Kinwalker

Michael Geis

Bioinformatics Group, Dept. of Computer Science
University of Leipzig

21th TBI Winterseminar in Bled

Questions Folding Kinetics attempts to answer

- What structure is a RNA random coil expected to fold into?
- What states does it pass through in the process?

Given the probabilities of conformations and the in- and outflow probabilities, the change in probability can be determined.
The **master equation** summarizes this.

Let $P_\alpha(t)$ denote the probability of state α at time t, k_+ the rate of inflow and k_- the rate of outflow. Then:

$$\frac{\delta P_\alpha}{\delta t}(t) = \sum_{\beta \neq \alpha} [k_+ P_\beta(t) - k_- P_\alpha(t)]$$

$$k_{+/-} = k_0 e^{\frac{-\Delta G}{KT}}$$

- For $t \rightarrow \infty$, transition rates become constant, yielding the eventual probability of each state.
- Only a few of all possible conformations are of interest. Computing all of them becomes soon too expensive. Kinetic models essentially differ in how they group states into macro-states and their modelling of ΔG for the transition between these.

The underlying assumption

For $t \to \infty$, the result of the folding process and the state predicted by the thermodynamic ensemble agree.

1. Therefore, models based on thermodynamic considerations can be used to predict the final structure.
2. This is only true, if the model is *ergodic*: It must be possible to reach every state from every other state.

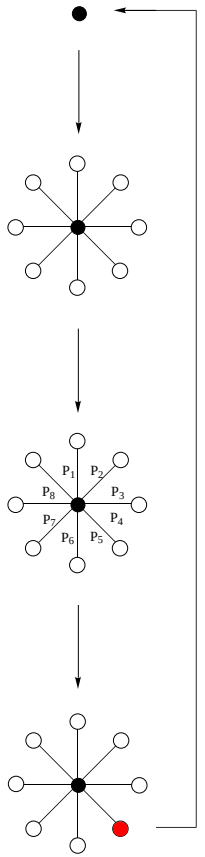
Kinetic Folding Algorithm

Sample folding paths using a Monte-Carlo based method to obtain transition rates.

Assign rates to each move, e.g.

$$P_i = \min\left\{1, \exp\left(-\frac{\Delta E}{kT}\right)\right\}$$

Select a move with probability proportional to its rate.



The Flooding Algorithm

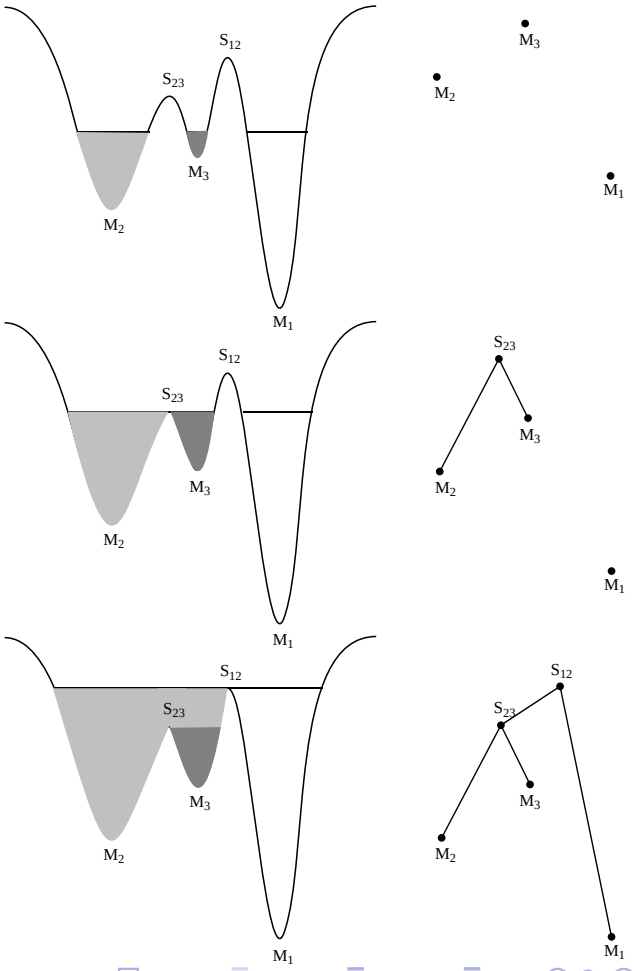
Refer to a set of points in the folding space for which a steepest descent leads to the same minimum as basin. The Flooding Algorithm then divides a landscape into basins.

