Energy Landscapes of Biopolymers

Michael Wolfinger

Institute for Theoretical Chemistry
and Structural Biology,
University Vienna

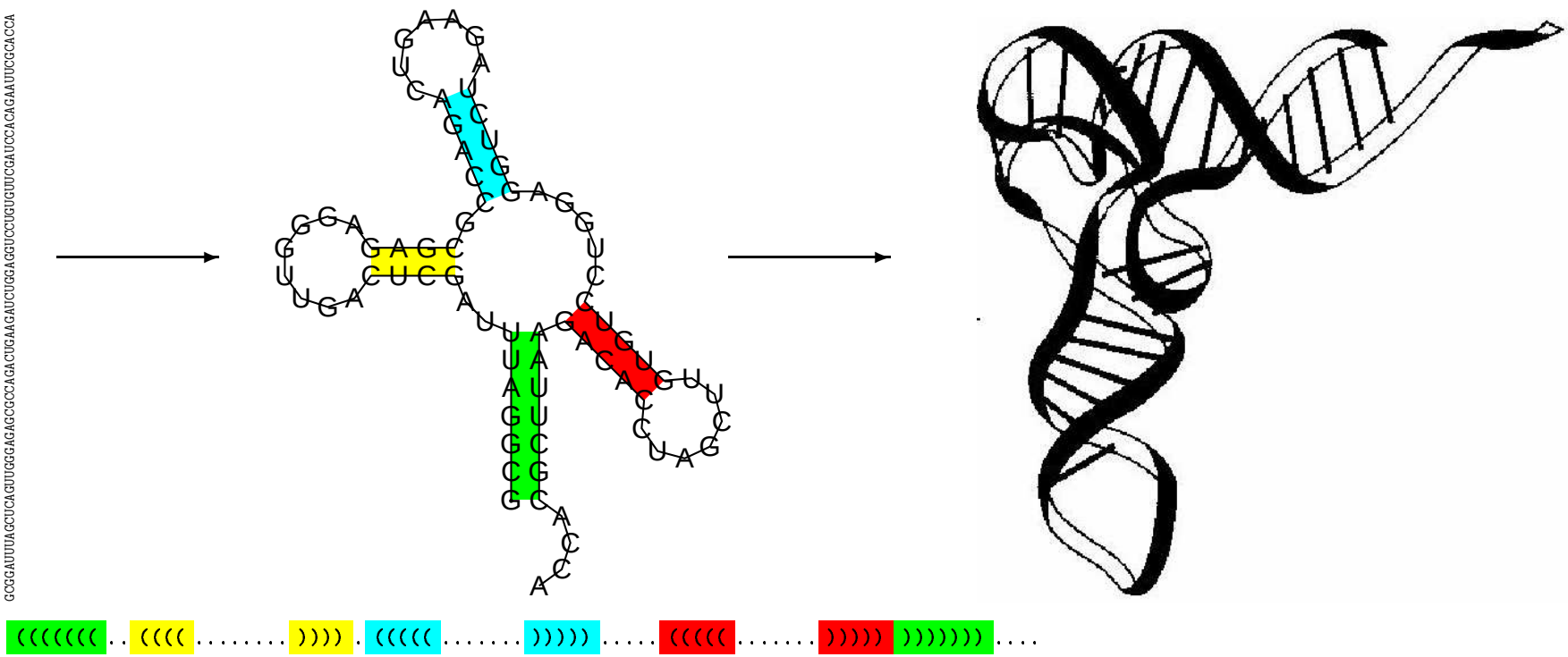
28th October 2004

tbi

Outline

tbi

The RNA model



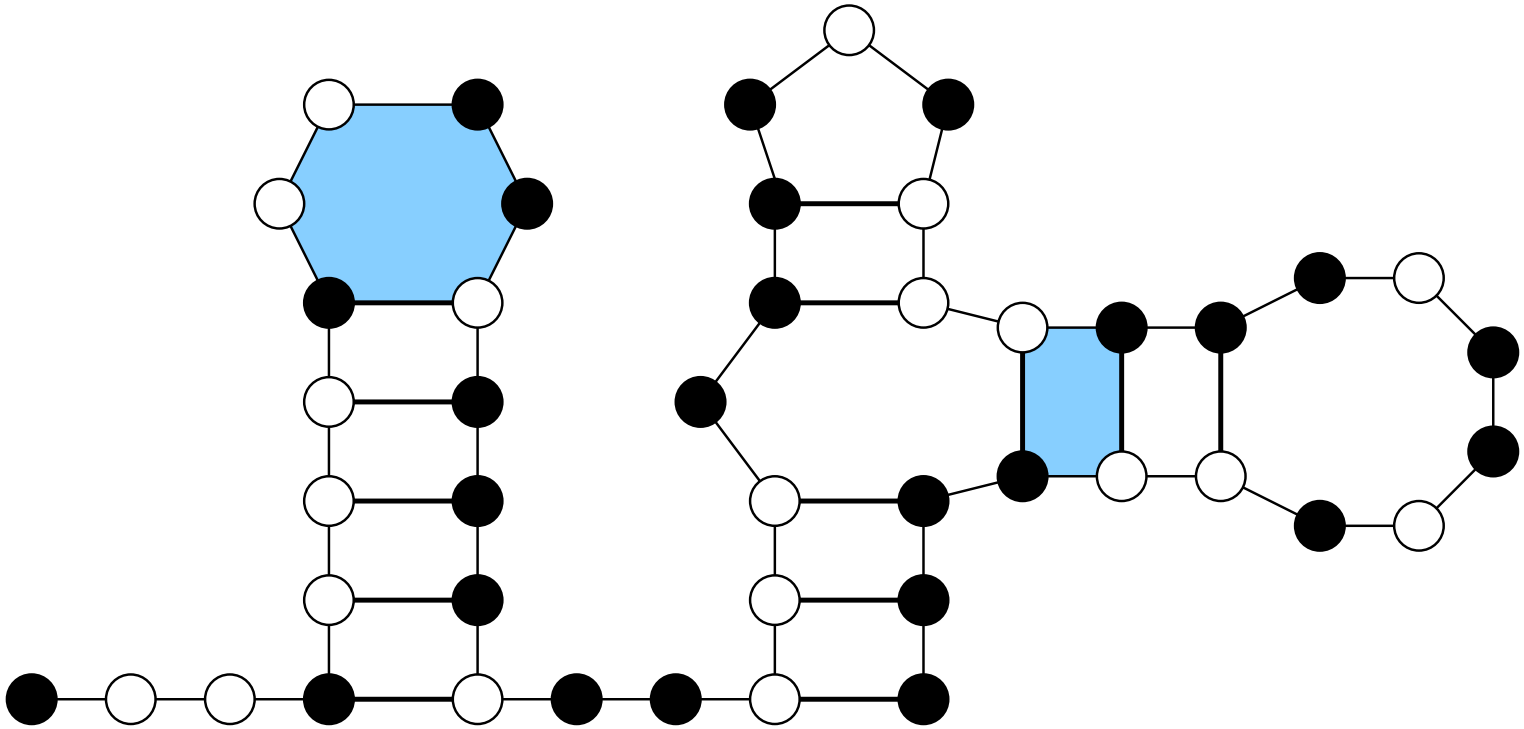
A secondary structure is a list of base pairs that fulfills two constraints:

