New approaches to understanding the spatial organization of bacterial genomes

Tung BK Le¹ and Michael T Laub¹,²

In all organisms, chromosomal DNA must be compacted nearly three orders of magnitude to fit within the limited volume of a cell. However, chromosomes cannot be haphazardly packed, and instead must adopt structures compatible with numerous cellular processes, including DNA replication, chromosome segregation, recombination, and gene expression. Recent technical advances have dramatically enhanced our understanding of how chromosomes are organized in vivo and have begun to reveal the mechanisms and forces responsible. Here, we review the current arsenal of techniques used to query chromosome structure, focusing first on single-cell fluorescence microscopy approaches that directly examine chromosome structure and then on population-averaged biochemical methods that infer chromosome structure based on the interaction frequencies of different loci. We describe the power of these techniques, highlighting the major advances they have produced while also discussing their limitations.

Addresses
¹ Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
² Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Corresponding author: Laub, Michael T (laub@mit.edu)

Microscopy and imaging of individual chromosomal loci

The spatial arrangement of chromosomes can be partially reconstructed by tracking the subcellular positions of specific loci using microscopy. One of the first methods developed for such visualizations was FISH, or fluorescence in situ hybridization. FISH involves the partial permeabilization of cells and subsequent addition of fluorescently-labeled DNA probes that hybridize to complementary regions of the chromosome (Figure 1a). Epi-fluorescence microscopy can then reveal the subcellular location of the labeled locus. For example, an early examination of 23 loci in Escherichia coli by FISH [¹¹] revealed that loci are arranged within the cell in the same approximate order as they appear in the genome, implying a highly-organized chromosome (Figure 1a). In addition, these studies indicated that the E. coli chromosome possesses two so-called macrodomains, called Ori and Ter, in which loci near the origin of replication (oriC) or terminus (dif), respectively, frequently co-localize (Figure 1a).

Although FISH allows the visualization of loci without modification of the genome, it requires the fixing and permeabilization of cells, treatments that may alter chromosome conformation. Moreover, because FISH requires cell fixation, it cannot be used to track the
Single-cell microscopic techniques. A summary of microscopic techniques (left) used to probe the spatial organization of bacterial chromosomes, with example discoveries (right). (a) Fluorescence in situ hybridization. The origin of replication (oriC) is shown as a red dot and the terminus (ter) as a blue line. The left and right arms of the chromosome are colored in green and orange, respectively. The Ori and Ter macrodomains are indicated by thick opaque red and blue lines, respectively. (b) Fluorescent repressor-operator systems. Proposed patterns of chromosome organization in replicating Bacillus subtilis, slow-growing Escherichia coli, and Caulobacter crescentus based on traditional fluorescent repressor-operator systems (top). The proposed spatial organization of replicating chromosomes in fast-growing E. coli cells, based on studies using an orthogonal ParB/parS system is also shown (bottom). Cross-sections show the radial distribution of the left and right chromosomal arms as well as oriC, ter, and replication forks (purple triangles). (c) Imaging of fluorescently-tagged nucleoid-associated proteins. Reconstruction of the nucleoid of slow-growing E. coli cells expressing HupA-mCherry indicated a coiled-shaped structure of no particular handedness. Nucleoid intensities, summarized from Z projections, also showed rapid fluxes of density along the length of the nucleoid.

dynamic movement of chromosomal loci in living cells. These limitations were overcome in part through the development of fluorescent repressor-operator systems (FROS) in which an array (~240 copies) of operator sites (e.g. lacO, tetO or λO1,1) is inserted near a locus of interest and cells are engineered to express a cognate, fluorescently-tagged DNA-binding protein (e.g. LacI, TetR, or λCI) [4,12,13]. The binding of multiple fluorescent proteins to an operator array produces a discrete focus detectable by fluorescence microscopy that can be tracked over time in living cells (Figure 1b).

