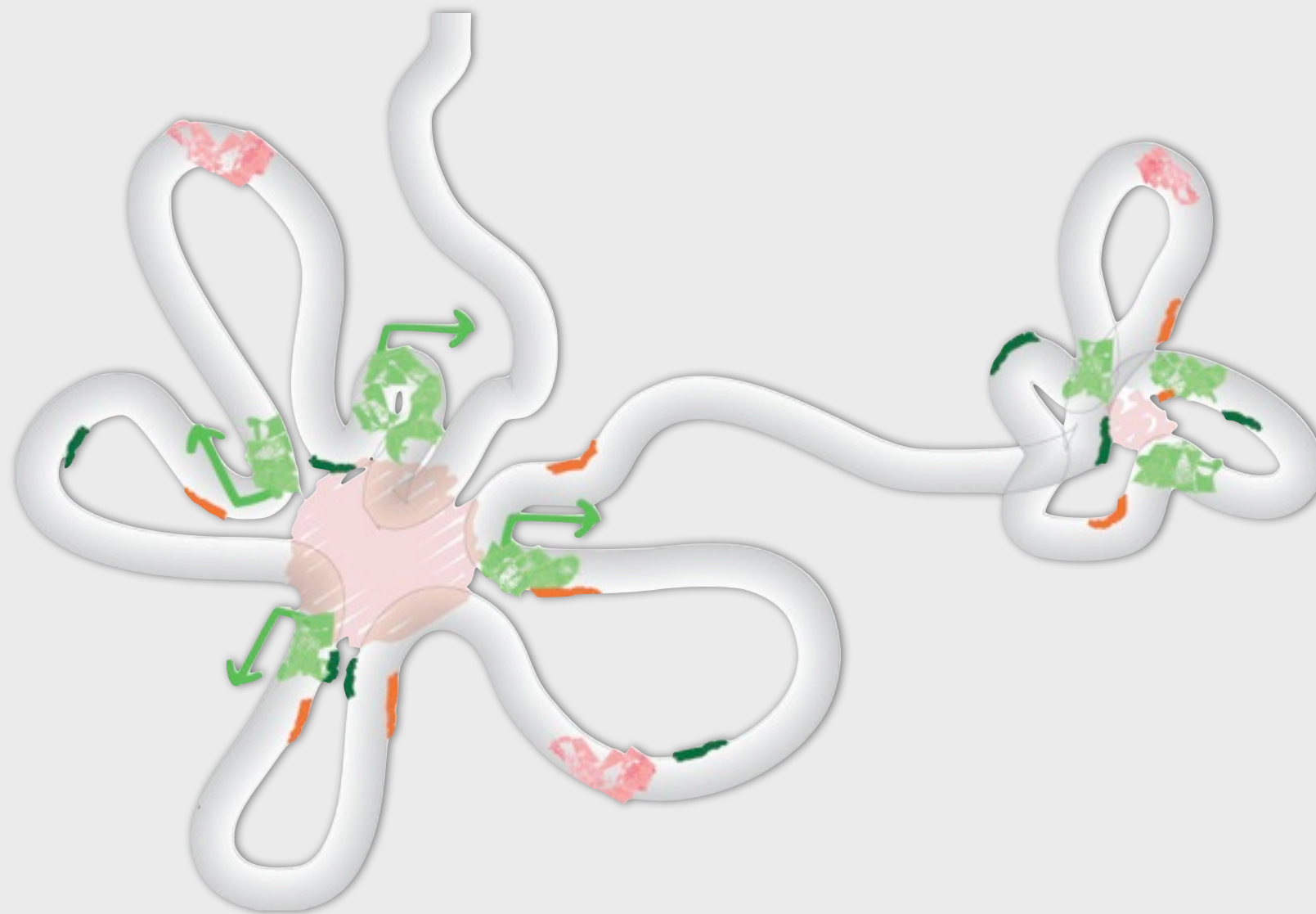


3D folding of chromosomal domains in relation to gene expression



Marc A. Marti-Renom

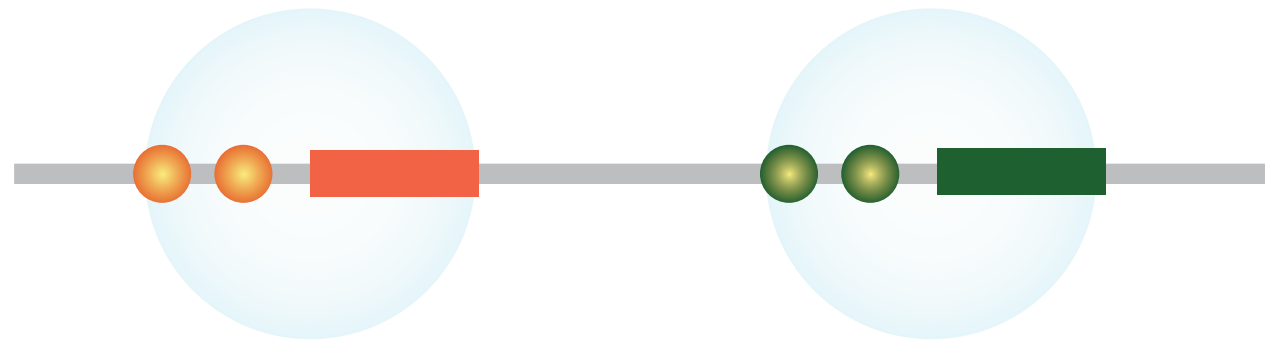
<http://sgu.bioinfo.cipf.es>



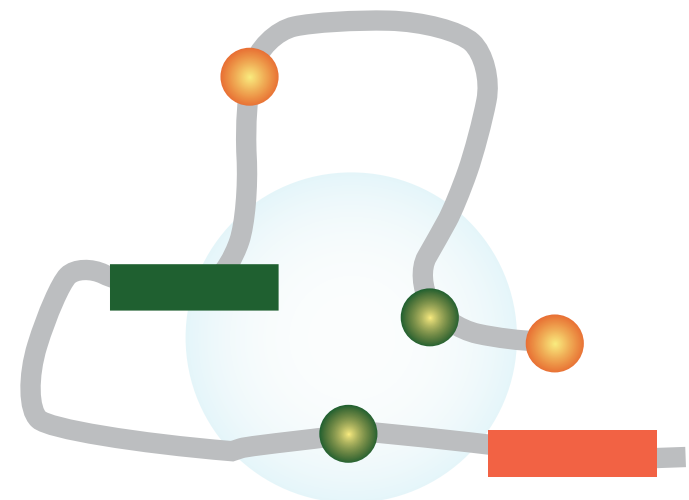
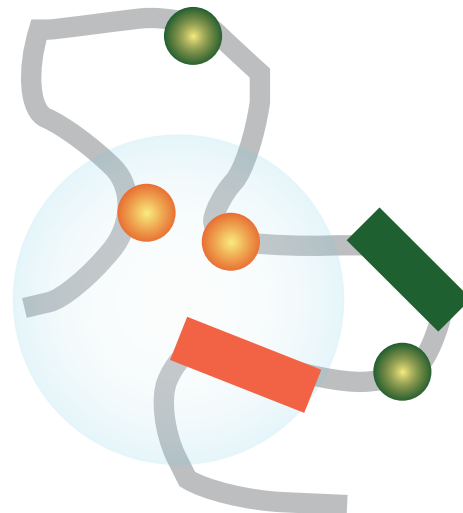
Aim

Can we relate structure and expression?

Simple genomes



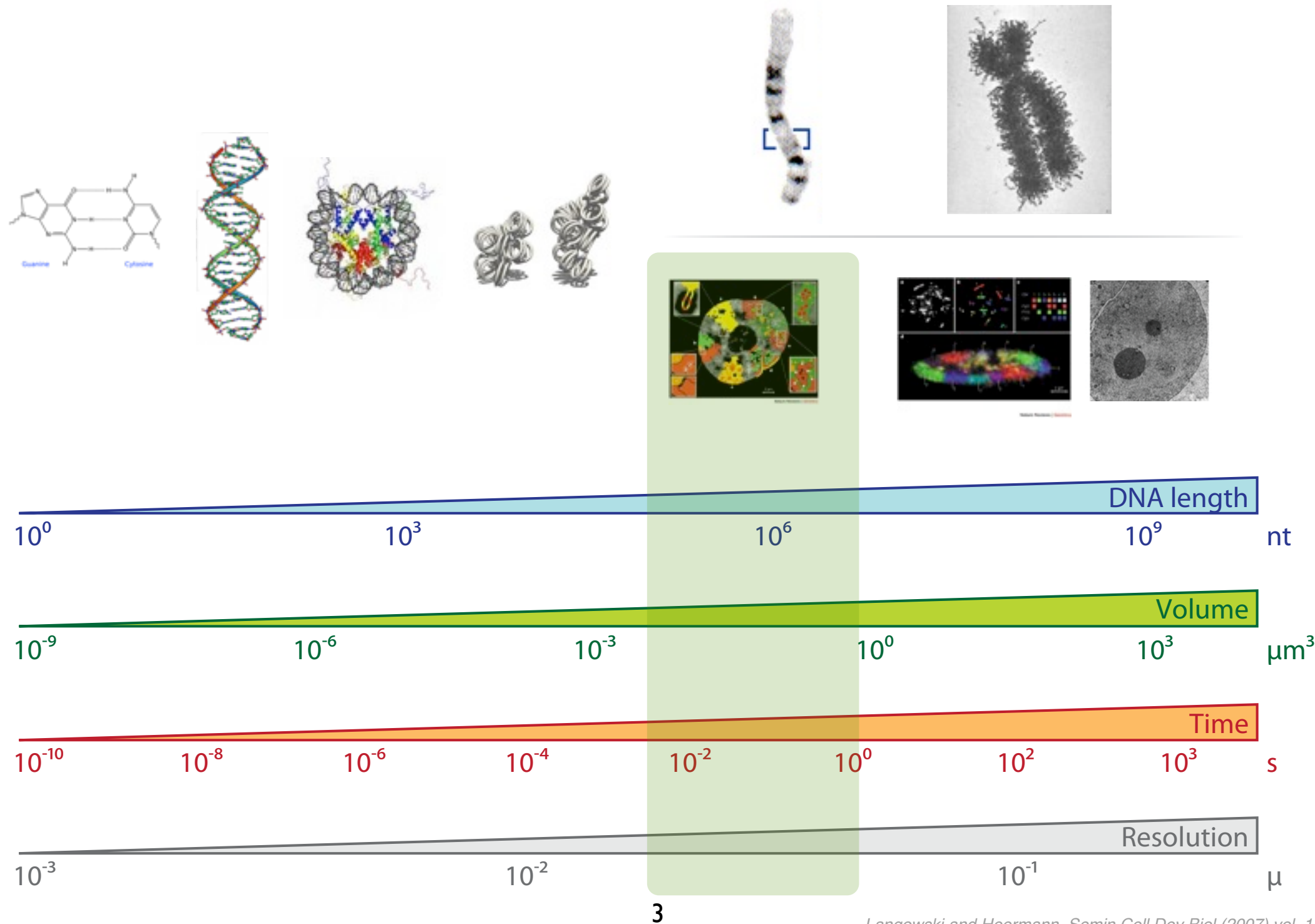
Complex genomes



Resolution

Limited knowledge...

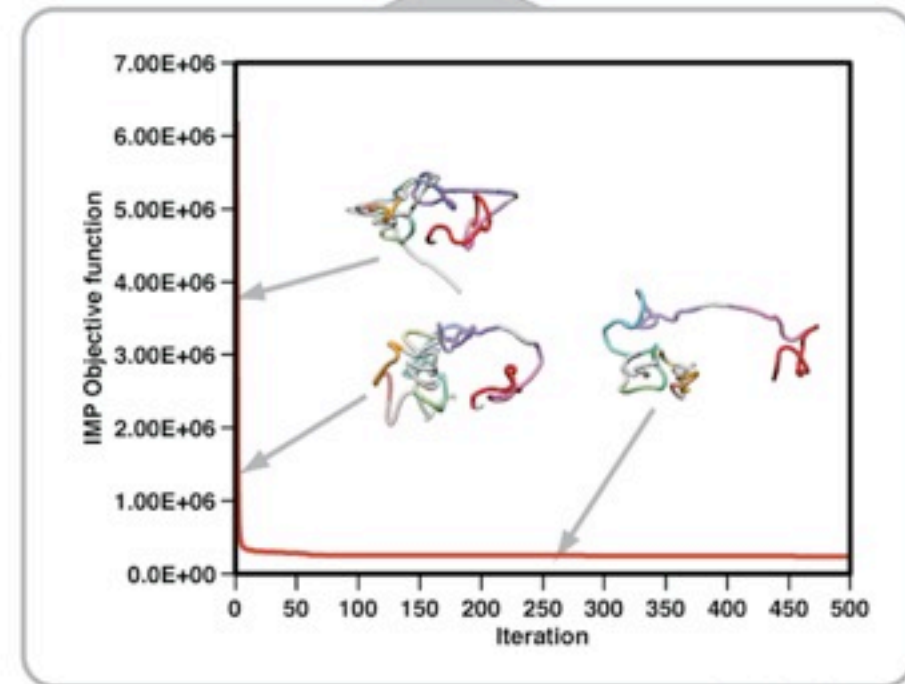
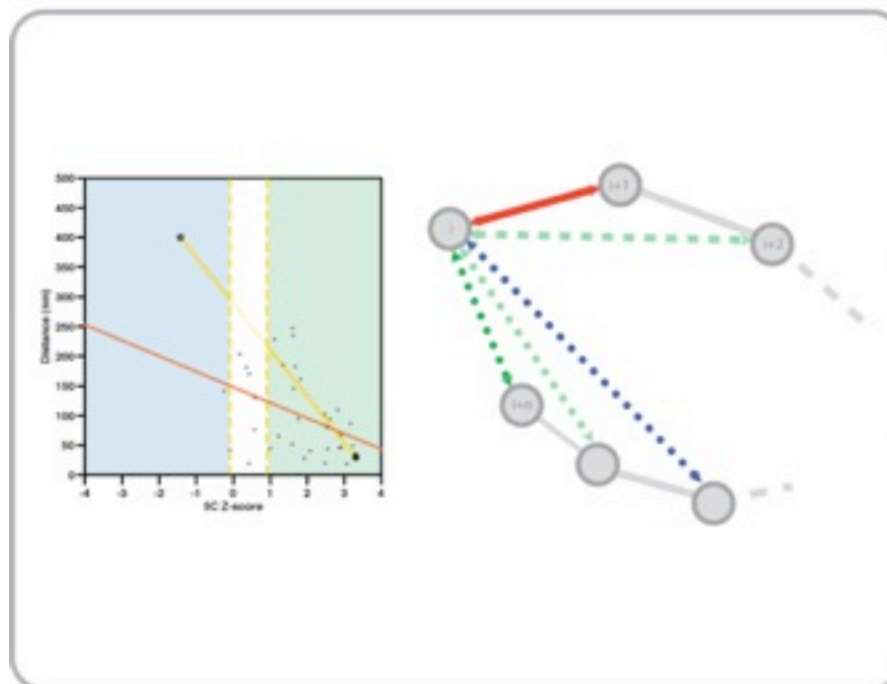
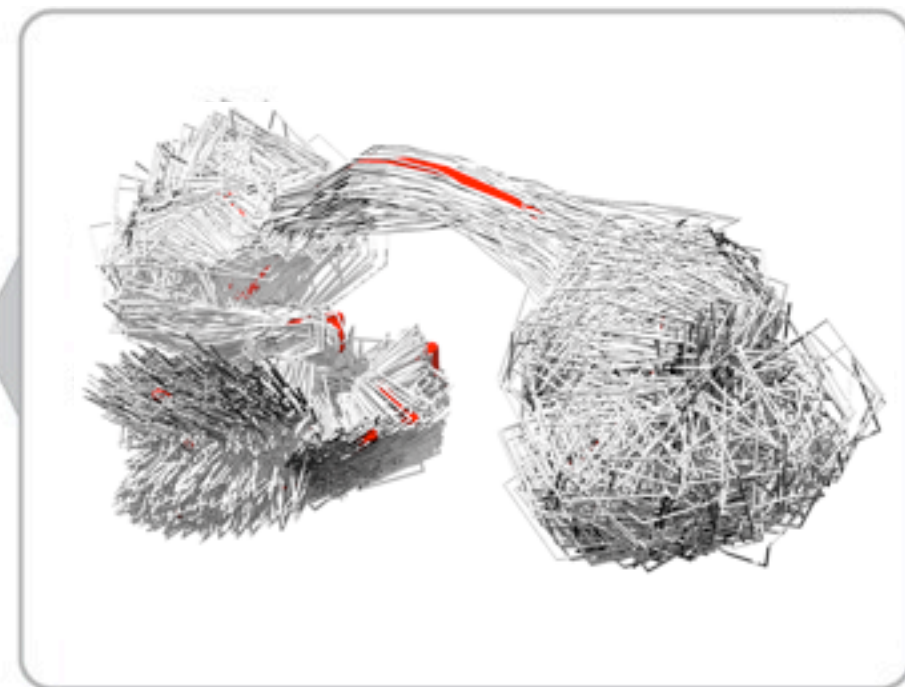
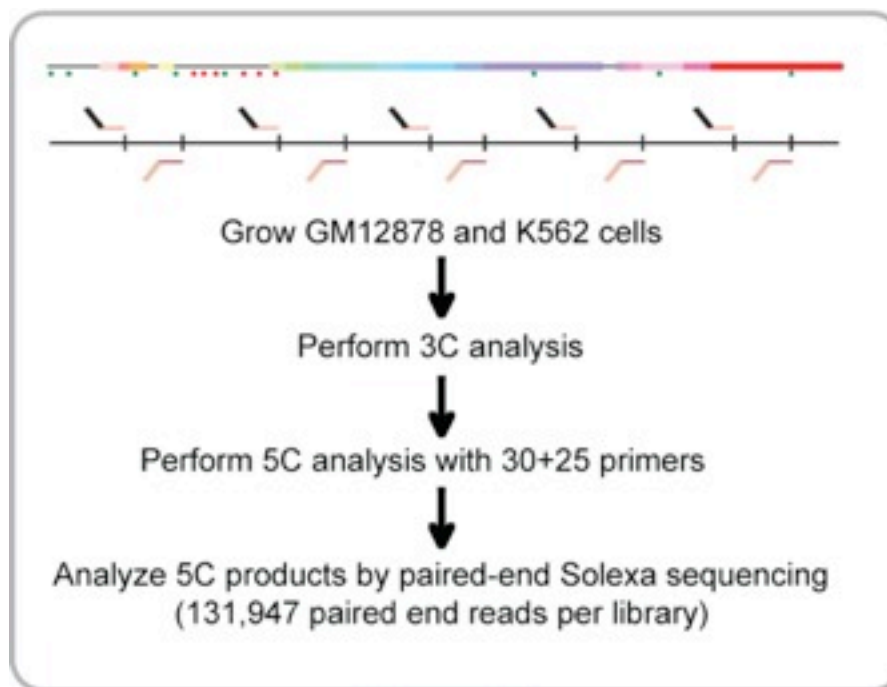
Knowledge



