COMP564: Introduction to Molecular Dynamics Simulations

Jérôme Waldispühl
School of computer science
McGill University

(based on slides from M. Smaoui)
Aims of the lecture

• Introduce you to what Molecular Dynamics (MD) is and why it is used

• Familiarize you with the basic terminology and algorithms of the molecular dynamics method

• How to analyze MD results
Outline

• Background
  – What is MD (and what is it not)
  – How is it useful?

• Modeling atomic systems
  – PDB files

• Newtonian physics
  – Motion equations

• Verlet algorithm
  – Trajectories of atoms in space

• MD data analysis
MD Vision

• “In the real world, this could eventually mean that most chemical experiments are conducted inside the silicon of chips instead of the glassware of laboratories. Turn off that Bunsen burner; it will not be wanted in ten years.”

- The Economist, reporting on the work of the 1998 Chemistry Nobel Prize Awardees
Simulating reality

- Many Physically-Based Simulations model easily observable real world phenomena.

- Molecular Dynamics Simulations model things too small for us to observe directly.
Nobel of Chemistry 2013

The Nobel Prize in Chemistry 2013

Martin Karplus
Université de Strasbourg, France and Harvard University, Cambridge, MA, USA

Michael Levitt
Stanford University School of Medicine, CA, USA

Arieh Warshel
University of Southern California, Los Angeles, CA, USA
Background

• Calculate how a system of particles evolves in time
• Consider a set of atoms with positions and velocities, and the potential energy function of the system
• Predict the next positions of particles over some short time interval by solving Newtonian mechanics
Modeling atomic systems

<table>
<thead>
<tr>
<th>ATOM</th>
<th>1</th>
<th>N</th>
<th>MET A 217</th>
<th>-5.308</th>
<th>4.213</th>
<th>-16.145</th>
<th>1.00</th>
<th>3.22</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOM</td>
<td>2</td>
<td>CA</td>
<td>MET A 217</td>
<td>-5.667</td>
<td>3.781</td>
<td>-14.799</td>
<td>1.00</td>
<td>75.03</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>3</td>
<td>C</td>
<td>MET A 217</td>
<td>-6.837</td>
<td>4.599</td>
<td>-14.260</td>
<td>1.00</td>
<td>23.11</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>4</td>
<td>O</td>
<td>MET A 217</td>
<td>-7.215</td>
<td>5.617</td>
<td>-14.839</td>
<td>1.00</td>
<td>41.31</td>
<td>O</td>
</tr>
<tr>
<td>ATOM</td>
<td>5</td>
<td>CB</td>
<td>MET A 217</td>
<td>-4.466</td>
<td>3.908</td>
<td>-13.861</td>
<td>1.00</td>
<td>65.23</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>6</td>
<td>CG</td>
<td>MET A 217</td>
<td>-3.764</td>
<td>5.254</td>
<td>-13.951</td>
<td>1.00</td>
<td>34.45</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>7</td>
<td>SD</td>
<td>MET A 217</td>
<td>-2.568</td>
<td>5.503</td>
<td>-12.625</td>
<td>1.00</td>
<td>32.14</td>
<td>S</td>
</tr>
<tr>
<td>ATOM</td>
<td>8</td>
<td>CE</td>
<td>MET A 217</td>
<td>-1.665</td>
<td>6.920</td>
<td>-13.245</td>
<td>1.00</td>
<td>74.11</td>
<td>C</td>
</tr>
<tr>
<td>ATOM</td>
<td>9</td>
<td>HA</td>
<td>MET A 217</td>
<td>-5.963</td>
<td>2.744</td>
<td>-14.853</td>
<td>1.00</td>
<td>31.03</td>
<td>H</td>
</tr>
<tr>
<td>ATOM</td>
<td>10</td>
<td>HB2</td>
<td>MET A 217</td>
<td>-4.801</td>
<td>3.768</td>
<td>-12.845</td>
<td>1.00</td>
<td>21.54</td>
<td>H</td>
</tr>
<tr>
<td>ATOM</td>
<td>12</td>
<td>HG2</td>
<td>MET A 217</td>
<td>-3.248</td>
<td>5.314</td>
<td>-14.897</td>
<td>1.00</td>
<td>72.00</td>
<td>H</td>
</tr>
<tr>
<td>ATOM</td>
<td>13</td>
<td>HG3</td>
<td>MET A 217</td>
<td>-4.508</td>
<td>6.035</td>
<td>-13.898</td>
<td>1.00</td>
<td>50.01</td>
<td>H</td>
</tr>
<tr>
<td>ATOM</td>
<td>14</td>
<td>HE1</td>
<td>MET A 217</td>
<td>-2.324</td>
<td>7.532</td>
<td>-13.843</td>
<td>1.00</td>
<td>41.20</td>
<td>H</td>
</tr>
<tr>
<td>ATOM</td>
<td>15</td>
<td>HE2</td>
<td>MET A 217</td>
<td>-1.290</td>
<td>7.499</td>
<td>-12.414</td>
<td>1.00</td>
<td>0.21</td>
<td>H</td>
</tr>
<tr>
<td>ATOM</td>
<td>16</td>
<td>HE3</td>
<td>MET A 217</td>
<td>-0.838</td>
<td>6.582</td>
<td>-13.851</td>
<td>1.00</td>
<td>41.45</td>
<td>H</td>
</tr>
</tbody>
</table>
MD run
Range & Time-scale of folding

