

Applications – MD / MC

Basic tools

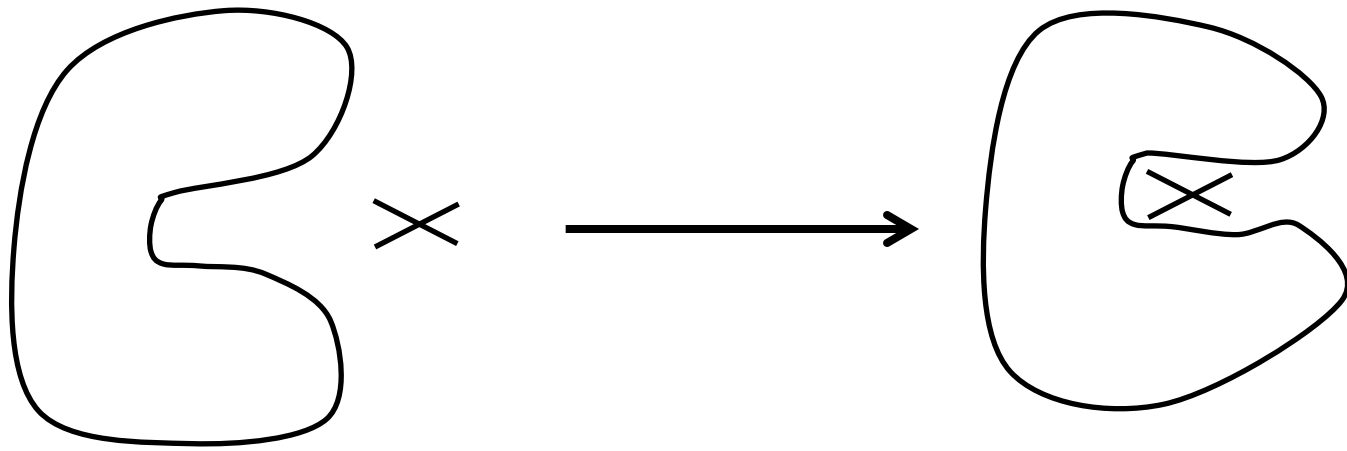
- Force field
- MD / MC

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 rad
 - time for decay in $A(t) = A(0) e^{-t/\tau}$
 - relaxation time
 - characteristic time
- times in proteins...

Some typical times in proteins

	Amplitude (Å)	$\log_{10} \tau(\text{s})$
bond vibration	0.01 – 0.1	–14 to –13
rotation of surface sidechain	5 – 10	–11 to –10
protein hinge bending	1 – 5	–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

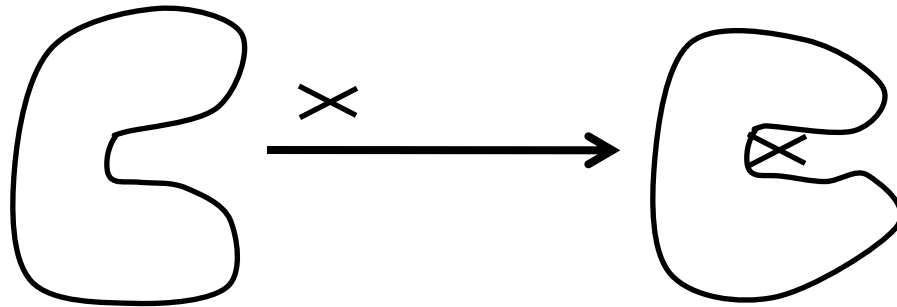
Timescales

- Typical big simulation $\approx 1\text{ns} = 10^{-9}\text{s}$
- Imagine event with characteristic time 10^{-9}s
 - may or may not be seen
- consider time 10^{-10}s
 - may be seen a few times
- What you would like
 - 100's or 1 000's of observations
- Limits of timescales
 - fast events $\tau \ll t_{\text{simulation}}$ OK
 - events $\tau < t_{\text{simulation}}$ poor statistics
 - $\tau \approx t_{\text{simulation}}$ no statistics
- Previous example (drug binding)
 - it is not enough to observe an event once (or few times)

