# Self-consistent mean field methods

General method for problems with

- multiple sites
- each site exists in different states
- each site interacts with other sites

History

- Ising spin model
- application different to this one

Aims

- find optimal set of states or
- find distribution of states at a given temperature

# examples

Relevant to us

- protein side chains
- RNA base pairing
- sequence design

Historic / simple

- spin systems

Not here

- wave functions (standard method)
- polymer properties

# Plan

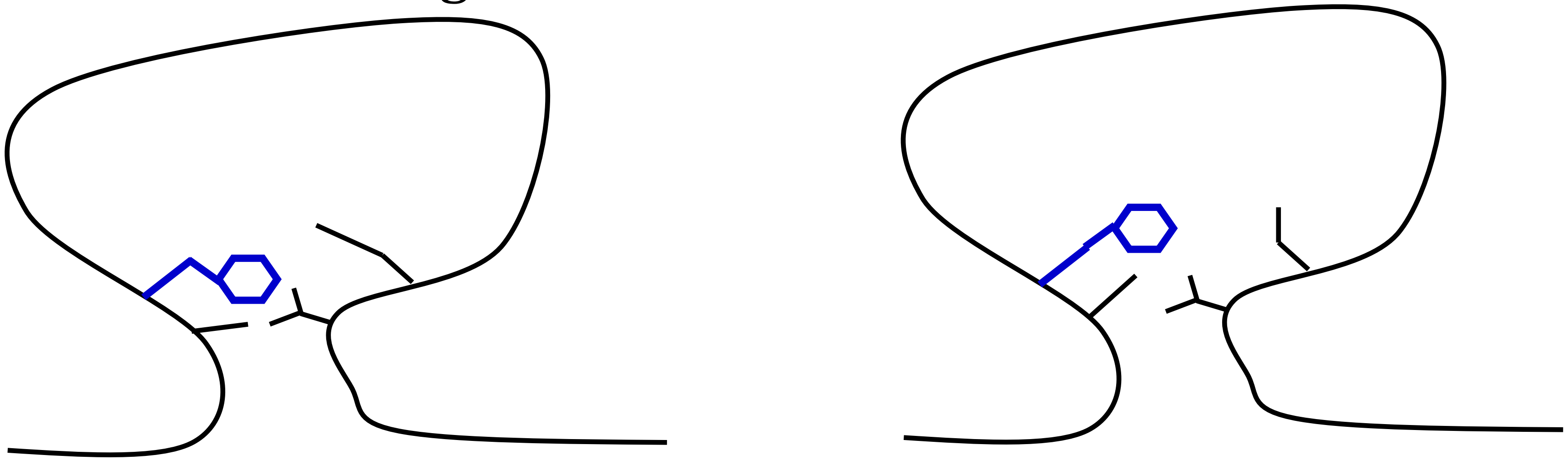
- some example problems
- Boltzmann relation
- examples in detail

Examples

- common feature
- parts of a system exist in some number of states
- parts of a system interact with each other

# Protein side chains

- optimise (energy) their coordinates
- each interacts with his neighbours

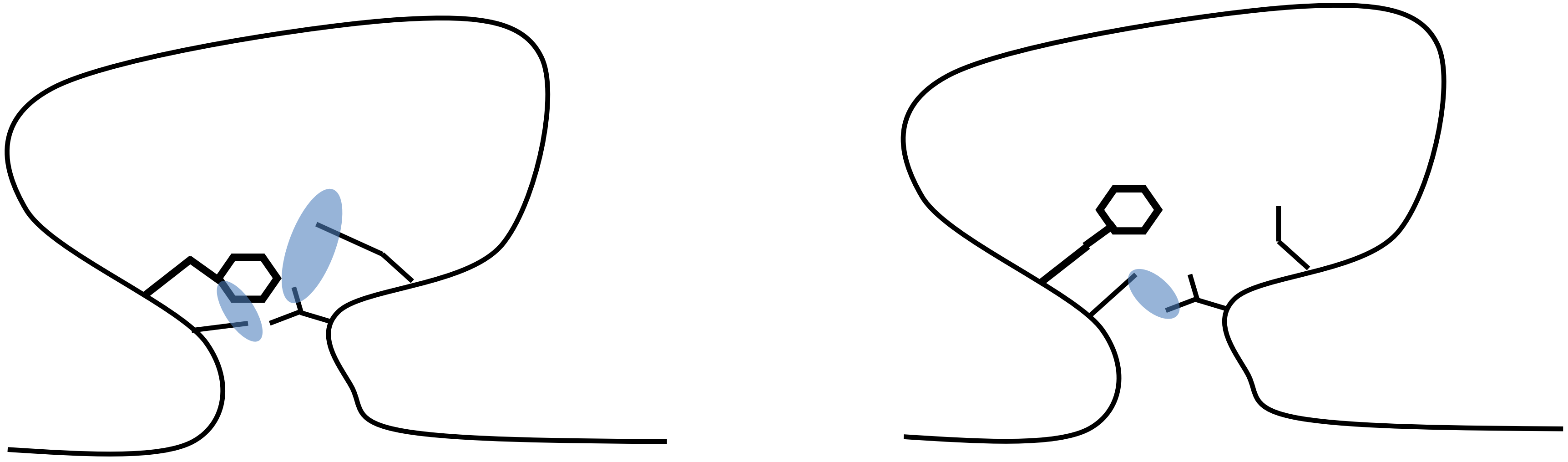


Simplification

- each sidechain can exist in one of $m$ positions
  - say $m = 3$

# Protein side chains

How many interactions ?



Make one interaction and break another

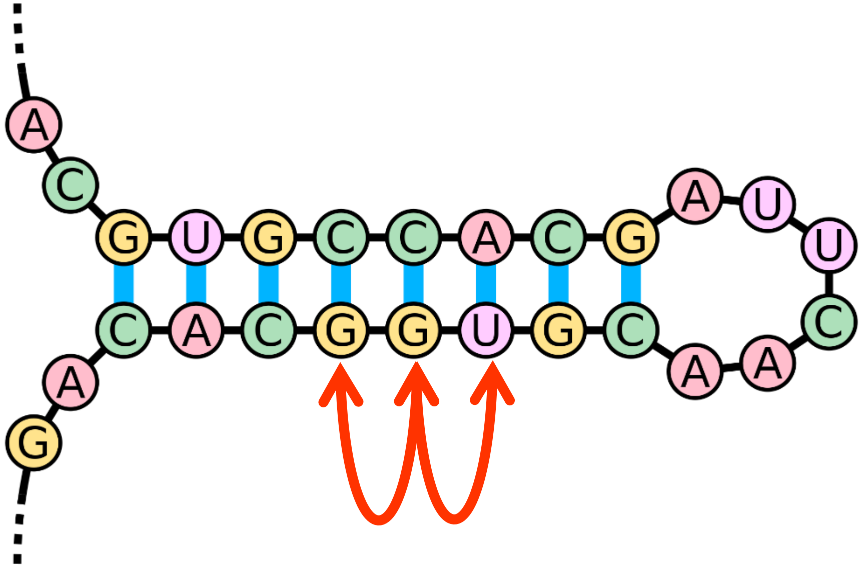
- what is the best combination ?

