Pulling of biomolecules: lessons from toy models

A. Prados

Física Teórica
Universidad de Sevilla

Mathematical Models for Physical and Biological
Nanoscale Systems and Their Applications,
Banff, Canada, Aug 28-Sep 2, 2016

In collaboration with:

L. L. Bonilla (UC3M), A. Carpio (UCM), C. A. Plata (US), F. Cecconi & M. Chinappi (La Sapienza)

Outline

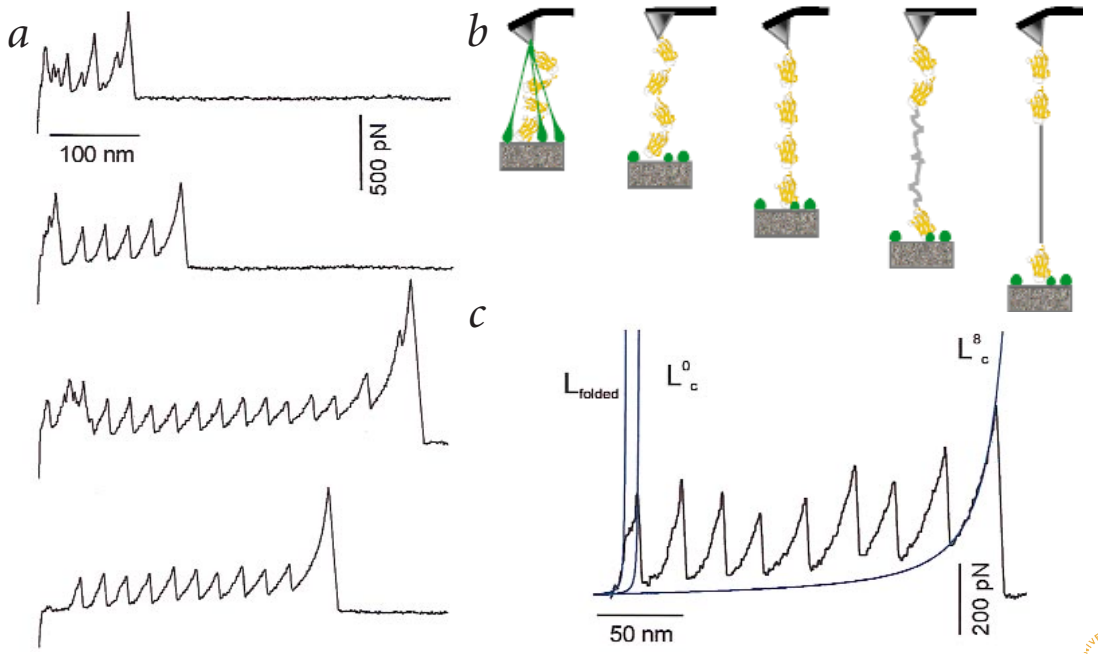
1. Force-extension curves
   - Motivation: experiments
   - Langevin approach (no spatial structure)
   - Conclusions (1)

2. Unfolding pathway
   - Some definitions and experimental results
   - Langevin approach (with spatial structure)
   - Conclusions (2)

Protein unfolding

Fisher, Marszalek and Fernandez, *Nat. Struct. Biol.* **7**, 719 (2000)

- Atomic Force Microscopy (AFM) experiments.
  - Molecule is stretched between { The tip of a microscopic cantilever. / Flat gold-covered substrate.
  - Forces acting on the molecule are transmitted to the cantilever, causing it to bend.
- *Simplest* ⇀ Modular protein comprising *N* identical domains.
- Force extension curve (FEC)
- Length-control: Sawtooth pattern



Main ideas for modelling

Bonilla, Carpio, Prados, PRE 91, 052712 (2015)

- N units characterised by a bistable free energy $A(\boldsymbol{\eta}; Y)$ for their extensions η_j. (Y intensive variables)

- Overdamped Langevin dynamics with noises verifying the fluctuation-dissipation theorem.

$$\gamma\dot{\eta}_j = F - \frac{\partial}{\partial\eta_j}A(\boldsymbol{\eta}; Y) + \sqrt{2T\gamma}\,\xi_j(t),$$

$$\langle\xi_j(t)\rangle = 0, \quad \langle\xi_j(t)\xi_l(t')\rangle = \delta_{jl}\,\delta(t-t'), \quad j = 1,\ldots,N.$$

- "Ideal" stretching (or pulling) of the biomolecule: additional force F.
  - Given for the force-controlled case.
  - Unknown (Legendre multiplier) for the length-controlled case: determined by $\sum_j \dot{\eta}_j = \dot{L}$, where $L(t)$ is the imposed end-to-end distance.

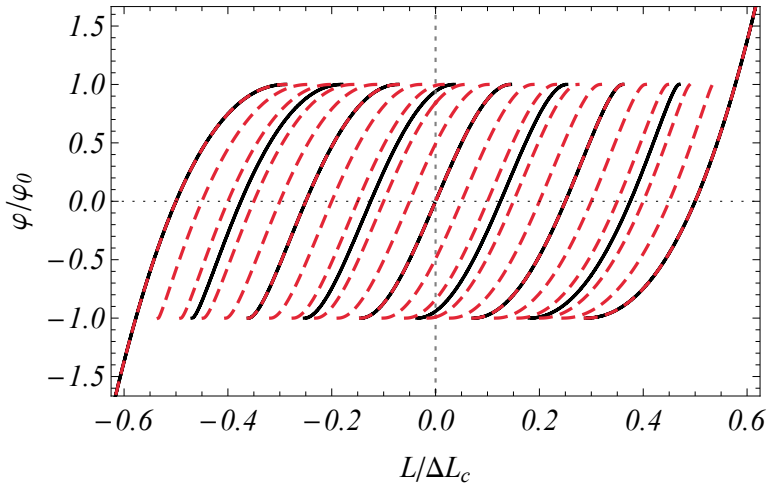
- Whether the force or the length is controlled, the FEC is recorded.

Simple quartic free energy

- Independent units with Landau-like bistable potentials:

$$A(\eta) = \sum_{j=1}^{N} a(\eta_j), \quad a(\eta) = F_c\eta - \alpha\eta^2 + \beta\eta^4.$$

- Results
  - Multi-stability in a range of forces around F_c.
  - Equilibrium branches.

