

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

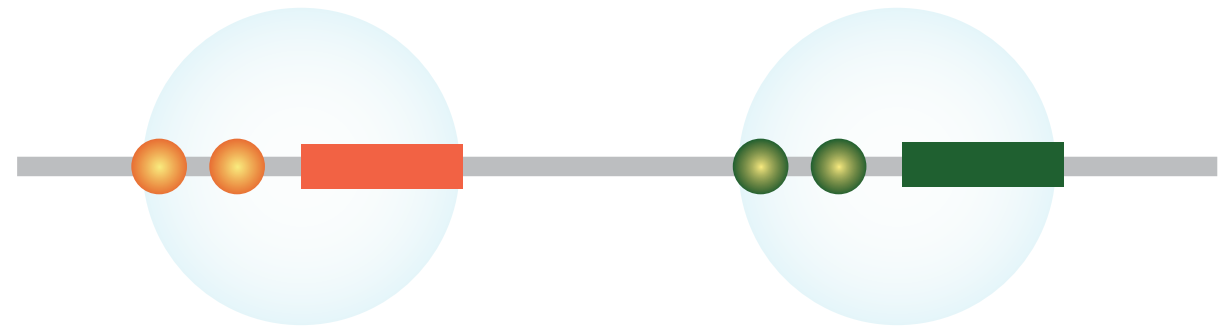
Marc A. Marti-Renom

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

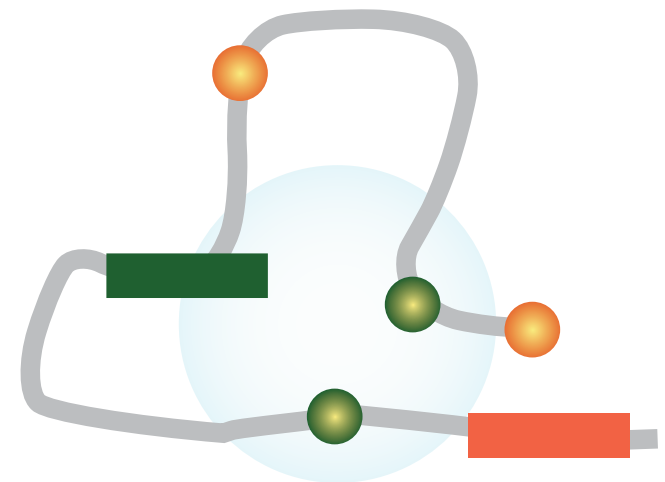
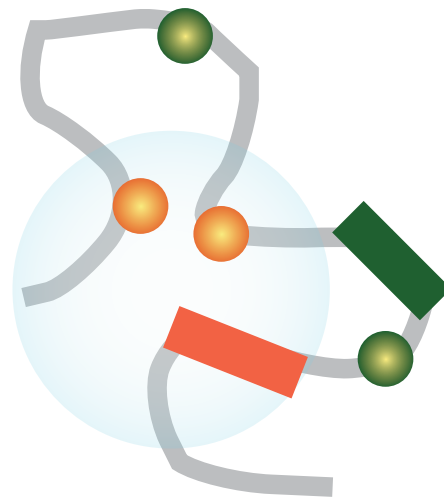
cnag



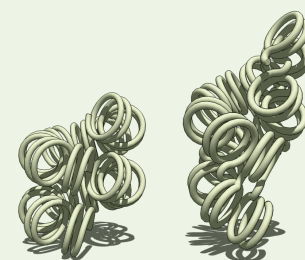
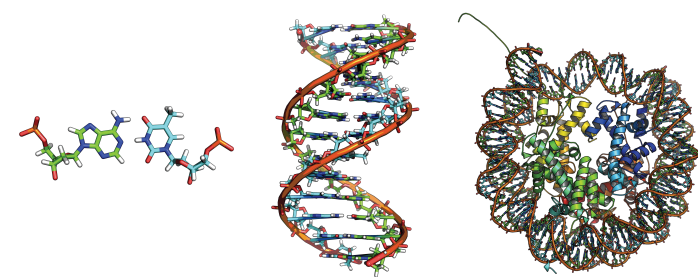
Simple genomes



Complex genomes

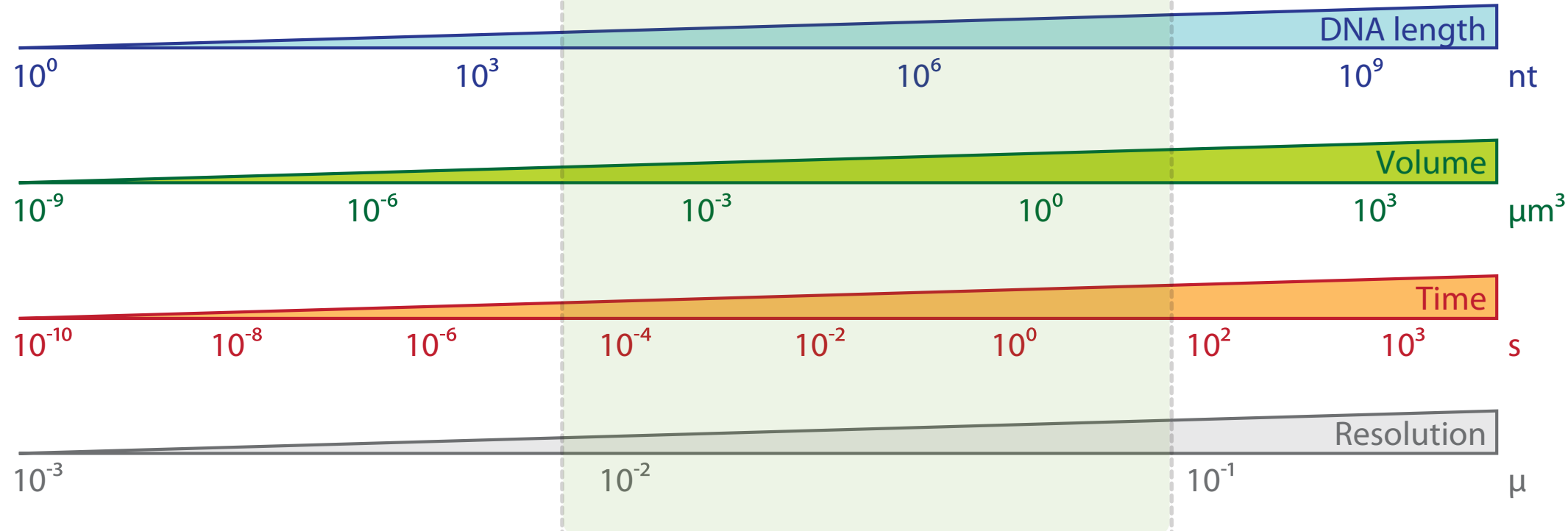
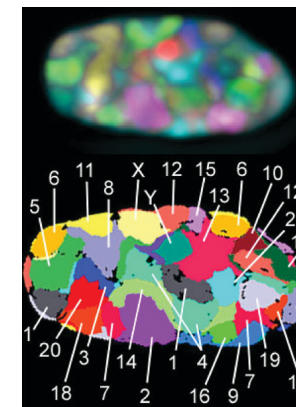
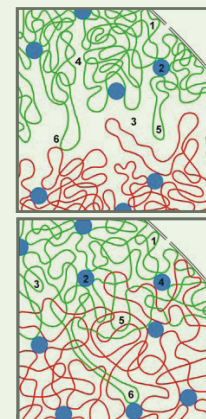


Knowledge



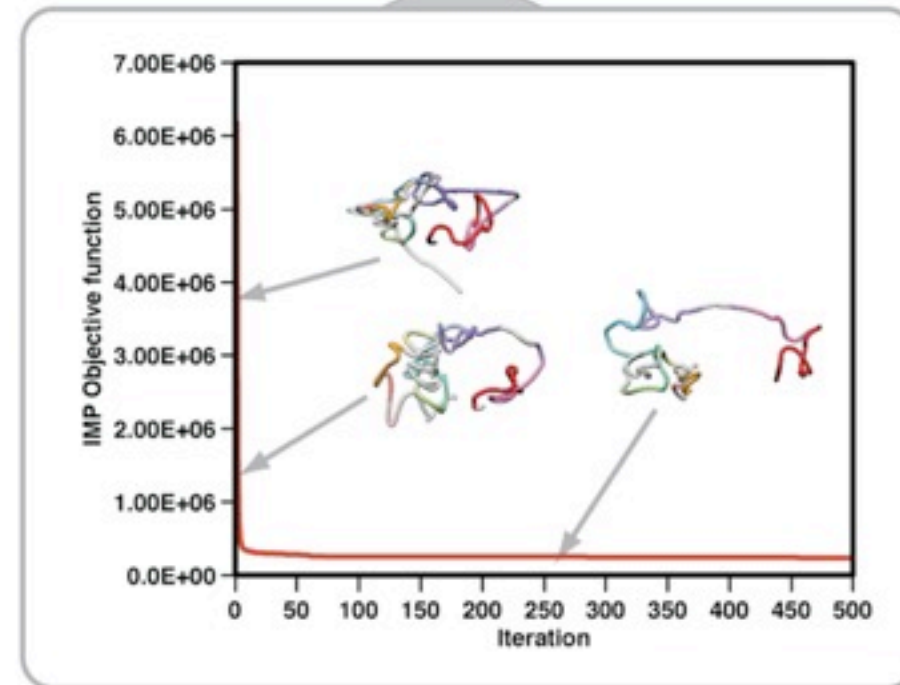
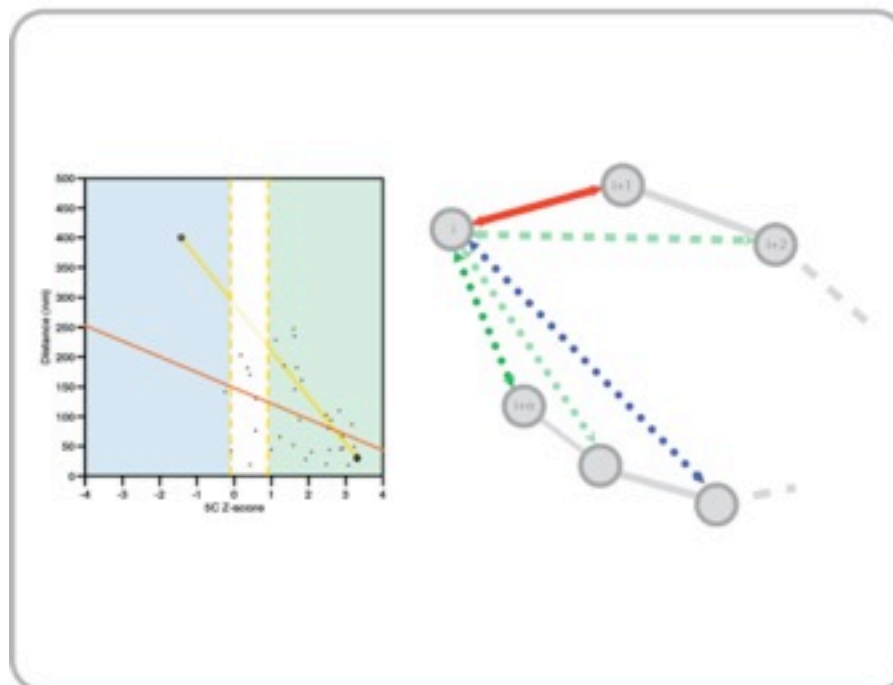
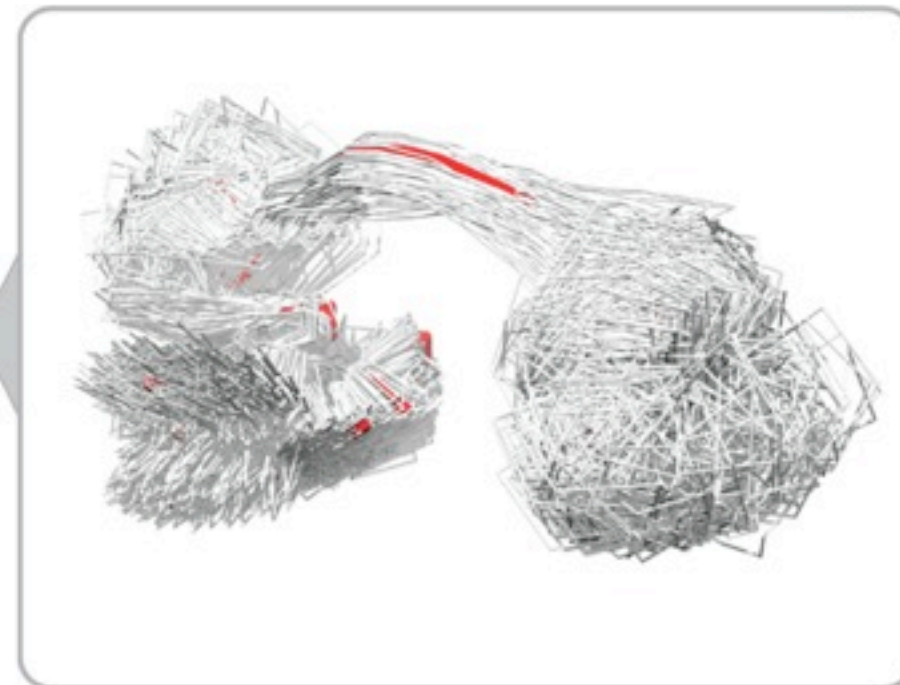
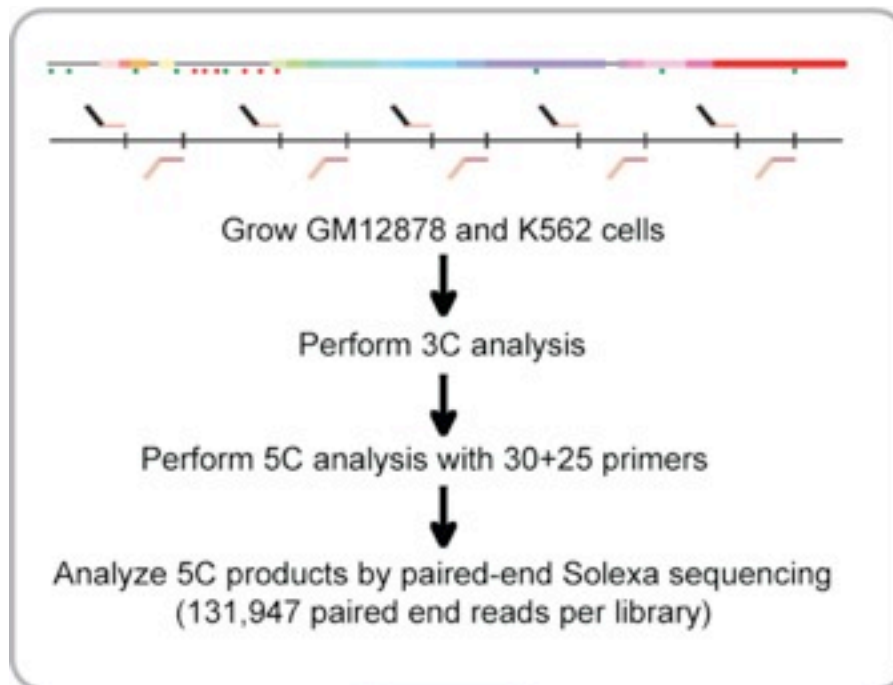
IDM

INM

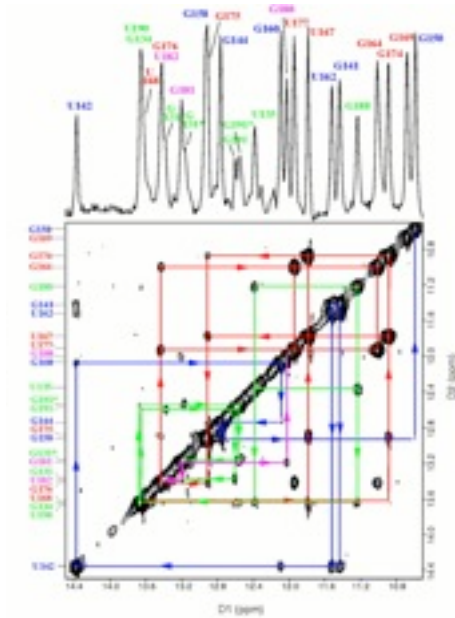
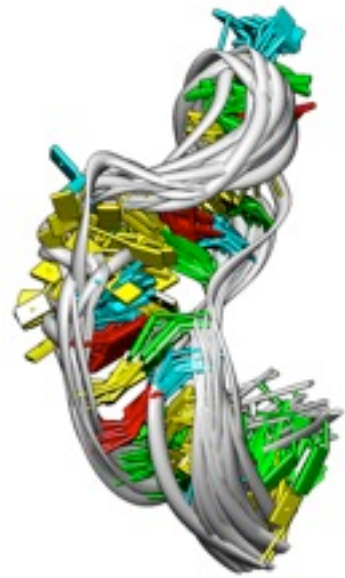


Adapted from:
Langowski and Heermann. *Semin Cell Dev Biol* (2007) vol. 18 (5) pp. 659-67

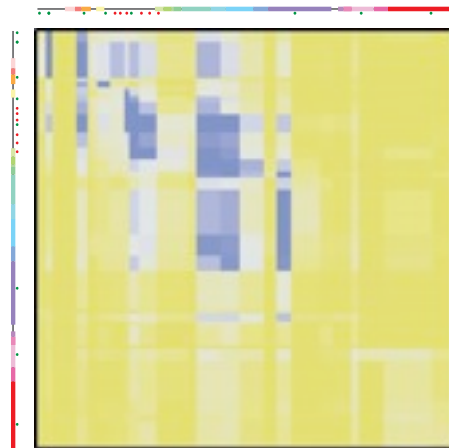
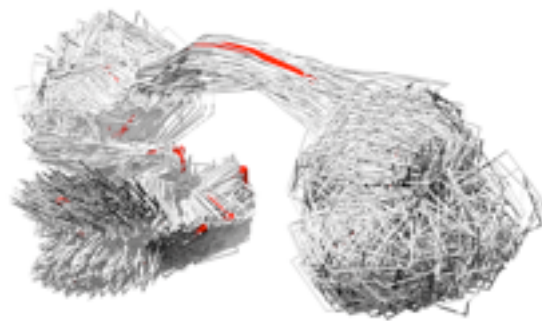
Experiments