How big is the search space ?

- $n$ sidechains each has $m$ configurations $= m^n$
- for m= 3 we have $3^n$ = very many

# Sequence Design

RNA, but could be proteins, DNA



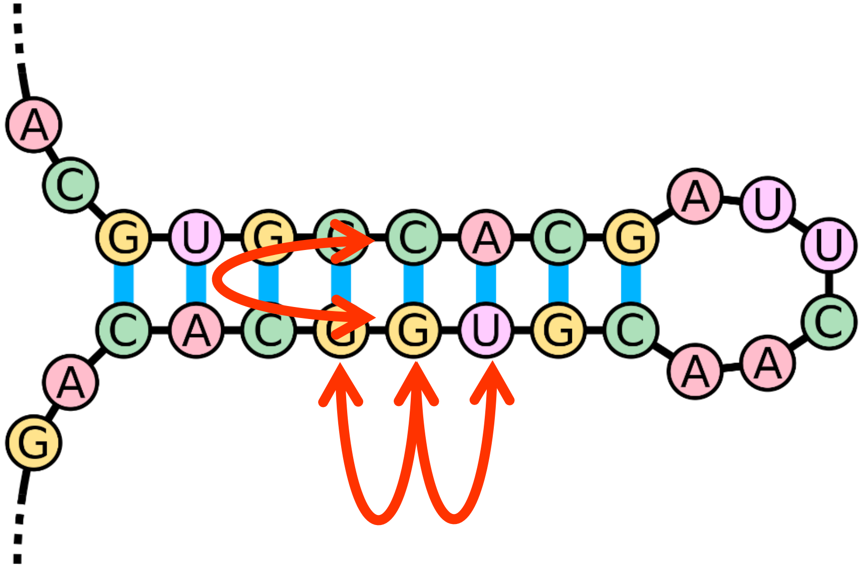
How are energies calculated ?

1. base pairs – across chain
2. sequence neighbours – base stacking

# Sequence Design

Best energy

- change one base
  - affects neighbours
    - across
    - along chain



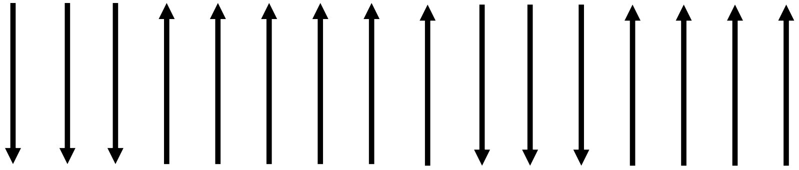
- $m = 4$ base types
- $n =$ length positions
- $m^n$ possibilities (search space)

# magnetism / spin models

Not bioinformatics ? Classic / historic

Energy (no external field)

$$V = -c \sum_{i=1}^{n-1} \sigma_i \sigma_{i+1}$$



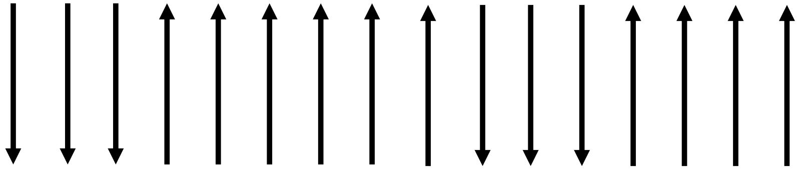
- $2^n$ possible arrangements
- flip one spin to fit to left neighbour
  - might break interaction to right neighbour

$c$ some constant
$\sigma_i$ vector – which way is spin $i$ pointing ?

# magnetism / spin models

Toy example ? You know the optimal answer(s)

Systems with more states / more complicated interactions



Do not always want the optimum

- distribution as a function of temperature

# More examples

Electronic configuration of a small system ($n$ electrons)

- shells $s$, $p$, $d$, …
- electrons have spin (↑↓)
- each electron interacts with every other electron
- put an electron in a certain $p$ orbital
  - changes probability of neighbours
  - changing their probabilities changes

# Common properties

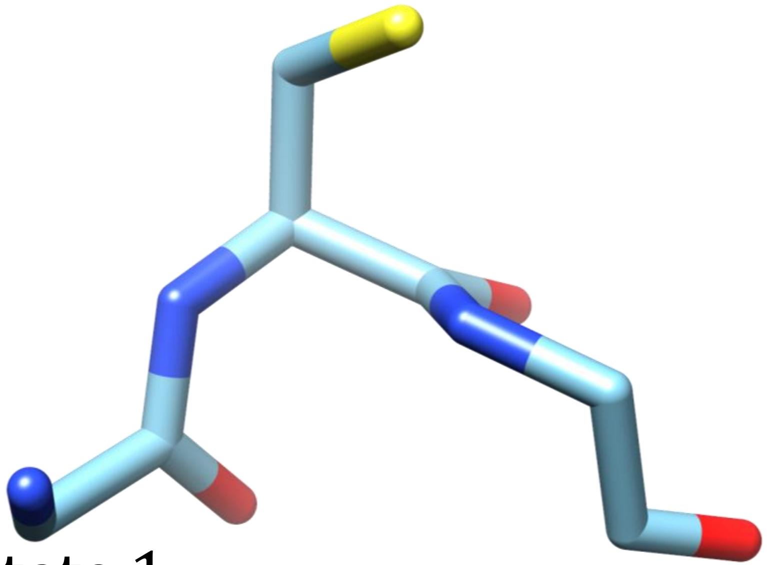
- $n$ sites
- $m$ states
  - $m^n$ search space
- changing state at $i$ affects site $j$ which affects site $k$ …
- sites are not independent
  - you cannot optimise $i$, then $j$, then $k$, …

General approach

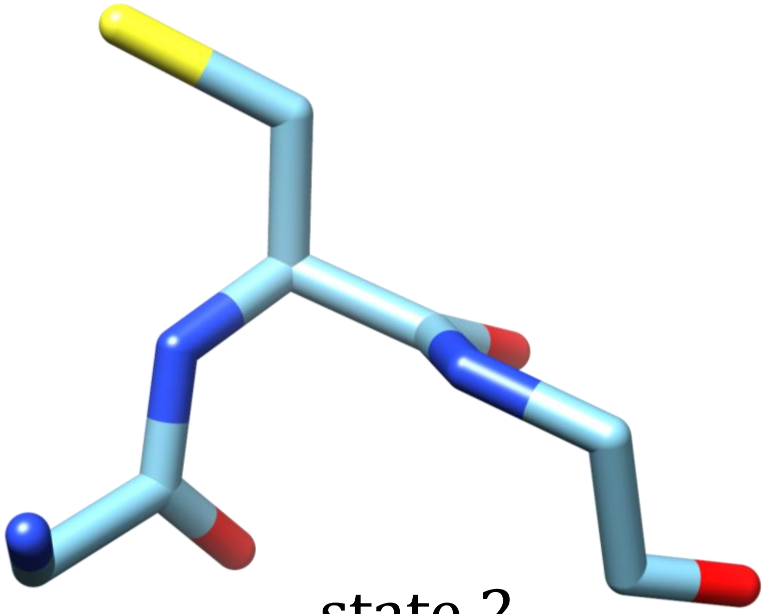
- mean field methods / self-consistent mean field methods

