Marek Cieplak
Institute of Physics, PAS, Warsaw, Poland

Stretching to understand proteins

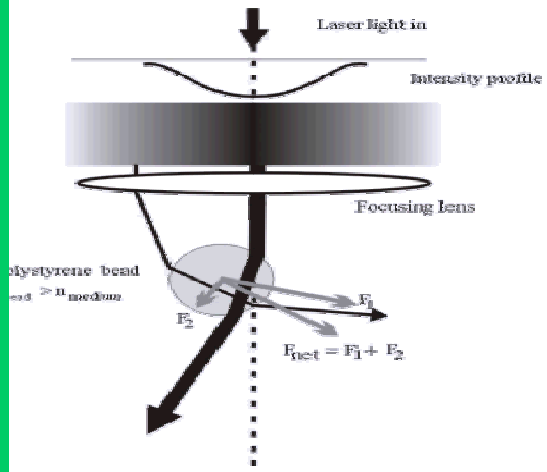
Stretching of bridge pylons

No rupture

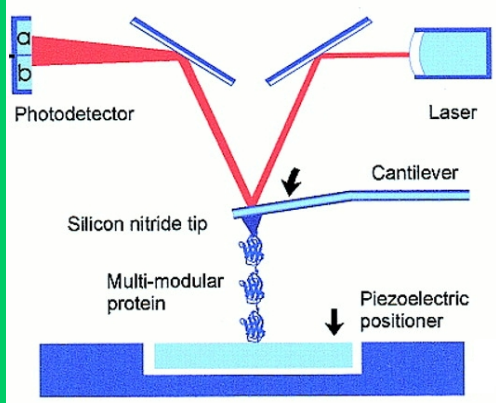
An adequate force is needed to generate rupture to learn about the structure

MANIPULATION WITH SINGLE BIOMOLECULES: 10-300 pN

Optical tweezers

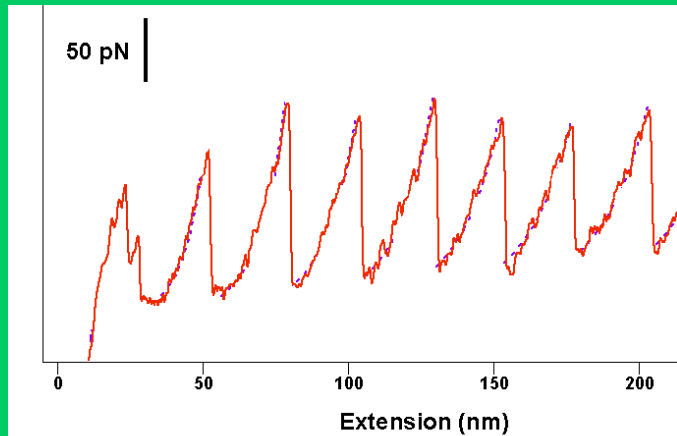


Atomic force microscope



Characteristic scale of the force

F_{max}: for titin ~ 200 pN



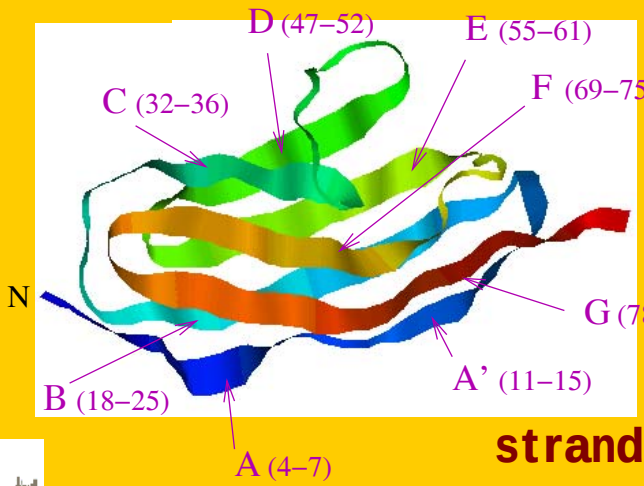
Fowler Best Toca-Herrera Rutherford
Steward Paci Karplus Clarke 2002

PROTEIN-DEPENDENT PATTERNS – NEE

TITIN

3.5 MDa – gigantic muscle protein

~ 300 globular domains

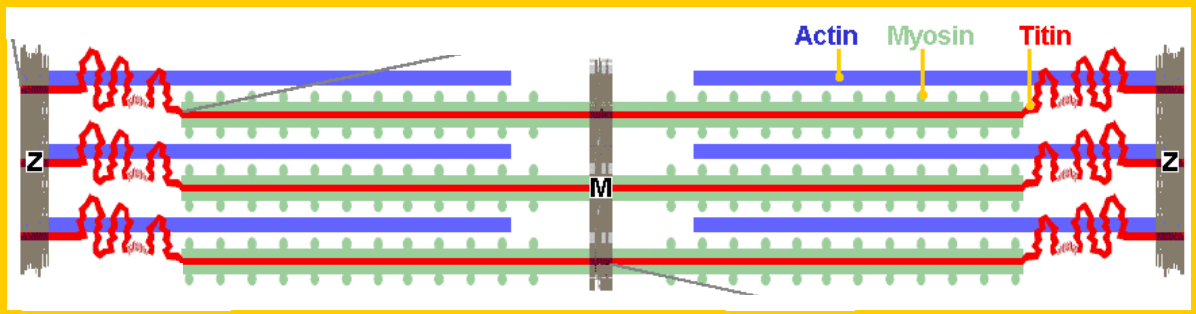


strands

β-sandwich architecture

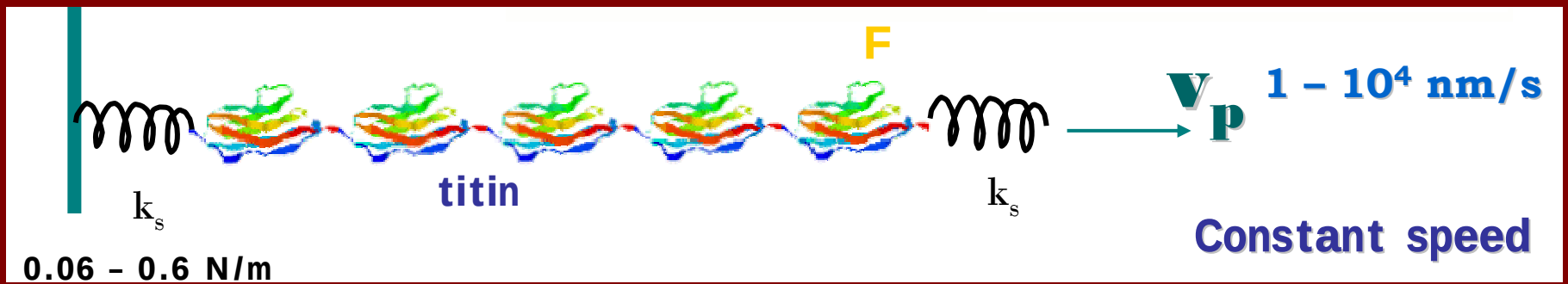
sarcomere

1μm long (x 4)



Z-disk

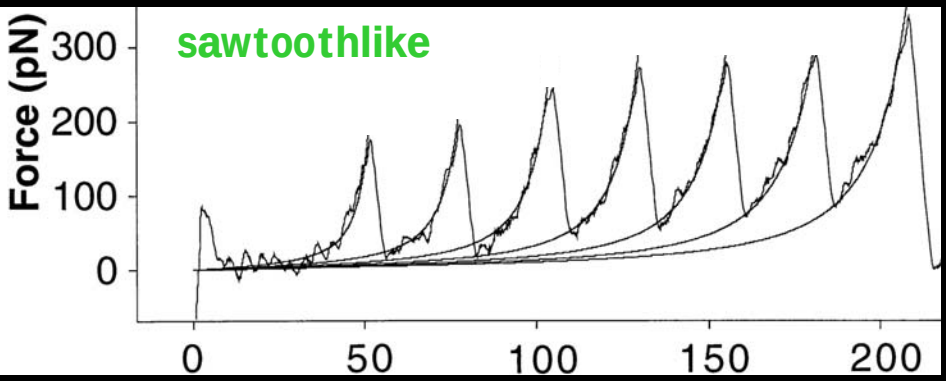
M-line

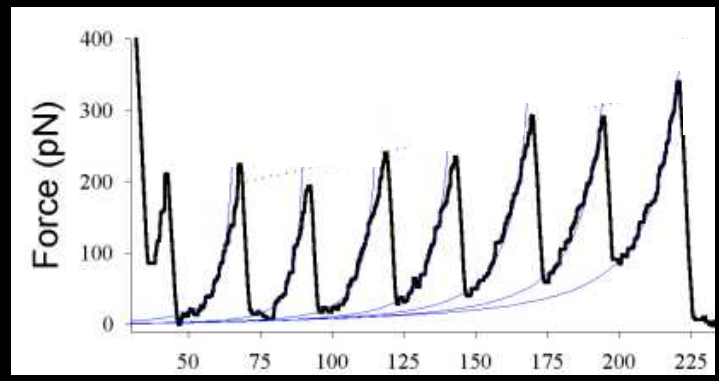


'TITIN IS A WEIRD SPRING' Erickson 1997

SERIAL UNWINDING

I27-I34





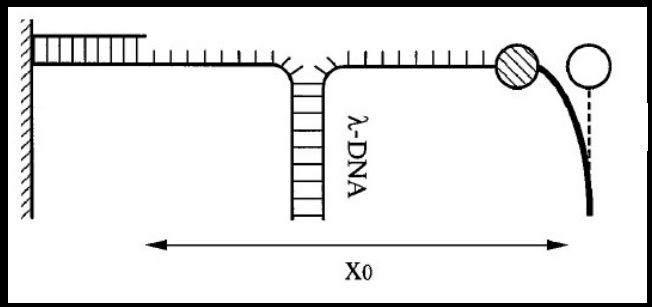
Extension [nm]

