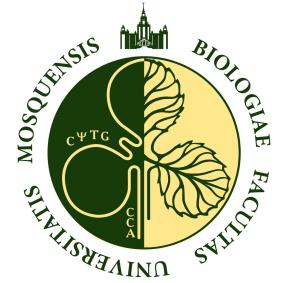




Lomonosov Moscow State University
Faculty of Biology
Dept. of Bioengineering
Integrative Biology Group



Using supercomputers to study molecular dynamics of DNA-protein complexes at multi-microsecond time scale

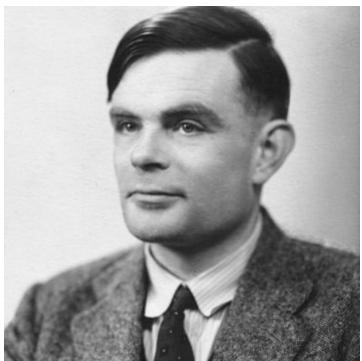
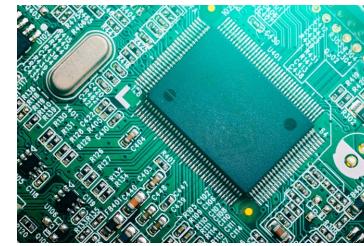
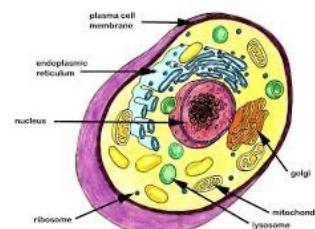
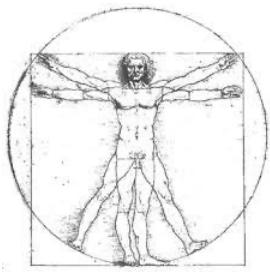
Alexey K. Shaytan, PhD, Dr. Sci.

alex@intbio.org

<http://intbio.org>

Russian supercomputing days, September 27-28, 2021

Computing and biology



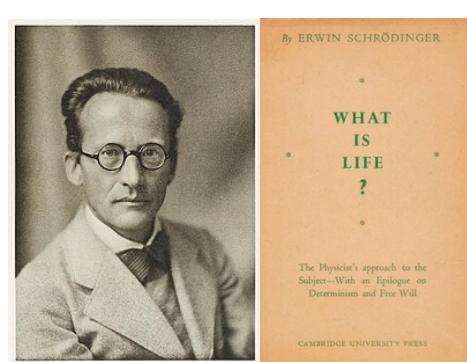
THE CHEMICAL BASIS OF MORPHOGENESIS

By A. M. TURING, F.R.S. University of Manchester

(Received 9 November 1951—Revised 15 March 1952)



Ronald Fisher

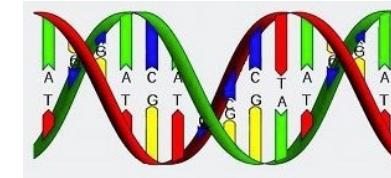
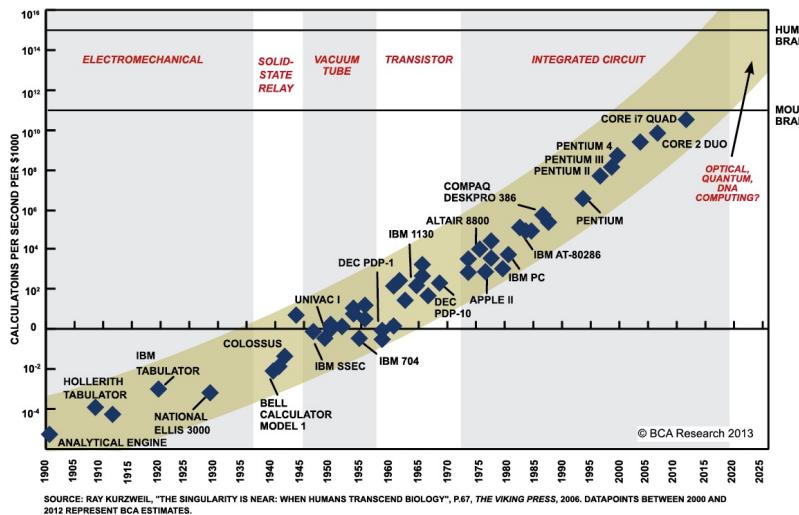


Erwin Schrodinger

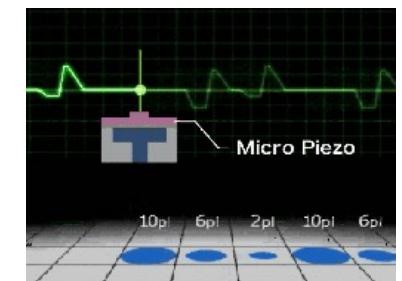
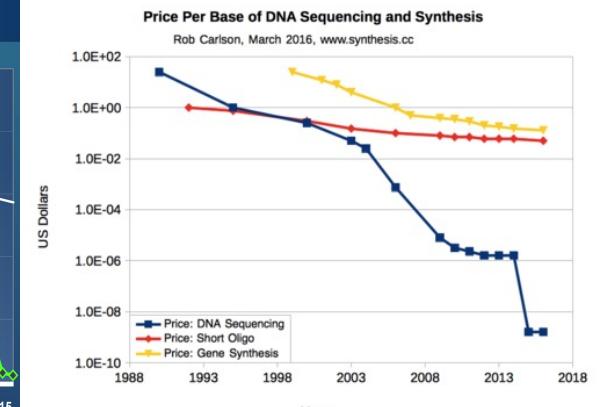
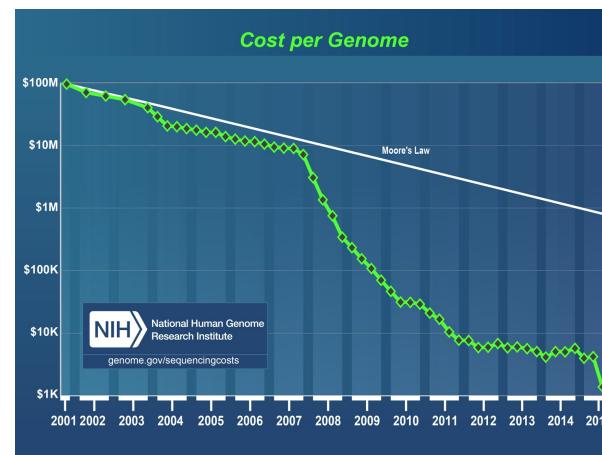


Claude Shannon

Computing and biology



DNA is a digital code which can be read and written.



Computers vs biosystems

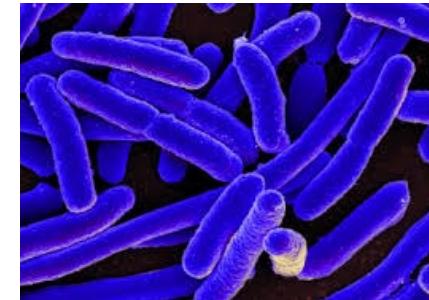


Human genome (2001)

3 billion nucleotides
262K pages



~ 1 Tb

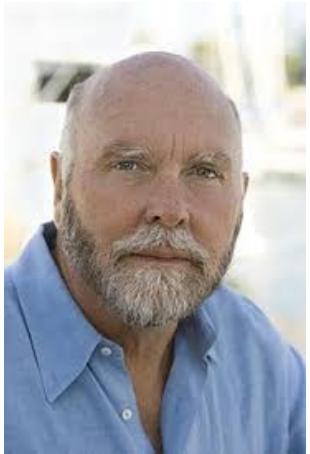


E. coli bacteria
Genome: ~ 5M bases



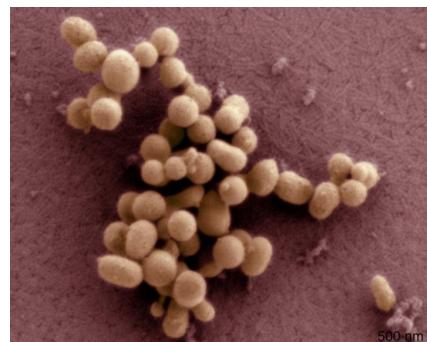
Intel Xeon Phi CPU
 8×10^9 transistors⁴

First «artificial life»



