Introduction to Simulation and Modeling of DNA Systems

Aleksei Aksimentiev
Department of Physics, University of Illinois at Urbana-Champaign
WHAT IS LIFE?
The Physical Aspect of the Living Cell

BY

ERWIN SCHRODINGER

SENIOR PROFESSOR AT THE DUBLIN INSTITUTE FOR ADVANCED STUDIES

Based on Lectures delivered under the auspices of the Institute at Trinity College, Dublin, in February 1943

CAMBRIDGE
AT THE UNIVERSITY PRESS

1948
DNA, the blueprint
A nucleus of a human cell contains 23x2 chromosomes

http://homepage.smc.edu/wissmann_paul/histology/
Human chromosomes

Chr 1  Chr 2  Chr 3  Chr 4  Chr 5  Chr 6  Chr 7  Chr 8

Chr 9  Chr 10  Chr 11  Chr 12  Chr 13  Chr 14  Chr 15  Chr 16

Chr 17  Chr 18  Chr 19  Chr 20  Chr 21  Chr 22  Chr X  Chr Y
Each chromosome is millimeters-long single fiber: chromatin

Entangled chromatin fiber (~mm)

Human genome: 3 billions of A, T, G and C
DNA is a highly negatively charged polymer that forms the human genome.

A human nucleus contains 23x2 chromosomes made of 12 billion nucleotides (~2 meters of DNA).

~10 µm

A human nucleus contains 23x2 chromosomes made of 12 billion nucleotides (~2 meters of DNA)

Animation by Chris Maffeo

Same sign charges ....

DNA is surrounded by counter ions

- +1e, sodium or potassium
- +2e, magnesium or calcium
- +3e, spermidine
- +4e, spermine

Effective attraction between DNA is observed when counterions have charge ≥ 2e
MD simulation of dense DNA arrays

What we control

- DNA density (or harmonic constraint radius)
- $[\text{Na}^+]_{\text{buf}} \sim 200$ mM
- $[\text{Mg}^{2+}]_{\text{buf}} \sim 0$ or $20$ mM

What we measure

- Pressure as a function of $[\text{ion}]$ & $[\text{DNA}]$
- DNA / ion distribution:
- DNA / ion diffusion inside the array

Cylindrical harmonic constraint (radius of 10 – 12 nm) only against DNA

Water / salt free to move
The standard MD force field fails to predict internal pressure of a DNA array

Too strong Na/Mg-phosphate attraction induces artificial DNA clusters!! $[\text{Na}] \sim 4\text{M}!!$

Recalibrate ion-DNA parameters using osmotic pressure data

• Osmotic pressure is directly related to ion-pair formation: \( \pi = \phi cRT \)

• Pros: modify only ion-DNA phosphate interaction, without altering ion-water interaction.

• Cons: nothing.

* Luo & Roux, JPCL (2009)
Improved parametrization of ion-DNA interactions

\[ [\text{Mg}]_{\text{buf}} \sim 20 \text{ mM} \]
\[ [\text{Na}]_{\text{buf}} \sim 200 \text{ mM} \]
CUFIX: Accurate parameterization of non-bonded interactions

http://bionano.physics.illinois.edu/CUFIX

Volume 26  Number 41  15 October 2014

Topical review

Close encounters with DNA

C Maffeo, J Yoo, J Comer, D B Wells, B Luan and A Aksimentiev
The **nucleotide sequence** contains biological information

Central dogma of molecular biology
The Human Genome Project

Duration:
October 1990 - 2003

Discovered ALL
20,000-25,000 human genes

Determined complete sequence of the 3 billion DNA bases

5’-ACCGGTGGGTGCATAGCTGTGCTGTAAGTGAAGTG
AGGCGGCCAGGTTGTTGAAAGTCGATGTAGTTCGTAG
GTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTG
GACAGGGTGACTAGTGAATCGATGCTAGCCTAGCTA
GTCAGTGGTGCTAGCTACGATCGATCGATTTCAGGCTGCT
... and ~ 3,000,000 more pages!

