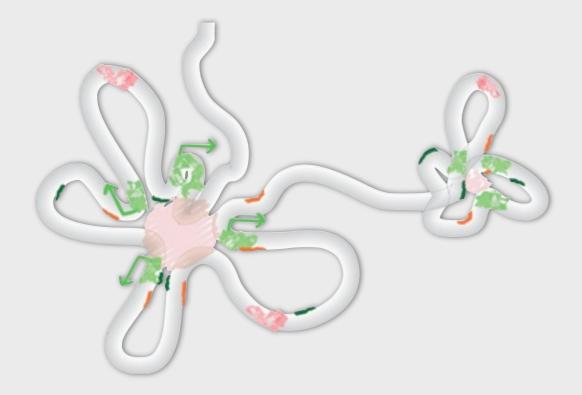
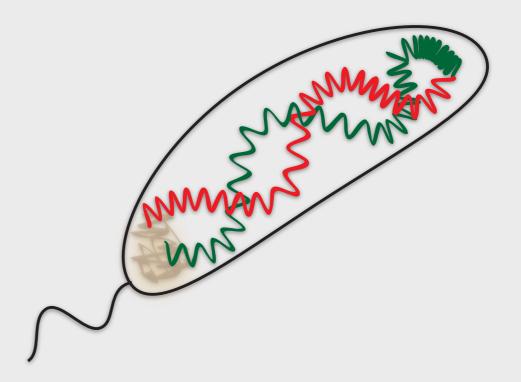
# Structure determination of genomes and genomic domains by satisfaction of spatial restraints





#### Marc A. Marti-Renom

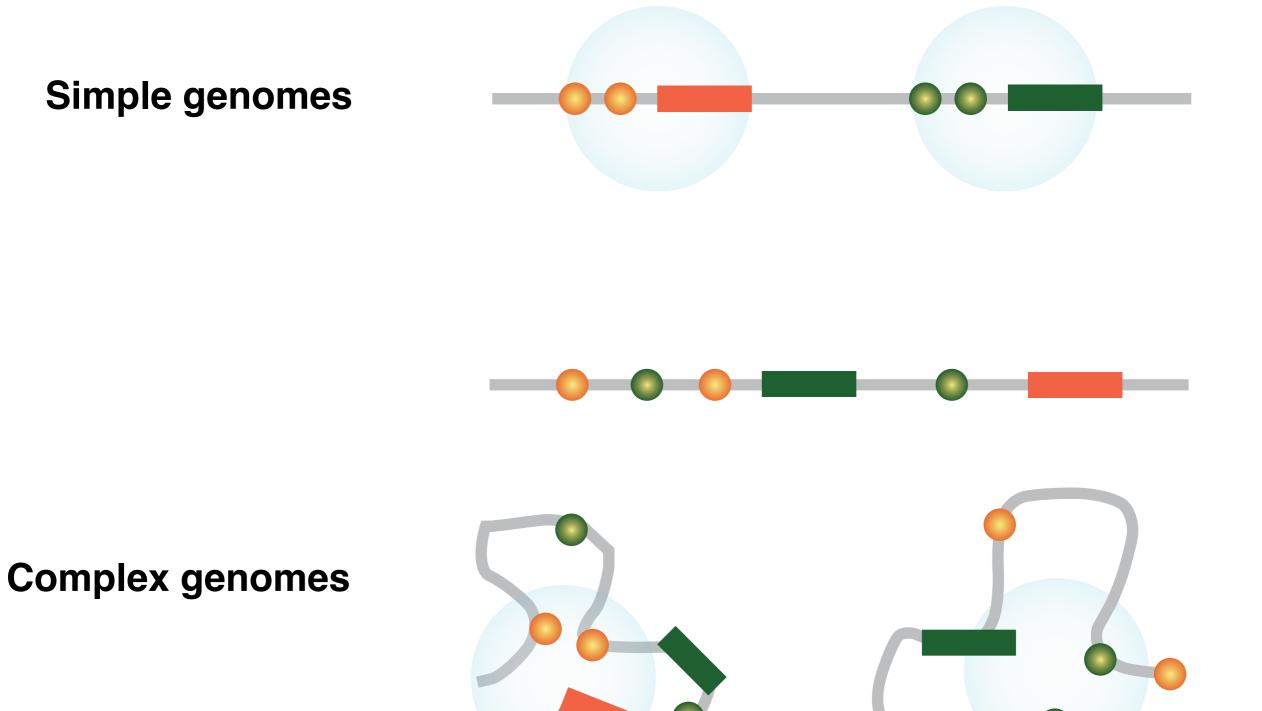


Structural Genomics Laboratory Bioinformatics & Genomics Department Prince Felipe Research Center (CIPF), Valencia, Spain



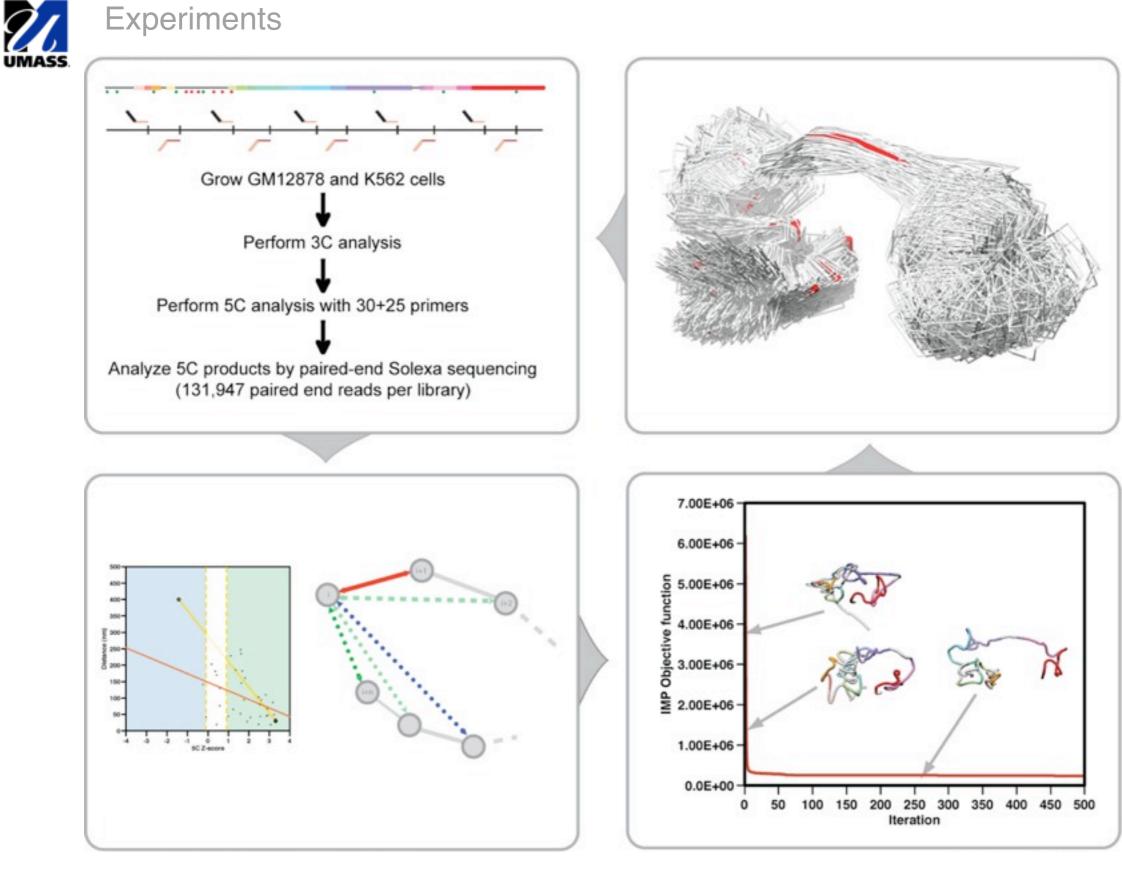
http://sgu.bioinfo.cipf.es

Friday, November 4, 11

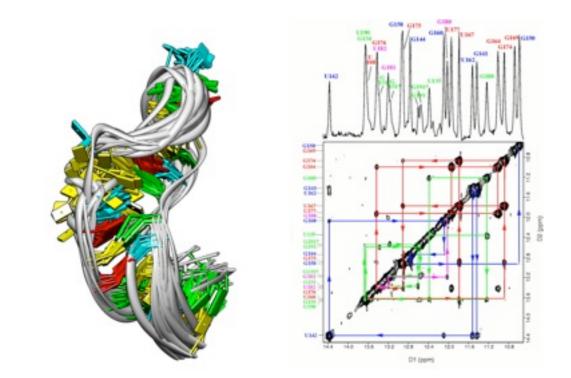


Know	vledge								
A A A A A					IDM			$\begin{array}{c} 6 \\ 11 \\ 5 \\ 1 \\ 20 \\ 3 \\ 18 \\ 7 \\ 2 \\ 18 \\ 7 \\ 2 \\ 16 \\ 9 \\ 7 \\ 16 \\ 9 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 7 \\ 16 \\ 9 \\ 18 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10$	7
10 <sup>0</sup>		10 <sup>3</sup>			10 <sup>6</sup>			DNA length 10 <sup>9</sup>	] nt
								Volume	]
10 <sup>-9</sup>		10 <sup>-6</sup>		10 <sup>-3</sup>		10 <sup>0</sup>		10 <sup>3</sup>	μm³
10 <sup>-10</sup>	10 <sup>-8</sup>	10 <sup>-6</sup>	10 <sup>-4</sup>	10 <sup>-2</sup>		10 <sup>0</sup>	10 <sup>2</sup>	Time 10 <sup>3</sup>	s
10 <sup>-3</sup>			10 <sup>-2</sup>				10 <sup>-1</sup>	Resolution	μ

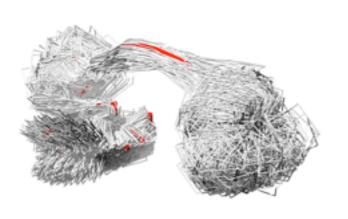
Adapted from: Langowski and Heermann. Semin Cell Dev Biol (2007) vol. 18 (5) pp. 659-67

