Maximum Entropy Model for Replicating Bacterial Chromosomes

Janni Harju†
Department of Physics
Princeton University
Princeton, NJ, USA
jh2214@princeton.edu

Joris J.B. Messelink†
Department of Physics
Ludwig-Maximilian-University
Munich, Germany
joris.messelink@gmail.com

Lucas Tröger
Department of Physics
Ludwig-Maximilian-University
Munich, Germany
lucastroeger@googlemail.com

Gregorz Gradziuk
Department of Physics
Ludwig-Maximilian-University
Munich, Germany
ggradziuk@gmail.com

Imesha Rathnayaka
Department Microbial Interactions
Friedrich-Schiller-Universität
Jena, Germany
imesha.rathnayaka@cup.lmu.de

Maria Billini
Department of Biology
University of Marburg
Marburg, Germany
billini@staff.uni-marburg.de

Martin Thanbichler
Department of Biology
University of Marburg
Marburg, Germany
thanbichler@uni-marburg.de

Muriel C.F. van Teeseling
Department Microbial Interactions
Friedrich-Schiller-Universität
Jena, Germany
murielvanteeseling@gmail.com

Chase P. Broedersz
Department of Physics and Astronomy
Vrije Universiteit
Amsterdam, Netherlands
c.p.broedersz@vu.nl

ABSTRACT

Chromosome conformation capture experiments (Hi-C) measure genome-wide spatial proximity between pairs of chromosomal loci, yielding contact frequency matrices that encode the statistical ensemble of three-dimensional chromosome configurations adopted by a population of cells. Extracting a meaningful, minimally biased model of chromosome organization from these data is a canonical statistical inference problem: given a set of empirical pairwise constraints, what is the least-assumptive probability distribution over chromosome configurations consistent with those constraints? The Maximum Entropy (MaxEnt) principle provides a rigorous approach for constructing such models: given a set of experimental constraints, we model the system using the distribution of states that maximizes the Shannon entropy.

Applying the MaxEnt principle to chromosome organization becomes substantially more difficult when the biological system is dynamic. Bacterial chromosomes are in continual motion, undergoing concurrent transcription, replication, and segregation throughout the cell cycle. Most Hi-C experiments measure average contact counts across populations of cells at different replication stages. Furthermore, Hi-C data from replicating systems conflate contacts between loci on the same DNA strand (*cis*) and on different replicated strands (*trans*), an ambiguity that has not previously been addressed in principled frameworks.

Here, we introduce a **4D Maximum Entropy (4D-MaxEnt)** approach that resolves these challenges to extract a time-resolved model of chromosome organization directly from time-course Hi-C and microscopy data, without imposing a structural model *a priori*. By applying the MaxEnt principle at each experimentally sampled time-point, we obtain the least-biased distribution of three-dimensional chromosome configurations across the bacterial cell cycle. We explicitly account for *cis/trans* contact ambiguity using microscopy-derived information on the positions of replication origins, which we incorporate as additional observational constraints within the MaxEnt framework.

We apply 4D-MaxEnt to *Caulobacter crescentus*, a bacterium whose cell cycle can be tightly synchronized, making it an ideal system for time-course Hi-C. After validating our inferred ensemble against independent microscopy data not used in model construction, we use the framework to extract quantitative, time-resolved information about chromosome organization across the replication cycle, including large-scale linear ordering of chromosomal regions, and local changes in compaction in the vicinity of replication forks. These data-driven inferences are then used to constrain and discriminate between mechanistic hypotheses for chromosome organization, linking the abstract statistical inference to concrete biological mechanisms.

Beyond the specific biological application, our work demonstrates how the MaxEnt principle provides a general, principled framework for extracting dynamic structural information from population-averaged, high-dimensional data, a challenge that arises broadly in biology whenever ensemble measurements must be interpreted in terms of an underlying distribution of configurations.

---

† These authors contributed equally.