(one month to show 24/7)

Just four letter:

~715 Mb

DNA code is billion times more efficient

2 bits

0 0
1 1

8 bits = 1b

4/8*3*10^9
Differences in the code are important

Among unrelated individuals, 99.4% of the sequence is similar
That is still over 1,000,000 differences.

You and chimpanzee: 99%

**Advanced diagnostics**
(early detection and, possibly, prevention of 4,000 genetic disorders)

**Research instrumentation**
(reconstruction of the tree of life, human history, psychology)

**Personal pharmaceutics**
(tailor drugs to an individual's genetic make-up)

**Cancer: disease of DNA**

**Prenatal diagnostics**

**Single cell sequencing**
The Sanger’s method

Nobel Prize in Chemistry 1980

As the DNA is synthesized, nucleotides are added on to the growing chain by the DNA polymerase.

The reactions start from the same nucleotide and end with a specific base.

Fluorescence-based sequence gel

http://bbrp.llnl.gov
Cost of sequencing a human genome (logarithmic scale)

Phase 1: $100K genome

$1000 genome project begins

Phase 2: $1000 genome

NIH National Human Genome Research Institute
genome.gov/sequencingcosts
DNA electrophoresis through a nanopore

- Isolates 1nm³ of volume
- Automatic loading and reloading
- Highly processive, single-file transport
- Compatible with several detection schemes
- No limit on the read length
The nanopore technology

3 nm
1000 times per minute
Nanopore sequencing of DNA

First described by Kasianowicz et al. PNAS 1996

The ionic current blockade reveals the sequence of the confined nucleotides

Nature Reviews Drug Discovery 1, 77-84 (January 2002)
Sequencing DNA using a biological nanopore

Experimentally measured ionic current blockades

Nature Biotech. 32: 829 - 834 (2014)
MinION (Oxford Nanopore Technologies)

MinION: 800 parallel detection wells
Read length: up to 200,000 nucleotides
Biological pore (R9, CsgG-derivative)
Helicase motor to control translocation
Direct readout of DNA epigenetic markers
Accuracy: 80 - 98%

Only 100g, great for in situ measurements
Biological nanopore: why does it work?

Bigger nucleotides block the current less!

Blockades are non-additive

Electric field

MD simulation neutravidin-anchored ssDNA in MspA

Manrao ... Gundlach, Nature Biotech. 30: 349 - 353 (2012)
All-atom molecular dynamics simulations: the computational microscope

Massive parallel computer
Blue Waters (UIUC): ~200,000 CPUs

Atoms move according to classical mechanics (F= ma)

Interaction between atoms is defined by molecular force field

Time scale: ~ 0.1-100 μs
Length scale: 10K - 100M atoms or (< 50 nm)$^3$
Time resolution: 2 fs
Spacial resolution: 0.1 A

Nanoscale 2:468 (2010)
Setting up a simulation is like cooking.
Setting up a simulation is like cooking

Components
- protein
- DNA
- lipid
- ions
- water

F = ma @ 300 K
Time step = 1 fs
Computing conductance of $\alpha$-hemolysin with molecular dynamics

Protein + lipid bilayer membrane + 1M water solution of KCl = ~300,000 atoms

Average electrostatic potential map
Current-voltage curve of $\alpha$-hemolysin

$Biophys. J.$ 88:3745 (2005)

$$I(t) = \frac{1}{\Delta t L_z} \sum_{i=1}^{N} q_i(z_i(t + \Delta t) - z_i(t))$$

Instantaneous current
Current-voltage curve of $\alpha$-hemolysin

$Biophys. J. 88:3745$ (2005)

$$I(t) = \frac{1}{\Delta t L_z} \sum_{i=1}^{N} q_i (z_i(t + \Delta t) - z_i(t))$$

Instantaneous current

Ionic current at 120mV:
Experiment: 120 pA
Simulations: 130 pA
MD simulations of current blockades in MspA

ACS Nano 2016, 10, 4644

DE Shaw’s Anton

Electric field

dNTP

DNA synthesis

Complete pore

Truncated pore

MD simulation ssDNA- DNA polymerase complex
(350,000 atoms, 150 ns)

