

Agenda 2nd Lecture

Non-Equilibrium Systems: some definitions

Umbrella Sampling

Jarzynski-Crook

Steered MD

applications

Fluctuations in large systems, are mostly determined by the conditions of the environment. Large deviations from the average behavior are hardly observable and the structural properties of the system cannot be inferred from the spectrum of fluctuations.

In contrast, small systems will display large deviations from their average behavior. These turn out to be less sensitive to the conditions of the surrounding environment (temperature, pressure, chemical potential) and carry information about the structure of the system and its non-equilibrium behavior.

From:https://www.researchgate.net/publication/1888827_Nonequilibrium_fluctuations_in_small_systems_From_physics_to_biology

Small Systems and energy fluctuations

“Small” systems are those in which the energy exchanged with the environment is a few times $k_B T$ and energy fluctuations are observable.

A few can be 10 or 1000 depending on the system.

A small system **must not necessarily be of molecular size or contain a few number of molecules.**

For example, a single polymer chain may behave as a small system although it contains millions of covalently linked monomer units.

At the same time, a molecular system may not be small if the transferred energy is measured over long times compared to the characteristic heat diffusion time. In that case the average energy exchanged with the environment during a time interval t can be as large as desired by choosing t large enough. Conversely, a macroscopic system operating at short time scales could deliver a tiny amount of energy to the environment, small enough for fluctuations to be observable and the system being effectively small.

Non-equilibrium systems

In general a non-equilibrium state is produced whenever the system properties change with time and/or the net heat/work/mass exchanged by the system and the bath is non zero. We can distinguish at **least three different types of non-equilibrium states**:

- 1 **Non-equilibrium transient state (NETS)**. The system is initially prepared in an equilibrium state and later driven out of equilibrium by switching on an external perturbation. The system quickly returns to a new equilibrium state once the external perturbation stops changing.
- 2 **Non-equilibrium steady-state (NESS)**. The system is driven by external forces (either time dependent or non-conservative) in a stationary non-equilibrium state where its properties do not change with time. The steady state is an irreversible non-equilibrium process, that cannot be described by the Boltzmann-Gibbs distribution, where the average heat that is dissipated by the system (equal to the entropy production of the bath) is positive.
- 3 **Non-equilibrium aging state (NEAS)**. The system is initially prepared in a non-equilibrium state and put in contact with the sources. The system is then let evolve alone but fails to reach thermal equilibrium in observable or laboratory time scales. In this case the system is in a non-stationary slowly relaxing non-equilibrium state called aging state and characterized by a very small entropy production of the sources. In the aging state two-times correlations decay slower as the system becomes older. Two-time correlation functions depend on both times and not just on their difference.

Some useful –maybe redundant- definitions

Free Energy Changes

In statistical mechanics, a **canonical ensemble** is the statistical **ensemble** that represents the possible states of a mechanical system in thermal equilibrium with a heat bath at some fixed temperature. The system can exchange energy with the heat bath, so that the states of the system will differ in total energy.

$$\Delta G = \Delta H - T \Delta S \quad (1)$$

$$S = \ln(\Omega) \quad (2)$$

where Ω is the number of accessible microstates.

Partition Function Z

Boltzmann Distribution

- Molecules will distribute themselves over the available energy levels according to the Boltzmann distribution.
- Physical interpretation: Z is a measure of the number of thermally accessible energy states. Note that it is a function of the pattern of energy levels (ε_i) and the temperature (T).

$$\begin{aligned} G &= -k_B T \ln(Z) \\ &= k_B T \ln(1/Z) \\ &= k_B T \sum_i \frac{1}{e^{-E_i/k_B T}} \\ &= k_B T \sum_i e^{E_i/k_B T} \\ &= k_B T \left\langle e^{E_i/k_B T} \right\rangle \end{aligned} \tag{3}$$

Umbrella Sampling I

In a canonical distribution, the probability to observe the microstate i is:

$$p_i = \frac{1}{Z} e^{-\beta E_i} \quad (4)$$

where $\beta = 1 / k_B T$. Z is the Partition Function, the sum over all accessible states j , and acts here as normalisation term:

$$Z = \sum_j e^{-\beta E_j} \quad (5)$$

The expectation value of the energy is:

$$\langle E \rangle = \frac{1}{Z} \sum_i E_i e^{-\beta E_i} \quad (6)$$

A nice mathematical trick is:

$$-\frac{\partial}{\partial \beta} Z = -\sum_i \frac{\partial}{\partial \beta} e^{-\beta E_i} = \sum_i E_i e^{-\beta E_i} \quad (7)$$

Umbrella Sampling II

and therefore by comparing equations 6 and 7:

$$\langle E' \rangle = -\frac{1}{Z} \frac{\partial}{\partial \beta} Z = -\frac{\partial \ln(Z)}{\partial \beta}$$

To relate the Free Energy to Z , the easiest way is to use the thermodynamic relation:

$$F = U - TS$$

Multiplying with β

$$\beta F = \beta U - S / k$$

and taking the derivative

$$\frac{\partial (\beta F)}{\partial \beta} = U = \langle E \rangle$$

Umbrella Sampling III

and comparing with equation 8 gives the expression for the Free Energy:

$$F = -k_B T \ln(Z) \quad (12)$$

and for the Entropy:

$$\begin{aligned} S &= -F / T - U / T \\ &= k_B \ln(Z) + k_B \beta \langle E \rangle \end{aligned} \quad (13)$$

This is a remarkable result, since we only need to get the density ρ of points in phase space and integrate over the phase space to obtain Z . Everything else follows from Z .

Umbrella Sampling IV

Umbrella sampling is the method of choice to compute the Free Energy along a reaction coordinate q for a transition between two states. The reaction coordinate is the dimension along which we study the process, for example the distance between donor and acceptor or ligand and receptor.

Umbrella Sampling IV

To express the reaction coordinate as independent vector, the original $3N$ degrees of freedom of N particles in 3D space (for example in the form of coordinates $[x_1, x_2, \dots, x_{3N}]$) can be transformed such that the new coordinate system is $[u_1, u_2, \dots, u_{3N-1}, q]$, where q is the reaction coordinate. A change in the original coordinates is related to the new coordinate system:

$$d\vec{x} = d\vec{u} dq \quad (14)$$

When we look for the free energy for a certain value of q , we average over all remaining degrees of freedom, i.e. we perform an MD and sample all degrees of freedom except for q . An example would be the free energy for the formation of an ion pair in solution, as shown in Fig. 3. For every value of q , a MD is performed to calculate the free energy for that value of the reaction coordinate.

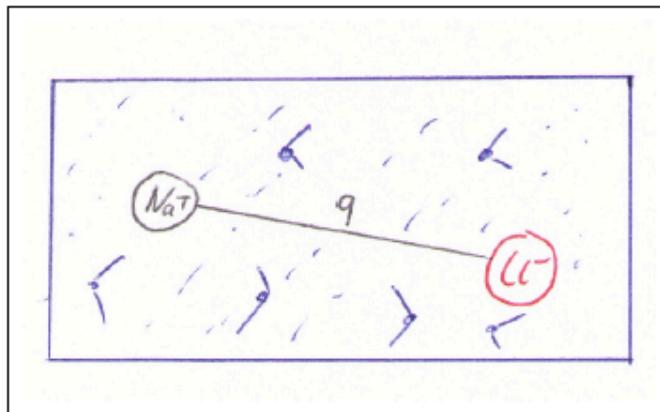


Figure Na^+ and Cl^- in water solution: the distance between the ions is the reaction coordinate q .

Umbrella Sampling V

The Free Energy is given by:

$$F = -k_B T \ln \left(\sum_i e^{-\beta E_i} \right) \quad (15)$$

We can choose to consider only states in which a specific reaction coordinate q_0 is attained by defining the Kronecker function

$$\delta(q - q_0) \quad (16)$$

Umbrella Sampling VI

In the following, the initial expression is expanded by Z and expressed in terms of the probability density ρ in phase space, which together with the Kronecker functions yields the density at q_0 , leading to the target expression for the Free Energy as a function of the reaction coordinate.

$$\begin{aligned} F(q_0) &= -k_B T \ln \sum_i \delta(q - q_0) e^{-\beta E_i} \\ &= -k_B T \ln \left(Z \sum_i \delta(q - q_0) \frac{e^{-\beta E_i}}{Z} \right) \\ &= -k_B T \ln \left(Z \sum_i \delta(q - q_0) \rho_i \right) \\ &= -k_B T \ln (Z \langle \delta(q - q_0) \rangle) \\ &= -k_B T \ln (Z) - k_B T \ln (\langle \delta(q - q_0) \rangle) \end{aligned} \tag{17}$$

In an MD simulation, we can count how often the system is at q_0 . To compute the Free Energy difference between states A and B :

Umbrella Sampling VII

We obtain the probability $P(q_0)$ for the system to be at coordinate q_0 .

$$P(q_0) = \sum_i \delta(q - q_0) \rho_i = \langle \delta(q - q_0) \rangle \quad (18)$$

Umbrella Sampling VIII

The Free Energy difference between two states A and B

$$\begin{aligned} F_B - F_A &= -k_B T \ln(Z) - k_B T \ln(\langle \partial(q - q_B) \rangle) + \\ &\quad + k_B T \ln(Z) + k_B T \ln(\langle \partial(q - q_A) \rangle) \\ &= -k_B T \ln\left(\frac{\langle \partial(q - q_B) \rangle}{\langle \partial(q - q_A) \rangle}\right) \\ &= -k_B T \ln\left(\frac{P(q_B)}{P(q_A)}\right) \end{aligned} \tag{19}$$

So, the task is clear: perform a MD, specify a coordinate, and then just count, how often the system is at special values of the reaction coordinate: the difference of these numbers gives the free energy difference!

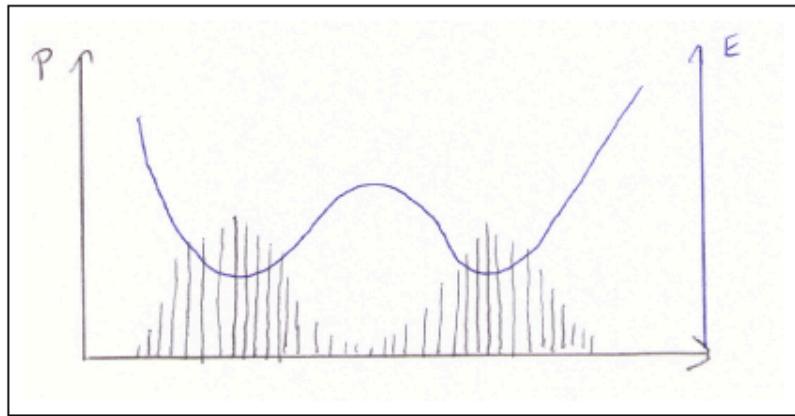


Figure Schematic energy profile along a reaction coordinate and the probability distribution. The barrier region is sampled poorly.

This is nice, but we also know the problem: If we have to cross a high barrier along the reaction coordinate to come from A to B, a pure MD will never do it. Therefore, we have to **drive** the system 'somehow': This can be done, by [applying an additional potential V!](#)

Umbrella Sampling IX

$$\begin{aligned} F(q_0) &= -k_B T \ln \left(\frac{\sum_i \partial(q - q_0) e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} \right) \\ &= -k_B T \ln \left(\frac{\sum_i \partial(q - q_0) e^{\beta V} e^{-\beta(E_i + V)}}{\sum_i e^{-\beta(E_i + V)}} \frac{\sum_i e^{-\beta(E_i + V)}}{\sum_i e^{-\beta E_i}} \right) \\ &= -k_B T \ln \left(\langle \partial(q - q_0) e^{\beta V} \rangle_{E+V} \frac{\sum_i e^{-\beta E_i + V}}{\sum_i e^{\beta V} e^{-\beta(E_i + V)}} \right) \\ &= -k_B T \ln \left(\langle \partial(q - q_0) e^{\beta V} \rangle_{E+V} \frac{1}{\langle e^{\beta V} \rangle_{E+V}} \right) \\ &= -k_B T \ln \left(e^{\beta V(q_0)} \langle \partial(q - q_0) \rangle_{E+V} \frac{1}{\langle e^{\beta V} \rangle_{E+V}} \right) \\ &= -k_B T \ln (\langle \partial(q - q_0) \rangle_{E+V}) - V(q_0) + k_B T \ln \left(\langle e^{\beta V} \rangle_{E+V} \right) \\ &= -k_B T \ln (P^*(q_0)) - V(q_0) + k_B T \ln \left(\langle e^{\beta V} \rangle_{E+V} \right) \end{aligned}$$

Umbrella Sampling X

The last equation has the form:

$$F(q) = -k_B T \ln(P^*(q)) - V(q) + K \quad (21)$$

We can use this scheme efficiently, when we move harmonic potentials along the reaction coordinate as shown in Fig. 5.

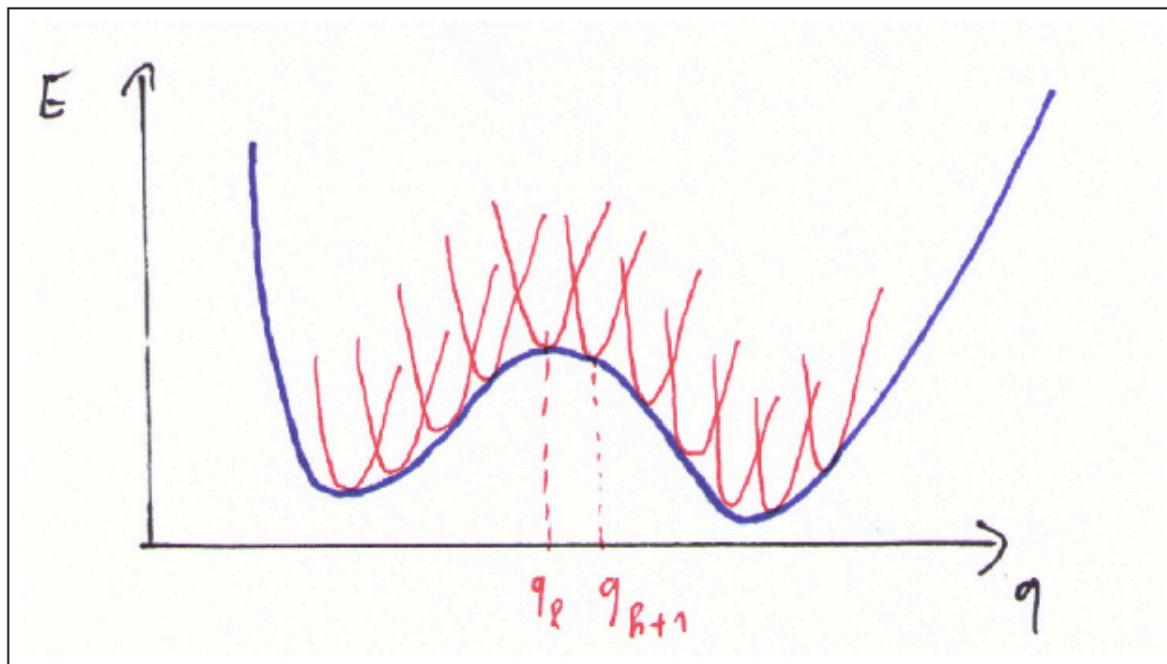


Figure Additional harmonic potentials to keep the system in the region of the desired value of the reaction coordinate