Computation



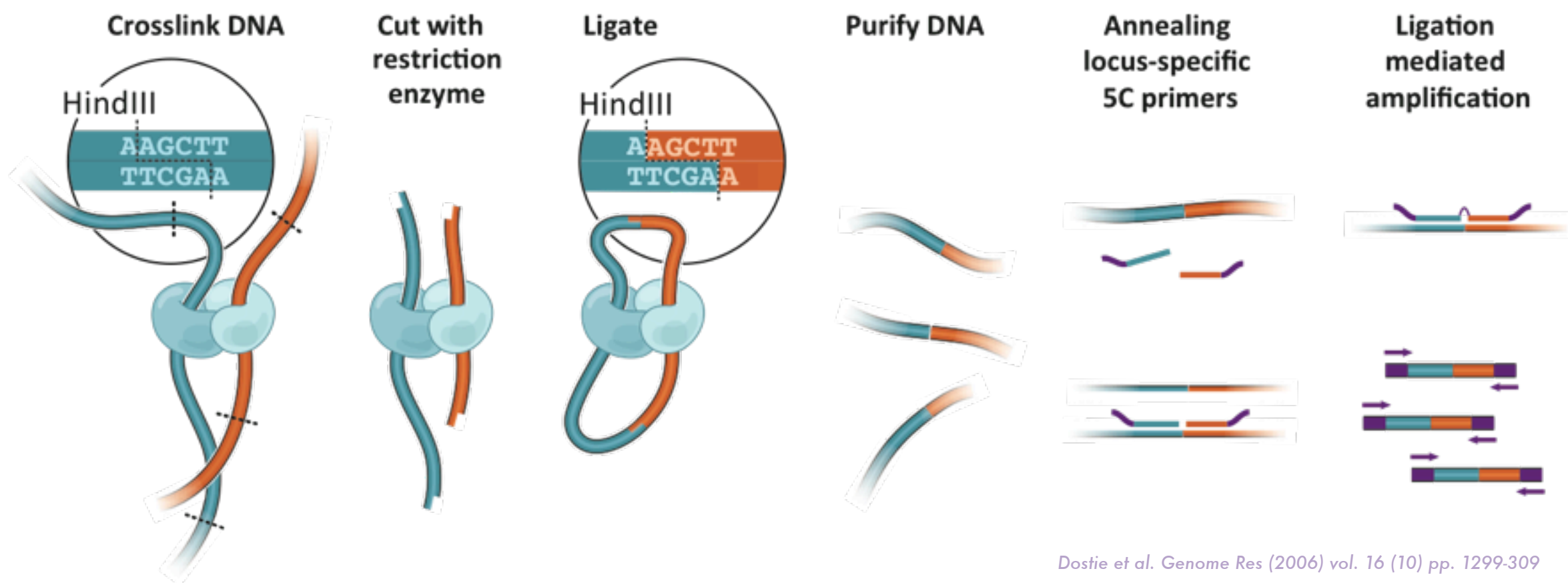
Biomolecular structure determination 2D-NOESY data



Chromosome structure determination 5C data

5C technology

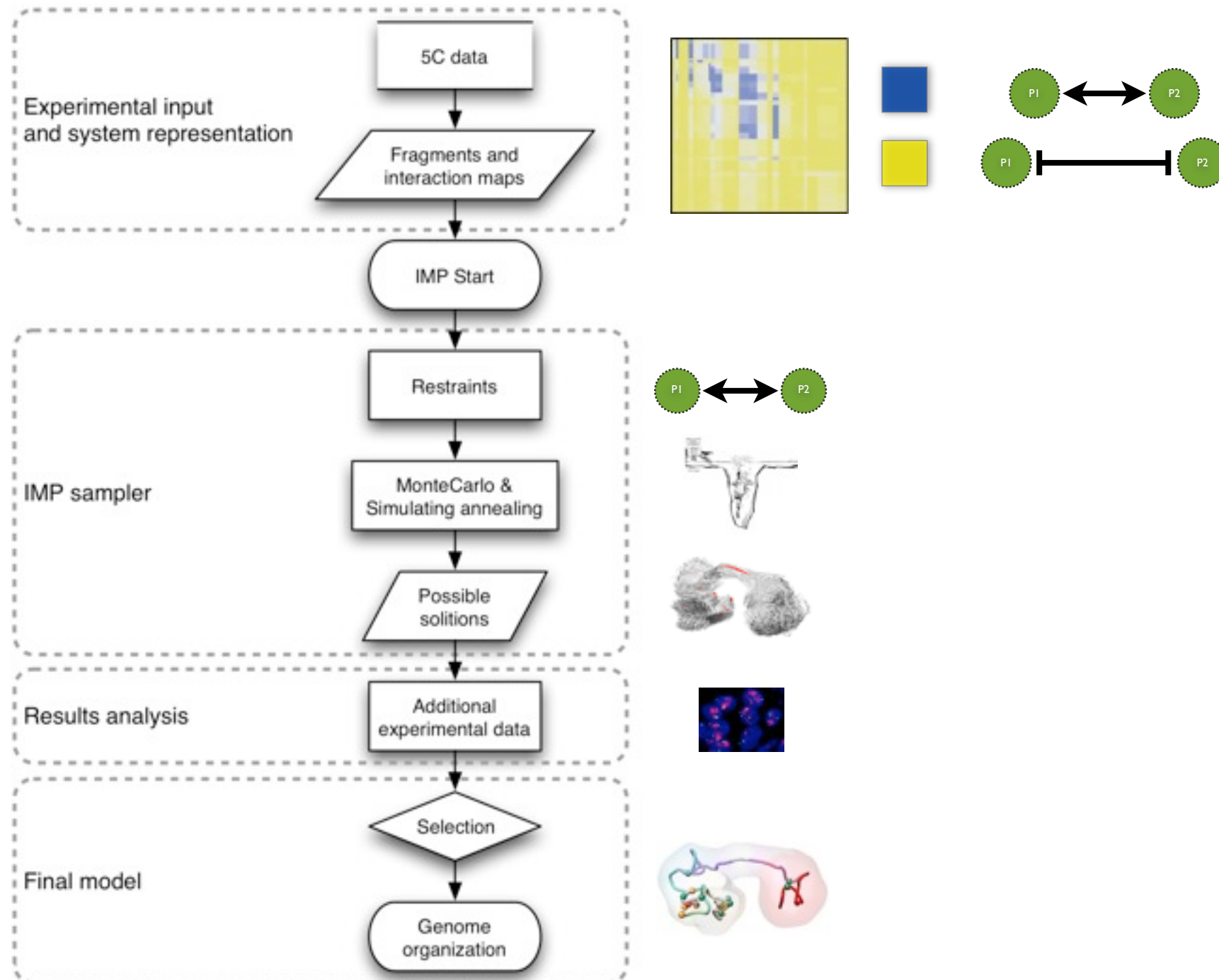
<http://my5C.umassmed.edu>



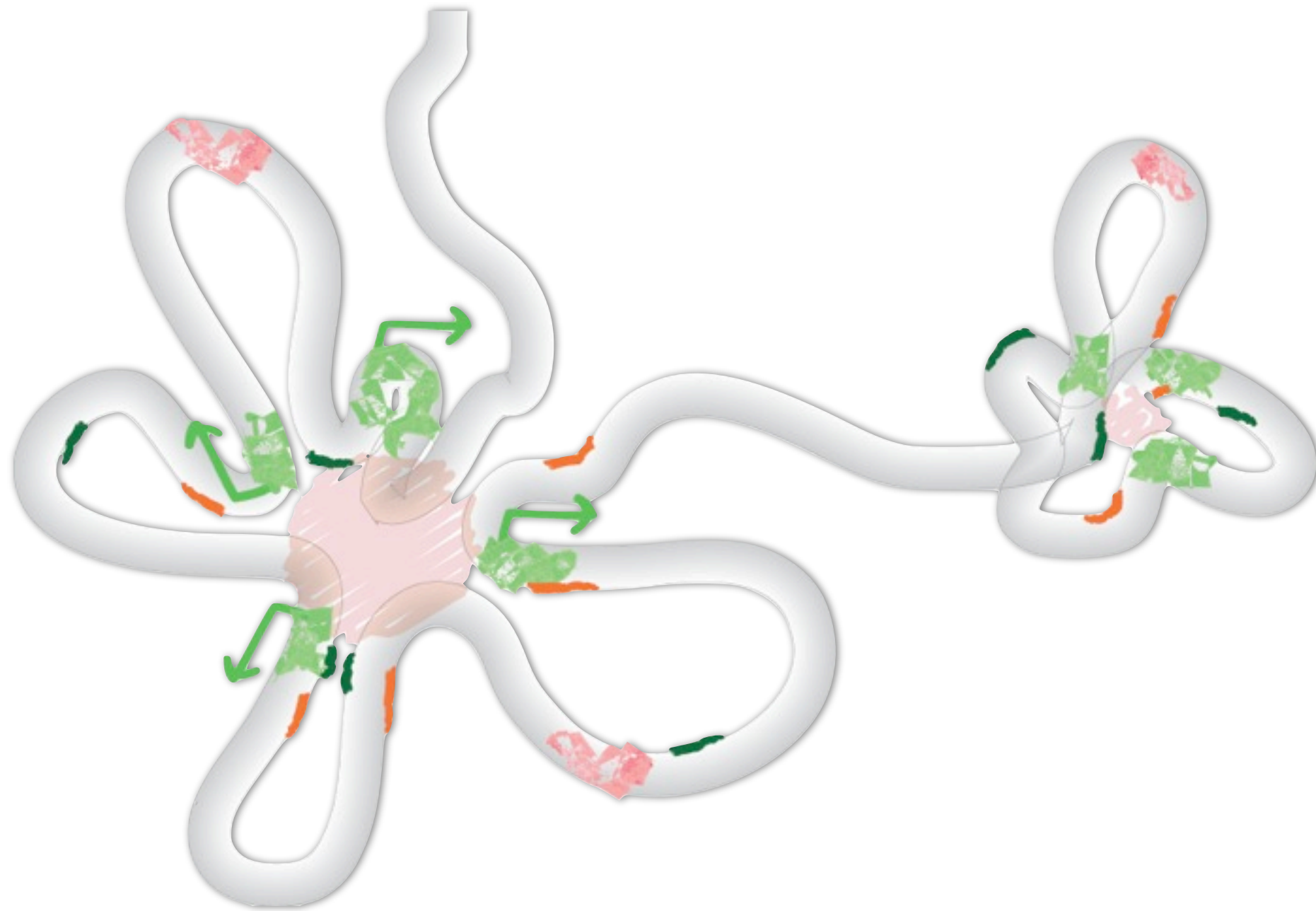
Dostie et al. *Genome Res* (2006) vol. 16 (10) pp. 1299-309

Integrative Modeling

<http://www.integrativemodeling.org>



Human α -globin domain



Human α -globin domain

ENm008 genomic structure and environment

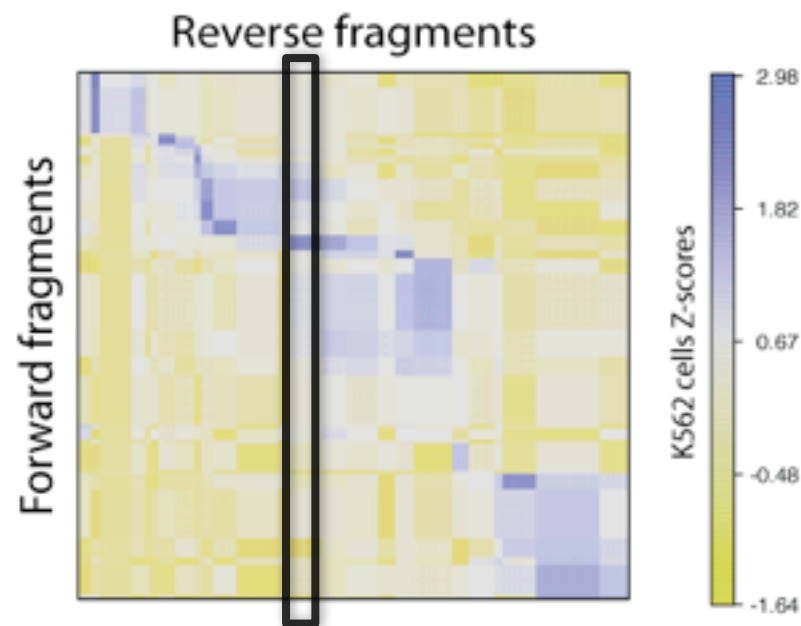
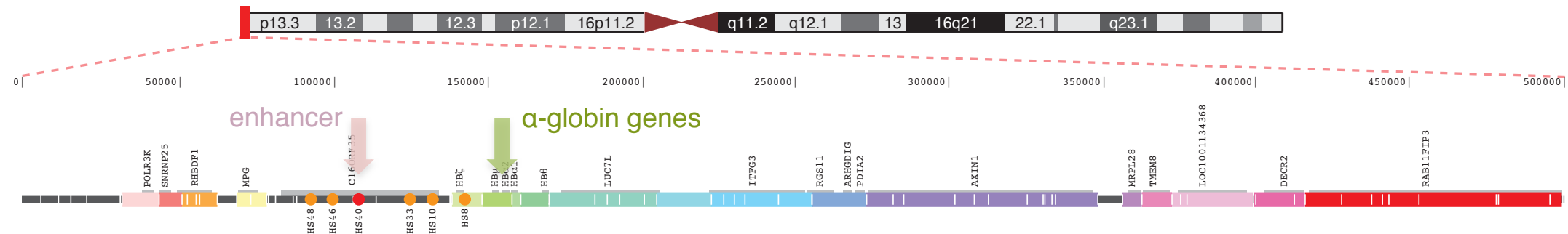


The ENCODE data for ENm008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute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

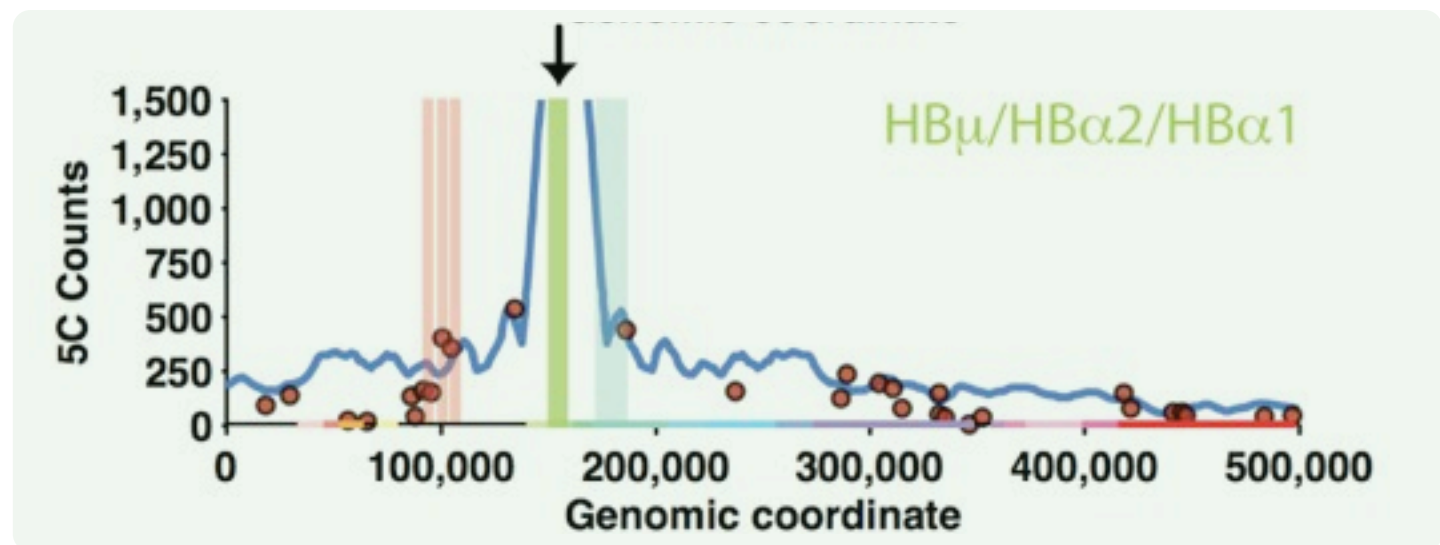
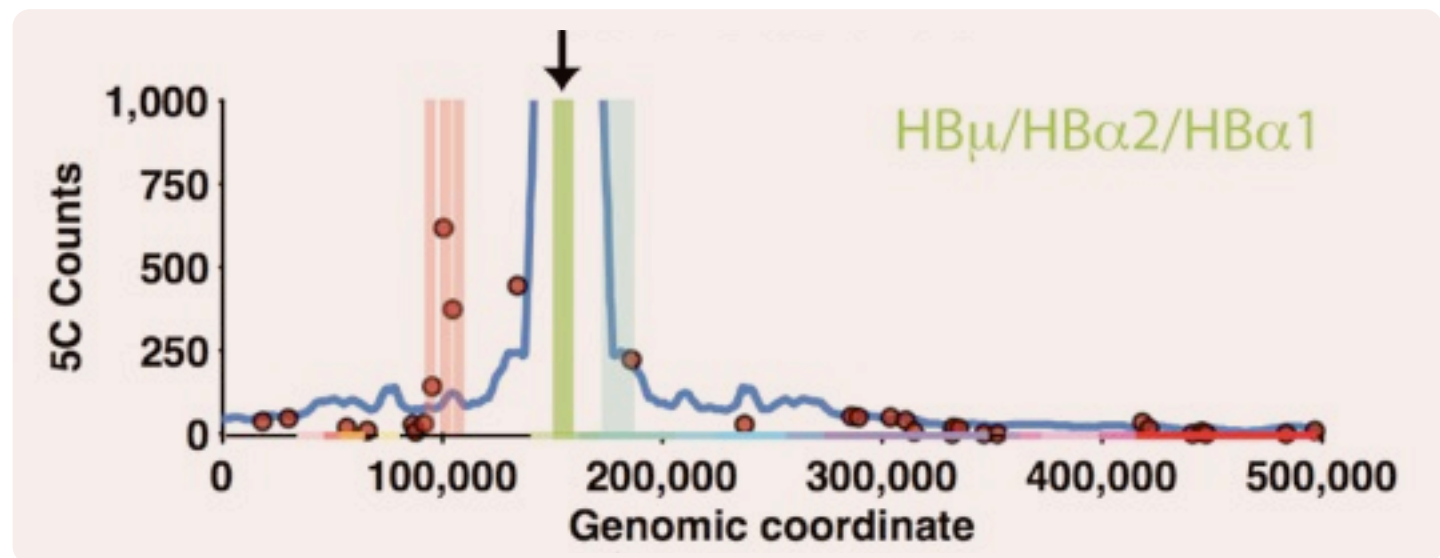
ENCODE Consortium. *Nature* (2007) vol. 447 (7146) pp. 799-816

