

Computational Biophysics

PHYS 462/562

Winter 2015-2016
TR 11-12:20pm, Disque 919

Prof. Luis Cruz
ccruz@drexel.edu

1

Preliminaries 1.

- Main purpose is to explain the physics behind computer simulations of many body systems
- Need to explain the “black box” packages out there, not to learn how to run them
- Depending on time and experience, there could be some actual programming and computer experimentation, but it is not the focus

2

Preliminaries 2.

- We will cover mostly the classical many-body problem
- Will do MD and some aspects of MC
- Topics will take off from the treatment of high-density mono atomic substances (liquids) and progress to polymers and proteins
- Time permitting: a birds-eye view of Computational Neuroscience (1-2 last lectures)

3

Preliminaries 3.

- Best way to make use of this course:
 - Take notes
 - i.e. scribble down your thoughts throughout lecture
 - Read chapter summaries
 - Go over the lecture slides
 - “experiment” with the course software
 - Ask questions
 - Be curious!

4

Root of the problem

- The 2+ body problem does not have an explicit analytical solution
- There are only a few idealized problems that have explicit solutions
- Real materials and situations demand heavy approximations to theory

5

Role of biophysical simulations

Role of simulations as a bridge between:

- models and theoretical predictions
- models and experimental results

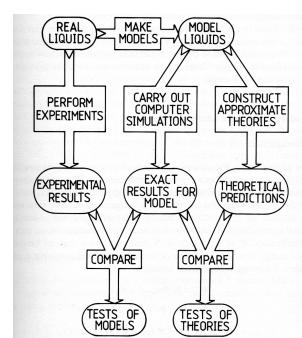


Fig. 1.2 The connection between experiment, theory, and computer simulation.

6

Caveats

- Simulations do not provide understanding, only numbers
- We have to infer conclusions based on these numbers
- These numbers contain statistical error
- Require a lot of computing time and costly computer resources

7

Opportunities

- All aspects of the behavior are open for inspection
- Can attack a problem considering all of its details *without* approximation
- Simulations can be used as predictive tools, but require careful preparation of the model and interpretation of the results
- Computer hardware has been getting better with time

8

Hardware Advances: Then...



- 1980s:
- single-user workstations
 - and CRAY mainframe

9

Now - #1 (2009) of top 500



Jaguar (CRAY XT) is the primary system in the ORNL Leadership Computing Facility (OLCF).

The XT5 partition contains 224,256 compute cores in addition to dedicated login/service nodes. Each compute node contains two hex-core AMD Opteron processors, 16GB memory, and a SeaStar 2+ router.

(www.nccs.gov - Nov. 2009)

10

Now - #1 (2015) of top 500

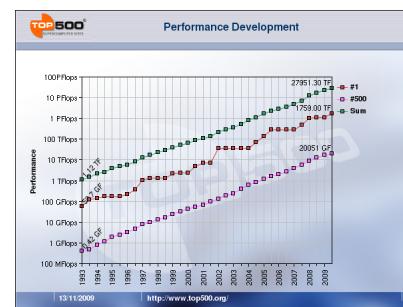


Tianhe-2 (China).

Uses Intel Xeon E5-2692 12C, 2.2GHz and Intel Xeon Phi.
3,120,000 cores
33.86 petaflop/s
17,800 kW
1,024,000 GB Memory

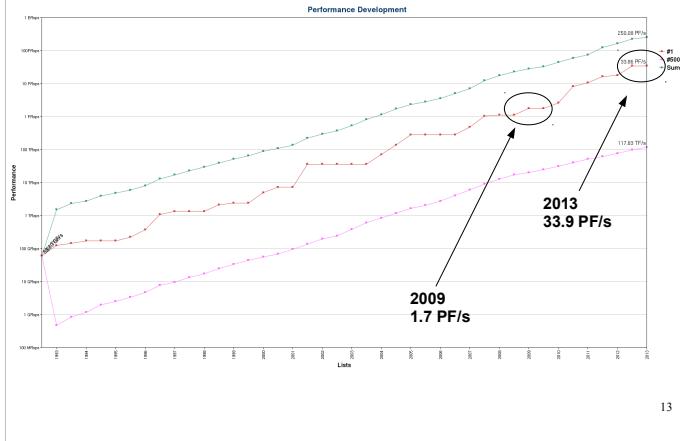
11

Ever increasing performance

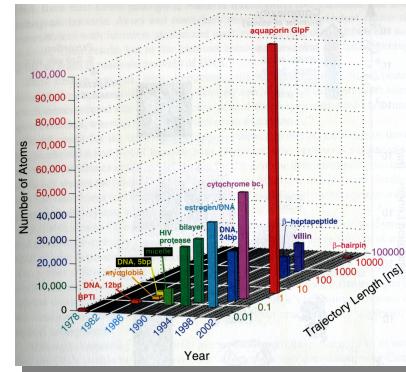


12

Ever increasing performance



Growth in time



- Evolution of molecular dynamics simulations with respect to system sizes and simulation lengths

Table 1.2. The evolution of molecular mechanics and dynamics.

