How does order grow out of disorder? – A journey inspired by Prof. C.C. Lin

Liu HONG

Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University

Prof. Lin’s Legend

- Three working periods:

  1940’s-60’s -- hydrodynamic stability (macroscopic scale)

  1960’s-80’s -- spiral structure in galaxies (cosmoscopic scale)

  2000’s-10’s -- protein folding (microscopic scale)
- To understand how the nature is organized and works in virtue of applied mathematics

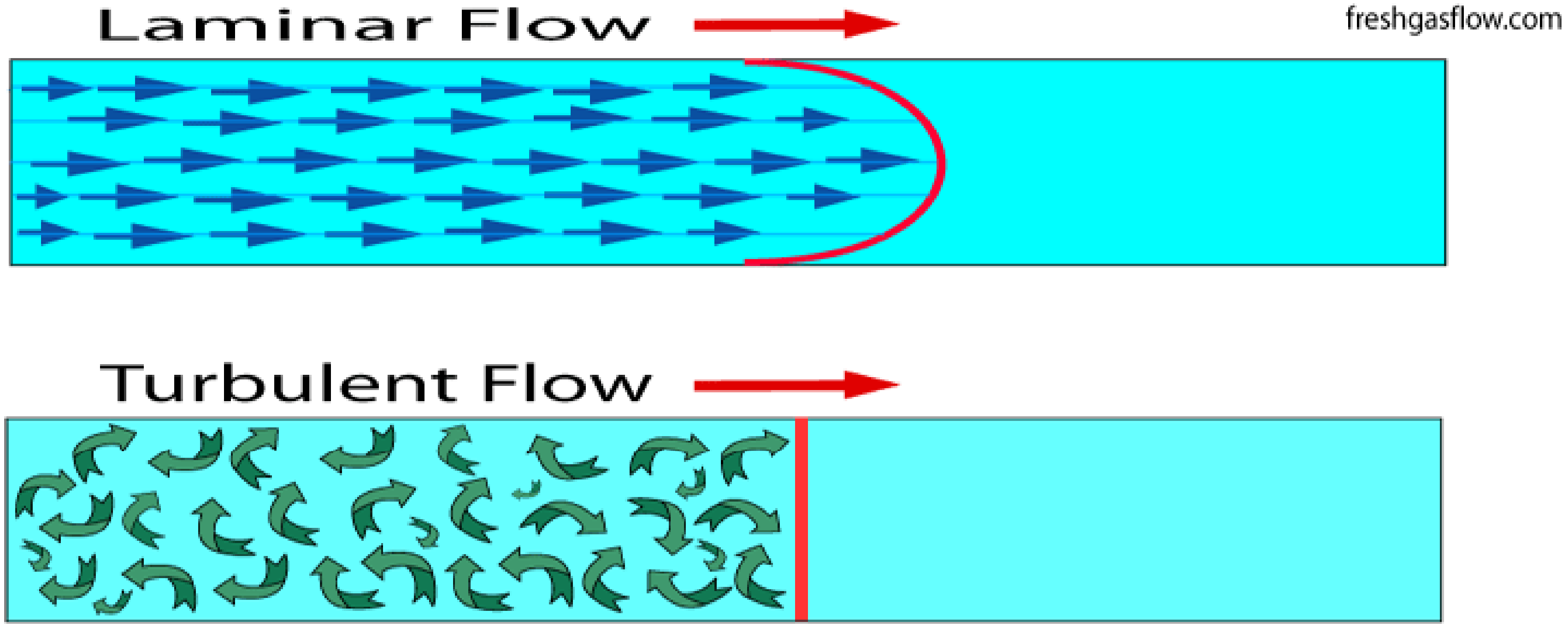
I. Hydrodynamic Stability

– From laminar flows to turbulence
in the macroscopic scale

2-D Navier-Stokes Equations

- 2-D incompressible, unsteady flow

$$\begin{cases} \dfrac{\partial u}{\partial x} + \dfrac{\partial v}{\partial y} = 0, \\ \dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} + v\dfrac{\partial u}{\partial y} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial x} + \upsilon\left(\dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2}\right), \\ \dfrac{\partial v}{\partial t} + u\dfrac{\partial v}{\partial x} + v\dfrac{\partial v}{\partial y} = -\dfrac{1}{\rho}\dfrac{\partial p}{\partial y} + \upsilon\left(\dfrac{\partial^2 v}{\partial x^2} + \dfrac{\partial^2 v}{\partial y^2}\right), \end{cases}$$


Laminar Flow
freshgasflow.com
Turbulent Flow

Equations for Perturbation

- Substitute u = U + u'; v = v'; p = P + p' (Poiseuille flow).
- Neglect quadratic terms in the perturbations and assume $\frac{\partial \mathrm{U}}{\partial \mathrm{x}} << \frac{\partial \mathrm{U}}{\partial \mathrm{y}}$

$$\begin{cases} \dfrac{\partial u'}{\partial x} + \dfrac{\partial v'}{\partial y} = 0, \\ \dfrac{\partial u'}{\partial t} + U\dfrac{\partial u'}{\partial x} + v'\dfrac{\partial U}{\partial y} = -\dfrac{1}{\rho}\dfrac{\partial p'}{\partial x} + \upsilon\left(\dfrac{\partial^2 u'}{\partial x^2} + \dfrac{\partial^2 u'}{\partial y^2}\right), \\ \dfrac{\partial v'}{\partial t} + U\dfrac{\partial v'}{\partial x} = -\dfrac{1}{\rho}\dfrac{\partial p'}{\partial y} + \upsilon\left(\dfrac{\partial^2 v'}{\partial x^2} + \dfrac{\partial^2 v'}{\partial y^2}\right), \end{cases}$$

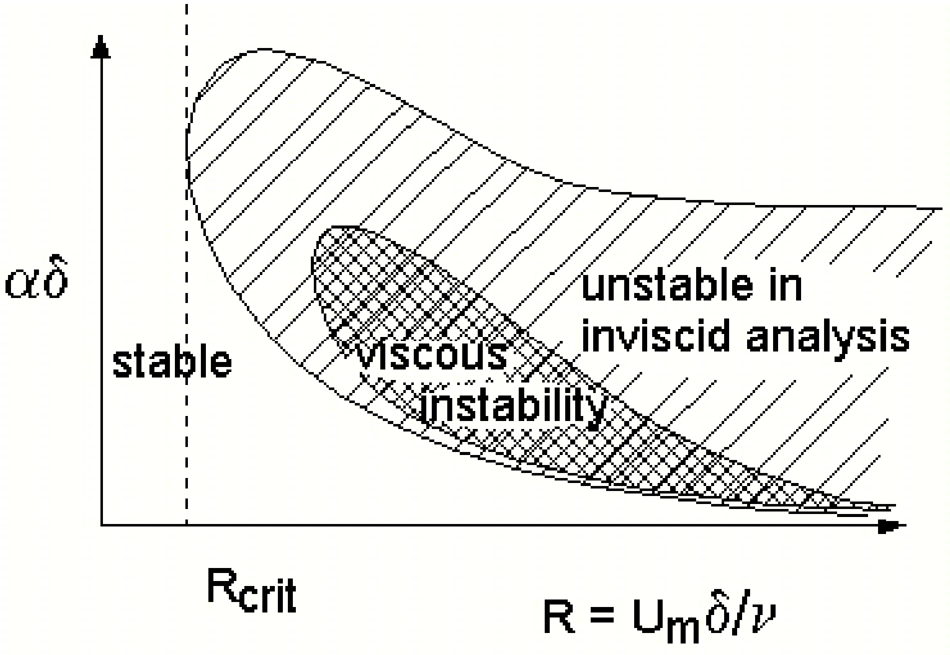
Orr-Sommerfeld Equation

- Assume a periodic disturbance (linear assumption)
- For 2-D flow, the stream-function can be defined as $\Psi(x,y,t) = \Phi(y)e^{i(\alpha x - \beta t)}$
- Then $u' \equiv \frac{\partial \varphi}{\partial y} = \phi'(y)e^{i(\alpha x - \beta t)}$ $\quad v' \equiv \frac{\partial \varphi}{\partial x} = -i\alpha\phi'(y)e^{i(\alpha x - \beta t)}$
- Orr-Sommerfeld equation

$$(u - c)(\phi'' - \alpha^2\phi) - U''\phi = -\frac{i}{\alpha \mathrm{Re}}(\phi'''' - 2\alpha^2\phi'' + \alpha^4\phi)$$

Thumb Curve

- Below some critical Reynolds number, all sorts of disturbances die out, so the shear layer is stable to any disturbance. For Reynolds numbers greater than this, there are instabilities which can amplify. As Reynolds number becomes very large, most disturbances result in turbulence.



History

- In 1924, Heisenberg guessed an approximate solution to the problem of the onset of turbulence from laminar flow between two parallel plates.
- In 1944, C.C. Lin verfied heisenberg‘s guess analytically in his Ph.D. Thesis, which later confirmed by J. von Neumann and L.H. Thomas numerically.

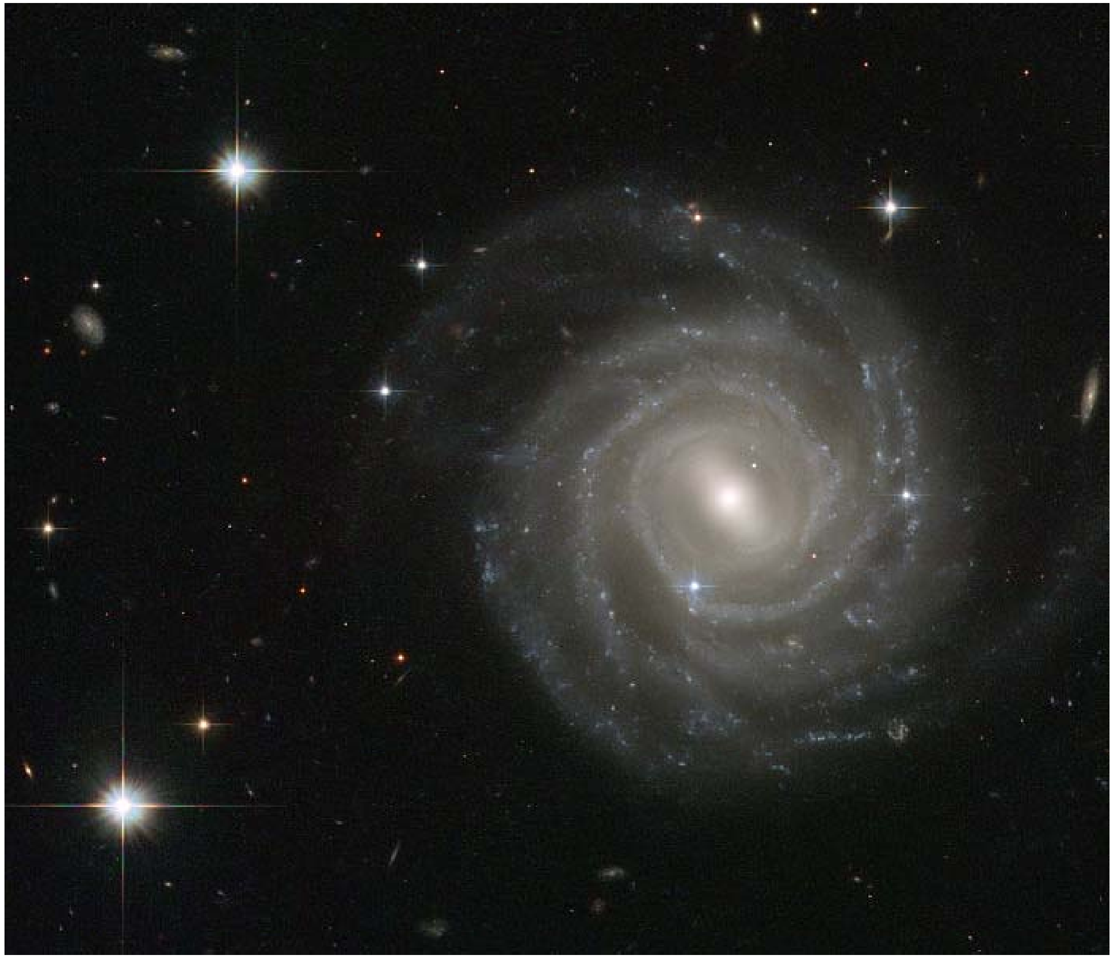
- On the stability of two-dimensional parallel flows, C.C. Lin, *PNAS*, 1944.
- *The Theory of Hydrodynamic Stability*, C.C. Lin, Cambridge, 1955.

II. Spiral Structure in Galaxies

– The formation and sustention of
spirial structures in the cosmoscopic
scale

Spiral Galaxies

- A flat, rotating disc of (mostly newly created) stars and interstellar matter;
- A central stellar bugle of mainly older stars, which resembles an elliptical galaxy;
- A near-spherical halo of stars, including many in globular clusters;
- A supermassive black hole at the very center of the central bulge



Barred spiral galaxy UGC 12158.

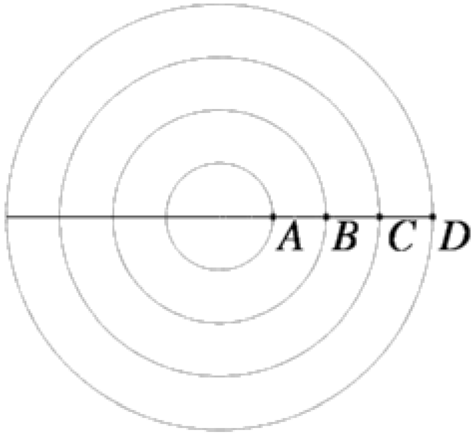
NGC 1300 in infrared light.