Human α -globin domain

ENm008 genomic structure and environment



K562 cells:
 α -globin genes active



Representation

Harmonic

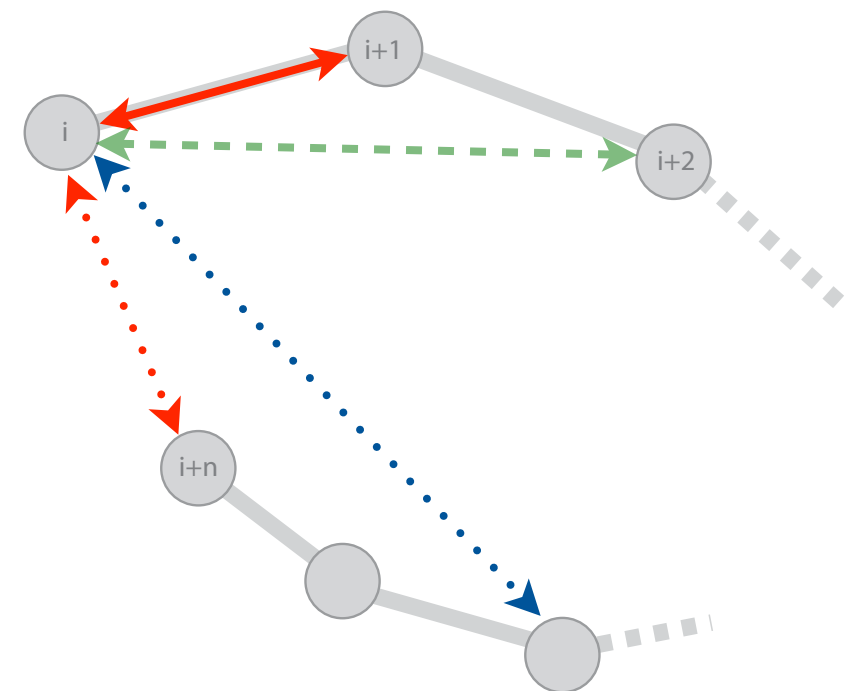
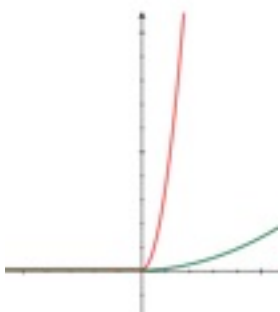
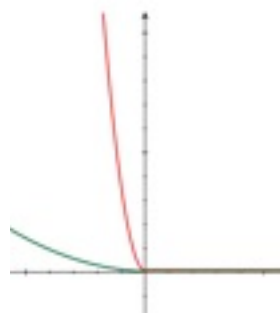
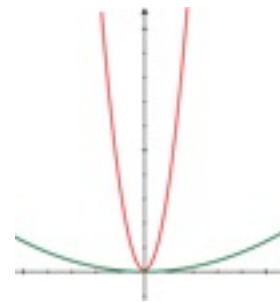
$$H_{i,j} = k(d_{i,j} - d_{i,j}^0)^2$$

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

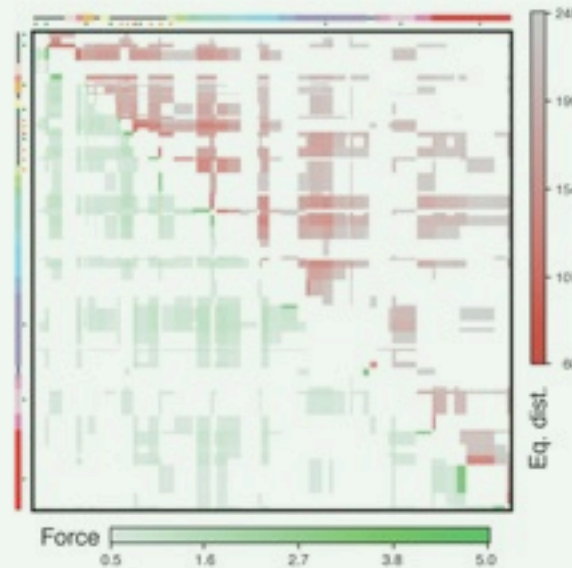
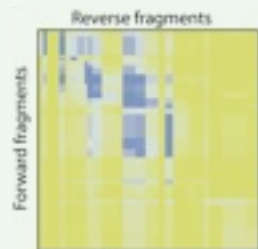
$$\begin{cases} \text{if } d_{i,j} \geq d_{i,j}^0; & ubH_{i,j} = k(d_{i,j} - d_{i,j}^0)^2 \\ \text{if } d_{i,j} < d_{i,j}^0; & ubH_{i,j} = 0 \end{cases}$$



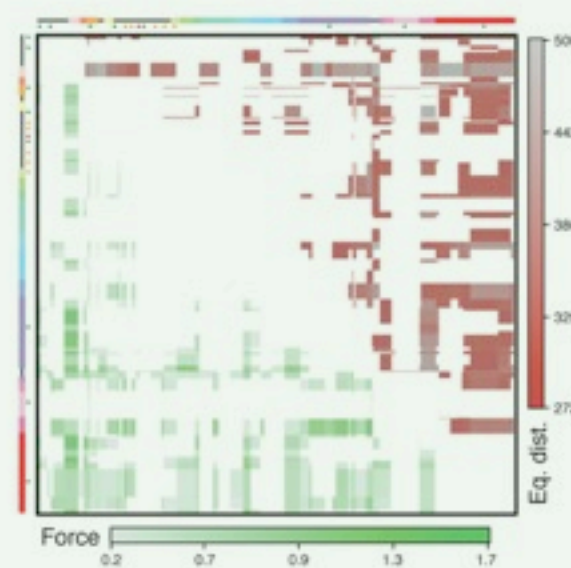
Scoring

GM12878

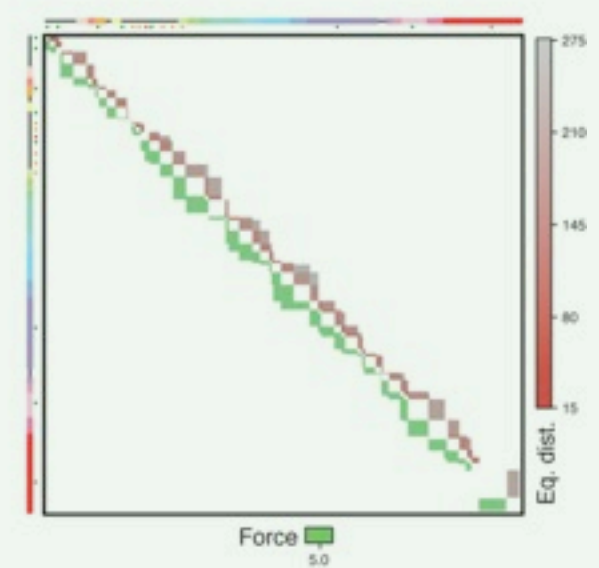
70 fragments
1,520 restraints



Harmonic



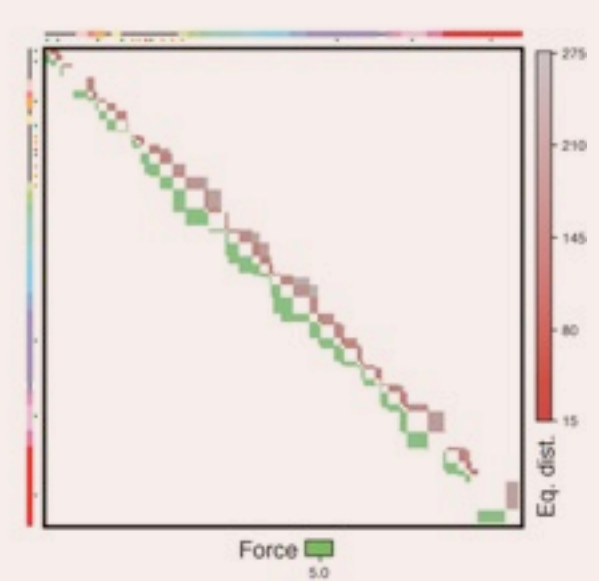
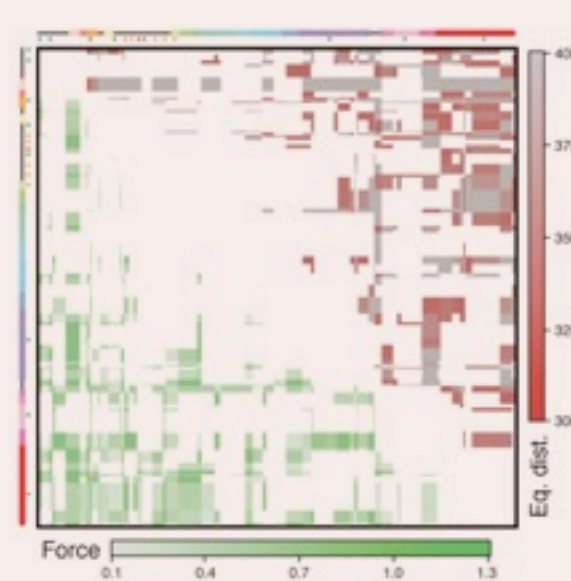
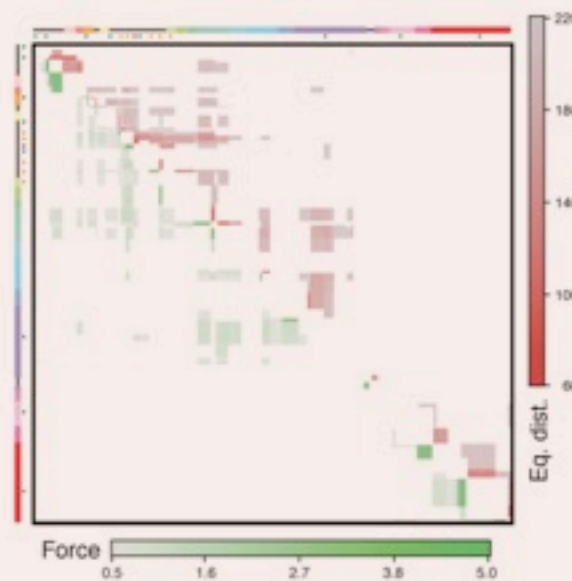
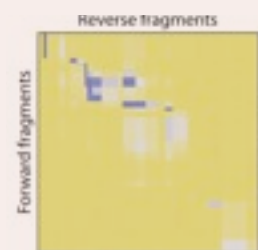
Harmonic Lower Bound



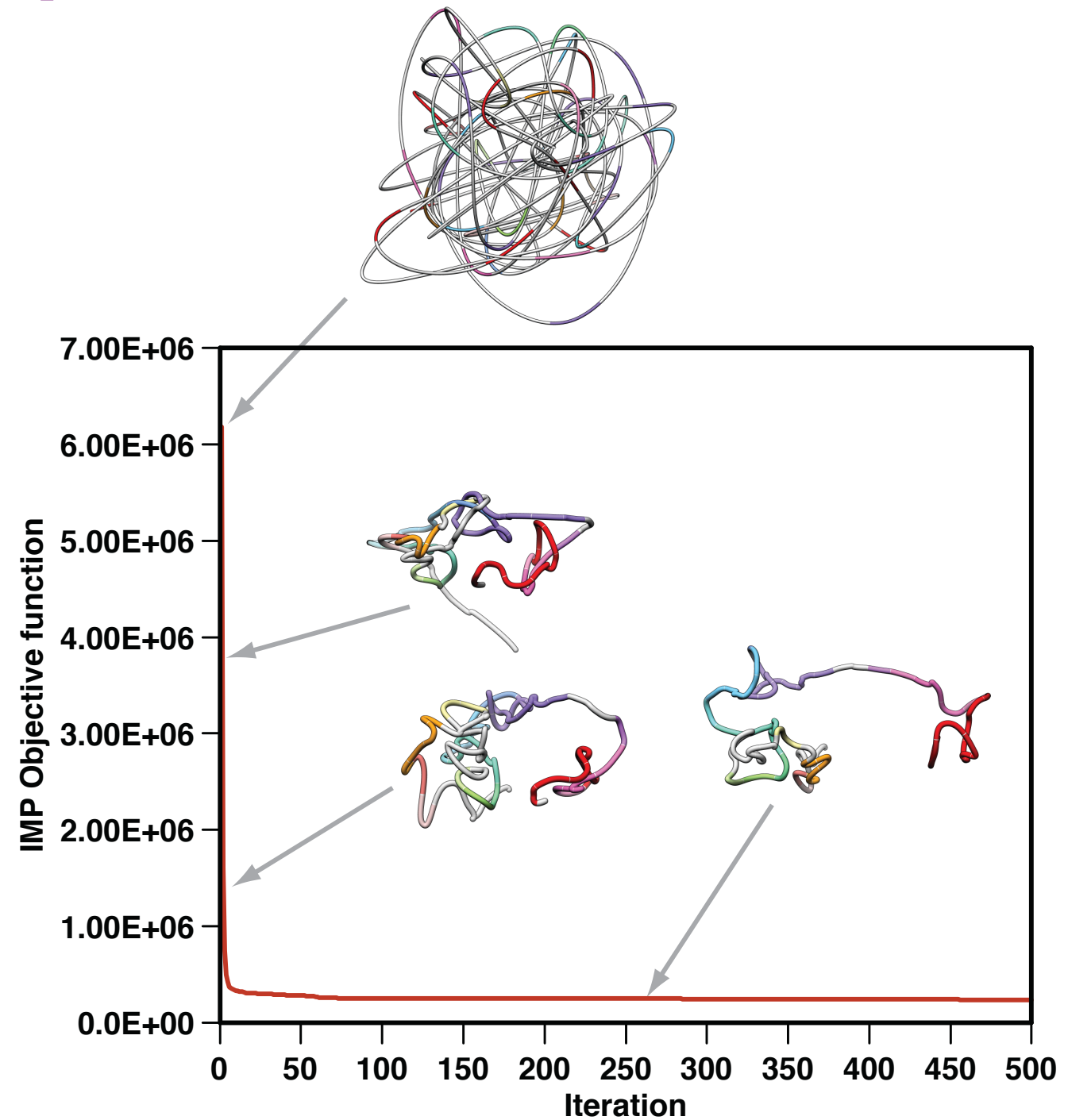
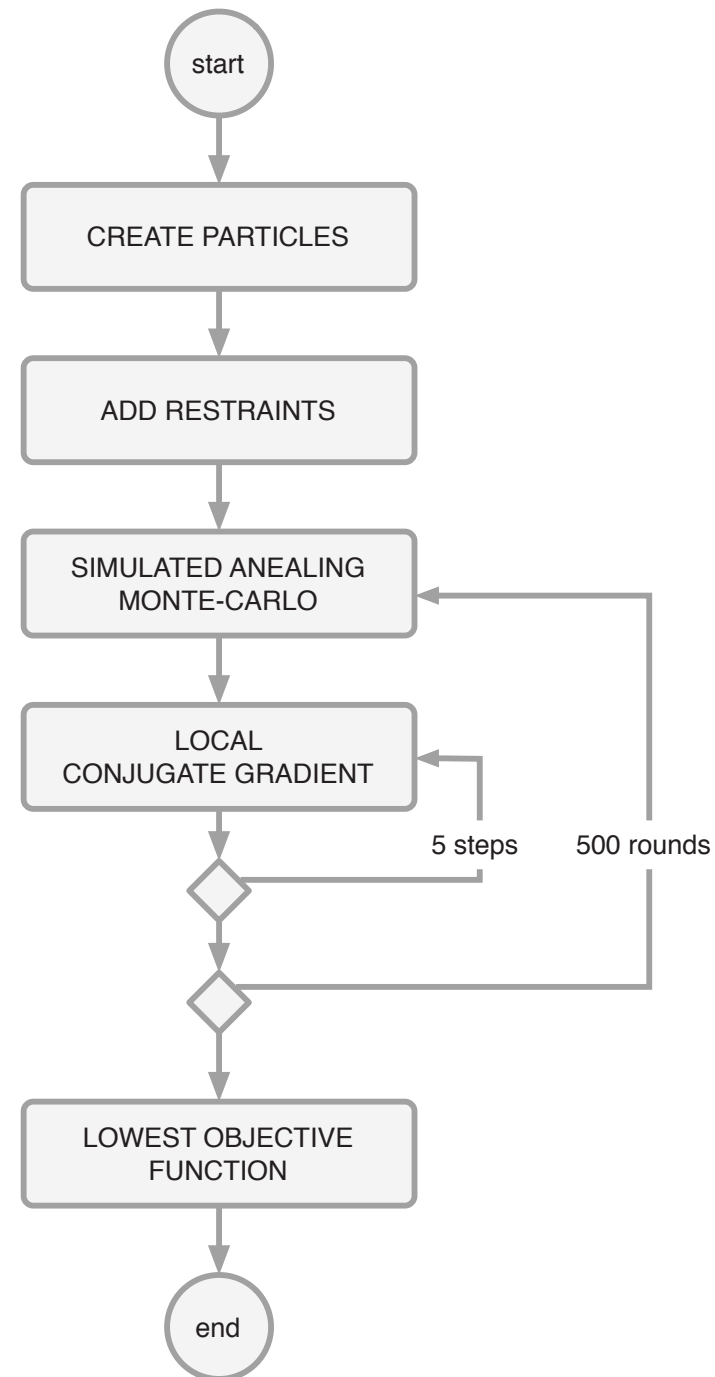
Harmonic Upper Bound

