# From HiC matrices to 3D structures

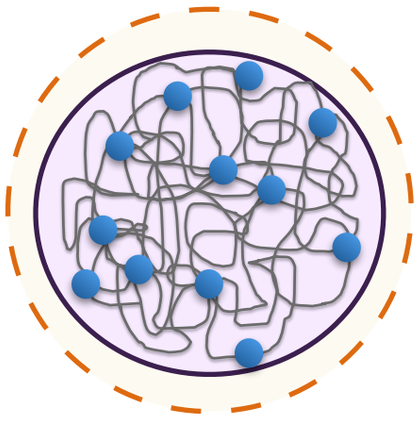
Francois Serra, Paula Soler, Marc A Marti-Renom
Genome Biology Group (CNAG)
Structural Genomics Group (CRG)

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

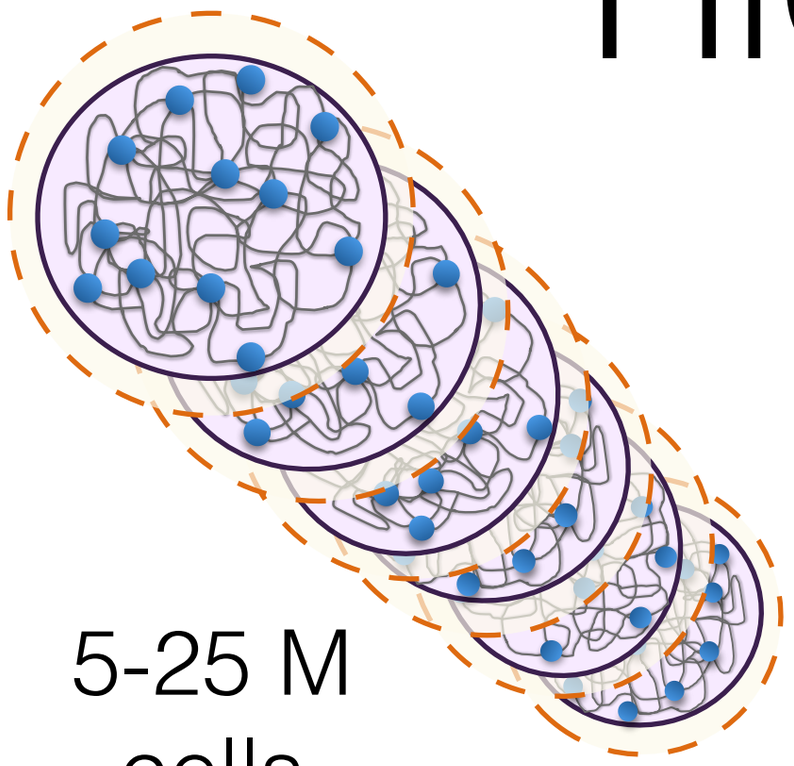
CRG
Centre for Genomic Regulation

http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu

# HiC experiment

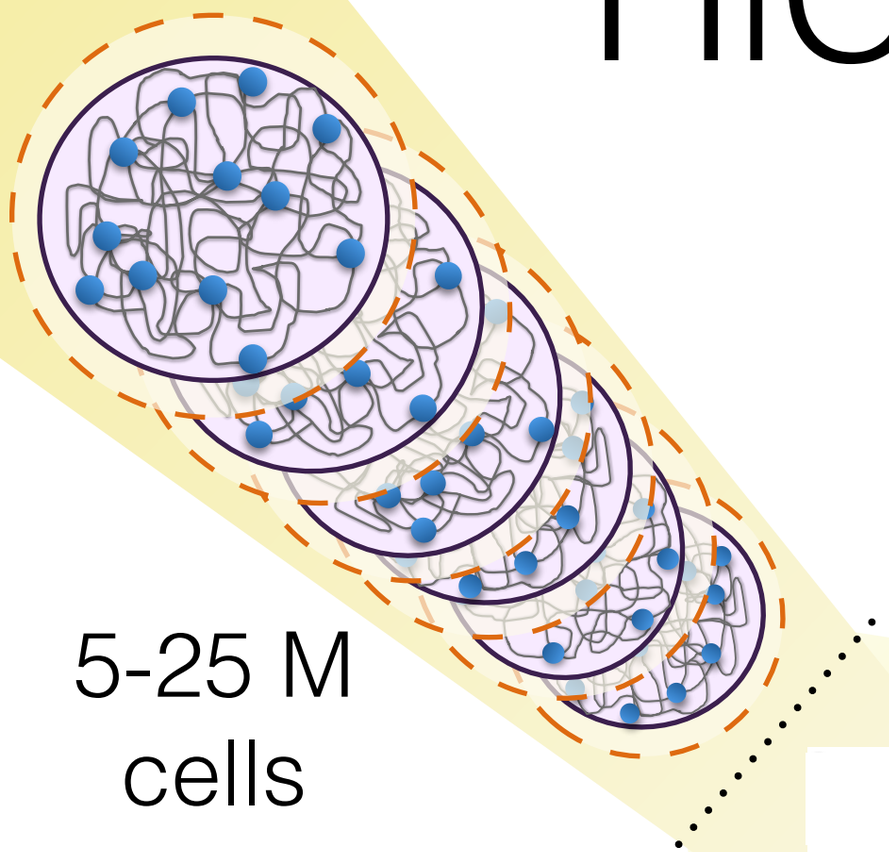


# HiC experiment



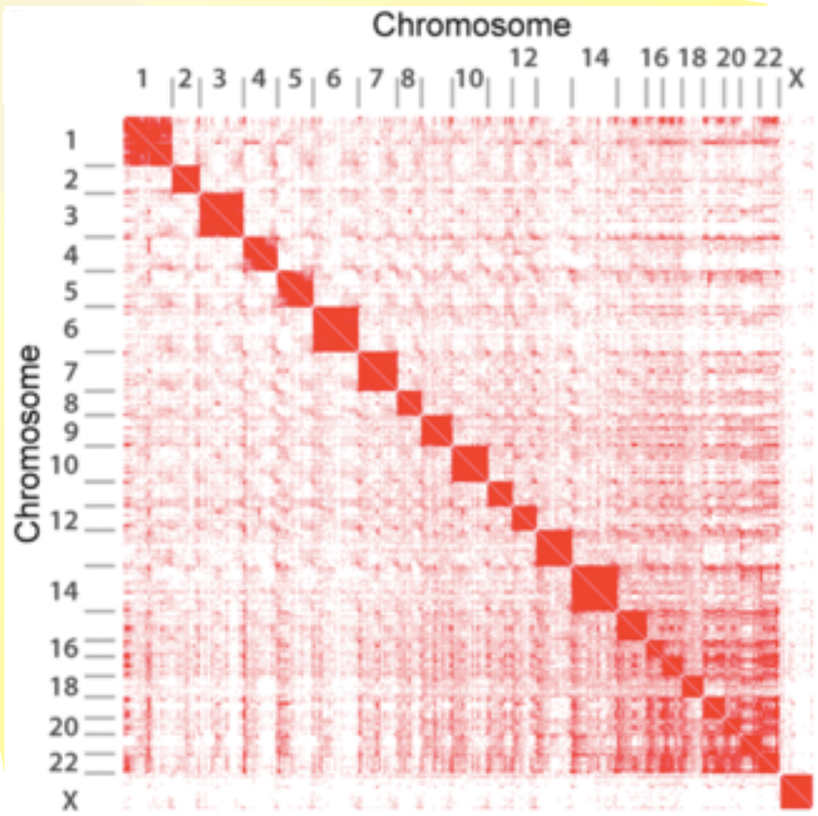
5-25 M cells

# HiC experiment



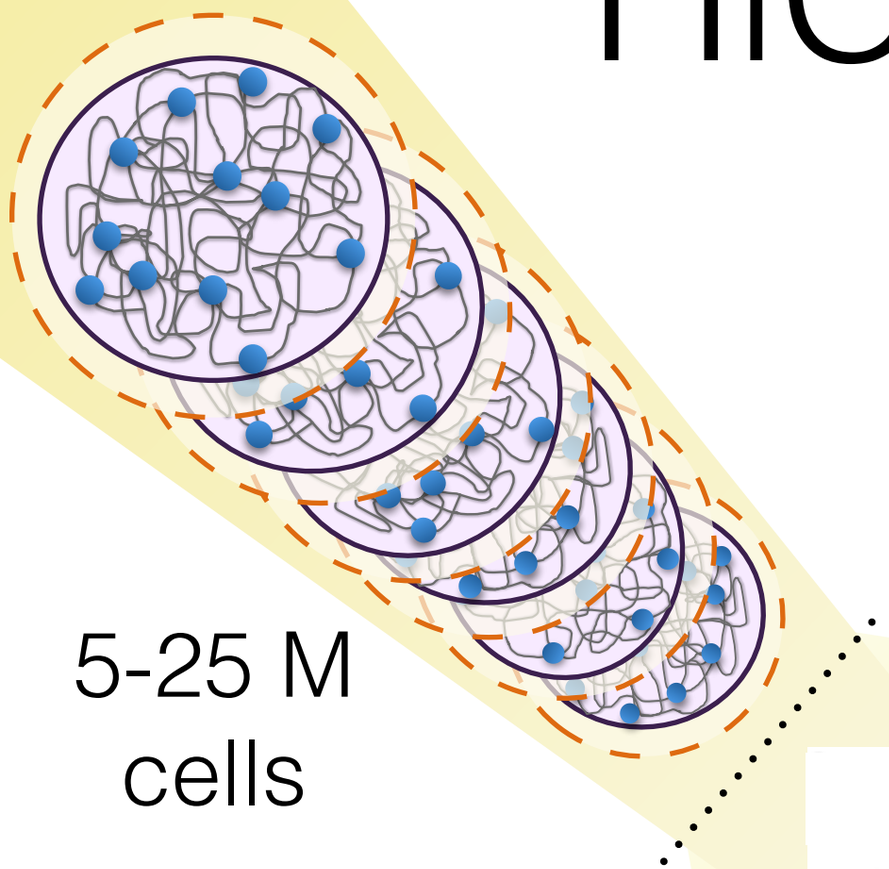
5-25 M cells

50-1000 M interactions

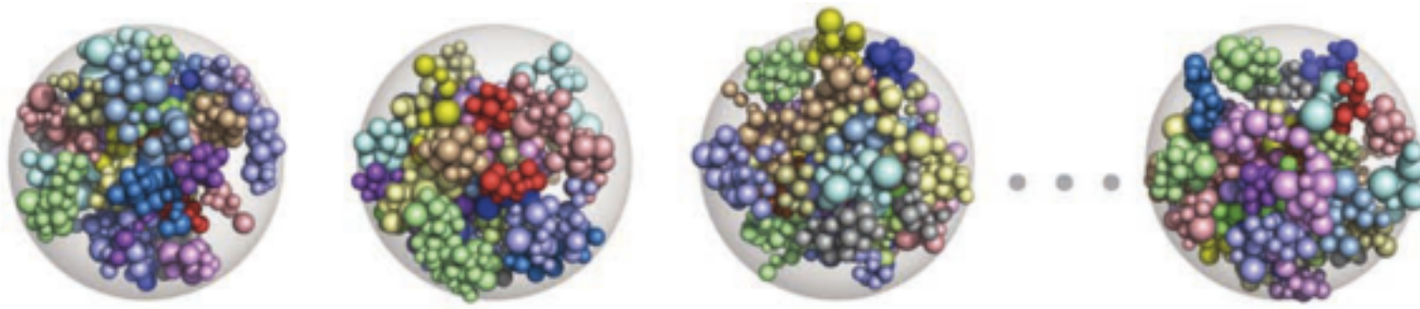



Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L. 2012.
Genome architectures revealed by tethered chromosome conformation capture and population-based modeling.
Nat Biotechnol 30: 90–8.

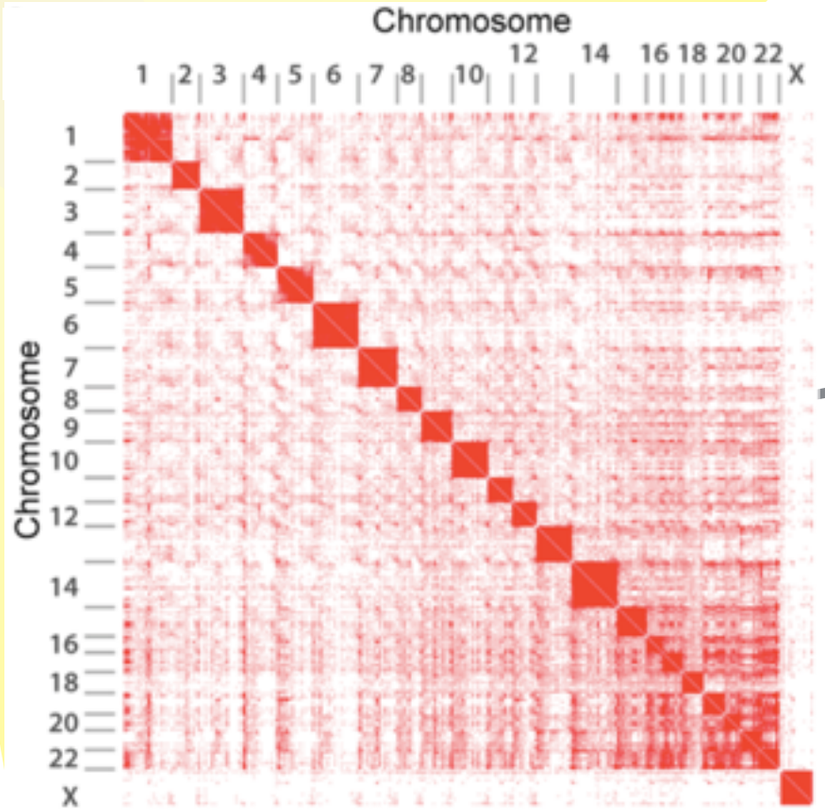
# HiC experiment



5-25 M cells



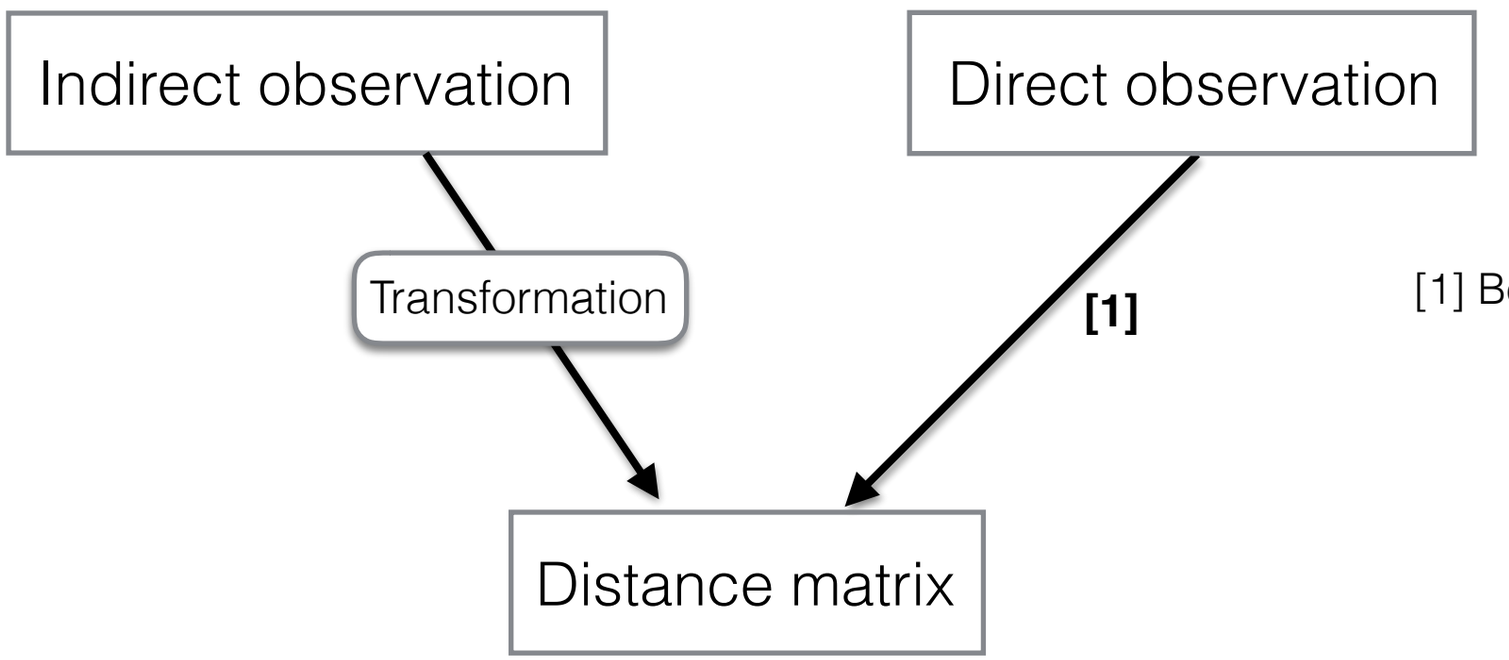
1-10,000 models




50-1000 M interactions

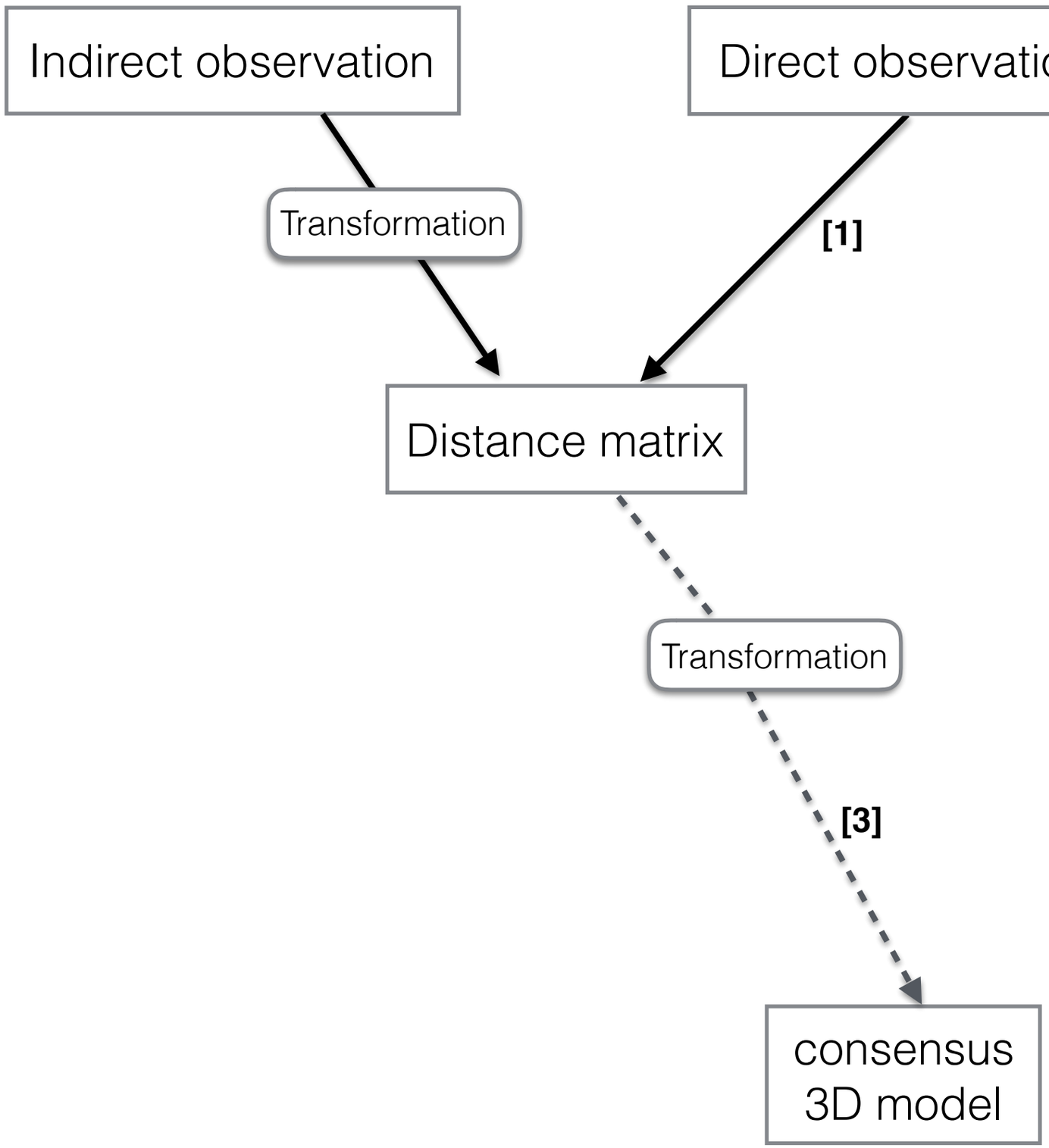
Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L. 2012.
Genome architectures revealed by tethered chromosome conformation capture and population-based modeling.
Nat Biotechnol 30: 90–8.

