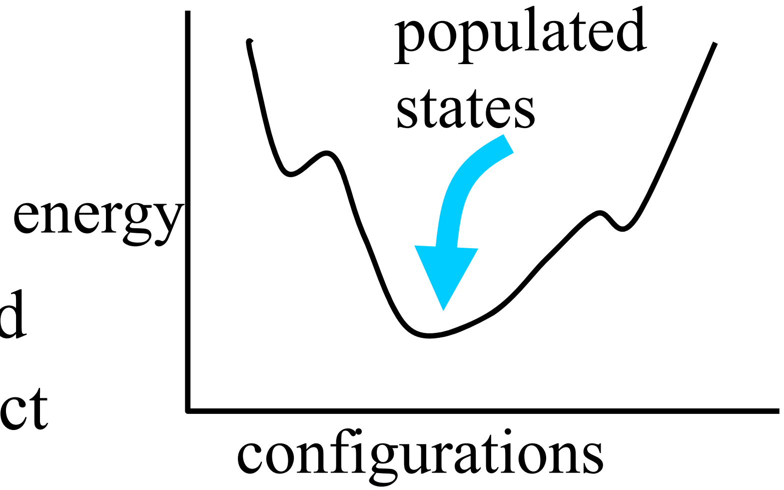
RNA Folding / Kinetics

Andrew Torda, 2009, RNA

- Predicted free energy conformations ?
- Does RNA find them ?
- Does RNA have some help ?

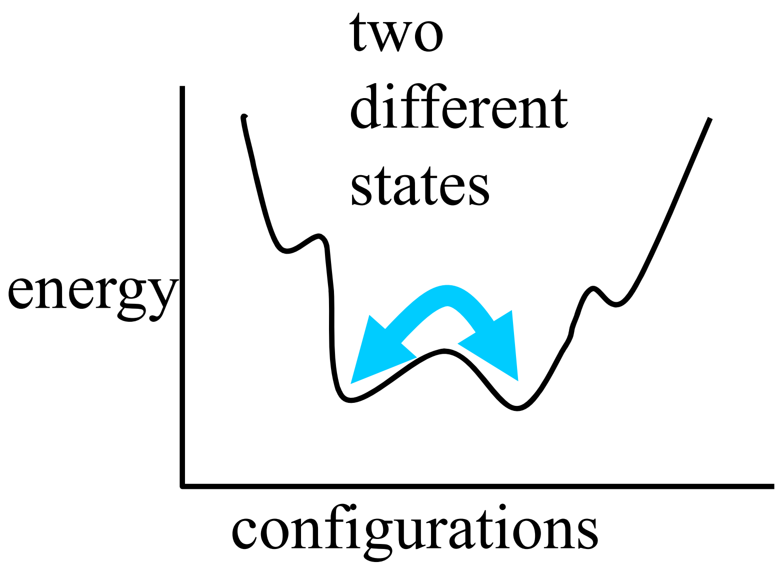
- First … rules

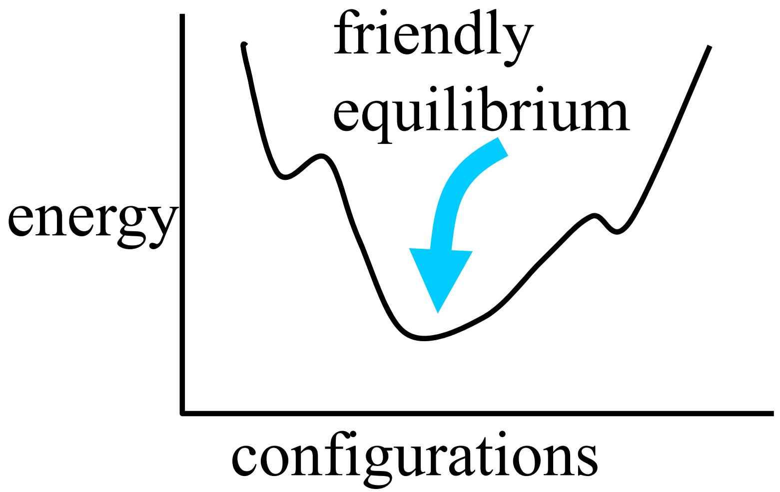
- Equilibrium / ideal world
  - lowest energy most populated
  - Boltzmann distribution perfect