# Boltzmann .. the detour

Site with two states



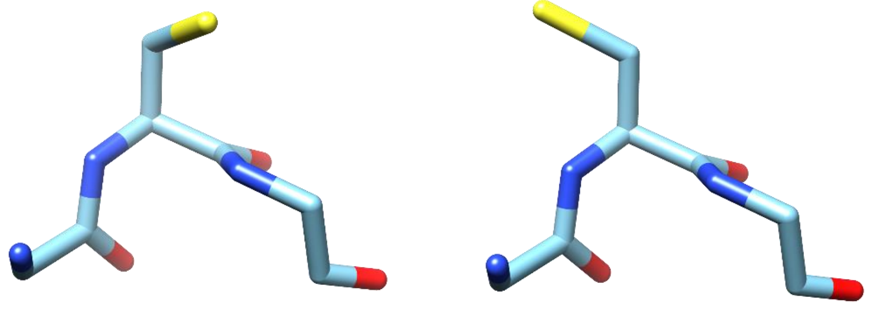
state 1
$E_1$



state 2
$E_2$

Energy difference $\Delta E = E_1 - E_2$
what is the ratio of populations ?

# Boltzmann



$$\frac{p_i}{p_j} = e^{-\Delta E/_{kT}}$$

why should you believe me ?

$$\ln \frac{p_i}{p_j} = -\frac{\Delta E}{kT}$$

$\Delta E = -kT \ln \frac{p_i}{p_j}$ which looks like $\Delta G = -RT \ln \frac{[\mathrm{A}]}{[\mathrm{B}]}$

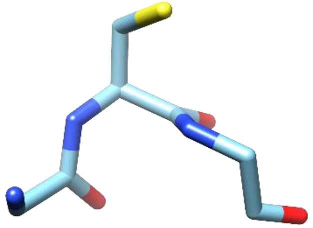
in the reaction $\mathrm{A} \rightleftharpoons \mathrm{B}$

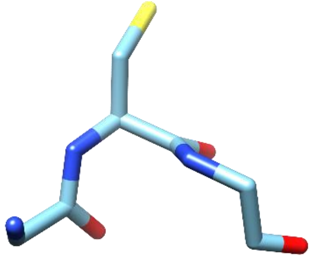
# Boltzmann – more states

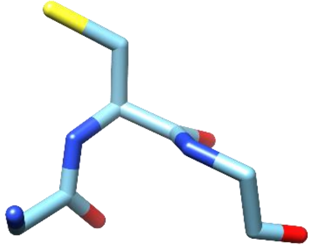
$\frac{p_i}{p_j} = e^{-\Delta E/_{kT}}$ but we could also say $\frac{p_i}{p_j} = \frac{\exp\left(\frac{-E_i}{kT}\right)}{\exp\left(\frac{-E_j}{kT}\right)}$

$e^{\frac{-E_i}{kT}}$ is the Boltzmann weight of $i$

What if I have three states ?







$$p_1 = \frac{\exp\left(\frac{-E_1}{kT}\right)}{\exp\left(\frac{-E_1}{kT}\right)+\exp\left(\frac{-E_2}{kT}\right)+\exp\left(\frac{-E_3}{kT}\right)}$$

what about $m$ states ?

# Boltzmann – $n$ states

$$p_1 = \frac{\exp\left(\frac{-E_1}{kT}\right)}{\exp\left(\frac{-E_1}{kT}\right)+\exp\left(\frac{-E_2}{kT}\right)+\exp\left(\frac{-E_3}{kT}\right)}$$

generalises to

$$p_i = \frac{\exp\left(\frac{-E_i}{kT}\right)}{\sum_j^m \exp\frac{-E_j}{kT}}$$

will be used over and over again

# Distributions

Simple system with two states $\frac{p_i}{p_j} = e^{-\Delta E/_{kT}}$

At $T = 0$, $\frac{-\Delta E}{kT}$ becomes huge, negative

all the probability goes to lowest energy state

At $T \gg 0$, $\frac{-\Delta E}{kT}$ goes towards 0, $e^0 = 1$

at high temperature, $p_i \approx p_j$

For in-between …

# Optima and Distributions

$T = 0$ or $T = 300$ K or $T = 10^{10}$ ?