# Modeling strategies




[1] Bolzer A et al. 2005. PLoS Biol 3: 0826–0842.

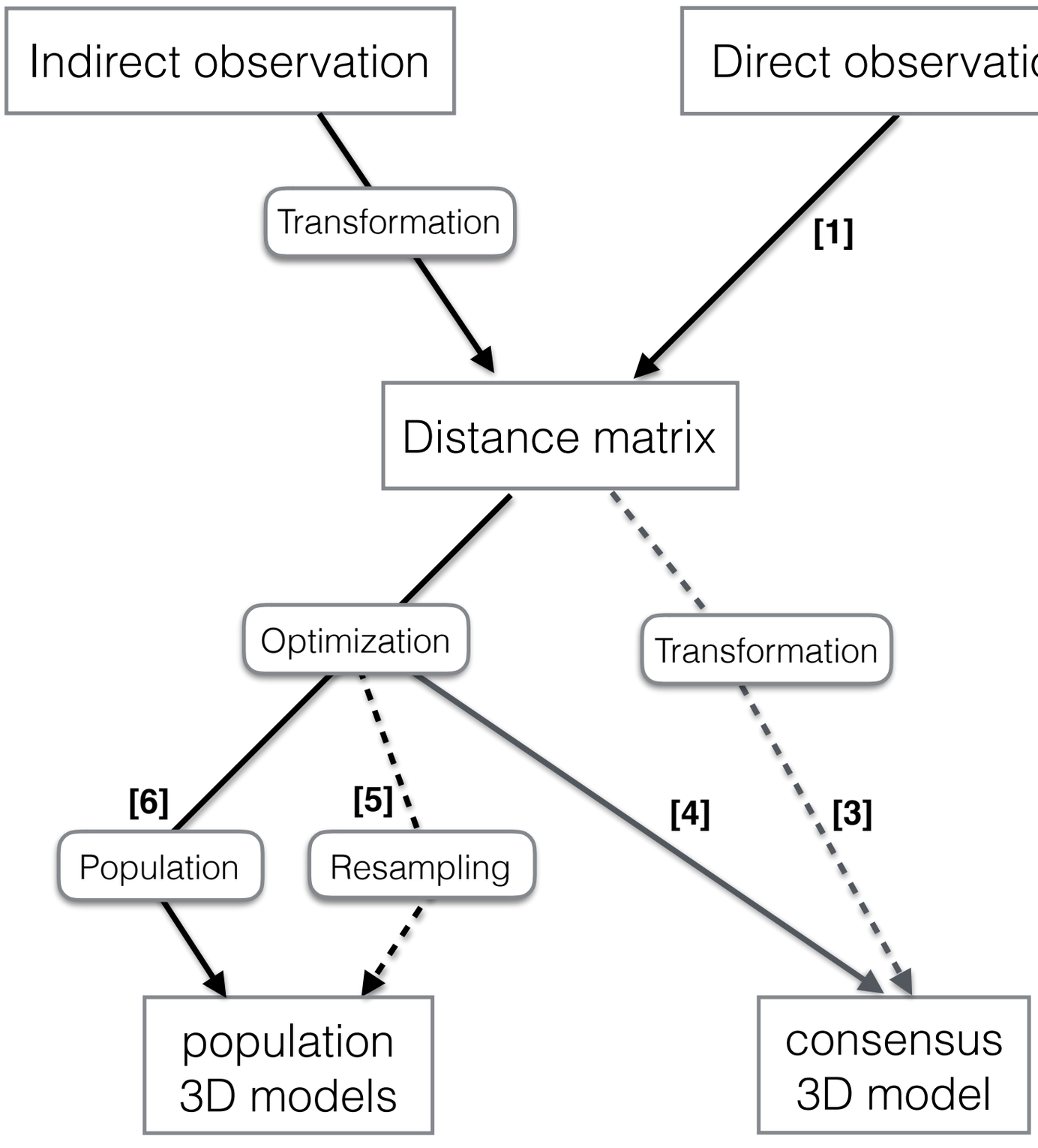
# Modeling strategies




[1] Bolzer A et al. 2005. PLoS Biol 3: 0826–0842.

[3] Zhang Z et al. 2013. J Comput Biol 20: 831–46.
Lesne A et al. 2014. Nat Methods 4: 10–13.

# Modeling strategies





[1] Bolzer A et al. 2005. PLoS Biol 3: 0826–0842.

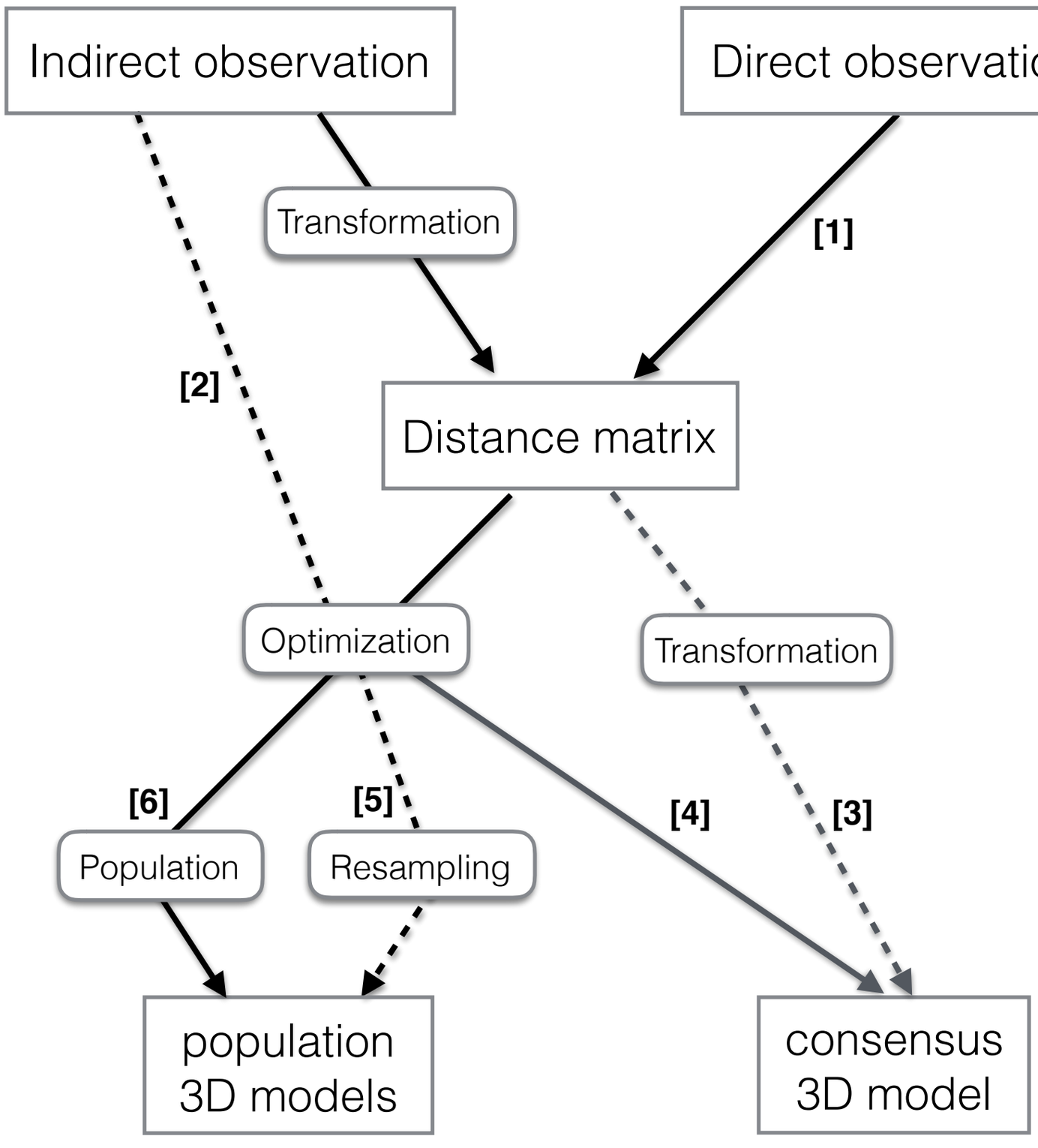
[3] Zhang Z et al. 2013. J Comput Biol 20: 831–46.
Lesne A et al. 2014. Nat Methods 4: 10–13.

[4] Peng C et al. 2013. Nucleic Acids Res 41: e183.

[5] Meluzzi D, Arya G. 2013. Nucleic Acids Res 41: 63–75.
Duan Z et al. 2010. Nature 465: 363–7.
Rousseau M et al. 2011. BMC Bioinformatics 12: 414.
Varoquaux N et al. 2014. Bioinformatics 30: i26–33.
Serra F et al. 2016. bioRxiv 1–29.

[6] Kalhor R et al. 2012. Nat Biotechnol 30: 90–8.
Giorgetti L et al. 2014. Cell 157: 950–963.
Hu M et al. 2013. PLoS Comput Biol 9: e1002893.

# Modeling strategies





[1] Bolzer A et al. 2005. PLoS Biol 3: 0826–0842.

[2] Meluzzi D, Arya G. 2013. Nucleic Acids Res 41: 63–75.
Kalhor R et al. 2012. Nat Biotechnol 30: 90–8.

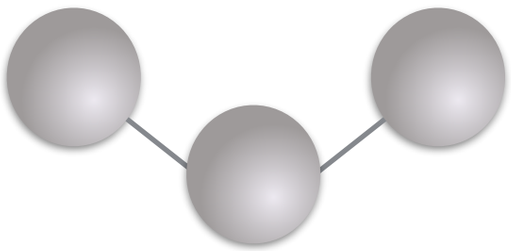
[3] Zhang Z et al. 2013. J Comput Biol 20: 831–46.
Lesne A et al. 2014. Nat Methods 4: 10–13.

[4] Peng C et al. 2013. Nucleic Acids Res 41: e183.

[5] Meluzzi D, Arya G. 2013. Nucleic Acids Res 41: 63–75.
Duan Z et al. 2010. Nature 465: 363–7.
Rousseau M et al. 2011. BMC Bioinformatics 12: 414.
Varoquaux N et al. 2014. Bioinformatics 30: i26–33.
Serra F et al. 2016. bioRxiv 1–29.

[6] Kalhor R et al. 2012. Nat Biotechnol 30: 90–8.
Giorgetti L et al. 2014. Cell 157: 950–963.
Hu M et al. 2013. PLoS Comput Biol 9: e1002893.

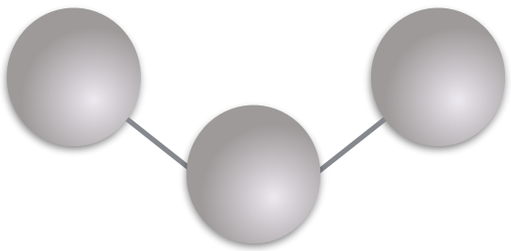
# The representation



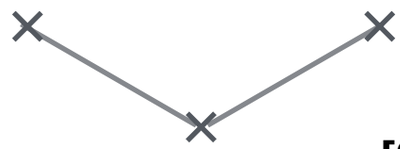
[1]


[1] Kalhor R et al. 2012. Nat Biotechnol 30: 90–8.
Giorgetti L et al. 2014. Cell 157: 950–963.
Duan Z et al. 2010. Nature 465: 363–7.
Peng C et al. 2013. Nucleic Acids Res 41: e183.
Meluzzi D, Arya G. 2013. Nucleic Acids Res 41: 63–75.
Serra F et al. 2016. bioRxiv 1–29.

# The representation



[1]

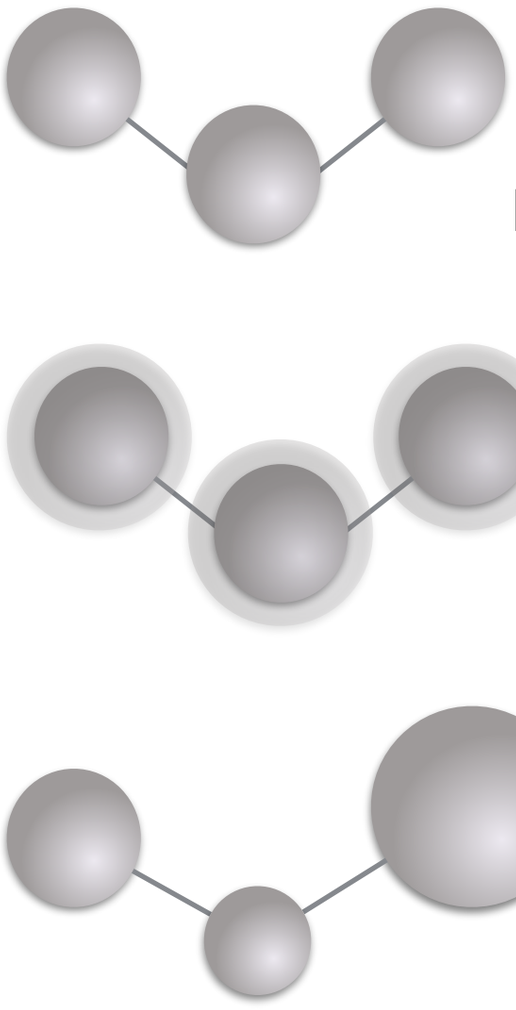


[2]

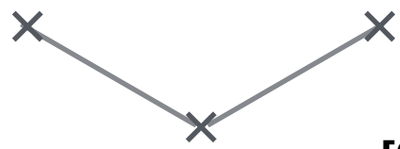

[1] Kalhor R et al. 2012. Nat Biotechnol 30: 90–8.
Giorgetti L et al. 2014. Cell 157: 950–963.
Duan Z et al. 2010. Nature 465: 363–7.
Peng C et al. 2013. Nucleic Acids Res 41: e183.
Meluzzi D, Arya G. 2013. Nucleic Acids Res 41: 63–75.
Serra F et al. 2016. bioRxiv 1–29.

[2] Zhang Z et al. 2013. J Comput Biol 20: 831–46.
Lesne A et al. 2014. Nat Methods 4: 10–13.
Hu M et al. 2013. PLoS Comput Biol 9: e1002893.
Varoquaux N et al. 2014. Bioinformatics 30: i26–33.
Rousseau M et al. 2011. BMC Bioinformatics 12: 414.

# The representation



[1]

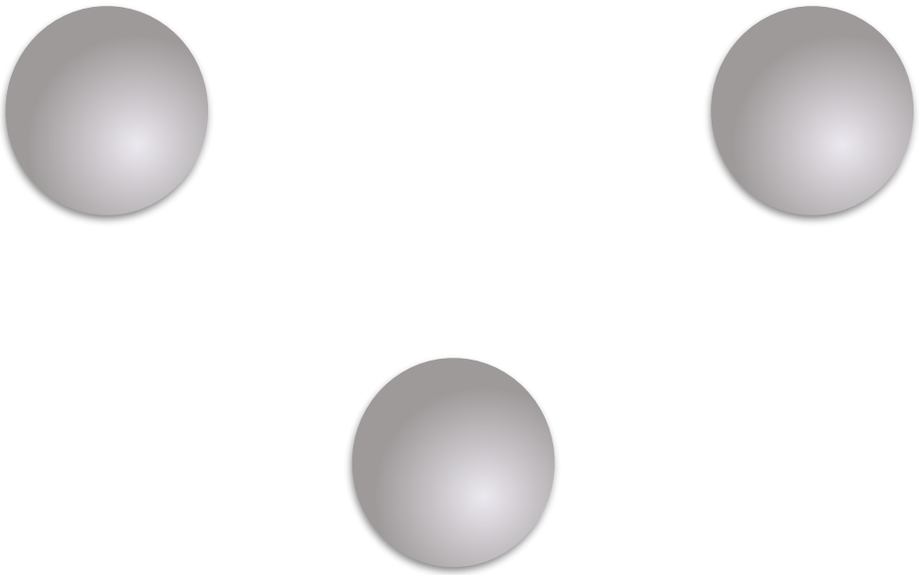


[2]


[1] Kalhor R et al. 2012. Nat Biotechnol 30: 90–8.
Giorgetti L et al. 2014. Cell 157: 950–963.
Duan Z et al. 2010. Nature 465: 363–7.
Peng C et al. 2013. Nucleic Acids Res 41: e183.
Meluzzi D, Arya G. 2013. Nucleic Acids Res 41: 63–75.
Serra F et al. 2016. bioRxiv 1–29.

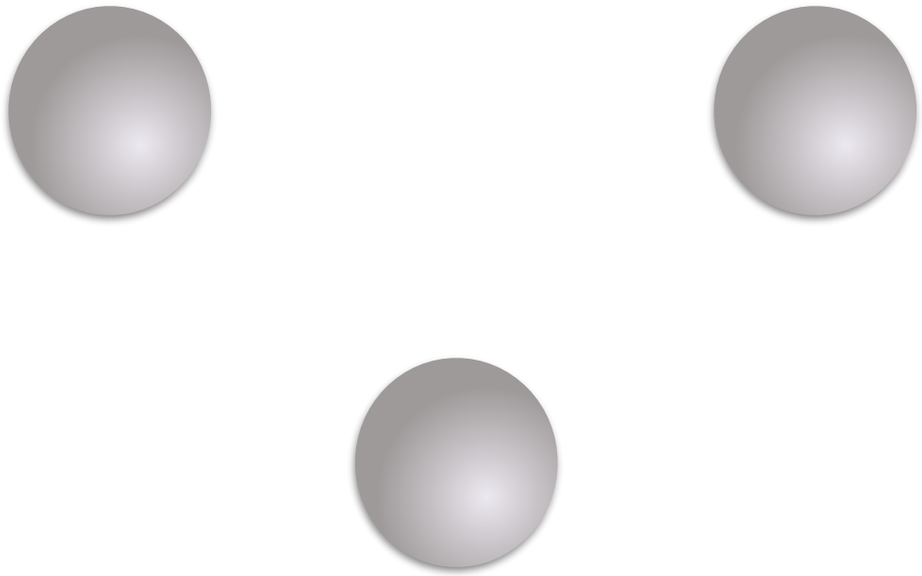
[2] Zhang Z et al. 2013. J Comput Biol 20: 831–46.
Lesne A et al. 2014. Nat Methods 4: 10–13.
Hu M et al. 2013. PLoS Comput Biol 9: e1002893.
Varoquaux N et al. 2014. Bioinformatics 30: i26–33.
Rousseau M et al. 2011. BMC Bioinformatics 12: 414.

