimal parameters using a grid-search where a limited number of models are built for each set of parameters. The three parameters are: the proximal distance between two non-interacting particles, a lower-bound cut-off to

define particles that do not interact frequently and an upper-bound cut-off defining particles that do interact frequently. The resulting models for each combination of parameters are then used to calculate a contact map to compare it to the input interaction matrix by calculating the Spearman correlation coefficient between the two matrices (here called IMPSCC). Thus, similarly to many restraint-based methods for 3D genome reconstruction, TADbit sampling aims at identifying a set of models that maximizes the similarity between the models contact map and the Hi-C interaction matrix. Once the optimal parameters are identified, restraints are applied to the particles. Pairs of particles with contact frequencies above the upper-bound threshold are restrained to be at a given equilibrium distance. Pairs of particles with contact frequencies below the lower-bound threshold are maintained further than an equilibrium distance. Finally, TADbit uses a Monte Carlo simulated annealing sampling procedure to identify a set of 3D models that best satisfy the imposed restraints.

d) Model accuracy

We assessed the structure similarity between the original toy genome architecture sets and the reconstructed models by computing two different measures. First, the distance Root Mean Square Deviation (dRMSD) between the best-reconstructed model and each of the 100 original selected structures was calculated after optimal superimposition of their structures by:

$$dRMSD = \sqrt{\sum_i \sum_j \left(O_{i,j} - R_{i,j}\right)^2}$$

where O_{ij} and R_{ij} are the distance vectors between particle i and j in the original structure and in the reconstructed model, respectively. The dRMSD is a measure that varies between 0, when the two structures are identical, and a large number, proportional to the size of the object measured, when the two structures are completely different. The maximum dRMSD depends on the size of the object and the number of particles compared. Therefore, the reconstructed models were scaled to have the same dimensions in the three axes as the toy structures before structural superimposing them. The scale factor was calculated as the average ratio between the maximum distances in x-, y- and z-axis of the reconstructed models and the toy structures. Second, a Spearman correlation coefficient (dSCC) between all pairwise distances of particles in the best-

reconstructed model and the corresponding ones in each of the 100 original toy structures was calculated. The dSCC measure varies between -1.0 and 1.0 for comparisons where the distances perfectly anti-correlate or correlate, respectively. Therefore, a model with a dSCC of 1.0 indicates good accuracy regardless of the scale of the compared structure.

e) Matrix Modeling Potential

With the aim of identifying *a priori* whether an interaction matrix has the potential of being use for modeling, we calculated from each of the 168 simulated Hi-C matrices three different measures: (i) the contribution of the significant eigenvectors from the matrix, (ii) the skewness and (iii) the kurtosis of the distribution of Z-scores in the matrix.

The contribution of the significant eigenvectors (SEV) score was obtained by first calculating the eigenvectors of the interaction matrix and the percentage of contribution of their corresponding eigenvalues. Next, we randomized 100 times the interaction matrix by shuffling the cells in the matrix that are equidistant from the diagonal. This shuffling strategy preserved the expected exponential decay of interactions as we go from the diagonal to the anti-diagonal corners of the matrix. From the 100 randomized matrices, we also calculated their eigenvectors and the percentage of contribution of their corresponding eigenvalues. We then set as "significant eigenvector" those with eigenvalues above the mean eigenvalue plus two standard deviations of the equivalent eigenvectors in the random set of matrices. The final SEV score was the sum of the differences of the contribution of eigenvalues of all significant eigenvectors:

$$SEV = \sum_{i} ev_i - \overline{rev}_i$$

where ev_i corresponds to the contribution of the eigenvalue of the significant eigenvector i in the interaction matrix and $\overline{\text{rev}}_\text{i}$ is the average contribution of the eigenvalue of the same eigenvector in the randomized 100 interaction matrices. Overall, large *SEV* scores are indicative of good potential for modeling. Intuitively, they indicate the presence of specific contacts that are not just the results of a random conformation of the chromosome.

The other two descriptive statistics were calculated directly from the distribution of Z-scores in the Hi-C matrices. First, the skewness statistic (SK) assesses in a single measure whether a score is skewed towards the right or left tails of its distribution. The

kurtosis statistic (KT) complements the interpretation of the skewness. For example, matrices with skewness close to zero may result from multi-modal distributions of Z-scores. In such cases, the distribution will result in large KT scores. Therefore, the SK score will indicate skewness of the matrix towards positive or negative Z-scores and the KT score will indicate whether a matrix results or not in single-peaked distribution of Z-scores. For optimal modeling in TADbit, we expect no skewness and a single peak in the Z-score distribution. Both the skewness and the kurtosis statistic where estimated using the SciPy python library (http://www.scipy.org). The SK and KT are calculated as:

$$SK = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^{2^{3/2}}} \quad \text{and} \quad KT = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{\sum_{i=1}^{N}(x_i - \bar{x})^{2^2}}$$

where N is the number of bins in the Z-score distribution and x_i corresponds to the frequency of a given bin i.

Finally, to calculate the Matrix Modeling Potential (MMP) score, we used the size (number of bins in the matrix), SEV, SK and KT for all 168 simulated Hi-C matrices as input to train a classifier with a linear regression kernel using Weka [245]. During the training of the classifier, we used the actual accuracy of the produced 3D models (that is, the dSCC measure) as a target goal. We decided to use the dSCC measure instead of the dRMSD accuracy measure because it is independent of the scale and size of the objects to compare. The classifier, thus, aims at identifying a linear combination of the four matrix measures to produce a final score that best correlates with the dSCC of the models. We trained the classifier with a 10-fold cross-validation procedure, which resulted in a correlation coefficient of 0.84 between the MMP score and the dSCC measure. The MMP score is calculated as:

$$MMP = -0.0002 * Size + 0.0335 * SK - 0.0229 * KU + 0.0069 * SEV + 0.8126$$

2.4 Results

a) Toy genome structures and derived matrices

We investigated the reconstruction efficiency of six types of toy genomes hereafter labeled by ch40, ch75, ch150, ch40_TAD, ch75_TAD and ch150_TAD depending on the bp density along the chromosome and on the presence, or not, of TAD-like organization. To this end, for each toy genome, we generated seven sets of 100 different conformations, corresponding to seven different structural variability levels. More precisely, the n^{th} set was generated by extracting 100 conformations separated by a time step of ($\Delta t = 10^n$) iterations in the corresponding worm-like chain simulation (Fig. 2). Altogether, for each toy genome we generated 700 different chromosome conformations that were distributed among seven different sets, with set 0 having the lowest structural variability ($\Delta t = 1$) and set 6 the highest ($\Delta t = 10^6$). Such structural sets were then used to derive four contact maps with varying levels of experimental noise (that is, with α = 50, 100, 150 and 200), which simulate the results of a hypothetical Hi-C experiment. Finally, the contact maps were input to TADbit to build 3D models using a previously implemented protocol [152]. The initial structural sets for the 6 tested toy genome architectures, their derived interaction matrices and the reconstructed 3D models are available at http://www.3DGenomes.org/datasets. Specific details on the construction of the toy genomes and the derived models are given in the Materials and Methods.

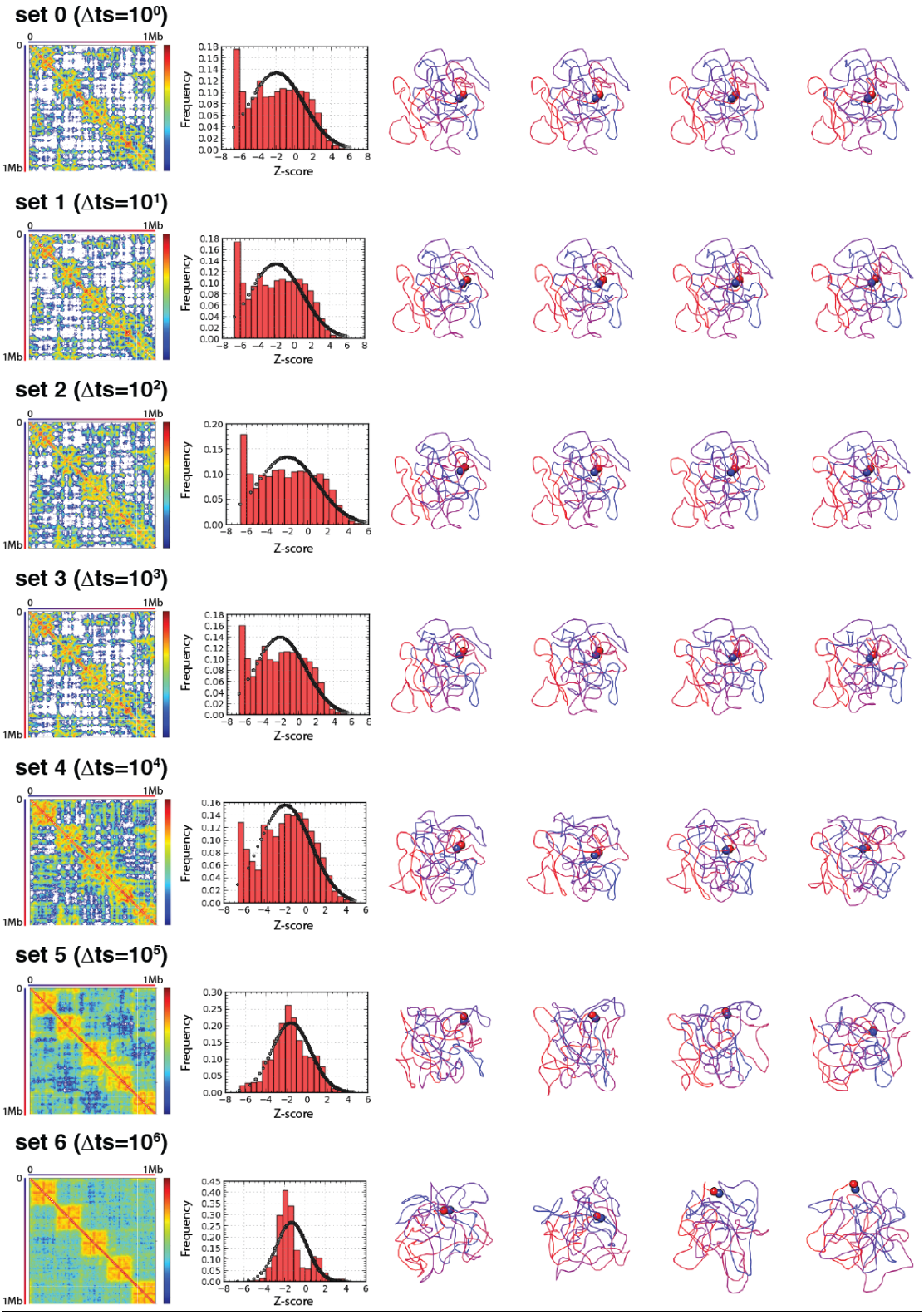

Figure 2. Simulated Hi-C interaction matrices. Simulated Hi-C interaction matrices for the toy genome architecture of chr75_TAD with noise levels α=50. Each row shows the calculated matrix, the distribution of Z-scores and four randomly selected input structures, which are colored from particle 1 in blue to particle *N* in red, the start and end particles are highlighted with spheres. From top to bottom the figure depicts the simulated matrices from sets 0 to 6 ($\Delta t = 1$ to $\Delta t = 1{,}000{,}000$).

b) Overall accuracy of the generated models

To assess the accuracy of the genomic 3D models built by TADbit, we calculated two different accuracy measures between the reconstructed models and the toy genomic structures (that is, the distance Root Mean Square Deviation (dRMSD) and the distance Spearman Correlation Coefficient (dSCC)). Both measures of accuracy were calculated for all reconstructed models and averaged over architecturally similar toy genomes (Table 1). In total, we generated 168 simulated Hi-C matrices for the 6 toy genome architectures (that is, 6 architectures with 7 levels of structural variability and each with 4 levels of noise in the data). The reconstructed architecture that best fitted the input structures corresponded to the 40 bp/nm density with a TAD-like architecture (chr40_TAD), with an average dRMSD of 60.5 nm and dSCC of 0.79. The architecture most difficult to reconstruct corresponded to 150 bp/nm density with no TAD-like features (chr150), with an average dRMSD of 86.4 nm and dSCC of 0.51. These values correspond to average measures over the 28 simulated Hi-C matrices per architecture, which include varying degrees of noise and structural variability. For example, within the chr40_TAD architecture, one of the best reconstructions corresponded to the matrix with mid noise level ($\alpha = 100$), and low structural variability ($\Delta t = 10$), which resulted in a 3D model with dRMSD of 32.7 nm and dSCC of 0.94 (Fig. 3A top). Similarly, for the low-resolution architecture 150T, the best result (dRMSD = 45.4 nm and dSCC = 0.86) corresponded to low level of noise ($\alpha = 50$) and low structural variability ($\Delta t = 1$) (Fig. 3A bottom). In summary, TADbit was able to produce accurate models for all six toy genome architectures with a varying degree of accuracy depending on the levels of noise and structural variability in the simulated Hi-C matrices.

Name	Dens. (bp/nm)	TAD	Size	<Restraints per particle>	<Spearman CC>	<dRMSD>	<dSCC>
Chr40	40	No	626	104.4	0.84	71.12	0.78
Chr40_TAD	40	Yes	626	113.7	0.86	60.49	0.79
Chr75	75	No	402	91.8	0.84	82.14	0.69
Chr75_TAD	75	Yes	402	79.9	0.86	68.56	0.74
Chr150	150	No	202	46.3	0.82	86.42	0.51
Chr150_TAD	150	Yes	202	53.5	0.86	72.63	0.64

Table 1. Toy genome architectures and overall reconstruction accuracy.

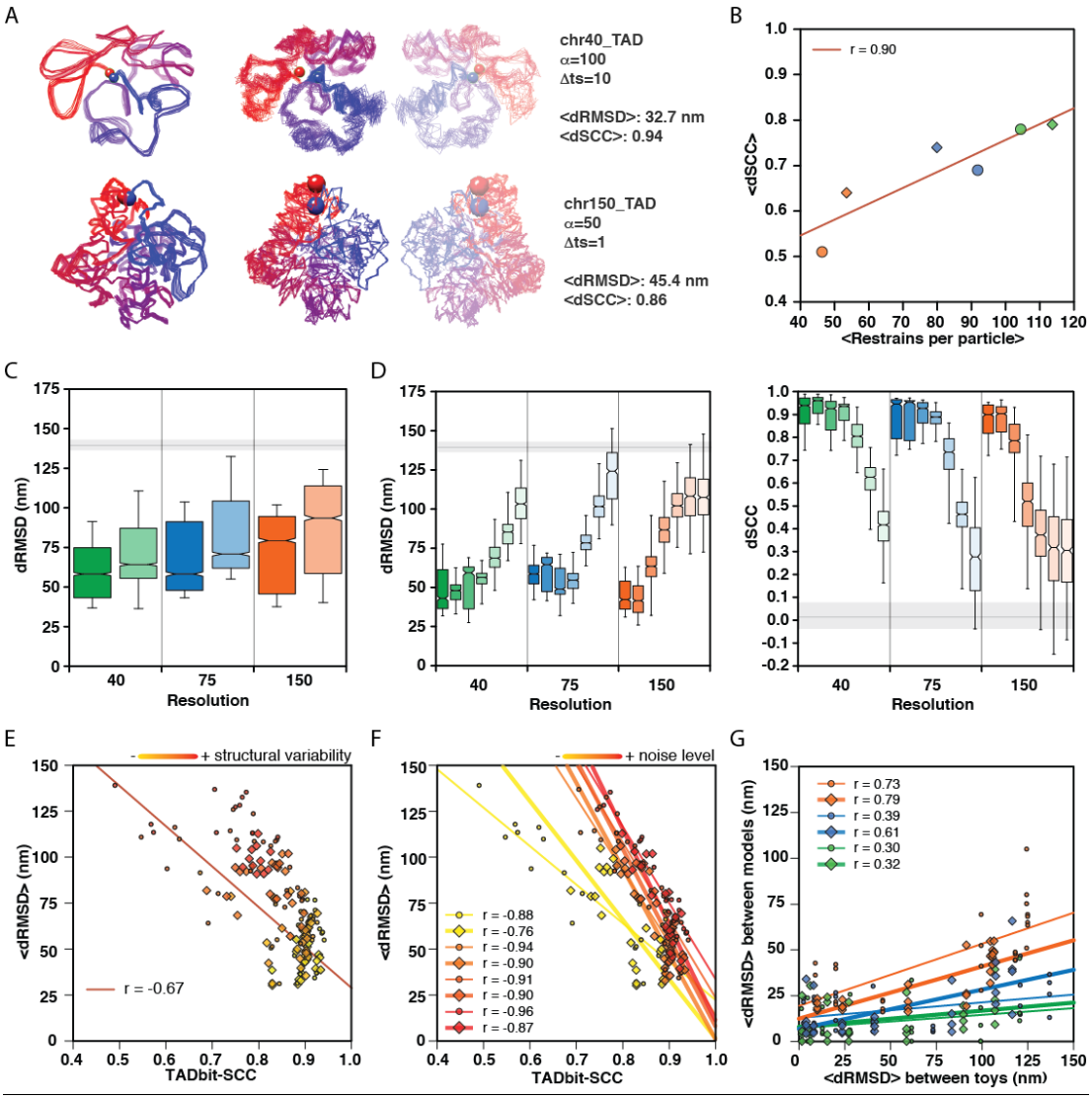

Figure 3. Model assessment. (**A**) Comparison of a 3D model ensemble of genome architectures for the chr40_TAD (top) and chr150_TAD (bottom) architectures. Superimposed input structures for set 0 (left models) and superimposed reconstructed 3D models (due to mirroring, TADbit generates right- and left-handed models [152]). Models are colored from particle 1 in blue to particle *N* in red, the start and end particles are highlighted with spheres. (**B**) Correlation between the restraints per particle and the accuracy of the reconstructed models as measured by the average dSCC score per architecture. Circle symbols correspond to non-TAD-like architectures. Rhomboid symbols correspond to TAD-like architecture. The colors indicate the toy genome density (green, blue and orange for 40, 75 and 150 bp/nm, respectively). (**C**) dRMSD distributions with respect to genome architecture. Colors correspond to the three density values with dark and pale colors corresponding to TAD-like and non-TAD-like architectures, respectively. Horizontal grey line and shade corresponds to the dRMSD distributing of comparing a "random genome" of the same size and number of particles as the reconstructed models but with randomized coordinates. (**D**) Model accuracy as measured by dRMSD (left) and dSCC (right) with respect to the model density. Each density is colored as in panel A and contains 7 distributions from the 7 sets of structures from set 0 ($\Delta t = 1$) to high structural variability set 6 ($\Delta t = 10^6$) with dark to pale colors, respectively. Horizontal grey lines and shade as in panel C. (**E**) Correlation between the dRMSD values per reconstructed models and the Spearman correlation coefficient of the contact map from the reconstructed models and the original toy-genome structures (TADbit- SCC). The points are colored proportional to the level of structural variability in the matrix (yellow to red from low set 0 ($\Delta t = 1$) to high structural variability set 6 ($\Delta t = 10^6$)). Shapes represented as in panel B. (**F**) Same as panel E but now the points are colored by the level of noise in the data (yellow to red for low to high levels of noise, that is from $\alpha = 50$ to 200). The regression coefficients indicate the correlation per noise level α. (**G**)

Correlation between structural variability in the toy genome structures and in the reconstructed models. Colors and shapes as in panel B.

c) Genome architecture and model accuracy

We tested two features of the toy genome architecture: its density (or resolution) and the presence or absence of local compact regions representing TADs. Models based on higher-resolution matrices resulted in a higher number of imposed restraints per particle in the reconstructed 3D models (Table 1). As expected, we observed a linear relationship between the number of restraints per particle imposed during modeling and the dSCC value (r=0.9, Fig. 3B), which in turn depends on the resolution of the input matrices determined by the density of the toy genomes. Despite the relative low accuracy of models for high-density genomes (i.e., low-resolution genomes), TADbit was able to generate topologies very similar to the input structures (Fig. 3A). Altogether, these results indicate that the choice of genomic density and, with it, the resolution representing the genome alter the accuracy of the reconstructed models. The existence of a TAD-like organization in the genome had also an effect on the accuracy of the reconstructed models. All simulated matrices with genome architectures at 40 bp/nm density with TAD-like architecture resulted in an average dRMSD of 60.5 nm while genome architectures with no TADs resulted in an average dRMSD of 71.1 nm (Table 1). This trend was observed for all resolutions where the TAD-like architecture resulted in lower average dRMSDs (t-test p-value <0.001, Fig. 3C). Overall, both high resolution simulated matrices and the existence of a TAD-like structures in the toy genomes resulted in more accurate reconstructed 3D models.

d) The accuracy of the models is sensitive to structural variability but robust to noise

3C-based experiments are performed on tens of millions of cells and thus are a population-based interrogation of the genome. It is therefore likely that the interrogated cell population harbors structurally different conformations of their genome, due to the unsynchronized cell cycle or to natural cell-to-cell variability, among many other factors. To simulate such situation, we increased the structural variability in the input matrices by selecting structures from the architectural genomes at different simulation time steps (Materials and Methods). Simulated Hi-C matrices with increasing variability provided less detail of local chromosome structuring but captured the large-scale organization of the toy genomes such as the existence of TADs (Fig. 2). As expected for any mean-field reconstruction method, the accuracy of our reconstructed genomes decreased with the increase in the input structural variability (Fig. 3D and E). For all toy genomes with different architectures, the accuracy of the models was maintained up to the structural variability set 3 ($\Delta t = 1000$). The models resulting from the sparse matrices based on the structural sets 4 to 6 ($\Delta t \gg 10{,}000$) had significantly higher dRMSD values as compared to the other models. Indeed, model reconstruction based on low-resolution matrices (150 bp/nm genomes) and large structural variability, resulted in models with poor accuracy (dRMSD > 90 nm). At the highest levels of structural variability (*i.e.*, set 4 to 6 or $\Delta t \gg 10{,}000$), the interaction matrices were predominantly populated in the proximity of the diagonal, or the TAD structures, as the only common interacting regions between the different input structures for both the non-TAD-like and TAD-like architectures, respectively (see for example Fig. 2 bottom rows). Interestingly, the reconstruction of 3D models with TADbit was robust to noise (Fig. 3F). In fact, the accuracy of the models was constant to mid levels of noise in the data (average dRMSD of 70.7, 71.5, 74.3 and 78.7 for α values of 50, 100, 150 and 200, respectively). Nevertheless, the correlation between the TADbit-SCC and the dRMSD was higher at the mid level of noise compared to the low level of noise (0.77 for $\alpha = 50$ and 0.87 for $\alpha >= 150$). In summary, the reconstruction of 3D models based on noisy data is robust but mean-field methods are sensitive to structural variability in the simulated Hi-C interaction matrices.

