Wako–Saitô–Muñoz–Eaton Model of Protein Folding

Alessandro Pelizzola, Dipartimento di Fisica, Politecnico di Torino

Pierpaolo Bruscolini (BIFI, Zaragoza)

Alberto Imparato (PoliTO)

Marco Zamparo (PoliTO)

The Model

Protein:

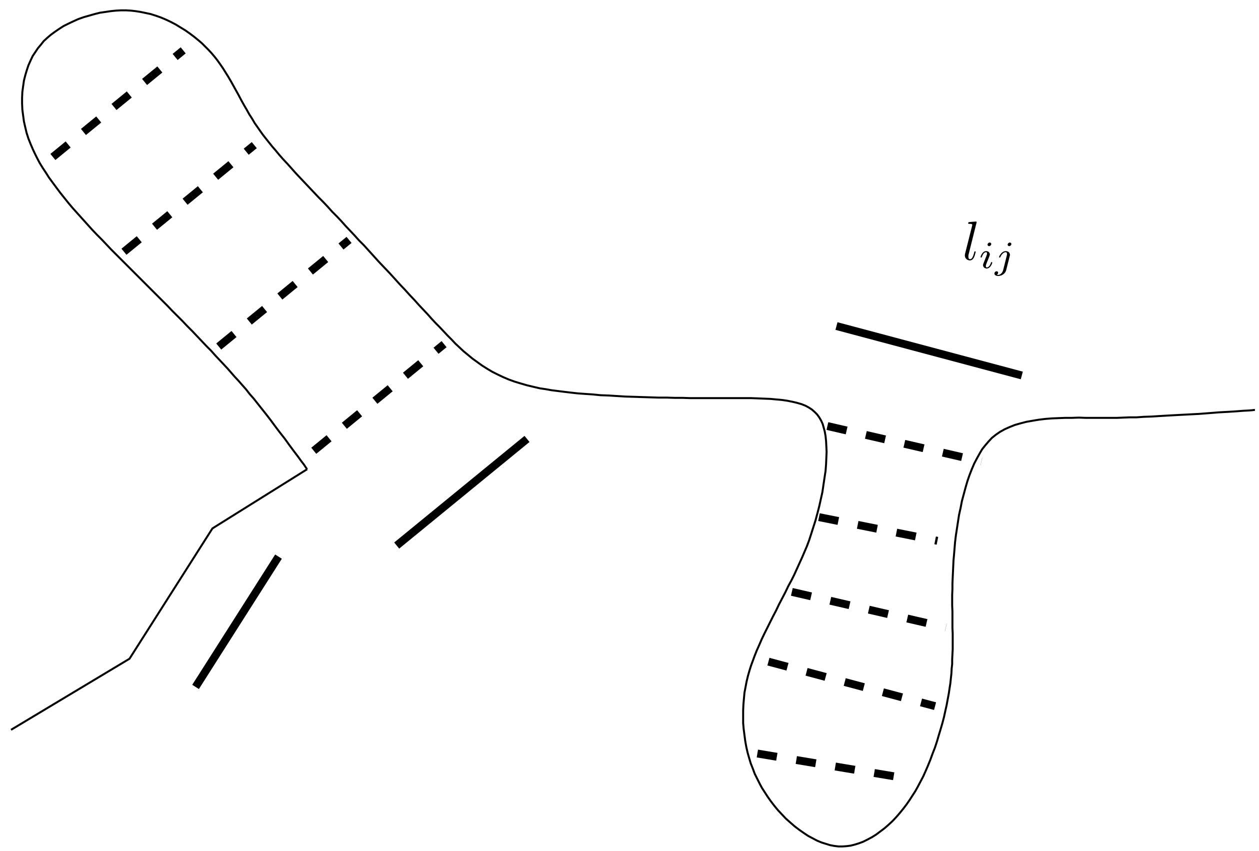
- sequence of aminoacids,
- connected by C–N peptidic bonds,
- with a well–defined low–energy, low–entropy native state,
- whose native structure is thought to be important for the kinetics

The Model

- A binary variable $m_i = 0, 1$ is associated to each peptidic bond $i = 1, \ldots N$.
- $m_i = 1$ for a native peptidic bond, 0 otherwise
- an entropy $q_i > 0$ is associated to each non–native bond
- two aminoacids can interact only if
  - they are in contact in the native state
  - all the peptidic bond between them are ordered (native)
- Effective free energy ("Hamiltonian")

$$H = -\sum_{i<j} \epsilon_{ij} \Delta_{ij} \prod_{k=i}^{j} m_k - T \sum_i q_i (1 - m_i)$$

- 1D model with long–range, many–body interactions


l_{ij}

Previous works on equilibrium

Wako and Saitô (1978):

- PBC, translation invariant
- Exact solution

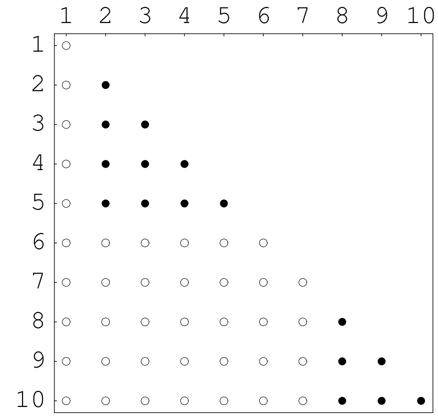
Eaton and co. (1997–1999):

- Single sequence approx $0\cdots01\cdots10\cdots0$
- Double sequence approx $0\cdots01\cdots10\cdots01\cdots10\cdots0$
- Triple sequence approx $\cdots$

Banavar, Flammini and Maritan (2002): (thanks!)

- Mean field
- Ideal α–helix: exact solution
- Ideal β–hairpin: exact solution

Equilibrium: exact solution in the non–homogeneous case



The mapping to the new binary variables $x_{i,j} = \prod_{k=i}^{j} m_k$ yields a 2D model on (a portion of) a square lattice, such that

- the new variables are non–interacting, except for
- the constraints $x_{i,j} = x_{i+1,j} x_{i,j-1}$,
- and the number of states per row is linear in N

[P. Bruscolini and A. Pelizzola, Phys. Rev. Lett. **88**, 258101 (2002)]

