Molecular dynamics simulations

Methods in Molecular Biophysics, Spring 2009

Basics of molecular mechanics and dynamics
  Statistical mechanics of liquids
  Basic ideas of continuum solvation
  The MM/PBSA model
1901 (and earlier?) ball and stick models

Baird & Tatlock 1901
1950s: wire models of proteins
- separate nuclei and electrons
- polarisation, electron transfer and correlation
- can specify electronic state
- can calculate formation energies
- can do chemistry (bond breaking and making)
- variationally bound
- computationally expensive
- typically ~10-100 atoms
- dynamics ~1 ps

QM MOLECULE
Electrons
Nuclei
- no explicit electrons, net atomic charges
- no polarisation, electron transfer or correlation
- conformational energies for ground state
- no chemistry
- semi-empirical force fields
- not variationally bound
- solvent and counterion representations
- typically ~1000-100000 atoms
- dynamics up to ~100 ns

MM MOLECULE
Some force field assumptions

1. **Born-Oppenheimer approximation** (separate nuclear and electronic motion)
2. **Additivity** (separable energy terms)
3. **Transferability** (look at different conformations, different molecules)
4. **Empirical** (choose functional forms and parameters based on experiment)
What does a force field look like?

\[
U = \sum_{b\text{onds}} K_b(b - b_{eq})^2 + \sum_{\theta\text{ngles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{impropers}} K_w w^2
\]

\[
+ \sum_{\text{nonbonded pairs}} \left\{ 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] + \frac{q_i q_j}{r} \right\}
\] (1)

[Diagram of formamide and water molecules]
Lennard-Jones energy curve
Distance dependence

Electrostatic

Lennard-Jones

Energy (kcal/mol)

Rij (angstroms)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>H bonded to nitrogen atoms</td>
</tr>
<tr>
<td>HC</td>
<td>H aliph. bond. to C without electrwd. group</td>
</tr>
<tr>
<td>H1</td>
<td>H aliph. bond. to C with 1 electrwd. group</td>
</tr>
<tr>
<td>H2</td>
<td>H aliph. bond. to C with 2 electrwd. groups</td>
</tr>
<tr>
<td>H3</td>
<td>H aliph. bond. to C with 3 electrwd. groups</td>
</tr>
<tr>
<td>HA</td>
<td>H arom. bond. to C without electrwd. groups</td>
</tr>
<tr>
<td>H4</td>
<td>H arom. bond. to C with 1 electrwd. group</td>
</tr>
<tr>
<td>H5</td>
<td>H arom. bond. to C with 2 electrwd. groups</td>
</tr>
<tr>
<td>HO</td>
<td>hydroxyl group</td>
</tr>
<tr>
<td>HS</td>
<td>hydrogen bonded to sulphur</td>
</tr>
<tr>
<td>HW</td>
<td>H in TIP3P water</td>
</tr>
<tr>
<td>HP</td>
<td>H bonded to C next to positively charged gr</td>
</tr>
</tbody>
</table>

**AMBER parm94 H atom types**
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>sp2 C carbonyl group</td>
</tr>
<tr>
<td>CA</td>
<td>sp2 C pure aromatic (benzene)</td>
</tr>
<tr>
<td>CB</td>
<td>sp2 aromatic C, 5&amp;6 membered ring junction</td>
</tr>
<tr>
<td>CC</td>
<td>sp2 aromatic C, 5 memb. ring HIS</td>
</tr>
<tr>
<td>CK</td>
<td>sp2 C 5 memb. ring in purines</td>
</tr>
<tr>
<td>CM</td>
<td>sp2 C pyrimidines in pos. 5 &amp; 6</td>
</tr>
<tr>
<td>CN</td>
<td>sp2 C aromatic 5&amp;6 memb. ring junct.(TRP)</td>
</tr>
<tr>
<td>CQ</td>
<td>sp2 C in 5 mem. ring of purines between 2 N</td>
</tr>
<tr>
<td>CR</td>
<td>sp2 arom as CQ but in HIS</td>
</tr>
<tr>
<td>CT</td>
<td>sp3 aliphatic C</td>
</tr>
<tr>
<td>CV</td>
<td>sp2 arom. 5 memb. ring w/1 N and 1 H (HIS)</td>
</tr>
<tr>
<td>CW</td>
<td>sp2 arom. 5 memb. ring w/1 N-H and 1 H (HIS)</td>
</tr>
<tr>
<td>C*</td>
<td>sp2 arom. 5 memb. ring w/1 subst. (TRP)</td>
</tr>
</tbody>
</table>