The use of fluorescent repressor-operator systems has produced significant advances in our understanding of chromosome organization and dynamics in bacteria. Early studies of Bacillus subtilis [14,15] analyzed four different
Although FROS has revolutionized the study of bacterial chromosome organization and been applied now to many different species [18,19,20,21], it has two important limitations. First, it requires the insertion of a large (~8 kbp) array into the genome, which can be difficult to engineer and, more importantly, can potentially perturb local genome architecture. The integrated operator arrays are also sometimes unstable, with recombination between operators leading to smaller arrays and diminished signal. Although not yet widely used, super-resolution microscopy can eliminate the need for large arrays; cells harboring arrays of just three tetO or lacO operators yield detectable signal [22,23].

The limitations of operator arrays can also be partially overcome by using ParB/parS systems derived from plasmid and chromosome partitioning systems. ParB proteins typically bind a single cognate parS site (~140–280 bp) and initiate the polymerization and spreading of additional ParB proteins on nearby DNA. Thus, insertion of a single parS site into a genome of interest and expression of a fluorescently-tagged ParB can enable the visualization of specific loci. ParB/parS derived from plasmids or chromosomes of different organisms, i.e. orthogonal ParB/parS systems, avoid interference with an endogenous ParB/parS. Further, the use of orthogonal ParB/parS systems with different recognition elements can enable the simultaneous examination of multiple loci in individual cells. The spatial organization of E. coli chromosomes was recently investigated during fast growth using ParB/parS systems from plasmid pMT1 and phage P1 to probe the positioning of 13 different loci. The left and right arms of replicating chromosomes were observed to form separate but parallel running structures in the outer shell of the nucleoid in each cell half (Figure 1b) [3]. By contrast, the replication forks, the origin of replication and the terminus were found at the center of the nucleoid (Figure 1b) [3]. The notion that different regions of the chromosome prefer different shells of the nucleoid, or positions along the short axis of a rod-shaped cells, had not been appreciated previously.

A second limitation of fluorescent repressor-operator systems is the tight binding (Kd ~ 1 nM) of TetR and LacI to multiple operator sites or the spreading of ParB around parS, which can produce a roadblock that impedes DNA replication and results in replication fork collapse or DNA stress [24,25]. Additionally, multimerization between the DNA-binding domains of repressors or between the fluorescent proteins fused to repressors can artificially increase cohesion of sister DNA loci [26]. This artifact can be partially reduced by using a variant of ParB (H.J. Nielsen, personal communication) or a fluorescent protein derivative with diminished propensity for multimerization [26].

**Global organization revealed by imaging DNA and DNA-binding proteins**

Another limitation of fluorescent repressor-operator systems is that only 2–3 loci can be visualized at a time. Assessing global properties of chromosome structure requires a different set of techniques. One common method involves imaging of cells stained with dyes such as 4’,6-diamidino-2-phenylindole (DAPI) that bind DNA non-specifically, thereby revealing the overall cellular distribution and gross structure of bacterial chromosomes *in vivo*. DAPI staining has demonstrated that the chromosomes of many bacteria, including E. coli and B. subtilis, do not fill the volume of the cell whereas in others, such as C. crescentus, staining occurs throughout the cell. In the former cases, co-staining for other proteins indicates that the nucleoids of these cells often exclude ribosomes, setting up a partial separation of transcription and translation [27]. In B. subtilis, DAPI-staining has also revealed major changes to chromosome morphology upon entry to sporulation, with the nucleoid transitioning from a globular state to an extended axial filament, which may facilitate pumping of one chromosome into the forespore [28].

