



Multiscale Complex Genomics

Coarse-grained DNA and chromatin model

Jürgen Walther – 14.10.2016



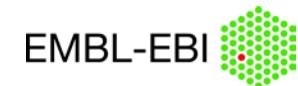
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.



INSTITUTE
FOR RESEARCH
IN BIOMEDICINE



UNITED KINGDOM • CHINA • MALAYSIA





Outline

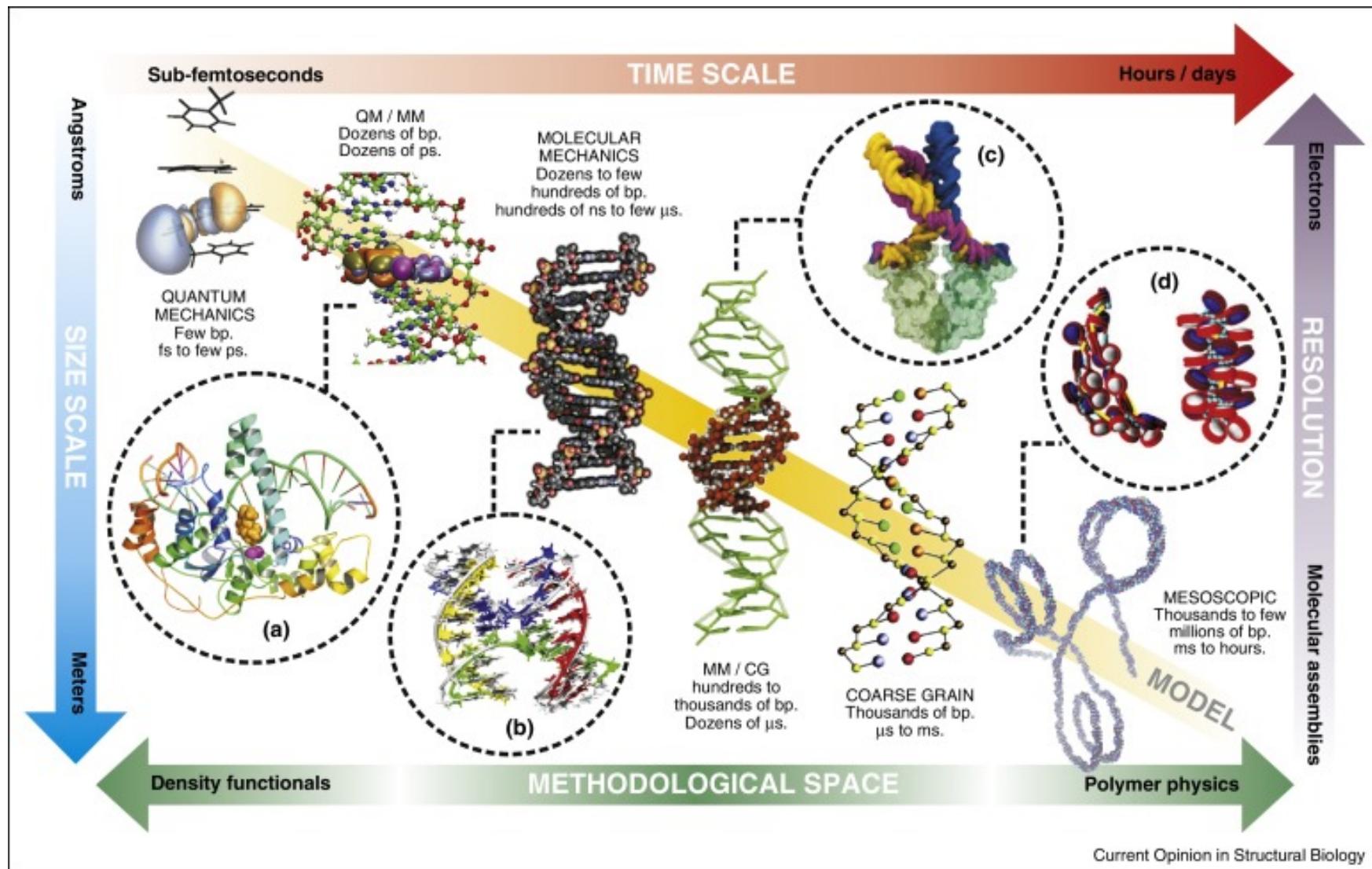
1) Introduction into DNA modeling

2) DNA model

3) Chromatin model

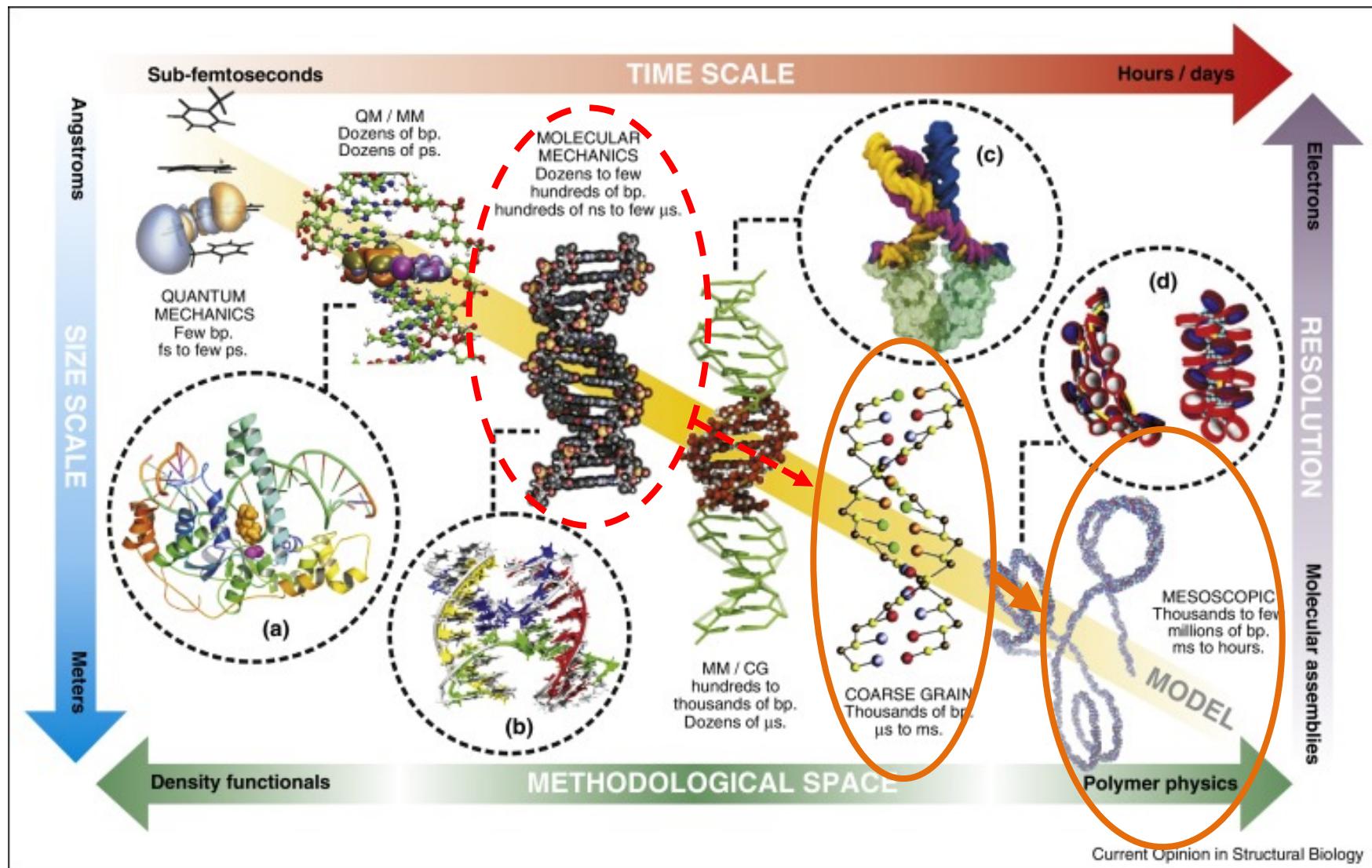
Multiscale simulation of DNA

Pablo D Dans^{1,2}, Jürgen Walther^{1,2}, Hansel Gómez^{1,2} and
Modesto Orozco^{1,2,3}



Multiscale simulation of DNA

Pablo D Dans^{1,2}, Jürgen Walther^{1,2}, Hansel Gómez^{1,2} and
Modesto Orozco^{1,2,3}





Multiscale Complex Genomics

Coarse-graining DNA

Describing movement of DNA with 6 parameters



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.



INSTITUTE
FOR RESEARCH
IN BIOMEDICINE



Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación



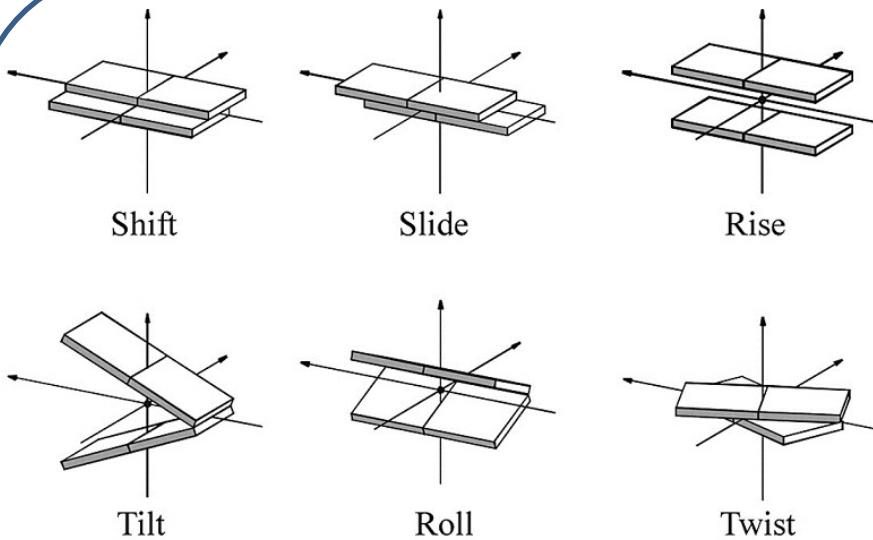
CRG[®]
Centra
for Genomic
Regulation



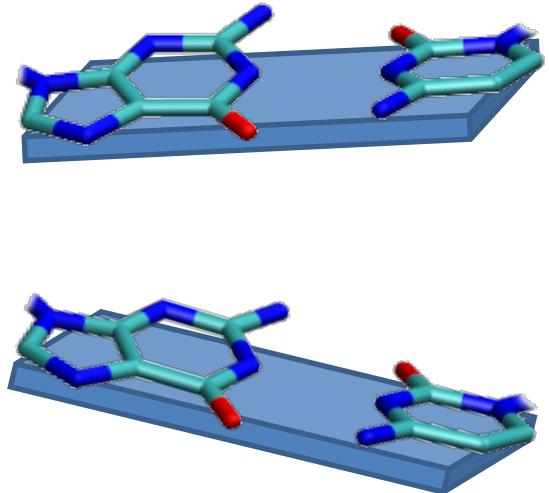
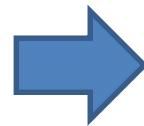
UNITED KINGDOM • CHINA • MALAYSIA



Coarse-graining the DNA

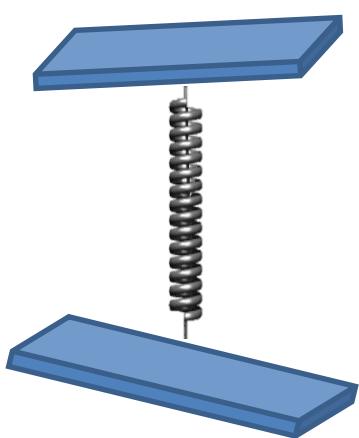


Base-pair step parameters



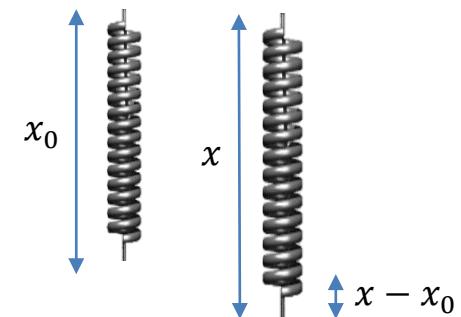
**Relative orientation and
position between adjacent
base-pairs**

DNA is a spiral staircase

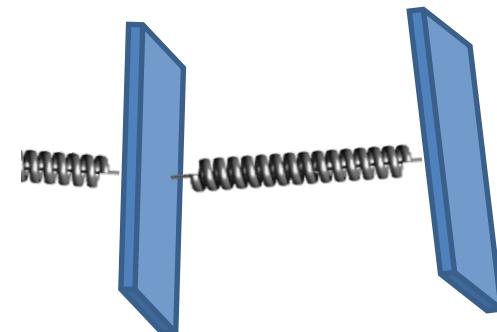
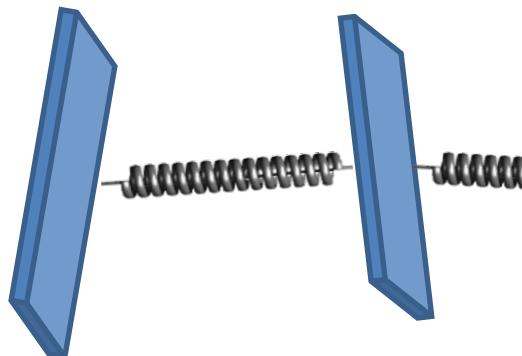