For simulations of the real world

$T = 300$ K

To find the optimum

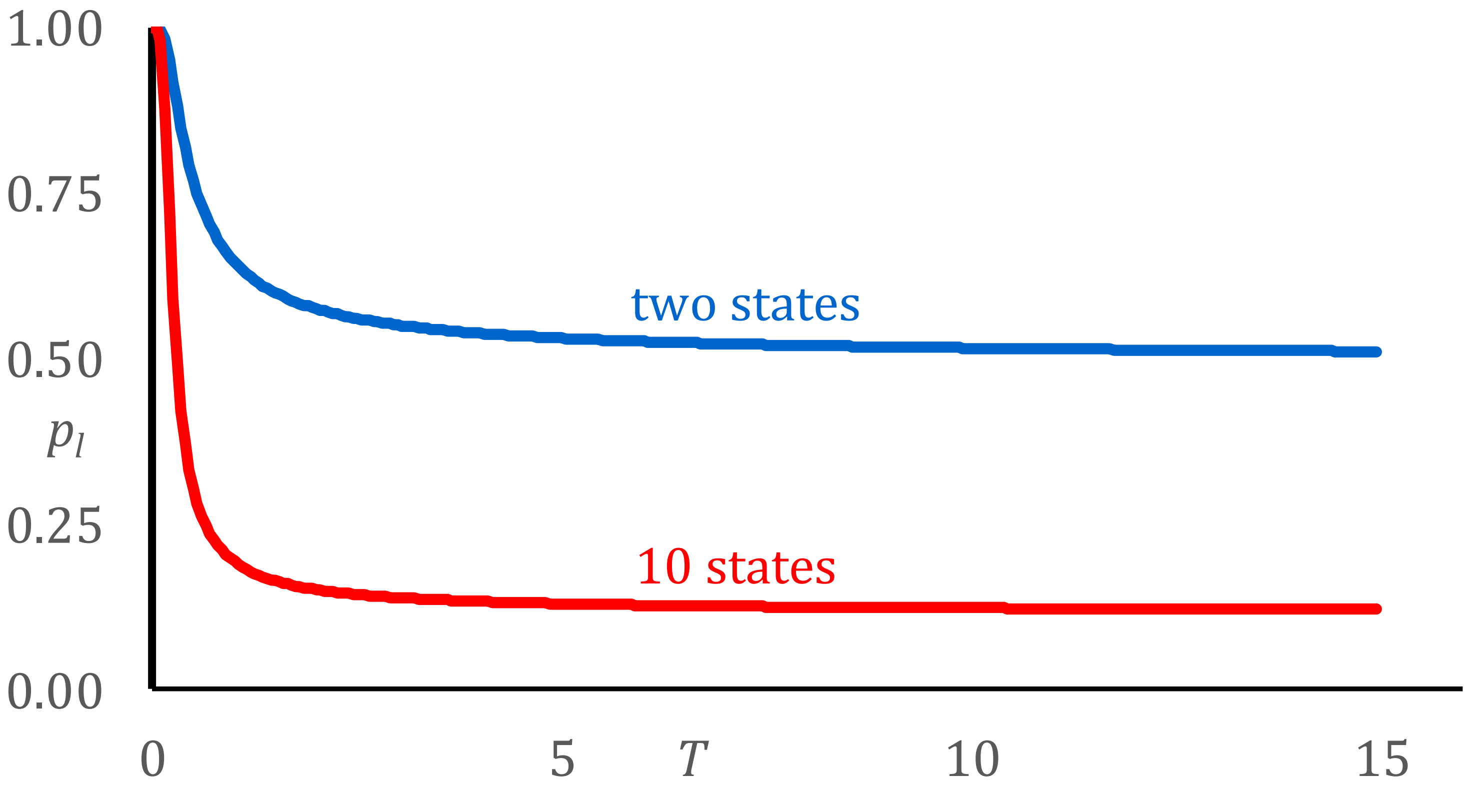
$T = 0$

$T$ is

- real temperature or
- a convergence parameter
  - as the system cools, it is pushed to lower energy states

# Probability as function of temperature

Probability of lowest energy state depend on $T$



$p_l$ probability of lowest energy state, $T$ temperature

# Real world or optimisation ?

Simulations ?

- distributions of states

Answers

- rotamer distributions
- base-pairing
- sequence design
  - just the optimum

For these lectures

- best solution at $T = 0$

# Philosophy

Start system at high temperature

- all states are equally likely
- each part of the system feels the average of its neighbours

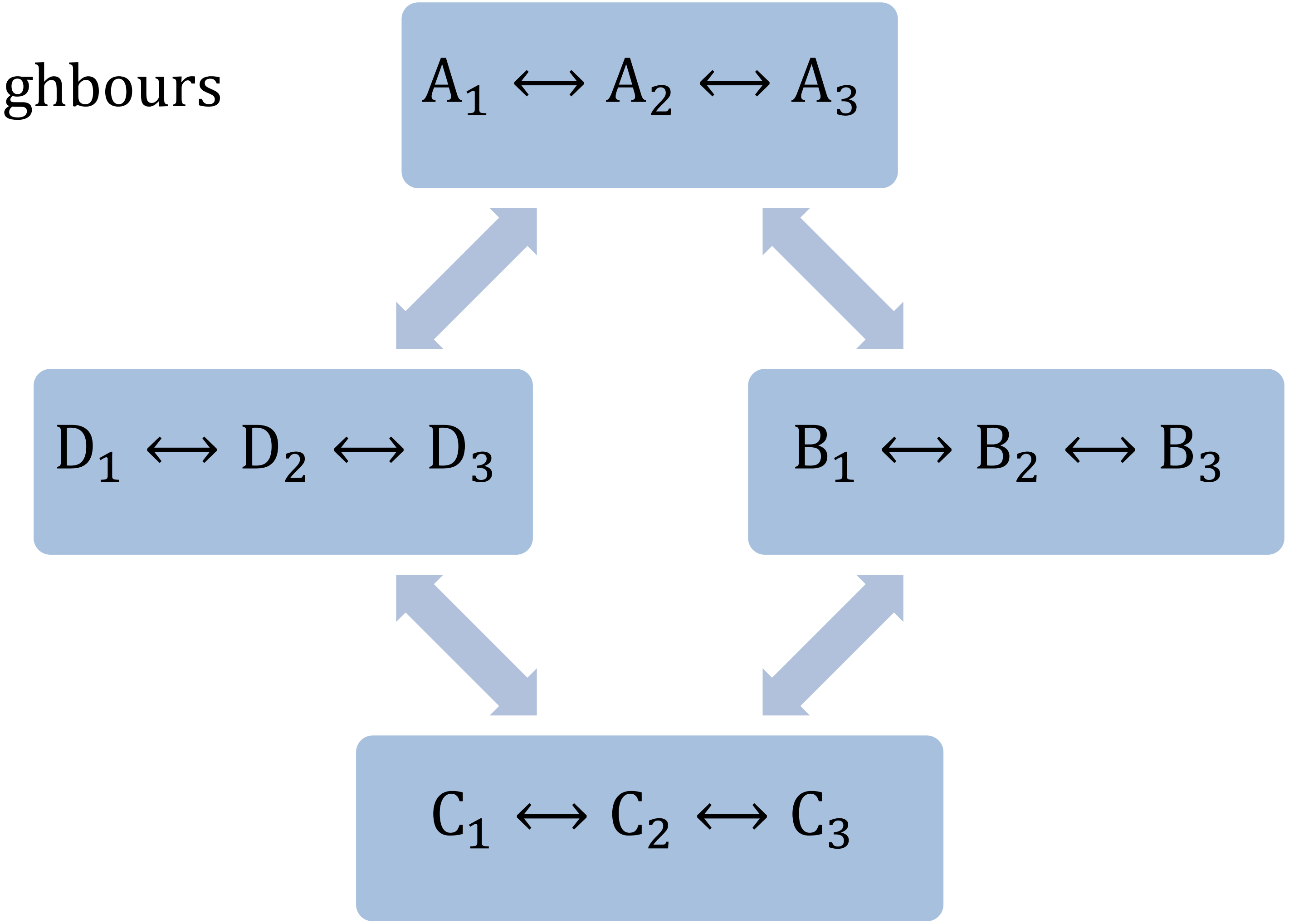
Gradually cool

  - each site moves to lowest energy states

Can we just look at lowest energy state in one step ?

