CLASSICAL MOLECULAR DYNAMICS
NUMERICAL METHODS FOR THE SIMULATION OF BIOLOGICAL SYSTEMS

Chris Chipot

Laboratoire International Associé CNRS-UIUC,
Unité Mixte de Recherche n° 7565, Université de Lorraine

Beckman Institute for Advanced Science and Technology,
Department of Physics
University of Illinois at Urbana-Champaign
Molecular dynamics simulations

Learn to:

- set up your molecular assays
- run your simulations
- perform your own analyses

... and understand what you are doing

Run your own simulations after only one day of training
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
Supply the microscopic detail that experiment cannot capture;
Attain key thermodynamic properties of the system,
e.g., $\Delta G_{\text{adsorption}}, \Delta G_{\text{hydration}}$;
Acquire structural information,
e.g., refinement of three-dimensional structures,
Acquire dynamic information,
e.g., $c(t), D(z), \ldots$.

The massive increase of computational resources in the past twenty years has opened the way to the investigation of sizable molecular assemblies of chemical, physical and biological interests, over representative time scales.
Complementing experiment — The computational microscope
Ideally, for an $N$-particle chemical system,

$$i \frac{\partial}{\partial t} \left| \Psi(r^N; t) \right\rangle = \mathcal{H} \left| \Psi(r^N; t) \right\rangle$$

In practice, the full quantum treatment of the system is virtually impossible for obvious overwhelming computational reasons.

At the Hartree-Fock level, for $N$ atoms, the CPU investment grows as $\mathcal{O}(N^3)$.

One possible solution: Linear-scaling approaches, or *divide and conquer* growing as $\mathcal{O}(N)$. Currently, they are limited to small systems with a sampling of a few ps. Moreover, they are restrained to the Born-Oppenheimer approximation, and are significantly more costly than molecular mechanics.

The Born-Oppenheimer approximation: The motion of the electrons and that of the nuclei are dissociated;

The de Broglie wavelength associated to a given particle is substantially smaller than the intermolecular distances in a liquid;

Quantum effects can be globally neglected. It is, therefore, legitimate to turn to the classical equations of motion in molecular dynamics simulations

Time averages and ensemble averages coincide for ergodic systems,

\[
\lim_{t \to \infty} \overline{A}_t = \langle A \rangle
\]

McQuarrie, D. A. Statistical mechanics, Harper and Row, 1976
We turn to molecular dynamics simulations to access the atomic detail not always accessible to experiment, to gain dynamic information and to predict thermodynamic and kinetic properties.

Molecular dynamics simulations rely on the equations of classical mechanics and, therefore, are restrained to the motion of the nuclei. They are limited to the description of molecular assemblies that do not involve the creation or rupture of chemical bonds.

Ergodicity implies coincidence of ensemble and time averages — and, hence, of properties inferred from Monte-Carlo and molecular dynamics simulations.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
**PERIODIC BOUNDARY CONDITIONS**

Bridge the macroscopic, or the mesoscopic world to the microscopic one;

Can we relate the properties of a system containing a number of particles on the order of $N_A$ to those of a system containing only a few hundreds to a few thousands particles?

How can we handle edge effects?


CellBasisVector1  70  0  0
CellBasisVector2  0  70  0
CellBasisVector3  0  0  70

Periodic boundary conditions

Periodic boundary conditions

minimum-image convention: each individual particle in the cell interacts with the closest image of the remaining particles in the assay.

Conventionally, beyond the spherical truncation, the interaction potential is nil.

Conventionally, beyond the spherical truncation, the interaction potential is nil. Yet, it is possible to include a reaction-field term, \( V_{\text{elec}}(r) = \sum_{i<j} q_i q_j \left[ \frac{1}{r_{ij}} + \left( \frac{\epsilon_{\text{RF}} - 1}{2\epsilon_{\text{RF}} + 1} \right) \frac{r_{ij}^2}{R_{\text{cut}}^3} \right] \).

In principle, spherical truncation is acceptable for electrostatic interactions in $1/r^n$, with $n \geq 3$. For $n < 3$, an appropriate treatment is required — e.g., Ewald, or Ladd.

Build a simple molecular assay using VMD

— Load the topology and Cartesian coordinates, \( \{ \mathbf{x} \} \), of the biological object
— Solvate in a pre-equilibrated solvent cell

Periodic boundary conditions leave enough headspace when solvating to avoid periodicity-induced artifacts

Leach, A. R. Molecular modeling. Longman, 1996


Chipot, C. in Les Nanosciences: Nanotechnologies et nanophysique, Belin, 2006
Adding ions to a solvated protein-ligand complex

— At what salinity was the experiment carried out?
— What ions were used?
— What is the excess charge of the complex?
Building a membrane protein assay

- equilibrated membrane patch
- insert randomly additives
- insert membrane protein
- add water if necessary
- insert ions if necessary

in what lipids was the experiment conducted?

were additives present in the experimental assay?

allow extracellular loops to extend

Building a membrane protein assay with different lipids

1. Equilibrated pure-lipid patch
2. Random insertion of lipids
3. Embed membrane protein
4. Add aqueous environment
5. Adjust salinity

CHARMM-GUI provides a web-based graphical user interface to generate various molecular simulation systems and input files to facilitate and standardize the usage of common and advanced simulation techniques. Currently, CHARMM-GUI supports CHARMM, NAMD, GROMACS, AMBER, and OpenMM simulation programs mostly based on the CHARMM force fields.

CHARMM-GUI is powered by CHARMM, an academic research program used worldwide for macromolecular dynamics and mechanics (http://www.charmm.org). Its development began in the research group of Professor Martin Karplus at Harvard University and continues throughout the world with contributing developers. CHARMM performs standard molecular dynamics and energy minimization with the potential energy functions for proteins, nucleic acids, lipids, carbohydrates, and various small molecules. In addition, CHARMM can be used for various chemical and conformational free energy calculations with many types of restraints.

The CHARMM-GUI team hopes that the tools and materials offered here are useful and helpful for your research and education.
Periodic boundary conditions offer a convenient framework to bridge the microscopic world to the macroscopic one, whilst being emancipated from edge effects.

Spherical truncation is in general valid only for short-range interactions.

Switching is recommended to ensure continuous derivative across the cut-off sphere.

When constructing molecular assays, leave enough headspace to avoid spurious periodicity-induced artifacts.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
Synopsis of a molecular dynamics simulation

**Synopsis of a molecular dynamics simulation**

- **Read starting point data**: \((x, p_x)\)
  - Compute energy
  - Compute forces:
    - intramolecular
    - intermolecular
    - constraints

- **Integration of trajectories**

- **Intermediate averages**

- **Compute physical quantities**

- **Results, analysis of trajectories**:
  - thermodynamic properties
  - dynamic properties

**References**

Goal: Follow the trajectory of an ensemble of particles with time:

\[
\begin{align*}
    m_i \frac{d^2 x_i(t)}{dt^2} &= f_i \\
    f_i &= -\frac{\partial V(x)}{\partial x_i}
\end{align*}
\]

Rely upon a potential energy function capable of describing the chemical system:

\[
V(x) = \sum_i v_1(x_i) + \sum_i \sum_{j>i} v_2(x_i, x_j) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3(x_i, x_j, x_k) + \ldots
\]
Goal: Follow the trajectory of an ensemble of particles with time:

\[ m_i \frac{d^2 x_i(t)}{dt^2} = f_i \]

\[ f_i = -\frac{\partial V(x)}{\partial x_i} \]

Rely upon a potential energy function capable of describing the chemical system:

\[ V(x) = \sum_i v_1(x_i) + \sum_i \sum_{j>i} v_{\text{effective}}^e(x_i, x_j) \]

Require a numerical integration of the Newtonian equations of motion, utilizing an infinitesimal time-step, namely \( \delta t = 1-2 \times 10^{-15} \) s.

why are we limited to such low \( \delta t \)’s?

\[ v = 1/2 \pi c (k/\mu)^{1/2} \]

Typical timescales explored in molecular dynamics and the computational resource required to access them.

<table>
<thead>
<tr>
<th>Time</th>
<th>ps</th>
<th>ns</th>
<th>µs</th>
<th>ms</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decay</td>
<td>$10^{-12}$</td>
<td>$10^{-9}$</td>
<td>$10^{-6}$</td>
<td>$10^{-3}$</td>
<td>1</td>
</tr>
</tbody>
</table>

1 gigaflop 1 teraflop 100 teraflops 10 petaflops
The error associated to the propagator should be bound, that is,

$$\lim_{n_{\text{steps}} \to \infty} \left( \frac{1}{n_{\text{steps}}} \right)^{n_{\text{steps}}} \sum_{k=1}^{n_{\text{steps}}} \left| \frac{\mathcal{E}(k\delta t) - \mathcal{E}(0)}{\mathcal{E}(0)} \right| \leq \varepsilon_{\text{MD}}$$

which is true for a symplectic propagator.

