



Fractional Gaussian Noise, Subdiffusion and Stochastic Networks in Biophysics

Samuel Kou
Department of Statistics
Harvard University

Single-Molecule Experiments

- Statistics & probability experienced fundamental change in the past 20 years
- Biophysics & chemistry also witnessed dramatic progress: **single-molecule experiments**
- Using nanotechnology, scientists can study biological processes on a single-molecule basis (eg. enzymatic kinetics, protein/DNA dynamics)

“Seeing images of single atoms is a religious experience”

--- Richard Feynman

New Aspects for Scientific Discovery

- Can measure molecular properties ***individually***, instead of inferring from population statistics
- If the reaction/kinetic time is slow, ensemble experiments become almost impossible due to the ***difficulty of synchronization***
- Single-molecule trajectory provides detailed ***dynamic information***
- Understanding the dynamics of individual is essential to unlock their biofunctions



Statistical & Probabilistic Challenges

- Require new stochastic modeling

Subdiffusion

Enzymatic reaction

- Data are noisier, and efficient inference methodology is needed

Brownian diffusion

- Since Einstein's 1905 paper, the theory of Brownian diffusion has revolutionized not only natural sciences but also social sciences.
- Brownian motion described by Langevin equation

$$m \frac{dv_t}{dt} = -\zeta v_t + F(t)$$

where $F(t)$ is white noise process satisfying

$$E\{F(t)F(t')\} = \zeta k_B T \cdot \delta(t - t')$$

by **fluctuation-dissipation** theorem.

- Solution: Ornstein-Uhlenbeck process $v(t)$ Gaussian
- Displacement (location) $x(t) = \int_0^t v(s) ds$

$$E\{x(t)^2\} \sim 2 \frac{k_B T}{\zeta} t, \text{ for large } t$$

--- Corner stone for statistical mechanics

Subdiffusion

- Brownian diffusion, however, cannot explain so called subdiffusion:

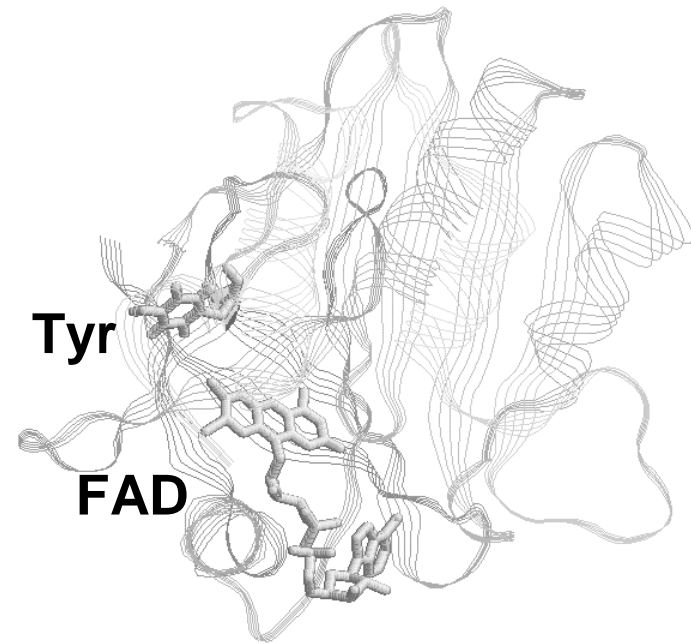
$$E\{x(t)^2\} \propto t^\alpha, \quad 0 < \alpha < 1$$

- Distance fluctuation within a single protein molecule (Yang *et al.* 2003, *Science*; Min *et al.* 2005, *Physical Review Letters*).
- Need tools beyond Langevin equation and BM.

Single-molecule fluorescence experiment on protein complex

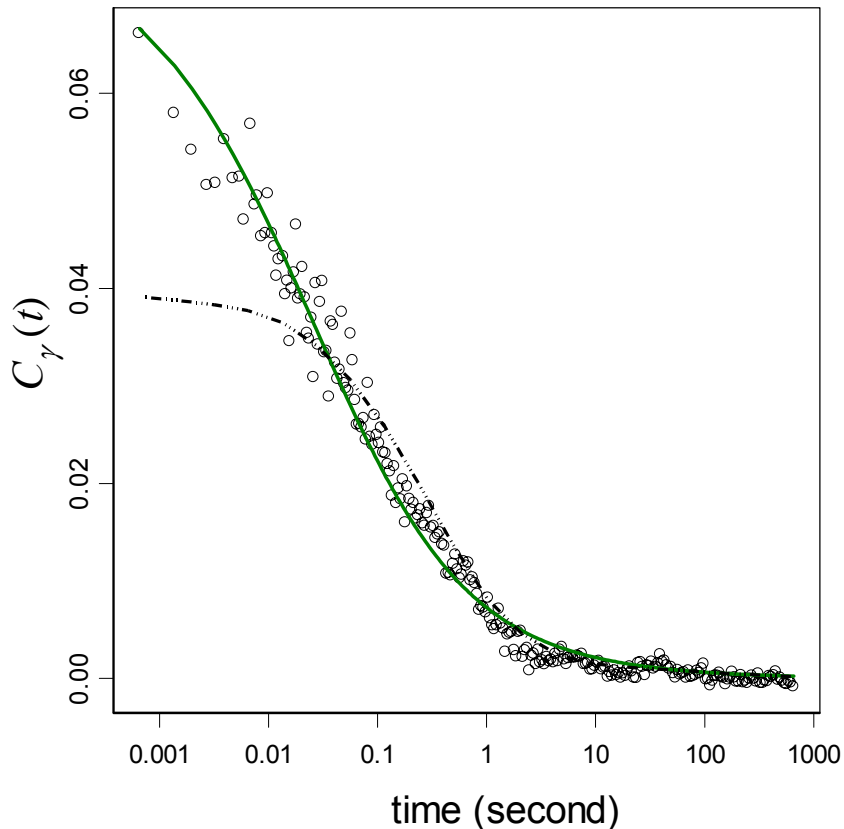
- Yang et al. (2003, *Science*) studied a protein-enzyme complex **Fre**, catalyzes the reduction of flavin
- Fre contains two substructures: a flavin adenine dinucleotide (FAD) and a tyrosine (Tyr).
- Fluorescence lifetime of FAD varies due to **distance fluctuation**
- Relationship between fluorescence lifetime and distance