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

Integrative and iterative approach

Experiments



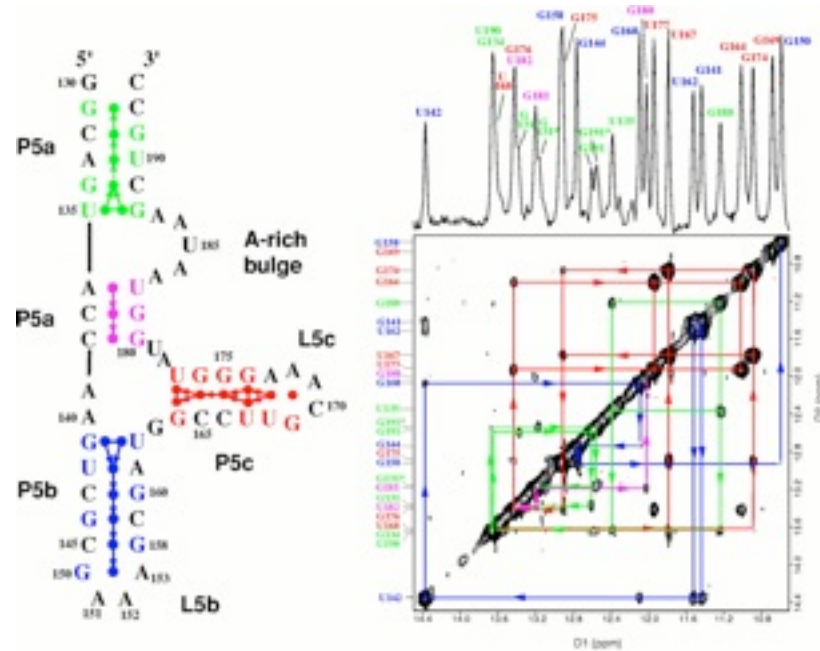
Computation 

Structure determination

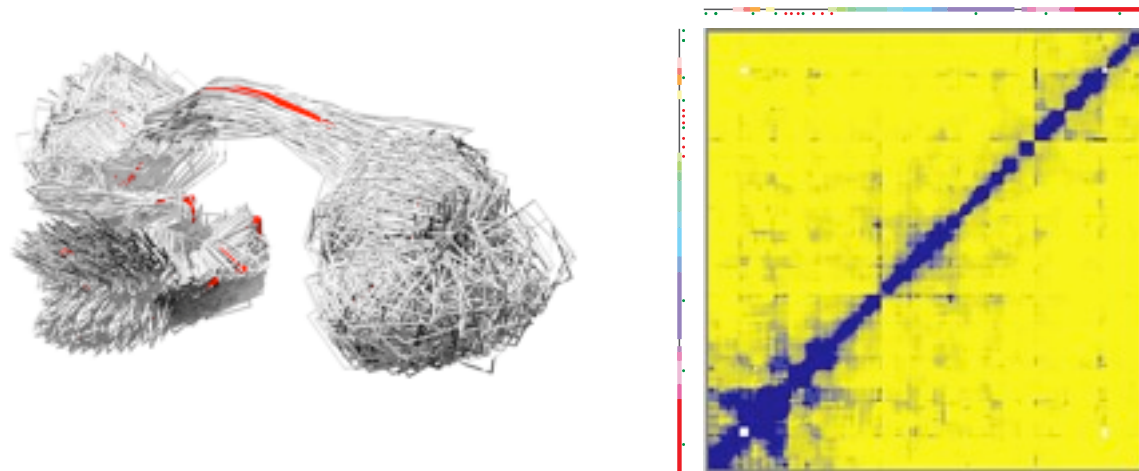
Integrative Modeling Platform

<http://www.integrativemodeling.org>

Alber et al. Nature (2007) vol. 450 (7170) pp. 683-94



Biomolecular structure determination
2D-NOESY data



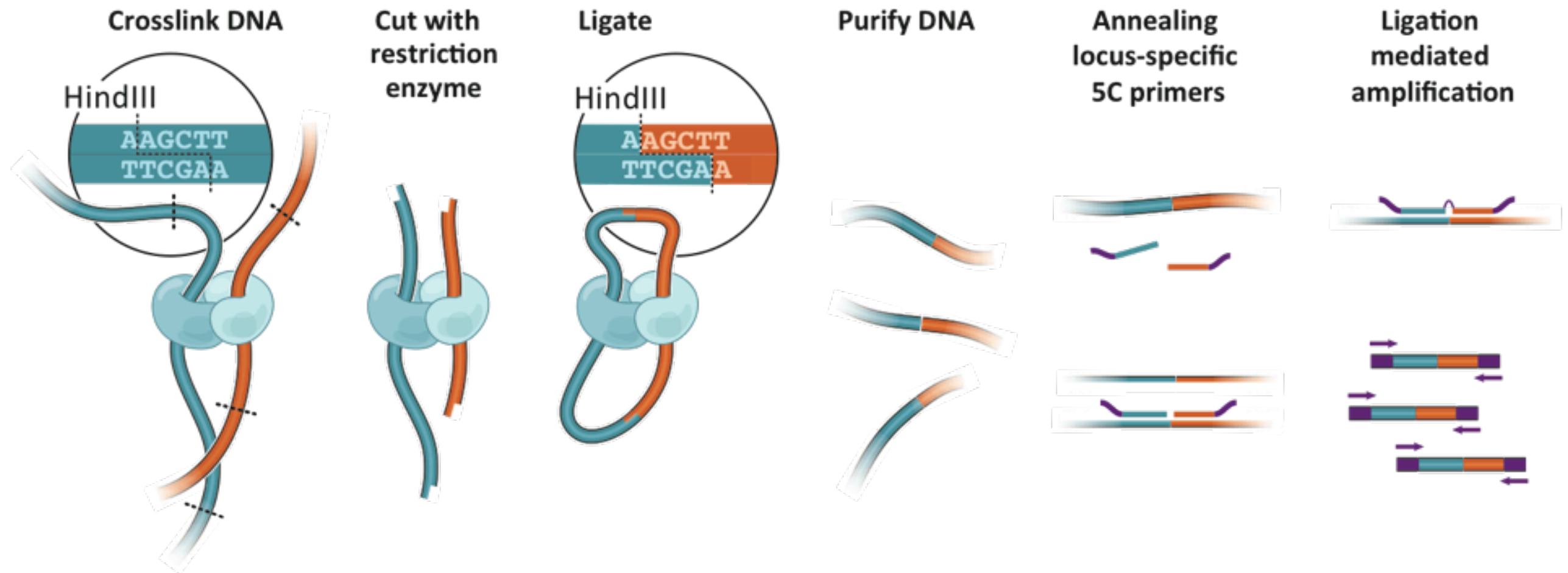
Chromosome structure determination
5C data

5C technology

Detecting up to millions of interactions in parallel

<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

Human α -globin domain

ENm008 genomic structure and environment

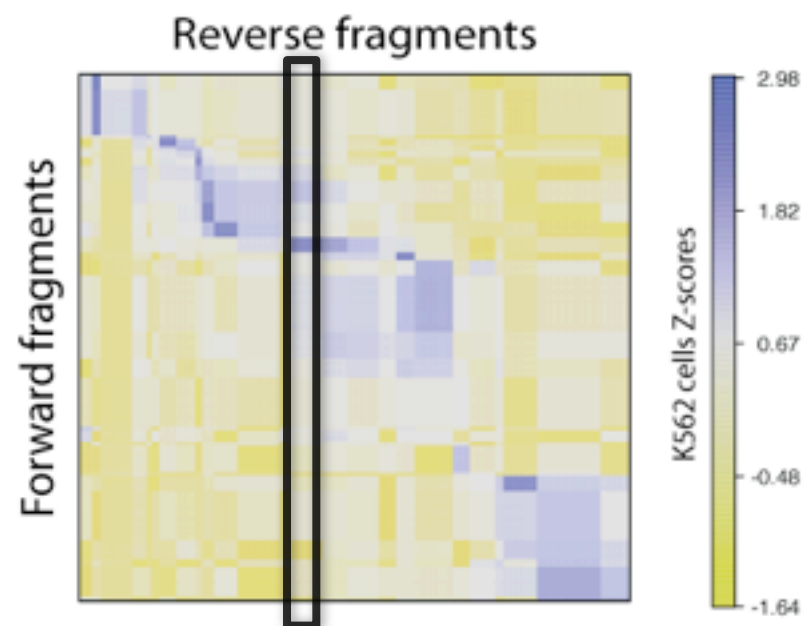
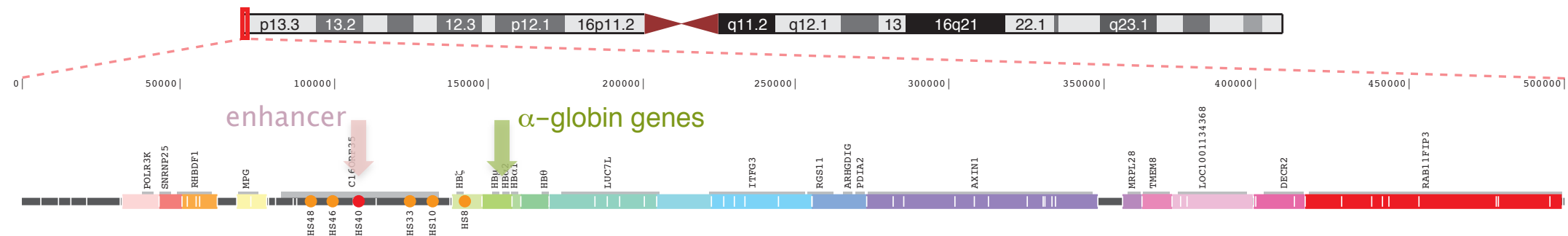
ENCODE Consortium. *Nature* (2007) vol. 447 (7146) pp. 799-816



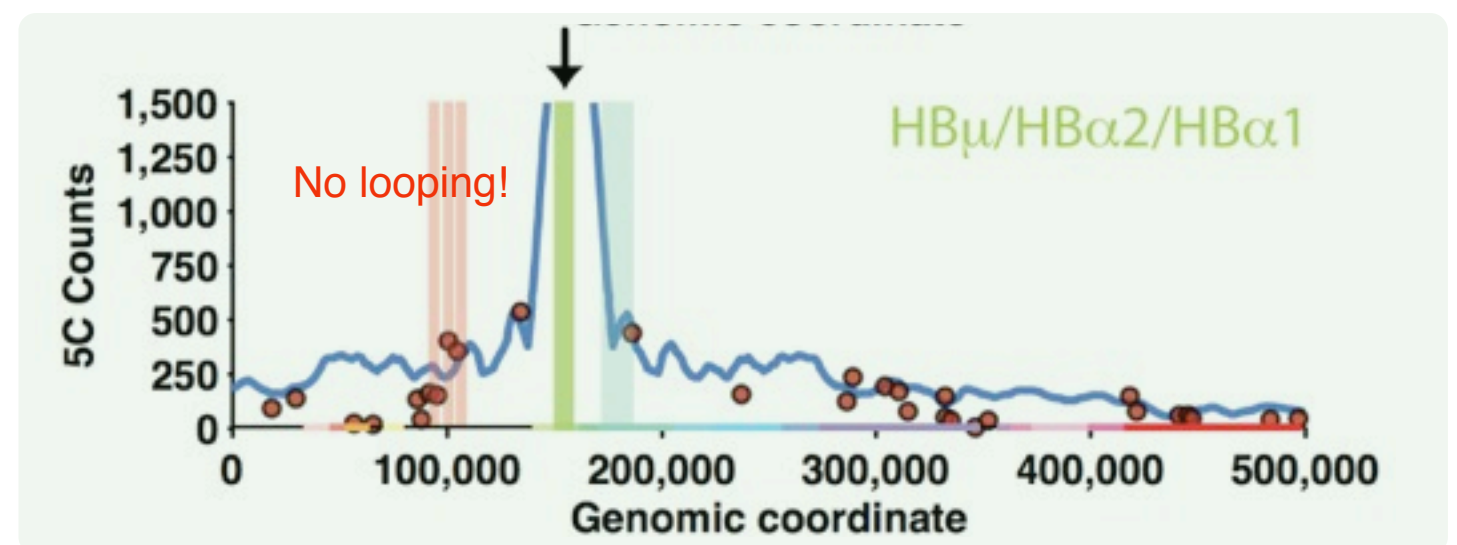
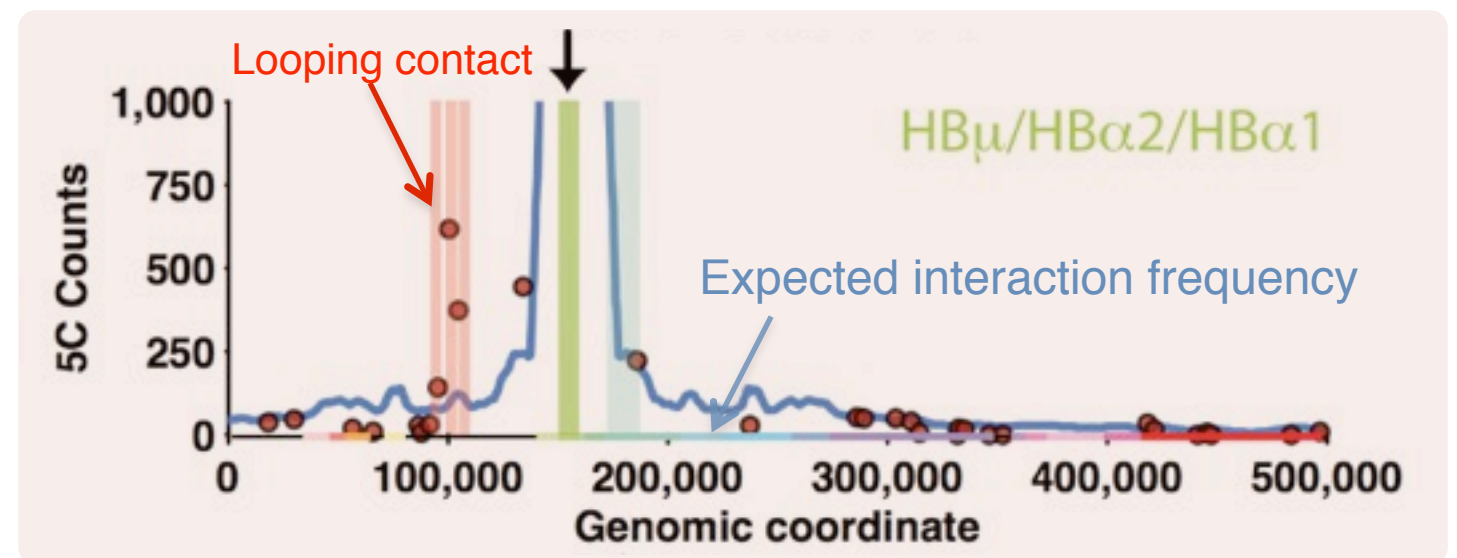
The ENCODE data for ENm008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute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

