Reconstructing the free energy landscape of a mechanically unfolded model protein

Stefano Luccioli

Università di Firenze

Alessandro Torcini

(ISC - CNR Firenze)

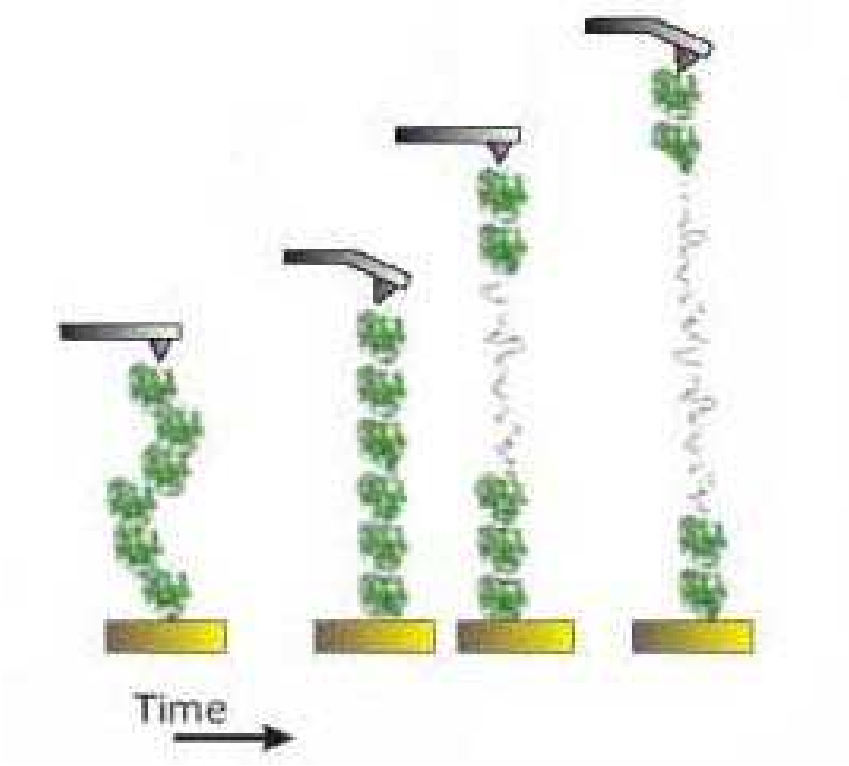
Alberto Imparato

(Politecnico di Torino)

Introduction

Reconstruction of the (equilibrium) free energy landscape of a protein from out-of-equilibrium mechanically unfolded configurations (Atomic Force Microscope) via two methods:

- Extended Jarzynski Equality
- thermodynamical averages over Inherent Structures (IS) of the protein
  ISs $\equiv$ local minima of the potential energy



Summary

- Jarzynski Equality (JE)
- Extended Jarzynski Equality (EJE)
- Protein model and thermodynamic features (T_θ,T_f,T_g)
- Pulling protocol
- EJE reconstruction
- Inherent structures (ISs)
- EJE versus ISs reconstruction
- Conclusions and perspectives

To be read:

- D.J. Wales, *Energy Landscapes* (Cambridge Univ. Press, 2003);
- C. Jarzynski *Phys. Rev. Lett.* 78, 2690 (1997)

Reversible Work

The system is described by the Hamiltonian $H(x,\mu)$, where x defines the state of the system and μ is an external parameter that can be manipulated.

- Within the canonical ensemble the equilibrium state is described by the Gibbs distribution $p^{eq}_{\mu}(x) = \frac{\mathrm{e}^{-H(x,\mu)\beta}}{Z_{\mu}}$
- The partition fuction is $Z_{\mu} = \int dx \mathrm{e}^{-H(x,\mu)\beta}$
  The free energy reads as $F_{\mu} = -\log Z_{\mu}/\beta$
- The derivative of F_{μ} with respect to the parameter μ gives:

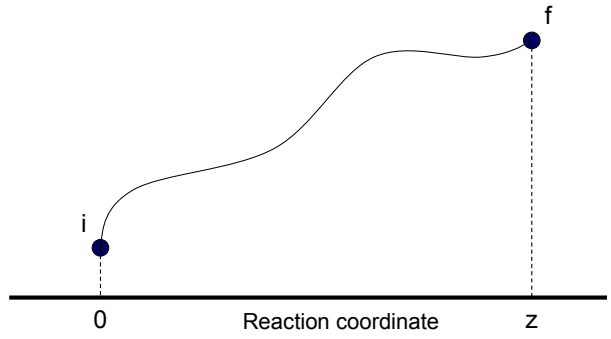
$$\frac{\partial F_{\mu}}{\partial \mu} = \int dx \; p^{eq}_{\mu}(x) \; \frac{\partial H}{\partial \mu} = \langle \frac{\partial H}{\partial \mu} \rangle_{\mu}$$

  where $< \cdot >_{\mu}$ is the average done within the canonical ensemble
- A finite variation of the parameter induces the following variation of F_{μ}
  $\Delta F = F_{\mu} - F_0 = \int_0^{\mu} d\mu' \; \langle \frac{\partial H}{\partial \mu} \rangle_{\mu'} \equiv W_{rev}$
- At equilibrium the reversible work done on the system is equal to the free energy variation $\Delta F = W_{rev}$
  W_{rev} does not fluctuate, since it is an equilibrium average of an observable. [a]

[a] Tolman, The principles of statistical mechanics (Oxford, 1938)– Imparato & Peliti (cond-mat/0706.1134)

Jarzynski equality (JE)

Jarzynski equality [a] relates the work done on the system during an out-of-equilibrium process to the difference of equilibrium free energy.



[a] C. Jarzynski *Phys. Rev. Lett.* 78, 2690 (1997)

(1) $$\langle e^{-\beta W_{if}} \rangle_{t_f} = e^{-\beta(F_{z(t_f)} - F_0)} \qquad \beta = 1/kT, \qquad F_0 = F_{z(0)}$$

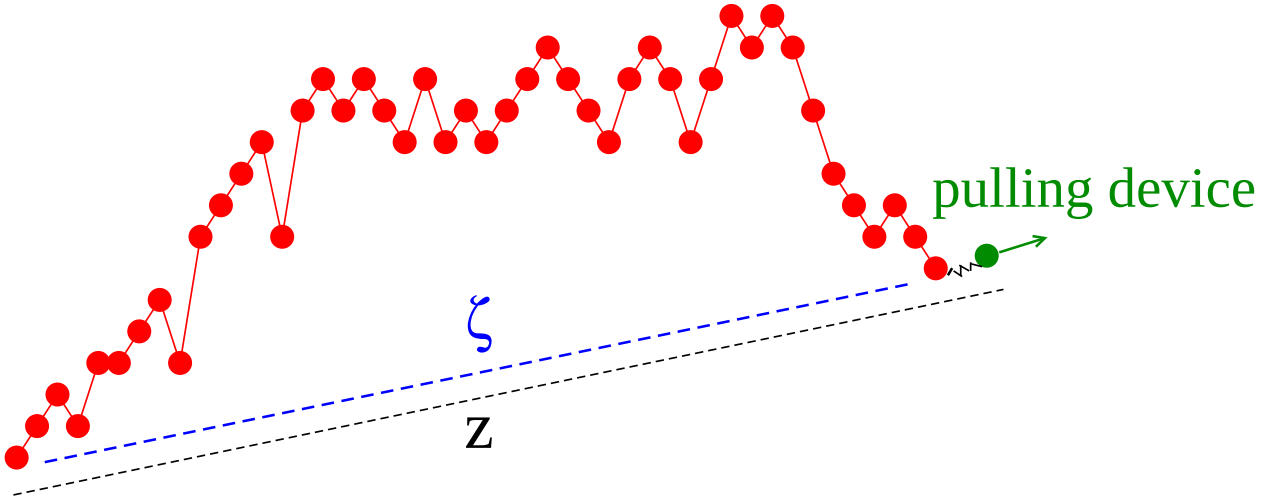
where:

- $\langle \ldots \rangle_{t_f} \rightarrow$ average over repetitions of the same experiment (protocol)
- Initial and final equilibrium states
- $W_{if} \rightarrow$ work done on the system - W_{if} fluctuates due to thermal fluctuations
  $W_{if} = \int dW = \int_{t=0}^{t=t_f} dt \ \dot{z} \ \frac{\partial H(x(t),z)}{\partial z}$
- $z \rightarrow$ externally controlled manipulation parameter (position of the pulling device)
- $z = z(t) \qquad t \in [0, t_f]$ manipulation protocol

Problem $\rightarrow$ JE gives the equilibrium F-profile + the pulling device as a function of z

Extended Jarzynski equality (EJE)

stretching of a polypeptidic chain



(2) $$\langle \delta(\zeta - \zeta(x)) e^{-\beta[W - U_{z(t)}(\zeta)]} \rangle_t = e^{-\beta(F(\zeta) - F_0)}$$

where: [a]

- $F(\zeta) = -kT \ln \int dx \delta(\zeta - \zeta(x)) e^{-\beta H(x)}$
- $U_{z(t)}(\zeta) = c[z(t) - \zeta]^2/2 \rightarrow$ coupling energy between device and protein
- ζ=end-to-end-distance $\rightarrow$ internal collective coordinate
- $z \rightarrow$ distance between the first bead and the pulling device

[a] G. Hummer and A. Szabo, *PNAS* **98**, 3658 (2001), A. Imparato and L. Peliti, *J. Stat. Mech.* 03005 (2006)

The protein model (I)

The simplified model assumes that the aminoacids (the residues) are represented by the C_α positioned along a one dimensional chain and the aminoacids are of three types only : B=hydrophobic, P=polar , N=neutral
The simplified interactions are :

- a stiff nearest-neighbour harmonic potential intended to maintain the bond distance almost constant : V^{harm};
- a three body interactions which accounts for the bond angles : V^{ang} ($\theta_0 = 105$);
- a four-body potential corresponding to the dihedral terms and responsible for the formation of secondary structures V^{dih} (in this case β-sheets are favourite);
- a long-range Lennard-Jones potential reproducing in an effective way the presence of the solvent V^{LJ} (hydrophobic and hydrophilic mediated interactions among residues);

This simple model has been widely studied in the last 17 years, because it reproduces some general feature of protein folding, in particular depending on the aminoacid sequence bad or good folders are observables, moreover it can lead to the formation of different secondary structures (α-elices or β-sheets).
This model with the parameter here studied favourites the formation of four stranded β-barrel native configurations.

The protein model (II)

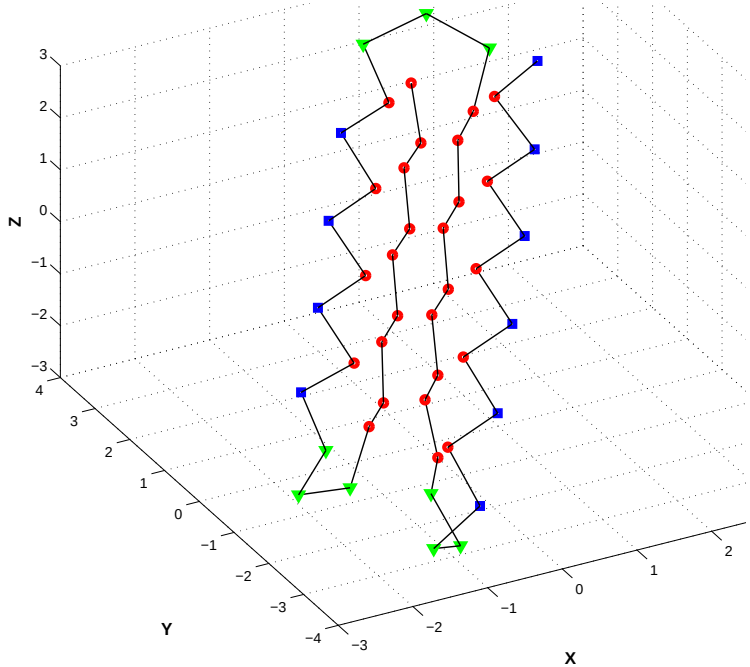
Model BPN (B=hydrophobic, P=polar, N=neutral) $N = 46$

Sequence: $B_9 N_3 (PB)_4 N_3 B_9 N_3 (PB)_5 P$

- Intramolecular potential [a]:

(3) $$V = \sum_{i=1}^{N-1} V_i^{harm} + \sum_{i=2}^{N-1} V_i^{ang} + \sum_{i=2}^{N-2} V_i^{dih} + \sum_{i=1}^{N-3} \sum_{j=i+3}^{N} V_{ij}^{LJ}$$

$$V_i^{harm} = \alpha (r_{i,i+1} - \sigma)^2$$
$$V_i^{ang} = A \cos(\theta_i) + B \cos(2\theta_i)$$
$$V_i^{dih} = A_i [1 + \cos(\phi_i)] + B_i [1 + \cos(3\phi_i)]$$
$$V_{ij}^{LJ} = C_{ij} [(\frac{\sigma}{r_{ij}})^{12} - D_{ij} (\frac{\sigma}{r_{ij}})^6)]$$



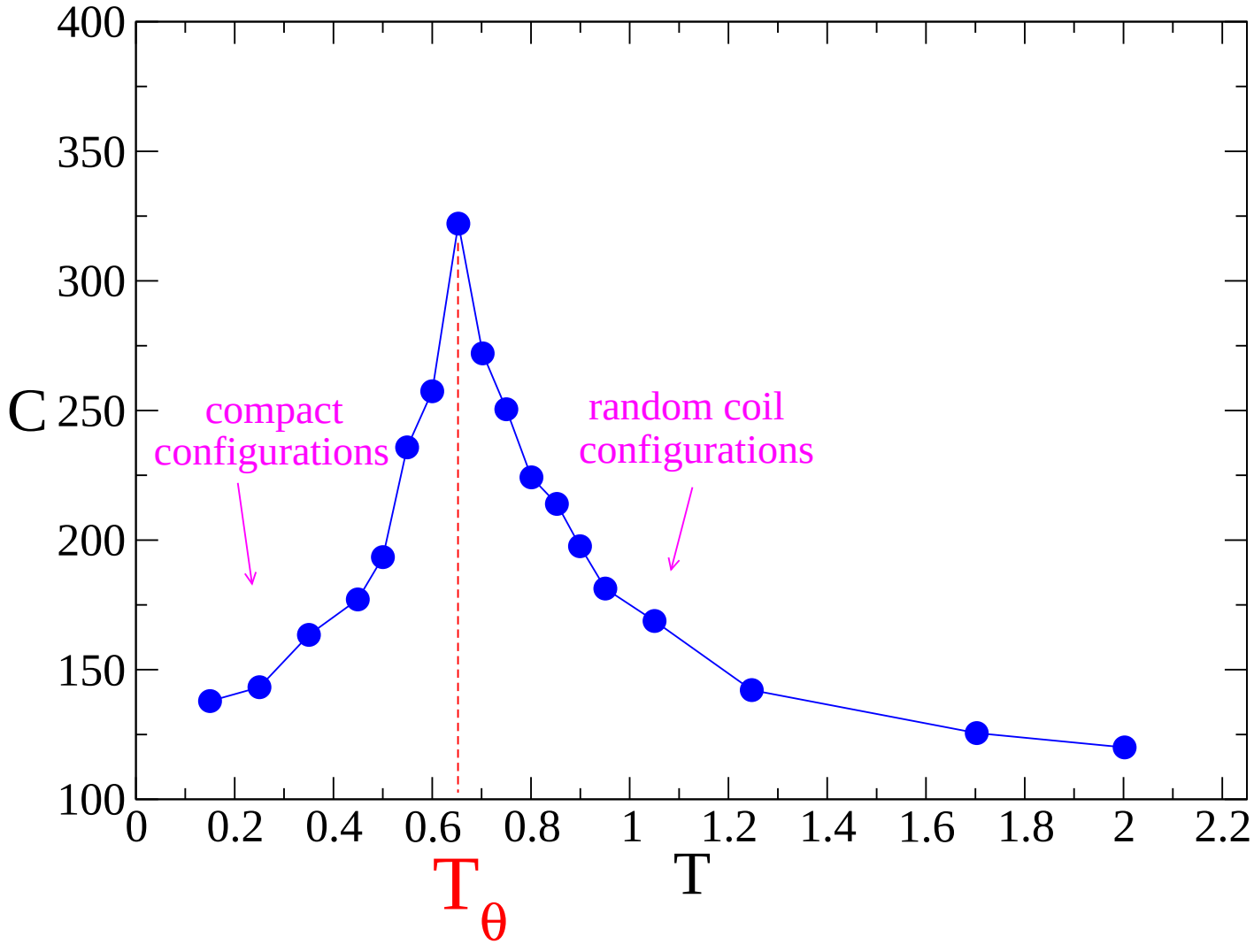
- Global minimum of $V \rightarrow$ native configuration

