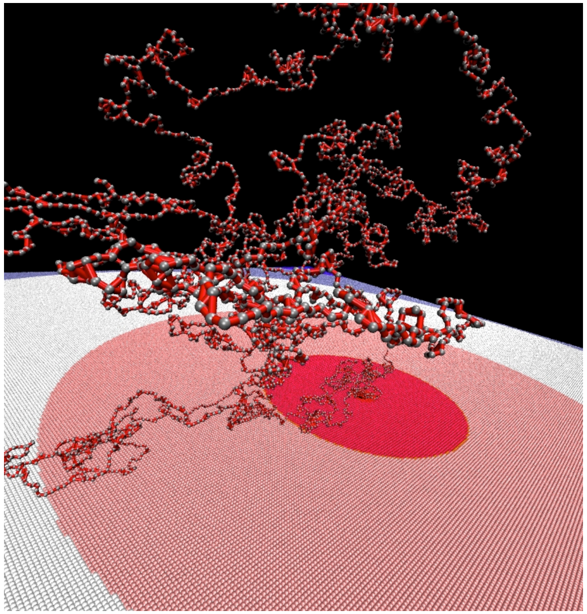
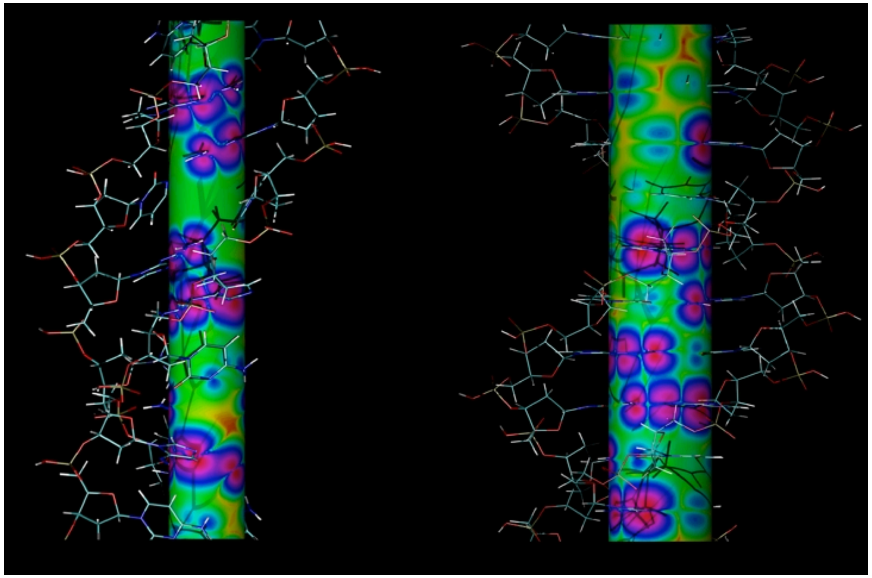
Computational Investigation of Complex Biomolecular Units

- Motivation: sequencing, epigenetics, replication and repair, ...
- Coarse-grained potential: model, simulations, validation
- Translocation dynamics through nanopores: electronic sequencing

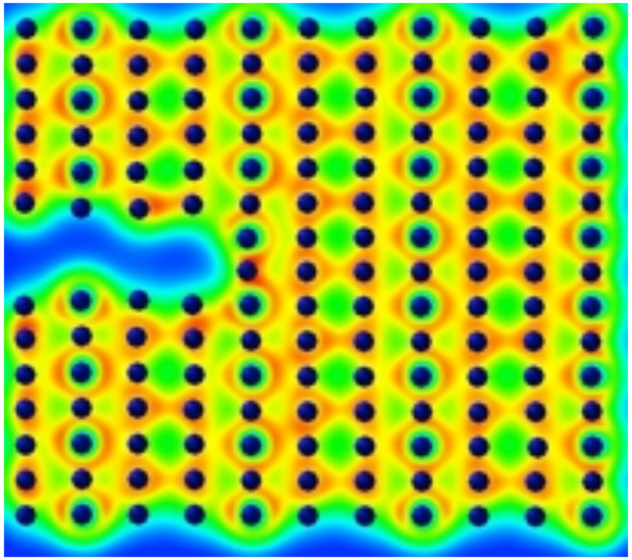
3D EUROPEAN PhD SUMMER SCHOOL AND WORKSHOP ON
"MATHEMATICAL MODELING OF COMPLEX SYSTEMS"
15 – 26 JULY , 2013
DEPARTMENT OF PHYSICS , UNIVERSITY OF CRETE



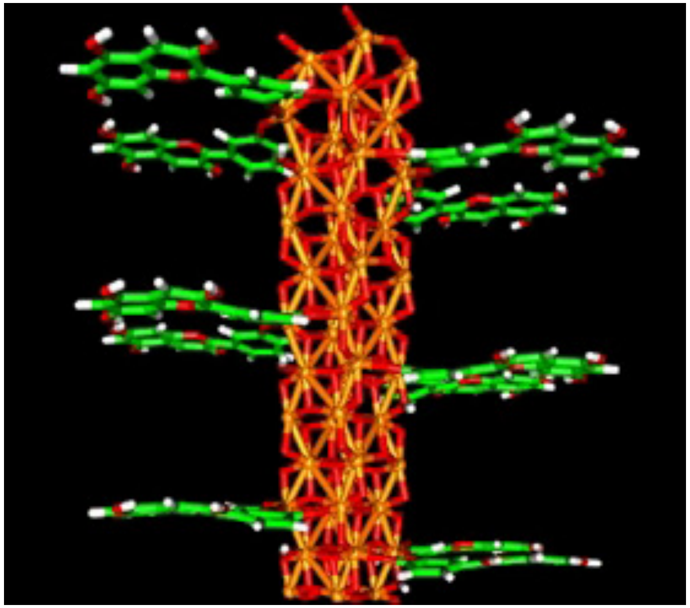
DNA translocation



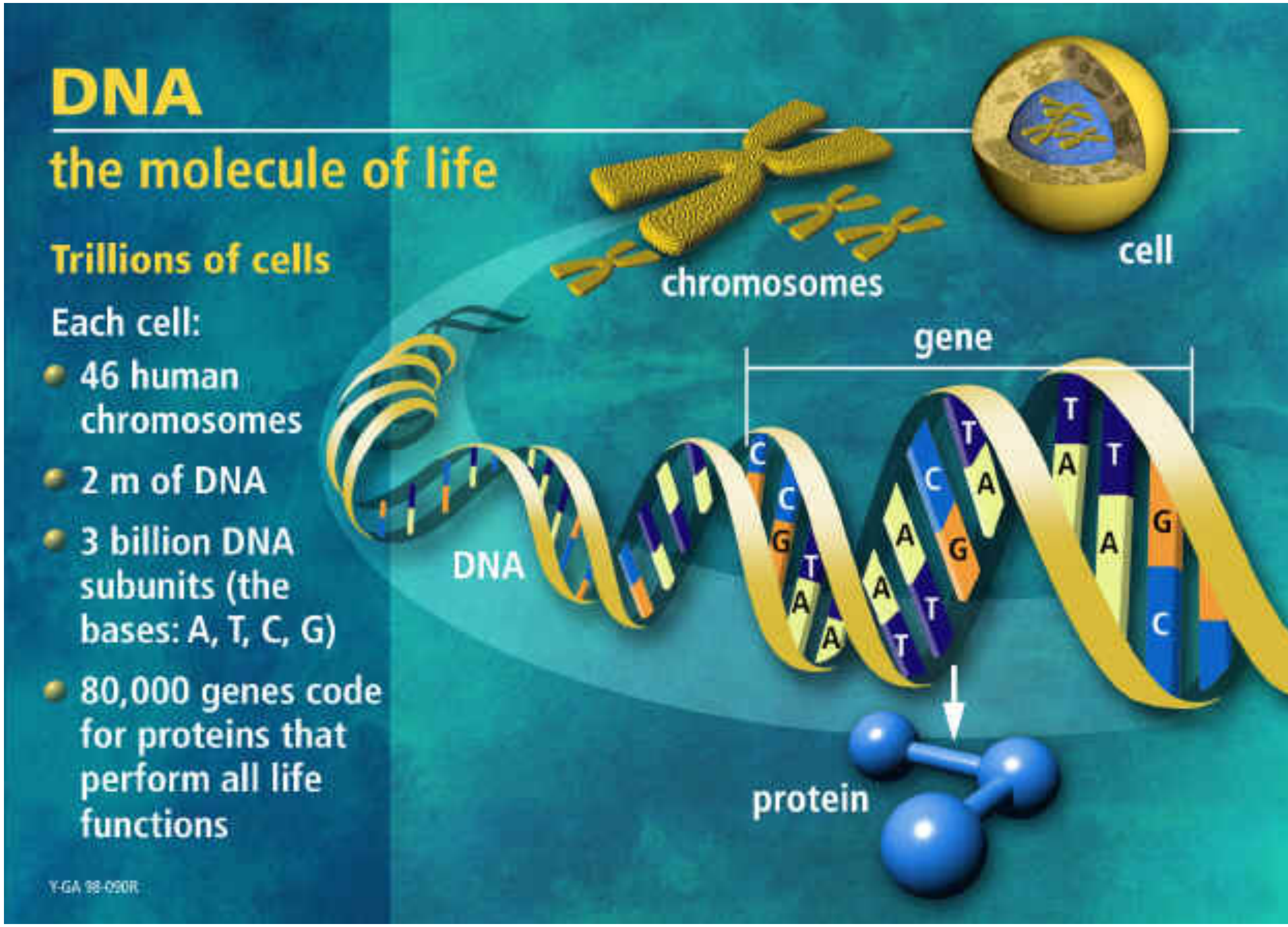
DNA structural/electronic properties

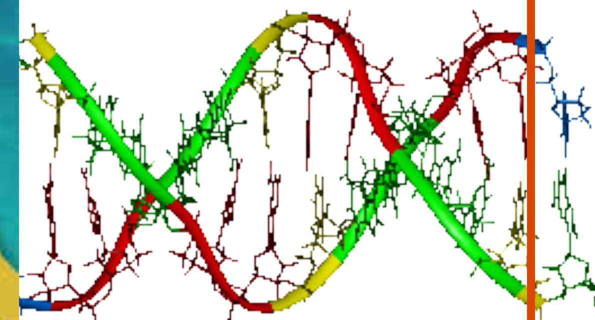


Mechanochemistry



Nanostructures

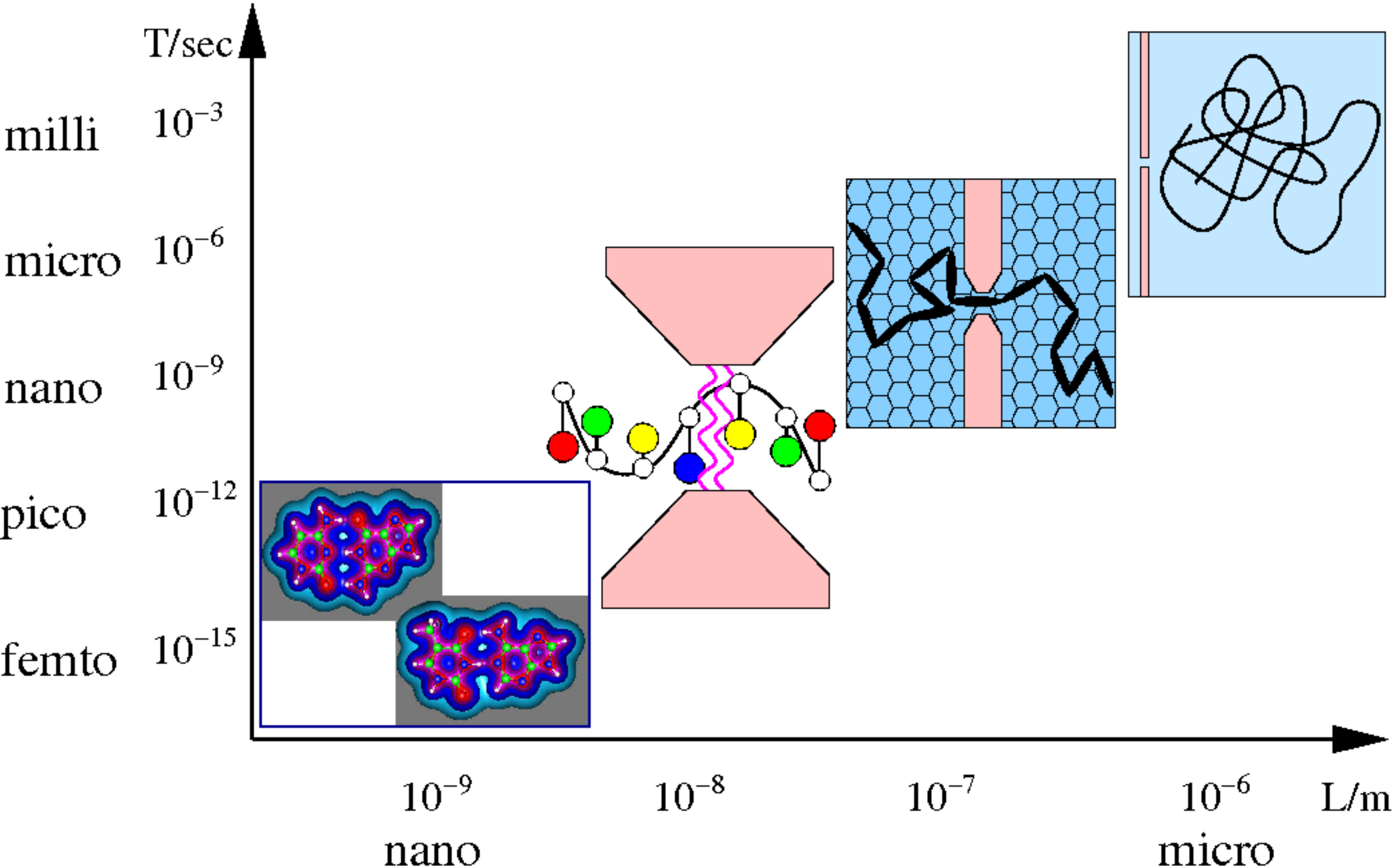




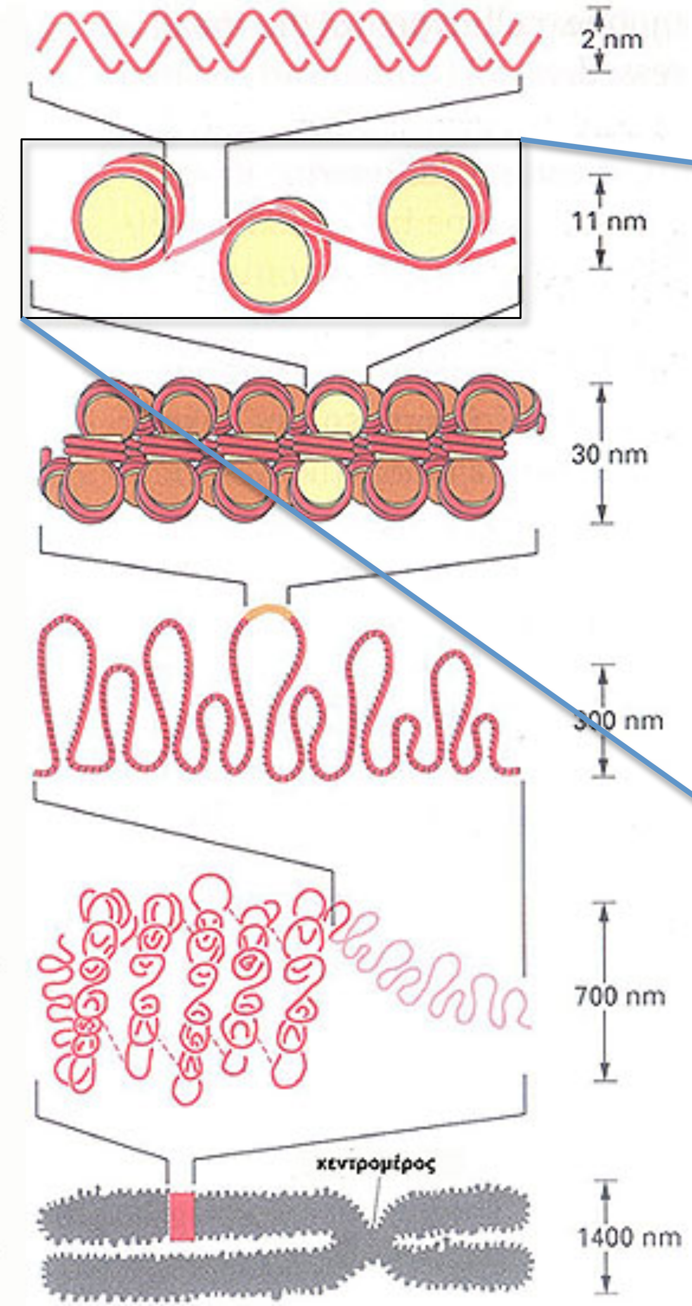
DNA sequencing: biochemical methods – cut pieces (ending at specific bases), measure by **gel electrophoresis**.

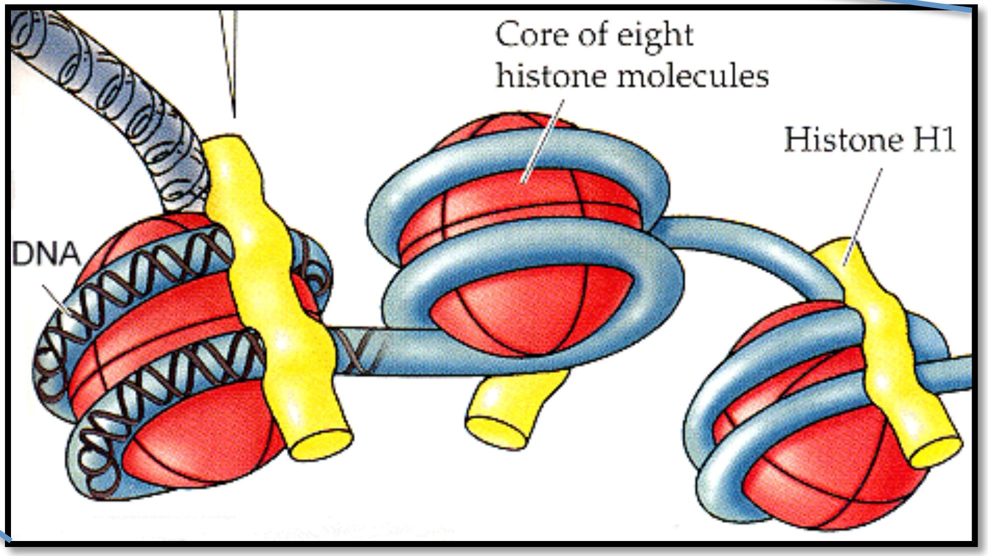
Instead, use **electronic signature** for sequencing – multiscale process

Multiple length/time scales: DNA electronic sequencing



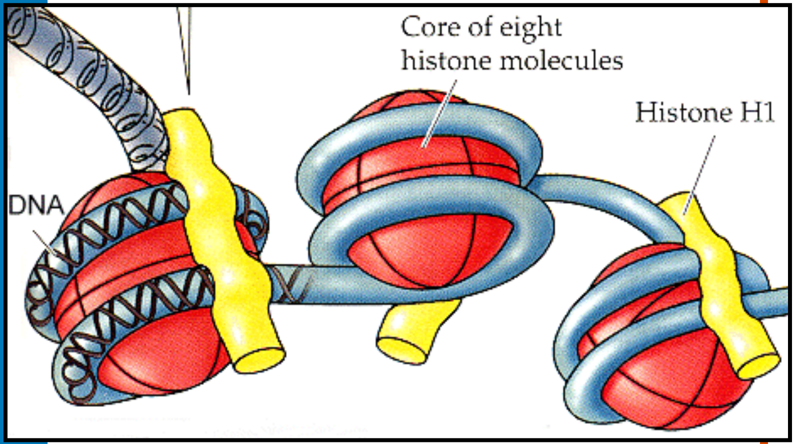
Structure of DNA on several scales: From NUCLEOSOME to CHROMOSOME

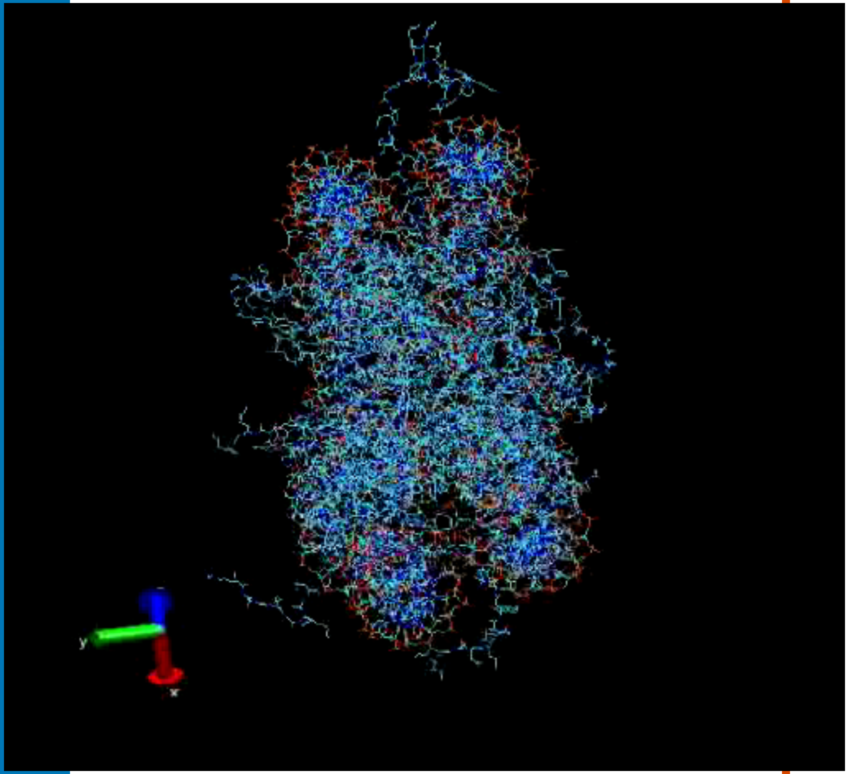




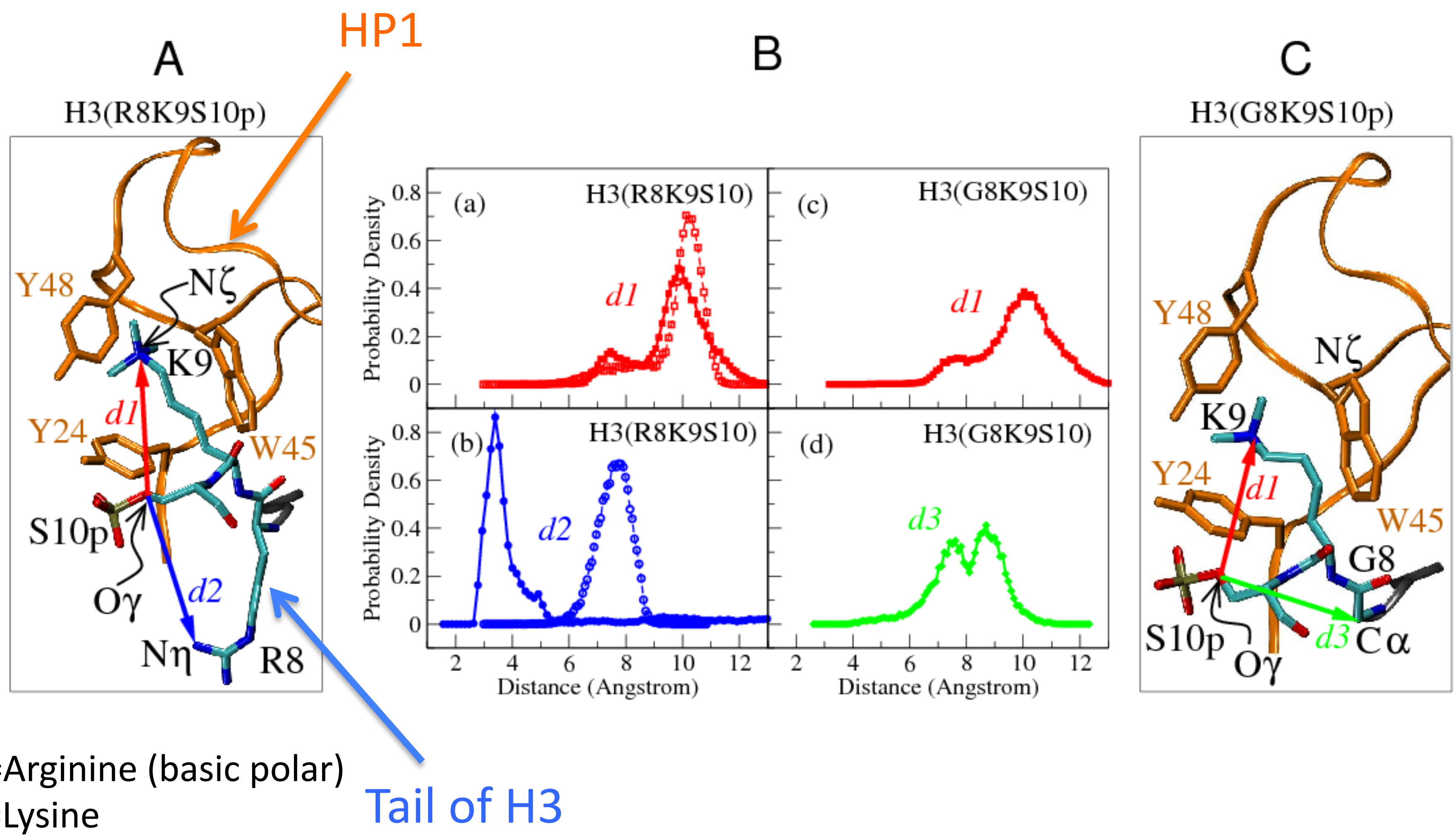
EPIGENETICS: passing genetic information **NOT** encoded in DNA base sequence

