Self-consistent mean field methods

General method for problems with

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

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

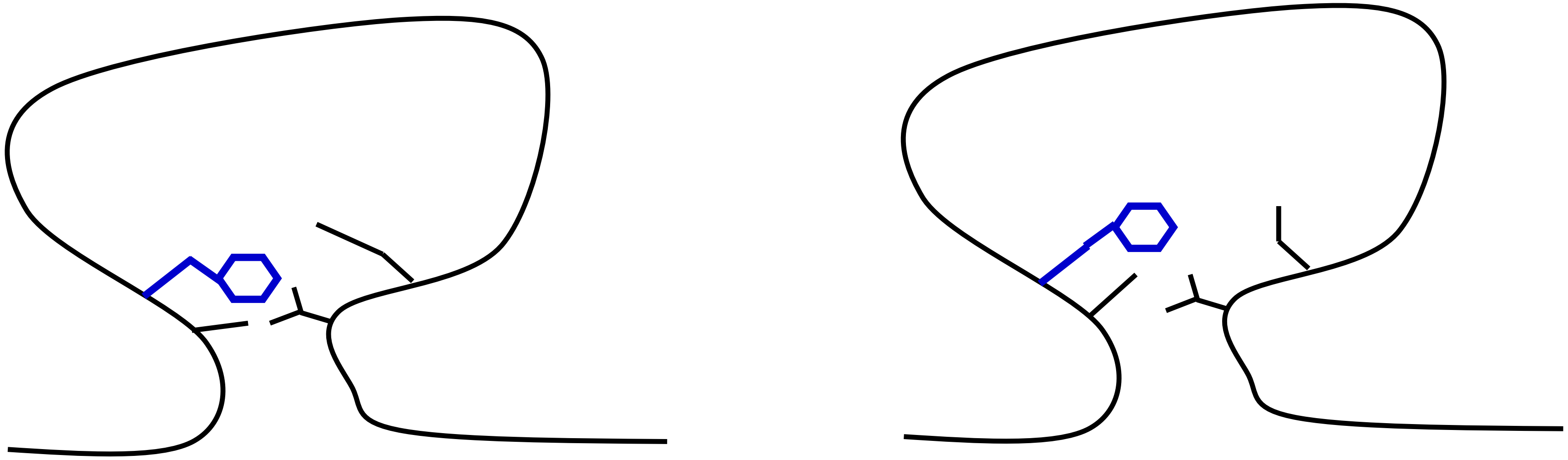
- some example problems
- Boltzmann relation
- examples in detail

Examples

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

Protein side chains

Optimise (energy) their coordinates



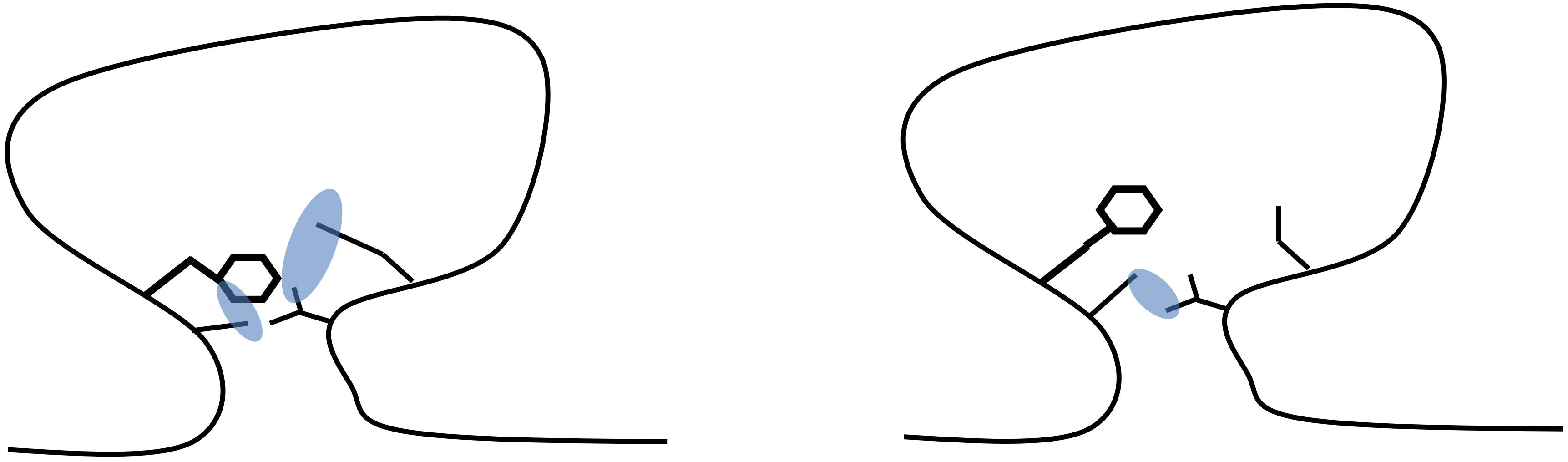
- each site interacts with his neighbours

