RNA base pairing
self consistent mean field approach

Andrew Torda, Hamburg

How to recognize a brave protein calculator ?

- rescue drowning babies ?
- take a job with a Liechtenstein Treuhander ?
- administrator in the Kenyan electoral commission ?
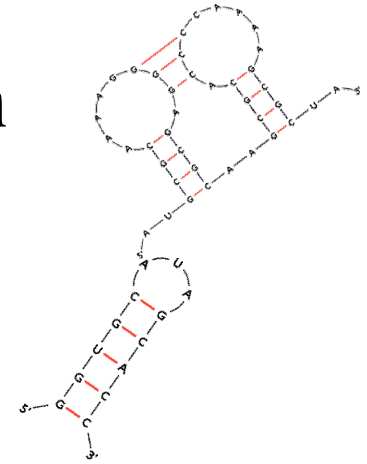- director of WestLB bank ?

RNA base pairing
self consistent mean field approach

Andrew Torda, Hamburg

How to recognize a brave protein calculator ?
- rescue drowning babies ?
- take a job with a Liechtenstein Treuhänder ?
- administrator in the Kenyan electoral commission ?
- director of WestLB bank ?

- give a talk on RNA to RNA people

    - a method for finding basepairs
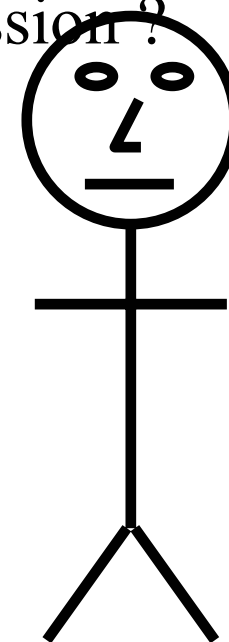   - pseudoknots ? give me more.
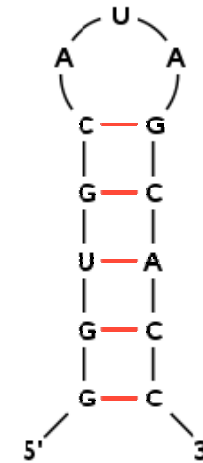
RNA base pairing
self consistent mean field approach

- work of Jens Kleesiek
- RNA secondary structure / base pairing
  - maybe first step to 3D structures
  - much easier to calculate
- energy functions
  - yes / no
  - horrible decomposition
    - interpret in terms of bases
    - calculated by pairs of pairs
  - discrete view of world
- using them..

Two Views Of Problem
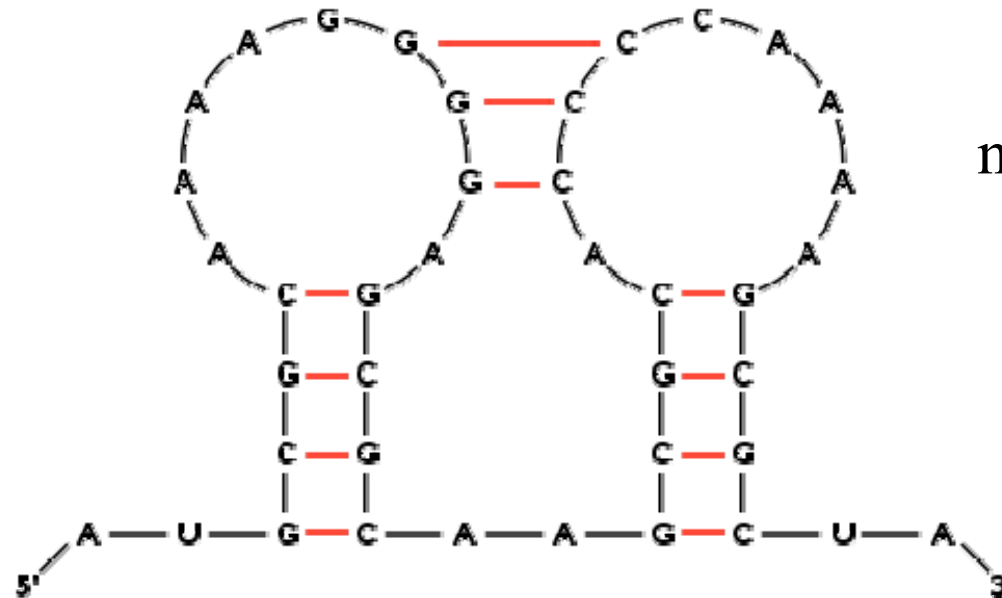
Impose some restrictions
- ordering / crossing of pairs
- dynamic programming problem
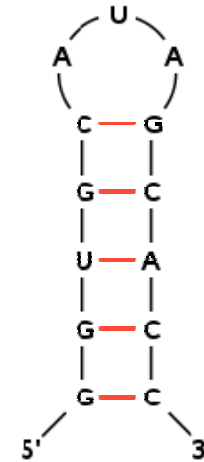
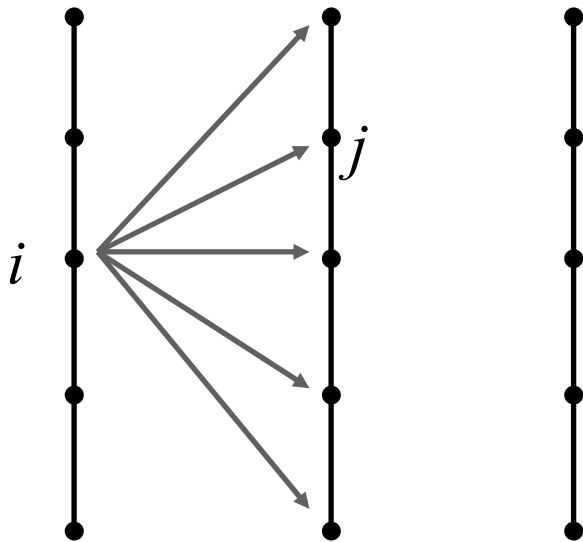easy