- A base may participate in at most one base pair.
- Base pairs must not cross, i.e., no two pairs (i,j) and (k,l) may have $i < k < j < l$.
  (no pseudo-knots)

The optimal as well as the suboptimal structures can be computed recursively.

tbi

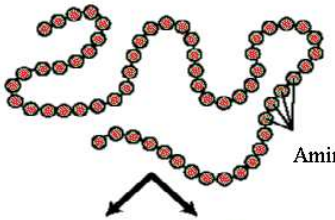
RNA energy model



The energy of a sequence and particular structure is given as the sum of contributions from the "loops" (planar faces). Stacks yield stabilizing contributions, all other loops lead to destabilizing energy contributions.

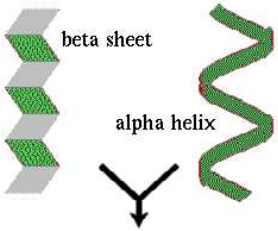
tbi

Levels of structure in proteins

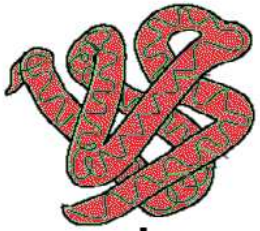


Primary protein structure
is sequence of amino acids

Amino Acids

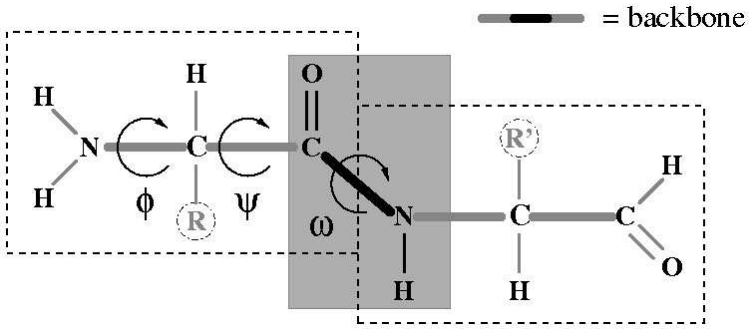


Secondary protein structure
local conformation
primarily stabilized by
hydrogen bonding



Tertiary protein structure
three dimensional conformation

Quaternary protein structure
Combination of multiple
polypeptide chains

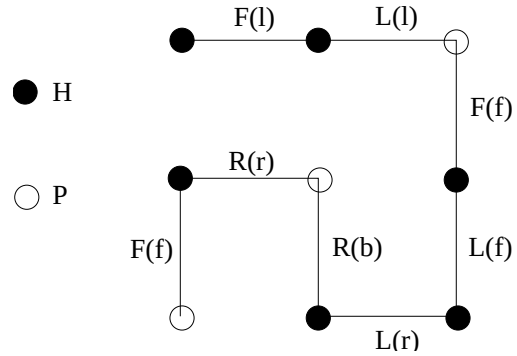


The HP-model

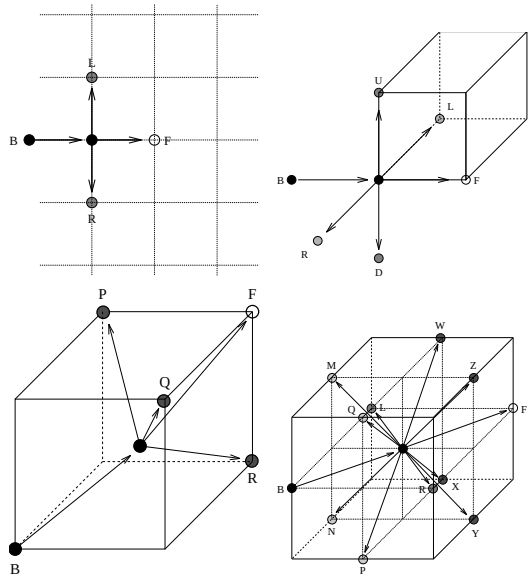
Suggested by Dill, Chan and Lau in the late 1980ies. In this *simplified model*, a conformation is a *self-avoiding walk (SAW)* on a given lattice in 2 or 3 dimensions. Each bond is a straight line, bond angles have a few discrete values. The 20 letter alphabet of amino acids (monomers) is reduced to a two letter alphabet, namely **H** and **P**. H represents hydrophobic monomers, P represents hydrophilic or *polar* monomers.