# The representation



$$S = U_{3C} + U_{Biol} + U_{Phys}$$

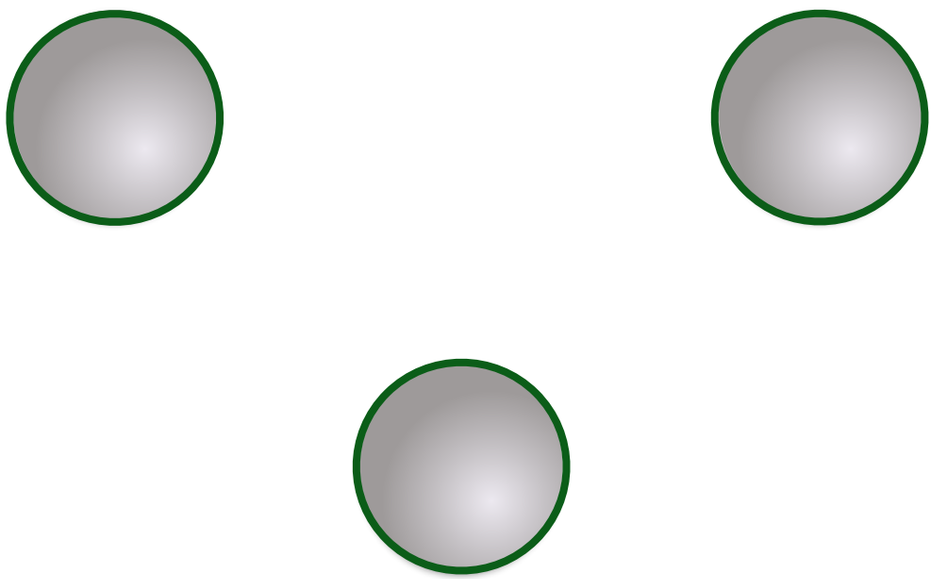
# The representation



$$S = U_{3C} + U_{Biol} + U_{Phys}$$

Tethering
Nucleus size

# The representation

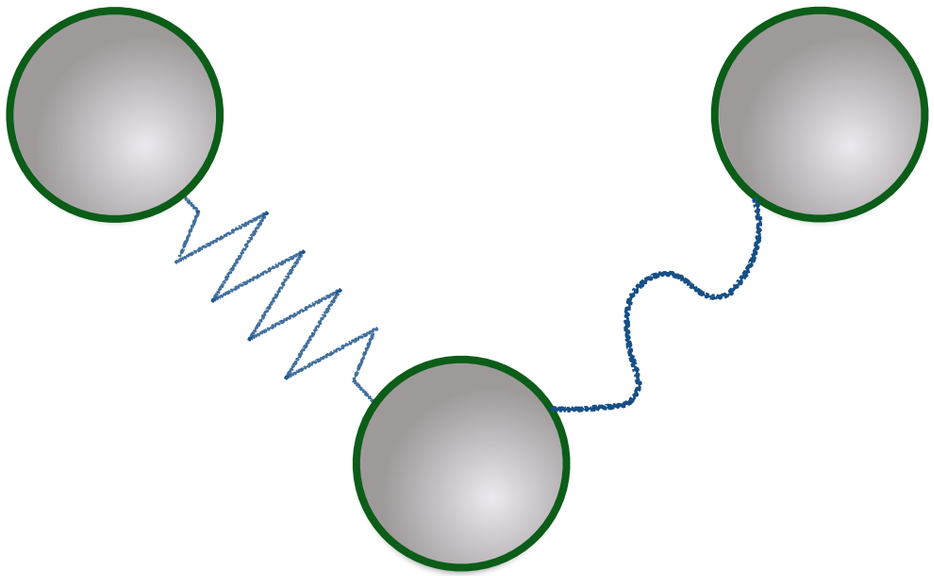


$U_{Excl}$

$S = U_{3C} + U_{Biol} + U_{Phys}$

Tethering
Nucleus size

# The representation

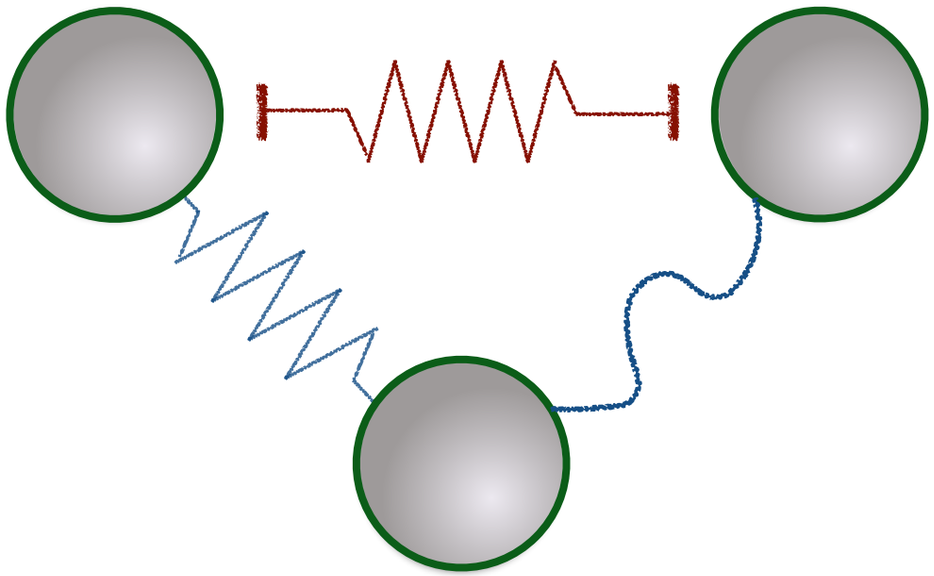


$U_{Excl} + U_{Bond}$

$$S = U_{3C} + U_{Biol} + U_{Phys}$$

Tethering
Nucleus size

# The representation

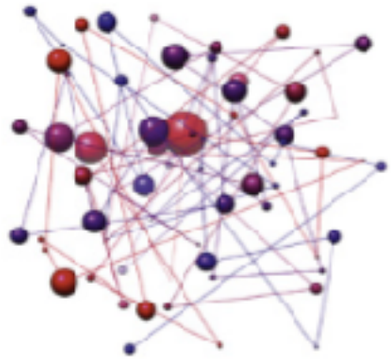


$$U_{Excl} + U_{Bond} + U_{Bend}$$

$$S = U_{3C} + U_{Biol} + U_{Phys}$$

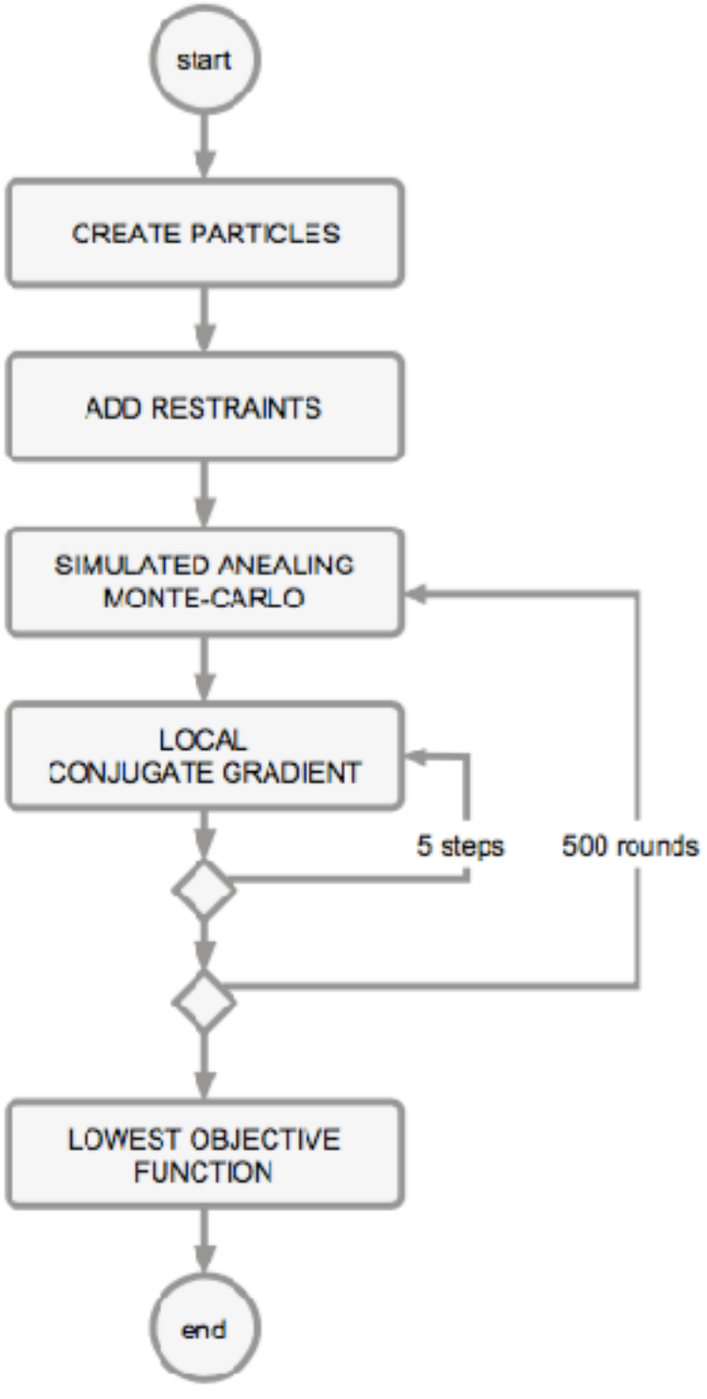
Tethering
Nucleus size

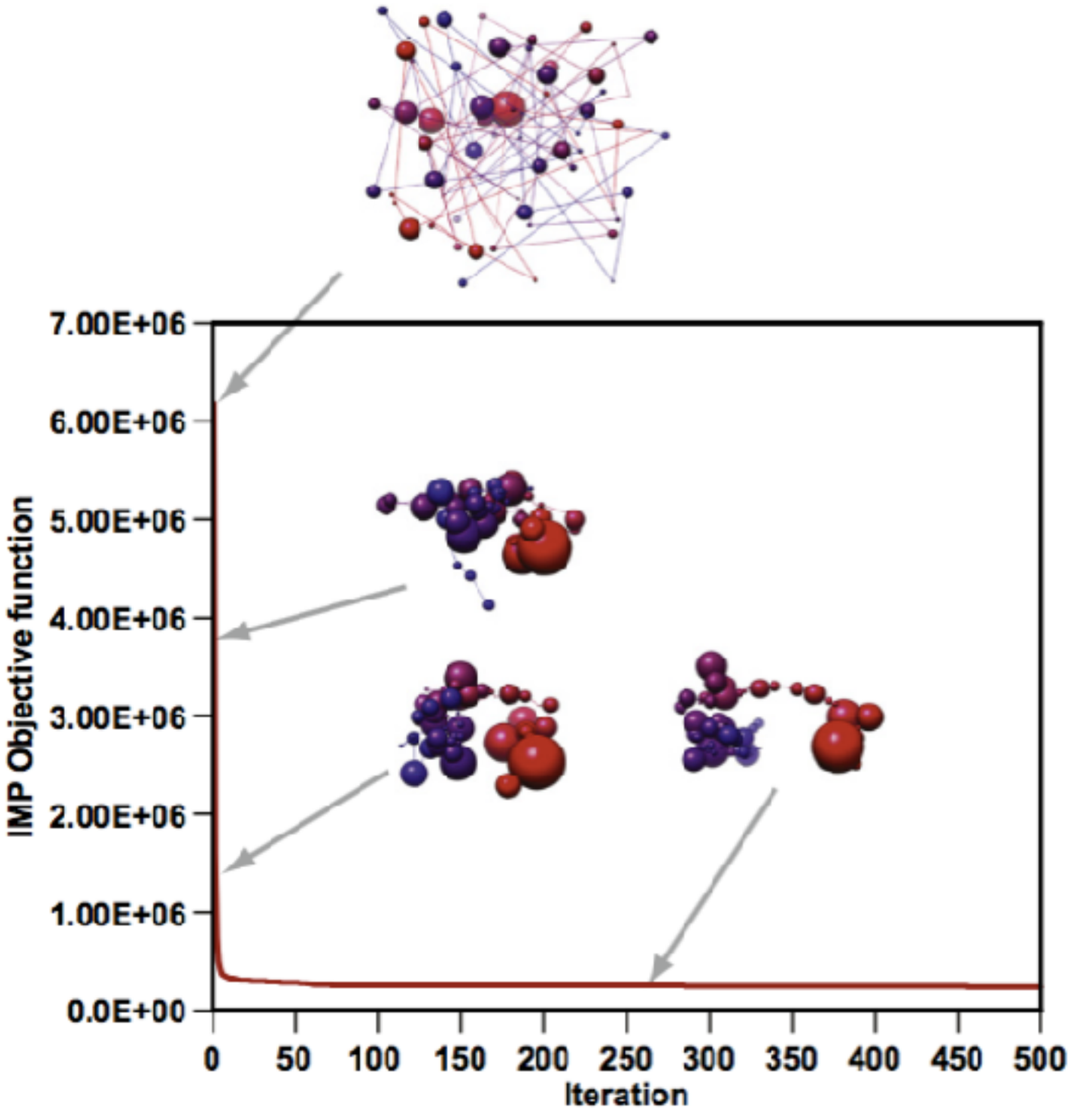
# Model optimization



Baù D, Marti-Renom MA. 2012.
Genome structure determination via 3C-based data integration by the Integrative Modeling Platform.
Methods 58: 300–6.

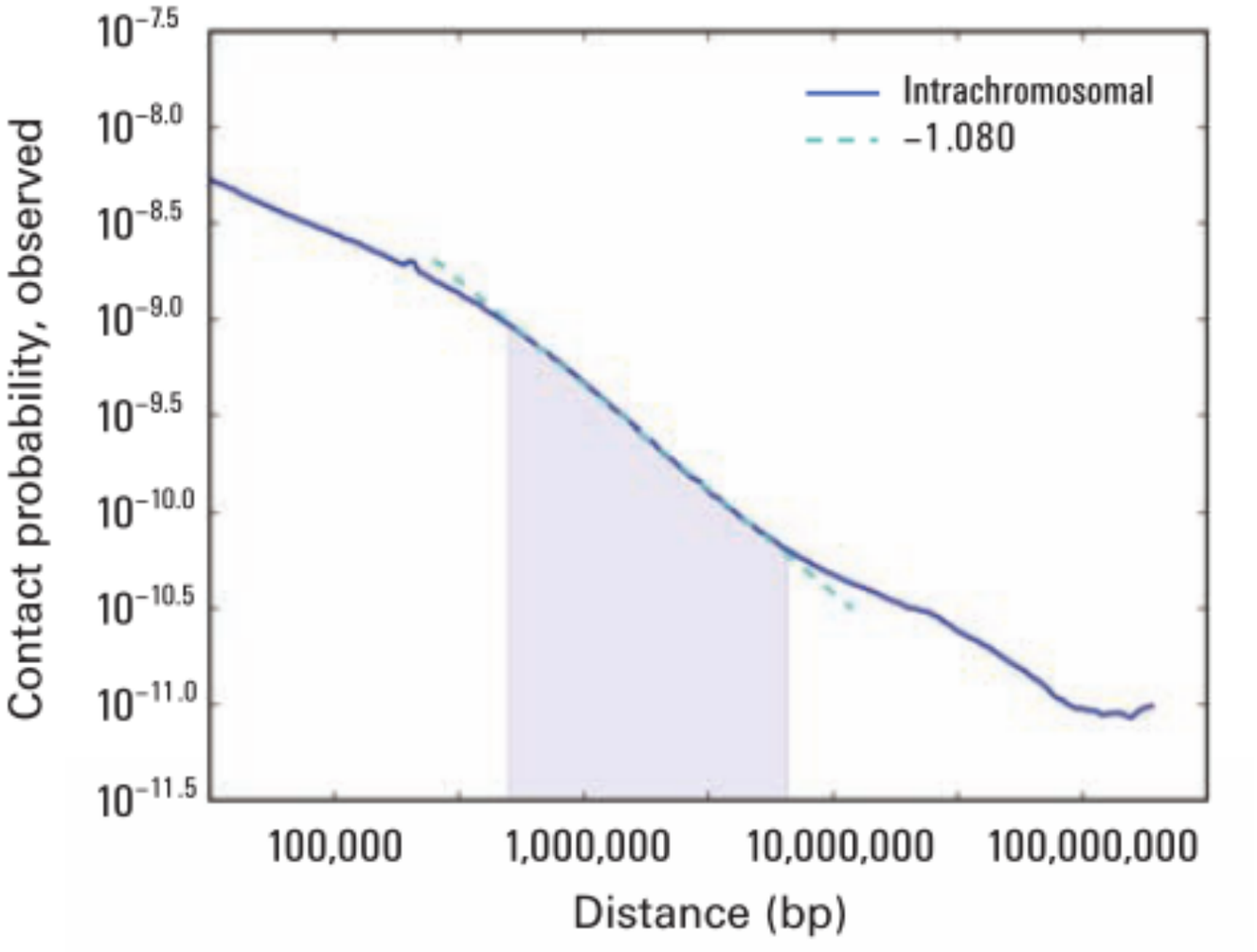
# Model optimization



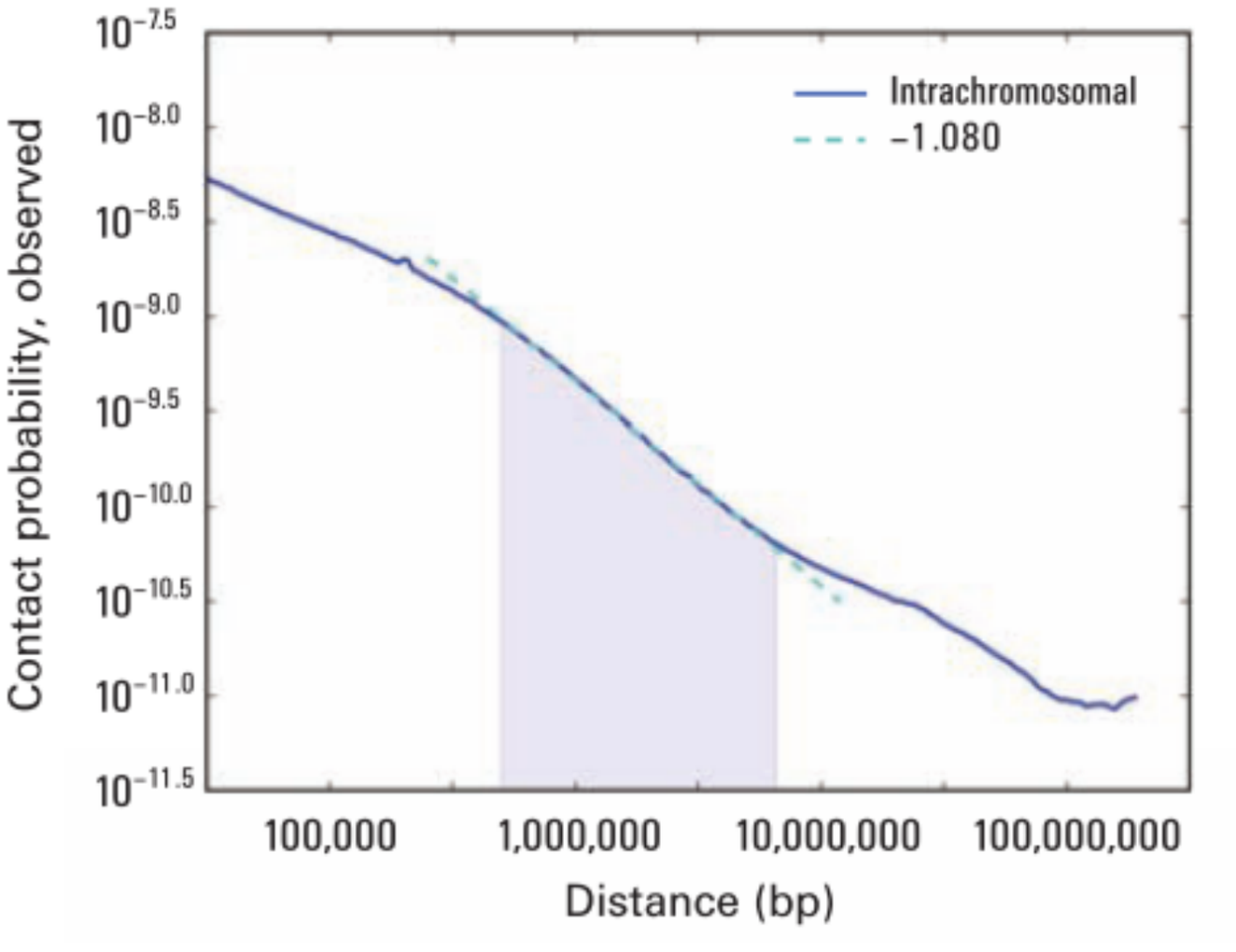

Baù D, Marti-Renom MA. 2012.
Genome structure determination via 3C-based data integration by the Integrative Modeling Platform.
Methods 58: 300–6.