- DNA methylation
- Histone modification (acetylation, methylation, phosphorylation, …)



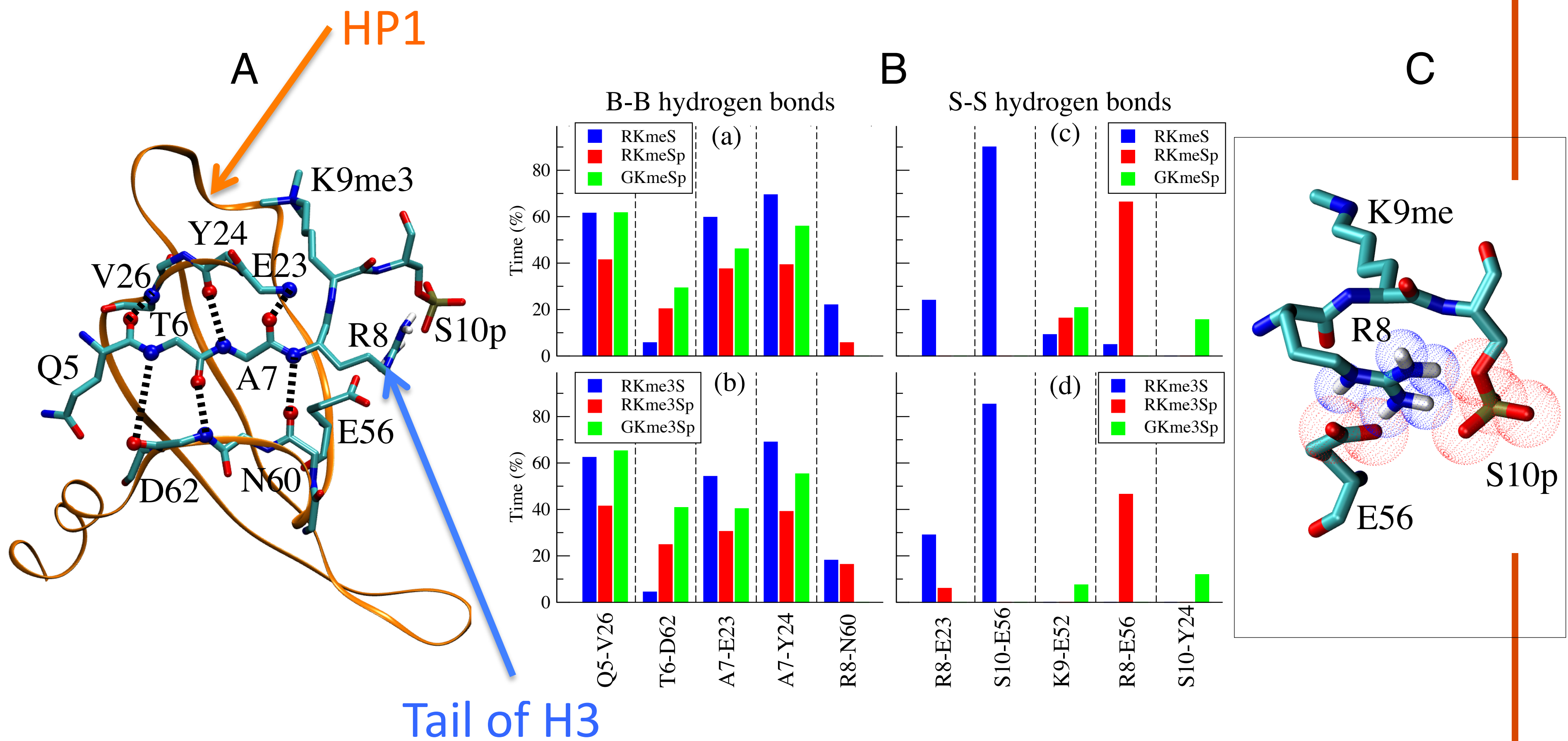


Papamokos *et al.*,
Biophysical J. (2012)



R=Arginine (basic polar)
K=Lysine
S=Serine
G=Glycine (non-polar)
W=Tryptophan
Y=Tyrosine

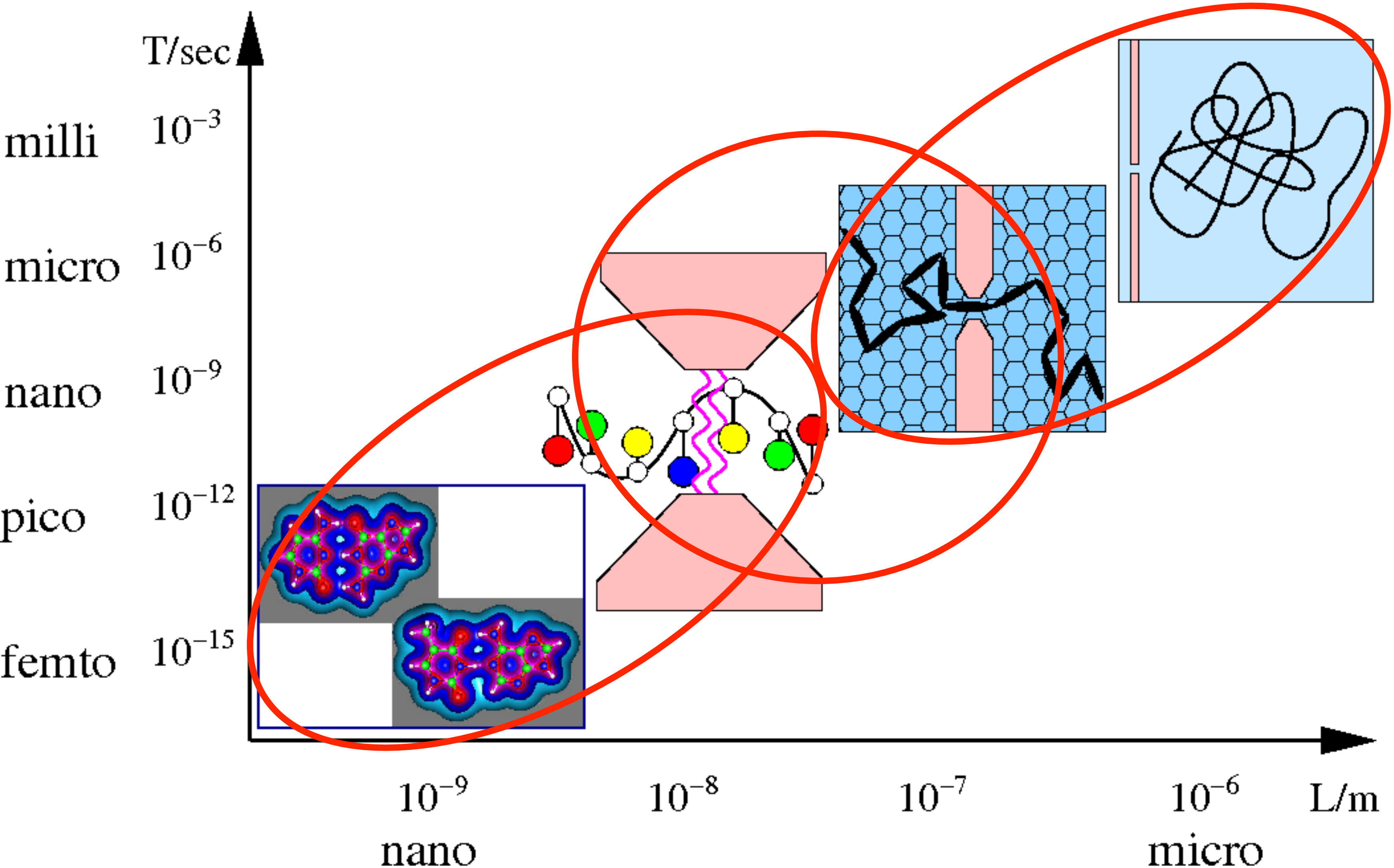
"Structural Role of RKS Motifs in Chromatin Interactions: An MD Study of HP1 Bound to a Variably Modified Histone Tail"
Papamokos *et al*., Biophysical J. (2012)



R=Arginine (basic polar)
K=Lysine
S=Serine
G=Glycine (non-polar)
W=Tryptophan
Y=Tyrosine

Papamokos *et al.*, Biophysical J. (2012)

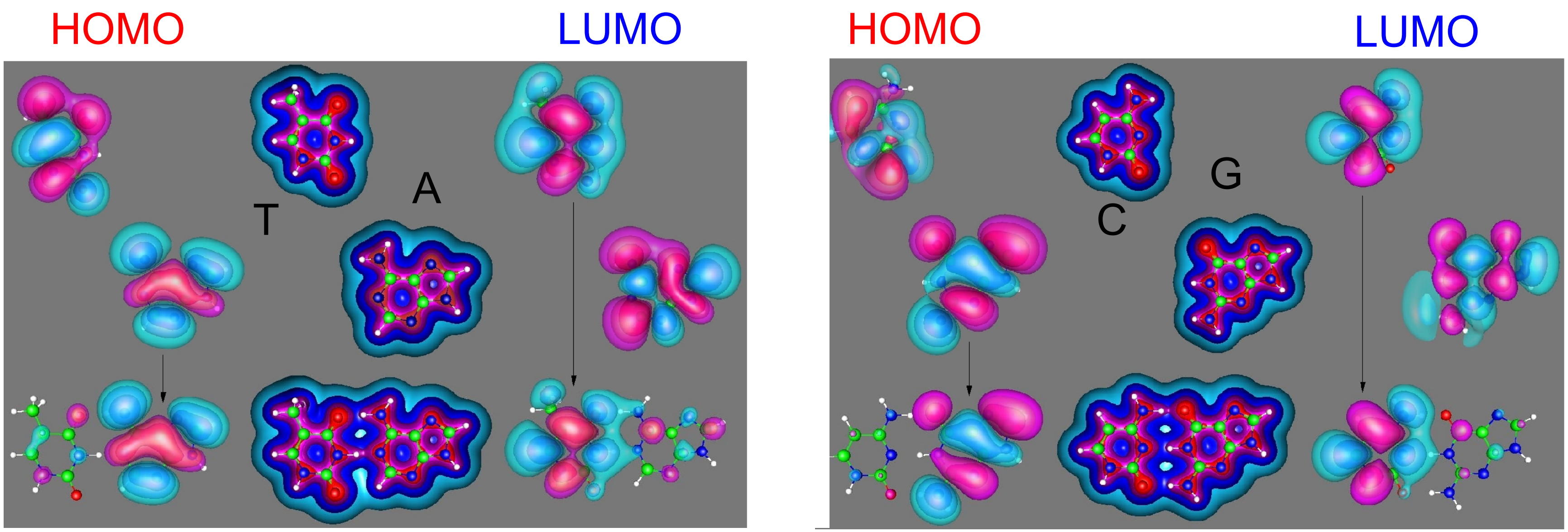
Multiple length/time scales: DNA electronic sequencing



Electronic states of stacked DNA base-pairs

"Multiscale model of electronic behavior and localization in stretched dry DNA"

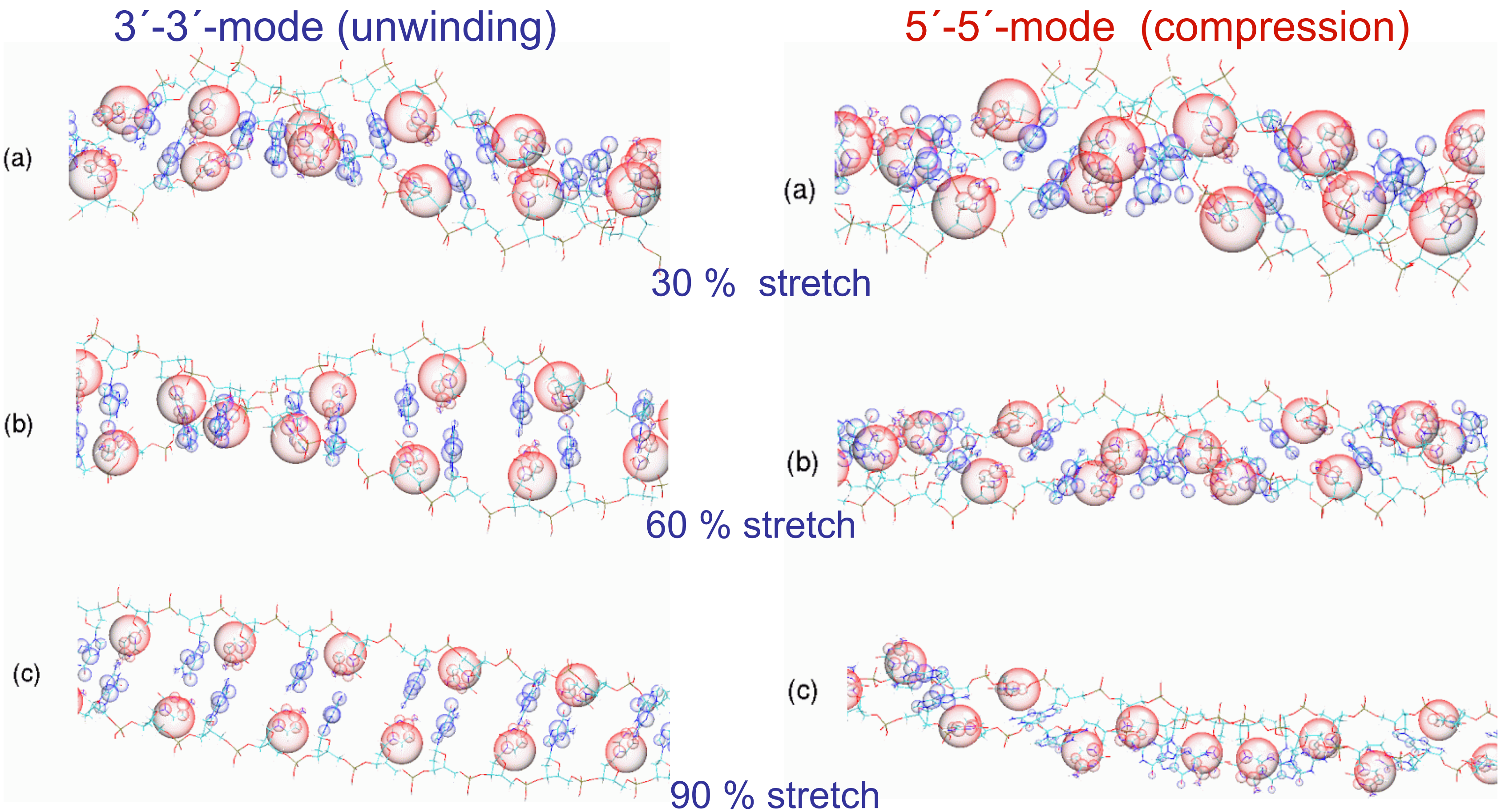
R. Barnett, P. Maragakis, A. Turner, M. Fyta, EK; *J. Mater. Sci.* (2007)



Results from first-principles electronic structure calculations using Density-Functional-Theory (DFT)

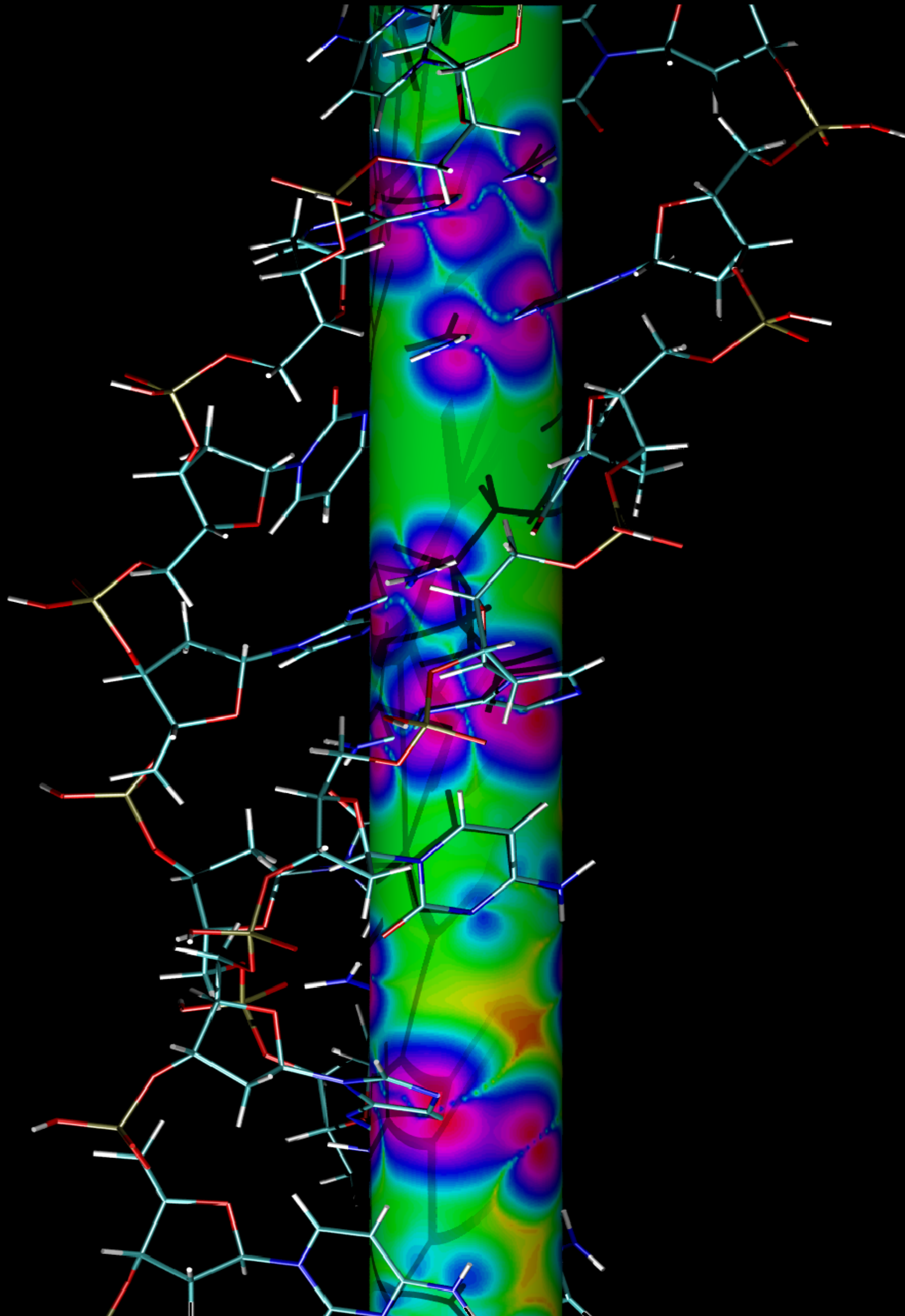
Multiscale issues: structure of double-helix (mesoscopic - microns) affects coupling of electronic states (microscopic - Angstroms)

Different stretching modes

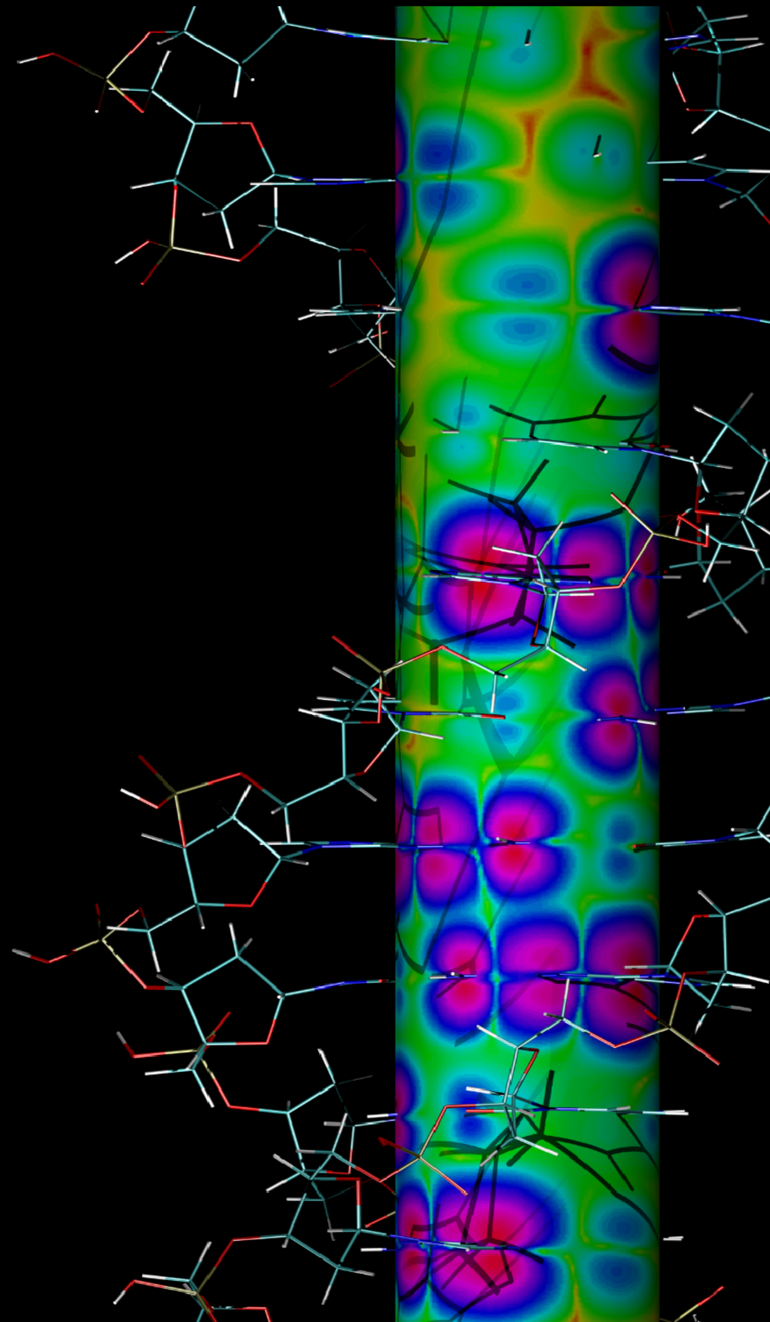


A. Lebrun and R. Lavery, *Nucleic Acids Research* 24, 2260 (1996)

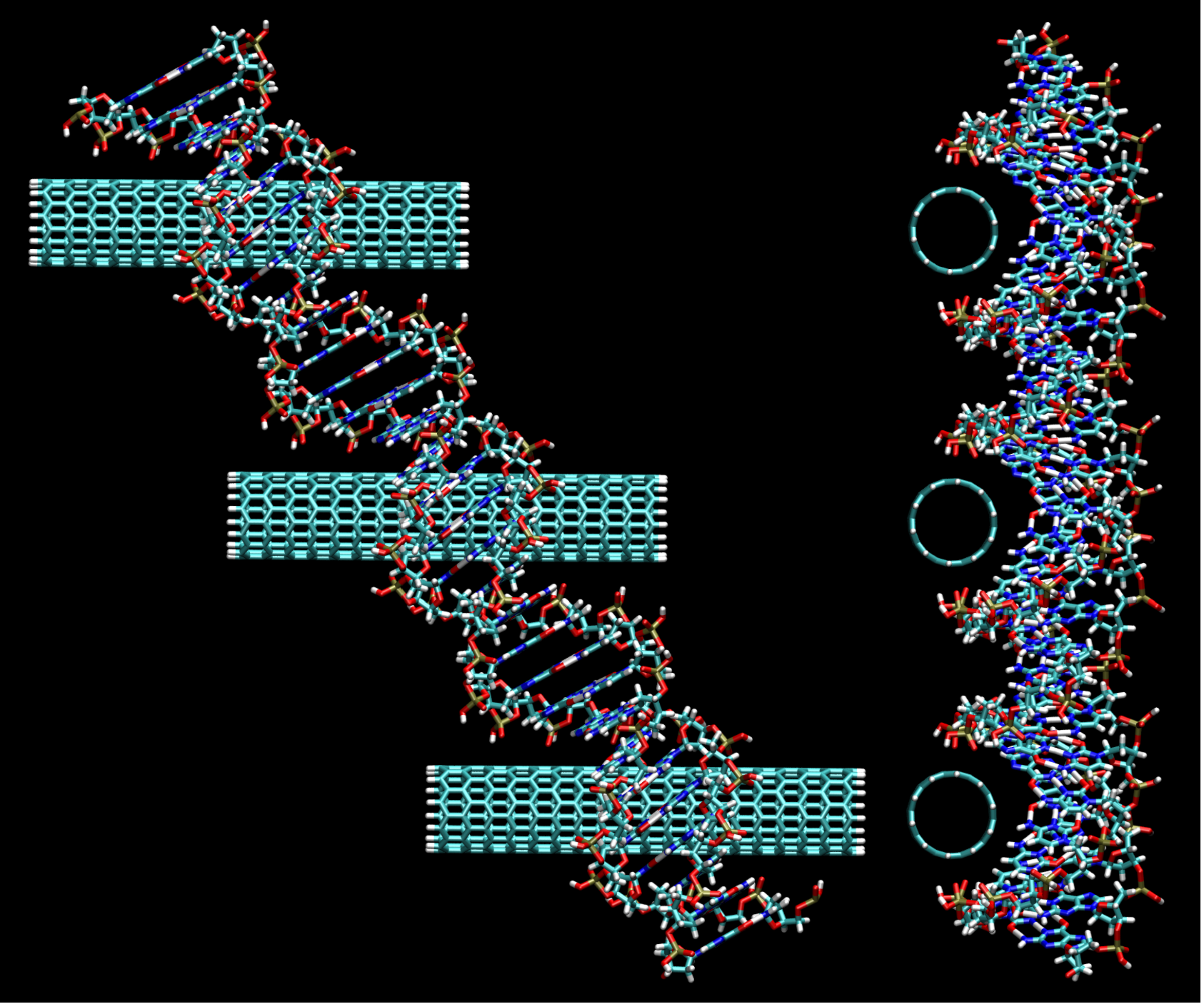
Structures from classical MD simulations (**empirical!**)
Electronic states from DFT calculations (**first-principles**)



