Applications – MD / MC

Basic tools

- Force field
- MD / MC

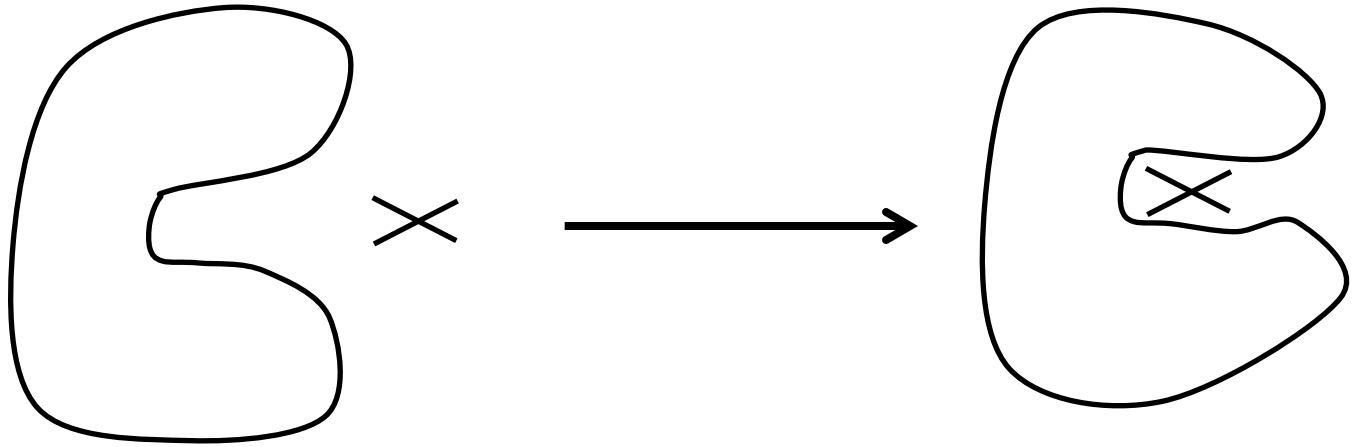
Some application areas

- 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 τ

- 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

	Amplitude (Å)	$\log_{10} \tau$(s)
bond vibration	0.01 – 0.1	–14 to –13
rotation of surface sidechain	5 – 10	–11 to –10
protein hinge bending	1 – 20	–11 to –7
rotation of sidechain in middle of a protein	5	–4 to 0
local loss of protein structure	5 – 10	–5 to +1

Numbers taken from McCammon, J.A. and Harvey, S.C., Dynamics of proteins and nucleic acids, Cambridge Uni Press, 1987

Timescales, simulations, statistics

Typical big simulation ≈ 100 ns = 10^{-7}s

- Imagine event with characteristic time 10^{-7}s - may or may not be seen

Consider time 10^{-8} s

- may be seen a few times

What you would like - 100's or 1000's of observations

fast events	$\tau \ll t_{simulation}$	OK
	$\tau < t_{simulation}$	poor statistics
slower events	$t \approx t_{simulation}$	no idea / very bad statistics

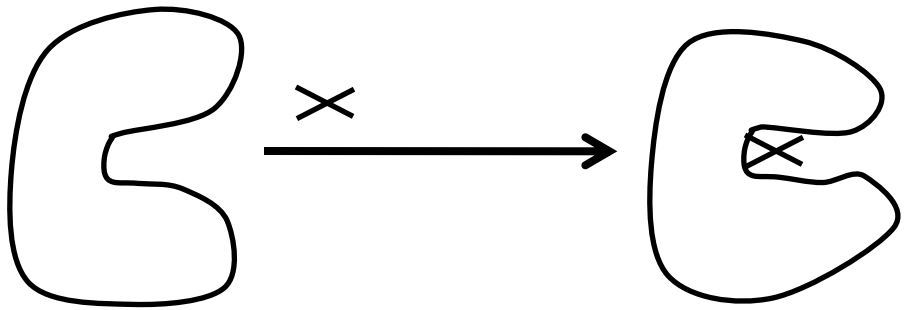
Previous example (drug binding)

- it is not enough to observe an event once (or few times)

Free Energy Calculations

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

$$= e^{\frac{-\Delta G}{RT}}$$



Contributing terms ?

