Introduction to Molecular Dynamics

Dr Ilian Todorov (STFC, UK) – DL_POLY
Dr Rossen Apostолов (KTH, SWE) – GROMACS
Computer modelling is used to provide insight and understanding of how complex systems behave beyond what theory and experiment could deliver separately. It bridges theory and experiment by solving state equations numerically.

Computer simulations are used as an assisting tool by scientists and engineers to verify and/or predict experimental observations as well as test and/or tune theoretical models.
Computational Science is a generic term for any field of science where computer simulation is used in conjunction with theory and experiment to model various aspects of “reality” (observed or sought after phenomena), to provide solution to problems and answer to questions that often cannot be satisfactory explained by theory and experiment alone.

Within natural sciences one could outline a number of fields which generally resolve behaviour of nature (the physical world surrounding us) on different time- and length-scales: Physics, Chemistry, Materials and Engineering.
Two primary roles:
• Test models which explain experiments
• Test theoretical predictions

Driving forces:
• Computers are fast enough for numerical experiments
• Most models are too complicated for purely theoretical reasoning
• There are phenomena which can not be observed directly by experiments

Allen&Tildesley
Computational Chemistry uses computers to aid chemical enquiry. It generally involves methodology and theory to model systems on the atomistic scale. It is generally divided into:

- **Ab Initio** methods that use rigorous Quantum Mechanics
  + accurate and versatile
  - computationally extremely expensive
  : give good results for very small systems $\sim 10^2$ atoms

- **Semi-Empirical** methods that use approximate Quantum Mechanics
  - rely on empirical or *ab initio* parameters
  + affordable and accurate
  : limited to the correct, well defined problems for systems with $\sim 10^4$ atoms

- **Molecular Mechanics** methods using Classical Mechanics
  - rely on empirical force fields without accounting for electronic properties (change in chemical nature – bond breaking or forming)
  + very affordable and used as virtual experiment
  : can handle extremely large systems $\sim 10^{12}$ atoms
Time & Length-Scales – I

Accuracy

g- geometrical
t- topological
q- qualitative

1 eV / 40 kT
0.1 eV / 4 kT
0.0001 eV / 0.004 kT

finite elements
Coarse grained molecular

Electrons

atoms

GAP in accuracy and speed
**Time & Length-Scales – II**

**Biological Experiments**

- Molecular dynamics
- QM simulations

**Coarse-grained models**

- Whole proteins, viruses

- **Molecular dynamics**
  - (Atomic detail)
  - (MD is fast!)

- **QM simulations**
  - (Electrons)
Examples of Model Systems

- Proteins solvation & binding
- DNA strands dynamics
- Membranes’ processes
- Drug polymorphs & discovery
- Crystalline & Amorphous Solids – damage and recovery
- Dynamic processes in Metal-Organic & Organic Frameworks
- Dynamics at Interfaces & of Phase Transformations
Molecular Dynamics

- MD is the solution of the classical equations of motion for atoms and molecules to obtain the time evolution of a system.

- It is applied to many-particle systems since a general analytical solution is not possible. Thus one must resort to numerical methods and computers.

- It does classical mechanics only since a fully fledged many-particle time-dependent quantum method is not yet available.

- It uses a Maxwell-Boltzmann averaging process for thermodynamic properties (i.e. time averaging).
Limitations of MD

• Interatomic interactions (called a Force Field) are based on parameters and are imperfect, because they are derived or fitted for fixed conditions (P, T, pH, environment). Transferability of force-fields for mixed environments is an open question.

• Phase space is not sampled exhaustively by single events and/or within small model systems. Usually goal is to sample equilibrium ensembles, not simulate individual particles!

• Example: Free energies of solvation for amino acids often have errors ~1kJ/mol. It is unlikely to calculate binding free energies more accurately than this. MD can only give free energy differences!

• Limited polarization effects; waters can reorient, but partial charges are fixed. Multipolar force fields are needed for more accurate representation such processes.
Example Properties

Ensemble average:

• Free energy (of binding, solvation, interaction) differences
• Diffusion coefficients, viscosity
• Reaction rates, phase transition properties
• Protein folding times
• Structure refinement

Single events for strongly non-equilibrium processes:

• Radiation damage
• Surface coating
• Nucleation (metadynamics)
• Just because you see something in a simulation does NOT mean it is real!

• Equilibrium properties are all about statistics!

• When you’ve seen it 10 times it’s significant - a single event is often not!

