Computational Chemistry - MD Simulations

P. Ojeda-May

Department of Chemistry/HPC2N, Umeå University,

901 87, Sweden.

June 17, 2016
# Table of contents

1. **Basics on MD simulations**
   - Force field
   - Ensembles
   - Beyond classical MD
     - Accelerated MD
     - Umbrella sampling
     - String method
     - Coarse graining
     - Alchemical method

2. **Abisko and Kebnekeise**

3. **Using GROMACS at HPC2N**
   - Setting up the system, minimization, solvation, equilibration, production and analysis.
Background
Background

- Physicist currently working on Chemistry.
- Experience with Supercomputers and HPC Clusters.
Background

- Physicist currently working on Chemistry.
- Experience with Supercomputers and HPC Clusters.
- Code developer of Chemistry at Harvard Molecular Mechanics (CHARMM) package:
  - Carr-Parrinello Molecular Dynamics
  - Multiple time steps algorithms
  - Modified Neglect of Diatomic Overlap (MNDO) QM/MM approach
  - Hybrid MPI/OpenMP approaches for QM/MM simulations
  - QM/MM-Cutoff methods.
Background

- Physicist currently working on Chemistry.
- Experience with Supercomputers and HPC Clusters.
- Code developer of Chemistry at Harvard Molecular Mechanics (CHARMM) package:
  - Carr-Parrinello Molecular Dynamics
  - Multiple time steps algorithms
  - Modified Neglect of Diatomic Overlap (MNDO) QM/MM approach
  - Hybrid MPI/OpenMP approaches for QM/MM simulations
  - QM/MM-Cutoff methods.
Simulations time scale

Figure: Accuracy w.r.t. time scale for different modeling approaches.
Computer simulation of protein folding

Michael Levitt* & Arieh Warshel*

Department of Chemical Physics, Weizmann Institute of Science, Rehovoth, Israel

A new and very simple representation of protein conformations has been used together with energy minimisation and thermalisation to simulate protein folding. Under certain conditions, the method succeeds in 'renaturing' bovine pancreatic trypsin inhibitor from an open-chain conformation into a folded conformation close to that of the native molecule.

protein, in this case myoglobin, was based on the packing of cylinders supposed to represent α-helices. The method was not implemented on a computer and cannot be applied more generally to other proteins not built entirely from helices.

Here we tackle the problem differently. First, we simplify the representation of a protein by averaging over the fine details. This is done both to make the calculations much more efficient and also to avoid having to distinguish between many

Figure : Nature, 253 (1975).
Current MD simulations

The Nobel Prize in Chemistry 2013

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".

Figure: Taken from: http://www.nobelprize.org.
Current MD simulations

EXPERIMENT  --------  THEORY

SIMULATIONS
Application of Parallel algorithms

Proteins

Clays

Figure: AdK enzyme in water.

Figure: Clay [JPC C, 118, 1001 (2014)].
Basics on MD simulations
Abisko and Kebnekeise
Using GROMACS at HPC2N

Application of Parallel algorithms

---

**Figure**: Asphalt research.

Molecular dynamics study of interfacial mechanical behavior between asphalt binder and mineral aggregate

Guangji Xu, Hao Wang
Department of Civil and Environmental Engineering, Rutgers University, Piscataway, NJ 08854, USA

---

**Figure**: Asphalt [Const. Build. Mat., 121, 246 (2016)].

---

*Fig. 3. Representative model for the MD simulations and schematic of tensile simulation.*
Newton’s equation

\[ \mathbf{F} = -\nabla U \quad \text{Newton’s Law (1687)} \quad (1) \]

solution of this equation requires the knowledge of an array of particles’ positions and velocities

\[ \mathbf{X} = (x_1^1, x_2^1, x_3^1, x_1^2, x_2^2, x_3^2, \ldots, x_1^N, x_2^N, x_3^N) \quad (2) \]

\[ \mathbf{V} = (v_1^1, v_2^1, v_3^1, v_1^2, v_2^2, v_3^2, \ldots, v_1^N, v_2^N, v_3^N) \quad (3) \]
Force fields