Energy drifts stem from inaccuracies in the integration of the equations of motion.

These drifts can be in principle monitored by means of a shadow Hamiltonian.

$$\widetilde{\mathcal{H}}(\mathbf{x}, \mathbf{p}_x) = \mathcal{H}(\mathbf{x}, \mathbf{p}_x) + \frac{\delta t^2}{2} h(\mathbf{x}, \mathbf{p}_x) + \mathcal{O}(\delta t^4)$$

$$h(\mathbf{x}, \mathbf{p}_x) = \frac{1}{6} \mathbf{p}_x^\top m^{-1} \frac{\partial^2 \mathcal{V}(\mathbf{x})}{\partial \mathbf{x}^2} m^{-1} \mathbf{p}_x$$

$$- \frac{1}{12} [\nabla \mathcal{V}(\mathbf{x})]^\top m^{-1} \nabla \mathcal{V}(\mathbf{x})$$

Lagrangian formalism,

\[ \mathcal{L}(q, p) = \mathcal{T}(p_x) - \mathcal{V}(q) \]

\[
\begin{aligned}
\mathcal{T}(p_x) &= \sum_i \frac{1}{2} \frac{p_i^2}{m_i} \\
\mathcal{V}(x) &= \sum_i v_1(q_i) \\
&+ \sum_i \sum_{j > i} v_2(q_i, q_j) + \ldots
\end{aligned}
\]

Lagrange’s equation of motion,

\[
\frac{d}{dt} \left( \frac{\partial \mathcal{L}(q, p)}{\partial \dot{q}} \right) - \frac{\partial \mathcal{L}(q, p)}{\partial q} = 0
\]

Derivative of the Lagrangian,

\[
\frac{d\mathcal{L}(q, p)}{dt} = \sum_i \frac{\partial \mathcal{L}(q, p)}{\partial q_i} \dot{q}_i + \sum_i \frac{\partial \mathcal{L}(q, p)}{\partial \dot{q}_i} \ddot{q}_i
\]

Goldstein, H. Classical mechanics. Addison-Wesley, 1980
Lagrange’s equation of motion,

\[
\frac{d}{dt} \left( \frac{\partial \mathcal{L}(\mathbf{q}, \mathbf{p})}{\partial \dot{\mathbf{q}}} \right) - \frac{\partial \mathcal{L}(\mathbf{q}, \mathbf{p})}{\partial \mathbf{q}} = 0
\]

Newton’s equations of motion,

\[
\begin{align*}
    m_i \mathbf{a}_i(t) & = f_i \\
    \mathbf{f}_i & = \nabla_{(r_i)} \mathcal{L}(\mathbf{q}, \mathbf{p}) = -\nabla_{(r_i)} \mathcal{V}(\mathbf{x})
\end{align*}
\]

Hamilton’s equations of motion,

\[
\begin{align*}
    \dot{q}_i & = \frac{\partial \mathcal{H}(\mathbf{q}, \mathbf{p}_x)}{\partial p_i} \\
    \dot{p}_i & = -\frac{\partial \mathcal{H}(\mathbf{q}, \mathbf{p}_x)}{\partial q_i}
\end{align*}
\]
Use of massively parallel architectures (e.g., NUMA, or PC clusters) to spread the task over several processors;

Moore's law suggests longer and more affordable simulations;

Robust methodology, allowing the modeler to work in realistic conditions and apt thermodynamic ensembles.
Historical perspective

A turning point in computational structural biology

First molecular dynamics simulation applied to a small protein, BPTI, over 8 ps.

Historical perspective
In its simplest acceptation, molecular dynamics consists in solving the Newton equations of motion, using small time increments.

Choice of the time increment is dictated by the fastest vibration in the molecular assay.

Discretization errors are bounded provided that a symplectic integrator is used.

Current molecular dynamics simulations attain the millisecond timescale and the hundred-million-atom size scale.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
THE POTENTIAL ENERGY FUNCTION

Modeling a complex reality with a rudimentary toolkit

Freddolino, P. L.; Schulten, K. *Biophys. J.* 2009, 97, 2338-2347

villin headpiece
THE POTENTIAL ENERGY FUNCTION

Modeling a complex reality with a rudimentary toolkit
Case example of the macromolecular AMBER force field,

\[ V(x) = \sum_{\text{bonds}} k_r (r - r_0)^2 + \sum_{\text{valence angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \frac{1}{k_1^{1-4}} \sum_{\{i, j\} \in 1-4} \varepsilon_{ij} \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \frac{1}{k_1^{1-4}} \sum_{\{i, j\} \in 1-4} q_i q_j \frac{4\pi \epsilon \epsilon_0}{r_{ij}} \]

+ \sum_{\{i, j\} > 1-4} \varepsilon_{ij} \left[ \left( \frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \sum_{\{i, j\} > 1-4} q_i q_j \frac{4\pi \epsilon \epsilon_0}{r_{ij}}

Other macromolecular force fields — CHARMM, MM3, OPLS.


Parametrization of a concise set of chemically unique atoms;

Underlying hypothesis: The parameters describing a given atom are essentially local, and may, therefore, be ported to different chemical environments.
Bond-stretching potential

The potential energy function $V(r)$ as a function of bond length $r$ (Å) is shown. The potential is calculated using the MP2/3–21G*//MP2/3–21G* method. The graph exhibits a minimum at a specific bond length, indicating the equilibrium bond length.

Leach, A. R. Molecular modeling. Longman, 1996
Bond-stretching potential

\[ V(r) = k_r (r - r_0)^2 \]

Leach, A. R. *Molecular modeling.* Longman, 1996
Bond-stretching potential

The potential energy function

\[ k_r (r - r_0)^2 \]

\[ k_r (r - r_0)^2 \times \left[ 1 - k'_r (r - r_0) + k''_r (r - r_0)^2 \right] \]

Leach, A. R. Molecular modeling. Longman, 1996
Bond-stretching potential

\[ k_r (r - r_0)^2 \]

\[ D_0 \left[ e^{-\alpha (r - r_0)} - 1 \right]^2 \]

\[ k_r (r - r_0)^2 \times \left[ 1 - k'_r (r - r_0) + k''_r (r - r_0)^2 \right] \]

Leach, A. R. Molecular modeling. Longman, 1996
Torsional potential — description of a complex potential requires several terms,

\[ V_1(\phi) \propto \cos\phi \]
Torsional potential — description of a complex potential requires several terms,

\[ V_3(\phi) \propto \cos 3\phi \]

Leach, A. R. Molecular modeling. Longman, 1996
Torsional potential — description of a complex potential requires several terms,
Modeling the electrostatic contribution

Two distinct philosophies:

**AMBER:** The electrostatic potential is the *fingerprint* of the molecule

\[
V_{\text{reference}}(r_k) = \sum_A \frac{Z_A}{|r_k - R_A|} - \sum_i \sum_{\mu} \sum_{\nu} c_{\mu i}^* c_{\nu i} \left\langle \phi_{\mu} \left| \frac{1}{|r_k - r|} \right| \phi_{\nu} \right\rangle
\]

**CHARMM:** The electrostatic term is parameterized from intermolecular interactions

\[
E_{\text{interaction}}^{\text{QM}} = E_{\text{molecule:water}}^{\text{QM}} - E_{\text{molecule}}^{\text{QM}} - E_{\text{water}}^{\text{QM}}
\]


Charges derived from the molecular electrostatic potential,

\[ f(\{q_k\}) = \sum_{i=1}^{N_{\text{points}}} \left[ V_{\text{reference}}(\mathbf{r}_i) - \sum_{j=1}^{N_{\text{atoms}}} \frac{q_j}{r_{ij}} \right]^2 \]


Beyond the simple and limited monopole approximation,

\[ g(\{q_k\}, \{\mu_{k'}\}) = \sum_{i=1}^{N_{\text{points}}} \bigg[ V_{\text{reference}}(\mathbf{r}_i) - \left( \sum_j q_j \frac{1}{r_{ij}} - \sum_{j'} \mu_{j'} \cdot \frac{\mathbf{r}_{ij'}}{r_{ij'}^3} \right) \bigg]^2 = \min \]

Including intramolecular induction effects,

\[ h(\{q_k\}) = \sum_{k=1}^{N_{\text{conformations}}} w_k \sum_{i=1}^{N_{\text{points}}} \left( V_{k_{\text{reference}}}(\mathbf{r}_i) - \sum_j q_j \frac{1}{r_{ij}} \right)^2 = \min \]

Including intermolecular induction effects,

\[ \mathcal{H} = \mathcal{H}^0 - \sum_{\ell m} \sum_{\ell' m'} Q_{\ell m} f^{\ell m}_{\ell' m'} \langle \Psi | Q_{\ell' m'} | \Psi \rangle \]

Modeling van der Waals interactions

Leach, A. R. Molecular modeling. Longman, 1996
Modeling van der Waals interactions

\[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \]

Leach, A. R. Molecular modeling. Longman, 1996
Modeling van der Waals interactions

The potential energy function

\[ V(r) = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{C_{ij}}{r_{ij}^9} - \frac{D_{ij}}{r_{ij}^6} \]

Leach, A. R. Molecular modeling. Longman, 1996
Modeling van der Waals interactions

The potential energy function

\[ \psi(r) = A' e^{-B'r_{ij}} - \frac{C'_{ij}}{r_{ij}^6} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \]

\[ C_{ij} \frac{1}{r_{ij}^9} - D_{ij} \frac{1}{r_{ij}^6} \]

Leach, A. R. Molecular modeling. Longman, 1996
Transferability of parameters constitutes one of the key assumptions of macromolecular force fields.

