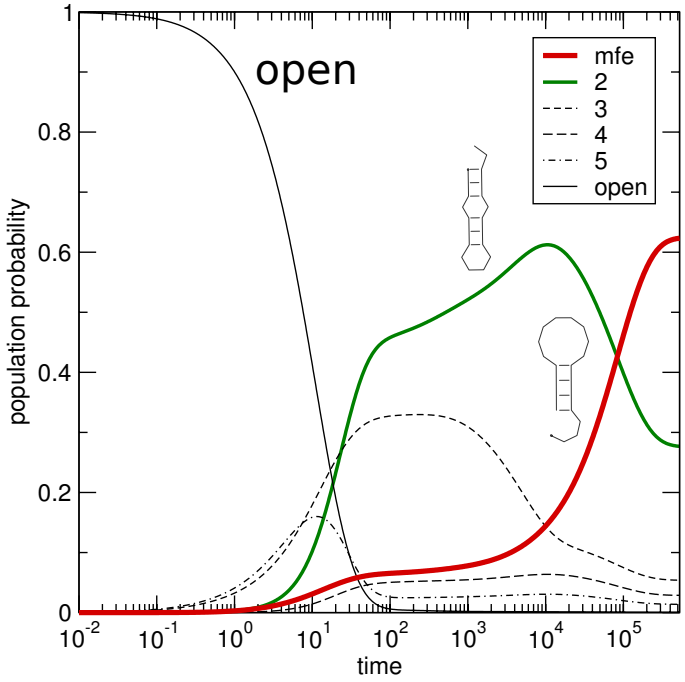
Energy Landscapes: Motivation/Goal

- change of molecule structure over time
- energy driven process

folding process: move through structure-space on energy landscape



Kinetics in contrast to *Thermodynamics*

Energy Landscapes: Idea

- states
- neighbors (of a state)
- energy (of a state)

Energy Landscapes

Definition (Energy Landscape)

An *energy landscape* (EL) consists of

1. a set of states $\mathcal{X}$
2. a notion of neighborhood, nearness, distance, accessibility on $\mathcal{X}$ (relation $\mathcal{N}$)
3. an (energy) function $\mathsf{E} : \mathcal{X} \to \mathbb{R}$.

(That is, it is a triple $(\mathcal{X}, \mathcal{N}, \mathsf{E})$).

Energy Landscapes

Definition (Energy Landscape)

An *energy landscape* (EL) consists of

1. a set of states $\mathcal{X}$
2. a notion of neighborhood, nearness, distance, accessibility on $\mathcal{X}$ (relation $\mathcal{N}$)
3. an (energy) function $\mathsf{E} : \mathcal{X} \to \mathbb{R}$.

(That is, it is a triple $(\mathcal{X}, \mathcal{N}, \mathsf{E})$).

Remarks

- here, states $\mathcal{X}$ are structures
  $\Rightarrow$ for our models of RNA, protein: discrete & finite
- however: continuous $\mathcal{X}$ possible
- physical folding process: energy function, energy minimization
- evolutionary process: fitness function, fitness maximization

EL Examples: RNA

EL of RNA sequence S

1. $\mathcal{X}$ = set of non-crossing RNA structures of S
2. P_1 and P_2 are neighbors ($P_1 \mathcal{N} P_2$) iff
   $P_1 \neq P_2$ and
   $\exists (i,j) : P_1 = P_2 \cup \{(i,j)\}$ or $P_2 = P_1 \cup \{(i,j)\}$
3. $\mathrm{E}(P) = E_S(P)$

similar: HP-proteins; define neighborhood by local moves, pivot moves, ...

Basic Properties: Neighborhood

discrete Neighborhood
defined by *neighbor function* $\mathsf{N} : \mathcal{X} \to \mathcal{P}(\mathcal{X})$

define
$x \in \mathcal{X}$ *has neighbor* y iff $y \in \mathsf{N}(x)$, write $x \mathcal{N} y$

often:
neighbor relation is symmmetric,
i.e. $x \mathcal{N} y$ iff $y \mathcal{N} x$.