Rief Gautel Oesterhelt Fernandez Gaub 1997

Li Linke Oberhauser Carrion-Vazquez Kerkviliet Lu Marszalek Fernandez 2002

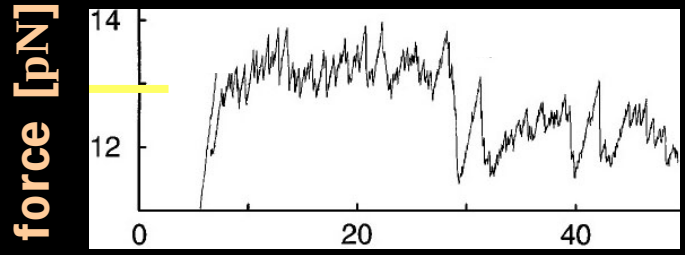
Apparatus effects at small extensions

DNA



Bockelmann Essevaz-Roulet Heslot 1997

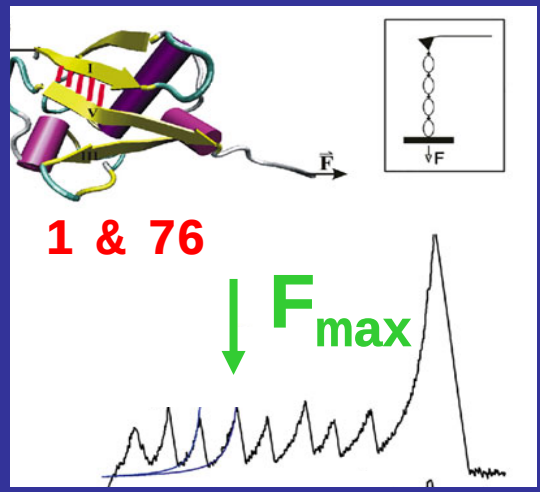
Each protein has its own pattern

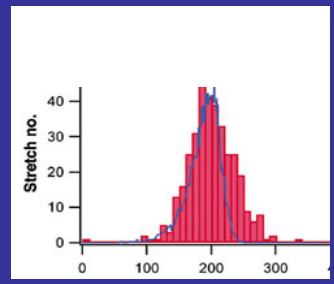


Linkage dependent elasticity

UBIQUITIN N=76

Studies usually involve homo- or hetero-linkages of modules



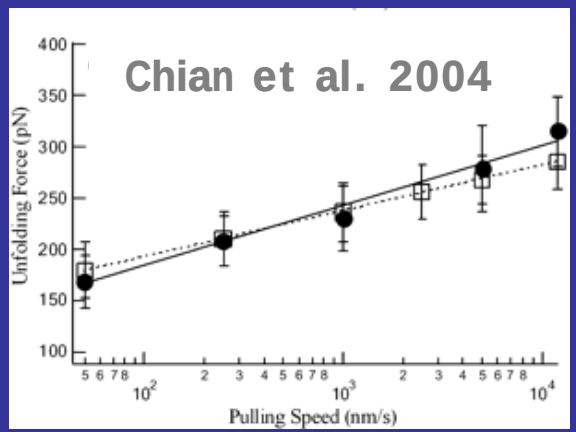


203±35 pN

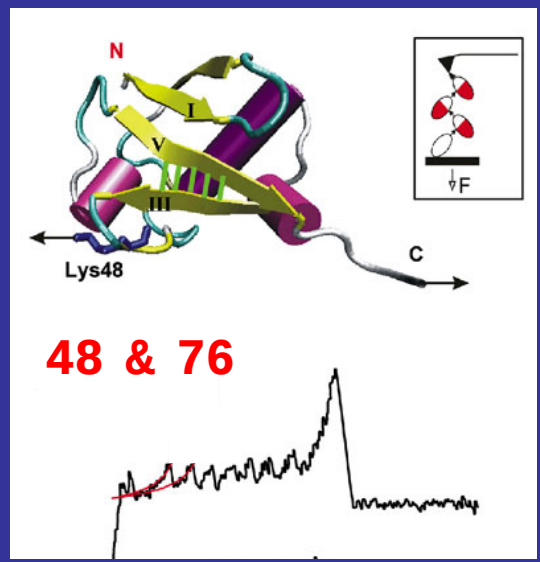
Assumption of seriality

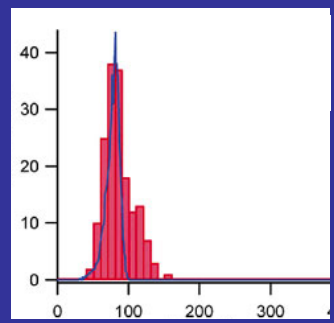
~ logarithmic dependence on v_p

Expected: a constant at v_p->0



would suggest ~600 pN at 10^{10} nm/s





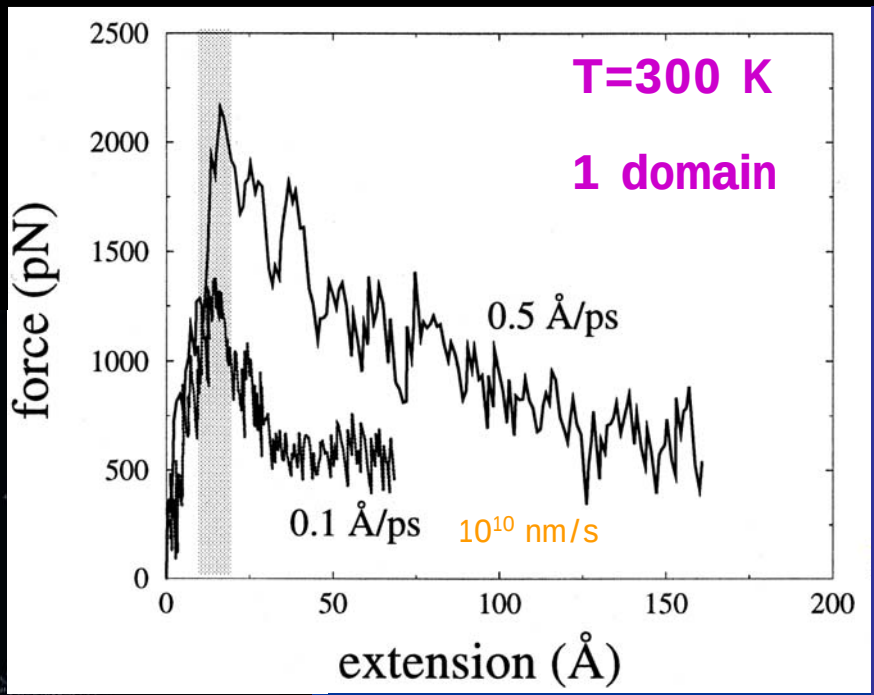
85±20 pN

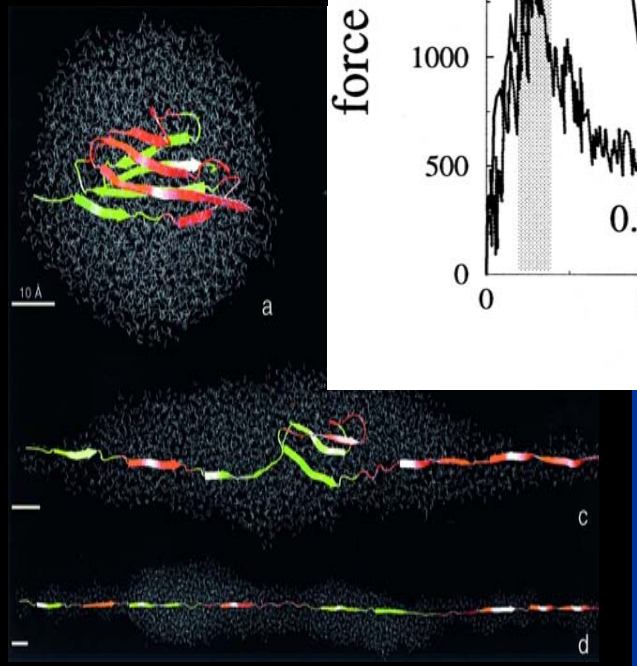
Carrion-Vazquez, Li, Lu, Marshalek, Oberhauser, Fernandez 2003

ALL-ATOM SIMULATIONS

Lu Schulten 2000 (Paci Karplus 2000)

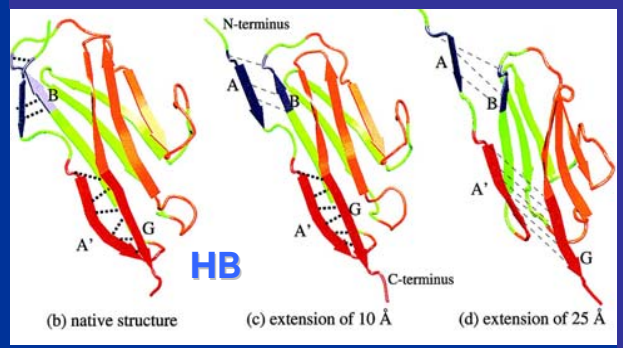
typically 10 ns time scales





(Hydrodynamic interactions reduce F_{max})

Pabon, Amzel 2006 – quasistatic ~500 pN

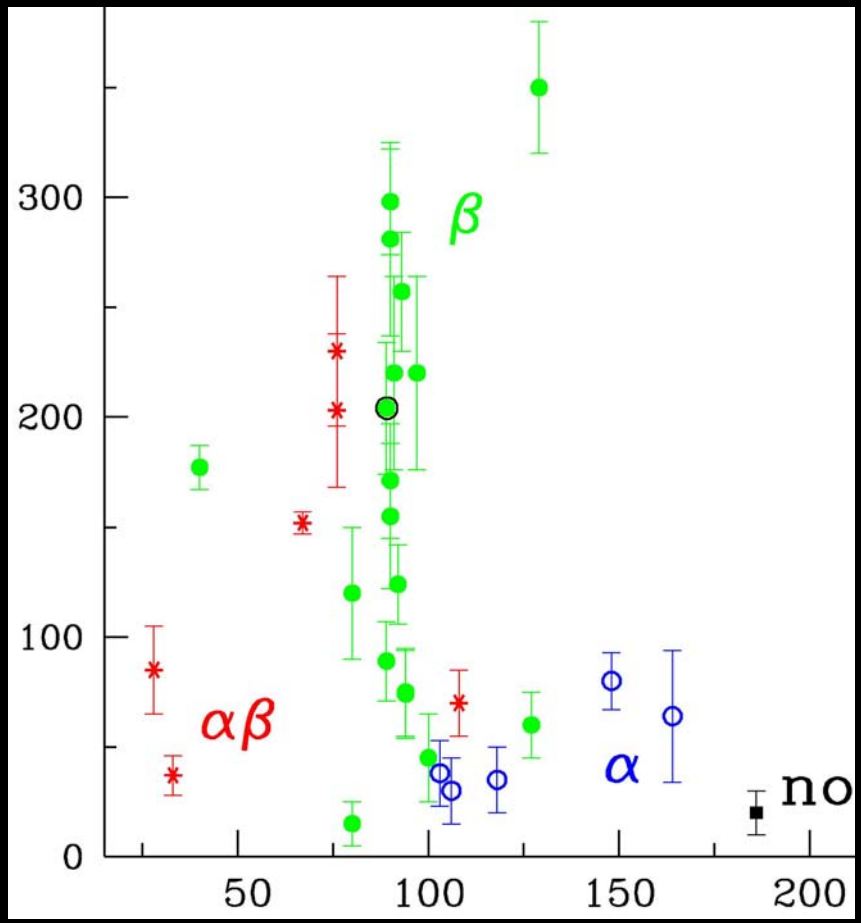


Difficult:

comparison of processes involving large conformational changes

studies of large sets of proteins

Experimental results on stretching at constant speed



F_{max} [pN]

N – sequential length

~55 proteins

All-atom simulations on ~ 21 proteins

A need for systematic studies across the PDB to generate understanding and explore the possibilities

J. Fernandez

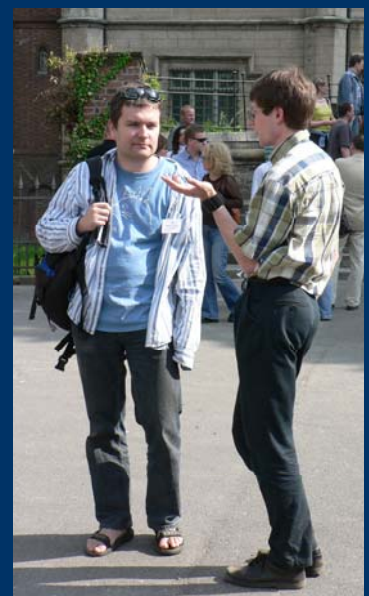
What proteins are strong and why?

Simplified Go-like models: Big proteins, many domains, variations of parameters, near-experimental v_p

1. *Theoretical survey of 7749 proteins within a coarse-grained Go-like model – stretched at constant speed* **Topical review in J. Phys.: Cond. Matt.**
2. *Stretching at constant force*
3. *Stretching by fluid flow*

Joanna I. Sułkowska Institute of Physics, PAS, Warsaw, Poland

Piotr Szymczak, Warsaw University, Poland



Trinh Xuan Hoang Institute of Physics, Hanoi, Vietnam ~2000

Mark O. Robbins Johns Hopkins University, Baltimore, MD, USA

Go models of proteins – coarse grained: only the C^{α} atoms

Constructed from the experimentally derived native structure

MODEL

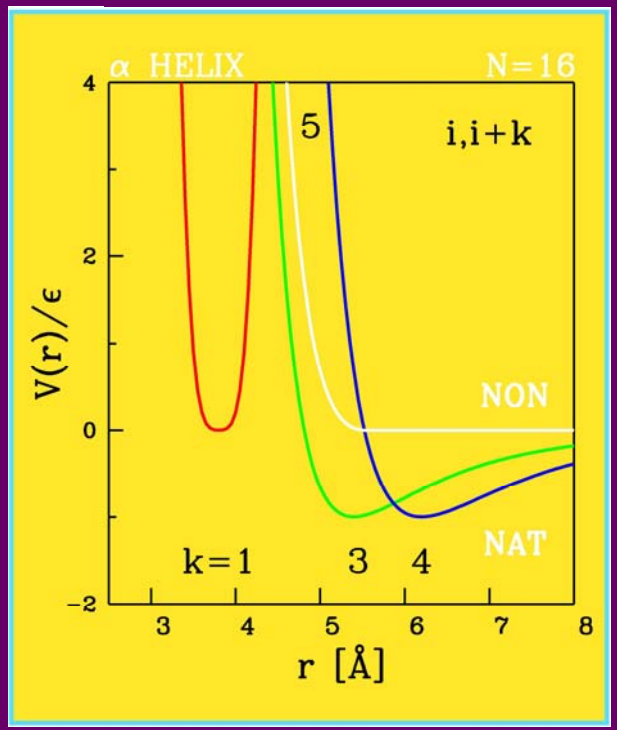
$$E_p(\{\mathbf{r}_i\}) = V^{BB} + V^{\mathrm{NAT}} + V^{\mathrm{NON}} + V^{\mathrm{CHIR}}$$

V^{BB} – TETHERING of consecutive beads at 3.8 Å $=d_0$

$$V^{\mathrm{NAT}} = \sum_{i<j}^{\mathrm{NAT}} 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]$$

ε: 1.6 kcal/mol ~ 800 K

Room T: 0.35 ε



σ_{ij} calculated based on the van der Waals radii of the atoms
Tsai Taylor Chothia Gerstein 1999

