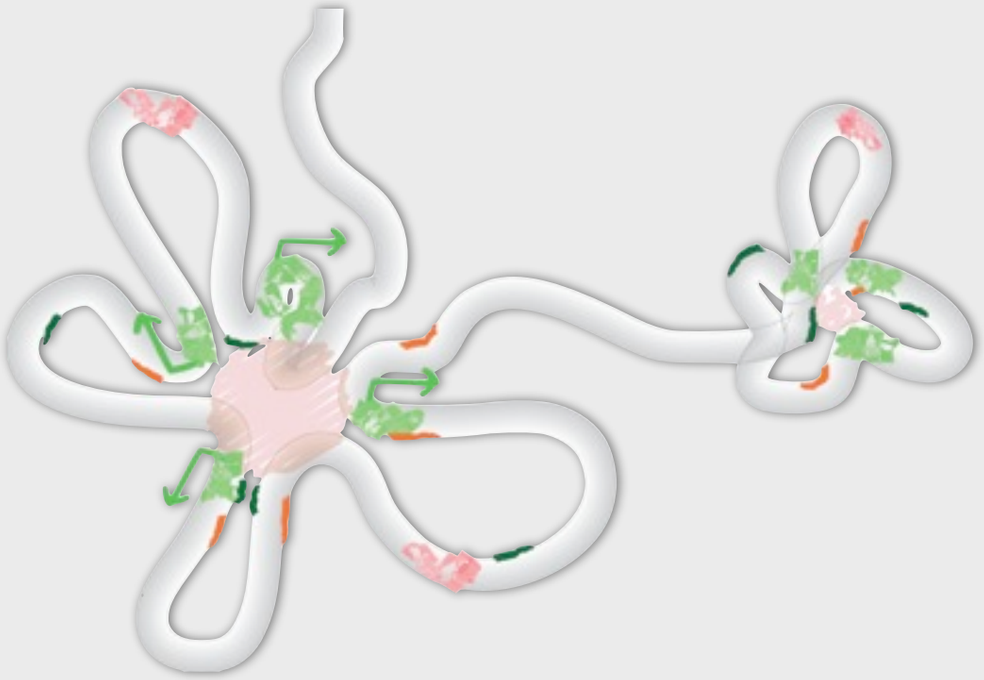
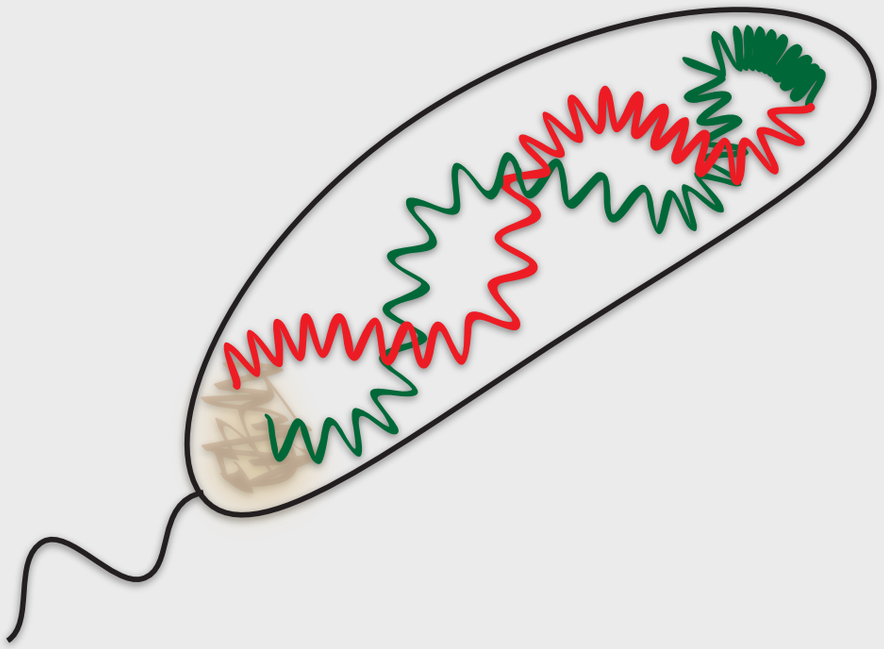
Structure determination of genomes and genomic domains by satisfaction of spatial restraints





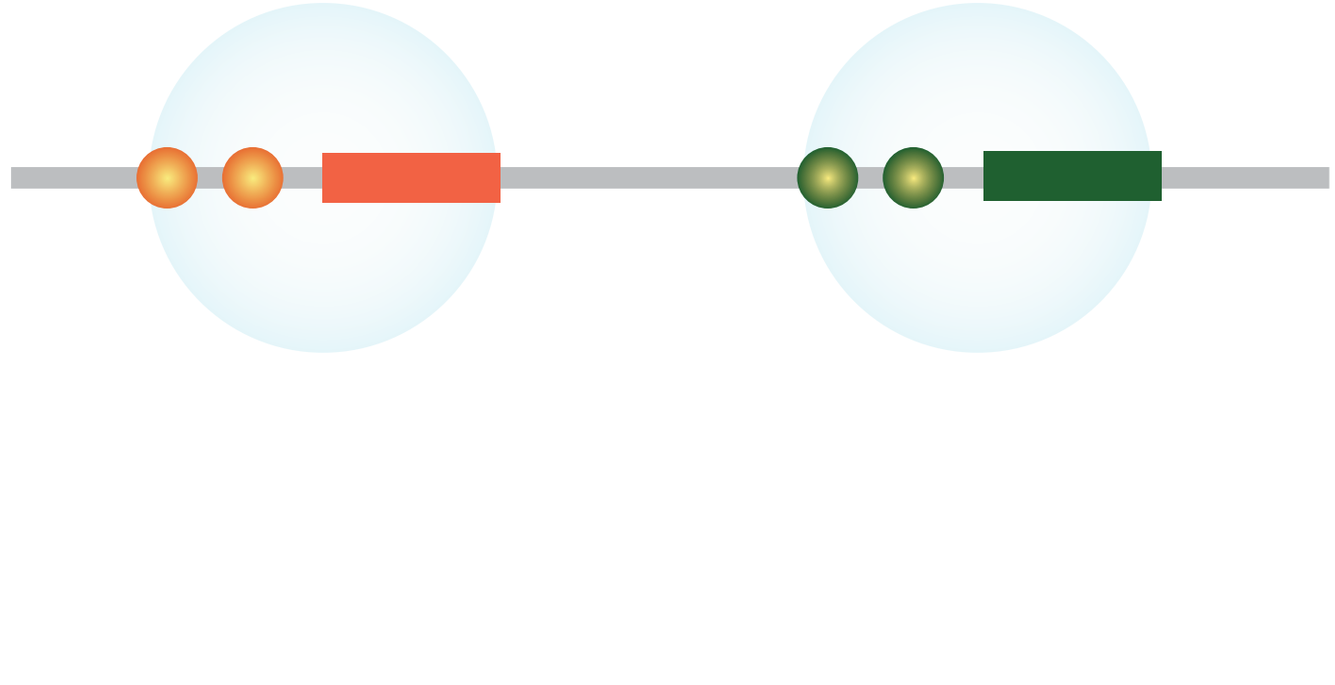
Marc A. Marti-Renom

Structural Genomics Laboratory
Bioinformatics & Genomics Department
Prince Felipe Research Center (CIPF), Valencia, Spain

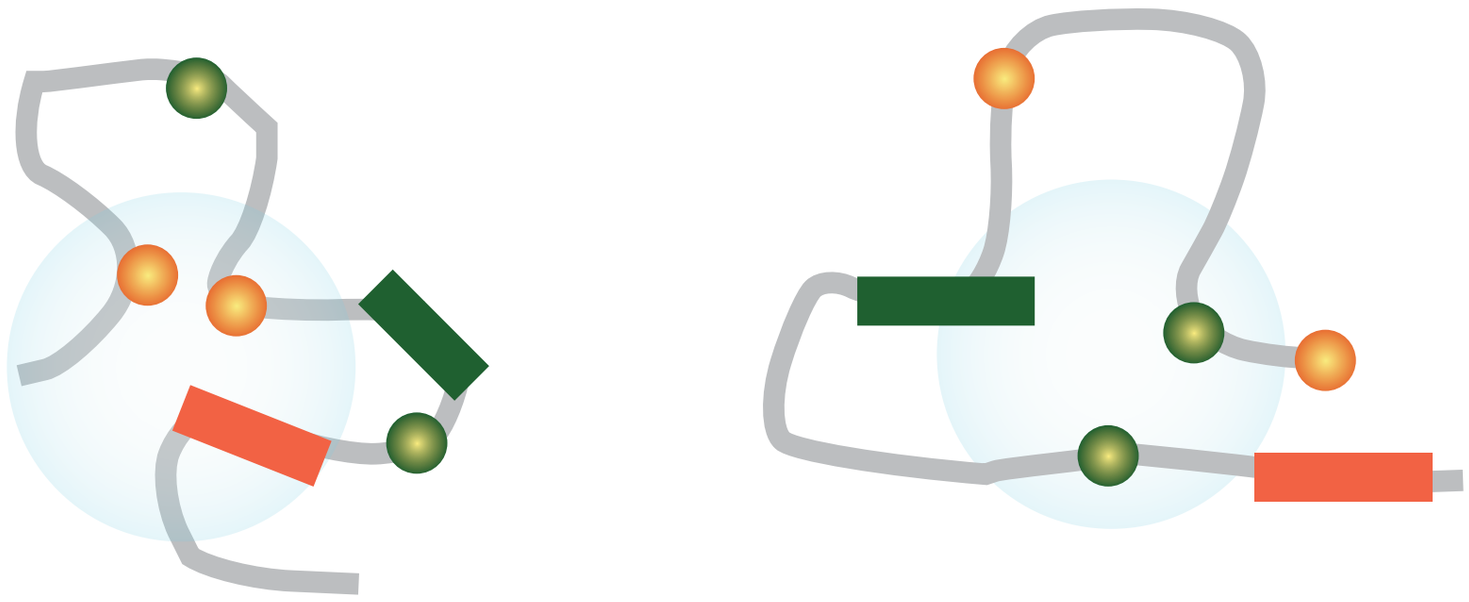


http://sgu.bioinfo.cipf.es

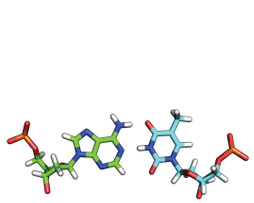
Simple genomes

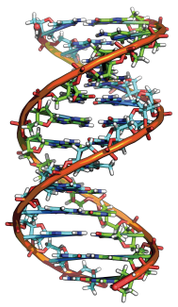


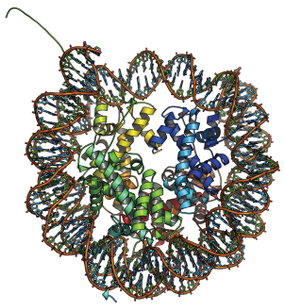
Complex genomes

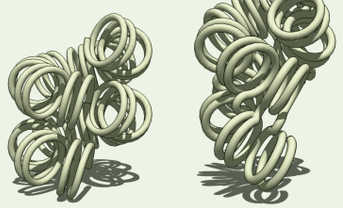


Knowledge



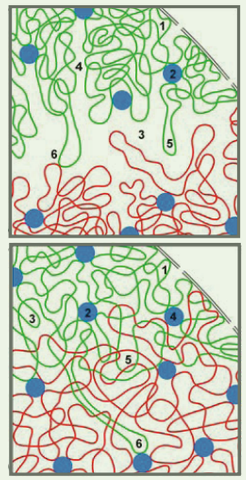


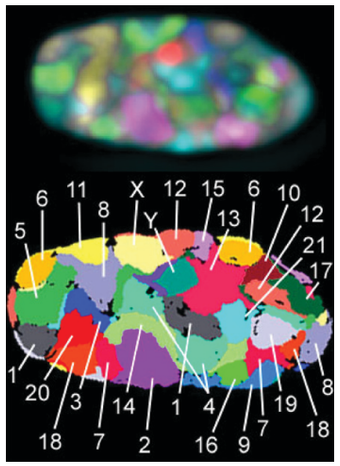




IDM

INM





DNA length

10^0 10^3 10^6 10^9 nt

Volume

10^{-9} 10^{-6} 10^{-3} 10^0 10^3 μm^3

Time

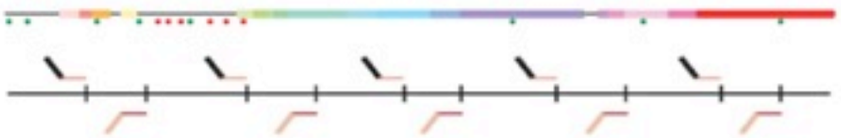
10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2} 10^0 10^2 10^3 s

Resolution

10^{-3} 10^{-2} 10^{-1} µ

Adapted from:
Langowski and Heermann. Semin Cell Dev Biol (2007) vol. 18 (5) pp. 659-67

Experiments



Grow GM12878 and K562 cells

↓

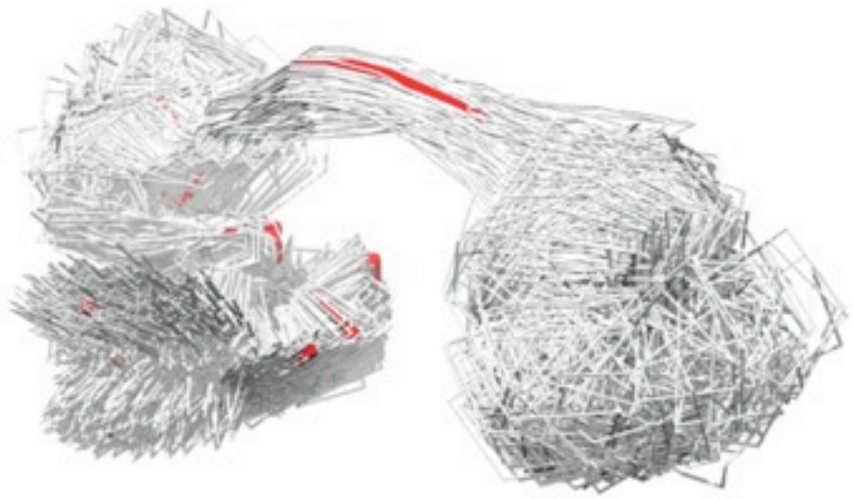
Perform 3C analysis

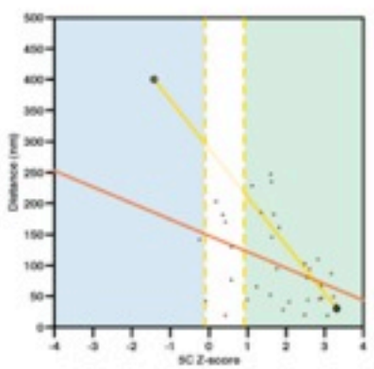
↓

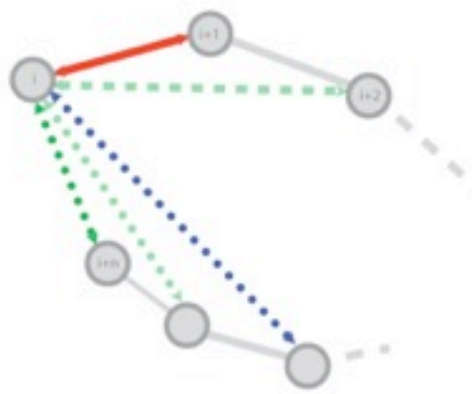
Perform 5C analysis with 30+25 primers

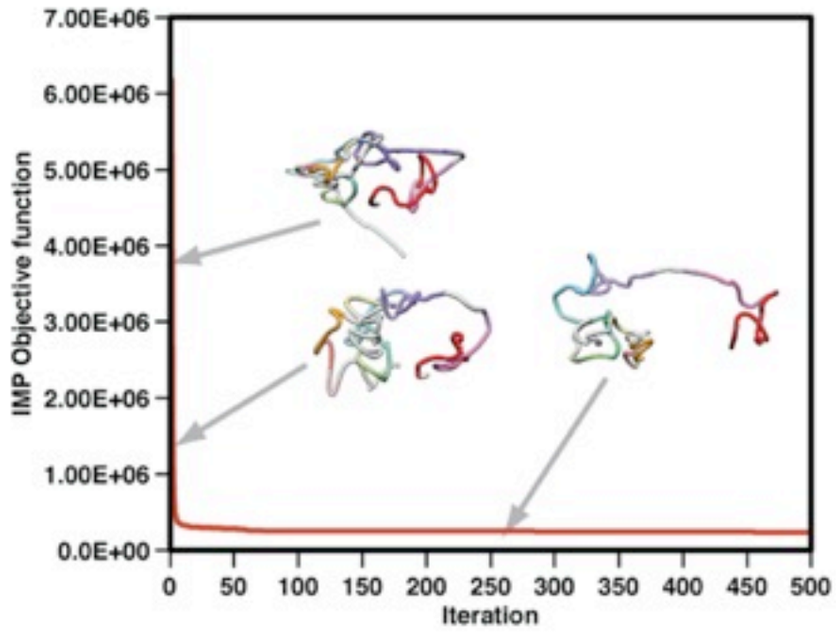
↓

Analyze 5C products by paired-end Solexa sequencing
(131,947 paired end reads per library)



