

# Integrative modeling applied to chromatin

**Davide Baù & François Serra**

Genome Biology Group (CNAG)  
Structural Genomics Group (CRG)

# Course outline

Theory Practice

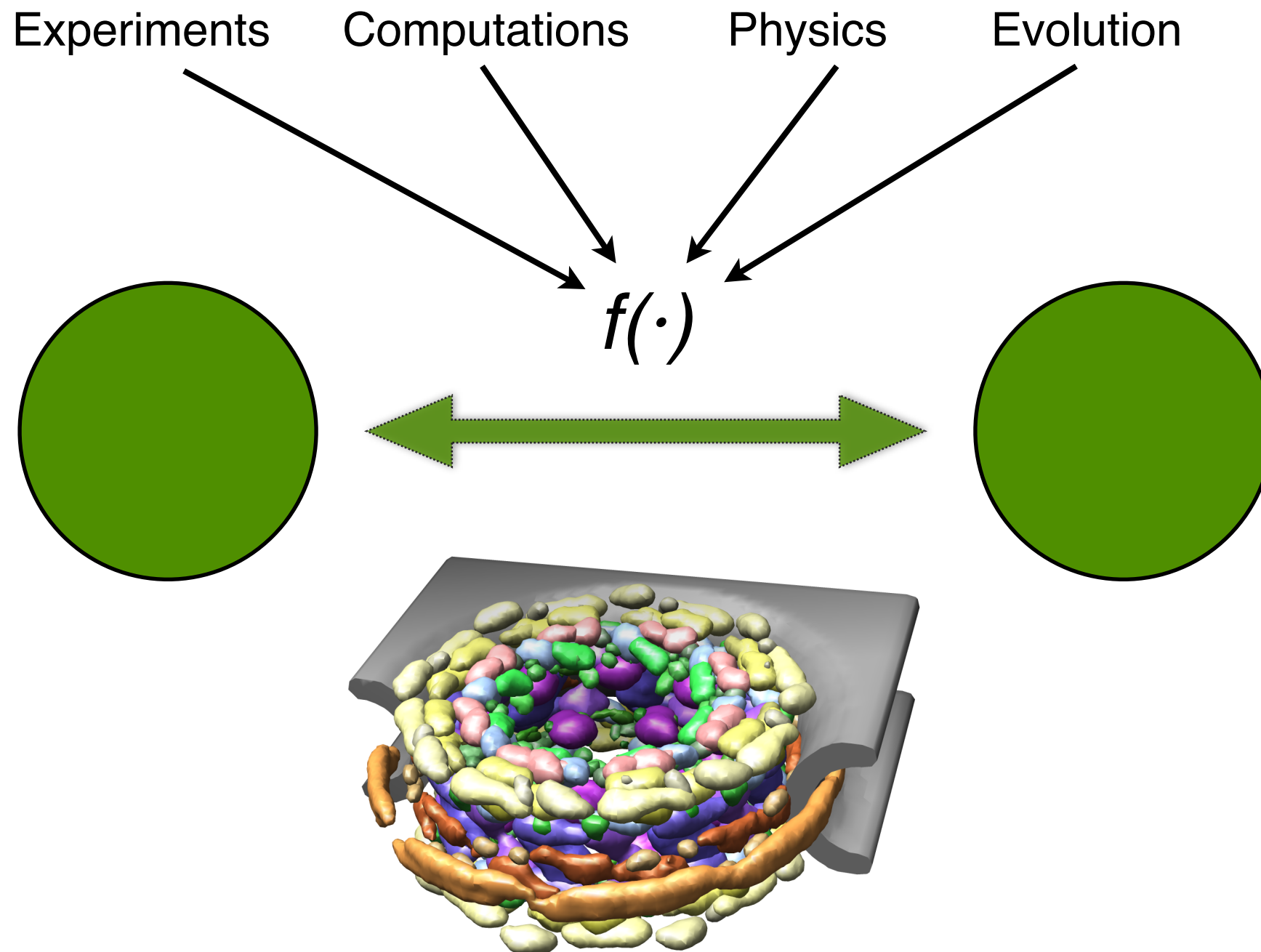
- Day 1** Introduction to structure determination  
Chromatin structure and Hi-C data  
Introduction to linux and python (FACULTATIVE)  
The Integrative Modeling Platform and Chimera
- Day 2** The Integrative Modeling Platform applied to chromatin  
TADbit introduction and installation  
Topologically Associated Domains detection and analysis
- Day 3** The TADbit documentation: examples and code snippets  
3D modeling of real Hi-C data  
Analysis of the results



# The Integrative Modeling Platform framework

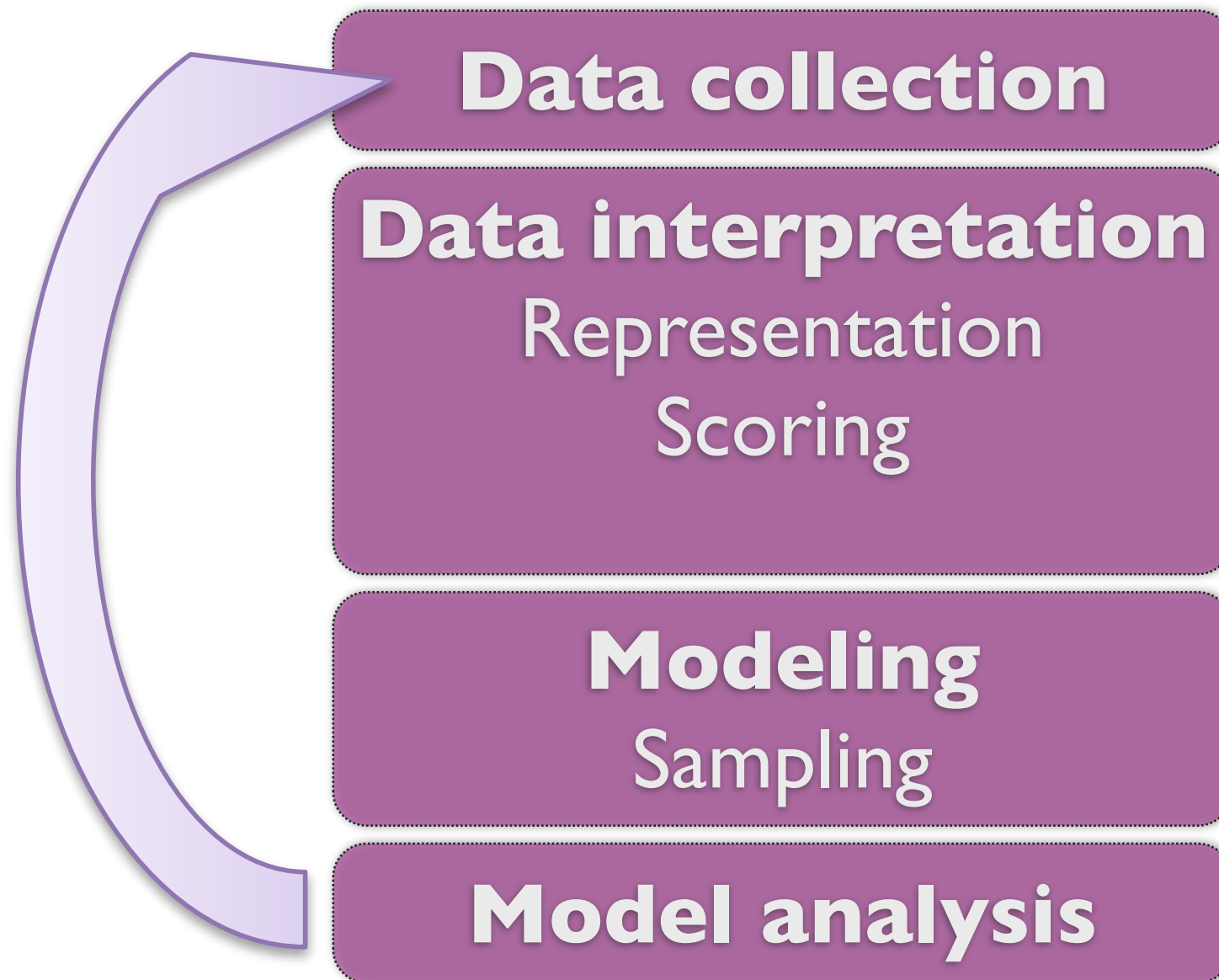
<http://www.integrativemodeling.org>

Russel, D. et al. PLOS Biology 10, e1001244 (2012).



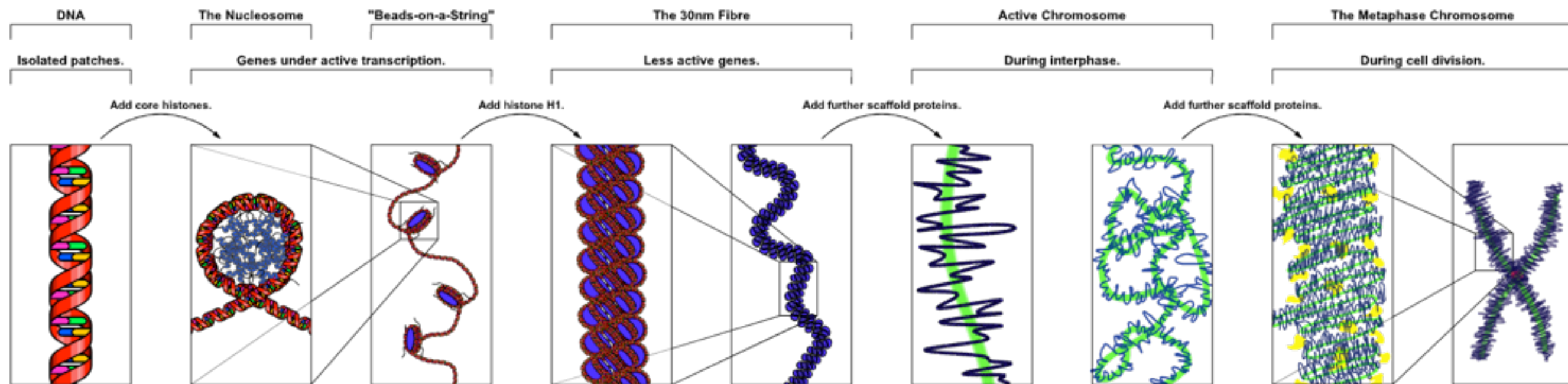
From Alber, F. et al. Nature 450, 695–701 (2007).

# Integrative modeling





# Chromatin structures



# Chromatin structure

**Chromatin = DNA + (histone) proteins**

**The genome is well organized and hierarchically packaged**

**Histone modifications affect chromatin structure and activity**

**3C-like data measure the frequency of interaction between distant loci**



# Linux commands

Command	Description	Example	Action
pwd	print working directory	pwd	path & name of dir. I am in now
ls	list contents of directory	ls	list contents of current dir.
		ls test/	list contents of the test dir. that hangs from the current working dir.
		ls -lh	vertical list of dir. contents
cd	change directory	cd	go to home directory
		cd /home/user/Docs	go to the Docs directory
		cd ..	go to parent directory
mkdir	make directory	mkdir test	creates directory test/
rmdir	remove directory	rmdir test	remove test/ if empty
cp	copy	cp fileA fileB	copy fileA to fileB
mv	move or rename file or directory	mv a b	change name from a to b
		mv a ..	move a to parent directory
more	see file contents	more a.txt	see contents of a.txt
gedit	simple text editor!	gedit a.txt	edit a.txt
firefox	a web and directory browser	firefox a.html or firefox a.jpg	use web browser to view file
info or man	information on a command	info ls	manual page for the 'ls' command

# Python definitions

- variables
- loops
- conditionals
- lists, tuples, dictionaries

```
a = 1  
b = 3.14  
c = 'charles'
```

```
for i in range(0, 10, 1):  
    print i
```

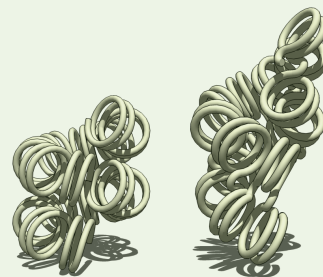
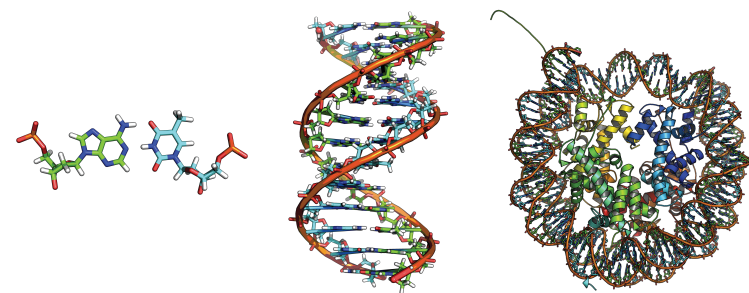
```
i = 0  
while i < 10:  
    # print i  
    i = i + 1
```

```
for i in range(0, 10, 1):  
    if i == 3:  
        print 'we have 3'  
    elif i > 3:  
        print 'we have many'  
    else:  
        print 'we have few'
```

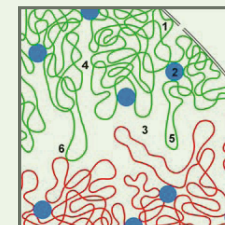
```
a = [0, 1, 2, 3, 4]  
b = (0, 1, 2, 3, 4)  
c = {'one': 11, 'two': 22  
     'three': 33, 'four': 79}
```



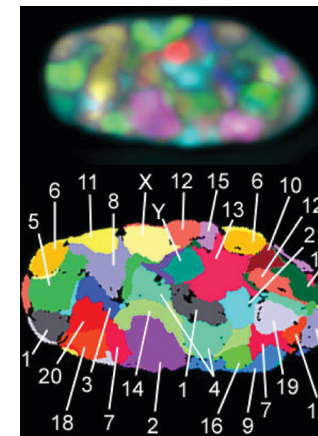
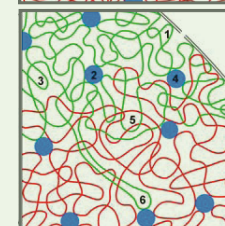
# Knowledge



IDM



INM



DNA length

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

Volume

$10^{-9}$   $10^{-6}$   $10^{-3}$   $10^0$   $10^3$   $\mu\text{m}^3$

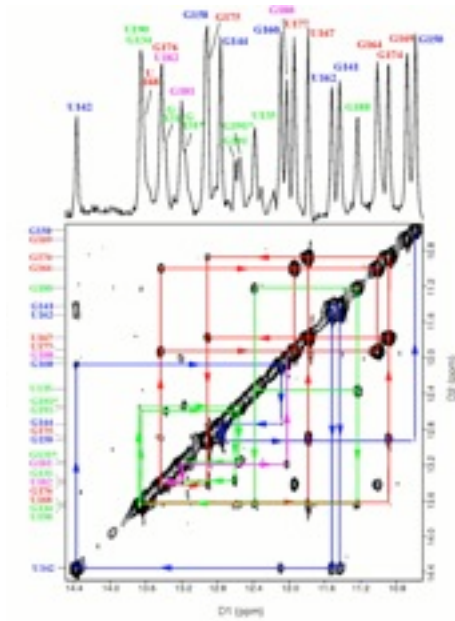
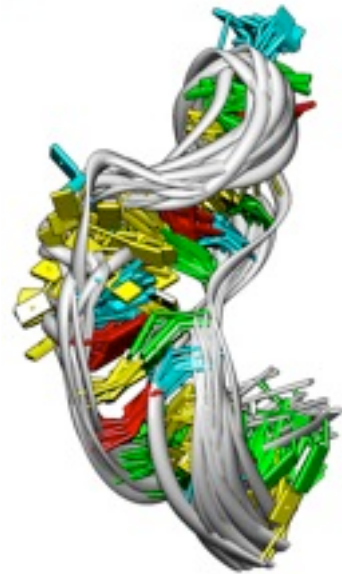
Time

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

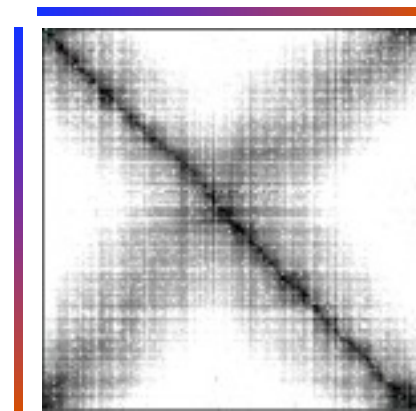
Resolution

$10^{-3}$   $10^{-2}$   $10^{-1}$   $\mu$

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



## Biomolecular structure determination 2D-NOESY data

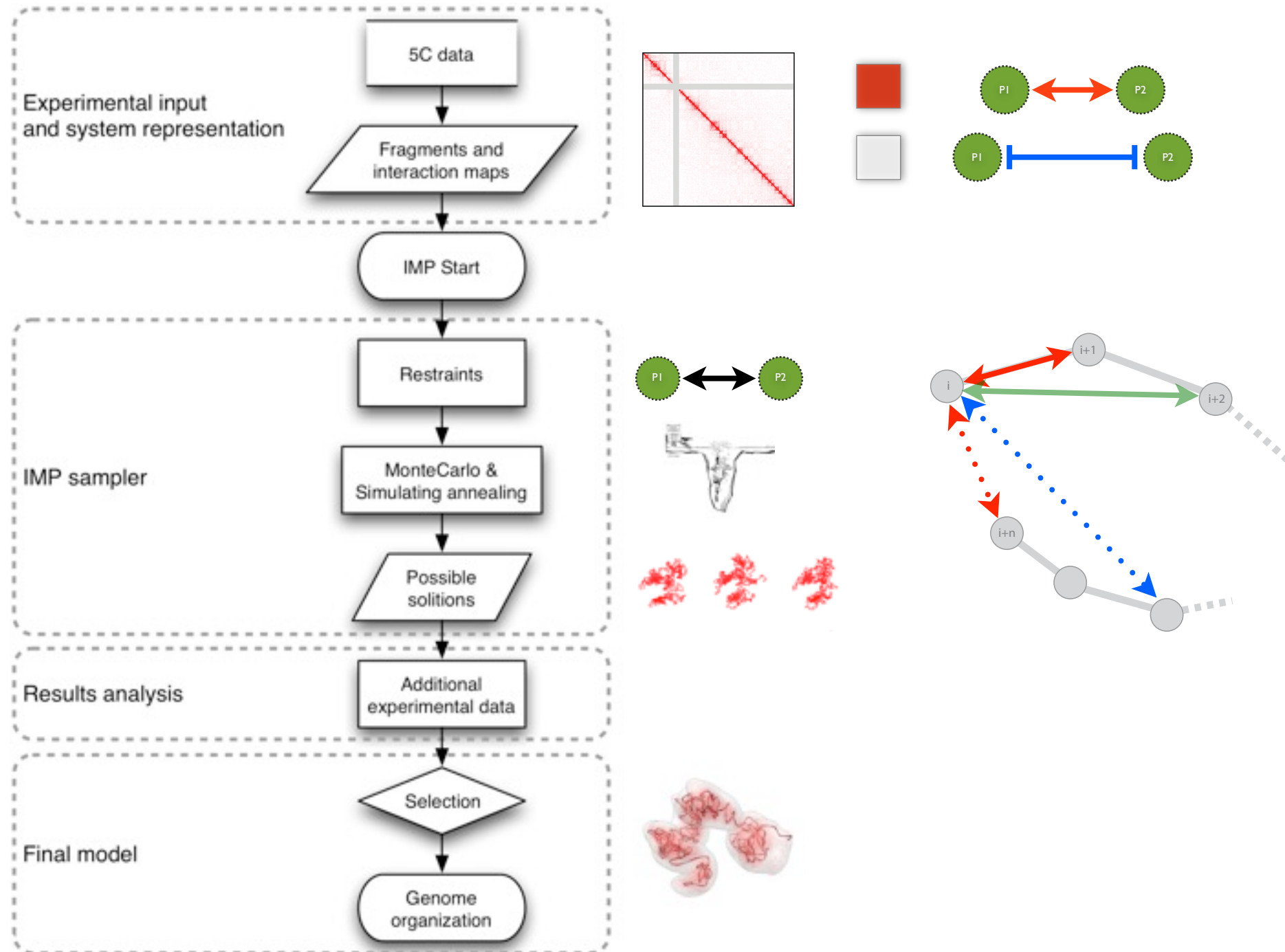


## Chromosome structure determination 5C data



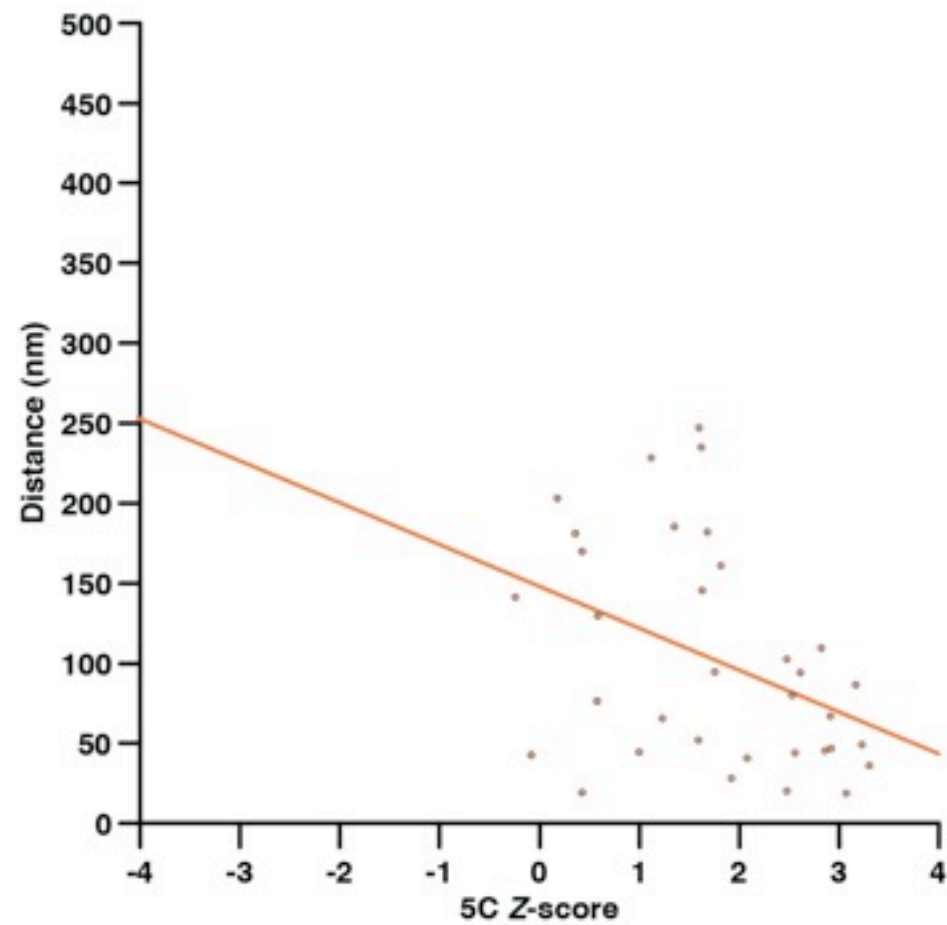
# Integrative modeling applied to chromatin

