Estimation of low-energy refolding paths
Visualization of Lattice Protein Dynamics

Michael Wolfinger

Institute for Theoretical Chemistry
University Vienna

May 23, 2006

tbi

Outline

1 Lattice Proteins

2 Conformation space

3 Energy landscapes

4 Refolding paths

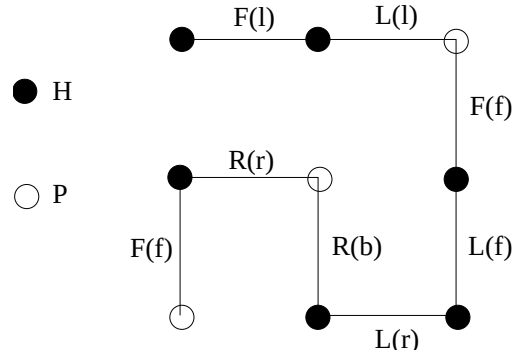
5 Dynamics

6 Visualization

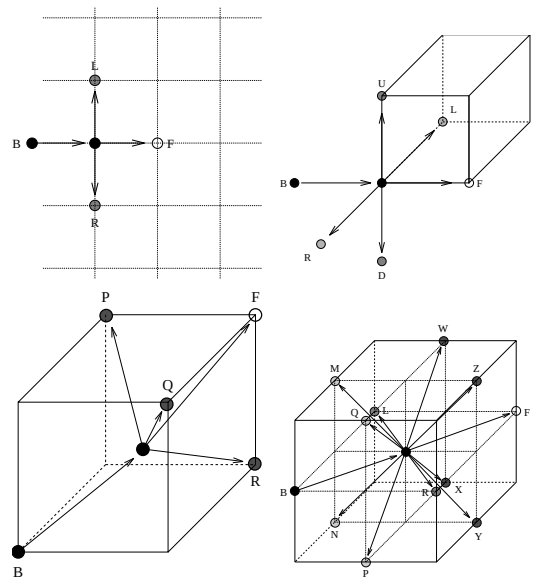
tbi

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.



FRRLLFLF



Advantages:

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

tbi

Contact Potentials

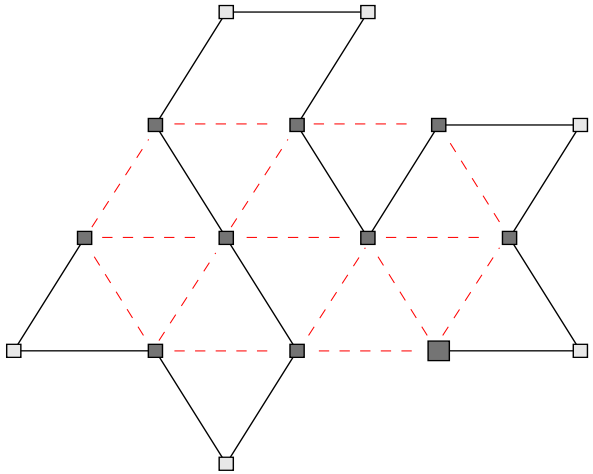
Generally, the energy function for a sequence with n residues $\mathfrak{S} = \mathfrak{s}_1\mathfrak{s}_2\ldots\mathfrak{s}_n$ with $\mathfrak{s}_i \in \mathscr{A} = \{a_1, a_2, \ldots, a_b\}$, the alphabet of b residues, and an overall configuration $x = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ on a lattice $\mathscr{L}$ can be written as the sum of pair potentials

$$E(\mathfrak{S}, x) = \sum_{\substack{i<j-1 \\ |\mathbf{x}_i - \mathbf{x}_j| = 1}} \Psi[\mathfrak{s}_i, \mathfrak{s}_j].$$

tbi

Lattice proteins

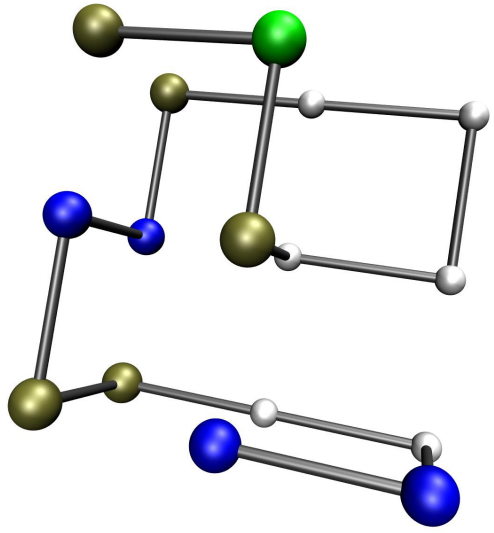
$\mathfrak{S} = \texttt{HPHPHHHPPHHHPHPH} \quad n = 16$



$E = -15$

	H	P
H	-1	0
P	0	0

$\mathfrak{S} = \texttt{NNHHPPNNPHHHHPXP} \quad n = 16$



$E = -16$

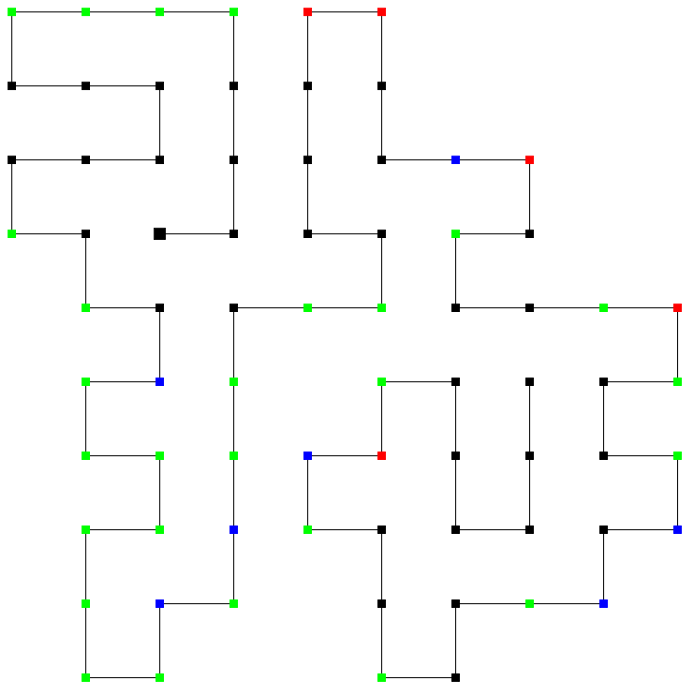
	H	P	N	X
H	-4	0	0	0
P	0	1	-1	0
N	0	-1	1	0
X	0	0	0	0

tbi

Lattice proteins - interaction scheme II

$\mathfrak{S} =$ HHHHNNNNHHHHHHHNHNHPNNNNNNNNPNPNNHNNHHHHHXXHHPXHNHHNXNHHNPHPNHHNHHNPXNHHHHHH

$n = 74$



$E = -120$

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

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.

Number of lattice protein structures

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

problem is NP-hard

In the RNA case

$$c_n \sim 1.86^n \cdot n^{-\frac{3}{2}}$$

dynamic programming algorithms available

dim	Lattice Type	μ	γ
2	SQ	2.63820	1.34275
	TRI	4.15076	1.343
	HEX	1.84777	1.345
3	SC	4.68391	1.161
	BCC	6.53036	1.161
	FCC	10.0364	1.162

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

- A set X of configurations
- a symmetric neighborhood relation $\mathfrak{N} : X \times X$
- an energy function $f : X \rightarrow \mathbf{R}$

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.

