



Multiscale Complex Genomics

MC-DNA and Chromatin Dynamics

Jürgen Walther – 21.09.2018



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



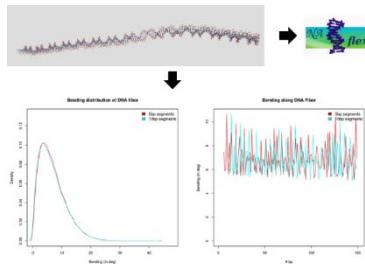
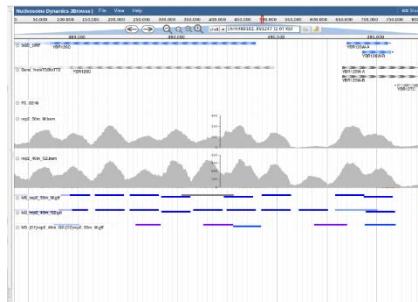


Outline

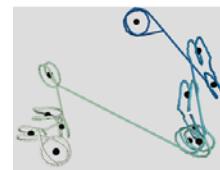
- 1) Introduction into DNA simulations
- 2) DNA model – “MC DNA”
- 3) Chromatin model – “Chromatin Dynamics”



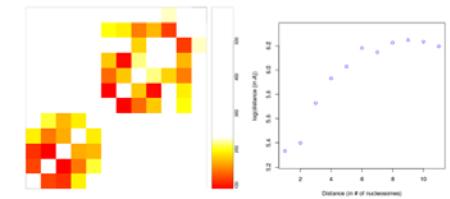
Storyline



3D structure

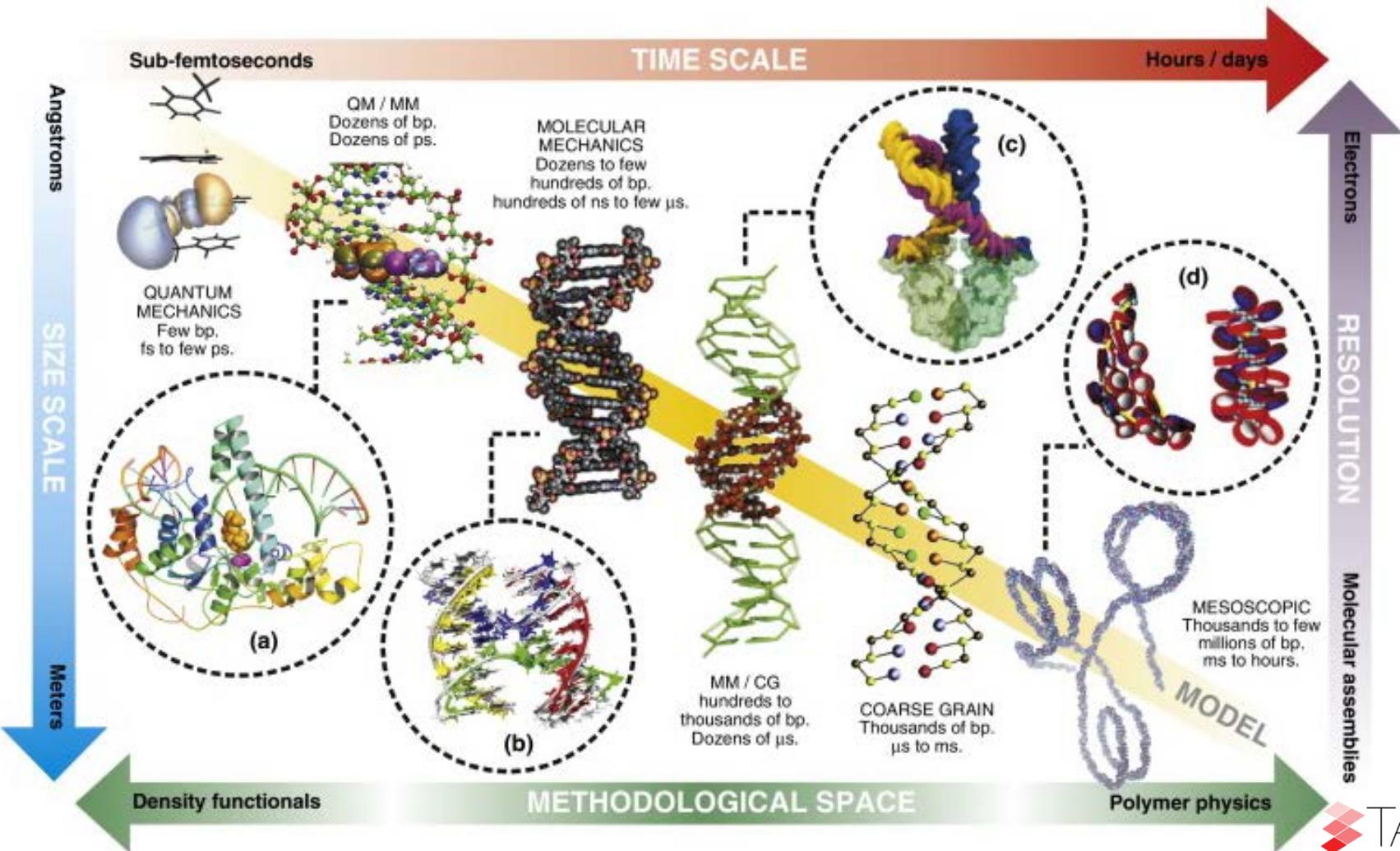


Distance matrix



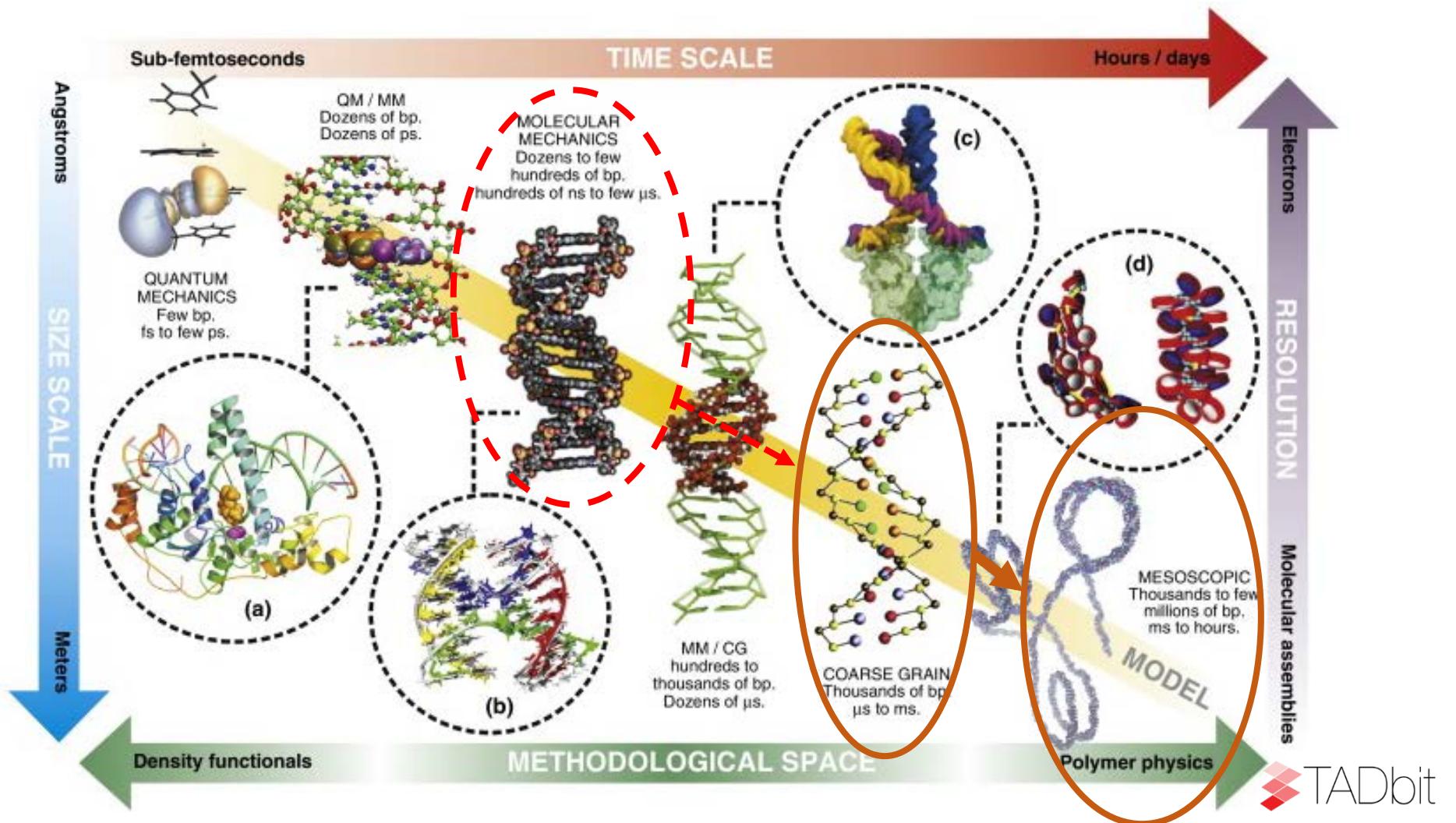


Biomolecular simulations: A multi-scale problem





Biomolecular simulations: A multi-scale problem





Biomolecular simulations: A multi-scale problem



Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Structural Biology



Multiscale simulation of DNA

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

DNA is not only among the most important molecules in life, but a meeting point for biology, physics and chemistry, being studied by numerous techniques. Theoretical methods can help in gaining a detailed understanding of DNA structure and function, but their practical use is hampered by the multiscale nature of this molecule. In this regard, the study of DNA covers a broad range of different topics, from sub-Angstrom details of the electronic distributions of nucleobases, to the mechanical properties of millimeter-long chromatin fibers. Some of the biological processes involving DNA occur in femtoseconds, while others require years. In this review, we describe the most recent theoretical methods that have been considered to study DNA, from the electron to the chromosome, enriching our knowledge on this fascinating molecule.

Addresses

¹ Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona, Spain

² Joint BSC-IRB Research Program in Computational Biology, Baldiri Reixac 10-12, 08028 Barcelona, Spain

³ Department of Biochemistry and Molecular Biology, University of Barcelona, 08028 Barcelona, Spain

Corresponding author: Orozco, Modesto
(modesto.orozco@irbbarcelona.org)

in the day time-scale (10^5 s); the local breathing of nucleobases occurs in the millisecond range (10^{-3} s), while electronic rearrangements take place in the sub-femtosecond time-scale ($<10^{-15}$ s).

During the last years we have witnessed the development of a wide repertoire of theoretical methods that aimed to reproduce the properties of DNA, either isolated or protein bound. Even if primitive, these methods allow researchers to consider the DNA at different resolution levels, and provide information of great value on the structure, dynamics, and interactions of this fascinating molecule. We will briefly summarize some of these most recent theoretical approaches, focusing our analysis on the contributions of the last three years, when the field has experienced a significant improvement.

For the sake of simplicity, throughout this manuscript we will classify theoretical methods in four groups, according to their level of resolution (Figure 1): firstly, electronic, secondly, atomistic, thirdly, coarse grained, and lastly, mesoscopic. It is worth noting that moving in the resolution space means moving also in the methodological space, since the basic physical models underlying the

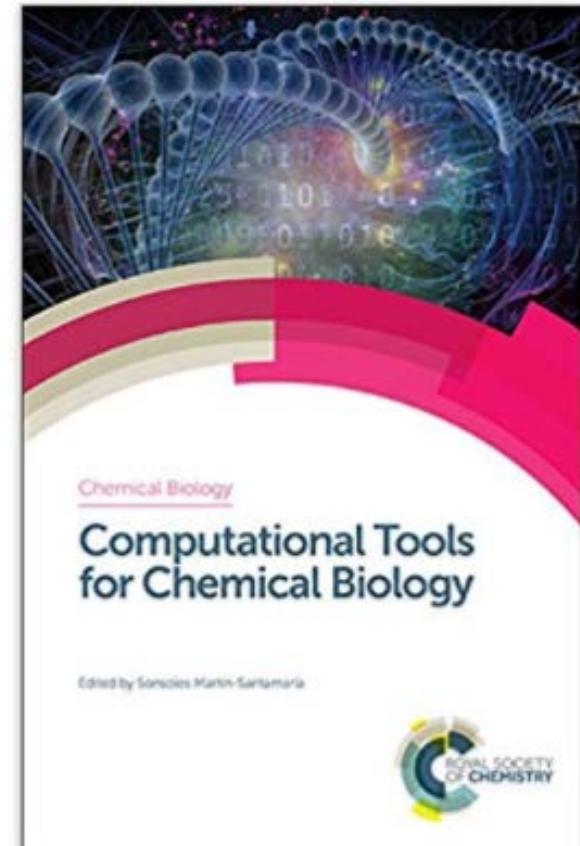
Multiscale simulation of DNA (2016)

PD Dans, J Walther, H Gómez, M Orozco

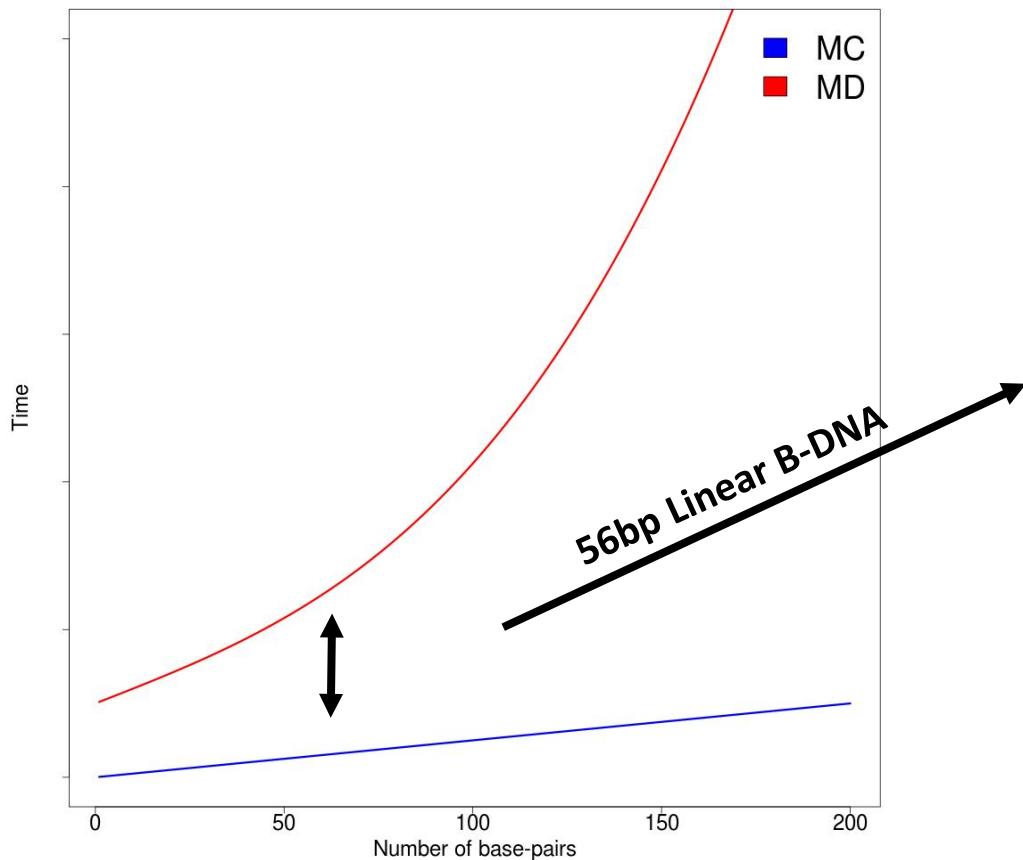
Current opinion in structural biology 37, 29-45

Molecular Modeling of Nucleic Acids (2017)

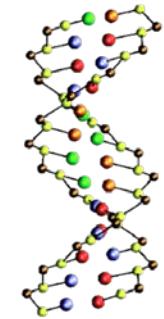
H Gómez, J Walther, L.Darré, I Ivani, PD Dans, M Orozco. In Computational Tools for Chemical Biology. RSC, ISBN: 1782627006



Calculation time: coarse-grained DNA (MC) vs atomistic DNA (MD) simulations



10K structures in
10 min
(1 CPU)



vs

650,000 CPU
hours in the
MareNostrum
supercomputer





Multiscale Complex Genomics

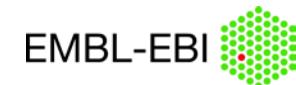


DNA model - MC DNA

A fast method to accurately probe DNA properties at base-pair level



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

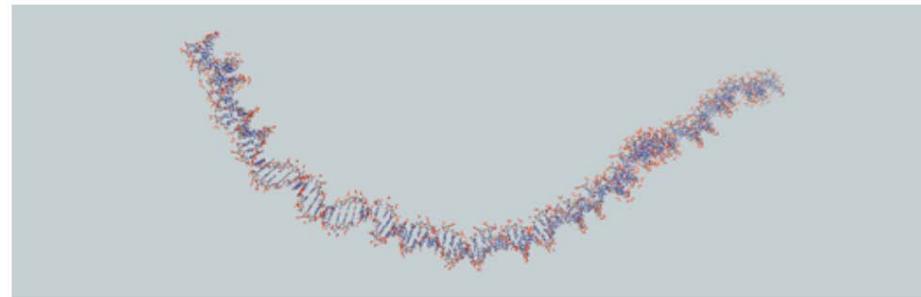




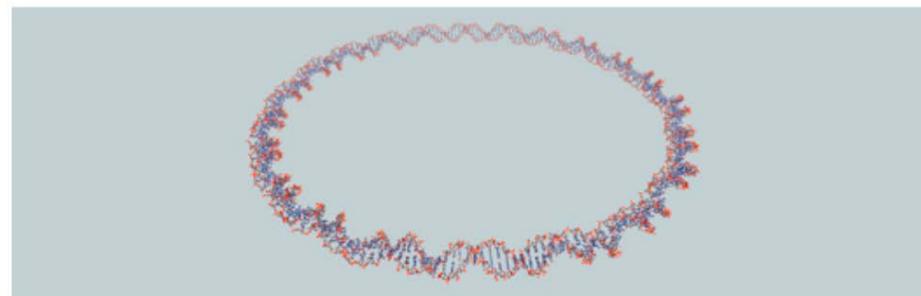
DNA

<https://mmbrbarcelona.org/MCDNA/>

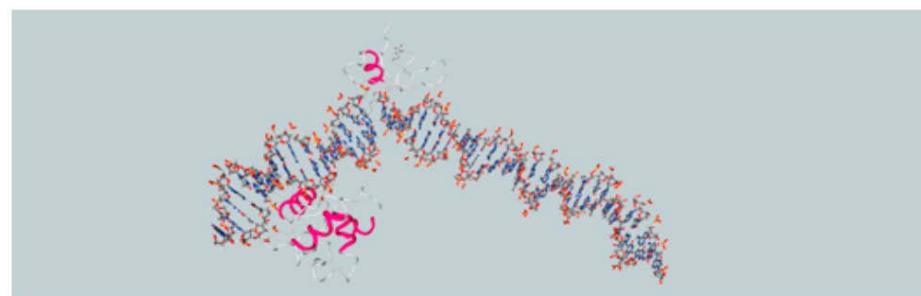
Free linear



Circular



Protein-coated





<https://mmb.irbbarcelona.org/MCDNA/>

147 users
engaged in
1279 sessions

Country	Users	% Users
1. Spain	91	50.56%
2. United States	31	17.22%
3. United Kingdom	11	6.11%
4. India	8	4.44%
5. Switzerland	7	3.89%
6. Germany	4	2.22%
7. France	4	2.22%
8. Netherlands	3	1.67%
9. Russia	3	1.67%
10. Sweden	3	1.67%