The underlying idea of macromolecular force fields is a minimalist description of complex chemical objects, based on a limited number of terms — often ad hoc.

Parametrization of different macromolecular force fields rely on distinct philosophies.

Parameters are not interchangeable between different potential energy functions.
Synopsis of a force-field parametrization

- **determine bonded terms**
  - gas-phase QM calculations, experiment

- **determine electrostatic term**
  - gas-phase QM calculations

- **determine van der Waals term**
  - liquid simulations

- **determine torsional term**
  - gas-phase QM calculations

- **Force field**

If parameters are missing from the force field

Option 1. Parametrize the missing terms following the philosophy of the force field

If parameters are missing from the force field

Option 2. Turn to a more comprehensive force field, e.g., OPLS

\[
\begin{align*}
\text{CA (len=4):} & \quad (3.5500,0.0700) \quad n = 571 \\
& \quad (3.5500,0.0500) \quad n = \quad 3 \\
& \quad (3.5000,0.0800) \quad n = \quad 7 \\
& \quad (3.5500,0.0760) \quad n = \quad 14
\end{align*}
\]
If parameters are missing from the force field

Option 3. Resort to servers relying upon ad-hoc inferences and educated guesses of missing parameters

Water models

One of the commonest molecules — and one of the hardest to model.

Main culprit — the electrostatics. A simple three-point charge model yields an error of about 50% in the regeneration of the potential.

This error drops to less than 10% with a four-point charge model.

Problem — Most force fields were developed with three-point charge water models.

<table>
<thead>
<tr>
<th>model</th>
<th>N</th>
<th>$r_e$ (Å)</th>
<th>$E_{pot}$ (kJ/mol)</th>
<th>$\rho$ ($g \text{ cm}^{-3}$)</th>
<th>$T$ (K)</th>
<th>$D$ ($\times 10^{-9} \text{ m}^2 \text{ s}^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPC/E$^a$ original</td>
<td>820</td>
<td>12.0</td>
<td>$-47.2(0.18)$</td>
<td>1.008</td>
<td>301(4.4)</td>
<td>2.7(0.12)</td>
</tr>
<tr>
<td>SPC/L$^b$ original</td>
<td>901</td>
<td>12.0</td>
<td>$-45.4(0.03)$</td>
<td>0.998</td>
<td>298.2(1.4)</td>
<td>2.8(0.06)</td>
</tr>
<tr>
<td>SPC$^a$ original</td>
<td>820</td>
<td>12.0</td>
<td>$-42.2(0.16)$</td>
<td>0.988</td>
<td>301(4.4)</td>
<td>4.2(0.08)</td>
</tr>
<tr>
<td>SPC$^b$ original</td>
<td>901</td>
<td>12.0</td>
<td>$-40.5(0.03)$</td>
<td>0.998</td>
<td>298.6(1.1)</td>
<td>4.2(0.08)</td>
</tr>
<tr>
<td>TIP3P$^a$ original</td>
<td>820</td>
<td>12.0</td>
<td>$-40.8(0.16)$</td>
<td>1.001</td>
<td>301(4.4)</td>
<td>5.4(0.14)</td>
</tr>
<tr>
<td>TIP3P$^b$ original</td>
<td>901</td>
<td>12.0</td>
<td>$-39.0(0.02)$</td>
<td>0.998</td>
<td>297.0(0.9)</td>
<td>5.6(0.08)</td>
</tr>
<tr>
<td>TIP3P$^b$ modified</td>
<td>901</td>
<td>12.0</td>
<td>$-39.8(0.02)$</td>
<td>0.998</td>
<td>299.2(1.0)</td>
<td>5.9(0.08)</td>
</tr>
<tr>
<td>SPC$^b$ refined</td>
<td>901</td>
<td>12.0</td>
<td>$-40.3(0.03)$</td>
<td>0.998</td>
<td>297.7(1.2)</td>
<td>4.2(0.10)</td>
</tr>
<tr>
<td>exptl</td>
<td></td>
<td></td>
<td>$-41.5$</td>
<td>0.997</td>
<td></td>
<td>2.3</td>
</tr>
</tbody>
</table>


dynamic properties are more difficult to reproduce accurately
How to handle induction effects?

How to handle induction effects?

Induction energy, $U_{\text{ind},k}$, resulting from the polarization of the molecule by a charge $q_k$, located at point $r_k$,

$$U_{\text{ind},k} = E_{\text{total},k} - \varepsilon^0 - q_k \ V_k$$

$\forall \ k = 1, \ldots, N_p$

Approximation of the induction energy, introducing $\alpha$,

$$\tilde{U}_{\text{ind},k} = -\frac{1}{2} \sum_{i,j} \sum_{\ell=0}^{N_\ell(i)} \sum_{\kappa=-\ell}^{\ell} \sum_{\ell'=0}^{N_{\ell'}(j)} \sum_{\kappa'=-\ell'}^{\ell'} q_k \ T^{ki}_{\ell\kappa} \ a^{ij}_{\ell\kappa,\ell'\kappa'} \ T^{jk}_{\ell'\kappa'} q_k$$

How to handle induction effects?

The Drude oscillator

Induced dipole moment, \( \mu = \frac{q_D^2 E}{k_D} \)

The primary computational cost of simulating classical Drude oscillators is the calculation of the local electrostatic field and the repositioning of the Drude particle at each step, which is done traditionally in a self-consistent fashion. This cost can be reduced by assigning a small mass to each Drude particle, applying a Lagrangian transformation and evolving the simulation in the generalized coordinates.

Main advantage: Simplicity of implementation, cost-effectiveness, charge-charge framework.

Drude, P. *Lehrbuch der Optik. 1. Ausgabe*. Verlag von S. Hirzel, 1900
Towards a simpler representation

The results of a 1-µs simulation of a 64-million atom assay shows that the HIV capsid acts as an osmotic regulator.
If atom types and parameters are missing from a macromolecular force field, one ought to follow verbatim its philosophy to update it. A force field is a very fragile construct — its elementary bricks cannot be swapped without any precaution.

The vast majority of macromolecular force fields are pairwise additive, i.e., they do not account for through-space induction phenomena, nor of charge transfers.

To reduce the computational effort and access longer timescales, the atomic description can be coarse-grained. One should, however, not simplify a model until one understands fully the phenomena at play.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
THE PROPAGATORS OF MOLECULAR DYNAMICS

Early experiments of propagating motion
The Verlet propagator

Taylor expansions at \( t + \delta t \) and \( t - \delta t \),

\[
\begin{align*}
\mathbf{r}_i(t + \delta t) &= \mathbf{r}_i(t) + \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2} \mathbf{a}_i(t) + \mathcal{O}(\delta t^3) \\
\mathbf{r}_i(t - \delta t) &= \mathbf{r}_i(t) - \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2} \mathbf{a}_i(t) + \mathcal{O}(\delta t^3)
\end{align*}
\]


The Verlet propagator

\[ \mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \mathbf{a}_i(t) \delta t^2 + \mathcal{O}(\delta t^4) \]

Taylor expansions at \( t + \delta t \) and \( t - \delta t \),

\[ \mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \delta t) - \mathbf{r}_i(t - \delta t)}{2 \delta t} + \mathcal{O}(\delta t^2) \]