# From interactions to distance

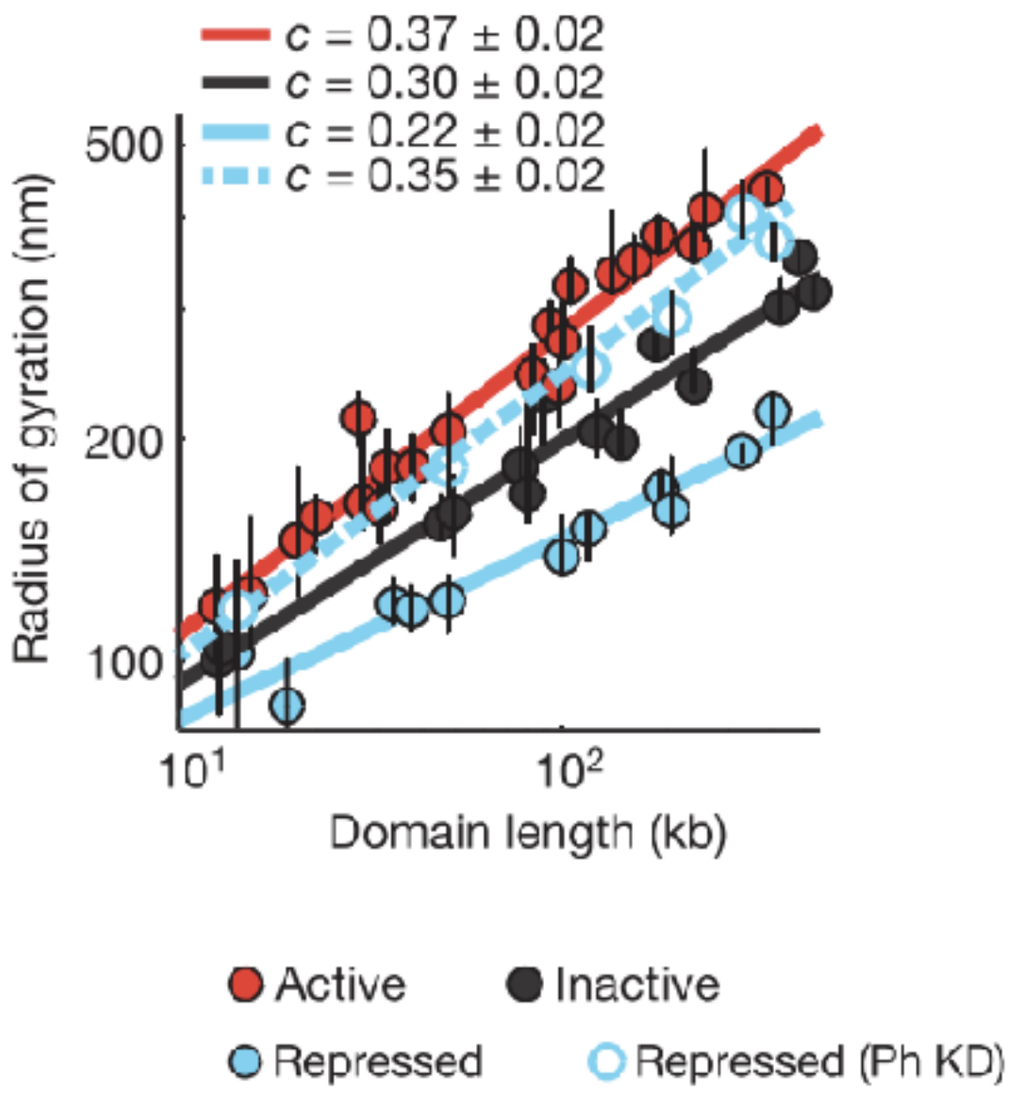



Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M V, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (80- ) 326: 289–93.

# From interactions to distance



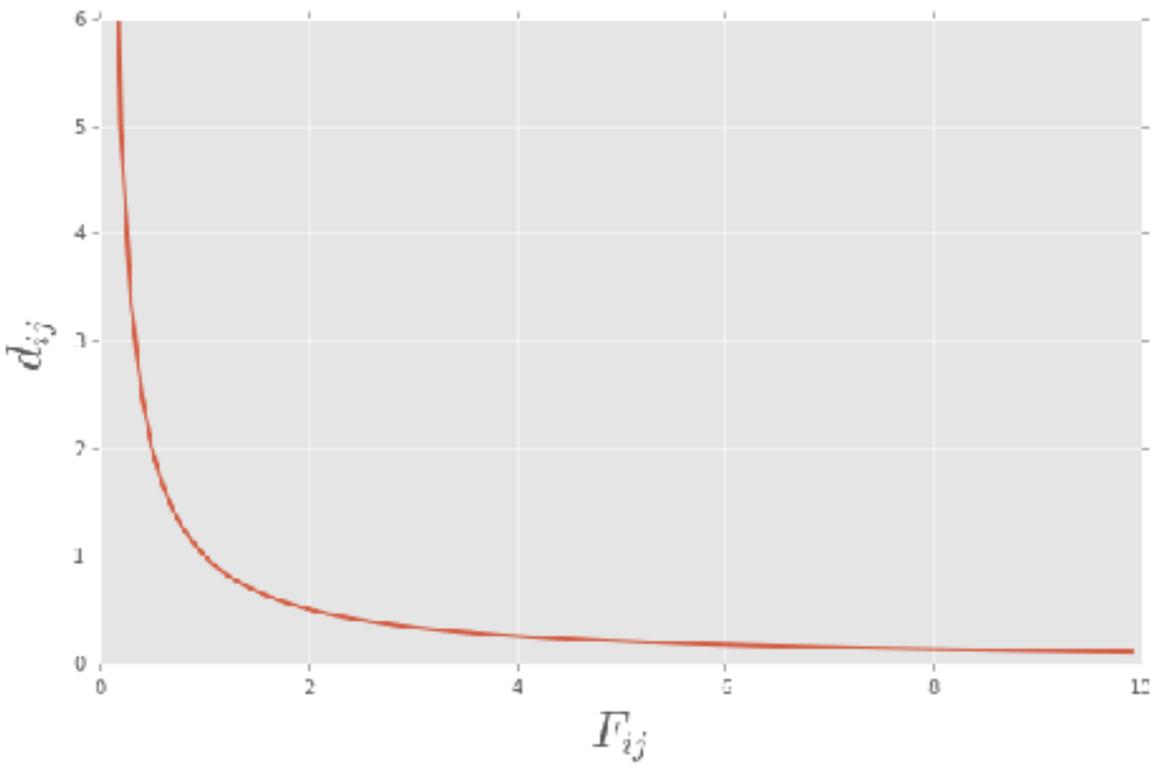

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M V, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009.
Comprehensive mapping of long-range interactions reveals folding principles of the human genome.
Science (80- ) 326: 289–93.




Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, Imakaev M, Mirny LA, Wu C, Zhuang X. 2016.
Chromatin Folding for Different Epigenetic States.
Nature 1–15.

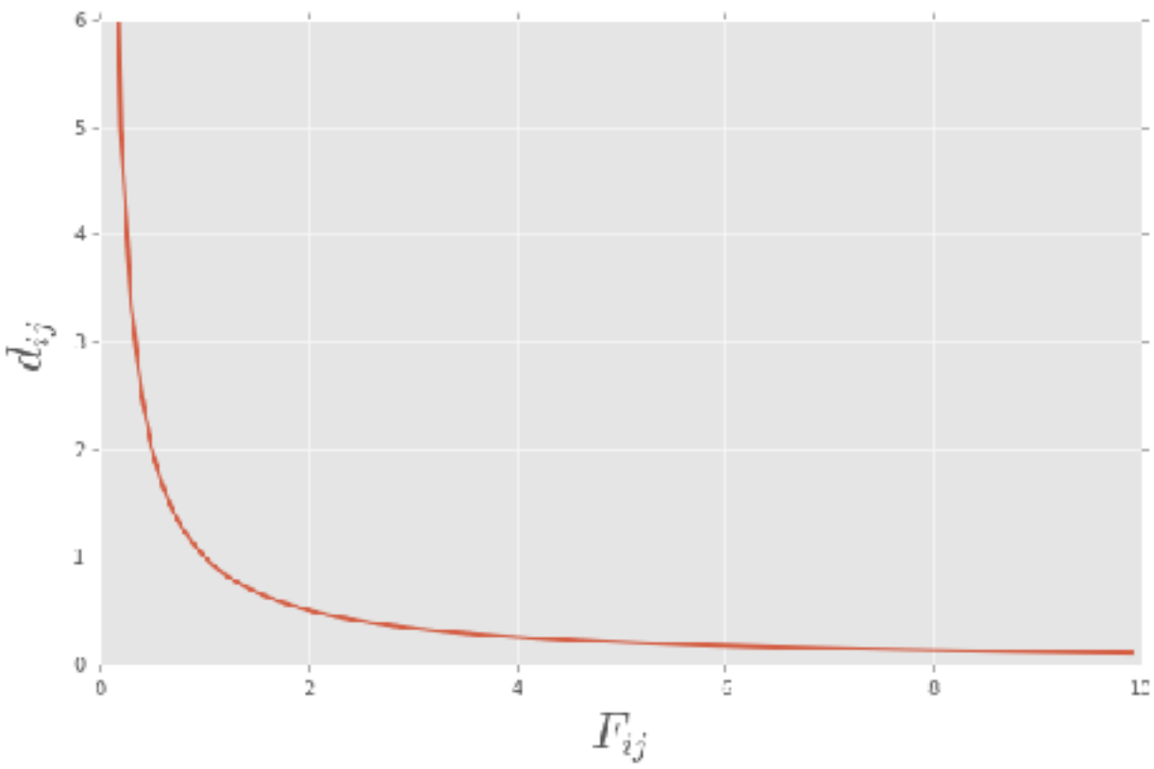
# From interactions to distance

$$d_{ij} \propto \left(\frac{1}{F_{ij}}\right)^{\alpha}$$

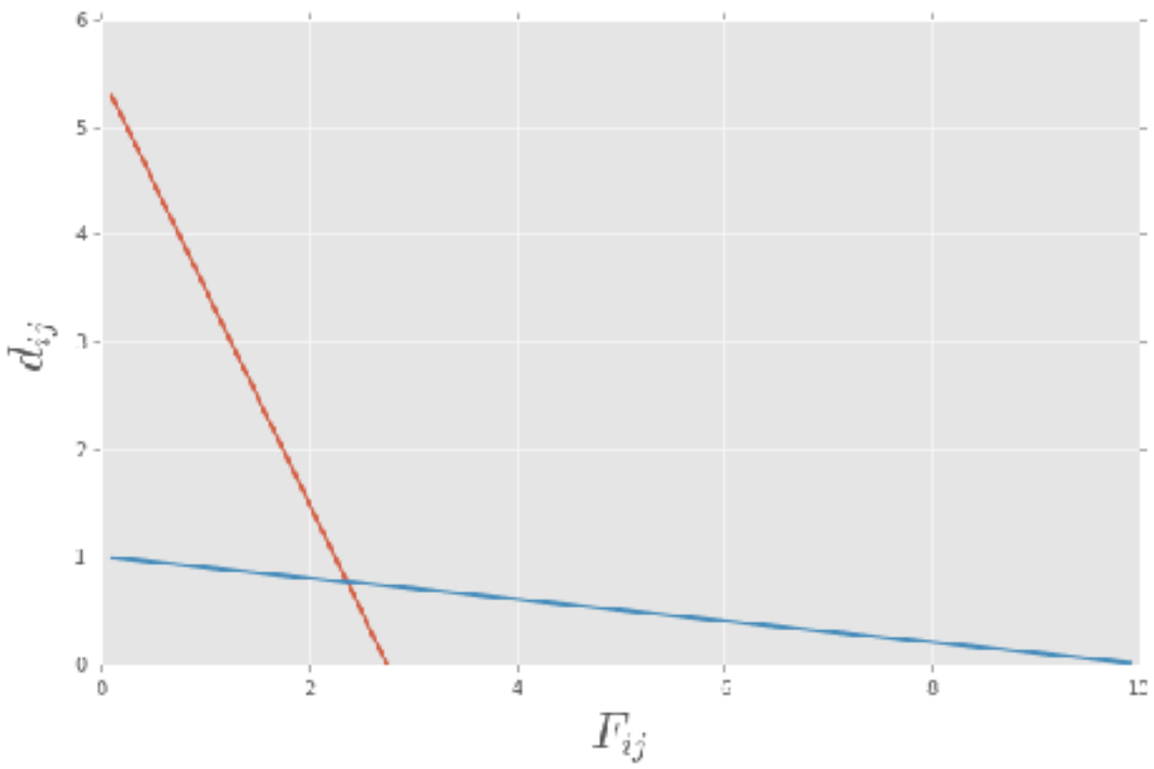


# From interactions to distance

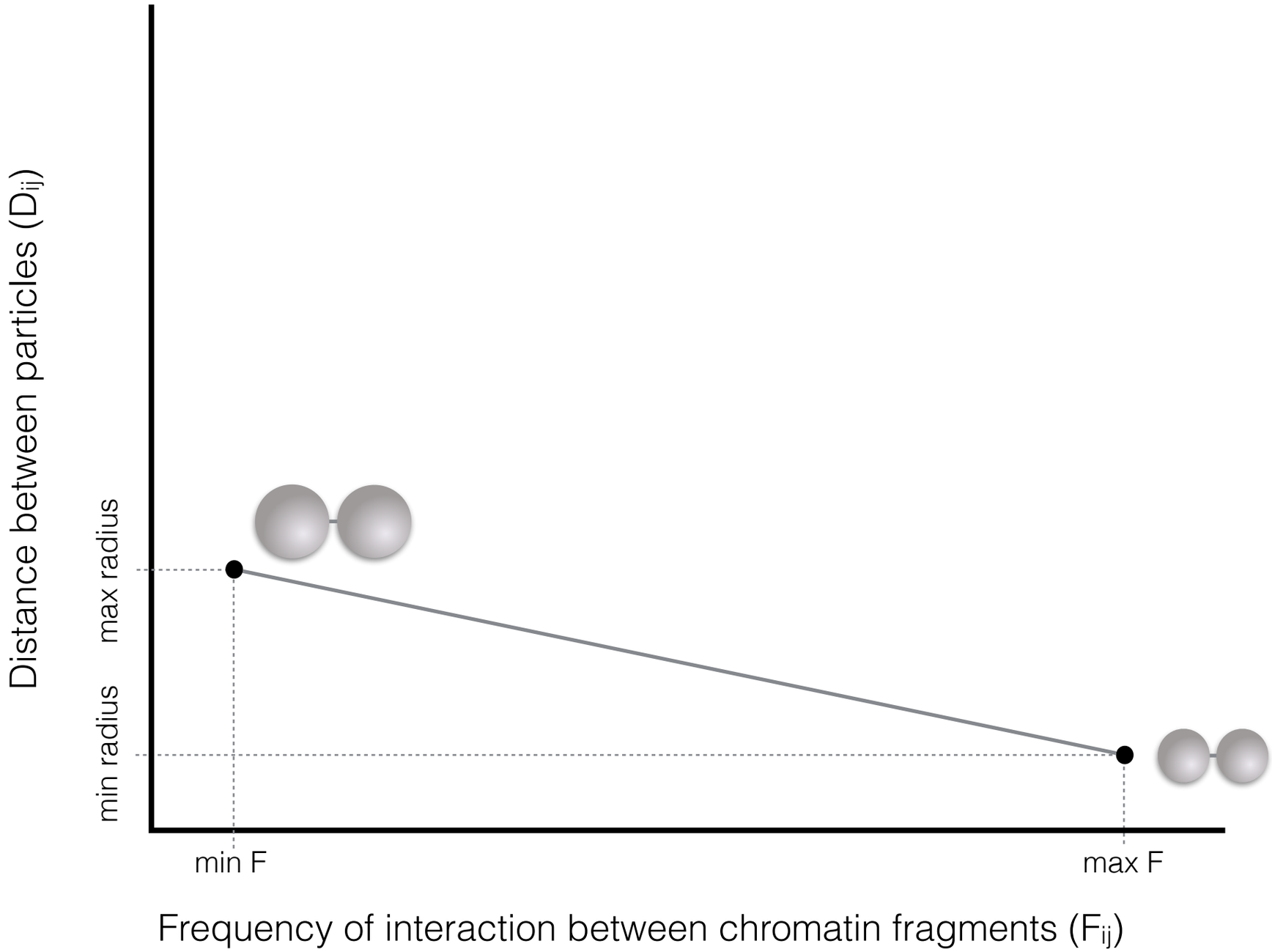
$$d_{ij} \propto \left(\frac{1}{F_{ij}}\right)^{\alpha}$$




