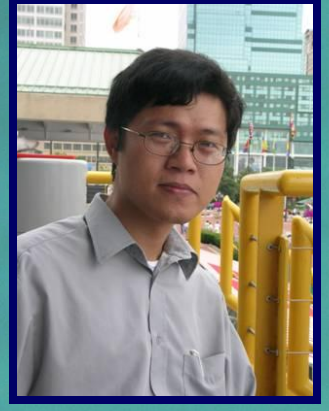
Marek Cieplak
Institute of Physics, Warsaw, Poland

Mechanical stability of proteins and virus capsids:

Coarse-grained structure based models

Developed in response to limitations of all-atom models, more qualitative

Since 1999



Trinh Xuan Hoang
Institute of Physics,
Hanoi, Vietnam

Piotr Szymczak,
Institute of Theoretical
Physics, Warsaw
Univeristy, Poland

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

Michal
Wojciechowski,

Szymon
Niewieczerzał

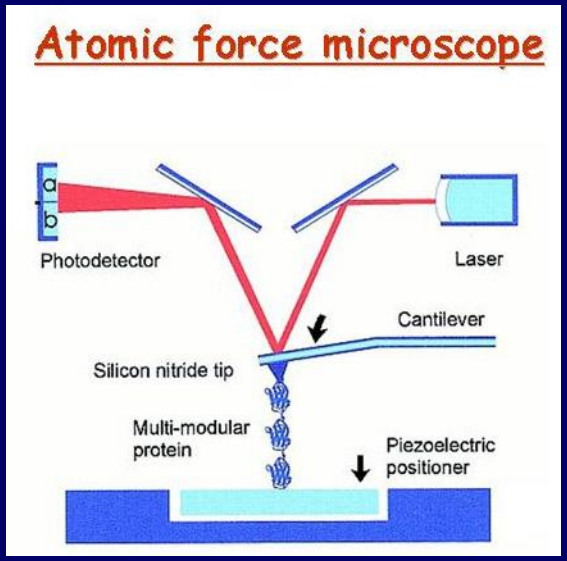
Mateusz
Sikora

Joanna I.
Sułkowska

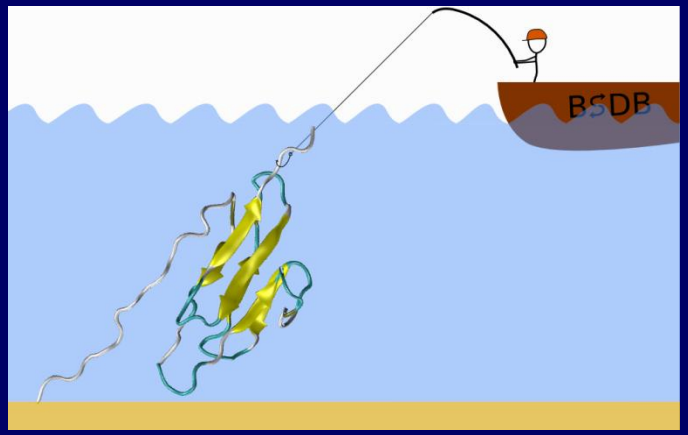
Piotr Sułkowski,
Insititute of Nuclear
Physics, Warsaw, Poland

S. Filipek (Warsaw), K. Krzyśko (Warsaw), H. Janovjak (Berkeley), P. Marszałek (Duke), A. Pastore (London), M. Carrion-Vazquez (Madrid)

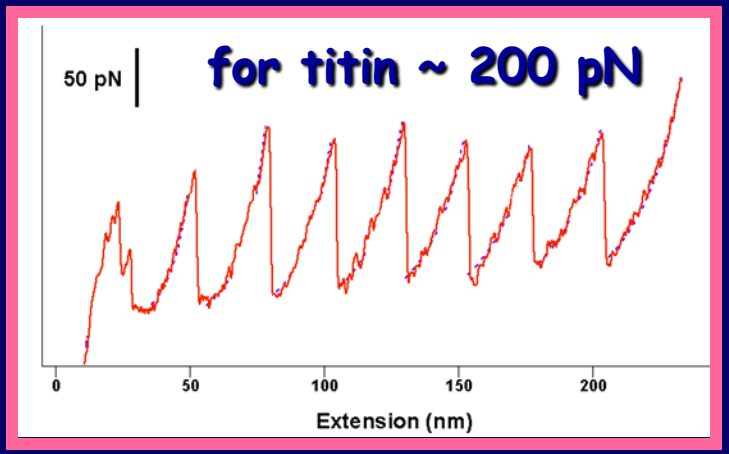
STRETCHING OF SINGLE MOLECULES



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



Characteristic scale of the force: F_{max}



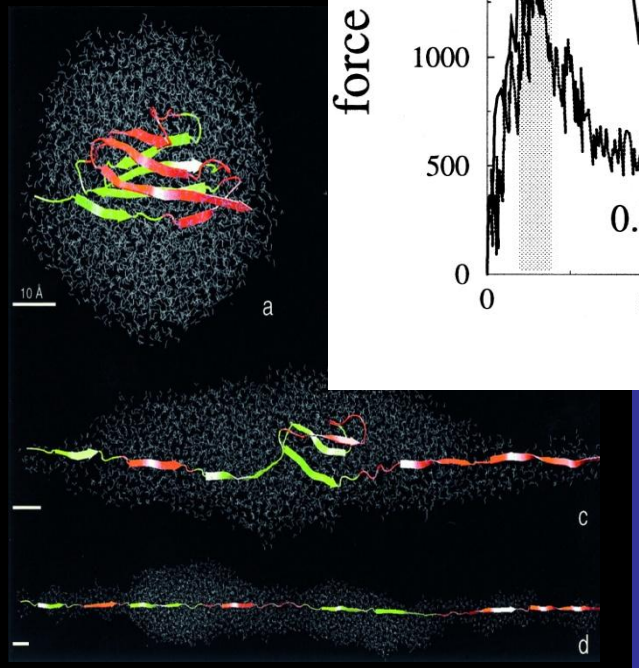
Stretching of bridge pylons

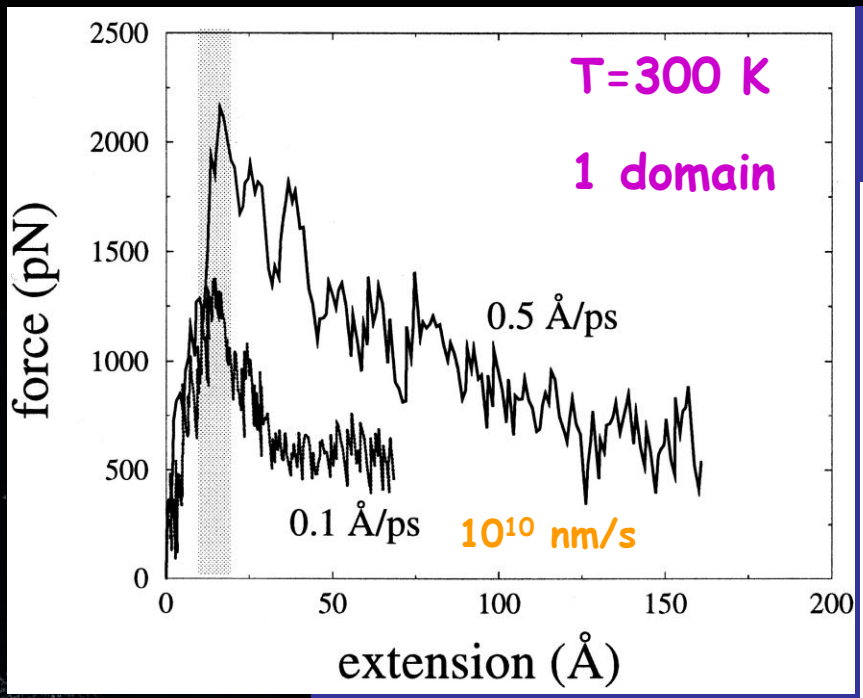
No rupture

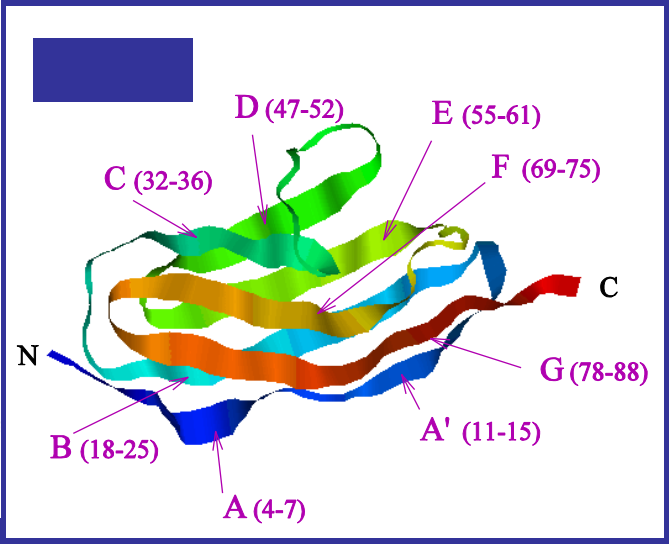
ALL-ATOM SIMULATIONS: titin

Lu Schulten 2000 (Paci Karplus 2000)

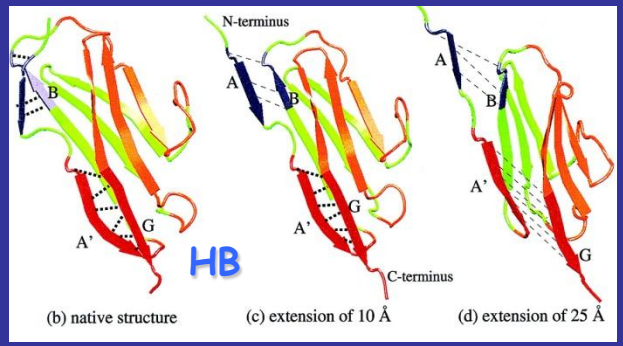
typically 10-100 ns time scales







Pabon, Amzel 2006 - quasistatic ~500 pN



Difficult:

processes involving large conformational changes

comparatory studies of many proteins

STRUCTURE-BASED MODELS OF PROTEINS

TAKE HOMOPOLYMERS –
chains of tethered beads:

C^{α}

ADD ATTRACTION
BETWEEN SOME BEADS to
shape the homopolymers into
the backbones of proteins

Idea: N. Go & H. Abe 1981

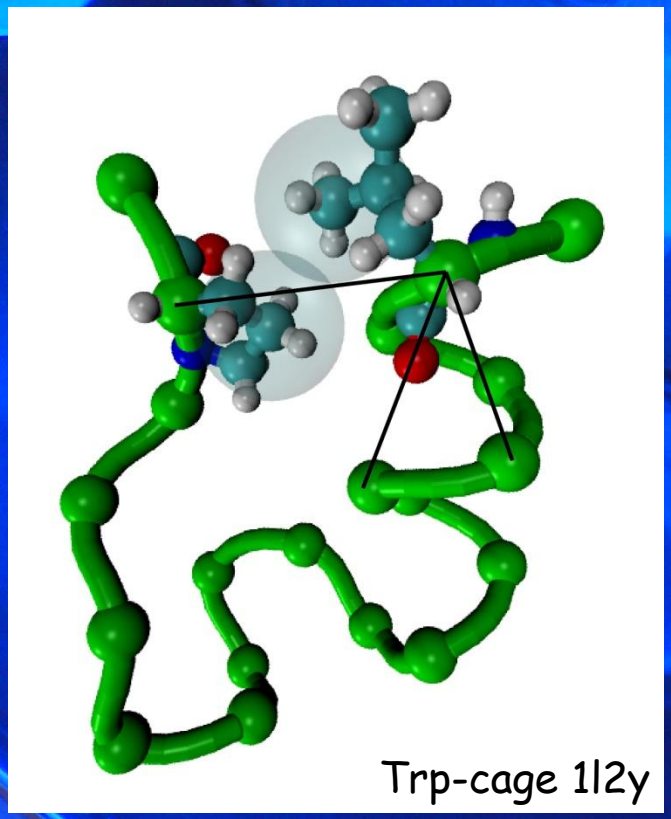
First MD implementation in a "minimalist model": J. Honeycutt & D. Thirumalai 1992

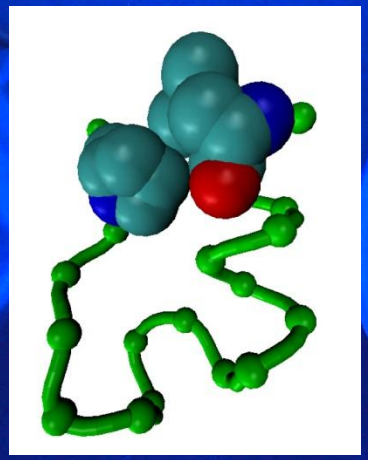
NATIVE CONTACTS

Defined by the conformation of the native state

Tsai Taylor Chothia Gerstein 1999

Overlap of the van der Waals spheres of the heavy atoms





GEOMETRY-BASED MODEL

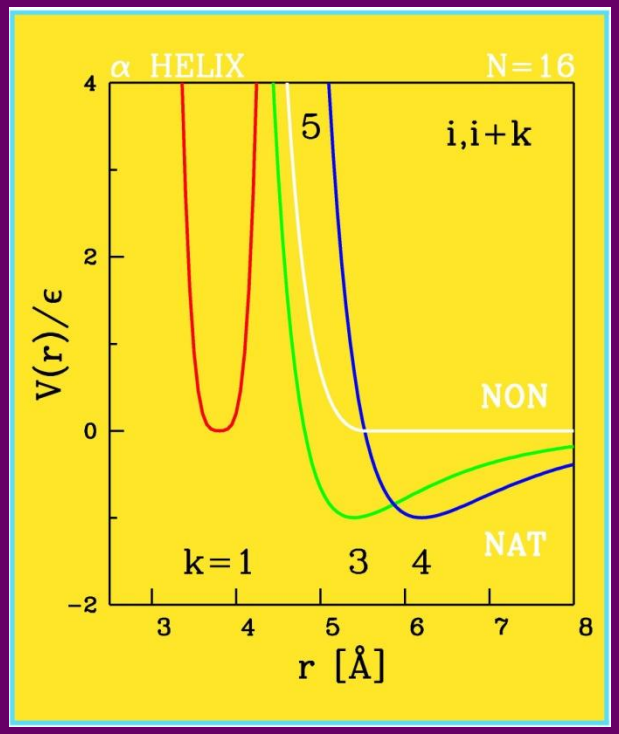
$$E_p(\{\mathbf{r}_i\}) = V^{BB} + V^{\mathrm{NAT}} + V^{\mathrm{NON}} + V^{S}$$

