Reconstruction of Vectorial Protein Folding Pathways by Atomic Force Microscopy and Molecular Dynamics Simulations

Piotr E. Marszalek
Center for Biologically Inspired Materials and Material Systems
Department of Mechanical Engineering
and Materials Science
Duke University
Box 90300
Durham, NC 27708
pemar@duke.edu
http://smfs.pratt.duke.edu/homepage/lab.htm

DUKE
ENGINEERING
EDMUND T. PRATT, JR.
SCHOOL OF ENGINEERING

"Can we predict how proteins will fold?"

This question was listed in 2005 as one of the 125 most important unsolved problems in science by the *Science* magazine

So much more to know.... *Science 309*, 78-102 (2005).

Atomic-Level Characterization of the Structural Dynamics of Proteins

David E. Shaw,[1,2]* Paul Maragakis,[1]† Kresten Lindorff-Larsen,[1]† Stefano Piana,[1]† Ron O. Dror,[1] Michael P. Eastwood,[1] Joseph A. Bank,[1] John M. Jumper,[1] John K. Salmon,[1] Yibing Shan,[1] Willy Wriggers[1]

SCIENCE VOL 330 15 OCTOBER 2010

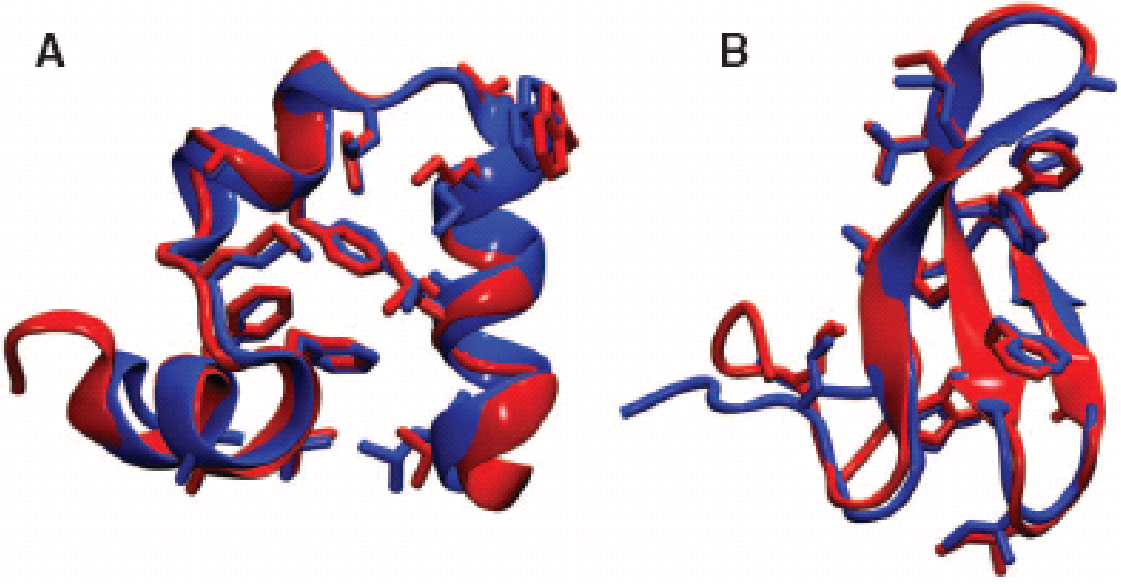


Fig. 1. Folding proteins at x-ray resolution, showing comparison of x-ray structures (blue) (*15*, *24*) and last frame of MD simulation (red): (**A**) simulation of villin at 300 K, (**B**) simulation of FiP35 at 337 K. Simulations were initiated from completely extended structures. Villin and FiP35 folded to their native states after 68 μs and 38 μs, respectively, and simulations were continued for an additional 20 μs after the folding event to verify the stability of the native fold.

Part I

Combining AFM-based single-molecule force spectroscopy (SMFS) with steered molecular dynamics simulations to examine vectorial folding of proteins

Part II

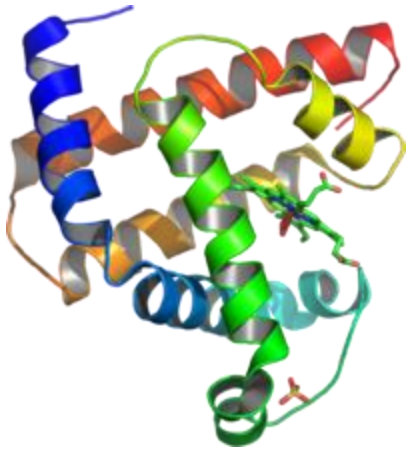
Developing protein-based SMFS probes for characterizing the strength of protein-protein interactions

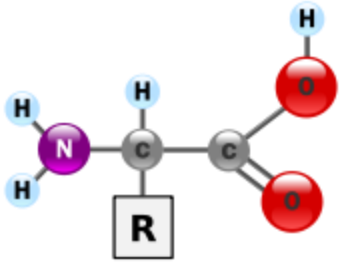
Part III

Creating novel protein constructs with unusual folding and mechanical properties and engineering protein based materials

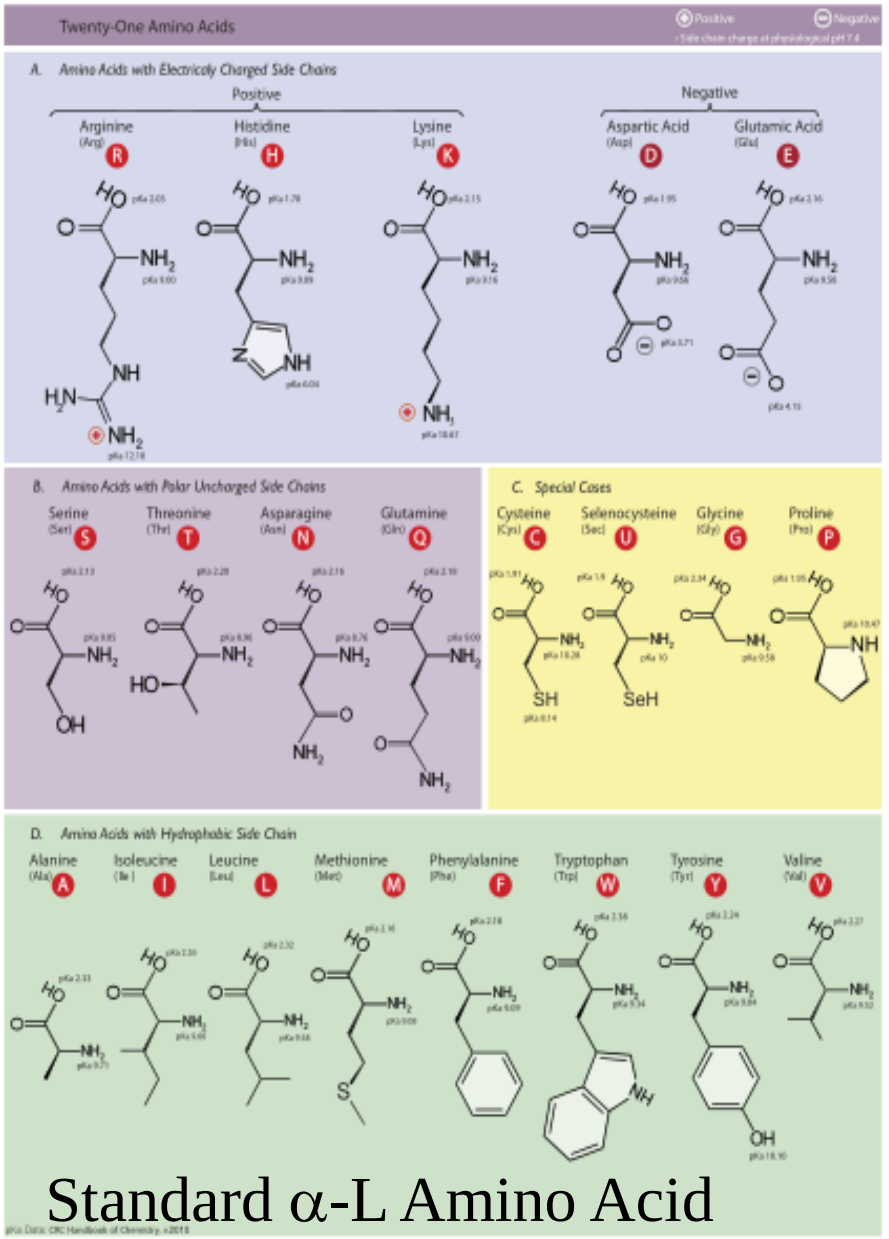
Overview of Proteins Structure

Protein Myoglobin

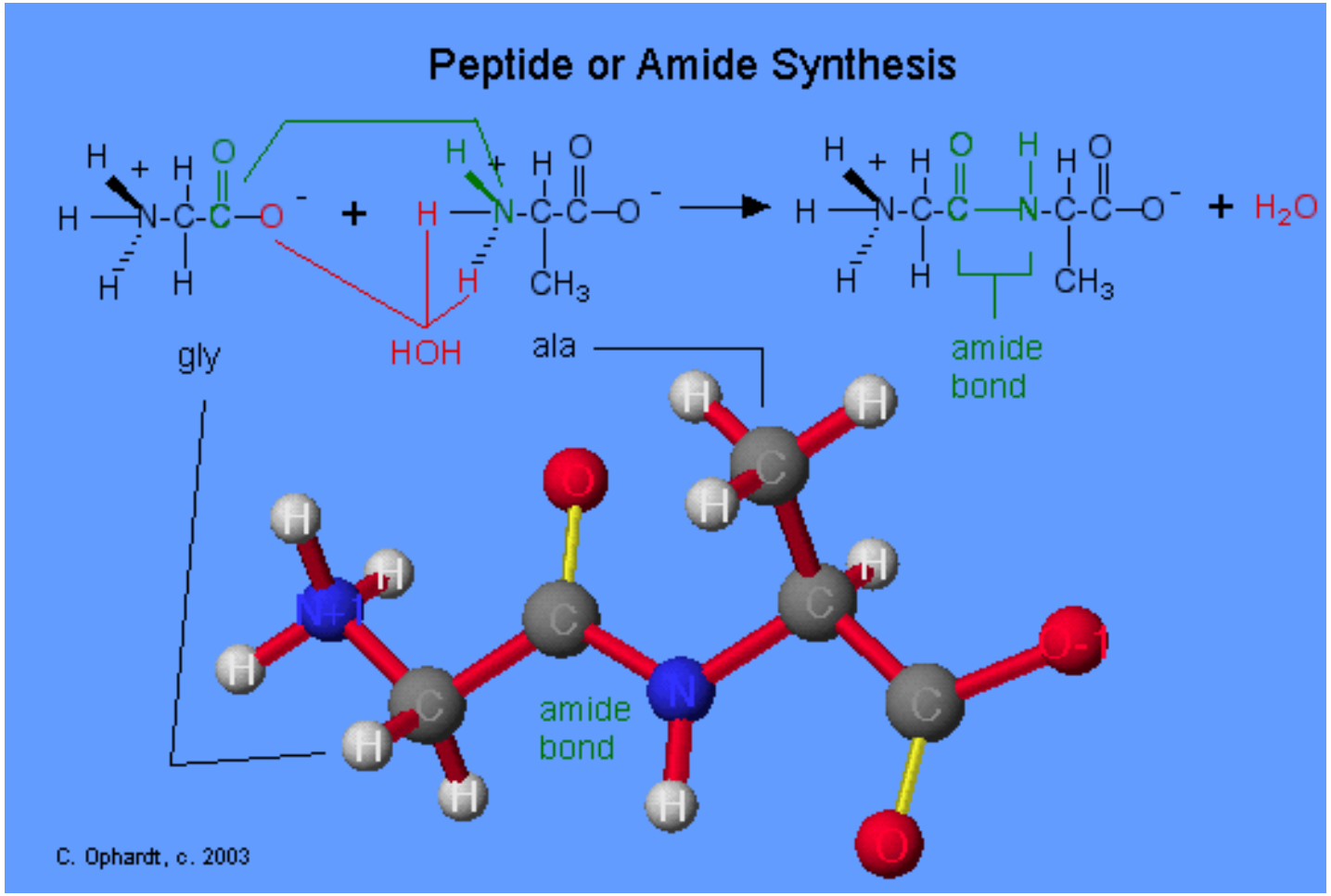




Amino acid

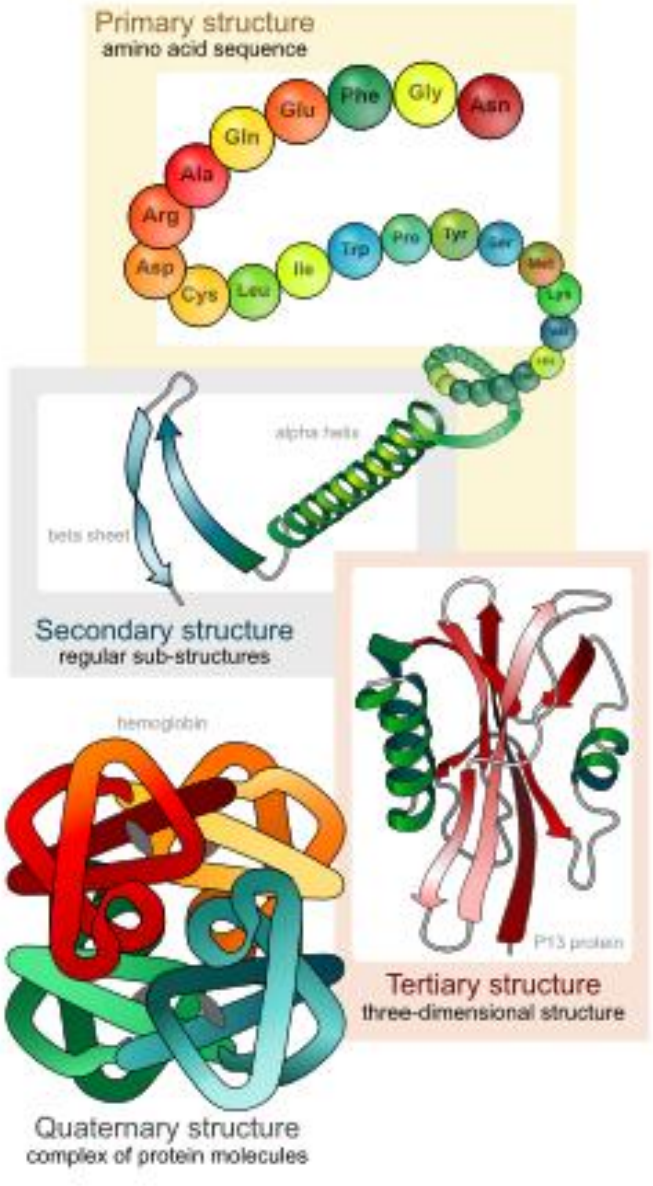


Standard α-L Amino Acid



Show Translation Movie

Primary, secondary, tertiary and quaternary structure



Protein Folding "Problem"
How do proteins acquire their unique 3D structures?