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

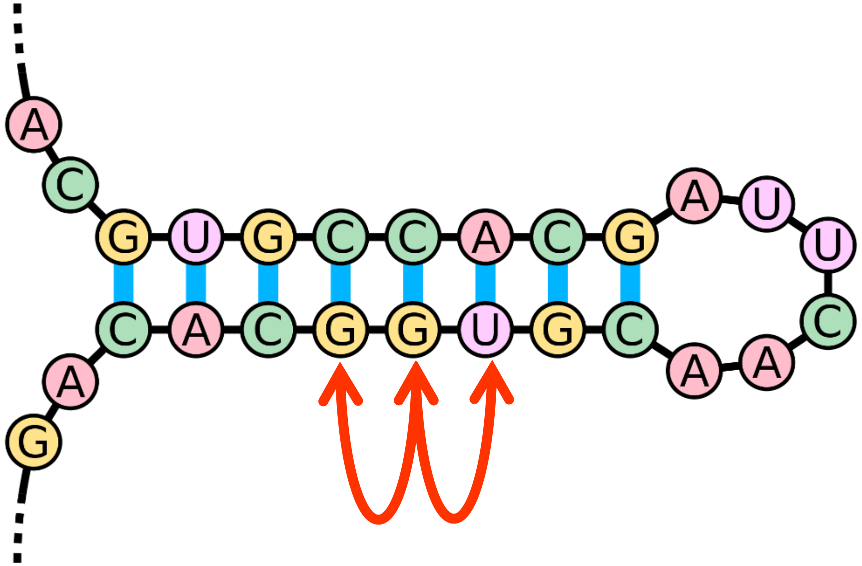
- what is the best combination ?

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



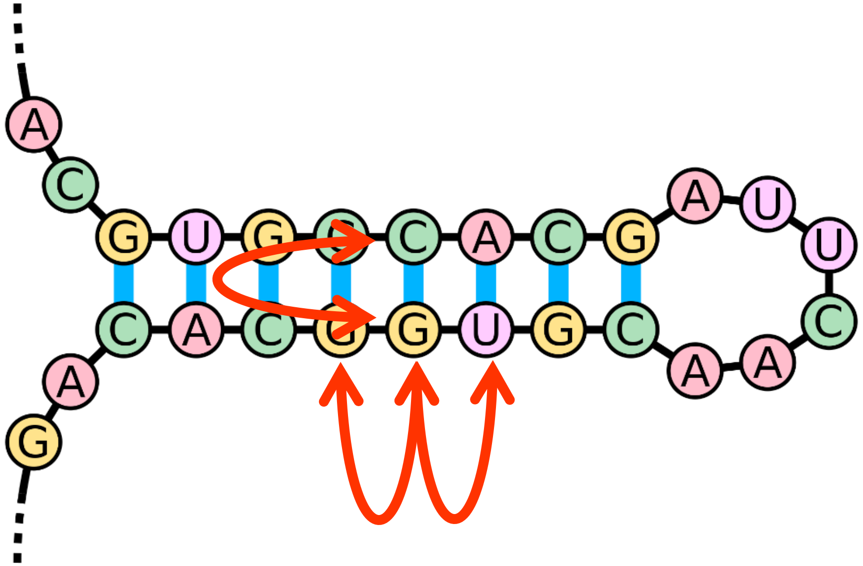
How are energies calculated ?