K562

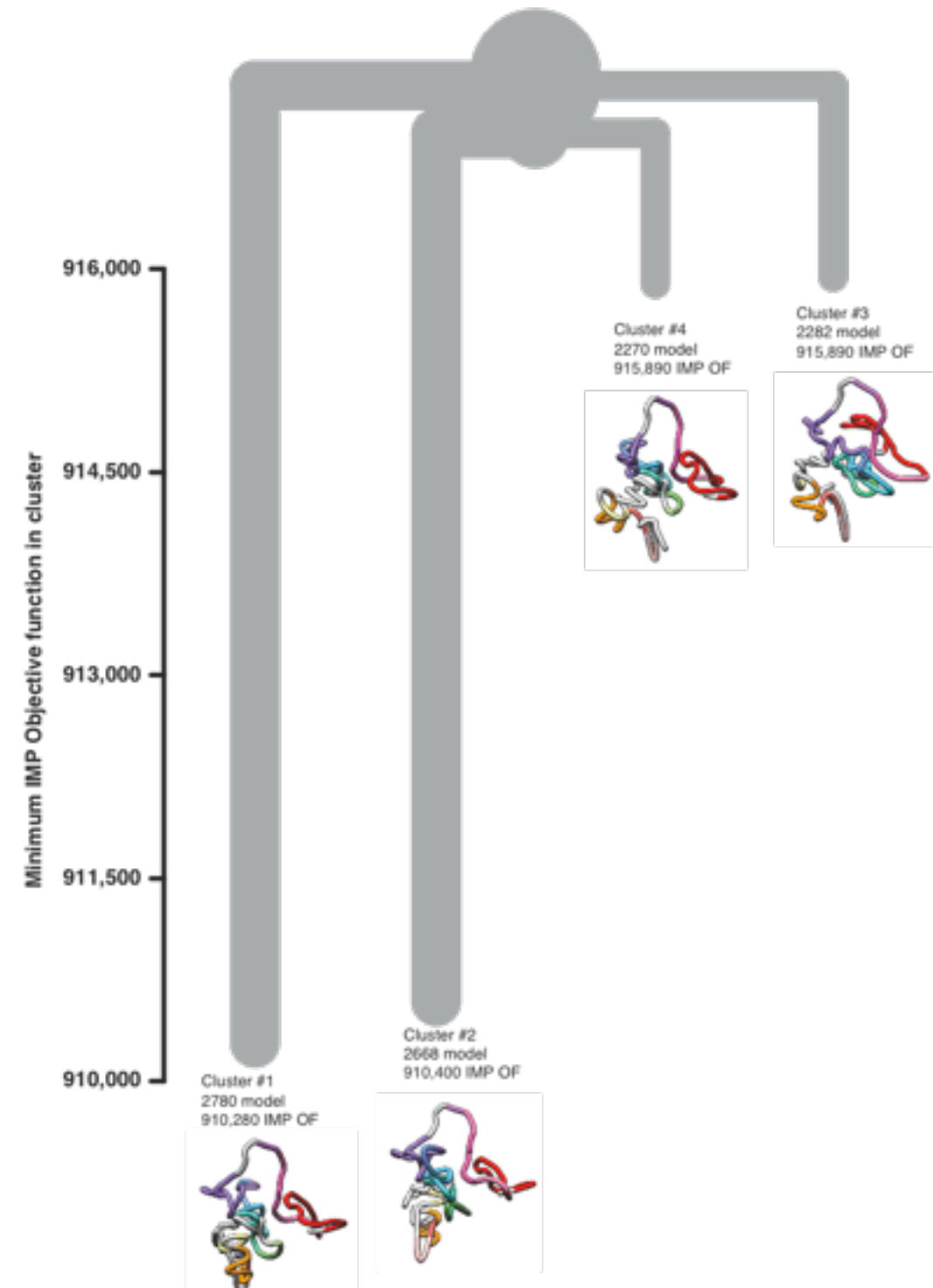
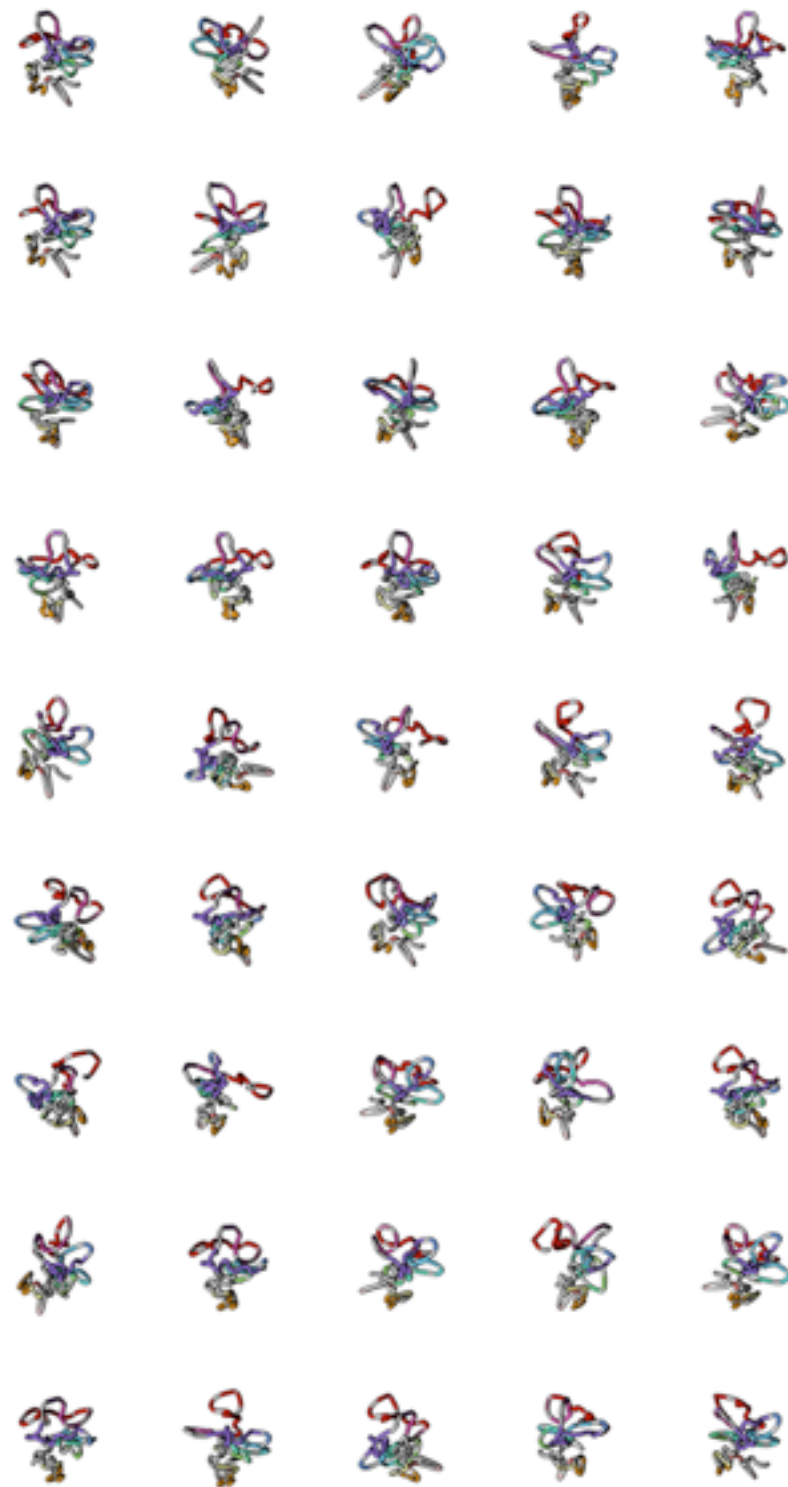
70 fragments
1,049 restraints



Optimization

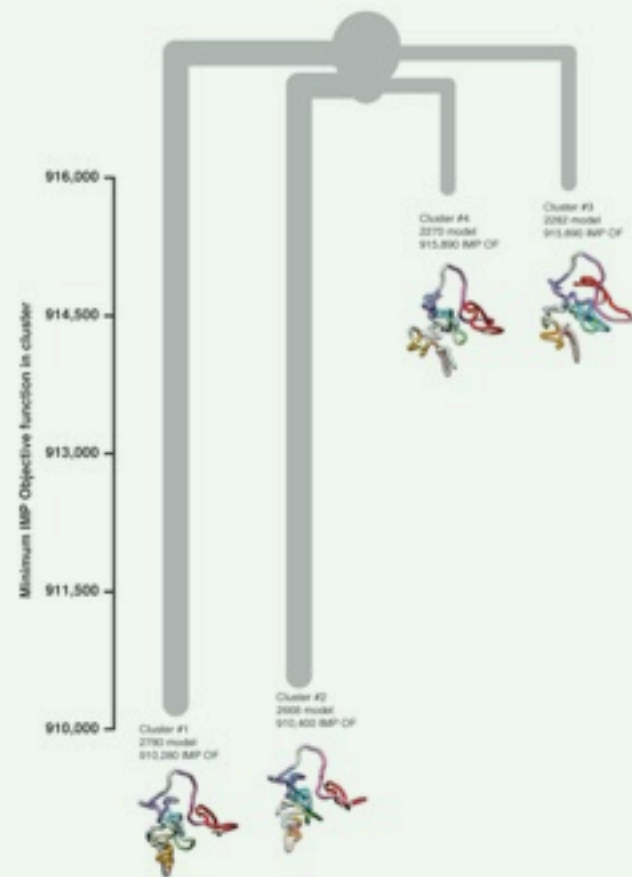
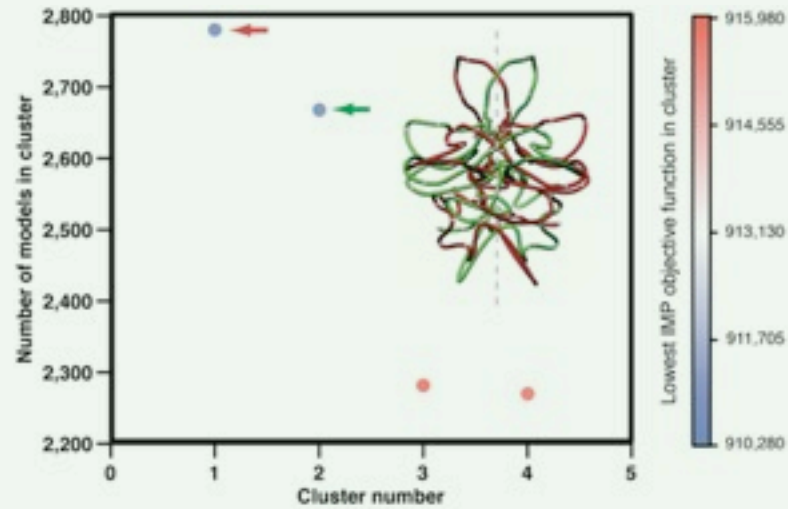


Clustering

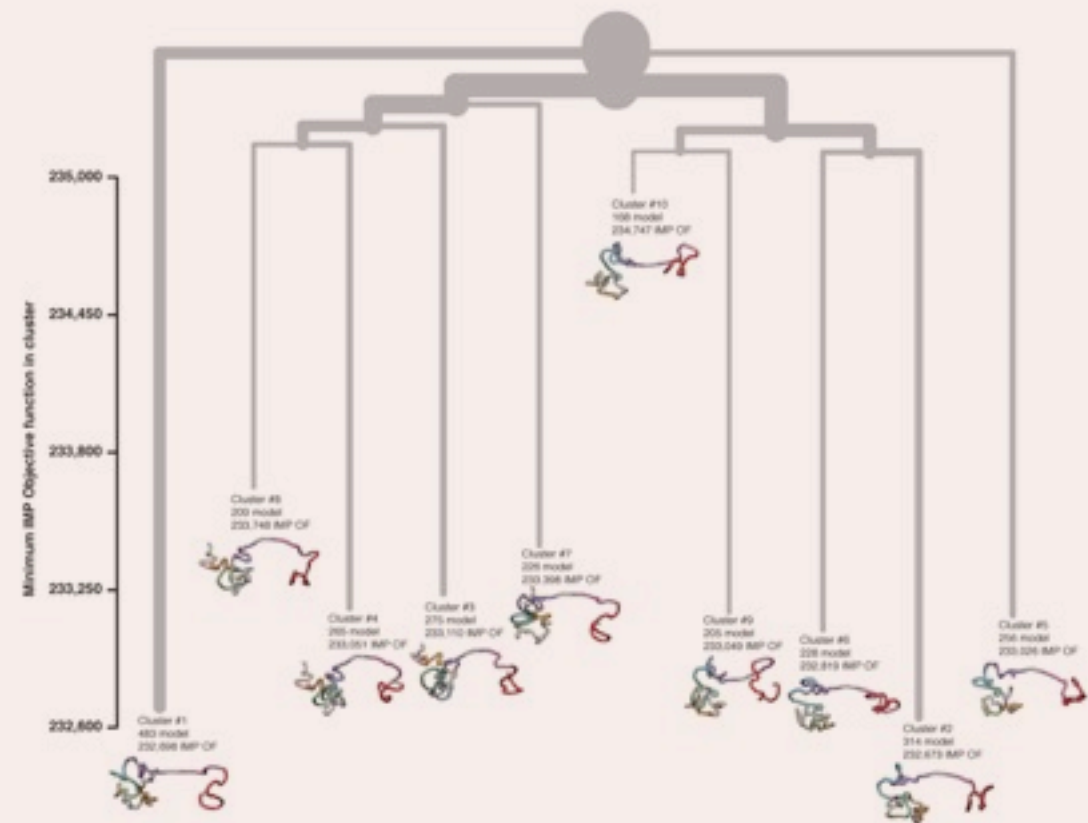
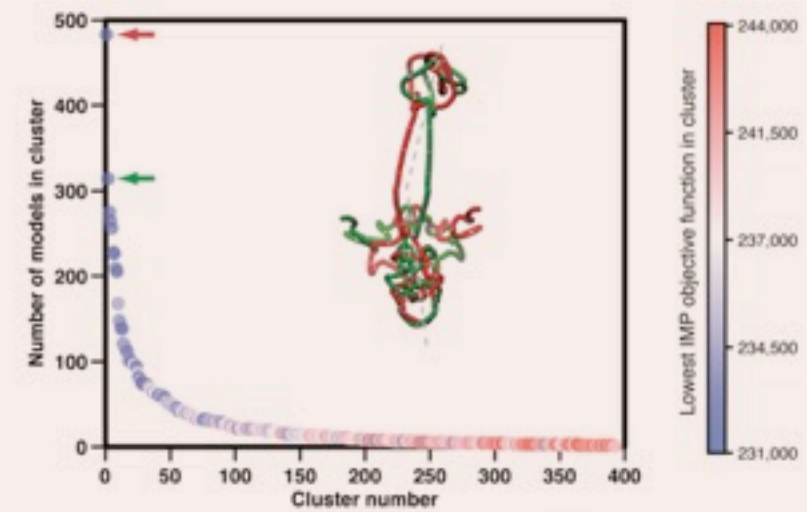


Not just one solution

GM12878



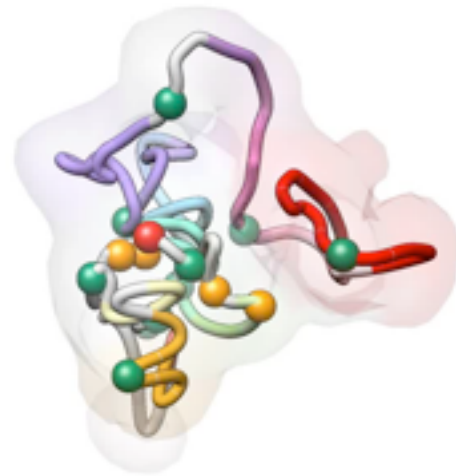
K562



Consistency

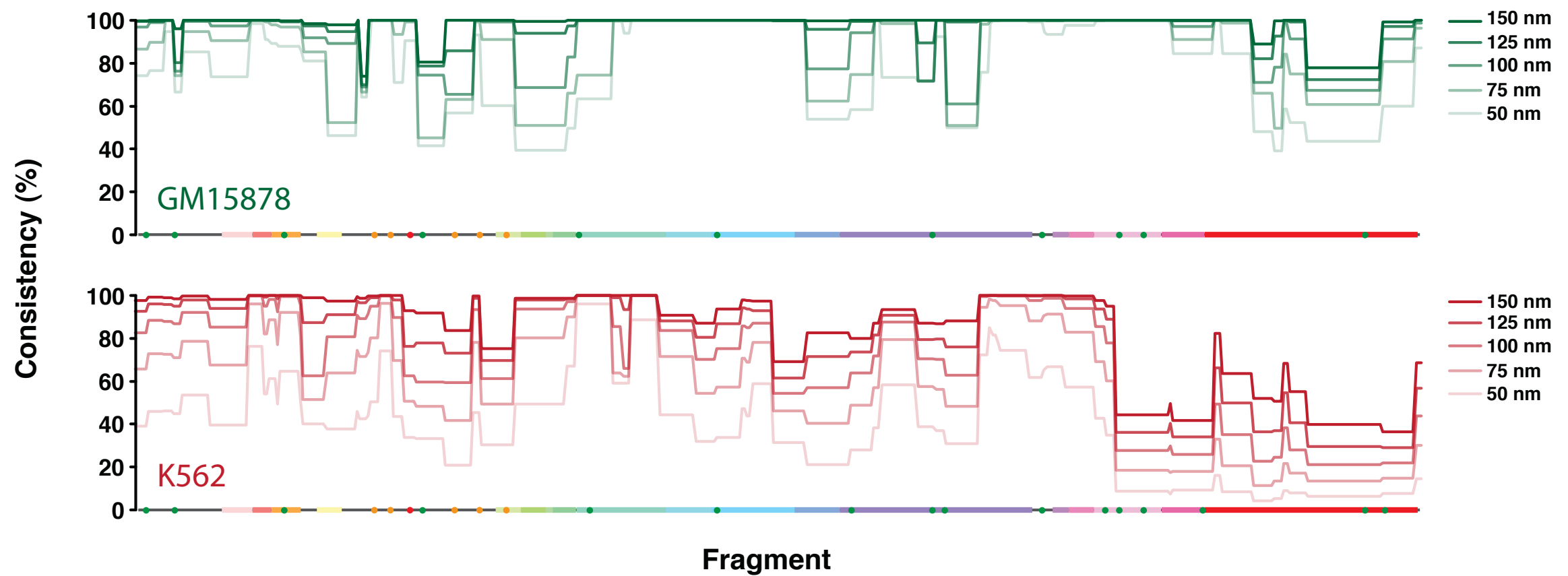
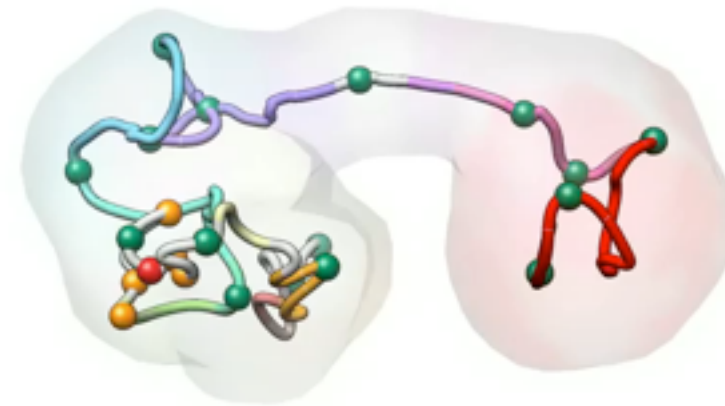
GM12878