-Levinthal's paradox (3^{100} x10^{-15} s)

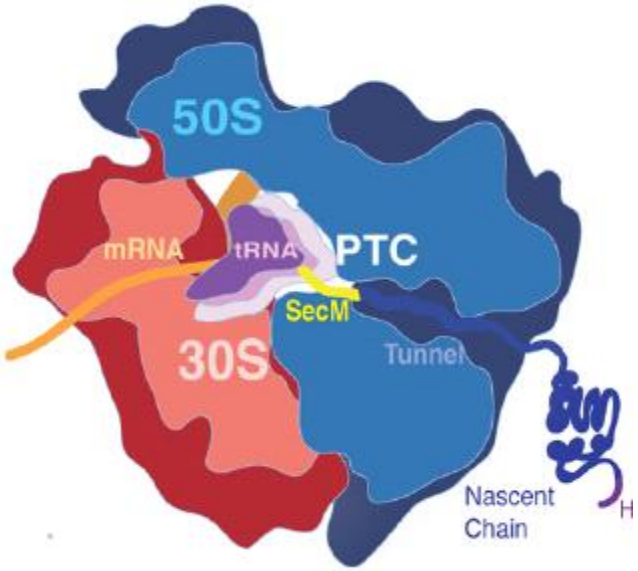
-Anfinsen's thermodynamic hypothesis

Native, unique structure corresponds to the minimum of the free energy

- uniquness
- stability
- kinetical accessibility (energy funnel)

-Folding in *vitro vs in vivo*; cotranslational folding?

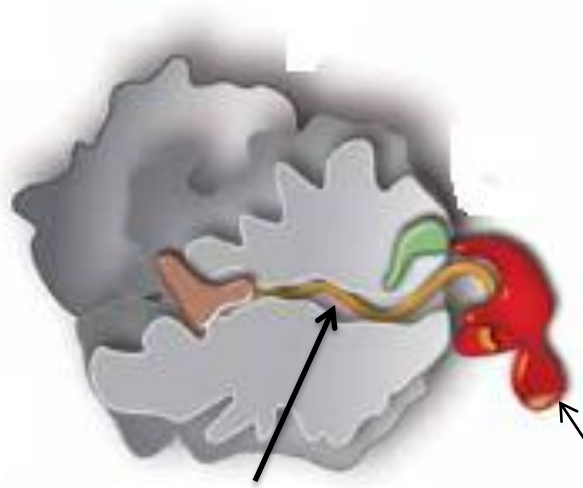
Nascent Polypeptide Chain (NPC)



Cabrita, Hsu, Launay, Dobson, Christodoulou. 2009.
PNAS **106, 22239–22244.**

Exit tunnel: 10 nm long
1-2 nm wide, accomodates
30 aa in the extended
conformation, up to 60 aa
(α –helix)

Kramer, Boehringer, Ban, Bukau. (2009).
NAT. STRUCT. & MOL. BIOL. 16 , 589.



NPC

Trigger Factor (TF)

Cabrita, Hsu, Launay, Dobson, Christodoulou. 2009. *PNAS* **106, 22239–22244.**
Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy

…….These findings are of particular interest when compared to force-induced mechanical unfolding experiments, which provide an in vitro representation of one form of vectorial folding that could be somewhat analogous to the behavior of an NC as it emerges from the ribosomal tunnel……

Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM

Matthias Rief, Mathias Gautel, Filipp Oesterhelt, Julio M. Fernandez, Hermann E. Gaub*

SCIENCE 276, 16 MAY 1997, pp. 1109-1112

Folding-unfolding transitions in single titin molecules characterized with laser tweezers

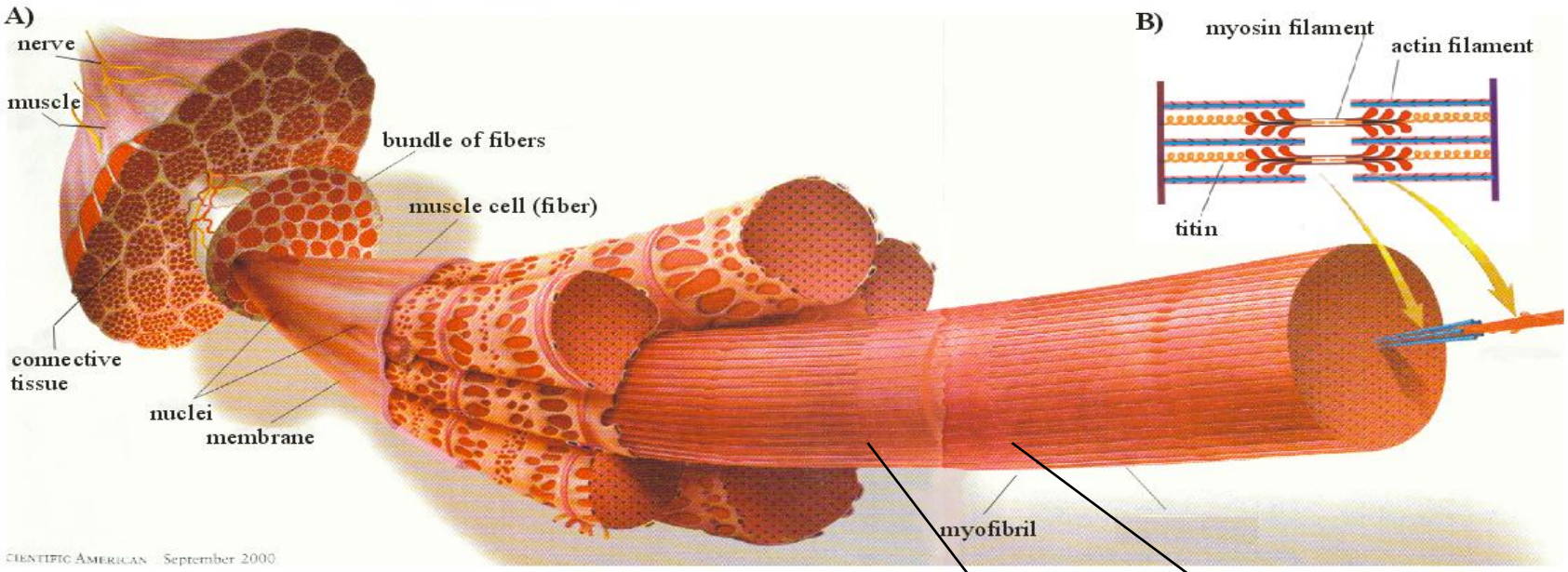
Kellermayer, M. S. Z. Smith, S. B. Granzier, H. L. Bustamante, C.

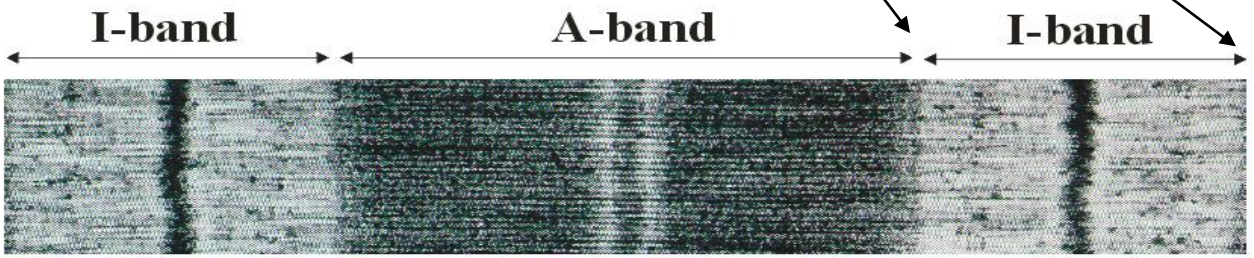
SCIENCE 276, 16 MAY 1997, pp. 1112-1116.

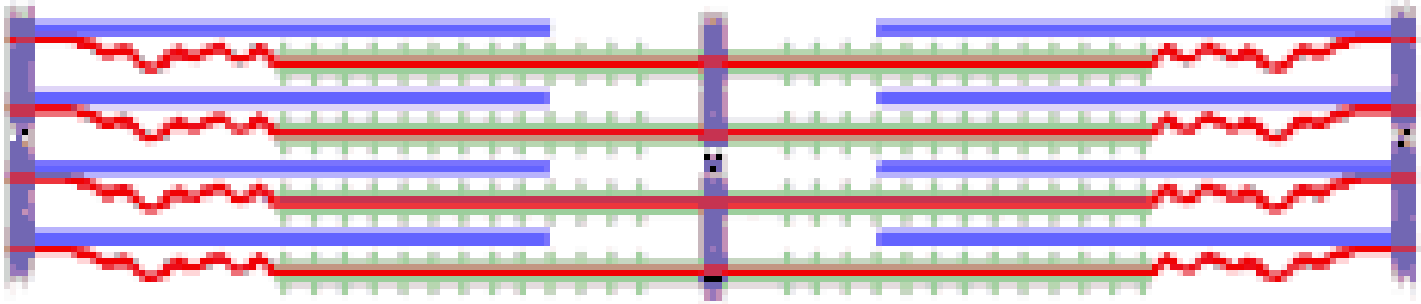
Elasticity and unfolding of single molecules of the giant muscle protein titin

L. Tskhovrebova, J. Trinick, J. A. Sleep, R. M. Simmons

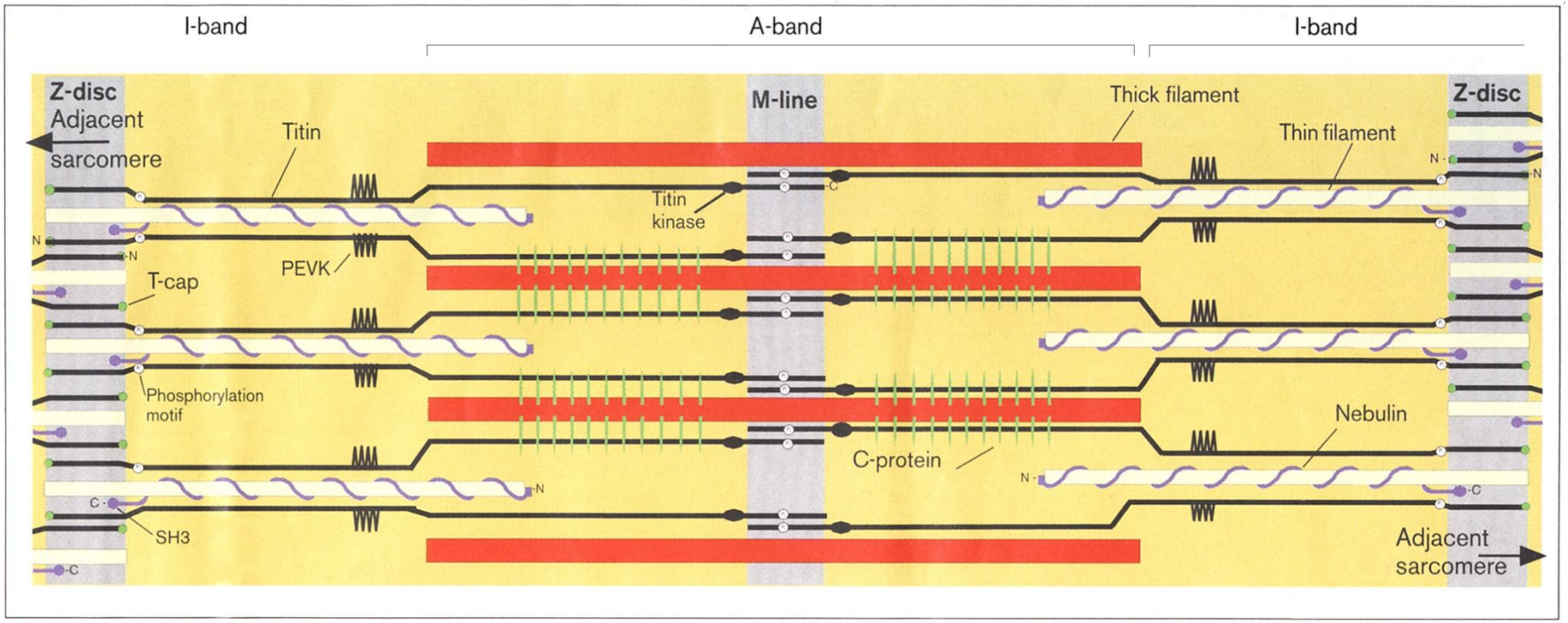
Nature **387**, 308-312 (15 May 1997)


A)
nerve
muscle
bundle of fibers
muscle cell (fiber)
connective tissue
nuclei
membrane
myofibril
B)
myosin filament
actin filament
titin
SCIENTIFIC AMERICAN September 2000