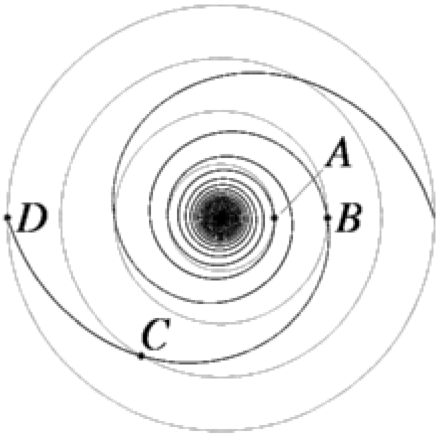
Spiral galaxy NGC 6384.

The Winding Problem

- Spiral rotation curves are not "solid body" rotation curves.

  Spiral rotation curves are flat: V_c ~ constant. Therefore the angular rotation rate goes like V_c/r, so that the outer edge takes longer to complete an orbit than the inner regions. So no solid arm could last very long -- it would quickly get "wound up“.





- what's wrong with this picture?

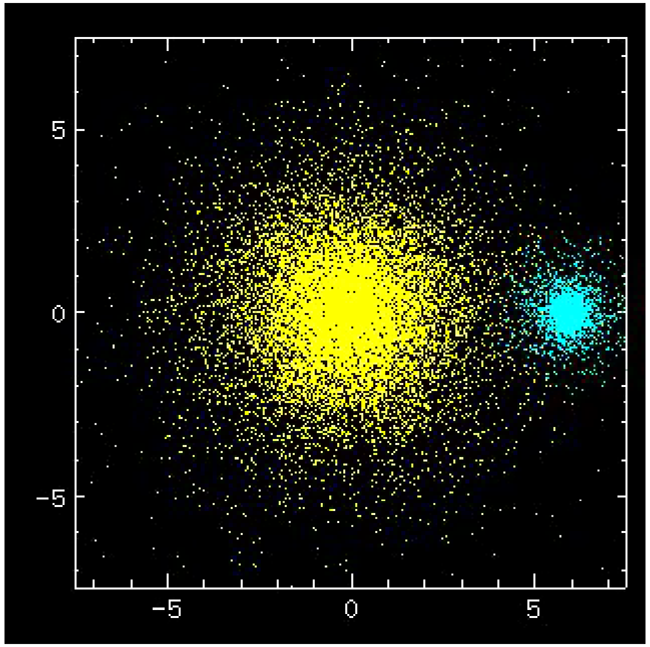
  It happens too fast. It happens in a few rotation periods (say, half a billion years). The age of the Universe is something like 10-15 billion years -- if spirals made their arms this way, they should all be so tightly wrapped we wouldn't see them as spiral arms anymore! So spiral arms can't be physical objects.

Spiral Density Waves I

- Imagine the spiral pattern is a density wave. It rotates through the galaxy at a fixed angular speed, called the pattern speed. In the inner parts of the galaxies, stars are moving faster than the pattern speed, and overtake the density. In the outer parts of the galaxies, stars move more slowly than the pattern speed, and the spiral arms overtake the stars.

Spiral Density Waves II

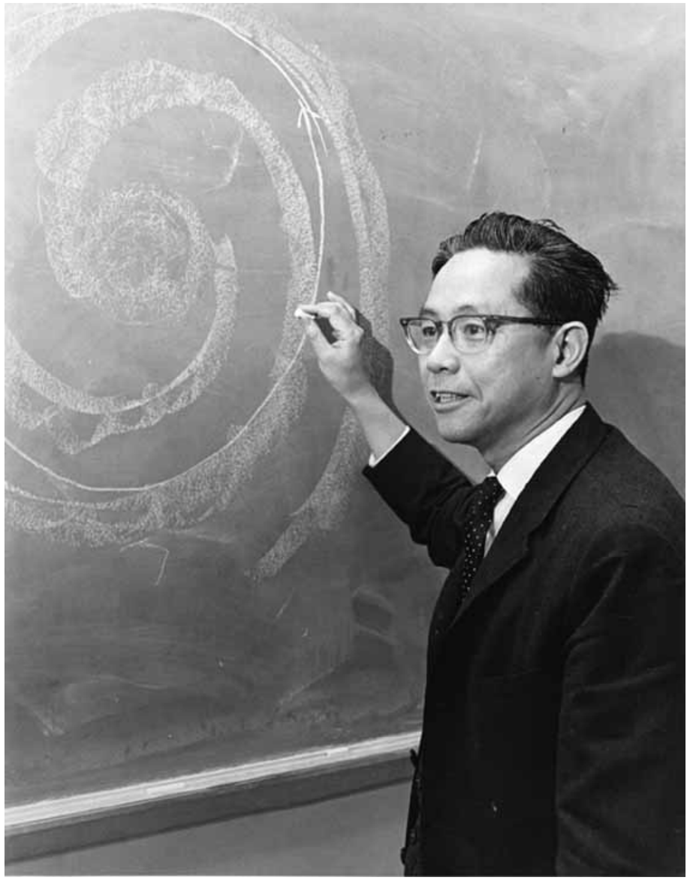
- When a star moves near the wave, it speeds up towards the density wave. After it passes through the wave, it slows down and leaves very slowly. So the star spends more time around the density wave than it otherwise would. We see this as an enhanced density of stars -- a spiral arm!



An example of how an encounter between a big galaxy and a small satellite companion can drive spiral structure.

History

- The pioneer of studies of the rotation of the Galaxy and the formation of the spiral arms was Bertil Lindblad in 1925.
- This idea was developed into density wave theory by C.C. Lin and Frank Shu in 1964.



- On the spiral structure of disk galaxies, C.C. Lin & F.H. Shu, *Astrophys. J.,* 1964, 1969; *PNAS*, 1966.
- *Spiral Structure in Galaxies: A Density Wave Theory*, G. Bertin and C.C. Lin, MIT Press, 1995.

III. Protein Folding and Misfolding

– The polymer dynamics and energy cascade in the microscopic scale

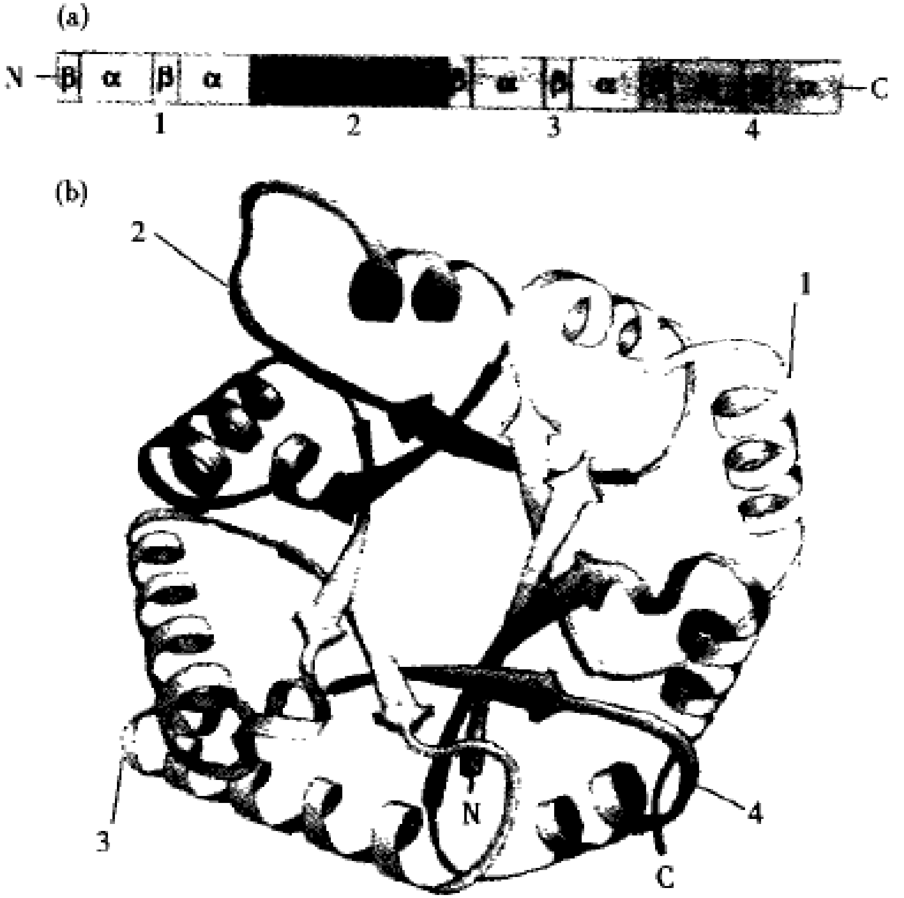


Fig.1 Motifs that are adjacent in the amino acid sequence are also usually adjacent in the three-dimensional structure. Triose-phosphate isomerase is built up from four β-α-β-α motifs that are consecutive both in the amino acid sequence (a) and in the three-dimensional structure (b)

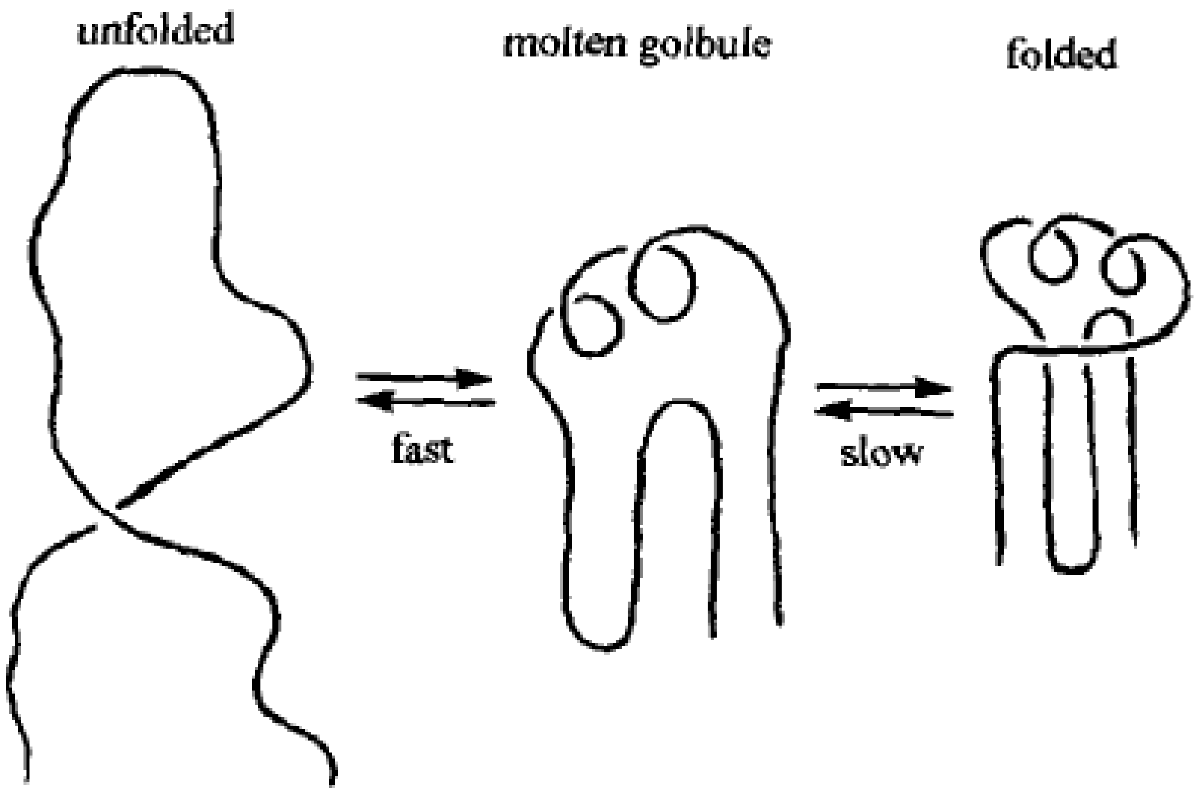


Fig.3 The molten globule state is an important intermediate in the folding pathway when a polypeptide chain converts from an unfolded to a folded state. The molten globule has most of the secondary structure of the native state but it is less compact and the proper packing interactions in the interior of the protein have not been formed

Energy Cascade Theory

- The secondary structures in protein are largely regular structures determined by the principle of minimum free energy, while the flexible(irregular) structures on other scales are primarily determined by a dynamic process.
- This dynamic process may be treated by the principles of statistical physics. Specifically, it is the kinetic theory for dissipative systems (in analogy with Heisenberg’s theory of isotopic turbulence).

Spectrum for Structure Energy

- By using arguments analogous to Kolmogorov’s, we can arrive at a similar prediction for the spectrum of the tertiary structure of the molecular of protein. The energy spectrum of the structure energy (and the kinetic energy of internal oscillations) should follow the formula $E(k) \sim G\varepsilon^{2/3}\kappa^{-5/3}$ at small scales.
- For the study of the structure and function of proteins, where the latter is generally believed to be directly related to the tertiary structures, the use of the spectral representation is especially convenient, since the local structure is most clearly described by the spectrum at large wave numbers.