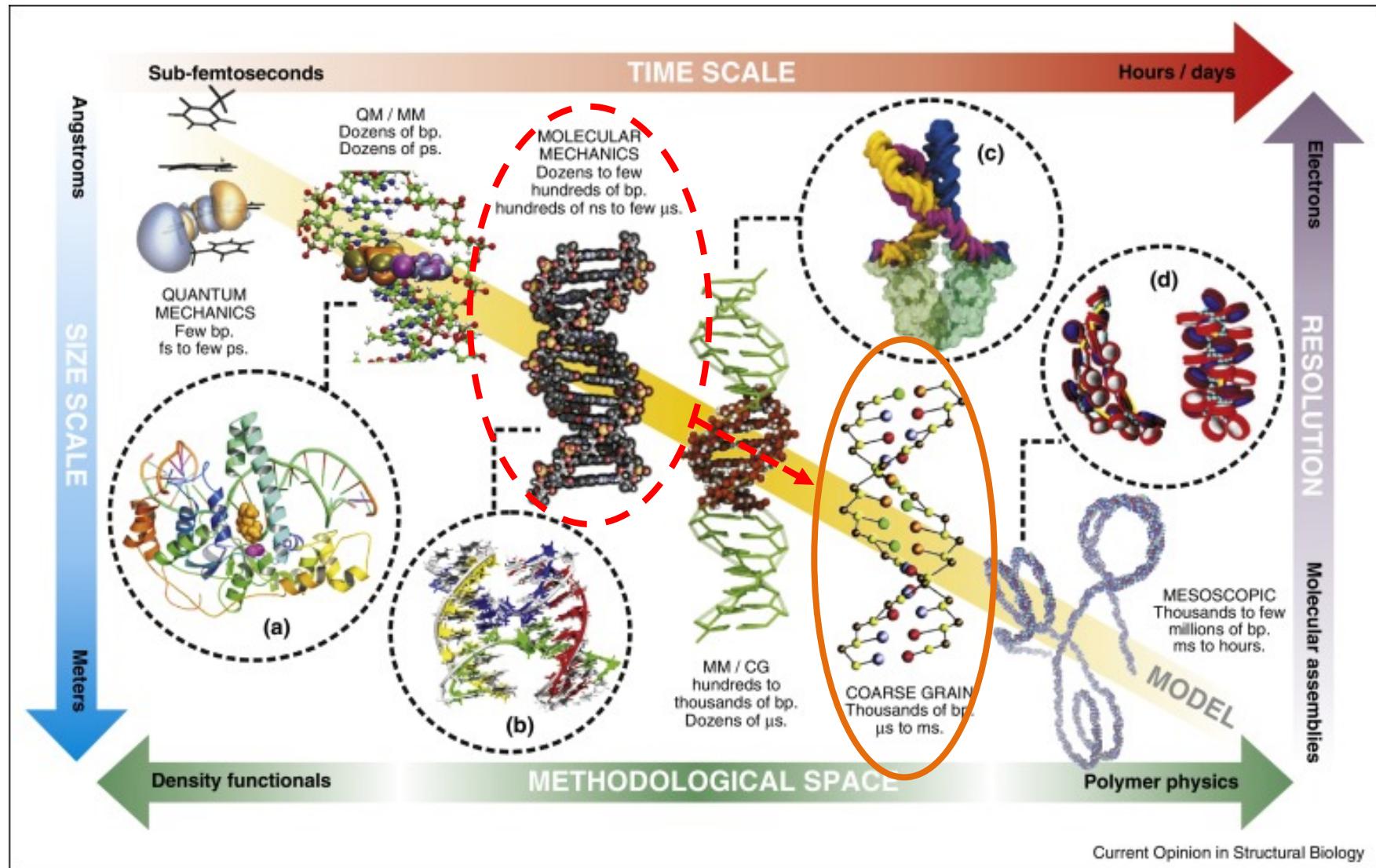
Base-pair

Energy of a spring

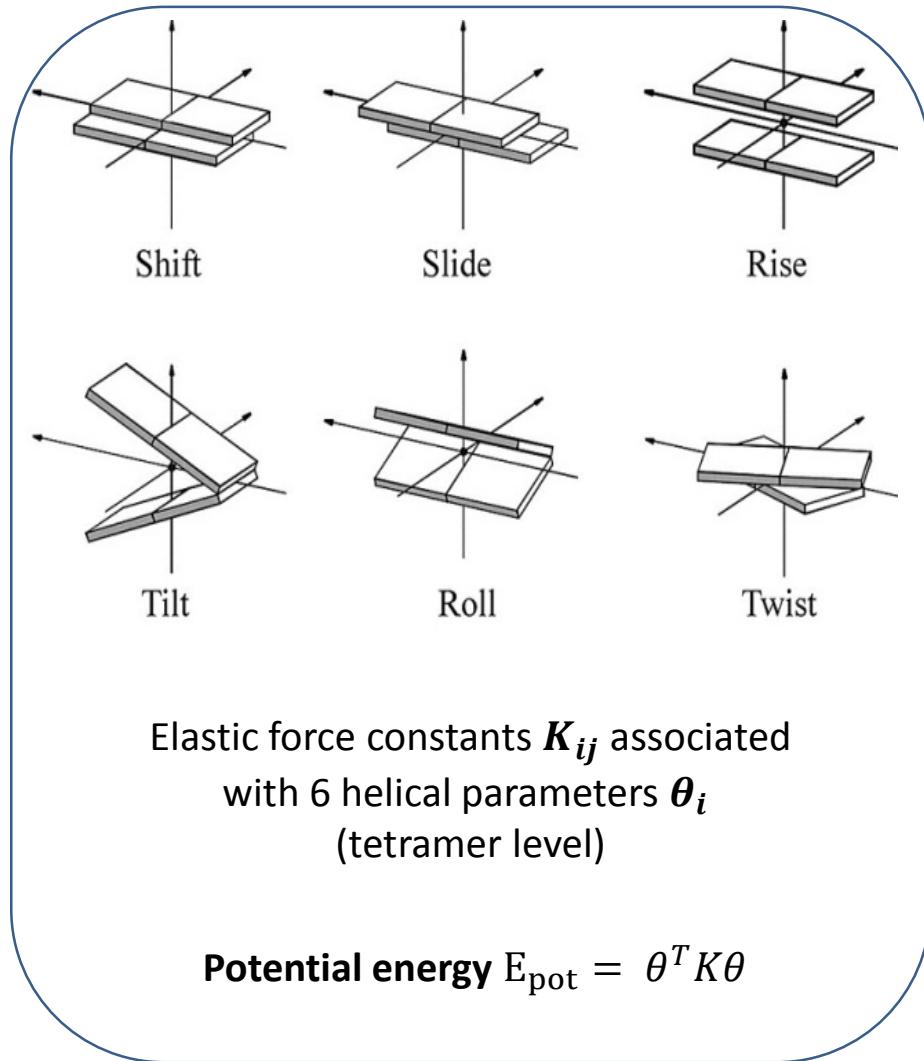
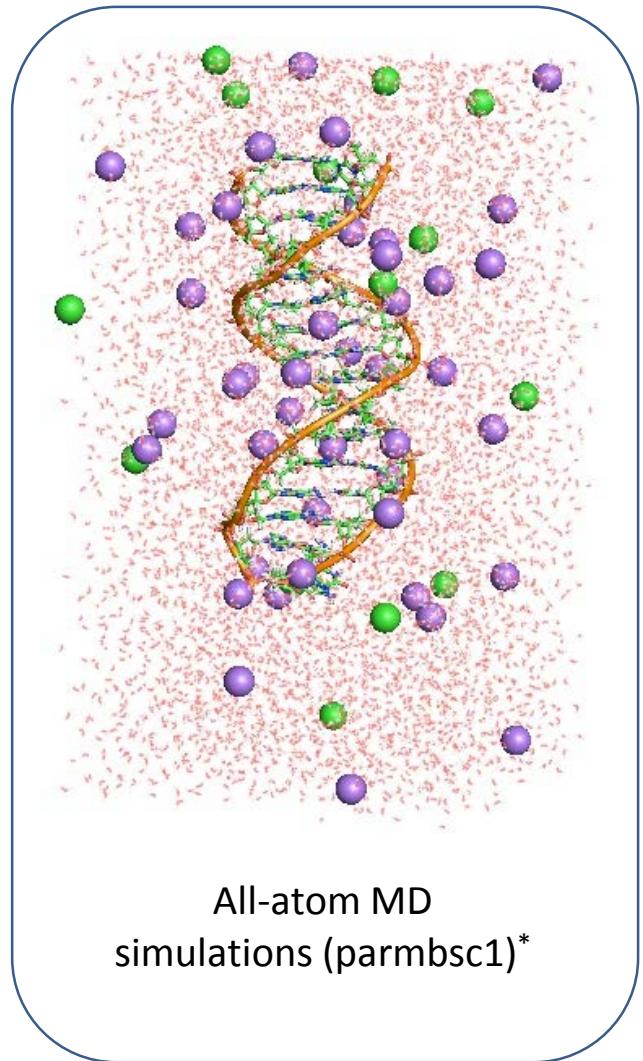


$$E = k(x - x_0)^2$$





Deriving spring constant K



The DNA model includes sequence-specific effects by elastic force constants K

DNA coarse-grain model on bp-step level

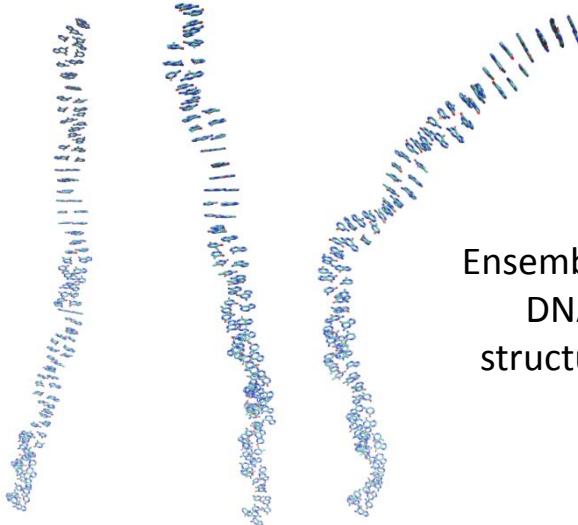
Metropolis Monte Carlo



Stochastic molecular simulation

Start from straight linear DNA
as starting structure

Calculate energy $E = \theta^T K \theta$



Ensemble of
DNA
structures

Randomly change
helical parameters
of random bp-step k

$$\theta_k \rightarrow \theta_{k \text{ new}}$$

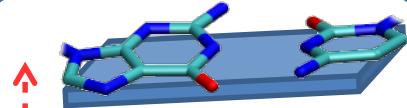
Calculate change
in potential energy

$$\Delta E = \theta_{k \text{ new}}^T K \theta_{k \text{ new}} - \theta_k^T K \theta_k$$

accept $\theta_{k \text{ new}}$ keep θ_k
yes no

$$\Delta E \geq 0: e^{-\Delta E/kT} > R ??$$

$$\Delta E < 0: \text{accept } \theta_{k \text{ new}}$$



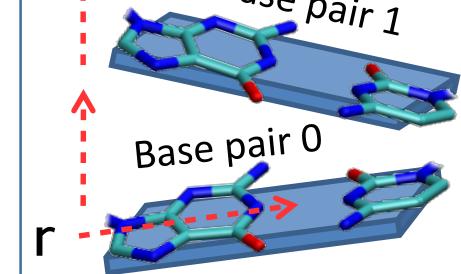
Base pair N

Rotation + Translation
(based on 6 helical
parameters)

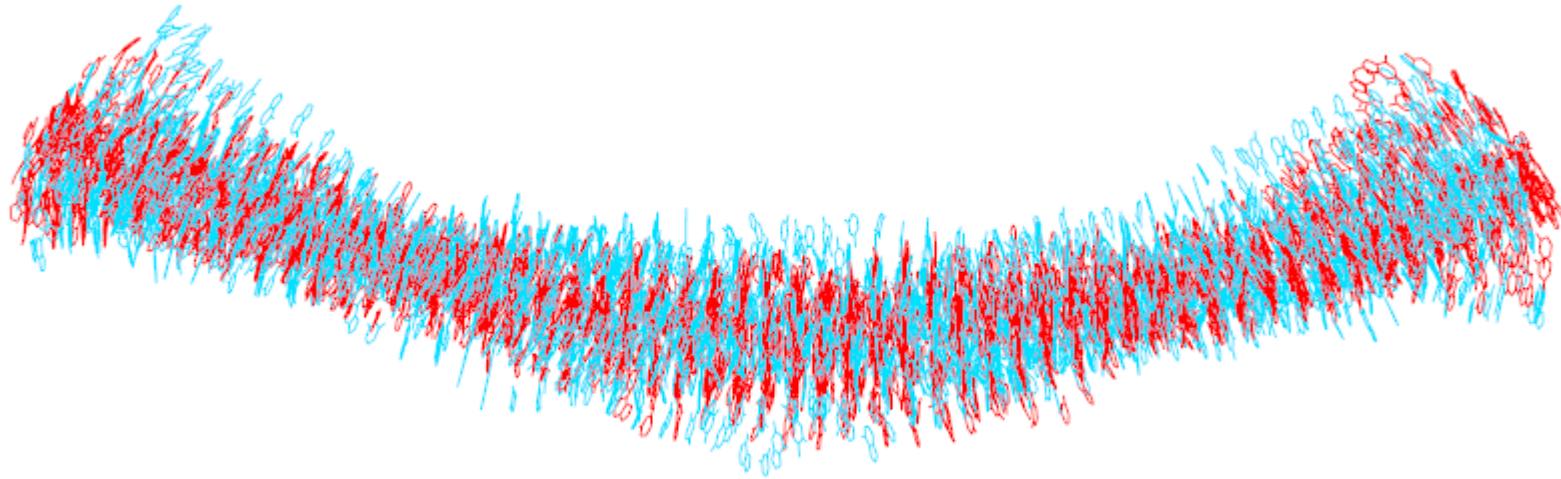
Base pair 1

Base pair 0

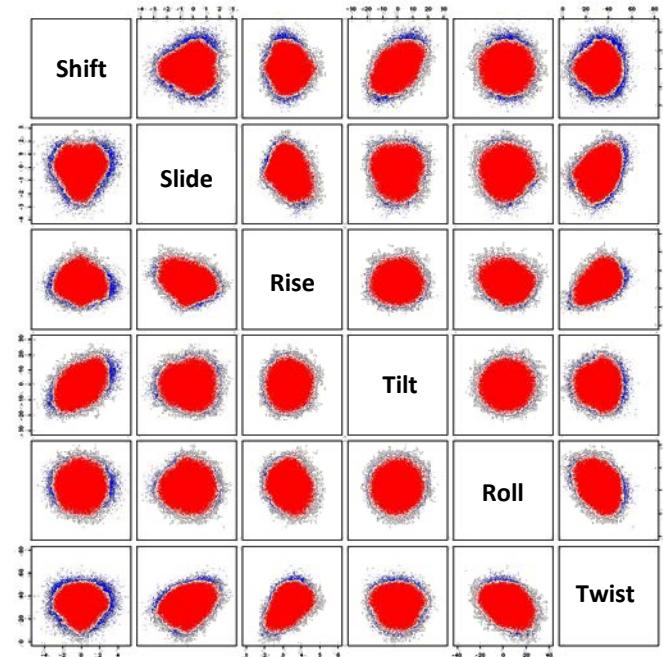
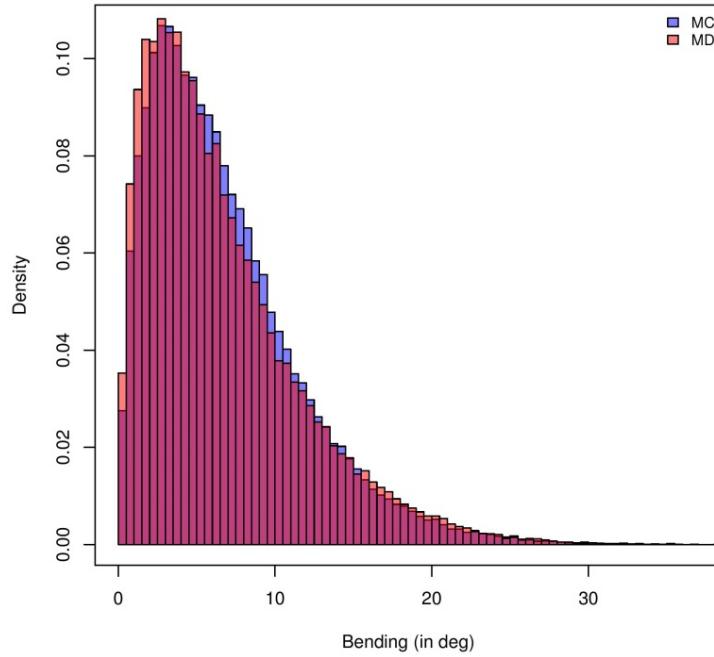
Cartesian
reconstruction