Problem ?
- pseudoknots

nasty

- alternative view ..

Two Views Of Problem

Impose some restrictions
- ordering / crossing of pairs
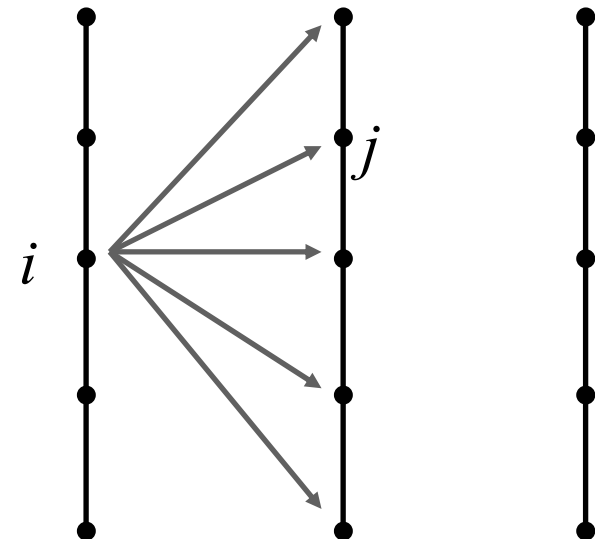- dynamic programming problem

Alternative

- i can pair with some j
- restricts other possibilities
  - nobody else can pair with j
- very limited set of states
- goal
  - best energies
  - consistent set of pairs

SCMF

- limited set of consistent states ?
  - just like wave functions or side chains
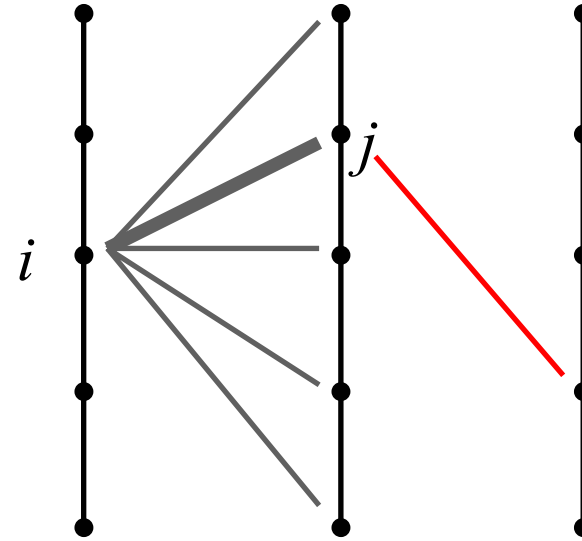- what is the probability of being in a state ?

$$p_i = \frac{e^{-E_i/kT}}{\sum_j e^{-E_j/kT}}$$

- high temperature
  - all states possible
  - system is in many states at once
- low temperature
  - lowest energy states more likely
  - very few states possible

Consistency

- being in one state means you are not in another
- probabilities propagate through system – takes time
- mean field ?