Free Energy Calculations

- Free energy is most important
- Predicting therapeutic efficacy

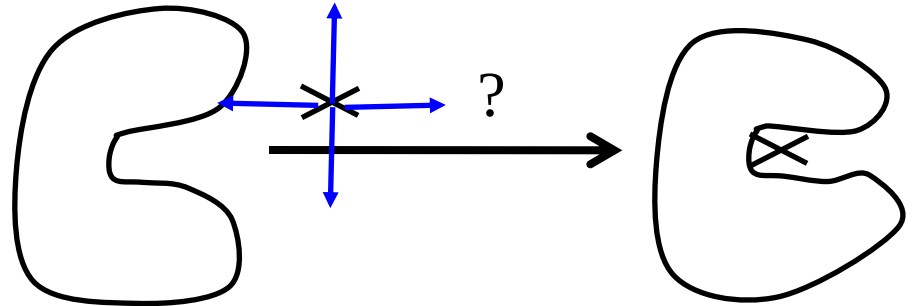
$$k_d = \frac{[\text{drug}][\text{protein}]}{[\text{drug-protein}]}$$
$$= e^{\frac{-\Delta G}{RT}}$$



- could we just look at energies ? What are contributing terms ?
 - ligand-water $\rightarrow$ ligand + water (many interactions, ΔS)
 - ligand+protein
 - ligand loss of entropy / water entropy change
- simulate ?

Free simulation for binding

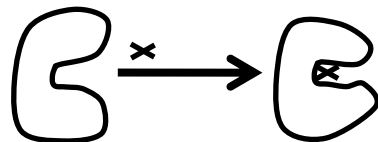
- if we simulate, where will the ligand go ?



- 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

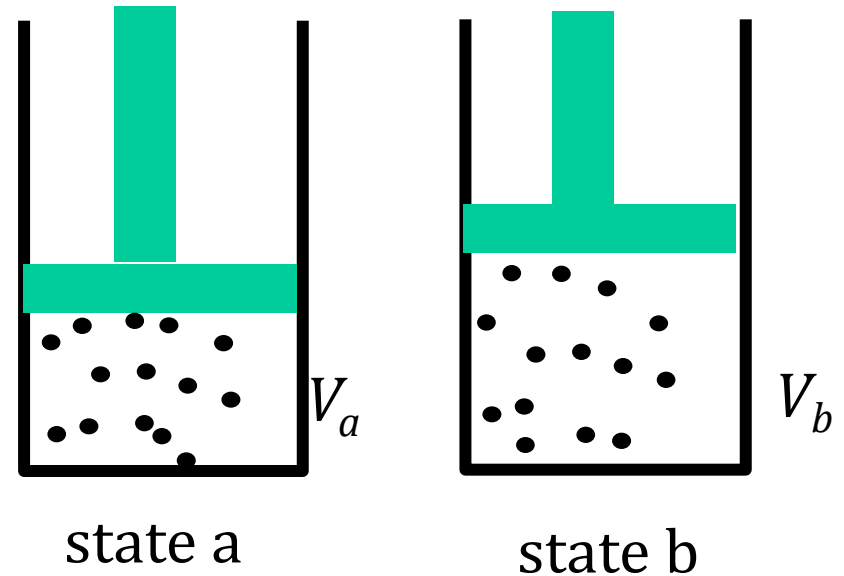
- $G = U - TS$
 - but $S = -k \sum_{i=1}^{N_{state}} p_i \ln p_i$
 - so we cannot really get S
 - some books write in terms of partition function
 - similar problem – especially visiting high energy regions
- forget absolute free energies
 - concentrate on ΔG
 - no problem – usually interesting property