$$\gamma^{-1}(t) = [k_0 e^{-\beta(X_{eq} + X_t)}]^{-1}$$



Autocorrelation function $E\{\Delta\gamma^{-1}(0)\Delta\gamma^{-1}(t)\}$

$$\Delta\gamma^{-1}(t) = \gamma^{-1}(t) - E\{\gamma^{-1}(t)\}$$



- Need tools beyond Langevin equation and BM.
- The model: Generalized Langevin equation with fractional Gaussian noise (GLE with fGn).

Generalized Langevin Equation with fGn

- Langevin equation $m \frac{dv_t}{dt} = -\zeta v_t + F(t)$
- Generalized Langevin equation $m \frac{dv_t}{dt} = -\zeta \int_{-\infty}^t v_u K(t-u) du + G_t,$
- **Fluctuation-dissipation theorem** links memory kernel $K(t)$ with fluctuating force $E\{G_t G_s\} = \zeta k_B T \cdot K(t-s)$
- Key question: How to introduce the noise structure?
- Understand the white noise: White noise is the derivative of the Wiener process $F_t = \frac{d}{dt} B_t$
- $B(t)$ is the unique process: (i) Gaussian, (ii) independent increment, (iii) stationary increment, (iv) self-similar

- Natural generalization: (i) Gaussian (ii) stationary increment (iii) self-similar.

- The **ONLY** candidate fBM $B^{(H)}(t)$, $0 < H < 1$

$E\{B_t^{(H)}\} = 0$, and covariance function

$$E\{B_t^{(H)} B_s^{(H)}\} = (|t|^{2H} + |s|^{2H} - |t-s|^{2H}) / 2 \quad \text{for } t, s \geq 0.$$

when $H = 1/2$, reduces to $B(t)$.

- Fractional Gaussian noise $F^{(H)}(t) = \frac{dB_t^{(H)}}{dt}$:
Gaussian & stationary.

- Memory kernel

$$K_H(t) = E\{F^{(H)}(0)F^{(H)}(t)\}$$

$$= H(2H-1)|t|^{2H-2}, \text{ for } t \neq 0$$

- Spectral density

$$\tilde{K}_H(\omega) = \int_{-\infty}^{\infty} e^{it\omega} K_H(t) dt = \Gamma(2H+1) \sin(H\pi) |\omega|^{1-2H}$$

Toward subdiffusion

- Applying Fourier transform on

$$m \frac{dv_t}{dt} = -\zeta \int_{-\infty}^t v_u K_H(t-u) du + F_t^{(H)},$$

$v(t)$ Gaussian $E\{v(t)\} = 0$,

$$C(t) = E\{v(0)v(t)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{it\varpi} \tilde{C}(\varpi) d\varpi$$

$$\tilde{C}(\varpi) = \frac{k_B T \zeta \tilde{K}_H(\varpi)}{\left| \zeta \tilde{K}_H^+(\varpi) - i\varpi m \right|^2}$$

- For displacement $X(t) = \int_0^t v(s) ds$

$$E\{X(t)^2\} = \int_0^t \int_0^t E\{v(s)v(u)\} du ds$$

$$E\{X(t)^2\} \sim \frac{k_B T}{\zeta} \frac{2 \sin(2H\pi)}{\pi H (2H-1)(2H-2)} t^{2-2H} \propto t^{2-2H}$$

- $H > 1/2$ leads to subdiffusion!

Harmonic potential

- GLE: $m \frac{dv_t}{dt} = -\zeta \int_{-\infty}^t v_u K_H(t-u) du + F_t^{(H)},$

- For external field $U(x)$, changes to

$$m\ddot{X}(t) = -\zeta \int_{-\infty}^t \dot{X}(u) K_H(t-u) du - U'(X_t) + F^{(H)}(t)$$

where $X(t) = \int_0^t v(s) ds$, $\dot{X}(t) = \frac{dv(t)}{dt}$.

For harmonic potential $U(x) = \frac{1}{2} m \omega^2 x^2$

$$m\ddot{X}(t) = -\zeta \int_{-\infty}^t \dot{X}(u) K_H(t-u) du - m\omega^2 X(t) + F^{(H)}(t)$$

- Fourier method gives $E\{X(t)X(s)\}, E\{X(t)v(s)\}$

Overdamped condition

- Acceleration negligible, GLE reads

$$m\omega^2 X(t) = -\zeta \int_{-\infty}^t \dot{X}(u) K_H(t-u) du + F^{(H)}(t)$$

$$m\omega^2 X(t) = -\zeta \int_{-\infty}^t \dot{X}(u) K_H(t-u) du + F^{(H)}(t)$$

- Solution: $X(t)$ stationary Gaussian $E\{X(t)\}=0$

$$\begin{aligned} C_{xx}(t) &= E\{X(0)X(t)\} \\ &= \frac{k_B T}{m\omega^2} E_{2-2H} \left(-\frac{m\omega^2}{\zeta \Gamma(2H+1)} t^{2-2H} \right) \end{aligned}$$

where $E_\alpha(z) = \sum_{k=0}^{\infty} z^k / \Gamma(\alpha k + 1)$

- For all H ,
- $$C_{xx}(0) = E\{X(0)X(0)\} = \frac{k_B T}{m\omega^2}$$

the **thermal equilibrium** value.

