

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

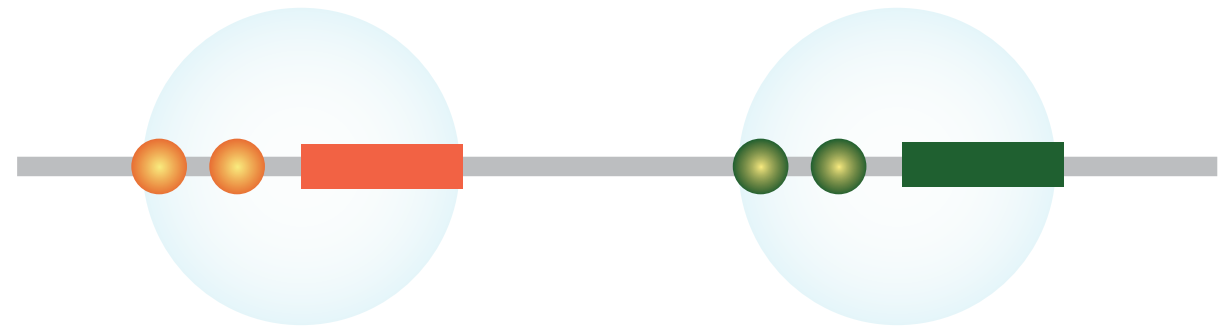
Marc A. Marti-Renom

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

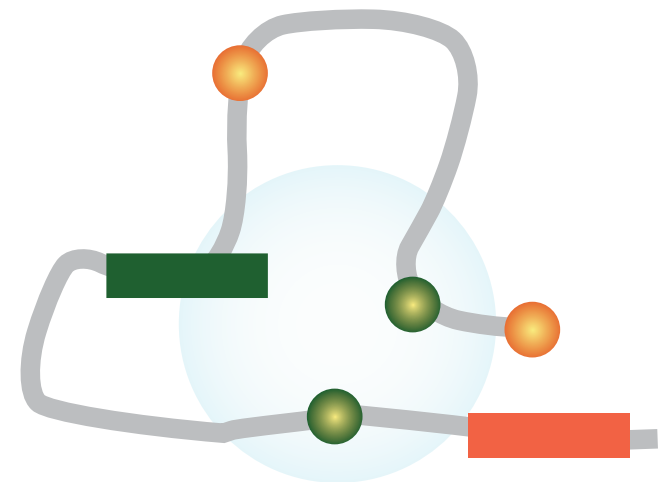
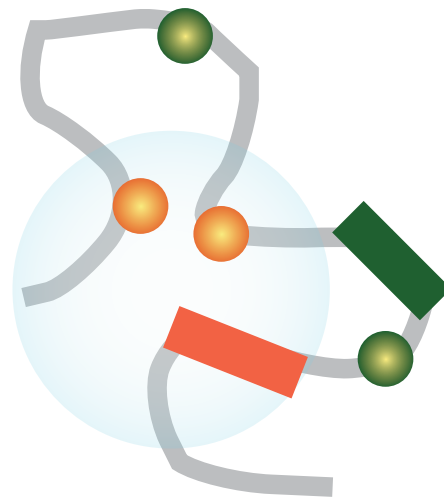
cnag



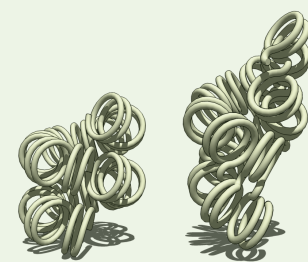
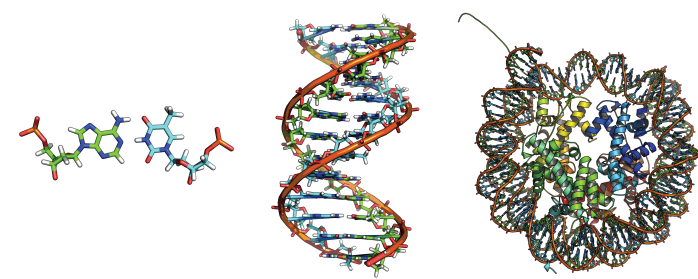
Simple genomes



Complex genomes

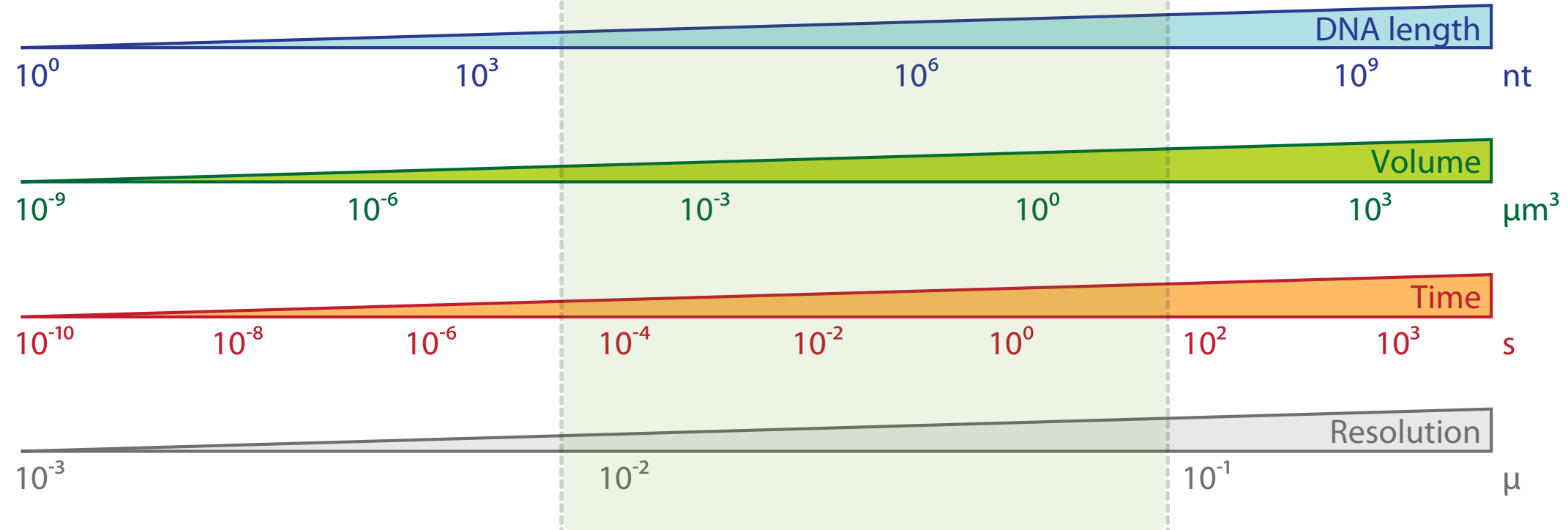
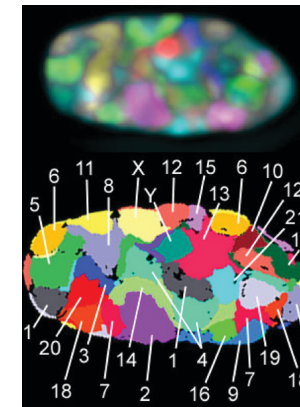
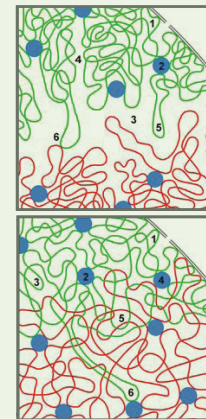


Knowledge



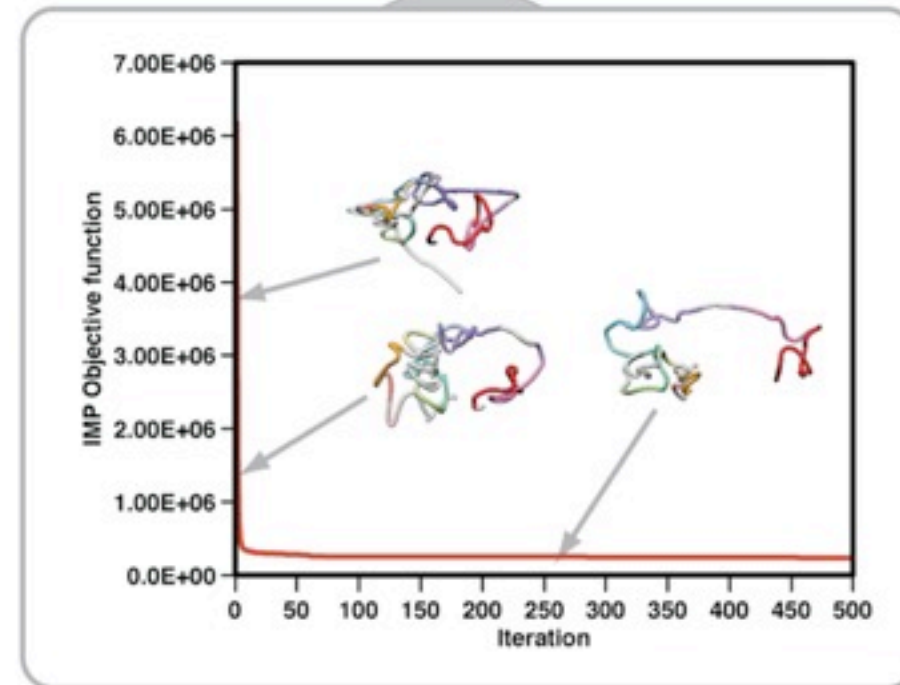
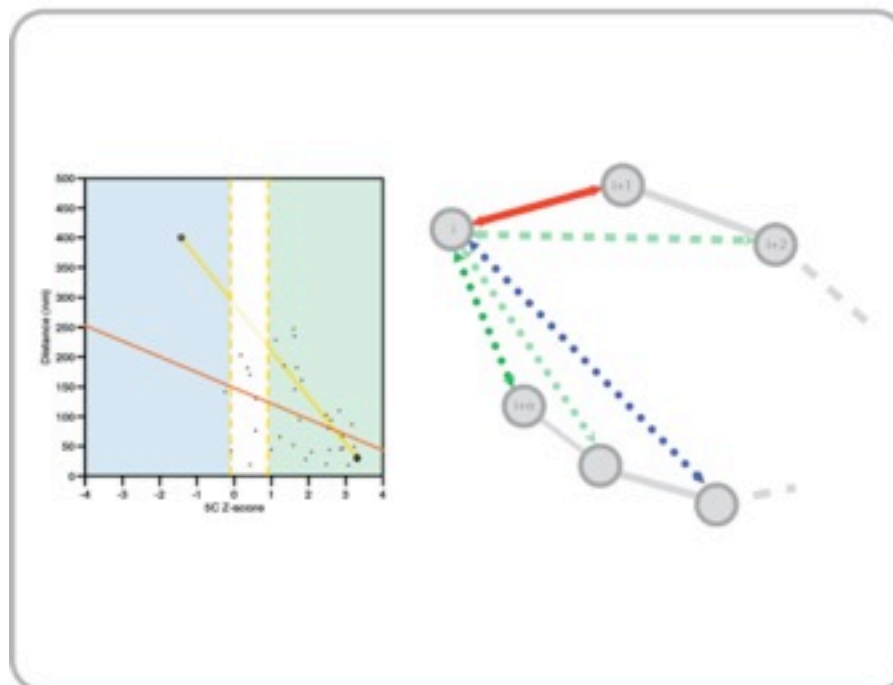
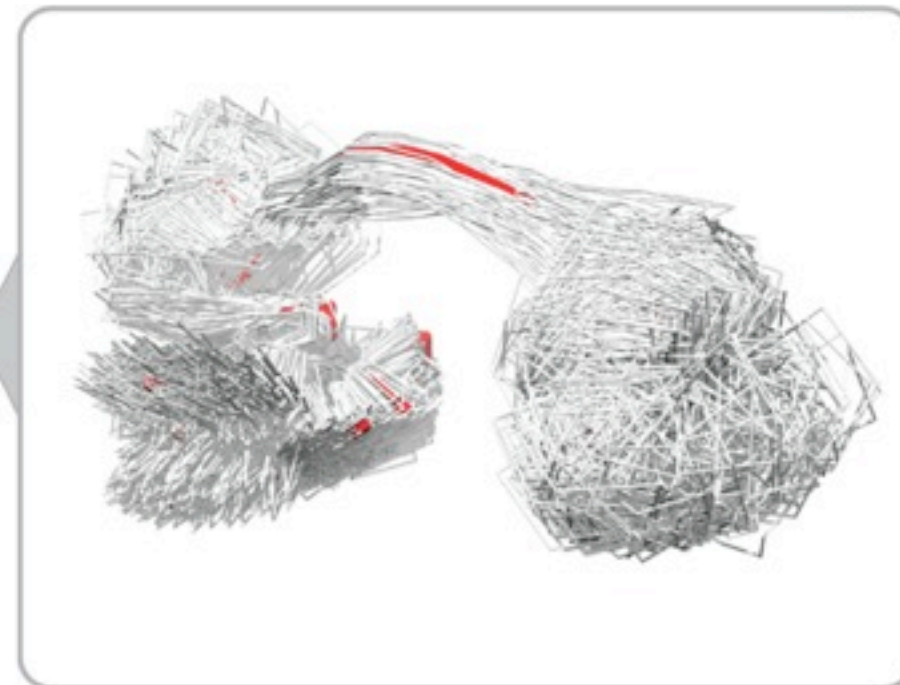
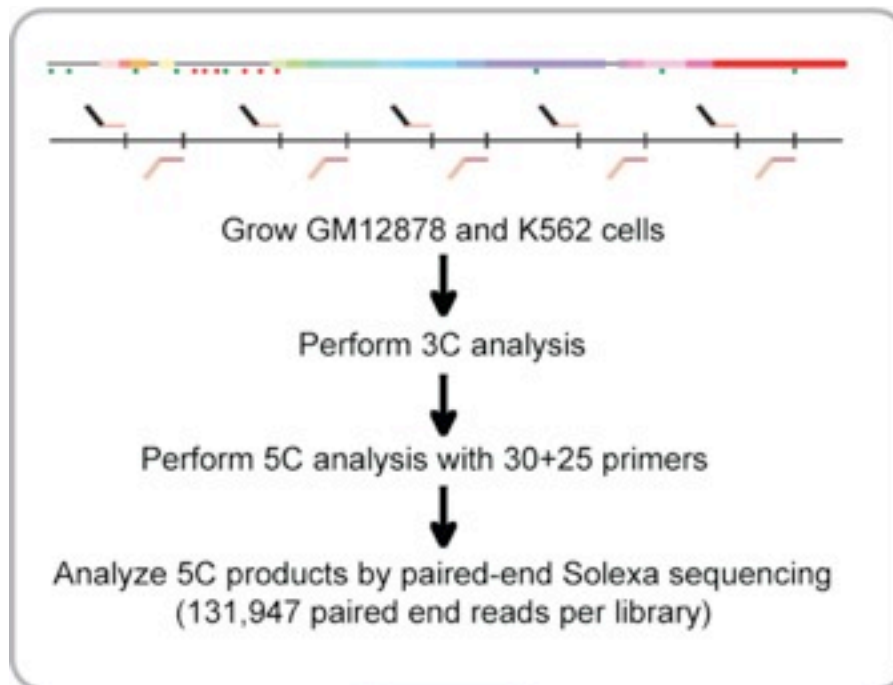
IDM

INM

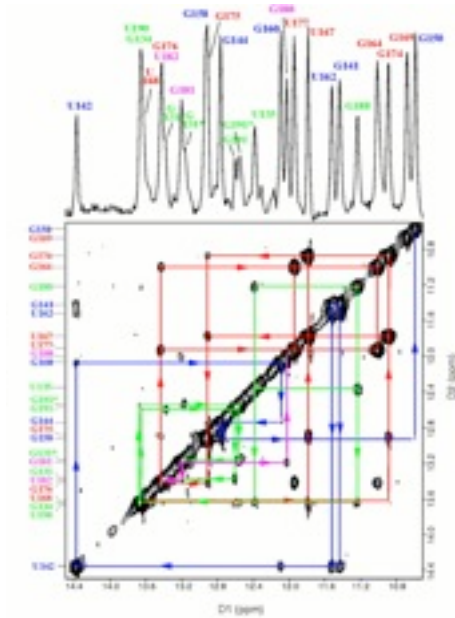
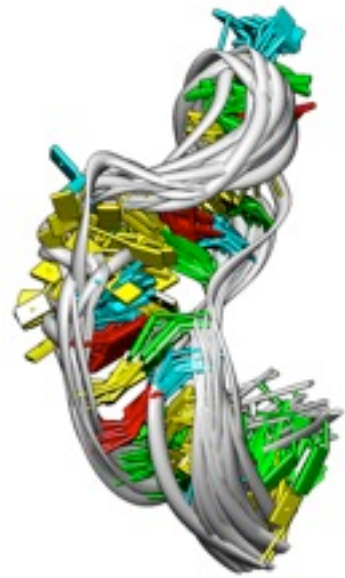


Adapted from:
Langowski and Heermann. *Semin Cell Dev Biol* (2007) vol. 18 (5) pp. 659-67

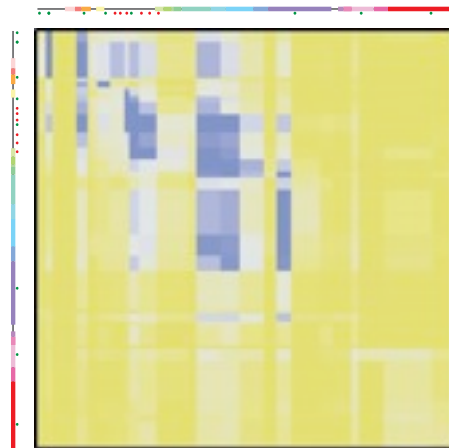
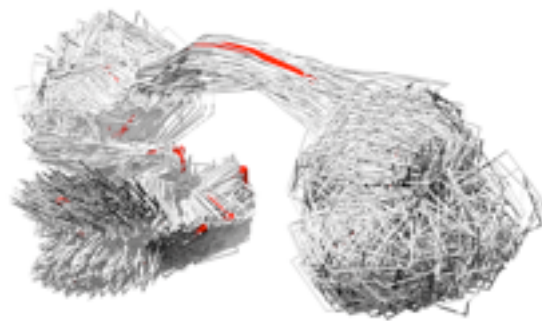
Experiments



