High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome

Massachusetts Institute of Technology

High-resolution mapping of the spatial organization of a

bacterial chromosome

Tung B.K. Le[1], Maxim V. Imakaev[2], Leonid A. Mirny[2,3*], Michael T. Laub[1,4*]

[1]Department of Biology

[2]Department of Physics

[3]Institute for Medical Engineering and Sciences

[4]Howard Hughes Medical Institute

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[*]Correspondence: laub@mit.edu, leonid@mit.edu

Abstract

Chromosomes must be highly compacted and organized within cells, but how this is achieved *in vivo* remains poorly understood. We report the use of Hi-C to map the structure of bacterial chromosomes. Analysis of Hi-C data and polymer modeling indicates that the *Caulobacter crescentus* chromosome consists of multiple, largely independent spatial domains likely comprised of supercoiled plectonemes arrayed into a bottlebrush-like fiber. These domains are stable throughout the cell cycle and re-established concomitantly with DNA replication. We provide evidence that domain boundaries are established by highly-expressed genes and the formation of plectoneme-free regions, whereas the histone-like protein HU and SMC promote short-range compaction and the cohesion of chromosomal arms, respectively. Collectively, our results reveal general principles for the organization and structure of chromosomes *in vivo*.

In all organisms chromosomal DNA must be compacted nearly three orders of magnitude to fit within the limited volume of a cell. Chromosomes must adopt structures that are compatible with critical cellular processes such as transcription, DNA replication, and chromosome segregation. While bacterial chromosomes are probably highly organized within cells (*1-6*), the resolution of previous studies is limited. For eukaryotes, chromosome conformation capture coupled with deep-sequencing, or Hi-C, has enabled higher resolution studies of chromosome structure *in vivo (7, 8)*. These studies have suggested that interphase chromosomes are organized into a series of topological or structural domains <1 Mb in size (*8-11*), but the factors that create, maintain, and influence these domains is presently unknown.

To study the organization of bacterial chromosomes with high-resolution, we used Hi-C on *Caulobacter* cells (fig. S1-S2). We performed Hi-C on swarmer cells that each contain a single circular and unreplicated chromosome. To analyze our Hi-C data, we divided the genome into 10-kilobase (kb) bins with interaction frequencies for each restriction fragment assigned to corresponding bins. We visualized interactions as a heat map where each matrix position, m_{ij}, reflects the relative frequency of interactions between loci in bins i and j. For a description of data processing, normalization, reproducibility, and comparability to a previous 5C study (*4*), see the SOM (figs. S3-S6).

The swarmer cell interaction matrix contains two prominent diagonals (Fig. 1A). The main diagonal reflects high frequency interactions between loci on the same chromosomal arm. The other, less prominent diagonal captures lower frequency inter-arm contacts, *i.e.* those between loci on one chromosomal arm and those on the opposite arm of the circular genome. These locus pairs are separated by substantial distances in the primary genome sequence, but the Hi-C data indicates they are often physically adjacent and capable of interacting. This overall pattern, also seen by 5C (*4*), is consistent with the *Caulobacter* chromosome adopting an elongated structure with the single origin anchored at one pole and the two chromosome arms running the length of the cell in close proximity.

Further inspection of the Hi-C interaction matrix revealed highly self-interacting regions, or chromosomal interaction domains (CIDs), of the genome that appear as squares along the main diagonal (Fig. 1A) or as triangles if rotating the contact map 45° clockwise (Fig. 1B, S7-S8). Loci within a CID interact preferentially with other loci within the same CID compared to other CIDs. Loci at the border of each CID strongly favor interactions with loci

on their left- or right-hand side but not both, whereas loci in the middle of a CID show high levels of interaction with loci to both sides. The Hi-C matrix exhibited variability in boundary sharpness and some nested domains (Fig. 1B, S8-S9). This hierarchical organization resembles the so-called topologically-associated domains (TADs) previously observed in eukaryotic Hi-C data (*8-11*).

To systematically map the boundaries of CIDs, we generated plots of directional preference as a function of genome position (Fig. 1B, S7-S9). There were 23 CIDs, ranging in length from 30 to 420 kb (table S1). CIDs were not artefacts of restriction site or sequencing read densities (fig. S10), and were independently verified using a recombination-based assay for interaction frequencies (fig. S11). The CIDs identified must be present in most cells, as Hi-C reflects interactions in a population of cells. Individual cells could have other, perhaps transient, domains.