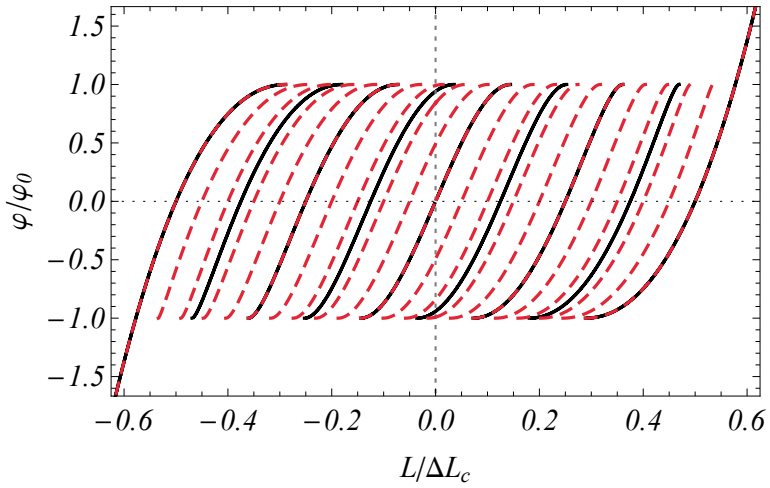


Simple quartic free energy

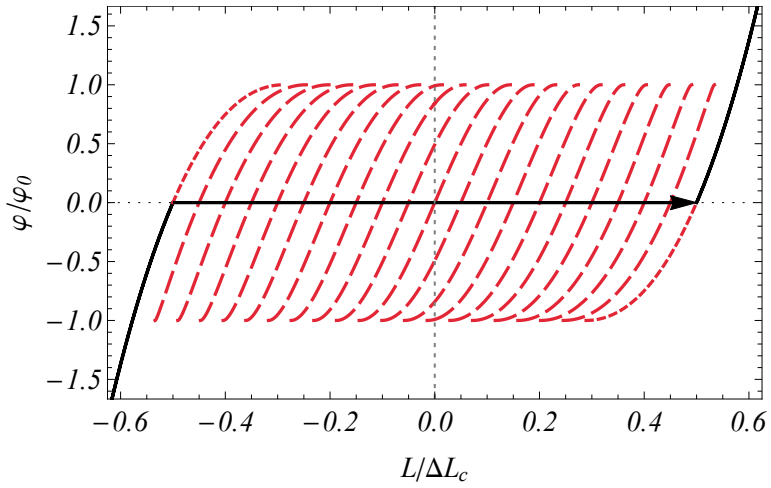
- Independent units with Landau-like bistable potentials:

$$A(\eta) = \sum_{j=1}^{N} a(\eta_j), \quad a(\eta) = F_c\eta - \alpha\eta^2 + \beta\eta^4.$$

- Results
  - Multi-stability in a range of forces around F_c.
  - Equilibrium branches.



  - Force-controlled situation: Independent units.
  - Equilibrium: Continuity of $G = A - FL$.

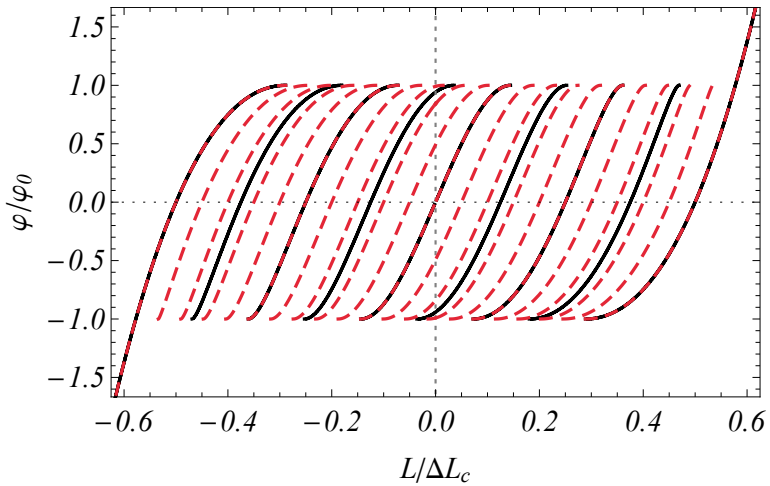


Simple quartic free energy

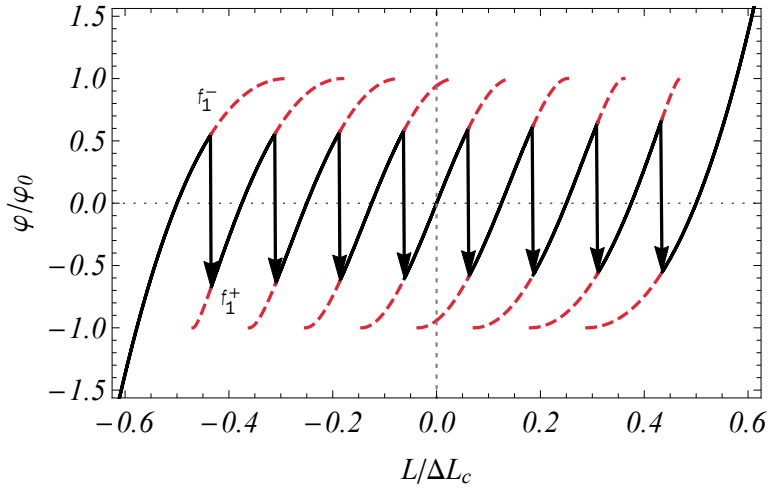
- Independent units with Landau-like bistable potentials:

$$A(\eta) = \sum_{j=1}^{N} a(\eta_j), \quad a(\eta) = F_c\eta - \alpha\eta^2 + \beta\eta^4.$$

- Results
  - Multi-stability in a range of forces around F_c.
  - Equilibrium branches.



  - Length-controlled situation: $\sum_j \eta_j = L$ is a global constraint.
  - Equilibrium: Continuity of A.