<http://www.integrativemodeling.org>

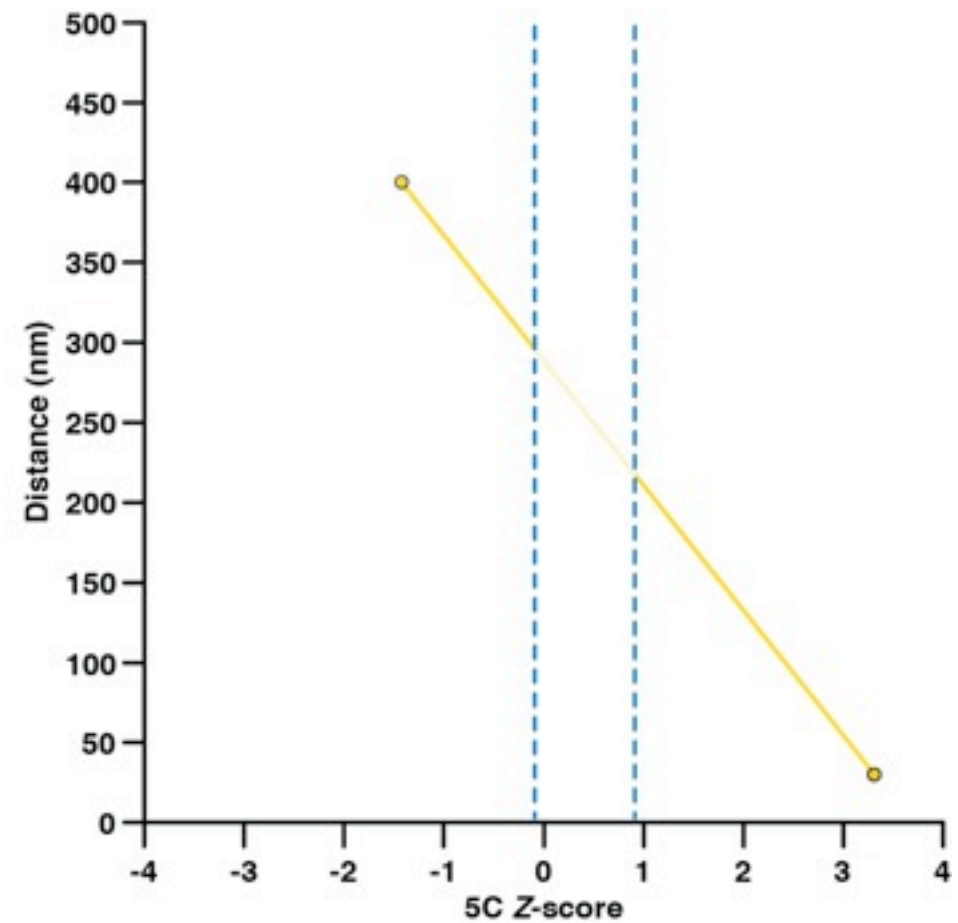


# From 5C data to spatial distances

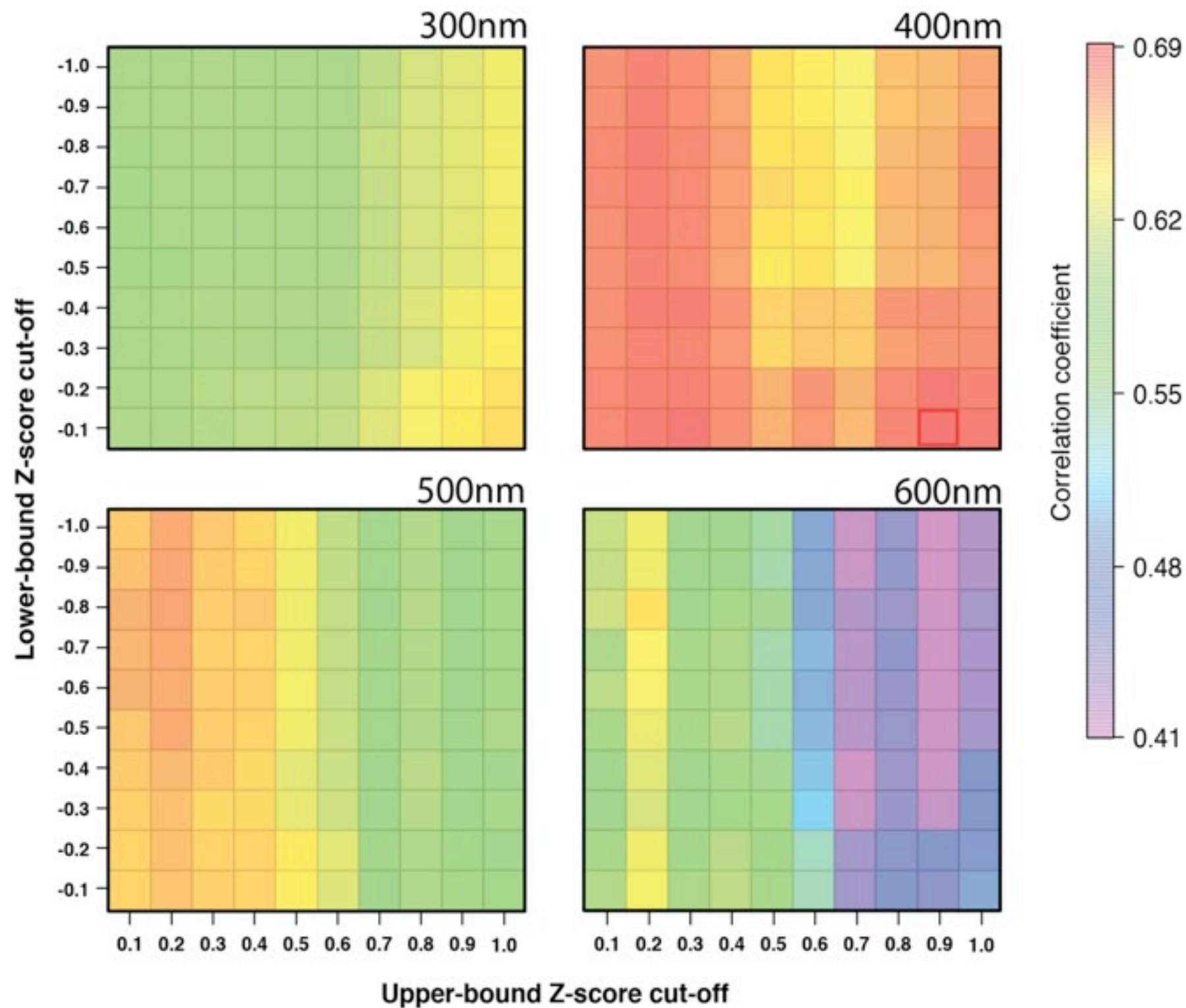
Neighbor fragments



Non-Nighbor fragments



# Parameter optimization



# Representation

Constituent parts of the molecule

## Harmonic

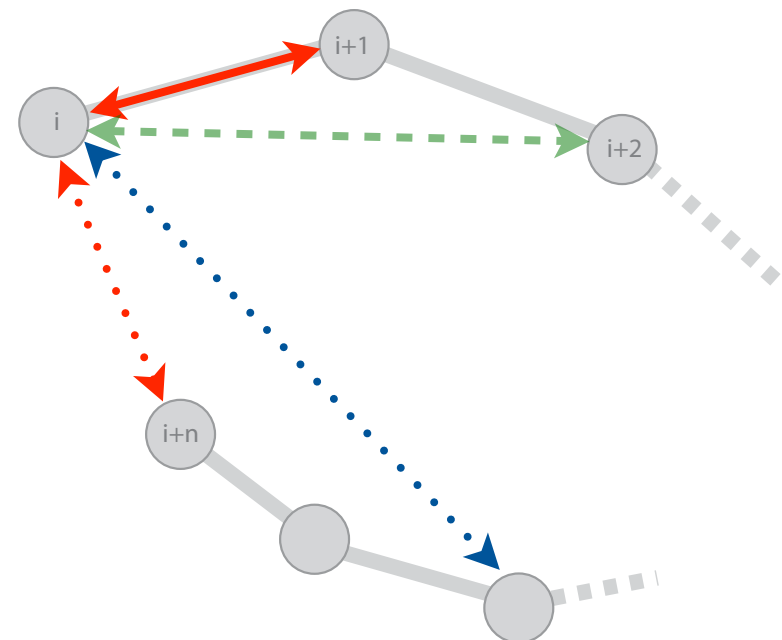
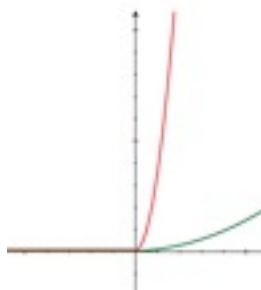
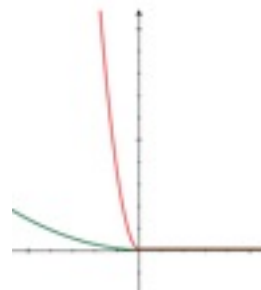
$$H_{i,j} = k(d_{i,j} - d_{i,j}^0)^2$$

## 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

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



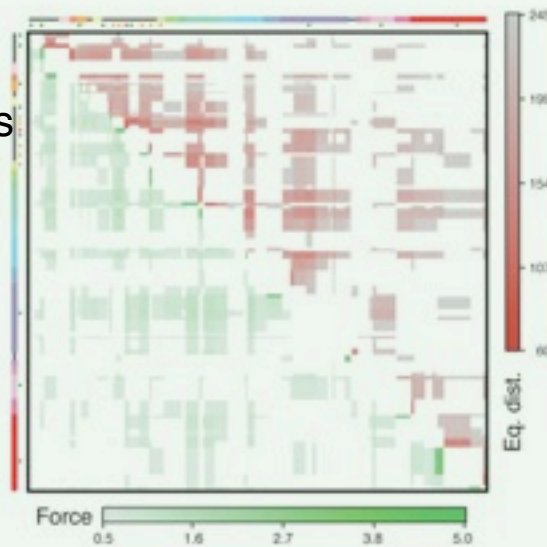


# Scoring

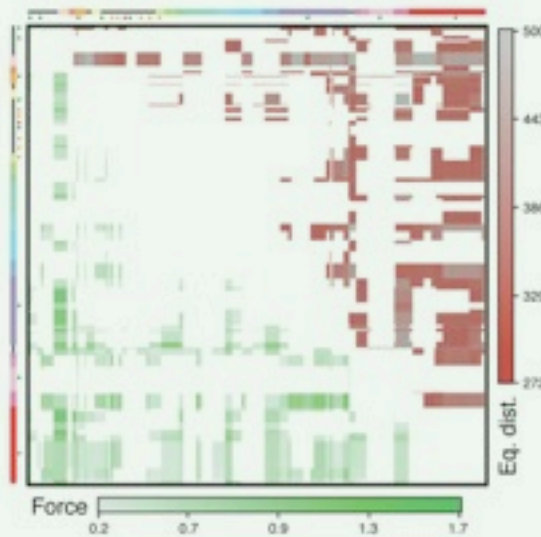
Individual spatial restraints encoding the data

## GM12878

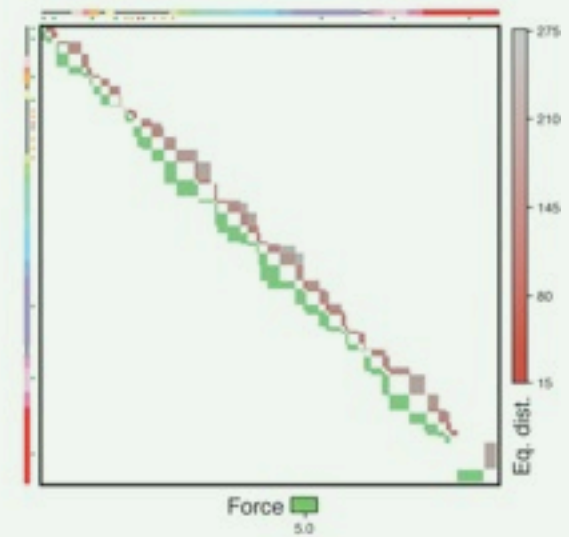
70 fragments  
1,520 restraints



Harmonic



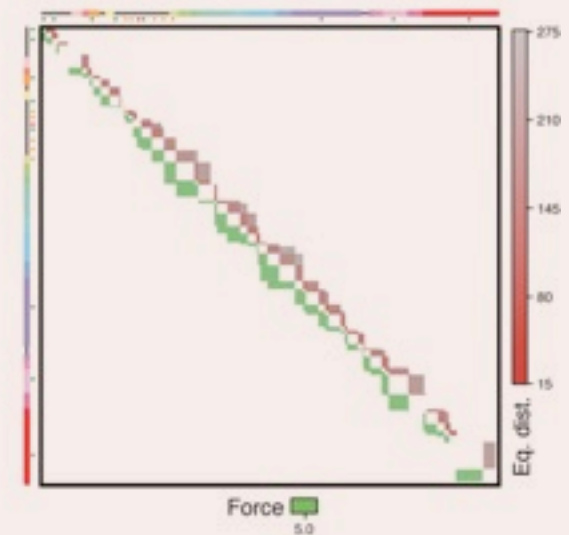
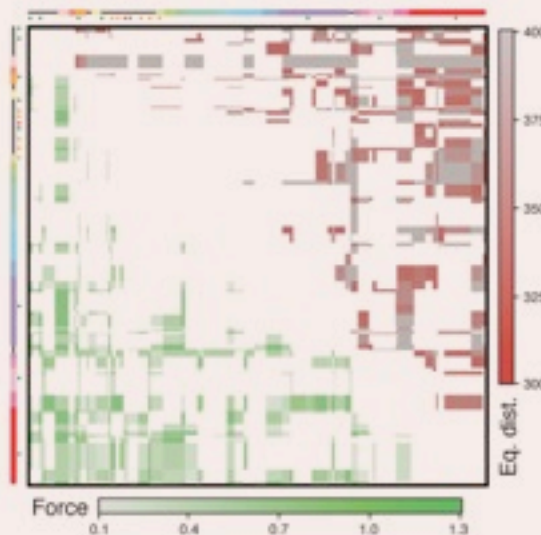
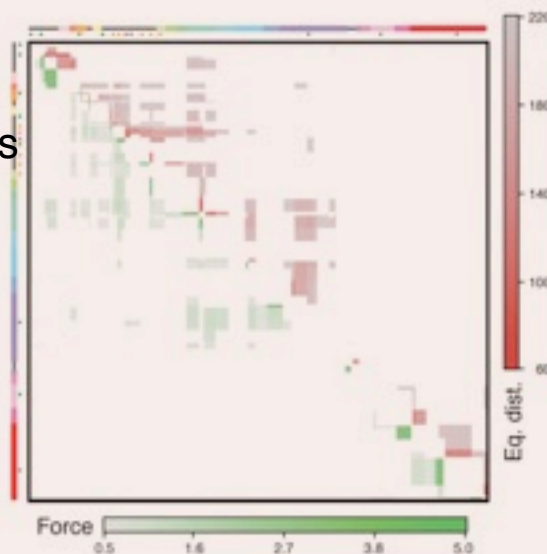
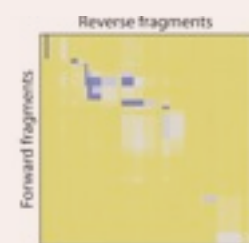
Harmonic Lower Bound



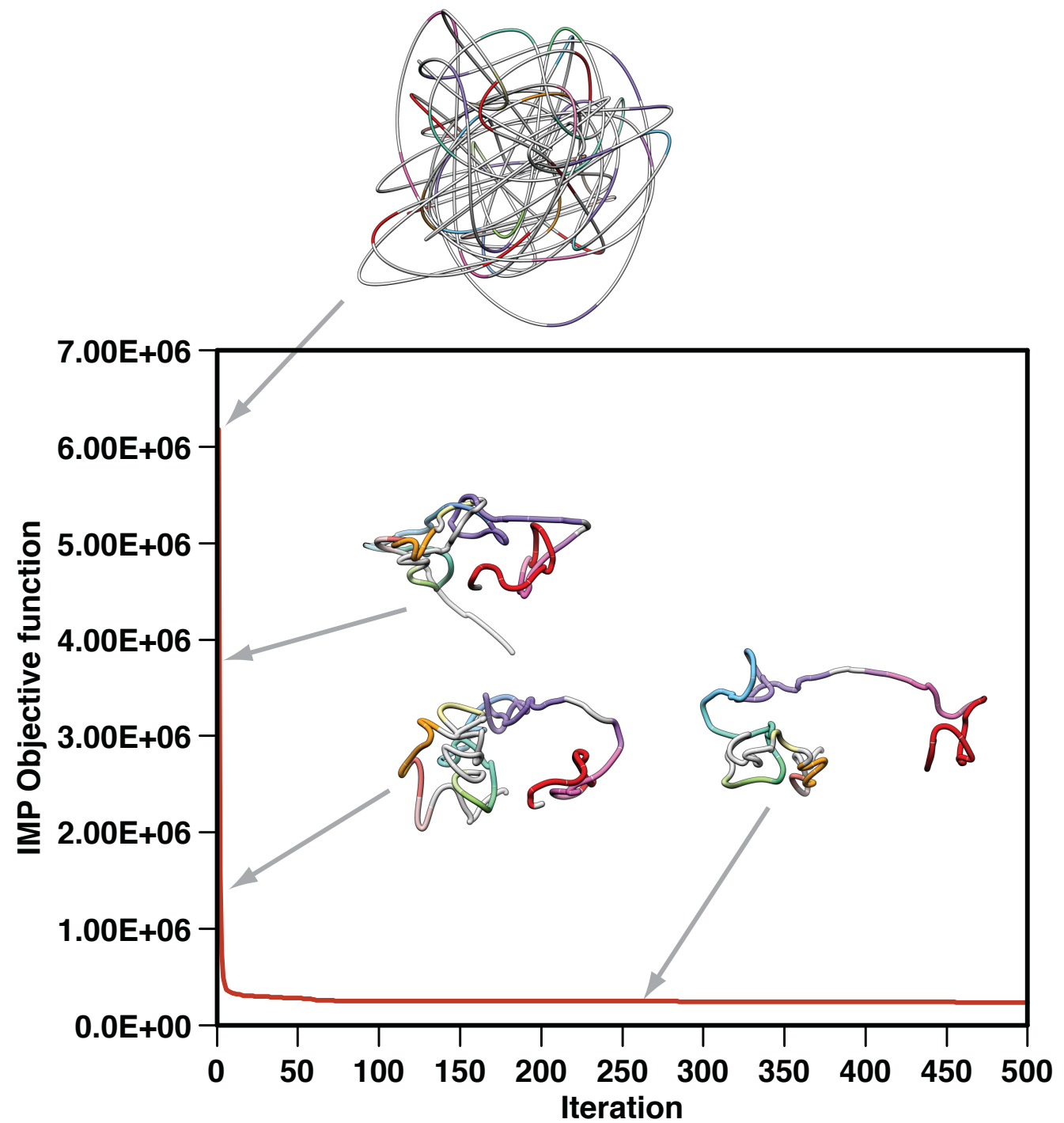
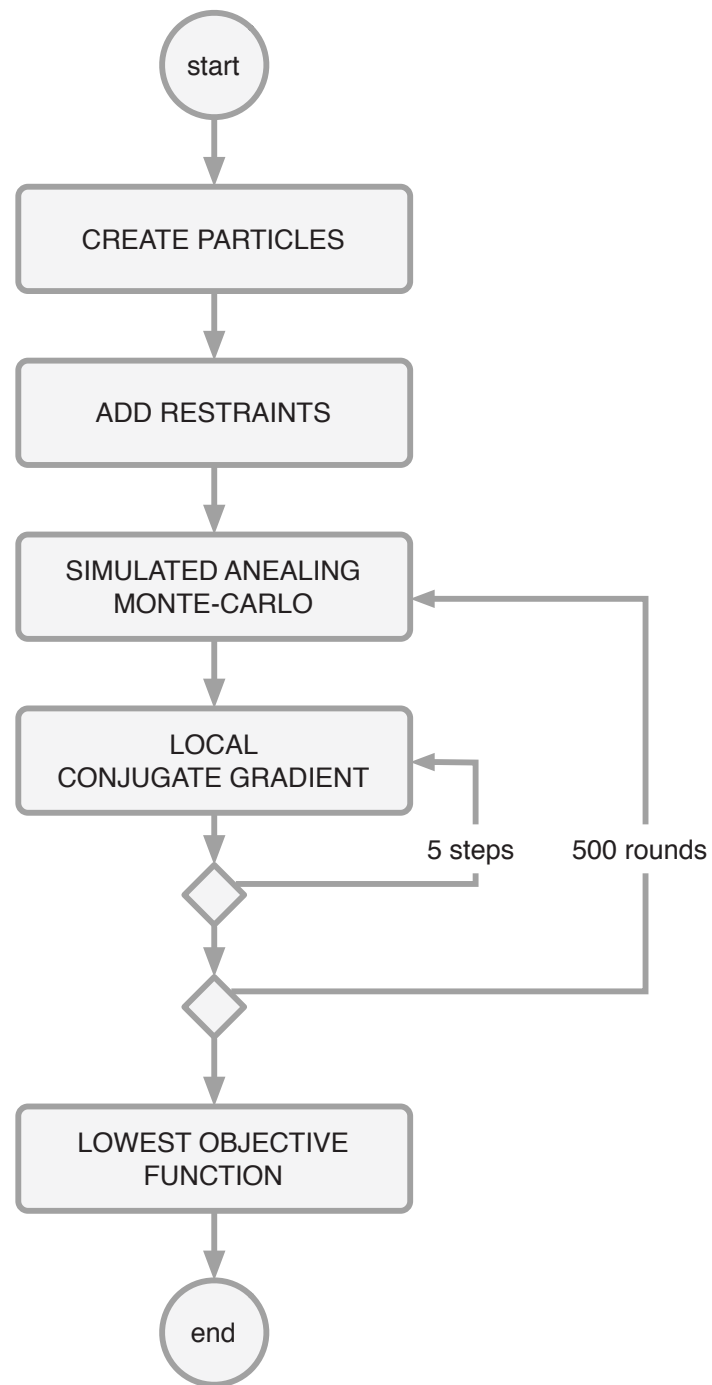
Harmonic Upper Bound

