Simple Models and Protein Folding: Exact Solution of the Muñoz-Eaton Model

P. B. & A. Pelizzola

INFM & Dipartimento di Fisica, Politecnico di Torino

- Two-state thermodynamics, two-state kinetics: what relationship?

- The Muñoz-Eaton model for protein folding

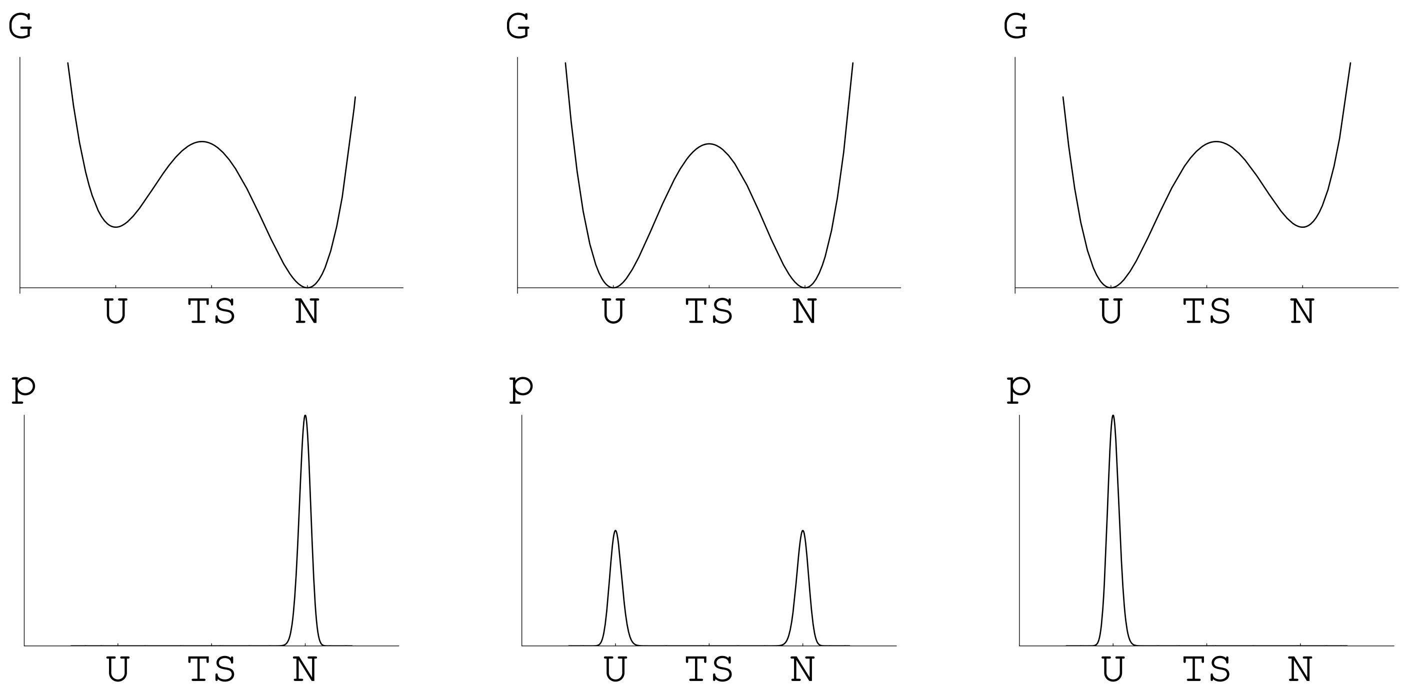
- (Previous) analysis of the Muñoz-Eaton model under simplifying assumptions