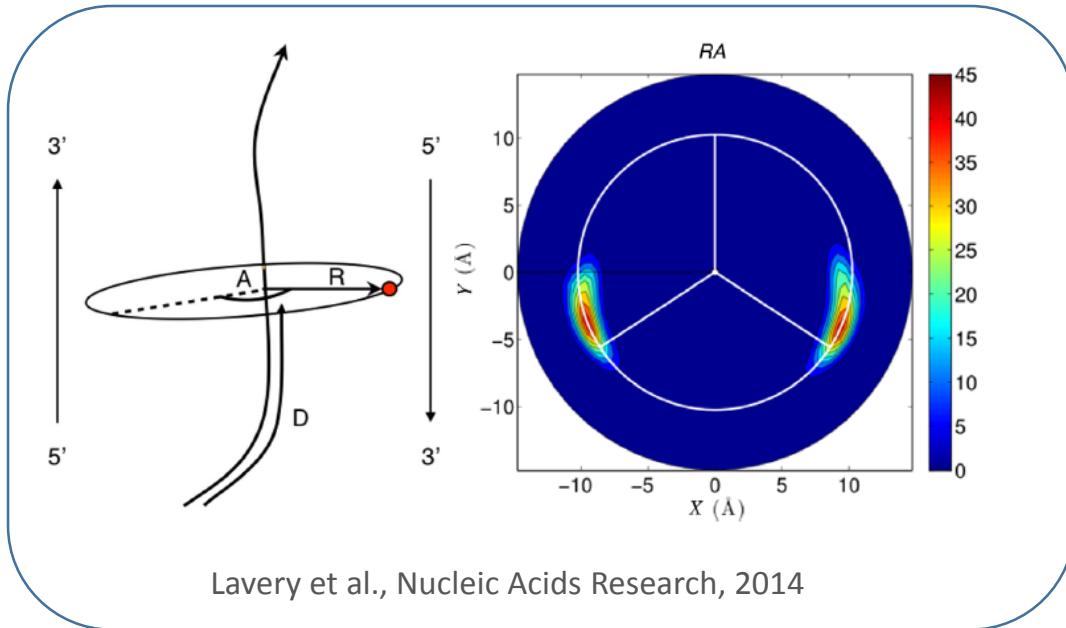
Validation by comparison with all atom **MD**



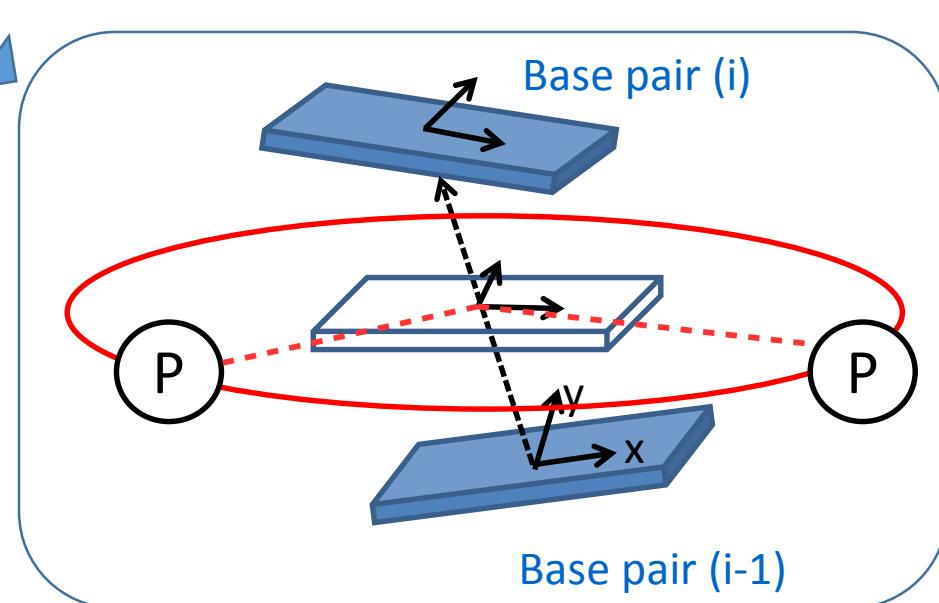
Bending of 10bp pieces



Inferring phosphate position from helical parameters

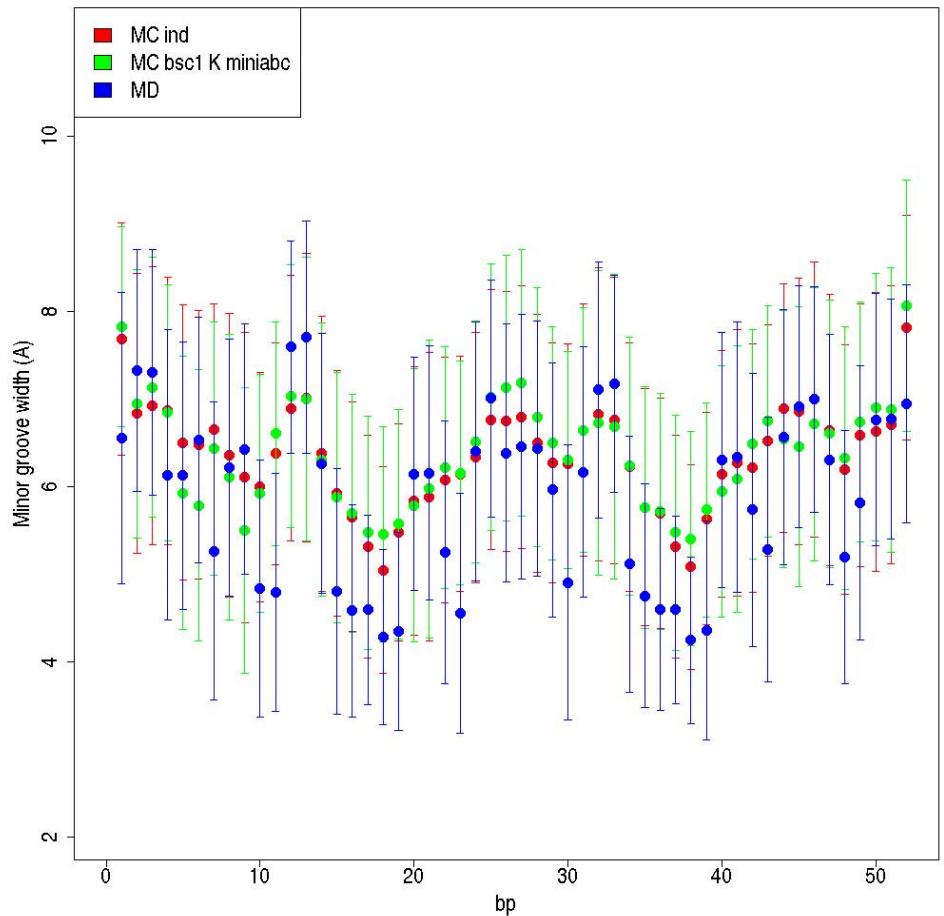


Position based on concentration maps
Computed with Curves+canion

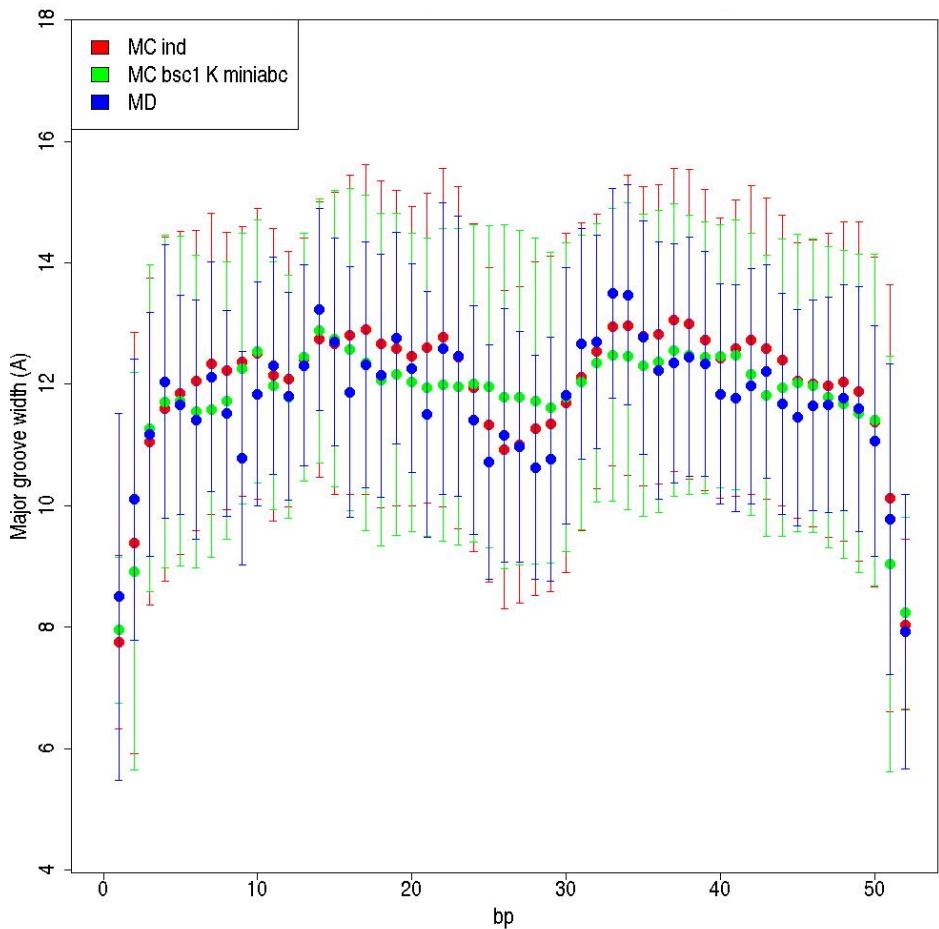


Major and minor groove width – 56mer

Minor groove



Major groove



CGCCGGCAGTAGCCGAAAAAAATAGGCGCGCTCAAAAAAATGCCCATGCCGCGC

CGCCGGCAGTAGCCGAAAAAAATAGGCGCGCTCAAAAAAATGCCCATGCCGCGC

Applications of the DNA model – Imposing distortions

Umbrella sampling – a method to obtain free energy profiles of distortions

Add elastic constraint to potential energy E_p of the system:

$$E_p \rightarrow E_p + \frac{k}{2}(x - x_0)^2$$

k : adjustable force constant

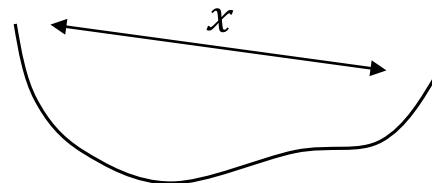
x : reaction coordinate

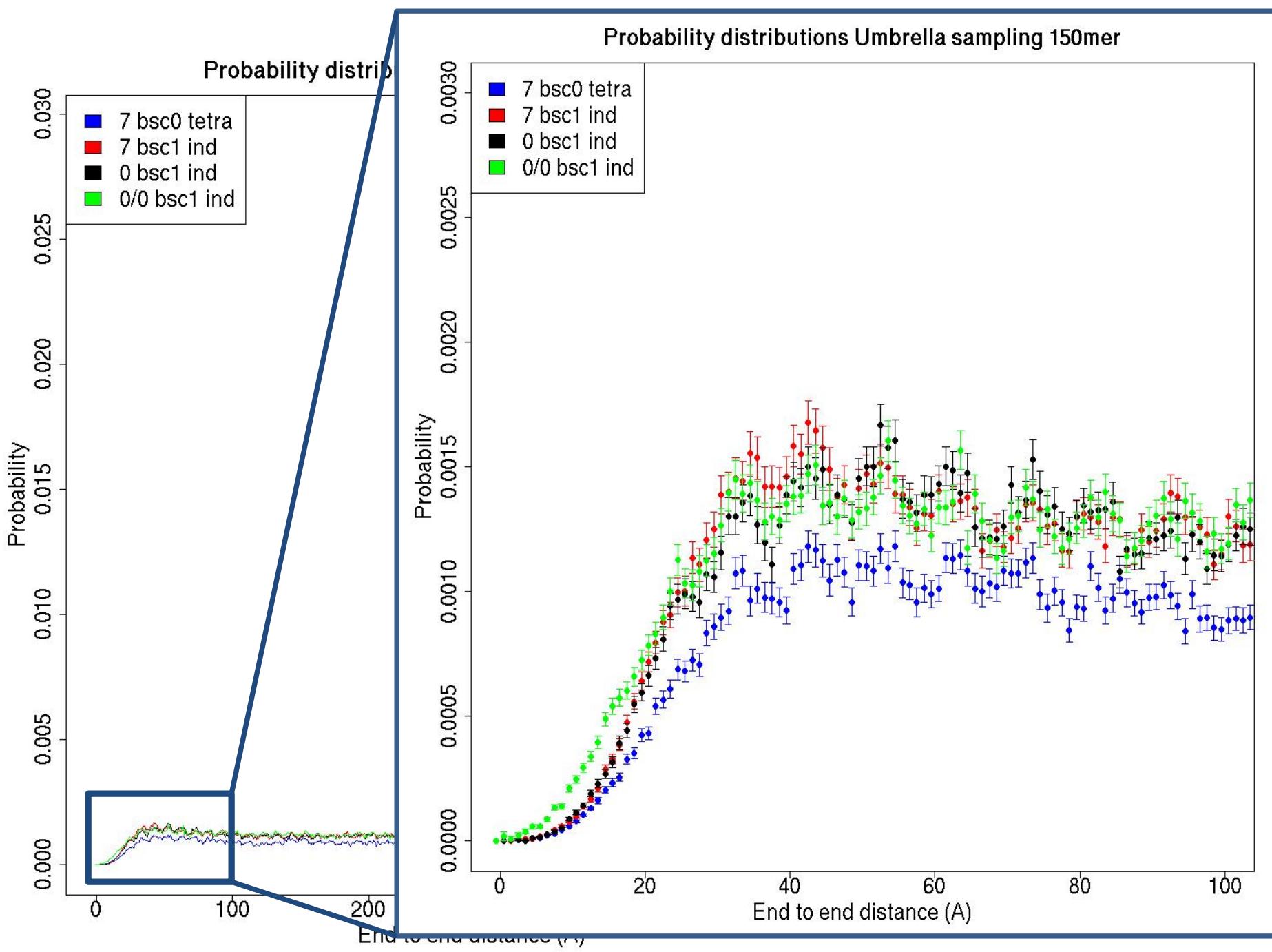
x_0 : desired value of reaction coordinate

Result: Probability and free energy distribution along reaction coordinate

Example: Cyclisation probability

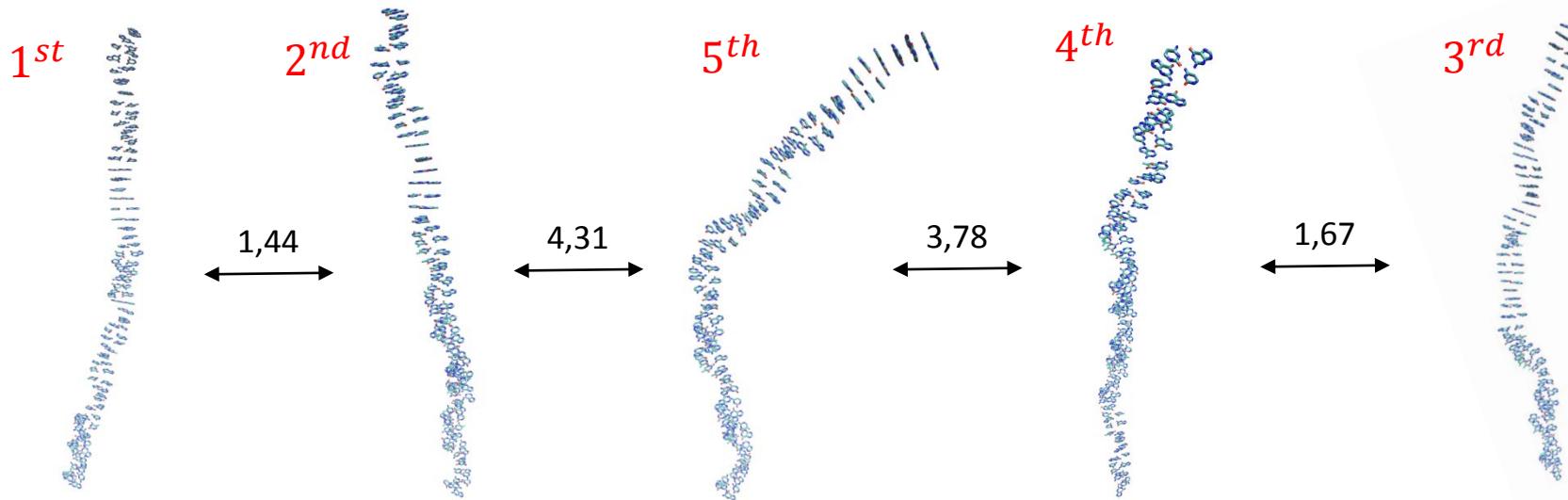
Set $x = |r_N - r_0|$ (end-to-end distance)





Create a kinetic series of DNA ensemble structures

RMSD (root mean squared distance) between structures



Structures	1	2	3	4	5
1		1,44	3,65	1,53	1,77
2			4,31	1,98	1,76
3				3,78	3,99
4					1,67
5					

Kinetic series – Example with 56 bp



Kinetic series – Example: Chromatin-like structure of 660 bp



How to use this tool

- Single sequence
- Multiple sequences to compare properties such as bending, major/minor groove width, sampling of conformational space, free energy profiles of distortions and the probability of certain distorted configurations

Give me a sequence
and I will make you
happy very quickly.





