Structure determination of genomes and genomic domains by satisfaction of spatial restraints
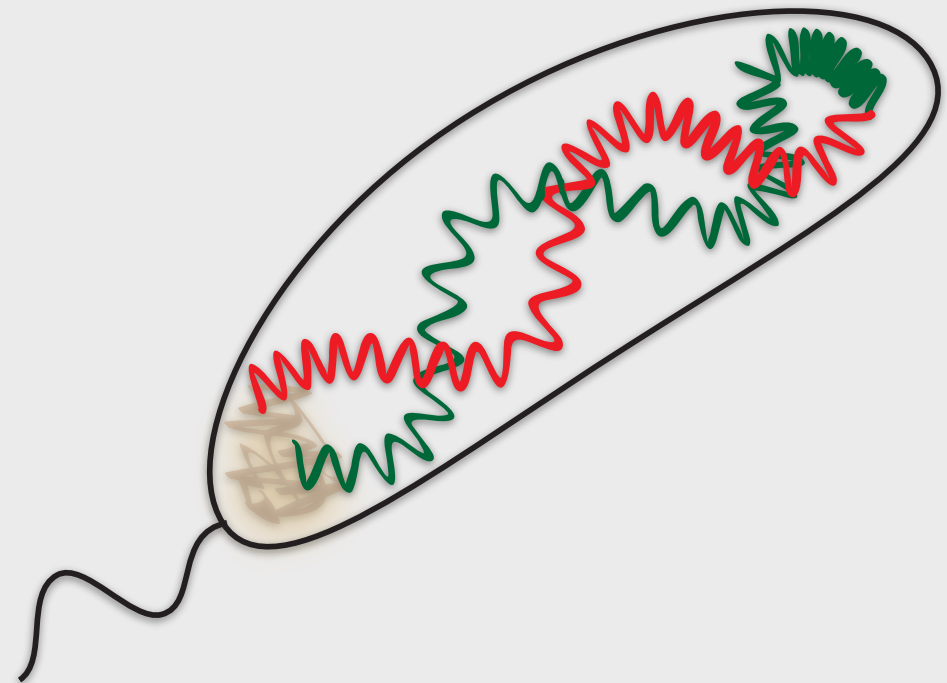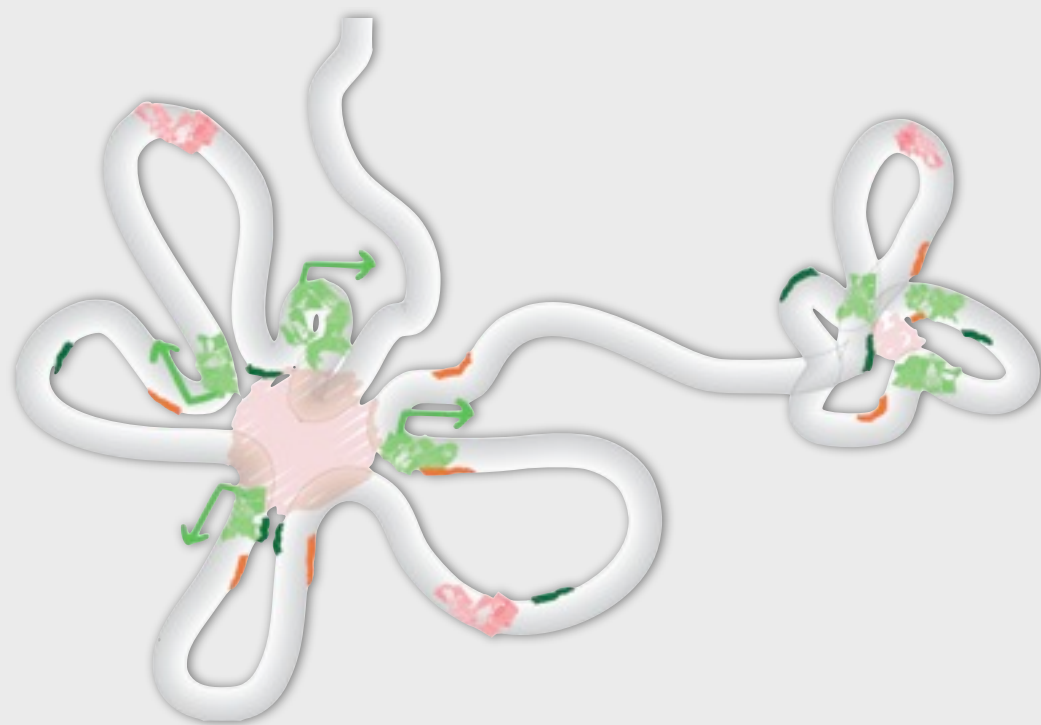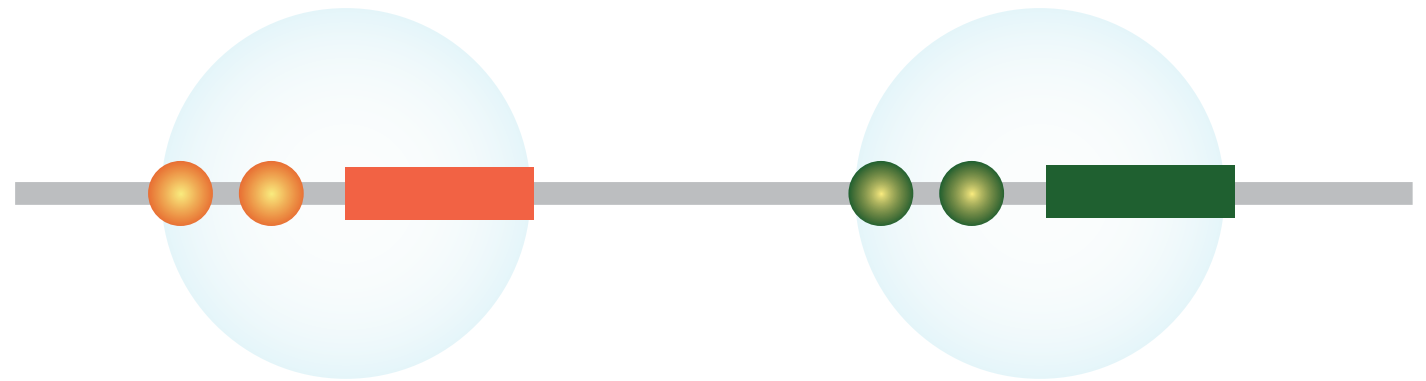
Marc A. Marti-Renom

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

http://sgu.bioinfo.cipf.es
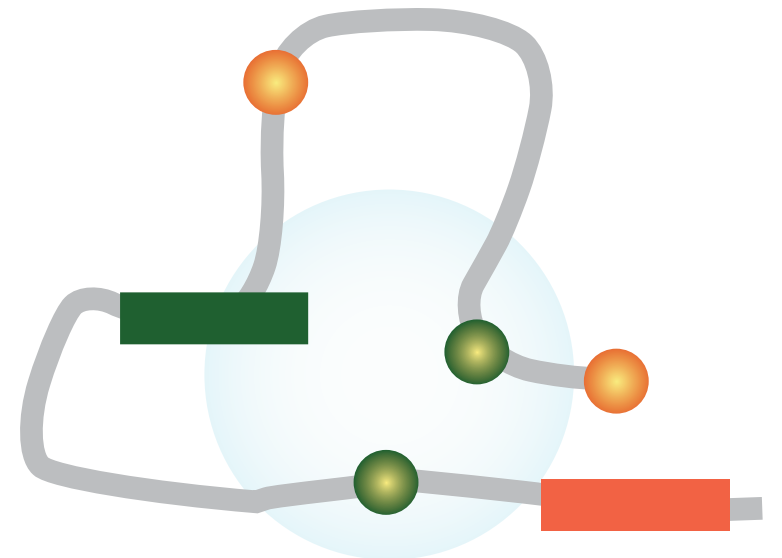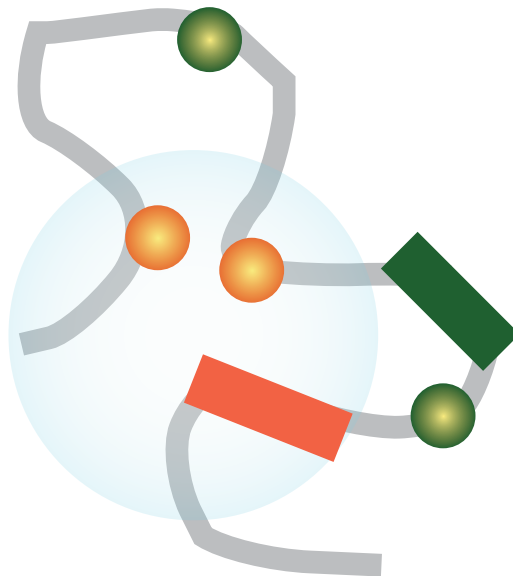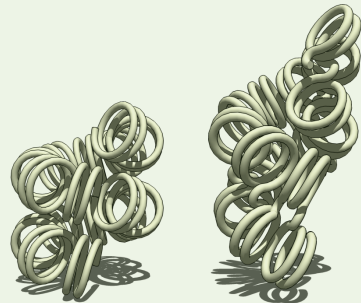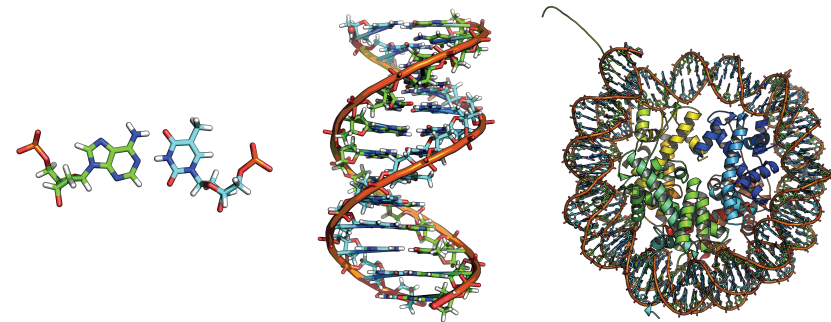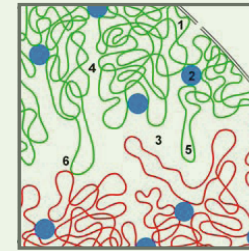
Simple genomes

Complex genomes

Knowledge

IDM

INM
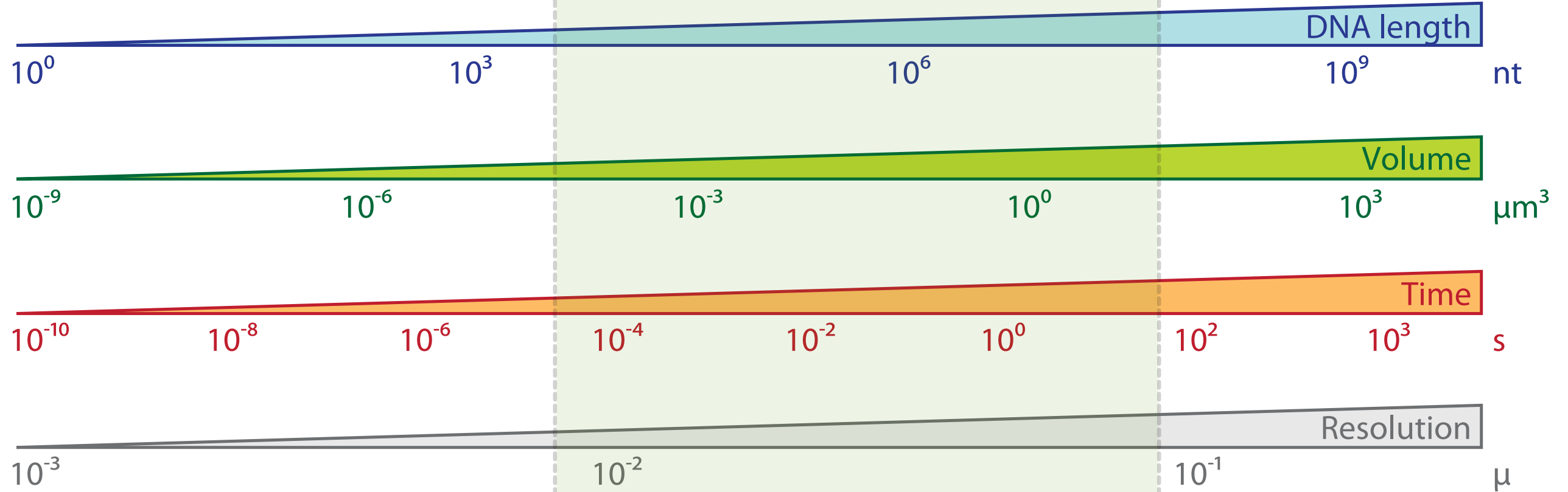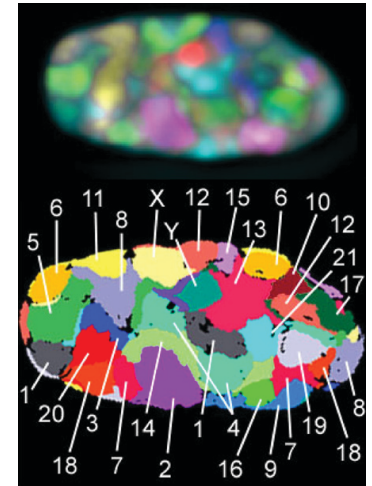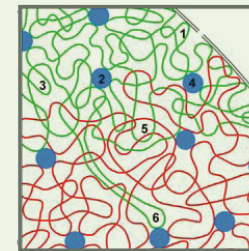
DNA length

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

Volume

10^{-9} 10^{-6} 10^{-3} 10^0 10^3 μm³

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

3

Grow GM12878 and K562 cells

↓

Perform 3C analysis

↓

Perform 5C analysis with 30+25 primers

↓

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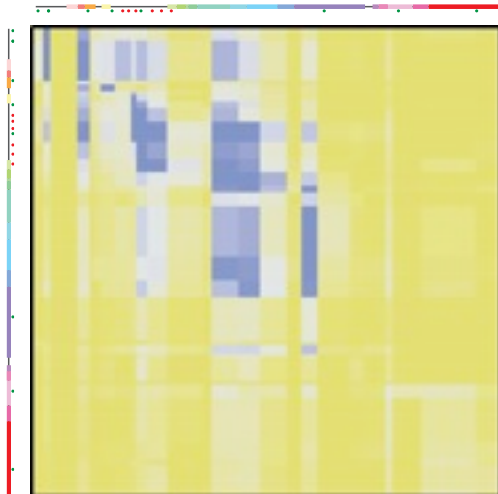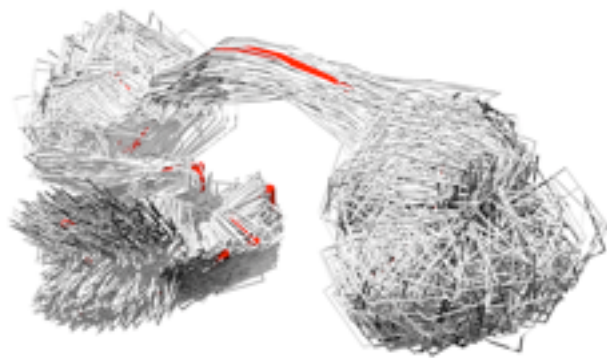(131,947 paired end reads per library)