- no

Each site affects his neighbours



$A_1 \leftrightarrow A_2 \leftrightarrow A_3$

$D_1 \leftrightarrow D_2 \leftrightarrow D_3$

$B_1 \leftrightarrow B_2 \leftrightarrow B_3$

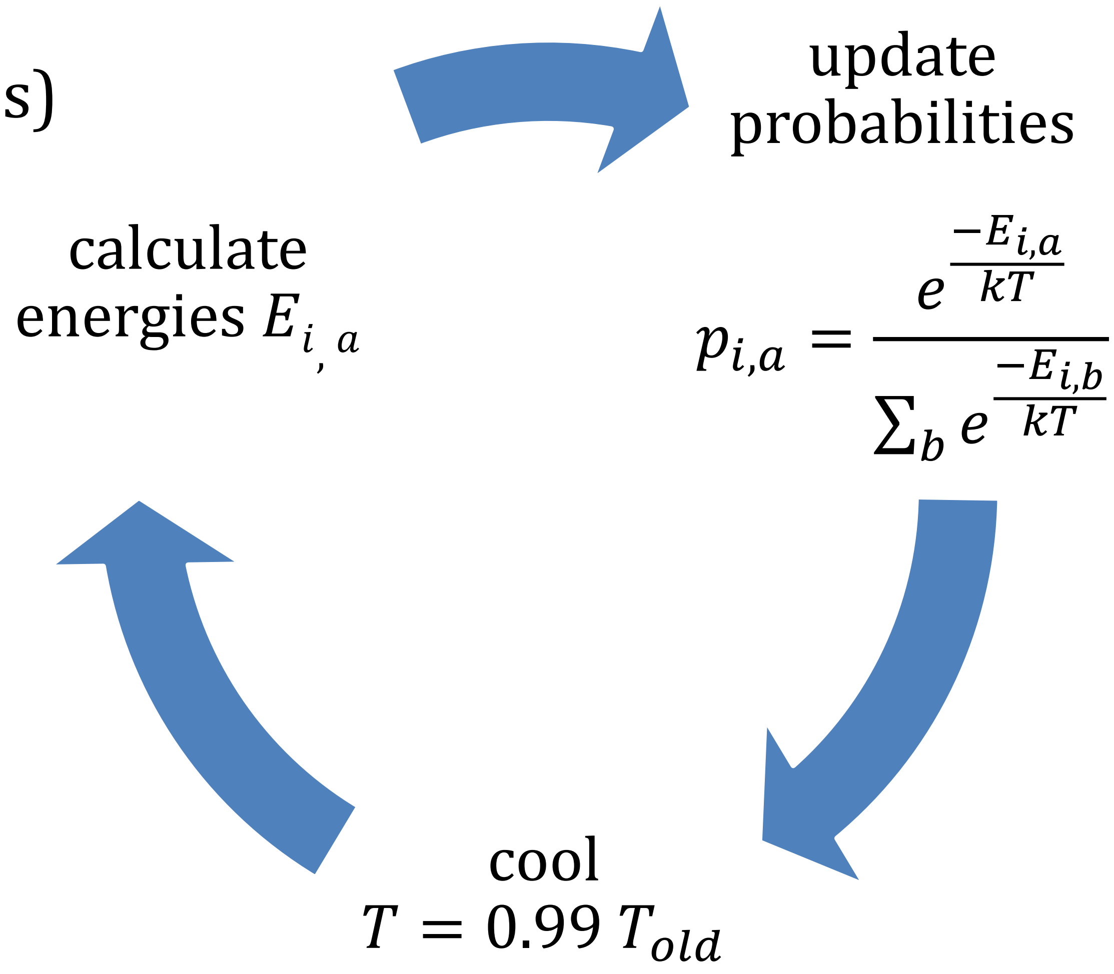
$C_1 \leftrightarrow C_2 \leftrightarrow C_3$

Cannot know the optimum of A since we do not yet know B, C or D

The state of A affects B affects …

We have much bigger networks (many sites)

- adjust one a little bit, cool a bit…



update probabilities

$$p_{i,a} = \frac{e^{\frac{-E_{i,a}}{kT}}}{\sum_b e^{\frac{-E_{i,b}}{kT}}}$$

cool

$T = 0.99\ T_{old}$

calculate energies $E_{i,a}$

# Examples – Sidechain conformations

Assume

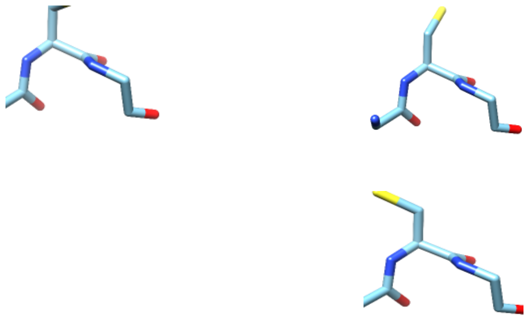
- some model for energy
- discretisation – sidechain rotamers
  - residue $i$ can exist in $m$ conformations

Energy depends on

- neighbours
- interactions with backbone

- Example .. side chain with 3 positions

# Work through a calculation



Use $a, b, c$ for states… Use $i, j, ..$ for sites

Consider one side chain at site $i$

- 3 states (for example)
- we want probability $p_{i,a}$ in each state $a$

What are the interactions of sidechain $i$ ? Consider neighbour $j$

- $j$ has a probability $p_{j,a}$ of being in state $a$ (for all the different $a$)
- use the mean field

# mean field

Say $E(i, j)$ is the energy of sites $i$ and $j$ interacting, but be more specific

$E(i_a, j_b)$ is the energy due to $i$ in state $a$ with $j$ in state $b$

We do not know the state of $j$, but we do know the probabilities

$$E(i_a, j) = \sum_{b}^{m_{states}} \left(p_{j,b}\, E(i_a, j_b)\right)$$

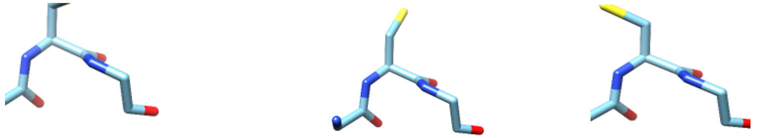
this is for one neighbour, but we want the total energy $E_{i,a}$

$$E_{i,a} = \sum_{j}^{n_{neighbour}} \left( \sum_{b}^{m_{states}} \left(p_{j,b}\, E(i_a, j_b)\right) \right)$$

Summation over all states of neighbours – mean field

Now have $E_{i,a}$

- repeat for each state $a$ use the Boltzmann rule to get the probabilities
- from $p_{i,a} = \dfrac{\exp\frac{-E_a}{kT}}{\sum_{b=1}^{N_{states}} \exp\left({}^{-E_b}/_{kT}\right)}$



