Applications – MD / MC

Basic tools
• Force field
• MD / MC

Some application areas
• timescales
• free energy calculations
• simulated annealing
• structure refinement
Simulating dynamics (optimistic / naïve)

Claim
• protein has a hinge which must open to bind ligand

Can one see rates?
• rates for different ligands?
Timescales

Most common quantity $\tau$
- time to rotate by 1 radian
- time for decay in $A(t) = A(0)e^{-\frac{t}{\tau}}$
  - relaxation time
  - characteristic time
- times in proteins...
# Typical times in proteins

<table>
<thead>
<tr>
<th>Event</th>
<th>Amplitude (Å)</th>
<th>$\log_{10} \tau$(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bond vibration</td>
<td>0.01 – 0.1</td>
<td>−14 to −13</td>
</tr>
<tr>
<td>rotation of surface sidechain</td>
<td>5 – 10</td>
<td>−11 to −10</td>
</tr>
<tr>
<td>protein hinge bending</td>
<td>1 – 20</td>
<td>−11 to −7</td>
</tr>
<tr>
<td>rotation of sidechain in middle of a protein</td>
<td>5</td>
<td>−4 to 0</td>
</tr>
<tr>
<td>local loss of protein structure</td>
<td>5 – 10</td>
<td>−5 to +1</td>
</tr>
</tbody>
</table>

Timescales, simulations, statistics

Typical big simulation $\approx 100 \text{ ns} = 10^{-7}\text{s}$
- Imagine event with characteristic time $10^{-7}\text{s}$ - may or may not be seen
Consider time $10^{-8}\text{s}$
- may be seen a few times
What you would like - 100's or 1000's of observations

<table>
<thead>
<tr>
<th>Events</th>
<th>Condition</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>fast events</td>
<td>$\tau &lt; t_{\text{simulation}}$</td>
<td>OK</td>
</tr>
<tr>
<td></td>
<td>$\tau &lt; t_{\text{simulation}}$</td>
<td>poor statistics</td>
</tr>
<tr>
<td>slower events</td>
<td>$t \approx t_{\text{simulation}}$</td>
<td>no idea / very bad statistics</td>
</tr>
</tbody>
</table>

Previous example (drug binding)
- it is not enough to observe an event once (or few times)
Free Energy Calculations

\[ k_d = \frac{[\text{drug}][\text{protein}]}{[\text{drug}−\text{protein}]} = \frac{[D][P]}{[DP]} \]

\[ = e^{\frac{-\Delta G}{RT}} \]

Contributing terms?
- ligand-water → ligand + water  (many interactions, \( \Delta S \))
- ligand+protein
- ligand loss of entropy / water entropy change
  - simulate?
Infinite time – free energy estimate

\[ \text{DP} \rightleftharpoons D + P \]

\[ \Delta G = kT \ln \frac{[D][P]}{[DP]} \]

Very simple - simulate for long time

- Ligand (drug) goes on and off protein
- Look at [D], [P] and [DP] - calculate \( \Delta G \) directly from concentrations

Will not work – cannot simulate long enough

Coming philosophy

- \( \text{DP} \rightleftharpoons D + P \) is too hard, find an alternative
Free simulation for binding

If we simulate, where will the ligand go?
What is the shape of the energy landscape?

May take years for ligand to find protein

Short cut?
- force ligand to protein
  - artificial force + corrections
  - very difficult – still requires rearranging water
  - entropy estimation very difficult
Estimating free energy differences

\[ G = U - TS \]

but \[ S = -k \sum_{i=1}^{N_{state}} p_i \ln p_i \]

- so we cannot really get \( S \)
- similar problem – especially visiting high energy regions

Forget absolute free energies
- concentrate on \( \Delta G \)
- no problem – usually interesting property
Summarise free energy problem so far

• Sounds easy, just estimate [D], [P], [DP] – will not work – no simulation long enough
• Cheat – push ligand in? System not at equilibrium, requires work
• Chemically difficult – lots of interactions
  • requires completely changing water configuration
  • breaking ligand-water interactions, finding the correct ligand-protein binding
  • big change in solvent entropy, ligand entropy, protein entropy

