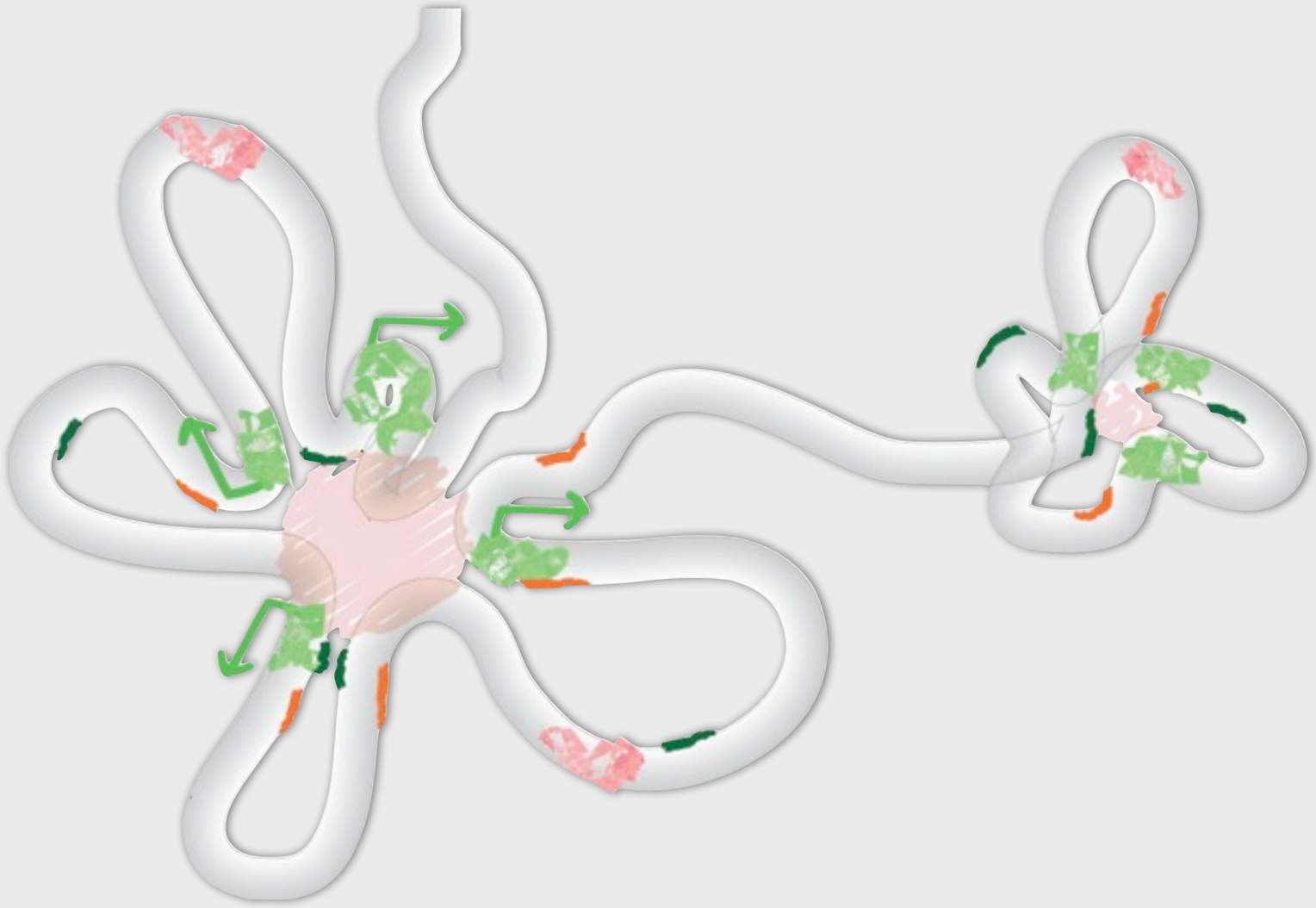
3D folding of chromosomal domains in relation to gene expression



Marc A. Marti-Renom

http://sgu.bioinfo.cipf.es

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

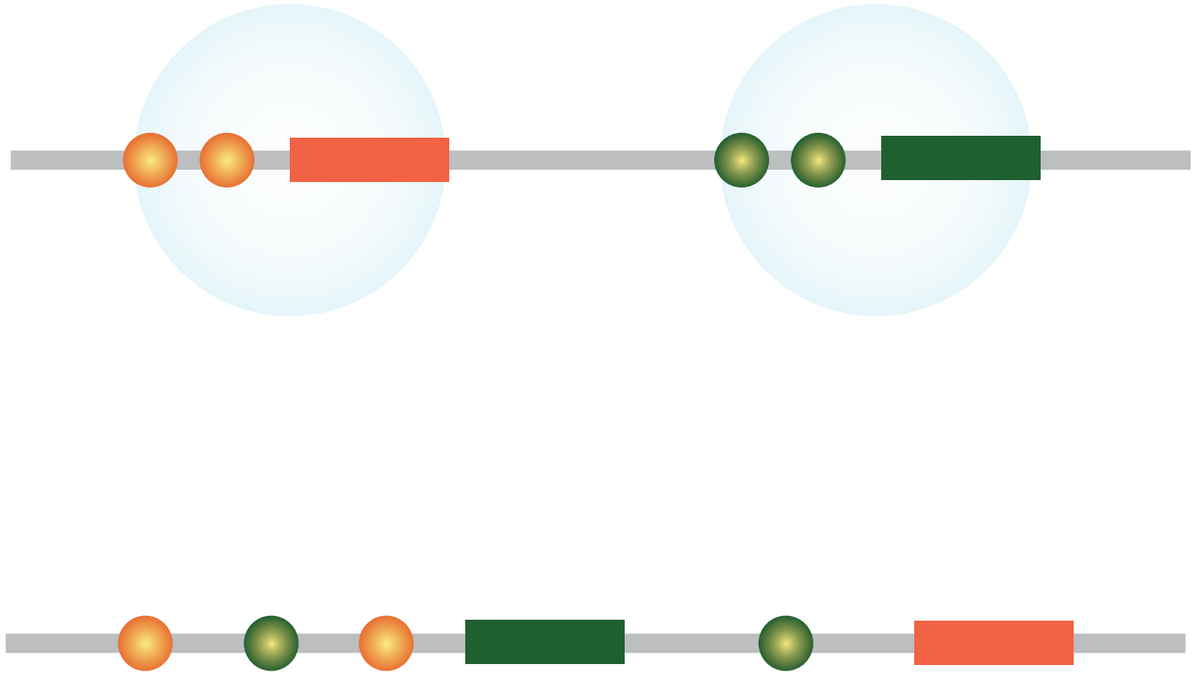
PRINCIPE FELIPE
CENTRO DE INVESTIGACION

UMASS MEDICAL SCHOOL

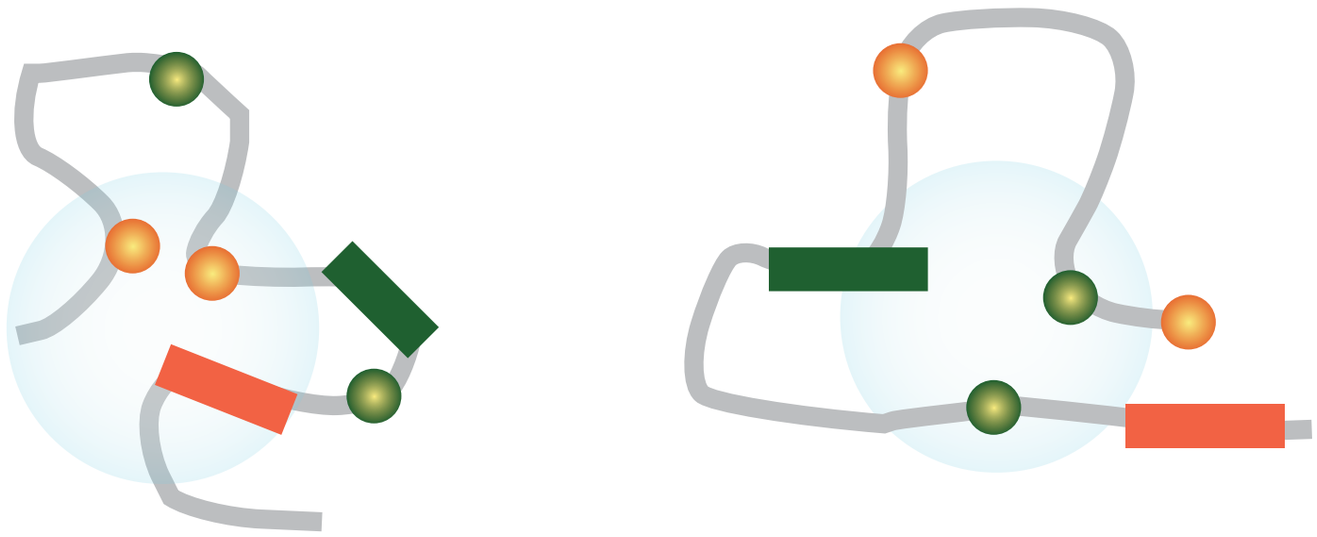
Aim

Can we relate structure and expression?

Simple genomes



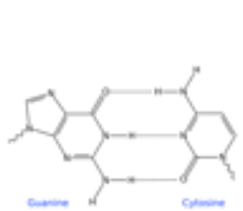
Complex genomes

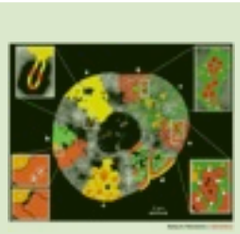


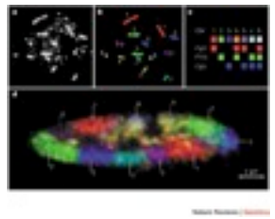
Resolution

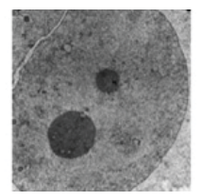
Limited knowledge...