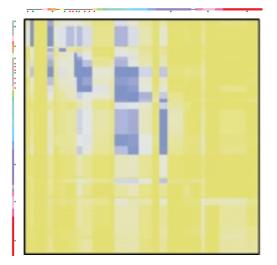


Computation PRINCIPE FELIPE



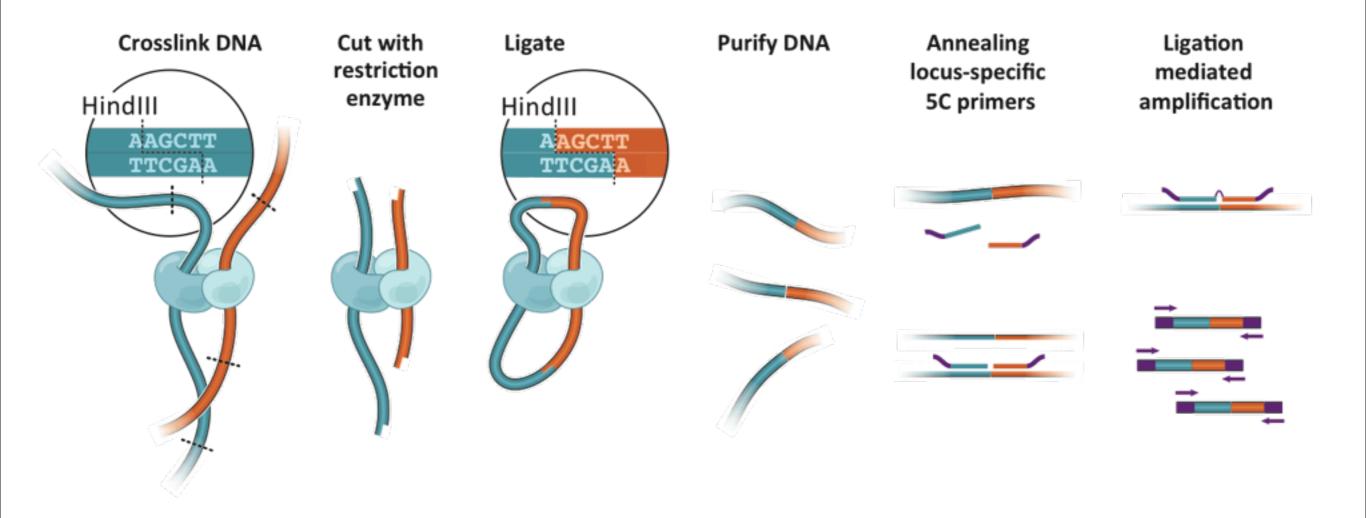
### Biomolecular structure determination 2D-NOESY data





## Chromosome structure determination 5C data





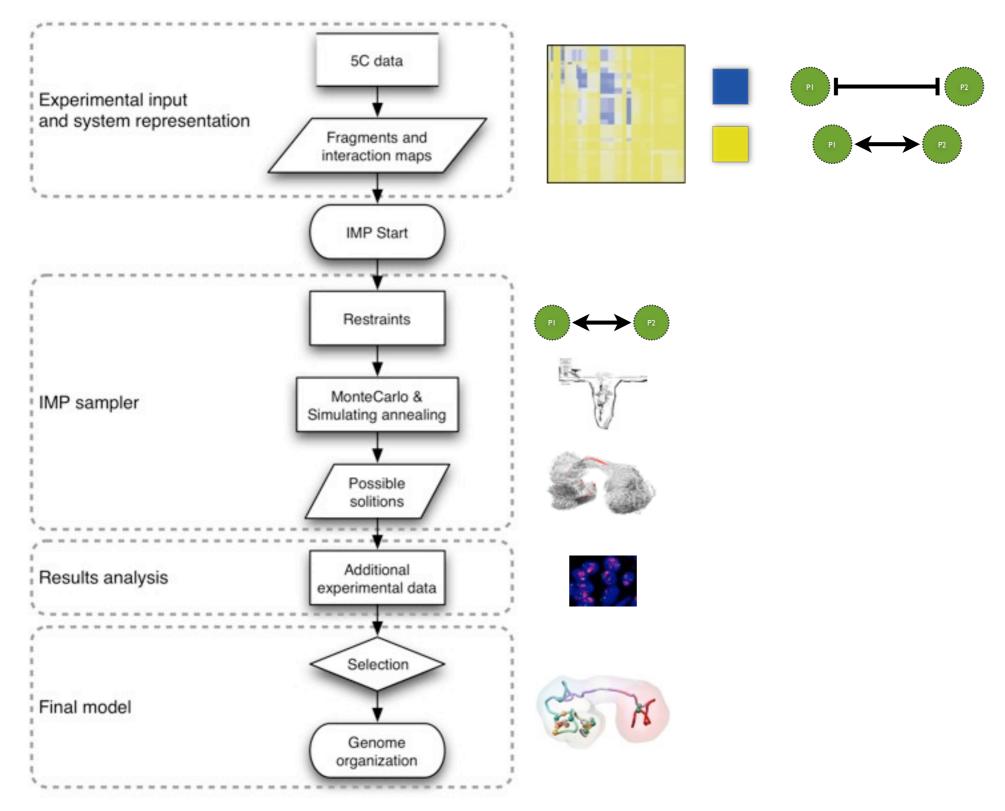
5C "copies" the 3C library into a 5C library containing only ligation junctions

Performed at high levels of multiplexing:

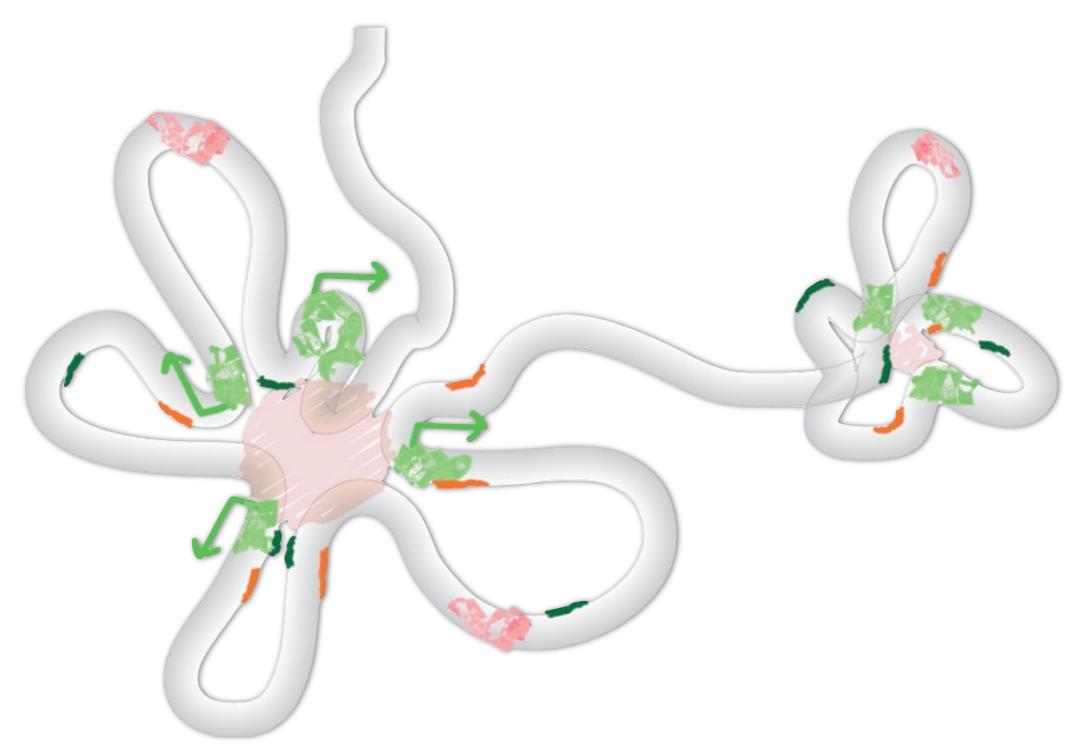
2,000 primers detect 1,000,000 unique interactions in 1 reaction

# **Integrative Modeling**

http://www.integrativemodeling.org



# Human $\alpha$ -globin domain



# Human $\alpha$ -globin domain

### ENm008 genomic structure and environment

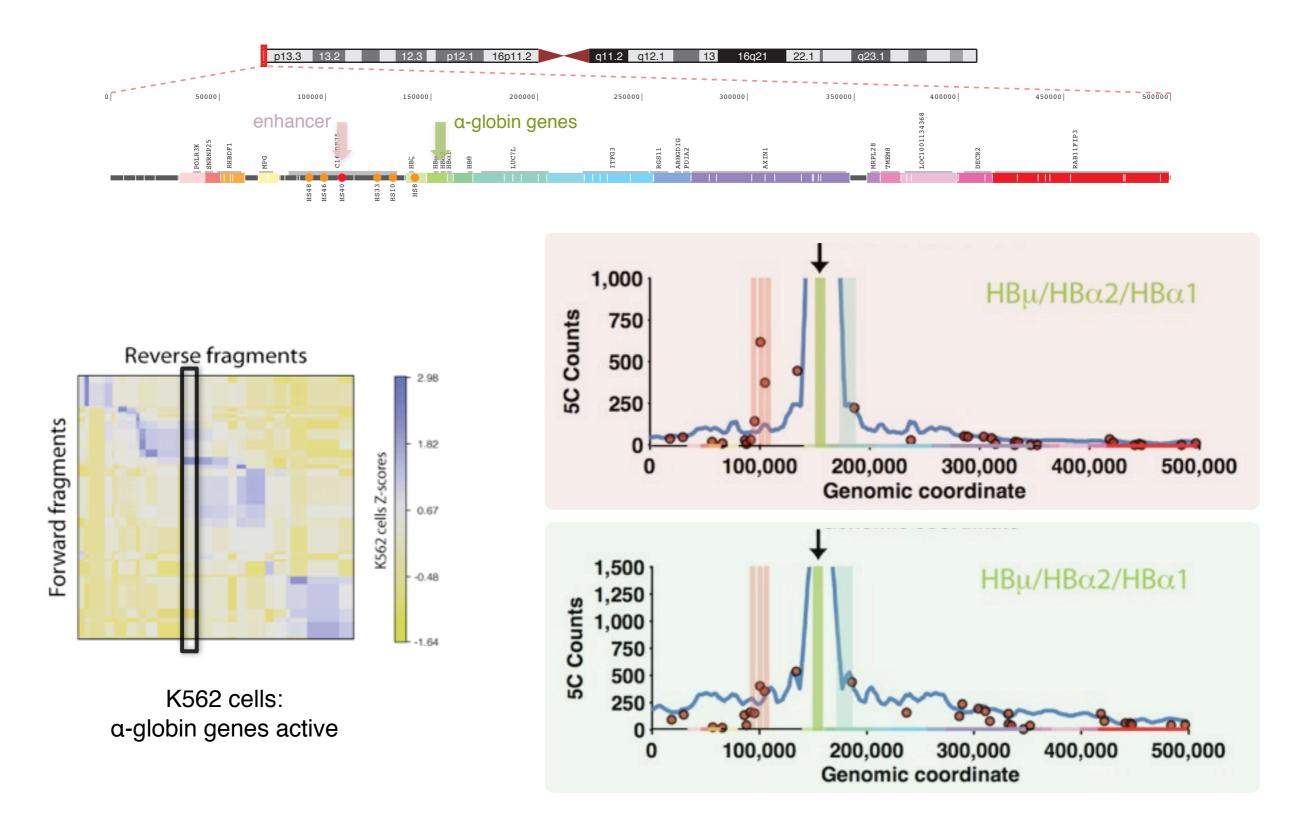
ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816



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 DNasel Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute ChIP-seq (Bernstein Group at Broad Institute of Harvard and MIT).

