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 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, ...)

Core of eight histone molecules
Histone H1
"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”

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

3´-3´-mode (unwinding)

5´-5´-mode (compression)

30 % stretch

60 % stretch

90 % stretch


Structures from classical MD simulations (*empirical*)
Electronic states from DFT calculations (*first-principles*)
30 % overstretched

B-DNA at natural length
“DNA – carbon nanotube interaction”
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 \textit{ab-initio} calculations*. Assumption: \textit{separable} interactions

- Hydrogen bonding (distance; dihedral, flip angles)
- Stacking interactions (distance; twist angle)
- Backbone interactions (distance; 3’-5’ orientation)
- Electrostatic interactions

* to the extent possible

\textbf{Data points:} DFT calculations, \textbf{Lines:} fits with simple curves
Coarse-grained potential – hydrogen bonding I

Distance between bases, dihedral angle
Coarse-grained potential – hydrogen bonding II

Flip angle

\[ E_{\text{hb}}(r) \]
Coarse-grained potential – stacking interactions

Distance between planes

\[ E_{st}(\text{eV}) \]

\[ r_{st}(\text{Å}) \]

\( \theta_{tw} \)

\( d_1 \)

\( d_2 \)

\( d_1' \)

\( d_2' \)

GC-TA

GC-CG

AT-TA

GC-AT

GC-GC

AT-AT
Coarse-grained potential – stacking interactions

(a) Twist angle between pairs

(b) Twist angle between pairs

 Twist angle between pairs

$E_{tw}$ (eV)

$\theta_{tw}$ (deg)

AT-AT

GC-GC

GC-AT

GC-TA

GC-CG

AT-TA
Coarse-grained potential – backbone interactions

(a) 

(b)
Coarse-grained potential – electrostatic interactions

Deoxyribonucleic Acid

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

\[ \varepsilon(r) = \varepsilon_{in} \quad r < r_0 \]

\[ \varepsilon(r) = \varepsilon_{in} e^{\alpha(r-r_0)} \quad r_0 < r < r_1 \]

\[ \varepsilon(r) = \varepsilon_{\infty} e^{-\kappa r} \quad r > r_1 \]

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

3 parameters!

\[ r_0, \alpha, \varepsilon_{in} \]
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

(a) Force versus Rise per base-pair (Å)
- Black dots: pulling 3’ end
- Red diamonds: pulling 5’ end

(b) Force versus Twist angle $\theta_{tw}$ (deg.)

Insertion image (a): Images of a DNA molecule with varying force applied.

Insertion image (b): Graph showing different regimes: entropic, enthalpic, and overstretches.

Graph (b): Worm-like chain model and overstretches transition.
Melting simulations
Major open issues:
- Systematic method for coarse graining DOF (normal mode analysis?)
- Temperature?
- Entropy?


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)
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

Mellor et al., *PNAS* (2000)

Storm et al., *NanoLetters* (2005)

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 $\alpha = 1.27 \pm 0.03$. 
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
**Lattice-Boltzmann Method (LBM)**

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

\( f_i(x,t), \ i=1,n: \) probability to find a particle at lattice site \( x \) at time \( t \) with speed \( c_i \)

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

local equilibrium:

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

\[
\rho(\vec{x}, t) = \sum_p f_p(\vec{x}, t) \quad \text{density}
\]

\[
\rho \vec{u}(\vec{x}, t) = \sum_p f_p(\vec{x}, t) \vec{c}_p \quad \text{flow speed}
\]

\[
\vec{P}(\vec{x}, t) = \sum_p f_p(\vec{x}, t) \vec{c}_p \vec{c}_p \quad \text{momentum flux}
\]
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 $\nu_p$:

$$m \frac{d\nu_i}{dt} = F^c_i + F^f_i + F^r_i + \lambda_i \partial_{\ddot{r}_i} \sigma$$

- bead-bead interactions
- random force
- constraint force
- Solute-solvent interactions

Coupling LB to MD:

$$F^f_i = -m\gamma (\ddot{u}_i - \ddot{\nu}_i)$$

$\nu$ : bead velocity
$u$ : fluid velocity

Coupling MD to LB:

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

Details of simulation

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

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

  [translocation time \(<\ll\) DNA relaxation time]

\[
\frac{F_{\text{pull}}R}{kT} \gg 50
\]
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

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

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

(SAW : \( \nu = 0.6 \))

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

translocated (\( T \)) part
untranslocated (\( U \)) part

\[ R_g / R_U(0) \]

\[ t/\tau \]

\[ \# \text{ of events} \]

\[ R_g / <R_U(0)> \]
Anisotropic radius of gyration

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

Untranslocated-transverse dominates
($\nu_U = 0.6$)
\[ S_H(t) = \frac{dW_H}{dt} = \gamma \left\langle \sum_i^N \vec{v}_i(t) \cdot \vec{u}_i(t) \right\rangle \]  
Hydrodynamic interactions

\[ S_E(t) = \frac{dW_E}{dt} = \left\langle \sum_i^N \vec{F}_{\text{drive},i}(t) \cdot \vec{v}_i(t) \right\rangle \]  
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!

\[ dW = P \, dV + \sigma_\gamma \, dA \]

\[ P = \frac{2\sigma_\gamma}{R} \]

: 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

\[ d = 5h \]
\[ d = 9h \]
\[ d = 17h \]
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)

\[
f_i(\bar{x} + \bar{c}_i \Delta t, t + \Delta t) = f_i(\bar{x}, t) - \omega \Delta t (f_i - f_i^{eq})(\bar{x}, t)
\]

\[
f_i^{eq} \propto \rho_w i \left[ 1 + \frac{\bar{c}_i \cdot \bar{u}}{c^2} + \frac{(\bar{c}_i \cdot \bar{u})^2 - c^2 u^2}{2c^4} \right]
\]

Bhatnagar-Gross-Krook algorithm

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

Fluid density

Momentum (flow)

Stress Tensor

Wall Stress

\[ \rho(\bar{x},t) = \sum_i f_i(\bar{x},t) \]

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

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

\[ S(\bar{x}_w,t) = \sqrt{(\bar{\sigma} : \bar{\sigma})(\bar{x}_w,t)} \]
Definition of “particles” (cells, proteins, ...)

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

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

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

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

\[ F^H = \sum_x \varphi = -\gamma (V - \tilde{u}) \]

\[ \tilde{u} = u \ast \tilde{\delta}_\xi \]

\[ \Delta f_p = -\frac{w_p}{c^2} c_p \cdot \sum_R \varphi \]
Equations of motion:

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

\[
\Phi^H_{6 \times 1} = \Gamma_{6 \times 6} \Psi^*_{6 \times 1} + \Delta_{6 \times 3 \times 3} : E_{3 \times 3}
\]

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

\begin{align*}
\Gamma & \quad \text{Grand Resistance matrix} \\
\Delta & \quad \text{Shear Resistance matrix} \\
E & \quad \text{Strain tensor} \\
u & \quad \text{Fluid velocity @ center} \\
\omega & = \frac{1}{2} \partial \times u \quad \text{Fluid vorticity}
\end{align*}

Brenner et al ‘72

\(\Gamma\) and \(\Delta\) depend on the whole configuration

Pair-wise superposition

\(O(N^3)\) complexity!

Brady & Bossis ‘89
Fluid-particle coupling:

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

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

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

Momentum exchange
(Newton’s restitution law)

Multiscale Simulation
of Nanobiological Flows
Red Blood Cell in Motion

Tumbling (Flipping Coin)

Tank Treading

Deformation

Lift $\alpha \gamma$

WALL
(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