- For $H = 1/2$, recovers the Brownian diffusion result

$$C_{xx}(t) = E\{X(0)X(t)\} = \frac{k_B T}{m\omega^2} e^{-2\frac{m\omega^2}{\zeta}t}$$

The Hamiltonian for GLE with fGn

GLE with fGn can be derived from the interaction between a particle and a harmonic oscillator heat bath.

Start from system Hamiltonian \Rightarrow

$$H_s = \frac{1}{2}mv^2 + \frac{1}{2}m\omega^2x^2$$
$$= \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2, \quad p = mv$$

and the heat bath Hamiltonian

$$\Rightarrow H_B = \sum_j \left(\frac{p_j^2}{2m_b} + \frac{1}{2}m_b\omega_j^2 \left(q_j - \frac{\gamma_j}{\omega_j^2}x \right)^2 \right)$$

where m_b the media molecule mass, ω_j individual frequency, and γ_j coupling strength of j th oscillator.

Equation of motion for $H_s + H_B$

$$\frac{dx}{dt} = \frac{\partial}{\partial p}(H_s + H_B), \quad \frac{dp}{dt} = -\frac{\partial}{\partial x}(H_s + H_B)$$
$$\frac{dq_j}{dt} = \frac{\partial}{\partial p_j}(H_s + H_B), \quad \frac{dp_j}{dt} = -\frac{\partial}{\partial q_j}(H_s + H_B)$$

- Solve it. Leads to (i) GLE

$$m\ddot{X}(t) = -\zeta \int_{-\infty}^t \dot{X}(u)K(t-u)du - m\omega^2 X(t) + F(t)$$

where

$$K(t) = m_b \sum_j \frac{\gamma_j^2}{\omega_j^2} \cos \omega_j t = m_b \int \frac{\gamma^2(\varpi)}{\varpi^2} \cos(\varpi t) g(\varpi) d\varpi$$

$$F(t) = \sum_j m_b \gamma_j (q_j(0) - \frac{\gamma_j}{\omega_j^2} x(0)) \cos(\omega_j t) + \sum_j \frac{\gamma_j}{\omega_j} p_j(0) \sin(\omega_j t)$$

and (ii) the fluctuation-dissipation theorem

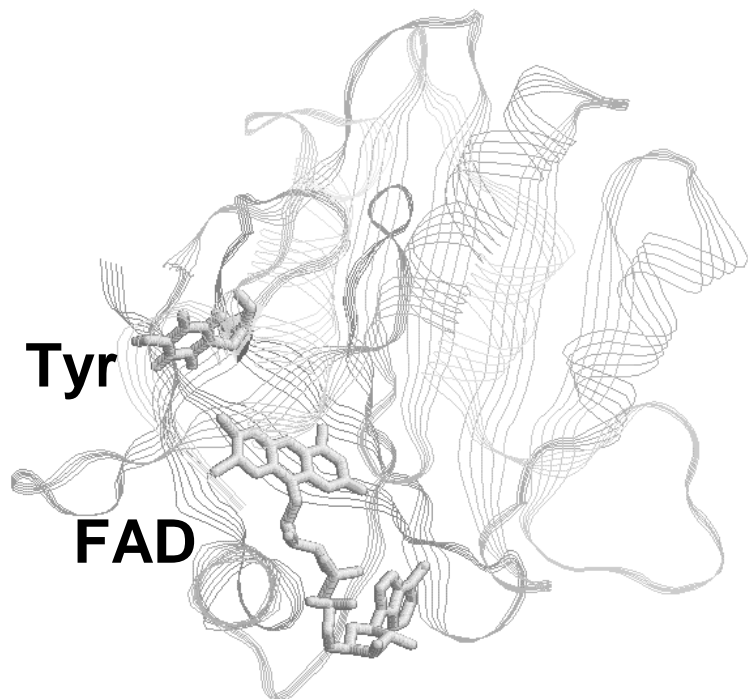
$$E\{F_t F_s\} = k_B T \zeta \cdot K(t-s)$$

- Furthermore, if we take

$$\gamma^2(\varpi)g(\varpi) = \frac{1}{\pi} \Gamma(2H+1) \sin(H\pi) |\varpi|^{3-2H} \implies \text{fGn memory kernel}$$