In words…

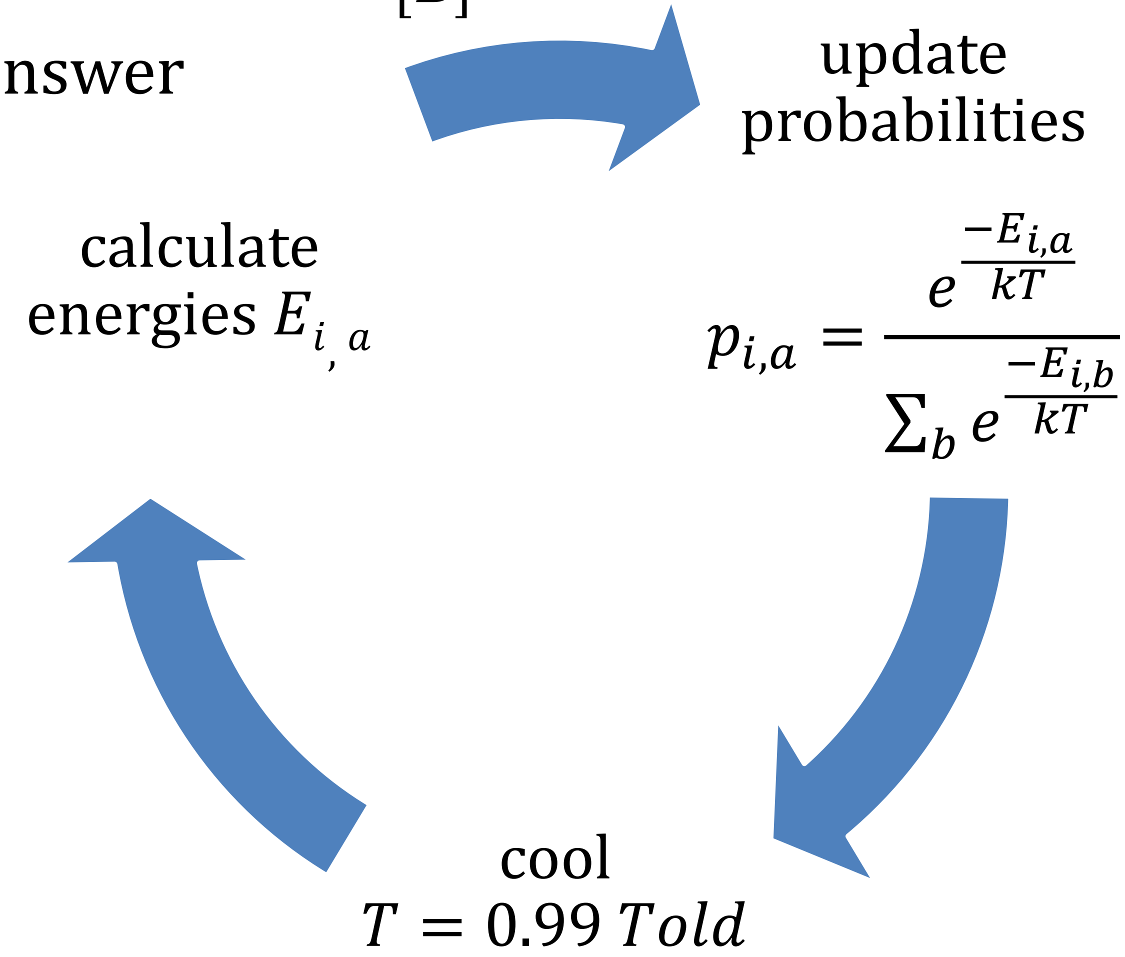
```
for each site $i$
    for each state $a$
        for each neighbour $j$
            for each state $b$
                add in $p_{j,b}\ E(i_a, j_b)$
                $E_{i,a} = E_{i,a} + p_{j,b}E(i_a, j_b)$
```

# Why cool ?

Remember $\Delta G = -RT \ln \frac{[\mathrm{A}]}{[\mathrm{B}]}$ in the reaction $\mathrm{A} \rightleftharpoons \mathrm{B}$ so $\frac{[\mathrm{A}]}{[\mathrm{B}]} = e^{-\frac{\Delta G}{RT}}$

- if $T \neq 0$, we get an equilibrium, not an answer
- reason for..

Not quite finished - initialisation



update probabilities

$$p_{i,a} = \frac{e^{\frac{-E_{i,a}}{kT}}}{\sum_b e^{\frac{-E_{i,b}}{kT}}}$$

cool

$T = 0.99\, Told$

calculate energies $E_{i,a}$

# Starting a calculation

- calculating $E_{i,a}$ requires knowing $p_{j,b}$ for each site $j$ in each state $b$
- at the start, set all $p_{j,b}$ to $^1/_{m_{state}}$

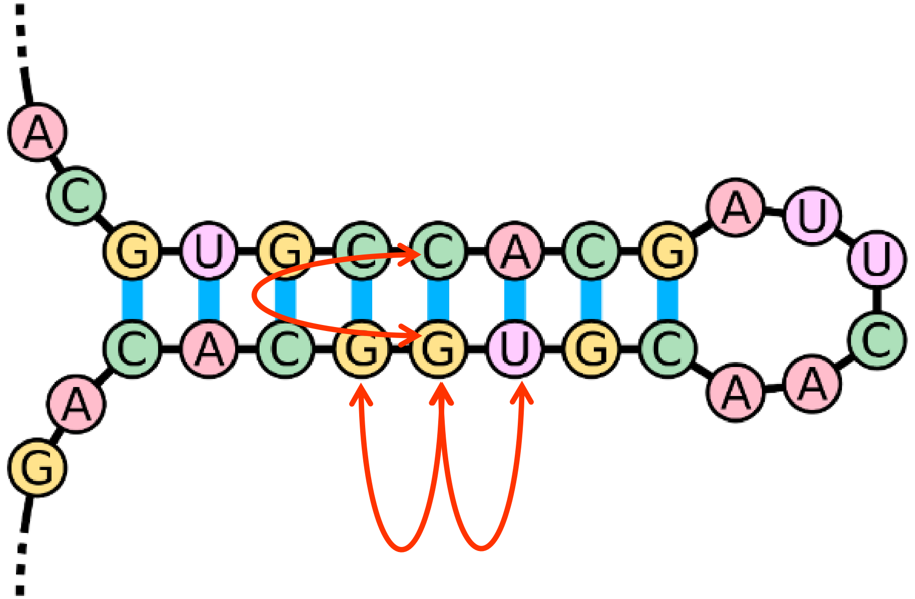
# Sequence optimisation - Another example

The question

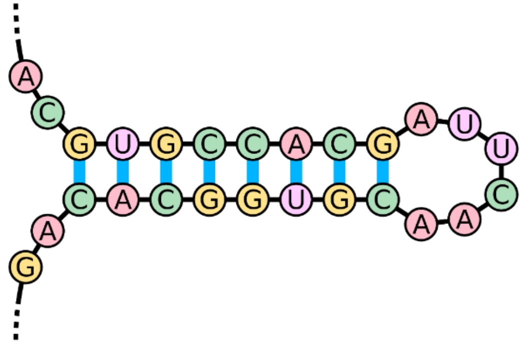
- Given a structure, find a better sequence for it
  - some energy / scoring scheme

RNA or protein ? RNA in 2D is easier – energies dominated by

- base pairing
- neighbours of $i$ ($i - 1$ and $i + 1$)