Figure: Taken from: http://www.lpwchem.org/force-field-development/
Force fields

\[
V = \sum_{\text{bonds}} \frac{1}{2} k_{\text{bonds}} (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_{\text{angle}} (\theta - \theta_0)^2
\]

\[
+ \sum_{\text{torsions}} \sum_j V_j (1 + \cos j \phi)
\]

\[
+ \sum_{\text{Coulomb}} \frac{q_i q_j}{r_{ij}} + \sum_{\text{VdW}} \left\{ 4 \epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right\}
\]

- Proteins and Hydrocarbons: GROMOS, OPLS-AA, AMBER, CHARMM.
- Clays: CLAYFF
- Coarse-graining: MARTINI
Force fields: Energy surface

Figure: Energy surface described by $V = \sin(x) \times \cos(x)$
Water models

Figure: 3-5 sites water models. Taken from:
http://www1.lsbu.ac.uk/water/water_models.html
### Water models

<table>
<thead>
<tr>
<th>Water Model</th>
<th>Slope 1.9–9.6 ps ($\times 10^{-9}$ m$^2$ s$^{-1}$)</th>
<th>Slope 4.0–20.0 ps ($\times 10^{-9}$ m$^2$ s$^{-1}$)</th>
<th>Temperature (K)</th>
<th>$D$ (25 °C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIP3P original$^a$</td>
<td>5.88$^d$ (0.10)$^e$</td>
<td>5.87$^d$ (0.09)$^e$</td>
<td>301.4$^d$ (1.7)$^e$</td>
<td>5.67</td>
</tr>
<tr>
<td>TIP3P original$^b$</td>
<td>5.59 (0.06)</td>
<td>5.59 (0.08)</td>
<td>297.0 (0.9)</td>
<td>5.65</td>
</tr>
<tr>
<td>TIP3P modified$^a$</td>
<td>5.92 (0.09)</td>
<td>5.92 (0.11)</td>
<td>301.2 (1.8)</td>
<td>5.73</td>
</tr>
<tr>
<td>TIP3P modified$^b$</td>
<td>5.83 (0.07)</td>
<td>5.85 (0.08)</td>
<td>299.2 (1.0)</td>
<td>5.78</td>
</tr>
<tr>
<td>SPC original$^a$</td>
<td>4.39 (0.05)</td>
<td>4.40 (0.06)</td>
<td>301.0 (1.7)</td>
<td>4.22</td>
</tr>
<tr>
<td>SPC original$^b$</td>
<td>4.22 (0.06)</td>
<td>4.24 (0.08)</td>
<td>298.6 (1.1)</td>
<td>4.20</td>
</tr>
<tr>
<td>SPC refined$^a$</td>
<td>4.49 (0.08)</td>
<td>4.48 (0.08)</td>
<td>301.0 (1.8)</td>
<td>4.30</td>
</tr>
<tr>
<td>SPC refined$^b$</td>
<td>4.26 (0.07)</td>
<td>4.24 (0.10)</td>
<td>297.7 (1.2)</td>
<td>4.26</td>
</tr>
<tr>
<td>SPC/E original$^a$</td>
<td>2.90 (0.06)</td>
<td>2.89 (0.08)</td>
<td>300.4 (1.9)</td>
<td>2.75</td>
</tr>
<tr>
<td>SPC/E original$^b$</td>
<td>2.78 (0.04)</td>
<td>2.77 (0.06)</td>
<td>298.2 (1.4)</td>
<td>2.76</td>
</tr>
</tbody>
</table>

$^a$ Nonbonded list 1 (see Methods). $^b$ Nonbonded list 2 (see Methods). $^c$ Temperature of the MD simulation. $^d$ Mean values. $^e$ Standard deviations. $^f$ Self-diffusion coefficients adjusted to 25 °C, using the slope 4.0–20.0 ps.

**Figure**: See for details: JPC A, **105**, 9954 (2001).
Figure: 20 natural amino acids. Taken from: goo.gl/YrYvvv
Protein systems

Figure: PDB information of AdK.

Figure: Structure of yeast AdK.
Periodic boundary conditions (PBC)

The systems we can study with MD simulations are tiny compared to real experimental setups ($10^{23}$ particles).