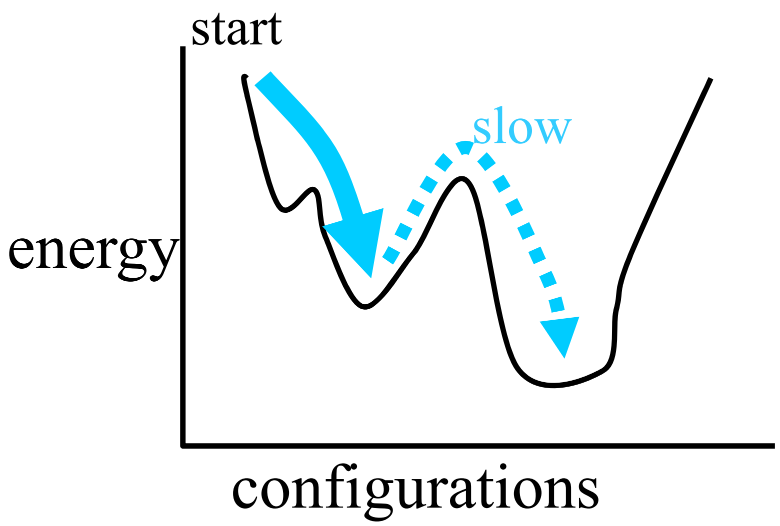


- other possibilities

Energy landscapes



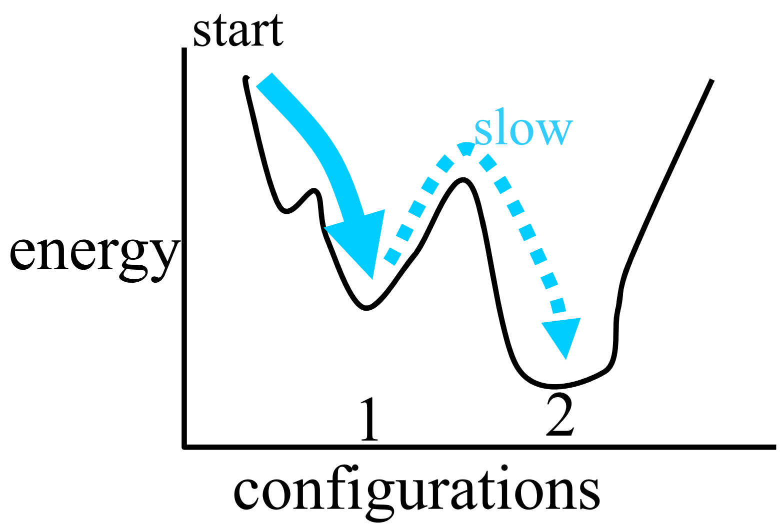




Nasty kinetics

possible

- 1 → 2 transition slow
  - never happens or
  - RNA is degraded



- consequences
  - predicted free energy minimum is not helpful
  - people try to estimate barrier heights

Predicting kinetics

- As in protein lectures
  - possible with simple models
  - still rather difficult
- Approaches
  - big simulations
  - big searches

- Kinetics
  - examples of more general methods

Brute force simulations

- MD or Monte Carlo style ?

- Energy model – is the classic Nussinov or nearest neighbour model friendly ?
  - in both forms $E = \begin{cases} 0 & \text{base pairs not formed} \\ \text{favourable} & \text{bases paired} \end{cases}$

  - not differentiable function – no forces – not friendly

- two possibilities for dynamic simulations
  1. different energy model (not discussed here)
  2. discrete methods (here)

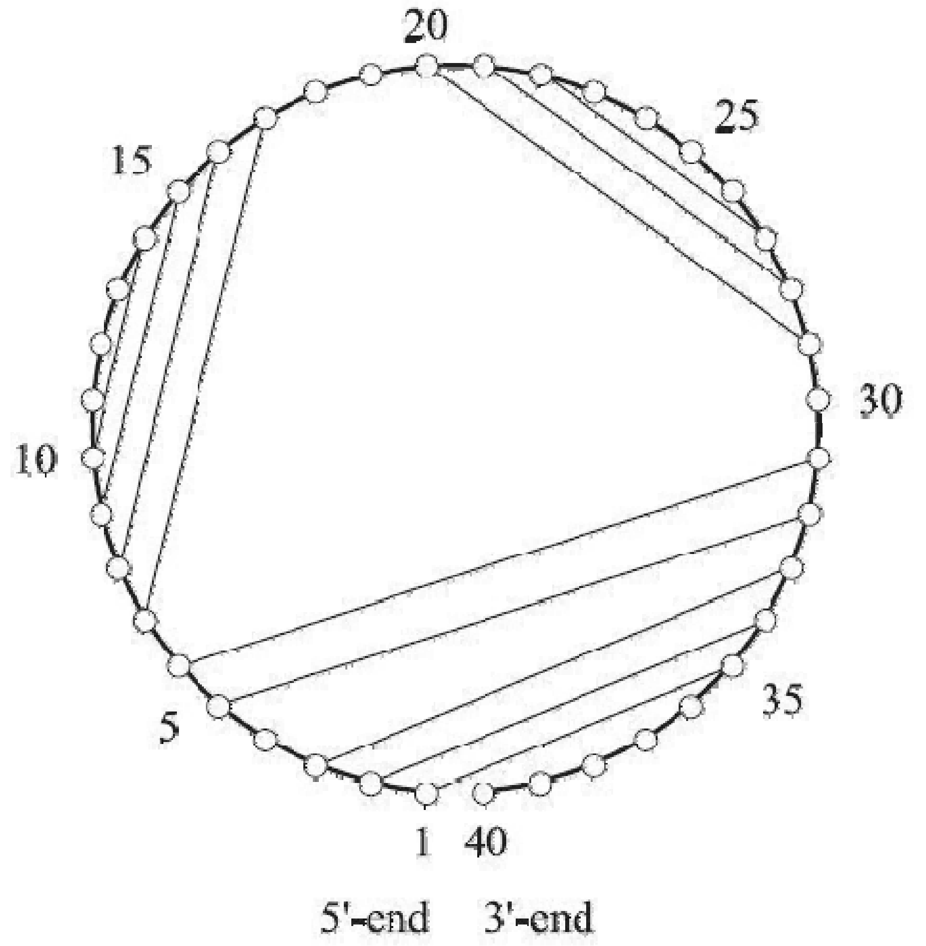
Monte Carlo like methods

- Normal Monte Carlo
  - any random, unbiased or non-physical moves OK
  - no attempt to model time – not normally relevant

- Claim – act of faith – belief – dream
  - select a move set which you believes models physical moves
  - the simulated system might reflect physical processes
  - what would the moves look like ?