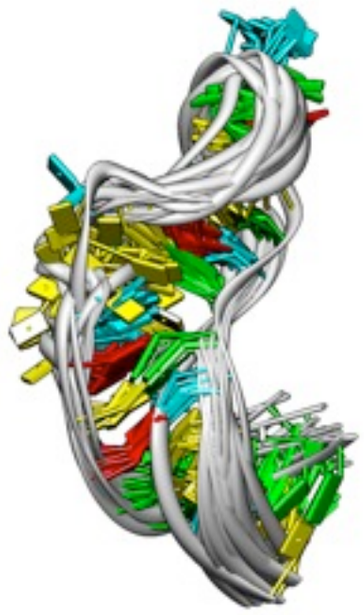


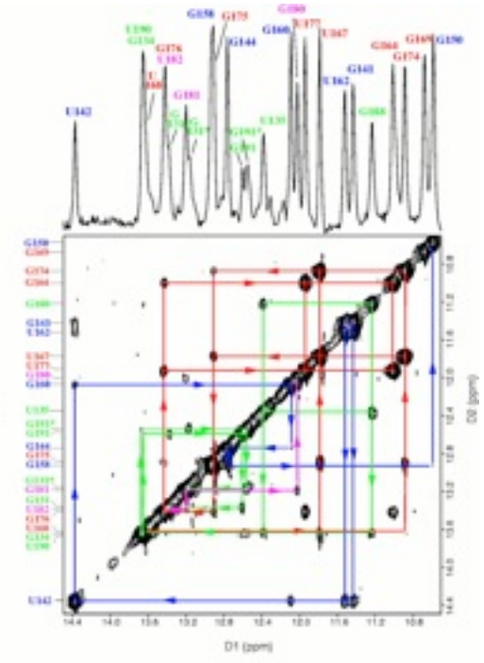




Computation

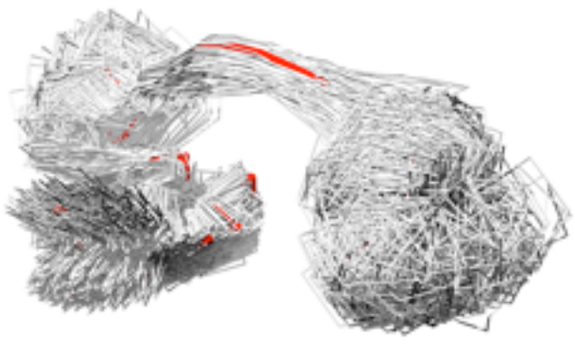
PRINCIPE FELIPE
CENTRO DE INVESTIGACION

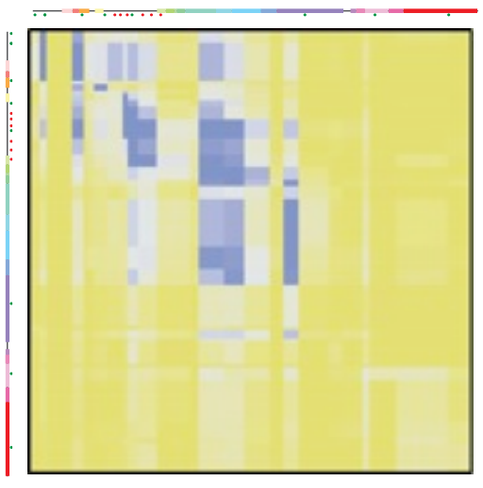




Biomolecular structure determination

2D-NOESY data





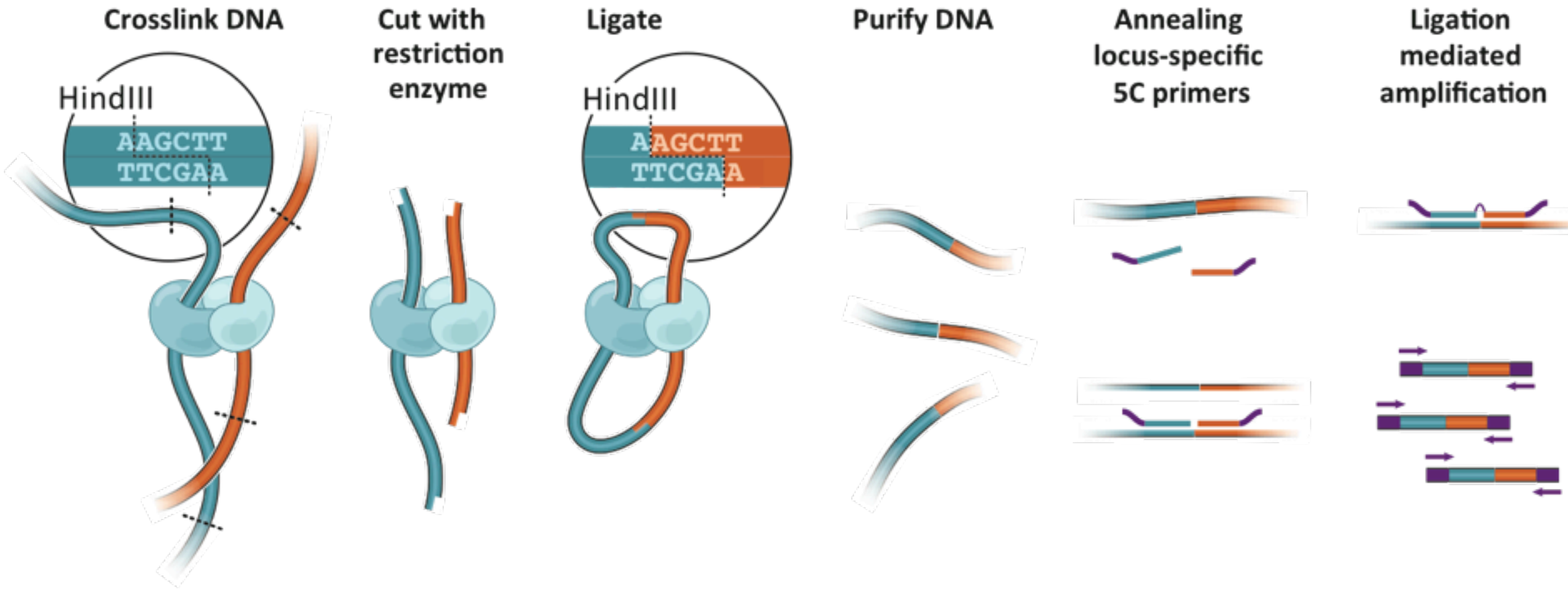
Chromosome structure determination

5C data

5C technology

http://my5C.umassmed.edu

Dostie et al. Genome Res (2006) vol. 16 (10) pp. 1299-309



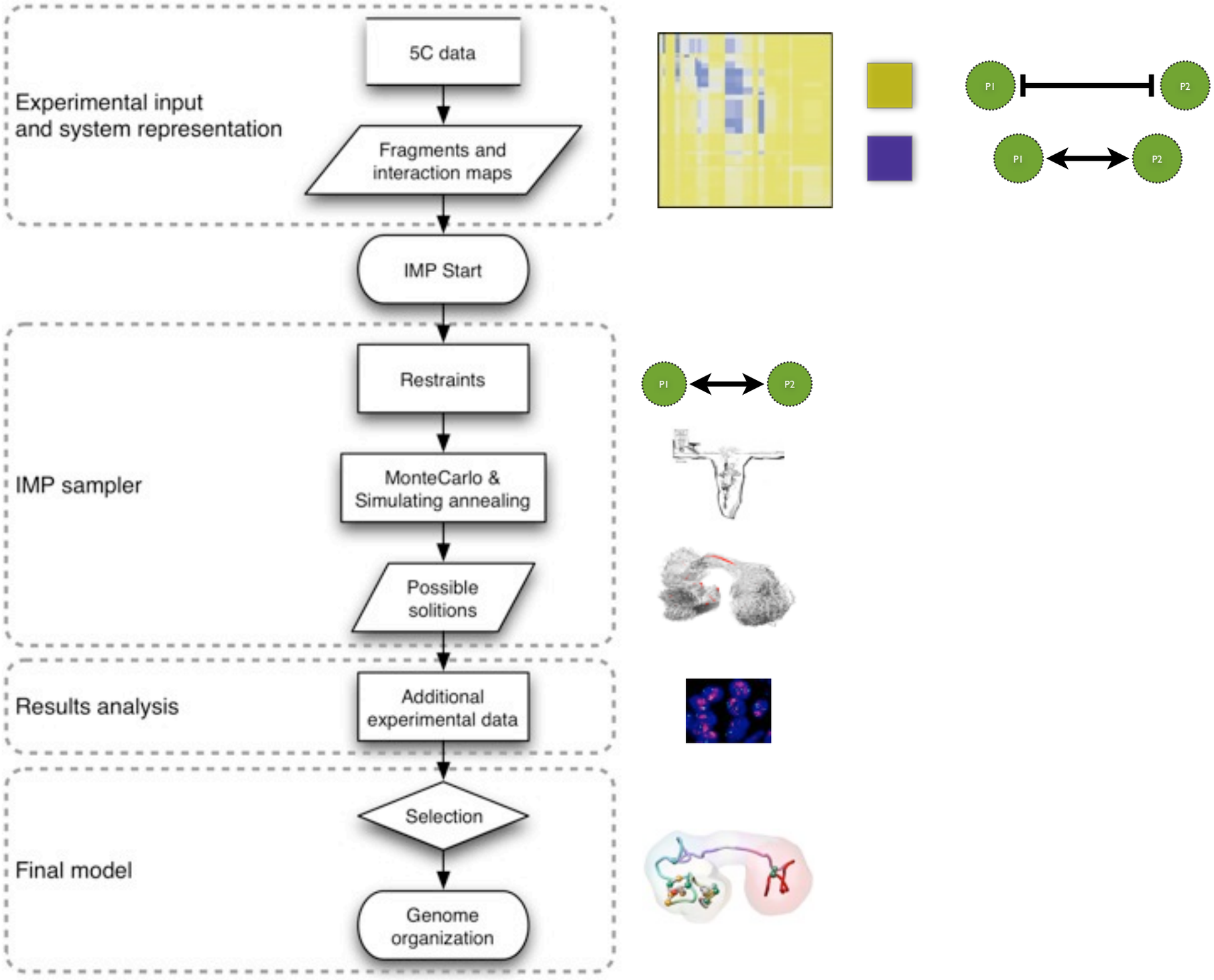
5C “copies” the 3C library into a 5C library containing only ligation junctions

Performed at high levels of multiplexing:

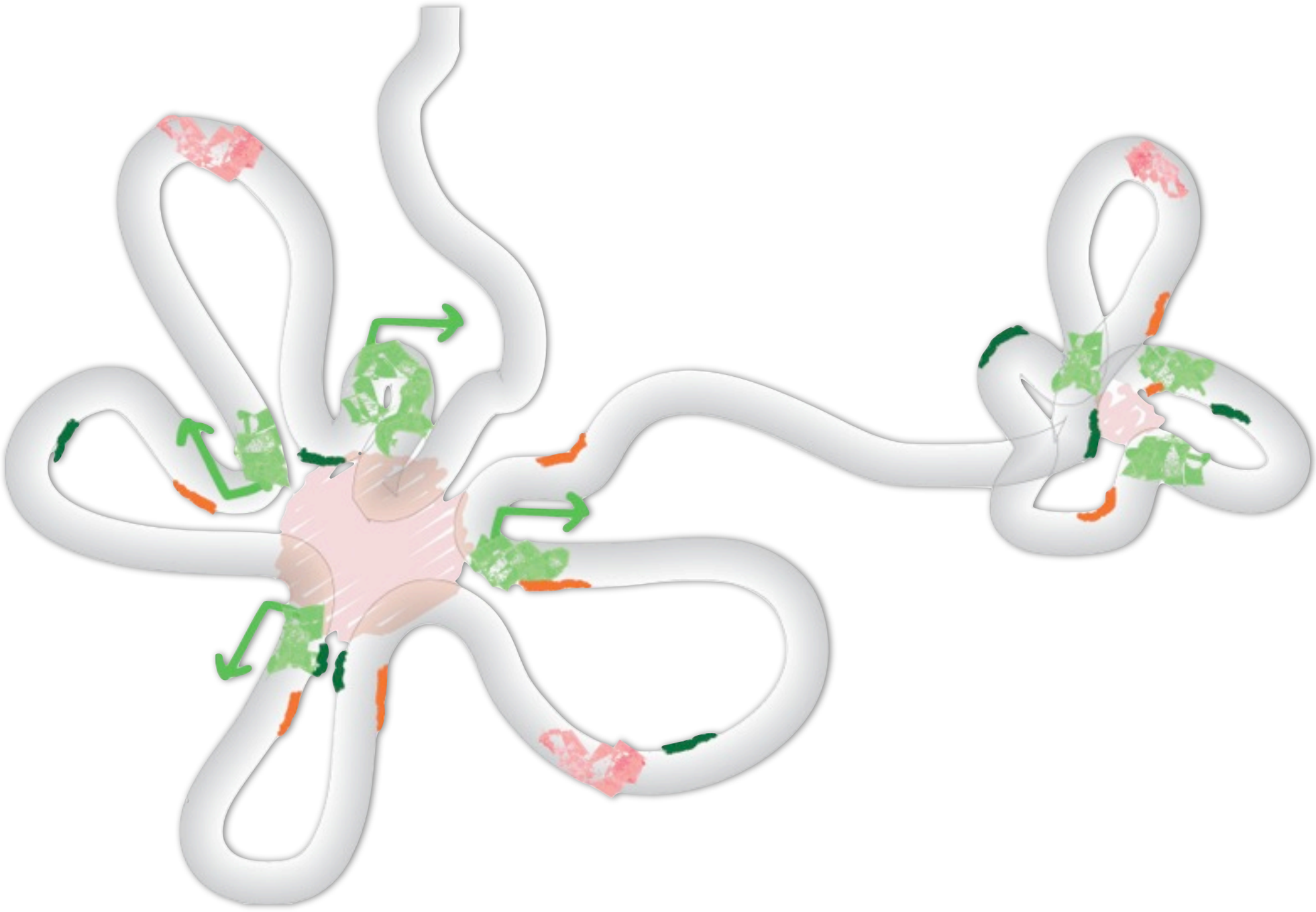
2,000 primers detect 1,000,000 unique interactions in 1 reaction

Integrative Modeling

http://www.integrativemodeling.org



Human α-globin domain



Human α-globin domain

ENm008 genomic structure and environment

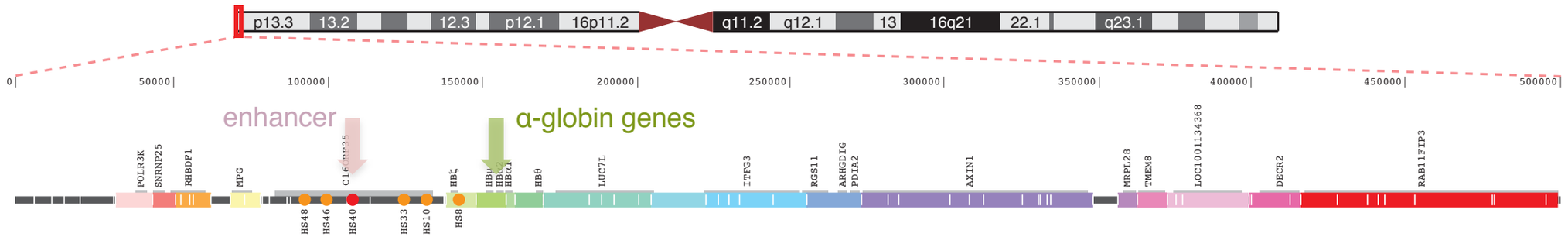
ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816

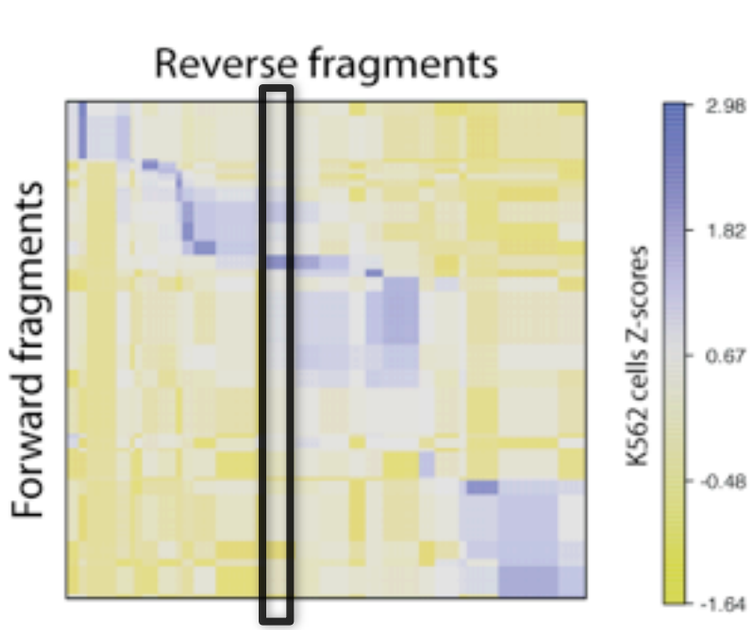


The ENCODE data for ENm008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute ChIP-seq (Bernstein Group at Broad Institute of Harvard and MIT).

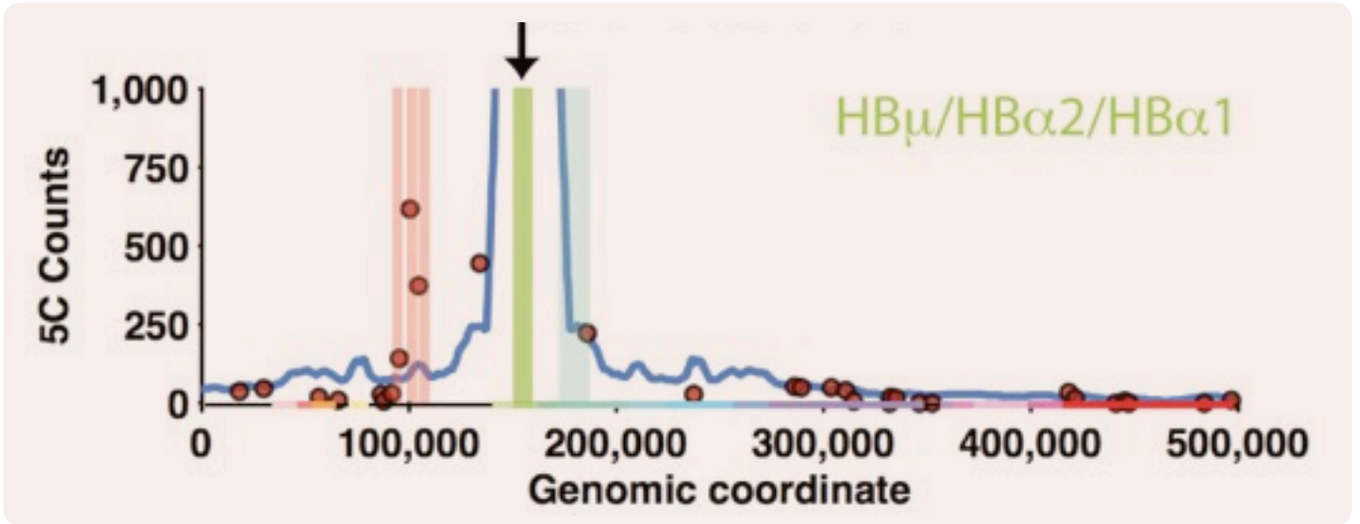
Human α-globin domain

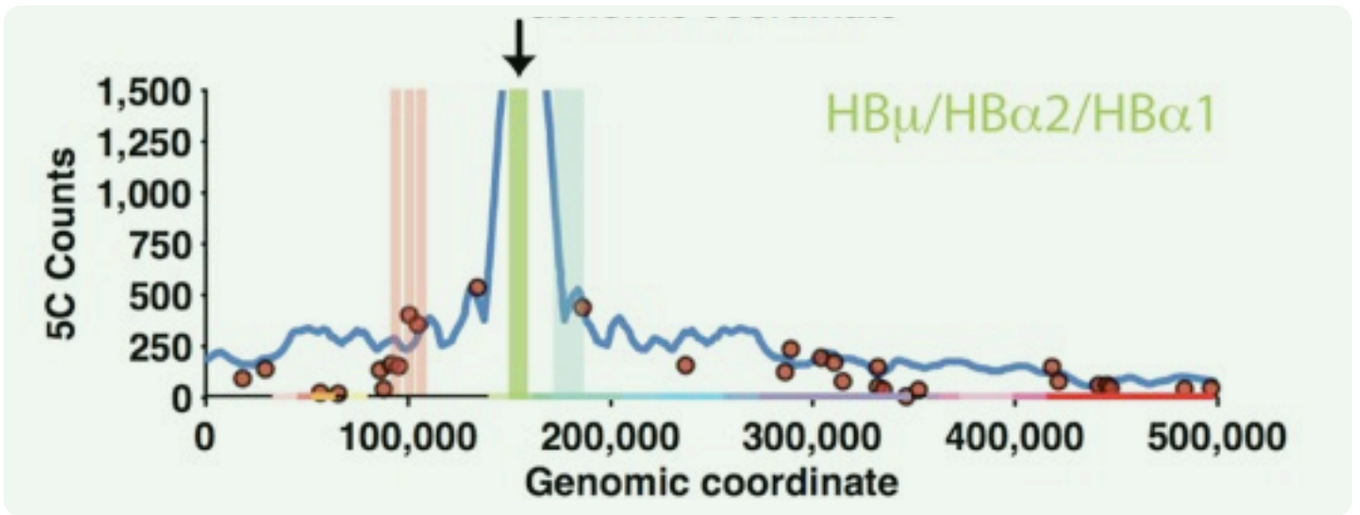
ENm008 genomic structure and environment