$$d_{ij} \propto \alpha F_{ij} + \beta$$

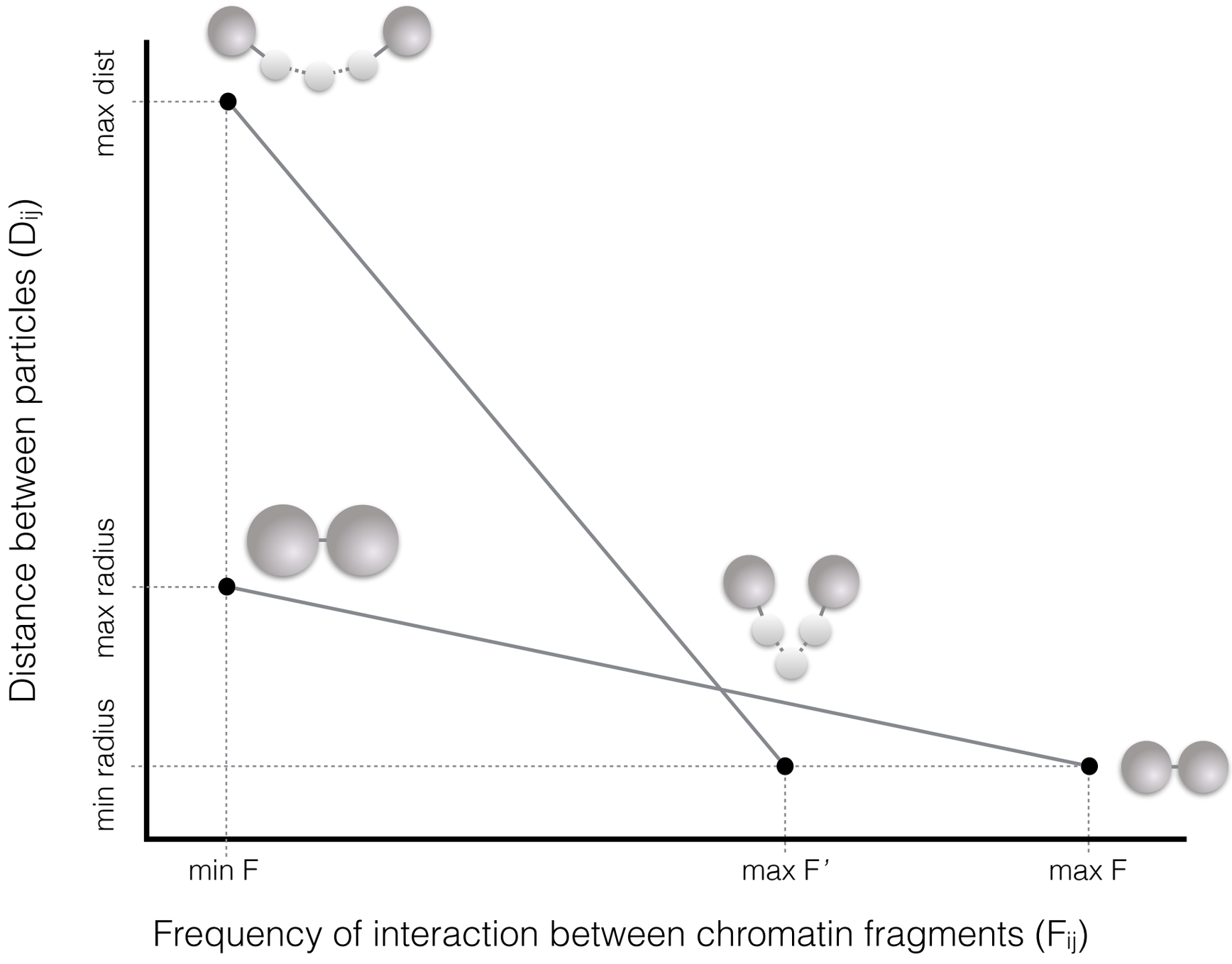


# From interactions to distance



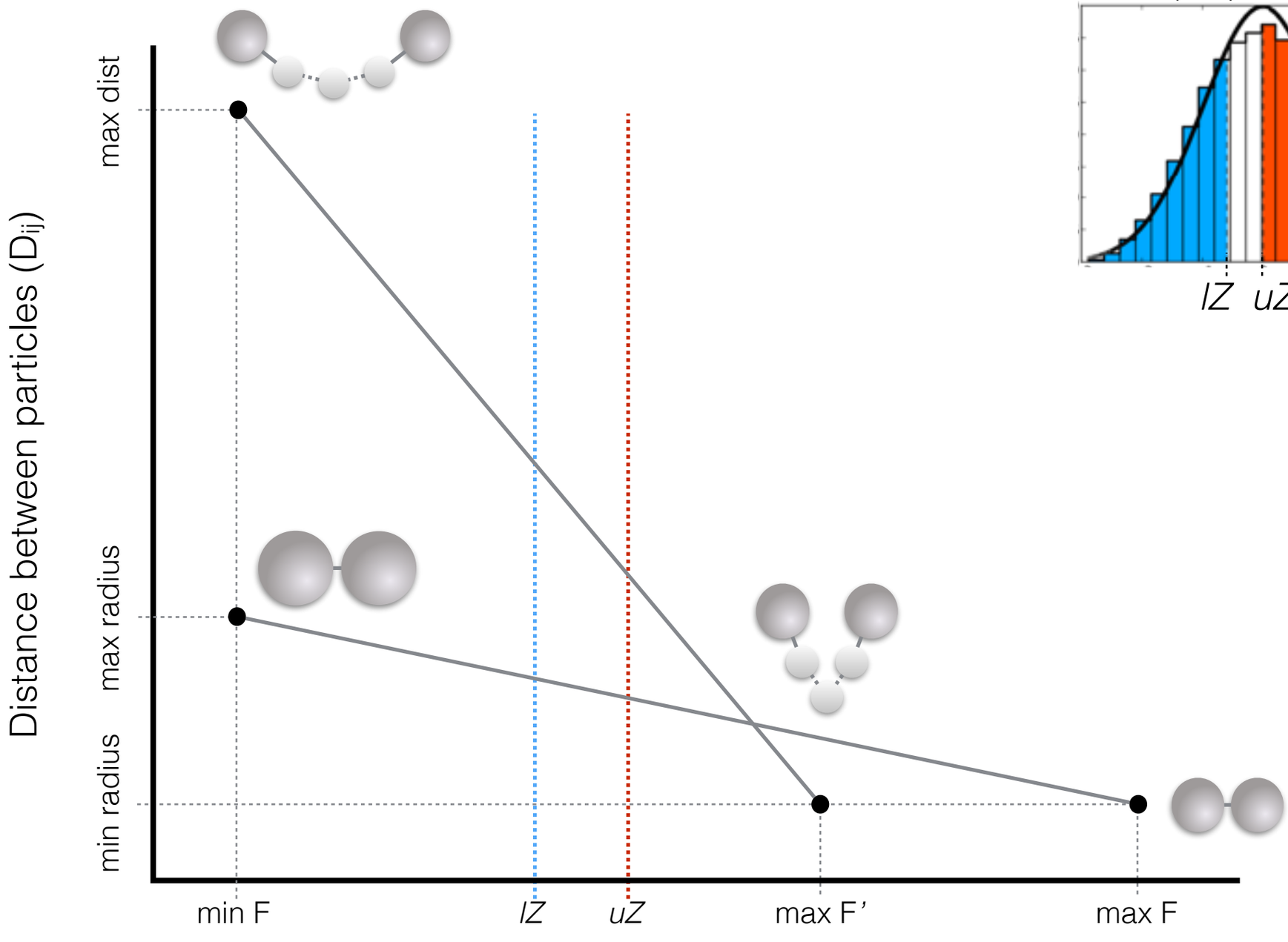

Baù D, Marti-Renom MA. 2012.
Genome structure determination via 3C-based data integration by the Integrative Modeling Platform.
Methods 58: 300–6.

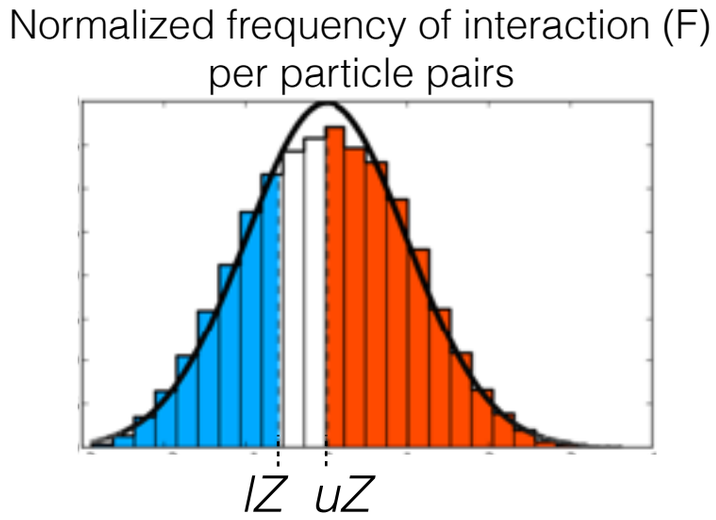
# From interactions to distance




Baù D, Marti-Renom MA. 2012.
Genome structure determination via 3C-based data integration by the Integrative Modeling Platform.
Methods 58: 300–6.

# From interactions to distance

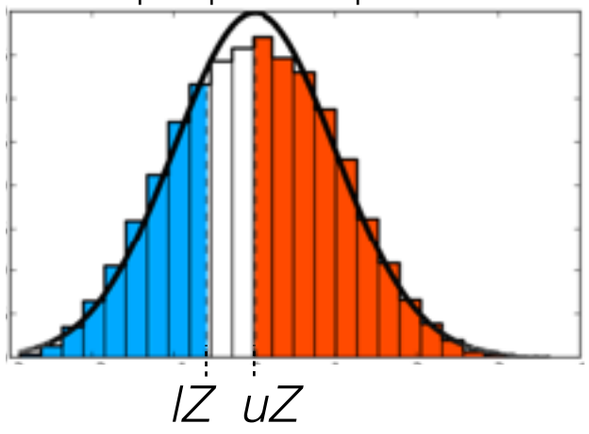


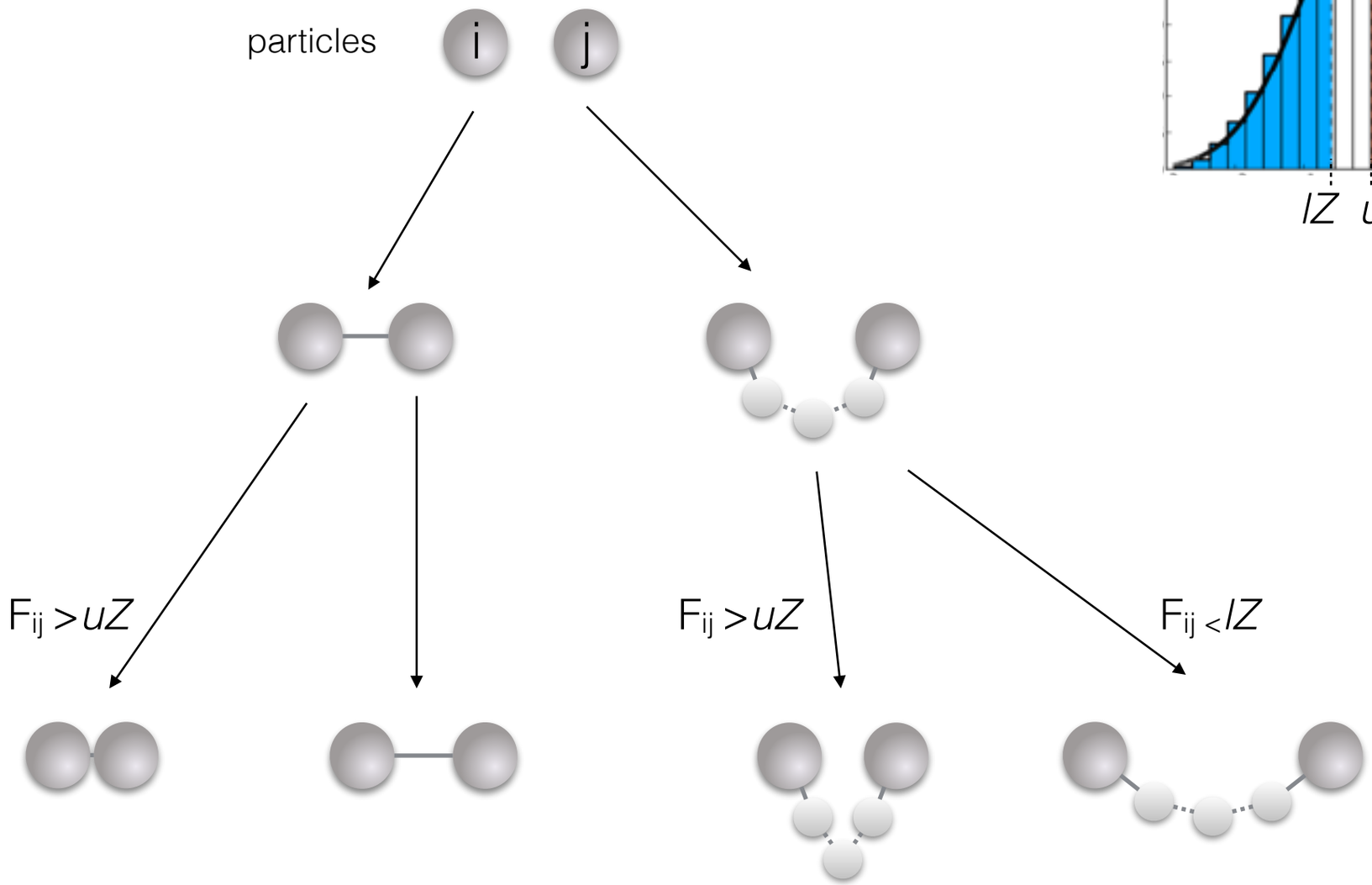




Baù D, Marti-Renom MA. 2012.
Genome structure determination via 3C-based data integration by the Integrative Modeling Platform.
Methods 58: 300–6.

# Restraints

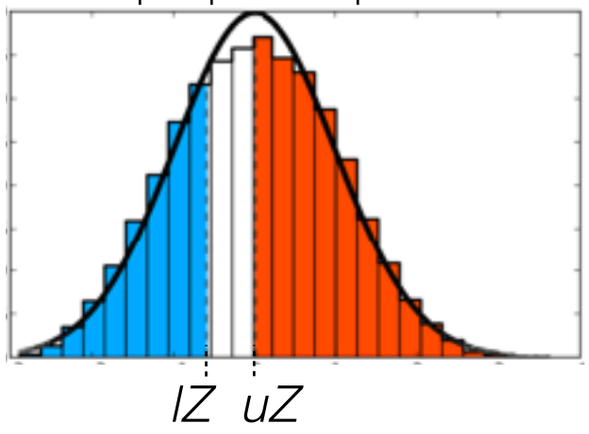
Normalized frequency of interaction (F)
per particle pairs

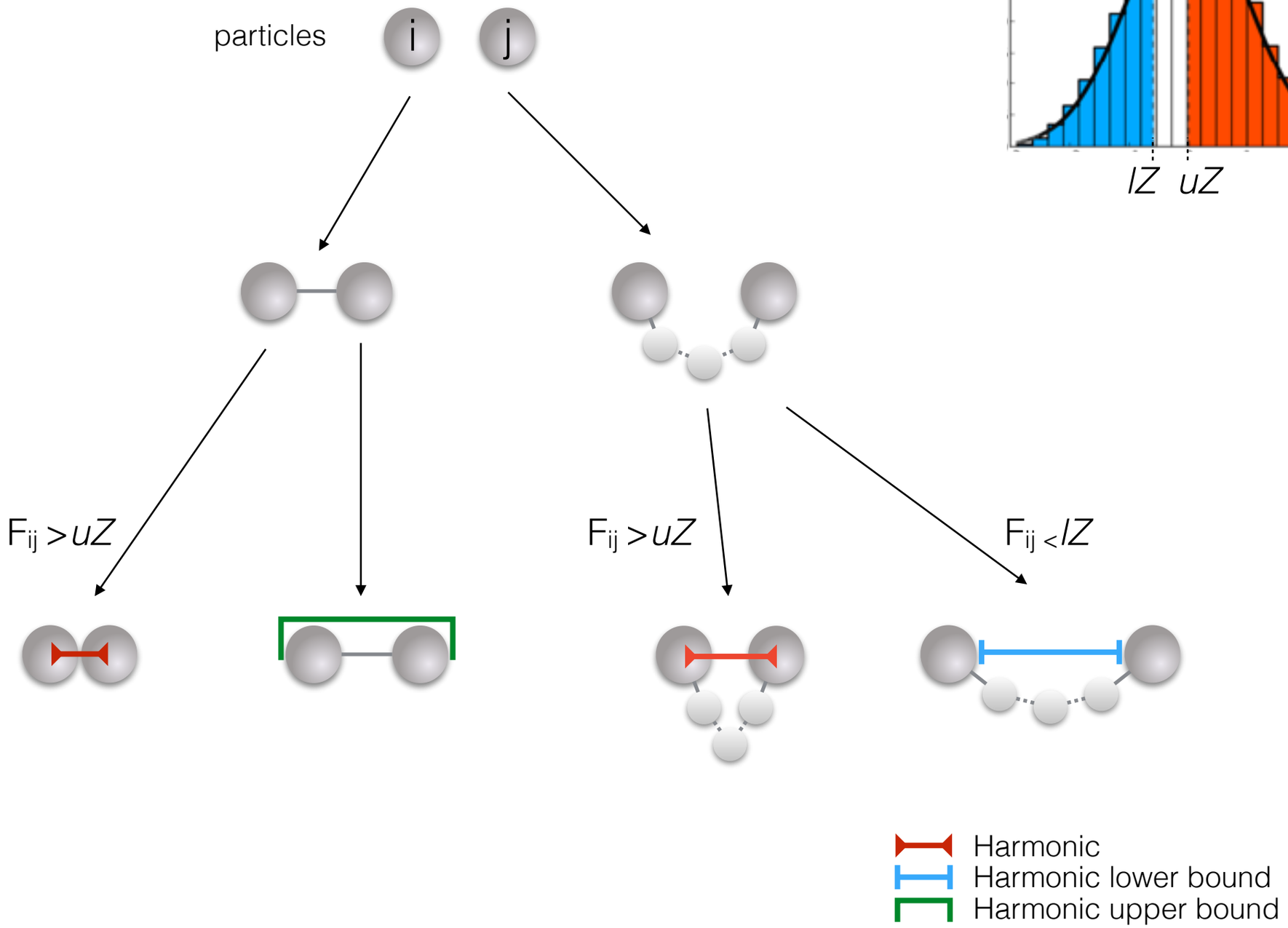


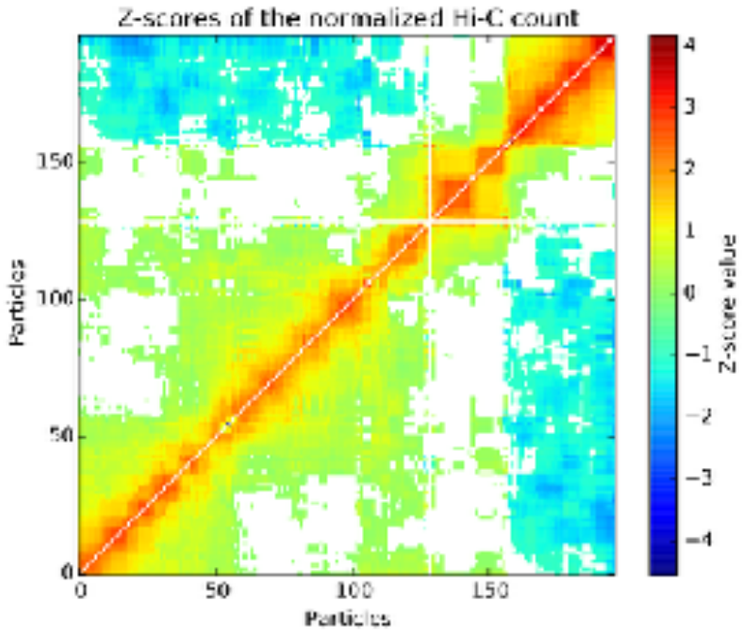
# Restraints

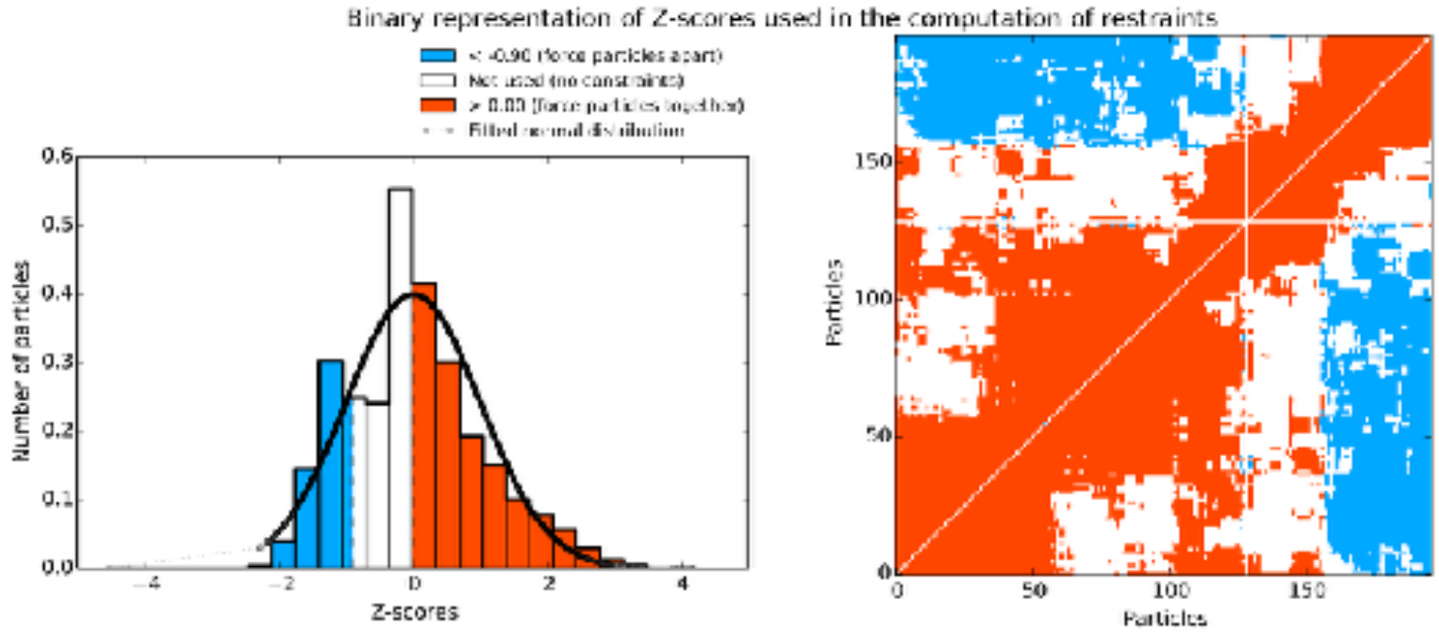
Normalized frequency of interaction (F)
per particle pairs