Figure: PBCs and minimum image convention [Allen & Tildesley, Comp. Sim. of Liquids]
The electrostatic energy for a periodic system can be written as

\[ E = \frac{1}{2} \sum_{m \in \mathbb{Z}^3} \sum_{i,j=1}^{N} \left( \sum_{m} \frac{q_i q_j}{|\mathbf{r}_{ij} + m\mathbf{L}|} \right) \]

where \( \mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j \), \( m \) refers to the periodic images. Primed summation means \( i = j \) interaction is excluded for \( m = 0 \). \( q_x \) is the partial charge on atom \( x \).

\(^{1}\text{Adv. Polym. Sci., 185, 59 (2005)}\)
The electrostatic energy for a periodic system can be written as \(^1\),

\[
E = \frac{1}{2} \sum_{m \in \mathbb{Z}^3} \sum_{i,j=1}^{N} \frac{q_i q_j}{|r_{ij} + mL|} 
\]

where \(r_{ij} = r_i - r_j\), \(m\) refers to the periodic images. Primed summation means \(i = j\) interaction is excluded for \(m = 0\). \(q_x\) is the partial charge on atom \(x\). The potential is splitted such that,

\[
\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r} 
\]

Electrostatic interactions: Ewald method

The electrostatic energy for a periodic system can be written as 

$$E = \frac{1}{2} \sum_{m \in \mathbb{Z}^3} \sum_{i,j=1}^{N} \frac{q_i q_j}{|r_{ij} + mL|}$$

(5)

where $r_{ij} = r_i - r_j$, $m$ refers to the periodic images. Primed summation means $i = j$ interaction is excluded for $m = 0$. $q_x$ is the partial charge on atom $x$. Giving rise to the total energy:

$$E = E^{(r)} + E^{(k)} + E^{(s)} + E^{(d)}$$

(6)

---

Electrostatic interactions

\[ E^{(r)} = \frac{1}{2} \sum_{m \in \mathbb{Z}^3} \sum_{i,j=1}^{N} q_i q_j \frac{\text{erfc}(\alpha |r_{ij} + mL|)}{|r_{ij} + mL|} \]  

\[ E^{(k)} = \frac{1}{2V} \sum_{k \neq 0} \frac{4\pi}{k^2} e^{k^2/4\alpha^2} |\bar{\rho}(k)|^2 \]  

\[ E^{(s)} = -\frac{\alpha}{\sqrt{\pi}} \sum_i q_i^2 \]  

\[ E^{(d)} = \frac{2\pi}{(1 + 2\epsilon')V} \left( \sum_i q_i r_i \right)^2 \]
Electrostatic interactions: Cutoff methods

The electrostatic energy for a periodic system can be written as\(^2\),

\[
E = \frac{1}{2} \sum_{m \in \mathbb{Z}^3} \sum_{i,j=1}^{N} \sum_{i,j=1}^{N} \frac{q_i q_j}{|r_{ij} + mL|} \tag{11}
\]

where \( r_{ij} = r_i - r_j \), \( m \) refers to the periodic images. Primed summation means \( i = j \) interaction is excluded for \( m = 0 \).

Can we truncate the interactions up to \( r = R_c \)?

\[
E = \frac{1}{2} \sum_i \sum_j \frac{1}{r_{ij}} + \Phi \tag{12}
\]

Electrostatic interactions: NaCl lattice

\[ E_i(R_c) = \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \quad (13) \]

\[ E^{Mad} = -3.495129...q^2/a \]

Figure: NaCl lattice. (source:goo.gl/Fa7tcL)

Figure: Single ion energy for NaCl lattice. (Wolf et al., JCP, 110, 8256 (1999))
Electrostatic interactions: NaCl lattice

Figure: NaCl lattice. (source:goo.gl/Fa7tcL)

\[ E_i(R_c) = \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \quad (13) \]

\[ E^{Mad} = -3.495129 \ldots q^2 / a \]