Period	System and Size ^a	Trajectory Length [ns]	CPU Time/Computer ^b
1973	Dineoleptide (GpC) in vacuum (8 flexible dihedral angles)	—	—
1977	BPTI (17 residues, 88 atoms)	0.01	—
1983	DNA, vacuum, 12 & 24 bp (754/1530 atoms)	0.09	several weeks each Vax 780
1984	Grasp, vacuum (12 residues, 161 atoms)	0.15	—
1985	Myoglobin, vacuum (1423 atoms)	0.30	50 days VAX 11/780
1985	DNA 5 bp (25 atoms)	0.50	20 hrs Cray X-MP
1989	Phospholipid Micelle (n = 7,000 atoms)	0.10	—
1992	HIV protease (55,000 atoms)	0.10	Cray Y-MP
1997	Estrogen/DNA (36,000 atoms, multipoles)	0.10	HP-735 (8) 22 days
1998	DNA 24 bp (25 atoms, PME)	0.50	1 year, SGI Challenge
1998	β -heptapeptide in methanol (n = 5000/9000 atoms)	200	8 months, SGI Challenge (3) 4 months, Cray T3D/proc. Cray T3D/proc.
1999	Villin headpiece (36 residues, 12,000 atoms)	1000	—
1999	Ac _n complex in phospholipid bilayer (n = 100,000 atoms, cutoffs)	1	75 days, 64 450-MHz- proc, Cray T3E
2001	C-terminal β -hairpin of protein C (77 residues, 1,000 atoms)	38000	~ 8 days, 5000- Folded state, Cray T3E
2002	β -hairpin, channel protein in lipid membrane (10,189 atoms, PME)	5	30 hrs, 500 proc, LeMieux terascle system; 50 days, 32 proc, Linux (Athlon)

^aThe examples for each period are representative. The first five systems are modeled in vacuum and the last one in a lipid bilayer. The size of the system in molecular dynamics (MD) is the system size for the β -heptapeptide [39] reflect two temperature-dependent simulations. See text for definitions of abbreviations and further entry information.

^bThe 38 μ s β -hairpin simulation in 2001 represents an ensemble (or aggregate) dynamics simulation, as accumulated over several short runs, rather than one long simulation [40].

The simulation length is given as possible estimates for the vacuum DNA, heptapeptides,

β -hairpin, and channel protein simulations [41, 39, 40, 42] were kindly provided by M. Levin, W. van Gunsteren, V. Pandé, and K. Schulten respectively.

15

Growth of system size vs year

The examples for each period are representative. The first five systems are modeled in vacuum and the last one in a lipid bilayer. The size of the system in molecular dynamics (MD) is the system size for the β -heptapeptide [39] reflect two temperature-dependent simulations. See text for definitions of abbreviations and further entry information.

The simulation length is given as possible estimates for the vacuum DNA, heptapeptides,

β -hairpin, and channel protein simulations [41, 39, 40, 42] were kindly provided by M. Levin, W. van Gunsteren, V. Pandé, and K. Schulten respectively.

14. 1. Biomolecular Structure and Modeling: Historical Perspective

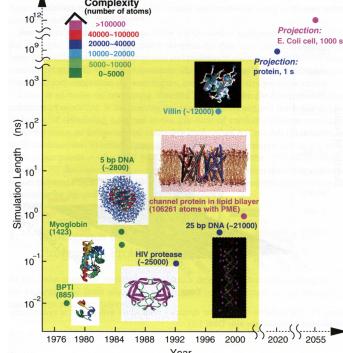
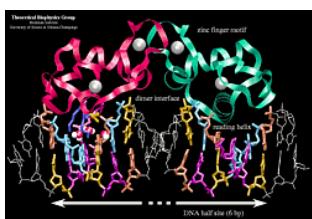


Figure 1.2. The evolution of molecular dynamics simulations with respect to simulation lengths (see also Table 1.2 and Figure 1.1). The data points for 2020 and 2055 represent extrapolations from the 1977 BPTI [56] and 1998 villin [70, 69] simulations, assuming a computational power increase by a factor of 10 every 3–4 years, as reported in [34].

