

МГУ имени М.В. Ломоносова биологический факультет кафедра биоинженерии группа интегративной биологии





Интегративное и суперкомпьютерное моделирование нуклеосом

Алексей Константинович Шайтан, к.ф.-м.н. alex@intbio.org

Семинар "Суперкомпьютерные технологии в науке, образовании и промышленности" 12 ноября 2019 г.

Компьютерные технологии и биология



Инженерная биология/Синтетическая биология

International Genetically Engineered Machine Competition (iGEM)





https://2019.igem.org/Team:Moscow







Randy Rettberg, Tom Knight

Rewiring Cells

How a handful of MIT electrical engineers pioneered synthetic biology.

Компьютерные технологии и биология



"Top-down"





Инженерные принципы

- Разделение дизайна и производства (decoupling)
- Стандартизация
- Абстракция
- Автоматизация/Роботизация



habr Публикации Новости Пользователи Хабы Компании Стать автором

• m1rko 9 мая 2019 в 16:17

Разработка белков в облаке с помощью Python и Transcriptic или Как создать любой белок за \$360

Автор оригинала: Brian Naughton

The Ethics of Synthetic Biology and Emerging Technologies

https://bioethicsarchive.georgetown.edu/pcsbi/sites/default/files/news/PCSBI-Synthetic-Biology-Report-12.16.10.pdf



Аналогии компьютеров и живых клеток

- Software boots into hardware
- Genetic code boots into cells





2010 *Mycoplasma mycoides* JCVI-syn1.0



Craig Venter



Первый искусственный геном (~\$40 млн)

Human	~20,000-25,000						
Escherichia coli (K12 strain)	~4500						
Syn 1.0	901						
Mycoplasma genitalium*	525						
Syn 3.0	473						

2016 Syn3.0



Минимальный геном

От программирования компьютеров к программированию жизни



	۰																				
			•																		
8886	000	16666	eee	leeee	20000	0000	9999	9996	9999	9999	888	9996	1888	6666	seee	eee	1000	000	9996	999	9999
1111		1111					111	1111	111	1111	111		111	1111						111	1111
2222	22222	2222	222	2223	2222	2222	2222	2222	222	2222	222	2222	222	2223	2222	2223	222	222	222	222	2222
3333	33333	33	3333	13333	3333	3333	333	3333	333	3333	333	3333	1333	3333	1333	3333	1333	3333	333	333	3333
4444	44484	4444	4414	4444	1444	1444	1444	****	444	6666	444	4444	1444	444	1444	444	1444	444	1444	444	4444
55	555	5555	8555	18555	15555	5555	5555	5555	555	5555	555	5555	555	5555	5555	5555	555	5555	5555	555	5555
666	66666	6666	6666	6664				6666	6666	6666	666	6666	666	6666	5666	6664		6661		666	6666
7773	77 77	7777	7777	777	\overline{m}	m	m	7777	777	7777	777	7777	777	777		777	\overline{m}	m	m	777	7777
		9899	9999				RRR	RRRR	RRR		888		1999	9999	1999				2888	RRR	RRRR



Ядро ОС Linux ~100 Мб











Синтетическая биология: конструирование живого

А.А. Грешнова, кандидат биологических наук Г.С. Глухов, кандидат физико-математических наук А.К. Шайтан

Мы научились создавать генно-модифицированные организмы, организмы с отредактированным геномом. Но модификации и редактирование — еще не конструирование с нуля, из стандартизированных обычно связан с постепенным усложнением. Каждый год элементов по известным правилам. Инженерия живых появляются более сложные электронные устройства, более систем — следующая цель современных биологических наук.

Еще одна инженерная дисциплина

Ход научно-технического развития человечества можно описать простой формулой «от познания — к созданию». Открытия фундаментальной науки пробуждают интерес к разработке на их основе новых устройств, машин, технологий. Так, открытый Гансом Эрстедом алюминий преобразовал металлургию и промышленность, благодаря катализаторам Карла Циглера и Джулио Натты стало возможным современное производство пластмасс. Открытие радиоволн привело к созданию радиосвязи, открытие свойств полупроводников к созданию современных компьютеров. При этом в областях знаний, которые принято называть техническими, прогресс

Синтетическая бактерия — Mycoplasma laboratorium JCVI-syn3.0

32



Хроматин



Number of cells in human body: ~ $30 * 10^{12}$

Total DNA length: 2.2* 30 * 10¹² = 66 billion km



Sun to Earth: 150 million km

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

(2 nm * 2 nm* 2*0.34 nm * 3.23 *10⁹)^{1/3}= **2** μm

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

 $R_G = rac{\sqrt{N} \, l}{\sqrt{6}}$ = 135 µm



Ncleosome core structure



Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Nature 1997, 389 (6648), 251–260.





"Plastic" model of a nucleosome https://github.com/molsim/nuclLEGO

Nucleosome compositional variability



Armeev, G. A.; Gribkova, A. K.; Pospelova, I.; Komarova, G. A.; **Shaytan, A. K.** Linking Chromatin Composition and Structural Dynamics at the Nucleosome Level. *Current Opinion in Structural Biology* **2019**, *56*, 46–55.