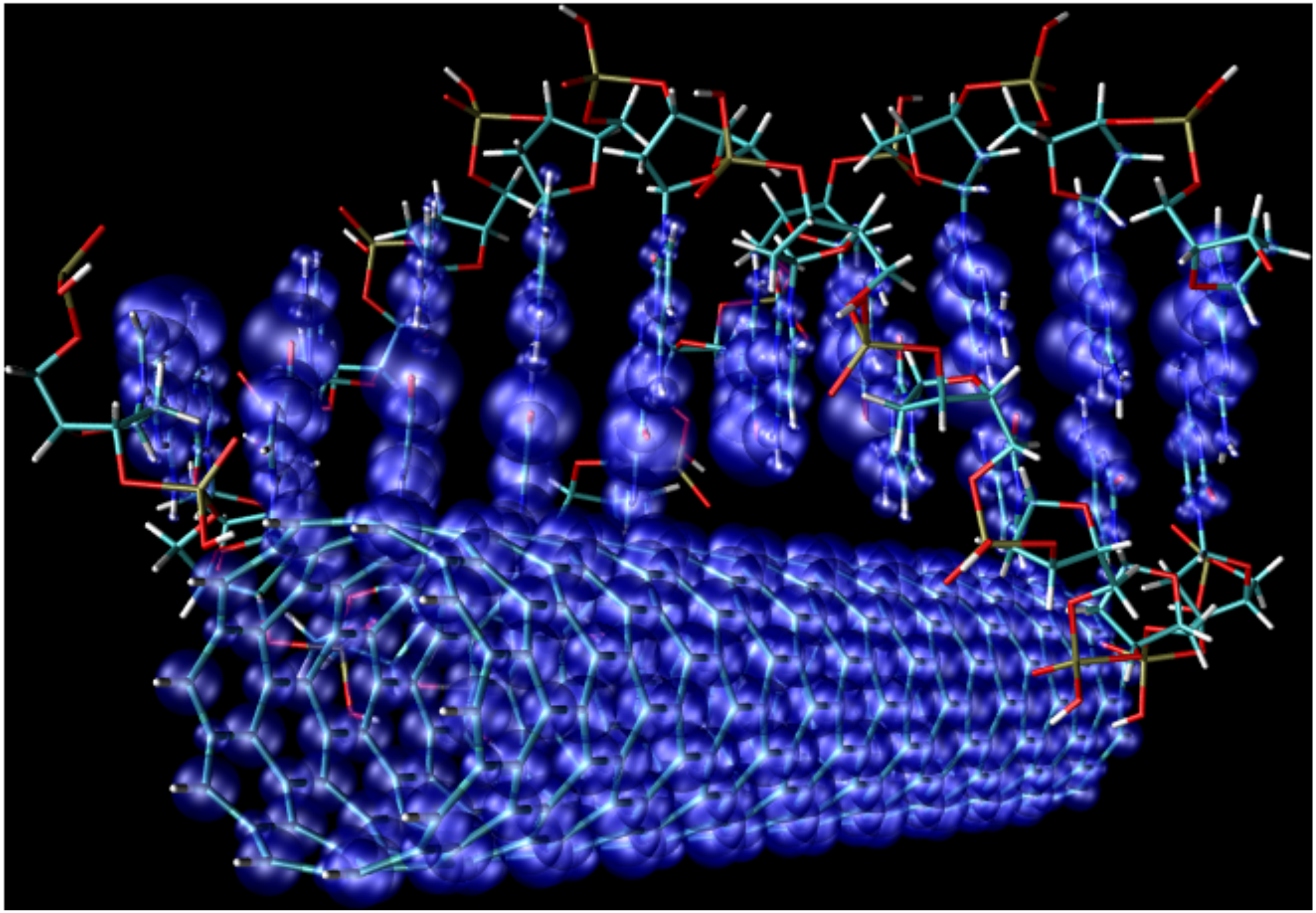
30 % overstretched



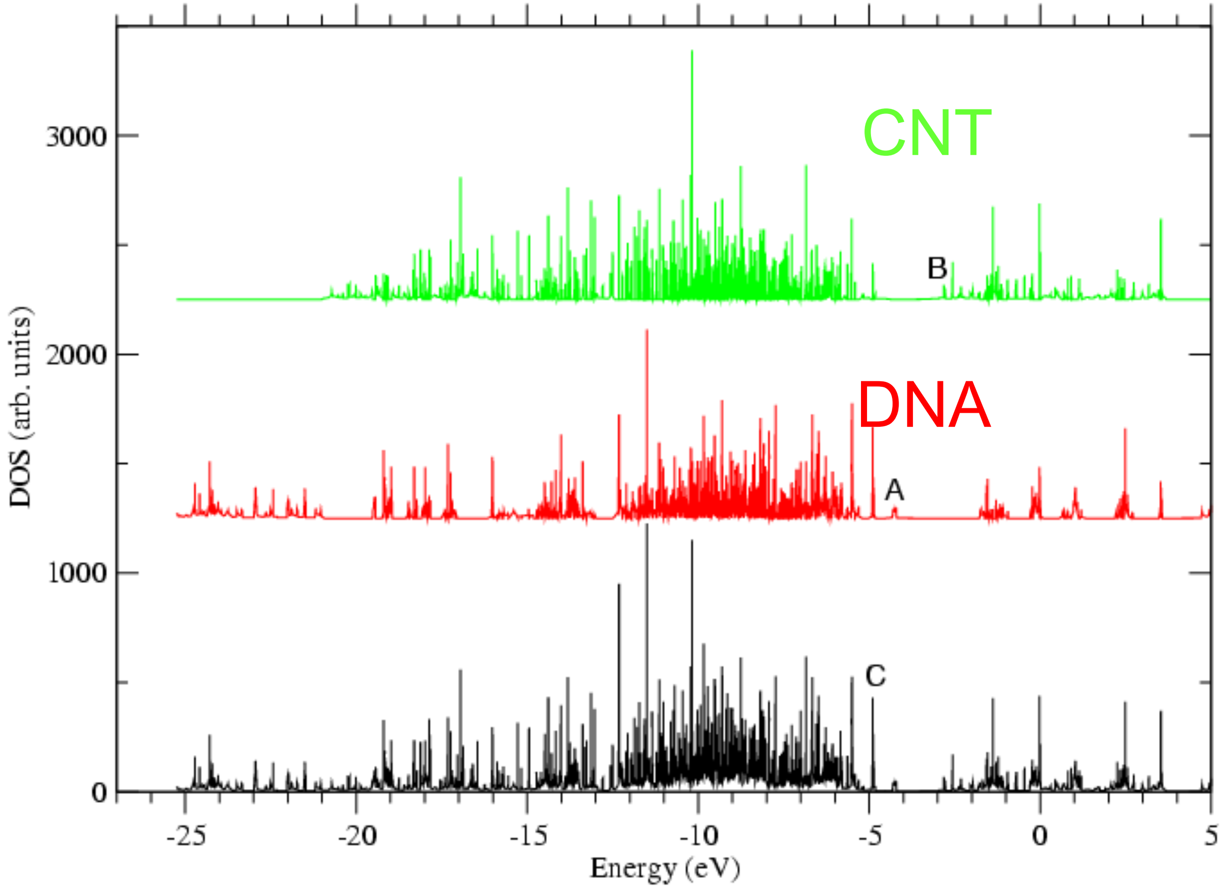
B-DNA at natural length



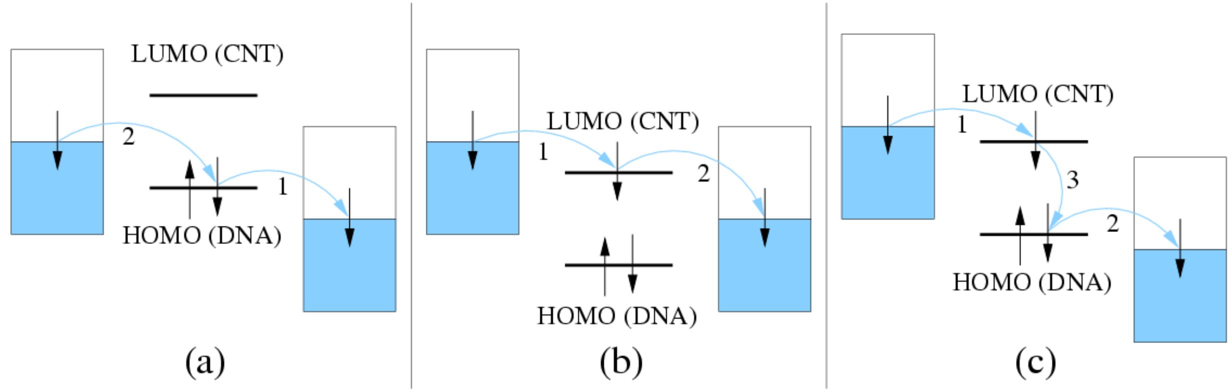
“DNA – carbon nanotube interaction”
G. Lu, P. Maragakis, E. Kaxiras, (NanoLett. 2005)



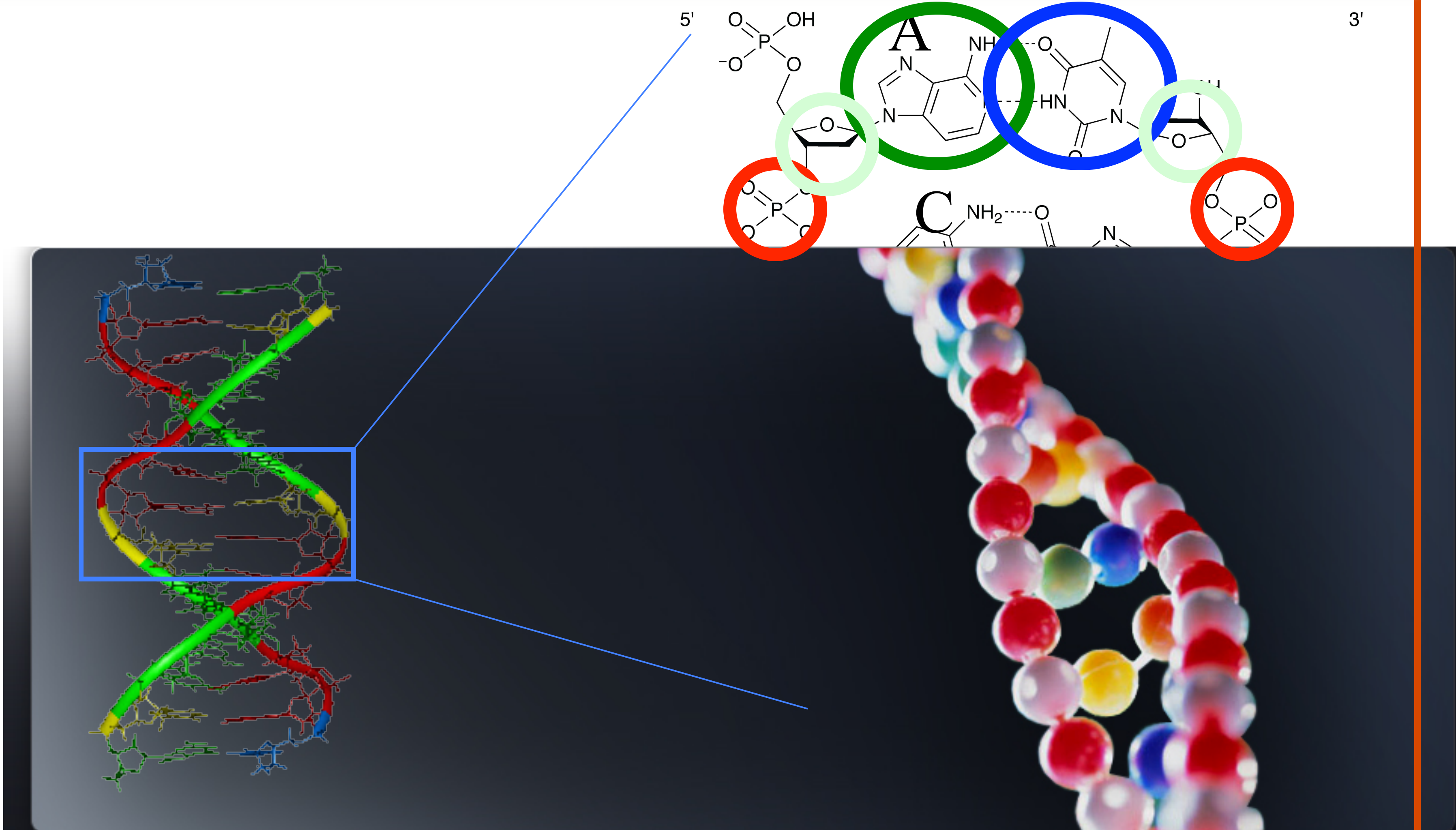
Hypothetical mechanism: tunneling between CNT/DNA
→ sequencing through intimate base-CNT contact



DNA-CNT electronic device:



Coarse-grained potential for DNA structure



W. Hsu, M. Fyta, G. Lakatos, S. Melchionna, EK (2012)

Goal: derive coarse-grained potential:
minimal (but sufficient) model;
all parameters from *ab-initio* calculations*.
Assumption: **separable** interactions

- Hydrogen bonding (distance; dihedral, flip angles)

- Stacking interactions (distance; twist angle)

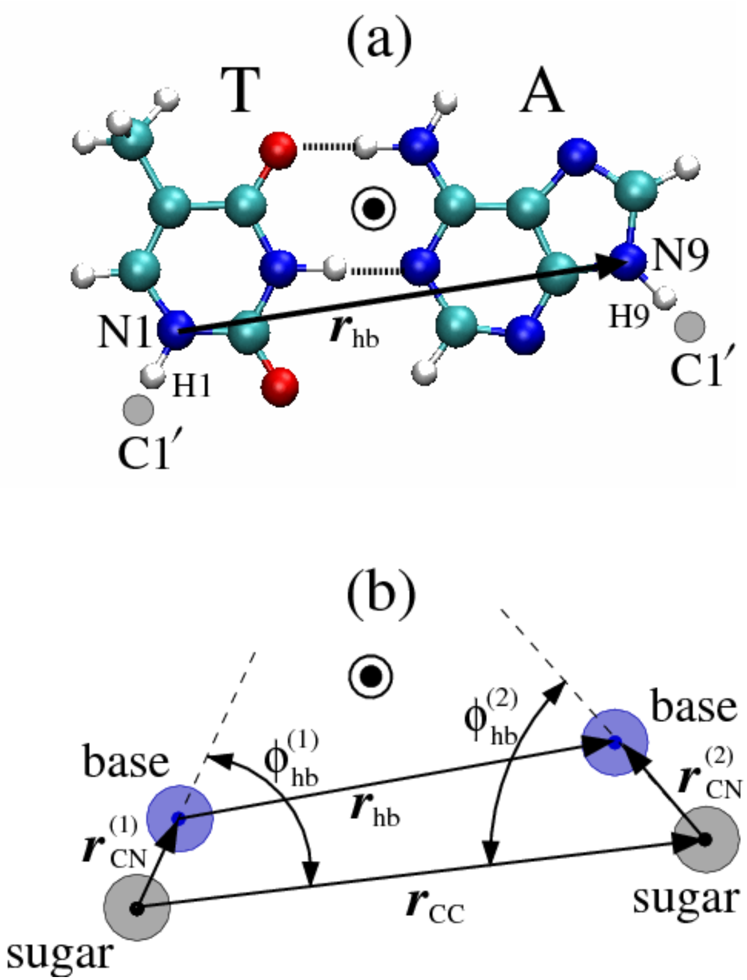
- Backbone interactions (distance; 3'-5' orientation)

- Electrostatic interactions

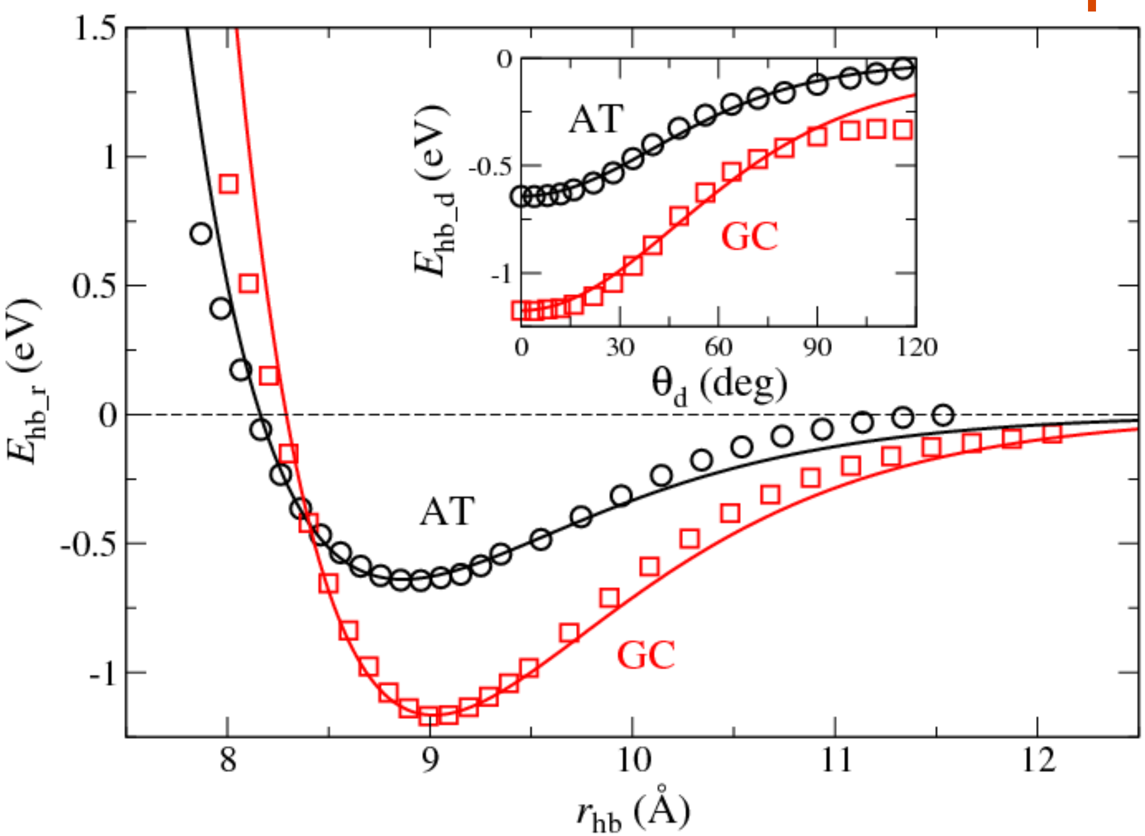
Data points: DFT calculations,
Lines: fits with simple curves

* to the extent possible

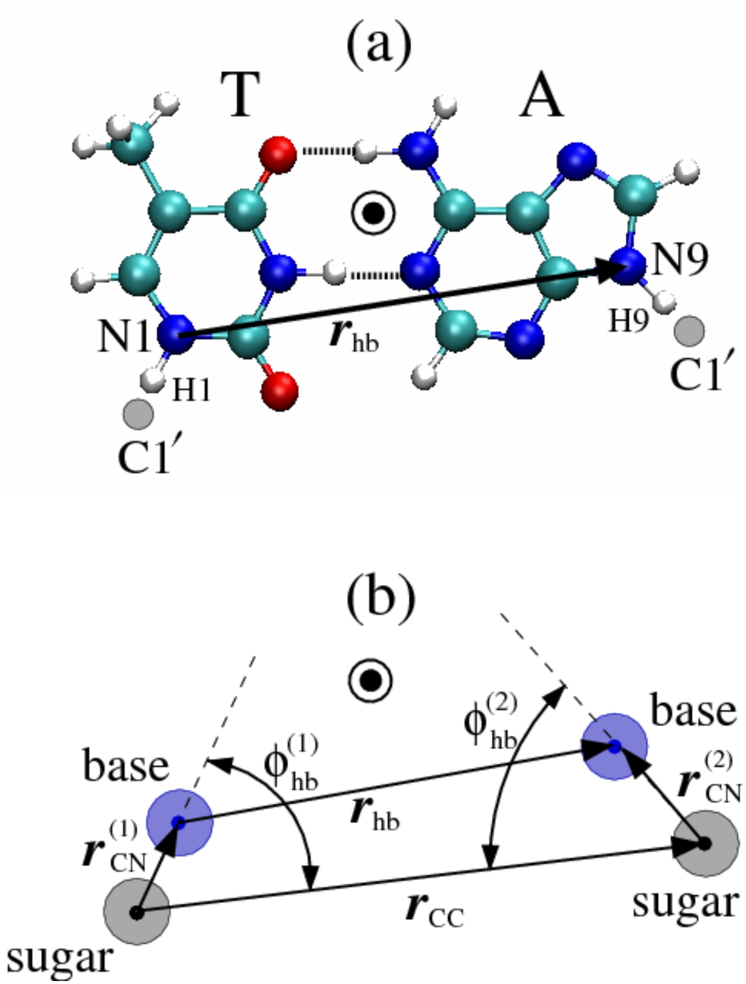
Coarse-grained potential – hydrogen bonding I