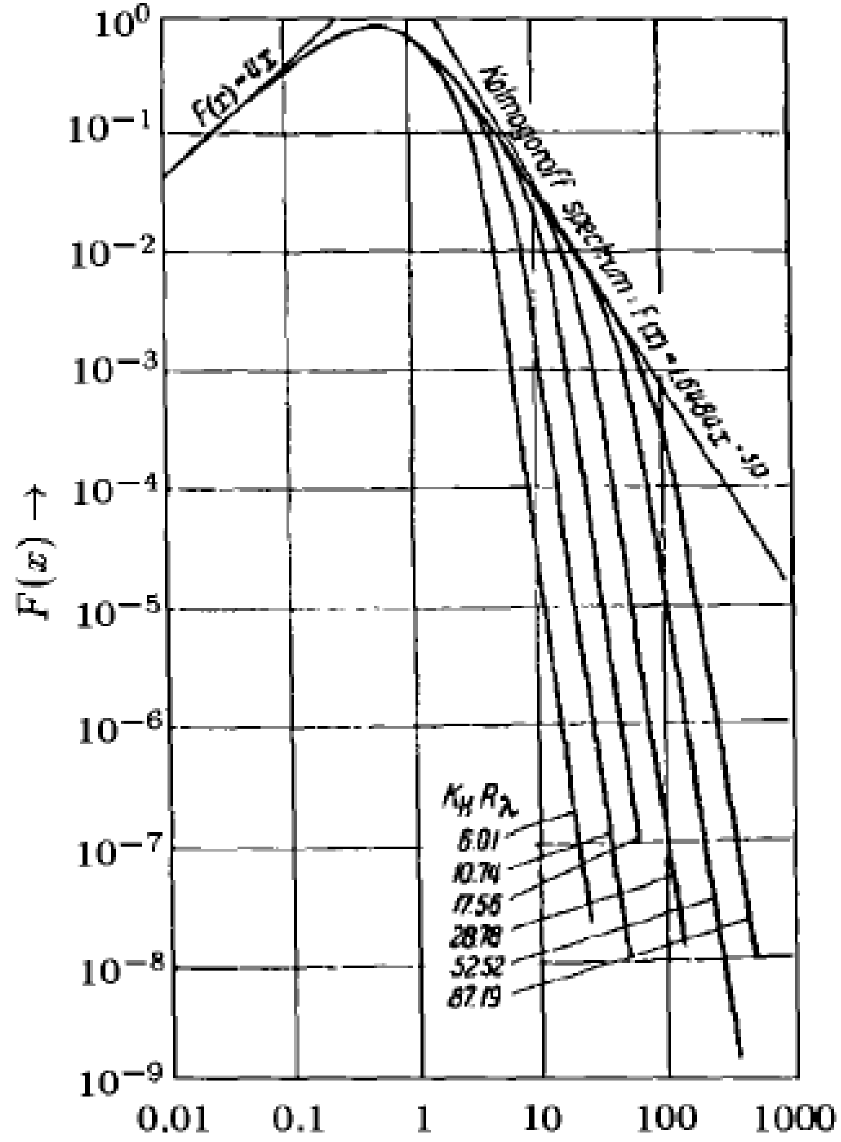


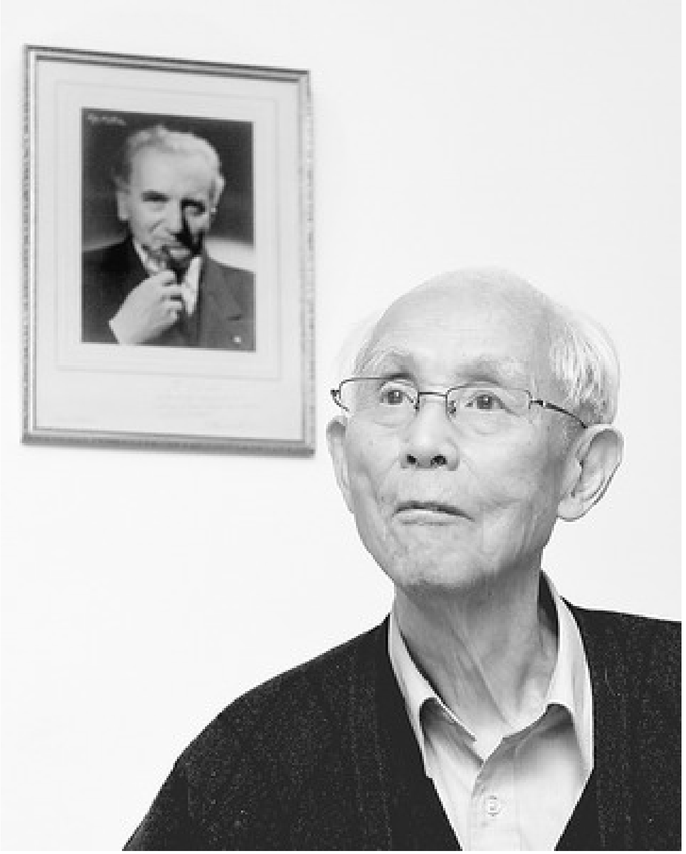
Fig.2 The energy spectrum according to Heisenberg's transfer theory with initial period similarity

Dynamic Equation

- In this description, the process of protein folding and unfolding may be described as the evolution of the spectrum E(k, t) (in seconds) by the schematic differential equation

$$\frac{dE(k,t)}{dt} = F - D$$

- In the above equation, E(k, t)is the energy spectrum at time t, k is the wave number, D is dissipative process which occurs primarily at large wave numbers, and F represents the processes of folding and initial conformational motions that leads to the formation of protein structure.



- On the evolution of applied mathematics, C.C. Lin, *ACTA MECHANICA SINICA*, 2003.

IV. Some Further Developments in Protein Folding and Misfolding

Outline

- 1. Scaling laws for protein structure (tertiary structure)
- 2. Theory for helix-sheet-coil transition (secondary structure)
- 3. Kinetic models for protein aggregation (protein misfolding and amyloid fiber formation)

1. Scaling laws for Tertiary Structure

- For homopolymers, it is well-known that

  $Rg \propto N^{3/5}$ in good solvent;

  $Rg \propto N^{1/3}$ in poor solvent.

- For natural proteins (heteropolymer), we found that $Rg \propto N^{2/5}$ under physiological conditions.

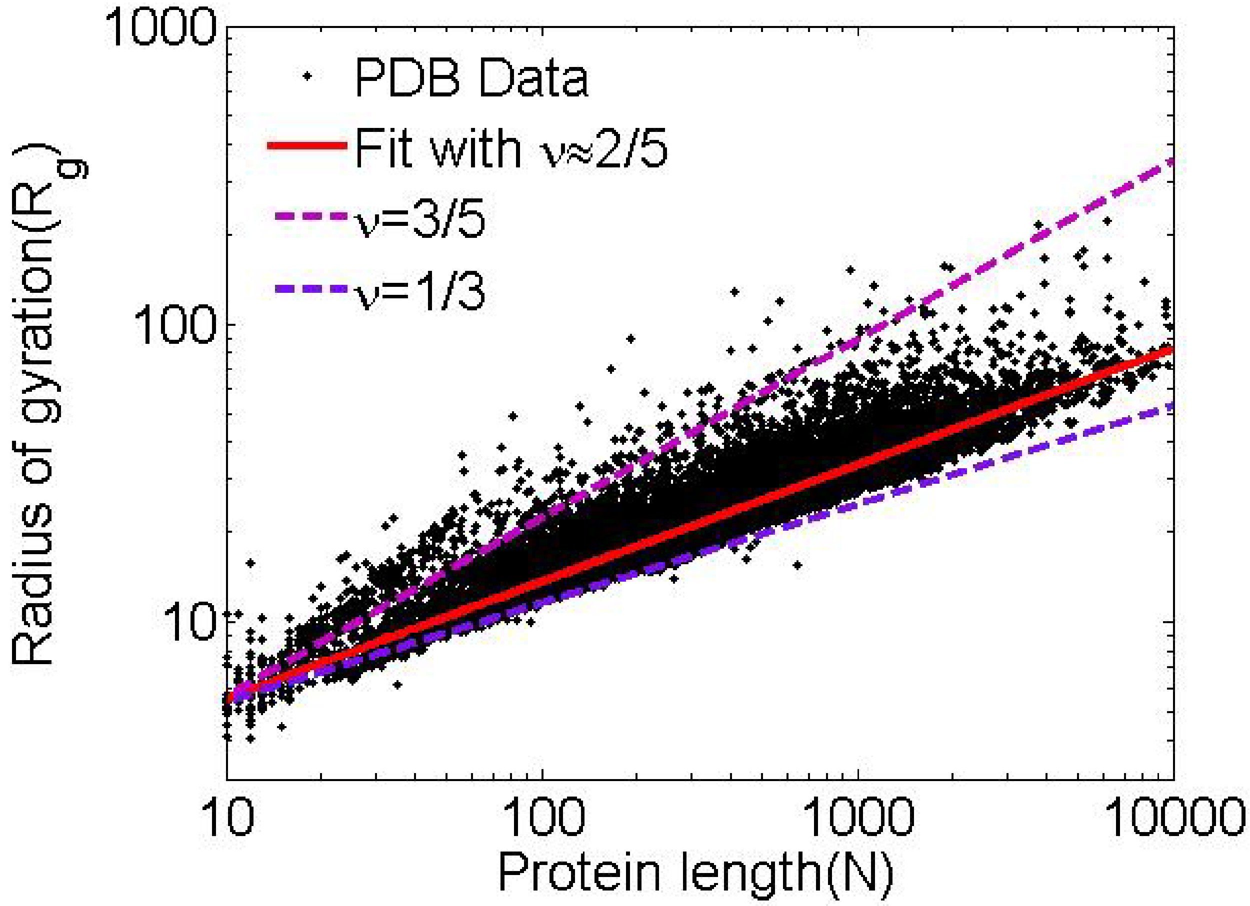


FIG. 1: Log-log plot of the radius of gyration vs. protein length for 37162 proteins in PDB. The red solid line is given by least-square linear fitting method with slope 0.3915 ± 0.0008. Two purple dashed lines are with slope $3/5$ and $1/3$ respectively.

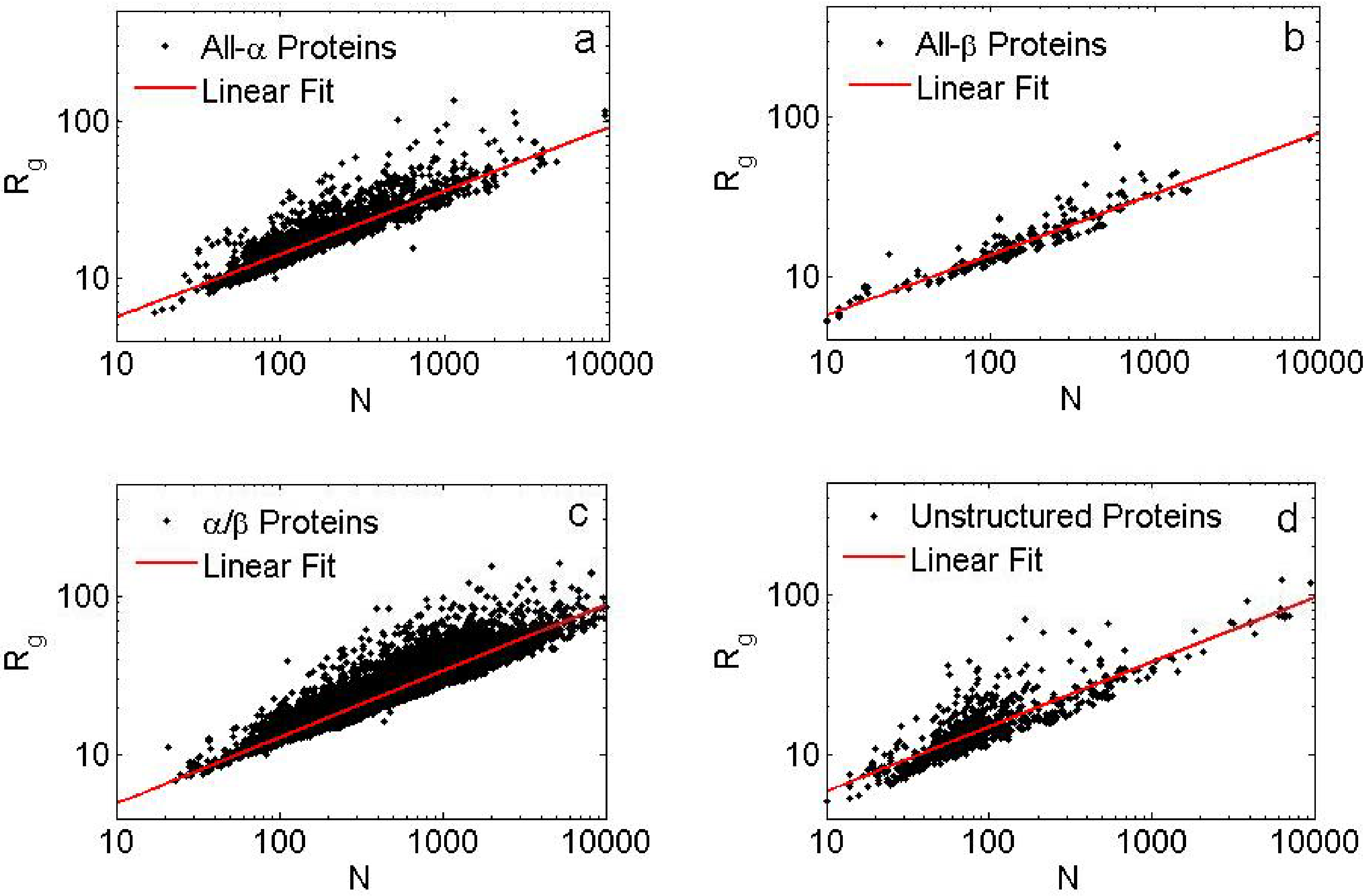
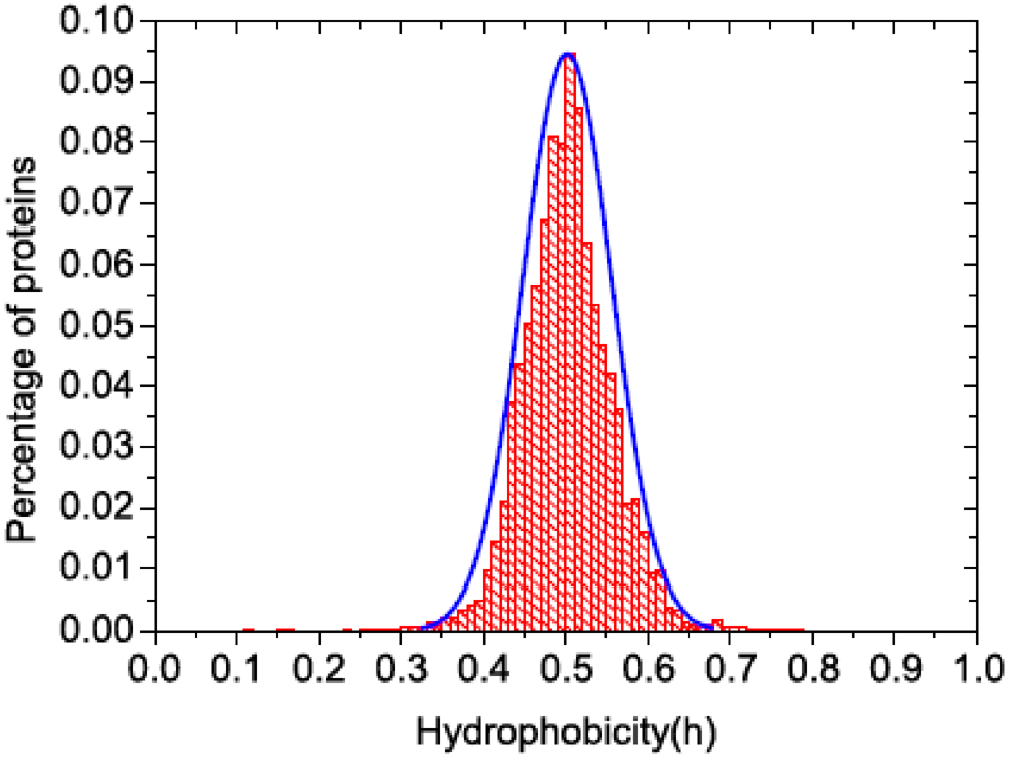