K562 cells:
α-globin genes active

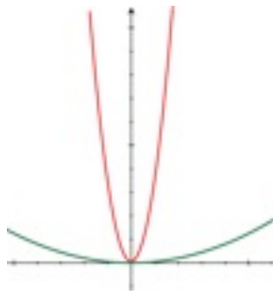




Representation

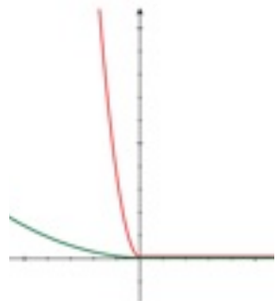
Harmonic

$$H_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2$$



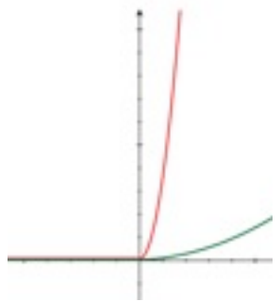
Harmonic Lower Bound

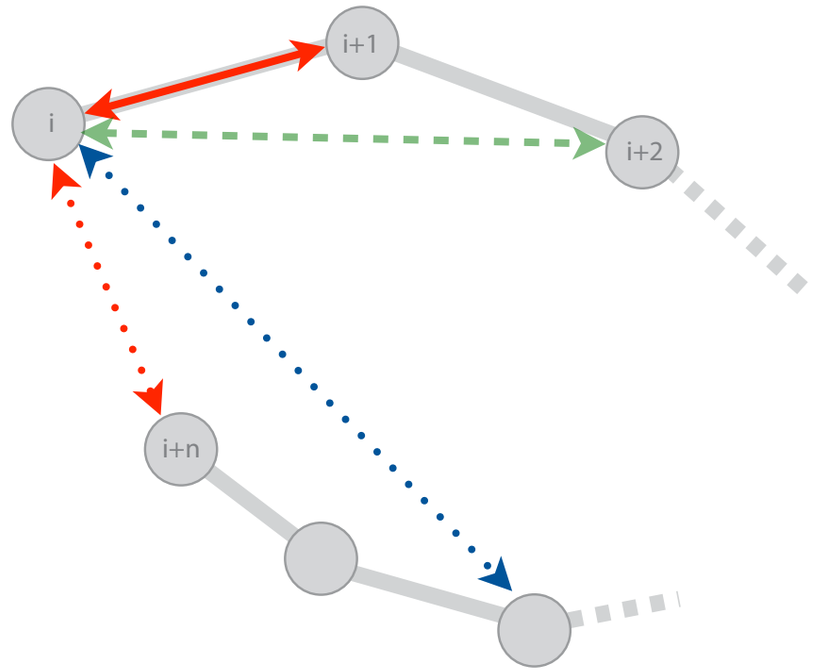
$$\begin{cases} if \;\; d_{i,j} \le d^0_{i,j}; & lbH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if \;\; d_{i,j} > d^0_{i,j}; & lbH_{i,j} = 0 \end{cases}$$



Harmonic Upper Bound

$$\begin{cases} if \;\; d_{i,j} \ge d^0_{i,j}; & ubH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if \;\; d_{i,j} < d^0_{i,j}; & ubH_{i,j} = 0 \end{cases}$$



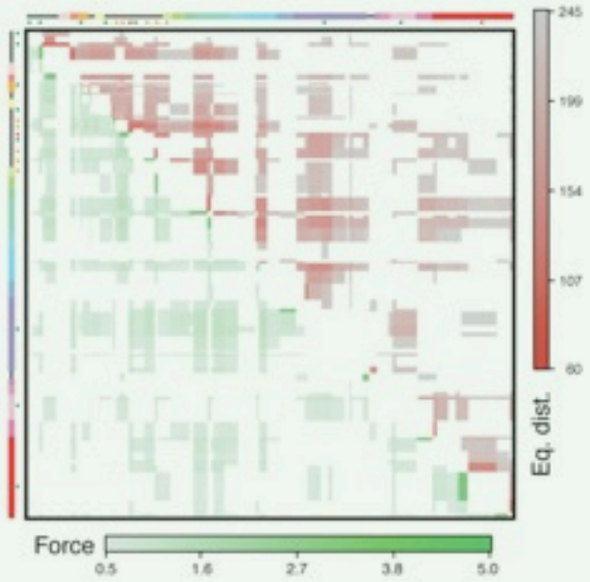


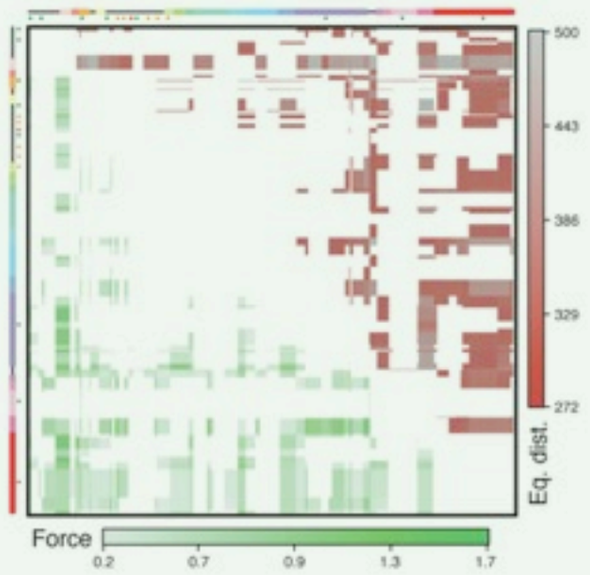
Scoring

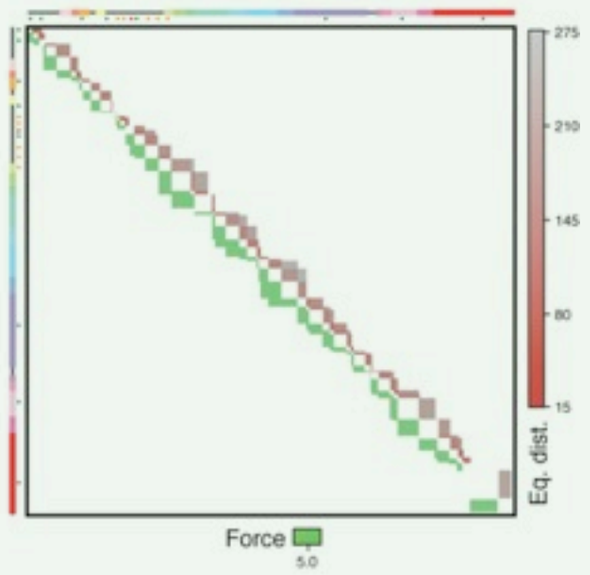
GM12878

70 fragments
1,520 restraints







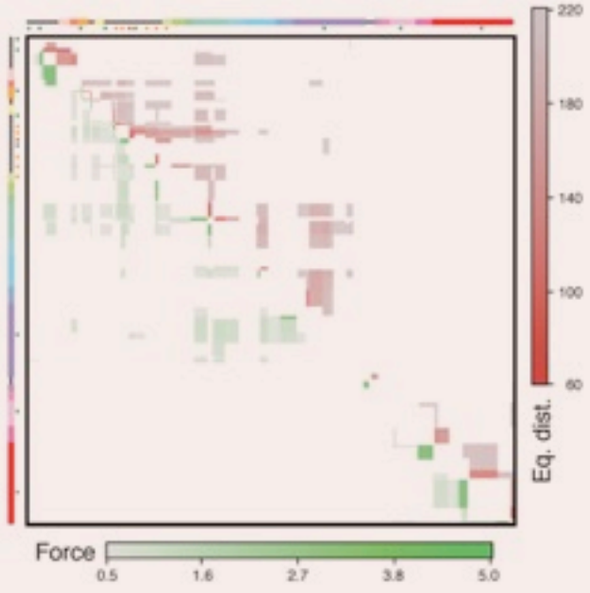


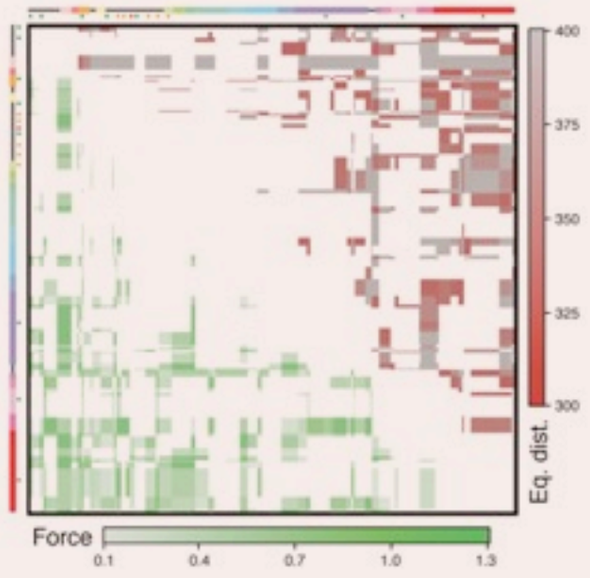
Harmonic | Harmonic Lower Bound | Harmonic Upper Bound