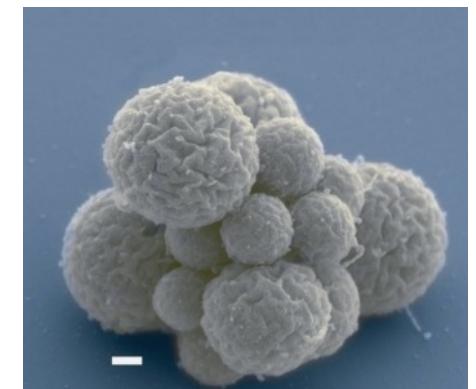
Craig Venter

2010 *Mycoplasma mycoides* JCVI-syn1.0



Human	~20,000–25,000
<i>Escherichia coli</i> (K12 strain)	~4500
Syn 1.0	901
<i>Mycoplasma genitalium</i> *	525
Syn 3.0	473

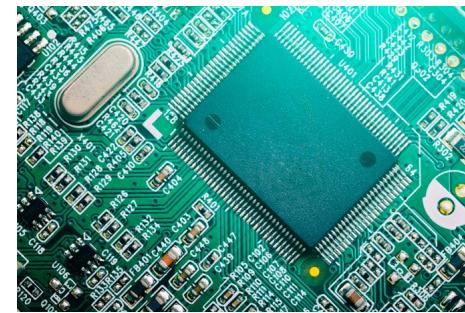
2016 Syn3.0



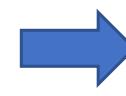
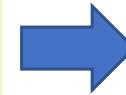
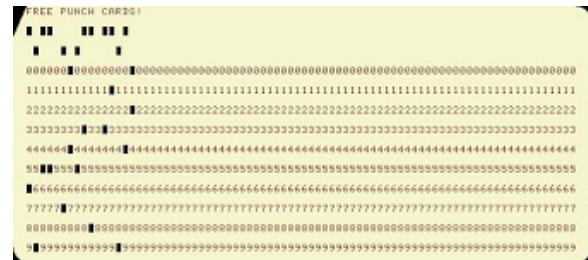
First artificial genome (~\$40 M)

Minimal genome

- Software boots into hardware
- Genetic code boots into cells



~1950s



????
Genetic CAD?
Genome compiler?

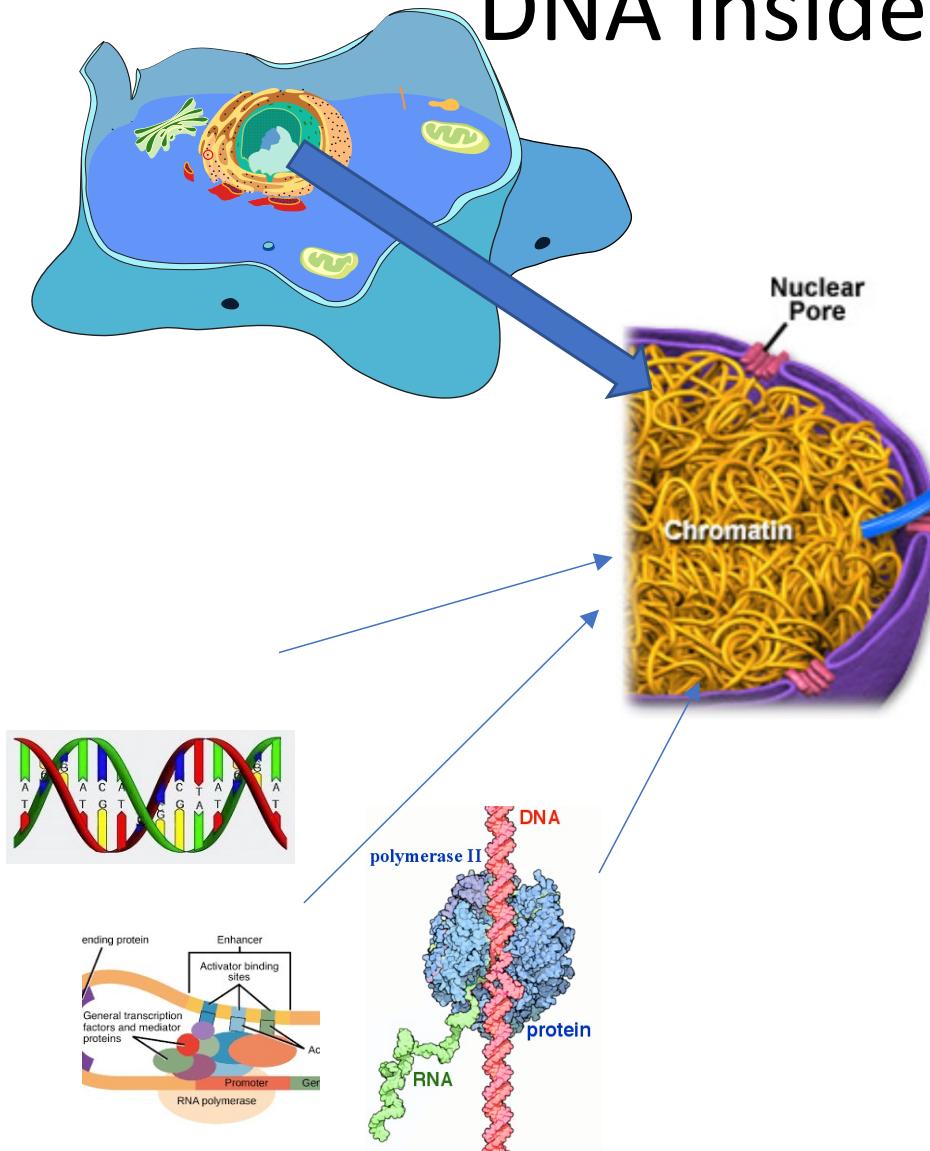


Bill Gates-backed cell programming startup Ginkgo Bioworks is going public via a \$15 billion SPAC deal



Tom Knight,
Software engineer from MIT,
“father of synthetic biology”

DNA inside cell nucleus



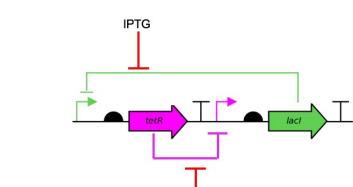
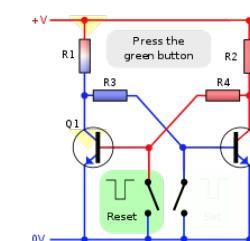
Memory (NV-RAM)



Memory (RAM)

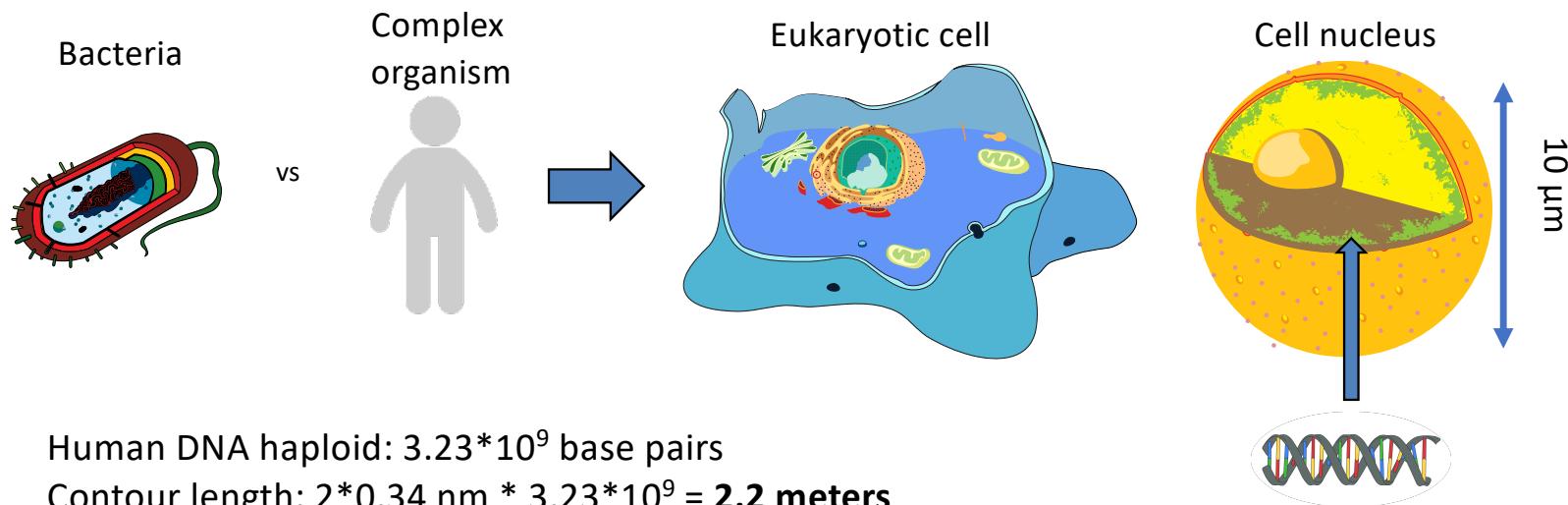


Information processing



Genetic circuits

Chromatin



Human DNA haploid: 3.23×10^9 base pairs

Contour length: $2 \times 0.34 \text{ nm} \times 3.23 \times 10^9 = 2.2 \text{ meters}$