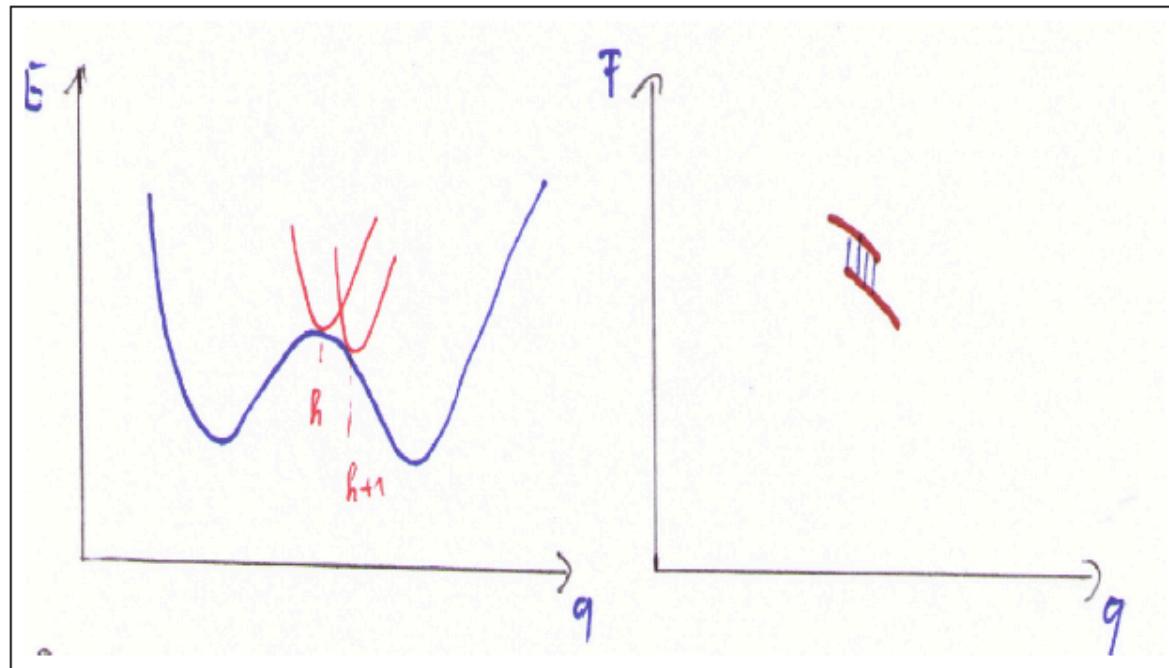
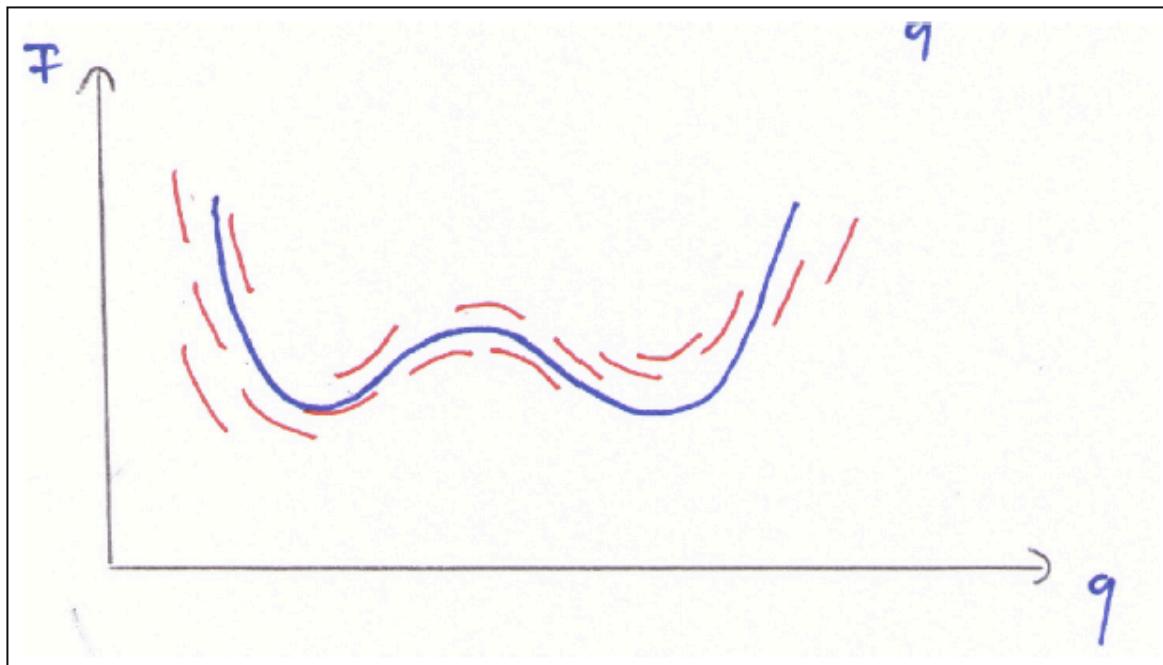


Figure Offset in free energy between two simulations k and $(k+1)$.
The offset is given by $K_k - K_{k+1}$



Figure

Matching of histograms from different simulations

Jarzynski and Crook

The maximum work theorem: a consequence of the second law of thermodynamics

The amount of work W performed on a system during a non-equilibrium transformation is larger than the free energy difference ΔF between the equilibrium states corresponding to the transition end points:

$$\langle W \rangle \geq \Delta F$$

As shown by Jarzynski in 1997, the work fluctuations resulting for microscopic systems can be accounted for in an exact way, transforming the maximum work theorem into an equality:

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}$$

This is referred to as the Jarzynski non-equilibrium work theorem and relates the statistics of irreversible work carried out on the system, while it is driven away from equilibrium, to an equilibrium free energy difference.

Jarzynski and Crook

A closely connected result is the **Crooks fluctuation theorem**, which relates the equilibrium free energy difference to the work distributions of the forward and reversed process.

In general, processes during which work is performed on or by the system drive the system away from equilibrium, such that the phase space distribution obtained at the end of the process may differ strongly from the equilibrium distribution to which the system relaxes after the external perturbation has been stopped.

For instance, a piston pushed quickly into a gas-filled cylinder generates non-equilibrium states with strong flows markedly different from the static equilibrium state to which the gas eventually relaxes after the piston has reached its final state.

At first sight, it is therefore surprising that equilibrium properties, such as free energy differences, can be extracted from non-equilibrium trajectories.

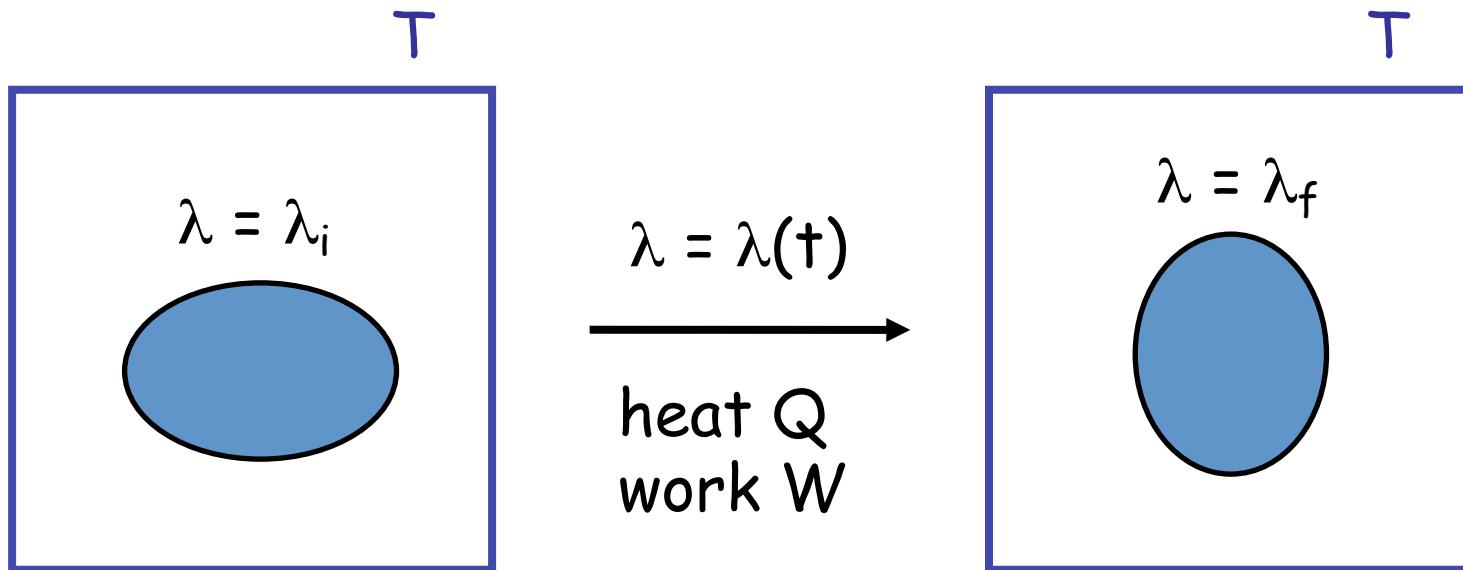
Jarzynski and Crook

Crooks has shown that for **dynamics that is microscopically reversible**, the work distributions $P(W)$ and $P_R(W)$ for the forward and reverse process, respectively, are related by

$$P(W) = P_R(-W) e^{\beta(W - \Delta F)}$$

This exact result, known as the Crooks fluctuation theorem, also serves as a basis for various free energy calculation methods.

Jarzynski's Equality



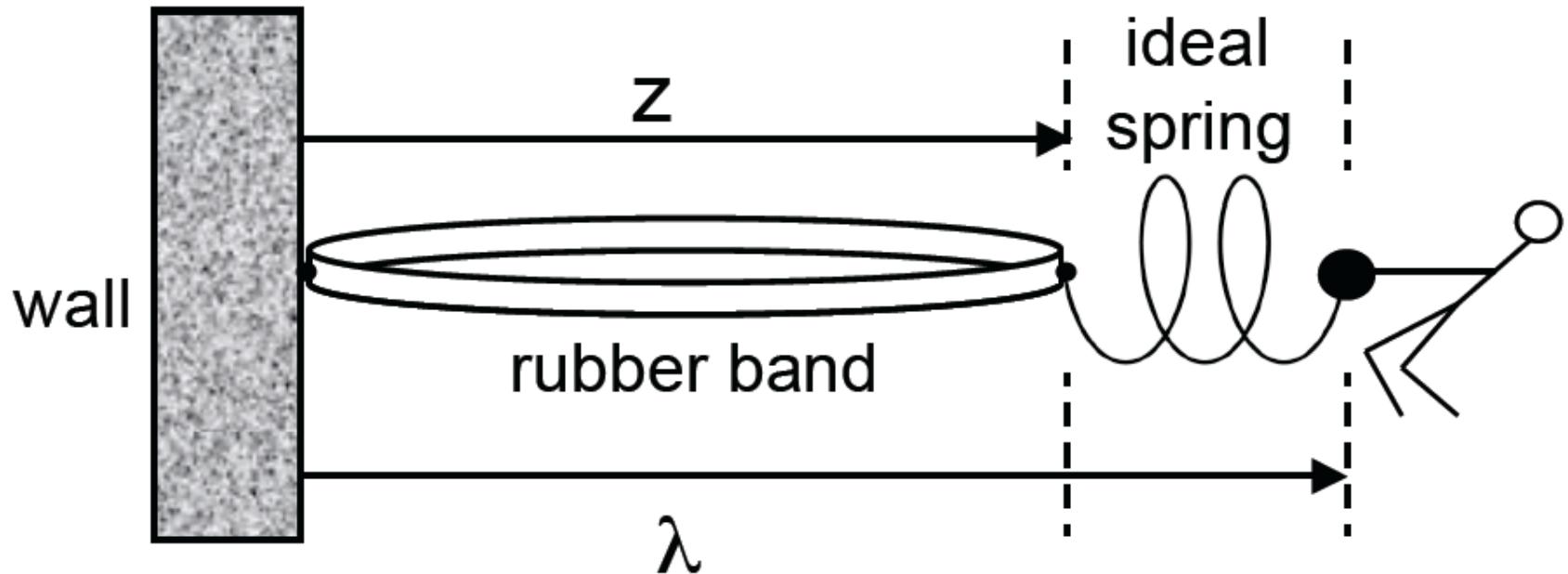
λ = end-to-end distance, position of substrate along a channel, etc.

2nd law of thermodynamics: $\langle W \rangle \geq \Delta F = F(\lambda_f) - F(\lambda_i)$

Jarzynski (1997): $\langle \exp(-\beta W) \rangle = \exp(-\beta \Delta F)$

difficult to estimate

Jarzynski work parameter



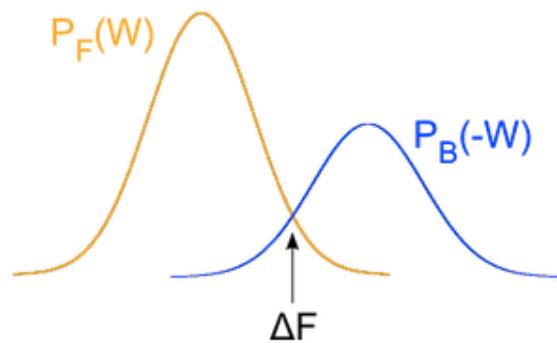
- system of interest = rubber band (+ spring)
- thermal environment = surrounding air ("heat bath", "reservoir")
- work parameter = λ

We act on the system by manipulating the work parameter.

λ fixed \rightarrow system relaxes to equilibrium state (λ, T).

- The Crooks theorem Yet another non-equilibrium relationship, even more fundamental, was put forward by Crooks in 1999. The expression is grounded on microscopic reversibility, and one useful theorem derived from it is the Transient Fluctuation Theorem (TFT). The former relates the ratio of the work distribution for a Markovian process that moves from state A to state B and from state B to a A with the dissipated work involved in the transformation.

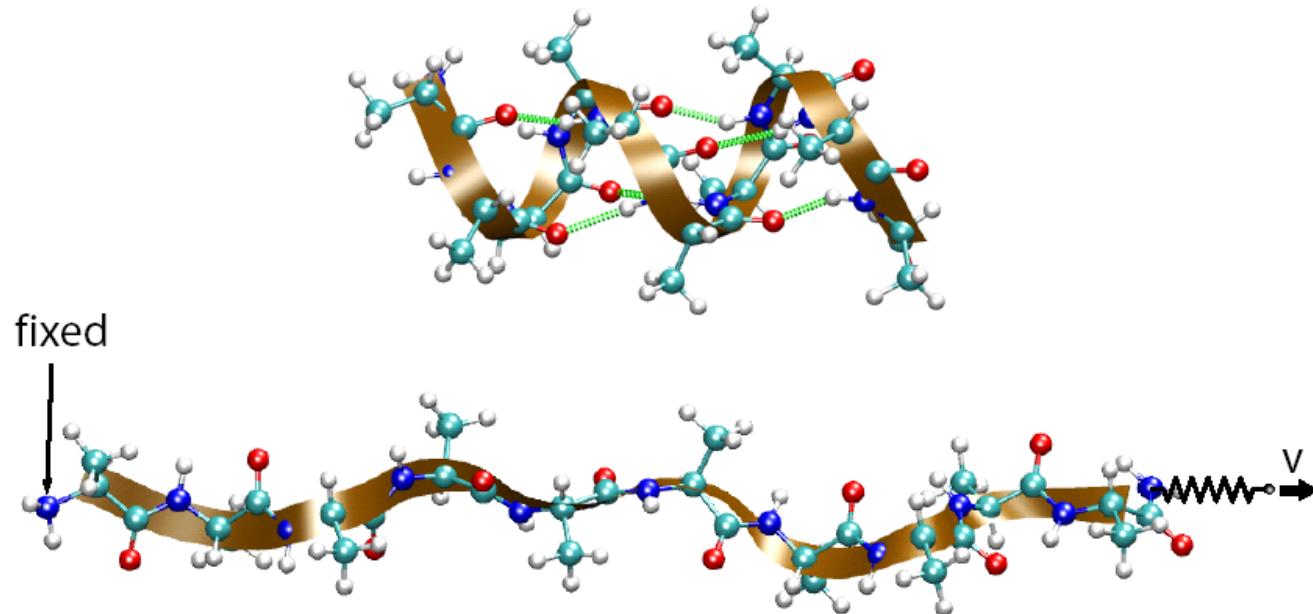
$$\frac{P_F(W)}{P_R(-W)} = e^{\beta(W - \Delta F)}$$



Crooks

$$\ln \left[\frac{P_F(W)}{P_R(-W)} \right] = \beta W - \beta \Delta F$$

Helix-Coil Transition of Deca-Alanine in Vacuum



Main purpose:

Systematic study of the methodology of free energy calculation

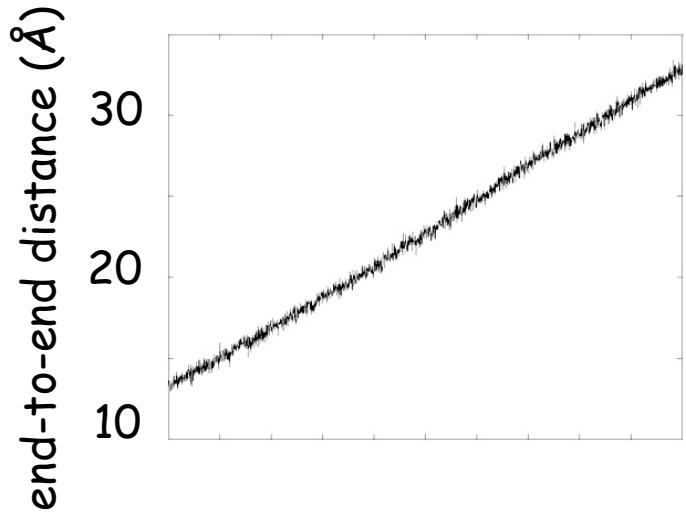
- Which averaging scheme works best
with small number (~ 10) of trajectories ?