CID boundaries were enriched for highly-expressed genes (p=7.7x10^{-5}, Fisher's exact test, fig. S10). Of the 23 CID boundaries, 17 contained one or more highly-transcribed genes (Fig. 1B, S8,S10). We hypothesized that high gene expression unwinds the DNA duplex and creates plectoneme-free regions (PFRs), which form barriers between CIDs. These PFRs likely prevent the diffusion of supercoils and physically separate CIDs, thereby decreasing the contact probabilities of loci in different domains, as also suggested in *Salmonella* (*1*).

To better understand the 3D organization of the *Caulobacter* chromosome, we developed a detailed polymer model (see SOM, figs. S12-S15). The chromosome was modeled as a circular polymer comprising a dense array of plectonemes that have no sequence specificity and are stochastic in length and location (Fig. 1C, S12). We generated an equilibrium ensemble of chromosome conformations, simulated the Hi-C procedure on 25,000 modeled chromosomes, and compared the resulting data to experimental Hi-C data. By systematically varying model parameters, we identified values that provided the best fit to the observed Hi-C contact frequencies (fig. S13-S14).

Our model reflects two broad levels of chromosomal organization. On one level, the DNA is arranged into a fiber of ~300 plectonemes separated by small spacers, resembling a bottlebrush. At a higher level, the bottlebrush fiber forms a circular chromosome tethered at the pole by an origin-proximal region with chromosomal arms in close proximity down the long-axis of the cell. We also used the model to examine the effects of PFRs on interactions between loci. A single PFR of ~2 kb created a space of ~100-200 nm between flanking loci.

3

This spacer reduced contacts between neighboring plectonemes and prevented diffusion of supercoils through the PFR in the simulations, recapitulating a CID boundary (Fig. 1D, S16). We then introduced PFRs into the chromosome model at the locations of the 20 most highly-expressed genes. Simulated Hi-C data generated a pattern of CIDs that resembled those observed experimentally (Fig. 1D, S17).

To probe the role of gene expression in chromosome structure, we performed Hi-C on swarmer cells treated for 30 minutes with rifampicin (rif), an inhibitor of transcription elongation (*12*). The interaction matrix for rif-treated cells was globally similar to that of untreated cells, indicating that the overall shape of the chromosome was unperturbed (Fig. 2A-B). However, CID boundaries were severely disrupted in rif-treated cells leading to a nearly domain-free organization (Fig. 2B, S18-19). Simulations of rif-treated chromosomes, performed by removing PFRs, also produced domain-free contact maps (fig. S19-21).

We also moved the highly-expressed gene *rsaA* to the *vanA* locus, a poorly expressed region of the genome (fig. S22). The native *vanA* locus normally resides within a CID, but insertion of *rsaA* generated a sharp new CID boundary at this position in the genome (Fig. 2C). Relocating *rsaA* to the *xylX* locus, ~1.7 Mb from the *vanA* locus, also created a new CID boundary at this location (fig. S23). We conclude that highly-expressed genes play a direct role in defining chromosomal domain boundaries.

We also used Hi-C to probe the effect of inhibiting supercoiling on chromosomal organization. Swarmer cells were incubated for 30 minutes with a sublethal dose of novobiocin (fig. S24) and then subjected to Hi-C analysis (Fig. 3A-B, S18). Novobiocin, which inhibits DNA gyrase and negative supercoiling (*13*), significantly reduced the frequency of interactions in the 20-200 kb range while modestly increasing interactions in the 200-800 kb range relative to the untreated wild type (fig. S25). Additionally, novobiocin reduced the sharpness and positions of CID boundaries (Fig. 3B, S18). The decrease in interaction frequencies in the 20-200 kb range does not result from the emergence of a subpopulation of cells with gross defects in chromosome organization (fig. S26).

To model the effect of novobiocin, we increased the spacing between duplexes in a plectoneme monomer and increased the average spacing between plectonemes five-fold. Subsequent simulations reproduced the partial loss of CID boundaries and changes in contact frequencies observed (fig. S19-20, S27). Taken together, our results demonstrate that

supercoiling is (i) critical to genome compaction in the 20-200 kb range and (ii) helps establish CIDs *in vivo*.

To investigate the role of nucleoid-associated proteins in chromosomal organization, we focused on the histone-like proteins HU1 and HU2 (*14, 15*). We isolated swarmer cells from a Δ*hup1*Δ*hup2* strain and performed Hi-C. The interaction matrix for Δ*hup1*Δ*hup2* was grossly similar to that of wild-type cells (Fig. 3C). The correlation between directional preferences for each bin along the chromosome of wild type and *Δhup1Δhup2* cells was high (r=0.73, p<10^{-15}) indicating that CID boundaries were retained, although they became less pronounced in the absence of HU1 and HU2 (fig. S28). The contact probability plot for Δ*hup1*Δ*hup2* cells revealed a significant decrease in short-range contacts, up to ~100 kb, compared to the wild type (fig. S25). These changes suggest that deleting HU may disrupt interactions within and between neighboring plectonemes without affecting inter-plectoneme spacing, which is critical for producing CID boundaries (Fig. 3C, S20). We modeled the effect of deleting HU by increasing the spacing between duplexes within a plectoneme; simulations based on this model recapitulated the Hi-C data (fig. S29). These analyses indicate that HU facilitates local, short-range compaction of the genome, possibly through the packing and stabilization of plectonemic DNA.