Figure: Energy convergence upon charge neutralization. (Wolf et al., JCP, 110, 8256 (1999))
Isotropc Periodic Sum method: NaCl lattice

\[ \epsilon_{ij}^{\text{IPS}}(r_{ij}) = \begin{cases} \epsilon_{ij}(r_{ij}) + \phi_{ij}(r_{ij}) & \text{if } r_{ij} \leq R_c \\ 0 & \text{otherwise} \end{cases} \]  

(14)

\( \epsilon_{ij} \) is the Coulombic term and \( \phi_{ij} \) is the long-range IPS correction whose operational expression is given by,

\[ \phi_{ij}(r_{ij}) = \frac{q_i q_j}{R_c} \left[ \sum_{k=1}^{6} b_{2k} \left( \frac{r_{ij}}{R_c} \right)^{2k} \right] \]  

(15)

**Figure**: Single ion energy using different cutoff methods. [JCP, 140, 164106 (2014), JCP, 122, 044107 (2005)]
Integration of Newton’s equation

We now know the force field and we know the law of motion:

\[ \mathbf{F} = m \mathbf{a} - \nabla U \quad \text{Newton's Law} \quad (16) \]

We need to integrate this equation, here we use the leap-frog scheme [Hockney, 1970],

\[ \mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t + \frac{1}{2} \delta t) \quad (17) \]

\[ \mathbf{v}(t + \frac{1}{2} \delta t) = \mathbf{v}(t - \frac{1}{2} \delta t) + \delta t \mathbf{a} \quad (18) \]

Velocities are updated according to,

\[ \mathbf{v}(t) = \frac{1}{2} \left( \mathbf{v}(t + \frac{1}{2} \delta t) + \mathbf{v}(t - \frac{1}{2} \delta t) \right) \quad (19) \]
Constraints

Collision of two diatomic molecules

**Figure:** Free collision.

**Figure:** SHAKE constraint.

See JCP, 112, 7919 (2000)
Constraints

Modern approaches to deal with constraints

**Figure:** ILVES method.

See *JCC*, **32**, 3039 (2011)
Techniques to speedup simulations

- MPI parallelization
- MPI+OpenMP parallelization
- Domain decomposition scheme
- Multiple communicators

\[
\text{do } i=1,\text{num}_\text{particles} \\
\text{x(i) = x(i) + f(i) } \cdot \text{dt} \\
\text{enddo}
\]
Ergodicity

\[ A_{\text{obs}} = \langle A \rangle_{\text{time}} \]
\[ = \langle A(\Gamma(t)) \rangle_{\text{time}} \]
\[ = \lim_{t_{\text{obs}} \rightarrow \infty} \int_{0}^{t_{\text{obs}}} A(\Gamma(t)) dt \]  
(20)

Figure: Coffee cup.
Statistical ensembles

- Microcanonical ensemble (NVE) partition function is [Allen & Tildesley, Comp. Sim. of Liquids],

\[ Q_{NVE} = \frac{1}{N!} \frac{1}{h^{3N}} \int \text{d}r \text{d}p \delta(\mathcal{H}(r, p) - E) \quad (21) \]

The thermodynamic potential is the negative of the entropy

\[ -S/k_B = -\ln Q_{NVE} \]

- In the case of the Canonical ensemble (NVT) the partition function is,

\[ Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \int \text{d}r \text{d}p \exp(-\mathcal{H}(r, p)/k_B T) \quad (22) \]

with thermodynamic potential \( A/k_B T = -\ln Q_{NVT} \).
Statistical ensembles

- Isothermal-isobaric ensemble (NPT) partition function is,

\[ Q_{NPT} = \frac{1}{N!} \frac{1}{h^{3N}} \frac{1}{V_0} \int dV \int dr dp \exp\left( -\frac{\mathcal{H}(r, p) + PV}{k_B T} \right) \]

the corresponding thermodynamic potential is

\[ G/k_B = -\ln Q_{NPT} \]

- Grand-canonical ensemble (\(\mu VT\)) partition function is,

\[ Q_{\mu VT} = \sum_N \frac{1}{N!} \frac{1}{h^{3N}} \exp\left( \frac{\mu N}{k_B T} \right) \int dr dp \exp\left( -\frac{\mathcal{H}(r, p)}{k_B T} \right) \]

the corresponding thermodynamic potential is

\[ -PV/k_B = -\ln Q_{\mu VT} \]
Thermostats

- NVE is obtained by solving NE.
- NVT can be achieved with the following thermostats: Berendsen, Velocity-rescaling, Nose-Hoover.

\[ H = \sum_{i=1}^{N} \frac{p_i}{2m_i} + U(r_1, r_2, \ldots, r_N) + \frac{p_\xi^2}{2Q} + N_f kT \xi \]  (25)

A better approach is Nose-Hoover chain.