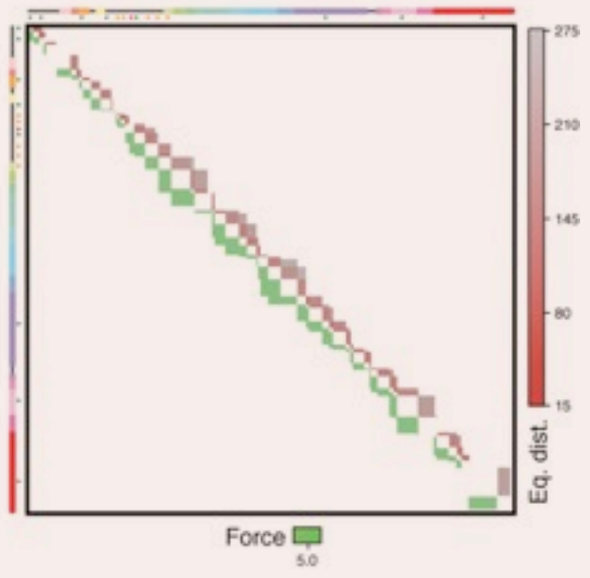
K562

70 fragments
1,049 restraints

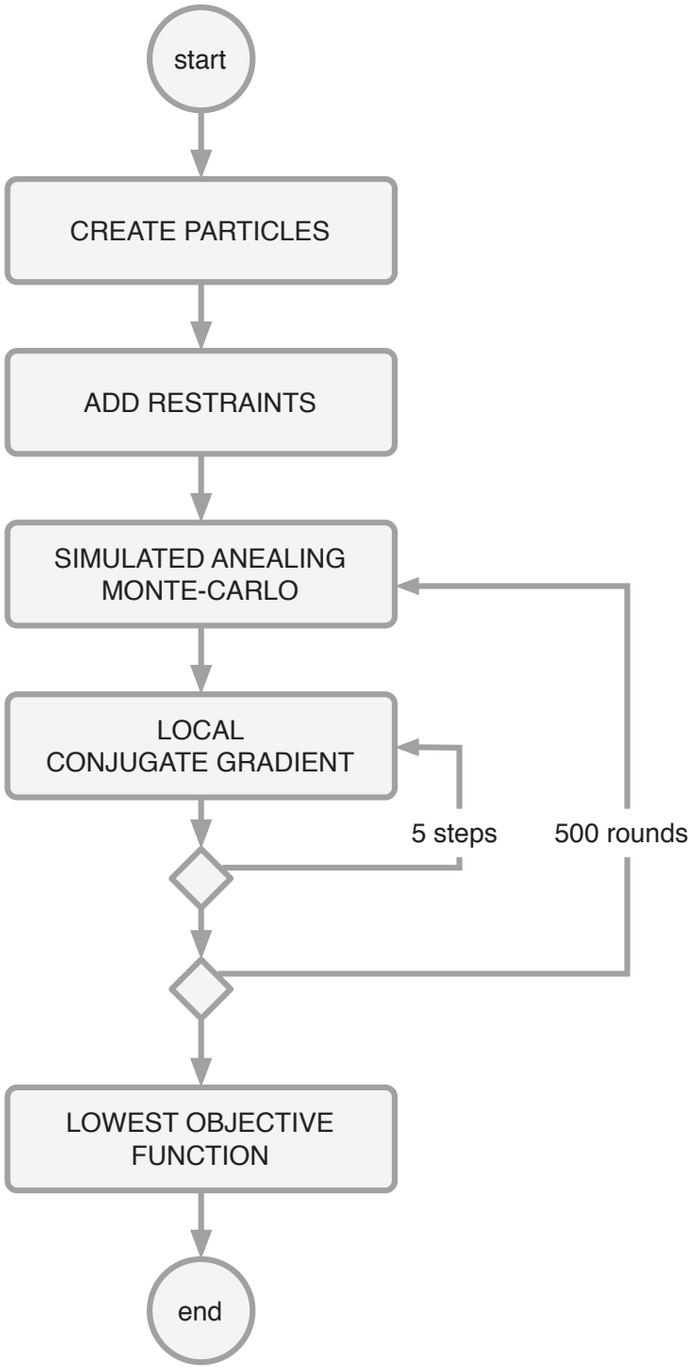


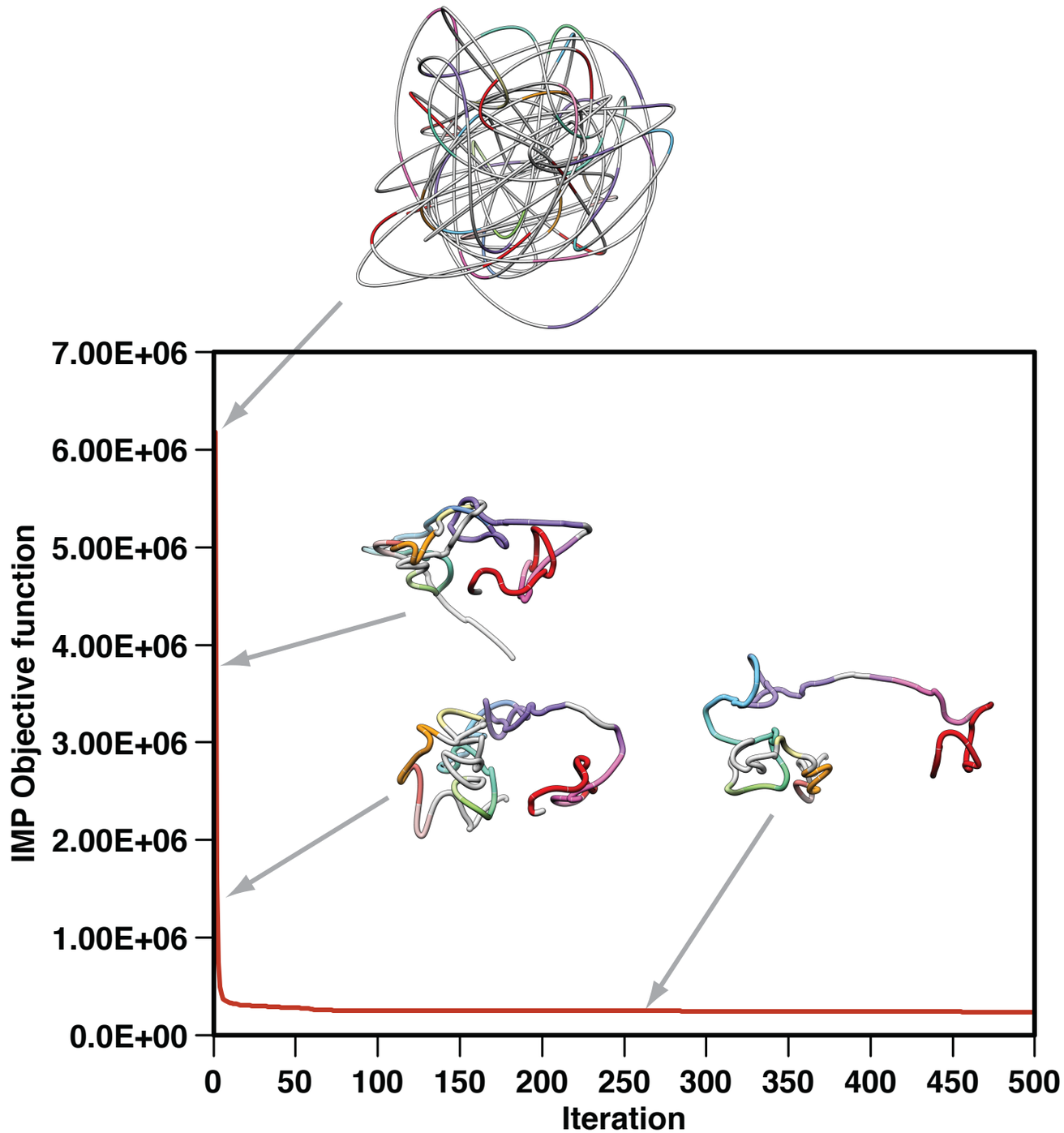




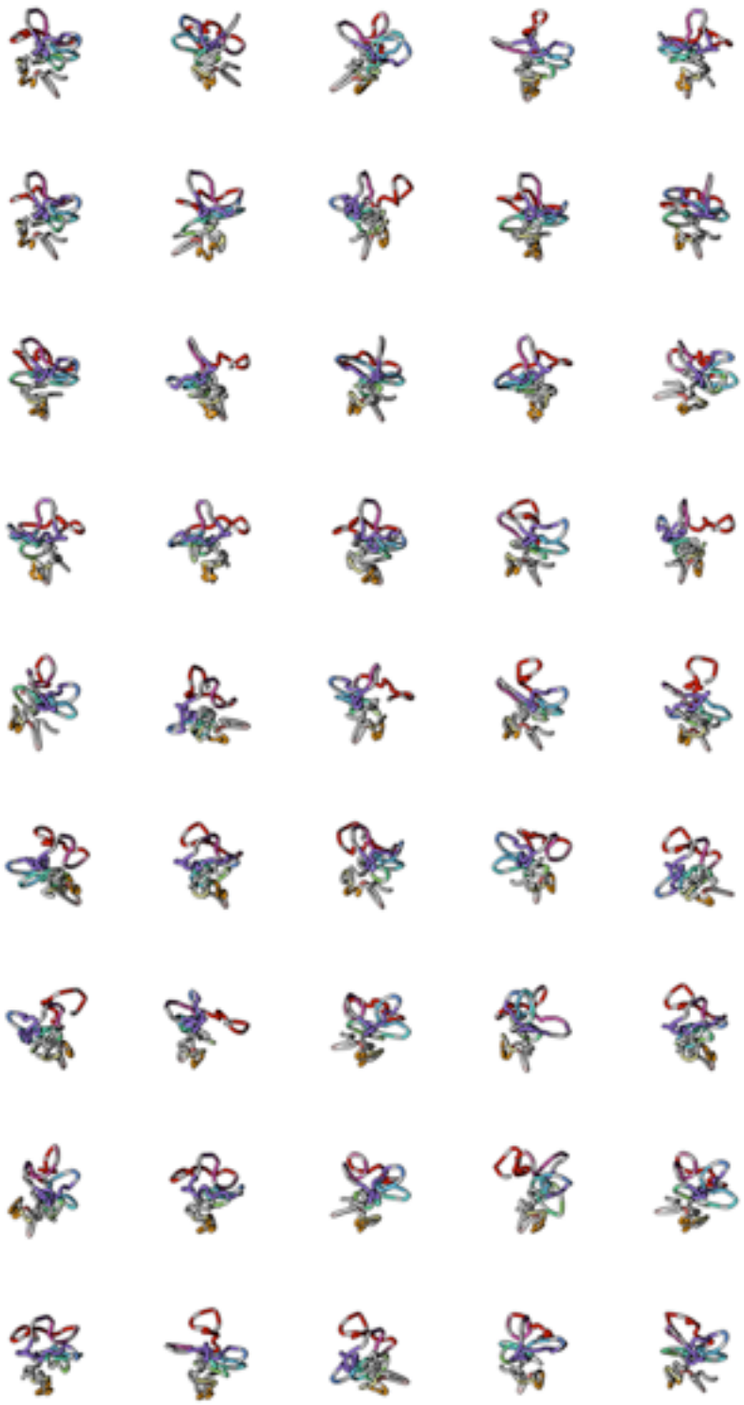


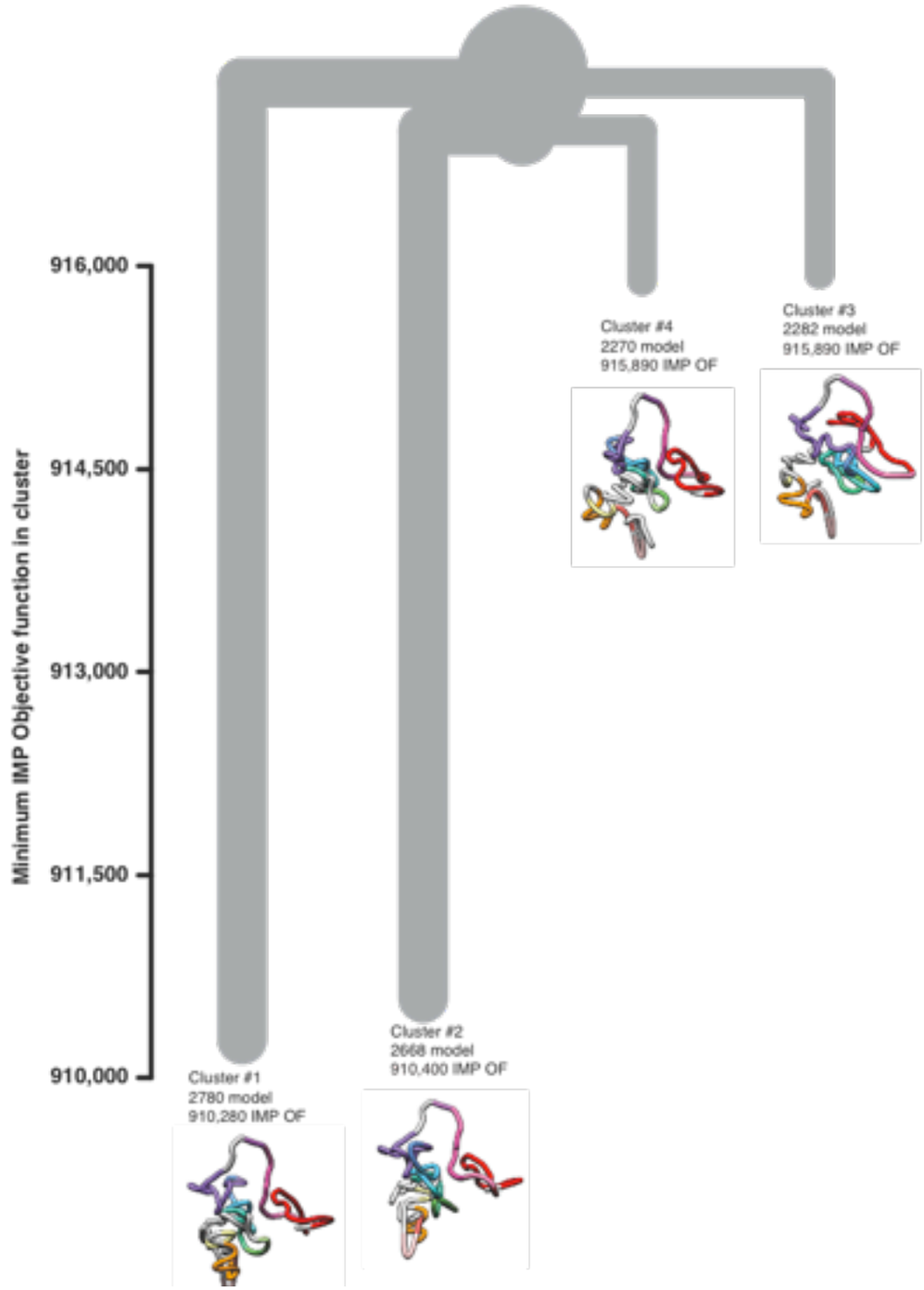
Optimization





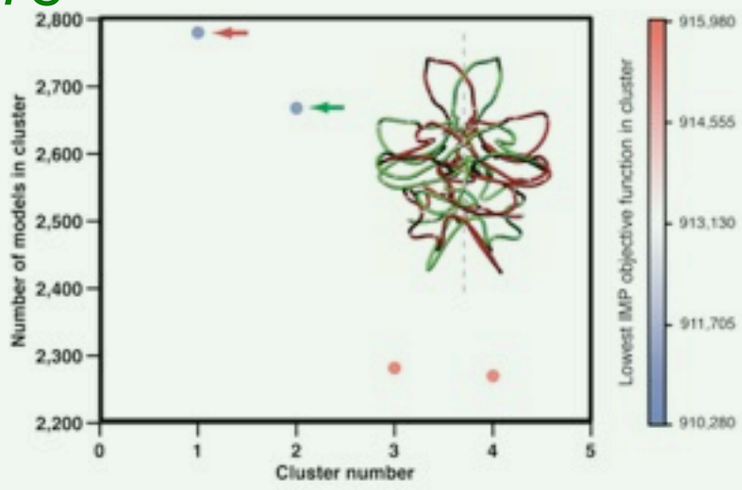
Clustering

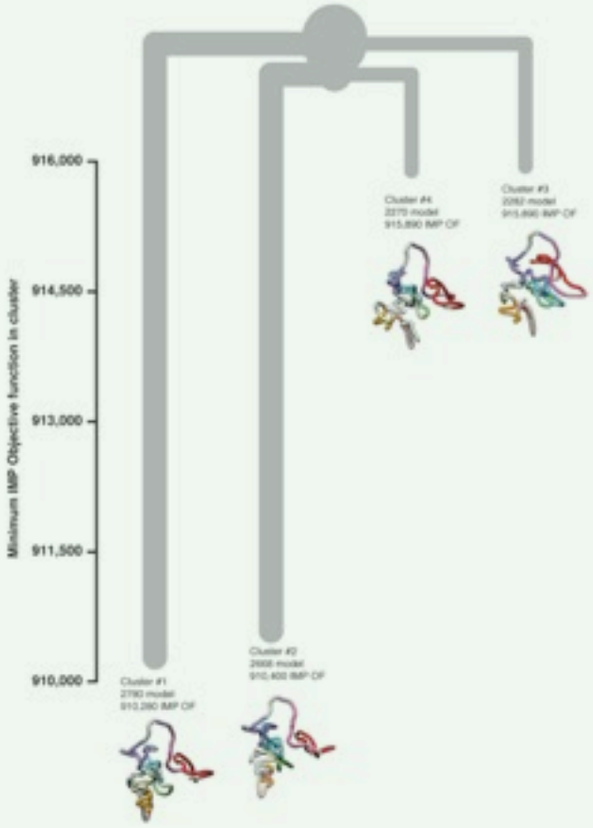



916,000
914,500
913,000
911,500
910,000
Minimum IMP Objective function in cluster
Cluster #4
2270 model
915,890 IMP OF
Cluster #3
2282 model
915,890 IMP OF
Cluster #1
2780 model
910,280 IMP OF
Cluster #2
2668 model
910,400 IMP OF