How can one minimise the problems?
• do an easier problem (soon)

First - small detour on work
Work and free energy changes

work done A to B
• free energy change
  • look at either state
    • real world automatically includes entropy

Work going from unbound \(\rightarrow\) bound
• \(\Delta G_{AB}\)
• what is B? what is A?
  • more later
• measuring work?
Measure the work needed to move from $A$ to $B$

\[ \Delta G = \frac{\partial \mathcal{H}}{\partial \lambda} \int_A^B \lambda d\lambda \quad \text{or} \quad \Delta G = \sum_{i=0}^{N_{\text{step}}} (H_{i+1} - H_i) \]

where $\mathcal{H}$ is again Hamiltonian ($E_{\text{kin}} + E_{\text{pot}}$)
Binding energy - feasibility

Would this approach work?

\[ \langle \frac{\partial H}{\partial \lambda} \rangle \] must be a good average (lots of fluctuations)

must change \( \lambda \) slowly

Chemistry problems: your simulation would

• get averages with all water molecules
• gradually remove water molecules (high energy?)
• find the correct binding
• get good averaging there

• states A and B are very different they must be well sampled
• intermediate (higher energy states) must also be sampled
• does not work well in practice
Paths / Energy differences (detour)

Problem – the path is too difficult – changes too big
• Energy differences depend on end states – not paths
• Look at $\Delta E_{1,2} = E_1 - E_2$
  • would it matter if we go $E_1 \rightarrow E_3 \rightarrow E_2$?
Can we take even stranger paths?
• go through non existent $E_4$?
  • no problem

Same reasoning applies to free energies
Applying different paths

Originally wanted (ligand A or B, protein P)

\[ A + P \leftrightarrow AP \quad \Delta G_A \]
\[ B + P \leftrightarrow BP \quad \Delta G_B \]

If I know \( \Delta G_B \)

\( \Delta \Delta G_{AB} \) is easier

\[ \Delta \Delta G_{AB} = \Delta G_A - \Delta G_B \]

What would \( \Delta \Delta G_{AB} \) mean?

• relative binding strength
Alternative routes

\( \Delta G_A \) and \( \Delta G_B \) too hard

- we would be happy with \( \Delta \Delta G_{AB} \)

\[
\Delta G_A + \Delta G_Y = \Delta G_B + \Delta G_X \\
\Delta G_A - \Delta G_B = \Delta G_X - \Delta G_Y \quad \text{remember} \quad \Delta \Delta G_{AB} = \Delta G_A - \Delta G_B
\]

So \( \Delta \Delta G_{AB} = \Delta \Delta G_{XY} \)

- why \( \Delta G_X \) easier ?
- why \( \Delta G_Y \) easier ?

A + P \( \xrightarrow{\Delta G_A} \) AP

\( \Delta G_X \) \|

B + P \( \xrightarrow{\Delta G_B} \) BP

\( \Delta G_Y \)
Easier free energy changes

if A/B are rather similar
  AP ↔ BP or
  B + P ↔ A + P               (free A ↔ B       forget the protein)

are small changes – smaller than
  • removing water order, removing water energy, finding protein...

Example
  • small change
Fictitious states

Remember formulae

\[ \Delta G = \int_{A}^{B} \left( \frac{\partial \mathcal{H}(p,r)}{\partial \lambda} \right)_{\lambda} d\lambda \quad \text{and} \quad \Delta G = \sum_{i=0}^{N_{\text{step}}} (H_{i+1} - H_{i}) \]

make chemistry a function of \( \lambda \)
λ dependence

• λ = 0  
  an OH group

• λ = 1  
  an OCH$_3$ group

• λ = 0.5  
  • charge of H – half of original charge
  • radius / size (σ, ε) half of real value and so on

Atoms gradually
  • appear in one direction
  • disappear in other

Description of system is now function of λ
Two simulations necessary
• $\lambda$ from 0.0 $\leftrightarrow$ 1.0 in protein
• $\lambda$ from 0.0 $\leftrightarrow$ 1.0 in water
• both from red $\leftrightarrow$ blue