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)

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

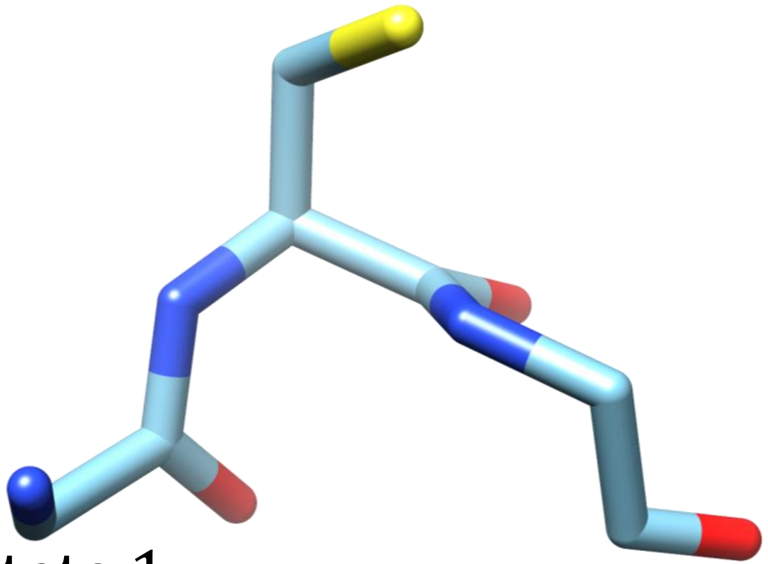
- n sites
- m states
    - m^n search space
- changing state at i affects site j which affects site k …
- sites are not independent
    - you cannot optimise i, then j, then k, …