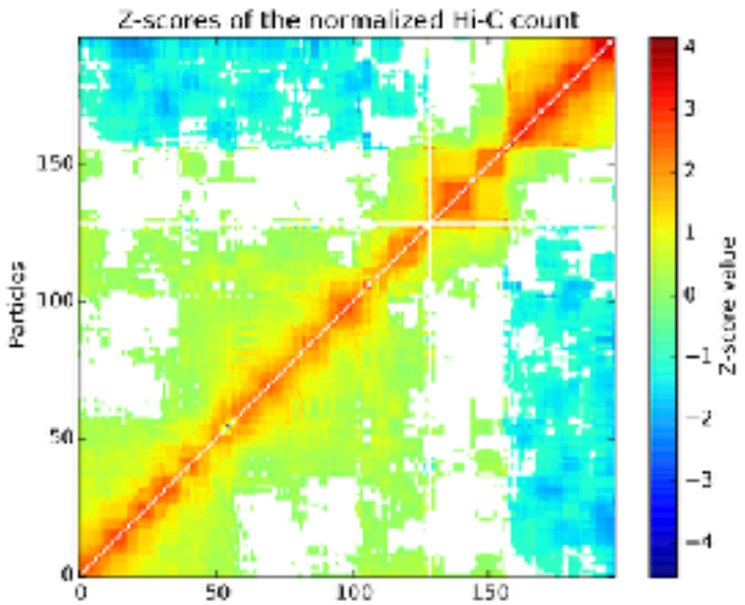



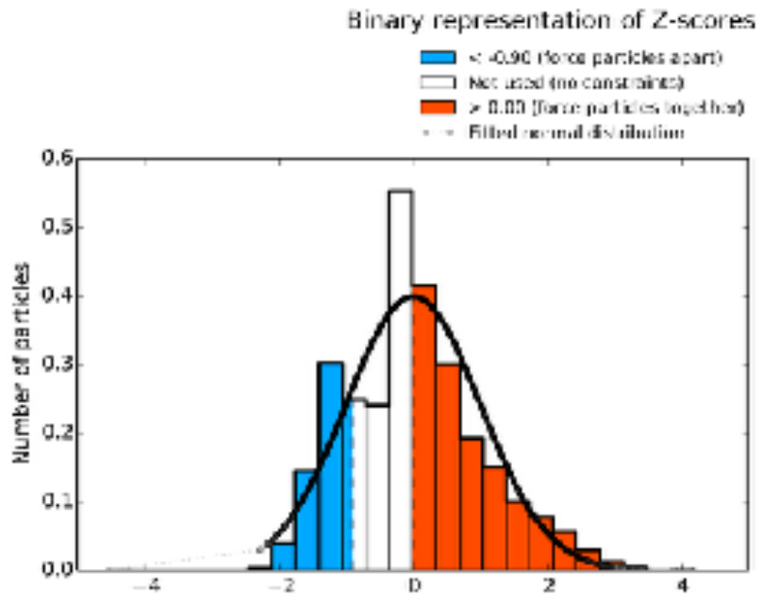
# Parameter optimisation

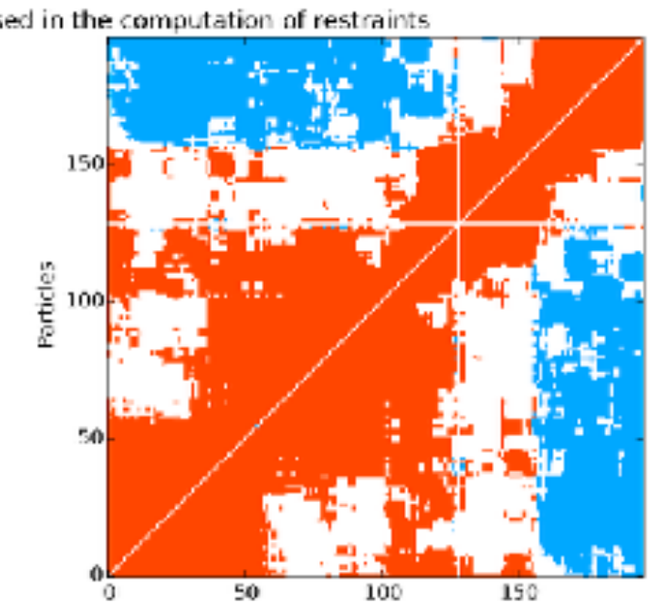


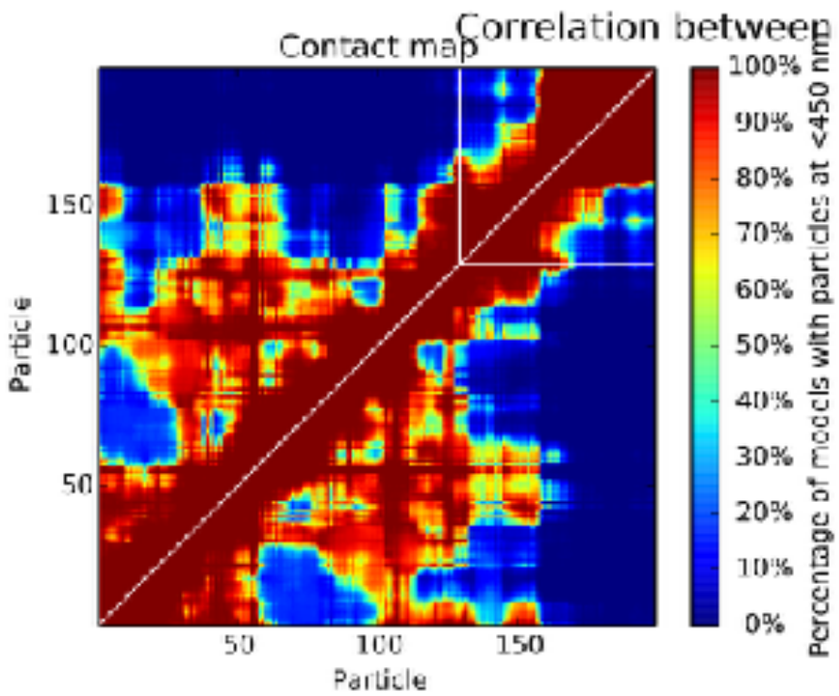



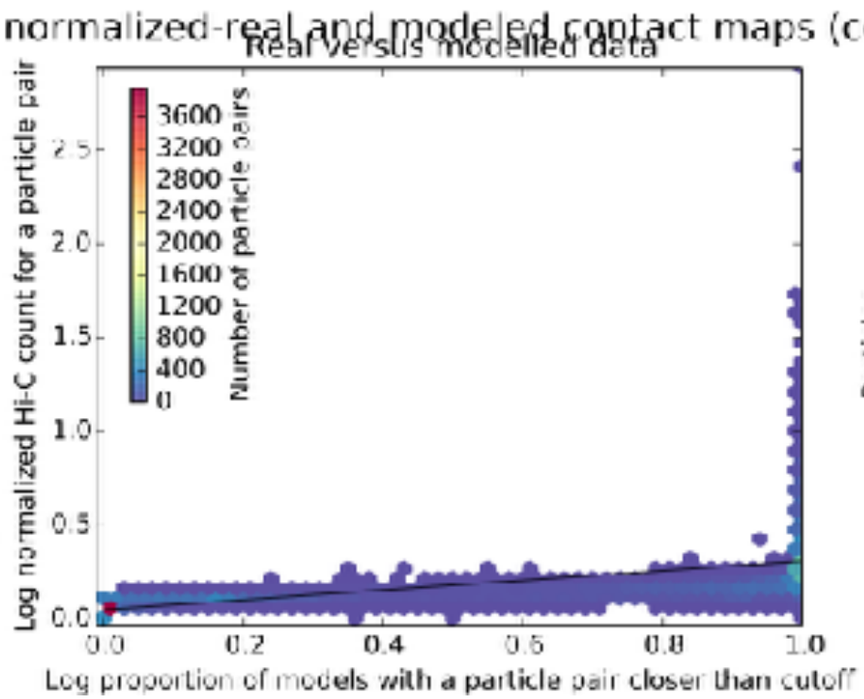
# Parameter optimisation

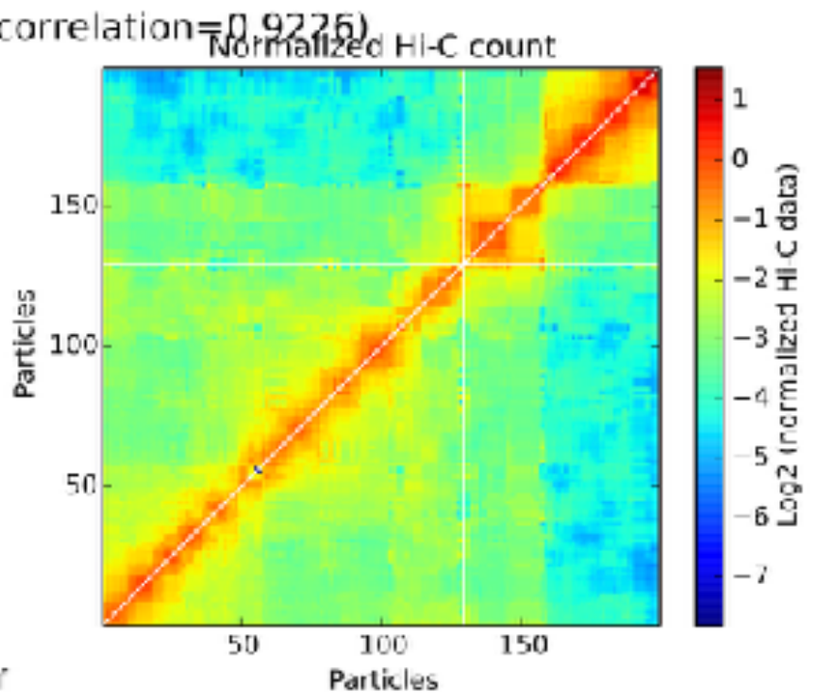


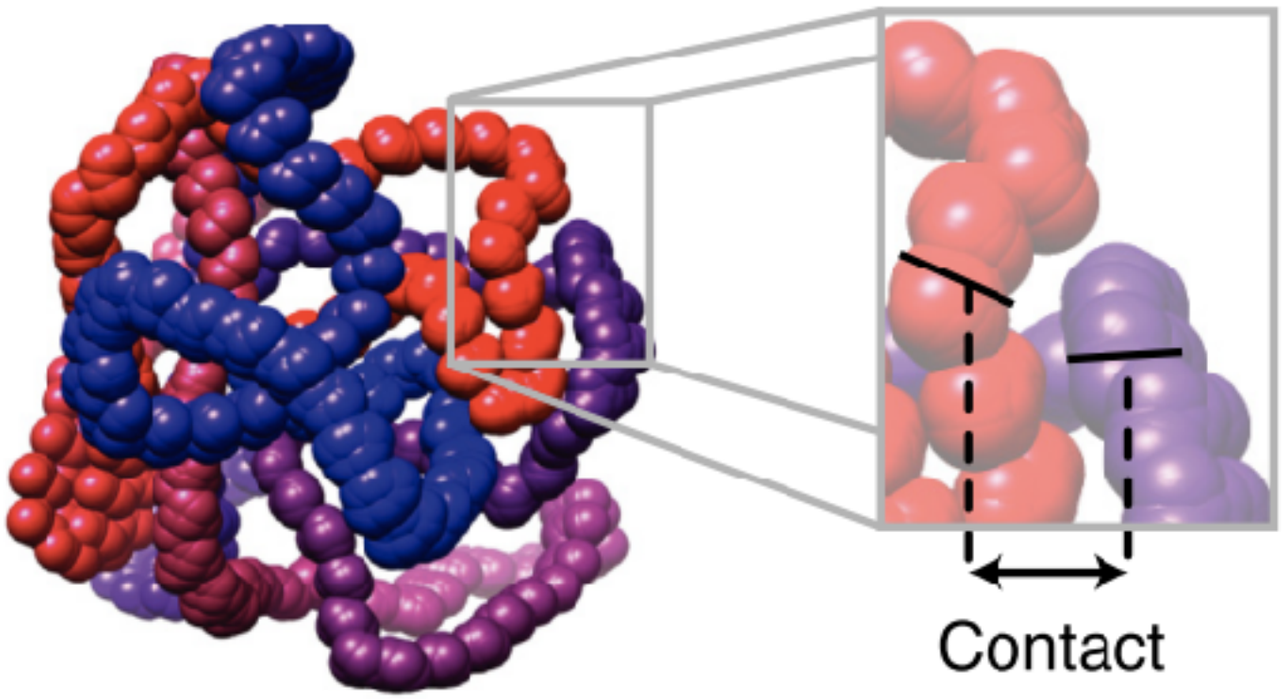



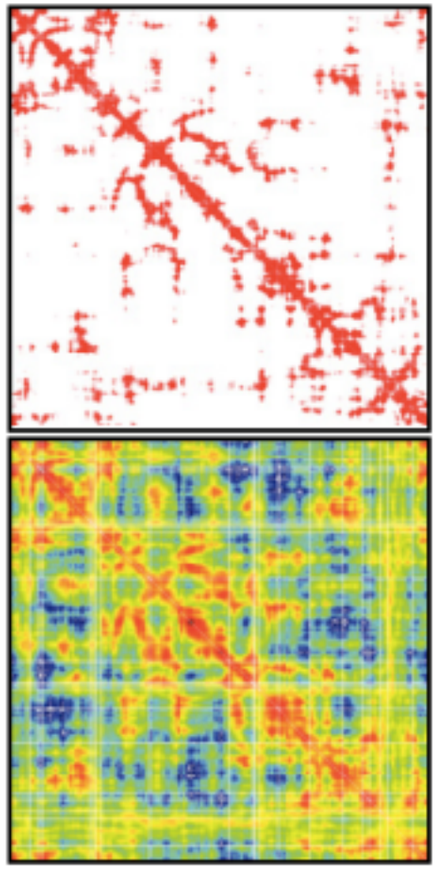




Correlation between normalized-real and modeled contact maps (correlation=0.9226)

# Parameter optimization



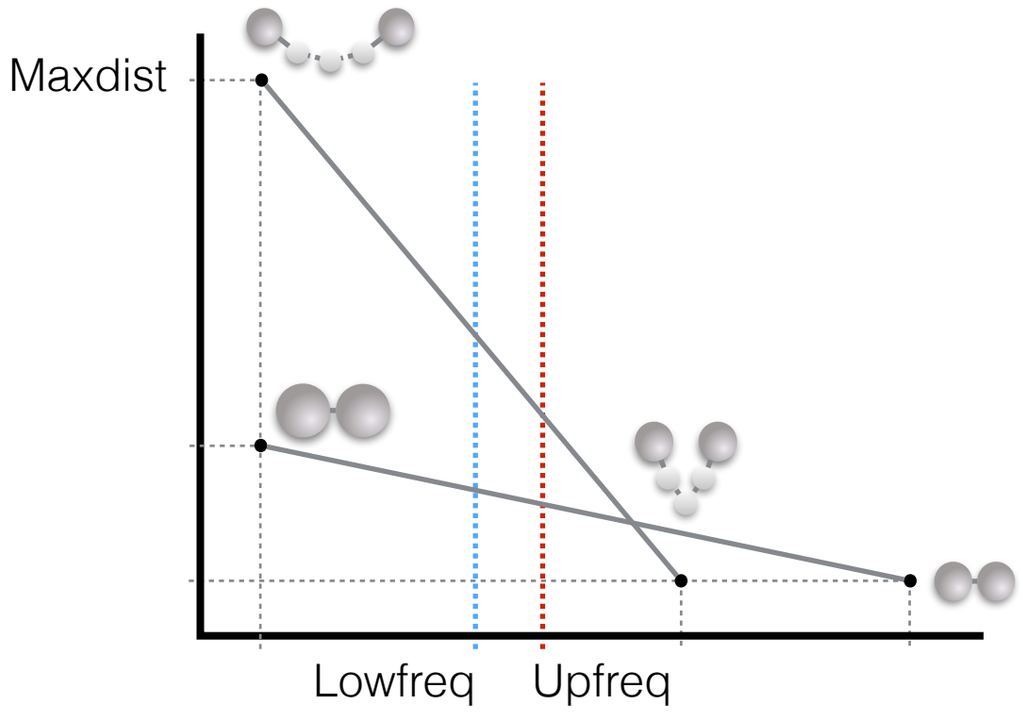
Contact matrix

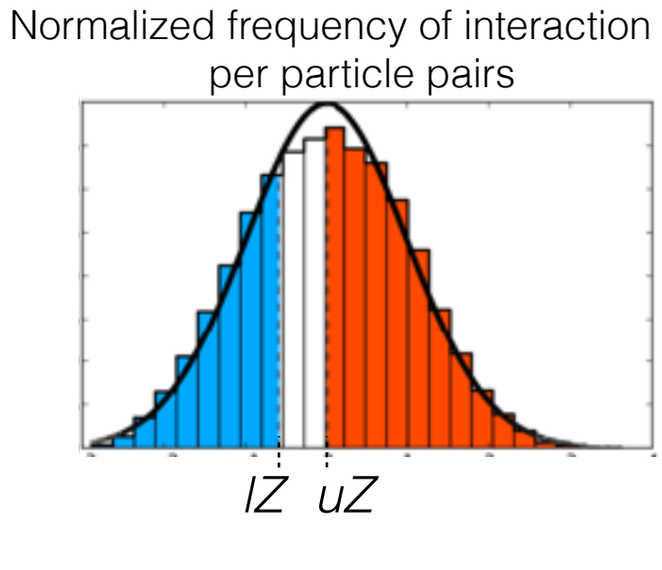
Hi-C data

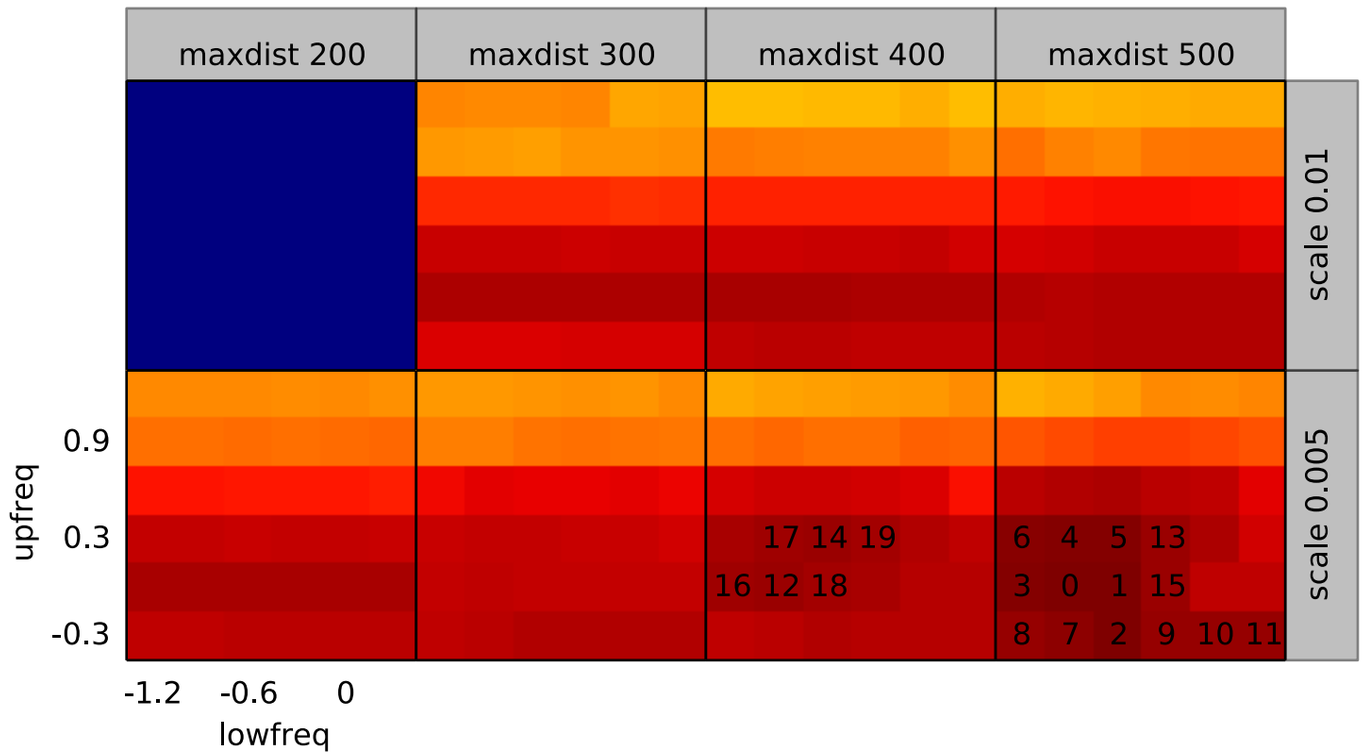
**Adapted from**

Trussart M, Serra F, Baù D, Junier I, Serrano L, Marti-Renom MA. 2015.
Assessing the limits of restraint-based 3D modeling of genomes and genomic domains.
Nucleic Acids Res 43: 3465–77