I-band
A-band
I-band


Z-line
M-line

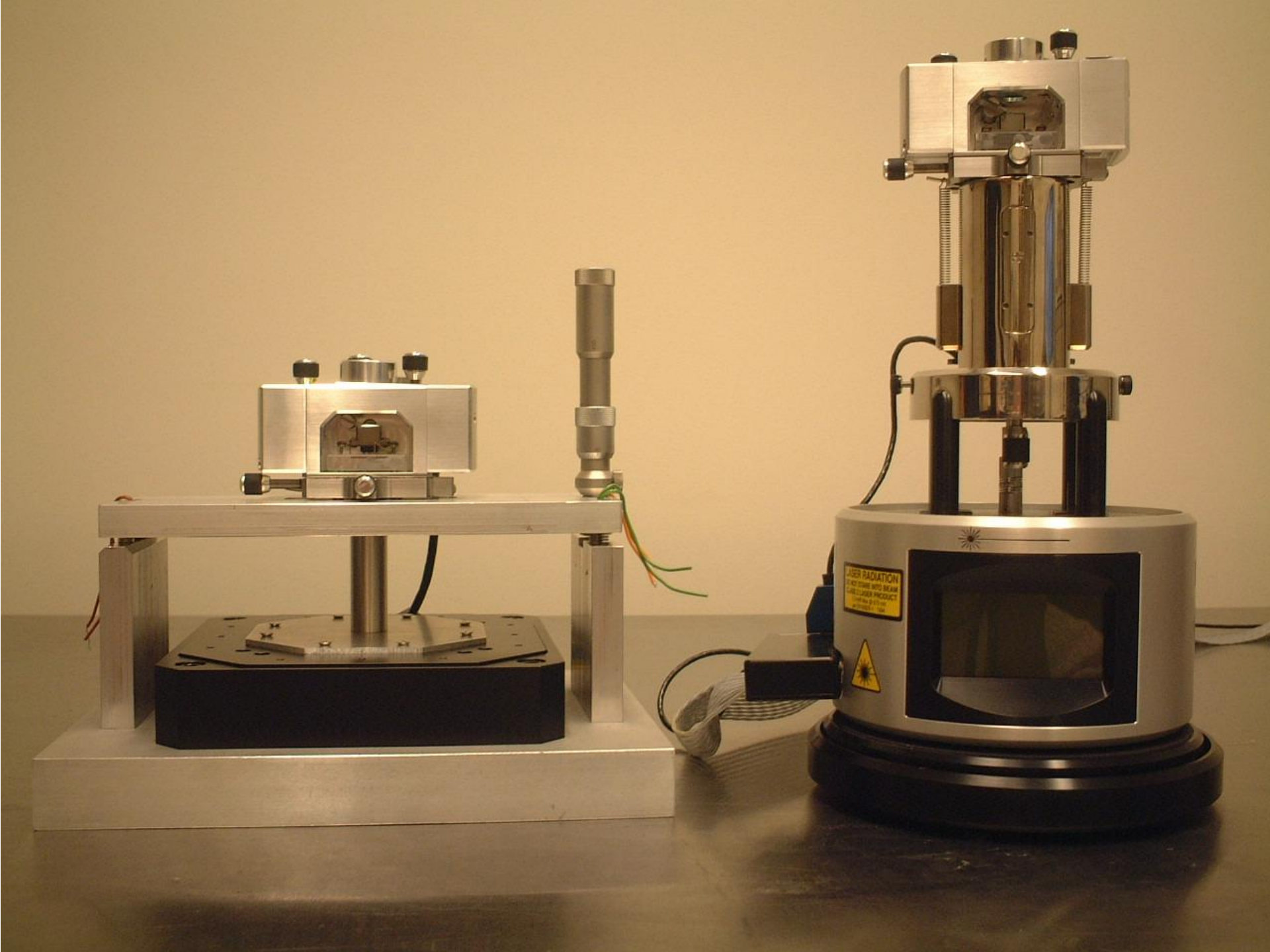
Sarcomere



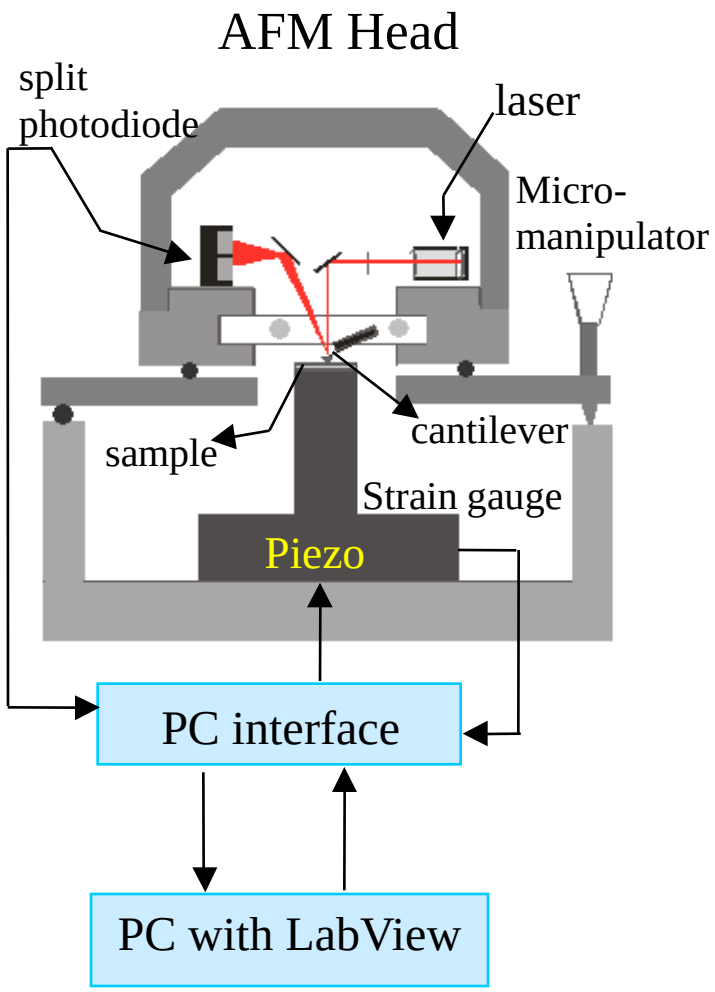
PEVK-unstructured
Entropic springs

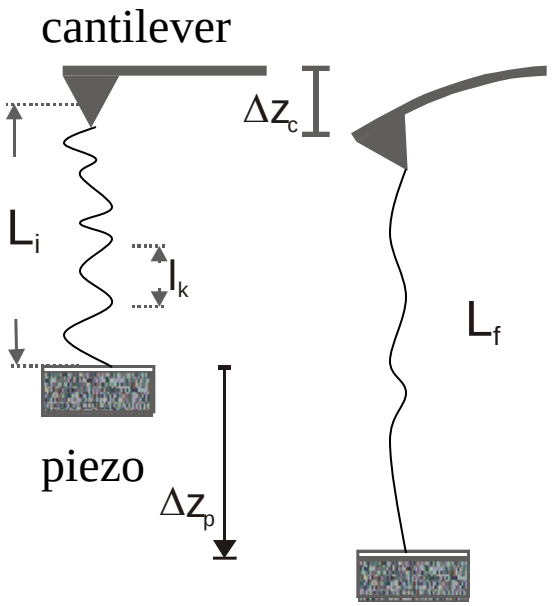
Ig domains

N2B-unstructured
Entropic springs


LASER RADIATION

Atomic Force Microscope





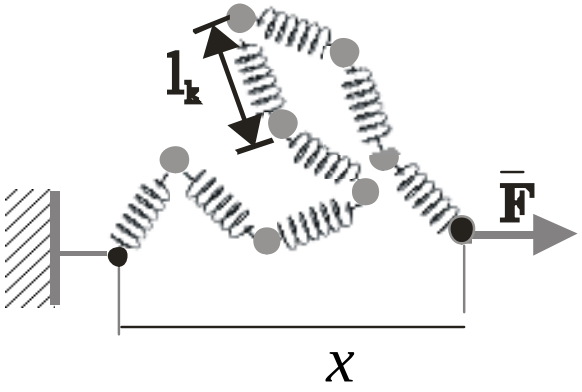
$$X = L_f - L_i$$

$$F = K_c \Delta Z_c$$

(K_c, spring constant)

Freely jointed chain with segment elasticity

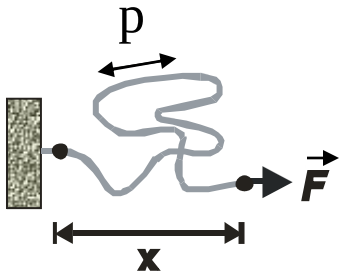
ssDNA polysaccharides



$$x(F) = \left(\coth\left(\frac{F l_K}{k_B T} \right) - \frac{k_B T}{F l_K} \right) \left(L_{con} + \frac{nF}{k_{segment}} \right)$$

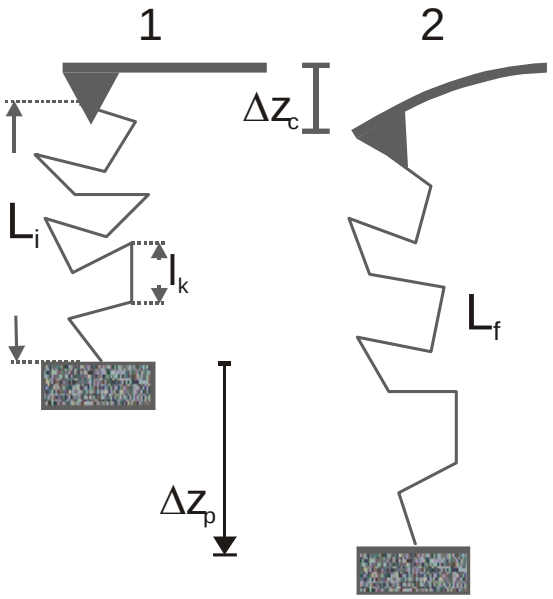
Worm-like chain

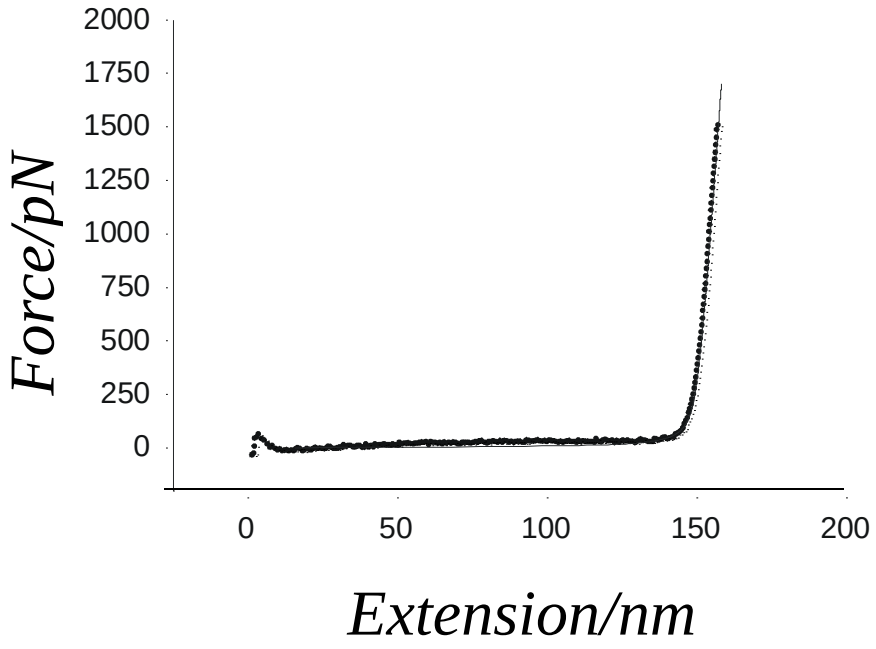
ds DNA modular proteins



$$F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{L_{con}} \right)^{-2} - \frac{1}{4} + \frac{x}{L_{con}} \right]$$