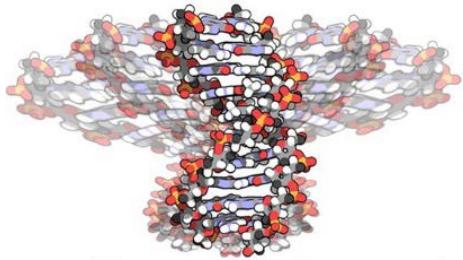
[BioExcel Webinar](#)
[Series #23: MC_DNA](#)

145 views of webinar

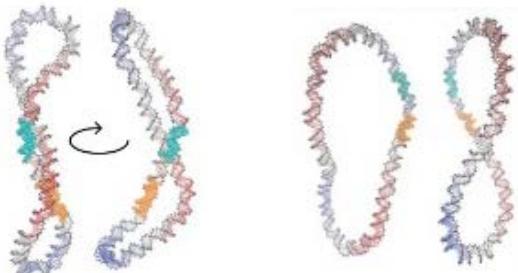


DNA

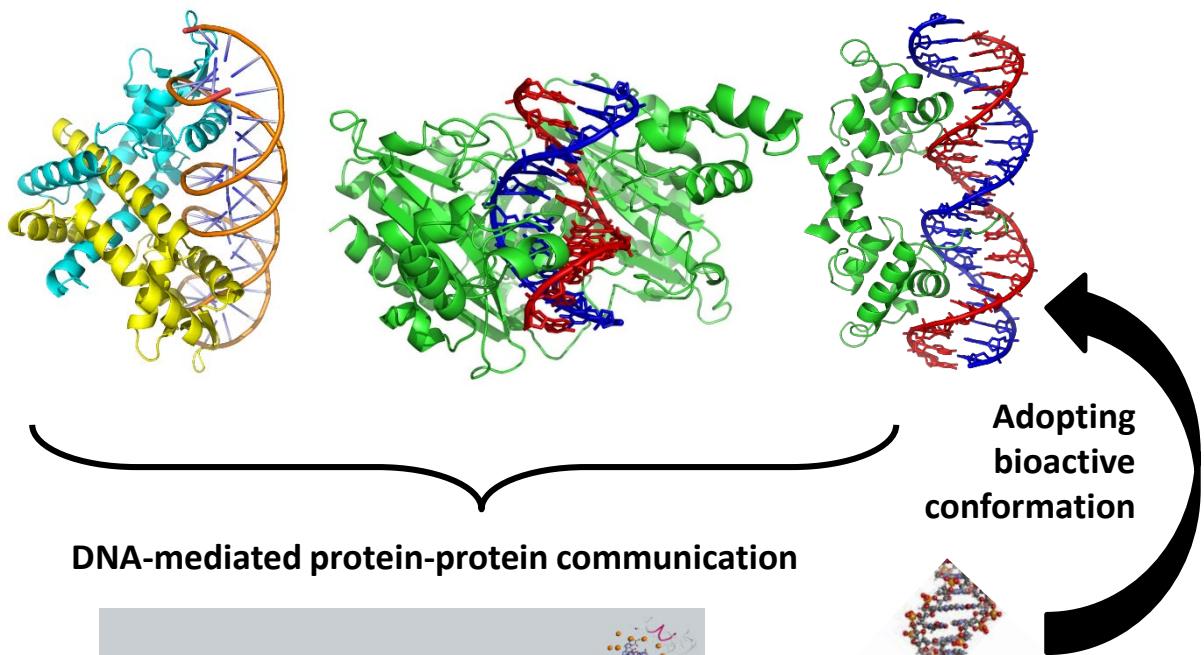
**Dynamics of free linear DNA
(MC_DNA)**



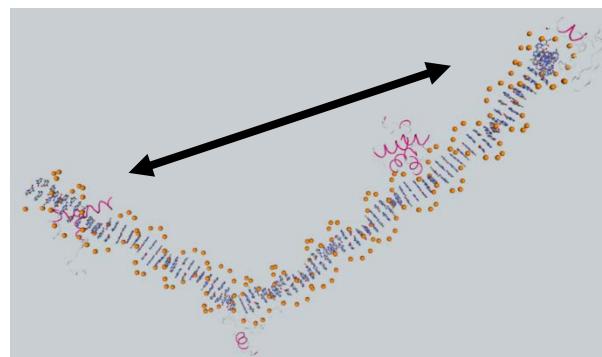
Distant contacts in a constrained environment (circular MC_DNA)



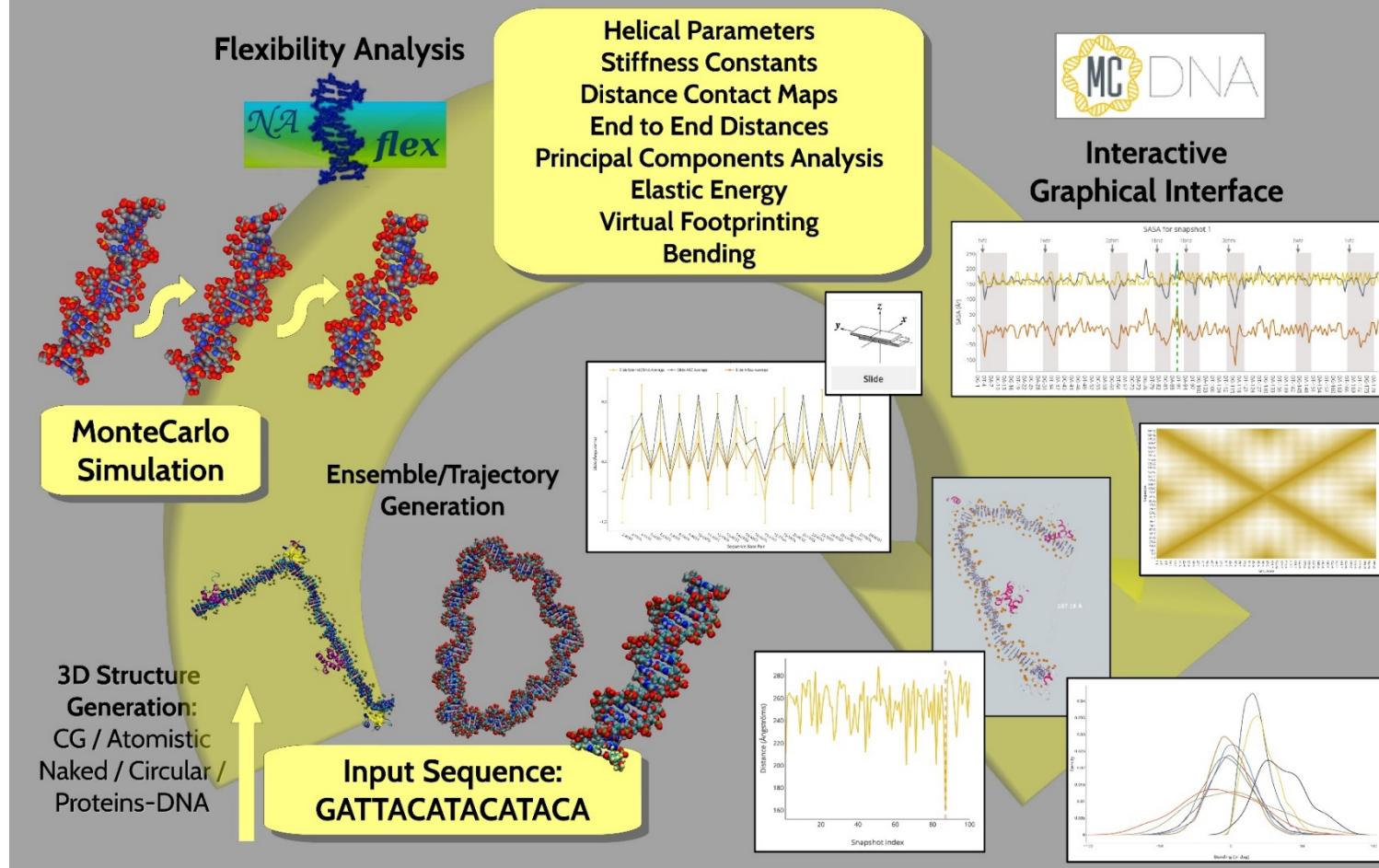
Accessibility of a protein-coated DNA fiber (MC_DNA + proteins)



DNA-mediated protein-protein communication



MCDNA: MonteCarlo Coarse-Grained Simulations.

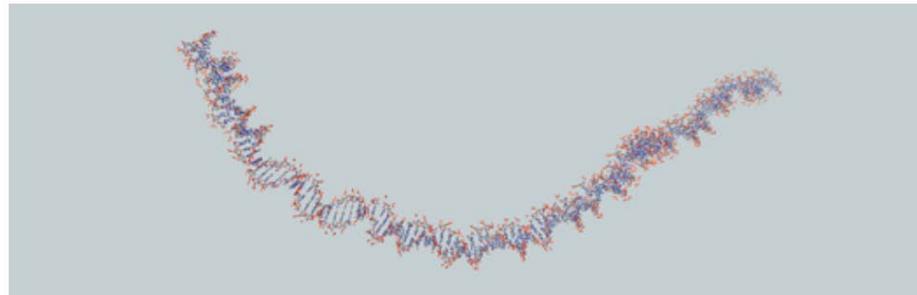




DNA

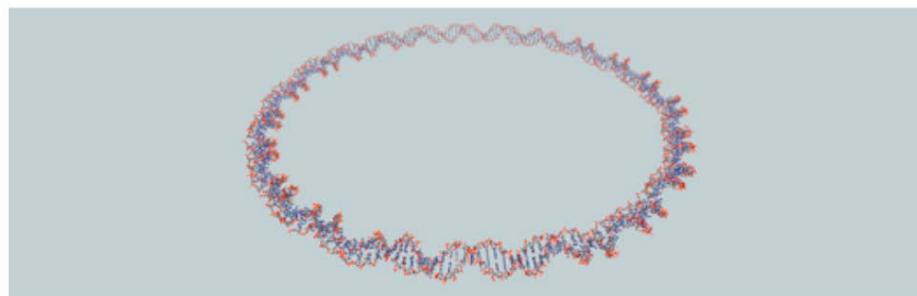
<https://mmb.irbbarcelona.org/MCDNA/>

Free linear

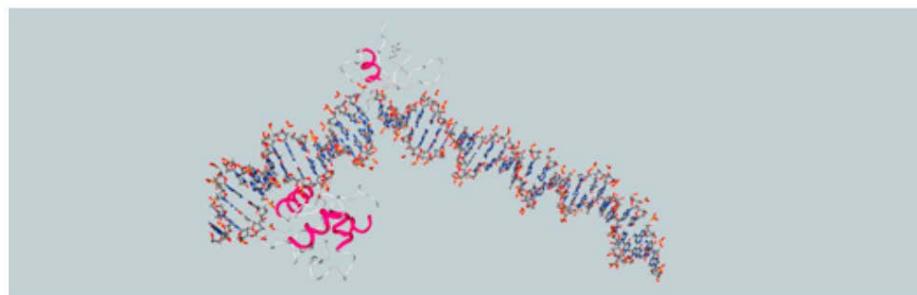


Multiscale
Complex
Genomics

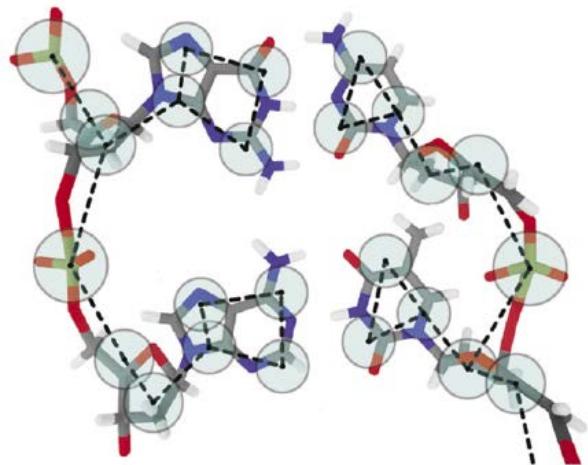
Circular



Protein-coated



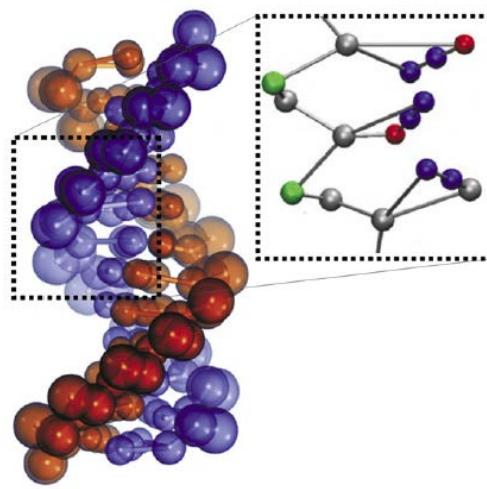
Particle-based



Martini

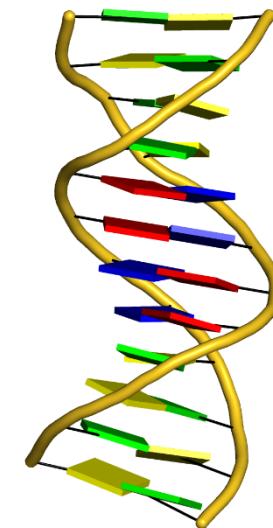
1 bead backbone
2 beads sugar
3 beads Y
4 beads R
(6/7 total).

Internal space-based

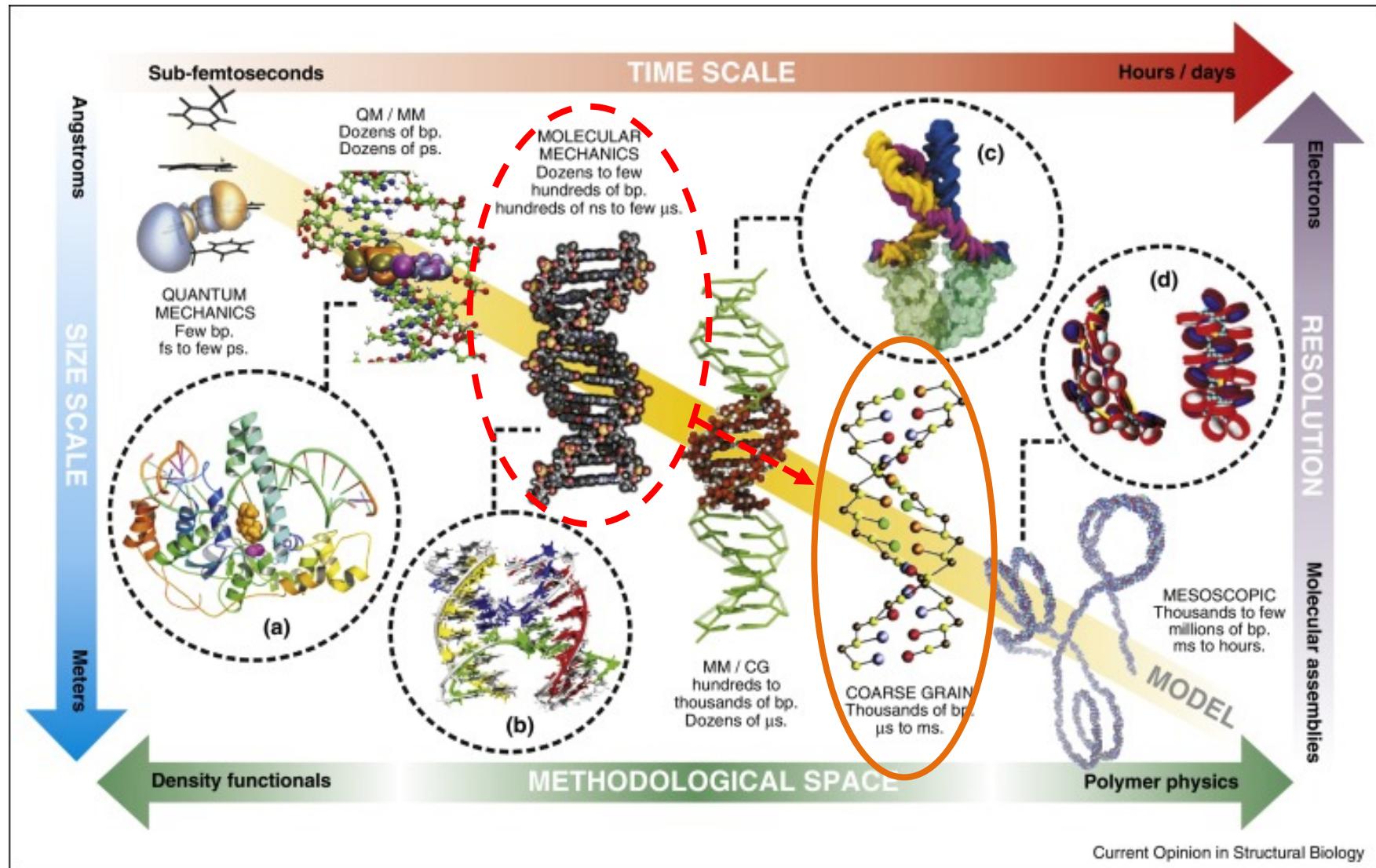


SIRAH

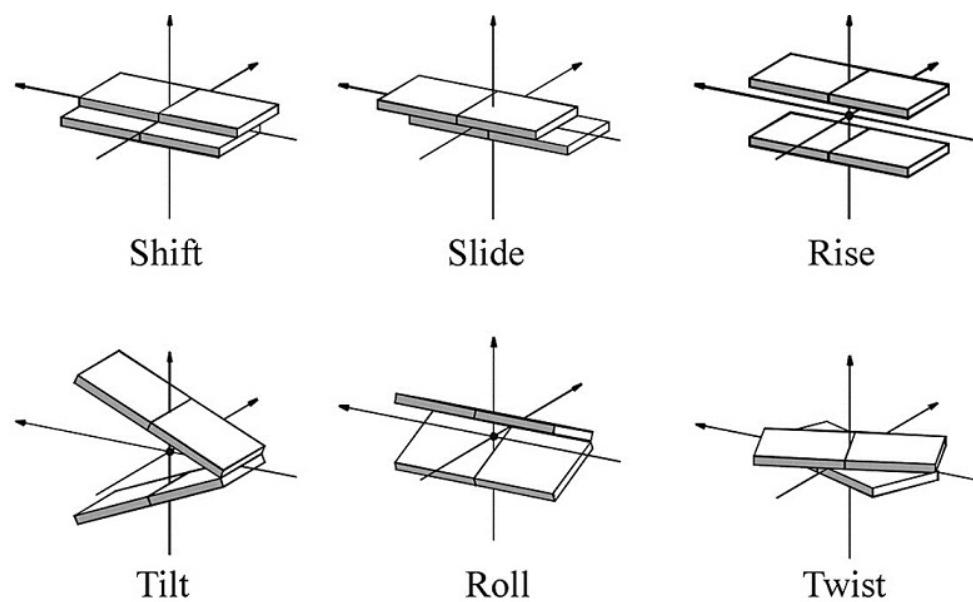
2 beads backbone
1 bead sugar
3 beads base
(6 total)



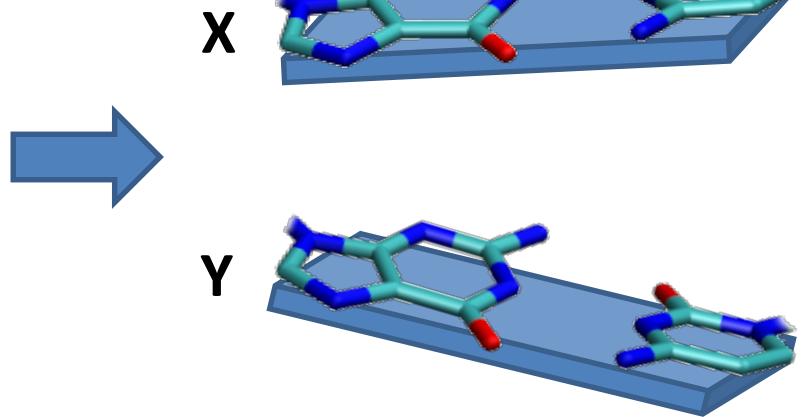
Bases or base pairs are rigid objects and behave based on coordinates in an internal conformational space



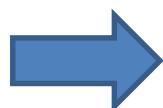
From Cartesian to Helical space



Base-pair step parameters (6)



Relative orientation and position between adjacent base-pairs



X^0

Equilibrium values for each of the 136 unique tetranucleotides (taken as the average of a given helical parameter along time).