The leap-frog propagator

\[
\begin{align*}
\mathbf{r}_i(t + \delta t) & = \mathbf{r}_i(t) + \mathbf{v}_i \left( t + \frac{\delta t}{2} \right) \delta t \\
\mathbf{v}_i \left( t + \frac{\delta t}{2} \right) & = \mathbf{v}_i \left( t - \frac{\delta t}{2} \right) + \mathbf{a}_i(t) \delta t
\end{align*}
\]


The velocity Verlet propagator

\[
\begin{align*}
\mathbf{r}_i(t + \delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t) \delta t + \frac{1}{2} \mathbf{a}_i(t) \delta t^2 \\
\mathbf{v}_i(t + \delta t) &= \mathbf{v}_i(t) + \frac{\mathbf{a}_i(t) + \mathbf{a}_i(t + \delta t)}{2} \delta t \\
\mathbf{v}_i \left( t + \frac{\delta t}{2} \right) &= \mathbf{v}_i(t) + \frac{1}{2} \mathbf{a}_i(t) \delta t \\
\mathbf{v}_i(t + \delta t) &= \mathbf{v}_i \left( t + \frac{\delta t}{2} \right) + \frac{1}{2} \mathbf{a}_i(t + \delta t) \delta t
\end{align*}
\]

This algorithm is time-reversible

The predictor-corrector algorithm

Stage 1: Estimate \( \{ \mathbf{r}_i(t + \delta t), \mathbf{v}_i(t + \delta t), \mathbf{a}_i(t + \delta t), \mathbf{b}_i(t + \delta t) \} \),

\[
\begin{align*}
\mathbf{r}_i(t + \delta t)^{\text{pred}} &= \mathbf{r}_i(t) + \mathbf{v}_i(t) \delta t + \frac{1}{2} \mathbf{a}_i(t) \delta t^2 + \frac{1}{6} \mathbf{b}_i(t) \delta t^3 \\
\mathbf{v}_i(t + \delta t)^{\text{pred}} &= \mathbf{v}_i(t) + \mathbf{a}_i(t) \delta t + \frac{1}{2} \mathbf{b}_i(t) \delta t^2 + \ldots \\
\mathbf{a}_i(t + \delta t)^{\text{pred}} &= \mathbf{a}_i(t) + \mathbf{b}_i(t) \delta t + \ldots \\
\mathbf{b}_i(t + \delta t)^{\text{pred}} &= \mathbf{b}_i(t) + \ldots
\end{align*}
\]

Stage 2: From the forces evaluated at \( t + \delta t \), using \( \{ \mathbf{r}_i(t + \delta t)^{\text{pred}} \} \), correct the accelerations and supply an estimate of the prediction error,

\[
\begin{align*}
\mathbf{r}_i(t + \delta t)^{\text{corr}} &= \mathbf{r}_i(t + \delta t)^{\text{pred}} + c_0 \Delta \mathbf{a}(t + \delta t) \\
\mathbf{v}_i(t + \delta t)^{\text{corr}} &= \mathbf{v}_i(t + \delta t)^{\text{pred}} + c_1 \Delta \mathbf{a}(t + \delta t) \\
\mathbf{a}_i(t + \delta t)^{\text{corr}} &= \mathbf{a}_i(t + \delta t)^{\text{pred}} + c_2 \Delta \mathbf{a}(t + \delta t) \\
\mathbf{b}_i(t + \delta t)^{\text{corr}} &= \mathbf{b}_i(t + \delta t)^{\text{pred}} + c_3 \Delta \mathbf{a}(t + \delta t)
\end{align*}
\]

Multiple time-step propagators

The Liouville equation of motion,

\[ \frac{\partial \rho_{NVT}(\mathbf{x}, p_x, t)}{\partial t} = -i \mathcal{L} \rho_{NVT}(\mathbf{x}, p_x, t) \]

\[ \Gamma(t) = e^{i \mathcal{L} t} \Gamma(0) \]

Trotter factorization,

\[ e^{i \mathcal{L} \Delta t} = e^{i \mathcal{L}_1 \frac{\Delta t}{2}} \left[ e^{i \mathcal{L}_2 \frac{\Delta t}{2k}} e^{i \mathcal{L}_3 \frac{\Delta t}{k}} e^{i \mathcal{L}_2 \frac{\Delta t}{2k}} \right]^k e^{i \mathcal{L}_1 \frac{\Delta t}{2}} \]

Separation of the short- and long-range contributions,

\[ e^{i \mathcal{H} \Delta t} = e^{i \mathcal{V}_{\text{long}}(\mathbf{x}) \frac{\Delta t}{2}} \left\{ e^{i \mathcal{V}_{\text{short}}(\mathbf{x}) \frac{\Delta t}{2}} \left[ e^{i \mathcal{V}_{\text{valence}}(\mathbf{x}) \frac{\Delta t}{2}} e^{i \mathcal{T}(p_x) \frac{\Delta t}{2}} e^{i \mathcal{V}_{\text{valence}}(\mathbf{x}) \frac{\Delta t}{2}} \right]^2 e^{i \mathcal{V}_{\text{short}}(\mathbf{x}) \frac{\Delta t}{2}} \right\}^2 e^{i \mathcal{V}_{\text{long}}(\mathbf{x}) \frac{\Delta t}{2}} \]
Integration of the equations of motion generally relies upon a Taylor expansion of the positions — possibly the velocities, at $t + \delta t$ and $t - \delta t$.

While in some algorithms knowledge of the velocities is not mandatory, the latter are required for the determination of the kinetic energy.

The cost of the simulation can be appreciably reduced, turning to a multiple-time step integrator, which exploits the fact different degrees of freedom relax over different time-scales and, thus, can be updated at different frequencies.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
The concept of biased molecular dynamics simulation

To push back the limitation of brute-force molecular dynamics and access timescales not amenable to conventional simulation times — due to significant free-energy barriers.

Adding additional forces to the conservative forces arising from the force field will result in preferential sampling of slow degrees of freedom ordinarily poorly explored in standard simulations.


Role of geometrical restraints

Basic idea: Reduce the volume of configurational space to be sampled by means of geometrical, harmonic restraints.

Application: Staging in geometrical free-energy calculations.
Role of geometrical restraints

Application: Steered molecular dynamics. The external force is exerted onto the system to force it to drift along the desired direction of Cartesian space.

In constant-force SMD simulations, a constant force on the order of $10^{-12}$ to $10^{-9}$ N, is exerted onto a subset of atoms, in addition to the conservative forces of the potential energy function.

Constant-velocity SMD simulations mimic the action of a mobile cantilever acting on a substrate in AFM experiments.

In AFM, the spring constant of the cantilever is $10^{-3}$ Nm$^{-1}$, translating to large thermal fluctuations in the position of the attached substrate. SMD utilizes spring constants $10^2$ stiffer, giving sharper spatial resolution and more detailed information about the interactions at play.

Until recently, the time scales explored by SMD were substantially shorter than those of AFM experiments, by a factor of $10^3$-$10^6$. 

Role of geometrical restraints

Application: Molecular dynamics flexible fitting

Basic idea: Fit atomic structures into density maps. The method consists of adding external forces proportional to the gradient of the density map into a molecular dynamics simulation of the atomic structure.

Combination of high-resolution atomic structures with low-resolution cryo-electron microscopy maps results in atomic models representing the conformational state captured by cryo-EM.

Role of holonomic constraints

Eliminate the hard degrees of freedom of the system, that correspond to high frequencies of vibration. In other words, constraints ≠ restraints.

Application: Increase $\delta t$

$$\chi_{ij}(t) = |\mathbf{r}_j(t) - \mathbf{r}_i(t)|^2 - d_{ij}^2$$

$$m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = f_i + g_i$$

Constraint force, \( g_i = - \sum_j \lambda_{ij}(t) \nabla_i \chi_{ij}(t) = -2 \sum_j \lambda_{ij}(t) \mathbf{r}_{ij}(t) \)

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2m_i} [f_i(t) + g_i(t)]$$

Role of holonomic constraints

\( d(1,2) \) and \( d(1,3) \) are fixed. \( \theta(1,2,3) \) may vary through the intramolecular potential.

Equations of motion,

\[
\begin{align*}
  m_1 \mathbf{a}_1 &= f_1 + g_1 \\
  m_2 \mathbf{a}_2 &= f_2 + g_2 \\
  m_3 \mathbf{a}_3 &= f_3 + g_3
\end{align*}
\]

Role of holonomic constraints

\( d_{(1,2)} \) and \( d_{(1,3)} \) are fixed. \( \theta_{(1,2,3)} \) may vary through the intramolecular potential.