- Scheme

start warm
while (not converged)
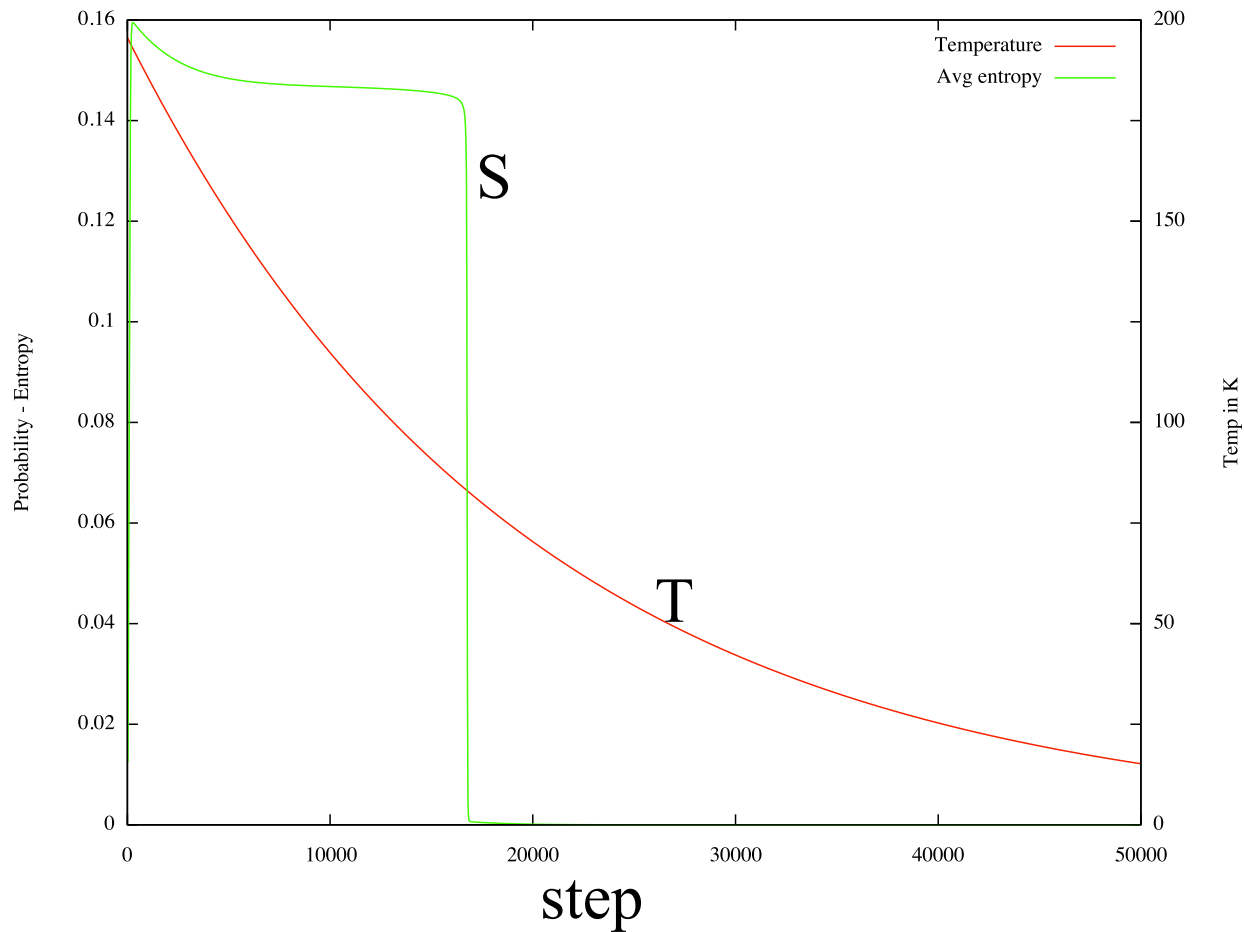   calculate energies of all states
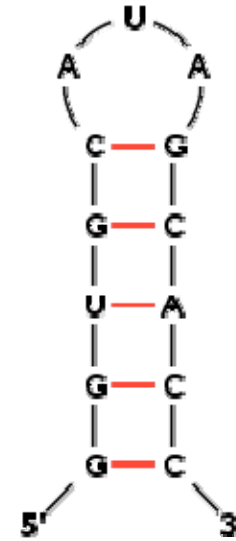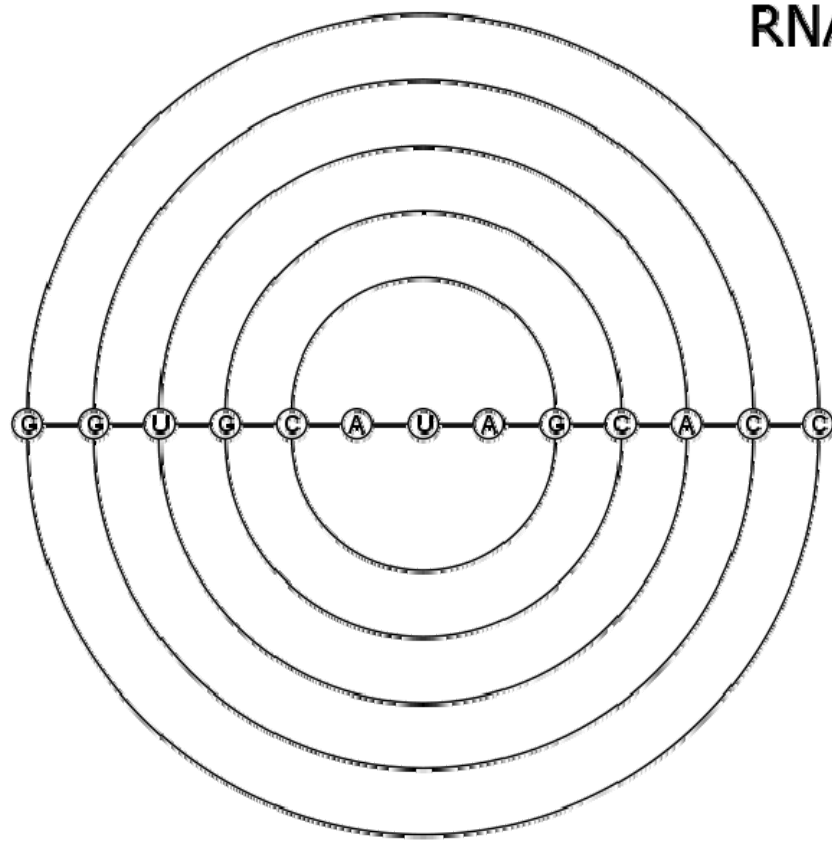   recalculate probabilities
    cool