# Human $\alpha$ -globin domain

ENm008 genomic structure and environment



# Representation

### Harmonic

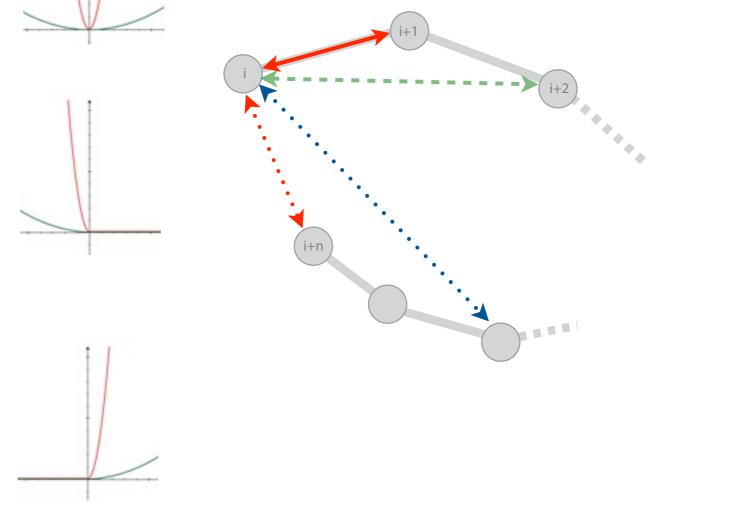
$$H_{i,j} = k \left( d_{i,j} - d_{i,j}^0 \right)^2$$