Equations of motion,

\[
\begin{align*}
    m_1 \, a_1 &= f_1 + g_1 \\
    m_2 \, a_2 &= f_2 + g_2 \\
    m_3 \, a_3 &= f_3 + g_3
\end{align*}
\]

Constraints to be satisfied,

\[
\begin{align*}
    \chi_{12} &= r_{12}^2(t) - d_{12} = 0 \\
    \chi_{13} &= r_{13}^2(t) - d_{13} = 0
\end{align*}
\]

Role of holonomic constraints

\[ d(1,2) \text{ and } d(1,3) \text{ are fixed. } \theta(1,2,3) \text{ may vary through the intramolecular potential.} \]

Equations of motion,

\[
\begin{align*}
    m_1 \, a_1 &= f_1 + g_1 \\
    m_2 \, a_2 &= f_2 + g_2 \\
    m_3 \, a_3 &= f_3 + g_3
\end{align*}
\]

Constraints to be satisfied,

\[
\begin{align*}
    \chi_{12} &= r_{12}^2(t) - d_{12} = 0 \\
    \chi_{13} &= r_{13}^2(t) - d_{13} = 0
\end{align*}
\]

Introduction of the Lagrange multipliers,

\[ g_a = \frac{1}{2} \lambda_{12} \nabla(r_a) \chi_{12} + \frac{1}{2} \lambda_{13} \nabla(r_a) \chi_{23} \]

The set of constraints ought to be satisfied at every time step, \( m_a \, a_a = f_a + g_a \simeq f_a + g_a^{(r)} \)

In the Verlet propagator,

\[ r_a(t + \delta t) = r'(t + \delta t) + \frac{\delta t^2}{m_a} \, g_a^{(r)}(t) \]

Role of holonomic constraints

Introduction of $g_a^{(r)}$ into the Verlet propagator,

\begin{align*}
g_1^{(r)} &= \lambda_{12} \, r_{12} \\
g_2^{(r)} &= \lambda_{23} \, r_{23} - \lambda_{12} \, r_{12} \\
g_3^{(r)} &= -\lambda_{23} \, r_{23}
\end{align*}

\begin{align*}
\begin{cases}
  r_{12}(t+\delta t) &= r'_{12}(t+\delta t) + \left( \frac{1}{m_1} + \frac{1}{m_2} \right) \lambda_{12} \, \delta t^2 \, r_{12}(t) - \frac{1}{m_2} \lambda_{23} \, \delta t^2 \, r_{23}(t) \\
  r_{23}(t+\delta t) &= r'_{23}(t+\delta t) + \frac{1}{m_2} \lambda_{12} \, \delta t^2 \, r_{12}(t) - \left( \frac{1}{m_2} + \frac{1}{m_3} \right) \lambda_{23} \, \delta t^2 \, r_{23}(t)
\end{cases}
\end{align*}

This is a quadratic system of two equations of unknowns, $\lambda_{12}$ and $\lambda_{13}$.

It is possible to rigidify the water molecule fully by enforcing an additional constraint on the H—H distance.

The SHAKE algorithm

Cyclic computation of all constraints, in an iterative fashion until each constraint be satisfied.

Modification of the Verlet propagator: At each step, \( \mathbf{g} \) is approximated in order to satisfy the set of constraints,

\[
\begin{align*}
\mathbf{r}_a(t + \delta t) &= \mathbf{r}'_a(t + \delta t) + \frac{1}{2} \frac{\delta t}{m_a} \mathbf{g}^{(r)}_a(t) \\
\mathbf{v}_a \left( t + \frac{\delta t}{2} \right) &= \mathbf{v}'_a \left( t + \frac{\delta t}{2} \right) + \frac{1}{2} \frac{\delta t}{m_a} \mathbf{g}^{(r)}_a(t)
\end{align*}
\]

Note: Constraint forces, \( \mathbf{g}^{(r)}_a(t) \), are collinear with \( \mathbf{r}_{ab}(t) \), and so are \( \mathbf{g}^{(r)}_a(t+\delta t) \).

Next:
\[
\mathbf{v}_a(t + \delta t) = \mathbf{v}'_a(t + \delta t) + \frac{1}{2} \frac{\delta t}{m_a} \mathbf{g}^{(r)}_a(t + \delta t)
\]

Note: Knowledge of \( \mathbf{v}'_a(t+\delta t) \) requires the preliminary evaluation of \( \mathbf{f}_a(t+\delta t) \).

The RATTLE algorithm — Iterative computation of the solutions to these equations.
Rigid-body molecular dynamics

{\{e^B\}}: Frame of reference of the rigid body.
{\{e^S\}}: Absolute frame of reference: \( e^B = S \cdot e^S \)

\[
S = \begin{pmatrix}
\cos \phi \cos \psi - \sin \phi \cos \theta \sin \psi & \sin \phi \cos \psi + \cos \phi \cos \theta \sin \psi & \sin \theta \sin \psi \\
-\cos \phi \sin \psi - \sin \phi \cos \theta \cos \psi & -\sin \phi \sin \psi + \cos \phi \cos \theta \cos \psi & \sin \theta \cos \psi \\
\sin \phi \sin \theta & -\cos \phi \sin \theta & \cos \theta
\end{pmatrix}
\]

\[
\tau_i = \sum_a (r_{ia} - r_i) \times f_{ia}
\]

\( \tau^B = S \tau^S \)

\( \omega^S = S^{-1} \omega^B \)

\[
\frac{dL^B}{dt} + \omega^B \times L^B = \tau^B
\]

\[
\frac{de^S}{dt} = \frac{de^B}{dt} + \omega^S \times e^S = \omega^S \times e^S \quad \text{which depends upon} \quad (\dot{\phi}, \dot{\theta}, \dot{\psi}) \quad \text{function of} \quad 1/\sin \theta.
\]

Rigid-body molecular dynamics

Quaternions: Four scalar quantities satisfying $q_1 + q_2 + q_3 + q_4 = 1$

$$\begin{align*}
q_0 &= \cos \frac{\theta}{2} \cos \frac{\phi + \psi}{2} \\
q_1 &= \sin \frac{\theta}{2} \cos \frac{\phi - \psi}{2} \\
q_2 &= \sin \frac{\theta}{2} \sin \frac{\phi - \psi}{2} \\
q_3 &= \cos \frac{\theta}{2} \sin \frac{\phi + \psi}{2}
\end{align*}$$

\[ S = \begin{pmatrix}
q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 + q_0 q_3) & 2(q_1 q_3 - q_0 q_2) \\
2(q_1 q_2 - q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 + q_0 q_1) \\
2(q_1 q_3 + q_0 q_2) & 2(q_2 q_3 - q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2
\end{pmatrix} \]

For each rigid molecule, the quaternions satisfy,

$$\begin{pmatrix}
\dot{q}_0 \\
\dot{q}_1 \\
\dot{q}_2 \\
\dot{q}_3
\end{pmatrix} = \frac{1}{2} \begin{pmatrix}
q_0 & -q_1 & -q_2 & -q_3 \\
q_1 & q_0 & -q_3 & q_2 \\
q_2 & q_3 & q_0 & -q_1 \\
q_3 & -q_2 & q_1 & q_0
\end{pmatrix} \begin{pmatrix} 0 \\
\omega_x^B \\
\omega_y^B \\
\omega_z^B
\end{pmatrix}$$

Application: Robust computation of root mean-square deviations.

Often utilized interchangeably — erroneously so, *restraints* and *constraints* are two different concepts.

Geometric restraints are introduced in a molecular dynamics simulation to restrict sampling to regions of interest of configurational space.

Holonomic constraints are introduced to remove degrees of freedom — to freeze them in the equations of motion. Constrained molecular dynamics implies integrating modified equations of motion featuring constraint forces.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
Thermodynamic ensembles and conserved properties

<table>
<thead>
<tr>
<th>Ensemble</th>
<th>Free energy</th>
<th>Conserved Hamiltonian, $\mathcal{H}(\mathbf{x}, p_x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N, V, \mathcal{E})$</td>
<td>$A = U - TS$</td>
<td>$\sum_i \frac{1}{2m_i} (p_{xi}^2 + p_{yi}^2 + p_{zi}^2) + \mathcal{V}(\mathbf{x})$</td>
</tr>
<tr>
<td>$(N, V, T)$</td>
<td></td>
<td>$U$</td>
</tr>
<tr>
<td>$(N, P_\perp, \mathcal{A}, T)$</td>
<td>$A = U - TS + P_\perp V$</td>
<td>$\sum_i \frac{1}{2m_i} (p_{xi}^2 + p_{yi}^2 + p_{zi}^2) + \mathcal{V}(\mathbf{x}) + P_\perp V$</td>
</tr>
<tr>
<td>$(N, P_\parallel, \mathcal{A}, T)$</td>
<td>$A = U - TS + P_\parallel V$</td>
<td>$\sum_i \frac{1}{2m_i} (p_{xi}^2 + p_{yi}^2 + p_{zi}^2) + \mathcal{V}(\mathbf{x}) + P_\parallel V$</td>
</tr>
<tr>
<td>$(N, V, \gamma, T)$</td>
<td>$A = U - TS - \gamma \mathcal{A}$</td>
<td>$\sum_i \frac{1}{2m_i} (p_{xi}^2 + p_{yi}^2 + p_{zi}^2) + \mathcal{V}(\mathbf{x}) - \gamma \mathcal{A}$</td>
</tr>
<tr>
<td>$(N, P_\perp, \gamma, T)$</td>
<td>$A = U - TS + P_\perp V - \gamma \mathcal{A}$</td>
<td>$\sum_i \frac{1}{2m_i} (p_{xi}^2 + p_{yi}^2 + p_{zi}^2) + \mathcal{V}(\mathbf{x}) + P_\perp V - \gamma \mathcal{A}$</td>
</tr>
</tbody>
</table>

Constant-temperature molecular dynamics

**Velocity rescaling**

Central idea: The velocities are rescaled periodically by a factor of $\sqrt{T/T_\mathcal{F}}$.