- ligand-water → ligand + water (many interactions, ΔS)
- ligand+protein
- ligand loss of entropy / water entropy change
  - simulate ?

Infinite time – free energy estimate

$$DP \rightleftharpoons D + P$$

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

Very simple - simulate for long time

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

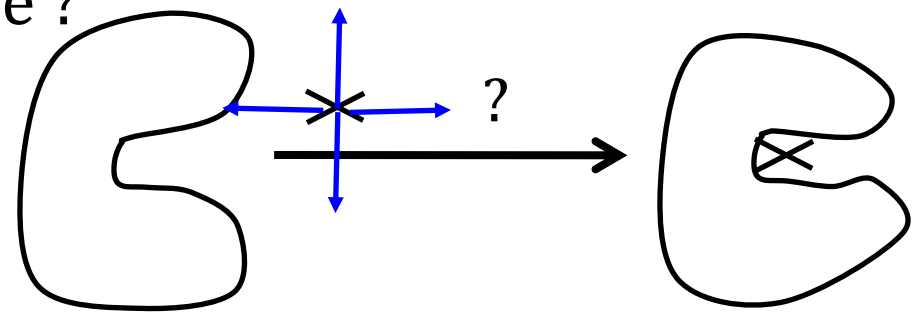
Will not work – cannot simulate long enough

Coming philosophy

- $DP \rightleftharpoons D + P$ 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

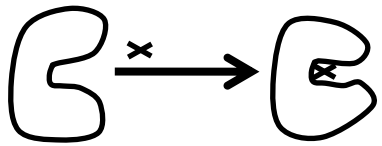
$G = U - TS$

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 ΔG
- 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 interations
  - 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 ?

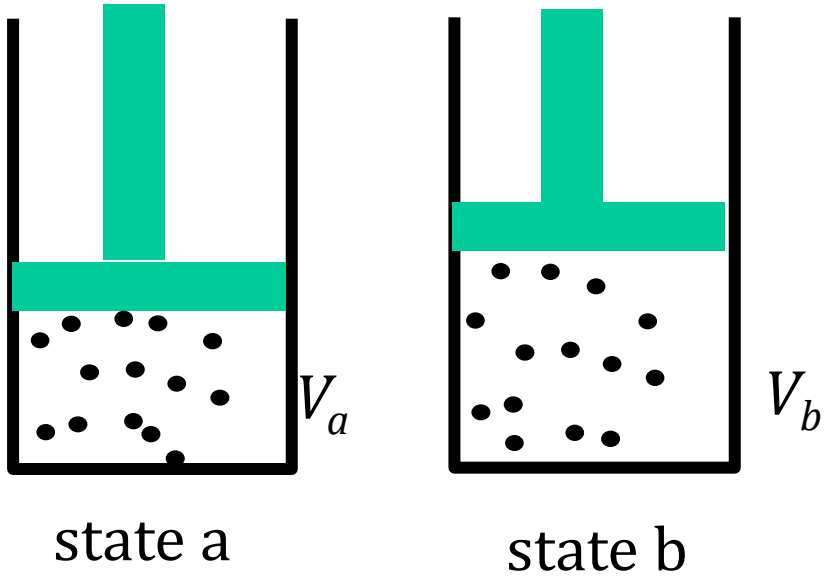
- do an easier problem (soon)

First - small detour on work

Work and free energy changes

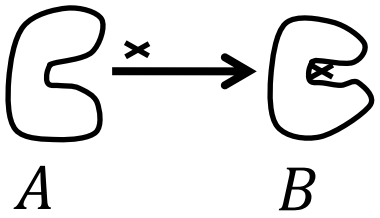
work done A to B

- free energy change
  - look at either state
    - real world automatically includes entropy