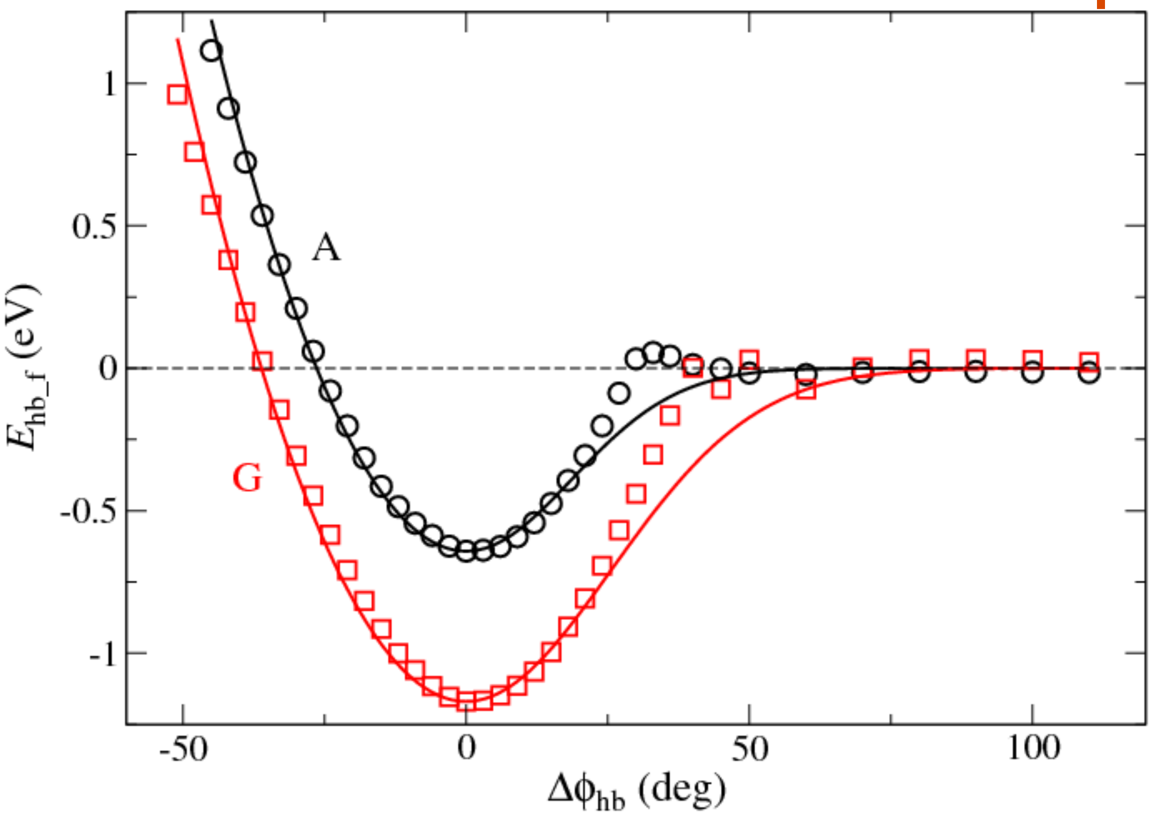
Distance between bases, dihedral angle



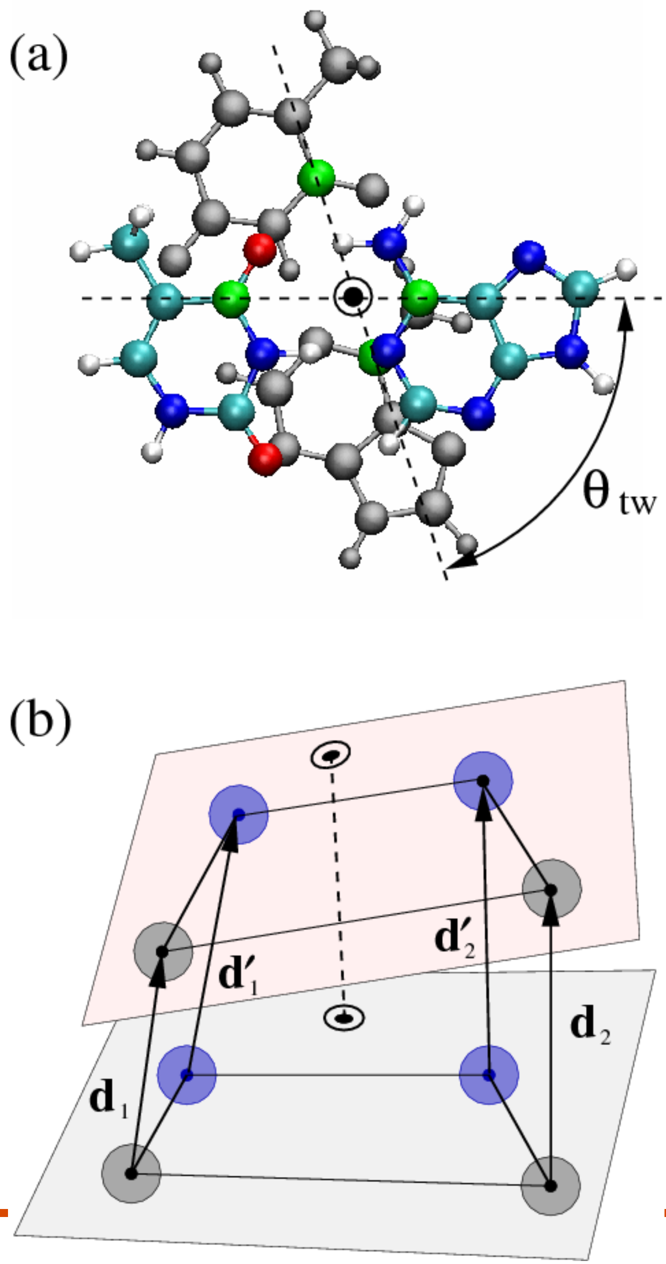
Coarse-grained potential – hydrogen bonding II



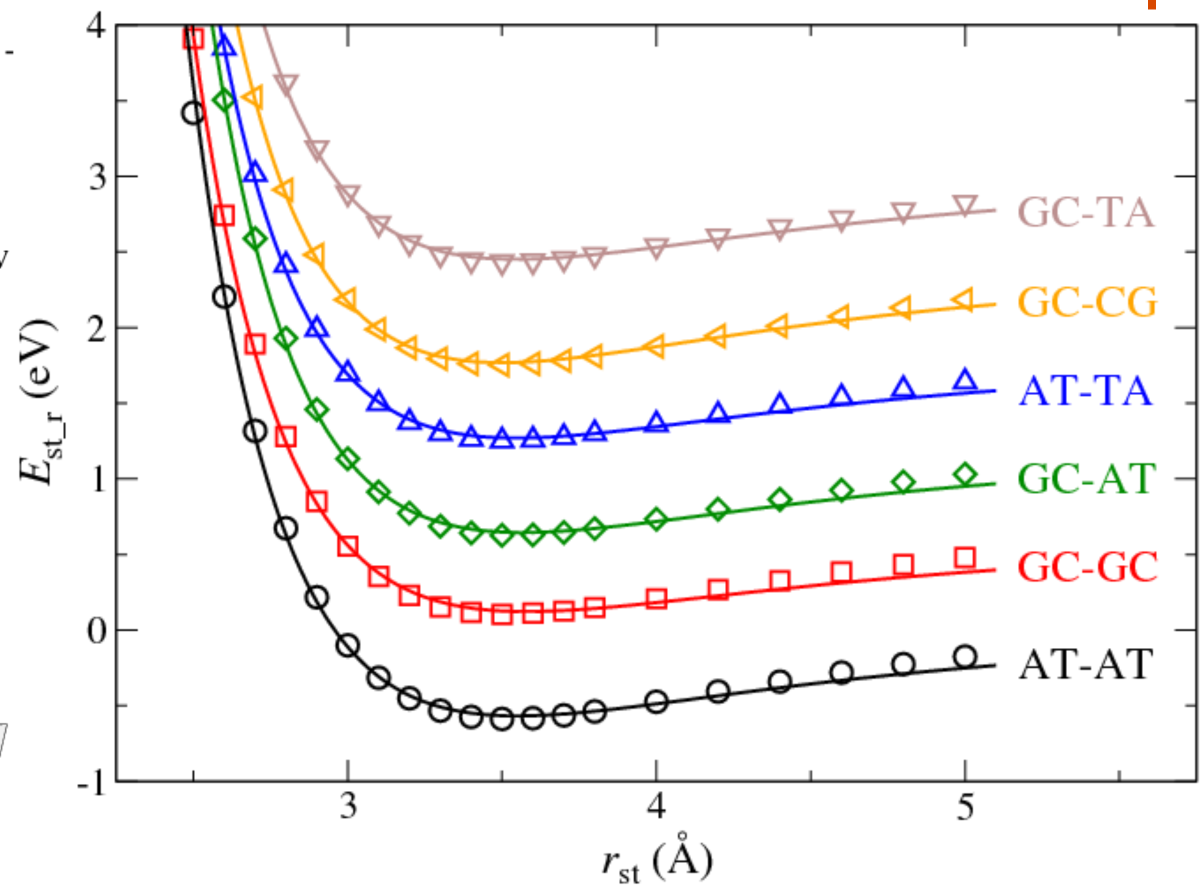
Flip angle



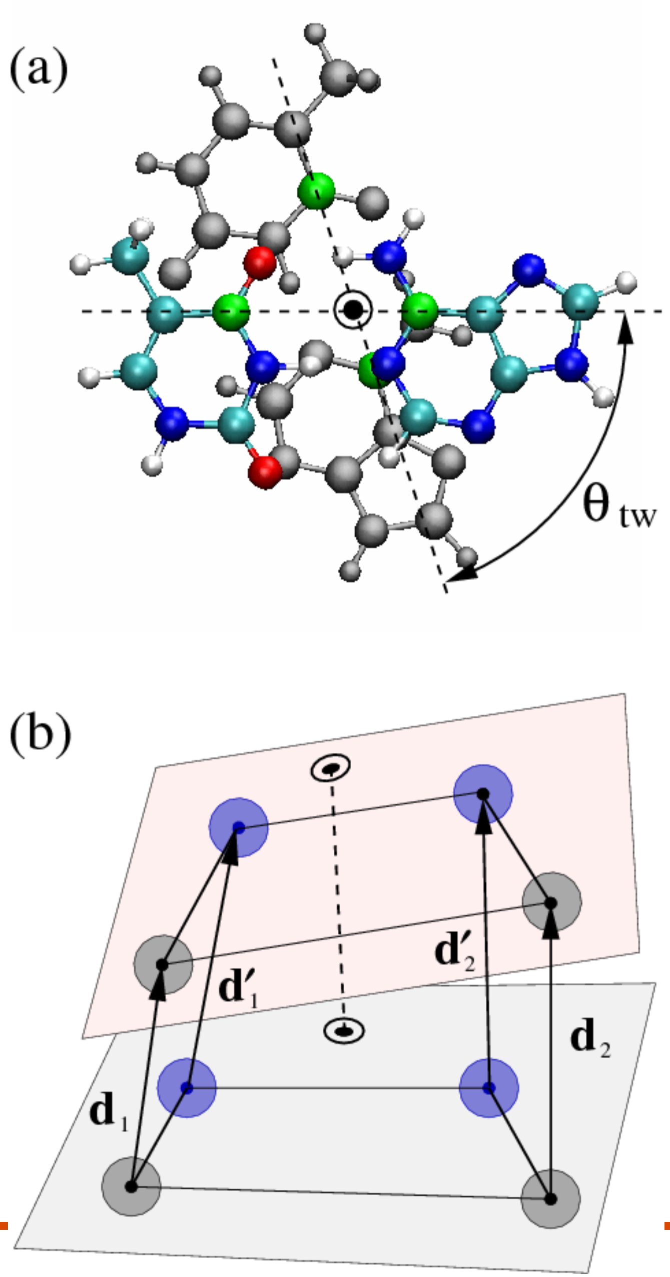
Coarse-grained potential – stacking interactions



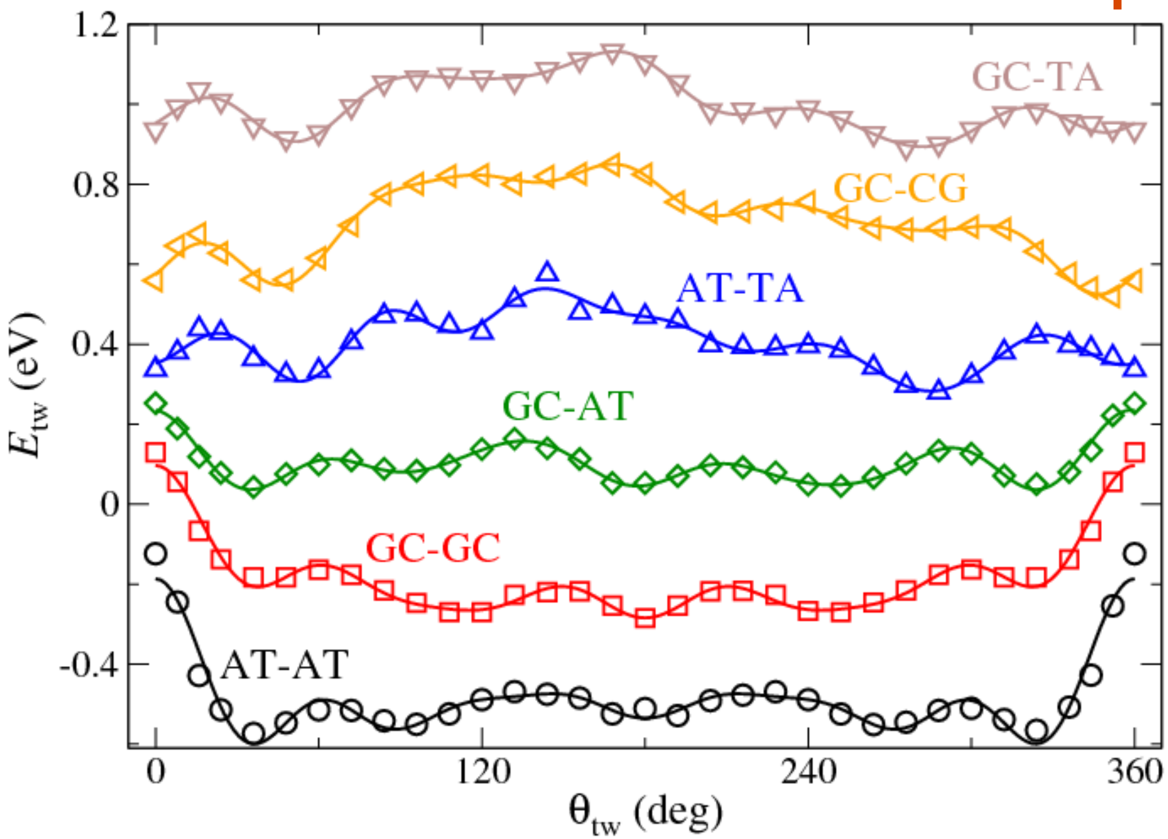
Distance between planes



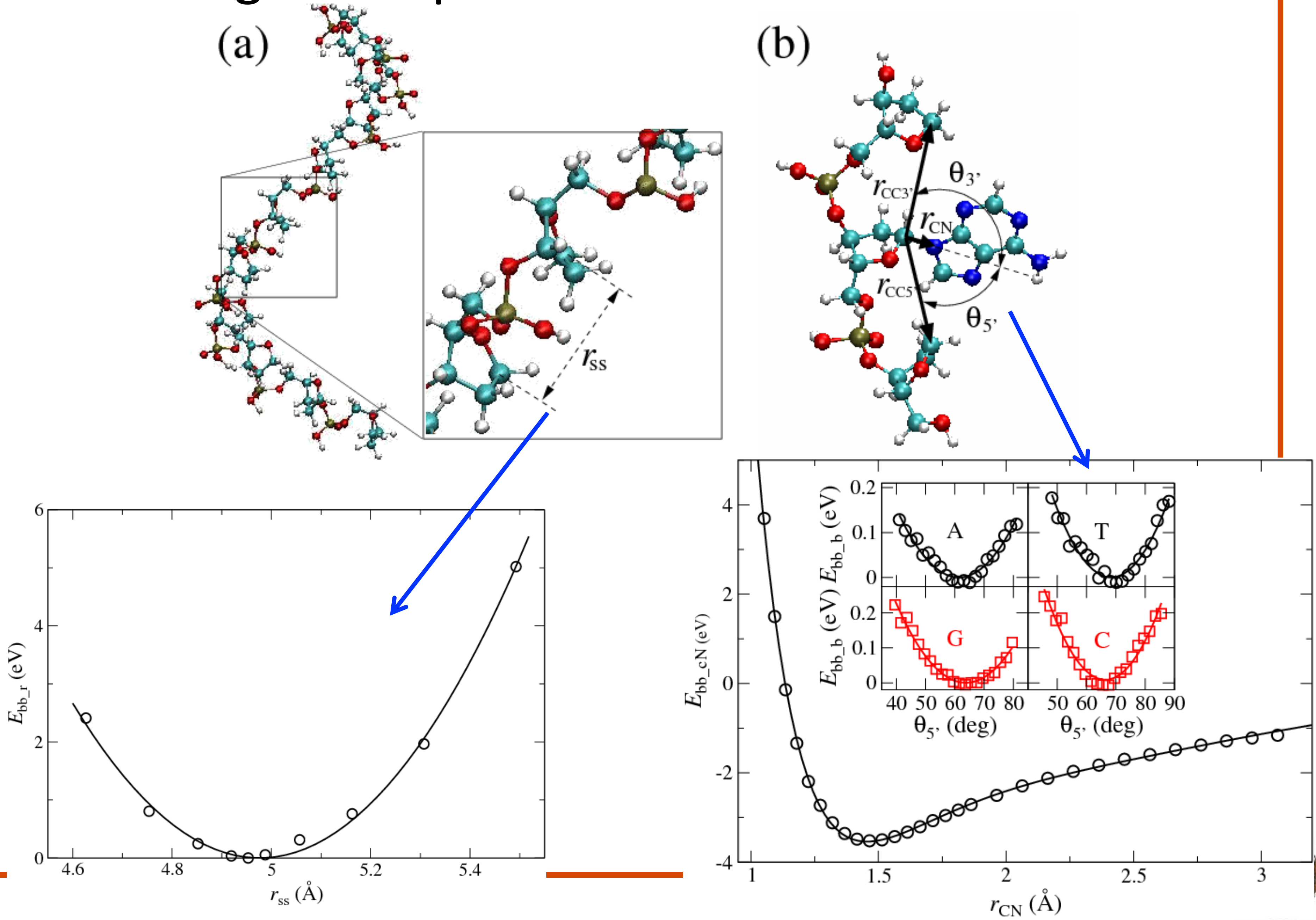
Coarse-grained potential – stacking interactions



Twist angle between pairs

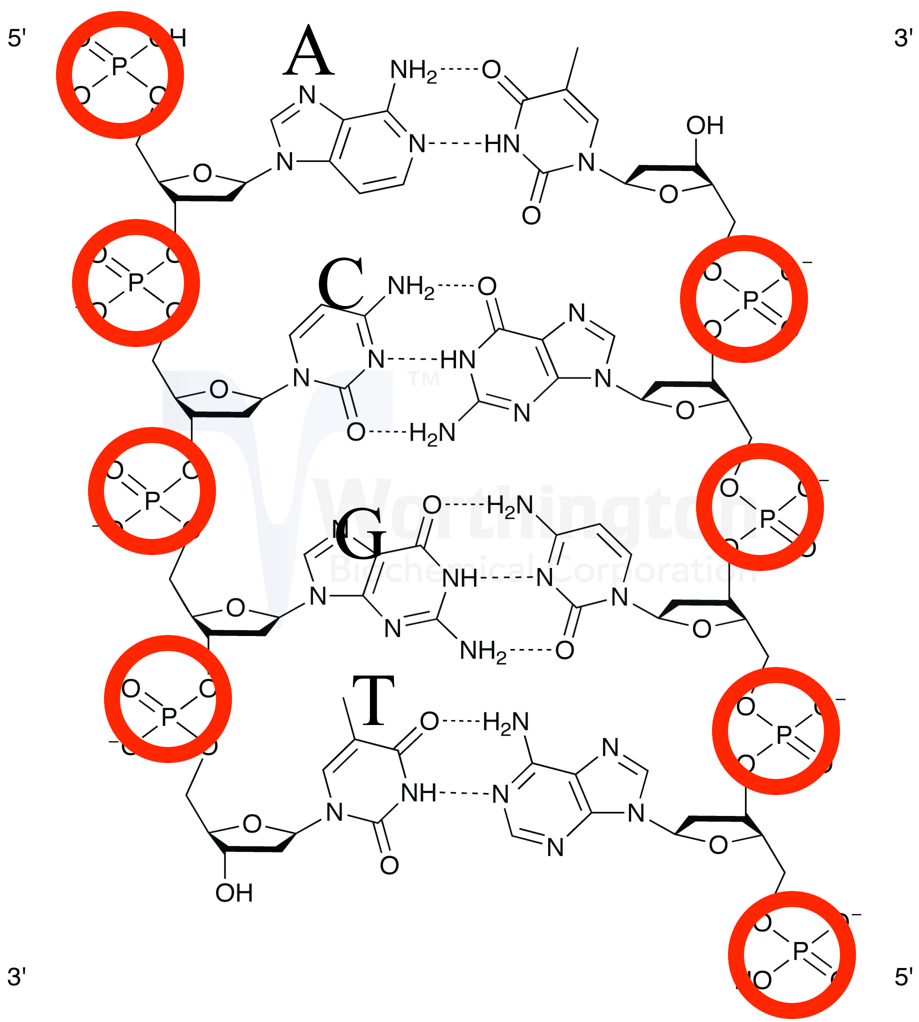


Coarse-grained potential – backbone interactions



Coarse-grained potential – electrostatic interactions

Deoxyribonucleic Acid



5'-Adenine-Cytosine-Guanine-Thymine-3'

$$E = \frac{1}{4\pi\varepsilon_0\varepsilon(r)}\frac{e^2}{r}$$

$$\varepsilon(r) = \varepsilon_{\text{in}} \qquad r < r_0$$

$$\varepsilon(r) = \varepsilon_{\text{in}} e^{\alpha(r-r_0)} \qquad r_0 < r < r_1$$

$$\varepsilon(r) = \varepsilon_{\infty} e^{-\kappa r} \qquad r > r_1$$

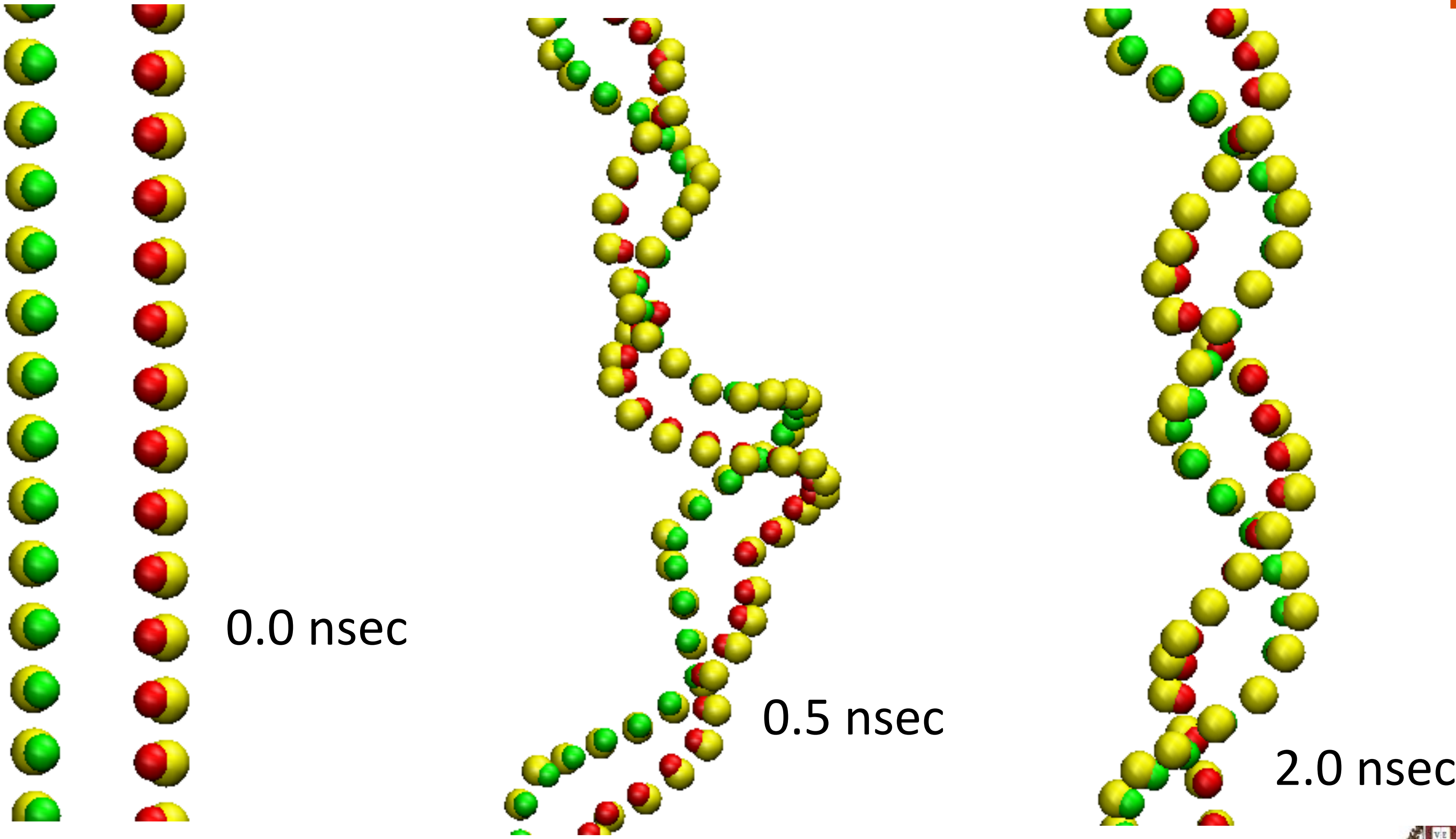
$$\kappa^{-1} = \sqrt{\frac{\varepsilon_0\varepsilon_\infty k_B T}{2N_A e^2 I}}$$

3 parameters !