Why decaalanine in vacuum?

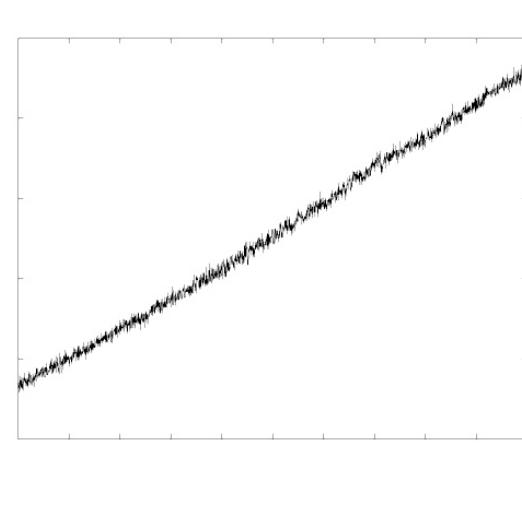
- small, but not too small: 104 atoms
- short relaxation time \rightarrow reversible pulling \rightarrow exact free energy

Typical Trajectories

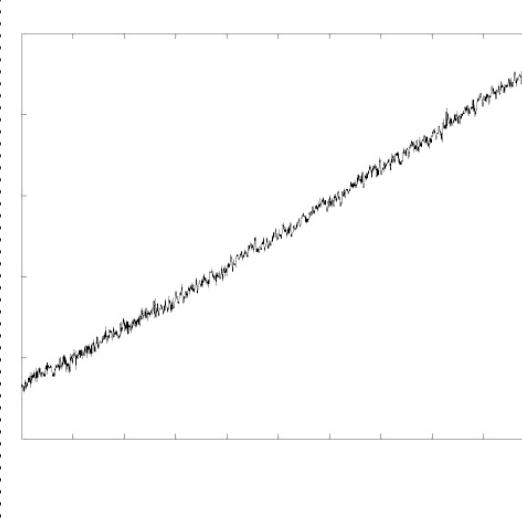
$v = 0.1 \text{ \AA/ns}$



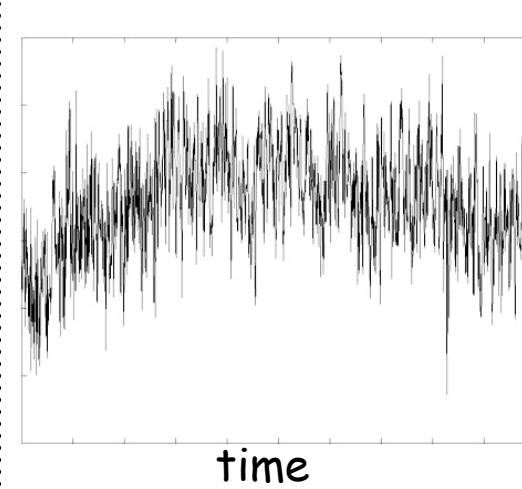
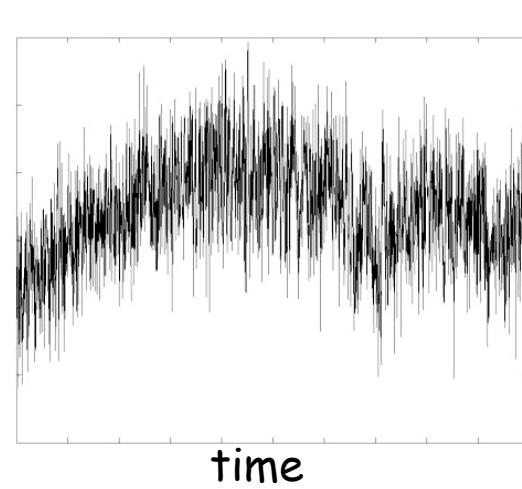
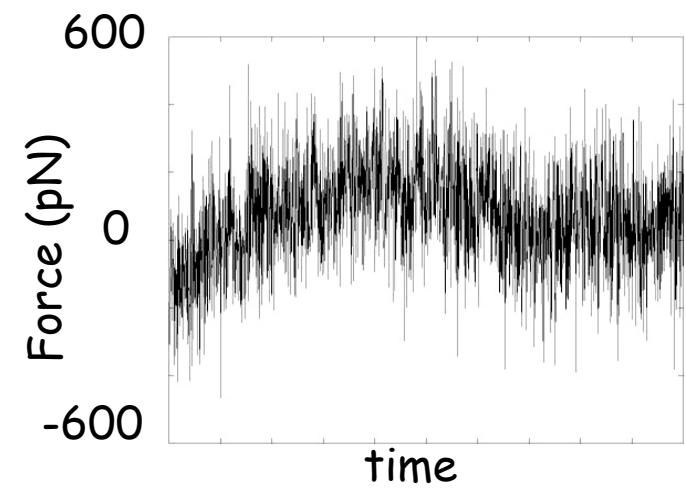
$v = 10 \text{ \AA/ns}$



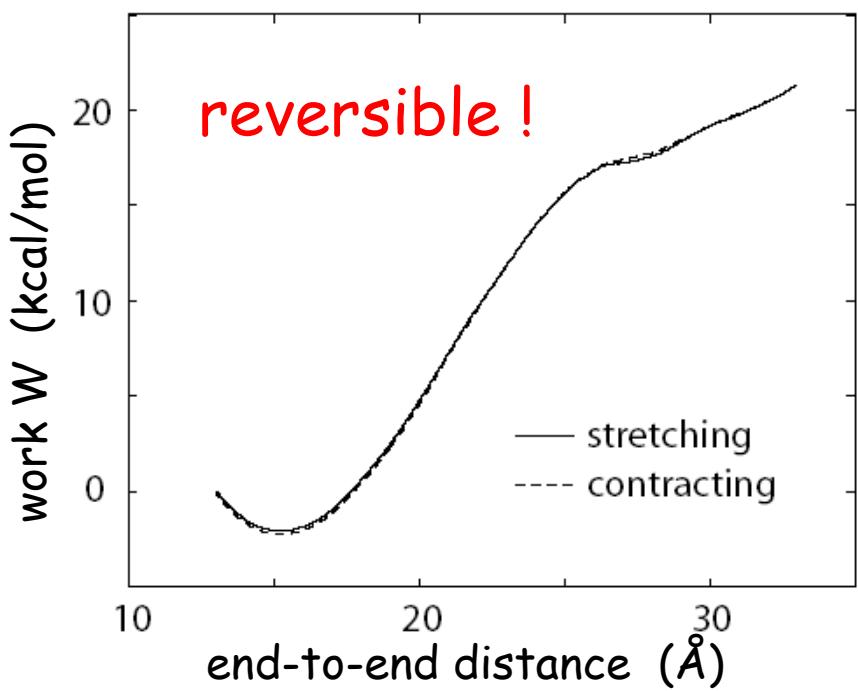
$v = 100 \text{ \AA/ns}$



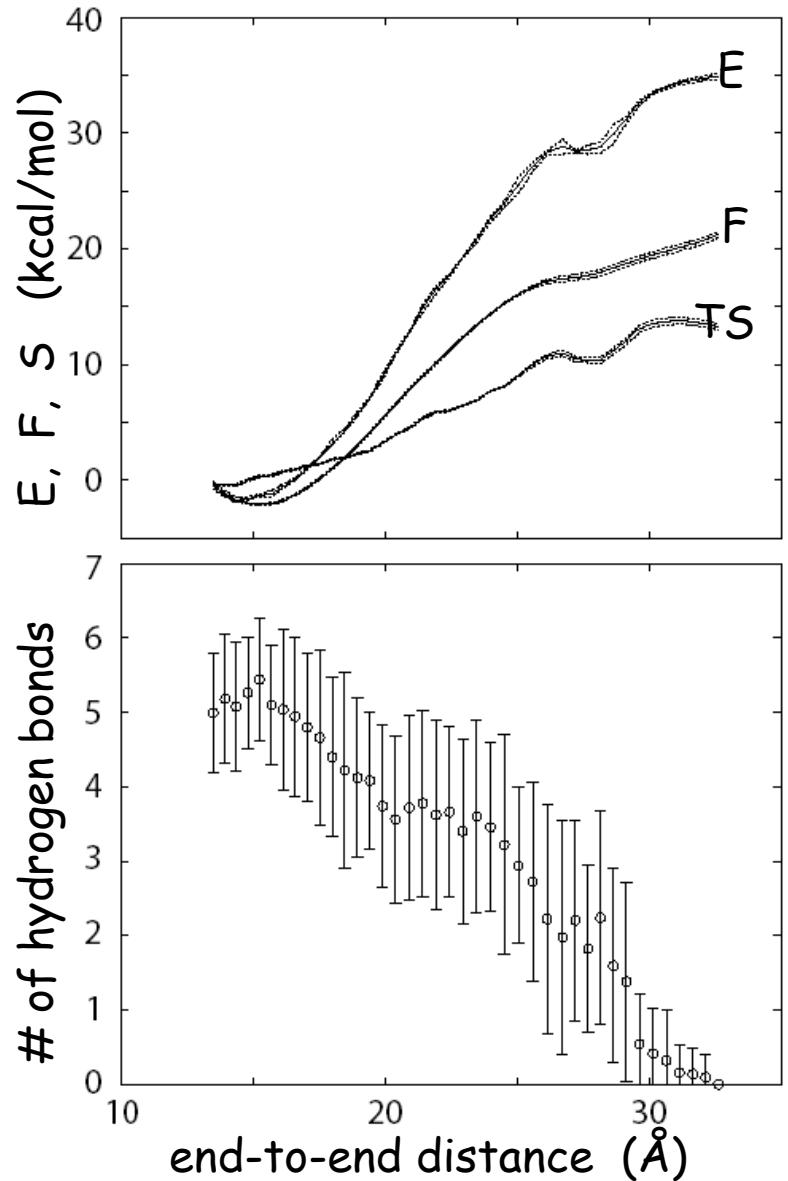
Force (pN)



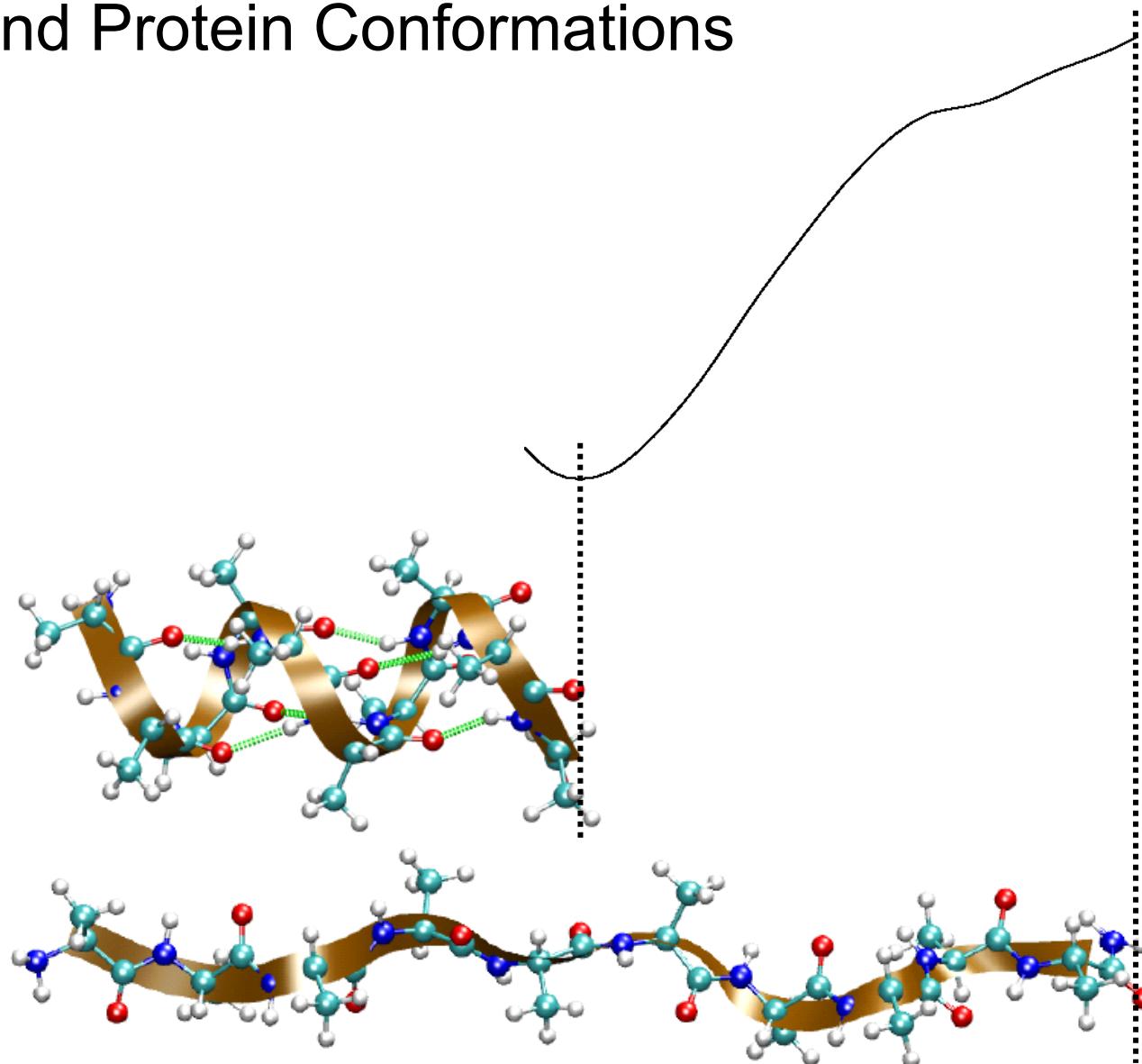
Reversible Pulling ($v = 0.1 \text{ \AA/ns}$)