Computation



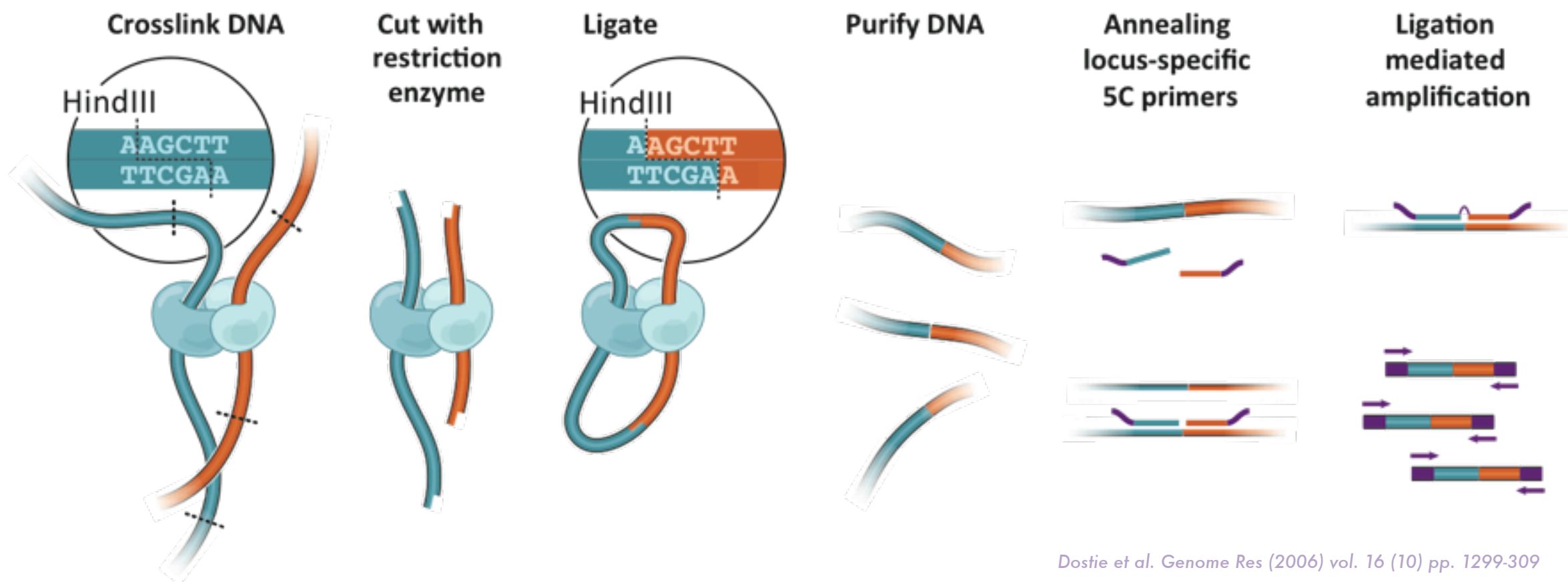
Biomolecular structure determination 2D-NOESY data



Chromosome structure determination 5C data

5C technology

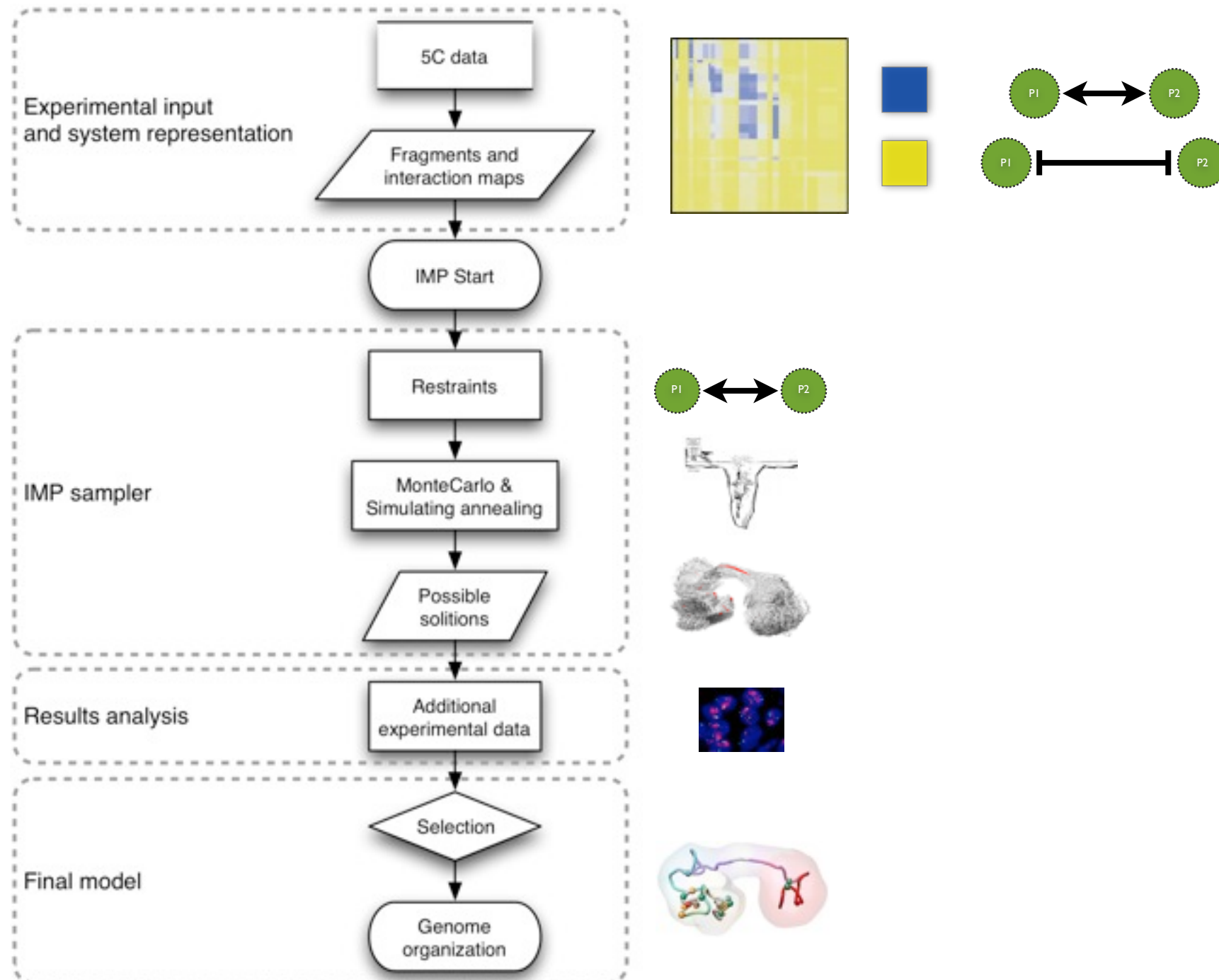
<http://my5C.umassmed.edu>



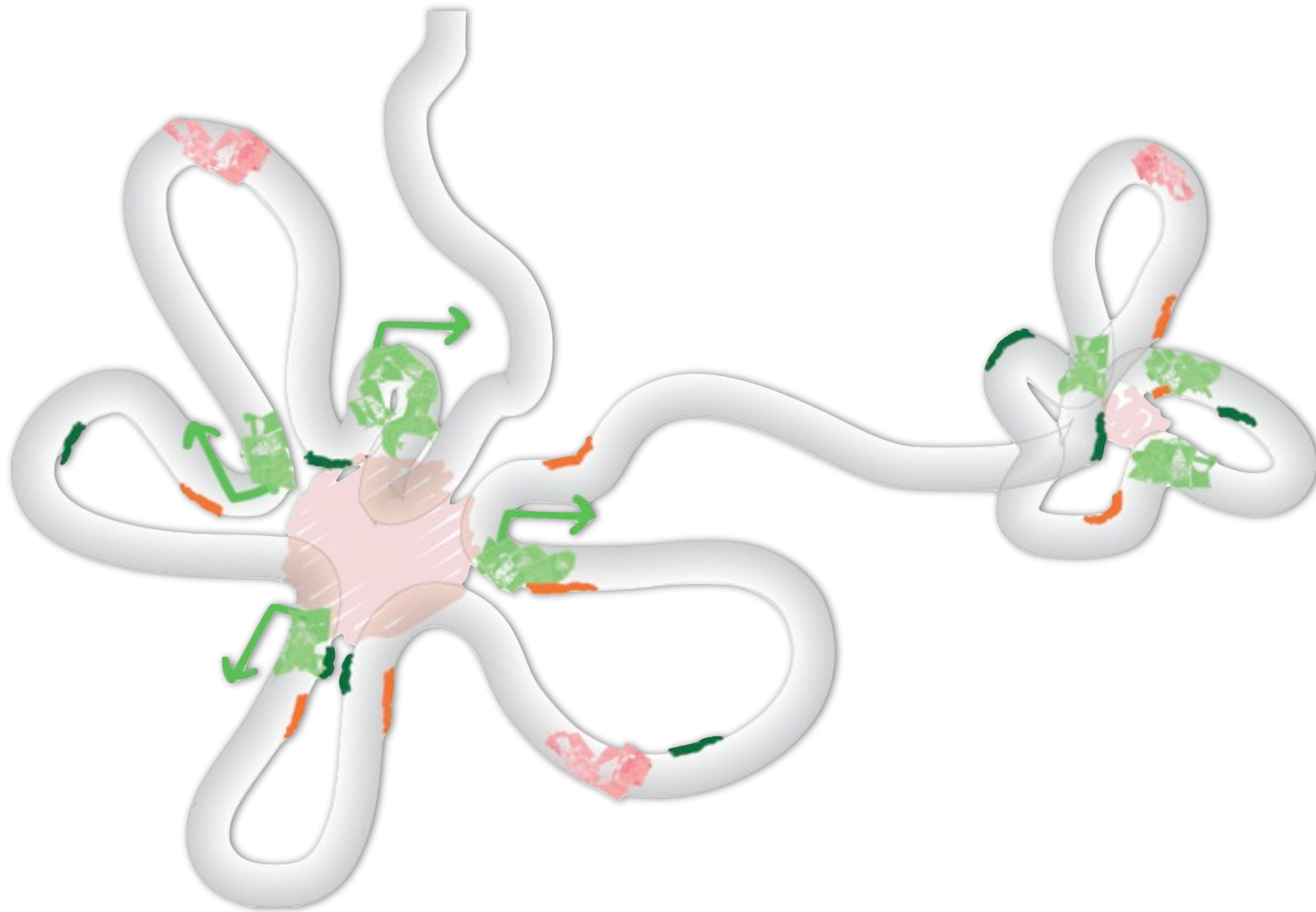
Dostie et al. *Genome Res* (2006) vol. 16 (10) pp. 1299-309

Integrative Modeling

<http://www.integrativemodeling.org>



Human α -globin domain



Human α -globin domain

ENm008 genomic structure and environment

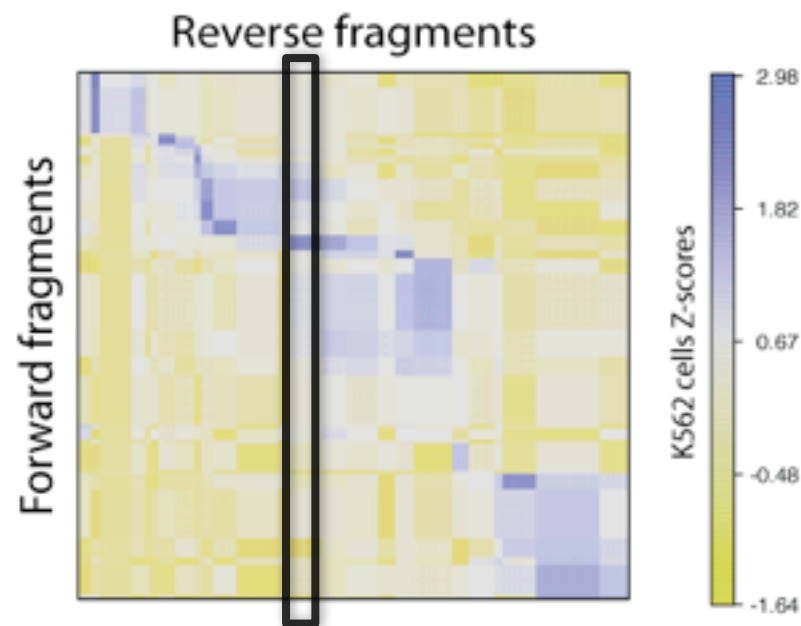
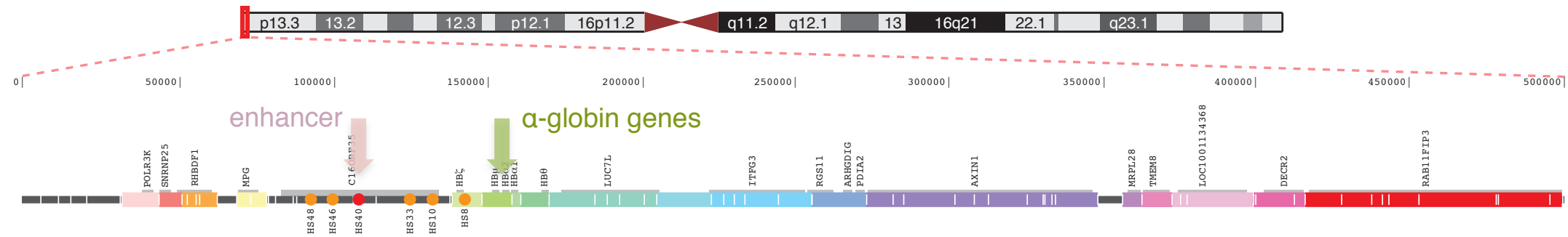


The ENCODE data for ENm008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute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

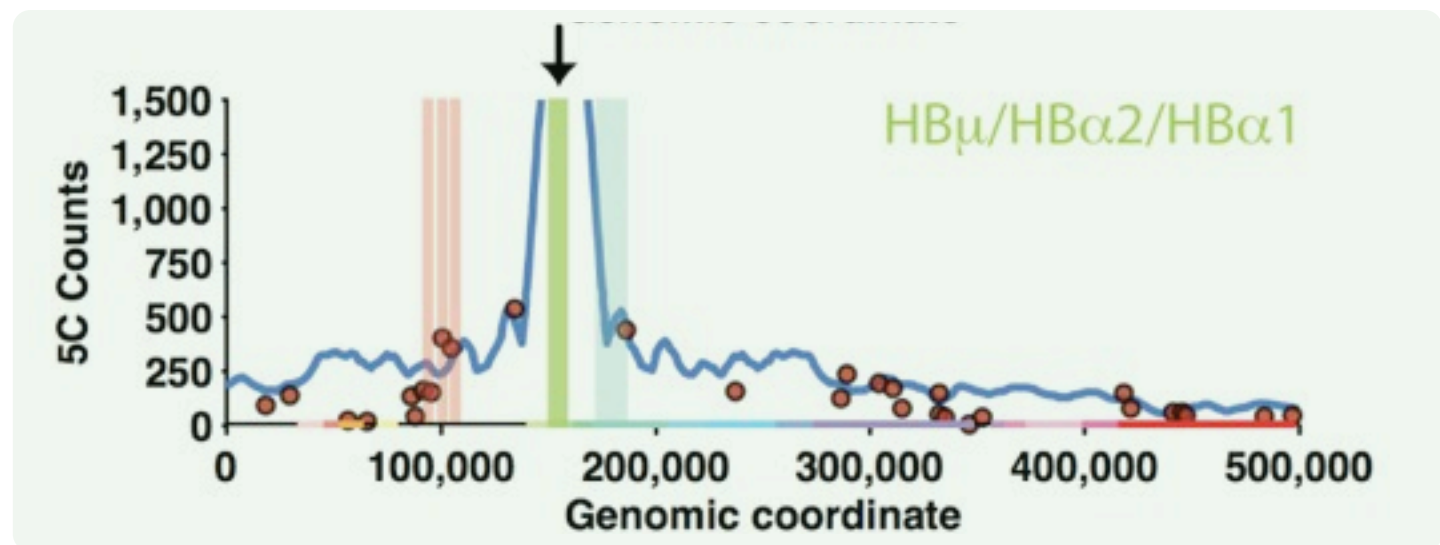
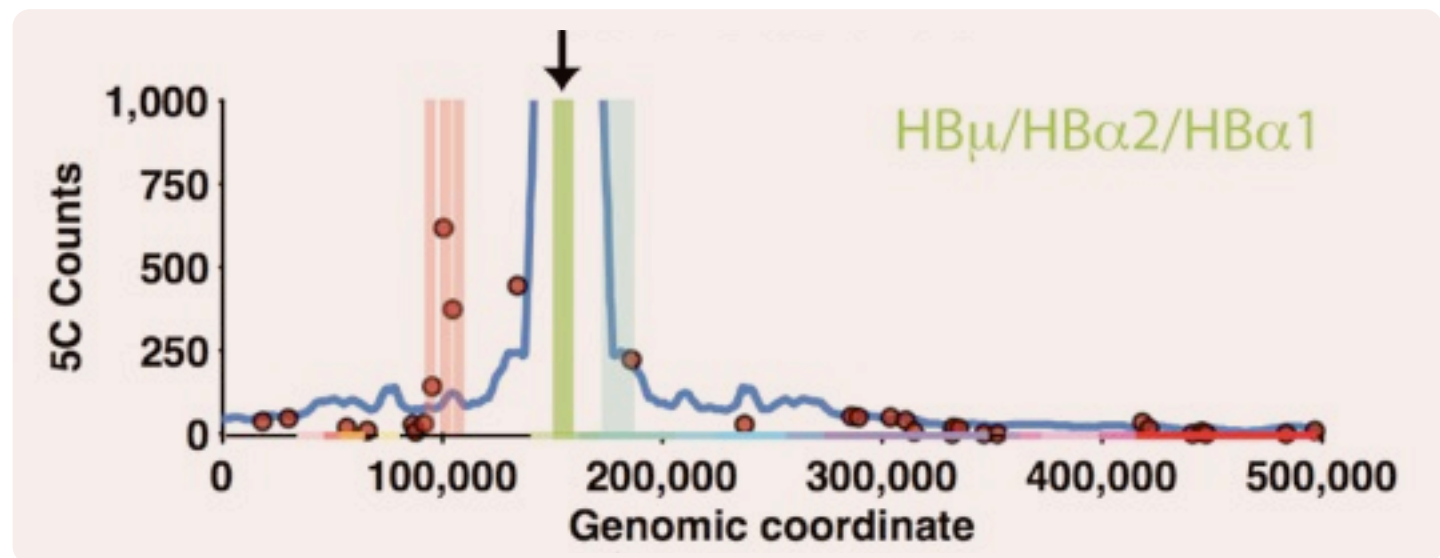
ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816

Human α -globin domain

ENm008 genomic structure and environment



K562 cells:
 α -globin genes active



Representation

Harmonic

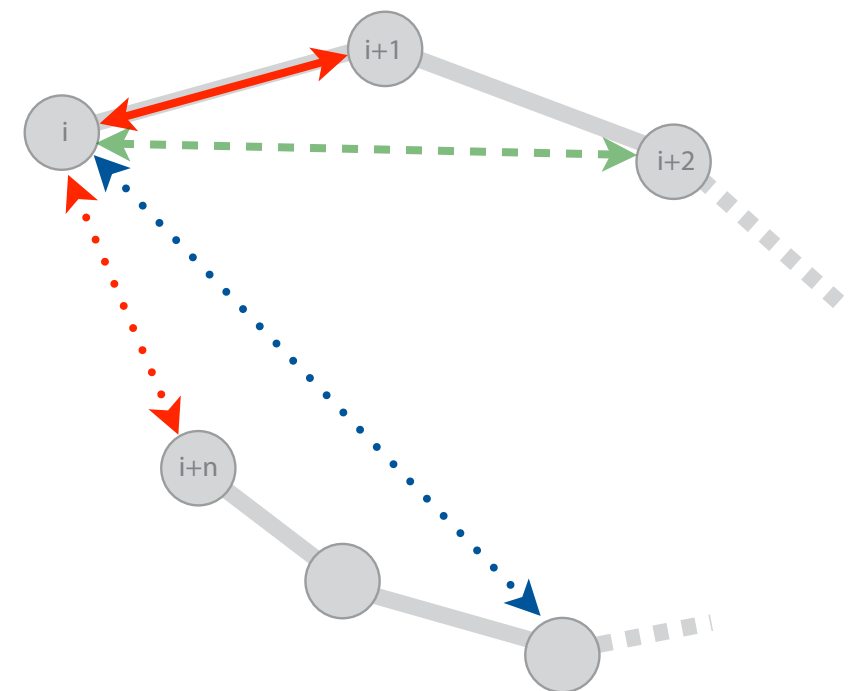
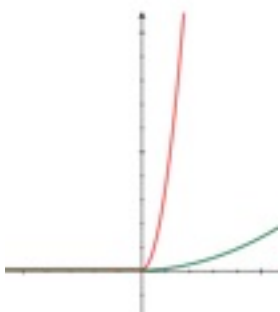
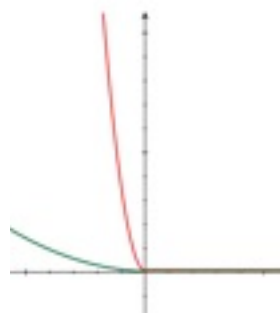
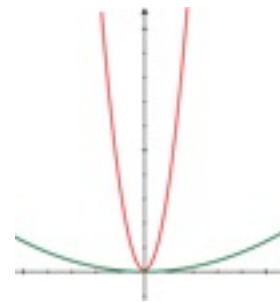
$$H_{i,j} = k(d_{i,j} - d_{i,j}^0)^2$$