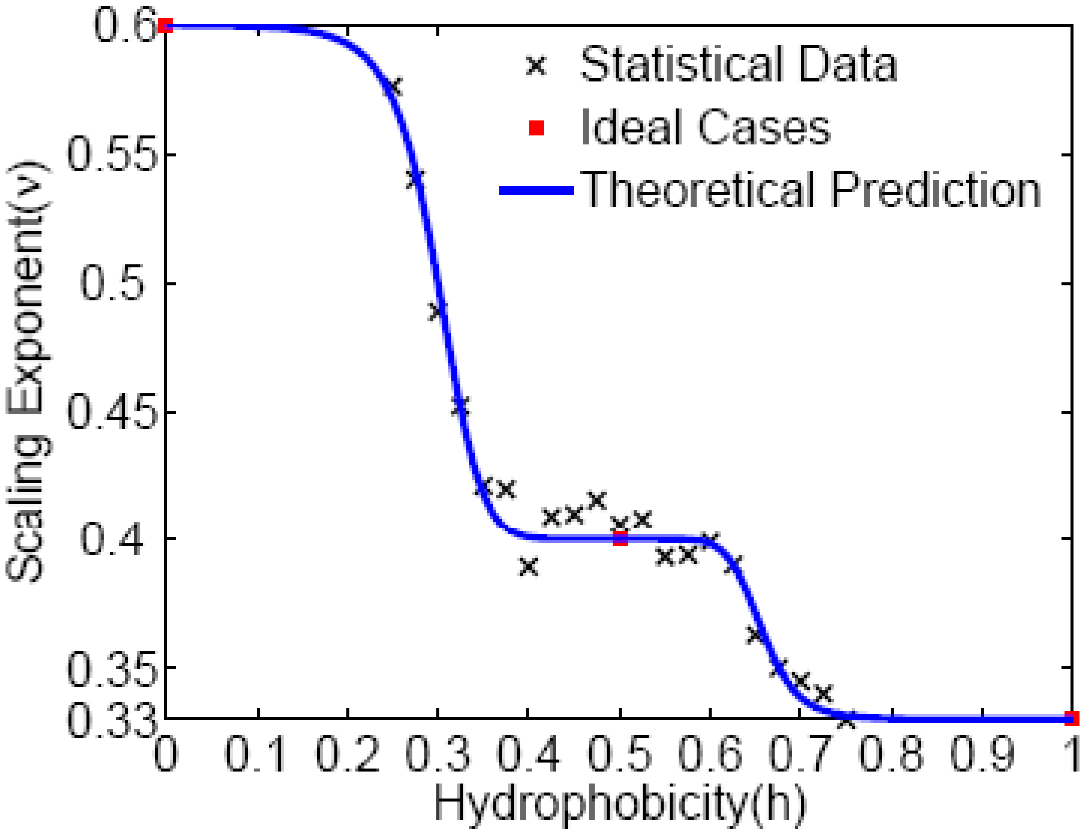
FIG. 2: Log-log plot of the radius of gyration vs. chain length for proteins with different folded types. (A) all-α proteins, (B) all-β proteins, (C) α/β proteins, (D) unstructured proteins. There are 3080 all-α proteins (single α structures are excluded), 334 all-β proteins, and 25804 α/β proteins, with more than 50% amino acids in secondary structures. And there are 839 intrinsically unstructured proteins with less than 20% amino acids in secondary structures. The data are fitted by lines with slope 0.4026 ± 0.0036 for all-α proteins, 0.3838 ± 0.0075 for all-β proteins, 0.4166 ± 0.0010 for α/β proteins, and 0.4037 ± 0.0065 for unstructured proteins.

Hydrophobicity of Proteins

- The hydrophobicity (fraction of hydrophobic residues) of natural proteins exhibits a Gaussian distribution N(0.5, 0.054).



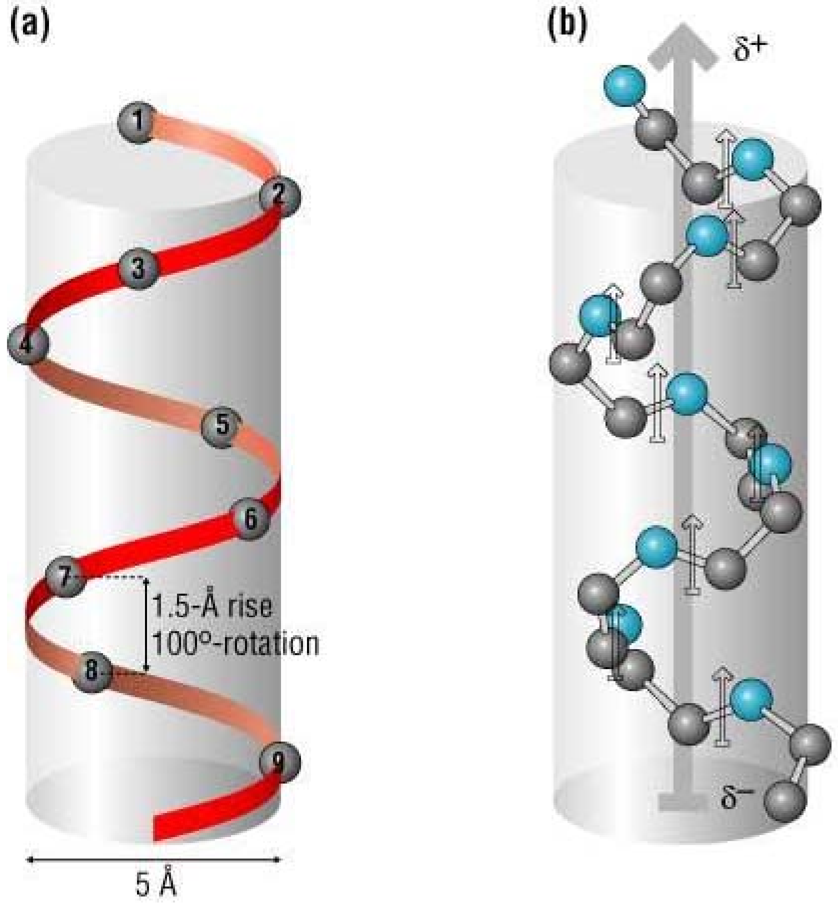
Influence of Hydrophobicity



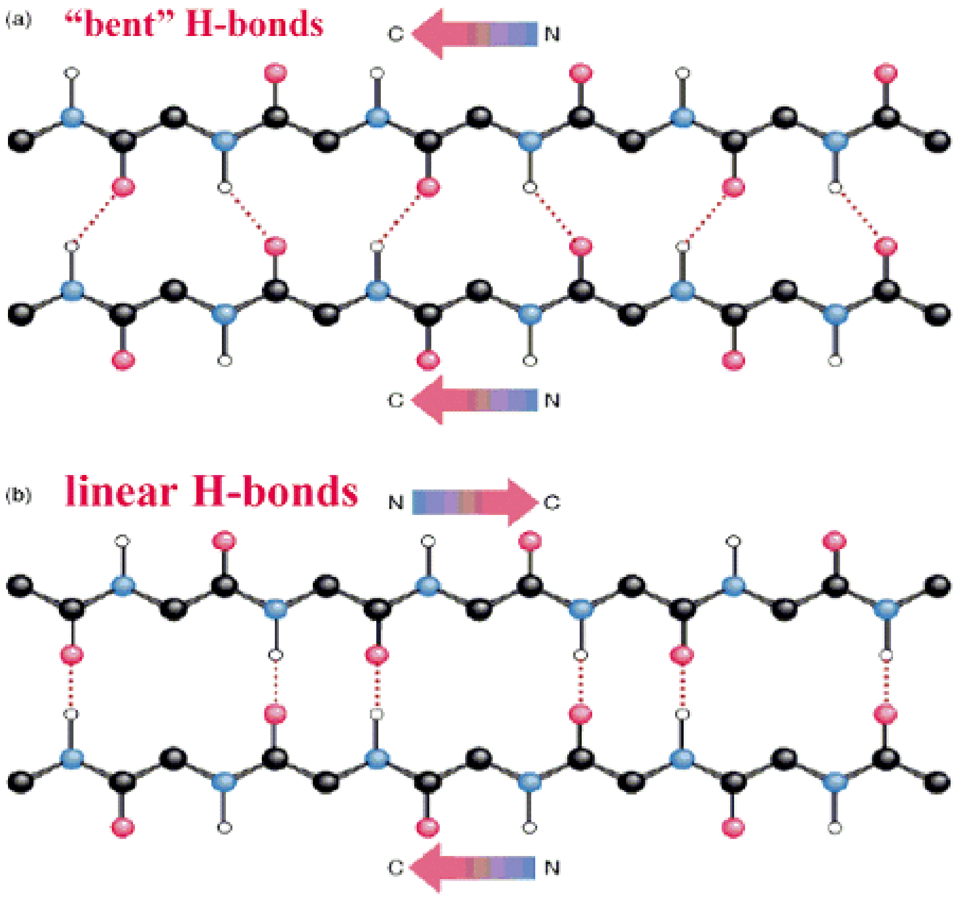
L. Hong & J. Z. Lei. Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. *J. Polym. Sci. B, Polym. Phys., 47:207–214, 2008.*

2. Secondary Structures of Proteins

α-helix



β-sheet



Theory for Helix-Sheet-Coil Transition

- Helix-coil transition theory

  Zimm-Bragg model (1959)

  Lifson-Roig model (1961)

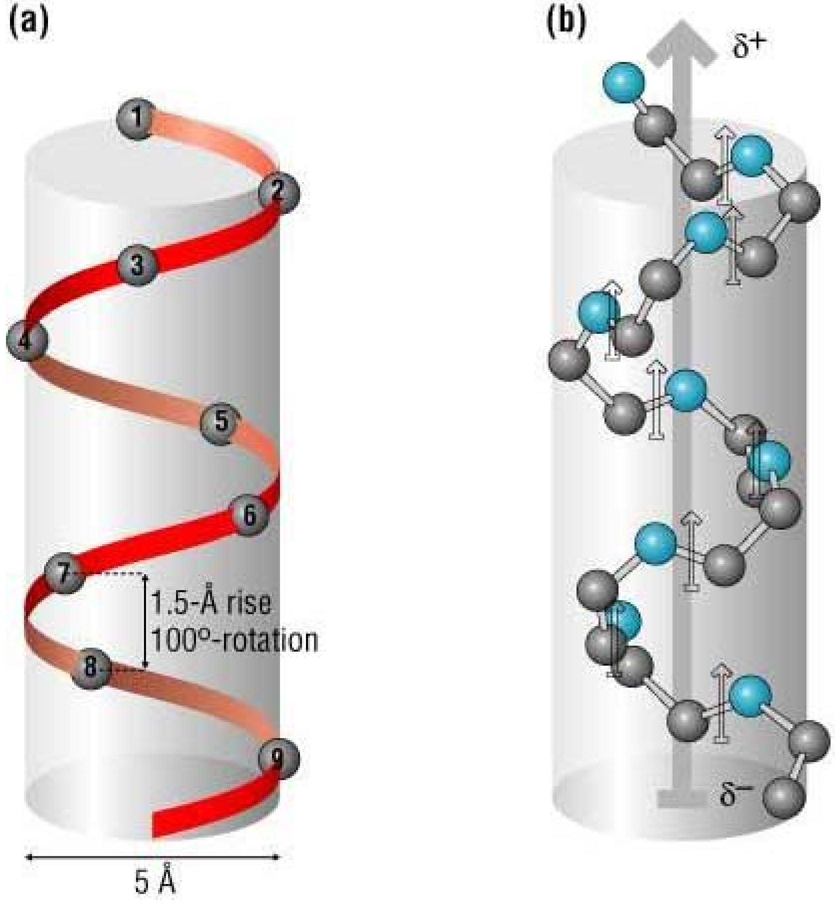
- Statistical models for beta structures