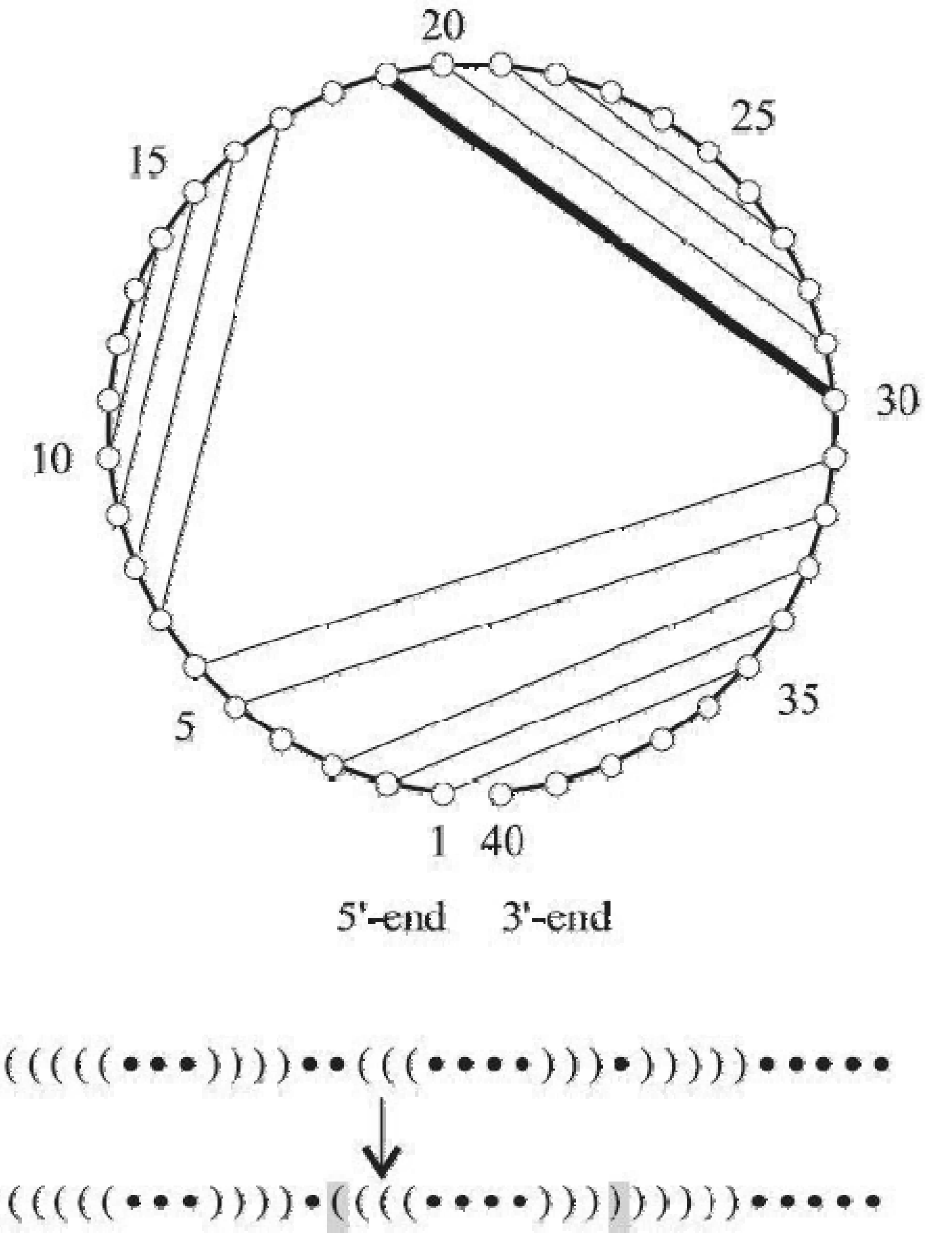
A move set for RNA

- add / remove a base pair



(((•((((((•••))))••(((••••)))•)))))•••••

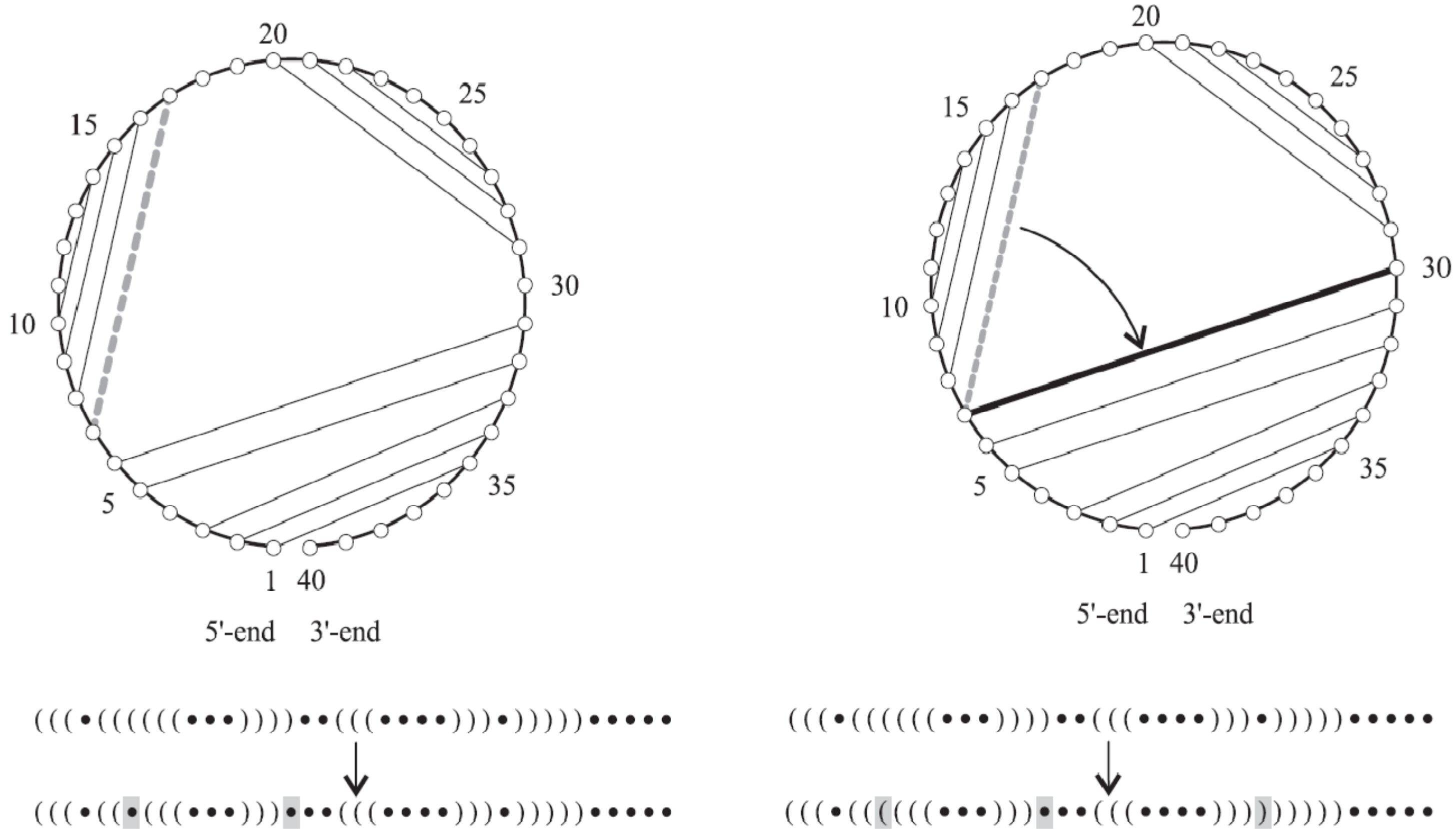
A move set for RNA