Cluster #1
2780 model



K562

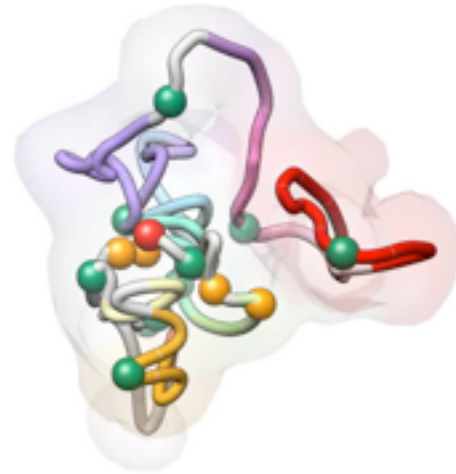
Cluster #2
314 model



Regulatory elements

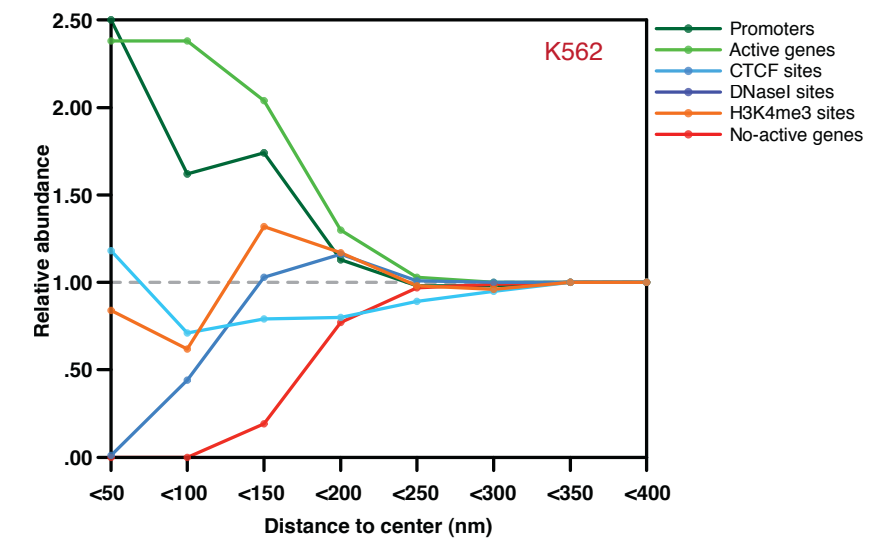
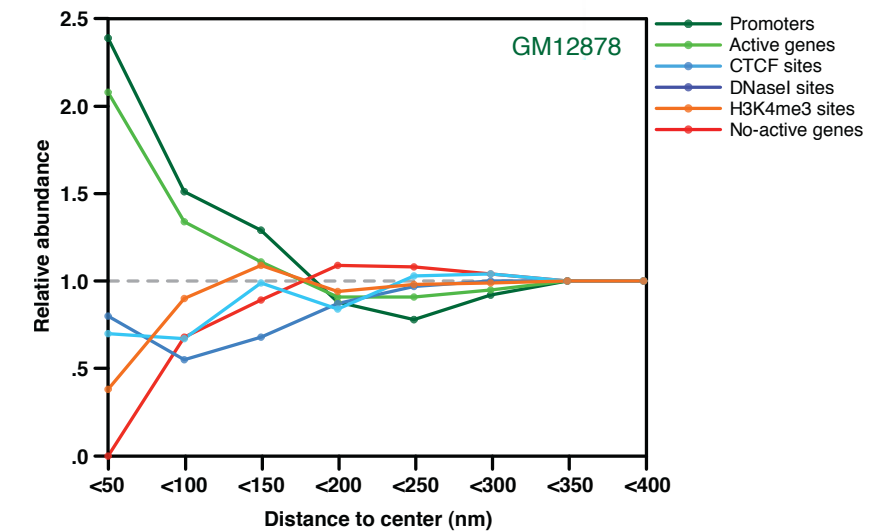
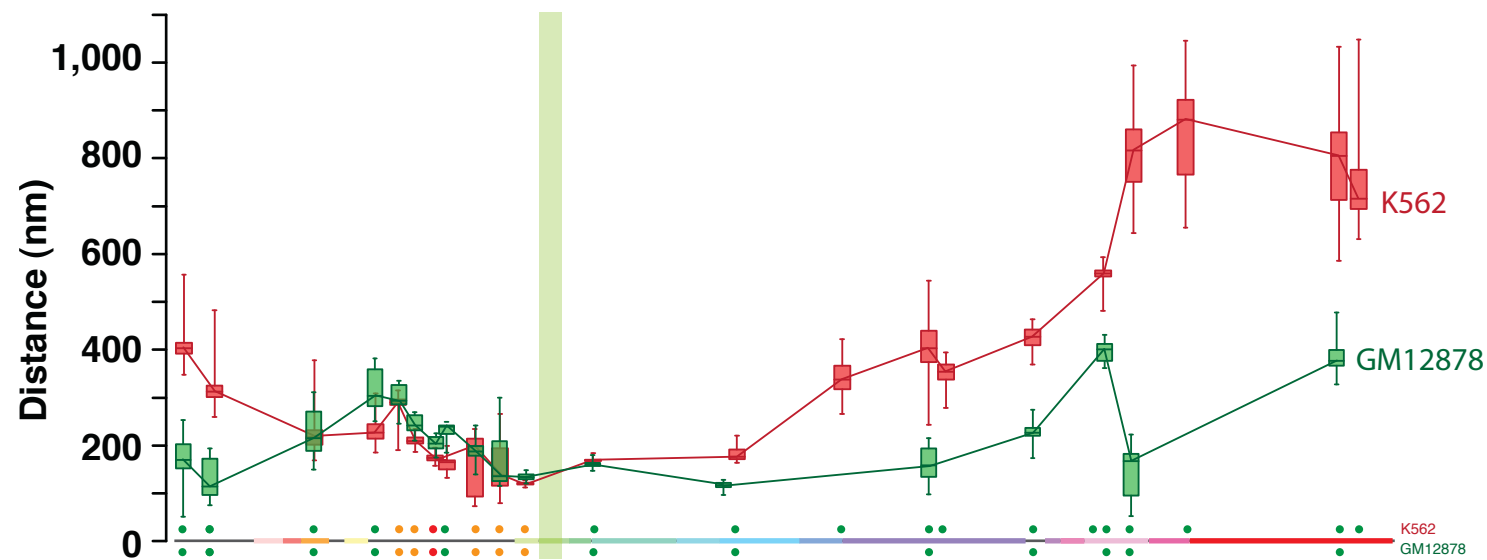
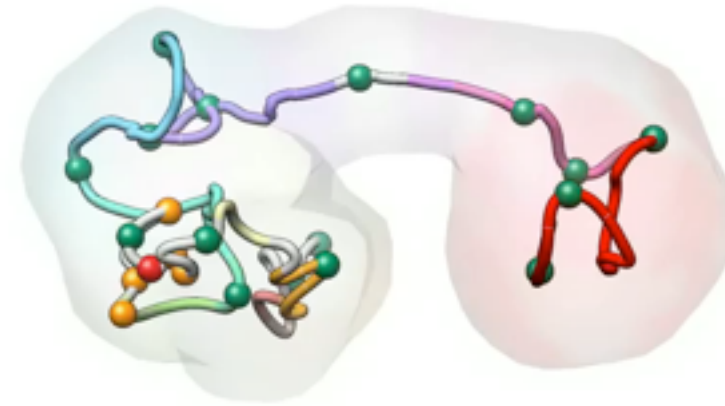
GM12878

Cluster #1
2780 model



K562

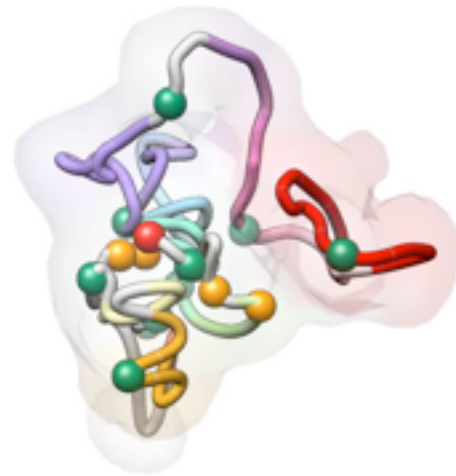
Cluster #2
314 model



Compactness

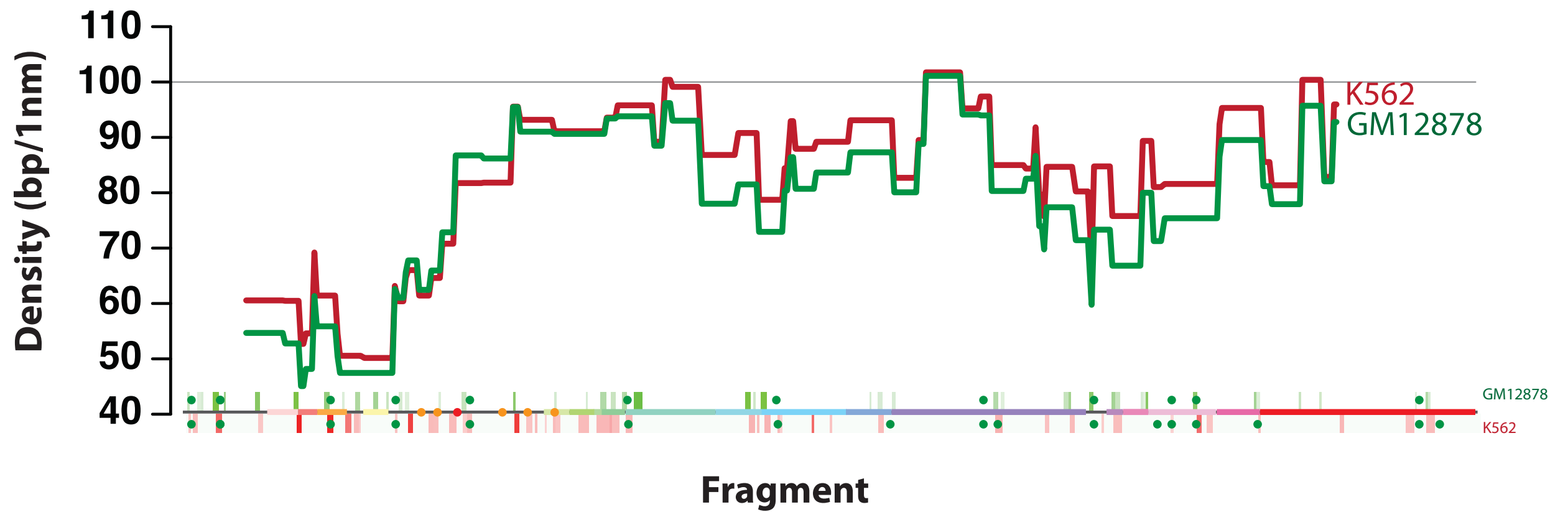
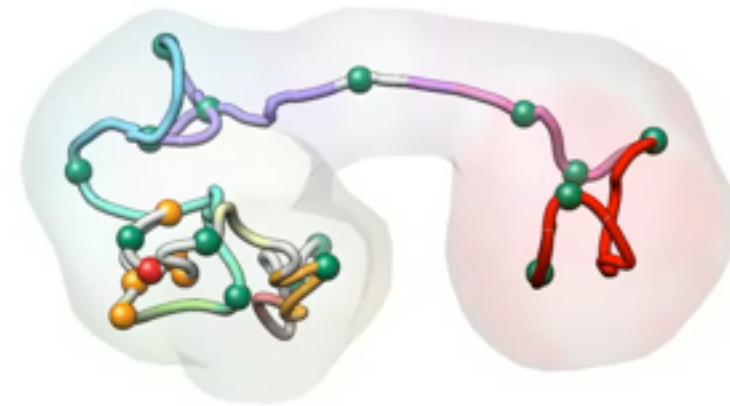
GM12878

Cluster #1
2780 model



K562

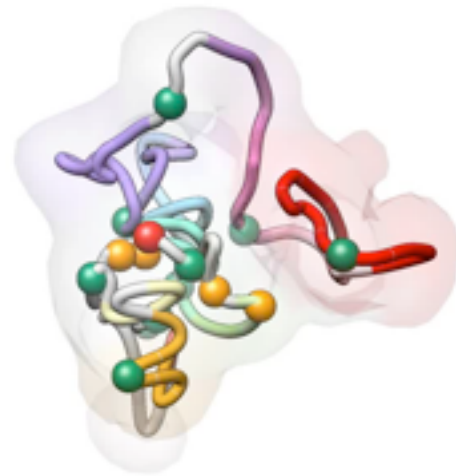
Cluster #2
314 model



Multi-loops

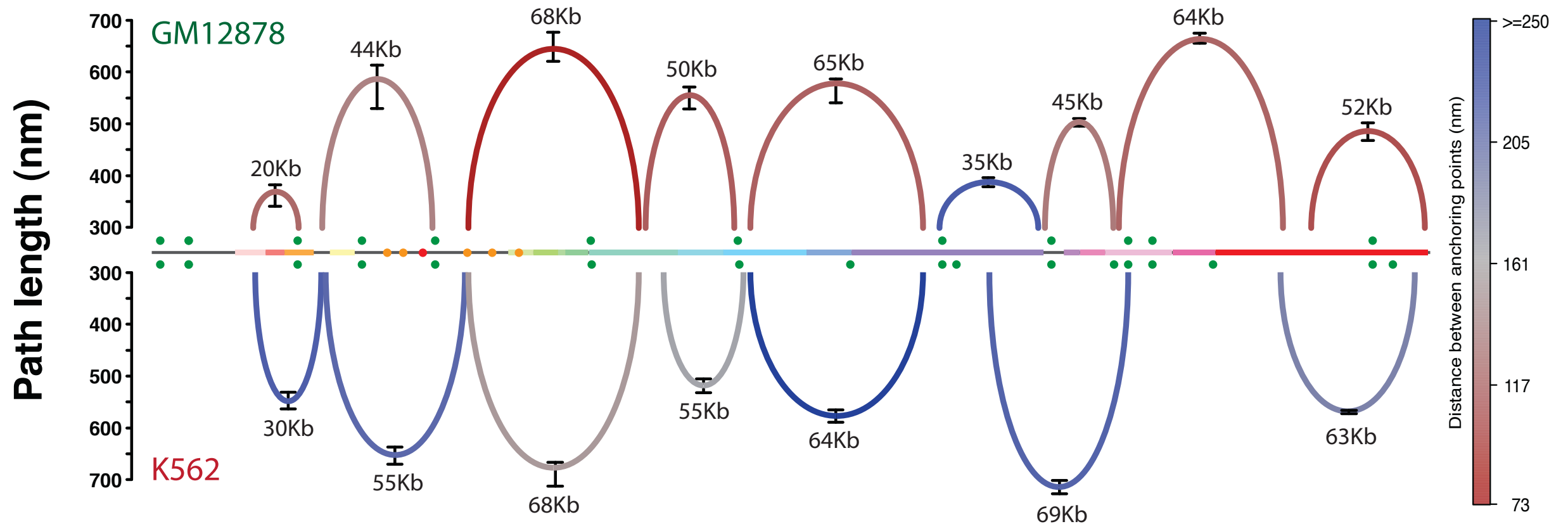
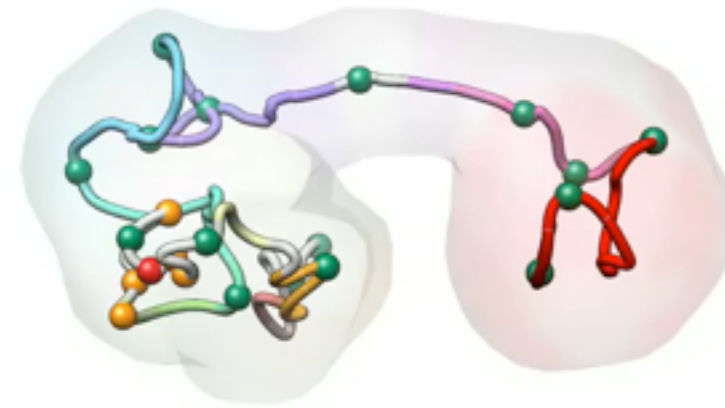
GM12878