Knowledge









DNA length

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

Volume

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

Time

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

Resolution

10^{-3} 10^{-2} 10^{-1} µ

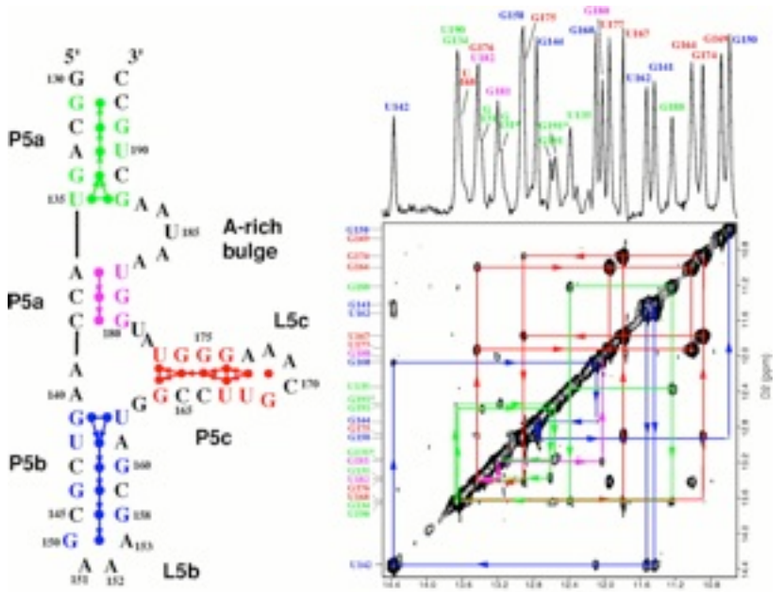
Adapted from:
Langowski and Heermann. Semin Cell Dev Biol (2007) vol. 18 (5) pp. 659-67

Structure determination

Integrative Modeling Platform

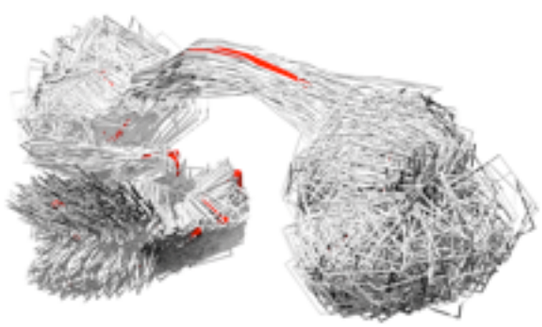
http://www.integrativemodeling.org

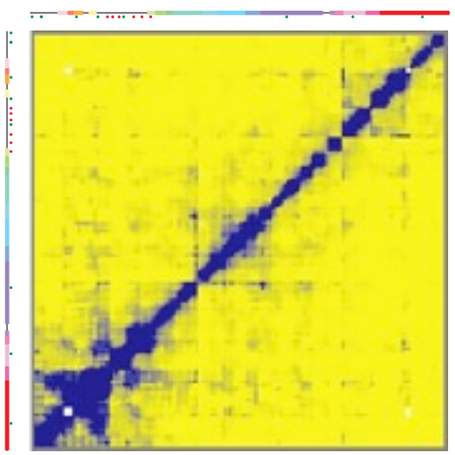
Alber et al. Nature (2007) vol. 450 (7170) pp. 683-94



Biomolecular structure determination

2D-NOESY data





Chromosome structure determination

5C data

Human α-globin domain

ENm008 genomic structure and environment

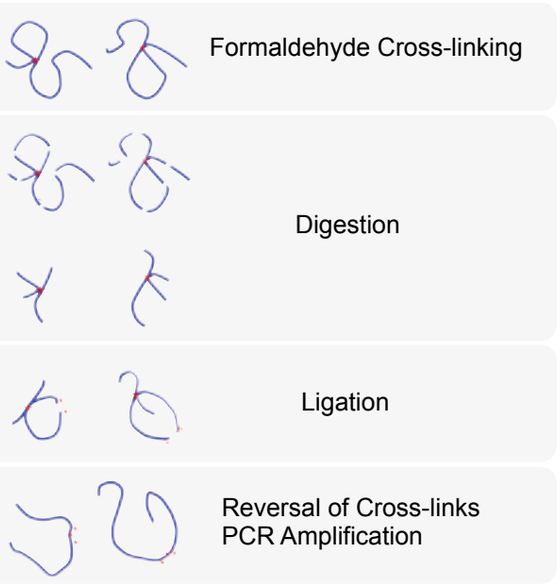
ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816

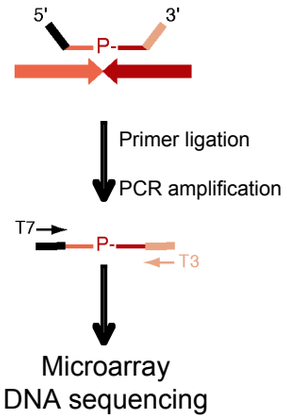
The ENCODE data for ENm008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute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

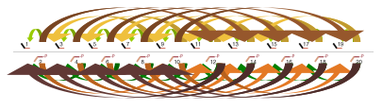
5C experiments

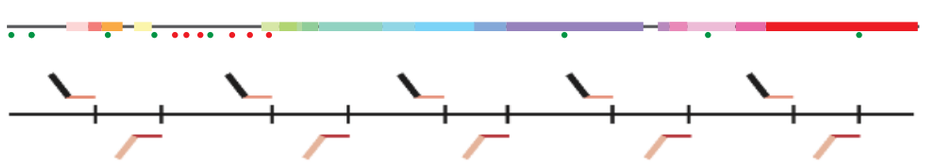
http://my5C.umassmed.edu

B. R. Lajoie, N. L. van Berkum, A. Sanyal et al., Nat Methods 6 (10), 690 (2009).









Grow GM12878 and K562 cells

↓

Perform 3C analysis

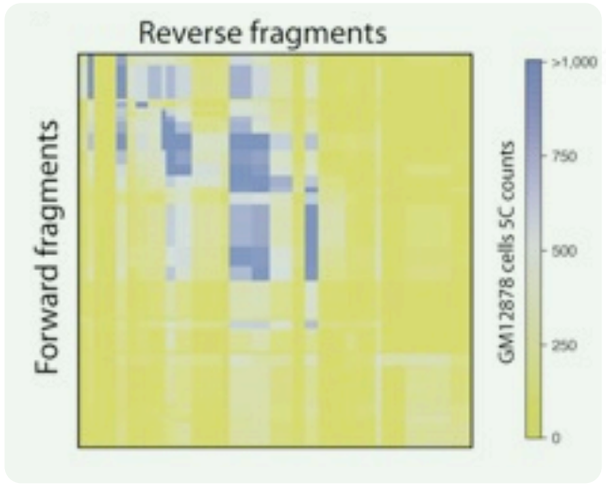
↓

Perform 5C analysis with 30+25 primers

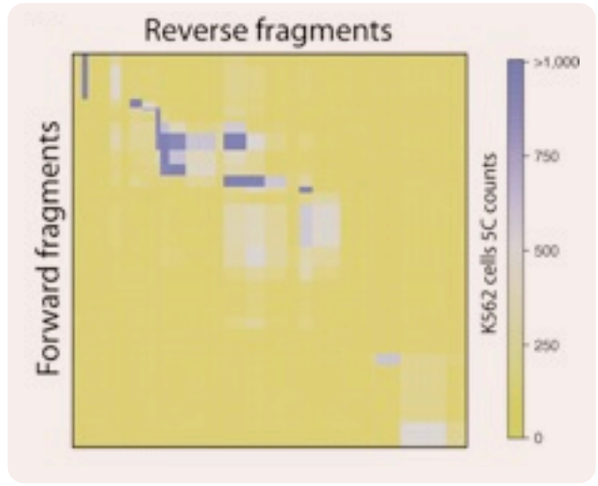
↓

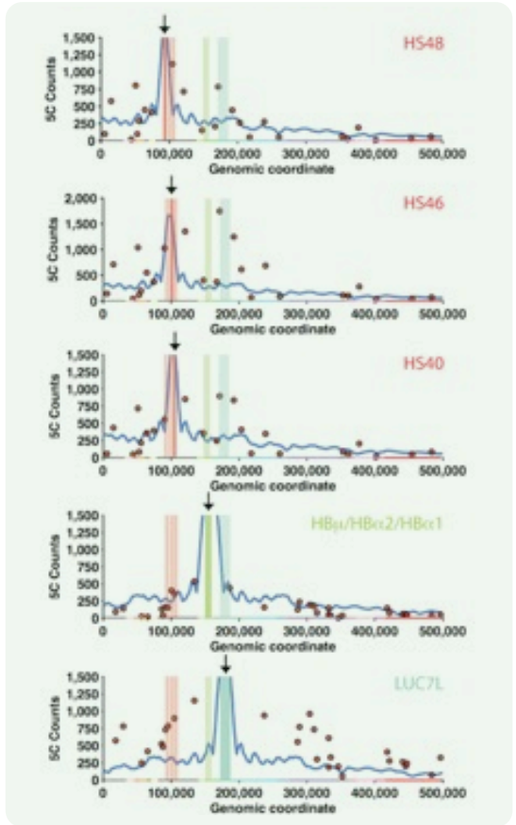
Analyze 5C products by paired-end Solexa sequencing
(131,947 paired end reads per library)