Non-native: repulsive with σ=4Å

Disulfide bonds enhanced

V^{CHIR}: angular terms locally favoring the native shape of the backbone

MOLECULAR DYNAMICS

$$m\ddot{\mathbf{r}} = -\gamma\dot{\mathbf{r}} + F_c + \Gamma$$

$$F_c = -\nabla_r E_p$$

$$\langle \Gamma(0)\Gamma(t) \rangle = 2\gamma k_B T \delta(t)$$

$$\tau = \sqrt{ma^2/\epsilon} \sim 3ps \qquad \text{(small } \gamma\text{)}$$

$$a = 5\mathring{A} \approx < \sigma_{ij} >$$

m=118m$_p$

Langevin noise as a thermostat and as an emulator of water

Large friction
Finite bead size

modify the effective time scale τ

Veitshans, Klimov, Thirumalai 1997: τ~3ns

τ=0.25 ns

when comparing with experiments

Stretching ~independent of γ

Folding time linear in γ

Use γ=2m/τ

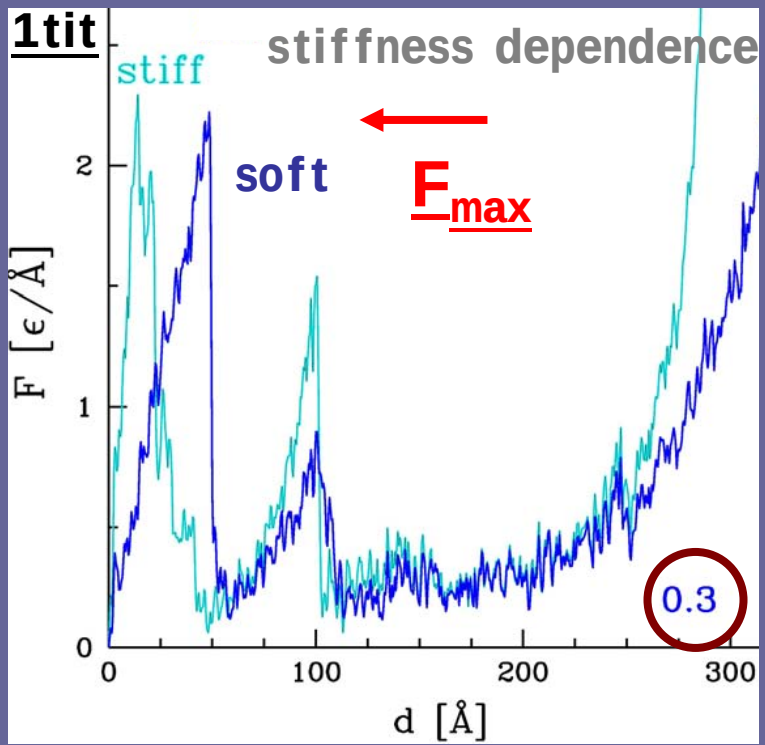
Water-like γ ~ 25 times bigger

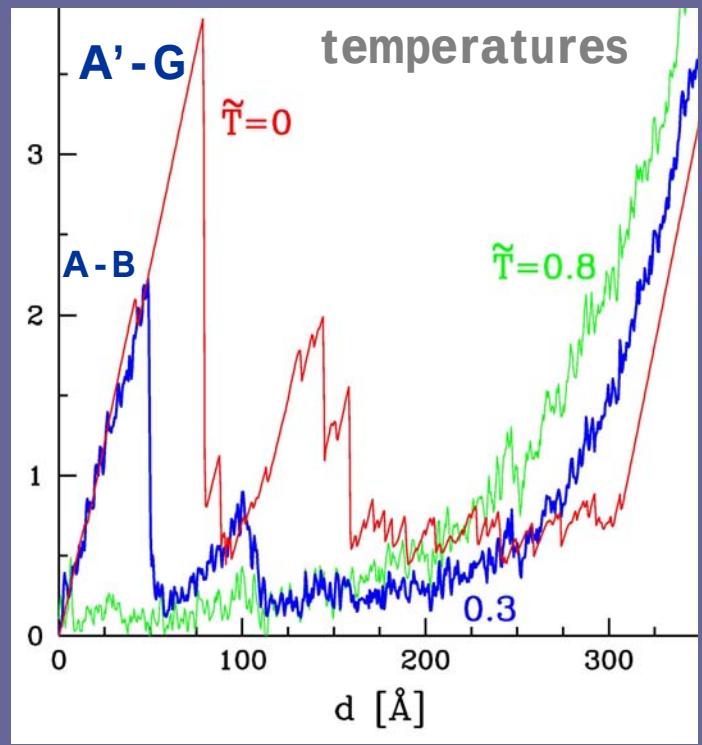
v_p=0.005 Å/τ ~ 10^6 nm/s

logarithmic shifts with v_p

F_{max} does not depend on the AFM stiffness but depends on T

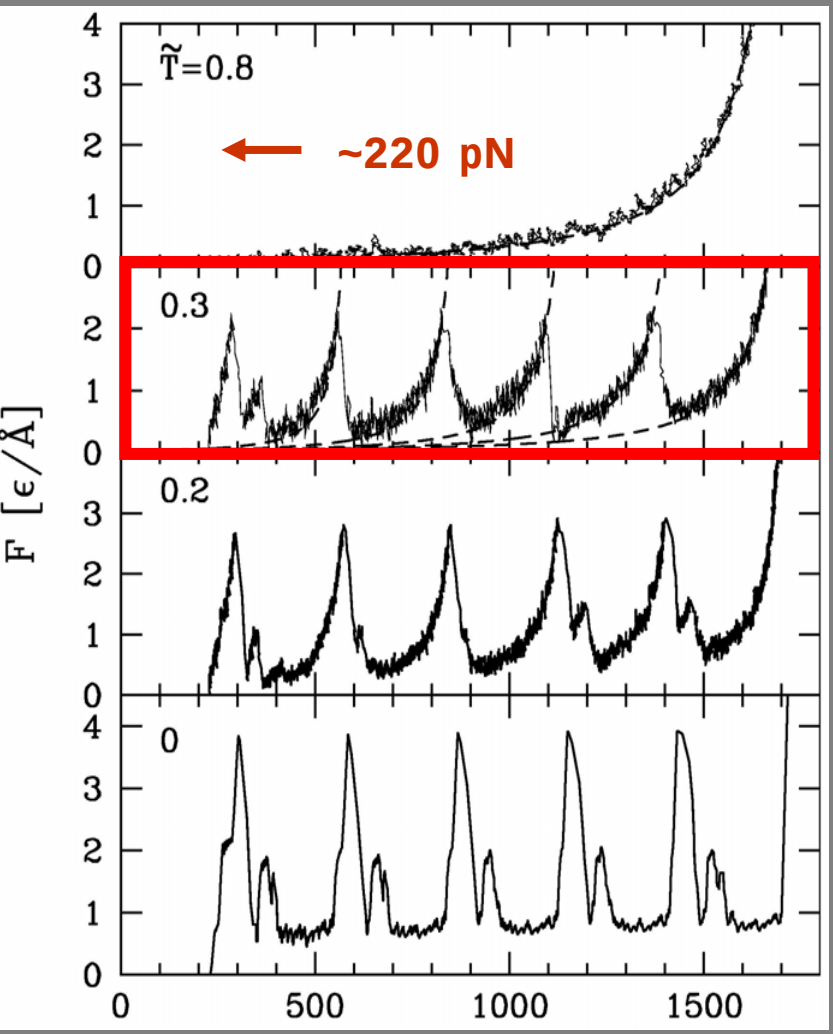
one domain



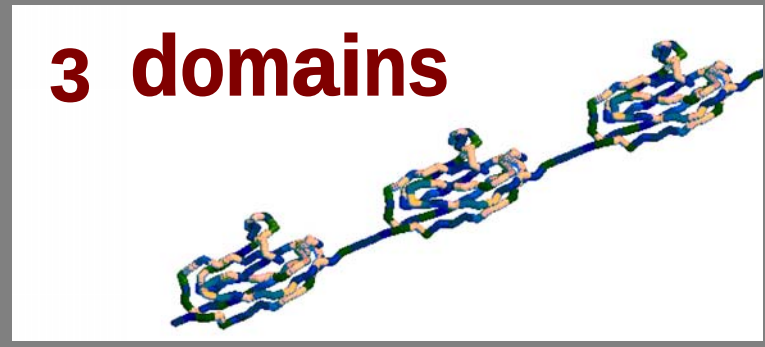