Start with an energy sorted list of conformations. For each structure x, there are three cases:

1. If no neighbors of x are known, x is a local minimum and starts a new basin.
2. x belongs to a basin B, if all its neighbors belong to it.
3. If x has neigbors in several basins, it merges these basins. The new basin retains the lowest minimum of the original basins.

Grouping States Into Macro-States

- Use the flooding algorithm to find basins (*macro-states*), i.e. sets of states sharing saddle points.
- Within a *macro-state* the transition rates are: $k_{\alpha\beta} = \Gamma_{\alpha\beta} \exp\left(-(E^*_{\alpha\beta} - G_\alpha)/kT\right)$
  G_α : energy of macro-state.
  $E^*_{\alpha\beta}$: energy of transition state.
- Only transition rates between *macro-states* require the master equation.



1. Helix Models relax state resolution to the level of helices. This prunes the state space by the number of conformations with basepairs outside of helices.
2. The difficulty lies in finding a good model for the energy difference between states. Kinwalker uses the Morgan-Higgs heuristic which provides a microscopic resolution of state transitions.

Example:

1. k_+ is comprised only of the entropic component of bending the structure into the shape required by the helix: $\Delta G^F{}_+ = T * \Delta S$
2. k_- is made up of an enthalphic component only, i.e. the energy that is released from opening base pairs: $\Delta G^F{}_- = \Delta H$

Lessons

I. Keep the set of states small.

II. Model state transitions at the level of base pairs.

Implementation

I. Only consider global minima on substrings of the RNA chain. There are less than $\frac{n^2}{2}$ of these.

II. Use the Morgan-Higgs approach for barrier calculations, which determines transition rates at base pair level.

Overview

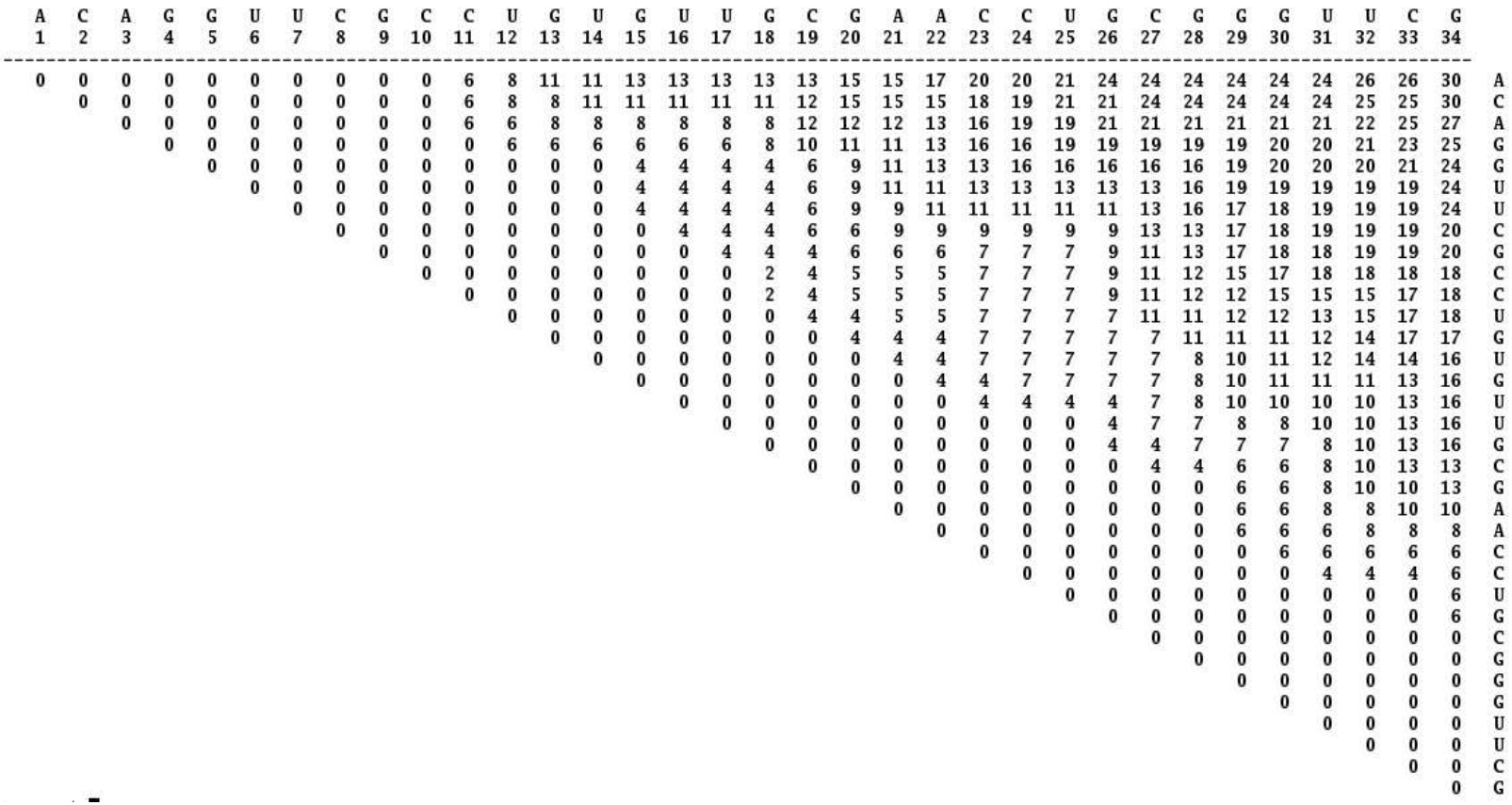
- Find the minimum free energy of all substructures and divide it by the bases in the substructure.
- Populate a matrix with these values and record its local minima.
- Group non-intersecting structures of local minima into an energy front.
- Continously extend the front subject to energy constraints
- Stop when the global mfE structure is reached.