Strictly speaking, this algorithm yields *non-Newtonian* dynamics.

**Constrained equations of motion**

$$\begin{cases} 
\dot{x}_i = \frac{p_i}{m_i} \\
\dot{p}_i = f_i - \xi(x; p_x) \cdot p_x 
\end{cases}$$

The friction coefficient guarantees that $\dot{T}_\mathcal{F} = 0$, and is chosen so that the perturbation of the chemical system is minimal,

$$\xi(x; p_x) = \frac{\sum_i p_i \cdot f_i}{\sum_i |p_i|^2}$$

Constant-temperature molecular dynamics

Extended system

Central idea: Introduce an additional degree of freedom, $s$, thermostat of the system.

\[
\begin{align*}
\mathcal{V}_s &= \frac{f + 1}{\beta} \ln s \\
\mathcal{I}_s &= \frac{1}{2} \mathcal{Q} \dot{s}^2
\end{align*}
\]

Velocity of particle $i$: $v_i = s \dot{r}_i = p_i/m_is$

Extended Lagrangian: $\mathcal{L}_s(x; p_x) = \mathcal{I}(p_x) + \mathcal{I}_s(p_x) - \mathcal{V}(x) - \mathcal{V}_s(x)$

Poorly chaotic for a limited number of degrees of freedom


Constant-temperature molecular dynamics

Extended system

Practical implementation:

\[ \begin{aligned}
\dot{p}' &= \frac{p}{s} \\
\dot{s}' &= \frac{s'}{s} \\
\dot{\xi} &= \frac{\sum_i \frac{p_i^2}{m_i} - \frac{f + 1}{\beta}}{2} \\
\dot{s} &= \frac{d \ln s}{dt} = \xi
\end{aligned} \]


Frenkel, D.; Smit, B. Understanding molecular simulations: From algorithms to applications, Academic Press, **2002**
Constant-temperature molecular dynamics

**Weak-coupling algorithm**

Central idea: Let the instantaneous kinetic temperature, $T_\mathcal{F}(t)$, “relax” towards the desired, reference temperature, $T$.

\[
\frac{dT_\mathcal{F}(t)}{dt} = \frac{T - T_\mathcal{F}(t)}{\tau_T}
\]

\[
\Delta \mathcal{F} = \frac{1}{2} (\chi^2 - 1) N k_B T_\mathcal{F}(t)
\]

\[
\chi = \left[1 + \frac{\delta t}{\tau_T} \left(\frac{T}{T_\mathcal{F}(t)} - 1\right)\right]^{1/2}
\]

Strictly speaking, this algorithm does not yield a canonical distribution.


Constant-temperature molecular dynamics

Langevin dynamics

Central idea: Add to the conservative forces a frictional force proportional to the velocity. The friction removes the kinetic energy from the system. A random force adds kinetic energy to the system.

\[
m_i \frac{d^2 x_i}{dt^2} = -\frac{dV(x)}{dx} - \gamma \frac{dx_i}{dt} + f(t)
\]

Gaussian white noise of zero mean, \( \langle f(t) \rangle = 0 \), obeying the fluctuation-dissipation theorem

\[
\langle f(t) f(t') \rangle = \frac{2 m_i \gamma}{\beta} \delta(t - t')
\]

which is precisely the condition to generate a canonical distribution.

Constant-temperature molecular dynamics

**Langevin dynamics**

Central idea: Add to the conservative forces a frictional force proportional to the velocity. The friction removes the kinetic energy from the system. A random force adds kinetic energy to the system.

\[
m_i \frac{d^2 x_i}{dt^2} = -\frac{d\mathcal{V}(x)}{dx} - \gamma m_i \frac{dx_i}{dt} + f(t)
\]

Gaussian white noise of zero mean, \( \langle f(t) \rangle = 0 \), obeying the fluctuation-dissipation theorem

\[
\langle f(t) f(t') \rangle = \frac{2 m_i \gamma}{\beta} \delta(t - t')
\]

which is precisely the condition to generate a canonical distribution.

**What should I use for the friction?**

Schlick, T. *Molecular modeling and simulation* Springer, 2002
Constant-temperature molecular dynamics

The Lowe-Andersen algorithm

Central idea: Couple the chemical system to a heat bath via stochastic forces that modify the kinetic energy of the atoms or molecules.

The time between collisions — or the number of collisions over a short-time interval, is decided randomly, based on the Poisson distribution, \( P(t) = \nu \exp(-\nu t) \).

Between collisions the system evolves at constant energy. During a collision event, the new momentum of the particles is chosen randomly from a Boltzmann distribution at a given temperature. The stochastic collision frequency can adopt any value, albeit there is an optimum choice,

\[
\nu = \frac{2a\kappa V^{1/3}}{3k_B N}
\]

Andersen, H. C. J. Chem. Phys. 1980, 72, 2384-2393
Instantaneous pressure and the virial theorem

Pair interactions:
\[ \sum_i x_i \cdot f_i = \sum_{i<j} x_i \cdot f_{ij} = \frac{1}{2} \sum_i \sum_{j \neq i} r_i \cdot f_{ij} + x_j \cdot f_{ji} \]

Third law of Newton, \( f_{ij} = -f_{ji} \)

\[ \sum_i x_i \cdot f_i = \frac{1}{2} \sum_i \sum_{j \neq i} f_{ij} \cdot (x_i - x_j) = \sum_{i<j} f_{ij} \cdot x_{ij} \]

Internal virial:
\[ \mathcal{W} = \frac{1}{3} \sum_i x_i \cdot f_i^{\text{intermolecular}} = -\frac{1}{3} \sum_i x_i \cdot \nabla (r_i) \mathcal{V}(x) \]

The external pressure arises from external forces,
\[ \frac{1}{3} \langle \sum_i x_i \cdot f_i^{\text{external}} \rangle = -PV = - \left( \frac{N}{\beta} + \langle \mathcal{W} \rangle \right) \]

Instantaneous pressure,
\[ P_\beta = \frac{\rho}{\beta} + \frac{\mathcal{W}}{V} = \frac{1}{V} \left( \frac{N}{\beta} - \frac{1}{3} \sum_i x_i \cdot f_i \right) \]
Constant-pressure molecular dynamics

Extended system

Central idea: Couple the system to an external, additional variable, \( V \), that could be related to the motion of a piston.

\[
\begin{align*}
\mathcal{V}_V &= \frac{1}{2} m_P \dot{V}^2 \\
\mathcal{T}_V &= P \dot{V}
\end{align*}
\]

Rescale: \( s = x/V^{1/3} \) and \( \dot{s} = v/V^{1/3} \)

Extended Lagrangian, \( \mathcal{L}_V(x; p_x) = \mathcal{T}(p_x) + \mathcal{T}_V(p_x) - \mathcal{V}(x) - \mathcal{V}_V(x) \)

Add a frictional term and a stochastic force to eliminate possible resonances between the piston and the chemical system.

\[
\begin{align*}
\ddot{s}_i &= \frac{f_i}{m_i V^{1/3}} - 2 \frac{\dot{s}_i \dot{V}}{V} \\
\ddot{V} &= \frac{P \mathcal{P} - P}{m_P} - \gamma \dot{V} + R(t)
\end{align*}
\]

Andersen, H. C. J. Chem. Phys. 1980, 72, 2384-2393

Constant-pressure molecular dynamics

Extended system

Central idea: Couple the system to an external, additional variable, $V$, that could be related to the motion of a piston.

$$
\begin{align*}
\mathcal{V}_V &= \frac{1}{2} m_P \dot{V}^2 \\
\mathcal{T}_V &= P V
\end{align*}
$$

Rescale: $s = x/V^{1/3}$ and $\dot{s} = v/V^{1/3}$

Extended Lagrangian, $\mathcal{L}_V(x; p_x) = \mathcal{T}(p_x) + \mathcal{T}_V(p_x) - V(x) - V_V(x)$

Add a frictional term and a stochastic force to eliminate possible resonances between the piston and the chemical system.

$$
\begin{align*}
\ddot{s}_i &= \frac{f_i}{m_i V^{1/3}} - C \frac{\ddot{s}_i}{V} \\
\ddot{V} &= \frac{P \mathcal{P} - P}{m_P} - \gamma \dot{V} + R(t)
\end{align*}
$$


Constant-pressure molecular dynamics

The weak -coupling algorithm

Central idea: The instantaneous pressure, $P(t)$, relaxes towards the reference value, $P$.

$$\frac{dP(t)}{dt} = \frac{P - P(t)}{\tau_P}$$

The atomic coordinates and the dimensions of the simulation cell are rescaled by a factor $\varsigma$.