Harmonic Lower Bound

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

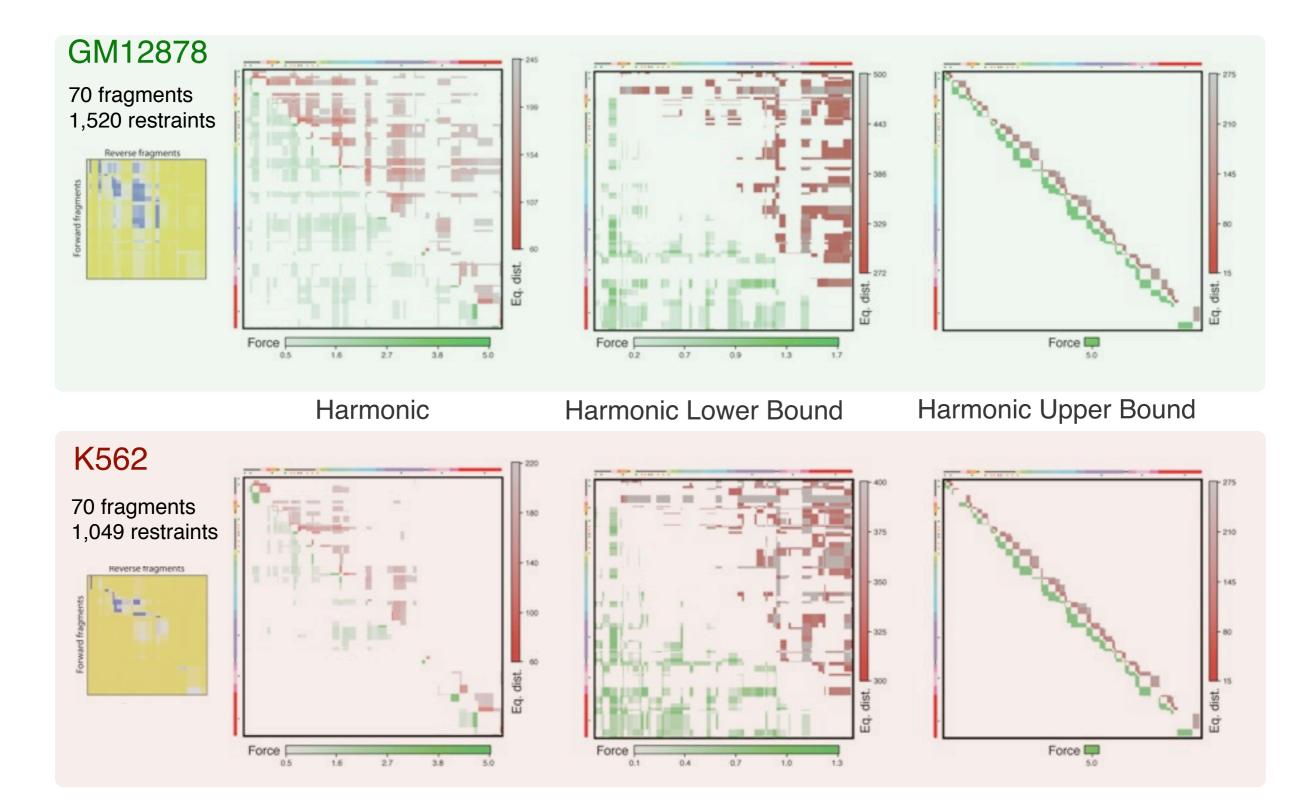
### Harmonic Upper Bound

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

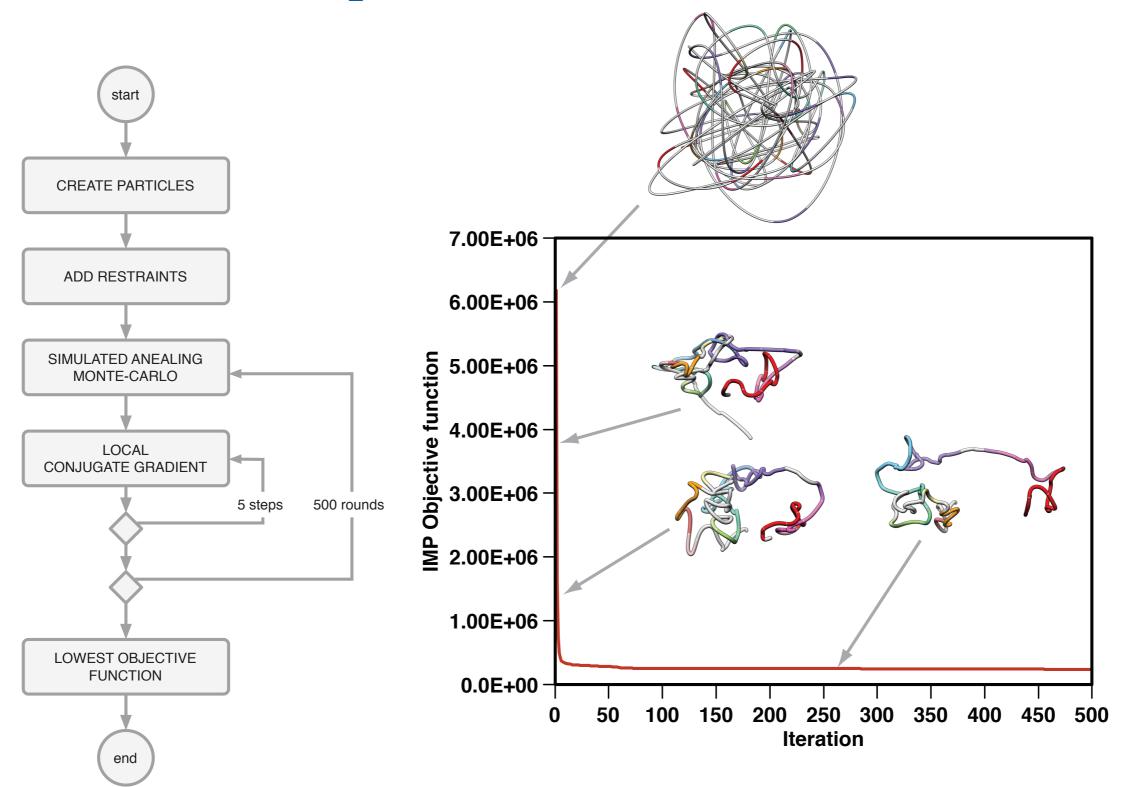


Friday, November 4, 11

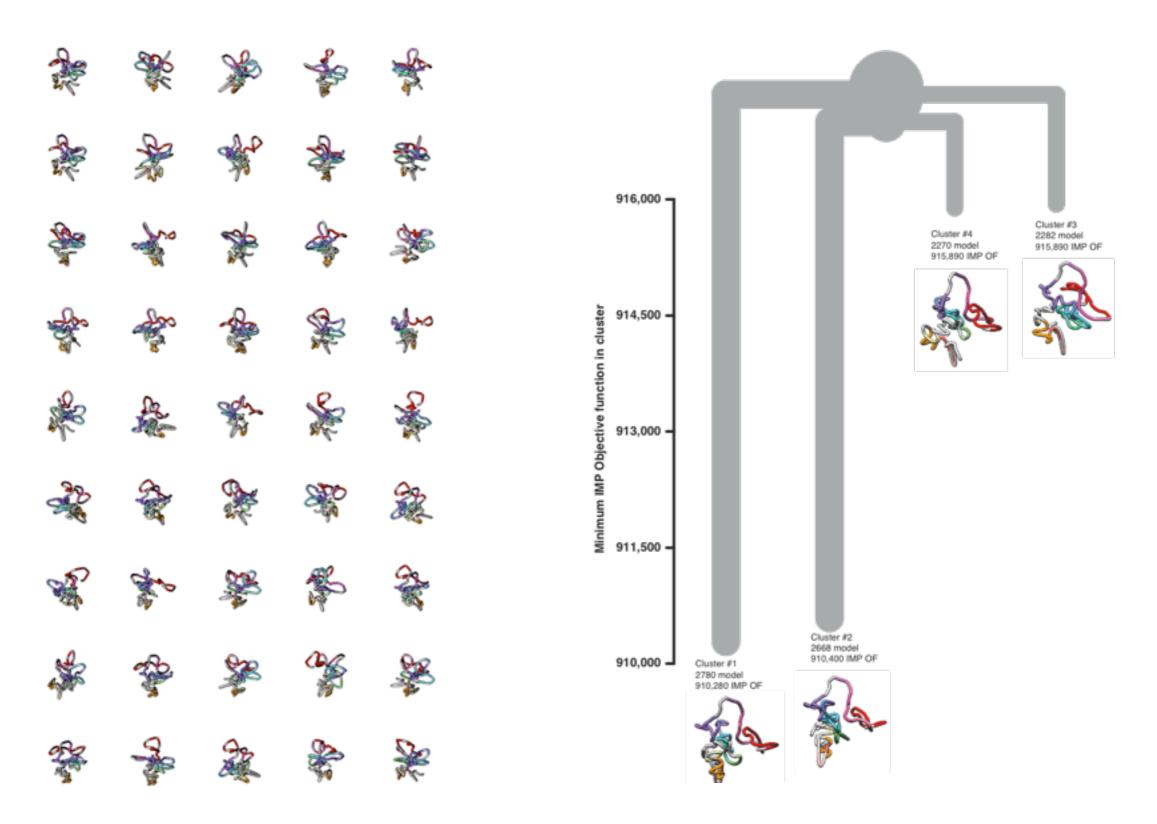
# Scoring



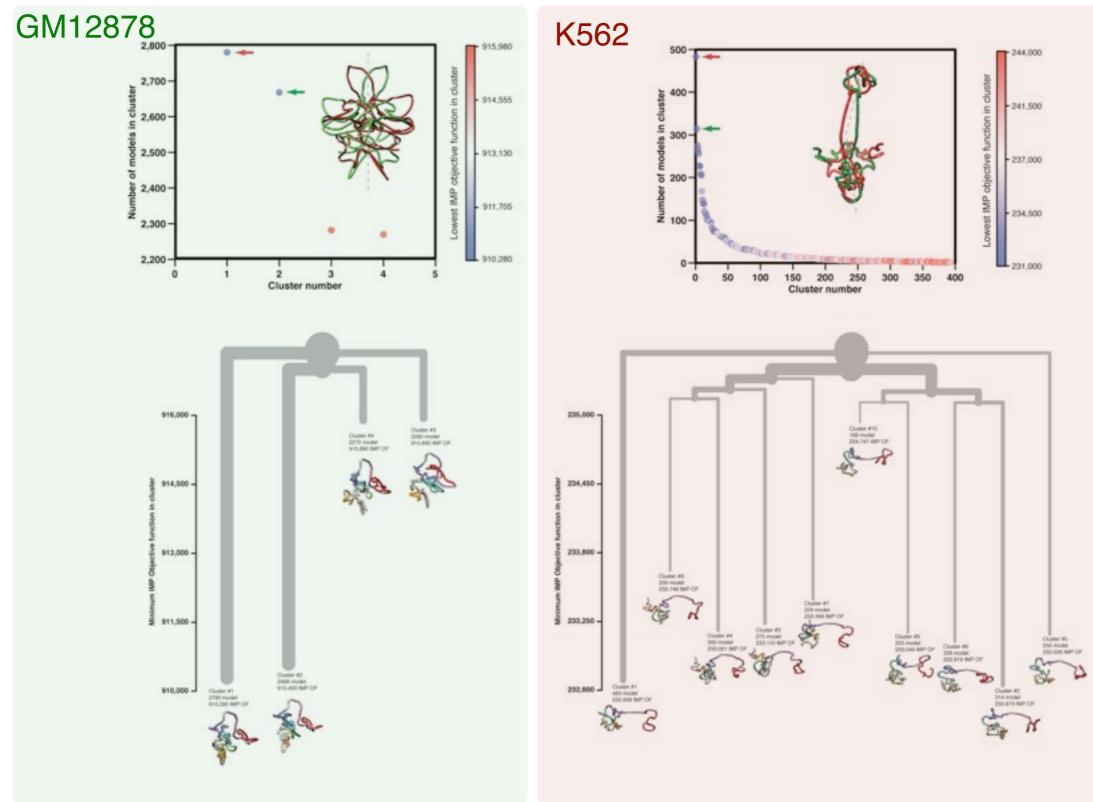
# Optimization



# Clustering

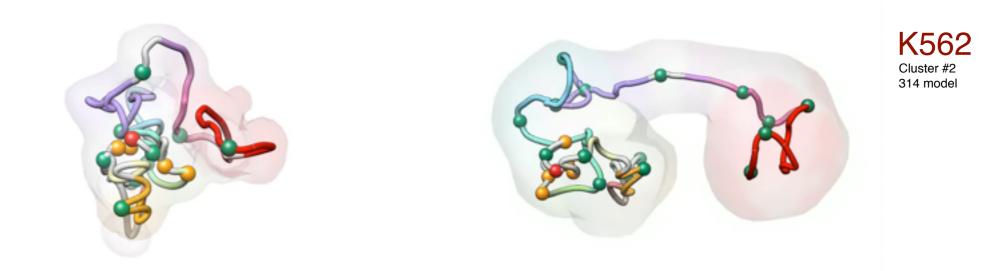


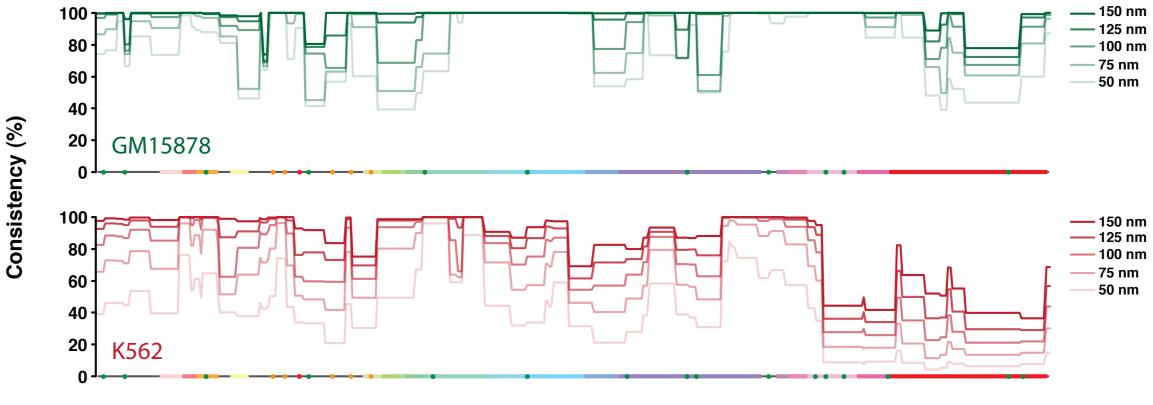
# Not just one solution



# Consistency

GM12878 Cluster #1 2780 model

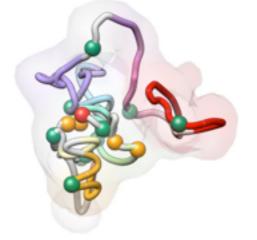


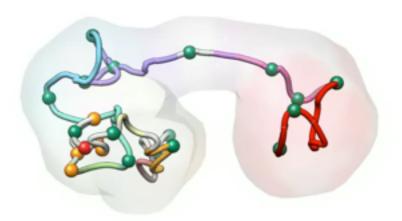


Fragment

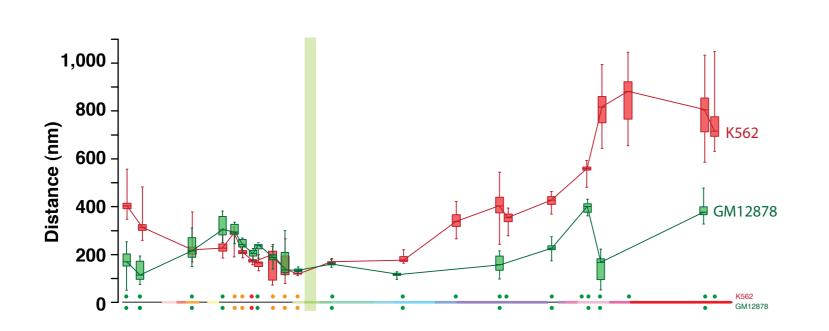
# **Regulatory elements**

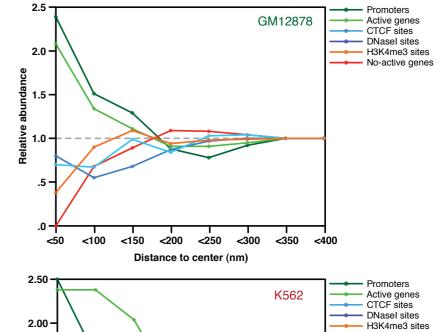
GM12878 Cluster #1 2780 model

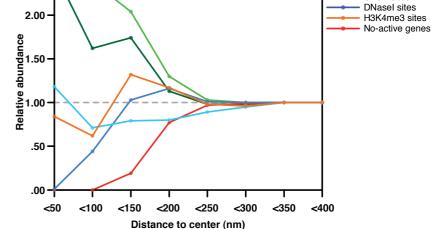




K562 Cluster #2 314 model