e) The TADbit-SCC is an accurate scoring function for modeling

TADbit model building depends on the imposed restraints for modeling, which in turn are determined by three optimized parameters. The three cut-offs are determined by maximizing the Spearman correlation coefficient between a contact map calculated from the reconstructed 3D models and the input simulated Hi-C matrix (here called TADbit-SCC). To test whether the TADbit-SCC measure is a good proxy for model accuracy, we compared it the dRMSD of the resulting reconstructed genomes. The results clearly indicate that the use of the TADbit-SCC as a scoring function to identify the best models is reasonable (Fig. 3E) as high values of TADbit-SCC are indicative of low dRMSD (r=-0.67). However, the relationship is not perfect and has two main properties that affect its adequacy for identifying good models: (i) a range of low dRMSD values may result in very similar TADbit-SCC, and (ii) the dRMSD value saturates for low TADbit-SCC values. Altogether, the analysis indicates that the use of Spearman correlation coefficient (TADbit-SCC) as a scoring function during modeling is a good proxy for model accuracy but needs to be complemented by additional measures (see below).

f) Reconstructed models capture part of the structural variability in the matrices

Mean-field restraint-based modeling methods assume that the interaction matrix reflects an average structure of the genome with a limited number of different conformations. Thus, such methods have intrinsic difficulties in capturing the variability of the data. To test whether our reconstructed models reflect the structural variability in the matrices, we calculated the dRMSD between the 100 input toy genome structures in each of the 168 matrices. We also calculated the dRMSD between 100 generated models per simulated matrix. In all the genomic architectures, we observed a correlation between the variability in the toy genome structures and the resulting variability in the reconstructed models (Fig. 3G). The captured variability decreased with the increased number of restraints per particle (Fig. 3B). That is, higher-resolution matrices that resulted in more restrained models, have less structural variability in the output structures. Importantly, the degree of variability is about two fold less in the resulting

models compared to the input toy structures. Nevertheless, and despite the intrinsic limitations, the resulting models capture part of the structural variability in the matrices.

g) Statistics of the input matrices correlate with the accuracy of the models

To assess which features from the interactions matrices could be useful to predict the accuracy of the reconstructed models, we have calculated three statistical measures from the simulated Hi-C matrices (Materials and Methods). In particular we measure the contribution of the significant eigenvectors from the matrix (SEV), the skewness (SK) and the kurtosis (KT) of the distribution of Z-scores. These three measures are indicative of the internal correlations in the matrix (SEV) and the deviation from normality of the distribution of interaction counts (SK and KT). These features are relevant for the modeling with the TADbit protocol since they determine the quantity and quality of the imposed restraints during modeling [152]. In principle, an input matrix with high contribution of the significant eigenvectors, skewness close to zero and low negative kurtosis is optimal for 3D reconstruction. For example, the toy genome architecture chr40_TAD, which results in accurate 3D reconstructed models (dRMSD = 47.2 nm and dSCC = 0.91), has a SEV of 32.3%, a SK of -0.32 and a KT of -0.69 (Fig. 4A). Indeed, the three statistical measures from the simulated Hi-C matrices correlate with the final accuracy of the reconstructed models (Fig. 4B). dRMSD correlates with SEV, SK and KT with a -0.53, 0.63 and 0.75 regression coefficient, respectively. dSCC correlates with SEV, SK and KT with an 0.70, -0.60 and -0.54 regression coefficient, respectively. Moreover, we observed that the SK statistic, which measures whether a matrix has a Z-score distribution skewed towards positive or negative values, could be used to discern between matrices with high structural variability from those with high experimental noise (Fig. 4C). All but one of the simulated Hi-C matrices with large noise content ($\alpha = 200$) and low structural variability (set 0) result in negative values of SK score. Similarly, all but two of the simulated Hi-C matrices with low noise content ($\alpha = 50$) and high structural variability (set 7) result in positive values of SK score. In summary, we introduced here three simple statistics from the Hi-C matrices that can help us assess the likeliness of an interaction matrix to result in accurate reconstructed models.

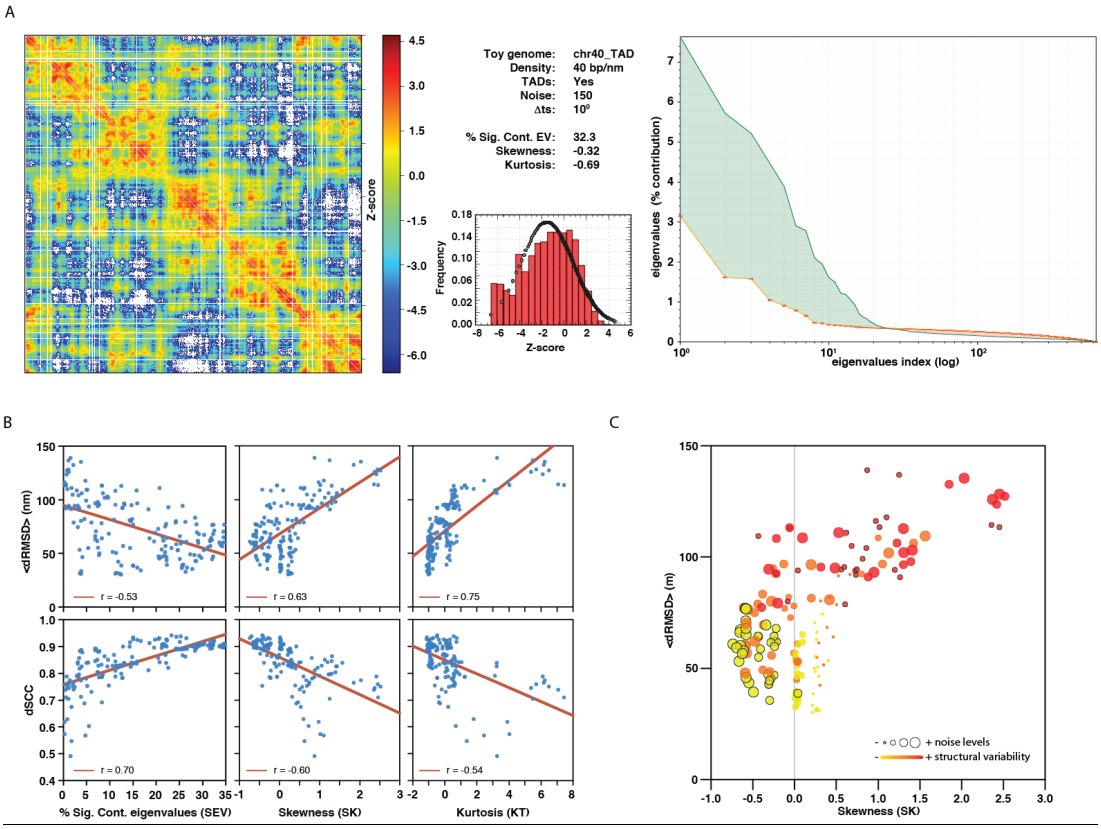

Figure 4. Matrix entropy. (A) Statistical measures for interaction matrix for chr40_TAD architecture. Simulated Hi-C matrix (left), statistical measures and Z-score distribution (middle) and eigenvalues plot (right). The eigenvalues plot shows the real distribution of values (green solid line) and the distribution for random matrices (orange solid line). The green area corresponds to the contribution of the significant eigenvalues (i.e., the SEV score). (B) Correlation of the statistical matrix properties (SEV, SK and KT) with the reconstructed model accuracy measures (dRMSD and dSCC). (C) Correlation plot between Skewness and dRMSD for the 168 simulated Hi-C matrices. Dots are colored with respect to the structural variability in the matrix (yellow to red for set 0 to set 6). The size of the dots is proportional to the level of noise in the matrix (small for $\alpha = 50$ and large for $\alpha = 200$). Highlighted red small dots indicate low noise and high structural variability matrices. Highlighted yellow large dots indicate high noise and low structural variability matrices.

h) The Matrix Modeling Potential (MMP) score

To assess whether we could *a priori* evaluate the adequacy of the input matrix for 3D reconstruction, we calculated a single score, here called Matrix Modeling Potential (MMP), combining four measures from the interaction matrices: its size, SEV, SK and KT values. We trained a linear regression with the four measures for the 168 simulated Hi-C matrices to obtain a single score that correlates the most with the dSCC accuracy measure of the 168 reconstructed models. The training set contains, thus, a variety of resolutions, experimental noise and structural variability. Using a 10-fold cross validation the MMP score resulted in a final correlation with the dSCC of the reconstructed models of 0.84 (Fig. 5A). The mean absolute error of the MMP score in predicting the dSCC accuracy of the models is 3.1%, which provides a clear predictive

power to the new score. Indeed, the MMP score behaves as expected (Fig. 5B). Simulated matrices built from toy genomes with TAD-like structure results in higher MMP score, which also increases with a slight presence of noise in the matrix and is clearly affected by the increase of structural variability. In summary, combining the three statistical scores from the simulated Hi-C matrices as well as its size into a single MMP score, provides a means to *a priory* evaluate the modeling potential of the matrix. Matrices with high MMP scores are likely to result in accurate 3D reconstructed models.

To test the applicability of our new score, we selected three datasets of real Hi-C experimental data from human [239], fly [246], and bacterial [113] genomes and calculated their MMP score (Fig. 5C). Of the three example matrices, the human genomic domain results in a MMP score of 0.82, which predicts a dSCC of 0.81 (0.69-0.92 at 95% confidence range). The best individual score for the human genomic domain is the skewness of the distribution, which approximates zero (SK = 0.20). However, the contribution of the significant eigenvalues is small (SEV = 3.61). Similarly, the *Caulobacter crescentus* genome matrix has good SK and KT values but poor SEV (0.26, -0.25 and 3.05, respectively). The resulting MMP score is 0.77, which predicts a dSCC of 0.73 (0.62-0.85 at 95% confidence range). Finally, the fly genomic domain is the one with the best MMP score (0.83) of the three real Hi-C matrices, which resulted in a predicted dSCC of 0.83 (0.72-0.94 at 95% confidence range). This result shows that, at different levels of predicted accuracy, real Hi-C matrices could be used in TADbit for 3D reconstruction of genomes and genomic domains.

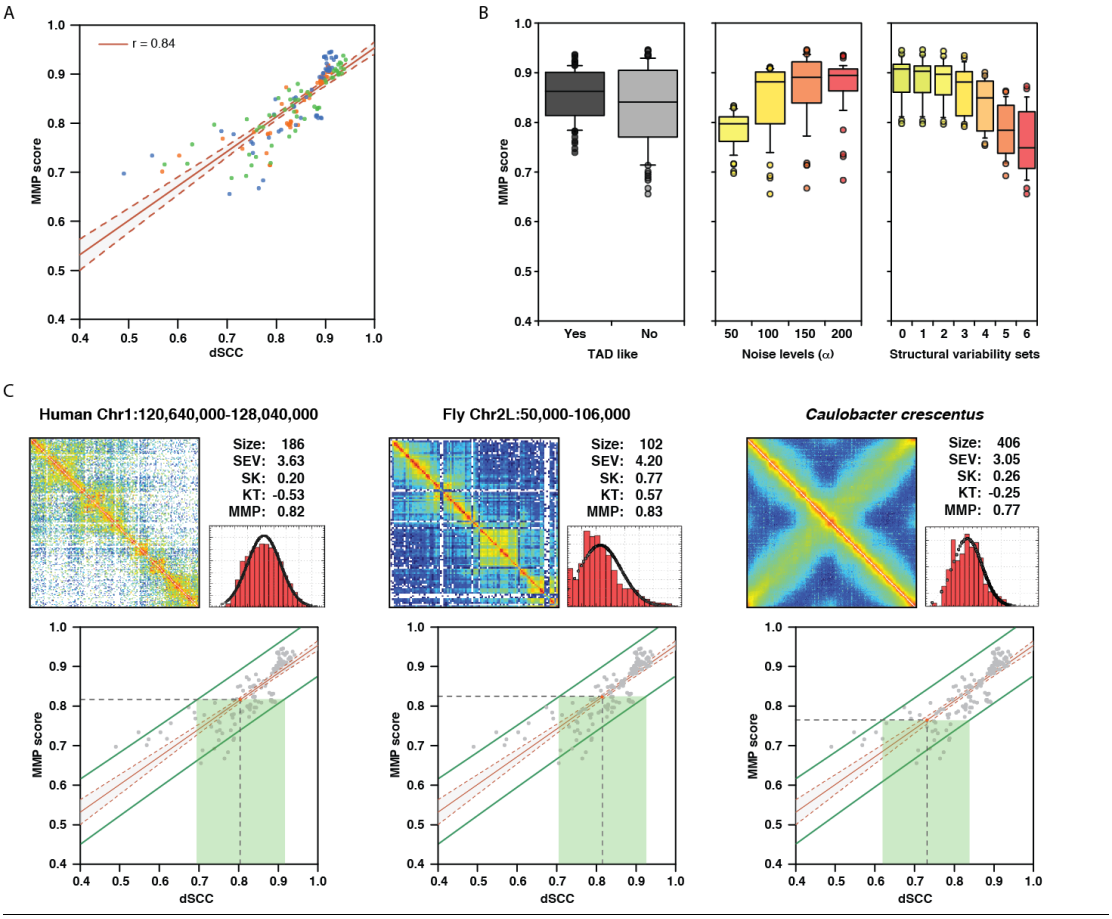

Figure 5. Predicting the accuracy of the reconstructed models. (A) Correlation between the MMP score and the dSCC accuracy measure. Points are colored by the density of the simulated Hi-C matrices (green 40, blue 75 and orange 150 bp/nm) Shaded area corresponds to the correlation confidence band. (B) MMP score distributions depending on genome architecture, noise and structural variability of the simulated Hi-C matrices. Panels from left to right show existence of TAD-like architecture, noise levels (yellow to red for α from 50 to 200), and structural variability (yellow to red from sets 0 to 6). (C) Example of MMP score and the predicted dSCC of the resulting models for genomic domains or entire genomes in real Hi-C matrices. For each panel we show the actual Hi-C matrix in Z-score scale (red to blue from positive to negative Z-scores), the four input statistics as well as the MMP score, the Z-score distribution shape and the predicted range of dSCC (green bar at 95% confidence level). Left panel for a genomic domain in chromosome 1 of the human genome [239], middle panel for a genomic domain in chromosome 2L of the fly genome [246], and right panel for the entire Caulobacter crescentus genome [113].

2.5 Discussion

Recently, chromatin interaction matrices from 3C-based experiments have been used for modeling the three-dimensional organization of genomes and genomic domains [151]. Those approaches aim at providing a 3D representation of the bi-dimensional interaction matrices that can be explored to extract biological insights. Here, we have introduced a comprehensive analysis of the limitations of chromatin model building using a restraint-based mean-field approach. To do so, we have derived a series of simulated Hi-C matrices where the genomic architectures are pre-defined and the amounts of noise and structural variability is controlled. The entire set of 168 simulated Hi-C matrices can be considered as a benchmark set for assessing the future developments of restraint-based methods for modeling genomes and genomic domains. To our knowledge, this is the first fully available dataset for benchmarking reconstruction methods, which can be freely accessed here: http://www.3DGenomes.org/datasets.

In our analysis, a total of six different genomic architectures were benchmarked. Those varied the resolution (or genomic density) as well as the presence of locally compacted regions resembling TADs observed for many organisms [239, 247, 248]. The overall accuracy of the reconstructed models points to three main conclusions. First, independently of the genomic architecture, restraint-based mean-field modeling can provide accurate models with dRMSD as low as 30nm and dSCC as high as 0.99 with the major variability in accuracy originating from the structural variability in the input matrices. Second, an increase of the matrix resolution (that is, low density models with larger proportion of restraints per particle) results in more accurate reconstructed models. Therefore, increasing the sequencing depth of a Hi-C experiment will result not only in higher-resolution models (*i.e.*, more bins in the interaction matrix) but also in models of higher overall accuracy. And third, the presence of a TAD-like architecture results in more accurate models at any levels of noise and structural variability. This increased accuracy can be interpreted as the result of a sharper structuring of the input Hi-C matrix for scales equal or larger than that of TADs, which are expected to be found under the form of compact globules [153]. In vertebrates, these globules are believed to be the result of multiple specific chromatin loops induced by the bridging of several protein complexes, with CTCF as a major factor [141, 249]. Indeed, such

specific loops can be easily integrated in polymer models of chromosomes [250], and should facilitate the inner reconstruction of TADs.

Typically, Hi-C experiments capture a limited number of all possible interactions in each cell [158, 251] and thus are performed on a population of tens of millions of cells. This results in interaction matrices that have two main sources of variability originating from noise in the experiment and/or the natural conformational differences between genomes in each cell. Here we have simulated these two sources of variability by first varying the probability of capturing an interaction from the toy models (experimental noise) and second by deriving simulated interaction matrices from models of varying structural similarity. The results of our test clearly indicate that restraint-based mean-field reconstruction is robust to experimental noise but sensitive to high levels of structural variability. Indeed, at all levels of experimental noise, our method was able to reconstruct accurate models when structural variability was low. However, the reconstruction of models degraded significantly when the level of structural variability was high, indicating that mean-field methods may have difficulties capturing the entire structural diversity of the input matrices. It is important to note that our simulated Hi-C matrices with high levels of structural variability (set 6) contain homogenous structural variability where each of the toy structures can be considered as a "single cell state" that is equally different to all other structures in the set. Despite these limitations, our approach was also able to capture part of the structural variability in the original set. Altogether, our results conclude that Hi-C interaction matrices from as homogenous as possible population of cells (*e.g.*, synchronized in cell cycle, same cell state, unique cell type, etc…) are more adequate for 3D reconstruction. Interestingly, we also show that experimental noise, which could originate from limitations in any of the four main steps in 3C-based methods (that is, cell fixation, DNA fragmentation, DNA ligation and read-out by sequencing), is not highly relevant for 3D reconstruction.

Most of the reconstruction approaches, either those mean-field or population-based approaches, have a scoring function to minimize. The specific scoring function varies between methods but all aim at correlating the observed 3C-based interactions with those obtained from the re-constructing models. In our approach we find optimal parameters for the simulation by maximizing the Spearman correlation coefficient (TADbit-SCC) between the input interaction matrix and a contact map obtain from the models. We have shown here that this scoring function is appropriate for 3D

reconstruction and that high TADbit-SCC result in accurate models, which validates our protocol for 3D reconstruction by TADbit. In practice, with our method the TADbit-SCC can be taken as a proxy of model accuracy. Additionally, we also provide, for the first time, a single measure (the Matrix Modeling Potential or MMP) calculated from the interaction matrix that highly correlates with the accuracy of the resulting models ($r = 0.84$, p-value < 0.001). The MMP score is composed of a weighted sum of four properties of the matrix (that is, its size, the percentage of contribution of significant eigenvalues in the interaction matrix as well as the skewness and kurtosis of the distribution of Z-scores in the interaction matrix). Interestingly, the skewness of the distribution has an additional property that allowed us to differentiate between matrices rich in experimental noise from those high in structural variability. Negative skewness matrices (that is, with a long positive tale) are likely to contain a large proportion of experimental noise. Positive skewness matrices (that is, with a long negative tail) are likely to be obtained from a population of cells with large structural variability. We applied our new MMP score to three published Hi-C interaction matrices. The results indicate that the 3D reconstruction of two genomic domains from the human and fly datasets as well as the entire *Caulobacter crescentus* genome could result in accurate models.

In summary, we provide a dataset of simulated toy structures and their respective Hi-C matrices that can be used for benchmarking restraint-based methods for 3D reconstruction. Our dataset was used to show that such methods are adequate for building 3D models of genomes and genomic domains. Moreover, we have shown that these methods are robust with respect to experimental noise but are more sensitive to structural variability in the input matrices. Experimentalists aiming to generate 3C-based interaction matrices for 3D reconstruction are thus encouraged to obtain the most homogenous cell population before performing the experiments. Finally, we provide for the first time a new score (here called MMP score) that allows predicting *a priori* the accuracy of the resulting models by calculating a limited number of properties of the input interaction matrices. Such score may probe very useful for defining whether a newly generated interaction matrix can be useful for obtaining accurate 3D models, which can then be more easily explored to extract biological insights.

2.6 Availability

The initial structural sets for the 6 tested toy genome architectures, their derived interaction matrices and the reconstructed 3D models are available at http://www.3DGenomes.org/datasets.

Marie Trussart*, Eva Yus, Sira Martinez, Davide Baù, Yuhei O Tahara, Thomas Pengo, Simon Kretschmer, Jim Swoger, Makoto Miyata, Marc A. Marti-Renom*, Maria Lluch-Senar*, Luís Serrano*: Defined chromosome structure in a minimal cell, submitted July 2015.

3. DEFINED CHROMOSOME STRUCTURE IN A MINIMAL CELL

3.1 Abstract

By combining electron microscopy, super-resolution localization microscopy and Hi-C, we have determined the 3D structure of the chromosome of the genome-reduced bacterium *M. pneumoniae* at 20 kb resolution. We find that despite having a reduced number of structural proteins, the chromosome of this bacterium still has a defined structure. There is a global symmetry between the two chromosomal arms connecting the Ori and Ter, which are located at the two opposite poles of the structure. Analysis of local structures at 5 kb resolution indicates that the chromosome is organized into domains ranging from 15 kb to 35 kb, establishing a fundamental principle of genome organization. We provide evidence that the genes within the same domain tend to be co-regulated, suggesting that chromosome organization could influence transcriptional regulation. The inhibition of supercoiling resulted in a decrease in domain sizes and interaction frequencies, indicating that supercoiling affects domain formation. This study extends the current understanding of bacterial genome organization and demonstrates that a defined chromosomal structure is a universal feature of living systems.

3.2 Introduction

Recent studies have revealed novel insights into chromatin dynamics and their effect on gene expression regulation and replication [1-4, 16, 235]. Such interplay suggests that chromatin organization might have a role in regulating gene expression at both the global and gene-specific level [6, 9-15]. In all kingdoms of life genome organization occurs in a functional and dynamic manner, packaging the genome into the nucleus in the case of eukaryotes and packing it into the cell in the case of bacteria. Simultaneously, DNA-based processes such as transcription, replication and repair are efficiently accommodated. Although technical limitations for chromosome visualization have hampered the understanding of the detailed organization of bacterial chromosomes, several levels of regulation have been identified. At the molecular level, bacteria have evolved mechanisms that condense their chromosomes, such as DNA supercoiling [24, 31] and nucleoid-associated proteins (NAPs) mediated folding [56, 57, 83]. Negative supercoiling forms plectonemic loops of 10 to 100 kilobases (kb) [28, 29], which are maintained by both gyrases and topoisomerases [28, 30] as well as the likely contribution of NAPs [115]. Moreover, NAPs also play a role in chromosome segregation and DNA repair [55-57]. It has been shown that changes in DNA supercoiling can control transcription in bacteria [117-120]. This could be more important in small-genome bacteria such as *Mycoplasma genitalium* [121] where, despite the absence of many structural DNA-binding proteins [122], both gyrases and topoisomerases are present to control gene expression through changes in the DNA local structure [121]. On a larger scale, it has been shown that the *Escherichia coli* genome consists of four macrodomain-like regions of about 1 megabase (Mb) each and two less constrained regions [130], all of which influence the segregation and mobility of the chromosome [131].