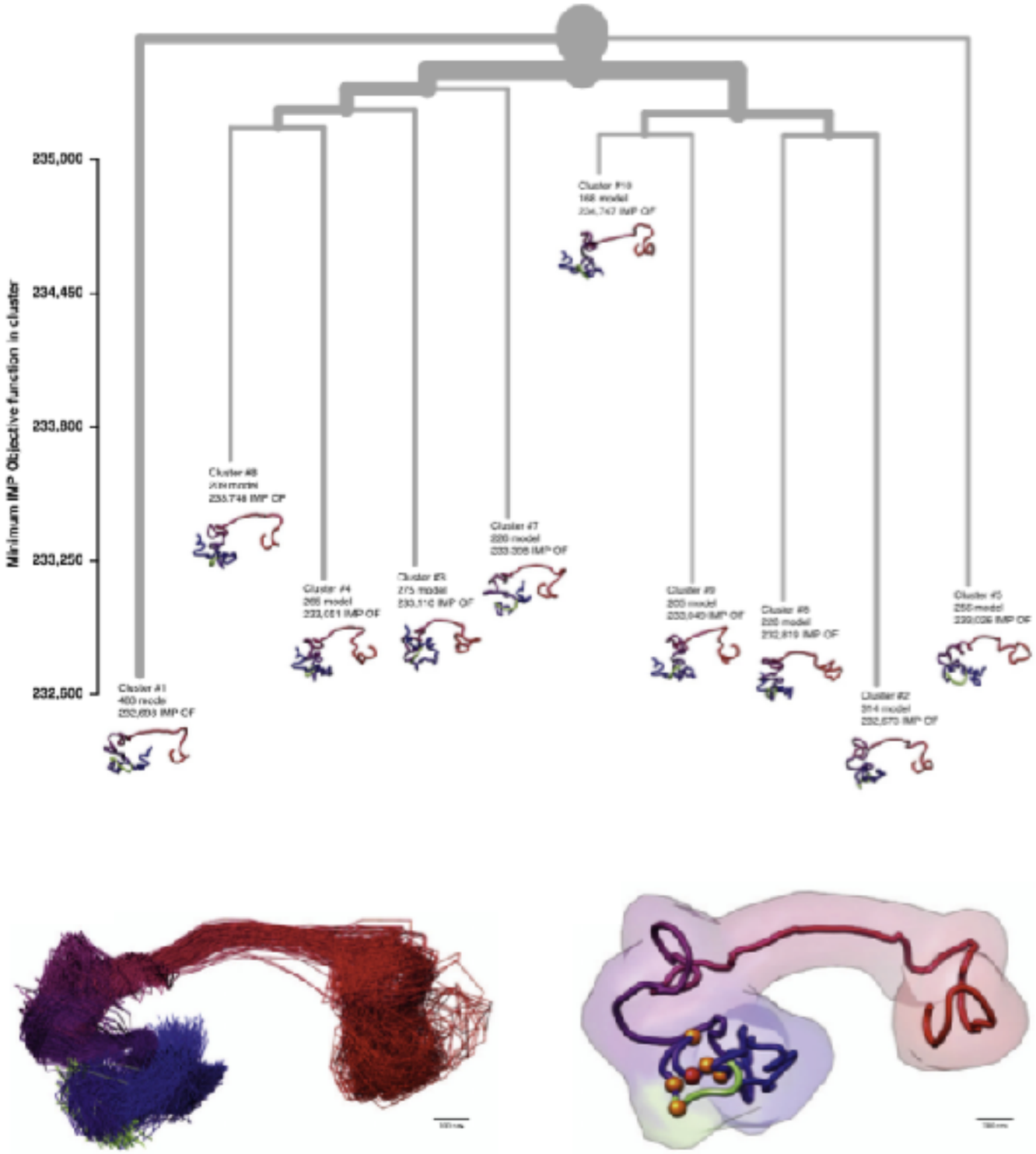
# Parameter optimization



# Model clustering



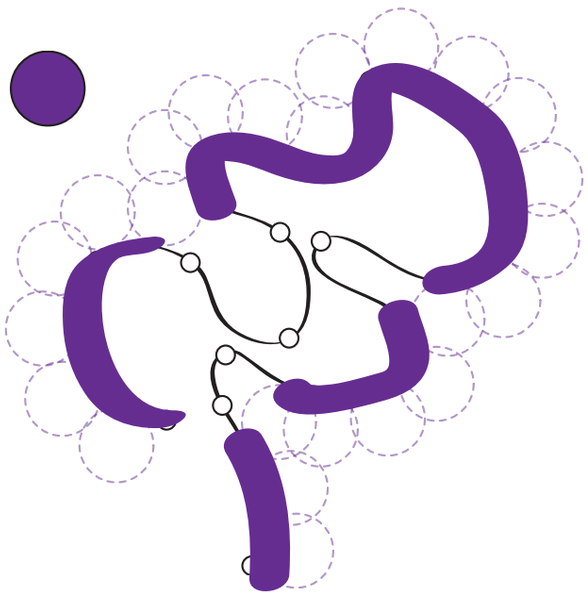

Baù D, Marti-Renom MA. 2012.
Genome structure determination via 3C-based data integration by the Integrative Modeling Platform.
Methods 58: 300–6.

# Population of models

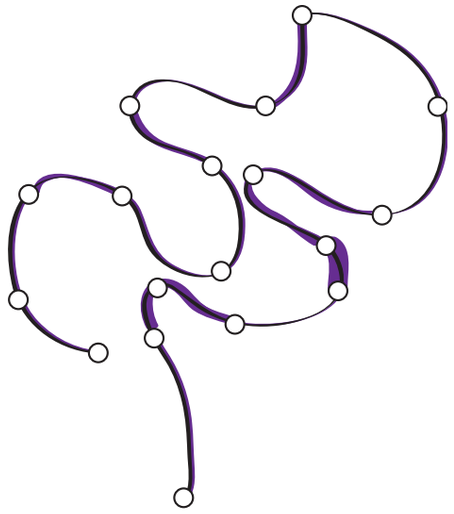
- Consensus model
  - single model
  - assumption of structural homogeneity (very fast)
- Resampling
  - independent model optimization
  - parameter optimization population based
  - assumption that structural heterogeneity is directly reflected in local minimums of the model optimization
- Population
  - models optimized in parallel
  - risk of overfitting, computationally expensive

# Model analysis

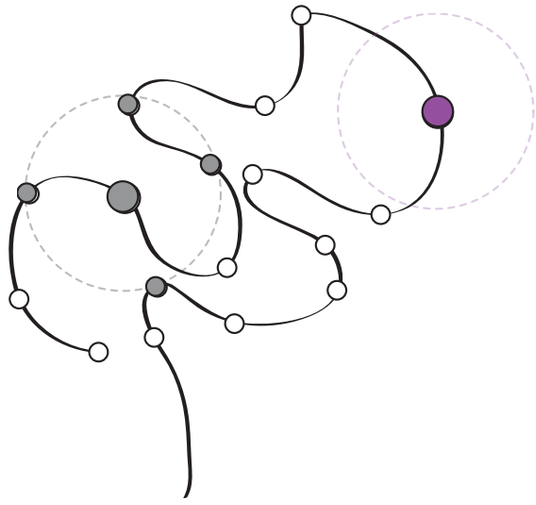
Accessibility (%)



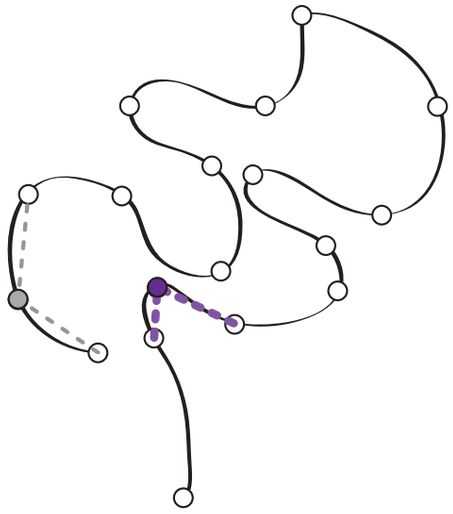
Density (bp/nm)



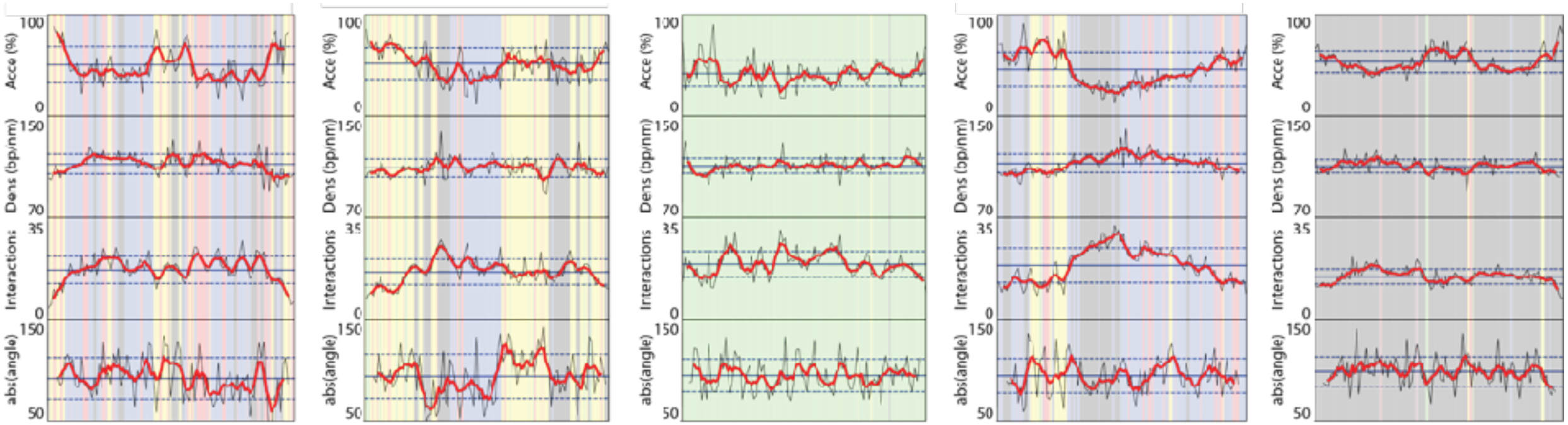
Interactions

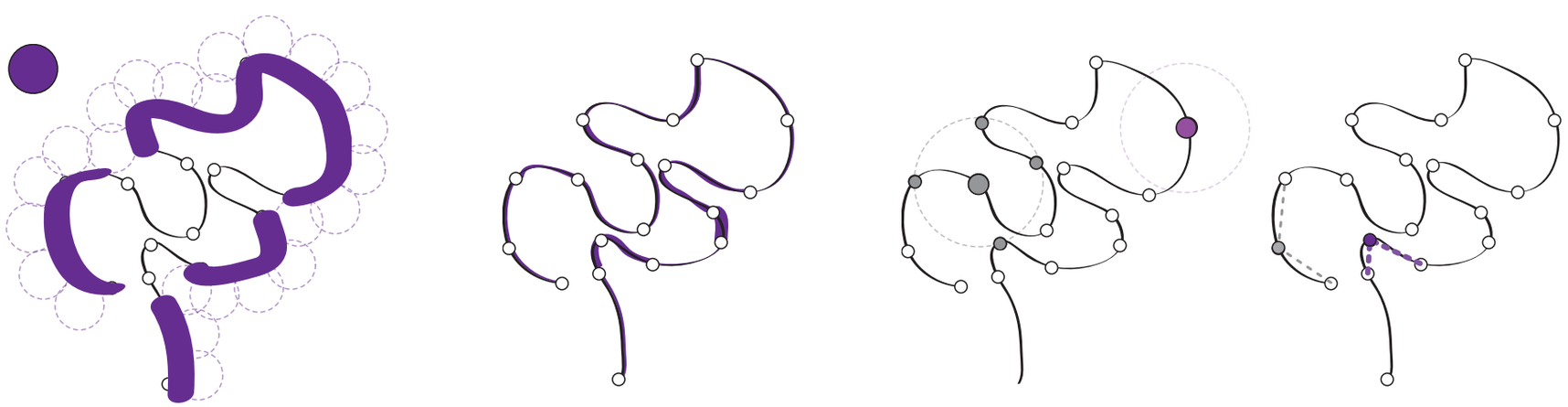


Angle

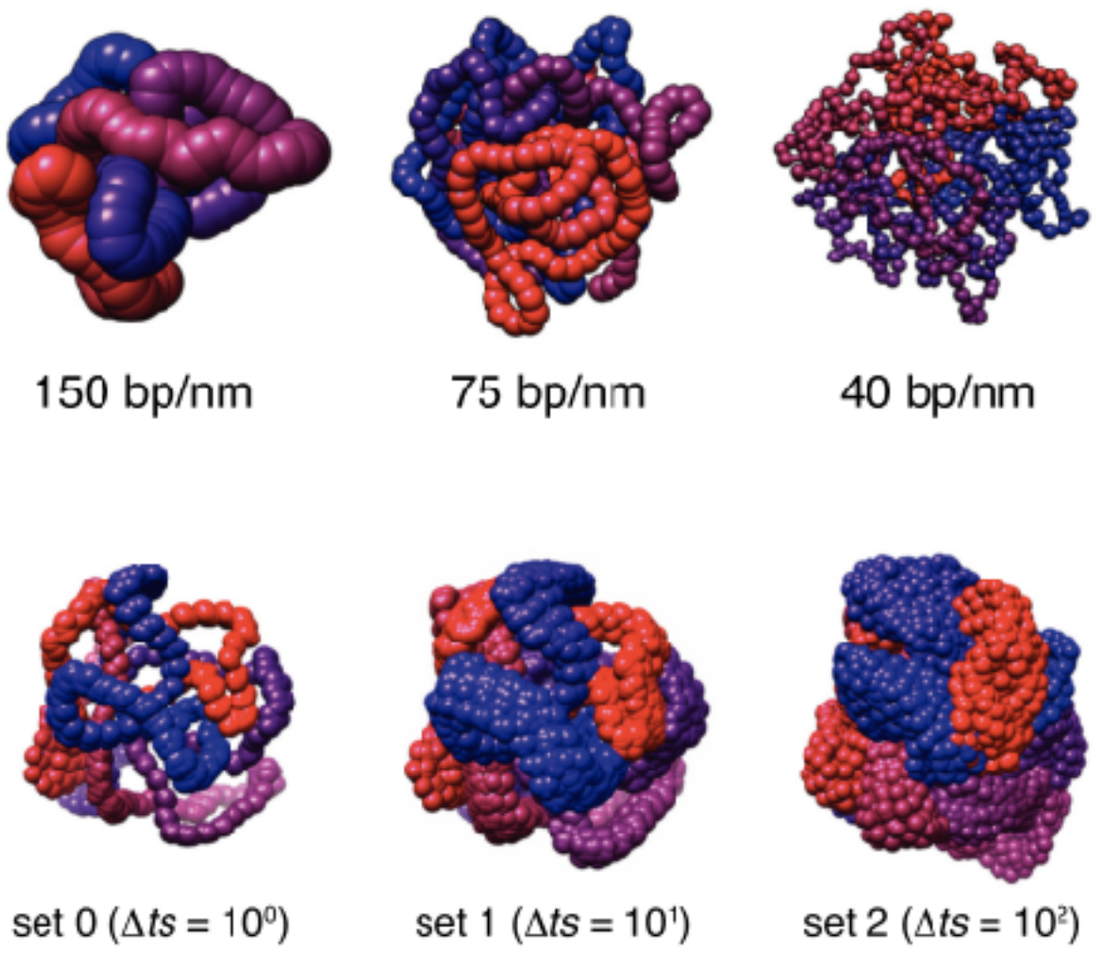


# Model analysis: structural features



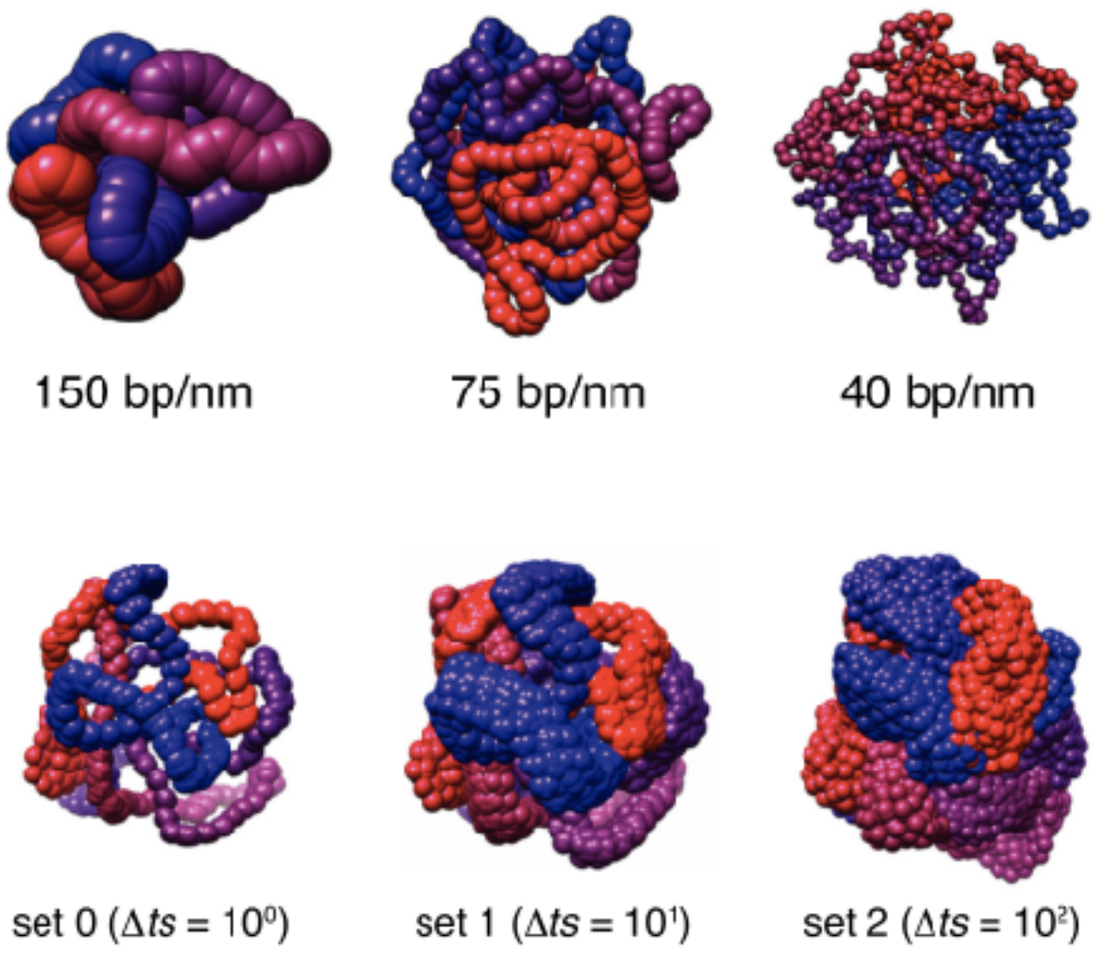



# Modeling potential




Trussart M, Serra F, Baù D, Junier I, Serrano L, Marti-Renom MA. 2015.
Assessing the limits of restraint-based 3D modeling of genomes and genomic domains.
Nucleic Acids Res 43: 3465–77