## K562

70 fragments  
1,049 restraints



# Optimization of the scoring function



# Human $\alpha$ -globin domain



Amartya Sanyal Bryan Lajoie

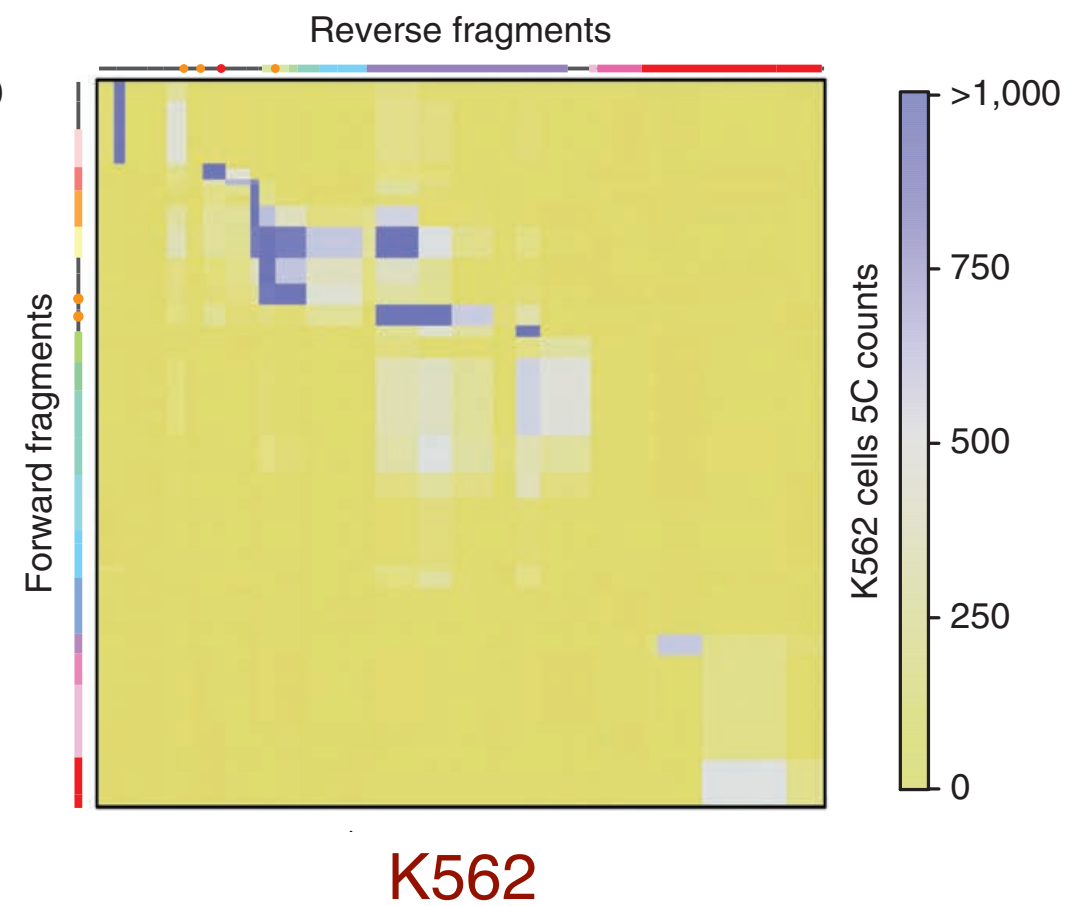
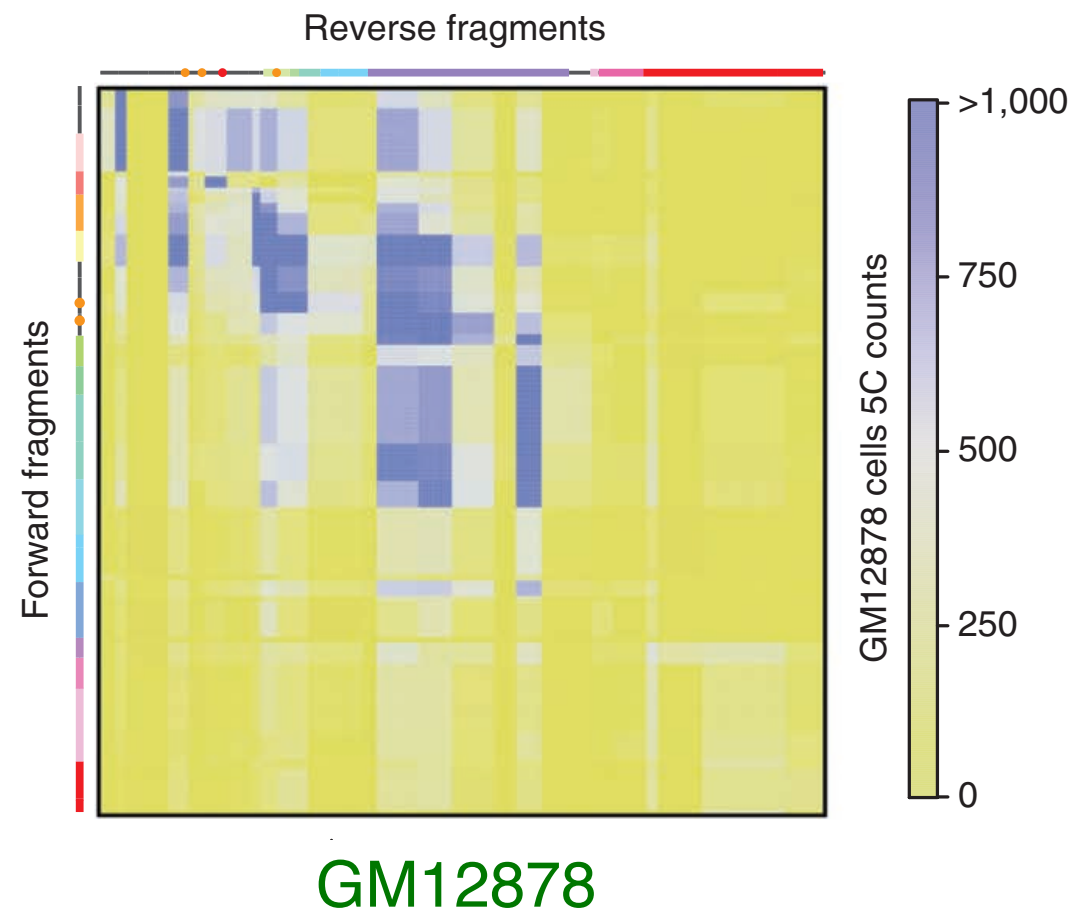
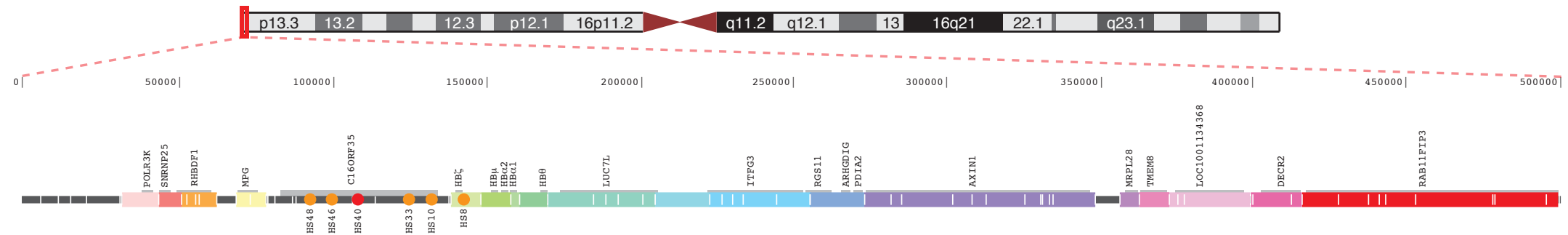


The ENCODE data for ENM008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute ChIP-seq (Bernstein Group at Broad Institute of Harvard and MIT).

# Human $\alpha$ -globin domain

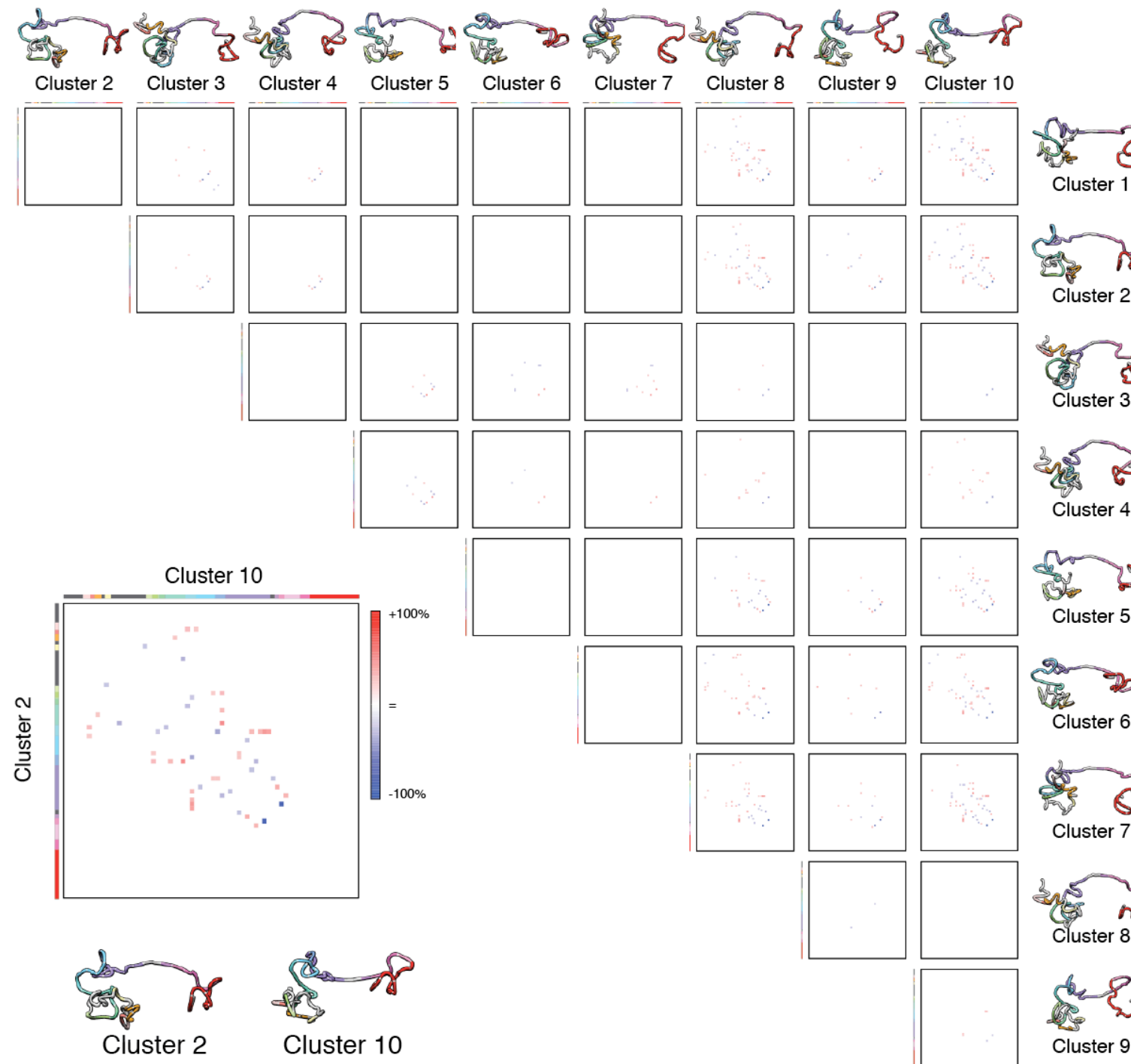
## ENm008 genomic structure and environment

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



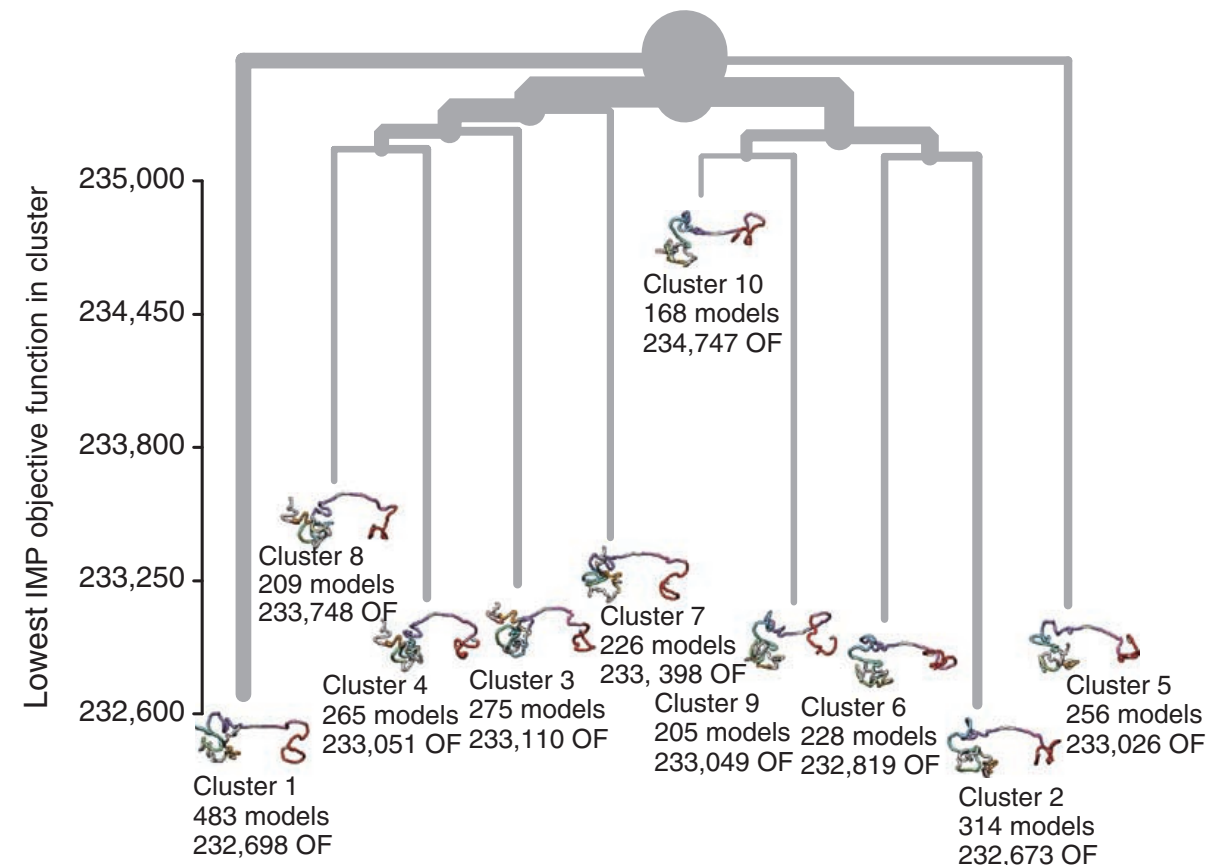
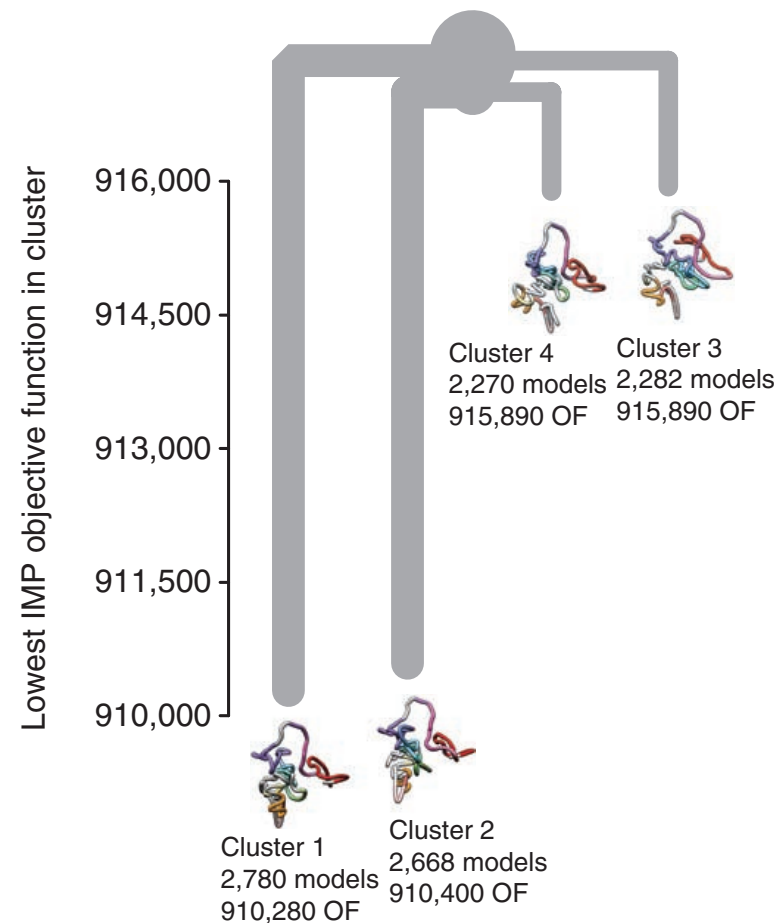
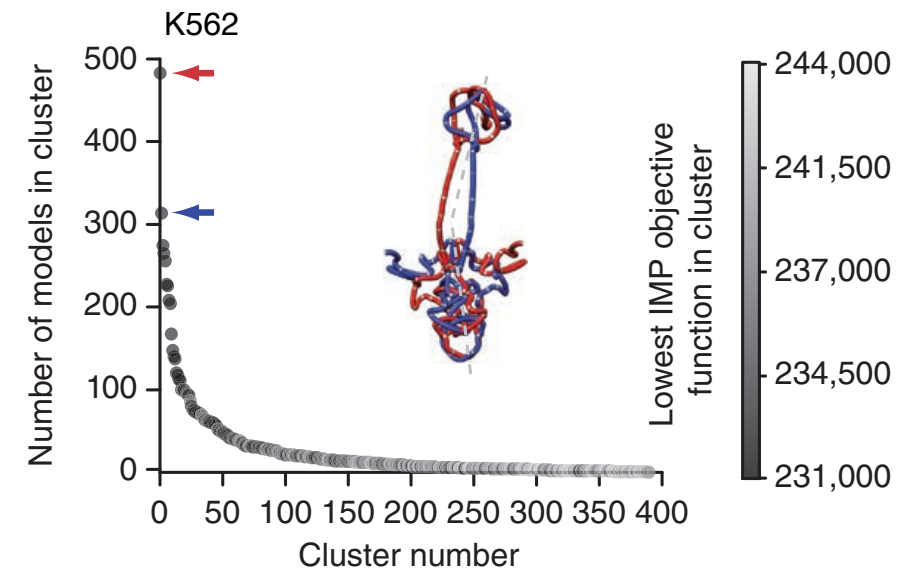
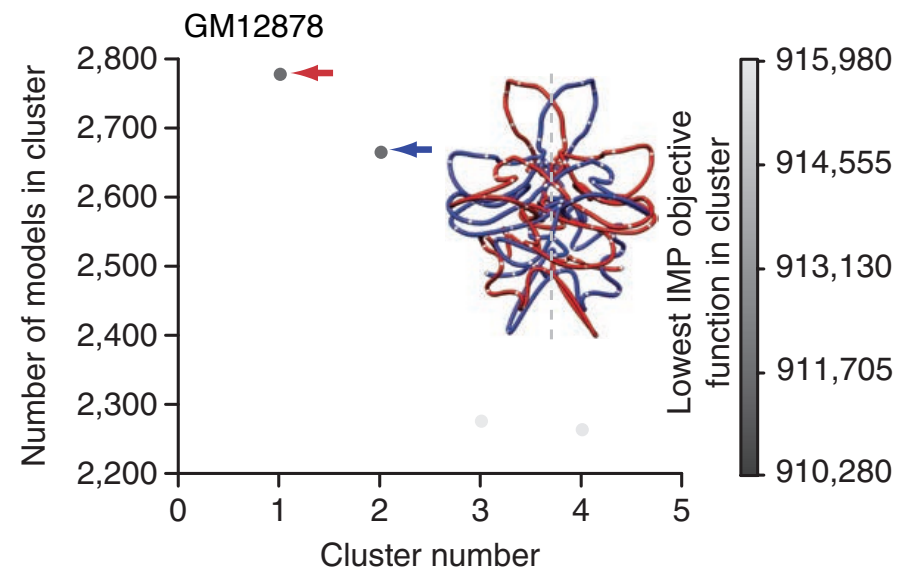
# Not just one solution

5C de-convolution analysis

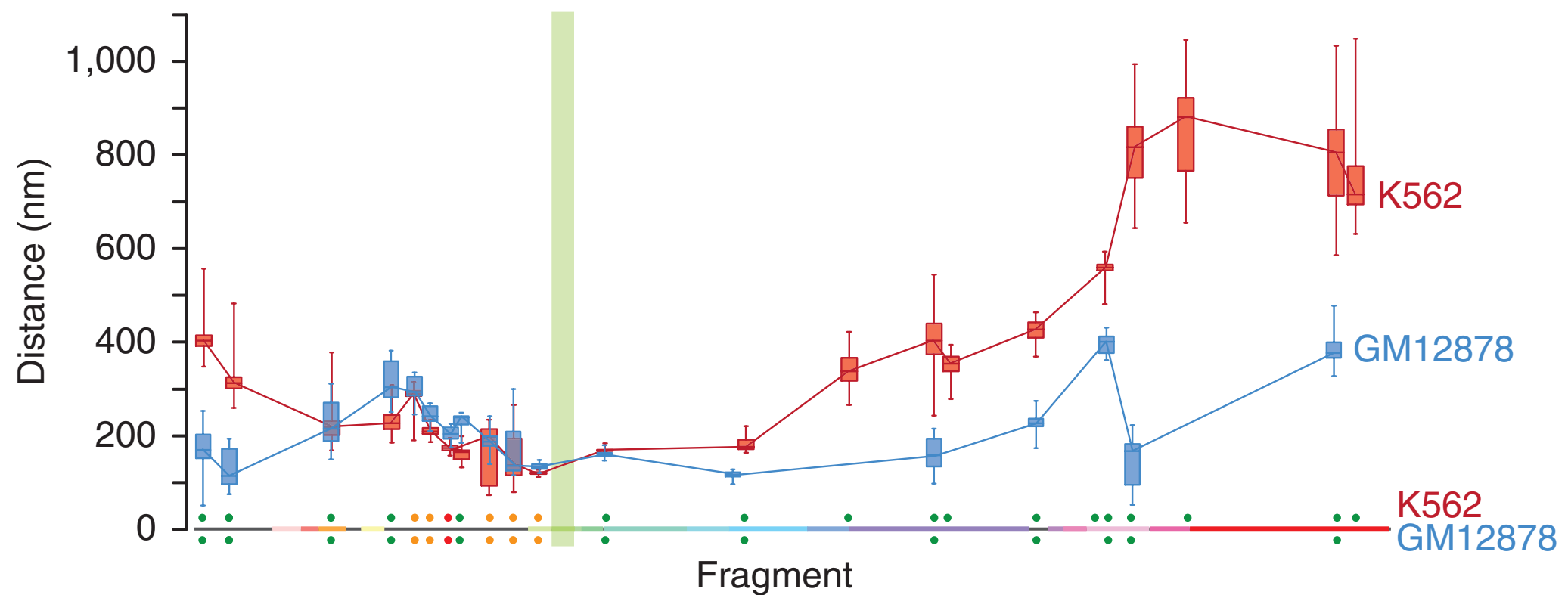
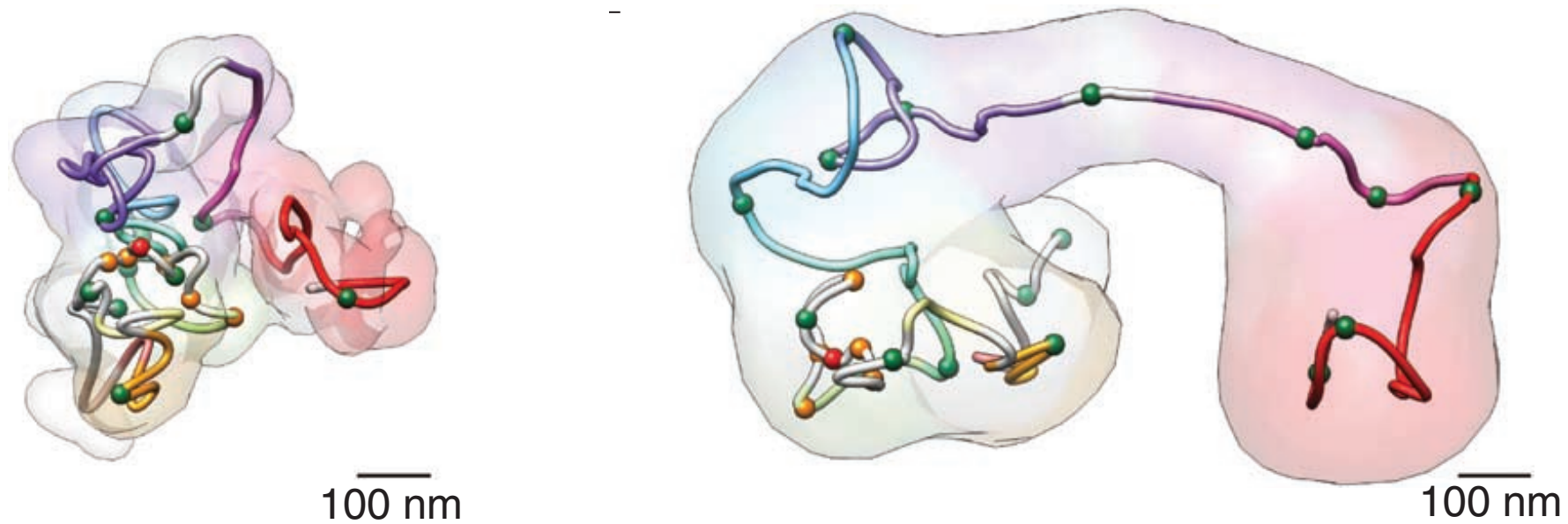