  The antiparallel and parallel β-sheet in proteins

  Helix-sheet-coil transition in homopolypeptide

Helix-Coil Transition Theory

α-helix

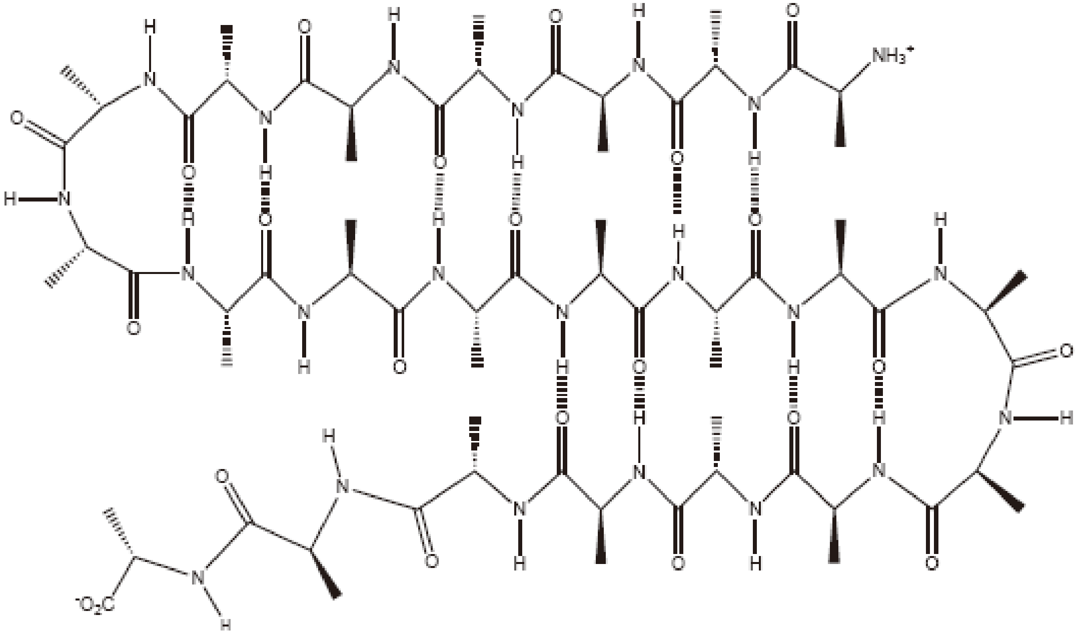


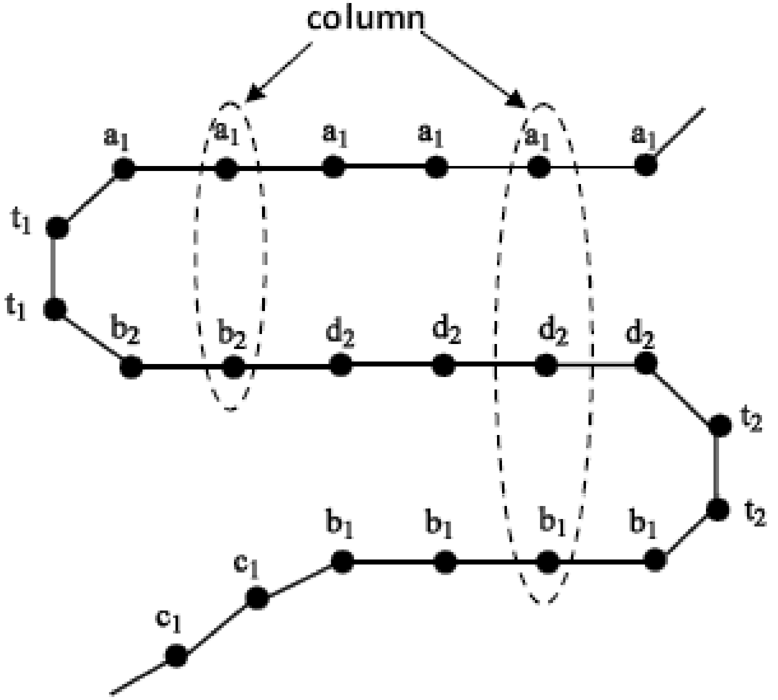
Z-B Model

$$h \; h \; h \; h \; h \; c \; c \; c \quad h \; h \; h \; h \; c \; c \; c$$

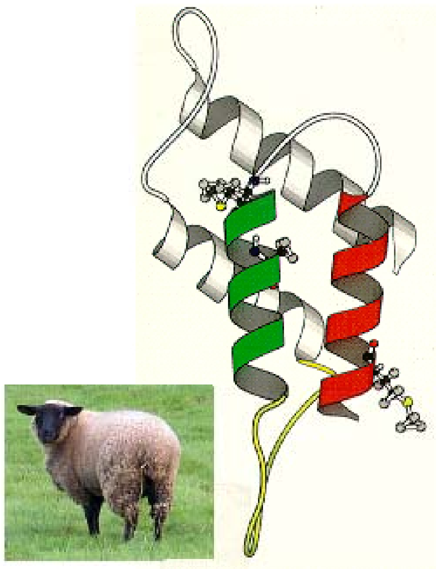
$$\sigma s \; s \; s \; s \; s \; 1 \; 1 \; 1 \; \sigma s \; s \; s \; s \; 1 \; 1 \; 1$$

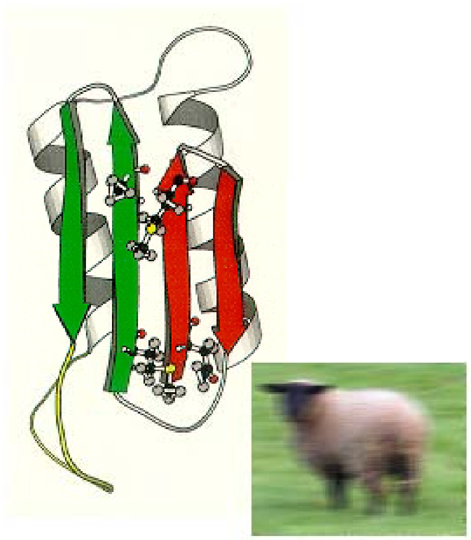
Antiparallel Beta-Sheet/Coil Transition

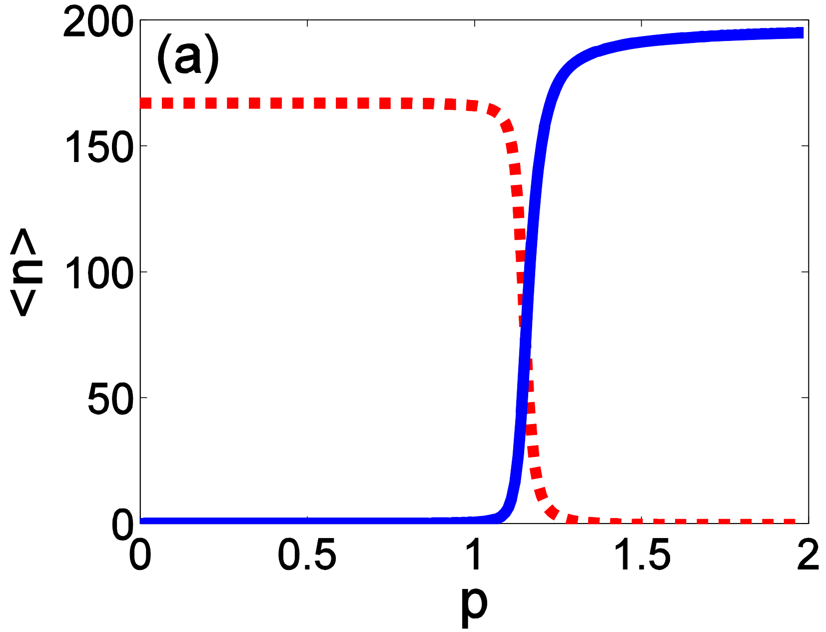




Helix-Sheet Transition

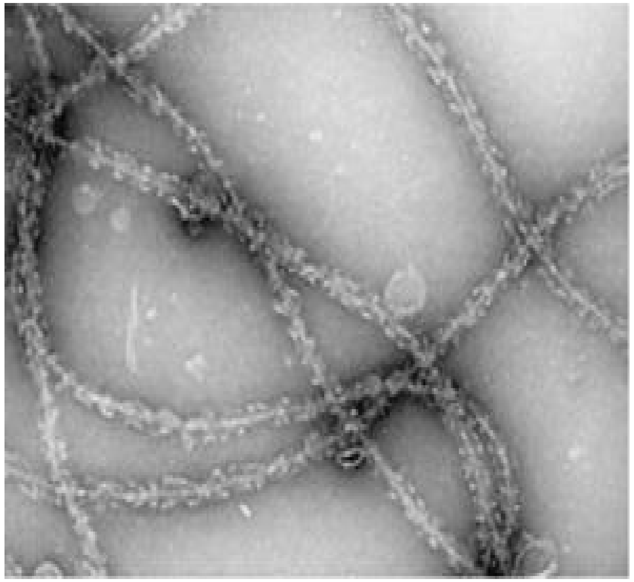

Prion normal : PrP


Prion anormal : PrPsc

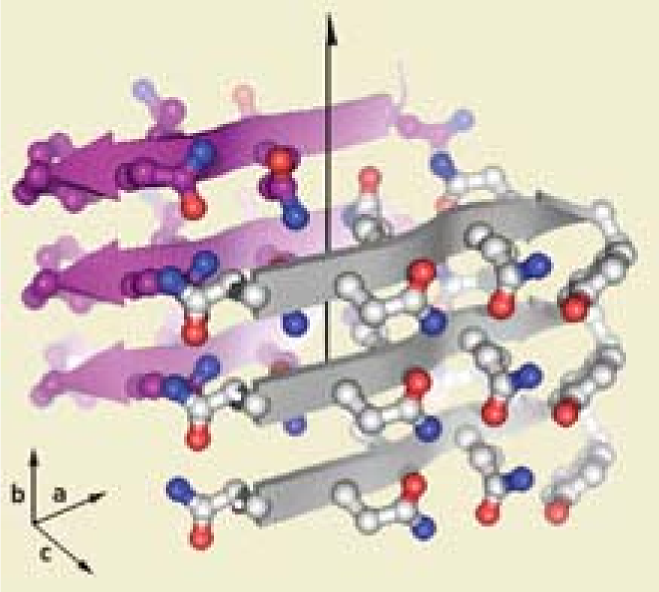


L. Hong & J. Z. Lei. Statistical mechanical model for helix-sheet-coil transitions in homopolypeptides. *Phys. Rev. E, 78:051904, 2008.*