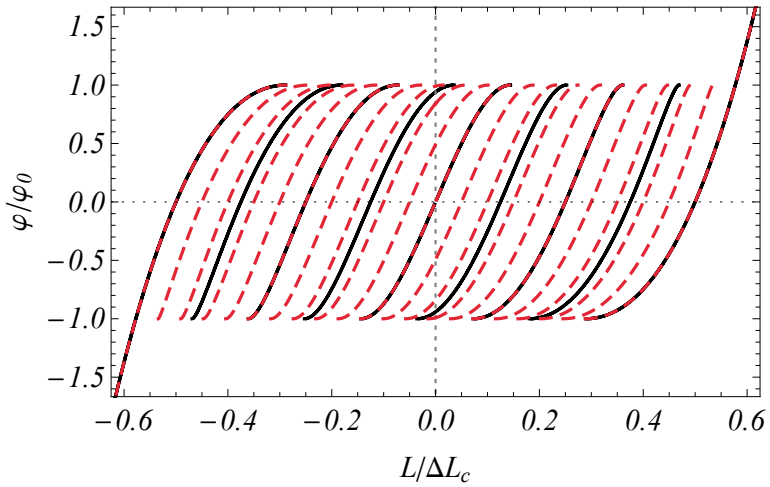
PC&B, PRE 88, 012704 (2013)

Simple quartic free energy

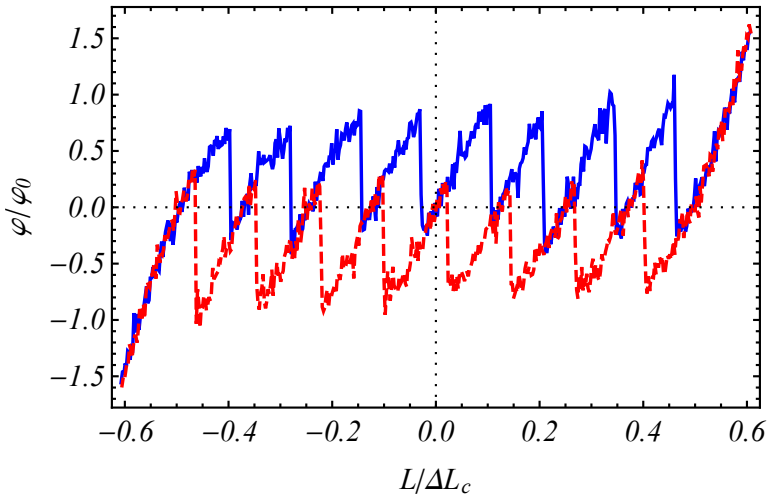
- Independent units with Landau-like bistable potentials:

$$A(\boldsymbol{\eta}) = \sum_{j=1}^{N} a(\eta_j), \quad a(\eta) = F_c\eta - \alpha\eta^2 + \beta\eta^4.$$

- Results
  - Multi-stability in a range of forces around F_c.
  - Equilibrium branches.



  - Length-controlled situation: $\sum_j \eta_j = L$ is a global constraint.
  - High pulling rate.

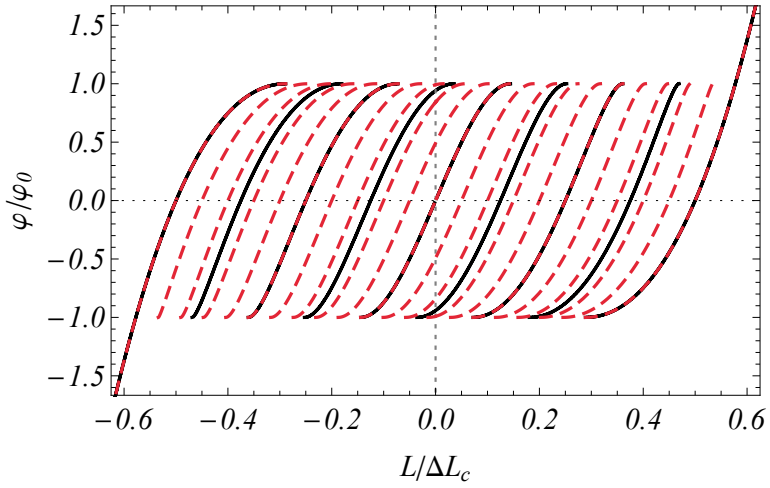


Simple quartic free energy

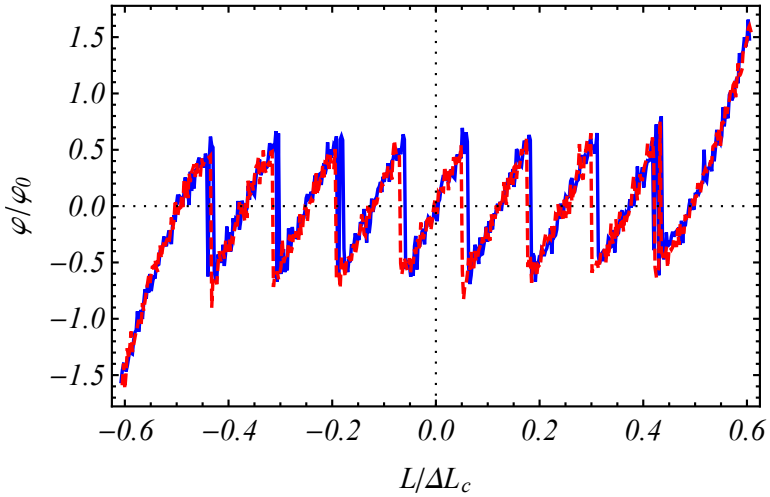
- Independent units with Landau-like bistable potentials:

$$A(\boldsymbol{\eta}) = \sum_{j=1}^{N} a(\eta_j), \quad a(\eta) = F_c\eta - \alpha\eta^2 + \beta\eta^4.$$

- Results
  - Multi-stability in a range of forces around F_c.
  - Equilibrium branches.

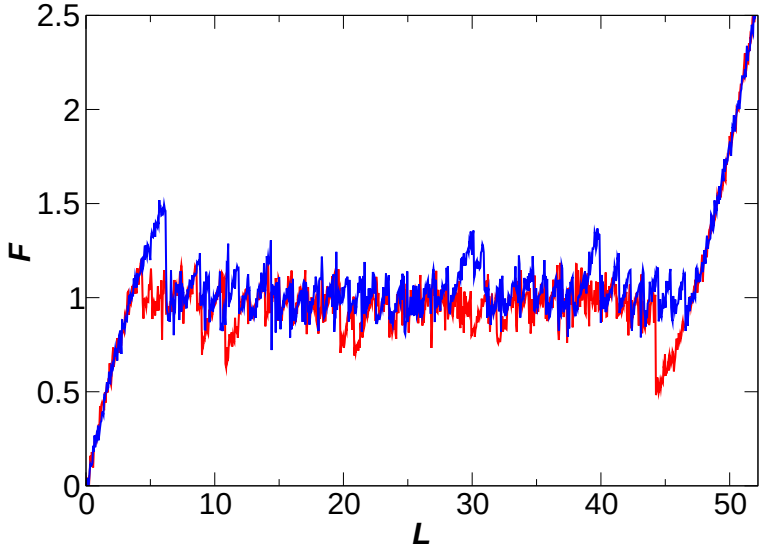


  - Length-controlled situation: $\sum_j \eta_j = L$ is a global constraint.
  - Slow pulling rate.