Convergence

- convenient measure

$$S = \sum_{j=1}^{N_{seq}} \sum_{i \in \text{states}} p_{i,j} \ln p_{i,j}$$
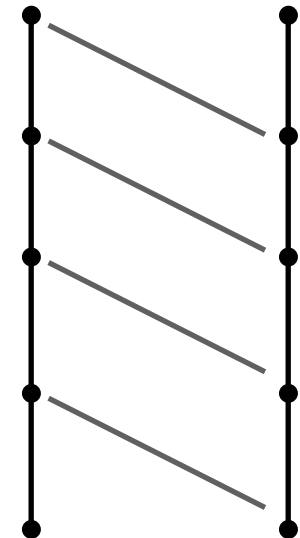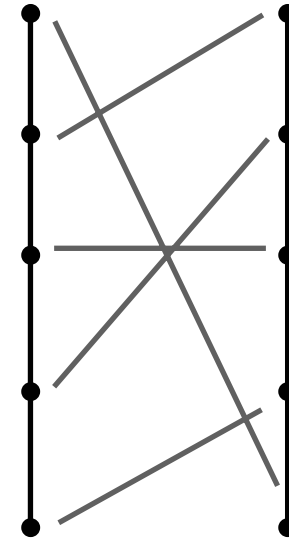
Does this work ?

RNAscmf

PDB
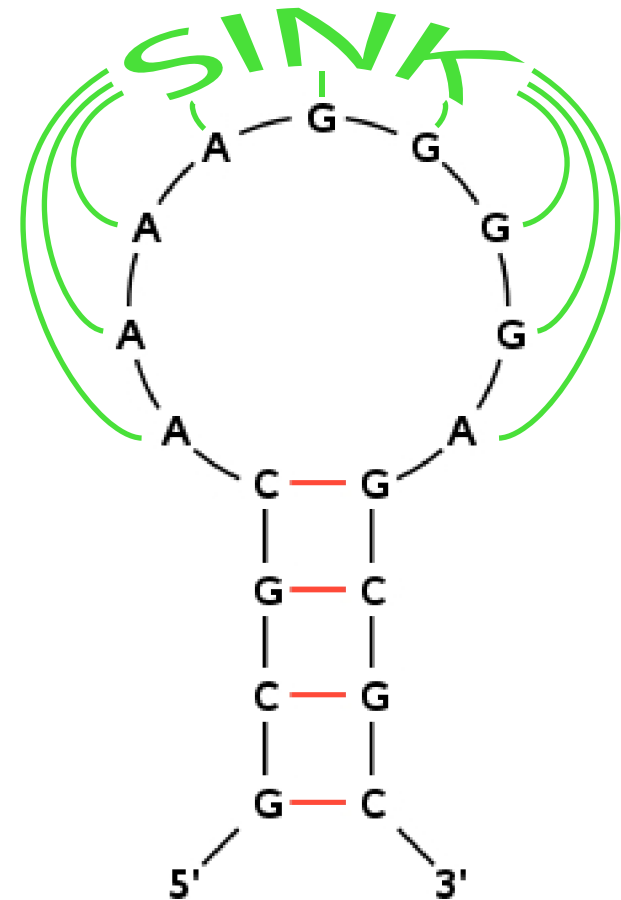
- not impressed ?
- more on energies

Energies and biases

- Basic energies
  - Matthews / Turner scheme
- what will happen naïvely ?