Reduced system (28,000 atoms)
MD simulations of current blockades in MspA

5’-poly(dT)  
Bhattacharya, Yoo, Aksimentiev, ACS Nano 10, 4644 (2016)

\[ \langle I/I_0 \rangle = 20.1 \pm 2.3 \]

5’-poly(dA)

\[ \langle I/I_0 \rangle = 31.6 \pm 3.0 \]
Water mediates DNA sequence recognition

Red: structured water
Blue: unstructured water

Current increase with the number of water molecules as $N^{1.7}$

Bhattacharya, Yoo, Aksimentiev, ACS Nano 10, 4644 (2016)
MinION (Oxford Nanopore Technologies)

MinION: **800 parallel detection wells**

Read length: up to 200,000 nucleotides

Biological pore (R9, CsgG-derivative)

Helicase motor to control translocation

Direct readout of DNA epigenetic markers

Accuracy: 80 - 98 %

Only 100g, great for *in situ* measurements
Silicon Nanotechnology for Sequencing DNA
(Back in 2003!)

- ultra-thin membranes

TEM X-section through a gate

polysilicon

0.7nm oxide

silicon

10nm (0.34nm = 1bp)

DNA

2nm

TEM (top-down projection)

T. Sorsch, V. Dimitrov, C. Ho
Imaging nanopores using MD

Transmission electron micrograph of a nanopore in Si$_3$N$_4$ (Timp Lab, UIUC)

All atom model of a nanopore in Si$_3$N$_4$ built by the Aksimentiev group

DNA electrophoresis through a nanopore

- Isolates 1nm$^3$ of volume
- Automatic loading and reloading
- Highly processive, single-file transport
- Compatible with several detection schemes
- No limit on the read length

Ionic current through pore measured

Current transients associated with passage of dsDNA

![Illustration of DNA electrophoresis through a nanopore with current transients](image-url)
Graphene Nanopores
Simulation of DNA translocation through graphene nanopore

Top view

Side view


14-A diameter pore (surface-to-surface); 3-layer graphite; poly(dT)$_{20}$; 500 mV bias
DNA transport is stepwise

14-A diameter pore (surface-to-surface); 3-layer graphite; poly(dT)$_{20}$; 500 mV bias

DNA-graphene interactions act as a stepping motor!

Can ionic current blockades can reveal the DNA sequence?

Atomic-Resolution Brownian Dynamics simulations of ionic current blockades in graphene nanopores
Sequencing proteins using graphene nanopores?

Modeling Nanopores for Sequencing DNA

- Biological Nanopores
  - α-Hemolysin
    - Tutorial Difficulty: Easy
  - Manish Shankla

- Solid-State Nanopores
  - Silicone Nitride
    - Tutorial Difficulty: Medium
  - Silica
    - Tutorial Difficulty: NIGHTMARE
What is the next step?

• Biology is understood

• Means to detect a problem developed

• How are we going to fix the problem?

  Target ONLY diseased cells

  Prevent degradation before reaching target

  Make sure the drug enters a cell
DNA origami

Building a structure with nanoscale precision by folding DNA

Viral DNA (scaffold)

Synthetic DNA (staples) ~30–40 nucleotides

2.5 nm
DNA origami
DNA origami

Idea: direct folding of a long single strand of DNA into desired shapes.

Paul Rothemund (2006), NATURE Vol 440:297
Design and characterization of DNA nanostructures

Computer-aided design of DNA origami with caDNAno (Shih group, Harvard U.)

Transmission electron microscopy and/or atomic force microscopy validates the design

Cryo-EM reconstruction, the only experimentally derived structural model
DNA origami box


- Designed with Cadnano plugin using the design from Zagdedan et al.
- DNA origami box is 18.0 nm by 13.5 nm by 17.5 nm.
- Solvated system contains ~2.8 million atoms, where ~70,000 are in the DNA origami box.
- The solution contains 10mM Mg ions; gemcitabine (anti-cancer nucleoside), deoxycytidine kinase (protein that activates gemcitabine) and siRNA
Cancer killing DNA robot (2012)

Science 335, 831–834 (2012).