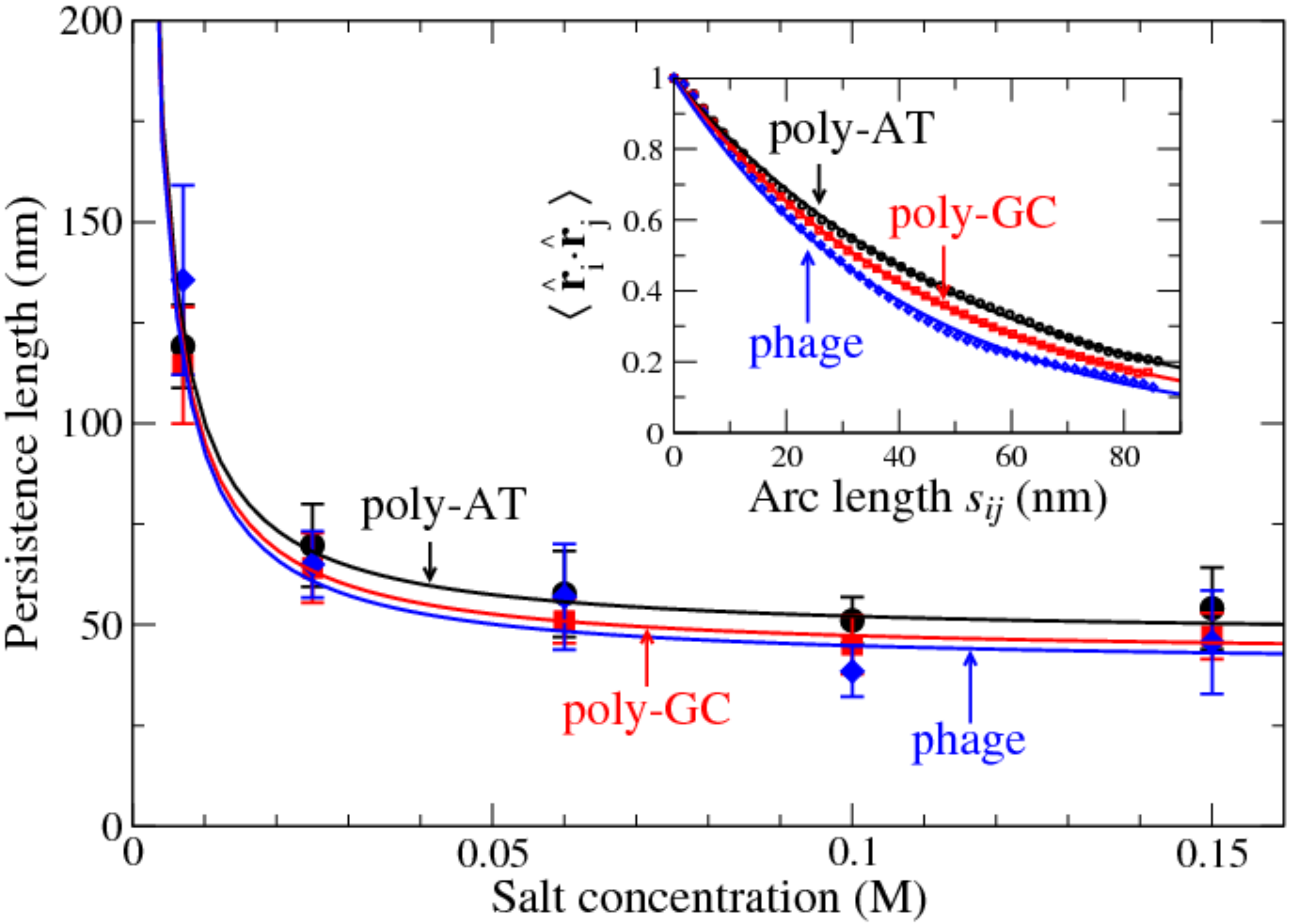
r_0, α, ε_{in}

Coarse-grained potential – validation

Two parallel strands
coil to form double-helix

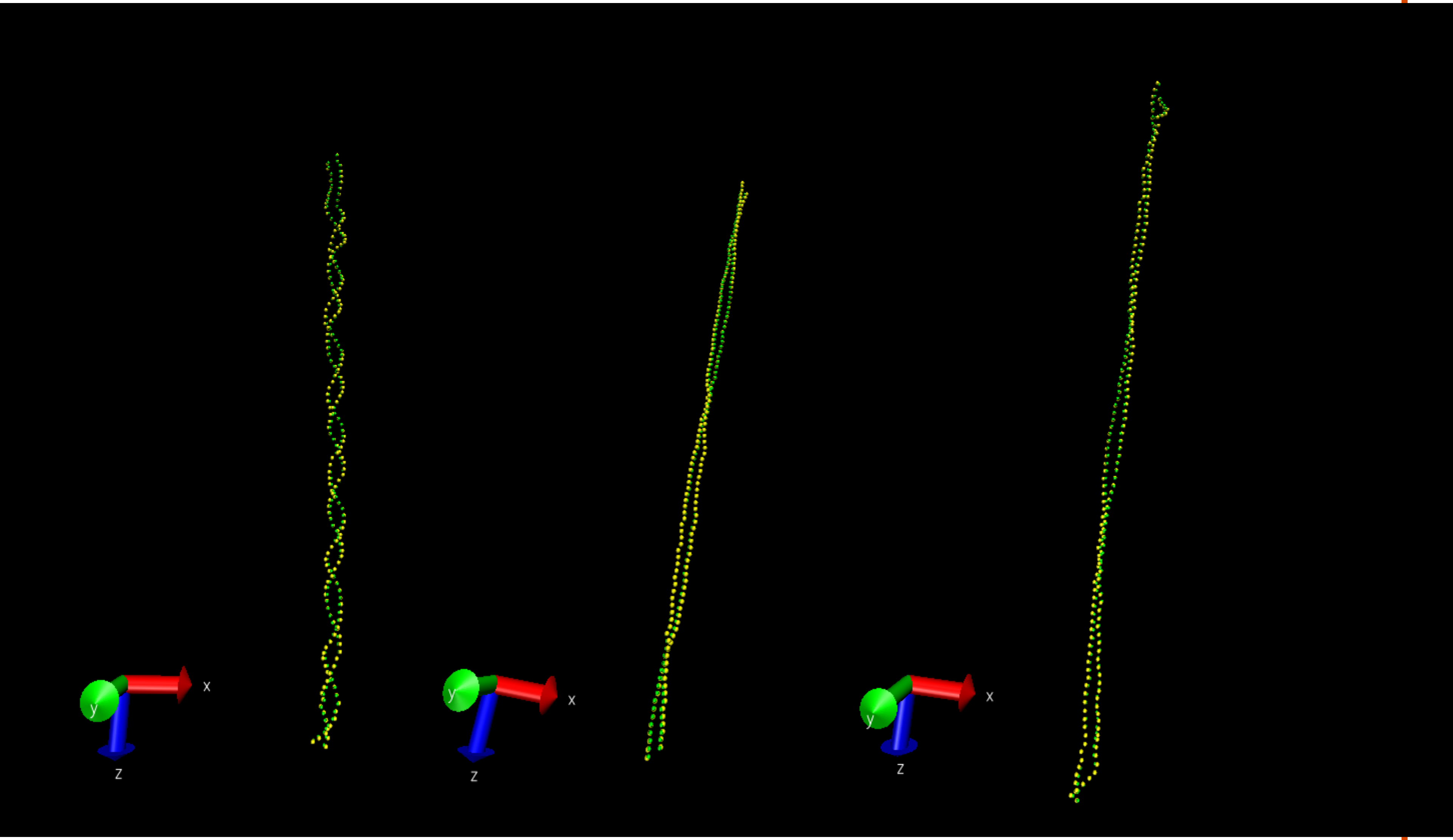


Coarse-grained potential – validation



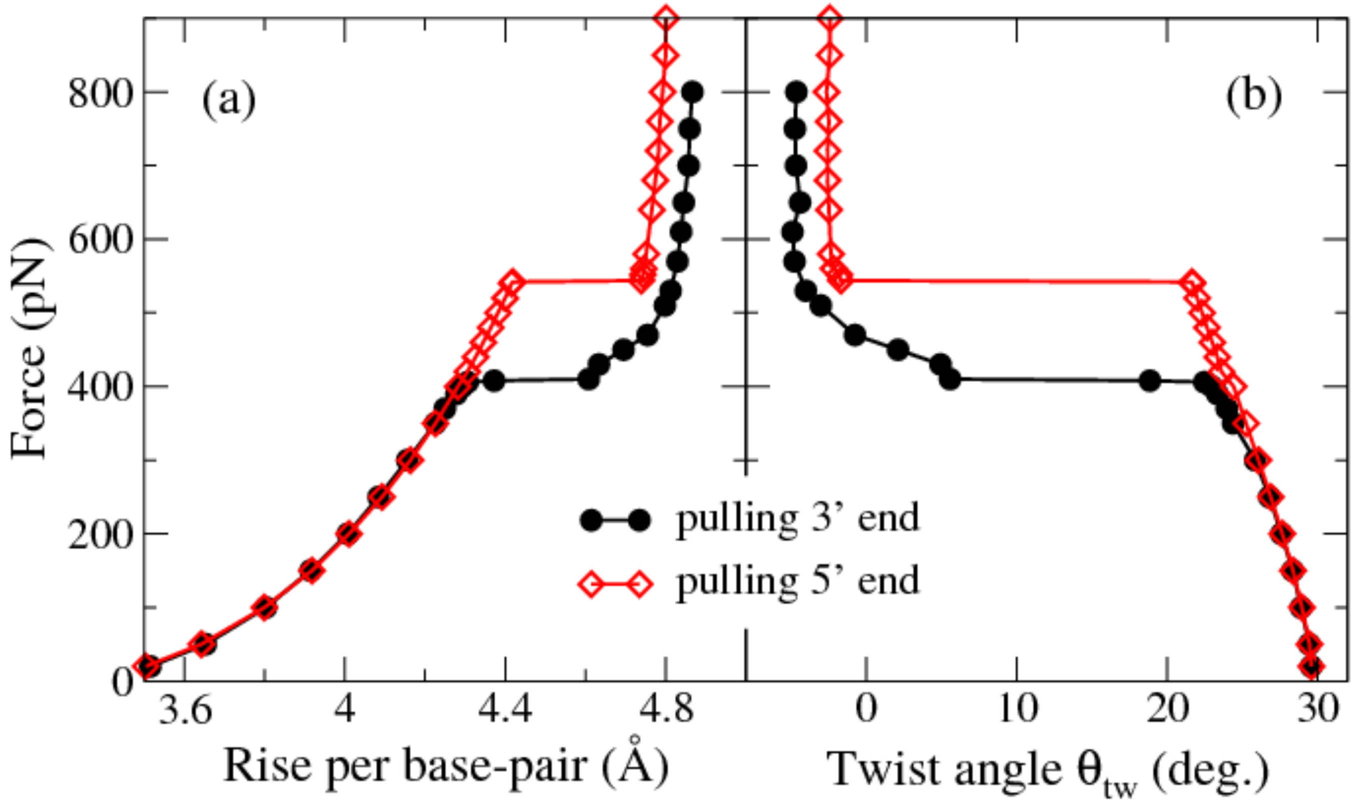
Persistence length (~50 nm)

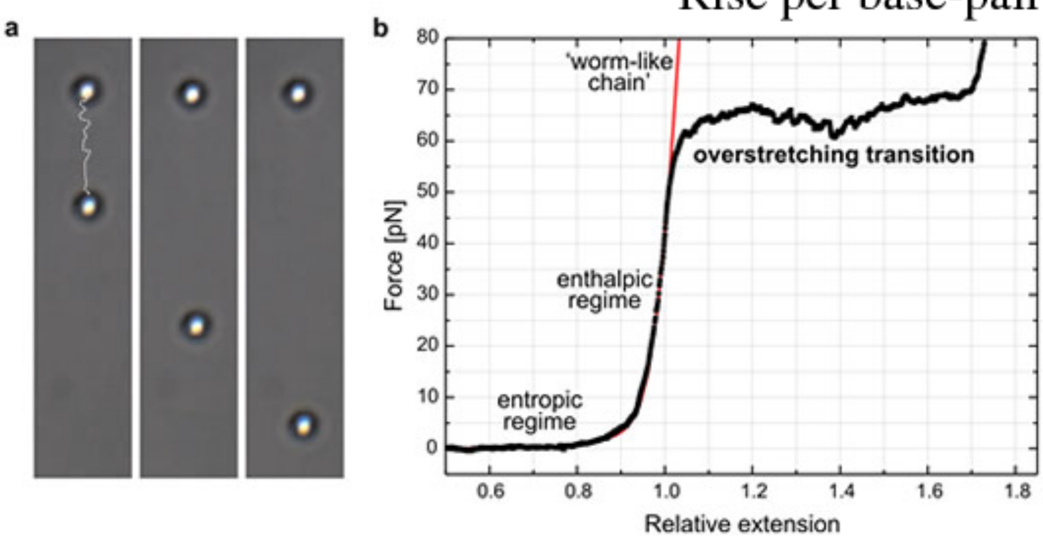
Force-extension simulations



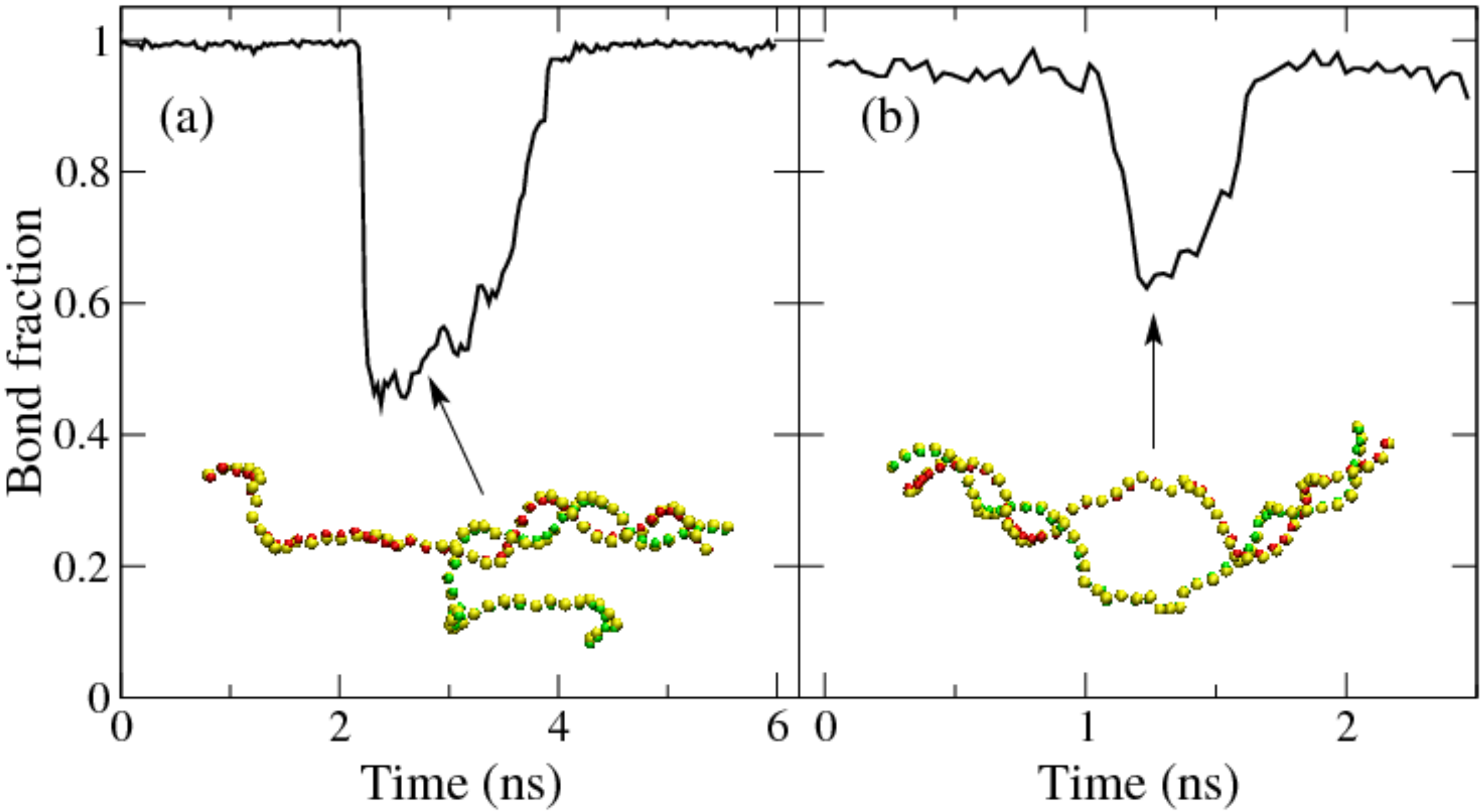
3’-3’ stretch 5’-5’ stretch

Force-extension simulations





Melting simulations



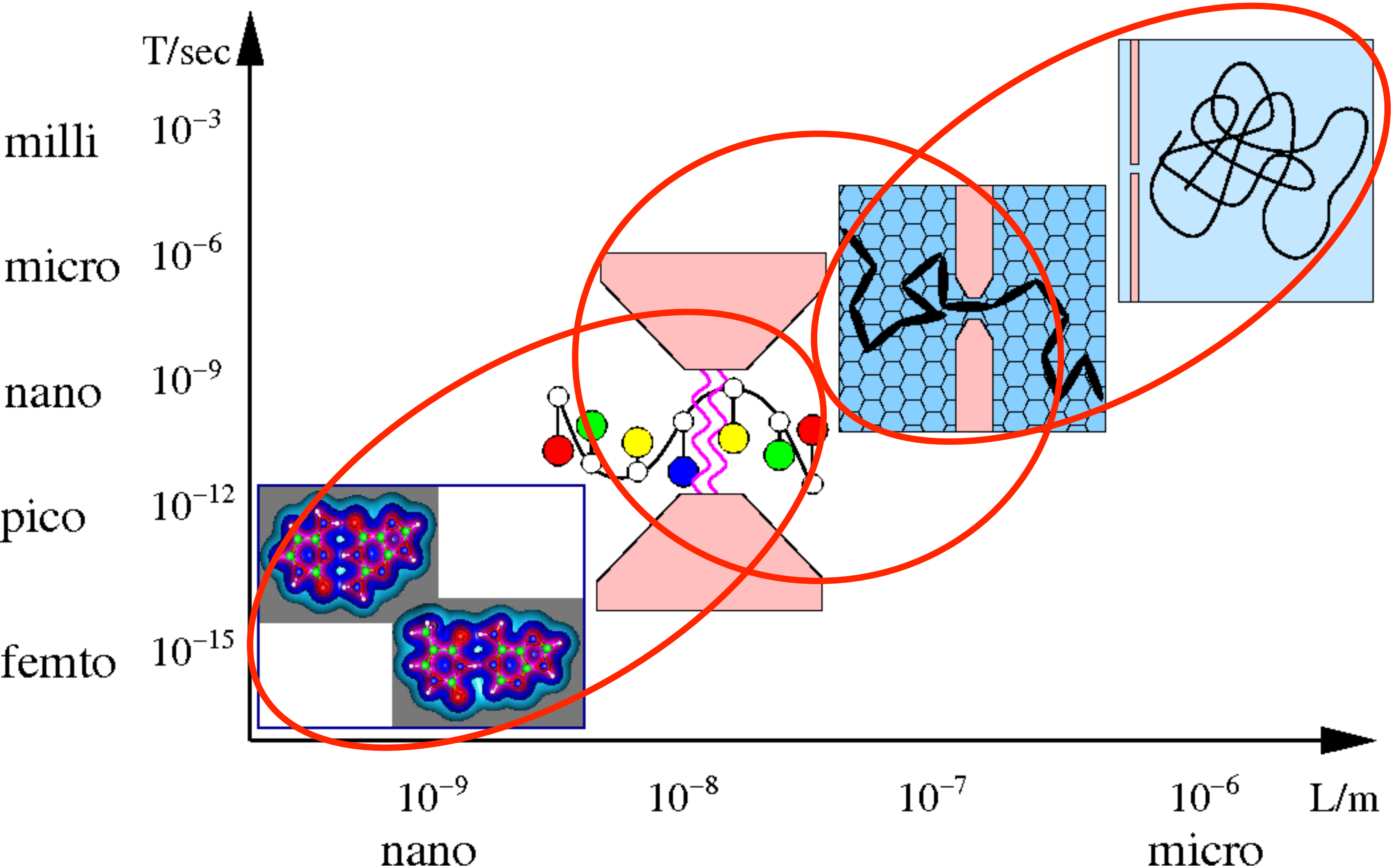
Major open issues:

- Systematic method for coarse graining DOF (normal mode analysis?)
- Temperature?
- Entropy?

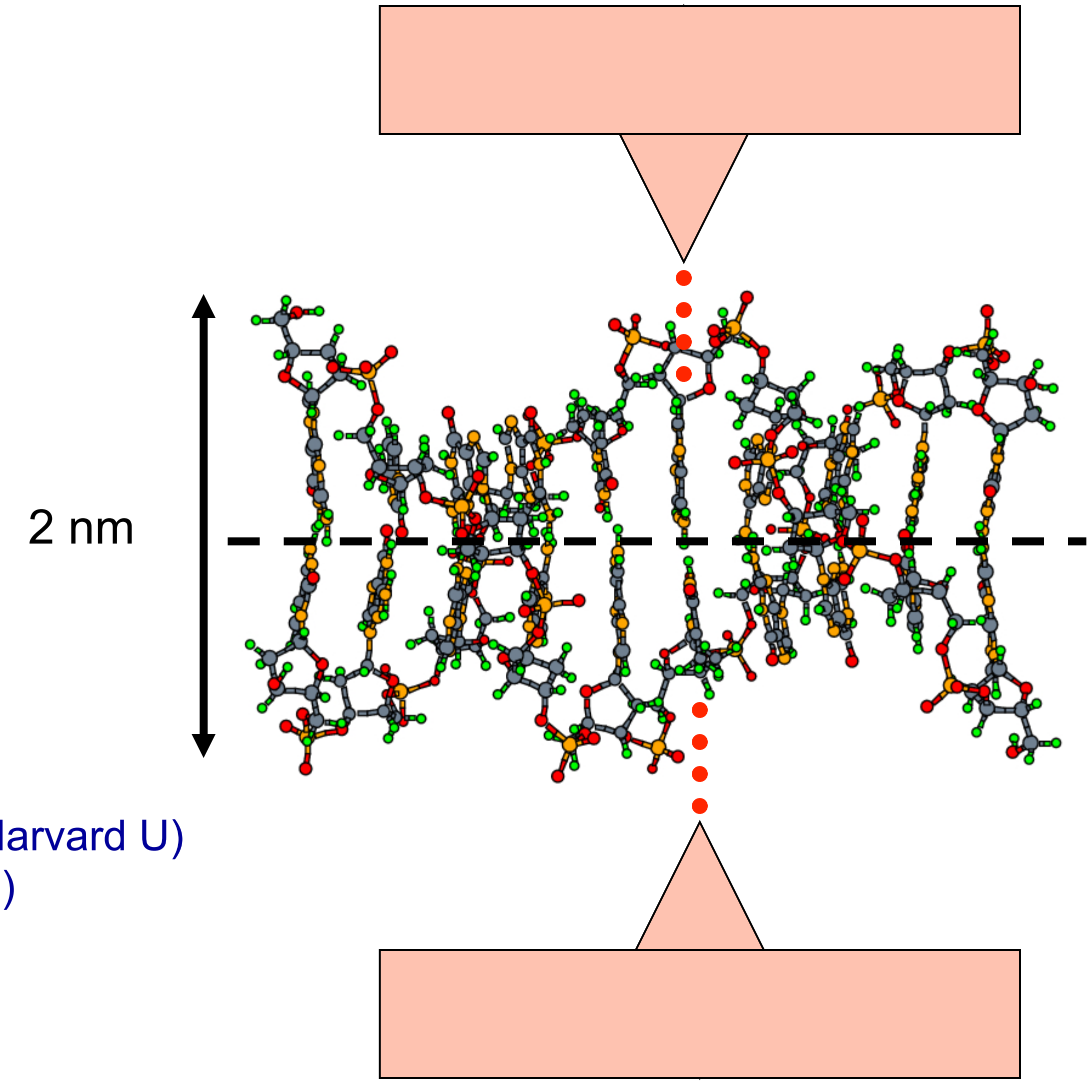
V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, and K. Kremer, *Macromolecules* **2006,** *39,* 6708-6719

D. Fritz *et al.,* Phys. Chem. Chem. Phys., 2011, **13**, 10412–10420

Multiple length/time scales: DNA electronic sequencing



Electronic sequencing of DNA



Motivation: ultrafast sequencing through electronic signals:
J. Golovchenko *et al.* (Harvard U)
C. Decker *et al.* (Delft U)

The New York Times

October 6, 2009

I.B.M. Joins Pursuit of $1,000 Personal Genome

By JOHN MARKOFF

One of the oldest names in computing is joining the race to sequence the genome for $1,000. On Tuesday, I.B.M. plans to give technical details of its effort to reach and surpass that goal, ultimately bringing the cost to as low as $100, making a personal genome cheaper than a ticket to a Broadway play.

The project places I.B.M. squarely in the middle of an international race to drive down the cost of gene sequencing to help move toward an era of personalized medicine. The hope is that tailored genomic medicine would offer significant improvements in diagnosis and treatment.