Basic Properties: Local Optima

Definition (global minimum)

$\hat{x}$ is a *global minimum* iff

$$\mathsf{E}(\hat{x}) = \min_{y \in States} \mathsf{E}(y).$$

Definition (local minimum)

$\hat{x}$ is a *local minimum* iff

$$\forall y \in \mathsf{N}(\hat{x}) : \mathsf{E}(\hat{x}) \leq \mathsf{E}(y).$$

Note

easy to show: global minima are local minima

Walks and Basins

Definition (Walks, Basin of attraction)

A *walk, or path,* $w \in \mathcal{X}$ is $w = w_1 \dots w_k \in$, s.t. $w_i \mathcal{N} w_{i+1}$ $(1 \leq i < k)$.

A walk is *adaptive* iff $\mathrm{E}(w_i) \geq \mathrm{E}(w_{i+1})$ $(1 \leq i < k)$.

A walk is called *gradient walk* iff $w_{i+1} = \arg\min_{x \in N(w_i)} \mathrm{E}(x)$ $(1 \leq i < k)$.

A *gradient walk of x* is a gradient walk starting in x and ending in a local minimum $\hat{x}$; x is *attracted* by $\hat{x}$.

The *basin (of attraction), or gradient basin* of a local minimum $\hat{x} \in \mathcal{X}$ is the set of all x attracted by $\hat{x}$.

Remarks

- are gradient walks unique?
- Degenerate EL: $\exists x, y \in \mathcal{X} : x \neq y \wedge \mathrm{E}(x) = \mathrm{E}(y)$.
- Assume non-degenerate energy landscape.

Barriers

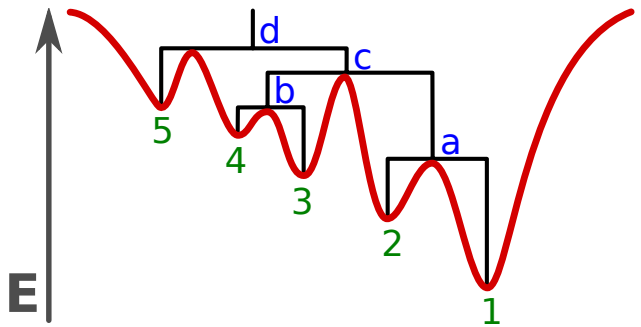
Non-degenerate case: Gradient basins partition the structure space

Definition (Barrier)

The *energy barrier* $\mathrm{E}[x, y]$ *from* x *to* y ($x, y \in \mathcal{X}$) is the minimum energy of a state z on any walk from x to y. z is called *saddle point from* x *to* y.

Remarks:

- $\mathcal{N}$ symmetric $\Longrightarrow$ energy barrier/saddle point symmetric ($\mathrm{E}[x, y] = \mathrm{E}[y, x]$).
- Assume symmetry
- Then, $\mathrm{E}[x, y]$ induces an additive distance on states, in particular local minima.
- $\Longrightarrow$ *barrier tree*, visualizes EL



Move Sets

Move sets define neighborhood of states/structures.

Definition (Move Set)

A *move set for* $\mathcal{X}$ is a function $N : \mathcal{X} \to \mathcal{P}(\mathcal{X})$.

As before: $x \mathcal{N} y$ iff $y \in N(x)$.

Most important properties: symmetry, ergodicity

Definition (Ergodicity)

A move set for $\mathcal{X}$ is *ergodic* iff for all $x, y \in \mathcal{X}$ there is a walk from x to y (with neighborship $\mathcal{N}$).

Equivalent in case of symmetric move set: Fix any state $x_0 \in \mathcal{X}$ (e.g. open chain). Ergodic iff all $x \in \mathcal{X}$ are connected to x_0 (by a walk).

