Energy Landscapes of Lattice Proteins

Michael Wolfinger

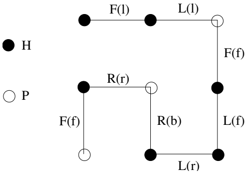
Institute for Theoretical Chemistry
University Vienna

July 26, 2005

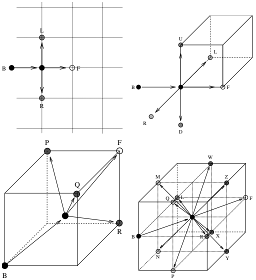
tbi

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

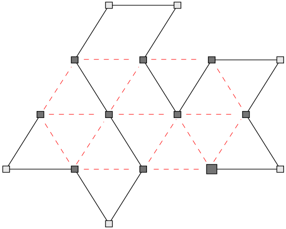


Advantages:

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

Lattice proteins

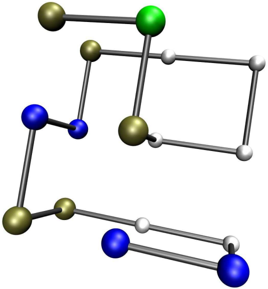
HPHPHHHPPHHHPHPH $n = 16$



$E = -15$

	H	P
H	-1	0
P	0	0

NNHHPPNNPHHHHPXP $n = 16$



$E = -16$

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

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

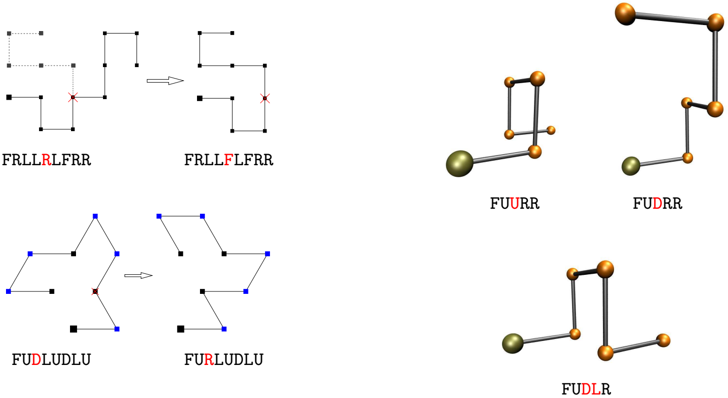
dim	Lattice Type	μ	γ
2	SQ	2.63820	1.34275
	TRI	4.15076	1.343
	HEX	1.84777	1.345
3	SC	4.68391	1.161
	BCC	6.53036	1.161
	FCC	10.0364	1.162