Harmonic Lower Bound

$$\begin{cases} \text{if } d_{i,j} \leq d_{i,j}^0; & lbH_{i,j} = k(d_{i,j} - d_{i,j}^0)^2 \\ \text{if } d_{i,j} > d_{i,j}^0; & lbH_{i,j} = 0 \end{cases}$$

Harmonic Upper Bound

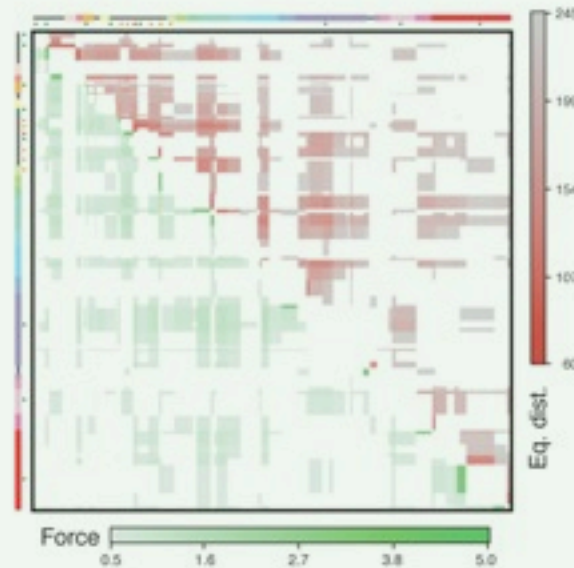
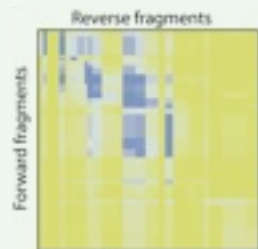
$$\begin{cases} \text{if } d_{i,j} \geq d_{i,j}^0; & ubH_{i,j} = k(d_{i,j} - d_{i,j}^0)^2 \\ \text{if } d_{i,j} < d_{i,j}^0; & ubH_{i,j} = 0 \end{cases}$$



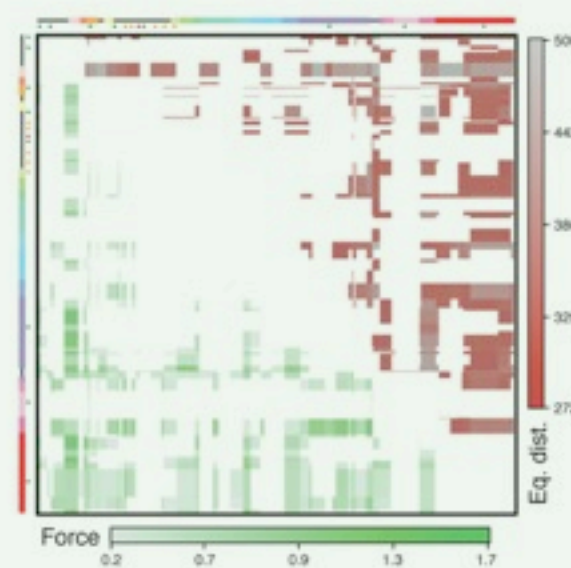
Scoring

GM12878

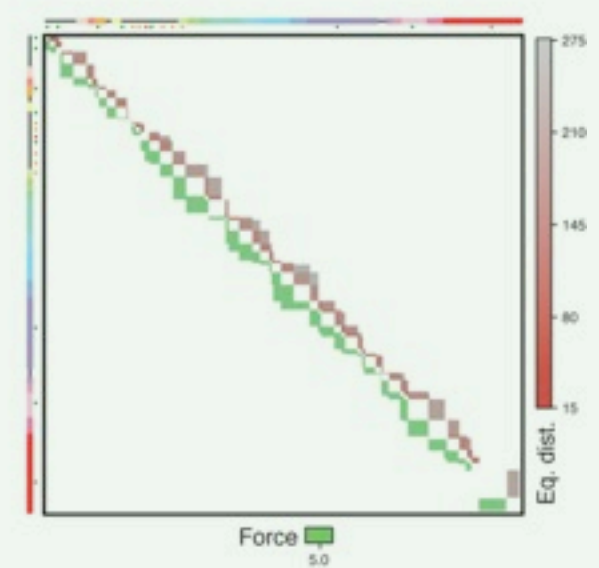
70 fragments
1,520 restraints



Harmonic



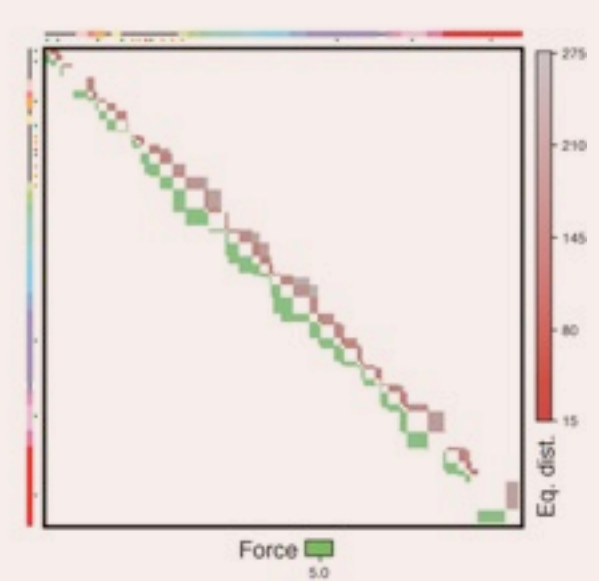
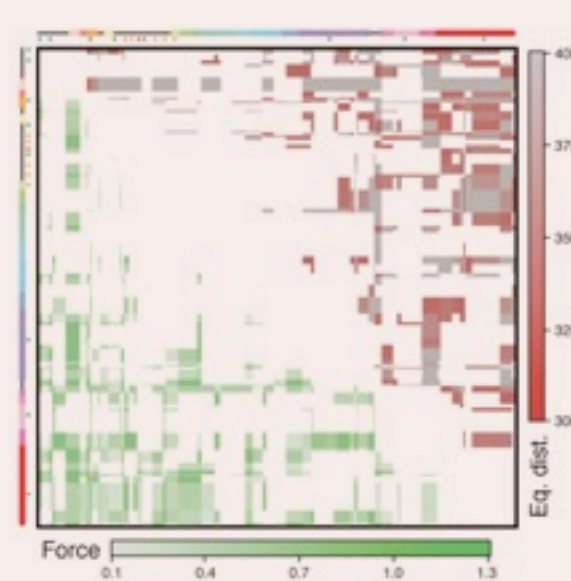
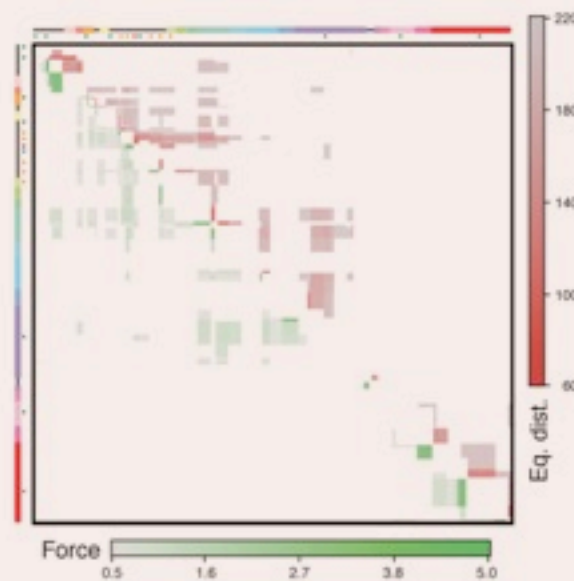
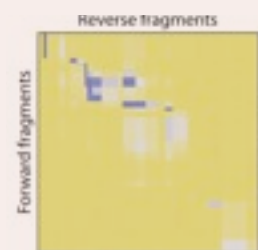
Harmonic Lower Bound



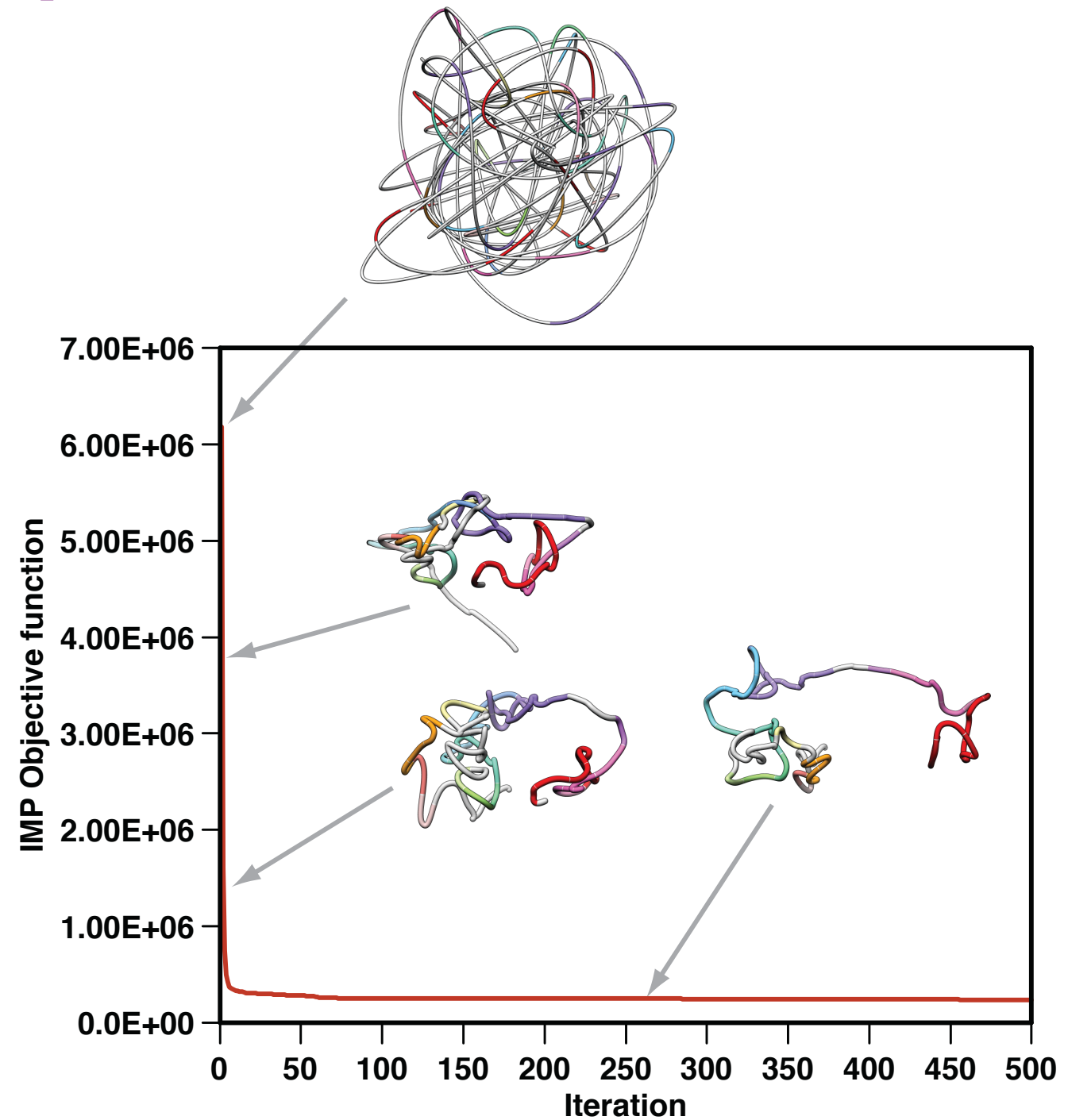
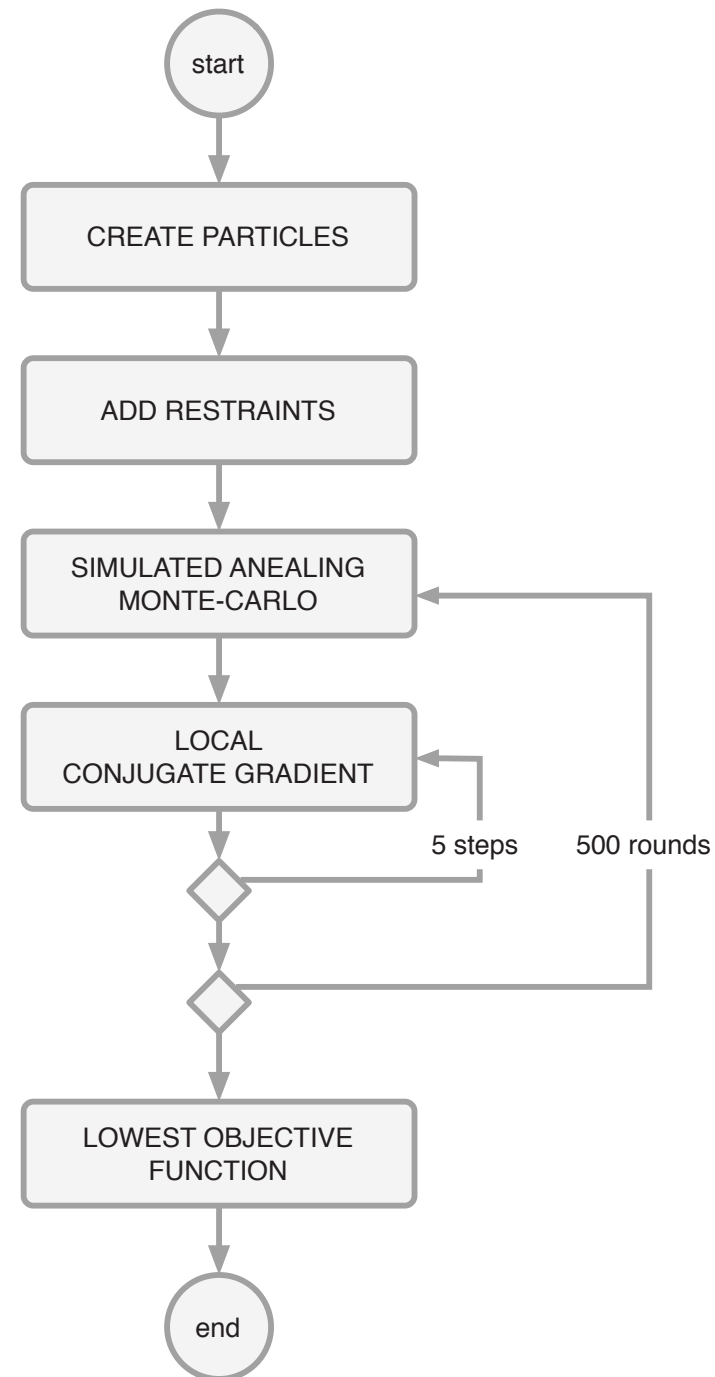
Harmonic Upper Bound

