Molecular Modeling of Biomolecules: How can GPUs Advance Research?

Jeffery B. Klauda

Lipid Gel & Ripple Phases

Drug Binding to Lipid Membranes
Laboratory of Molecular & Thermodynamic Modeling

Energy-based Research

- Geological Modeling of Hydrates
- CO₂ Storage
- DOE-Fossil Energy

Biomolecular/Membrane Research

- Lipid Membranes
- Transmembrane Proteins
- Peripheral Proteins

Hydrotrpoe Stabilizing nanodroplet of oil
Research Methods & Design

- Research Design

  * ab initio QM

  Molecular Simulations & Models

  Macroscopic Properties

  QM=Quantum Mechanics
Molecular Dynamics (MD)

- Governing Equations
  - Newton’s Laws of Motion
    \[ f_i = m_i a_i = m_i \ddot{r}_i \] where \( i \) is a molecule or atom
  - Force drives the motion of a system
    \[ f_{ij} = -\frac{\partial w_{ij}}{\partial \vec{r}_{ij}} \] where \( w \) is the inter- and intramolecular potential

- Accurate force fields are required for realistic simulations

Lennard-Jones Potential (van der Waals/non-bonded forces)
\[ w(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \]

- Why use MD?
  - Probe biomolecules at atomic resolution without introduction of artificial labels or expensive equipment
  - Aid experiments (diffraction, NMR, spin labeling) in determining what is measured\(^1\text{-}^4\)
  - Dynamical understanding of membrane function

**Biomolecular Force Field**

\[
V(\mathbf{R}) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{cross \ UB} K_{UB} (r_{1,3} - r_{1,3}^0)^2 + \sum_{improper} K_im (1 - \cos(2\phi))
\]

\[
+ \sum_{dihedrals} \left[ \sum_j K_{\phi,j} \left( 1 + \cos(n_j \phi - \delta_j) \right) \right] + \sum_{\text{nonbonded pairs}} \epsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right] + \sum_{\text{nonbonded pairs}} \frac{q_i q_j}{r_{ij}}
\]

- Many terms to describe intra- and intermolecular interactions
- The $r^{-12}$, $r^{-6}$ and $r^{-1}$ terms are the most computationally demanding terms

**Typical System Sizes Required for Simulations**

<table>
<thead>
<tr>
<th></th>
<th>Liquid Simulation (small molecule)</th>
<th>Lipid Membrane (lipid only)</th>
<th>Protein with Lipid Membrane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
<td>8-10 nm$^3$</td>
<td>125-700 nm$^3$</td>
<td>500-1500 nm$^3$</td>
</tr>
<tr>
<td># atoms</td>
<td>3,000-5,000</td>
<td>20,000-70,000</td>
<td>50,000-150,000</td>
</tr>
</tbody>
</table>

- Efficient codes that run these large systems is **crucial**
MD Simulation Programs

- CHARMM (Chemistry at HArvard Macromolecular Mechanics)\(^1\)
  - Came out of Prof. Martin Karplus’ group at Harvard
  - A comprehensive code that contains many cutting-edge techniques in addition to traditional simulation techniques
  - GPUs: CHARMM/OpenMM interface

- NAMD (Scalable Molecular Dynamics)\(^2\)
  - NIH-supported code from Prof. Klaus Schulten’s group (UI-UC)
  - Less focus on functionality and more parallel scalability
  - GPUs & MICs: directly available in NAMD code

- GROMACS (Groningen University)\(^3\)
  - Development spawned from Prof. Herman Berendsen group
  - European analog to CHARMM
  - GPUs: Built-in functionality for GPUs

- Other Commonly Used Programs
  - AMBER, TINKER, DESMOD, and LAMMPS

---

\(^1\)www.charmm.org \hspace{1cm} \(^2\)www.ks.uiuc.edu/Research/namd/
\(^3\)www.gromacs.org
Computational Equipment/Resources

- Local UMD Computational Clusters (UMD/IT)
  - **Deepthought**: Dell Linux cluster with ~4000 cores
  - **Deepthought2**: Dell Linux cluster with 9200 cores and 40 nodes with dual GPUs

- XSEDE (NSF Supported)
  - **Stampede**: Dell Linux cluster with 100,000+ Cores (10 PetaFlops) (All have MIC & some with GPUs)
  - **K20**
Lipids

- Complex biomolecules
  - Contain a fatty acid chains and head group
  - Classified into 8 categories\(^1\)

\[^1\text{Fahy et al. J. Lipid. Res. 46: 839 (2005).}\]
Lipid Self-Assembly

- Self-assembly into phases depending on water content

- Lower concentration of lipid form spherical micelles
- Higher concentrations form bilayer structures (common in cells)

1S.A. Sefran. Statistical Thermodynamics of Surfaces, Interfaces and Membranes (Addison-Wesley, NY, 1994).
Lipid Bilayer Phases

Phase Transitions

• Fluid or liquid crystalline ($L_\alpha$) bilayer phases are most common and have high chain disorder.

• Certain lipids go through a pretransition as the temperature is lowered to a ripple phase with interdigitation.

• This short pretransition ($\sim 10^\circ C$) leads to an ordered gel phase ($L_\beta$).

• Introduction of cholesterol leads to a liquid ordered phase (sometimes existing as a lipid raft).