Number of cells in human body: $\sim 30 \times 10^{12}$

Total DNA length: $2.2 \times 30 \times 10^{12} = 66 \text{ billion km}$

- The human DNA could be densely packed into a cube of size:

$$(2 \text{ nm} \times 2 \text{ nm} \times 2 \times 0.34 \text{ nm} \times 3.23 \times 10^9)^{1/3} = 2 \mu\text{m}$$

- The human DNA would form a random coil with gyration radius:

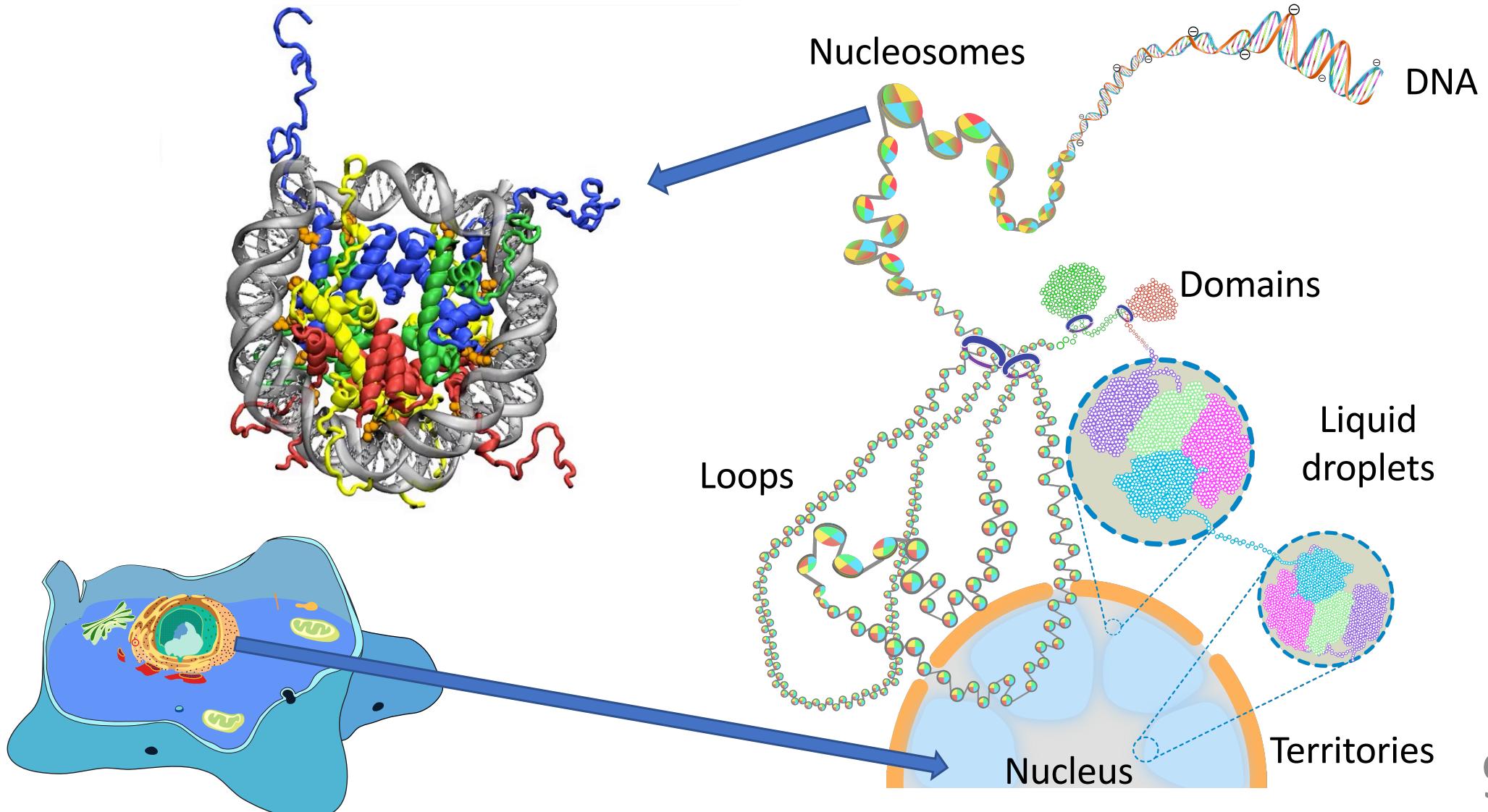
A schematic diagram of a random coil polymer segment, showing a series of vectors \vec{r}_i originating from a common point, representing the bond lengths and angles of the coil. A dashed line indicates the overall radius of gyration \bar{R} .

$$R_G = \frac{\sqrt{N} l}{\sqrt{6}} = 135 \mu\text{m}$$



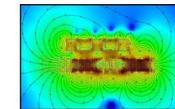
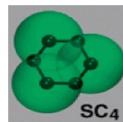
Sun to Earth: 150 million km

The complexity of chromatin structure and dynamics



Supercomputer molecular dynamics (MD) simulations

$$\hat{H}\Psi = E\Psi \quad \vec{F} = m\vec{a}$$



Quantum
chemistry

Molecular
mechanics

Coarse grain
models Continuum
models

Accurate

Approximate



2013 Chemistry Prize



Taking the Experiment to Cyberspace

The Nobel Prize in Chemistry 2013 was awarded jointly to [Martin Karplus](#), [Michael Levitt](#) and [Arieh Warshel](#) "for the development of multiscale models for complex chemical systems".

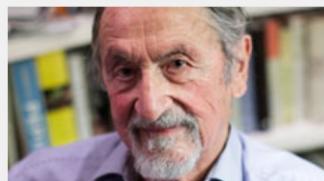


Photo © Harvard University

Martin Karplus

Martin Karplus, U.S. and Austrian citizen. Born 1930 in Vienna, Austria. Ph.D. 1953 from California Institute of Technology, CA, USA. Professeur Conventionné, Université de Strasbourg, France and Theodore William Richards Professor of Chemistry, Emeritus, Harvard University, Cambridge, MA, USA



Photo: S. Fisch

Michael Levitt

Michael Levitt, U.S., British and Israeli citizen. Born 1947 in Pretoria, South Africa. Ph.D. 1971 from University of Cambridge, UK. Robert W. and Vivian K. Cahill Professor in Cancer Research, Stanford University School of Medicine, Stanford, CA, USA.



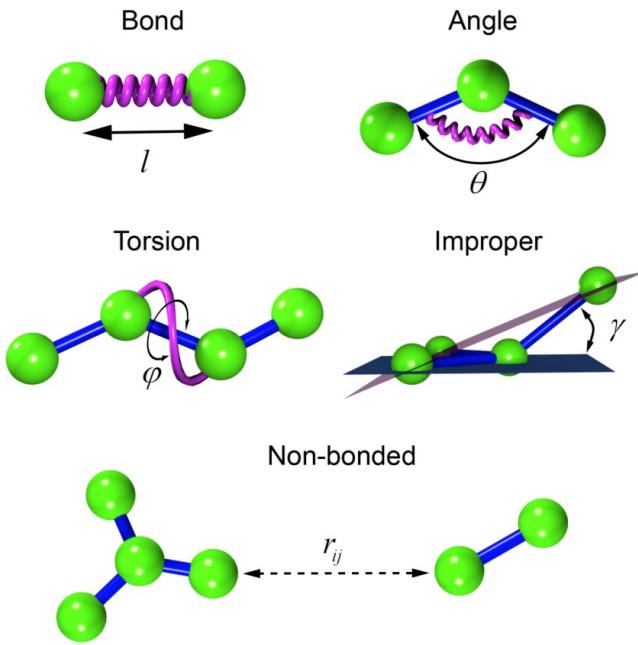
Photo: Wikimedia Commons

Arieh Warshel

Arieh Warshel, U.S. and Israeli citizen. Born 1940 in Kibbutz Sde-Nahum, Israel. Ph.D. 1969 from Weizmann Institute of Science, Rehovot, Israel. Distinguished Professor, University of Southern California, Los Angeles, CA, USA.