Number of lattice protein structures

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

problem is NP-hard

In the RNA case

$$c_n \sim 1.86^n \cdot n^{-\frac{3}{2}}$$

dynamic programming algorithms available

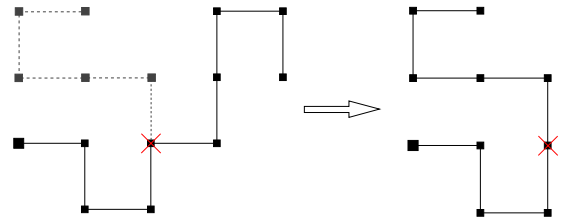
dim	Lattice Type	μ	γ
2	SQ	2.63820	1.34275
	TRI	4.15076	1.343
	HEX	1.84777	1.345
3	SC	4.68391	1.161
	BCC	6.53036	1.161
	FCC	10.0364	1.162

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

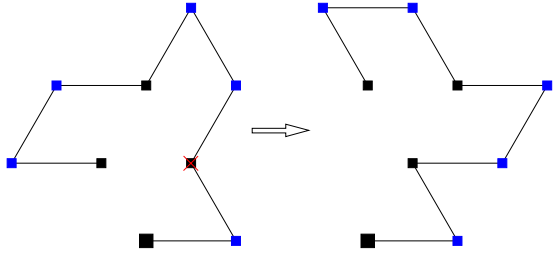
- A set X of configurations
- a symmetric neighborhood relation $\mathfrak{N} : X \times X$
- an energy function $f : X \to \mathbf{R}$

tbi

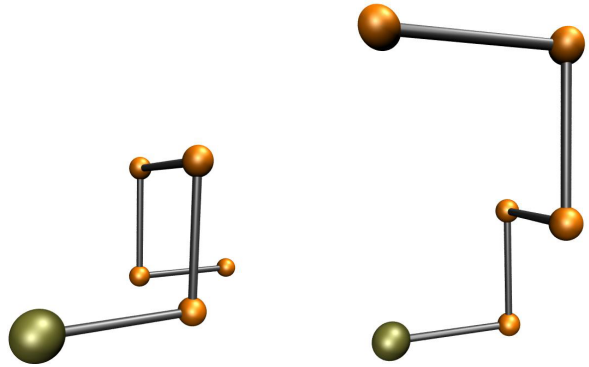
The move set



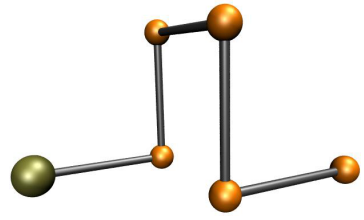
FRLLRLFRR FRLLFLFRR



FUDLUDLU FURLUDLU



FUURR FUDRR

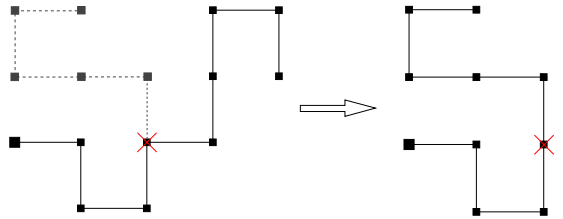


FUDLR

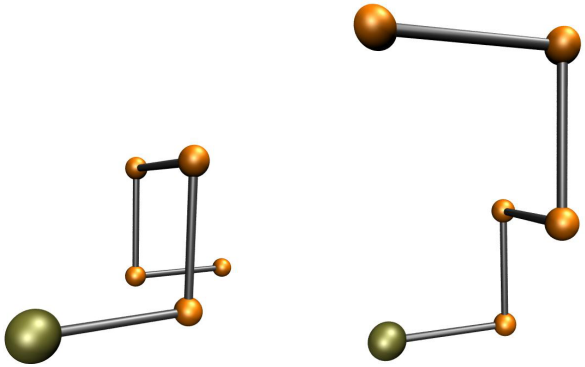
- For each move there must be an inverse move
- Resulting structure must be in X
- Move set must be *ergodic*

tbi

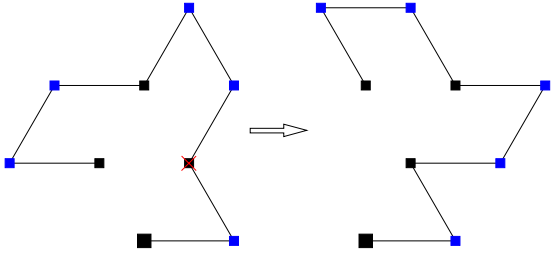
The move set



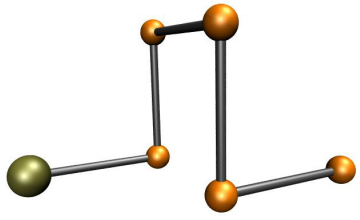
FRLLRLFRR FRLLFLFRR



FUURR FUDRR



FUDLUDLU FURLUDLU



FUDLR

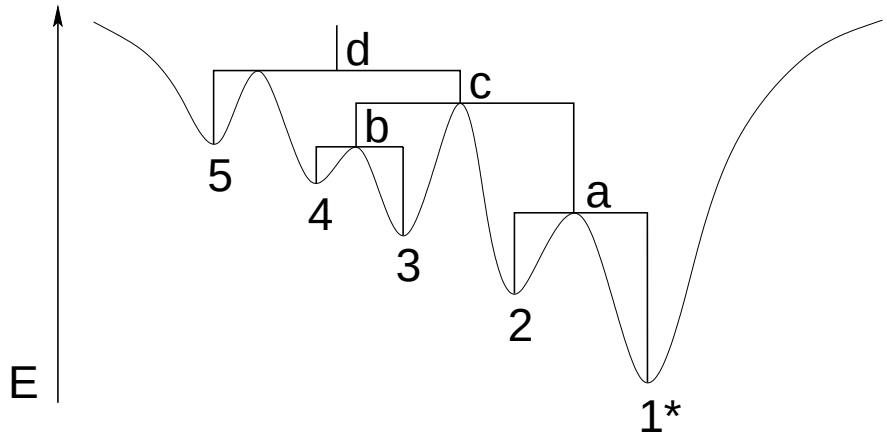
- For each move there must be an inverse move
- Resulting structure must be in X
- Move set must be *ergodic*

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



We further define

- a walk between two conformations x and y as a list of conformations $x = x_1 \dots x_{m+1} = y$ such that $\forall 1 \leq i \leq m : \mathfrak{N}(x_i, x_{i+1})$
- the lower part of the energy landscape (written as $X^{\leq \eta}$) as *all* conformations x such that $E(\mathfrak{S}, x) \leq \eta$ (with a predefined threshold η).

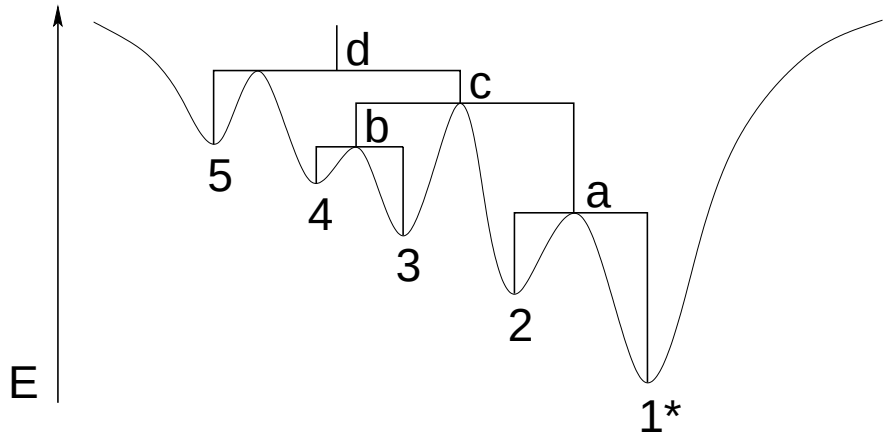
C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger.
Barrier trees of degenerate landscapes.
Z. Phys. Chem., 216:155–173, 2002.

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



We further define

- a walk between two conformations x and y as a list of conformations $x = x_1 \dots x_{m+1} = y$ such that $\forall 1 \leq i \leq m : \mathfrak{N}(x_i, x_{i+1})$
- the lower part of the energy landscape (written as $X^{\leq \eta}$) as *all* conformations x such that $E(\mathfrak{S}, x) \leq \eta$ (with a predefined threshold η).

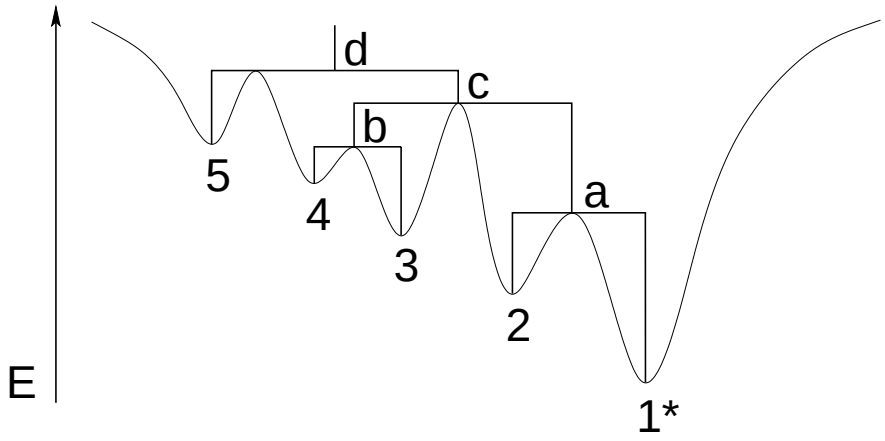
C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger.
Barrier trees of degenerate landscapes.
Z. Phys. Chem., 216:155–173, 2002.