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 C^{α}_i-C^{α}_j native distance

Non-native: repulsive with σ=4Å

Disulfide bonds like peptide bonds

V^{S}: angular terms locally favoring the native shape of the backbone: local stiffness

MOLECULAR DYNAMICS

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

$$F_c = -\nabla_r E_p$$

Brownian Dynamics if the inertia term is neglected

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

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

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

$m = 118 m_p$

Langevin noise as a thermostat and as an emulator of water

Large friction
Finite bead size

modify the effective time scale τ

$\tau \sim 1$ ns

diffusional time to move by ~ 5 Å

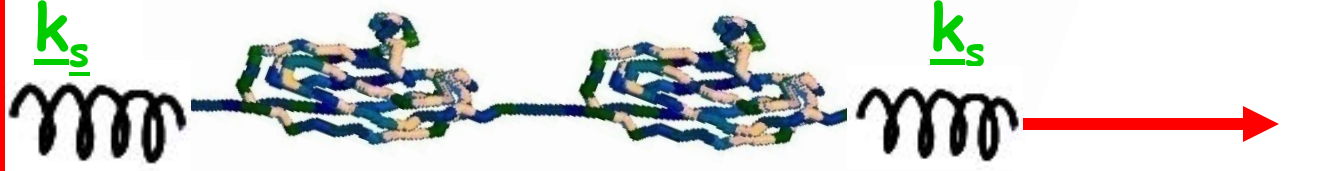
Veitshans, Klimov, Thirumalai 1997: τ~3ns

when comparing with experiments

Use $\gamma = 2m/\tau$

v_p=0.005 Å/τ ~ 10^6 nm/s

experiment < 10^4 nm/s

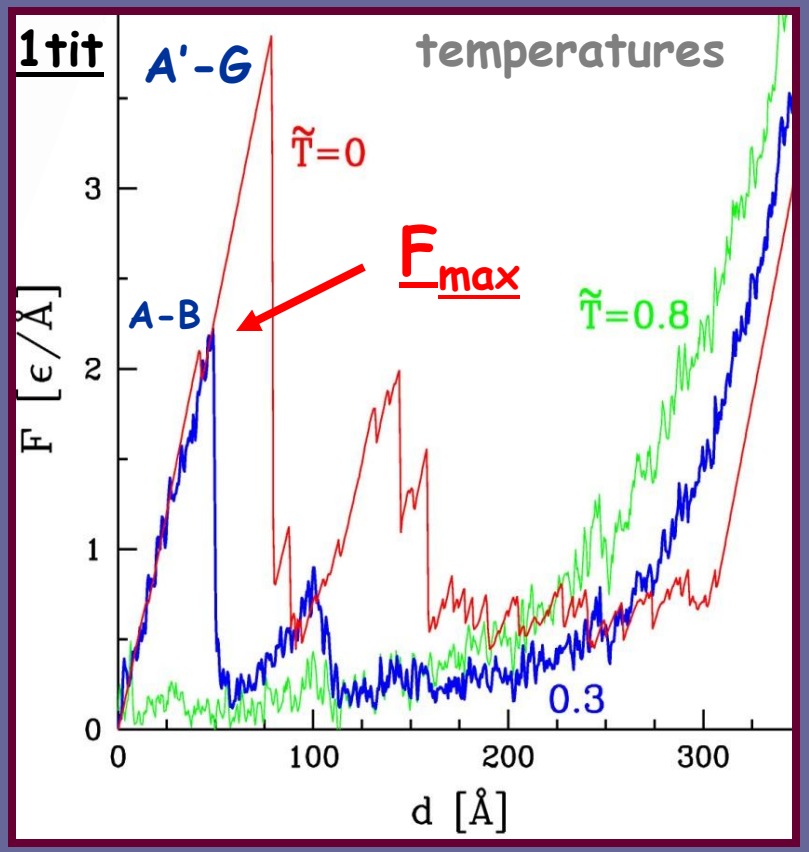


soft: k_s=0.12 ε/Å^2 ~0.08 N/m

stiff: k_s=30 ε/Å^2

logarithmic shifts with v_p

one domain



F_{max} depends on T

OTHER GO-LIKE MODELS

model = { V^{NAT}, S, M, E, C^{α}/C^{β} }

504 variants enumerated, 62 studied

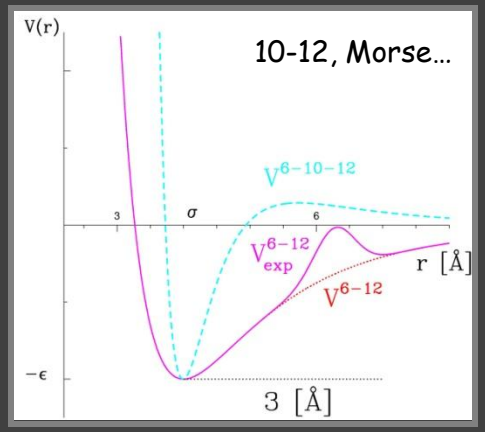
native contact potential
6-12, 10-12, Morse …

local backbone stiffness
Chirality or Angular (bond & dihedral)

contact map
cutoff, vdW, CSU

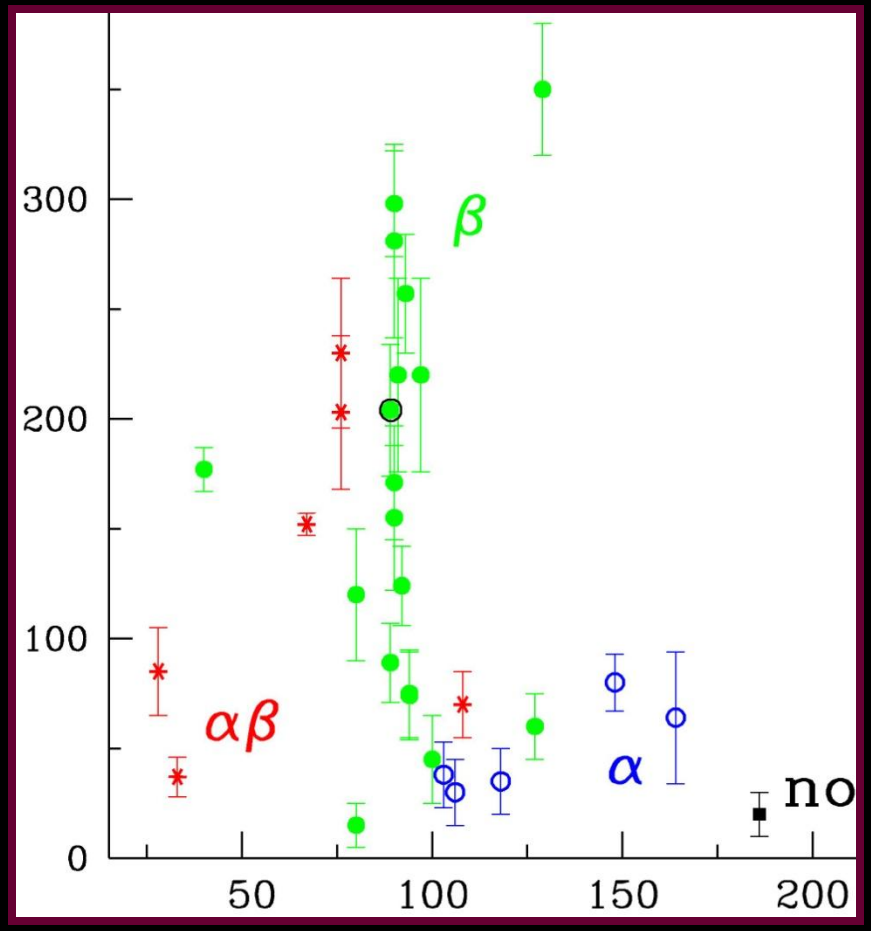
energy scale
uniform or not, hydrogen bonds different …

if side groups represented by C^{β}



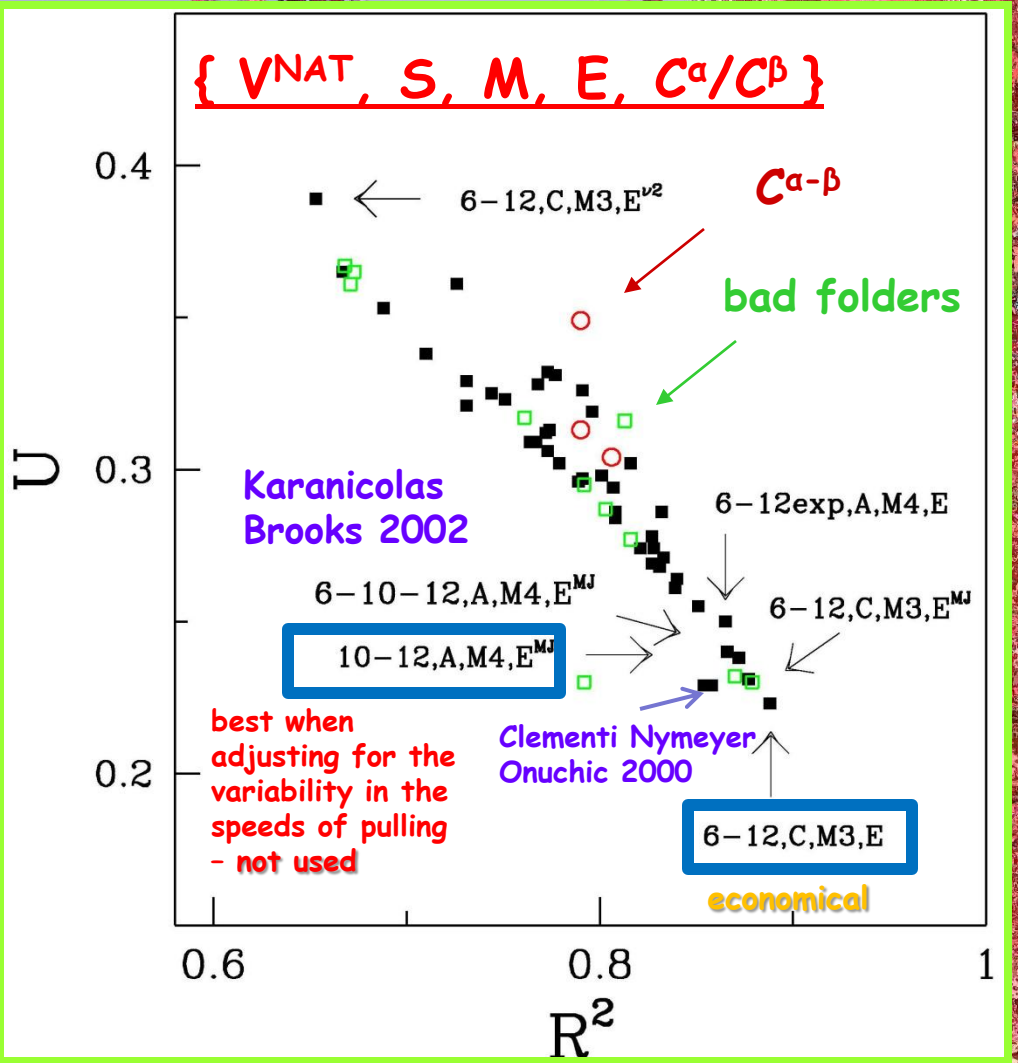
WHICH MODEL IS OPTIMAL?