- Using general and local thermostats.
- NPT can be simulated with Berendsen and Parrinello-Rahman methods.
Accelerated MD simulations

The original potential energy surface $V(r)$ is modified according to,

$$V^*(r) = \begin{cases} V(r), & V(r) \geq E, \\ V(r) + \Delta V(r), & V(r) < E. \end{cases}$$

(26)

Figure: Modified potential energy surface [JCP, 120, 11919 (2004)].
Accelerated MD simulations

the biasing term is,

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))}$$

(27)

Figure: Free energy landscape of Alanine dipeptide [JCP, 120, 11919 (2004)].
Umbrella sampling (US) simulations

The potential energy is modified as follows JCP, 23, 187 (1977):

\[ E^b(r) = E^u(r) + w_i(\xi) \]

with \( w_i(\xi) = K/2(\xi - \xi_{i \text{ ref}})^2 \)

For each window the free energy is given by,

\[ A_i(\xi) = -(1/\beta) \ln P_i^b(\xi) - w_i(\xi) + F_i \]
String method (SM) simulations

Define a set of collective variables $z_j$ and effective forces as follows

$$\frac{k}{T} \int_0^T (z_j - \theta_j(t))dt \sim \frac{\partial F(z)}{\partial z_j}$$

The free energy along the string is computed by PRB, 66, 052301 (2002),

$$F(z(\alpha)) - F(z(0)) = \int_0^\alpha \sum_{i=1}^N \frac{dz_i(\alpha')}{d\alpha'} \frac{\partial F(z(\alpha'))}{\partial z_i} d\alpha'$$
String method (SM) simulations

Figure: Free energy surface of Alanine dipeptide.
Coarse-grain simulations

Figure: Reduction of the degrees of freedom [Annu. Rev. Biophys., 42, 73 (2013)].
Alchemical simulations

![Thermodynamic cycle for binding of two protein ligands $L_1$ and $L_2$, [JCC, 30, 1692 (2009)].](image)

$$
\Delta \Delta G_{L_i \rightarrow L_j}^{bind} = \Delta G_{L_j}^{bind} - \Delta G_{L_i}^{bind} = \Delta G_{RL_i \rightarrow RL_j}^{prot} - \Delta G_{L_i \rightarrow L_j}^{solv} \quad (28)
$$

The Hamiltonian is modified according to,

$$
H = T_x + (1 - \lambda) V_0 + \lambda V_1 \quad (29)
$$
Alchemical simulations

Figure: Thermodynamic cycle for binding of two protein ligands $L_1$ and $L_2$, [JCC, 30, 1692 (2009)].

The free energy difference going from $\lambda = 0$ to $\lambda = 1$ is,

$$
\Delta G_{\lambda=0 \to \lambda=1} = \sum_{\lambda=0}^{1} -\frac{1}{\beta} \ln \langle \exp \left( -\beta (H(\lambda+\delta\lambda) - H(\lambda)) \right) \rangle 
$$

(30)
Abisko and Intro to Kebnekeise

Birgitte Brydsoe
GROMACS on GPU/Phi

GROMACS 5.0: Phi vs. Kepler K40 fastest GPU!

GROMACS 5.0 RC1 (ns/day) on K40 with Boost Clocks and Intel Phi 192K Waters Benchmark (CUDA 6.0)
GROMACS

- Setting up the system
- minimization
- solvation
- neutralization
- equilibration
- production
- analysis
GROMACS files

- *.gro, *.pdb (coordinates)
- *.top (topology)
- *.tpr (binary input file)
- *.mdp (parameter file for simulation)