Analysis of Molecular Simulation Experiments

Nuno Micaêlo, PhD
micaelo@quimica.uminho.pt
http://simulation.quimica.uminho.pt

Universidade do Minho, 2010
Analysis of molecular dynamics experiments

- **Objectives**
  - Understand that simulation studies is mostly about planning experiences and analysing your results
  - Overview of some algorithms and analysis methods
  - Learn how to validate and ensure the quality of your simulations
Analysis of molecular dynamics experiments

- **Output from molecular dynamics simulations**
  - Atom coordinates trajectory (frequency 1 frame/ps)
  - Energy, pressure, volume, temperature (frequency: same as coordinates)
  - **REMEMER:** What we want is to have a collection of states (conformation + energy of each conformation)

- **Challenge:**
  - Explain macroscopic observations with data obtained from atomic molecular simulations
  - Discover novel phenomena difficult or impossible to observe experimentally
Analysis of molecular dynamics experiments

- **Visualization**
  - Gives you a first insight of your simulation
  - Use your favourite visualization software
    - Pymol (my favourite)
    - VMD
    - UCSF Chimera
    - Rasmol

- **Advices**
  - Stick to one (or two) software and learn all its features
Analysis of molecular dynamics experiments

- **Visualization**
  - Observe the “time evolution” of your system
  - See if geometries are as expected. This is important when you are developing parameters for new molecules
  - Large conformational changes can be easily observed
  - Short ranged interactions can be observed directly (hydrogen bonds, salt bridges)
  - Structure visualization is important, however you should try to quantify your observations
Analysis of molecular dynamics experiments

- **Potential energy**
  - It is the sum of all terms of the FF function. Check if your system is equilibrated by looking to the system’s potential energy, kinetic energy and total energy
Analysis of molecular dynamics experiments

- **Density**
  - Easily compared with experimental data
  - Some systems may need more or less time to converge towards the equilibrium
Analysis of molecular dynamics experiments

- **Atomic root mean square displacement (RMSD)**
  - Is an average distance between a conformation and a reference structure
  - RMSD will depend on the atoms used for fitting
  - Mass weighted RMSD

\[
RMSD = \sqrt{\frac{\sum_i m_i (r_i^{\text{ref}} - r_i)^2}{\sum_i m_1}}
\]
Analysis of molecular dynamics experiments

- Atomic root mean square displacement (RMSD)
Analysis of molecular dynamics experiments

- Atomic root mean square displacement (RMSD)
  - Molecules suffer conformational alterations (bond, angle and dihedral changes) but also whole molecule rotation and translation.
  - In most cases, we can eliminate the translation and rotational movement of our system using a fitting procedure.
  - The fitting methods translates and rotates the systems relative to a reference structure position and minimizes the RMSD between them.
Analysis of molecular dynamics experiments

- Fitting the conformation at time \( t=n \) against the initial structure \( t=0 \)
  - After fitting you can see more clear the conformational changes of your molecule
Analysis of molecular dynamics experiments

- **Atomic root mean square displacement (RMSD)**

Hen egg white lysozyme as a function of MD simulation.
1) Dashed line: simulation in *vacuum* using the GROMOS87 force field
2) Dot - Dashed line: simulation in *water* using the GROMOS87 force field
3) Solid line using the GROMOS96 force field

Analysis of molecular dynamics experiments

- **Radius of gyration (Rg)**
  - Is the root mean square distance of the atoms about the molecule x, y and z axes
  - It can give you a measurement of the size and “compactness” of your system
  - Can be correlated with X-ray or NMR experimental data

Hen egg white lysozyme as a function of MD simulation.

1) Dashed line: simulation in **vacuum** using the GROMOS87 force field
2) Dot - Dashed line: simulation in **water** using the GROMOS87 force field
3) Solid line using the **GROMOS96** force field

Analysis of molecular dynamics experiments

- **Radius of gyration (Rg) in one dimension**

\[
R_g = \frac{1}{N} \sum_{k=1}^{N} \left( r_k - r_{\text{mean}} \right)^2
\]

N is the number of atoms
Analysis of molecular dynamics experiments

- **Molecular Surface**
  - Important to understand many molecular processes and properties:
    - Folding, Solvation free energy, docking, continuum electrostatic calculations
  - Two concepts
    - Accessible surface
    - Molecular surface
Analysis of molecular dynamics experiments

- **Solvent accessible surface**
  - The solvent accessible surface is traced out by the probe sphere center as it rolls over the protein.
  - The surface is given by the center of the probe sphere.

Analysis of molecular dynamics experiments

- **Solvent accessible surface**

Hen egg white lysozyme as a function of MD simulation.