Obtaining the force constants from the covariance matrix in the helical space

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 11163–11168, September 1998
Biophysics

DNA sequence-dependent deformability deduced from protein-DNA crystal complexes

WILMA K. OLSON^{*†}, ANDREY A. GORIN[‡], XIANG-JUN LU^{*}, LYNETTE M. HOCK^{*}, AND VICTOR B. ZHURKIN^{†§}

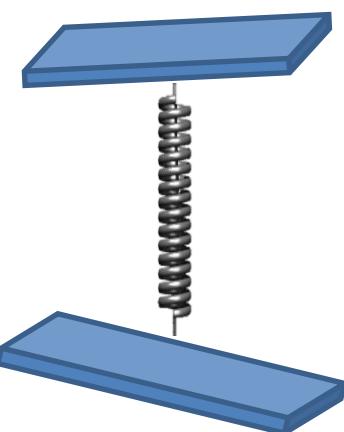
^{*}Department of Chemistry, Rutgers University, New Brunswick, NJ 08903; [‡]Sloan-Kettering Cancer Center, New York, NY 10021; and [§]National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

Communicated by Donald M. Crothers, Yale University, New Haven, CT, June 30, 1998 (received for review April 13, 1998)

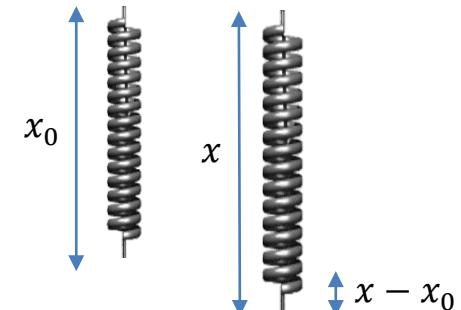
$$K^i = k_B T C_h^{-1} = \begin{bmatrix} k_{\text{twist}} & k_{i-r} & k_{i-l} & k_{i-i} & k_{i-s} & k_{i-d} \\ k_{i-r} & k_{\text{roll}} & k_{r-l} & k_{r-i} & k_{r-s} & k_{r-d} \\ k_{i-l} & k_{r-l} & k_{\text{tilt}} & k_{l-i} & k_{l-s} & k_{l-d} \\ k_{i-i} & k_{r-i} & k_{l-i} & k_{\text{rise}} & k_{i-s} & k_{i-d} \\ k_{i-s} & k_{r-s} & k_{l-s} & k_{i-s} & k_{\text{shift}} & k_{s-d} \\ k_{i-d} & k_{r-d} & k_{l-d} & k_{i-d} & k_{s-d} & k_{\text{slide}} \end{bmatrix}$$

$$E_{\text{el}}^i = (x_i - x_i^0)^T K^i (x_i - x_i^0) \quad x_i \in \mathbb{R}^6, K_i \in \mathbb{R}^{6 \times 6}$$

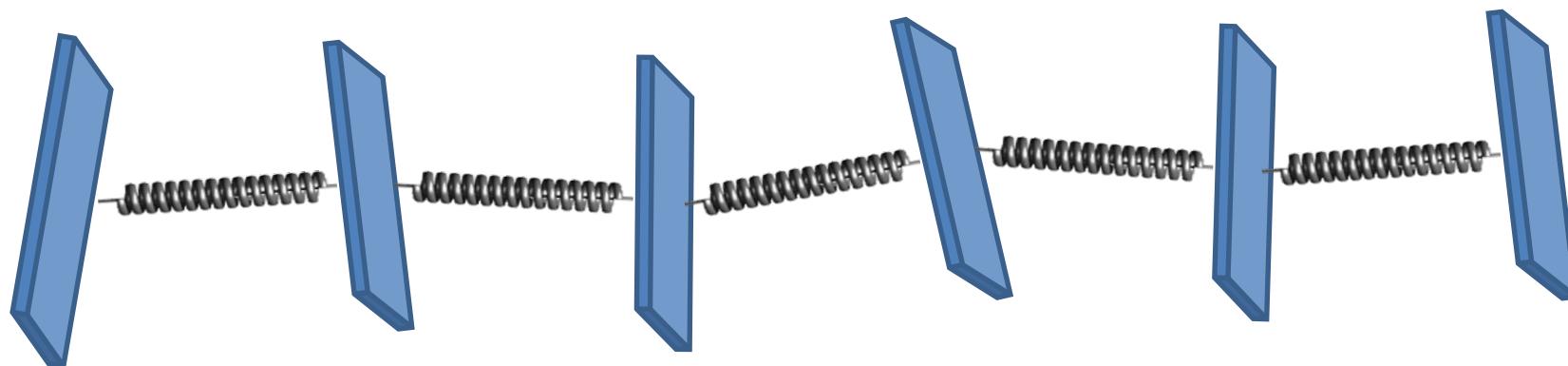
DNA is a spiral staircase - each step a base-pair



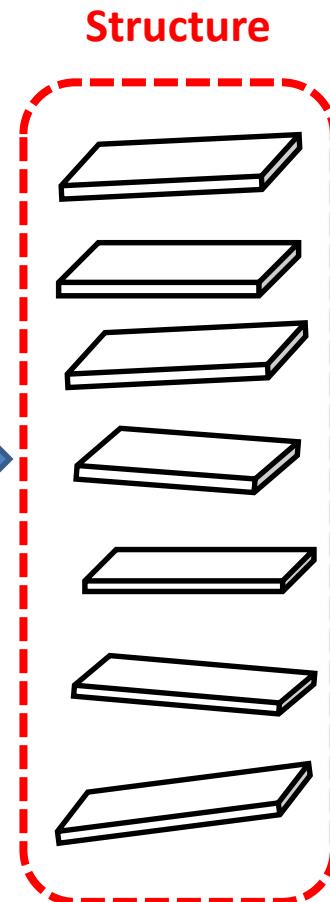
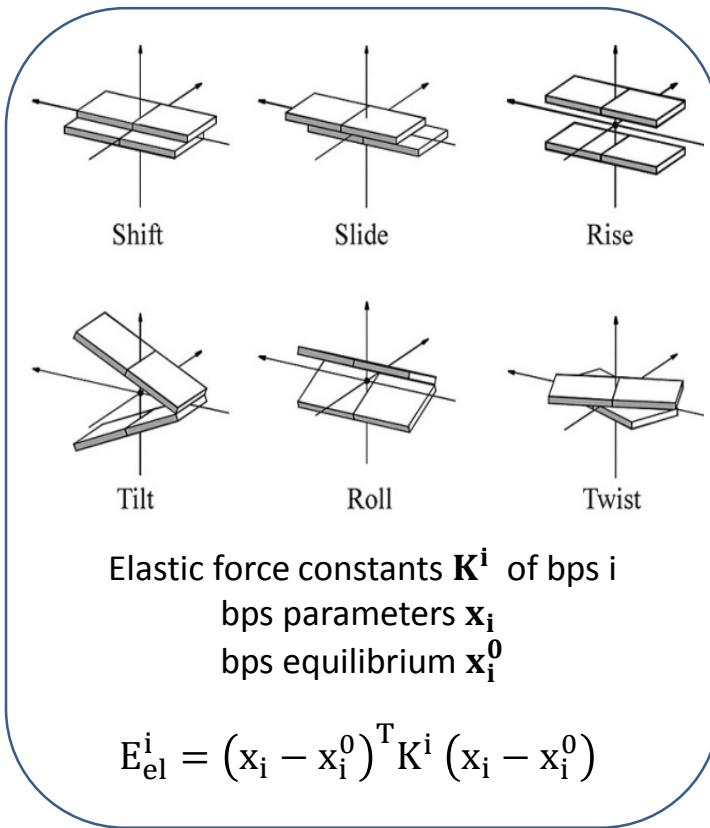
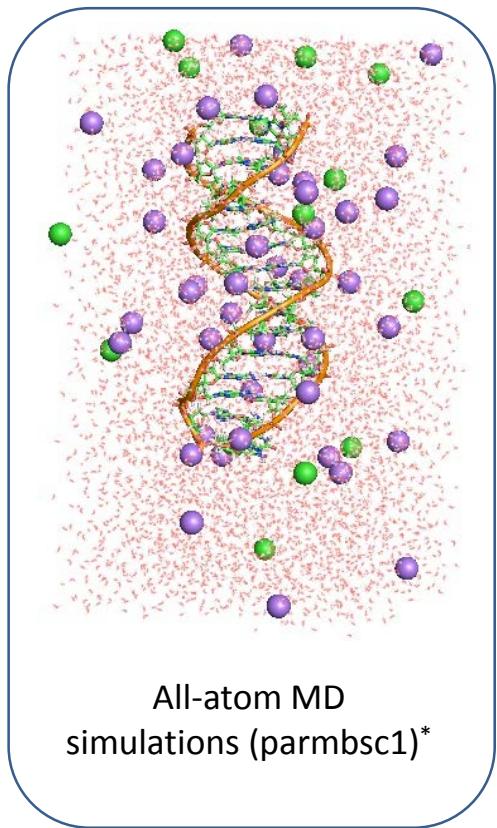
Interaction between
two adjacent base
pairs described by
virtual spring with
stiffness k



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



MC_DNA - The method: next-nearest neighbor model



Sequence-specific effects (NN model) included in elastic force constants K

MC_DNA - The method

Stochastic
molecular
simulation



Metropolis Monte Carlo

Randomly change
helical parameters
of random bp-step k

$$x_k \rightarrow x_{k \text{ new}}$$

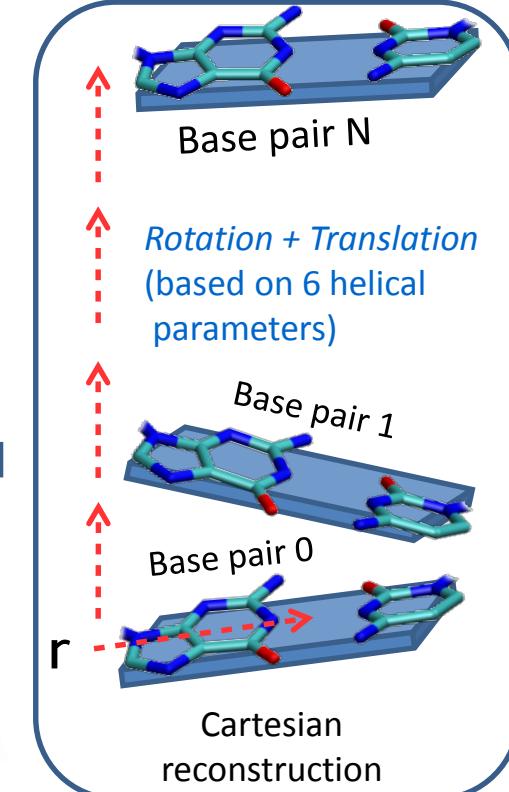
Calculate change
in potential energy

$$\Delta E = x_{k \text{ new}}^T K x_{k \text{ new}} - x_k^T K x_k$$

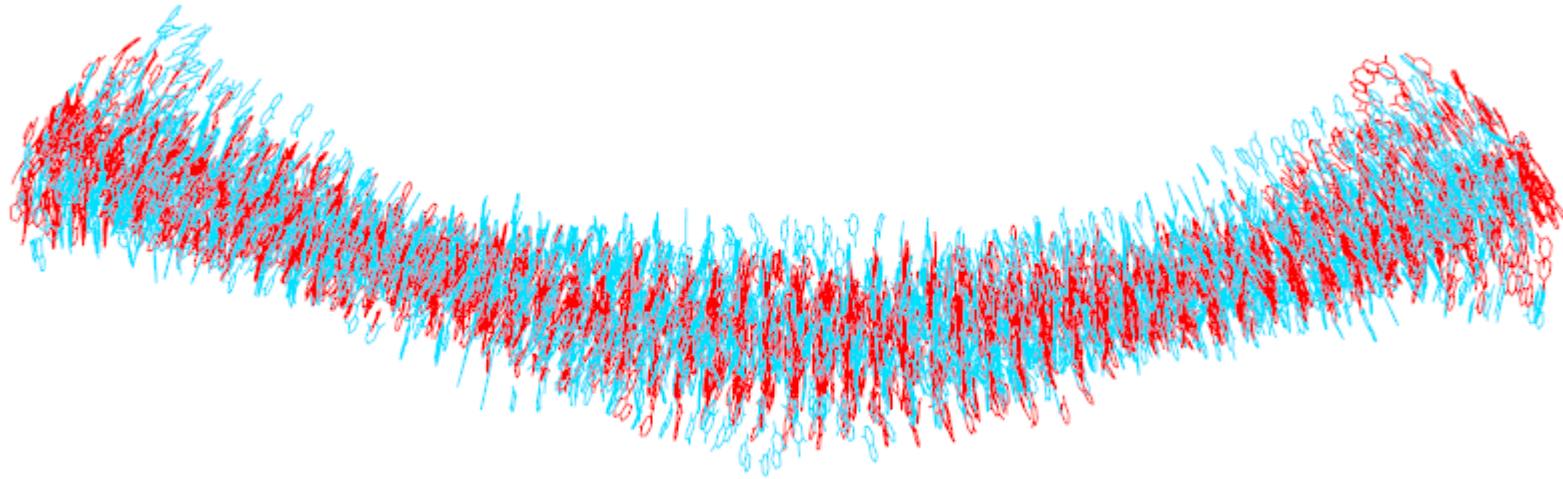
accept $x_{k \text{ new}}$ keep x_k
yes no

$\Delta E \geq 0:$ $e^{-\Delta E/kT} > R ??$
 $\Delta E < 0:$ accept $x_{k \text{ new}}$

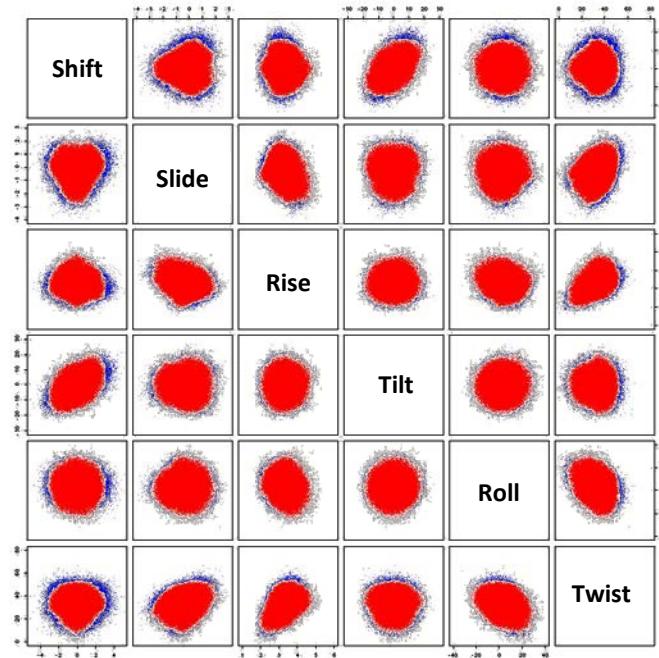
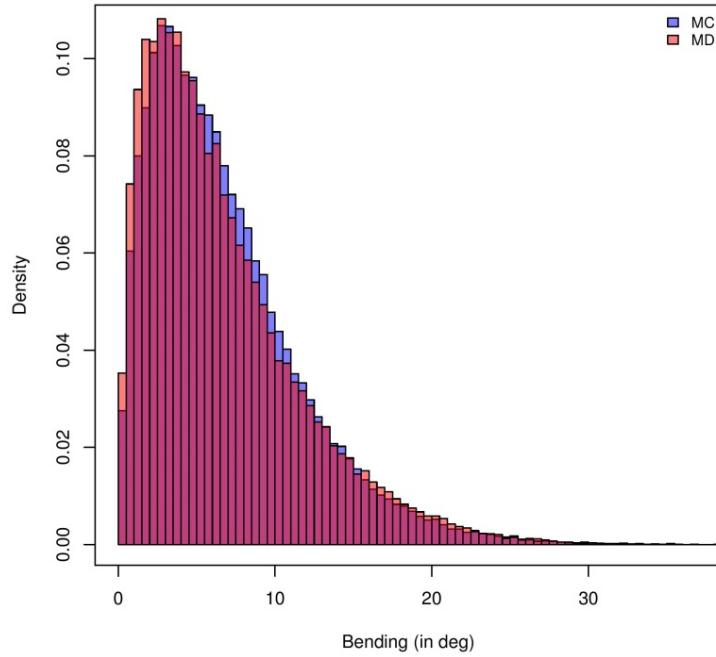
Trajectory
Ensemble of
DNA
structures



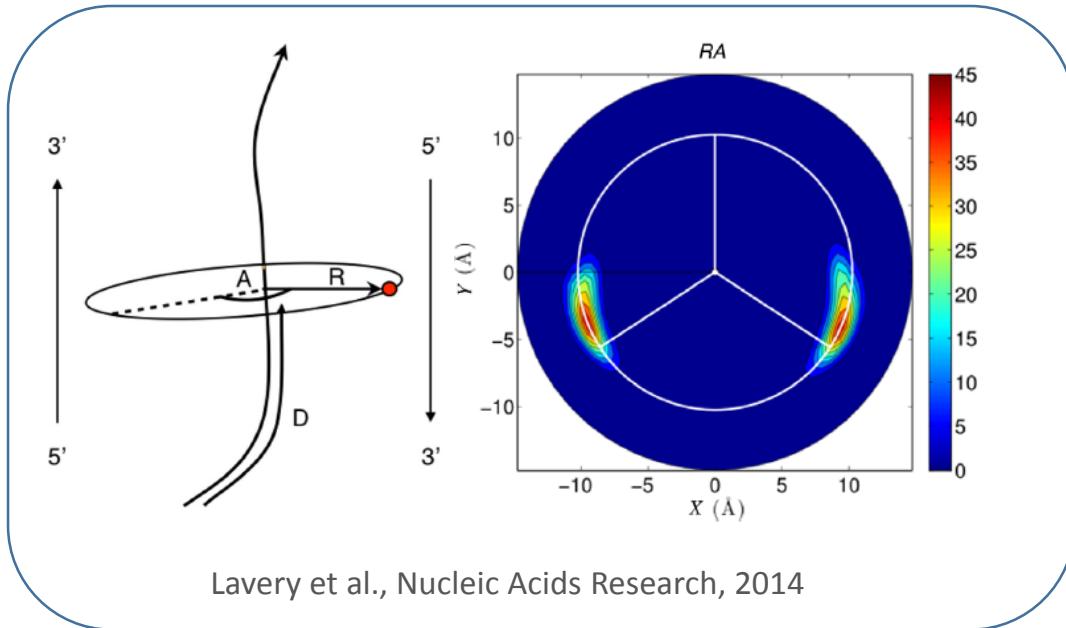
Validation by comparison with all atom **MD**