In the past, diffraction-limited resolution has impaired the detailed characterization of chromosome structure. However, more recent developments in super-resolution localization microscopy [252-255] and chromosome conformation capture (3C)-based techniques [132] have enabled the determination of global chromosome organization of some bacteria [8, 18, 256]. High-throughput derivations of genome-wide 3C-based assays such as Hi-C technologies [139] have been used to generate high-resolution contact maps of genomes, which when combined with modeling, can provide three-dimensional (3D) representations of the genome structure [113, 162, 163, 240]. Such

studies of bacterial chromosome organization and regulation have been carried out in *Caulobacter crescentus* with a Hi-C map at 13 kb [162], *E. coli* at 20 kb [167] and *Bacillus subtilis* at 30 kb resolution [166]. These studies have shown that genome structure is globally related to the processes of chromosome segregation in *C. crescentus* and to DNA replication and transcription in *E. coli*. More recently, a higher resolution Hi-C map of *C. crescentus* at 10 kb [113] revealed that its genome is divided into 23 chromosome interacting domains (CIDs) or highly self-interacting regions, similarly to the topologically associating domains (TADs) found in eukaryotes [19, 20], but with a size ranging from 30 to 400 kb [113]. In *C. crescentus,* the strongest determinant of these domain boundaries was the presence of highly expressed genes, whereas surprisingly the absence of the NAP heat unstable (HU) histone-like proteins and structural maintenance of chromosomes (SMC) proteins did not affect domain boundaries significantly [113]. No such domains were described in the lower resolution Hi-C map of *B. subtilis* and *E. coli*. Nevertheless, it was found that histone-like proteins such as factor for inversion simulation (Fis), integration host factor (IHF) and histone-like nucleoid structuring (H-NS) do not contribute to the global organization of the *E. coli* genome [167].

The above mentioned bacteria all have large and complex genomes with sizes above 4 Mb coding for hundreds of transcription factors (TFs) [168, 169], multiple DNA structural proteins, and several sigma factors that play key roles in the response to physiological and environmental signals [170]. How this structural organization is achieved and what its impact is in transcriptional regulation remains an open question. Furthermore, whether smaller bacteria with reduced genomes and few structural proteins keep a defined chromosome structure is also undetermined. To address this question, we studied the chromosome organization of the genome-reduced bacterium, *Mycoplasma pneumoniae*, which has minimal genetic components and several structural DNA-binding proteins absent [122]. *M. pneumoniae* is one of the smallest self-replicating organisms [189] that causes atypical pneumonia in humans [171]. This bacterium does not have a cell wall and possesses an attachment organelle (AO) that is located at one cell pole [211] and is involved in adherence, motility and cell division [197, 201, 205, 257]. It has no defined nucleoid, but rather the chromosome occupies the available space [211]. *M. pneumoniae* only has a few known NAPs compared to other bacteria (Table 1): with MPN529, IHF-HU possibly affecting DNA topology [123]; MPN426, SMC family; MPN 229, SSB binding single stranded DNA

(ssDNA) [124]; MPN 554, binding ssDNA [125] and possible evidence for a homolog of CbpA, MPN002, Xdj1. It also has very few TFs and only two sigma factors are found in its genome (Table 1) [258]. In addition, *M. pneumoniae* has been systematically characterized in a quantitative manner by transcriptomics, proteomics and metabolomics studies [223, 227-232].

Gene number	**Gene name**	**Protein name**
MPN002	*cbpA*	Curved DNA-binding protein CbpA
MPN003	*gyrB*	DNA gyrase subunit B
MPN004	*gyrA*	DNA gyrase subunit A
MPN122	*parB*	DNA topoisomerase 4 subunit B
MPN123	*parC*	DNA topoisomerase 4 subunit A
MPN124	*hrcA*	Heat-inducible transcription repressor hrcA
MPN229	*ssbA*	SSB binding single stranded DNA (ssDNA)
MPN239	*gntR*	Probable HTH-type transcriptional regulator gntR
MPN241	*whiA*	Transcription factor with WhiA C-terminal domain
MPN266	*spxA*	Transcriptional regulator Spx
MPN275	*ybaB*	DNA-binding protein, YbaB/EbfC family
MPN294	*araC*	AraC-like transcriptional regulator
MPN332	*lon*	ATP-dependent protease La (EC 3.4.21.53)
MPN352	*sigA*	RNA polymerase sigma factor rpoD (Sigma-A) (EC 2.7.7.6)
MPN424	*ylxM*	Putative helix-turn-helix protein, YlxM/p13-like protein
MPN426	*smc*	SMC family, chromosome/DNA binding/protecting functions
MPN478	*yrbC*	YebC family protein (transcription factor of the tetR family)
MPN529	*ihf*	Histone-like bacterial DNA-binding protein
MPN554	*ssbB*	Putative single-stranded DNA-binding protein
MPN572	*pepA*	Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase) (LAP) (Leucyl aminopeptidase)
MPN608	*phoU*	Transcriptional regulator involved in phosphate transport system
MPN626	*mpn626*	Alternative sigma factor
MPN686	*dnaA*	Chromosomal replication initiator protein dnaA
MPN688	*SpoJ/ParA*	Member of the ParA family of ATPases involved in plasmid and chromosomal segregation

Table 1: List of assigned transcription factors, sigma factors and structural proteins

Here, by combining electron microscopy, super-resolution light microscopy and Hi-C, we have determined the 3D structure of *M. pneumoniae* chromosome at 20 kb resolution and 5 kb for local structures. We observed a general symmetry along the axis of the origin (Ori) and terminus (Ter) of replication and found that Ori and Ter are located at

the two opposite poles of the chromosome structure. Moreover we detected that the chromosome is organized into 24 CIDs, ranging from 15 kb to 35 kb, which are smaller than the CIDs previously described for *C. crescentus* [113]. Inhibiting supercoiling induced a decrease in the domain sizes and interaction frequencies, suggesting that supercoiling might play a role in the regulation of these domains. Interestingly, we provide the first evidence that genes inside CIDs tend to be co-regulated, suggesting that chromosome organization could influence transcriptional regulation. Our results, together with previous 3D structures of other bacterial chromosomes and data on eukaryotes, indicate that chromosome organization in cells is a widespread phenomenon in life.

3.3 Material and Methods

a) Overview of Methodology

With the aim of reconstructing the 3D genome structure of the *M. pneumoniae* chromosome, we first performed Hi-C, enabling the purification of ligation products and subsequent massive parallel sequencing [139]. Next, all fragments of reads were mapped to the *M. pneumoniae* genome with the iterative mapping pipeline ICE hiclib [148], which were further filtered and normalized as previously described [148] to obtain a genome-wide chromatin contact map. Next the MMP score of the matrix [259] was computed to assess its modeling potential. Finally 3D models of the *M. pneumoniae* genome were generated using TADbit [240]. To validate the obtained 3D architecture of the genome, fluorescent and electron microscopy was performed to estimate the cell dimensions and volume as well as distances between different chromosomal regions.

b) Chromosome conformation capture with next generation sequencing

Hi-C protocol with a 6-cutter [139]

To fix the long range DNA interactions, $3 \cdot 10^9$ *M. pneumoniae* M129 cells were grown in 150 cm^2 dishes for 6h (exponential phase) or for 96 h (stationary phase) and were cross-linked with 1% formaldehyde (methanol free, Pierce) for 10 min at room temperature (RT). The reaction was stopped with 0.125 M glycine and cells were washed prior to lysis. Four mL of lysis buffer (10 mM Tris·HCl pH 8.0, 10 mM NaCl, 0.2% NP-40, protease inhibitors from Roche, 1 mM EGTA) was added and cells were broken with the help of a syringe/G25 needle (5x). The lysate was distributed into four tubes and spun in a tabletop centrifuge at 5,000 rpm for 5 min. The supernatant was removed and three pellets frozen for later use. One chromatin pellet was washed twice with 1.4 mL NEBuffer 2/3 (HindIII). After resuspension in 1 mL NEBuffer 2/3, 10 µL 10% SDS was added, mixed carefully and incubated at 65ºC for 10 min to allow accessibility of restriction enzymes. Tubes were placed back on ice immediately after incubation. SDS was quenched by adding 110 µL 20% Triton X-100 and mixed carefully. Chromatin was digested by adding 100 µL 20,000 U/mL HindIII + 5 mM EGTA and incubated at 37ºC overnight (O/N) while shaking. The next steps include marking the DNA ends with biotin and performing blunt-end ligation of cross-linked

fragments. This last step allows ligation junctions to be purified later. To fill in the restriction fragment overhangs and mark the DNA ends with biotin, 5 μL of a mixture containing 10 mM dATP, dGTP and dTTP, 62.5 μL 0.4 mM biotin-14-dCTP, and 41 μL 2 U/μL Klenow was added to the Hi-C tubes, mixed carefully and incubated for 45 min at 37ºC. To inactivate the enzymes, 250 μL 10% SDS was added to the Hi-C tubes, before incubation at 65ºC for exactly 30 min and placed on ice immediately afterwards. The ligation is performed under extremely dilute conditions in order to favor ligation events between cross-linked fragments. Working on ice, 9 mL ligation mix (0.5 mL 20% Triton X-100, 1 mL 10x T4 ligation buffer, and 7.5 mL water) was added to a 50 mL falcon tube and the digested chromatin was incorporated into the mixture to the corresponding tube. After mixing by inverting the tubes, the ligation was performed for 4 h at 17ºC. Cross-links were reversed and proteins degraded by adding 50 μl 20 mg/mL proteinase K per Hi-C tube and incubating the tubes O/N at 65ºC. An additional 50 μl 20 mg/mL proteinase K was added per tube the next day and incubated at 65ºC for another 2 h. The reaction mixture was cooled to RT and DNA was purified by performing an extraction in Maxtract tubes (Qiagen) with one volume phenol pH 8.0 and then with phenol/chloroform/IAA (25:25:1) (at each step the tube was vortexed for 2 min, spun for 5 min, 1500g, RT and carefully as much of the aqueous phase as possible and was transferred to a new 50 mL tube). Then DNA was precipitated by adding 2 μl glycogen, 0.1x volume of 3 M sodium acetate, pH 5.5 and 2x volumes ethanol, left 30 min at -20°C and spun 25 min at 12,000 rpm (Beckman-Coulter 25,50 rotor) at 4ºC. The pellet was washed with ~5 mL 75% ethanol and air-dried before dissolving it in 400 μL TE (10 mM Tris·HCl pH 8.0, 1 mM EDTA). The DNA mixture was transferred to a clean 1.5 mL centrifuge tube and an agarose gel was run as a control. Another round of purification was performed by doing one phenol/chloroform/IAA extraction and DNA precipitation by adding 0.1x volume of 3 M sodium acetate, 2x volume of ethanol and incubating 30 min at -80ºC. After spinning down the precipitated DNA, the DNA pellet was washed with 70% ethanol and resuspended in 25 μL TE. To degrade any RNA that might be present, 1 μL 1 mg/mL RNAse A was added per tube and incubated for 30 min at 37ºC. Some fragments do not get ligated: to avoid pulling them down later, biotin was removed from these unligated ends using the exonuclease activity of T4 DNA polymerase, as follows: Ca. 5 μg (~25 μL) of Hi-C DNA was mixed with 1 μL 10 mg/ml BSA, 10 μL 10x NEBuffer 2, 1 μL

10 mM dATP, 1 μL 10 mM dGTP and 5 U T4 DNA polymerase in a total volume of 100 μL and incubated at 12ºC for 2 h. The reaction was stopped by adding 2 μL of 0.5 M EDTA pH 8.0. To purify the DNA, a phenol/chloroform/IAA extraction was done followed by ethanol precipitation as above. The supernatant was discarded and the DNA pellets resuspended in 50 μL water. Then the DNA was sheared and size selected, to obtain a uniform size suitable for high-throughput sequencing. The DNA must be sheared to a size of 300-500 bp with a Covaris nebulizer (10% duty cycle, intensity: 2,200 cycles, 45 s at 4ºC) in a minimum of 55 μL TE. The size was checked on a 2% agarose gel and the concentration was measured with Qubit (DNA, High sensitivity, Invitrogen). To repair the sheared DNA ends, the Next (NEB) protocol was followed (blunting and A tailing). Subsequently the junctions were enriched by biotin pull-down, thus allowing for the identification of interacting chromatin fragments by paired-end sequencing, as follows: Ligation junctions were purified from the DNA pool, first, 150 μL resuspended streptavidin Dynabeads (Invitrogen) beads were washed twice with 400 μL Tween Buffer (TB: 5 mM Tris·HCl pH 8.0, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween). All washes were done in the same manner: (i). buffer added to the beads, (ii) sample rotated for 3 min at RT, (iii) sample spun briefly to collect all of the suspension, (iv) beads reclaimed using a magnetic particle concentrator, and (v) supernatant removed and beads were resuspended in 600 μL No Tween Buffer (NTB: 5 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 M NaCl) plus Hi-C DNA (~500 ng). Binding was allowed by incubating the mixture at RT for 15 min with rotation, and reclaiming the DNA bound streptavidin beads as above, before washing in 400 μL NTB followed by 100 μL T4 ligase buffer (NEB). Finally the beads were resuspended in 50 μL ligation buffer and Illumina paired end adapters were ligated (ratio: 1 μL 2 μM primers per 10 ng DNA) with 1,200 Units T4 DNA Ligase (NEB) for 2 h at RT. Non-ligated adapters were removed by reclaiming the Hi-C DNA bound beads and washing the beads twice with 400 μL TB, once with 200 μL NTB, and finally once with 200 μL and then 50 μL NEBuffer 2. After the last wash, the beads were resuspended in 25 μL NEBuffer 2. The library was PCR amplified with Phusion (Next kit, NEB), 2 μl of the suspension in a 50 μL reaction, and 1.0 and 2.1 Illumina primers (1 μL 10 uM), for 16 cycles and sequenced in the HiSeq Illumina platform.

Hi-C protocol with a 4-cutter [167]

Chromatin was prepared as above. When indicated, 100 μg/mL novobiocin (Sigma) was added directly to the medium 30 min prior to fixation. Cells were lysed with 4 mL Hi-C Lysis buffer (10 mM Tris·HCl pH 8.0, 10 mM NaCl, 0.2% NP-40, 5 mM EGTA and protease inhibitors) at 4ºC, passed 5x through a syringe/G25 needle and chromatin was collected by centrifugation (5,000 rpm for 5 min, 4ºC, tabletop centrifuge). Only one pellet was used (the rest were frozen at -80ºC), and was washed twice in 1 mL of NEBuffer 1 plus 5 mM EGTA at 4ºC. Before digestion chromatin was solubilized by adding 300 μL NEBuffer 1, 5 mM EGTA and 0.1% SDS and incubated 1 h at 37ºC, and stopped with Tx-100 (2% final). Afterwards, 100 U HpaII was added and incubated O/N at 37ºC. The reaction was stopped adding SDS to a final concentration of 1.3% and incubated for 1 h at 50ºC. Half of the sample was ligated by adding 5 mL 10x NEB T4 Ligase buffer, 2.5 mL 20% Tx-100, 0.5 mL 0.5M EGTA, in a final volume of 50 mL; and ligated with 20U (50 μL) T4 DNA ligase, O/N at 16ºC. The sample was decrosslinked with 375 μL proteinase K, 2h at 65ºC and purified with phenol extraction and Maxtract resin and ethanol precipitated as above. Further fragmentation was performed with Covaris to reduce the size of DNA to ~200-500 bases (Duty cycle: 10%, int.: 2, 200 cycles, 20 s at 4ºC). DNA was submitted to the CRG Ultrasequencing facility for standard Illumina library prep and paired-end sequencing.

Genomic DNA prep

For the controls without formaldehyde fixation, genomic DNA was prepared as in [229] and digested and religated as above (without the need of decrosslinking). The same equivalent concentration was used in order to keep the infinite dilution conditions.

c) Generation of contact matrix

To construct the interactions maps of the *M. pneumoniae* genome, read pairs of 50bp were uniquely mapped to the MPN129 reference genome (NC_000912, NCBI) covering 816,394 bp, using Bowtie2 [145] and following the iterative mapping strategy ICE from hiclib Python library [148]. The optimal start and end positions for mapping were determined using the fastq quality of the read and set to 4 and 44, and the minimal size for mapping was set to 25 bp. We constructed a genome wide matrix M of different resolutions 5, 10, 15 and 20 kb by dividing the genome into 5, 10, 15 and 20 kb bins,

pooling interactions into their corresponding bins. To correct for possible artifacts of Hi-C experiments, the matrix was then filtered and normalized using the methodology of iterative correction with hiclib Python library [148] as done in a previous study in *C. crescentus* [113]. The total number of reads before and after filtering are shown in Table S2. In addition, using a control library without formaldehyde fixation, we filtered interactions off-diagonal and off-diagonal plus one, that are not due to 3D contacts in the chromosome, representing about 8% of the total number of cells in the matrix. These interactions were found in two regions with a high sequence similitude computed by the Needleman-Wunsch global sequence alignment with EMBOSS Needle [260], which justifies possible PCR artifact amplification for repetitive sequences. The affected bins were: 1, 3, 4, 5, 6, 7, 9, 10, 11, 13, 17, 18, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, and 41.

d) Reproducibility of Hi-C data

To analyze reproducibility between HindIII1 and HindIII2, and HpaII and HindIII1 in stationary phase, and HpaII in exponential and stationary phase, we decomposed the two-dimensional matrices of normalized and filtered datasets into two one-dimensional vectors row-by-row and computed Pearson correlation between the two vectors with R.

e) Matrix Modeling Potential using MMP score

We computed the matrix modeling potential (MMP) score of the matrix to assess its potential for modeling, using the MMPscore python script [259]. This score is based on the matrix size, the contribution of significant eigenvectors in the matrix and the skewness and kurtosis of the z-scores distribution of the matrix.

f) Integrative 3D Modeling with TADbit

The HpaII Hi-C matrix was used for modeling at a resolution of 20 kb after filtering by hiclib methodology [148] and additional filtering using a control library as previously mentioned. To build the 3D models, we applied a restraint-based modeling approach using the TADbit python library [240, 259]. The genome was defined by 41 particles, determined by the resolution of the contact map at 20 kb. Each particle had a radius of 36 nm that was determined empirically with the scale parameter of 0.0036 nm/bp. First, TADbit identified empirically three optimal parameters using a grid search: (i) the

proximal distance between two non-interacting particles set as 100nm, (ii) a lower-bound cut-off to define particles that do not interact frequently, set as -0.6 and (iii) an upper-bound cut-off to define particles that do interact frequently, set as -0.2. Subsequently, considering an inverse relationship between the frequencies of interactions of the contact map and the corresponding spatial distances, TADbit translated the frequencies of interactions into spatial restraints between particles. Two consecutive particles were spatially restrained at an equilibrium distance, which corresponds to the sum of their radii. Non-consecutive particles with contact frequencies above the upper-bound cut-off were restrained at an equilibrium distance, while those below the lower-bound cut-off were maintained further than an equilibrium distance. Second, TADbit used a Monte Carlo simulated annealing sampling procedure to identify 3D models that best satisfy all of the imposed restraints. A contact map obtained from the final models resulted in a Pearson correlation of 0.83 to the input Hi-C interaction matrix, which is indicative of good model accuracy [259].

g) TEM imaging

Mycoplasma cells were recovered into fresh growth medium at 50-fold the concentration of the original culture. The cell suspension was put on a 3×3 mm piece of glass and left at 37ºC for 15 min. The cells on the glass were fixed with 1% glutaraldehyde in PBS consisting of 75 mM sodium phosphate (pH 7.3) and 68 mM NaCl for 3 min at RT, rinsed with 0.2 mL PBS once and then thoroughly washed in water. The fixed cells were frozen at a liquid nitrogen temperature using CryoPress (Valiant Instruments, St. Louis, MO), deep-etched, rotary-shadowed by platinum at an angle of 30 degrees, and backed with carbon in a JFDV freeze-etching machine (JEOL Ltd, Akishima, Japan). Replicas were floated from the glass by slow immersion along the surface of full-strength hydrofluoric acid, cleaned with a commercial bleach, rinsed in water, and picked up onto Formvar-coated 400-mesh copper grids as described [261]. The series of replica images were taken by tilting the sample stage for 30 degrees to both sides with 5 degrees intervals, by a transmission electron microscope (JEM1010, JEOL) at 80 kV.

h) 3D reconstruction and cell volume

TEM images tilted with angles –30, 25, 20, 15, 10, 0, 5, 10, 15, 20, 25, 30 were registered by cross-correlation with Matlab and rotated to ensure a vertical rotation axis. The outline of the cell in each image was determined by thresholding using Fiji, followed by manual removal of background contamination and filling of gaps inside the bacteria. From these binarized images the sample area could be extracted by counting pixels within the cell area.

The volume of the cell was then calculated by assuming each cell is rotationally symmetric along its long axis (Video S2). The cell was segmented into cylinders and cones along this axis, and the volume was computed as the sum of the cylinders'volumes and the cones' volumes as follows:

$V_{cyl} = \pi \times r^2 \times h$ and $V_{cone} = \frac{1}{3} \pi \times r^2 \times h$

where V_{cyl} is the volume of a cylinder of height h and radius r and V_{cone} is the volume of a cone of height h and radius r.

To reduce the error due to inaccuracies in the manual image editing or a lack of rotational symmetry, the final volume was calculated as the average of the volume of the individual images.

i) Estimation of chromosome dimensions and volume

The mean length, mean width and mean volume of the chromosome were computed over the 1000 models lengths, widths and volumes. The length of the chromosome in each model was estimated as the distance between the two most distant particles of the model and the width as the double of the radius of gyration of the model. To calculate the volume of the model, we first computed the diagonal of a cube where the height, length and width were given by the difference between the minimum and maximum coordinates for each the x, y and z axis, and we computed the volume of the cube where the height, length and width were equal to this diagonal, that would enclose the whole model.

j) Fluorescence In Situ Hybridization (FISH) combined with Immunofluorescence

We estimated the distances between four regions of interest corresponding to the four quarters of the circular genome of *M. pneumoniae,* namely Ori, Right, Ter and Left as shown in Table S3, performing FISH of those regions combined with immunofluorescence of the P1 protein localized in the AO [196, 197, 201].

The resolution of a regular fluorescence microscope image is limited by diffraction of about half the wavelength of the emitted light, which due to the small size of *M. pneumoniae* does not allow localizing the region marked by the fluorescent probe. To overcome the diffraction limit, we used a high-resolution microscope with ground-state depletion followed by individual molecule return (GSDIM) that improves resolution down to 20 nm [252-255]. The principle resides in ensuring that only a few illuminated fluorophores are able to emit simultaneously, allowing each one to be localized individually with a resolution below the diffraction limit. To do so, a strong continuous excitation light source is used so that most of the fluorophores instantly go into a temporary dark state and only a few switch stochastically to an active state and are able to fluoresce [252-254]. The microscope records the precise position of the fluorophores over a series of imaging cycles. Because in our setup each color needs to be imaged sequentially, we have not been able to observe two genomic probes of the genome marked by FISH simultaneously, as the second probe was not resistant to two consecutive sessions of strong illumination. To overcome this technical limitation, we combined one genomic probe marked by FISH with the immunofluorescence of the protein P1 adhesin of the attachment organelle. We first observed the genomic probe and then the region marked by immunofluorescence in a second session, which proved to be resistant to photobleaching. We found that the Ter probe was close to the AO, and therefore we could deduce the distances between Ori-Ter, Right-Ter and Left-Ter, approximated as their median distance to AO minus the median distance between Ter-AO.