# Ensemble analysis

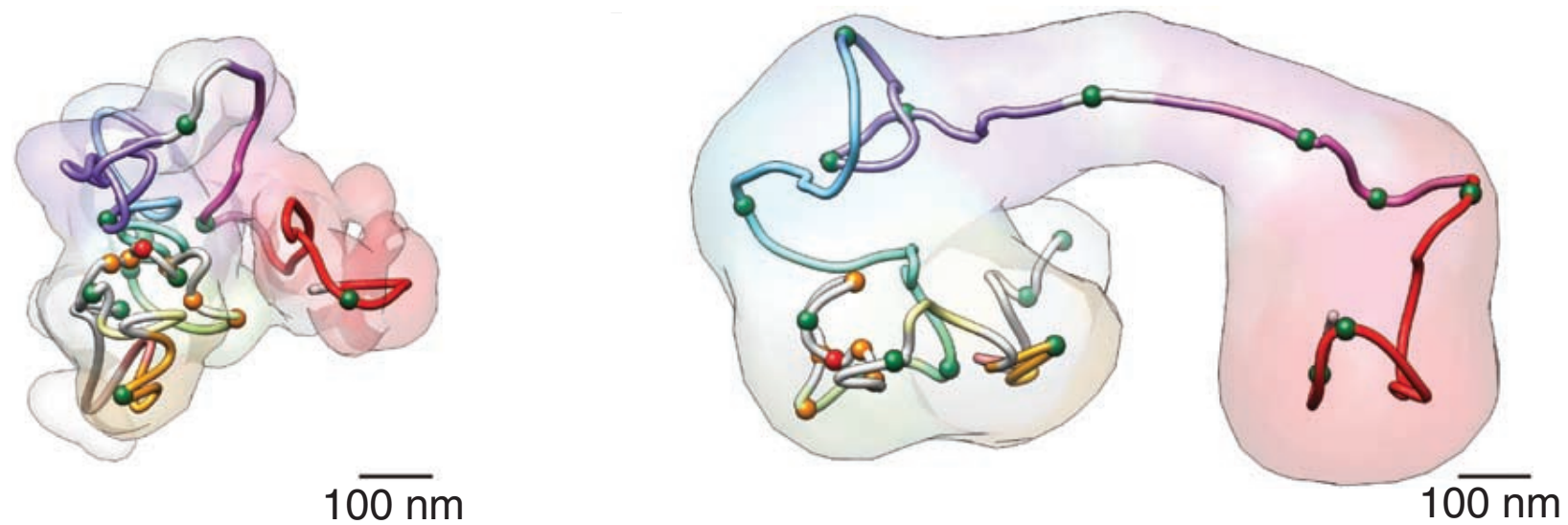


# Long-range interactions

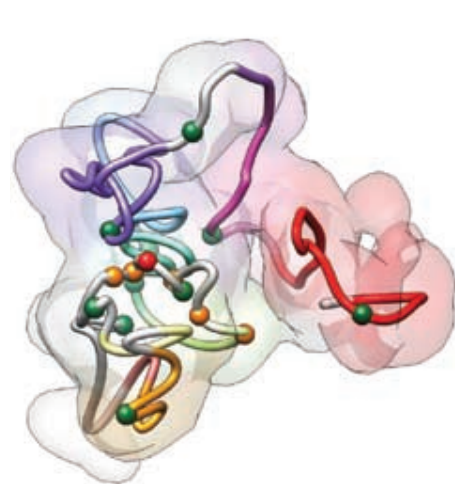


# Chromatin globules

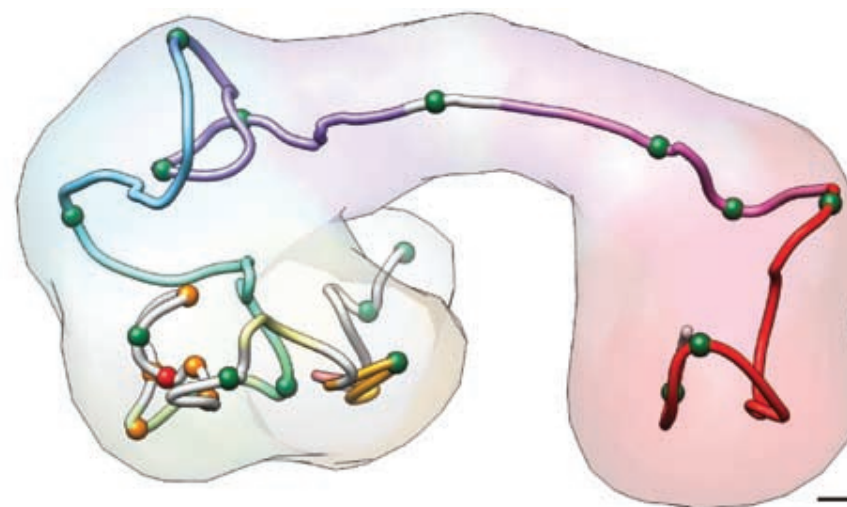
Frequency contact map differences



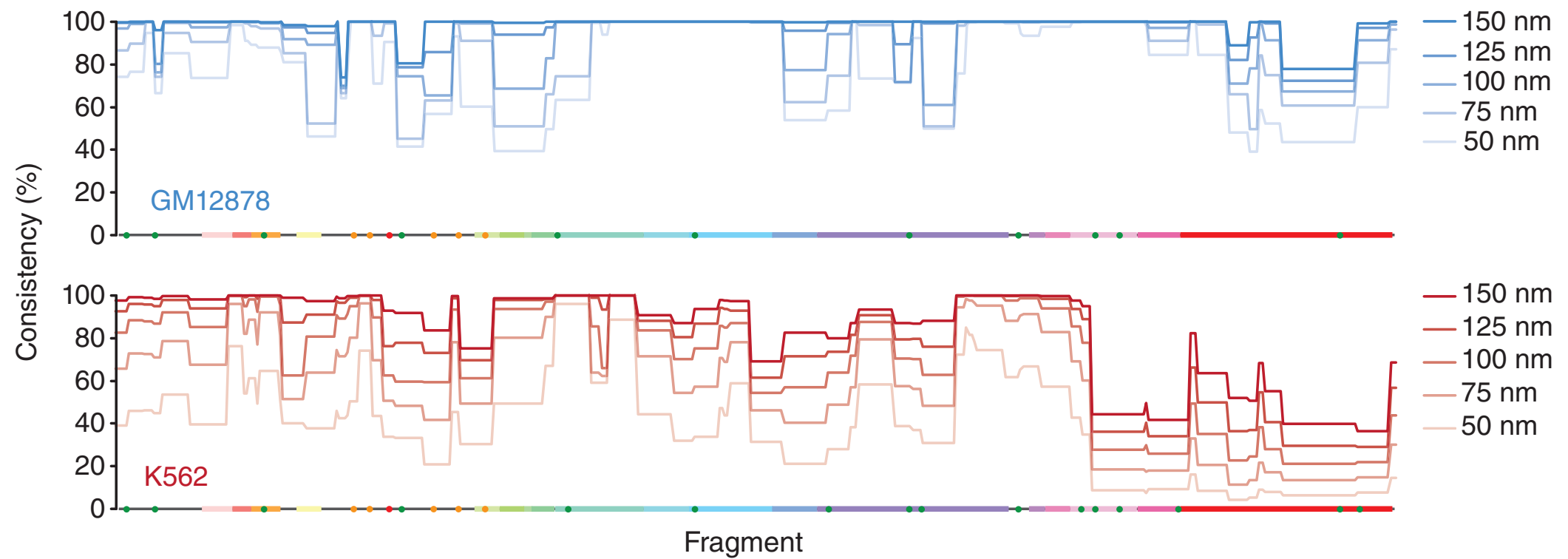
# Consistency



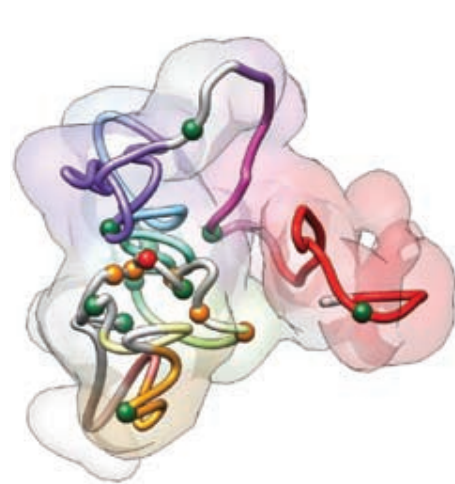
100 nm



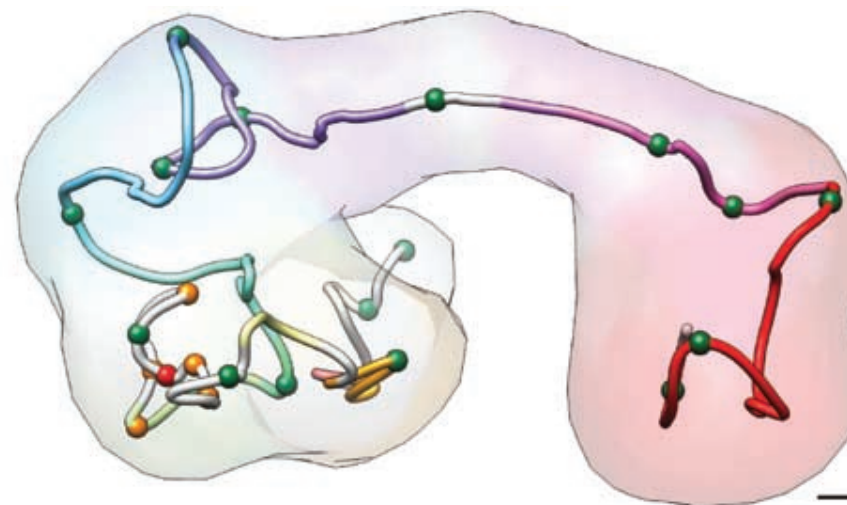
100 nm



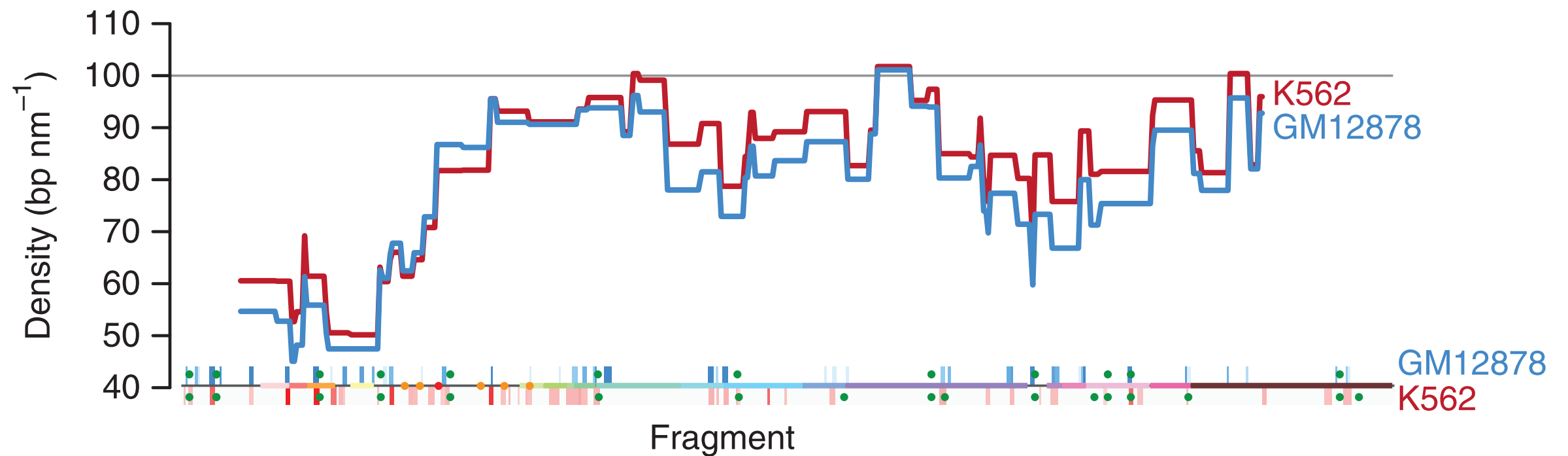
# Chromatin compaction



100 nm

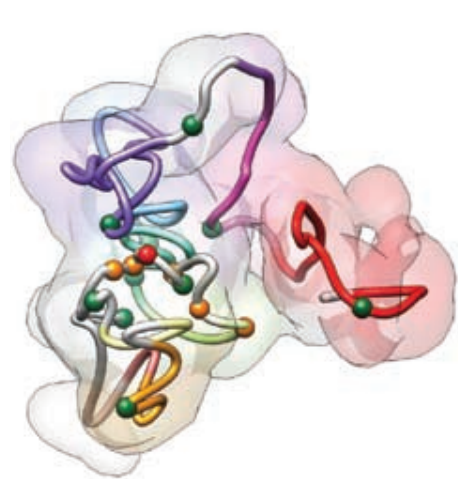


100 nm

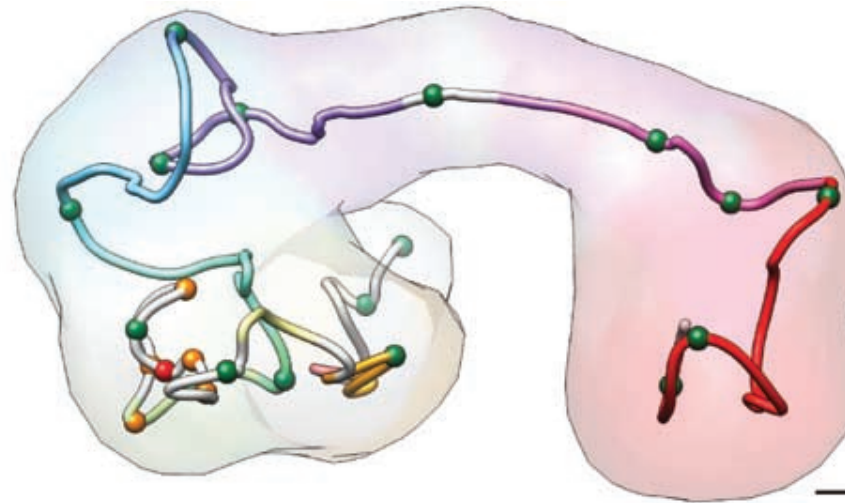




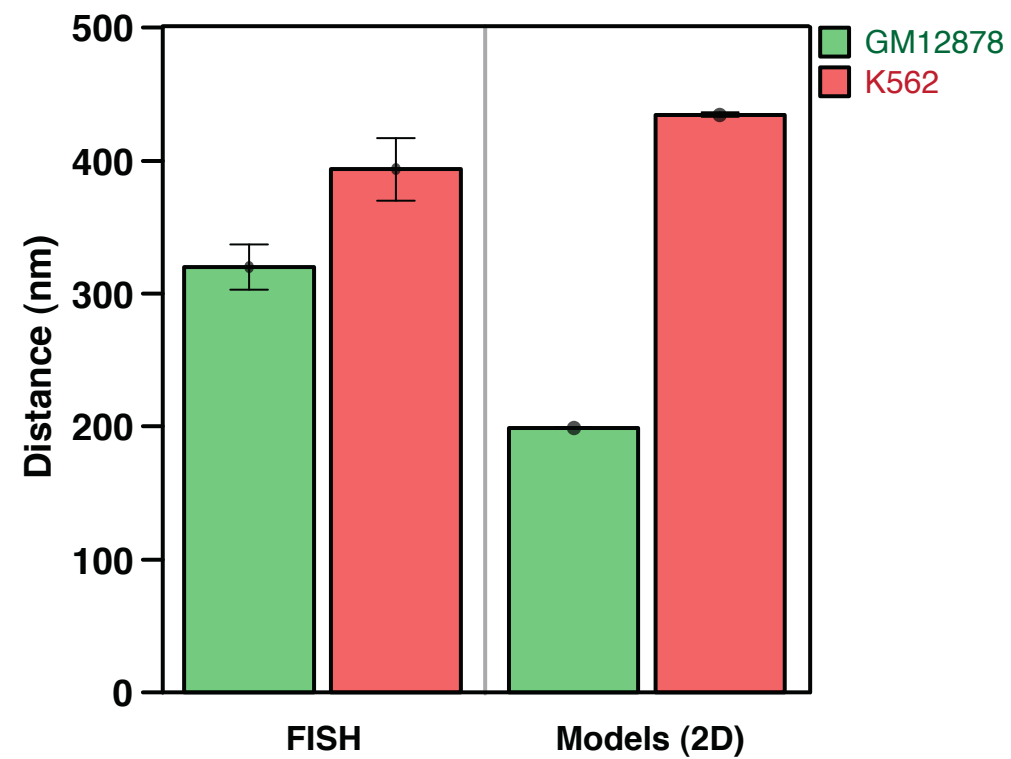
# Model validation



100 nm

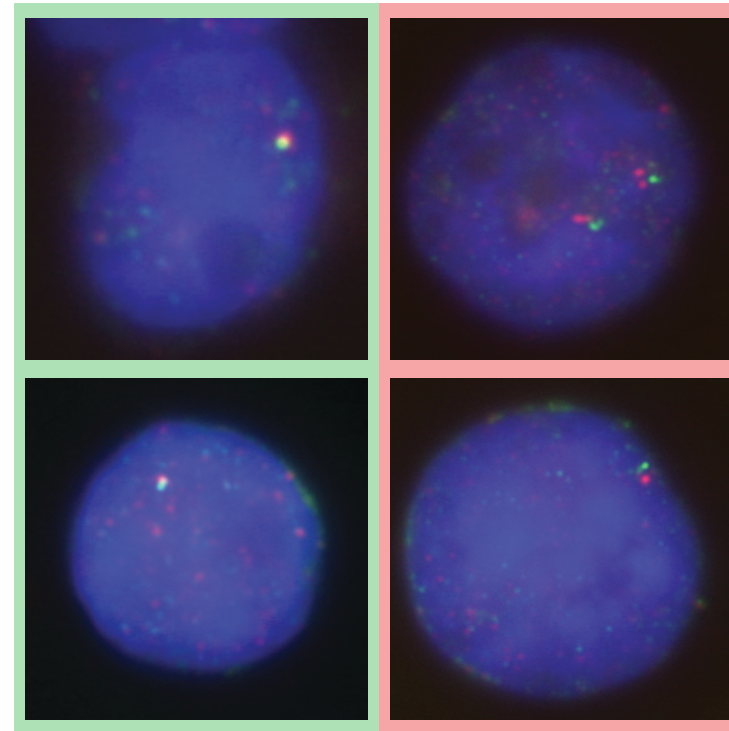


100 nm



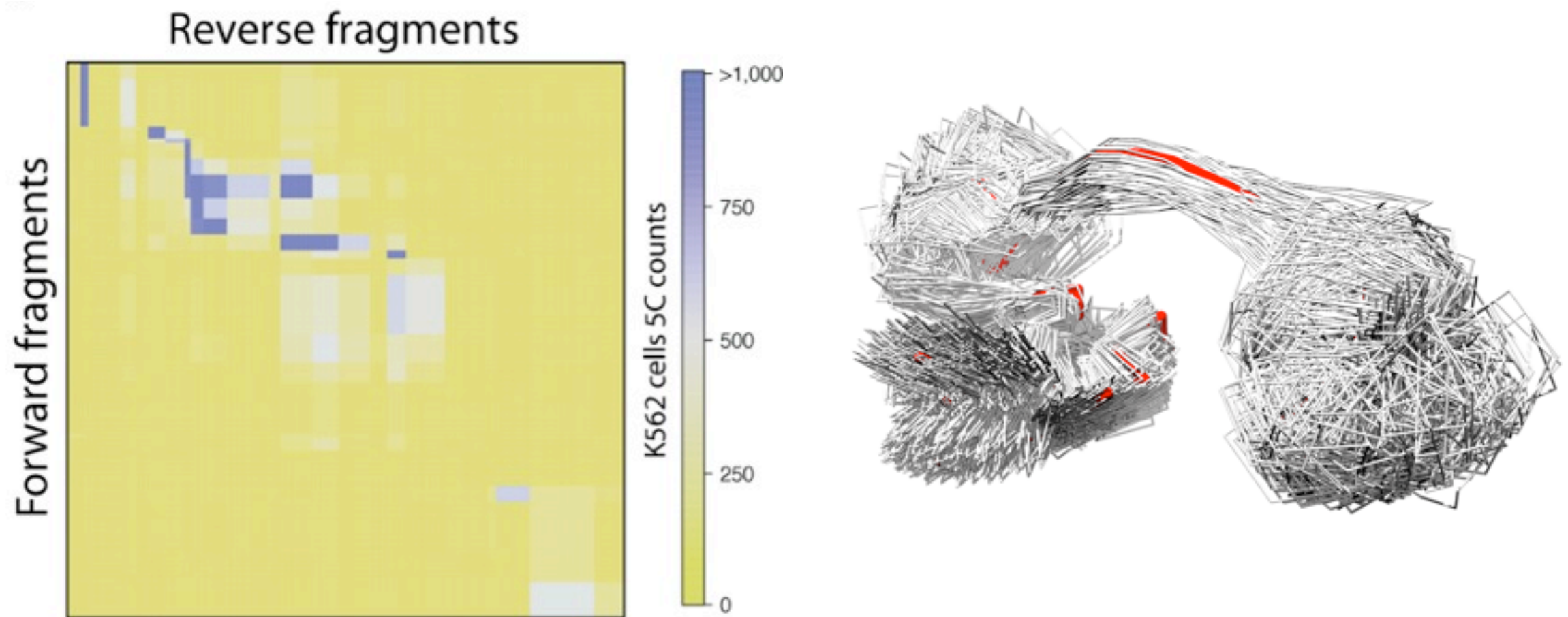
GM12878

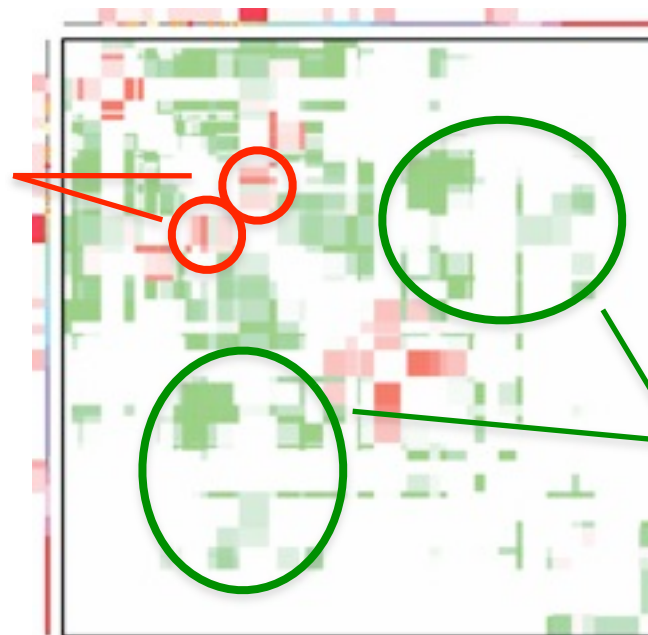
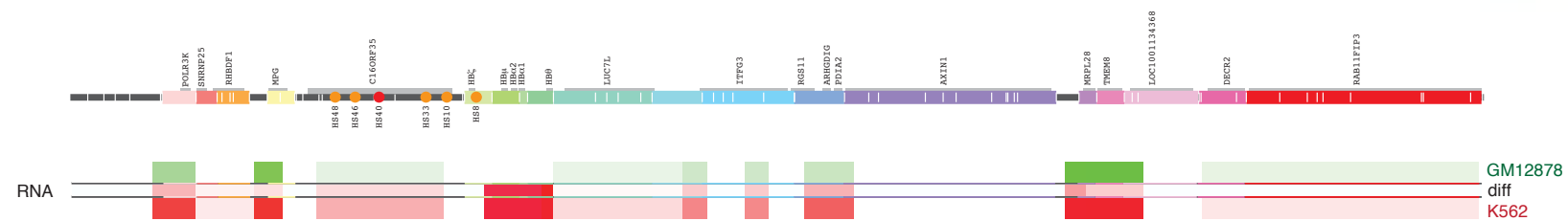
K562



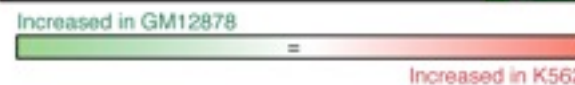
# Summary

5C data results in consistent 3D models





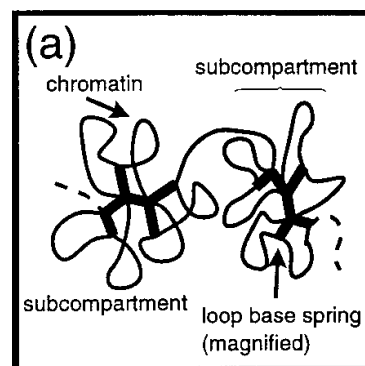
## ▶ Looping interaction In GM12878 cells



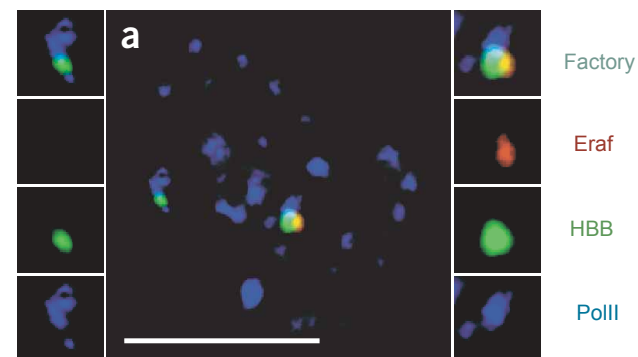
# The “Chromatin Globule” model

D. Baù et al. *Nat Struct Mol Biol* (2011) 18:107-14

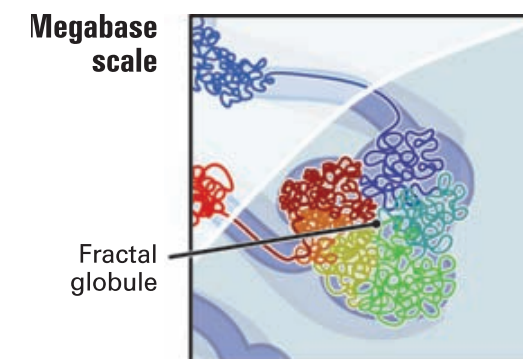
A. Sanyal et al. *Current Opinion in Cell Biology* (2011) 23:325–33.