Work and free energy changes

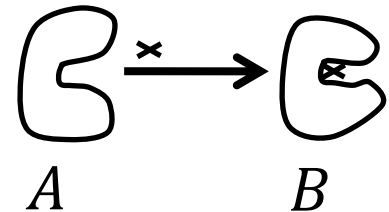
work done A to B

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

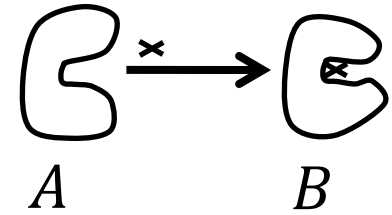


work going from unbound $\rightarrow$ bound

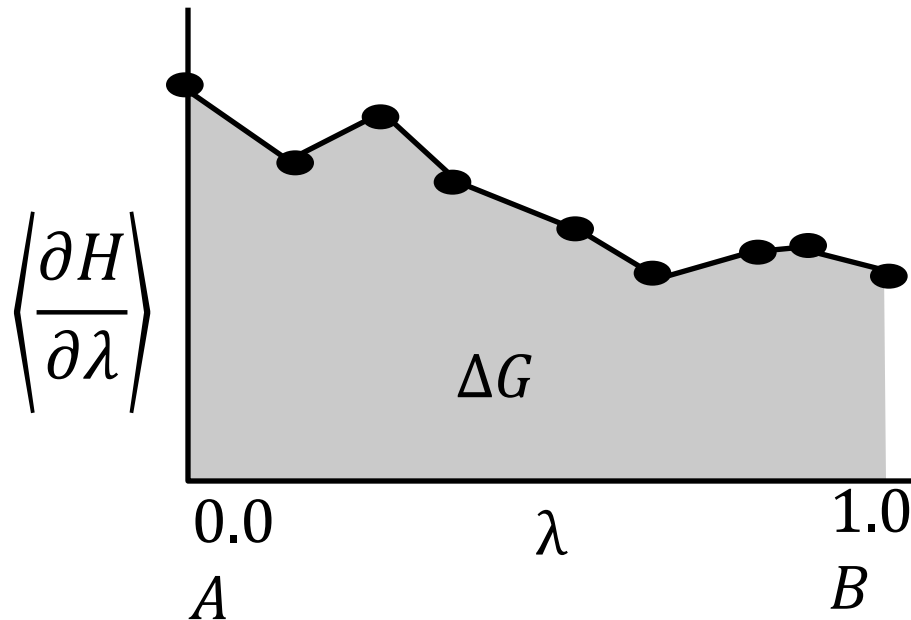
- ΔG_{AB}
- what is B ? what is A ?
 - more later
- measuring work ?



Work and free energy



- measure the work needed to move from A to B



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

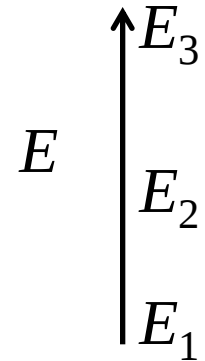
- $$\Delta G = \int_A^B \left\langle \frac{\partial H(\mathbf{p}, \mathbf{r})}{\partial \lambda} \right\rangle_\lambda d\lambda \quad \text{or} \quad \Delta G = \sum_{i=0}^{N_{step}} (H_{i+1} - H_i)$$

Binding energy - feasibility

- Would this approach work ?
 - $\langle \partial H / \partial \lambda \rangle$ 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)
 - $A + P \leftrightarrow AP$ ΔG_A
- what if I know $B + P \leftrightarrow BP$? ΔG_B
- $\Delta\Delta G_{AB}$ is easier
 - $\Delta\Delta G_{AB} = \Delta G_A - \Delta G_B$