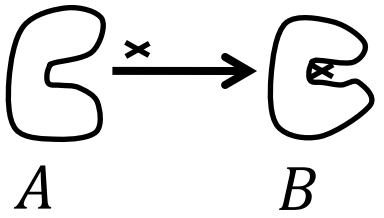


Work going from unbound →bound

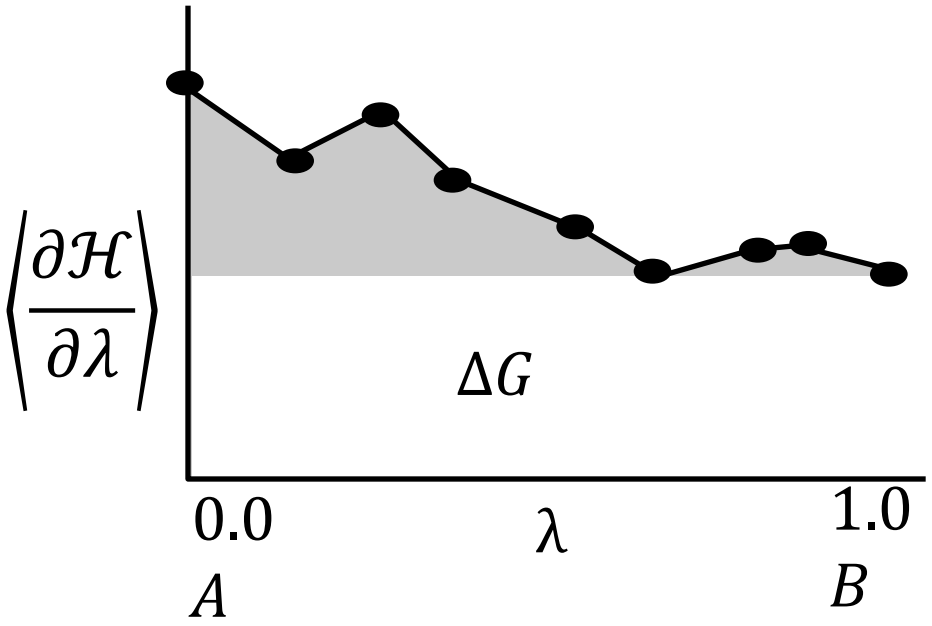
- ΔG_{AB}
- what is B ? what is A ?
  - more later
- measuring work ?



Work and free energy



Measure the work needed to move from A to B



where $\mathcal{H}$ is again Hamiltonian ($E_{kin} + E_{pot}$)

$$\Delta G = \int_A^B \left\langle \frac{\partial \mathcal{H}(\mathbf{p},\mathbf{r})}{\partial \lambda} \right\rangle_\lambda d\lambda \quad \text{or} \quad \Delta G = \sum_{i=0}^{N_{step}} (H_{i+1} - H_i)$$

Binding energy - feasibility

Would this approach work ?

$\left\langle {}^{\partial \mathcal{H}}/_{\partial \lambda} \right\rangle$ must be a good average (lots of fluctuations)

must change λ 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)

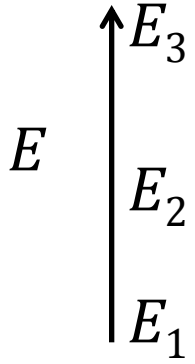
Problem – the path is too difficult – changes too big

- 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 ↔ AP $\quad\quad \Delta G_{\mathrm{A}}$

B + P ↔ BP ? $\quad\quad \Delta G_{\mathrm{B}}$

If I know ΔG_{B}

$\Delta\Delta G_{AB}$ is easier

$\Delta\Delta G_{AB} = \Delta G_A - \Delta G_B$

$$\mathrm{A + P} \xrightarrow{\Delta G_{\mathrm{A}}} \mathrm{AP}$$

What would $\Delta\Delta G_{AB}$ mean ?

- relative binding strength

$$\mathrm{B + P} \xrightarrow{\Delta G_{\mathrm{B}}} \mathrm{BP}$$

Alternative routes

$\Delta\ G_A$ and $\Delta\ G_B$ too hard

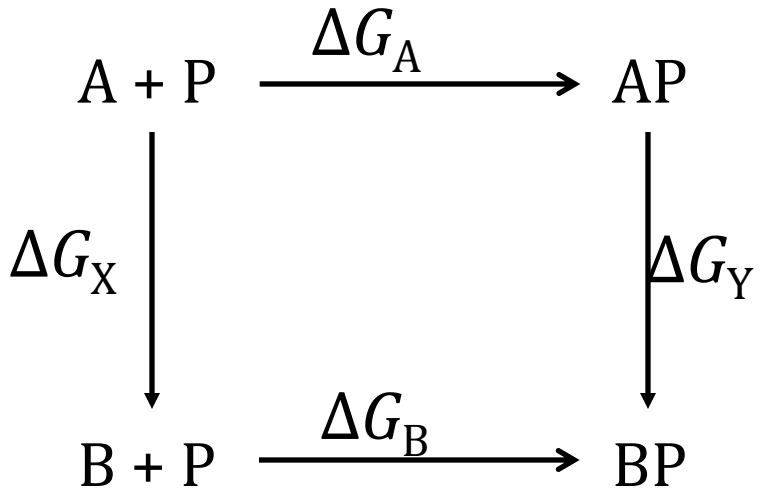
- we would be happy with $\Delta\ \Delta G_{AB}$