Münkel et al. *JMB* (1999)



Osborne et al. *Nat Genet* (2004)



Lieberman-Aiden et al. *Science* (2009)



# Caulobacter crescentus 3D genome

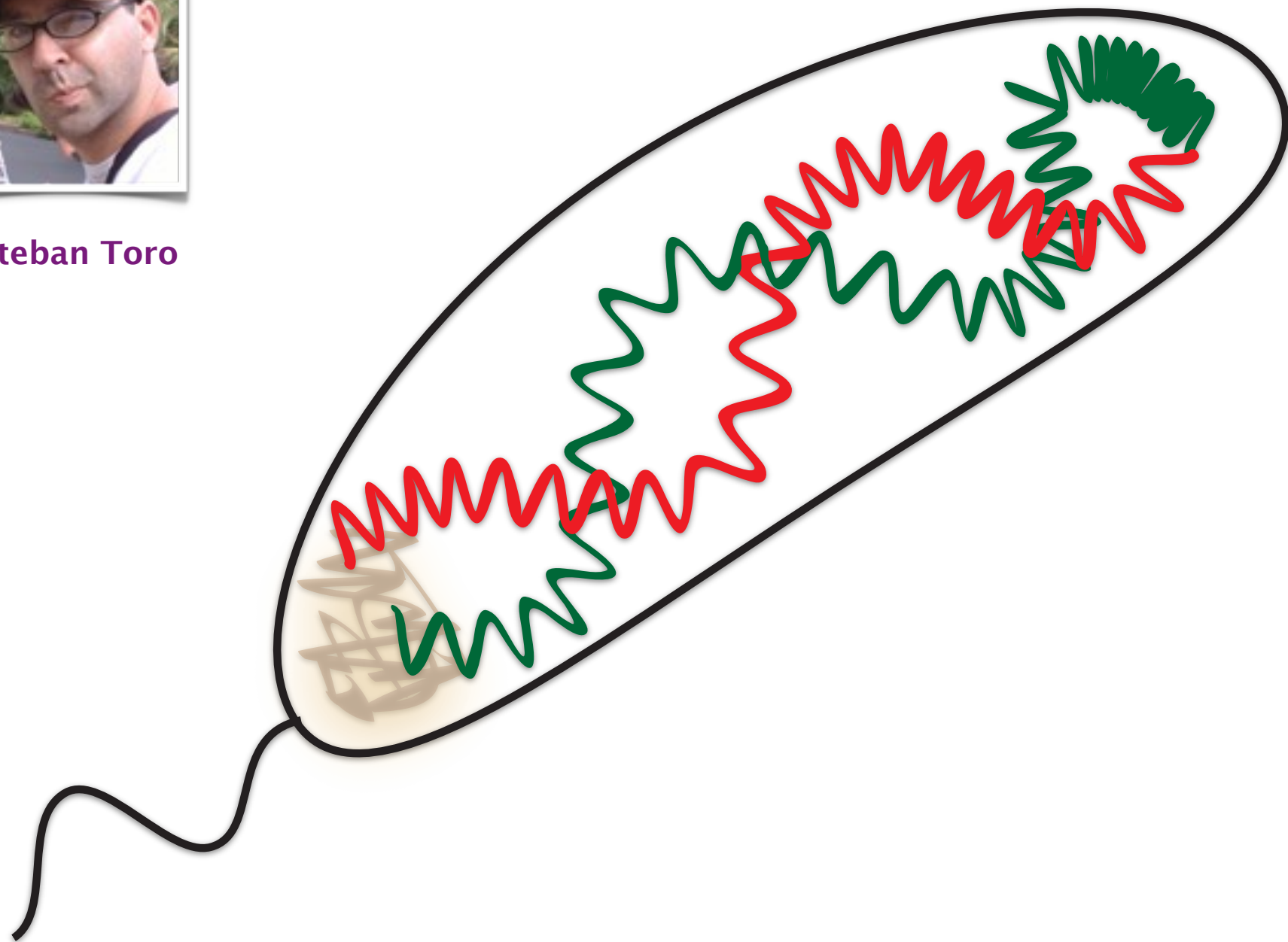
M.A. Umbarger, et al. Molecular Cell (2011) 44:252–264



Mark Umbarger



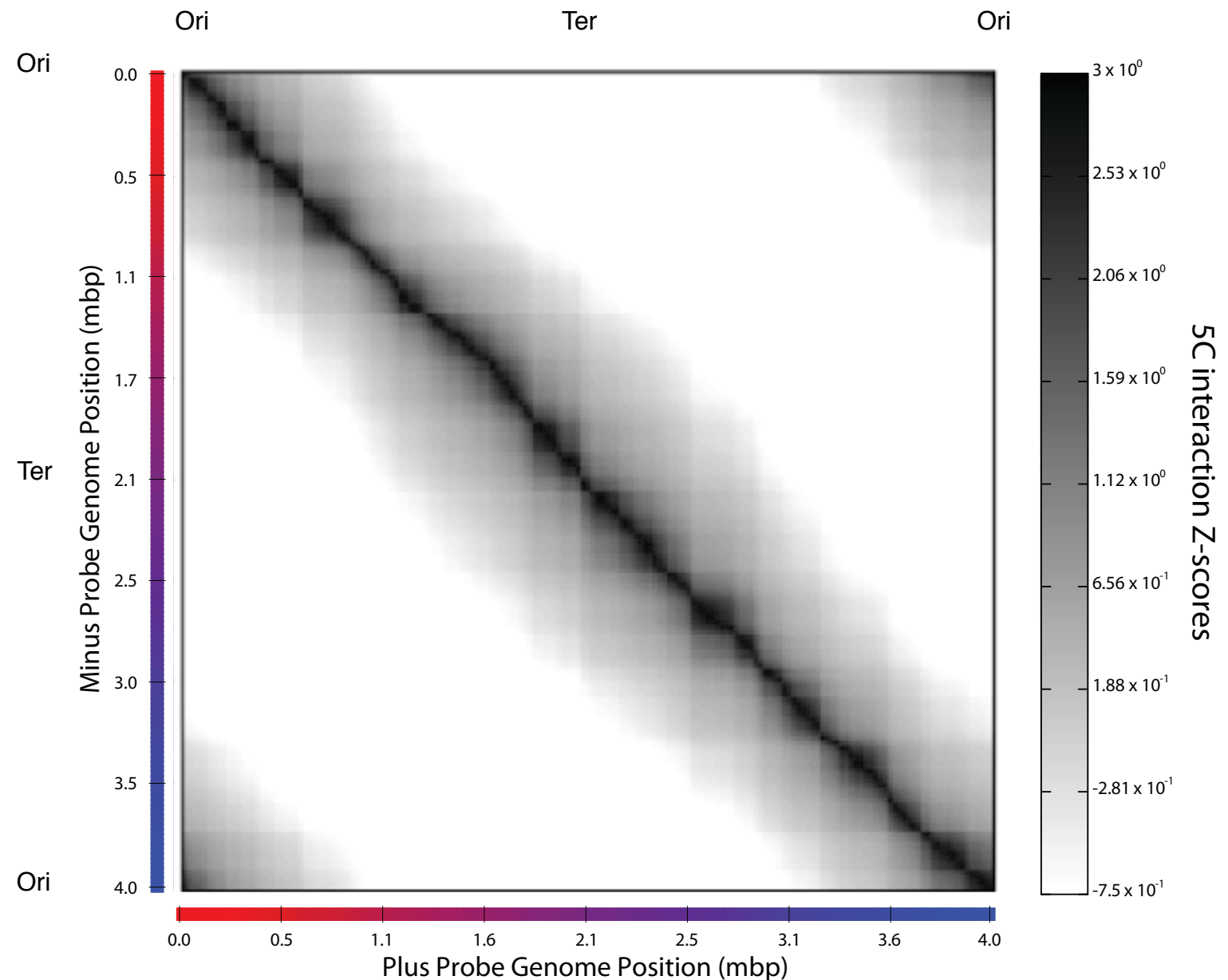
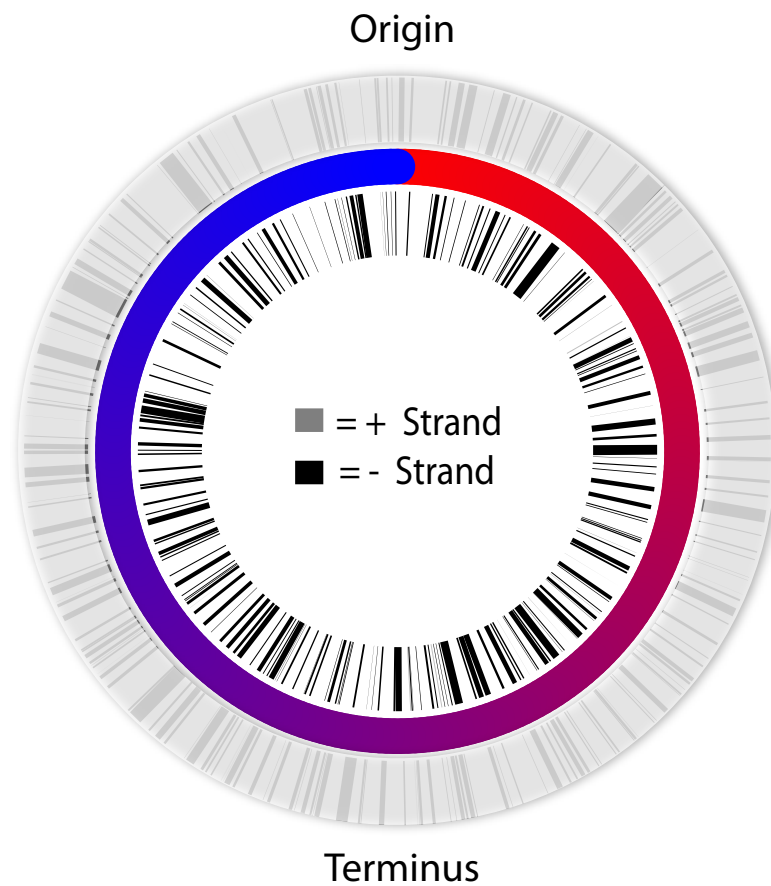
Esteban Toro





# The 3D architecture of *Caulobacter Crescentus*

4,016,942 bp & 3,767 genes

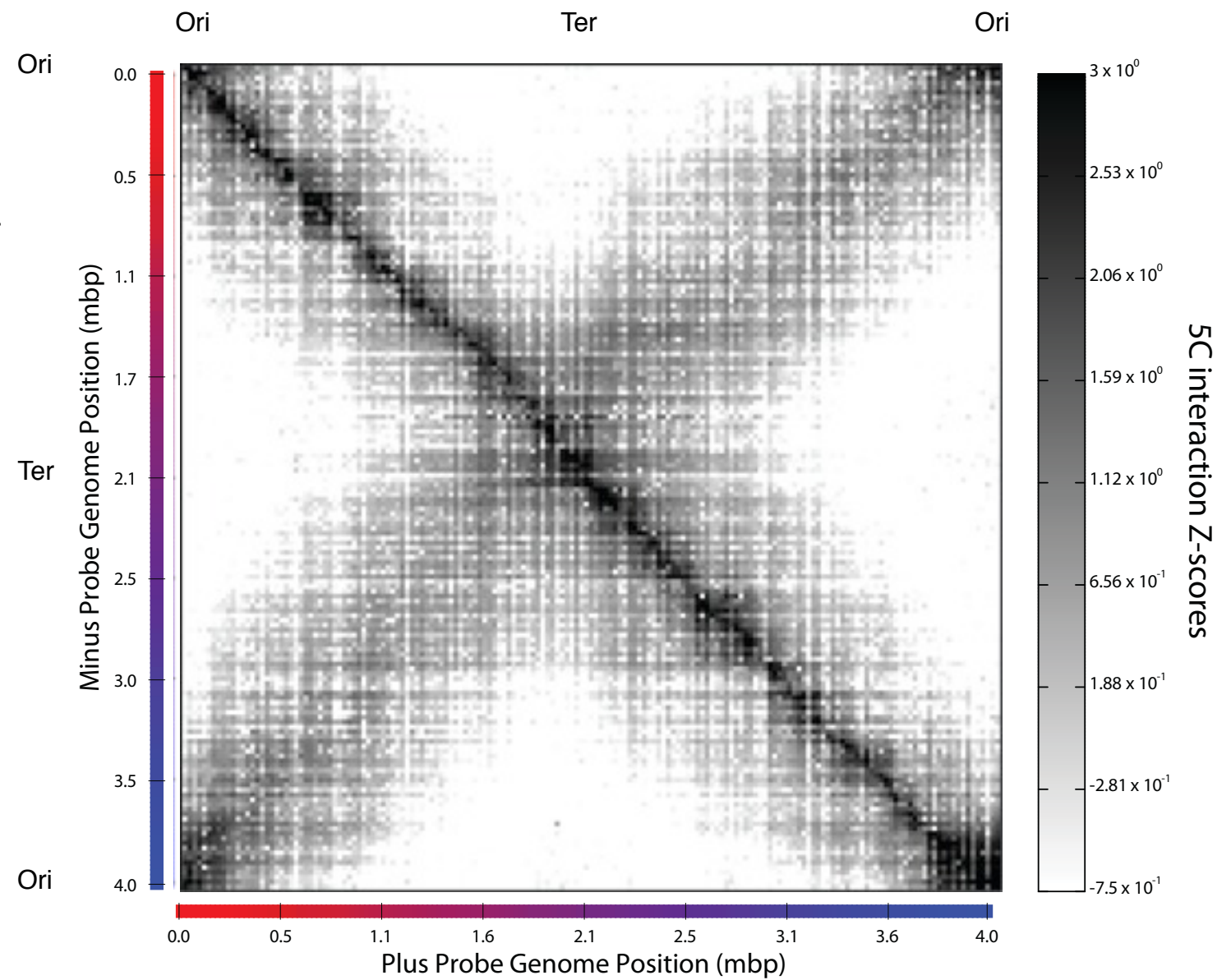
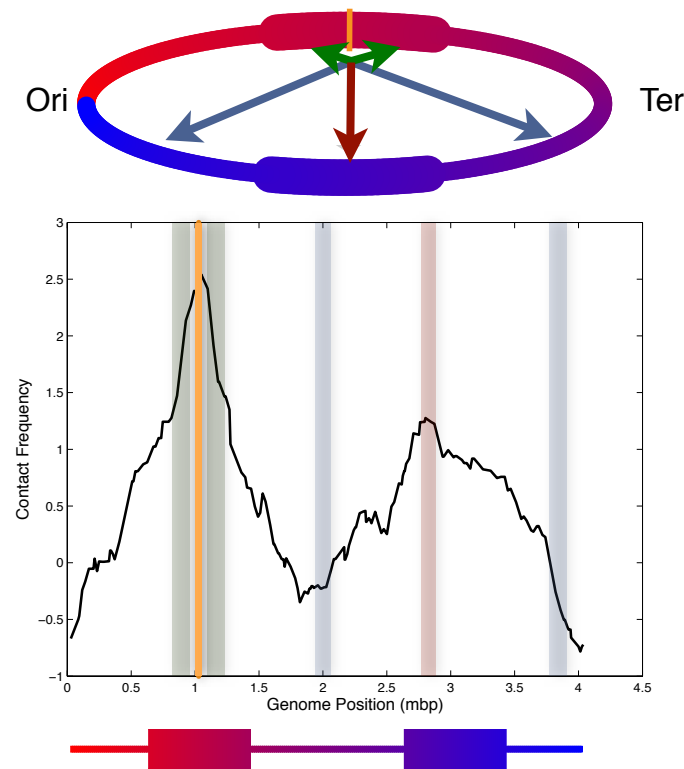
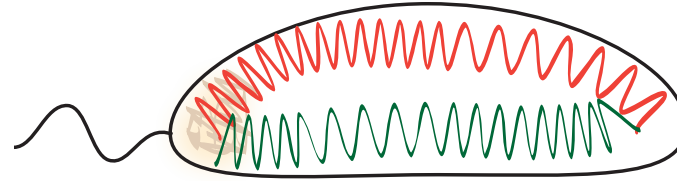


169 5C primers on + strand  
170 5C primers on - strand  
**28,730 chromatin interactions**

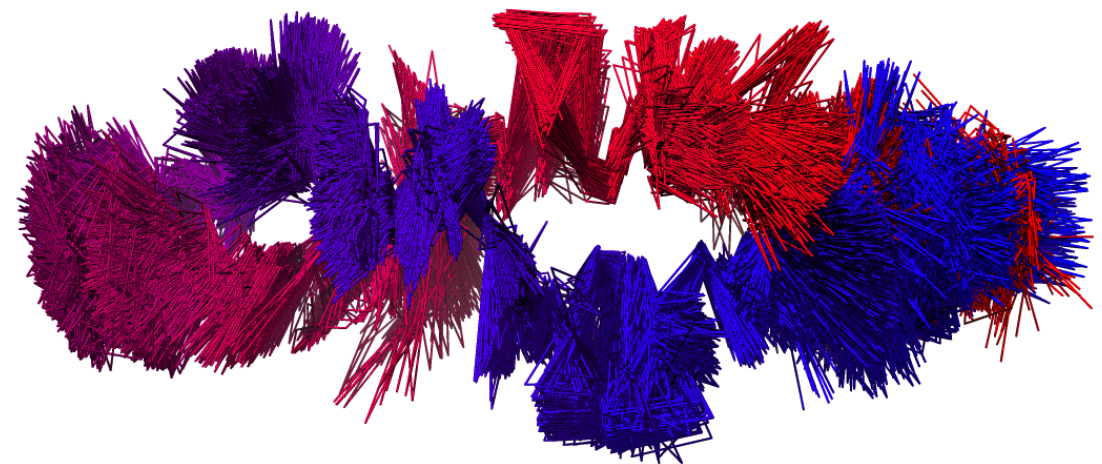
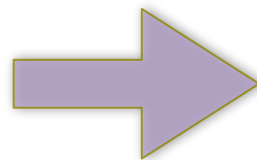
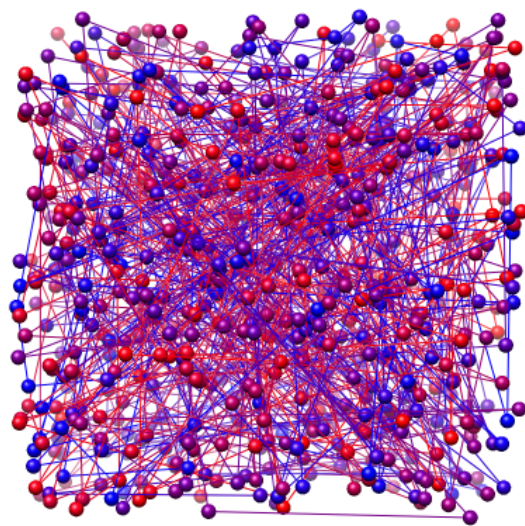
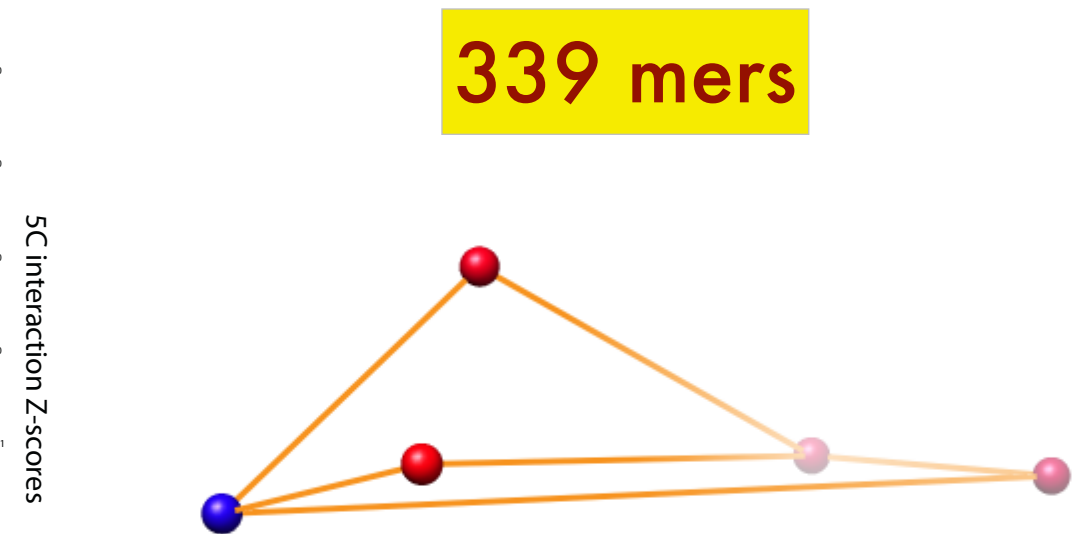
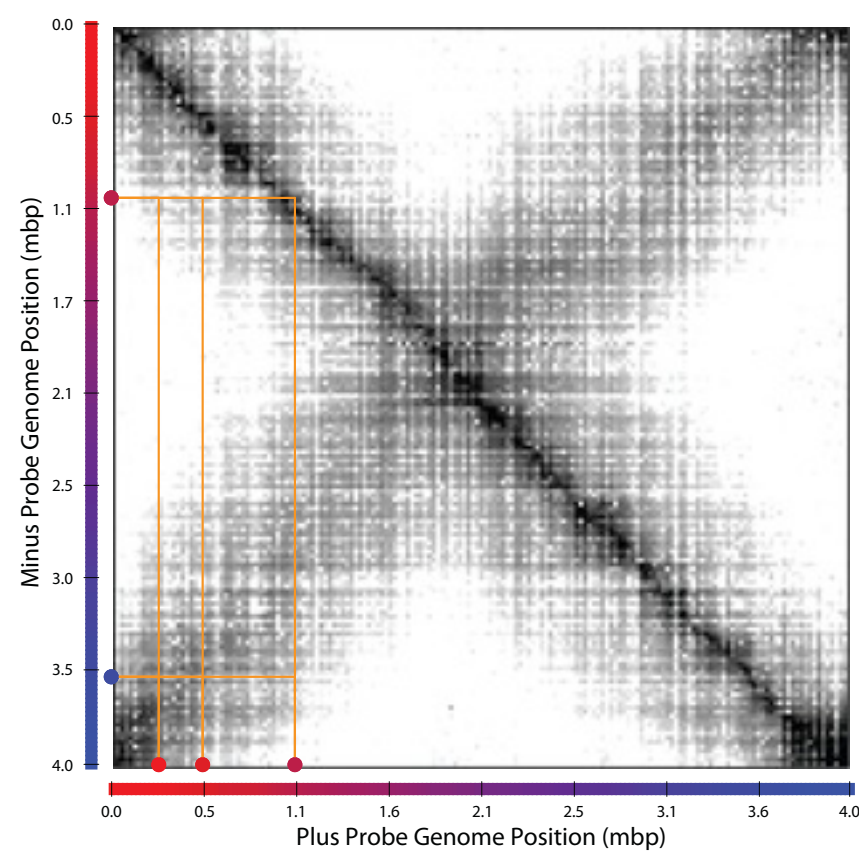
**~13Kb**