- Langevin dynamics:

(4) $$m\ddot{\mathbf{r}}_i = -\nabla V - \gamma \dot{\mathbf{r}}_i + \eta(t) \qquad i = 1, N$$

[a] J.D. Honeycutt and D. Thirumalai, *PNAS* **87**, 3526 (1990), R.S. Berry *et al*, *PNAS* **94**, 9520 (1999)

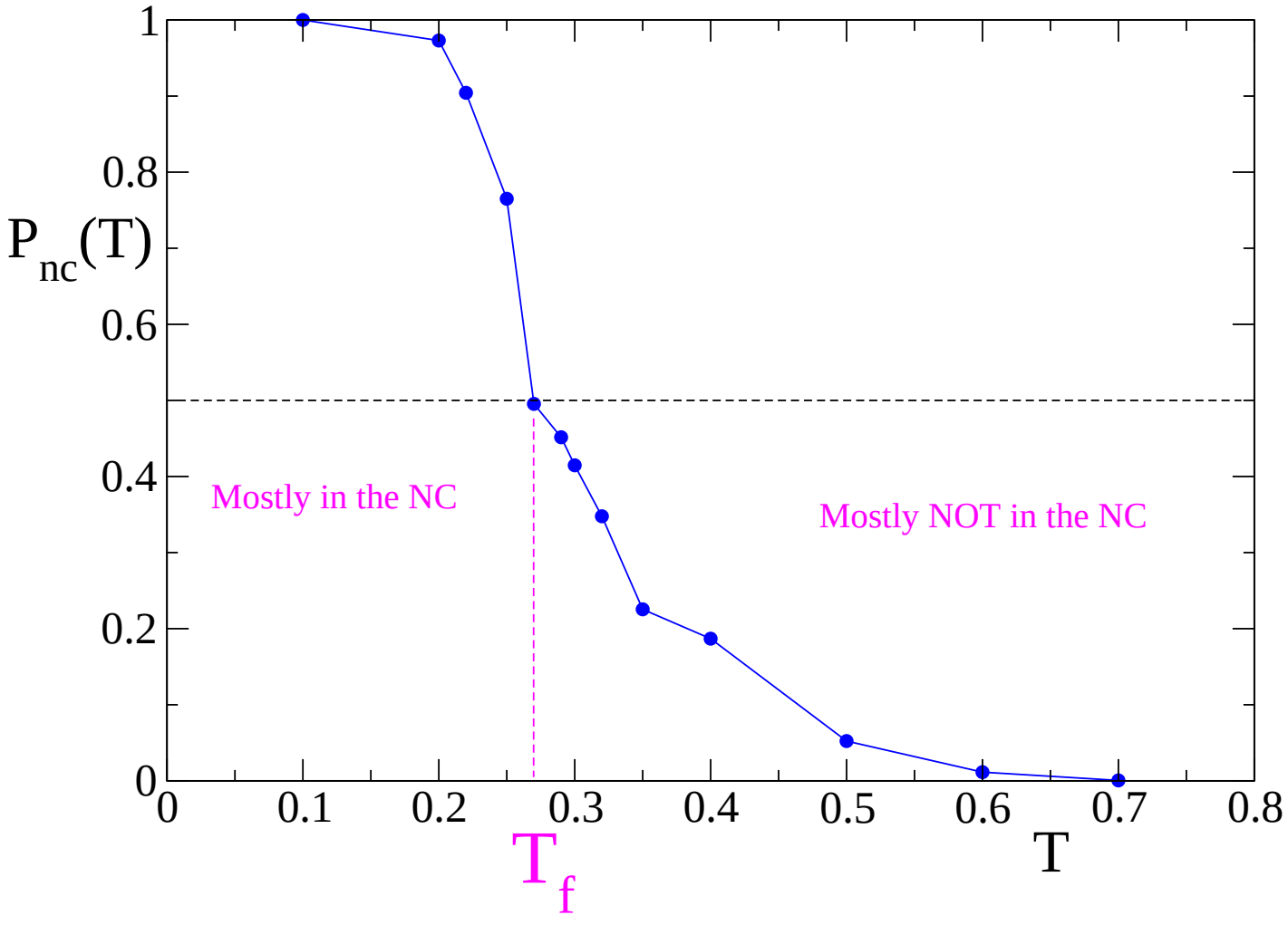
Hydrophobic collapse temperature



$C(T_\theta) = C^{max} \rightarrow T_\theta = 0.65(1)$ where $C(T) = \frac{\langle E^2 \rangle - \langle E \rangle^2}{T^2}$ [a]