I.B.M. already has a wide range of scientific and commercial efforts in fields like manufacturing supercomputers designed specifically for modeling biological processes. The company's researchers and executives hope to use its expertise in semiconductor manufacturing, computing and material science to design an integrated sequencing machine that will offer advances both in accuracy and speed, and will lower the cost.

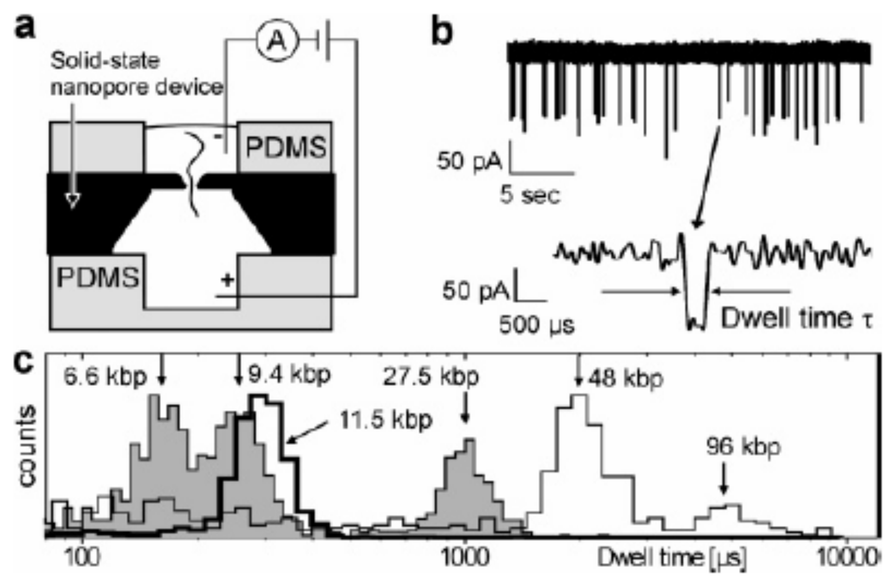
"More and more of biology is becoming an information science, which is very much a business for I.B.M.," said Ajay Royyuru, senior manager for I.B.M.'s computational biology center at its Thomas J. Watson Laboratory in Yorktown Heights, N.Y.

DNA sequencing began at academic research centers in the 1970s, and the original Human Genome Project successfully sequenced the first genome in 2001 and cost roughly $1 billion.

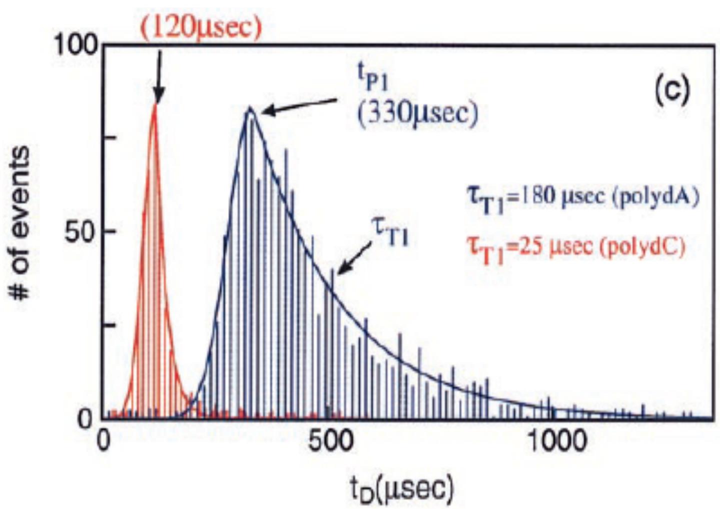
Since then, the field has accelerated. In the last four to five years, the cost of sequencing has been falling at a rate of tenfold annually, according to George M. Church, a Harvard geneticist. In a recent presentation in Los Angeles, Dr. Church said he expected the industry to stay on that curve, or some fraction of that improvement rate, for the foreseeable future.

At least 17 startup and existing companies are in the sequencing race, pursuing a range of third-generation technologies. Sequencing the human genome now costs $5,000 to $50,000, although Dr. Church emphasized that none of the efforts so far had been completely successful and no research group had yet

Double-stranded DNA (6 - 96 kbp)
passing through solid pore
forced by electric field at pore



Single-stranded DNA (100 bp)
passing through α-hemolysin
forced by ionic current



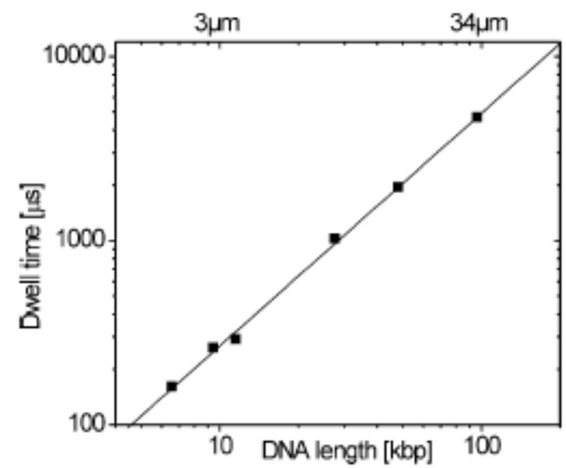


Figure 2. Dwell time versus DNA length. The line shows the result of a power-law fit to the data, with a best-fit exponent of $\alpha = 1.27 \pm 0.03$.

Mellor et al., ***PNAS*** **(2000)**

Storm et al., ***NanoLetters*** **(2005)**

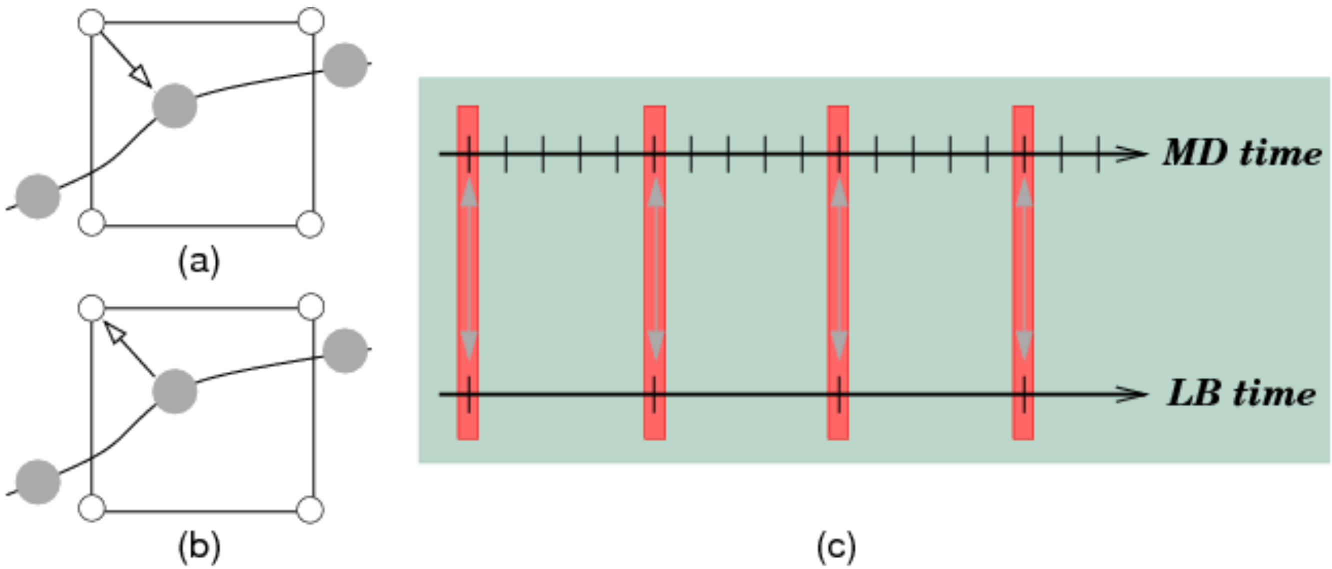
Two-scale approach to DNA translocation

– ***Molecular Dynamics*** for DNA:

course-grained molecules (~30 bp/bead)

– ***Lattice Boltzmann Equation*** for the solvent:

- advantages in describing arbitrary shapes
- fluid dynamics in particle language



spatial temporal

Lattice-Boltzmann Method (LBM)

Lattice Boltzmann Equation : discrete distribution functions $f_i(x,t)$

$f_i(\mathbf{x},t)$, i=1,n: probability to find a particle at lattice site **x** at time t with speed $\mathbf{c}_i$

$$f_p(\vec{x}+\vec{c}_p\Delta t, t+\Delta t) = f_p(\vec{x},t) - \omega\Delta t(f_p\text{-}f_p^{eq})(\vec{x},t) + G_p\Delta t$$

polymer-fluid back reaction

local equilibrium :

$$f_p^{eq} = w_p\left[\frac{1}{kT}\vec{u}\cdot\vec{c}_p + \frac{1}{2kT}(\vec{u}\,\vec{u}\cdot(\vec{c}_p\vec{c}_p\text{-}kT\overleftrightarrow{I}))\right]$$

$$\rho(\vec{x},t) = \sum_p f_p(\vec{x},t)$$ density

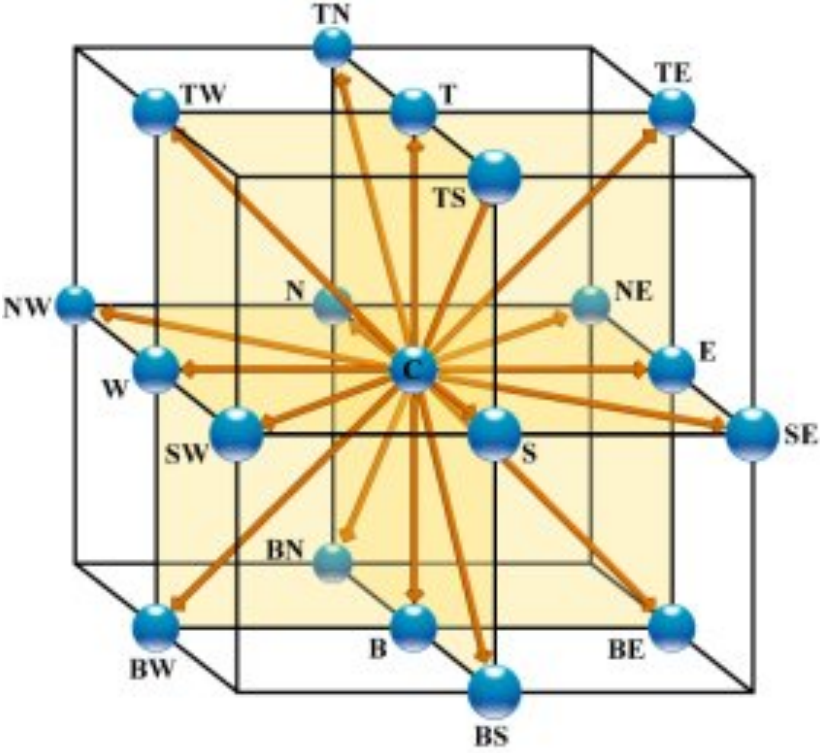
$$\rho\vec{u}(\vec{x},t) = \sum_p f_p(\vec{x},t)\vec{c}_p$$ flow speed

$$\overleftrightarrow{P}(\vec{x},t) = \sum_p f_p(\vec{x},t)\vec{c}_p\,\vec{c}_p$$ momentum flux

Lattice-Boltzmann Method (LBM)

Fluid particles move only along trajectories prescribed by the lattice directions

(in 3D: 19-speed lattice)



Molecular (Langevin) Dynamics (MD)

DNA with N beads at positions r_p with velocities υ_p :

$$m\frac{d\vec{\upsilon}_i}{dt} = \vec{F}_i^{\,c} + \vec{F}_i^{\,f} + \vec{F}_i^{\,r} + \lambda_i \partial_{\vec{r}_i} \sigma$$

bead-bead interactions ($\vec{F}_i^{\,c}$)

Solute-solvent interactions ($\vec{F}_i^{\,f}$)

random force ($\vec{F}_i^{\,r}$)

constraint force ($\lambda_i \partial_{\vec{r}_i} \sigma$)

$$\sigma = |\vec{r}_{i+1} - \vec{r}_i|^2 - r_0^2 = 0$$

implemented by SHAKE algorithm

Coupling LB to MD:

$$\vec{F}_i^{\,f} = -m\gamma(\vec{u}_i - \vec{\upsilon}_i)$$

υ : bead velocity

u : fluid velocity

Coupling MD to LB:

$$G_p(\vec{r},t) = w_p\,\beta \sum_{i \in D(\vec{r})} [\vec{F}_i^{\,f} + \vec{F}_i^{\,r}] \cdot \vec{c}_i$$

M. Fyta, S. Melchionna, E. Kaxiras, S. Succi, *Multiscale Model. Sim.* (2006)

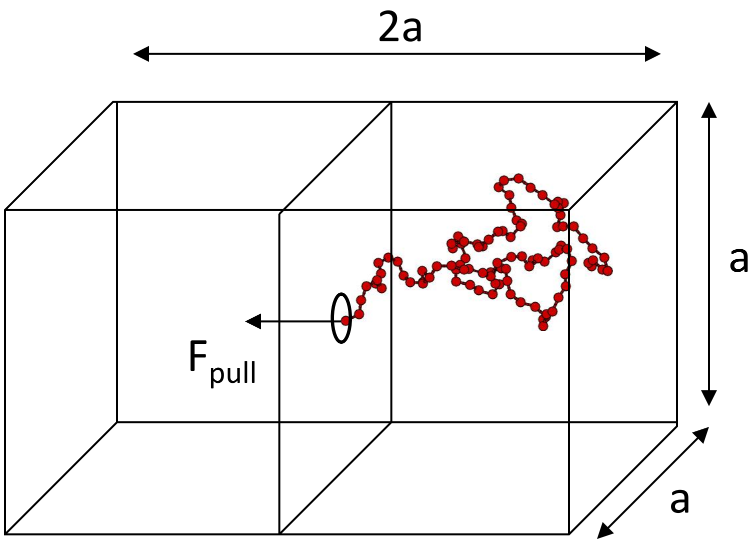
Details of simulation

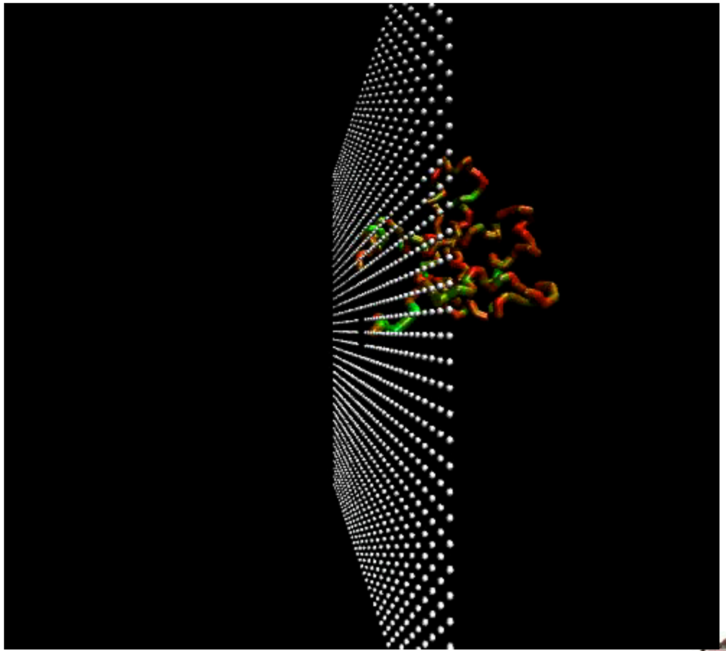
- 3D box of (2a x a x a) size
- hole size = 6 nm
- lattice spacing Δx = 3 nm

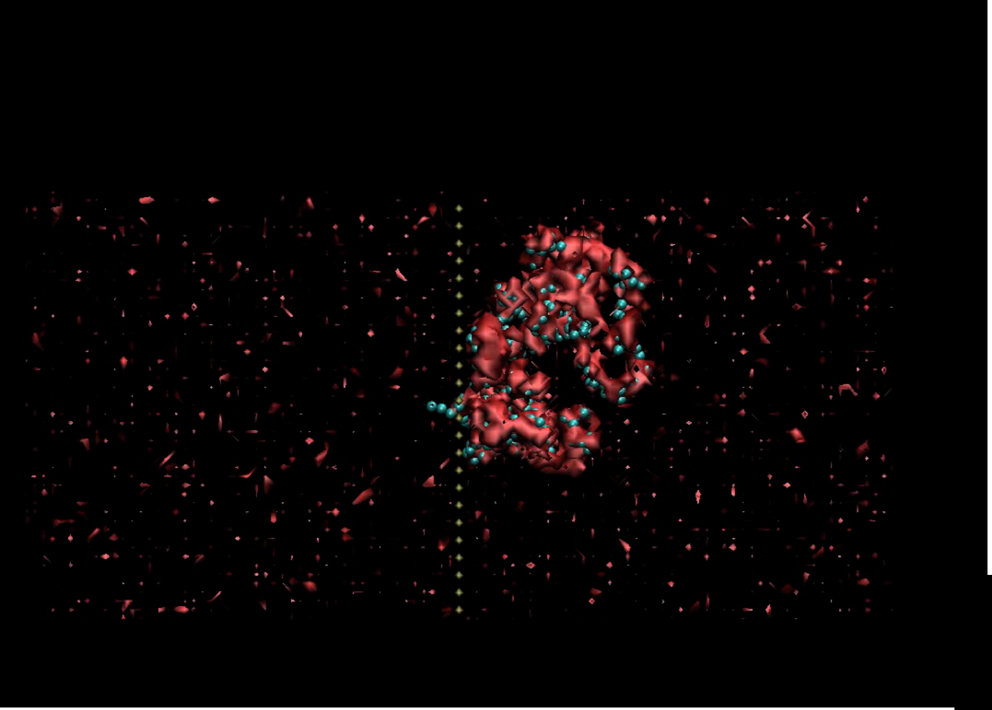
- F_{pull} =0.02, kT =10^{-4}
- Fast translocation regime :

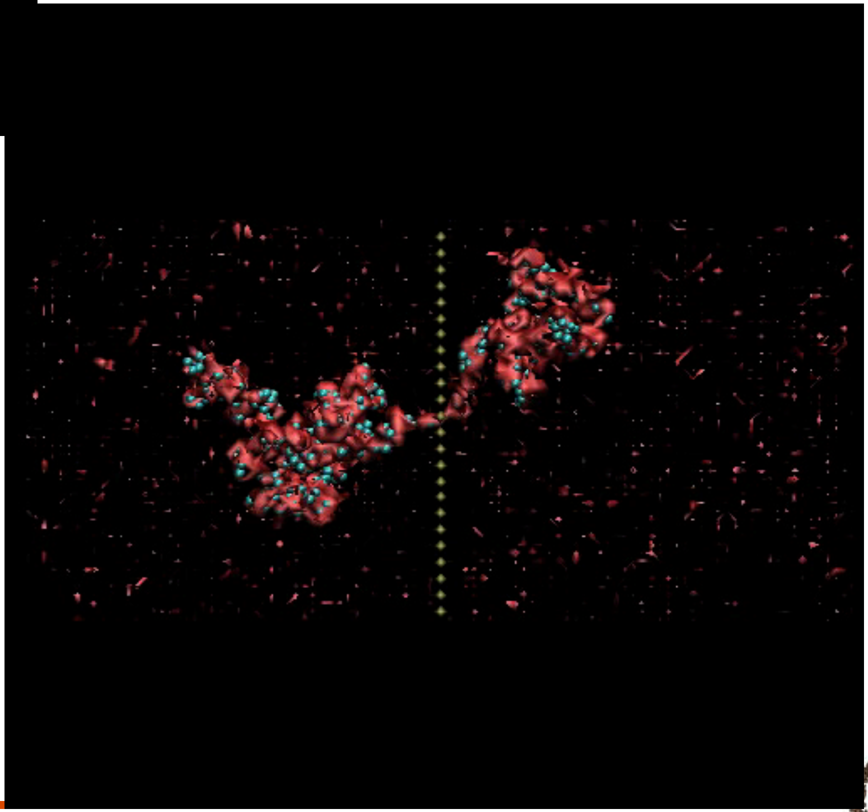
[translocation time << DNA relaxation time]