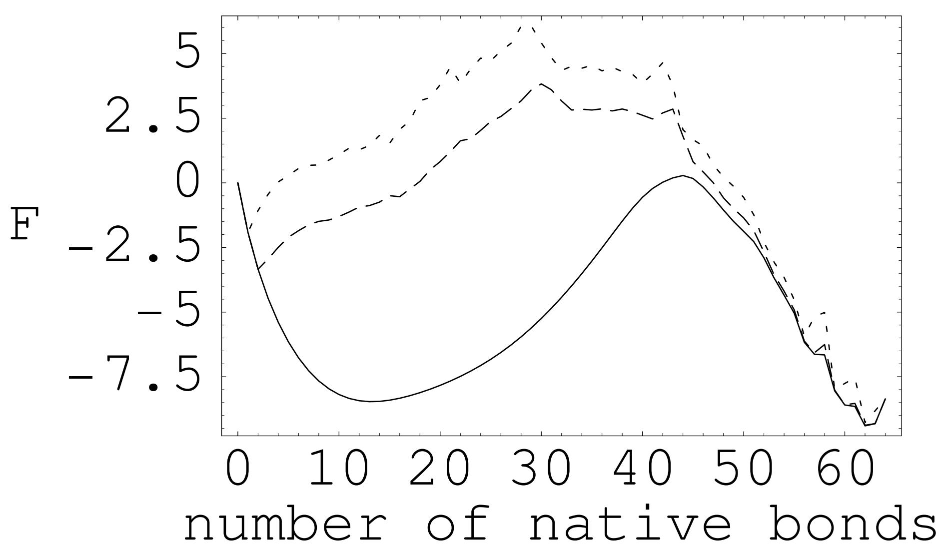
Equilibrium probability factors:

$$p = \prod_{\text{plaqs}} p_{\text{plaq}} \prod_{\text{links}} p_{\text{link}}^{-1} \prod_{\text{sites}} p_{\text{site}}$$

[A. Pelizzola, J. Stat. Mech. P11010 (2005)]

Single sequence, double sequence, . . . approximations severely underestimate entropy of disordered state