Bending of 10bp pieces

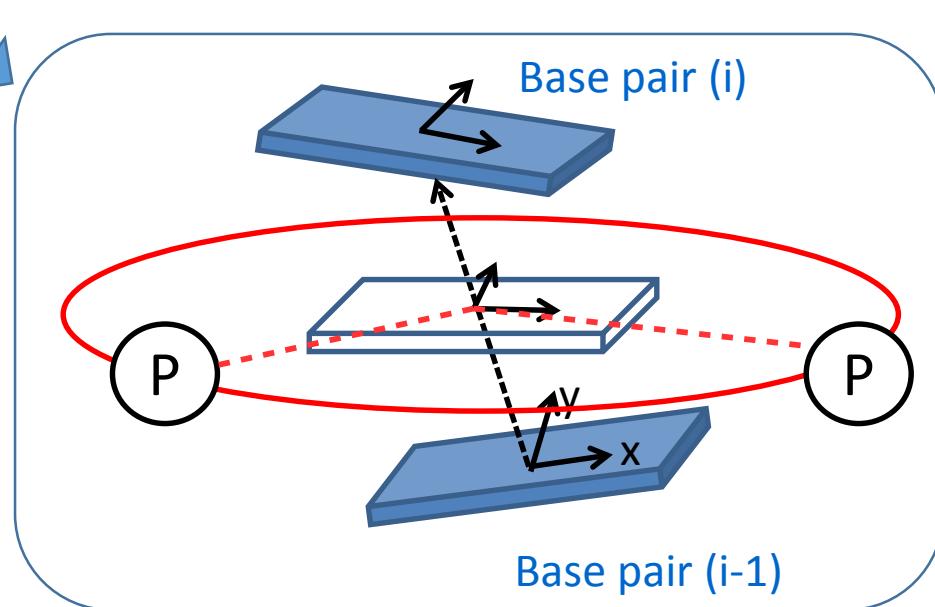


Inferring phosphate position from helical parameters



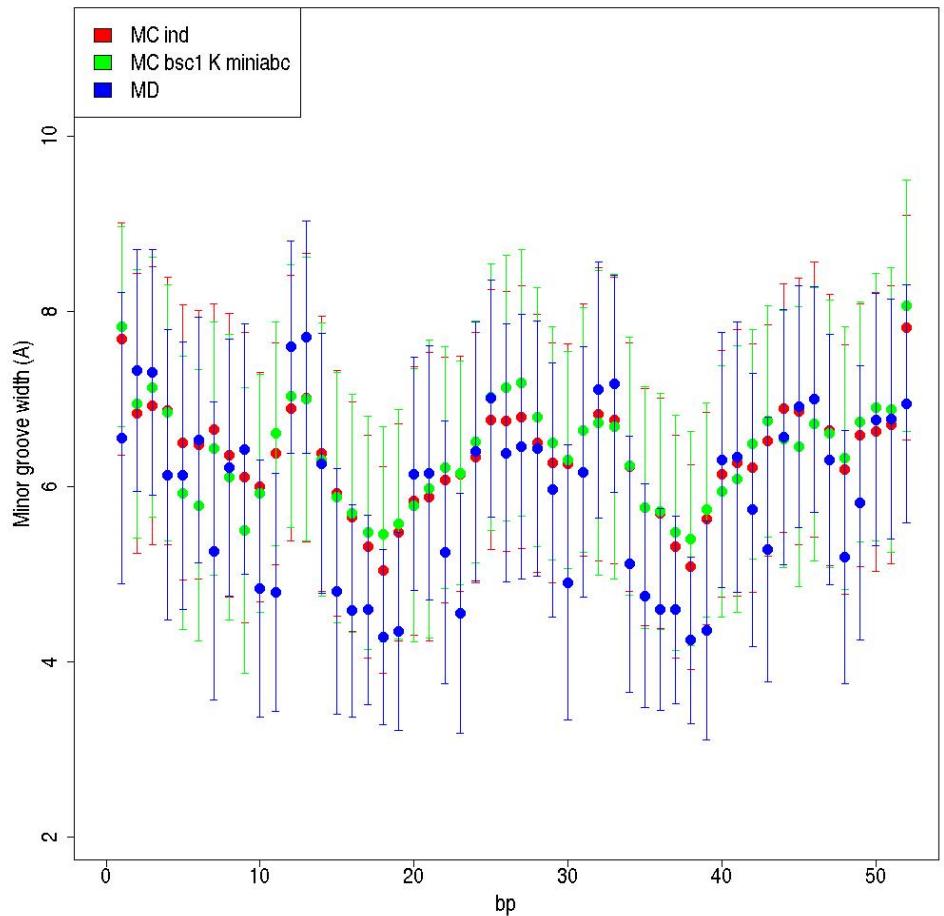
Lavery et al., Nucleic Acids Research, 2014

Position based on concentration maps
Computed with Curves+canion

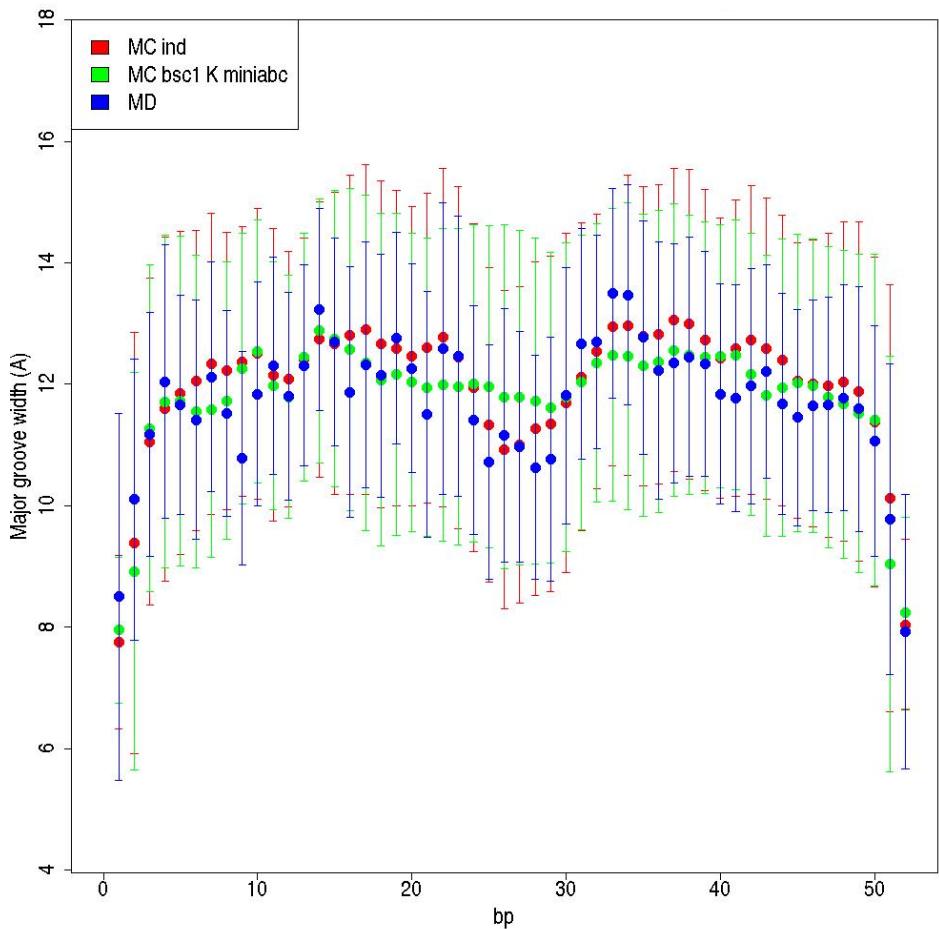


Major and minor groove width

Minor groove



Major groove



CGCCGGCAGTAGCCGAAAAAAATAGGCGCGCGCTCAAAAAAATGCCCATGCCGCGC

CGCCGGCAGTAGCCGAAAAAAATAGGCGCGCGCTCAAAAAAATGCCCATGCCGCGC

Creating kinetic series of the simulation





Workflow

Input: DNA sequence in txt file (f.ex: ACGTGCTAATCGCGCGCGTATCTAGCTA)

Create Structure: Creates a single structure of DNA in a relaxed state

Create Trajectory: Creates a certain number of DNA structures



User case

Genomic region

chrII:489181..491246

(Gene TPS1 with its promoter region)

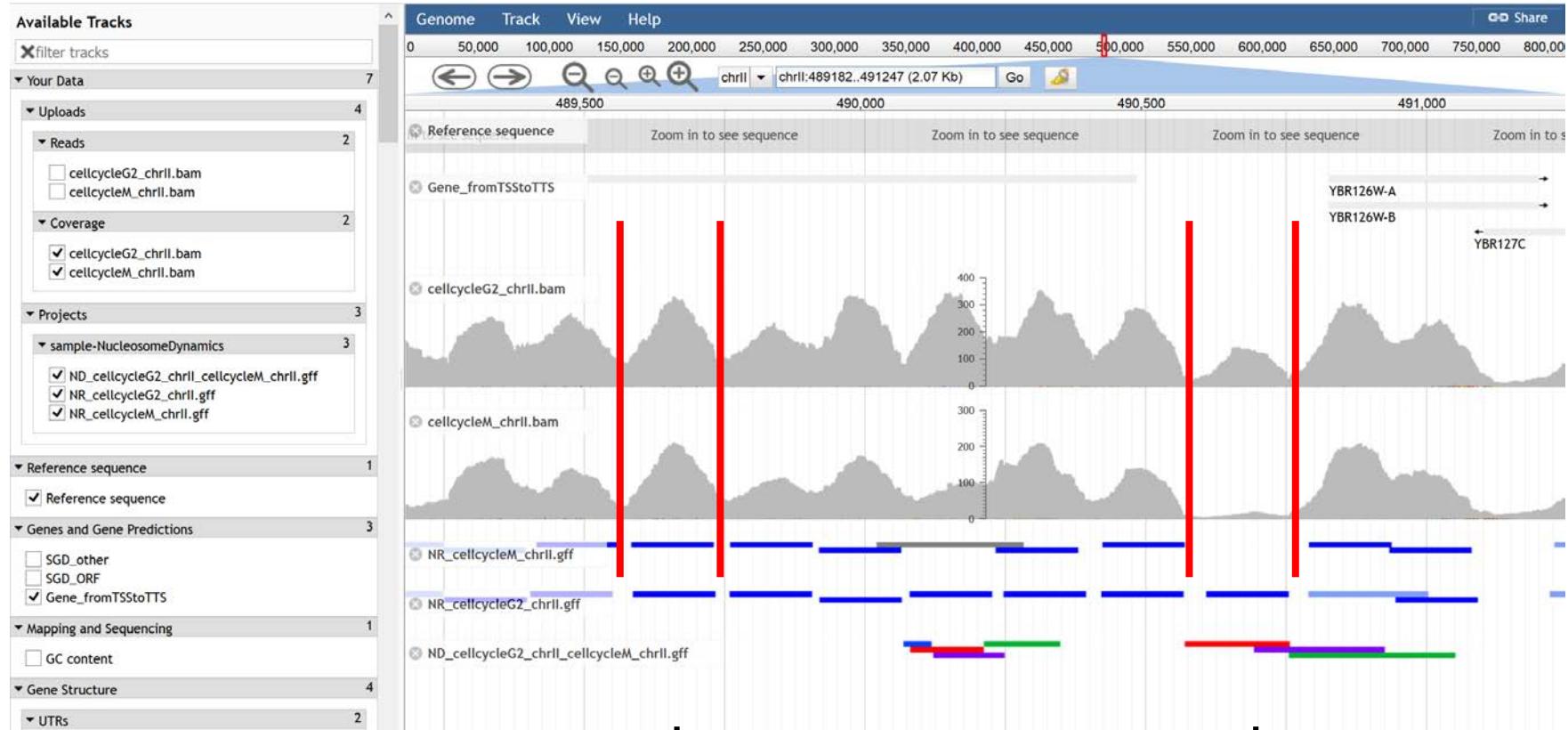
- Visualize genomic region with outputs from Nucleosome Dynamics in Jbrowse
- Select sequence to simulate

MC-DNA_noNucl
MC-DNA_nucl



User case

chrII:489181..491246



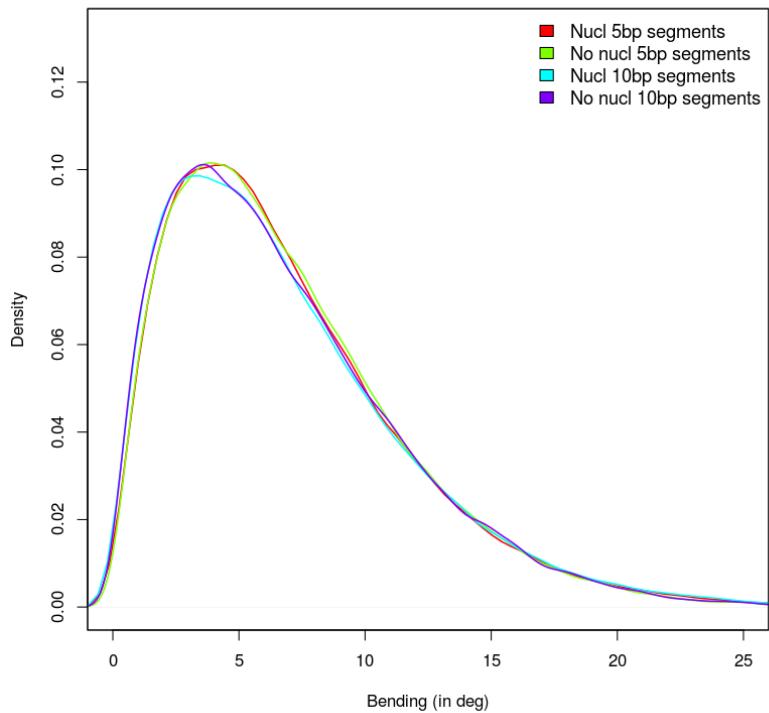
seq_nucl.txt



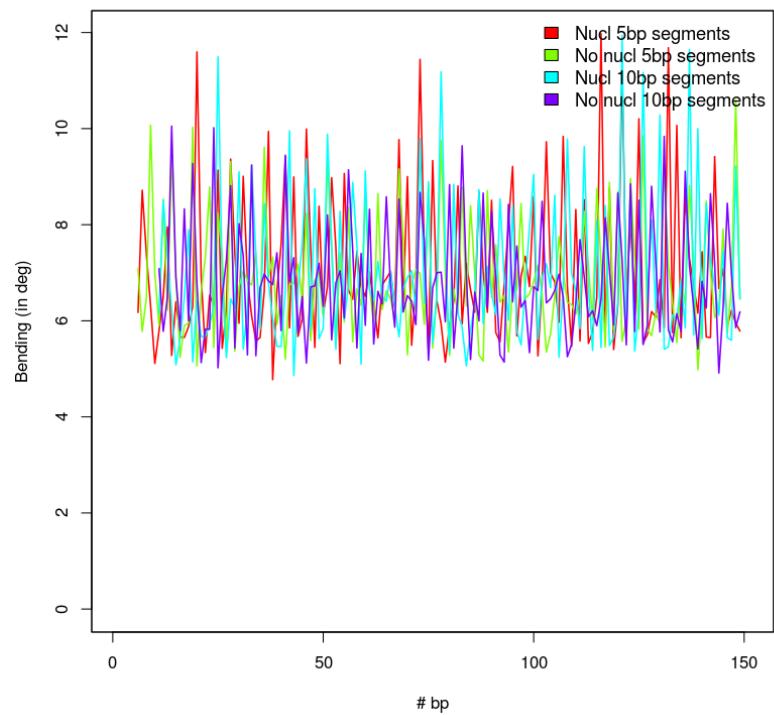
seq_no_nucl.txt

Combining results for trajectory for seq_nucl.txt and seq_no_nucl.txt *(not part of the VRE)*

Bending distribution of DNA fiber



Bending along DNA Fiber





Workflow

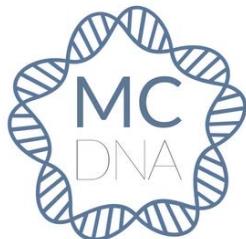
Virtual Research Environment Test

Homepage User Workspace Get Data External Links Forum User Admin This is a BETA version of MuG VRE

							Tools	Actions
	chromdyn_str.pdb	PDB	workshop_chromdyn_50m_12N	2017/03/31 15:05	177.34 K			
	workshop_chromdyn_gff_G2	PDB	workshop_chromdyn_gff_G2	2017/04/01 16:40	177.65 K			
	chromdyn_str.pdb	PDB	workshop_chromdyn_gff_G2	2017/04/01 16:40	149.50 K			
	workshop_chromdyn_gff_M	PDB	workshop_chromdyn_gff_M	2017/04/01 16:30	172.18 K			
	chromdyn_str.pdb	PDB	workshop_chromdyn_gff_M	2017/04/01 16:30	148.39 K			
	workshop_mc-dna_no_nucl	PDB	workshop_mc-dna_no_nucl	2017/03/31 17:05	145.45 M			
	mc_dna_eq_str.pdb	PDB	workshop_mc-dna_no_nucl	2017/03/31 16:56	625.32 K			
	mc_dna_str.dcd	MDCRD	workshop_mc-dna_no_nucl	2017/03/31 17:05	54.70 M			
	mc_dna_str.pdb	PDB	workshop_mc-dna_no_nucl	2017/03/31 17:05	625.32 K			
	mc_dna_str.top	PARMTOP	workshop_mc-dna_no_nucl	2017/03/31 17:05	3.62 M			
	workshop_mc-dna_nucl	PDB	workshop_mc-dna_nucl	2017/03/31 17:06	146.39 M			
	mc_dna_eq_str.pdb	PDB	workshop_mc-dna_nucl	2017/03/31 16:56	627.48 K			
	mc_dna_str.dcd	MDCRD	workshop_mc-dna_nucl	2017/03/31 17:05	54.89 M	MD Energy Refinement		
	mc_dna_str.pdb	PDB	workshop_mc-dna_nucl	2017/03/31 17:05	627.48 K	NAFlex		
	mc_dna_str.top	PARMTOP	workshop_mc-dna_nucl	2017/03/31 17:05	3.63 M	pyDock DNA		

Showing 261 to 277 of 277 entries

x 10 11 12 13 14





Multiscale Complex Genomics

Chromatin Dynamics



Implementing nucleosomes into MC DNA



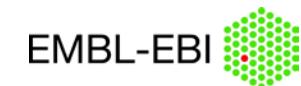
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.