AMBER parm94 C atom types
**Force fields in Amber**

- **ff94**: widely used ("Cornell et al."), pretty good nucleic acid, too much $\alpha$-helix for proteins
- **ff99**: major recalibration by Junmei Wang and others; basis for most current Amber ff’s
- **ff99SB**: recalibration of backbone potentials for proteins by Carlos Simmerling ("SB")
- **ff02r1**: polarizable extension for ff99
- **ff03**: new charge model (Yong Duan) + backbone torsions for proteins
- **ff03ua**: united atom extension
Periodic boundary conditions
Basics of the Ewald approach

direct, short-ranged
smooth, use FFT
Minimization and simulated annealing
The Simplex algorithm
Eq. (1) is the original Verlet propagation algorithm; Eqs. 2 and 3 are the “leap-frog” version of that. Remember that

\[ a = \frac{d^2x}{dt^2} = F/m = \frac{\partial V/\partial x}{m}. \]

See pp. 42-47 in Becker & Watanabe.
Regulating temperature

“Temperature” is a measure of the mean kinetic energy. The instantaneous KE is

\[ T(t) = \frac{1}{k_B N_{dof}} \sum_{i} m_i v_i^2 \]

(cf. classical rule of thumb: “\(k_B T/2\) of energy for every squared degree of freedom in the Hamiltonian”)

Suppose the temperature is not what you want. At each step, you could scale the velocities by:

\[ \lambda = \left[ 1 + \frac{\hbar}{2\tau} \left( \frac{T_0}{T(t)} - 1 \right) \right]^{1/2} \]

This is the “Berendsen” or “weak-coupling” formula, that has a minimal disruption on Newton’s equations of motion. But it does not guarantee a canonical distribution of positions and velocities. See Morishita, J. Chem. Phys. 113:2976, 2000; and Mudi and Chakravarty, Mol. Phys. 102:681, 2004.
Consider the stochastic, Langevin equation:

\[ \frac{dv}{dt} = -\zeta v + A(t) \]

By Stokes’ law, the friction coefficient is related to the viscosity of the environment: \( \zeta = 6\pi a \eta / m \). At long times, we want this system to go to equilibrium at a temperature \( T \), which is a Maxwell-Boltzmann distribution:

\[ W(v, t; v_0) \sim \exp \left[ -\frac{mv^2}{2k_B T} \right] \]

for every value of \( v_0 \). This places restraints on the properties of the stochastic force \( A(t) \). It can be shown that

\[ \zeta = (\beta / m) < A^2 > \]

where we have assumed that \( < A > = 0 \) and \( < A(0)A(t) > = < A^2 > \delta(t) \).
More coarse-grained potentials

United atom approximation

Remove non-polar hydrogens

H - C

H - C - H

H - C - H

H

C'

C''

C'''

United atom approximation
Go model for protein folding

- square-well potential
- native contacts "+1"
- non-native contacts "-1"
- cannot represent frustration during folding
Gaussian network model

\[ r_{ij}^n < R_{cut} \]

\[ E_{ij} = k(r_{ij} - r_{ij}^n)^2 \]
Getting free energies

\[ \Delta A = -k_B T \ln \frac{\rho(B)}{\rho(A)} \]

\[ W = -k_B T \ln \rho(\delta) \]  \hspace{1cm} (5)
Knowledge-based potentials

- Start from set of known protein structures
- Assume energy can be decomposed into residue pair interactions
- Assume that frequency of interactions within the ensemble ≈ frequency of interactions within the equilibrium ensemble of a single protein
- Derive potential of mean force for residue pairs from observed occurrence probabilities
- Knowledge-based potentials are used in both threading and folding

Free energy profiles

\[ \rho(\delta) = \frac{\int \exp(-\beta U) d\Sigma}{\int \exp(-\beta U) d\delta d\Sigma} \]  

(6)

Here \( \beta = 1/k_B T \) and \( d\Sigma \) represents an integration over all remaining degrees of freedom except \( \delta \). Now add a biasing potential \( U^*(\delta) \) which depends only upon \( \delta \):

\[
\begin{align*}
\rho^*(\delta) &= \exp[-\beta U^*(\delta)] \frac{\int \exp(-\beta U) d\Sigma}{\int \exp(-\beta [U + U^*]) d\delta d\Sigma} \\
&= \rho(\delta) \exp[-\beta U^*(\delta)] / \langle \exp(-\beta U^*) \rangle 
\end{align*}
\]