- The exact solution; equilibrium properties

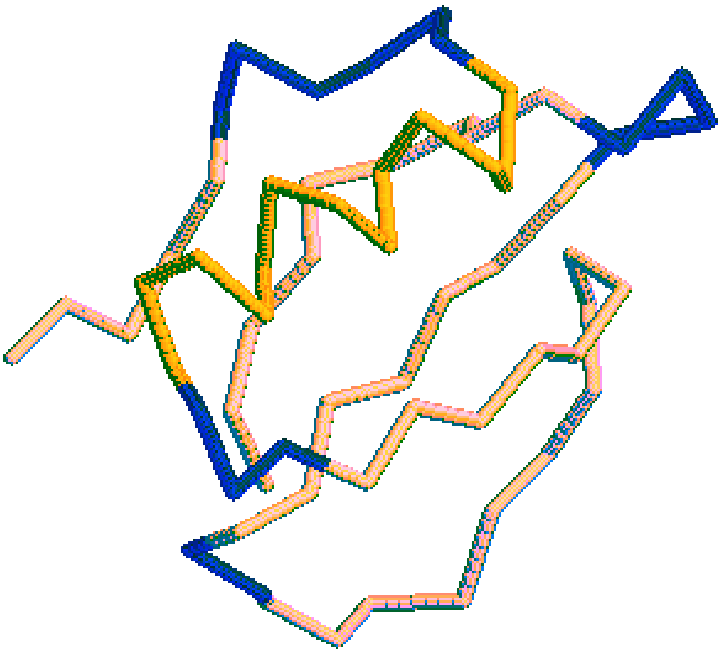
- Results for a β-hairpin and protein CI-2

Two-State Thermodynamics

Low T | $T = T^{\ddagger}$ | High T



Protein CI-2 (pdb code: 1COA)



Two-State Kinetics

Consider the process: U $\underset{k_f}{\overset{k_u}{\leftrightharpoons}}$ N

In general, given $N(t)$ = fraction of molecules in state N:

$$\frac{dN(t)}{dt} = -k_u(t)N(t) + k_f(t)U(t)$$

If $k_f(t) = k_f$, $k_u(t) = k_u \Rightarrow$ **simple (two-state) kinetics**:

$$\frac{\Delta N(t)}{\Delta N(0)} = \frac{N(t) - N_{eq}}{N(0) - N_{eq}} = e^{-k\,t}$$

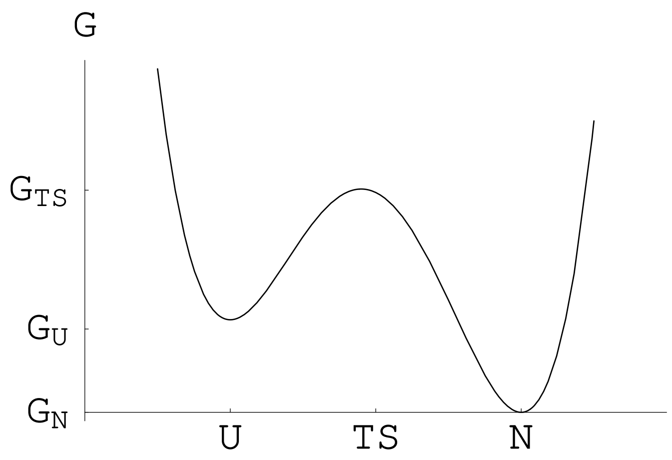
where $k = k_u + k_f$, $N_{eq} = N(t \to \infty)$

Conditions for the existence of a rate constant k:

- existence of a reaction coordinate
- existence of a barrier $\Delta G_{\ddagger} \gg RT$
- separation of time scales: during motion along the reaction coordinate, all other modes can be considered at equilibrium

If these conditions are fulfilled:

$$k_f = \tilde{k}_f(T)\, e^{-\Delta G_{\text{TS-U}}(T)\,/\,RT}$$
$$k_u = \tilde{k}_u(T)\, e^{-\Delta G_{\text{TS-N}}(T)\,/\,RT}$$



$\Rightarrow$ Kinetics is related to equilibrium properties!
(warning: $\tilde{k}_{f,u}$ account for microscopic dynamics, and could vary during folding)

But wait a moment...

In a generic polypeptide, one would expect:

- indefinitely many local minima, kinetic traps, transition states
- all time scales
- glassy thermodynamics and dynamics

so, **why proteins show just "two-states"?**

- what is the reaction coordinate?
- what is the denatured state like?
- what is the transition state?

Experiments say:

- denatured state: more or less compact or structured (depends on protein)

- transition state: and ensemble of states, with a folding nucleus (some fragments of secondary and tertiary structure)

- "nucleation-condensation" mechanism

Homeworks for the theorist:

- Write down a (minimal) polymer model showing two-state thermodynamics & kinetics ...

- Preliminary tasks:

  1. try describing folding, rather than predicting it:
     use the knowledge of the native structure.

  2. perhaps only native interactions are relevant along the reaction coordinate:
     try disregarding the other interactions.

