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 $F_{\text{max}}$: for titin $\sim 200$ pN

---

**PROTEIN-DEPENDENT PATTERNS – NEEI**

Fowler Best Toca-Herrera Rutherford Steward Paci Karplus Clarke 2002
**TITIN**

3.5 MDa – gigantic muscle protein

~ 300 globular domains

**sarcomere**

1 μm long (x 4)

**Z-disk**

**M-line**

**β-sandwich architecture**

**strands**

**Constant speed**

0.06 – 0.6 N/m

1 – 10⁴ nm/s
'TITIN IS A WEIRD SPRING' Erickson 1997

SERIAL UNWINDING

I27-I34

sawtoothlike

Force (pN)

Extension [nm]

Rief Gautel Oesterhelt Fernandez Gaub 1997

Apparatus effects at small extensions

Li Linke Oberhauser Carrion-Vazquez Kerkvliet Lu Marszalek Fernandez 2002

Each protein has its own pattern

DNA

Bockelmann Essevaz-Roulet Heslot 1997

force [pN]

displacement [µm]
Linkage dependent elasticity

UBIQUITIN N=76

Studies usually involve homo- or hetero-linkages of modules

F_{max} 203±35 pN

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

Assumption of seriality

Chian et al. 2004

~ logarithmic dependence on \( v_p \)

Expected: a constant at \( v_p \rightarrow 0 \)

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

\( 1 \) & \( 76 \)

\( 48 \) & \( 76 \)

\( 85±20 \) pN

\( \downarrow F_{max} \)
**ALL-ATOM SIMULATIONS**

Lu Schulten 2000  (Paci Karplus 2000)

Typically 10 ns time scales

\[ \text{T} = 300 \text{ K} \]

1 domain

\[ \text{force (pN)} \]

\[ \text{0.5 Å/ps} \]

\[ \text{0.1 Å/ps} \]

\[ 10^{10} \text{ nm/s} \]

(Hydrodynamic interactions reduce \( F_{\text{max}} \))

**Difficult:**

Comparison of processes involving large conformational changes

Studies of large sets of proteins

Pabon, Amzel 2006 - quasistatic ~500 pN
Experimental results on stretching at constant speed

~55 proteins

All-atom simulations on ~ 21 proteins

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

What proteins are strong and why?

J. Fernandez
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
2. Stretching at constant force
3. Stretching by fluid flow
Go models of proteins – coarse grained: only the $C^\alpha$ atoms

Constructed from the experimentally derived native structure
\[
E_p(\{r_i\}) = V^{BB} + V^{\text{NAT}} + V^{\text{NON}} + V^{\text{CHIR}}
\]

\[V^{BB} - \text{TETHERING of consecutive beads at } 3.8 \, \text{Å} = d_0\]

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

\[\epsilon: 1.6 \, \text{kcal/mol} \sim 800 \, \text{K}\]

\[\text{Room T: } 0.35 \, \epsilon\]

\[\sigma_{ij} \text{ calculated based on the van der Waals radii of the atoms}\]

Tsai Taylor Chothia Gerstein 1999

Non-native: repulsive with \( \sigma=4\,\text{Å} \)

Disulfide bonds enhanced

\[V^{\text{CHIR}}: \text{angular terms locally favoring the native shape of the backbone}\]
Large friction
Finite bead size modify the effective time scale $\tau$

Veitshans, Klimov, Thirumalai 1997: $\tau \sim 3$ns

When comparing with experiments

Stretching $\sim$ independent of $\gamma$
Folding time linear in $\gamma$

Use $\gamma = 2m/\tau$

Water-like $\gamma \sim 25$ times bigger

MOLECULAR DYNAMICS

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

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

$$\tau = \sqrt{m a^2 / \epsilon} \sim 3 \text{ps} \quad \text{(small $\gamma$)}$$

$$a = 5 \text{Å} \approx < \sigma_{ij} >$$

$m = 118 m_p$

Langevin noise as a thermostat and as an emulator of water
$v_p = 0.005 \text{ Å/Å} \sim 10^6 \text{ nm/s}$

logarithmic shifts with $v_p$

$k_s = 0.12 \varepsilon/\text{Å}^2 \sim 0.08 \text{ N/m}$

soft: $k_s = 0.12 \varepsilon/\text{Å}^2 \sim 0.08 \text{ N/m}$

stiff: $k_s = 30 \varepsilon/\text{Å}^2$

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

one domain

$\tau = 0$

$\tilde{\tau} = 0.8$

$A' - G$ temperatures

$A - B$
The pattern depends on \( T \) and at \( T=0.3 \) it is similar to the experimental results.