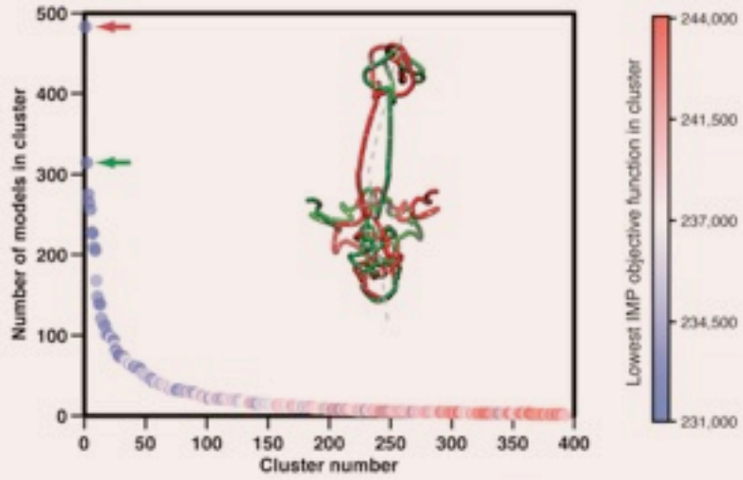
Not just *one* solution

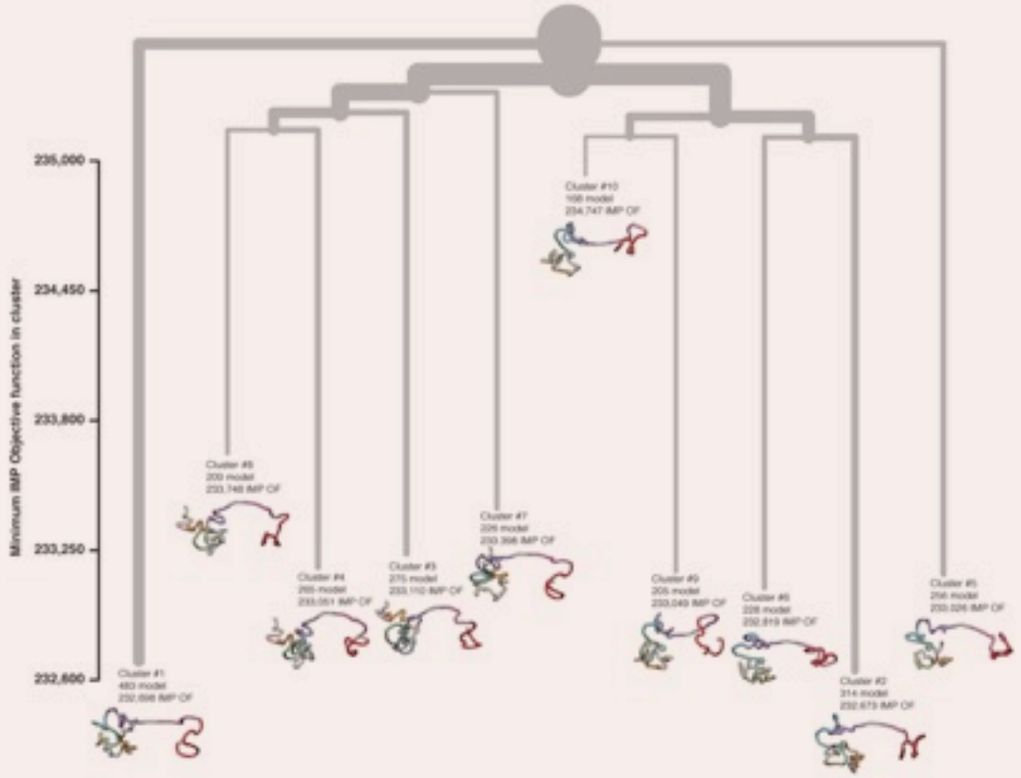
GM12878





K562

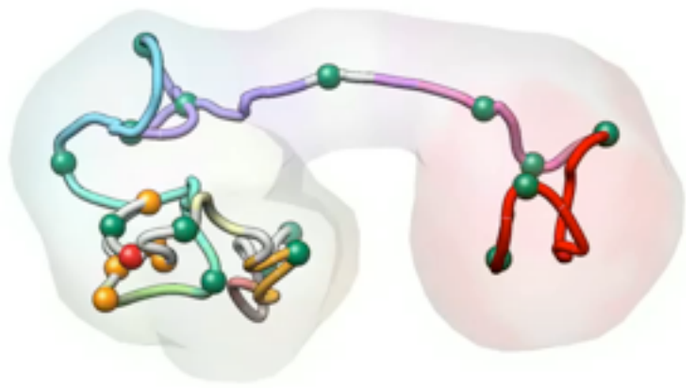




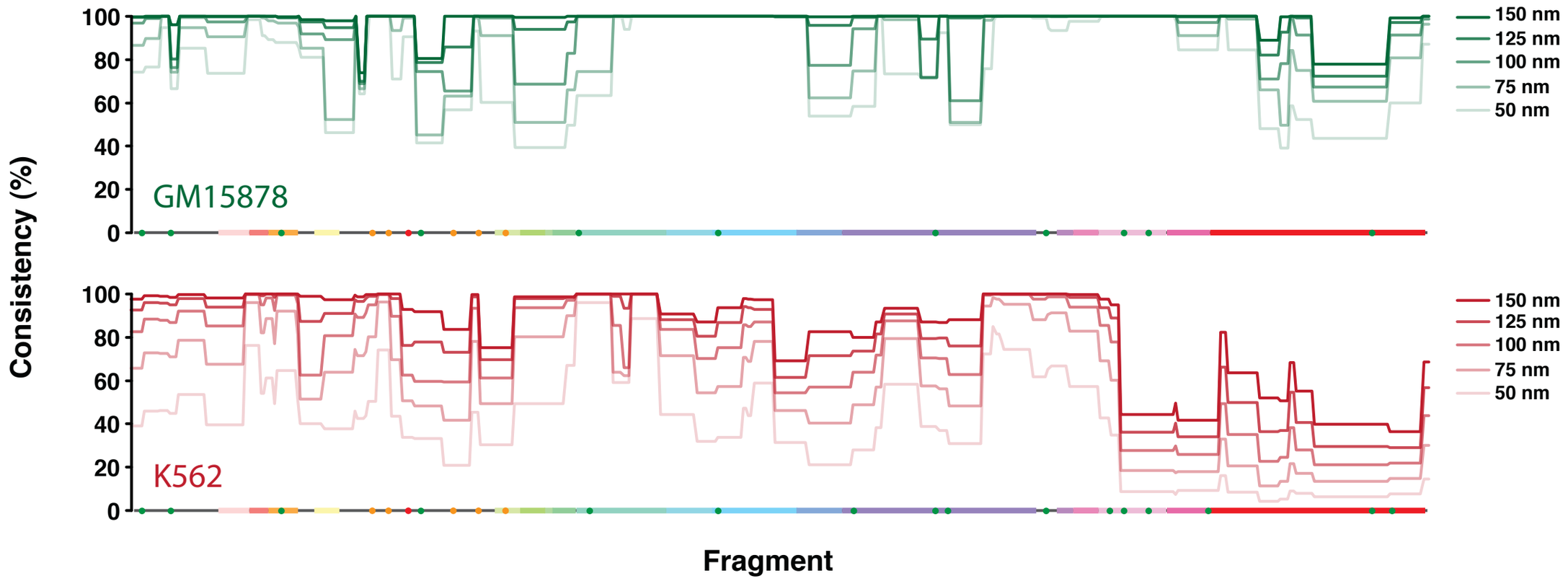
Consistency

GM12878
Cluster #1
2780 model





K562
Cluster #2
314 model

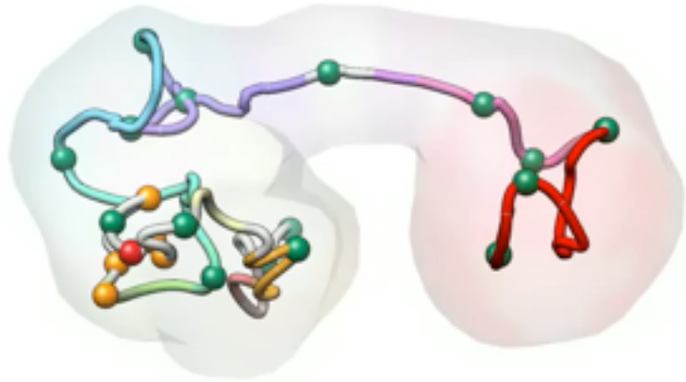


Regulatory elements

GM12878

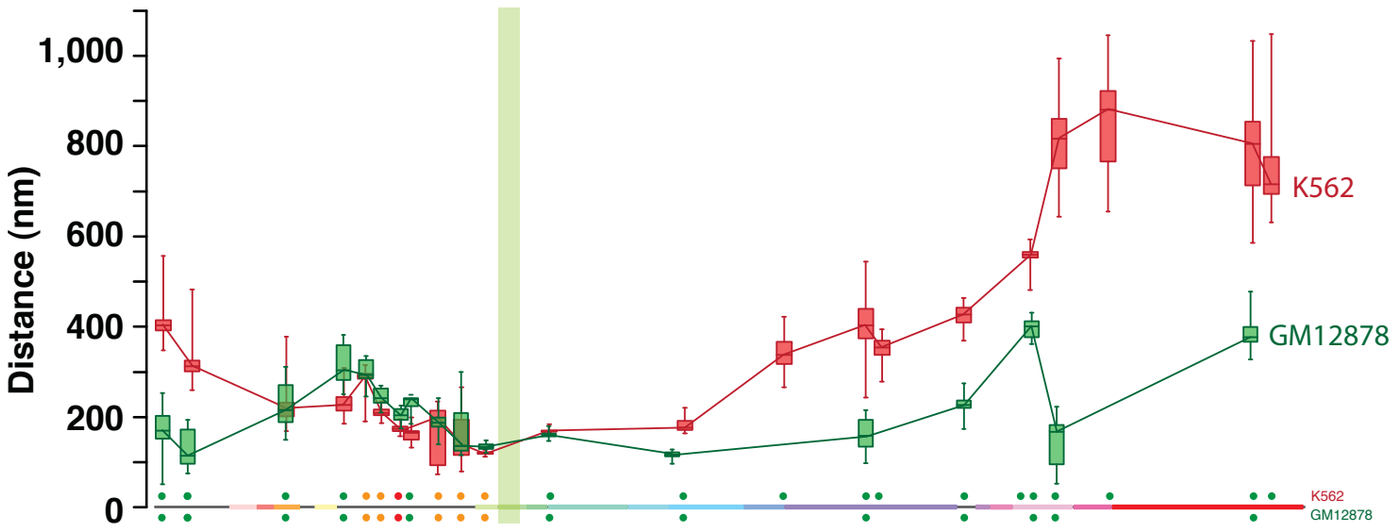
Cluster #1
2780 model

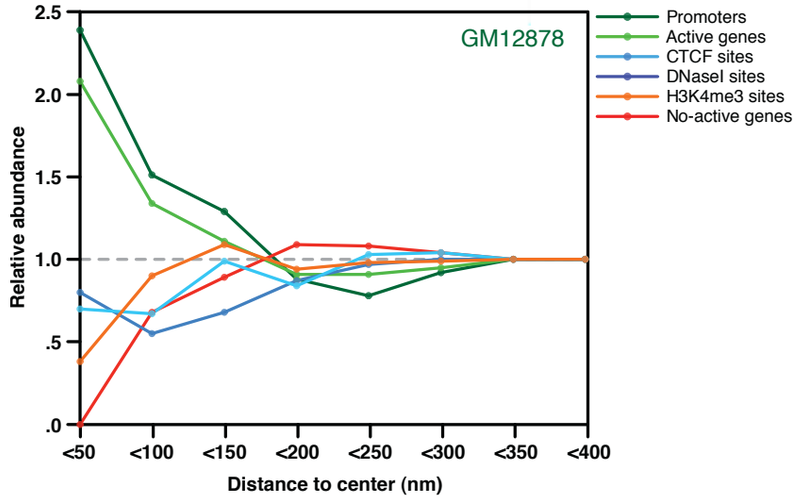


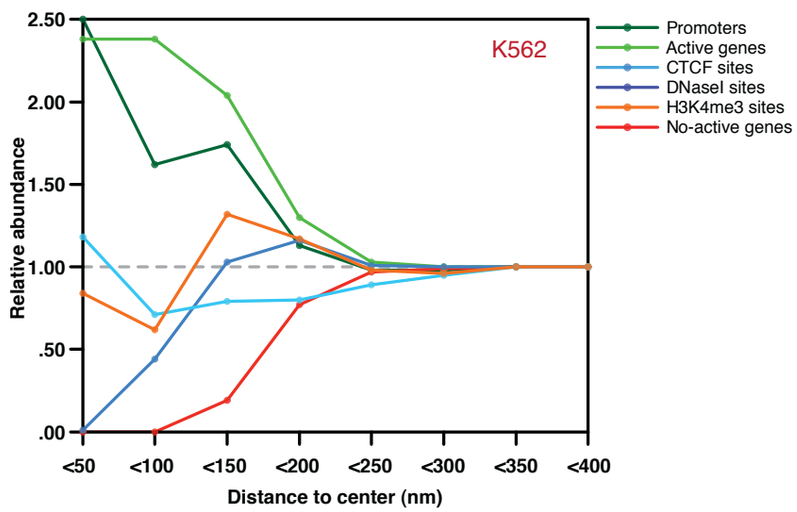


K562

Cluster #2
314 model





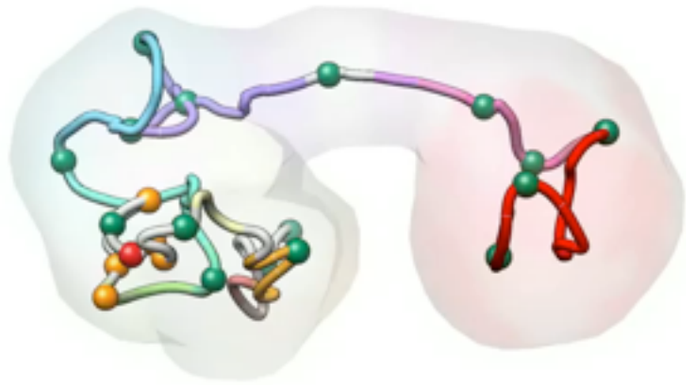


Compactness

GM12878

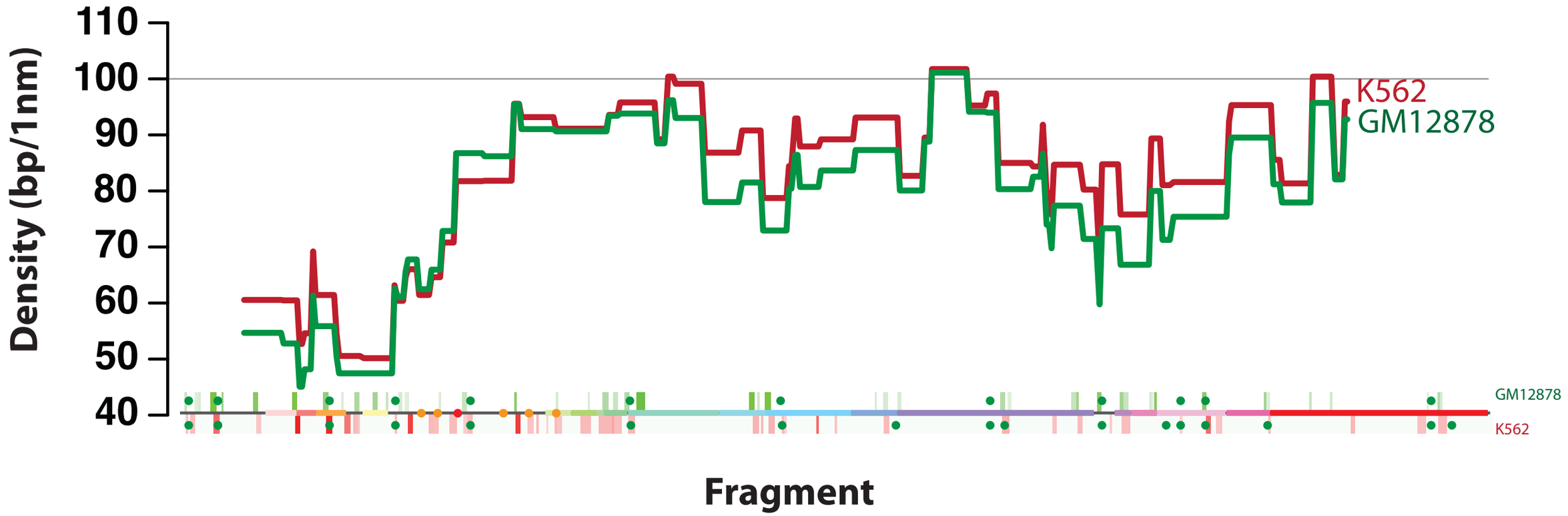
Cluster #1
2780 model





K562

Cluster #2
314 model

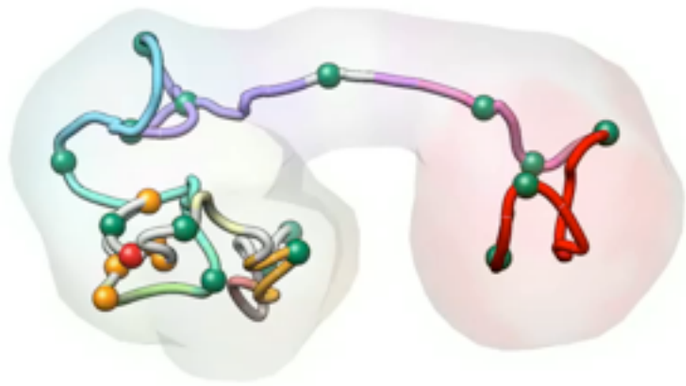


Multi-loops

GM12878

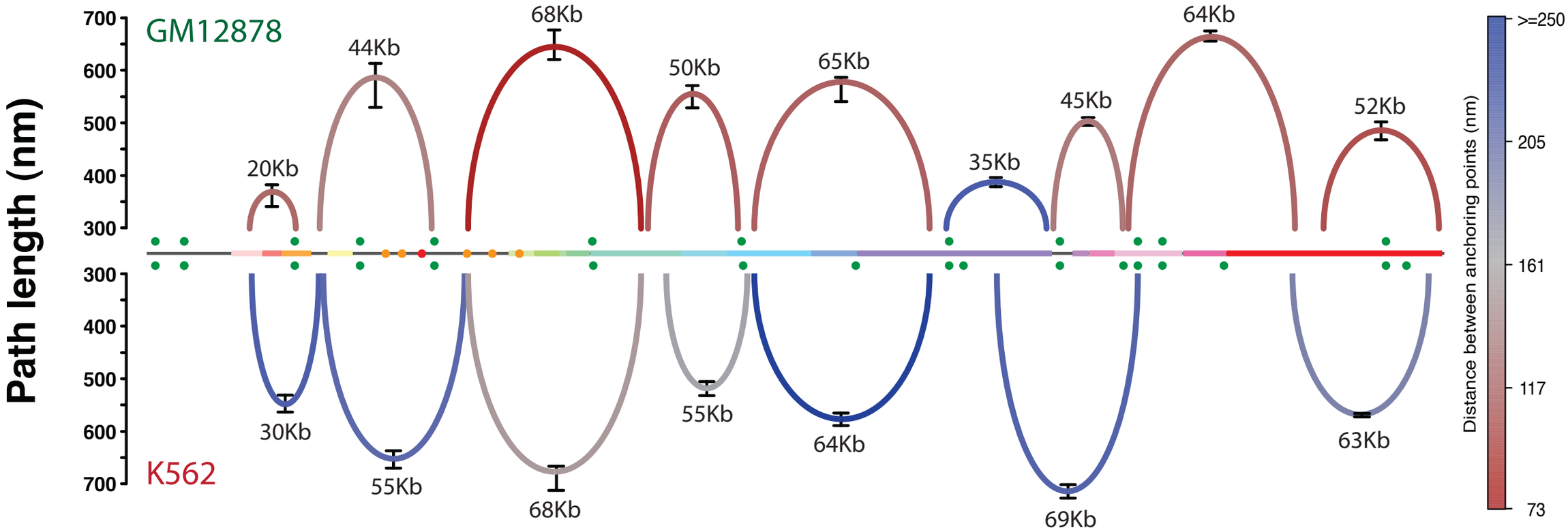
Cluster #1
2780 model





K562

Cluster #2
314 model

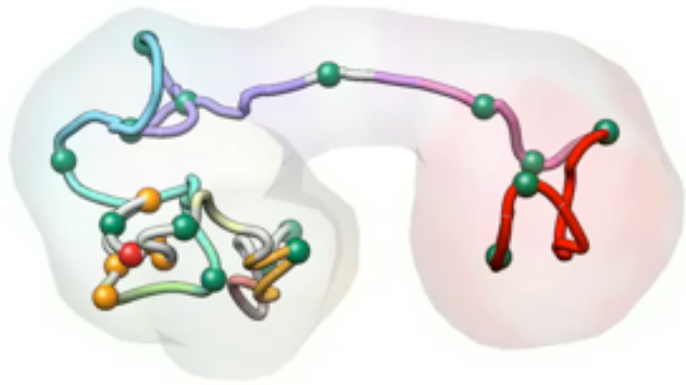


Expression

GM12878

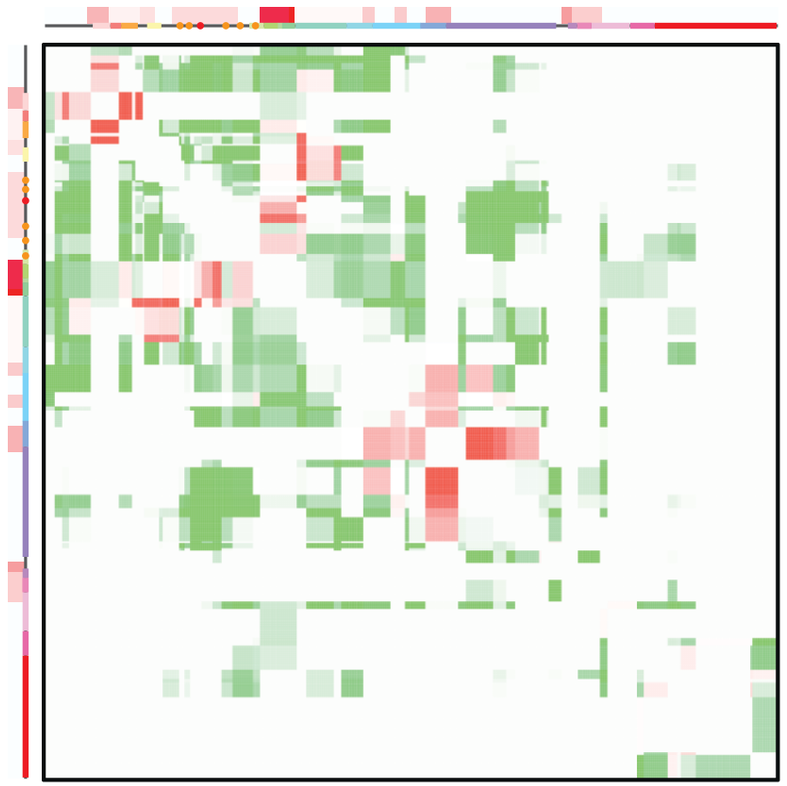
Cluster #1
2780 model





K562

Cluster #2
314 model



Increased in GM12878

=

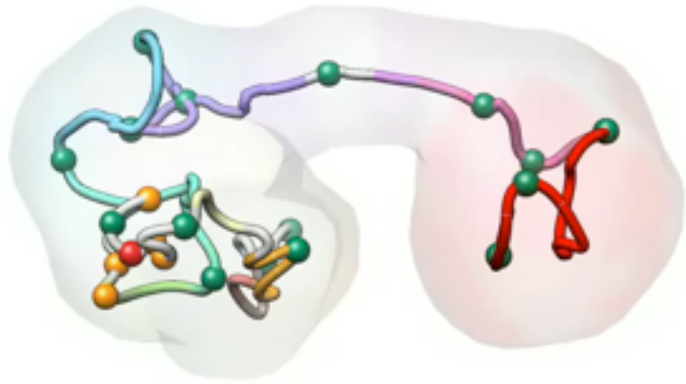
Increased in K562

FISH validation

GM12878

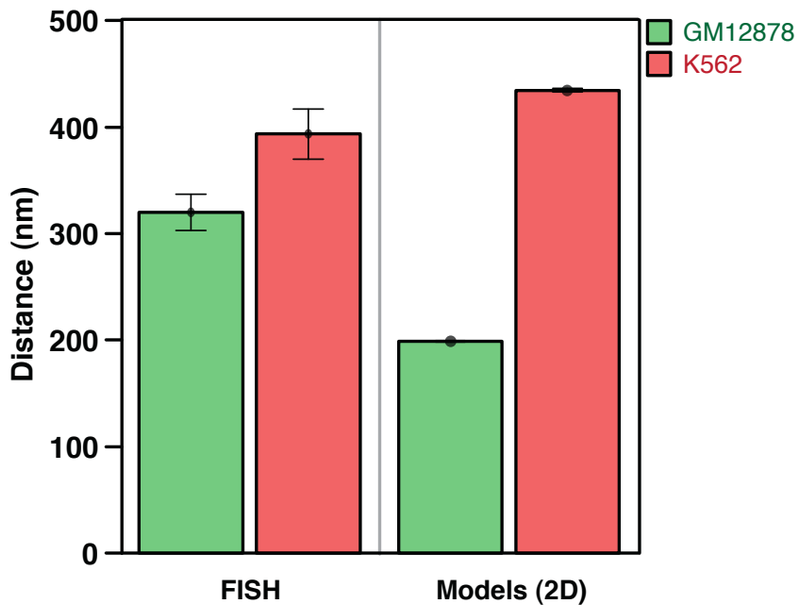
Cluster #1
2780 model

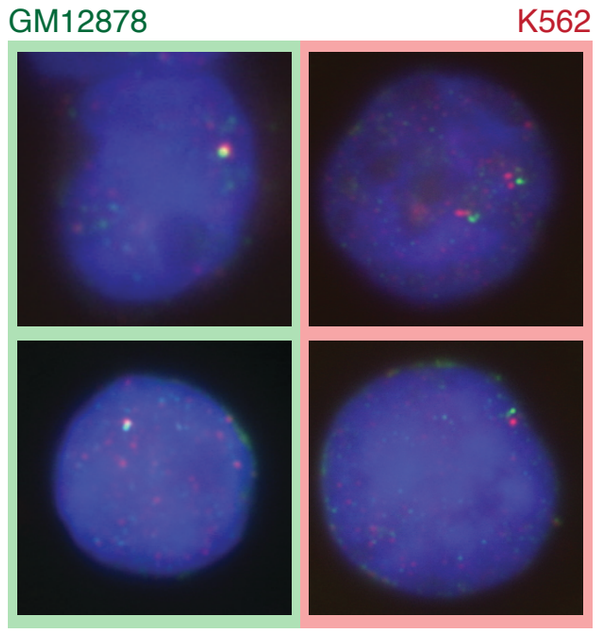




K562

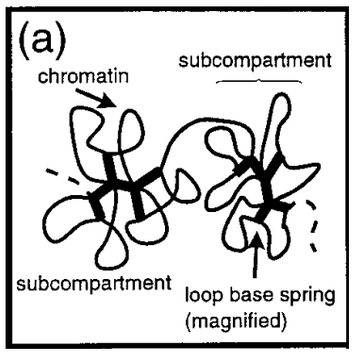
Cluster #2
314 model



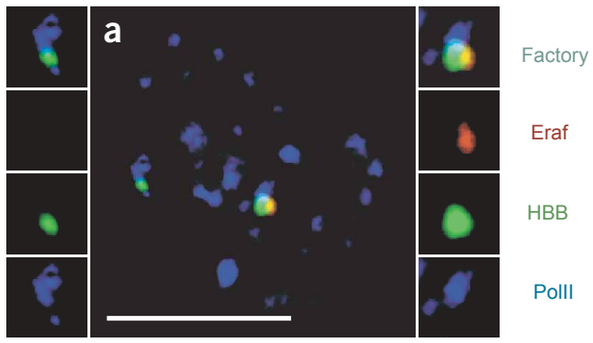


The “Chromatin Globule” model

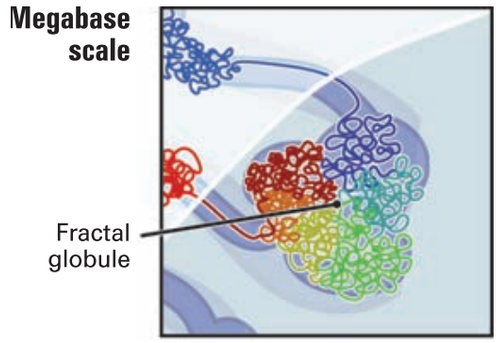




Münkel et al. JMB (1999)



Osborne et al. Nat Genet (2004)

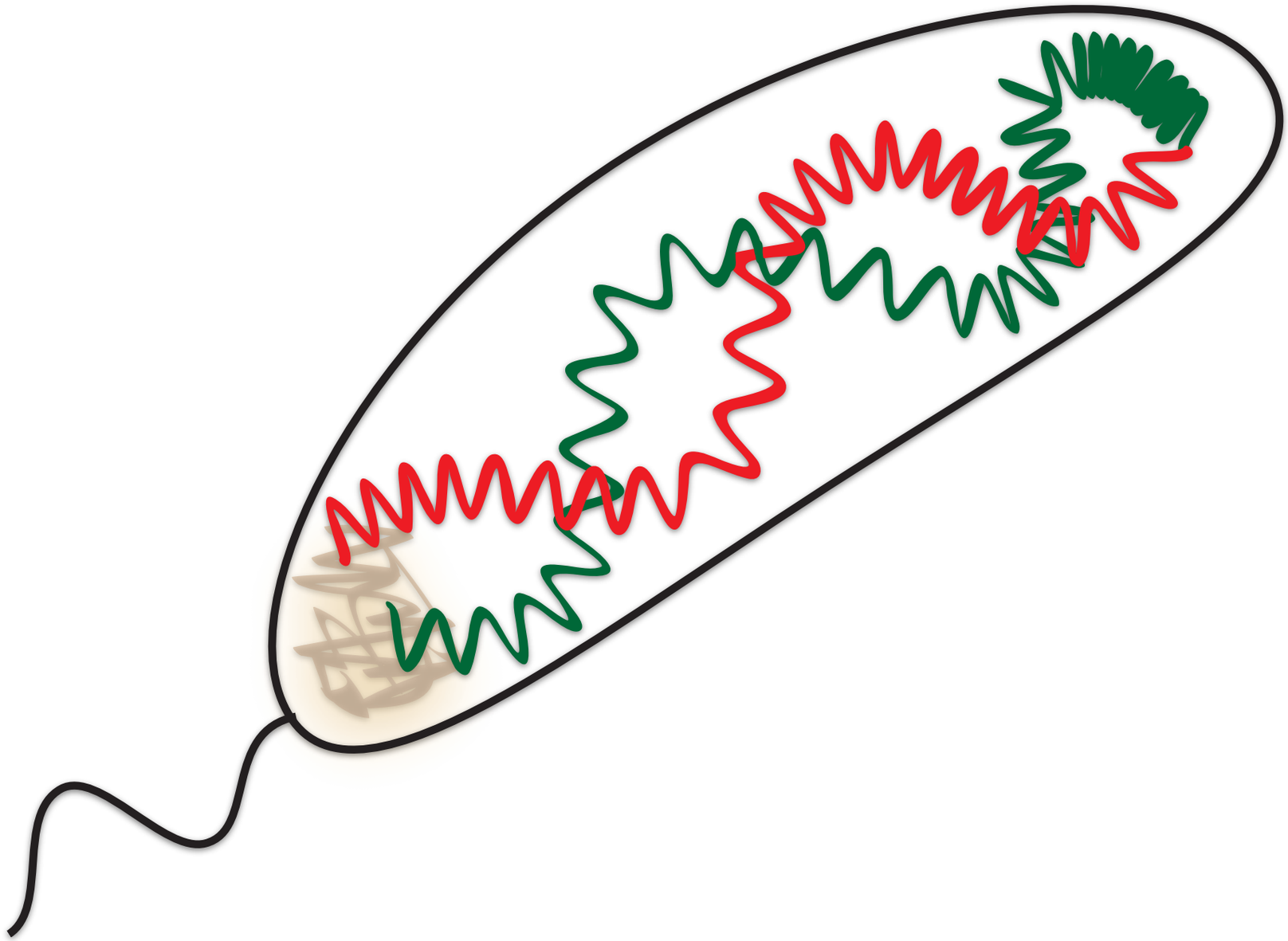


Lieberman-Aiden et al. Science (2009)

D. Baù *et al.* **Nat Struct Mol Biol** (2011) 18:107-14

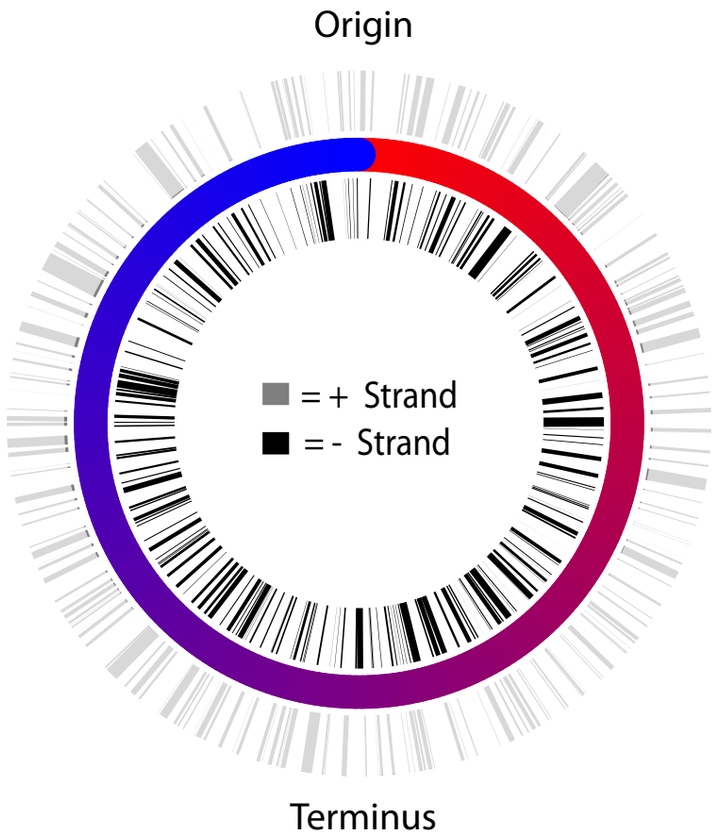
A. Sanyal *et al.* **Current Opinion in Cell Biology** (2011) 23:325–33.