$$\Delta G_A + \Delta G_Y = \Delta G_B + \Delta G_X$$

$$\Delta G_A - \Delta G_B = \Delta G_X - \Delta G_Y$$ remember $\Delta\Delta\ G_{AB} = \Delta G_A - \Delta G_B$

So $\Delta\Delta G_{AB} = \Delta\Delta G_{XY}$

- why ΔG_X easier ?
- why ΔG_Y easier ?



Easier free energy changes

if A/B are rather similar

AP ↔ BP or

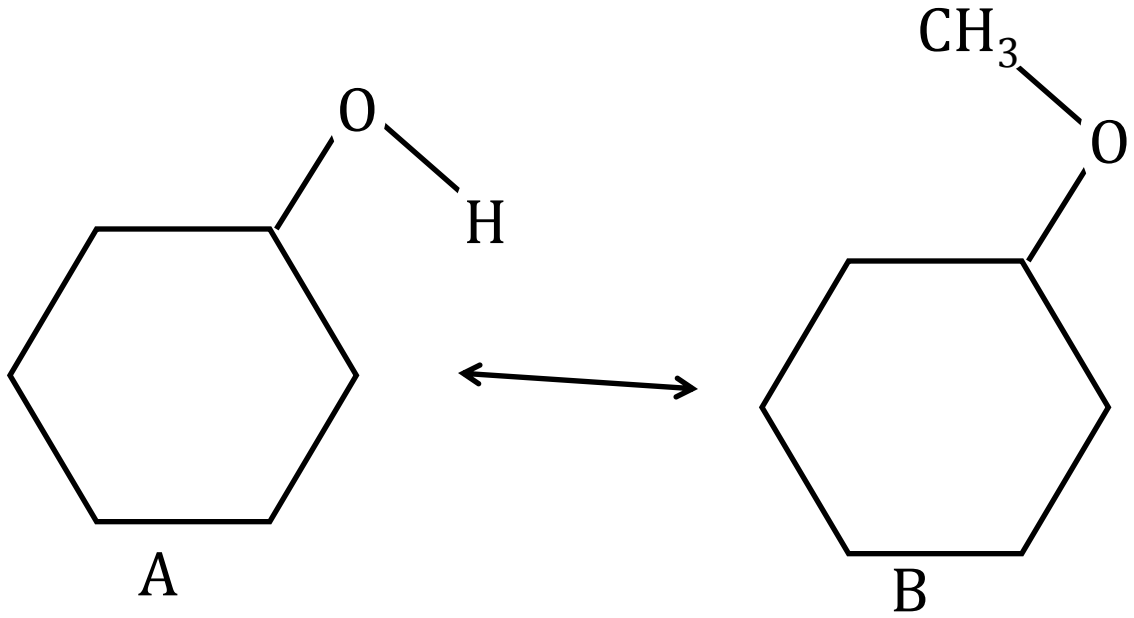
B + P ↔ A + P (free A ↔ B forget the protein)