5C on the human α -globin domain

ENm008 genomic structure and environment

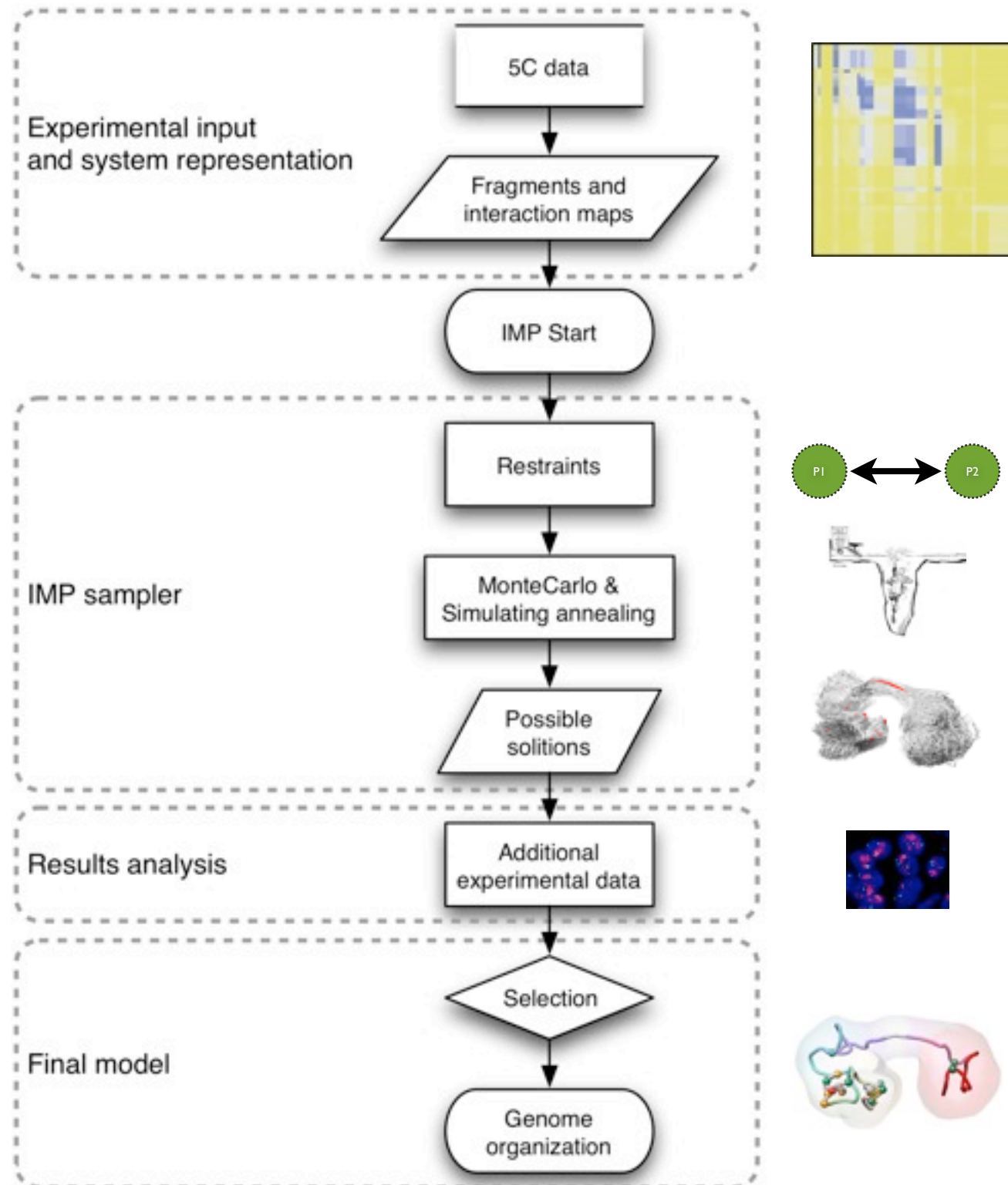


K562 cells:
 α -globin genes active



Integrative Modeling

<http://www.integrativemodeling.org>



Representation

Harmonic

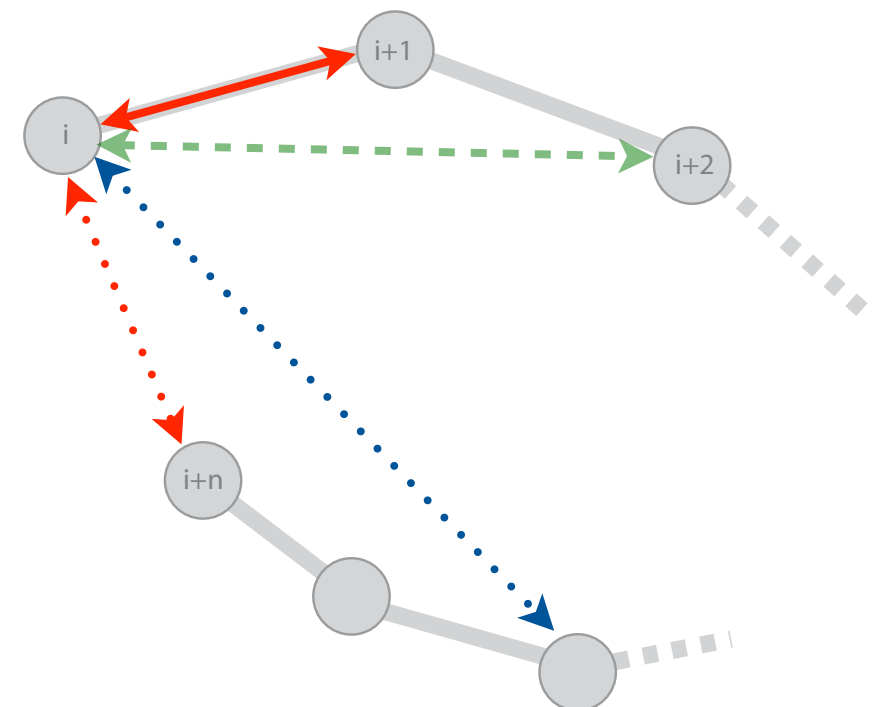
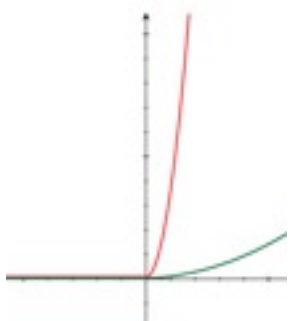
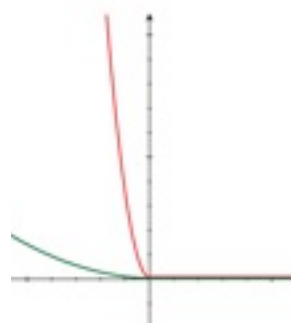
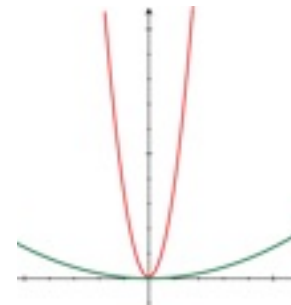
$$H_{i,j} = k(d_{i,j} - d_{i,j}^0)^2$$

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

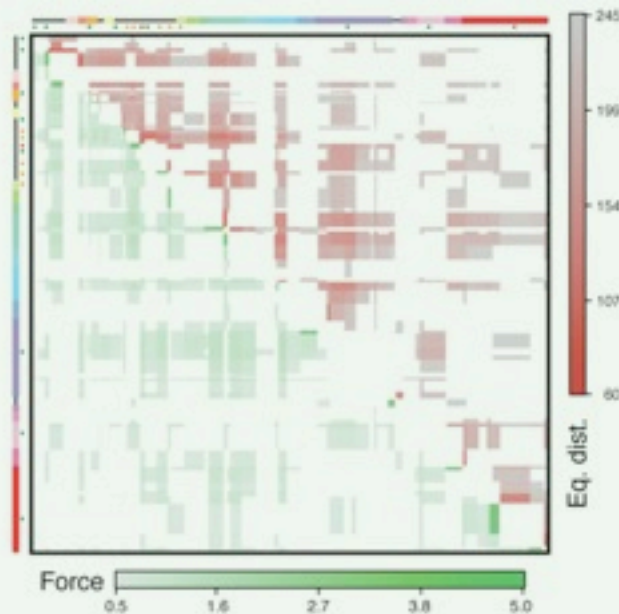
$$\begin{cases} \text{if } d_{i,j} \geq d_{i,j}^0; & ubH_{i,j} = k(d_{i,j} - d_{i,j}^0)^2 \\ \text{if } d_{i,j} < d_{i,j}^0; & ubH_{i,j} = 0 \end{cases}$$



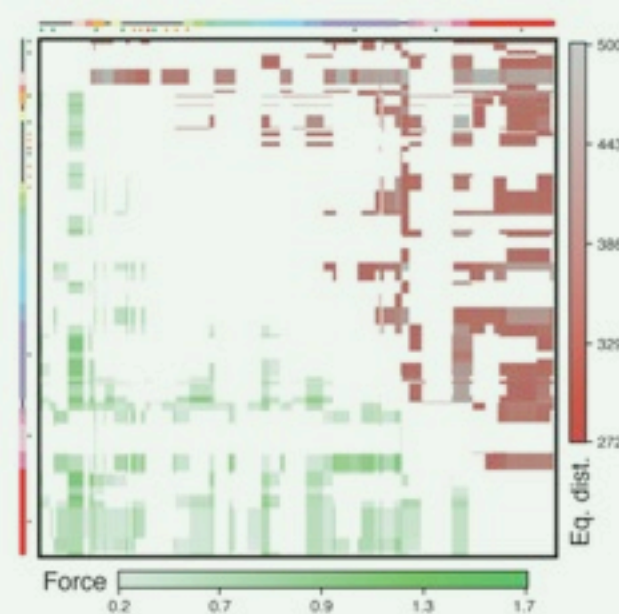
Scoring

GM12878

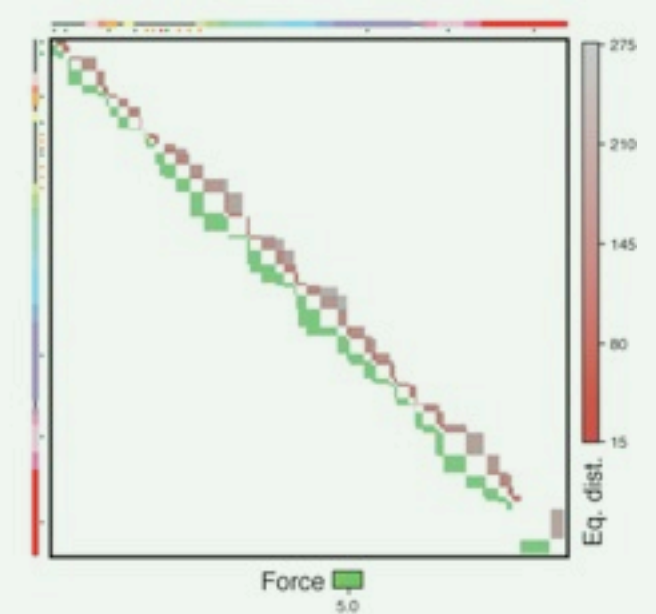
70 fragments
1,520 restraints



Harmonic



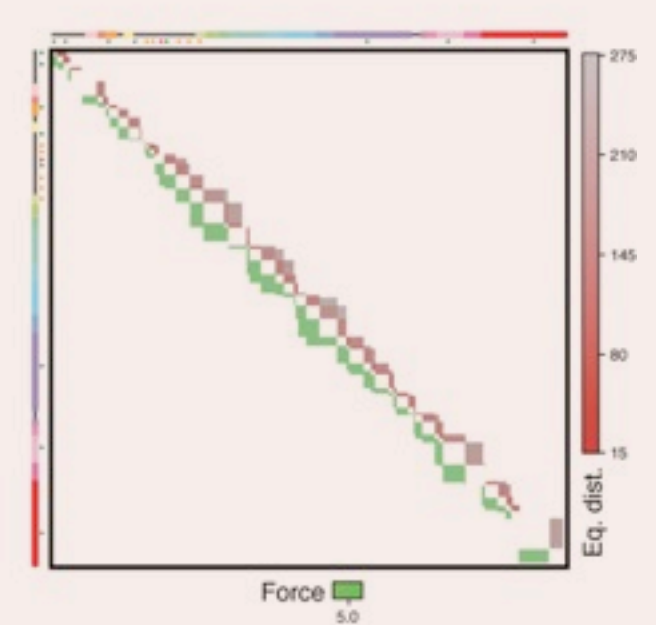
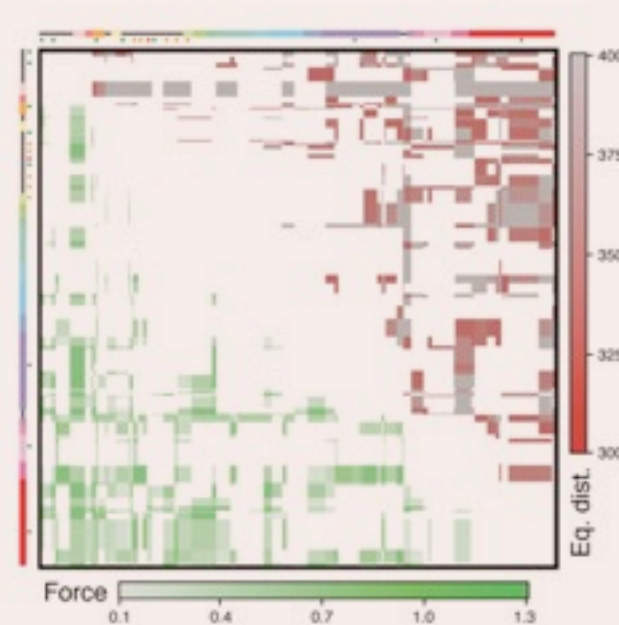
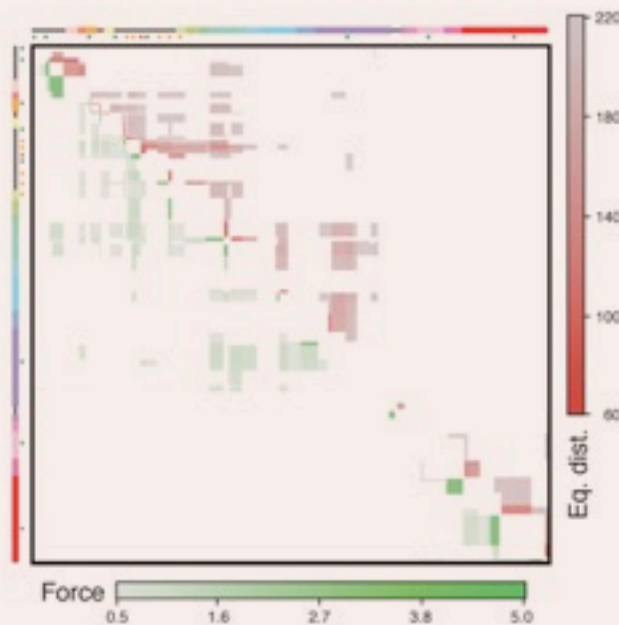
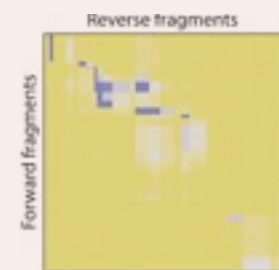
Harmonic Lower Bound