- biases
  - base i has some choices
    - consistent with partner
  - bias
    - neighbour has p associated with a some pair
    - you would like to pair so as to form a helix
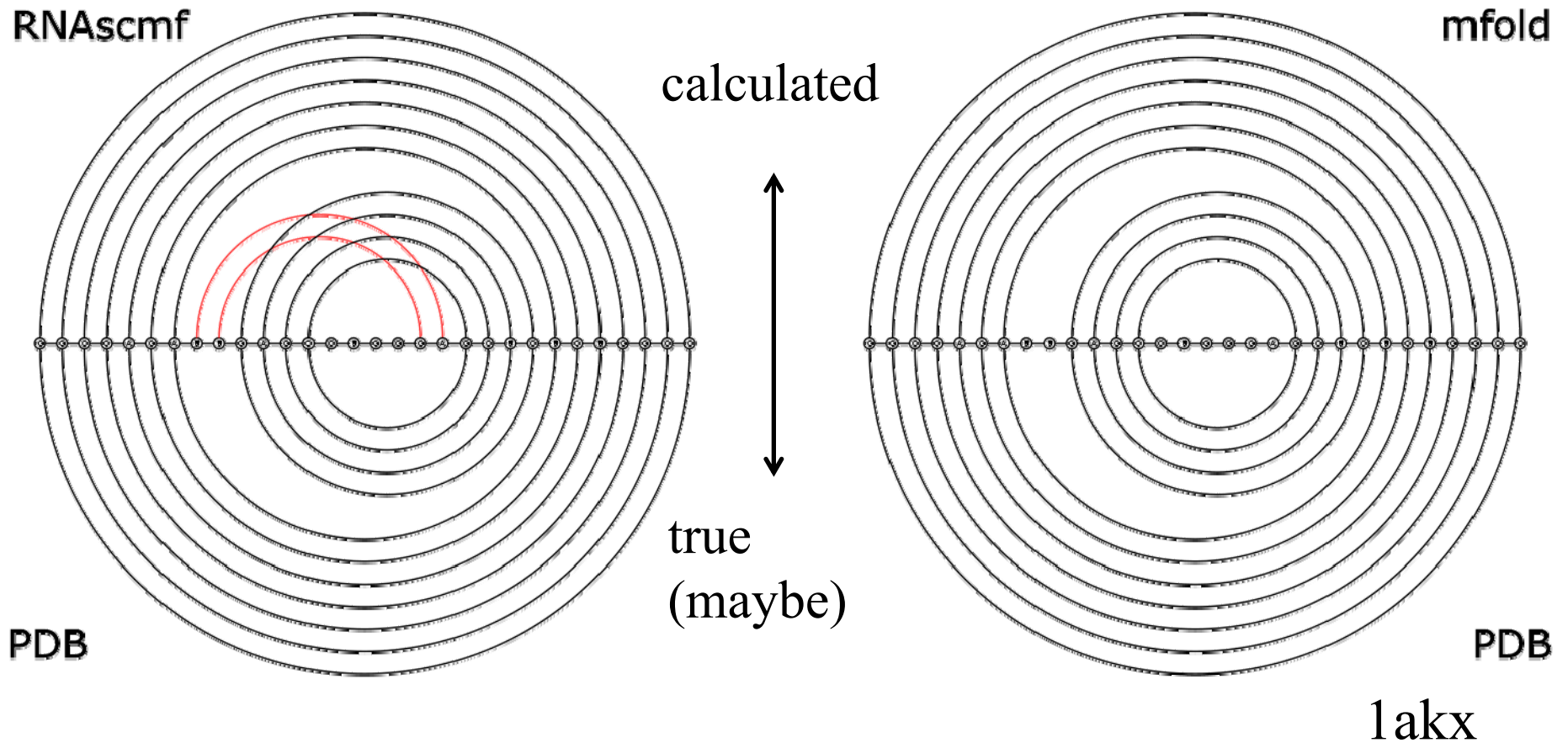- final recipe

Energies and biases

- Literature energies (almost)
- helix bias
- loop bias
- $i \rightarrow j = j \rightarrow i$
- sink
  - energy from not forming a pair

- running time O(mn^2)

- more parameters
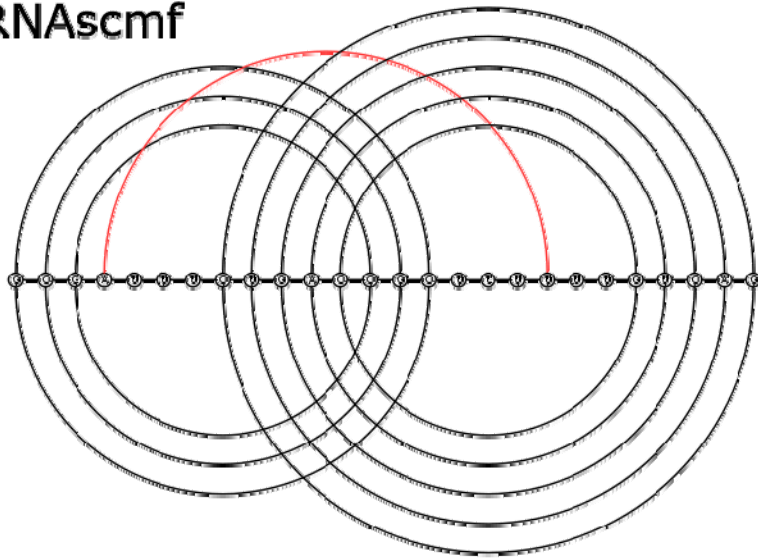 - cooling scheme
  - memory / damping