DNA probes preparation for FISH

Standard PCRs were performed with genomic DNA to amplify four regions of interest, with the following pairs of primers Ori (F_Ori,R_Ori), Ter (F_Ter, R_Ter), Right (F_90C, R_90C) and Left (F_270C, R_270C) (Table S3). The different amplified

fragments were labeled by adapting the protocol from the Random Primed DNA Labeling Kit (Roche). Briefly, the probes were denatured by incubating 10 min at 100ºC and mixed with the following reagents: 5 µL 10x Klenow buffer, 0.25 µL 100 mM dATP, dCTP, dGTP, 0.16 µL 100m dTTP, 2.5 µL 1 mM ChromaTide Alexa Fluor 568-5-dUTP (Life Technologies), 1.6 µL 3 µg/µl random hexamers and 0.25 µL 10 U/µL Klenow fragment, before O/N incubation at 37ºC. The reaction was stopped by adding 2 µl of 0.2M EDTA and incubated 10 min at 65ºC. The labeled probes were then purified by ethanol precipitation.

FISH labeling

M. pneumoniae cells were grown in a 75 cm^2 flask in Hayflick medium for 4 days under standard conditions. After 4 days, the medium was removed and cells were re-suspended in 5 mL of fresh medium (Hayflick), then scrapped and collected. Cells were then passed through a syringe/G25 needle (10x) and a filter (0.45 µm) and mixed with 5 mL of 6% gelatin. Then cells were grown on borosilicate and CC2 coverglass slides (Thermo Scientific) for 6 h, as replicates. Cells were fixed in a final concentration of 4% formaldehyde (Pierce) for 20 min at RT followed by 40 min at 4°C, before further fixing with cold methanol at -20°C O/N.

After washing twice with PBS, two washes were done with 2x SSC /Tween-20 for 5 min, then with 2x SSC/formamide at 37ºC for 30 min. Each genomic probe was then mixed with a hybridization buffer (2x SSC, 50% formamide 100 µg/mL salmon DNA sperm) and warmed at 95ºC for 10 min. In parallel the slides were also warmed at 95ºC for 2 min before adding the probes to the slides and incubated at 42ºC O/N.

Several washes were then done, twice with 2x SSC/formamide for 30 min at 37ºC, then with SSC 2X/ 25% Form for 10min, 3x with 2x SSC for 10 min and finally briefly with PBS.

When immunofluorescence localization of P1 protein was required, samples were blocked during 1 h by using 2% Elisa reagent Blocking solution (Roche). Then, a primary antibody from rabbit recognizing P1 adhesin protein of the Attachment Organelle (Organelle 65114, provided by Prof. Richard Herrmann) was incubated for 1 h at RT with blocking solution. After three washes with PBS 1x/Tween-20 0.05% for 15 min at RT, the secondary antibody (anti-rabbit marked with Alexa 488) was added to the slides in blocking solution for 1h at RT. After three washes with PBS 1x/Tween-20

0.05% of 15min at RT, the removable slide chambers were dried and mounted on glass slides using the Prolong Gold antifade reagent (Life Technologies).

k) FISH imaging acquisition and processing

The super-resolution microscope used to acquire the data was a ground-state depletion (GSD) Multiline TIRF microscope (Leica, Wetzlar, Germany) using the proprietary Leica software, equipped with a 1.46 NA 100x TIRF objective and an Andor iXon EMCCD camera. We processed the data using rapidSTORM [262] and used PALMsiever [263] for filtering and rendering. Finally, using Fiji [264], the distances were calculated between the center of mass of the observed probes.

l) Domain detection on Hi-C contact map

The filtered raw HpaII matrix at 5 kb resolution was used for domain detection. First, TADbit program normalized the matrix with a single iteration of ICE. Then TADbit returned the optimal segmentation of the chromosome under BIC-penalized likelihood. The algorithm for the domains detection uses a change-point algorithm, inspired by methods used to detect copy number variations in CGH experiments [265]. The model assumes that counts have a Poisson distribution and that the expected value of the counts decreases like a power-law to the linear distance on the chromosome. The details of the implementation of TADbit will be further defined elsewhere (Serra et al., manuscript in preparation).

m) Co-expression levels analysis (RNA-seq)

A co-expression tendency was computed based on RNA-seq expression over 282 conditions and represents the fraction of conditions in which the pair of genes vary in the same direction minus the conditions in which they vary in opposite directions (Junier, Ünal, Yus, manuscript in preparation).

n) HpaII sites number on domains borders

We computed the number of sites on the 48 domain borders, with each domain border being defined by two bins, the last bin of the previous domain and the first bin of the next domain. We evaluated whether it is significant using a permutation test, where all

48 domain border positions were shifted across the genome, but conserved both the size and number of domains. We obtained a number of sites per domain border for each permutation that we could compare to the original case. We obtained an empirical p-value, calculated as the ratio between the number of values that are higher than or equal to the observed value in the original domain border case.

o) High co-expression levels within domains

We computed the mean co-expression levels between pairs of genes within the same domain, compared to pairs of genes where the first gene is localized within one domain and the second gene in a different domain. Then we computed the Mann–Whitney test p-value to compare the two distributions.

p) Low co-expression levels surrounding domains borders

To assess whether there is a significant low co-expression in the domain borders, we performed a permutation test, where all domain border positions were shifted across the genome, but conserving both the size and number of genomic domains. Then, for each permutation, we calculated the mean co-expression levels of the genes present in the domain borders. Finally we computed the empirical p-value as the ratio between the number of values that are lower than or equal to the observed value in the original domain border case.

3.4 Results

a) Generation of the Hi-C map of the *M. pneumoniae* genome

Here, we have studied the *M. pneumoniae* genome during stationary phase, using the four-cutter enzyme HpaII and the six-cutter enzyme HindIII, which have average cutting frequencies of 450 bp and 1810 bp, respectively. Although the Hi-C interaction maps obtained at exponential and stationary phase display similar features, the analysis of the chromosome structure at exponential phase could be hampered by heterogeneity, as it is not possible to synchronize *M. pneumoniae*. Therefore we concentrated on the stationary phase samples. To analyze the Hi-C datasets, the paired-end library reads were first mapped and then further filtered and normalized as previously described [148] (Methods). Next, the interaction frequencies were grouped into 41 genomic loci of 20 kb each, called bins. The final frequencies of interactions were represented as two dimensional matrices where M(i,j) indicates the relative frequency of interactions between fragments in bins i and j. The decision to use the resolution of 20 kb was determined based on the correlation between all replicates at 5, 10, 15 and 20 kb resolutions (Fig. S1) as well as the matrix modeling potential (MMP) score [259] of the resulting matrices (Table S1). The HindIII dataset resulted in an MMP score of 0.80, with a predicted model accuracy of 0.78 (0.66-0.89 at 95% confidence interval). The HpaII dataset resulted in an MMP score of 0.80 with a predicted model accuracy of 0.79 (0.67-0.90 at 95% confidence interval). Additionally, the matrices were highly reproducible between biological replicates (r=0.93) as well as between HindIII and HpaII datasets (r=0.95). Even though no significant differences were found between the two enzymes, since the HpaII enzyme has a higher frequency of cutting, we decided to use the HpaII interaction matrices at 20 kb resolution for modeling and subsequent analysis.

The resulting Hi-C interaction map had two diagonals intersecting near the center of the chromosome (Fig. 1a). The potential Ori was predicted by the position of the dnaA boxes [266] but the exact localization of the Ter has not been experimentally determined in this bacterium. Since in bacteria the Ter is located opposite to the Ori [267, 268], we predicted it to be at this intersection point (~400 kb). The main prominent diagonal, characteristic of Hi-C maps, results from the local contact of proximal genomic regions. The second less prominent diagonal (from the lower right to

the upper left corner) reflects both the circularity of the genome and the interactions between fragments located on the opposite arm of the chromosome. All together this indicates that the chromosome has a global symmetry within the cell extending from the Ori-Ter axis. Interestingly, such symmetry is also observed in the linear organization of the genome where genes are distributed in opposite strands both in the left and the right arms of the chromosome (Fig. 1b).

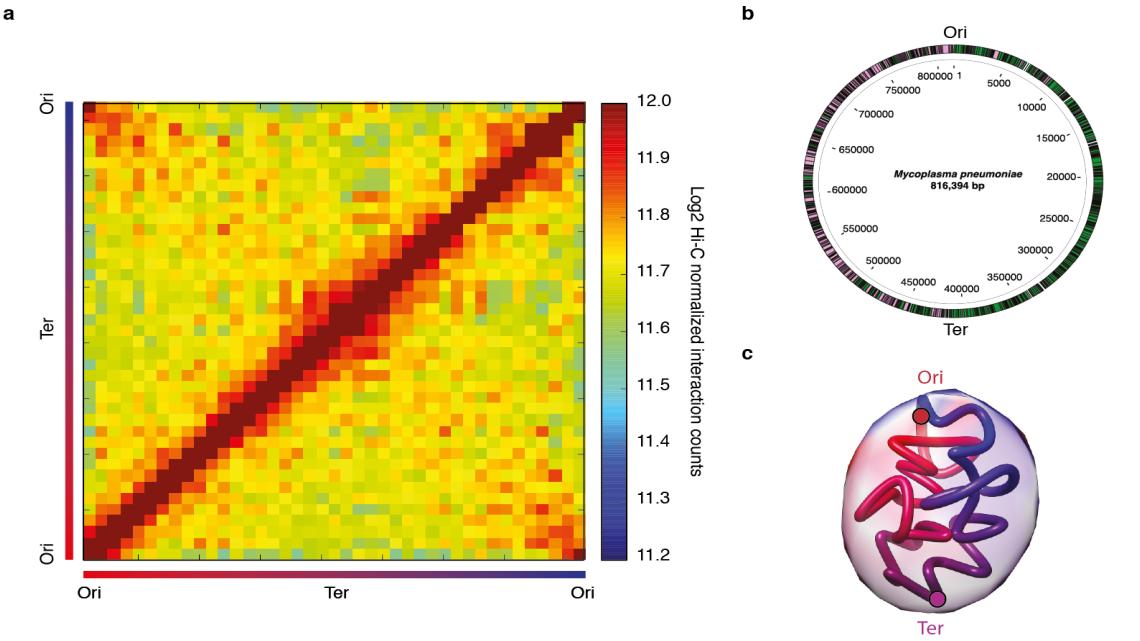

Figure 1: Hi-C matrix and 3D models of the *M. pneumoniae* chromosome reveal a global symmetry with Ori and Ter located at the two opposite poles. **(a)** Normalized HpaII Hi-C contact map of *M. pneumoniae,* in stationary phase with 20 kb resolution. The frequency of interactions between a given pair of bins is found at the intersection of the row and column corresponding to those bins. The color of the contact map, from blue to red, indicates the log2 contact frequency. The bar underneath indicates genome position with Ori being located at a genome coordinate of 0 and Ter located at ~ 400 kb. **(b)** Simplified genomic map showing the gene distribution across the chromosome, with black lines delimitating the genes. The color indicates the strand position, with pink being the - strand and green the strand +. **(c)** 3D density map representations of the first cluster of *M. pneumoniae* genome models with Ori and Ter represented by red and purple circles, respectively. A color tube shows the centroid model, following the same color code of the bar as in (a) and the lighter color represents the space occupied by all the models in the cluster, i.e. the variability across the cluster.

b) 3D modeling reveals a chromosome structure with Ori and Ter localized at the two opposite poles

To assess whether the overall symmetry is also reflected in the spatial organization of the genome, we built 3D models of the genome of *M. pneumoniae* based on the filtered and normalized Hi-C matrix at 20 kb. Briefly, based on the hypothesis that chromatin interaction frequencies are a proxy for spatial proximities between loci [139, 269], we

used TADbit [240, 259] to search for 3D conformations that best satisfy the spatial distances between genomic loci inferred from the frequencies of our Hi-C matrix. A total of 5,000 models were built by an optimized protocol where the loci were initially placed randomly in 3D space and their positions were modified iteratively using simulated annealing and Monte-Carlo sampling to satisfy as many restraints as possible. Finally, we selected the 1,000 models with minimal penalty for not satisfying the imposed restraints and clustered them based on their structural similarity. We found two main clusters corresponding to mirror images of each other, one containing 510 models and the other 490 models. The variability of the models within the cluster is homogeneous along the entire chromosome (Fig. 1c). Similarly to what was observed for the 3D organization of *C. crescentus* [113, 162], the circular chromosome obtained has a global symmetry between the two chromosomal arms connecting the Ori and Ter, which are located at the two opposite poles of the structure (Fig. 1c, Video S1).

c) Chromosome occupancy is about two-thirds of the total cell volume

To ensure that the predicted dimensions and volumes of the models fit within the cell of *M. pneumoniae*, we examined cells by transmission electron microscopy (TEM) using the quick-freeze deep-etch replica method [270] (Fig. 2a). We measured a median cell length of 1,389 nm (with standard deviation of 337 nm) and width of 359 nm (sd of 65 nm), compared to a median chromosome length of 268 nm and mean width of 194 nm obtained from the 3D models (Fig. 2b). We also acquired tilted TEM images of cells that were used for 3D reconstruction (Video S2). Although technical limitations did not allow for a full 3D reconstruction of the whole cell, we could still detect a rotational symmetry along the long cell axis thereby allowing us to estimate the cell volume. In stationary phase, the chromosome volume estimated from the 3D models is 0.050 μm^3, which when compared to a median cell volume of 0.075 μm^3 is about two thirds of the total cell volume. Altogether, in contrast to what was previously shown [211], these results indicate that the modeled 3D genome fits within the cell without occupying the entire available space.

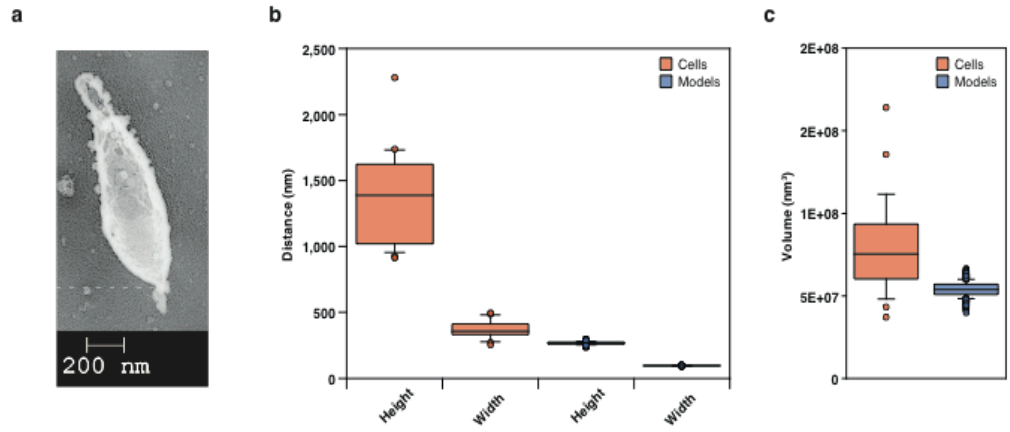

Figure 2: Chromosome occupancy is approximately two thirds of the total cell volume. (a) Quick-freeze deep-etch replica TEM imaging of a *M. pneumoniae* cell. **(b)** Comparison of the estimated heights and widths of cells and chromosome models in nm. Boxplot distribution and median values of height and width over 25 cells, here shown in red, estimated from TEM imaging. Boxplot distribution and median values of height and width, estimated over 1000 chromosome models, shown in blue. **(c)** Distribution and median volume estimated over 25 cells as shown in red, and estimated volume over 1000 chromosome models as shown in blue, in nm^3.

d) Validation of 3D models with fluorescent imaging

The orientation of the chromosome in the cell body was assessed by using 4'6-diamidino-2-phenylindole (DAPI) combined with immunofluorescence of the P1 adhesin protein localized at the AO [201, 204]. We analyzed cells in exponential phase since in stationary phase *M. pneumoniae* clumps in large aggregates. The resulting exponential and stationary contact matrices significantly correlated ($r=0.85$, $p<0.0001$, Fig. S2) suggesting that the overall conformation of the chromosome does not significantly change between the two states. Distances between different genomic regions were determined by super-resolution localization microscopy (Fig. 3a and 3b). We measured distances between fluorescent DNA probes (Fluorescence *in situ* hybridization imaging; FISH) mapped to the Ori (0-1 kb), Right (204 kb-205 kb), Ter (390 kb-391 kb) and Left (612 kb-613 kb) loci (Table S3), in conjunction with the immunofluorescence localization of P1 adhesin (Fig. 3a).

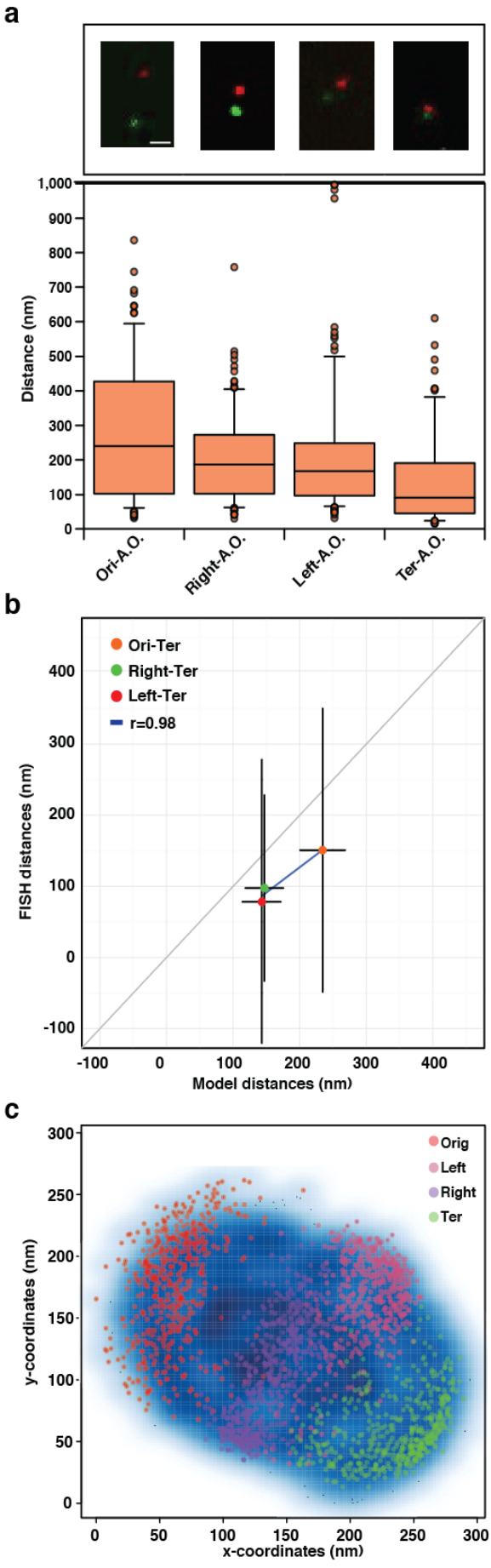
a
1,000
900
800
700
600
500
400
300
200
100
0
Distance (nm)
Ori-A.O.
Right-A.O.
Left-A.O.
Ter-A.O.
b
Ori-Ter
Right-Ter
Left-Ter
r=0.98
400
300
200
100
0
-100
FISH distances (nm)
-100
0
100
200
300
400
Model distances (nm)
c
300
250
200
150
100
50
0
y-coordinates (nm)
Orig
Left
Right
Ter
0
50
100
150
200
250
300
x-coordinates (nm)

Figure 3: Validation of 3D models with super-resolution imaging. (a) (Top) FISH imaging with red (Alexa Fluor 568) indicating genomic probes Ori, Right, Left, Ter respectively, and green (Alexa Fluor 488) representing the P1 adhesin, attachment organelle protein. Scale bar = 200nm. (Bottom) Boxplot distribution and median distances estimated between the genomic probes and AO over about 100 cells. **(b)** Pearson correlation between the Ori-Ter, Right-Ter and Left-Ter estimated distances from chromosome models in the x-axis and experimental FISH imaging in the y-axis. Black lines indicate the variability within the estimated distribution. **(c)** 2D map representation of chromosomal models from the first cluster shown in blue, with x and y coordinate positions shown in the x-axis and y-axis, respectively. Ori, Left, Right and Ter positions across the first cluster of chromosome models are shown in red, pink, purple and green, respectively.

The Ter-AO measurements displayed the smallest median separation distance of 91nm, while the Left-AO and Right-AO distances were larger (167 nm and 186 nm, respectively). Finally, the Ori-AO distances had the largest separation with a median distance of 240 nm. Additionally, we calculated approximate distances between Ter-Left, Ter-Right and Ter-Ori from their respective distances to the AO by deducting the median Ter-AO distance of 91nm (Fig. 3b). The fact that the Ori-AO measurements have a larger variability compared to the others, particularly to the Ter-AO measurements, probably suggests that after duplication the Ori moves toward the opposite pole, whereas the unduplicated Ter remains fixed throughout the replication process, similar to what was observed during *B. subtilis* replication [127, 271].

Next, to assess whether the distances obtained from the 3D models were congruent with the distances obtained experimentally, we computed the Euclidian distances between Ter-Ori, Ter-Right and Ter-Left given their respective 3D coordinates in the models. The Ter-Left and Ter-Ori distances were respectively the smaller and larger median distances estimated both from the 3D models (144 nm and 236 nm) and the imaging data (76 nm and 149 nm) (Fig. 3b). Although the distances estimated from the 3D models are overall larger than the experimental ones, a Pearson correlation of 0.98 is found between the median distances Ter-Left, Ter-Right and Ter-Ori estimated from both the 3D models and the fluorescent imaging (Fig. 3b). Overall, our imaging data qualitatively validate our 3D models and we conclude that, as was shown in *B. subtilis* and *C. crescentus* [86, 127, 272, 273], the folding of the chromosome is consistent with the linear order of genes along the DNA (Fig. 3c).

e) Genes are co-expressed within chromosome interaction domains

In Hi-C interaction maps, a significant proportion of the signal lies in the vicinity of the diagonal where most of the interactions occur. We have used this property to further increase the resolution of our maps to 5 kb bins. Although such maps have relatively

low scores for accurate 3D modeling (the MMP score of the maps at 5 kb was 0.65, Table S1), they can be used for studying the local organization of chromatin, omitting long-range interaction data. We used the TADbit program to segment the HpaII matrix at 5 kb resolution into CIDs (Fig. 4a).