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



We further define

- a walk between two conformations x and y as a list of conformations $x = x_1 \dots x_{m+1} = y$ such that $\forall 1 \leq i \leq m : \mathfrak{N}(x_i, x_{i+1})$
- the lower part of the energy landscape (written as $X^{\leq \eta}$) as *all* conformations x such that $E(\mathfrak{S}, x) \leq \eta$ (with a predefined threshold η).

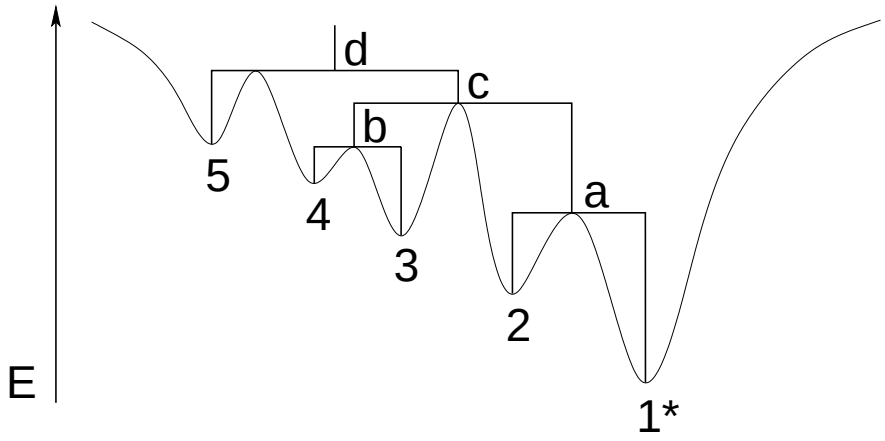
C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger.
Barrier trees of degenerate landscapes.
Z. Phys. Chem., 216:155–173, 2002.

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



We further define

- a walk between two conformations x and y as a list of conformations $x = x_1 \ldots x_{m+1} = y$ such that $\forall 1 \leq i \leq m : \mathfrak{N}(x_i, x_{i+1})$
- the lower part of the energy landscape (written as $X^{\leq \eta}$) as *all* conformations x such that $E(\mathfrak{S}, x) \leq \eta$ (with a predefined threshold η).

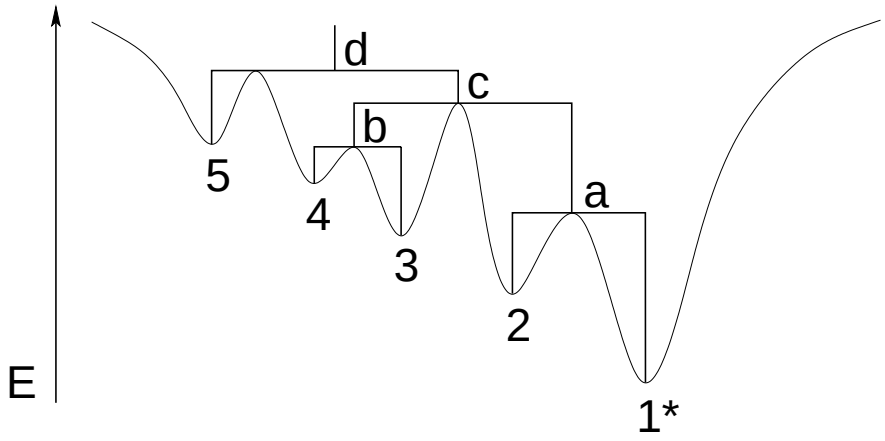
C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger.
Barrier trees of degenerate landscapes.
Z. Phys. Chem., 216:155–173, 2002.

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



We further define

- a walk between two conformations x and y as a list of conformations $x = x_1 \dots x_{m+1} = y$ such that $\forall 1 \leq i \leq m : \mathfrak{N}(x_i, x_{i+1})$
- the lower part of the energy landscape (written as $X^{\leq \eta}$) as *all* conformations x such that $E(\mathfrak{S}, x) \leq \eta$ (with a predefined threshold η).

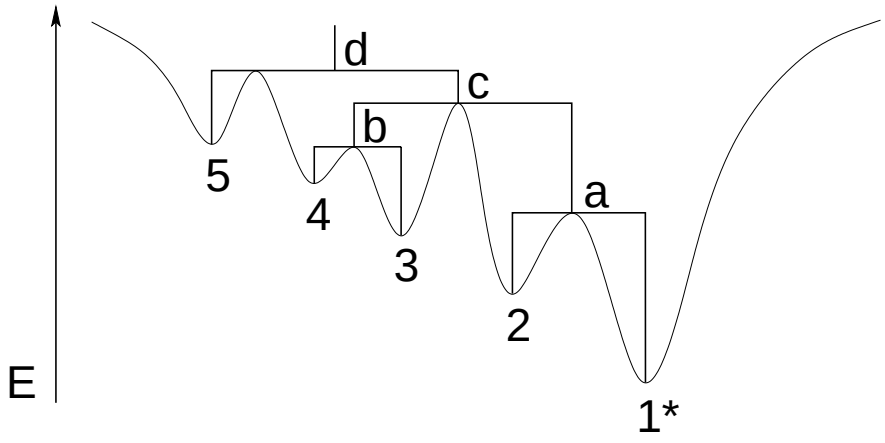
C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger.
Barrier trees of degenerate landscapes.
Z. Phys. Chem., 216:155–173, 2002.

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



We further define

- a walk between two conformations x and y as a list of conformations $x = x_1 \ldots x_{m+1} = y$ such that $\forall 1 \leq i \leq m : \mathfrak{N}(x_i, x_{i+1})$
- the lower part of the energy landscape (written as $X^{\leq \eta}$) as *all* conformations x such that $E(\mathfrak{S}, x) \leq \eta$ (with a predefined threshold η).

C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger.
Barrier trees of degenerate landscapes.
Z. Phys. Chem., 216:155–173, 2002.

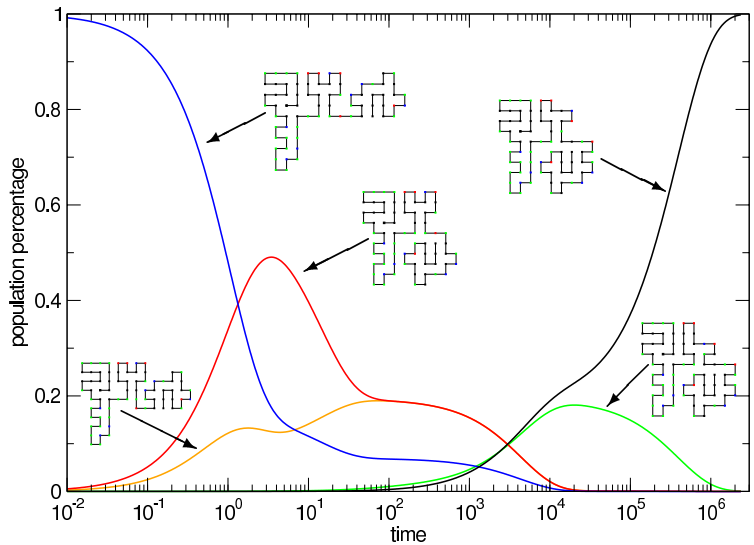
tbi

Information from the barrier tree

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

This information can be used to approximate the dynamics of biopolymers, i.e. transition rates between different macrostates (basins in the barrier tree)

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



The lower part of the energy landscape

Two conformations x and y are mutually accessible at the level η (written as $x \leftrightarrow\!\!\!\!\underline{\rho\ \eta\ }\!\!\!\!\rightarrow y$) if there is a walk from x to y such that all conformations z in the walk satisfy $E(\mathfrak{S}, z) \leq \eta$. The saddle height $\hat{f}(x, y)$ of x and y is defined by

$$\hat{f}(x, y) = \min\{\eta \mid x \leftrightarrow\!\!\!\!\underline{\rho\ \eta\ }\!\!\!\!\rightarrow y\}$$

Given the set of all local minima $X_{\text{min}}^{\leq \eta}$ below threshold η, the lower energy part $X^{\leq \eta}$ of the energy landscape can alternatively be written as