• Can MD simulations on all-atom force fields see this? Requires a significant amount of computational time.

---

MD Simulations of DMPC/DPPC Bilayers

- **Details of the Simulation**
  - Force field and composition: CHARMM36\(^1\) and 50% DMPC
  - Program & #atoms: NAMD with 16,704
  - Deepthought2 with GPUs for 300ns of simulation time

- **Benchmarks**

  **CPU-only**

<table>
<thead>
<tr>
<th>#Cores</th>
<th>hr/ns</th>
<th>ns/day</th>
<th>%Eff</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1.99</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1.06</td>
<td>22.7</td>
<td>94.0</td>
</tr>
<tr>
<td>60</td>
<td>0.74</td>
<td>32.4</td>
<td>89.7</td>
</tr>
<tr>
<td>80</td>
<td>0.63</td>
<td>38.4</td>
<td>79.6</td>
</tr>
</tbody>
</table>

- Two K20m GPUs on a single node results in 1.24 hr/ns or 19.4 ns per day!
- More significant speedup for larger systems

- **DMPC/DPPC at 20°C**

  - Two weeks to get 300ns with GPUs and this took over two months on older generation HPC.

\(^1\)Klauda, J.B. et al. *JPCB*. **114**: 7830 (2010).
Gel Phase DMPC/DPPC Bilayers Formation

50/50% of DMPC/DPPC at 20°C (300ns)

• Starts with a L$_\alpha$ phase that shortly transitions to a ripple-like phase before gelling
• Chain alignment and tilt between leaflets exists in agreement with experiment
Ripple Phase DMPC/DPPC Bilayers Formation

25/75% of DMPC/DPPC at 25°C (300ns)

- Starts with a $L_\alpha$ phase that slowly transitions to a ripple-like phase
- Leaflet interdigitation and lipid buckling promotes the ripple-like phase.
Drug Binding to Lipid Bilayers

- Drug Partitioning in Lipids

  - Many drugs and toxins are lipophilic that is they like lipids over water phases
  - Precursor to full transport into/out of cell via membrane transport proteins
  - Alternating Access Model of substrate transport with transmembrane proteins

MD Simulations of Ethidium Binding to a Lipid Bilayer

- Details of the Simulation
  - Force field and composition: CHARMM36\(^1\) and POPC/POPG bilayer (simple bacterial model)
  - Program & #atoms: NAMD with 30,000
  - Deepthought2 with GPUs for 200ns of simulation time

- Partitioning into Membrane

  **0.012% Ethidium**

  **0.047% Ethidium**

  - Benchmark: 18 ns per day with GPU+CPU
  - Quickly sample partitioning and dynamics of antibiotic binding to lipid membranes with the use of GPU+CPU

\(^1\)Kluda, J.B. et al. *JPCB.* **114**: 7830 (2010).
Movie of Ethidium Binding

0.047% Ethidium in water with POPC/POPG Bilayer (200ns)

- Quickly determine the extent of drug binding to the membrane
- Ethidium binds to the hydrophobic/philic interface but cannot easily go across the bilayer without the aid of a transport protein
Summary

• MD simulations at the atomic level can probe a wide range of self-assembly and biological problems

• MD simulations require a high amount of computational resources that benefit from GPUs

• Most MD software has been optimized with CUDA programming

• Lipid phase changes are complex but our use of CPU+GPU on DT2 has allowed us to probe gel and ripple phase formation

• Our C36 lipid force field\(^1\) accurately represents the phase transition temperature of PC lipid mixtures

• Many drugs and toxin partition into lipid membranes and fat cells of the body

• The use of GPU nodes has allowed us to quickly determine the tendency of drugs to bind to membranes and their location

• All of these projects are currently being applied to protein-related research in drug transport and understanding of diseases

Acknowledgments

Outside University of Maryland

Richard Pastor (NIH/NHLBI)
Rick Venable (NIH/NHLBI)
Will Prinz (NIH/NIDDK)
Klaus Gawrisch (NIH/NIAAA)
Mary Roberts (Boston College)
Wonpil Im (University of Kansas)
Alex MacKerell (UMB)
Katie Henzler-Wildman (WU-St. Luis)
Bryan Berger (Lehigh U.)
Hirsh Nanda (NIST)
Yuji Sugita (Riken/Japan)
Cor Peters (TU-Delft/Petroleum Institute)

University of Maryland

Brent Rogaski (M.S. 2010/Industry)
Pushkar Pendse (Ph.D./Postdoc)
Viviana Monje (Ph.D. Student)
Pouyan Khakbaz (Ph.D. Student)
Xiaohong Zhuang (Ph.D. Student)
Joe Lim (undergrad/MIT)
Diana Villanueva (undergrad/GSK)
Chris Boughter (undergrad-phys)
Sylvia Kang (undergrad-BioE)
John Daristotle (undergrad)
Sook Wong (undergrad)
Ryan Konas (undergrad)
Connor Welch (undergrad)
Francis Bacarisas (undergrad-BioE)
Prof. Mikhail Anisimov (ChBE/IPST)

Funding and Computational Resources

NSF CAREER (MCB-1149187)
NSF/BIO (DBI-1145652)
National Institutes of Health (intramural)
Anton at NRBSC/PSC [PSCA000009P]
XSEDE [TG-MCB100139]/OIT (UMD)