16

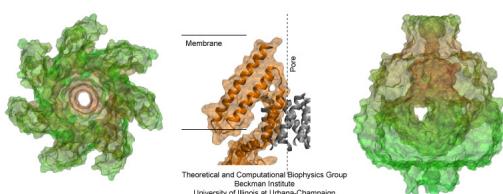
Ex. 1. DNA-dimer interaction



Conformational change –
a bend in the DNA

17

Ex. 2. Mechanosensitive channels



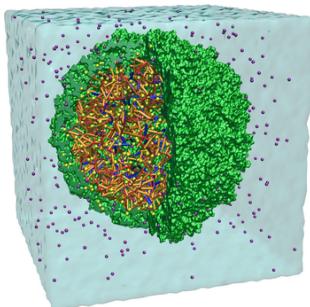
Mechanosensitive channels are membrane proteins that open and close in response to mechanical forces produced by sound, gravity or osmotic pressure, among other mechanical stimuli. In the open state, these proteins allow passage of ions across the membrane, thus generating an ionic current that eventually becomes an electrical signal (mechanotransduction) or that simply helps, for instance, in the regulation of cell volume.

18

Ex. 3. Tobacco Mosaic Virus

Viruses are very small intracellular parasites that invade the cells of virtually all known organisms.

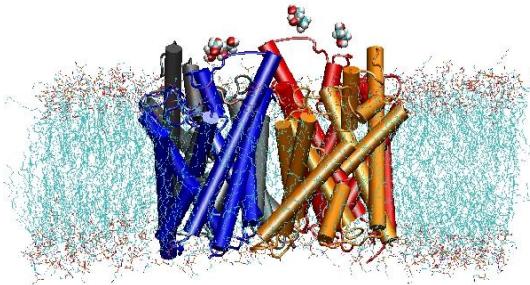
The complete STMV structure is stable, while the STMV capsid alone is not.



Theoretical and Computational Biophysics Group
Beckman Institute
University of Illinois at Urbana-Champaign

19

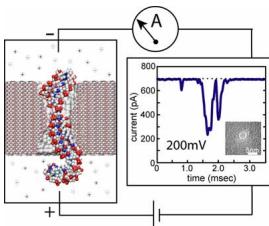
Ex. 4. Water through aquaporins



Aquaporins are membrane water channels that play critical roles in controlling the water contents of cells.

20

Ex. 5. DNA permeation

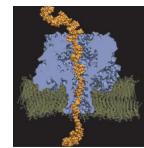


Electric detection of individual DNA molecules with a nanopore.

21

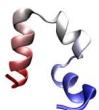
Ex. 6. DNA through membrane

By measuring the translocation of DNA through alpha-hemolysin, a membrane protein with a narrow pore, researchers discovered that directed single stranded DNA moves much faster when entering the pore 3' end first, rather than 5' end first.



22

Ex. 7. Villin Headpiece Folding



Fragment of the Villin protein, an actin binding protein. Folds independently from the full protein, stabilized by hydrophobic interactions.

23

Formalism

- Because our treatment will be classical,
- Can write down a classical Hamiltonian,
- With corresponding KE and PE
- Extra terms will be added to account for higher levels of complexity
 - Particle interactions (soft, hard, long-range)
 - Geometry (bonds, angles)
 - vibrations

24

Hamiltonian

System is specified by coordinates and momenta:

$$q = (q_1, q_2, \dots, q_N) \quad p = (p_1, p_2, \dots, p_N)$$

With the hamiltonian

$$H(q, p) = K(p) + V(q)$$

Where the kinetic energy is given by

$$K = \sum_{i=1}^N \sum_{\alpha} \frac{p_{i\alpha}^2}{2m_i}$$

25

Hamiltonian

And the potential energy is given by

$$V = \sum_i v_1(r_i) + \sum_i \sum_{j>i} v_2(r_i, r_j) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3(r_i, r_j, r_k) + \dots$$

First term is due to external fields. Second term is pair-wise interactions, usually expressed in terms of distances between particles. Third term is the potential energy due to triplets of particles.

Third term accounts usually for about 10% of interaction energy, but for reasons of computational efficiency is usually absorbed into the second term.