$$K(t) = H(2H-1)t^{2H-2}$$

Back to experiment



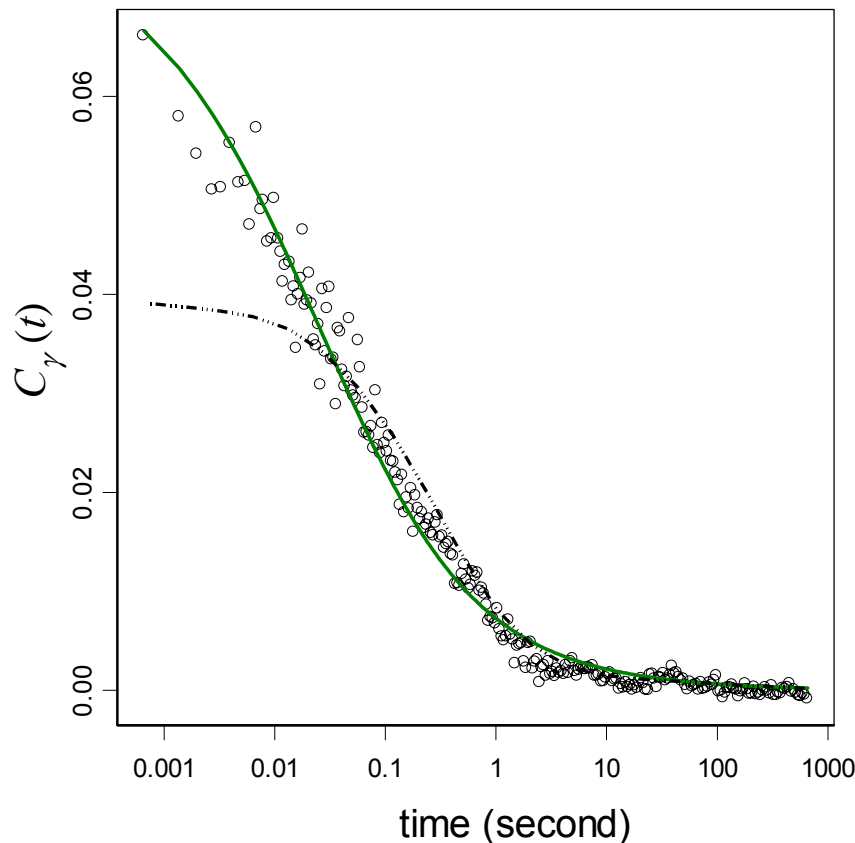
Fluorescence lifetime of FAD depends on the distance between FAD and Tyr

$$\gamma^{-1}(t) = [k_0 e^{-\beta(X_{eq} + X_t)}]^{-1}$$

- ⇒ Model $X(t)$ by GLE with fGn under harmonic potential
easy calculation of lifetime autocorrelation

$$Cov\{\gamma^{-1}(0), \gamma^{-1}(t)\} = k_0^2 e^{2\beta X_{eq} + \beta^2 C_{xx}(0)} (e^{\beta^2 C_{xx}(t)} - 1)$$

Fitting experimental autocorrelation



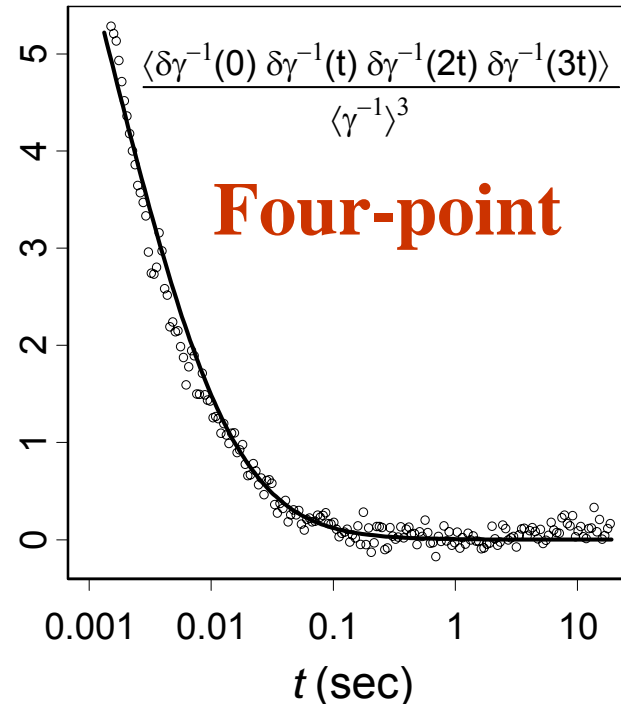
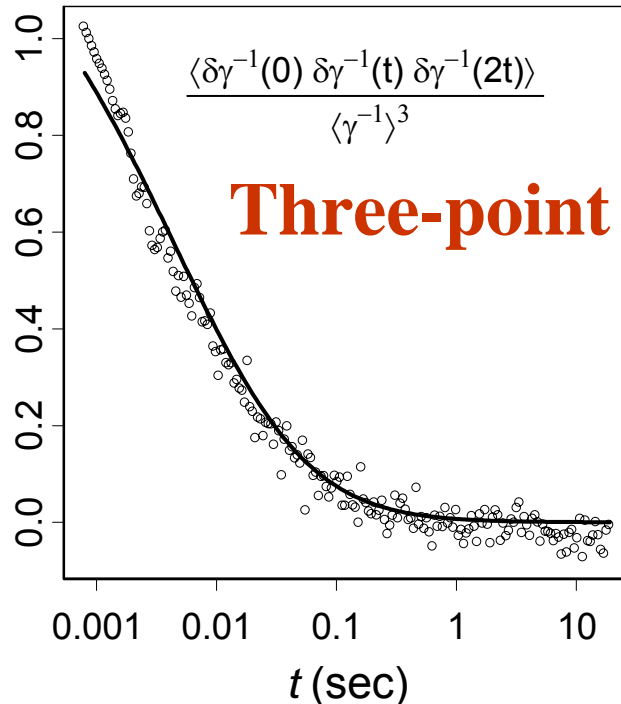
$$H \quad \frac{\zeta}{m\omega^2} \quad \frac{k_B T}{m\omega^2} \beta^2$$

0.74 0.40 0.81

Kou and Xie, *Phys. Rev. Lett.*, **93**,
180603 (2004).