1) Dashed line: simulation in **vacuum** using the GROMOS87 force field
2) Dot - Dashed line: simulation in **water** using the GROMOS87 force field
3) Solid line using the **GROMOS96** force field

Analysis of molecular dynamics experiments

- **Molecular surface**
  - The molecular surface is the surface traced by the inward-facing surface of the probe sphere

Analysis of molecular dynamics experiments

- **Radial distribution function.** $g(r)$
  - Useful to describe the structure of a molecular system
  - The pair distribution function, $g(r)$, gives the probability of finding an atom at distance $r$ from another atom
  - The neighbors around each atom are sorted into distance bins, and averaged over the entire simulation
  - Can be compared with experimental data. X-ray, NMR, neutron diffraction

Analysis of molecular dynamics experiments

-Radial distribution function of the blue atoms around the red atom (1 angstrom shells).
  -The number of atoms found on each shell is normalised with the expected number of atoms in that slice.

\[ g(r) = \frac{n(r)}{\rho 4\pi r^2 \delta r} \]

<table>
<thead>
<tr>
<th>r</th>
<th>n(r)</th>
<th>g(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2.4</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2.1</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

- \( \rho \) - system density
- \( r \) - shell radius
- \( \delta r \) - shell thickness
Analysis of molecular dynamics experiments

- **Liquid and solid g(r)**
  - In the liquid state, it is two times more likely that two atoms are separated by 1.2 Å than in the ideal solution.
  - Solids have long range structural ordering.
Analysis of molecular dynamics experiments

- **Mean square displacement. MSD**
  - The MSD contains information on the atomic diffusivity
  - Provides an easy way to compute the diffusion constant

\[
\text{MSD} = \left\langle \left| \mathbf{r}(t) - \mathbf{r}(0) \right|^2 \right\rangle
\]

Analysis of molecular dynamics experiments

- **Diffusion coefficient.** $D$
  - If the system is liquid, MSD grows linearly with time
  - Diffusion is calculated using the Einstein relation
  - Can be compared with NMR data

\[
D = \lim_{x \to \infty} \frac{1}{6t} \left\langle |r(t) - r(0)|^2 \right\rangle
\]
Shear viscosity. $\eta$

- Resistance of a fluid that is being deformed by shear stress
- In simple liquids, results from the nonbonded intermolecular interactions
- Determined using the periodic perturbation method
- Can be compared with experimental data

Analysis of molecular dynamics experiments

- **Periodic perturbation method**
  - Non-equilibrium MD simulation
  - At each time step an external force in the x-direction is applied to each molecule
  - As a result, a spatially periodic velocity profile is developed

\[
a_x(z) = A \cos\left(\frac{2\pi z}{l_z}\right)
\]

Analysis of molecular dynamics experiments

- **Clustering**
  - Clustering is very useful to group ‘similar’ molecular structures
  - There is no correct algorithm for clustering
  - Clustering methods rely on ‘similarity’ (or dissimilarity) measures between structures
    - RMSD between pairs of structures
    - Torsion angle ‘distance’
Analysis of molecular dynamics experiments

- **Clustering methods**
  - **Single linkage**
    - Add a structure to a cluster when its distance to any element of the cluster is less than a specific cut-off (ex: 2 angstroms)
    - Fails in some cases

Analysis of molecular dynamics experiments

- **Clustering methods**
  - **Jarvis Patrick**
    - Uses a nearest neighbor approach
    - Two conformations are considered to be in the same cluster if they satisfy the following criteria:
      - They are in each other’s list of \( m \) nearest neighbors
      - They have \( p \) (\( p < m \)) nearest neighbors in common

Analysis of molecular dynamics experiments

- Clustering methods
  - Jarvis Patrick - Example
    - Three nearest neighbors \((m = 3)\)
    - Two out of three nearest neighbors in common \((p = 2)\)

Analysis of molecular dynamics experiments

- **Enthalpy of vaporization.** $\Delta H_{\text{vap}}$
  - Energy required to overcome the intermolecular interactions in the liquid state
  - Important to parameterize new liquids
  - Calculated from the potential energies in the gaseous ($U^{\text{gas}}$) and liquid phases ($U^{\text{liq}}$). Two simulations
  - Easily compared with experimental data

\[
\Delta H_{\text{vap}} = U^{\text{gas}}_{\text{pot}} - U^{\text{liq}}_{\text{pot}} + RT
\]
Analysis of molecular dynamics experiments

- **Root mean square fluctuation. RMSF**
  - Measure of the deviation between the position of the atoms and some reference structure
  - Locate regions with high/low mobility
Analysis of molecular dynamics experiments

- Root mean square fluctuation. RMSF

Active site loop
Analysis of molecular dynamics experiments

- B-factors. B
  - Calculated from RMSF
  - Comparison with X-ray B-factors

\[ B = \frac{8\pi^2}{3} \text{RMSF} \]
Analysis of molecular dynamics experiments

- **Hydrogen bonds**
  - Donor acceptor distance 3.5 Å. Angle 30°.
  - Responsible for many short range molecular properties

  Transition state of 1-phenyl-ethanol
  
  Transition state of 2-phenyl-propanol

  - Catalysis of R
  - NO catalysis of S!
  - Catalysis R and S
Analysis of molecular dynamics experiments