General approach

- mean field methods / self-consistent mean field methods

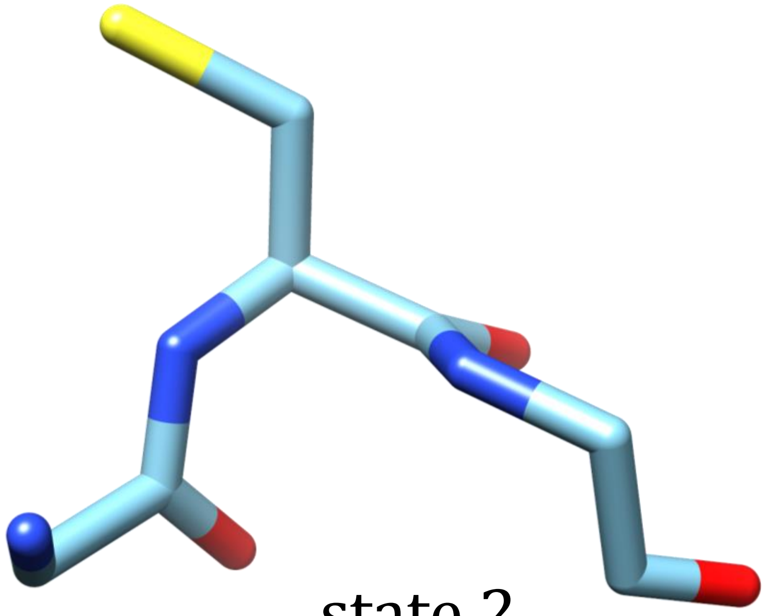
Boltzmann .. the detour

Site with two states



state 1

E_1



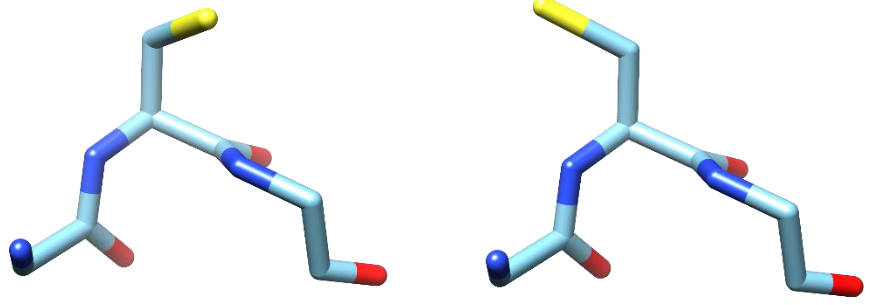
state 2

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 $\Delta G = -RT \ln \frac{[\mathrm{A}]}{[\mathrm{B}]}$

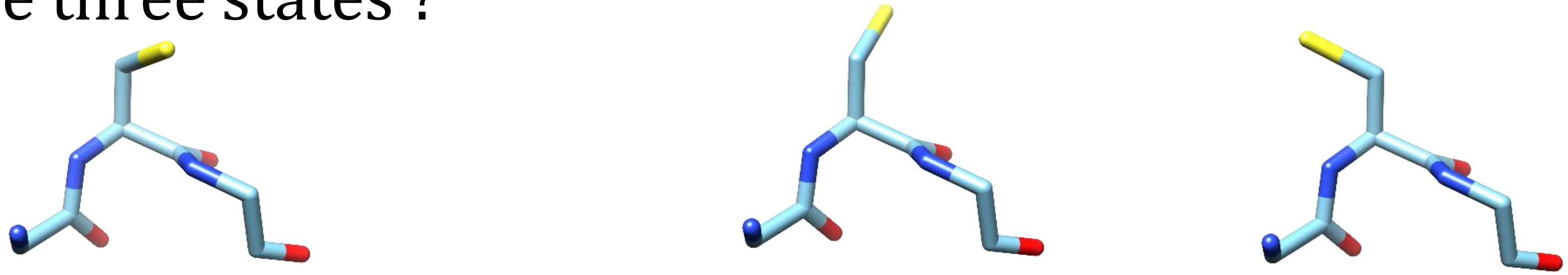
in the reaction $\mathrm{A} \rightleftharpoons \mathrm{B}$

Boltzmann – more states

$$\frac{p_i}{p_j} = e^{-\Delta E/kT} = e^{-\left(E_j - E_i\right)/kT} = \frac{\exp\left(\frac{-E_i}{kT}\right)}{\exp\left(\frac{-E_j}{kT}\right)}$$