Additional results

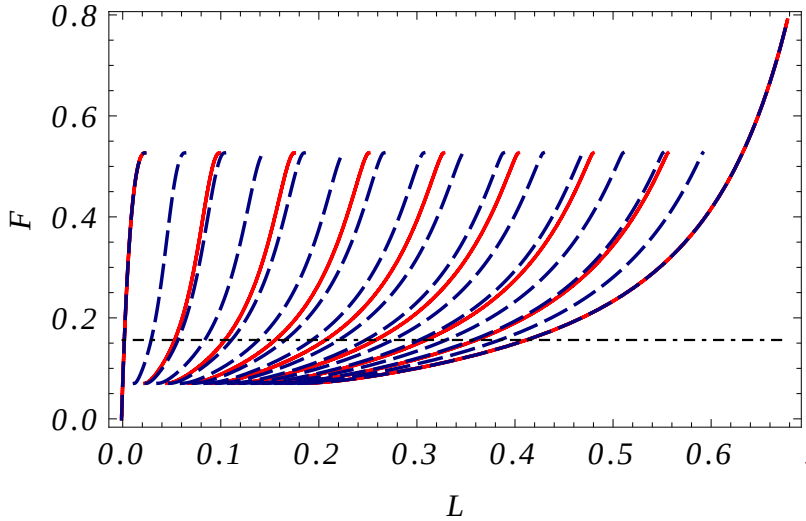
- Harmonic interaction $\frac{k}{2}(\eta_j - \eta_{j-1})^2$.
  - Almost reversible behaviour for low pulling rate but
    - ★ Intrinsic hysteresis in the first and last rip (seen in DNA hairpins).
  - Extra energetic cost for creating the first "domain wall".



- Realistic potential (no interaction).
  - Morse potential at short extension plus WLC at long extension (Berkovich et al.)

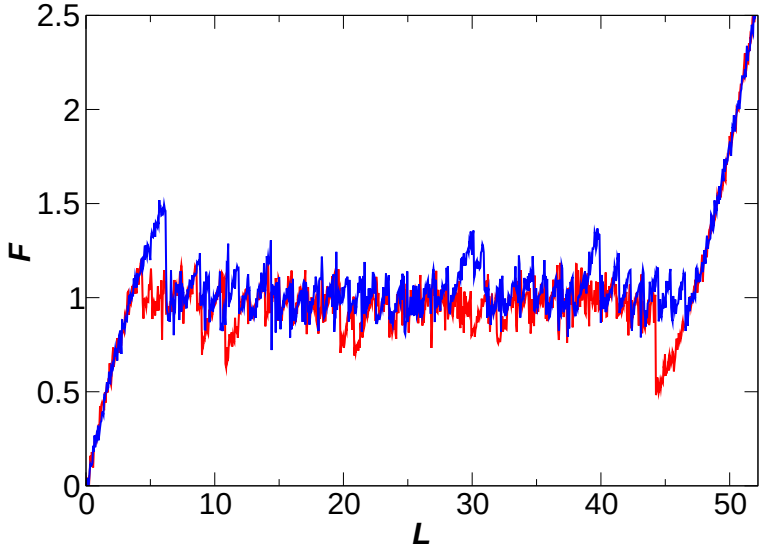
$$a(\eta) = U_0\left[\left(1 - e^{-2b(\eta - R_c)/R_c}\right)^2 - 1\right] + \frac{k_B T L_c}{4P}\left(\frac{1}{1 - \frac{\eta}{L_c}} - 1 - \frac{\eta}{L_c} + \frac{2\eta^2}{L_c^2}\right).$$

  - Equilibrium branches: non-symmetric!



Additional results

- Harmonic interaction $\frac{k}{2}(\eta_j - \eta_{j-1})^2$.
  - Almost reversible behaviour for low pulling rate but
    - ★ Intrinsic hysteresis in the first and last rip (seen in DNA hairpins).
  - Extra energetic cost for creating the first "domain wall".

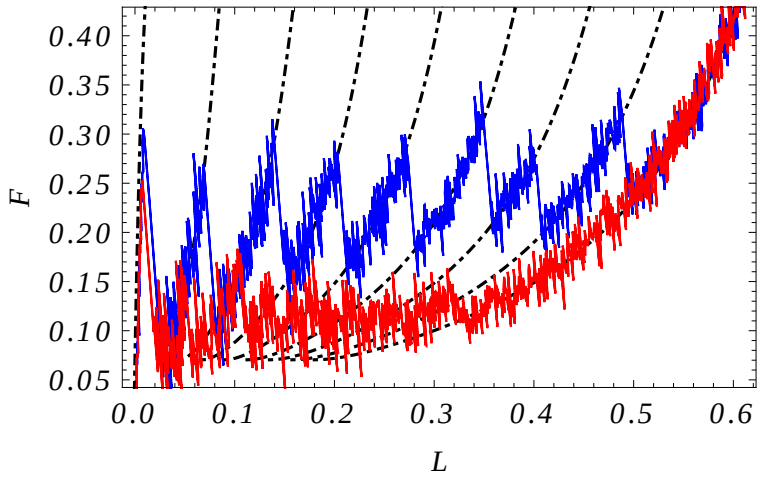


- Realistic potential (no interaction).
  - Morse potential at short extension plus WLC at long extension (Berkovich et al.)

$$a(\eta) = U_0\left[\left(1 - e^{-2b(\eta - R_c)/R_c}\right)^2 - 1\right] + \frac{k_B T L_c}{4P}\left(\frac{1}{1 - \frac{\eta}{L_c}} - 1 - \frac{\eta}{L_c} + \frac{2\eta^2}{L_c^2}\right).$$

  - Stretching-relaxing: more marked rips during unfolding!