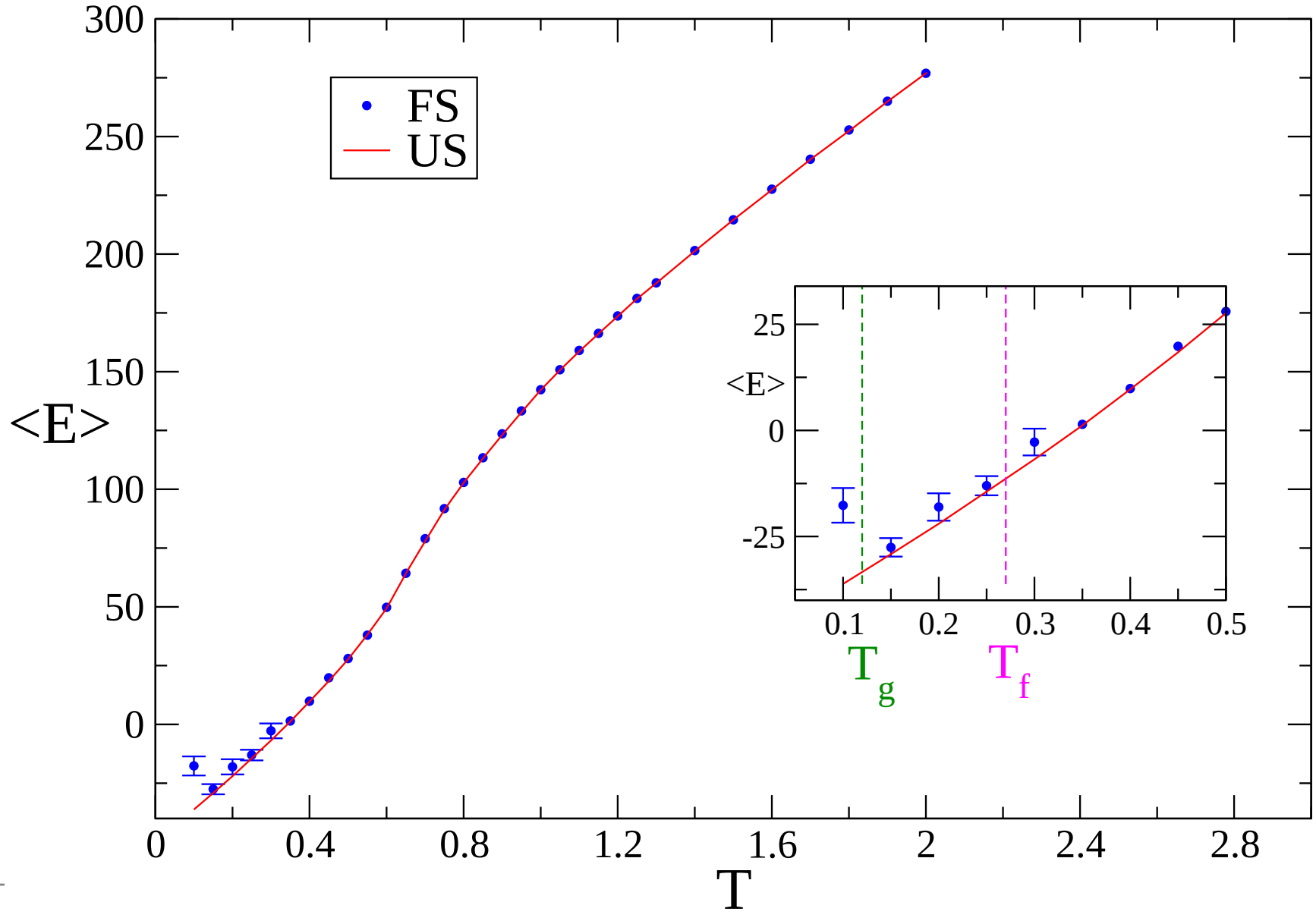
[a] De Gennes, Scaling concepts in polymer physics (1979),

Folding temperature

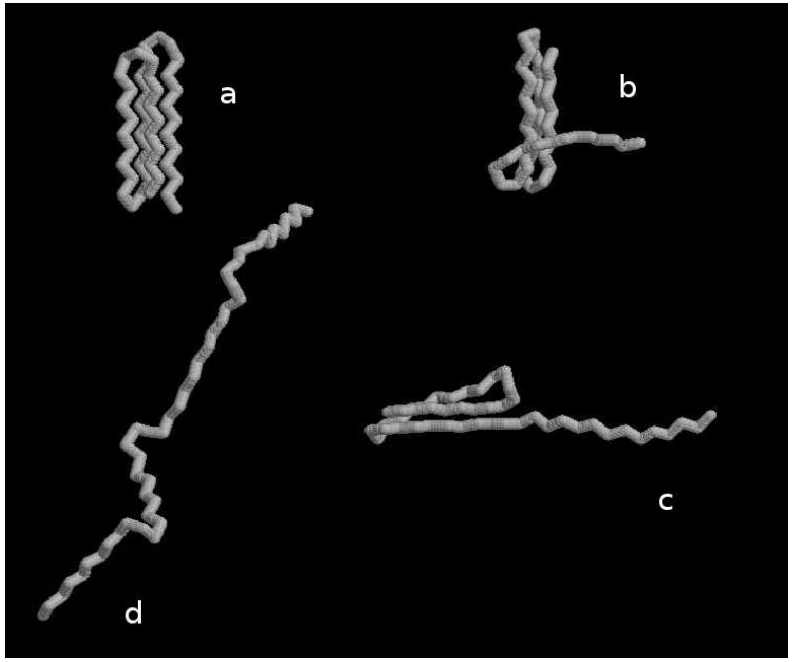


$P_{nc}(T_f) = 0.5 \rightarrow T_f = 0.27(1)$

Glassy temperature



Pulling protocol



- the first bead is kept fixed and the last is attached to the pulling device (a stiff spring) moving along a fixed direction with the law: [a]

  $z(t) = z(0) + v_p t \qquad t \in [0, t_f]$ linear protocol (v_p=constant velocity)

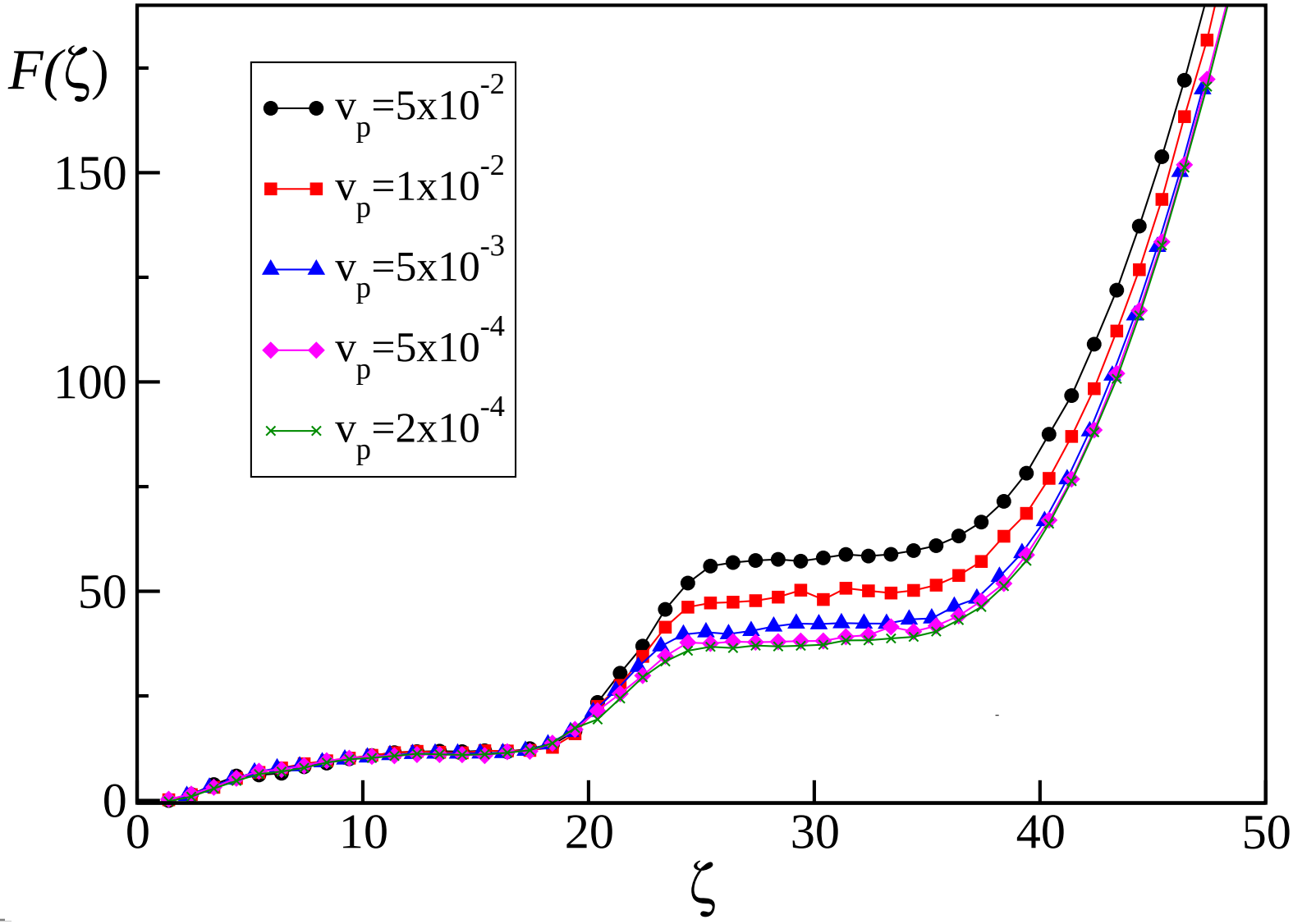
  $U(\xi) = k/2(z(t) - \xi(t))^2$ external potential