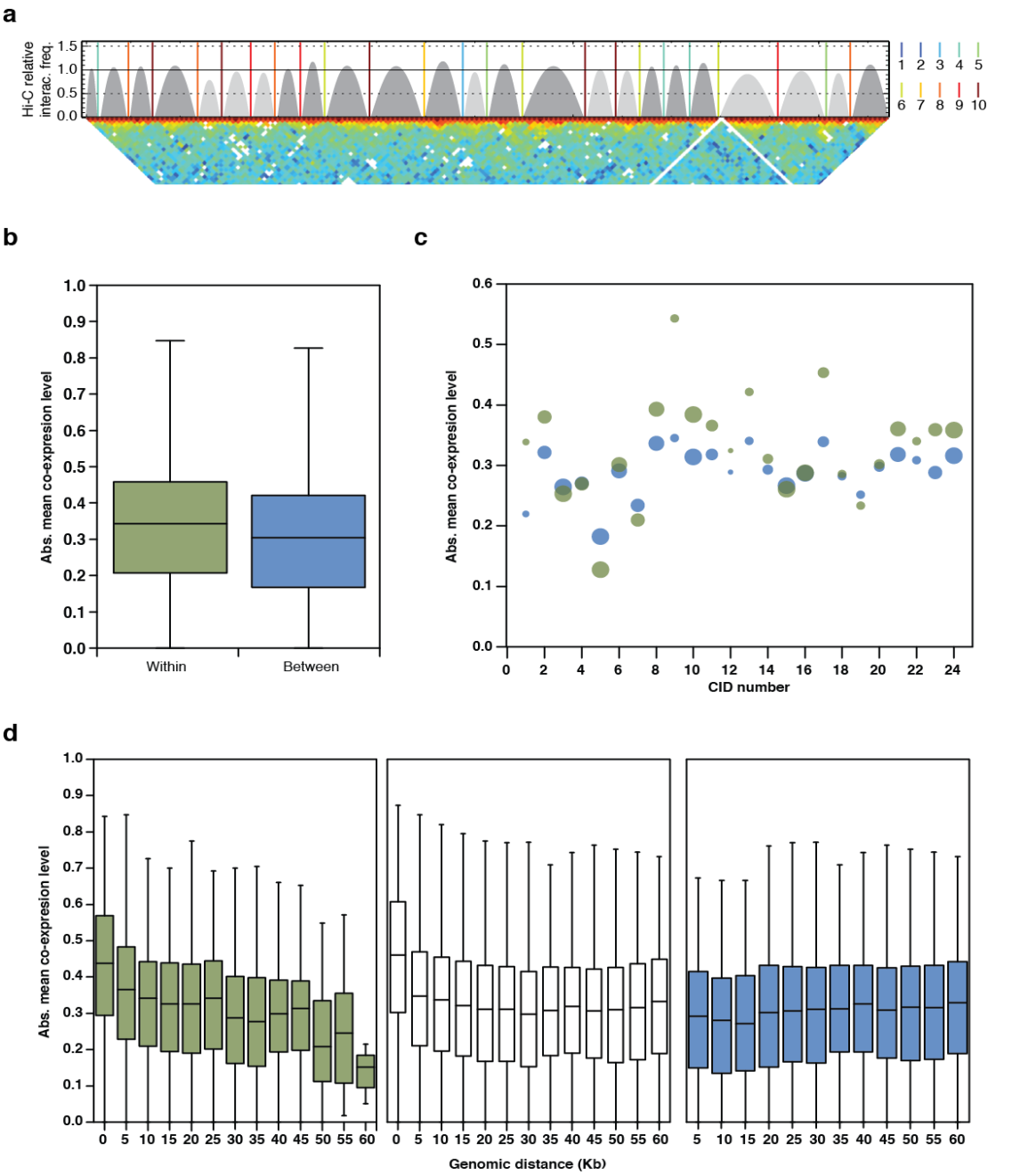

Figure 4: *M. pneumoniae* chromosome is partitioned into domains of co-expressed genes. (a) Hi-C HpaII filtered and normalized contact map at 5 kb resolution, rotated 45º with domain density plots. Each domain is represented by a grey-filled arc and delimited by a colored line. The color code from blue to red, numbered 1 to 10, indicates the border strength or confidence score of the identification of domains. The y-axis displays the relative Hi-C interaction frequencies and the horizontal line at y=1 indicates the expected frequency, given the domain size. **(b)** Absolute mean co-expression distribution of gene pairs, when both genes are located within the same domain as shown in green, or genes between two different domains as shown in blue. **(c)** Detailed absolute mean co-expression distribution across the 24 domains. Point sizes are proportional to domains scores. The color depicts, as before, the two cases of genes pairs

within the same domain, shown in green, and genes pairs between different domains, shown in blue. **(d)** Absolute mean co-expression distribution as a function of genomic distance, with distances between genes pairs smaller than 60 kb for the same two cases as in (a) and any genes pairs across the whole genome, as shown in white**.**

Moreover, TADbit assigned a confidence score to each domain border ranging from 1 to 10, the higher the score, the higher the confidence (Fig. 4a). The CID sizes ranged from 15 kb to 65 kb, with a median of 30 kb, smaller than those previously found in *C. crescentus* which range from 30 to 420 kb [113]. To ensure that the identified domain borders were not artifacts arising from the localization of restriction sites, we calculated the number of HpaII sites present in borders. The analysis confirmed that the domain borders were not significantly enriched with HpaII sites (permutation test p-value=0.66). We also looked at other properties such as gene localization, COG function, TFs enrichment, termination of genes, methylation levels and convergent and divergent pairs of genes, but none of them were found to be significantly enriched with the permutation test. Only GC content was found to change significantly with a lower percentage value at domain borders (permutation test p-value=0.012).

To assess whether the local organization of the *M. pneumoniae* genome into CIDs is related to transcriptional regulation, we compared the absolute mean co-expression of pairs of genes within and between domains. Interestingly, we found that genes are significantly co-expressed within domains (t-test p-value<0.0001) (Fig. 4b). Specifically, higher mean absolute co-expression values were observed for genes in 18 domains out of the 24 domains (Fig. 4c). These results also indicated that the higher co-expression levels for genes within CIDs are not only an effect of genomic linear proximity (Fig. 4d). Indeed, independently of the CID in which the genes are located, proximal genes have higher co-expression than distant genes (> 5 kb). However, the correlation trends are reversed when comparing gene co-expression within or between CIDs (Fig. 4d). Interestingly, the correlation of expression is stronger for proximal genes (< 5 kb) when those genes are located within the same CID, while it is weaker for genes localized in different CIDs (that is, separated by a CID border). Moreover, a border permutation test confirmed that the genome is partitioned into domains of co-expressed genes and a significantly low co-expression is found at the border of these domains (p-value<0.0001). Altogether, our results suggest that the mean co-expression level observed for genes of the same domain is higher than the mean co-expression level observed for genes in close proximity.

f) Inhibiting supercoiling decreases domain sizes and interaction frequencies

To study the effect of inhibiting supercoiling on chromosome structure, we performed Hi-C on cells treated with novobiocin, a drug that inhibits DNA gyrase and DNA negative supercoiling. The outcome of inhibiting DNA gyrase with novobiocin is the relaxation of the DNA [274]. The novobiocin treated cells resulted in Hi-C interaction maps with 27 CIDs (Fig. 5a), ranging from 20 kb to 50 kb and with a median size of 25 kb (Fig. 5b). Interestingly, the TADbit median confidence score for the border domains was 5 out of 10 (the higher the score, the higher the confidence) compared to 8 for non-treated cells (Fig. 5b). Similarly, the domain densities, associated to the relative number of interactions given the domain size, were lower in novobiocin-treated cells, with 12 out of 27 domains having densities lower than expected and a median density of 1.05 compared to 1.02 in the other case (Fig. 5b). Taken together, and as reported in *C. crescentus* [113], novobiocin reduces the CID sizes, the sharpness of its borders and the frequency of interactions within the domains, therefore suggesting that supercoiling may regulate domain formation in bacteria.

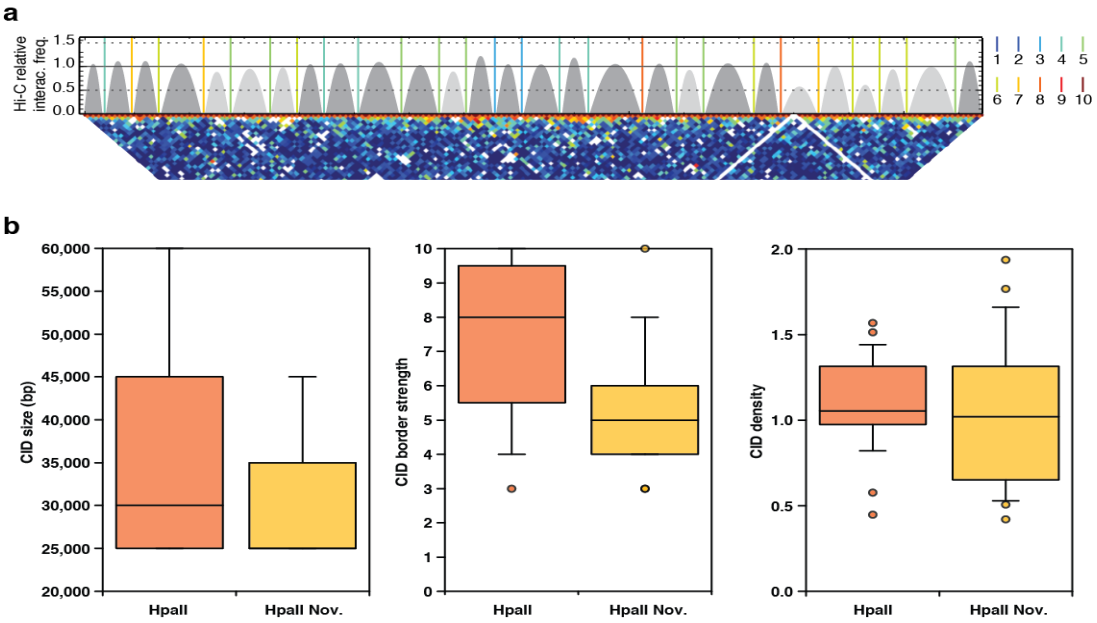

Figure 5: Inhibiting supercoiling decreases domain sizes and interaction frequencies. (a) Same as Fig. 4a, but with Hi-C HpaII Novobiocin-treated contact map at 5 kb resolution. **(b)** Comparison of CID size, border strength with median values, and density distribution in wild-type (red) and Novobiocin-treated (yellow) cells.

3.5 Discussion

Chromosome conformation capture based experiments coupled with deep sequencing have been used to explore bacterial genome organization and its role in transcriptional regulation [113, 162, 166, 167]. Here, we have analyzed the 3D genome organization of the *M. pneumoniae* bacterium, a model organism with a small genome size and a simplified gene regulatory network. Indeed, compared to other bacteria, *M. pneumoniae* has a lower number of TFs and only two sigma factors to coordinate gene transcription. By combining Hi-C and super-resolution fluorescent imaging, we were able to identify fundamental principles of genome organization such as the partitioning of a reduced genome into domains. Furthermore, we provide evidence that genes inside CIDs tend to be co-regulated, indicating that the chromosome structure has a role in transcriptional regulation.

The *M. pneumoniae* genome contact map revealed a double diagonal intersecting near the center of the chromosome, corresponding to the Ter and reflecting the global symmetry of the genome along the Ori-Ter axis. The 3D models generated of the genome conformation resulted in the Ori and Ter loci being located at the two opposite poles of the structure. In addition, our TEM images indicate that the 3D chromosome models fit within the cell of *M. pneumoniae*. DAPI staining of the chromosome showed that *M. pneumoniae* does not have a defined nucleoid in its center but rather the chromosome occupies the available volume [211]. However, we estimated that the chromosome is only occupying about two-thirds of the total cell volume. This can be explained by the limited resolution of our models (20 kb) from which it cannot easily be assessed the actual occupancy of a bin within the cell.

Using super-resolution fluorescent imaging, we corroborated our 3D models of chromosome conformation. Imaging indicates that the Ter locus is the closest of all tested loci to the AO. In *M. pneumoniae*, the duplication of the AO was reported to be coordinated with cell division, which occurs by binary fission [182]. Moreover, during cell replication and before nucleoid separation, the migration of the AO to the opposite pole of the cell was observed in fixed cells, which suggests a coordination between AO duplication and DNA replication [211]. Once a new organelle is formed, it remains attached to the surface, and the old attachment organelle pulls the dividing cell away from the nascent organelle, positioning itself at the opposite pole [220]. Similarly, as described for the analogous species *Mycoplasma. gallisepticum* [219], our findings

suggest possible anchoring of the DNA near the Ter to the AO. Unfortunately, the observed cell-to-cell variability in Ori-AO and Ter-AO distances could not demonstrate that the AO is attached to a specific chromosome region as division occurs. Technical limitations of the FISH protocol only allowed the study of fixed cells, limiting a deeper understanding of cell division in *M. pneumoniae*.

Previously, a double diagonal was observed in the contact map of the phylogenetically closely related gram-positive bacterium *B. subtilis* [166] and also in two other gram-negative bacteria *C. crescentus* [113, 162] and *Vibrio cholera* [166]. Interestingly, this symmetry observed along the two replichores was not observed in *E. coli,* which has an open chromosome structure (Fig. 6a) [166, 167], likely due to the orientation of the chromosome within the cell. Indeed in *B. subtilis* and *C. crescentus,* the Ori and Ter are preferentially located at opposite poles early in the cell cycle [86, 273], with the Ter situated at the new pole and the Left and Right extending along the cell in *C. crescentus* [162] (Fig. 6b,c).

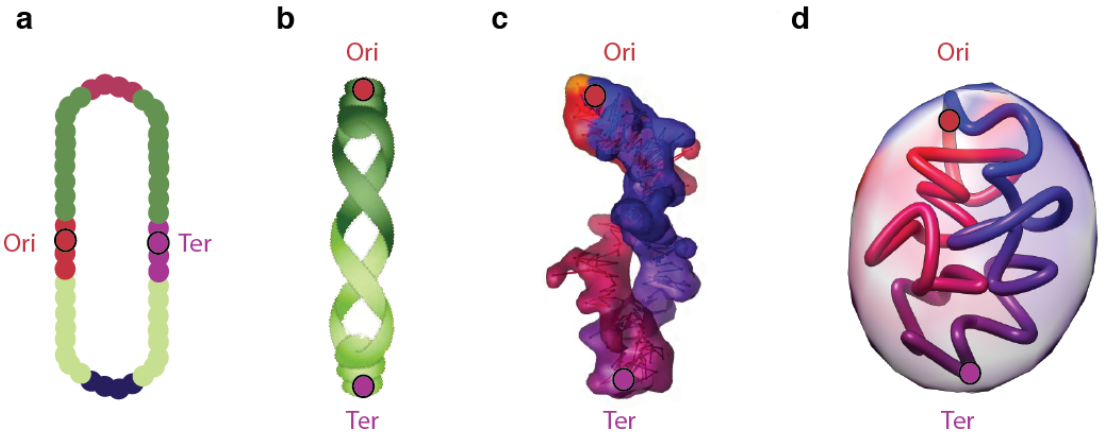

Figure 6: Models of bacterial chromosome organization. Models of nucleoid organization with Ori and Ter represented by red and purple circles **(a)** Model of *E. coli* genome with the four macro-domains Ori, Ter, Left, Right, represented by circles in red, purple, pink, nd blue respectively. **(b)** Model of *B. subtilis* genome adapted from [273]. **(c)** 3D models of *C. crescentus* genome conformation [162]. **(d)** 3D models of *M. pneumoniae* genome conformation.

Similarly they both have an origin proximal region parS (partition system) that assists in the orientation of the chromosome during replication [275, 276], whereas in *E. coli* Left and Right are situated toward the two poles and Ori and Ter are close to the middle of the cell [277, 278] (Fig. 6a). The chromosome organization of *C. crescentus* is similar to that of *M. pneumoniae,* with the Ori and Ter being localized at the two opposite poles (Fig. 6d), but in contrast has an ellipsoidal form with periodically arranged arms twisting around each other (Fig. 6c) [162].

Several published studies have previously shown that mammalian genomes are partitioned into the so-called topological associating domains (TADs) [19, 20], which range from 200 kb to 1 Mb and are conserved across different species and cell types. These studies suggest thus, that chromosomal domains could be a fundamental principle of genome organization. Our analysis allowed the detection of bacterial TAD-like domains (CIDs) for the first time in *M. pneumoniae*. Indeed, *M. pneumoniae* is partitioned into a total of 24 CIDs ranging from 15 kb to 65 kb, which are smaller to those reported in *C. crescentus* [113]. We also observed as previously reported in *C. crescentus* [113], that inhibition of supercoiling by novobiocin significantly reduced the sharpness and sizes of CIDs. Our finding suggests that supercoiling could be regulating domain formation in bacteria. Interestingly, our analysis also indicates that genes inside CIDs tend to be co-regulated, with lower co-expression levels between genes being detected at the domain boundaries. It was previously reported in *C. crescentus* that domain borders correlated with the presence of highly expressed genes, being as the

DNA is kept free of plectonemic loops by active transcription [113]. Although we did not observe such findings, we established that borders are characterized by low GC content levels. This could be related to the physical properties of DNA such as DNA curvature, which has previously been linked to the AT content [279]. Similarly, it has also been reported in *E. coli* and *Salmonella typhimurium* that the localization of domain loop boundaries is found in AT-rich regions [95]. Even though the contribution of NAP and SMC in the global genome organization was recently refuted in *E. coli* [167] as well as their role in the formation of CIDs in *C. crescentus* [113], the formation of these domains was consistent with the distribution of histone-like proteins binding sites, such as H-NS, HU, Fis and IHF [90, 95]. As *M. pneumoniae* however has few copies of the histone-like IHF protein [230], thereby making it difficult to maintain the CID boundaries, it is likely that additional factors contribute to the formation of such domain loops. Since *M. pneumoniae* only has a handful of DNA-binding proteins and very few TFs (Table 1), it is intriguing that it can establish a well-defined chromosome structure as well as maintain its CID boundaries. We speculate that very few factors may be necessary to define a 3D chromosome structure and provided evidence that other elements like supercoiling could regulate these domain boundaries, which are characterized by low GC-content and might be related to the physicochemical properties of DNA.

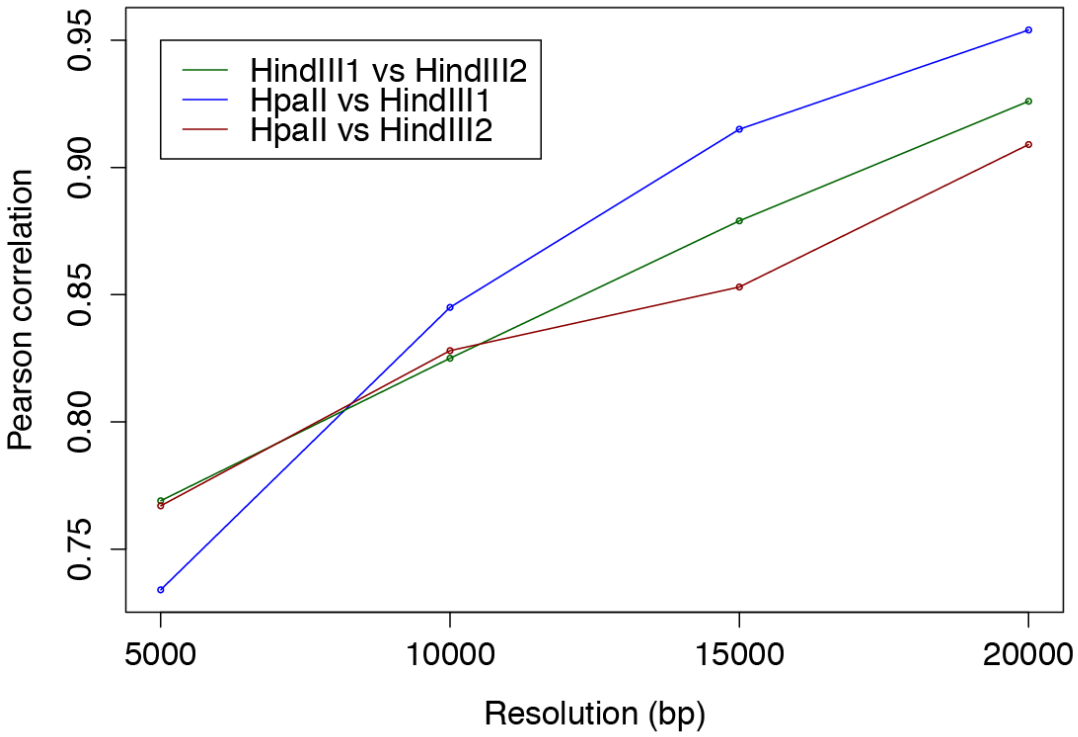

Figure S1: Pearson correlation between HpaII and HindIII datasets at different resolutions. Comparison of normalized and filtered Hi-C matrices of HindIII1, HindIII2 and HpaII datasets across 5, 10, 15 and 20 kb resolutions, shown in the x-axis, by Pearson correlation, shown in the y-axis.

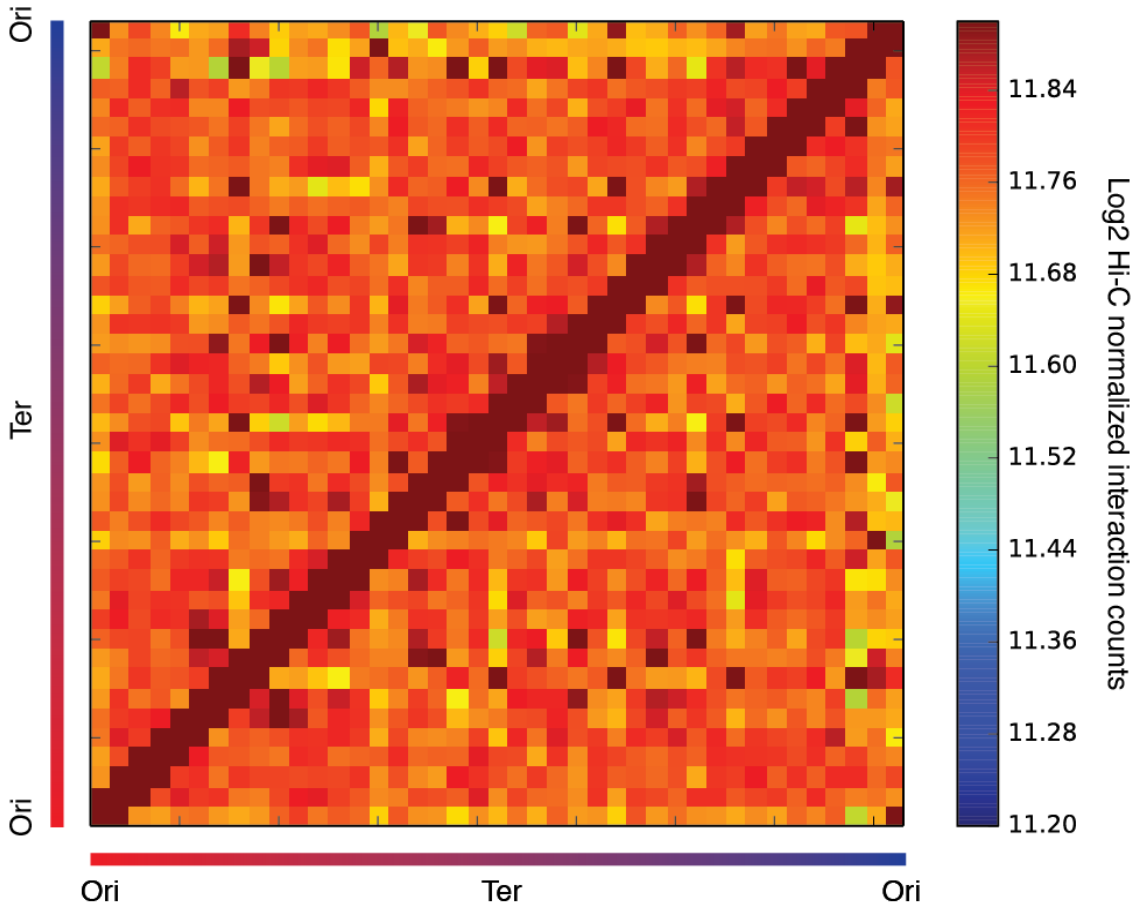

Figure S2: Normalized HpaII Hi-C contact map of *M. pneumoniae*, in exponential phase with 20 kb resolution. The frequency of interactions between a given pair of bins is found at the intersection of the row and column corresponding to those bins. The color of the contact map, from blue to red, indicates the log2 contact frequency. The bar underneath indicates genome position with Ori being located at a genome coordinate of 0 and Ter located at ~ 400 kb.