$$\Delta F = \langle W \rangle$$
$$TS = E - F$$

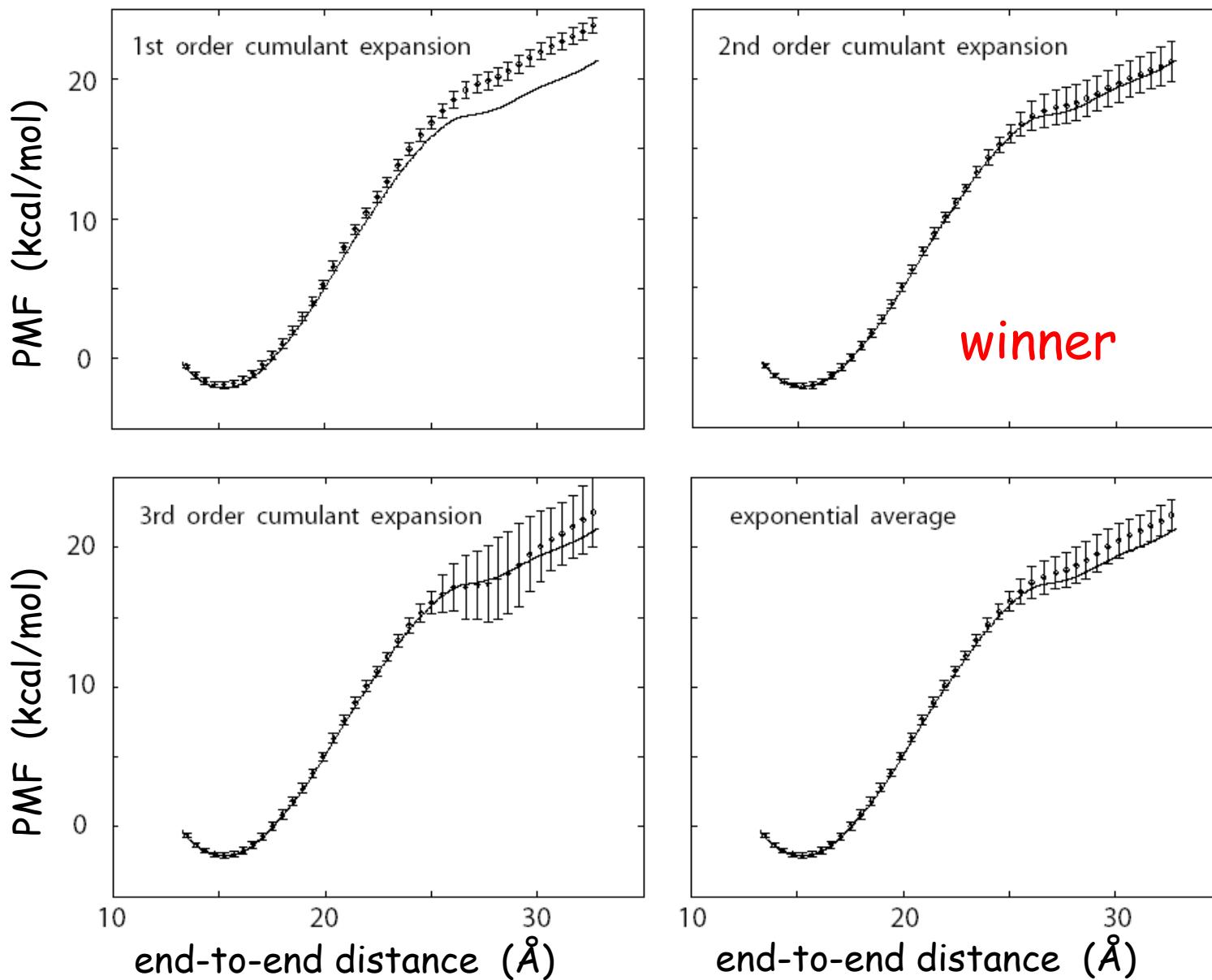


PMF and Protein Conformations



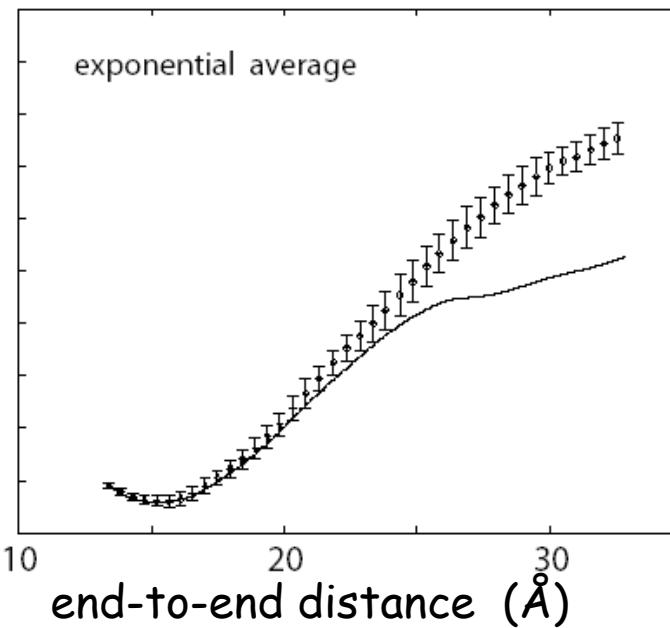
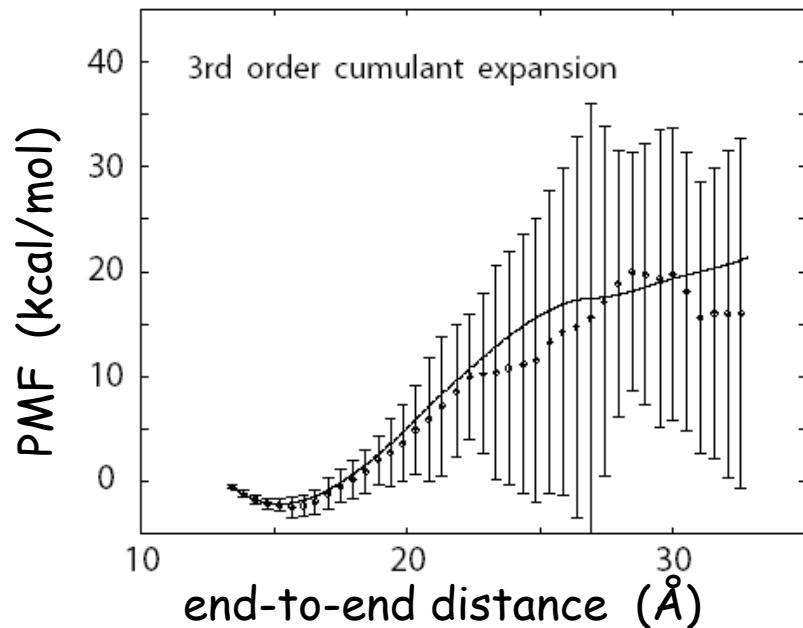
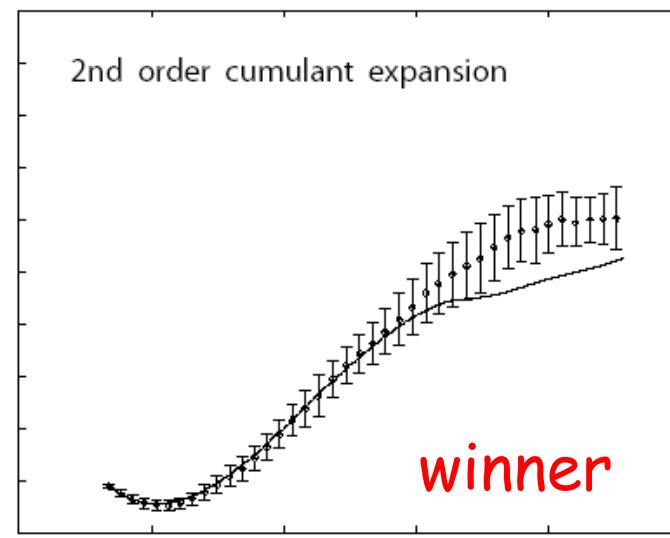
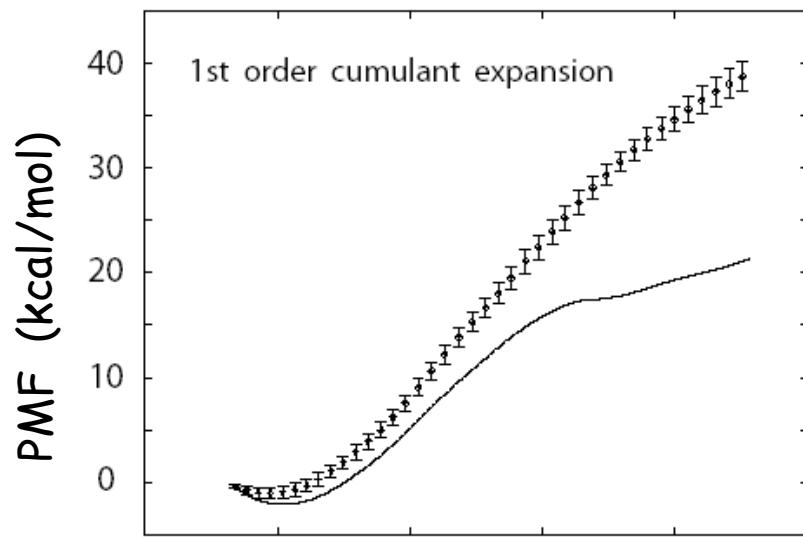
Irreversible Pulling ($v = 10 \text{ \AA/ns}$)

10 blocks of 10
trajectories

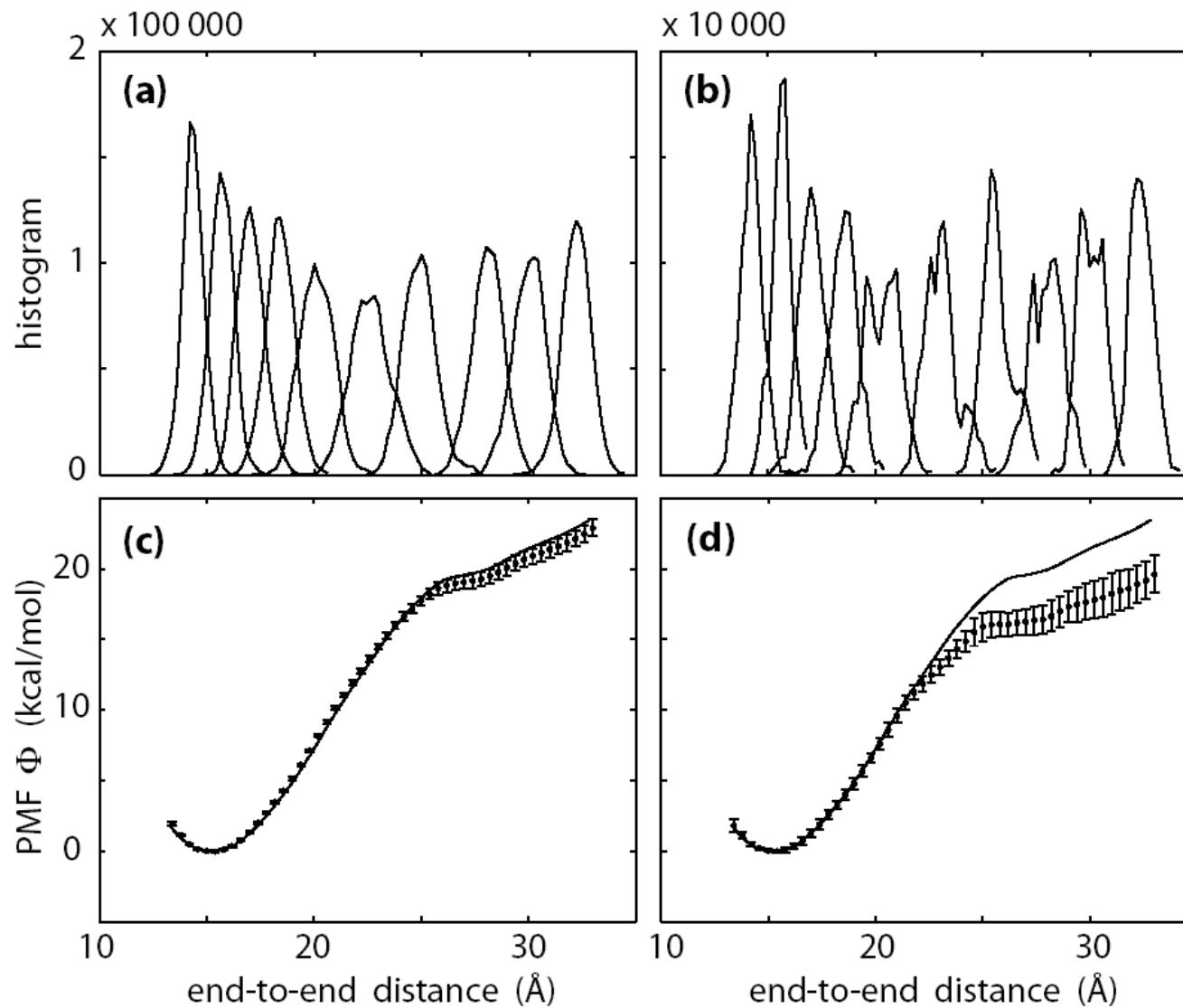


Irreversible Pulling ($v = 100 \text{ \AA/ns}$)

10 blocks of 10
trajectories

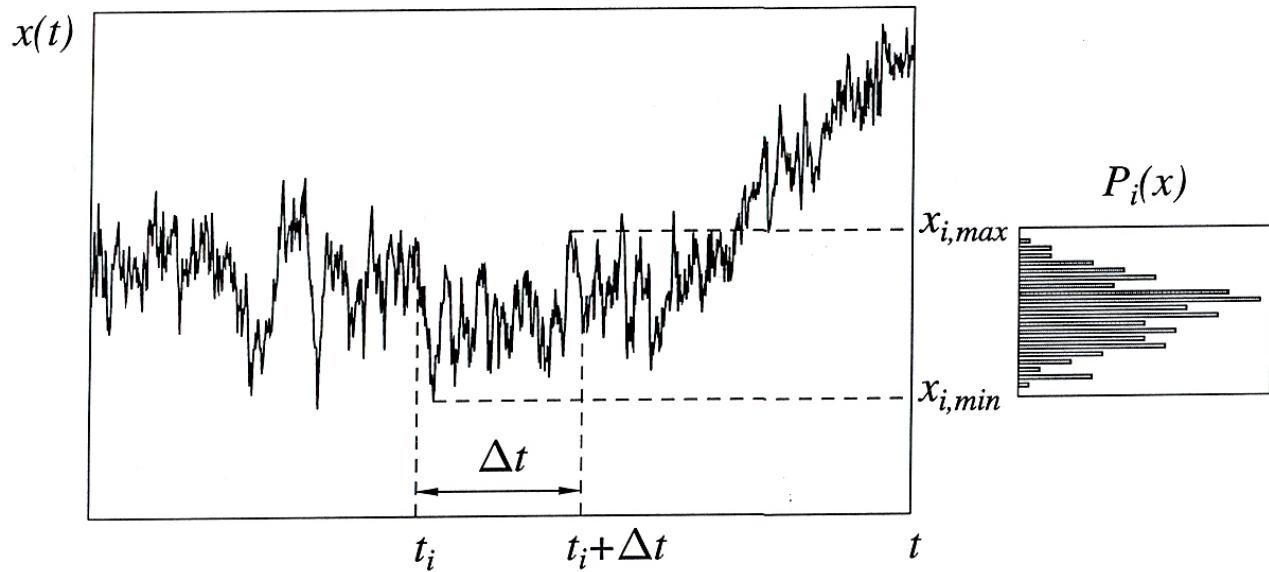


Umbrella Sampling w/ WHAM



Weighted Histogram Analysis Method

SMD
trajectory



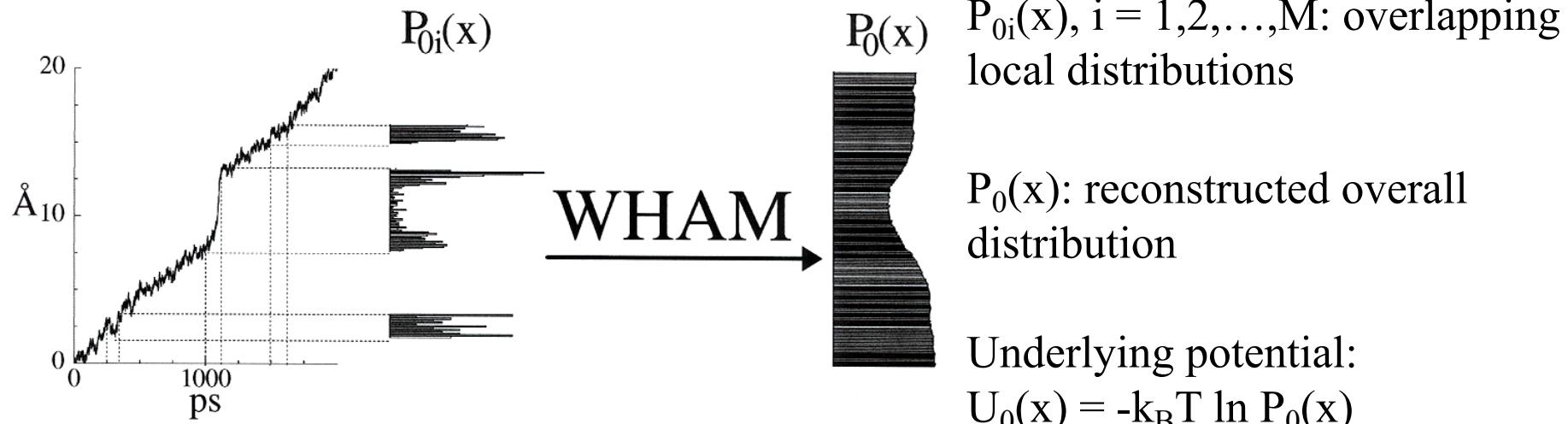
Biasing potential:

$$U_i(x) = \frac{1}{\Delta t} \int_{t_i}^{t_i + \Delta t} \frac{k}{2} (x - vt)^2 dt = \frac{k}{2} \left(x - vt_i - \frac{v\Delta t}{2} \right)^2 + \frac{k(v\Delta t)^2}{24}$$

Choice of Δt :

$$v\Delta t = \delta x, \quad \text{such that} \quad \exp\left(-\frac{k\delta x^2}{2k_B T}\right) \leq \varepsilon \rightarrow 0$$

Weighted Histogram Analysis Method



To reconstruct $P_0(x)$ from $P_{0i}(x)$ ($i=1,2,\dots,M$)

$$P_0(x) = \frac{\sum_{i=1}^M P_{0i}(x) N_i}{\sum_{i=1}^M \frac{Z_0}{Z_{0i}} P_i(x) N_i} ; \quad \frac{Z_0}{Z_{0i}} = \int_{x_0}^{x_f} P_0(x) P_i(x) dx ,$$