• You should always try to calculate error estimates for predicted properties!
Each force field has their own water potential parameters that works best (AMBER & Charmm: TIP3P). Choose the one that maches.
Avoid System Size Effects

Small system:
Number of water molecules: $10 \times 10 \times 10 = 1000$
Number of ‘surface’ waters: $1000 - 8 \times 8 \times 8 = 488$
or 48% of the bulk

Large system:
Number of water molecules: $10^2 \times 10^2 \times 10^2 = 10^6$
Number of ‘surface’ waters: $10^6 - (10^2-2)^3 \sim 6 \times 10^4$
or ~6% of the bulk
Our model systems are still too small especially with respect to Avogadro’s number! To avoid surface over bulk domination effects we resort to periodic boundaries, pretending that boundaries don’t exist (like in pacman)!

2D cubic periodic
3D Periodic Boundaries

- Triclinic
- Hexagonal prism
- Truncated octahedron
- Rhombic dodecahedron
Boundary Conditions

- **None** – biopolymer simulations
- **Stochastic boundaries** – biopolymers
- **Hard wall boundaries** – pores, capillaries
- **Periodic boundaries** – most MD simulations
Periodic Boundaries – Consequences

- The system no longer has a *surface*.

- The system becomes *pseudo-periodic* (used to advantage for Ewald sums). It is incorrect to impose cutoff on long range interactions. However, the Ewald method can deal with infinite number of periodic images – specifically Particle Mesh Ewald (PME).

- Correlations in space beyond *half cell width* (L/2) are artificial. For this reason, the cut-off $r_{\text{cut}}$ is usually no greater than L/2.

- Correlations in time beyond $t=\frac{L}{c}$ are (in principle) subject to *recurrence*. In practice this does not seem to be the case.

- Use with *Minimum Image* convention.
Minimum Image Convention

Use $r_{ij'}$ not $r_{ij}$

$$x_{ij} = x_{ij} - L \cdot Nint(x_{ij}/L)$$

$Nint(a) =$ nearest integer to $a$
Interactions – I

- **Bonded interactions** – also referred as *intra*-molecular
  
  Bond vibration
  
  Angle vibration
  
  Torsion potentials

- **Non-bonded interactions** – also referred as *inter*-molecular
  
  
  van der Waals interactions
  \[ U \propto \frac{1}{r^{12}} - \frac{1}{r^6} \]  
  short-ranged

  Electrostatics
  \[ U \propto \frac{q_1 q_2}{r} \]  
  long-ranged
Interactions – II

Point ions and atoms

Polarisable ions (core+shell)

Flexible molecules

Constraint bonds

Flexible linked rigid molecules

Rigid molecules

Rigid bond linked rigid molecules
Different ‘atom types’ for a given element (e.g. carbon) may have different partial charges depending on the environment. Force-fields and force-field builders do account for this!
A set of interactions constitutes a force field (FF)

Do not mix different FFs!
\[ V(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) = \sum_{i,j}^N U_{\text{pair}}(|\vec{r}_i - \vec{r}_j|) + \frac{1}{4\pi\varepsilon_0} \sum_{i,j}^N \frac{q_i q_j}{|\vec{r}_i - \vec{r}_j|} + \]

\[ \sum_{i,j,k}^{N'} U_{\text{Tersoff}}(\vec{r}_i, \vec{r}_j, \vec{r}_k) + \sum_{i,j,k}^{N'} U_{3\text{-body}}(\vec{r}_i, \vec{r}_j, \vec{r}_k) + \sum_{i,j,k,n}^{N'} U_{4\text{-body}}(\vec{r}_i, \vec{r}_j, \vec{r}_k, \vec{r}_n) + \]

\[ \varepsilon_{\text{metal}} \left( \sum_{i,j}^{N'} V_{\text{pair}}(|\vec{r}_i - \vec{r}_j|) + \sum_{i}^{N} F\left( \sum_{i,j}^{N'} \rho_{ij}(|\vec{r}_i - \vec{r}_j|) \right) \right) + \]

\[ \sum_{i_{\text{bond}}}^{N_{\text{bond}}} U_{\text{bond}}(i_{\text{bond}}, \vec{r}_a, \vec{r}_b) + \sum_{i_{\text{angle}}}^{N_{\text{angle}}} U_{\text{angle}}(i_{\text{angle}}, \vec{r}_a, \vec{r}_b, \vec{r}_c) + \]

\[ \sum_{i_{\text{dihed}}}^{N_{\text{dihed}}} U_{\text{dihed}}(i_{\text{dihed}}, \vec{r}_a, \vec{r}_b, \vec{r}_c, \vec{r}_d) + \sum_{i_{\text{invers}}}^{N_{\text{invers}}} U_{\text{invers}}(i_{\text{invers}}, \vec{r}_a, \vec{r}_b, \vec{r}_c, \vec{r}_d) + \]

\[ \sum_{i_{\text{tether}}}^{N_{\text{tether}}} U_{\text{tether}}(i_{\text{tether}}, \vec{r}_t, \vec{r}_{t=0}) + \sum_{i_{\text{core-shell}}}^{N_{\text{core-shell}}} U_{\text{core-shell}}(i_{\text{core-shell}}, |\vec{r}_i - \vec{r}_j|) + \sum_{i=1}^{N} \Phi_{\text{external}}(\vec{r}_i) \]
Example: Simulation of Ar