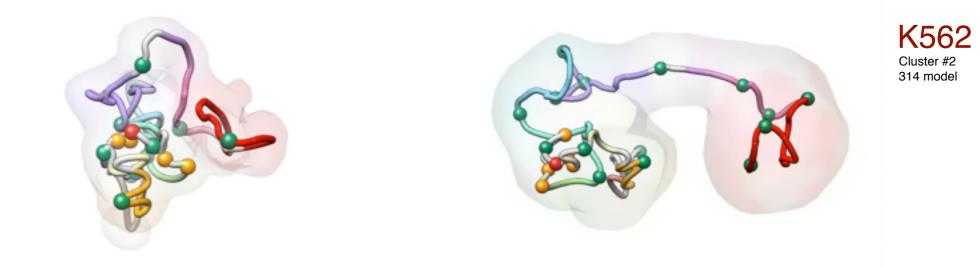


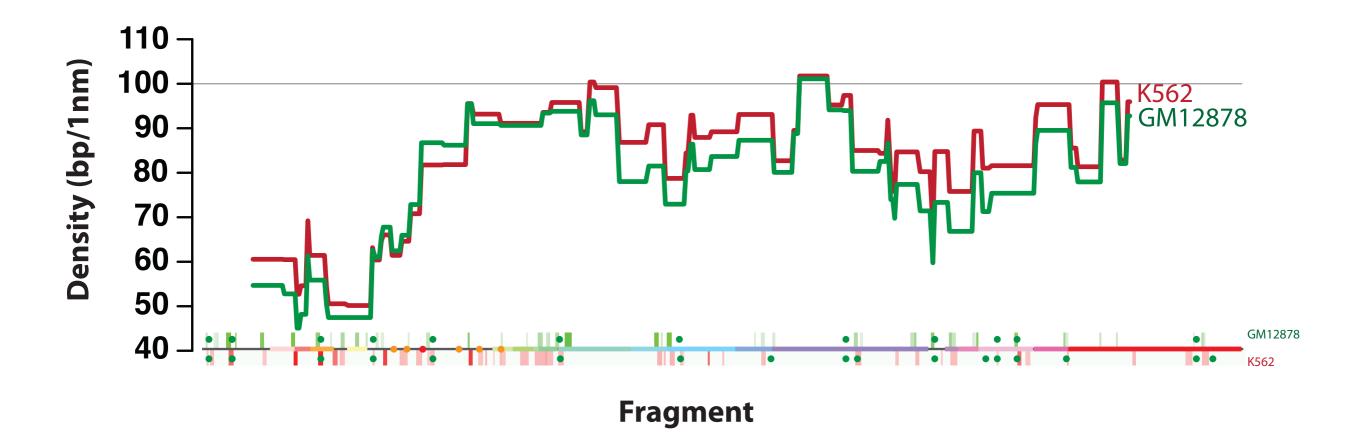


Friday, November 4, 11

# Compactness

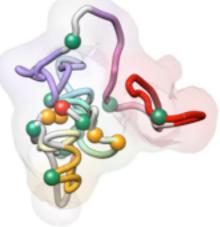
GM12878 Cluster #1 2780 model

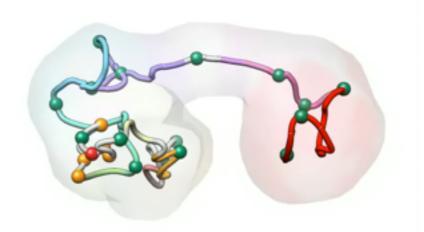




# Multi-loops

GM12878 Cluster #1 2780 model

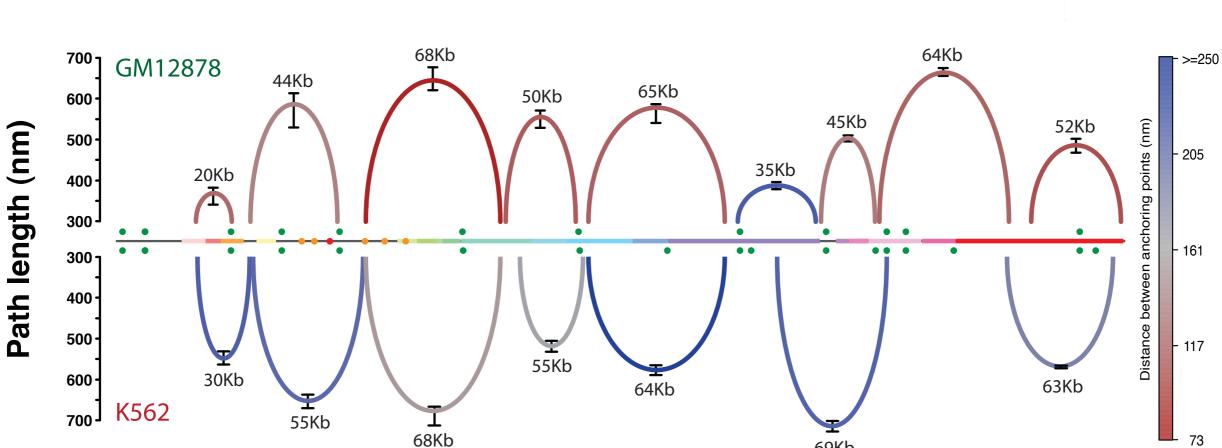




69Kb

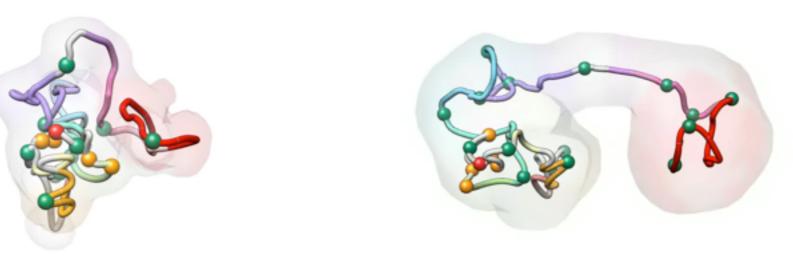
K562 Cluster #2

314 model



# **Expression**

GM12878 Cluster #1 2780 model

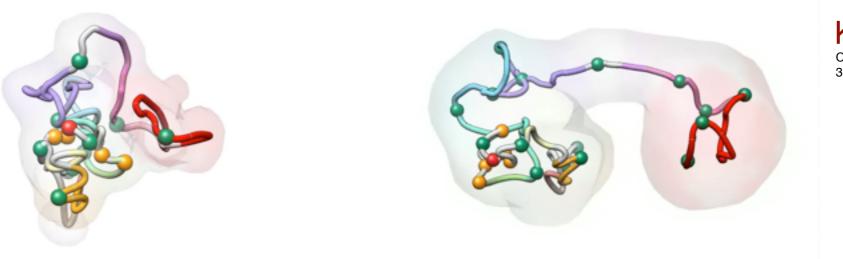






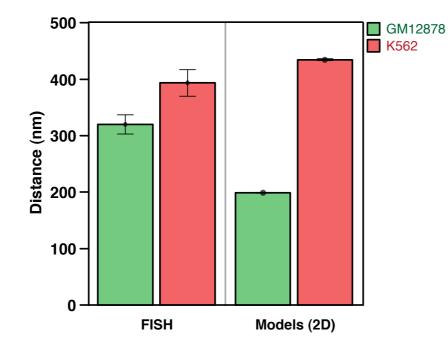
# **FISH validation**

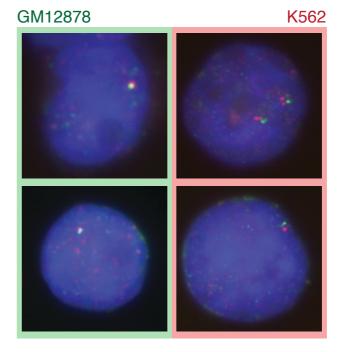






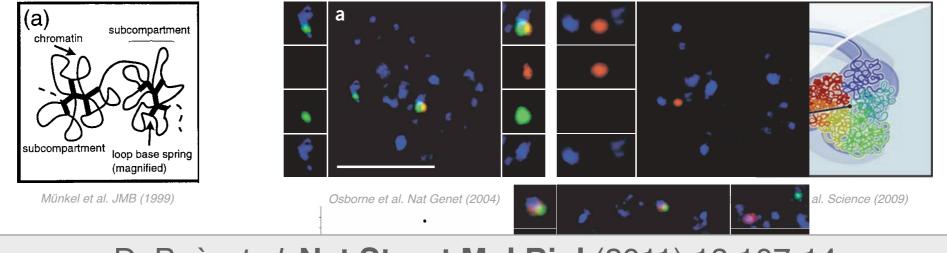






## The "Chromatin Globule" model



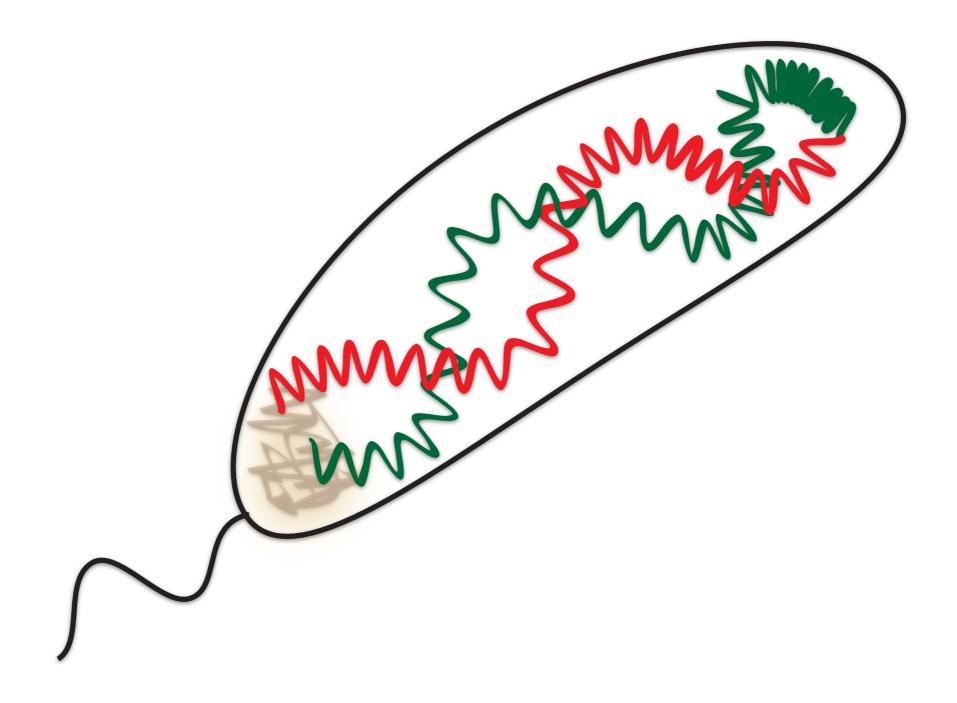


D. Baù *et al.* Nat Struct Mol Biol (2011) 18:107-14 A. Sanyal *et al.* Current Opinion in Cell Biology (2011) 23:325–33.