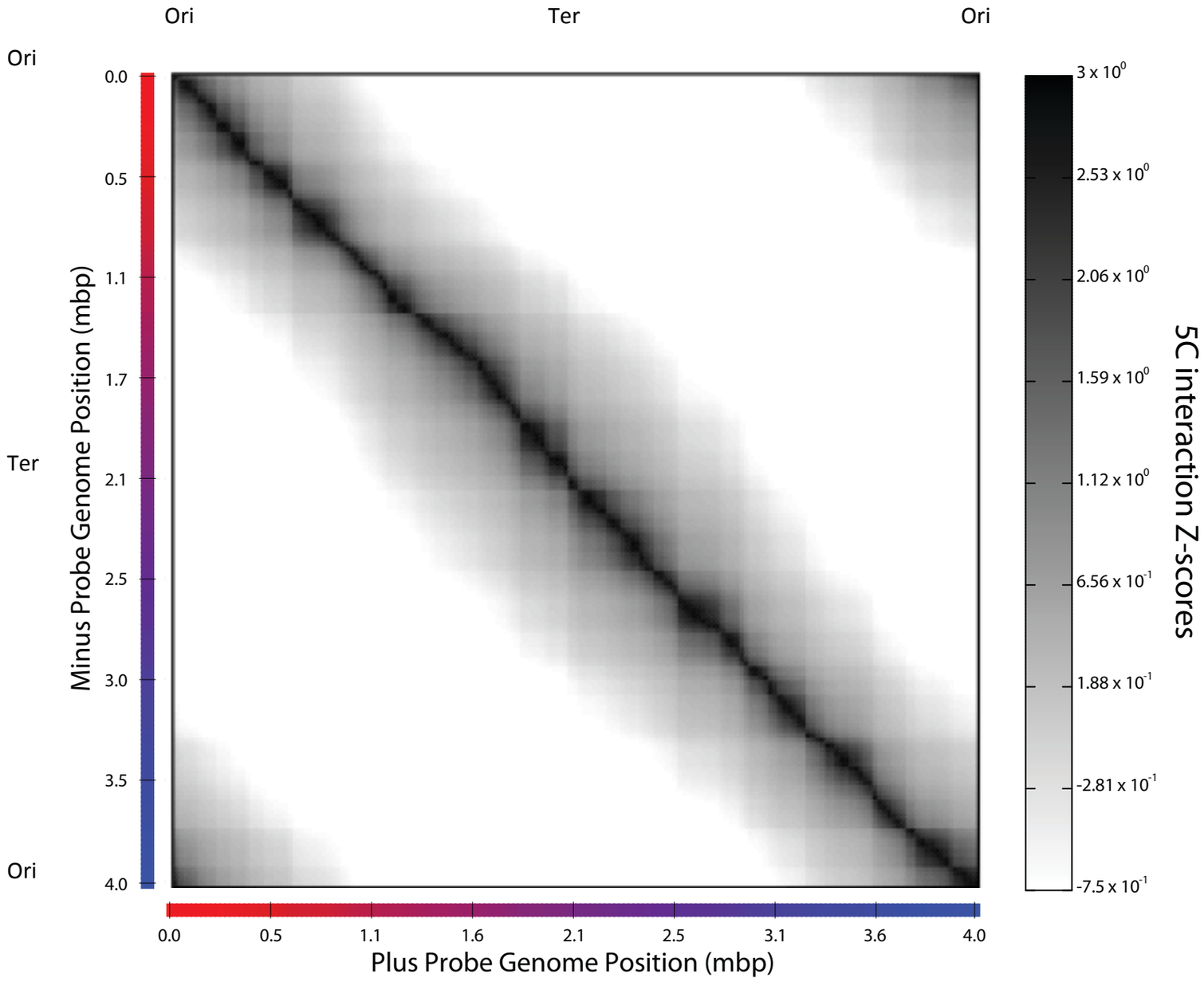
Caulobacter crescentus genome



The 3D architecture of *Caulobacter Crescentus*

4,016,942 bp & 3,767 genes





169 5C primers on + strand

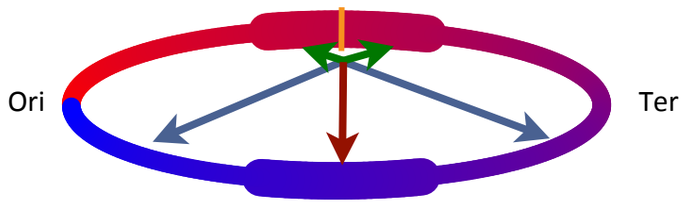
170 5C primers on – strand

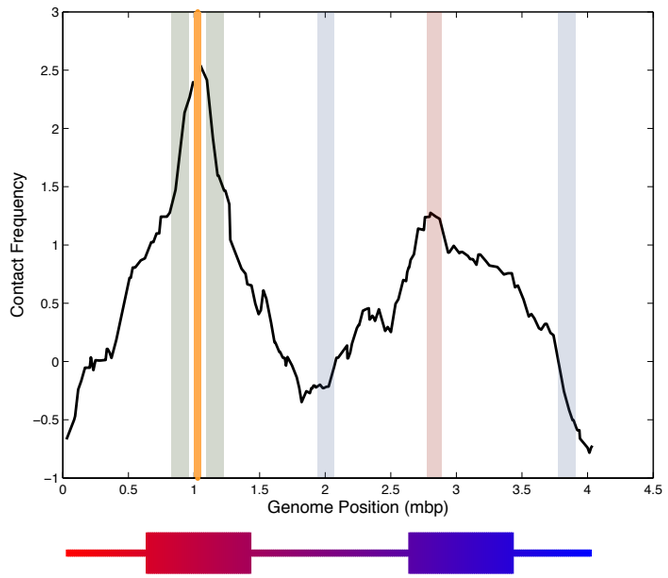
28,730 chromatin interactions

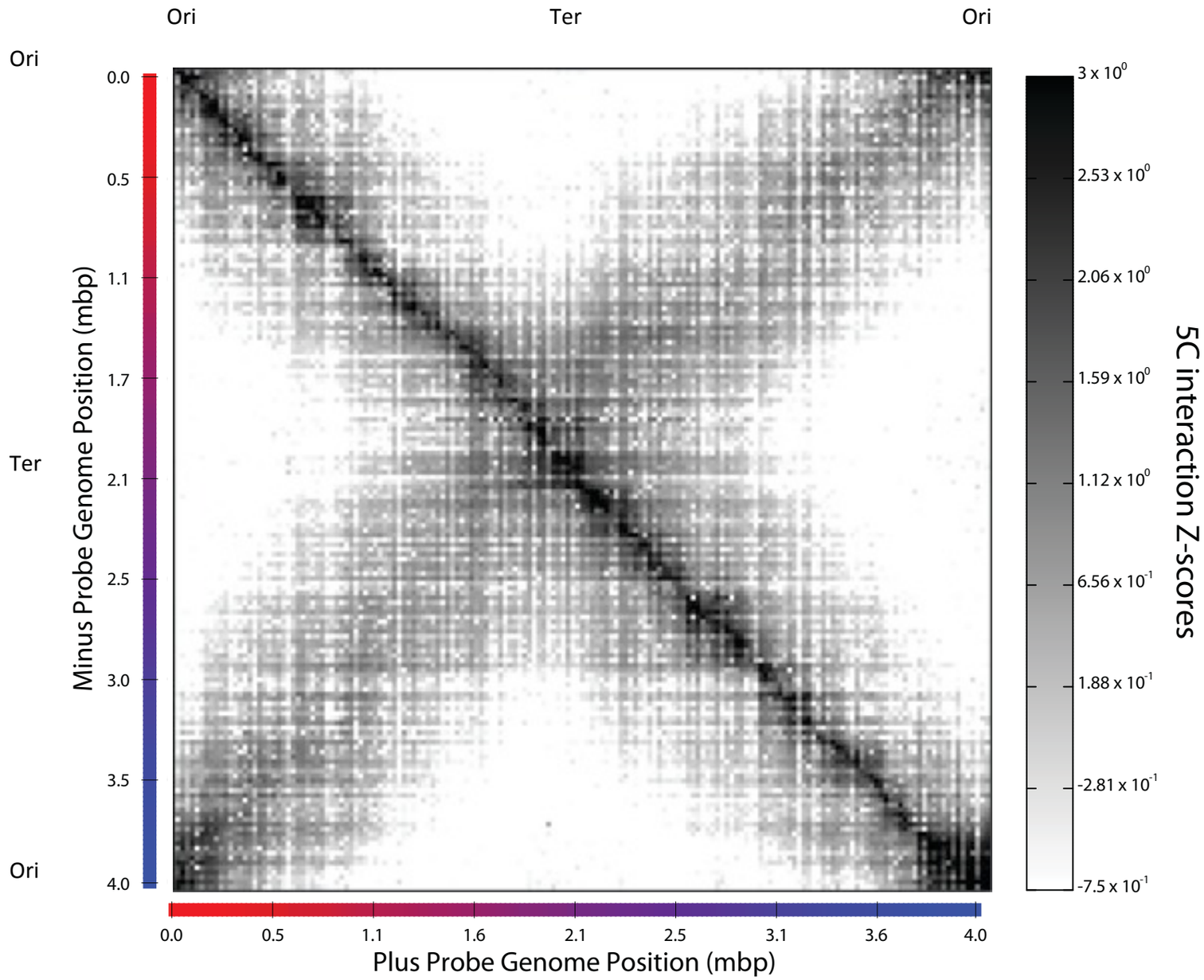
~13Kb

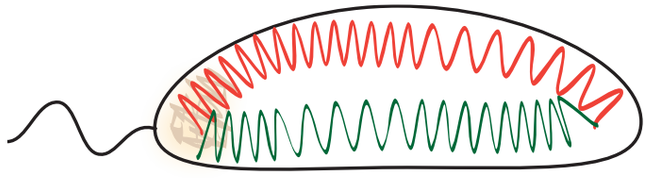
5C interaction matrix

ELLIPSOID for Caulobacter cresentus

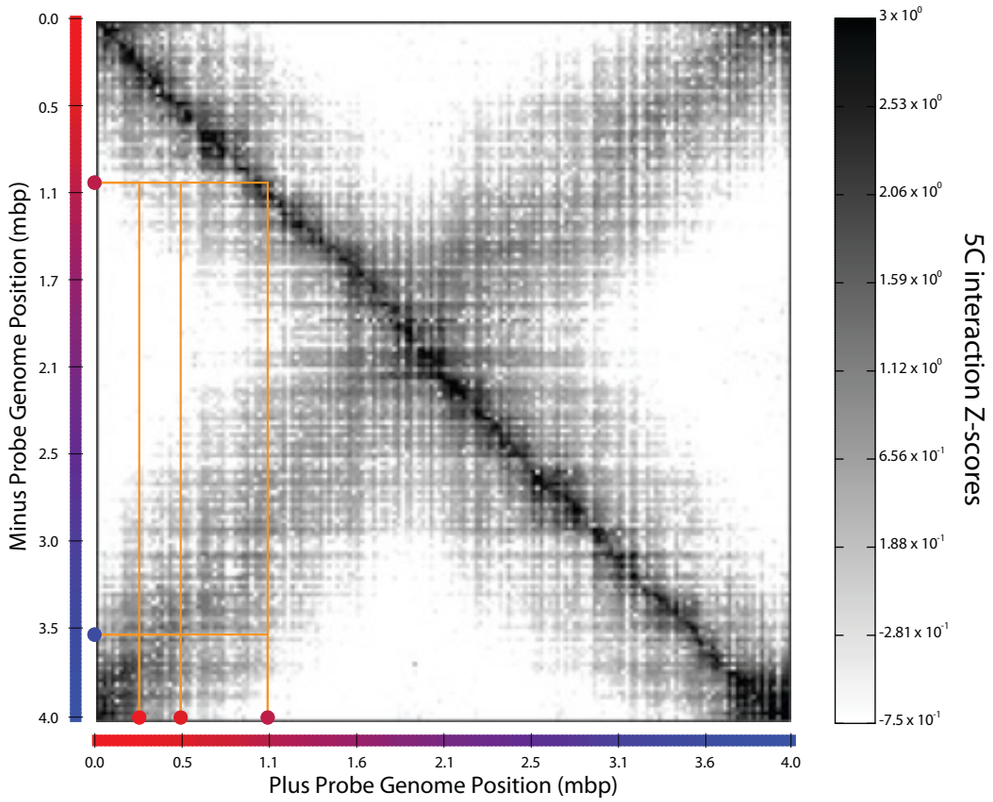




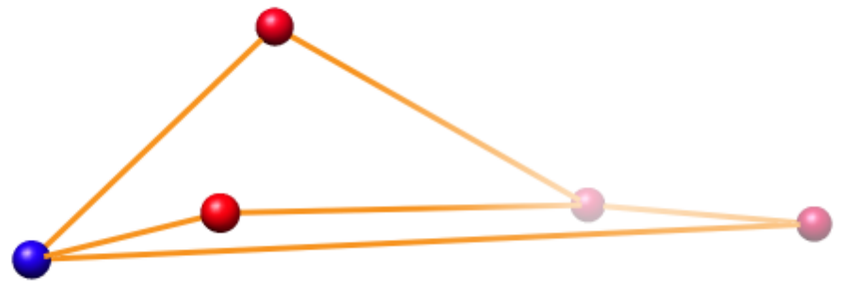


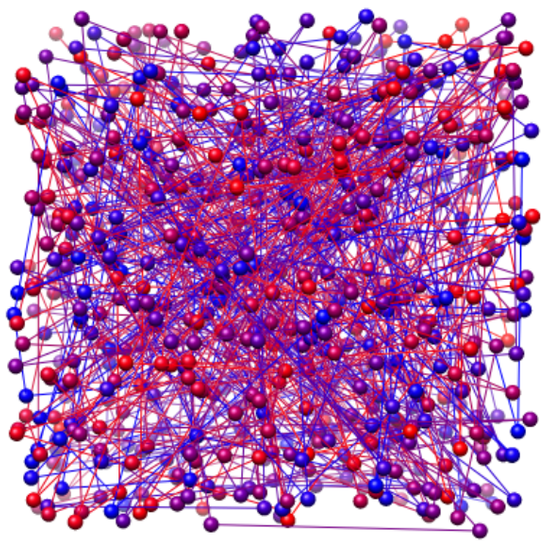


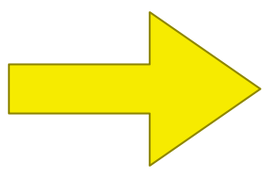
3D model building with the 5C + IMP approach