are small changes – smaller than

- removing water order, removing water energy, finding protein…

Example

- small change

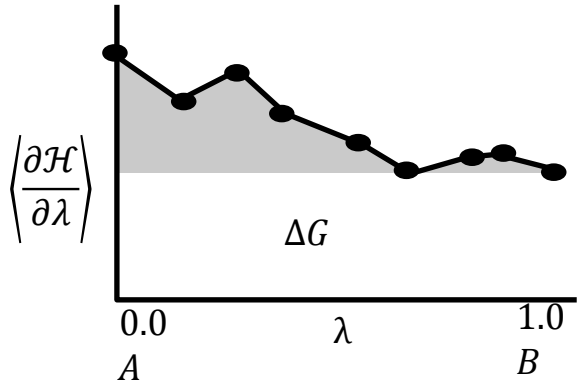


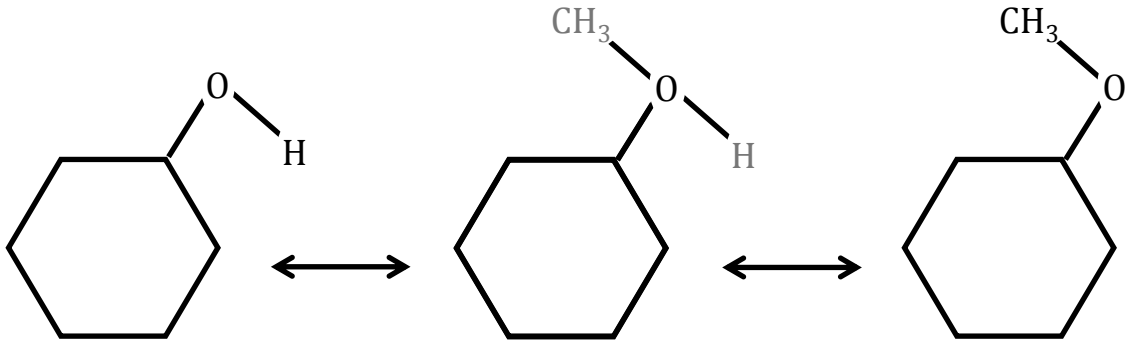
Fictitious states

Remember formulae

$$\Delta G = \int_A^B \left\langle \frac{\partial \mathcal{H}(\mathbf{p},\mathbf{r})}{\partial \lambda} \right\rangle_\lambda d\lambda \quad \text{and} \quad \Delta G = \sum_{i=0}^{N_{step}} (H_{i+1} - H_i)$$

make chemistry a function of λ





A
λ=0

A/B
λ=0.5

B
λ=1

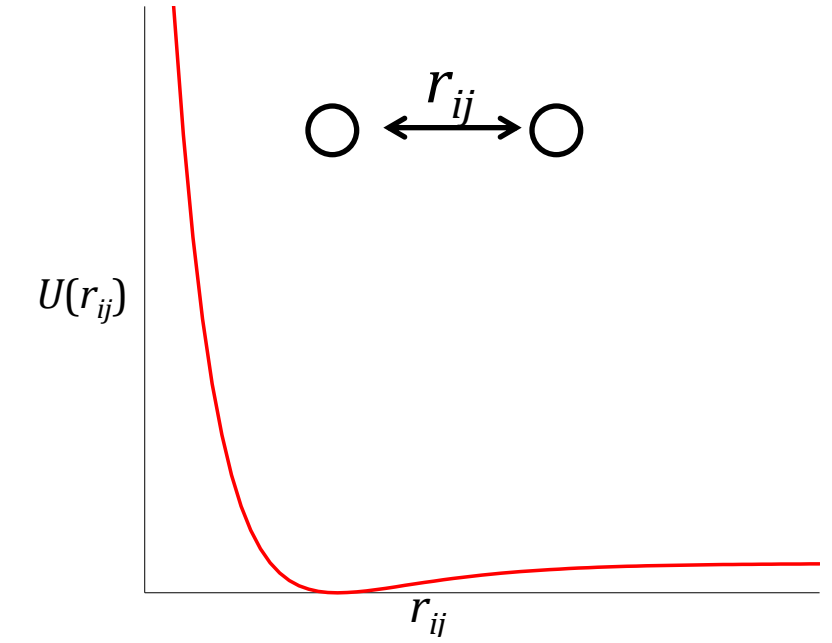
λ dependence

- $\lambda = 0$ an OH group
- $\lambda = 1$ an OCH_3 group
- $\lambda = 0.5$
  - charge of H – half of original charge
  - radius / size (σ, ε) half of real value and so on