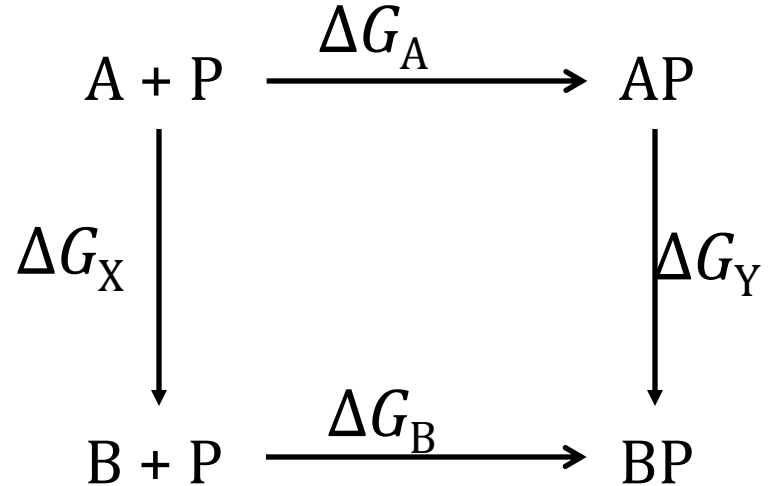
- what would $\Delta\Delta G_{AB}$ mean ?
 - what is relative binding strength ?



Alternative routes

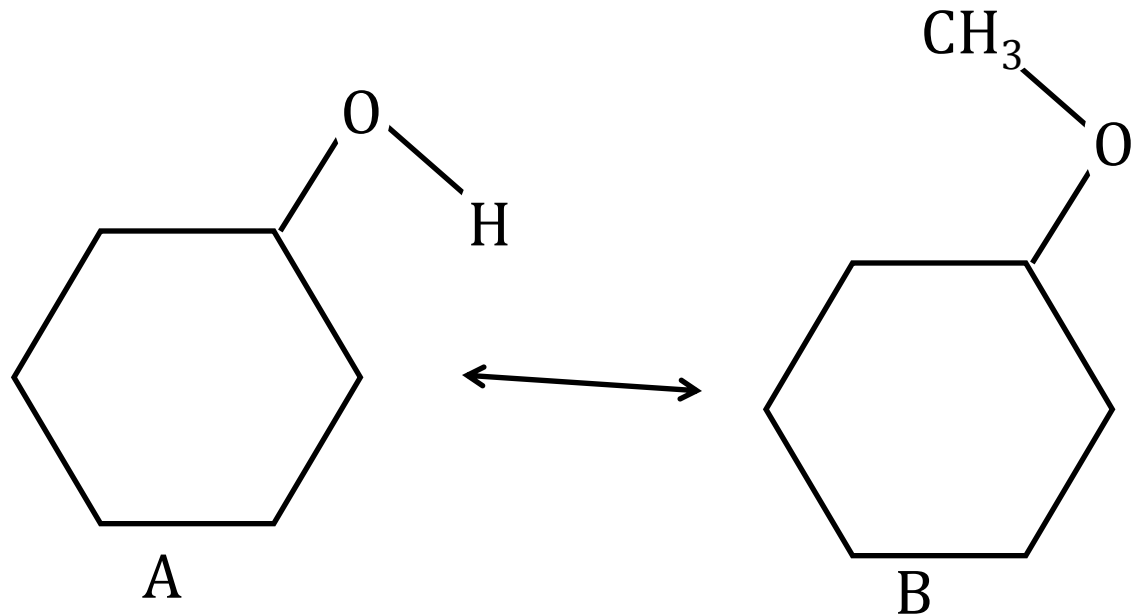
- ΔG_A and ΔG_B too hard
- we would be happy with $\Delta \Delta G_{AB}$
- $\Delta G_A + \Delta G_Y = \Delta G_B + \Delta G_X$
- $\Delta G_A - \Delta G_B = \Delta G_X - \Delta G_Y$ remember $\Delta \Delta G_{AB} = \Delta G_A - \Delta G_B$

- 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
 - $AP \leftrightarrow BP$ or
 - $B + P \leftrightarrow A + P$ (free $A \leftrightarrow B$)
- are small changes – smaller than
 - removing water order, removing water energy, finding protein...
- example
 - small change