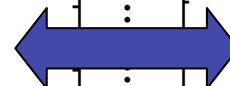
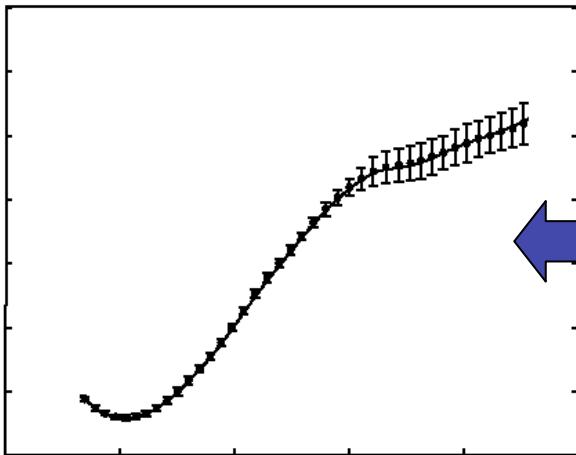
N_i = number of data points in distribution i ,

$$P_i(x) = \frac{1}{\Delta t} \int_{t_i}^{t_i + \Delta t} \exp[-U_s(x, t)/k_B T] dt$$

Biasing potential: $U_s(x, t) = k(x - vt)^2$

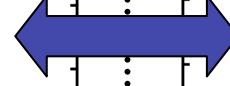
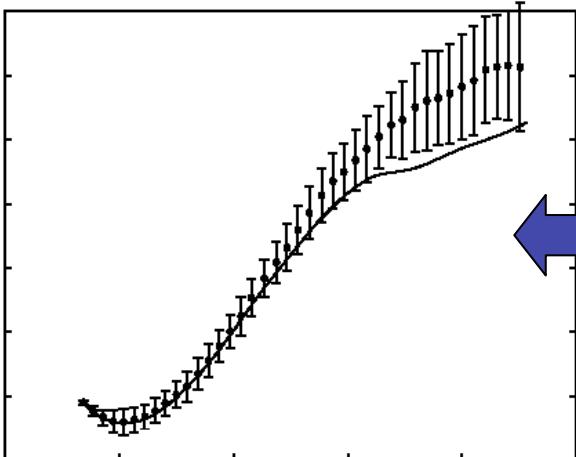
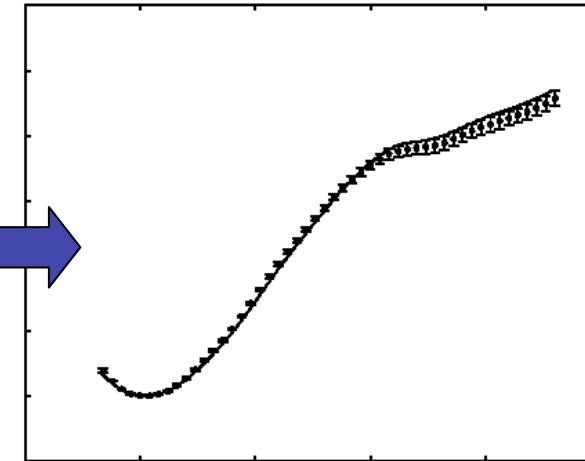
SMD-Jarzynski

(equal amount of simulation time)



Umbrella Sampling

(equal amount of simulation time)



simple analysis: coupled nonlinear equations (WHAM)

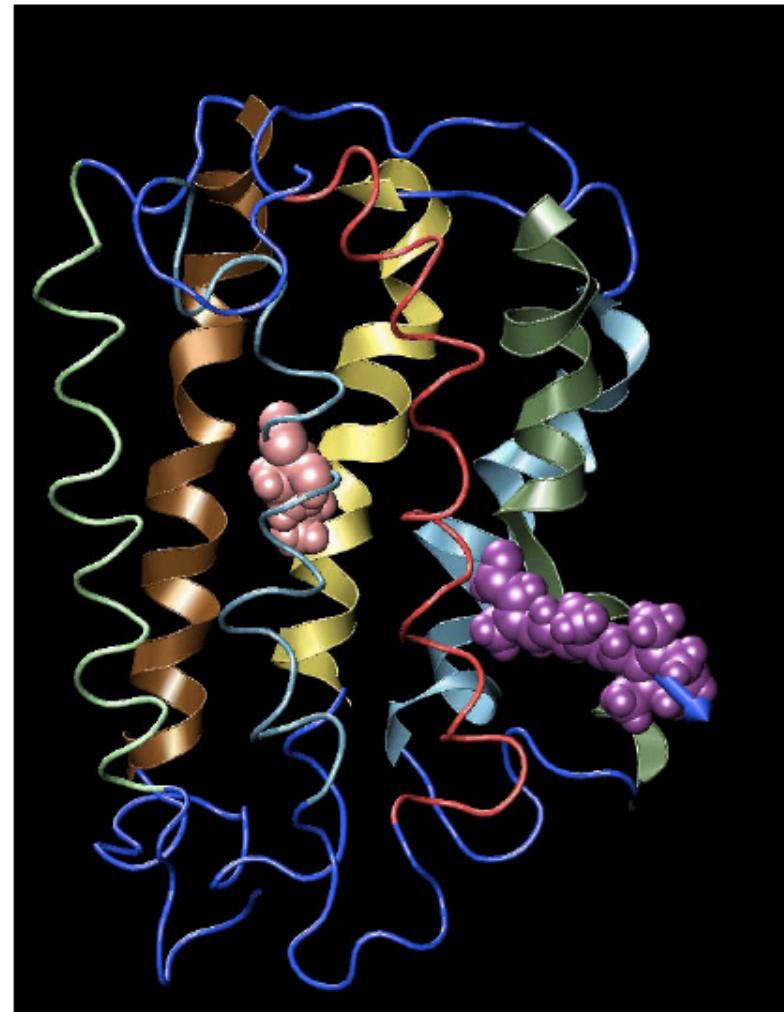
uniform sampling of the reaction coordinate: nonuniform sampling of the reaction coordinate

Steered Molecular Dynamics (SMD)

- ▶ SMD accelerates conformational changes in biomolecular systems through the application of external forces.

Why SMD ?

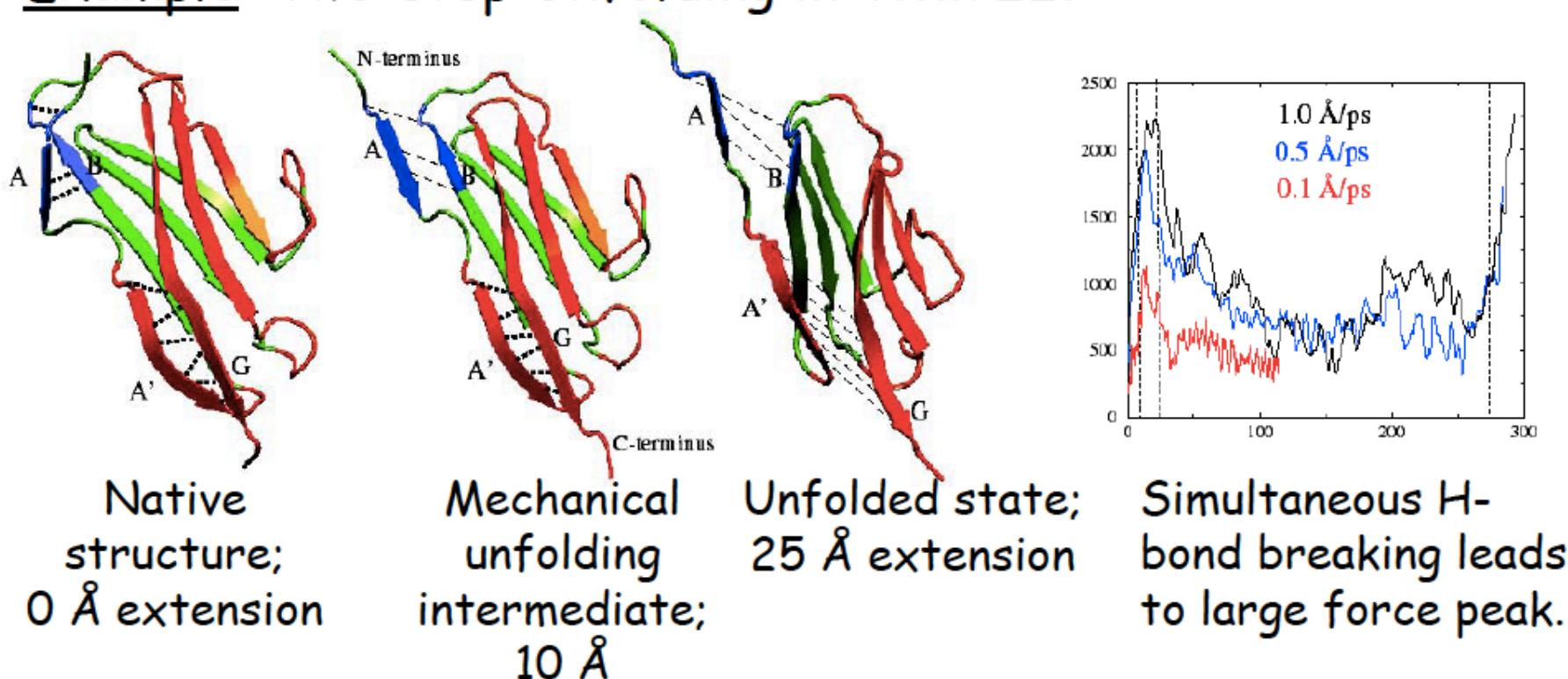
1. Explain single-molecule experiments;
2. Overcome slow natural time scales;
3. Explore putative conformational pathways.



Force Application in SMD

- ▶ Linear forces: a moving harmonic restraint (e.g., pulling with **constant velocity**) or **constant forces** drag selected atoms in a given direction

Example: Two-step Unfolding in Titin I27



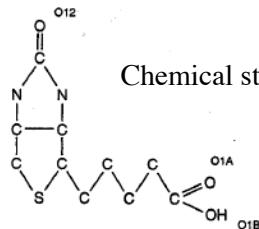
Steps in a typical SMD project

1. *Generate a hypothesis for the conformational transition (i.e., "guess" the reaction coordinate)*
2. *Apply forces that induce the hypothetical process during reasonable wall clock time*
3. *Analyze results in terms of energetic barriers and specific (un)binding events to test the working hypothesis*

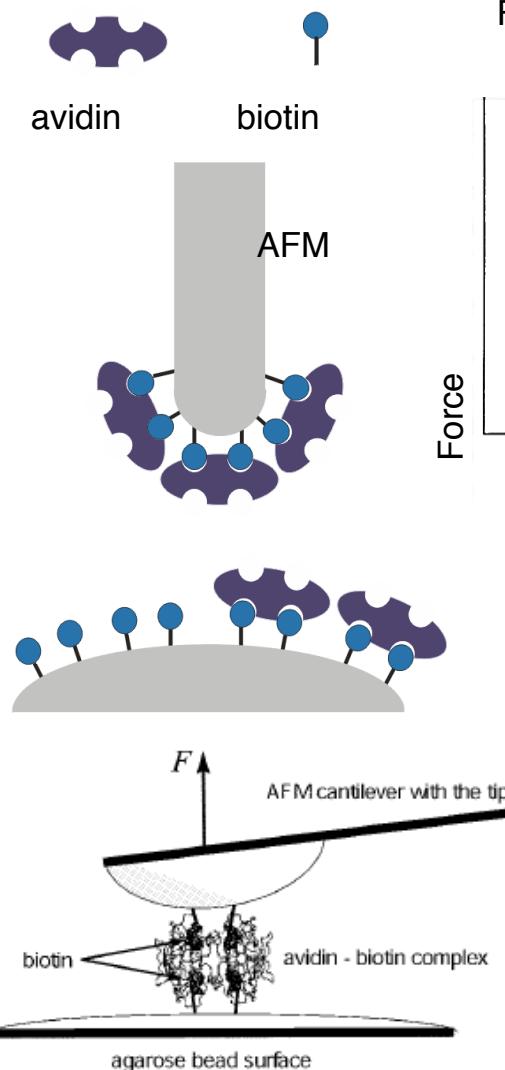
Atomic Force Microscopy Experiments of Ligand Unbinding



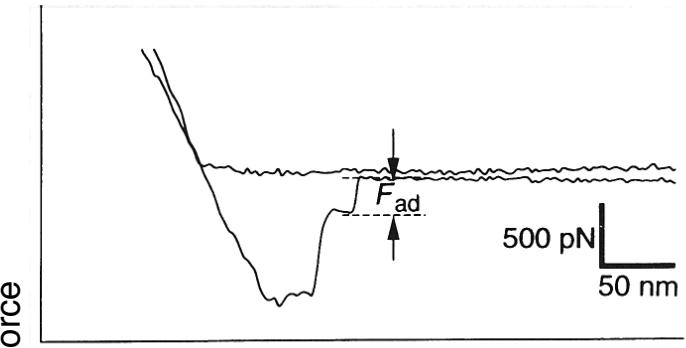
Biotin



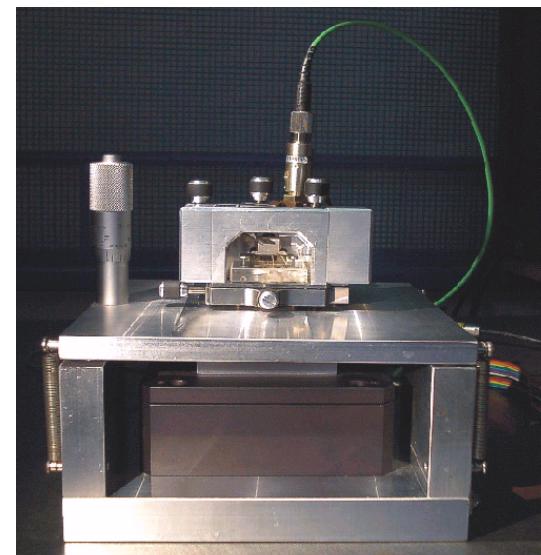
Chemical structure of biotin



Florin et al., Science 264:415 (1994)

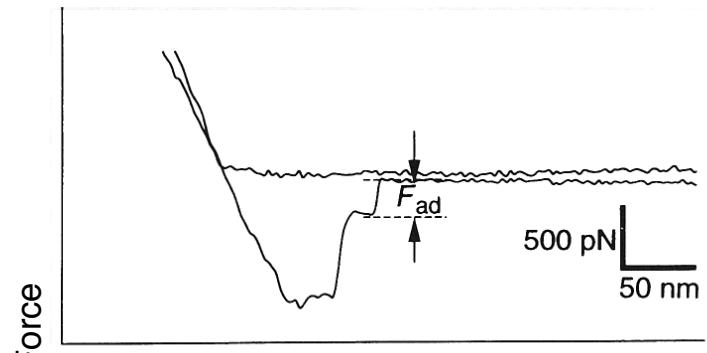
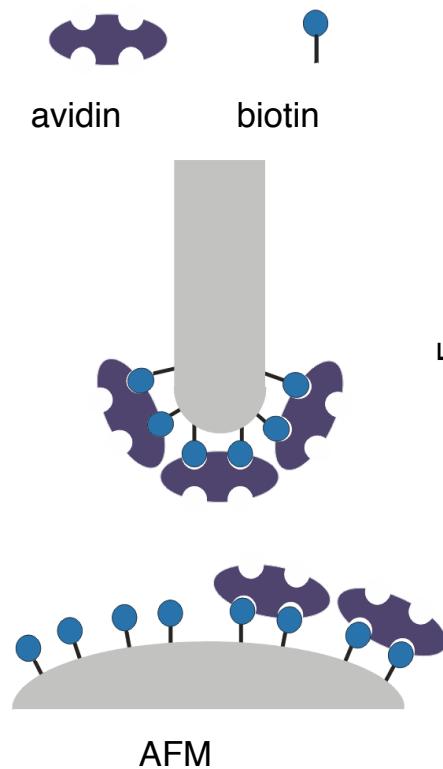
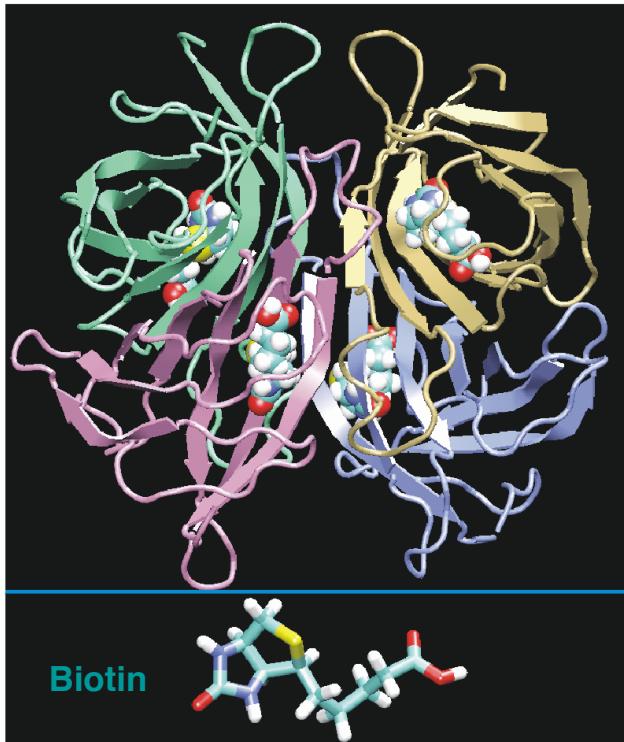