Multiscale Complex Genomics

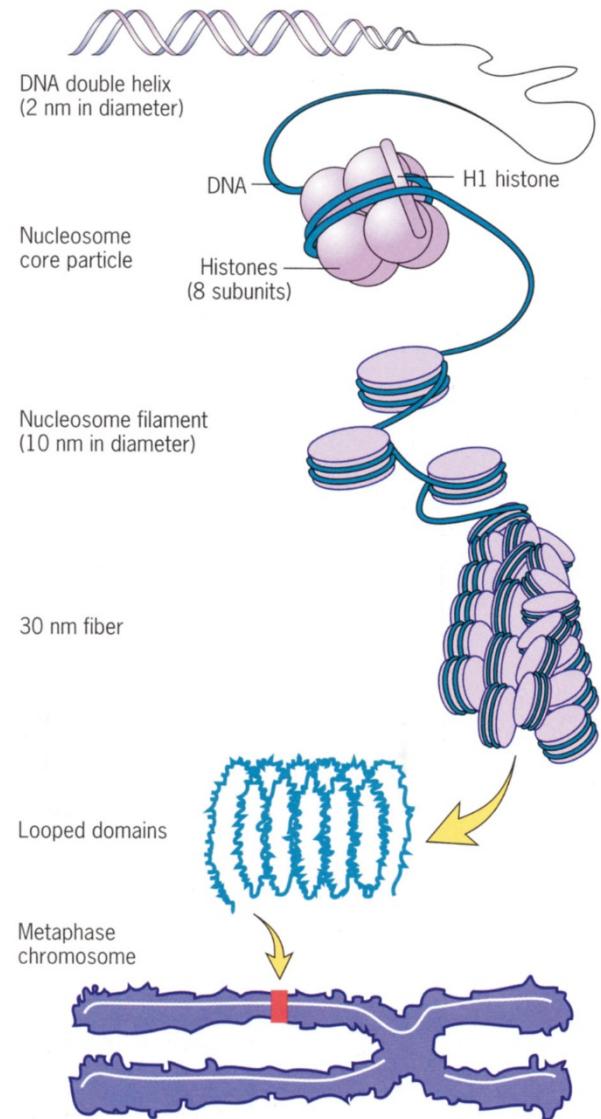
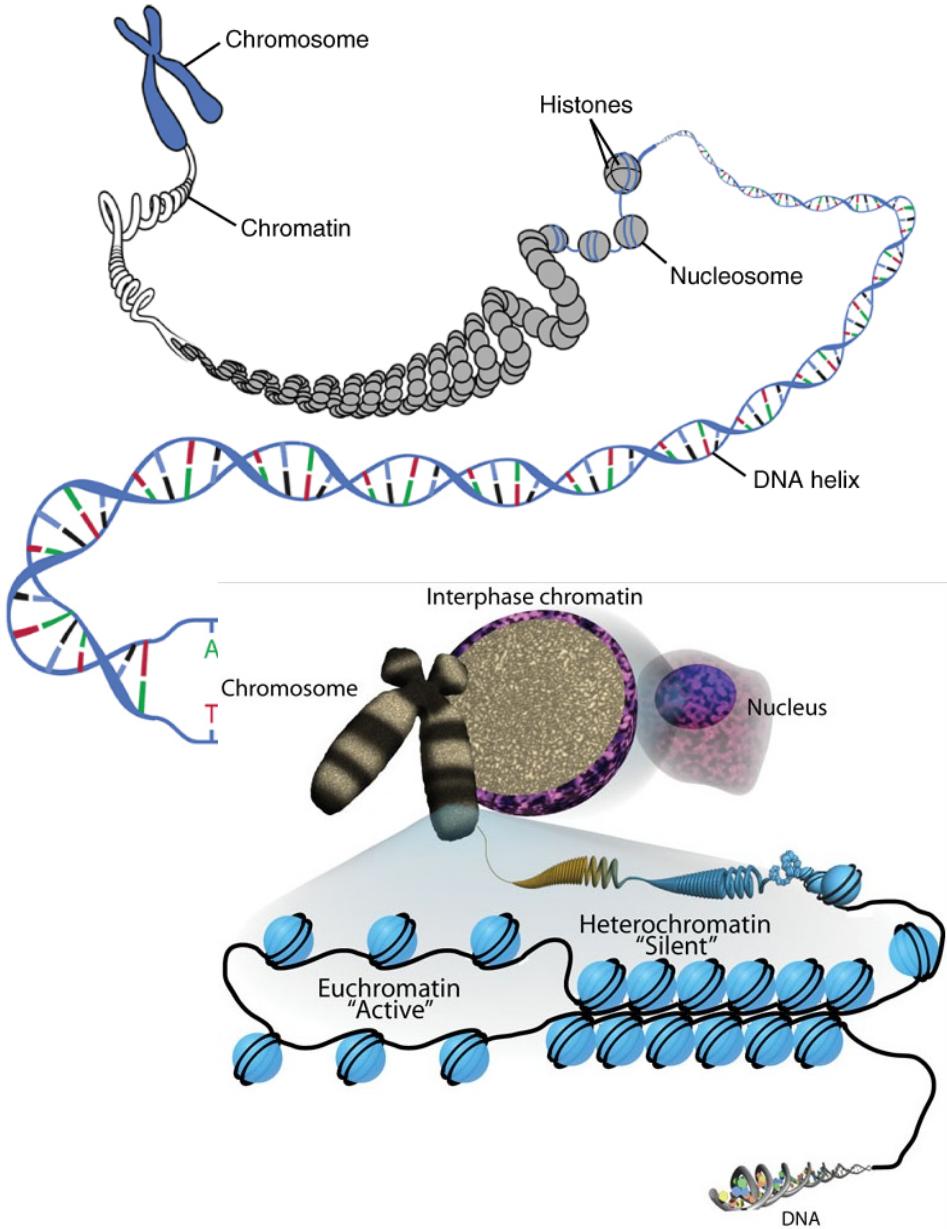
Coarse-grained chromatin model

Adding nucleosomes to the 6 parameter DNA model



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.

Chromatin – a 30nm fiber??



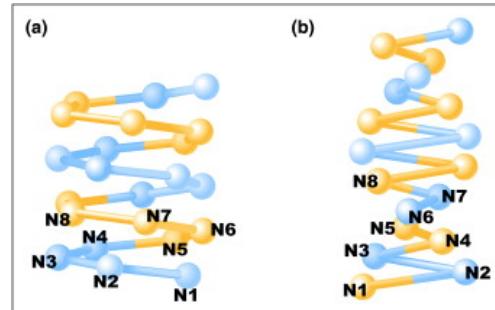
Chromatin structure: does the 30-nm fibre exist *in vivo*?

Maeshima et al. *Curr Opin Cell Biol* (2010)

In-vitro

P.J. Robinson et al. PNAS (2006)

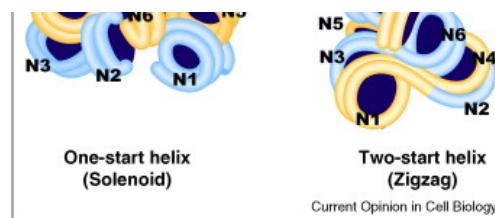
Cryo-EM of regular artificial chromatin fibers



T. Schalch et al. Nature (2005)

X-ray structure of tetra-nucleosome

"The precise spontaneous secondary structure of chromatin depends on the cell type and other internal and external factors, and is still under debate." (Özer et al. *Curr Opin Struct Biol* (2015))



In-vivo

J. Dubochet, N. Sartori Blanc. *Micron* (2001)

C. Bouchet-Marquis et al. *Histochem Cell Biol* (2006)

J. Dekker. *J Biol Chem* (2008)

cryo-EM
cryo-EM
3C in yeast

No regular 30nm fiber visible

Bottom-up



Short region of DNA double helix

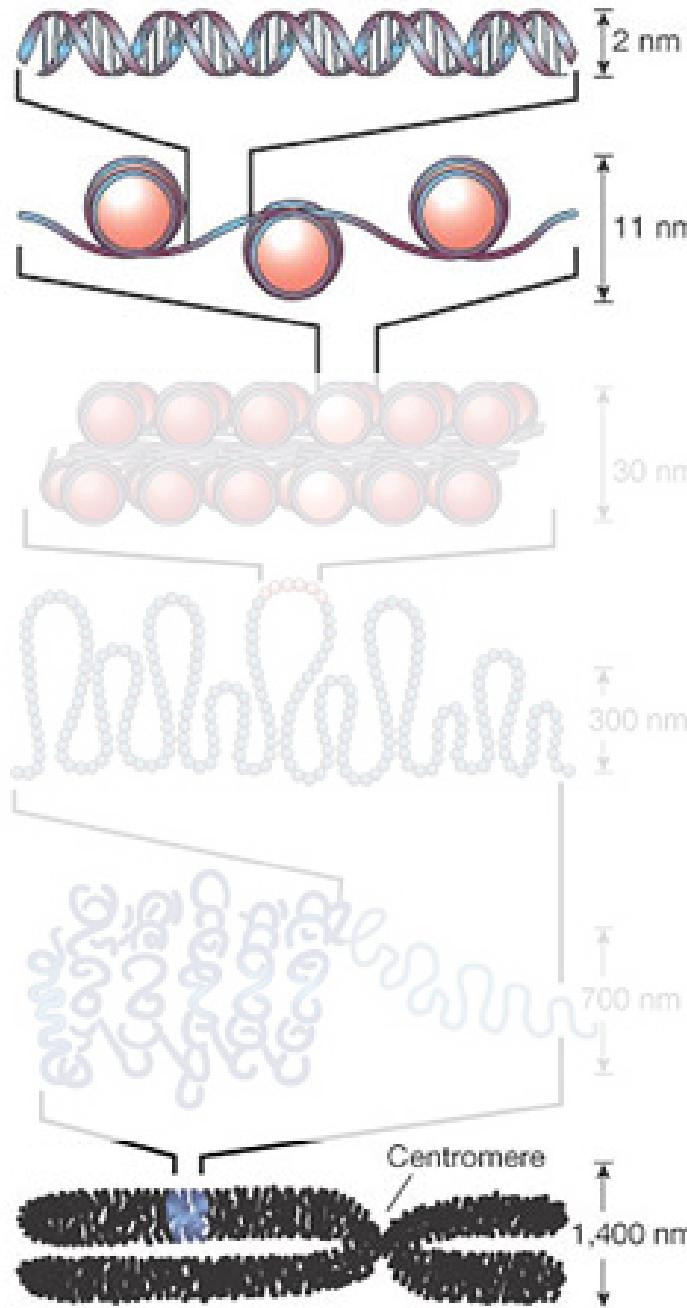
"Beads on a string" form of chromatin

30-nm chromatin fibre of packed nucleosomes

Section of chromosome in an extended form

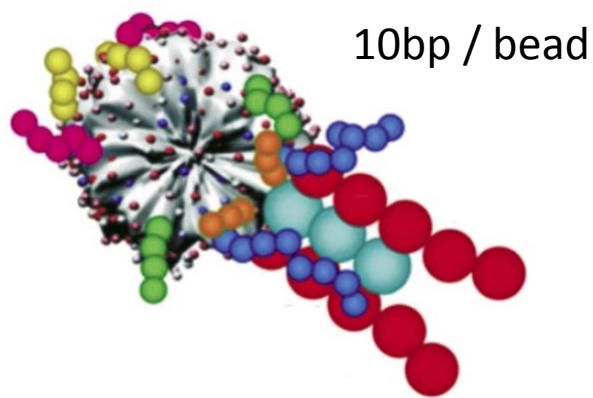
Condensed section of chromosome

Entire mitotic chromosome

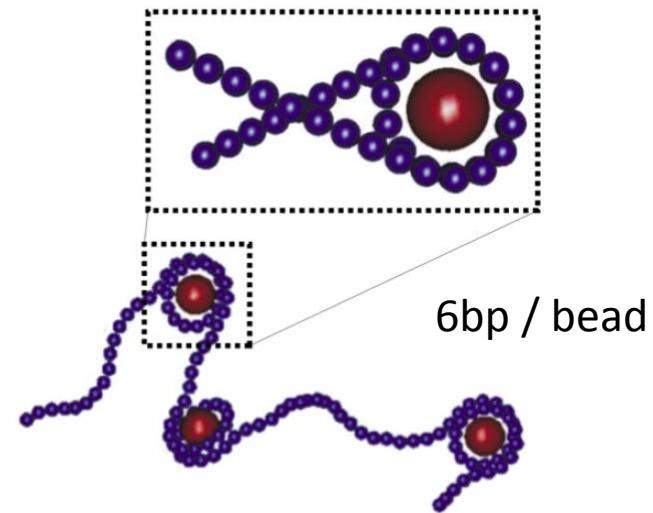


Top-down

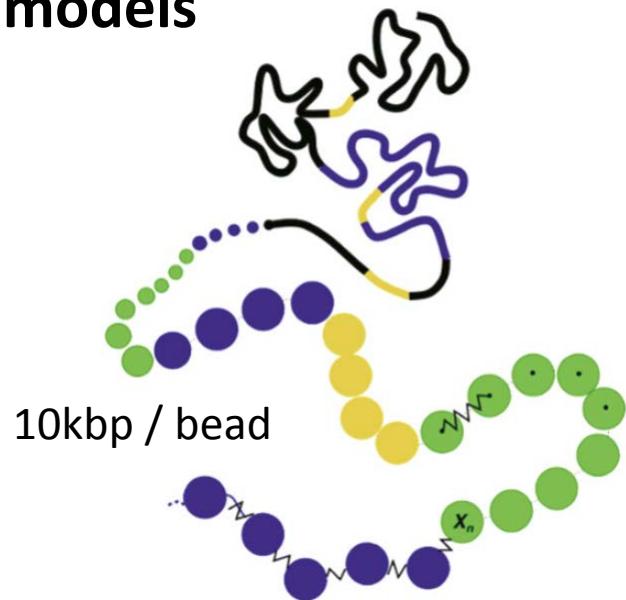
Bottom-up models



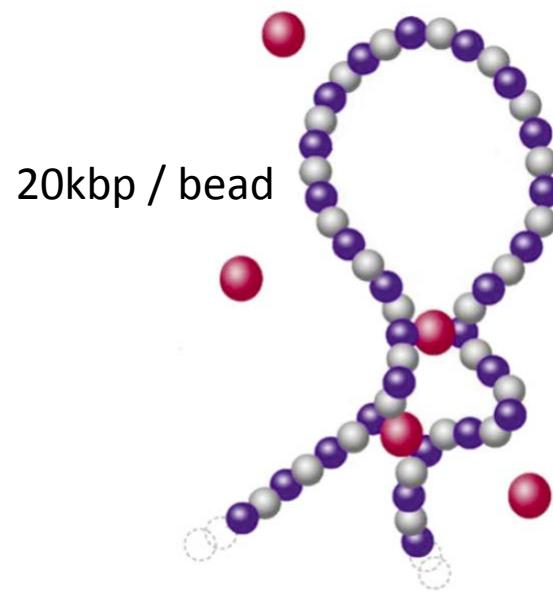
Collepardo-Guevara R, Schlick T , PNAS. 2014