Advantages:

- lattice-independent folding algorithms
- simple energy function
- hydrophobicity can be reasonably modeled



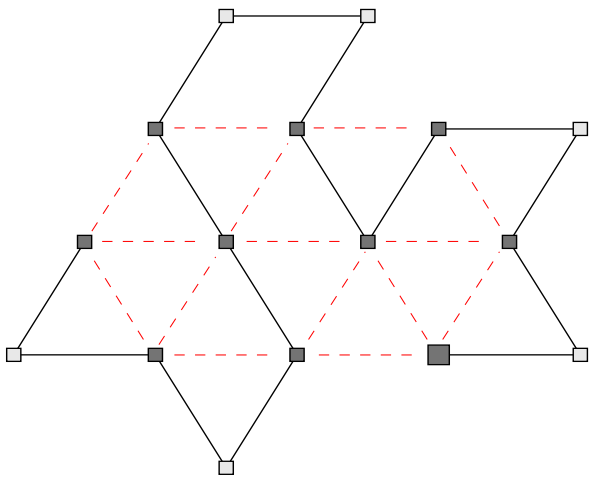
FRRLLFLF



tbi

Lattice proteins

HPHPHHHPPHHHPHPH $n = 16$



15 contacts

	H	P
H	-1	0
P	0	0

	H	P	N	X
H	-4	0	0	0
P	0	1	-1	0
N	0	-1	1	0
X	0	0	0	0

tbi

Folding kinetics

Biomolecules may have kinetic traps which prevent them from reaching equilibrium within the lifetime of the molecule. Long molecules are often trapped in such meta-stable states during transcription.

Possible solutions are

- Stochastic folding simulations (predict folding pathways)
- Predicting structures for growing fragments of the sequence
- Analysis of the energy landscape based on complete suboptimal folding

Folding landscape - energy landscape

The energy landscape of a biopolymer molecule is a complex surface of the (free) energy versus the conformational degrees of freedom.

RNA

$$c_n \sim \alpha^n \cdot n^{-\frac{3}{2}}$$

dynamic programming algorithms available

Lattice proteins

$$c_n \sim \mu^n \cdot n^{\gamma-1}$$

problem is NP-hard

Formally, three things are needed to construct an energy landscape:

- A set X of configurations
- a notion $\mathfrak{M}$ of neighborhood, nearness, distance or accessibility on X, and
- an energy function $f : X \to \mathbf{R}$

tbi

Folding landscape - energy landscape

The energy landscape of a biopolymer molecule is a complex surface of the (free) energy versus the conformational degrees of freedom.

RNA

$$c_n \sim \alpha^n \cdot n^{-\frac{3}{2}}$$

dynamic programming algorithms available

Lattice proteins

$$c_n \sim \mu^n \cdot n^{\gamma-1}$$

problem is NP-hard

Formally, three things are needed to construct an energy landscape:

- A set X of configurations
- a notion $\mathfrak{M}$ of neighborhood, nearness, distance or accessibility on X, and
- an energy function $f : X \to \mathbf{R}$

tbi

Folding landscape - energy landscape

The energy landscape of a biopolymer molecule is a complex surface of the (free) energy versus the conformational degrees of freedom.

RNA

$$c_n \sim \alpha^n \cdot n^{-\frac{3}{2}}$$

dynamic programming algorithms available

Lattice proteins

$$c_n \sim \mu^n \cdot n^{\gamma-1}$$

problem is NP-hard

Formally, three things are needed to construct an energy landscape:

- A set X of configurations
- a notion $\mathfrak{M}$ of neighborhood, nearness, distance or accessibility on X, and
- an energy function $f : X \to \mathbf{R}$

tbi

Folding landscape - energy landscape

The energy landscape of a biopolymer molecule is a complex surface of the (free) energy versus the conformational degrees of freedom.

RNA

$$c_n \sim \alpha^n \cdot n^{-\frac{3}{2}}$$

dynamic programming algorithms available

Lattice proteins

$$c_n \sim \mu^n \cdot n^{\gamma-1}$$

problem is NP-hard

Formally, three things are needed to construct an energy landscape:

- A set X of configurations
- a notion $\mathfrak{M}$ of neighborhood, nearness, distance or accessibility on X, and
- an energy function $f : X \to \mathbf{R}$

tbi

Folding landscape - energy landscape

The energy landscape of a biopolymer molecule is a complex surface of the (free) energy versus the conformational degrees of freedom.

RNA

$$c_n \sim \alpha^n \cdot n^{-\frac{3}{2}}$$

dynamic programming algorithms available

Lattice proteins

$$c_n \sim \mu^n \cdot n^{\gamma-1}$$

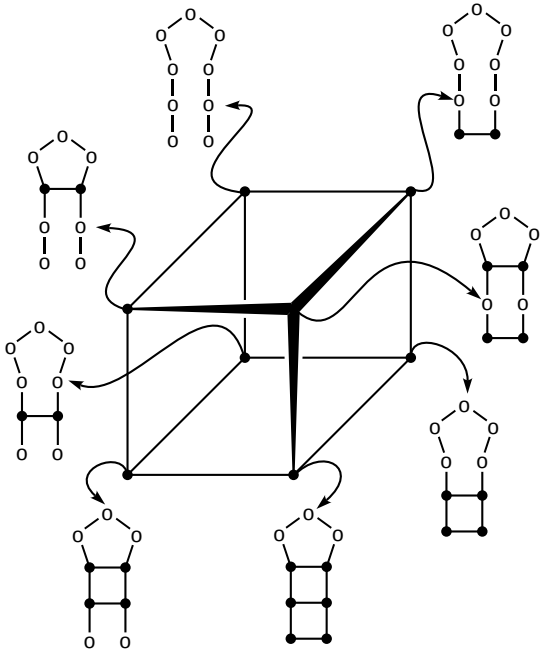
problem is NP-hard

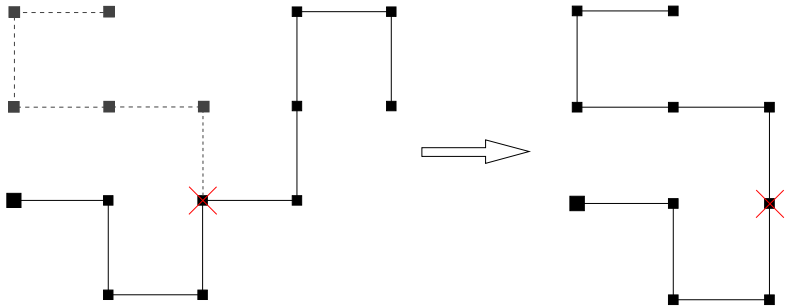
Formally, three things are needed to construct an energy landscape:

- A set X of configurations
- a notion $\mathfrak{M}$ of neighborhood, nearness, distance or accessibility on X, and
- an energy function $f : X \to \mathbf{R}$

tbi

The move set

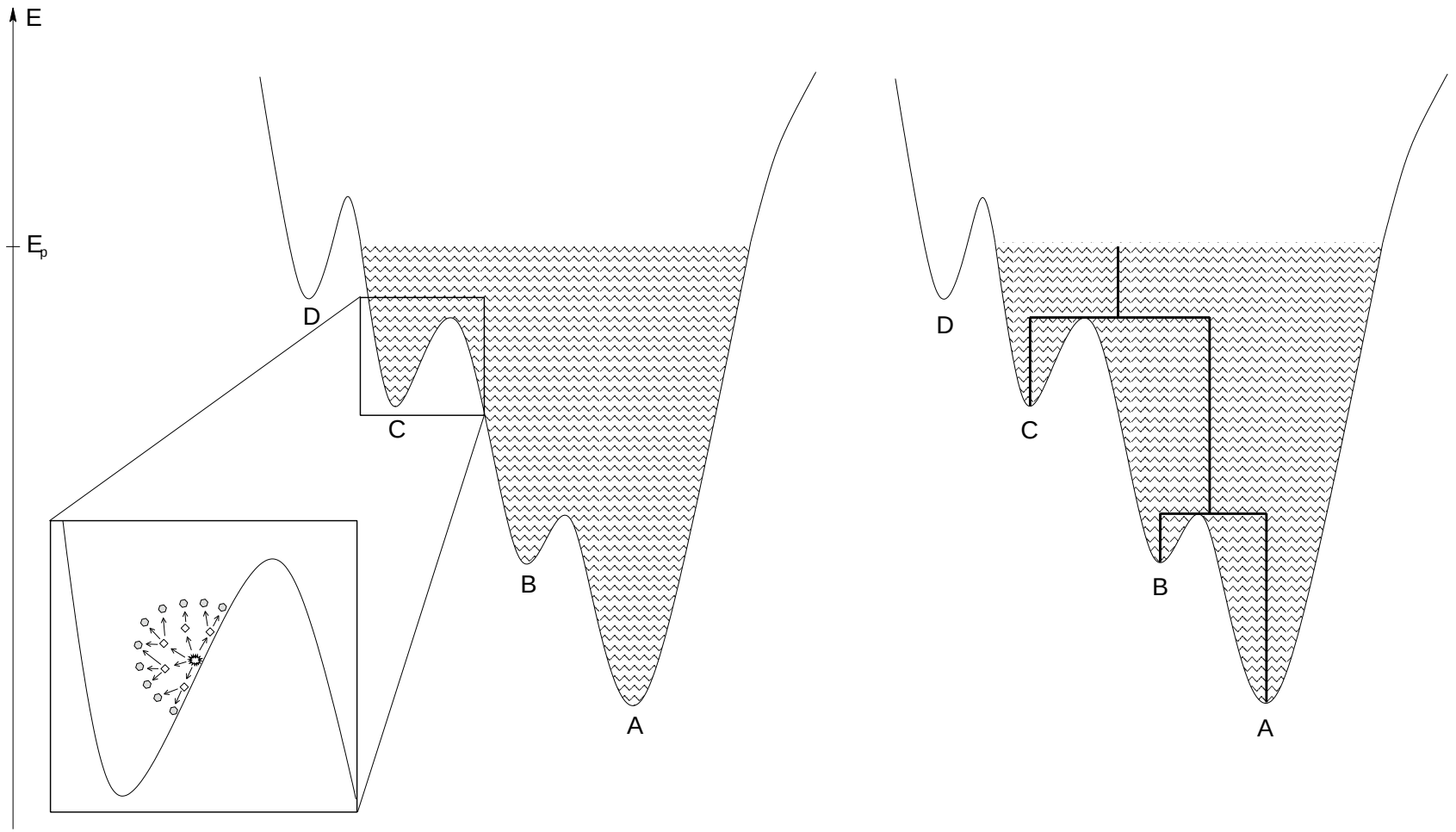




- For each move there must be an inverse move
- Resulting structure must be in X
- Move set must be *ergodic*

tbi

Low-energy states of lattice proteins



Kinetic Folding Algorithm

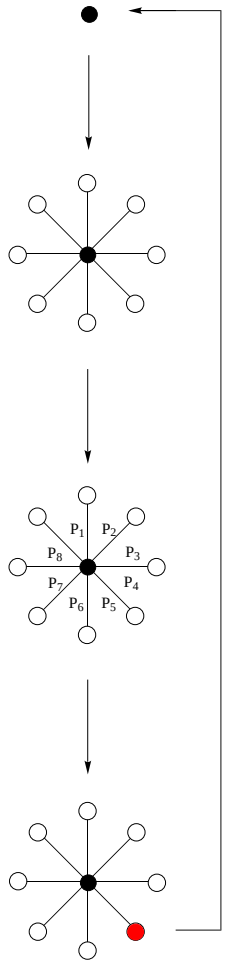
Simulate folding kinetics by a rejection-less Monte-Carlo type algorithm:

Generate all neighbors using the move-set

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
Advance clock $1/\sum_i P_i$.