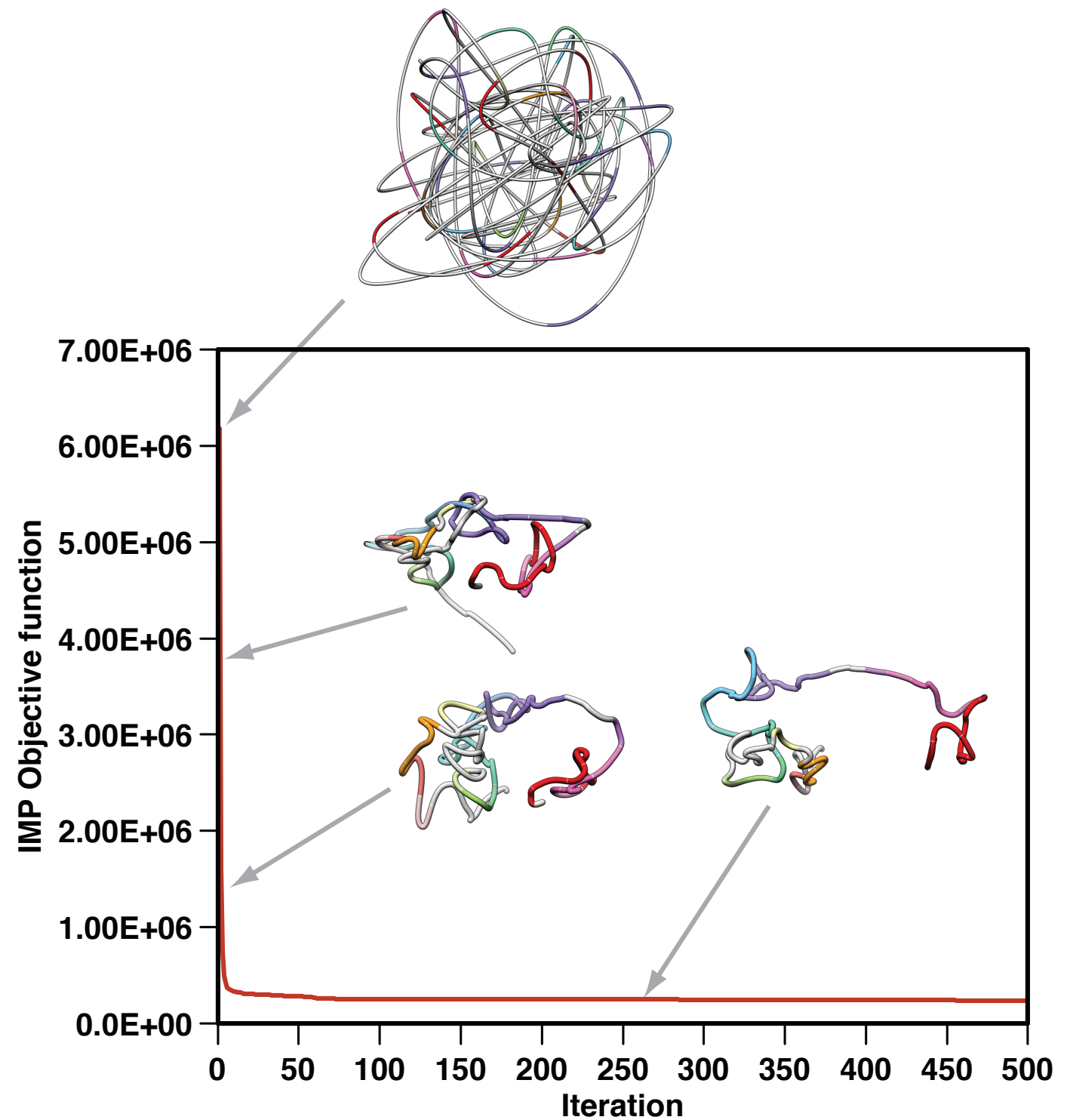
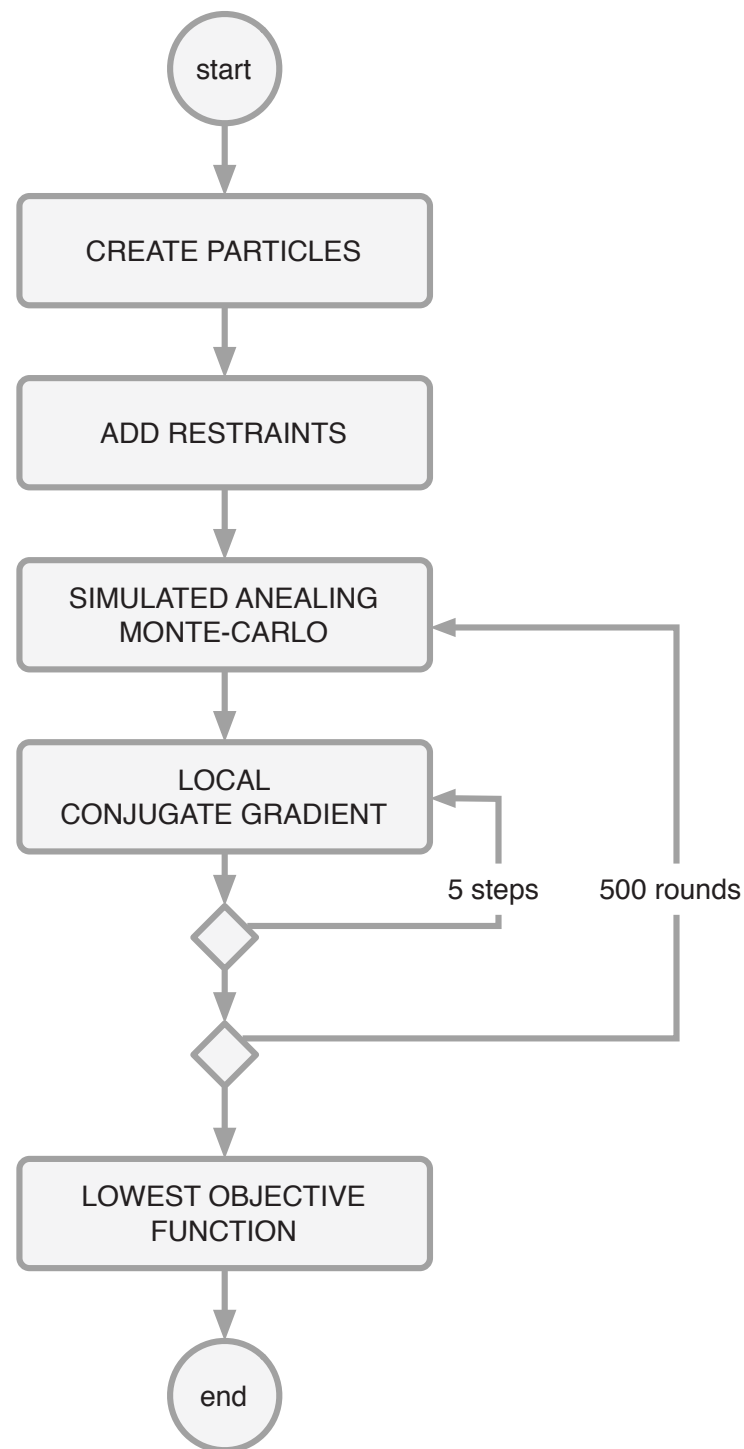
Harmonic Upper Bound

K562

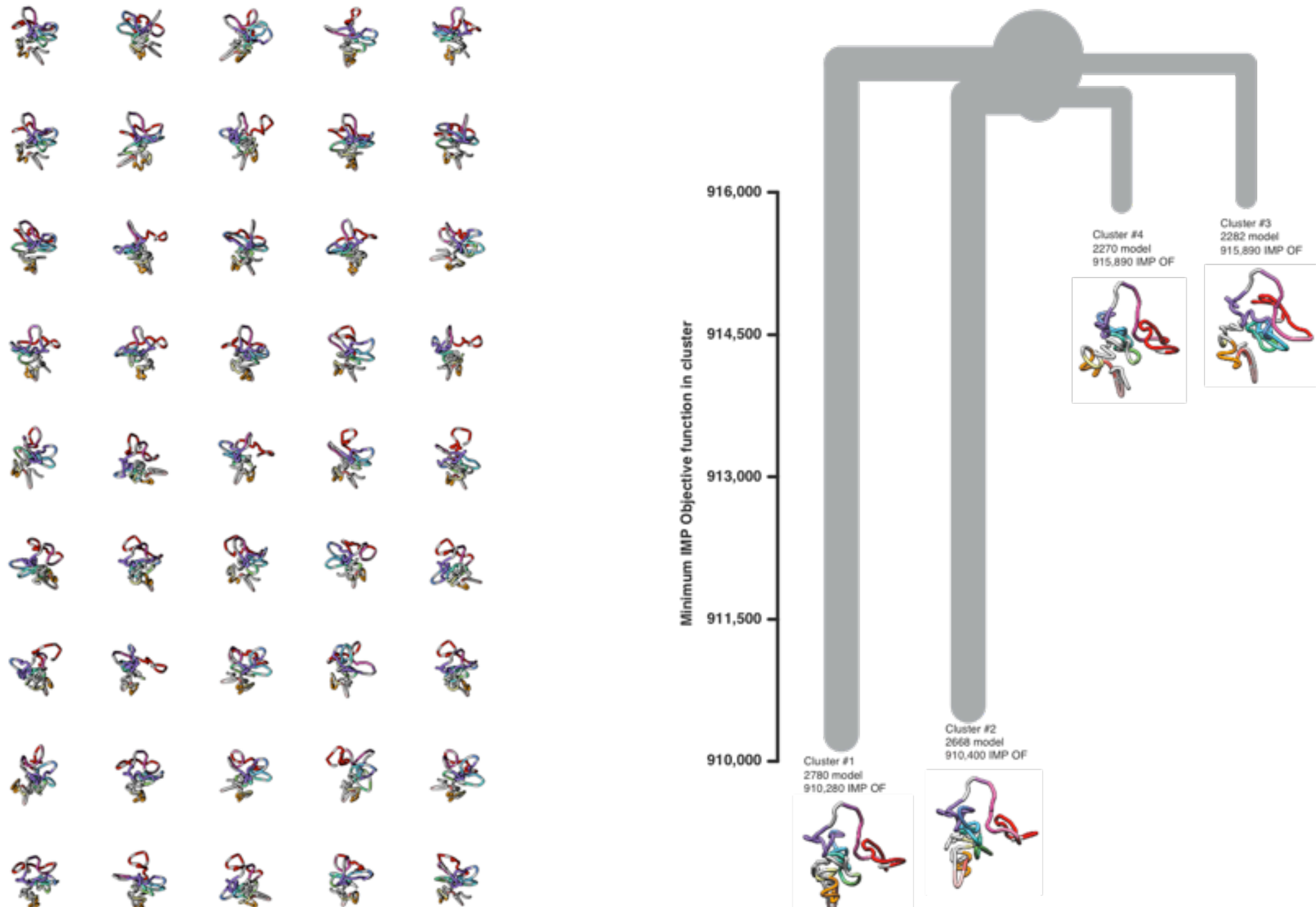
70 fragments
1,049 restraints



Optimization

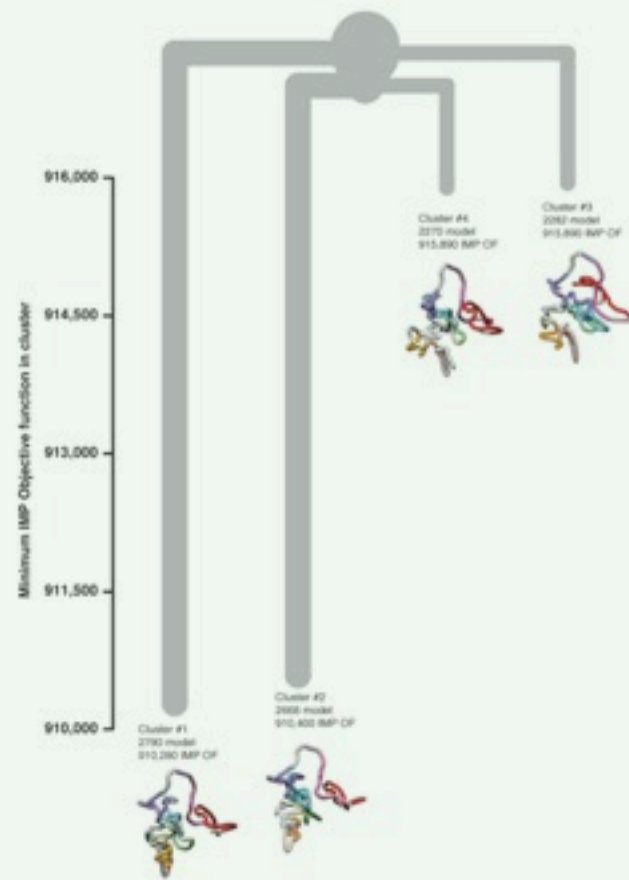
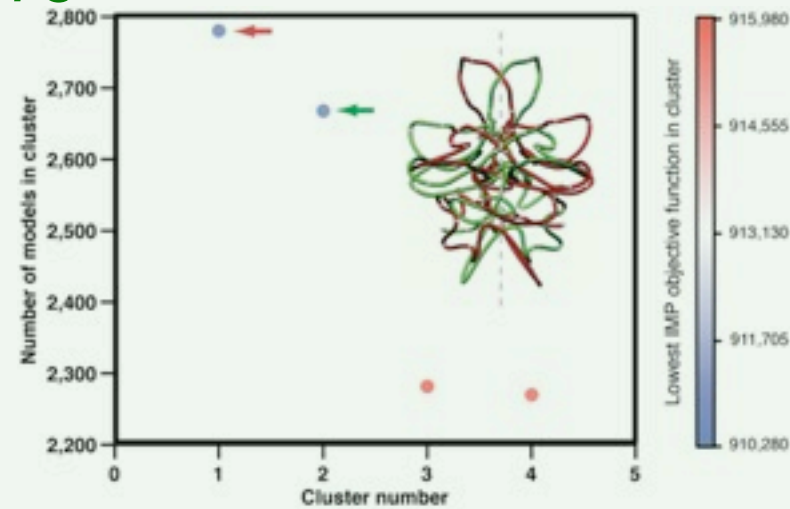


Clustering

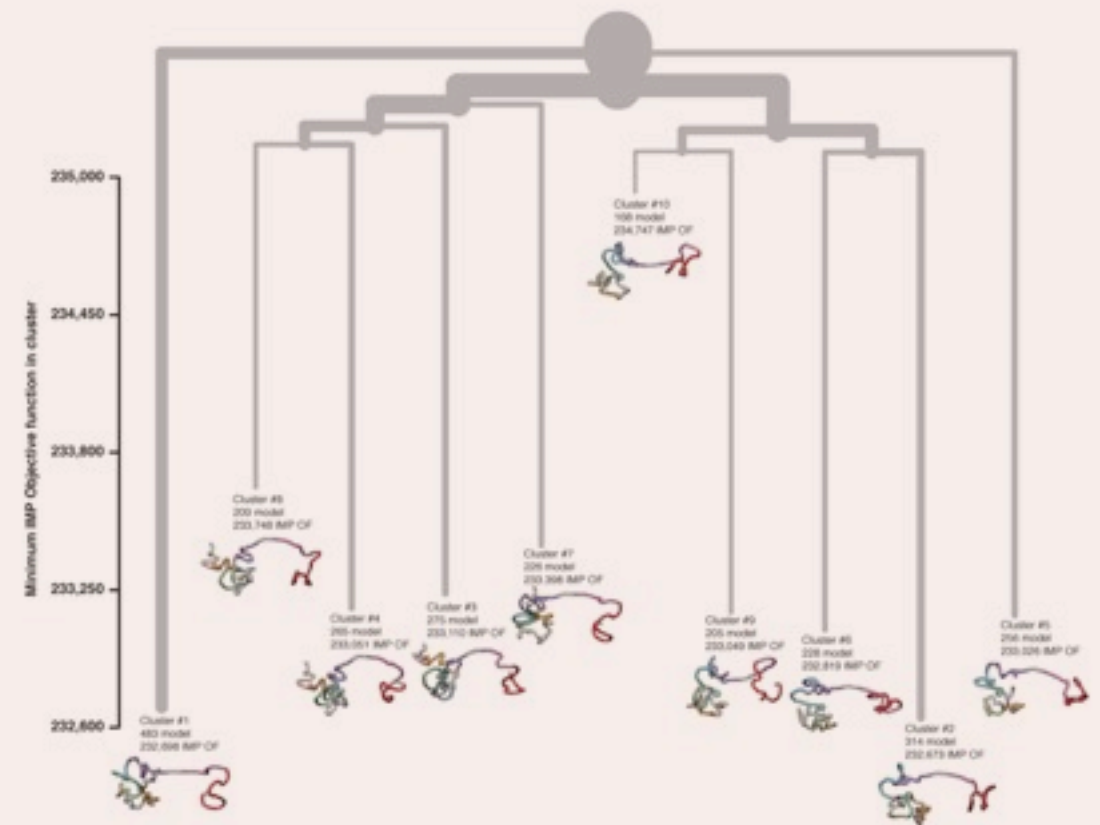
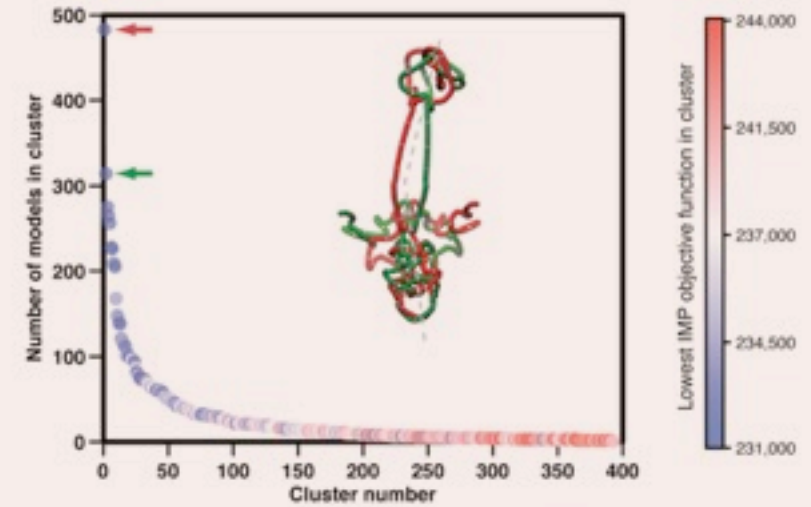