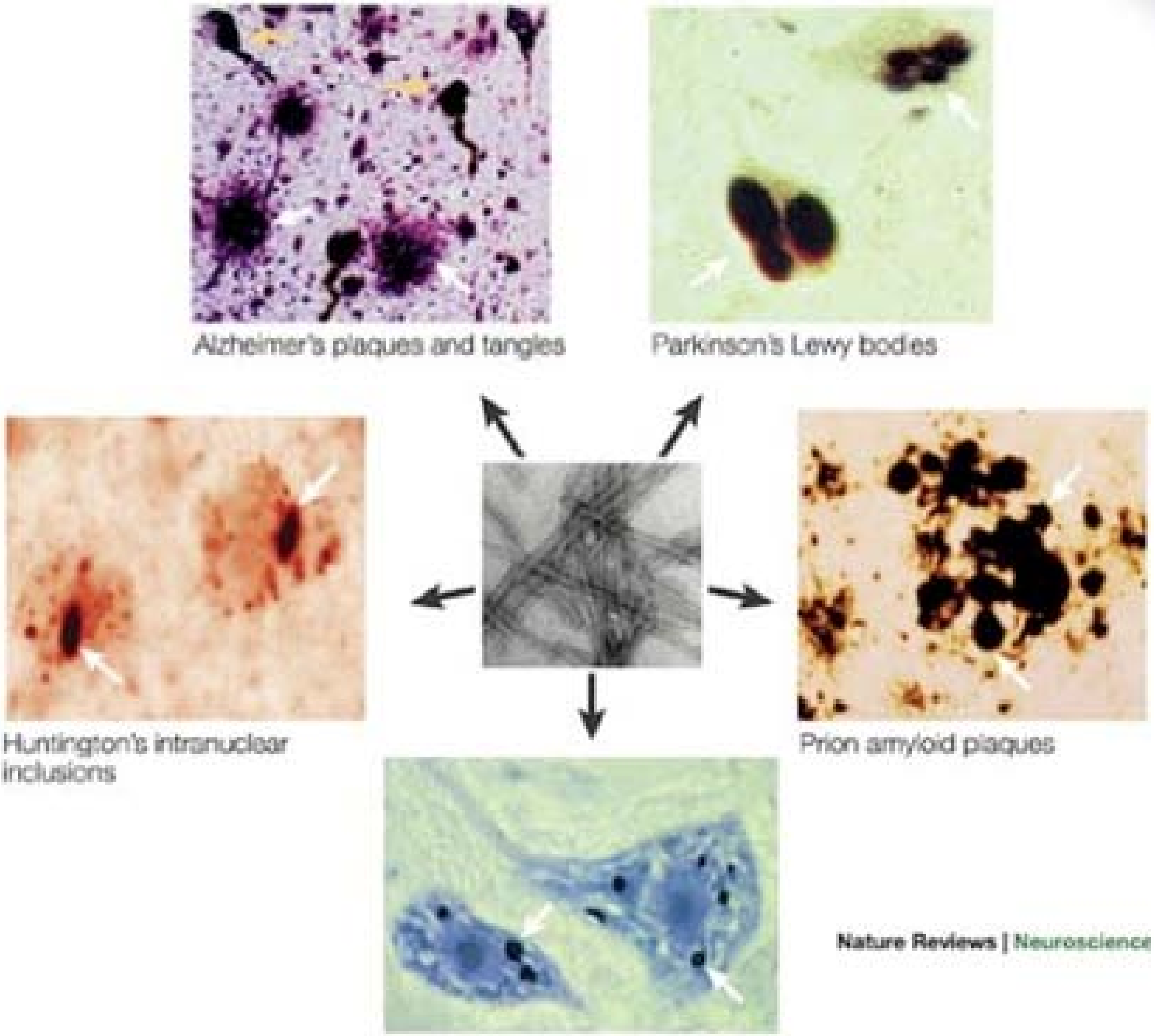
3. Protein Misfolding & Amyloid Fiber



Yeast prion Sup35,
Shorter Lab



Eisenberg, Nature, 2005



Alzheimer's plaques and tangles

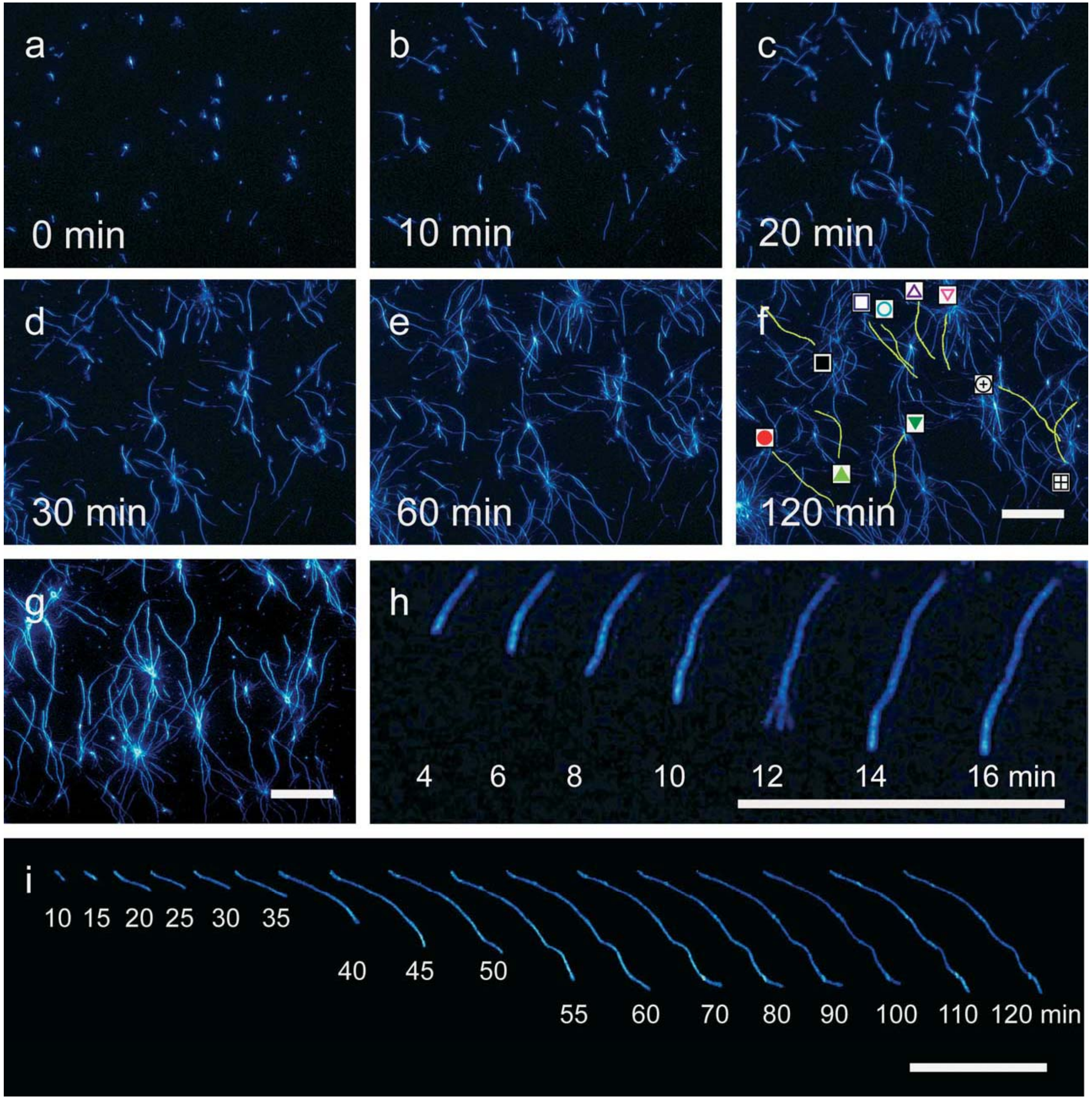
Parkinson's Lewy bodies

Huntington's intranuclear inclusions

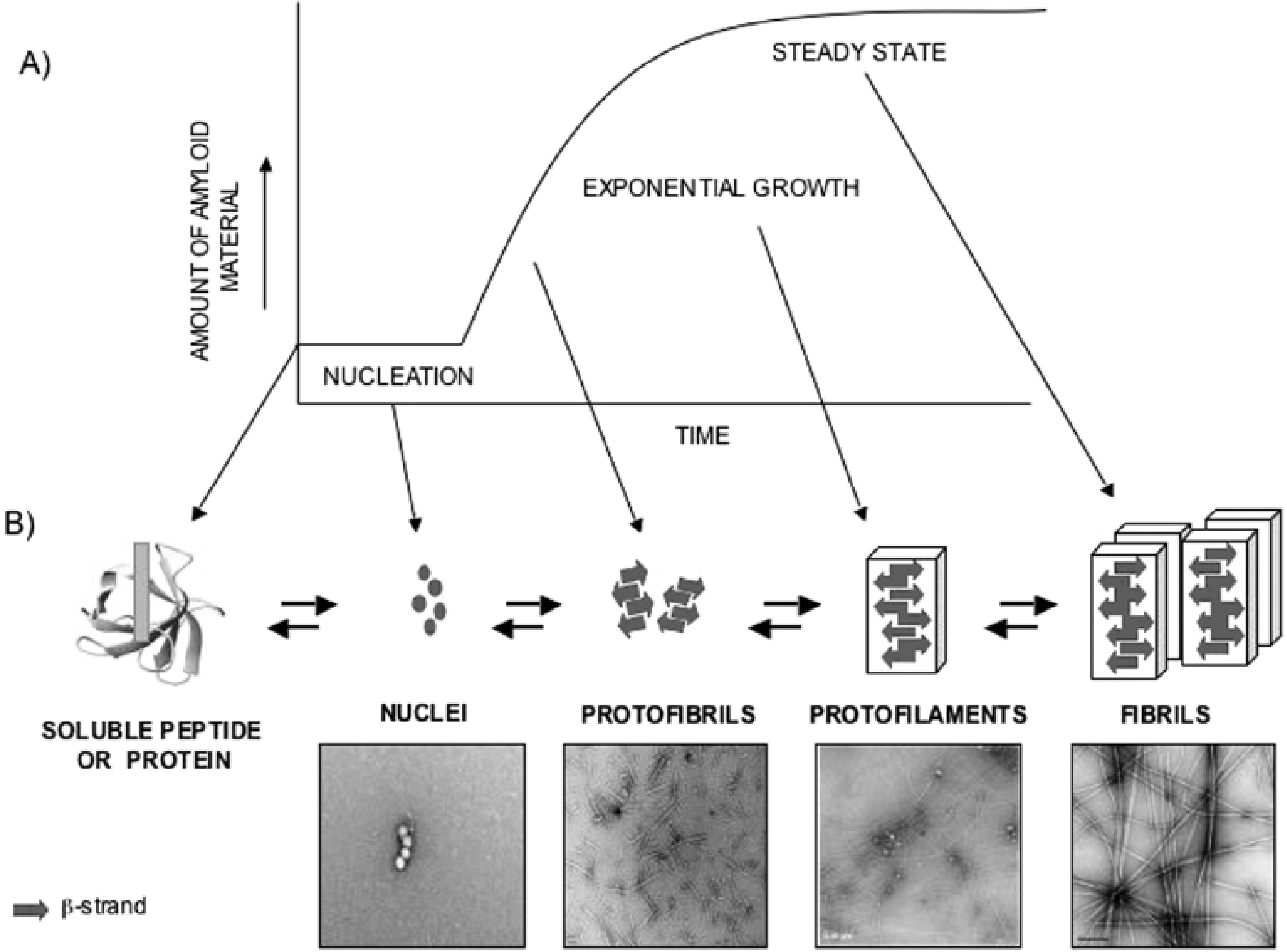
Prion amyloid plaques

Amyotrophic lateral sclerosis aggregates

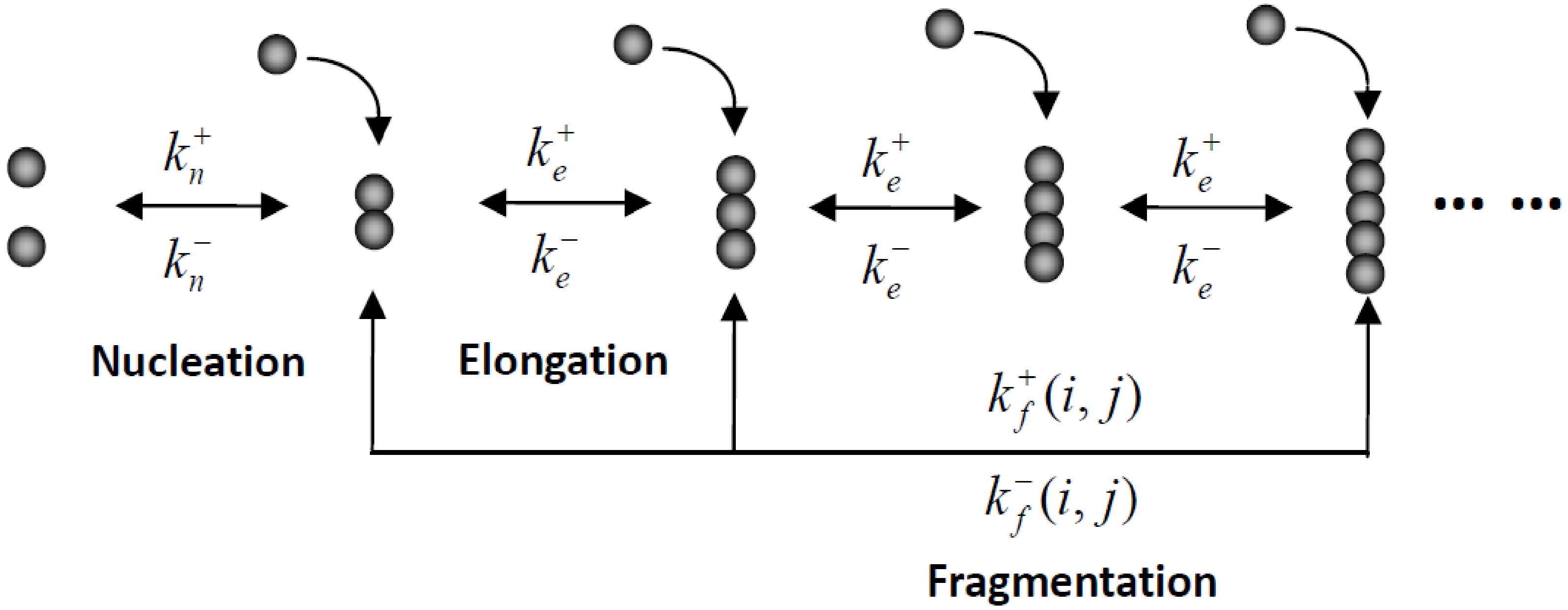
Nature Reviews | Neuroscience



Direct observation of Aβ(1-40) amyloid fibril growth by TIRFM. By Ban *et al.* JMB, 344, 2004.


A)
AMOUNT OF AMYLOID MATERIAL
STEADY STATE
EXPONENTIAL GROWTH
NUCLEATION
TIME
B)
SOLUBLE PEPTIDE OR PROTEIN
NUCLEI
PROTOFIBRILS
PROTOFILAMENTS
FIBRILS
β-strand

NEF Model



$$n_c A_1 \underset{k_{no}^-}{\overset{k_{no}^+}{\rightleftharpoons}} A_{n_c},$$

$$A_i + A_1 \underset{k_e^-}{\overset{k_e^+}{\rightleftharpoons}} A_{i+1}, \qquad (i \ge n_c)$$

$$A_{i+j} \underset{k_f^-(i,j)}{\overset{k_f^+(i,j)}{\rightleftharpoons}} A_i + A_j. \qquad (i, j \ge n_c)$$

Chemical Mass-Action Equations

$$\begin{cases} \dfrac{d}{dt}[A_1] = -n_c k_{no}^{+}[A_1]^{n_c} - 2k_e^{+}[A_1]\sum_{i=n_c}^{\infty}[A_i] + n_c k_{no}^{-}[A_{n_c}] + 2k_e^{-}\sum_{i=n_c+1}^{\infty}[A_i] \\ \dfrac{d}{dt}[A_i] = 2k_e^{+}[A_1]([A_{i-1}]-[A_i]) - 2k_e^{-}([A_i]-[A_{i+1}]) + 2\sum_{j=n_c+i}^{\infty} k_f^{+}(i,j-i)[A_j] \\ \qquad -\sum_{j=n_c}^{i-n_c} k_f^{+}(j,i-j)[A_i] - 2\sum_{j=n_c}^{\infty} k_f^{-}(i,j)[A_i][A_j] + \sum_{j=n_c}^{i-n_c} k_f^{-}(j,i-j)[A_j][A_{i-j}] \\ \qquad +\left(k_{no}^{+}[A_1]^{n_c} - k_{no}^{-}[A_i] - 2k_e^{+}[A_1][A_{i-1}] + 2k_e^{-}[A_i]\right)\delta_{i,n_c}, (i \ge n_c) \end{cases}$$

where $\left[A_i\right]$ is the fiber length distribution.