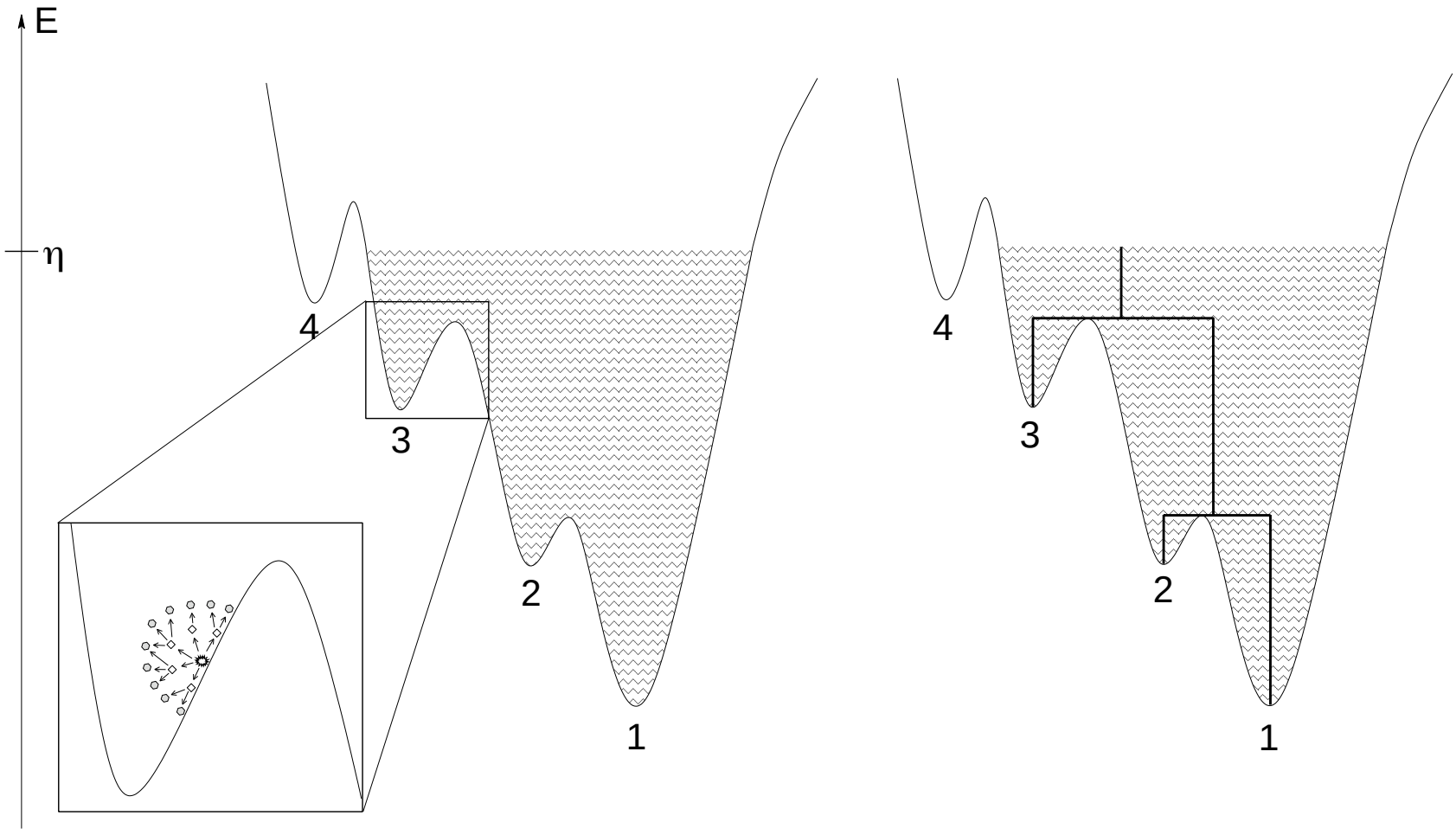
$$X^{\leq \eta} = \{y \mid \exists x \in X_{\text{min}}^{\leq \eta} : \hat{f}(x, y) \leq \eta\}$$

Given a restricted set of low-energy conformations, X_{init}, and a reasonable value for η, the lower part of the energy landscape can be calculated.

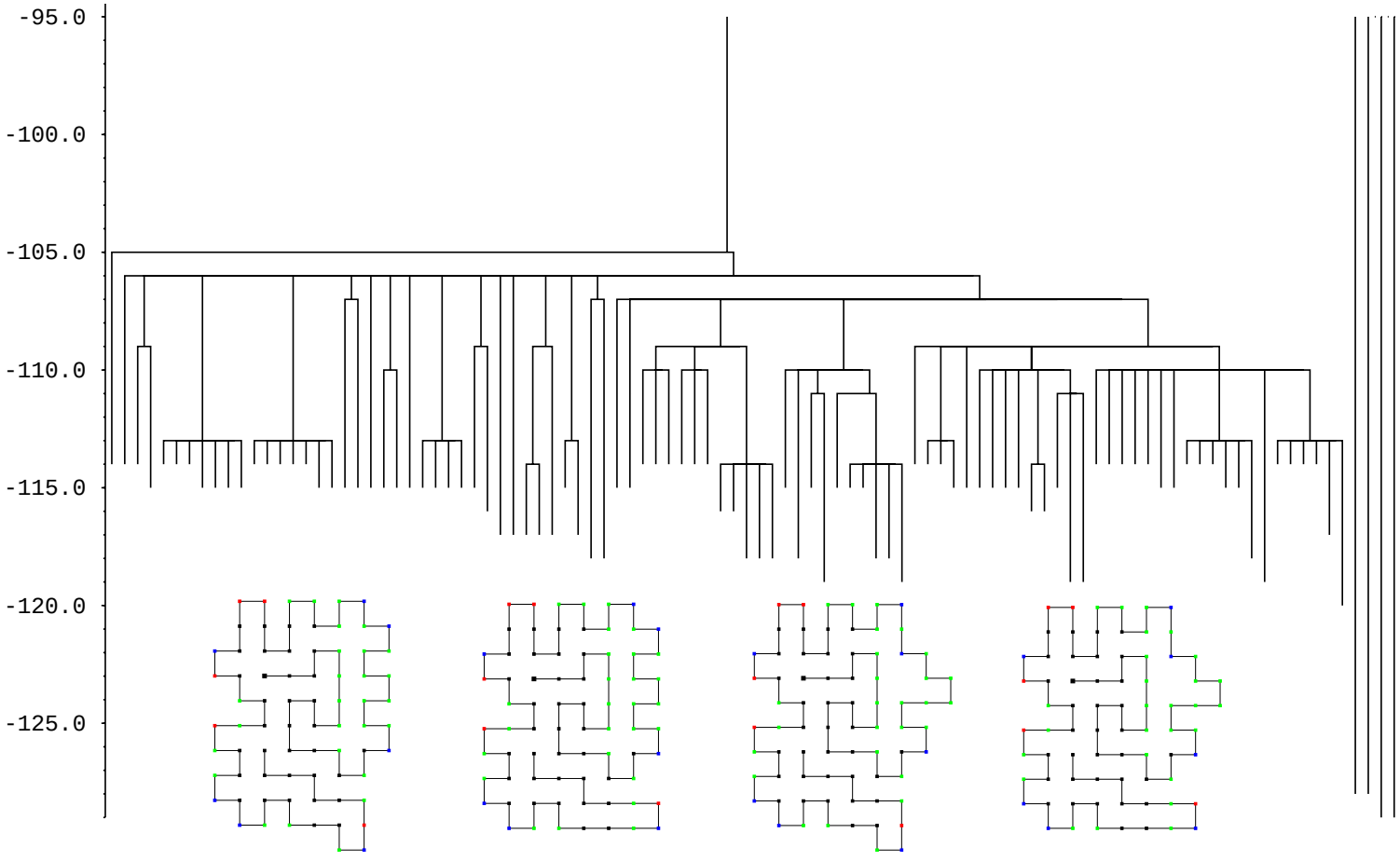
M. T. Wolfinger, S. Will, I. L. Hofacker, R. Backofen, and P. F. Stadler.
Exploring the lower part of discrete polymer model energy landscapes.
Europhys. Lett., 2006.