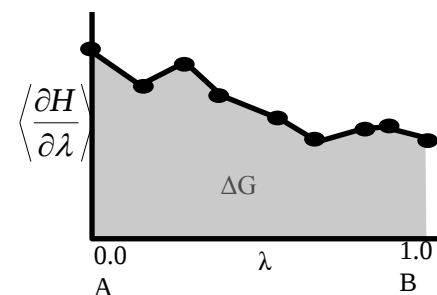
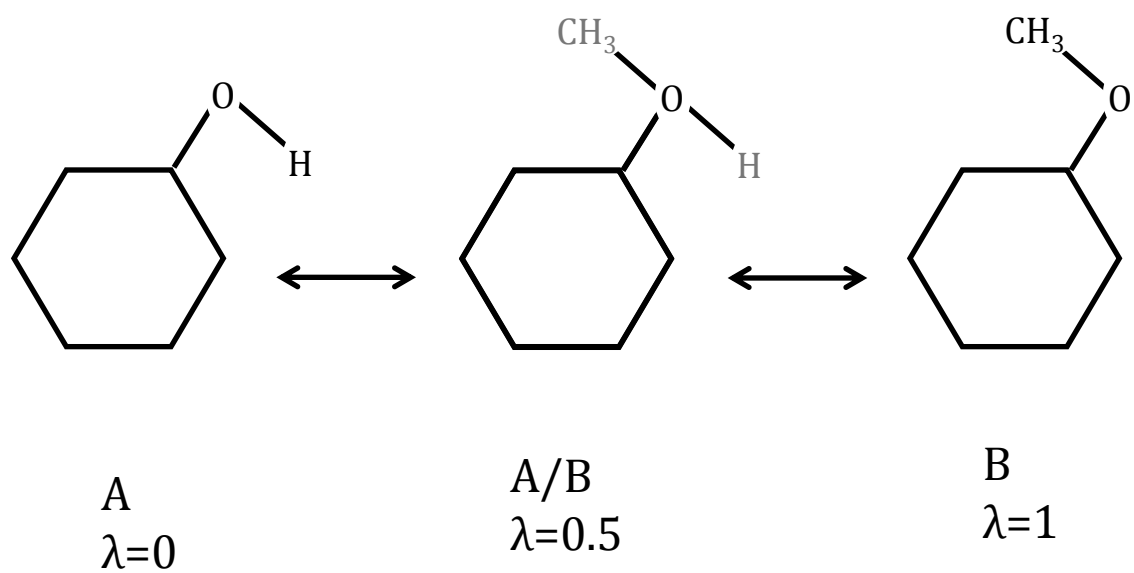


Fictitious states

- remember formulae
- we need to make chemistry a function of λ

$$\Delta G = \int_A^B \left\langle \frac{\partial H(\mathbf{p}, \mathbf{r})}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