Entropic elasticity

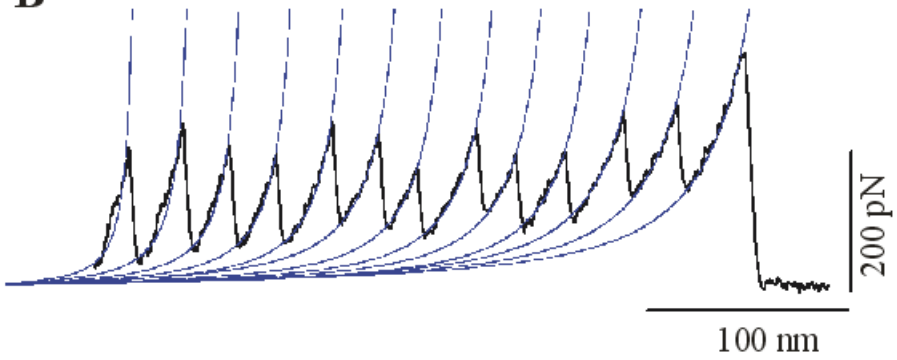




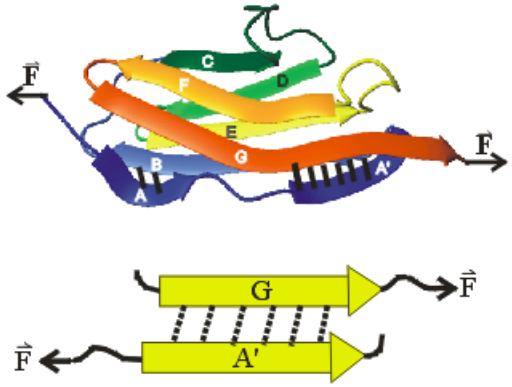
Advantages of polyproteins for SMFS

A

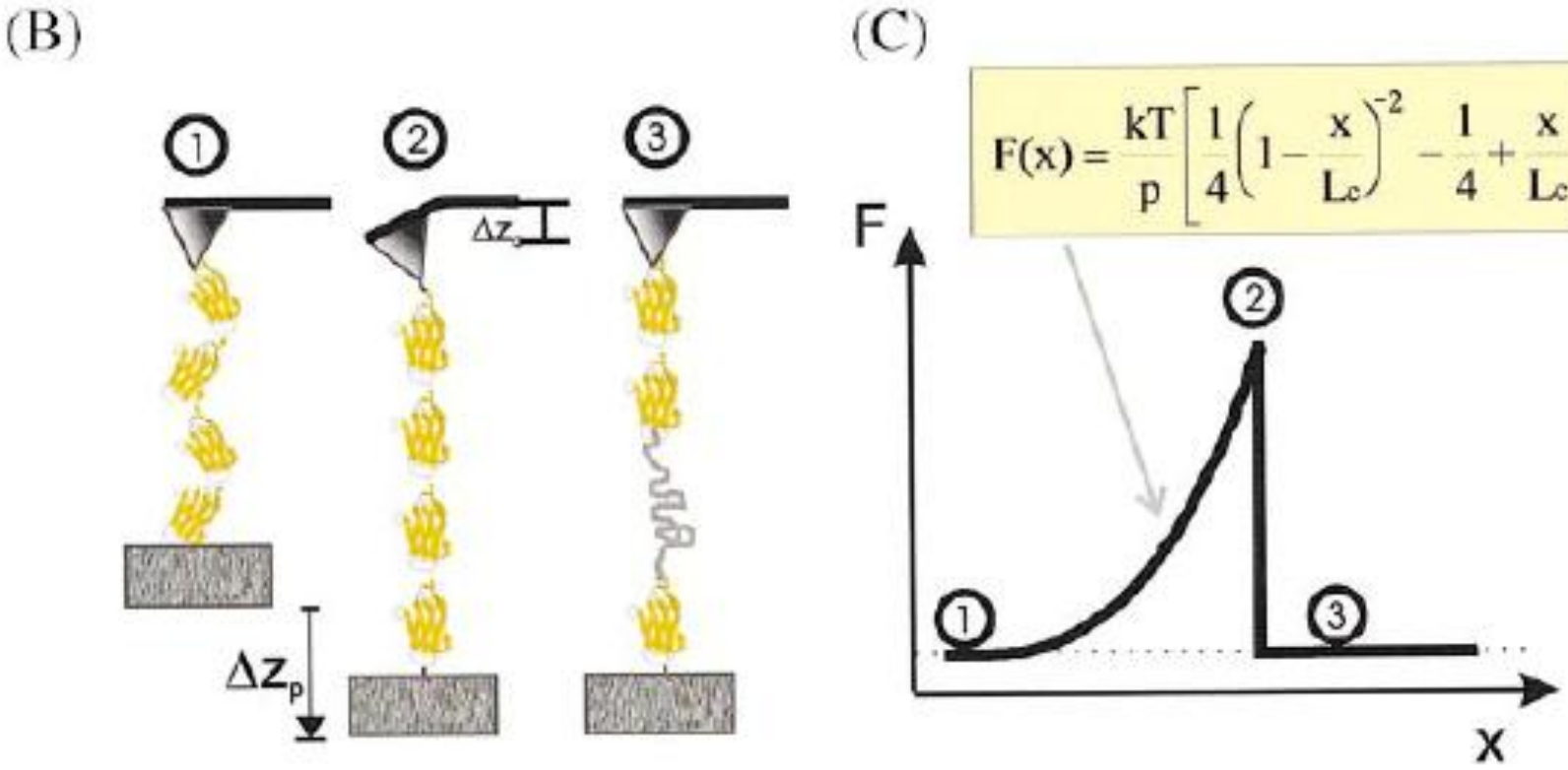
B



C

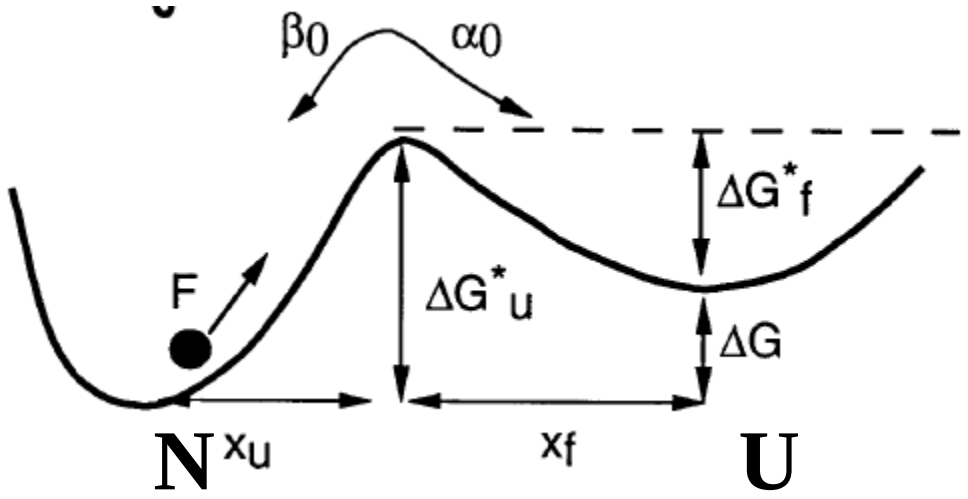


Carrion-Vazquez, Oberhauser, Fisher, Marszalek, Li & Fernandez. (2000). *Prog. Biophys. Mol. Biol.* 74, 63-91 (and references cited therein)



Show I27 unfolding animation

Two-state model of protein unfolding/folding



$N \leftrightarrow U$

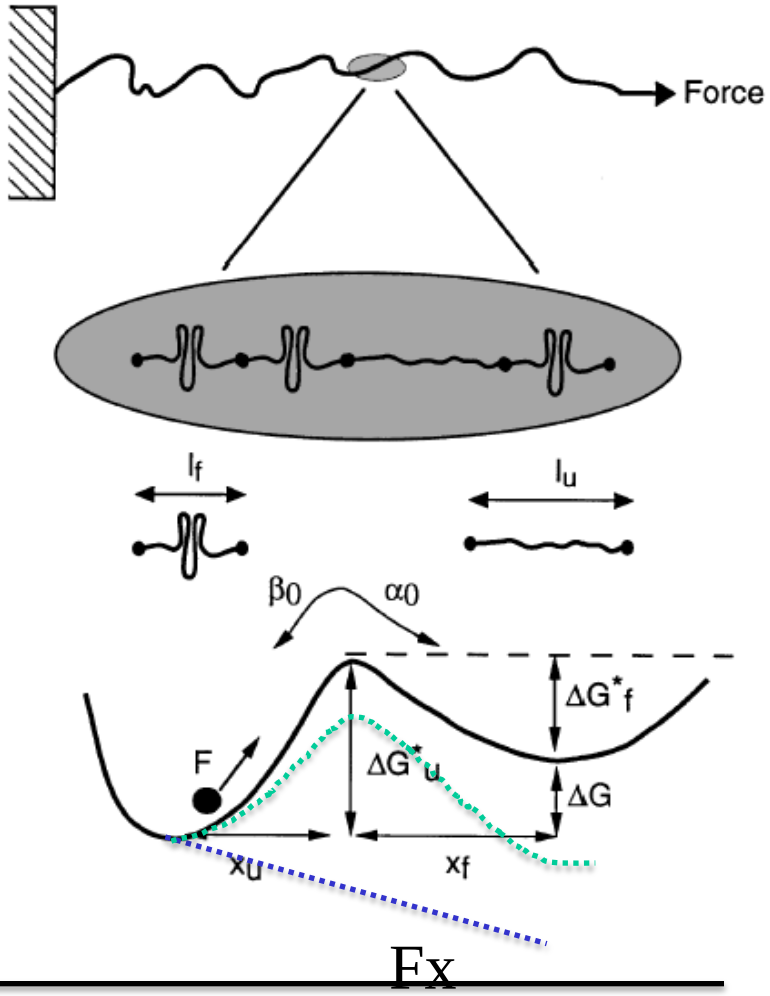
$$\alpha_0 = \omega e^{-\Delta G_u^*/k_B T}$$

$$\beta_0 = \omega e^{-\Delta G_f^*/k_B T}$$

$dP_N/dt = \beta_0 P_U - \alpha_0 P_N \quad P_U + P_N = 1$

Elastically Coupled Two-Level Systems as a Model for Biopolymer Extensibility

Matthias Rief,[1] Julio M. Fernandez,[2] and Hermann E. Gaub[1]



$$F(x) = \frac{k_B T}{p}\left(\frac{1}{4(1 - x/L)^2} - \frac{1}{4} + \frac{x}{L}\right). \quad (1)$$

$$\alpha_0 = \omega e^{-\Delta G^*_u / k_B T}$$

$$\beta_0 = \omega e^{-\Delta G^*_f / k_B T}$$

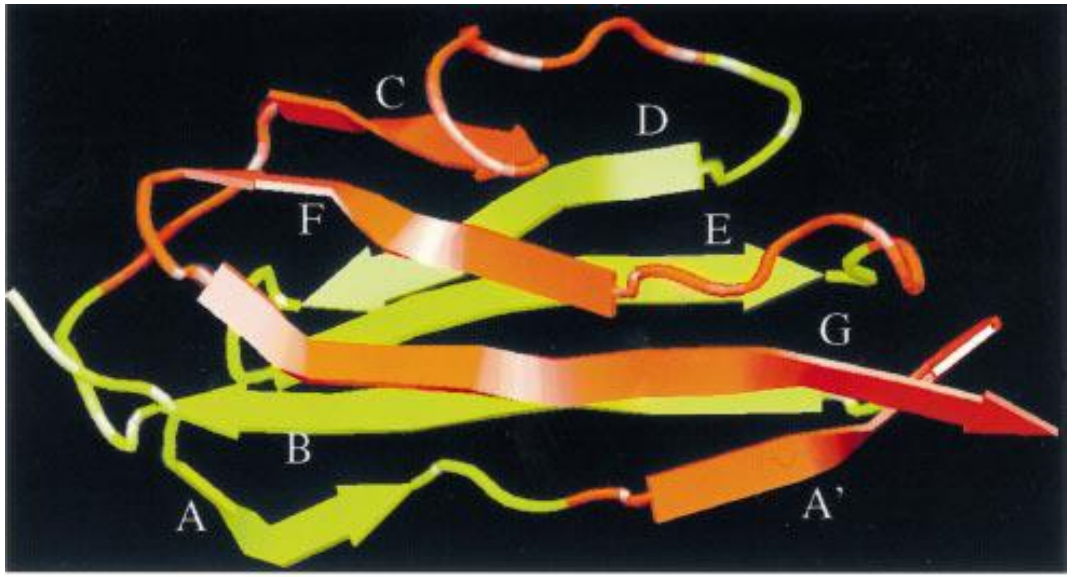
$$\alpha(F) = \omega e^{-(\Delta G^*_u - F x_u)/k_B T} = \alpha_0 e^{F x_u / k_B T} \quad (4)$$

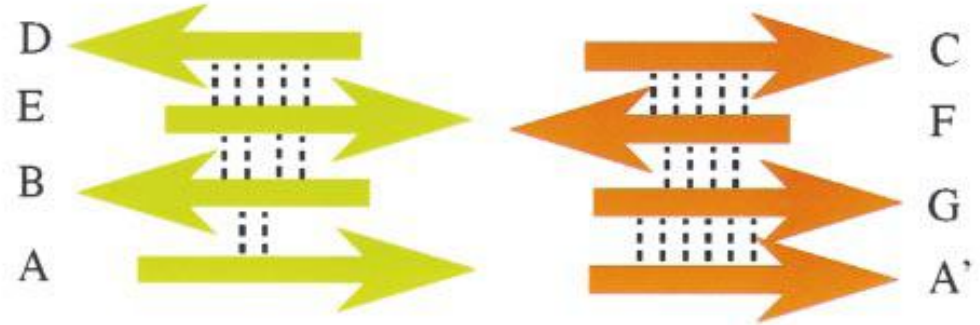
$$\beta(F) = \omega e^{-(\Delta G^*_f + F x_f)/k_B T} = \beta_0 e^{-F x_f / k_B T} \quad (5)$$

Reaction Coordinate (extension)

G. I. Bell, Science **200, 618 (1978)**

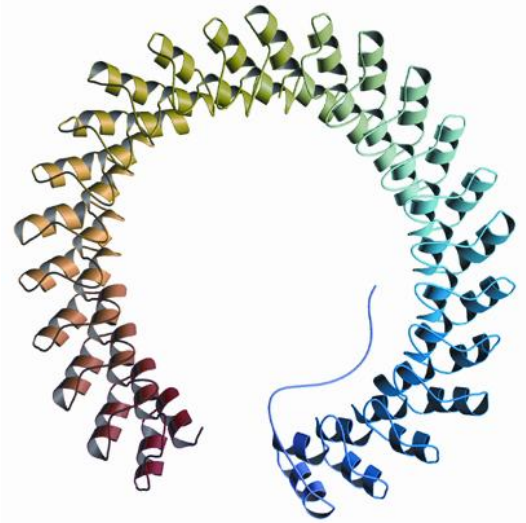
The I27 domain of titin and its network of backbone hydrogen bonds