K562

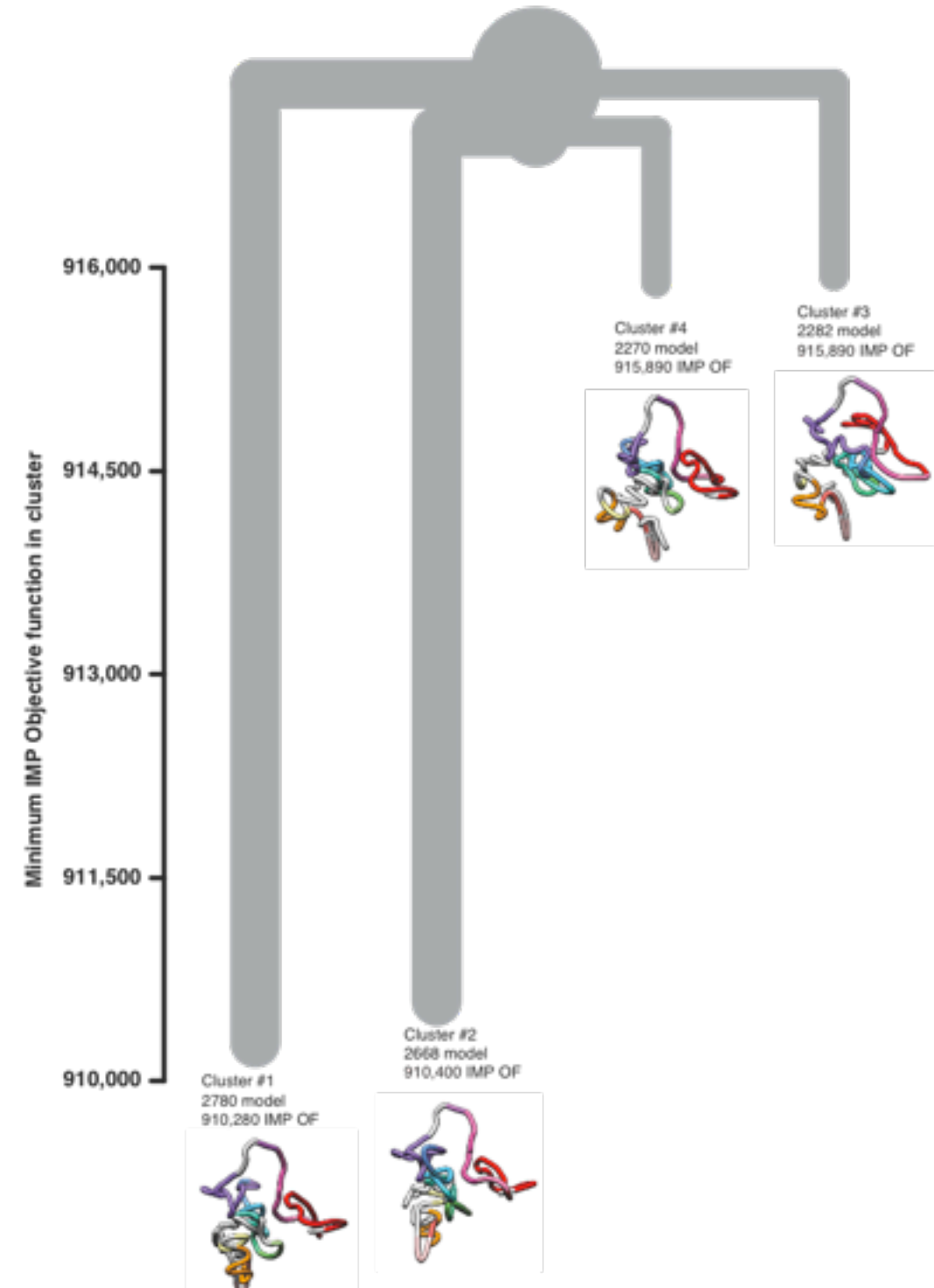
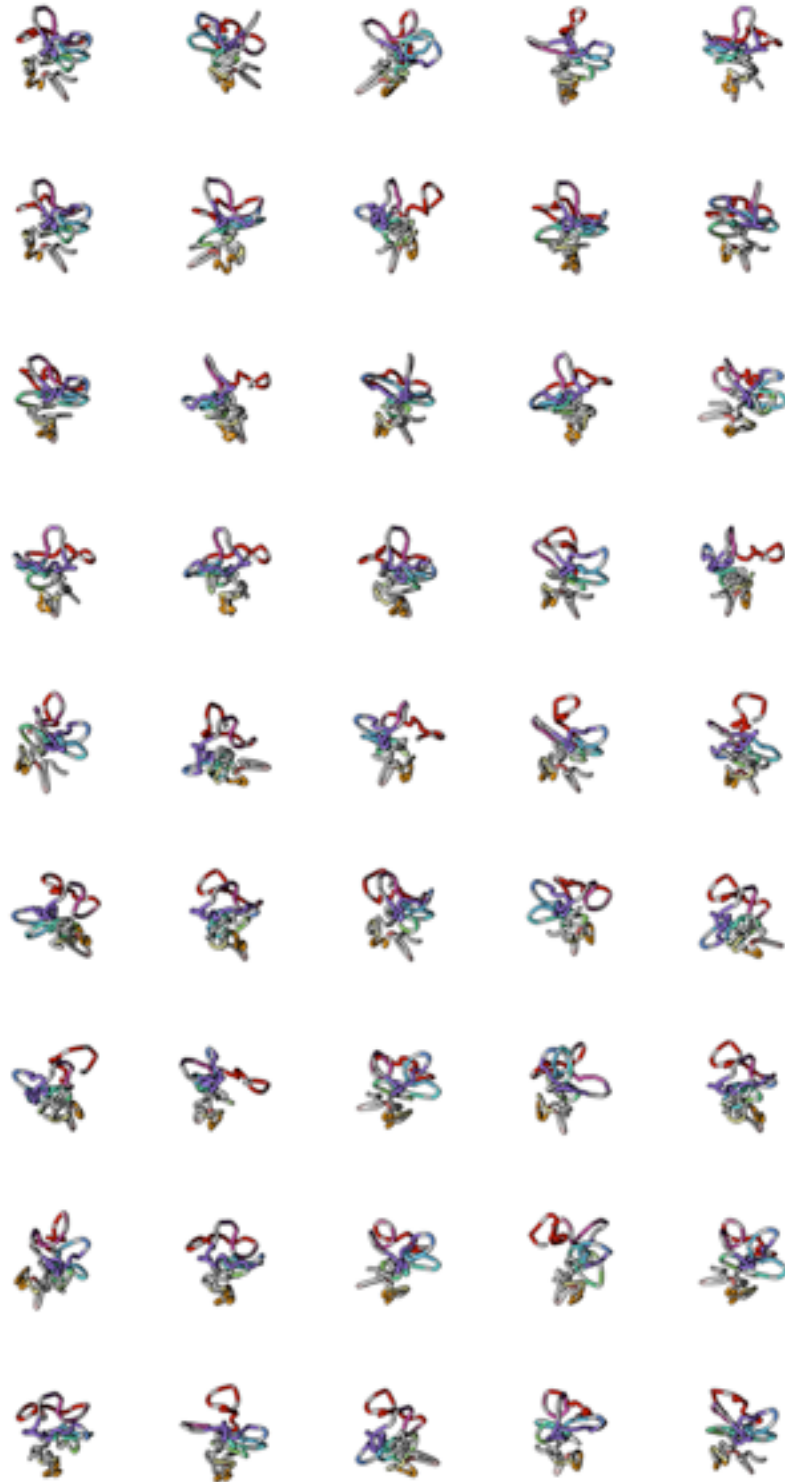
70 fragments
1,049 restraints



Optimization

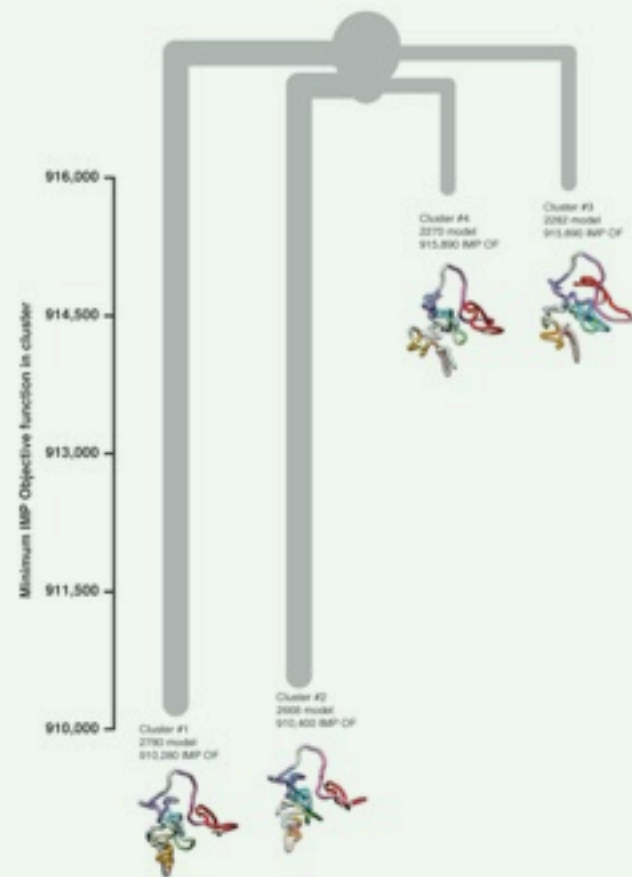
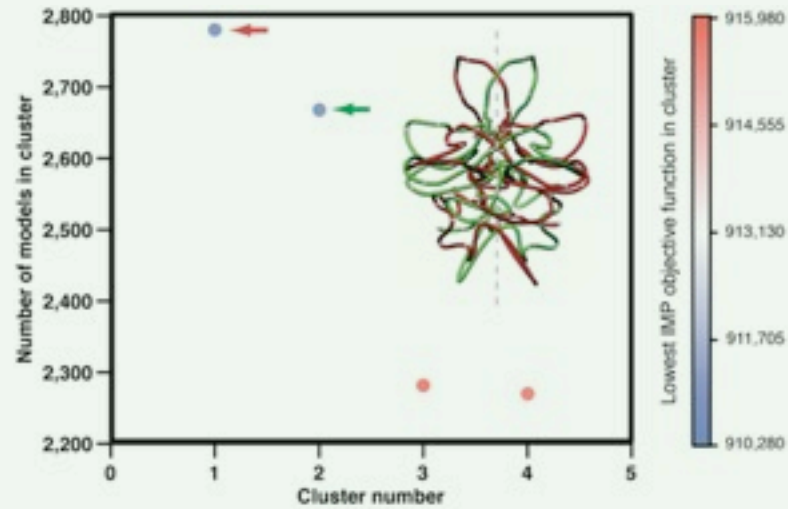


Clustering

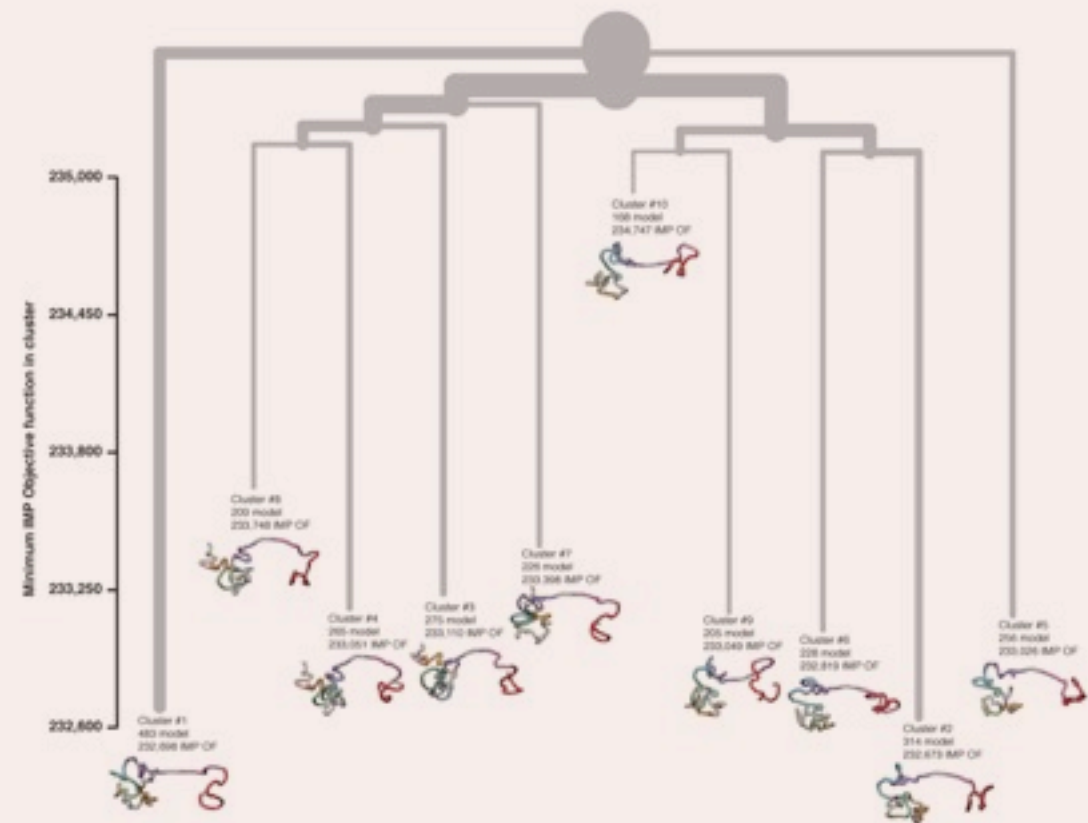
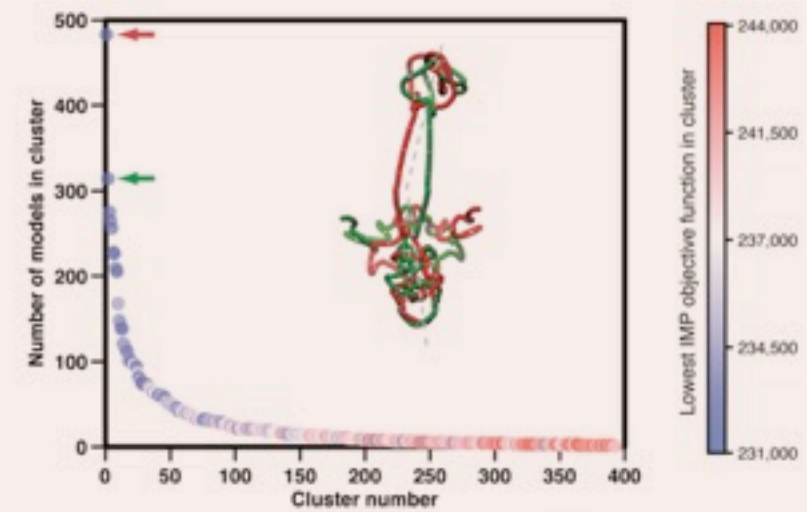


Not just one solution

GM12878



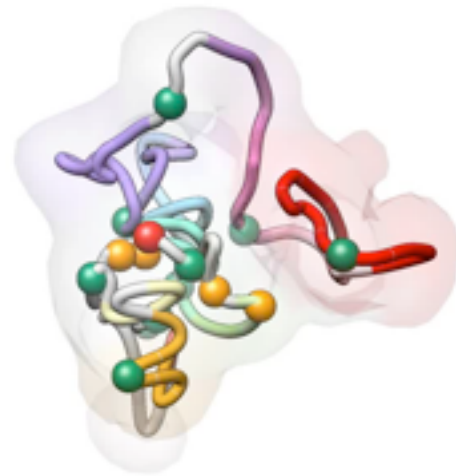
K562



Consistency

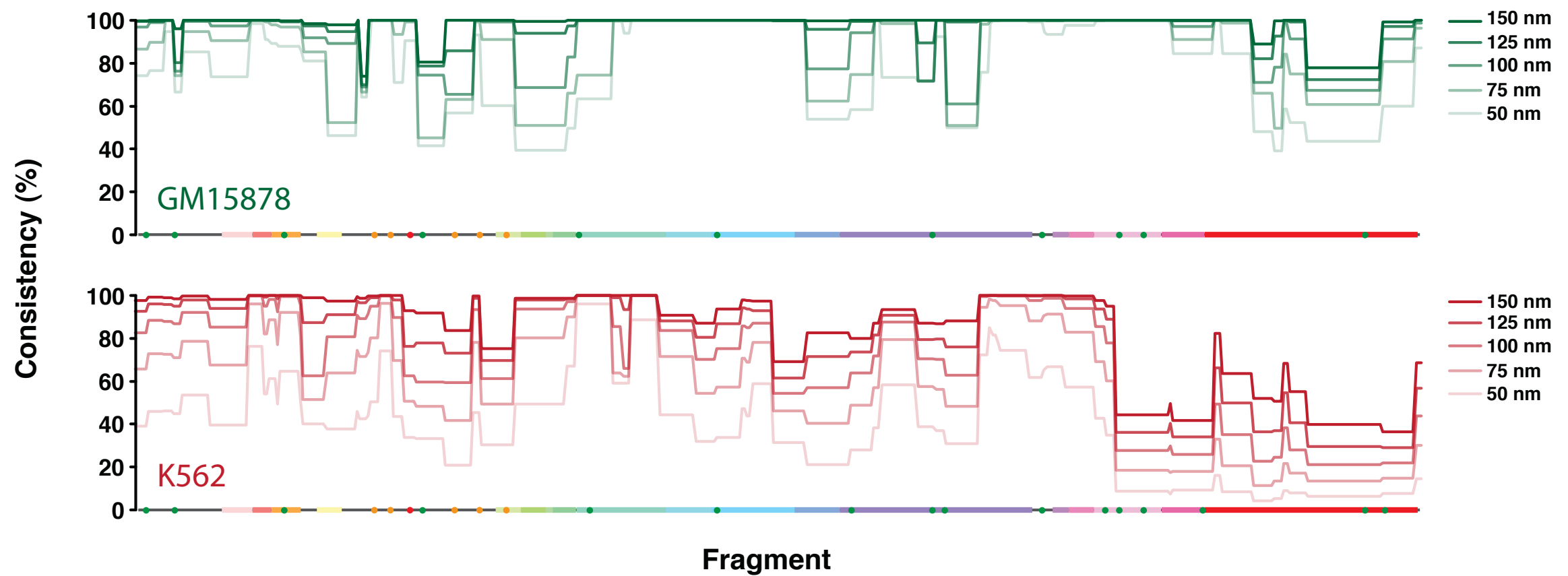
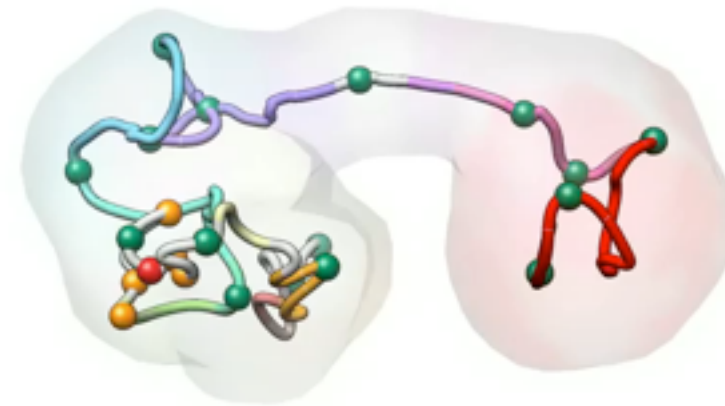
GM12878