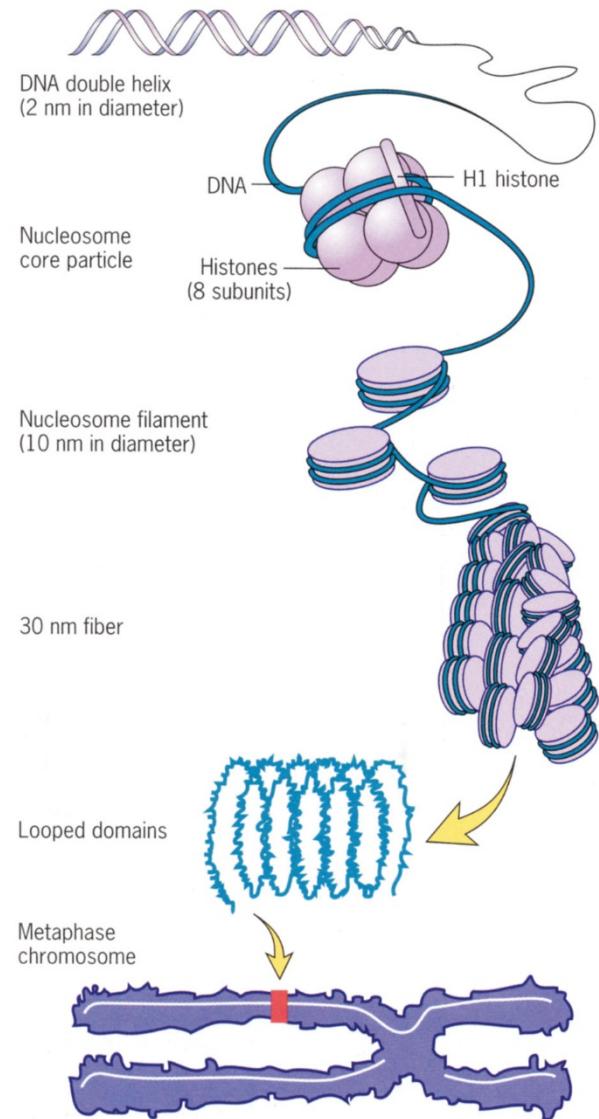
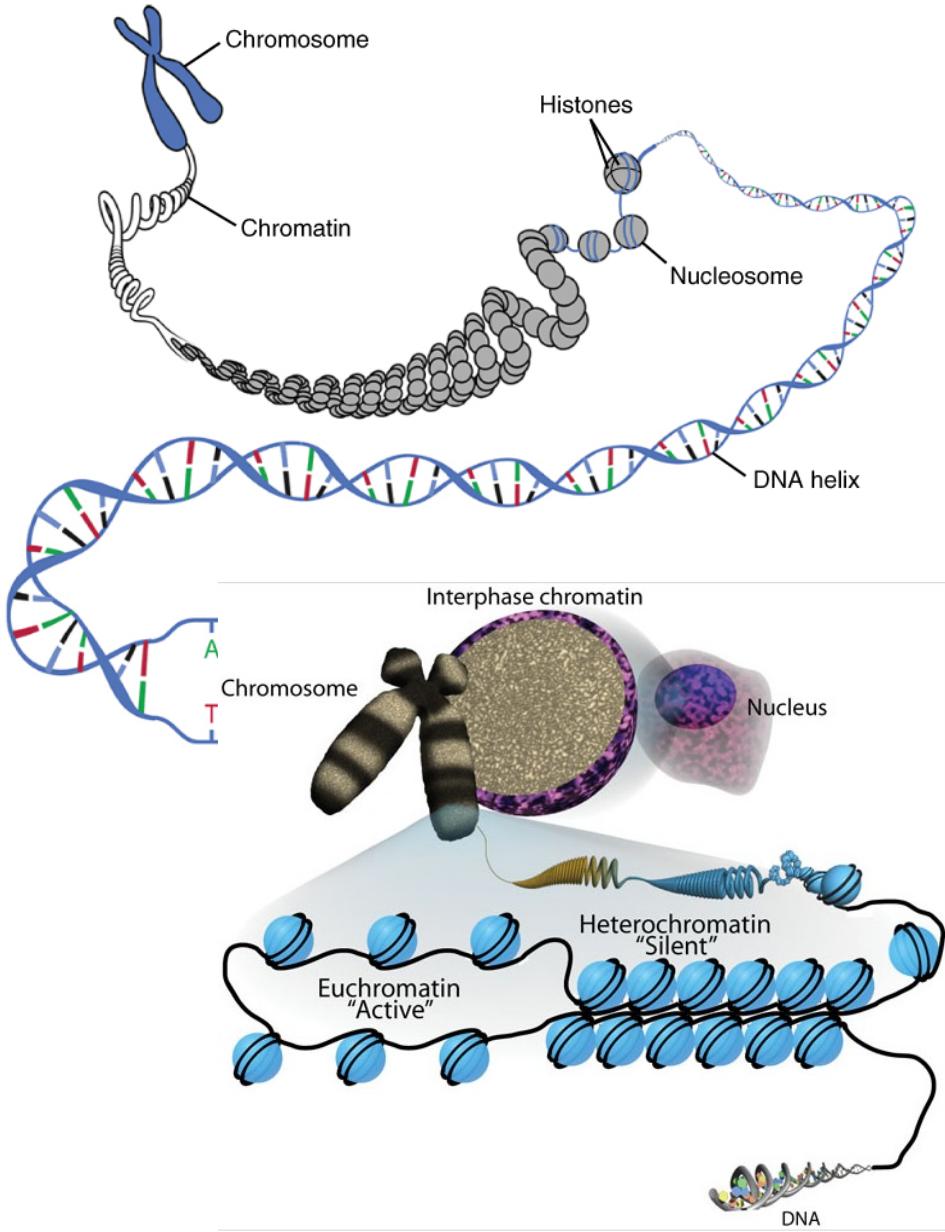
INSTITUTE
FOR RESEARCH
IN BIOMEDICINE



UNITED KINGDOM • CHINA • MALAYSIA



Chromatin – a 30nm fiber?

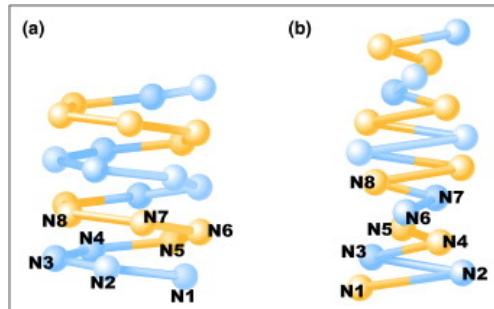


Does the 30-nm fibre exist *in vivo*?

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)

TH. Hsieh et al. *Cell* (2015)

Ricci et al. *Cell* (2015)

cryo-EM

cryo-EM

3C in yeast

micro-C in yeast

STORM

No regular 30nm fiber visible

Bottom-up



Short region of DNA double helix

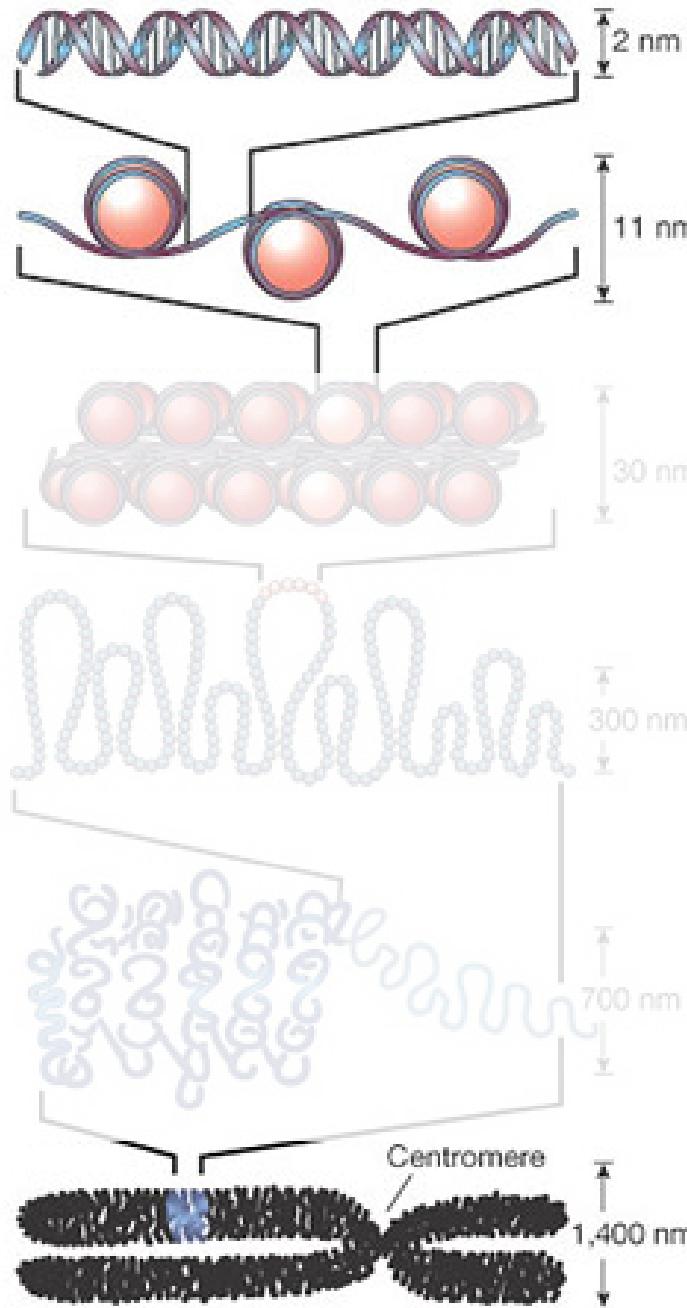
"Beads on a string" form of chromatin

30-nm chromatin fibre of packed nucleosomes

Section of chromosome in an extended form

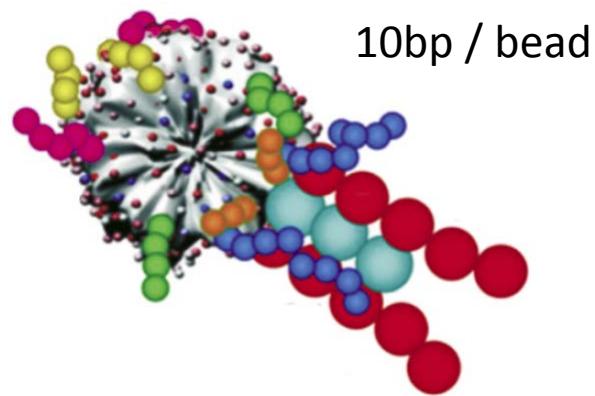
Condensed section of chromosome

Entire mitotic chromosome

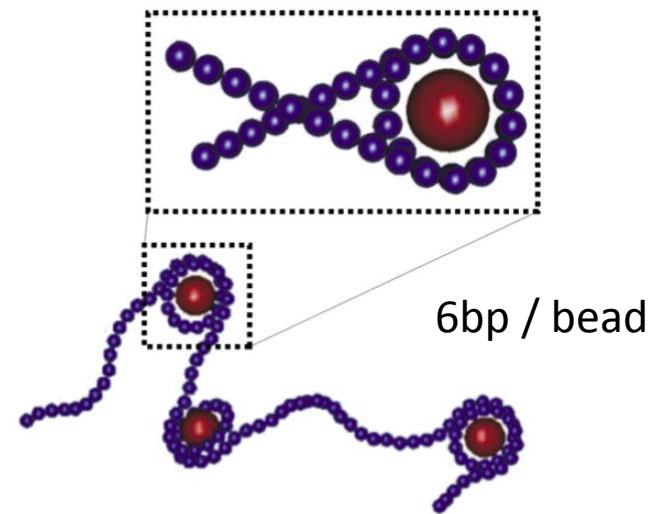


Top-down

Bottom-up models

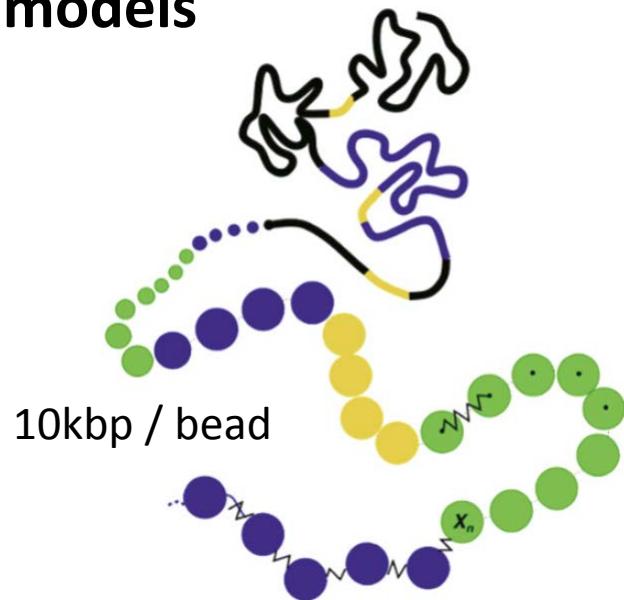


Collepardo-Guevara R, Schlick T , PNAS. 2014

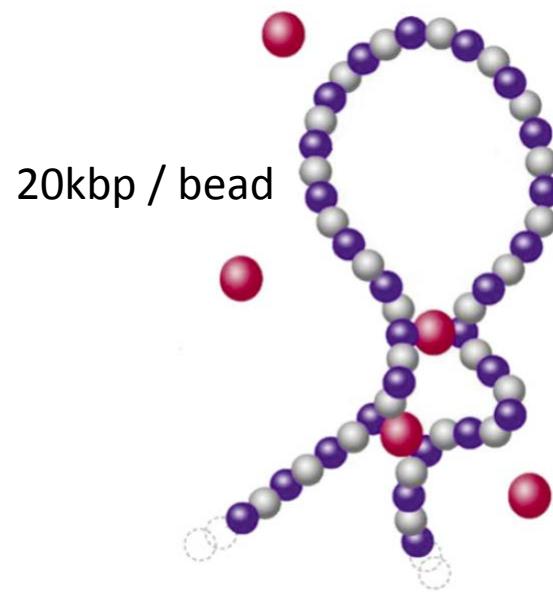


Kimura et al. J Biochem. 2013

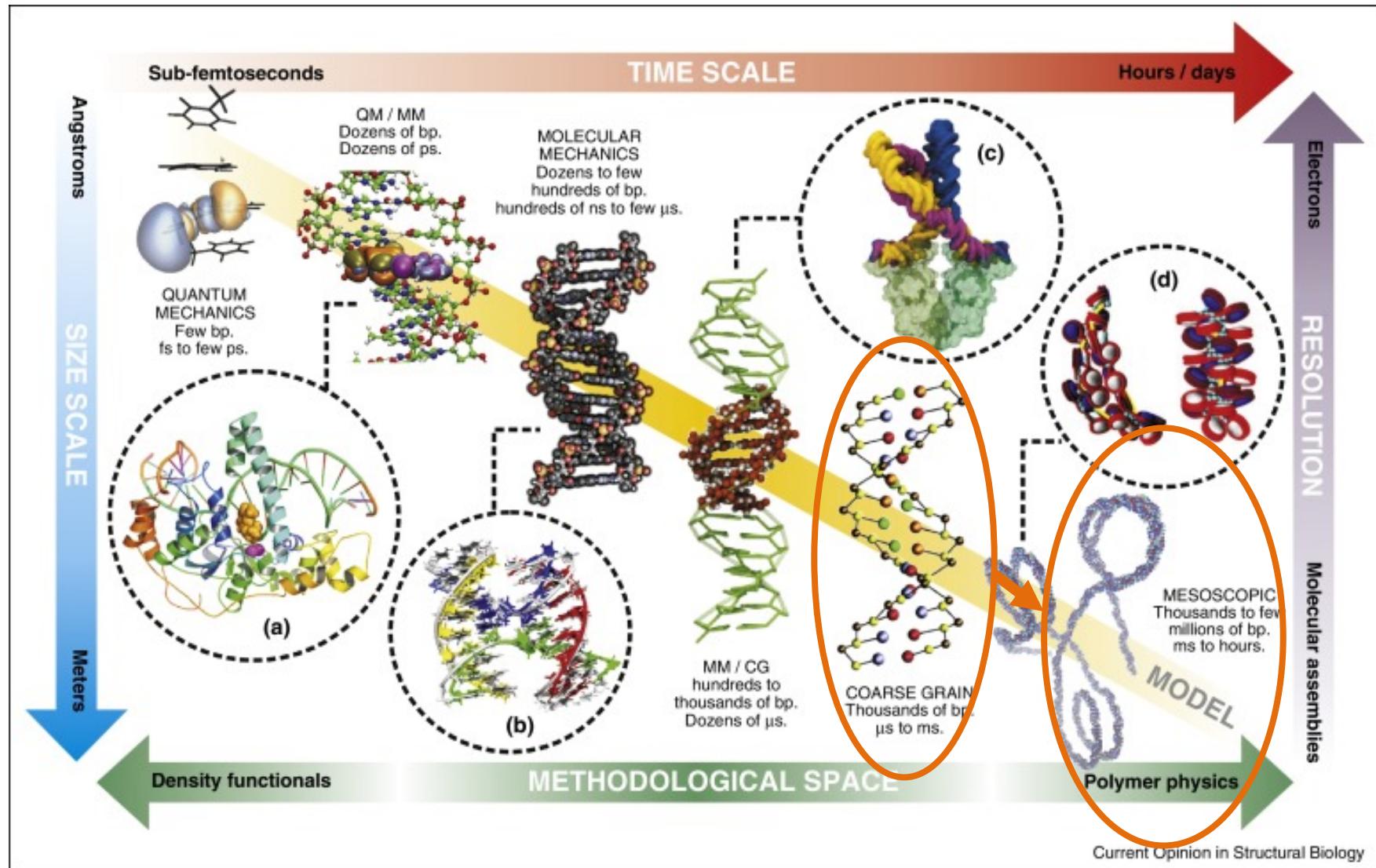
Top-down models



Jost et al., NAR. 2014



Barbieri et al., PNAS. 2012



From DNA to chromatin

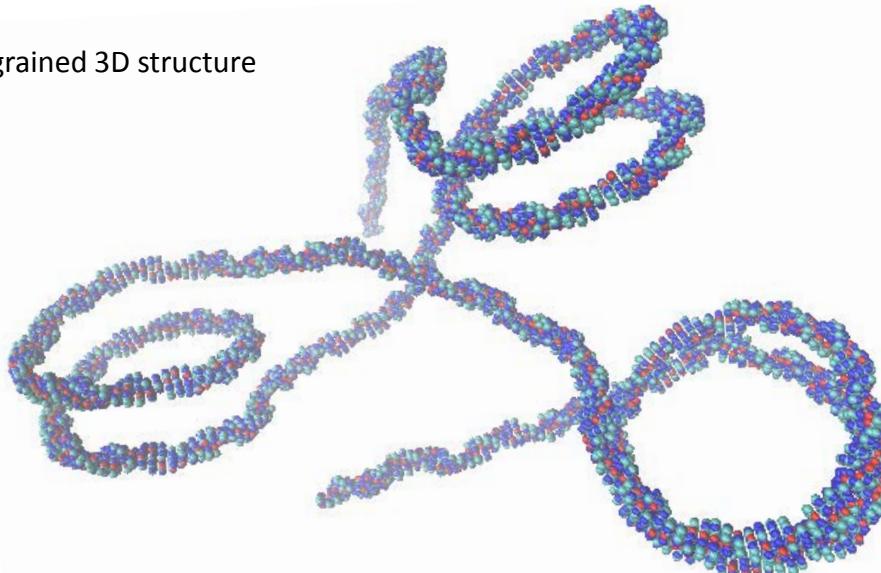
INPUT

Linker sequence
Nucleosome positions

Nucleosome 1 Nucleosome 2 Nucleosome 3
↓ ↓ ↓
ACGTACTGACGTTACTAACCTGCGC TGCACTAGCAAACGGTTGACCACCGTGAGATCTCTAACG CGGATTGGATCCGACTACGATCAACC GATAGCAGCGCGCGCGAA