  3. an all-atom description entails a huge configuration space (and painful simulations):
     try discretization of the monomer states.

The Muñoz-Eaton Model

- Only two states of peptide units:

$$m_i = \begin{cases} 1 & \text{native (i.e. native dihedral angles)} \\ 0 & \text{unfolded} \end{cases}$$

  A native peptide yields an entropic cost Δs_i

- Known native structure (contact map Δ_{ij}^{nat}):

$$\Delta_{ij}^{nat} = \begin{cases} 1 & \text{if residues i,j are in contact in} \\ & \text{the native state} \\ 0 & \text{otherwise} \end{cases}$$

- Only native interactions: ϵ_{ij} = energy associated to the contact ij in the native state

- Contact ij is active only if the whole stretch ij is native ($\prod_{k=i}^{j} m_k = 1$)

$$H_{\text{eff}}(\{m_k\}) = \sum_{i<j=2}^{N} \epsilon_{ij} \, \Delta_{ij}^{nat} \prod_{k=i}^{j} m_k - T \sum_{i=1}^{N} \Delta s_i \, m_i$$

Previous Analysis of the Model

Approximations to thermodynamics:

$$\mathcal{Z} = \sum_{\{m_1,\ldots,m_N\}=0,1} \exp[-\beta H_{\text{eff}}(\{m_1,\ldots,m_N\})]$$

- single/double sequence approximation: only consider configurations with one/two contiguos stretches of native residues (Eaton & co.)
- mean field and Monte carlo simulations (Maritan & co.)

Approximations for kinetics:

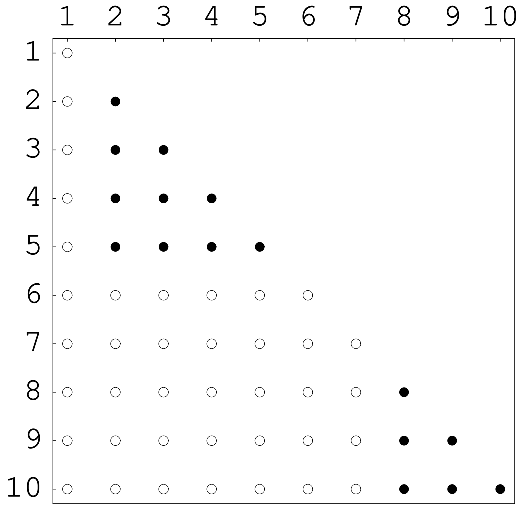
- elementary dynamics of adding/removing a residue from a native stretch; simulations within single/double seq. approx.
- kinetic rates prediction: try to identify a reaction coordinate and free energy barriers; evaluate folding/unfolding rate from equilibrium properties

Exact Solution of the Muñoz-Eaton Model

- Introduce $m_{j,i} = \prod_{k=i}^{j} m_k$:
  1d model, nonlocal interactions $\to$ 2d, local:

$$H_{\mathrm{eff}}(\{m_k\}) = \sum_{i<j=2}^{N} \epsilon_{ij}\, \Delta_{ij}^{nat} m_{j,i} - T \sum_{i=1}^{N} \Delta s_i\, m_{i,i}$$

- Local constraints $m_{j,i} = m_{j-1,i}\, m_{j,i+1} = 0, 1 \to$ "triangles" in the configuration space:



$\bullet = 1$, $\circ = 0$

Key observation: any row state is of the kind:

$$(\overbrace{0,0,\dots,0,\underbrace{1,1,\dots,1}_{k}}^{j}) \equiv \mathbf{v}_k^{(j)} = (\overbrace{0,\dots,0,\underset{k+1}{1},0,\dots,0}^{j+1})$$

2^N states on the diagonal, but just $(j+1)$ states $(k=0,\dots,j)$ on row j!