Like FISH, DAPI staining involves fixed cells, precluding analyses of living cells. Thus, many labs have begun imaging fluorescently-tagged proteins that bind throughout the chromosome, including RNA polymerase subunits and nucleoid-associated proteins (NAPs) which, like histones in eukaryotes, often play important roles in structuring the chromosome [1,29,30]. The subcellular distribution of these proteins serves as a reasonable proxy for the chromosome and has driven significant new insights. One recent study of E. coli used an mCherry fusion to the nucleoid-associated protein HupsA to image chromosomes at high spatial and temporal resolution (Figure 1c). This work revealed an overall helical structure of the E. coli nucleoid,
although with no handedness preference (Figure 1c) [1]. This study also revealed the dynamic nature of the E. coli nucleoid. Even in non-replicating G1 cells, nucleoid density waves fluxed back and forth along the longitudinal axis of the cell (Figure 1c) [1]. In replicating cells, sister chromosomes segregated end-to-end in sequential, discontinuous pulses [1], with periodic (~20 min intervals) accumulation and relief of intra-nucleoid tethers. The nature of such tethers and the mechanisms responsible for the dynamic motions observed are currently unknown but may be crucial to chromosome segregation.

Although many NAPs, including HU, Fis, and StpA, are relatively evenly distributed across the E. coli genome [23], super-resolution imaging studies have suggested that some are not. For instance, H-NS appears to form discrete subcellular structures that may help to establish or reinforce chromosome structure, but the precise function of H-NS in genome organization remains unclear [23]. Still in its infancy, super-resolution imaging promises to provide important new insights into chromosome organization in the coming years.

**Inferring chromosome organization by recombination frequencies and Hi-C**

Although fluorescence microscopy-based methods have been the mainstay of chromosome biology, these approaches are limited in spatial resolution, even using super-resolution techniques. An alternative class of approaches involves the inference of chromosome structure through assays that report on the frequency of collisions between different loci. Collision, or contact, frequencies in a population of cells will strongly reflect interlocus distances, although they are not always equivalent.

One contact-frequency method uses site-specific recombination systems. Typically, two DNA elements, such as λ *attP* and *attB*, the sites recognized by λ integrase, are engineered at different chromosomal loci and the frequency of recombination between them is measured following induction of the recombinase (Figure 2a). This approach was used to measure recombination rates between *att* sites inserted at various loci around the E. coli genome (Figure 2a), confirming and extending the existence of macrodomains discovered initially by FISH [31**]. Loci within a given macrodomain recombined at higher frequencies with other loci in the same macrodomain.

A complementary system involves site-specific resolvases derived from transposon Tn3. Whereas λ Int can report on long-distance interactions, these resolvase systems, such as γ6 Res, are well-suited to probing shorter-range contacts as they rely on negative supercoiling and plec-tone slithering to bring resolution (res) sites together for recombination rather than random collisions (Figure 2b). Studies using the γ6 Res system in Salmonella typhimurium revealed the maximal length of a plec-tone, i.e. the longest distance over which res site can be brought together, to be ~100 kb [32]. This work further showed that barriers to plec-tone diffusion appear stochastically distributed within the chromosome [33], although regions of high transcription often formed major barriers to super-coil diffusion, suggesting that the chromosome may be partitioned somewhat deterministically (Figure 2b) [34**,35].

Although powerful, the throughput of recombination-based systems for probing chromosome structure is limited and dependent on an ability to insert the relevant sites throughout a genome of interest. However, a related class of approaches based on the technique chromosome conformation capture (3C) enables much greater throughput and has recently driven major advances in understanding chromosome structure in bacteria and eukaryotes [7*,36]. For 3C, cells are typically treated with formaldehyde to crosslink proteins to DNA and DNA to DNA, thereby preserving the conformation of the chromosome (Figure 2c). The chromosome is then restriction digested, followed by a ligation reaction under dilute conditions, which favors the joining of loci that were crosslinked together. These ligation events fuse loci that were in close spatial proximity when formaldehyde was added. Ligated junctions are then isolated for analysis. The original 3C assay uses PCR to query a junction of interest; 5C uses multiplexed PCR and sequencing of many different junctions; Hi-C involves the incorporation of a biotinylated nucleotide before ligation, a subsequent streptavidin-based precipitation of ligated regions, and next-generation sequencing to query all possible junctions (Figure 2c).