[► Interviews with Chemistry](#)

Typical biomolecular force field

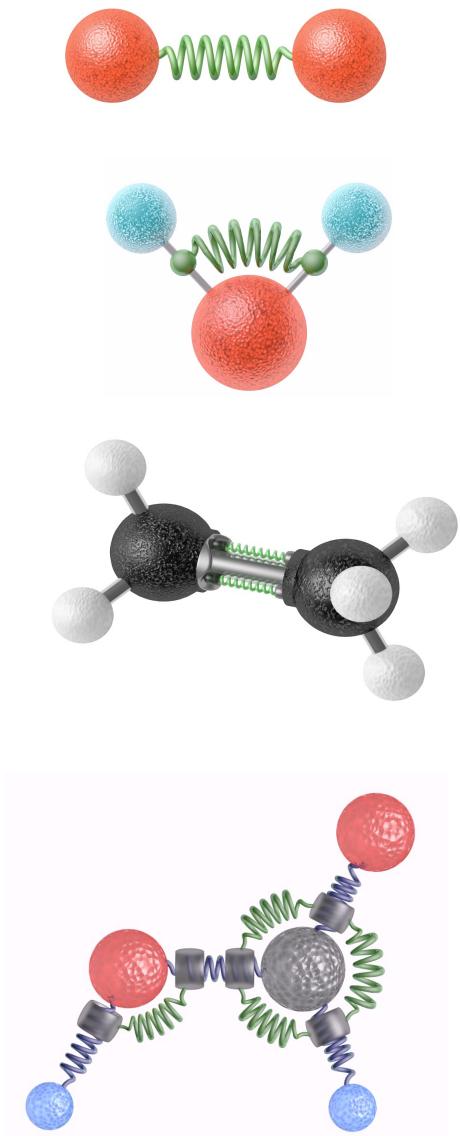


$$\vec{F} = m * \vec{a}$$

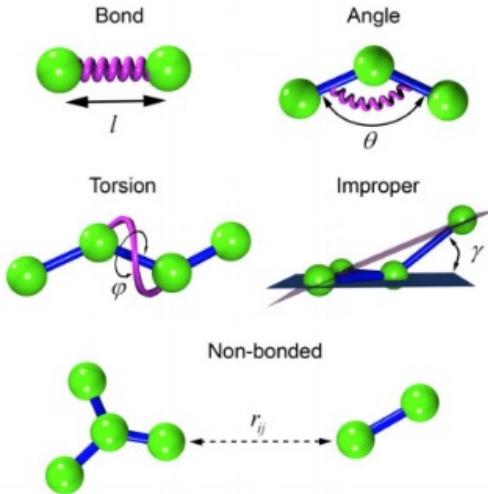
$$\frac{d^2\vec{r}_i}{dt^2} = \vec{a} = \frac{\vec{F}}{m} = -\frac{1}{m_i} \nabla_i U(\{\vec{r}_i\})$$

$$\begin{aligned}\vec{r}(t + \Delta t) &= \vec{r}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 \\ \vec{v}(t + \Delta t) &= \vec{v}(t) + \frac{\vec{a}(t) + \vec{a}(t + \Delta t)}{2}\Delta t\end{aligned}$$

$$\begin{aligned}U(\{\vec{r}_i\}) &= \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\varphi - \varphi_0)] \\ &+ \sum_{impropers} \frac{1}{2} k_\gamma (\gamma - \gamma_0)^2 + \sum_{j=1}^{N-1} \sum_{i=j+1}^N \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} f_{ij}\end{aligned}$$



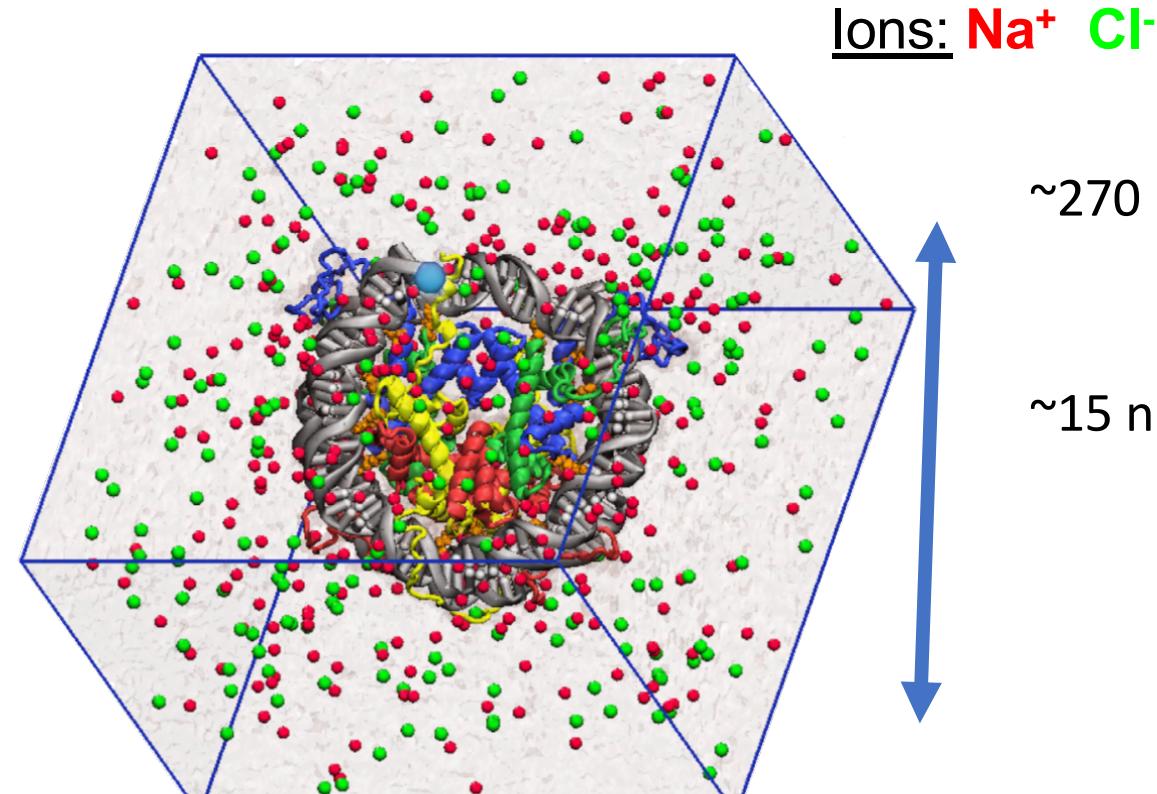
All-atom MD simulations of nucleosomes



- Force field:
AMBER14SB + parmbsc1 + CUFIX + TIP3P
- 150 mM NaCl
- GROMACS2019 (GPU)



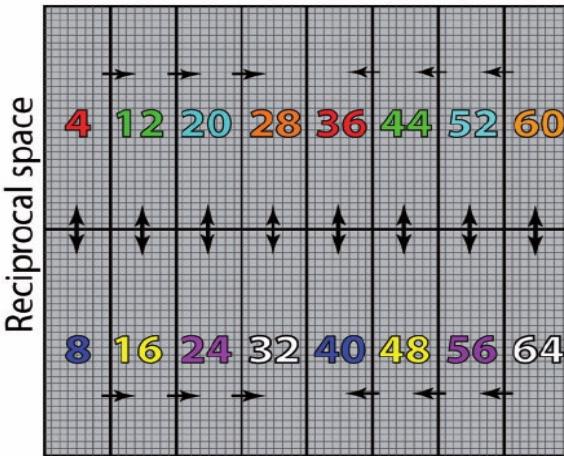
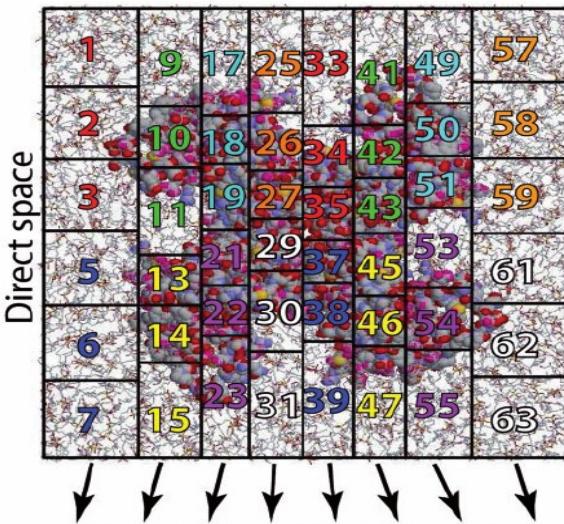
Lomonosov-2
supercomputer



- Simulations up to 15 microseconds long.