Remark: ergodic $\equiv$ connected

Move Sets for RNA

Fix RNA sequence S. $\mathcal{X}$ is the set of non-crossing RNA structures of S.

- Single Base Pair Moves
  insert or remove a single base pair
- Stem Moves
  insert or remove a stem (set of stacked bp)
- Shift Moves
  move one end of a base pair
  (combine with single base pair moves)

Remarks

- Properties: Symmetry and Ergodicity
- Move Set Hierarchy
- Effect of move set on EL

Move Sets for (Lattice) Proteins

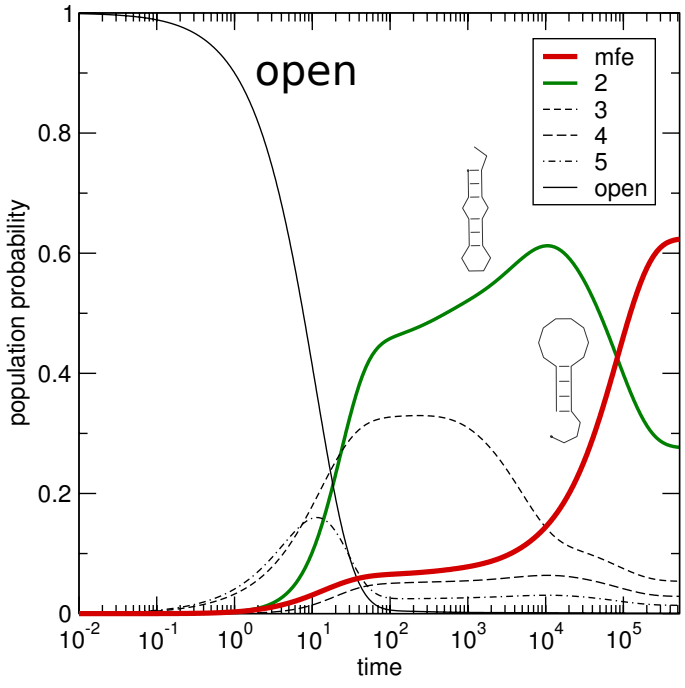
Fix seqeuence S.
Recall: state/structure is a vector $\omega = (\omega_1, \ldots, \omega_n) \in L^n$, ω self-avoiding walk!

- k-Local Moves
  change position of $k' \leq k$ consecutive monomers $i, \ldots, i + k' - 1$ (s.t. result is self-avoiding walk)
- Pivot Moves
  Apply transformation (lattice automorphism) to monomers $1, \ldots, i$ (s.t. result is self-avoiding walk)

Remarks

- Properties: Ergodicity! frozen structures
  k-local moves: not ergodic; pivot-moves: ergodic.
- Effect of move set on EL
- Other ergodic move sets: e.g. Pull moves

Back to our Goal



How do the probabilities of single structures change over time?
(different from "probabilities in equilibrium", cf. McCaskill)

We need a probabilistic model of the folding process.

Stochastic Process

The physical folding process is described as a *stochastic process*.

Define a *random function* X, where X(t) is a *random variable*
$X(t) =$"state at time t".

A physical process has "no history" $\equiv$ *Markov property*

Excursion: (Time-homogenous) Markov Chain

- states $\mathcal{X} = \{1, \ldots, n\}$
- random variables $X_0, X_1, \ldots$
- initial probabilities $\pi_x^0 = Pr[X_0 = x]$
- transition probabilities
  general case, after history $\vec{y} = y_0, \ldots, y_{t-1}$ to x:
  $Pr[X_t = x | X_{t-1} = y_{t-1}, X_{t-2} = y_{t-2}, \ldots]$
  $=_{\textit{no history}} Pr[X_t = x | X_{t-1} = y_{t-1}]$
  $=_{\textit{time-homogenous}} p_{xy}$ "transition to x from y"
  Transition matrix $P = (p_{xy})_{1 \leq x,y \leq n}$