CI2: free energy vs number of native bonds at denaturation temperature



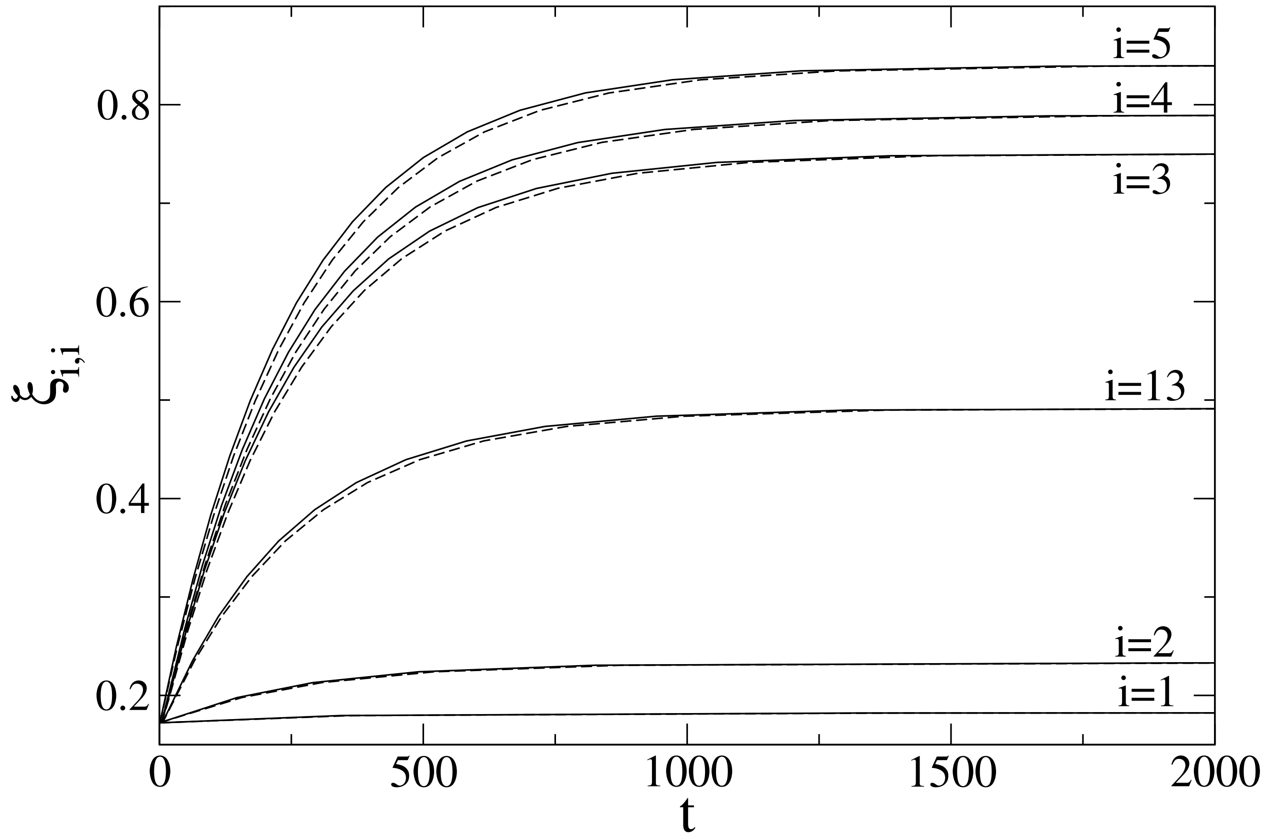
Local Equilibrium Kinetics

Assume factorization of the probability at any time

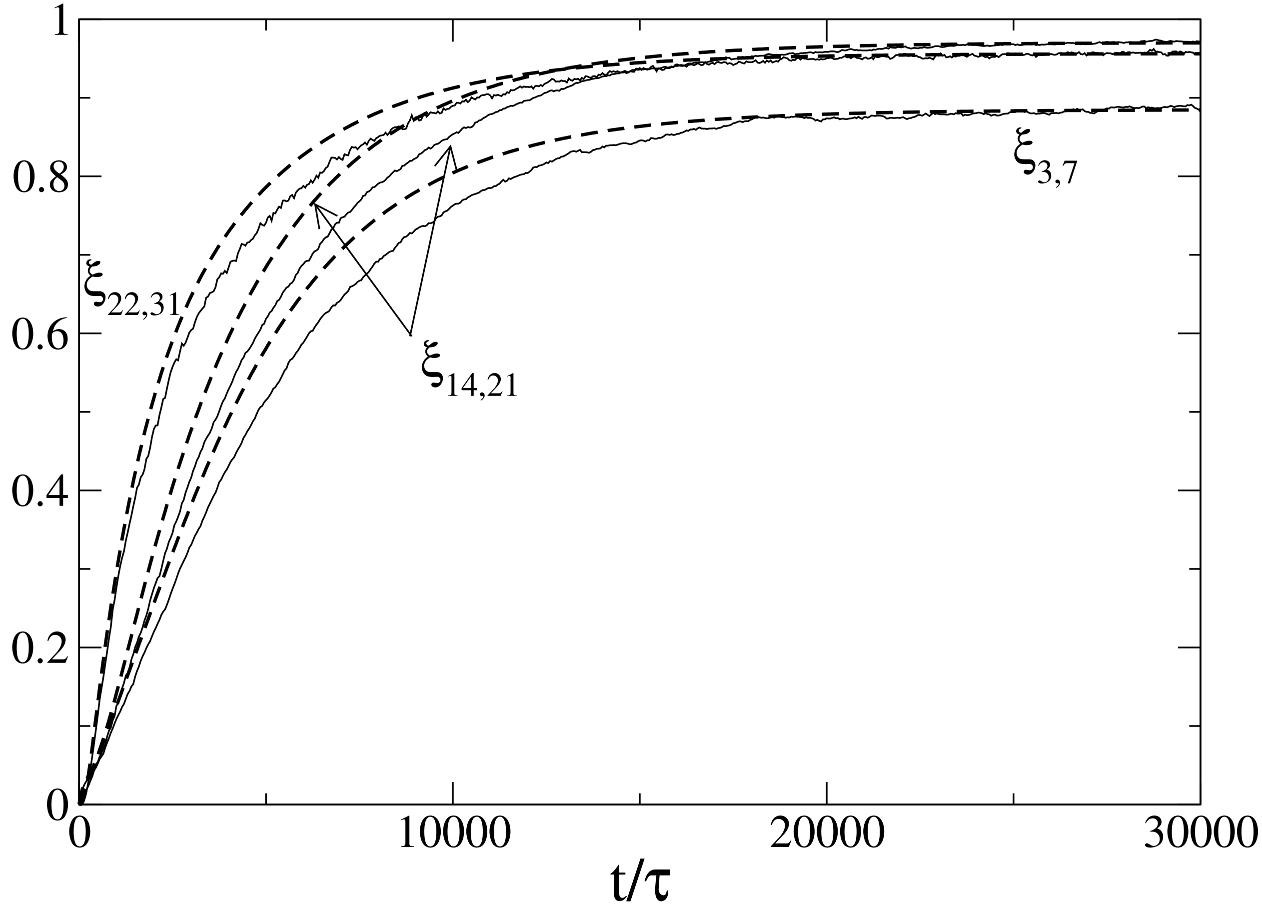
Master equation for local probabilities has polynomial complexity for "simple" kinetics (e.g. single "bond–flip")

- free energy decreases
- exact equilibrium is reached
- equilibration rate is an upper bound of the exact one

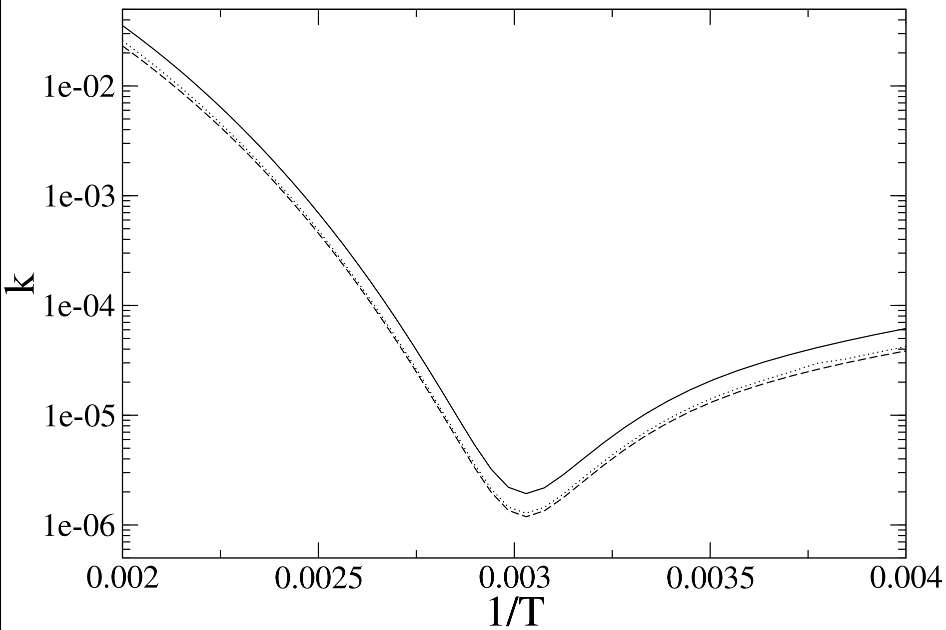
[M. Zamparo and A. Pelizzola, Phys. Rev. Lett. **97**, 068106 (2006)]

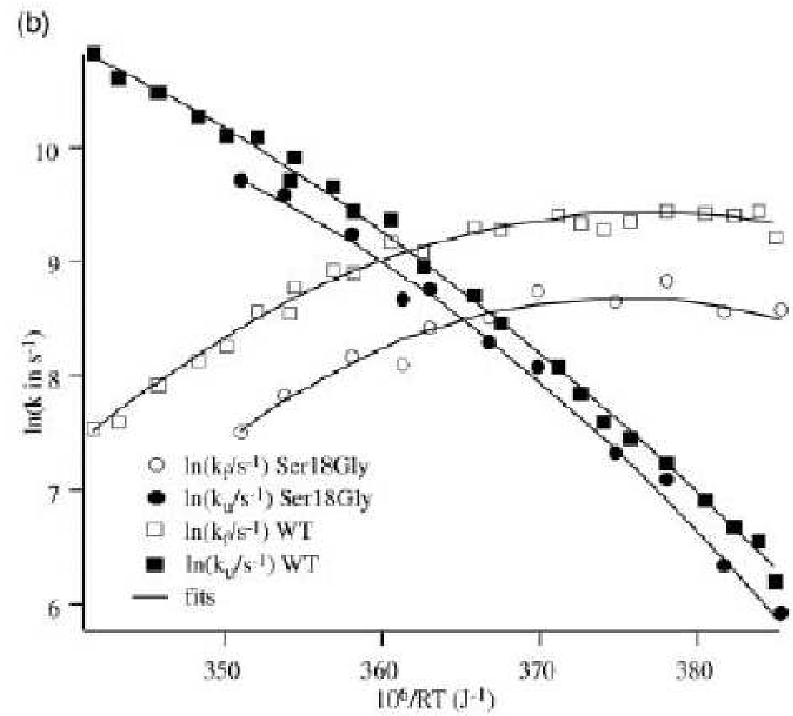


Probability of β-hairpin bond i being native, versus time. Solid lines: our results, dashed line: exact results.