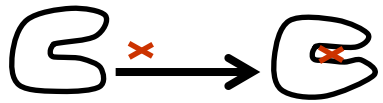
Atoms gradually

- appear in one direction
- disappear in other

Description of system is now function of λ

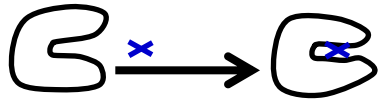


λ dependent simulations



Two simulations necessary

- λ from 0.0 ↔ 1.0 in protein
- λ from 0.0 ↔ 1.0 in water
- both from red ↔ blue



As λ 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

- ΔG 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

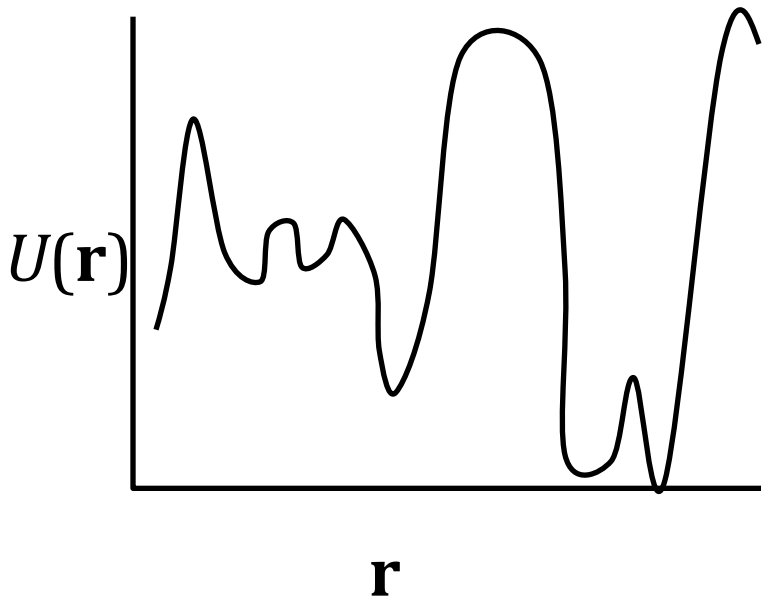
- MC or MD
  - with control of temperature (temperature bath)

Use : difficult optimisation problem

- chip layout
- travelling salesman problem
- protein structure

Optimisation problem

- several dimensional (2 to 2 000)
- many local minima



Kirkpatrick, Gelatt and Vecchi, Science, 220, 671-680, "Optimization by Simulated Annealing" (1983)

Procedure

while ($T > T_{end}$)

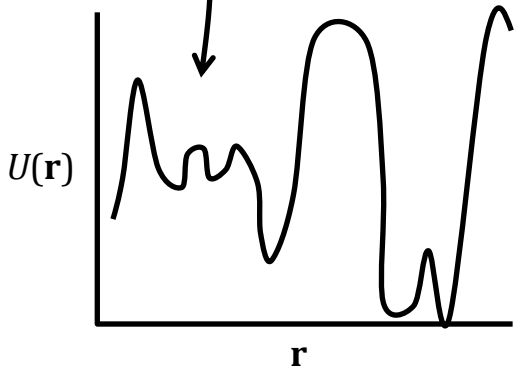
$T(t) = T_0 \, e^{-ct}$

move system (Monte Carlo)

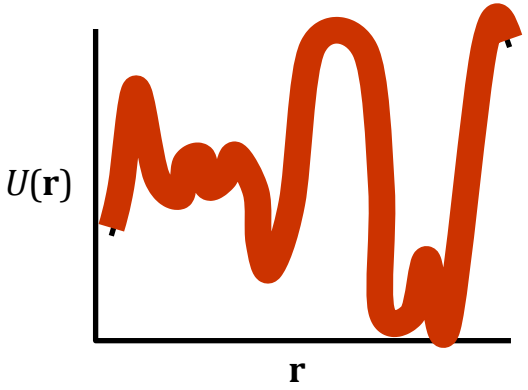
- T_0 initial temperature is hot
- c is decay rate (cooling of system)
- cost function is
  - E_{pot} in chemistry
  - path length in travelling salesman
  - board cost in chip layout problem …
- why may this work ?