Markov chain models discrete time. Next: continous time

Markov Process

Definition (Continuous-Time Markov Process)

A *(continous-time, time-homogenous, finite state) Markov Process* modeling a random function $X : \mathbb{R} \to \mathcal{X},\ t \mapsto X(t)$ is a triple $(\mathcal{X}, \pi^0, P)$, where

- $\mathcal{X} = \{1, \ldots, n\}$ set of states
- π^0 vector of initial probabilities
- $P(t)$ matrix of probabilities of transitions $p_{xy}(t)$ to x from y in time t

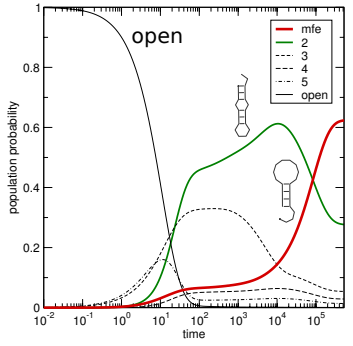
$$P(t) = \begin{pmatrix} p_{11}(t) & \ldots & p_{1n}(t) \\ \vdots & \ddots & \vdots \\ p_{n1}(t) & \ldots & p_{nn}(t) \end{pmatrix}$$

that satisfy the *(strong) Markov property*

$$Pr[X(t+s) = x | X(s) = y] = Pr[X(t) = x | X(0) = y] = p_{xy}(t).$$

Markov Process Allows Studying Folding Behavior

For example, our main goal:



Definition (Probabilities of a state over time)

$$\pi_x(t) := Pr[\text{“State } x \text{ at time } t\text{”}]$$

$$\pi_x(t) = \sum_y \pi_y^0 p_{xy}(t)$$

Yet, we need to construct/define the Markov Process for an EL:
What are the transition probabilities?

Markov Process of an Energy Landscape

EL $(\mathcal{X}, N, \mathrm{E})$

Idea: specify Markov Process

- of the same states $\mathcal{X}$
- by *rates* between neighbored states $x \mathcal{N} y$. Rates tell how fast the system moves from state to state. Rate k_{xy} determined by *energy change* $E(x) - E(y)$.

Review on folding kinetics approaches

Christoph Flamm and Ivo Hofacker. Beyond energy minimization: approaches to the kinetic folding of RNA. Chemical Monthly, 2008.

The Master Equation

Definition (Master Equation)

The *master equation* of a Markov process $(\mathcal{X}, \pi^0, P)$ with state distribution $\pi(t)$ at time t and rate matrix K is

$$\frac{d}{dt}\pi(t) = K\pi(t)$$

Equivalently:

$$\frac{d}{dt}\pi_x(t) = \sum_{y \neq x} \pi_y(t) k_{xy} - \sum_{y \neq x} \pi_x(t) k_{yx}$$

Note: since $\sum_x \pi_x(t) = 1$, $k_{xx} = -\sum_{y \neq x} k_{yx}$.

Properties of Folding Markov process

- Irreducible

$$p_{xy}(t) > 0$$

  for all x,y,t (cf. ergodicity).

- Detailed Balance

$$\pi^*_y k_{xy} = \pi^*_x k_{yx}$$

  for stationary distribution π^*.

- Stationary Distribution = Boltzmann Distribution

$$\pi^*_x = \frac{\exp(-\mathrm{E}_x/(RT))}{Z}$$

  since we want to model the folding process.