- **Correlation function.** $C_{xy}$
  - Determines if two properties are correlated
  - The correlation coefficient $c_{xy}$ between two data sets, $X$ and $Y$, of the same size with $M$ elements is given by:
    \[
    c_{xy} = \frac{1}{M} \sum_{i=1}^{M} x_i y_i \sqrt{\left( \frac{1}{M} \sum_{i=1}^{M} x_i^2 \right) \left( \frac{1}{M} \sum_{i=1}^{M} y_i^2 \right)} = \frac{\langle x_i y_i \rangle}{\sqrt{\langle x_i^2 \rangle \langle y_i^2 \rangle}}
    \]
  - $-1 \leq c_{xy} \leq 1$. A value of 0 indicates no correlation and an absolute value of 1 indicates high degree of correlation

Analysis of molecular dynamics experiments

- **Time correlation coefficients**
  - Determines if the value of one property at some instant \( y(0) \) is correlated with another property at latter time \( x(y) \)
  
  \[ C_{xy}(t) = \langle x(t)y(0) \rangle \]

  - If the \( x \) and \( y \) properties are different, \( C_{xy}(t) \) is referred as **cross-correlation function**
  - If the \( x \) and \( y \) properties are the same, \( C_{xx}(t) \) is referred as **autocorrelation function**

Analysis of molecular dynamics experiments

- Correlated data
  - Example

![Graph showing property vs. time](image-url)
Analysis of molecular dynamics experiments

- **Autocorrelation function**
  - Determines how well does the system “remember” its previous state.
  - $-1 \leq c_{xx} \leq 1$. Has initial value of 1 and decays to 0 ($t \rightarrow \infty$)
  - The time taken to loose the correlation ($c_{xx} = 0.1$) is called the **correlation time**, or the **relaxation time** ($t_{corr}$)

Analysis of molecular dynamics experiments

- **Correlation time/Relaxation time**
  - Your simulation must be longer than the correlation time of the property you are interested to measure.
  - Longer simulations provide better correlation time statistics by using different \((M)\) time origins \((t_j)\).

\[
c_{xx}(t) = \frac{1}{N} \sum_{j=0}^{M} \sum_{i=1}^{N} \frac{\langle x_i(t) \cdot x_i(t_j) \rangle}{\langle x_i(t_j) \cdot x_i(t_j) \rangle}
\]

Analysis of molecular dynamics experiments

- **Estimating errors**
  - As in real experiments, computer simulation experiments are subject to systematic and statistical errors.
  - Most of the time we want to calculate the average of some property B over a finite data set (energy, density).
  - Assuming that our data is normally distributed, we would like to know the variance of this distribution.
  - However, for some properties, data collected at very short intervals could be correlated. If this happens, variance will be underestimated.

Analysis of molecular dynamics experiments

- Estimating errors
  - Example

Correlated data

Uncorrelated data
Analysis of molecular dynamics experiments

- **Estimating errors**
  - In a trajectory with $M$ frames, if the value of property $X$ is statistically independent from the previous one, the variance of the mean is given by:

$$\sigma^2_{\text{mean}} = \frac{\sigma^2}{M}$$

where $\sigma^2$

$$\sigma^2 = \frac{1}{M} \sum_{t=1}^{M} (X(t) - \langle X \rangle)^2$$

Analysis of molecular dynamics experiments

- **Estimating errors**

  - However, if our data has a correlation time \( t_{\text{corr}} \) higher than 1, the variance is underestimated and must be corrected.

  - In this context, this means that only about one configuration in every \( t_{\text{corr}} \) steps recorded contributes with new information to the average.

\[
\sigma_{\text{mean}}^2 = \frac{\sigma^2}{M/t_{\text{corr}}}
\]

Analysis of molecular dynamics experiments

- **Other analysis**
  - The analysis of your data can be one of the most time-consuming steps of your research projects.
  - Most of the properties you want to analyze are “buried” in the simulation output (trajectory, energies).
  - Learn and study how microscopic properties can explain the observed experimental facts.
  - Write your own software analysis tools (python, awk, C).
Validating your simulations

- Establish a clear working hypothesis and research plan. Think first, then simulate
- You must ensure that your simulations are valid
- Just because the atoms move doesn't mean that the simulation is well done!
- If the result of a simulation is novel, unexpected or strange, it means that:
  - A new phenomenon has been found, AND/OR
  - The results are wrong

Validating your simulations

- The results could be wrong because:
  - The model is inappropriate
  - The force field is inadequate
  - The results have not converged due to poor sampling
  - The software contain bugs
  - The software was used incorrectly
Validating your simulations

- Agreement between simulation and experiment can be the result of:
  - The simulation adequately reflects the experimental system
  - The property examined is insensitive to the details of the simulation
  - Error compensation

Validating your simulations

- No agreement between simulation and experiment can be the result of:
  - The simulation does not reflect the experimental system
    - Model
    - Force field
    - Sampling/convergence
    - Software bug
    - Incorrect software usage
  - The experimental data are incorrect

Validating your simulations

- **Publishing**
  - Refer to the theory and methodologies used
  - Refer the force field used, version and modifications
  - Present the time evolution/correlation of key properties in order to judge the degree of sampling and convergence
  - Refer to the software used, version and modifications
  - Specify the chosen input parameters of the simulation