Higher order autocorrelation functions

$$\delta\gamma^{-1}(t) := \gamma^{-1}(t) - \langle \gamma^{-1}(t) \rangle$$



Same parameters:

H	$\frac{\zeta}{m\omega^2}$	$\frac{k_B T}{m\omega^2} \beta^2$
0.74	0.40	0.81

A prediction for three-point autocorrelation

- GLE with fGn predicts time symmetry

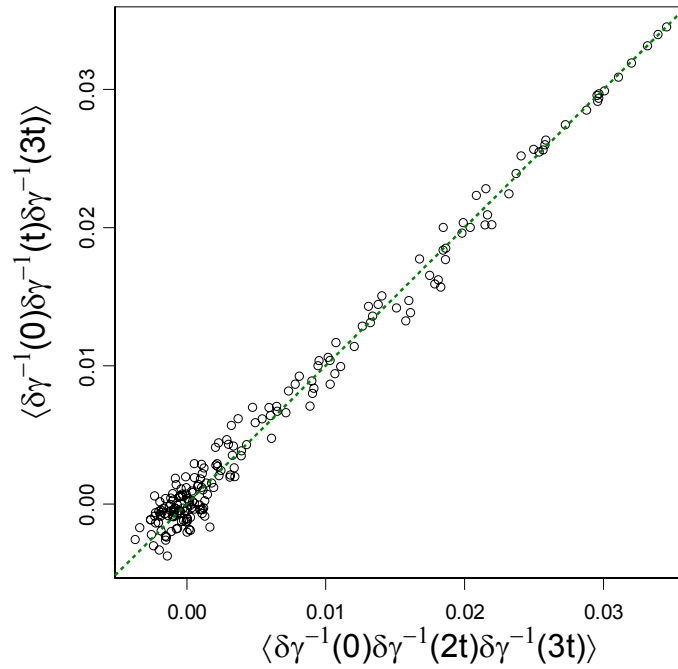
$$\begin{aligned} & E\{\delta\gamma(0)^{-1} \delta\gamma(t_1)^{-1} \delta\gamma(t_1 + t_2)^{-1}\} \\ &= E\{\delta\gamma(0)^{-1} \delta\gamma(t_2)^{-1} \delta\gamma(t_1 + t_2)^{-1}\} \end{aligned}$$

- In particular



$$\begin{aligned} & E\{\delta\gamma^{-1}(0) \delta\gamma^{-1}(t) \delta\gamma^{-1}(3t)\} \\ &= E\{\delta\gamma^{-1}(0) \delta\gamma^{-1}(2t) \delta\gamma^{-1}(3t)\} \text{ for all } t \end{aligned}$$

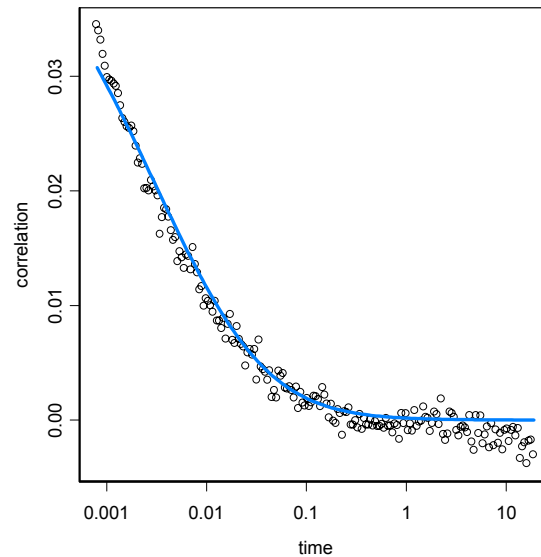
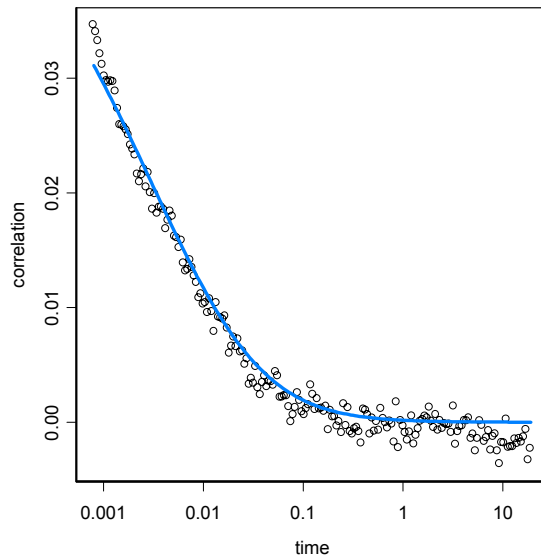
- Check with experiments:



$E\{\delta\gamma^{-1}(0)\delta\gamma^{-1}(t)\delta\gamma^{-1}(3t)\}$
 versus
 $E\{\delta\gamma^{-1}(0)\delta\gamma^{-1}(2t)\delta\gamma^{-1}(3t)\}$

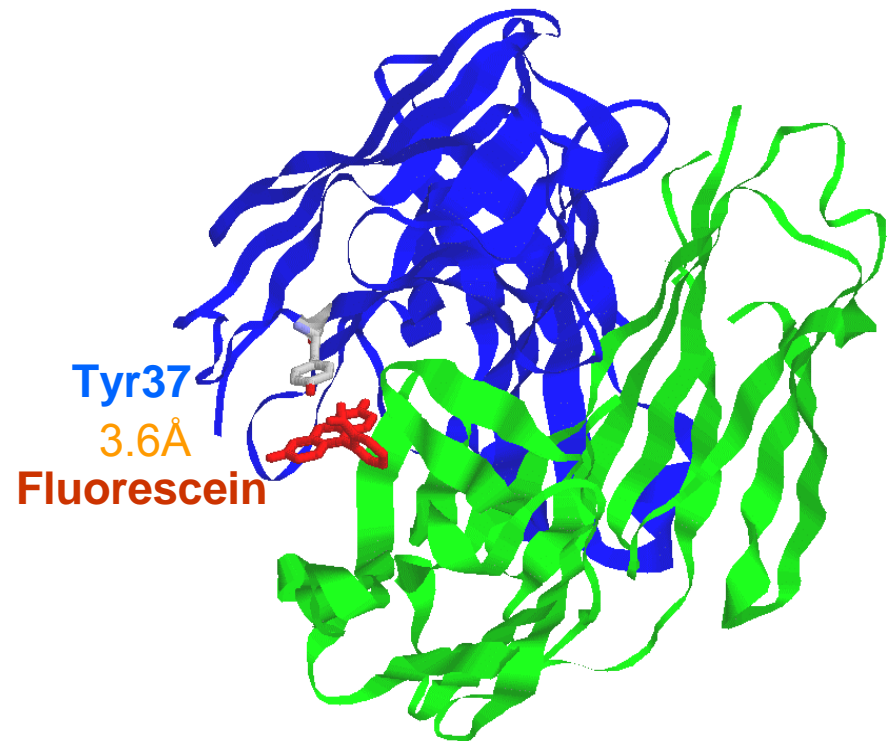
$\langle \delta\gamma^{-1}(0)\delta\gamma^{-1}(2t)\delta\gamma^{-1}(3t) \rangle$