Computation

4
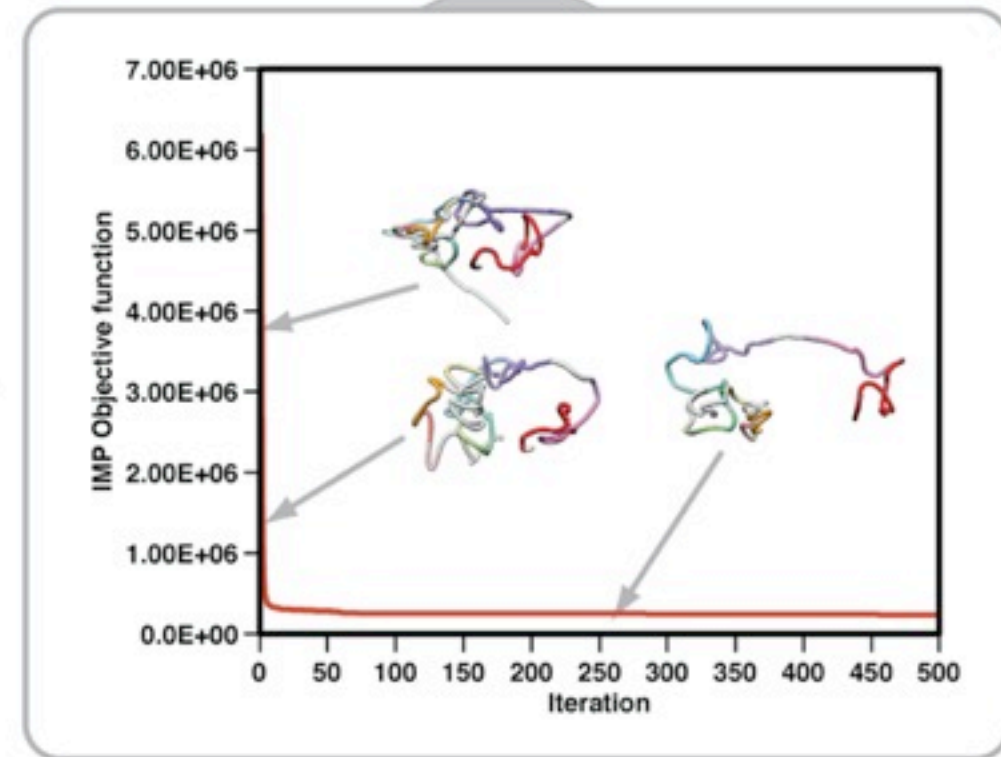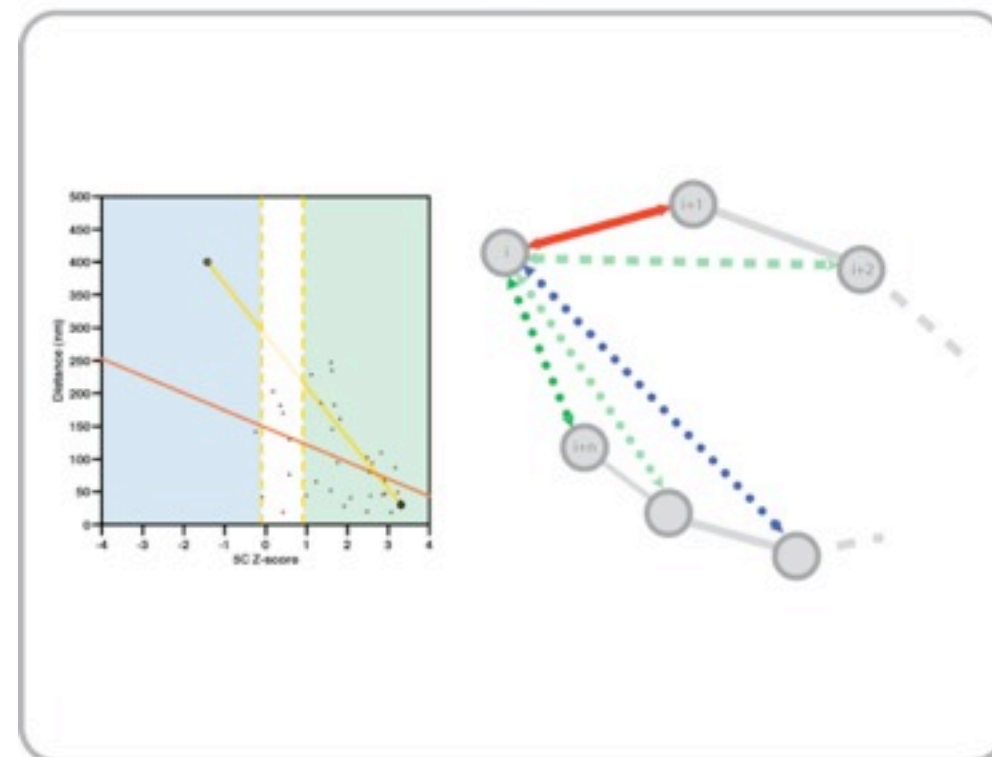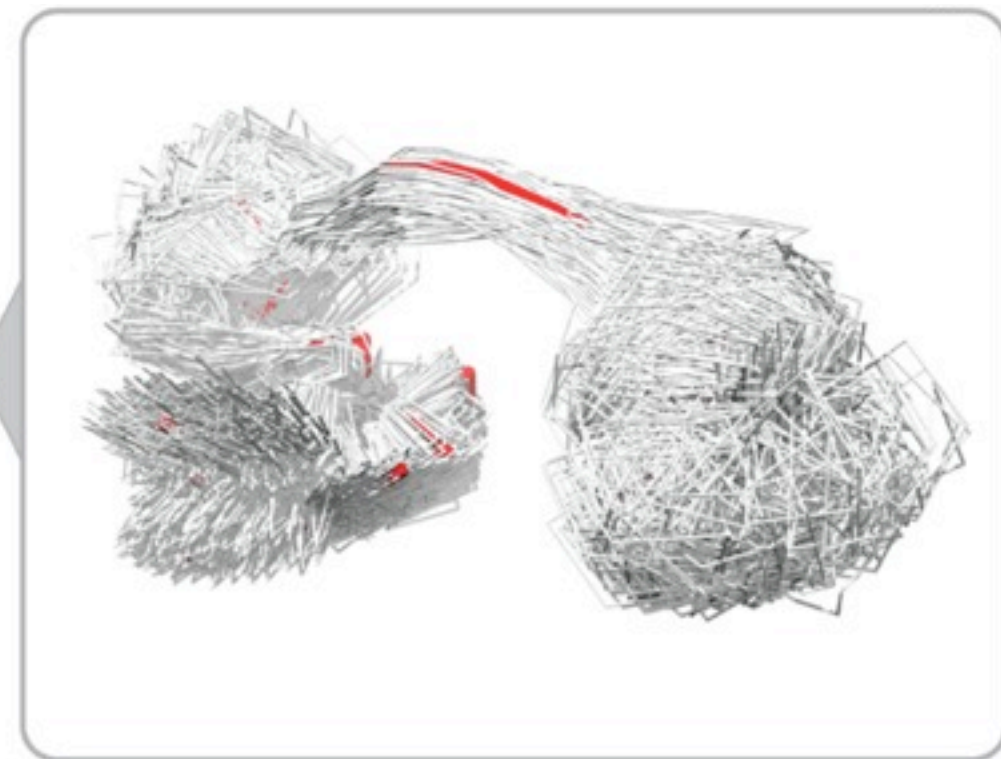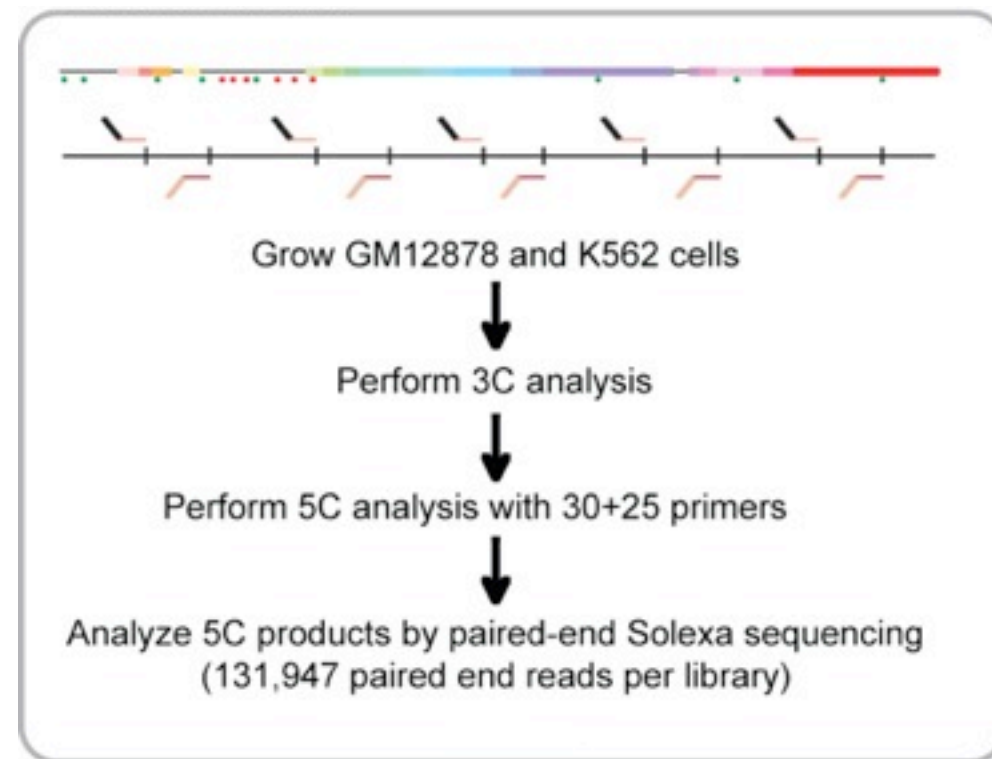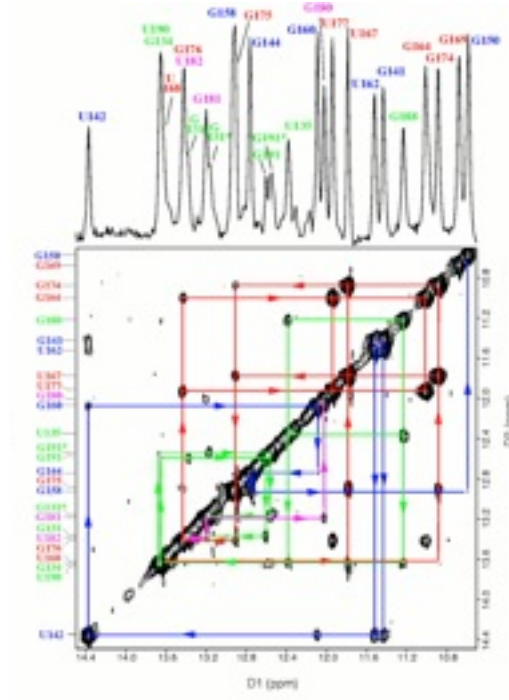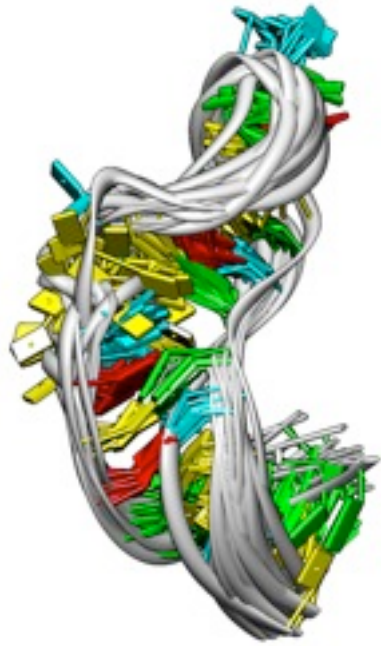
Biomolecular structure determination
2D-NOESY data

Chromosome structure determination
5C data

5C technology

Crosslink DNA — HindIII
AAGCTT
TTCGAA

Cut with restriction enzyme

Ligate — HindIII
AAGCTT
TTCGAA

Purify DNA

Annealing locus-specific 5C primers

Ligation mediated amplification
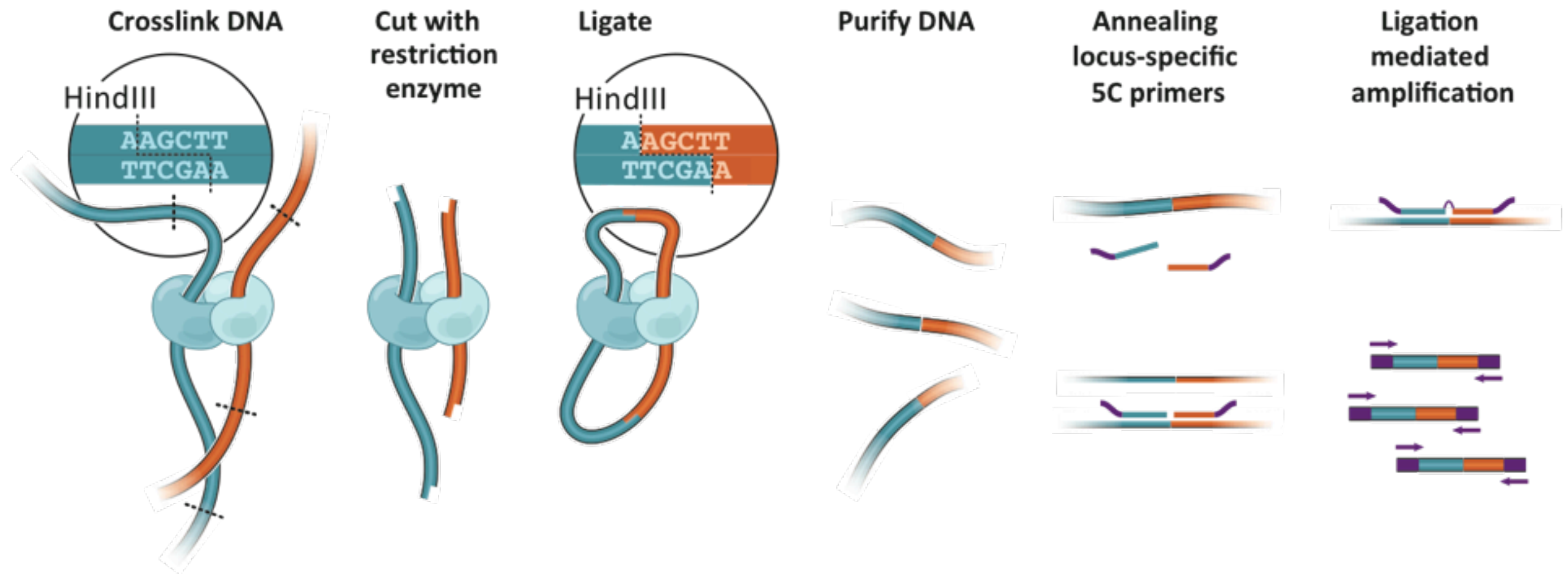
5C "copies" the 3C library into a 5C library containing only ligation junctions

Performed at high levels of multiplexing:

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

Thursday, October 13, 11

Integrative Modeling

http://www.integrativemodeling.org
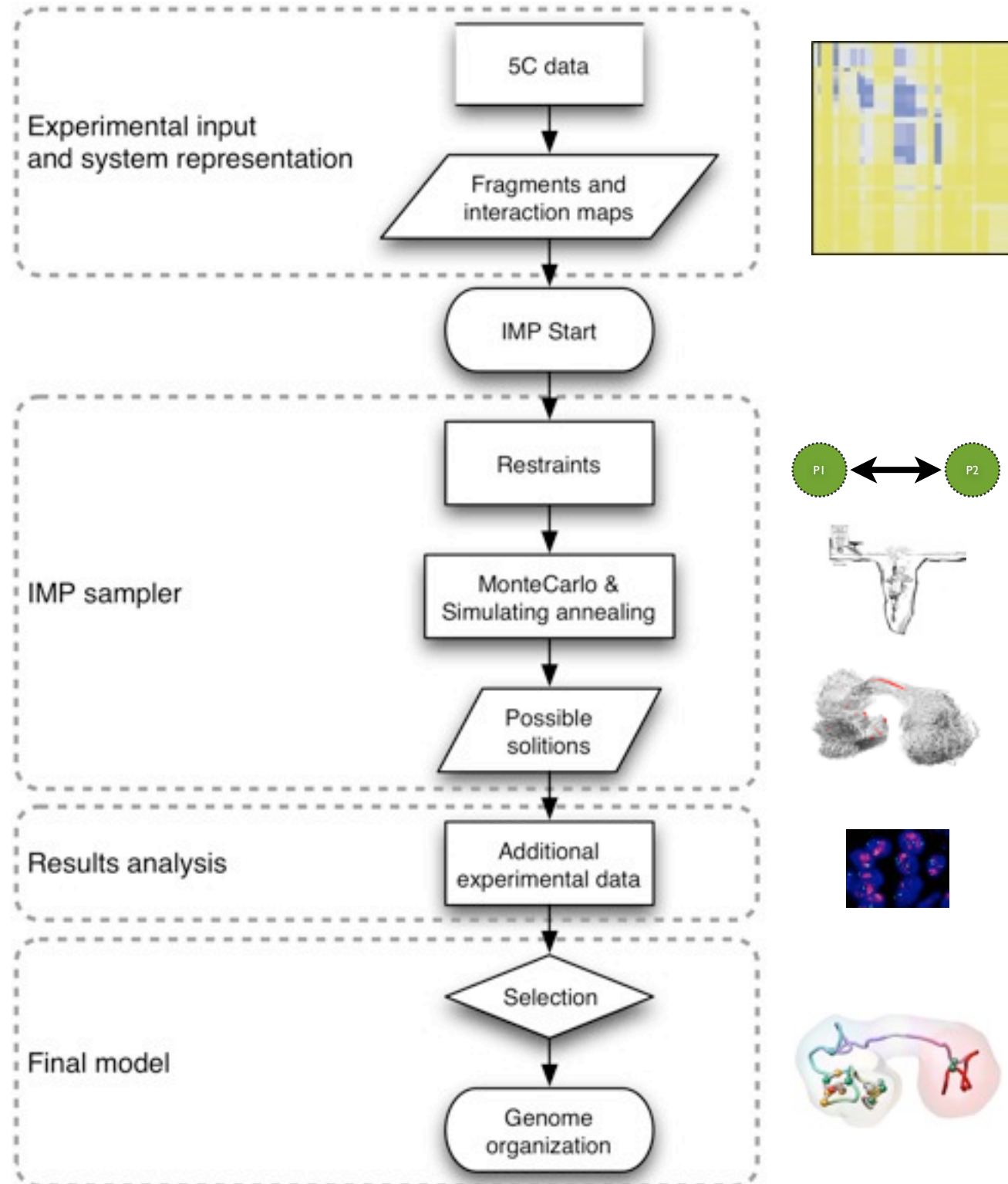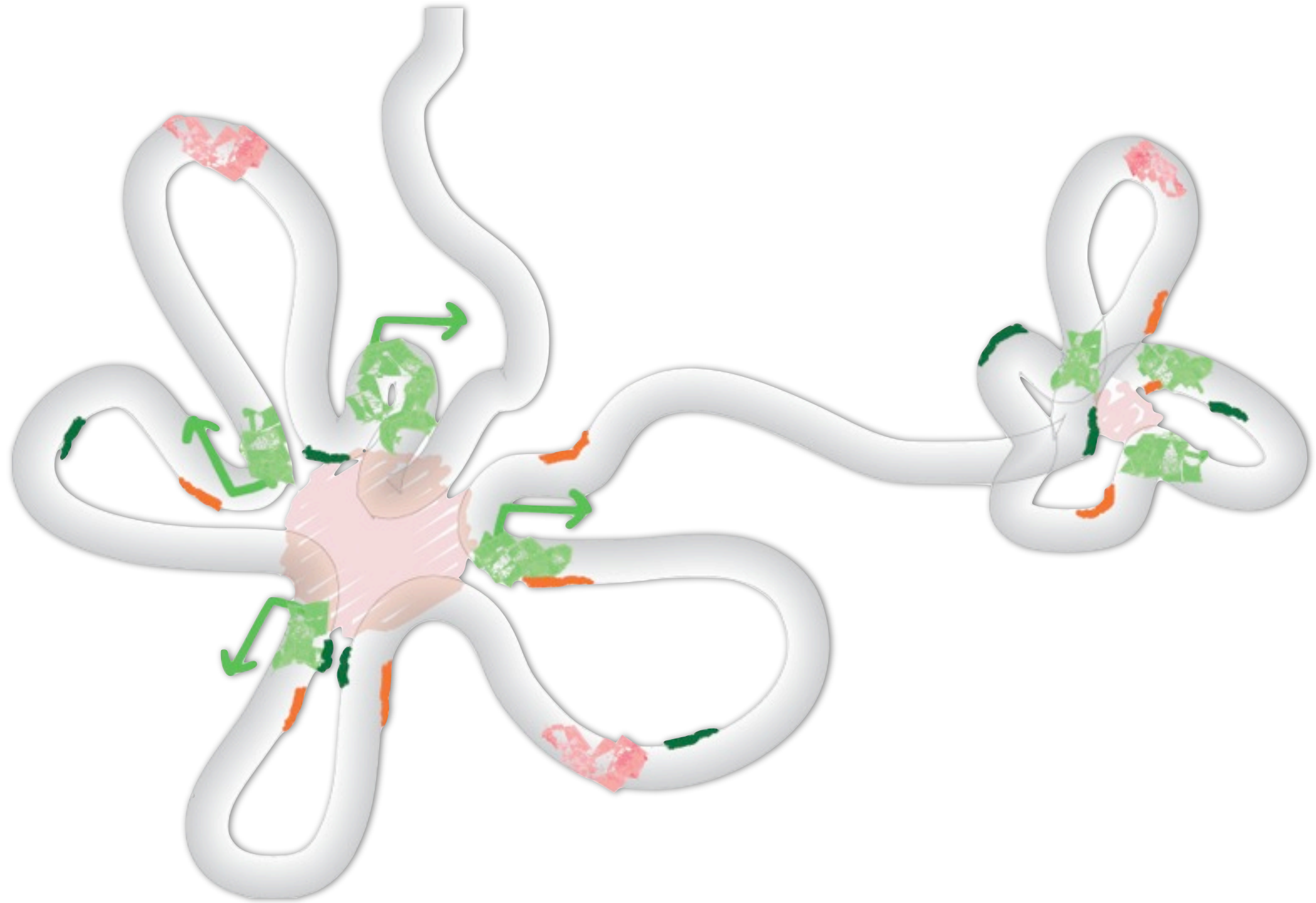
Human α-globin domain

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).