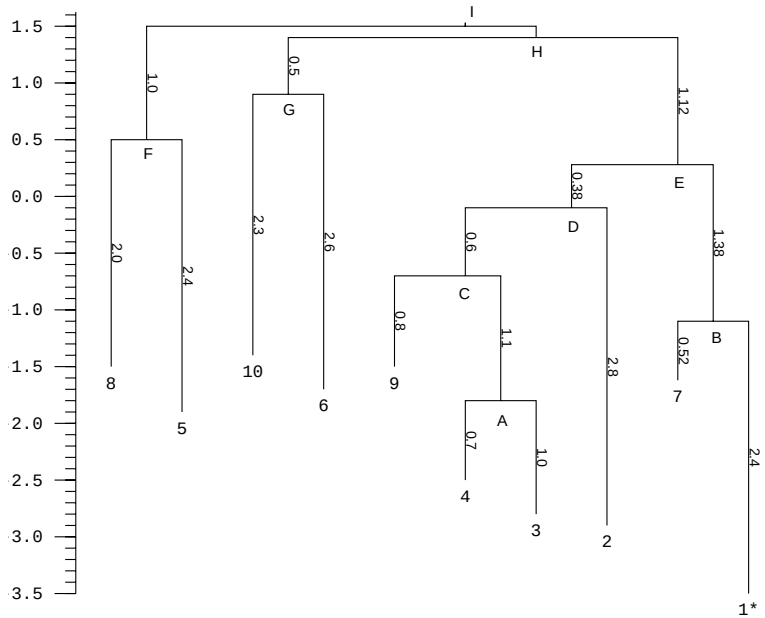


tbi

Energy barriers and barrier trees

Some topological definitions:
A structure is a

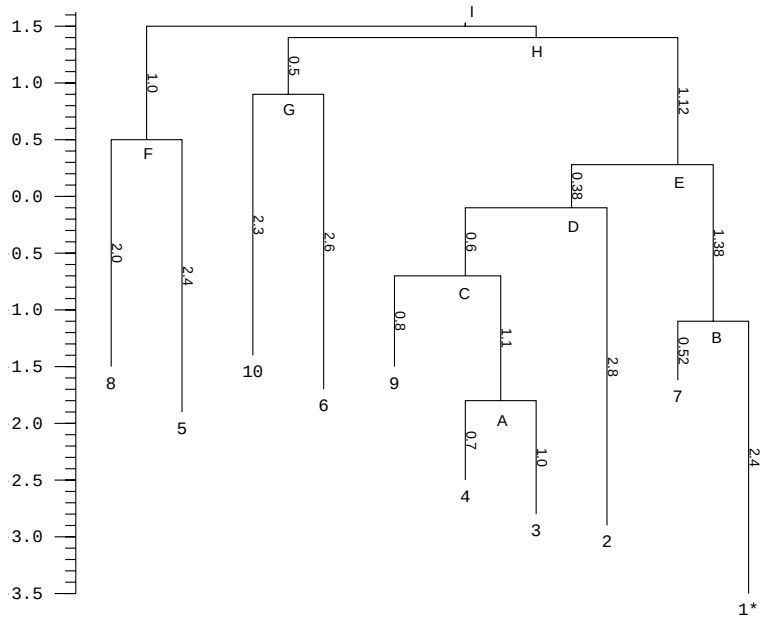
- local minimum if its energy is lower than the energy of **all** neighbors
- local maximum if its energy is higher than the energy of **all** neighbors
- saddle point if there are at least two local minima thar can be reached by a downhill walk starting at this point



Energy barriers and barrier trees

Some topological definitions:
A structure is a

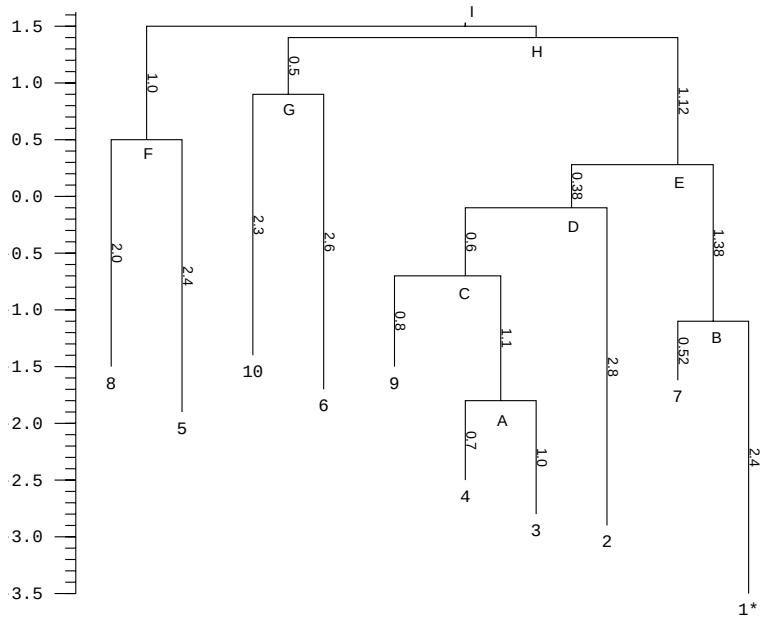
- local minimum if its energy is lower than the energy of **all** neighbors
- local maximum if its energy is higher than the energy of **all** neighbors
- saddle point if there are at least two local minima thar can be reached by a downhill walk starting at this point



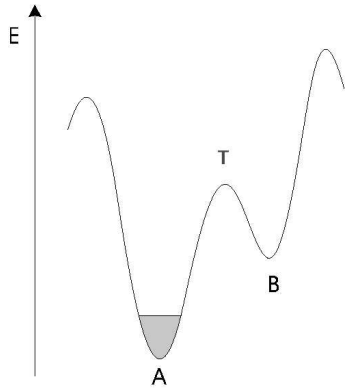
Energy barriers and barrier trees

Some topological definitions:
A structure is a

- local minimum if its energy is lower than the energy of **all** neighbors
- local maximum if its energy is higher than the energy of **all** neighbors
- saddle point if there are at least two local minima thar can be reached by a downhill walk starting at this point