The pattern depends on T and at T=0.3 it is similar to the experimental results

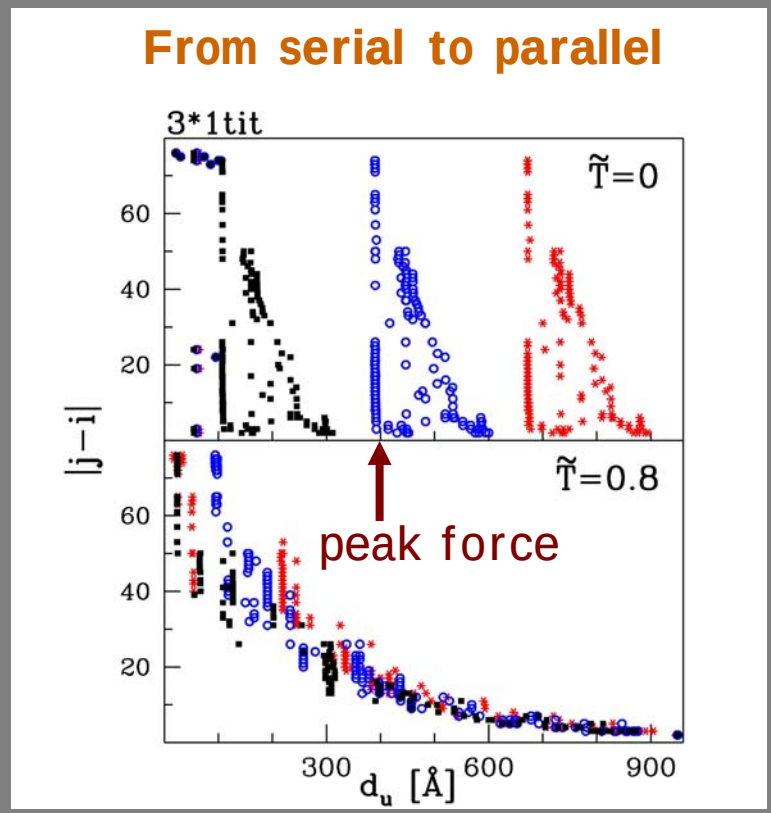
5 x 127



end-to-end distance



breaking native contacts



tip displacement

contacts – identified by the sequential distance

Validation of the Go model for stretching

Uniform ε

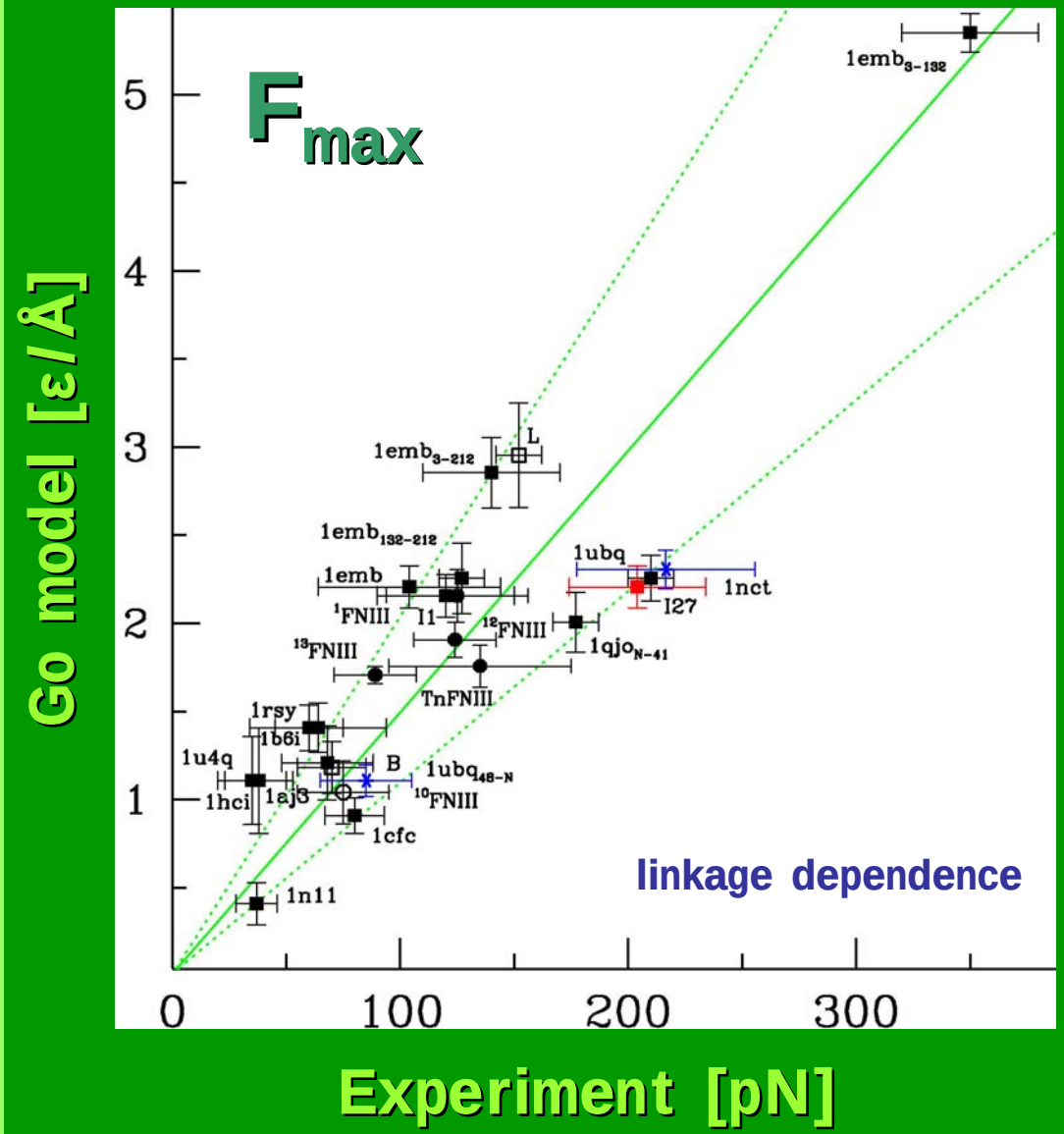
Constant speed pulling

Correct contact map – should work close to the native state

ε=1.6 kcal/mol

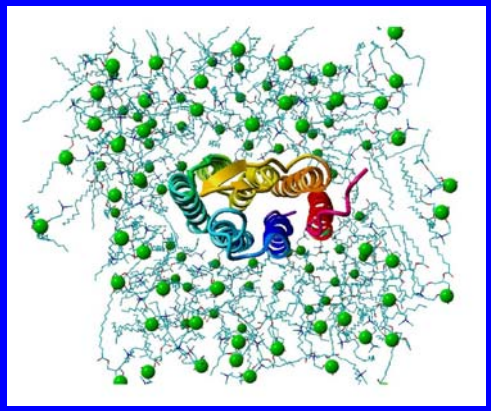
→ ε/Å=110 pN

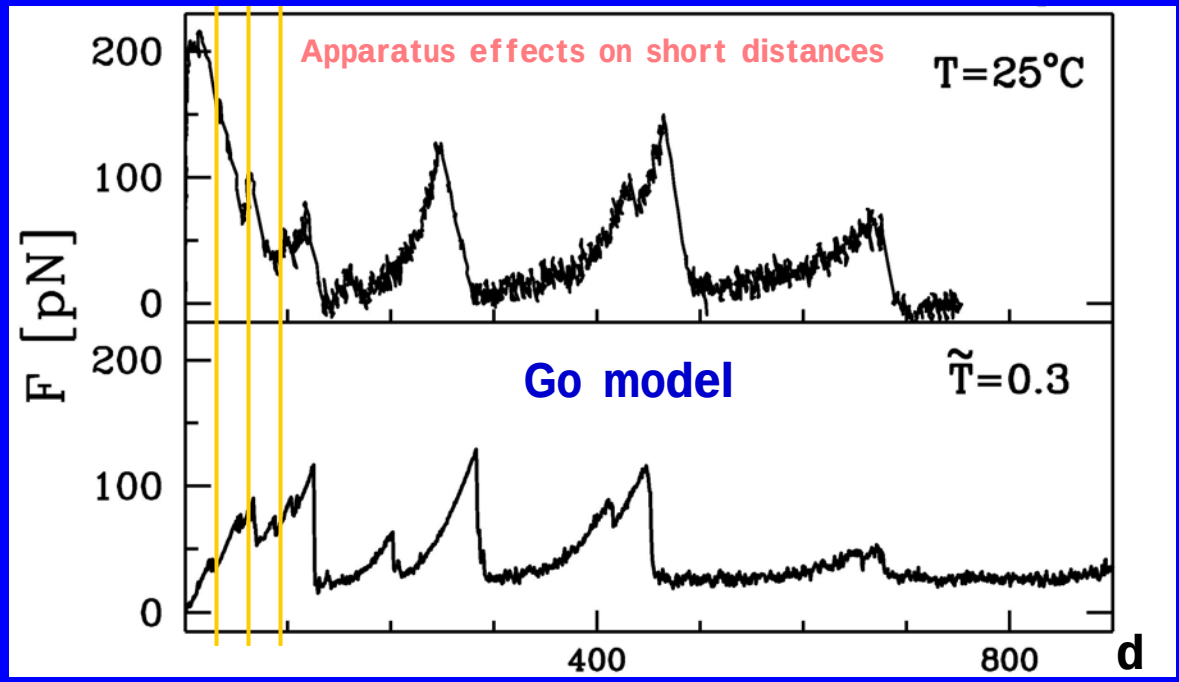
@ T=0.3ε/k_B



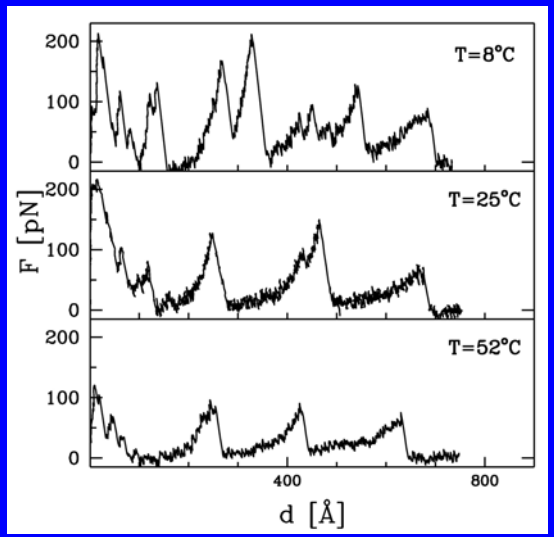
Bacteriorhodopsin pulled out of a membrane

(By the C-terminus)





Janovjak et al. 2003



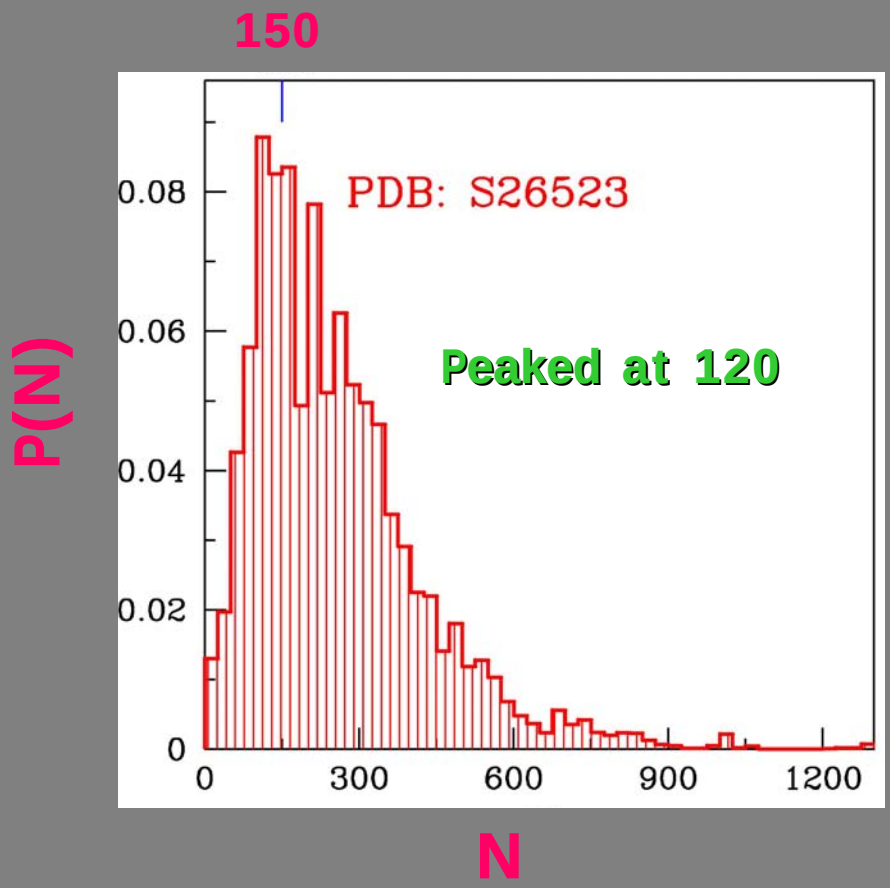
with S. Filipek, K. Krzyśko, H. Janovjak - 2006

Protein Data Bank: 29385 structures on July 26 2005

~15000 proteins not in complexes

Studied:

7 510 proteins with 40 ≤ N ≤150: set S7510



239 with 150 < N ≤ 851

7 749

3 813 with 40 ≤ N ≤150:

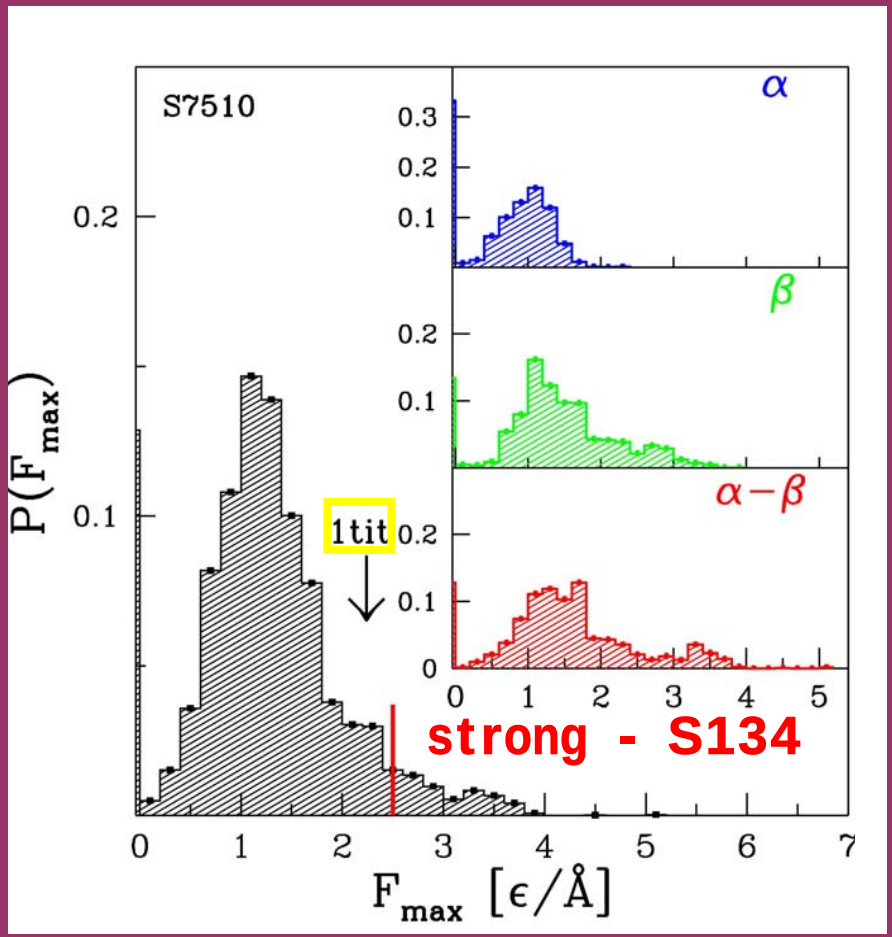
CATH-based assignment of topology to the fold available

Hierarchy:

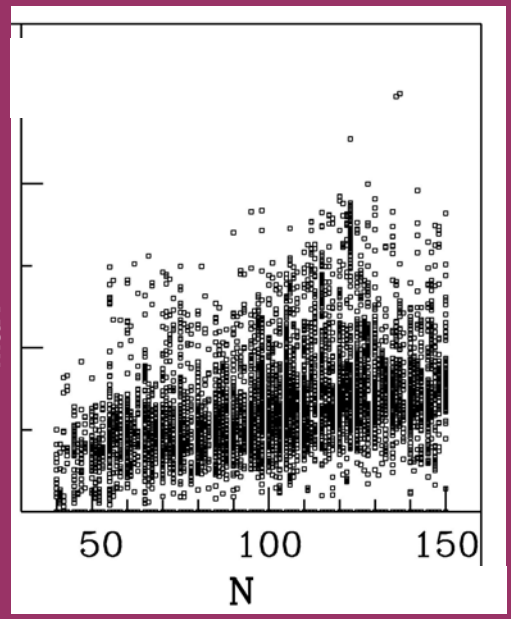
Class
Architecture
Topology
Homology

Probability distribution of F_{max}

short **CLASS**

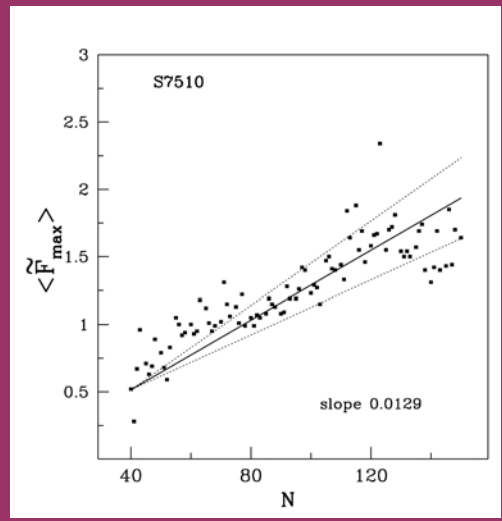


no strong α proteins



Pulling by the termini

Average F_{max} for N

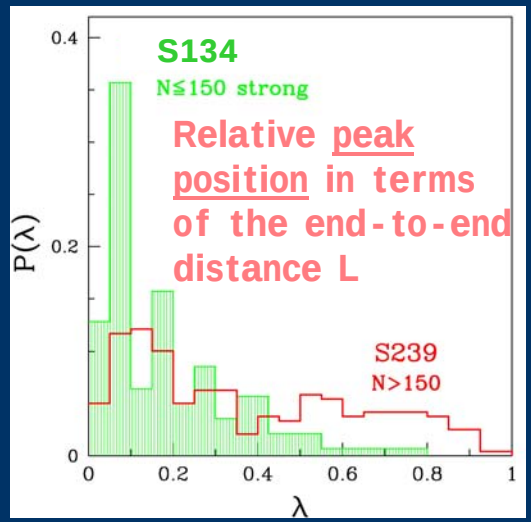


The spread in F_{max} depends on N only weakly, but the larger the N the bigger the chance of a large peak force

THE PREDICTED LIST

The strongest proteins with N≤150

n	PDB	F	λ	%CATH
1	1c4p	5.1	18	3.10.20
2	1qqr	5.1	19	3.10.20
3	1g1k	3.9	6	2.60.40
4	1c76	3.8	17	3.10.20
5	1c77	3.8	25	3.10.20
6	1c79	3.8	25	3.10.20
7	1aoh	3.7	6	2.60.40
8	1c78	3.7	25	3.10.20
9	2sak	3.7	18	3.10.20
10	1nam	3.7	9	2.60.40
11	1so9	3.6	18	2.60.370
12	1ppx	3.5	40	3.90.79
13	1ssn	3.5	34	3.10.20
14	1rnz^s	3.4	45	3.10.13
15	1ie5^s	3.4	14	2.60.40
16	1b88	3.4	7	2.60.40
17	3rsk^s	3.4	45	3.10.130
18	1npu	3.4	8	2.60.40
19	2ncm	3.3	8	2.60.40
20	1anu	3.3	7	2.60.40
21	1rlf	3.3	9	3.10.20
22	1eaj	3.3	8	2.60.40
23	1oo2	3.3	26	2.60.40
24	1h5b^s	3.2	8	2.60.40
25	1i9e	3.2	4	2.60.40
26	1mvf	3.1	12	2.60.40
27	1f5w	3.1	7	2.60.40
28	1sp0	3.1	16	2.60.370
29	1amx	3.1	31	2.60.40
30	1i3o	3.1	49	3.40.50
31	1tfp	3.1	21	2.60.40
32	1ves	3.1	6	
33	1sn0	3.1	21	2.60.40
34	1oau	3.1	5	2.60.40
35	1sn2	3.0	21	2.60.40
36	1oax	3.0	5	2.60.40
37	1oar	3.0	5	2.60.40
38	1pun	3.0	41	3.90.79
39	1j05	3.0	5	2.60.40
40	1lve^s	3.0	6	2.60.40
41	1fvc	2.9	5	2.60.40
42	1p7e	2.9	14	3.10.20
43	1jhl	2.9	6	2.60.40
44	1gke	2.9	27	2.60.40
45	1etb	2.9	26	2.60.40
46	1vhp	2.9	3	2.60.40
	1tit	2.1	4	2.60.40
	1ubq	2.2	6	3.40.50



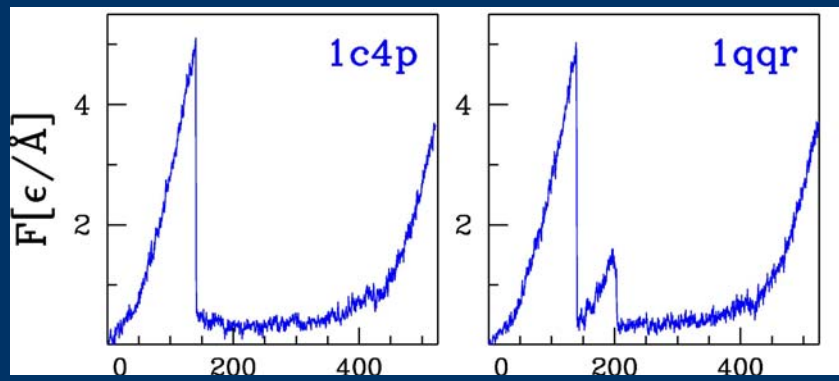
Top 134 proteins:

6 architectures

9 topologies

1c4p & 1qqr: β domain of streptokinase

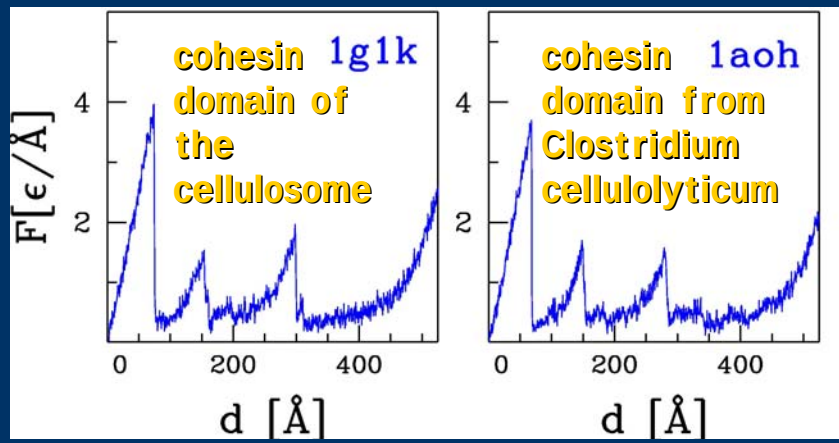
(blood clotting - different functions)