Macro-Equations

- Number concentration of filaments $P \equiv \sum_{i=n_c}^{\infty} [A_i]$
- Mass concentration of filaments $M \equiv \sum_{i=n_c}^{\infty} i \cdot [A_i]$

$$\begin{cases} \dfrac{dP}{dt} = k_n^+ \left(m_{tot} - M\right)^{n_c} - k_n^- \left[A_{n_c}\right] + \sum\limits_{i=n_c}^{\infty} \sum\limits_{j=i+n_c}^{\infty} k_f^+ \left(i, j-i\right) \left[A_j\right] \\ \qquad\qquad - \sum\limits_{i=n_c}^{\infty} \sum\limits_{j=n_c}^{\infty} k_f^- \left(i, j\right) \left[A_i\right] \left[A_j\right], \\ \dfrac{dM}{dt} = n_c k_n^+ \left(m_{tot} - M\right)^{n_c} + 2k_e^+ \left(m_{tot} - M\right) P - 2k_e^- P - \left(n_c k_n^- - 2k_e^-\right) \left[A_{n_c}\right], \end{cases}$$

- How to make above equations closed?

Moment-Closure Method

- The minimization of the free energy under certain constraints

min $F\left(\left[A_i\right]\right)$

s.t. $\sum_{i=n_c}^{\infty}\left[A_i\right]=P,\ \sum_{i=n_c}^{\infty} i\cdot\left[A_i\right]=M,\ \left[A_1\right]+\sum_{i=n_c}^{\infty} i\cdot\left[A_i\right]=m_{tot}.$

in which the free energy is

$$F=\varepsilon_n\sum_{i=n_c}^{\infty}\left[A_i\right]+\varepsilon_e\sum_{i=n_c}^{\infty}\left(i-n_c\right)\cdot\left[A_i\right]$$

$$+k_BT\left[\left(\left[A_1\right]\ln\left[A_1\right]-\left[A_1\right]\right)+\sum_{i=n_c}^{\infty}\left(\left[A_i\right]\ln\left[A_i\right]-\left[A_i\right]\right)\right].$$

Moment-Closure Equations

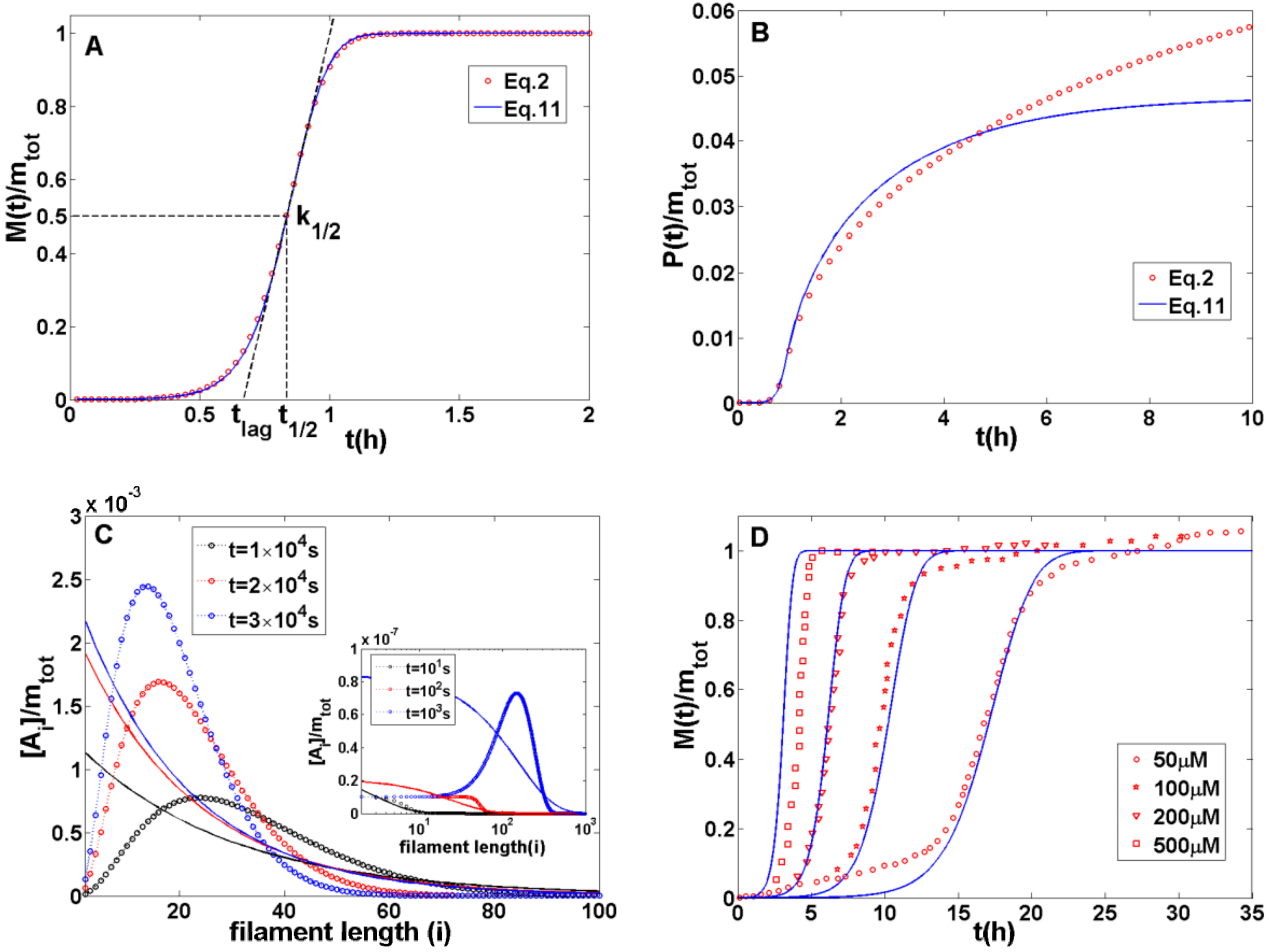
- Moment-closure equations

$$\begin{cases} \dfrac{dP}{dt} = k_n^+ \left(m_{tot} - M\right)^{n_c} - k_n^- \left(1-\theta\right) P + k_f^+ \left(1-\theta\right) P \theta^{n_c} \Xi_{n-1} - k_f^- \left(1-\theta\right)^2 P^2 \Lambda_1, \\ \dfrac{dM}{dt} = n_c k_n^+ \left(m_{tot} - M\right)^{n_c} + 2k_e^+ \left(m_{tot} - M\right) P - 2k_e^- P - \left(n_c k_n^- - 2k_e^-\right)\left(1-\theta\right) P, \end{cases}$$

where $\theta \equiv \left(M - n_c P\right) \big/ \left[M - \left(n_c - 1\right) P\right] \in [0,1)$

$$\Xi_{n-1} = \frac{n_c^n}{2^n}\left[\frac{\ln n_c}{n_c^2(-\ln\theta)^2} + \frac{(n-2)\ln n_c + 1}{n_c^3(-\ln\theta)^3} + \frac{(3n-5)(n-4)\ln n_c + 6(n-3)}{4n_c^4(-\ln\theta)^4} + \cdots\right]$$

$$\Lambda_1 = \frac{\ln n_c}{n_c^2(-\ln\theta)^2}$$



L. Hong & W. A. Yong. Simple Moment-Closure Model for the Self-Assembly of Breakable Amyloid Filaments. *Biophys. J., 104:533-540, 2013.*

Classification of amyloid proteins

$$\boxed{t_{1/2} \propto m_{tot}^{-\mu}, k_{app} \propto m_{tot}^{\nu}}$$

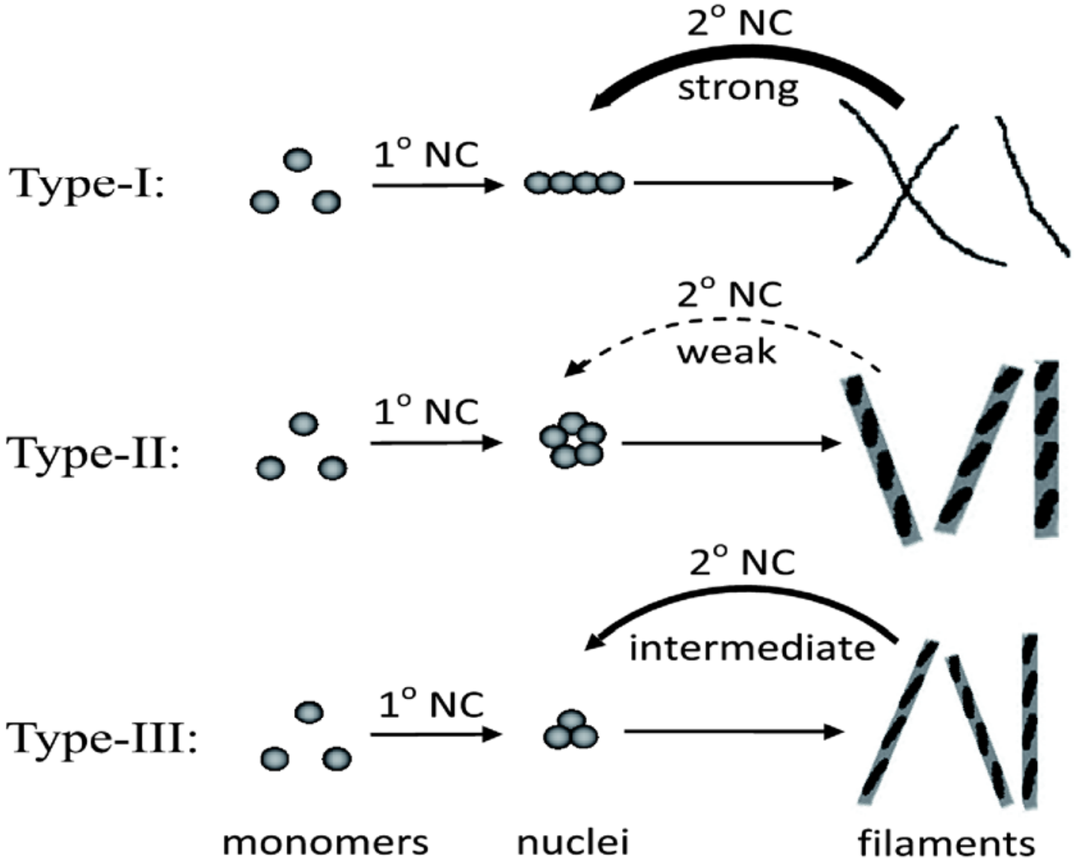
(1) $\mu = \nu = \frac{1}{2}$ independent of nucleus size;

Yeast prion Sup35 NW region, Csg $\mathrm{B}_{\mathrm{trunc}}$,Ure2, β2-icroglobulin, Stefin B, α-synuclein, WW domain, insulin.

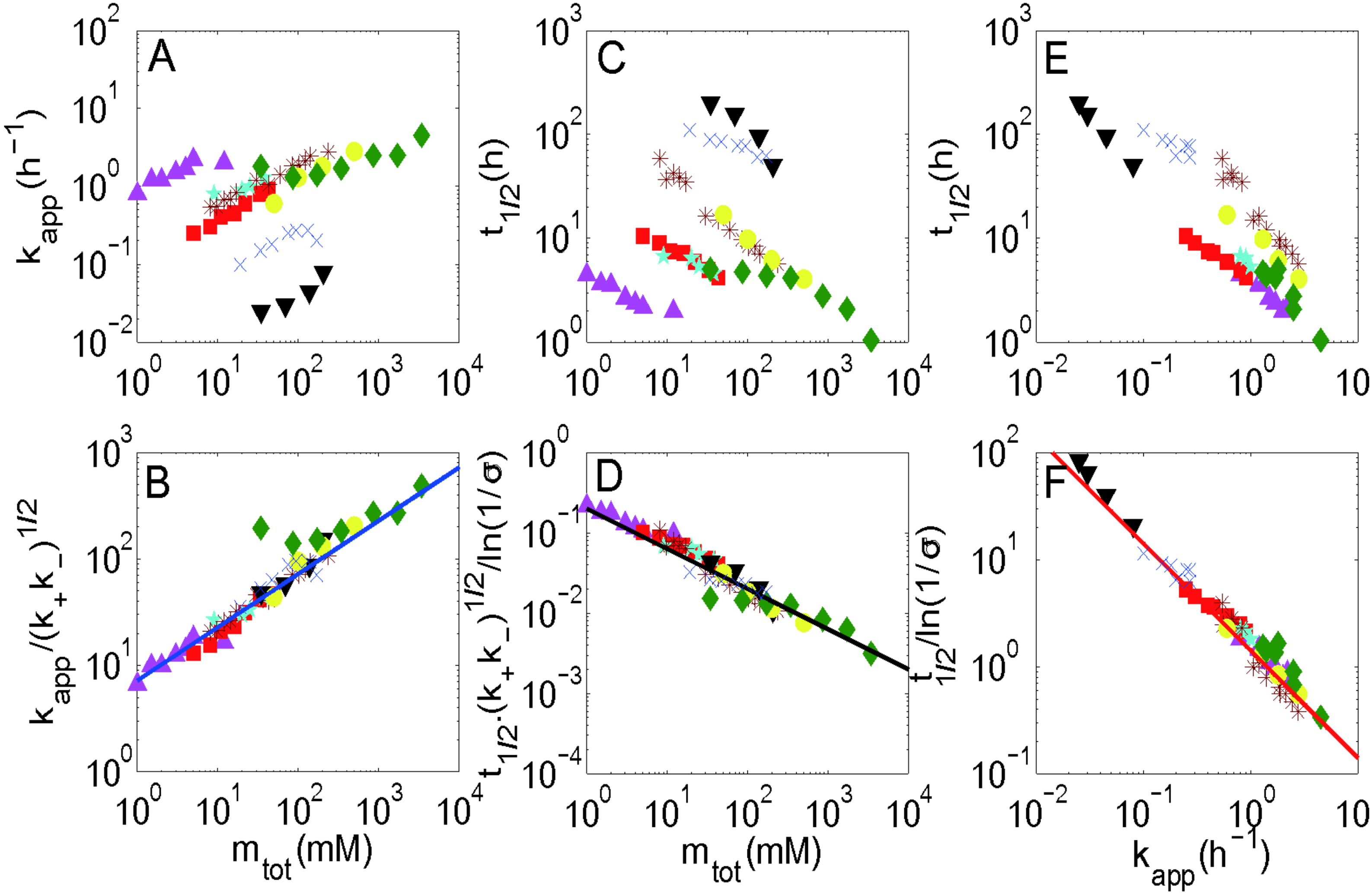
(2) μ or ν depends on nucleus size;

Aβ (M1-42), γC-crystallin, Apo C-II, TI I27.