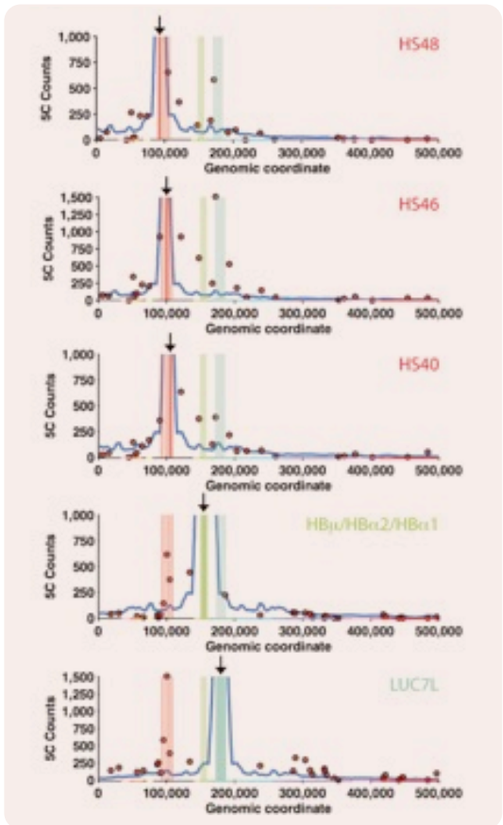
GM12878



K562

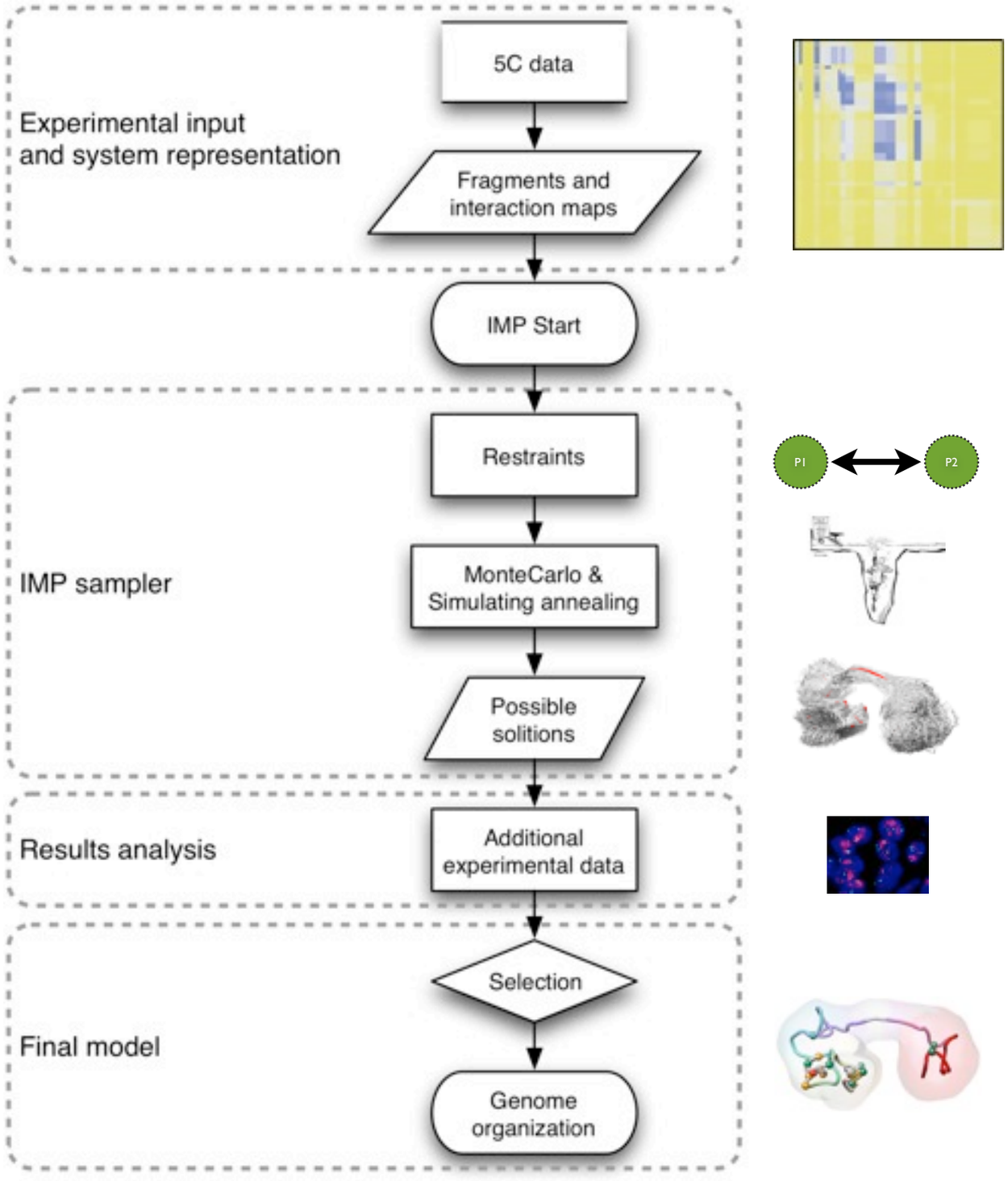






Integrative Modeling

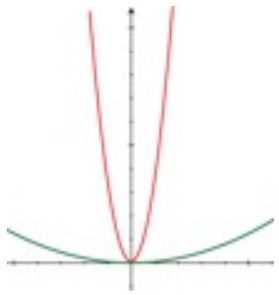
http://www.integrativemodeling.org


Experimental input
and system representation
5C data
Fragments and
interaction maps
IMP Start
IMP sampler
Restraints
MonteCarlo &
Simulating annealing
Possible
solitions
Results analysis
Additional
experimental data
Final model
Selection
Genome
organization
P1
P2

Representation

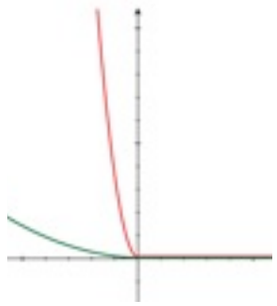
Harmonic

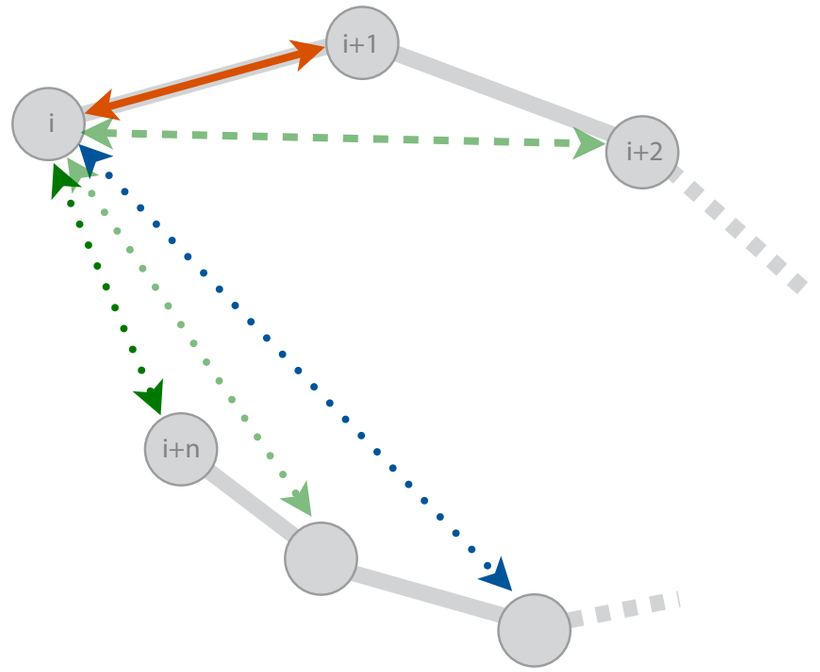
$$H_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2$$



Harmonic Lower Bound

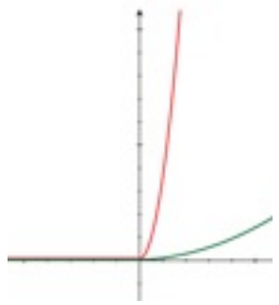
$$\begin{cases} if \;\; d_{i,j} \le d^0_{i,j}; & lbH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if \;\; d_{i,j} > d^0_{i,j}; & lbH_{i,j} = 0 \end{cases}$$





Harmonic Upper Bound

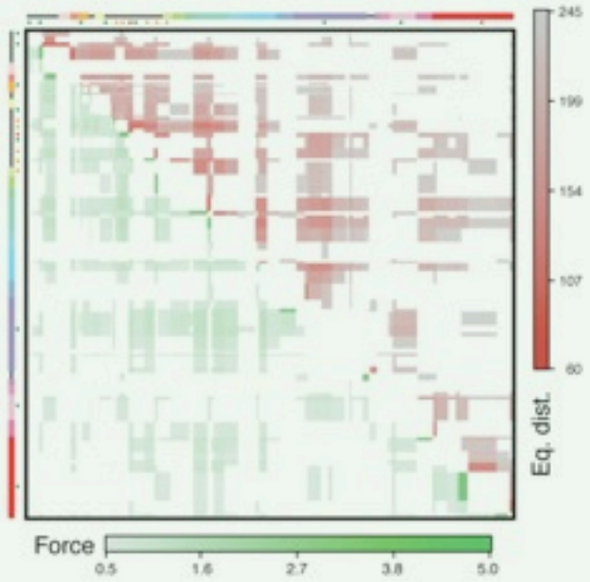
$$\begin{cases} if \;\; d_{i,j} \ge d^0_{i,j}; & ubH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if \;\; d_{i,j} < d^0_{i,j}; & ubH_{i,j} = 0 \end{cases}$$

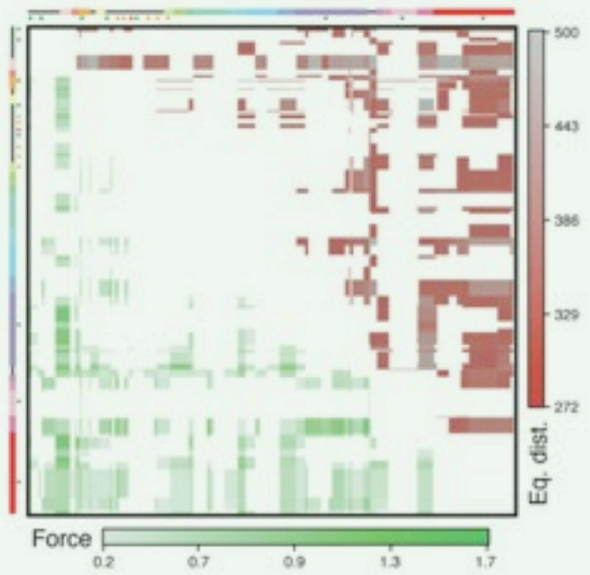


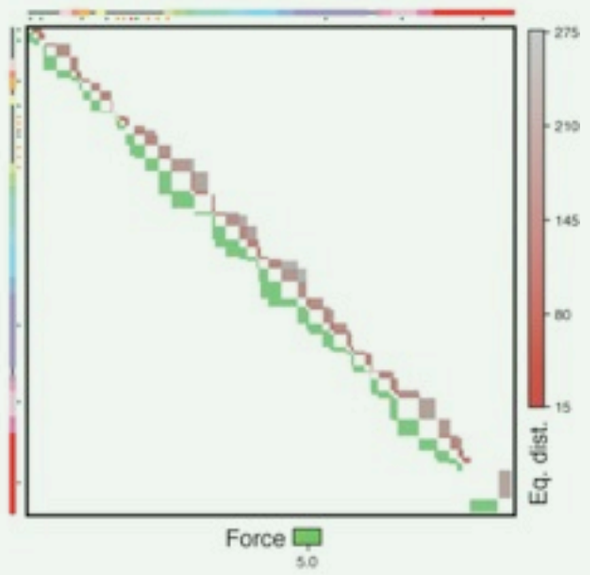
Scoring

GM12878

70 fragments
1,520 restraints



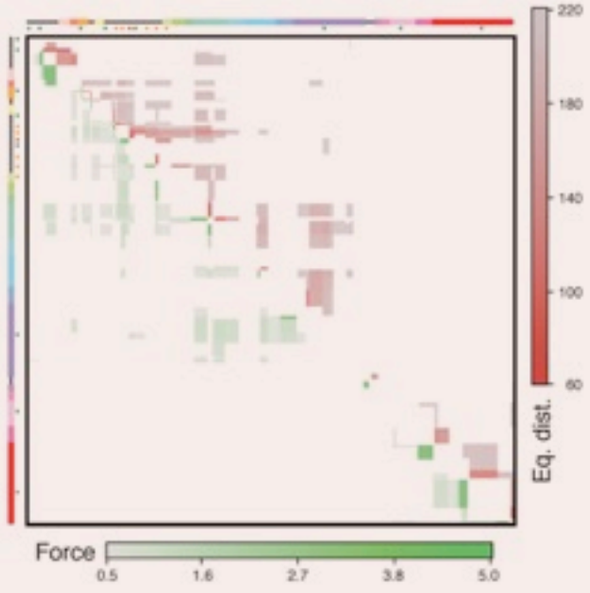


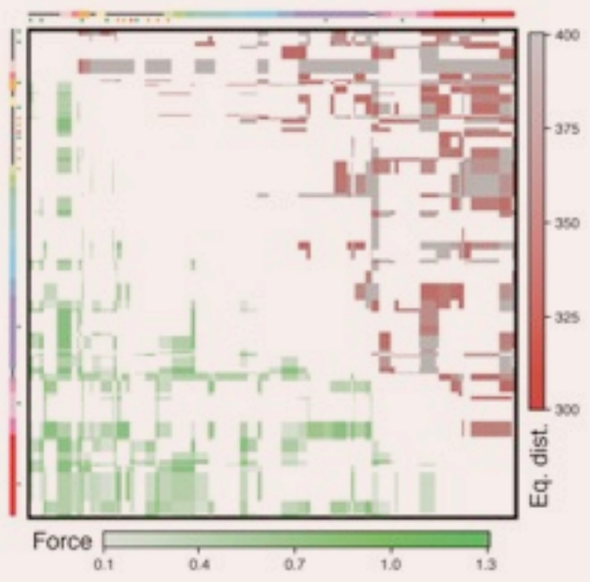


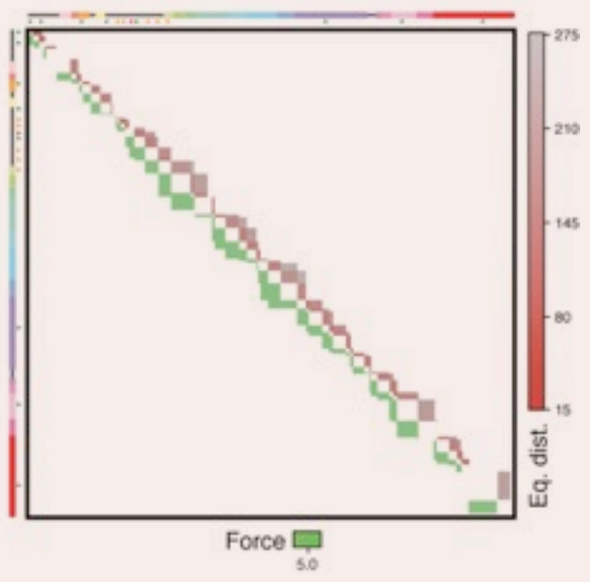
Harmonic | Harmonic Lower Bound | Harmonic Upper Bound