Displacement of AFM tip

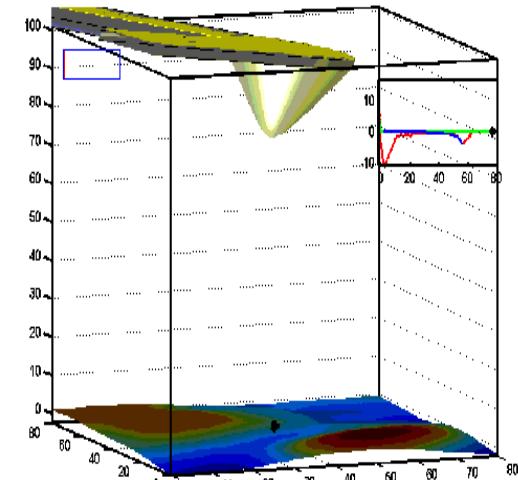


Atomic Force Microscopy Experiments of Ligand Unbinding

Florin et al., Science 264:415 (1994)



Displacement of AFM tip



Test system

Ig-like TITIN I27

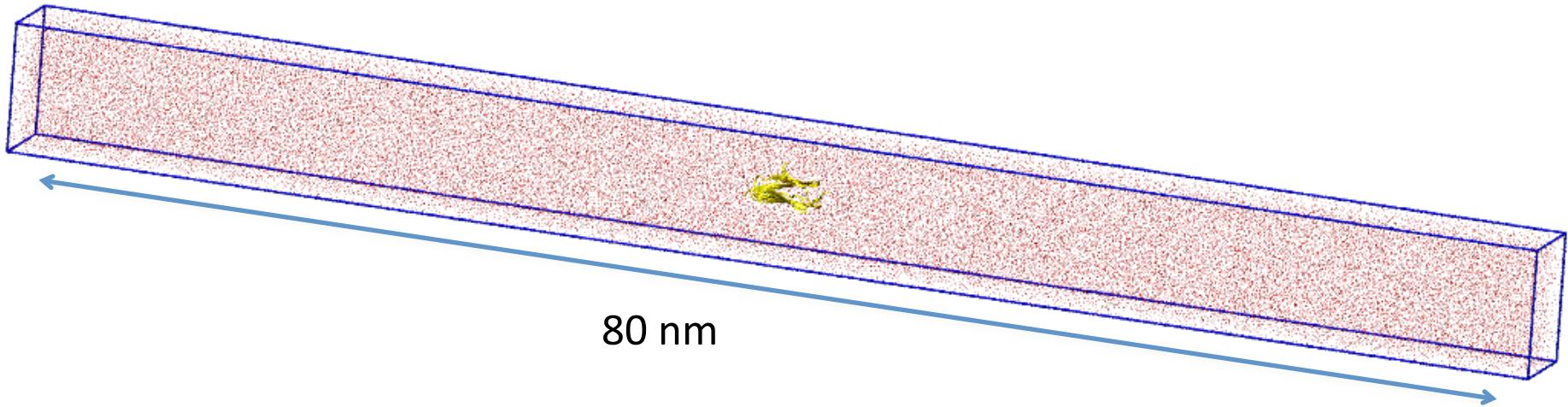
Titin is a giant protein that functions as a molecular spring which is responsible for the passive elasticity of muscle.

It is composed of 244 individually folded protein domains connected by unstructured peptide sequences. These domains unfold when the protein is stretched and refold when the tension is removed



Pulling Ig-like TITIN I27

Initial structure for Pulling: 50 ns standard MD and cluster analysis



System and simulation set up

BOX (6.560 4.362 80.000) nm

228704 atoms

75946 waters

6 Na⁺

GROMOS96 53a6 force field

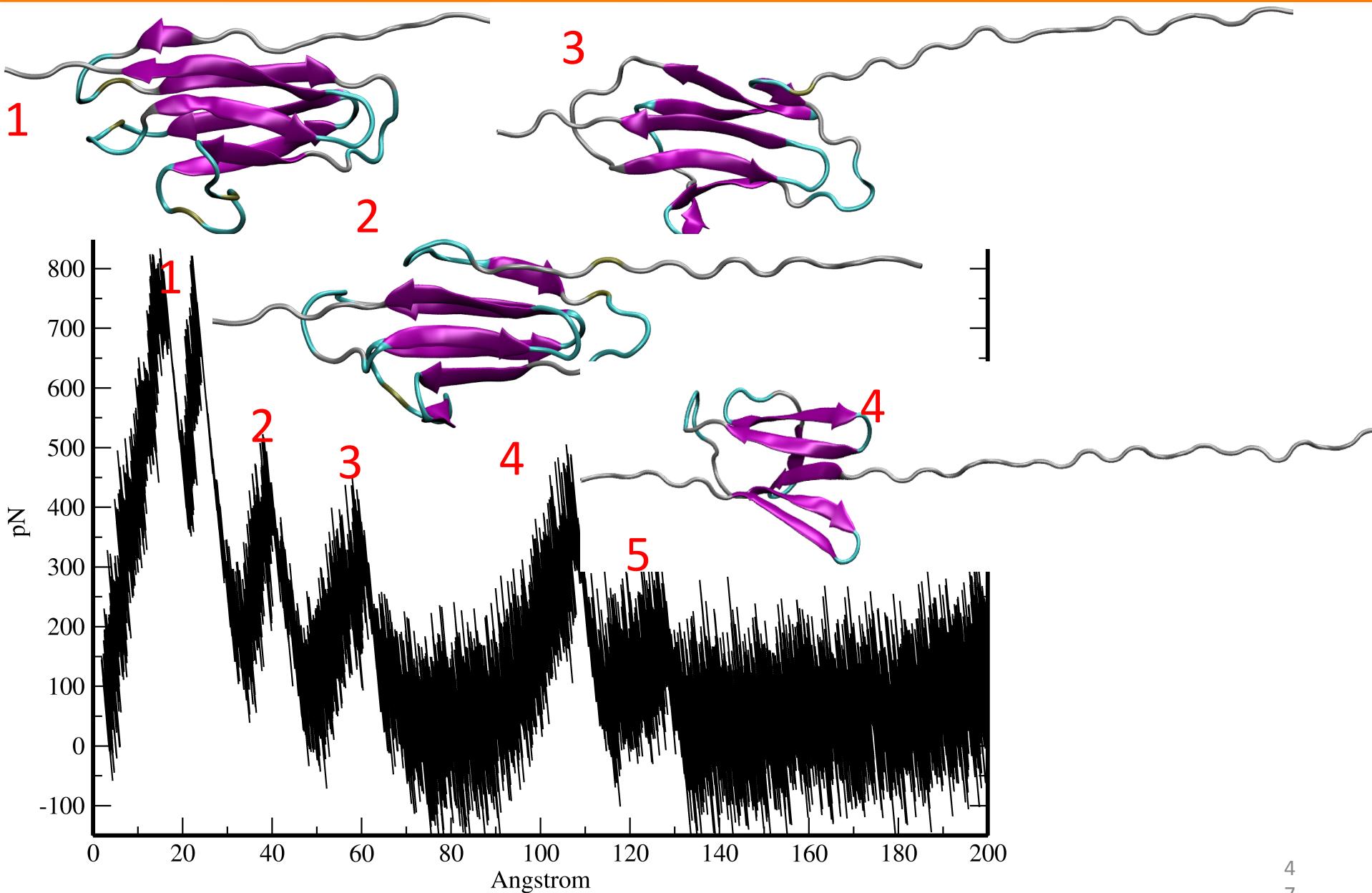
V pulling = 0.0004 nm/ps

$k_0 = 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$

MD with Gromacs 4.5.6

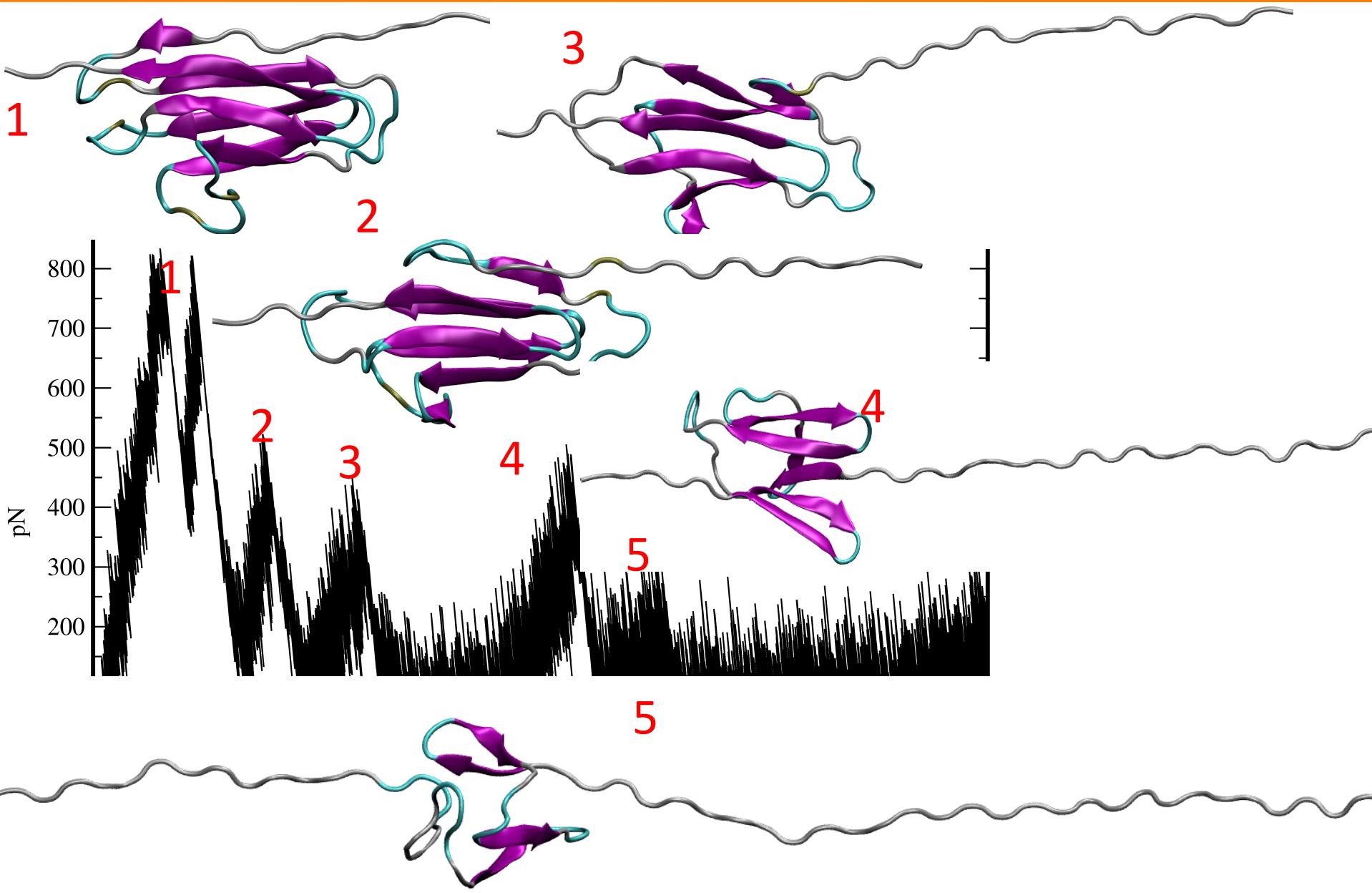
Force-Extension profile

Pulling Ig-like TITIN I27



Force-Extension profile

Pulling Ig-like TITIN I27



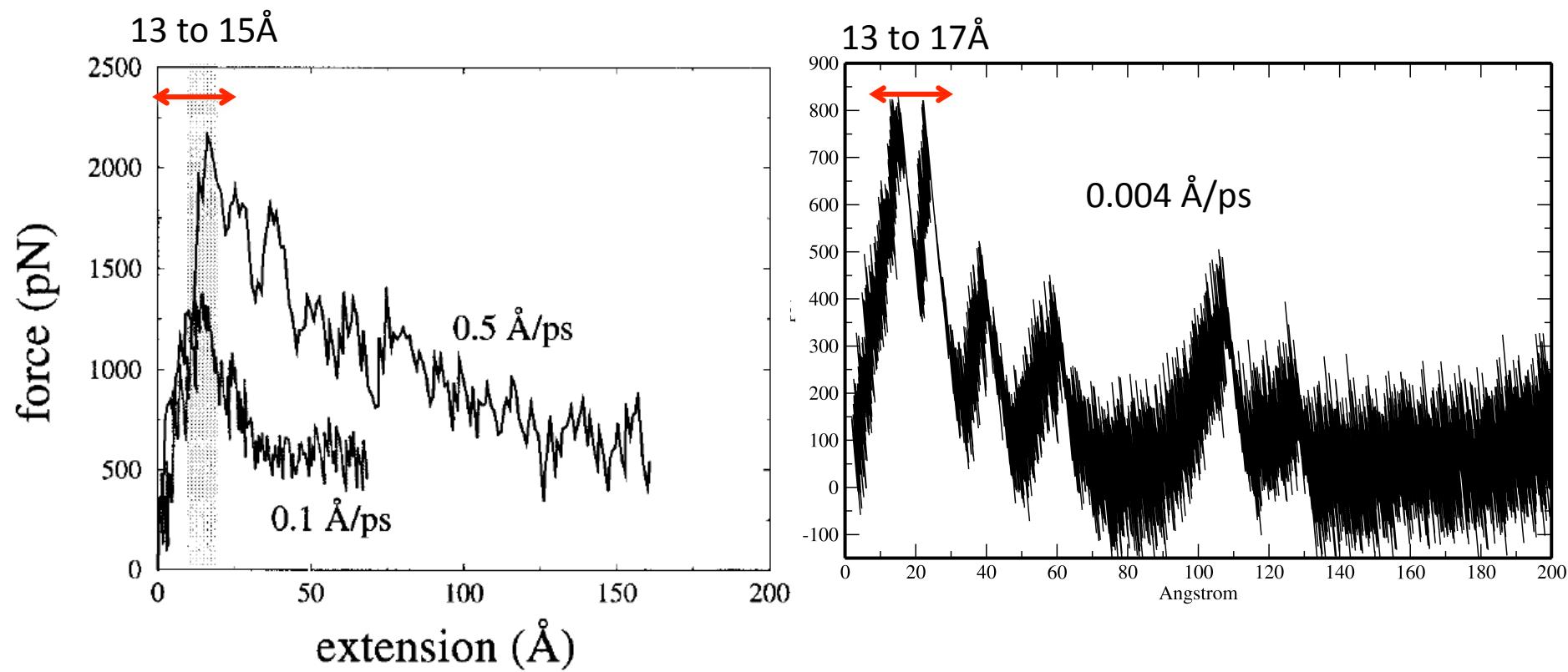
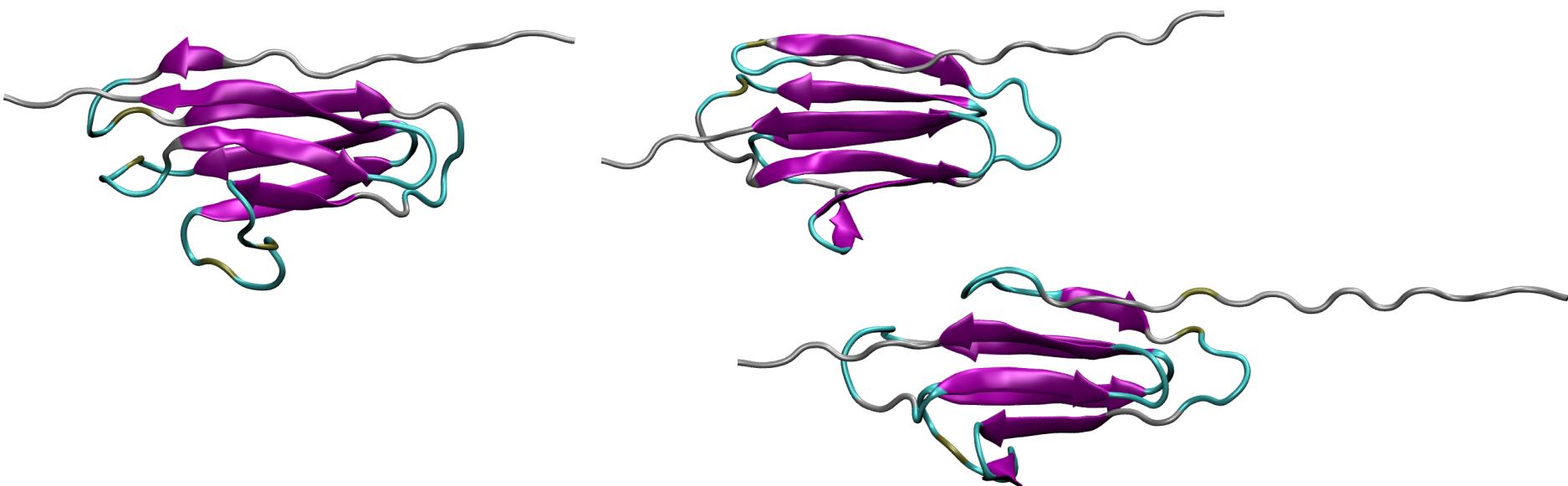
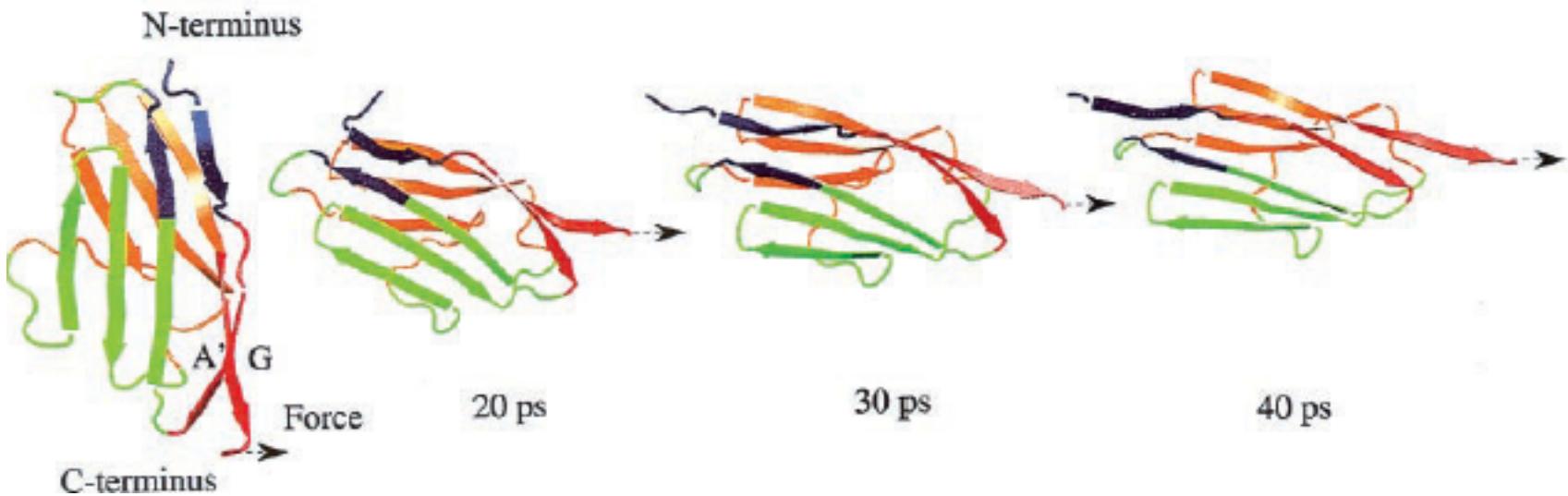
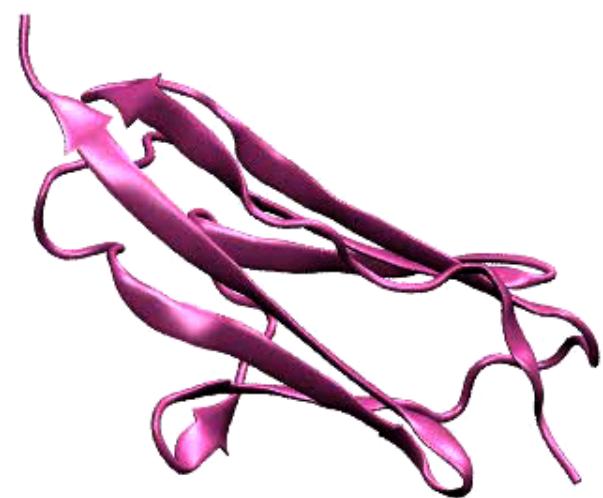