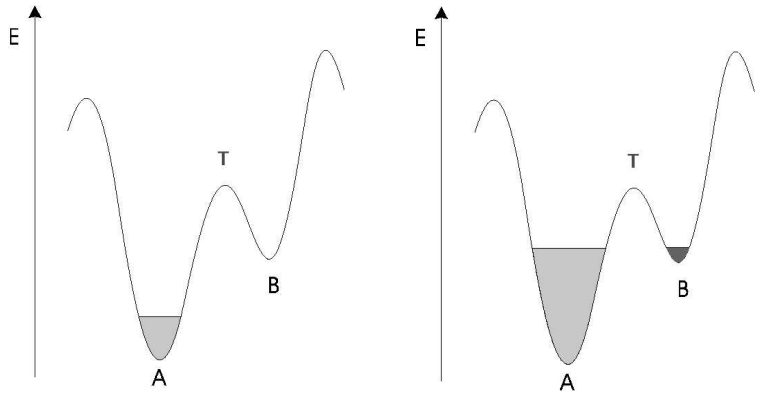


The flooding algorithm

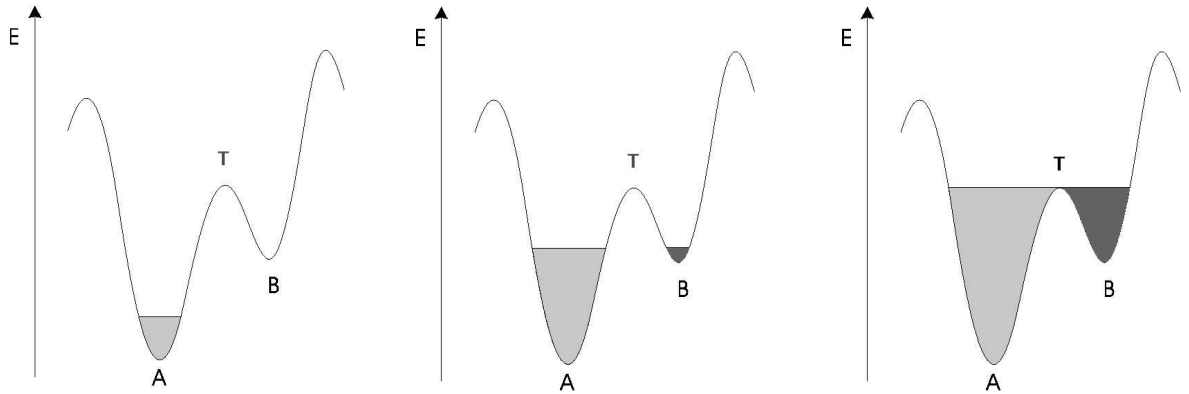


1

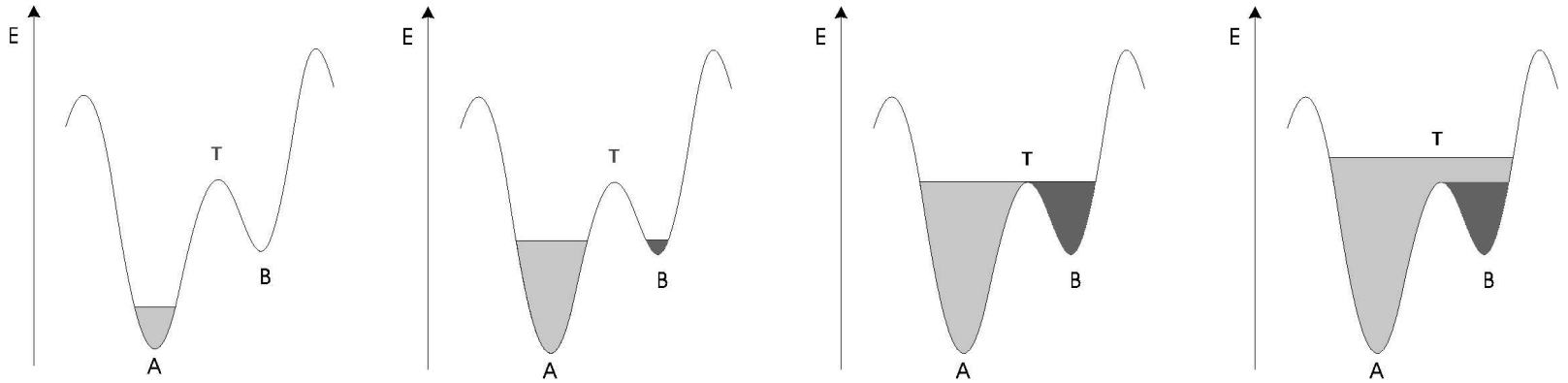
The flooding algorithm



The flooding algorithm



The flooding algorithm



Information from the barrier trees

- Local minima
- Saddle points
- Barrier heights
- Gradient basins
- Partition functions and free energies of (gradient) basins

N.B.: A *gradient basin* is the set of all initial points from which a gradient walk (steepest descent) ends in the same local minimum.

tbi

Dynamics of biopolymers

The probability distribution P of structures as a function of time is ruled by a set of forward equations, also known as the master equation

$$\frac{dP_t(x)}{dt} = \sum_{y \neq x} [P_t(y) k_{xy} - P_t(x) k_{yx}]$$

Given an initial population distribution, how does the system evolve in time? (What is the population distribution after n time-steps?)

$$\frac{d}{dt} P_t = \mathbf{U} P_t \quad \Longrightarrow \quad P_t = e^{t\mathbf{U}} P_0$$

tbi

Barrier tree kinetics

For a reduced description we need

- macro-states that form a partition of full configuration space
- transition rates between macro-states, e.g.