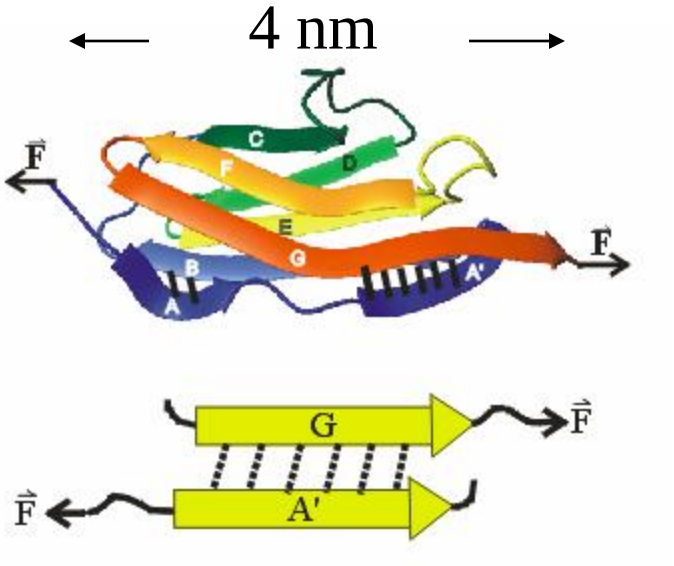


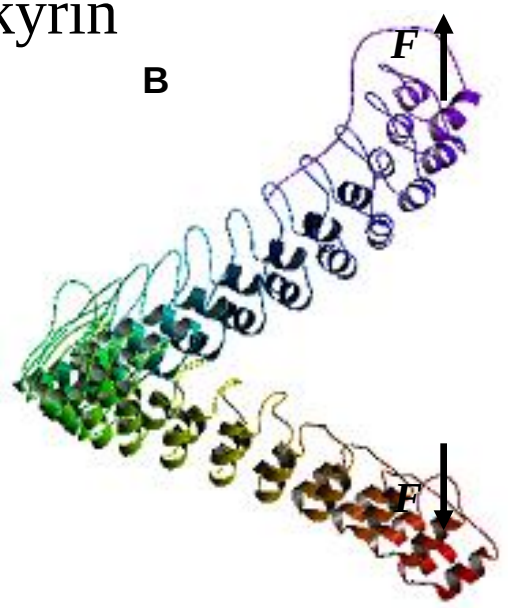
Nanomechanics of modular vs repeat (spiral) proteins

titin

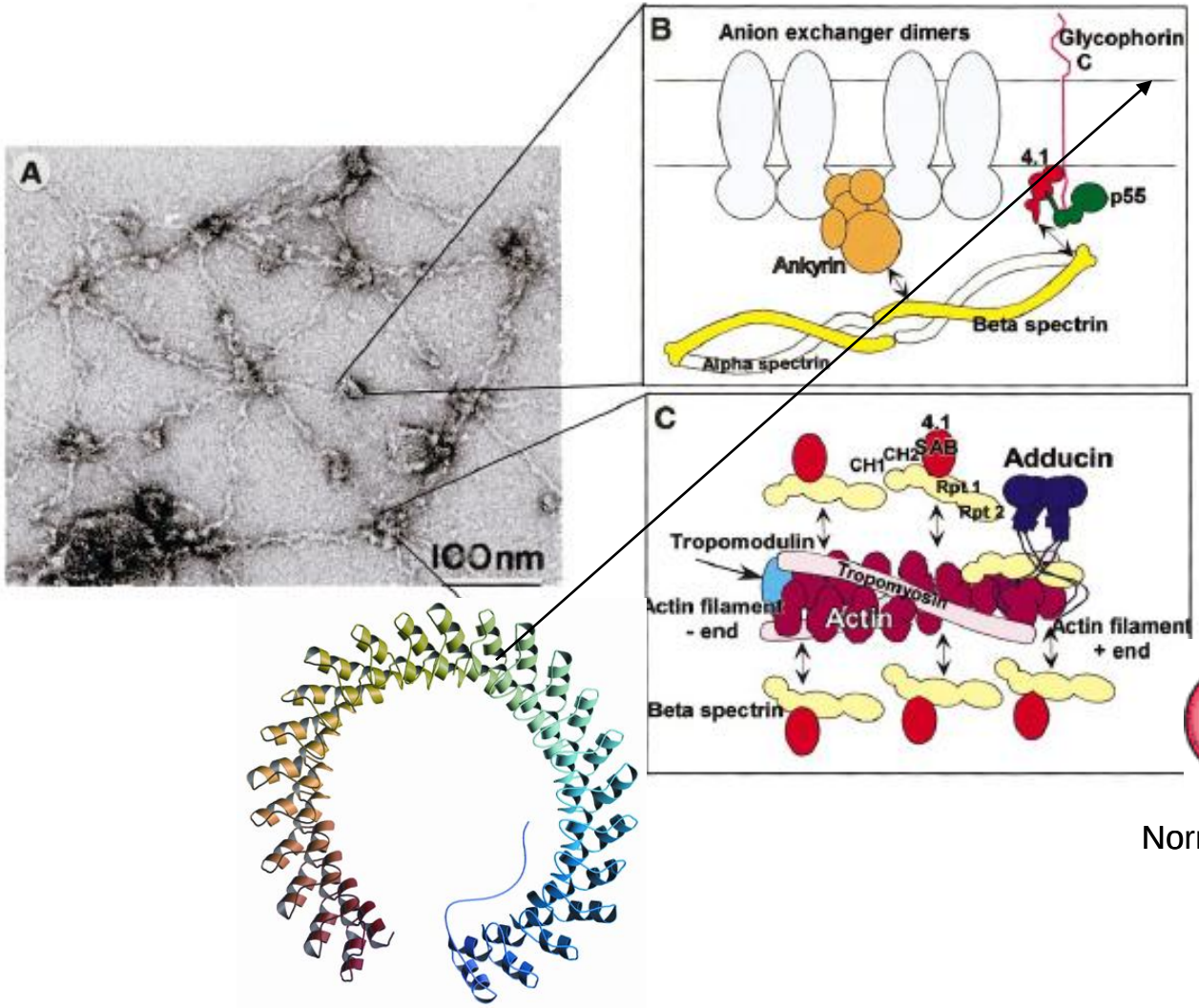


ankyrin

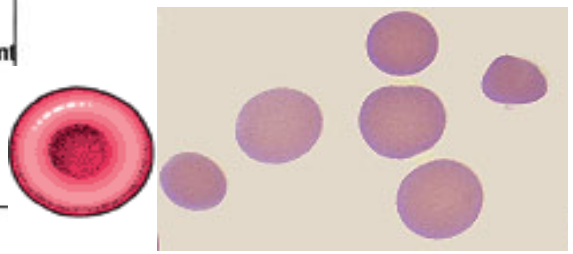




Organization of the RBC membrane skeleton



Bennett, V., andA.J. Baines. 2001. *Physiol. Rev.* 81(3):1353-1392.