Nucleosome structural variability



Armeev, G. A.; Gribkova, A. K.; Pospelova, I.; Komarova, G. A.; **Shaytan, A. K.** Linking Chromatin Composition and Structural Dynamics at the Nucleosome Level. *Current Opinion in Structural Biology* **2019**, *56*, 46–55.

Виды моделирования



Использование

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".

- Press release
- Popular information
- Advanced information



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.

More on Martin Karplus



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.

Have a look at Michael Levitt's photo gallery



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 Laureate Arieh Warshel

Integrative modeling approaches



Interpreting experimental data with low information content

Integrative models



H2A/H2B/H3/H4 RNAP ★ - DNasel-hypersensitive sites



Transcription through nucleosomes

Gaykalova, D. A. et al.. PNAS 112, E5787–E5795 (2015).



Unwrapping by chaperones

Valieva, M. E. *et al.*. *Nat. Struct. & Mol. Biol.* **23**, 1111–1116 (2016).



Centromeric nucleosomes

Shaytan, A. K. *et al. Nucleic Acids Research* **45**, 9229–9243 (2017)



Octamer deformation during remodeling

Hada, A et al. Cell Reports 2019, 28 (1), 282-294.e6.

Виды моделирования





Typical biomolecular force field





Применение термостата

Potential Energy -

Current Position

0.05

150

Time, fs

0.1

E 0.125

Length, 0.12

σ Bor 0.115

0.13

0.11

200



All-atom MD simulations of nucleosomes



250-350 thousand of atoms Box 15x15x(15-20) nm

- Force field: AMBER12SB + parmbsc1 + CUFIX + TIP3P
- 150 mM NaCl, TIP3P water model
- GROMACS2019 (GPU)
- Truncated octahedron simulation box, 2 nm distance to the box walls



Lomonosov-2



GROMACS GPU



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





169387 atoms



267450 atoms

45 ns/day 1.36 using 4 nodes microsec/month using 4 nodes





All-atom MD simulations scaling



Раздел Compute	
Общее число узлов	1487
Центральный процессор	Intel Haswell-EP E5-2697v3, 2.6 GHz, 14 cores
Графический ускоритель	NVidia Tesla K40M
Объем оперативной памяти на узле	64 GB
Основная сеть	Infiniband FDR
Сеть І/О	Infiniband FDR
Управляющая сеть	Gigabit Ethernet
Операционная система	CentOS 7



Lomonosov-2

```
#!/bin/bash -1
 1
   #SBATCH -t 72:00:00
 2
3 #SBATCH -p compute
4 #SBATCH -J jobname
  #SBATCH -o ogmx.%j
 5
6 #SBATCH -e eqmx.%j
7 #SBATCH -N 10
   #SBATCH --ntasks-per-node=2
 8
   #SBATCH --cpus-per-task=6
9
10
   \# N - number of nodes,
11
   # --ntasks-per-node - amount of MPI tasks to run on one node
12
   # -- cpus-per-task - amount of MP threads per one MPI task
13
   export OMP NUM THREADS=7
14
15
   mpirun -np $(($SLURM JOB NUM NODES * 2)) gmx mpi mdrun -ntomp $OMP NUM THREADS -gputasks 00 -pme cpu -nb gpu -deffnm $1
16
17
```

Equilibrium MD simulations of nucleosomes and what we can learn from them



Equilibrium MD simulations of nucleosomes and what we can learn from them

MD simulations of 1kx5 NCP, alpha—sat DNA, original full tails symmet

Histones H3 Histones H4 Histones H2A Histones H2B Min groove ARG DNA AT pairs GC pairs O5'DNA chain L

Time:



Dynamics smoothed with a 100 ns window

Функционально важные динамические моды



Перемещение нуклеосом (Позиционирование нуклеосом)

Отворачивание ДНК

Динамика хвостов гистонов

Spontaneous DNA unwrapping

MD simulations of H2A.Z containing NCP, 601 DNA, tails truncated

Histones H3 Histones H4 Histones H2A.Z Histones H2B Min groove ARG DNA AT pairs GC pairs O5'DNA chain I





Spontaneous DNA unwrapping

MD simulations of H2A.Z containing NCP, 601 DNA, tails truncated, sr



Dynamics smoothed with a 100 ns window

DNA rewrapping simulations

MD simulations of 31z0 NCP 20 bp DNA unwrapped, 601 DNA, tails tru

Histones H3 Histones H4 Histones H2A Histones H2B Min groove ARG DNA AT pairs GC pairs O5'DNA chain I





DNA rewrapping simulations



DNA rewrapping simulations



How do nucleosomes move?

What perturbations in nucleosome structure are needed for them to move?