As $\lambda$ slowly moves from 0.0
• water gradually feels more/less influence of some atoms
• system should not have to rearrange itself too much

When does method work best?
• when changes are small
  • comparison of similar ligands in a protein
Summary of free energy calculations

From first principles: free energy differences, equilibria
- easy to calculate
- in practice impossible (sampling not possible)

Forget absolute free energies
- $\Delta G$ determine most phenomena in the world

Processes like binding still too difficult to simulate
- slow, too many conformations / states to visit

Most calculations use $\Delta \Delta G$
- aim to get relative binding strengths
Simulated Annealing

Classic reference – in stine

Basic tools

• MC or MD
  • with control of temperature (temperature bath)

Use: difficult optimisation problem

• chip layout
• travelling salesman problem
• protein structure

Optimisation problem

• several dimensional (2 to 2,000)
• many local minima

Procedure

while \( T > T_{\text{end}} \)
  \[ T(t) = T_0 \, e^{-ct} \]
  move system (Monte Carlo)

- \( T_0 \) initial temperature is hot
- \( c \) is decay rate (cooling of system)
- cost function is
  - \( E_{\text{pot}} \) in chemistry
  - path length in travelling salesman
  - board cost in chip layout problem ...
- why may this work?
Simulated Annealing concept

initial (poor) guess

initial high $T$ distribution

cooler $T$

cold $T$
Properties, practical issues

Admit that there may not be a best solution
• not worth spending effort between many very good solutions

Some problems have "phase transitions"

How hot should \( T_0 \) be ?
• infinite ? No : look at barriers

How slow should cooling be (\( c \)) ?
• system should be at equilibrium
• very slow

Cool exponentially ?
• best first guess
• should certainly cool more slowly at transition points
Anneal with MC or MD?

Historic use of Monte Carlo
• easiest to apply to many problems

Use MD?
• provides expected advantages (efficiency)
• uses available gradient / derivative information

Implementation
• Couple to temperature bath, make T time dependent

Use in practice?
• simulated annealing in
  • most MD codes, refinement packages, ...
Refinement of Structures (NMR / X-ray)

• Story from first semester
Problem: generate protein coordinates from NMR information (or X-ray)
• distance geometry gives an initial guess, but
  • distance geometry methods spread error across all distances
  • errors are spread across bonds, measured distances
  • chirality may be broken (causes distance problems)

Belief
• coordinates are not bad, but could be improved
Pseudo – energy terms

For some distance measurement $i$ between some pair of atoms

- $r_0$ measured distance
- $r(t)$ distance between particles at time $(t)$
- say $U_i(r) = c_i (r(t) - r_0)^2$
- add this to normal force field

$$U_{tot}(\mathbf{r}) = U_{phys}(\mathbf{r}) + \sum_{i=1}^{N_{restraints}} U_i(\mathbf{r})$$

$U_{phys}(\mathbf{r})$ normal force field - atomistic (bonds, electrostatics...)

Andrew Torda
26/05/2015 [28]
result?

System moves to low energy + low fake energy
• gradually moves to agree with experimental data

Practical issues

\[ U_{tot}(\mathbf{r}) = U_{phys}(\mathbf{r}) + \sum_{i=1}^{N_{restraints}} U_i(\mathbf{r}) \]

\[ U_i(\mathbf{r}) = c_i (r(t) - r_0)^2 \]

• big \( c \) very artificial
• small \( c \) system will be slightly biased to agree with experimental data
Fake Energies - examples

Refinement of
• X-ray structures (common)
• NMR (often)
• others: microwave spectroscopy, ...

Modelling problems
• you want to put a bond in a model
  • putting it in directly
    • high energy bond
    • system stuck in minimum
  • introduce a distance restraint
    • gradually increase associated constant $c$
Summary

What one can do with related methods

• look at timescales of motions (very superficial)
• free energy calculations – important for problems such as binding of ligands
• simulated annealing – methods used as minimizers, not necessarily to get an ensemble
• pseudo-(potential) energies (X-ray, NMR, …)