The partition function in a tranfer matrix formalism

$$\mathcal{Z} = \sum_{\{m_1 \ldots m_N\}=0,1} e^{-\beta H_{\text{eff}}(m_1 \ldots m_N)} = \sum_{\{m_1 \ldots m_N\}=0,1} w(m_1 \ldots m_N)$$

Consider a generic configuration:

$$w(0,1,1,0,0,1,1,1,1,0) = \bar{w}_1 \, w_{2,3} \, \bar{w}_4 \, \bar{w}_5 \, w_{6,9} \, \bar{w}_{10}$$

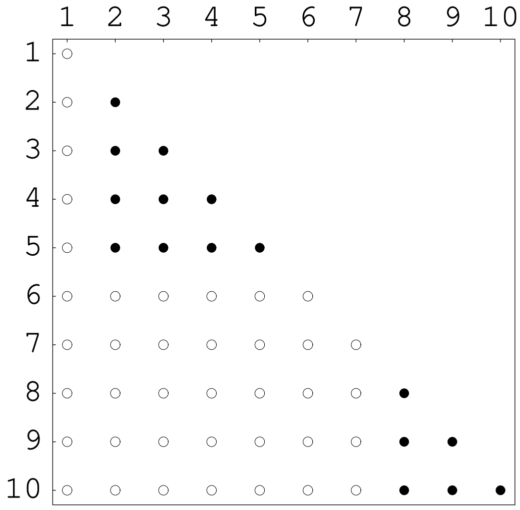
$\bar{w}_i$ = Boltzmann weight of $m_i = 0$.
$w_{i,j}$ = Boltzmann weight of native stretch i-j

To evaluate $w(0,1,1,0,0,1,1,1,1,0)$:

- start from the last position
- any time you find a native stretch:
  - assign the proper weight to its last element,
  - then do nothing on the others:

$$\begin{array}{cccccccccc}
\bar{w}_1 & & w_{2,3} & \bar{w}_4 & \bar{w}_5 & & & & w_{6,9} & \bar{w}_{10} \\
0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10
\end{array}$$

Now imagine to do this for all configurations...
With the transfer matrix formalism, configurations are processed in parallel!



Transfer matrix from row $(j+1)$ to j, Q^{j}_{j+1}:

$$Q^{j}_{j+1}\,\mathbf{v}^{(j+1)}_{k} \;=\; \mathbf{v}^{(j)}_{k-1} \qquad\qquad k \neq 0$$

$$\begin{aligned} Q^{j}_{j+1}\,\mathbf{v}^{(j+1)}_{0} \;=\; & \bar{w}_j \mathbf{v}^{(j)}_{0} + w_{j,j}\mathbf{v}^{(j)}_{1} + w_{j-1,j}\mathbf{v}^{(j)}_{2} \\ & + \ldots + w_{1,j}\mathbf{v}^{(j)}_{j} \end{aligned}$$

$\bar{w}_j = \exp(\Delta s_j/R) =$ Boltzmann weight of $m_j = 0$
$w_{i,j}$ = Boltzmann weight of native stretch i–j

Free energy $F = -RT \log \mathcal{Z}$, where

$$\mathcal{Z} = \mathbf{v}^{(0)}_{0}\, Q^{0}_{1} \ldots Q^{j-1}_{j} \ldots Q^{N}_{N+1}\, \mathbf{v}^{(N+1)}_{0}$$

In the same fashion, free energy profiles, correlation functions, ...