Rates of the Folding Process

Detailed balance and stationary distribution leaves much freedom!
Only fixed ratio:

$$k_{xy}/k_{yx} = \pi^*_x/\pi^*_y = \exp(-(\mathrm{E}_x - \mathrm{E}_y)/(RT))$$

Usually defined in the form of *Arrhenius rates* assuming *transition state* $\tau(x, y)$; then, activation energy (from y to x): $\mathrm{E}_{\tau(x,y)} - \mathrm{E}_y$

$$k_{xy} := \gamma \exp(-(\mathrm{E}_{\tau(x,y)} - \mathrm{E}_y)/(RT))$$

Metropolis rates $[\mathrm{E}_{\tau(x,y)} = \max(E_x, E_y)]$

$$k_{xy} := \gamma \begin{cases} 1 & \text{if } \mathrm{E}_x \leq \mathrm{E}_y \\ \exp(-(\mathrm{E}_x - \mathrm{E}_y)/(RT)) & \text{otherwise} \end{cases}$$
$$= \gamma \min\{1, \exp(-(\mathrm{E}_x - \mathrm{E}_y)/(RT))\}$$

Kawasaki rates $[\mathrm{E}_{\tau(x,y)} = \frac{1}{2}(E_x + E_y)]$

$$k_{xy} := \gamma \exp(-(\mathrm{E}_x - \mathrm{E}_y)/(2RT))$$

Example Markov Process for RNA

- Energy Landscape $(\mathcal{X}, N, \mathrm{E})$
  - $\mathcal{X}$ non-crossing RNA structures
  - N simple base pair moves
  - E loop-based free energy
- Markov process $(\mathcal{X}, \pi^0, P)$
  - $\pi^0_x = \begin{cases} 1 & x = \text{open chain} \\ 0 & \text{otherwise} \end{cases}$
  - P specified by rate matrix K

$$k_{xy} = \gamma \min\{1, \exp(-(\mathrm{E}_x - \mathrm{E}_y)/(RT))\}$$

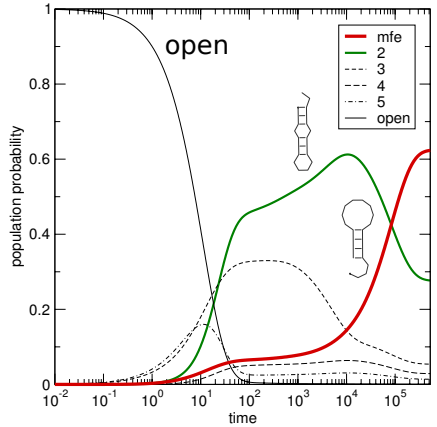
Determine $\pi(t)$

- Solve master equation $\frac{d}{dt}\pi(t) = K\pi(t)$
  - numerical solution, after solving the differential equation

$$\pi(t) = exp(Kt)\pi^0$$

for example solve by diagonalizing K: $K = UDU^{-1}$ and D diagonal, then $\exp(Kt) = U\exp(Dt)U^{-1}$ [exponential of diagonal matrix: element-wise]
$\Longrightarrow$ only small systems (several thousand states)

for example, xbix=`CUGCGGCUUUGGCUCUAGCC`, 20 nucleotides, 3886 structures



- usually too expensive $\Longrightarrow$ Simulation, Coarse Graining, ...

Monte Carlo Simulation with Metropolis Criterion (Rejection-based)

- $x =$ initial conformation (random according to π^0)
- for $t = 1$ to $t_{\max}$ do
  - choose move $x \to x'$ with probability $\mathcal{A}(x \to x')$
  - accept with probability $\mathcal{P}(x \to x')$: $x = x'$

Remarks

- transition probability $x \to x'$ is

$$\mathcal{A}(x \to x')\mathcal{P}(x \to x')$$

- Metropolis criterion:

$$\mathcal{P}(x \to x') = \min(1, exp(-(E_{x'} - E_x)/(RT)))$$

- In general: no detailed balance! $\implies$ this does not simulate the folding process

Rejection-less Monte Carlo Simulation

- $x =$ initial conformation (random according to π^0)
- $t = 0$
- for $i = 1$ to $i_{\max}$ do
  - evaluate all possible moves from x and compute "rate out of x"

$$\kappa_x := \sum_{\text{move } x \to x''} k_{x''x}$$

  - choose move $x \to x'$ with probability

$$\mathcal{P}(x \to x') = k_{x'x}/\kappa_x$$

  - accept always: $x = x'$
  - sample "waiting time" Δt from exponential distribution with average rate κ_x
  - increment time: $t = t + \Delta t$