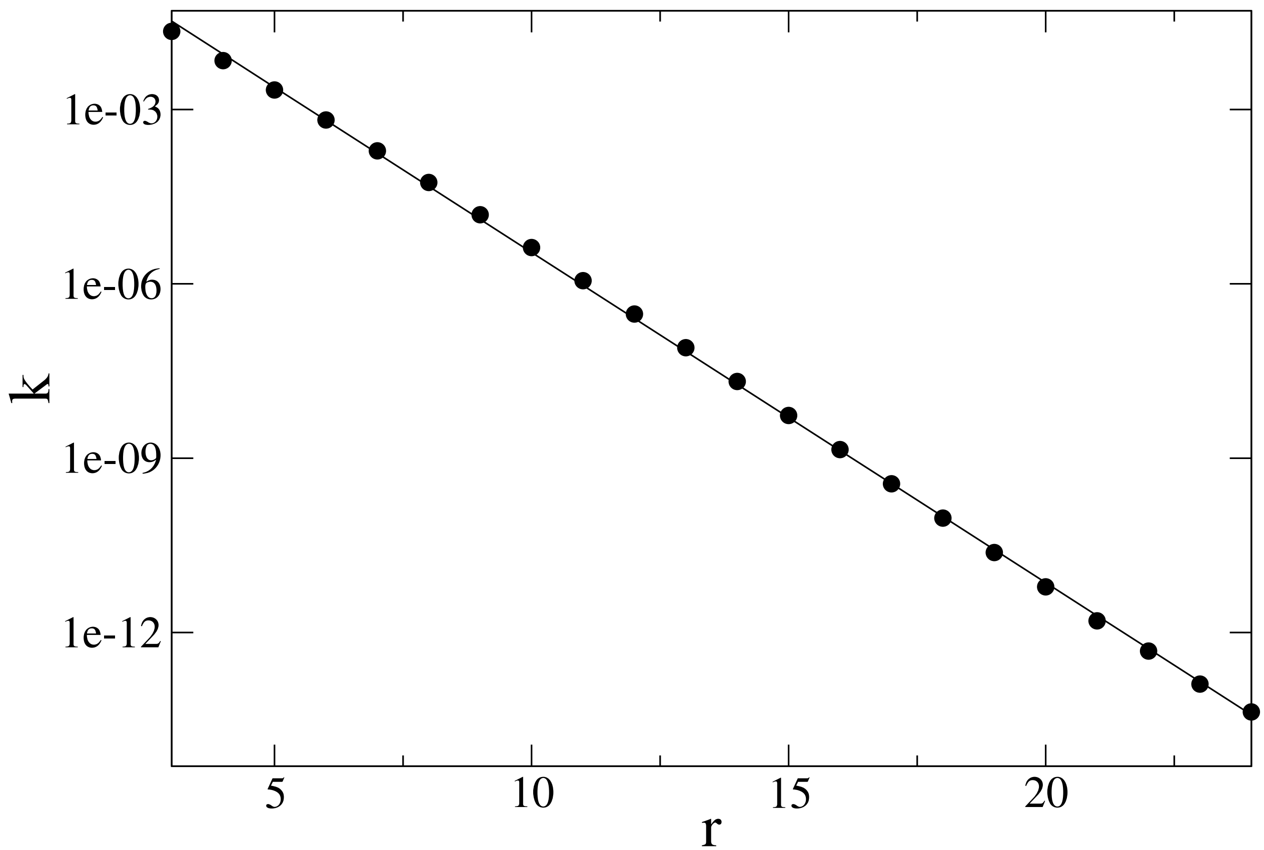


Probability of villin headpiece α–helices being native, versus time. Solid lines: Monte Carlo, dashed line: local equilibrium





PIN1 equilibration rate versus inverse temperature. Solid lines: Glauber kinetics, dashed line: Metropolis kinetics, dotted line: simplifying kinetics.



Equilibration rate for a β–sheet of four strands versus absolute contact order. Dots: our results, line: exponential fit.

Stretching under external force

Recall simplest polymer model (rubber):

$s_i = +1(-1)$ if monomer parallel (antiparallel) to force

$$L = \sum_i s_i \qquad H = -fL \qquad \text{(paramagnet)}$$

Hooke's law

Contraction by heating

PROTEIN $\equiv$ sequence of rigid stretches

For each stretch: native length l_{ij}, orientation $\sigma_{ij} = \pm 1$

$$H(m, \sigma) = H_0(m) - fL(m, \sigma),$$

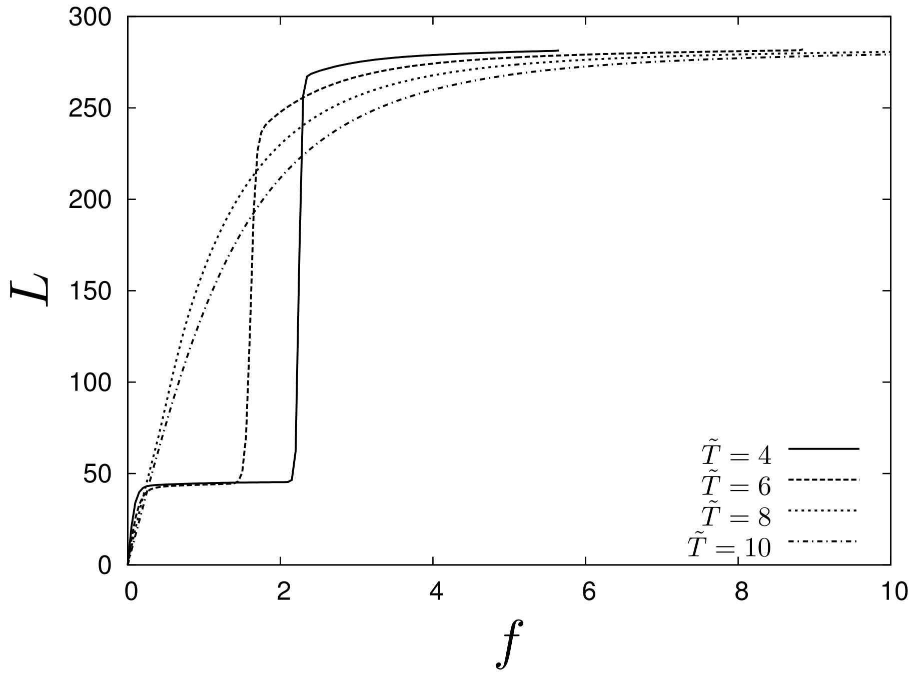
$$\begin{aligned} H_0(m) &= -\sum_{i<j} \epsilon_{ij} \Delta_{ij} \prod_{k=i}^{j} m_k \\ L(m, \sigma) &= \sum_{0 \le i < j \le N+1} l_{ij} \sigma_{ij} (1 - m_i)(1 - m_j) \prod_{k=i+1}^{j-1} m_k \end{aligned}$$