USE: Experimental results on stretching at constant speed



Unlike folding – stretching starts around the native conformation: good for testing Go-like models

All-atom simulations on ~ 22 proteins



D=28 proteins

$$R^2 = 1 - \sqrt{\frac{1}{D}\sum_{\lambda=1}^{D}\left(\frac{F_\lambda^t - F_\lambda^e}{F_\lambda^e}\right)^2}$$

t – theoretical e – experimental

$$W_{\lambda+1} = \frac{F_{\lambda+1}^t - F_\lambda^e}{F_\lambda^e} \qquad w_{\lambda+1} = \frac{F_{\lambda+1}^e - F_\lambda^e}{F_\lambda^e}$$

predicted actual

$$U = \sqrt{\frac{\sum_{\lambda=1}^{D-1}(W_{\lambda+1} - w_{\lambda+1})^2}{\sum_{\lambda=1}^{D-1}(w_{\lambda+1})^2}}$$

Theil

Karanicolas Brooks: ε in hydrogen bonds, Miyazawa-Jernigan-like modulation in other contacts

Like the Autumn leaves, they're all pretty, but this one is my favorite

Validation of the Go model for stretching

Model 6-12,C,M3,E uniform ε

Constant speed pulling

v=0.005 Å/τ ~ 500 000 nm/s

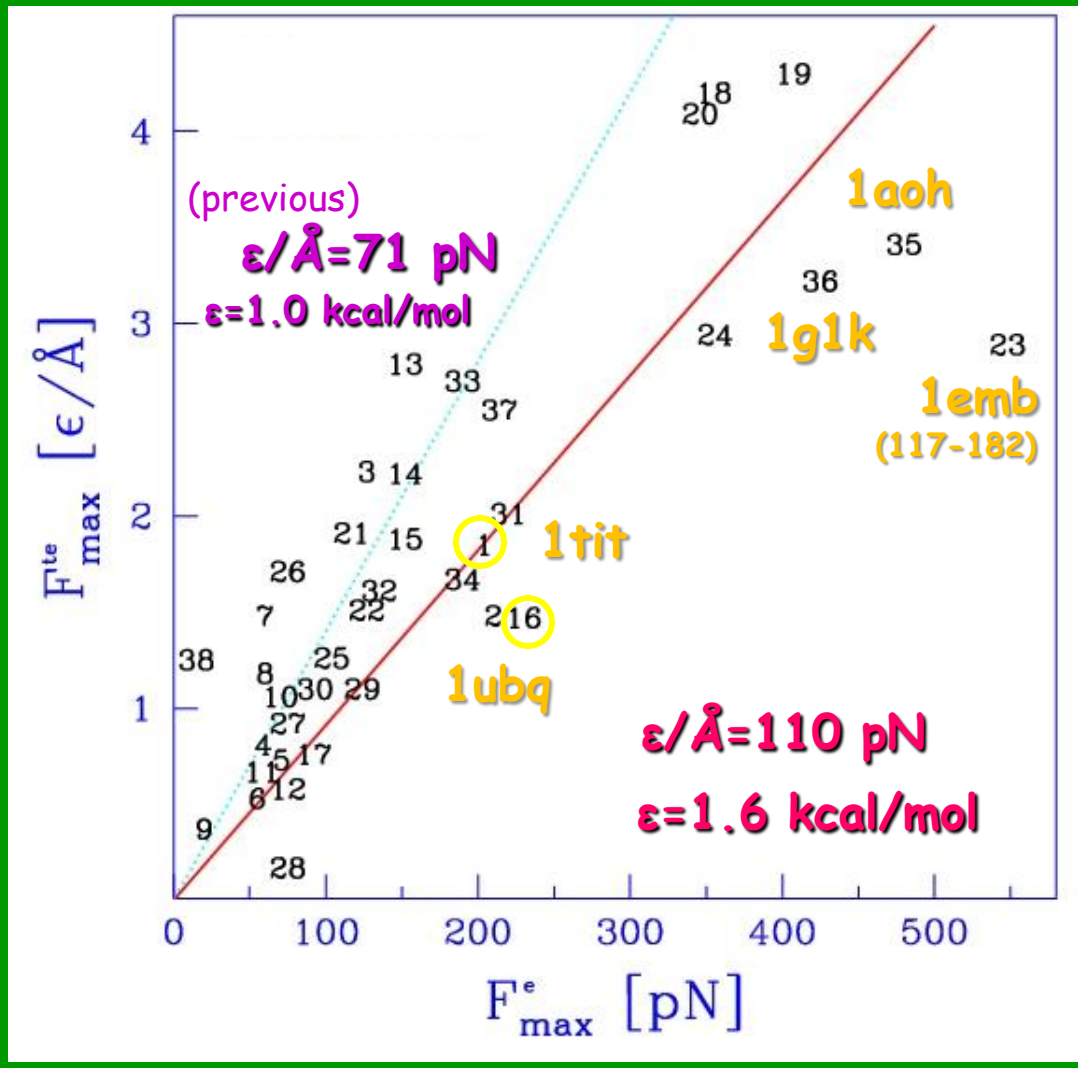
R^2= 0.89

extrapolation to experimental speeds & more points

R^2=0.83 U=0.28

@ $T=0.3\varepsilon/k_B$

linkage dependence



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

Protein Data Bank: 29385 structures on July 26 2005

54807 - on December 18 2008

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

J. Fernandez

What proteins are strong and why?

How does F_{max} correlate with structure?

CATH-based structure classification

Class (4)
Architecture
Topology
Homology

SCOP-based classification

Class (10)
Fold
Superfamily
Family
Protein

Hierarchy

Three theoretical surveys of proteins, within the uniform ε Go-like model, – stretching at constant speed

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

Model 6-12,C,M2,E

239 with 150 < N ≤ 851

7 749

J. Phys.: Cond. Matt. 2007

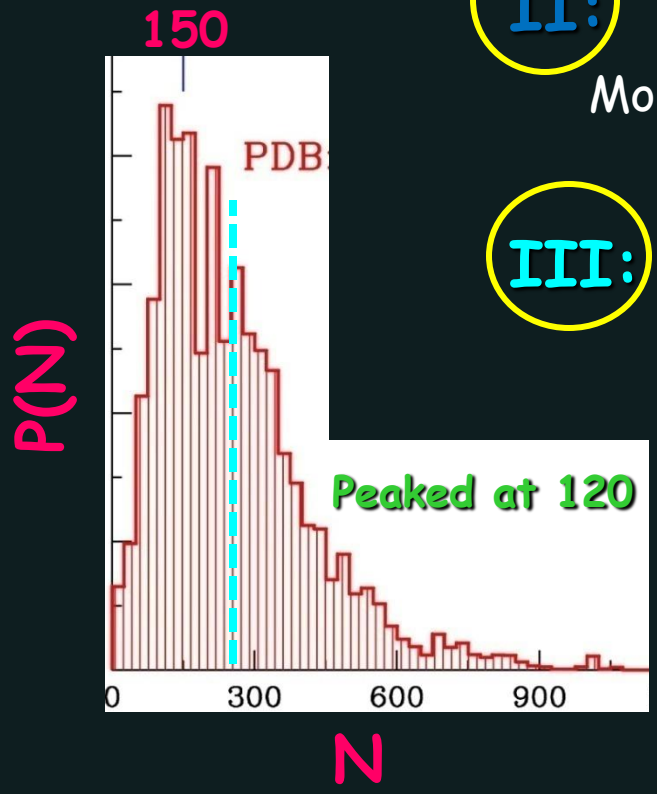
II: 7 510 proteins with 40 ≤ N ≤150

Model 6-12,C,M3,E (the better variant) Biophys. J. 2008

III: 17 134 proteins with 40 ≤ N ≤250

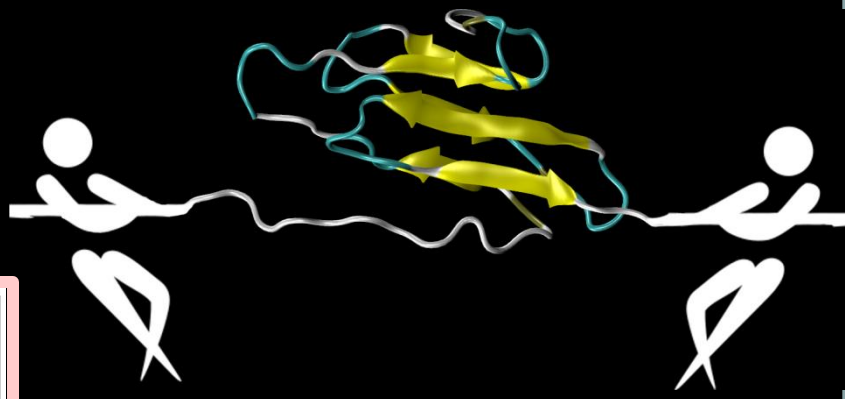
Model 6-12,C,M3,E

PLoS Computational Biology 2009

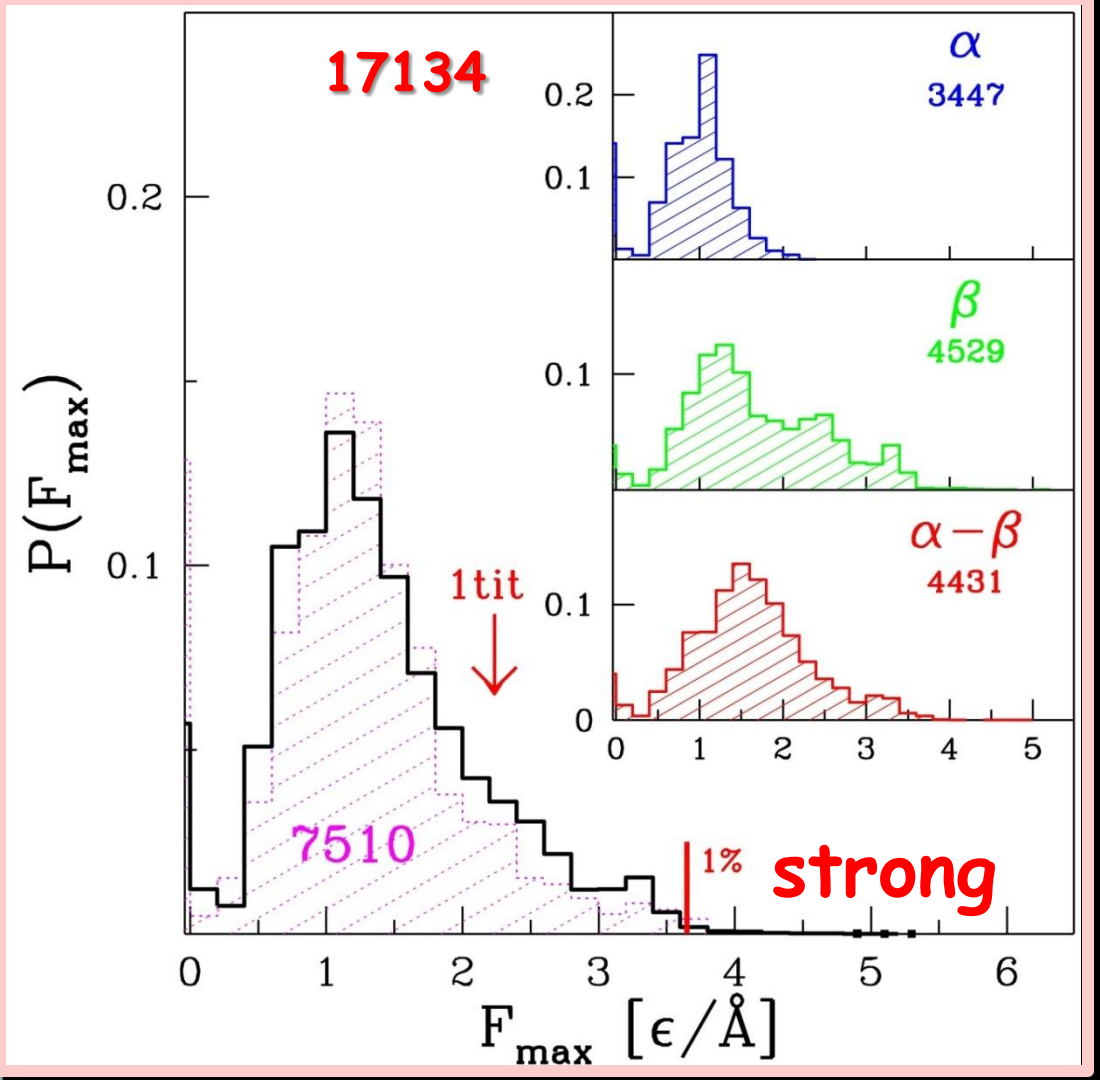


with J. Sułkowska and M. Sikora

Probability distribution of F_{max}



≤ 250 CLASS



no strong α proteins

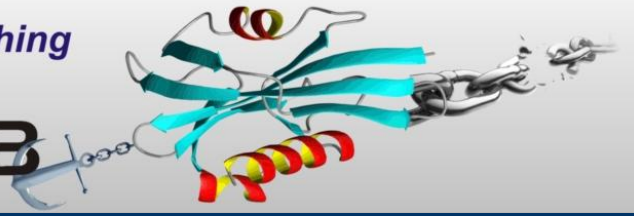
Pulling by the termini

All 17 134 in

THE PREDICTED LIST

Bio-molecule Stretching Database

BSDB



info.ifpan.edu.pl/BSDB/

n	PDBid	N	F_{max} [ε/Å]	L_{max} [Å]	λ	CATH	SCOP
1	1bmp	104	10.2	23.2	0.01	2.10.90.10	g.17.1.2
2	1qty	95	8.9	72.1	0.11	2.10.90.10	b.1.1.4
3	2bhk	119	7.3	26.5	0.67		
4	1lxi	104	7.3	22.5	0.01		g.17.1.2
5	1cz8	107	6.4	76.5	0.13	2.10.90.10	b.1.1.1
6	2gh0	219	5.8	25.9	0.06		
7	1wq9	100	5.5	72.0	0.10	2.10.90.10	g.17.1.1
8	1flt	107	5.5	75.6	0.12	2.10.90.10	b.1.1.4
9	1fzv	117	5.4	90.4	0.12	2.10.90.10	g.17.1.1
10	2gyz	100	5.4	14.4	0.01		
11	1rew	103	5.3	21.7	0.01	2.10.90.10	g.7.1.3
12	1m4u	139	5.3	52.1	0.07	2.10.90.10	g.17.1.2
13	1vpf	94	5.3	68.1	0.11	2.10.90.10	g.17.1.1
14	1c4p	137	5.1	106.0	0.12	3.10.20.180	d.15.5.1
15	1qqr	138	5.0	110.3	0.12	3.10.20.180	d.15.5.1
16	3bmp	114	5.0	33.0	0.03	2.10.90.10	g.17.1.2
17	1j8s	193	4.9	77.9	0.03	2.60.40.1370	b.2.3.3
18	1wq8	96	4.9	82.6	0.11	2.10.90.10	g.17.1.1
19	1j8r	193	4.8	77.7	0.03	2.60.40.1370	b.2.3.3
20	1f3y	165	4.8	284.7	0.43	3.90.79.10	d.113.1.1
36	1aoh	147	4.3	77.1	0.01	2.60.40.680	b.2.2.2
3144	1ubq	76	2.2	47.9	0.04	3.10.20.90	d.15.1.1
3580	1tit	89	2.1	55.3	0.04	2.60.40.10	b.1.1.4