$\langle \delta\gamma^{-1}(0)\delta\gamma^{-1}(t)\delta\gamma^{-1}(3t) \rangle$



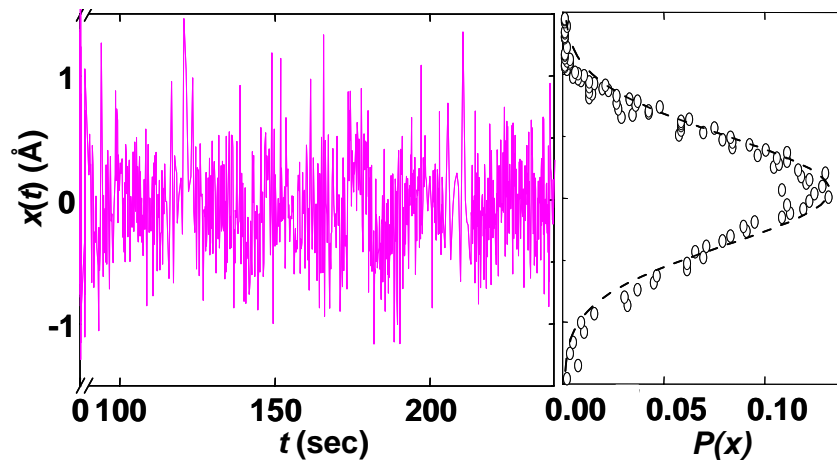
Another system

A protein complex formed between fluorescein (FL) and monoclonal anti-fluorescein (anti-FL)

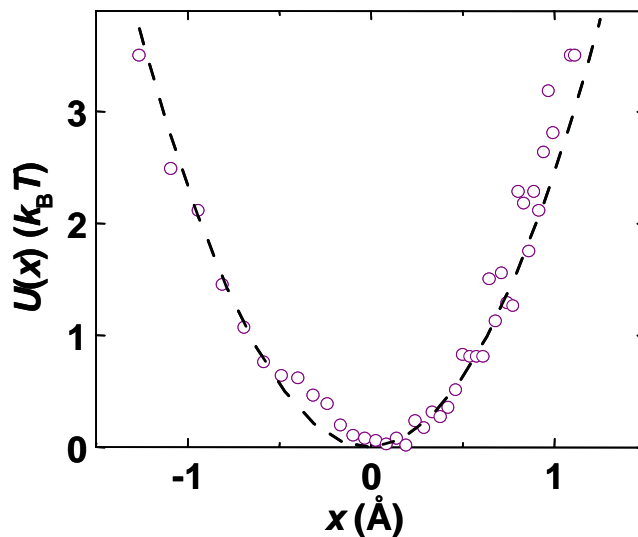


Min, *et al. Phys. Rev. Lett.* **94**, 198302 (2005).

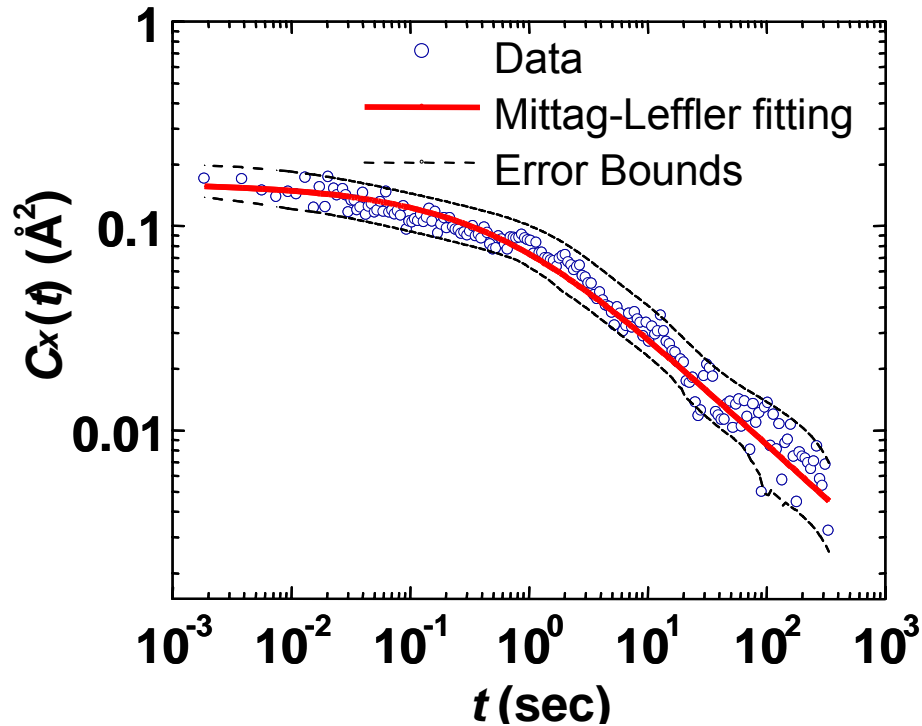
- Obtain distance fluctuation from $\gamma^{-1}(t) = [k_0 e^{-\beta(X_{eq} + X_t)}]^{-1}$