Not just *one* solution

GM12878

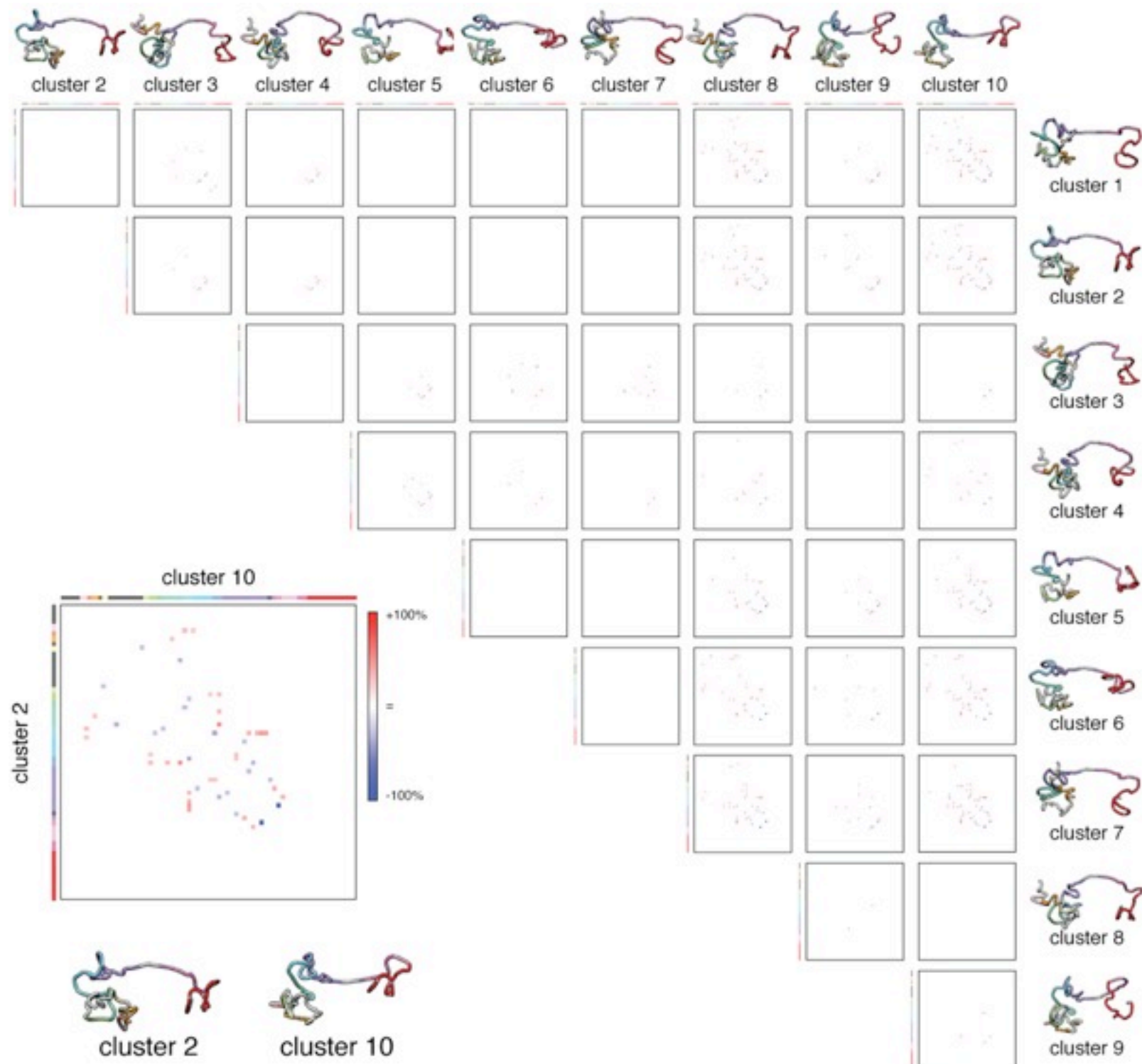


K562



Not just *one* solution

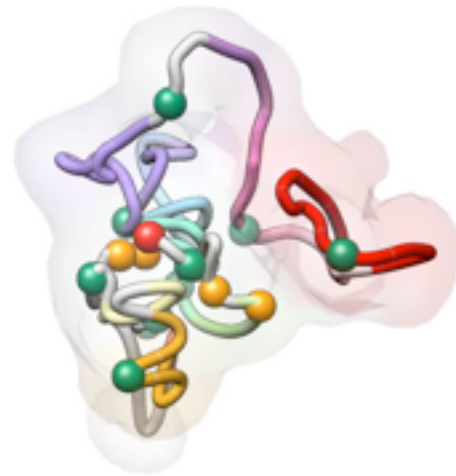
and we can de-convolute them!



Consistency

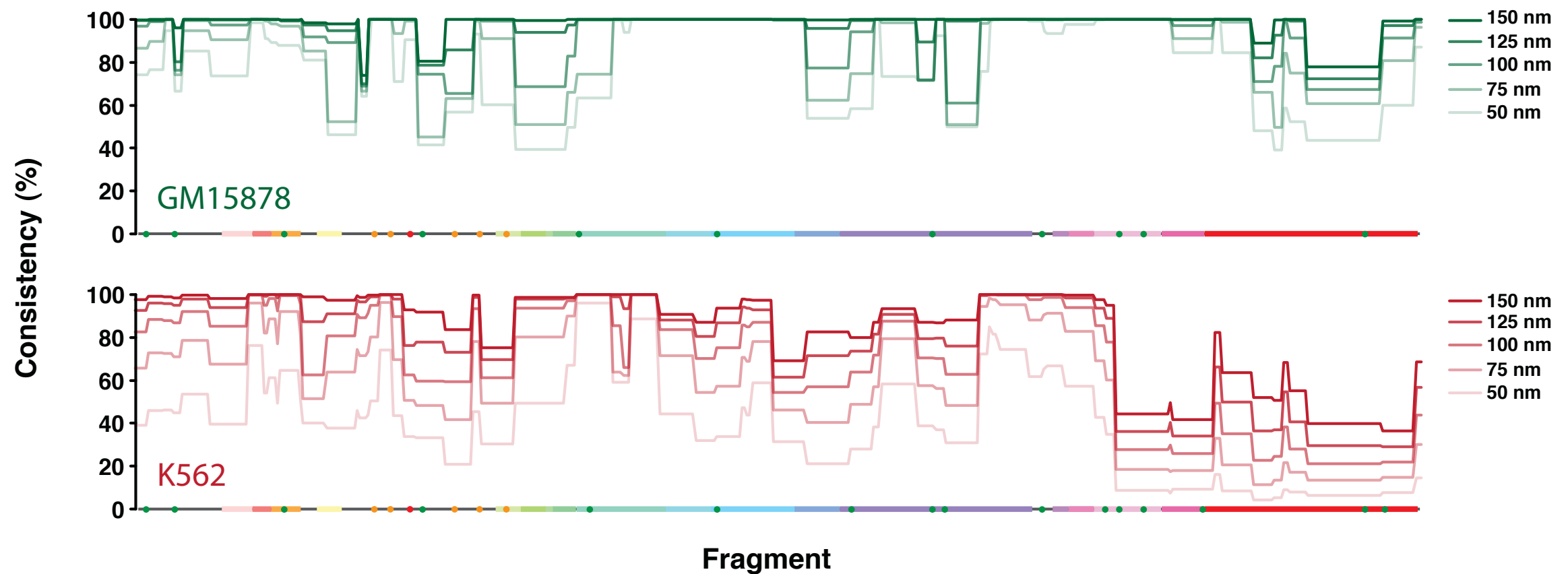
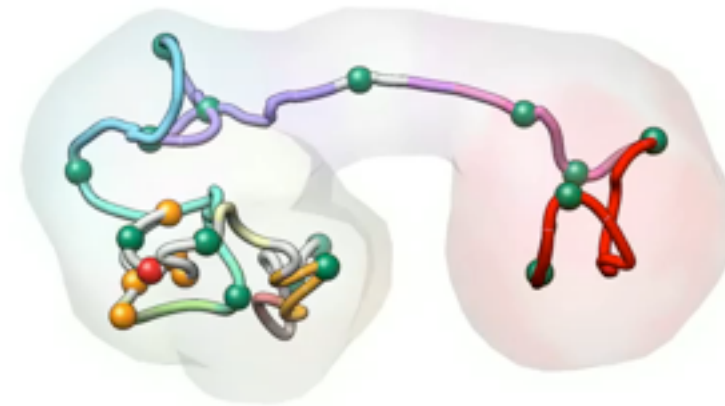
GM12878

Cluster #1
2780 model



K562

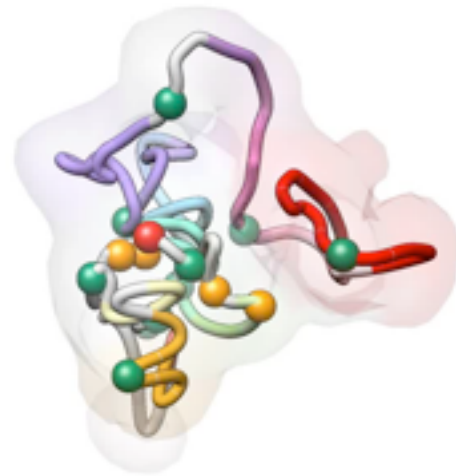
Cluster #2
314 model



Regulatory elements

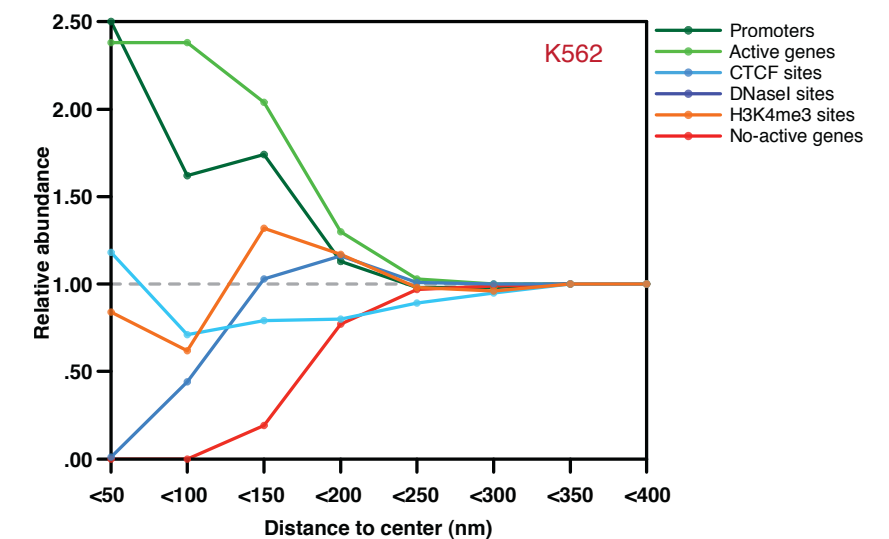
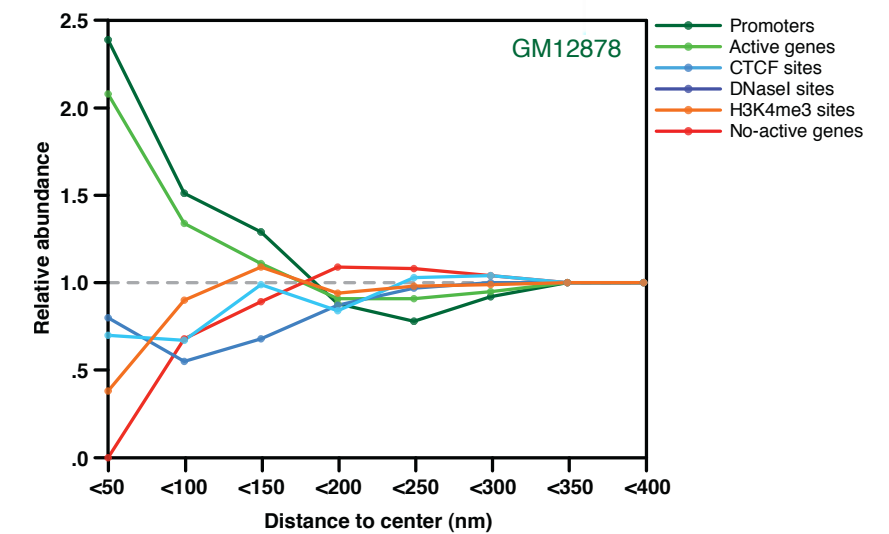
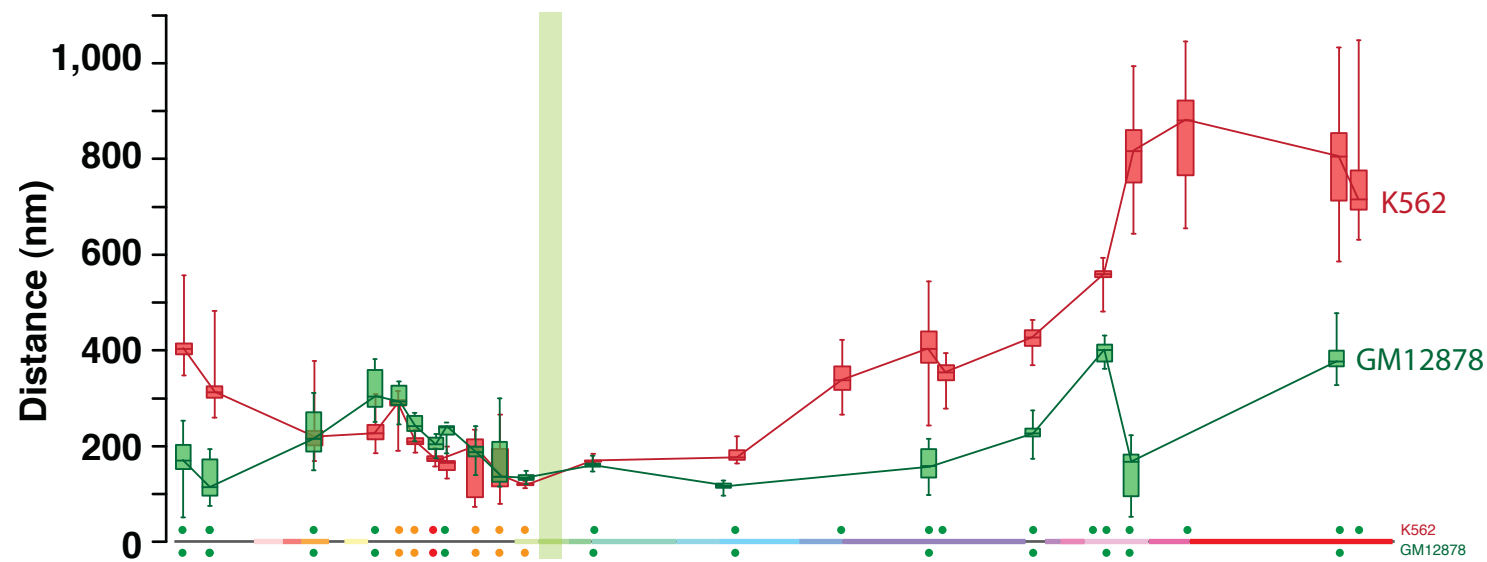
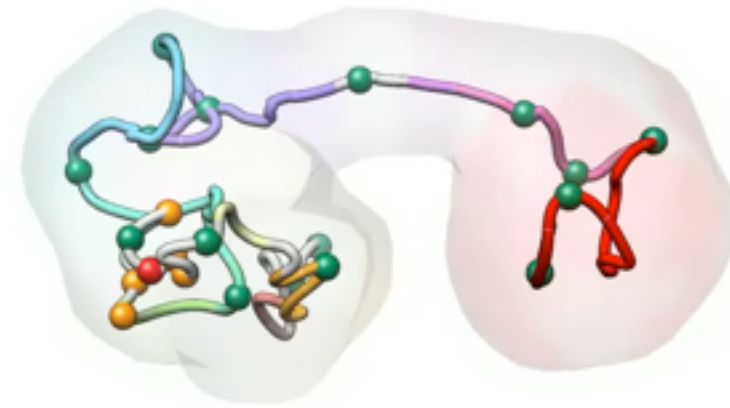
GM12878