α-β rolls 30%

β sandwiches 60%

3-layer sandwiches



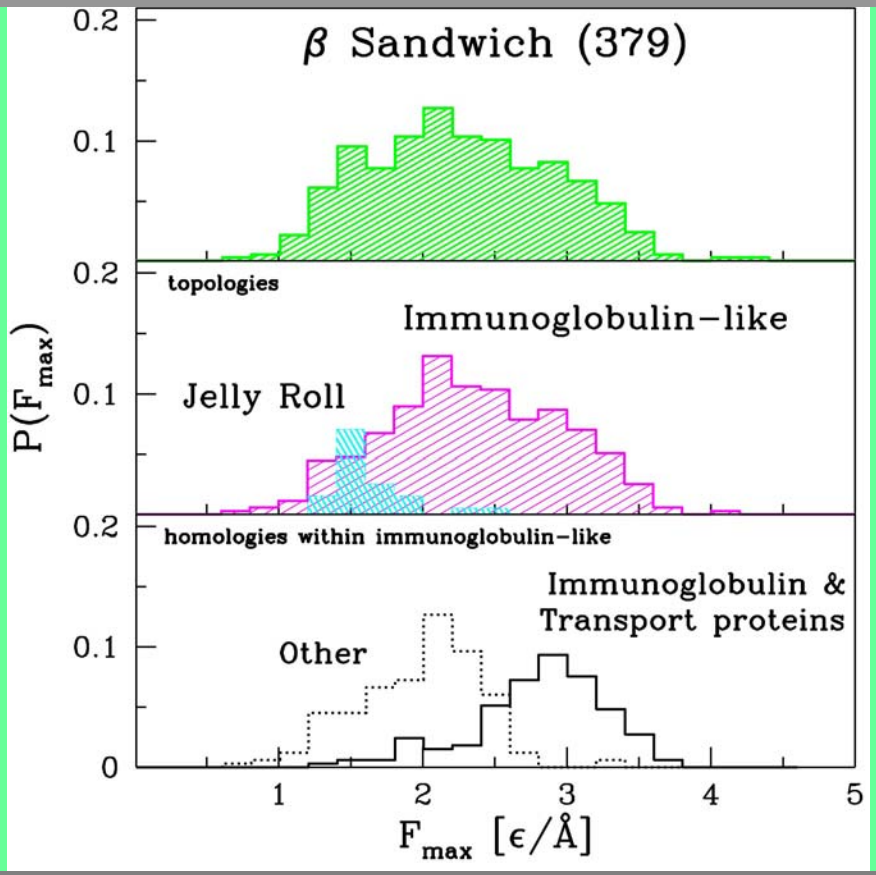
number of peaks

Resolving architectures into topologies

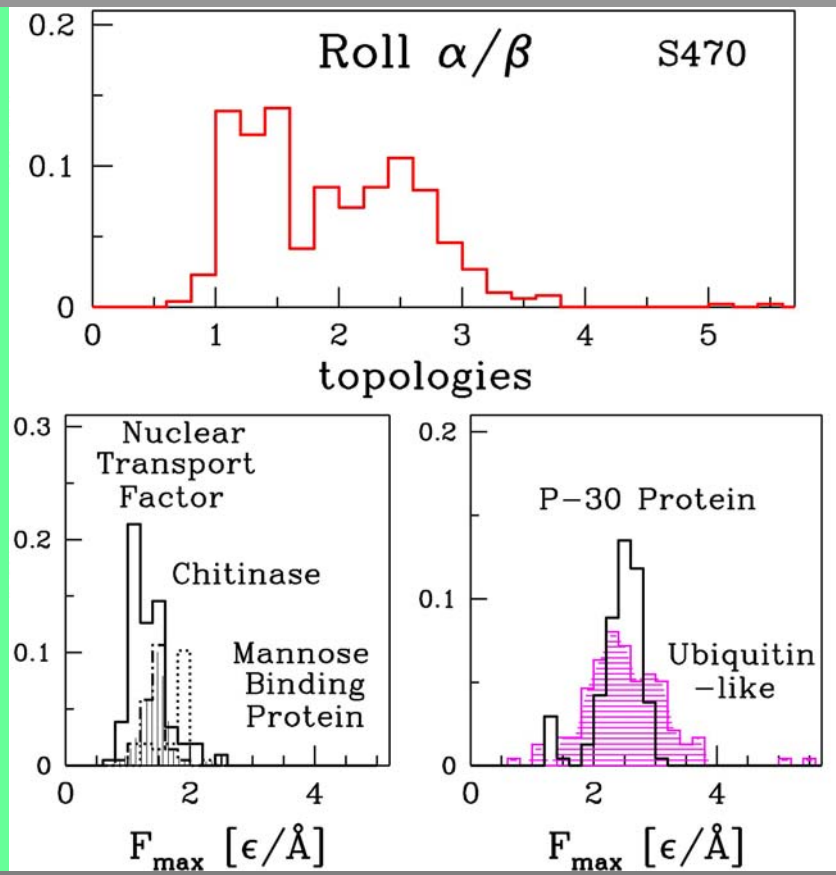
A

T

H



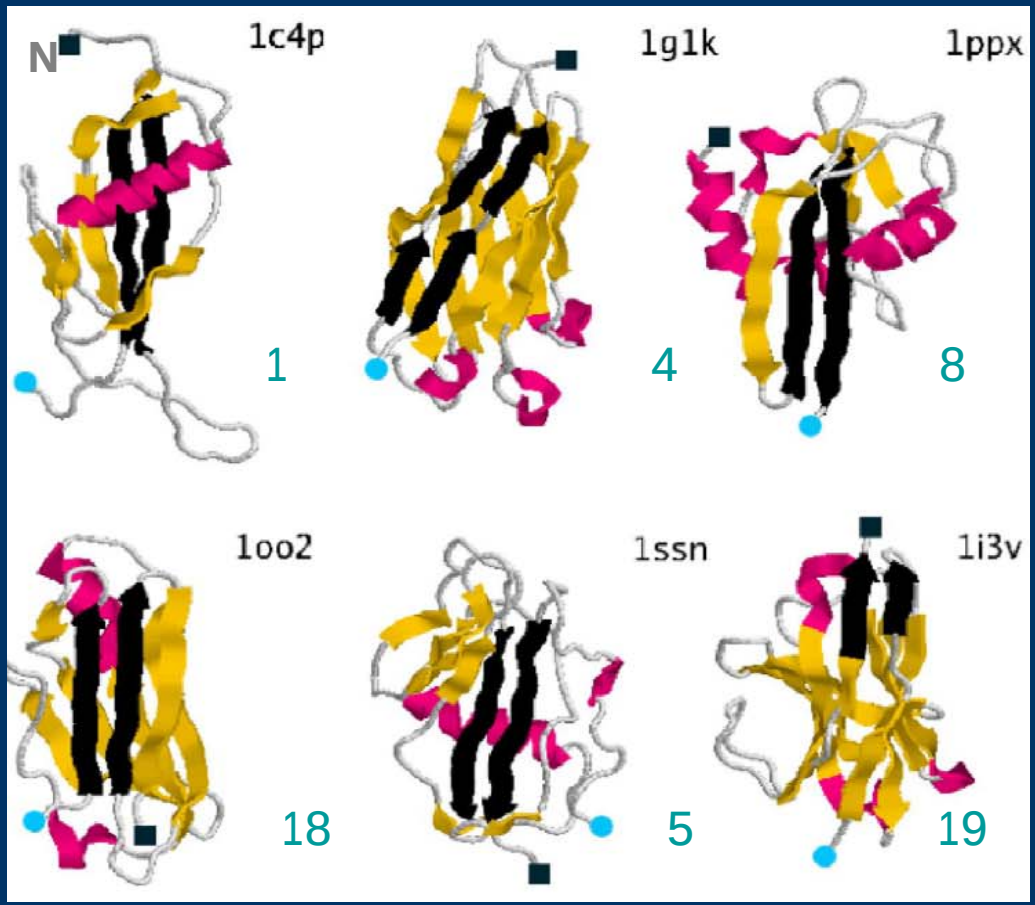
like titin



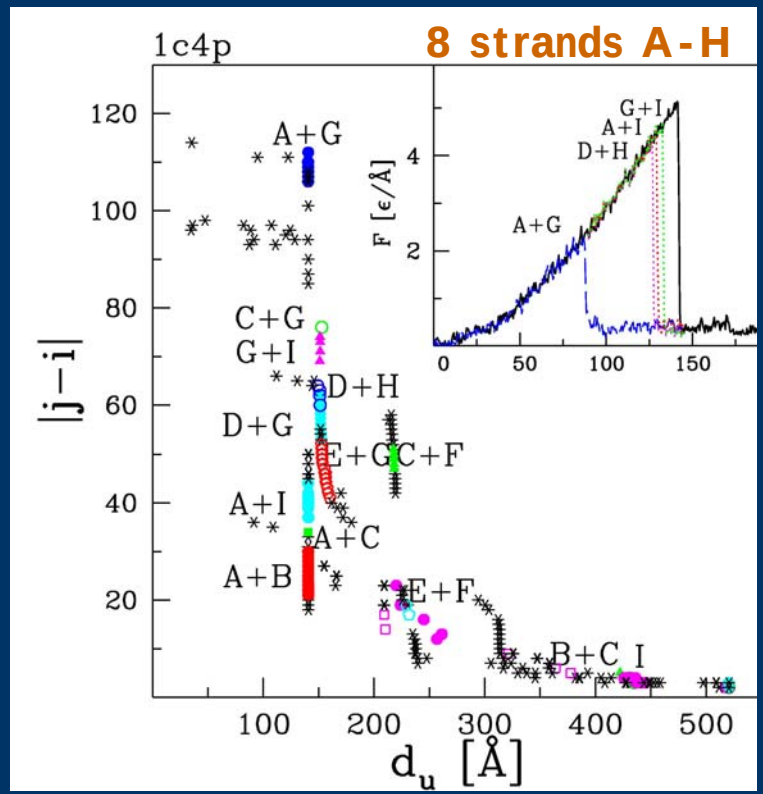
like ubiquitin

Mechanisms of rupture in strong short proteins

95%: shearing of hydrogen-bonded parallel β-strands shown in black



Unfolding scenario diagram –
more detailed than Q – the fraction of the total number of native contacts



Strength depends on the length of the mechanical clamp and on the environment of the clamp

Novel kinds of mechanical clamps

Antiparallel β-strands (50% of the force)

Unstructured clamps

A Box structure:

two antiparallel strands

and two antiparallel helices

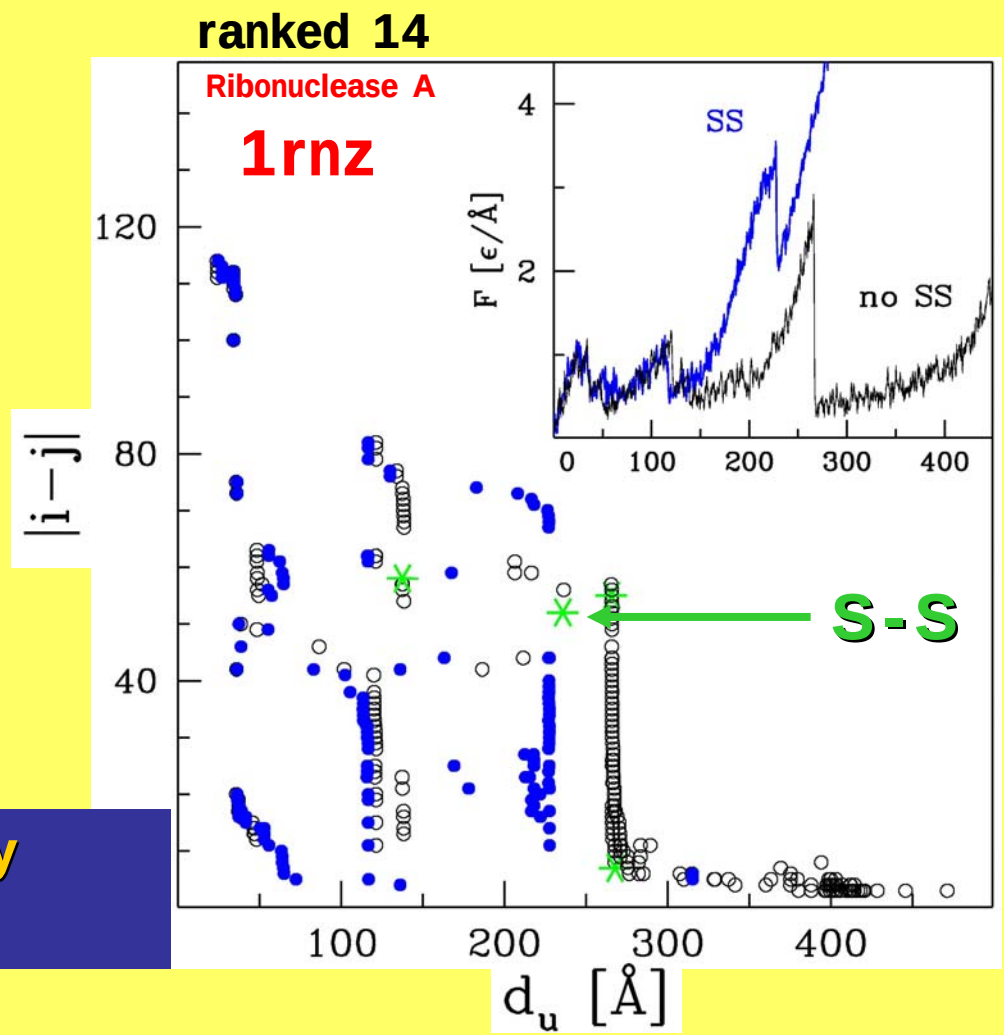
Delocalized clamps

Sulphide bonds cannot be ruptured

S-S contacts enhanced by the factor of 20.

Minor shifts of 9 proteins in S134

Convert S-S to S-H bonds by using the reducing agent DTT

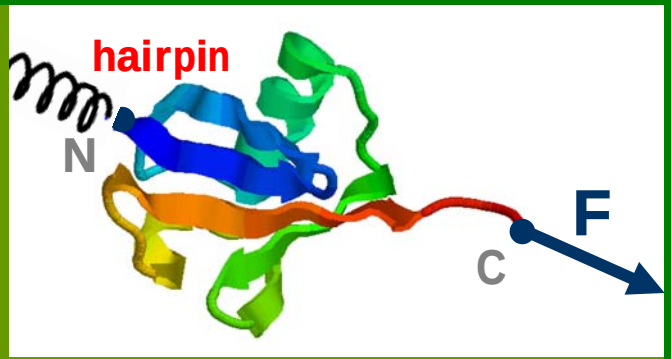


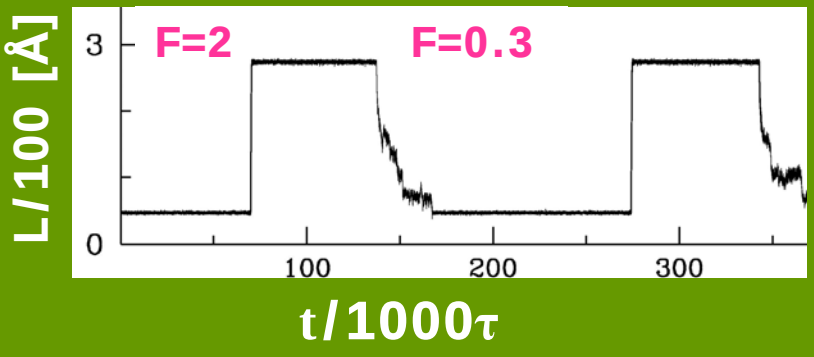
Dischер et al. 2001, 2004: Ig-CAM, CAM, VACM - comparisons