- Since $P(x) \propto \exp(-U(x) / k_B T)$
 Potential function $\hat{U}(x) = -k_B T \ln \hat{P}(x)$



Autocorrelation of distance fluctuation



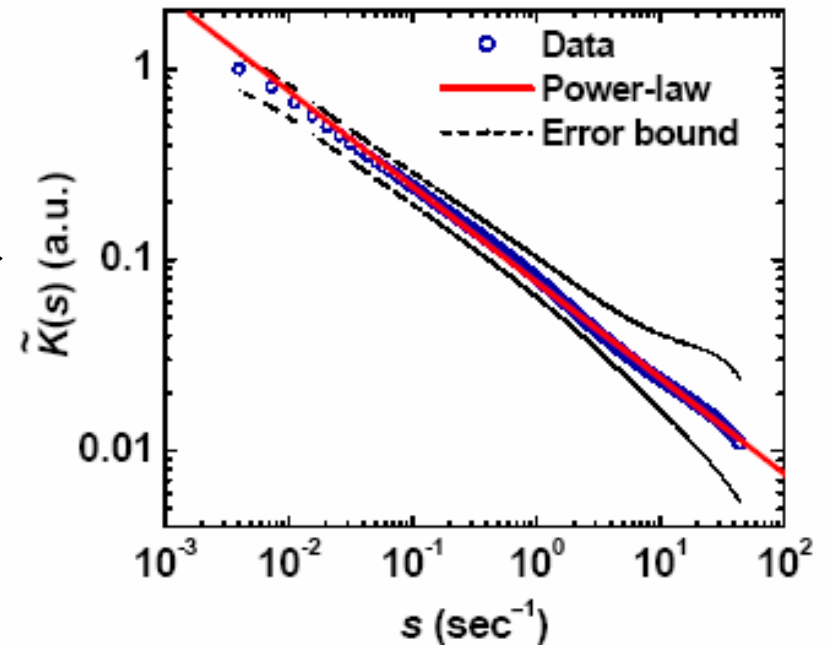
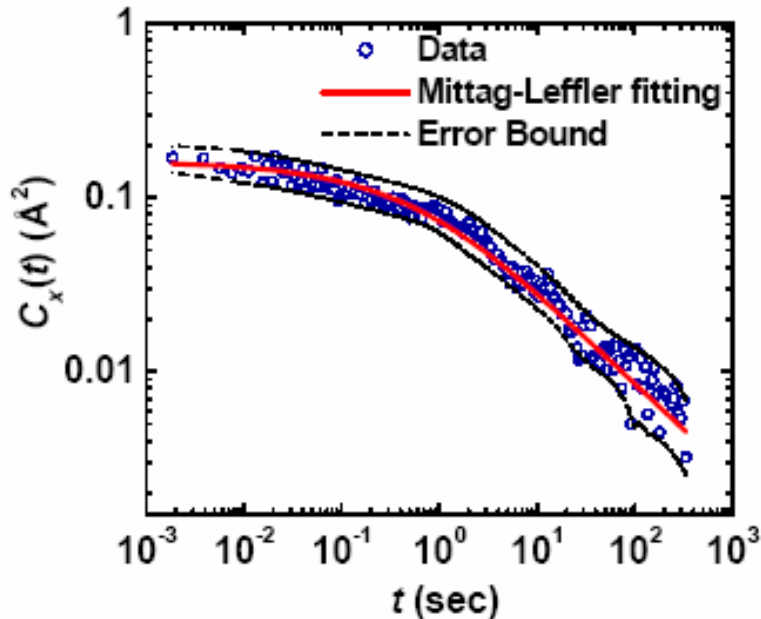
$$m\omega^2 x(t) = -\zeta \int_{-\infty}^t \dot{x}(u) K(t-u) du + F(t)$$

→ **one-to-one correspondence**

$$\tilde{K}(s) = \frac{m\omega^2}{\zeta} \frac{\tilde{C}_x(s)}{C_x(0) - s\tilde{C}_x(s)}$$

Experimental Memory kernel

$$\tilde{K}(s) = \frac{m\omega^2}{\zeta} \frac{\tilde{C}_x(s)}{C_x(0) - s\tilde{C}_x(s)}$$

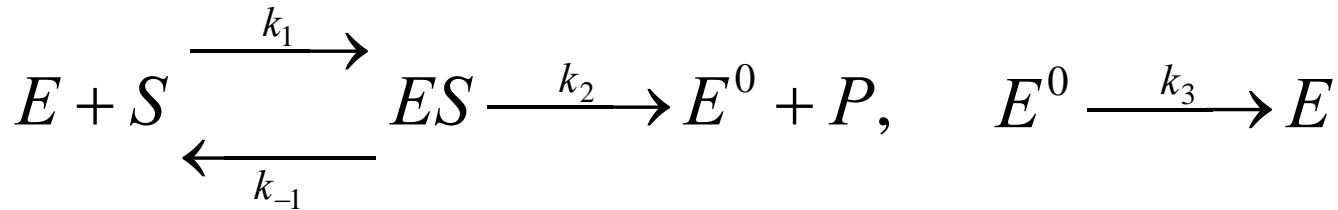


$$K(t) \propto t^{-\alpha-1} = t^{-0.51 \pm 0.07}$$

Brief Recap

- Propose Generalized Langevin Equation with fGn to explain subdiffusion
- Explains the observed conformational dynamics
- One set of parameters fits all
- Key model assumptions verified from experiments