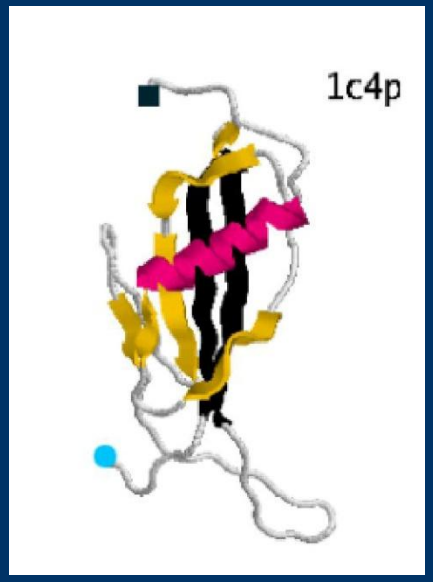
Strongest found in the previous survey

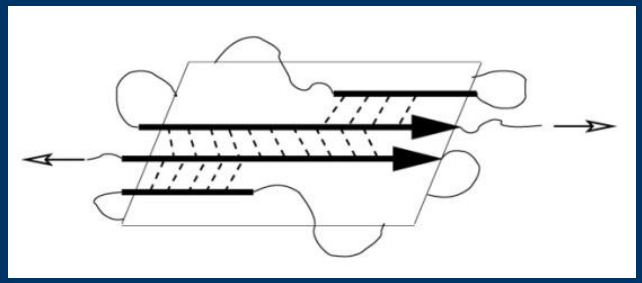
470 pN

λ - relative peak position in terms of the end-to-end distance L

The predominant source of strength in short proteins

shearing of hydrogen-bonded parallel β-strands





1c4p (14) & 1qqr (15): β domain of streptokinase

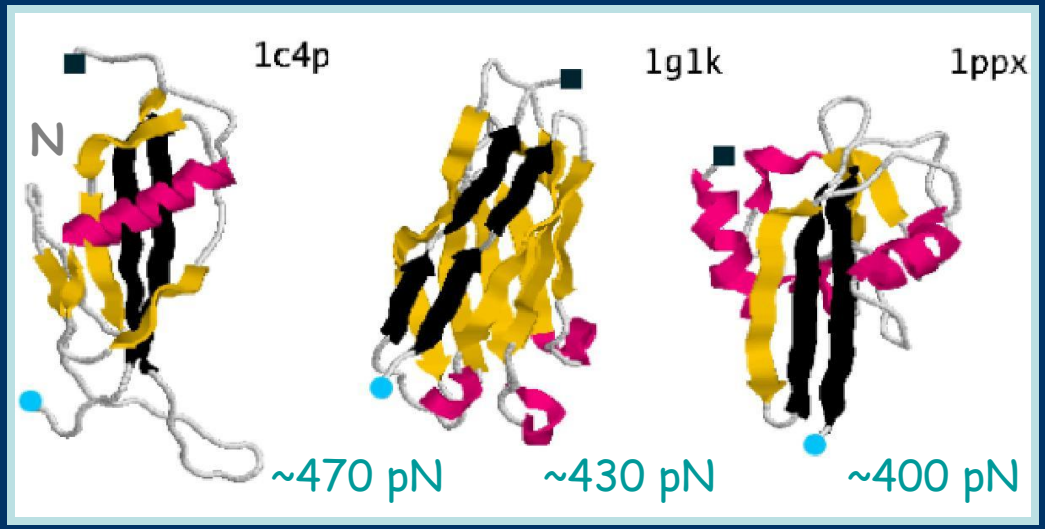
(blood clotting - different functions)

Scaffoldins: structural proteins of the cellulosome (degradation of cellulose):

		measured	predicted
(36)	1aoh (c7A) from Clostridium thermocellum	480 pN	470
(78)	1g1k (c1C) from Clostridium celluloticum	425 pN	350

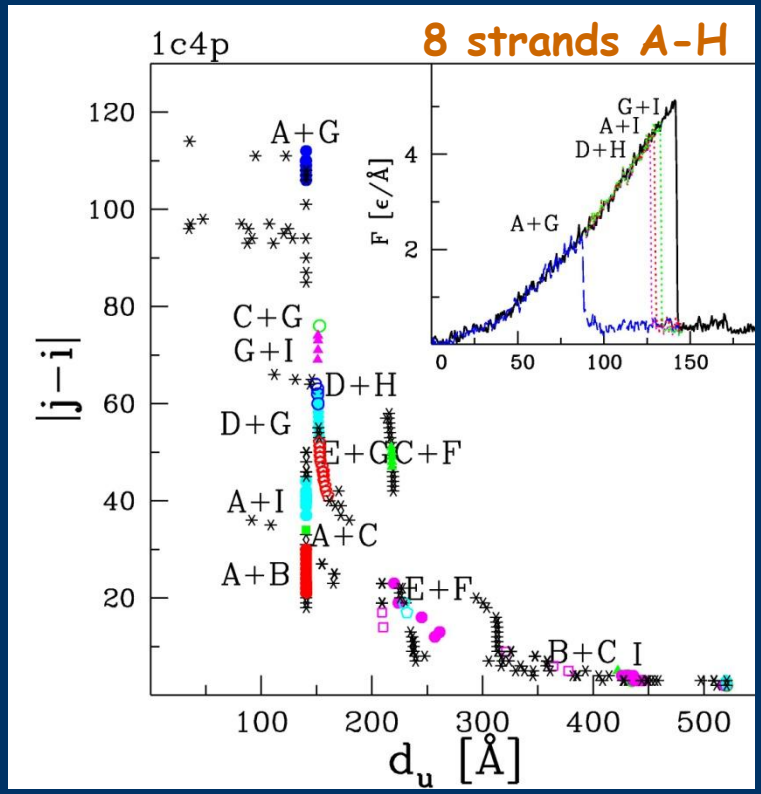
Valbuena … Carrion-Vazquez, 2009

shearing of hydrogen-bonded parallel β-strands



the relevant β-strands shown in black

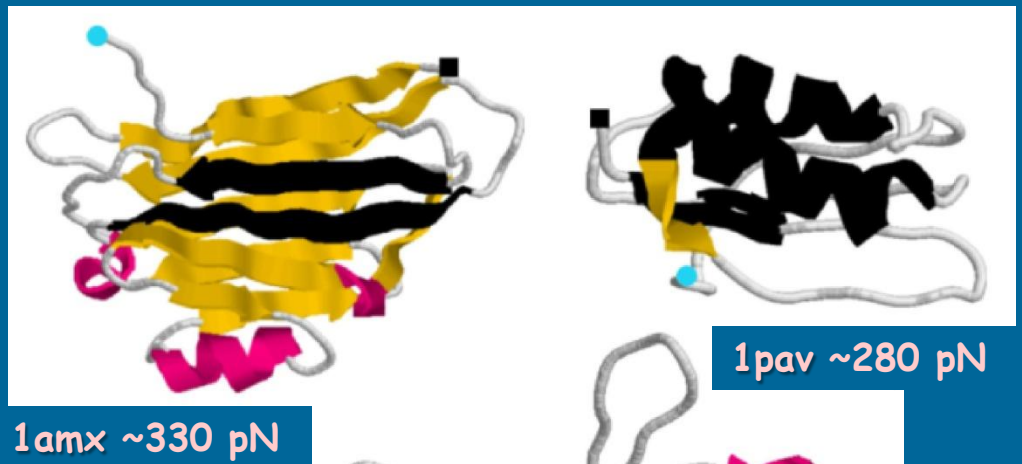
Unfolding scenario diagram – more detailed than Q – fraction of the total number of native contacts



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

Other kinds of mechanical clamps

Antiparallel β-strands



1amx ~330 pN

1pav ~280 pN

1qp1 ~300 pN

Unstructured clamps

A Box structure:

two antiparallel strands

and two antiparallel helices

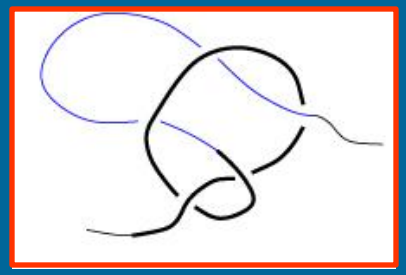
Delocalized clamps

CYSTEINE SLIPKNOTS

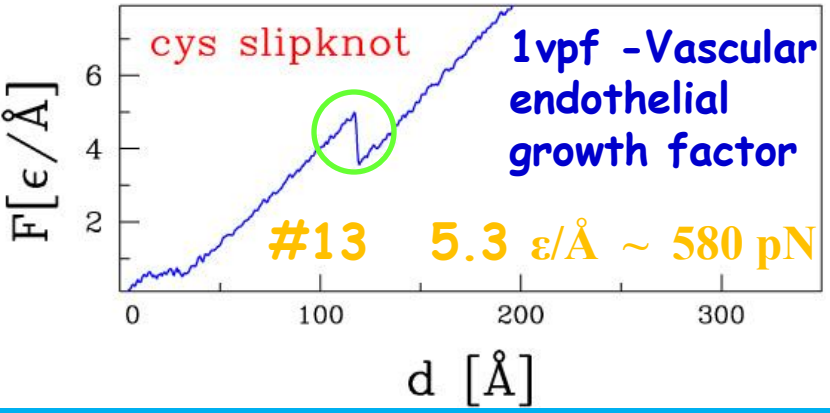
disulphide bridges
unusual entanglements

1vpf: vascular endothelial growth factor

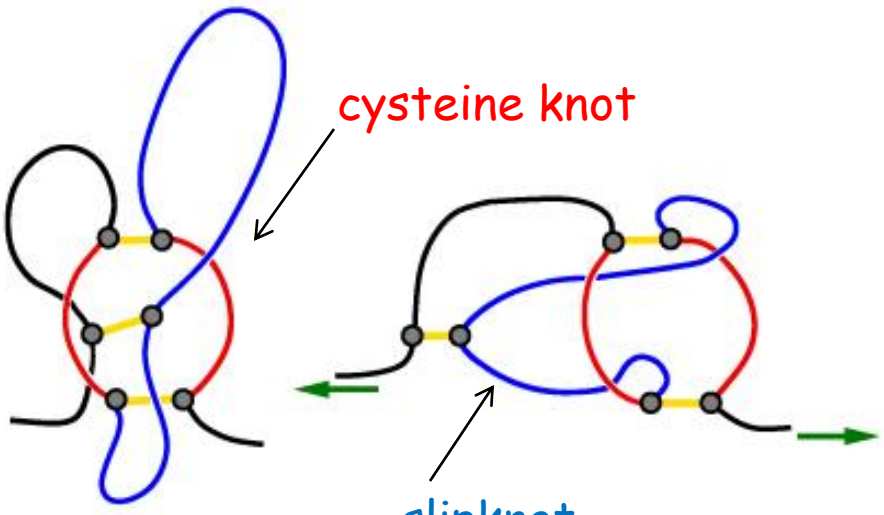
Cellular component: extracellular region



native slipknot



Slipknot resulting from pulling



1100 pN in 1bmp (#1)