17

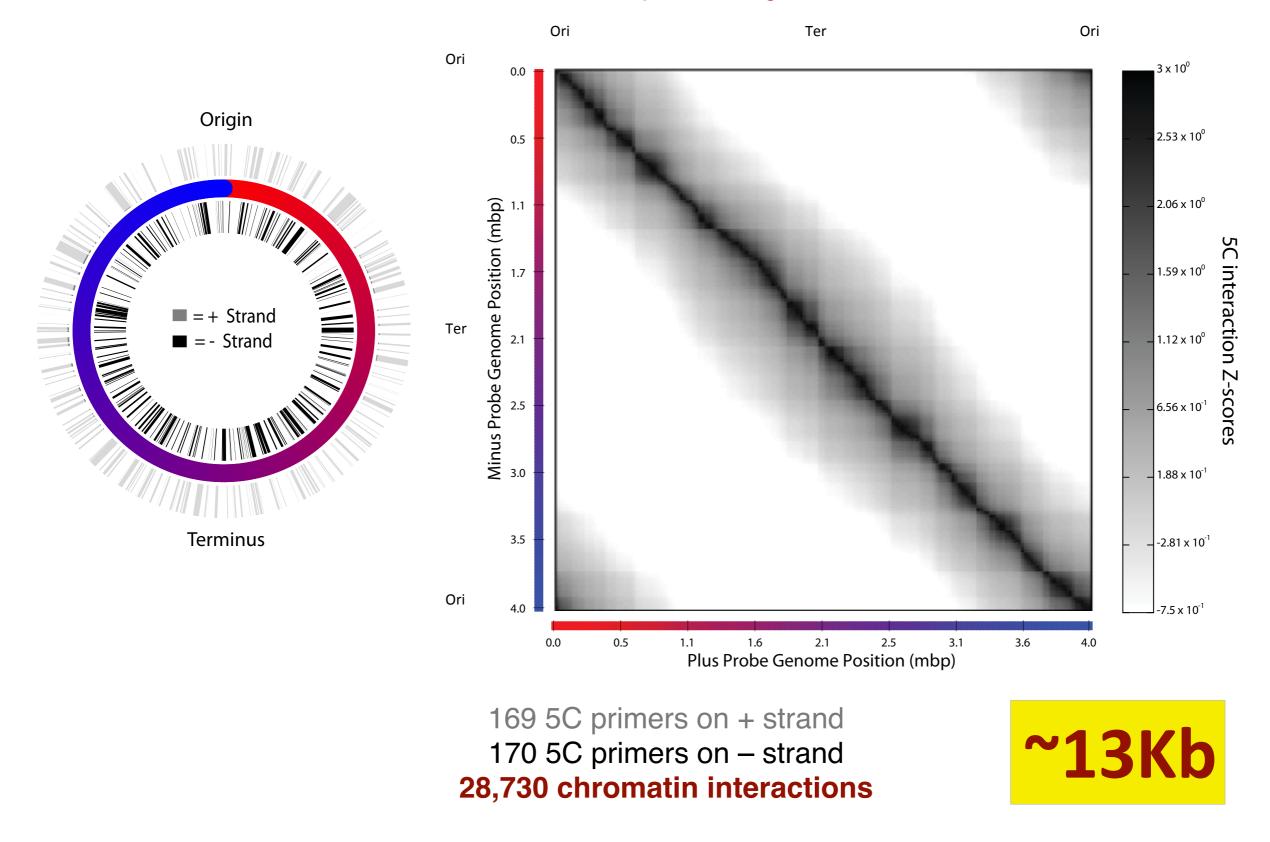
Friday, November 4, 11

## Caulobacter crescentus genome



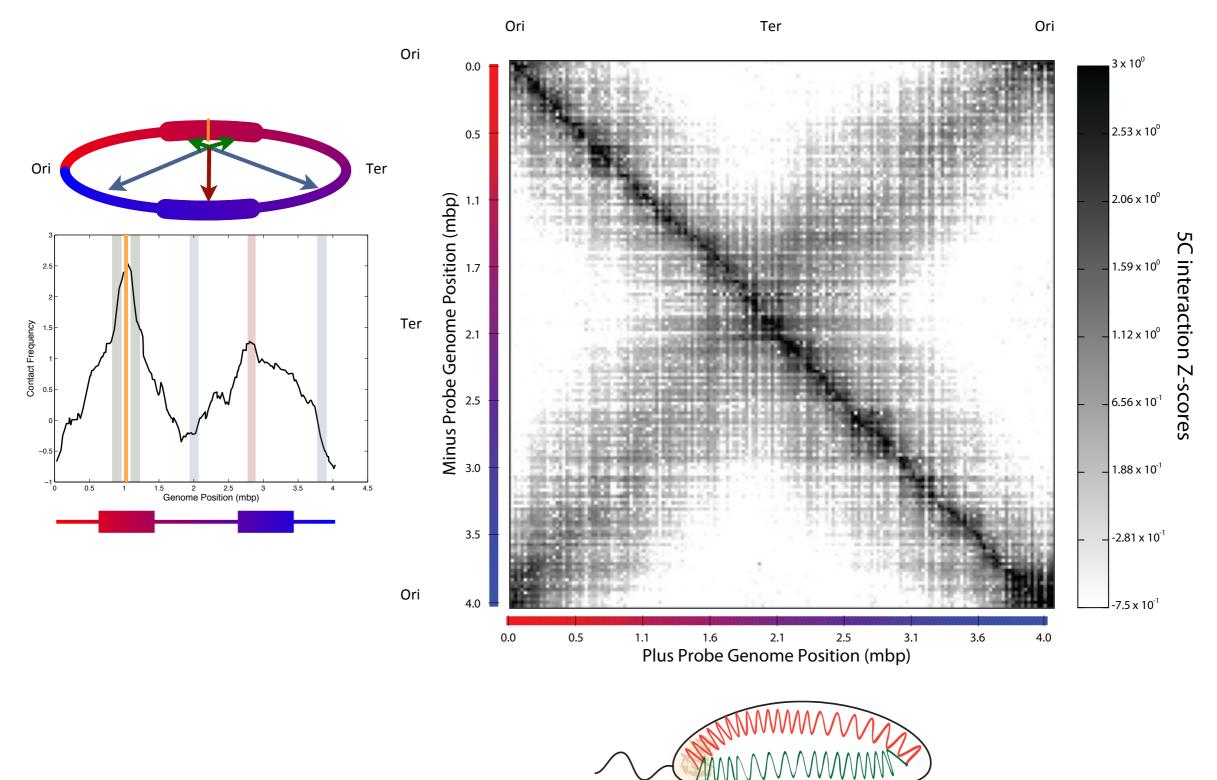
## The 3D architecture of Caulobacter Crescentus