# Sequence design – the philosophy



- sequence length $n$
- each site $i$
  - can be in one of four states (A, C, G, U)
  - each of the four states has equal probability $p_{i,a} = {}^{1}/_{n_{state}} = {}^{1}/_{4}$
- for each site $i$ in sequence, calculate energy interacting with neighbours in each state
  - for one neighbour

$$E(i_a, j) = \sum_{b}^{m_{states}} \left(p_{j,b}\, E(i_a, j_b)\right)$$

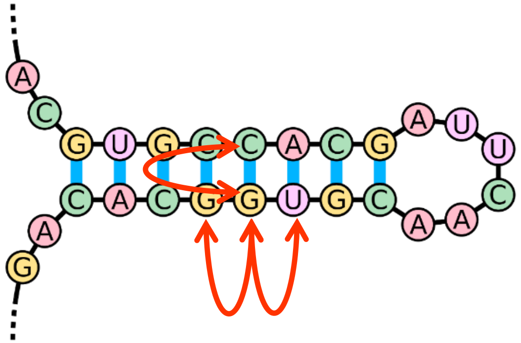
  - where the summation runs over $m_{states}$ (A, C, G, U)

- for all neighbours

$$E_{i,a} = \sum_{j}^{n_{neighbour}} \left( \sum_{b}^{m_{states}} \left(p_{j,b}\, E(i_a, j_b)\right) \right)$$

- neighbours are very clear here..
- then probabilities of states, at each site *i*

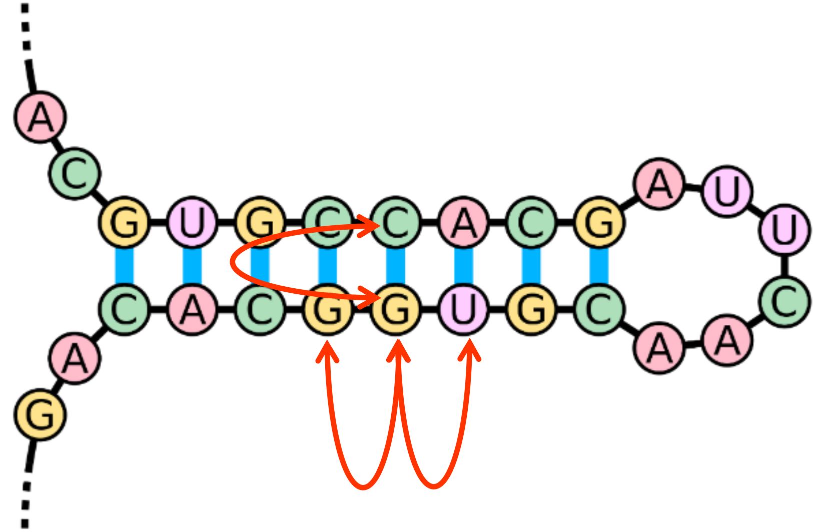
$$p_{i,a} = \frac{\exp\frac{-E_{i,a}}{kT}}{\sum_b \exp\frac{-E_{i,b}}{kT}}$$



- look at loops explicitly..

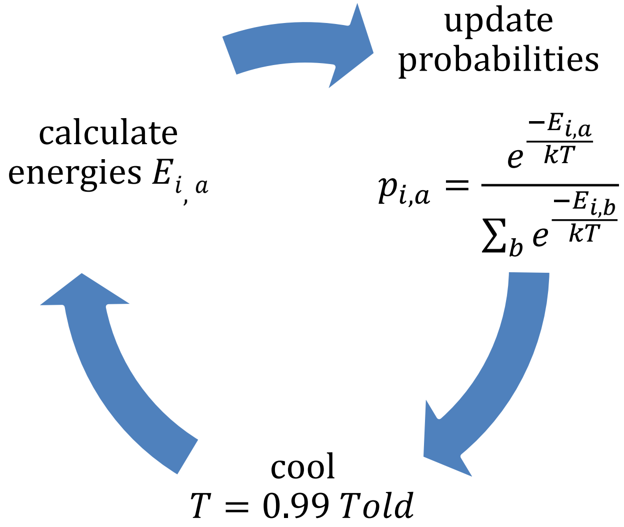
# loops for sequence optimisation

```
for each $i$
    for each state $a$ of $i$
        for each neighbour $j$
            for each state $b$ of $j$
                calculate $p_{j,b}\ E(i_a, j_b)$
                store $E_{i,a} = E_{i,a} + p_{j,b} E(i_a, j_b)$
```



Conceptually

- at the start probabilities are all equal – no sequence
- sequence emerges at $T \rightarrow 0$



update probabilities

$$p_{i,a} = \frac{e^{\frac{-E_{i,a}}{kT}}}{\sum_b e^{\frac{-E_{i,b}}{kT}}}$$

cool

$T = 0.99\,Told$

calculate energies $E_{i,a}$

# Practicalities – problems - cooling

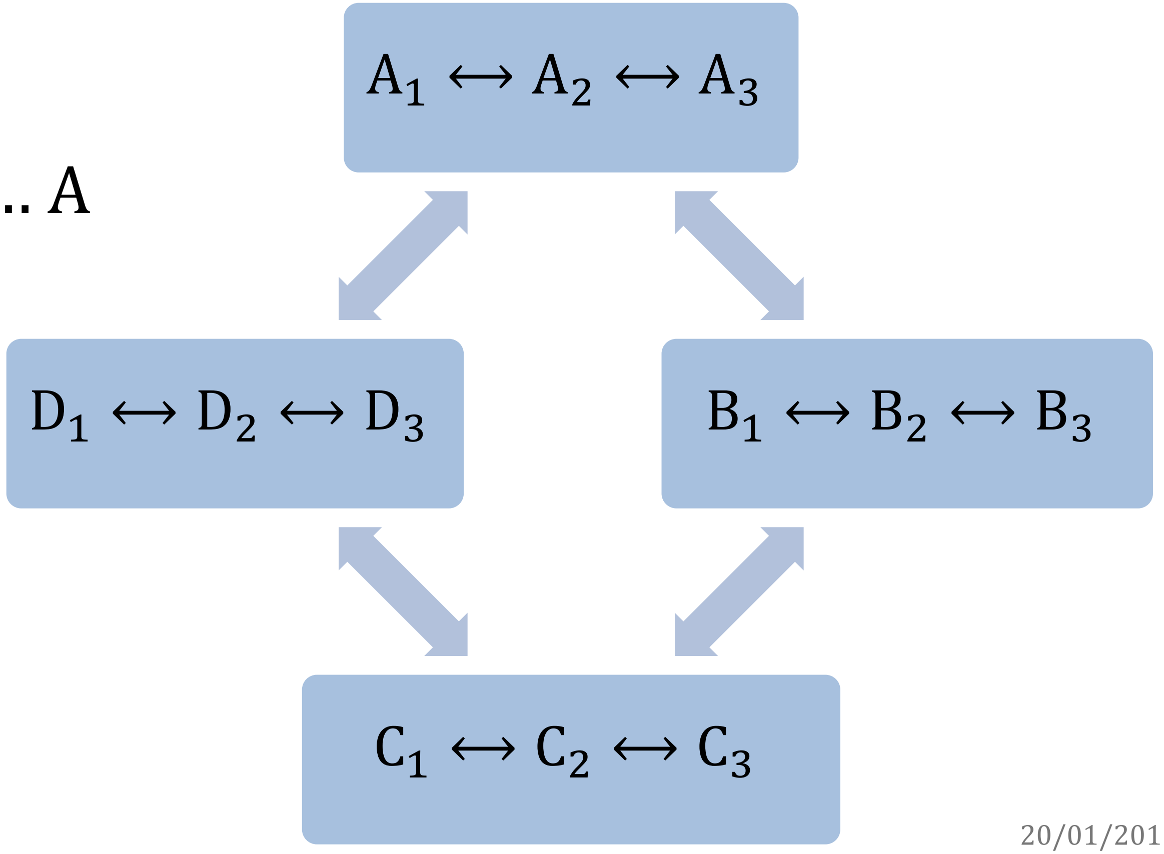
How fast does one cool ?