tbi

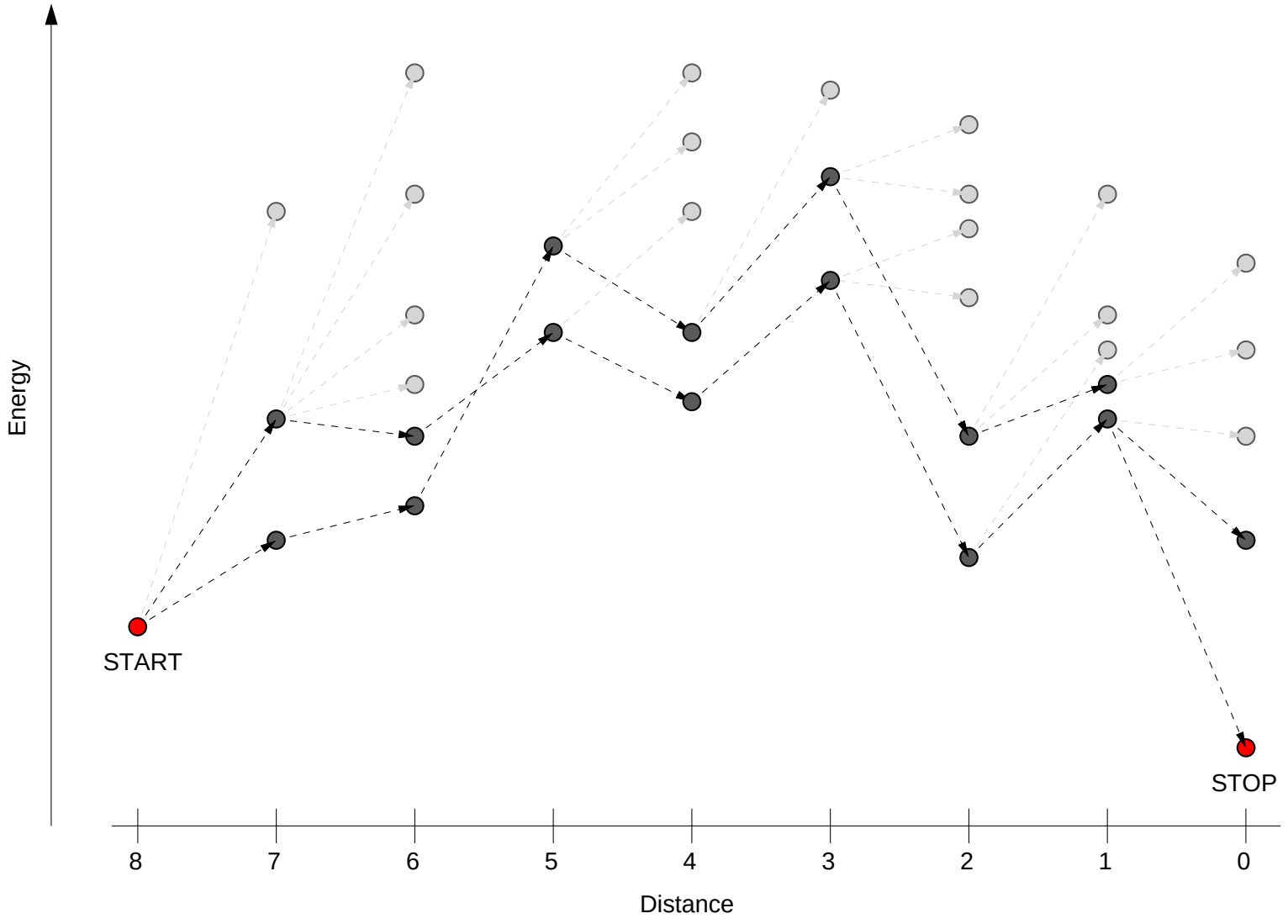
The Flooder approach



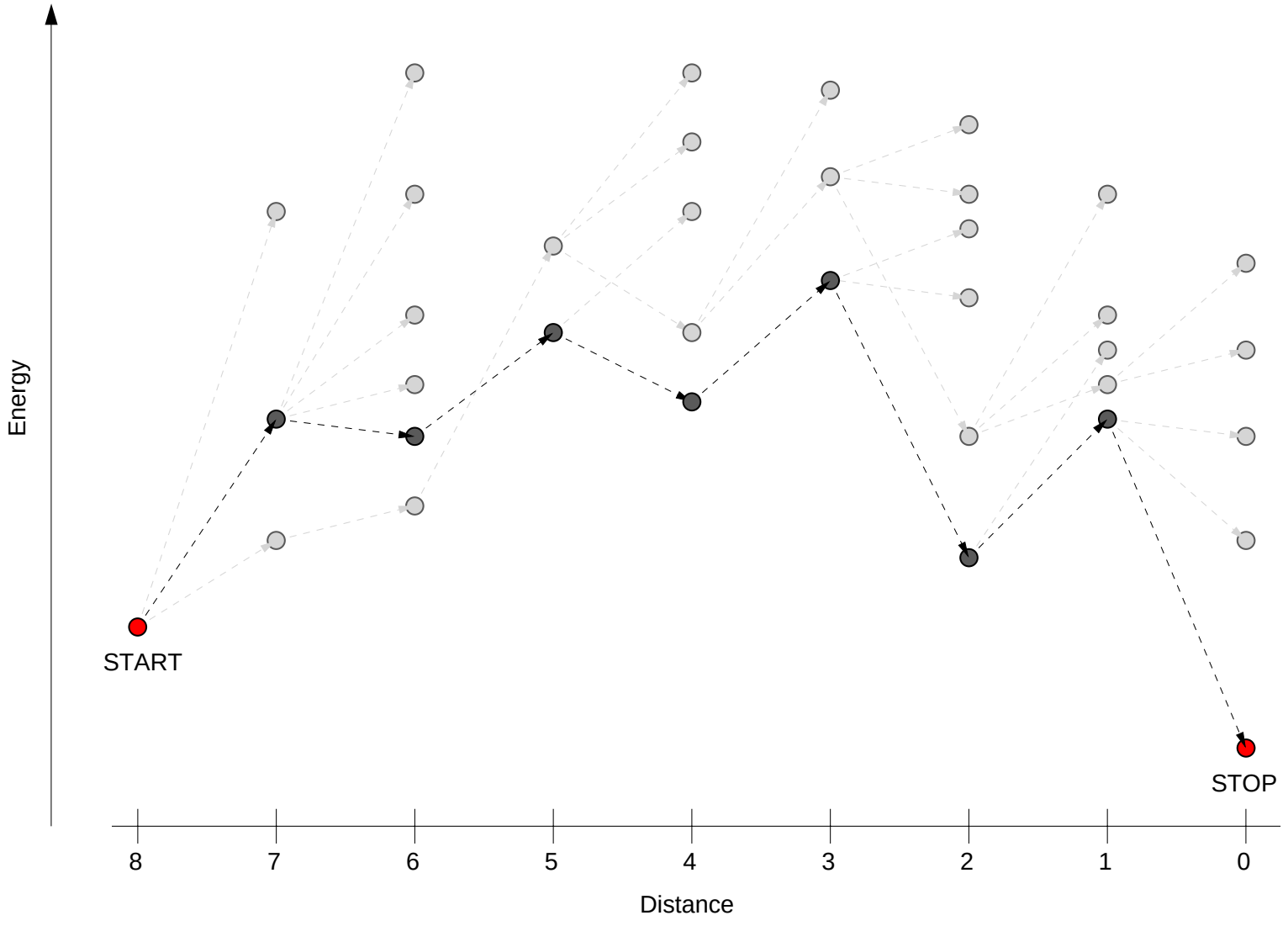
!Connected



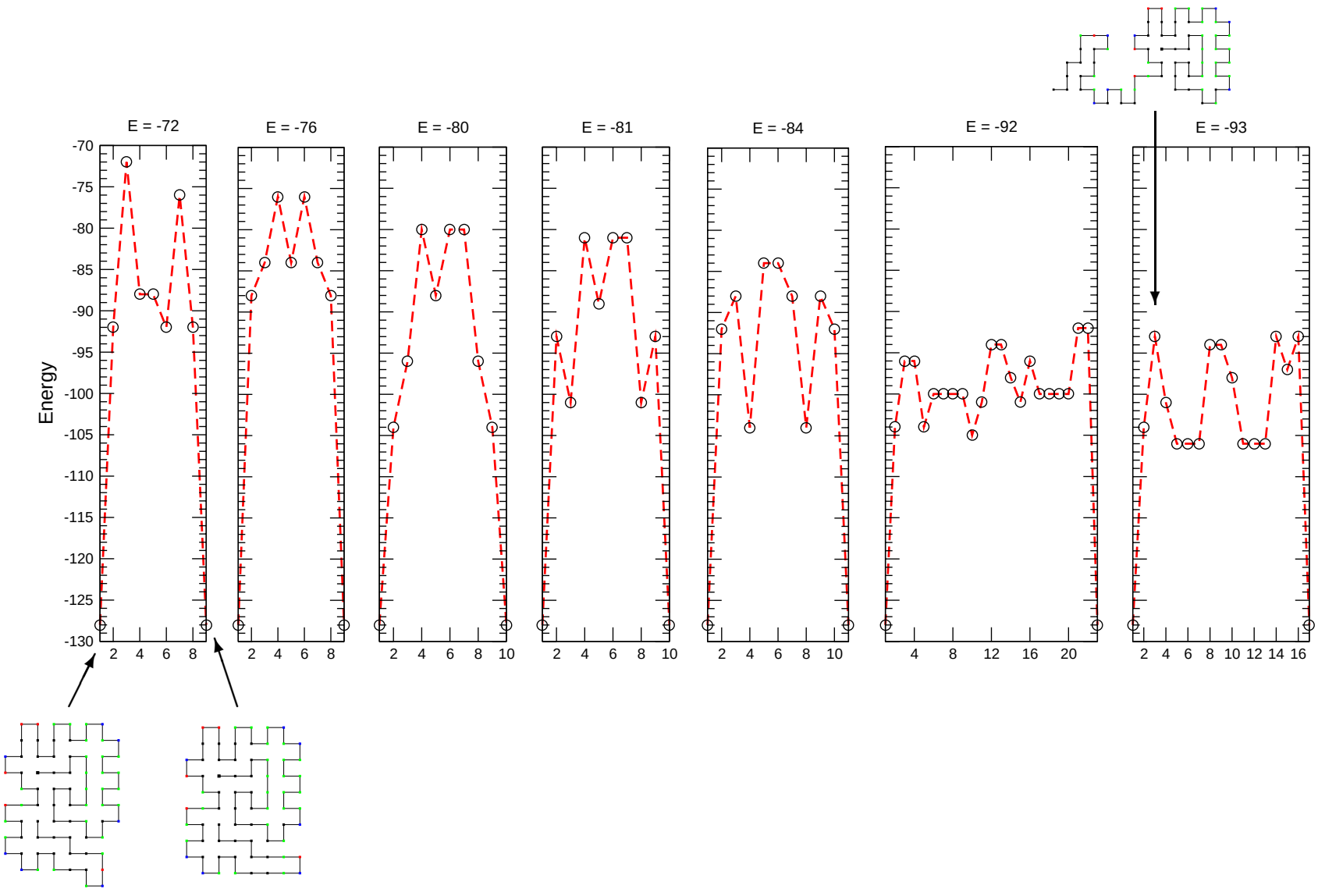
PathFinder - illustration



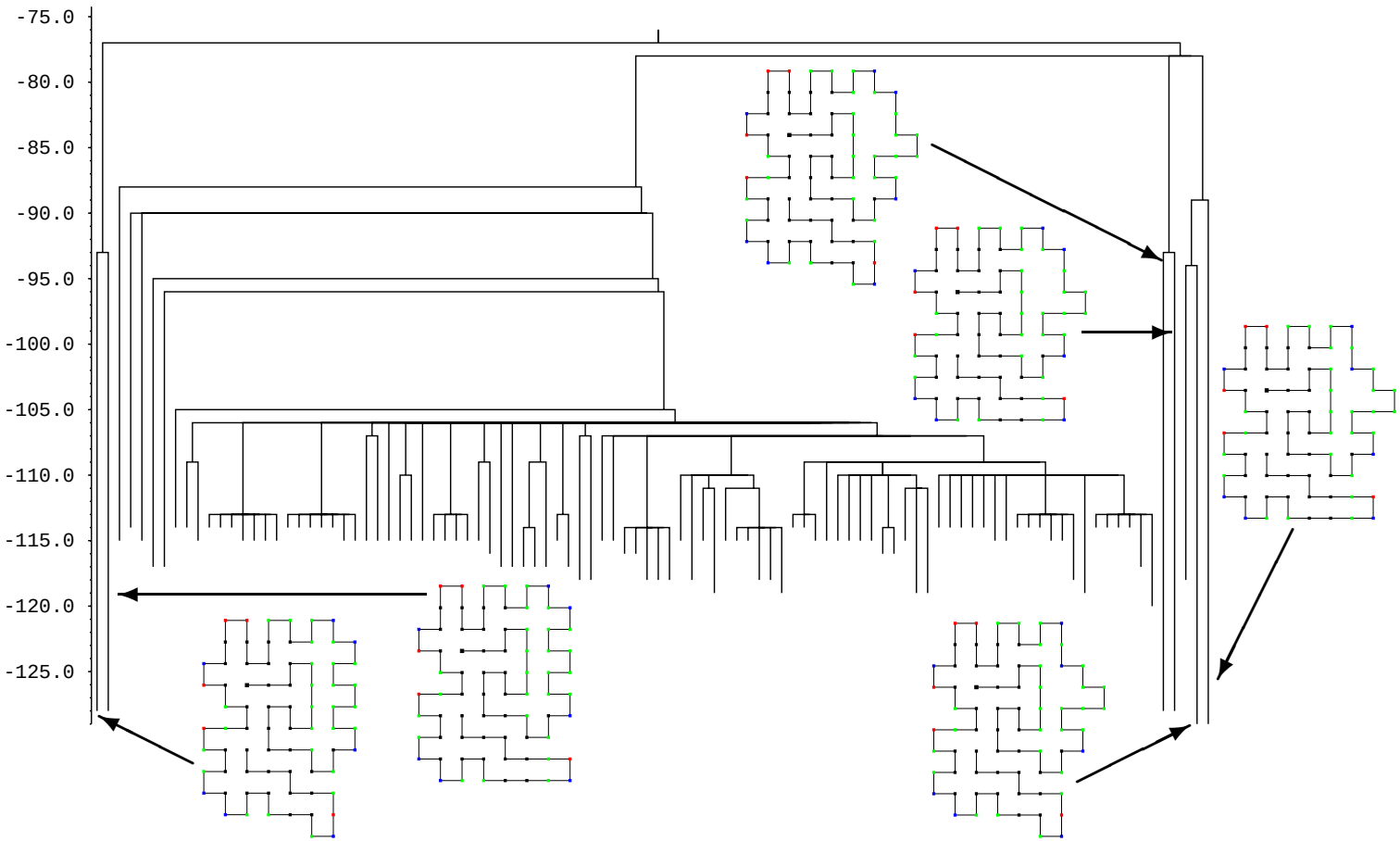
PathFinder - illustration



Refolding profiles



Connected!



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?)

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