- forced unfolding performed at constant temperature via Langevin dynamics

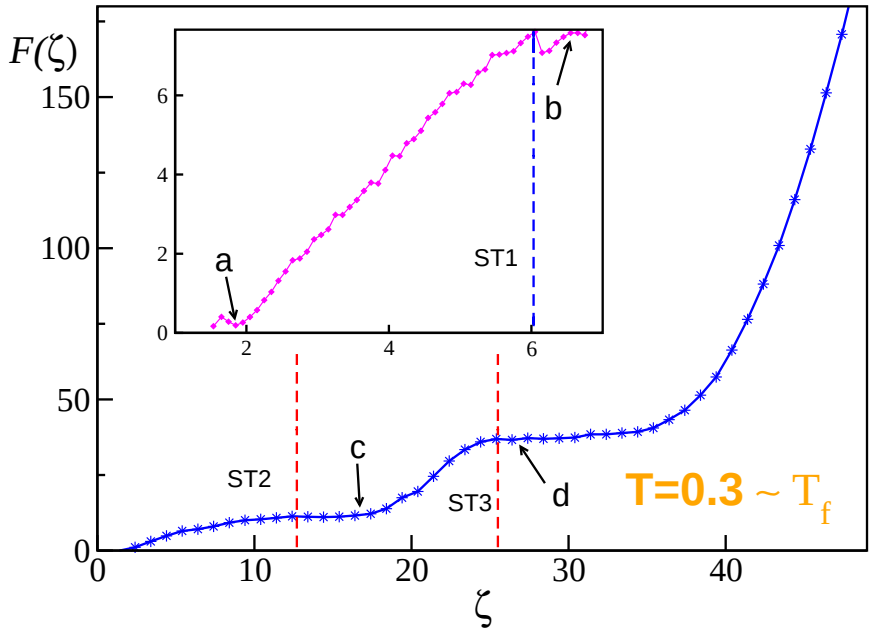
[a] Analogous to experimental setups N.C. Harris *et al.* PRL (2007)

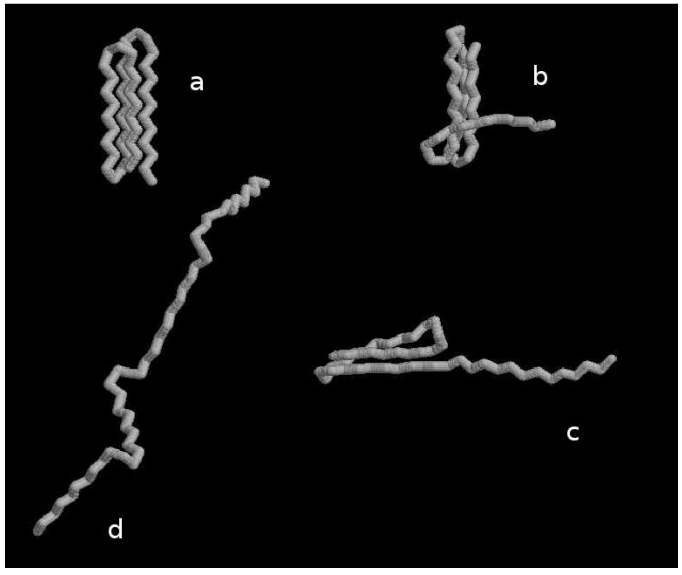
EJE: asymptotic reconstruction

T=0.3



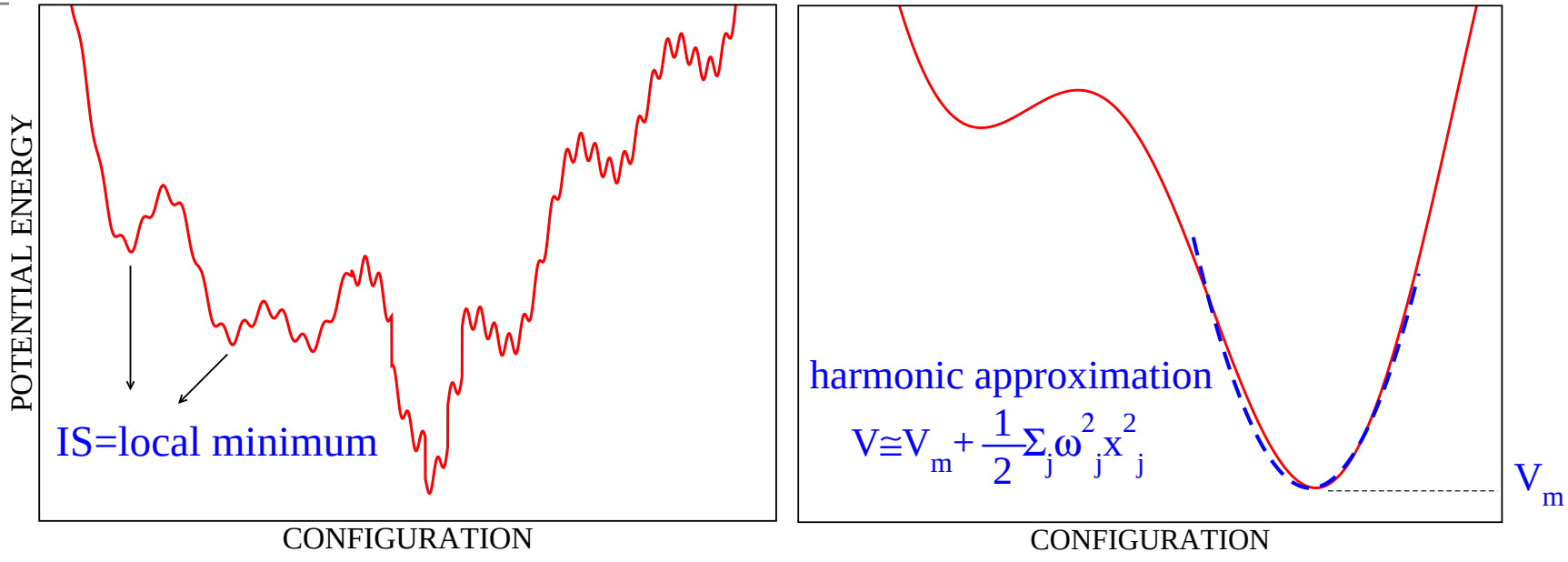
EJE reconstruction





- ST1 $\longrightarrow$ Escape form the native valley, below $\zeta \sim 6$ the configurations are similar to the NC
- ST2 corresponds to pull completely out of the β-barrel the last strand (i.e. $(PB)_5P$), the plateau $13 < \zeta < 18.5$ is due to the stretching of the strand (no work done);
- ST3 is associated to the complete destabilization of the core of the protein induced by pulling out the third strand, the plateau is associated to configurations similar to (d).
- The final quadratic rise corresponds to the stretching of bond angles and distances beyond their equilibrium values ($\zeta > \zeta_{trans}$)