STRETCHING AND FOLDING IN A FORCE-CLAMP

1tit: Oberhauser, Hansma, Carrion-Vazquez, Fernandez 2001

1ubq: Schlierf, Fernandez 2004 Fernandez, Li 2004

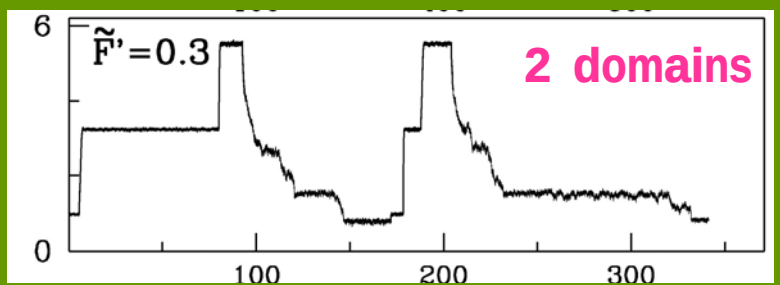


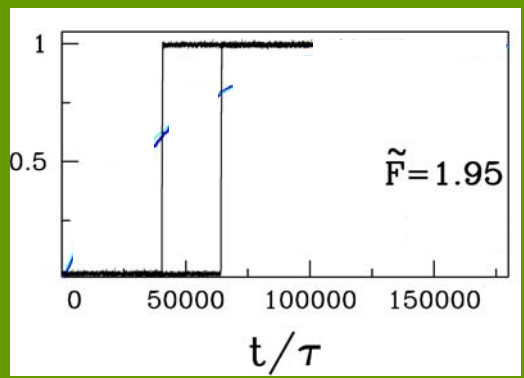


1ubq – Go model

unfolding in a single kinetic step refolding – in multiple steps

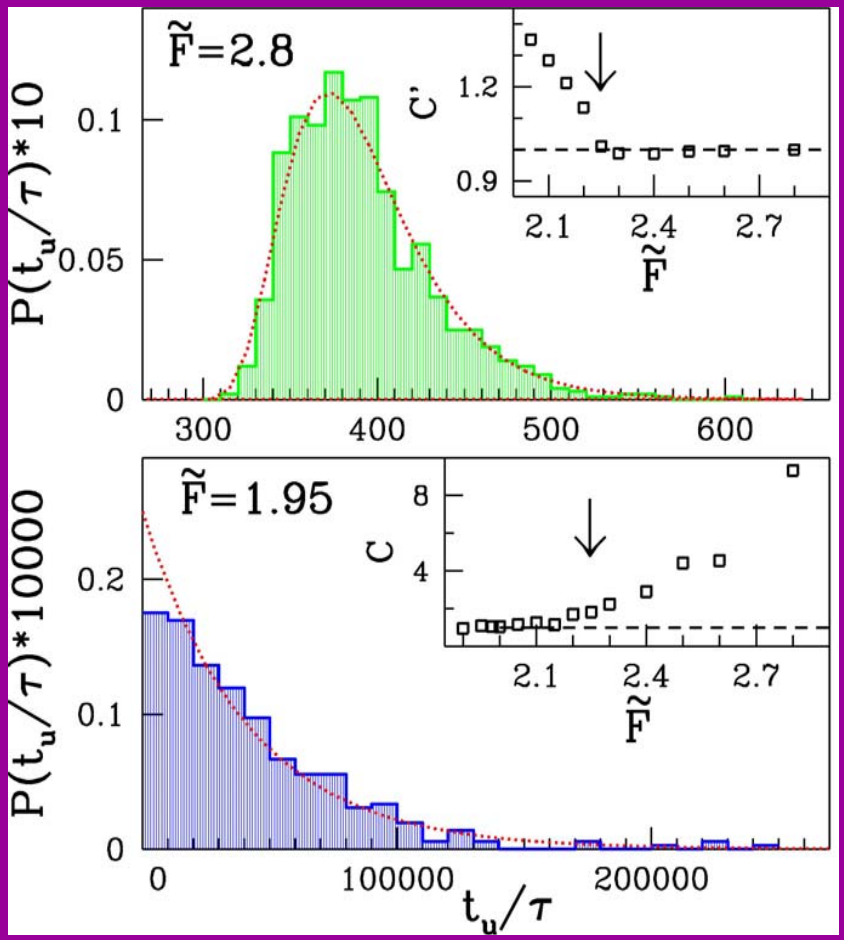
Fractional extension for 2 trajectories





DISTRIBUTION OF UNFOLDING TIMES

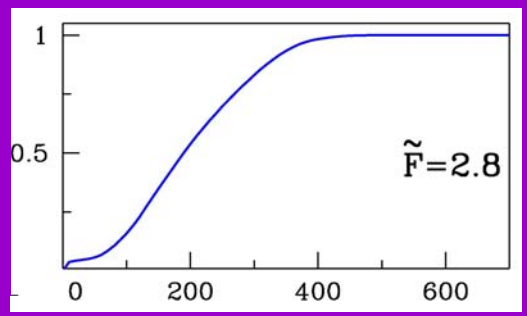
Log-normal above F_{max}

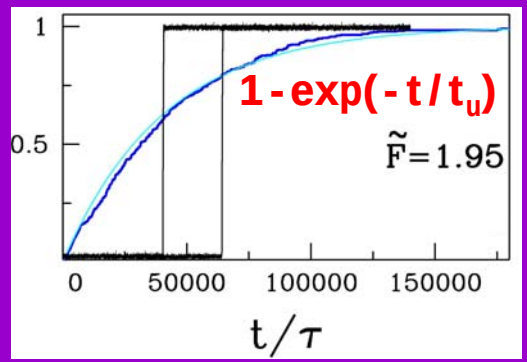


Exponential below F_{max}

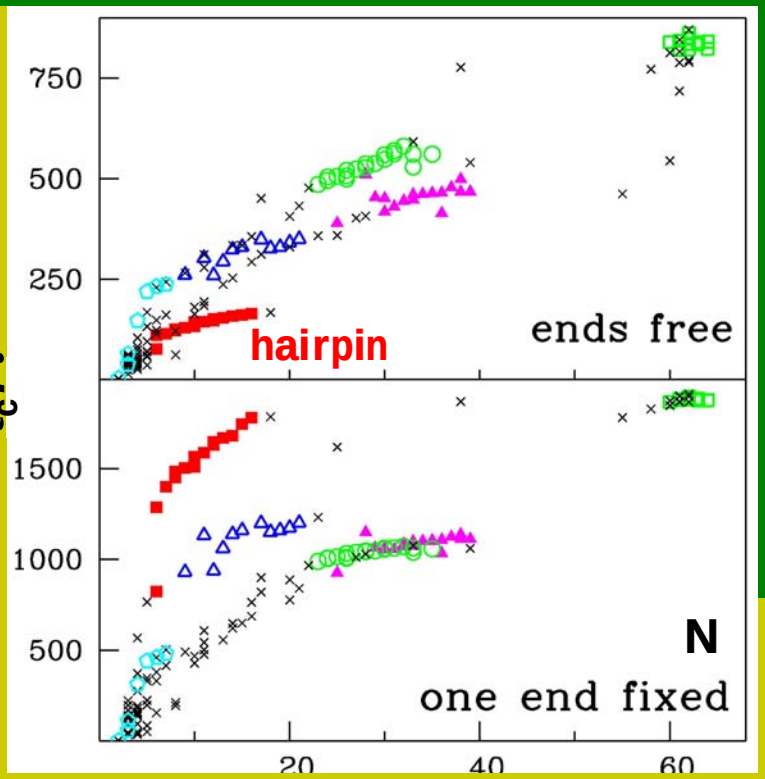
$C=\langle t\rangle/\sigma_t \quad C'=\langle t^2\rangle^3/(\langle t\rangle^3\langle t^3\rangle)$

Fractional extension averaged over many realizations

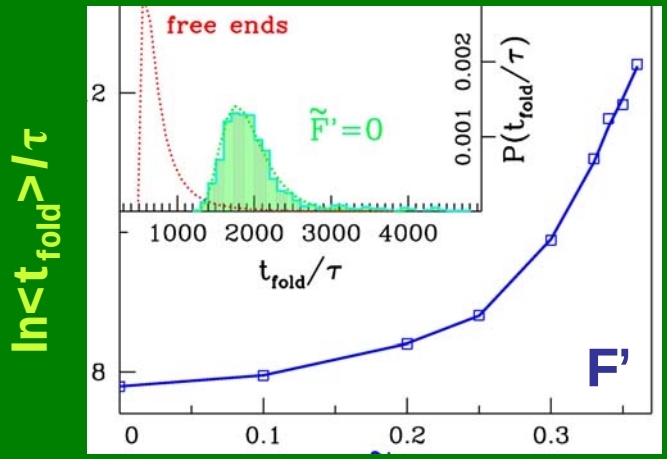




(two exponents more accurate)

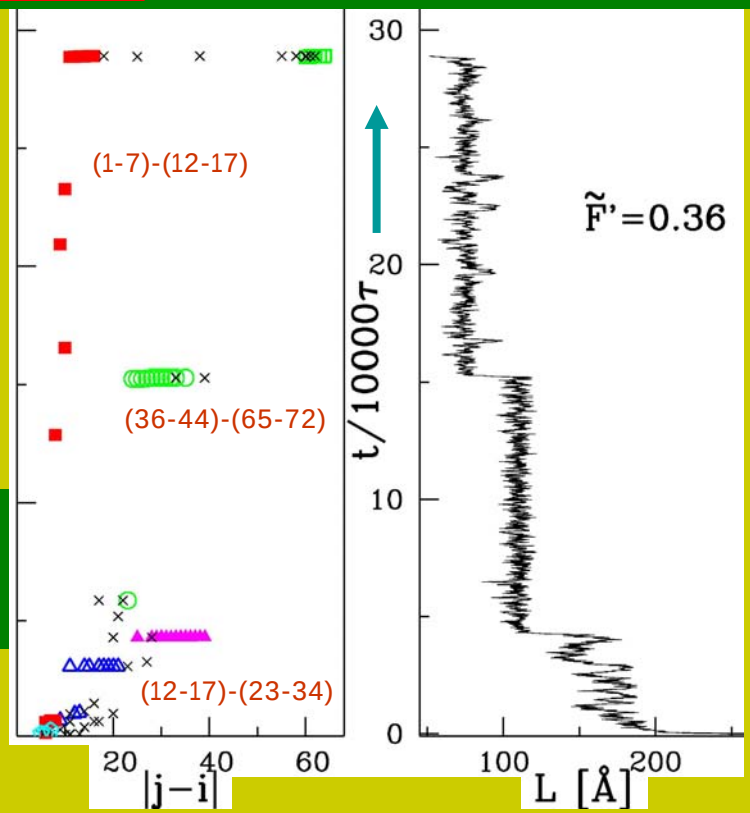


Log-normal distribution of folding times



FOLDING

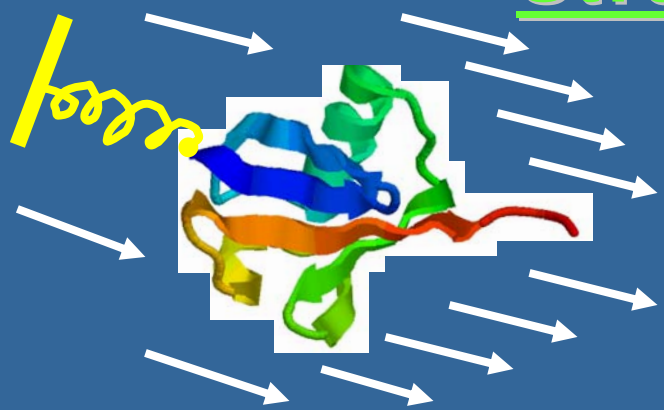
Force-clamp microscope indeed probes the folding process



Folding scenarios in and out of the force-clamp are distinct

Best, Hummer 2005 – studies of Q

Stretching in a uniform fluid flow



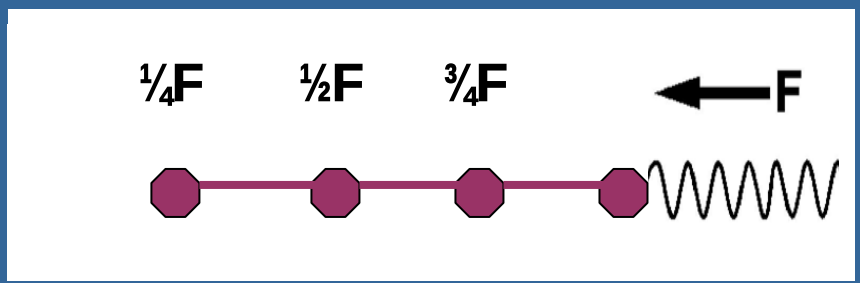
$$m\ddot{\mathbf{r}}_i = -\gamma(\dot{\mathbf{r}}_i - \mathbf{u}(\mathbf{r}_i)) + F_i^c + \Gamma$$