OUTPUT

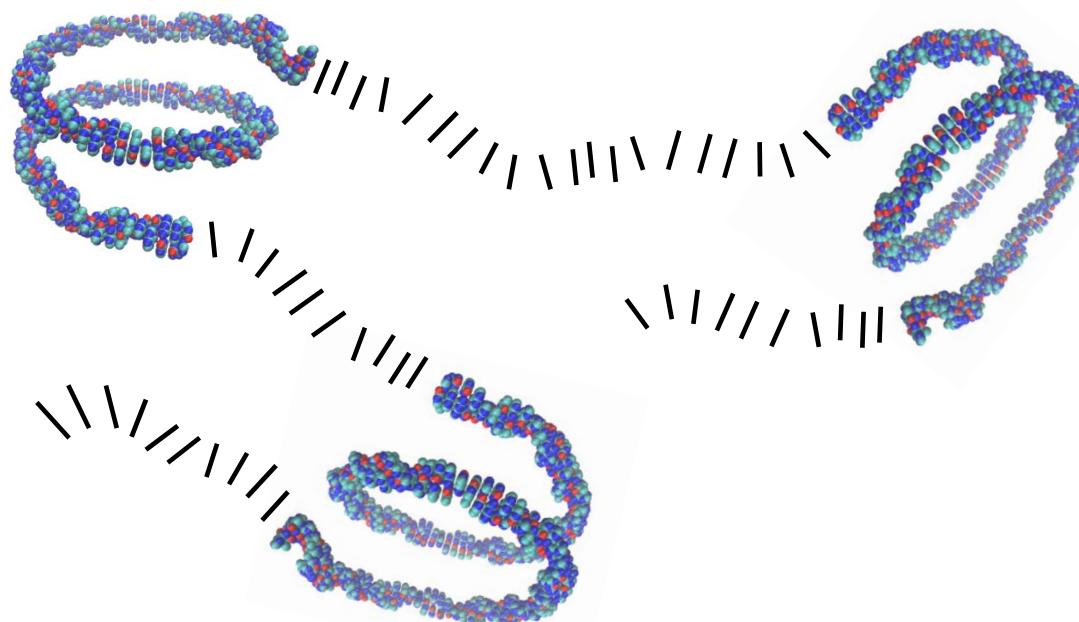
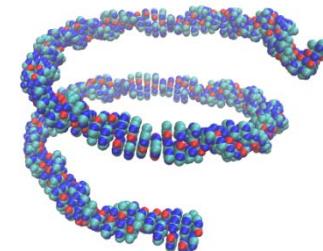
Coarse-grained 3D structure

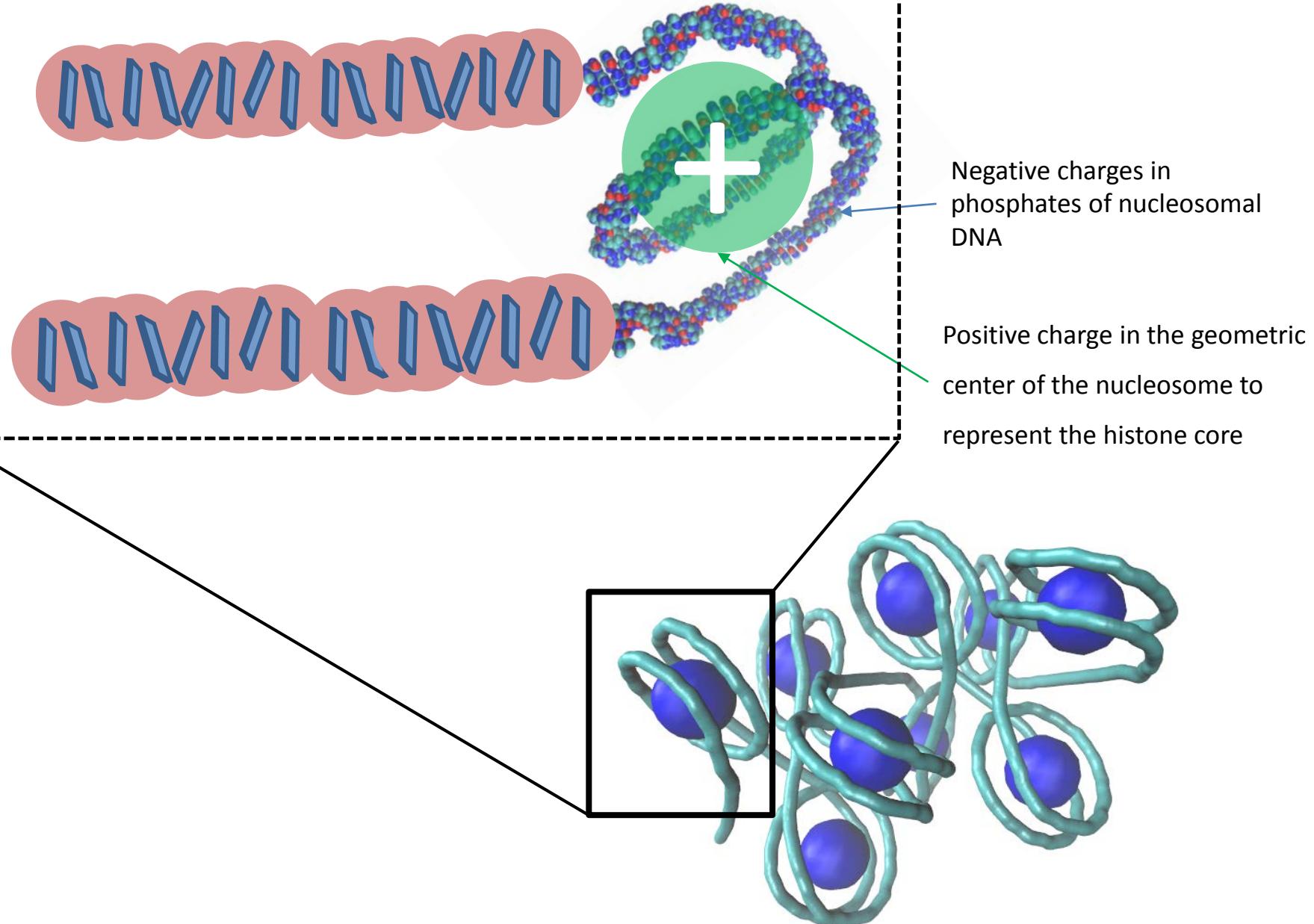


Linker DNA at bp-level with helical coordinates

Nucleosome rigid; spatial pathway of 1kx5

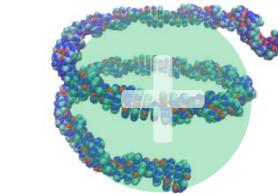
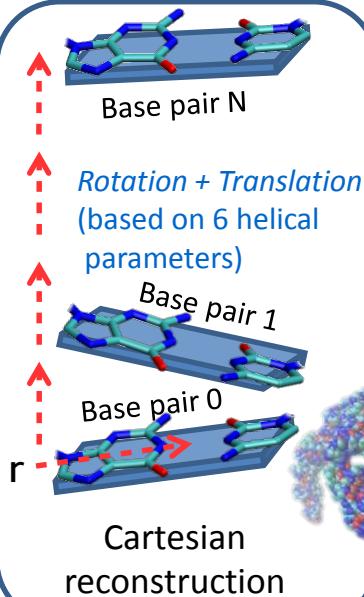
Exclude nucleosomal DNA from sampling



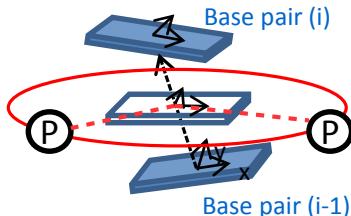




Random move on helical parameters of a bp step



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.

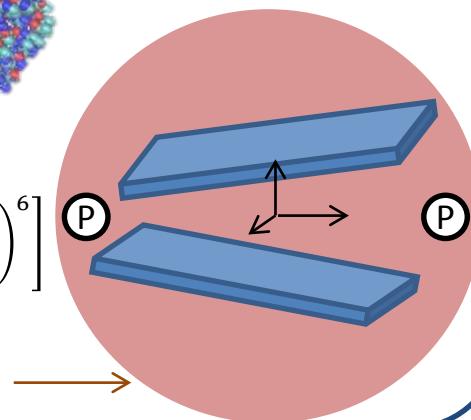
If accepted, repeat all steps with new linker configuration. If rejected repeat all steps with previous linker configuration

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

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

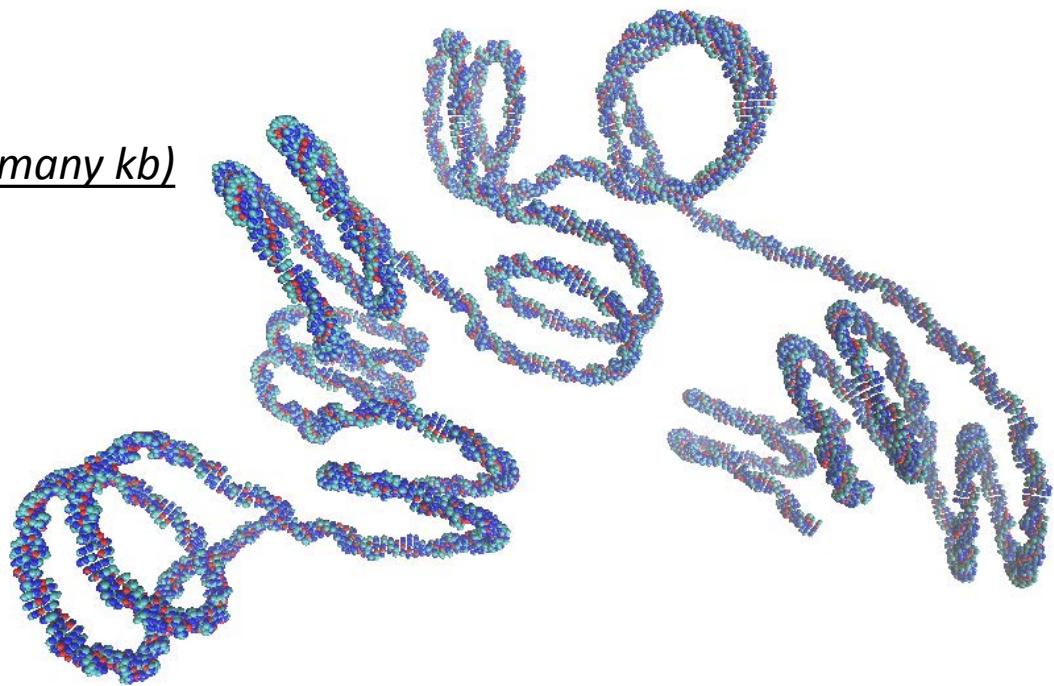
$$E^{LJ} = \sum_{i=1}^M \sum_{j=i}^M 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right]$$

Excluded Volume



Generate chromatin structure (0 – many kb)

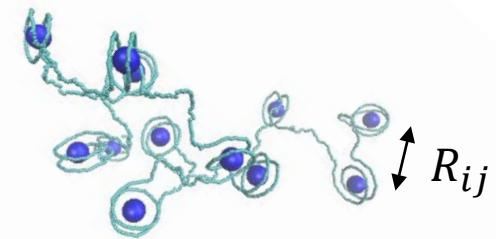
- Variable linker length
- Variable linker sequence



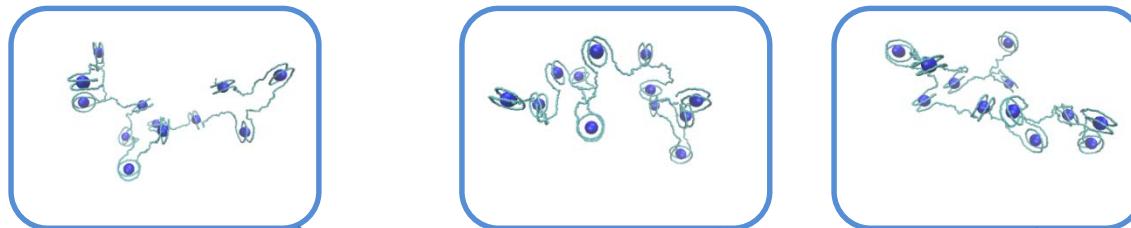
Validation: Salt-dependence of chromatin compaction

Sedimentation coefficient of nucleosomes

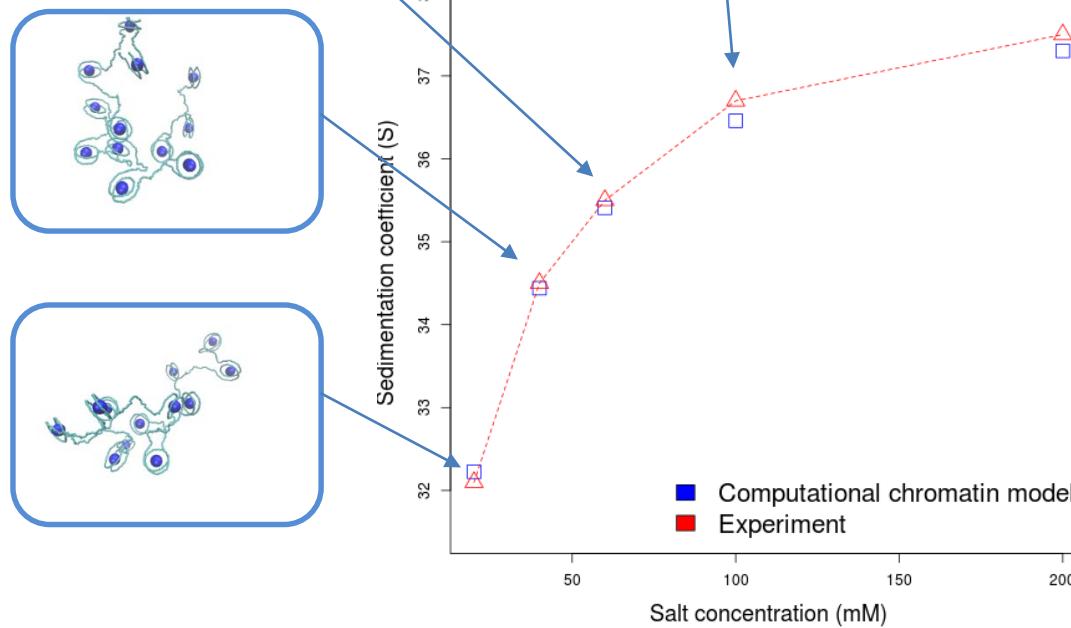
$$S = S_1 \left(1 + \frac{2R}{N} \sum_i^N \sum_{j>i}^N \frac{1}{R_{ij}} \right)$$



Experimental structure: 12 nucleosomes with 62 bp of linker DNA



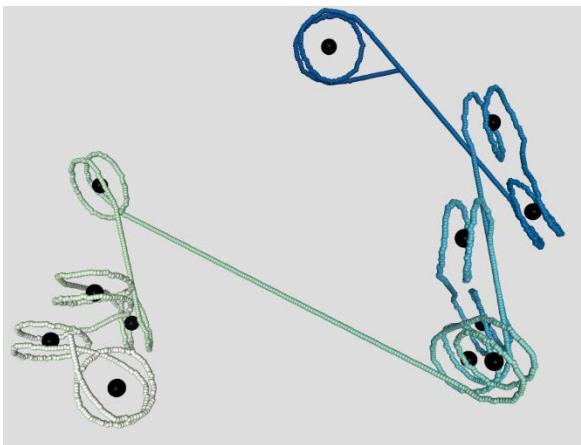
Salt dependence of sedimentation coefficient



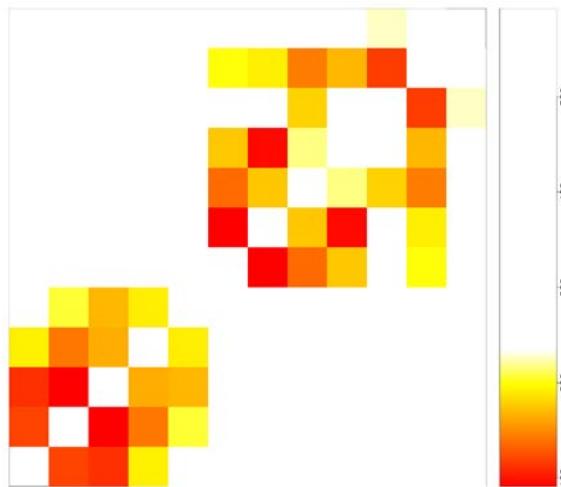
Analysis of chromatin fiber



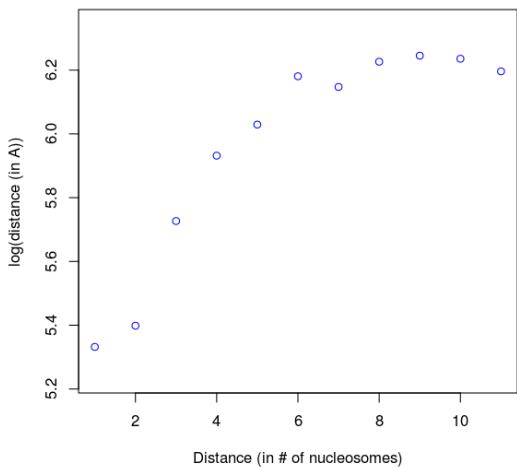
3D structure



Distance matrix



Internucleosomal distance

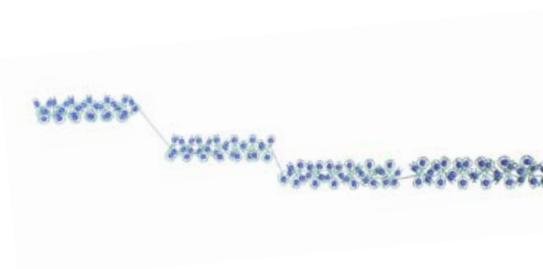


Possible applications

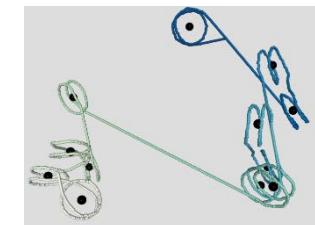
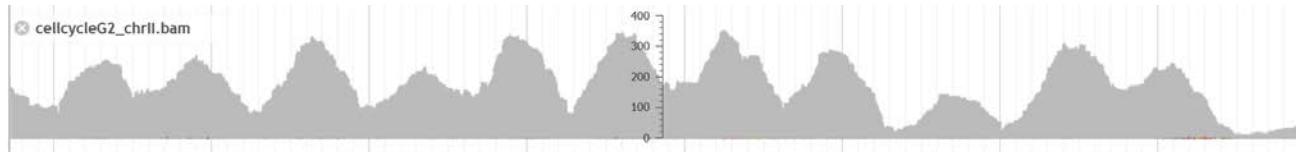


Nucleosome positions and linker DNA are based on ...