FIGURE 4 Force extension profiles from SMD simulations with constant velocity stretching. In both cases, the dominant force peak is located around extensions of 13 to 15 Å. Simulation SMD-(0.5 Å/ps) had been stopped after 320 ps, when the extension reached 160 Å. Simulation SMD-(0.1 Å/ps) had been stopped after 700 ps, when the domain extension reached 70 Å. The gray area highlights the region where force peak occurs.

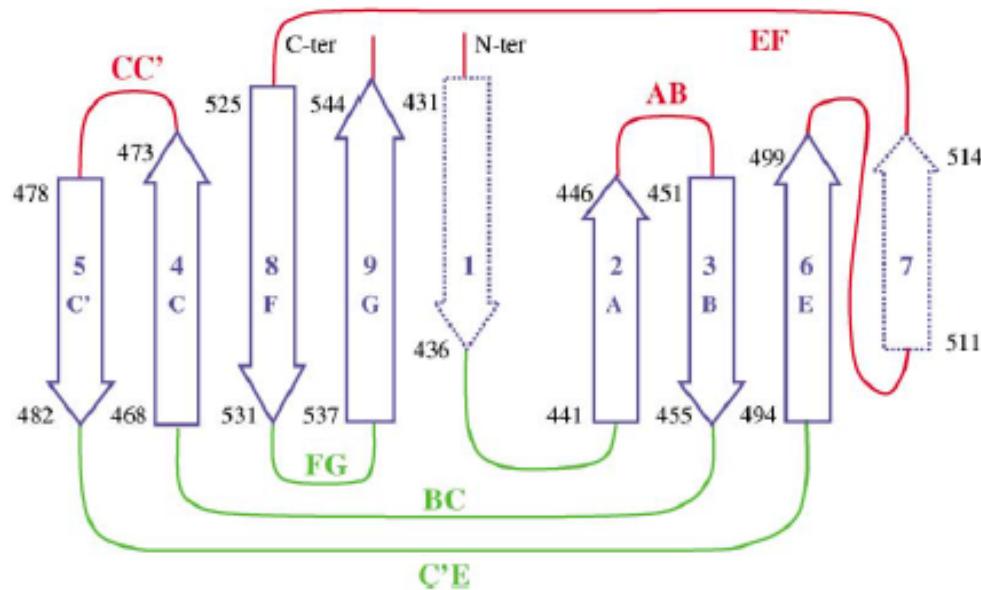
AFM Experimental value 150 to 300 pN

Hui Lu and Klaus Schulten, Biophysical Journal V 79 (2000) p 51–65
 Rief, M. et al., Science V 276 (1997) p 51–65





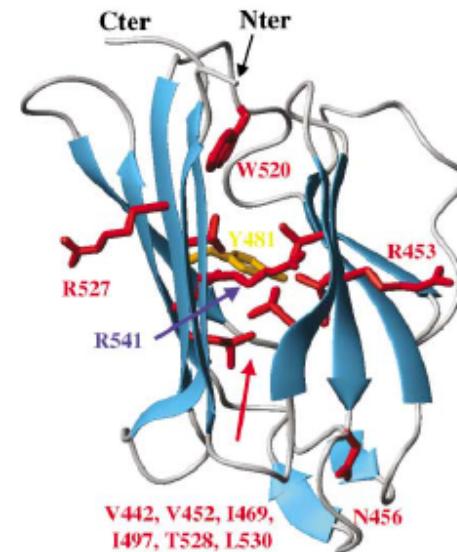
The Ig-like Structure of the C-Terminal Domain of Lamin A/C



It consists of 9 β strands, forming 2 β sheets of 4 and 5 strands, respectively, packed into a classical β sandwich

Strands 2, 3, 6, and 7 form the second β sheet, β 7 being parallel to β 6. Strands 2, 3, 4, 5, 6, 8, and 9 compose the classical Ig fold of type s and correspond to strands A, B, C, C', E, F, and G respectively, following the Ig fold nomenclature. Strands 1 and 7 are additional strands typical of the lamin fold

Mutations assigned to each phenotype



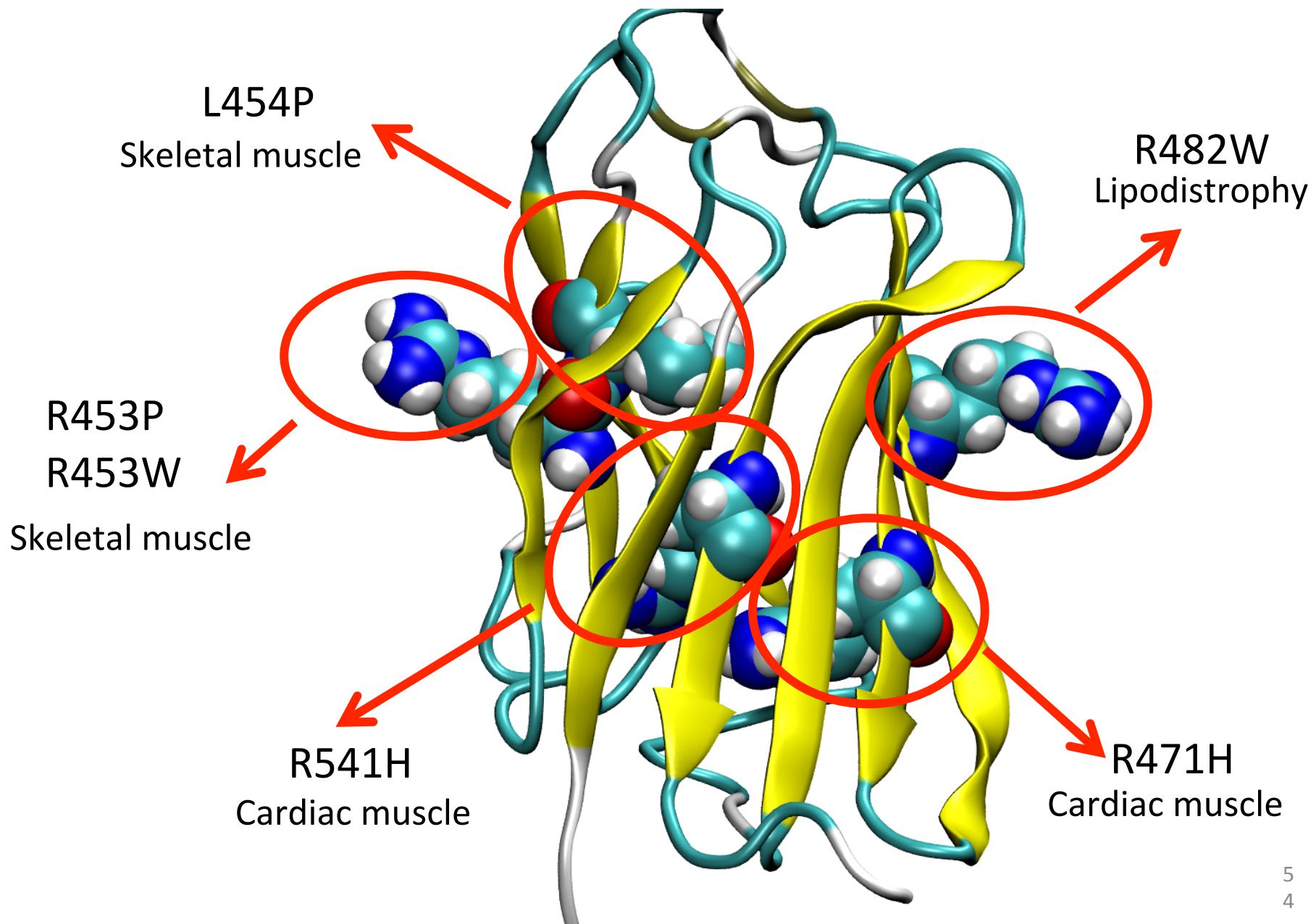
Skeletal muscle

Buried subgroup	Skeletal muscle cluster	Cardiac muscle	Lipodystrophy
21 G449D, G449V, R453P, L454P, N456H, N456D, N456I, N456K, W467R, I469T, Y481H, L489P, W489R, W498C, L512P, W514R, W520G, R527P, T528K, T528R, L530P	6 D446V, R453W, R455P, N459S, W520S, R545C	10 V440M, A441V, D461Y, R471H, T488P, G523R, R541S, R541G, R541C, R541H	9 R439C, G465D, R471G, R482W, R482Q, R482L, P485R, K486N, H506D,