K562

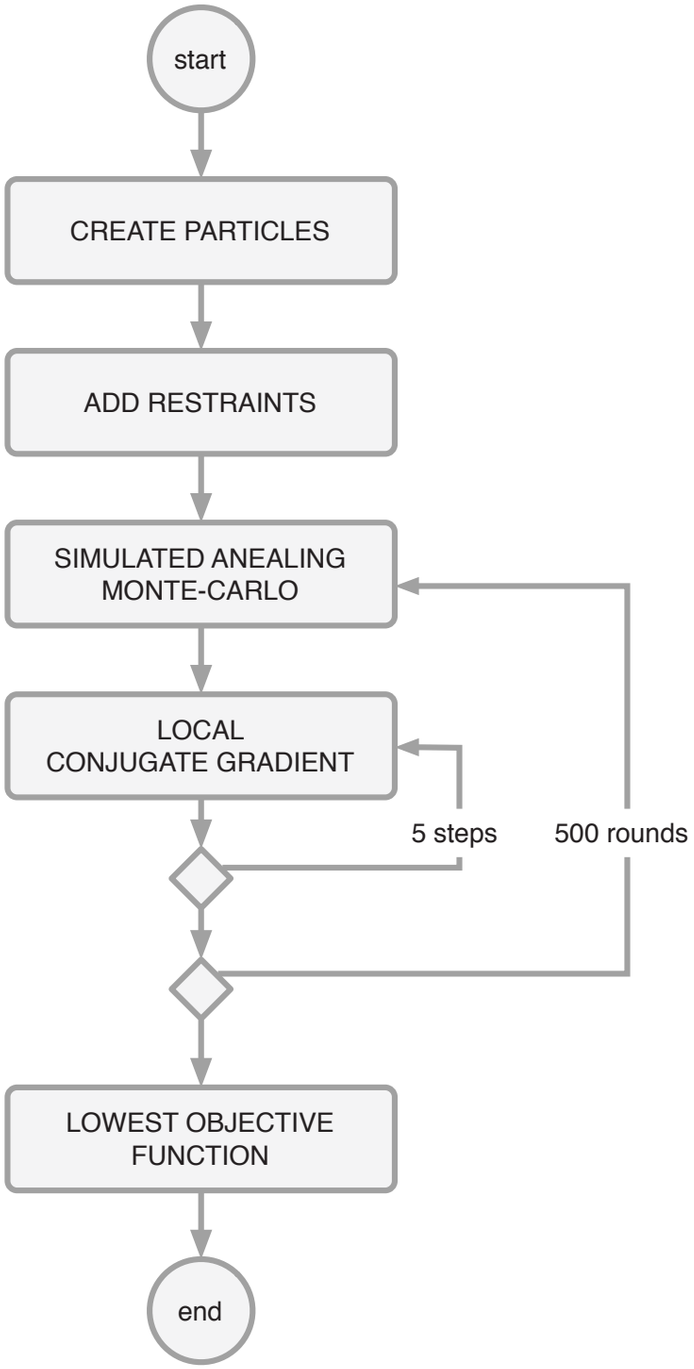
70 fragments
1,049 restraints

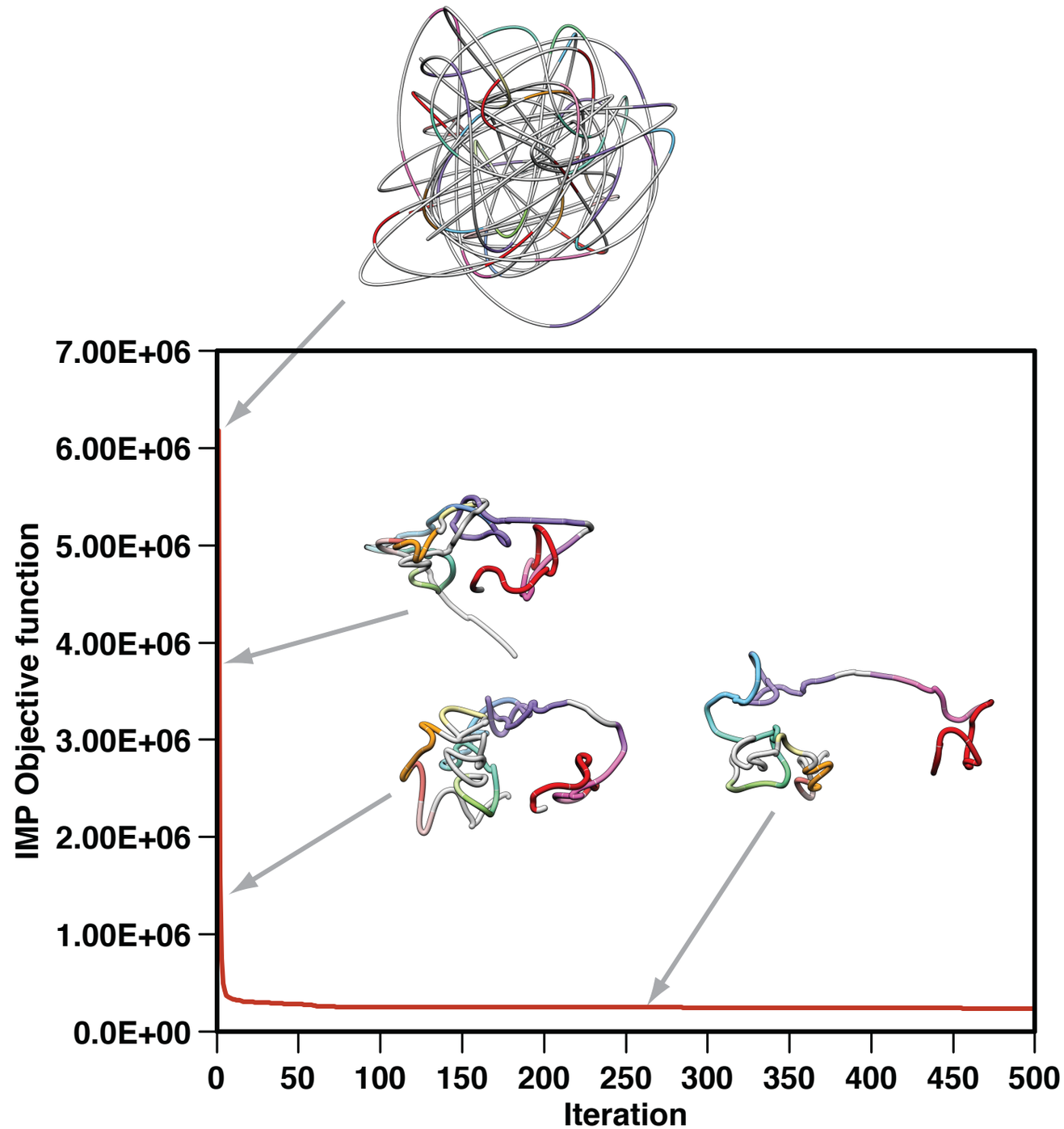






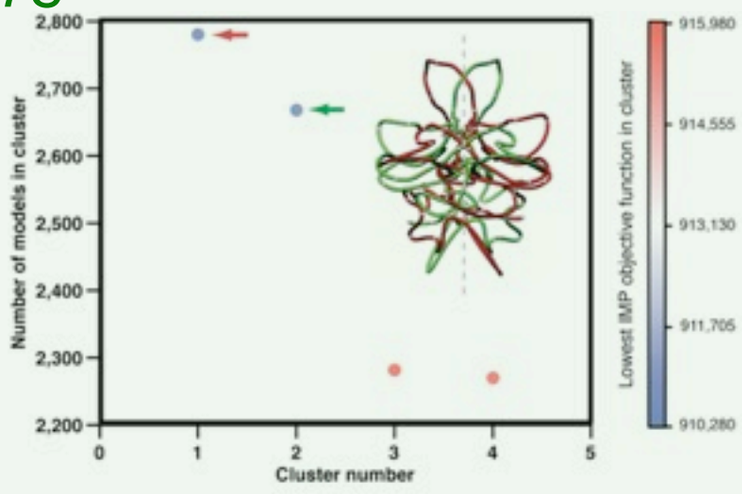
Optimization

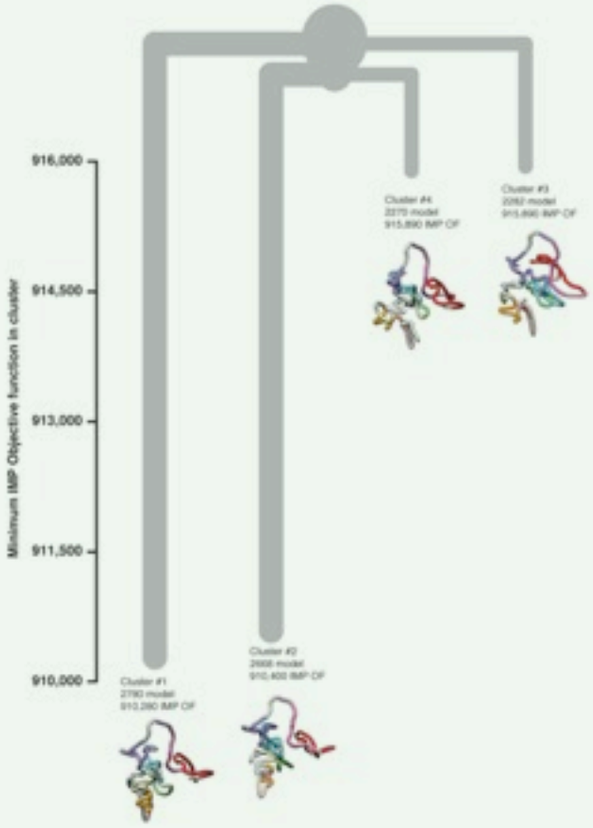




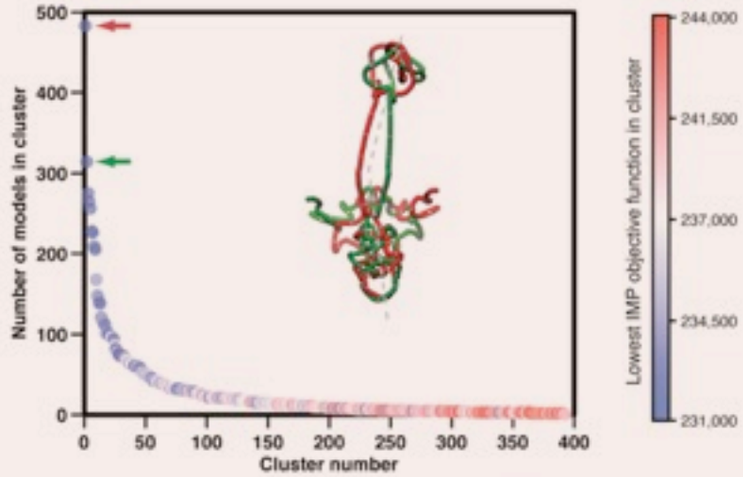