Cluster #1
2780 model



K562

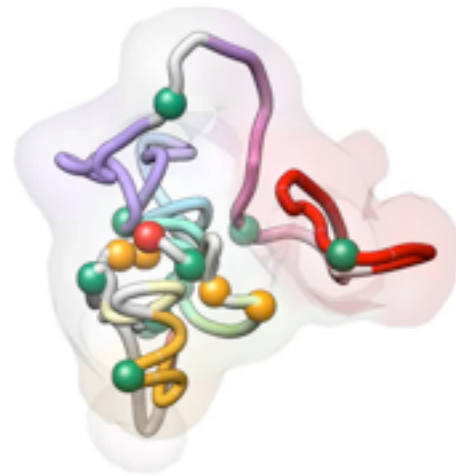
Cluster #2
314 model



Compactness

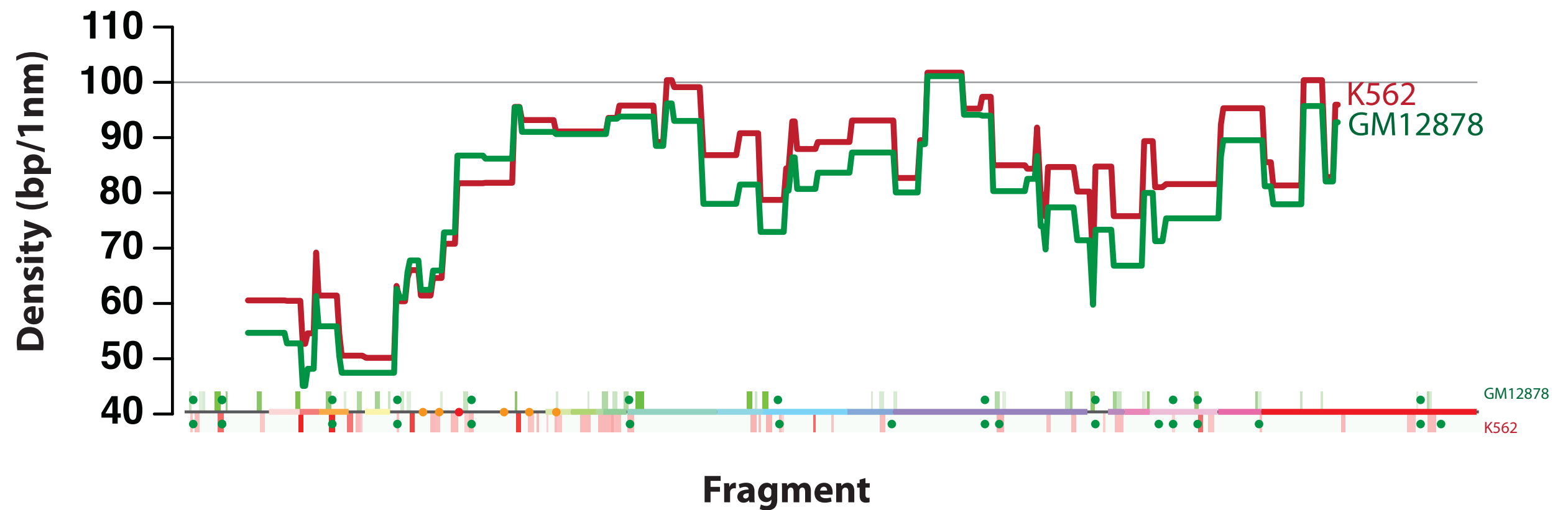
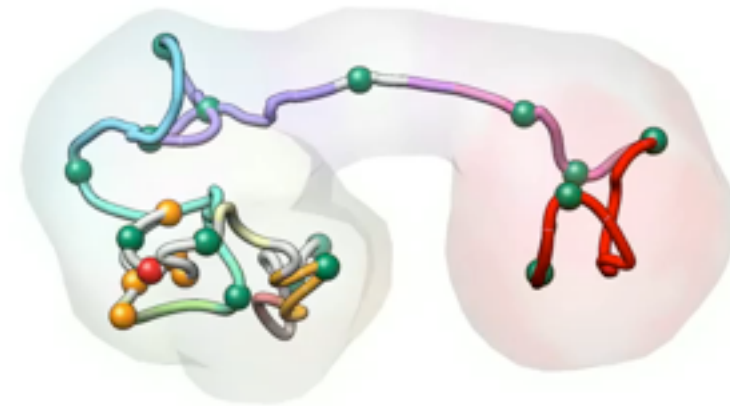
GM12878

Cluster #1
2780 model



K562

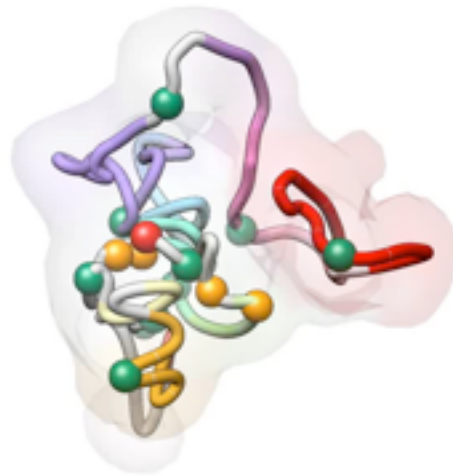
Cluster #2
314 model



Multi-loops

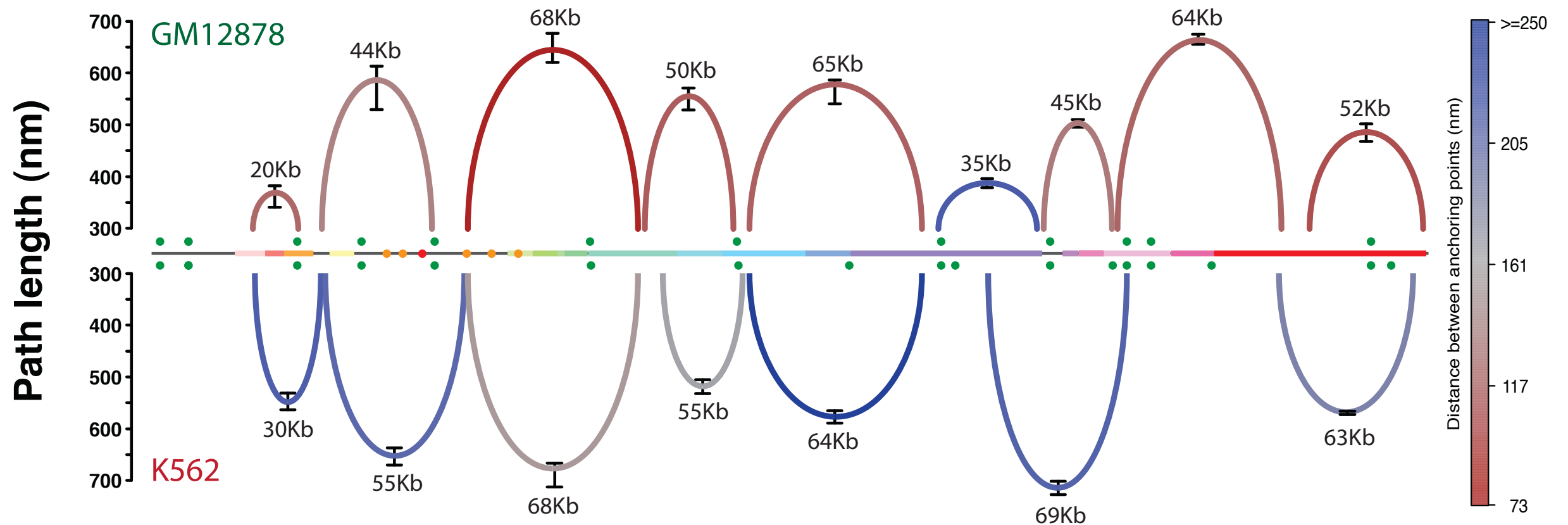
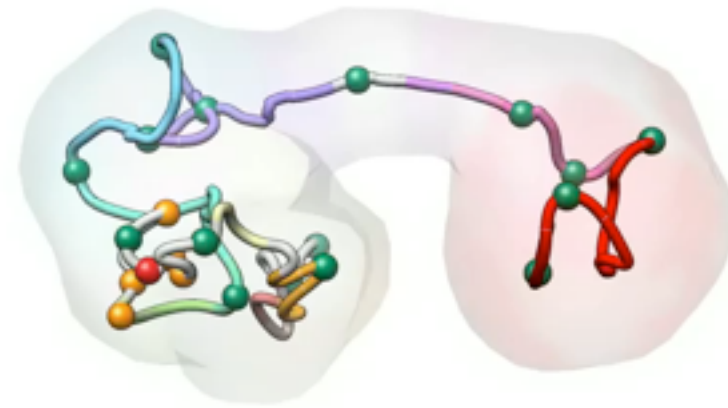
GM12878

Cluster #1
2780 model



K562

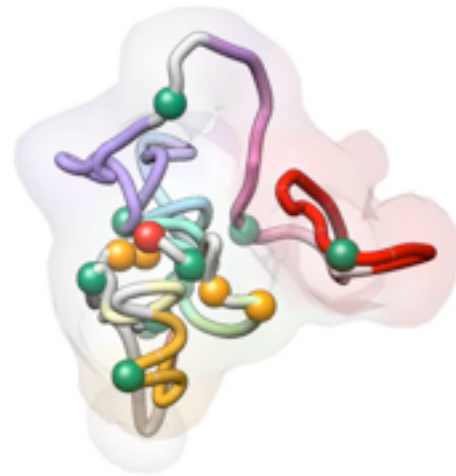
Cluster #2
314 model



Expression

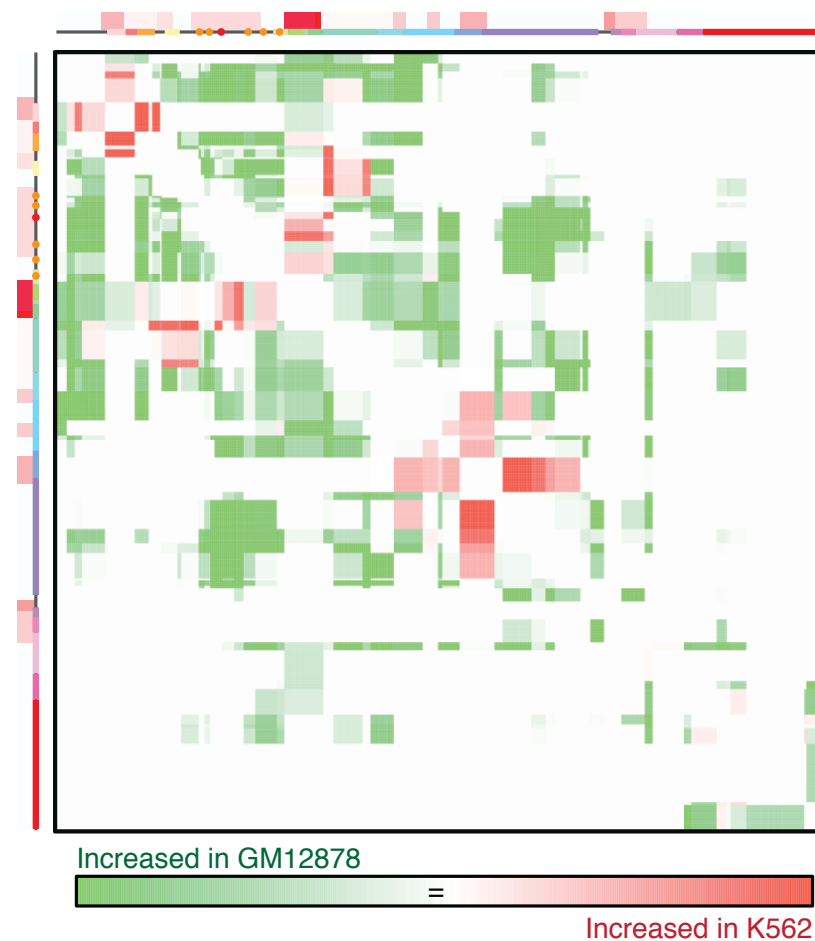
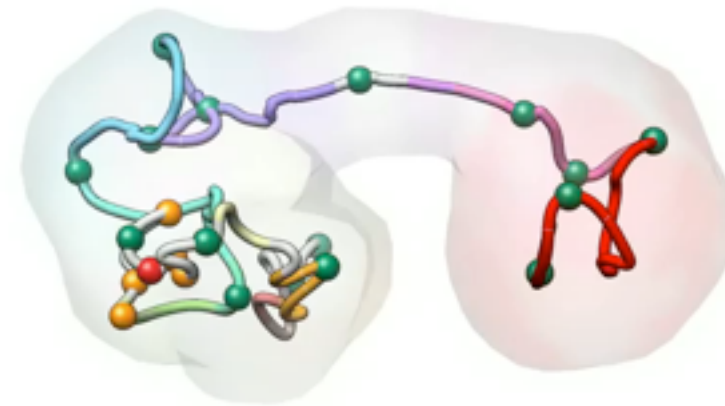
GM12878

Cluster #1
2780 model



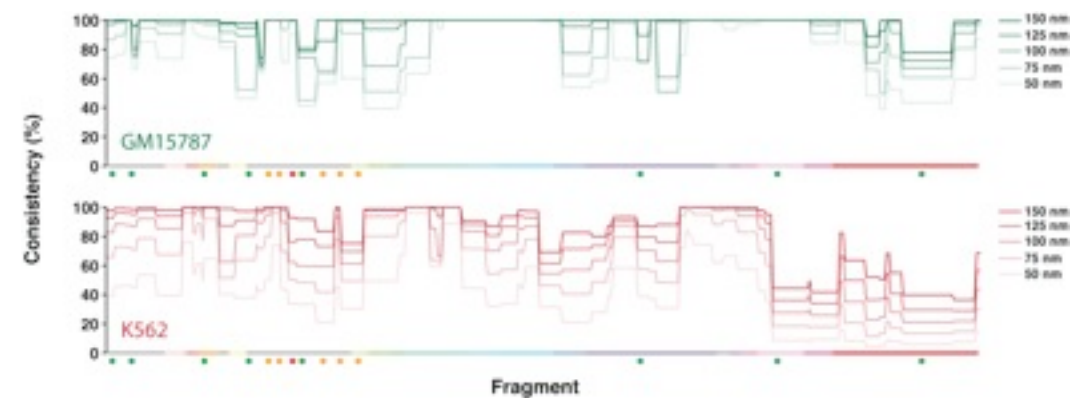
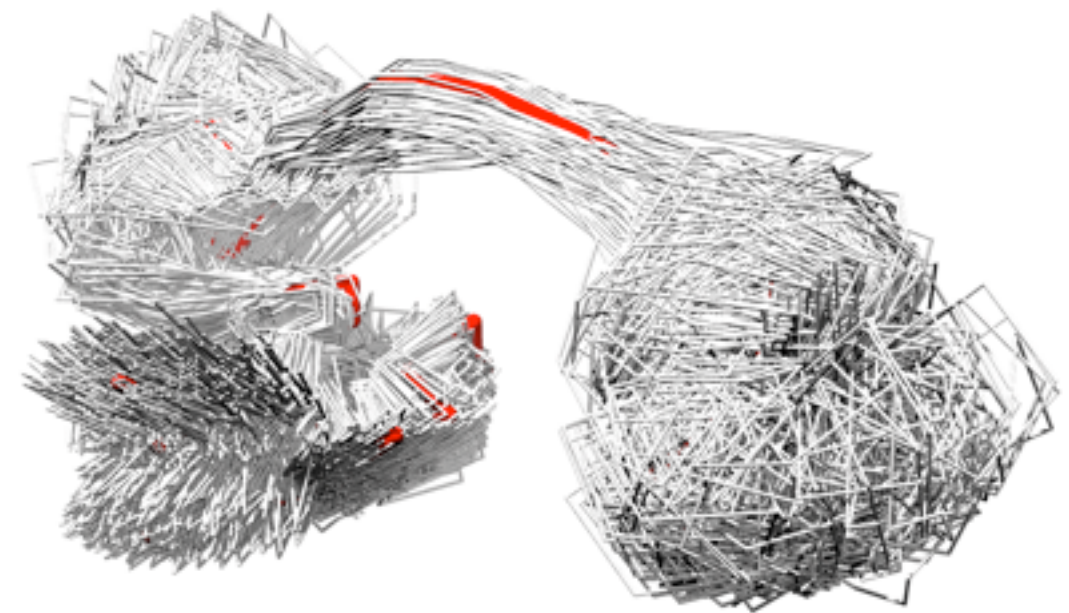
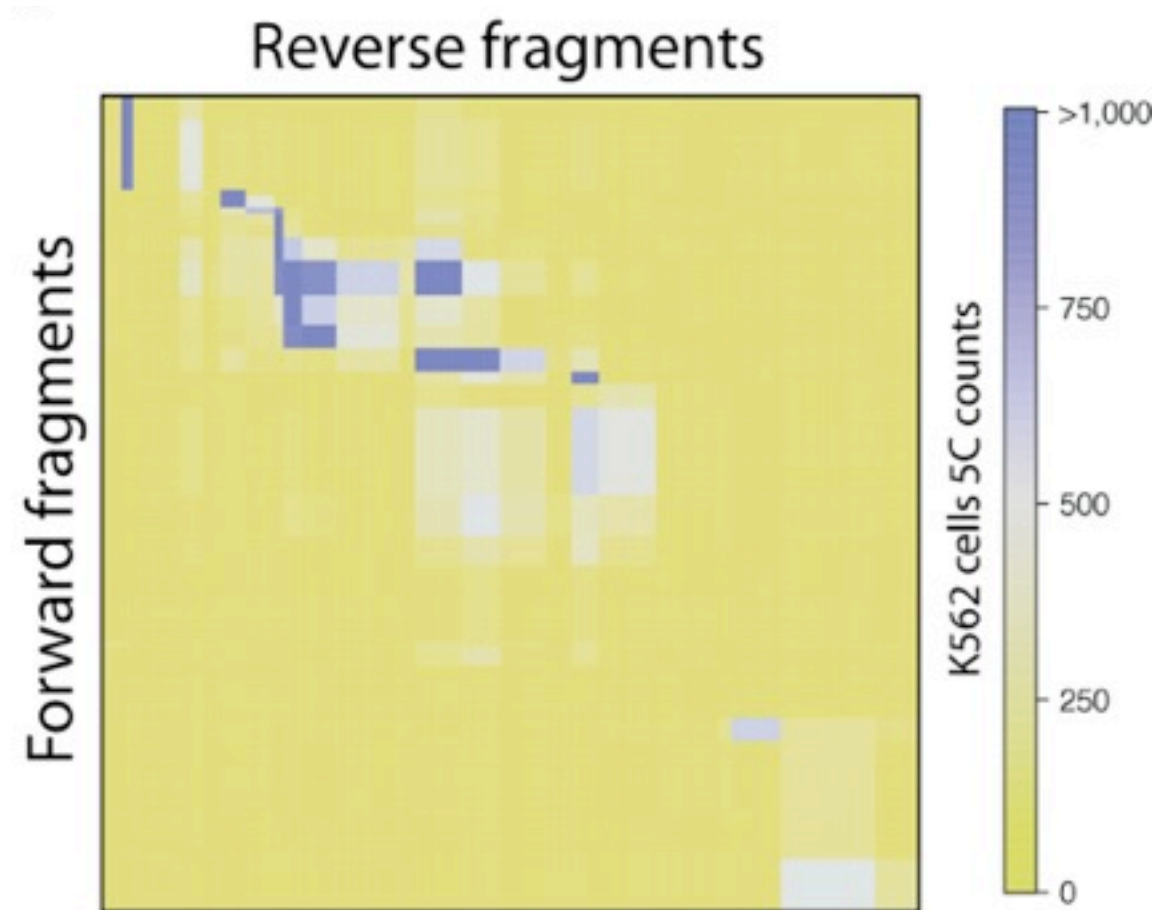
K562