Find optimal substructures

- Calculate the minimum free energies for substructures $[i, j]$, where $1 \leq i \leq j \leq n$.
- Populate an upper triangular matrix $C(i, j)$ with these $\frac{n*(n+1)}{2}$ entries.
- Divide each entry by $j - i + 1$.

This yields the minimum free energy contribution per base for each substructure.

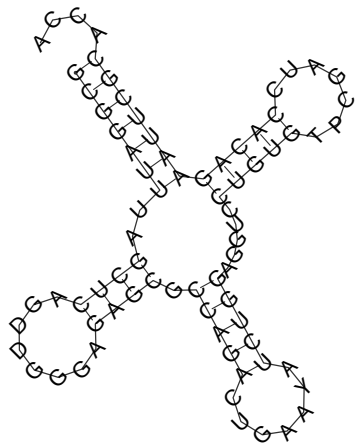
The relative free energy matrix



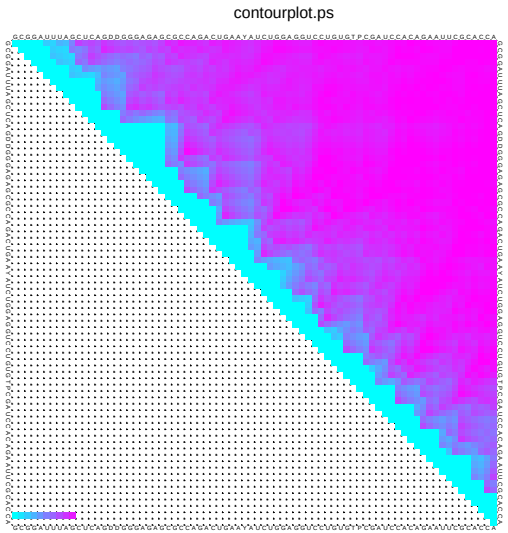
tRNA phe

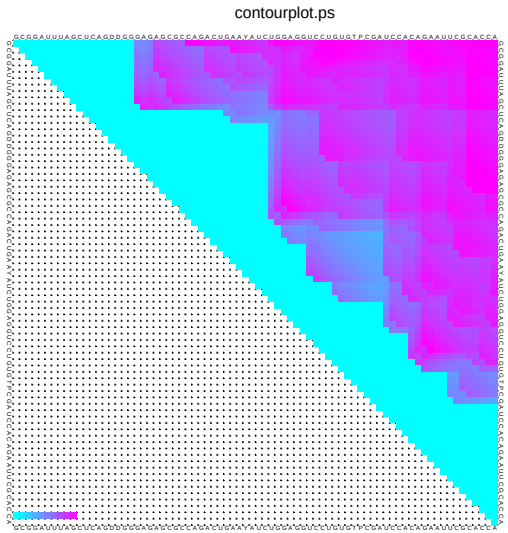
Primary structure:
GCGGAUUUAGCUCAGDDGGGAGAGCGCCAGACUGAAYA
UCUGGAGGUCCUGUGTPCGAUCCACAGAAUUCGCACCA

Secondary structure:
(((((((..((((.....)))).((((.....)))))).....(((((.....)))))))))))....