Cluster #1
2780 model



K562

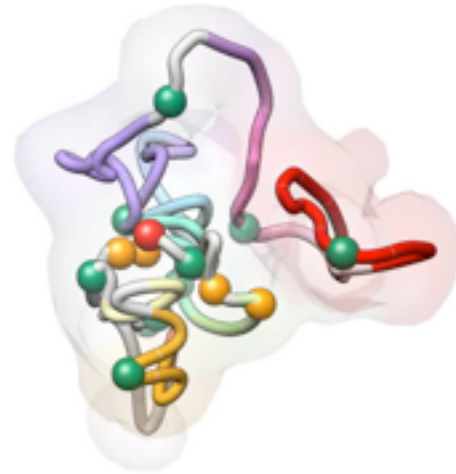
Cluster #2
314 model



Regulatory elements

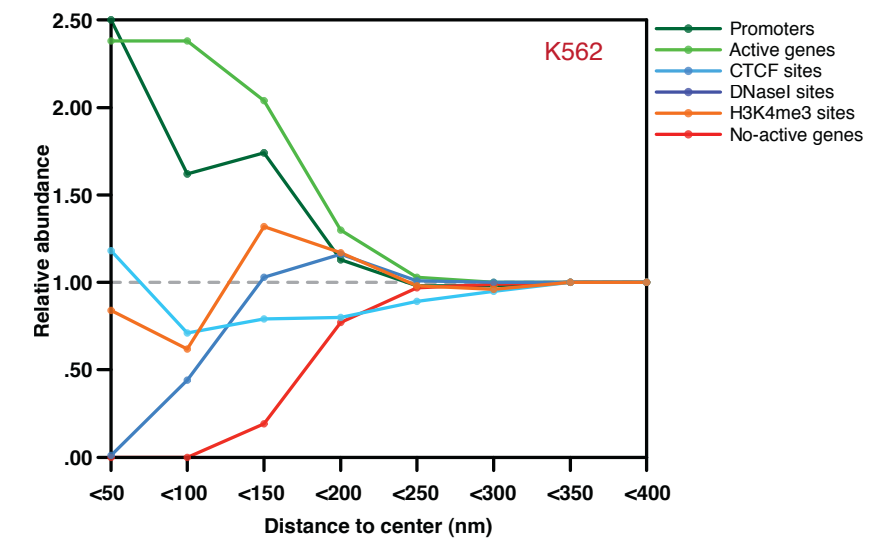
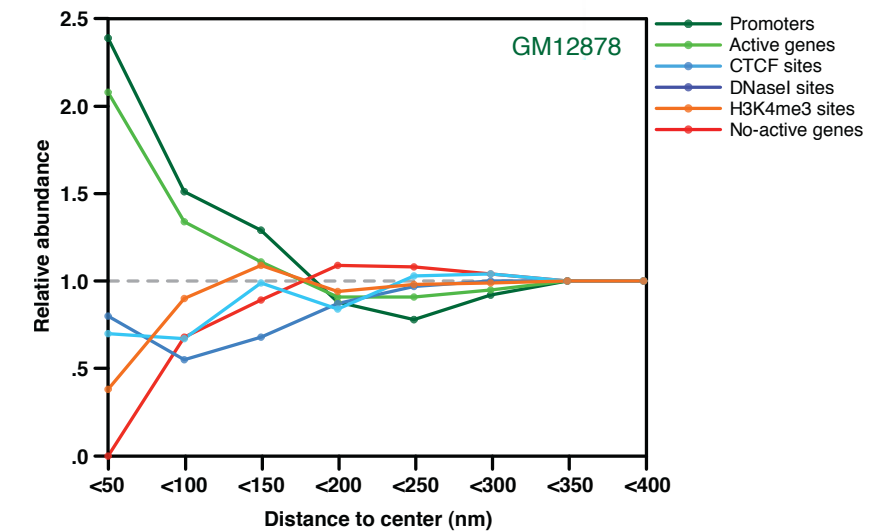
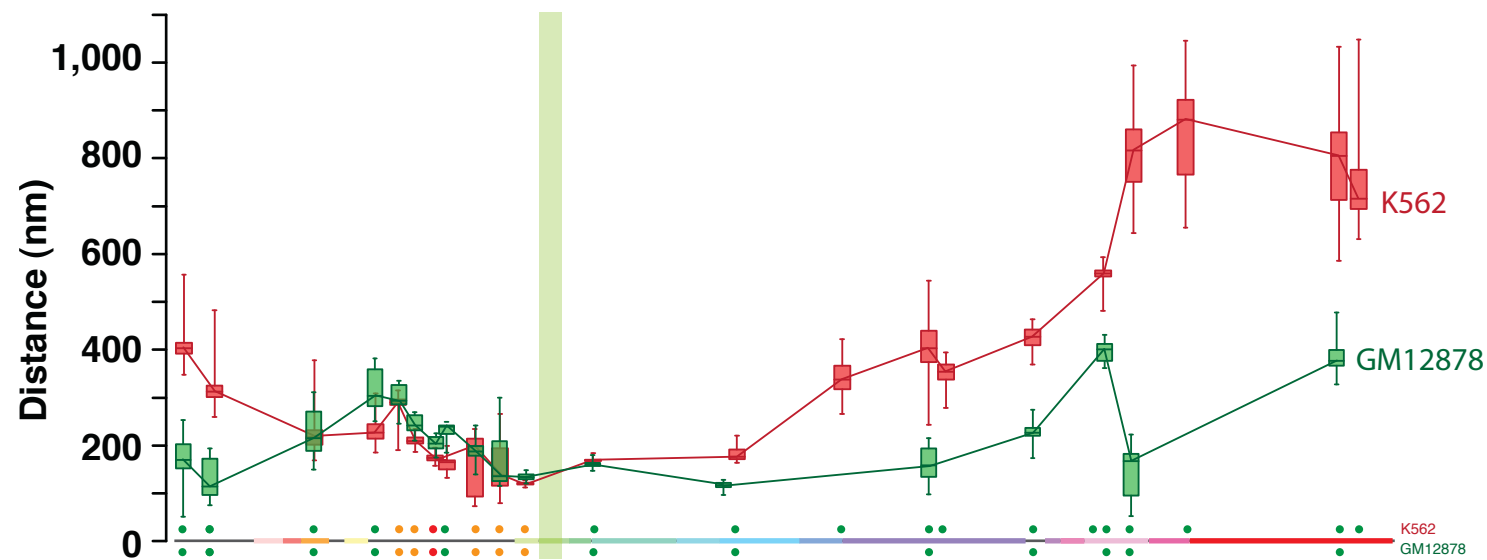
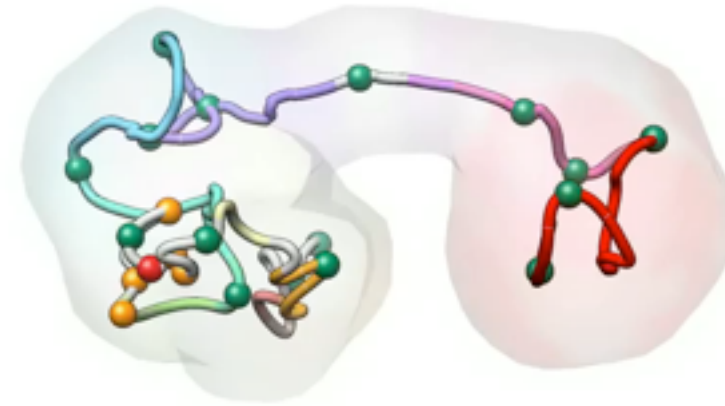
GM12878

Cluster #1
2780 model



K562

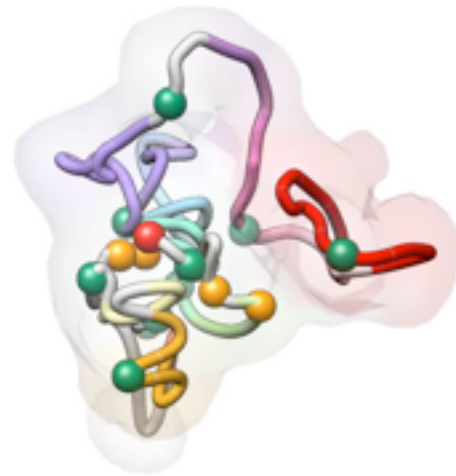
Cluster #2
314 model



Compactness

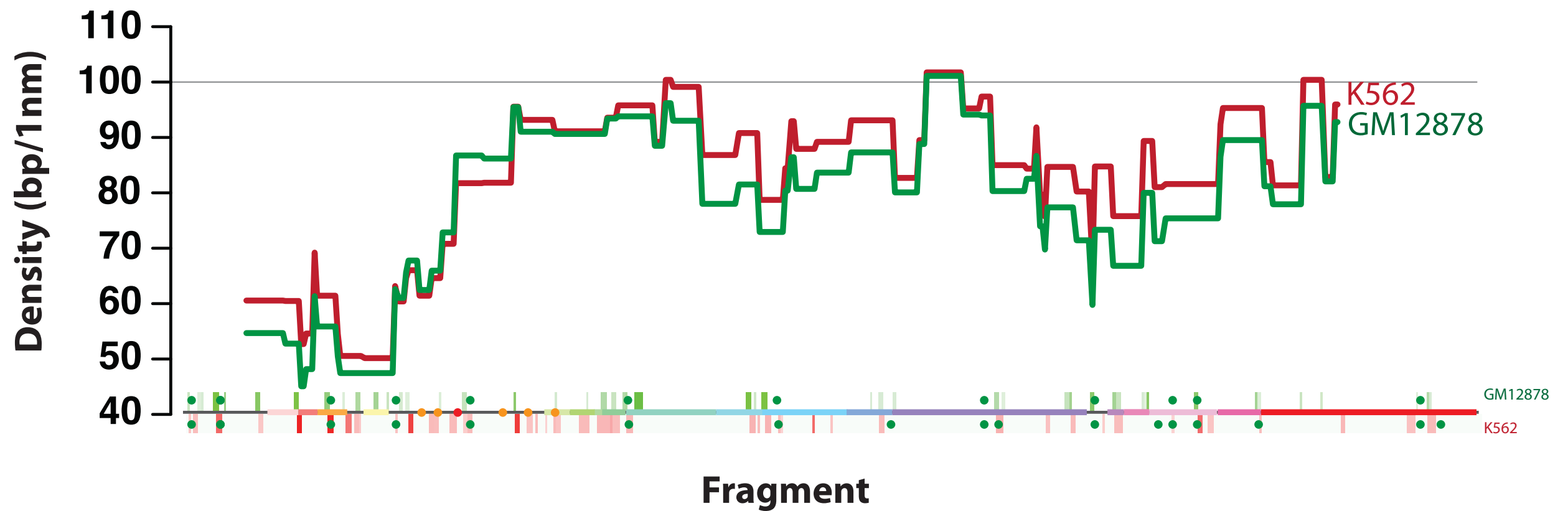
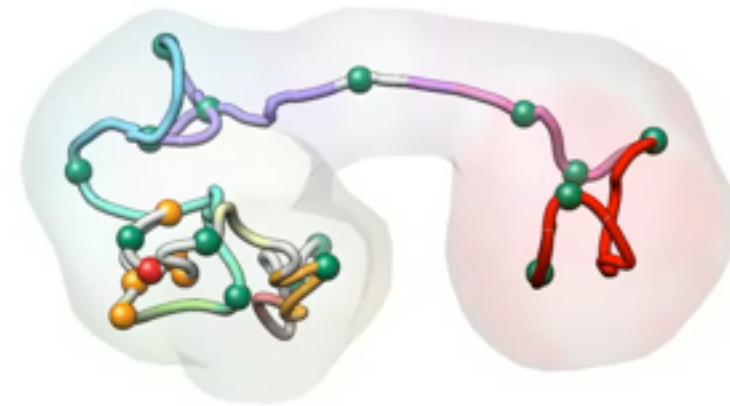
GM12878

Cluster #1
2780 model



K562

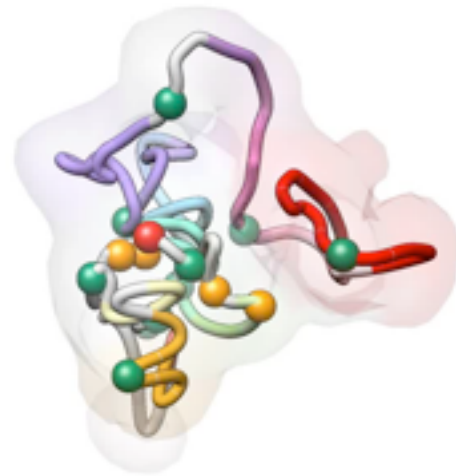
Cluster #2
314 model



Multi-loops

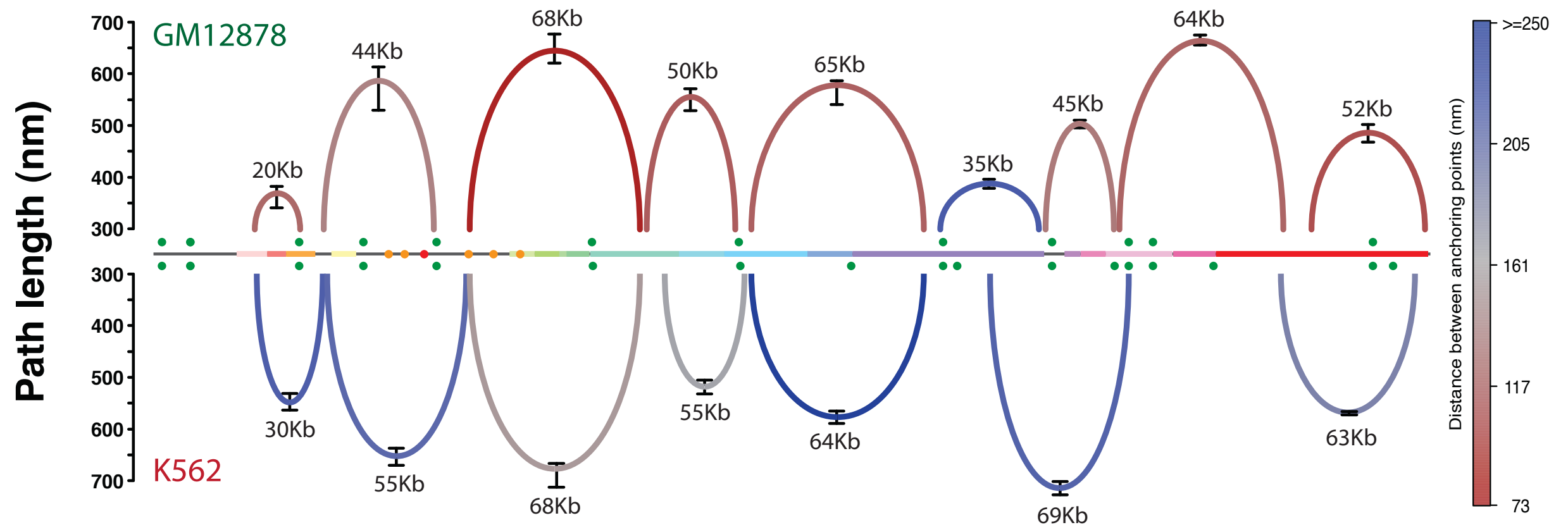
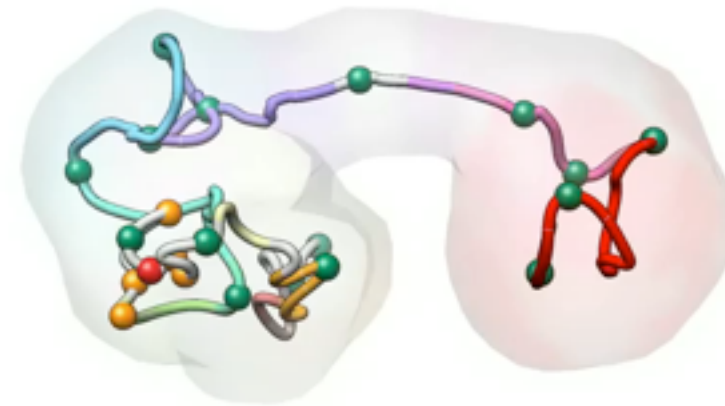
GM12878

Cluster #1
2780 model



K562

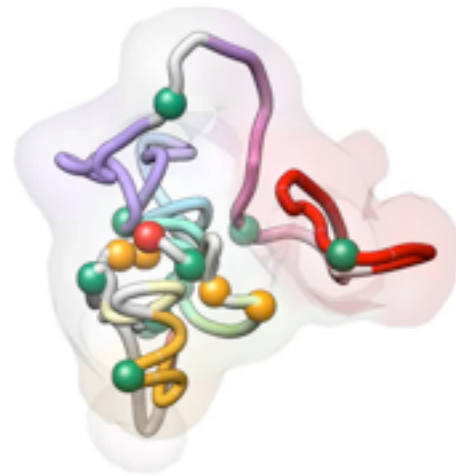
Cluster #2
314 model



Expression

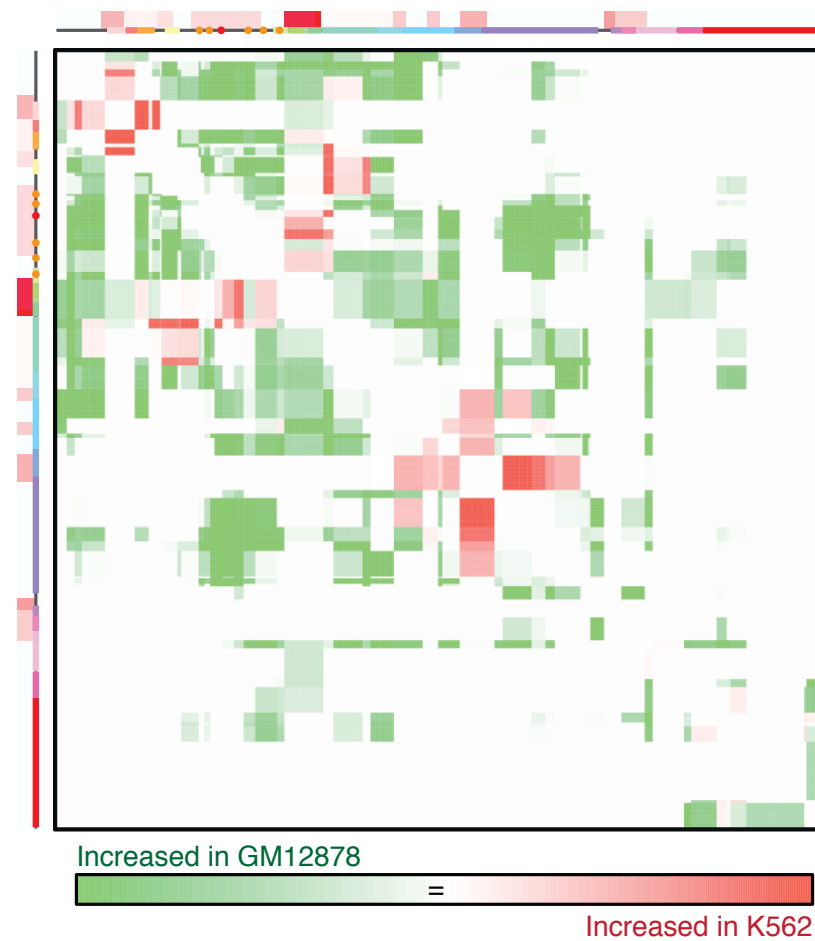
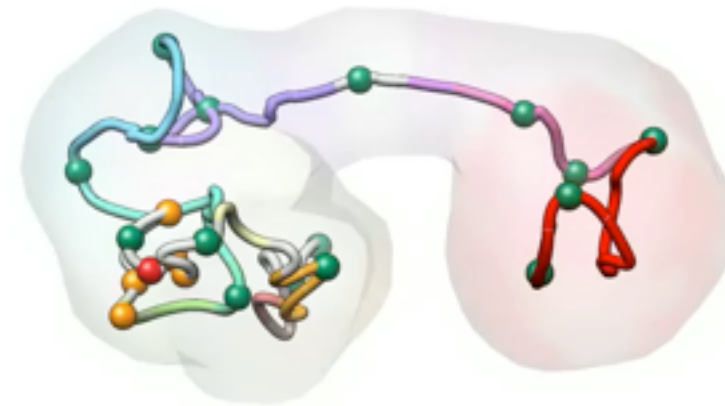
GM12878