Not just *one* solution

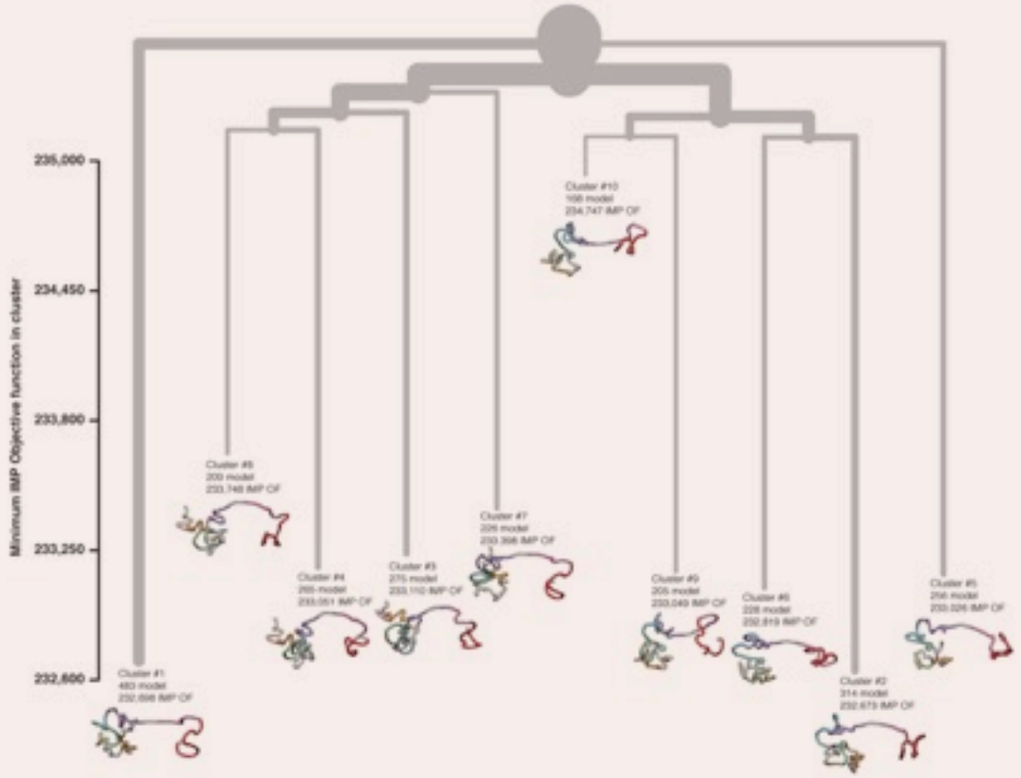
GM12878





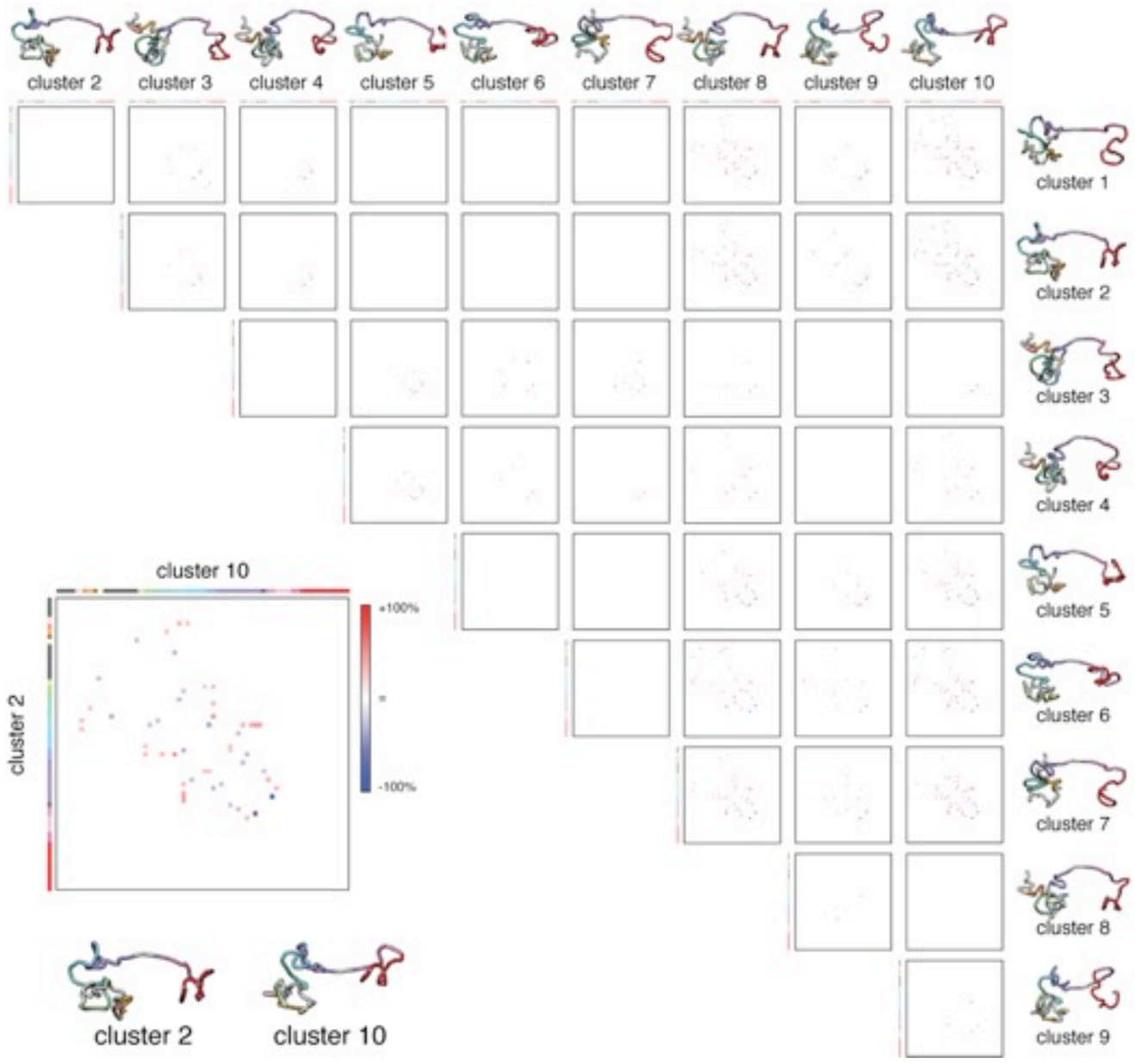
K562





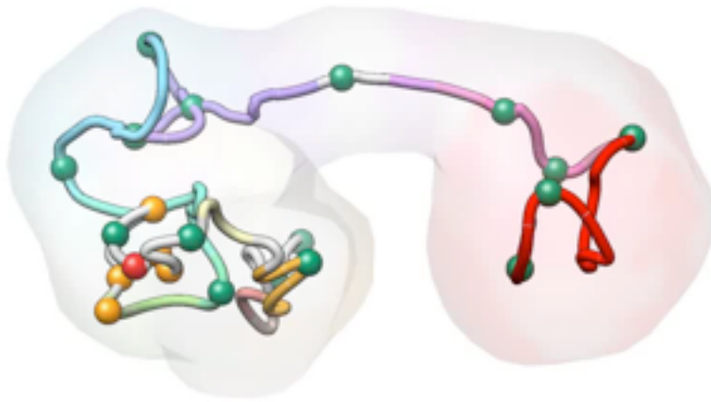
Not just *one* solution

and we can de-convolute them!