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

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

tbi

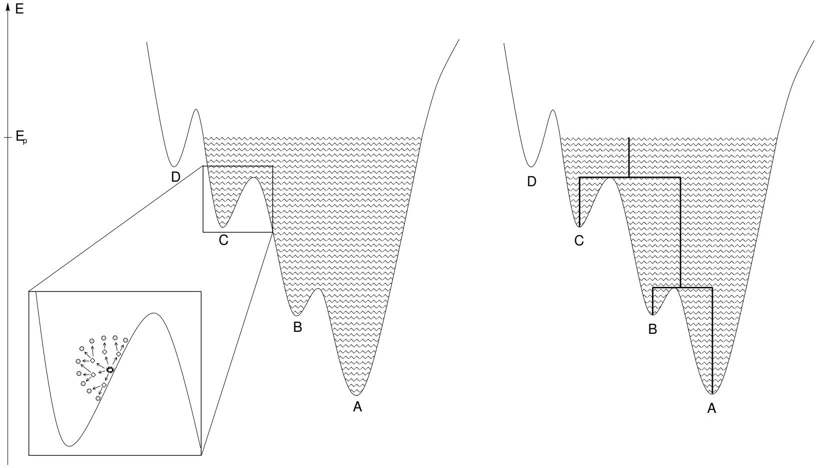
The move set



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

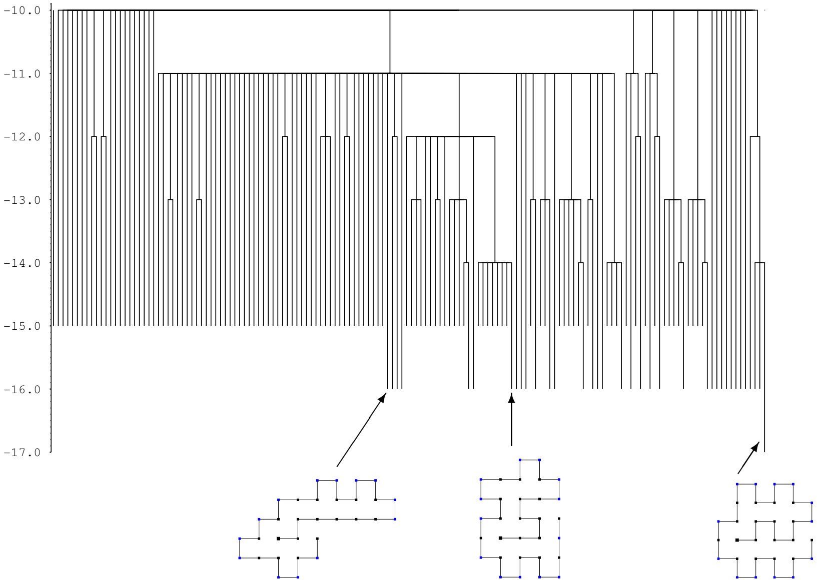
tbi

Low-energy states of lattice proteins



tbi

LP Energy Landscape



Kinetic Folding Algorithm

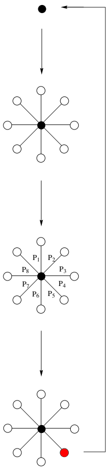
Simulate folding kinetics by a rejection-less Monte-Carlo type algorithm:

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 $1/\sum_i P_i$.



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.

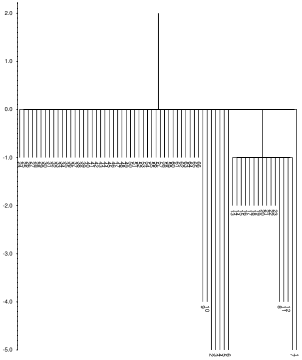
$r_{\beta\alpha} = \Gamma_{\beta\alpha} \exp\left(-(E^*_{\beta\alpha} - G_\alpha)/kT\right)$ or

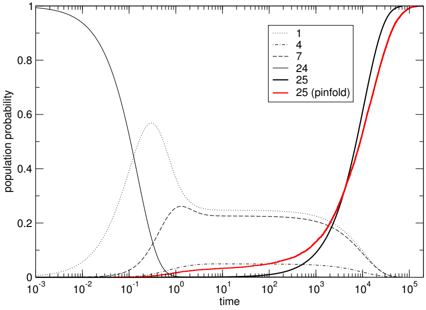
$r_{\beta\alpha} = \sum_{y\in\beta} \sum_{x\in\alpha} r_{yx} \mathrm{Prob}[x|\alpha] \quad \text{for } \alpha \neq \beta$ with $r_{yx} = \begin{cases} e^{\frac{-\Delta E}{kT}} & if\, \Delta E > 0 \\ 0 & y \notin \mathcal{N}(x) \\ 1 & \end{cases}$

All relevant quantities can be computed via the flooding algorithm.

Dynamics of lattice proteins: HEX lattice

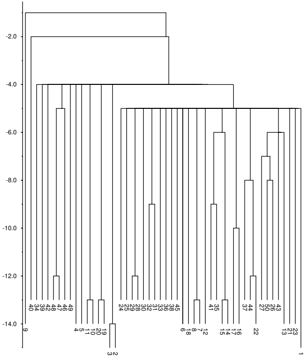
NNHHPPNNPHHHHPXP $n = 16$

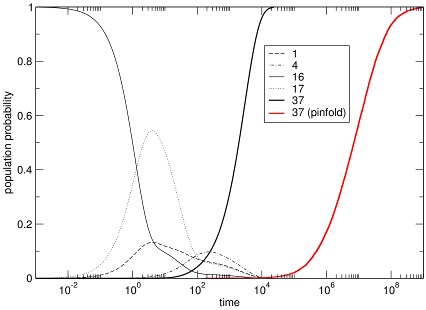




Dynamics of lattice proteins: TET lattice

NNHHPPNNPHHHHPXP $n = 16$

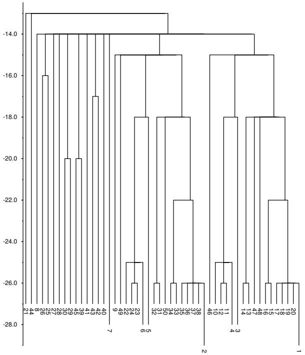


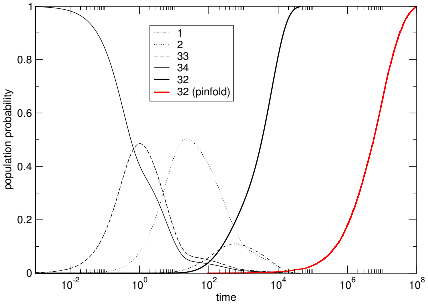


tbi

Dynamics of lattice proteins: TRI 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 mostly sufficient for lattice proteins.
- This newly generated framework provides a powerful method for further refinement of biopolymer folding landscapes.

Thanks

Peter Schuster

Peter Stadler

Ivo Hofacker

Christoph Flamm