$e^{\frac{-E_i}{kT}}$ 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

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

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

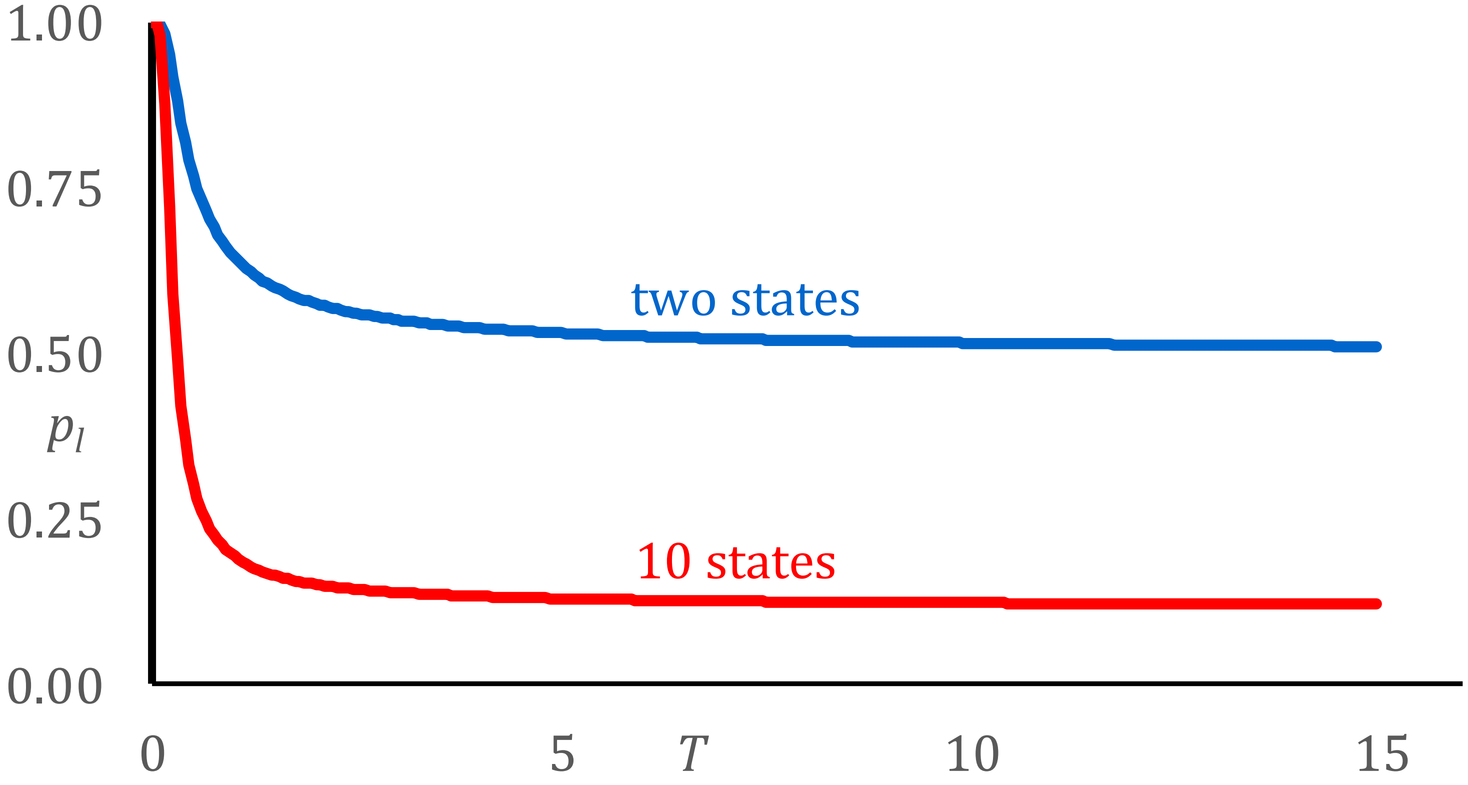
$T = 0$

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

Optimisation

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

For these lectures – aim is

- 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

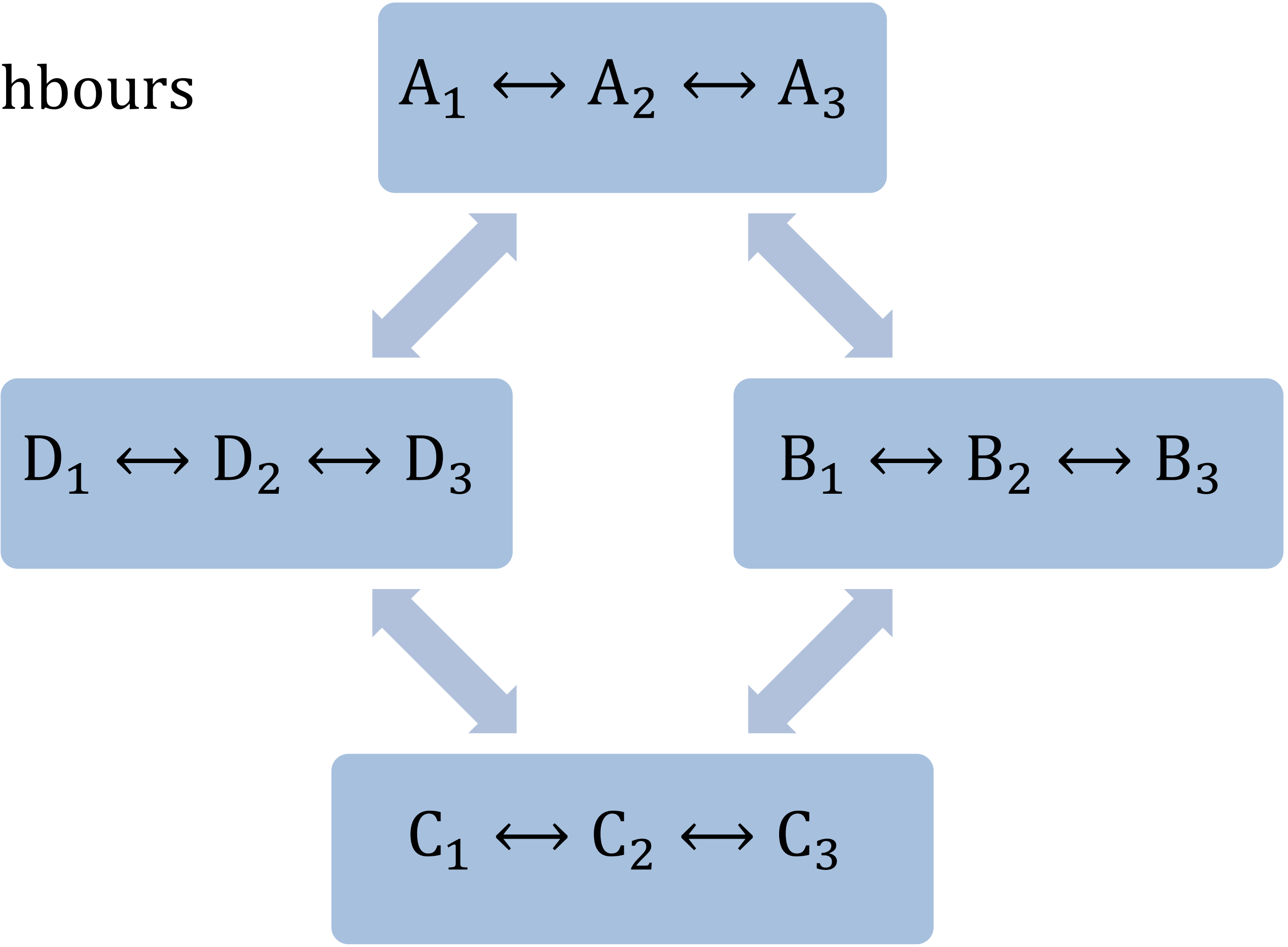
Gradually cool

- each site moves to lowest energy states

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