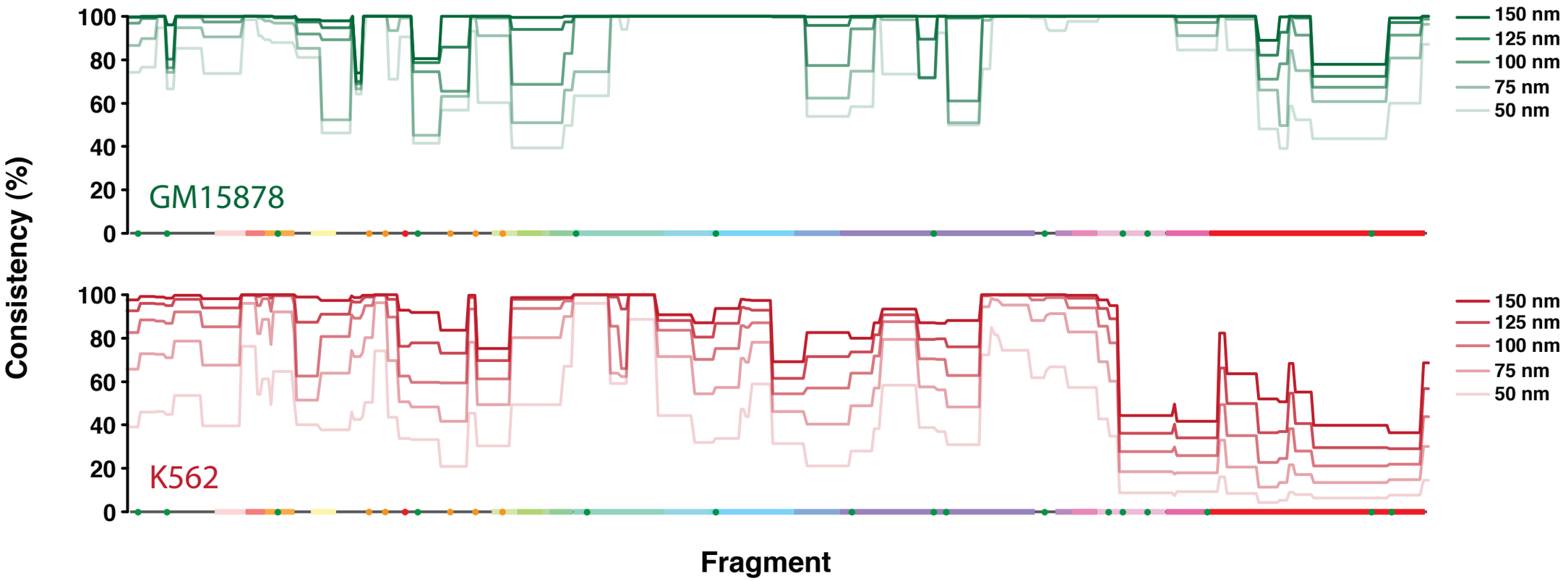


Consistency

GM12878
Cluster #1
2780 model



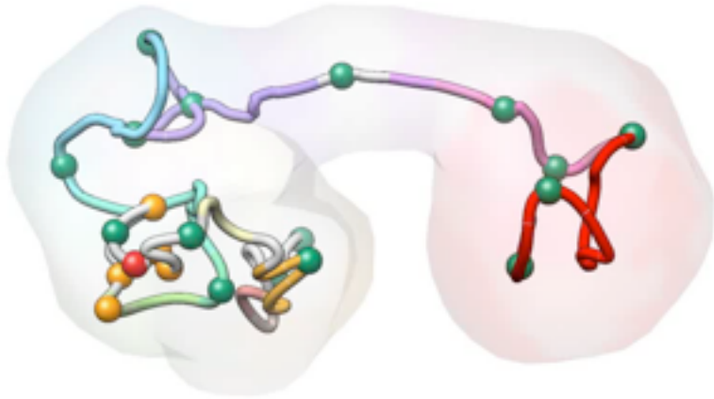
K562
Cluster #2
314 model



Regulatory elements

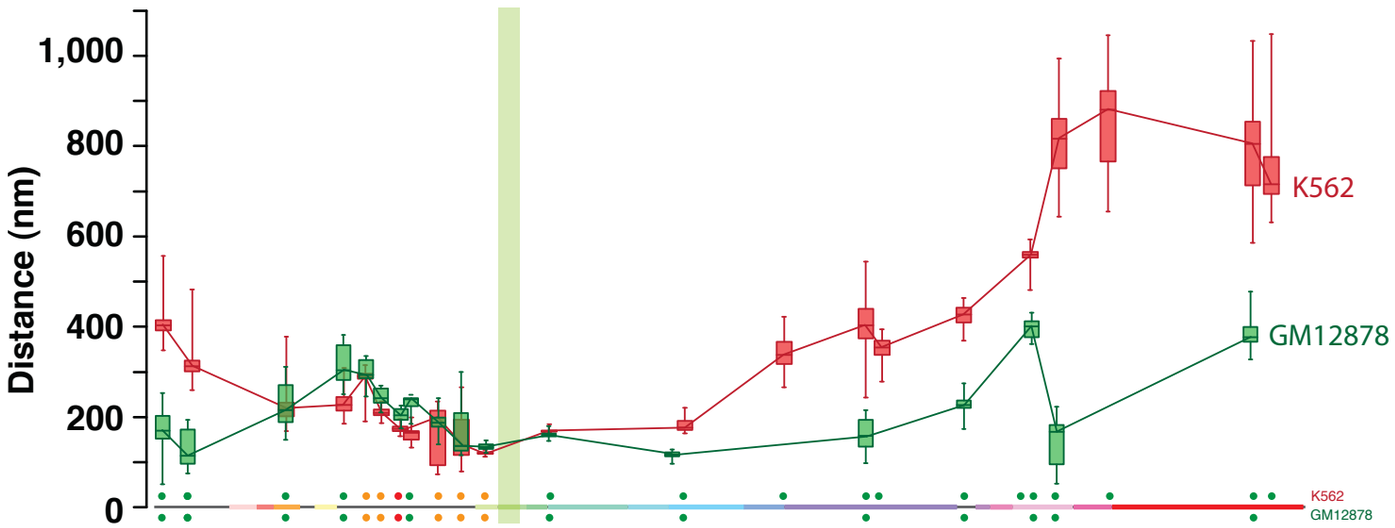
GM12878

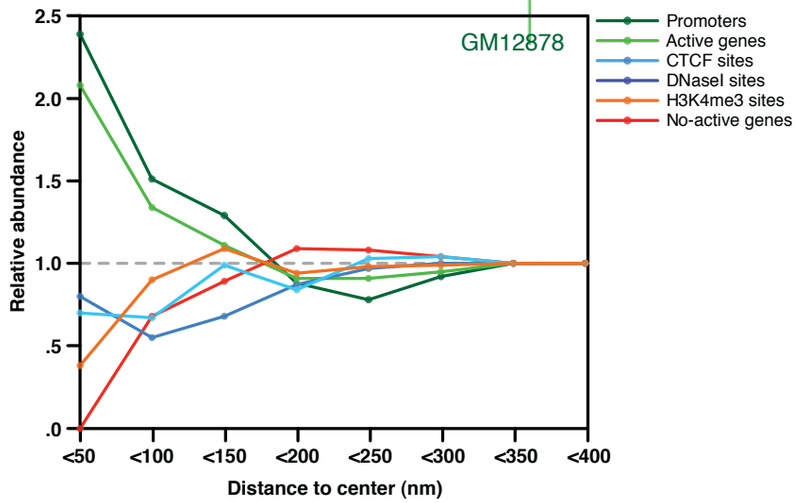
Cluster #1
2780 model

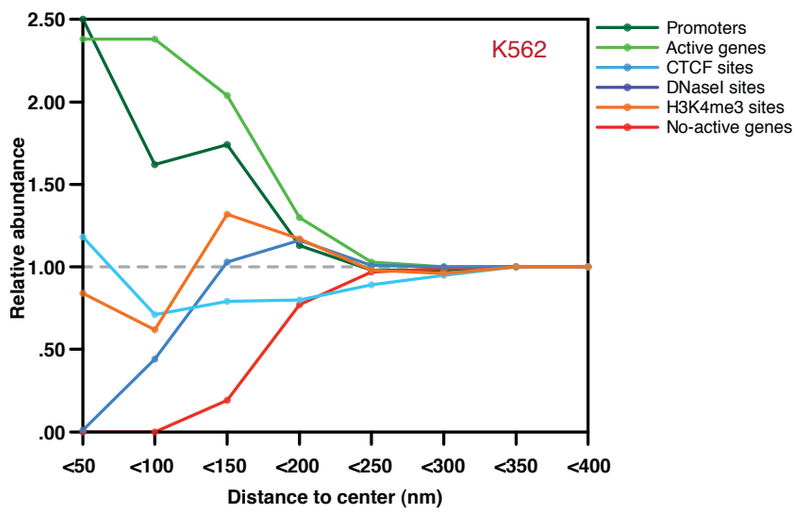


K562

Cluster #2
314 model

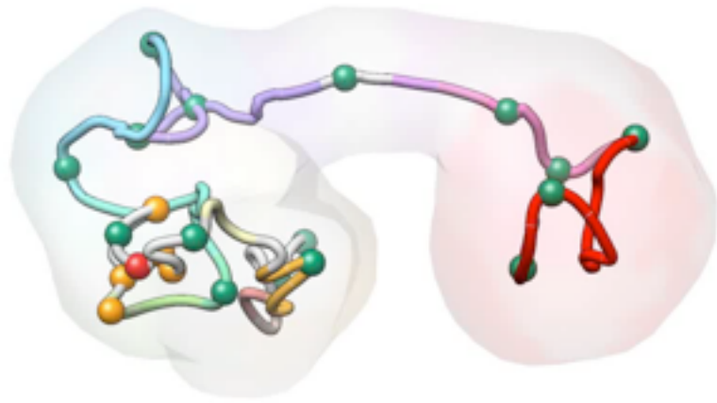




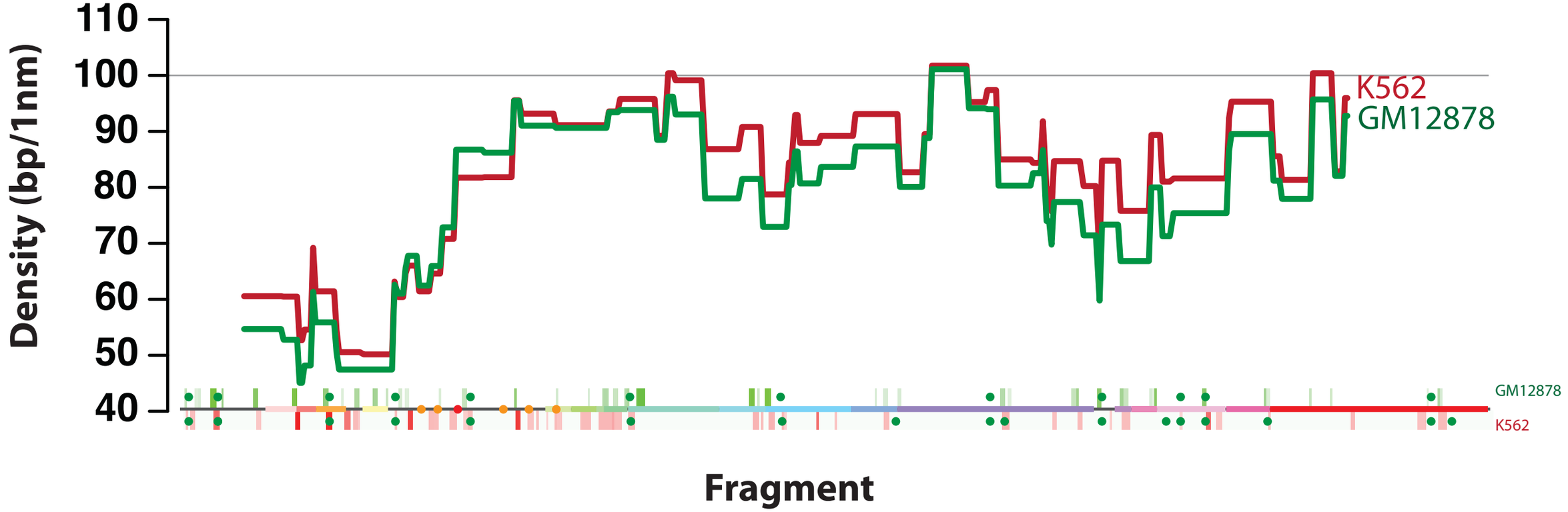


Compactness

GM12878
Cluster #1
2780 model



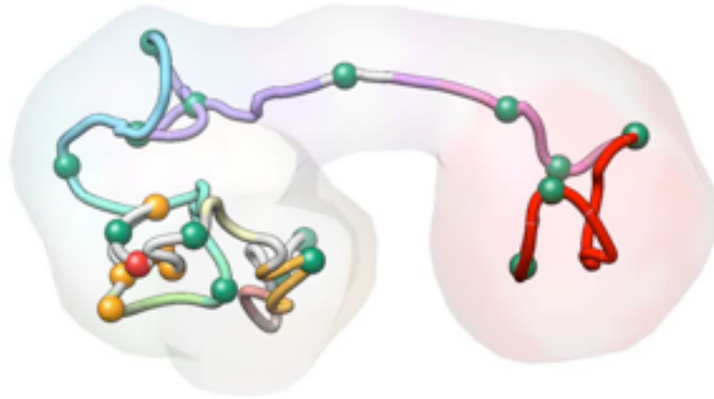
K562
Cluster #2
314 model



Multi-loops

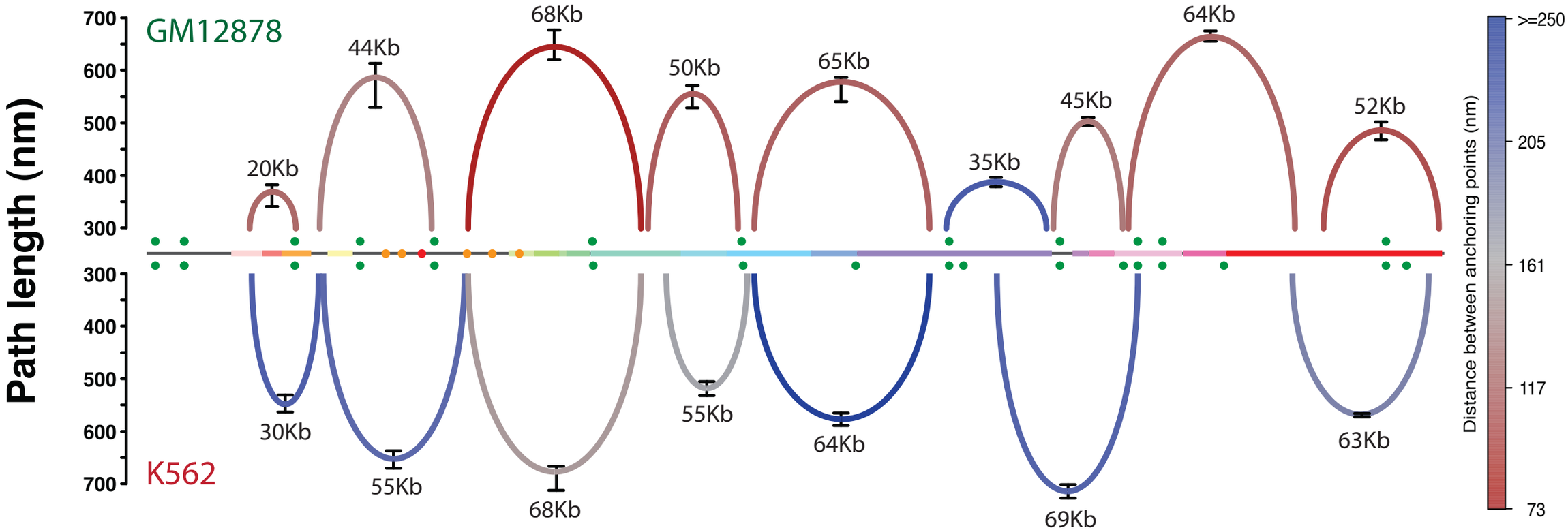
GM12878