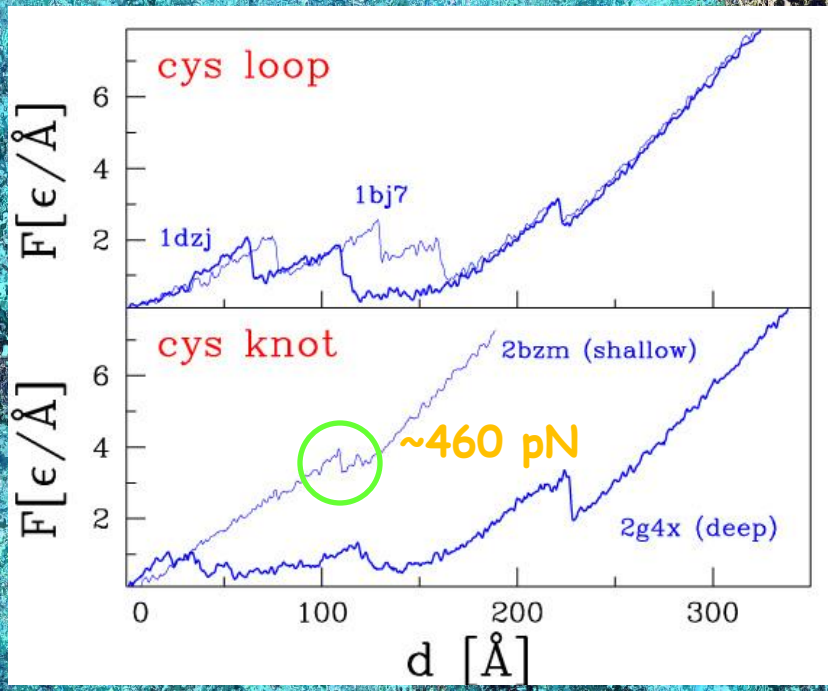
Bone morphogenetic protein-7

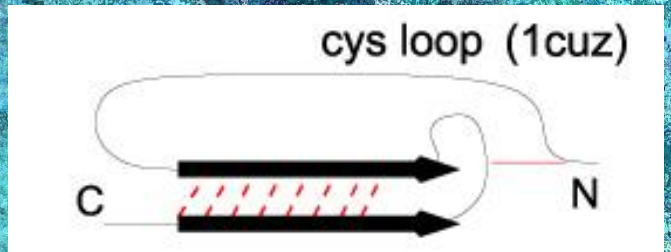
Function: growth factor activity

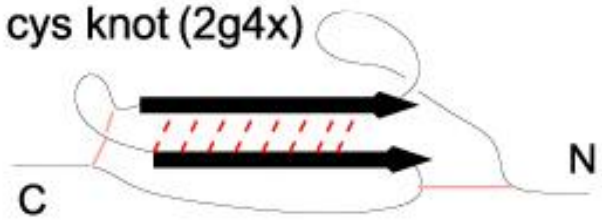
Cellular component: protein binding

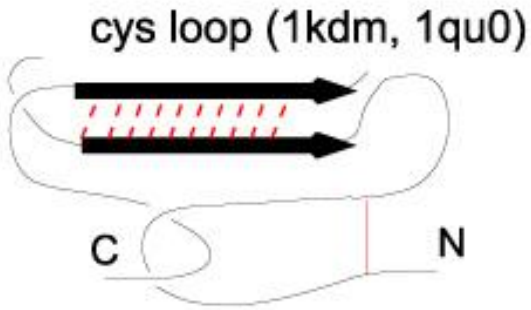
From Genetic Ontology

Other force clamps involving disulphide bridges



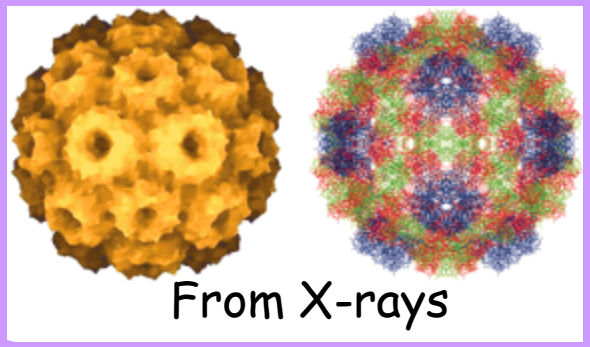




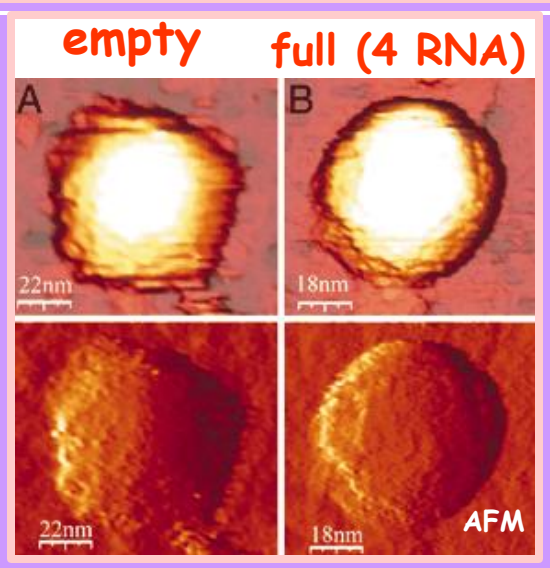


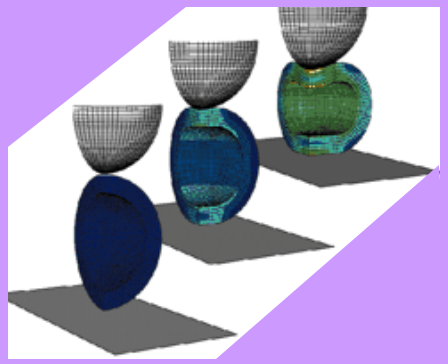
NANOINDENTATION OF VIRUS CAPSIDS

Self-assembled nanostructures consisting of a protein shell to protect the genetic material inside

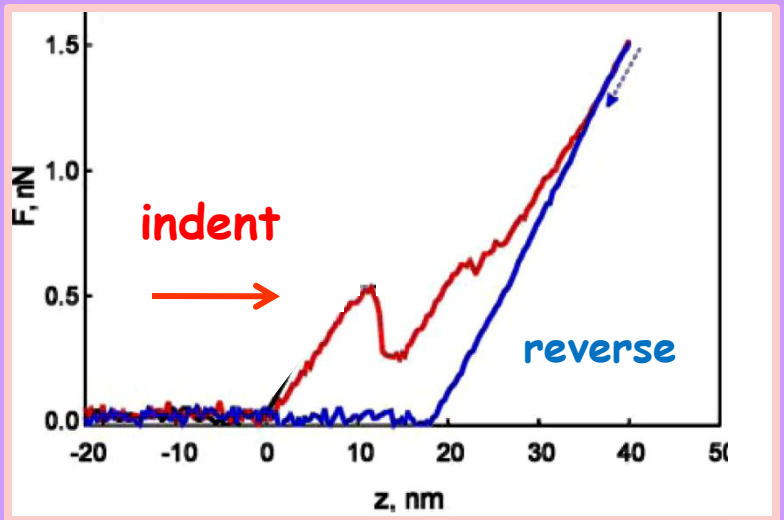


CCMV – cowpea (black-eyed pea) chlorotic mottle virus



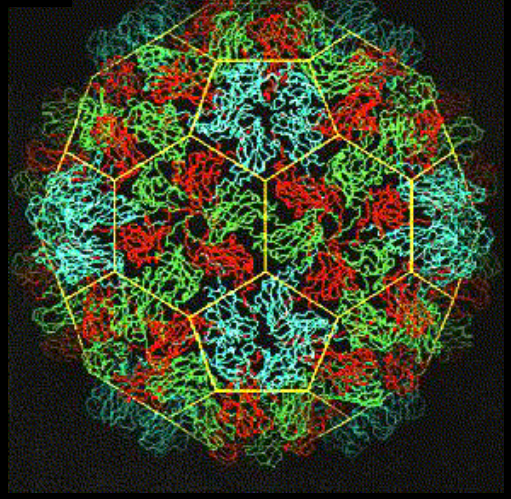


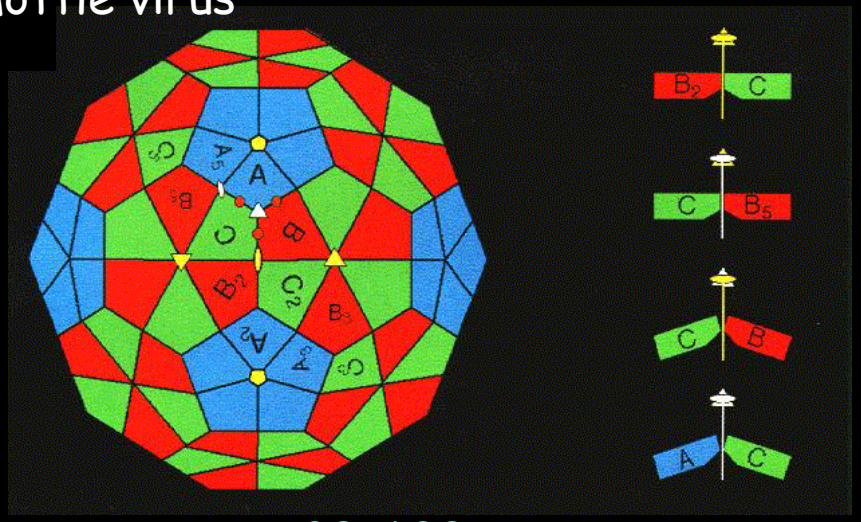
Michel, Ivanovska, Gibbons, Klug, Knobler, Wuite, Schmidt; Bruinsma - 2006



with M. O. Robbins

CCMV cowpea chlorotic mottle virus

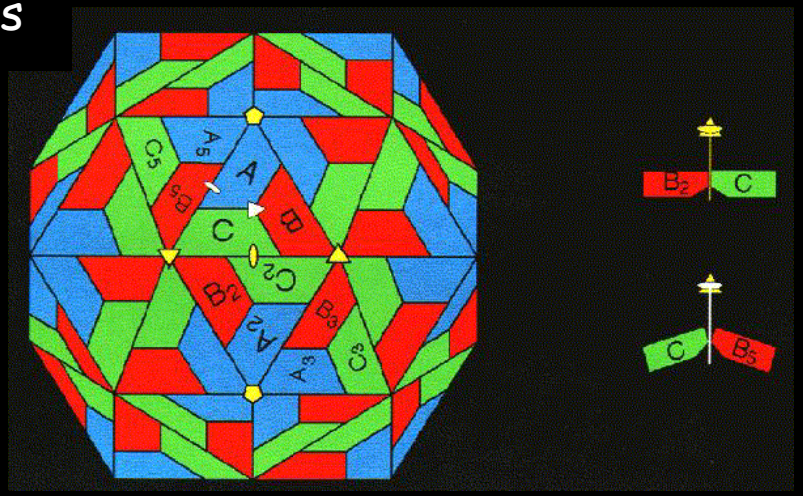




a truncated icosahedron model 20

28 620 C^{α} atoms, 62460 contacts

CPMV cowpea mosaic virus

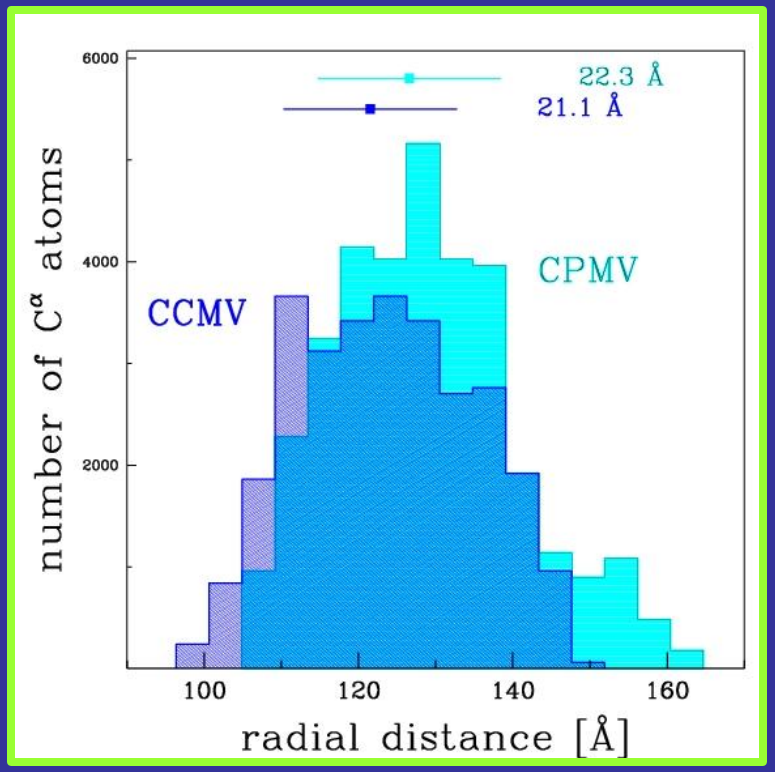


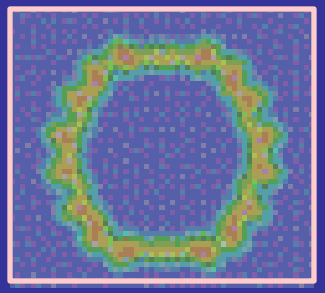
a rhombic triacontahedron model 30

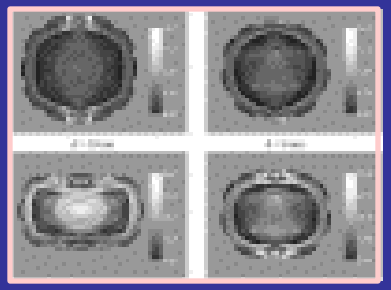
Both ~ 300 000 heavy atoms

33 480 C^{α} atoms, 90420 contacts

180 sequentially identical chains that self assemble







Model of a nonuniform elastic shell: Gibbons & Klug 2008

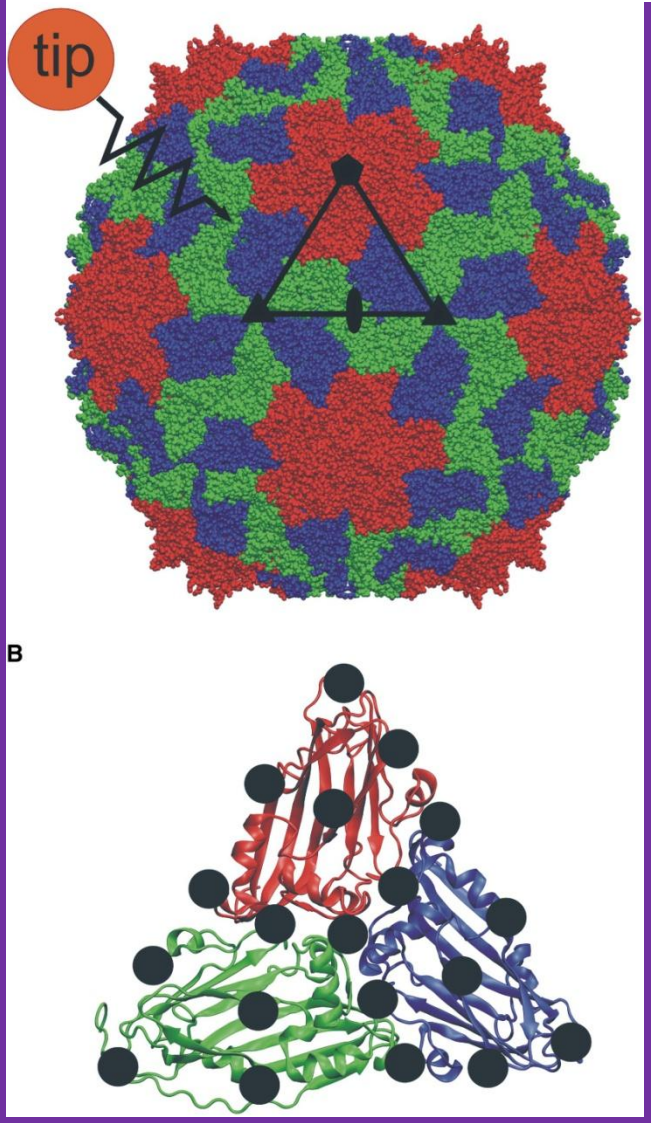
Roughly consistent with experimental behavior

Hard to distinguish bonds within the proteins from bonds between proteins.