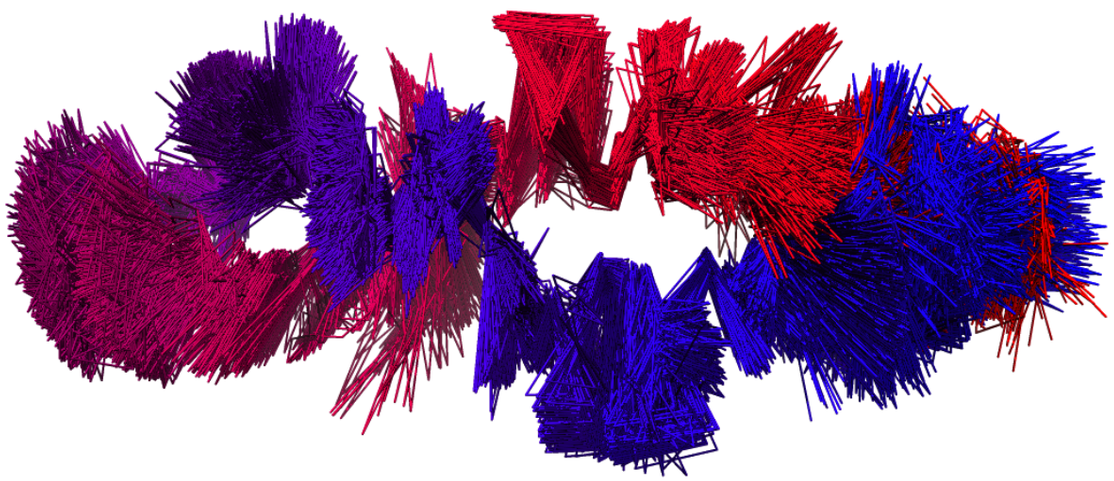


339 mers







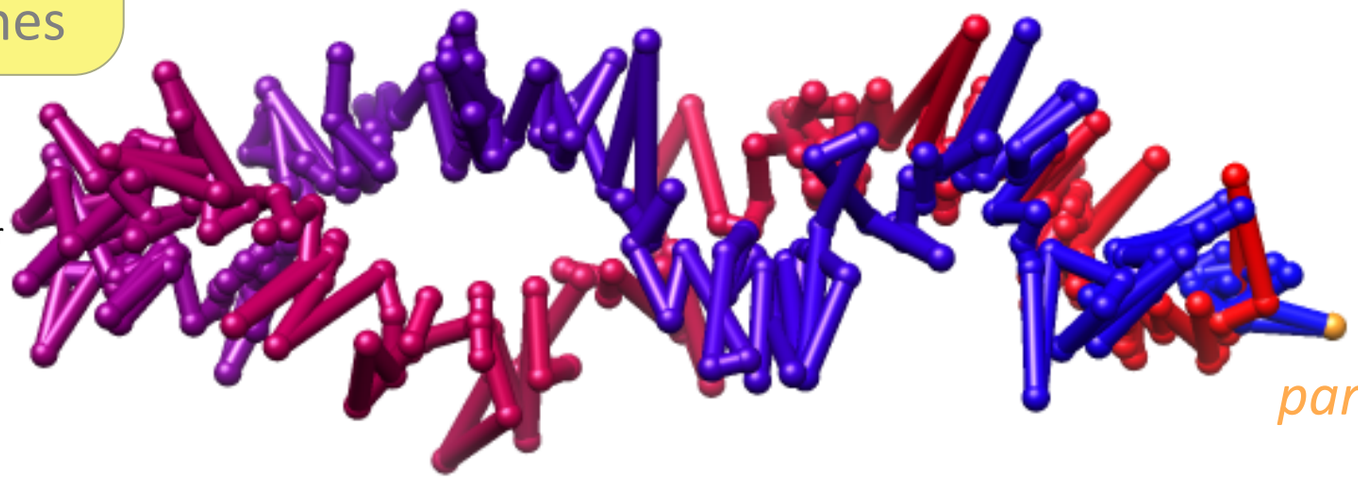


Genome organization in *Caulobacter crescentus*

Resolution of chromosomes

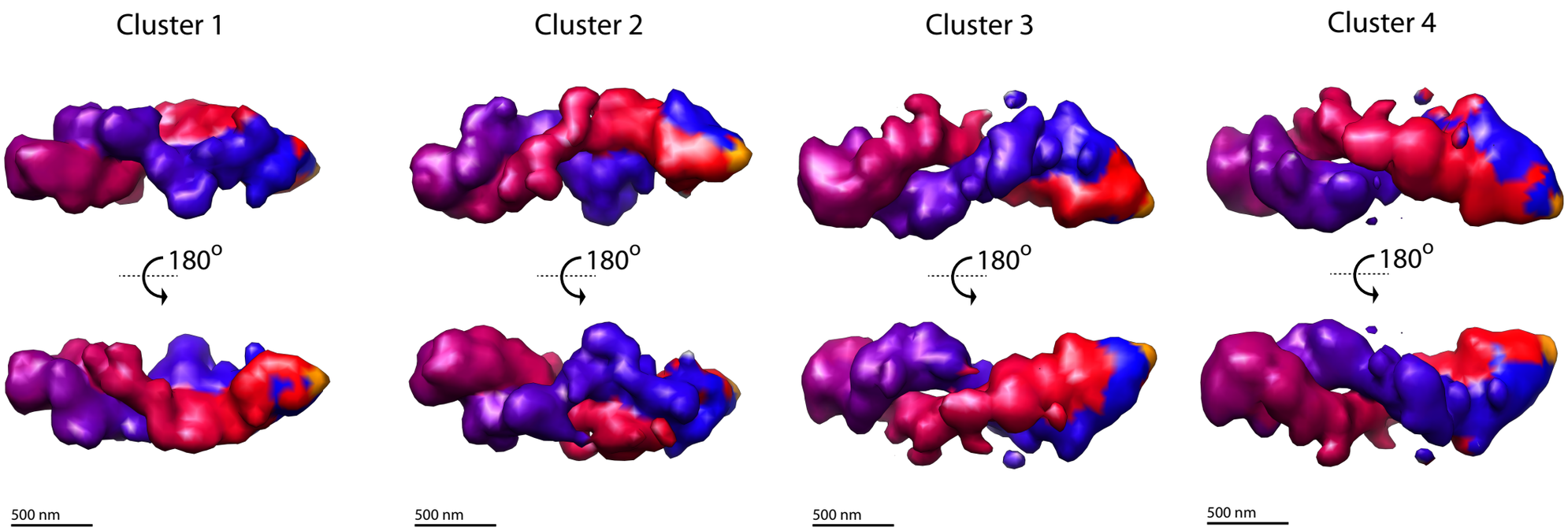
dif site 47±17Kb from Ter

Arms are helical



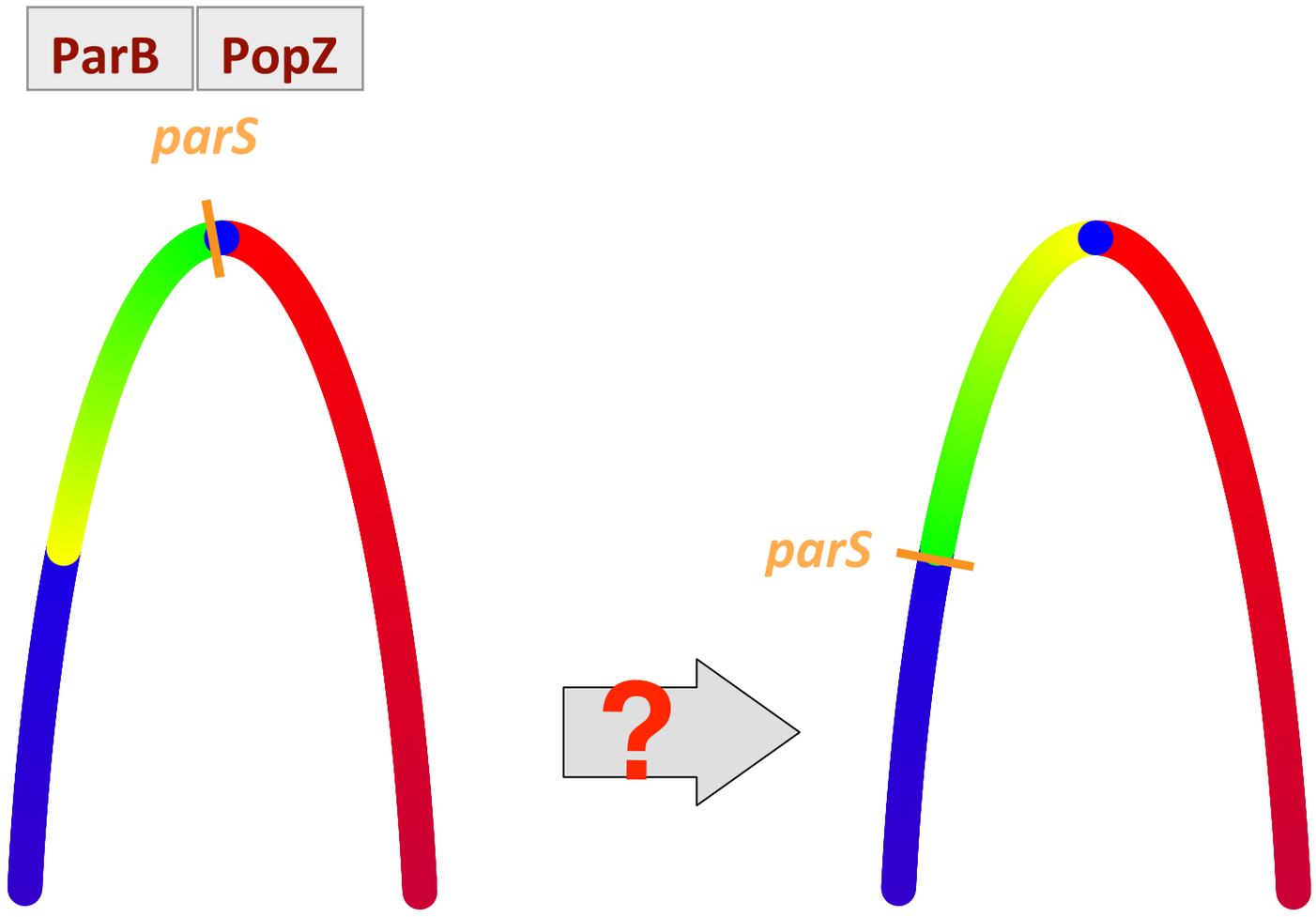
Centromer-like

parS sites 25±17Kb from Ori



MIRRORS!

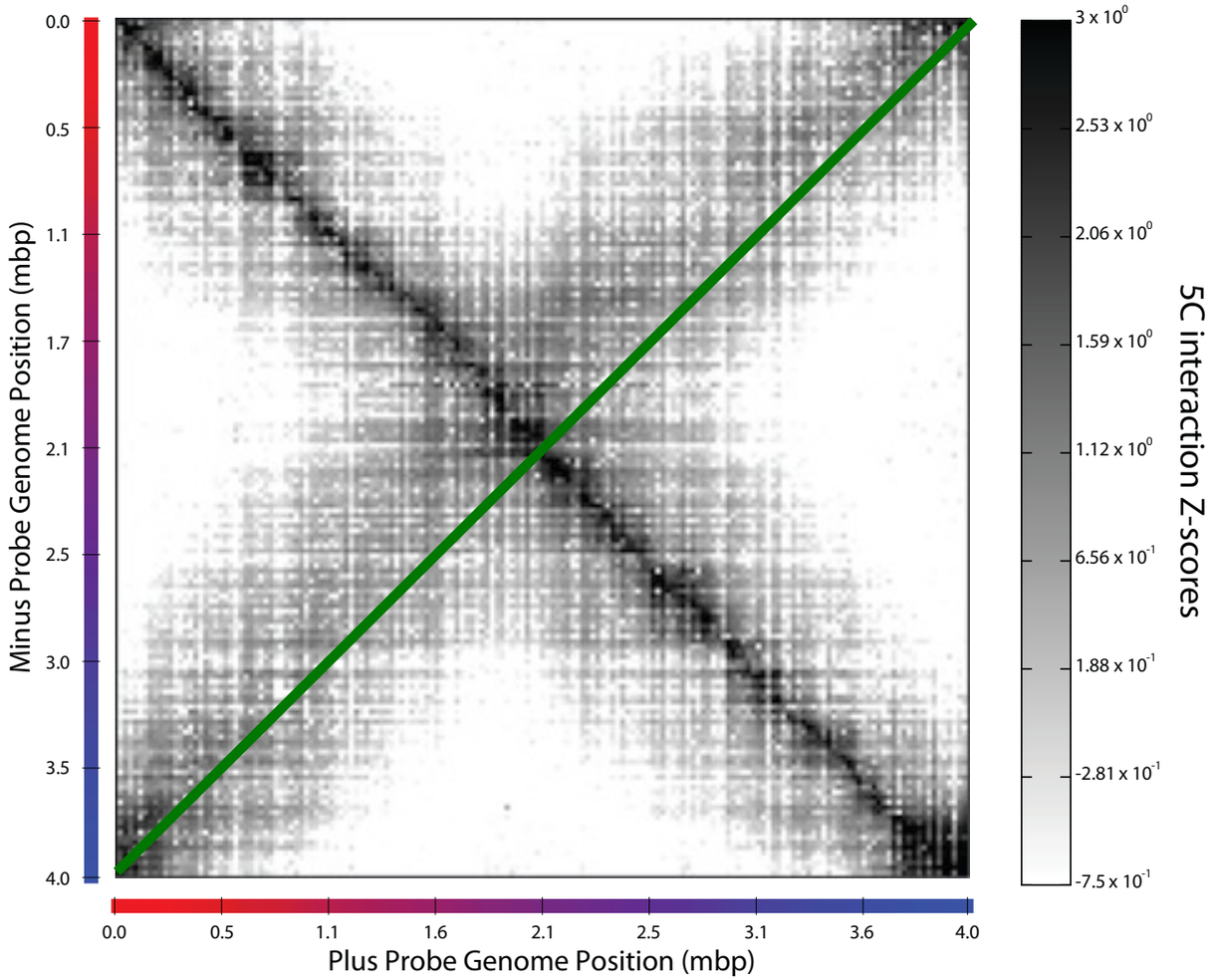
Moving the *parS* sites 400 Kb away from Ori

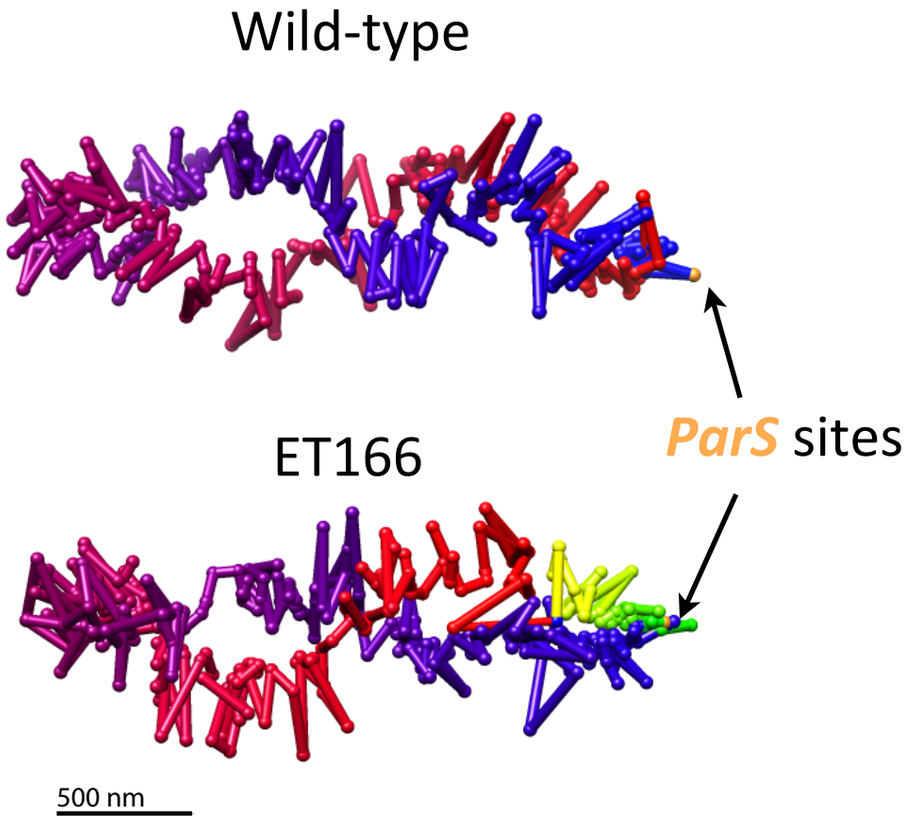


Wild-type

ET166

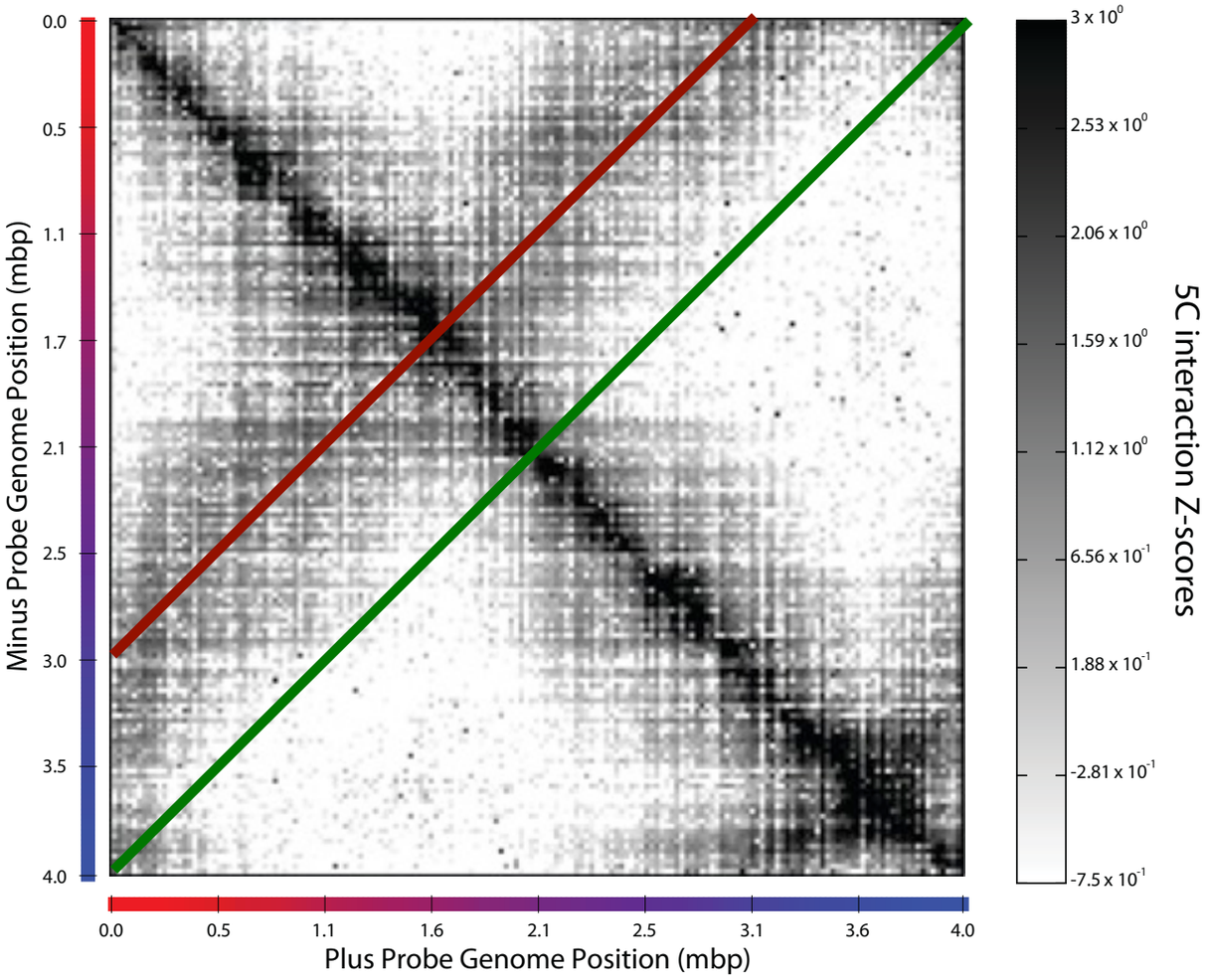
Moving the *parS* sites results in whole genome rotation!

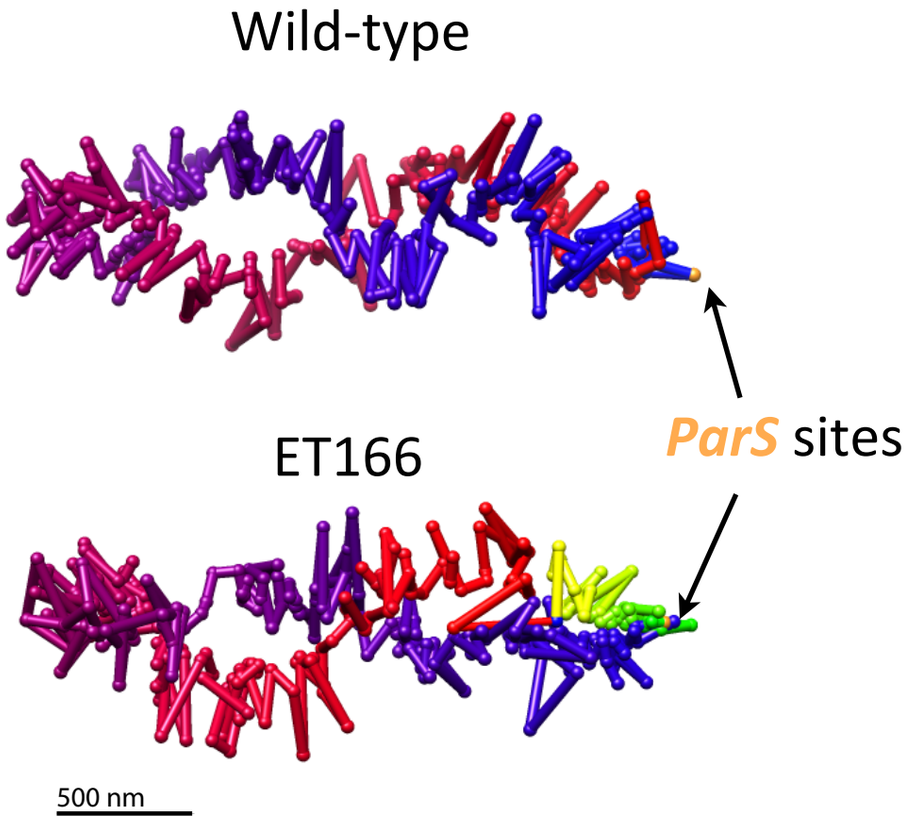




Arms are **STILL** helical

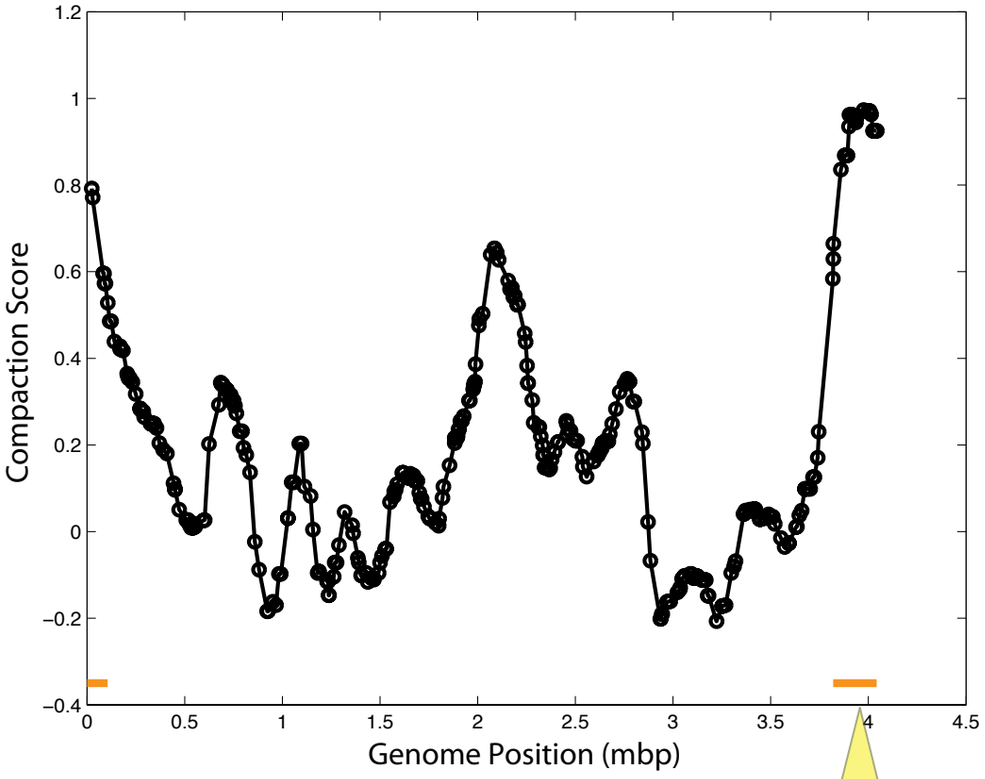
Moving the *parS* sites results in whole genome rotation!