SMC (structural maintenance of chromosomes) homologs are found in all domains of life, and can form ring-like structures that facilitate chromosome cohesion or compaction (*16-18*). Hi-C analysis of *Caulobacter Δsmc* swarmer cells showed a clear drop in the frequency of inter-chromosomal arm interactions (Fig. 3D-E). Concomitantly, loci typically interacted with a wider range of loci on the opposite chromosomal arm compared to wild-type cells. In contrast, the frequencies of intra-arm interactions (fig. S25) and CID boundaries (fig. S28) were largely unaffected in the absence of SMC. Although many bacterial SMC proteins may compact DNA (*19, 20*), our data suggest that *Caulobacter* SMC contributes significantly to the cohesion of chromosome arms in swarmer cells.

To study chromosome organization changes during cell-cycle progression, we performed Hi-C on synchronized cells collected at regular intervals during the cell cycle (Fig. 4). After replication initiation in *Caulobacter*, one origin remains near the stalked pole while the other moves to the opposite pole (*6*). As the cell cycle progressed, the Hi-C contact maps indicated progressively more interactions between origin- and terminus-proximal loci (Figs. 4, S30), particularly at the 30 and 45 minute time-points when the newly translocated origin is close

to the polarly-localized terminus (Fig. 4). However, the frequency of these interactions was still nearly 16-fold less than an average interaction between loci within 100 kb. This difference implies that the translocating chromosome is largely insulated from the anchored chromosome despite their physical proximity, which is likely to be advantageous to the segregation process by preventing the entanglement of chromosomes destined for different daughter cells. Also, the CIDs identified in swarmer cells (Fig. 1A-B) remained intact throughout the cell cycle (Fig. 4) indicating that CIDs must get re-established concurrently with, or shortly after, DNA replication (fig. S31-32), which may also help newly-replicated chromosomes from becoming entangled.

CIDs are reminiscent of the topologically-associated domains (TADs) documented in eukaryotes (*9-11, 21*) suggesting that domains on a 100-kb length-scale are a fundamental unit of chromosome structure in all organisms. Like TADs, *Caulobacter* CIDs often appear as nested domains, which may be related to the Mb-scale macrodomains identified in *E. coli (3, 5)*. The identification of CIDs indicates that many domain barriers in *Caulobacter* are relatively fixed; however, within each CID there could be additional barriers that arise and dissipate, perhaps stochastically, as suggested in *Salmonella* and *E. coli (1, 22)*, and CID boundaries may change in different growth conditions. Importantly, although chromosomal domains have now been documented in several organisms, the factors determining domain boundaries had been unclear. Although the DNA-binding proteins HU and SMC contribute to chromosome organization, they do not significantly affect CIDs. Instead, our work points to supercoiling and highly-expressed genes as critical determinants of domain formation in bacteria, and we suspect similar mechanisms contribute to creating TADs in higher organisms where boundaries may also be enriched for highly-expressed genes *(9)*.

Figure Legends

Fig. 1. Partitioning of the *Caulobacter* chromosome into chromosomal interaction domains (CIDs). (A) Normalized NcoI Hi-C contact map for *Caulobacter* swarmer cells displaying contact frequencies for pairs of 10-kb bins across the genome. Axes indicate the genome position of each bin. Inset: simplified genomic map showing the origin of replication (*ori*) and terminus (*ter*) along with the right (black) and left (grey) chromosomal arms. **(B)** Hi-C contact map for one arm of the chromosome rotated 45° clockwise with directional preference plots below. Left- and right-ward preferences are shown as green and red bars, respectively. CIDs are outlined in yellow and numbered. Highly-expressed genes at CID boundaries are listed (hypothetical genes designated by GenBank ID). **(C)** Polymer chromosome model indicating polarly-anchored origin (magenta), chromosome backbone (black), and plectonemes (grey, with every 10th plectoneme on one arm in a color). **(D)** Comparison of experimental and simulated Hi-C contact maps, indicating that plectoneme-free regions (PFRs) can account for CIDs.