Hard to parametrize the buckling instability at higher forces

No thermal fluctuations

All-atom simulations, Zink & Grubmuller 2009



(southern bean mosaic capsid, with water)

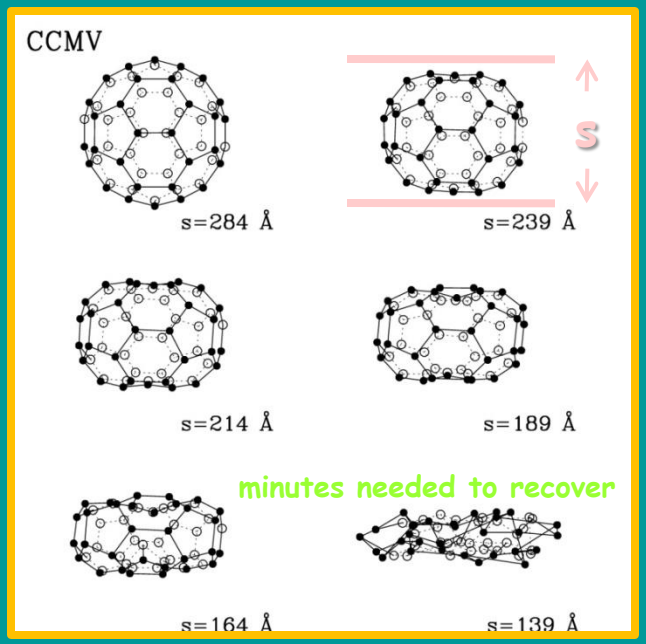
Nanoindentation times of order 1 ns compared to experimental 10 ms or more – much faster than the structural relaxation rates

The tip repreented by an atom that moves toward the center of mass. Short elastic region (1nm) followed by a rapid drop in F

The instability associated with bond breaking, but the bonds rapidly reform as the tip enters into the capsid

Coarse-grained structure-based model

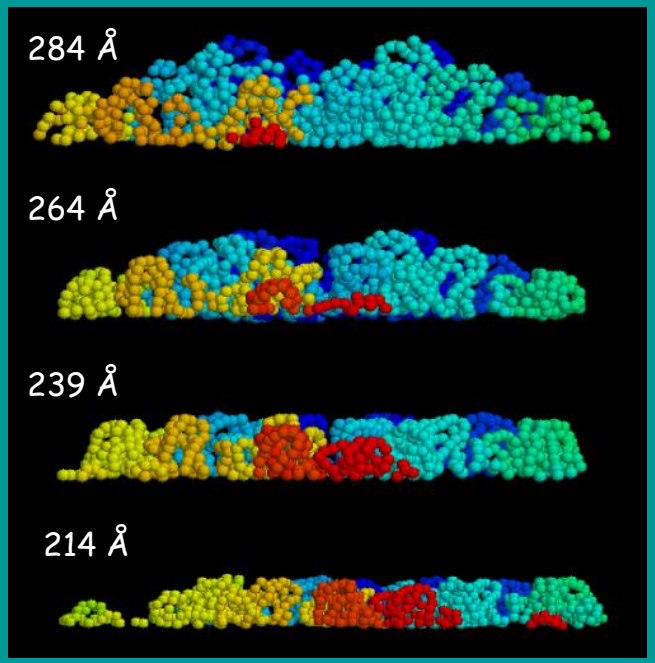
Compression by 2 plates, combined speed ~ 500 μs. Slow enough that stress can be transmitted across the capsid before the separation has changed substantially

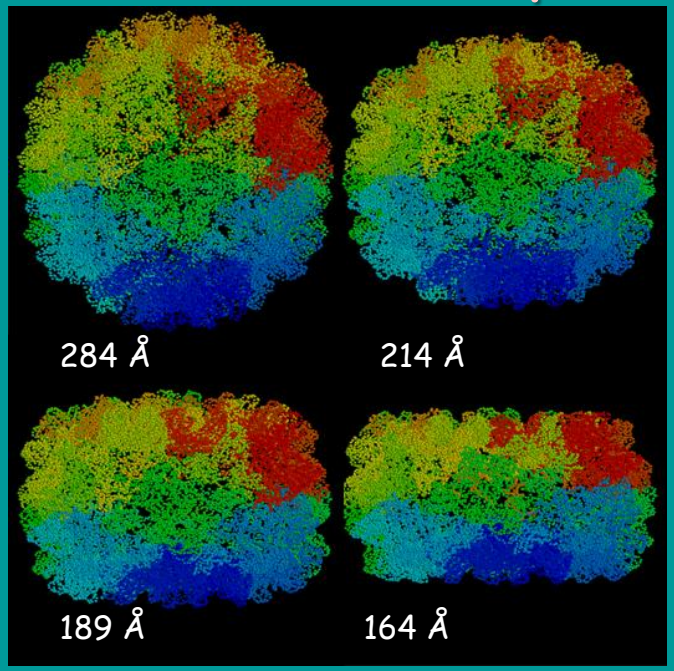


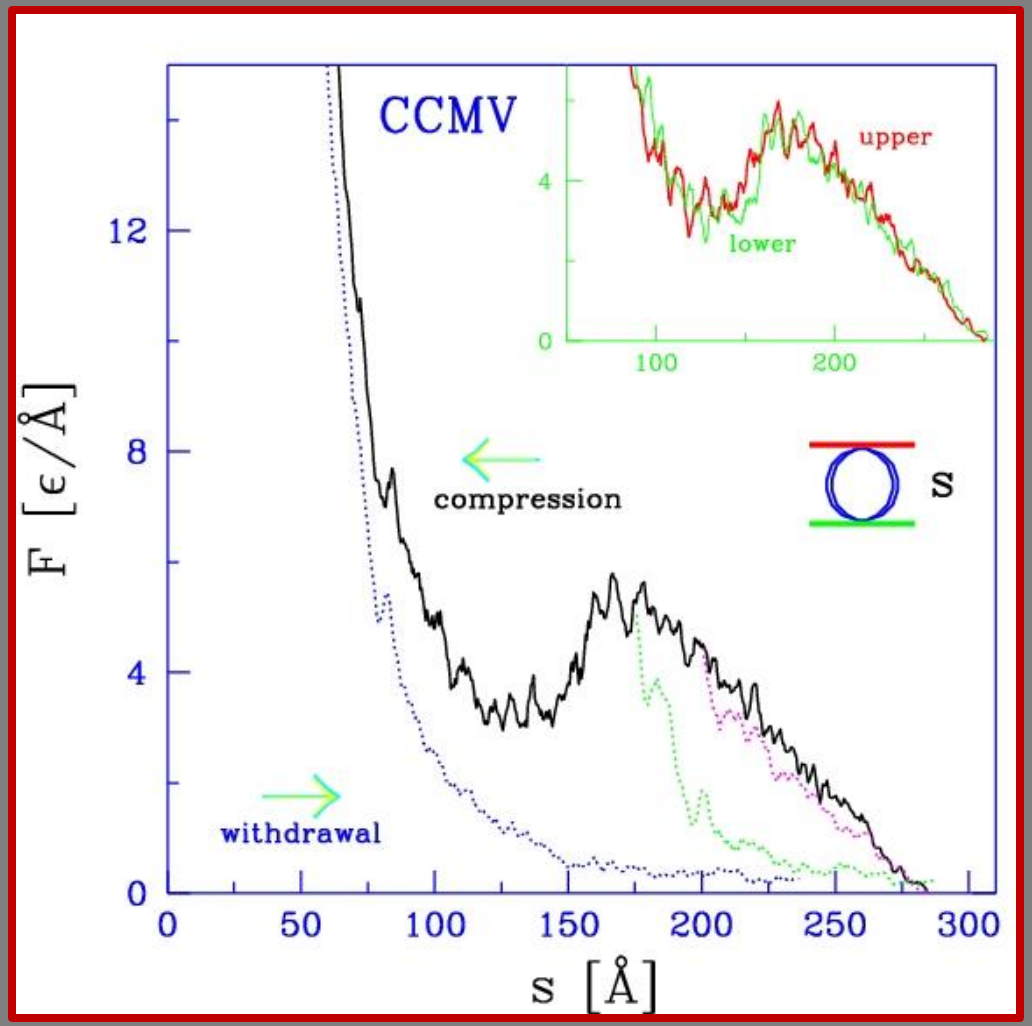
some clockwise rotation

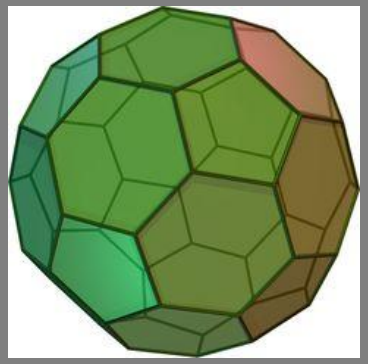
s – separation between the plates

Top 2150 C^{α} atoms



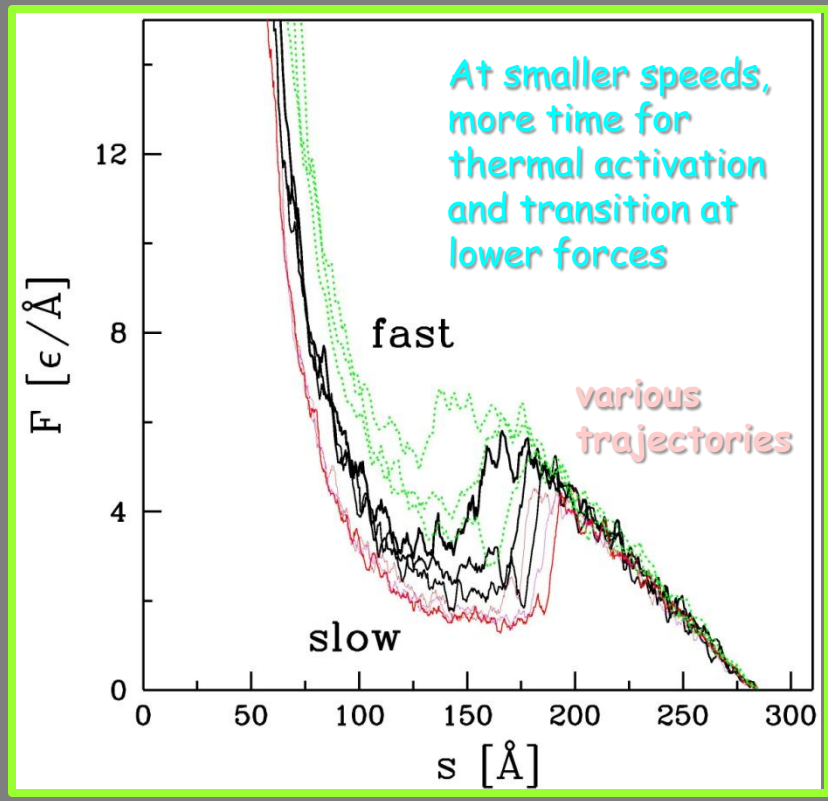






ACTIVATED TRANSITION

Breaking of a few bonds leads to stress transfer and a cascade of additional bonds

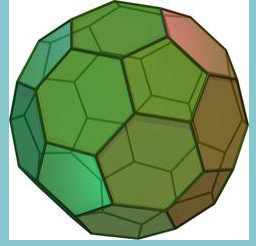


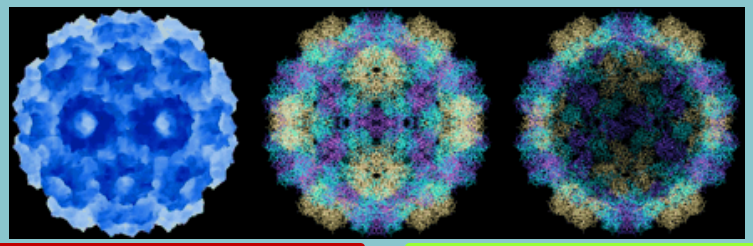
$k \sim 0.05\ \varepsilon/Å^2 \sim 0.055$ N/m