Cluster #1
2780 model



K562

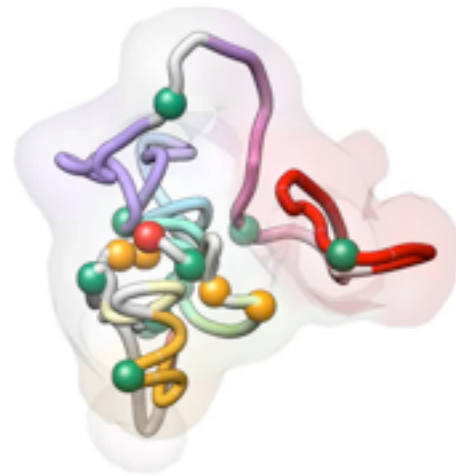
Cluster #2
314 model



Expression

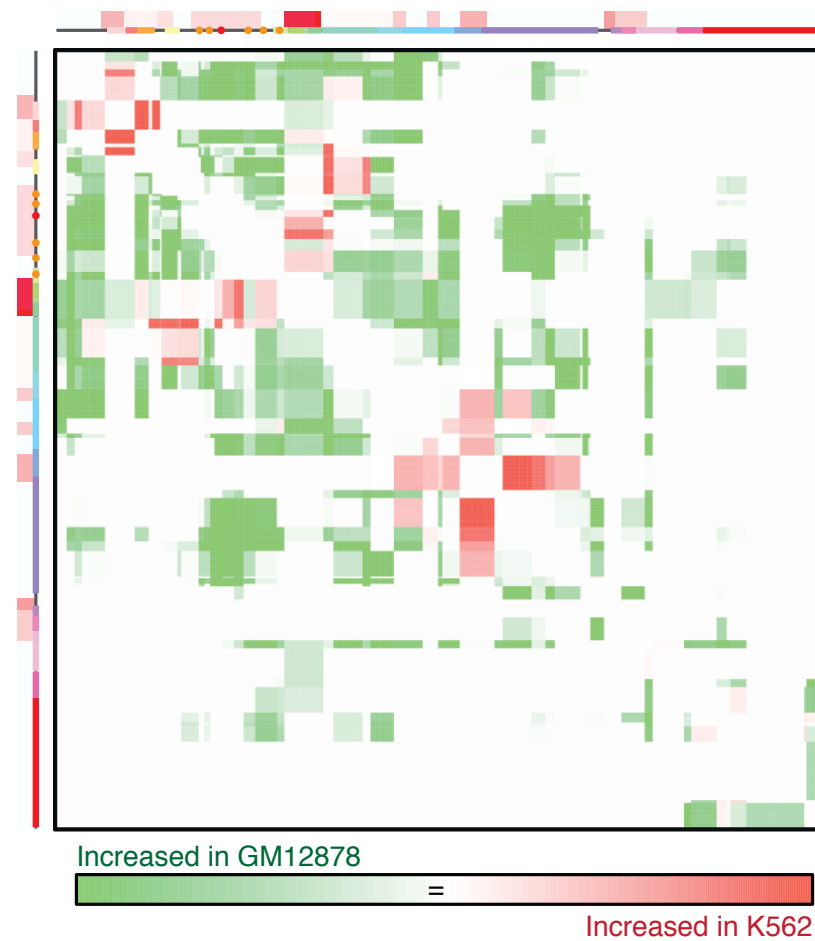
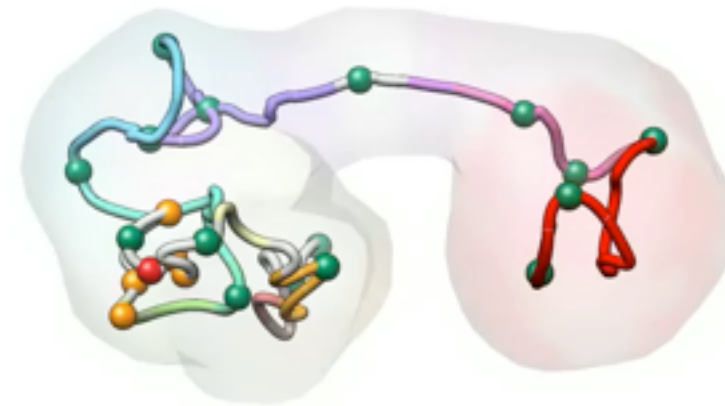
GM12878

Cluster #1
2780 model



K562

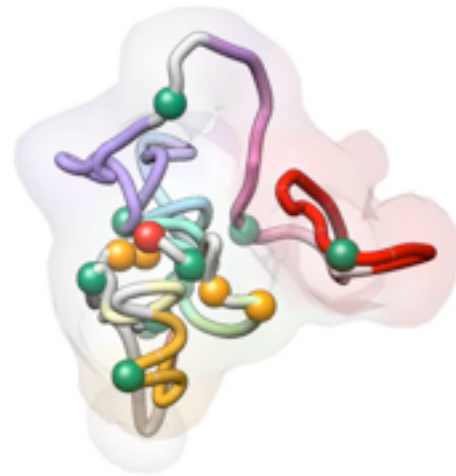
Cluster #2
314 model



FISH validation

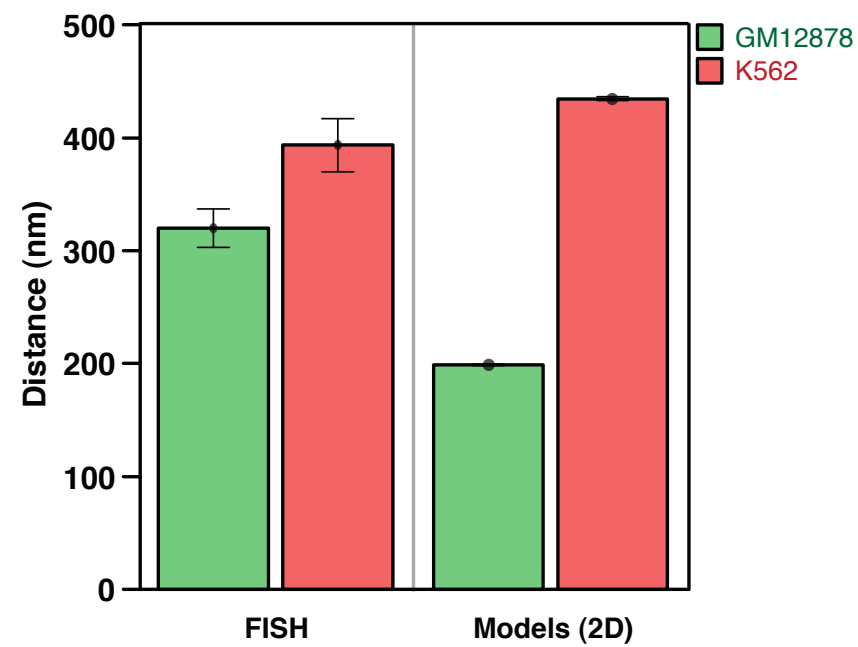
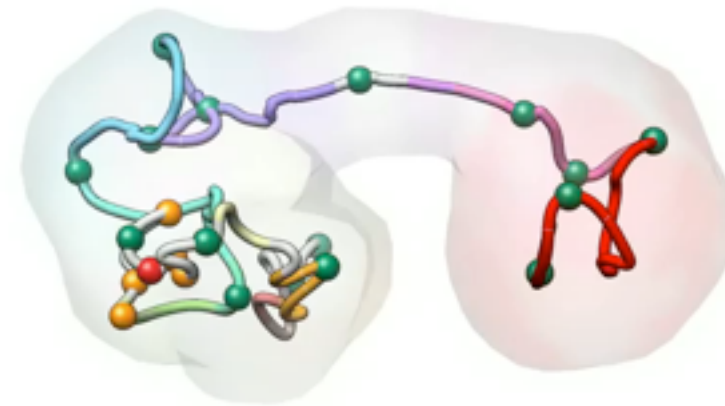
GM12878

Cluster #1
2780 model



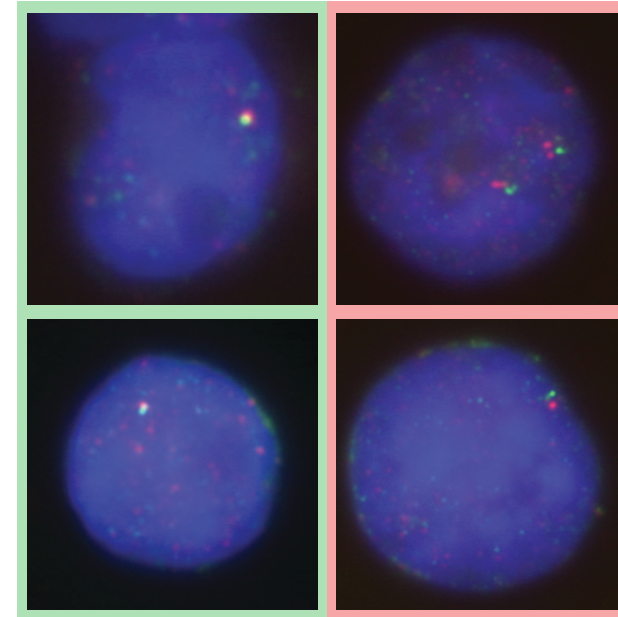
K562

Cluster #2
314 model

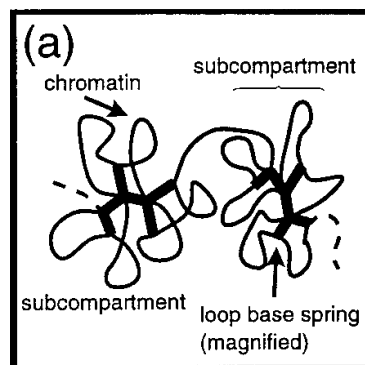


GM12878

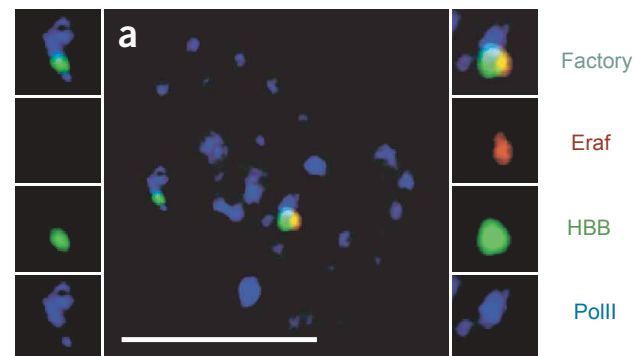
K562



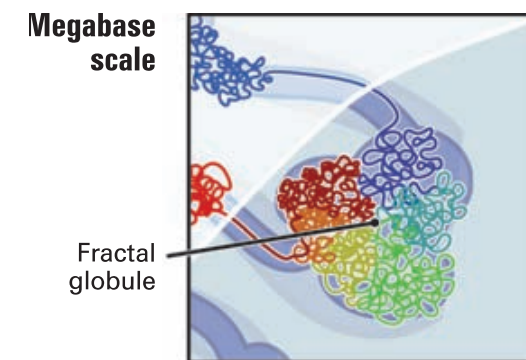
The “Chromatin Globule” model



Münkel et al. JMB (1999)



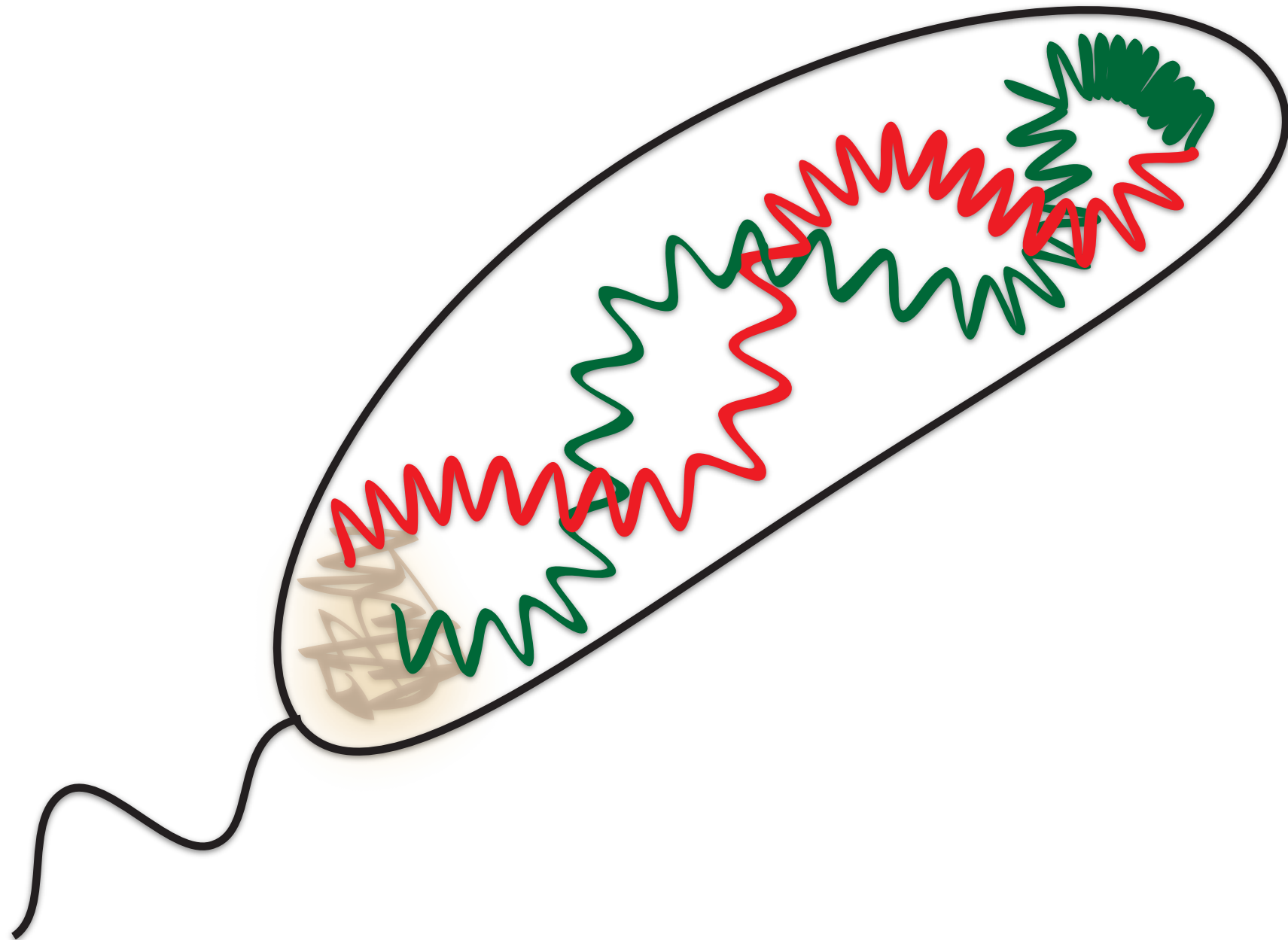
Osborne et al. Nat Genet (2004)



Lieberman-Aiden et al. Science (2009)

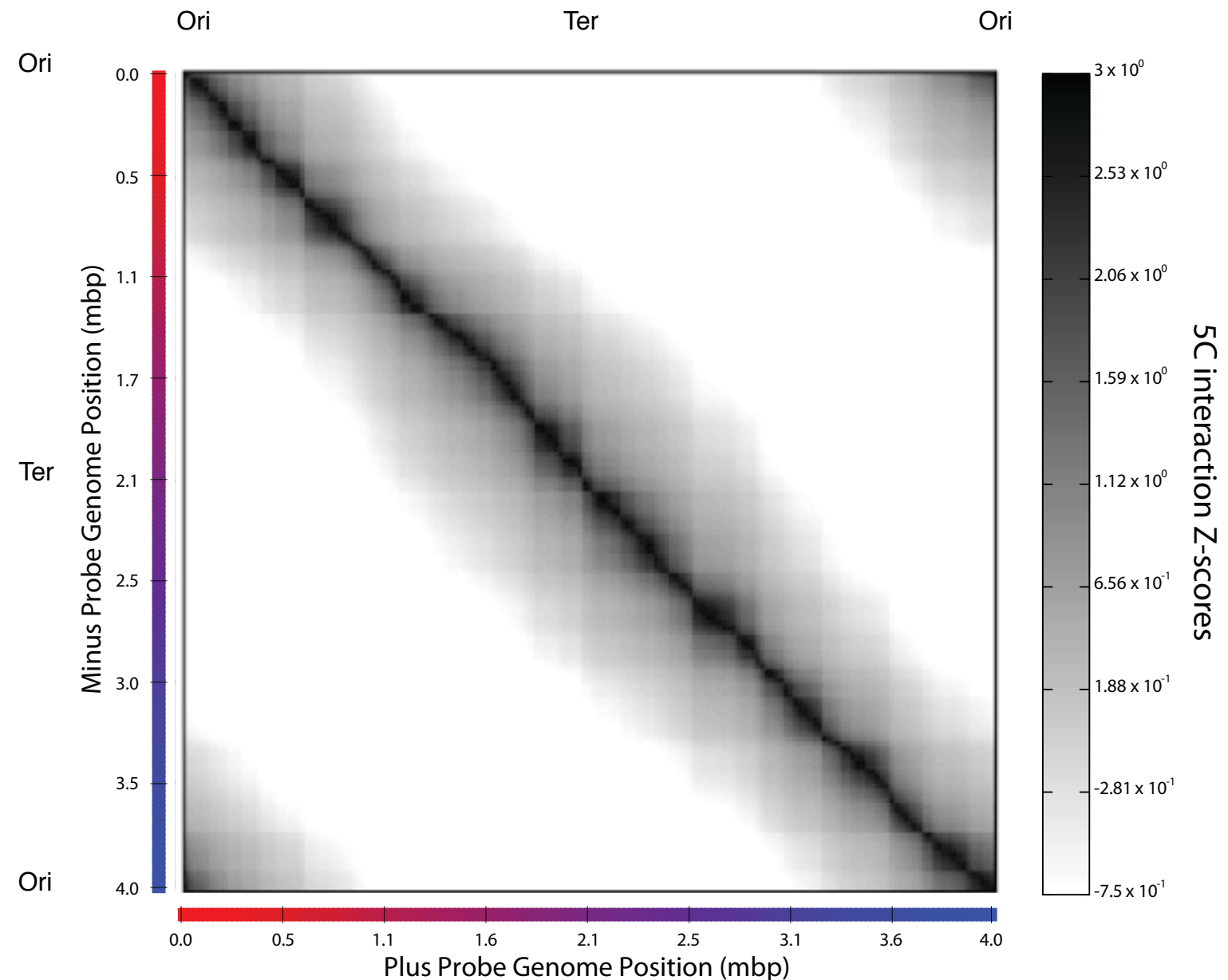
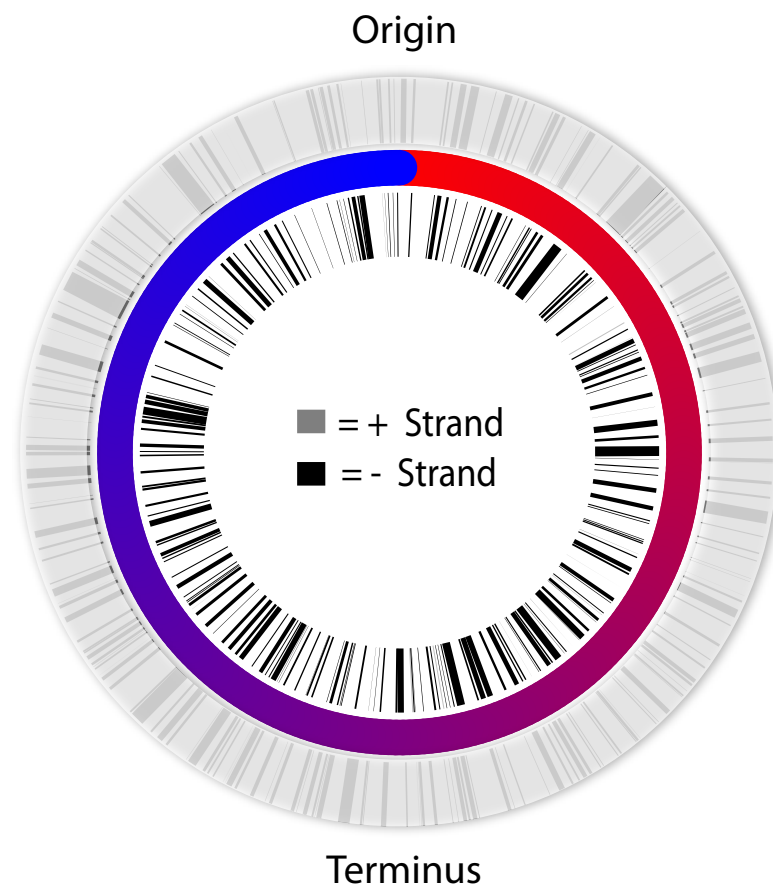
D. Baù et al. **Nat Struct Mol Biol** (2011) 18:107-14
 A. Sanyal et al. **Current Opinion in Cell Biology** (2011) 23:325–33.