Remarks

detailed balance due to time correction; correctly models folding process; a.k.a. Gillespie-algorithm or Boltz-Kalos-Liebowitz method; simulation still slow (average thousands of trajectories); for example, simulation tool `kinfold` (C. Flamm)

Coarse Grained Processes

- General idea: define *macro states* and *macro state process*
- For example, macro states = basins of attraction
- energy of macro state α: ensemble energy

$$Z_\alpha = \sum_{x \in \alpha} exp(-E_x/(RT)); E_\alpha = -RT \ln Z_\alpha$$

- macro rates (from macro state β to α): Arrhenius rates
  - energy of transition state (ensemble)

$$Z_{\alpha\beta} := \sum_{x \in \alpha, y \in \beta, \text{move } y \to x} exp(-\mathrm{E}_{\tau(x,y)}/(RT))$$

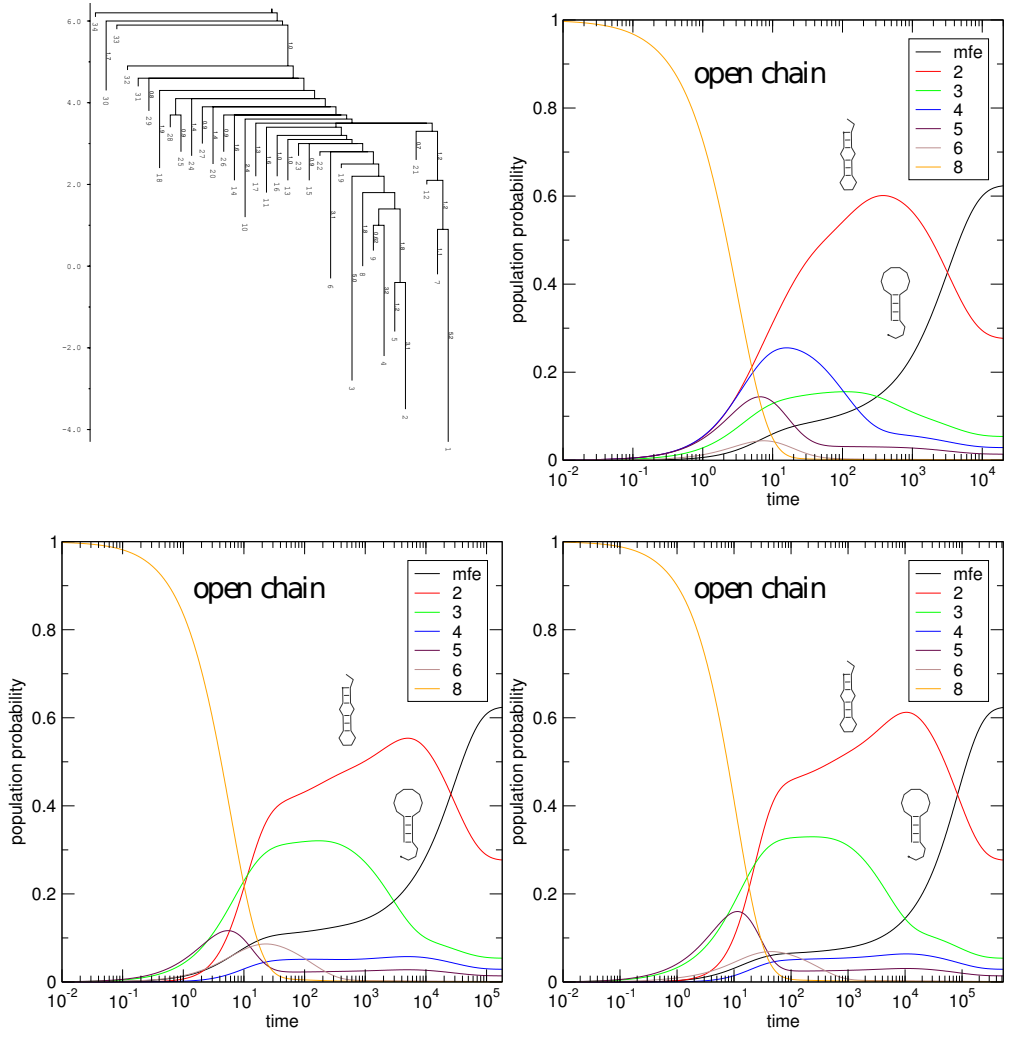
$$E_{\tau(\alpha,\beta)} = -RT \ln Z_{\alpha\beta}$$