**Pair Potential:**

\[ \Phi (\mathbf{r}) \begin{cases} 0 & (r < r_{\text{cut}}) \\ -\frac{1}{2} \epsilon (r - r_{\text{cut}})^2 & (r \geq r_{\text{cut}}) \end{cases} \]

Models the Pauli exclusion principle (repulsive) and van der Waals forces (attractive)

**Lagrangian:**

\[ L = \sum_{i<j} \Phi (\mathbf{r}_{ij}) \]
Lennard-Jones Potential

Models the Pauli exclusion principle (repulsive) at short distances & the van der Waals forces (attractive) at long ones.

\[ V(r) = 4 \varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \]

The pair-wise radial distance, \( r_{\text{cut}} \), is usually chosen at 3-5 times \( \sigma \) - 8-15 Å.
Equations of Motion

Lagrange Equation – time evolution

\[
\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{\alpha}_i} \right) - \frac{\partial L}{\partial \alpha_i} = 0
\]

Newton

\[
m_i \ddot{a}_i = \vec{F}_i
\]

Force Evaluation – particle interactions

\[
\vec{F}_i = \sum_{j \neq i}^{N} \vec{f}_{ij}
\]

\[
\vec{f}_{ij} = -\vec{\nabla}_i V (r_{ij})
\]
Ewald Summation

The method offers an elegant solution to solving the full electrostatic problem by splitting it in two parts – one in real space and one in reciprocal space. In real space, complying with the cutoff concept, a convenient screening function is added around all charges to make their interactions decay very fast at $r_{\text{cut}}$. The added screening functions can be subtracted in reciprocal space due to the periodic boundary condition by using Fourier transforms.
Potential Cut-off: Consequences

- There is a **step** in the potential at $r_{\text{cut}}$, which gives rise to artificial energy fluctuations.
- There is a **discontinuity** in the force at $r_{\text{cut}}$, which, if large enough, will show up in the space and force correlations.
- The above problems can be fixed by using shifted force potentials or by tapering the potential at short distance before the cutoff to the same effect.
- A long-range correction is required for both potential energy and virial in condensed phase studies.
Key Stages in a Simulation

- Set up initial system
- Calculate atomic forces
- Calculate atomic motion
- Calculate physical properties
- Repeat!
- Produce final summary
Integration Algorithms

**Essential Requirements:**

- Computational speed
- Low memory demand
- Accuracy
- Stability (energy conservation, no drift)
- Useful property - time reversibility
- Extremely useful property - Symplecticness
Integration: Essential Idea

\[ \mathbf{r}(t) \rightarrow \mathbf{r}(t+\Delta t) \]

\[ \mathbf{v}(t) \Delta t \]

\[ f(t)\Delta t^2/2m \]

\[ \mathbf{r}''(t+\Delta t) \]

\[ [\mathbf{r}(t), \mathbf{v}(t), f(t)] \rightarrow [\mathbf{r}(t+\Delta t), \mathbf{v}(t+\Delta t), f(t+\Delta t)] \]
Integration Algorithms: Velocity Verlet

\[ \mathbf{r}_{n+1} = \mathbf{r}_n - \Delta t \mathbf{\dot{r}}_n + \frac{\Delta^2}{2m_i} (\mathbf{F}_n + \mathbf{Q} \mathbf{\dot{r}}_n^2) \]

\[ \mathbf{\dot{r}}_{n+1} = \mathbf{\dot{r}}_n + \frac{\Delta}{2m_i} (\mathbf{F}_n + \mathbf{F}_{n+1}^2) + \mathbf{Q} \mathbf{\dot{r}}_n^2 \]

**Application in Practice**

\[ \mathbf{\dot{v}}_{i}^{n+1/2} = \mathbf{\dot{v}}_i^n + \frac{\Delta t}{2m_i} \mathbf{F}_i^n \]

\[ \mathbf{\dot{r}}_i = \mathbf{\dot{r}}_i^n + \Delta t \mathbf{\dot{v}}_{i}^{n+1/2} \]

\[ \mathbf{\dot{v}}_i^{n+1} = \mathbf{\dot{v}}_i^{n+1/2} + \frac{\Delta t}{2m_i} \mathbf{F}_i^{n+1} \]
Integration Algorithms: Variety