Change in the volume, $\Delta V = (\varsigma^3 - 1) V$, and variation of the pressure:

$$\Delta P = \frac{\Delta V}{\beta \gamma V}$$

$$\varsigma = \left[ 1 - \beta \gamma \delta t \frac{P - P(t)}{\tau_P} \right]^{1/3}$$

Strictly speaking, this algorithm does not yield an isobaric distribution.

Constant-pressure molecular dynamics

Modifying the shape of the cell

Metric transform: \( \mathbf{r} = \mathbf{H} \mathbf{s} \), where \( \mathbf{H} \equiv (h_1; h_2; h_3) \)

Volume of the cell: \( V = \| \mathbf{H} \| = h_1 \cdot h_2 \times h_3 \)

Extended Lagrangian:

\[
\begin{align*}
\mathcal{L}_V &= P \, V \\
\mathcal{F}_V &= \frac{1}{2} Q \sum_{\alpha} \sum_{\beta} \| \mathbf{H}_{\alpha\beta} \|^2
\end{align*}
\]

\[
m \frac{d^2 \mathbf{s}}{dt^2} = \frac{1}{\mathbf{H}} \mathbf{f} - m \frac{\dot{\mathbf{G}}}{\mathbf{G}} \frac{d \mathbf{s}}{dt}
\]

where \( \mathbf{G} = \mathbf{H}^T \mathbf{H} \)

\[
Q \, \ddot{\mathbf{G}} = (P_{\mathcal{D}} - 1P) \, V \left( \frac{1}{\mathbf{H}} \right)^T
\]

\[
P_{\mathcal{D}}^{\alpha\beta} = \frac{1}{V} \left[ \sum_i m_i \left( \mathbf{H} \frac{d \mathbf{s}_i}{dt} \right)^T \left( \mathbf{H} \frac{d \mathbf{s}_i}{dt} \right)^T + \sum_i \sum_{j > i} (\mathbf{H} \mathbf{s}_{ij})_{\alpha} (\mathbf{f}_{ij})_{\beta} \right]
\]

Constant-pressure molecular dynamics

How to handle anisotropic chemical systems?

Historically, two conflicting approaches for handling the pressure:

A membrane is a self-assembled system. Consequently, $\gamma = 0$, because:

$$\gamma = \int_{z_1}^{z_2} dz \left[ P_\perp(z) - P_\parallel(z) \right]$$

Since only a small patch of the “real” membrane is simulated, a component of the capillary wave and bending motion spectra is missing. $\gamma$ is, therefore, finite.

More accurate force fields reconcile the two views.


Molecular dynamics simulations can be carried out in a variety of ensembles.

The choice of the thermodynamic ensemble is dictated by the conditions in which the experiment was performed — and the nature of the molecular assembly.

Several algorithms are available to maintain the temperature and the pressure constant. Only a few of them supply the correct distribution.

The parameters of the thermostat and of the barostat have a direct impact on the behavior of the simulation and the convergence properties thereof.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
Why should we turn to lattice sums?

Central idea: For an electrically neutral system, replace the conditionally convergent,

$$\mathcal{V}_{\text{Coulomb}}(\mathbf{r}) = \sum_{i < j} \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon_1 r_{ij}}$$

by,

$$\sum_n \frac{1}{|n|} \mathcal{F}(n) + \sum_m \frac{1}{|m|} [1 - \mathcal{F}(m)]$$

Gaussian charge distribution,

$$\rho_i(\mathbf{r}) = q_i \alpha^3 \exp\left(-\alpha^2 r^2\right)$$

Ewald, P. Ann. Phys. 1921, 64, 253-287
Why should we turn to lattice sums?

\[
\sum_n \frac{1}{|n|} \mathcal{F}(n) + \sum_m \frac{1}{|m|} [1 - \mathcal{F}(m)]
\]

The interactions of charges are screened by the charge distribution, limiting their contribution to a short-range term. Convergence of the sum of charges and their images in the direct space is fast.

In the reciprocal space, a Fourier transform is used for solving \( \nabla^2 \mathcal{V}_i(r) = -4\pi q_i(r) \). The transform decreases rapidly and the sum converges equally so.

\[
\mathcal{V}_{\text{Ewald}}(r) = \frac{1}{2V\varepsilon_0} \sum_{k\neq 0} \frac{\exp(-k^2/4\alpha^2)}{k^2} \left[ \sum_j q_j \exp(-i\mathbf{k} \cdot \mathbf{r}_j) \right] \left[ \sum_j q_j \exp(i\mathbf{k} \cdot \mathbf{r}_j) \right] 
\]

\[
+ \frac{1}{4\pi\varepsilon_0} \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij} \text{erfc}(\alpha r_{ij})} - \frac{\alpha}{4\pi^{3/2}\varepsilon_0} \sum_i q_i^2 
\]

\[
+ \frac{1}{4\pi\varepsilon_0} \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}} 
\]

Ewald, P. Ann. Phys. 1921, 64, 253-287
Why should we turn to lattice sums?

Define a grid filling the simulation Cartesian space. Interpolate the charges on the grid and evaluate the charge distribution, $\varrho(r)$;

Fourier transform to obtain $\hat{\varrho}(k)$ and estimate of $\hat{V}_{\text{long}}(k)$; Inverse transform to get $V_{\text{long}}(r)$;

Numerical evaluation of the forces from the potential;

Interpolation of the electric field towards the position of the particles.

Why should we turn to lattice sums?

\[
G_K(R) = \frac{\left\langle \sum_{i,j; r_{ij} < R} \mu_i \cdot \mu_j \right\rangle}{N \mu^2}
\]

Access to dielectric permittivity.

Why should we turn to lattice sums?

\[
\Delta G(r) = -\frac{1}{\beta} \ln \frac{\int \delta(r'(x) - r) \ e^{-\beta V(x)}}{\int dx \ e^{-\beta V(x)}}
\]

Summary

Long-range electrostatics cannot be properly handled by a mere spherical truncation.

Lattice-sum algorithms address the long-range nature of electrostatic interactions and the conditional convergence of the Coulomb sum by splitting the latter into a real-space and a Fourier-space contribution.

Contrary to traditional Ewald sum, particle-mesh approaches handle the Fourier-space part of the Coulomb sum on grids over which the position of the atoms is interpolated.

The usual $N \ln N$ cost of particle-mesh algorithms can be modulated by using a coarser discretization of space at the expense of accuracy.

Alternate schemes, e.g., the fast multipole method, are available, but have proven effective so far only for large arrays of particles.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
Data replication

The particles are scattered onto the different processors.


HANDS-ON WORKSHOP ON COMPUTATIONAL BIOPHYSICS
DALIAN INSTITUTE OF CHEMICAL PHYSICS, DALIAN, CHINA, JULY 2018
Force-loop splitting

The intermolecular forces are scattered onto the different processors.

Super linear behavior for a small number of cores, but rapid collapse of performance.

Spatial decomposition

The simulation cell is broken into subcells, scattered onto the different processors, optimizing communication.

*Crucial role of load balancing for optimal performance.*


Towards ergodic sampling — parallel tempering

\[ p = \min \left( 1, e^{(U_i - U_j)(\beta_i - \beta_j)} \right) \]
Molecular dynamics is parallelizable in nature. Domain-decomposition strategies currently offer the best performance on multipurpose parallel computer architectures.

GPU acceleration offers a cost-effective alternative to conventional CPU-based machines. Further increase of the performance implies design of specific, dedicated processors.

Efficient parallelization on massively parallel architectures dilates the size scale more than it dilates the time scale.