(7)

\[
\langle \exp(-\beta U^*) \rangle = \frac{\int \exp(-\beta U^*) \exp(-\beta U) d\delta d\Sigma}{\int \exp(-\beta U) d\delta d\Sigma} 
\]

(8)

Taking logarithms, the potential of mean force in the presence of the umbrella potential, \( W^* \), is related to that in an unbiased simulation by:

\[
W^*(\delta) = W(\delta) + U^*(\delta) - C' 
\]

(9)

where \( C' = -k_B T \ln \langle \exp(-\beta U^*) \rangle \) is a constant independent of \( \delta \).
Example of explicit solvation setup
Basic ideas of continuum solvent models


Born Approximation: (1929)

\[ \Delta W = - \frac{1}{2} \left( 1 - \frac{1}{\varepsilon} \right) \frac{q^2}{\rho} \]
Conductor-like Screening Model

\[ E = E_{\text{gas}} + \int \frac{z}{r-r'} \, q + \frac{1}{2} \int \frac{1}{r-r'} \, q' \]

\[ = E_{\text{gas}} + zB_q + \frac{1}{2} q A_q \]

\[ \frac{dE}{dq} = 0 \Rightarrow A_q = -B_q \quad \text{or} \quad q = -\tilde{A} \tilde{B} \]

molecule-solvent interaction:

\[ -z B \tilde{A}' B \tilde{B} = -z \phi^{RF} \]

solvent-solvent interaction:

\[ \frac{1}{2} z B \tilde{A} \tilde{A}' B \tilde{B} = \frac{1}{2} z B \tilde{A} \tilde{B} \tilde{B} \]
Defining the continuum solvent model

Simplest model has “high” $\varepsilon_{\text{ext}}$ outside (white) and “low” $\varepsilon_{\text{in}}$ where solvent is excluded:
Generalized Born model

The solvation energy can be computed by quadrature if one adopts the Coulomb field approximation:

\[ W = \frac{1}{8\pi} \int \mathbf{E} \cdot \mathbf{D} dV = \frac{1}{8\pi} \left[ \int_{\text{in}} \frac{q^2}{\varepsilon_{\text{in}} r^4} dV + \int_{\text{ext}} \frac{q^2}{\varepsilon_{\text{ext}} r^4} dV \right] \]

\[ \Delta G = W(\varepsilon_{\text{ext}} = 80) - W(\varepsilon_{\text{ext}} = 1) \]

\[ \Delta G_{GB} = -\frac{1}{2} \left( 1 - \frac{1}{\varepsilon_{\text{ext}}} \right) \frac{q^2}{R_{\text{eff}}} ; \quad \text{or} \quad -\frac{1}{2} \left( 1 - \frac{1}{\varepsilon_{\text{ext}}} \right) \frac{q_i q_j}{f_{GB}(R_{\text{eff}}^i, R_{\text{eff}}^j, r_{ij})} \]

\[ R_{\text{eff}}^{-1} = \frac{1}{4\pi} \int_{\text{ext}} r^{-4} dV \]

Effects of added salt

\[
\left( 1 - \frac{1}{\varepsilon} \right) \rightarrow \left( 1 - \frac{e^{-\kappa G_B(d_{ij})}}{\varepsilon} \right)
\]

B-A energy differences for r,d(CCAACGTTGG)$_2$

<table>
<thead>
<tr>
<th></th>
<th>DNA</th>
<th>RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coulomb</td>
<td>-293.0</td>
<td>-266.9</td>
</tr>
<tr>
<td>PB</td>
<td>286.6</td>
<td>240.2</td>
</tr>
<tr>
<td>GB</td>
<td>288.1</td>
<td>242.2</td>
</tr>
<tr>
<td>vdw</td>
<td>-7.7</td>
<td>18.7</td>
</tr>
<tr>
<td>bad</td>
<td>-7.0</td>
<td>17.6</td>
</tr>
<tr>
<td>- TΔS</td>
<td>2.9</td>
<td>0.5</td>
</tr>
<tr>
<td>total</td>
<td>-21.0</td>
<td>9.8</td>
</tr>
<tr>
<td>0.1M salt</td>
<td>5.2</td>
<td>3.4</td>
</tr>
<tr>
<td>1.0M salt</td>
<td>6.0</td>
<td>3.9</td>
</tr>
</tbody>
</table>

Srinivasan, Cheatham, Kollman, Case, JACS 120, 9401 (1998)