Armeev, G. A.; Kniazeva, A. S.; Komarova, G. A.; Kirpichnikov, M. P.; Shaytan, A. K. *Nature Commun.* **2021**, 12.

Parallel simulations, domain decomposition (GROMACS)

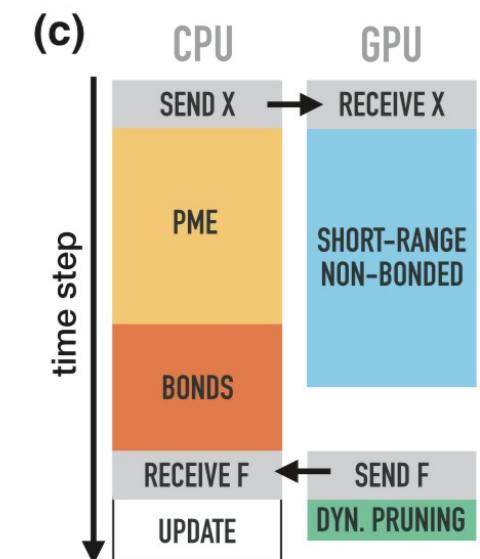
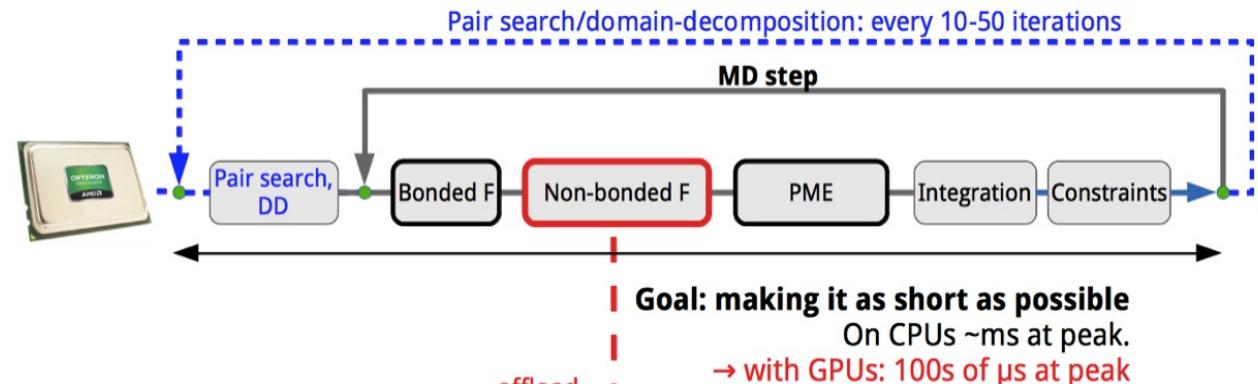


Domain decomposition

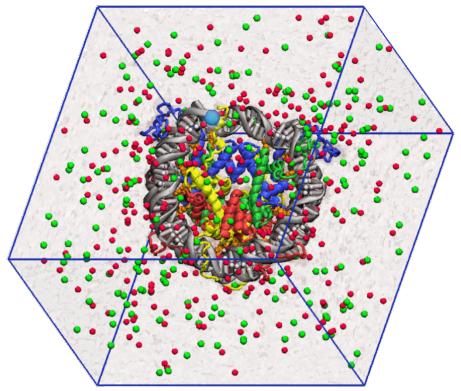
http://www.gromacs.org/GPU_acceleration

Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; Groot, B. L. de; Grubmüller, H.

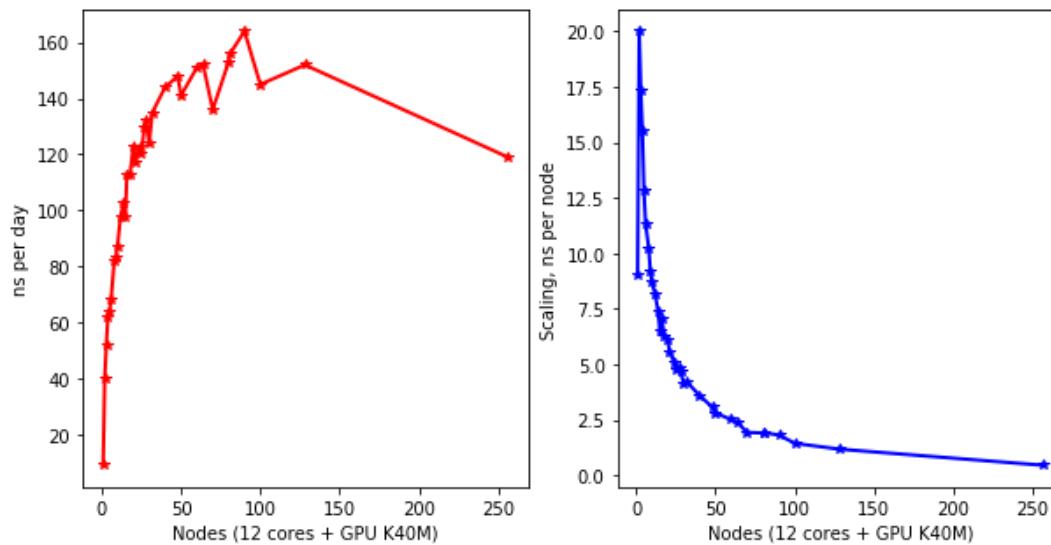
More Bang for Your Buck: Improved Use of GPU Nodes for GROMACS 2018. *Journal of Computational Chemistry* 2019, 40 (27), 2418–2431. <https://doi.org/10.1002/jcc.26011>.