# 5C interaction matrix

ELLIPSOID for *Caulobacter crescentus*



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



# Genome organization in *Caulobacter crescentus*

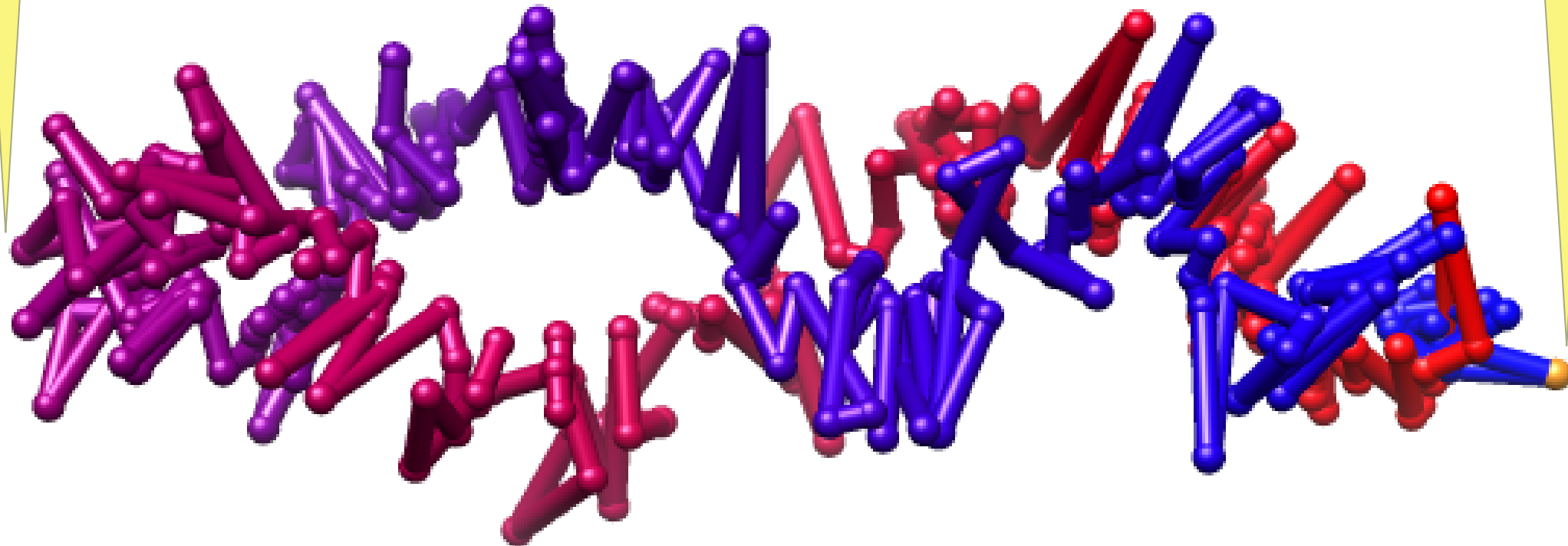
Arms are helical

*dif* site  $47 \pm 17$  Kb from Ter

*parS* sites  $25 \pm 17$  Kb from Ori

Resolution

Centromer-like



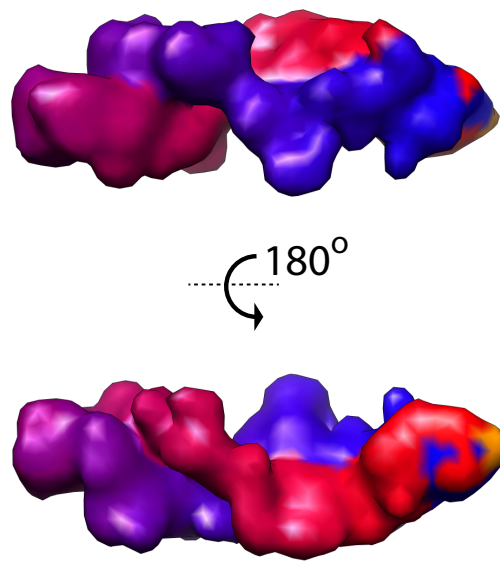


# Genome organization in *Caulobacter crescentus*

Arms are helical



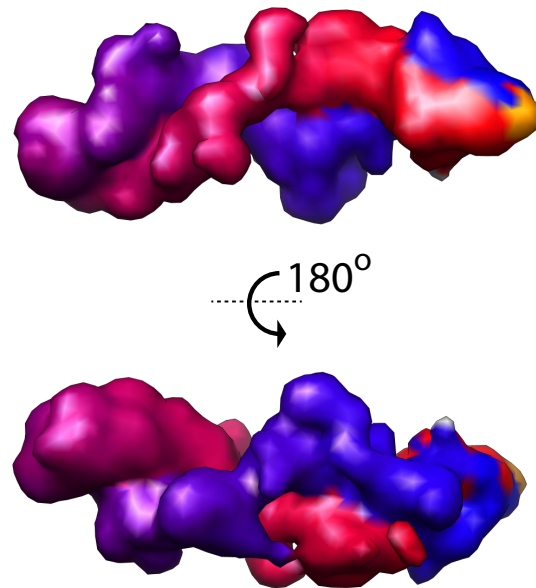
Cluster 1



180°

500 nm

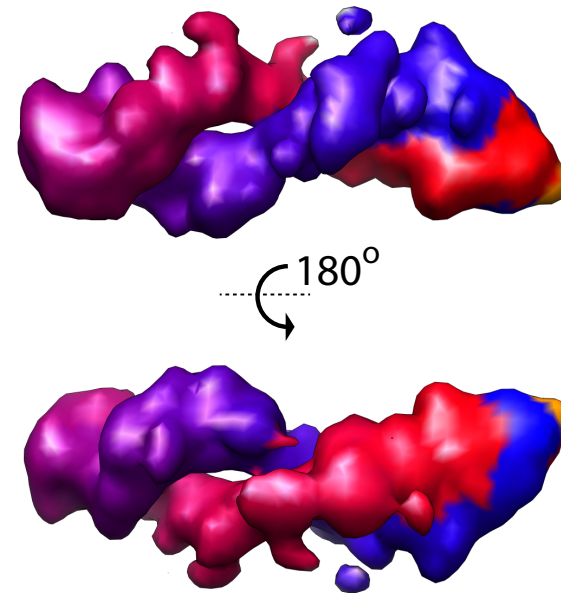
Cluster 2



180°

500 nm

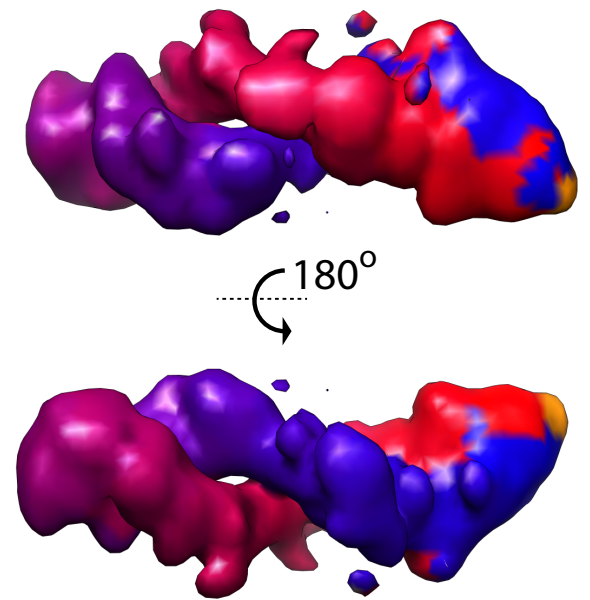
Cluster 3



180°

500 nm

Cluster 4



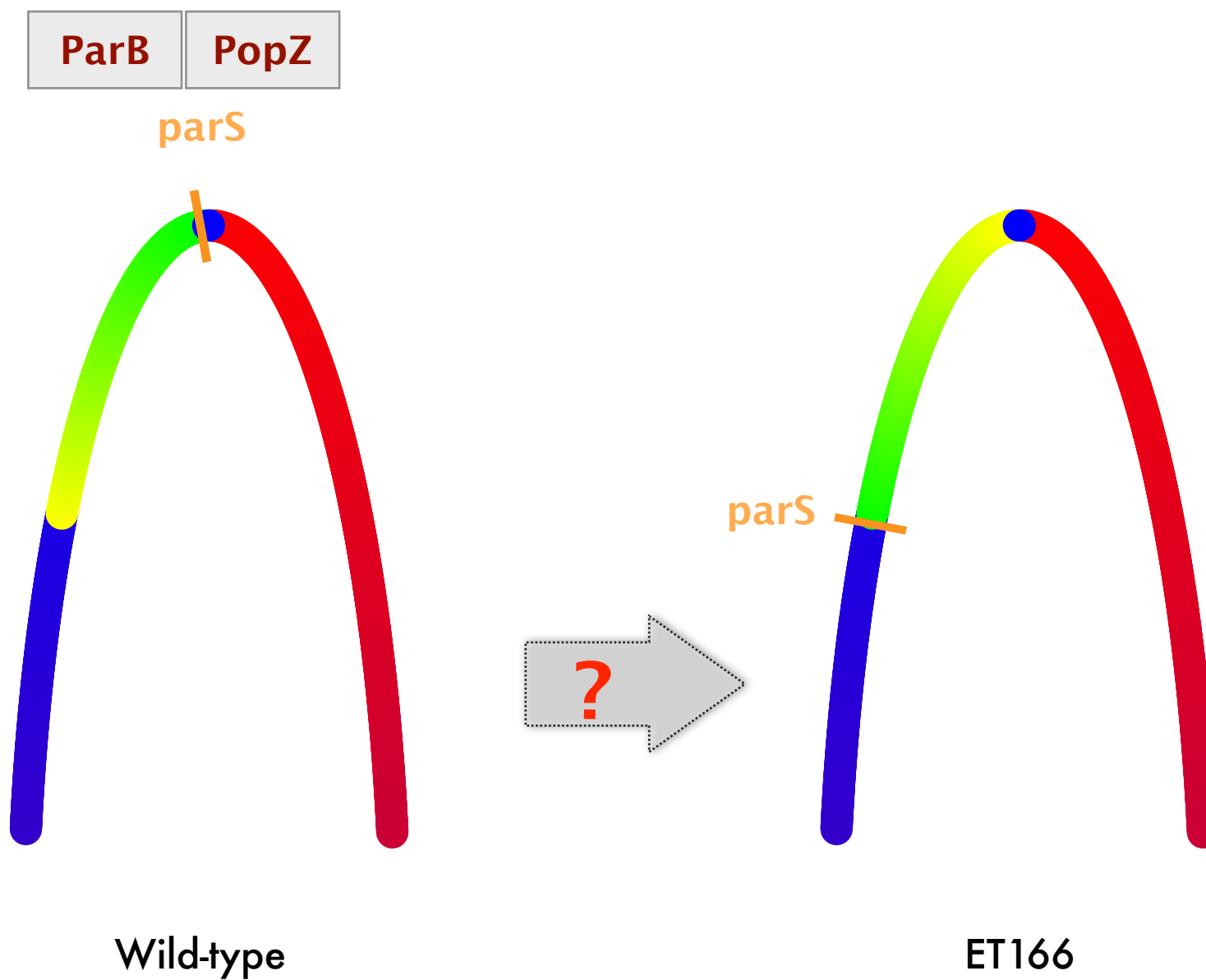
180°

500 nm

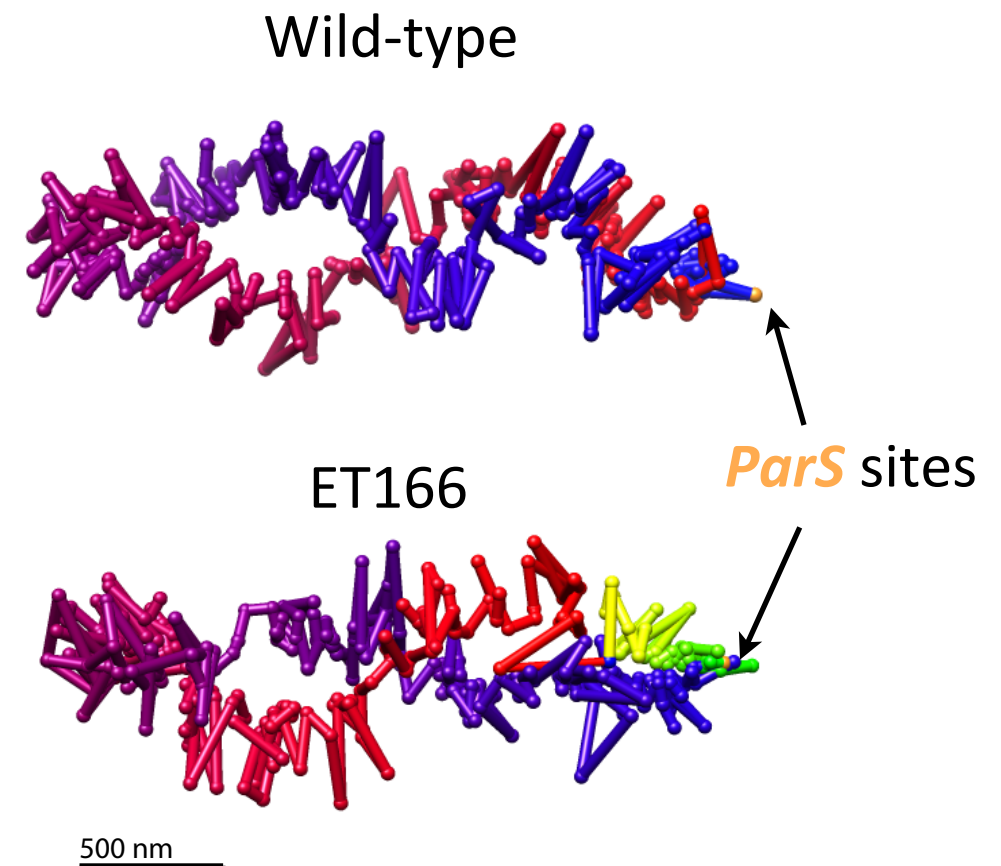
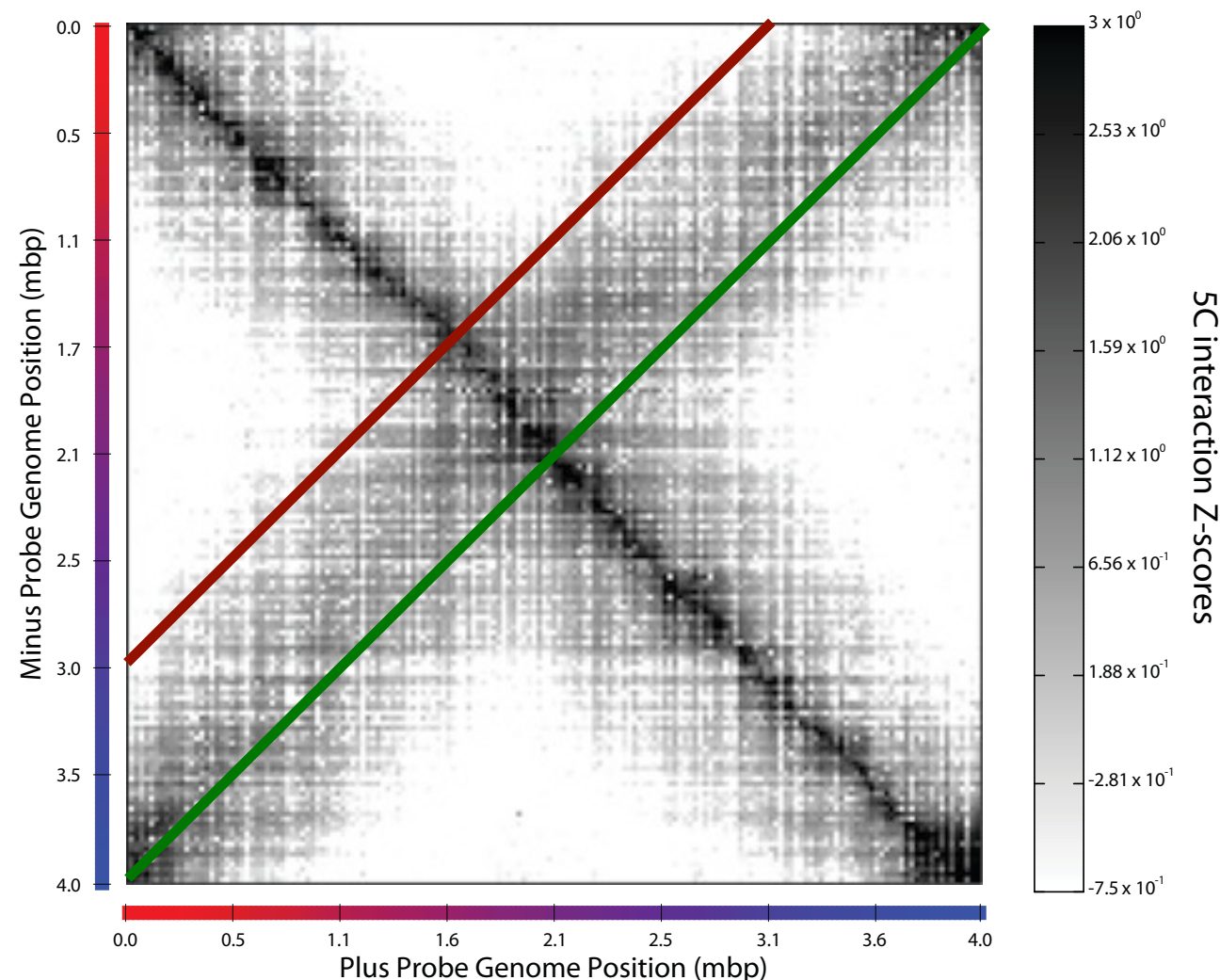
**MIRRORS!**



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



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

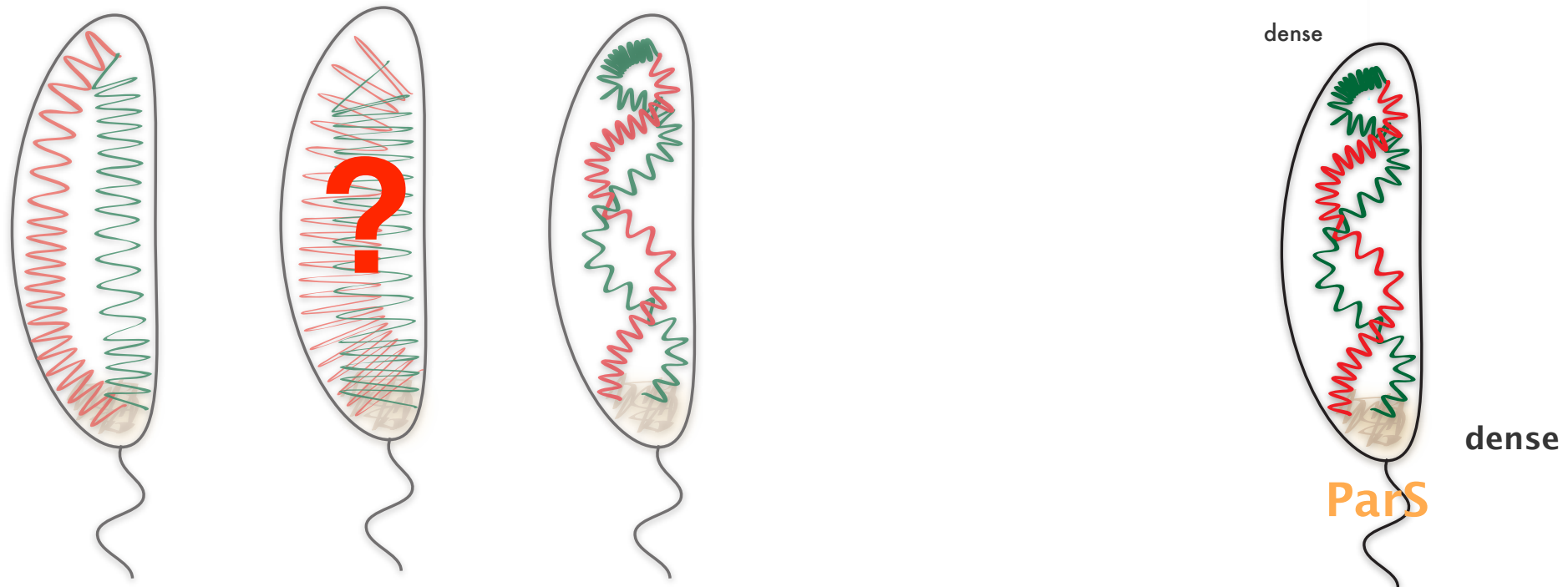
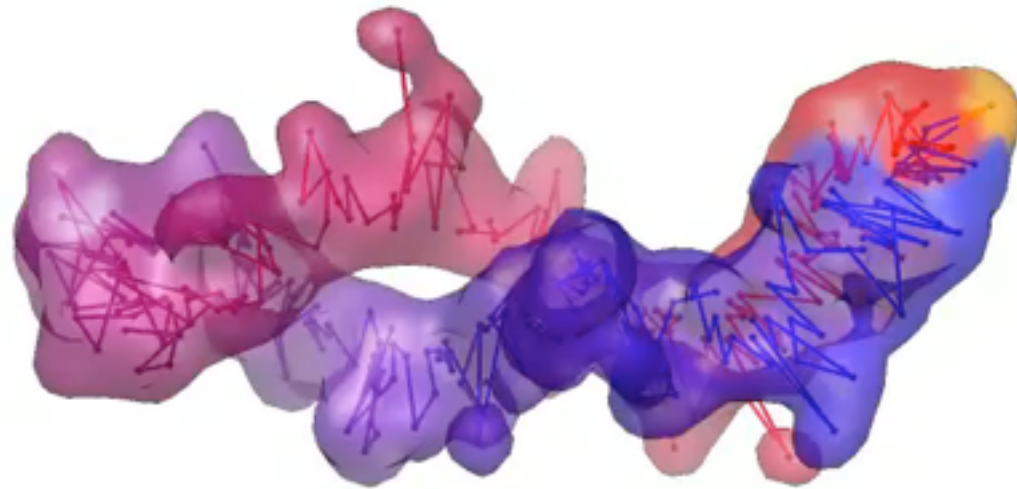