Thursday, October 13, 11

Human α-globin domain

ENm008 genomic structure and environment

ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816
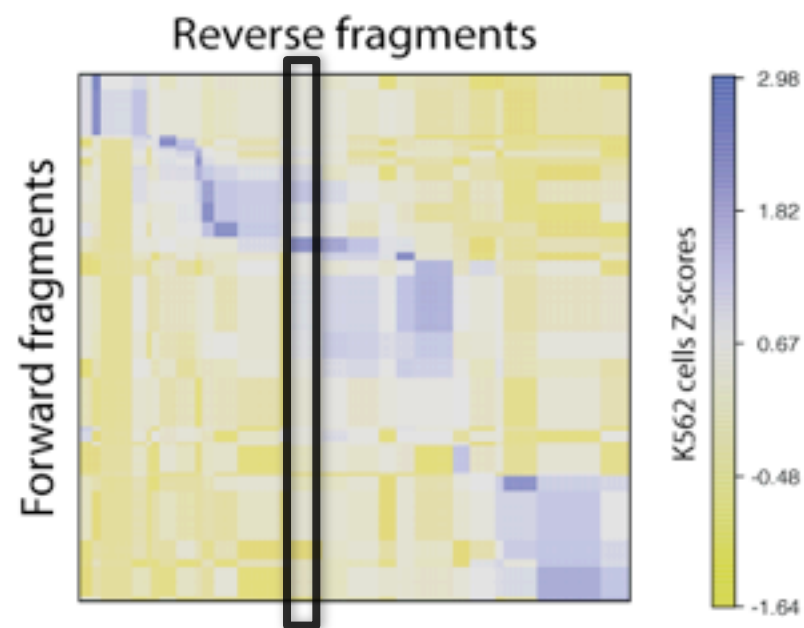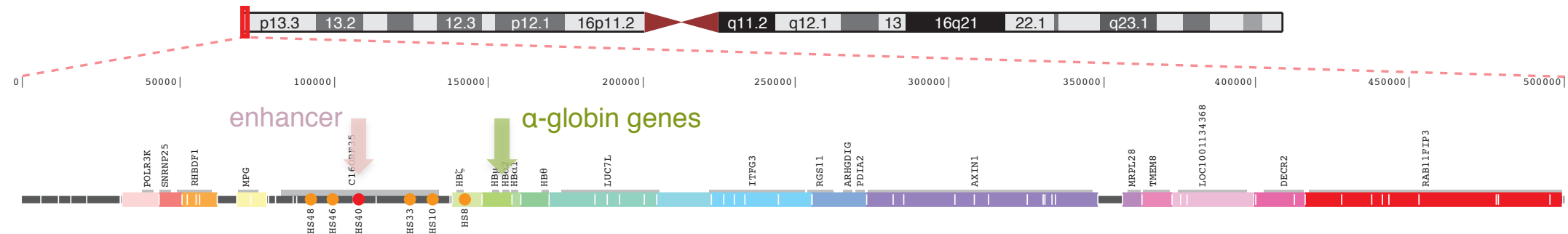
K562 cells:
α-globin genes active
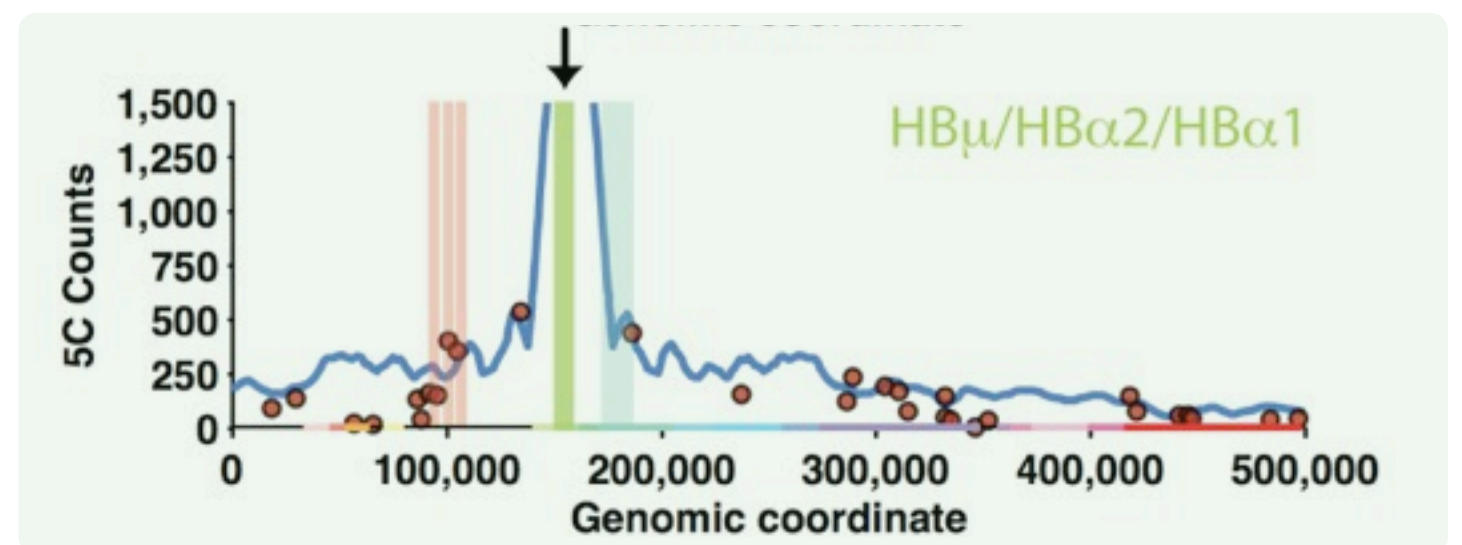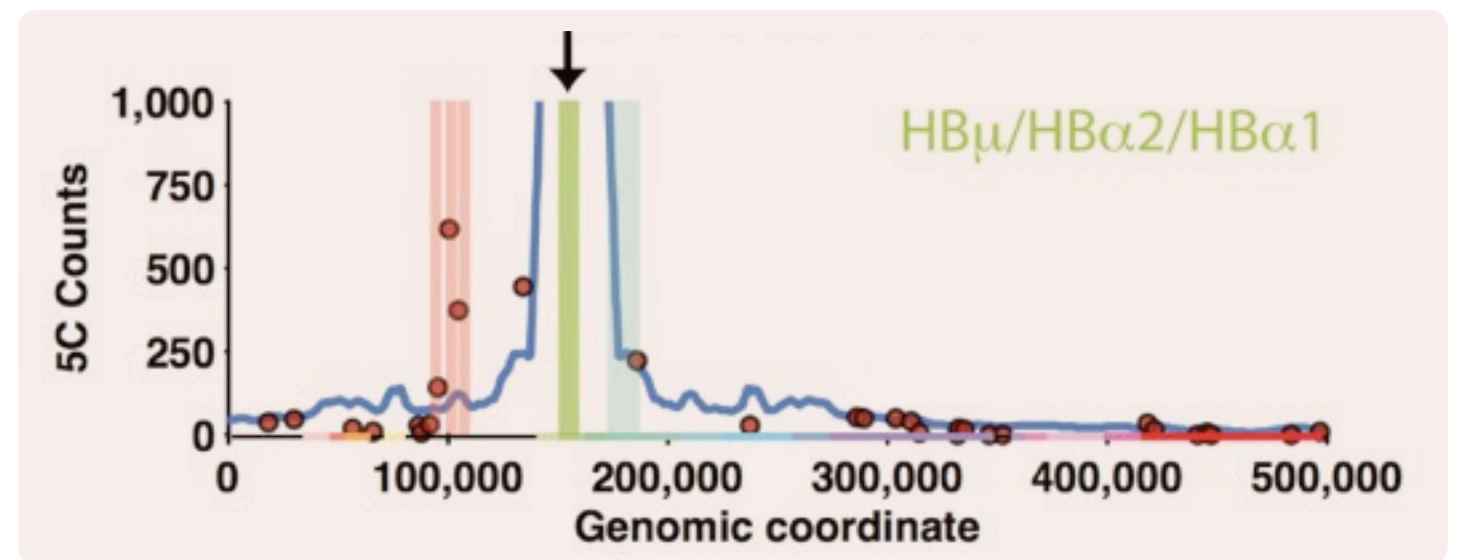
Representation

Harmonic

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

Harmonic Lower Bound

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

Harmonic Upper Bound

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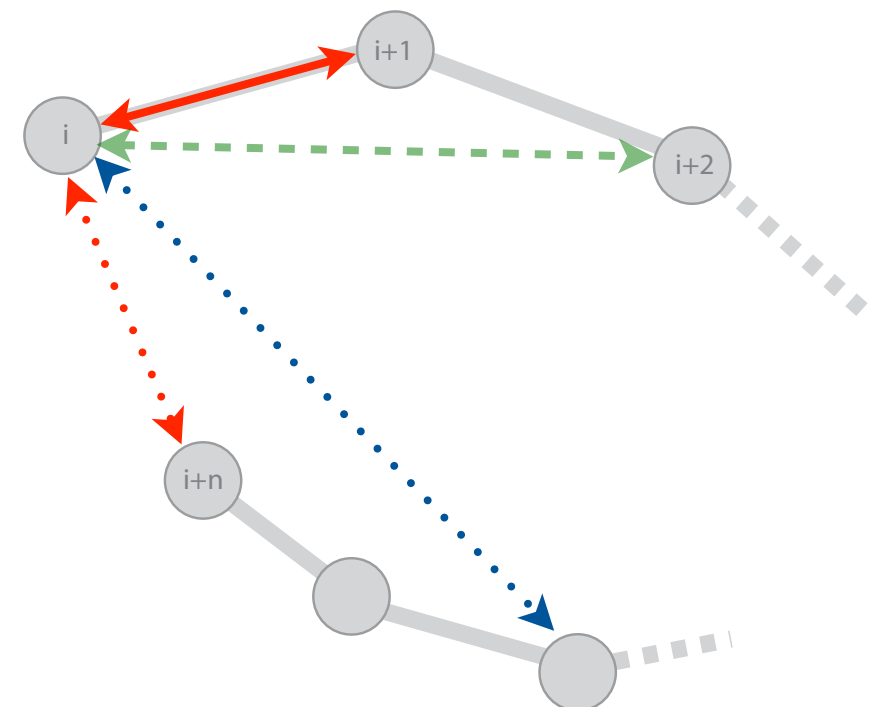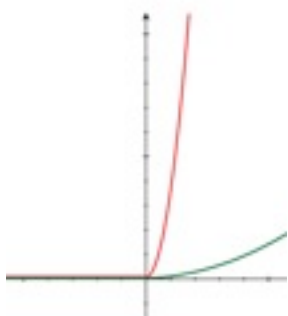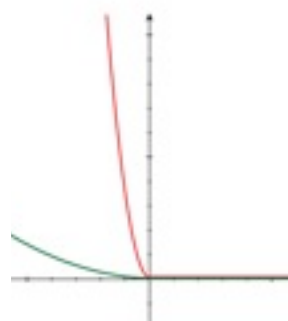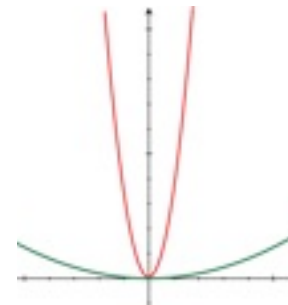
Scoring

GM12878

70 fragments
1,520 restraints
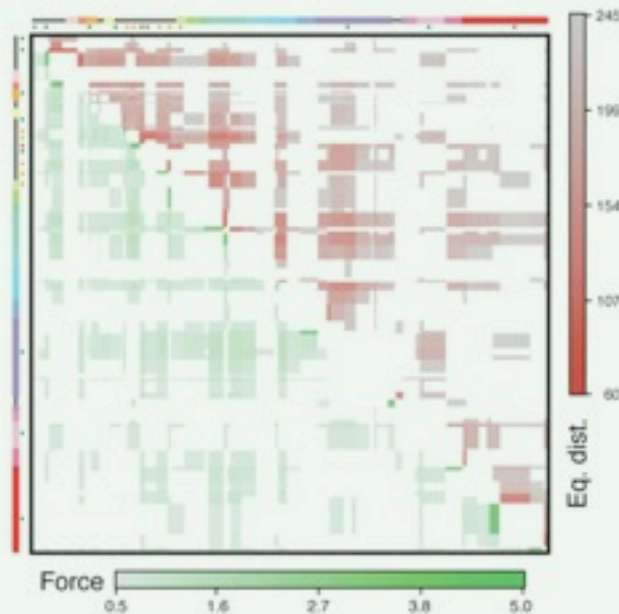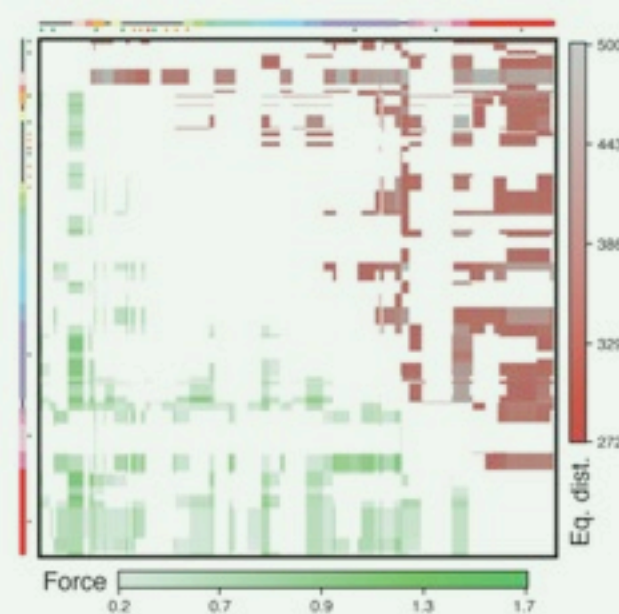
Harmonic   Harmonic Lower Bound   Harmonic Upper Bound
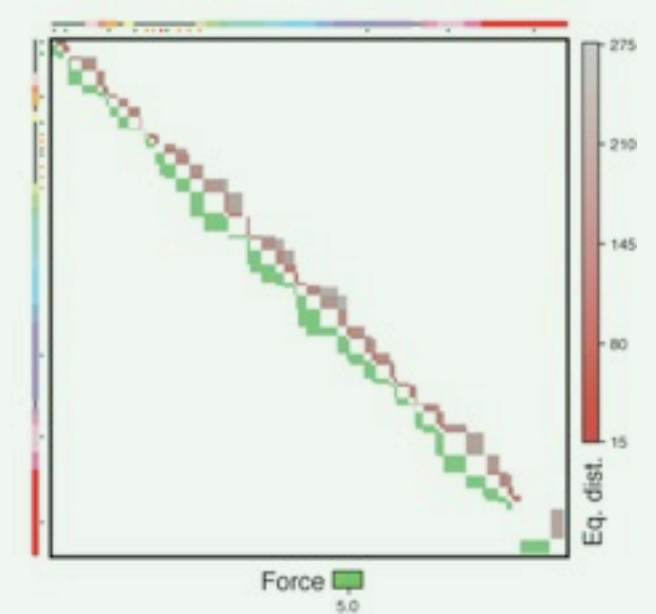
K562

70 fragments
1,049 restraints
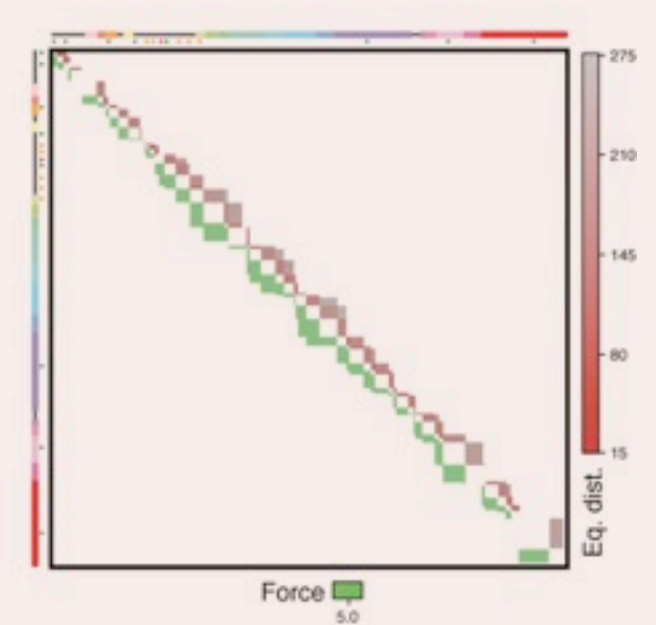
Optimization

Clustering

14
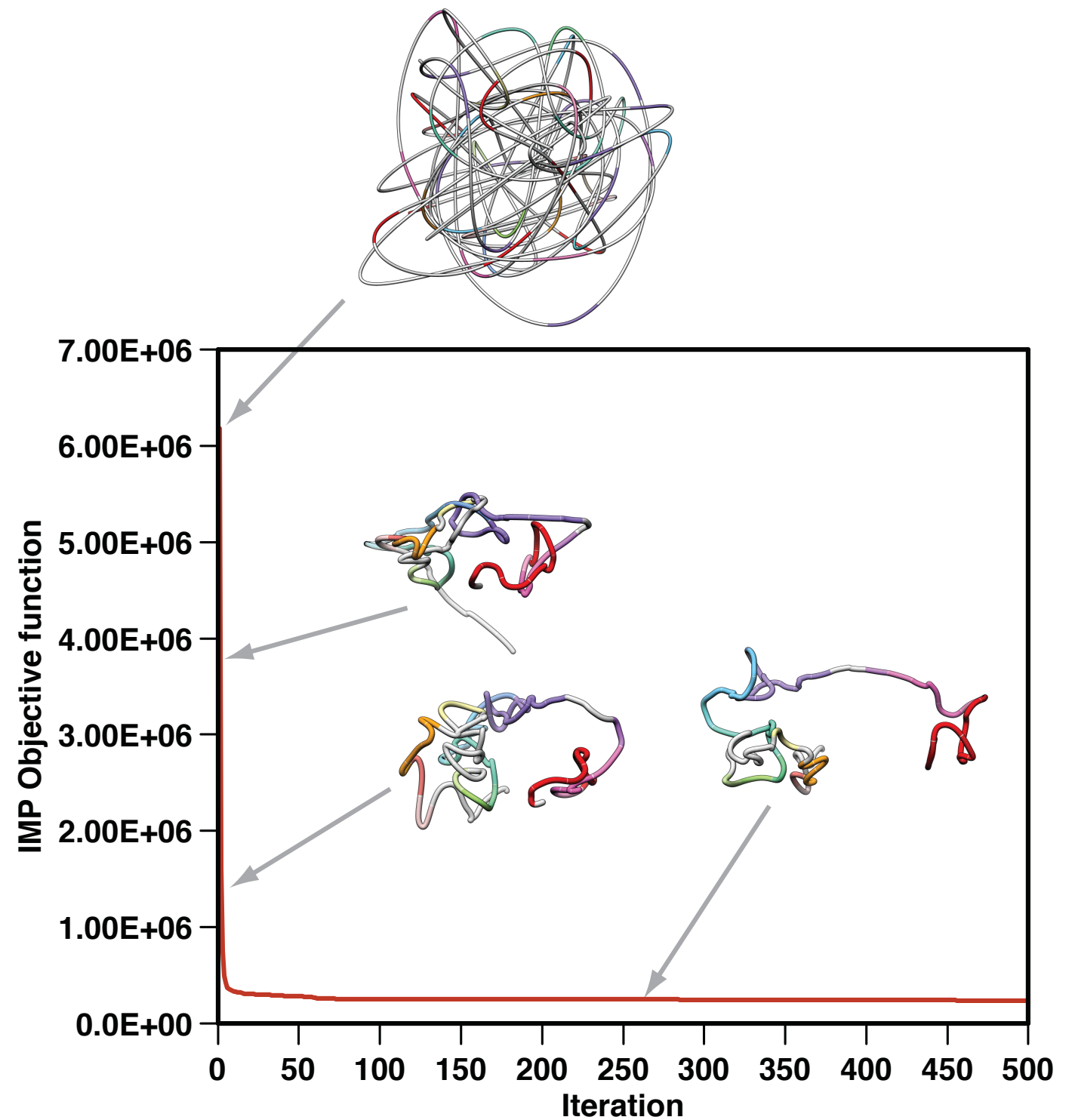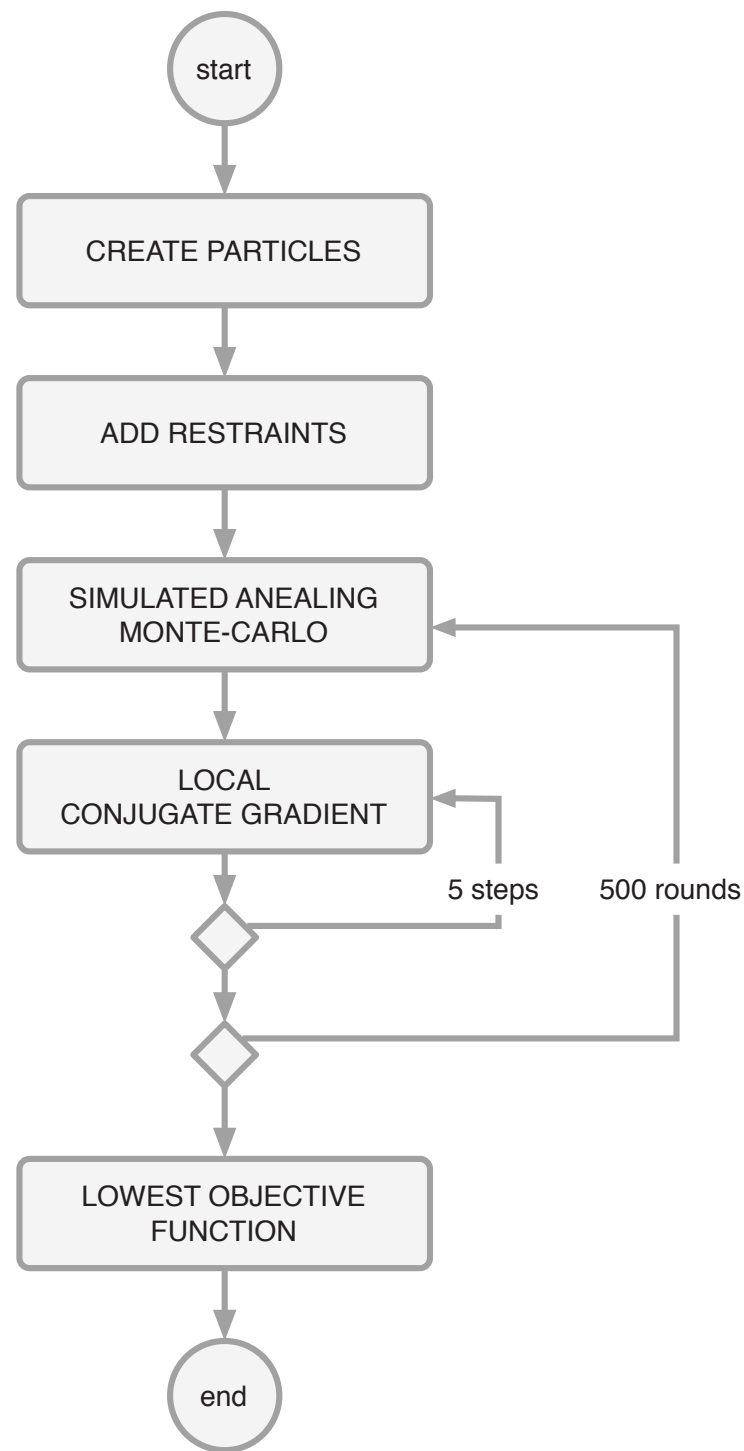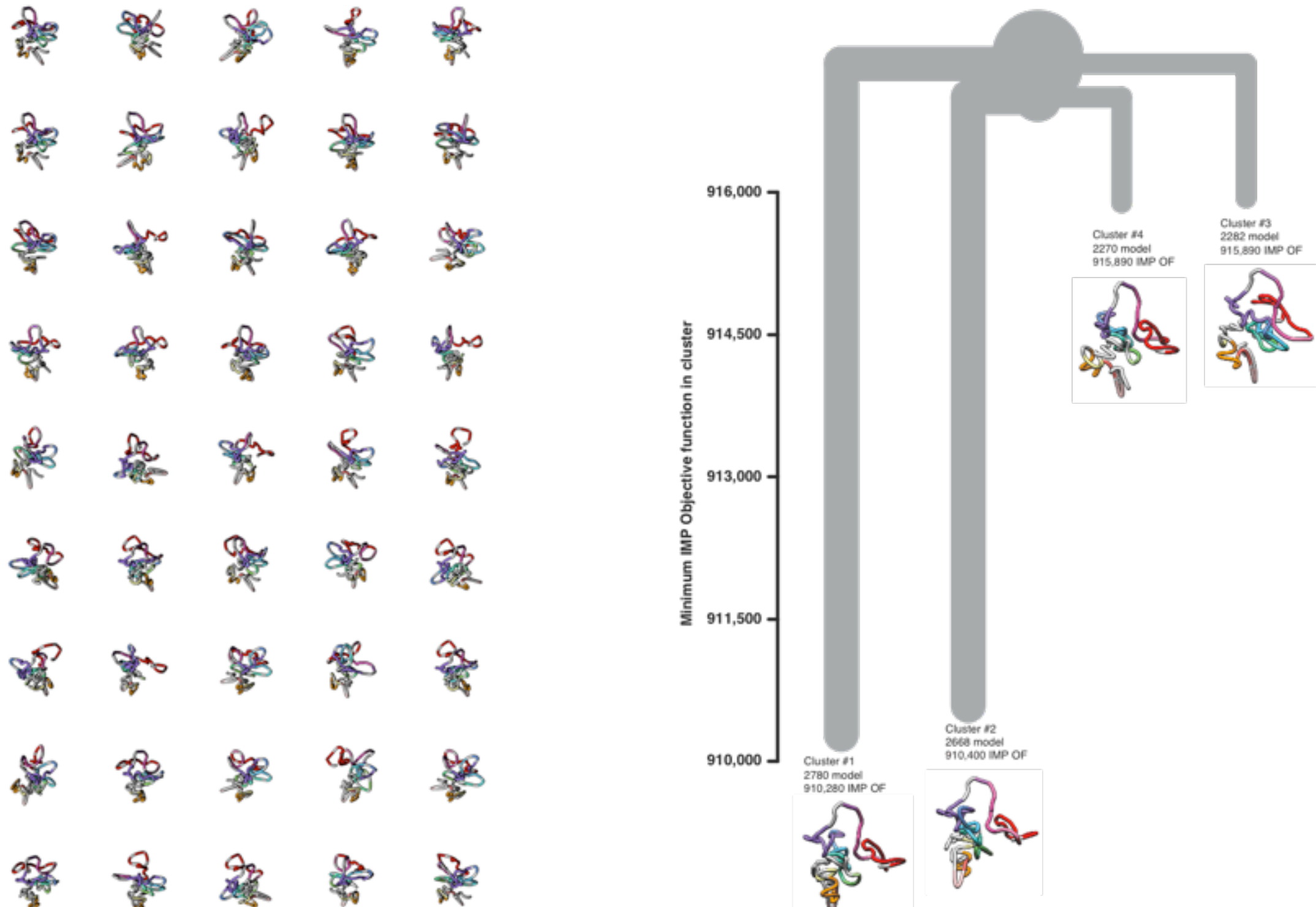
Not just *one* solution