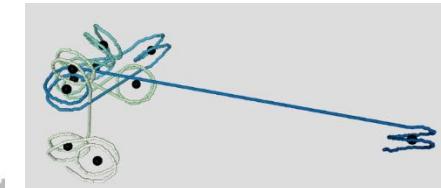
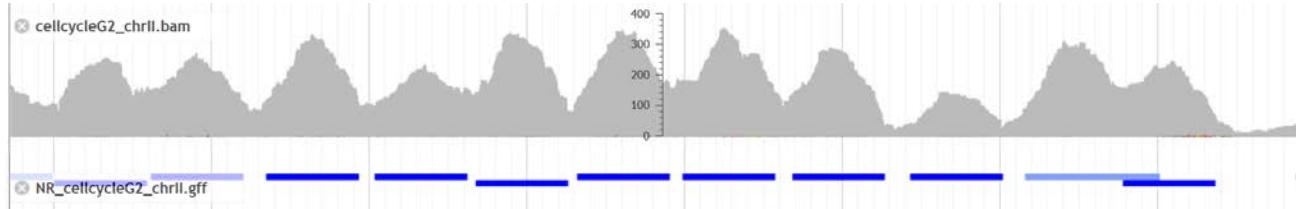
... average properties of an organism



... individual user input of a specific genomic region



... outputs of NucleR of a specific genomic region



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

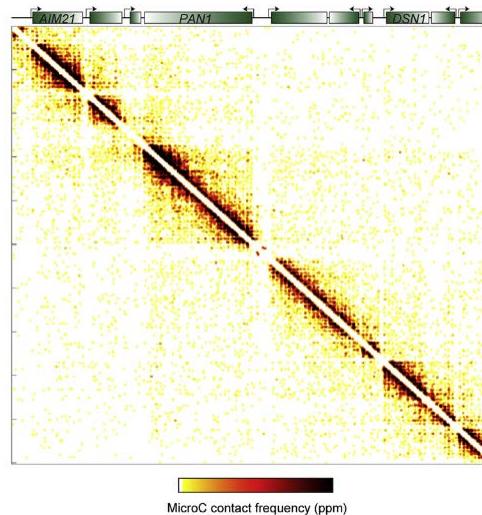
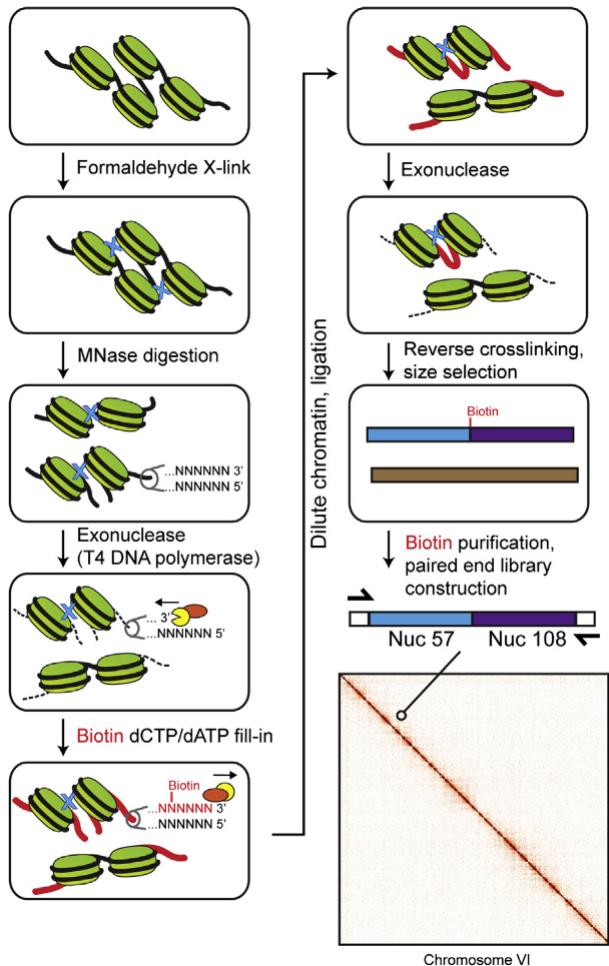
²School of Computer Science and Engineering, The Hebrew University, Jerusalem 91904, Israel

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

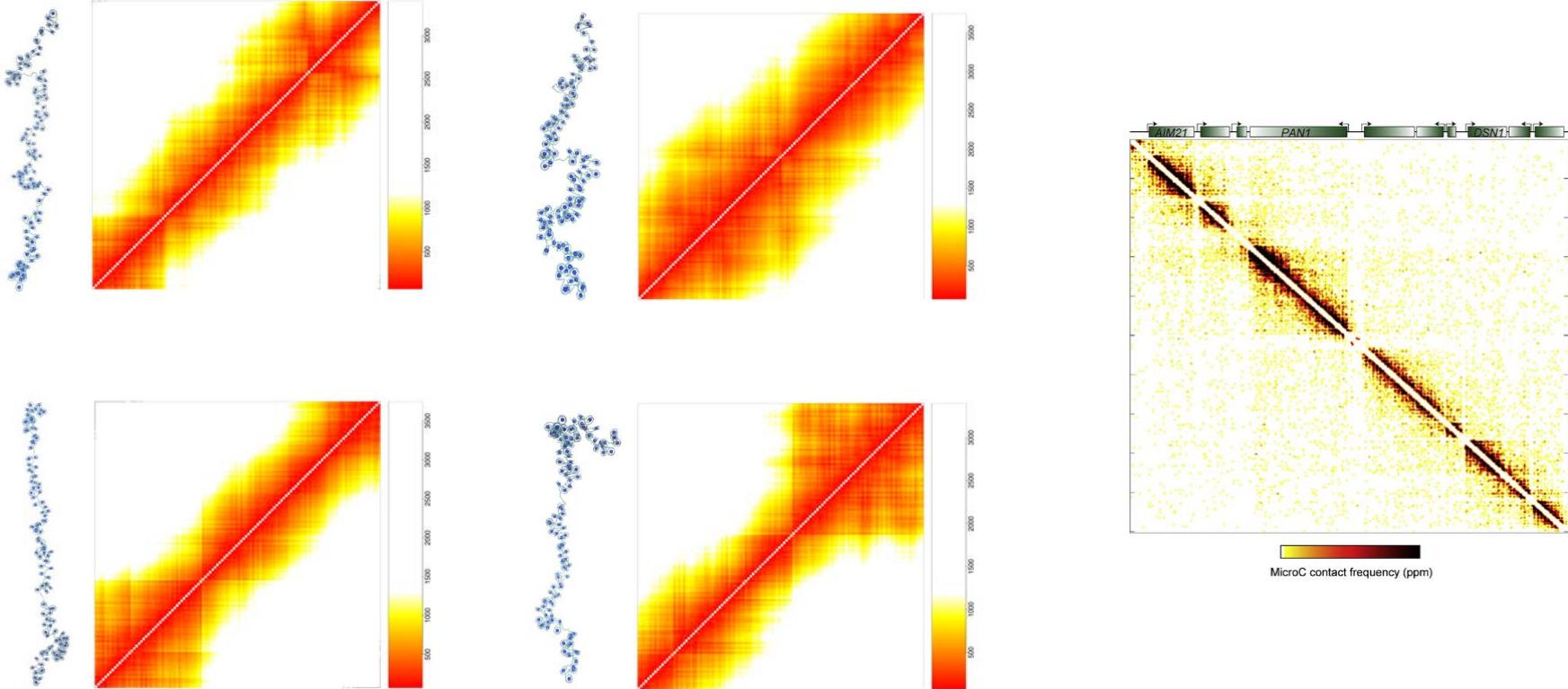
<http://dx.doi.org/10.1016/j.cell.2015.05.048>



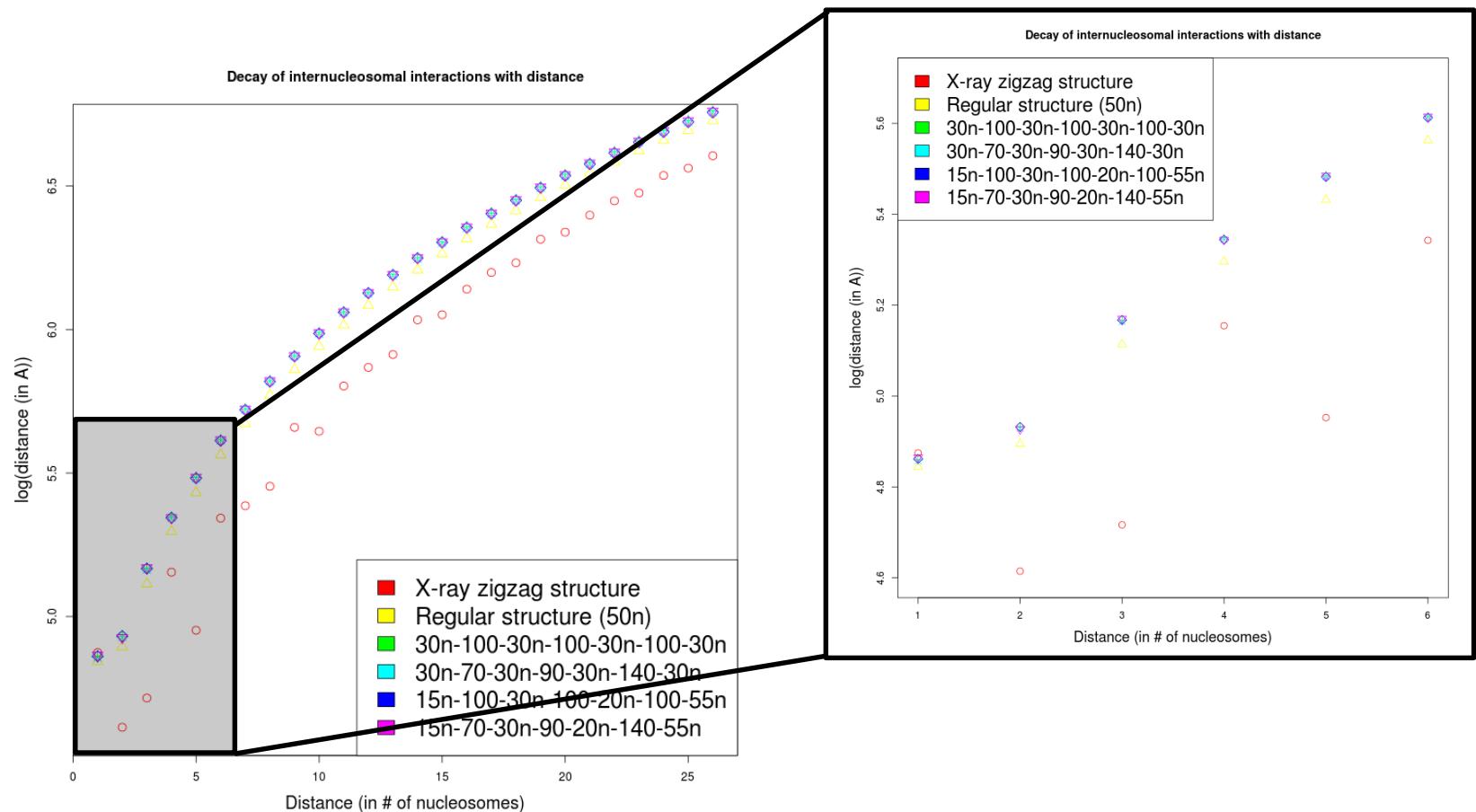
- Self-associating domains 1-5 genes (ca. 2-10kb)
- Boundaries of self-associating domains enriched in nucleosome-depleted regions
- Support for a common local motif of zig-zag arrangement of nucleosomes

Creating yeast-like nucleosome arrangements

- Average linker length between nucleosomes: 20bp
- Average length of nucleosome free region (NFR): 100bp
- Average size self-associating domains (SAD): $\approx 5\text{kb} (= 30n)$



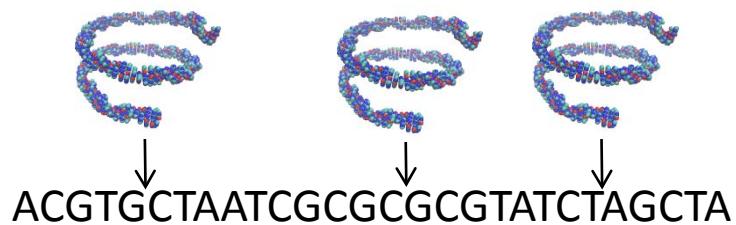
Creating yeast-like nucleosome arrangements





Workflow (1)

- Input:
- DNA linker sequence in txt file (f.ex: ACGTGCTAATCGCGCGCGTATCTAGCTA)
 - Positions of nucleosomes along linker sequence in txt file (f.ex. 5 15 23)

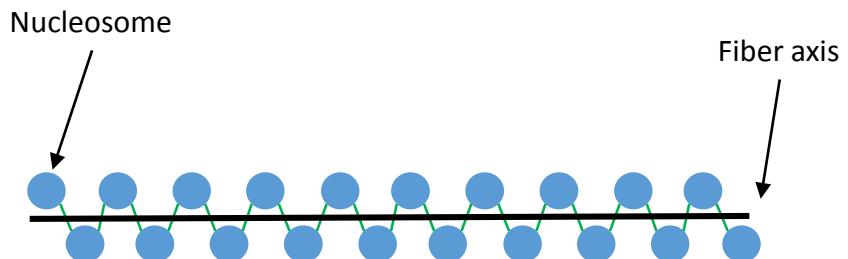


Create Structure: Creates a single structure of chromatin with straight linker DNA

Create Trajectory: Creates a certain number of simulated chromatin structures
(only if the 3D structure with straight linker DNA is not overlapping)



Workflow (1)



Contour length: length of the fiber axis

End-to-end distance: distance between first and last nucleosome of the fiber

Packing ratio: number of nucleosomes per 11nm of fiber length (relative to the fiber axis)

Fiber diameter: thickness of the fiber (average distance of a nucleosome to the fiber axis)

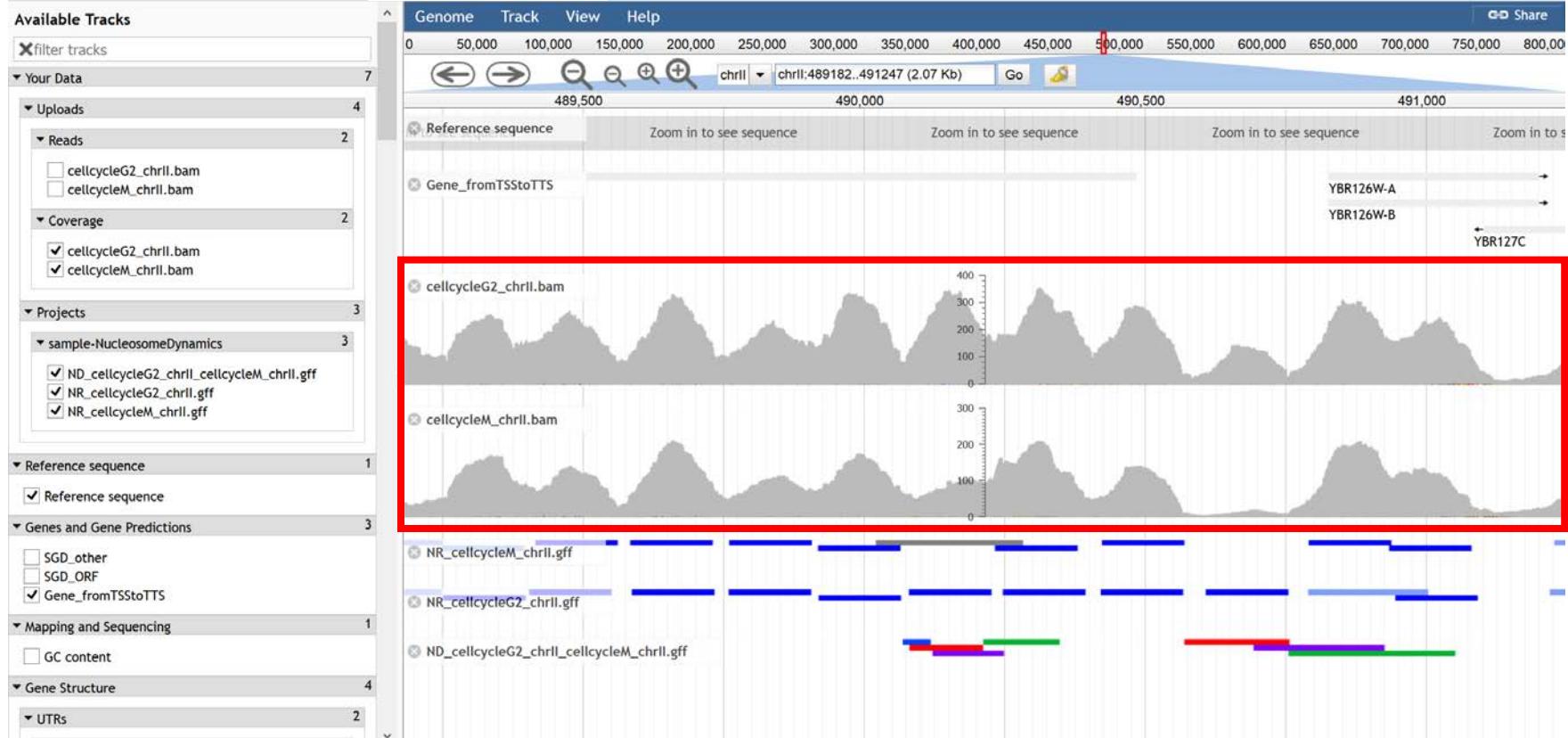
Radius of gyration: Volume occupancy of the fiber calculated with the positions of the geometric center of the nucleosomes



Workflow (1)

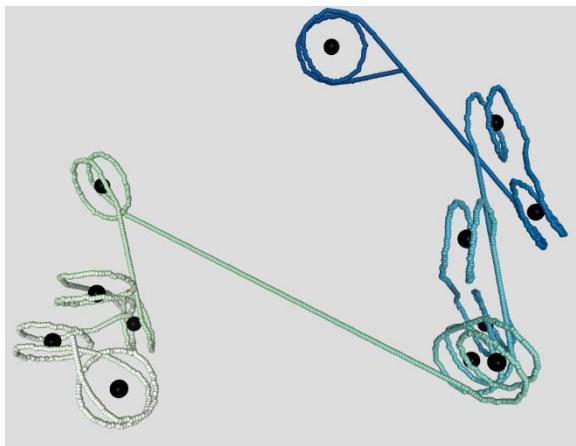
chromDyn_40m_12N
chromDyn_50m_12N

chrII:489181..491246

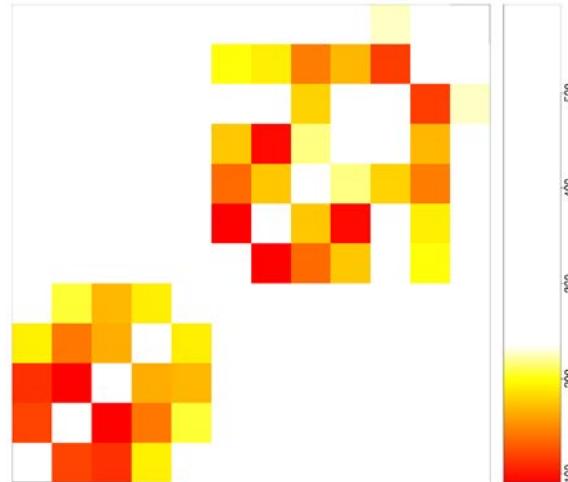


Determine nucleosome positions according to NucleR results manually (non-overlapping in 3D space)
 -> G2 like
 -> M like