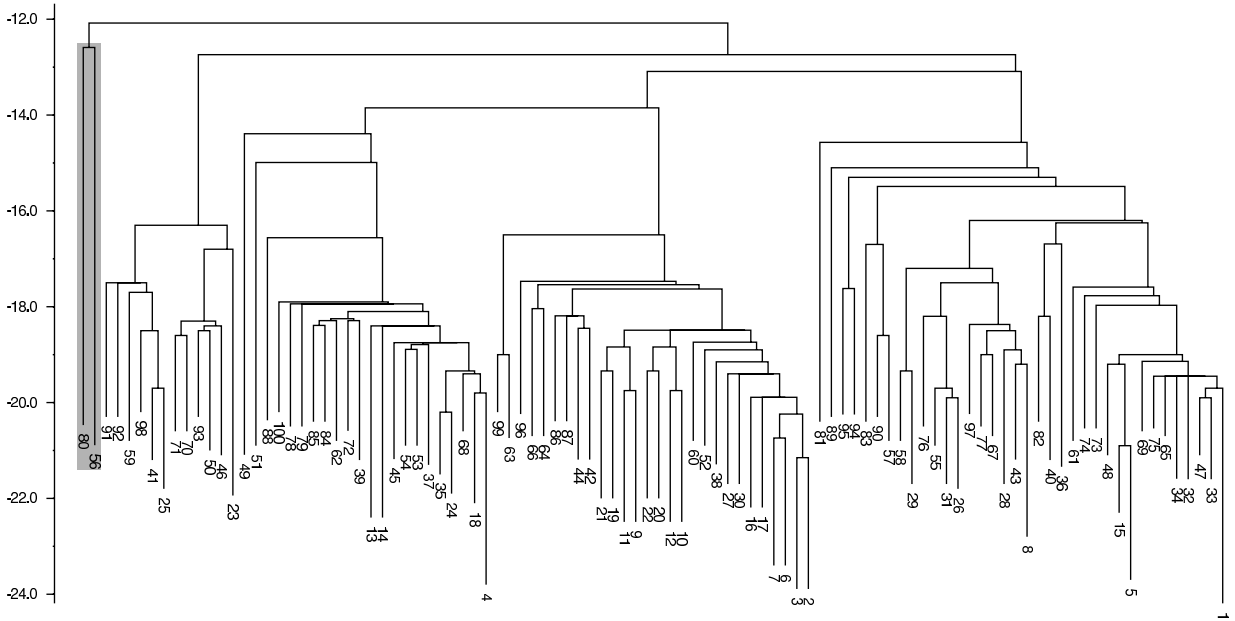
$$r_{\beta\alpha} = \Gamma_{\beta\alpha} \exp\left(-(E^*_{\beta\alpha} - G_\alpha)/kT\right)$$

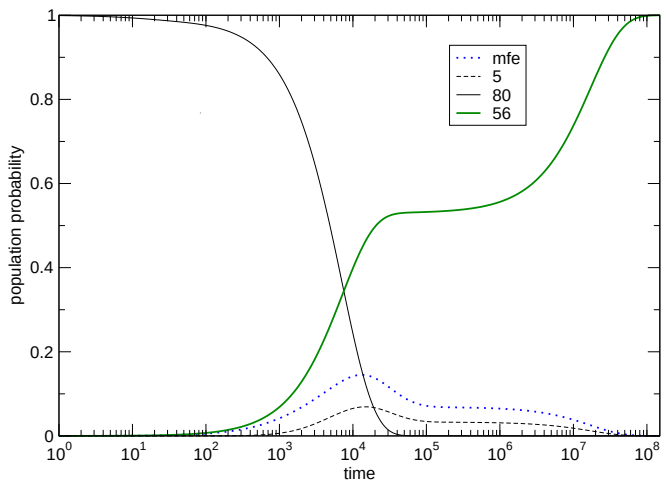
All relevant quantities can be computed via the flooding algorithm.