  - transition rate

$$k_{\alpha\beta} := \gamma \exp(-(E_{\tau(\alpha,\beta)} - E_\beta)/(RT)) = \gamma Z_{\alpha\beta}/Z_\beta$$

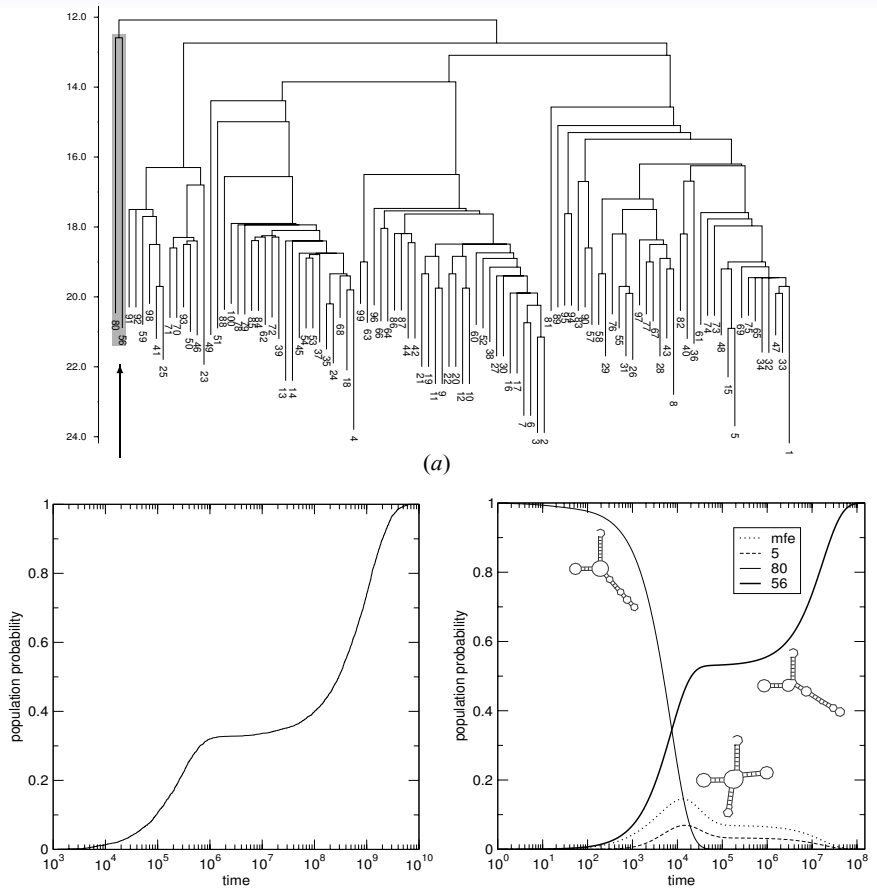
Equivalently, $k_{\alpha\beta} = \gamma \sum_{x \in \alpha, y \in \beta} k_{xy} \Pr[y \mid \beta]$,
since $Z_{\alpha\beta}/Z_\beta = \sum_{x \in \alpha, y \in \beta, y \to x} \exp(-\mathrm{E}_{\tau(x,y)}/(RT))/Z_\beta =$
$\sum_{x \in \alpha, y \in \beta, y \to x} \exp(-(\mathrm{E}_{\tau(x,y)} - \mathrm{E}_y)/(RT))) \exp(-\mathrm{E}_y)/Z_\beta$