Cluster #1
2780 model



K562

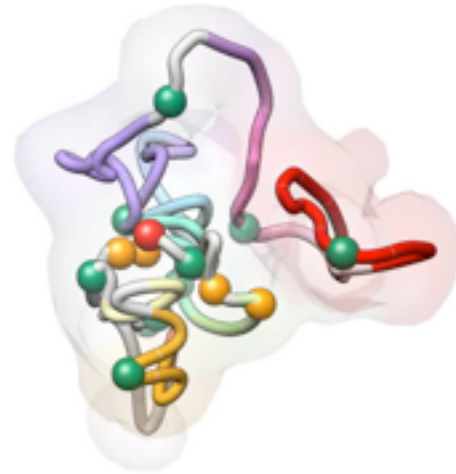
Cluster #2
314 model



FISH validation

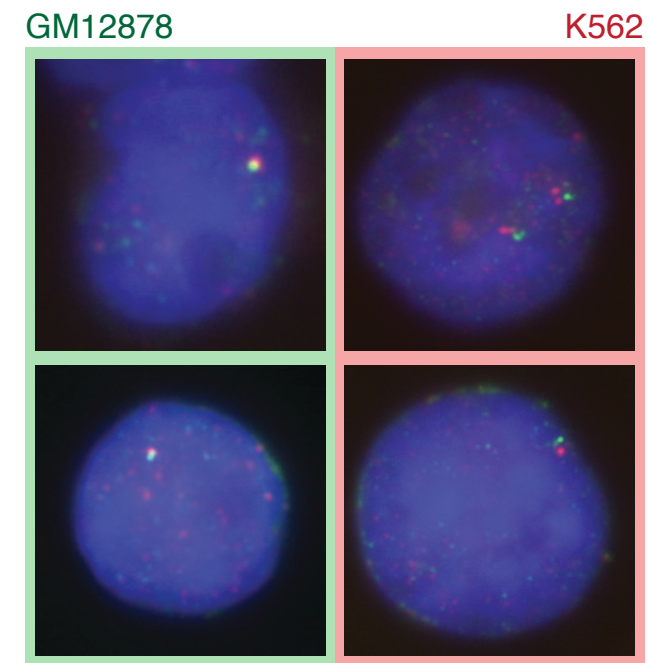
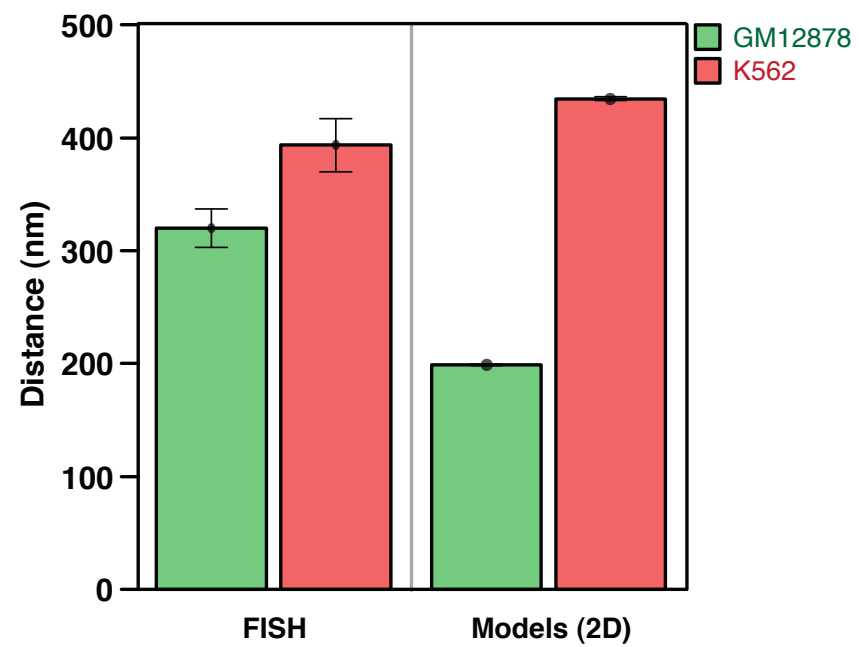
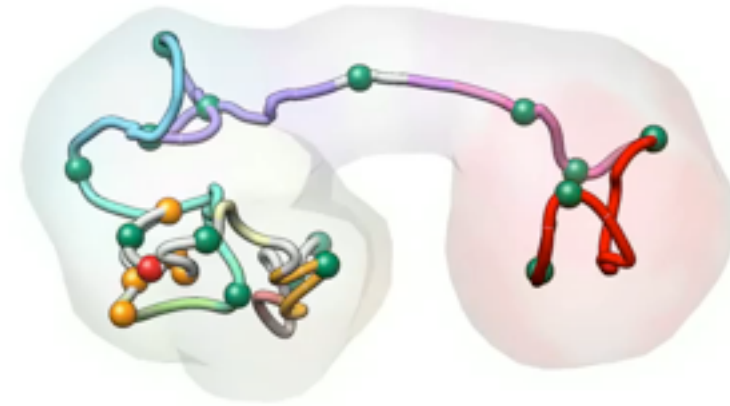
GM12878

Cluster #1
2780 model

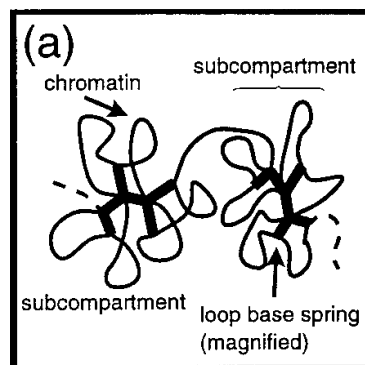


K562

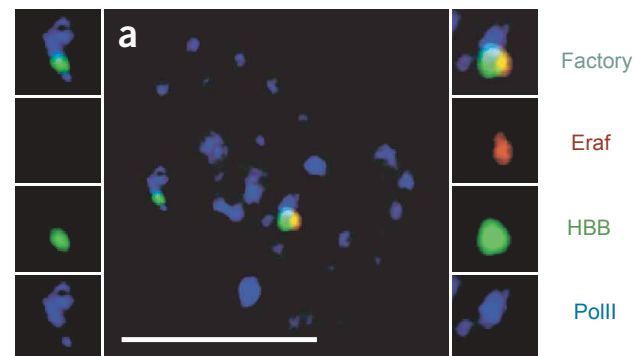
Cluster #2
314 model



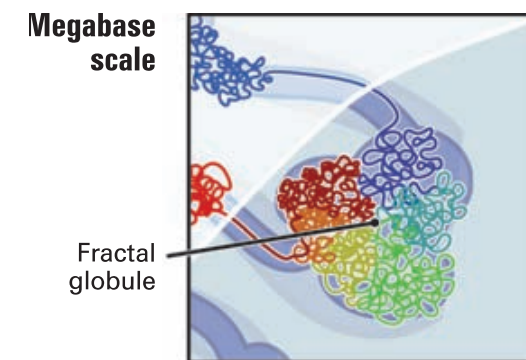
The “Chromatin Globule” model



Münkel et al. JMB (1999)



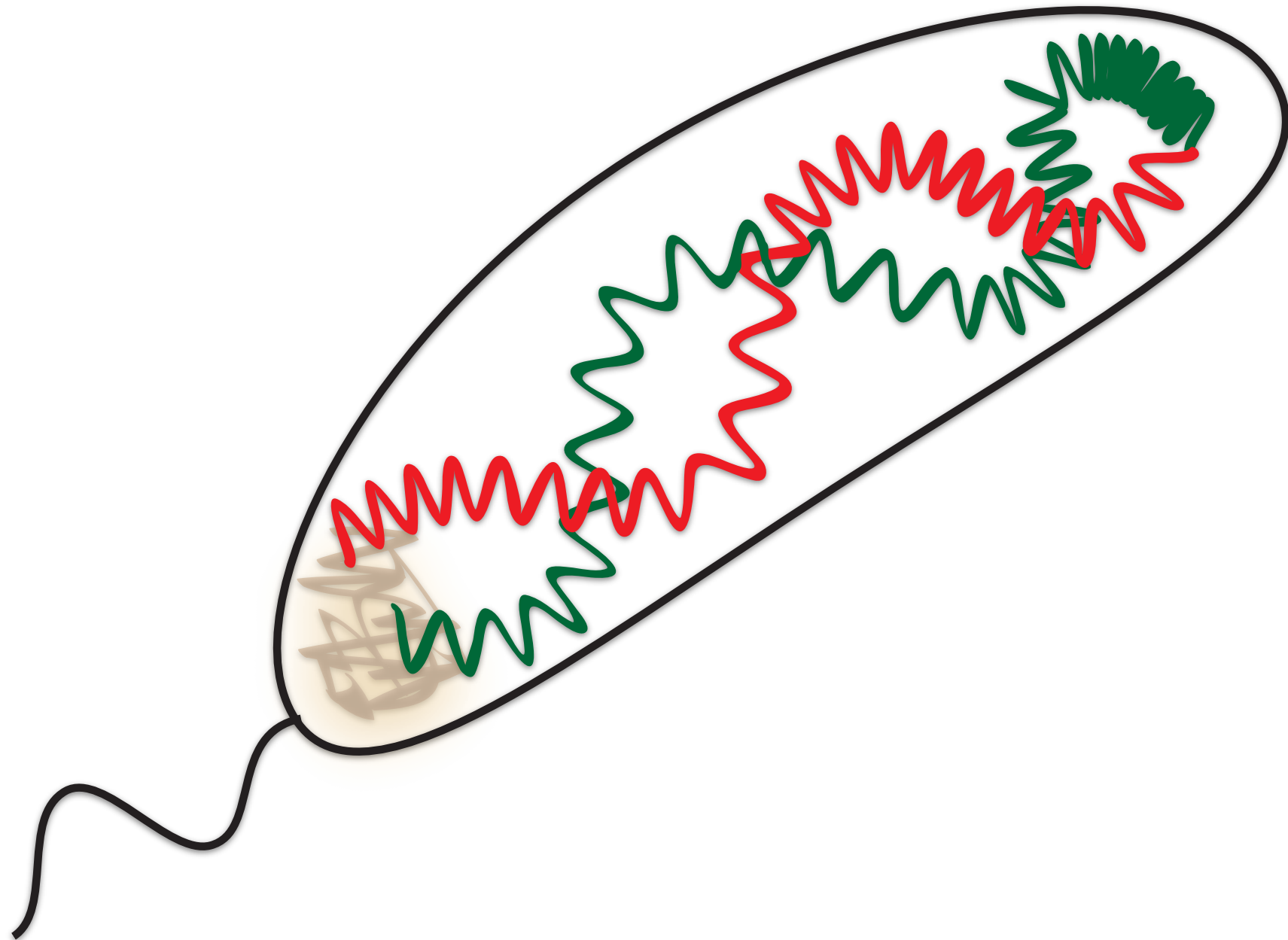
Osborne et al. Nat Genet (2004)



Lieberman-Aiden et al. Science (2009)

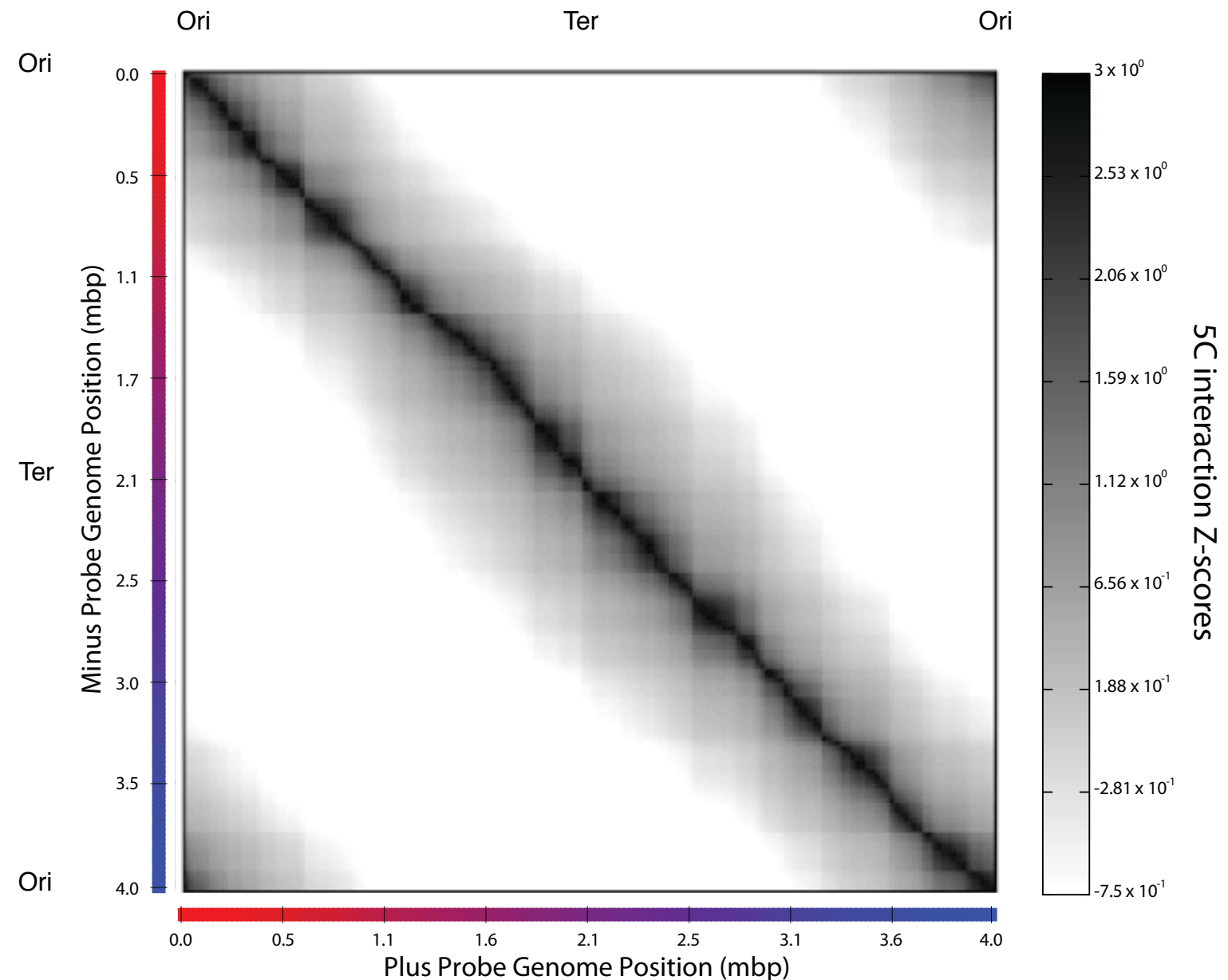
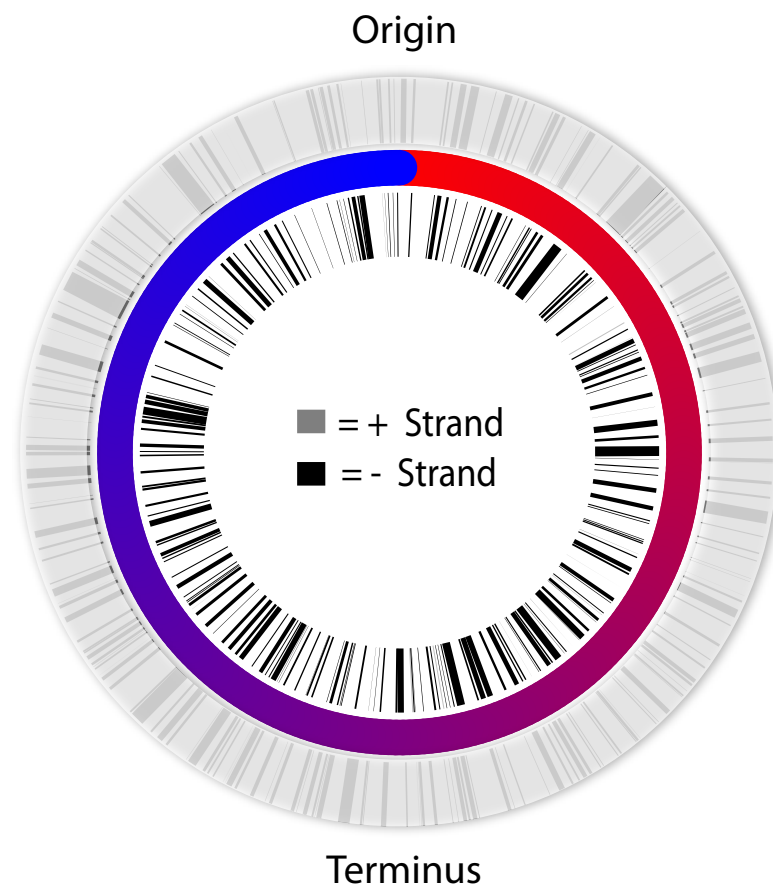
D. Baù et al. **Nat Struct Mol Biol** (2011) 18:107-14
 A. Sanyal et al. **Current Opinion in Cell Biology** (2011) 23:325–33.

Caulobacter crescentus genome



The 3D architecture of *Caulobacter Crescentus*

4,016,942 bp & 3,767 genes

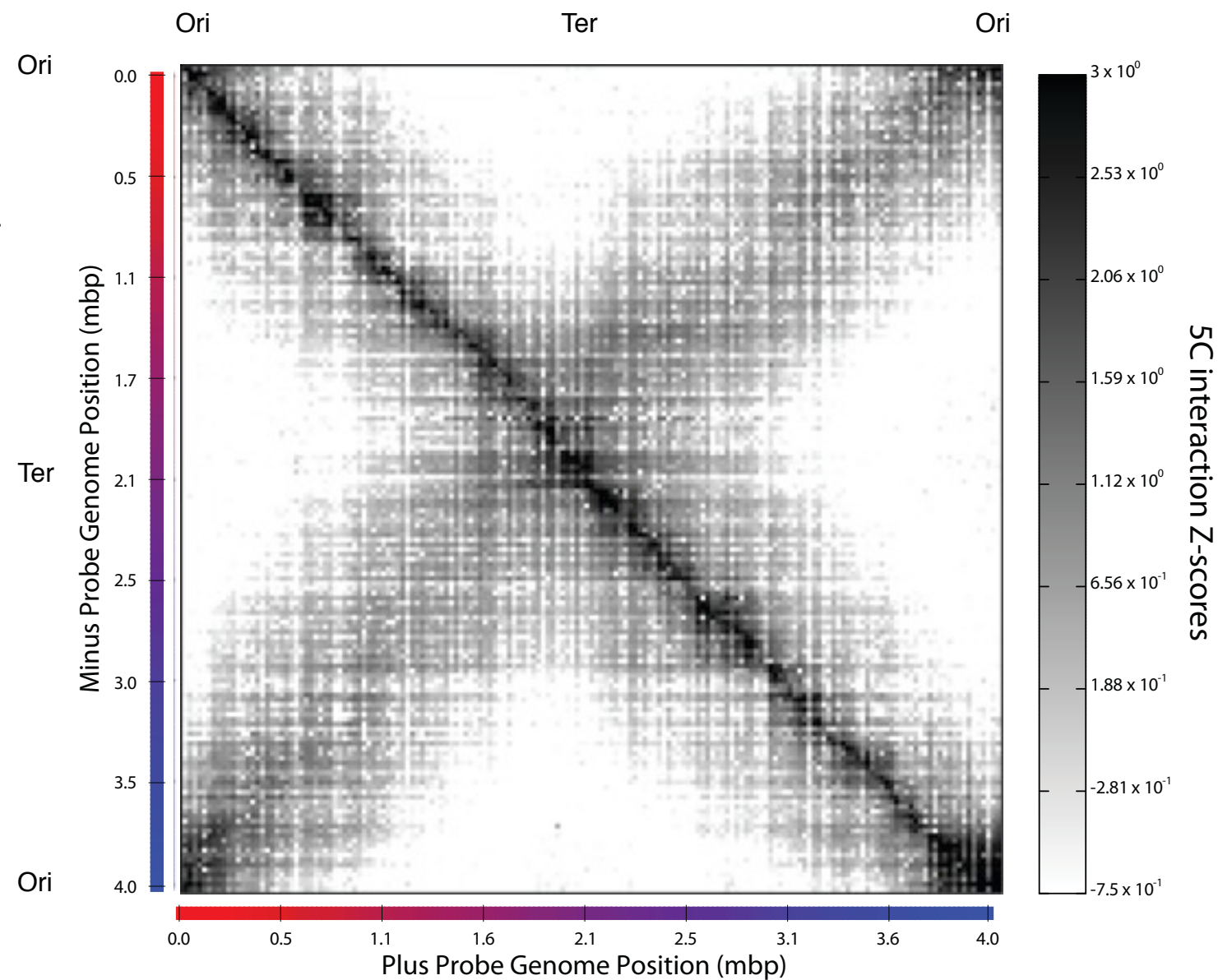
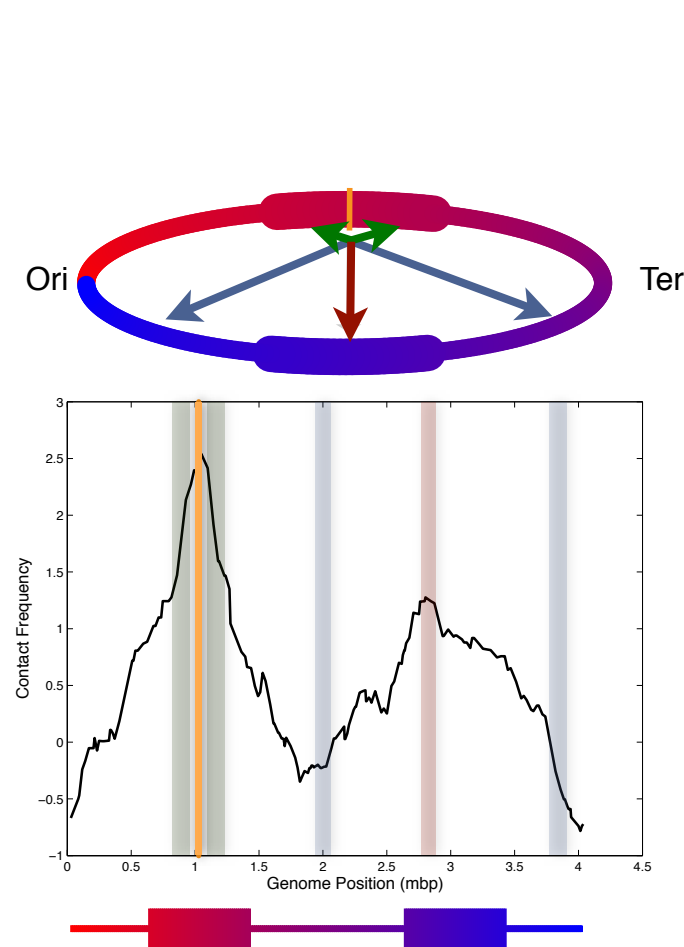
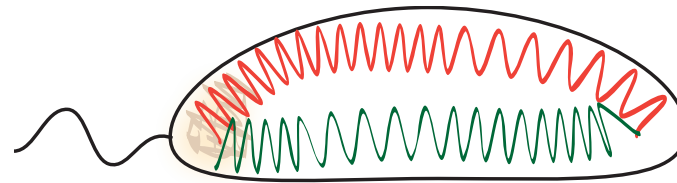