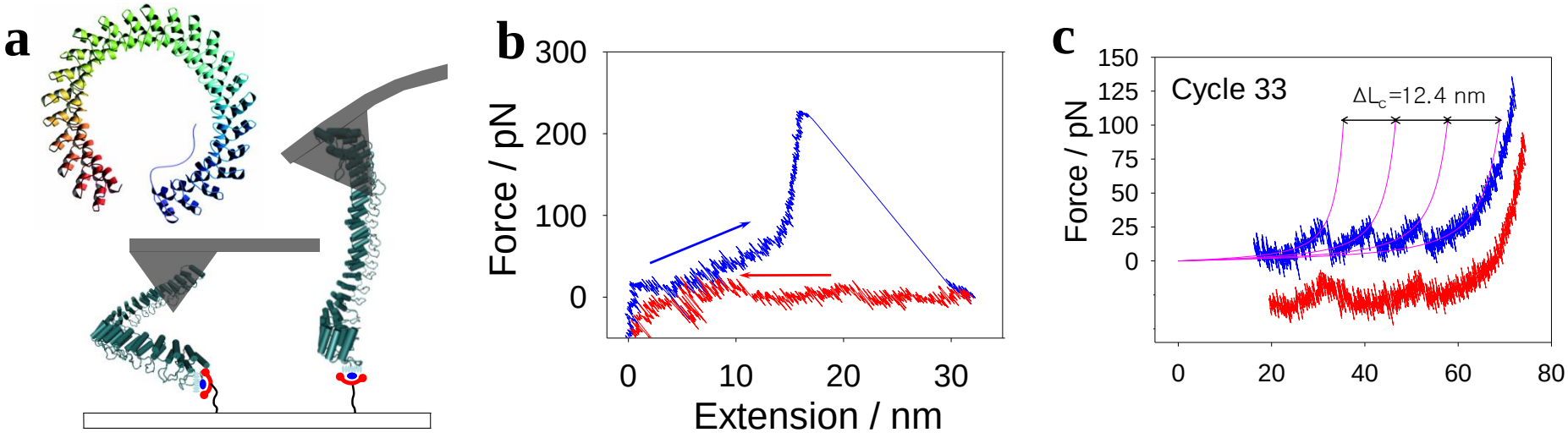


Normal RBC Ankyrin defects: Spherocytes

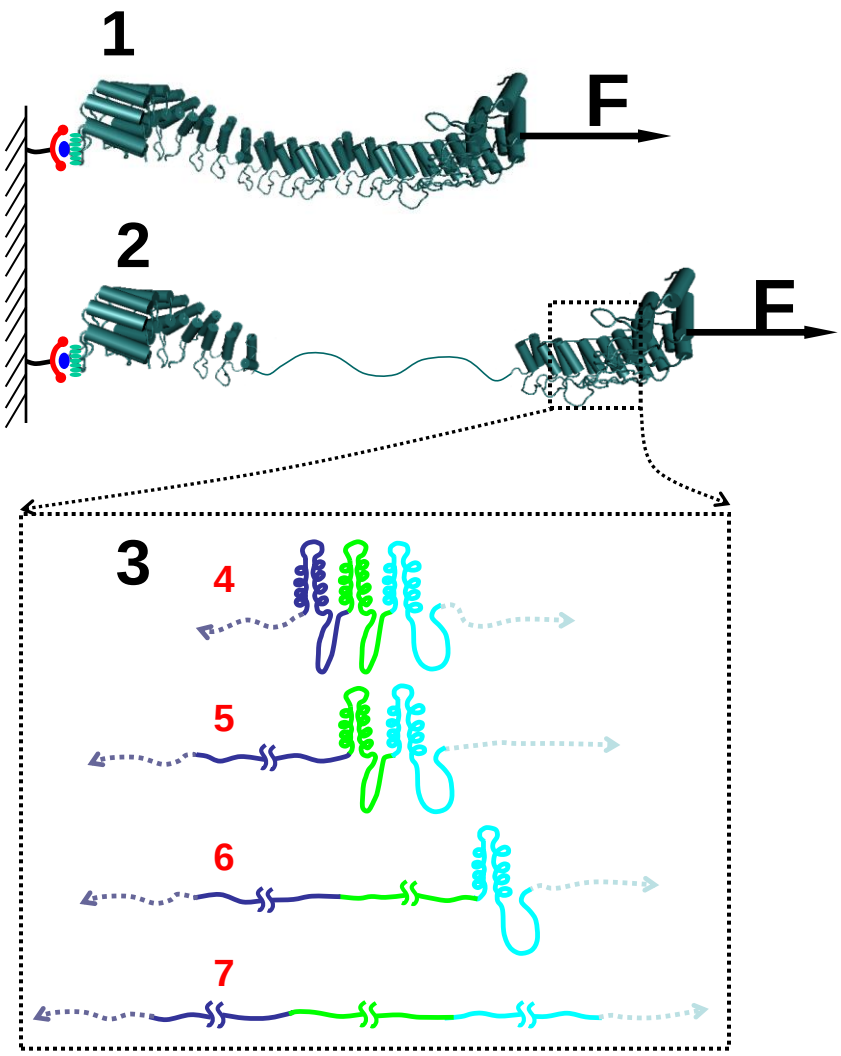
H277R

Ankyrin membrane-binding domain: 24 ANK repeats

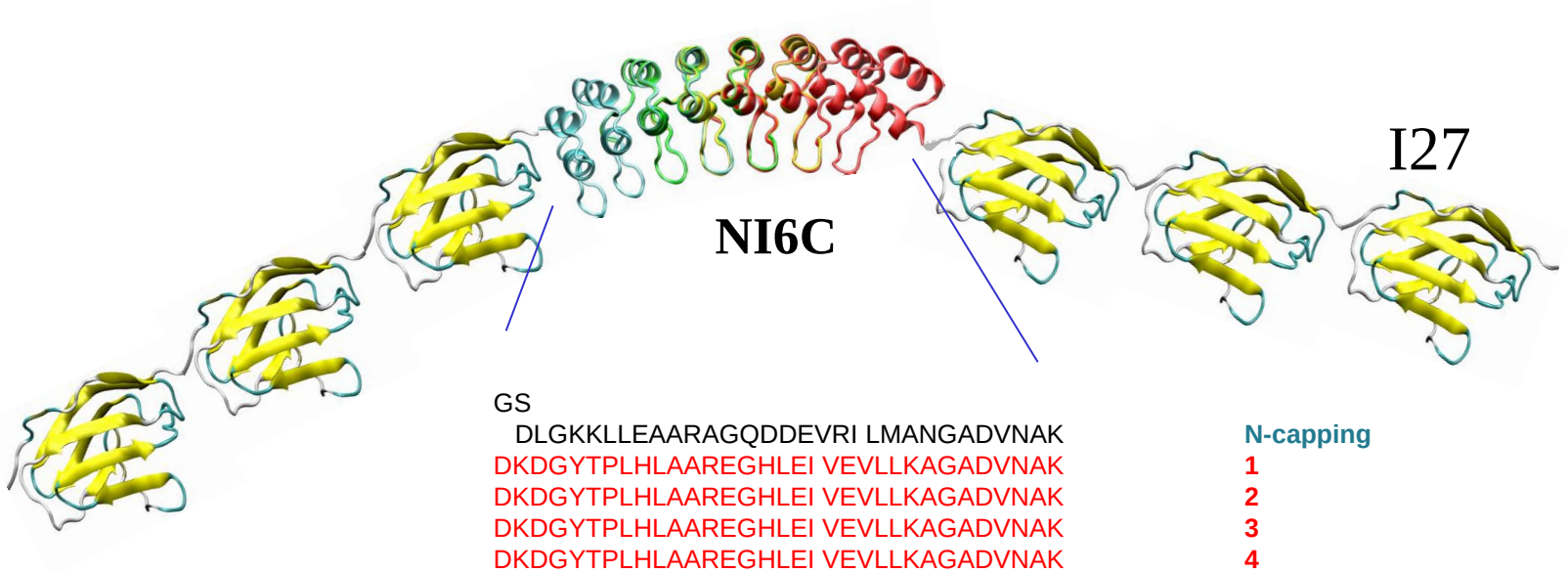
Nanospring Behavior of Ankyrin Repeats



G. Lee, K. Abdi, Y. Jiang, P. Michaely, V. Bennett & P.E. Marszalek. (2006). Nature 440, 246-249.


1
F
2
F
3
4
5
6
7

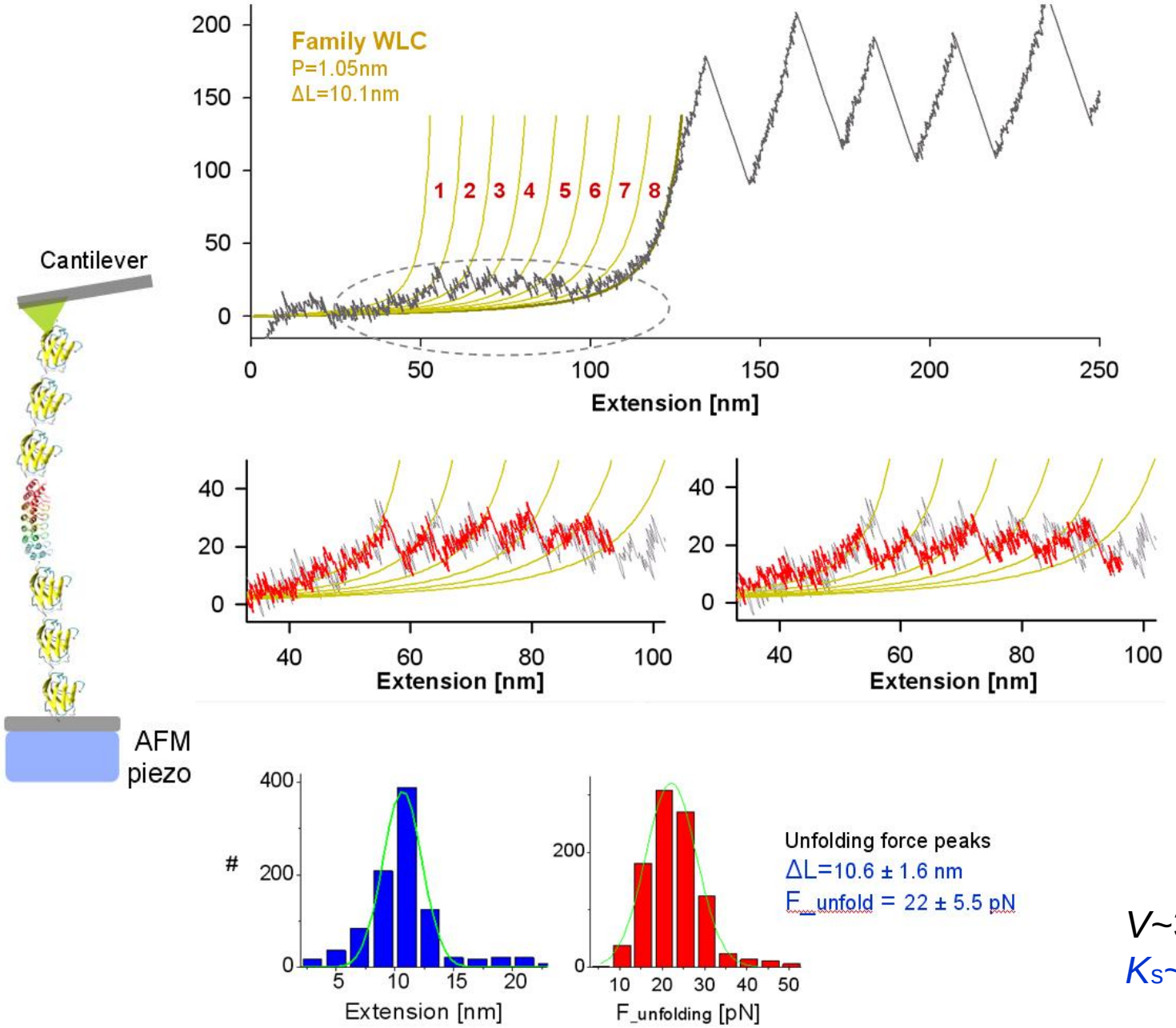
Consensus Ankyrin (NI6C) flanked by six I27 modules



GS
DLGKKLLEAARAGQDDEVRI LMANGADVNAK **N-capping**
DKDGYTPLHLAAREGHLEI VEVLLKAGADVNAK **1**
DKDGYTPLHLAAREGHLEI VEVLLKAGADVNAK **2**
DKDGYTPLHLAAREGHLEI VEVLLKAGADVNAK **3**
DKDGYTPLHLAAREGHLEI VEVLLKAGADVNAK **4**
DKDGYTPLHLAAREGHLEI VEVLLKAGADVNAK **5**
DKDGYTPLHLAAREGHLEI VEVLLKAGADVNAQ **6**
DKFGKTAFDISIDNGNEDLAEILQ **C-capping**

8 Ankyrin repeats
253 aa Stretched
length ~ 92nm

Svava K. Wetzel, Giovanni Settanni, Manca Kenig, H. Kaspar Binz and Andreas Plückthun. *J. Mol. Biol.* (2008) 376, 241–257.

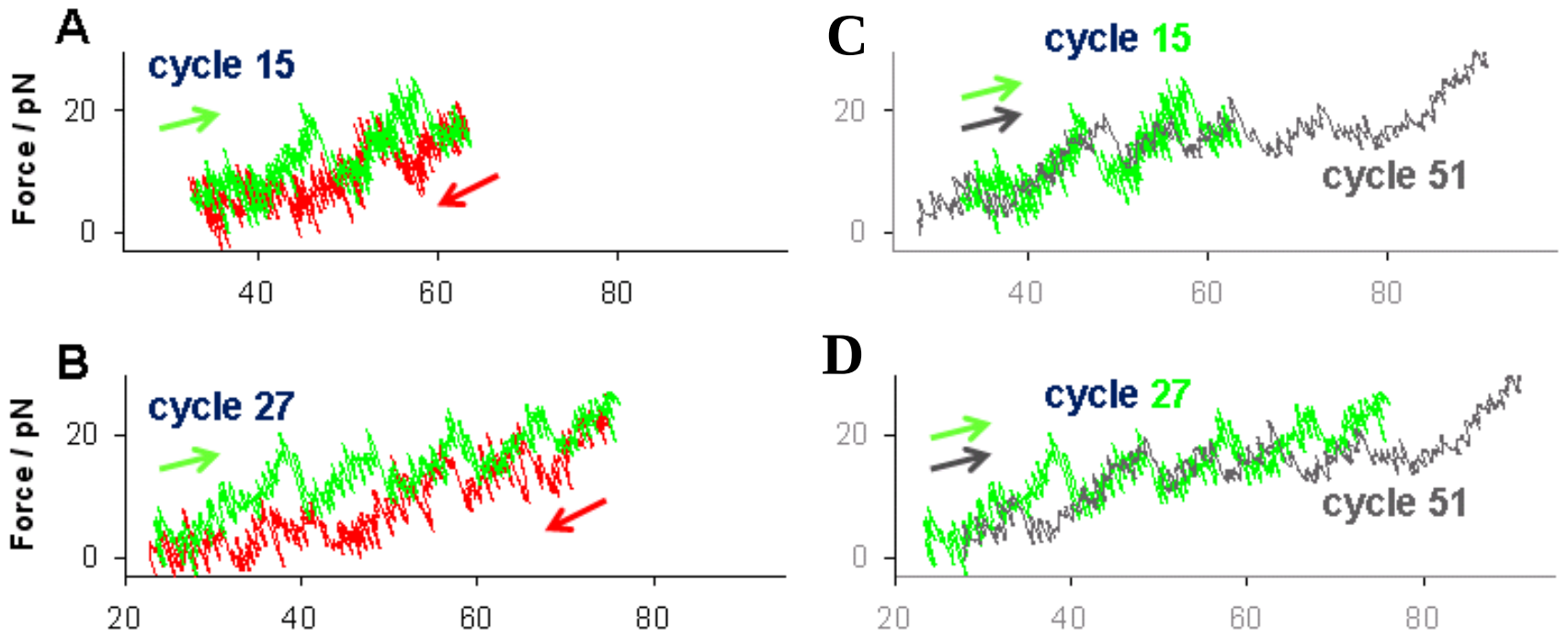


V~30nm/ms

K_S~6pN/nm

Lee , Zeng, Zhou, Bennett, Yang, Marszalek. (2010). **J Biol Chem 285, 38167-38172.**

Cyclic stretch-relax measurements



Protein mechanics und unfolding can be studied by computer simulations (molecular dynamics)

An Introduction to Molecular Dynamics Simulations

Macroscopic properties are often determined by molecule-level behavior.

Quantitative and/or qualitative information about macroscopic behavior of macromolecules can be obtained from simulation of a system at atomistic level.

Molecular dynamics simulations calculate the motion of the atoms in a molecular assembly using Newtonian dynamics to determine the net force and acceleration experienced by each atom. Each atom i at position r_i, is treated as a point with a mass m_i and a fixed charge q_i.

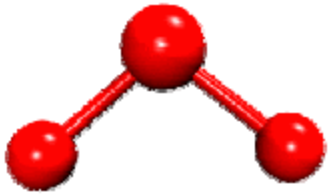
www.ks.uiuc.edu

Professor Klaus Schulten, Univ. of Illinois, Urbana-Champaign

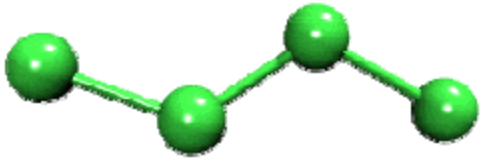
Energy Terms Described in the CHARMm Force Field

Bond

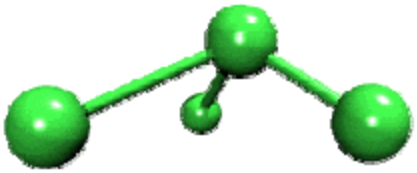
Angle



Dihedral



Improper



Energy Functions

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

U_{bond} = oscillations about the equilibrium bond length
U_{angle} = oscillations of 3 atoms about an equilibrium angle
$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

MD: Verlet Method

Energy function: $U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R})$

used to determine the force on each atom:

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R})$$

Newton's equation represents a set of N second order differential equations which are solved numerically at discrete time steps to determine the trajectory of each atom.

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \vec{F}_i(t)$$

Steered Molecular Dynamics

$$U_{tot} = U_{internal} + U_{harmonic\ (AFM)}$$

$$U_{harmonic} = \frac{1}{2} k \left[vt - \left(\vec{r} - \vec{r_o} \right) \cdot \vec{n} \right]^2$$

Go-Like Model : Structure-based Model

Weitao Yang's group

- Onuchic's Go Model
  - Clementi C, Nymeyer H & Onuchic JN (2000) "Topological and energetic factors: What determines the structural details of the transition state ensemble and En-route intermediates for protein folding? An Investigation for small globular proteins." *J. Mol. Biol.* **298**, 937-953.
  - http://sbm.ucsd.edu/cgi-bin/GenTopGro.pl

$$E(\Gamma,\ \Gamma_0) = \sum_{\text{bonds}} K_r(r - r_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2$$

$$+ \sum_{\text{dihedral}} K_\phi^{(n)}[1 + \cos(n \times (\phi - \phi_0))]$$

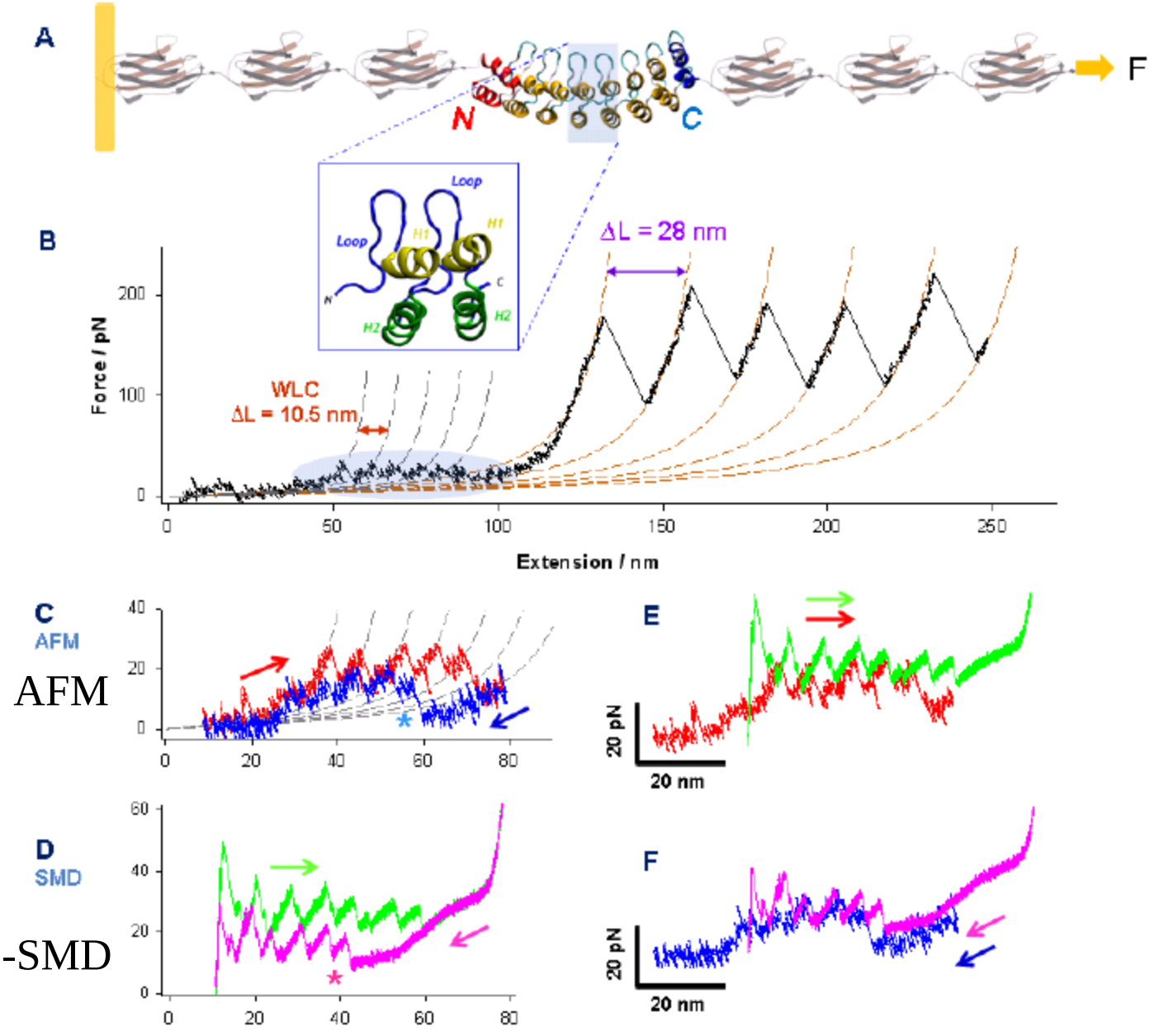
$$+ \sum_{i<j-3} \left\{ \varepsilon(i,j)\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] \right.$$

Native cont.