Finite stiffness



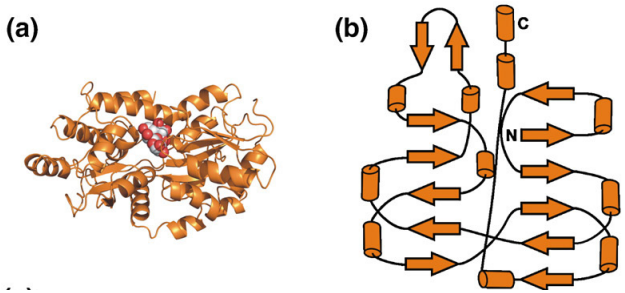
Summarising

- Length-control: global constraint introduces coupling.
- Metastable equilibrium behaviour: In the dynamical situations, the system is sweeping the metastable parts of the equilibrium branches (not actually in a far-from-equilibrium state).
- Some strong similarities with other physical systems composed of bistable units such as:
  - Semiconductor superlattices.
  - Li-ion storage systems.

Unfolding through intermediates

- Some proteins unfold through intermediates.
- Each jump corresponds to one unfoldon (a certain subunit).
- Example: Maltose Binding Protein which has four unfoldons M1, M2, M3, M4.
- Order in which the unfolding takes place: unfolding pathway.

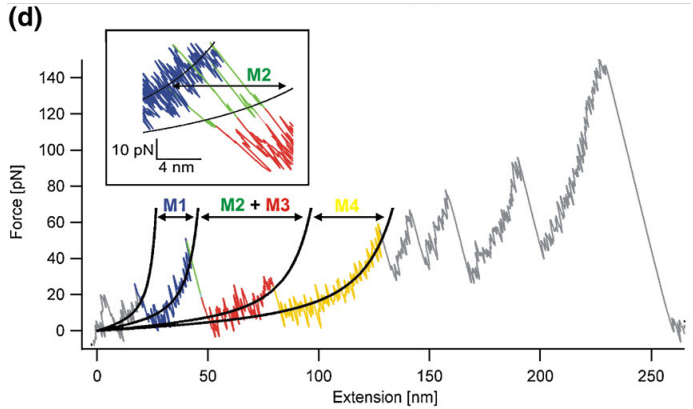
Bertz & Rief, J. Mol. Biol. 378, 447 (2008)



Unfolding through intermediates

- Some proteins unfold through intermediates.
- Each jump corresponds to one unfoldon (a certain subunit).
- Example: Maltose Binding Protein which has four unfoldons M1, M2, M3, M4.
- Order in which the unfolding takes place: unfolding pathway.

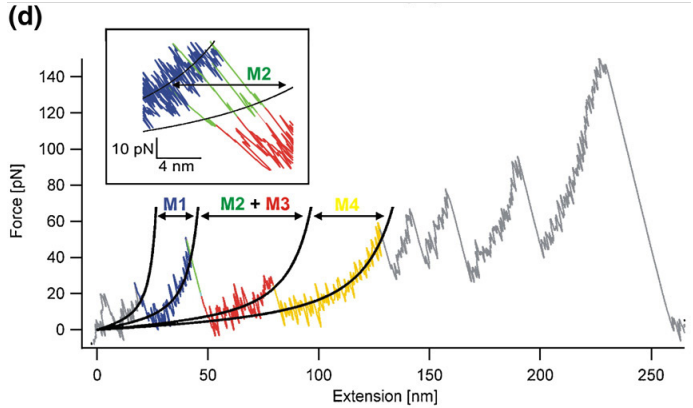
Bertz & Rief, J. Mol. Biol. 378, 447 (2008)



Unfolding through intermediates

- Some proteins unfold through intermediates.
- Each jump corresponds to one unfoldon (a certain subunit).
- Example: Maltose Binding Protein which has four unfoldons M1, M2, M3, M4.
- Order in which the unfolding takes place: unfolding pathway.

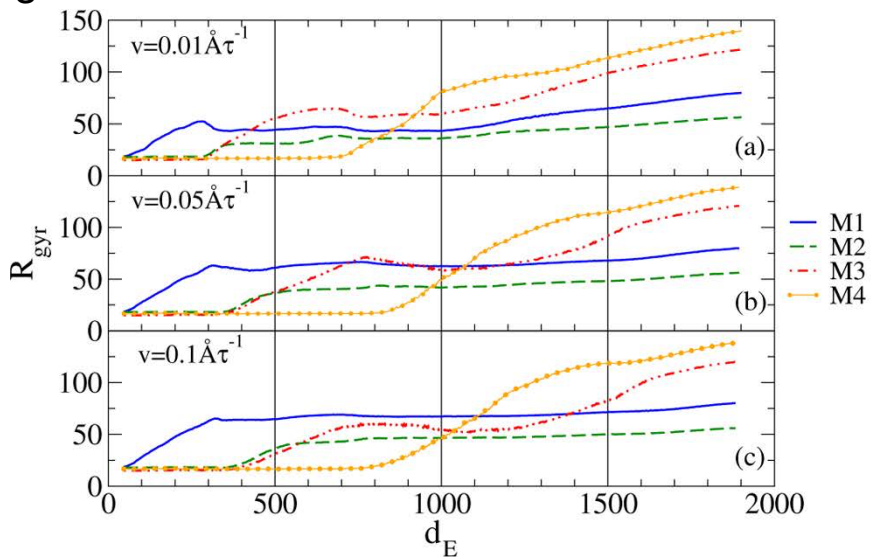
Bertz & Rief, J. Mol. Biol. 378, 447 (2008)



- Linear structure:

- The unfolding pathway may depend on:
  - C-pulling or N-pulling
  - Pulling velocity.

C-pulling

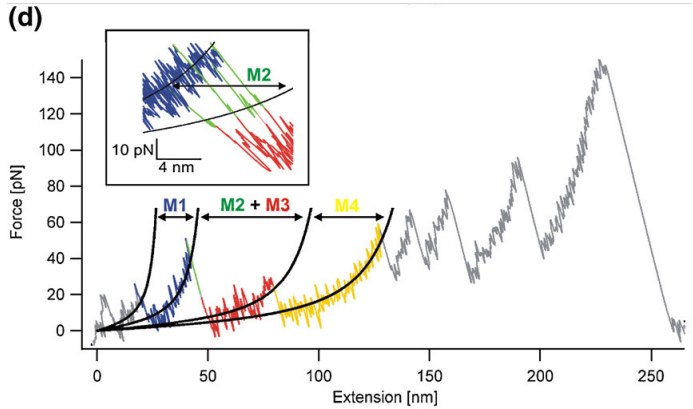


Guardiani, Di Marino, Tramontano, Chinappi, Cecconi, JCTC 10, 3589 (2014)

Unfolding through intermediates

- Some proteins unfold through intermediates.
- Each jump corresponds to one unfoldon (a certain subunit).
- Example: Maltose Binding Protein which has four unfoldons M1, M2, M3, M4.
- Order in which the unfolding takes place: unfolding pathway.