Twist-defect propagation hypothesis

Spontaneous diffusion ATP-dependent nucleosome remodeling

ATP-dependent nucleosome remodeling by SNF2



Li, M.; Xia, X.; Tian, Y.; Jia, Q.; Liu, X.; Lu, Y.; Li, M.; Li, X.; Chen, Z. Mechanism of DNA Translocation Underlying Chromatin Remodelling by Snf2. *Nature* **2019**, *567* (7748), 409–413. <u>https://doi.org/10.1038/s41586-019-1029-2</u>.

Is internal dynamics of the histone octamer important?



Concept of nucleosome plasticity

1. Bilokapic, S.; Strauss, M.; Halic, M. Structural Rearrangements of the Histone Octamer Translocate DNA. *Nature Communications* **2018**, *9* (1)

2. Sinha, K. K.; Gross, J. D.; Narlikar, G. J. Distortion of Histone Octamer Core Promotes Nucleosome Mobilization by a Chromatin Remodeler. *Science* **2017**, *355* (6322), eaaa3761.

3. Hada, A.; Hota, S. K.; Luo, J.; Lin, Y.; Kale, S.; Shaytan, A. K.; Bhardwaj, S. K.; et al. Histone Octamer Structure Is Altered Early in ISW2 ATP-Dependent Nucleosome Remodeling. *Cell Reports* **2019**, *28* (1), 282-294.e6

Octamer plasticity



Octamer plasticity



3lz0_ntm: H2A alpha2 helices CA-atom positions in XY-plane





3lz0_ntm_30unw: H2A alpha2 helices CA-atom positions in XY-plane





symmetry axis

(dyad axis)

Shaytan, A. K.; Xiao, H.; Armeev, G. A.; Wu, C.; Landsman, D.; Panchenko, A. R. Hydroxyl-Radical Footprinting Combined with Molecular Modeling Identifies Unique Features of DNA Conformation and Nucleosome Positioning. *Nucleic Acids Research* **2017**, *45* (16), 9229–9243. <u>https://doi.org/10.1093/nar/gkx616</u>.

Enhancing sampling: biased dynamics



Define a collective variable (CV)

$$d = f(\{x_i, y_i, z_i\})$$

Biasing potentials

Harmonic restraint and constant force $V_b = \frac{k}{2}(d-d_0)^2 + m*(d-d_0)$

Moving restraint

$$V_b(\vec{d},t) = \frac{1}{2}\kappa(t)(\vec{d}-\vec{d_0}(t))^2$$

Adiabatic-bias MD



$$V(\rho(t)) = \begin{cases} \frac{K}{2} \left(\rho(t) - \rho_m(t)\right)^2, & \rho(t) > \rho_m(t) \\ 0, & \rho(t) \le \rho_m(t), \end{cases}$$
$$\rho(t) = \left(d(t) - d_0\right)^2$$
$$\rho_m(t) = \min_{0 \le \tau \le t} \rho(\tau) + \eta(t)$$



Tetramer under stress



Metadynamics



Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci U S A. 2002 Oct 1;99(20):12562-6.



$$V(\vec{s,t}) = \sum_{k\tau < t} W(k\tau) \exp\left(-\sum_{i=1}^{d} \frac{(s_i - s_i(q(k\tau)))^2}{2\sigma_i^2}\right)$$
$$\bar{F}(\vec{s}) = -\frac{1}{t_{\rm sim} - t_{\rm diff}} \int_{t_{\rm diff}}^{t_{\rm sim}} V_{\rm bias}(\vec{s}, t) dt + C$$

Estimated energy profile

Well-tempered metadynamics

$$W(k\tau) = W_0 \exp\left(-\frac{V(s(q(k\tau)), k\tau)}{k_B \Delta T}\right)$$

$$V(\vec{s,t} \to \infty) = -\frac{\Delta T}{T + \Delta T}F(\vec{s)} + C$$

M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia and M. Parrinello. *PLUMED: a portable plugin for free energy calculations with molecular dynamics*, Comp. Phys. Comm. 180, 1961 (2009),





Metadynamics simulations with multiple walkers

Metadynamics simulations



RNA polymerase (RNAP) is a processive molecular motor capable of generating forces of 25-

4.1 pN-nm



Bancaud, A.; Wagner, G.; Silva, N. C. e; Lavelle, C.; Wong, H.; Mozziconacci, J.; Barbi, M.; Sivolob, A.; Cam, E. L.; Mouawad, L.; et al. Nucleosome Chiral Transition under Positive Torsional Stress in Single Chromatin Fibers. *Molecular Cell* **2007**, *27* (1), 135–147 Sivolob, A.; De Lucia, F.; Alilat, M.; Prunell, A. Nucleosome Dynamics. VI. Histone Tail Regulation of Tetrasome Chiral Transition. A Relaxation Study of Tetrasomes on DNA Minicircles. *J. Mol. Biol.* **2000**, *295* (1), 55–69.

Tetramer under torsional stress



"Reversosome"





Metadynamics



Acknowledgements

Dr. Grigory Armeev Anastasiia Kniazeva



Yunona Pospelova Anna Gribkova



Supported by HPC computing resources at Lomonosov Moscow State University.



grant 18-74-10006

Спасибо за внимание!!!