4,016,942 bp & 3,767 genes

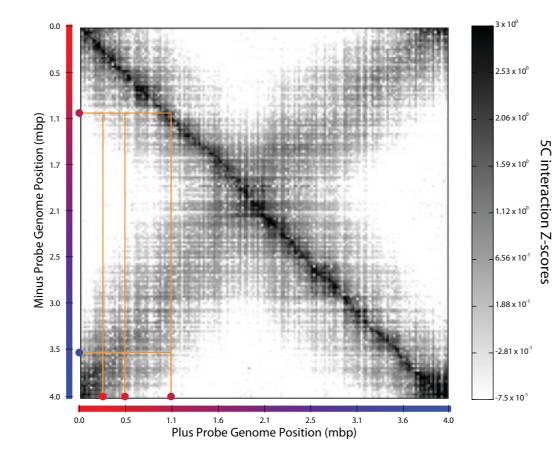


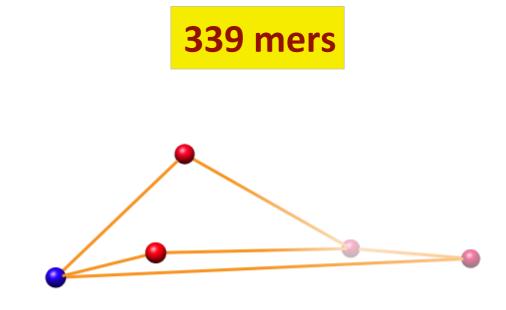
## **5C interaction matrix**

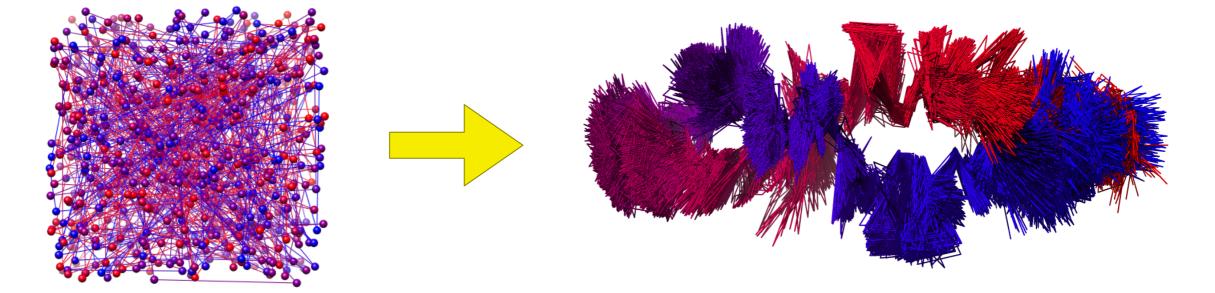
**ELLIPSOID** for Caulobacter cresentus



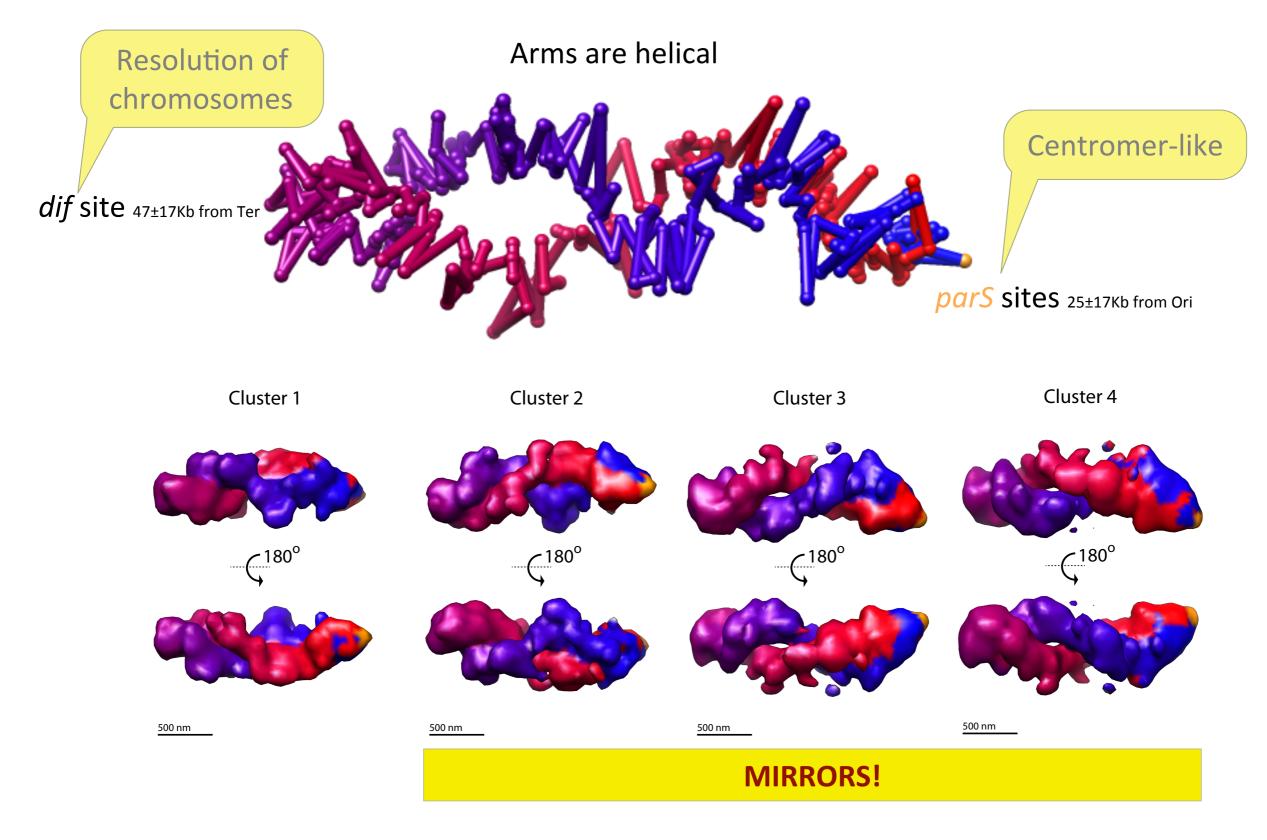
## 3D model building with the 5C + IMP approach



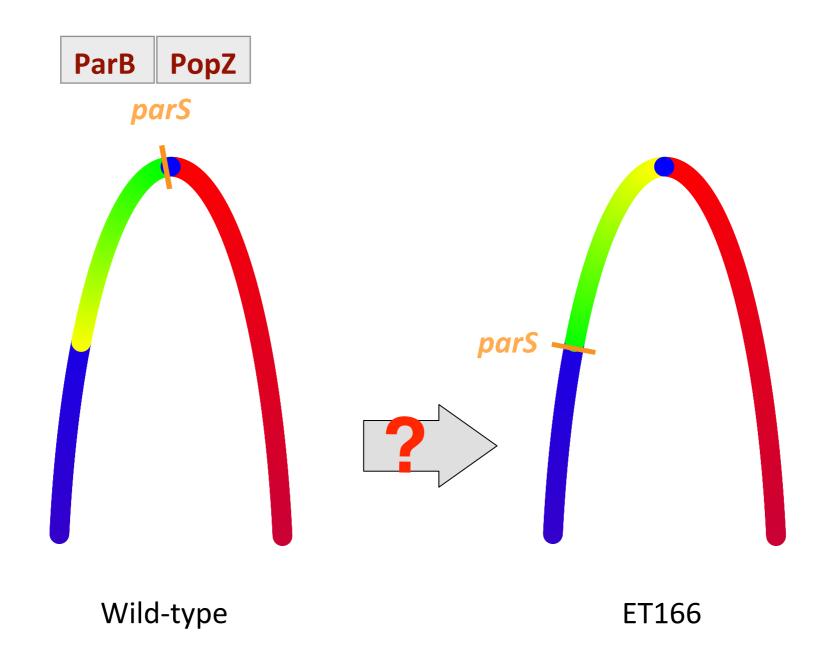




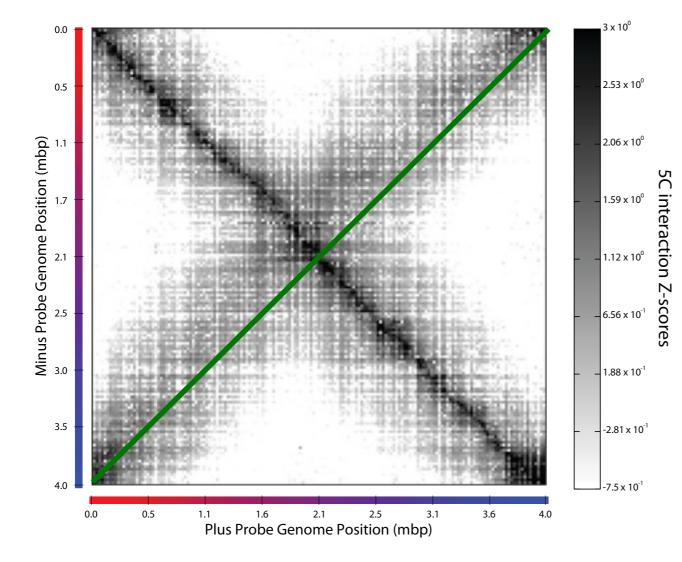
## Genome organization in Caulobacter crescentus

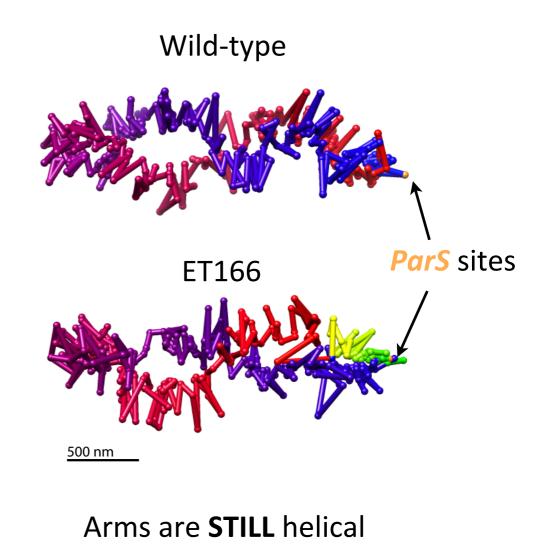


## Moving the *parS* sites 400 Kb away from Ori

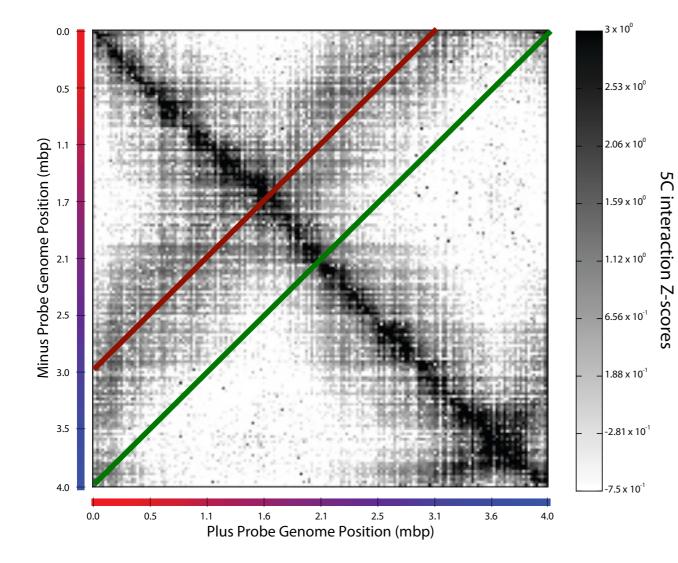


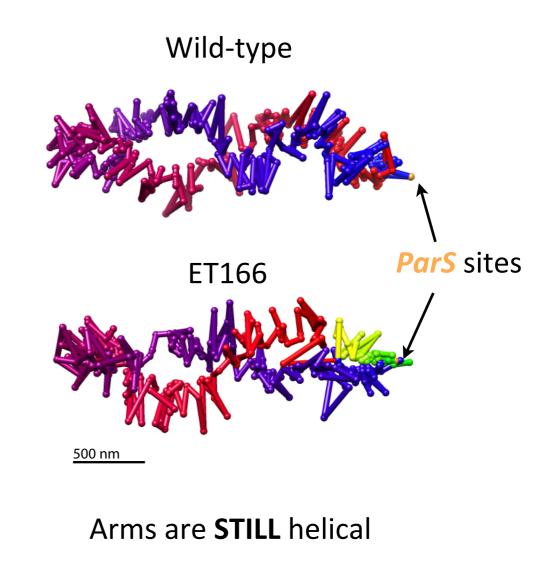
## Moving the *parS* sites results in whole genome rotation!