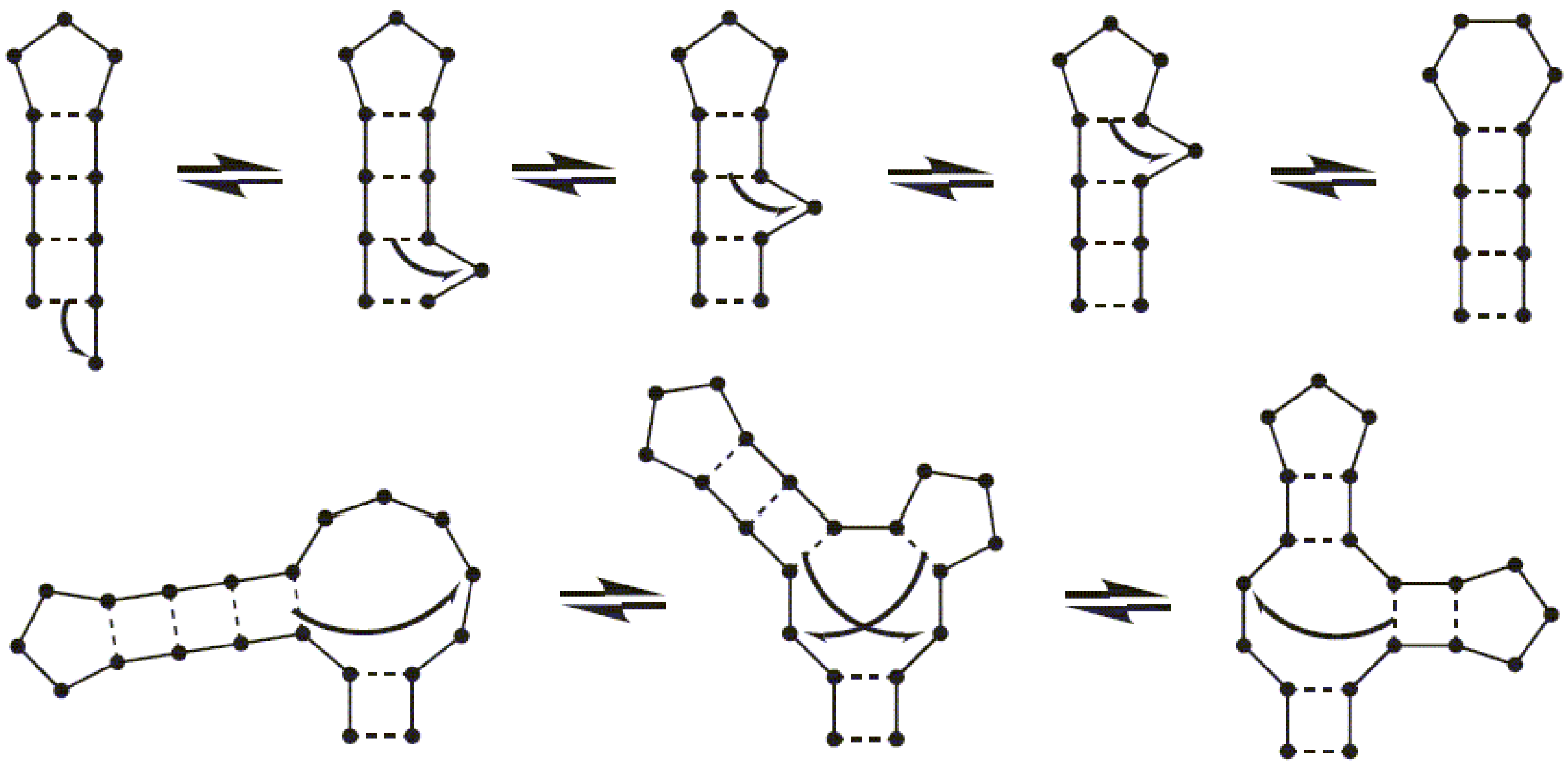
Flamm, C., Fontana, W., Hofacker, IL, Schuster, P, RNA, 6, 325-338, RNA Folding at elementary step resolution

Flip a partner

- looks easy
  - how much of a rearrangement does it mean ?