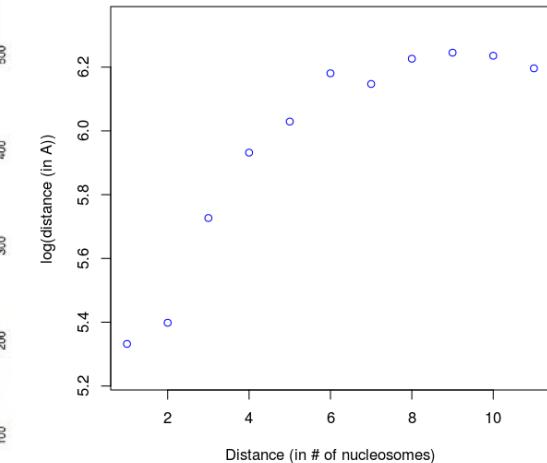
3D structure



Distance matrix



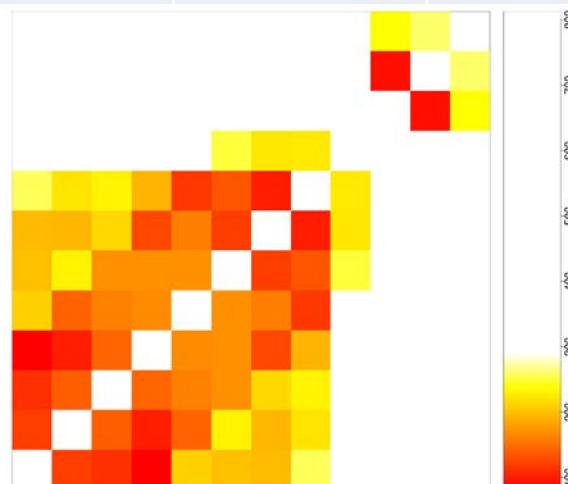
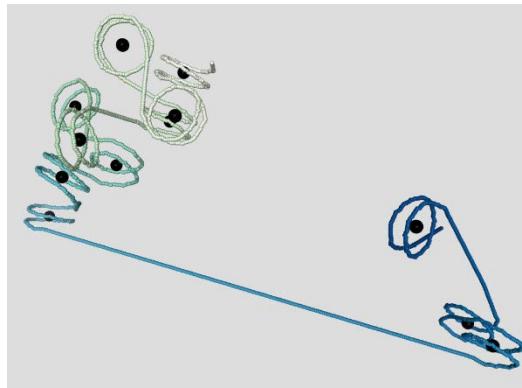
Internucleosomal distance



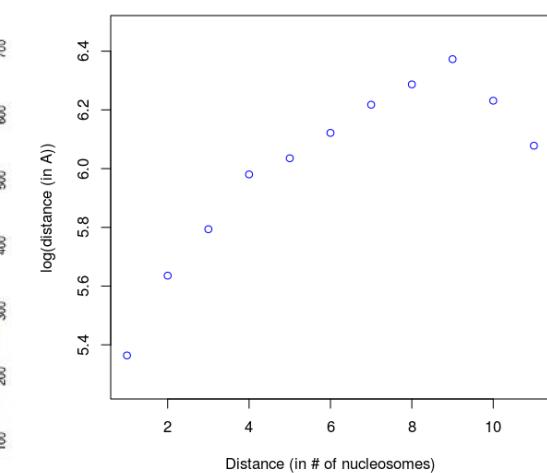
G2
like

	Contour length (in nm)	End-to-end distance (in nm)	Packing ratio	Fiber diameter (in nm)	Radius of gyration (in nm)
G2	225.866	49.086	0.58	1.472	26.315
M	234.415	43.614	0.56	1.285	29.932

M
like



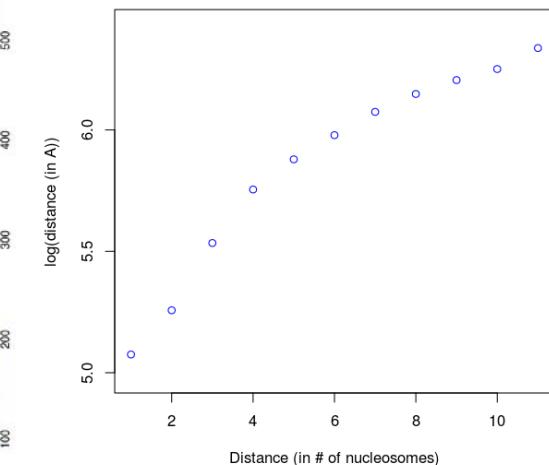
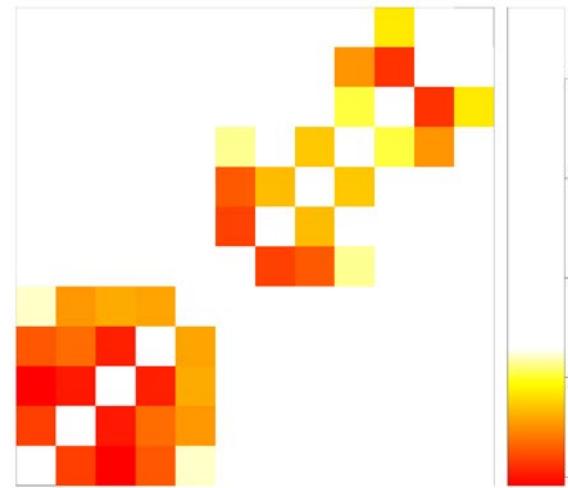
Decay of internucleosomal interactions with distance



Ensemble averaged results:

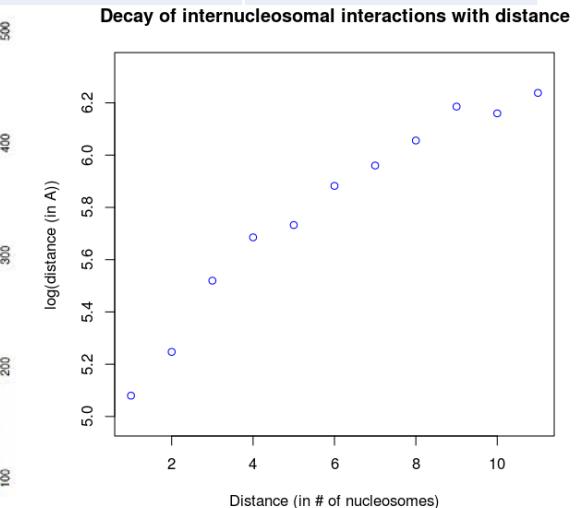
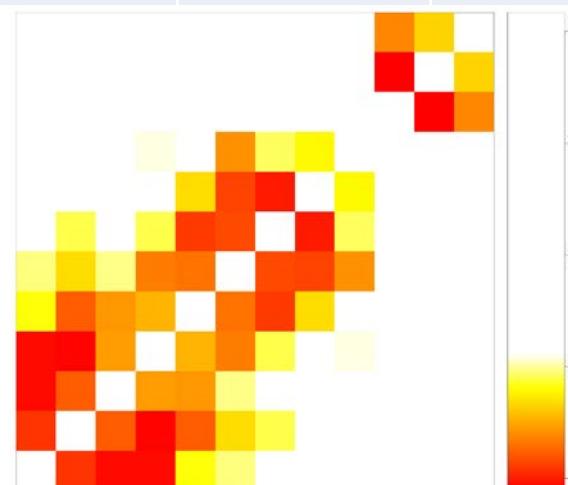
Distance matrix

Internucleosomal distance



G2 like

	Contour length (in nm)	End-to-end distance (in nm)	Packing ratio	Fiber diameter (in nm)	Radius of gyration (in nm)
G2	162.661	51.183	0.82	1.383	22.467
M	162.831	56.52	0.82	1.47	23.31

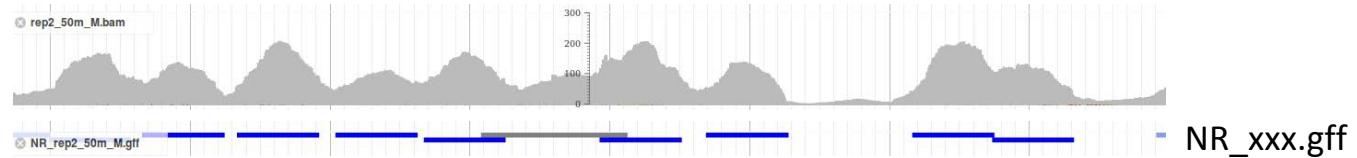


M like



Workflow (2)

Input: - Nucleosome positions computed by NucleR (in the workspace: NR_xxx.gff)



- Genomic region (chrII:489181..491246)

Create Structure: Creates a single structure of chromatin with straight linker DNA



Virtual Research Environment

Test

Homepage

User Workspace

Get Data

External Links

Forum

User

Admin

Name: Nucleosome_Dynamics_to_Chromatin_Dynamics

Description: Write a short description here...

Chromatin Dynamics from NucleR

File inputs

NucleR output: uploads / NR_M_chrl1.gff

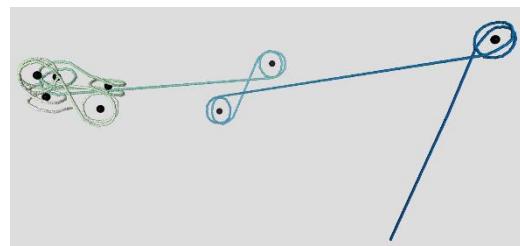
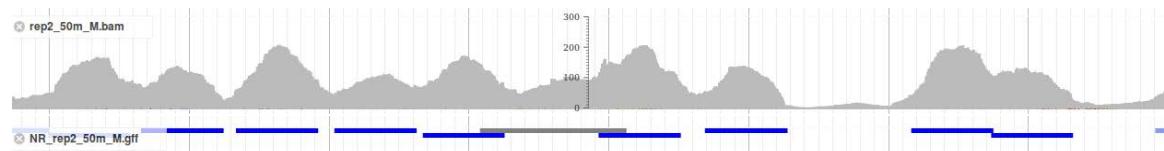
Settings

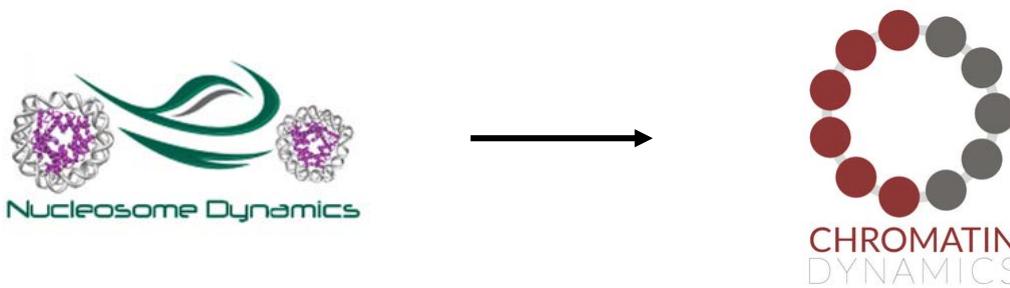
Operations: Create 3D Structure

3D structure from Nucleosome Dynamics: chrll:489181..491246

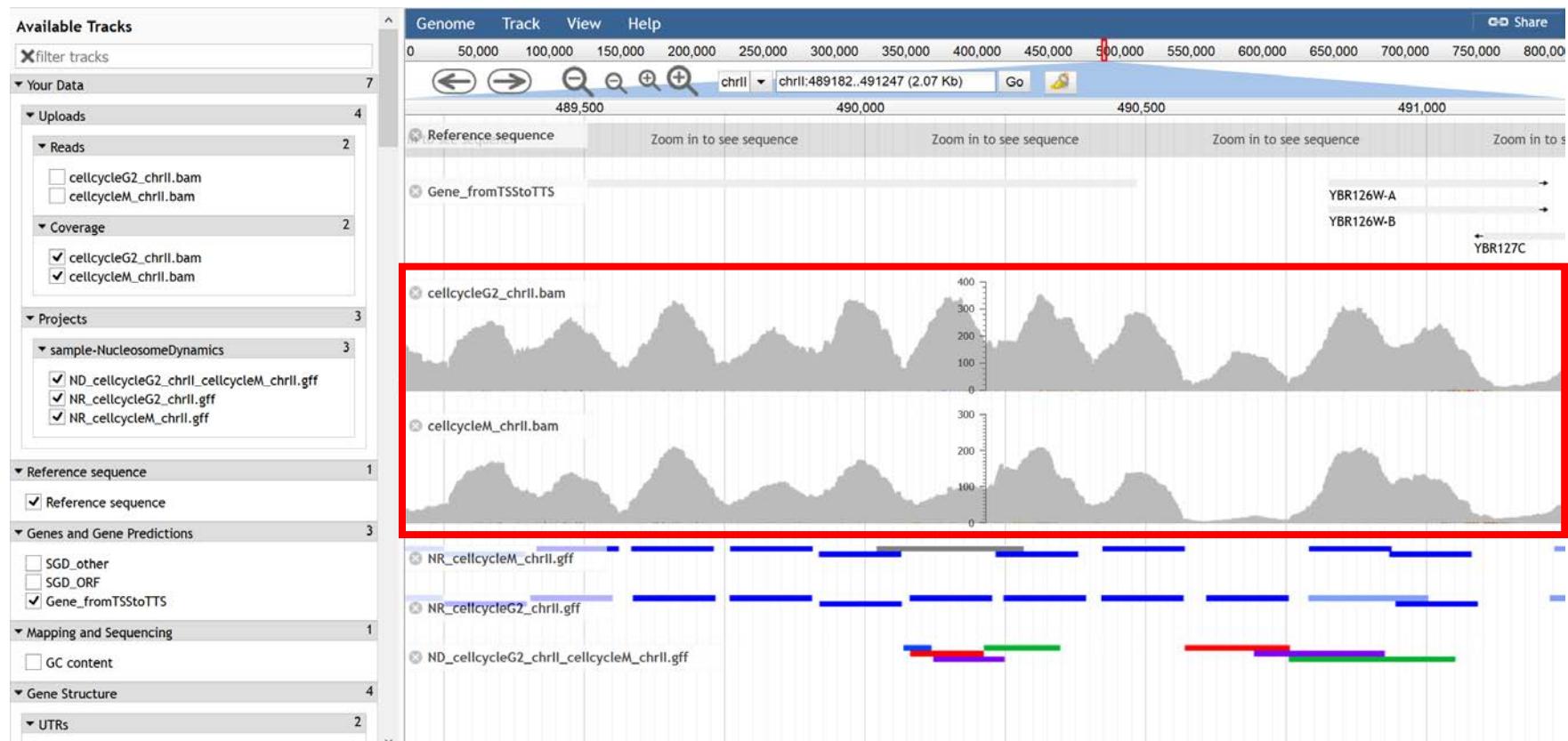
Compute

This screenshot shows a user interface for a virtual research environment. On the left, a sidebar lists various tools and resources: Homepage, User Workspace (selected), Get Data, External Links, Forum, User, and Admin. The main workspace is titled "Chromatin Dynamics from NucleR". It contains several input fields and controls. At the top, there are fields for "Name" (set to "Nucleosome_Dynamics_to_Chromatin_Dynamics") and "Description" (a placeholder text area). Below this is a section for "File inputs" with a dropdown menu showing "uploads / NR_M_chrl1.gff". Under "Settings", there is a "Create 3D Structure" button. In the bottom right corner, there is a "Compute" button with a checkmark icon. A yellow highlight box surrounds the "chrll:489181..491246" entry in the "3D structure from Nucleosome Dynamics" field.

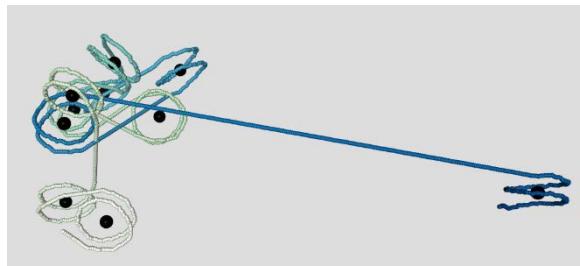




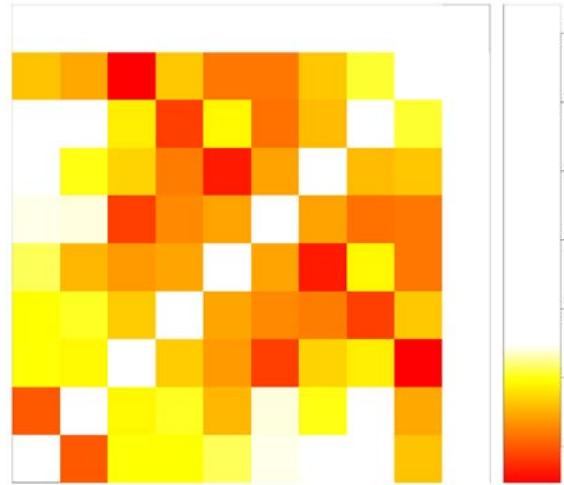
chrII:489181..491246



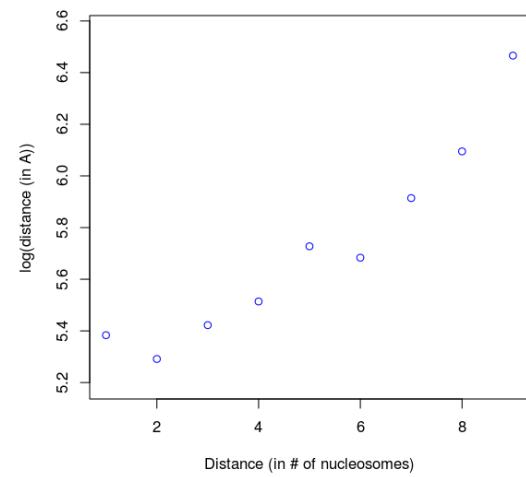
3D structure



Distance matrix

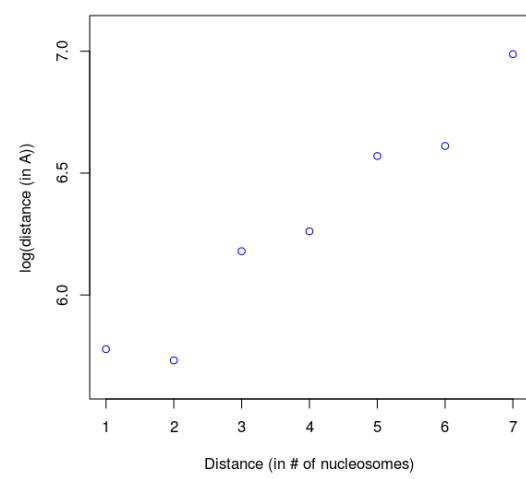
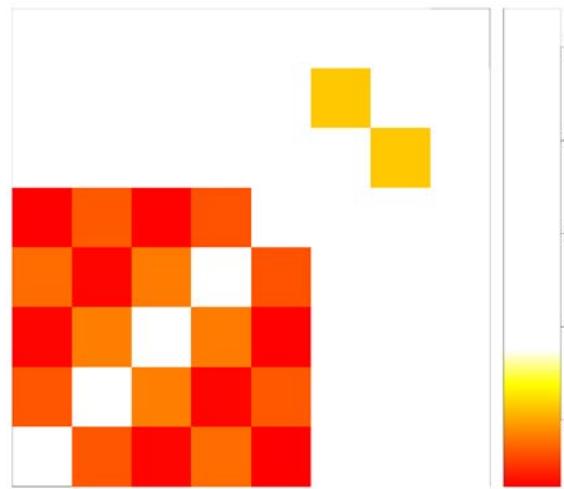
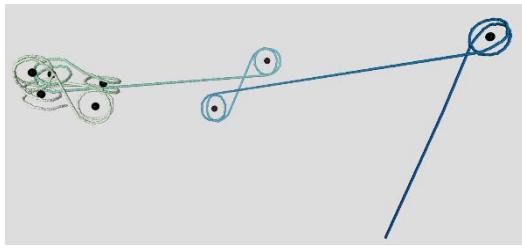


Internucleosomal distance



G2

M





Multiscale Complex Genomics

Thank you

juergen.walther@irbbarcelona.org



www.multiscalegenomics.eu



irbinfo.mug@irbbarcelona.org



@MuG_genomics



Acknowledgements



Modesto Orozco



Pablo Dans



Collaborators

Marco Pasi (University of Nottingham)

Richard Lavery (IBCP Lyon)

John Maddocks (EPFL)



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



Molecular Modelling & Bioinformatics Group



European Union, H2020, MuG: 676556

MuG Team