Fig. 2. Effect of inhibiting transcription on CID boundaries. Normalized BglII Hi-C contact maps for **(A)** untreated and **(B)** rifampicin-treated swarmer cells. **(C)** Hi-C contact maps for wild-type, Δ*rsaA*, and Δ*rsaA* + *van*::P*rsaA*-*rsaA*. Only the region of the genome containing the *van* locus (dashed line) is shown.

Fig. 3. Contribution of supercoiling, HU, and SMC to chromosome organization. Normalized BglII Hi-C contact maps for **(A)** wild type, **(B)** novobiocin-treated wild type, **(C)** *Δhup1Δhup2*, and **(D)** *Δsmc* swarmer cells. Only the upper left regions of the symmetric Hi-C maps are shown. A region from 2-3 Mb is enlarged and shown below each map. A cartoon summarizing the effects of each perturbation is shown with neighboring domains of plectonemic DNA in red and green and plectoneme-free regions in blue. **(E)** Hi-C scores for the diagonal (indicated in inset) of each contact map.

Fig. 4. Dynamics of chromosomal organization during cell cycle progression. Plots show a section of the Hi-C interaction map for the cell-cycle time points indicated.

References

1. B.M. Booker, S. Deng, N.P. Higgins, *Mol Microbiol* **78**, 1348 (2010).
2. J.K. Fisher *et al.*, *Cell* **153**, 882 (2013).
3. H. Niki, Y. Yamaichi, S. Hiraga, *Genes Dev* **14**, 212 (2000).
4. M.A. Umbarger *et al.*, *Mol Cell* **44**, 252 (2011).
5. M. Valens, S. Penaud, M. Rossignol, F. Cornet, F. Boccard, *EMBO J* **23**, 4330 (2004).
6. P.H. Viollier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9257 (2004).
7. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, *Science* **295**, 1306 (2002).
8. E. Lieberman-Aiden *et al.*, *Science* **326**, 289 (2009).
9. J.R. Dixon *et al.*, *Nature* **485**, 376 (2012).
10. E.P. Nora *et al.*, *Nature* **485**, 381 (2012).
11. T. Sexton *et al.*, *Cell* **148**, 458 (2012).
12. E.A. Campbell *et al.*, *Cell* **104**, 901 (2001).
13. M. Gellert, M.H. O'Dea, T. Itoh, J. Tomizawa, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 4474 (1976).
14. S. Dillon, C.J. Dorman, *Nat Rev Microbiol* **8**, 185 (2009).
15. F. Guo, S. Adhya, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4309 (2007).
16. A. Badrinarayanan, R. Reyes-Lamothe, S. Uphoff, M.C. Leake, D.J. Sherratt, *Science* **338**, 528 (2012).
17. T. Hirano, *Nat Rev Mol Cell Biol* **7**, 311 (2006).
18. M.A. Schwartz, L. Shapiro, *Molecular Microbiology* **82**, 1359 (2011).
19. P.L. Graumann, T. Knust, *Chromosome Res* **17**, 265 (2009).
20. A. Volkov, J. Mascarenhas, C. Andrei-Selmer, H.D. Ulrich, P.L. Graumann, *Mol Cell Biol* **23**, 5638 (2003).
21. C. Hou, L. Li, Z.S. Qin, V.G. Corces, *Mol Cell* **48**, 471 (2012).
22. L. Postow, C.D. Hardy, J. Arsuaga, N.R. Cozzarelli, *Genes Dev* **18**, 1766 (2004).
23. R. Jensen, L. Shapiro, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10661 (1999).
24. M. Imakaev *et al.*, *Nat Methods* **9**, 999 (2012).
25. S. Hong, *et al., Proc. Natl. Acad. Sci. U.S.A.* **110**, 1674 (2012).

Supplementary Materials

Materials and Methods
Fig. S1-S34
Tables S1-S4
References (23-25)

Figure 1

A

genome position (kb)

genome position (kb)

CC3763 0 kb CC0001
ori
3000 kb   1000 kb
ter
2000 kb

Hi-C interaction scores

0   0.01   0.02   0.03

B

0   500   1000   1500   2000 kb

directional preference

2
0
-2

1   2   3   4 5 6 7   8   9   10   11   12

CC0162/3   SucCDA   CC0472/3 L28 L34   RsaA   Ef-Tu L13/S9   Hfq
                                                            S10..L15
L7/L12/L25/L10 GroEL/ES   ribosomal protein cluster
Rpoββ'

C

● origin of replication   backbone   plectoneme

D

experimental Hi-C

1 Mb   2 Mb

simulated Hi-C

1 Mb   2 Mb

PFR

Figure 2

A

no treatment

B

+ rifampicin

C

wild type

$\Delta rsaA$ van::empty vector

$\Delta rsaA$ van::P_{rsaA}-$rsaA$

Figure 3

Figure 4