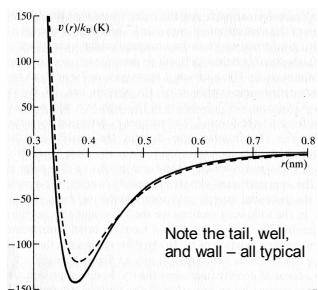
$$V \approx \sum_i v_1(r_i) + \sum_i \sum_{j>i} v_2^{eff}(r_j)$$

26

Experiment vs model

- Solid line is an experimental estimate of two argon atom interaction
- Dashed line is from a Lennard Jones model potential

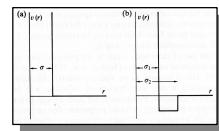
$$v^{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$



27

Other potentials

$$v^{HS}(r) = \begin{cases} \infty & (r < \sigma) \\ 0 & (\sigma \leq r) \end{cases}$$



$$v^{SW}(r) = \begin{cases} \infty & (r < \sigma_1) \\ -\epsilon & (\sigma_1 \leq r < \sigma_2) \\ 0 & (\sigma_2 \leq r) \end{cases}$$

- Oversimplified step potentials could very well provide a lot of science depending on the problem
- They contain advantages on the computation speed

28

Step wise potentials

- Step potentials can indeed yield information about transitions and differences between solid and fluid phases

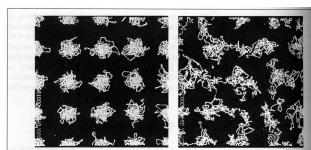
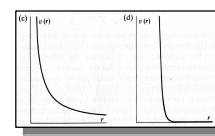


Fig. 7.3: Molecular graphics representation of the solids provided by 3D hard spherical particles in the solid (left) and fluid (right) phase. Reproduced from Alder, B. J. and T. F. Wainwright 1959. Studies in Molecular Dynamics I. General Method. Journal of Chemical Physics 31: 459-466.

29

Other potentials

- Soft-sphere potentials provide next step of complexity.
- “hardness” of potential increases with increasing v



$$v^{SS}(r) = \epsilon (\sigma/r)^v = a r^{-v}$$

30

Ionic contribution

- Long-range Coulombic interactions have to be included separately

$$v^{ions}(r_{ij}) = \frac{z_i z_j}{4\pi \epsilon_0 r_{ij}}$$

31

Molecular systems

- When considering molecules, can use algorithms from atom-atom interactions, but need to define points of interaction
- Bond lengths can be considered as constant, discarding high freq. (low amplitude) motion

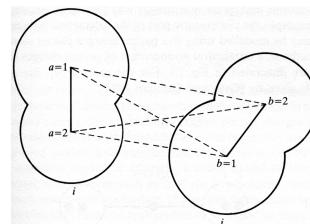


Fig. 1.6 An atom-atom model of a diatomic molecule.

32

Molecular systems

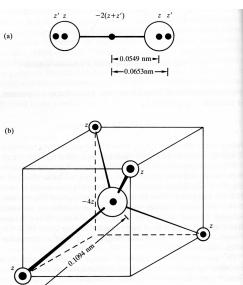


Fig. 1.7 Partial charge models. (a) A one-charge model for N_2 . There is one charge at the bond centre, two at the position of the nitrons, and two more displaced beyond the nuclei. Typical values are (in units of the magnitude of the electronic charge) $z = +5.2366$, $z' = -4.0695$, giving $Q = 1.1671 \times 10^{-19} C$. (b) A three-dimensional representation of a partial charge model for a molecule. There is one charge at the centre; and four others at the positions of the hydrogen nuclei. A typical value is $z = 0.143$ giving $Q = 3.77 \times 10^{-19} C m^3$ [Kirklin, Makii, and Klein 1981].

33

Molecular systems

- Points of interaction need not coincide all with nuclei positions
- Fictitious points could also have “partial” charges to account for multipolar interactions

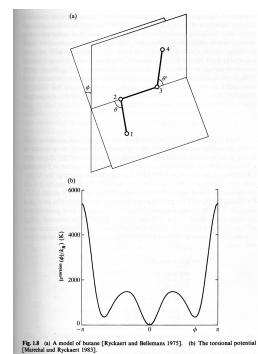


Fig. 1.8 (a) A model of butane [Ryckaert and Bellemans 1977]. (b) The torsional potential [Mortaj and Ryckaert 1981].

34

2013 Chemistry Nobel Prize

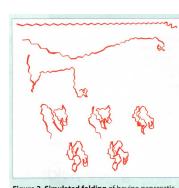
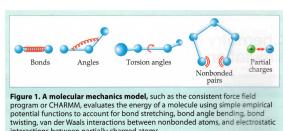


Figure 2. Simulated folding of bovine pancreatic trypsin inhibitor. Each of the protein's 38 amino-acid side chains was accommodated as a dummy atom (not shown) that carries some of the main backbone. (Adapted from ref. 5.)

Physics Today, December 2013, pp. 13-16

35