All-atom MD simulations scaling



~300K atoms
using GROMACS
2018 and 2020



~ 50 ns/day using 4 nodes

~ 70 ns/day on 1 node

Intel Haswell-EP E5-
2697v3, 2.6 GHz, 14
cores + NVIDIA Tesla
K40M

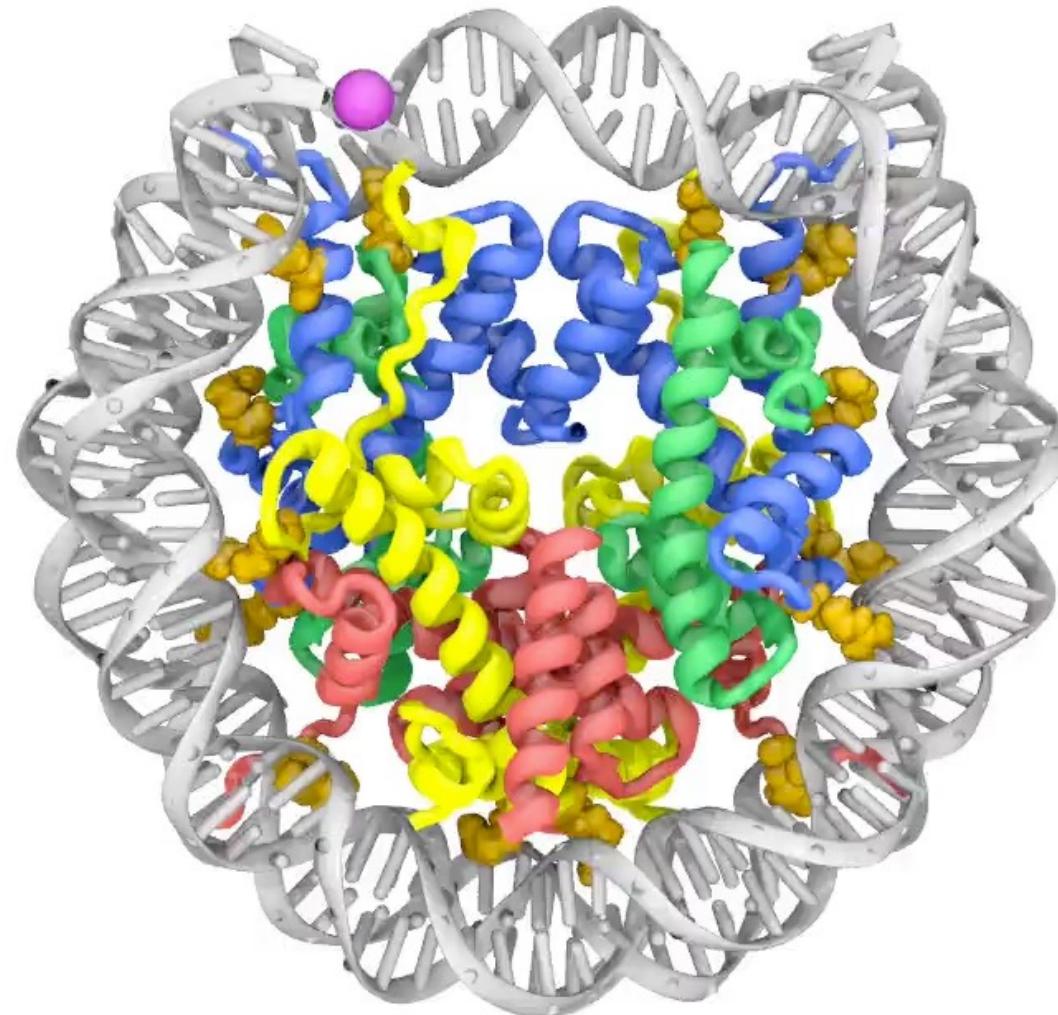
~ 110 ns/day using 4 GPU

Intel Xeon Gold 6240, 18 cores,
2.60GHz - 1497.6 GFlop/s + Nvidia
Tesla V100

NVIDIA DGX2

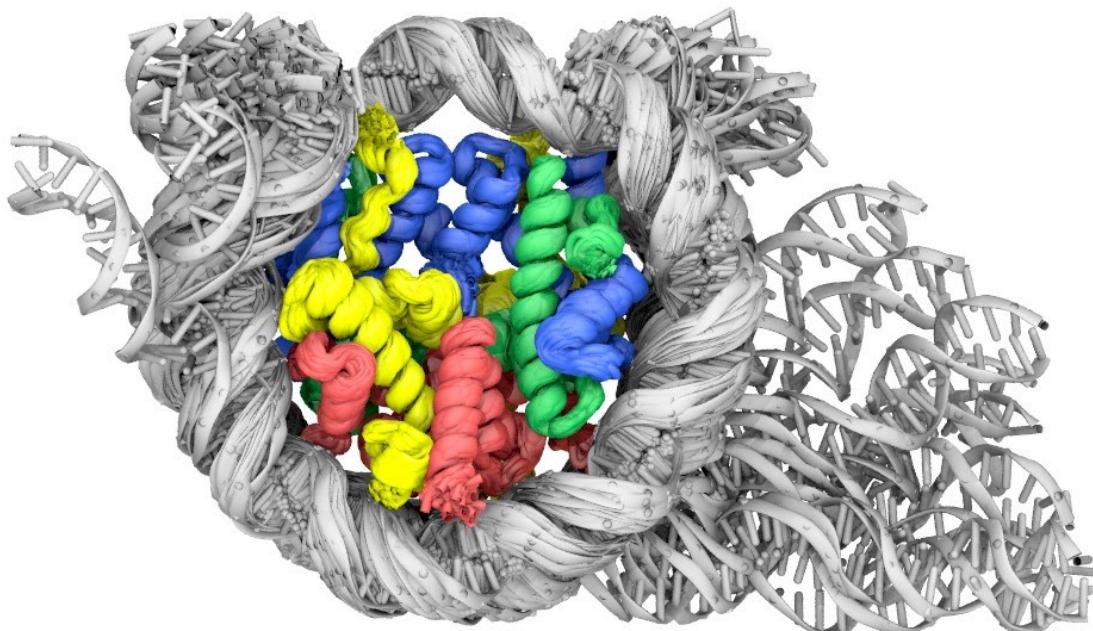
NCP_{145}^{tt}

Histones H3
Histones H4
Histones H2A
Histones H2B
Min groove ARG
DNA
05'DNA chain I