$$\frac{F_{pull}R}{kT} >> 50$$

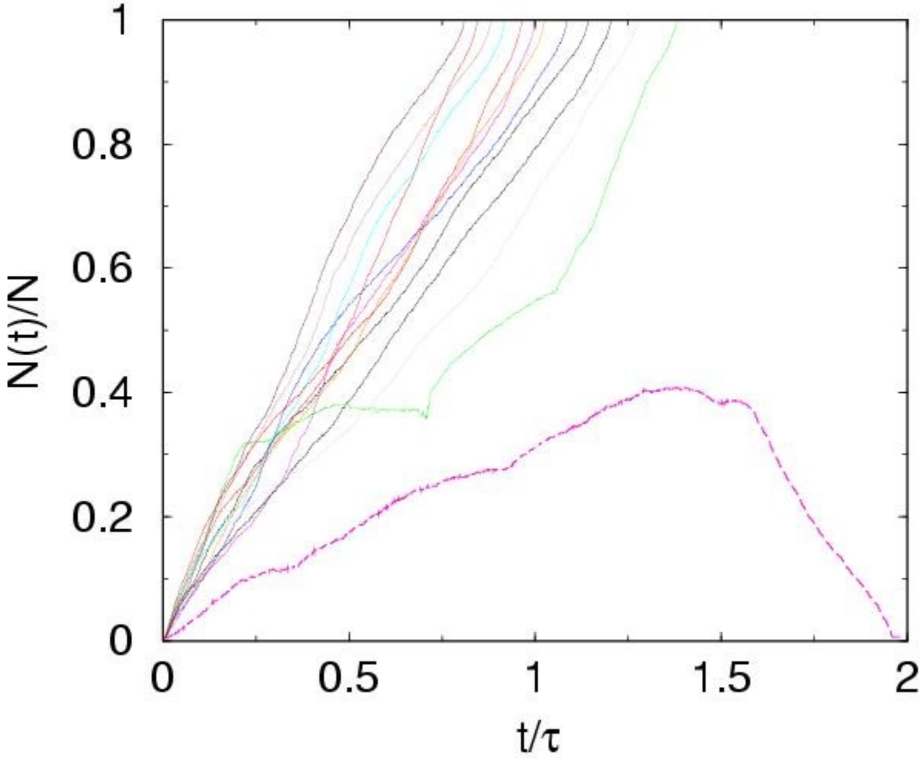


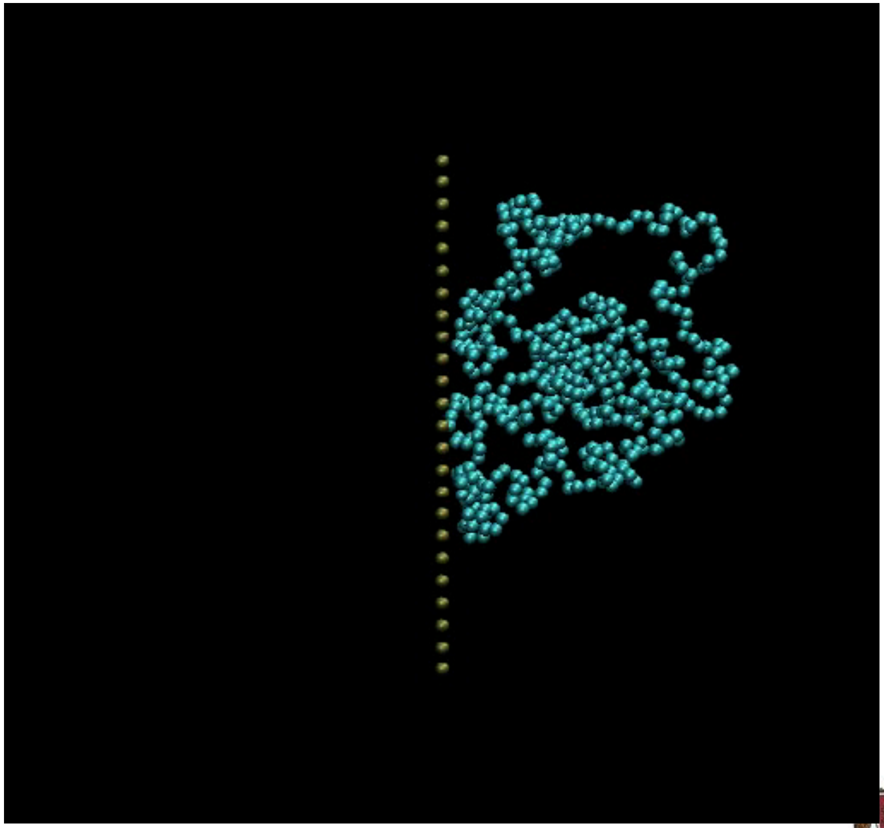


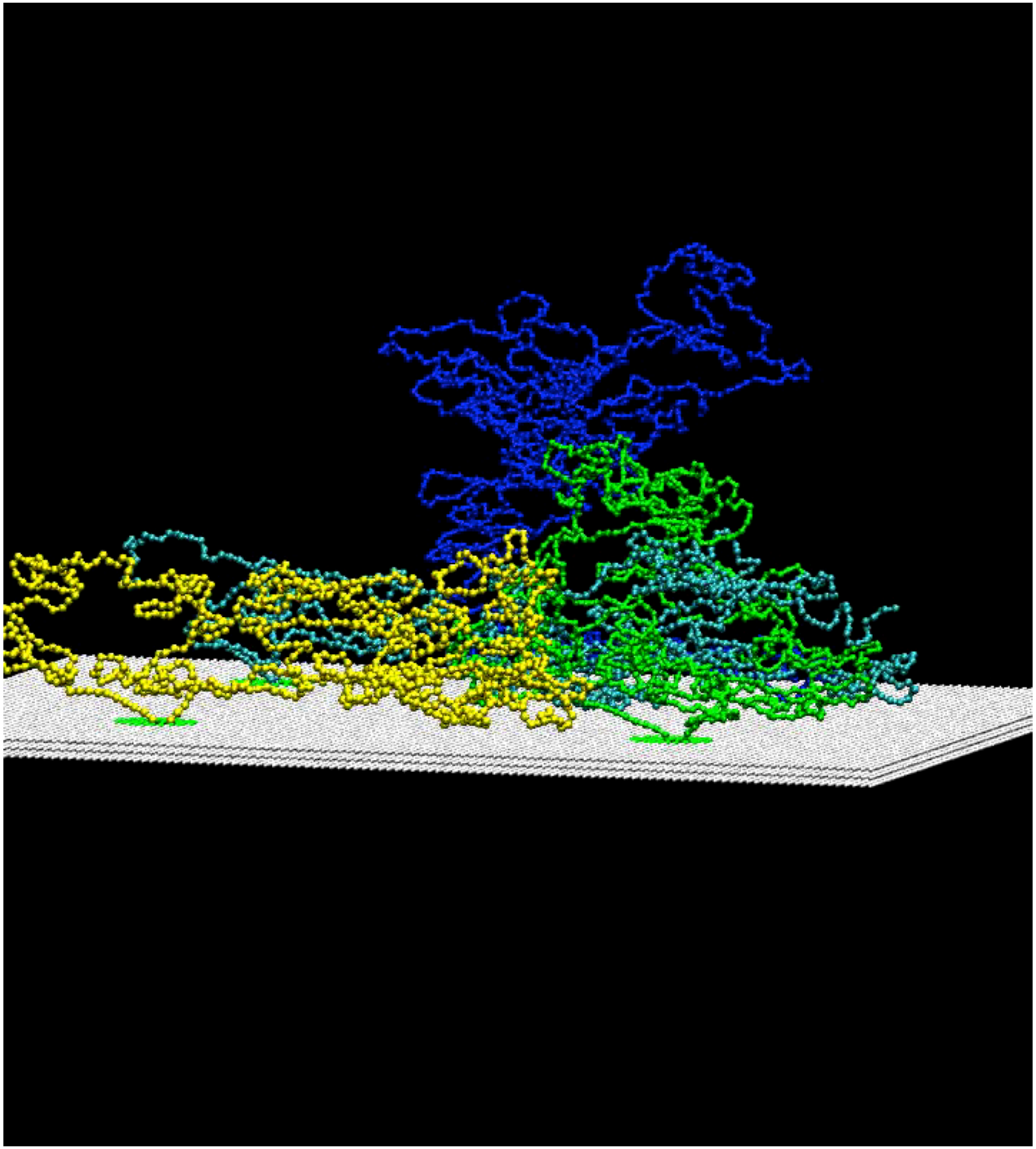




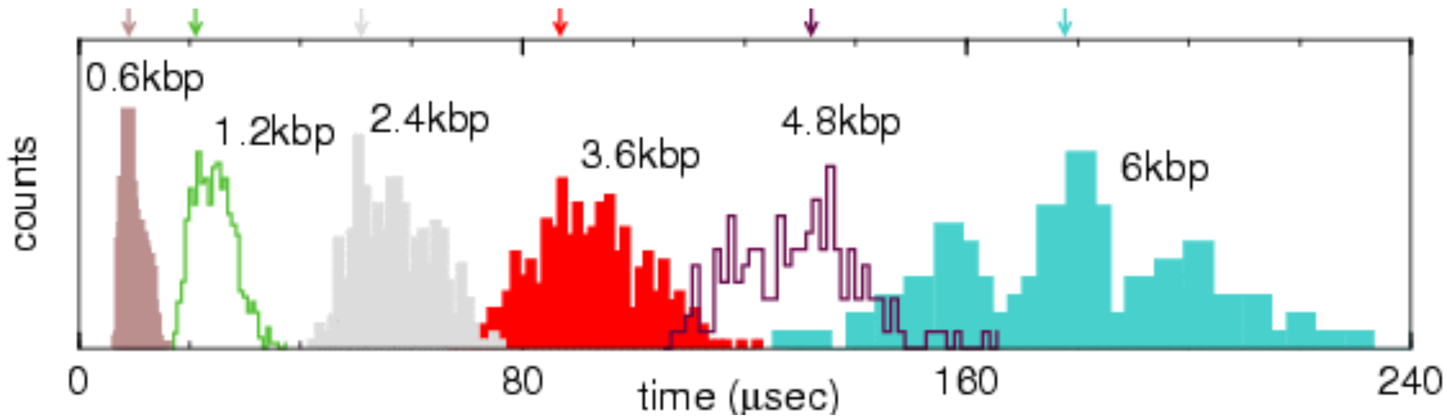
Rare events: retraction

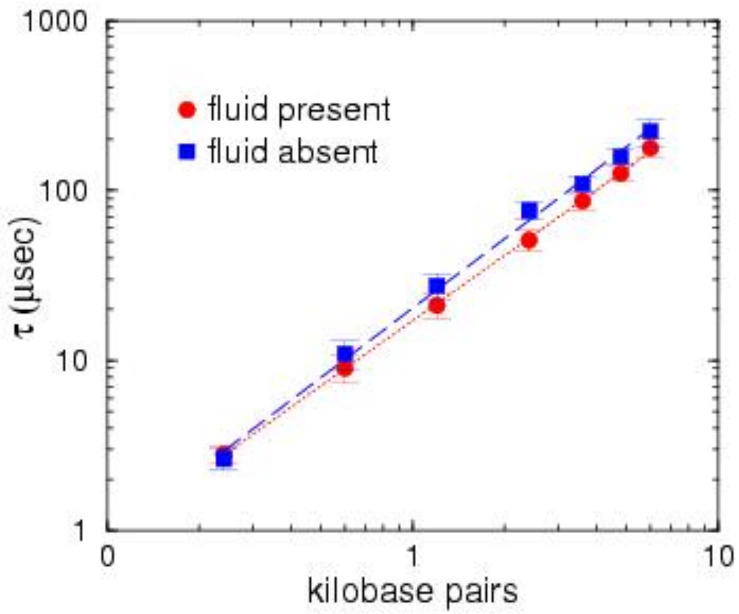






Translocation time – Statistics and scaling





Theory: $\tau \sim N^{1.28\pm0.01}$ with fluid

$\tau \sim N^{1.36\pm0.03}$ without fluid

Experiment: $\tau \sim N^{1.27\pm0.03}$

Phenomenological theory: radius of gyration

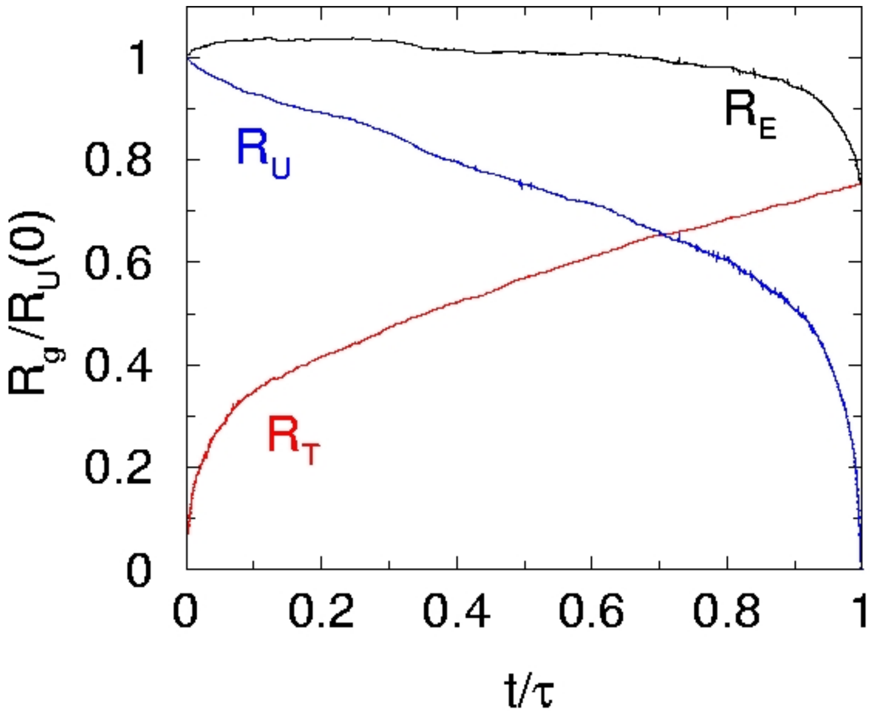
$$R = \sqrt{\sum_i |\vec{r}_i|^2}$$

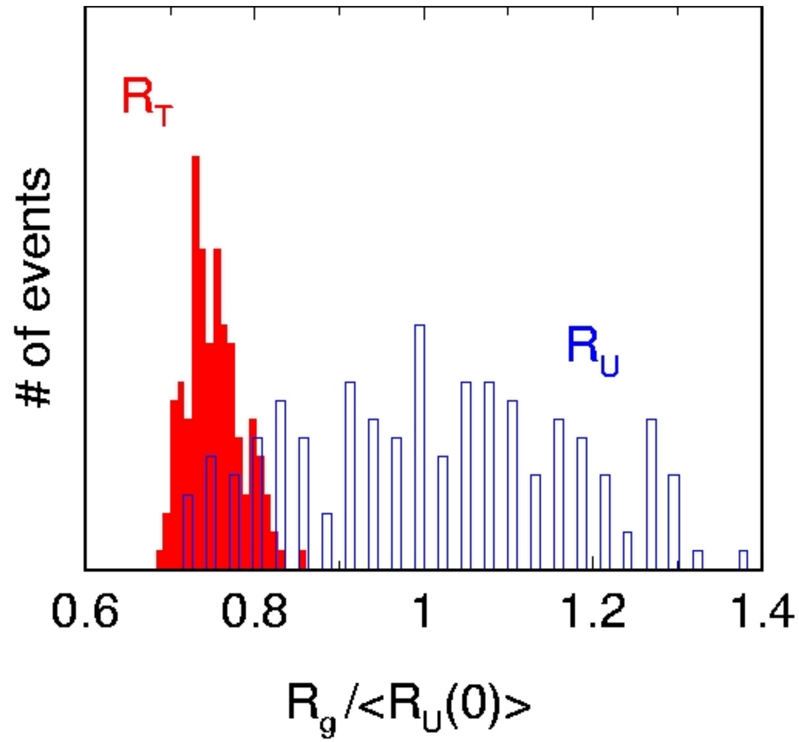
$$R_E(t) = [R_T(t)^{1/\nu} + R_T(t)^{1/\nu}]^\nu$$

$$R(t) \sim [N(t)]^\nu$$

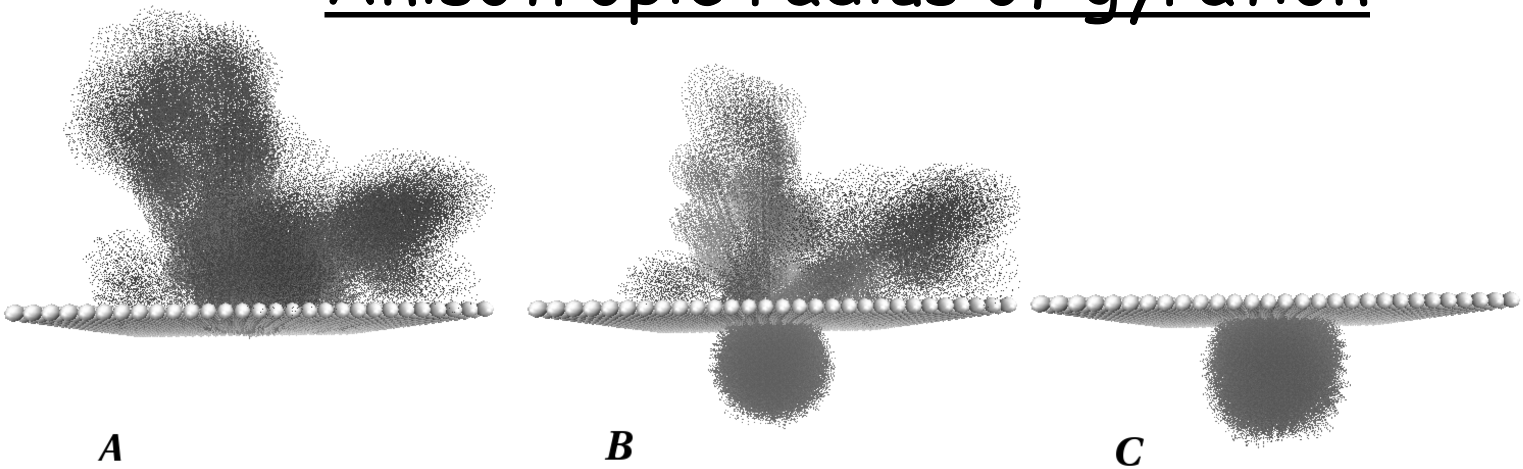
translocated (T) part
untranslocated (U) part

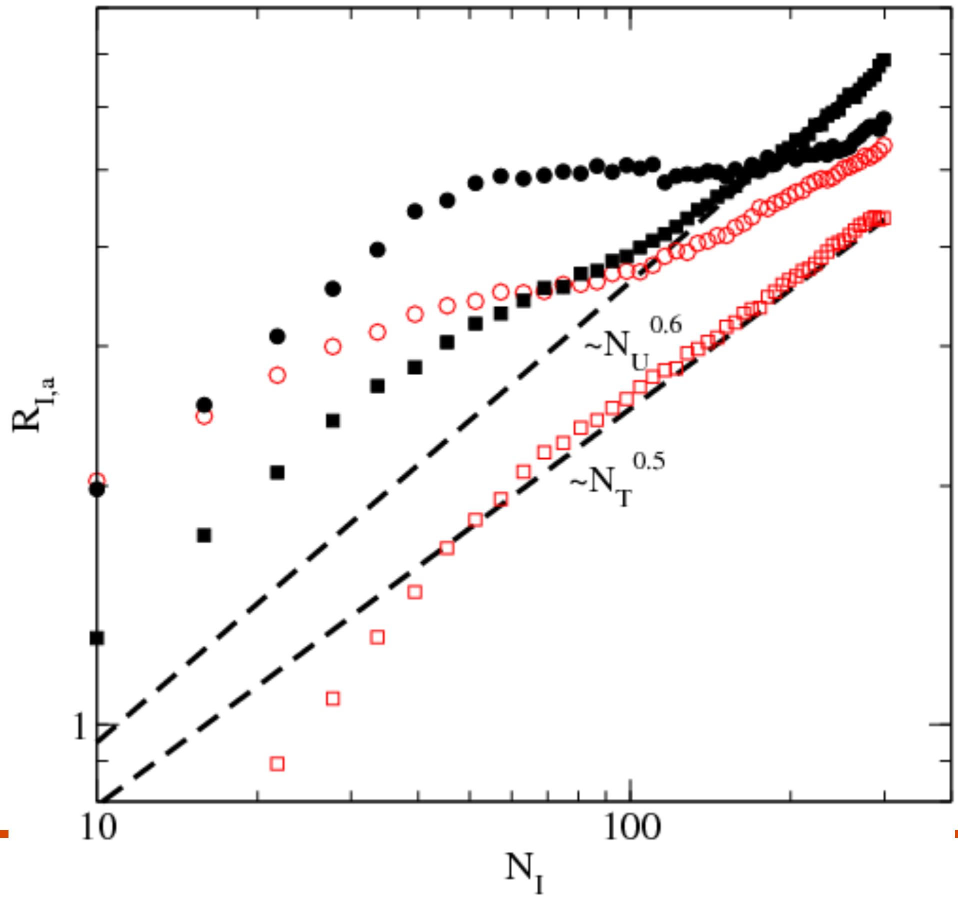
(SAW : $\nu = 0.6$)





Anisotropic radius of gyration



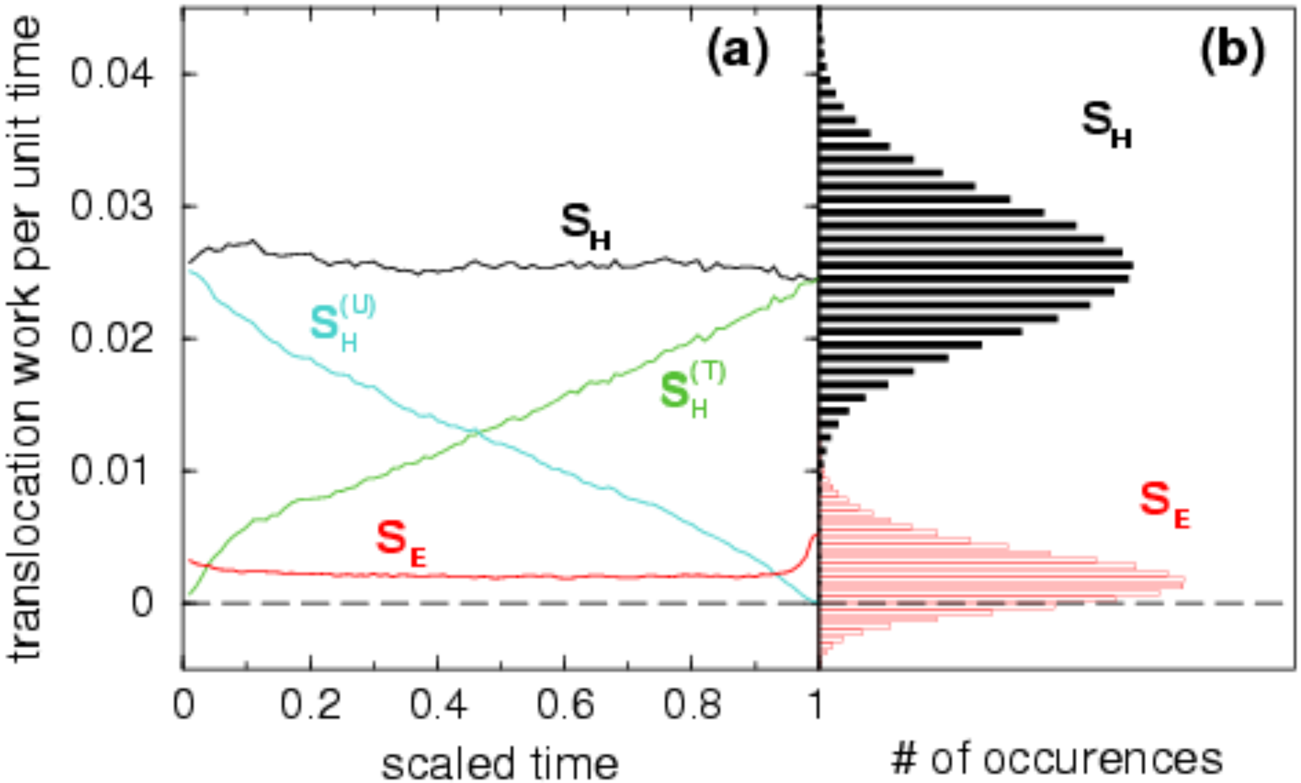


squares: transverse (*t*)
circles: longitudinal (*l*)
black: untranslocated (*U*)
red: translocated (*T*)

Untranslocated-transverse dominates
($\nu_U = 0.6$)

$$S_H(t) = \frac{dW_H}{dt} = \gamma \left\langle \sum_i^N \vec{\upsilon}_i(t) \cdot \vec{u}_i(t) \right\rangle$$ Hydrodynamic interactions

$$S_E(t) = \frac{dW_E}{dt} = \left\langle \sum_i^N \vec{F}_{drive,i}(t) \cdot \vec{\upsilon}_i(t) \right\rangle$$ External driving field



Phenomenological theory: rate of work

From Eq.s of Motion: $$\frac{dW}{dt} = S_H(t) - 2\gamma K(t) + S_E(t)$$ **: const!**

$$dW = P\,dV + \sigma_\gamma\,dA \qquad P = \frac{2\sigma_\gamma}{R}$$

: for both translocated (T) and untranslocated (U) parts

Young-Laplace equation

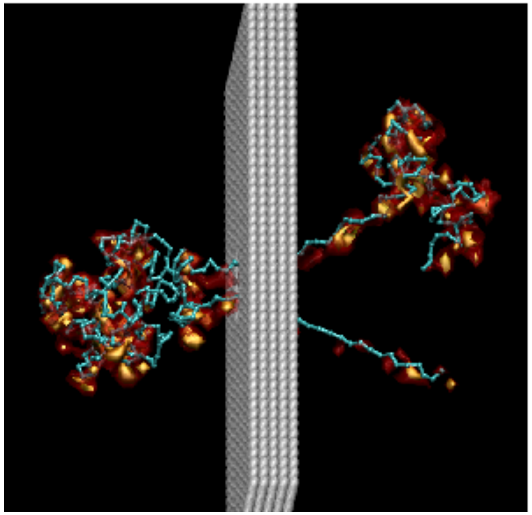
$$r(t) = \frac{N_T(t)}{N_0} \Rightarrow \frac{dW}{dt} \sim \left[N_0^{2\nu_T} r^{2\nu-1} \frac{dr}{dt} - N_0^{2\nu_U} (1-r)^{2\nu-1} \frac{dr}{dt} \right]$$ **: const!**

Integration over time:

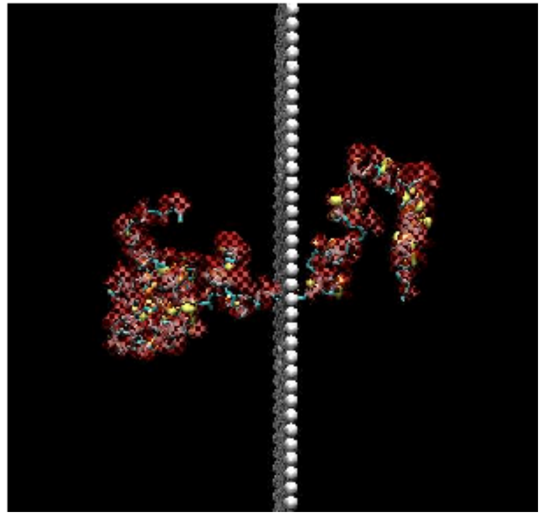
$$r(t) \in [0,1], \quad t \in [0,\tau] \Rightarrow \tau \sim N_0^{2\nu_U} \Rightarrow \alpha = 2\nu_U = 1.2$$