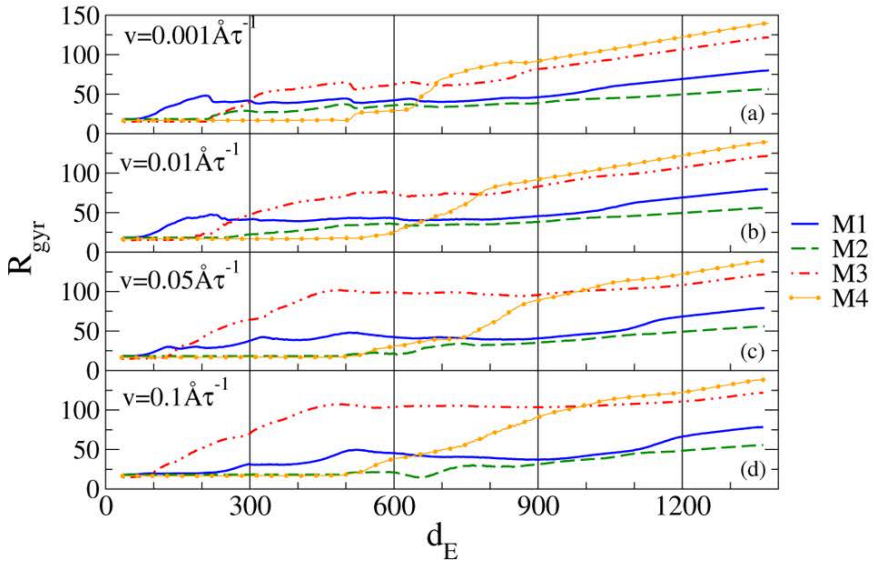
Bertz & Rief, J. Mol. Biol. 378, 447 (2008)



- Linear structure:

- The unfolding pathway may depend on:
  - C-pulling or N-pulling
  - Pulling velocity.

N-pulling

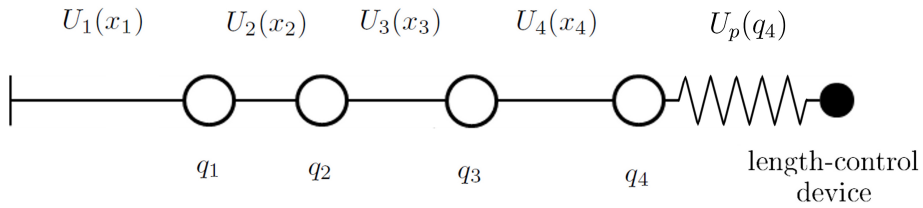


Guardiani, Di Marino, Tramontano, Chinappi, Cecconi, JCTC 10, 3589 (2014)

Main ideas for modelling

Plata, Cecconi, Chinappi, Prados, JSTAT P08003 (2015)

- Unfoldons: "Units" or monomers extending from q_{j-1} to q_j.
- Variant of the models in BC&P (2015) and Guardiani et al. (2014).
- Main difference with the model for analyzing the force-extension curves: spatial structure.



j-th unit extension: $x_j = q_j - q_{j-1}$.

- Langevin equations for q_j:

$$\gamma \dot{q}_j = -\frac{\partial}{\partial q_j} U(\boldsymbol{q}) + \sqrt{2T\gamma}\,\xi_j(t),$$

$$\langle \xi_j(t) \rangle = 0,\ \langle \xi_j(t)\xi_l(t') \rangle = \delta_{jl}\,\delta(t-t'),\ j = 1, \ldots, N.$$

- The system free energy is

$$U(\boldsymbol{q}) = \sum_{j=1}^{N} U_j(x_j) + U_p(q_N).$$

Perturbative theory

- Approximations: $\left\{ \begin{array}{l} \text{zero noise (deterministic approximation).} \\ \text{infinite stiffness (Lagrange multiplier } F\text{) device.} \end{array} \right.$
- Definitions: Pulling speed $v_p \equiv \dot{L}$. Length per unit $\ell = L/N$ measures time.
- More complex (coupled) equations due to the spatial structure:

$$\begin{aligned} \gamma \dot{x}_1 &= -U_1'(x_1) + U_2'(x_2), \\ \gamma \dot{x}_i &= -2U_i'(x_i) + U_{i+1}'(x_{i+1}) + U_{i-1}'(x_{i-1}), \\ \gamma \dot{x}_N &= -2U_N'(x_N) + U_{N-1}'(x_{N-1}) + F, \\ F &= \gamma v_p + U_N'(x_N). \end{aligned}$$

  - *Asymmetry* (disorder) in the potential:

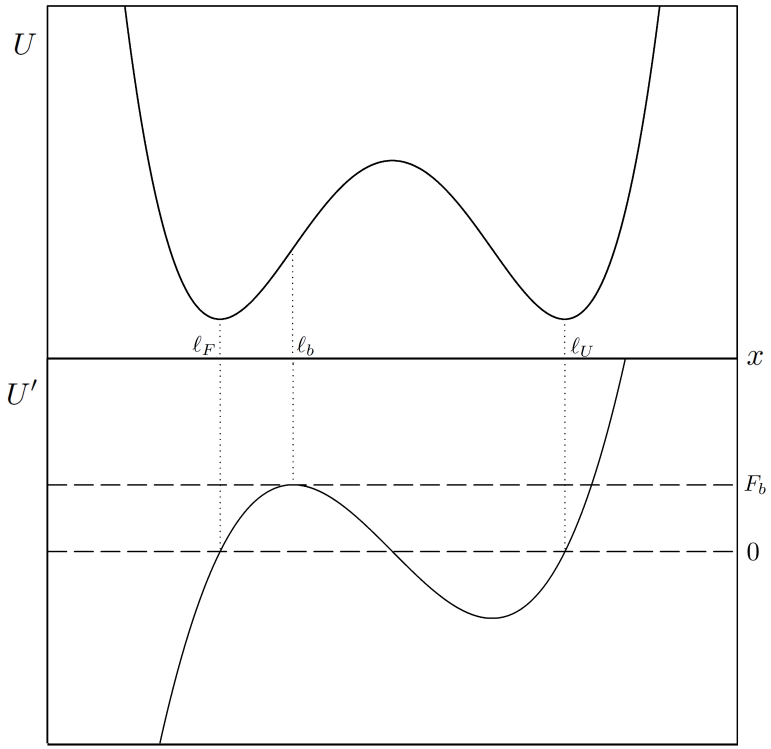
$$U_j(x) = U(x) + \xi\, \delta U_j(x)$$

  - Solution: Perturbative expansion in the pulling speed v_p ($d/dt \to v_p d/d\ell$) and the asymmetry ξ of the potential.