GM12878
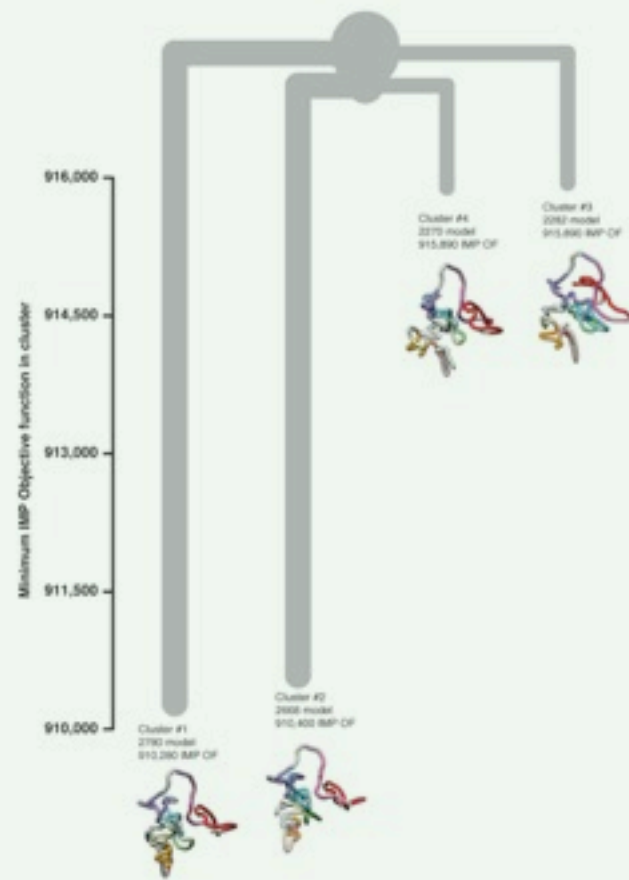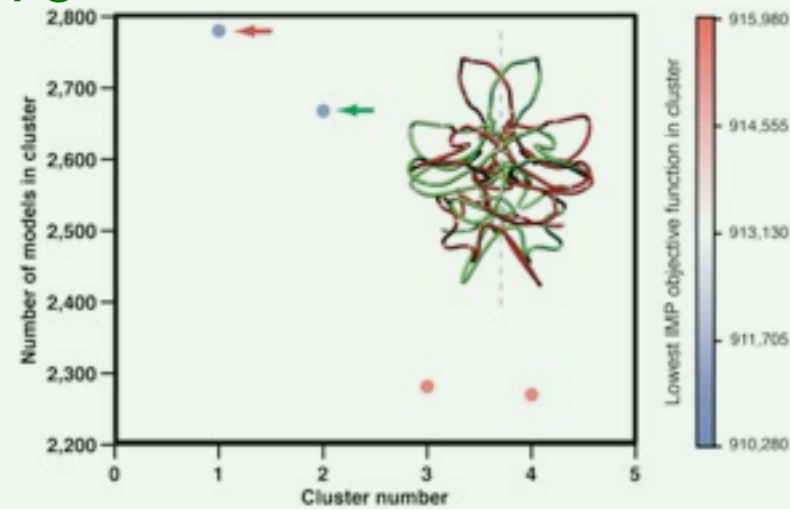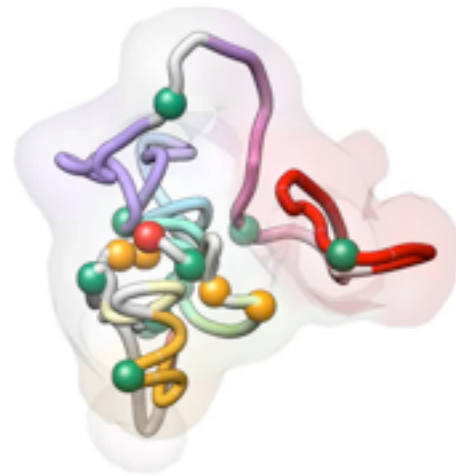
K562
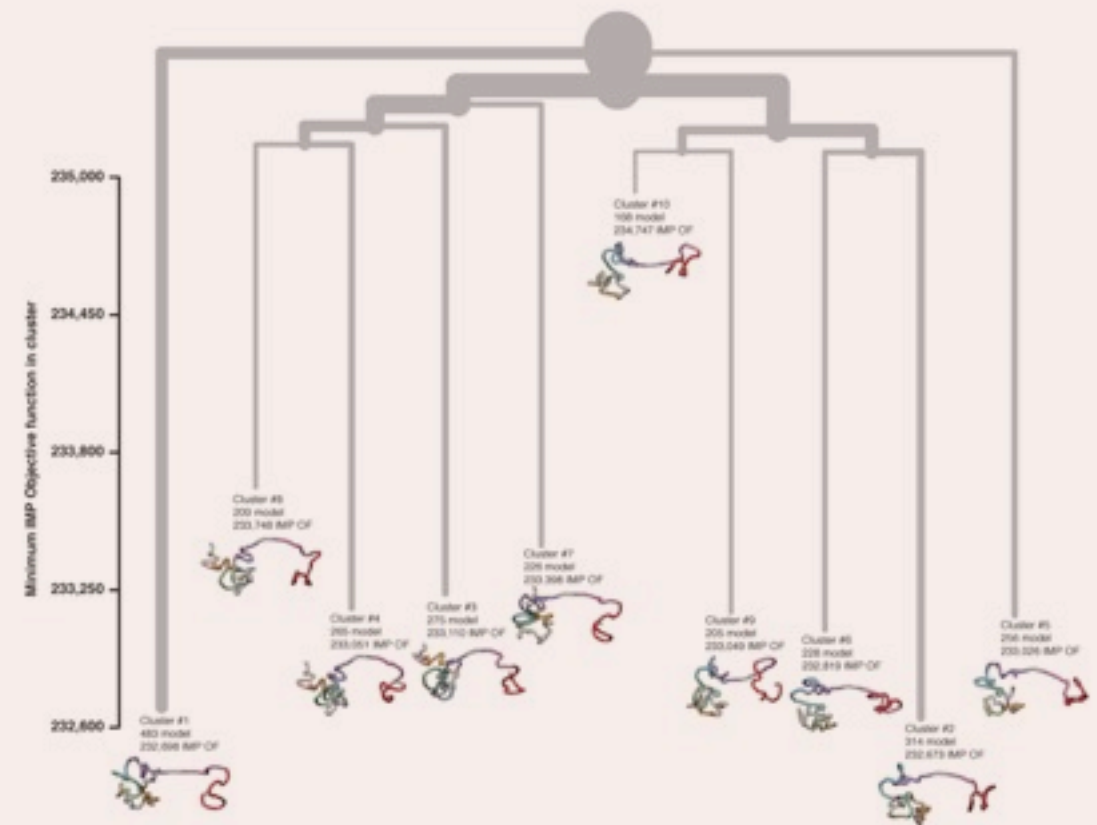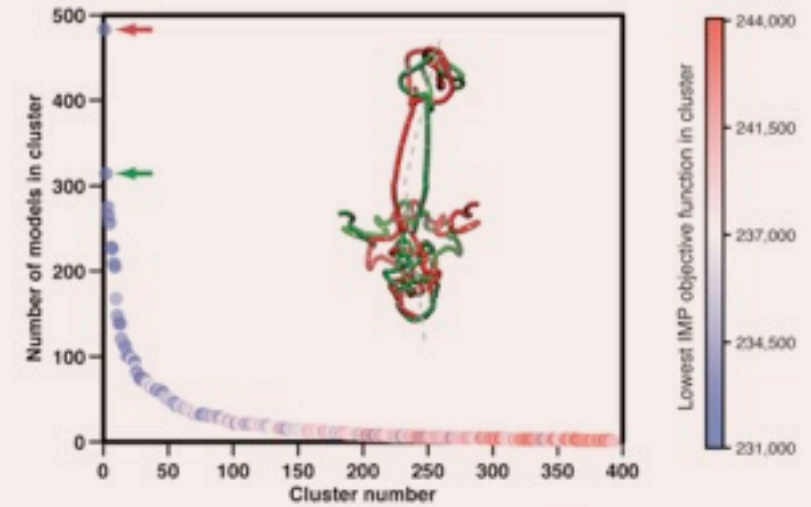
Consistency
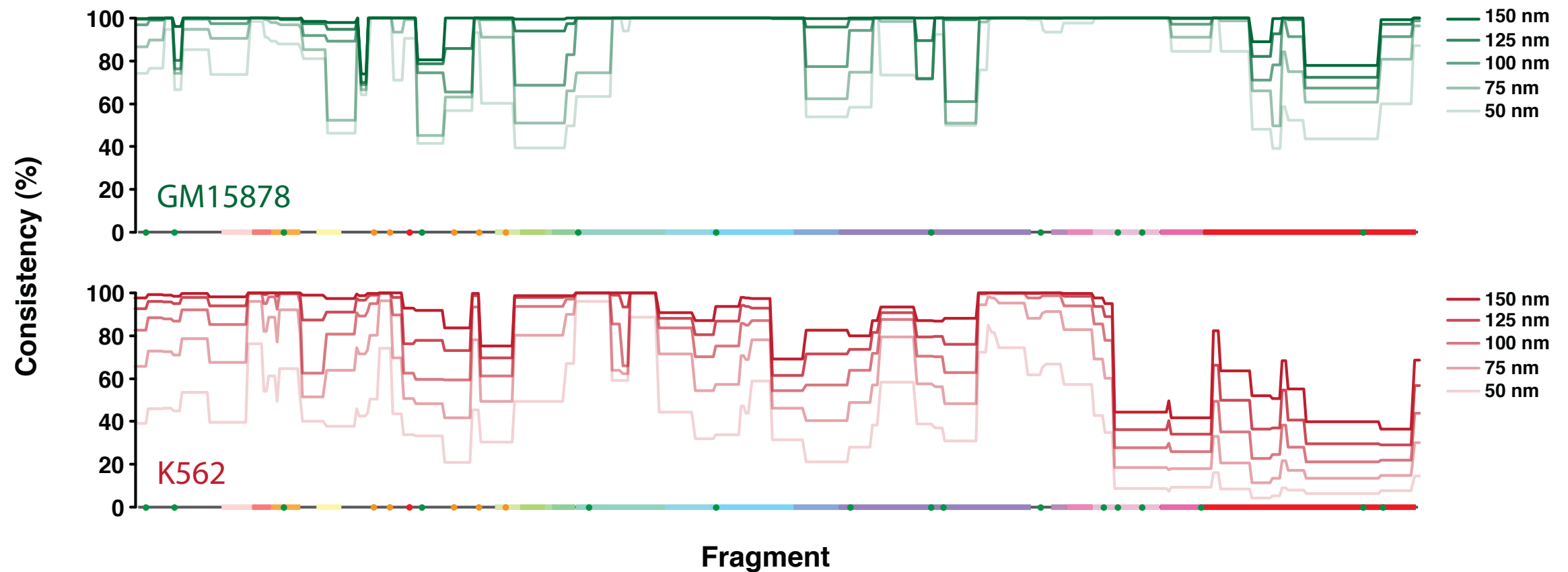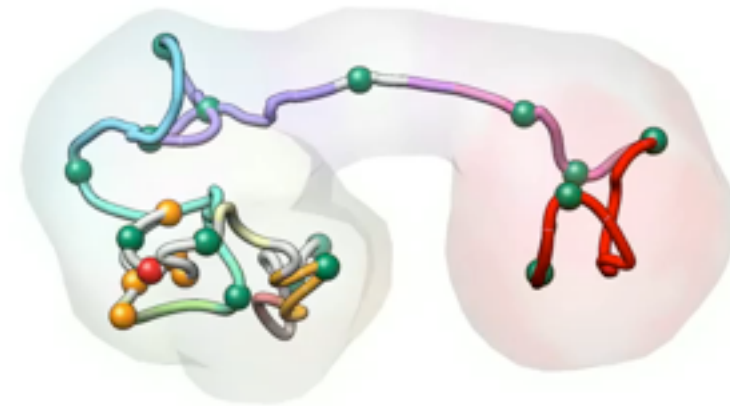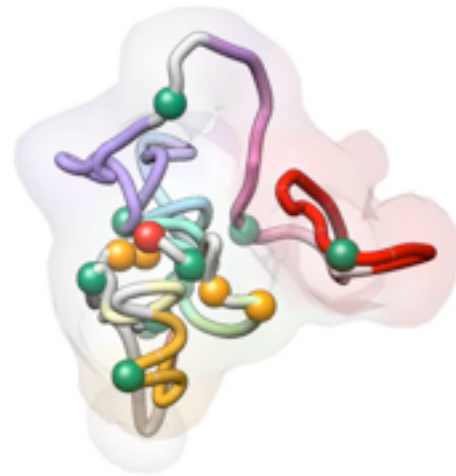
Regulatory elements
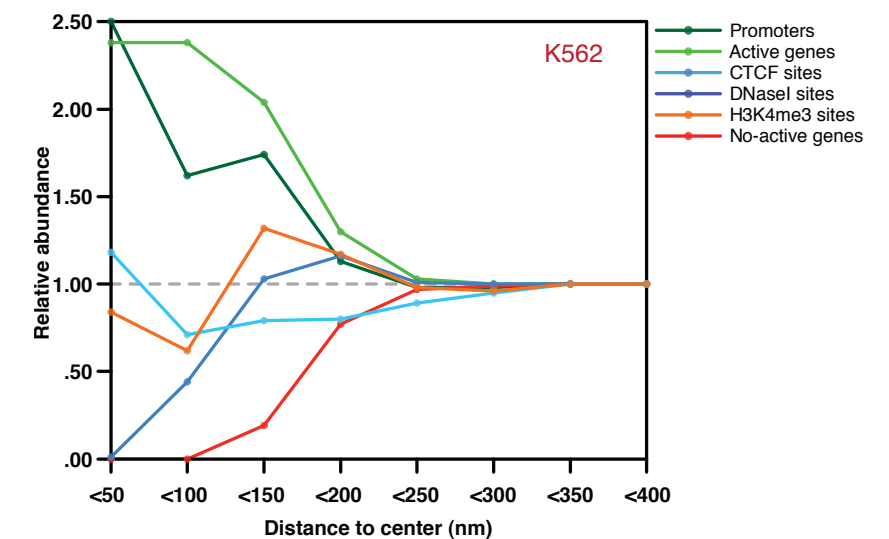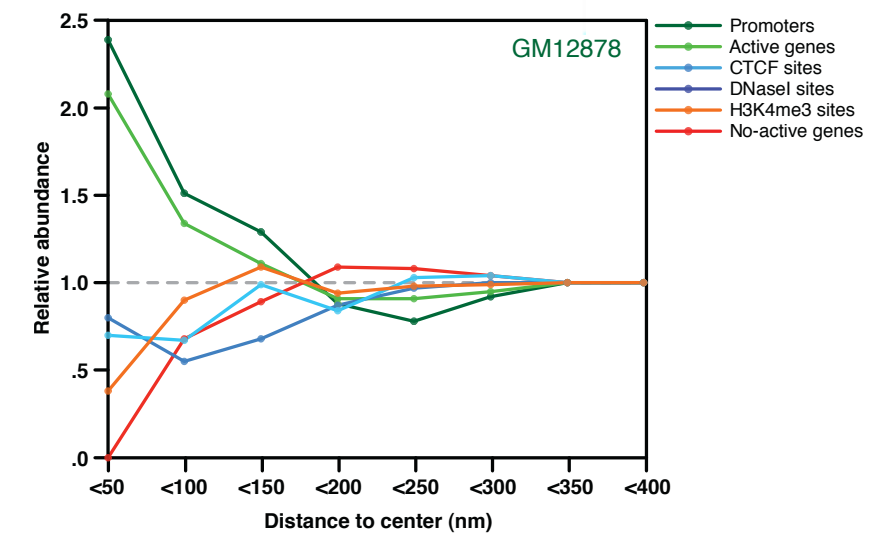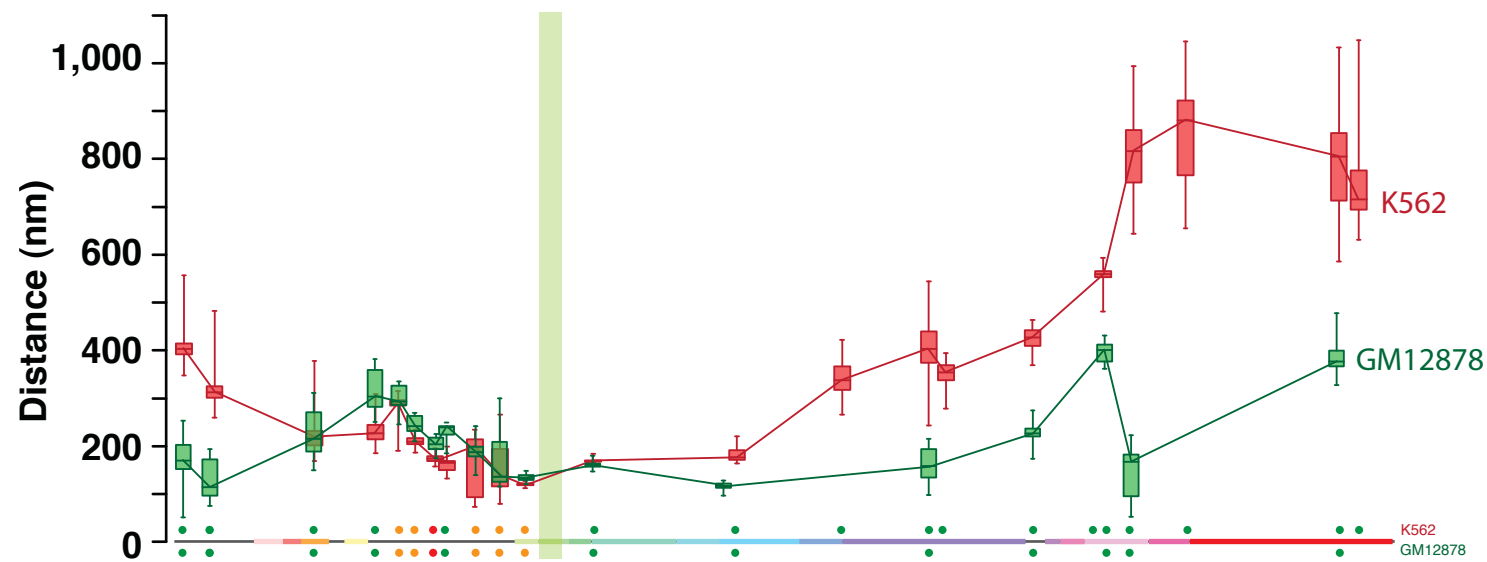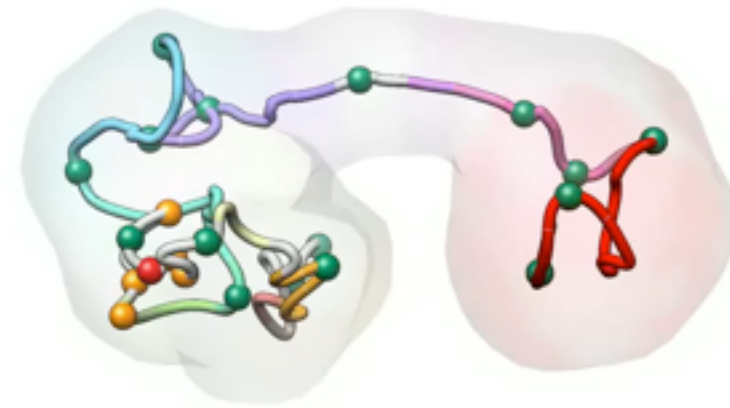
GM12878
Cluster #1
2780 model