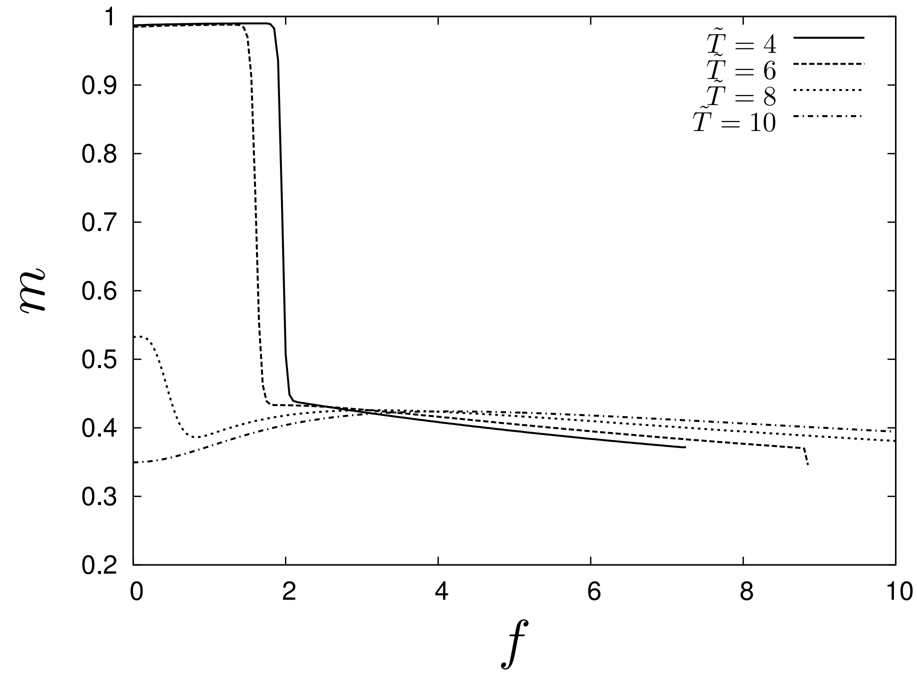
$$\sum_{\{\sigma\}} \exp(-\beta H(m, \sigma)) = \exp(-\beta H_{eff}(m))$$