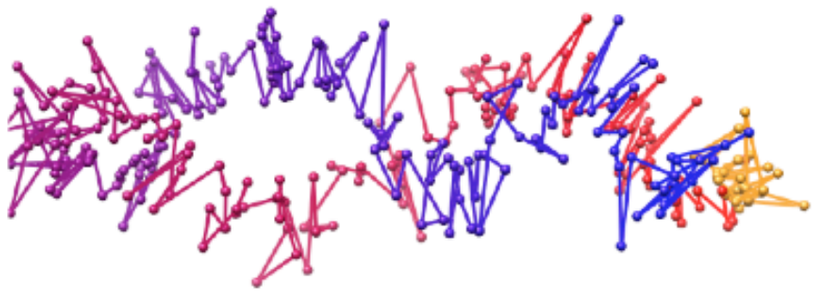


Arms are **STILL** helical

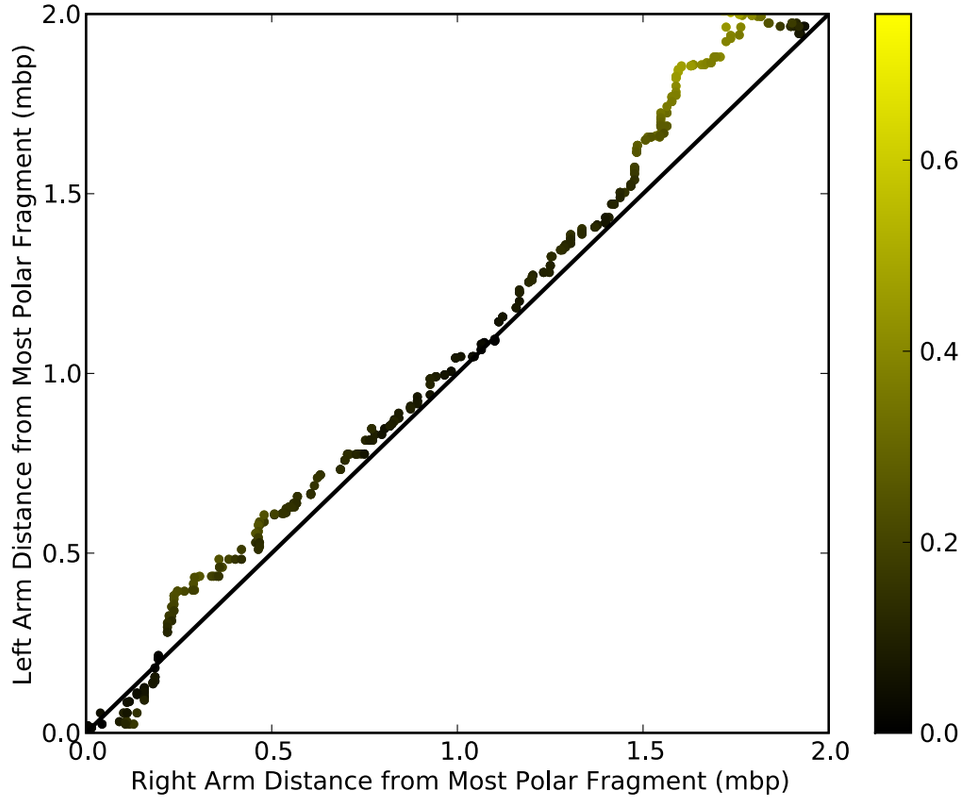
parS sites initiate compact chromatin domain

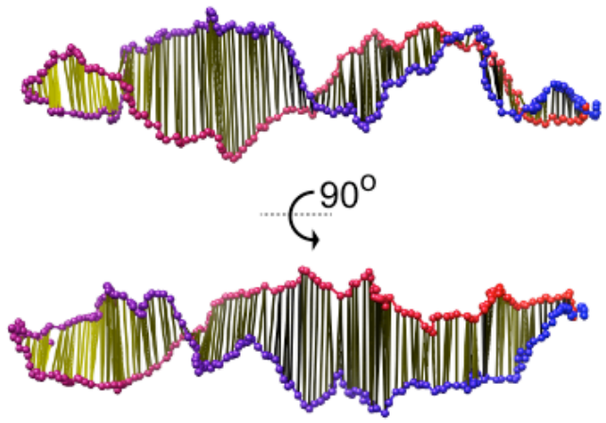


100-200Kb

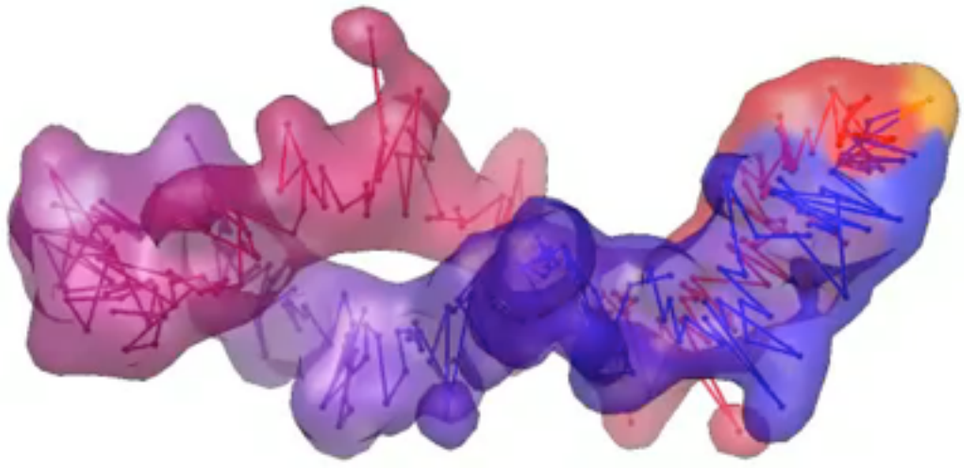


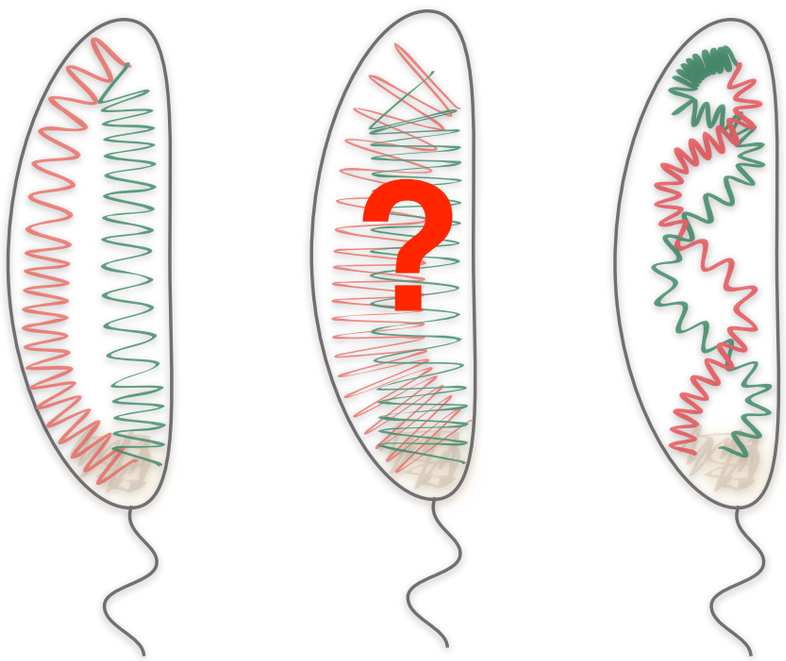
Chromosome arms are equidistant to the cell center

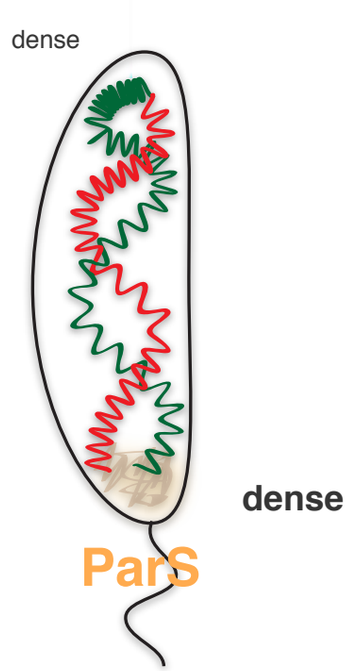




Genome architecture in *Caulobacter*





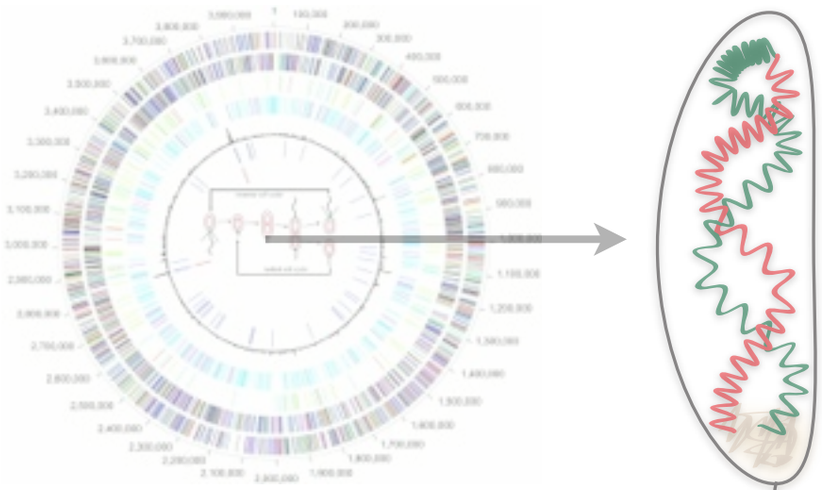


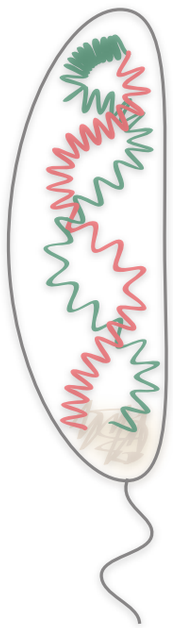
M.A. Umbarger, *et al.* **Molecular Cell** (2011) **44**:252–264

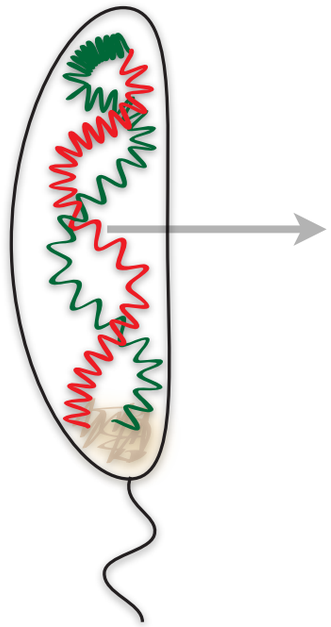
From Sequence to Function

5C + IMP

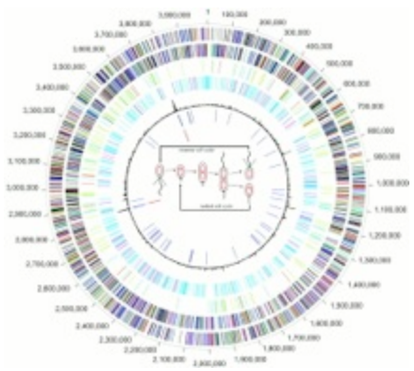
Technology







Hypothesis



Function!

D. Baù and M.A. Marti-Renom **Chromosome Res** (2011) 19:25-35.

PLoS CB Outlook

Marti-Renom MA, Mirny LA (2011) PLoS Comput Biol 7(7): e1002125.

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Review

Bridging the Resolution Gap in Structural Modeling of 3D Genome Organization

Marc A. Marti-Renom[1]*, Leonid A. Mirny[2]

1 Structural Genomics Laboratory, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain, **2** Harvard-MIT Division of Health Sciences and Technology, and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

Abstract: Over the last decade, and especially after the advent of fluorescent *in situ* hybridization imaging and chromosome conformation capture methods, the availability of experimental data on genome three-dimensional organization has dramatically increased. We now have access to unprecedented details of how genomes organize within the interphase nucleus. Development of new computational approaches to leverage this data has already resulted in the first three-dimensional structures of genomic domains and genomes. Such approaches expand our knowledge of the chromatin folding principles, which has been classically studied using polymer physics and molecular simulations. Our outlook describes computational approaches for integrating experimental data with polymer physics, thereby bridging the resolution gap for structural determination of genomes and genomic domains.

This is an "Editors' Outlook" article for *PLoS Computational Biology*

Recent experimental and computational advances are resulting in an increasingly accurate and detailed characterization of how genomes are organized in the three-dimensional (3D) space of the nucleus (Figure 1) [1]. At the lowest level of chromatin organization, naked DNA is packed into nucleosomes, which forms the so-called chromatin fiber composed of DNA and proteins. However, this initial packing, which reduces the length of the DNA by about seven times, is not sufficient to explain the higher-order folding of chromosomes during interphase and metaphase. It is now accepted that chromosomes and genes are non-randomly and dynamically positioned in the cell nucleus during the interphase, which challenges the classical representation of genomes as linear static sequences. Moreover, compartmentalization, chromatin organization, and spatial location of genes are associated with gene expression and the functional status of the cell. Despite the importance of 3D genomic architecture, we have a limited understanding of the molecular mechanisms that determine the higher-order organization of genomes and its relation to function. Computational biology plays an important role in the plethora of new technologies aimed at addressing this knowledge gap [2]. Indeed, Thomas Cremer, a pioneer in studying nuclear organization using light microscopy, recently highlighted the importance of computational science in complementing and leveraging experimental observations of genome organization [2]. Therefore, computational approaches to integrate experimental observations with chromatin physics are needed to determine the architecture (3D) and dynamics (4D) of genomes.

We present two complementary approaches to address this challenge: (i) the first approach aims at developing simple polymer models of chromatin and determining relevant interactions (both physical and biological) that explain experimental observations; (ii) the second approach aims at integrating diverse experimental observations into a system of spatial restraints to be satisfied, thereby constraining possible structural models of the chromatin. The goal of both approaches is dual: to obtain most accurate 3D and 4D representation of chromatin architecture and to understand physical constraints and biological phenomena that determine its organization. These approaches are reminiscent of the protein-folding field where the first strategy was used for characterizing protein "foldability" and the second was implemented for modeling the structure of proteins using nuclear magnetic resonance and other experimental constraints. In fact, our outlook consistently returns to the many connections between the two fields.

What Does Technology Show Us?

Today, it is possible to quantitatively study structural features of genomes at diverse scales that range from a few specific loci, through chromosomes, to entire genomes (Table 1) [3]. Broadly, there are two main approaches for studying genomic organization: light microscopy and cell/molecular biology (Figure 2). Light microcopy [4], both with fixed and living cells, can provide images of a few loci within individual cells [5,6], as well as their dynamics as a function of time [7] and cell state [8]. On a larger scale, light microscopy combined with whole-chromosome staining reveals chromosomal territories during interphase and their reorganization upon cell division. Immunofluorescence with fluorescent antibodies in combination with RNA, and DNA fluorescence *in situ* hybridization (FISH) has been used to determine the co-localization of loci and nuclear substructures.

Using cellular and molecular biology, novel chromosome conformation capture (3C)-based methods such 3C [9], 3C-on-chip or circular 3C (the so-called 4C) [10,11], 3C carbon copy (5C) [12], and Hi-C [13] quantitatively measure frequencies of spatial contacts between genomic loci averaged over a large

Citation: Marti-Renom MA, Mirny LA (2011) Bridging the Resolution Gap in Structural Modeling of 3D Genome Organization. PLoS Comput Biol 7(7): e1002125. doi:10.1371/journal.pcbi.1002125

Editor: Philip E. Bourne, University of California San Diego, United States of America

Published July 14, 2011

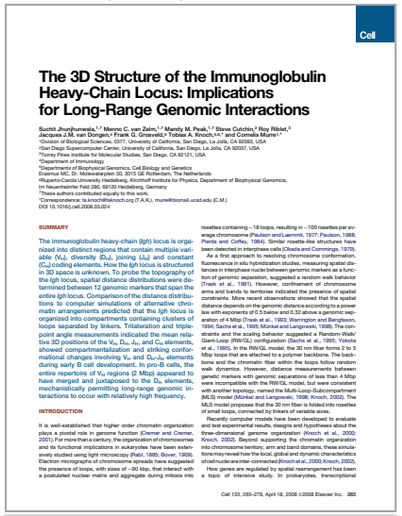
Copyright: © 2011 Marti-Renom, Mirny. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: MAM-R acknowledges support from the Spanish Ministry of Science and Innovation (BFU2010-19310). LM is acknowledges support of the NCI-funded MIT Center for Physics Sciences in Oncology. The funders had no role in decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

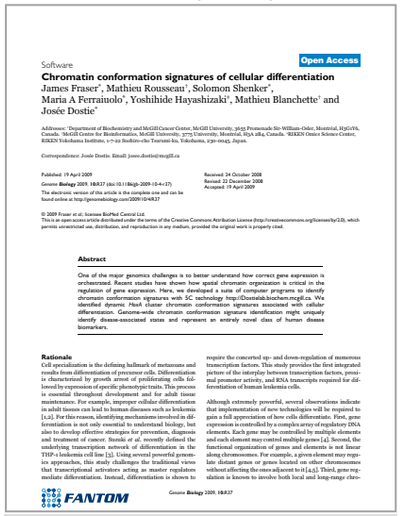
* E-mail: mmarti@cipf.es

MURRE
Cell (2008) **133**:265-79



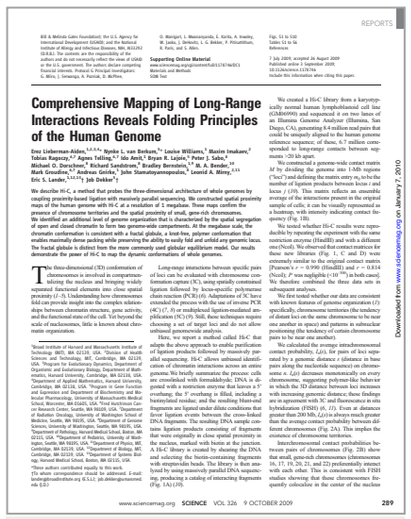
The 3D Structure of the Immunoglobulin Heavy-Chain Locus: Implications for Long-Range Genomic Interactions

DOSTIE/BLANCHETTE
Genome Biol (2009) **10**: R37



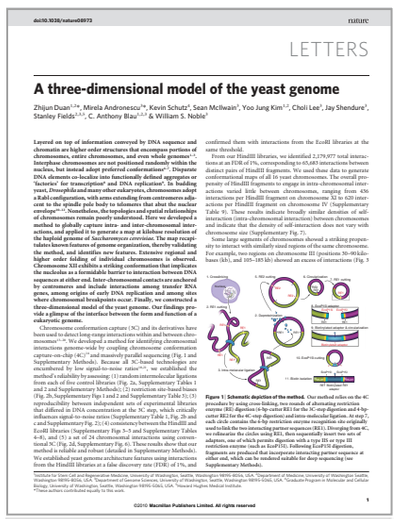
Chromatin conformation signatures of cellular differentiation

DEKKER/LANDER/MIRNY
Science (2009) **326**:289-93



Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

NOBLE
Nature (2010) **465**: 363-7



A three-dimensional model of the yeast genome

DEKKER/MARTI-RENOM
NSMB (2011) **18**:107-14



The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules

OPEN POSITIONS IN THE LAB
Starting early 2012

Acknowledgments



Bryan R Lajoie
Bioinformatician
UMASS



Amartya Sanyal
Postdoctoral fellow
UMASS



Meg Byron
Research Associate
UMASS



Mark Umbarger
PhD fellow
Harvard



Esteban Toro
PhD fellow
Stanford



Davide Baù
Postdoctoral fellow
CIPF

Job Dekker

Program in Gene Function and Expression
Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
Worcester, MA, USA

Jeanne Lawrence

Department of Cell Biology
University of Massachusetts Medical School
Worcester, MA, USA

George M. Church

Department of Genetics,
Harvard Medical School,
Boston, MA. USA

Lucy Shapiro

Department of Developmental Biology,
Stanford University School of Medicine,
Stanford, CA. USA

Marc A. Marti-Renom

Structural Genomics Laboratory
Bioinformatics and Genomics Department
Centro de Investigación Príncipe Felipe
Valencia, Spain

http://sgu.bioinfo.cipf.es
http://integrativemodeling.org