Easy cases

us ←――――――――――――――――――→ them

RNAscmf   mfold

calculated

true
(maybe)

PDB   PDB

1akx

- red = bad
- do not forget extra pairs
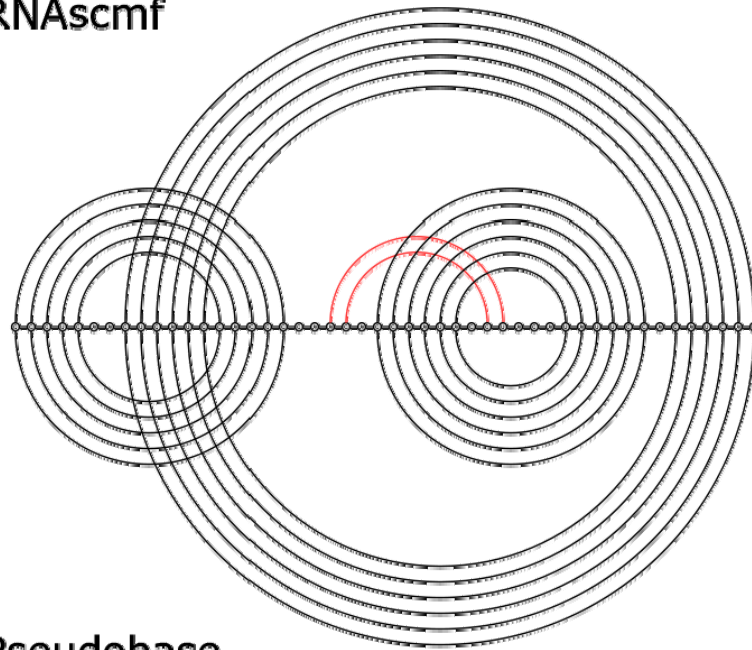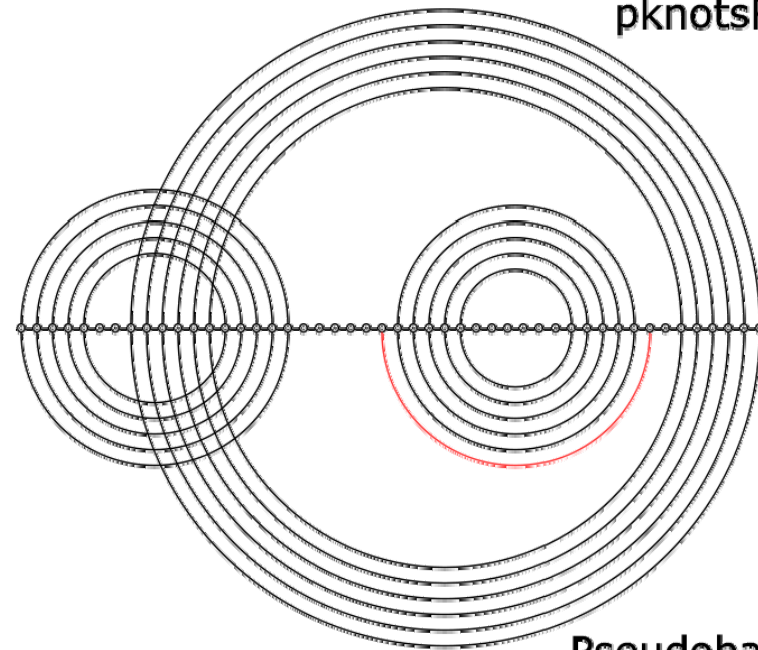
Nasty