can be performed analytically $\Rightarrow H_{eff}$ still exactly solvable

For zero force: H_{eff} = WSME model with $q_i = \ln 2$

1TIT: 27th immunoglobulin domain of titin



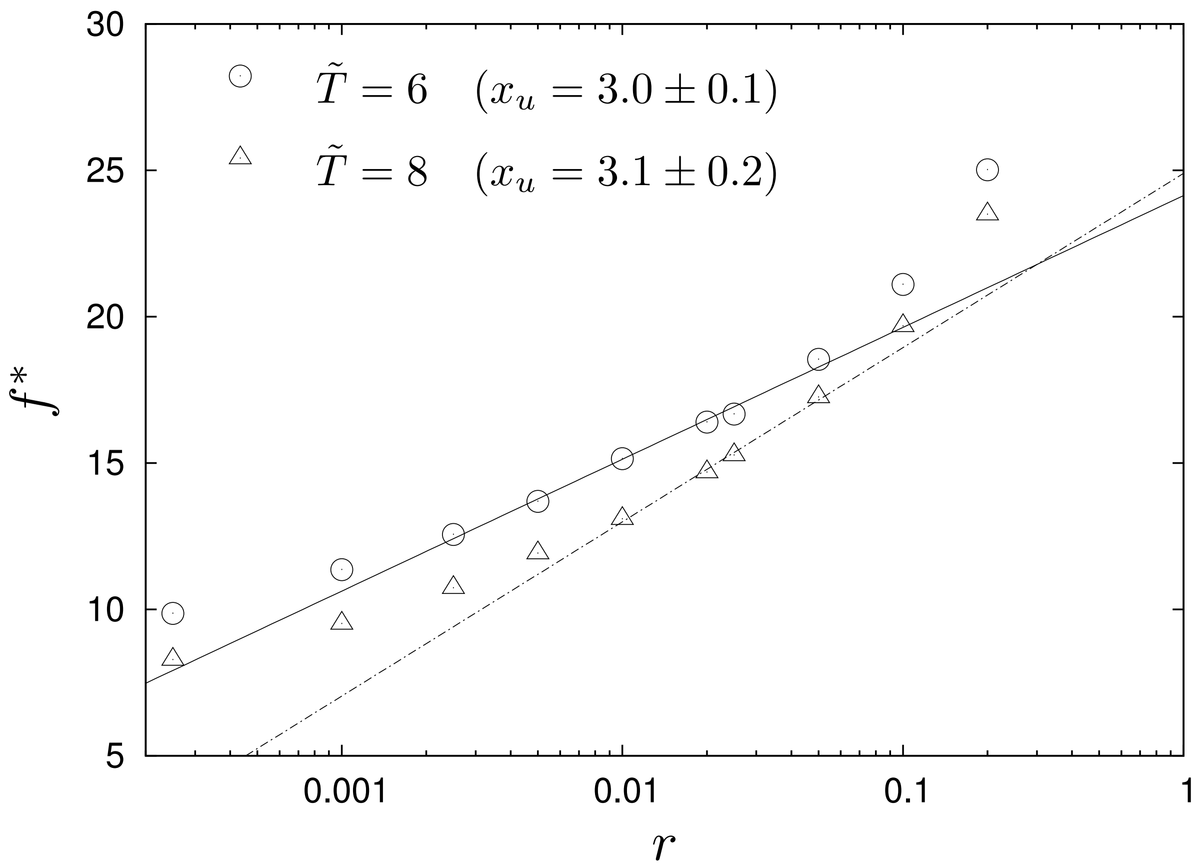


1TIT under dynamic loading: $f(t) = rt$

$$f^* = \frac{k_B T}{x_u} \ln\left[\frac{r x_u \tau_0}{k_B T}\right]$$

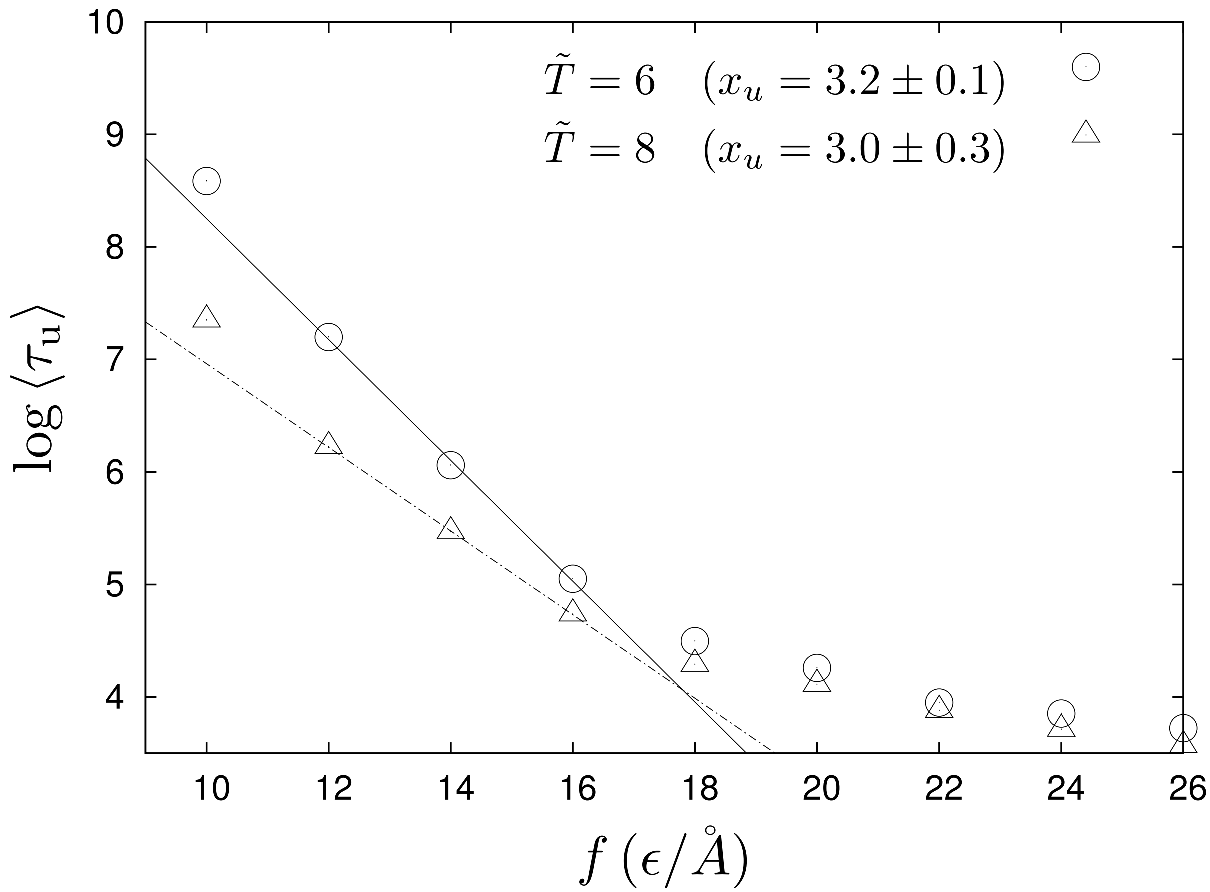
[E. Evans, 2001; D. Derényi et al, 2004]

x_u = temperature independent unfolding length



1TIT under force clamp

$$\langle \tau_{\mathrm{u}} \rangle = \frac{\exp\left[\beta\left(\Delta E_{\mathrm{u}} - f x_{\mathrm{u}}\right)\right]}{\omega_0},$$

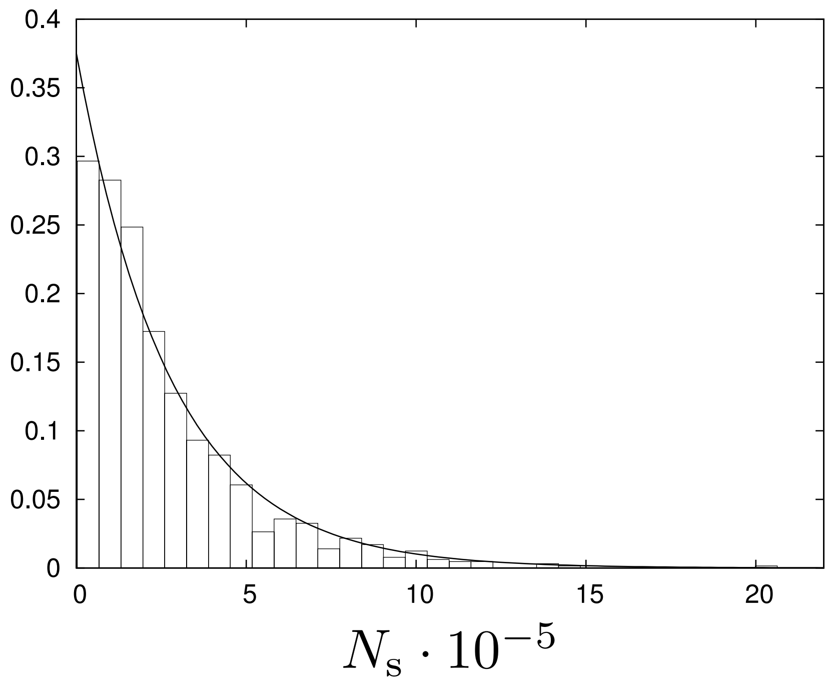


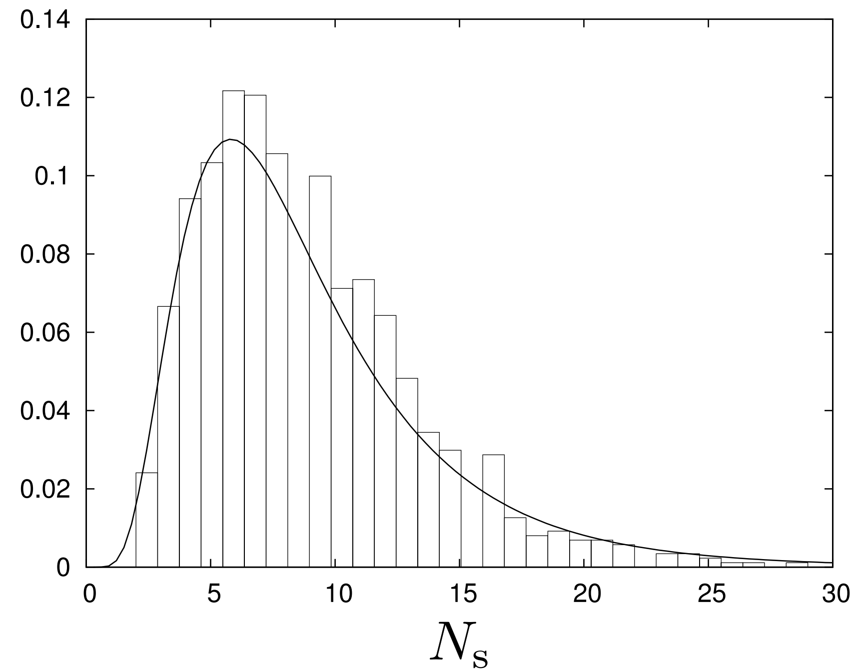
1BBL under force clamp: distribution of unfolding times