Cluster #2
314 model



Summary

5C data results in comprehensive interaction matrices to build a consistent 3D model



Summary

Models allow for 5C data de-convolution



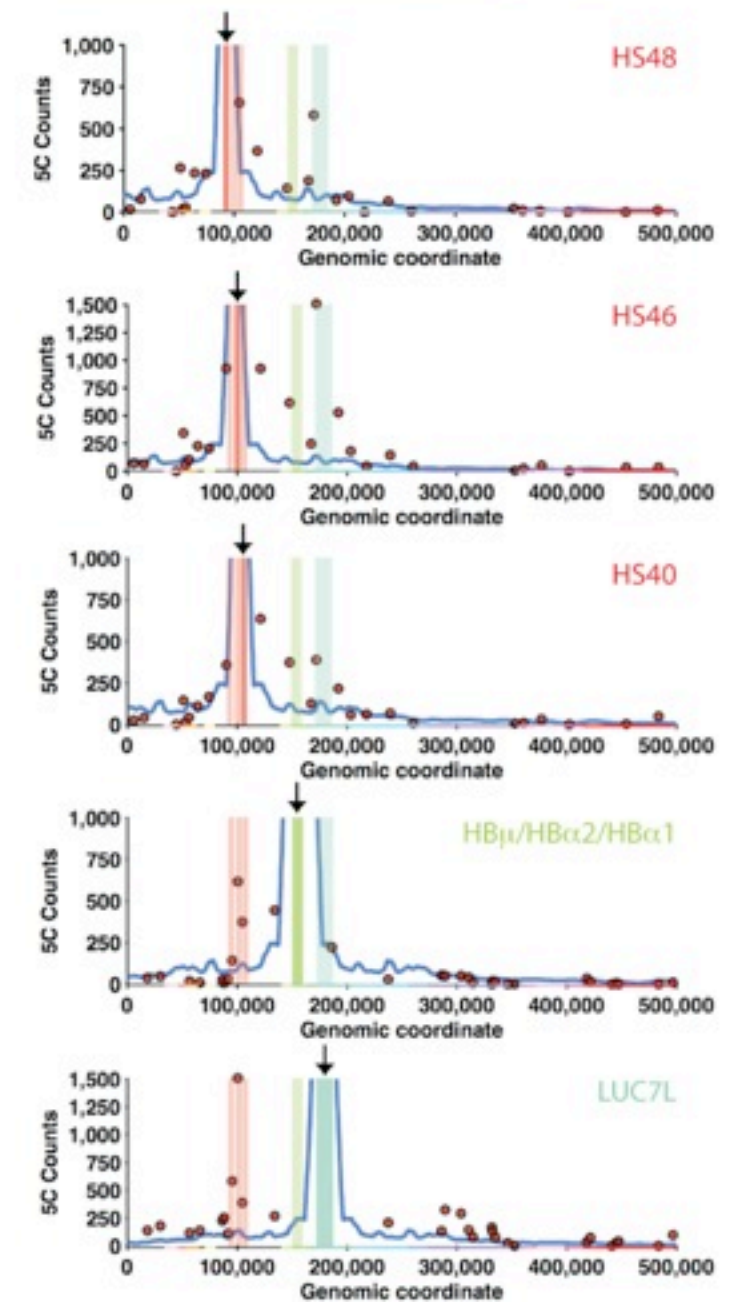
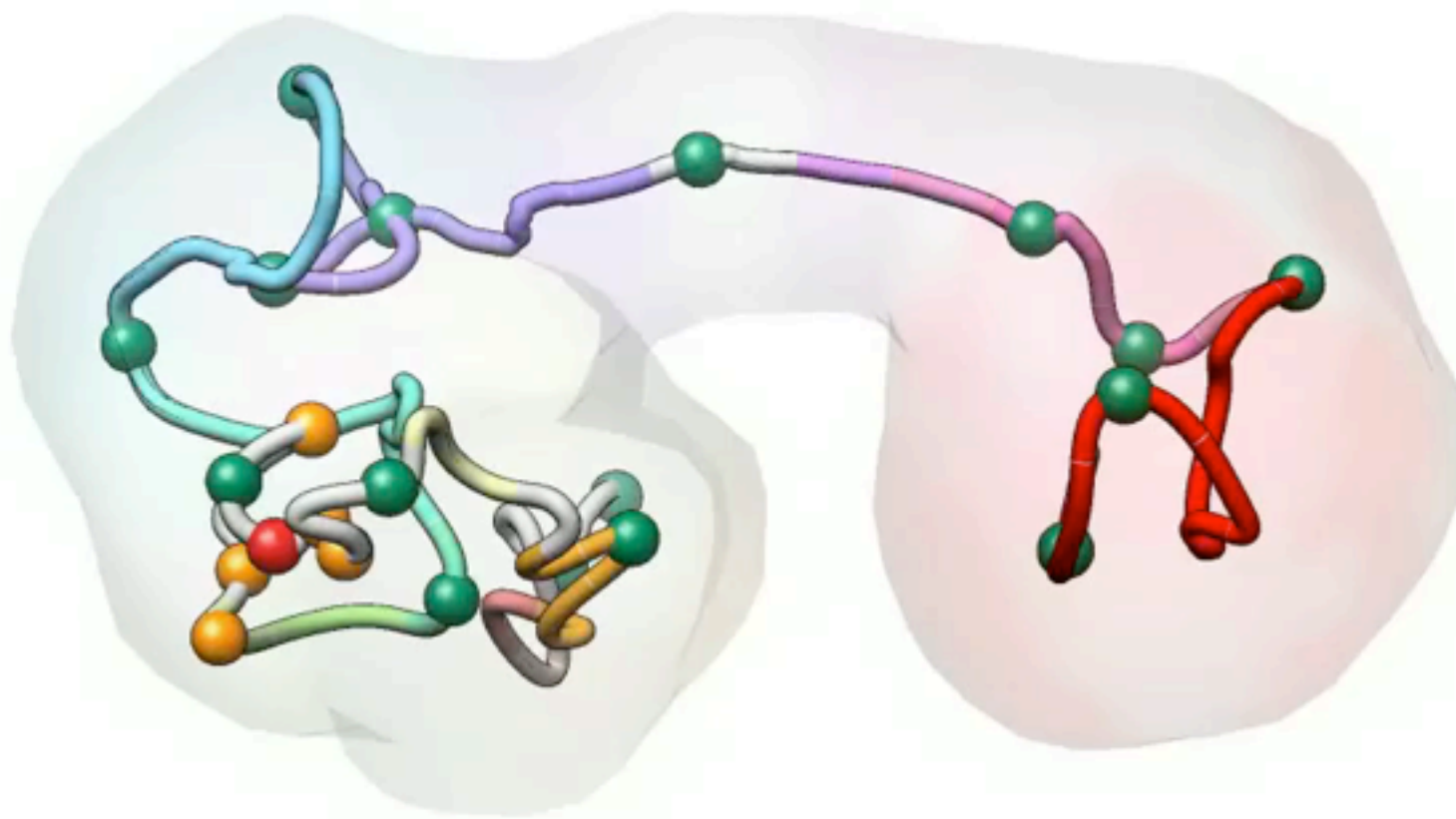
Summary

Models allow for 5C data de-convolution



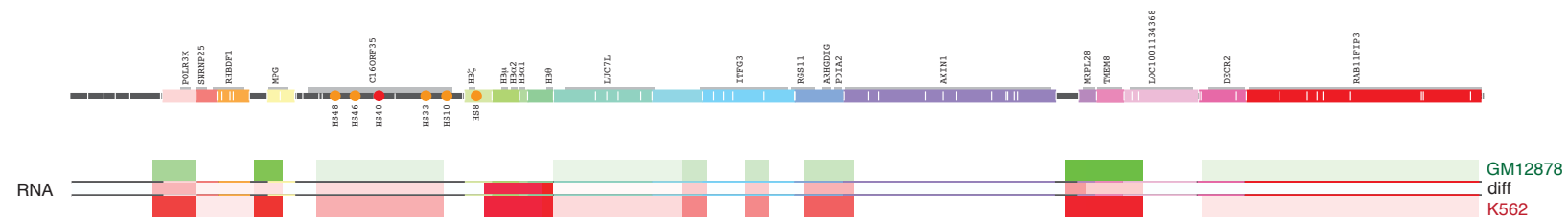
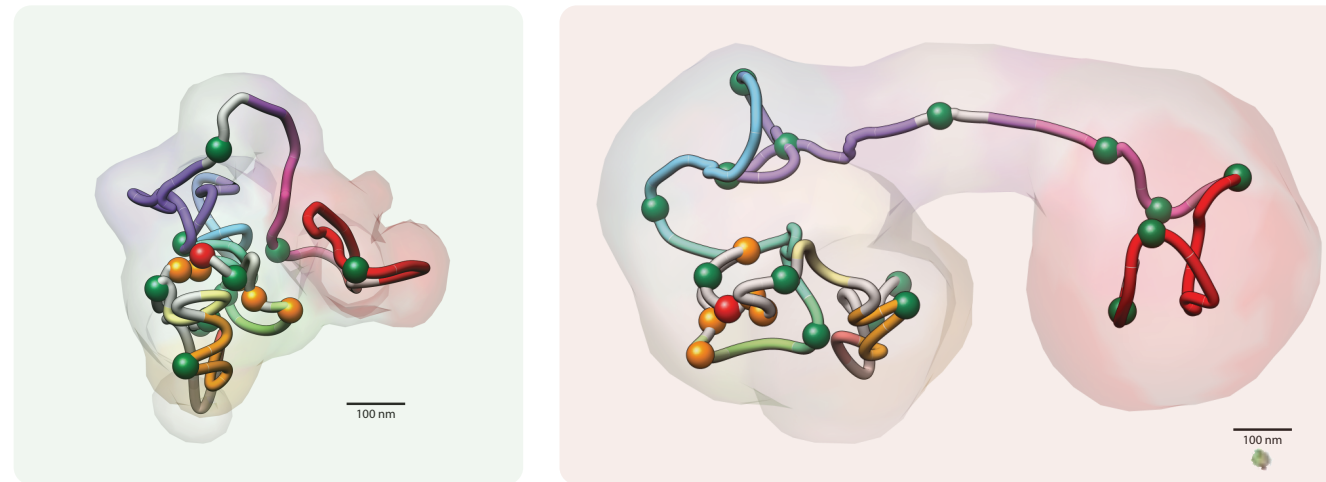
Summary

Selected models reproduce known (**and new**) interactions



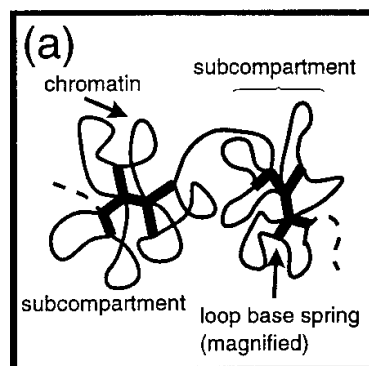
Summary

Large-scale changes in conformation correlate with gene expression of resident genes

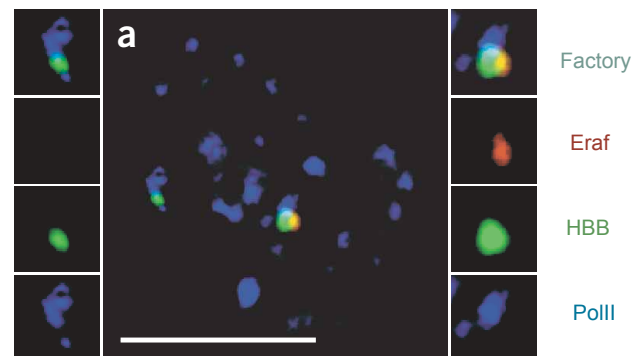


Summary

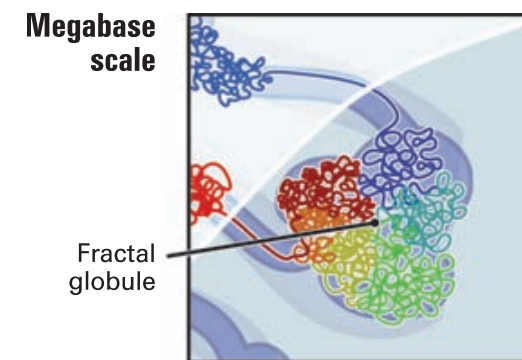
“Chromatin Globule” model



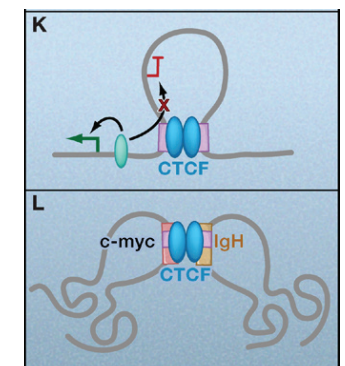
Münkel et al. JMB (1999)



Osborne et al. Nat Genet (2004)



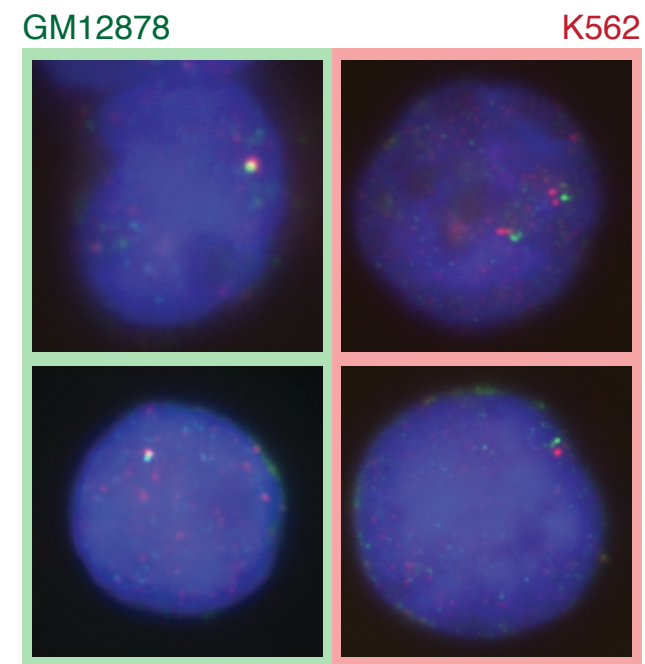
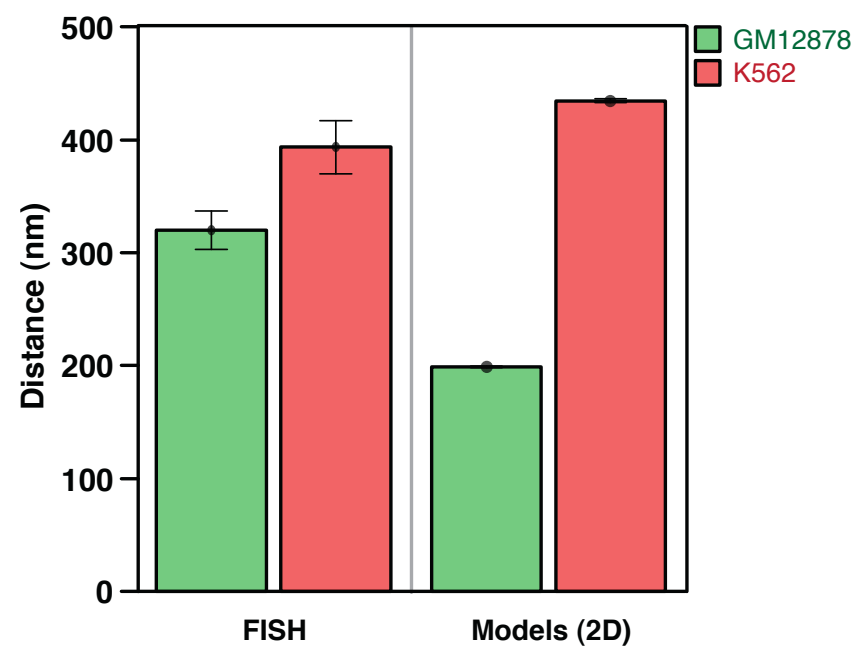
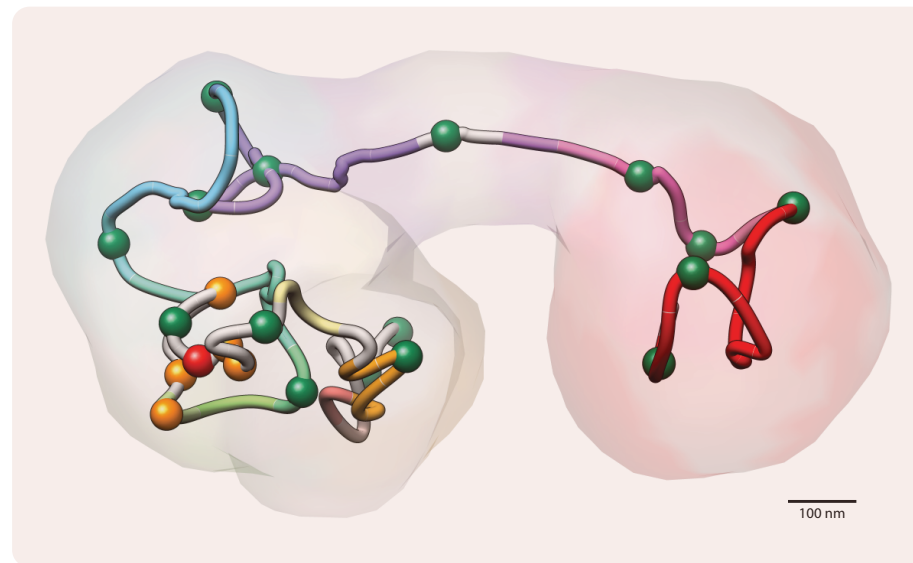
Lieberman-Aiden et al. Science (2009)