$$\Delta G = \sum_{i=0}^{N_{step}} (H_{i+1} - H_i)$$

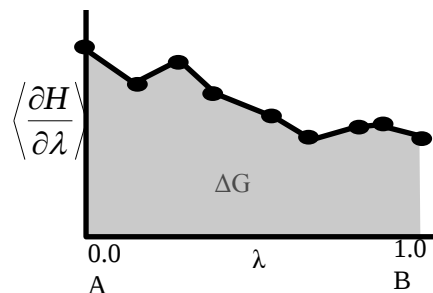
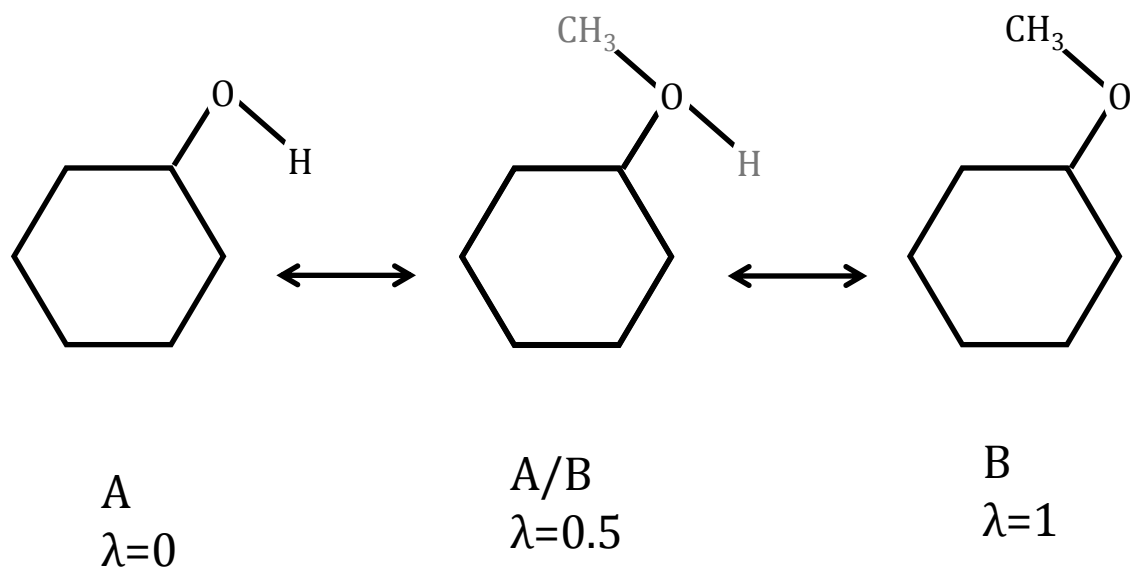


Fictitious states

- remember formulae

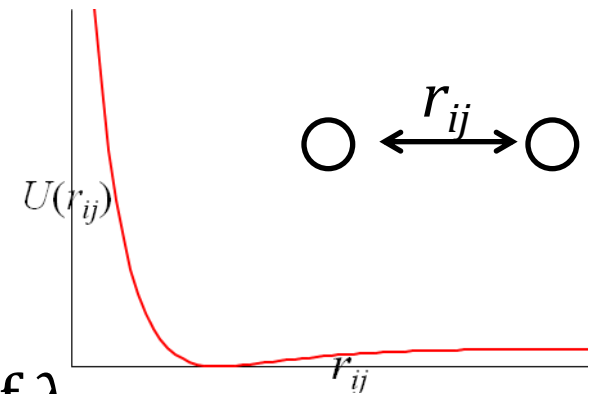
$$\Delta G = \int_A^B \left\langle \frac{\partial H(\mathbf{p}, \mathbf{r})}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad \text{and} \quad \Delta G = \sum_{i=0}^{N_{step}} (H_{i+1} - H_i)$$

- make chemistry a function of λ

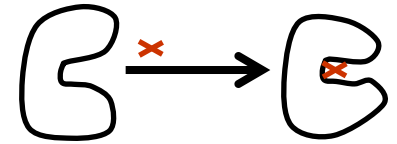


λ dependence

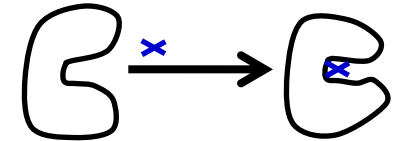
- $\lambda = 0$ an OH group
- $\lambda = 1$ an OCH₃ group
- $\lambda = 0.5$
 - charge of H – half of original charge
 - radius / size (σ , ϵ) half of real value and so on
- atoms gradually
 - appear in one direction
 - disappear in other
- description of system is now function of λ



λ dependent simulations



- two simulations necessary
 - λ from 0.0 $\leftrightarrow$ 1.0 in protein
 - λ from 0.0 $\leftrightarrow$ 1.0 in water
 - both from red $\leftrightarrow$ 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