Cluster #1
2780 model



K562

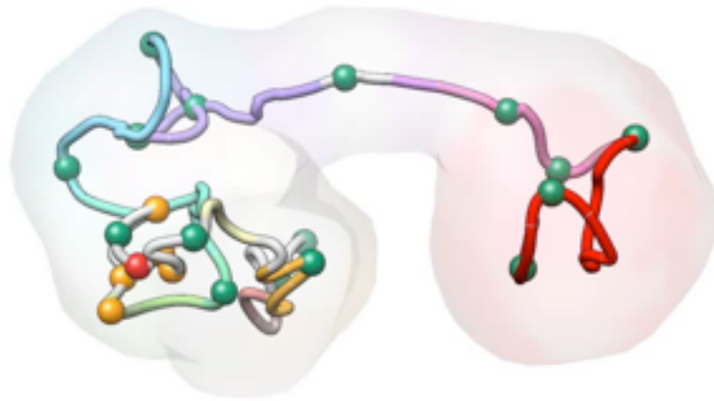
Cluster #2
314 model



Expression

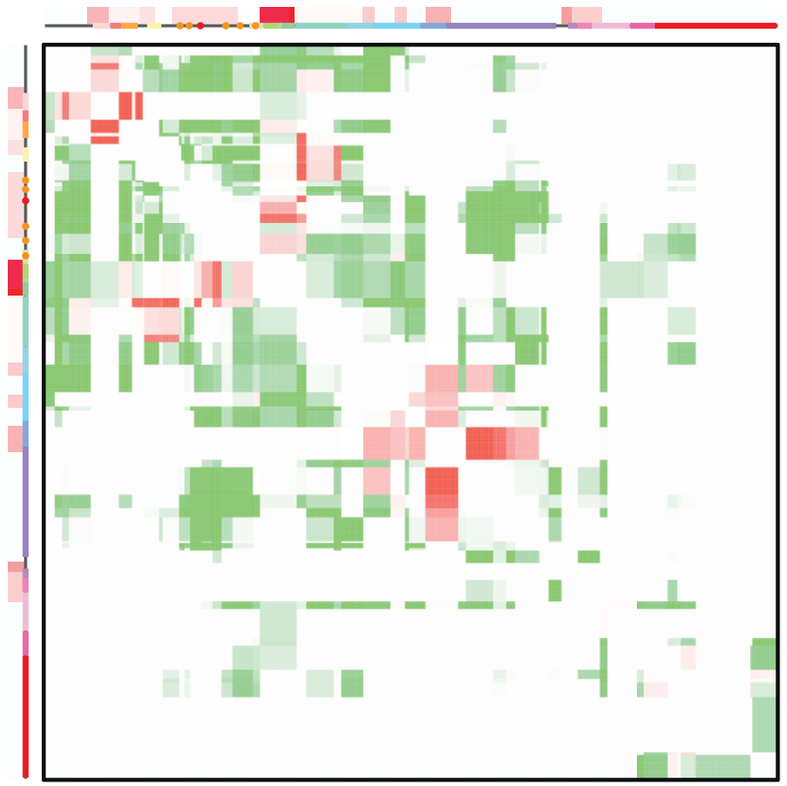
GM12878

Cluster #1
2780 model



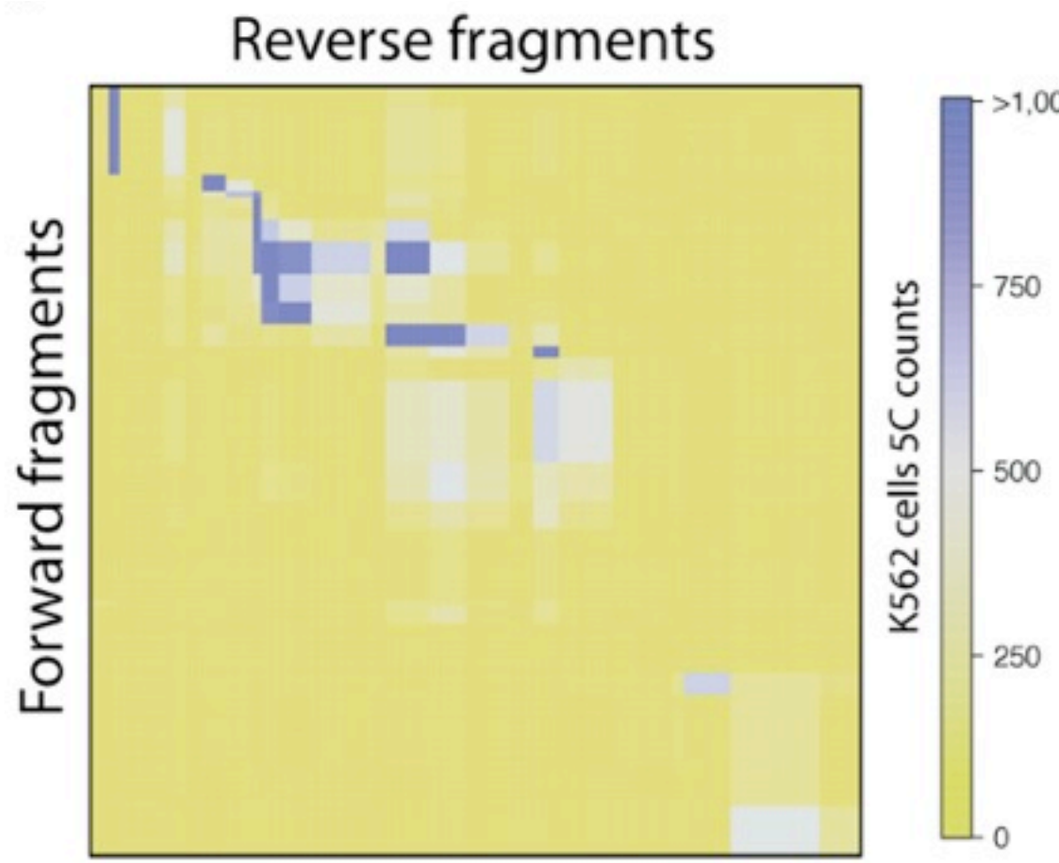
K562

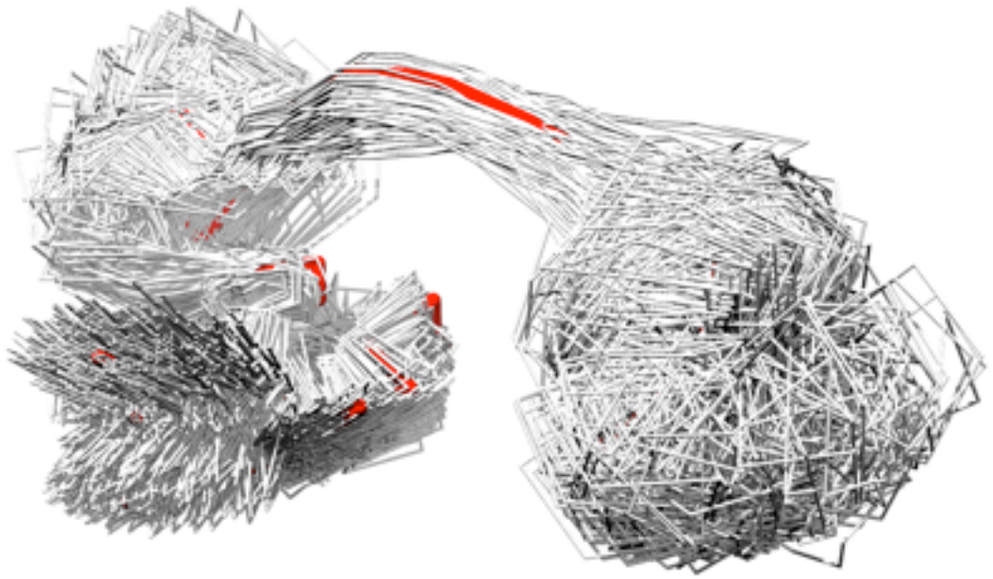
Cluster #2
314 model

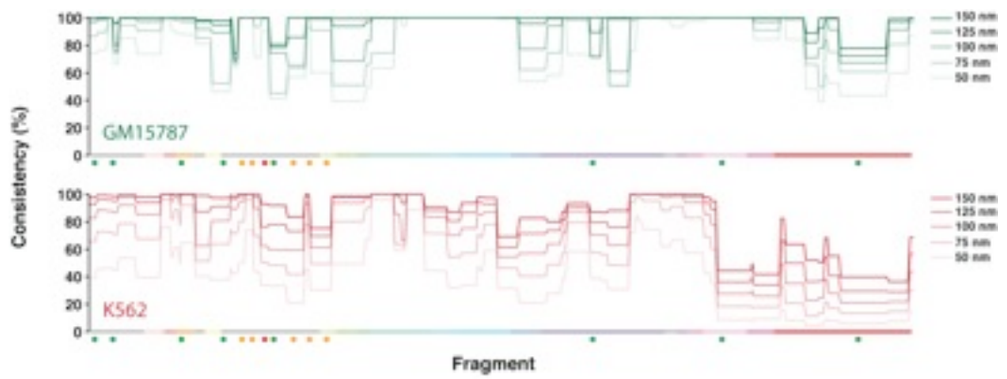


Summary

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







Summary

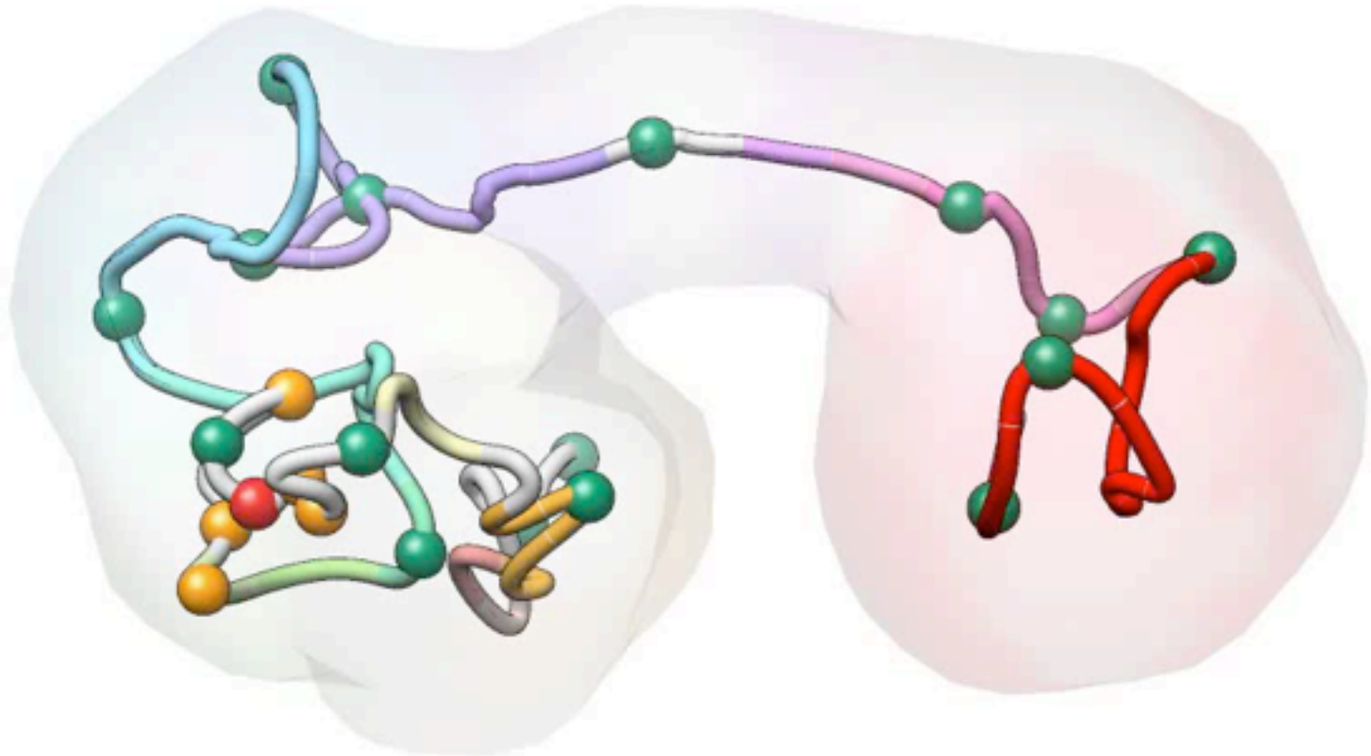
Models allow for 5C data de-convolution

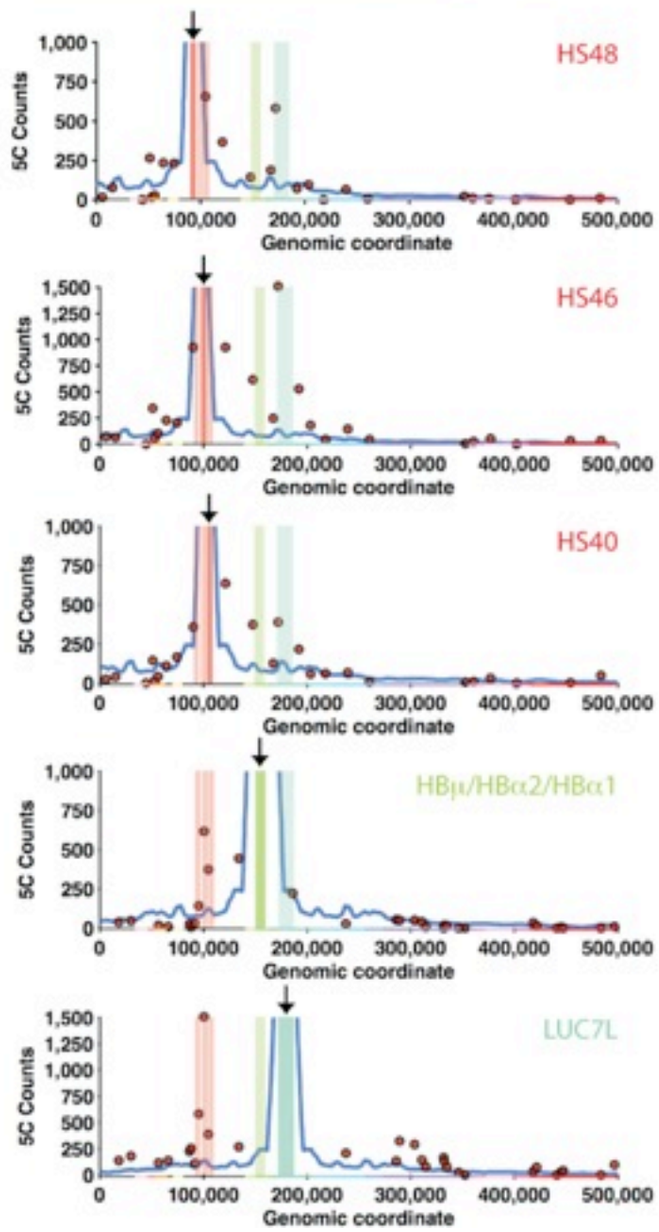
Summary

Models allow for 5C data de-convolution

Summary

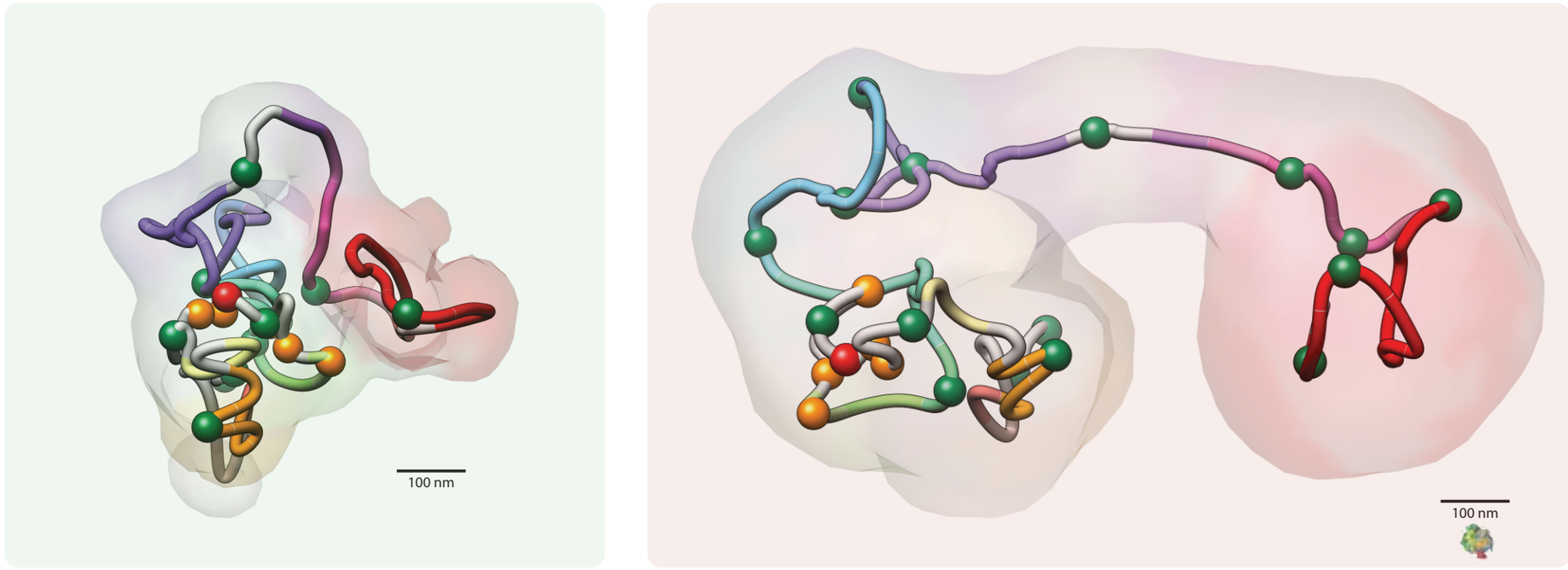