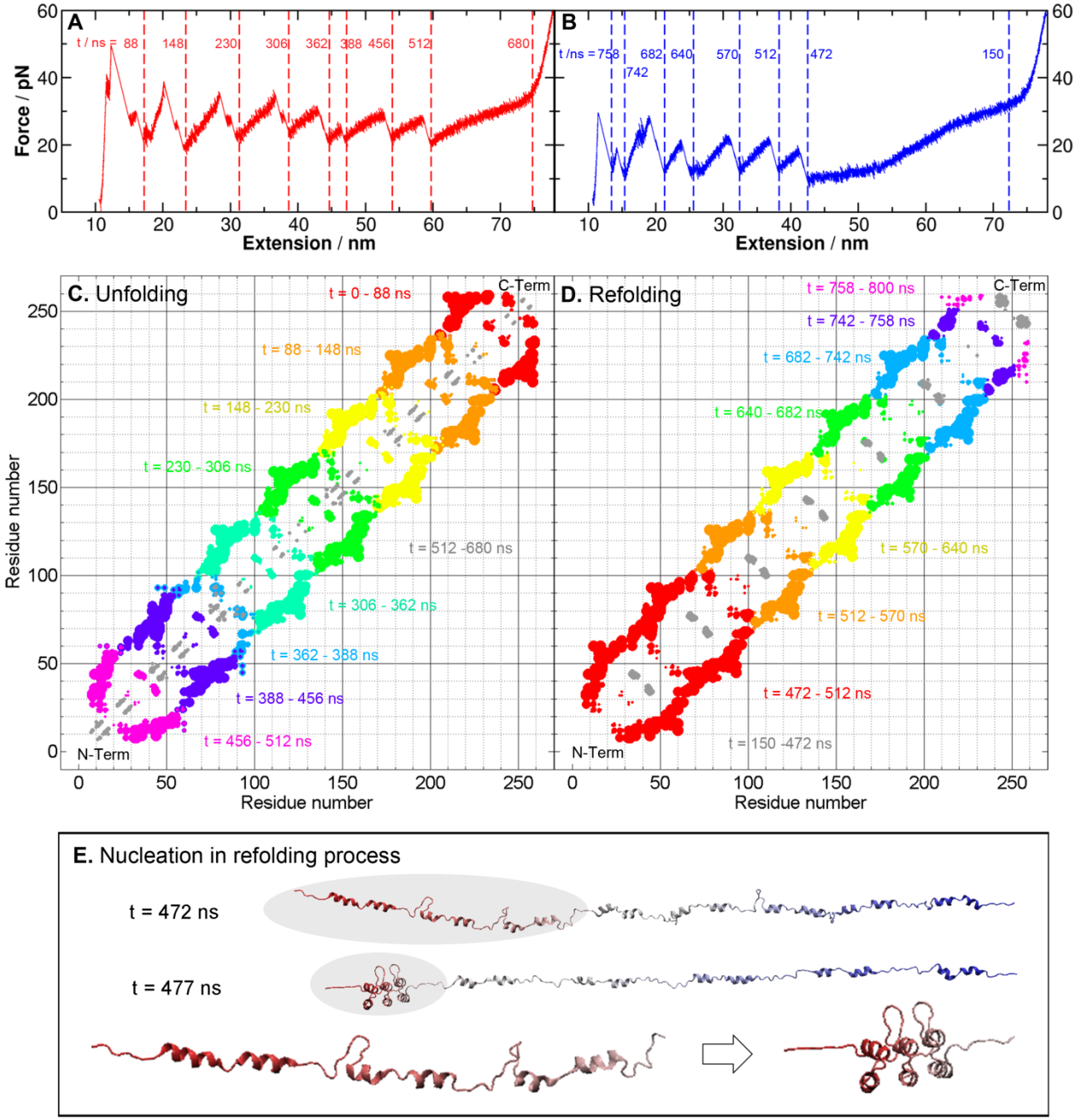
$$\left. + \varepsilon_2(i,j)\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} \right\}$$

Non-native cont.

Lee , Zeng, Zhou, Bennett, Yang, Marszalek. (2010).
J Biol Chem 285, 38167-38172.



NI6C contact map during unfolding and refolding

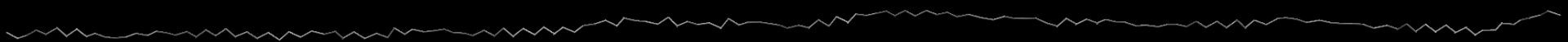


Count of native contacts in α-helical domains (H1 and H2) in the N-terminus, C- terminus, and internal repeats of NI6C

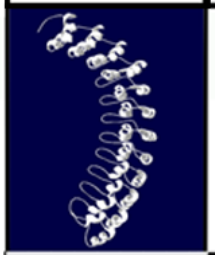
Domain	Sequence*	# of residues	# Of contacts†
N-terminal	Asp7 - Gln30	24	85
C-terminal	Thr241 - Gln259	19	70
Internal	Thr109 - Ala129	21	129

Vectorial unfolding of Consensus Ankyrin Repeats

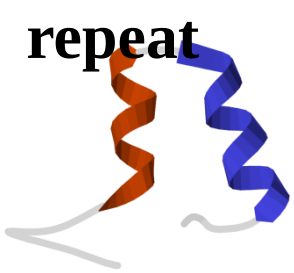
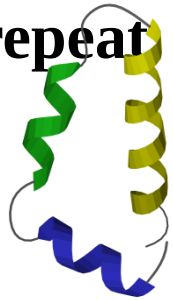
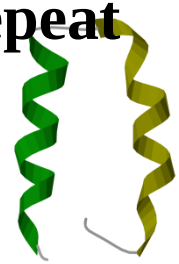
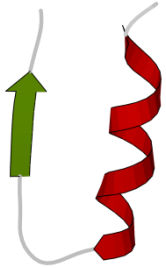
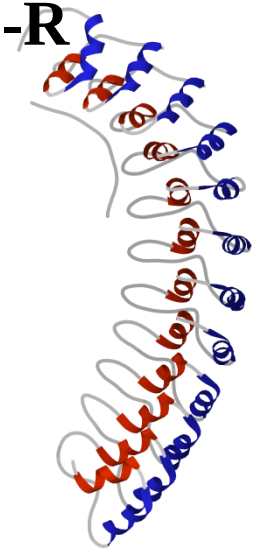
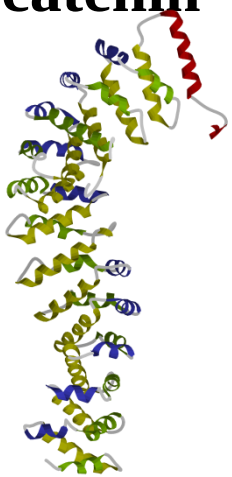
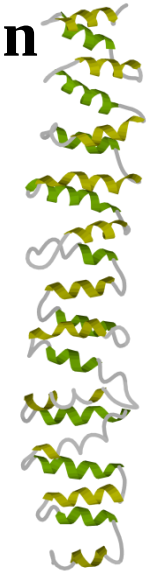
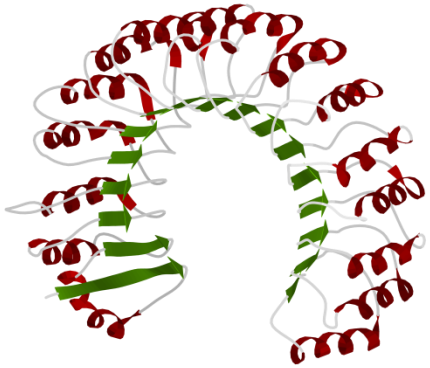
Vectorial Refolding of Consensus Ankyrin Repeats



Structures and functions of repeat proteins

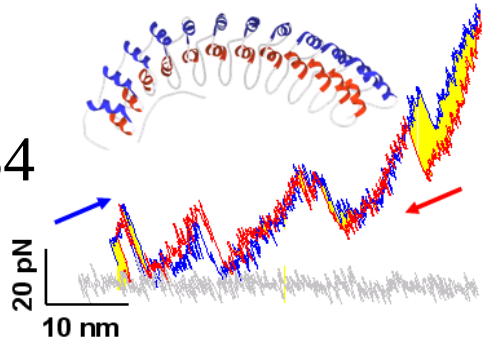
Crystal	Protein	Biological function	Type of Repeat	Occurence in H. sapeins	Additional examples bearing repeats
	ANKYRIN-R D13-24	Membrane adaptor for transmembrane localization	ANK REPEAT	3338	Ankyrin-1,2,3 BCL3 p19ink4d Notch IkappaBalpha
	CLATHRIN HEAVY CHAIN	Formation of small vesicles for intracellular transport	HEAT REPEAT	267	PP2A, subunit A Importin beta-2 Integrator complex subunit 4 Ran binding protein 5 Tbp-associated factor 172
	BETA-CATENIN	Plasma membrane adaptor for E-cadherin and transcriptional cofactor during development	ARM REPEAT	357	Importin subunit alpha-1-7 alpha & gamma catenin Plakoglobin APC tumor suppressor protein Importin subunit beta-1
	RIBONUCLEASE INHIBITOR	Inhibition of RNAse molecules	LEUCINE-RICH REPEAT	1745	Leucine rich repeat shoc-2 Toll like receptor 1,5 Slit homolog protein 2 Insulin like growth factor 1 G-protein coupled receptor 67

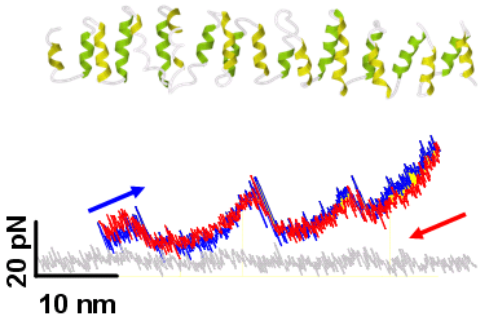
Repeat Proteins

Type of Repeat	**ANK repeat**	**ARM repeat**	**HEAT repeat**	**LR repeat**
				
Crystal	**Ankyrin -R**	**ß-catenin**	**Clathrin**	**RNAse Inhibitor**
				

Mechanical Properties of Repeat Proteins

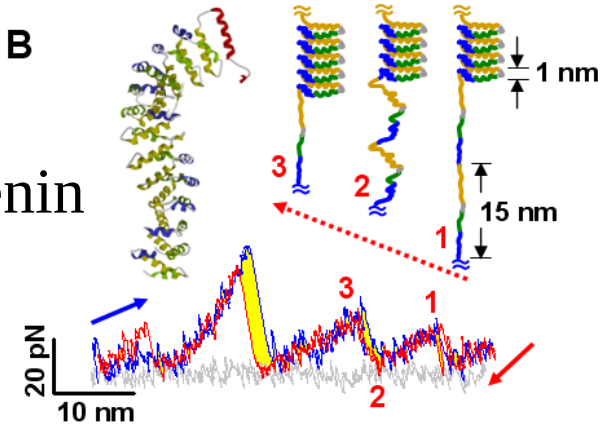
ANK-D34

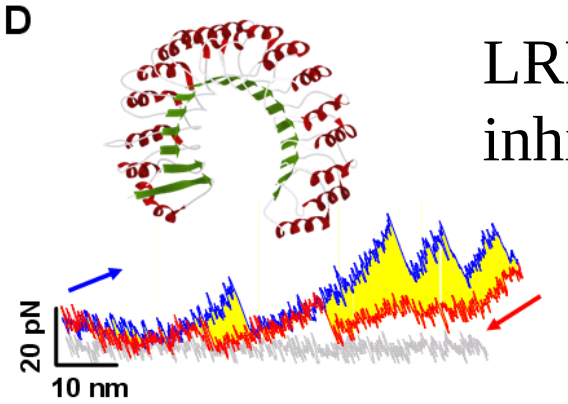




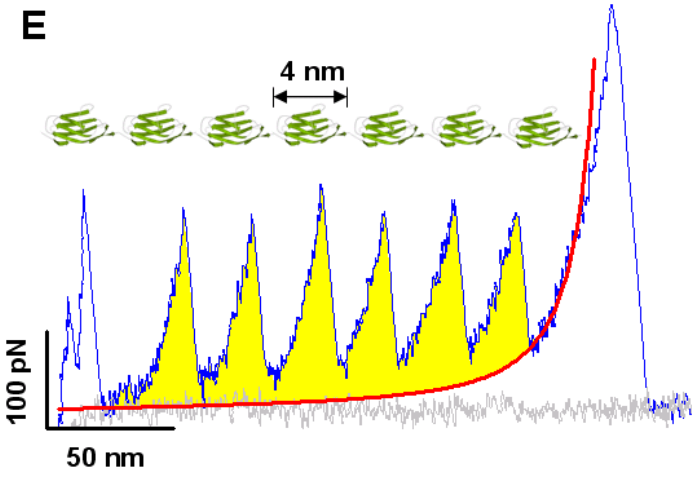
HEAT-Clathrin

ARM
β-catenin





LRR- Ribonuclease inhibitor



Kim, Abdi, Rabbi, Lee, Yang, Schofield, Bennett, Marszalek (2010). Biophys. J. **98** 3086.