- no

Each site affects his neighbours



$A_1 \leftrightarrow A_2 \leftrightarrow A_3$

$D_1 \leftrightarrow D_2 \leftrightarrow D_3$

$B_1 \leftrightarrow B_2 \leftrightarrow B_3$

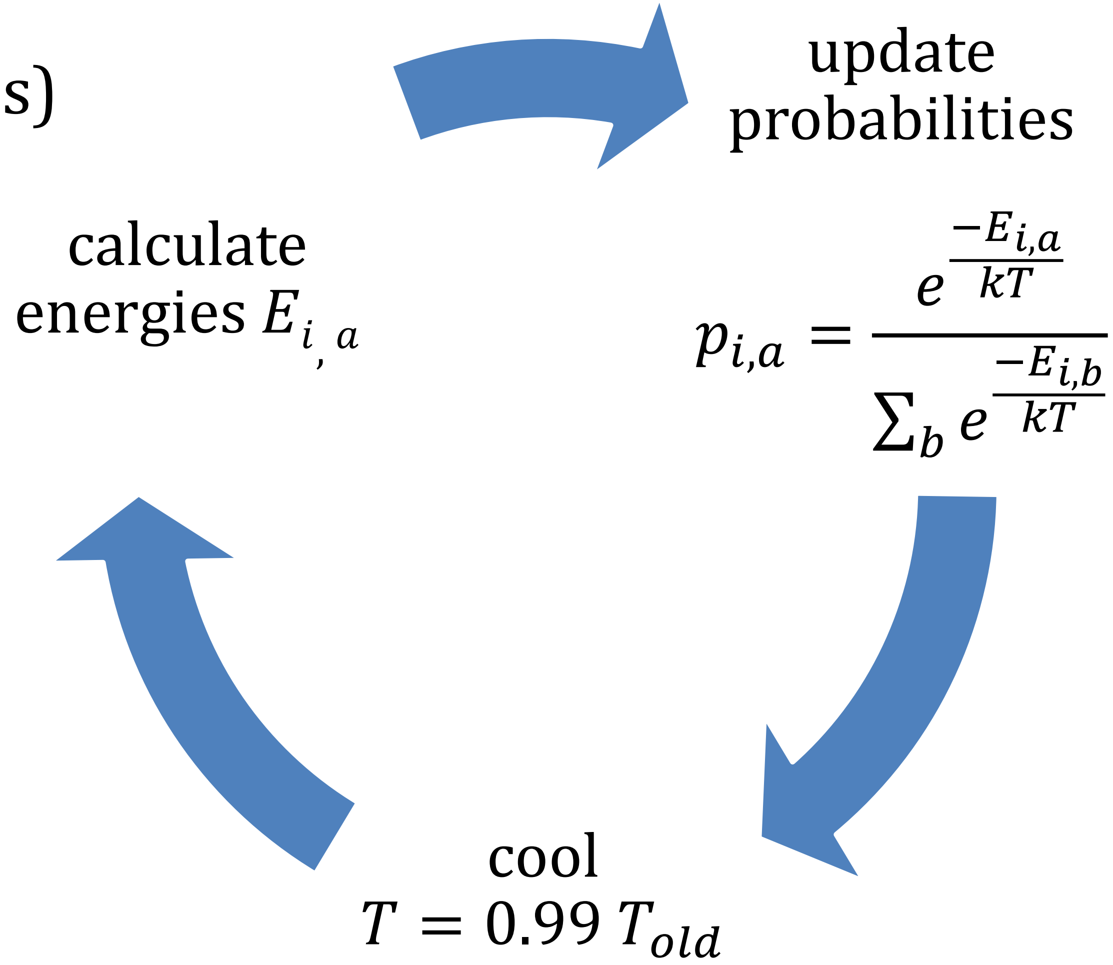
$C_1 \leftrightarrow C_2 \leftrightarrow C_3$

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…



calculate
energies $E_{i,a}$

update
probabilities

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

cool
$T = 0.99\, T_{old}$

Examples – Sidechain conformations

Assume

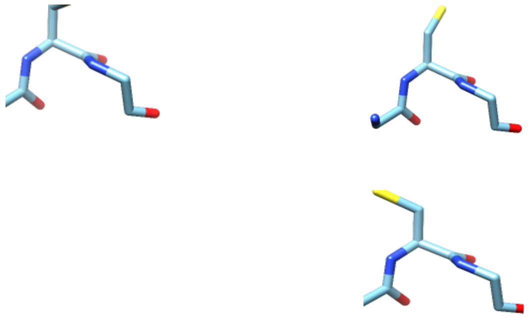
- some model for energy
- discretisation – sidechain rotamers
    - residue i can exist in m conformations

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 at site i in state a with j in state b

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

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

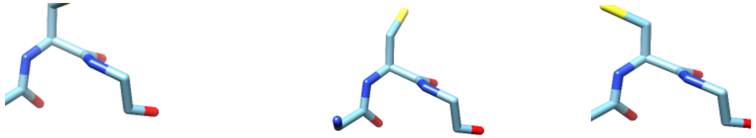
this is for one neighbour, but we want the total energy $E_{i,a}$

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

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 $p_{i,a} = \frac{\exp\frac{-E_a}{kT}}{\sum_{b=1}^{N_{states}} \exp\left({}^{-E_b}/_{kT}\right)}$



