

## 4-D Model of Bacterial Chromosome Structure

Matthew Wright  
Harvard Medical School  
200 Longwood Avenue  
Boston, MA 02115

mwright@genetics.med.harvard.edu

Daniel Segre  
Harvard Medical School  
200 Longwood Avenue  
Boston, MA 02115

dsegre@genetics.med.harvard.edu

George Church  
Harvard Medical School  
200 Longwood Avenue  
Boston, MA 02115

church@arep.med.harvard.edu

### ABSTRACT

Massive effort has been devoted to understanding biopolymer structure at the nanometer and micrometer scales. Very little success has been achieved studying their structure at intermediate scales. We outline a method to integrate data on the three dimensional structure of whole chromosomes and genomes. The method is based on evolutionary optimality. We hypothesize that the three dimensional structure of the chromosome is optimal for many of the functions that it performs. These functions determine geometrical and dynamical constraints that can be expressed mathematically. Our goal is to gather enough constraints to create a 4D model of the chromosome. Changes in the structure of the chromosome over time (the 4<sup>th</sup> dimension) will reflect the varying functional constraints on the genome during different phases of the cell cycle.

We model the bacterium *Mycoplasma pneumoniae*. It is nearly a minimal cell with a genome that is 816 kbp long and only 688 genes. It has limited metabolism, no known regulation, and very few DNA binding proteins [1]. Since it is so simple, we expect chromosome structure to be under strong evolutionary selection. We use transcriptional efficiency to motivate two sets of constraints for our model. First, we expect most transmembrane genes to be close to the membrane [2] and second we anticipate that ribosome component genes can be spatially colocalized (as they are in nucleoli). Both sets of constraints reflect the efficiency of producing proteins near their targets. Additional constraints are provided by experimental observations of bacterial nucleoids and considerations of chromosome replication. Microscopy shows that the origin and terminus of replication are localized to opposite poles of a bacterium during most of the cell cycle [3]. We constrain these regions to opposite poles in our model. Furthermore, we use a model of replication that requires proximity of regions that are symmetrically opposite the origin of replication [4-7]. This symmetry is enforced in our model.

Using this set of constraints we write a cost function that reflects the degree to which the constraints are met by a three dimensional structure. This cost function is then used to find optimal structures.

We use a combination of algorithms to search the structure space. One algorithm (HeliChrom) performs random walks in parameter space for a class of supercoiled helices. The parameters include helical rise, supercoil frequency, helix amplitude, supercoil sine and cosine amplitudes, and major helix envelope frequency. These parameters allow a wide diversity of structures from simple helices to very complex structures. We begin with randomly generated initial structures then randomly

change parameters to generate test structures for each iteration. The structures that result after many iterations of Montecarlo sampling are of significantly lower energy than the initial structures, and cluster into a small number of structure kinds.

The strict order imposed on these parameterized helical structures is relaxed by allowing local optimization through our second algorithm (BestChrom). The genome is discretized into N points and a random displacement of the structure is attempted at each iteration. Through Montecarlo sampling the structure is refined and the cost is further minimized. The structures obtained after several iterations maintain to some extent the overall low level of entanglement obtained through HeliChrom, but may display highly complex local organization.

The final structures can be compared with each other using rotationally invariant representations. We are in the process of calculating statistics from structure comparisons to find conserved features. We intend to use statistics from optimizations of genomes with randomly generated constraints to evaluate the significance of the constraints used in the cost function.

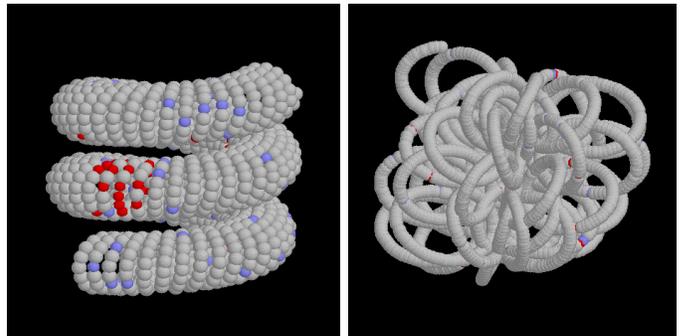


Figure 1a simple supercoil

Figure 1b complex supercoil

The model generates testable hypotheses. It predicts sets of point-point and point-membrane distances which are not initial constraints. These points can be visualized using arrays of lac operons and GFP lac fusions [3]. Alternatively, intramolecular circularization ligation, [8] can be used to test multiple conformational constraints per chromosomal molecule.

It is important to note that the constraints implemented in the cost function will probably require update as new experimental and theoretical data become available. New kinds of constraints may be incorporated as well. Such constraints could be derived from other protein complexes such as transcription machinery and carboxysomes, or from the structure of enzyme

factories [9-10] as well as from quantitative proteomic studies. We are currently exploring theoretical models of the dynamics of protein complex formation in order to better understand their dependence on the spatial distribution of the component genes. An important next step will be to incorporate replication and dynamics into the chromosome model. We envision an interplay of prediction and experimental testing to improve the quality of the model.

We believe that this work introduces an important new strategy for modeling large-scale structures in the cell and potentially provides a starting point for future cellular level studies based on evolutionary optimality.

## REFERENCES

- [1] Dandekar T, Huynen M, Regula JT, Ueberle B, Zimmermann CU, Andrade MA, Doerks T, Sanchez-Pulido L, Snel B, Suyama M, Yuan YP, Herrmann R, Bork P. Re-annotating the *Mycoplasma pneumoniae* genome sequence: adding value, function and reading frames. *Nucleic Acids Res.* 2000 Sep 1;28(17):3278-88
- [2] Potter MD, Nicchitta CV. Endoplasmic reticulum-bound ribosomes reside in stable association with the translocon following termination of protein synthesis. *J Biol Chem.* 2002 Jun 28;277(26):23314-20
- [3] Jensen RB, Shapiro L. Chromosome segregation during the prokaryotic cell division cycle. *Curr Opin Cell Biol.* 1999 Dec;11(6):726-31
- [4] Dingman CW. Bidirectional chromosome replication: some topological considerations. *J Theor Biol.* 1974 Jan;43(1):187-95
- [5] Lemon KP, Grossman AD. Movement of replicating DNA through a stationary replisome. *Mol Cell.* 2000 Dec;6(6):1321-30
- [6] Sundin O, Varshavsky A. Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. *Cell.* 1980 Aug;21(1):103-14
- [7] Hearst JE, Kauffman L, McClain WM. A simple mechanism for the avoidance of entanglement during chromosome replication. *Trends Genet.* 1998 Jun;14(6):244-7
- [8] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science.* 2002 Feb 15;295(5558):1306-11
- [9] Bouffard GG, Rudd KE, Adhya SL. Dependence of lactose metabolism upon mutarotase encoded in the gal operon in *Escherichia coli*. *J Mol Biol.* 1994 Dec 2;244(3):269-78
- [10] Huang X, Holden HM, Raushel FM. Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annu Rev Biochem.* 2001;70:149-80