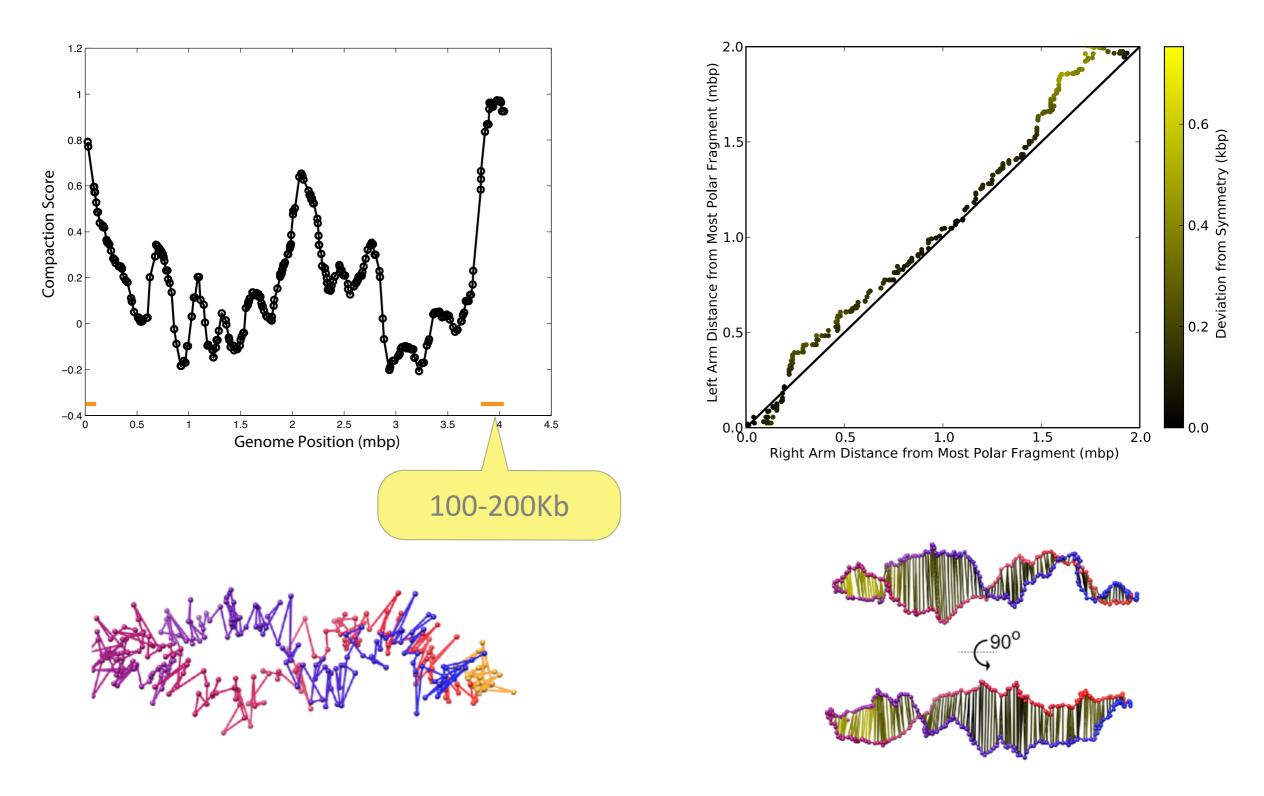
## Moving the *parS* sites results in whole genome rotation!



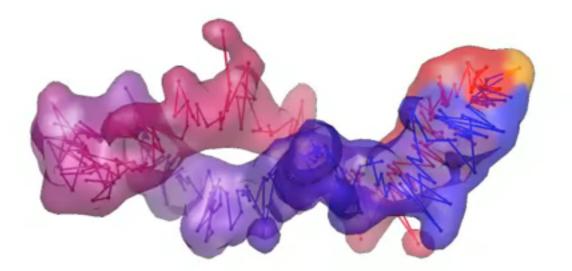


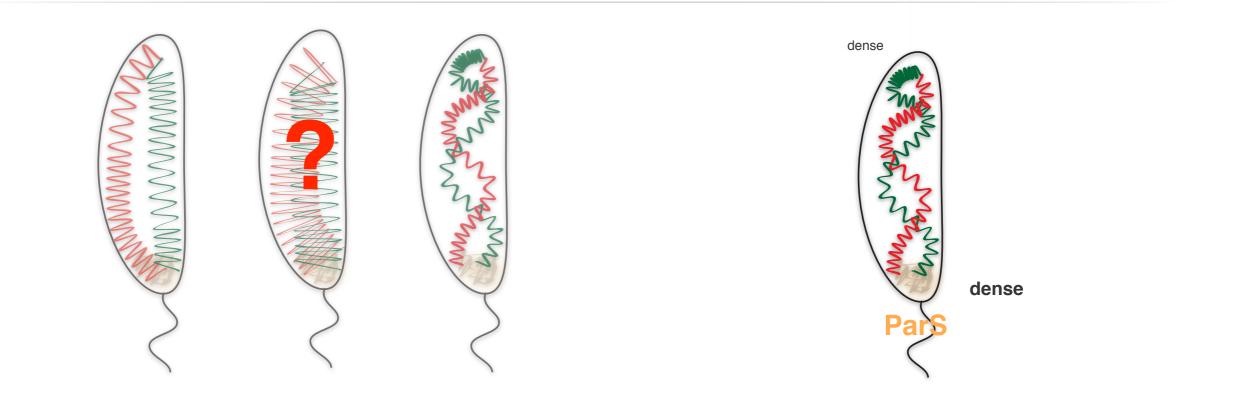
# *parS* sites initiate compact chromatin domain

## Chromosome arms are equidistant to the cell center



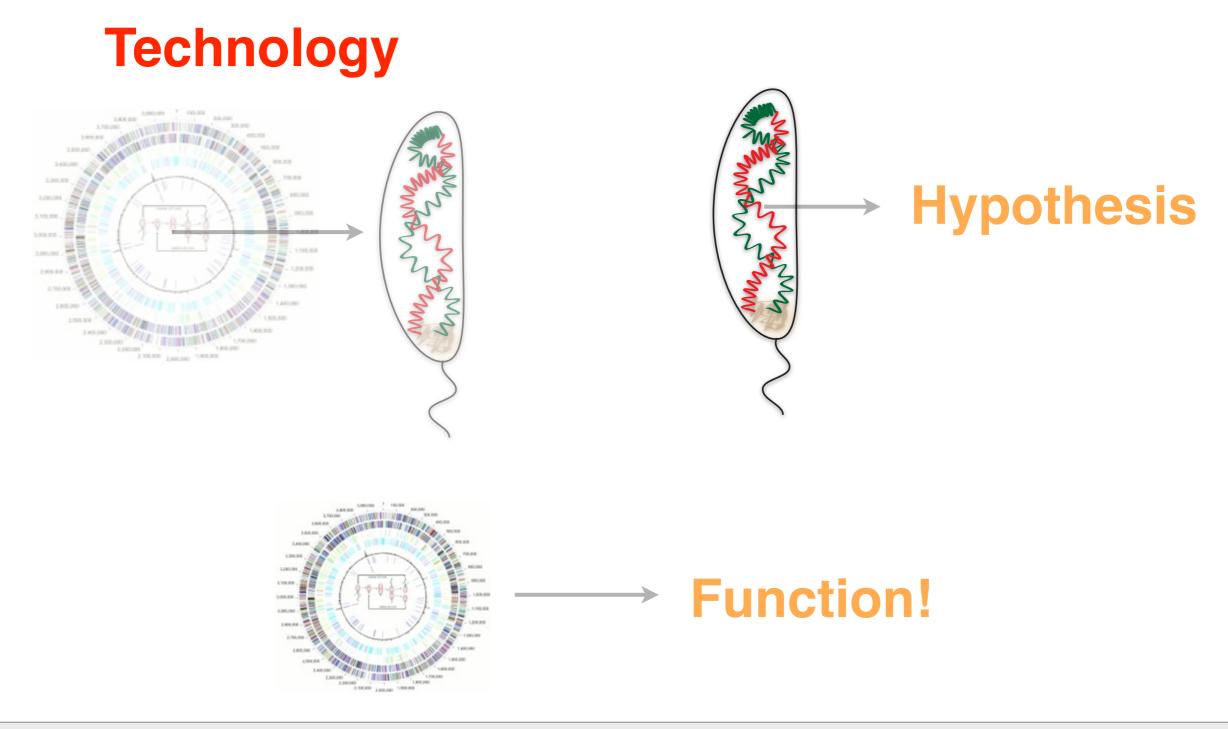
## Genome architecture in Caulobacter





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

## From Sequence to Function 5C + IMP



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

## **PLoS CB Outlook**

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

#### OPEN d ACCESS Freely available online

#### PLOS COMPUTATIONAL BIOLOGY

#### Review

### Bridging the Resolution Gap in Structural Modeling of 3D Genome Organization

#### Marc A. Marti-Renom<sup>1</sup>\*, Leonid A. Mirny<sup>2</sup>

1 Structural Genomics Laboratory, Bioinformatics and Genomics Department, Centro de Investigación Príncipe 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 has 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.

#### This is an "Editors' Outlook" article for PLoS Computational Biology

Recent experimental and computational advances are resulting in an increasingly accurate and detailed characterization of how genomes are organized in the three-dimensional (3D) space of the nucleus (Figure 1) [1]. At the lowest level of chromatin organization, naked DNA is packed into nucleosomes, which forms the so-called chromatin fiber composed of DNA and proteins. However, this initial packing, which reduces the length of the DNA by about seven times, is not sufficient to explain the higher-order folding of chromosomes during interphase and metaphase. It is now accepted that chromosomes and genes are non-randomly and dynamically positioned in the cell nucleus during the interphase, which challenges the classical representation of genomes as linear static sequences. Moreover, compartmentalization, chromatin organization, and spatial location of genes are associated with gene expression and the functional status of the cell. Despite the importance of 3D genomic architecture, we have a limited understanding of the molecular mechanisms that determine the higher-order organization of genomes and its relation to function. Computational biology plays an important role in the plethora of new technologies aimed at addressing this knowledge gap [2]. Indeed, Thomas Cremer, a pioneer in studying nuclear organization using light microscopy, recently highlighted the importance of computational science in complementing and leveraging experimental observations of genome organization [2]. Therefore, computational approaches to integrate experimental observations with chromatin physics are needed to determine the architecture (3D) and dynamics (4D) of genomes.

We present two complementary approaches to address this challenge: (i) the first approach aims at developing simple polymer models of chromatin and determining relevant interactions (both

. PLoS Computational Biology | www.ploscompbiol.org

echnology, Cambridge, Massachusetts, United States of America 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 proteinfolding 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 microcopy [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 colocalization of loci and nuclear substructures.

Using cellular and molecular biology, novel chromosome conformation capture (3C)-based methods such 3C [9], 3C-onchip 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:** 0 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. \* F-mail: mmarti@cinf.es

1 July 2011 | Volume 7 | Issue 7 | e1002125

#### (ii). PloS Computational Biology | www.ploscomphic.org.

764 2011 | Volume 7 | 1564 7 | e1002135

Equivarial observations with discretizin physics are resulted to experiment of a minimum SD and dynamics (201) of generation  $W_{2}$  present (see our We present (see our dynamics) approaches in address dis observange () for first approach and discretizing relevant in physics module of external and discretizing relevant infermation (bells module of external and discretizing relevant infermation (bells

and particular the second second

will make a properties of the material sector of sector sector of the sector sector of the sector sector



# OPEN POSITIONS 12 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 Postdoctoral fellow CIPF



### PRINCIPE FELIPE CENTRO DE INVESTIGACION

### Marc A. Marti-Renom

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

### Job Dekker

UMASS.

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

## http://sgu.bioinfo.cipf.es http://integrativemodeling.org