Selected models reproduce known (**and new**) interactions

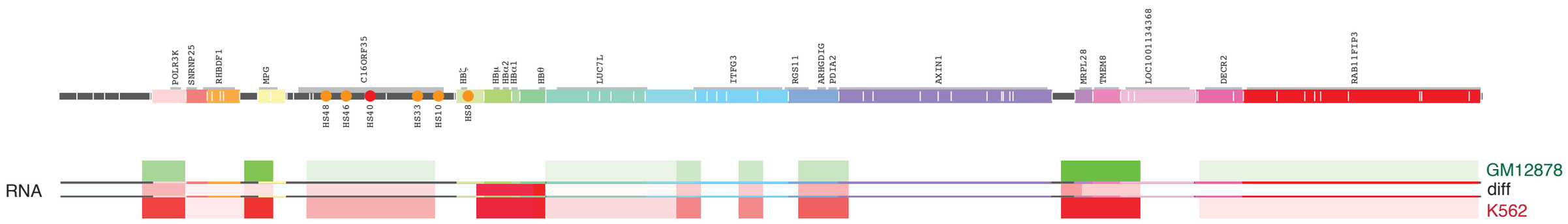




Summary

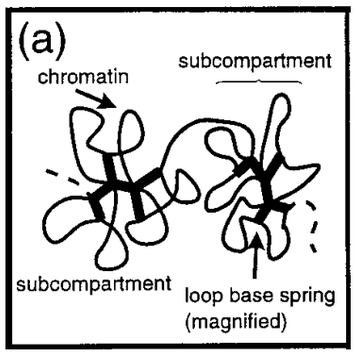
Large-scale changes in conformation correlate with gene expression of resident genes



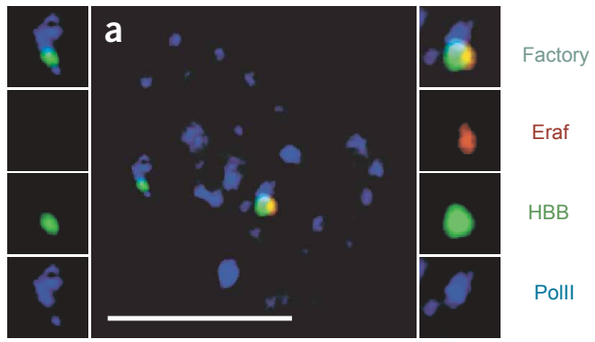


Summary

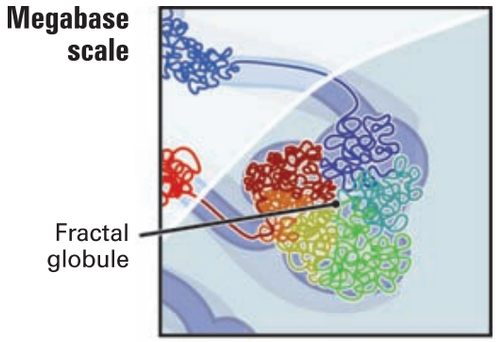
"Chromatin Globule" model



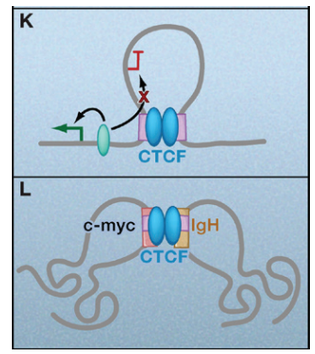
Münkel et al. JMB (1999)



Osborne et al. Nat Genet (2004)

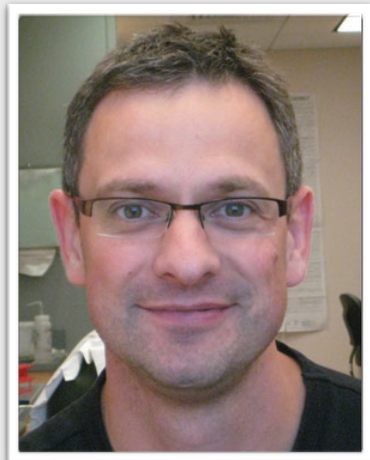


Lieberman-Aiden et al. Science (2009)



Phillips and Corces. Cell (2009)

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