- **N** – number of particles, **V** – volume, **E** – energy,
- **T** – temperature (requires a coupled thermostat equation),
- **P** – pressure (requires a barostat equation)

- **NVE** – micro-canonical
- **NVT (E_{kin})** Evans
- **NVT** Andersen, Langevin, Berendsen, Nosé-Hoover, GST
- **NPT** Langevin, Berendsen, Nosé-Hoover, Martyna-Tuckerman-Klein
- **N_σT/NPnAT/NPnγT** Langevin, Berendsen, Nosé-Hoover, Martyna-Tuckerman-Klein
Integration Algorithms: Thermostat & Barostat

- Thermostats and barostats have their own time scales for coupling to the dynamics of the particles that need to be slow enough to control oscillations in temperature and pressure (volume) but not too slow in order to reach equilibrium at the targeted temperature and pressure in a reasonable amount of time.

- Typically relaxation constants for bio-chemical simulations are 0.1ps for the thermostat (T) and 5ps for the barostat (P)

beware of oscillations!
System Properties: Static (1)

• Kinetic Energy:

\[ KE = \frac{1}{2} \sum_{i=1}^{N} m_i v_i^2 \]

• Temperature:

\[ T = \frac{2}{3N} KE \]

• Configuration Energy:

\[ U_c = \sum_{i,j} V_{ij} \]

• Pressure:

\[ P = \frac{1}{3N} \sum_{i,j} F_{ij} \]

• Specific heat:

\[ C_v = \frac{3}{2} R A \]
**System Properties:**
**Static (2)**

- **Pair correlation (Radial Distribution Function):**
  
  \[
  \rho(r) = \sum_{ij} \delta(\Delta r_{ij}) \left( \frac{4\pi}{\rho} \right) \sin^2 \left( \frac{kr}{2} \right)
  \]

- **Structure factor:**
  
  \[
  S(k) = \int_{0}^{\infty} \rho(\rho) \sin(k\rho) \, d\rho
  \]

*Note: S(k) available from x-ray diffraction and neutron scattering*
Typical RDF
The bulk of these are in the form of Correlation Functions:

\[ \langle \hat{C}(t) \rangle = \frac{1}{T} \int_{-T}^{T} \langle f(t + \tau) \hat{f}(\tau) f(0) f(\tau) \rangle d\tau \]

or

\[ \langle \hat{C}(t) \rangle = \langle f(t)f(0) \rangle \langle f(\tau) f(\tau) \rangle^{2} \]
System Properties: Dynamic (2)

**Single correlation functions**

- **Mean squared displacement**
  (Einstein relation)
  \[
  2Dt = \frac{1}{3} \langle |r_i(t) - r_i(0)|^2 \rangle
  \]

- **Velocity Autocorrelation**
  (Green-Kubo relation)
  \[
  D = \frac{1}{3} \int_0^\infty \langle v_i(t) \cdot v_i(0) \rangle dt
  \]
Typical MSDs

\[ \langle |r(t) - r(0)|^2 \rangle \ (\text{A}^2) \]

\( t \) (ps)

Liquid

Solid
Typical VAF
Limitations of MD

• Parameters are imperfect
• Phase space is not sampled exhaustively
• Example: Free energies of solvation for amino acids often have errors ~1kJ/mol
• Likely impossible to calculate binding free energies more accurately than this
• Limited polarization effects; waters can reorient, but partial charges are fixed

However, MD simulations are cheaper than experiments as more easy to set up, repeat with changes to the models system, force-field and/or initial conditions and thus serve as an invaluable testing tool for scientists!
How to Use MD

• Think first and then simulate
• Ask specific questions
  • Which of a set of molecules binds best?
  • Which material recovers better, faster under irradiation?
• Simulate 10 models: do they act the same?
• Try to answer A/B type questions:
  • Does a His-Arg mutation affect stability?
  • Does replacement of Zr with Pb in pyrochlores affect the recovery processes?
Beyond the Classical MD scales

The atomic scale is not always best choice when trying to access phenomena occurring on a larger time- or length scales. There ways to use MD strengths and modify the equations of motion and reduce the degrees of freedom as in Disipative Particle Dynamics or just the do the latter use Coarse Grained MD. These simulations will be cheaper to run but some of the fine-grainness (chemistry and small scale details) will be lost as well as will the accurate timing. Particles will be a congregations of atoms and may even have a shape and interact via non-spherically symmetric potentials.

Another way to ask questions for rare processes and speed up time-scale (nucleation, vacancy hopping, defects annihilation, protein folding, crystal phases in emulsion formation) is to extend dynamics using advanced methodologies such as temperature accelerated dynamics, hyper-dynamics and meta-dynamics.

Think about what you want to get from the simulation: predictions or explanations!
Recommended Textbooks