Diffusion of bulge



Flamm, C., Fontana, W., Hofacker, IL, Schuster, P, RNA, 6, 325-338, RNA Folding at elementary step resolution

Very naïve method

- Given this move set
  - start from unfolded RNA
  - try to fold it – see how fast a predicted structure is formed

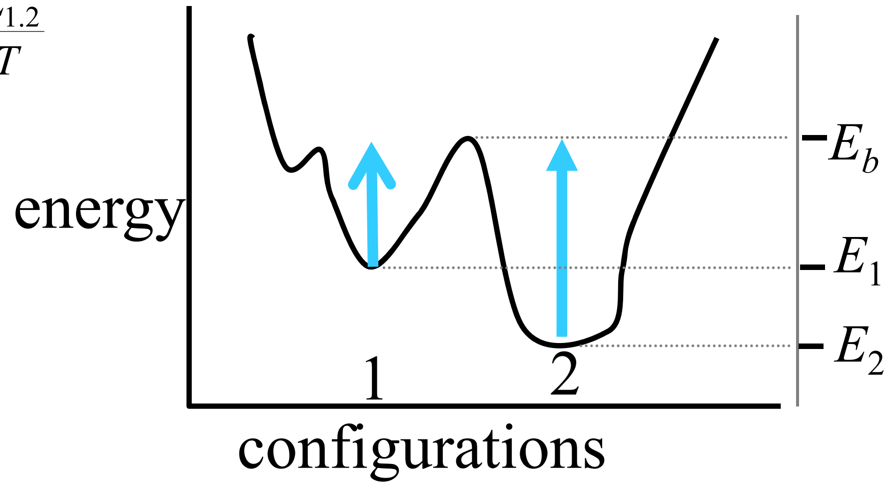
- more specific questions
  - from conformation 1 or 2 how fast is 2 or 1 populated ?

- will not work well..

Why is kinetics difficult

- different to earlier lectures
  - talk about rates
  - equilibrium

$$\frac{p_1}{p_2} = e^{\frac{\Delta E_{1,2}}{kT}}$$

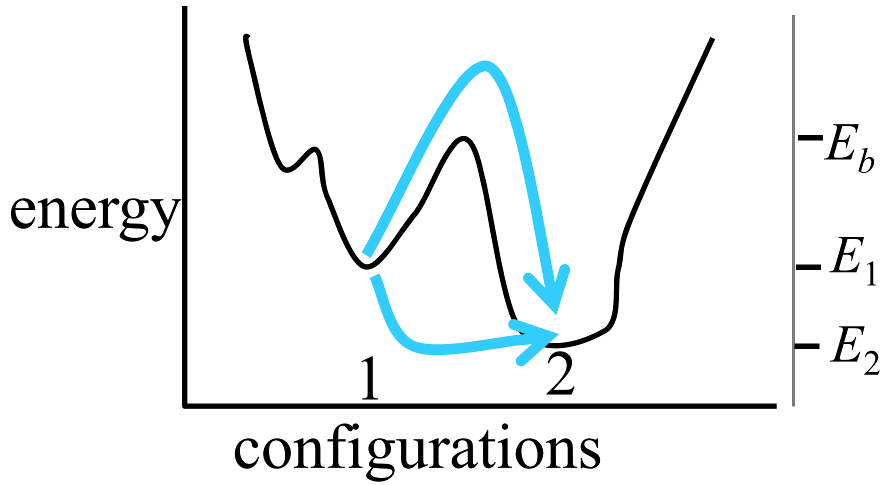


- $p_{1,2}$ depends on $e^{E_b - E_2 / kT}$
- $p_{2,1}$ depends on $e^{E_b - E_1 / kT}$
- but does one know $E_b - E_1$?

Kinetics is very difficult

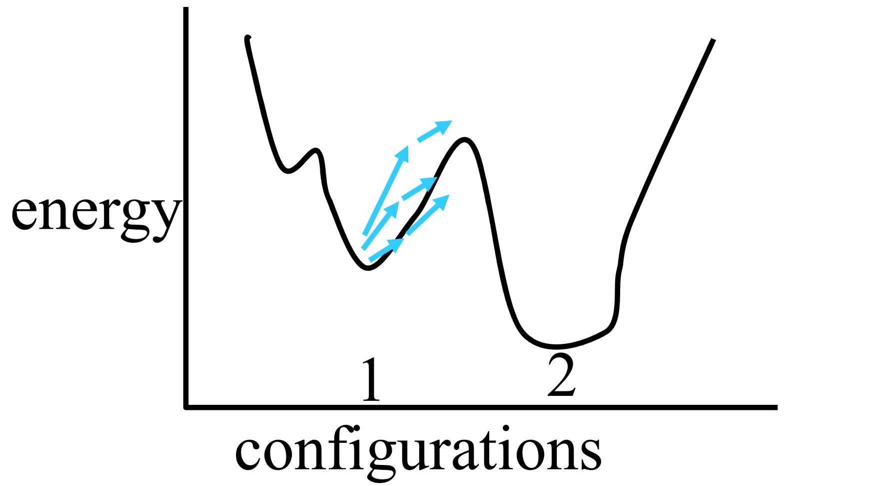
Idea of one barrier is not realistic

- lots of possible routes
  - each has its own rate
- final rate depends on flux through every path



The answer

- transition matrix / rate matrix / master equation…



A matrix approach

- Probability p_{jk} of going from k to j
- rows must sum to 1

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{bmatrix}$$

Example use

- at time t, system has vector $\mathbf{v}_t$ of being in each state
- at next time step $\mathbf{v}_{t+\delta t} = \mathbf{P}\mathbf{v}_t$

- If I apply this infinitely, we get an equilibrium distribution
  - interesting, but not helpful here

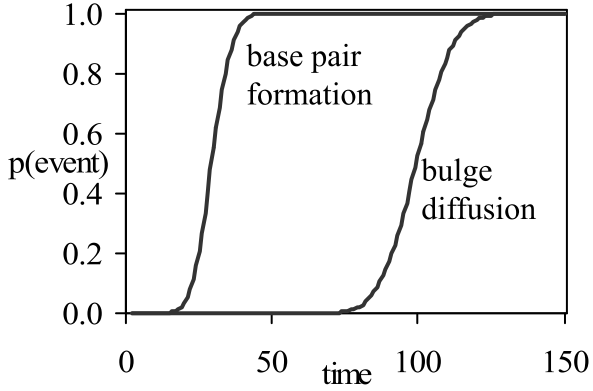
- Can we easily guess at the rate of transitions from i to j ?
  - not really .. how many states do we have ?
  - how would I get rates across all different paths ?