Inherent structures (ISs)



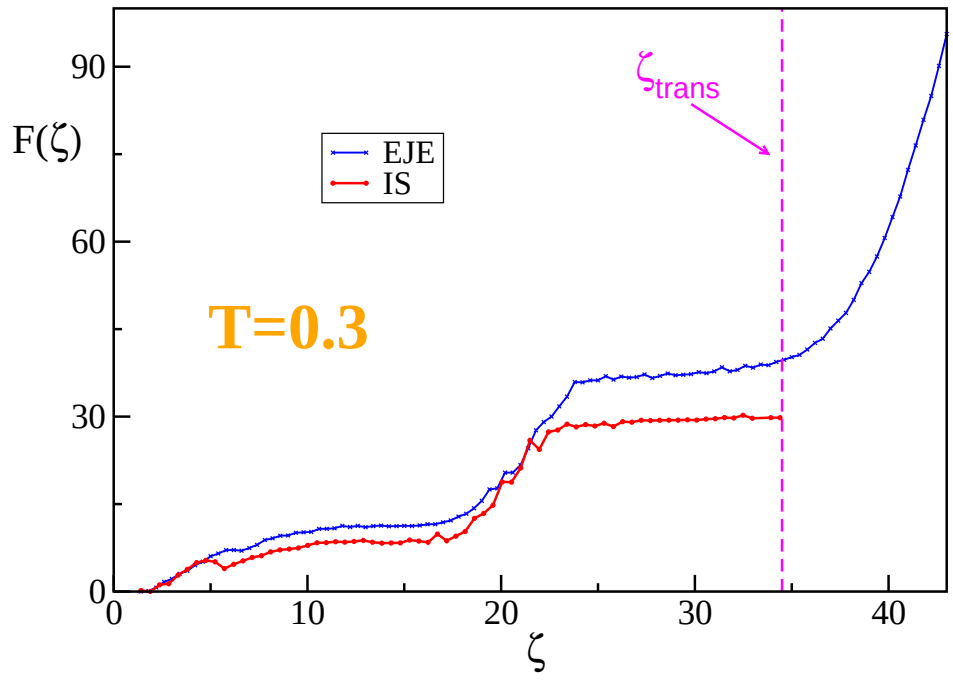
Within the IS formalism and assuming harmonic basins of attraction: [a]

(5) $$\mathrm{e}^{-\beta F_{IS}} = Z_{IS} = \sum_m \mathrm{e}^{-\beta(V_m + W_m)} \propto \sum_m \mathrm{e}^{-\beta V_m} \prod_{j=1}^{3N-6} (k_B T/\omega_m^j) \qquad \text{where :}$$

- V_m (resp. W_m) → potential (resp. vibrational free) energy of the IS;
- $\{\omega_m^j\}$ → frequencies of the vibrational modes.

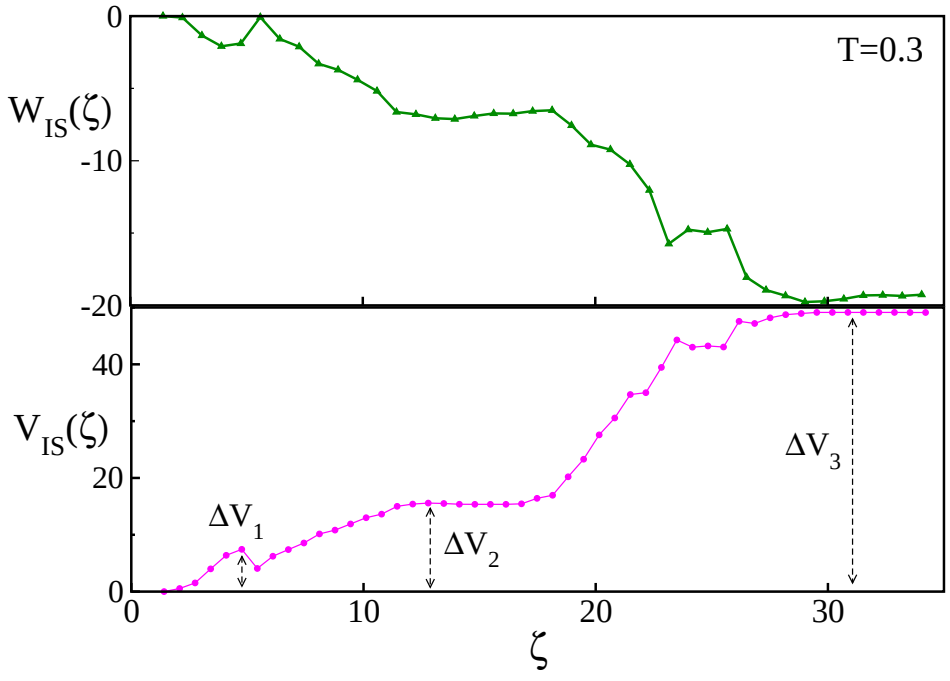
[a] Wales, *Energy Landscapes* (2003); Nakagawa & Peyrard, PNAS (2006)

EJE versus ISs reconstruction



- Good agreement up to $\zeta \sim 20$, up to this end-to-end distance the protein unfolds along the funnel jumping from one minima to another ;
- At higher ζ the underestimation given by the IS reconstruction should be noticeably reduced by including also the saddles in the in the IS analysis.
- $\zeta > \zeta_{trans}$ no more minima in the landscape, only saddles

Energetic and entropic barriers



(6)
$$V_{IS}(\zeta) = \sum_m{}' V_m \mathrm{e}^{-\beta(V_m+W_m)}/Z_{IS}(\zeta)$$

where $\sum'_m$ is limited to IS with end-to-end distance within $[\zeta, \zeta + \delta\zeta]$.

ΔV_i=energetic barrier transition temperature $\rightarrow T_t^i = \frac{2\Delta V_i}{3N}$ $i = 1, 2, 3$

$$T_t^1 = 0.11(1) \sim T_g \qquad T_t^2 = 0.23(2) \sim T_f \qquad T_t^3 = 0.72(1) \sim T_\theta$$

Conclusions and perspectives

The equilibrium free energy landscape for a good folder sequence has been reconstructed as a function of an internal coordinate of the system (the end-to-end distance ζ) via two independent methods [a]

- the agreement between the IS and the EJE reconstruction suggests that the two methodologies are consistent and able to reproduce equilibrium properties of the model;
- the structural transitions induced by pulling can be related to thermodynamical aspects of folding, thus indicating that ζ is a good reaction coordinate at least for this model protein;

Recent pubblication of the first experimental free energy reconstruction using the EJE for a Titin I27 domain: N.C. Harris *et al.* PRL (2007)

Future plans:

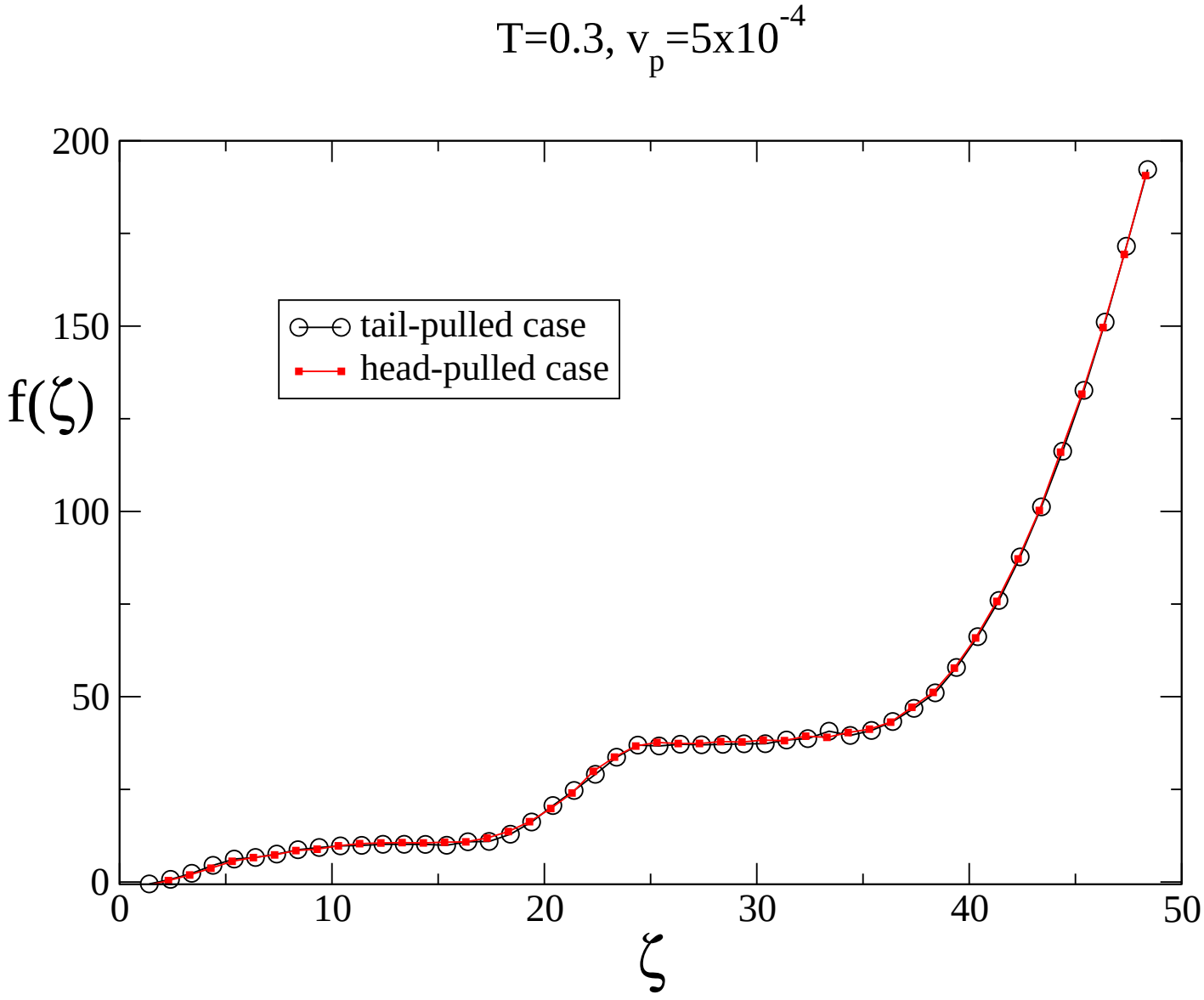
- Application of the two methods to reconstruct the free energy landscape of a bad folder (same number and types of residues of the good folder but random sequence).
- Analysis of the protein pulling with the constant force protocol.

[a] A. Imparato, S. Luccioli, A.T, PRL (2007)

THANK YOU
FOR YOUR ATTENTION!

http://www.fi.isc.cnr.it/users/alessandro.torcini/

Tail-pulled versus head-pulled case



Agree with F.-Y Li *et al.*, Phys. Rev. E 63 021905 (2001)

Langevin dynamics

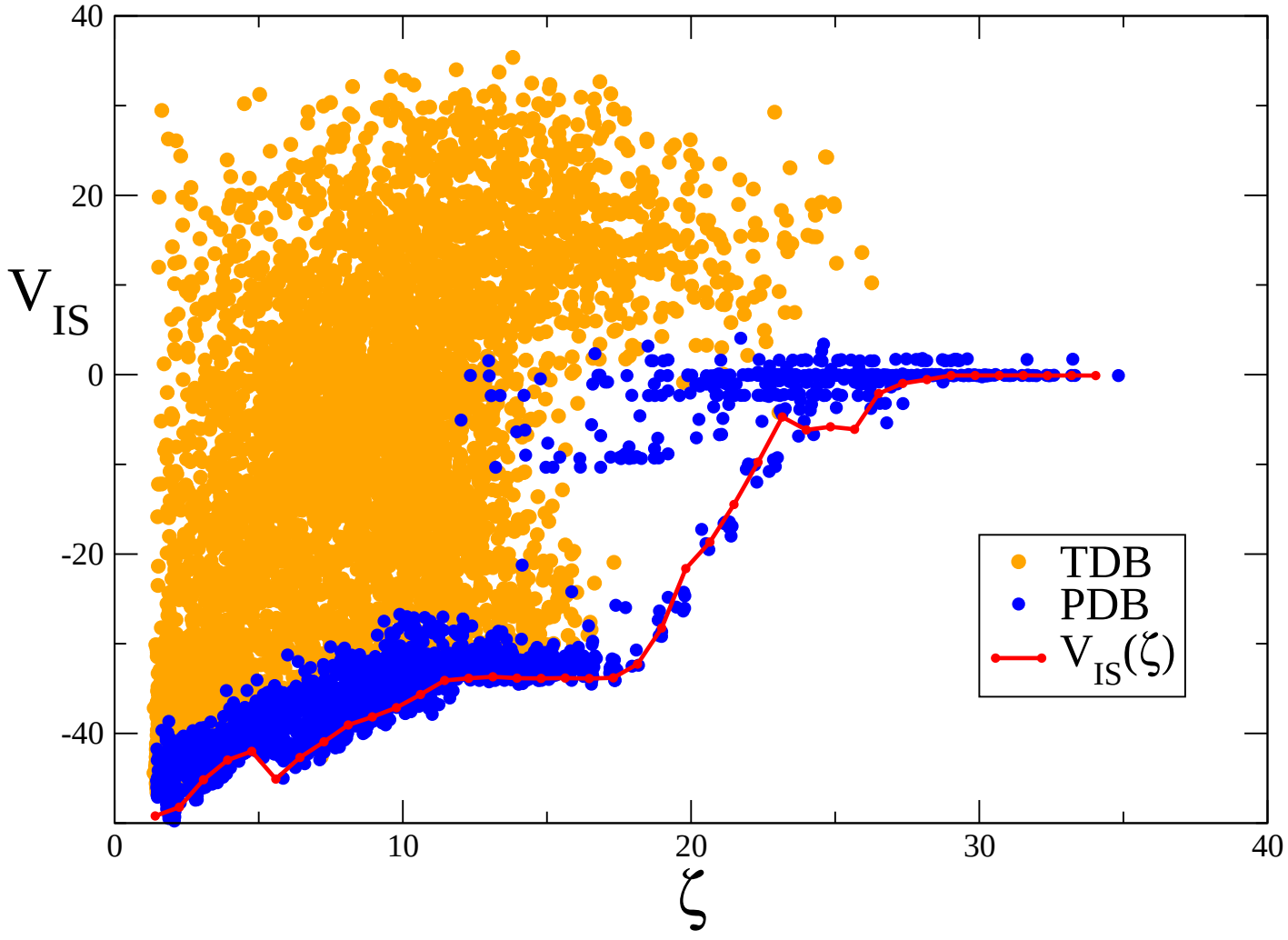
Canonical dynamics:

(7) $$m\ddot{\mathbf{r}}_i = \mathbf{F}(\mathbf{r}_i) - \gamma\dot{\mathbf{r}}_i + \eta(t) \qquad i = 1, N$$

where:

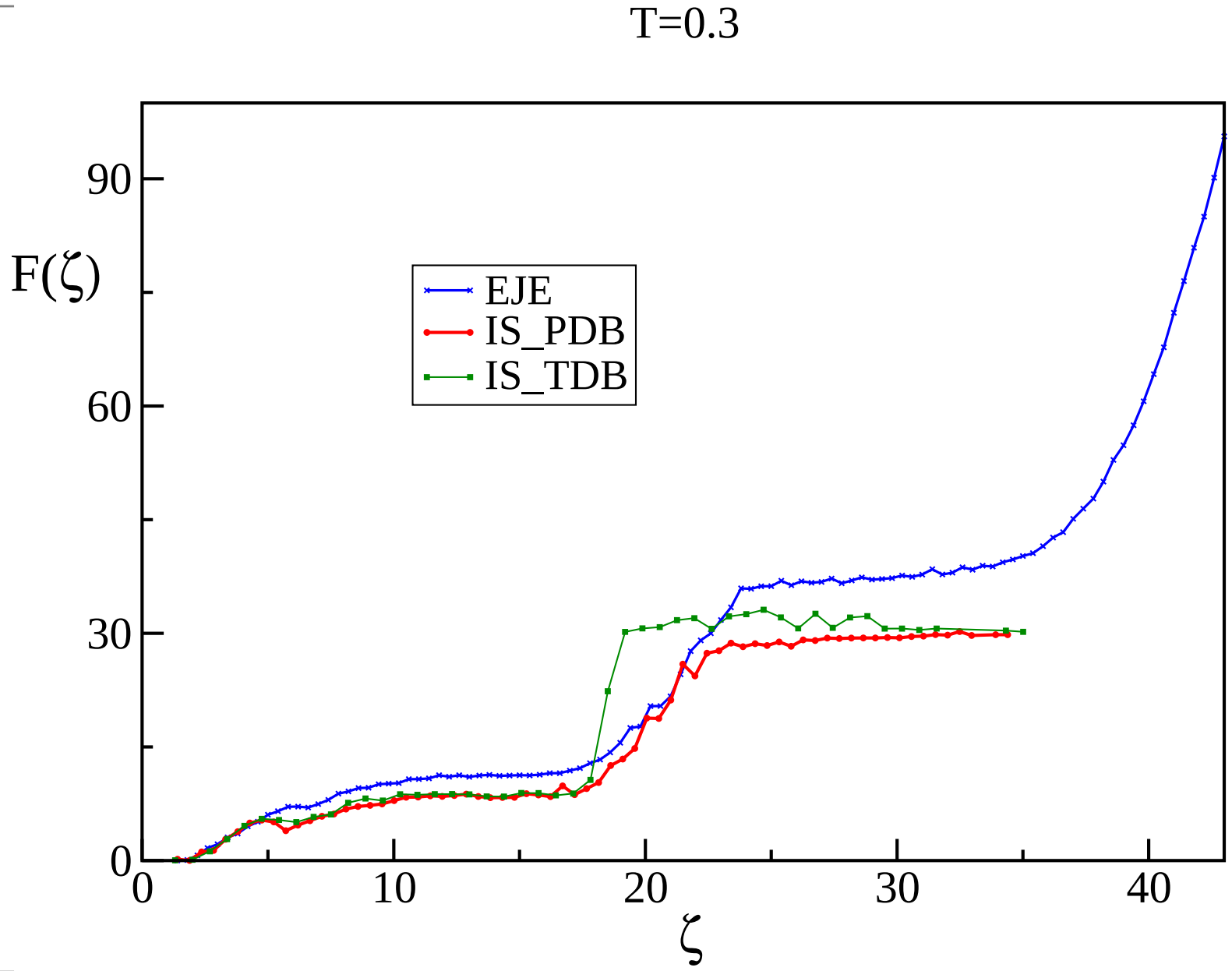
- $\langle\eta(t)\rangle = 0 \quad \langle\eta_\alpha(t)\eta_\beta(t')\rangle = (2K_BT\gamma/m)\delta(t-t')\delta_{\alpha,\beta} \qquad \alpha,\beta = x,y,z$
- $\mathbf{F} = -\nabla V \qquad \gamma \rightarrow$ friction coefficient

ISs data banks

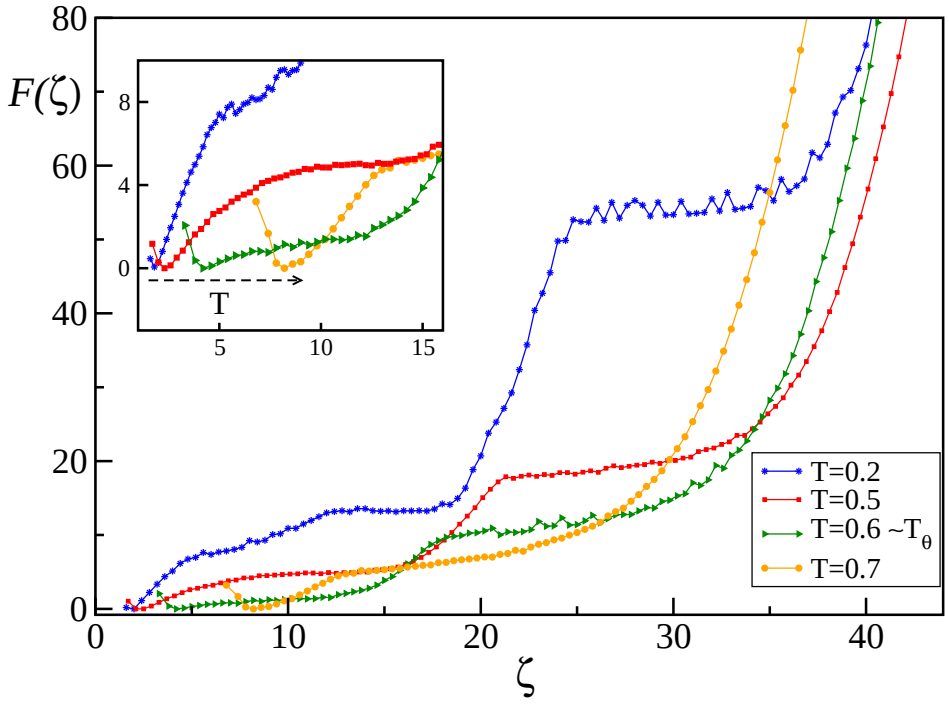


Two data bank of ISs (thermal data bank - TDB and pulling data bank - PDB) sampling the configurations visited in MD simulations and by relaxing via a steepest descent dynamics.

EJE versus ISs reconstruction



EJE: various temperatures



- $T \leq T_f$ $\longrightarrow$ the absolute minimum of $F(\zeta)$ is associated to the NC with $\zeta_0 \sim 2$;
- $T_f < T < T_\theta$ the free energy exhibits minima at $\zeta > \zeta_0$: the NC is no more the most favourite configuration, however the ST2 and ST3 barriers are lower but still present;
- $T > T_\theta$ only the ST2 barrier remains, the protein is mainly in extended configurations like (c) with some residual barrel structure.