- Local Motions (0.01 to 5 Å, $10^{-15}$ to $10^{-1}$ s)
  - Atomic fluctuations
  - Sidechain motions
  - Loop motions
- Rigid Body Motions (1 to 10Å, $10^{-9}$ to 1s)
  - Helix motions
  - Domain motions
  - Subunit motions
- Large-Scale Motions (> 5Å, $10^{-7}$ to $10^{4}$ s)
  - Helix coil transitions
  - Dissociation/Association
  - Folding and Unfolding
Newton’s First Law

An object either remains at rest or moves at a constant velocity, unless acted upon by an external force.

The force fields that affects motion of atomic particles are the following:

– Covalent bonds
– Van der Waals Potential
– Electrostatic potential
– Solvent model
Covalent bonds

\[ E_{bonded} = E_{cov}^r + E_{cov}^\theta + E_{cov}^\phi \]

\[ E_{cov}^r = \mathcal{K}_r \cdot (r - r_{eq})^2 \]

\[ E_{cov}^\theta = \mathcal{K}_\theta \cdot (\theta - \theta_{eq})^2 \]

\[ E_{cov}^\phi = \sum_{n=1}^{3} \frac{\nu_n}{2} (1 + \cos(n\phi - \gamma)) \]
Van der Waals

• Atoms with no net electrostatic charge will still tend to attract each other at short distances, as long as they don’t get too close.

• Once the atoms are close enough to have overlapping electron clouds, they will repel each other with astounding force

• Lennard-Jones Potential
Lennard-Jones potential

\[ E_{vdw} = E_m \left[ -\left( \frac{2R_m}{r_{ij}} \right)^{12} + 2 \left( \frac{2R_m}{r_{ij}} \right)^6 \right] \]

<table>
<thead>
<tr>
<th>Atom</th>
<th>( E_m )</th>
<th>( R_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.12</td>
<td>1.85</td>
</tr>
<tr>
<td>O</td>
<td>0.20</td>
<td>1.60</td>
</tr>
<tr>
<td>H</td>
<td>0.02</td>
<td>1.00</td>
</tr>
<tr>
<td>N</td>
<td>0.16</td>
<td>1.75</td>
</tr>
<tr>
<td>P</td>
<td>0.20</td>
<td>2.10</td>
</tr>
<tr>
<td>S</td>
<td>0.20</td>
<td>2.00</td>
</tr>
</tbody>
</table>
Implementation of LJ potential

**Truncation**: interactions set to zero for distances greater than the cutoff.

**SHIFT cutoff method**: modifies the entire potential energy surface such that at the cutoff distance the interaction potential is zero.