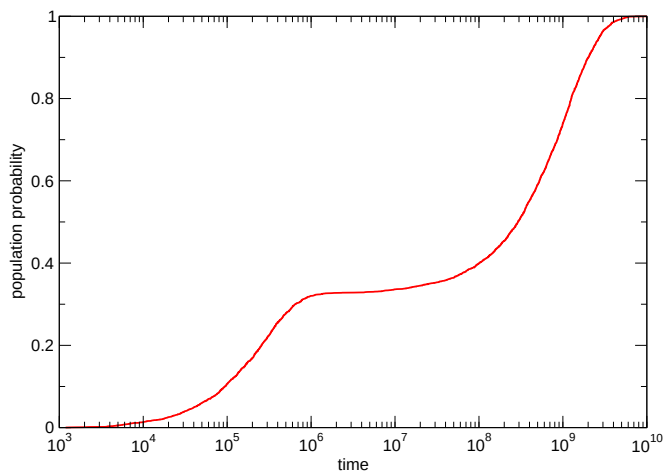
tbi

Dynamics of tRNA

GCGGAUUUAGCUCAGDDGGGAGAGCGCCAGACUGAAYAUCUGGAGGUCCUGUGTPCGAUCCACAGAAUUCGCACCA

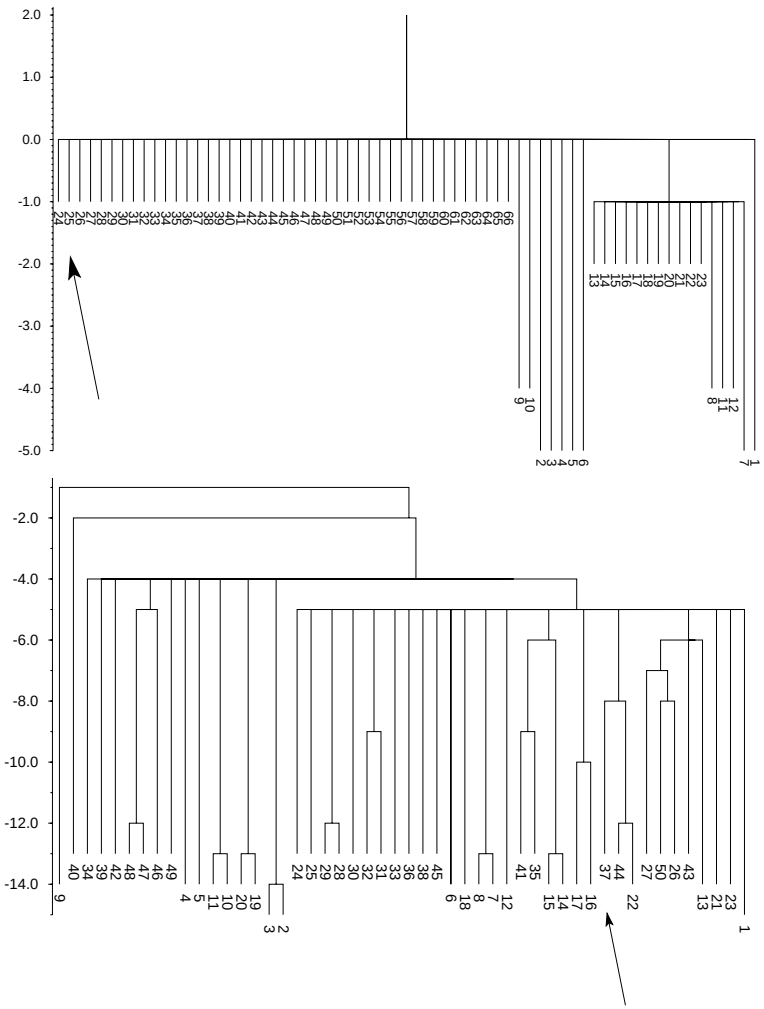


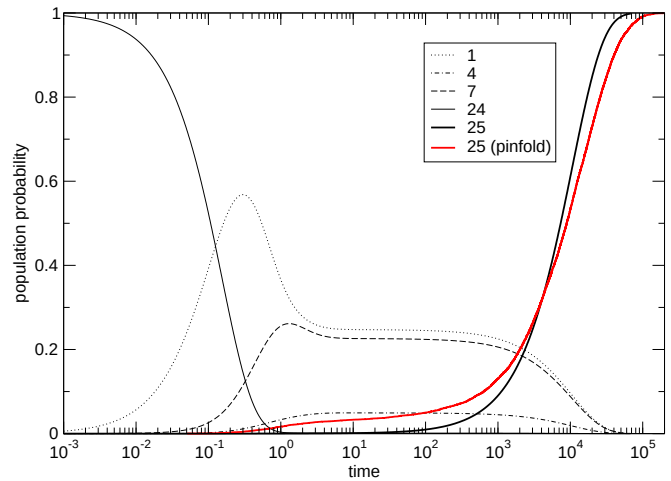


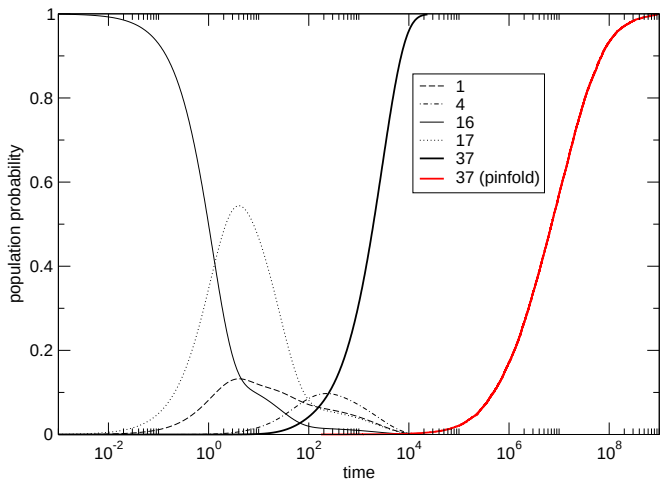


Dynamics of lattice proteins: HEX/TET lattice

NNHHPPNNPHHHHPXP $n = 16$







Conclusion

- Discrete models allow a detailed study of the energy surface.
- Barrier trees approximate the landscape topology and folding kinetics.
- A macrostate approach of folding kinetics reduces simulation time drastically.
- The accuracy of the model is very high in the case of RNA and mostly sufficient for lattice proteins.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

Conclusion

- Discrete models allow a detailed study of the energy surface.
- Barrier trees approximate the landscape topology and folding kinetics.
- A macrostate approach of folding kinetics reduces simulation time drastically.
- The accuracy of the model is very high in the case of RNA and mostly sufficient for lattice proteins.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

Conclusion

- Discrete models allow a detailed study of the energy surface.
- Barrier trees approximate the landscape topology and folding kinetics.
- A macrostate approach of folding kinetics reduces simulation time drastically.
- The accuracy of the model is very high in the case of RNA and mostly sufficient for lattice proteins.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

Conclusion

- Discrete models allow a detailed study of the energy surface.
- Barrier trees approximate the landscape topology and folding kinetics.
- A macrostate approach of folding kinetics reduces simulation time drastically.
- The accuracy of the model is very high in the case of RNA and mostly sufficient for lattice proteins.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

Conclusion

- Discrete models allow a detailed study of the energy surface.
- Barrier trees approximate the landscape topology and folding kinetics.
- A macrostate approach of folding kinetics reduces simulation time drastically.
- The accuracy of the model is very high in the case of RNA and mostly sufficient for lattice proteins.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

Thanks

Peter Stadler

Ivo Hofacker

Christoph Flamm

Peter Schuster

the audience

tbi