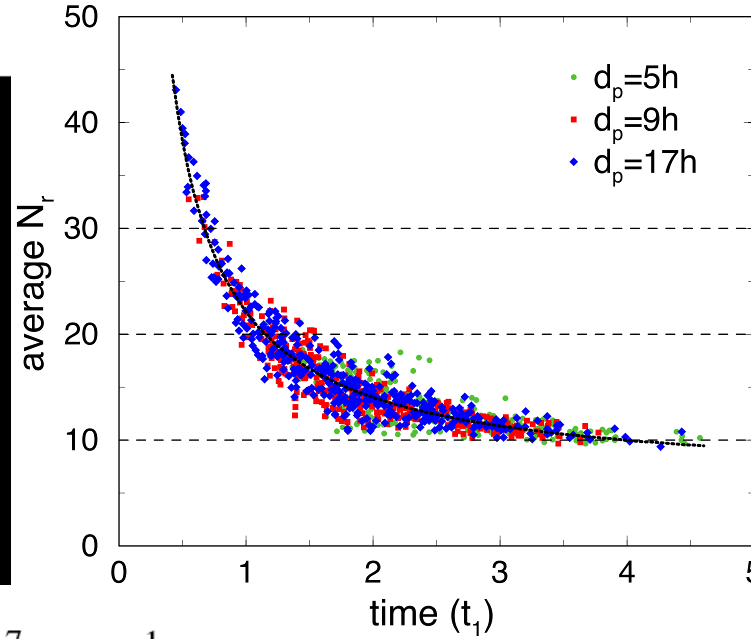
Multifile translocation: simulation

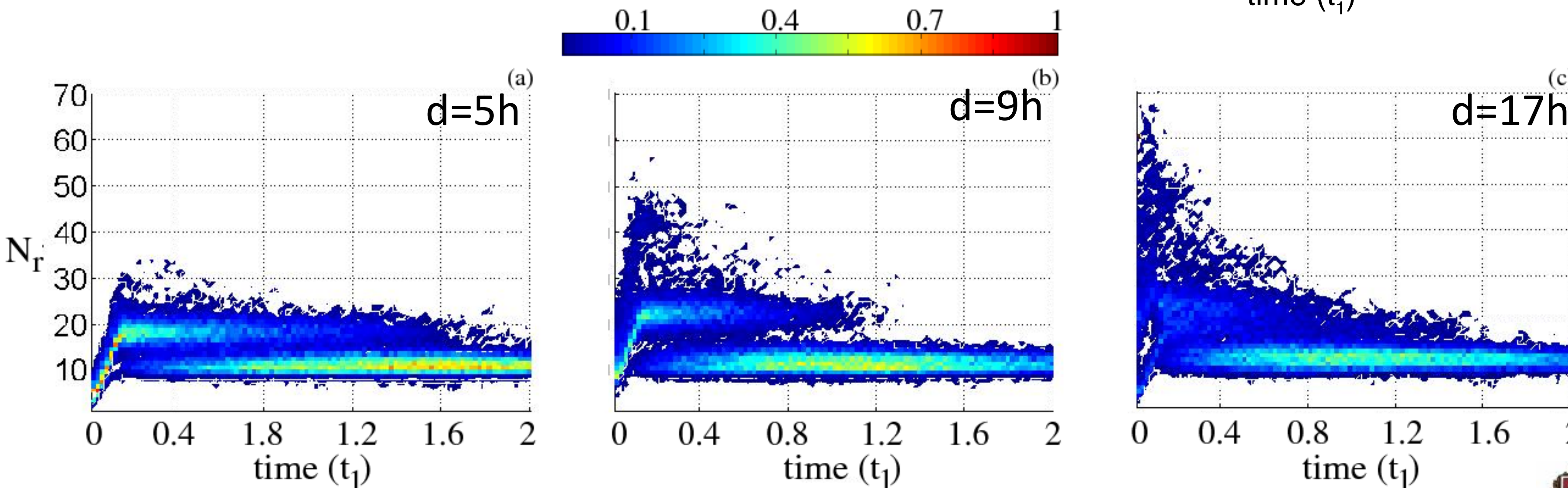
wide, thick pore



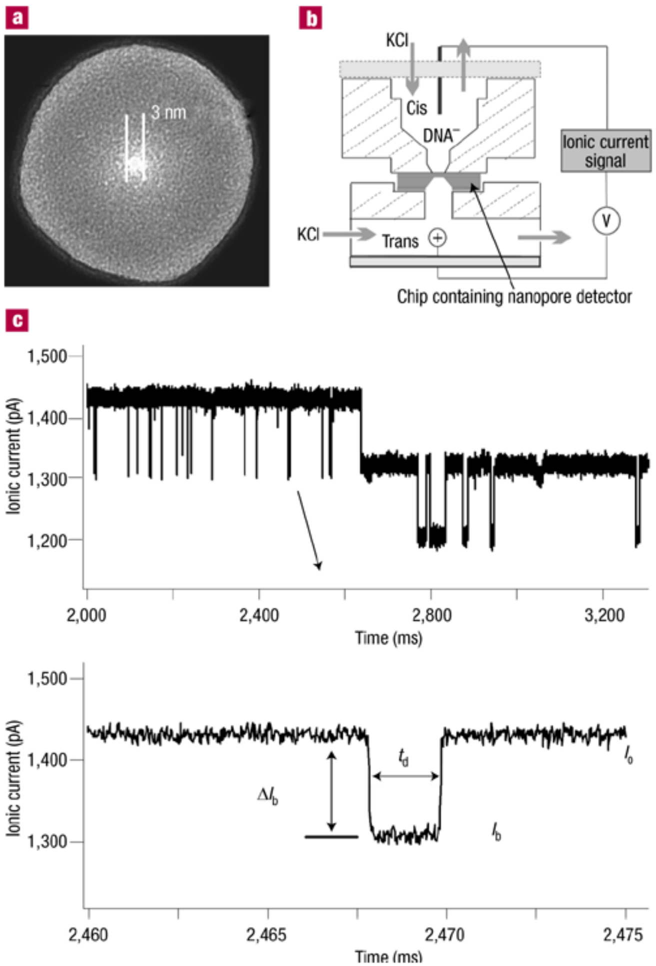
narrow, thin pore

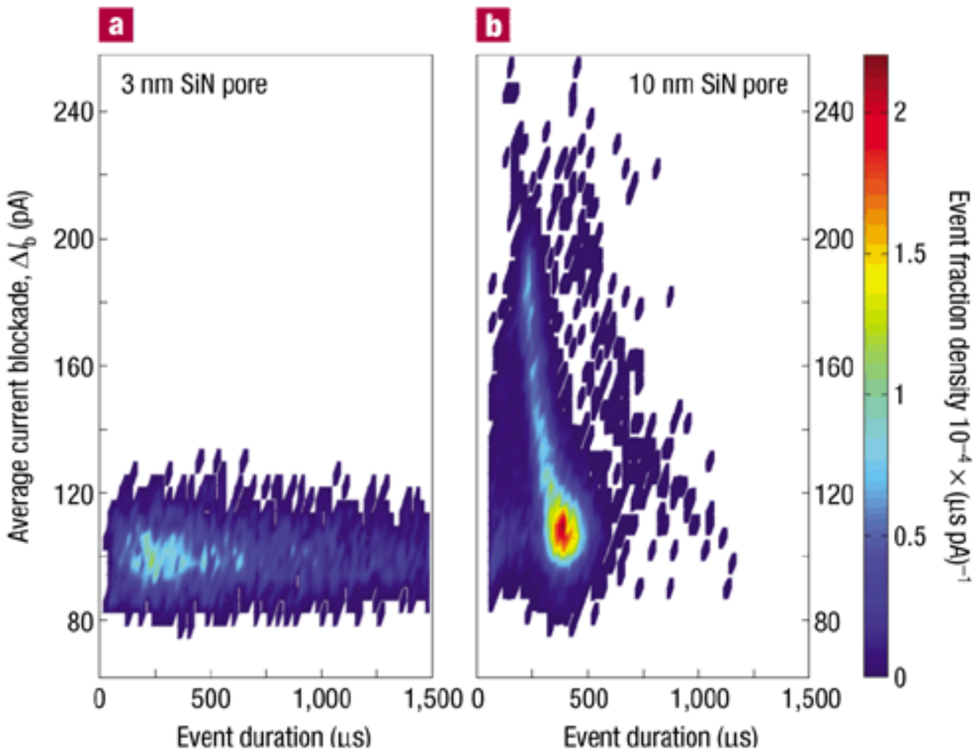




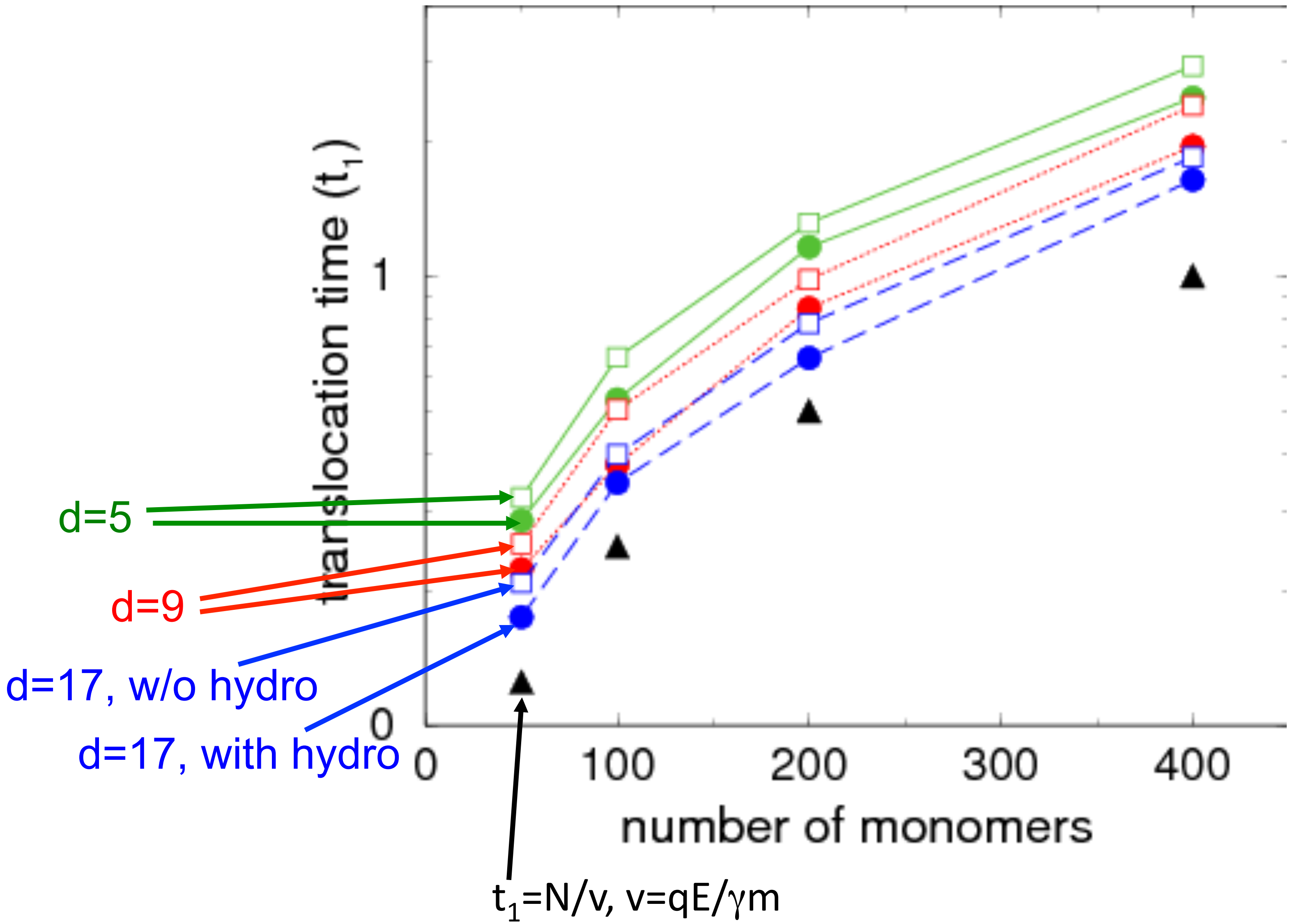


Multifile translocation: experiment





Li, Gershow, Stein, Brandin, Golovchenko,
Nature Materials, 2003

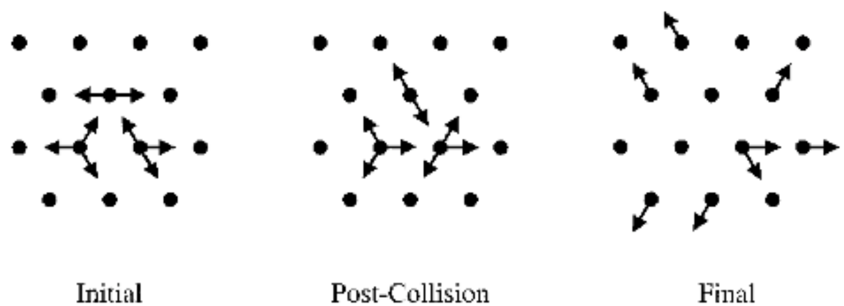


Hydrodynamics almost doubles the effective diameter of the pore!

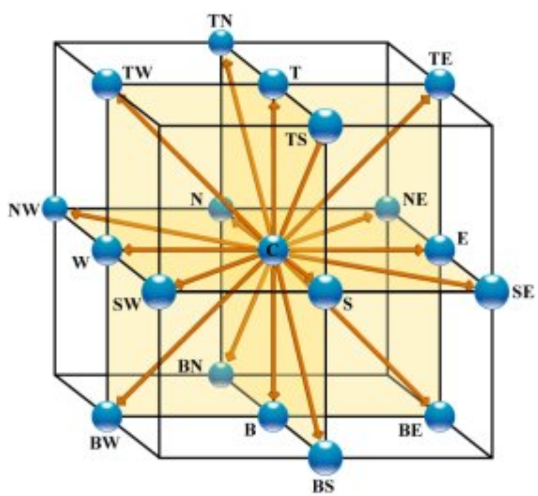
Fluid dynamics by cellular automata : Lattice Boltzmann Equation (LBE)

$$f_i(\vec{x}+\vec{c}_i\Delta t, t+\Delta t) = f_i(\vec{x},t) - \omega\Delta t(f_i - f_i^{eq})(\vec{x},t)$$

$$f_i^{eq} \propto \rho w_i\left[1+\frac{\vec{c}_i\cdot\vec{u}}{c^2}+\frac{(\vec{c}_i\cdot\vec{u})^2 - c^2u^2}{2c^4}\right]$$



Bhatnagar-Gross-Krook algorithm



Reproduces the physics of fluid dynamics (Navier-Stokes equation)

Fluid properties :

Fluid density

$$\rho(\vec{x},t) = \sum_i f_i(\vec{x},t)$$

Momentum (flow)

$$\rho(\vec{x},t)\vec{u}(\vec{x},t) = \sum_i f_i(\vec{x},t)\vec{c}_i$$

Stress Tensor

$$\overleftrightarrow{\sigma}(\vec{x},t) = \frac{\nu\omega}{c_s^2}\sum_i \vec{c}_i\vec{c}_i\left[f_i - f_i^{eq}\right](\vec{x},t)$$

Wall Stress

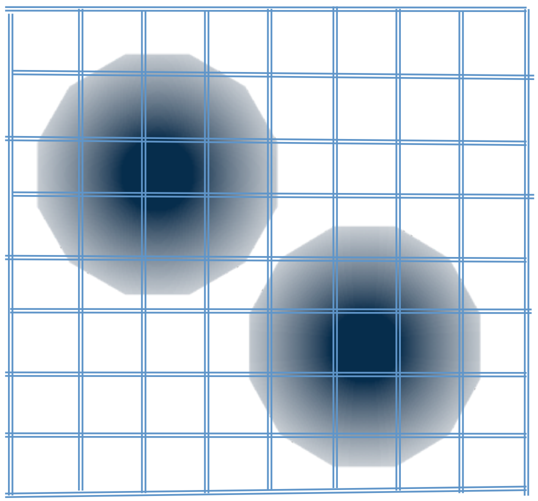
$$S(\vec{x}_w,t) = \sqrt{\left(\overleftrightarrow{\sigma} : \overleftrightarrow{\sigma}\right)\left(\vec{x}_w,t\right)}$$

Definition of "particles" (cells, proteins, …)

$$\tilde{\delta}_\xi(x-R) = \prod_{\alpha=x,y,z} \tilde{\delta}_\xi(x_\alpha - R_\alpha)$$

$$\sum_x \tilde{\delta}_\xi(x-R) = 1$$

$$\tilde{\delta}_\xi(a) = \begin{cases} \frac{1}{2\xi}\left(1+\cos\left(\frac{\pi|a|}{\xi}\right)\right) & 0 \le |a| \le \xi \\ 0 & \xi \le |a| \end{cases}$$



$$\varphi(x,R) = -\gamma\big(V - u(x)\big)\tilde{\delta}_\xi(x-R)$$

$$F^H = \sum_x \varphi = -\gamma\left(V - \tilde{u}\right)$$

$$\tilde{u} = u * \tilde{\delta}_\xi$$

$$\Delta f_p = -\frac{w_p}{c^2} c_p \cdot \sum_R \varphi$$

Equations of motion:

$$\Xi \frac{d\Psi}{dt} \equiv \begin{pmatrix} M \frac{dV}{dt} \\ I \frac{d\Omega}{dt} \end{pmatrix} = \begin{pmatrix} F + F^H \\ T + T^H \end{pmatrix} \equiv \Phi + \Phi^H$$

$$\Phi^H_{6\times1} = \Gamma_{6\times6} \Psi^*_{6\times1} + \Delta_{6\times3\times3} : E_{3\times3}$$

$$\Psi^* \equiv \begin{pmatrix} V - u \\ \Omega - \omega \end{pmatrix}$$

Γ	Grand Resistance matrix
Δ	Shear Resistance matrix
E	Strain tensor
u	Fluid velocity @center
$\omega = \frac{1}{2} \partial \times u$	Fluid vorticity

Brenner et al '72

Γ and Δ depend on the whole configuration

Pair-wise superposition

$O(N^3)$ complexity!

Brady & Bossis '89

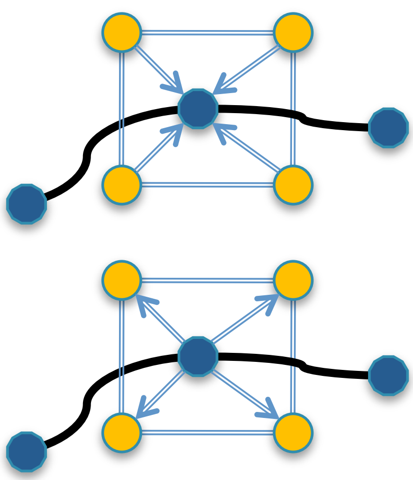
Fluid-particle coupling:

$$(\partial_t + v \cdot \partial_x) f = -\omega (f - f^{eq}) - \frac{1}{M} \sum_R F^H \cdot \partial_v f$$

$$\frac{d}{dt} V = \frac{1}{M} (F + F^H)$$

$$F^H = -\gamma \left[V - u(x, \{R, V\}) \right] \delta(x - R)$$

Momentum exchange
(Newton's restitution law)

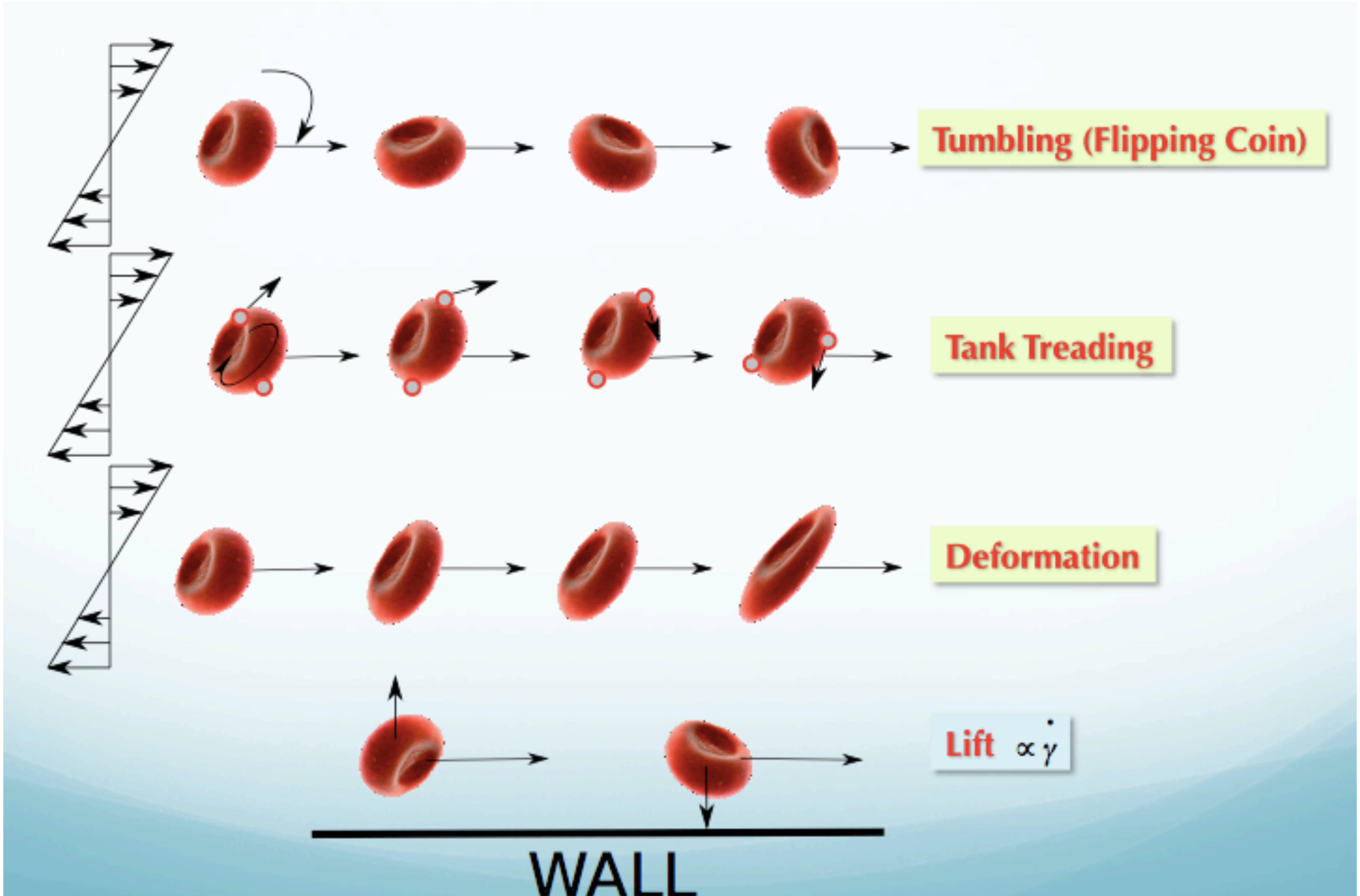


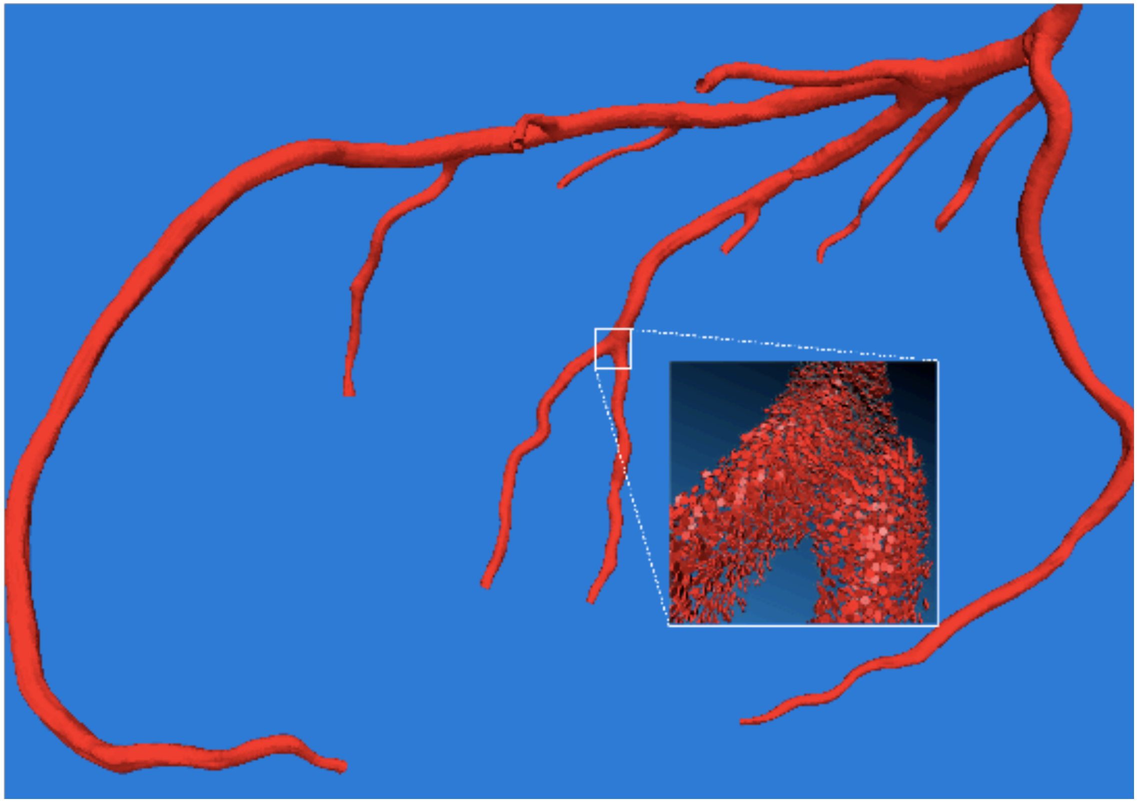

Computing in Science & Engineering

Multiscale Simulation of Nanobiological Flows

Maria Fyta and Efthimios Kaxiras
Harvard University
Simone Melchionna
University of Rome
Sauro Succi
National Research Council, Italy

Red Blood Cell in Motion





(geometry from cadaver)

MOVIES!

Students:	Wade Hsu Ryan Barnett Ari Turner Mauro Bisson Amanda Peters Michelle Borkin
Postdocs:	Maria Fyta Greg Lakatos Simone Meclhionna Paul Maragakis Gang Lu Sheng Meng
Visitors:	Sauro Succi Costas Papaloukas George Papamokos Massimo Bernaschi