Part I CONCLUSIONS

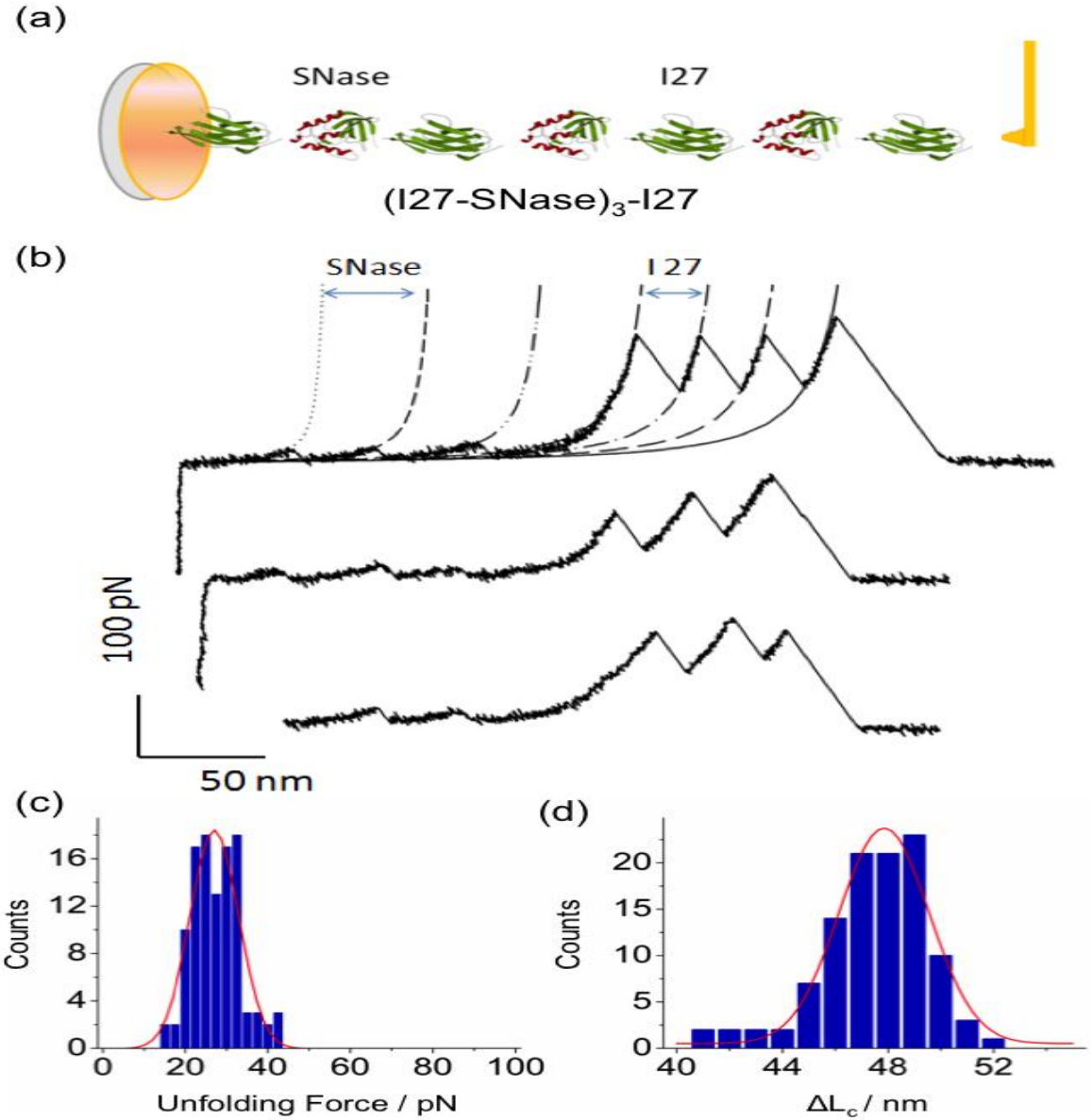
Vectorial, sequential folding may be a common feature of alpha helical stacked repeat proteins

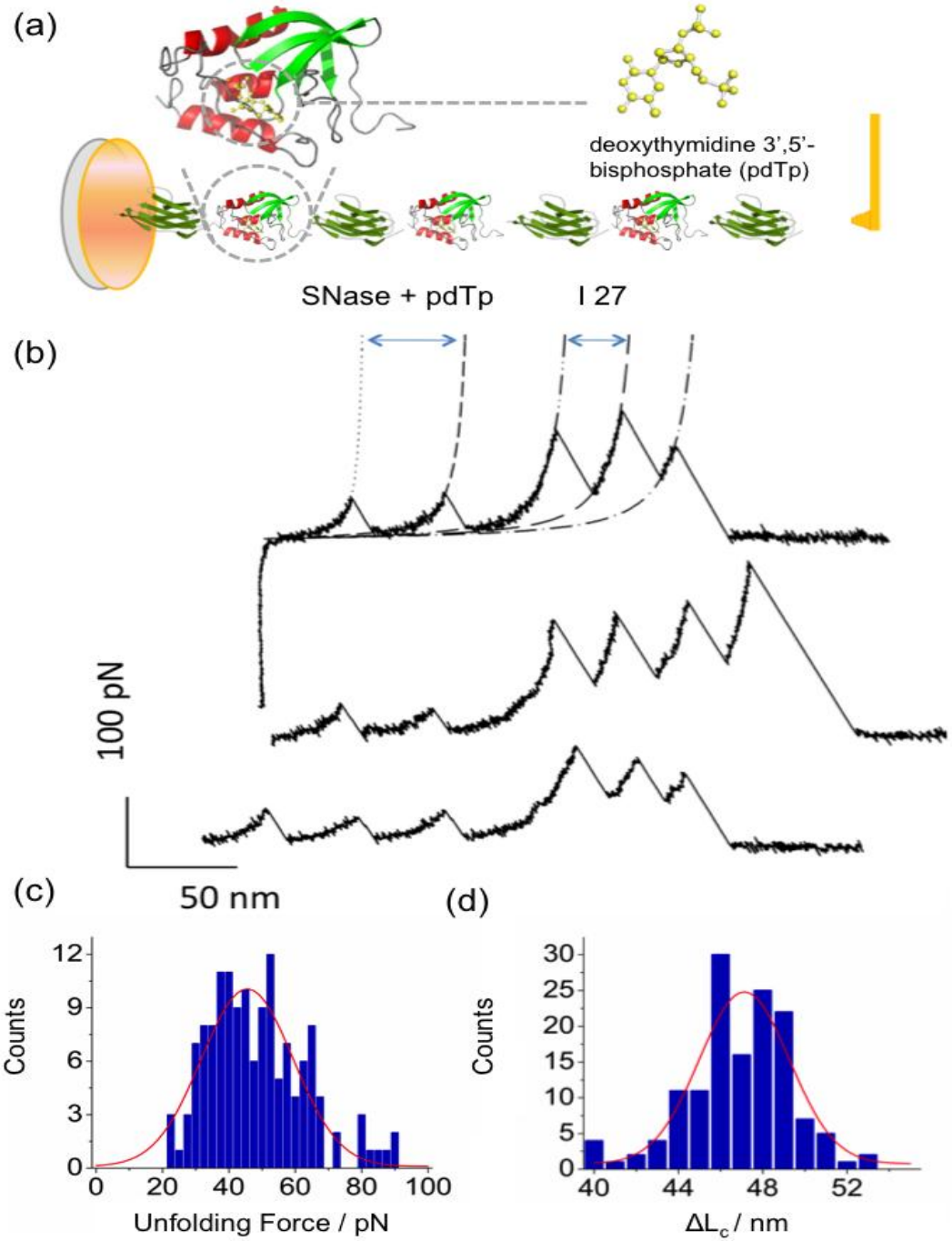
Native contacts topology dictates folding pathways of ankyrin repeat proteins

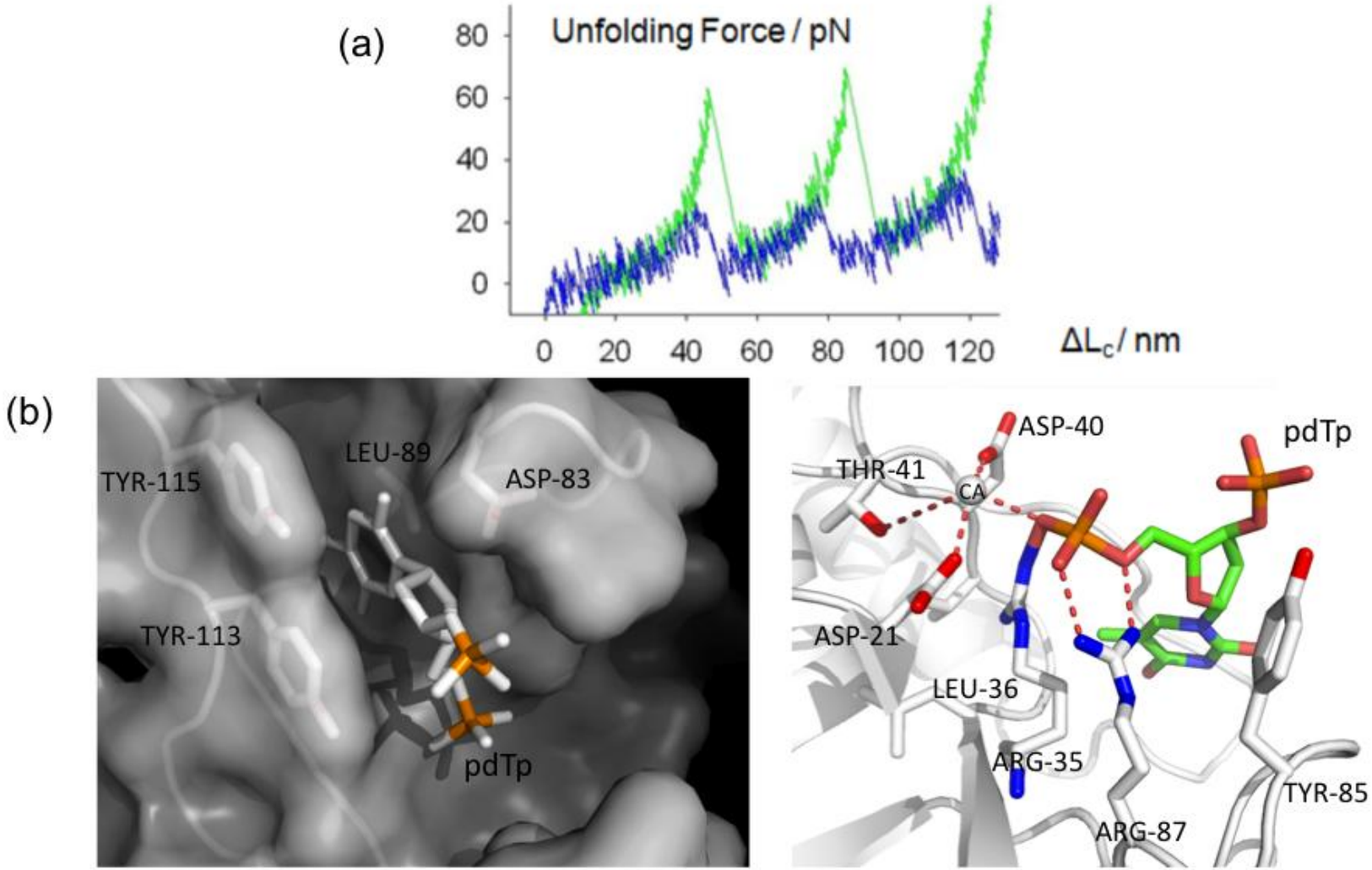
Ankyrin repeat proteins fold via nucleation of several repeats (nucleation may depend on the residual structure in the unfolded chain)

AFM refolding of repeat proteins occurs under 1D constraints, therefore it may reproduce the folding of the Nascent Polypeptide Chain

Protein reference force probes with strong and weak modules




(a)
deoxythymidine 3',5'-
bisphosphate (pdTp)
SNase + pdTp
I 27
(b)
100 pN
50 nm
(c)
Counts
Unfolding Force / pN
(d)
Counts
ΔL$_c$ / nm



Wang, C-C., Tsong, T-Y., Hsu, Y-H., Marszalek, P.E. (2011). Inhibitor Binding Increases the Mechanical Stability of Staphylococcal Nuclease. Biophysical J. 100: 1094-1099.

Acknowledgements

Columbia University
Julio M. Fernandez

UBC/Vancouver
Hongbin Li

Cajal Institute
Mariano Carrion-Vazquez

University of Texas, Galveston
Andres F. Oberhauser

Mayo Clinic
Yuan-Ping Pang

Nicholaus Copernicus University
Wiesław Nowak

Univ. of Illinois, Urbana-Champaign
Klaus Schulten

UNC/Chapel Hill
Piotr A. Mieczkowski

Duke /*Chemistry*
Weitao Yang
Zhenyu Lu
Xiancheng (Fox) Zeng

Florida State University
Huan-Xiang Zhou

Duke/Biochemistry
Paul Modrich
Celia Baitinger

Duke/Cell Biology
Vann Bennett
Khadar Abdi

NC State/Biochemistry
Hanna Gracz

PEM Lab:
Gwangrog Lee, Qingmin Zhang,, Changhong Ke , Michael Humeniuk
Yong Jiang, Whasil Lee, Minku Kim, Mahir Rabbi, Anna Loksztejn

NSF: MCB-0450835 and 0717770 , NIH: GM071197, GM079563