K562
Cluster #2
314 model

Compactness
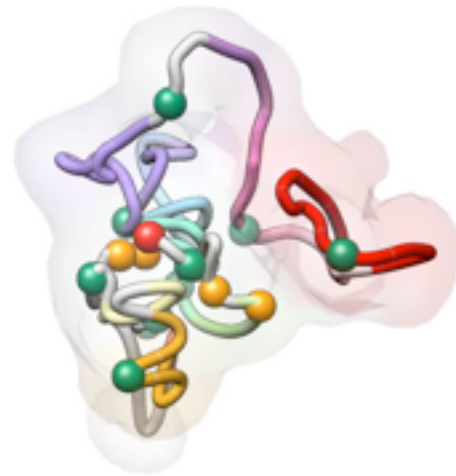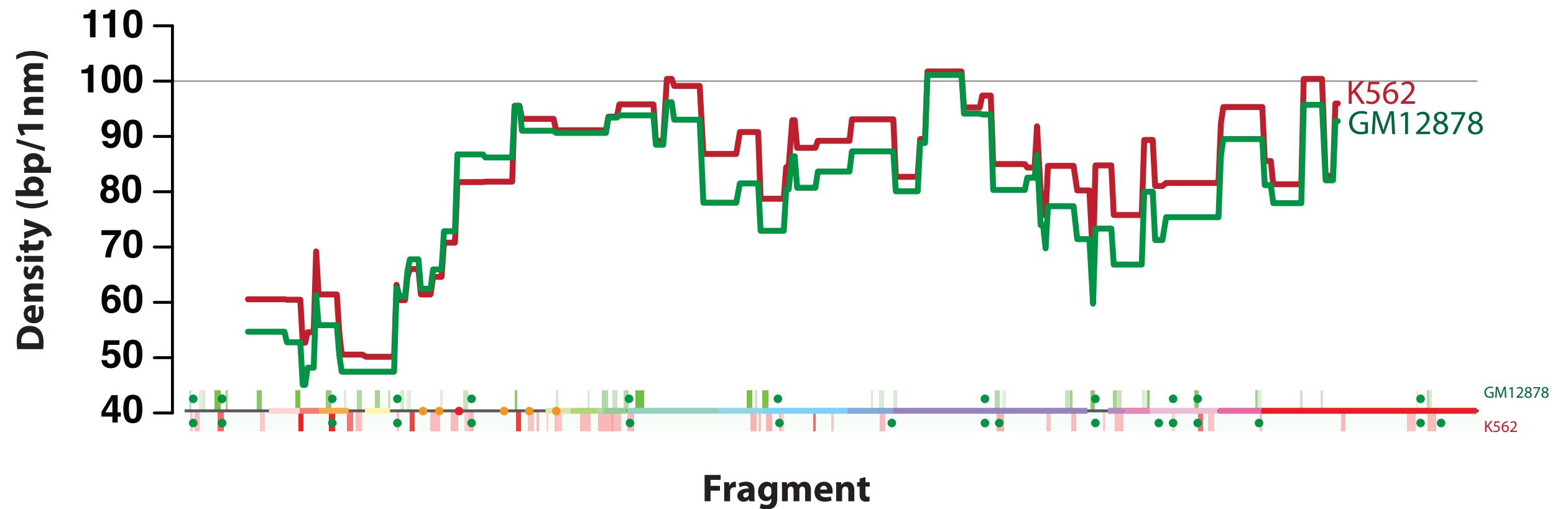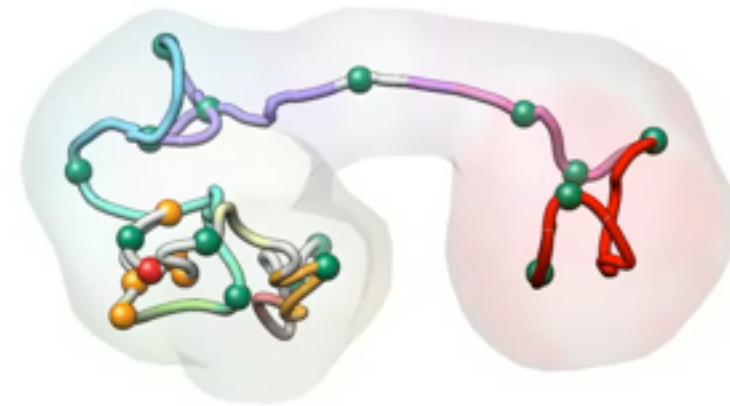
Multi-loops
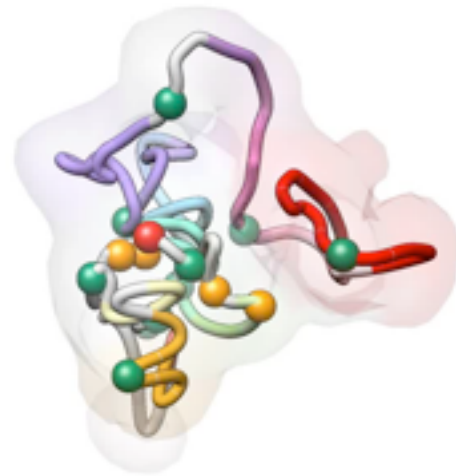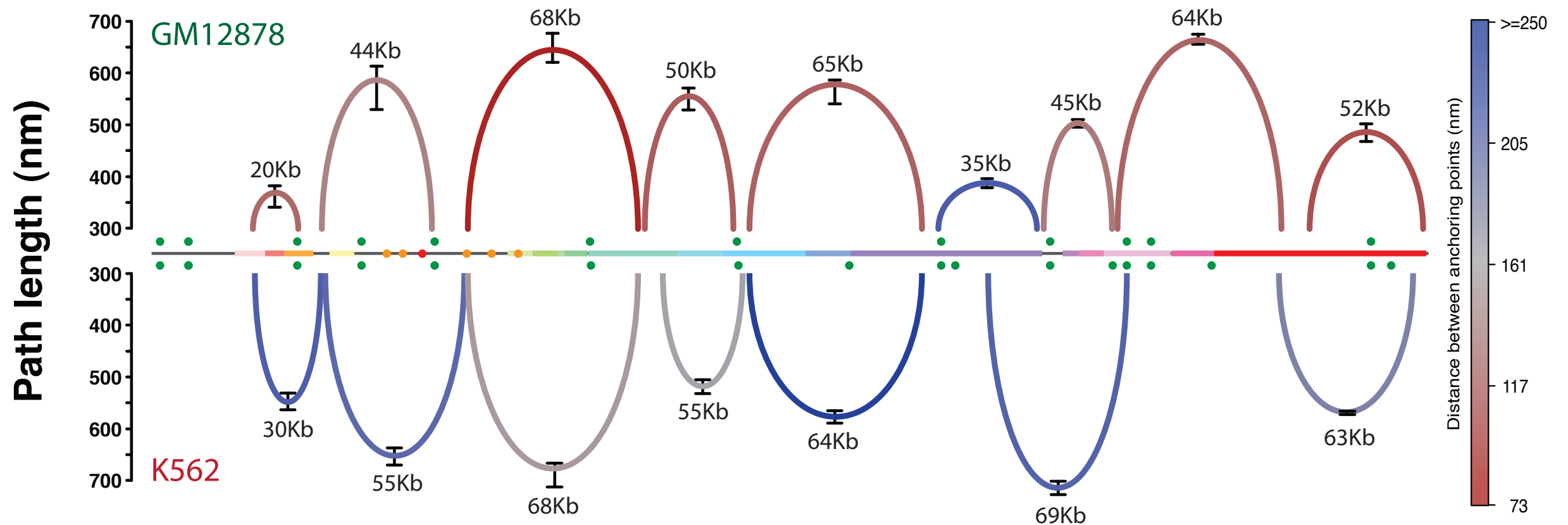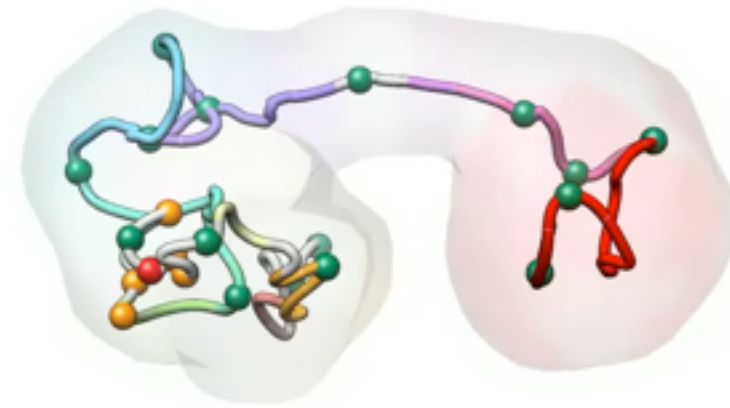
Expression

GM12878
Cluster #1
2780 model

K562
Cluster #2
314 model

Increased in GM12878
=
Increased in K562

16
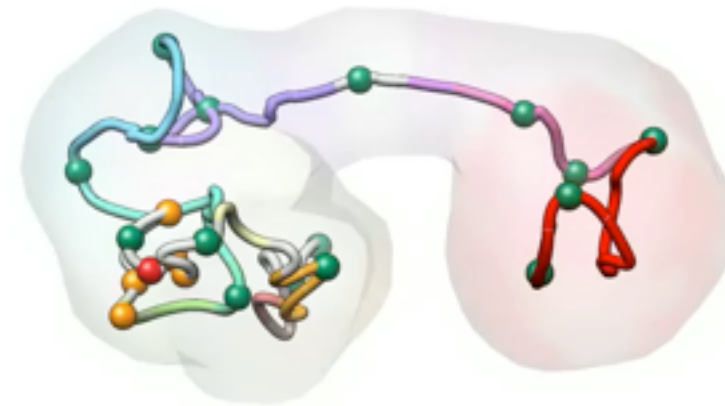
FISH validation

GM12878
Cluster #1
2780 model
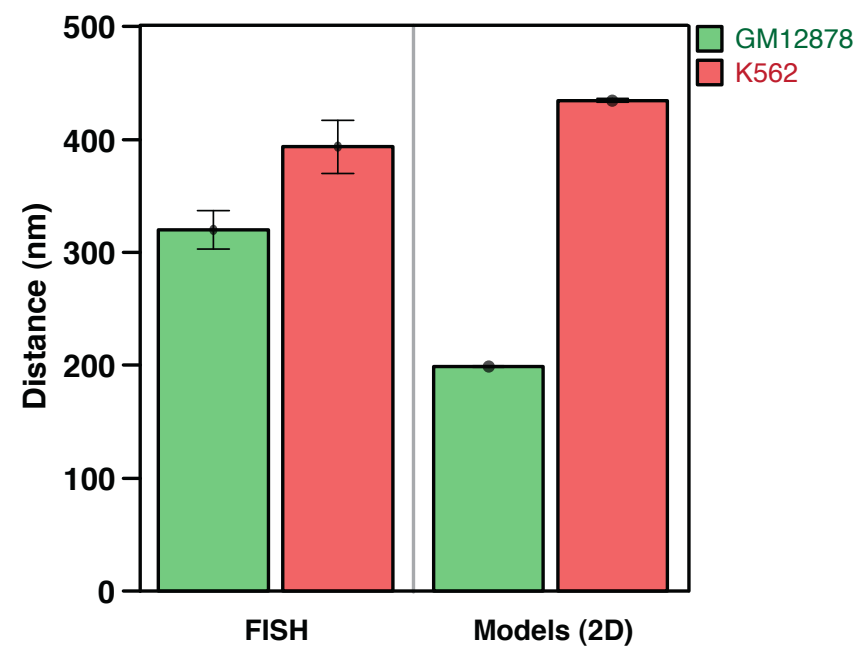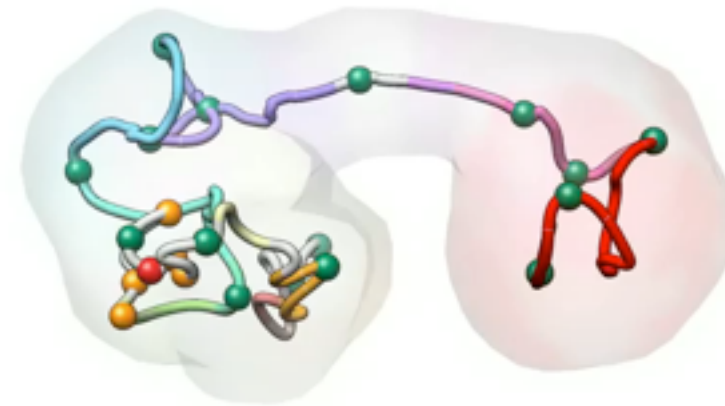
K562
Cluster #2
314 model
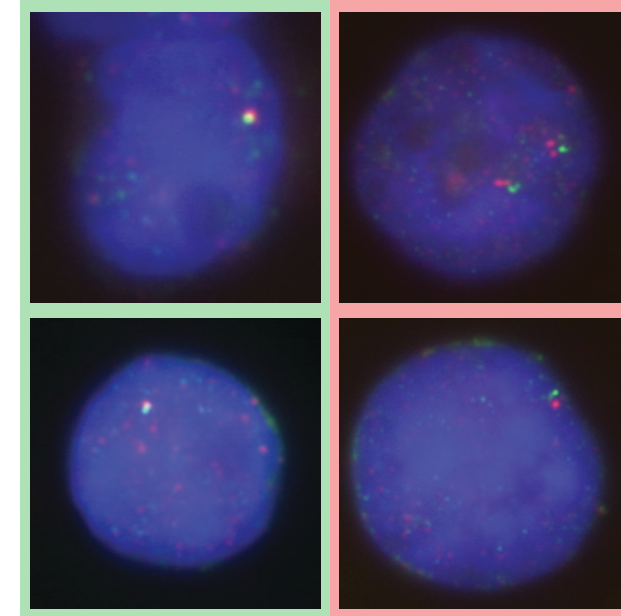
The "Chromatin Globule" model

Active genes
Inactive genes
CTCF sites
HS sites
Globule core

(a)
chromatin
subcompartment
subcompartment
loop base spring
(magnified)
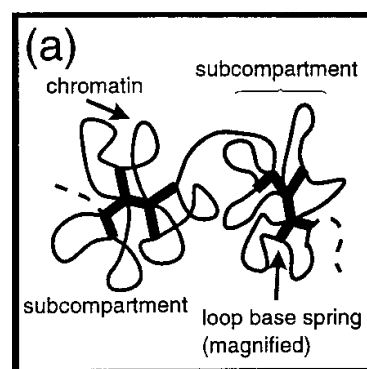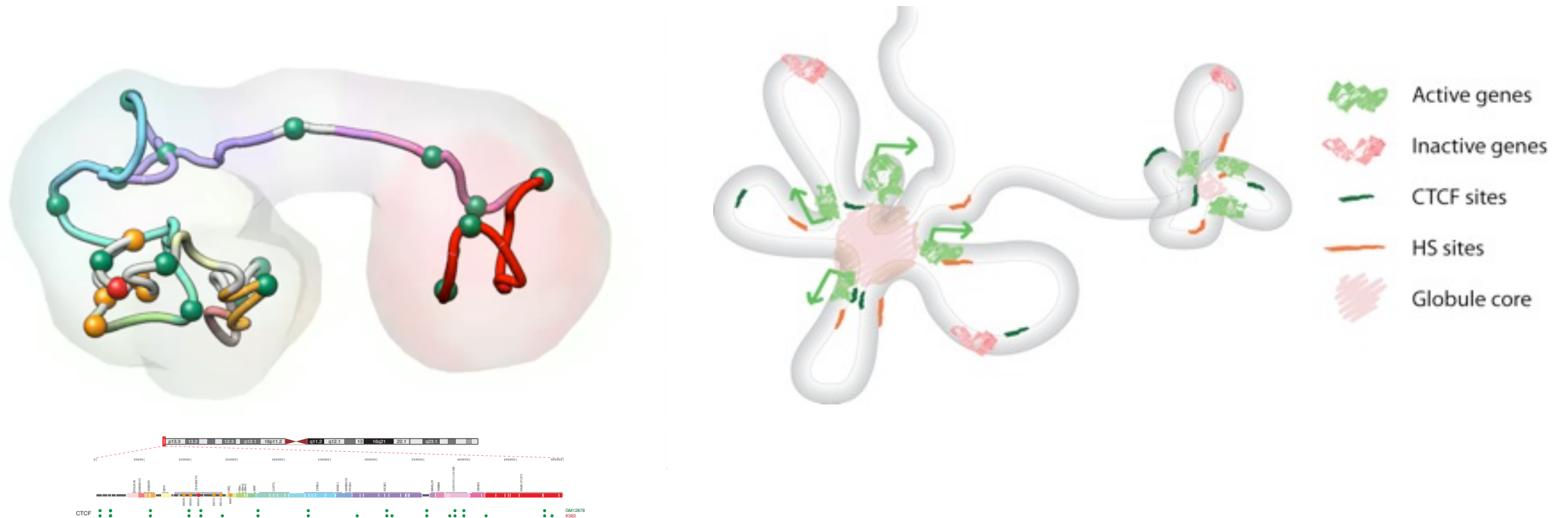
Münkel et al. JMB (1999)

a
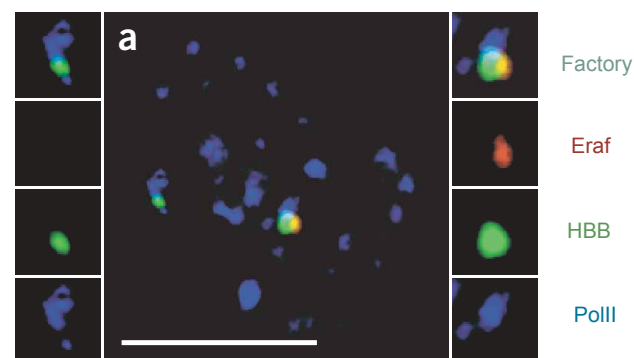
Osborne et al. Nat Genet (2004)
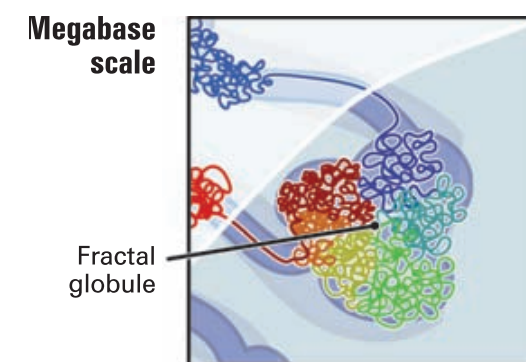
al. Science (2009)