$$\mathbf{F} = \gamma \sum_i^N \mathbf{u}(\mathbf{r}_i)$$

Force at the fixed end

$$F = N\gamma u$$

Tension is non-uniform: increases towards the anchored end



Ubiquitin: many intermediates

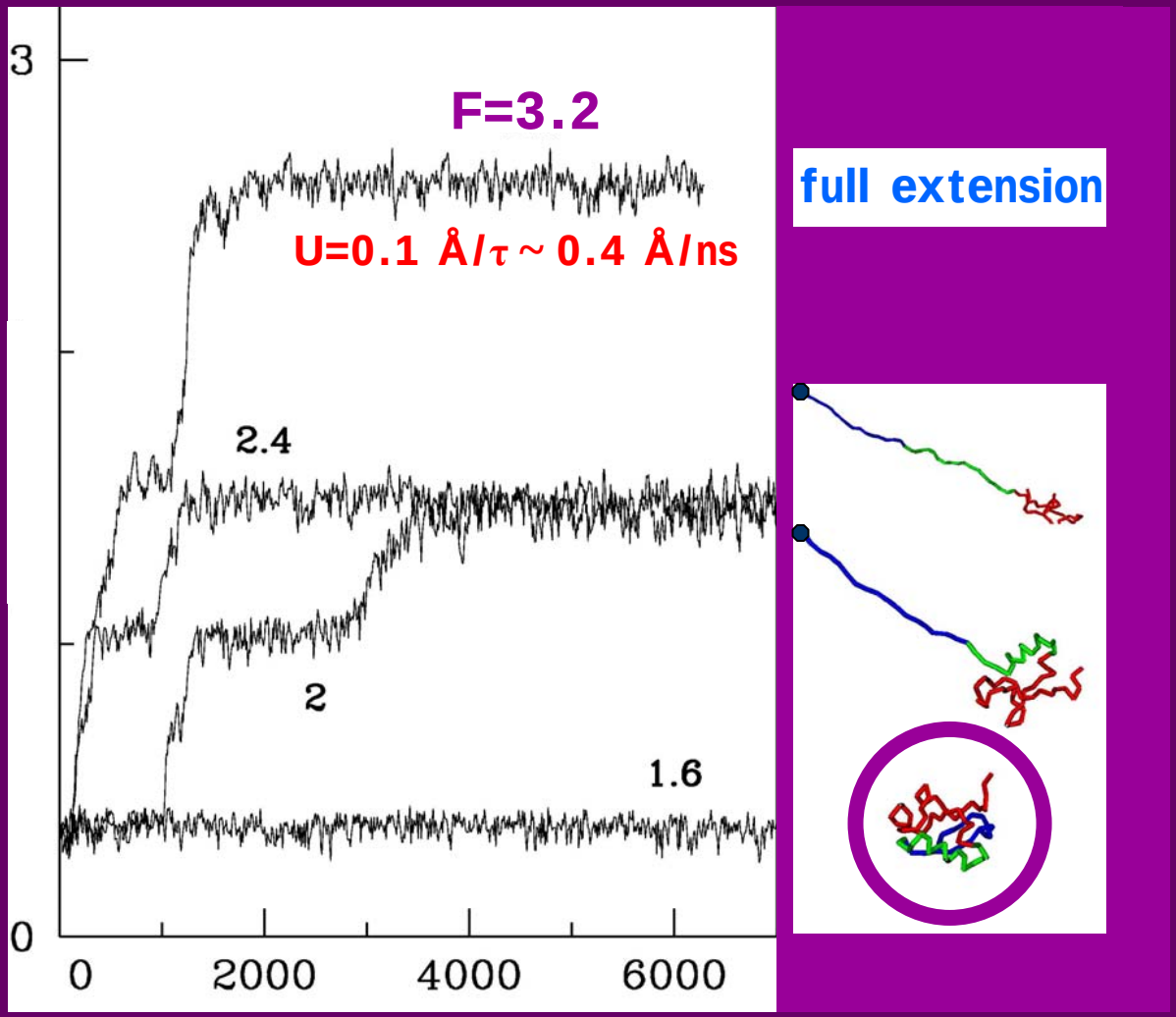
Unlike the force clamp case

Dependence on the choice of the anchored terminus

Can get more diagnostics of structure

Unwinding begins in the region closest to the anchor

Ubiquitin

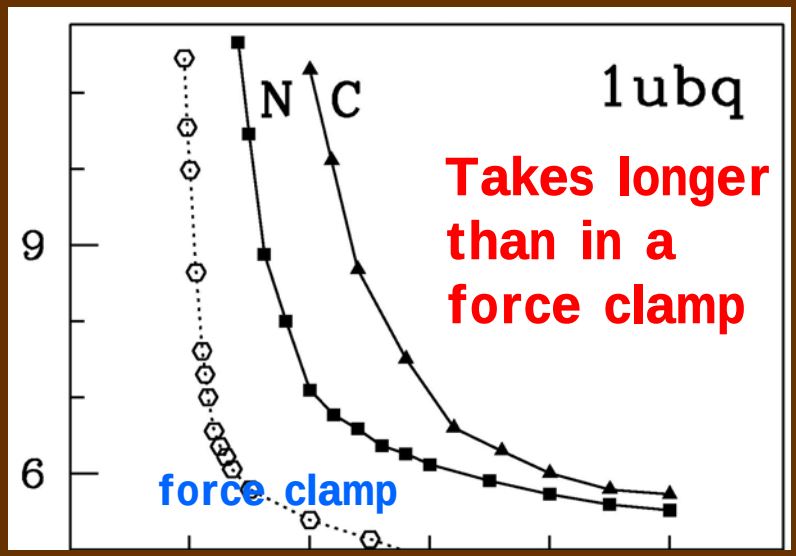


full extension

N terminus fixed

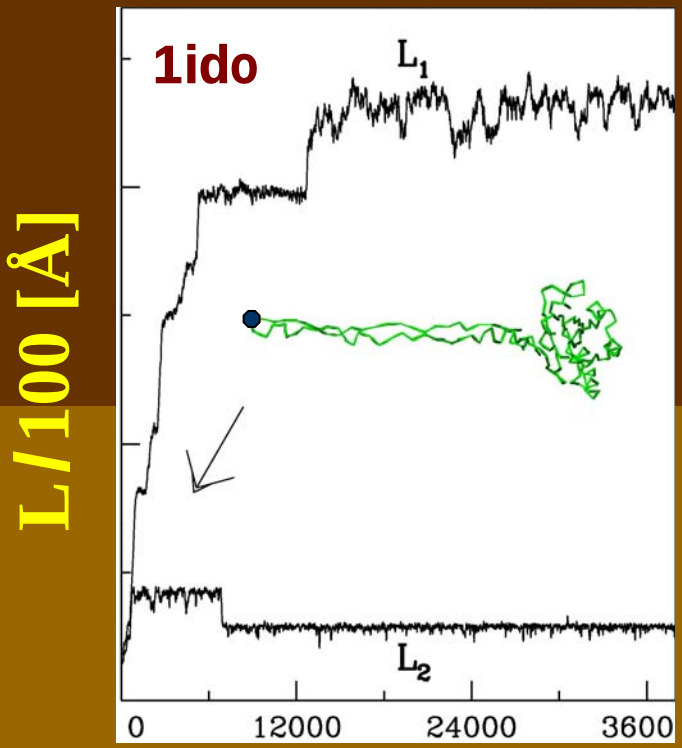
Other theoretical studies:
Lemak et al. 2003

Average time to get 90% of the full extension - depends on the terminal



Non-terminal attachment

F=4 K148 fixed



Polydomains - nonserial unwinding

Refolding after stopping the flow - misfolds

Time scales from the Peclet number

$Pe = UR_g/D$

flow vs. diffusion

U – characteristic velocity

R_g – radius of gyration

D – diffusion coefficient

Ubiquitin - simulation:

$D = 5\ Å^2/\tau$ $R_g = 11.5\ Å$ $U = 0.1\ Å/\tau$

Ubiquitin - experiment:

$D = 1.7 cm^2/s$

$Pe = 0.2$

same Pe →

$U = 4\ cm/s$

1000 faster than for DNA

→ τ corresponds to 0.25 ns

conclusions

Simple Go models can elucidate the microscopic picture of unwinding.

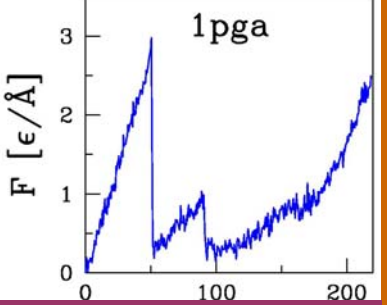
Scenarios represented on the time-contact order plane provide a detailed and useful description of large scale conformational changes.

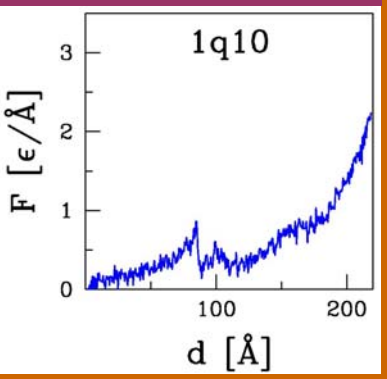
Unwinding of modular proteins need not be serial in nature – also controlled by T.

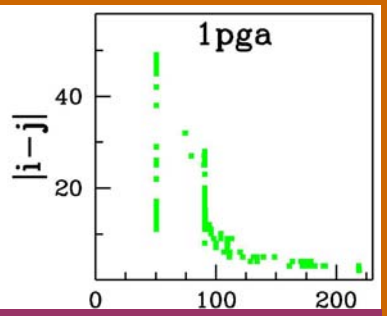
CONSTANT SPEED: survey of the PDB, determination of F_{max}, proposed list of strong proteins, correlations with the type of structure, identification of mechanical clamps.

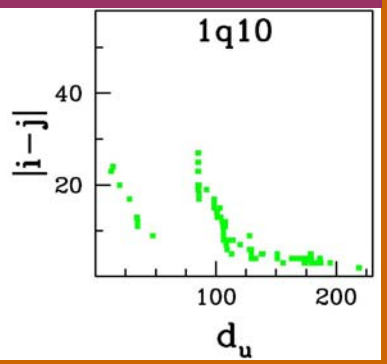
CONSTANT FORCE: exponential unfolding statistics below F_{max} and lognormal above it, refolding different than in the absence of the clamp.

UNIFORM FLOW: more intermediates than in force clamps, dependence on the choice of the anchor, may offer more diagnostic data than AFM.





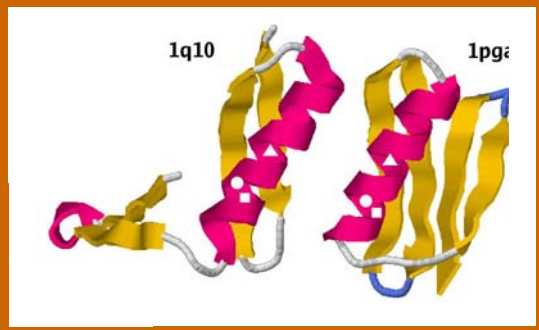




Proteins with the same CATH index may differ in resistance to pull – inadequacy of this classification scheme

In S134 - 10 proteins 3.10.20.10

α/β, roll, ubiquitin-like, immunoglobulin binding



Other in S3813

Two dynamical sets: weak and strong

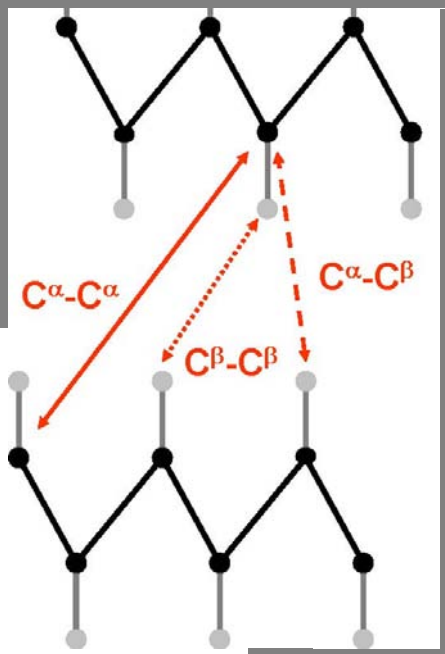
30F→V 33Y→F 34A→F

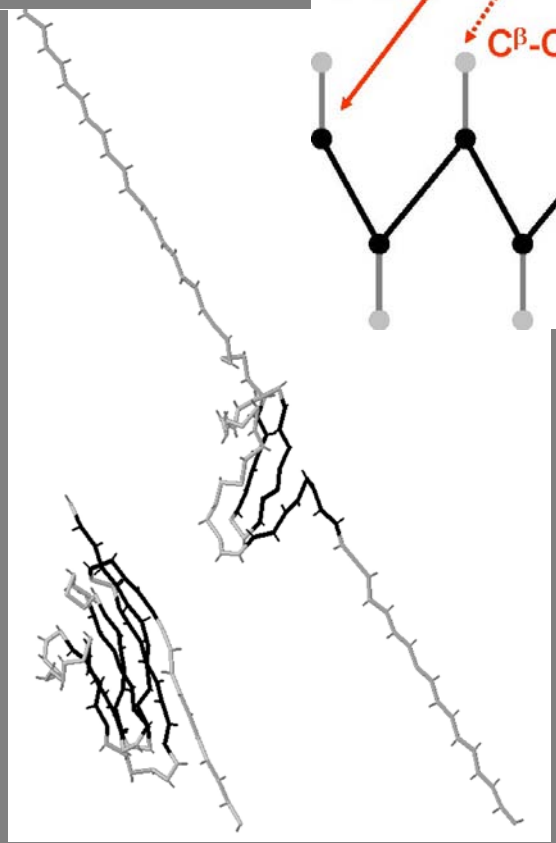
Differ in RMSD by 1.9Å

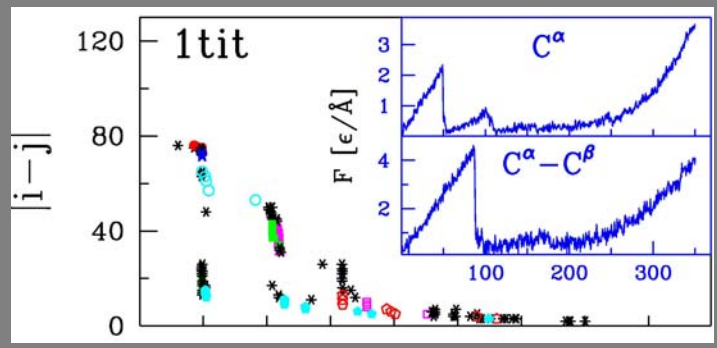
Crucial long-range contacts missing

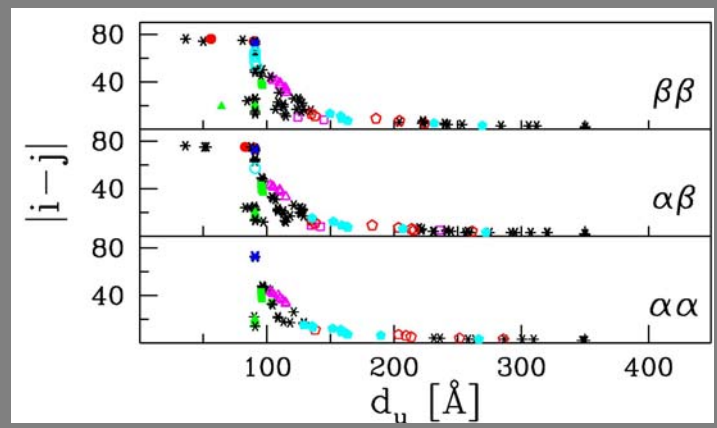
A model with the side groups,

as represented by the C^{β} atoms









A strength-modulated contact map

Frequent elimination of the secondary force peaks

Certain reshuffling of the ranking