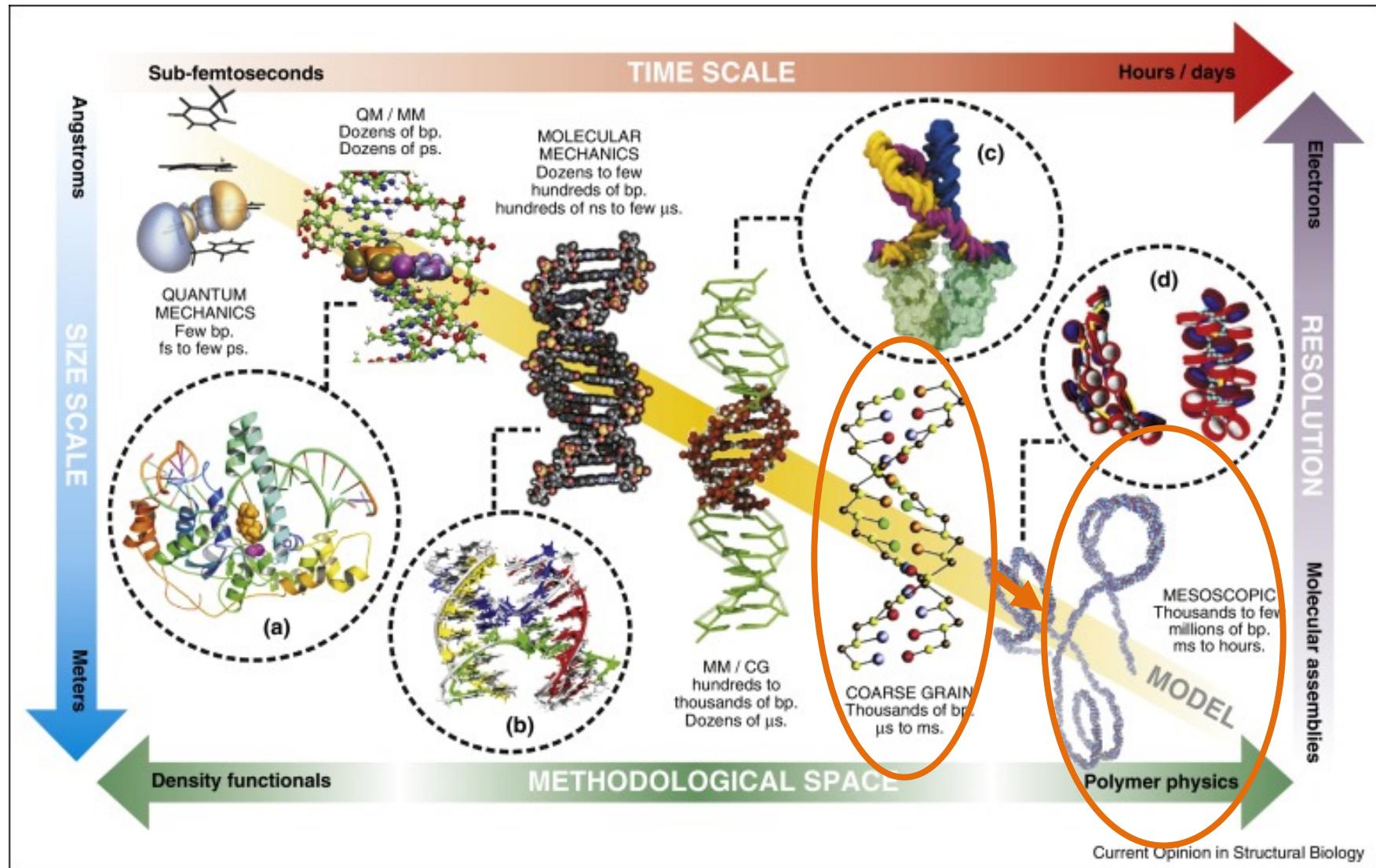
Top-down models



Jost et al., NAR. 2014



Barbieri et al., PNAS. 2012



Extension to a coarse-grained chromatin model...

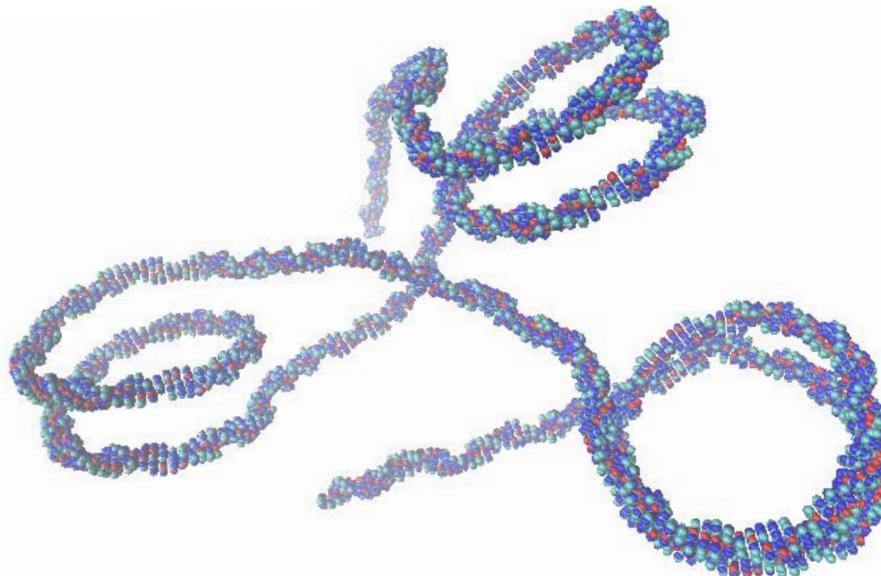
INPUT

Nucleosome 1 Nucleosome 2 Nucleosome 3

↓ ↓ ↓

ACGTACTGACGTTACTAACCTGCGC TGCACTAGCAAACGGTTGACCACCGTGAGATCTCCTAACG CGGATTGGATCCGACTACGATCAACC GATAGCAGCGCGCGCGCGA

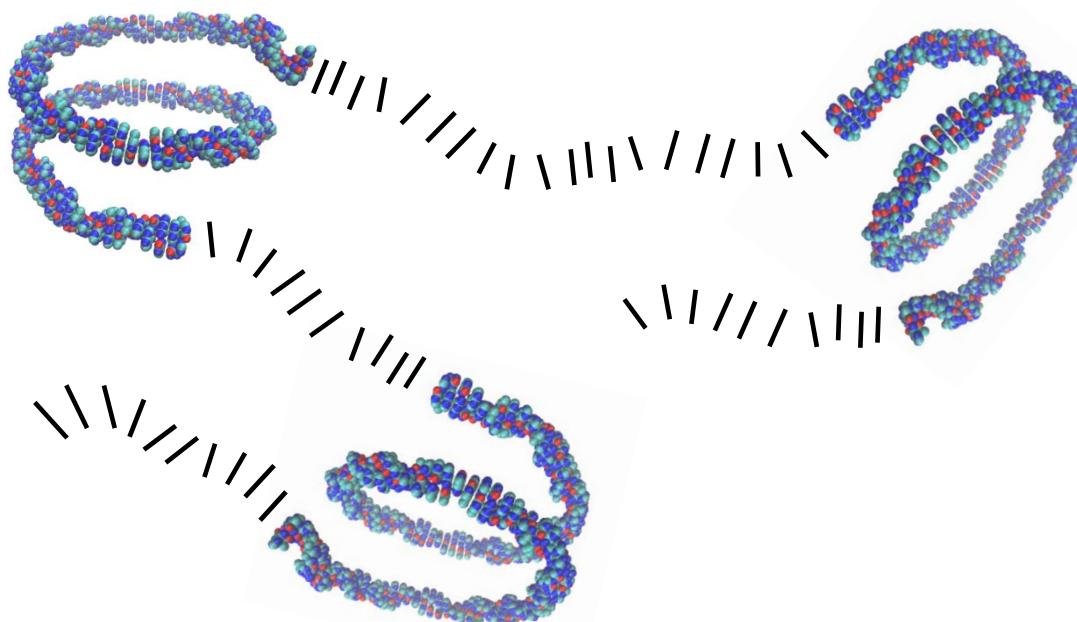
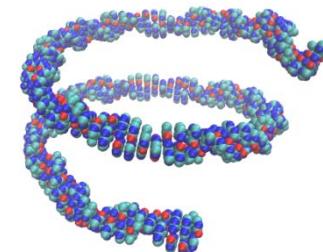
OUTPUT

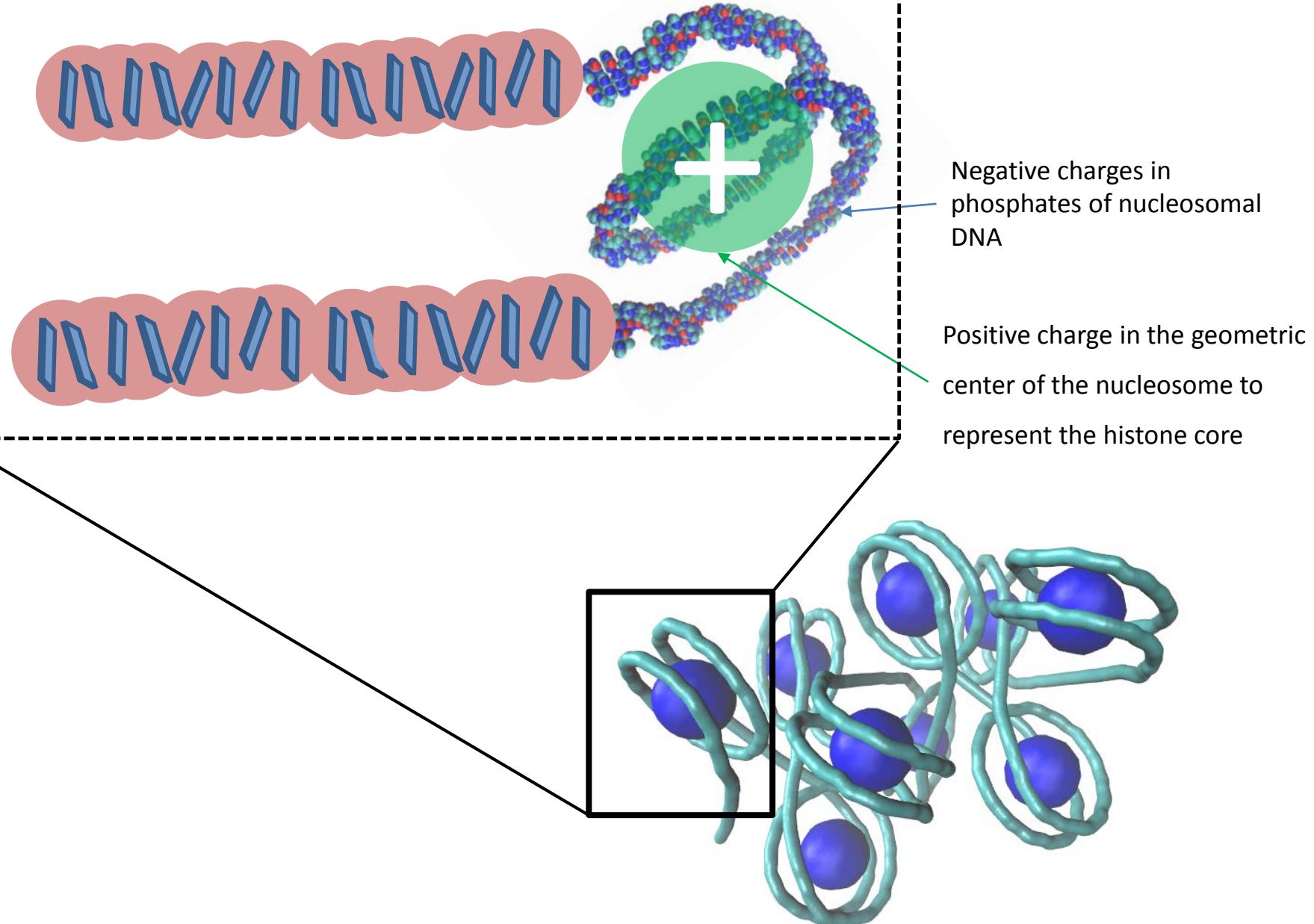


Linker DNA at bp-level with helical coordinates

Nucleosome rigid; spatial pathway of 1kx5

Exclude nucleosomal DNA from sampling



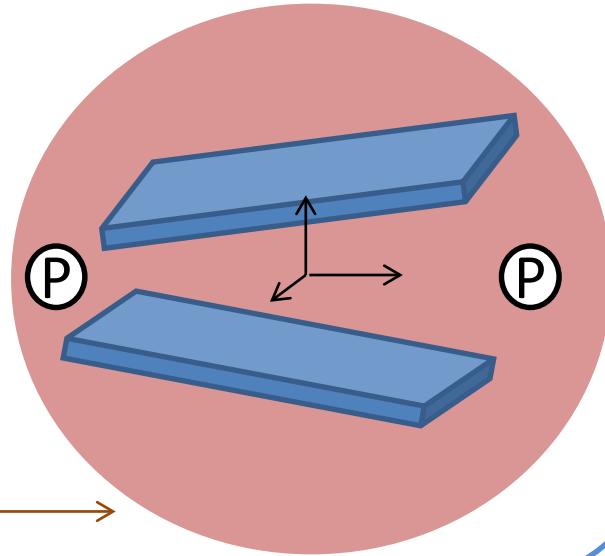


Long range electrostatic and excluded volume potential for linker DNA, nucleosomal DNA and nucleosome cores

$$E^{DH} = \sum_{i=1}^N \sum_{j=i}^N \frac{q_i q_j}{4\pi \epsilon \epsilon_0 r_{ij}} e^{-\kappa r_{ij}}$$

$$E^{LJ} = \sum_{i=1}^M \sum_{j=i}^M 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right]$$

Excluded Volume →



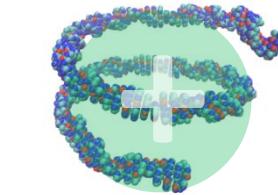
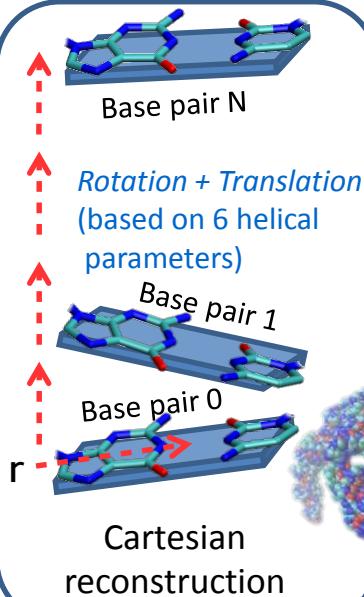
Random move on helical parameters gets accepted if the total energy

$$E_{tot} = E_{elast} + E^{DH} + E^{LJ}$$

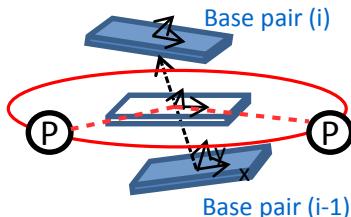
satisfies the Metropolis criterion.