In words…

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

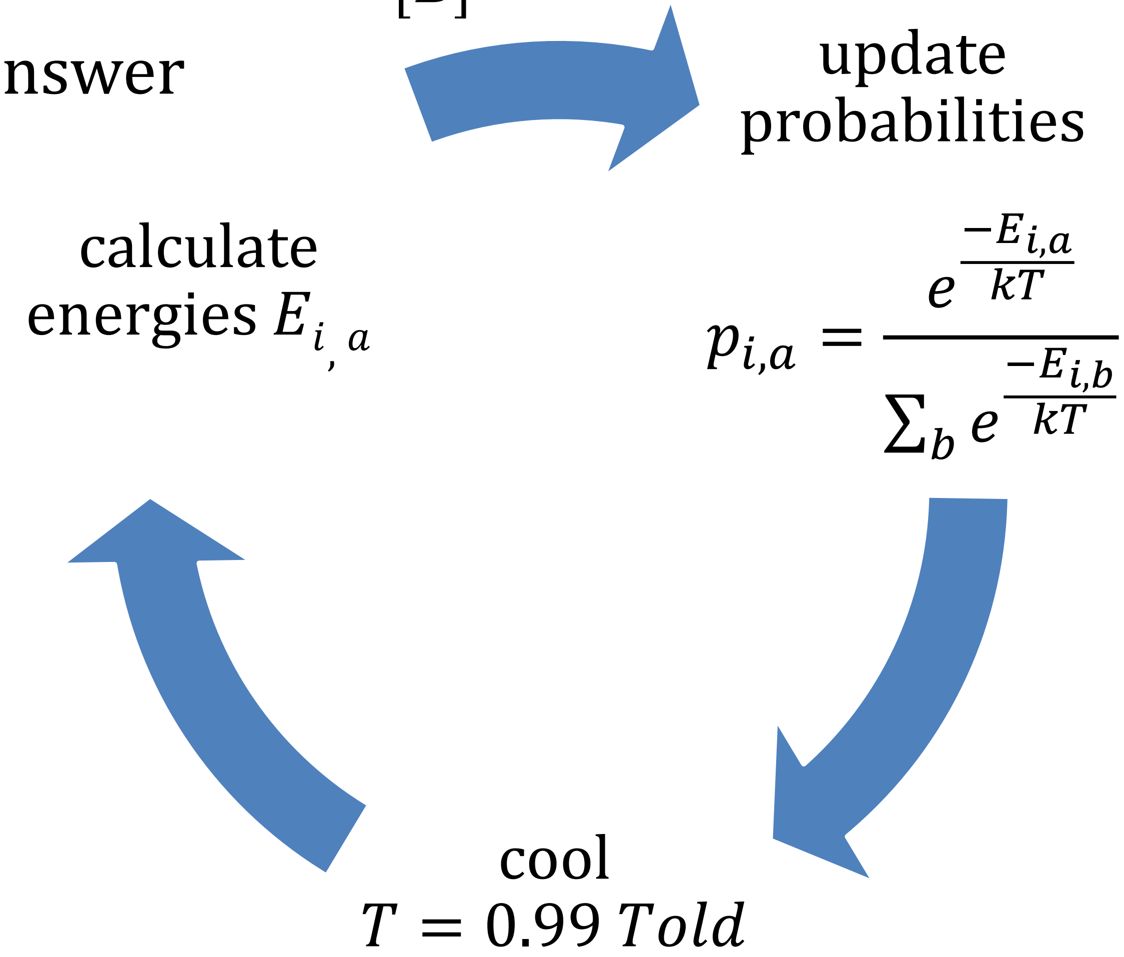
still not finished

Why cool ?

Remember $\Delta G = -RT \ln \frac{[\mathrm{A}]}{[\mathrm{B}]}$ in the reaction A ⇌ 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 ${}^{1}/_{m_{state}}$

Ende für Wintersemester