Classification of amyloid proteins

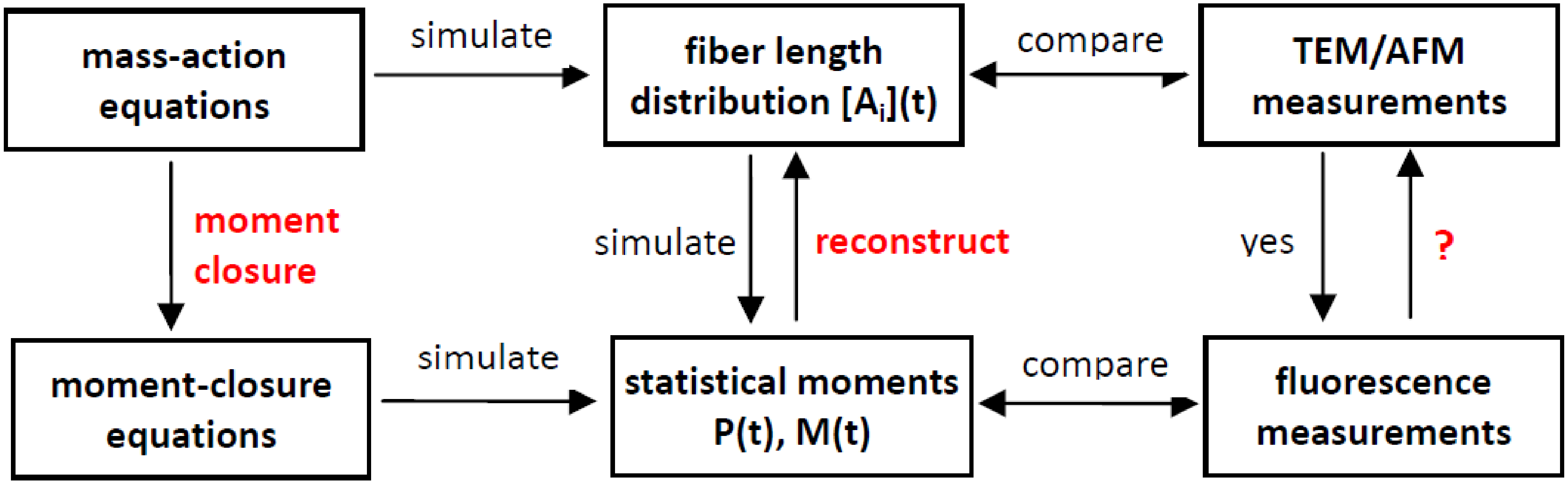


L. Hong *et al*. Dissecting the Kinetic Process of Amyloid Fiber Formation through Asymptotic Analysis. *JPCB, 116:6611-17, 2012.*



the yeast prion Sup35 NW region (purple triangles up), Csg B_{trunc} (red squares), Ure2 protein (cyan pentacles), β2-microglobulin (brown stars), stefin B (blue cross), α-synucleins (black triangles down), WW domain (yellow circles), and insulin (green diamonds)

From Macro to Micro



mass-action equations

simulate

fiber length distribution $[A_i](t)$

compare

TEM/AFM measurements

moment closure

simulate

reconstruct

yes

?

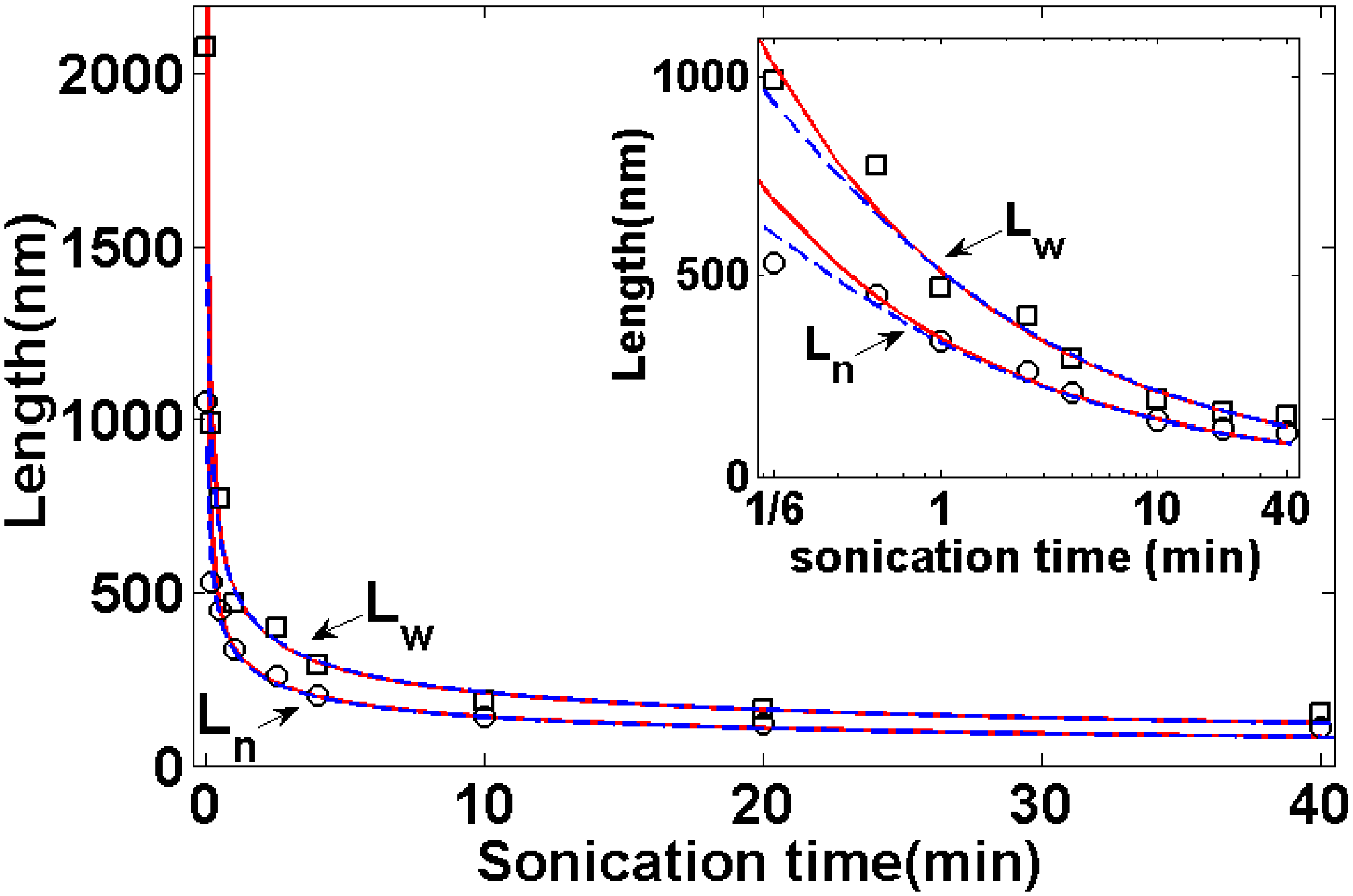
moment-closure equations

simulate

statistical moments $P(t)$, $M(t)$

compare

fluorescence measurements


Length(nm)
0
500
1000
1500
2000
Sonication time(min)
0
10
20
30
40
L_w
L_n
Length(nm)
0
500
1000
sonication time (min)
1/6
1
10
40
L_w
L_n

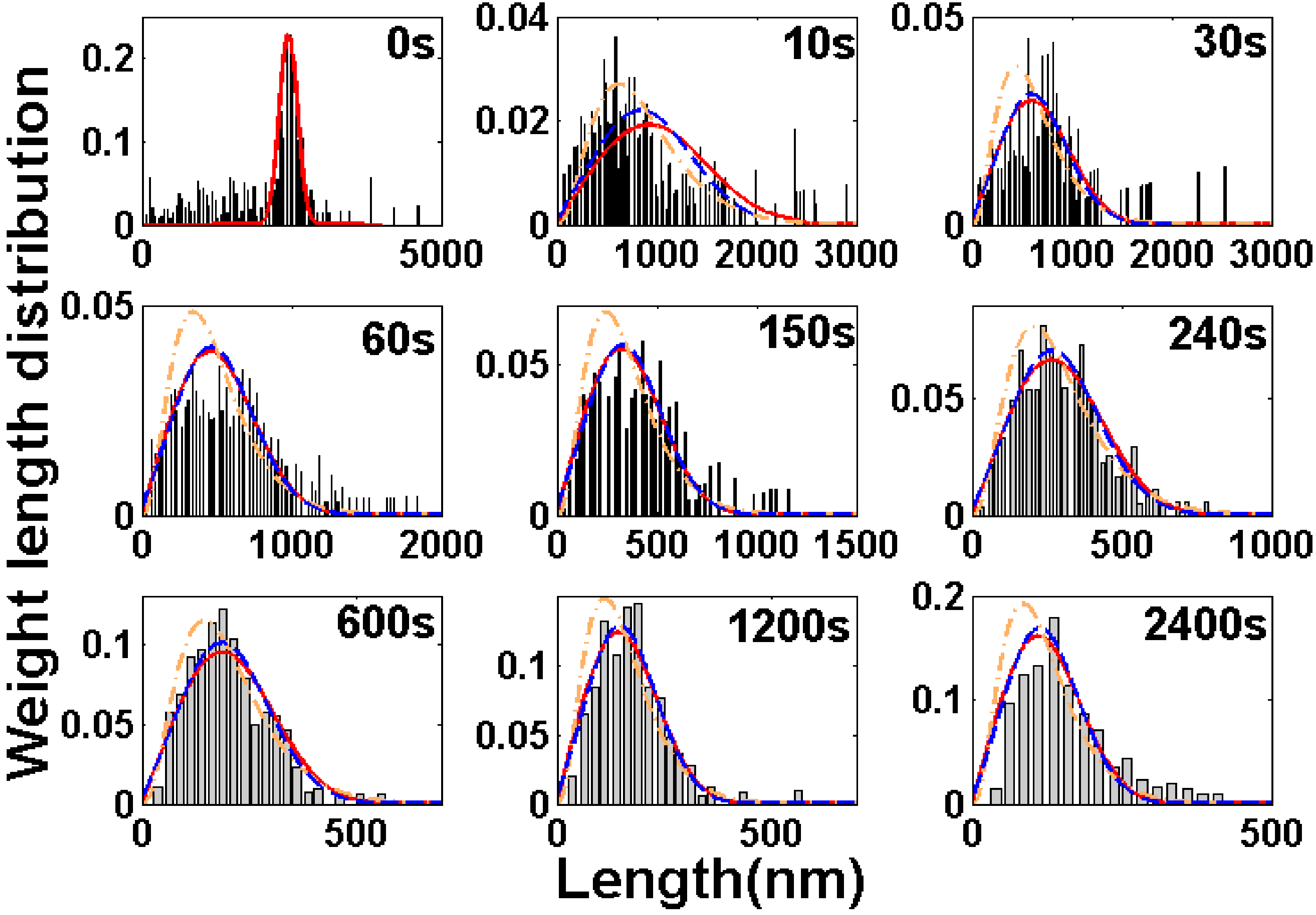
Construct Fiber Length Distribution

- We assume the fiber length distribution satisfies following form

$$[A_i] = \beta i^{x-3} \exp\left[-(i-i_0)^2 / \sigma^2\right],$$

in which the three unknown parameters are determined through normalization conditions

$$\begin{cases} \sum_{i=1}^{\infty} [A_i] = P, \\ \sum_{i=1}^{\infty} i \cdot [A_i] = m_{tot}, \\ \sum_{i=1}^{\infty} i^2 \cdot [A_i] = m_{tot} L_w \end{cases}$$



P.Z. Tan & L. Hong. Modeling Fibril Fragmentation in Real-Time. *arXiv:1212.5865, 2013.*

Thank You!