5 x I27

3 domains

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

\[ 5 \times I27 \]

\[ \tilde{T}=0.8 \]

\[ \sim 220 \text{ pN} \]

\[ \tilde{T}=0.8 \]

\[ \tilde{T}=0 \]

From serial to parallel

End-to-end distance

Breaking native contacts

Tip displacement

Contacts - identified by the sequential distance

Peak force
Validation of the Go model for stretching

Correct contact map - should work close to the native state

\[ \varepsilon = 1.6 \text{ kcal/mol} \]
\[ \Rightarrow \varepsilon / \text{Å} = 110 \text{ pN} \]

@ \( T = 0.3 \varepsilon / k_B \)
Bacteriorhodopsin pulled out of a membrane (By the C-terminus)

Janovjak et al. 2003

Apparatus effects on short distances

$\tilde{T} = 0.3$

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 $40 \leq N \leq 150$: set $S_{7510}$

239 with $150 < N \leq 851$

7,749

3,813 with $40 \leq N \leq 150$:

CATH-based assignment of topology to the fold available

Hierarchy:

Class
Architecture
Topology
Homology
The spread in $F_{\text{max}}$ depends on $N$ only weakly, but the larger the $N$ the bigger the chance of a large peak force.
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

- **β Sandwich (379)**
  - Immunoglobulin-like
  - Jelly Roll

- **Roll α/β**
  - Nuclear Transport Factor
  - Chitinase
  - Mannose Binding Protein
  - Ubiquitin-like
  - P-30 Protein

- **Other**
  - Immunoglobulin & Transport proteins

Like titin

Like ubiquitin
Mechanisms of rupture in strong short proteins

95%: shearing of hydrogen-bonded parallel \( \beta \)-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

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

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

Ranked 14

Ribonuclease A

1rnz

S-S bonds cannot be ruptured

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

Discher 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

Fractional extension for 2 trajectories

unfolding in a single kinetic step refolding – in multiple steps

Fractional extension for 2 trajectories

2 domains
DISTRIBUTION OF UNFOLDING TIMES

Log-normal above $F_{\text{max}}$

Exponential below $F_{\text{max}}$

Fractional extension averaged over many realizations

$$C = \frac{\langle t \rangle}{\sigma_t} \quad C' = \frac{\langle t^2 \rangle^3}{\langle t \rangle^3 \langle t^3 \rangle}$$
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

Tension is non-uniform: increases towards the anchored end

\[ m\ddot{r}_i = -\gamma(\dot{r}_i - u(r_i)) + F_i^c + \Gamma \]

\[ F = \gamma \sum_i^N u(r_i) \]

Force at the fixed 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

\[ \frac{F}{L/100 \text{ Å}} \]

\[ \frac{U}{t/\tau} \approx 0.4 \text{ Å/ns} \]

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

\[ Pe = \frac{UR_g}{D} \]

**flow vs. diffusion**

**Ubiquitin - simulation:**
- \( D = 5 \text{ Å}^2/\tau \)
- \( R_g = 11.5 \text{ Å} \)
- \( U = 0.1 \text{ Å}/\tau \)

**Ubiquitin - experiment:**
- \( D = 1.7 \text{ cm}^2/\text{s} \)
- \( \tau \) corresponds to 0.25 ns

**Experiment:**
- \( U = 4 \text{ cm/s} \)
- \( Pe = 0.2 \)
- 1000 faster than for DNA
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 $F_{\text{max}}$, proposed list of strong proteins, correlations with the type of structure, identification of mechanical clamps.

**CONSTANT FORCE**: exponential unfolding statistics below $F_{\text{max}}$ 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 $C^\beta$ atoms.

A strength-modulated contact map

Frequent elimination of the secondary force peaks

Certain reshuffling of the ranking