Resolution (bp)	MMP score
1000	0.191
3000	0.611
5000	0.652
8000	0.696
10000	0.708
15000	0.762
20000	0.802

Table S1: MMP score at different resolutions for the HpaII dataset. MMP score of normalized and filtered HpaII datasets was computed for the given resolutions.

Enzyme	HpaII	HindIII Rep.1	HindIII Rep.2
Total reads numbers	111,545,820	107,768,945	186,752,960
Final interaction numbers after hiclib filtering	3,289,717	1,361,843	3,035,901

Table S2: Hi-C datasets statistics. Total reads numbers and interaction numbers after filtering, obtained with hiclib Python library.

Primers	Annotation	Start position	End position	Genomic Sequence
F_Ori	Ori forward strand	1	26	TATTTACCGACGAAATTAATACCATC
R_Ori	Ori reverse strand	974	1000	TTTTGTTTTGACTAAAAGAGTTTGATC
F_90C	Right forward strand	204000	204020	TTGCACCAACTCCAGCAAGAC
R_90C	Right reverse strand	204974	205000	TGCTTGTCAATCATGTACTCAATTAAC
F_Ter	Ter forward strand	390000	390021	CGTAACATAAAAGAAGCACGTG
R_Ter	Ter reverse strand	390982	391000	GTTGTTTAGCGCGGGCTTC
F_270C	Left forward strand	612000	612016	CAAGCGCTCGCCTGGTC
R_270C	Left reverse strand	612971	613000	AATTTGAACAATTTCAACTAATTTATCAAC

Table S3: Regions of interest marked by FISH. Primer sequences and respective positions of the four regions marked by FISH.

Video S1: 3D model of the first cluster of *M. pneumoniae* genome models.

Video S2: 3D reconstruction of a *M. pneumoniae* cell from EM imaging.

3.6 Acknowledgments

We thank Dr. Besray Ünal and Dr. Ivan Junier for providing the co-expression data. We also thank Dr. Ivan Junier and Dr. Francois Serra for helpful suggestions and we are also grateful to Dr. Jae-Seong Yang for fruitful discussions. The research leading to these results was funding by the European Union Seventh Framework Programme (FP7/2007-2013 to L.S.), through the European Research Council, under grant agreement 232913 to L.S. and 609989 to M.A.M-R., the Fundación Botín to L.S., the Spanish Ministry of Economy and Competitiveness (BIO2007-61762 to L.S. and BFU2013–47736-P to M.A.M.-R.), the National Plan of R+D+i, the ISCIII - Subdirección General de Evaluación y Fomento de la Investigación- (PI10/01702 to L.S.), the Human Frontiers Science Program (RGP0044 to M.A.M.-R.), and the European Regional Development Fund (ERDF) to the ICREA Research Professor L.S. We acknowledge support from the Spanish Ministry of Economy and Competitiveness, 'Centro de Excelencia Severo Ochoa 2013-2017' (SEV-2012-0208). Library preparation and sequencing was done in the CRG Genomics Unit and high-resolution light microscopy analysis in the CRG microscopy unit.

3.7 Author contributions

M.T. performed super-resolution imaging experiments and EM experiments, collected and analyzed data and wrote the manuscript; E.Y. designed, optimized and performed Hi-C experiments, obtained the gene expression data and reviewed the manuscript; C.M. assisted with Hi-C experiments and performed the FISH for super-resolution imaging experiments; D.B. implemented the simulation of 3D models and reviewed the manuscript; Y.T. performed EM experiments; T.P designed a pipeline to analyze super-resolution images; S.K. performed initial Hi-C experiments and reviewed the manuscript; J.S performed 3D reconstruction from EM images; M.M. designed and supervised EM experiments and reviewed the manuscript; M.A.M.R. supervised the computational 3D modeling and reviewed the manuscript, M.L.S designed and supervised super-resolution imaging experiments and reviewed the manuscript and L.S. supervised the study and reviewed the manuscript.

Discussion

Here, our first objective was to evaluate 3D modeling of genome conformation using a RB mean-field approach, called TADbit. Such objective was essential for using this approach to determine the 3D structure of *M. pneumoniae* genome. After that, our second objective was to evaluate the impact of chromosome organization in the transcriptional regulation of a genome-reduced bacterium *M. pneumoniae*.

In the second chapter of the thesis, we used TB modeling to simulate population of "toy genomes" with different architecture, different resolution and different levels of heterogeneity within the populations. We then derived Hi-C contact matrices from those "toy genomes" population, adding increasing levels of noise that mimic Hi-C experimental artifacts and biases. Therefore, we generated the first dataset of simulated toy genome structures and their respective Hi-C matrices, for benchmarking RB modeling approaches, which comprises 168 simulated Hi-C matrices. Next, using TADbit, we determined the 3D genome structure based on those simulated Hi-C matrices to evaluate and compare the 3D models obtained to the original "toy genomes".

Resolution

The main conclusions are that RB mean-field modeling reproduces with accuracy the 3D genome structure and especially the TADbit protocol for 3D reconstruction and scoring function were validated. Interestingly, the accuracy of the reconstructed models is dependent on the resolution of the Hi-C map, as a result of the proportion of restraints per particle. As the sequencing depth experimentally defines the resolution, it is strongly recommended to increase the sequencing depth of Hi-C experiments that will result not only in higher-resolution models but also in models with higher overall accuracy.

Chromosomal architecture

Additionally, the comparison of Hi-C matrices of genomes with and without TAD-like

architecture revealed that having a genome partitioned into TAD resulted in more accurate reconstructed models. One could think that genomes with TAD architecture would result in a higher proportion of restraints per particle compared to genomes without TAD, as domains are defined as highly interacting. However we observed that this is not always the case as local interactions might also have high interactions, resulting in a higher number of restraints. A recent study identified chromatin loop using high-resolution matrices at 1kb, that often demarcate domains, which are anchored in regions associated to CTCF sites [141]. The fact that loop anchors occur at domain boundaries limits the possible conformations adopted for those loops and therefore the possible conformation of the domain at the boundary. A limited number of conformations would facilitate the 3D reconstruction of TAD, which further supports our findings.

Variability

In this study, we have considered two sources of variability with four levels of experimental noise as well as seven levels of structural heterogeneity between genomes over a population of cells. The first one was generated by varying the probability of capturing an interaction, which simulated Hi-C experimental artifacts affecting the detection of interacting fragments. The second source of variability was obtained by considering toy model genome populations of varying structural similarity. Although high levels of heterogeneity in the population alter the reconstruction of genome structure, TADbit is robust to high levels of experimental noise but sensitive to structural variability. Even if RB approaches are able to handle the simulations of different chromatin fiber properties compared to TB approaches, future optimization should be done to reproduce the heterogeneity of populations of cells, as for example non-synchronised cells where genomes would be at a different state of the cellular cycle. Population-based approaches [140] are addressing this limitation where sub-populations are representing the experimental variability, but the prior characterization of heterogeneity within Hi-C dataset remains an important challenge.

MMP score

In addition, relating the input Hi-C matrix to the corresponding accuracy of the resulting

models, allow us to compute a score, called MMP, to *a priori* evaluate any Hi-C matrix and predict its potential for accurate 3D modeling. Indeed this is the first evaluation that has been proposed to guide experimentalist as well as modelers in their choice of datasets. Up to now, the evaluation of Hi-C matrices was solely based on the correlation between technical replicates and between different restriction enzymes, when available. As an alternative, the MMP score provides a complementary characterization of Hi-C matrix and gives insight into the structural variability and experimental noise present in the matrix, evaluated by the skewness of the distribution. Additionally, the MMP score can guide the choice of resolution that is a critical step in the preprocessing of Hi-C datasets and not always obvious to distinguish between an adequate or sparse matrix at a given resolution that would result in accurate or inaccurate 3D models.

Summary

In summary, we proposed here the first framework to evaluate 3D modeling approaches, assessing how those approaches cope with heterogeneity and experimental noise. The outcomes of such evaluation is useful to understand the advantages and limitations of any restraint-based 3D modeling approaches, especially to guide experimentalist to *a priori* select Hi-C matrices that would result in accurate 3D models. The benchmarking of others modeling approaches using the same dataset would complement our conclusions based on TADbit modeling approach, and evaluate which modeling approaches are more appropriate according to the type of data, as for example single-cell 3C-based or heterogeneous Hi-C cell populations studies.

In the third chapter of the thesis, we selected *M. pneumoniae* as a model organism with small genome size and simplified regulatory network, to unveil the role of chromosome structure in transcriptional regulation. Indeed, compared to other bacteria, transcriptional regulation in *M. pneumoniae* appears simple, with a small number of DNA-binding proteins, few transcription factors and two sigma factors, leaving few alternatives for this the bacterium to modulate its transcriptional activity. Nevertheless, *M. pneumoniae* exhibits specific and complex transcriptional regulation in response to different stimuli [227, 229]. Understanding how this transcription is achieved in one of the smallest self-replicating genome is fundamental not only from the perspective of synthetic biology, but also to understand the evolution of gene regulatory circuits in "reduced genomes" and give insights in the fundamental requirements to control gene expression. Here, we focus on the role of chromosome structure to identify whether it can influence gene expression through changes in the local structure of the DNA or global rearrangements. To do so we determined *M. pneumoniae* 3D genome structure by combining electron microscopy, high-resolution light microscopy (STORM) and Hi-C. The resulting 3D models obtained with TADbit elucidated the global organization of the chromosome that present symmetry along the Ori–Ter axis with Ori and Ter located at opposite poles.

Cell Volume

Although DAPI staining of the chromosome revealed that *M. pneumoniae* does not have a defined nucleoid but rather the chromosome occupies the available volume [211], we estimated, by reconstruction of EM imaging, that the chromosome occupies about two thirds of the total cell volume. These differences can be explained by the limited resolution of DAPI imaging to accurately estimate the nucleoid occupancy. Moreover, the limited resolution of our models at 20kb could also explain this apparent shrinking of the chromosome dimensions, as the actual occupancy of a bin is not determined. The resolution was indeed set to 20kb to depict the global genome organization, limited by the total number of interactions obtained after filtering of Hi-C data. Finally, we cannot discard that the chromosome or cell volume changes between stationary and exponential phases. Since *M. pneumoniae* clumps in large aggregates in stationary phase, our experimental validation was done in exponential phase. We have experimentally calculated its cell volume, as well as measured distances between regions of the

chromosome by FISH in exponential phase, although the 3D modeling was done based on the Hi-C contact map in stationary. Indeed the analysis of the chromosome structure at exponential phase could be hampered by heterogeneity, as it is not possible to synchronize *M. pneumoniae*. However, the resulting exponential and stationary contact matrices significantly correlated (r=0.85) suggesting that the overall conformation of the chromosome does not significantly change between the two states, still the overall compaction could be different [57].

Validation by FISH

Using super-resolution imaging, we estimated by FISH the distances between the selected four quarters of the chromosome Ori, Right, Ter and Left and AO. Therefore, we deduced approximated distances between Ter-Ori, Ter-Right and Ter-Left to validate our 3D models of chromosome conformation, which resulted consistent with high-resolution fluorescent imaging. The distances estimated from 3D models obtained from Hi-C data in stationary phase were overall larger than the ones obtained from super-resolution imaging on cells in exponential phase. Indeed in stationary phase, there is less transcriptional activity and a corresponding relaxation of looped domain structure of the nucleoid [57]. Moreover, the limitation of super-resolution light microscopy imaging is that cells cannot be identified by phase contrast or bright field, which introduces a bias in the analysis towards short distances, compared to larger distances that might be discarded as they could belong to two different cells. As the 3D genome models are scaled according to the distances obtained by super-resolution imaging, the dimensions of the chromosome could also appear smaller than they are. This would also explain the difference previously mentioned between chromosome volume estimated from the model and DAPI staining.

Cell division

Additionally, we observed that Ori-AO measurements have a larger variability compared to Ter-AO, which would suggest that after duplication the Ori move toward the opposite pole whereas the unduplicated Ter remains located throughout the replication process. This dynamics of Ori and Ter were also observed in *B. subtilis* replication [127, 271].

In analogous species *Mycoplasma gallisepticum* [219], attachment organelles were found enriched for newly synthesized DNA, suggesting a possible interaction between DNA and AO during cell division. A previous study suggested that the migration of the AO to the opposite pole of the cell is coordinated with DNA replication [204]. One possibility is that the AO is anchored to the Ori and the AO is driving the new replicated DNA towards the opposite pole during its migration. Another possibility is that the AO is anchored to the Ter and once the DNA is fully replicated, the AO migrates toward the opposite pole to efficiently segregate the new replicated DNA from the old one. However, although imaging indicates that the Ter locus is the closest loci to the AO, the observed variability for Ori-AO and Ter-AO could not demonstrate that the AO is specifically attached to a chromosome region as division occurs. To address this question it would be interesting to track the Ori and Ter regions with respect to the AO, during DNA replication, to understand whether the migration of the nascent AO to the opposite pole is helping in DNA segregation by gliding motility and how such mechanism happens. Unfortunately the FISH protocol only allows studying fixed cells and as a consequence, we were not able to investigate in details the dynamic relationship between chromosome orientation and cell division in *M. pneumoniae.*

Comparison with other bacteria

The interaction map of *M. pneumoniae* is comparable to that of the phylogenetically closer related gram-positive bacterium *B. subtilis* with a double diagonal, suggesting a similar genome organization [166]. In fact, imaging of *B. subtilis* as well as *C. crescentus* showed that chromosomes are arranged linearly with Ori and Ter having preference for opposite poles [86, 127, 272, 273]. Similarly they both have an origin proximal region parS that assists the orientation of the chromosome during replication. Whereas in *E. coli*, Ori and Ter are located close to the middle of the cell and no double diagonal were observed in the *E. coli* contact map [128, 129].

C. crescentus genome structure was resolved combining similar approaches of 5C with light microscopy imaging [162] and its genome shares similar organization with *M. pneumoniae* but has an ellipsoidal form with periodically arranged arms that twist around each others. In conclusion, the linear ordering of loci seems to be a common principle in bacteria but the orientation of bacterial genome within the cell differs. This could be related to the different manners in which chromosome replication takes place

in different bacteria [162, 167].

Resolution

The fact that we could not model at higher resolution is mainly related to the sequencing depth and restriction enzymes used, but it could also be due to structural heterogeneity or experimental noise. Interestingly, as described in the first chapter, the skewness of the Hi-C matrices distribution allowed us to differentiate between matrices rich in experimental noise from those high in structural variability. The positive skewness obtained from HpaII dataset of 2.4 might suggest that the Hi-C matrix is richer in structural variability than experimental noise. Indeed positive skewness matrices with a long negative tail are likely to be obtained from a population of cells with large structural variability. However, we obtained a high MMP score of 0.8, which predicts a good potential for the 3D reconstruction of this HpaII matrix at 20kb.

Transcriptional domains

Several published studies have previously shown that mammalian genomes are partitioned into TADs [19, 20], which range from 200 kb to 1 Mb and are conserved across different species and cell types. Analysis of local chromatin structures revealed that *M. pneumoniae* is organized into 24 CIDs ranging from 15 kb to 65 kb, which are smaller than the CIDs reported in *C. crescentus* [113]. Those findings suggest that domains are a fundamental principle of genome organization. We further demonstrated that those domains are constituted by genes that tend to co-regulate and lower co-expression levels are found at the domains boundaries. Our findings indicate that physical clustering within CIDs may be used to coordinate gene expression.

It was previously reported for *C. crescentus* bacteria that domain borders correlated with the presence of highly expressed genes, where the DNA is kept free of plectonemic loops by active transcription [113]. Although we did not observe such findings, we established that a low GC content level characterizes borders. This could be related to physical properties of DNA such as DNA curvature, which has previously been linked the percentage of AT content [279]. Similarly, it has also been reported in *E. coli* and *S. typhimurium,* that the domains loops boundaries are found in AT-rich regions [90, 95]. Even though the contribution of NAP and SMC in the global genome organization was

recently refuted in *E. coli* [167], as well as their role in the formation of CIDs in *C. crescentus* [113], those domains formation were related to the distribution of NAPs [90, 95]. Nevertheless *M. pneumoniae* has few NAPs and few TFs, which make it difficult to maintain the CIDs boundaries and suggests that additional factors should contribute to the formation of those domain loops.

Topological domains are also thought to be dynamic with domains boundaries that would be formed as the result of transcription, translation or replication, without any particular protein binding [29]. Therefore it has been proposed that the local chromatin structure could be organized by transiently associated factors such as the coupled transcription and translation of membrane proteins, called transertion [21-23] or the transcription of some highly active promoters, in the absence of membrane translocation [110, 111]. Following this hypothesis, small dynamic domains that are non-specific would allow the chromosome to be compacted to fit within the cell and at the same permit efficient transcription and replication. In that sense, the chromosome will be structured without imposing rigidity [29].

In addition, although the global genome organization does not change with inhibition of supercoiling by novobiocin, we observed as previously reported in *C. crescentus* [40] that the sharpness and positions of CIDs were significantly reduced. Our results indicate that supercoiling is related to the local chromatin structure and could be regulating those domains formation in bacteria. Indeed, negative supercoiling is forming plectonemes loops that are maintained by gyrases and topoisomerases [27-29] to prevent relaxation of the entire chromosome. Nevertheless some domains borders remain intact after novobiocin treatment, indicating that domain formation might arise not only from the supercoiling but from a combination of the factors previously mentioned.

Furthermore, Hi-C data reflects the averaged dynamics of a population of millions of cells and it is unclear what does a CID represent in a single cell. Domains would appear randomly distributed over a population of cells, even though they might have some sequence specificity. Super-resolution imaging would assess whether CIDs are stable and present in every cell within a population or whether domains are randomly formed and vary across a cell population. Such analysis has been done in eukaryotes using fluorescent probes spanning several hundreds of kilobases across TADs and revealed that they do differ in size and degree of clustering from one cell to another [20]. More recently, using 3D FISH in mouse, the chromatin conformation within a TAD was

revealed to be highly variable, though not random and they proposed that structural fluctuations within TADs contribute to transcriptional variability by stochastically modulating interactions between regulatory sequences [156]. Future work of super-resolution imaging of CIDs could estimate how variable are CIDs in prokaryotic cell compared to the previous results of TADs in eukaryotes.

Summary

In summary, this study expands the current understanding of bacterial genome organization. We defined here fundamental principles of genome organization with the partition of a reduced genome into domains and provide evidence that genes inside CIDs tend to be co-regulated, indicating that the chromosome structure has a role in transcriptional regulation by defining the limits of regulatory neighborhood. One of the smallest replicating bacterium that has few DNA-binding proteins can establish a defined chromosomal structure, as well as maintain the CIDs boundaries. We speculate that few factors may be necessary to determine a 3D chromosome structure and we provided evidence that other element like supercoiling could be regulating those domains boundaries. Although we have shown that CIDs have a role in local chromatin folding and transcriptional regulation, it will be interesting to know whether disruption or deletion of CIDs boundaries are accompanied by long-range transcription changes, similarly to the deletion of TADs boundaries in eukaryotes [20]. Additionally, estimating whether CIDs are stable in every cell or variable across the cell population and compare it to the TADs variability observed in eukaryotes would determine whether similar mechanisms are determinants of domains formation in bacteria and eukaryotes. Future work will clarify the mechanism underlying those domains formation and to what extent they contribute to the transcriptional regulation.

References

1. Jackson, D.A., et al., *Visualization of focal sites of transcription within human nuclei.* The EMBO Journal, 1993. **12**(3): p. 1059-1065.
2. Jackson, D.A., et al., *Numbers and Organization of RNA Polymerases, Nascent Transcripts, and Transcription Units in HeLa Nuclei.* Molecular Biology of the Cell, 1998. **9**(6): p. 1523-1536.
3. Hozak, P., et al., *Visualization of replication factories attached to nucleoskeleton.* Cell, 1993. **73**(2): p. 361-73.
4. Cook, P.R., *The organization of replication and transcription.* Science, 1999. **284**(5421): p. 1790-5.
5. Lewis, P.J., S.D. Thaker, and J. Errington, *Compartmentalization of transcription and translation in Bacillus subtilis.* EMBO J, 2000. **19**(4): p. 710-8.
6. Jin, D.J. and J.E. Cabrera, *Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in Escherichia coli.* J Struct Biol, 2006. **156**(2): p. 284-91.
7. Jin, D.J., C. Cagliero, and Y.N. Zhou, *Role of RNA polymerase and transcription in the organization of the bacterial nucleoid.* Chem Rev, 2013. **113**(11): p. 8662-82.
8. Endesfelder, U., et al., *Multiscale spatial organization of RNA polymerase in Escherichia coli.* Biophys J, 2013. **105**(1): p. 172-81.
9. Sproul, D., N. Gilbert, and W.A. Bickmore, *The role of chromatin structure in regulating the expression of clustered genes.* Nat Rev Genet, 2005. **6**(10): p. 775-781.
10. Carpentier, A.S., et al., *Decoding the nucleoid organisation of Bacillus subtilis and Escherichia coli through gene expression data.* BMC Genomics, 2005. **6**: p. 84.
11. Sexton, T., et al., *Gene regulation through nuclear organization.* Nat Struct Mol Biol, 2007. **14**(11): p. 1049-1055.
12. Janga, S.C., H. Salgado, and A. Martínez-Antonio, *Transcriptional regulation shapes the organization of genes on bacterial chromosomes.* Nucleic Acids Research, 2009. **37**(11): p. 3680-3688.
13. Junier, I., J. Herisson, and F. Kepes, *Genomic organization of evolutionarily correlated genes in bacteria: limits and strategies.* J Mol Biol, 2012. **419**(5): p. 369-86.
14. Dorman, C.J., *Genome architecture and global gene regulation in bacteria: making progress towards a unified model?* Nat Rev Microbiol, 2013. **11**(5): p. 349-55.
15. Stancheva, I. and E.C. Schirmer, *Nuclear envelope: connecting structural genome organization to regulation of gene expression.* Adv Exp Med Biol, 2014. **773**: p. 209-44.
16. Cremer, T. and C. Cremer, *Chromosome territories, nuclear architecture and gene regulation in mammalian cells.* Nat Rev Genet, 2001. **2**(4): p. 292-301.
17. Lieberman-Aiden, E., et al., *Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.* Science, 2009. **326**(5950): p. 289-293.