# Modeling potential

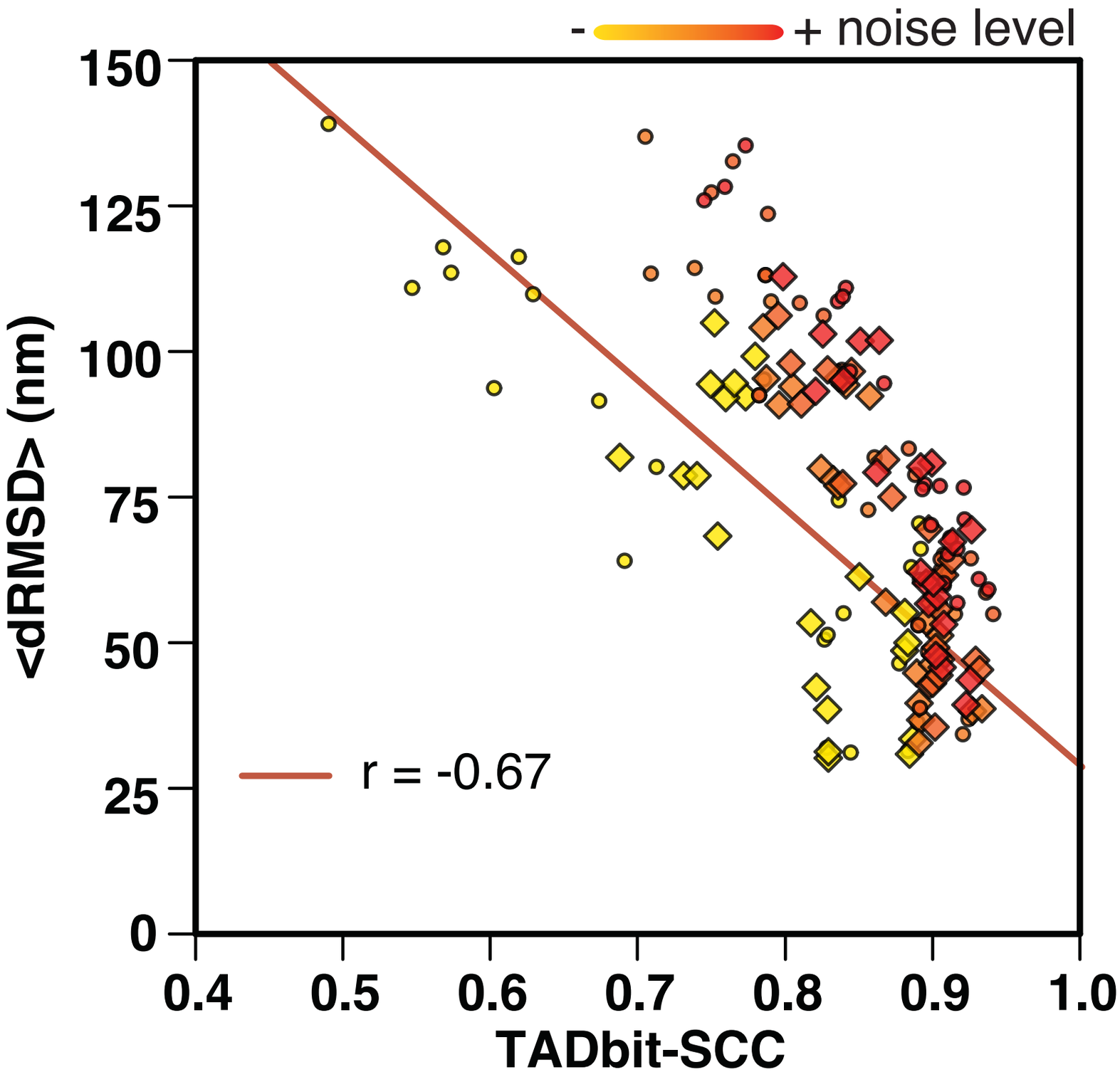



Study the effect of:

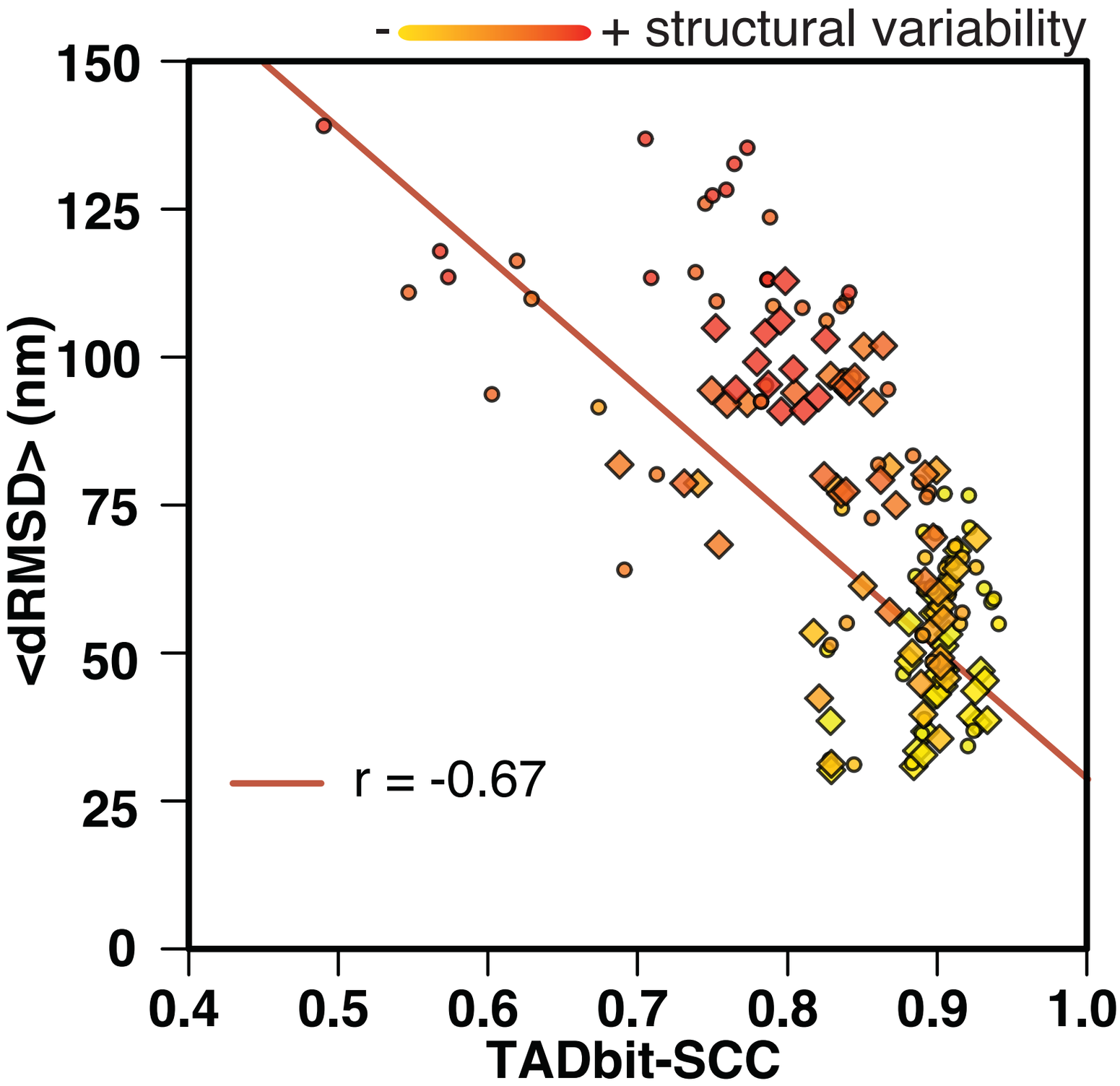
- **structural variability**
- **noise**
- number of particles
- presence of TADs

Trussart M, Serra F, Baù D, Junier I, Serrano L, Marti-Renom MA. 2015.
Assessing the limits of restraint-based 3D modeling of genomes and genomic domains.
Nucleic Acids Res 43: 3465–77

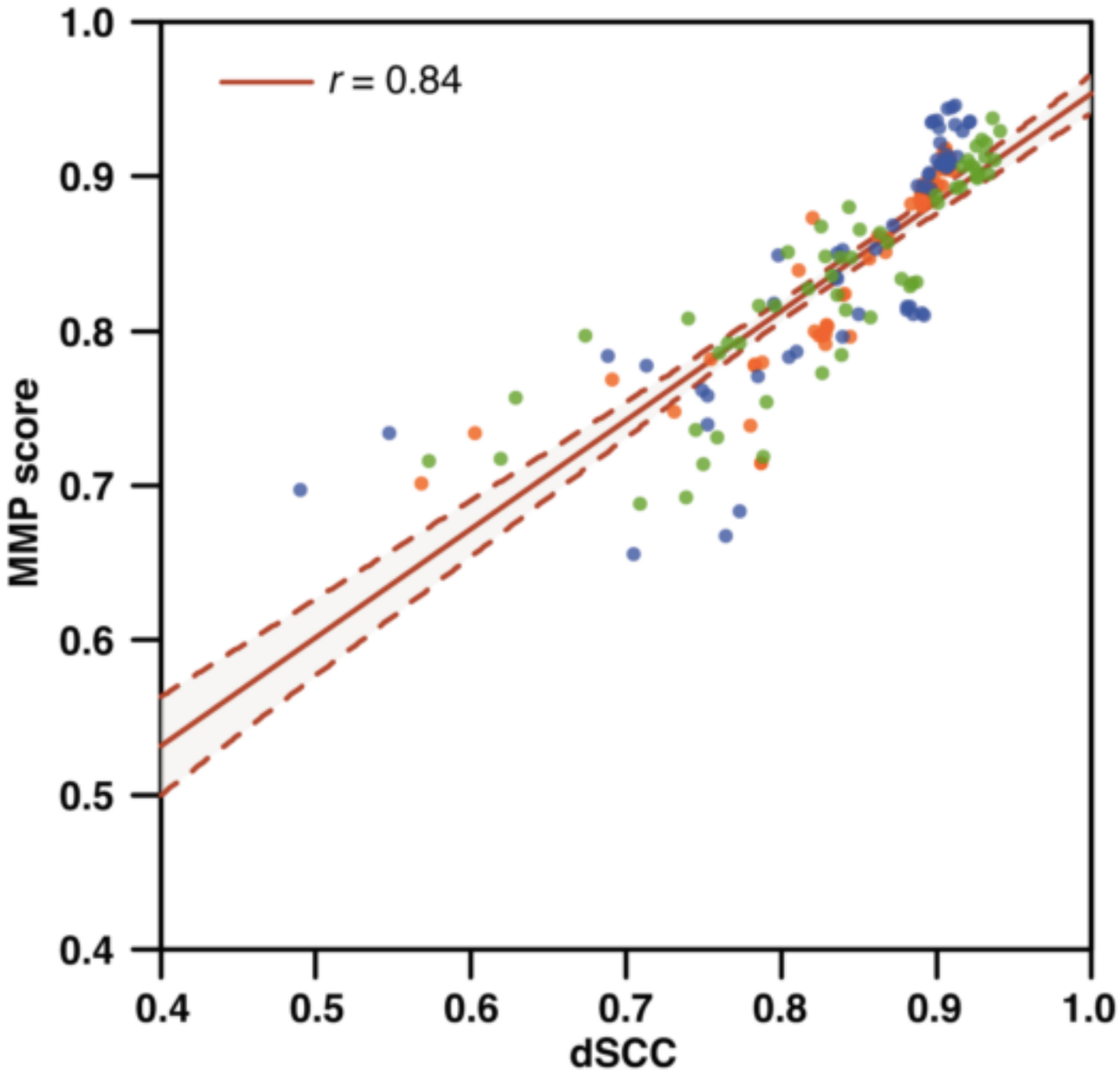
# Noise is "OK"


- + noise level
150
125
100
75
50
25
0
<dRMSD> (nm)
r = -0.67
0.4
0.5
0.6
0.7
0.8
0.9
1.0
TADbit-SCC

# Structural variability is "NOT OK"

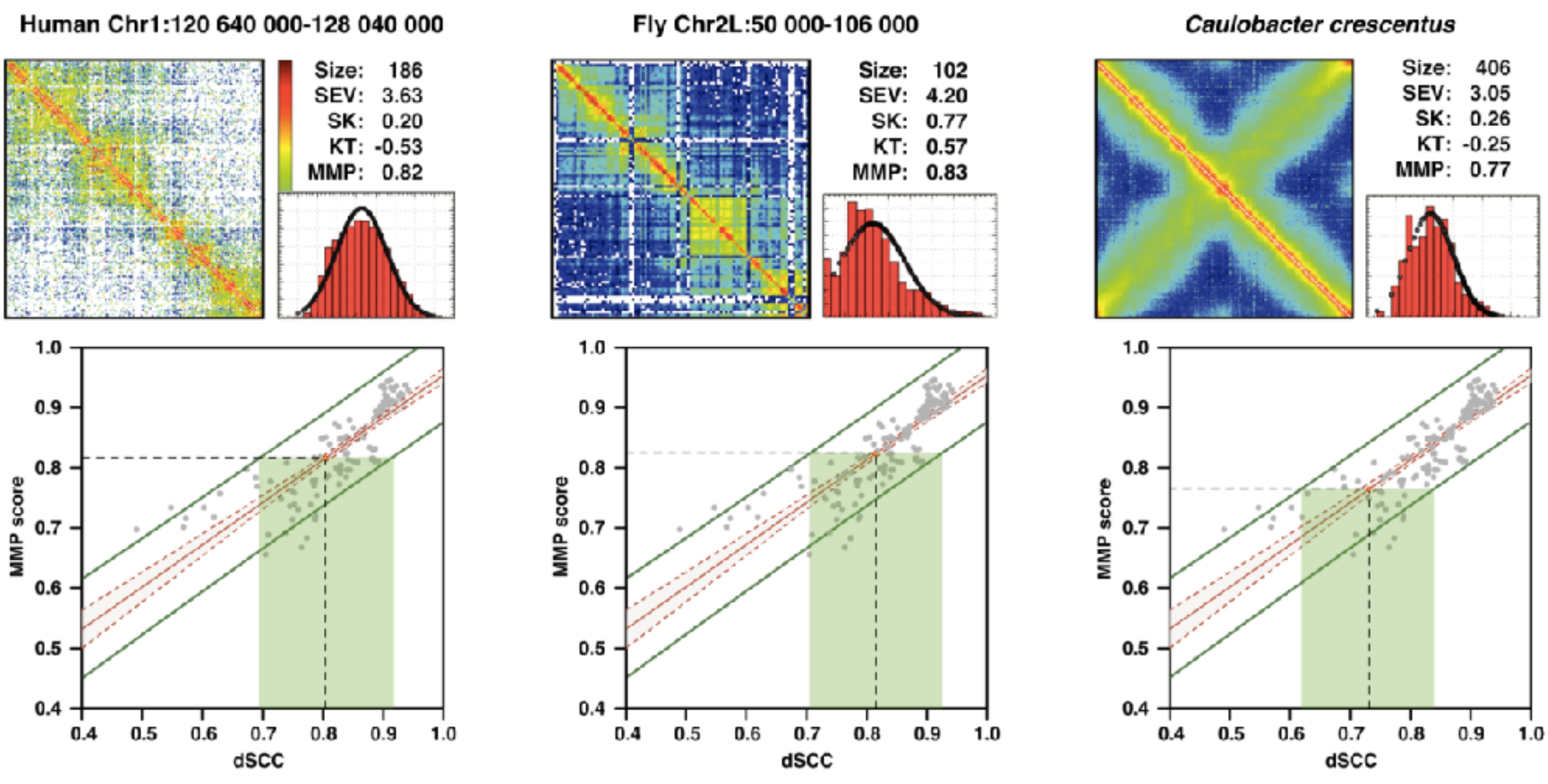


# Modeling potential




Trussart M, Serra F, Baù D, Junier I, Serrano L, Marti-Renom MA. 2015.
Assessing the limits of restraint-based 3D modeling of genomes and genomic domains.
Nucleic Acids Res 43: 3465–77

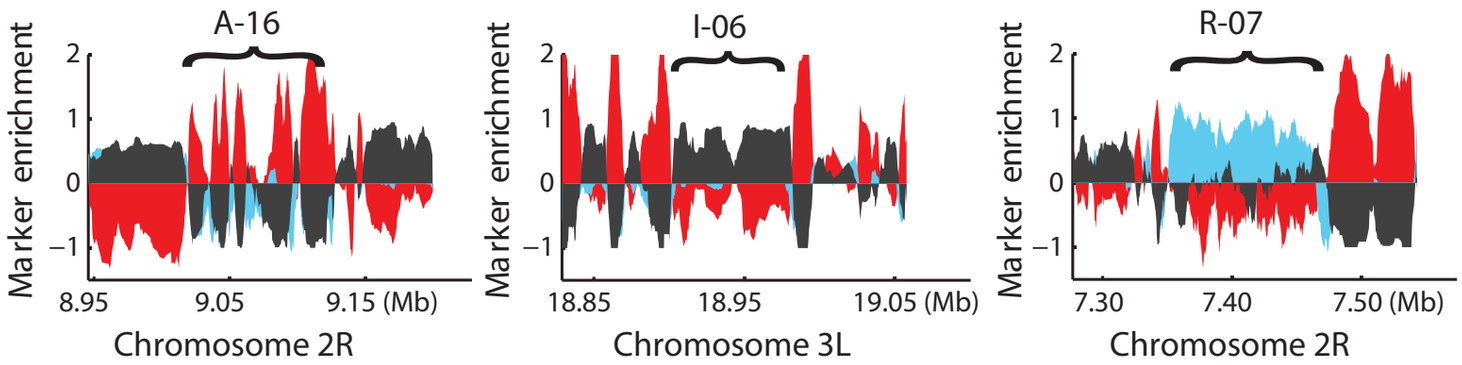
# Modeling potential

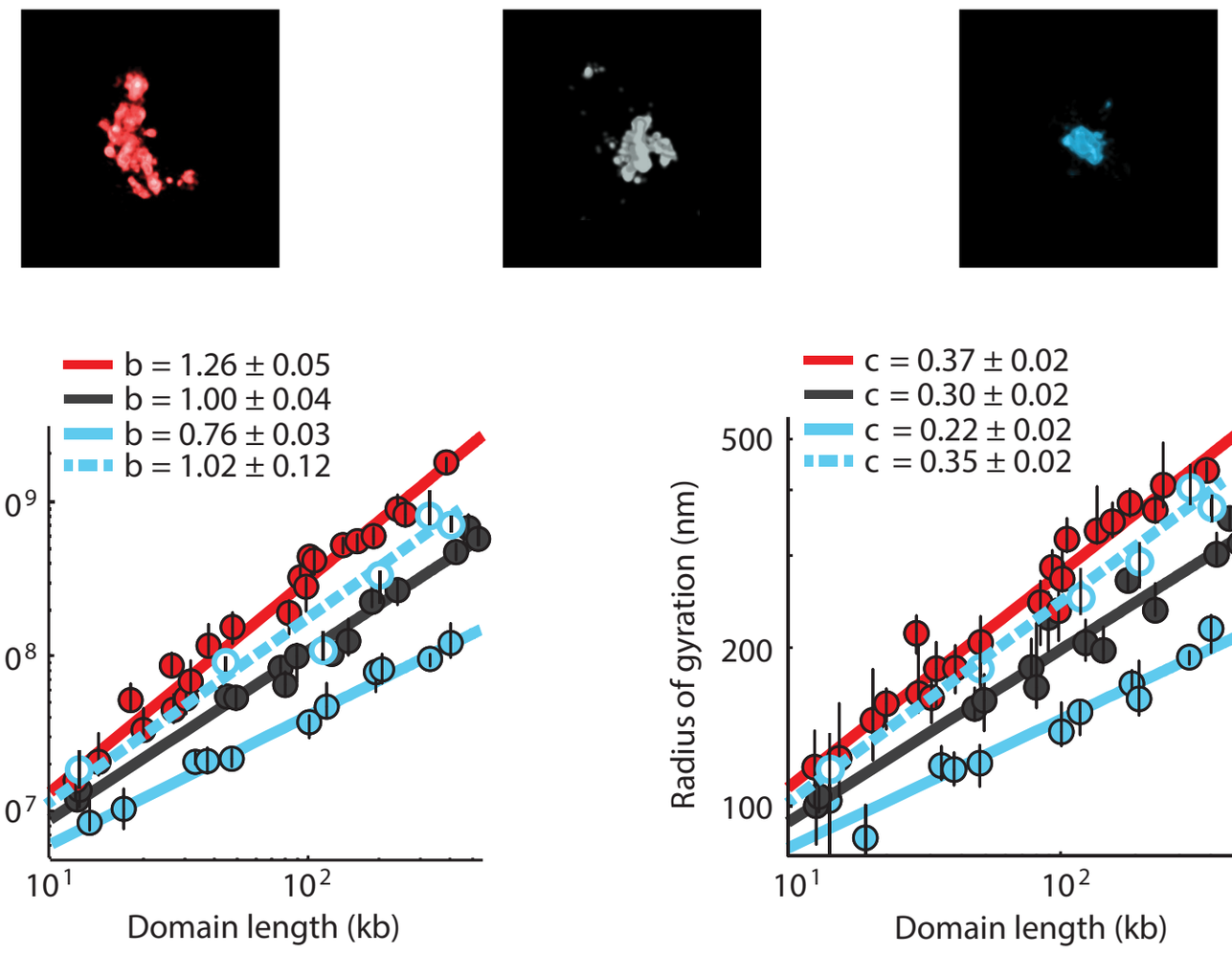



Trussart M, Serra F, Baù D, Junier I, Serrano L, Marti-Renom MA. 2015.
Assessing the limits of restraint-based 3D modeling of genomes and genomic domains.
Nucleic Acids Res 43: 3465–77

# Model accuracy

Boettiger, A. N., et al. (2016). Nature, 529, 418–422.



# Model accuracy (fly@2Kb)

Boettiger, A. N., et al. (2016). Nature, 529, 418–422.