$$x_j(\ell) = \ell + v_p\, \Delta x_j^{(k)}(\ell) + \xi\, \Delta x_j^{(d)}(\ell).$$

Order of unfolding

- The unfolding starts from the unit that first reaches the limit of stability F_b, corresponding to an extension $\ell_b \left\{ \begin{array}{l} U'(\ell_b) = F_b \\ U''(\ell_b) = 0 \end{array} \right\}$.



- Competition between the $\left\{ \begin{array}{l} \text{kinetic correction which favors the pulled unit.} \\ \text{asymmetry correction which favors the weakest.} \end{array} \right.$

Results

- Simplest: fixed end the weakest (similar to N-pulling in the MBP)
- Weakest and pulled units reach ℓ_b at the same time when
  - $x_1(\ell_c) = x_N(\ell_c) = \ell_b$ $\left\{ \begin{array}{l} \text{“time” } \ell_c \\ \text{velocity } v_c \text{ as a function of } \xi \end{array} \right.$

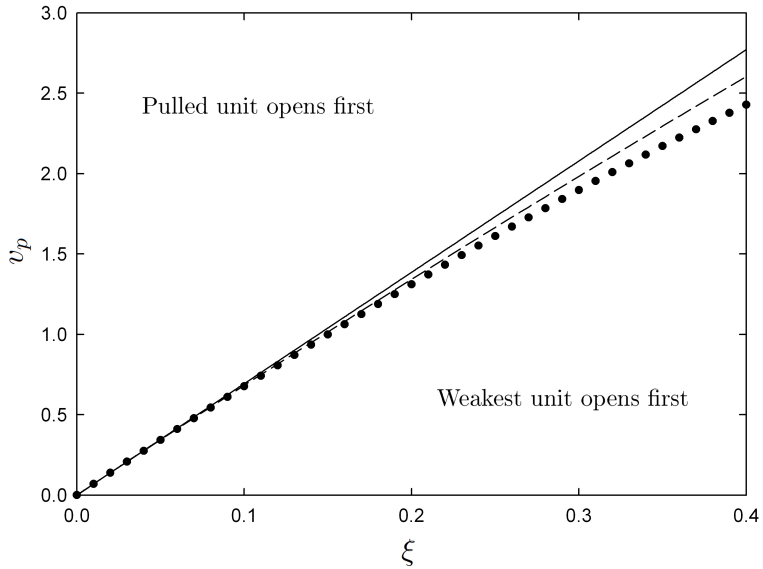
$$\frac{\gamma v_c}{\xi} = \frac{2}{N-1} \left[\delta U_N'(\ell_b) - \delta U_1'(\ell_b) \right] .$$

Results

- Simplest: fixed end the weakest (similar to N-pulling in the MBP)
- Weakest and pulled units reach ℓ_b at the same time when
  - $x_1(\ell_c) = x_N(\ell_c) = \ell_b \begin{cases} \text{“time” } \ell_c \\ \text{velocity } v_c \text{ as a function of } \xi \end{cases}$

$$\frac{\gamma v_c}{\xi} = \frac{2}{N-1}\left[\delta U_N'(\ell_b) - \delta U_1'(\ell_b)\right].$$

- **Pulling velocity-asymmetry phase diagram**
  - ***Critical velocity*** v_c**:**
    - ★ $v_p < v_c$: weakest unit unfolds first.
    - ★ $v_p > v_c$: pulled unit unfolds first.
  - Simple quartic potential.

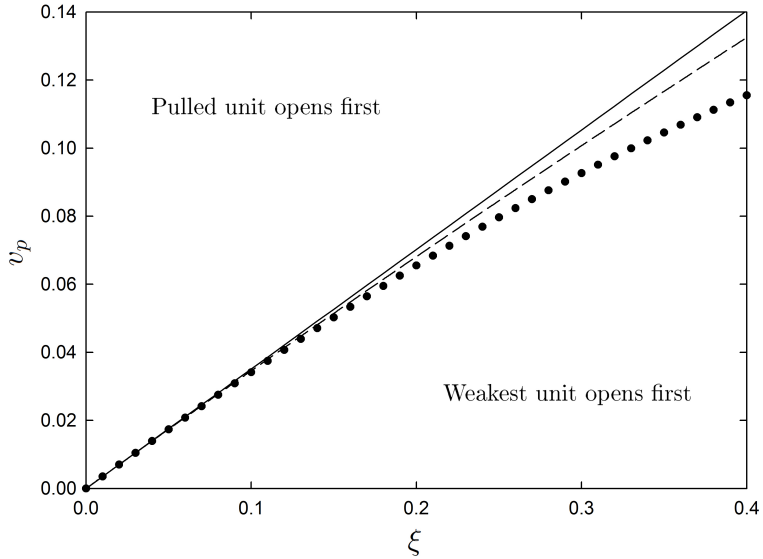


Results

- Simplest: fixed end the weakest (similar to N-pulling in the MBP)
- Weakest and pulled units reach ℓ_b at the same time when
  - $x_1(\ell_c) = x_N(\ell_c) = \ell_b \begin{cases} \text{“time” } \ell_c \\ \text{velocity } v_c \text{ as a function of } \xi \end{cases}$

$$\frac{\gamma v_c}{\xi} = \frac{2}{N-1}\left[\delta U_N'(\ell_b) - \delta U_1'(\ell_b)\right].$$

- **Pulling velocity-asymmetry phase diagram**
  - ***Critical velocity*** v_c**:**
    - ★ $v_p < v_c$: weakest unit unfolds first.
    - ★ $v_p > v_c$: pulled unit unfolds first.
  - Realistic potential.