Small force: exponential

Large force: lognormal

[cfr. MD by Szymczak and Cieplak, 2006]

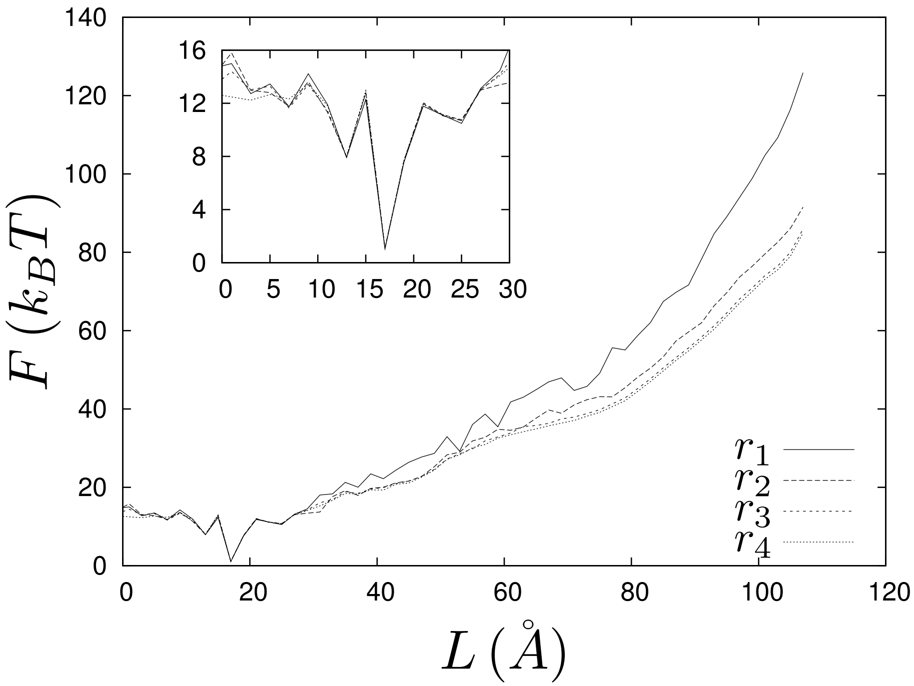




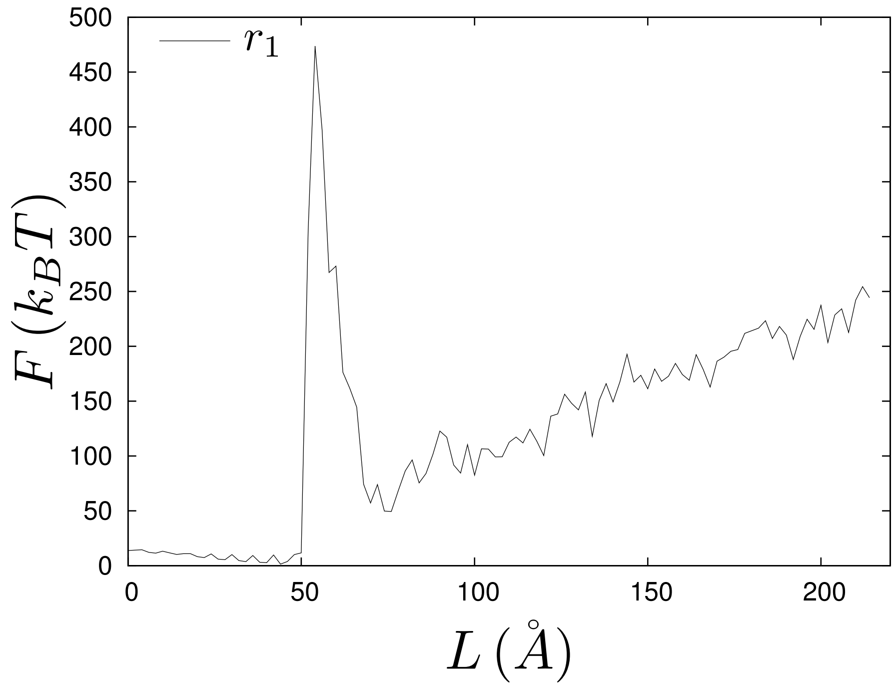
Energy landscapes

Jarzynski equality + weighted histograms $\Rightarrow F(L)$ [Imparato and Peliti, 2006]