DNA origami syringe

That's the question from experimentalist.

Experiment: Keyser lab (Cambridge, UK)  
caDNAno

• Computer-Aided Design of DNA origami made by Shih group at Harvard.

• Designed as a plugin of Autodesk MAYA.

• Limited to a design of antiparallel DNA helices in a honeycomb or square lattice.

From caDNAno to all-atom

- caDNAno returns topology (json) and sequence (csv) information.

- cadnano2pdb.pl combines json and csv files into a PDB file.

- CHARMM36 force field
- Explicit water
- [MgCl₂] ~ 10 mM
- NAMD
- 1 to 3M atoms
- 500 to 1,000 CPUs
Structural dynamics
Structural fluctuations reveal local mechanical properties

MD trajectories allow us to compute natural bending and torsion as well as persistence length

- Inter-DNA distance in color map
- Chicken wire frame represents center line of helices & junction

Our simulations predict higher rigidity for honeycomb-lattice design.

Tiled DNA nanostructures

DNA Origami + =

Supplementary Information for Molecular mechanics of DNA bricks: 

30 AAACCTGTATTGAATCAAGCGAACCAGACCGGACC
29 TTGCCCTTCACAACCCGAAAGAGAATGACCATAAA
28 GACGGCCGCTTTCCACAACATACGAGCCGTAGGA
27 GGCGGTGATGGTGACGTCAAACACTATTAAAGAACGT
26 GGGCTCGAATTGCAAAGCGCGT
25 TCAATTCTACTAAAAAATTTT
24 AACTGTAGCTCAACATGTTTTAAATATCAGAAGCACTCAGAGCATAAA
23 ATATATATTTTCAATGCCTGAGTAATGCGGAGACAGAG
22 CATACGGGGATGTGTTTTCCCCCCCAAAATAAATC
21 CTGATGGCTTATCCCAATTCTGCGAACGAGTAGAT
20 CACTTCACCAGGAGCTTCACCCCTCAAATGCTTTA
19 AAACTGCGCAATCTAGAGGGGATTCTCTAGCCAGC
18 GGCGCGAGCAGTACGGTGTCT
17 ACCATTAGATGCAGCAAGGGCCTCTTCCAGTGCCATC
16 TCGGCAAAATCCCCAAGAGTCGTCAGGATTAGAG
15 CCGGAAACCAG
14 GCAACTAATGAAAAGGTGGCA
13 GGAAGTTTCATTCCATATAATGTTAGCAAT
12 TTAGTAGGTTTGATAAGAGGTCATTTTAATTCTGA
11 GAGTGTTGTTCCAGTTTGGAATTATAAATAGGCG
10 AATTGCTGAAAGAGGAAGGGCAAAAGACTAAC
9 TTTAAGCCCCAGCCAAAAGAACAGCCAGCTTTCC
7 GCGGATCCCTGACTATTATAGT
6 TATCGCGTTTTTGCGGAGAGAGTTACATTTCGACG
5 TCAAAAATCATGAGTGAGATTAGCAAACGCCAGGGCT
4 AACAGTTCCTCACTGCGCAACAGCAACGACGGCGC
3 ATCGTCATAAATATTCCGTGCCAGCAGGGTGGAGG
2 AATGGGCGAAAAACCGTCTGGACTCCAGTT
1 ATCAATAGGGG

Continued on next page

3.2.1 Template Design: Carving a Structure

Included with this tutorial are several template files, which are collections of the initial M13mp18 sequence provided by CaDNAno program [2].

We have several tutorials on how to build structures with LegoGen, depending on your amount of skill with CAD software and would prefer to rapidly build using a template, several templates are provided and their use is explained in Section 3.2.1. If you would like a full understanding of the abilities of LegoGen, please consult all three sections. If you feel you are familiar with manual and tutorials for it, giving many ways to build structures as you wish.