Arms are **STILL** helical

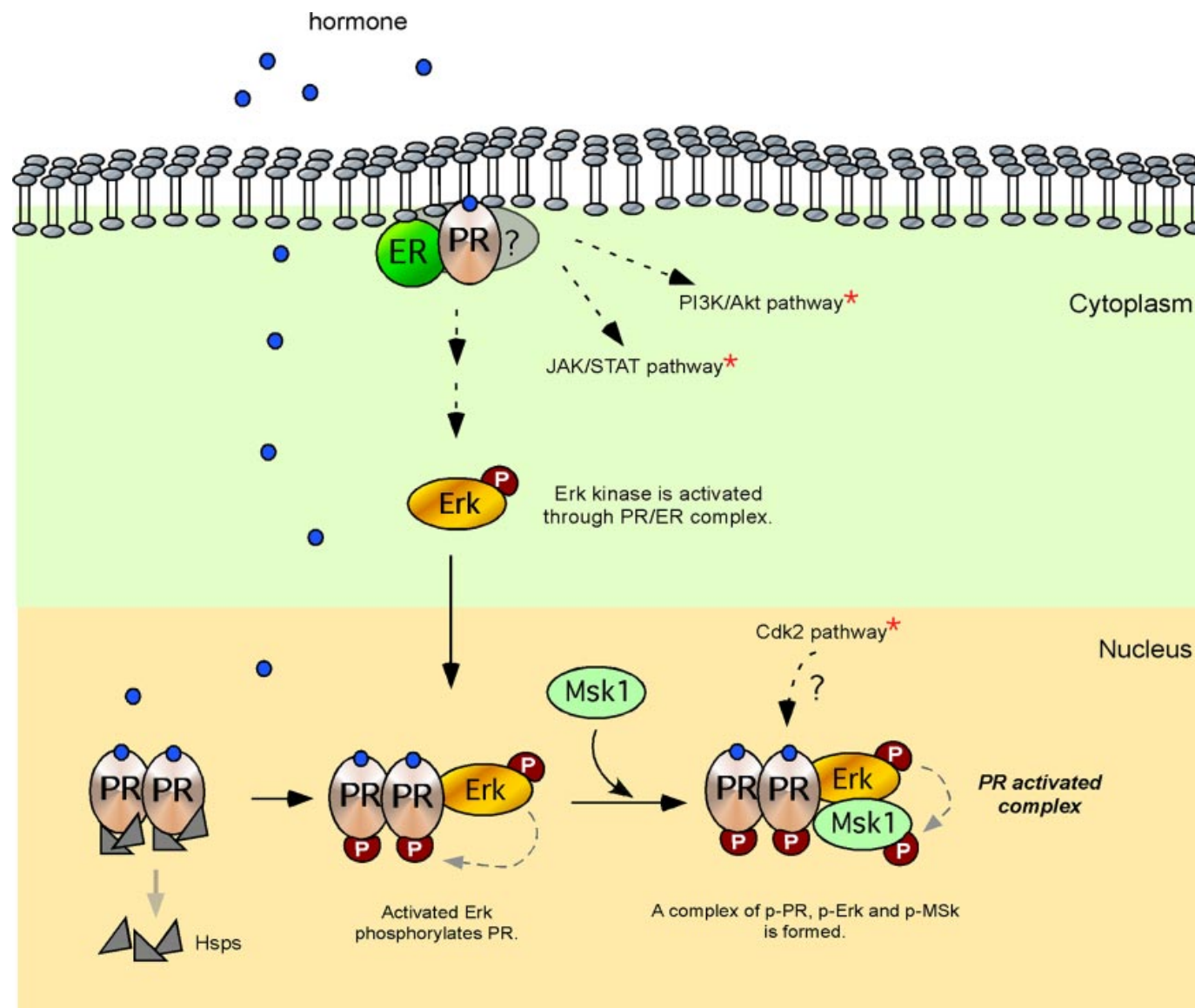
**Structure & function PRESERVED!!!**

# Genome architecture in *Caulobacter*

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



# Progesterone-regulated transcription in breast cancer



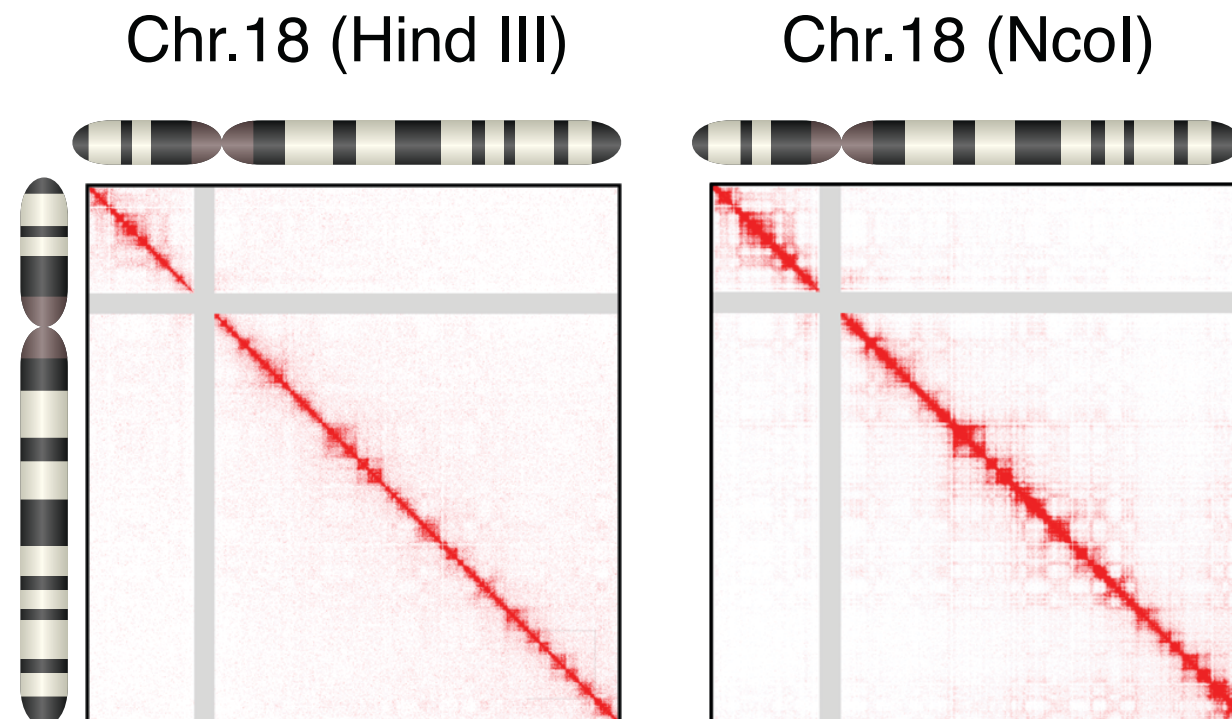
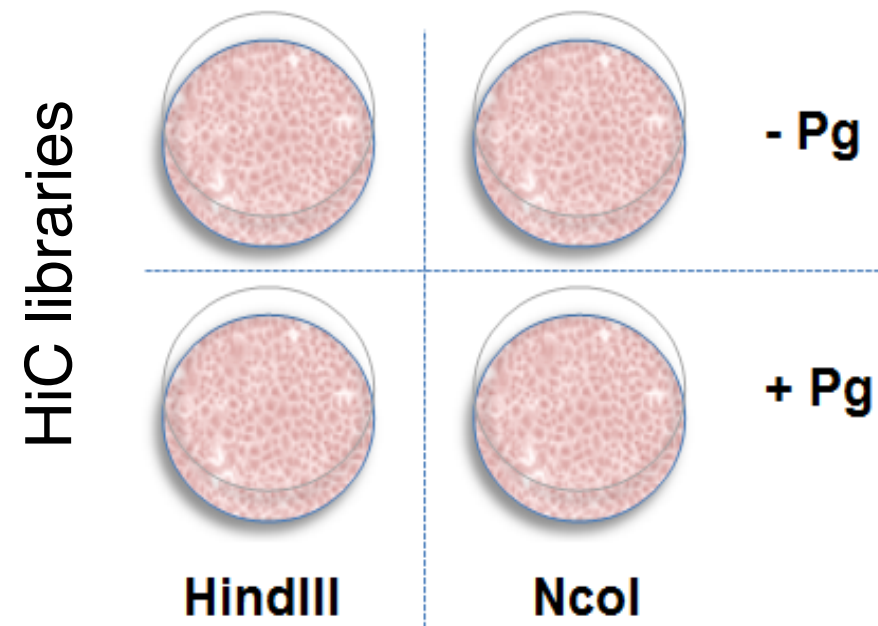
> 2,000 genes **Up**-regulated  
> 2,000 genes **Down**-regulated

**Regulation in 3D?**

Vicent *et al* 2011, Wright *et al* 2012, Ballare *et al* 2012



# Hi-C experimental design





# Scoring

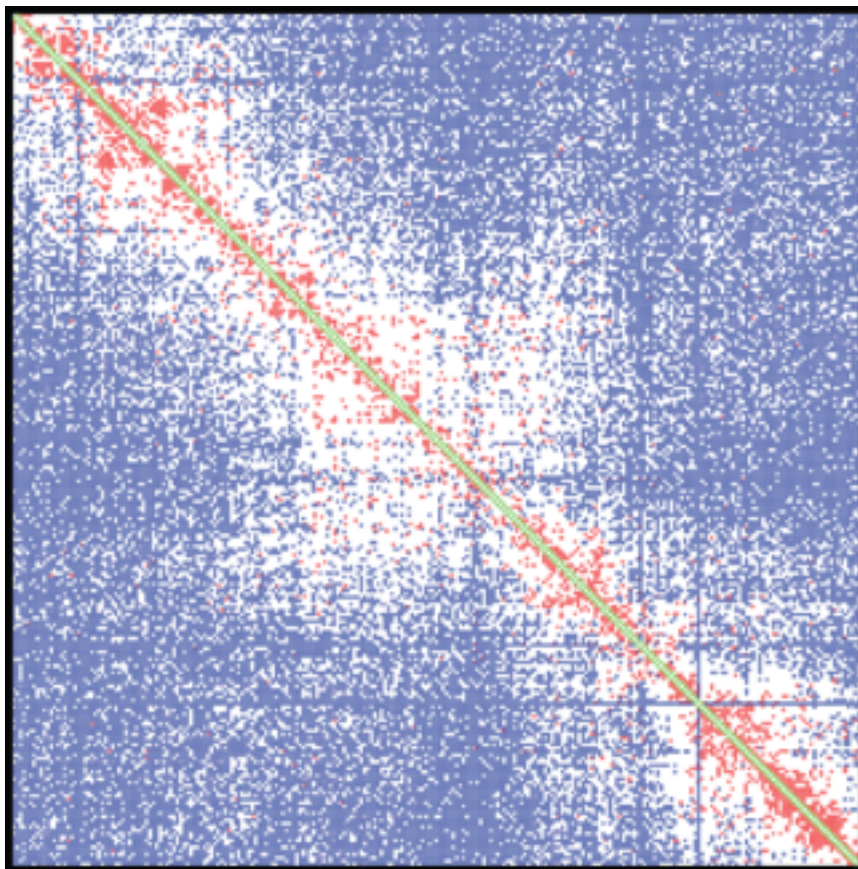
Individual spatial restraints encoding the data

R1 (8,480,000 - 13,460,000)

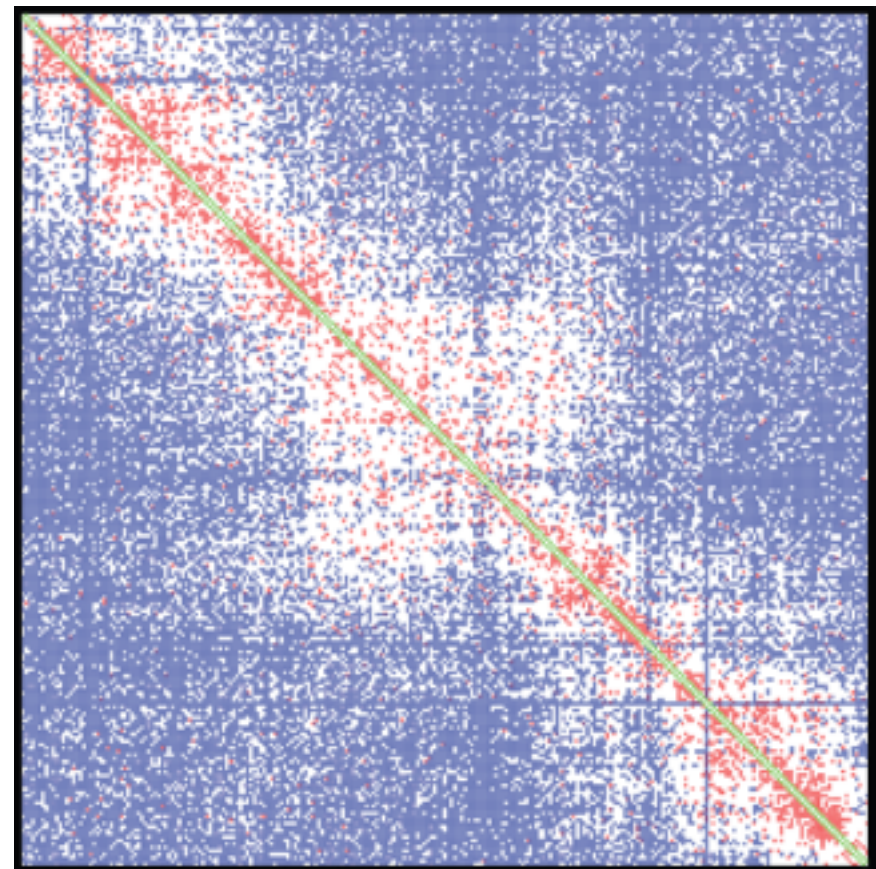


Harmonic · Harmonic Lower Bound · Harmonic Upper Bound

T0

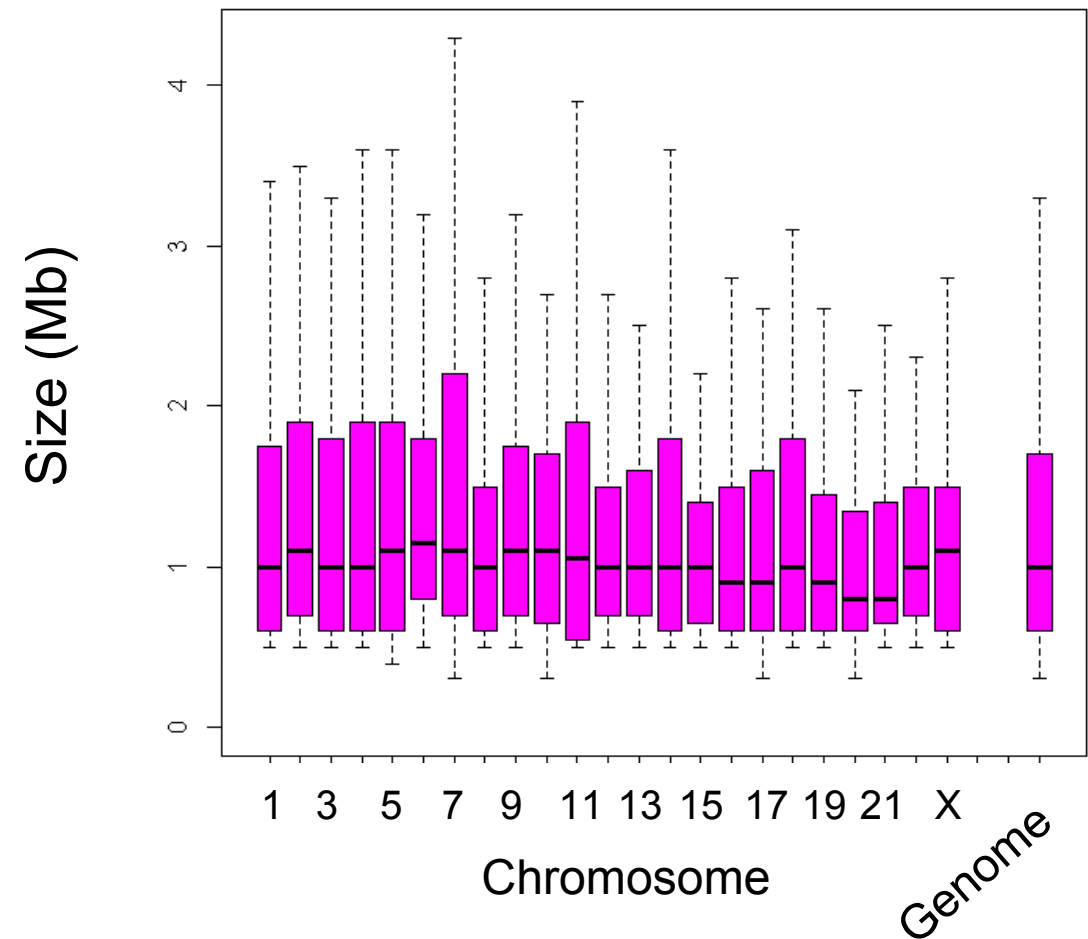
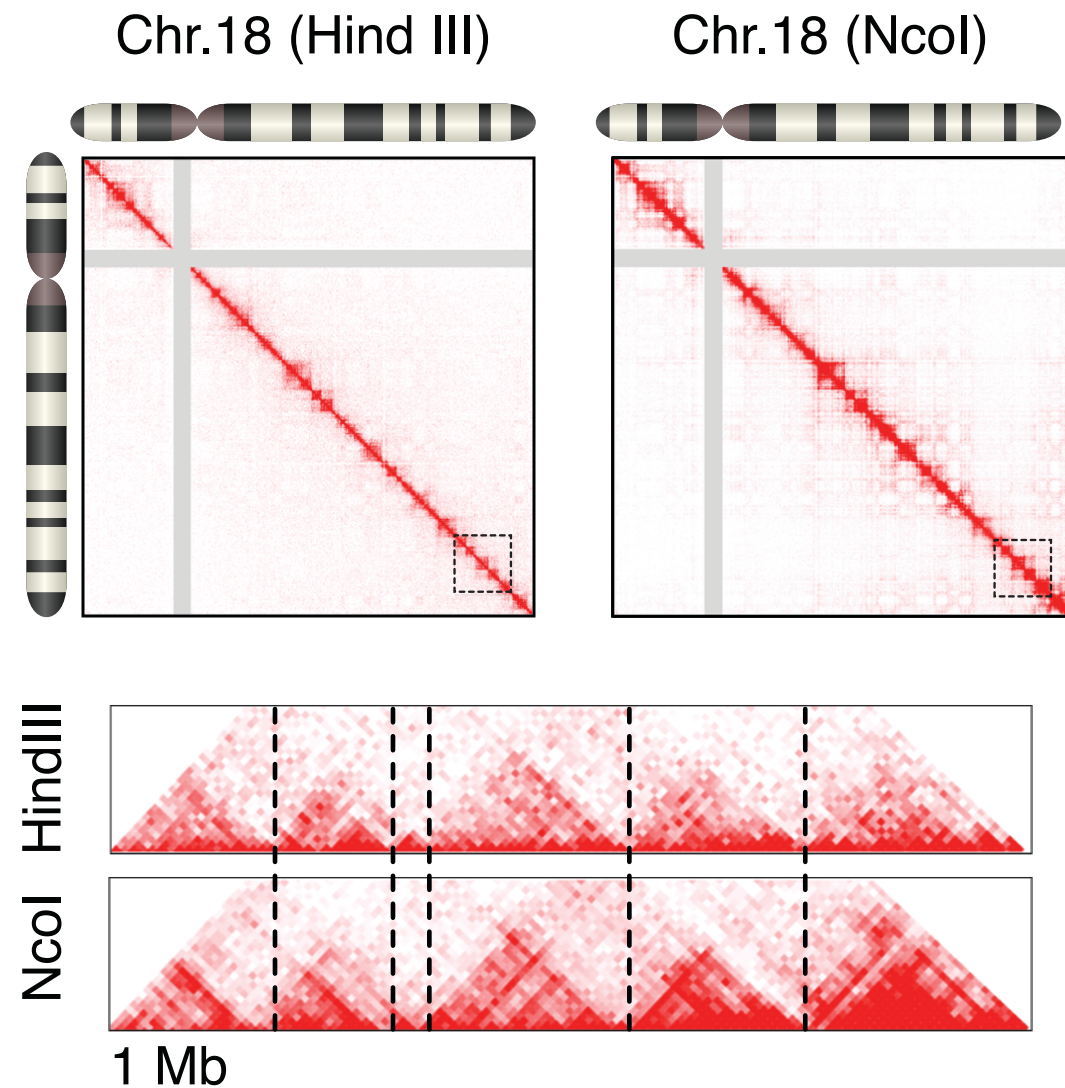


T60





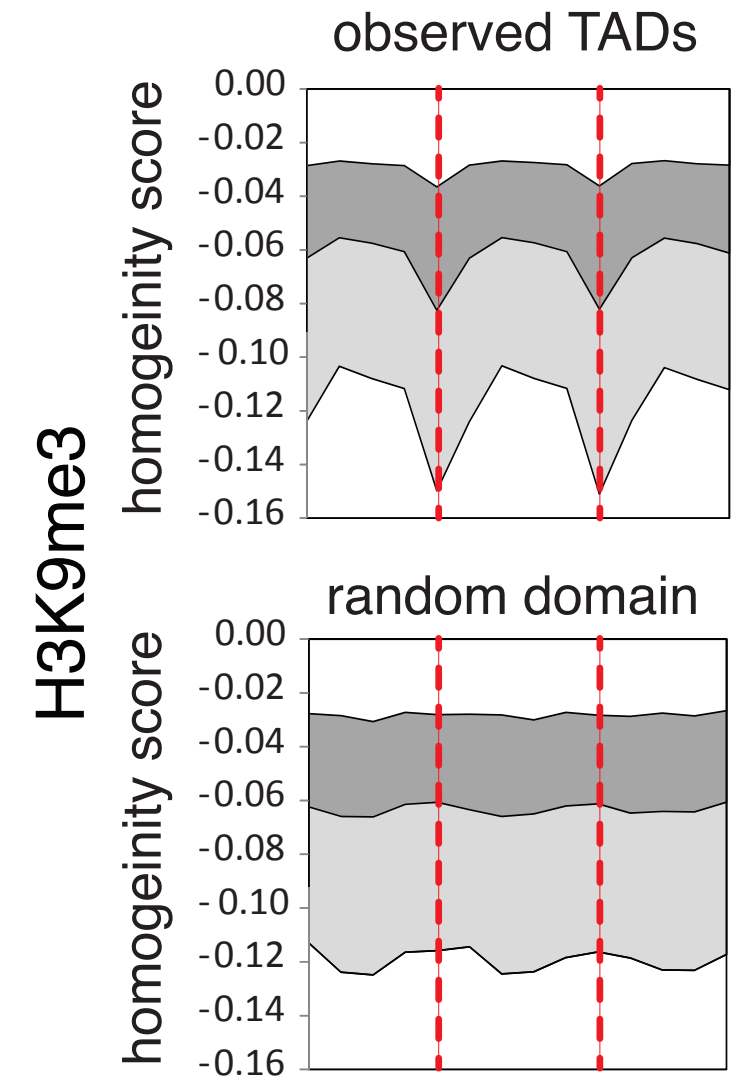
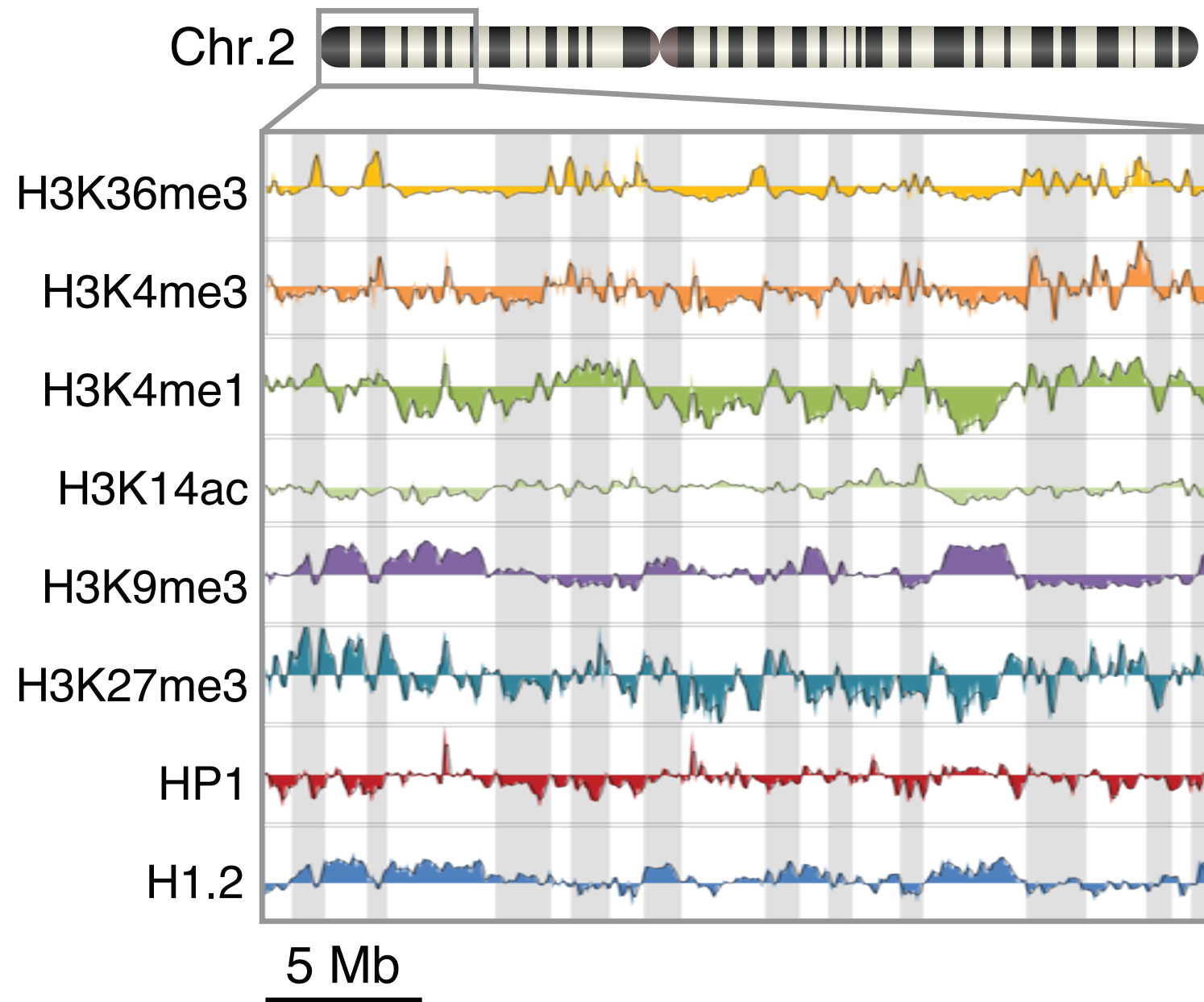
# Chromosomes are organized into Topologically Associated Domains (TADs)