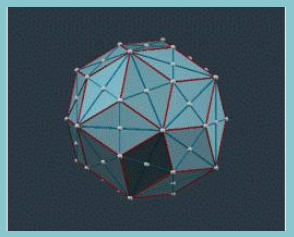
$F_m \sim 5.5\ \varepsilon/Å \sim 600$ pN

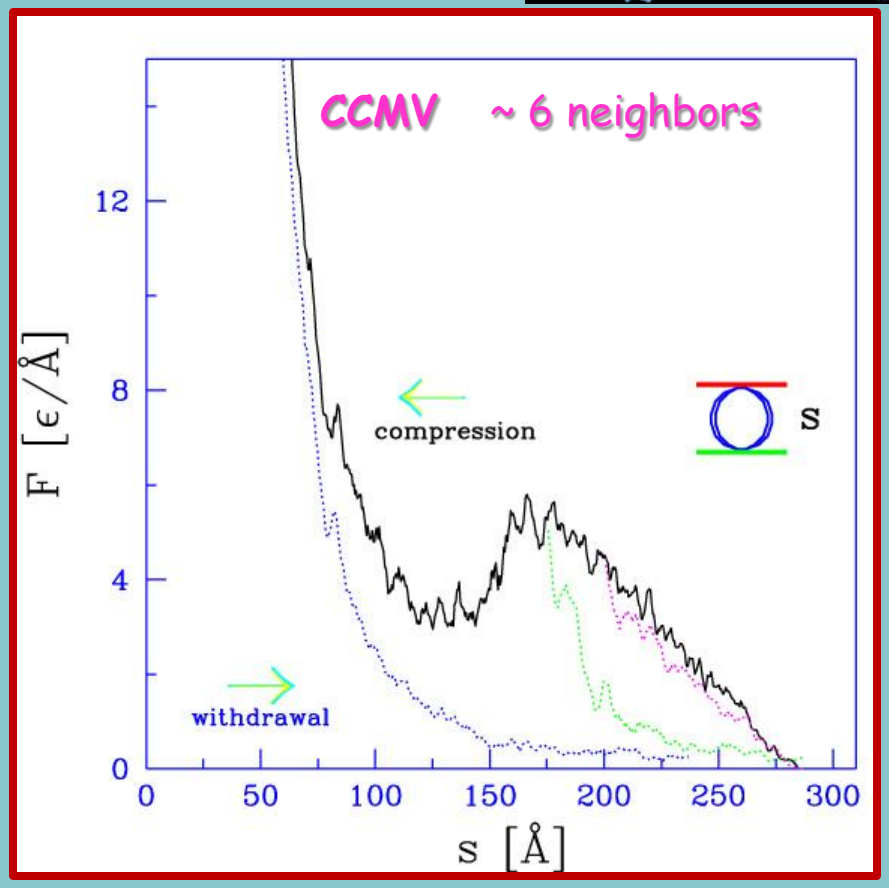
Experimental: 0.14 N/m & 500 pN

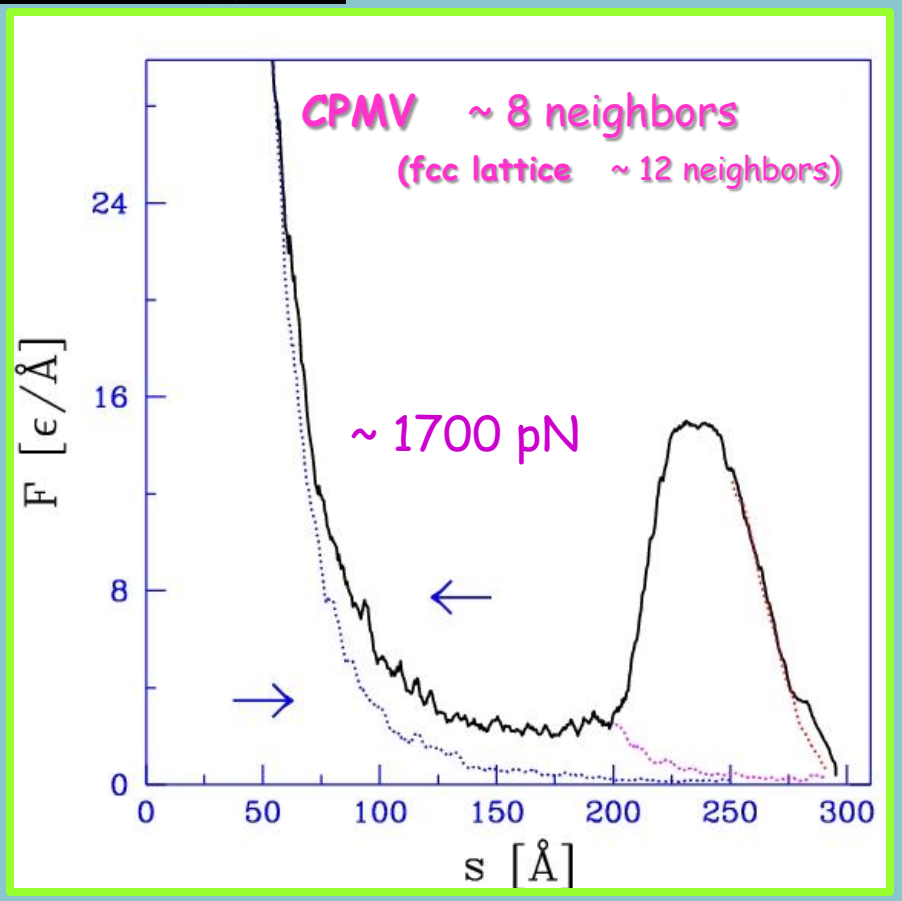
LJ radius in the potential ~ size of an amino acid (not atom) : softer



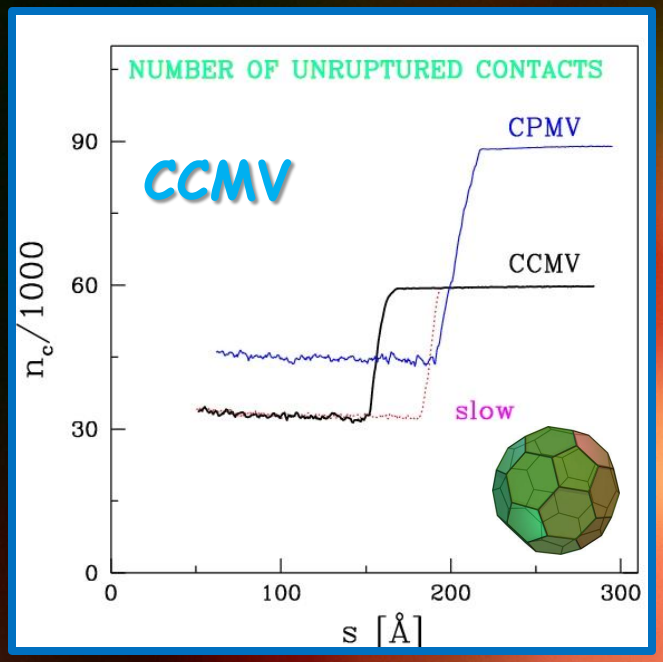








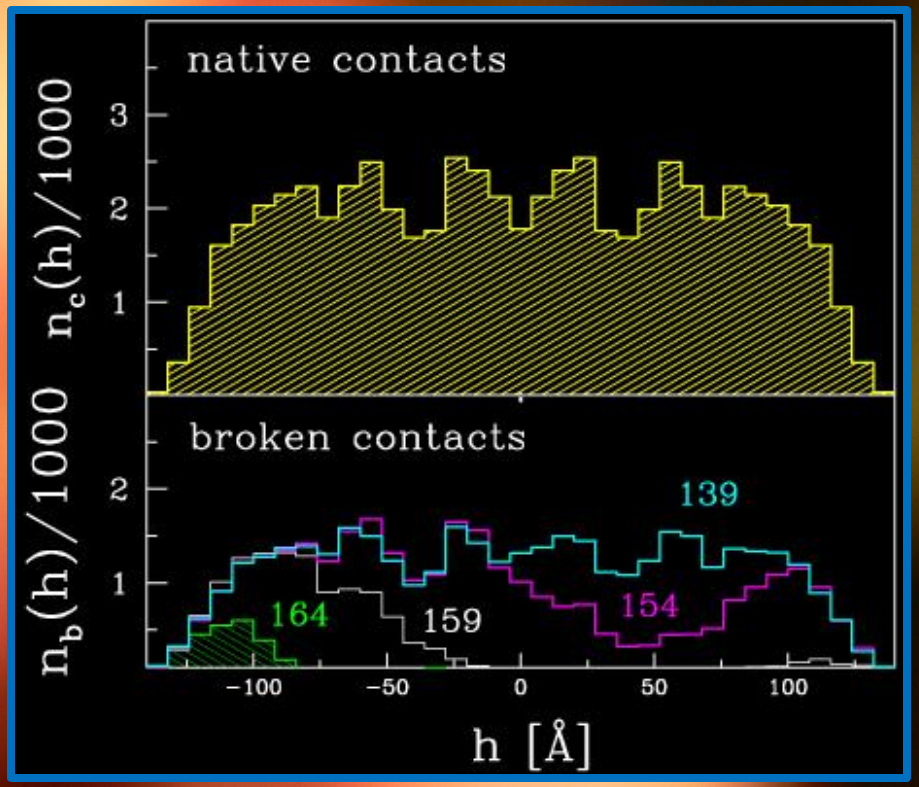
An order of magnitude bigger k & F_m 3 times as big despite comparable radius and shell thickness

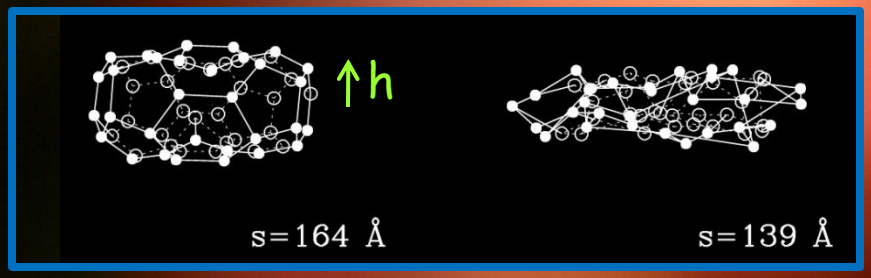


A rapid rupture of
~1/2 of the contacts

62000

Nearly all of them are
between the proteins





h- native z-coordinate of a C^{α} atom

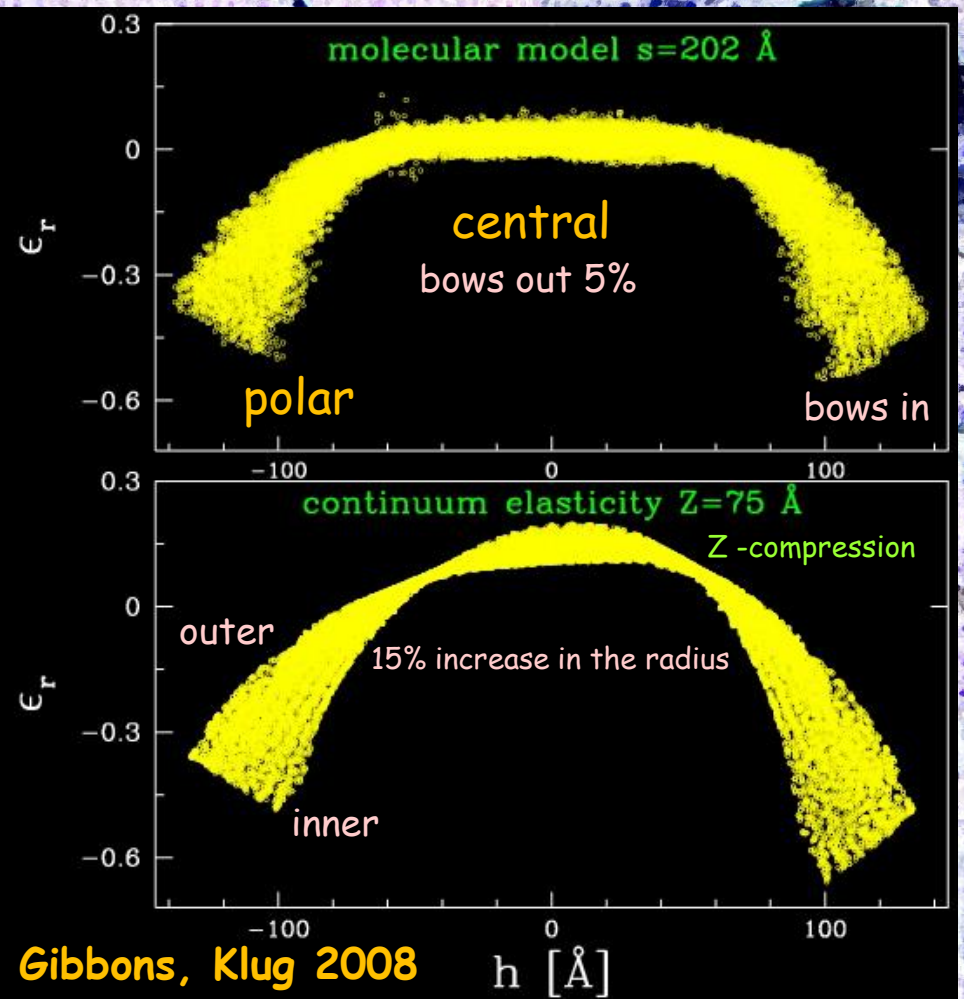
Starts at the bottom for this trajectory