- $T_{t+\delta t} = 0.99\, T_t$ ?  No. Just an example
- as in simulated annealing, cool as slowly as necessary

How slow ? Remember

- State at A affects B affects C affects D … A

- cool slowly enough to let changes propagate / diffuse



$A_1 \leftrightarrow A_2 \leftrightarrow A_3$

$D_1 \leftrightarrow D_2 \leftrightarrow D_3$

$B_1 \leftrightarrow B_2 \leftrightarrow B_3$

$C_1 \leftrightarrow C_2 \leftrightarrow C_3$

# Practicalities

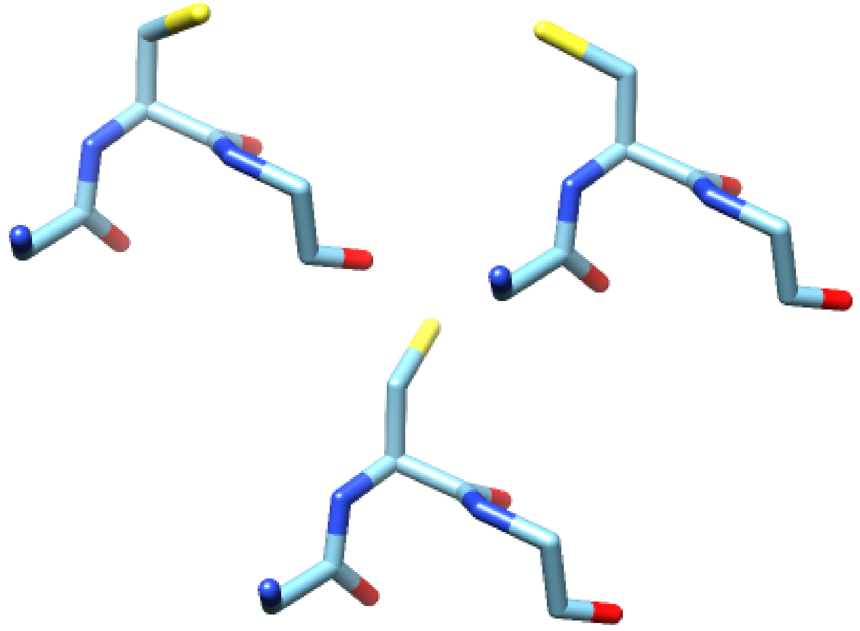
Is $T$ a real temperature ?

- here .. No.
- in some problems.. could be

Convergence – guaranteed ?

- no

# Symmetries



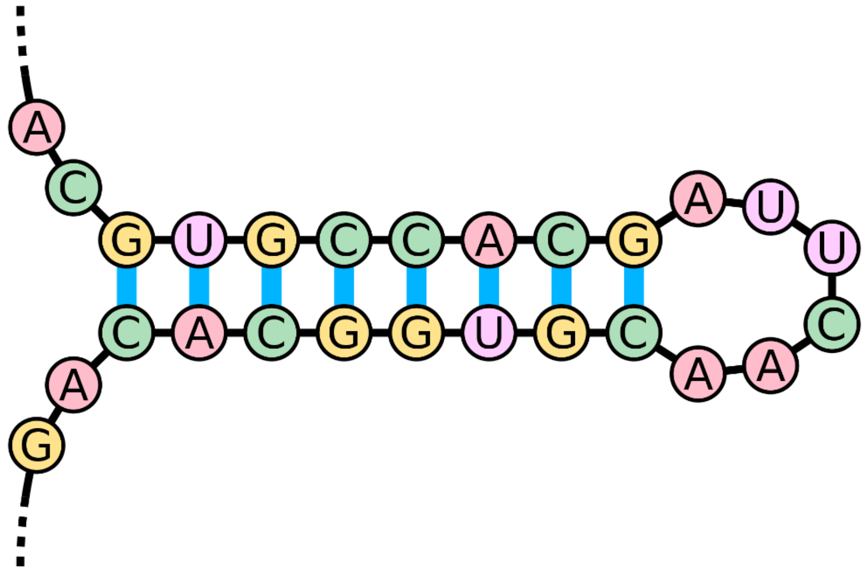
What if all rotamers have the same energy ?

- will it happen ? No
- think of interactions with backbone
  - some $E_a$ will be better than others

RNA problem

- not as symmetric as it looks ?

What if it happens anyway ?
(same energies for different states)



# Symmetries

Symmetry problems

$$p_{i,a} = \frac{1}{n_a}$$

but use

$$p_{i,a} = \frac{1}{n_a} \pm \delta \qquad \text{for some very small } \delta$$

- this is enough to make one solution preferred and dominate

Lots more problems – not here

- cooling, phase transitions, oscillations, ..

# Convergence

While cooling, monitor convergence

- how frozen is a system ?

In these systems, easy to measure

- at any site, can measure entropy

$$S = -\sum_{a=1}^{m} p_a \log p_a$$

$m$ is the number of states
$p_a$ is probability of state $a$

maximum entropy if all states equal

$$S_{max} = -m\, p_a \log p_a = -m \frac{1}{m} \log \frac{1}{m} = \log m$$

minimum

$$S_{min} = 0$$

# Convergence

entropy

- sum over all positions
- for fun – consider $\log_m p_a$ - gives nice normalisation

# what else can one do ?

- RNA base-pairing
- sequence alignments (very difficult)
- graph problems – knapsack, bipartition

- time to stop

# Summarise

Easiest on systems

- which can be discretised
- probabilities can be calculated

Works well on

- systems with many interacting parts – too big to tackle by other means
- lots of graph problems

Philosophy

- system visits all states at start
- cooling persuades it to find an answer