Both 3C and Hi-C reveal the identities of many interacting loci and the frequency of each interaction. The resolution of these techniques depends on the distribution and number of restriction sites and on the depth of sequencing performed. Although 3C-based methods do not require any genetic modifications, and are thus potentially applicable to any organism, success can hinge on the restriction enzyme used and the nature of the cell lysate produced following formaldehyde treatment. Because the initial steps of 3C-based methods are performed with a crude cell lysate, endogenous exonucleases or restriction enzyme inhibitors can significantly impact experimental outcomes.

The first 5C performed on bacteria was for C. crescentus swarmer cells that harbor a single circular chromosome [6*]. These data revealed that loci typically interacted most strongly with other loci within ~100 kb in primary sequence on the same chromosome arm, with weaker interactions occurring with loci at the same approximate
Techniques for studying bacterial chromosome structure by Le and Laub

Population-averaged biochemical methods based on locus-locus collision frequencies. As in Figure 1, a summary of techniques (left) used to probe the spatial organization of bacterial chromosomes, with example discoveries (right). (a) $\lambda$ Int recombination. Recombination between attP and attB sites causes the disruption of a reporter gene lacZ and creation of two new attL and attR sites. The direction of recombination is controlled by Int and Xis protein. The schematic of the E. coli genome shows oriC and ter as a red dot and a blue line, respectively. The left and right arms of the chromosome are colored green and orange. Four macrodomains: Ori, Ter, Left and Right are shown as thick opaque red, blue, green and orange lines, respectively. Two additional unstructured regions are shown as fuzzy green and orange lines. (b) $\gamma\delta$ Res recombination system. The slithering of supercoiled plectoneme brings $\gamma\delta$ res sites (blue and orange squares or circles) together for synapsis, and subsequent excision of intervening DNA (red). The $\gamma\delta$ Res system has shown that high levels of transcription (blue arrows) inhibits supercoil diffusion (black circles with arrowheads, preventing synapsis between flanking loci. (c) Genome-wide chromosome conformation capture assays (5C and Hi-C) have revealed the architecture of the C. crescentus chromosome. A Hi-C heat map uses colors to indicate the frequency of interactions between locus pairs across the genome. The main diagonal captures interactions within chromosomal arms (black dashed line) and the opposite diagonal indicates interactions between arms (grey dashed lines). The Hi-C map also indicates (see zoomed region) the presence of chromosomal interaction domains (CID) (orange triangles). A plectoneme-free region (blue) formed through high gene expression at domain boundaries spatially insulates DNA in flanking domains (green and red).

Position on the opposite arm of the chromosome (Figure 2c). This pattern reinforced FISH and FROS data showing that the Caulobacter origin is anchored at one pole of the cell with the two arms running in parallel down the long axis of the cell. These 5C data were also used to model the global configuration of the chromosome by (i) assuming that interaction frequencies reflect physical distance between loci and then (ii) searching for overall conformations that satisfy these inferred spatial constraints. The resulting ensemble of models suggested
that the two chromosomal arms may adopt a helical shape with the two arms gently twisted around one another [6], reminiscent of structures observed microscopically in *E. coli* [1,30], *B. subtilis* [37], and *Edellocitrio bactériocoros* [38].

This 5C study queried 169 × 170 locus interactions. The recent application of Hi-C to Caulobacter expanded the number of interactions queried to 700 × 700 and 2025 × 2025 interactions using BglII and NcoI, respectively [5**]. The increase in interactions examined produced contact maps with significantly higher resolution and, consequently, revealed new features of the Caulobacter chromosome. In particular, Hi-C revealed ~23 chromosomal interaction domains (CID), regions of the genome within which loci interact more frequently with each other than with loci in other domains (Figure 2c) [5**]. CIDs range in size from 30 to 400 kb with a mean of ~166 kb, and many appeared nested within larger domains that could correspond to macrodomains (Figure 2c). Importantly, because Hi-C signal comes from a population of cells, the CIDs documented are probably present in most or all cells; additional domains could be present in individual cells, but escape detection by Hi-C.