Standard MD and Pulling of WT and mutations comparing with Titin IG domain

IG domain

LAMIN WT and some mutations

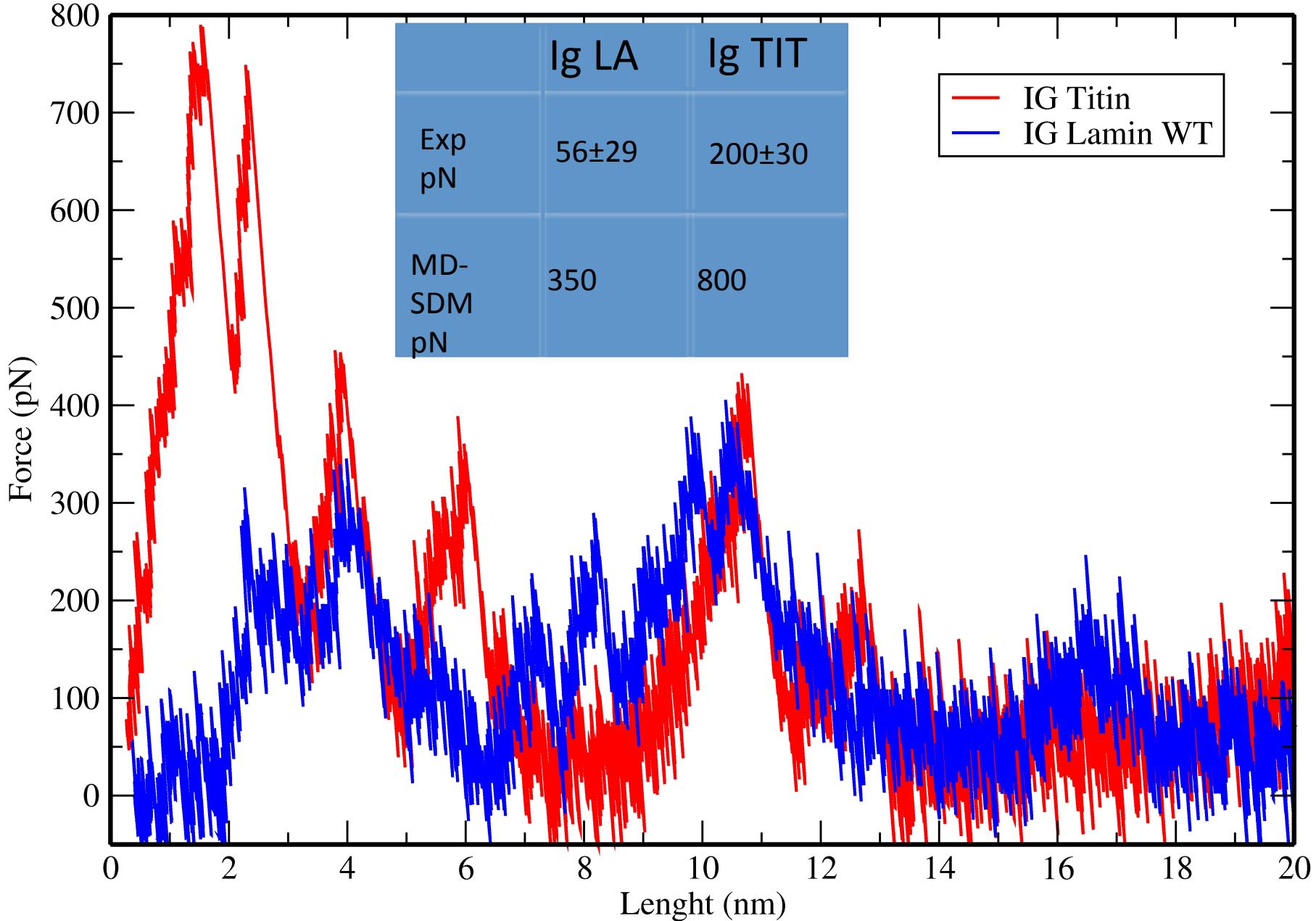


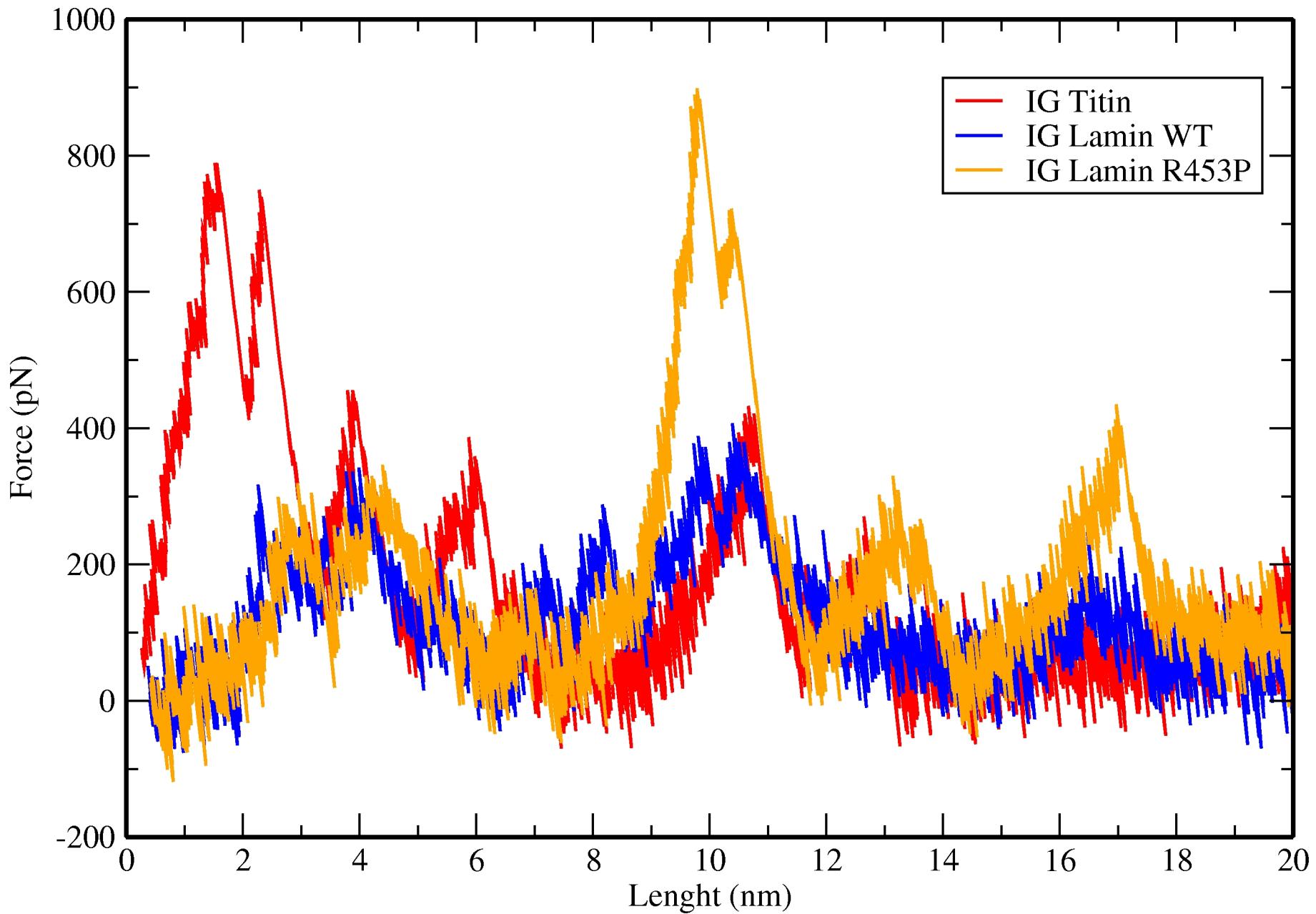
PDB ID: 1ZC0
Ligand: 2'-deoxyguanosine triphosphate





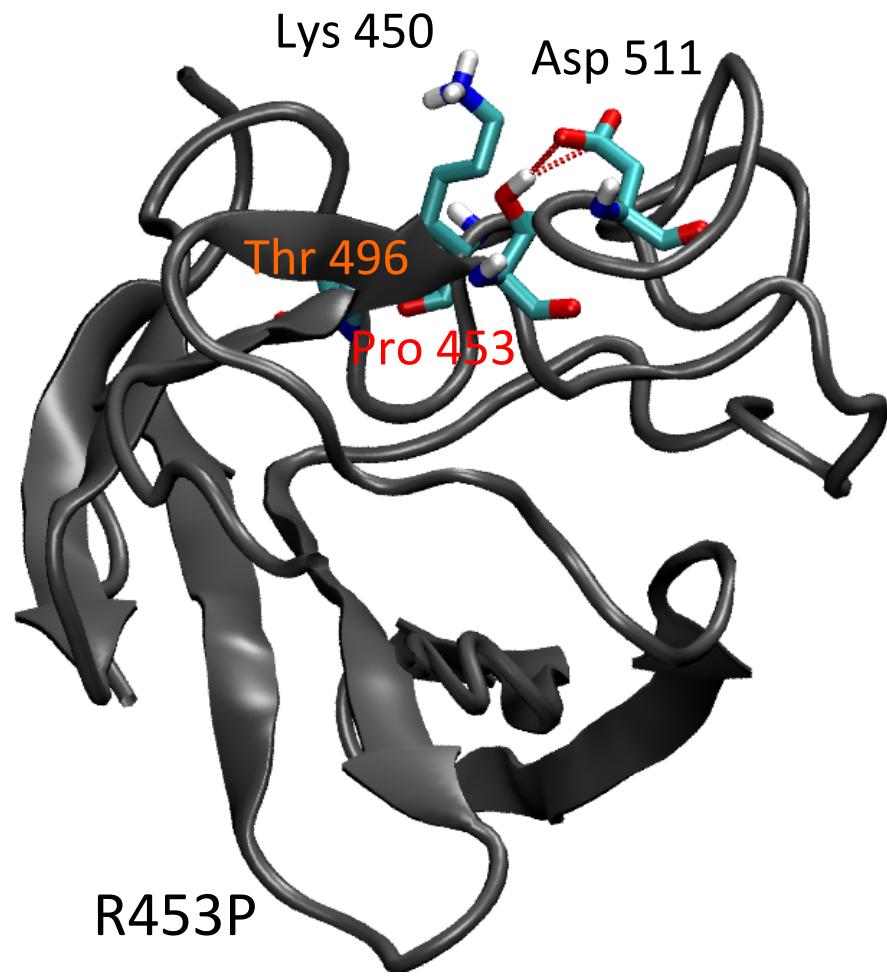
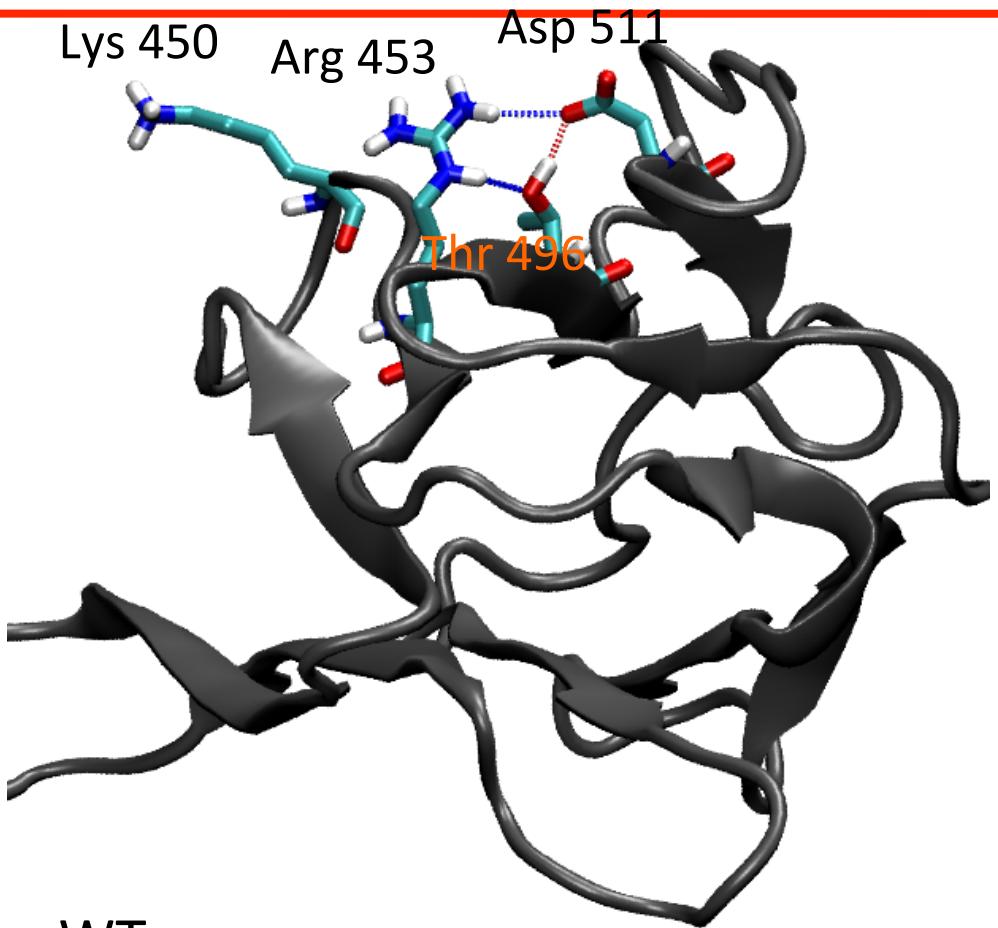
Mechanical Unfolding of lamin Ig domain and stability of Ig-emerin complex





IG domain

LAMIN WT and some mutations

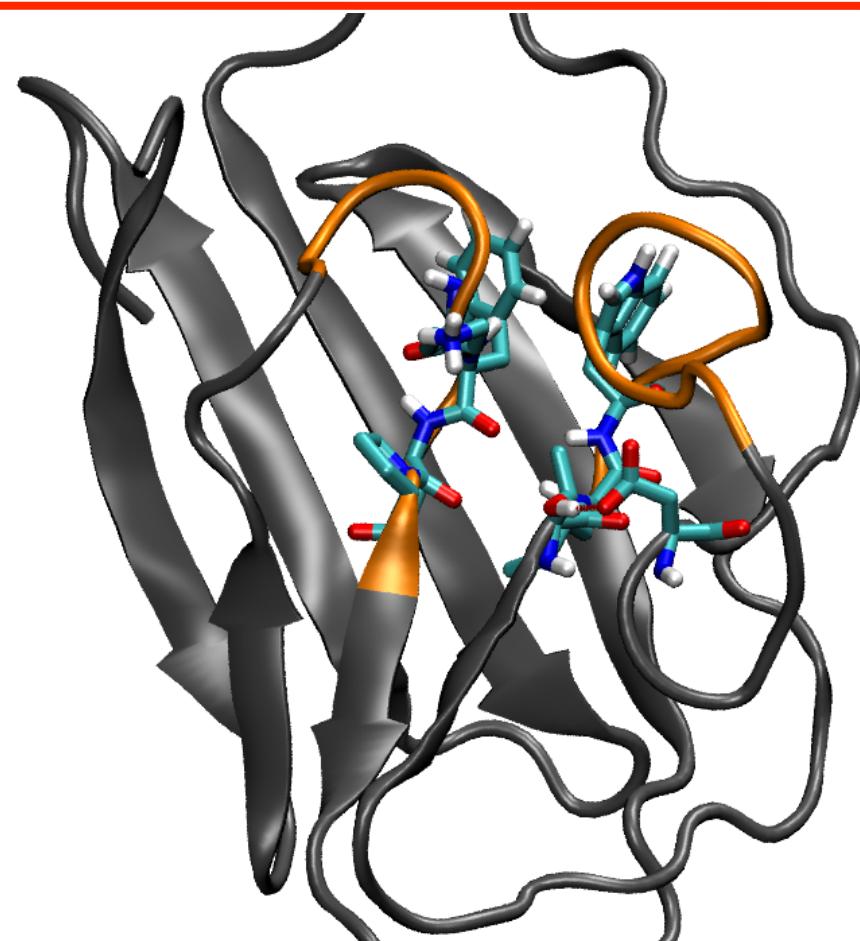
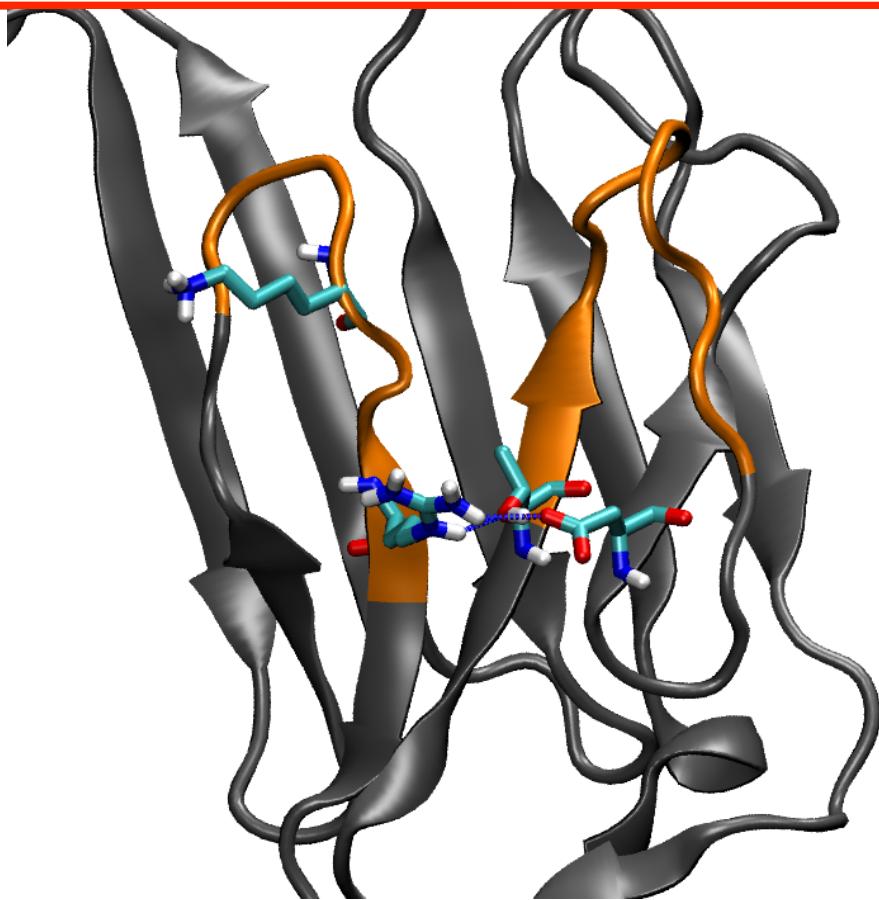


Interaction Arg 453 – Asp 511
Arg 453 – Thr 496

Collaps of the region GLY449, LYS450, PHE 451, VAL 452 / ILE497, TRP498, ALA499, ALA 500, GLY501, ALA502, GLY503, ALA504, THR505, HIS 506

IG domain

LAMIN WT and some mutations



Formation of new backbone-backbone H bonds
interactions

PHE451-TRP458

VAL452-ILE497

Readings

Jarzynski and Crooks equations in non-equilibrium systems

- Crooks, G. E. (1998), "Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems", *J. Stat. Phys.* **90**: 1481,
- Jarzynski, C. (1997), "Nonequilibrium equality for free energy differences", *Phys. Rev. Lett.* **78**: 2690, [arXiv:cond-mat/9610209](https://arxiv.org/abs/cond-mat/9610209), Bibcode:[1997PhRvL..78.2690J](https://doi.org/10.1103/PhysRevLett.78.2690), doi:[10.1103/PhysRevLett.78.2690](https://doi.org/10.1103/PhysRevLett.78.2690)
- Jarzynski, C. (1997), "[Equilibrium free-energy](#) differences from nonequilibrium measurements: A master-equation approach", *Phys. Rev. E* **56**: 5018, [arXiv:cond-mat/9707325](https://arxiv.org/abs/cond-mat/9707325), Bibcode:[1997PhRvE..56.5018J](https://doi.org/10.1103/PhysRevE.56.5018), doi:[10.1103/PhysRevE.56.5018](https://doi.org/10.1103/PhysRevE.56.5018)

Steered Molecular Dynamics

http://www.ks.uiuc.edu/Research/smd_imd/

Park S, Schulten K. Calculating potentials of mean force from steered molecular dynamics simulations. *J Chem Phys.* 2004 Apr 1;120(13):5946-61.

Sotomayor M, Schulten K. Single-molecule experiments in vitro and in silico. *Science.* 2007 May 25;316(5828):1144-8.

Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol.* 2001 Apr;11(2):224-30.