Phillips and Corces. Cell (2009)

Initial validation

FISH experiments by the Lawrence Lab (UMASS)



Jeanne Lawrence (UMASS)

Other applications

Caulobacter crescentus whole genome



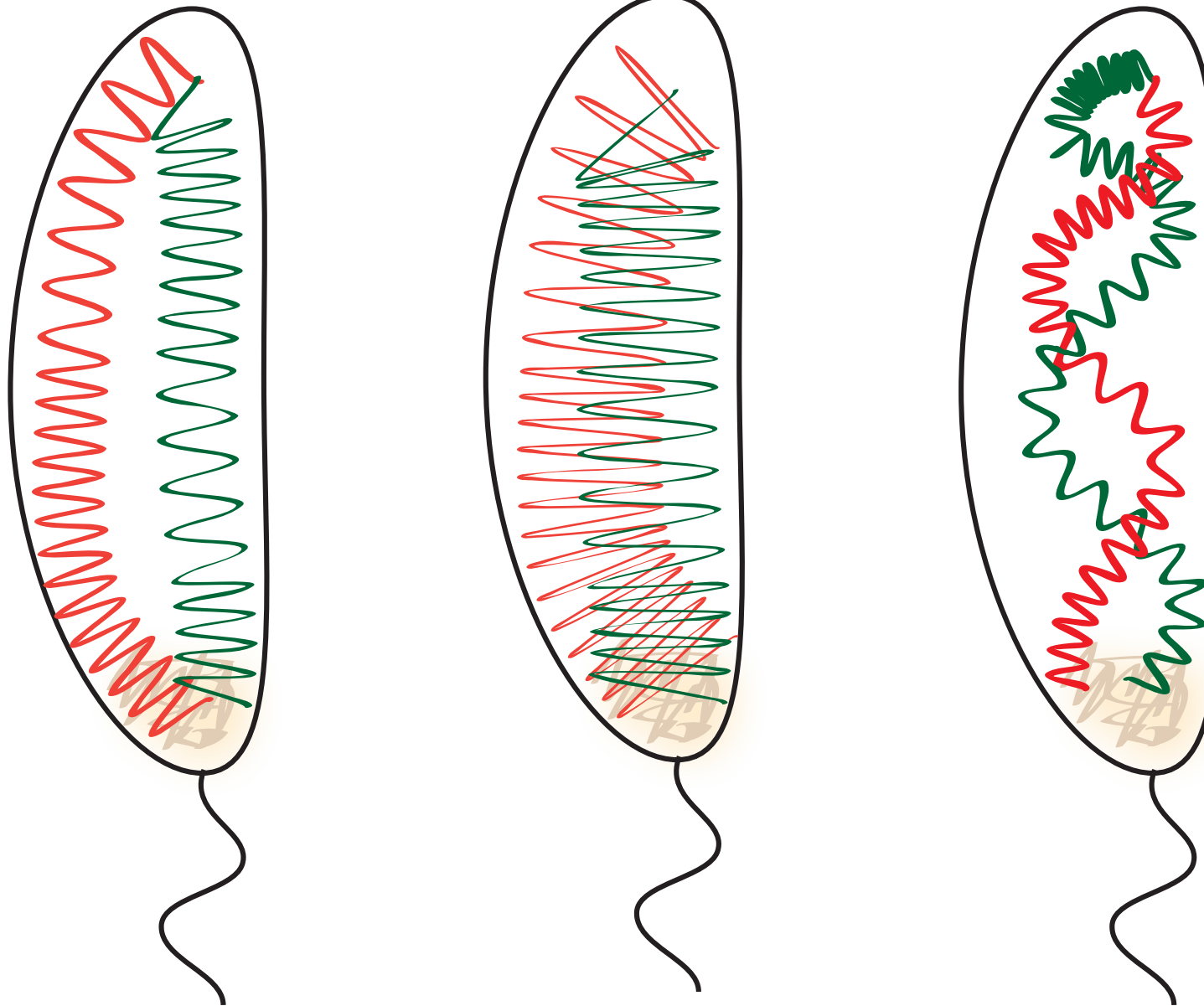
Human Chromosome 21 & mRNA chimeras



Other applications...

Caulobacter crescentus whole genome

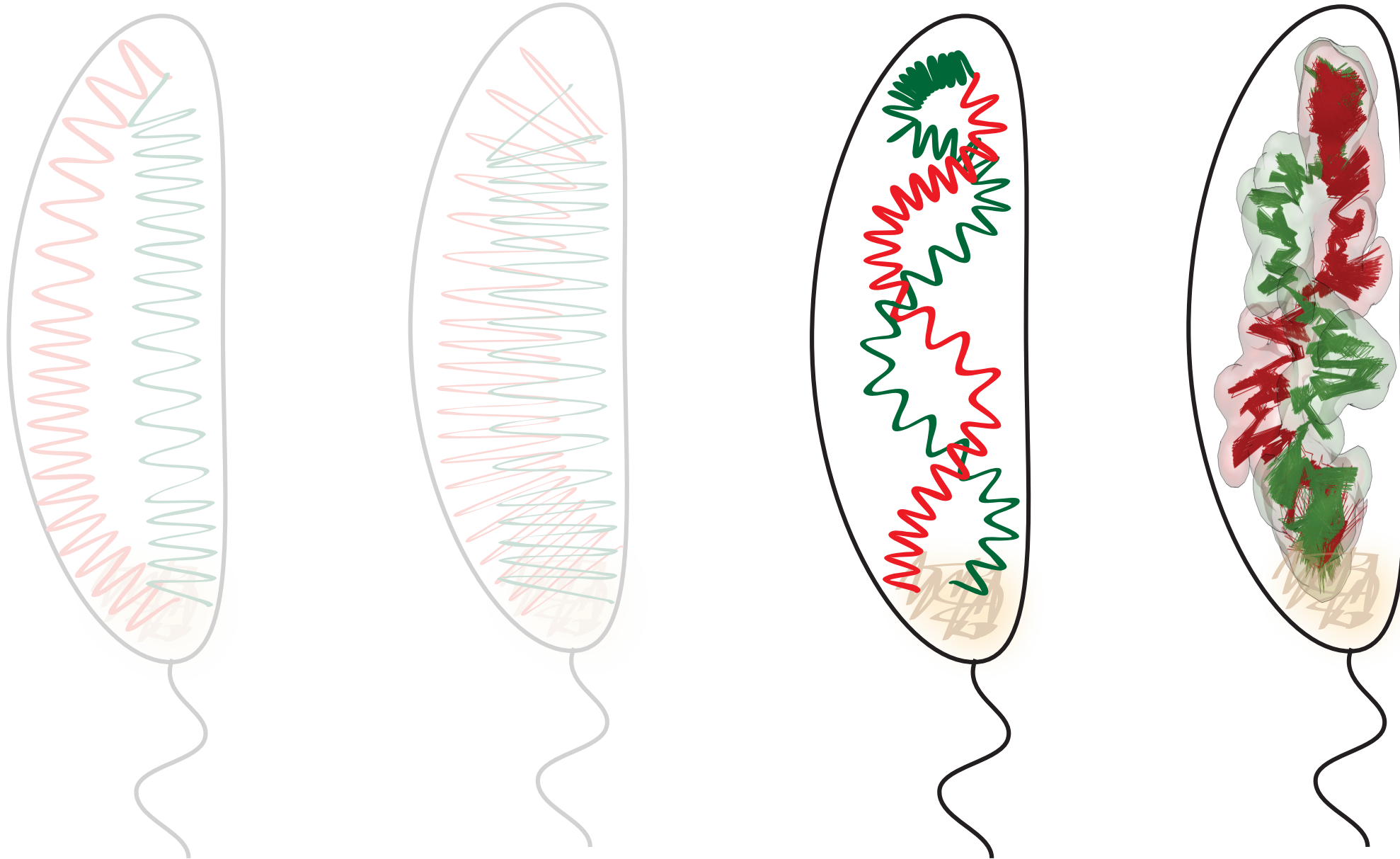
Mark A. Umbarger



Other applications...

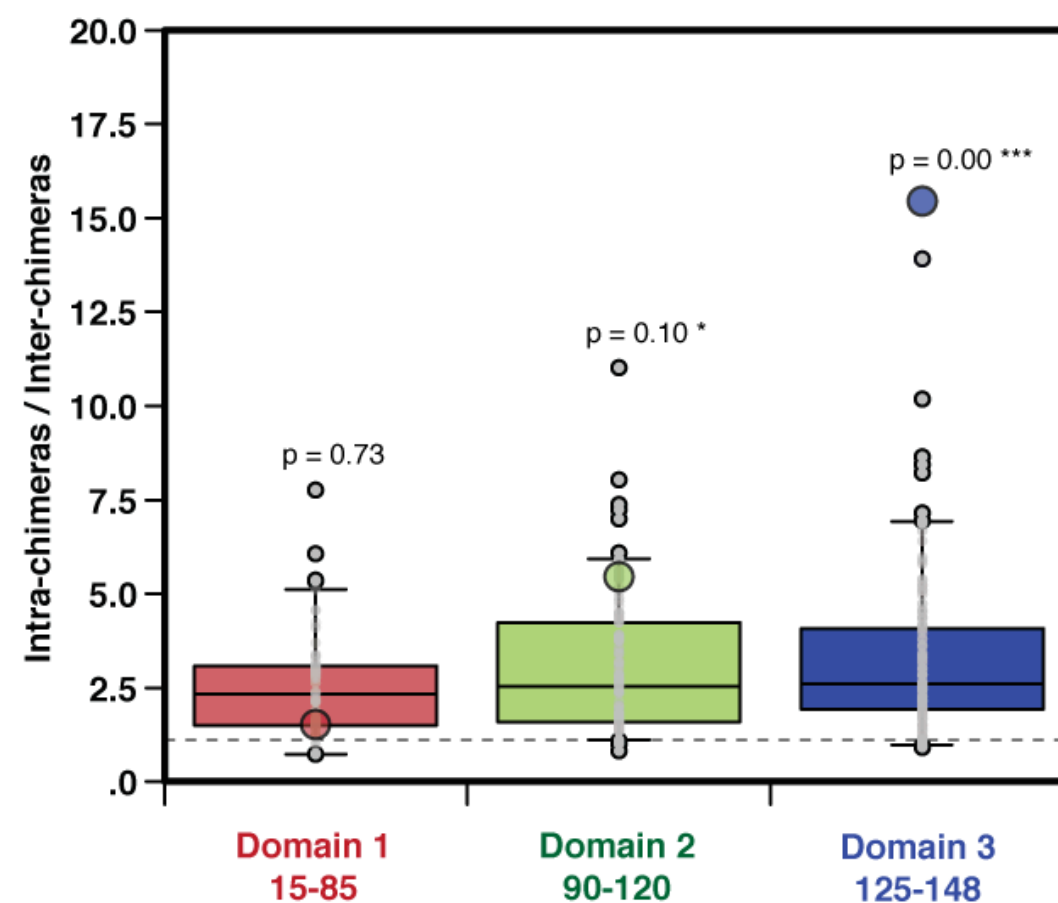
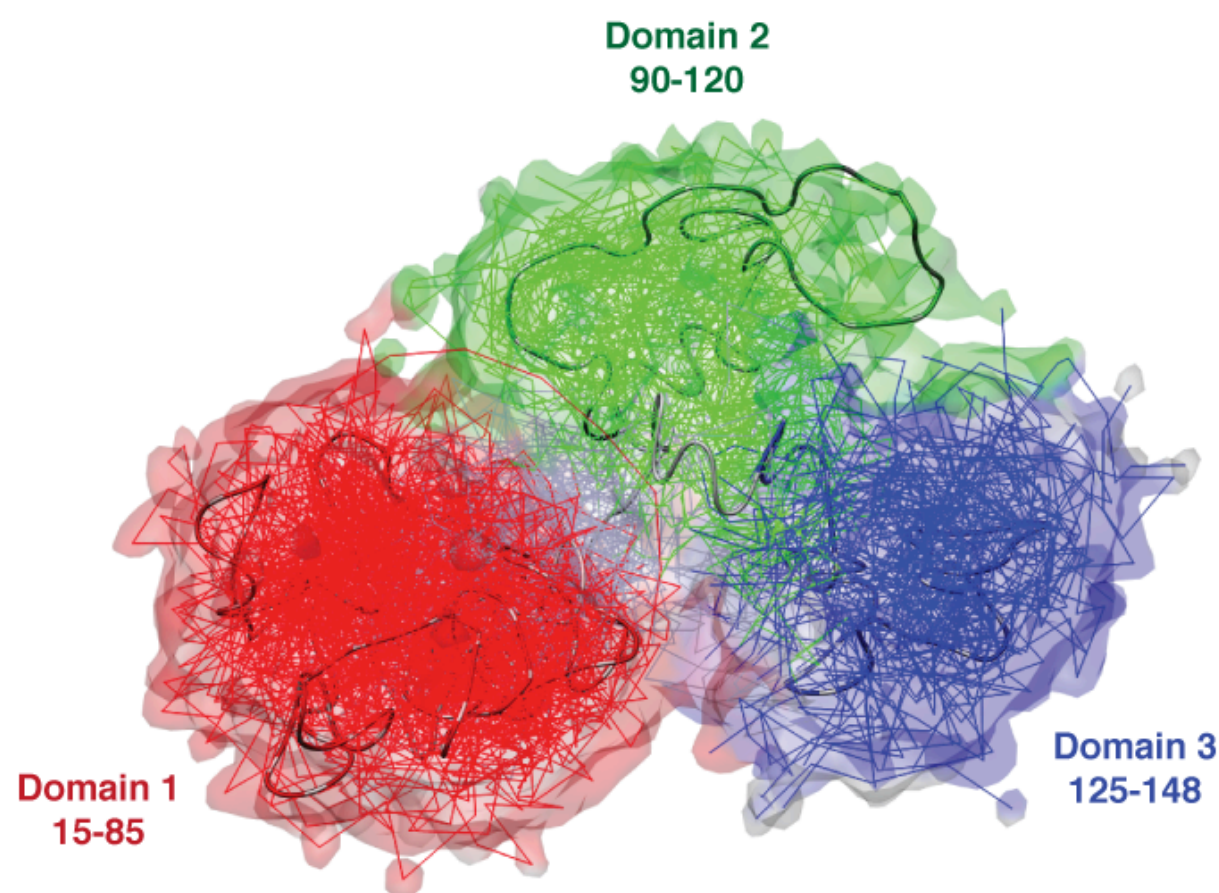
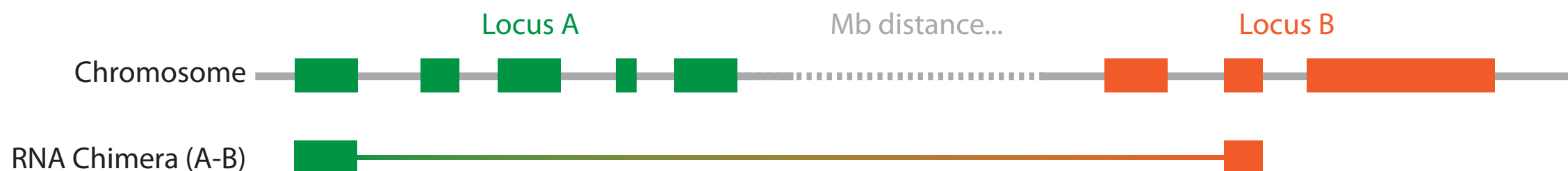
Caulobacter crescentus whole genome

Mark A. Umbarger



Other applications...

Whole (low-res) human chromosome 21 and RNA chimeras
Sarah Djebali



Acknowledgments

Job Dekker

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



Marc A. Marti-Renom

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



Amartya Sanyal

Postdoctoral Fellow
Dekker Lab



Davide Baù

Postdoctoral fellow
Structural Genomics Unit



Bryan Lajoie

Bioinformatician
Dekker Lab



Emidio Capriotti

Postdoctoral fellow
Structural Genomics Unit