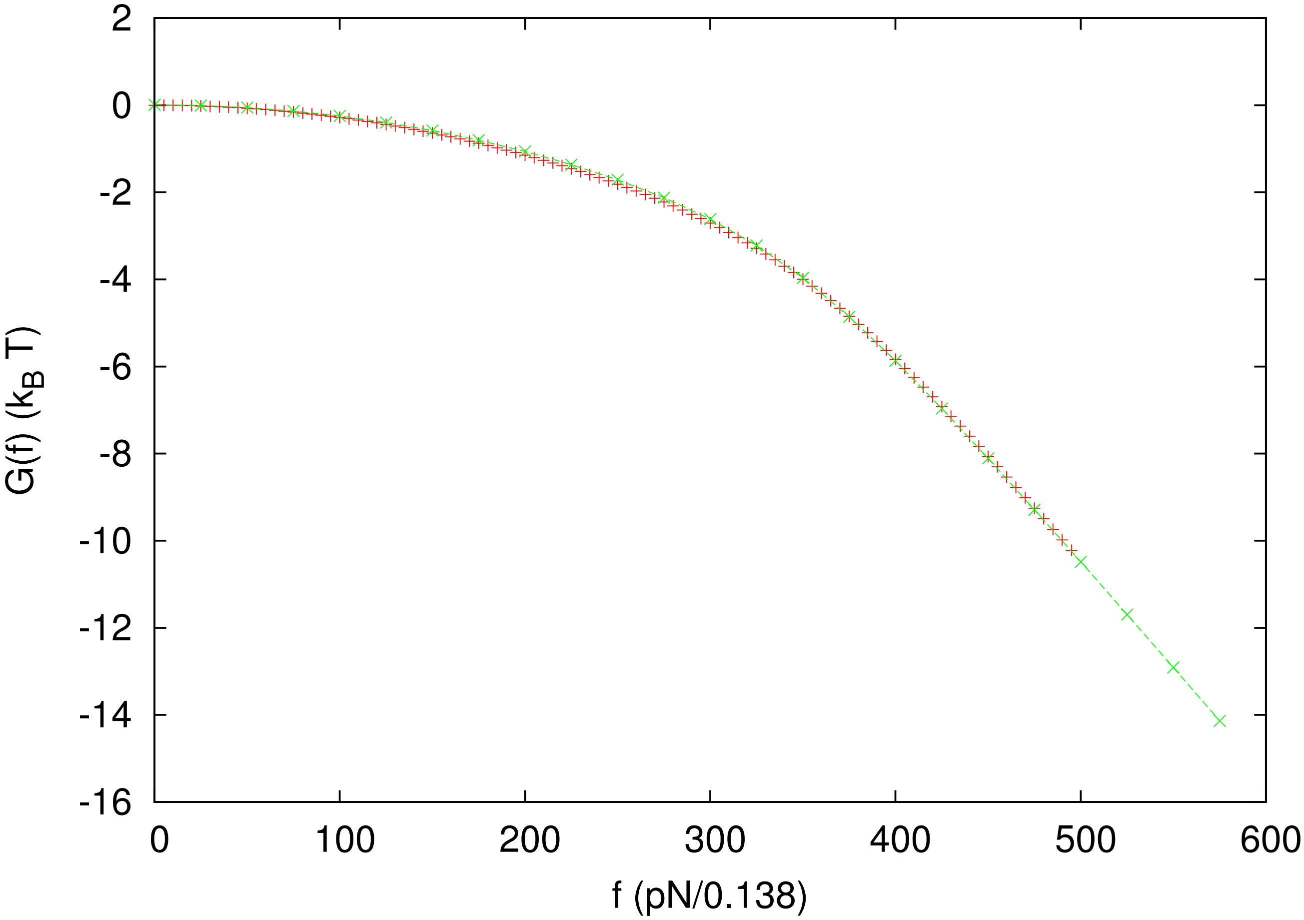
1BBL



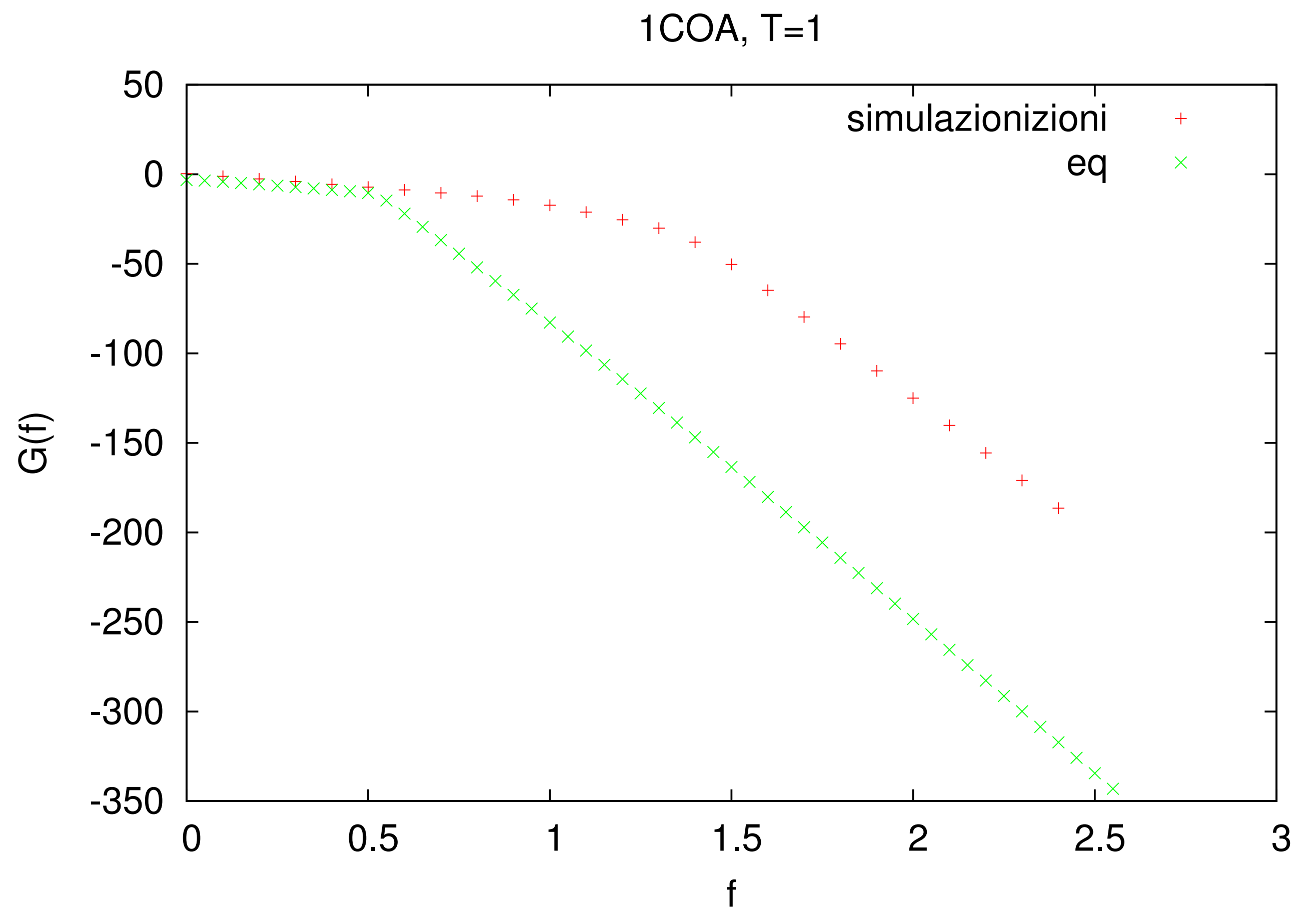
1TIT



Check: $\exp[-\beta G(f)] = \sum_L \exp(\beta f L) \exp[-\beta F(L)]$



G(f) (k_B T)
f (pN/0.138)
2
0
-2
-4
-6
-8
-10
-12
-14
-16
0
100
200
300
400
500
600



1COA, T=1
simulazionizioni
eq
G(f)
f
50
0
-50
-100
-150
-200
-250
-300
-350
0
0.5
1
1.5
2
2.5
3

CONCLUSIONS

Highly simplified model yields:

- Exact equilibrium
- "Quasi–exact" kinetics
- Two–state, single exponential kinetics
- Rates, relation to contact order
- Theoretical vs experimental rates, Φ–values, pathways, transition state ensembles (TO DO!)
- Stretching