169 5C primers on + strand
170 5C primers on - strand
28,730 chromatin interactions

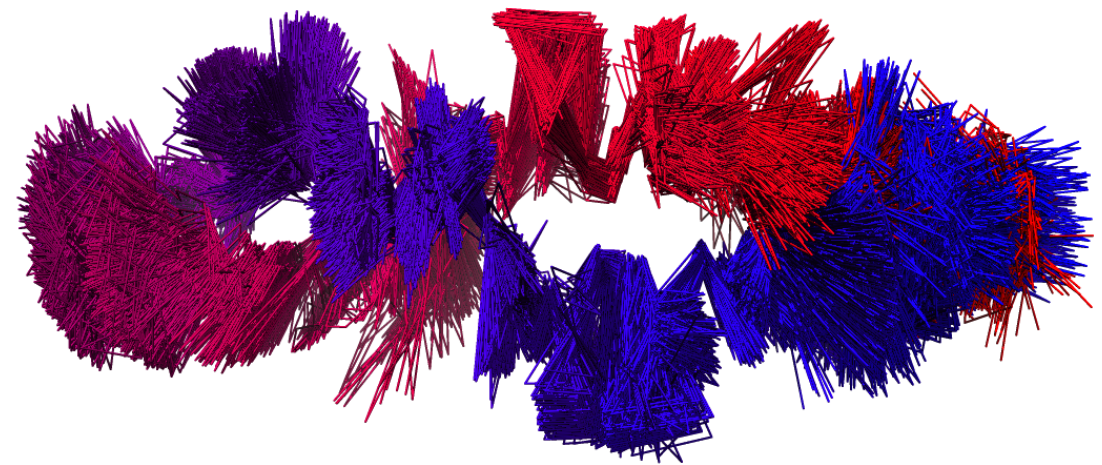
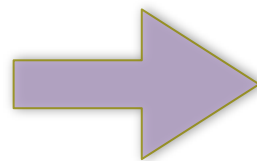
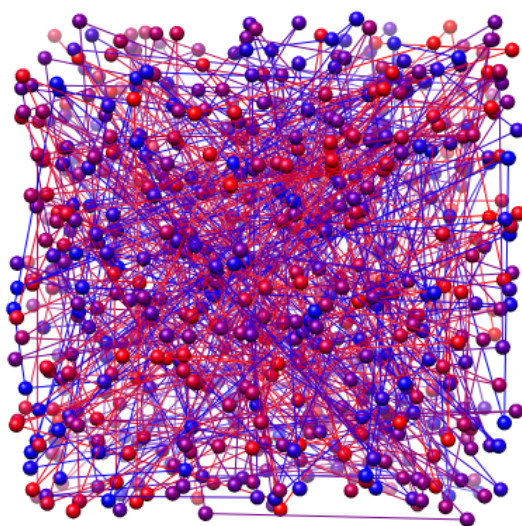
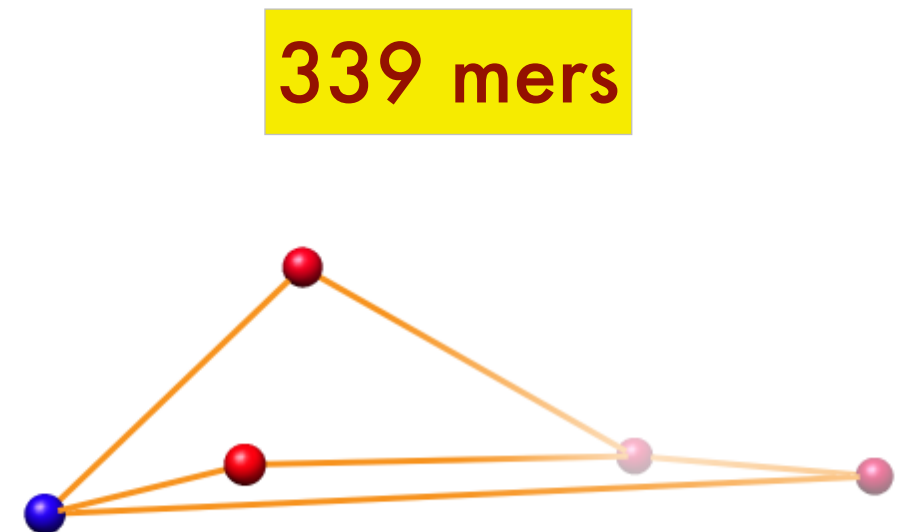
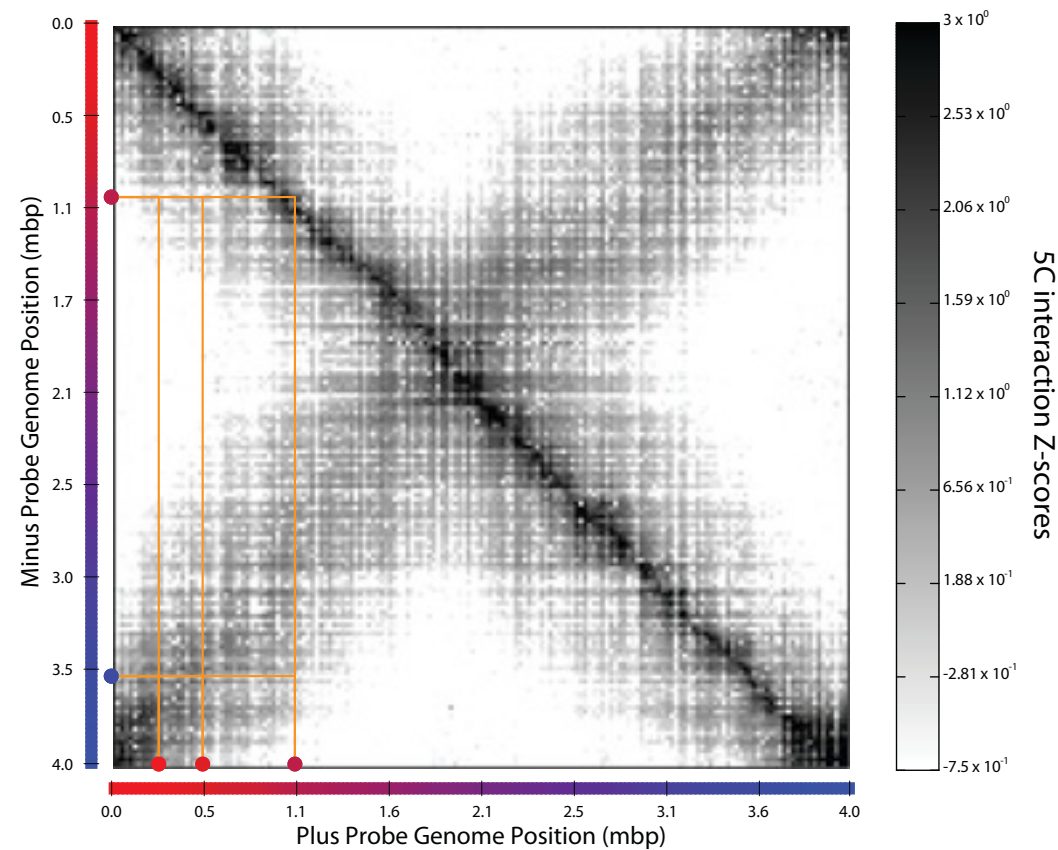
~13Kb

5C interaction matrix

ELLIPSOID for *Caulobacter crescentus*



3D model building with the 5C + IMP approach



Genome organization in *Caulobacter crescentus*

Arms are helical

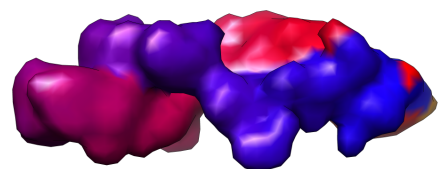
Resolution

dif site 47 ± 17 Kb from Ter

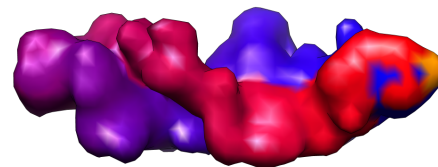
Centromer-like

parS sites 25 ± 17 Kb from Ori

Cluster 1

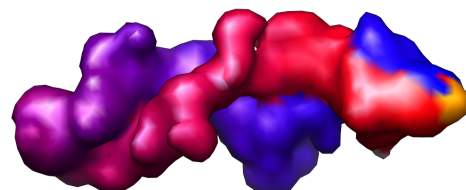


180°

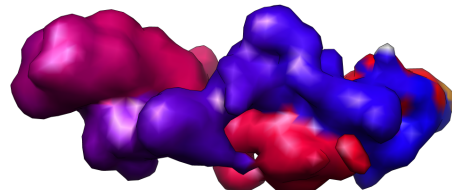


500 nm

Cluster 2

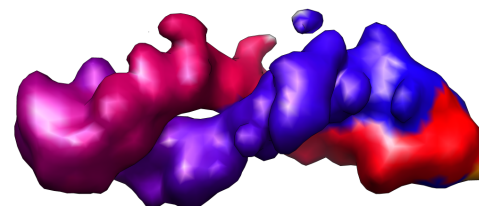


180°

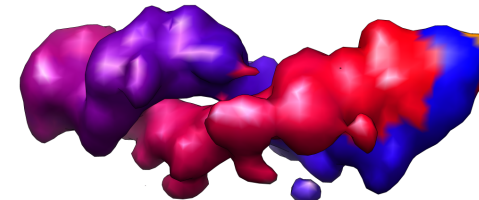


500 nm

Cluster 3

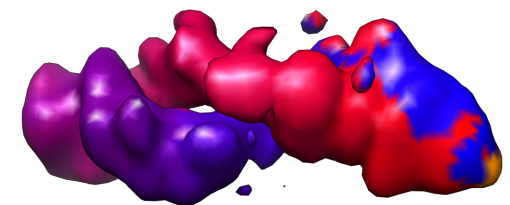


180°



500 nm

Cluster 4



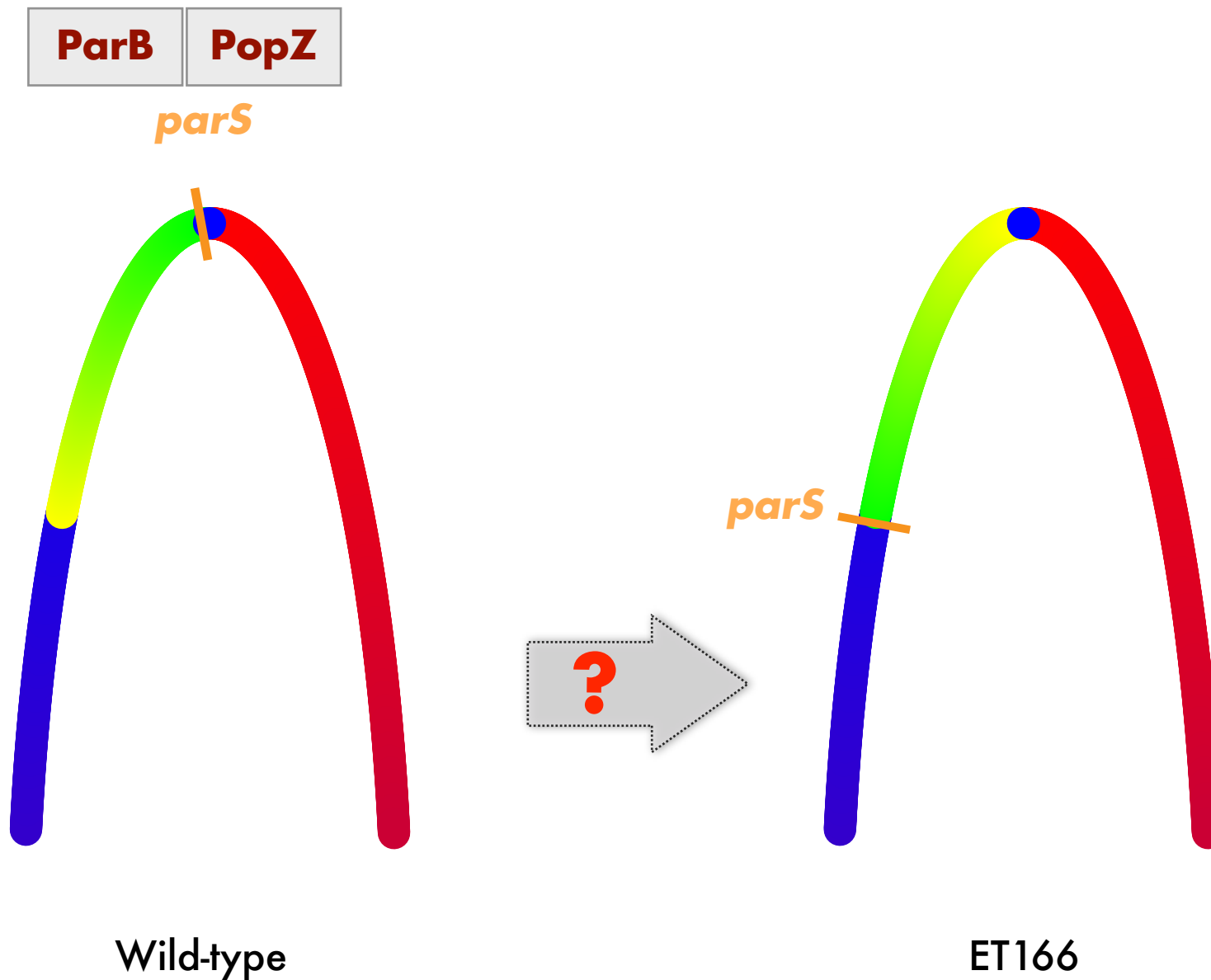
180°



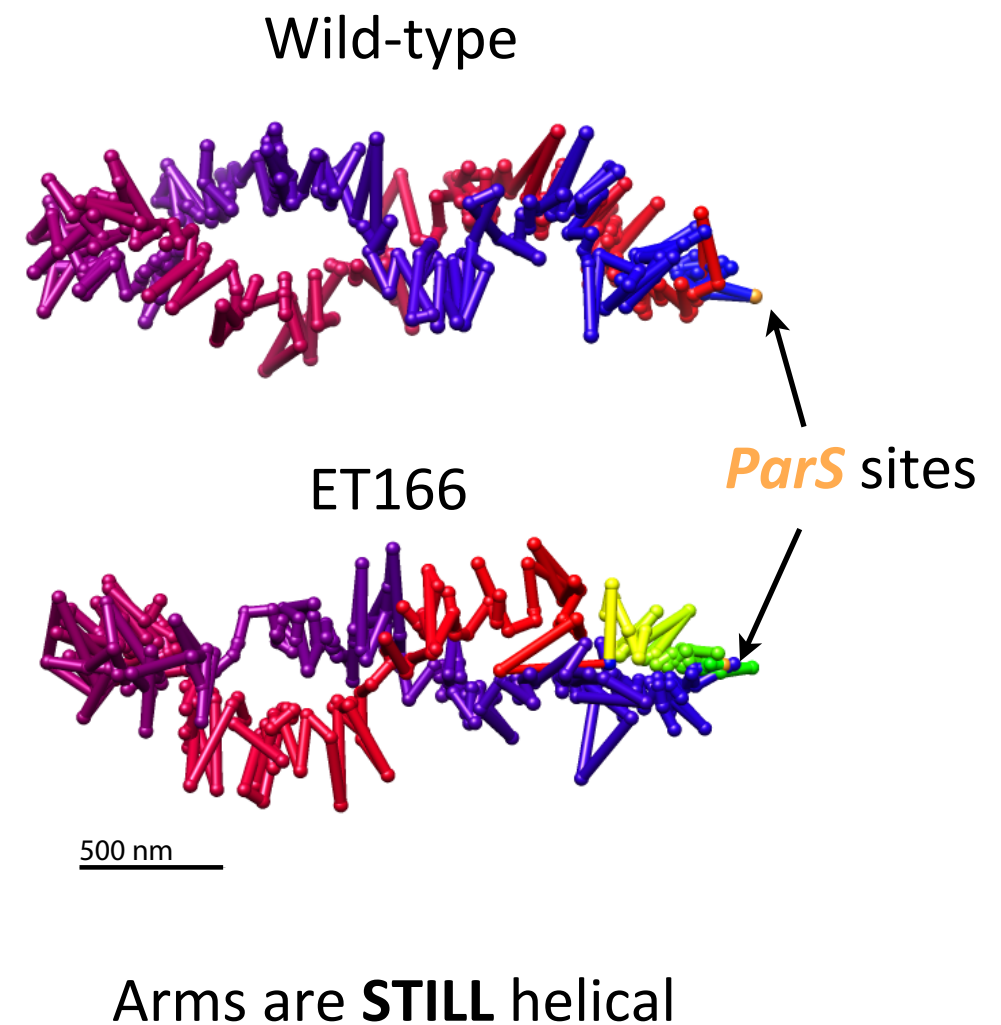
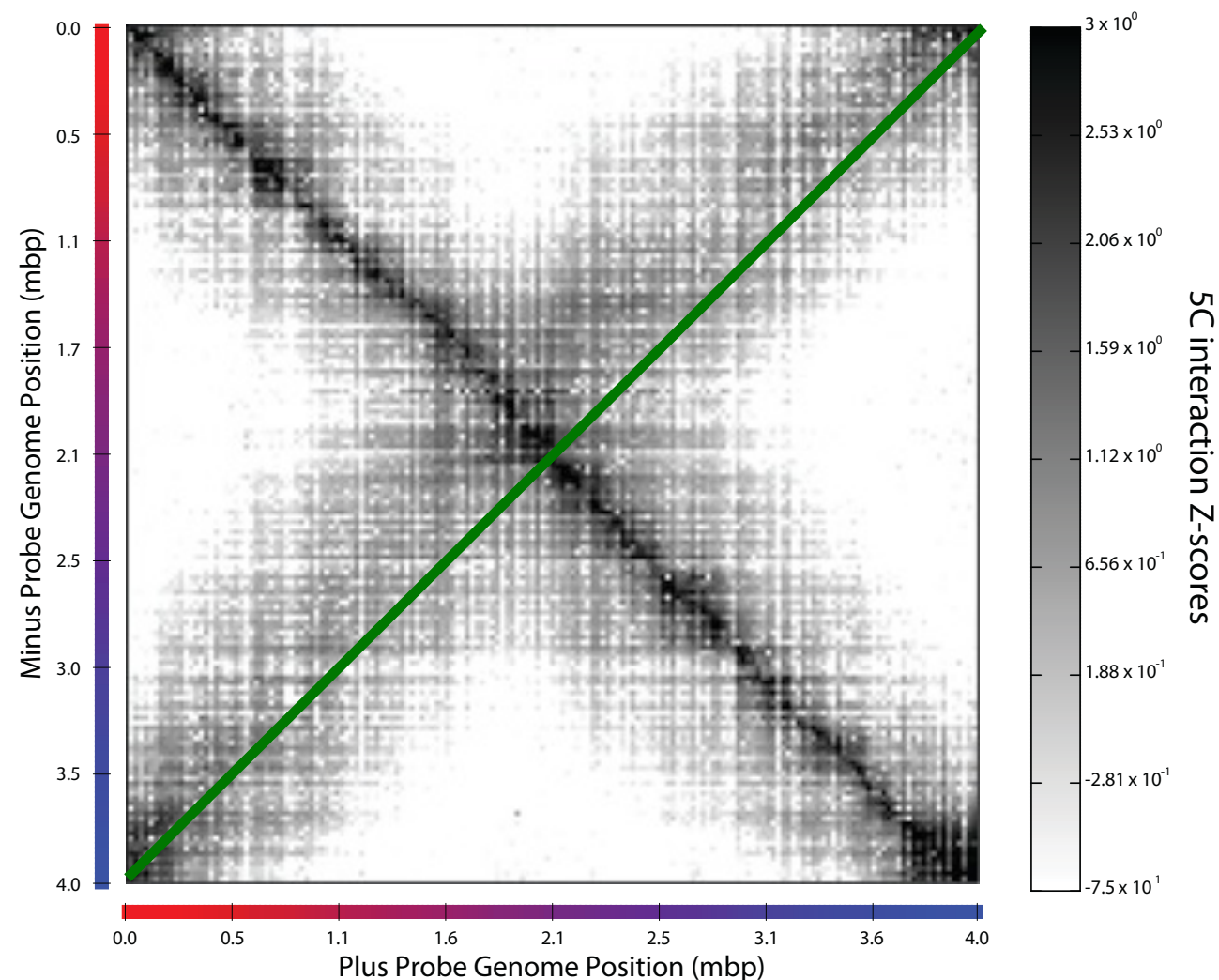
500 nm