Replica-exchange based algorithms enhance ergodic sampling by making high-temperature configurations available to low-temperature simulations.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
Appropriate questions that ought to be asked —

- The results of the simulation are unexpected
- A new phenomenon has been discovered
- The results are simply erroneous
  - The model used for this application is unadapted
  - The force field is inadequate
  - The simulation has not converged
  - The program contains undocumented bugs
  - The program has been misused
  - Detailed description of the models?
  - Detailed description of the force field?
  - Time series?
  - Checkup of the code?
  - Detailed description of simulation setup?
Appropriate questions that ought to be asked —

*Why is it of paramount importance to thermalize the molecular assembly prior to accrue data for analysis?*

First, because evolution of the temperature and the pressure ought to reflect the proper thermodynamic ensemble.

Second, because large fluctuations in the volume of the simulation cell may reflect suboptimal packing.

Poorly thermalized assays may result in artifactual observations due, for instance, to spurious conformational changes.

Molecular dynamics simulations are a source of numerous errors of different nature.

Insufficient thermalization is at the origin of spurious behaviors, often difficult to discern.

Thermalization ought to be performed in a stepwise fashion, confining the position of the different components of the molecular assay prior to unrestrained sampling.

Other factors can impact the outcome of the simulation. A clear understanding of the different parameters at play and their respective effects is highly desirable.
SYNOPSIS

1. Introduction
2. Periodic boundary conditions
3. Synopsis of a molecular dynamics simulation
4. The potential energy function
5. The propagators of molecular dynamics
6. Restraints versus constraints
7. In which ensemble should the simulation be performed?
8. Lattice sums: The Ewald–Kornfeld approach
9. Molecular dynamics on parallel architectures
10. Guidelines
11. Properties accessible from the trajectories
Radial distribution functions

The pair correlation function,

\[ g(x_1; x_2) = \frac{N(N-1)}{\rho^2} \int e^{-\beta \mathcal{V}(x_1, \ldots, x_N)} \, dx_1 \ldots dx_N \int e^{-\beta \mathcal{V}(x_1, \ldots, x_N)} \, dx_3 \ldots dx_N \]

can be restated as an ensemble average,

\[ g(r) = \frac{V}{N^2} \left\langle \sum_i \sum_{j \neq i} \delta(x - x_{ij}) \right\rangle \]

which after normalization reads,

\[ g_{ij}(r) = \frac{n_j(r; r + \delta r)}{4\pi \rho_j r^2 \, dr} \]

Chandler, D. Introduction to modern statistical mechanics, Oxford University Press, 1987
Autocorrelation functions

Correlation coefficient of observable $\mathcal{B}$,

$$c_{\mathcal{BB}'} = \frac{\langle \delta \mathcal{B} \, \delta \mathcal{B}' \rangle}{\sigma(\mathcal{B}) \, \sigma(\mathcal{B}')}$$

with $\delta \mathcal{B} = \mathcal{B} - \langle \mathcal{B} \rangle$

$$c_{\mathcal{BB}'}(t) = \frac{\langle \delta \mathcal{B}(t) \, \delta \mathcal{B}'(0) \rangle}{\sigma(\mathcal{B}) \, \sigma(\mathcal{B}')}$$

$$c_{\mathcal{BB}}(t) = \frac{\langle \delta \mathcal{B}(t) \, \delta \mathcal{B}(0) \rangle}{\langle \delta \mathcal{B}(0) \, \delta \mathcal{B}(0) \rangle}$$

Diffusion coefficients

From autocorrelation functions,

\[ \gamma = \int_0^\infty \left\langle \dot{\mathbf{B}}(t) \dot{\mathbf{B}}(0) \right\rangle \, dt \]

\[ 2t\gamma = \left\langle \left( \dot{\mathbf{B}}(t) - \dot{\mathbf{B}}(0) \right)^2 \right\rangle \]

\[ D = \frac{1}{3} \int_0^\infty \left\langle \mathbf{v}(t) \cdot \mathbf{v}(0) \right\rangle \]

which, over long times, coincides with,

\[ 2tD = \frac{1}{3} \left\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \right\rangle \]

Structural properties

\[ -S_i^{\text{mol}} = \frac{1}{2} \langle 3 \cos^2 \chi_i - 1 \rangle \]

Structural properties

Properties accessible from the trajectories

**Structural properties**

\[ \Phi(z) = \int_{0}^{z} d\zeta' \varrho(\zeta') \]

\[ \mathcal{E}(z) = \int_{0}^{z} d\zeta' \varrho(\zeta') \]

Structural properties

Time series of the distance root mean-square deviation with some reference structure,

\[
\text{RMSD}(\{x_i(t)\}, \{x_{i}^{\text{ref}}\}) = \left( \frac{1}{N} \sum_{i} |S[x_i(t) - x_{\text{COM}}(t)] - (x_{i}^{\text{ref}} - x_{\text{COM}}^{\text{ref}})|^2 \right)^{1/2}
\]

Find the optimal rotation, \(S\), which minimizes,

\[\text{Leach, A. R. Molecular modeling. Longman, 1996}\]
Structural properties

Time series of the radius of gyration, \( R_0^k = \sqrt{\frac{I^k}{M}} \)

\[
I = \sum_i m_i \left[ (\mathbf{r}_i \cdot \mathbf{r}_i) \mathbf{e} - \mathbf{r}_i \times \mathbf{r}_i \right]
\]

Asymmetry, \( A_3 = \frac{\sum_{i>j} \left( \left( R_{0i}^2 - R_{0j}^2 \right) \right)^2}{2 \left( \left( \sum_i R_{0i}^2 \right)^2 \right)} \)

Simplified expressions —

\[
R_g^2 := \frac{1}{N} \sum_i (\mathbf{r}_i - \bar{\mathbf{r}})^2
\]

Goldstein, H. Classical mechanics. Addison-Wesley, 1980

Rudnick, J.; Gaspari, G. Science 1987, 237, 384-389
Structural properties

Folding dynamics of villin headpiece unveiled in molecular dynamics simulations exploring key folding transitions never seen hitherto.

Freddolino, P. L.; Schulten, K. Biophys. J. 2009, 97, 2338-2347
CLASSICAL MOLECULAR DYNAMICS

Properties accessible from the trajectories

QwikMD
Gateway to Easy Simulation

www.ks.uiuc.edu/Research/qwikmd

QwikMD is freely available in VMD 1.9.3 and later


HANDS-ON WORKSHOP ON COMPUTATIONAL BIOPHYSICS
DALIAN INSTITUTE OF CHEMICAL PHYSICS, DALIAN, CHINA, JULY 2018
Structure refinements

The shape of the chromatophore is determined by LH1-RC
The curvature of LH1-RC was inferred by MDFF

The radius of curvature consistent with experiment;
The membrane curvature exhibits both anisotropy and twist.

Structure refinements

Structure determined from experimental data through modeling and simulation.


Free energies and other thermodynamic properties

(1) Methods based on histograms
\[ \Delta A(\xi) = -\frac{1}{\beta} \ln P(\xi) + \Delta A_0 \]

(2) Non-equilibrium work simulations
\[ \exp(-\beta \Delta A) = \langle \exp(-\beta w) \rangle \]

(3) Perturbation theory
\[ \exp(-\beta \Delta A) = \langle \exp(-\beta \Delta U) \rangle_0 \]

(4) Integration of the gradient
\[ \frac{dA(\xi)}{d\xi} = \left\langle \frac{\partial U}{\partial \xi} - \frac{1}{\beta} \frac{\partial \ln |J|}{\partial \xi} \right\rangle_\xi \]

A variety of properties, structural, thermodynamic or dynamic, can be inferred from molecular dynamics simulations.

Because analysis of these properties is ordinarily performed post facto, it is recommended to store configurations as frequently as possible.

Convergence of the calculation of these properties may vary appreciably — a free energy versus a diffusivity.

Many properties can be readily compared to experiment. Accurate reproduction of the latter is subservient to the reliability of the force field and finite-length effects. Estimation of the associated error is strongly recommended.
CONCLUDING REMARKS

The computational microscope — In recent years, molecular dynamics simulations have become an important tool to complement experiment, offering unprecedented resolution and access to valuable dynamic information.

The inherent parallelizable nature of molecular dynamics makes it an ideal candidate for harnessing the constantly growing capacity of novel computer architectures, dilating both the size- and timescale tackled by simulations.

Molecular dynamics rests on the equations of classical mechanics and the use of empirical potential energy functions, the construction of which relies on stringent assumptions and approximations.

As macromolecular force fields become more sophisticated — and, thus, less tractable, accounting for induction and charge-transfer phenomena, one must choose between longer simulations with a more approximate description and shorter, more accurate simulations.