pkb81

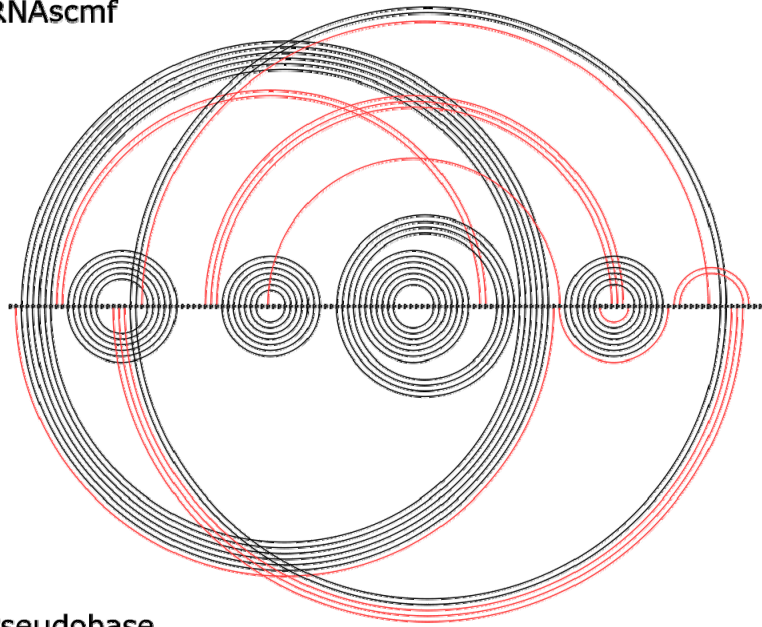- extra base pairs
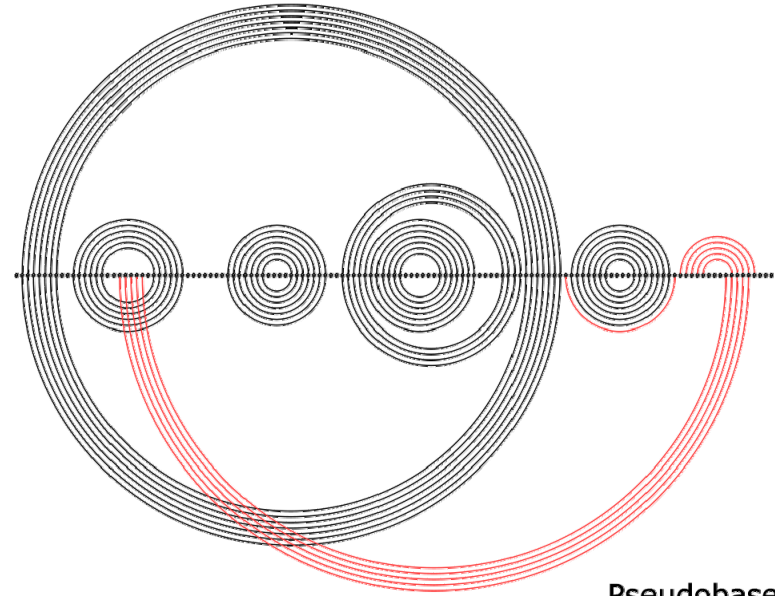- our energy ?
- more fair – do not use mfold

RNAscmf

pknotsRG

Pseudobase

Pseudobase

pkb131

- extra base pairs
- our energy ?
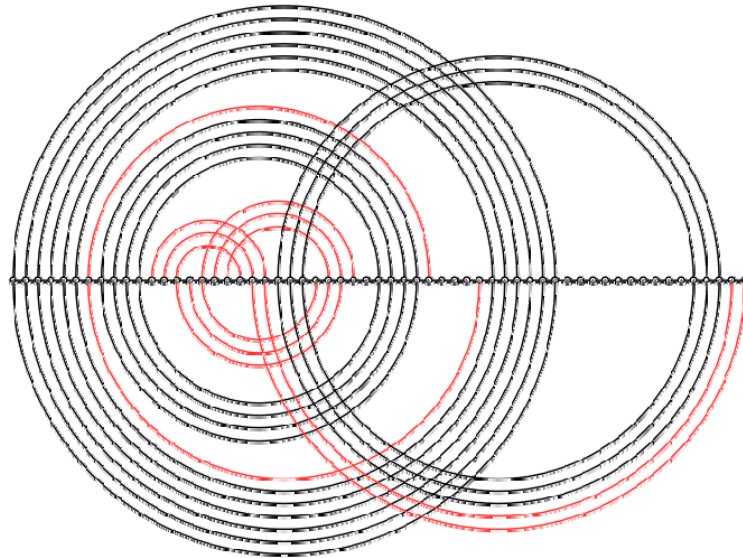- pknotsRG can be totally wrong