Random move on helical parameters of a bp step



Nucleosome as a rigid body
(spatial pathway according to
the structure of nucleosome)
Positive charge in geometric
center of the nucleosome



Determine position of phosphates in the backbone

Random move on helical parameters accepted if the total energy

$$E_{tot} = E_{elast} + E^{DH} + E^{LJ}$$

satisfies the Metropolis criterion.

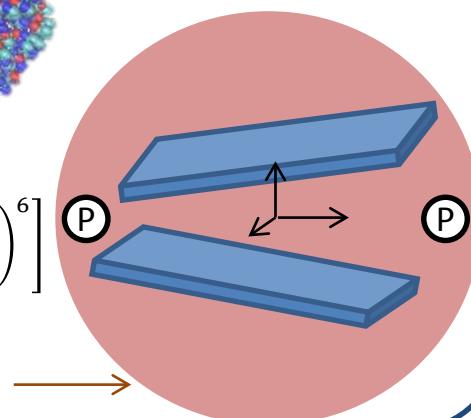
If accepted, repeat all steps with new linker configuration. If rejected repeat all steps with previous linker configuration

Long range electrostatic and excluded volume potential for linker DNA, nucleosomal DNA

$$E^{DH} = \sum_{i=1}^N \sum_{j=i}^N \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} e^{-\kappa r_{ij}}$$

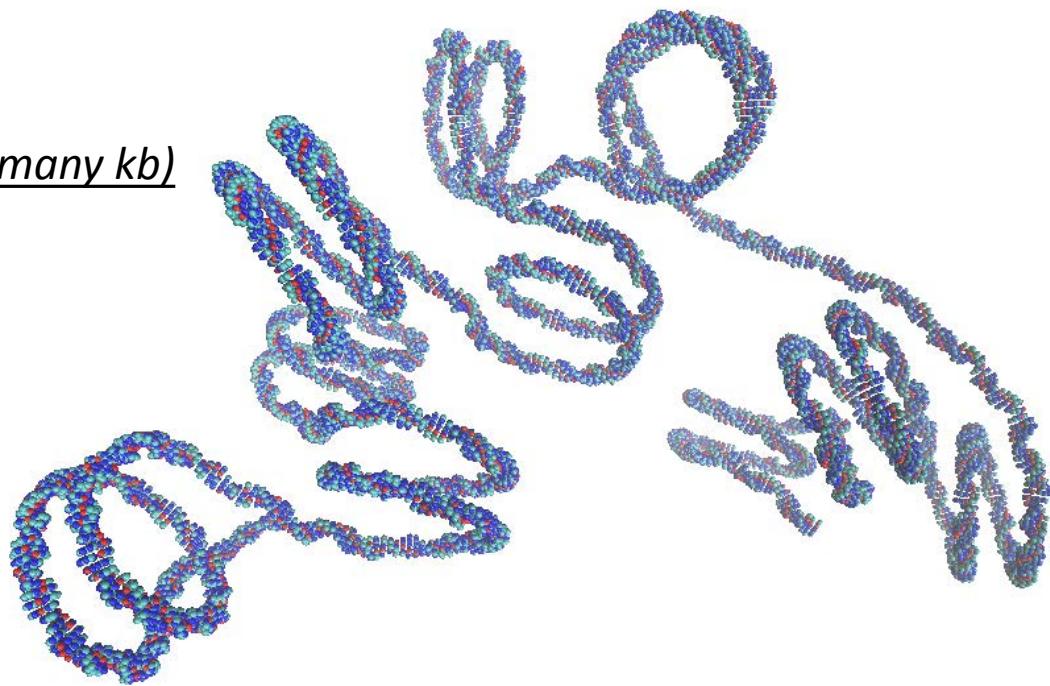
$$E^{LJ} = \sum_{i=1}^M \sum_{j=i}^M 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right]$$

Excluded Volume



Generate chromatin structure (0 – many kb)

- Variable linker length
- Variable linker sequence



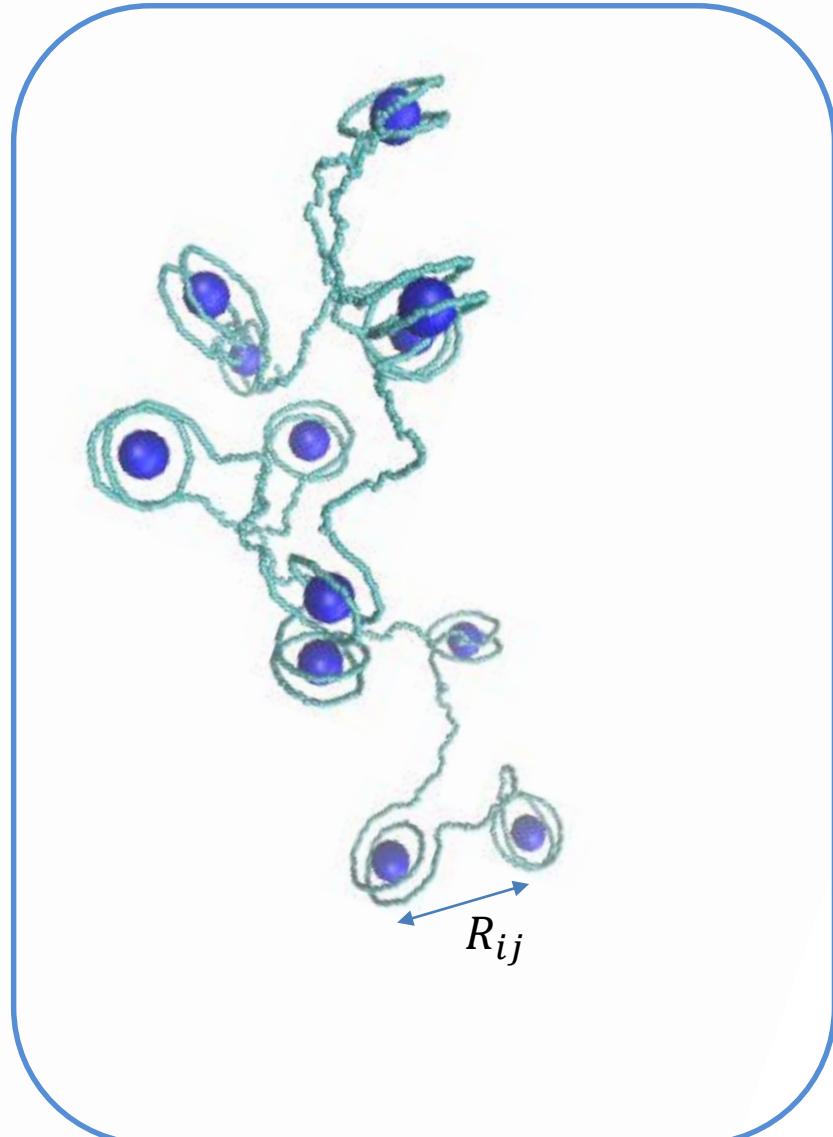
Validation: Salt-dependence of chromatin compaction

Sedimentation coefficient of nucleosomes

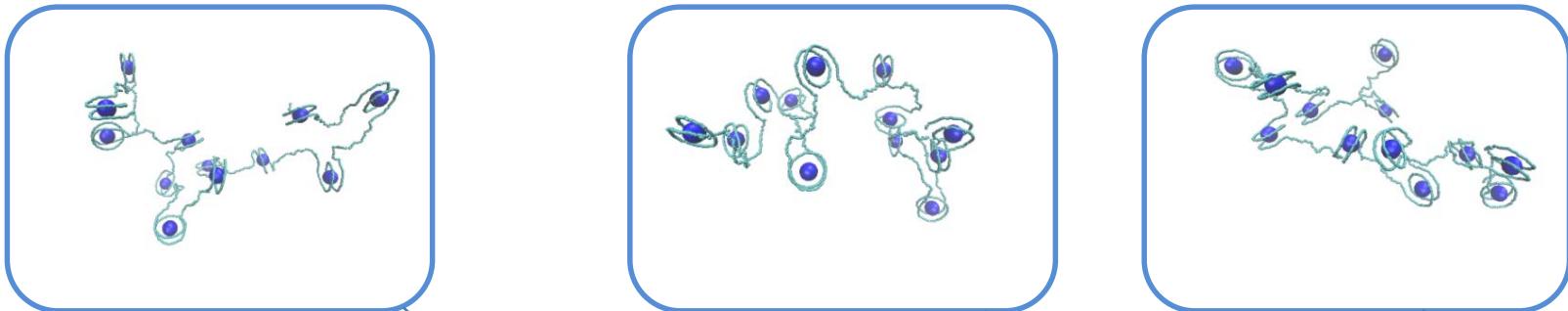
$$S = S_1 \left(1 + \frac{2R}{N} \sum_i^N \sum_{j>i}^N \frac{1}{R_{ij}} \right)$$

Experimental structure¹

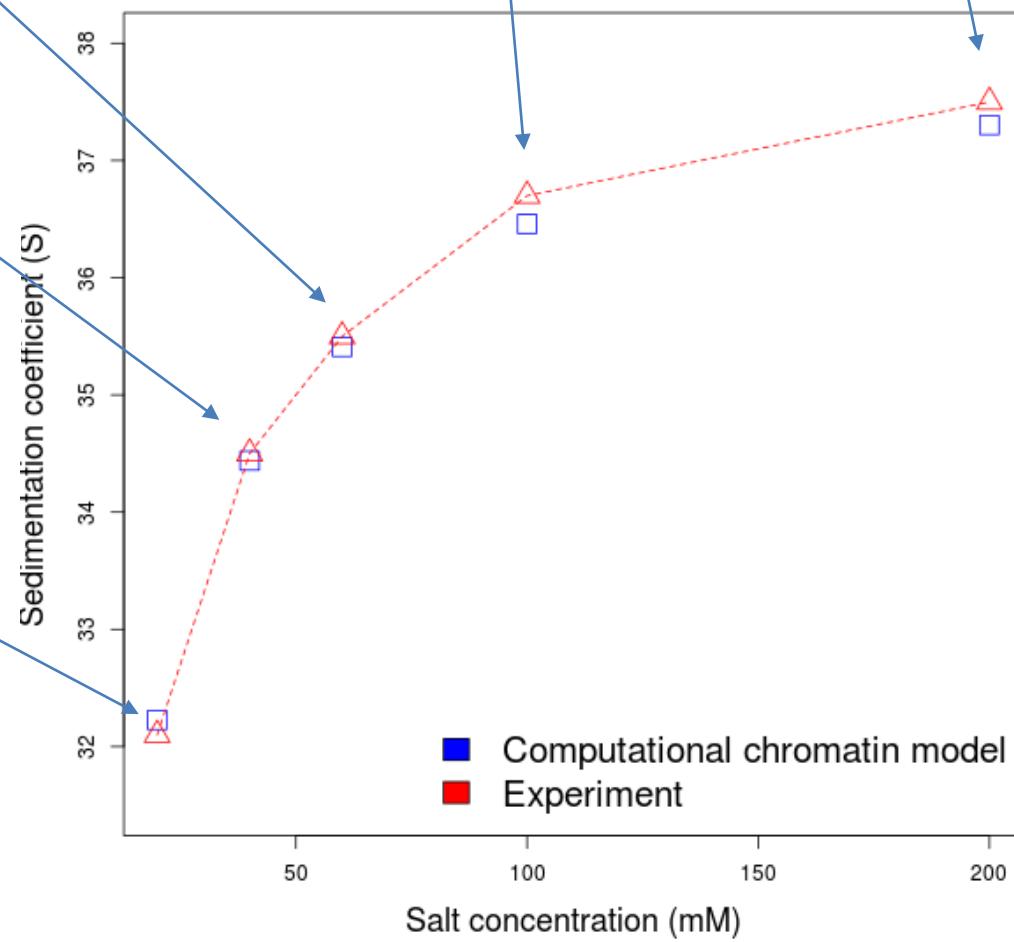
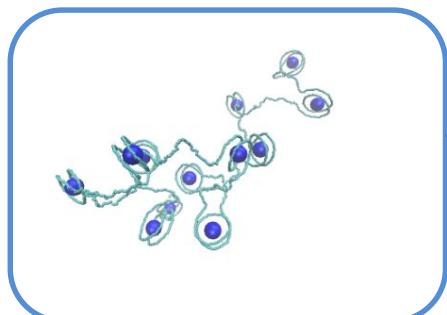
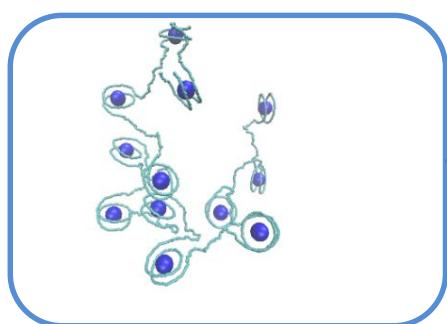
12 nucleosomes with 62 bp
of linker DNA



(1) Hansen et al., Biochemistry, 1989



Salt dependence of sedimentation coefficient



Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C

Tsung-Han S. Hsieh,¹ Assaf Weiner,^{2,3} Bryan Lajoie,^{1,4} Job Dekker,^{1,4} Nir Friedman,^{2,3} and Oliver J. Rando^{1,*}

¹Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

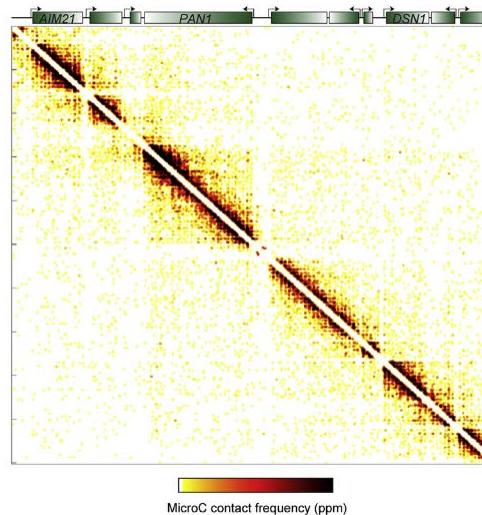
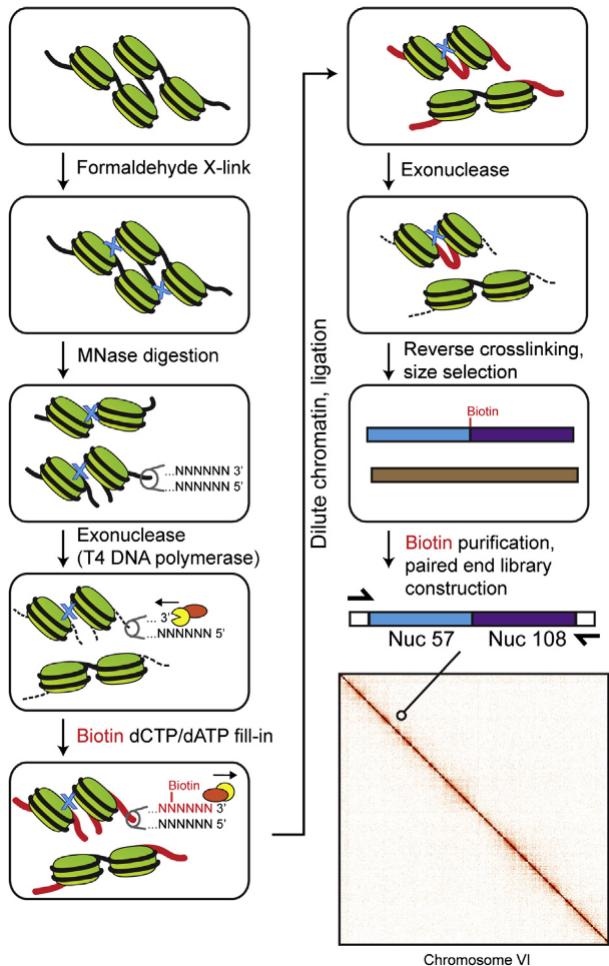
²School of Computer Science and Engineering, The Hebrew University, Jerusalem 91904, Israel