MIRRORS!

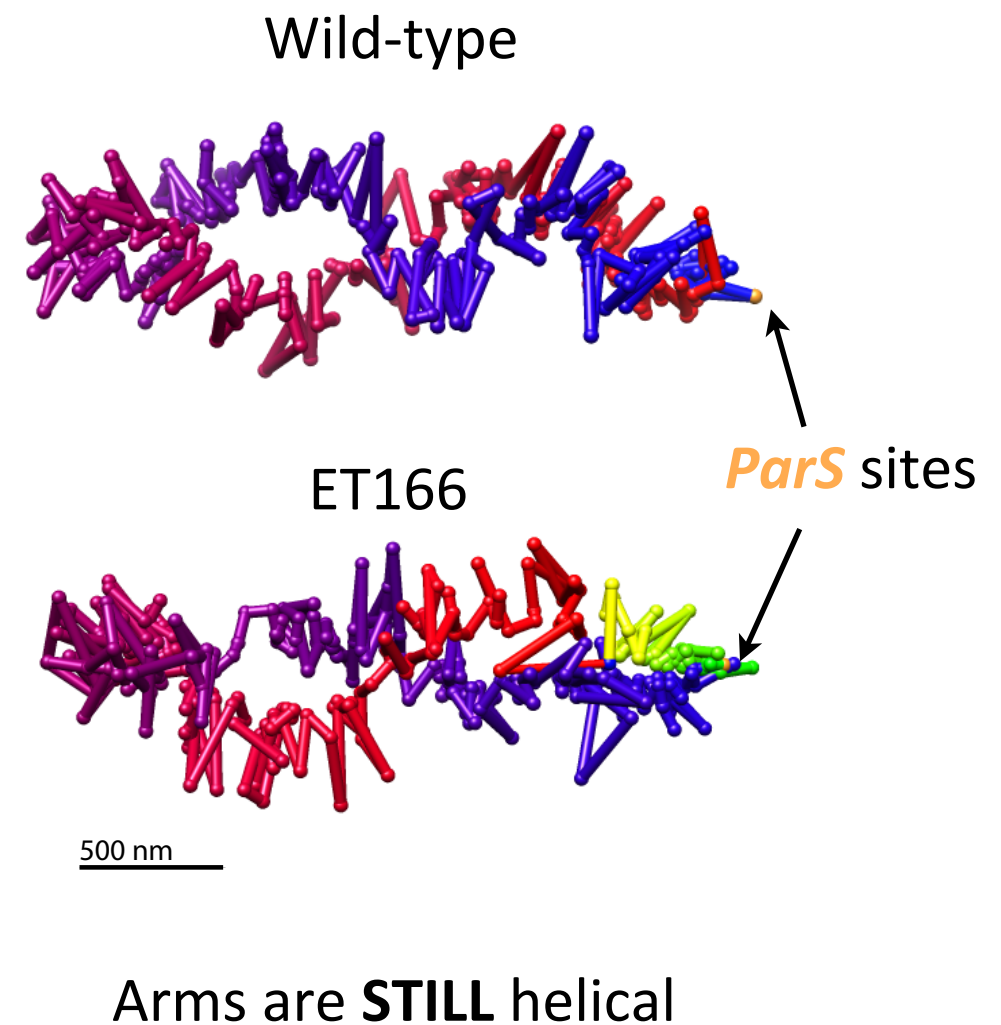
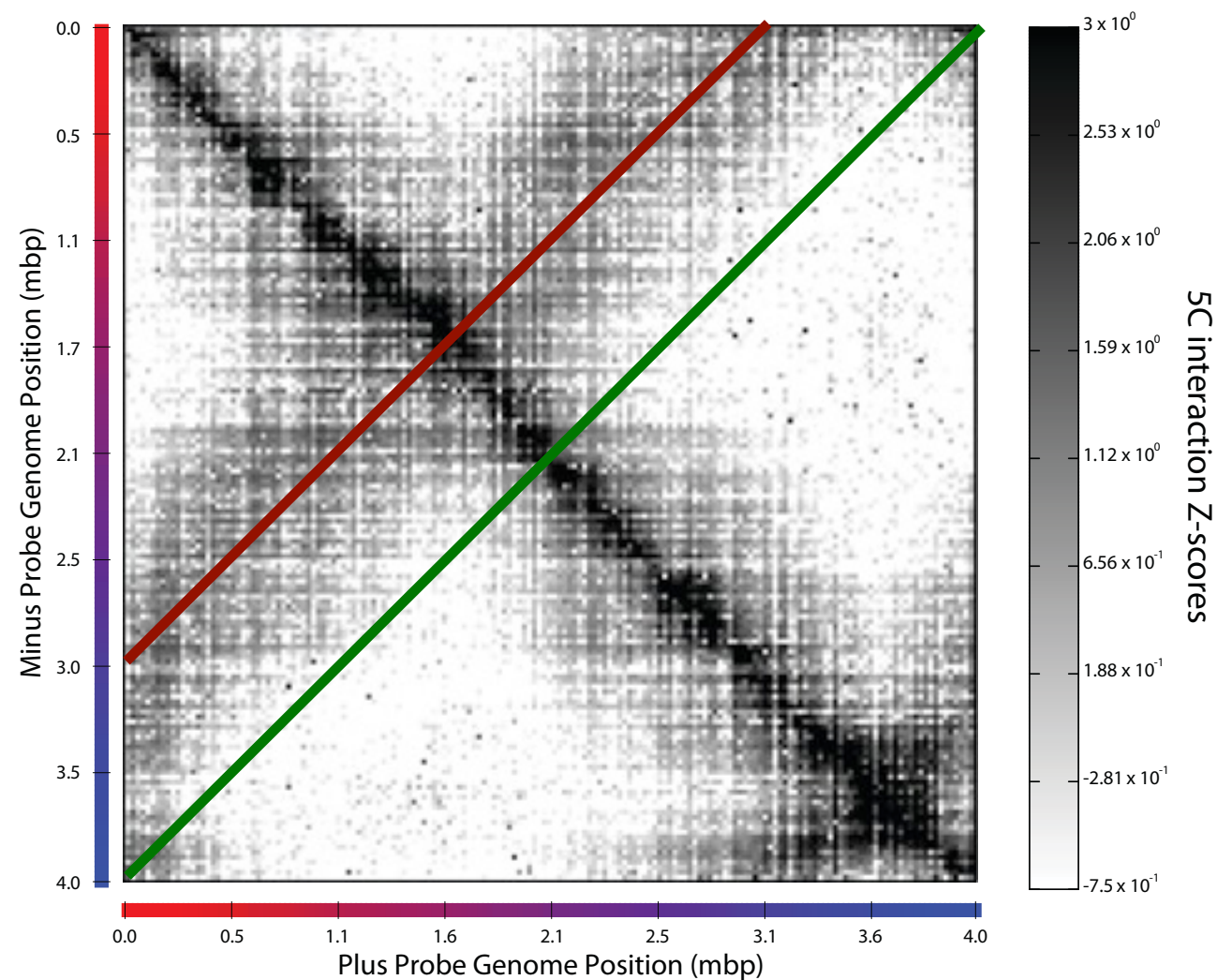
Moving the **parS** sites 400 Kb away from Ori



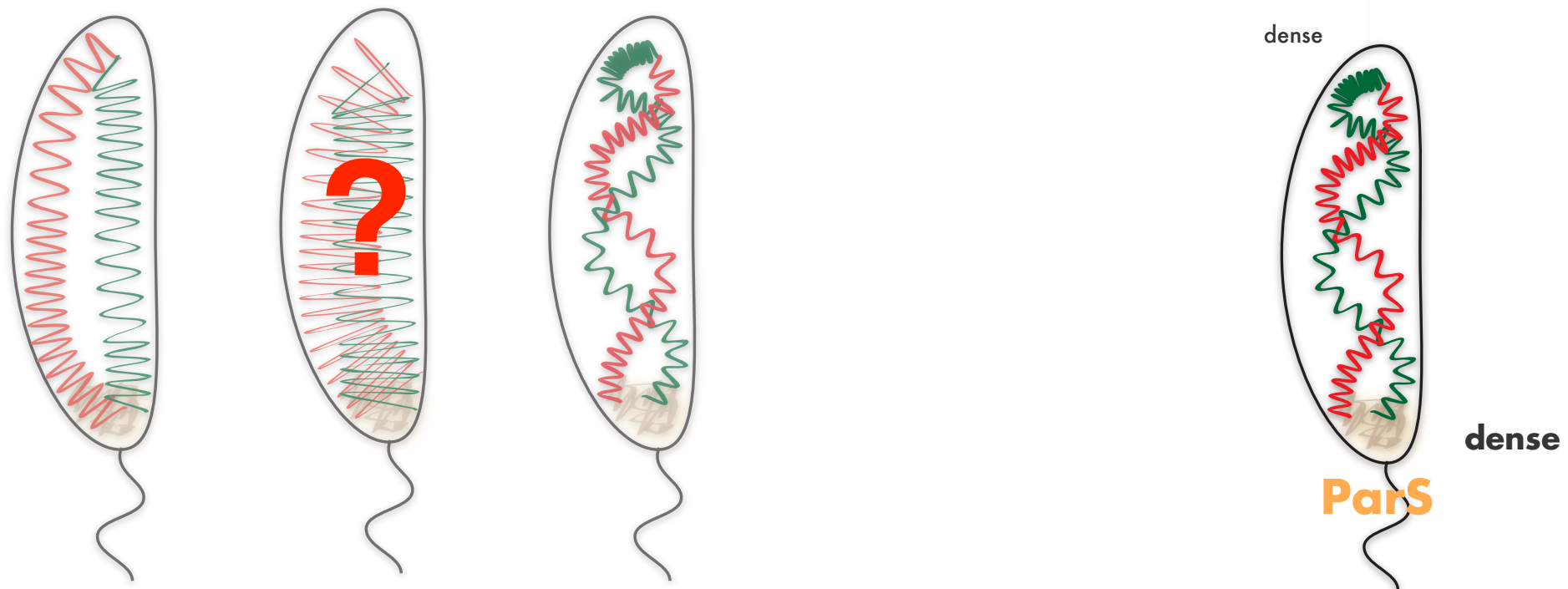
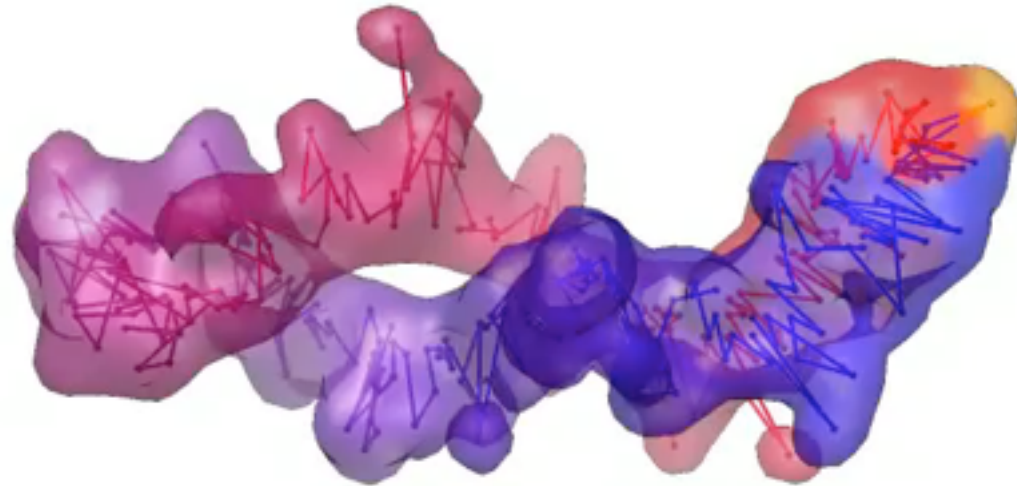
Moving the **parS** sites results in whole genome rotation!



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



Genome architecture in *Caulobacter*

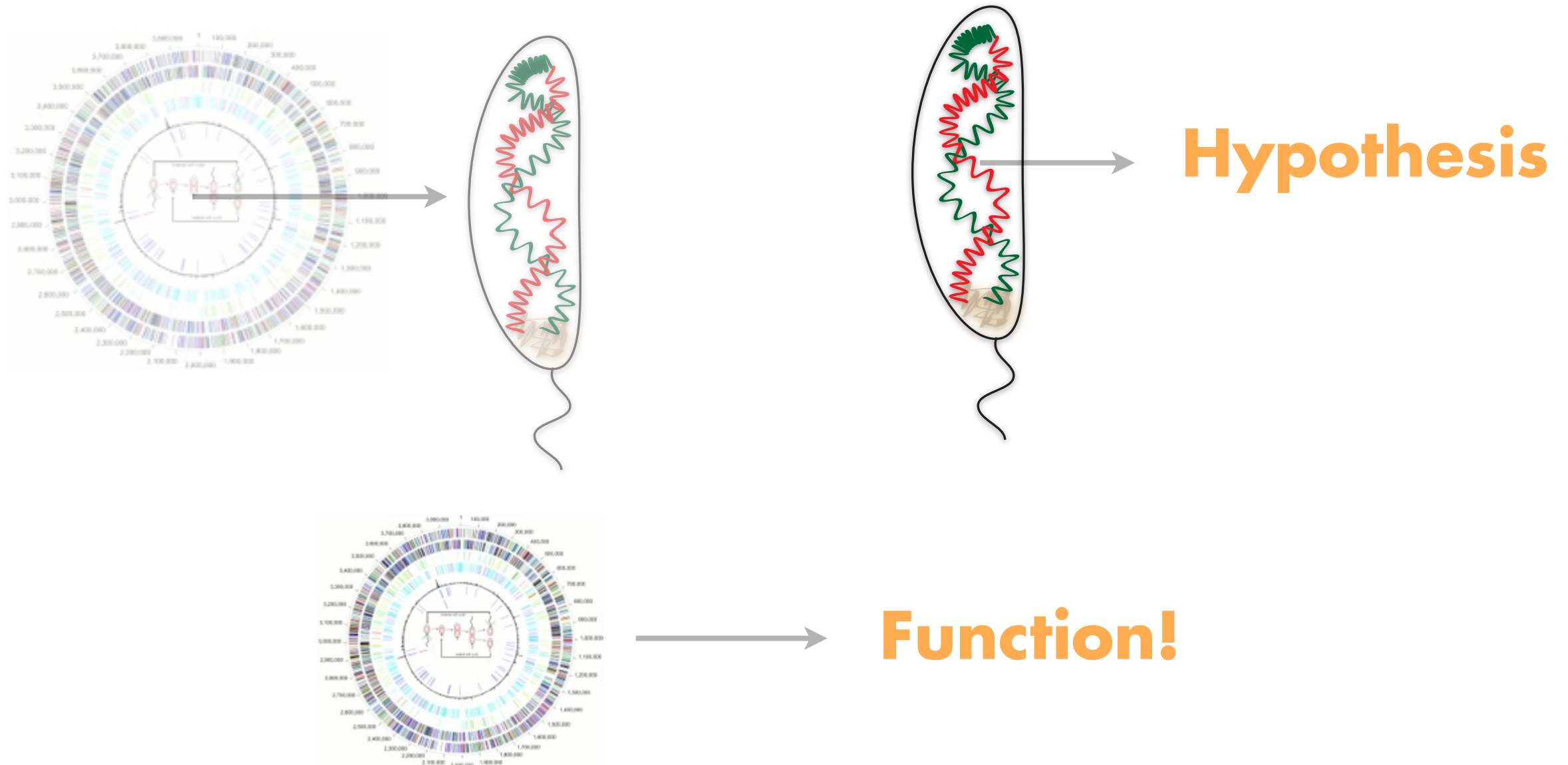


M.A. Umbarger, et al. **Molecular Cell** (2011) 44:252–264

From Sequence to Function

5C + IMP

Technology



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



OPEN POSITIONS!
Starting spring 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

Staff Scientist
CNAG



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



centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

Marc A. Martí-Renom

Genome Biology Group
Structural Genomics Team
National Center for Genomic Analysis
Barcelona, Spain.

<http://cnag.cat>
<http://marciuslab.org>
<http://integrativemodeling.org>