18. Cagliero, C., Y.N. Zhou, and D.J. Jin, *Spatial organization of transcription machinery and its segregation from the replisome in fast-growing bacterial cells.* Nucleic Acids Res, 2014. **42**(22): p. 13696-705.
19. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-380.
20. Nora, E.P., et al., *Spatial partitioning of the regulatory landscape of the X-inactivation centre.* Nature, 2012. **485**(7398): p. 381-385.
21. Liu, L.F. and J.C. Wang, *Supercoiling of the DNA template during transcription.* Proc Natl Acad Sci U S A, 1987. **84**(20): p. 7024-7.
22. Lynch, A.S. and J.C. Wang, *Anchoring of DNA to the bacterial cytoplasmic membrane through cotranscriptional synthesis of polypeptides encoding membrane proteins or proteins for export: a mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I.* Journal of Bacteriology, 1993. **175**(6): p. 1645-1655.
23. Norris, V., *Hypothesis: chromosome separation in Escherichia coli involves autocatalytic gene expression, transertion and membrane-domain formation.* Mol Microbiol, 1995. **16**(6): p. 1051-7.
24. Vologodskii, A.V. and N.R. Cozzarelli, *Conformational and thermodynamic properties of supercoiled DNA.* Annu Rev Biophys Biomol Struct, 1994. **23**: p. 609-43.
25. Odijk, T., *Osmotic compaction of supercoiled DNA into a bacterial nucleoid.* Biophys Chem, 1998. **73**(1-2): p. 23-9.
26. Drlica, K., *Control of bacterial DNA supercoiling.* Mol Microbiol, 1992. **6**(4): p. 425-33.
27. Zechiedrich, E.L., et al., *Roles of topoisomerases in maintaining steady-state DNA supercoiling in Escherichia coli.* J Biol Chem, 2000. **275**(11): p. 8103-13.
28. Staczek, P. and N.P. Higgins, *Gyrase and Topo IV modulate chromosome domain size in vivo.* Molecular Microbiology, 1998. **29**(6): p. 1435-1448.
29. Postow, L., et al., *Topological domain structure of the Escherichia coli chromosome.* Genes & Development, 2004. **18**(14): p. 1766-1779.
30. Dorman, C.J. and C.P. Corcoran, *Bacterial DNA topology and infectious disease.* Nucleic Acids Research, 2009. **37**(3): p. 672-678.
31. Baranello, L., et al., *The importance of being supercoiled: how DNA mechanics regulate dynamic processes.* Biochim Biophys Acta, 2012. **1819**(7): p. 632-8.
32. Cook, P.R., *A chromomeric model for nuclear and chromosome structure.* J Cell Sci, 1995. **108 (Pt 9)**: p. 2927-35.
33. Wu, H.Y., et al., *Transcription generates positively and negatively supercoiled domains in the template.* Cell, 1988. **53**(3): p. 433-40.
34. Menzel, R. and M. Gellert, *Regulation of the genes for E. coli DNA gyrase: homeostatic control of DNA supercoiling.* Cell, 1983. **34**(1): p. 105-13.
35. Tse-Dinh, Y.C., *Regulation of the Escherichia coli DNA topoisomerase I gene by DNA supercoiling.* Nucleic Acids Res, 1985. **13**(13): p. 4751-63.
36. Kouzine, F., et al., *The functional response of upstream DNA to dynamic supercoiling in vivo.* Nat Struct Mol Biol, 2008. **15**(2): p. 146-54.
37. Seila, A.C., et al., *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.
38. Travers, A. and G. Muskhelishvili, *A common topology for bacterial and eukaryotic transcription initiation?* EMBO Rep, 2007. **8**(2): p. 147-51.

39. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
40. Felsenfeld, G., D. Clark, and V. Studitsky, *Transcription through nucleosomes.* Biophys Chem, 2000. **86**(2-3): p. 231-7.
41. Bowman, G.D., *Mechanisms of ATP-dependent nucleosome sliding.* Curr Opin Struct Biol, 2010. **20**(1): p. 73-81.
42. Zimmerman, S.B., *Macromolecular crowding effects on macromolecular interactions: some implications for genome structure and function.* Biochim Biophys Acta, 1993. **1216**(2): p. 175-85.
43. Hancock, R., *A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus.* J Struct Biol, 2004. **146**(3): p. 281-90.
44. Jun, S. and B. Mulder, *Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome.* Proc Natl Acad Sci U S A, 2006. **103**(33): p. 12388-93.
45. Cunha, S., C.L. Woldringh, and T. Odijk, *Polymer-mediated compaction and internal dynamics of isolated Escherichia coli nucleoids.* J Struct Biol, 2001. **136**(1): p. 53-66.
46. Sunkara, P.S., C.C. Chang, and N.J. Prakash, *Role of polyamines during chromosome condensation of mammalian cells.* Cell Biol Int Rep, 1983. **7**(6): p. 455-65.
47. Gosule, L.C. and J.A. Schellman, *DNA condensation with polyamines I. Spectroscopic studies.* J Mol Biol, 1978. **121**(3): p. 311-26.
48. Becker, M., et al., *Spermine-DNA complexes build up metastable structures. Small-angle X-ray scattering and circular dichroism studies.* Nucleic Acids Res, 1979. **7**(5): p. 1297-309.
49. Tabor, C.W. and H. Tabor, *Polyamines in microorganisms.* Microbiol Rev, 1985. **49**(1): p. 81-99.
50. Thomas, T.J. and T. Thomas, *Polyamine-induced Z-DNA conformation in plasmids containing (dA-dC)n.(dG-dT)n inserts and increased binding of lupus autoantibodies to the Z-DNA form of plasmids.* Biochem J, 1994. **298**(Pt 2): p. 485-91.
51. Feuerstein, B.G., N. Pattabiraman, and L.J. Marton, *Spermine-DNA interactions: a theoretical study.* Proc Natl Acad Sci U S A, 1986. **83**(16): p. 5948-52.
52. Panagiotidis, C.A., et al., *Polyamines alter sequence-specific DNA-protein interactions.* Nucleic Acids Res, 1995. **23**(10): p. 1800-9.
53. Pelta, J., F. Livolant, and J.L. Sikorav, *DNA aggregation induced by polyamines and cobalthexamine.* J Biol Chem, 1996. **271**(10): p. 5656-62.
54. Saminathan, M., et al., *Ionic and structural specificity effects of natural and synthetic polyamines on the aggregation and resolubilization of single-, double-, and triple-stranded DNA.* Biochemistry, 1999. **38**(12): p. 3821-30.
55. Graumann, P.L. and T. Knust, *Dynamics of the bacterial SMC complex and SMC-like proteins involved in DNA repair.* Chromosome Res, 2009. **17**(2): p. 265-75.
56. Browning, D.F., D.C. Grainger, and S.J. Busby, *Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression.* Curr Opin Microbiol, 2010. **13**(6): p. 773-80.

57. Dillon, S.C. and C.J. Dorman, *Bacterial nucleoid-associated proteins, nucleoid structure and gene expression.* Nat Rev Microbiol, 2010. **8**(3): p. 185-95.
58. Drlica, K. and J. Rouviere-Yaniv, *Histonelike proteins of bacteria.* Microbiol Rev, 1987. **51**(3): p. 301-19.
59. Dorman, C.J. and P. Deighan, *Regulation of gene expression by histone-like proteins in bacteria.* Curr Opin Genet Dev, 2003. **13**(2): p. 179-84.
60. Afflerbach, H., O. Schröder, and R. Wagner, *Effects of the Escherichia coli DNA-binding protein H-NS on rRNA synthesis in vivo.* Mol Microbiol, 1998. **28**(3): p. 641-53.
61. Ali Azam, T., et al., *Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid.* J Bacteriol, 1999. **181**(20): p. 6361-70.
62. Bahloul, A., F. Boubrik, and J. Rouviere-Yaniv, *Roles of Escherichia coli histone-like protein HU in DNA replication: HU-beta suppresses the thermosensitivity of dnaA46ts.* Biochimie, 2001. **83**(2): p. 219-29.
63. Ussery, D., et al., *Genome organisation and chromatin structure in Escherichia coli.* Biochimie, 2001. **83**(2): p. 201-12.
64. Ali, B.M., et al., *Compaction of single DNA molecules induced by binding of integration host factor (IHF).* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10658-63.
65. Schröder, O. and R. Wagner, *The bacterial regulatory protein H-NS--a versatile modulator of nucleic acid structures.* Biol Chem, 2002. **383**(6): p. 945-60.
66. Shimizu, M., et al., *Characterization of the binding of HU and IHF, homologous histone-like proteins of Escherichia coli, to curved and uncurved DNA.* Biochim Biophys Acta, 1995. **1264**(3): p. 330-6.
67. Vis, H., et al., *Solution structure of the HU protein from Bacillus stearothermophilus.* J Mol Biol, 1995. **254**(4): p. 692-703.
68. van Noort, J., et al., *Dual architectural roles of HU: formation of flexible hinges and rigid filaments.* Proc Natl Acad Sci U S A, 2004. **101**(18): p. 6969-74.
69. Guo, F. and S. Adhya, *Spiral structure of Escherichia coli HUalphabeta provides foundation for DNA supercoiling.* Proc Natl Acad Sci U S A, 2007. **104**(11): p. 4309-14.
70. Kamashev, D., et al., *HU binds and folds single-stranded DNA.* Nucleic Acids Res, 2008. **36**(3): p. 1026-36.
71. Skoko, D., et al., *Mechanism of chromosome compaction and looping by the Escherichia coli nucleoid protein Fis.* J Mol Biol, 2006. **364**(4): p. 777-98.
72. Cho, B.K., et al., *Genome-wide analysis of Fis binding in Escherichia coli indicates a causative role for A-/AT-tracts.* Genome Res, 2008. **18**(6): p. 900-10.
73. Swinger, K.K. and P.A. Rice, *IHF and HU: flexible architects of bent DNA.* Curr Opin Struct Biol, 2004. **14**(1): p. 28-35.
74. Mumm, J.P., A. Landy, and J. Gelles, *Viewing single lambda site-specific recombination events from start to finish.* EMBO J, 2006. **25**(19): p. 4586-95.
75. Dame, R.T., M.C. Noom, and G.J. Wuite, *Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation.* Nature, 2006. **444**(7117): p. 387-90.
76. Tapias, A., G. López, and S. Ayora, *Bacillus subtilis LrpC is a sequence-independent DNA-binding and DNA-bending protein which bridges DNA.* Nucleic Acids Res, 2000. **28**(2): p. 552-9.

77. de los Rios, S. and J.J. Perona, *Structure of the Escherichia coli leucine-responsive regulatory protein Lrp reveals a novel octameric assembly.* J Mol Biol, 2007. **366**(5): p. 1589-602.
78. Niki, H., et al., *The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of E. coli.* EMBO J, 1991. **10**(1): p. 183-93.
79. Britton, R.A., D.C. Lin, and A.D. Grossman, *Characterization of a prokaryotic SMC protein involved in chromosome partitioning.* Genes Dev, 1998. **12**(9): p. 1254-9.
80. Almirón, M., et al., *A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli.* Genes Dev, 1992. **6**(12B): p. 2646-54.
81. Zhang, A., et al., *Escherichia coli protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics.* EMBO J, 1996. **15**(6): p. 1340-9.
82. Chenoweth, M.R. and S. Wickner, *Complex regulation of the DnaJ homolog CbpA by the global regulators sigmaS and Lrp, by the specific inhibitor CbpM, and by the proteolytic degradation of CbpM.* J Bacteriol, 2008. **190**(15): p. 5153-61.
83. Azam, T.A. and A. Ishihama, *Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity.* J Biol Chem, 1999. **274**(46): p. 33105-13.
84. Riley, S.P., et al., *Borrelia burgdorferi EbfC defines a newly-identified, widespread family of bacterial DNA-binding proteins.* Nucleic Acids Res, 2009. **37**(6): p. 1973-83.
85. Castang, S., et al., *H-NS family members function coordinately in an opportunistic pathogen.* Proc Natl Acad Sci U S A, 2008. **105**(48): p. 18947-52.
86. Webb, C.D., et al., *Bipolar localization of the replication origin regions of chromosomes in vegetative and sporulating cells of B. subtilis.* Cell, 1997. **88**(5): p. 667-74.
87. Mukherjee, A., P.J. DiMario, and A. Grove, *Mycobacterium smegmatis histone-like protein Hlp is nucleoid associated.* FEMS Microbiol Lett, 2009. **291**(2): p. 232-40.
88. Morikawa, K., et al., *Bacterial nucleoid dynamics: oxidative stress response in Staphylococcus aureus.* Genes Cells, 2006. **11**(4): p. 409-23.
89. Dame, R.T., et al., *Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the rrnB P1.* J Biol Chem, 2002. **277**(3): p. 2146-50.
90. Grainger, D.C., et al., *Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome.* Nucleic Acids Res, 2006. **34**(16): p. 4642-52.
91. Schmid, M.B., *More than just "histone-like" proteins.* Cell, 1990. **63**(3): p. 451-3.
92. Nozaki, S., Y. Yamada, and T. Ogawa, *Initiator titration complex formed at datA with the aid of IHF regulates replication timing in Escherichia coli.* Genes Cells, 2009. **14**(3): p. 329-41.
93. Arfin, S.M., et al., *Global gene expression profiling in Escherichia coli K12. The effects of integration host factor.* J Biol Chem, 2000. **275**(38): p. 29672-84.

94. Shao, Y., L.S. Feldman-Cohen, and R. Osuna, *Functional characterization of the Escherichia coli Fis-DNA binding sequence.* J Mol Biol, 2008. **376**(3): p. 771-85.
95. Noom, M.C., et al., *H-NS promotes looped domain formation in the bacterial chromosome.* Curr Biol, 2007. **17**(21): p. R913-4.
96. Gordon, B.R., et al., *Lsr2 of Mycobacterium represents a novel class of H-NS-like proteins.* J Bacteriol, 2008. **190**(21): p. 7052-9.
97. Nye, M.B. and R.K. Taylor, *Vibrio cholerae H-NS domain structure and function with respect to transcriptional repression of ToxR regulon genes reveals differences among H-NS family members.* Mol Microbiol, 2003. **50**(2): p. 427-44.
98. Afflerbach, H., O. Schröder, and R. Wagner, *Conformational changes of the upstream DNA mediated by H-NS and FIS regulate E. coli RrnB P1 promoter activity.* J Mol Biol, 1999. **286**(2): p. 339-53.
99. Pul, U., et al., *LRP and H-NS--cooperative partners for transcription regulation at Escherichia coli rRNA promoters.* Mol Microbiol, 2005. **58**(3): p. 864-76.
100. Dame, R.T., *The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin.* Mol Microbiol, 2005. **56**(4): p. 858-70.
101. Hirano, T., *At the heart of the chromosome: SMC proteins in action.* Nat Rev Mol Cell Biol, 2006. **7**(5): p. 311-22.
102. Hirano, M. and T. Hirano, *Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions.* Mol Cell, 2006. **21**(2): p. 175-86.
103. Sullivan, N.L., K.A. Marquis, and D.Z. Rudner, *Recruitment of SMC by ParB-parS organizes the origin region and promotes efficient chromosome segregation.* Cell, 2009. **137**(4): p. 697-707.
104. Minnen, A., et al., *SMC is recruited to oriC by ParB and promotes chromosome segregation in Streptococcus pneumoniae.* Mol Microbiol, 2011. **81**(3): p. 676-88.
105. Travers, A. and G. Muskhelishvili, *Bacterial chromatin.* Curr Opin Genet Dev, 2005. **15**(5): p. 507-14.
106. Broyles, S.S. and D.E. Pettijohn, *Interaction of the Escherichia coli HU protein with DNA. Evidence for formation of nucleosome-like structures with altered DNA helical pitch.* J Mol Biol, 1986. **187**(1): p. 47-60.
107. Hsu, Y.H., M.W. Chung, and T.K. Li, *Distribution of gyrase and topoisomerase IV on bacterial nucleoid: implications for nucleoid organization.* Nucleic Acids Res, 2006. **34**(10): p. 3128-38.
108. Hardy, C.D. and N.R. Cozzarelli, *A genetic selection for supercoiling mutants of Escherichia coli reveals proteins implicated in chromosome structure.* Mol Microbiol, 2005. **57**(6): p. 1636-52.
109. Zimmerman, S.B., *Cooperative transitions of isolated Escherichia coli nucleoids: implications for the nucleoid as a cellular phase.* J Struct Biol, 2006. **153**(2): p. 160-75.
110. Scheirer, K.E. and N.P. Higgins, *Transcription induces a supercoil domain barrier in bacteriophage Mu.* Biochimie, 2001. **83**(2): p. 155-9.
111. Deng, S., R.A. Stein, and N.P. Higgins, *Transcription-induced barriers to supercoil diffusion in the Salmonella typhimurium chromosome.* Proc Natl Acad Sci U S A, 2004. **101**(10): p. 3398-403.

112. Cabrera, J.E. and D.J. Jin, *The distribution of RNA polymerase in Escherichia coli is dynamic and sensitive to environmental cues.* Mol Microbiol, 2003. **50**(5): p. 1493-505.
113. Le, T.B.K., et al., *High-resolution mapping of the spatial organization of a bacterial chromosome.* Science (New York, N.Y.), 2013. **342**(6159): p. 731-734.
114. Johnson, R.C., et al., *Major Nucleoid Proteins in the Structure and Function of the Escherichia coli Chromosome*, in *The Bacterial Chromosome*. 2005, American Society of Microbiology.
115. Fritsche, M., et al., *A model for Escherichia coli chromosome packaging supports transcription factor-induced DNA domain formation.* Nucleic Acids Research, 2012. **40**(3): p. 972-980.
116. Woldringh, C.L., *The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation.* Mol Microbiol, 2002. **45**(1): p. 17-29.
117. Hatfield, G.W. and C.J. Benham, *DNA topology-mediated control of global gene expression in Escherichia coli.* Annu Rev Genet, 2002. **36**: p. 175-203.
118. Peter, B.J., et al., *Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli.* Genome Biol, 2004. **5**(11): p. R87.
119. Dorman, C.J., *DNA supercoiling and bacterial gene expression.* Sci Prog, 2006. **89**(Pt 3-4): p. 151-66.
120. Cameron, A.D., D.M. Stoebel, and C.J. Dorman, *DNA supercoiling is differentially regulated by environmental factors and FIS in Escherichia coli and Salmonella enterica.* Mol Microbiol, 2011. **80**(1): p. 85-101.
121. Zhang, W. and J.B. Baseman, *Transcriptional regulation of MG_149, an osmoinducible lipoprotein gene from Mycoplasma genitalium.* Molecular Microbiology, 2011. **81**(2): p. 327-339.
122. Herrmann, R. and B. Reiner, *Mycoplasma pneumoniae and Mycoplasma genitalium: a comparison of two closely related bacterial species.* Curr Opin Microbiol, 1998. **1**(5): p. 572-9.
123. Mouw, K.W. and P.A. Rice, *Shaping the Borrelia burgdorferi genome: crystal structure and binding properties of the DNA-bending protein Hbb.* Mol Microbiol, 2007. **63**(5): p. 1319-30.
124. Sluijter, M., et al., *The Mycoplasma pneumoniae MPN229 gene encodes a protein that selectively binds single-stranded DNA and stimulates Recombinase A-mediated DNA strand exchange.* BMC Microbiol, 2008. **8**: p. 167.
125. Das, D., et al., *Crystal structure of a novel single-stranded DNA binding protein from Mycoplasma pneumoniae.* Proteins, 2007. **67**(3): p. 776-82.
126. Viollier, P.H., et al., *Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(25): p. 9257-9262.
127. Teleman, A.A., et al., *Chromosome arrangement within a bacterium.* Curr Biol, 1998. **8**(20): p. 1102-9.
128. Wang, X., et al., *The two Escherichia coli chromosome arms locate to separate cell halves.* Genes Dev, 2006. **20**(13): p. 1727-31.

129. Wiggins, P.A., et al., *Strong intranucleoid interactions organize the Escherichia coli chromosome into a nucleoid filament.* Proc Natl Acad Sci U S A, 2010. **107**(11): p. 4991-5.
130. Valens, M., et al., *Macrodomain organization of the Escherichia coli chromosome.* The EMBO Journal, 2004. **23**(21): p. 4330-4341.
131. Espeli, O., R. Mercier, and F. Boccard, *DNA dynamics vary according to macrodomain topography in the E. coli chromosome.* Mol Microbiol, 2008. **68**(6): p. 1418-27.
132. Dekker, J., et al., *Capturing chromosome conformation.* Science, 2002. **295**(5558): p. 1306-11.
133. van Steensel, B. and J. Dekker, *Genomics tools for unraveling chromosome architecture.* Nat Biotechnol, 2010. **28**(10): p. 1089-1095.
134. Belton, J.M., et al., *Hi-C: a comprehensive technique to capture the conformation of genomes.* Methods, 2012. **58**(3): p. 268-76.
135. Vora, T., A.K. Hottes, and S. Tavazoie, *Protein occupancy landscape of a bacterial genome.* Mol Cell, 2009. **35**(2): p. 247-53.
136. Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).* Nat Genet, 2006. **38**(11): p. 1348-54.
137. Zhao, Z., et al., *Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions.* Nat Genet, 2006. **38**(11): p. 1341-7.
138. Dostie, J., et al., *Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements.* Genome Res, 2006. **16**(10): p. 1299-309.
139. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.
140. Kalhor, R., et al., *Genome architectures revealed by tethered chromosome conformation capture and population-based modeling.* Nat Biotechnol, 2012. **30**(1): p. 90-8.
141. Rao, S.S., et al., *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.* Cell, 2014. **159**(7): p. 1665-80.
142. Fullwood, M.J., et al., *An oestrogen-receptor-alpha-bound human chromatin interactome.* Nature, 2009. **462**(7269): p. 58-64.
143. Hughes, J.R., et al., *Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment.* Nat Genet, 2014. **46**(2): p. 205-12.
144. Kolovos, P., et al., *Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements.* Epigenetics Chromatin, 2014. **7**: p. 10.
145. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Meth, 2012. **9**(4): p. 357-359.
146. Marco-Sola, S., et al., *The GEM mapper: fast, accurate and versatile alignment by filtration.* Nat Methods, 2012. **9**(12): p. 1185-8.
147. Li, H., J. Ruan, and R. Durbin, *Mapping short DNA sequencing reads and calling variants using mapping quality scores.* Genome Res, 2008. **18**(11): p. 1851-8.