Simulated Annealing concept

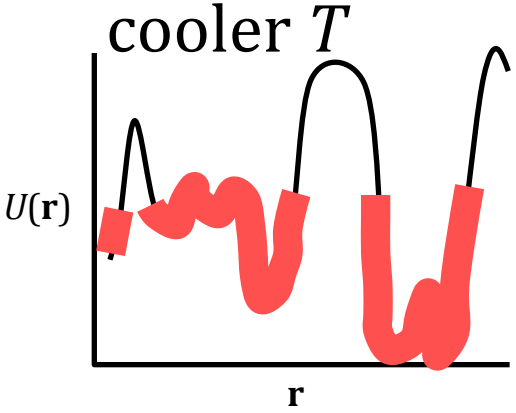
initial (poor) guess



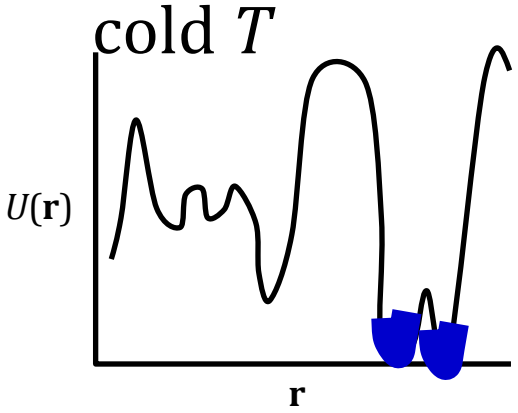
initial high T distribution



cooler T



cold T

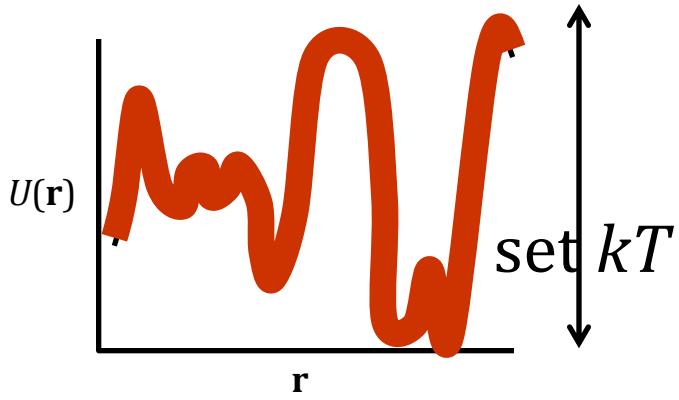


Properties, practical issues

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

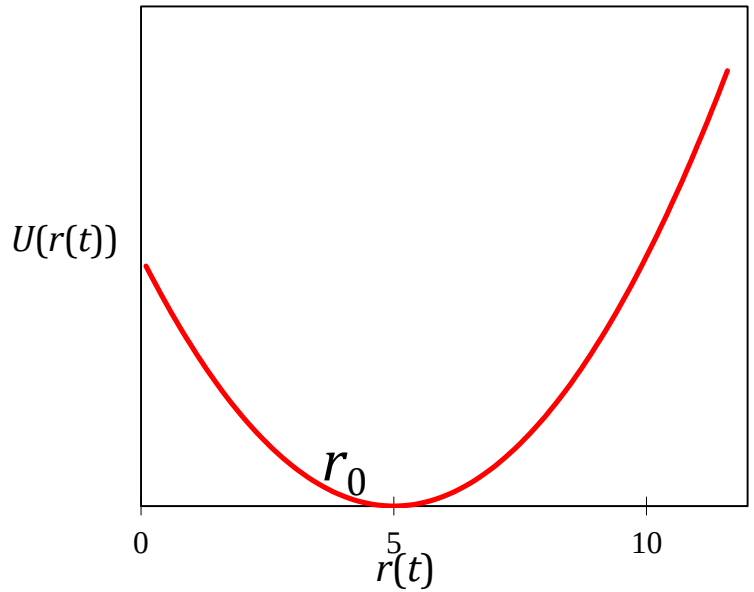
- coordinates are not bad, but could be improved

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…)

result ?

System moves to low energy + low fake energy

- gradually moves to agree with experimental data

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

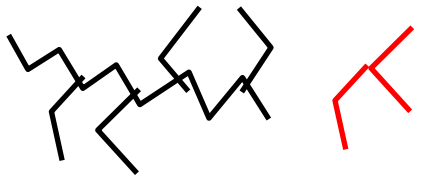
$U_i(r) = c_i\,(r(t) - r_0)^2$

- big c very artificial
- small c system will be slightly biased to agree with experimental data

Fake Energies - examples

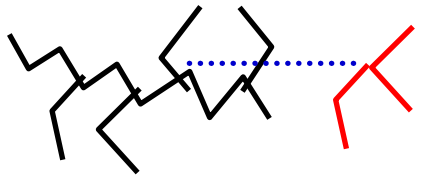
Refinement of

- 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, …)