Dynamics of RNA xbix



barrier tree; process of local minima via saddle point energies; macro state process; full process

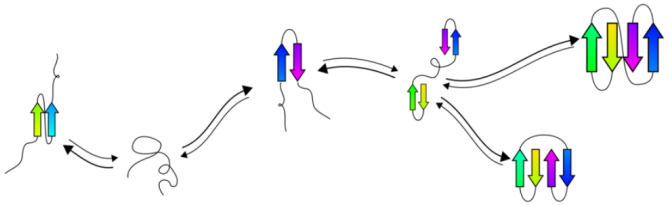
Dynamics of a tRNA



barrier tree; `kinfold` simulation;
macro state process (absorbing state 56)

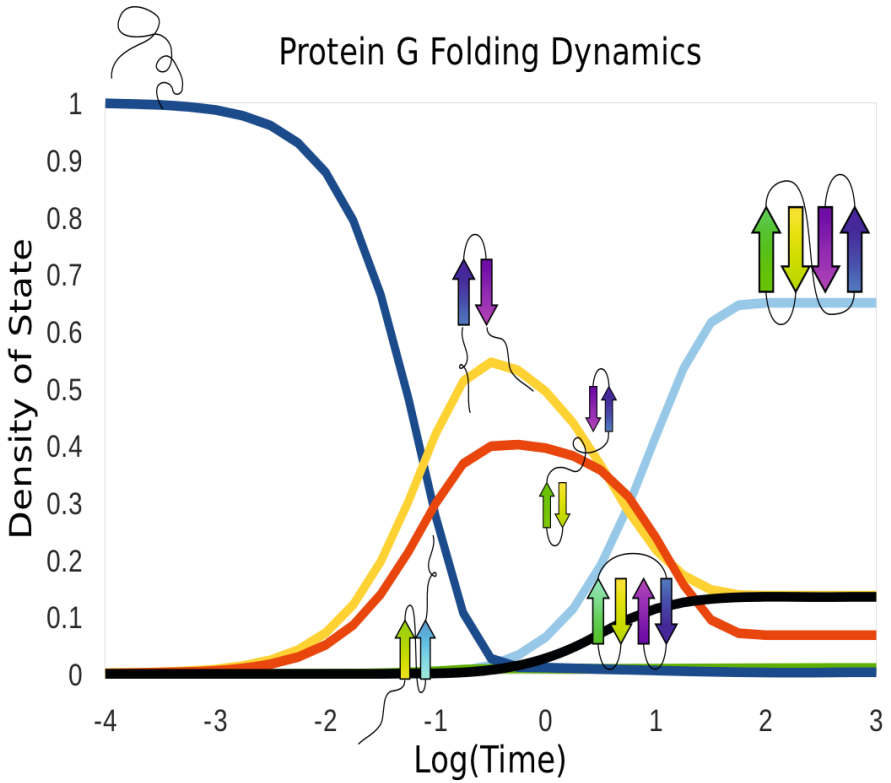
Wolfinger, Svrcek-Seiler, Flamm, Hofacker, Stadler Efficient computation of RNA folding dynamics. J.Phys. A, 2004

Dynamics of β-sheet proteins: tFolder



http://csb.cs.mcgill.ca/tfolder/

Coarse graining:
macro state =
sub-ensemble of a specific β-strand interaction



Shenker, O'Donnell, Devadas, Berger, Waldispühl. Efficient traversal of protein folding pathways using ensemble models. RECOMB 2011.