17

Caulobacter crescentus genome
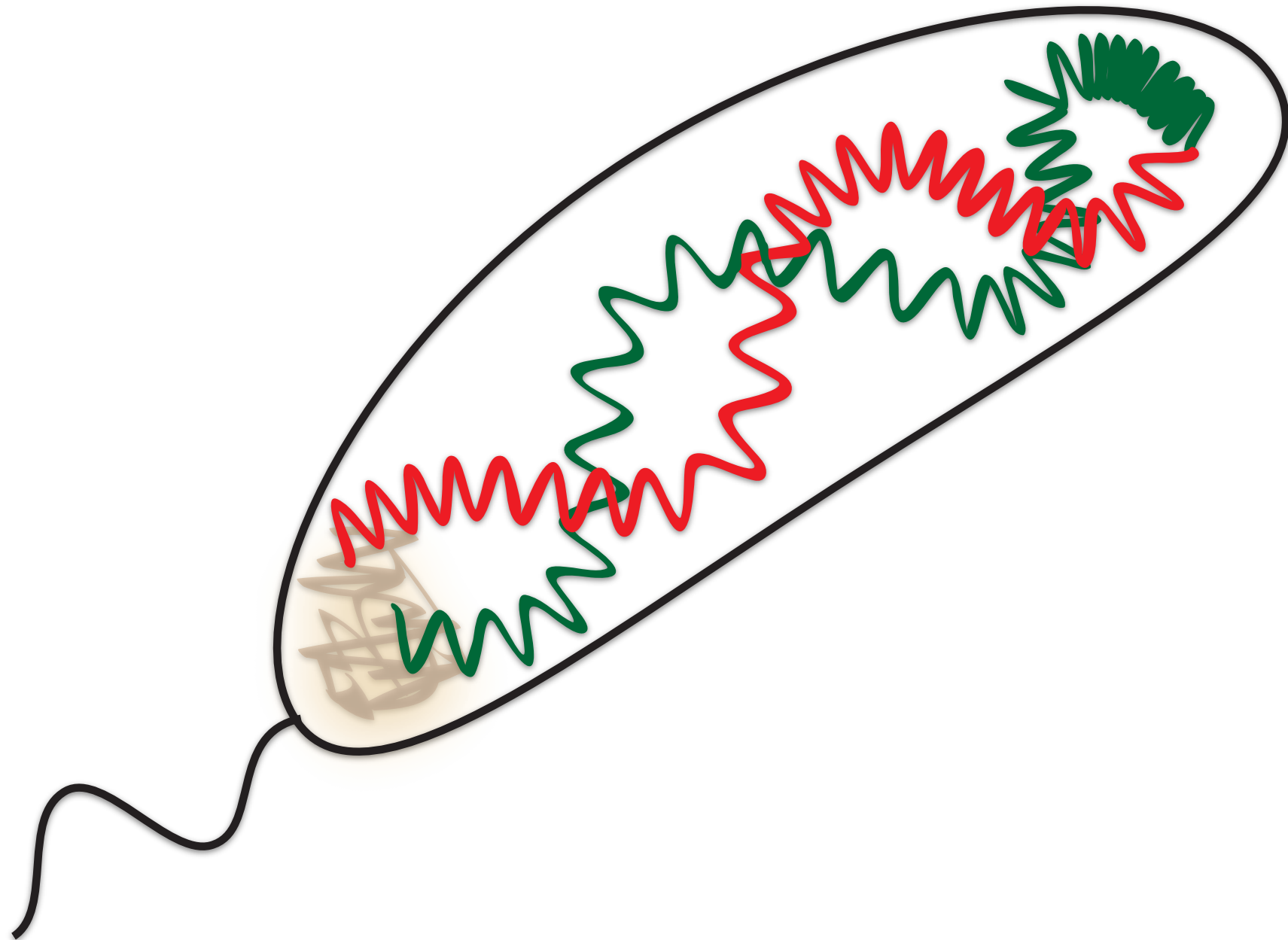
The 3D architecture of *Caulobacter Crescentus*

4,016,942 bp & 3,767 genes
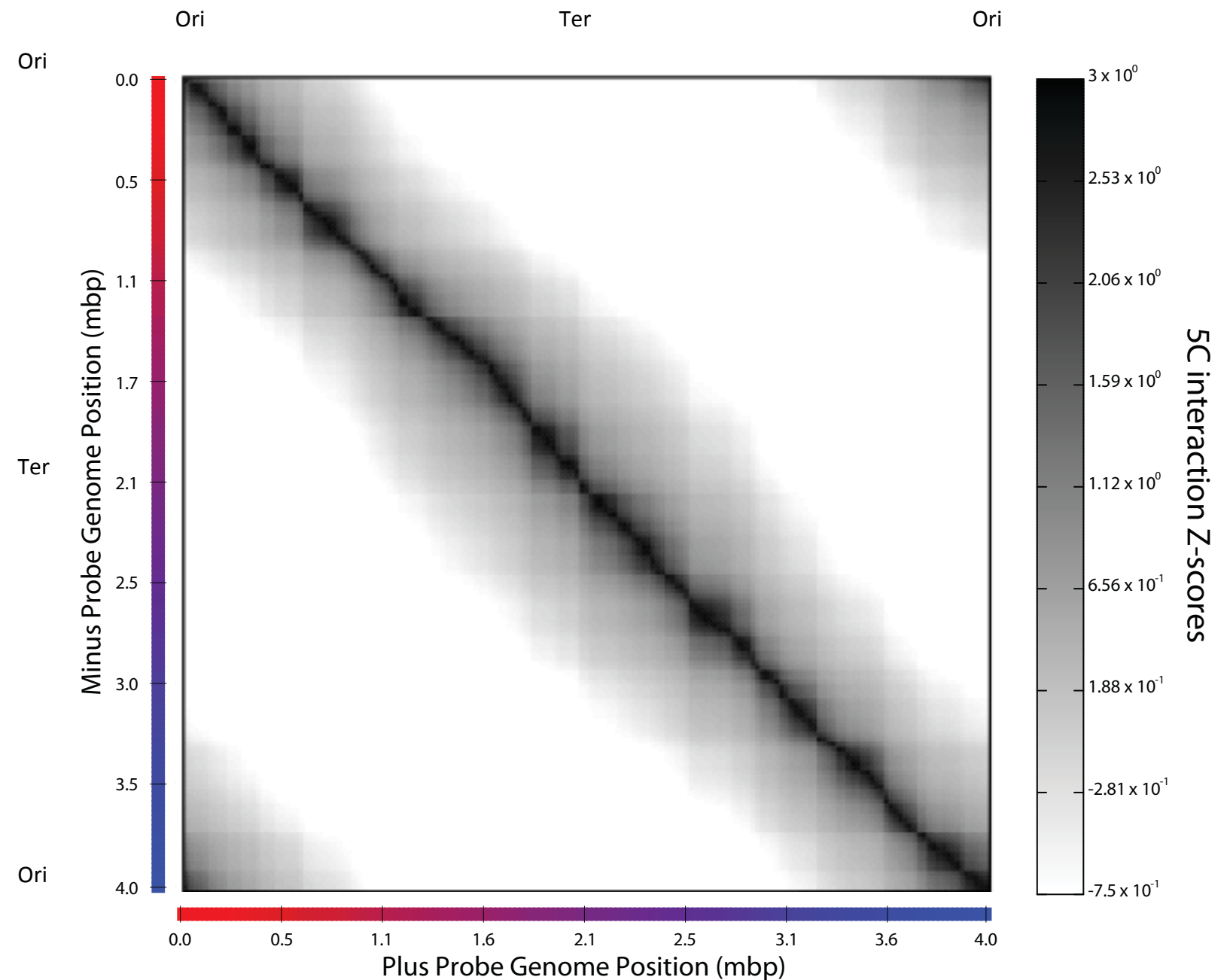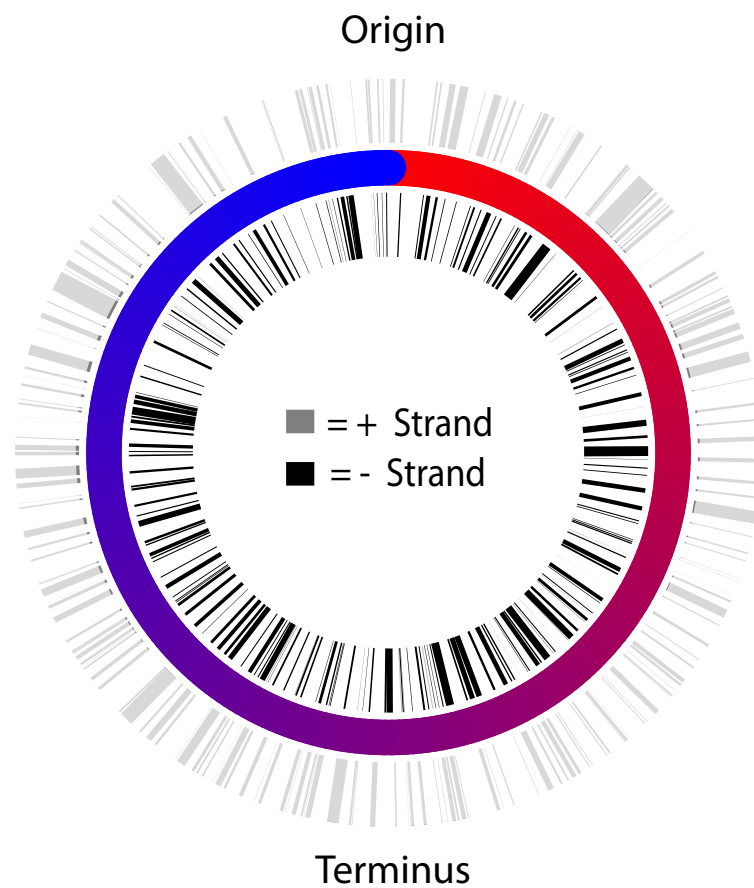
169 5C primers on + strand

170 5C primers on – strand

28,730 chromatin interactions

~13Kb

19

5C interaction matrix
ELLIPSOID for Caulobacter cresentus
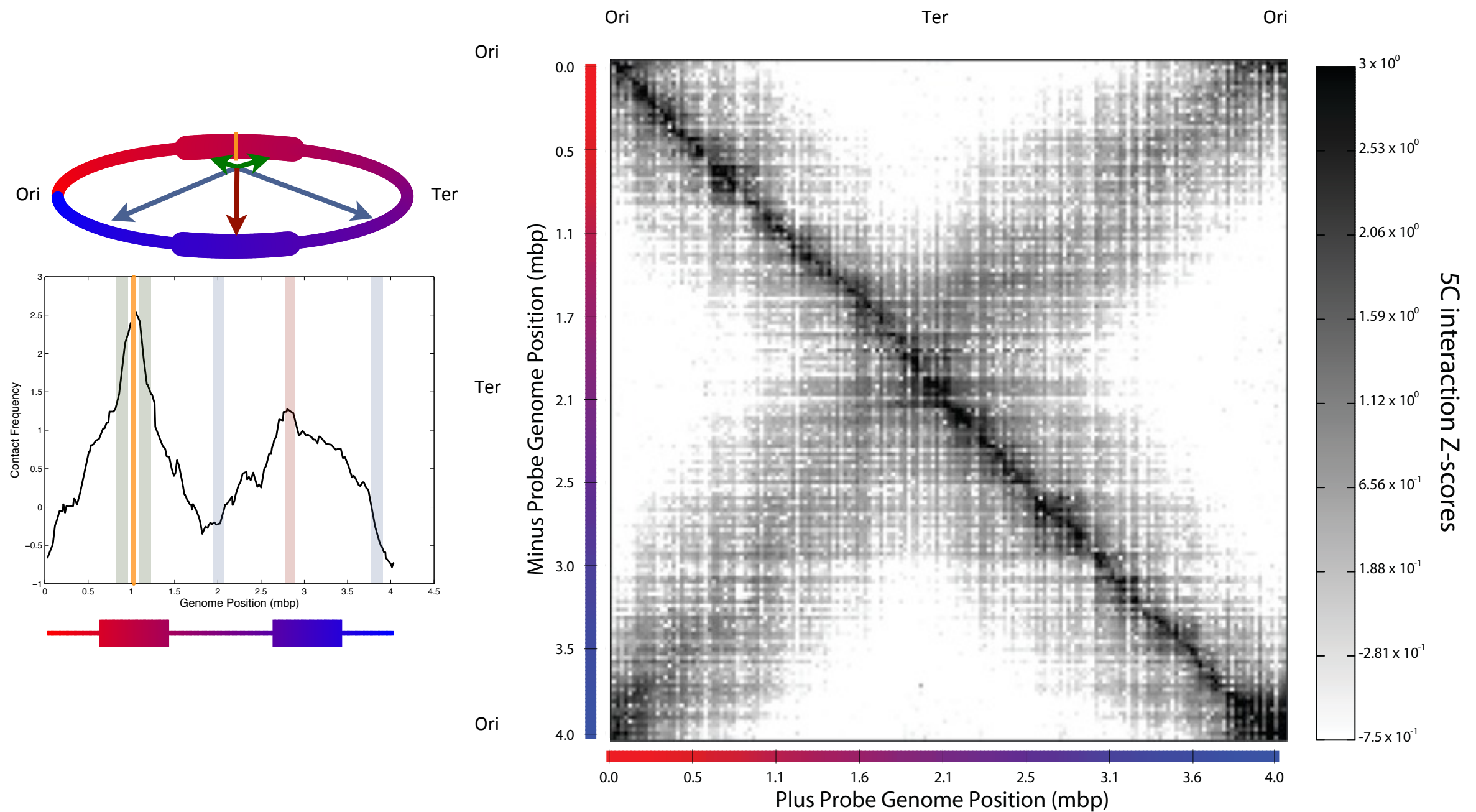
3D model building with the 5C + IMP approach
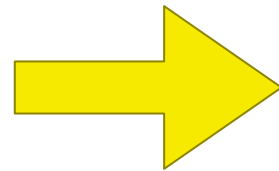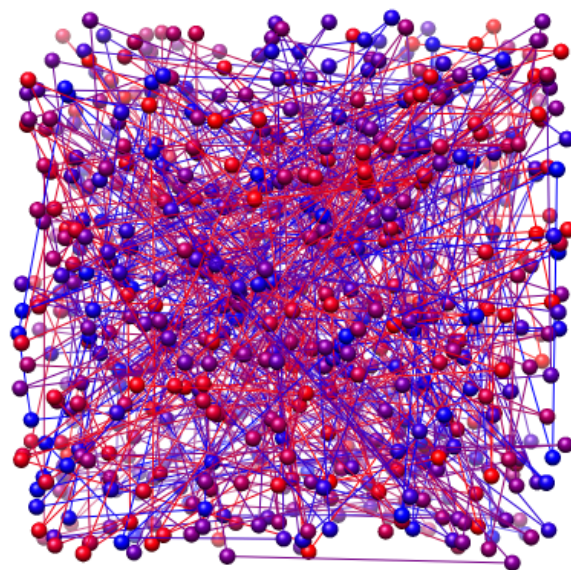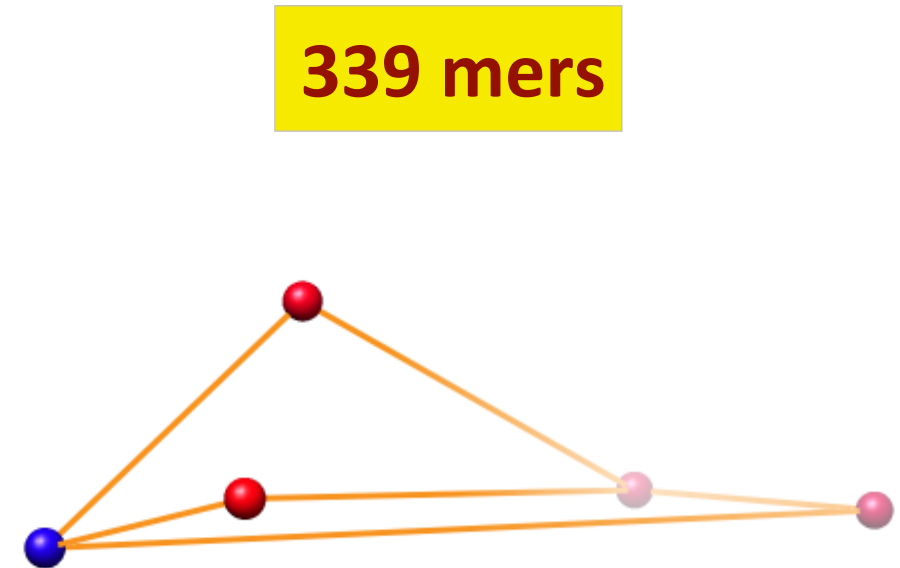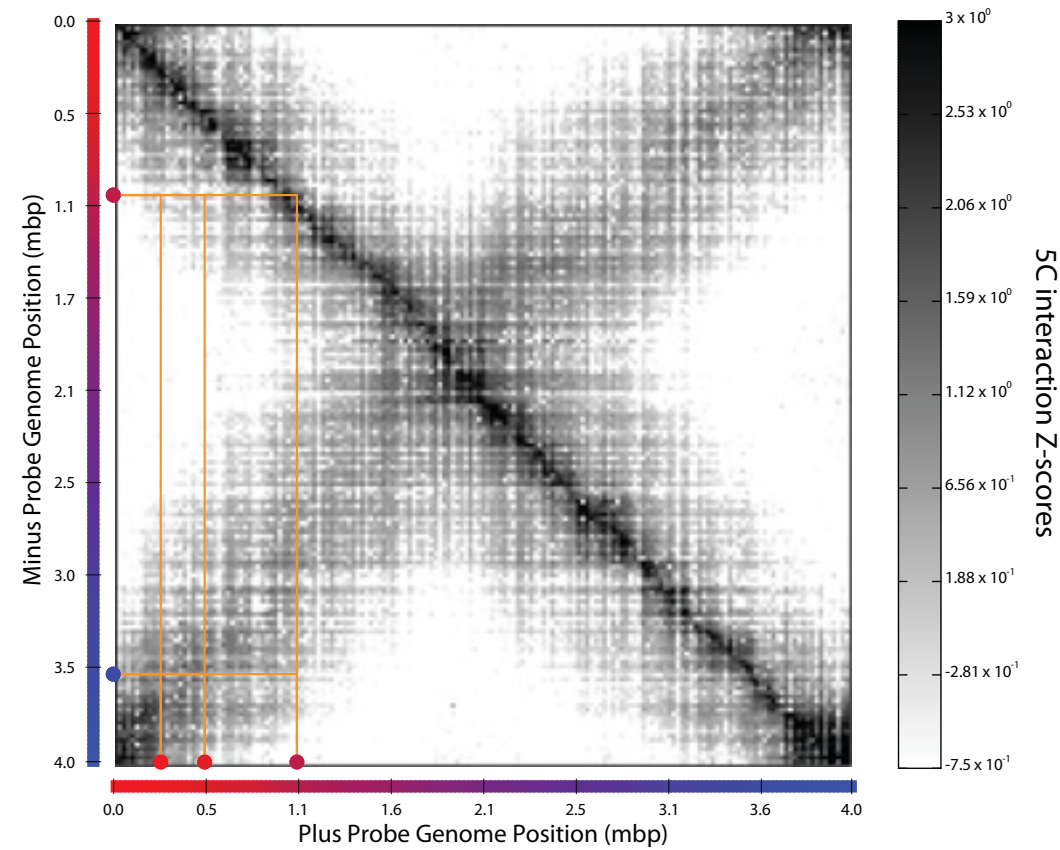
Genome organization in *Caulobacter crescentus*

Arms are helical

Resolution of chromosomes

Centromer-like

dif site 47±17Kb from Ter

parS sites 25±17Kb from Ori

Cluster 1   Cluster 2   Cluster 3   Cluster 4

180°   180°   180°   180°

500 nm   500 nm   500 nm   500 nm

MIRRORS!