³Alexander Silberman Institute of Life Sciences, The Hebrew University, Jerusalem 91904, Israel

⁴Program in Systems Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA

*Correspondence: oliver.rando@umassmed.edu

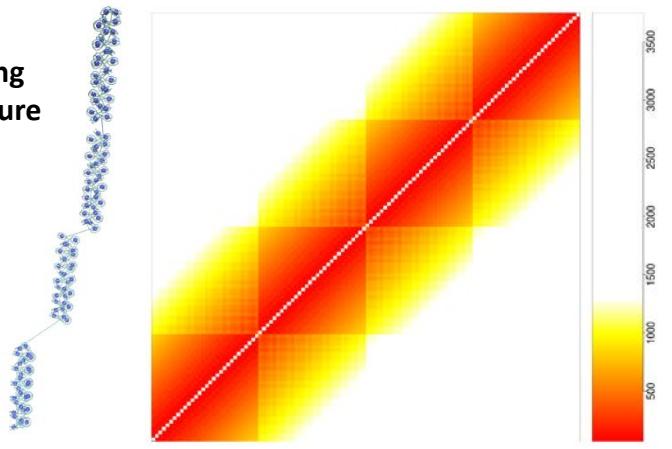
<http://dx.doi.org/10.1016/j.cell.2015.05.048>



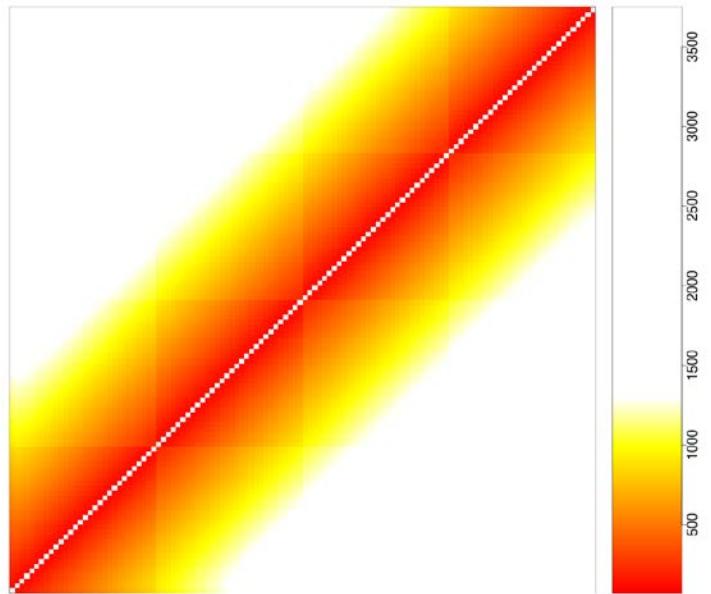
- Self-associating domains 1-5 genes (ca. 2-10kb)
- Boundaries of self-associating domains enriched in nucleosome-depleted regions
- Support for a common motif of zig-zag arrangement of nucleosomes

Example: Yeast-like chromatin (30n – 100 - 30n – 100 – 30n – 100 – 30n)

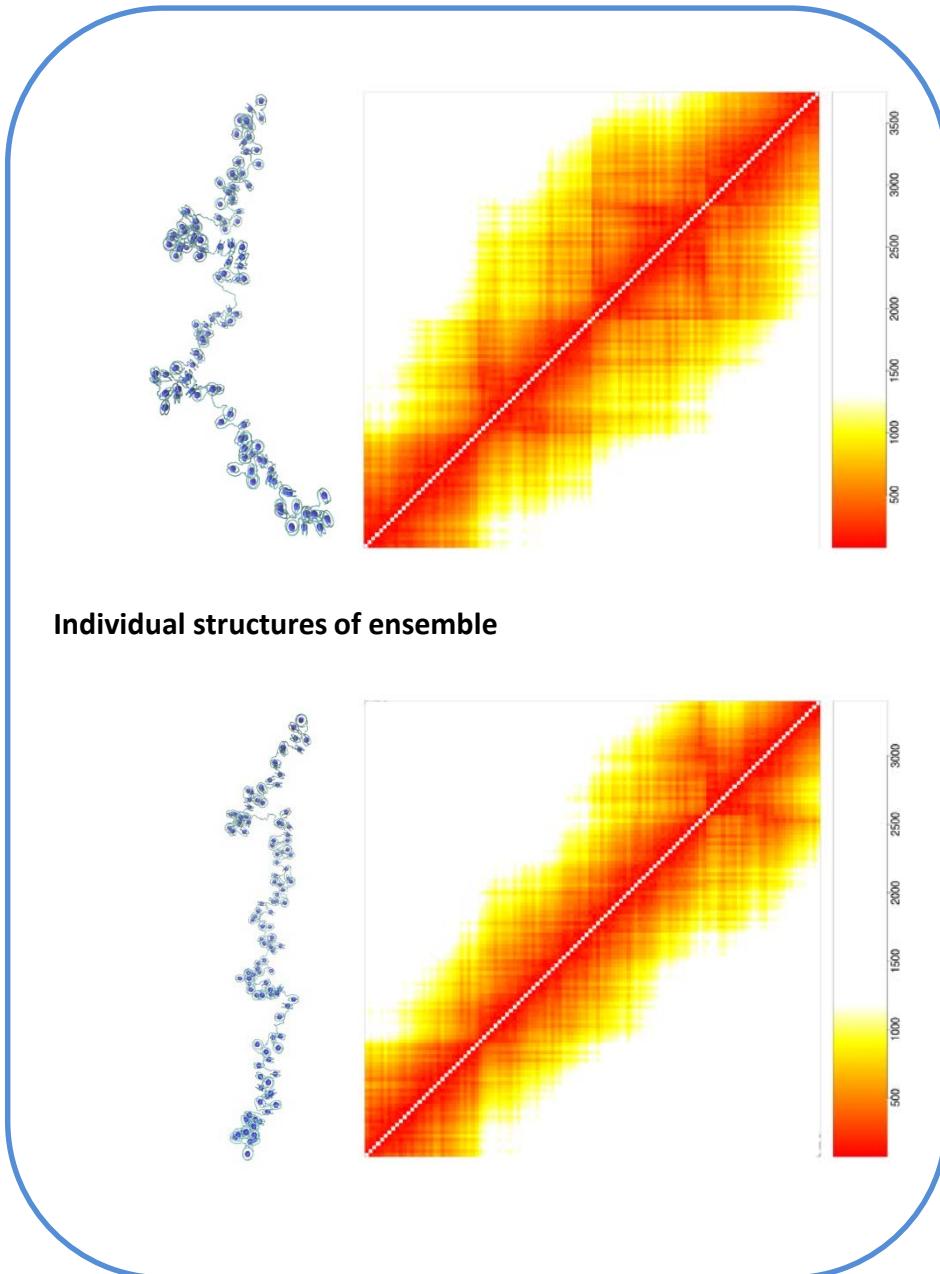
Starting structure



Ensemble average of nucleosome distances

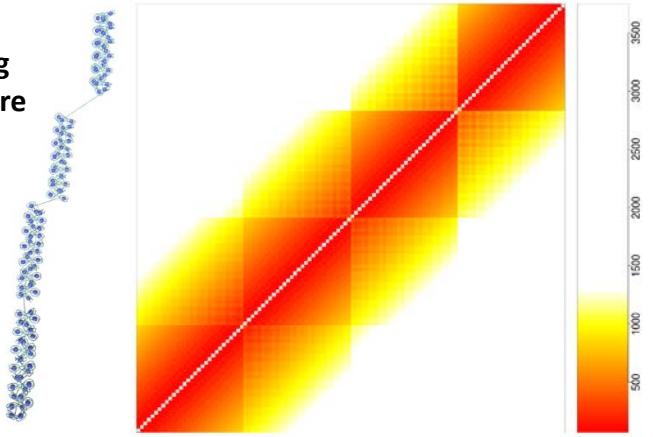


Individual structures of ensemble

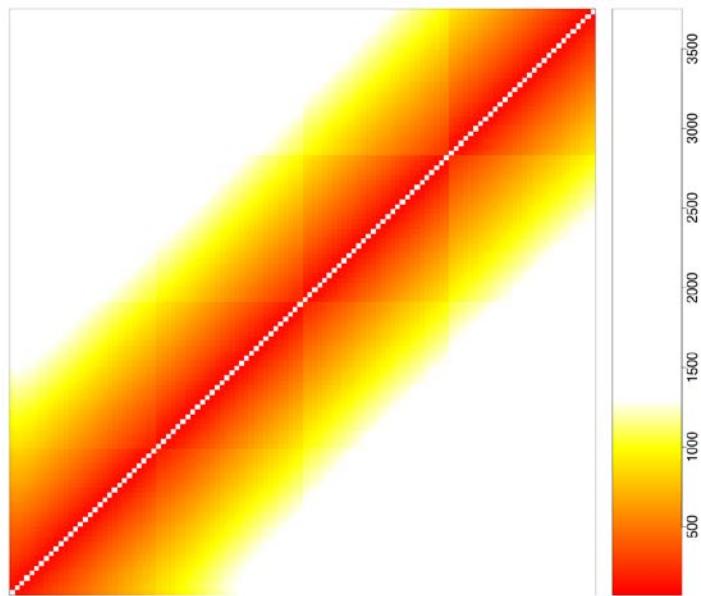


Example: Yeast-like chromatin (30n – 70 - 30n – 90 – 30n – 140 – 30n)

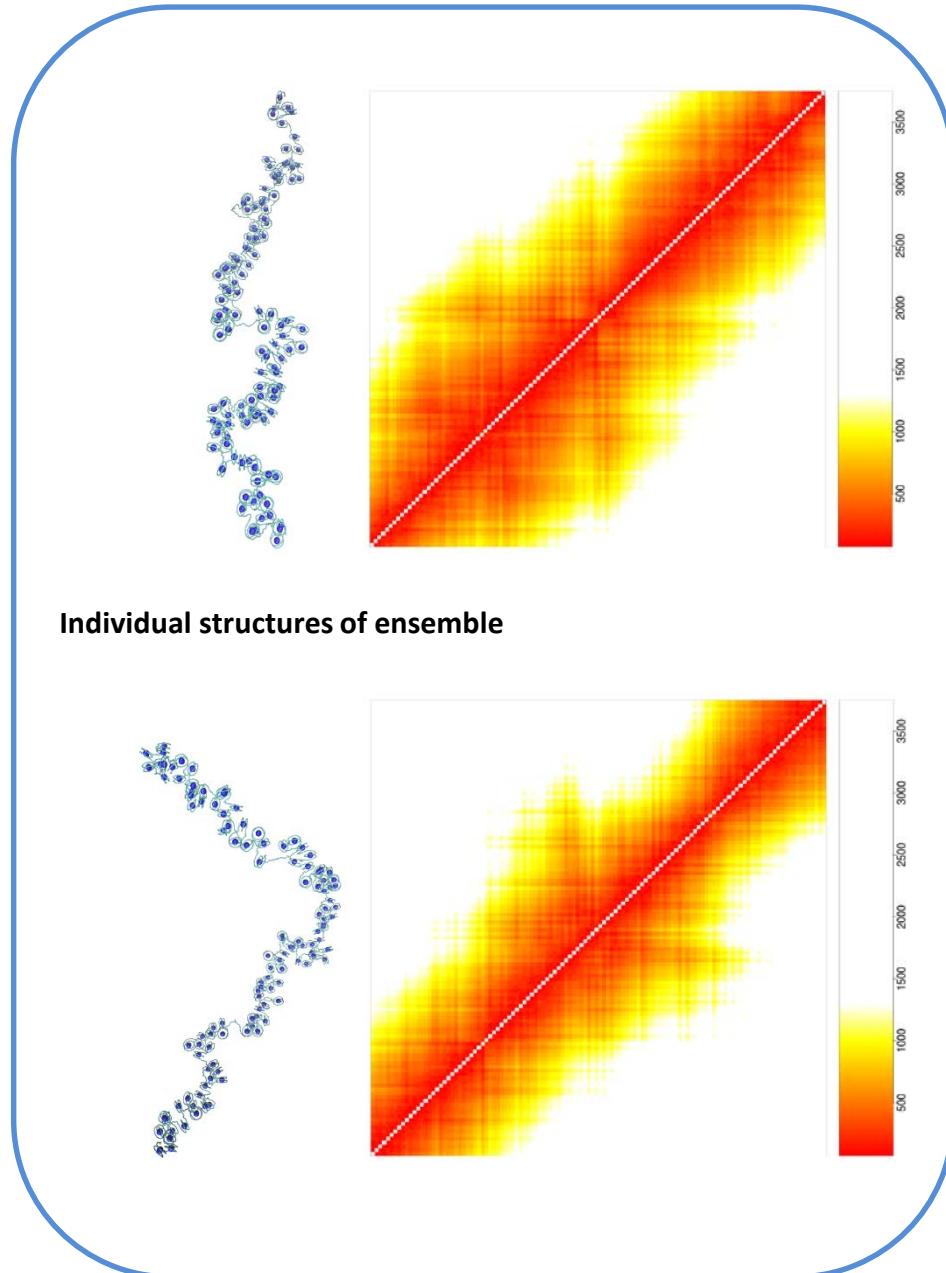
Starting structure



Ensemble average of nucleosome distances

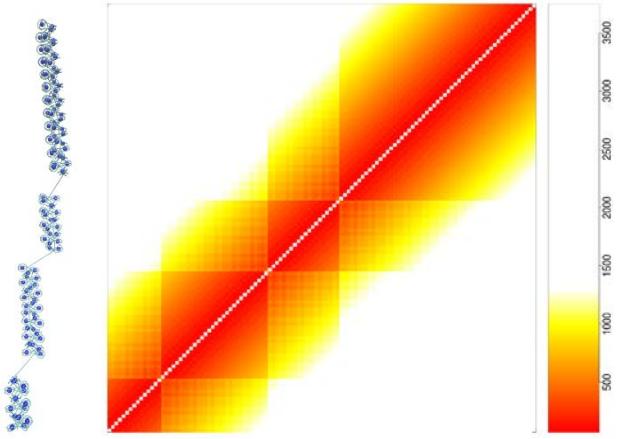


Individual structures of ensemble

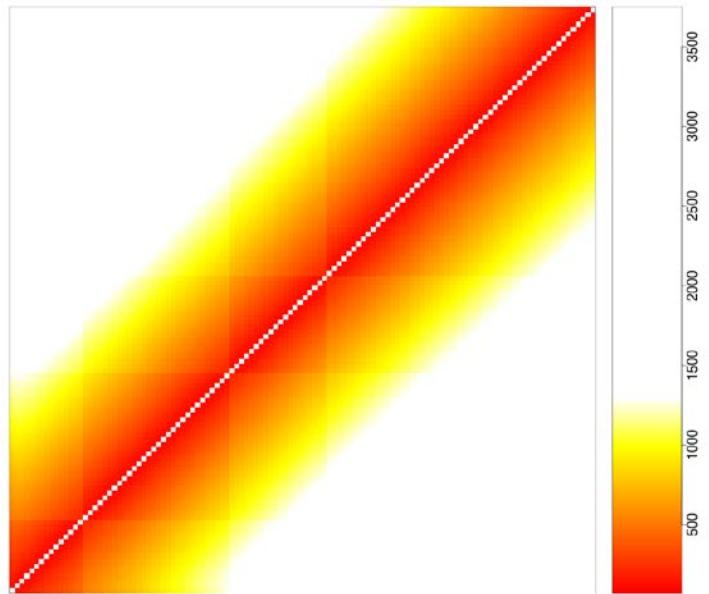


Example: Yeast-like chromatin (15n – 100 - 30n – 100 – 20n – 100 – 55n)