**The SWITCH cutoff method**: This method tapers the interaction potential over a predefined range of distances.
Electrostatic potential

- Opposite Charges Attract
- Like Charges Repel
- The force of the attraction is inversely proportional to the square of the distance
- Coulomb potential

\[ E_{\text{electrostatic}} = \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon_1 r_{ij}} \]
Electrostatic interactions are computed explicitly if distance between charges is below a threshold $R_{\text{cut}}$. Otherwise, a general potential is used.
**Explicit model**: Water molecules are modelled individually. It is more accurate and model the hydration layer at the surface of proteins but the number of molecules and thus the computational become prohibitive.

**Implicit model** Water effect is modelled as an external potential. Fast but less accurate.
Classical mechanics

To model the motion of atoms as a function of time, we need to solve for the following variables:

• Position \( (r) \)
• Momentum \( (m + v) \)
• Charge \( (q) \)
• Bond information (which atoms, bond angles, etc.)

With this information, we can deduce the force that will be exerted on the atoms of the system.
Verlet Algorithm

What do we know?

• State of the system at time $t$:
  – Position $x(t)$
  – Speed $x'(t)$

• Newton’s equation of motion:
  $$m \cdot x''(t) = F(x(t)) = -\nabla(x(t))$$

What do we want?

• State of the system at time $t + \Delta t$:
  – Position $x(t + \Delta t)$
  – Speed $x'(t + \Delta t)$
Verlet Algorithm

• Approximation using Taylor series:
  \[ x(t + \Delta t) = x(t) + x'(t) \cdot \Delta t + \frac{1}{2} \cdot x''(t) \cdot \Delta t^2 + O(\Delta t^3) \]
  \[ x(t - \Delta t) = x(t) - x'(t) \cdot \Delta t + \frac{1}{2} \cdot x''(t) \cdot \Delta t^2 + O(\Delta t^3) \]

• Sum up the two equations:
  \[ x(t + \Delta t) + x(t - \Delta t) = 2 \cdot x(t) + x''(t) \cdot \Delta t^2 + O(\Delta t^3) \]

• Plug in the Newton’s equation:
  \[ x(t + \Delta t) + x(t - \Delta t) = 2 \cdot x(t) - \frac{\nabla(x(t))}{m} \cdot \Delta t^2 + O(\Delta t^3) \]
Verlet Algorithm

New position calculated as:

\[ x(t + \Delta t) = 2 \cdot x(t) - x(t - \Delta t) - \frac{\nabla(x(t))}{m} \cdot \Delta t^2 \]

Speed can be obtained with:

\[ v(v + \Delta t) = \frac{x(t + \Delta t) - x(t - \Delta t)}{2 \cdot \Delta t} \]

**Advantages:**
- easy to implement
- low storage requirement

**Disadvantages:**
- precision
- velocity not explicit
- initial configuration (2 time points)
Velocity Verlet Algorithm

Approximation using Taylor series:

\[ x(t + \Delta t) = x(t) + v(t) \cdot \Delta t + \frac{1}{2} \cdot a(t) \cdot \Delta t^2 \]

Get velocity at mid-time step:

\[ v \left( t + \frac{\Delta t}{2} \right) = v(t) + a(t) \cdot \frac{\Delta t}{2} \]

Get acceleration from position:

\[ a(t + \Delta t) = -\frac{1}{m} \cdot \nabla V(x(t + \Delta t)) \]

Get velocity at next time step:

\[ v(t + \Delta t) = v \left( t + \frac{\Delta t}{2} \right) + a(t + \Delta t) \cdot \frac{\Delta t}{2} \]

Advantage: Position, velocity and acceleration are synchronized.
MD run
MD Analysis

RMSD
1AKI, Backbone
MD Analysis

RMS Fluctuation

![Graph showing RMS Fluctuation with residue number on the x-axis and RMSF (nm) on the y-axis.]