Domain boundaries frequently coincide with very highly-expressed genes, suggesting a potential causal relationship. Indeed, treating cells with rifampicin, which inhibits RNA polymerase, demonstrated that active gene expression is necessary for boundary formation. Further, insertion of a highly-expressed gene, *rsdA*, within a domain was sufficient to induce a new boundary [5**]. Highly-transcribed genes probably form plectoneme-free regions that are less compacted than neighboring regions. A polymer model of the chromosome suggested that plectoneme-free regions produce boundaries because their decompaction physically separates flanking domains and because they block the diffusion of supercoils, a primary driver of short-range interactions.

Hi-C has also driven new insights into how NAPs contribute to chromosome organization in vivo. Hi-C analysis of cells lacking HU showed that it specifically bolsters DNA interactions at short-length scales but without impacting CID formation, consistent with biophysical studies indicating that HU can bend and compact DNA in vitro [5**]. By contrast, Hi-C analysis indicated that the structural maintenance of chromosomes (SMC) protein, a homolog of eukaryotic cohesin and condensin [39,40], does not impact short-range interactions, and instead promotes the colinearity of chromosomal arms. Whether SMC physically tethers the arms together or somehow acts locally on individual arms to keep them in register remains unclear. Collectively, these Hi-C studies are providing important new insight into how specific proteins and other factors, such as supercoiling, contribute to chromosome organization at different length-scales.

### Concluding remarks

Recent advances in microscopy and the advent of new genomic techniques such as Hi-C have opened new windows into the organization and structure of chromosomes inside bacterial cells. The combination of population-averaged biochemical methods with single-cell microscopy studies, as well as classical molecular genetic studies, is revealing both global and local patterns of organization. Despite great progress, there is still much to be learned, and we anticipate that efforts to improve the spatial and temporal resolution of both imaging techniques and Hi-C approaches will spur new insights and answer unsolved questions about the structure, dynamics, and functioning of bacterial chromosomes.

### Acknowledgements

This work is supported by NIH grant R01GM082899 (M.T.L.) and by the Gordon and Betty Moore Foundation and a Life Sciences Research Foundation fellowship GBMF2550.02 (T.B.K.L.).

### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


3. Early use of an orthogonal ParB/parS system to label and track DNA loci in live cells. Revealed a strikingly different organization of chromosomes in slow-growing *E. coli* compared to other rod-shaped bacteria.


5. Examined the chromosome conformation of fast-growing *E. coli* cells using ParB/parS systems and showed that different parts of the chromosome reside in different shells of the nucleoid.


8. First use of Hi-C to produce a high-resolution DNA interaction map, revealing ~23 chromosomal interaction domains (CID). Demonstrated that highly-expressed genes are critical to domain boundary formation, and illustrated the power of Hi-C for investigating the in vivo functions and roles of SMC and NAPs.


10. Used genome-wide chromosome conformation capture (SC) to investigate bacterial chromosome structure. Together with live imaging and computational modeling, produced 3D models of the Caulobacter chromosome showing intertwined arms.


First report describing macrodomains in the E. coli chromosome based on FISH studies.


Using fluorescent repressor-operator systems, this paper (along with ref. 2) reported the general pattern of chromosome organization for slow-growing E. coli.


Used fluorescent repressor-operators systems to show that 112 different chromosomal loci reside at specific locations and in the same approximate order along the long axis of the cell as in the genome sequence. Newly-replicated DNA loci segregate sequentially to their final locations in the growing cell.


Employed the Int recombination system in E. coli to reveal four macrodomains and two less-structured (NS) regions, extending the prior discovery of the Ori and Ter macrodomains by FISH.


The yqf Res recombination system was used to show that a region with high rates of transcription can form a barrier to plasmid invasion.