22
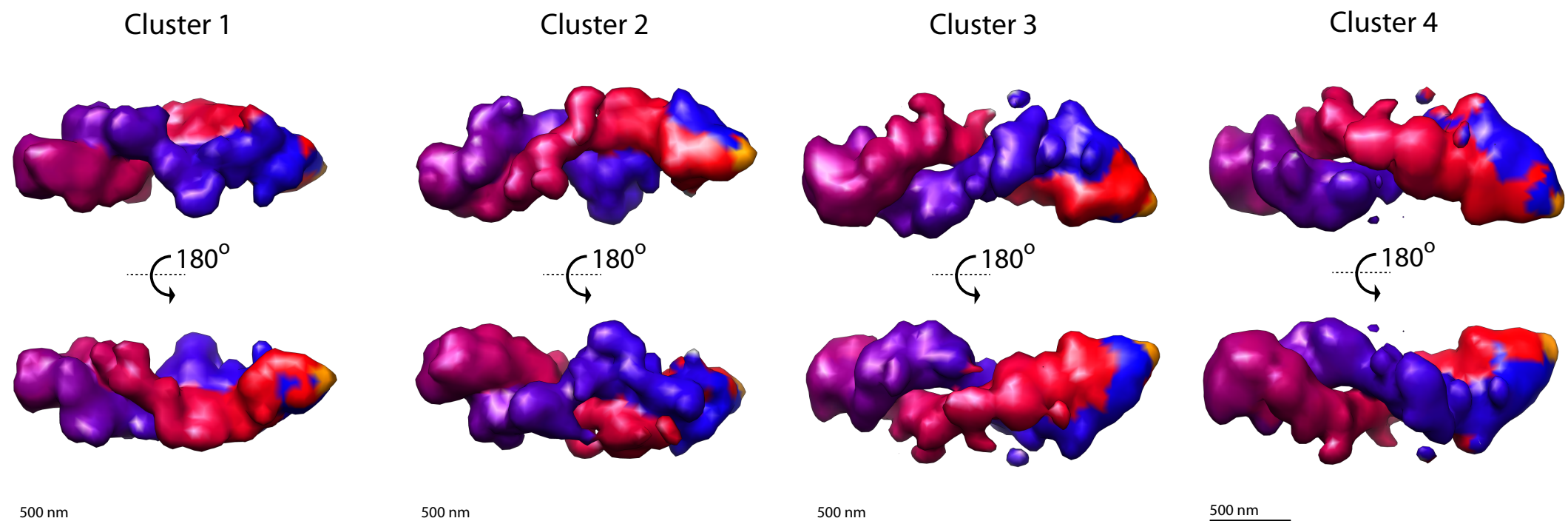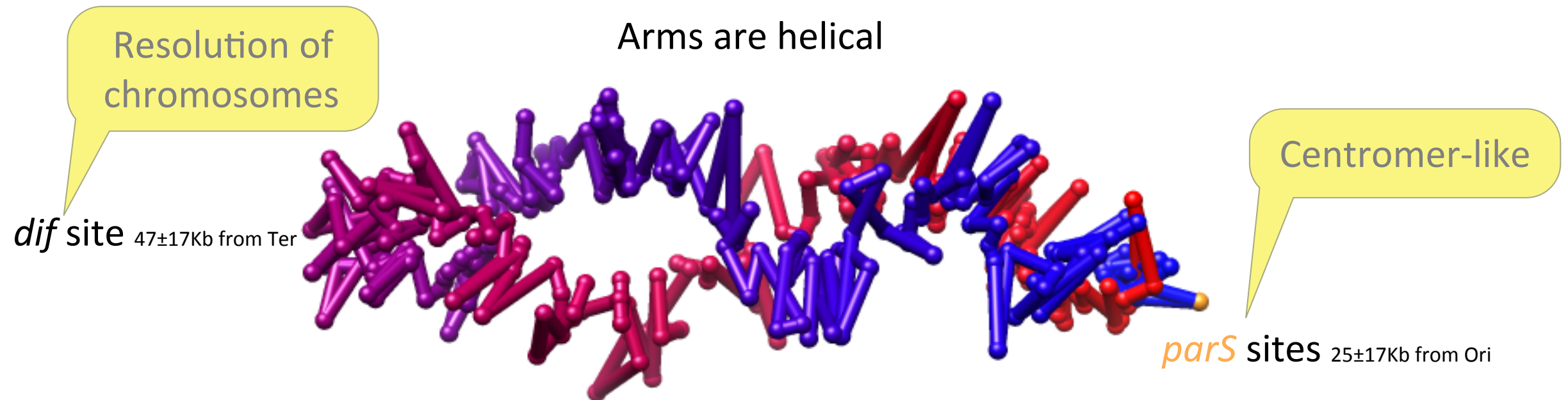
Moving the *parS* sites 400 Kb away from Ori
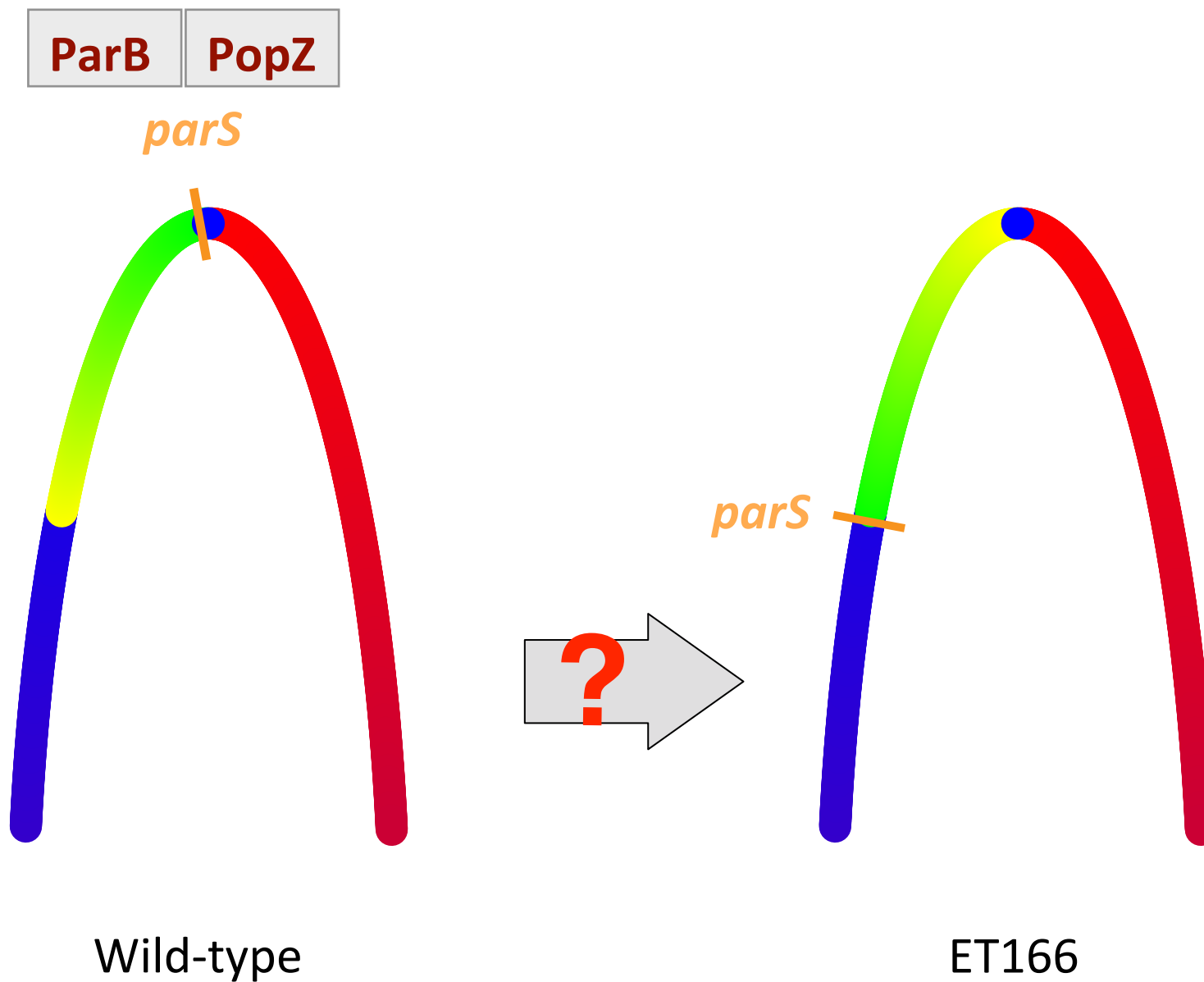
ParB | PopZ

parS

parS

?

Wild-type

ET166

23

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

Wild-type

ParS sites

ET166

500 nm

Arms are **STILL** helical

24
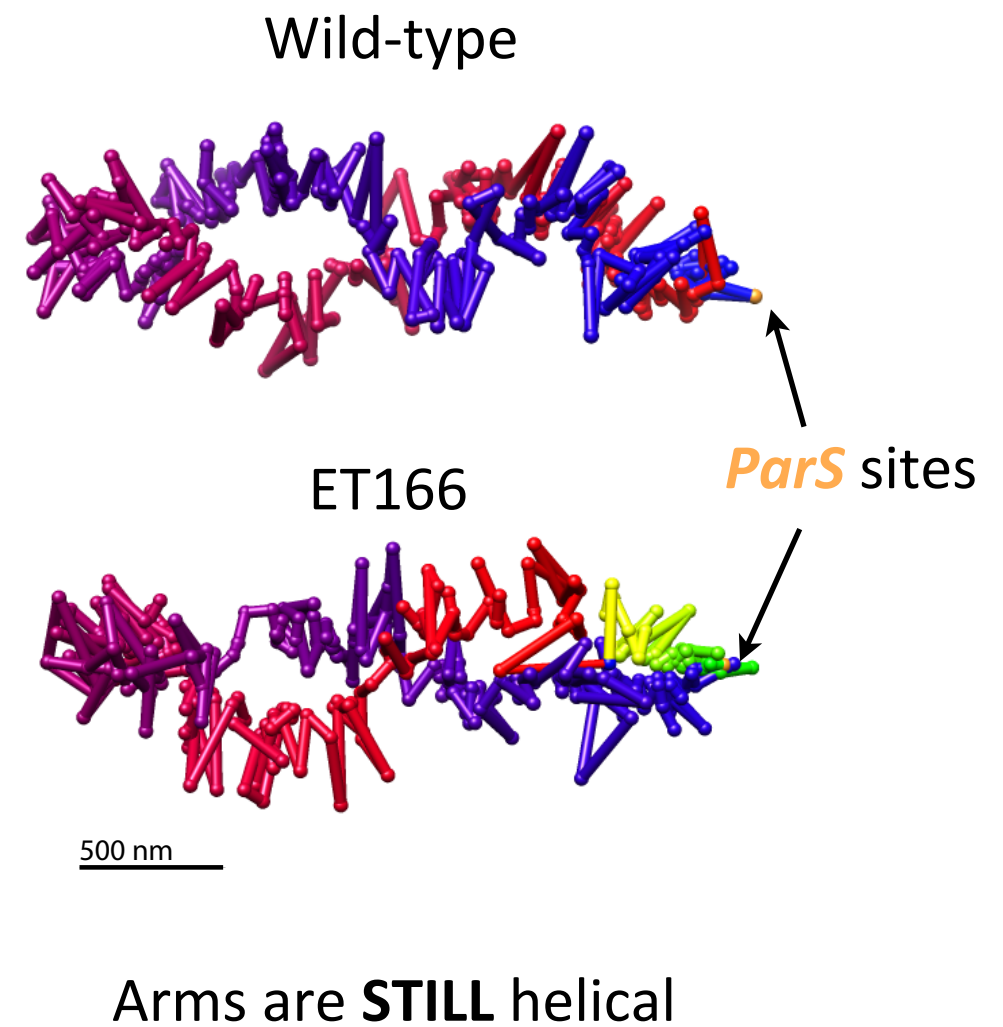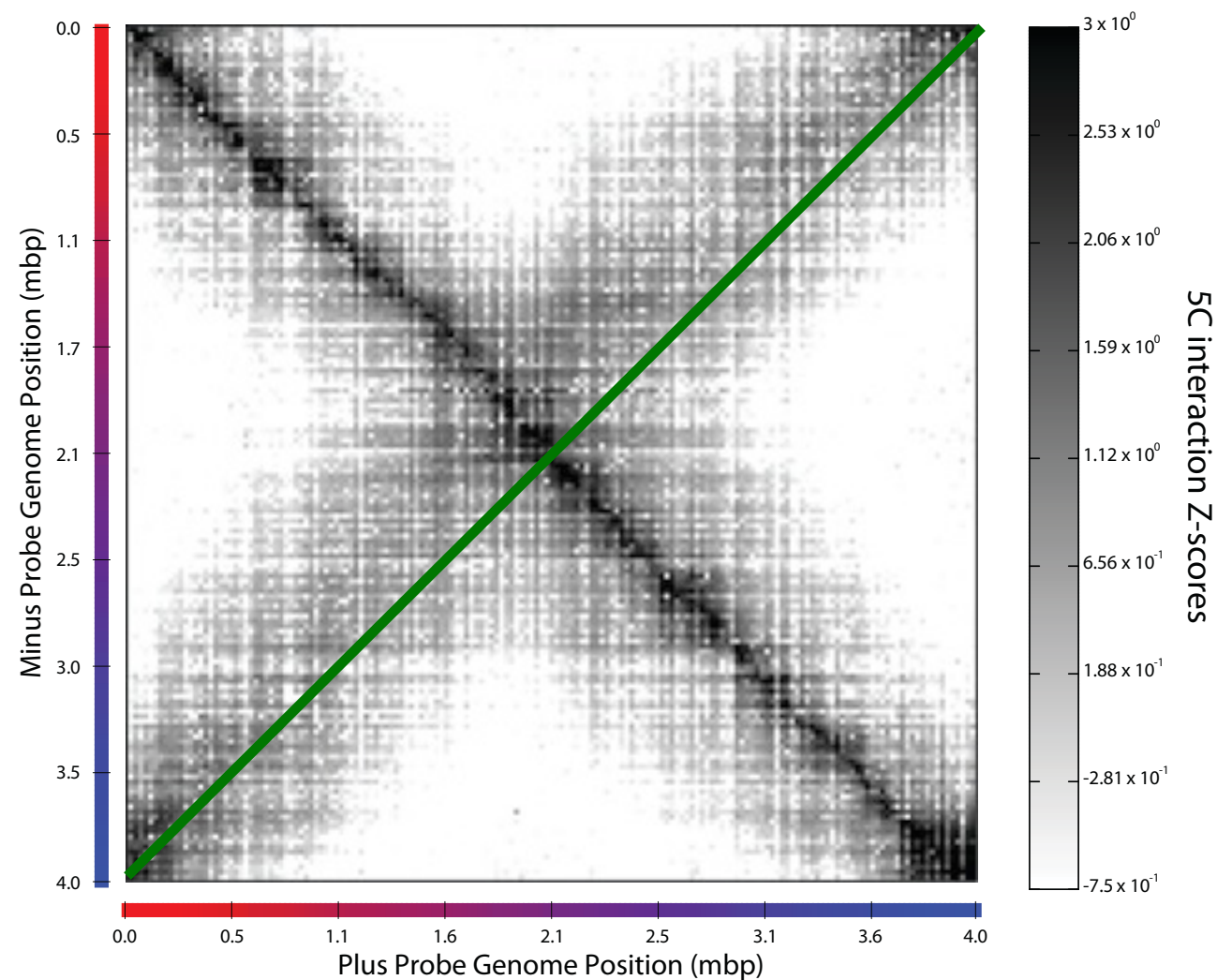
Moving the *parS* sites results in whole genome rotation!
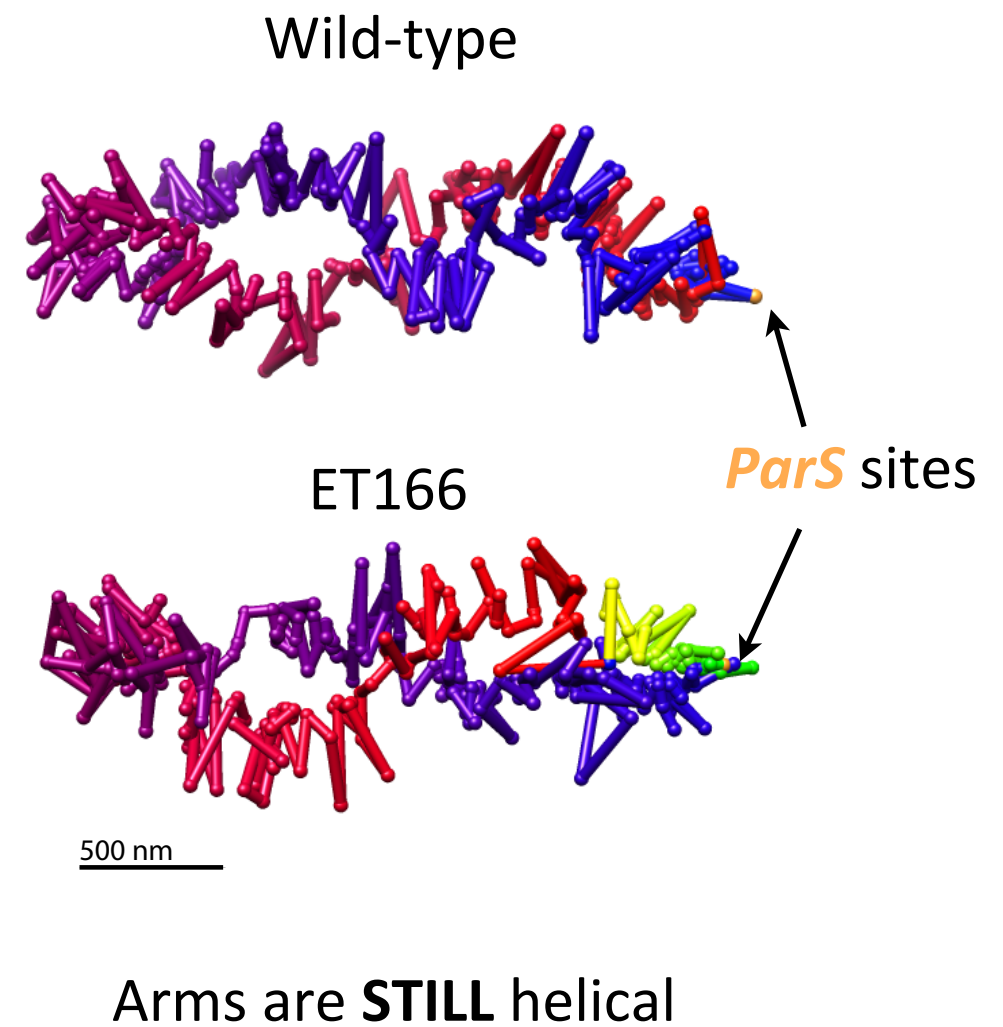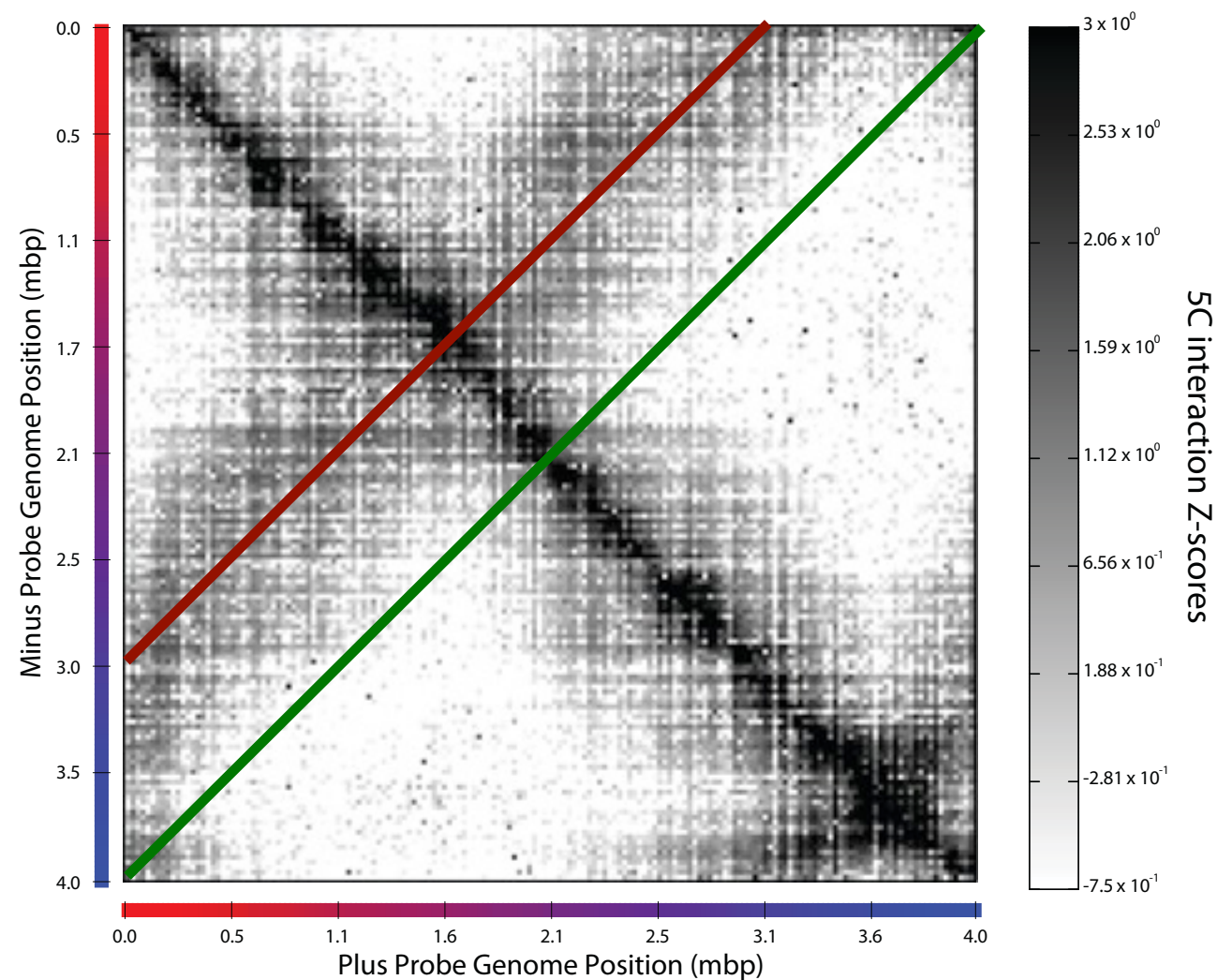
parS sites initiate compact chromatin domain
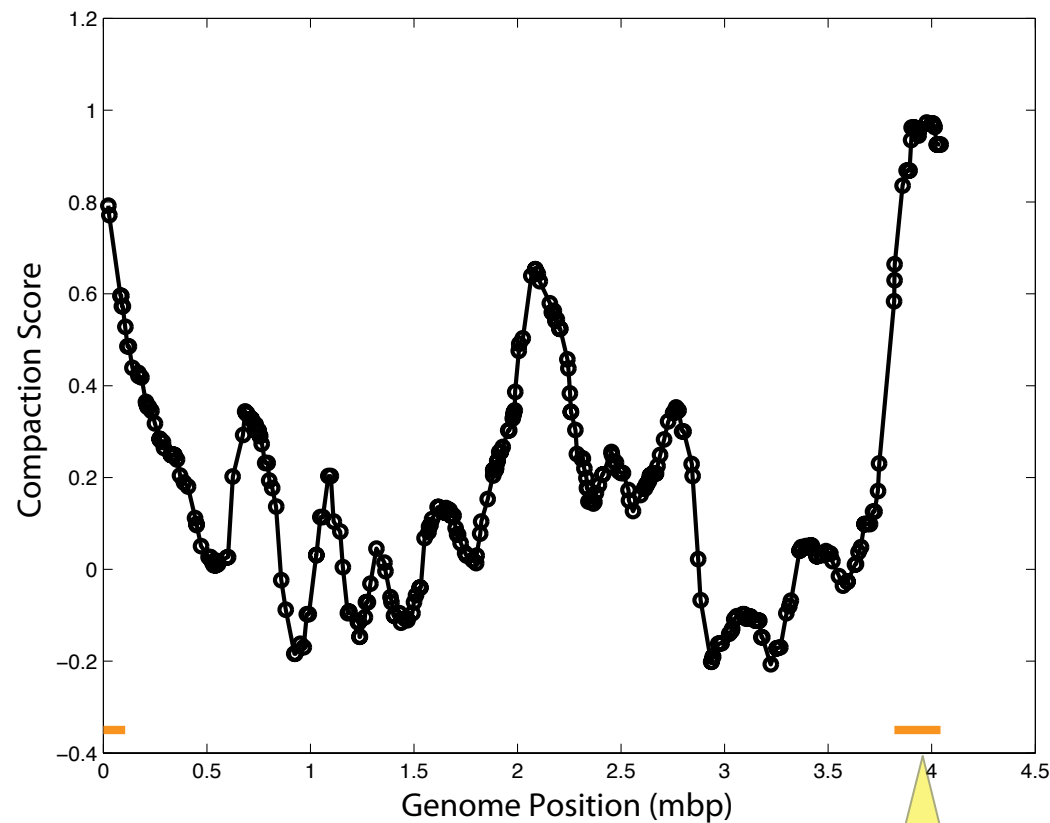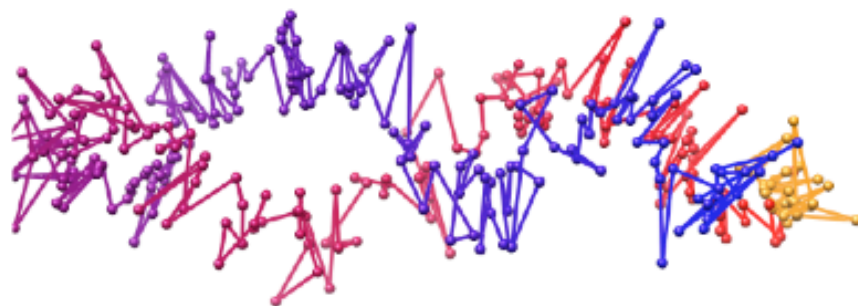
Chromosome arms are equidistant to the cell center