Contourplot of tRNA phe for stacksize 1 and 3

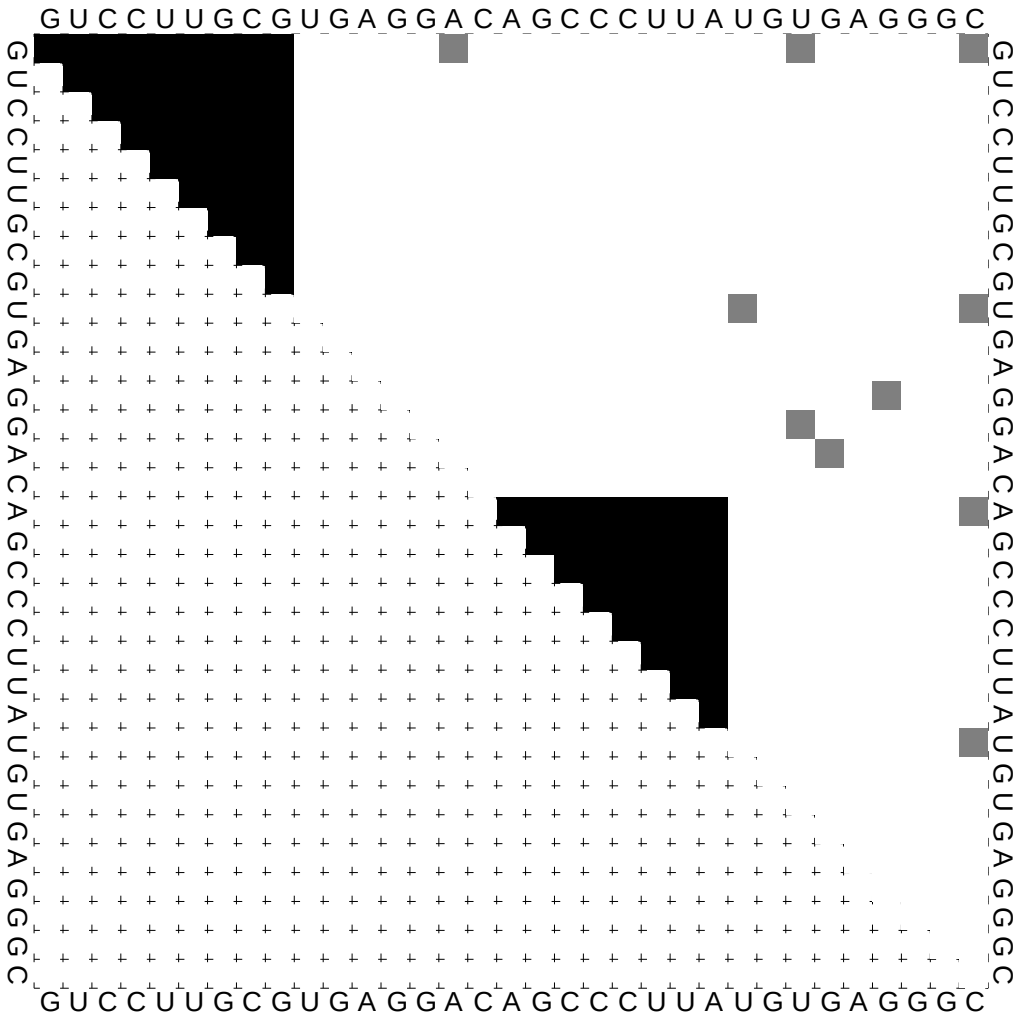




Initial Energy Front

- The state of the structure for a given thermal energy is reflected in the energy front.
- This front is a set of local minima in the relative energy matrix, i.e. disjoint optimal substructures.
- The algorithm progresses through rounds, trying to update the energy front in each turn.
- The order in which minima are tested for admissibility is given by:
  a. Increasing index of the diagonal
  b. Increasing distance from the edge of the matrix
  c. 5' before 3'

frontplot 0.ps

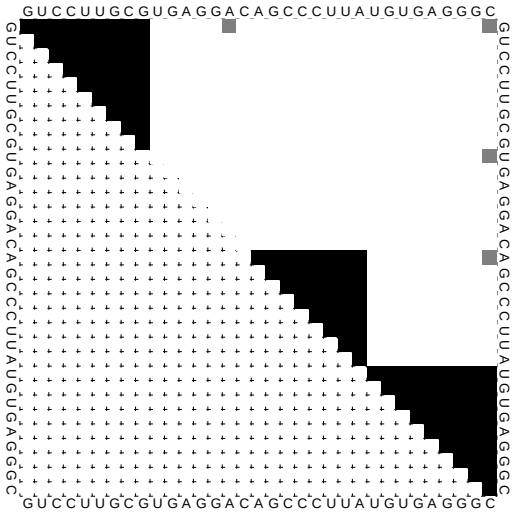


Energy Front Progression

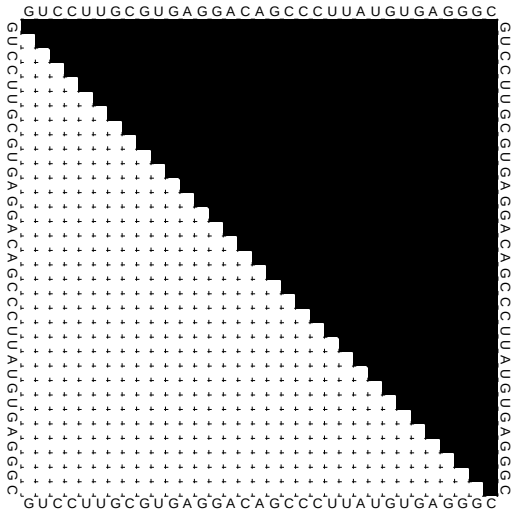
- The algorithm extends the energy front by local minima until the mfE structure is reached.
- In each round the maximal energy barrier between the front and admissible extension points is incremented.
- The Morgan-Higgs heuristic determines the lowest saddle point between a minimum and the energy front.
- The minimum is added to the energy front, if $E(front) - E(saddle) <= energy_barrier$.

Energy Front Progression

frontplot 1.ps



frontplot 2.ps



The Morgan-Higgs heuristic aims at determining the saddle point between two conformations A and B. It only considers direct routes, that is routes that only change contacts in the symmetric difference $A \triangle B$ of A and B.

Algorithm

Let $A_{add} = B \backslash A$ be the contacts to add to get from A to B.

Let $A_{remove} = A \backslash B$ be the contacts to remove to get from A to B.

Sort A_{add} by ascending number of conflicts with A_{remove}.

For each basepair $p \in A_{add}$:

Remove the conflicts of p in A_{remove} from the structure.

Add all elements in A_{add} without conflicts to the structure.

Record the maximum of the old and the new energy.

Comments:

1. The Morgan-Higgs heuristic returns the energy barrier of the lowest traversed path. There is no guarantee that the choice of routes includes the lowest direct route.

2. When there are several base pairs with an equal number of conflicts, paths for each possible ordering must be calculated.

Acknowledgments

Christoph Flamm, Peter Stadler