β-hairpin from protein G B1

16 C-terminal residues (15 peptide units)

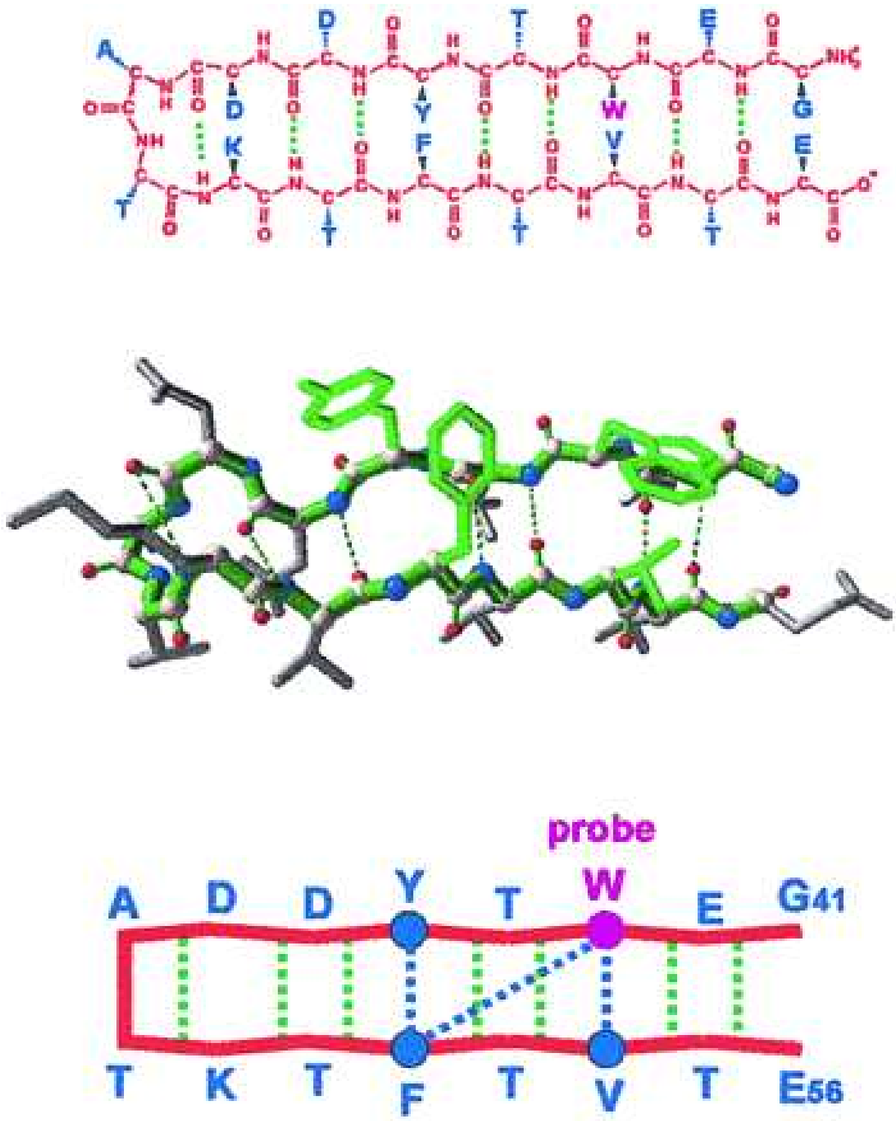
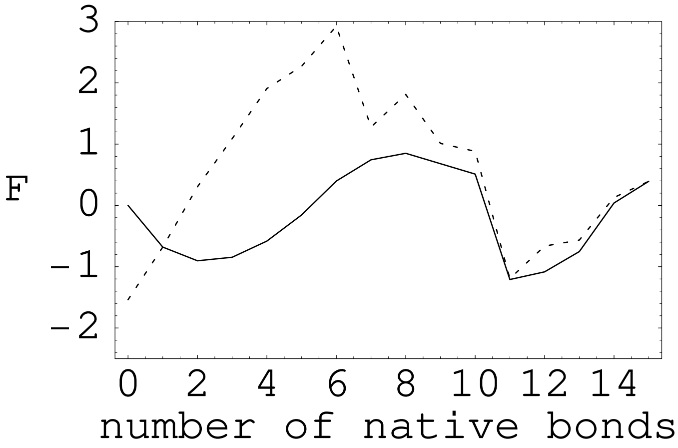


Figure from Muñoz et al., PNAS **95**, 5872 (1998).

Green dashed lines: hydrogen bonds; blue dashed lines: hydrophobic contacts

Results for the β-hairpin

Free energy profile (F in Kcal/mol)

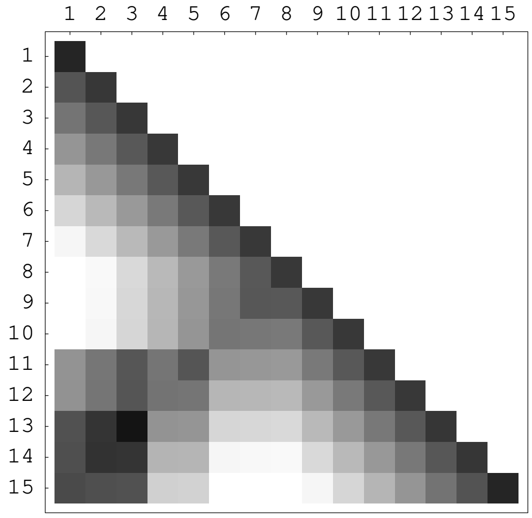


Solid = exact solution; dotted= single sequence approx.