Radial strain different than for a continuum shell

$$\varepsilon_r = \vec{r} \cdot \vec{dr} / \vec{r} \cdot \vec{r}$$

$\vec{r}$ the initial position relative to the center of mass

$\vec{dr}$ the change in this vector



Gibbons, Klug 2008

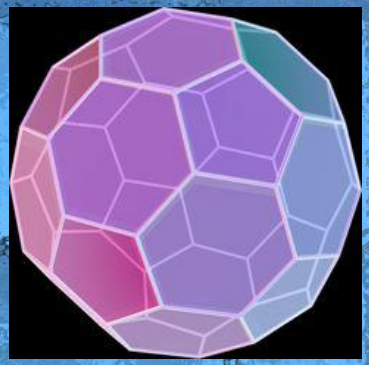
A nearly constant and small expansion in the center and a rapid change in the slope for |h|> 60 Å

Mean rotation ~ 4°

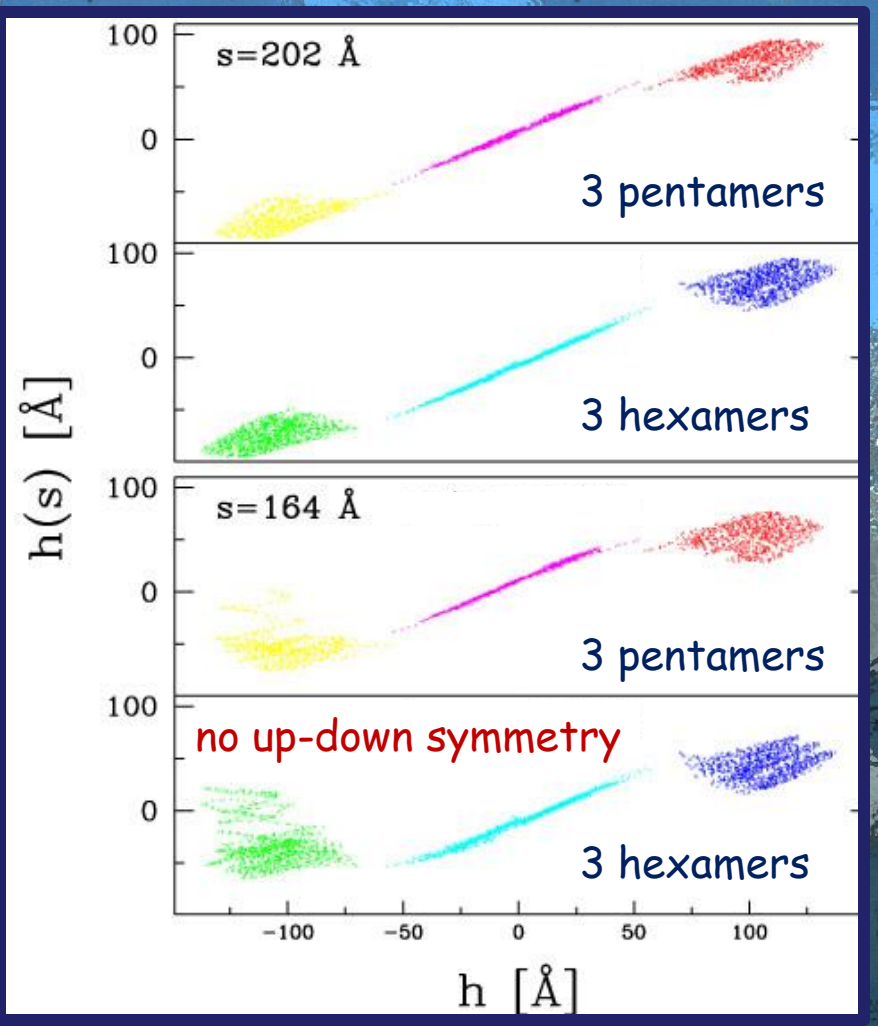
Some symmetry breaking due to buckling on one side

Vertical position of atoms belonging to particular n-mers

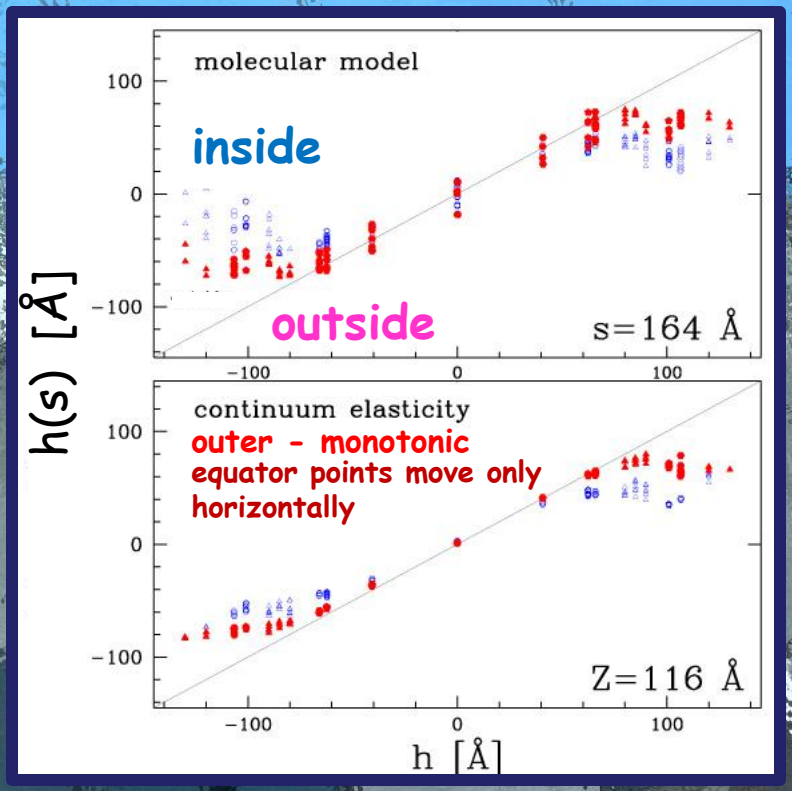
at separation s versus native



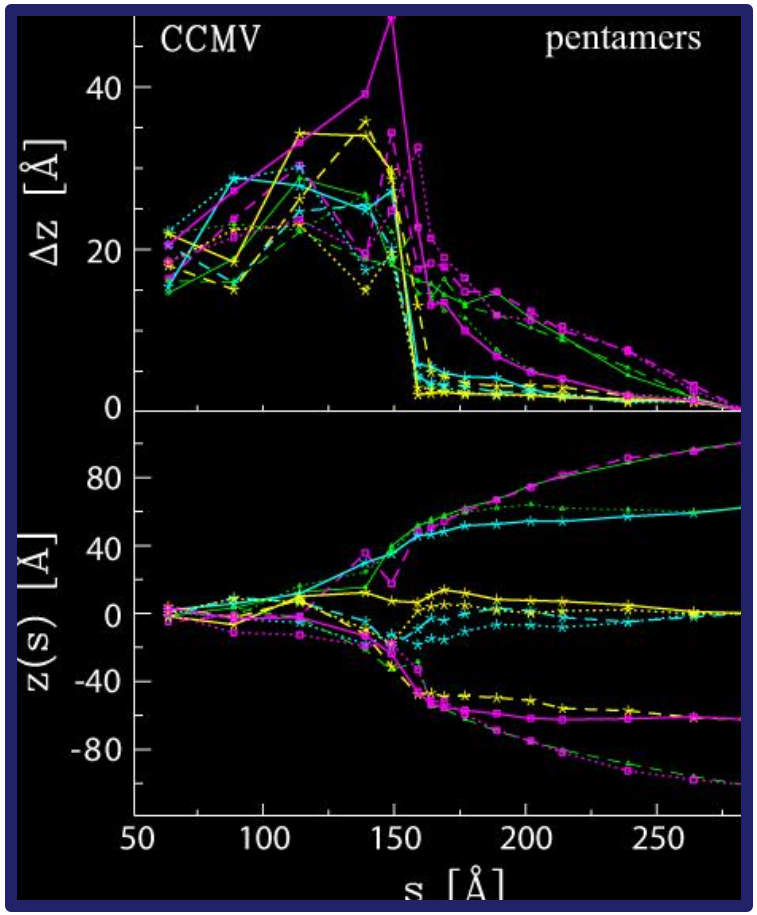
Central region nearly undeformed,
slope < 1 (~0.93) due to compression



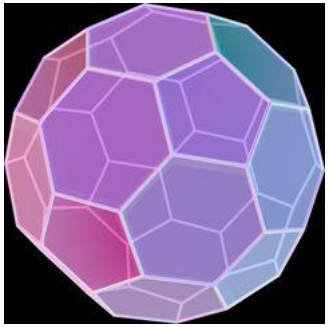
Outer n-mers pushed inwards and atoms displaced across the entire thickness.
A rapid change in the surface normal



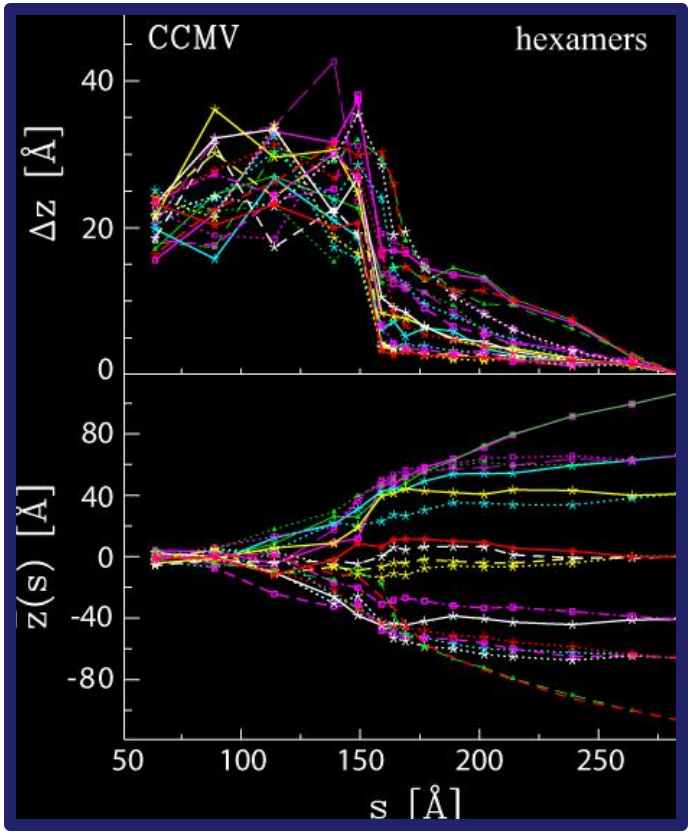
Not like an elastic shell



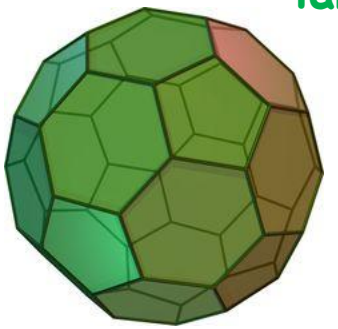
Center of mass height & rms variations for atoms belonging to particular n-mers



A smooth transition to the sandwich state with buckling: relative heights of some of the n-mers change sign



Polar n-mers undergo large rms variations



C^{α}-based description of empty CCMV & CPMV capsids.

Nanoindentation by a large tip modeled as compression between parallel plates.

Qualitatively consistent with continuum model. However, the details depend on the specifics of the molecular structure. A 30% increase in the number of contacts results in a 3-fold larger yield point and shorter elastic region – difficult to capture in continuum models

Elastic region followed by an irreversible activation transition to the sandwich state – related to rupturing nearly all of the bonds between capsid proteins

The molecular model undergoes a gradual symmetry breaking rotation and accomodates more strain near the walls

Simple geometry-based models miss many molecular details and yet can elucidate the microscopic picture of processes involving large conformational changes of biomolecules in an efficient way and with a large statistics of molecular dynamics trajectories.

Only certain variants of such models perform well when confronted with experimental data on stretching and also lead to folding.

(Handy for clarifying basic issues like: Thermal unfolding, on average, is reverse to folding and is unrelated to mechanical unfolding).

Scenarios represented on the time-contact order plane provide a detailed and useful description of the average time evolution.

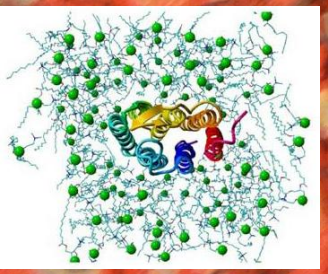
CONSTANT SPEED: survey of the PDB, determination of F_{max}, proposed list of strong proteins, correlations with the type of structure, identification of mechanical clamps.

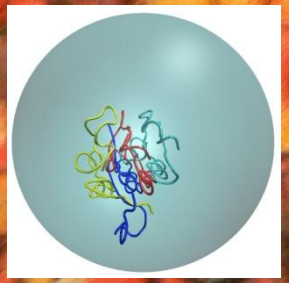
CONSTANT FORCE: exponential unfolding statistics below F_{max} and lognormal above it, refolding different than in the absence of the clamp.

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

HYDRODYNAMIC INTERACTIONS: affect time scales of processes – facilitate folding and mechanical unravelling but hinder unravelling by flow

topoisomerase





Models ready to tackle biological systems of a larger scale