A practical approach

- What is the value of a matrix element ?

$$\frac{k_{ij}}{k_{ji}} = e^{\frac{-\Delta G_{ij}}{kT}}$$

- looks like normal Monte Carlo
- Add kinetic element
  - assign characteristic time (distribution) to each move
    - forming a base pair is fast
    - moving a bulge is slow
- simulation scheme



Simulating with time

pick starting conformation

```
while (not finished)
    choose δt from poisson distribution
    from n move types calc p_i that move type i happened
    choose move according to p_i
    try move - accept/reject
```

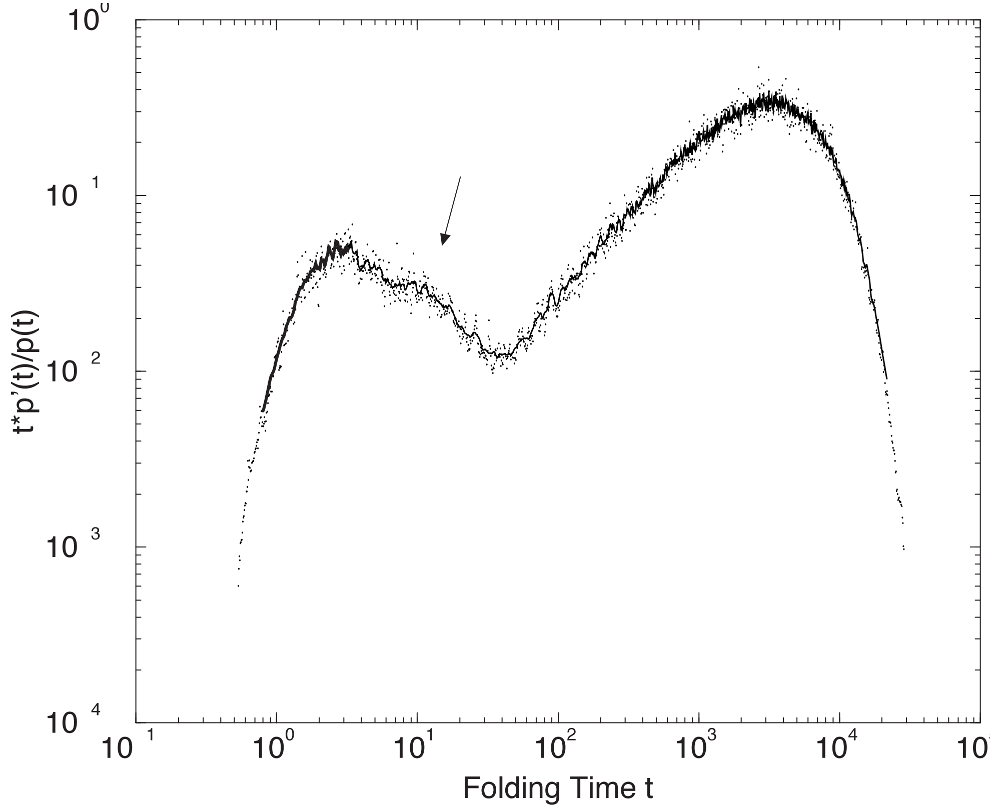
- result ?
  - from many short steps only a frequent (base pair formation) is tried
  - occasionally a less frequent step is tried

Beliefs

- energies – as in any scheme
- time scales – very difficult
- have you really captured the correct moves ?

Example result

- folding of a hairpin
- two dominant paths to final structure



Flamm, C., Fontana, W., Hofacker, IL, Schuster, P, RNA, 6, 325-338, RNA Folding at elementary step resolution

Landscape approach

- What would one like ?
  - complete picture of energy landscape
- Simplify
  - of the astronomical number of conformations
  - only a finite number are relevant

- Ingredients
  - literature model for energies
  - method to find all N_{low} structures within x kJmol^{-1} of best
    - N_{low} may be 10^6 or 10^7
  - sort this list

Landscape / barrier approach

- set up long list of conformations (10^6 or 10^7)
- set up list of basins (minima)
- set up list of transition / saddle points