Wrapping things up

- Perturbative approach for small (enough) pulling velocity and asymmetry ("disorder" in the free energies).

- The unfolding pathway depends on the pulling velocity and there appears a critical velocity v_c.
  - For low pulling velocity, the force is homogeneous across the protein and the weakest unit unfolds first.
  - As the pulling velocity is increased, the force is not homogeneous across the protein: the pulled unit feels a higher force and unfolds first for a large enough v_p.
  - **Taking into account the spatial structure is mandatory: otherwise, the force would be homogeneously distributed throughout the chain**.

- Limitations: $\begin{cases} \text{Thermal fluctuations have been neglected.} \\ \text{Independence of the units.} \end{cases}$

- Perspective:
  - Checking our theory in modular proteins by engineering one module to be weaker than the rest.
  - In progress with P. Marszalek's group (experiments and SMD simulations).

Thanks for your attention!

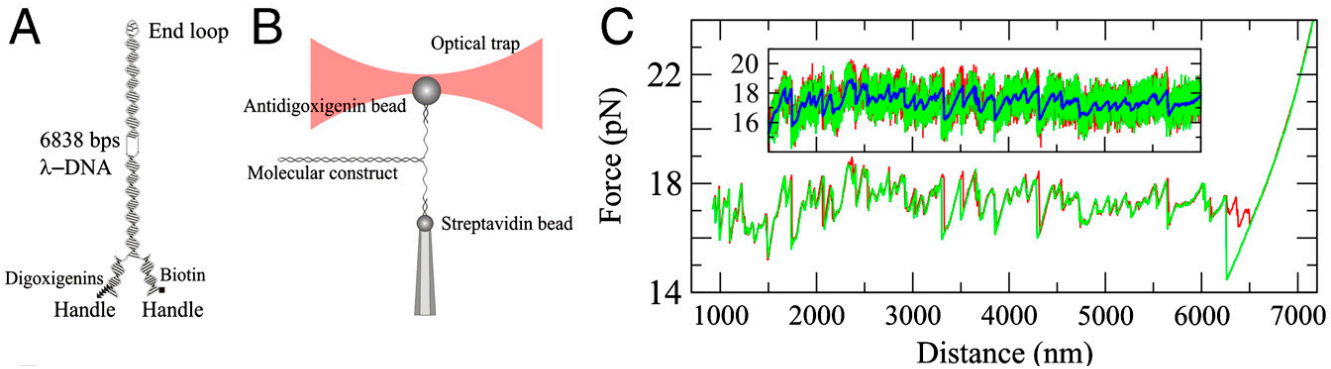
Comments and questions are welcome.

References for our work:

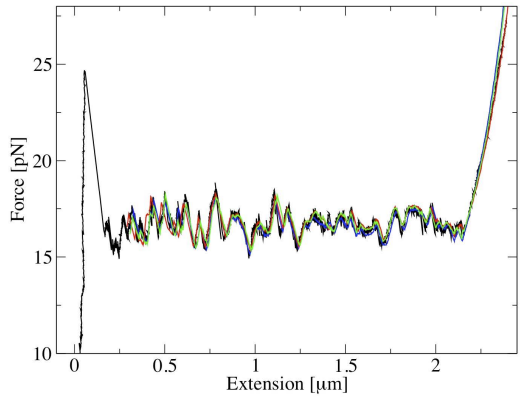
1. A. Prados, A. Carpio, and L. L. Bonilla, Spin-oscillator model for the unzipping of biomolecules by mechanical force, PRE 86, 021919 (2012).
2. A. Prados, A. Carpio, and L. L. Bonilla, Sawtooth patterns in force-extension curves of biomolecules: an equilibrium-statistical-mechanics theory, PRE 88, 012704 (2013).
3. L. L. Bonilla, A. Carpio, and A. Prados, Protein unfolding and refolding as transitions through virtual states, EPL 108, 28002 (2014).
4. L. L. Bonilla, A. Carpio, and A. Prados, Theory of force-extension curves for modular proteins and DNA hairpins, PRE 91, 052712 (2015) [Highlighted in Revista Española de Física 29 (3), 29 (2015)].
5. C. A. Plata, F. Cecconi, M. Chinappi, and A. Prados, Understanding the dependence on the pulling speed of the unfolding pathway of proteins, J. Stat. Mech. P08003 (2015).

Length-controlled unzipping/rezipping of DNA J. M. Huguet et al., PNAS 107, 15431 (2010)

- Nucleic acids are *simpler* than proteins: 4 different nucleotides vs. 20 amino acids.
- Technique: Laser Optical Tweezers.
- Unzipping/rezipping of a 7000bp-long DNA hairpin in a length-controlled experiment at very low pulling rates (10nm/s).
- Almost reversible force-extension curves are obtained.

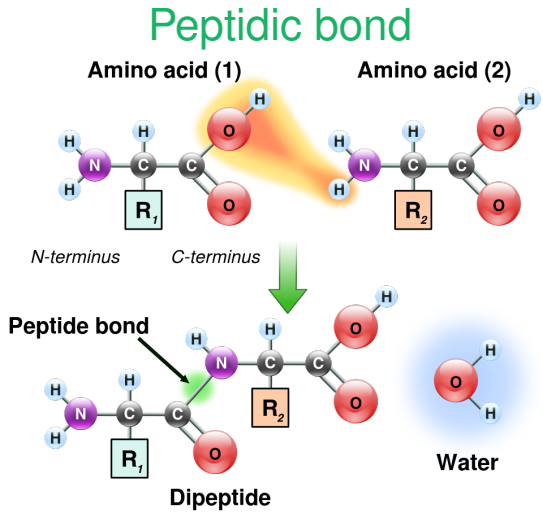


- There is some intrinsic hysteresis in the last (first) rip.



More detailed look at a protein

- Building blocks of proteins: amino acids.
  - C-terminus: carboxylic acid group COOH.
  - N-terminus: amino group.
- Different radicals (or substituents) groups *R*.



- Scheme of the 3d structural organisation of a protein.
  - Primary structure: Amino acid sequence.
  - Secondary structure: Highly regular *local* substructures (hydrogen bonding).
    - ★ α-helices.
    - ★ β-sheets.
  - Tertiary structure: 3d folding (hydrophobic interactions).