Energy parameters: $\Delta H_{hb} = -0.96$ Kcal mol^{-1}, $\Delta G_{sc} = -1.94$ Kcal mol^{-1}, $\Delta S_{conf} = -2.74$ cal K^{-1} mol^{-1}.

Probability of native stretches

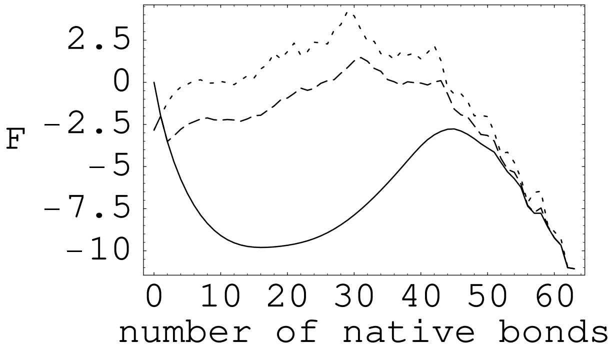
$$\mu_{j,i} = \langle (1-m_{i-1})(\prod_{k=i}^{j} m_k)(1-m_{j+1}) \rangle$$



– In $\mu_{j,i}$: darker regions = more probable stretches

Results for Protein CI-2

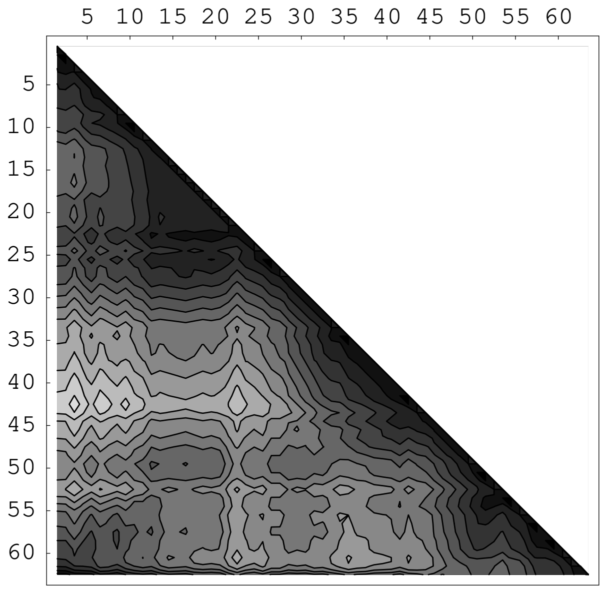
Free energy profile (F in Kcal/mol)



Solid = exact solution; dotted= single sequence; dashed= double sequence approx.

Energy parameters: $\epsilon = -0.521$ Kcal mol^{-1}, $\Delta s_0 = -1.2$ cal K^{-1} mol^{-1}, $\Delta s_1 = -3.3$ cal K^{-1} mol^{-1}

Probability of native stretches



Darker regions = more probable stretches

Conclusions

- Short proteins often show very simple (two-state) behavior in thermodynamics and dynamics, while random polymers do not. Why?

- Macroscopic explanation: nucleation-condensation folding mechanism, with one (or just a few) big free energy barrier along the folding pathway

- Detailed microscopic, theoretical description: still lacking. Promising approaches include the use of very simple models.

- The Muñoz-Eaton model and its exact solution: very simple and fast to evaluate, can be applied to any protein $\rightarrow$ in principle, a "first tool" to explain the main characteristics of the folding process

- More detailed studies are needed, though, to identify a reaction coordinate, characterize the folding barriers, find the relationship between kinetic rates and equilibrium properties.

References

On simple thermodynamics:

- Chan, Proteins, Struct. Func. Genet. **40**,543 (2000)
- http://www.biophysics.org/img/Bloomfield.V-Thermo.pdf

On simple kinetics:

- M. Karplus, J. Phys. Chem. B. **2000**, 104, 11-27

On the Muñoz-Eaton model:

- V. Muñoz *et al*, Nature **390**, 196 (1997).
- V. Muñoz *et al*, PNAS **95**, 5872 (1998).
- V. Muñoz and W. A. Eaton, PNAS **96**, 11311 (1999).
- A. Flammini, J.R. Banavar and A. Maritan, submitted to Europhys. Lett.

On the exact solution:

- P. Bruscolini and A. Pelizzola, submitted to Phys. Rev. Lett (ask the speaker for the preprint)