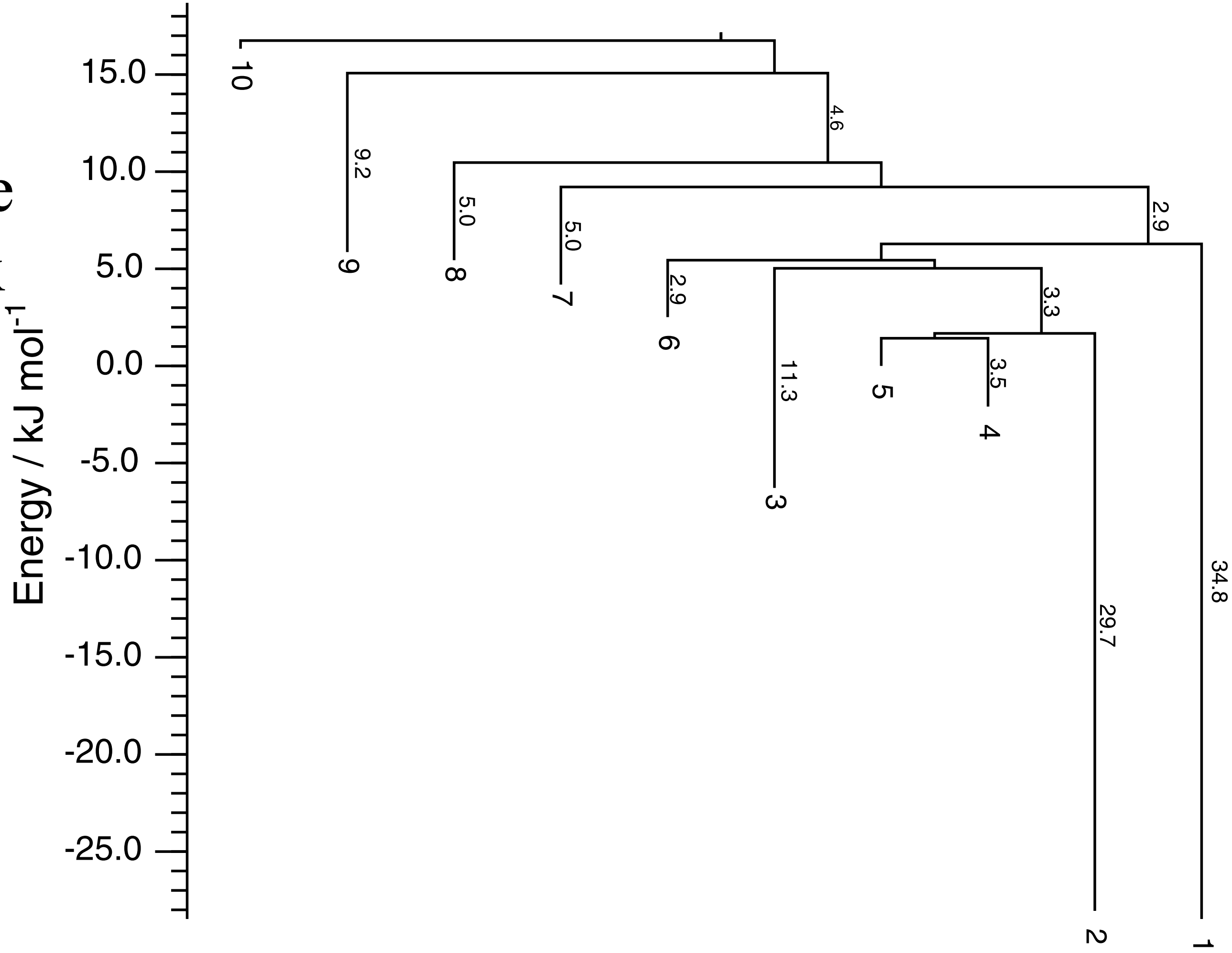
```
for each point x in sorted order
    build list L of neighbours (structures with single b.p.
  change)
        if all members of L are new
            add x to basins
        else
            add x to saddle points
```

- result
  - a list of minima with connecting saddle points

Landscape barriers

- we have a list of likely conformations
- we have a list of likely barriers

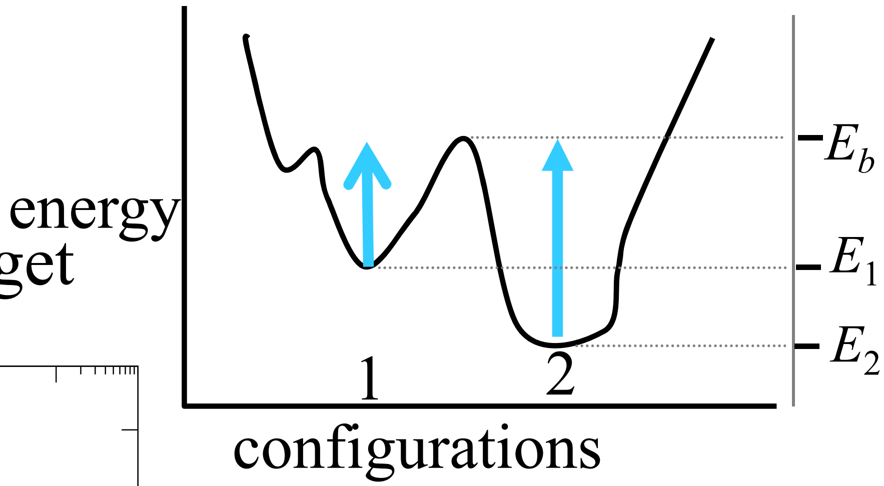
- now can really use a transition matrix approach

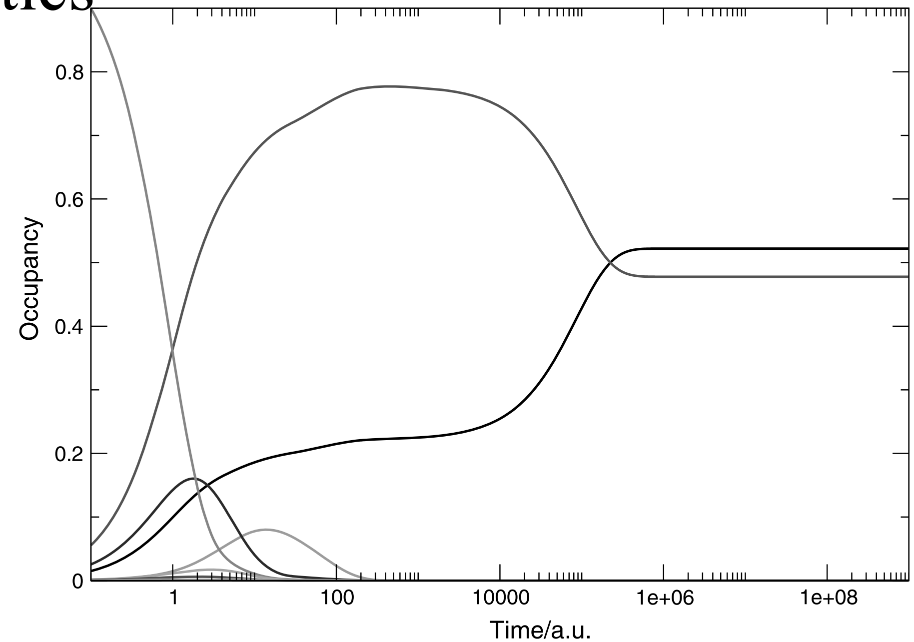


Using landscape barriers

- for any pair of minima we have a barrier height
- can calculate $p_{1,2} = e^{E_b - E_1 / kT}$