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.

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

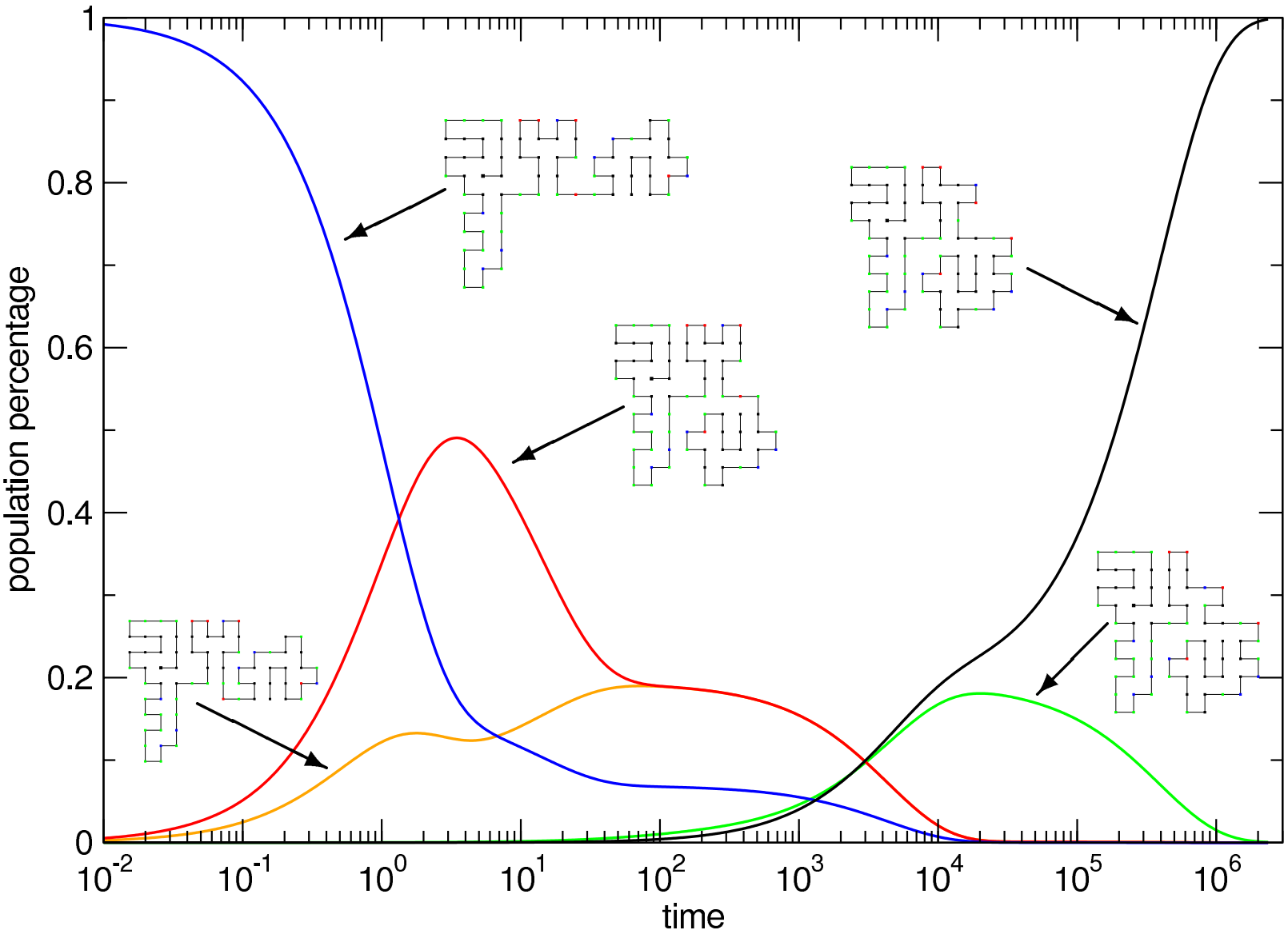
$$r_{\beta\alpha} = \sum_{y\in\beta} \sum_{x\in\alpha} r_{yx} \mathrm{Prob}[x|\alpha] \quad \text{for } \alpha \neq \beta \text{ with } r_{yx} = \begin{cases} e^{\frac{-\Delta E}{kT}} & \text{if } \Delta E > 0 \\ 0 & y \notin \mathcal{N}(x) \\ 1 & \end{cases}$$