100-200Kb
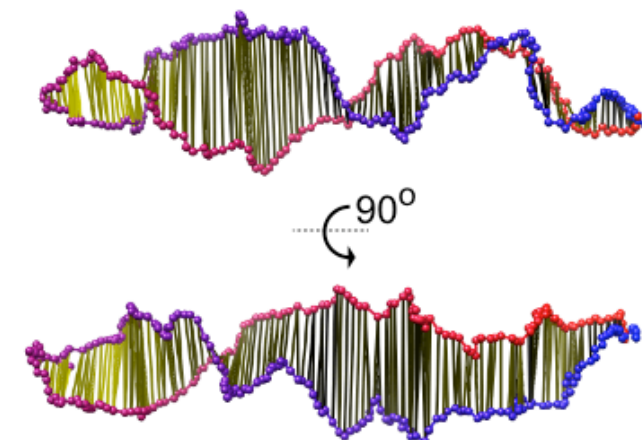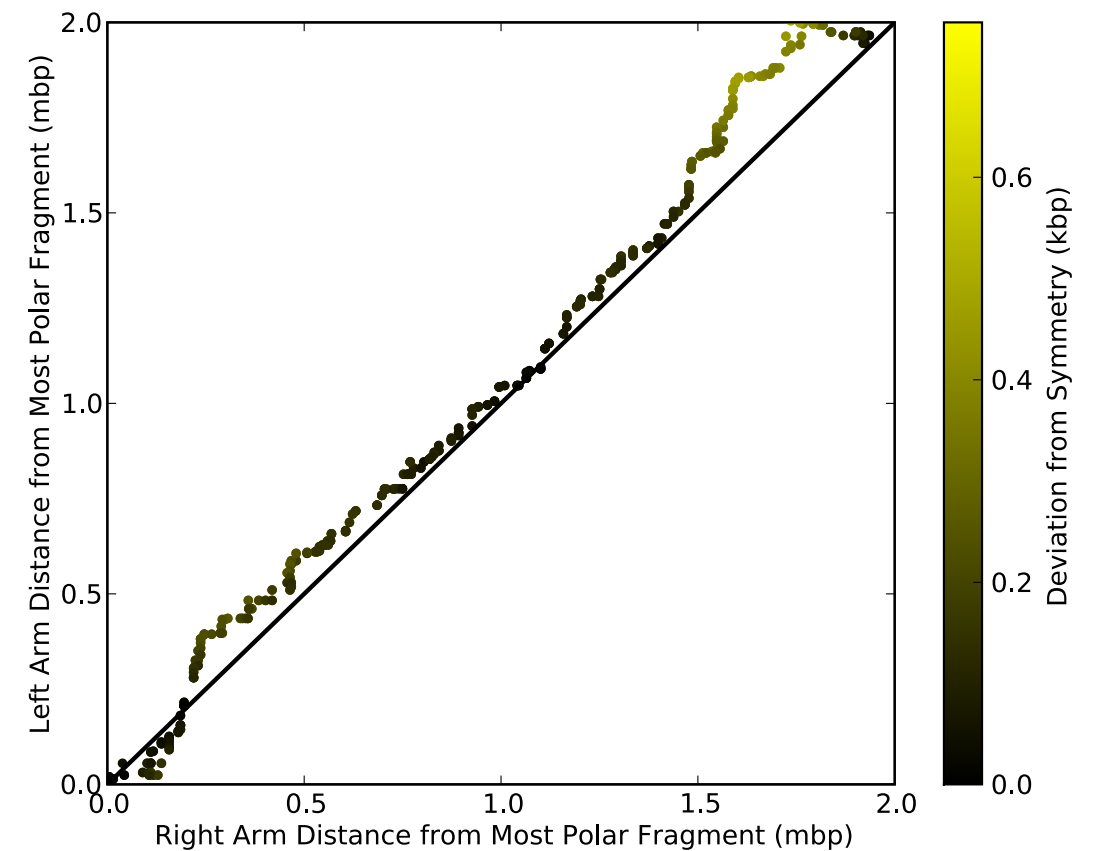
Genome architecture in *Caulobacter*

dense

dense
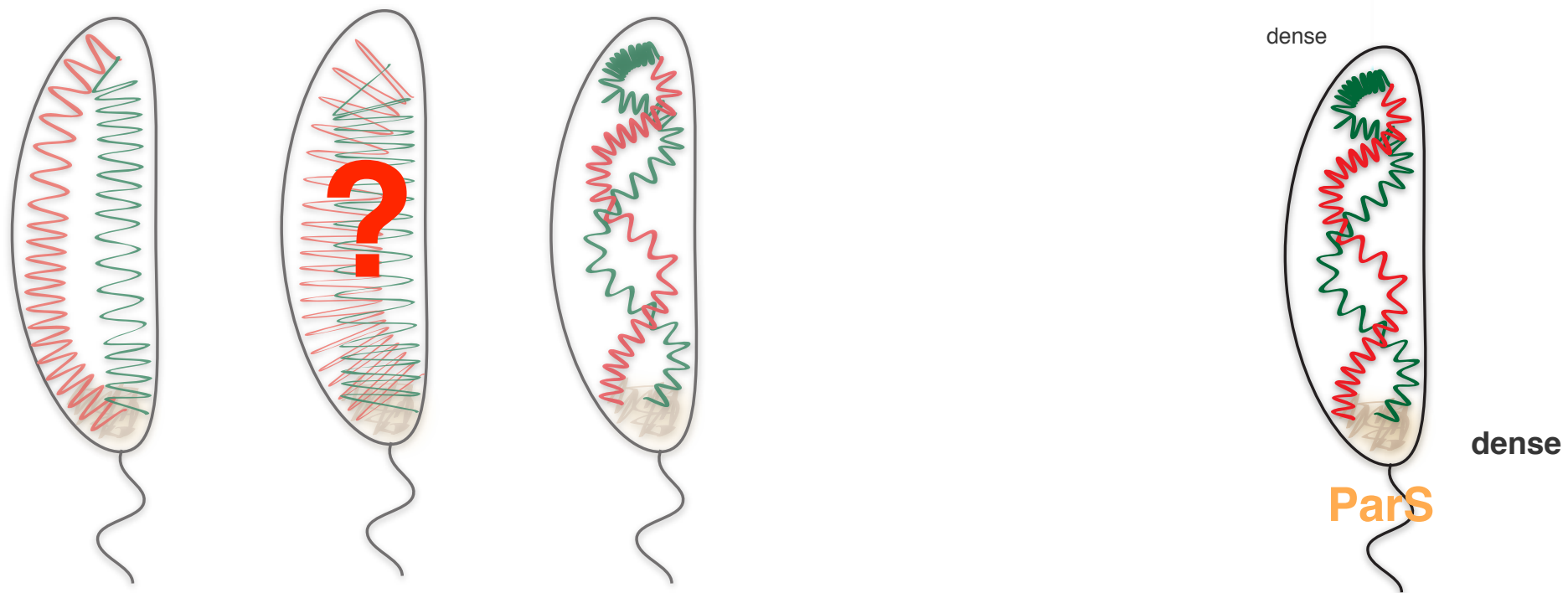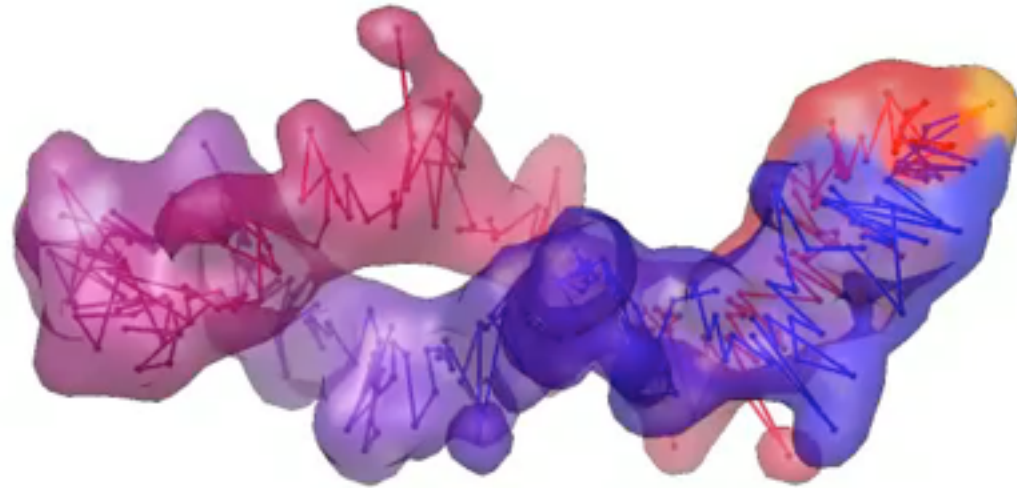
ParS

?

From Sequence to Function
5C + IMP

Technology

Hypothesis

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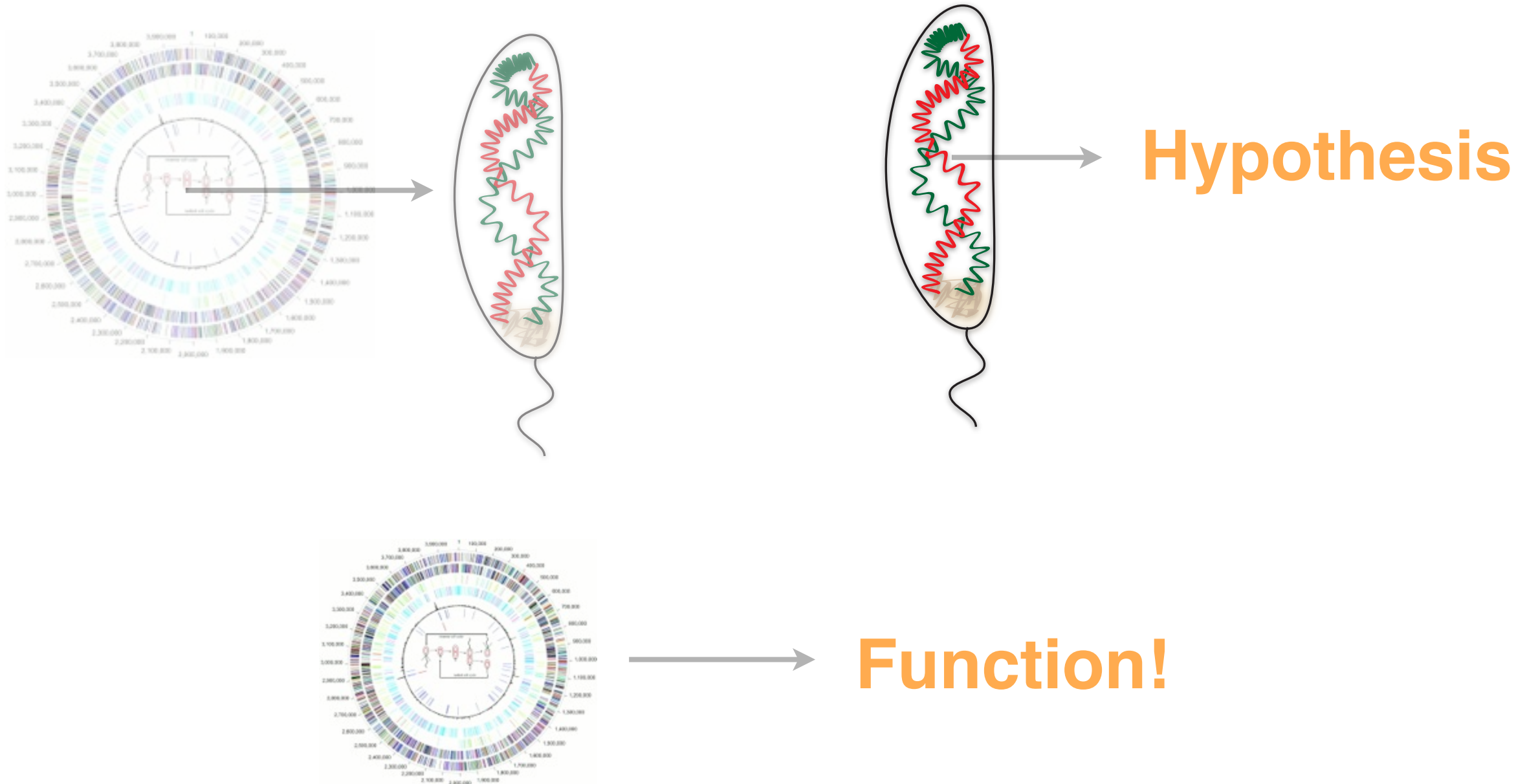
D. Baù and M.A. Marti-Renom **Chromosome Res** (2011) 19:25-35.

Thursday, October 13, 11

PLoS CB Outlook

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

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

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

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

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

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

28