- pknotsRG misses some helical regions
- we find too many   u68074
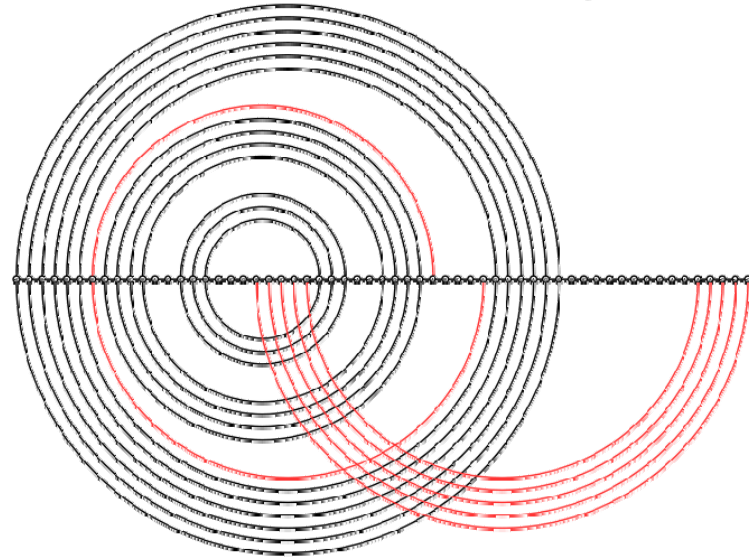
- and the same
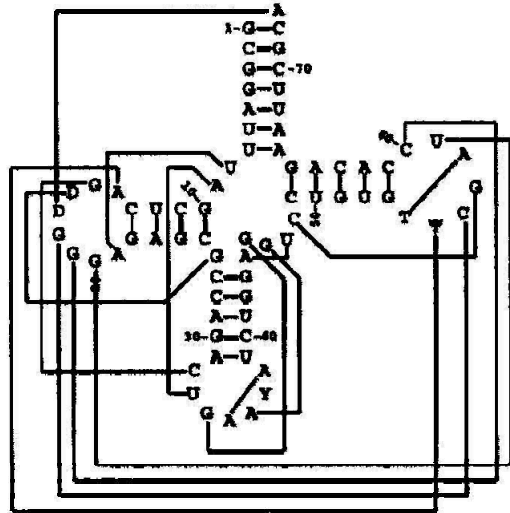
RNAscmf — Pseudobase

pknotsRG — Pseudobase

u67074

- exactly as before

- many more

- what are we doing wrong ?

Too many base pairs
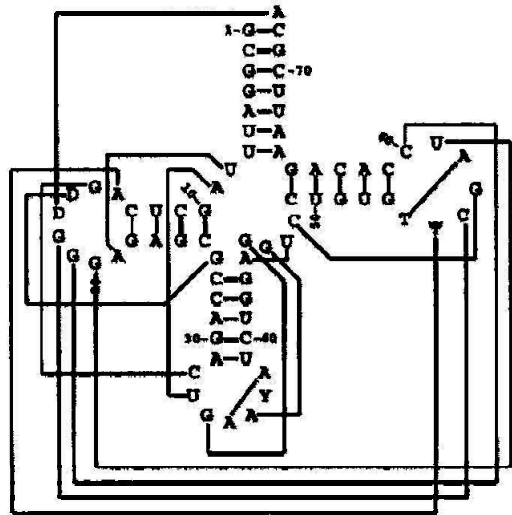
- are we the first to see this ?

- maximal graph matching approach

- given the problem
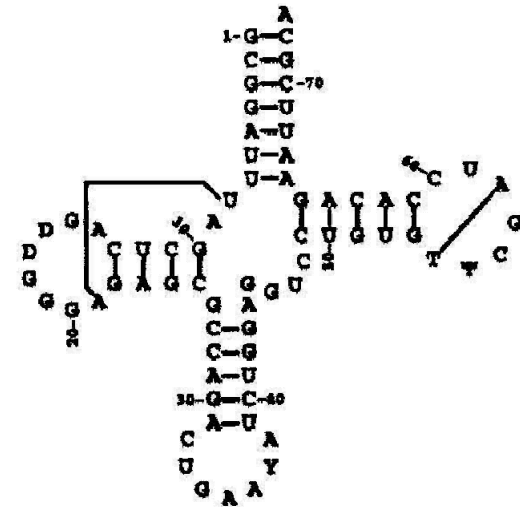 - this is a good solution
- Fixed ?

Tabaska, J.E., Carey, R. B., Gabow, H.N., Stormo, G, N, Bioinformatics, 14, 691-698 (1998)

Too many base pairs

- are we the first to see this ?

 filters →

- elegant ?

Tabaska, J.E., Carey, R. B., Gabow, H.N., Stormo, G, N, Bioinformatics, 14, 691-698 (1998)

PUKE

- are we the first to see this ?

AUSKOTZEN

BARF

filters →

SICH ÜBERGEBEN

"gag me with a spoon"

-Moon Unit Zappa

Tabaska, J.E., Carey, R. B., Gabow, H.N., Stormo, G, N, Bioinformatics, 14, 691-698 (1998)

Broken

- Why so ugly ?
  - generate wrong pairs and clean up
  - If pairs are known to be wrong
    - they should be detected as less likely
    - changes probability trajectory of system
- How broken / ugly are we ?
  - not too ugly
  - bit broken
    - helix, loop and sink terms are built in from start
- Fundamental problem

Problems

- Fundamental
  - we have excellent energies
  - energy model pushed too hard
- Global optima ?   maybe not quite
- Parameters can be tuned for any system
  - can it be done rationally / justifiably ?
- Some details to play with
  - cooling
  - size dependence
- Some arbitrary terms to come
  - not easy
- Where is truth and beauty ? pseudobase ?

- Till next time