Caulobacter crescentus genome



The 3D architecture of *Caulobacter Crescentus*

4,016,942 bp & 3,767 genes

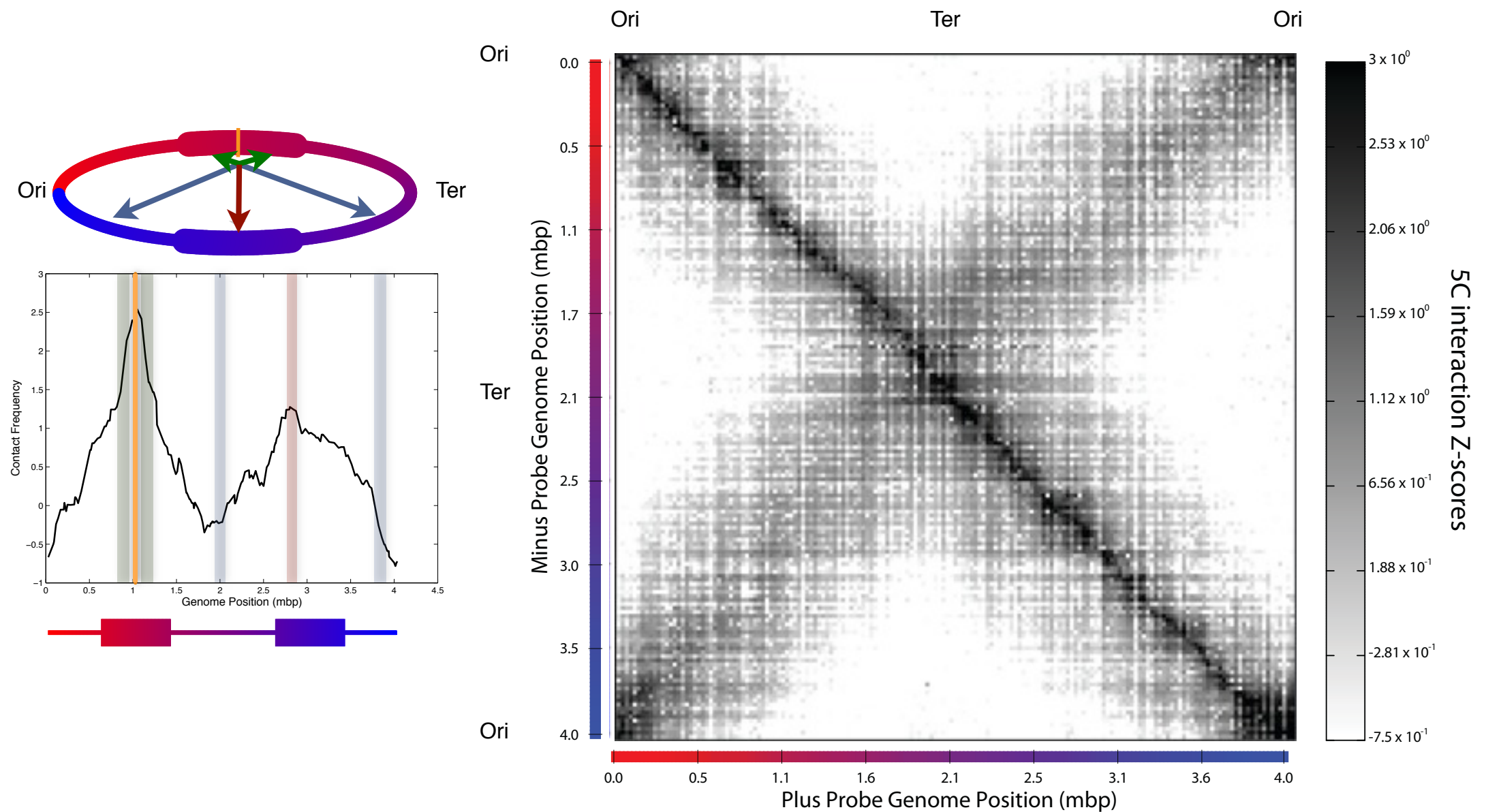
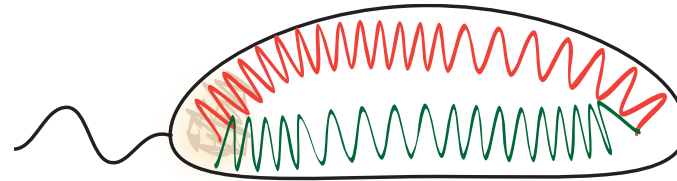


169 5C primers on + strand
170 5C primers on - strand
28,730 chromatin interactions

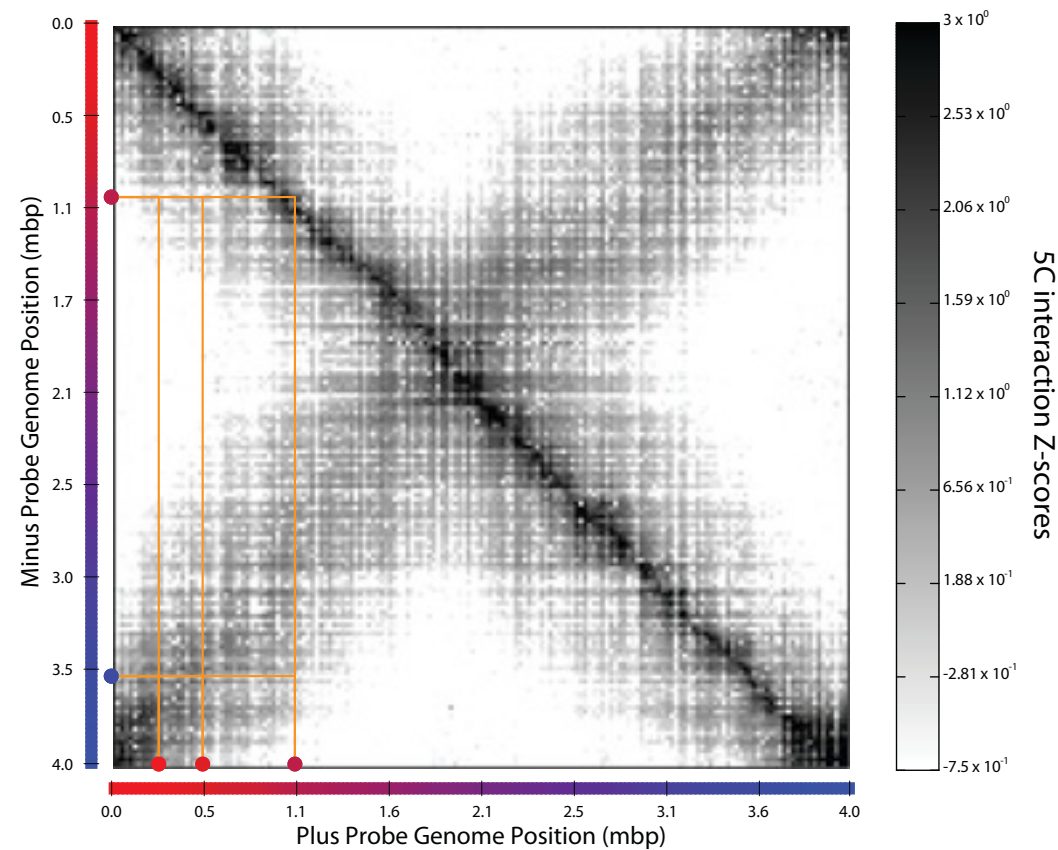
~13Kb

5C interaction matrix

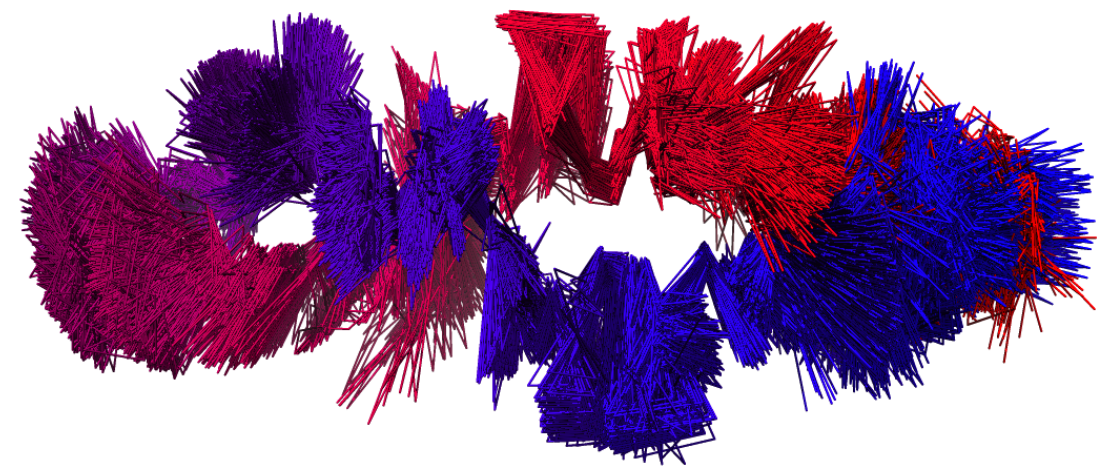
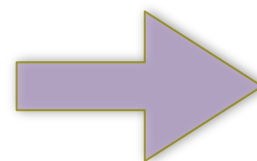
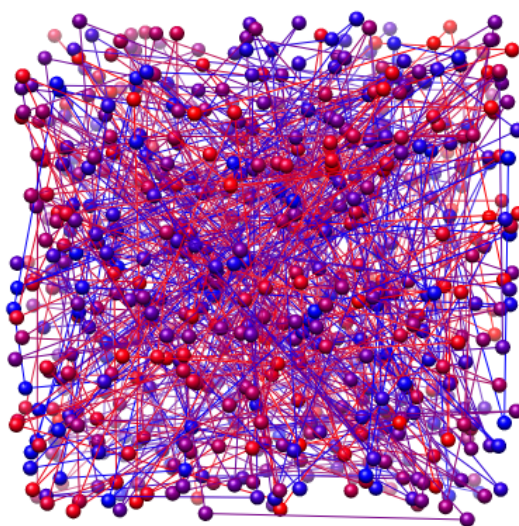
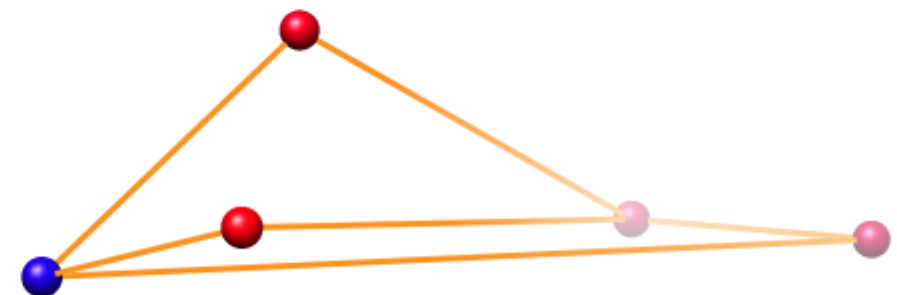
ELLIPSOID for *Caulobacter crescentus*