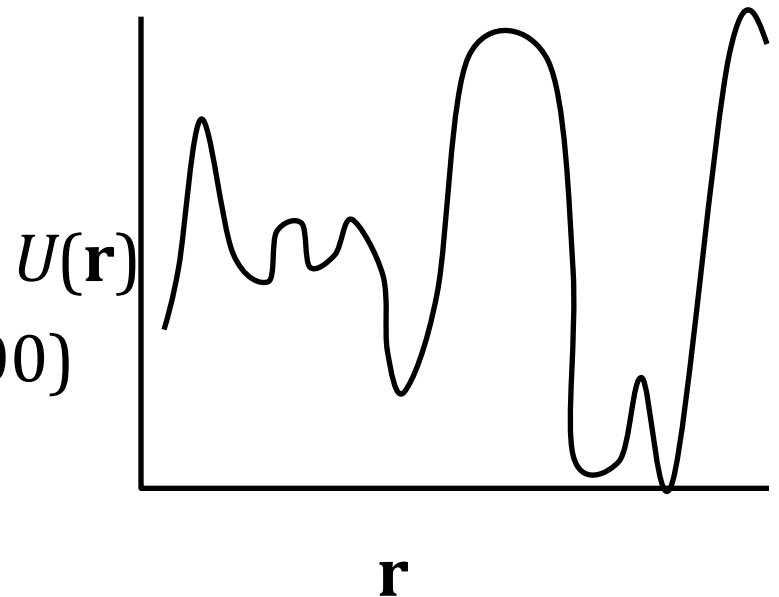


Summary of free energy calculations

- 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
- Use : difficult optimisation problem
 - chip layout
 - travelling salesman problem
 - protein structure
- Optimisation problem
 - several dimensional (2 to 2 000)
 - many local minima



Procedure

while ($T > T_{\text{end}}$)

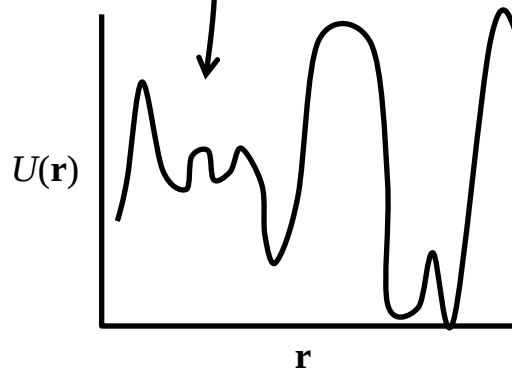
$$T(t) = T_0 e^{-ct}$$

move system (Monte Carlo)

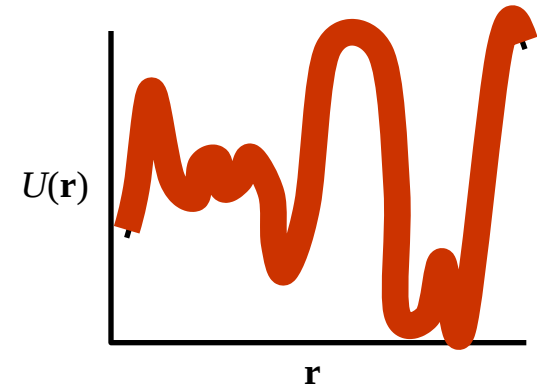
- T_0 initial temperature is hot
- c is decay rate (rate of decrease)
- cost function is
 - E_{pot} in chemistry
 - path length in travelling salesman
 - board cost in chip layout problem ...
- why may this work ?