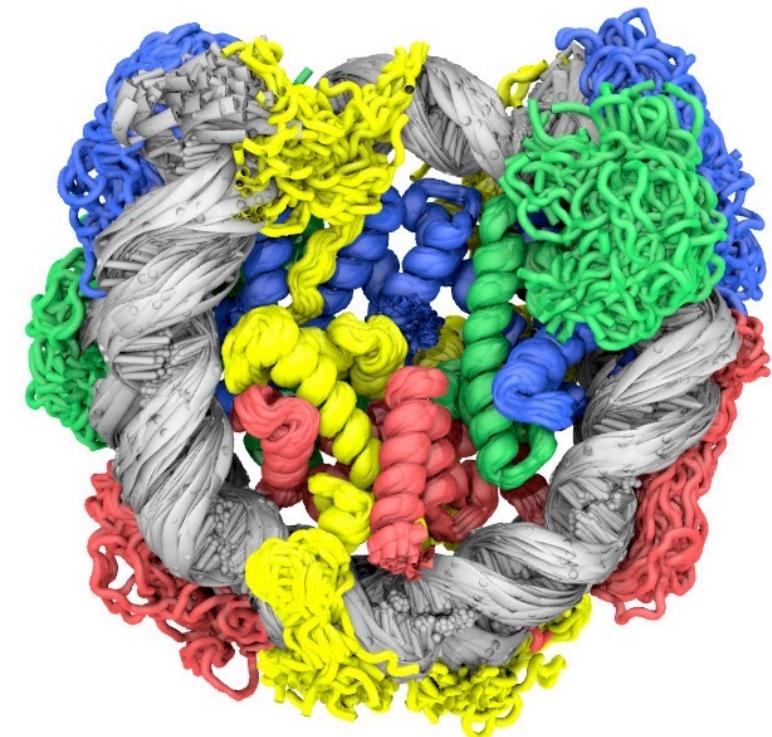


Time: 0.0 ns

Dynamics overview at 15 microsecond scale

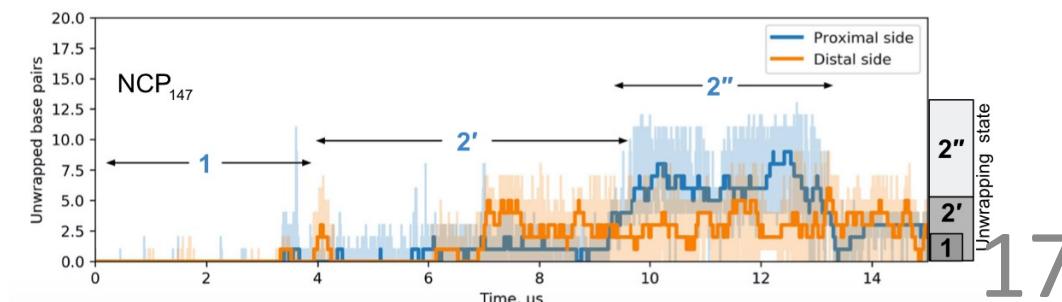
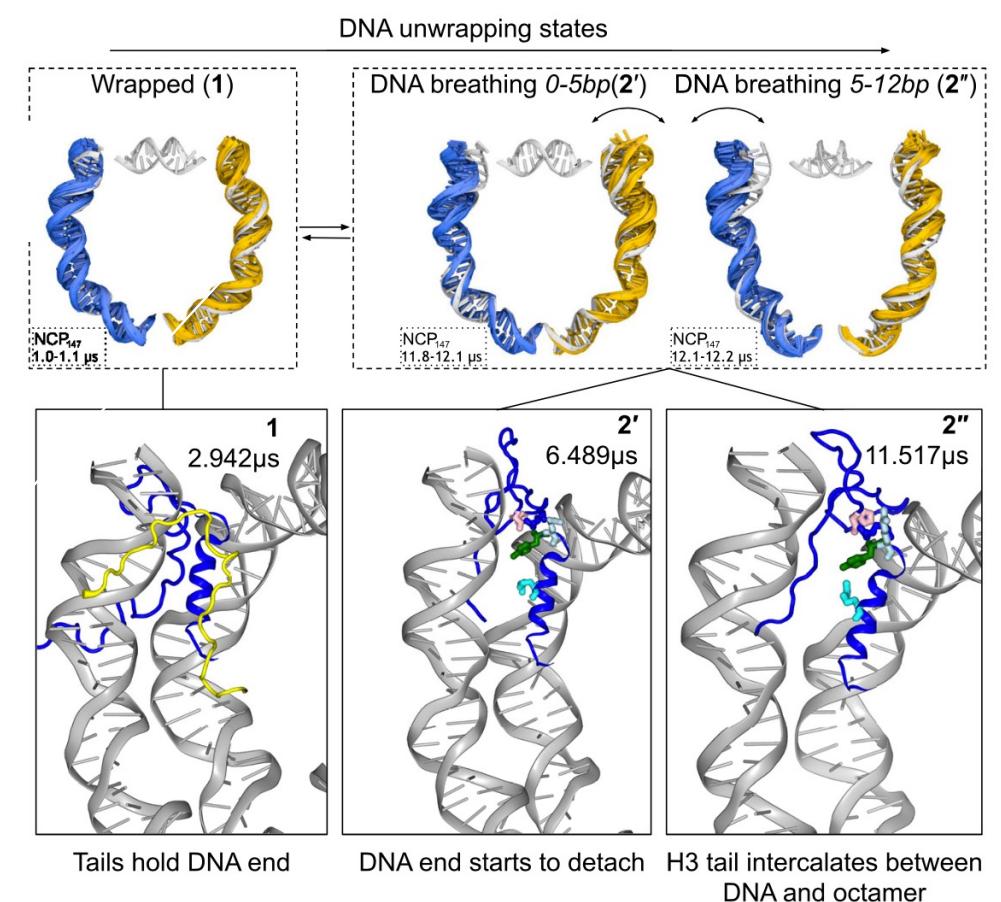
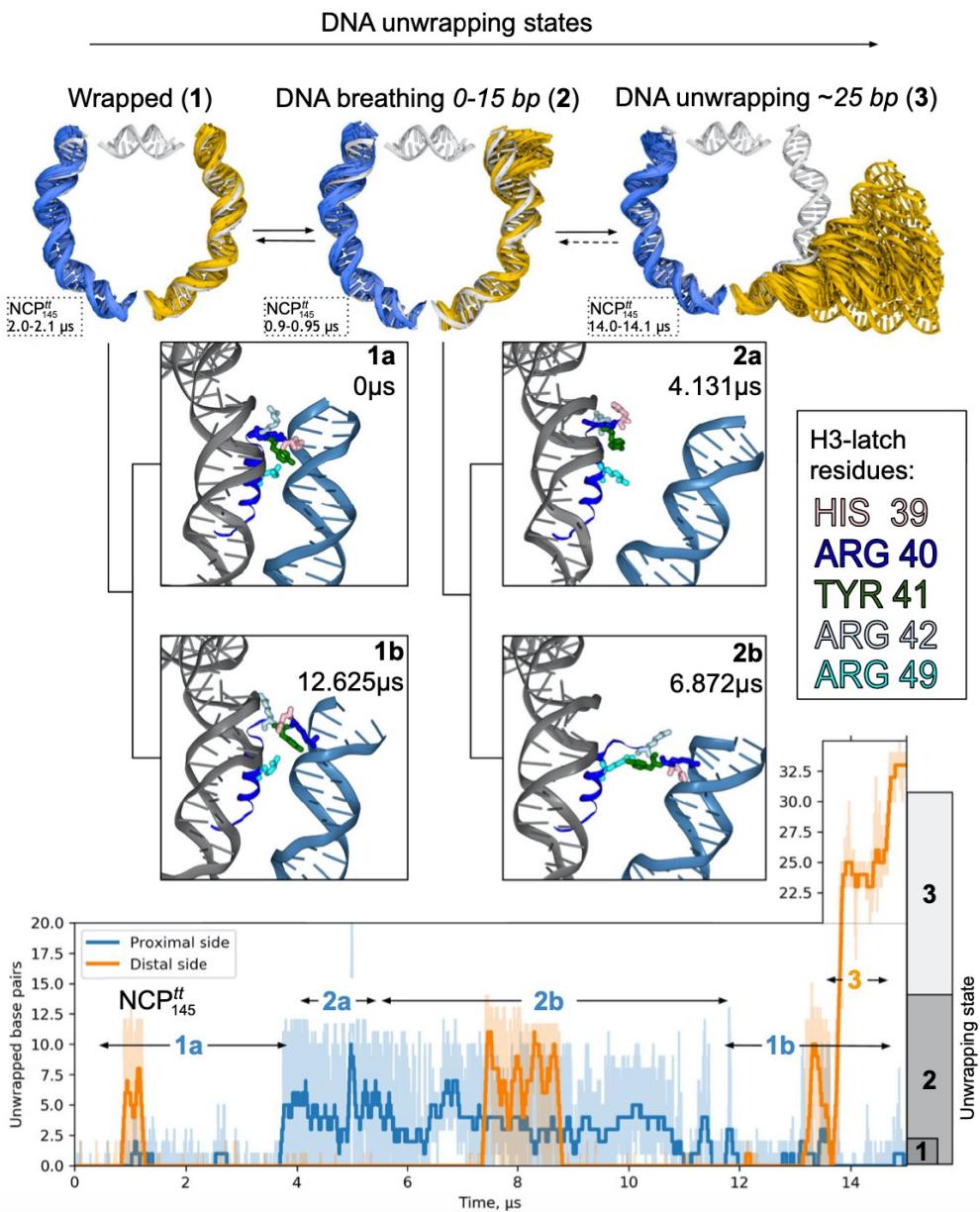


Nucleosome core particles
without histone tails



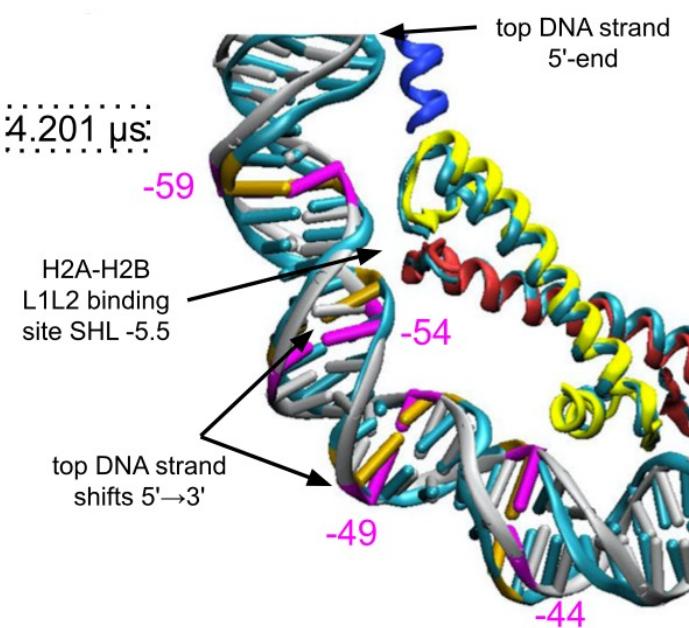
Nucleosome core particles with
full length histone tails

https://intbio.org/Armeev_et_al_2021/ - interactive trajectory viewer

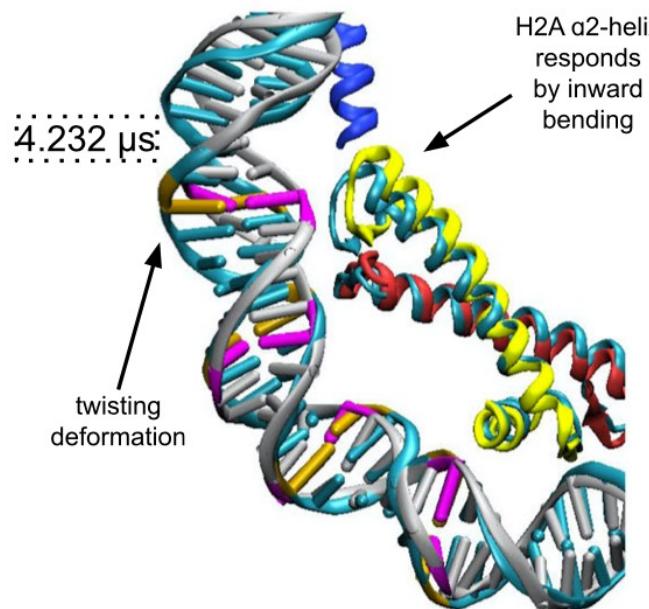


Relaxation of DNA twist-defects in nucleosome

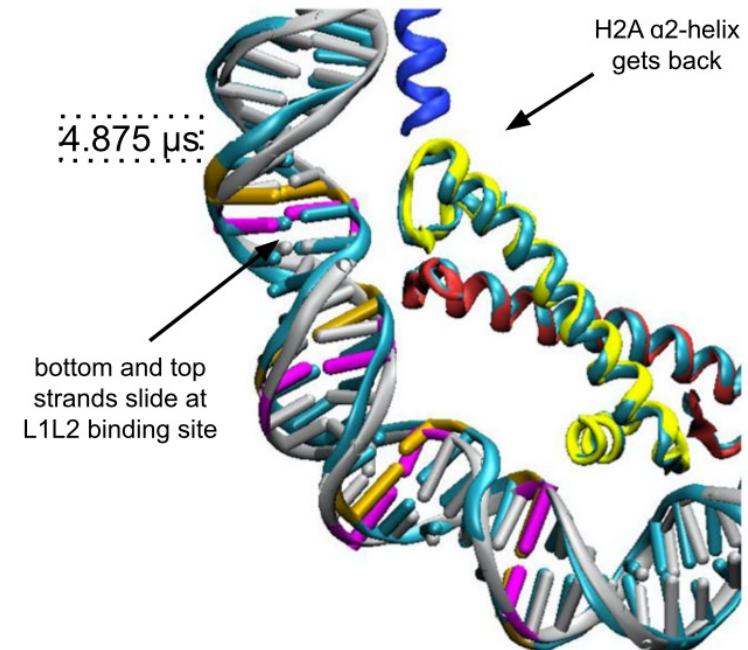
Twist-defect distortion
through top DNA strand sliding