3D model building with the 5C + IMP approach



339 mers



Genome organization in *Caulobacter crescentus*

Arms are helical

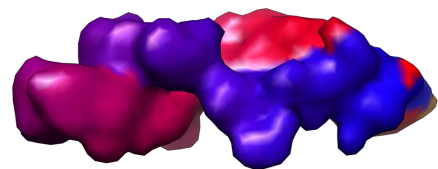
Resolution

dif site 47 ± 17 Kb from Ter

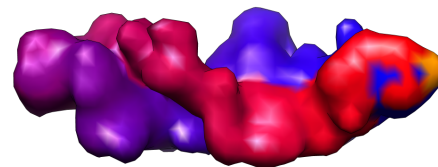
Centromer-like

parS sites 25 ± 17 Kb from Ori

Cluster 1

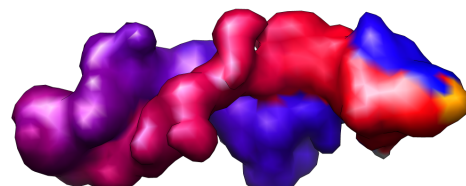


180°

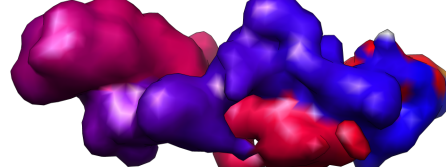


500 nm

Cluster 2

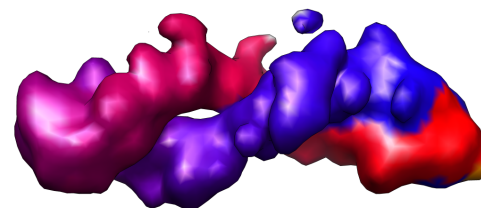


180°

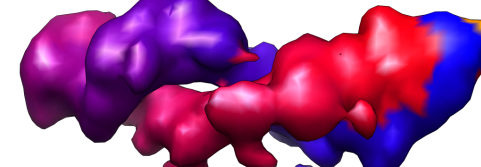


500 nm

Cluster 3

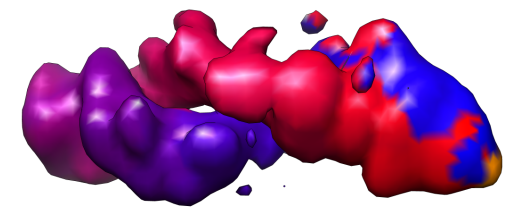


180°



500 nm

Cluster 4



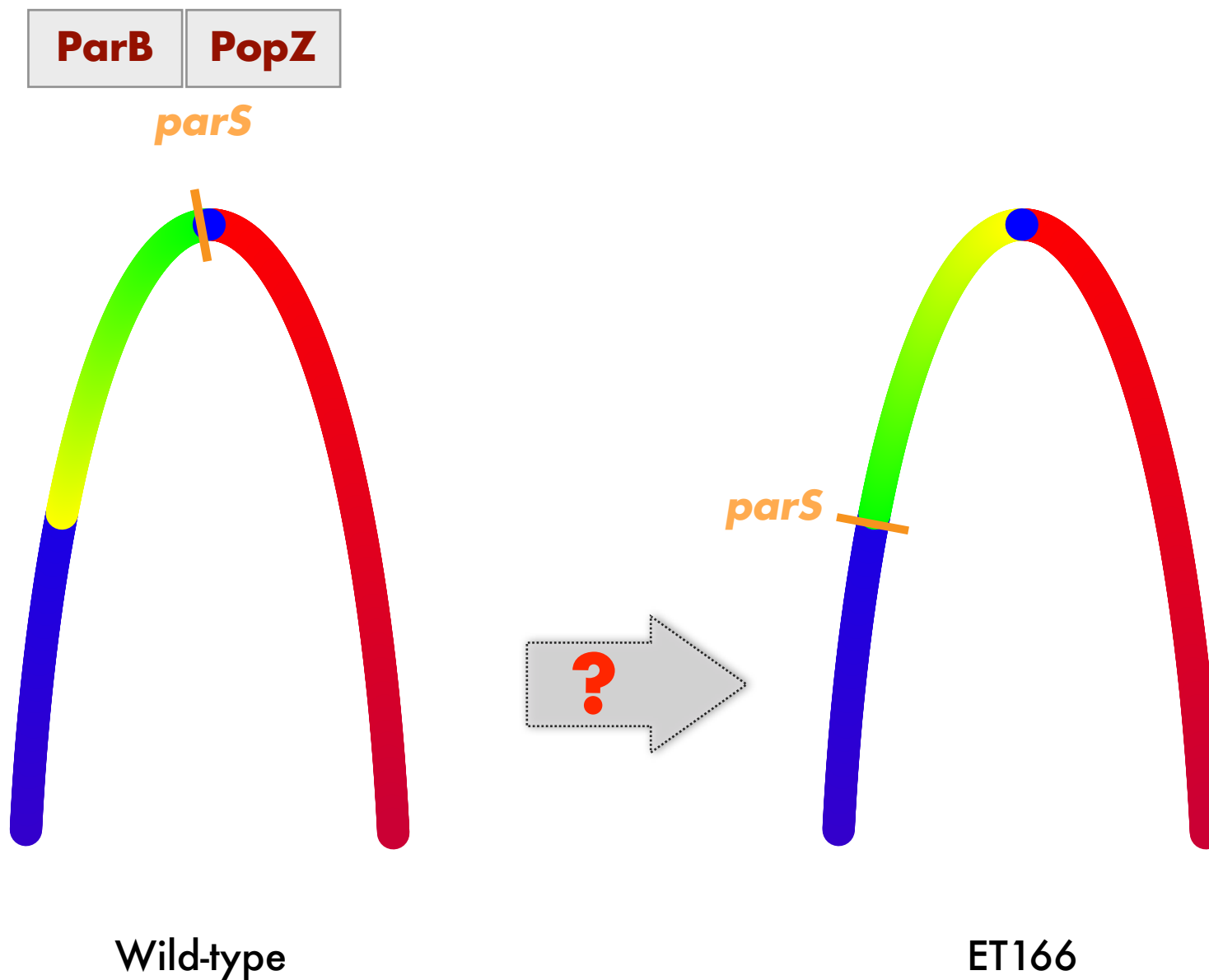
180°



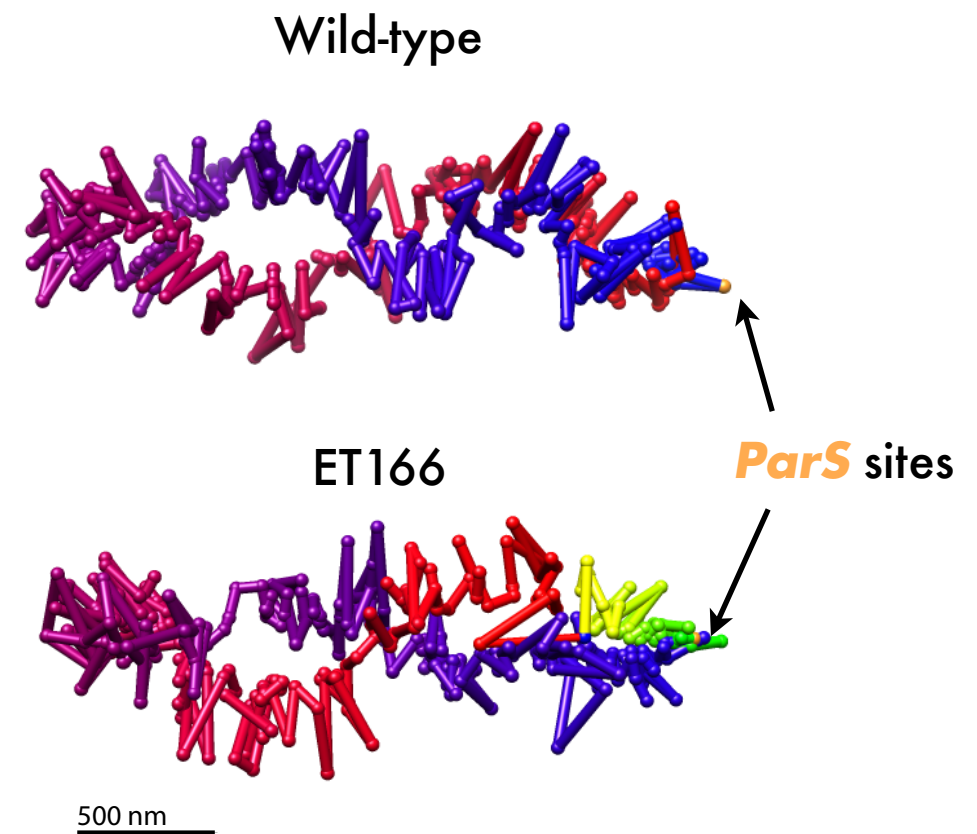
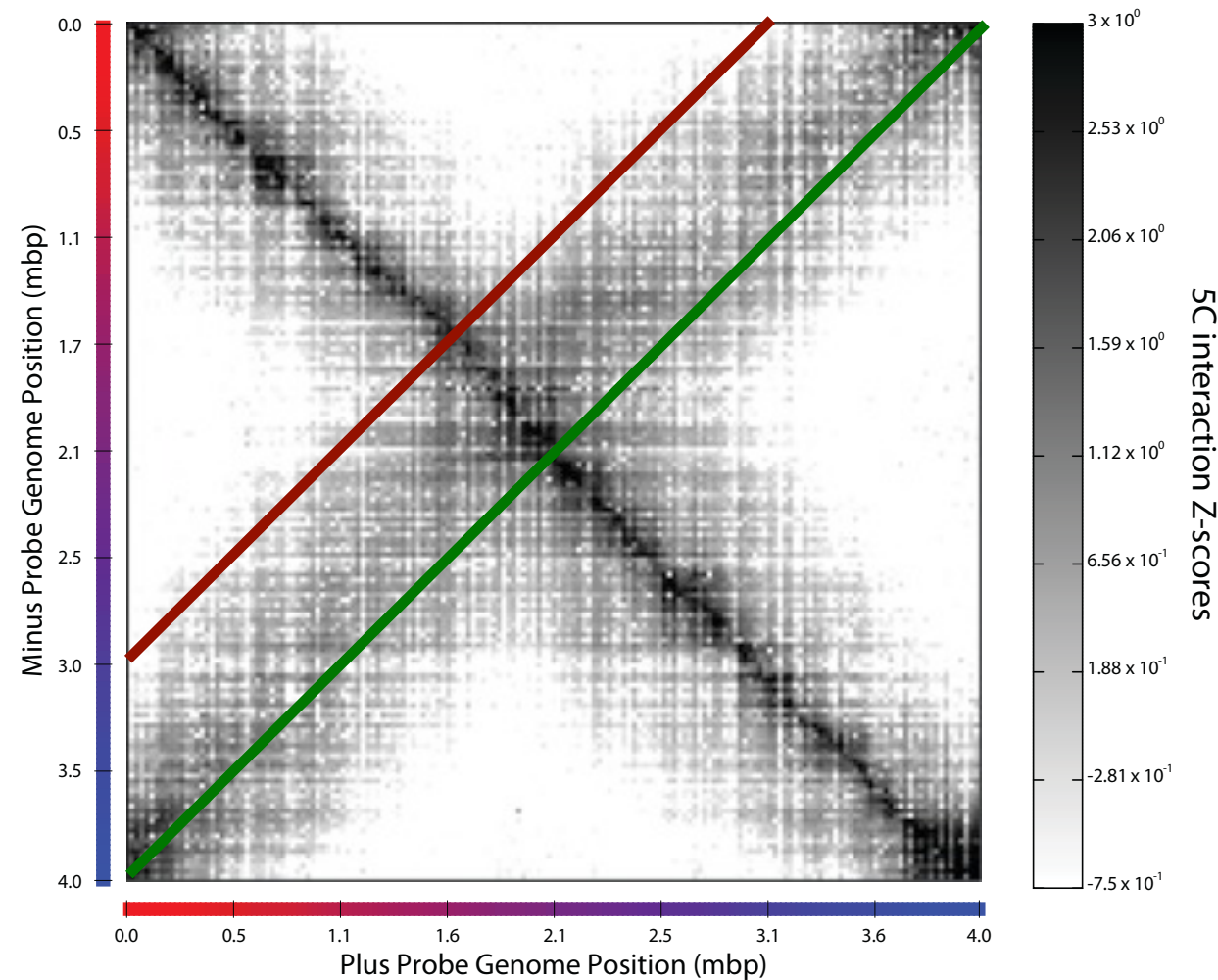
500 nm

MIRRORS!

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

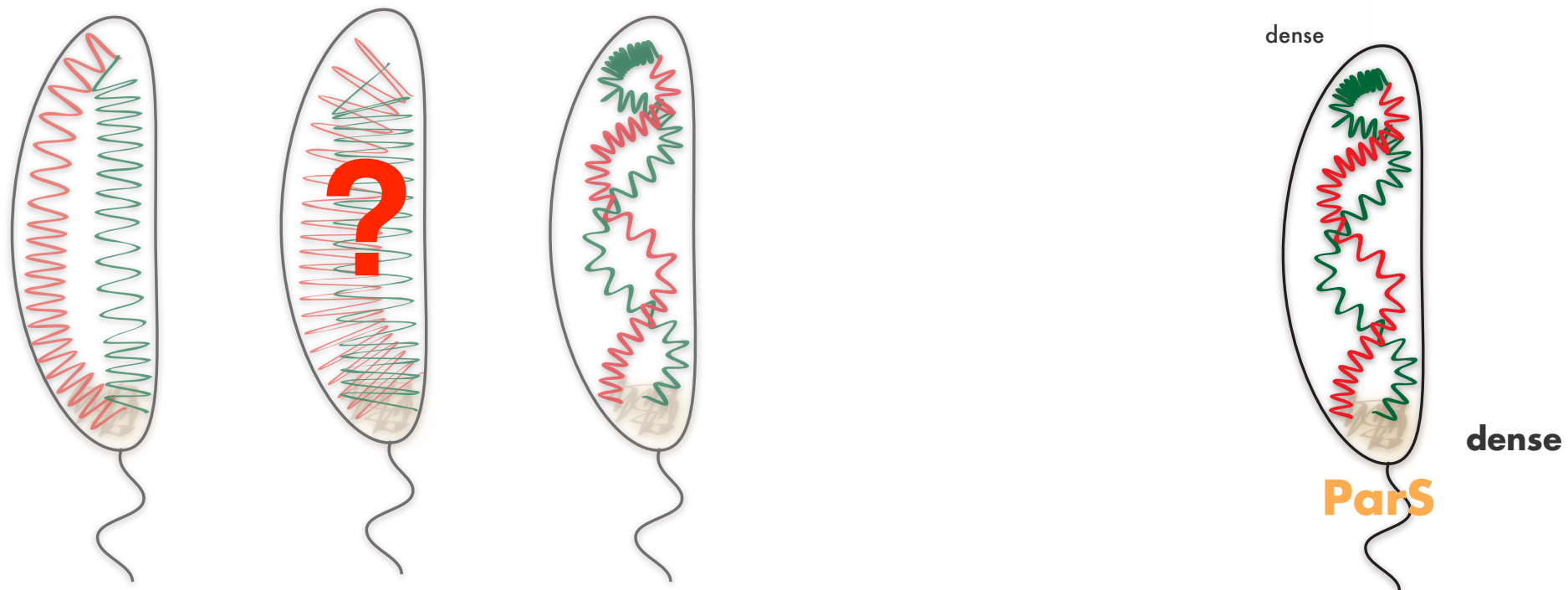
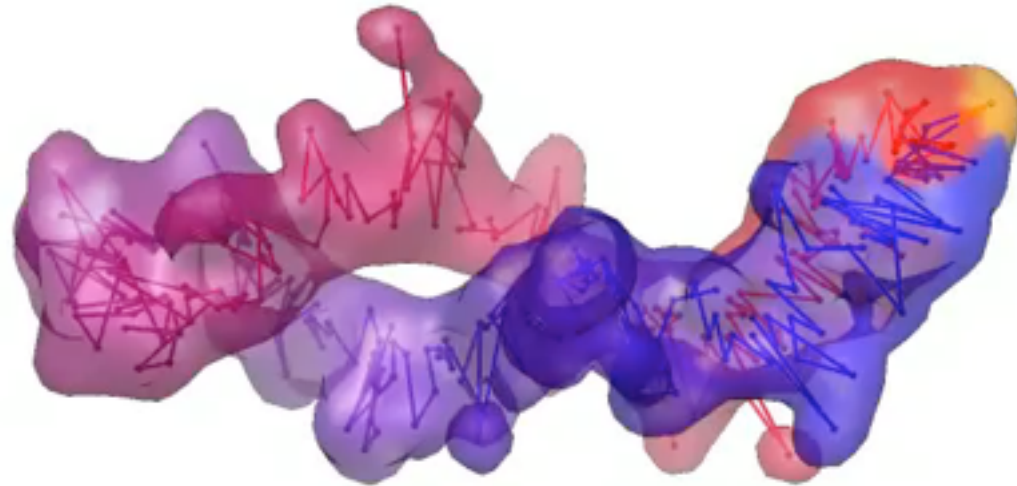


Moving the **parS** sites results in whole genome rotation!



Arms are **STILL** helical

Genome architecture in *Caulobacter*

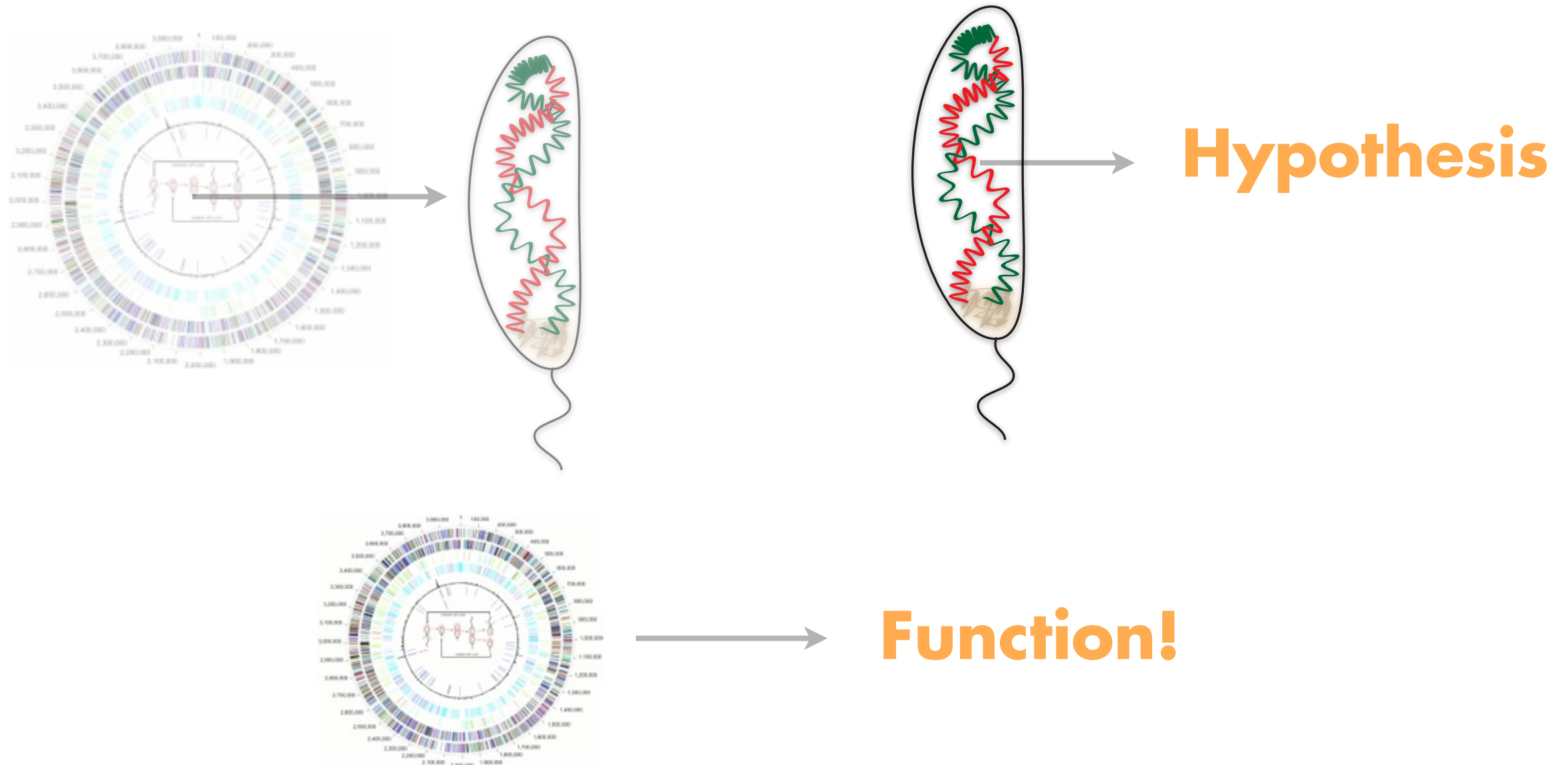


M.A. Umbarger, et al. **Molecular Cell** (2011) **44**:252–264

From Sequence to Function

5C + IMP

Technology



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

Marc A. Marti-Renom^{1*}, Leonid A. Mirny²

1 Structural Genomics Laboratory, Bioinformatics and Genomics Department, Centro de Investigación Principe 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 have 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.

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 microscopy [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 as 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 acknowledging 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.

 PLoS Computational Biology | www.ploscompbiol.org

July 2011 | Volume 7 | Issue 7 | e1002125

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

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

Suechi Thirumangalakudi,¹ Menna C. van Zelm,¹ Marely M. Peak,¹ Steve Cutchen,² Ray Ribbin,³ Jacques J.M. van Gorp,⁴ Frank G. Graessens,⁴ Tobias A. Knoch,¹ and Gert Mulder¹ *Department of Biological Sciences, 0677, University of California, San Diego, La Jolla, CA 92037, USA*
San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92037, USA
Torrey Pines Institute for Molecular Studies, San Diego, CA 92121, USA
Department of Immunology, Department of Biophysical Chemistry, Cell Biology and Genetics
Erasmus MC, Dr. Molenaarsdijk 26, 3015 GD Rotterdam, The Netherlands
Plant Genetics University of Bayreuth, 95492 Bayreuth, Germany
INM-Heinrich Heine University D-40225 Düsseldorf, Germany
These authors contributed equally to this work.
Correspondence: to.knoch@ucsd.edu (T.A.K.), mulder@biomc.erasmusmc.nl (G.M.)
DOI: 10.1016/j.celrep.2015.05.010

SUMMARY

[illegible]

Cell 120, 265-279, April 18, 2008 ©2008 Elsevier Inc. 265

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

Chromatin conformation signatures of cellular differentiation
James Fraser¹, Mathieu Rousseau¹, Solomon Shenker²,
Maria A Ferraiuolo³, Yoshihide Hayashizaki¹, Mathieu Blanchette² and
Josée Dostie²

Address: *Department of Biochemistry and McGill Cancer Center, McGill University, 3655 Promenade Sir-William-Osler, Montreal, H3G 2Y4, Canada; *McGill Centre for Biomedicine, McGill University, 3775 University, Montreal, H3A 2B4, Canada; *RIKEN Omics Science Center, RIKEN Yokohama Institute, 1-7-42 Suitoh-cho Tsukuba-In, Yokohama, 250-0445, Japan.

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

The electronic version of this article is the complete one and can be found online at <http://genesandbiology.com/2008/10/4/R27>

[illegible]

Received: 29 October 2008
 Revised: 22 December 2008
 Accepted: 19 April 2009

© 2009 Fraser et al.; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abu

One of the major genomics challenges is to better understand how correct gene expression is orchestrated. Recent studies have shown how spatial chromosome organization is critical in the regulation of gene expression. Here, we develop a suite of computer programs to identify chromosome conformation signatures with 5C technology (<http://CostelloBiochem.nyu.edu>). We identified dynamic Hi-C cluster configurations and chromosome signatures associated with cellular differentiation. Genome-wide chromosome conformation signatures identification might uniquely identify disease-associated states and represent an entirely novel class of human disease biomarkers.

Rationale

Cad specialization in the defining habitat of *maizetum* and results from differentiation of precursor cells. Differentiation of precursor cells into specialized cell types is a process that is well understood in the laboratory, but is poorly understood of specific phytochemicals. This process is the basis of the differentiation of specialized cells in the maintenance. For example, *lipoxygenase* differentiation of such tissues can lead to human diseases such as leukemia [10]. The process of differentiation of precursor cells into specialized cells is not only essential to understanding biology, but also to develop effective strategies for precursor cell differentiation [11]. The process of differentiation of precursor cells into specialized cells is not only essential to understanding biology, but also to develop effective strategies for precursor cell differentiation [11]. The process of differentiation of precursor cells into specialized cells is not only essential to understanding biology, but also to develop effective strategies for precursor cell differentiation [11].

regulate the *switched-up* and *down-regulation* of numerous transcription factors. This study provides the first in-depth picture of the interplay between transcription factors, general promoter activity, and RNA transcripts required for differentiation of human leukemia cells.

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

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

DEKKER/MARTI-RENO
NSMB (2011) 18:107-14[illegible][illegible][illegible]



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. Marti-Renom

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

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