148. Imakaev, M., et al., *Iterative correction of Hi-C data reveals hallmarks of chromosome organization.* Nat Meth, 2012. **9**(10): p. 999-1003.
149. Yaffe, E. and A. Tanay, *Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture.* Nat Genet, 2011. **43**(11): p. 1059-65.
150. Hu, M., et al., *HiCNorm: removing biases in Hi-C data via Poisson regression.* Bioinformatics, 2012. **28**(23): p. 3131-3.
151. Dekker, J., M.A. Marti-Renom, and L.A. Mirny, *Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data.* Nat Rev Genet, 2013. **14**(6): p. 390-403.
152. Baù, D. and M.A. Marti-Renom, *Genome structure determination via 3C-based data integration by the Integrative Modeling Platform.* Methods, 2012. **58**(3): p. 300-6.
153. Hu, M., et al., *Bayesian inference of spatial organizations of chromosomes.* PLoS Comput Biol, 2013. **9**(1): p. e1002893.
154. Lesne, A., et al., *3D genome reconstruction from chromosomal contacts.* Nat Methods, 2014. **11**(11): p. 1141-3.
155. Varoquaux, N., et al., *A statistical approach for inferring the 3D structure of the genome.* Bioinformatics, 2014. **30**(12): p. i26-33.
156. Giorgetti, L., et al., *Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription.* Cell, 2014. **157**(4): p. 950-63.
157. Zhang, Z., et al., *3D chromosome modeling with semi-definite programming and Hi-C data.* J Comput Biol, 2013. **20**(11): p. 831-46.
158. Nagano, T., et al., *Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.* Nature, 2013. **502**(7469): p. 59-64.
159. Duan, Z., et al., *A three-dimensional model of the yeast genome.* Nature, 2010. **465**(7296): p. 363-7.
160. Meluzzi, D. and G. Arya, *Recovering ensembles of chromatin conformations from contact probabilities.* Nucleic Acids Res, 2013. **41**(1): p. 63-75.
161. Bau, D., et al., *The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules.* Nat Struct Mol Biol, 2011. **18**(1): p. 107-14.
162. Umbarger, M.A., et al., *The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation.* Mol Cell, 2011. **44**(2): p. 252-64.
163. Le Dily, F., et al., *Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation.* Genes Dev, 2014. **28**(19): p. 2151-62.
164. Segal, M.R., et al., *Reproducibility of 3D chromatin configuration reconstructions.* Biostatistics, 2014. **15**(3): p. 442-56.
165. Russel, D., et al., *Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies.* PLoS Biol, 2012. **10**(1): p. e1001244.
166. Marbouty, M., et al., *Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms*, ed. G. McVean. Vol. 3. 2014.

167. Cagliero, C., et al., *Genome conformation capture reveals that the Escherichia coli chromosome is organized by replication and transcription.* Nucleic Acids Research, 2013. **41**(12): p. 6058-6071.
168. Perez-Rueda, E. and J. Collado-Vides, *The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12.* Nucleic Acids Res, 2000. **28**(8): p. 1838-47.
169. Sierro, N., et al., *DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information.* Nucleic Acids Res, 2008. **36**(Database issue): p. D93-6.
170. Helmann, J.D., *The extracytoplasmic function (ECF) sigma factors.* Adv Microb Physiol, 2002. **46**: p. 47-110.
171. Waites, K.B. and D.F. Talkington, *Mycoplasma pneumoniae and its role as a human pathogen.* Clin Microbiol Rev, 2004. **17**(4): p. 697-728, table of contents.
172. Atkinson, T.P., M.F. Balish, and K.B. Waites, *Epidemiology, clinical manifestations, pathogenesis and laboratory detection of Mycoplasma pneumoniae infections.* FEMS Microbiol Rev, 2008. **32**(6): p. 956-73.
173. Wilson, M.H. and A.M. Collier, *Ultrastructural study of Mycoplasma pneumoniae in organ culture.* Journal of Bacteriology, 1976. **125**(1): p. 332-339.
174. Razin, S., *Molecular biology and genetics of mycoplasmas (Mollicutes).* Microbiol Rev, 1985. **49**(4): p. 419-55.
175. Edward, D.G.f. and E.A. Freundt, *The Classification and Nomenclature of Organisms of the Pleuropneumonia Group.* Journal of General Microbiology, 1956. **14**(1): p. 197-207.
176. Maniloff, J., *Phylogeny and Evolution*, in *Molecular Biology and Pathogenicity of Mycoplasmas*, S. Razin and R. Herrmann, Editors. 2002, Springer US. p. 31-43.
177. Johansson, K.-E. and B. Pettersson, *Taxonomy of Mollicutes*, in *Molecular Biology and Pathogenicity of Mycoplasmas*, S. Razin and R. Herrmann, Editors. 2002, Springer US. p. 1-29.
178. Weisburg, W.G., et al., *A phylogenetic analysis of the mycoplasmas: basis for their classification.* J Bacteriol, 1989. **171**(12): p. 6455-67.
179. Sirand-Pugnet, P., et al., *Evolution of mollicutes: down a bumpy road with twists and turns.* Res Microbiol, 2007. **158**(10): p. 754-66.
180. Jensen, J.S., *Mycoplasma genitalium: the aetiological agent of urethritis and other sexually transmitted diseases.* J Eur Acad Dermatol Venereol, 2004. **18**(1): p. 1-11.
181. Levisohn, S. and S.H. Kleven, *Avian mycoplasmosis (Mycoplasma gallisepticum).* Rev Sci Tech, 2000. **19**(2): p. 425-42.
182. Balish, M.F. and D.C. Krause, *Mycoplasmas: a distinct cytoskeleton for wall-less bacteria.* J Mol Microbiol Biotechnol, 2006. **11**(3-5): p. 244-55.
183. May, M., et al., *Mycoplasma insons sp. nov., a twisted mycoplasma from green iguanas (Iguana iguana).* FEMS Microbiol Lett, 2007. **274**(2): p. 298-303.
184. MARE, C.J. and W.P. SWITZER, *NEW SPECIES: MYCOPLASMA HYOPNEUMONIAE; A CAUSATIVE AGENT OF VIRUS PIG PNEUMONIA.* Vet Med Small Anim Clin, 1965. **60**: p. 841-6.
185. Biberfeld, G. and P. Biberfeld, *Ultrastructural Features of Mycoplasma pneumoniae.* Journal of Bacteriology, 1970. **102**(3): p. 855-861.

186. Kammer, G.M., J.D. Pollack, and A.S. Klainer, *Scanning-Beam Electron Microscopy of Mycoplasma pneumoniae.* Journal of Bacteriology, 1970. **104**(1): p. 499-502.
187. Kim, C.K., R.M. Pfister, and N.L. Somerson, *Electron microscopy of Mycoplasma pneumoniae microcolonies grown on solid surfaces.* Applied and Environmental Microbiology, 1977. **34**(5): p. 591-594.
188. Boatman, E.S., *3 - MORPHOLOGY AND ULTRASTRUCTURE OF THE MYCOPLASMATALES*, in *The Mycoplasmas*, M.F.B. Razin, Editor. 1979, Academic Press. p. 63-102.
189. Razin, S., D. Yogev, and Y. Naot, *Molecular biology and pathogenicity of mycoplasmas.* Microbiol Mol Biol Rev, 1998. **62**(4): p. 1094-156.
190. Meng, K.E. and R.M. Pfister, *Intracellular structures of Mycoplasma pneumoniae revealed after membrane removal.* Journal of Bacteriology, 1980. **144**(1): p. 390-399.
191. Göbel, U., V. Speth, and W. Bredt, *Filamentous structures in adherent Mycoplasma pneumoniae cells treated with nonionic detergents.* The Journal of Cell Biology, 1981. **91**(2): p. 537-543.
192. Hatchel, J.M. and M.F. Balish, *Attachment organelle ultrastructure correlates with phylogeny, not gliding motility properties, in Mycoplasma pneumoniae relatives.* Microbiology, 2008. **154**(Pt 1): p. 286-95.
193. Regula, J.T., et al., *Defining the mycoplasma 'cytoskeleton': the protein composition of the Triton X-100 insoluble fraction of the bacterium Mycoplasma pneumoniae determined by 2-D gel electrophoresis and mass spectrometry.* Microbiology, 2001. **147**(4): p. 1045-1057.
194. Seybert, A., R. Herrmann, and A.S. Frangakis, *Structural analysis of Mycoplasma pneumoniae by cryo-electron tomography.* Journal of Structural Biology, 2006. **156**(2): p. 342-354.
195. Henderson, G.P. and G.J. Jensen, *Three-dimensional structure of Mycoplasma pneumoniae's attachment organelle and a model for its role in gliding motility.* Mol Microbiol, 2006. **60**(2): p. 376-85.
196. Baseman, J.B., et al., *Molecular basis for cytadsorption of Mycoplasma pneumoniae.* J Bacteriol, 1982. **151**(3): p. 1514-22.
197. Bredt, W., *Motility and multiplication of Mycoplasma pneumoniae. A phase contrast study.* Pathol Microbiol (Basel), 1968. **32**(6): p. 321-6.
198. Abu-Zahr, M.N. and M. Butler, *Ultrastructural features of Mycoplasma gallisepticum in tracheal explants under transmission and stereoscan electron microscopy.* Res Vet Sci, 1978. **24**(2): p. 248-53.
199. Bradbury, J.M., et al., *Mycoplasma imitans sp. nov. is related to Mycoplasma gallisepticum and found in birds.* Int J Syst Bacteriol, 1993. **43**(4): p. 721-8.
200. Tully, J.G., et al., *Titers of antibody to Mycoplasma in sera of patients infected with human immunodeficiency virus.* Clin Infect Dis, 1993. **17 Suppl 1**: p. S254-8.
201. Feldner, J., U. Gobel, and W. Bredt, *Mycoplasma pneumoniae adhesin localized to tip structure by monoclonal antibody.* Nature, 1982. **298**(5876): p. 765-767.
202. Hu, P., et al., *Mycoplasma pneumoniae infection: role of a surface protein in the attachment organelle.* Science, 1982. **216**(4543): p. 313-315.

203. Baseman, J.B., et al., *Identification of a 32-kilodalton protein of Mycoplasma pneumoniae associated with hemadsorption.* Isr J Med Sci, 1987. **23**(5): p. 474-9.
204. Seto, S. and M. Miyata, *Attachment organelle formation represented by localization of cytadherence proteins and formation of the electron-dense core in wild-type and mutant strains of Mycoplasma pneumoniae.* J Bacteriol, 2003. **185**(3): p. 1082-91.
205. Krause, D.C., *Mycoplasma pneumoniae cytadherence: unravelling the tie that binds.* Molecular Microbiology, 1996. **20**(2): p. 247-253.
206. Seto, S., et al., *Involvement of P1 adhesin in gliding motility of Mycoplasma pneumoniae as revealed by the inhibitory effects of antibody under optimized gliding conditions.* J Bacteriol, 2005. **187**(5): p. 1875-7.
207. Morrison-Plummer, J., D.K. Leith, and J.B. Baseman, *Biological effects of anti-lipid and anti-protein monoclonal antibodies on Mycoplasma pneumoniae.* Infect Immun, 1986. **53**(2): p. 398-403.
208. Romero-Arroyo, C.E., et al., *Mycoplasma pneumoniae protein P30 is required for cytadherence and associated with proper cell development.* J Bacteriol, 1999. **181**(4): p. 1079-87.
209. Hasselbring, B.M., J.L. Jordan, and D.C. Krause, *Mutant analysis reveals a specific requirement for protein P30 in Mycoplasma pneumoniae gliding motility.* J Bacteriol, 2005. **187**(18): p. 6281-9.
210. Krause, D.C., et al., *Transposon mutagenesis reinforces the correlation between Mycoplasma pneumoniae cytoskeletal protein HMW2 and cytadherence.* J Bacteriol, 1997. **179**(8): p. 2668-77.
211. Seto, S., et al., *Visualization of the attachment organelle and cytadherence proteins of Mycoplasma pneumoniae by immunofluorescence microscopy.* J Bacteriol, 2001. **183**(5): p. 1621-30.
212. Krause, D.C. and M.F. Balish, *Structure, function, and assembly of the terminal organelle of Mycoplasma pneumoniae.* FEMS Microbiology Letters, 2001. **198**(1): p. 1-7.
213. Balish, M.F., et al., *Localization of Mycoplasma pneumoniae cytadherence-associated protein HMW2 by fusion with green fluorescent protein: implications for attachment organelle structure.* Mol Microbiol, 2003. **47**(1): p. 49-60.
214. Krause, D.C. and M.F. Balish, *Cellular engineering in a minimal microbe: structure and assembly of the terminal organelle of Mycoplasma pneumoniae.* Mol Microbiol, 2004. **51**(4): p. 917-24.
215. Bredt, W., *5 - MOTILITY*, in *The Mycoplasmas*, M.F.B. Razin, Editor. 1979, Academic Press. p. 141-155.
216. Uenoyama, A. and M. Miyata, *Identification of a 123-kilodalton protein (Gli123) involved in machinery for gliding motility of Mycoplasma mobile.* J Bacteriol, 2005. **187**(16): p. 5578-84.
217. Jordan, J.L., et al., *Protein P200 is dispensable for Mycoplasma pneumoniae hemadsorption but not gliding motility or colonization of differentiated bronchial epithelium.* Infect Immun, 2007. **75**(1): p. 518-22.
218. Hasselbring, B.M. and D.C. Krause, *Cytoskeletal protein P41 is required to anchor the terminal organelle of the wall-less prokaryote Mycoplasma pneumoniae.* Mol Microbiol, 2007. **63**(1): p. 44-53.

219. Quinlan, D.C. and J. Maniloff, *Membrane Association of the Deoxyribonucleic Acid Growing-Point Region in Mycoplasma gallisepticum.* J Bacteriol, 1972. **112**(3): p. 1375-9.
220. Hasselbring, B.M., et al., *Terminal organelle development in the cell wall-less bacterium Mycoplasma pneumoniae.* Proc Natl Acad Sci U S A, 2006. **103**(44): p. 16478-83.
221. Himmelreich, R., et al., *Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.* Nucleic Acids Research, 1996. **24**(22): p. 4420-4449.
222. Blattner, F.R., et al., *The complete genome sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-62.
223. Lluch-Senar, M., et al., *Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium.* Mol Syst Biol, 2015. **11**: p. 780.
224. Consortium, I.H.G.S., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
225. Fraser, C.M., et al., *The minimal gene complement of Mycoplasma genitalium.* Science, 1995. **270**(5235): p. 397-403.
226. Morowitz, H.J., *The completeness of molecular biology.* Isr J Med Sci, 1984. **20**(9): p. 750-3.
227. Guell, M., et al., *Transcriptome complexity in a genome-reduced bacterium.* Science, 2009. **326**(5957): p. 1268-71.
228. Kuhner, S., et al., *Proteome organization in a genome-reduced bacterium.* Science, 2009. **326**(5957): p. 1235-40.
229. Yus, E., et al., *Impact of genome reduction on bacterial metabolism and its regulation.* Science, 2009. **326**(5957): p. 1263-8.
230. Maier, T., et al., *Quantification of mRNA and protein and integration with protein turnover in a bacterium.* Mol Syst Biol, 2011. **7**: p. 511.
231. Wodke, J.A., et al., *Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling.* Mol Syst Biol, 2013. **9**: p. 653.
232. Lluch-Senar, M., et al., *Comprehensive methylome characterization of Mycoplasma genitalium and Mycoplasma pneumoniae at single-base resolution.* PLoS Genet, 2013. **9**(1): p. e1003191.
233. Wodke, J.A., et al., *MyMpn: a database for the systems biology model organism Mycoplasma pneumoniae.* Nucleic Acids Res, 2015. **43**(Database issue): p. D618-23.
234. Sexton, T., et al., *Gene regulation through nuclear organization.* Nat Struct Mol Biol, 2007. **14**(11): p. 1049-55.
235. Misteli, T., *Beyond the sequence: cellular organization of genome function.* Cell, 2007. **128**(4): p. 787-800.
236. Kalhor, R., et al., *Genome architectures revealed by tethered chromosome conformation capture and population-based modeling.* Nat Biotechnol, 2011. **30**(1): p. 90-8.
237. Marti-Renom, M.A. and L. Mirny, *Bridging the resolution gap in structural modeling of 3D genome organization.* PLoS Comput Biol, 2011. **in press**.
238. Russel, D., et al., *Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies.* PLoS Biol, 2012. **10**(1): p. e1001244.

239. Le Dily, F., et al., *Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation.* Genes Dev, 2014. **28**(19): p. 2151-62.
240. Baù, D. and M. Marti-Renom, *Structure determination of genomic domains by satisfaction of spatial restraints.* Chromosome Research, 2011. **19**(1): p. 25-35.
241. Junier, I., F. Boccard, and O. Espeli, *Polymer modeling of the E. coli genome reveals the involvement of locus positioning and macrodomain structuring for the control of chromosome conformation and segregation.* Nucleic Acids Res, 2014. **42**(3): p. 1461-73.
242. Junier, I., O. Martin, and F. Kepes, *Spatial and topological organization of DNA chains induced by gene co-localization.* PLoS Comput Biol, 2010. **6**(2): p. e1000678.
243. Binder, K. and D.W. Heermann, *Monte Carlo simulations in statistical physics*. 2010: Springer.
244. Imakaev, M., et al., *Iterative correction of Hi-C data reveals hallmarks of chromosome organization.* Nat Methods, 2012. **9**(10): p. 999-1003.
245. Hall, M., et al., *The WEKA Data Mining Software: An Update.* SIGKDD Explorations., 2009. **11**(1).
246. Hou, C., et al., *Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains.* Mol Cell, 2012. **48**(3): p. 471-84.
247. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.
248. Nora, E.P., J. Dekker, and E. Heard, *Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods?* Bioessays, 2013. **35**(9): p. 818-28.
249. Ong, C.T. and V.G. Corces, *CTCF: an architectural protein bridging genome topology and function.* Nat Rev Genet, 2014. **15**(4): p. 234-46.
250. Junier, I., et al., *CTCF-mediated transcriptional regulation through cell type-specific chromosome organization in the beta-globin locus.* Nucleic Acids Res, 2012. **40**(16): p. 7718-27.
251. Dekker, J. and L. Mirny, *Biological techniques: Chromosomes captured one by one.* Nature, 2013. **502**(7469): p. 45-6.
252. Rust, M.J., M. Bates, and X. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM).* Nat Meth, 2006. **3**(10): p. 793-796.
253. Betzig, E., et al., *Imaging intracellular fluorescent proteins at nanometer resolution.* Science, 2006. **313**(5793): p. 1642-5.
254. Hess, S.T., T.P.K. Girirajan, and M.D. Mason, *Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy.* Biophysical Journal, 2006. **91**(11): p. 4258-4272.
255. Fölling, J., et al., *Fluorescence nanoscopy by ground-state depletion and single-molecule return.* Nat Methods, 2008. **5**(11): p. 943-5.
256. Wang, W., et al., *Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria.* Science, 2011. **333**(6048): p. 1445-1449.
257. Balish, M.F., *Subcellular structures of mycoplasmas.* Front Biosci, 2006. **11**: p. 2017-27.

258. Himmelreich, R., et al., *Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.* Nucleic Acids Res, 1996. **24**(22): p. 4420-49.
259. Trussart, M., et al., *Assessing the limits of restraint-based 3D modeling of genomes and genomic domains.* Nucleic Acids Research, 2015.
260. Needleman, S.B. and C.D. Wunsch, *A general method applicable to the search for similarities in the amino acid sequence of two proteins.* J Mol Biol, 1970. **48**(3): p. 443-53.
261. Miyata, M. and J.D. Petersen, *Spike Structure at the Interface between Gliding Mycoplasma mobile Cells and Glass Surfaces Visualized by Rapid-Freeze-and-Fracture Electron Microscopy.* J Bacteriol, 2004. **186**(13): p. 4382-6.
262. Wolter, S., et al., *rapidSTORM: accurate, fast open-source software for localization microscopy.* Nat Meth, 2012. **9**(11): p. 1040-1041.
263. Pengo, T., S.J. Holden, and S. Manley, *PALMsiever: a tool to turn raw data into results for single-molecule localization microscopy.* Bioinformatics, 2015. **31**(5): p. 797-798.
264. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Meth, 2012. **9**(7): p. 676-682.
265. Pique-Regi, R., et al., *Sparse representation and Bayesian detection of genome copy number alterations from microarray data.* Bioinformatics, 2008. **24**(3): p. 309-318.
266. Hilbert, H., et al., *Sequence analysis of 56 kb from the genome of the bacterium Mycoplasma pneumoniae comprising the dnaA region, the atp operon and a cluster of ribosomal protein genes.* Nucleic Acids Res, 1996. **24**(4): p. 628-39.
267. Mrázek, J. and S. Karlin, *Strand compositional asymmetry in bacterial and large viral genomes.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3720-5.
268. Rocha, E.P., *Order and disorder in bacterial genomes.* Curr Opin Microbiol, 2004. **7**(5): p. 519-27.
269. Miele, A. and J. Dekker, *Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C).* Methods Mol Biol, 2009. **464**: p. 105-21.
270. Heuser, J.E., *The origins and evolution of freeze-etch electron microscopy.* J Electron Microsc (Tokyo), 2011. **60**(Suppl 1): p. S3-S29.
271. Lin, D.C., P.A. Levin, and A.D. Grossman, *Bipolar localization of a chromosome partition protein in Bacillus subtilis.* Proc Natl Acad Sci U S A, 1997. **94**(9): p. 4721-6.
272. Jensen, R.B. and L. Shapiro, *The Caulobacter crescentus smc gene is required for cell cycle progression and chromosome segregation.* Proc Natl Acad Sci U S A, 1999. **96**(19): p. 10661-6.
273. Berlatzky, I.A., A. Rouvinski, and S. Ben-Yehuda, *Spatial organization of a replicating bacterial chromosome.* Proceedings of the National Academy of Sciences, 2008. **105**(37): p. 14136-14140.
274. Gellert, M., et al., *Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase.* Proc Natl Acad Sci U S A, 1976. **73**(12): p. 4474-8.
275. Lee, P.S., et al., *Effects of the chromosome partitioning protein Spo0J (ParB) on oriC positioning and replication initiation in Bacillus subtilis.* J Bacteriol, 2003. **185**(4): p. 1326-37.

276. Toro, E., et al., *Caulobacter requires a dedicated mechanism to initiate chromosome segregation.* Proc Natl Acad Sci U S A, 2008. **105**(40): p. 15435-40.
277. Bates, D. and N. Kleckner, *Chromosome and replisome dynamics in E. coli: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation.* Cell, 2005. **121**(6): p. 899-911.
278. Nielsen, H.J., et al., *The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves.* Mol Microbiol, 2006. **62**(2): p. 331-8.
279. Trifonov, E.N., *Curved DNA.* CRC Crit Rev Biochem, 1985. **19**(2): p. 89-106.