LegoGen is a tool for converting CAD models into DNA nanostructures through the DNA brick method. It provides .pdb and .pdf files for all-atom molecular dynamics simulations in NAMD, as well as a list of downloadable sequences for use in experiments. The tool also outputs the coordinates of DNA brick units, called voxels, and a Panel folder (contains) that can be used to build subsequent structures from the same mesh library. CAD models can be easily made with LegoGen with the included 3D model editor. Encoder. Once your models are in WRL format they can be converted to a DNA brick structure, mechanical properties and ionic conductivity.

To use LegoGen, the following requirements must be met:

- A voxel model is required, which is defined as a set of connected voxels that form a solid structure, defined by a bounding box. The box size is determined by the number of voxels in each dimension of the structure.
- A voxel unit is equal to 2.5 nm (a single voxel) or 5 nm (two voxels), depending on the size of the structure.
- The box size must be a power of 2, which determines the number of voxels in each dimension of the structure.
- A residue mapping file is required, which maps each voxel to a specific DNA sequence.
- A DNA sequence file is required, which contains the DNA sequence for each voxel.

To convert a voxel model to a DNA brick structure, use the following steps:

1. Load the voxel model into LegoGen.
2. Select the voxel unit size and the box size.
3. Load the residue mapping file.
4. Load the DNA sequence file.
5. Use the conversion tool to generate the DNA brick structure.

Once the DNA brick structure is generated, it can be exported to a file format compatible with NAMD, such as .pdb or .pdf.

4.2 Use of a Residue Mapping File and Preserving Sequence

In situ structure, mechanical properties and ionic conductivity.

4 CONVERTING MODELS WITH LEGOGEN

Converting models with LegoGen involves the following steps:

- Load the voxel model into LegoGen.
- Select the voxel unit size and the box size.
- Load the residue mapping file.
- Load the DNA sequence file.
- Use the conversion tool to generate the DNA brick structure.
- Export the DNA brick structure to a file format compatible with NAMD, such as .pdb or .pdf.

Once the DNA brick structure is generated, it can be used for simulations in NAMD.
Cryo-EM reconstruction versus all-atom simulation

Cryo-EM reconstruction versus all-atom simulation

Cryo-EM reconstruction versus all-atom simulation

MD simulation of the cryo-EM object starting from a caDNAno design


7M atom solvated model
130 ns MD trajectory
MD simulation of the cryo-EM object starting from a caDNAno design


7M atom solvated model
130 ns MD trajectory
MD simulation of the cryo-EM object starting from a caDNAno design

Simulation on Blue Waters (UIUC)

Time scale: 200 ns / Size ~7,000,000 atoms

Maffeo, Yoo & Aksimentiev, NAR 44: 3013
Cryo-EM, Dietz group, PNAS (2012)
http://bionano.physics.illinois.edu/origami-structure
De novo prediction of DNA origami structure


Maffeo, Yoo & Aksimentiev, *NAR* 44: 3013

Modest computational cost
(10 hours single workstation)

Server implementation has been requested by experimentalists

nanoHub implementation pending

http://bionano.physics.illinois.edu/origami-structure
CG ENRG MD can be used routinely during the design process.
Modeling biomolecular systems with atomic-resolution Brownian dynamics (ARBD)

**Large biological systems**

- **Chromatin:** gene-compacting and regulating protein–DNA complex

- **Chromatophor:** light harvesting organelle

**Nanotechnological systems**

- DNA sequencing via nanopores

**Grid-based representations of molecules with shape**

**Point-particle representations of polymers & small molecules**

[Link to Development/Download/download.cgi](http://www.ks.uiuc.edu/Development/Download/download.cgi)
http://bionano.physics.illinois.edu/dna-nanotechnology
Acknowledgements

Jeff Comer
Maxim Belkin
Rogan Carr
David Wells
Manish Shankla
Swati Bhattacharya
Jejoong Yoo
Chen-Yu Li
Chris Maffeo
Shu-Han Chao
Derek Vandamme
Jim Wilson

Swati Bhattacharya
Manish Shankla
Jejoong Yoo
David Wells
Chen-Yu Li
Chris Maffeo