Simulated Annealing concept

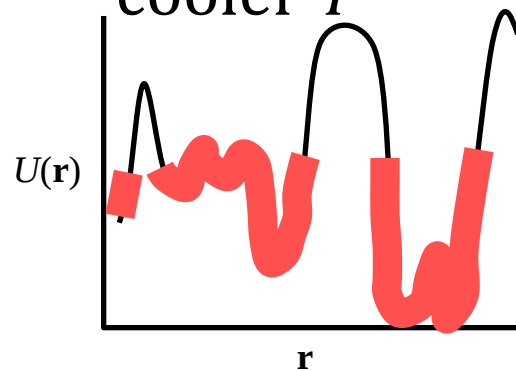
initial (poor)
guess



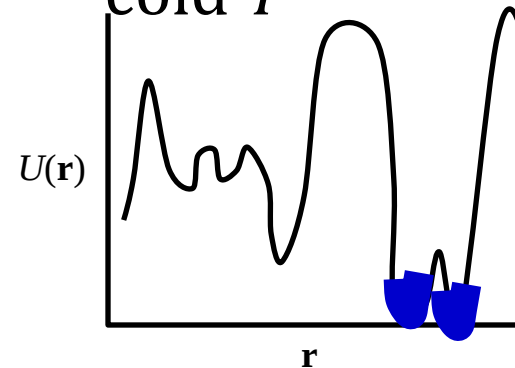
initial high T
distribution



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 ?

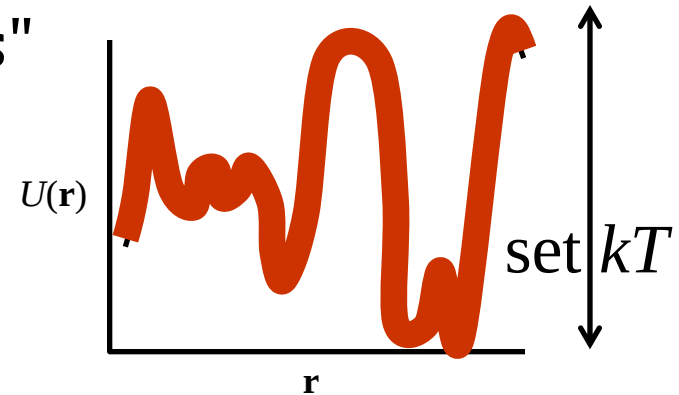
- infinite ? No : look at barriers

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, ...

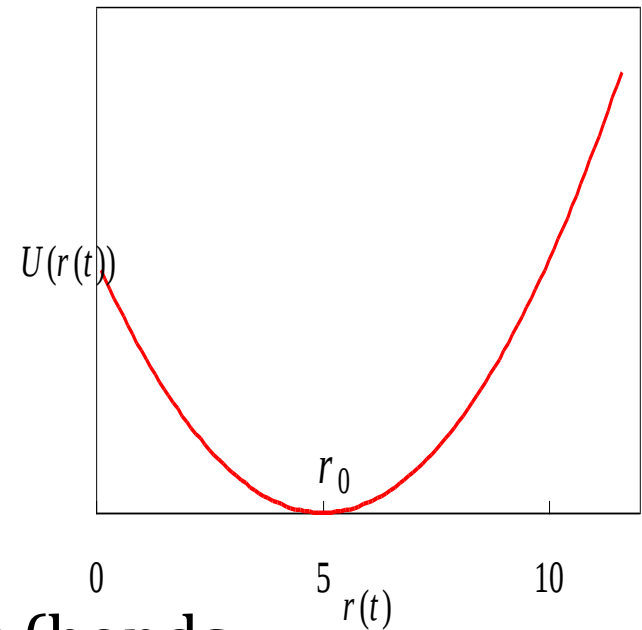
Refinement of Structures (NMR / X-ray)

- 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...)

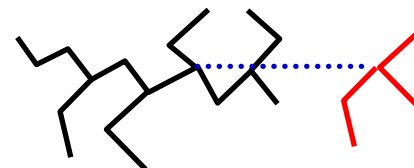
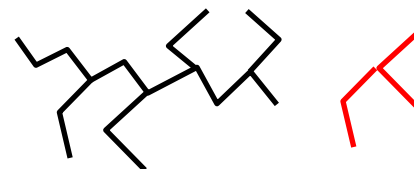
result ?

- 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

Fake energies for many purposes

- 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, ...)