- use the transition rate matrix to get kinetics





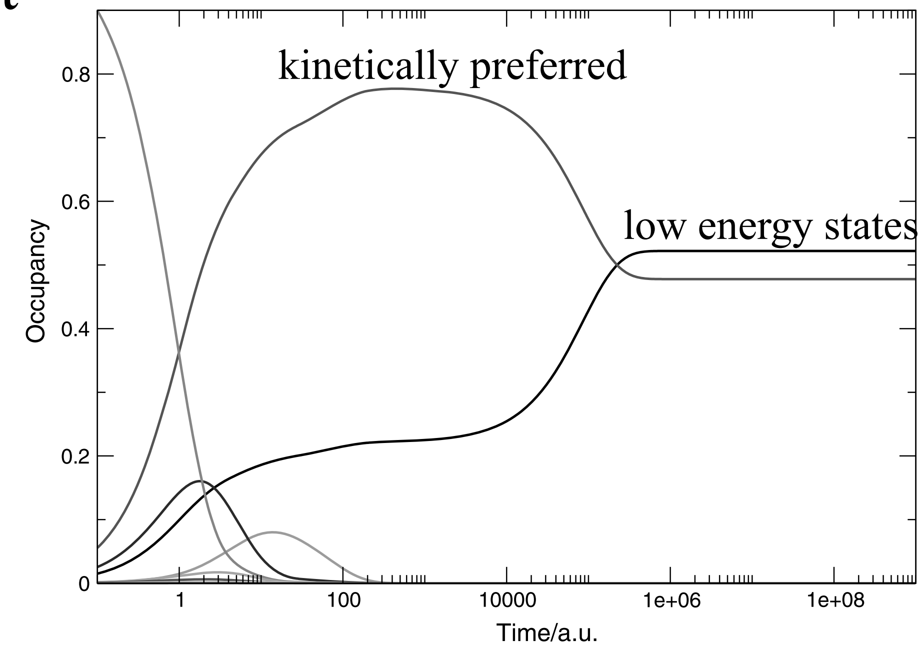
Flamm, C & Hofacker, I.L., Monatsh Chem 139, 447-457 (2008) Beyond energy minimization …

Assumptions / Implications

- assume we have not neglected too many relevant states
- great trust of energy model

- important
- the lowest energy state may not be the most populated

- if RNA is degraded ?



Biochemical complications

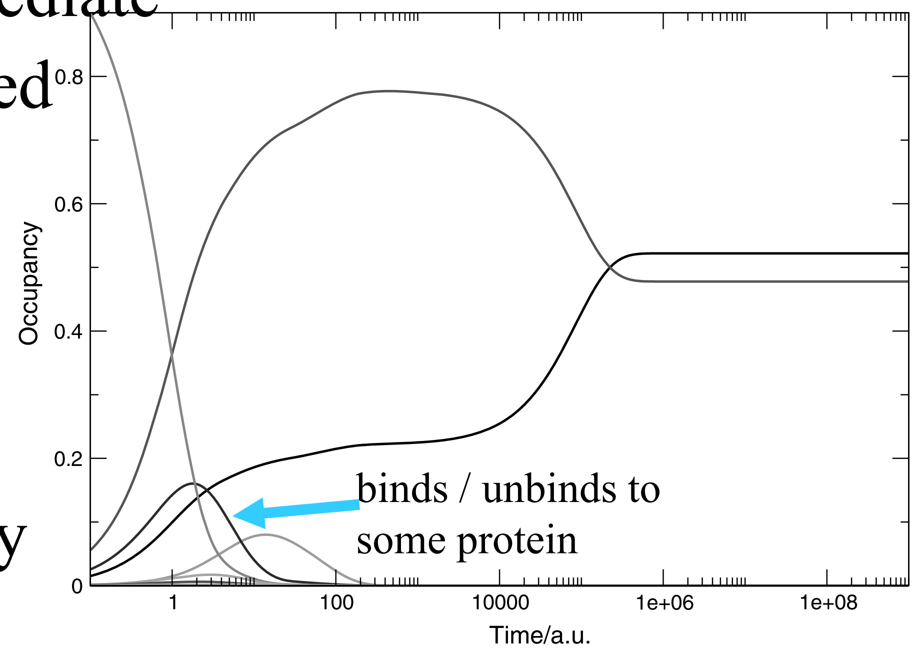
- If RNA folds by itself, one can try to model folding / kinetics

- RNA chaperones
  - very popular belief

- Rules
  - if they do not consume ATP (energy)
    - they cannot disturb equilibria
    - they could disturb pathways

Example kinetic complication

- fictitious
- if protein binds to some intermediate
  - some pathway may be slowed



- stories
  - chaperone "destabilises" misfolded structures
  - hard to justify on free energy terms
    - implies distortion of energy surface

Summary

- even good energy models for RNA are similar to discrete models in protein / polymer world
  - heavily discretised
- major assumption
  - one can either
    - simulate the system directly
    - obtain kinetics from simple transition matrix
- regardless of details
  - minimum (free) energy structure prediction may not be sufficient