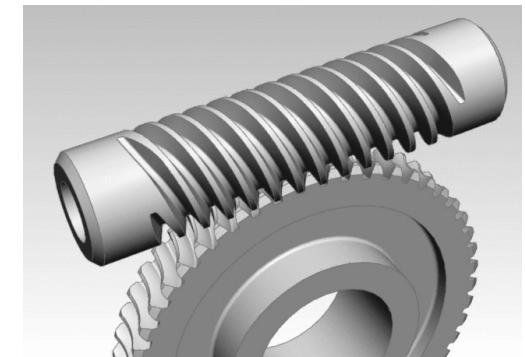
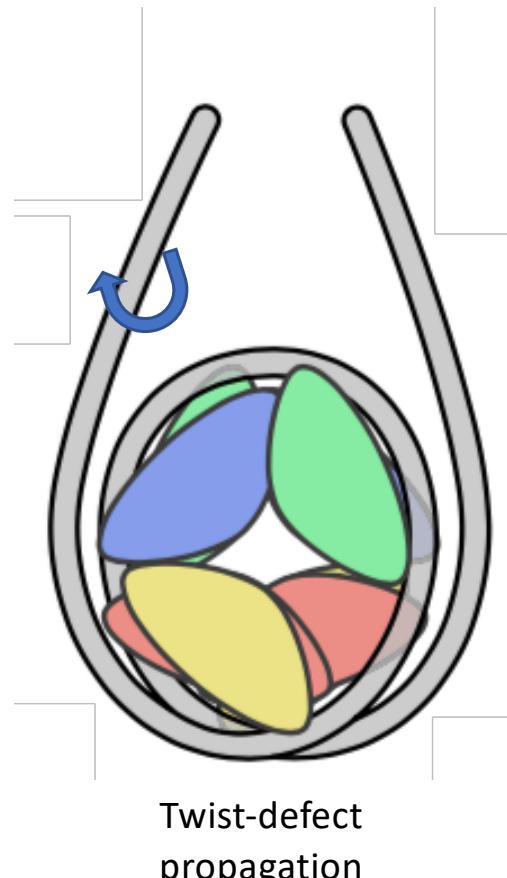
⇒ Overtwisting at SHL -5.5



⇒ DNA slides past SHL -5.5



DNA sliding in nucleosomes



Worm-drive mechanism

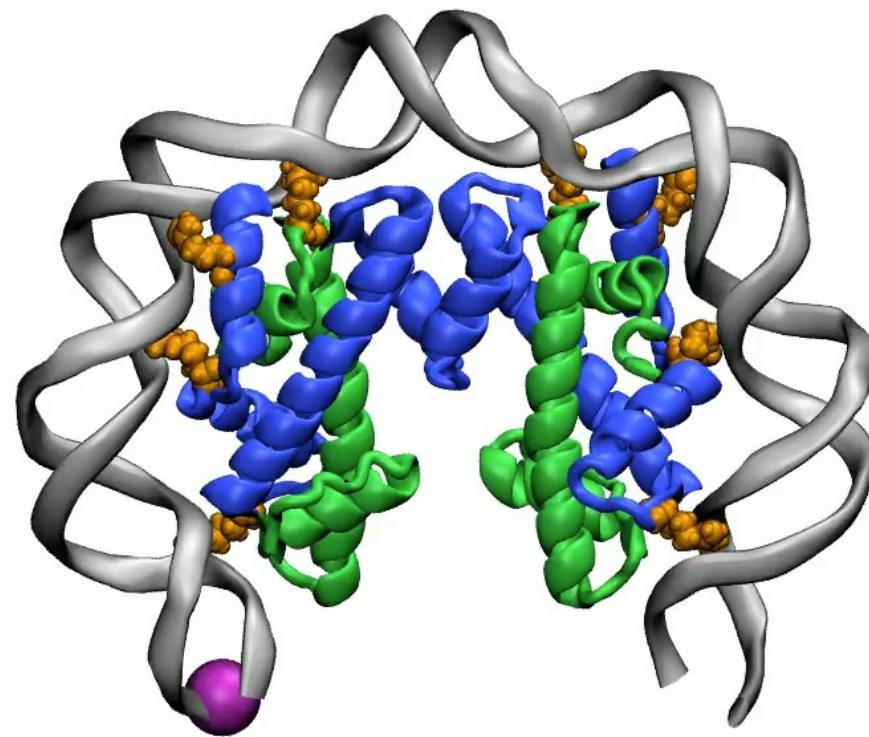


Inchworm mechanism

Steered MD simulations: unpeeling DNA from histones

DNA unpeeling, ABMD simulation

Histones H3
Histones H4
Histones H2A
Histones H2B
Min groove ARG
DNA
AT pairs
GC pairs
05'DNA chain I



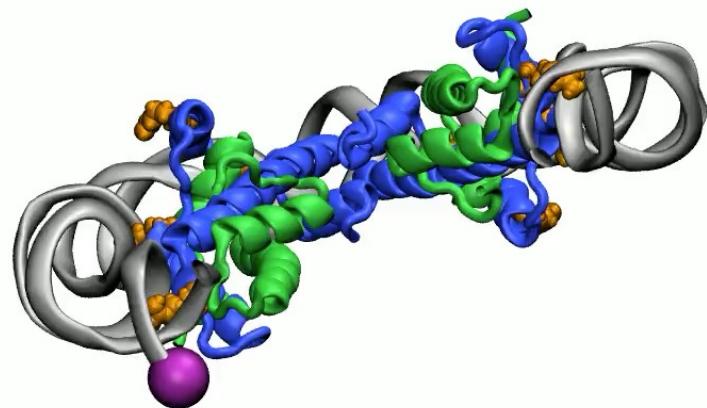
Time: 0.0 ns

Steered MD simulations: DNA supercoiling simulations

Positive supercoiling
stress ("Reversosome")

Biased dynamics: positive supercoiling

Histones H3
Histones H4
Histones H2A
Histones H2B
Minor groove ARG
DNA
AT pairs
GC pairs
5' DNA chain I

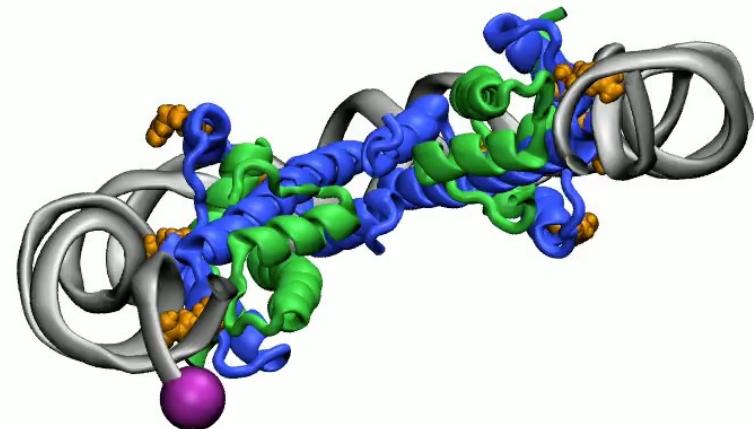


Time: 0.0 ns

Negative supercoiling stress

Biased dynamics: negative supercoiling

Histones H3
Histones H4
Histones H2A
Histones H2B
Minor groove ARG
DNA
AT pairs
GC pairs
5' DNA chain I



Time: 0.0 ns

Thank you!



Dr. Grigoriy Armeev



Anastasiia Kniazeva

Supported by HPC computing resources
at Lomonosov Moscow State University.



Russian
Science
Foundation

RSF grant # 18-74-10006
RFBR grant # 20-34-70039