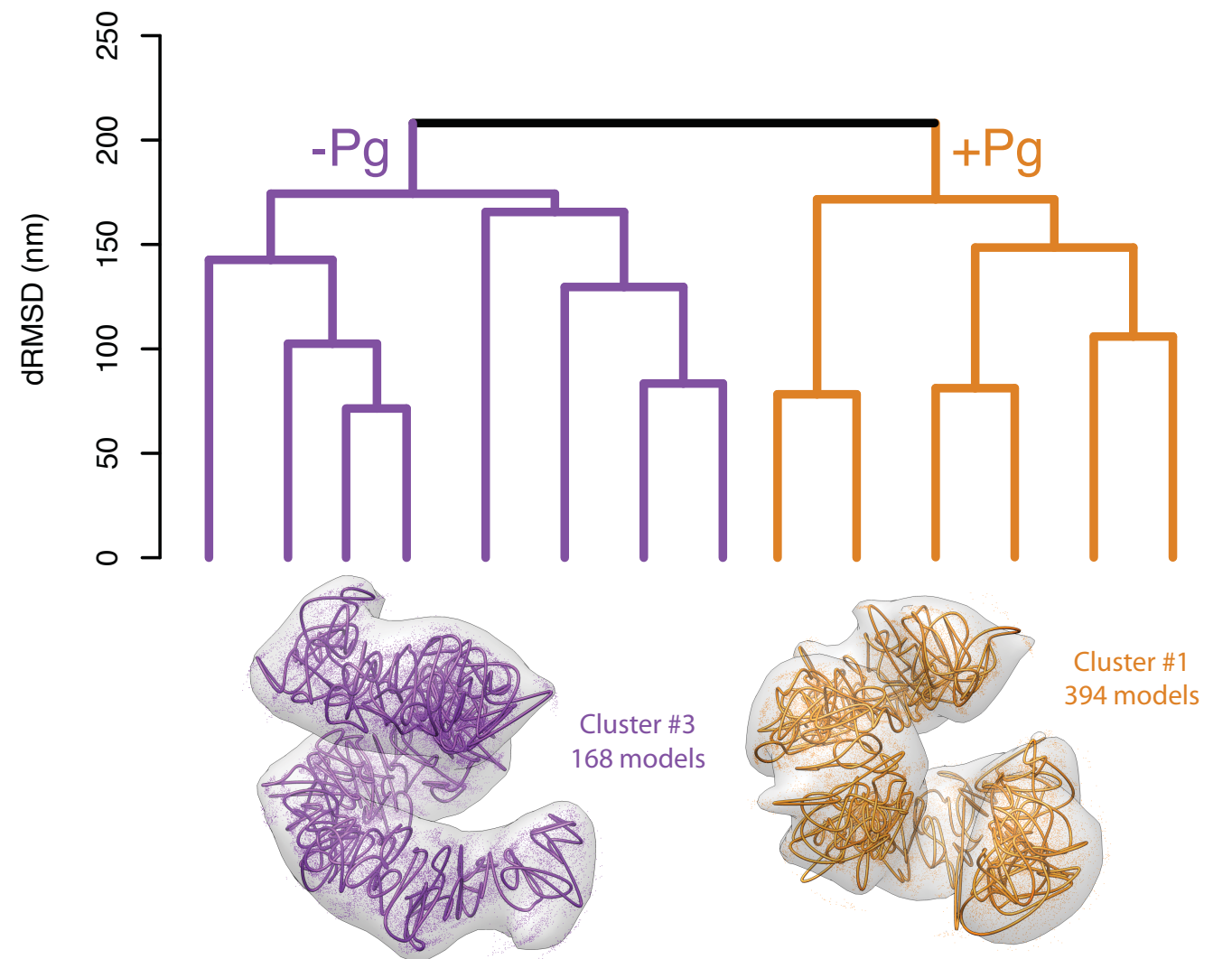
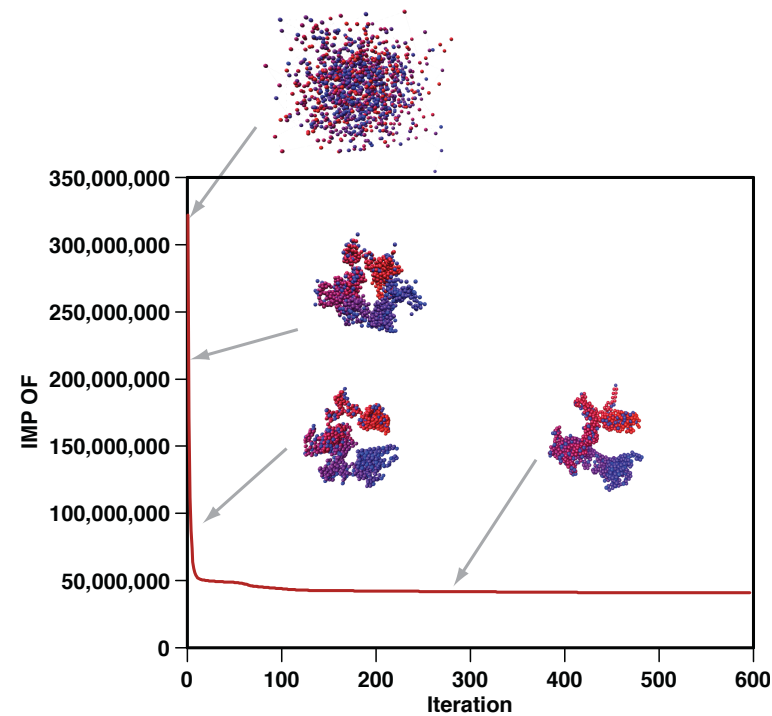
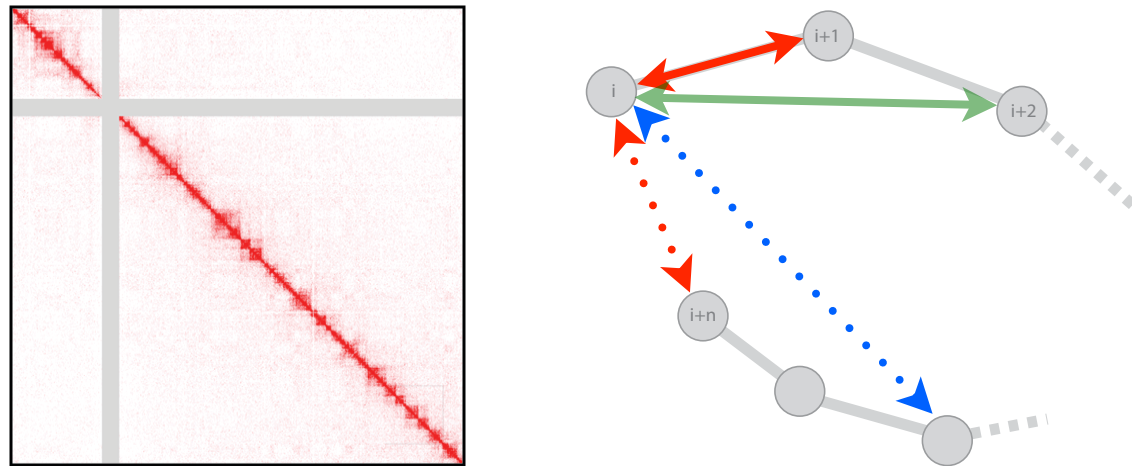


# Are TADs homogeneous?



# TAD 3D structure modeling

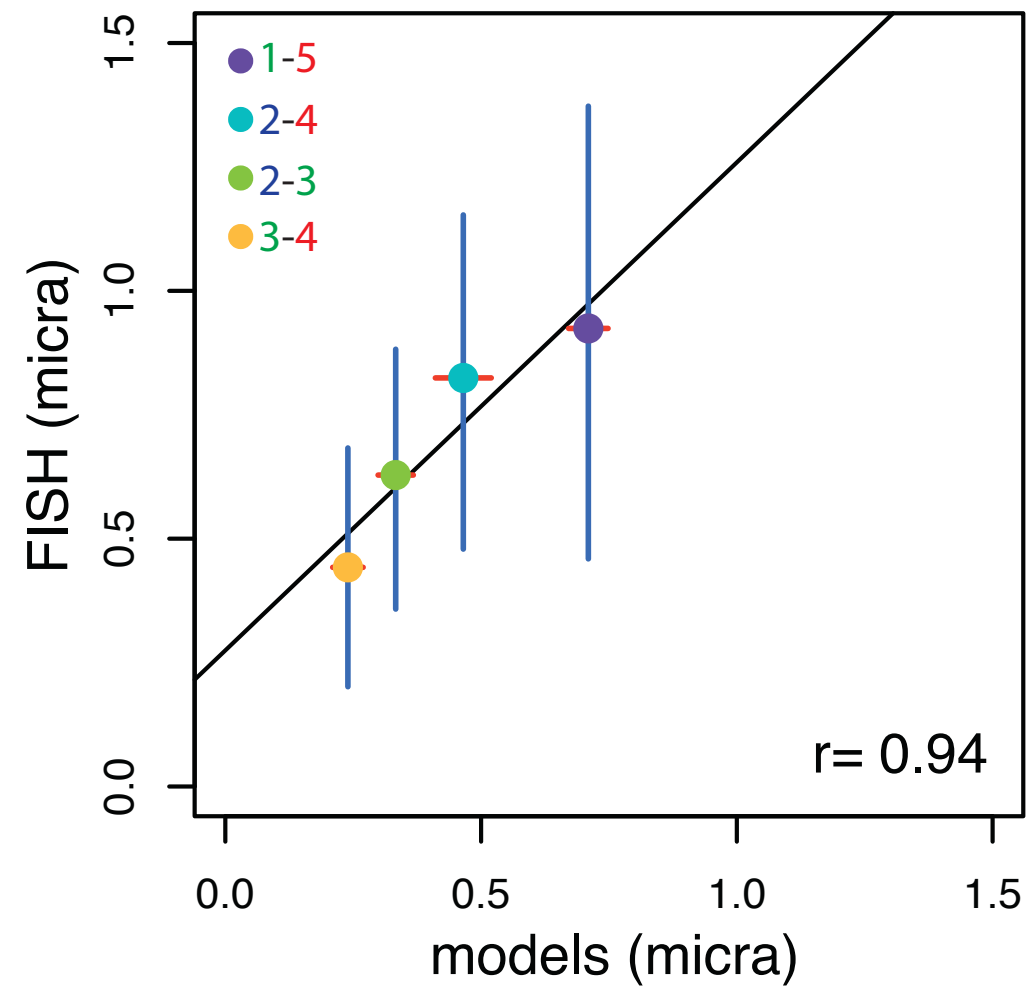
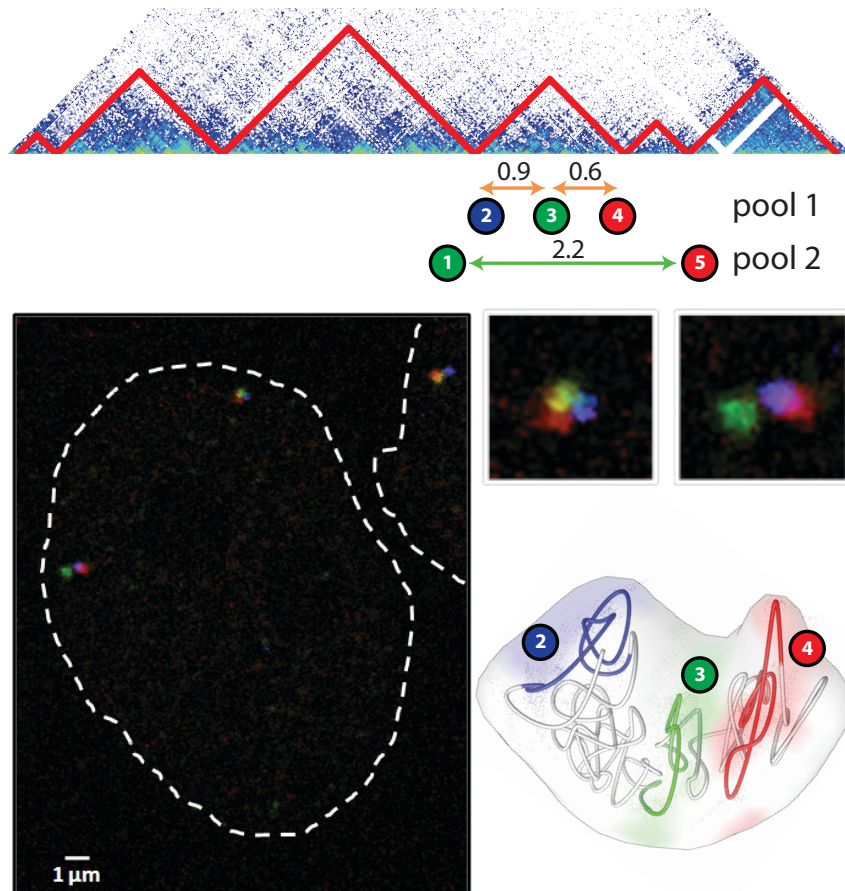
We modeled 15 genomic regions with a total of 105 TADs





# FISH validation of the 3D models

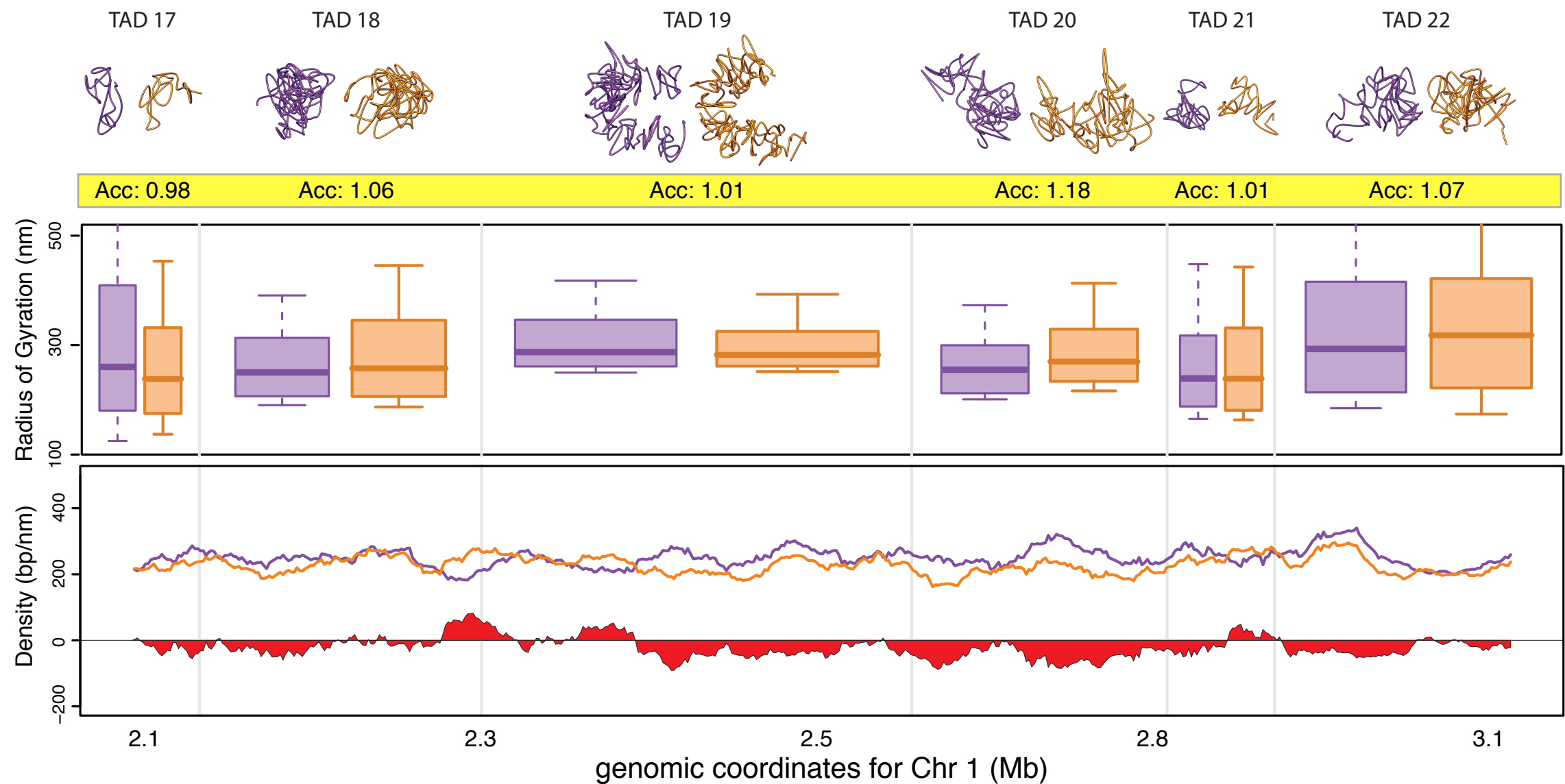
Chr1:21000000-31400000  
6 TADs



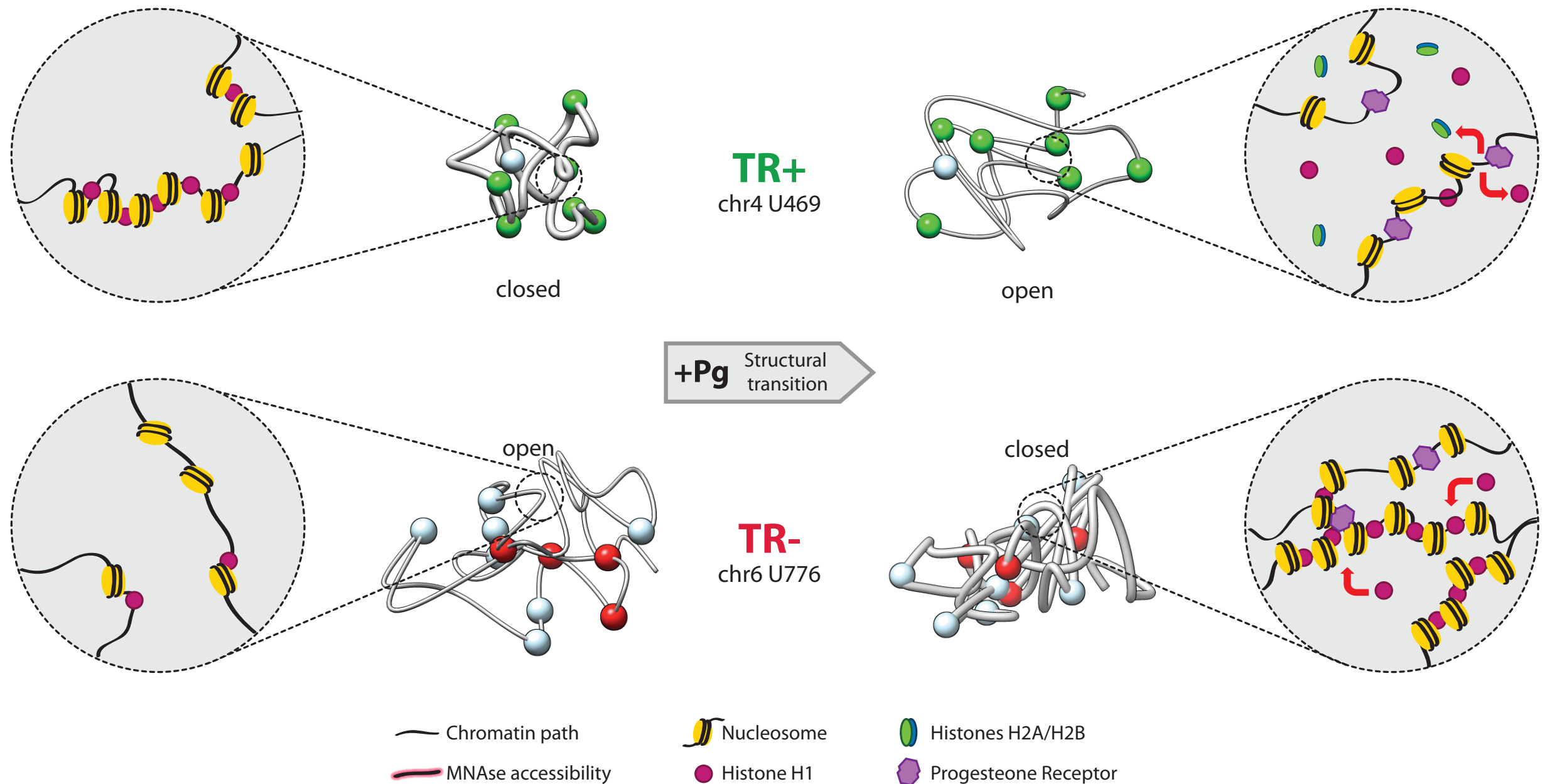


# Accessibility changes of TADs

Chr1:21000000-31400000  
6 TADs



# Proposed model for TAD regulation



# Acknowledgments



Marc A. Marti-Renom

François le Dily

François Serra



Miguel Beato

Chromatin and Gene Expression Group (CRG)



Guillaume Filion

Genome Architecture Group (CRG)

Travel funding to ISMB/ECCB 2013 was generously provided by ISCB

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