All relevant quantities can be computed via the flooding algorithm.

M. T. Wolfinger, W. A. Svrcek-Seiler, C. Flamm, I. L. Hofacker, and P. F. Stadler.
Efficient computation of RNA folding dynamics.
J. Phys. A: Math. Gen., 37(17):4731–4741, 2004.

tbi

Barrier tree kinetics - example



tbi

Kinetic Folding Algorithm

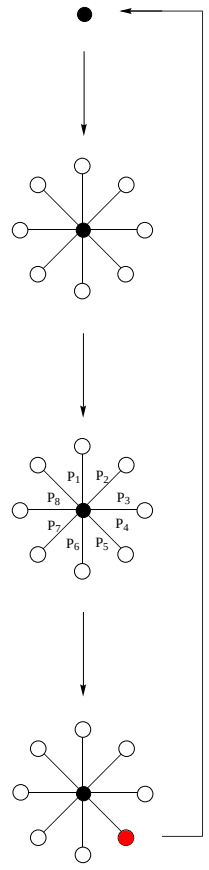
Simulate folding kinetics by a rejection-less Monte-Carlo type algorithm:

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 $1/\sum_i P_i$.



C. Flamm, W. Fontana, I. Hofacker, and P. Schuster.
RNA folding kinetics at elementary step resolution.
RNA, 6:325–338, 2000.

tbi

Visualization

- Visualize Pinfold output
- Support data analysis
- Emphasize possible relationships
- Provide simulation comparison
- Provide data analysis and exploration
- Uncover regularities
- Expose the unseen
- Speed up coginition

Shneiderman's mantra:

"Overview first, zoom and filter, details on demand"

- Start with an overview
- Let the user filter out interesting data
- Show details only on demand for different data

S. Pötzsch, G. Scheuermann, M. T. Wolfinger, C. Flamm, and P. F. Stadler.
Visualization of lattice-based protein folding simulations.
In *10th International Conference on Information Visualization (IV06)*, 2006.

tbi

Visualization

- Visualize Pinfold output
- Support data analysis
- Emphasize possible relationships
- Provide simulation comparison
- Provide data analysis and exploration
- Uncover regularities
- Expose the unseen
- Speed up coginition

Shneiderman's mantra:

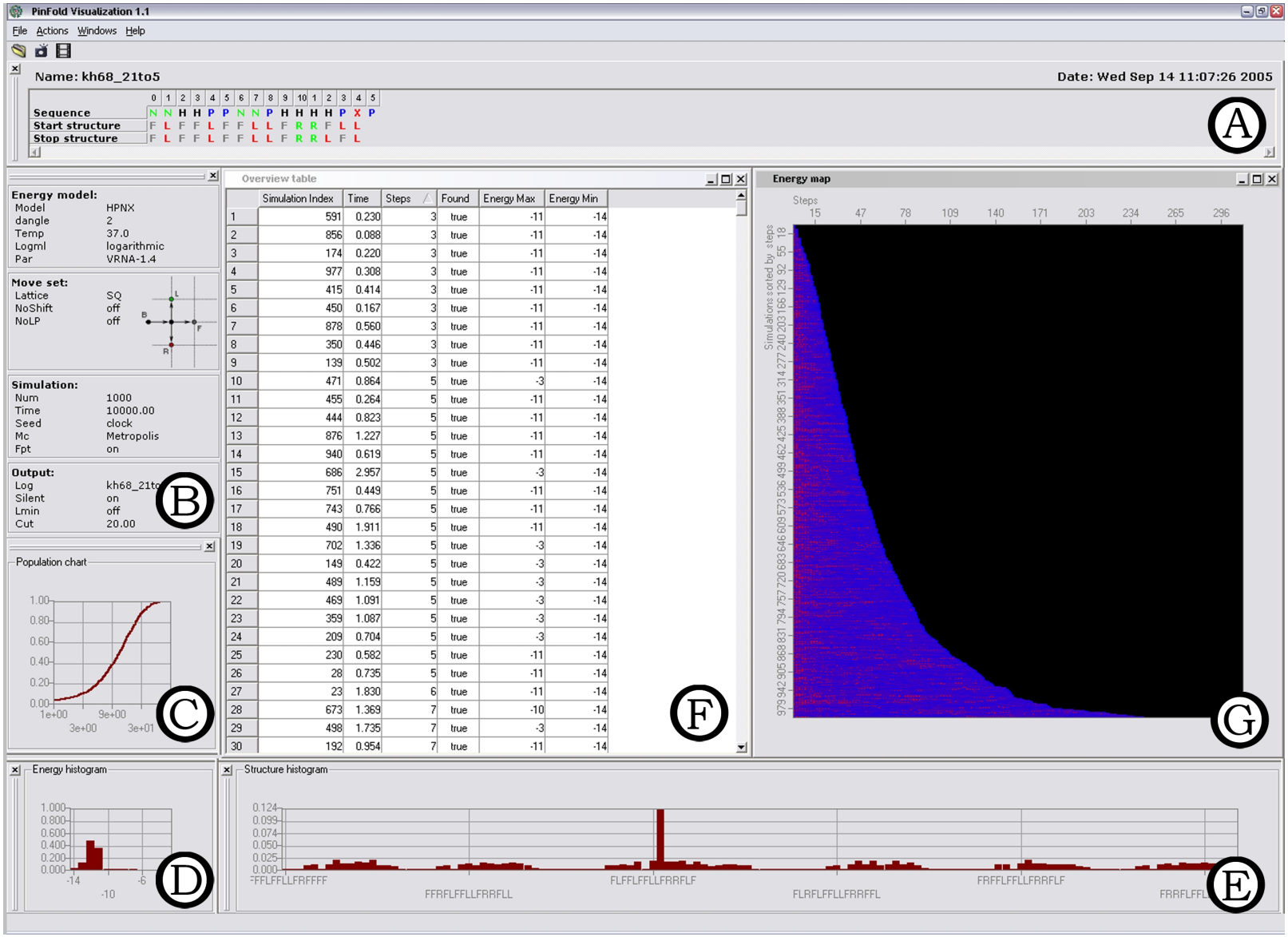
"Overview first, zoom and filter, details on demand "

- Start with an overview
- Let the user filter out interesting data
- Show details only on demand for different data

S. Pötzsch, G. Scheuermann, M. T. Wolfinger, C. Flamm, and P. F. Stadler.
Visualization of lattice-based protein folding simulations.
In *10th International Conference on Information Visualization (IV06)*, 2006.