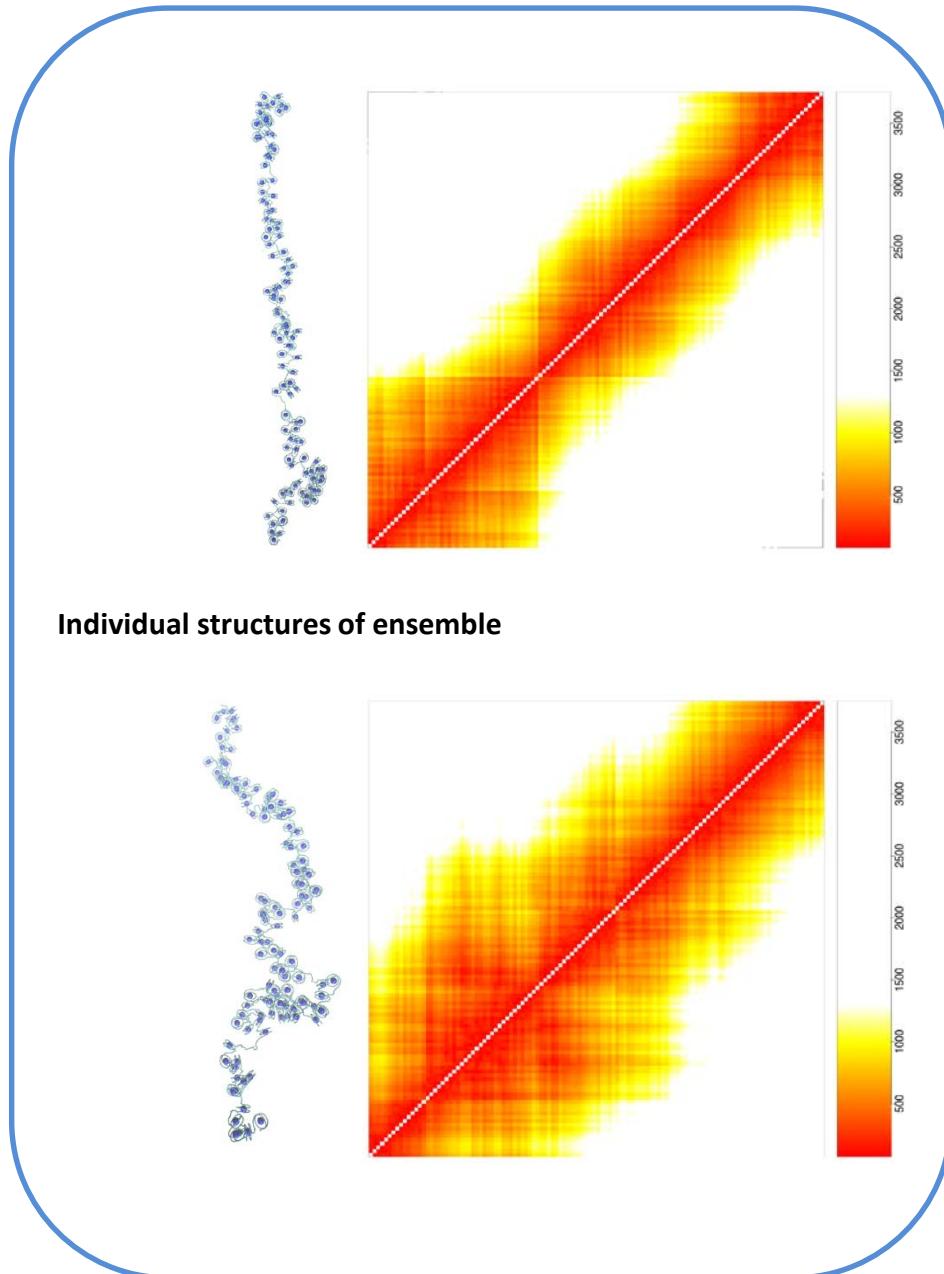
Starting structure



Ensemble average of nucleosome distances

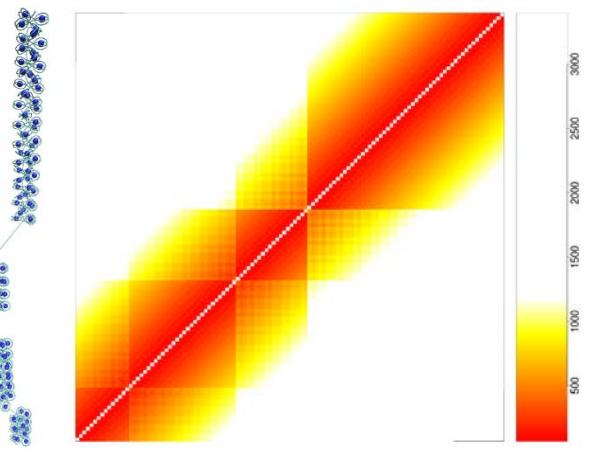


Individual structures of ensemble

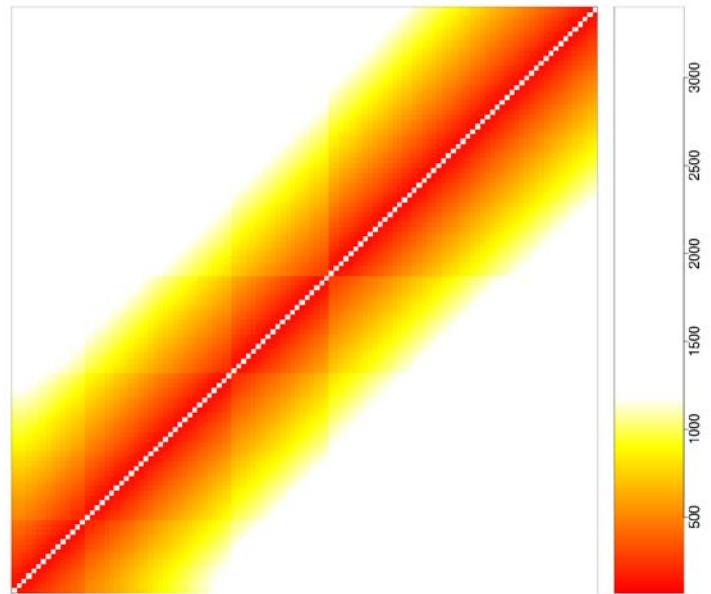


Example: Yeast-like chromatin (15n – 70 - 30n – 90 – 20n – 140 – 55n)

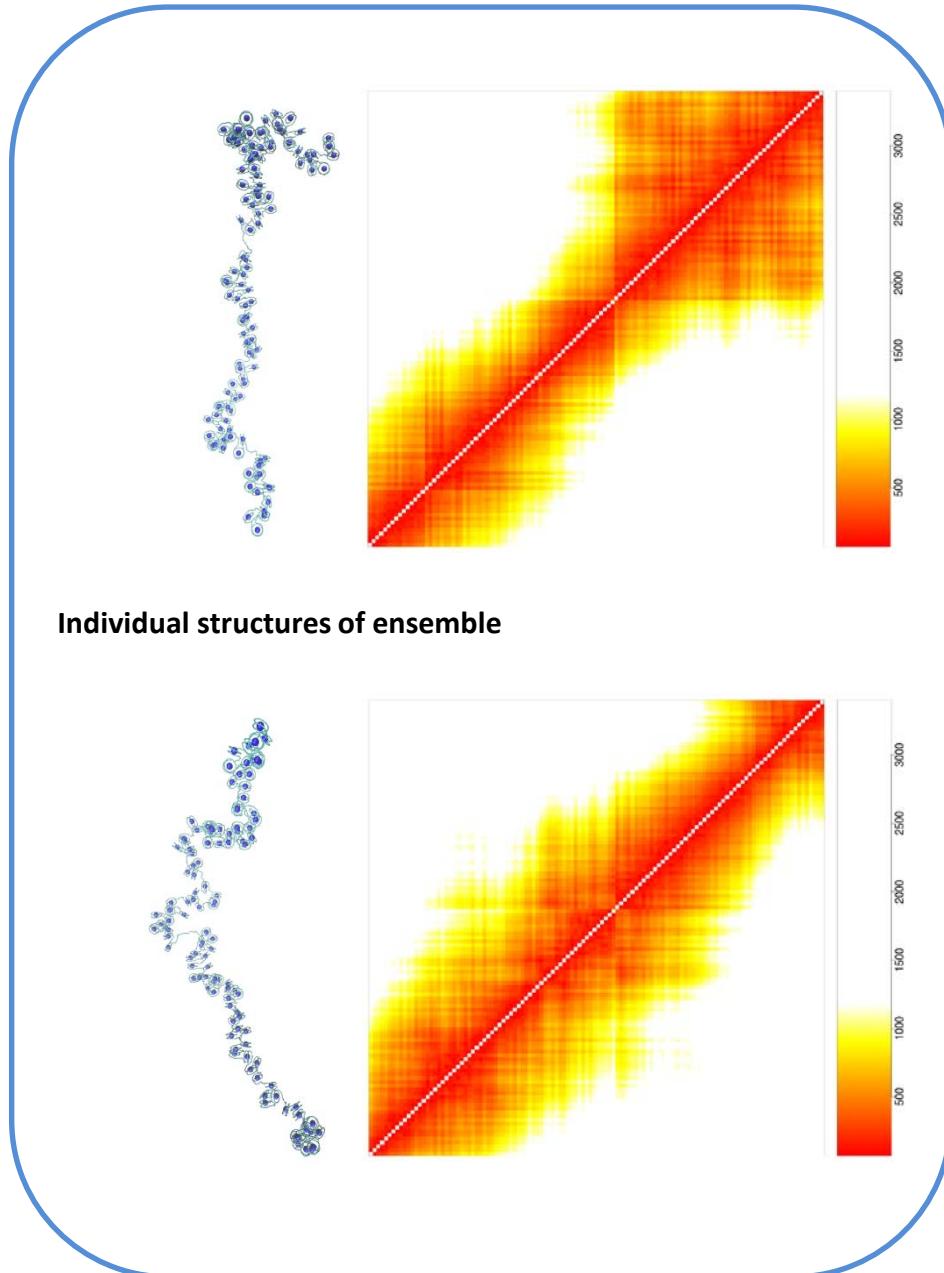
Starting structure



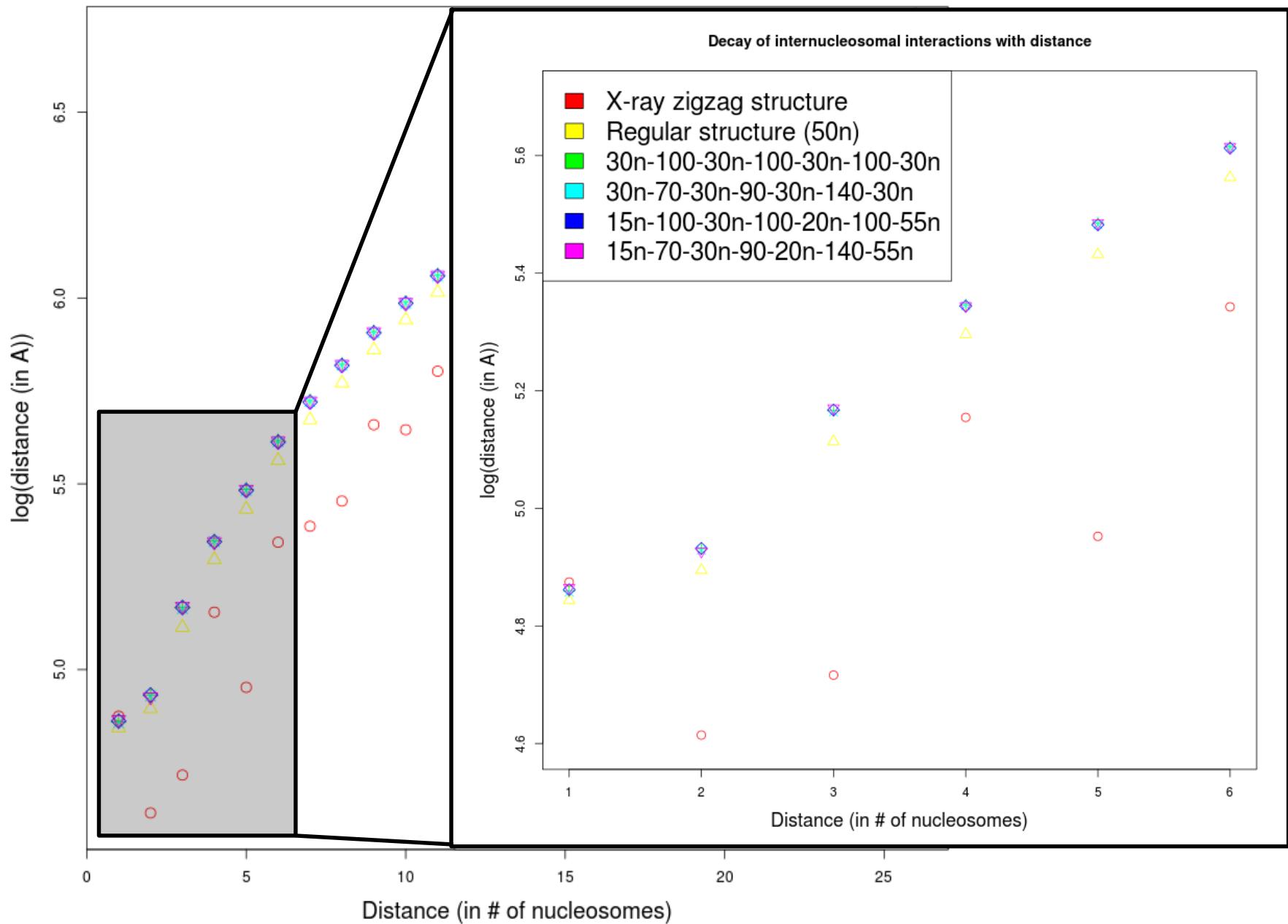
Ensemble average of nucleosome distances



Individual structures of ensemble



Decay of internucleosomal interactions with distance



How to use this tool

- Single sequence with nucleosome positions
- Multiple sequences with nucleosome positions to compare distance matrices, single contacts within the distance matrix, long range order ...

Give me a sequence
and I will make you
happy (not as quickly
as before).



Acknowledgements



Modesto Orozco



Pablo Dans

Collaborators



Jose Lluís Gelpí, Felipe Cano, Víctor López, Federica Battistini, Ricard Illa, Diana Buitrago



Marco Pasi (University of Nottingham)



Multiscale
Complex
Genomics

Richard Lavery (IBCP Lyon)

European Union, H2020, MuG: 676556



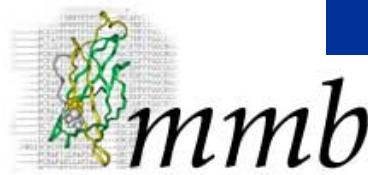
UNIÓN EUROPEA
“Una manera de hacer Europa”



MINISTERIO
DE ECONOMÍA
Y COMPETITIVIDAD



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



Molecular Modelling & Bioinformatics Group

MuG Team