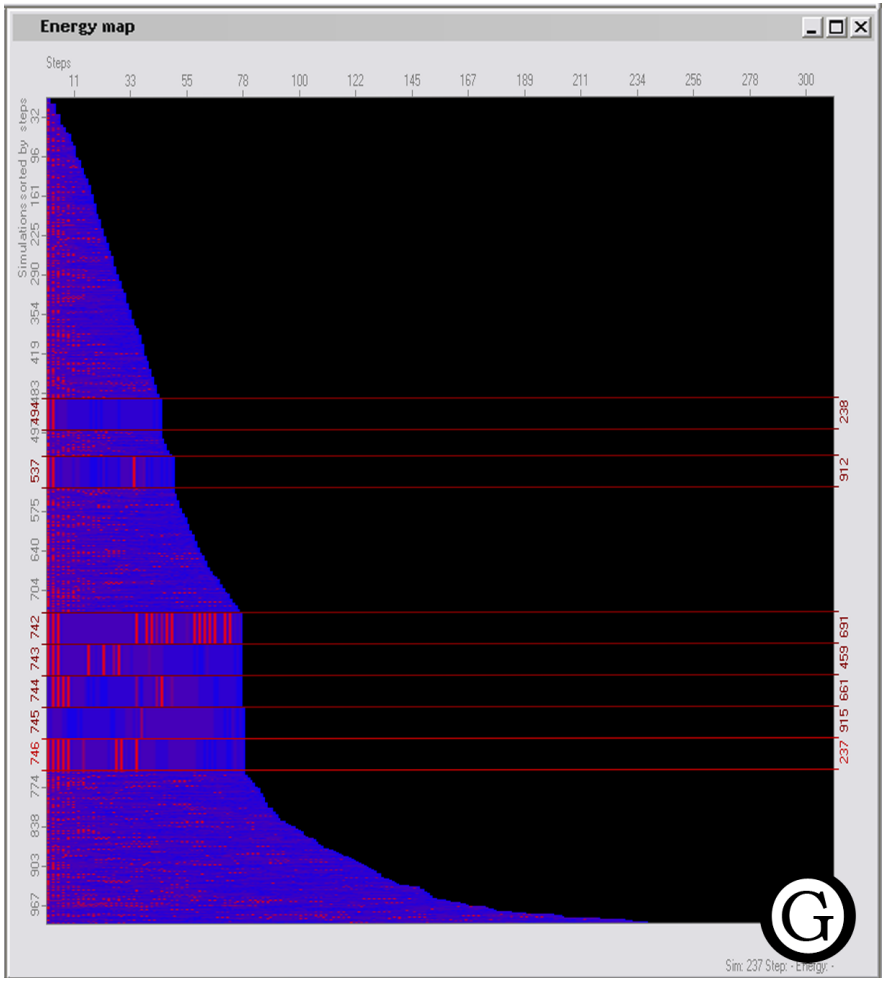
tbi

Overview first

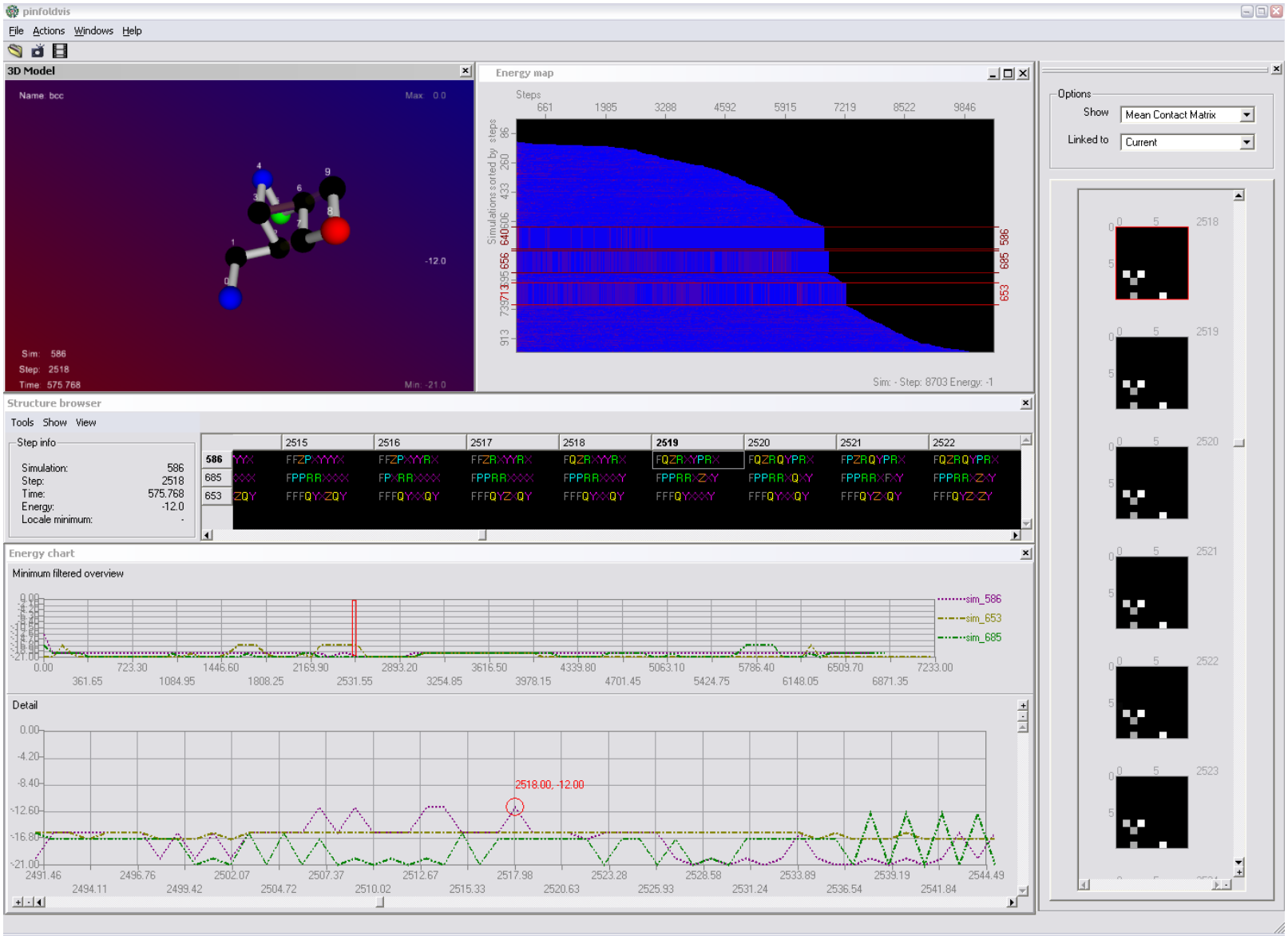


Zoom and filter

- Energy map
- Focus & context technique
- Huge data sets, limited screen size
- Details and overview in one window



Details on demand



Conclusion

- Barrier trees approximate the landscape topology and folding kinetics.
- A heuristic approach allows to sample low-energy refolding paths between different structures
- A macrostate approach of folding kinetics reduces simulation time drastically.
- A tool for visualization of folding trajectories enables a thorough investigation of folding kinetics simulations.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

tbi

Conclusion

- Barrier trees approximate the landscape topology and folding kinetics.
- A heuristic approach allows to sample low-energy refolding paths between different structures
- A macrostate approach of folding kinetics reduces simulation time drastically.
- A tool for visualization of folding trajectories enables a thorough investigation of folding kinetics simulations.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

tbi

Conclusion

- Barrier trees approximate the landscape topology and folding kinetics.
- A heuristic approach allows to sample low-energy refolding paths between different structures
- A macrostate approach of folding kinetics reduces simulation time drastically.
- A tool for visualization of folding trajectories enables a thorough investigation of folding kinetics simulations.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

tbi

Conclusion

- Barrier trees approximate the landscape topology and folding kinetics.
- A heuristic approach allows to sample low-energy refolding paths between different structures
- A macrostate approach of folding kinetics reduces simulation time drastically.
- A tool for visualization of folding trajectories enables a thorough investigation of folding kinetics simulations.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

tbi

Conclusion

- Barrier trees approximate the landscape topology and folding kinetics.
- A heuristic approach allows to sample low-energy refolding paths between different structures
- A macrostate approach of folding kinetics reduces simulation time drastically.
- A tool for visualization of folding trajectories enables a thorough investigation of folding kinetics simulations.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

tbi

Thanks

Ivo Hofacker

Christoph Flamm

Sebastian Will

Sebastian Pötzsch

Gerik Scheuermann

Rolf Backofen

Peter Stadler

Peter Schuster

tbi