Michaelis-Menten Mechanism



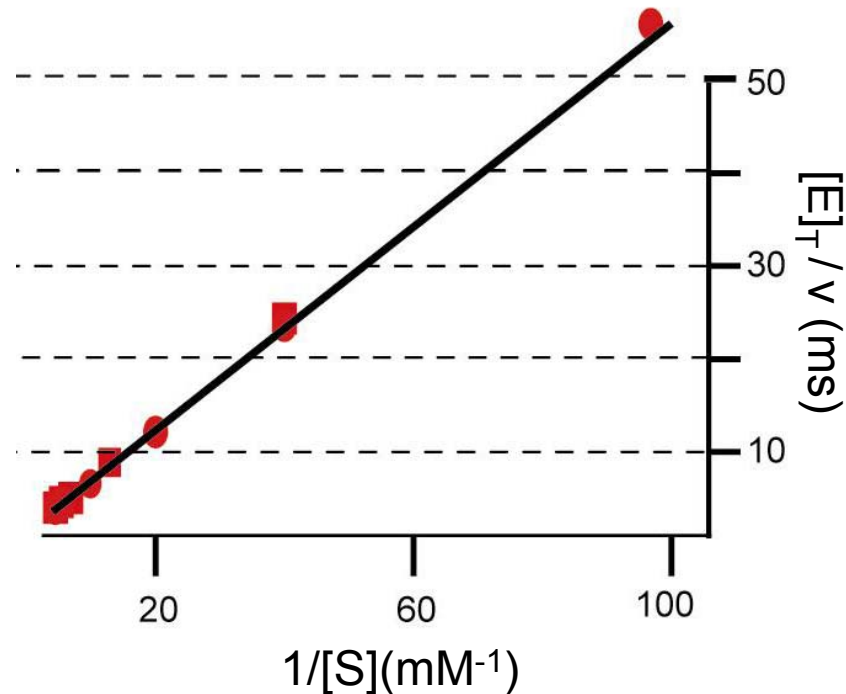
$$\begin{aligned} \frac{d[E]}{dt} &= -k_1[E][S] + k_{-1}[ES] \\ \frac{d[ES]}{dt} &= k_1[E][S] - (k_{-1} + k_2)[ES] \\ \frac{d[E^0]}{dt} &= \frac{d[P]}{dt} = k_2[ES] \end{aligned}$$



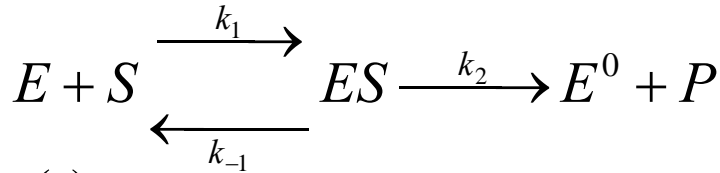
Classical Michaelis-Menten equation

$$v = \frac{v_{\max} [S]}{[S] + K_M}, \quad \begin{aligned} v_{\max} &= k_2 ([E] + [ES]) \\ K_M &= (k_{-1} + k_2) / k_1 \end{aligned}$$

Lineweaver-Burke plot



Single-molecule case



$$\frac{dP_E(t)}{dt} = -k_1[S]P_E(t) + k_{-1}P_{ES}(t)$$

$$\frac{dP_{ES}(t)}{dt} = k_1[S]P_E(t) - (k_{-1} + k_2)P_{ES}(t)$$

$$\frac{dP_{E^0}(t)}{dt} = k_2P_{ES}(t)$$

→ Turnover time distribution

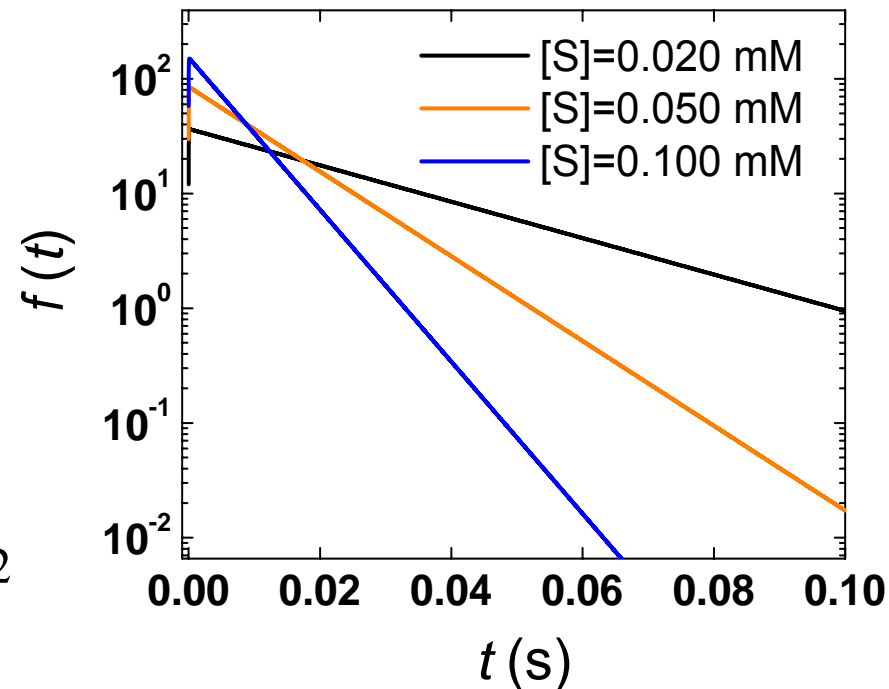
$$f(t) = \frac{k_1 k_2 [S]}{2A} [\exp(A+B)t - \exp(B-A)t]$$

$$A = \sqrt{B^2 - k_1 k_2 [S]}, \quad B = -(k_1[S] + k_{-1} + k_2)/2$$

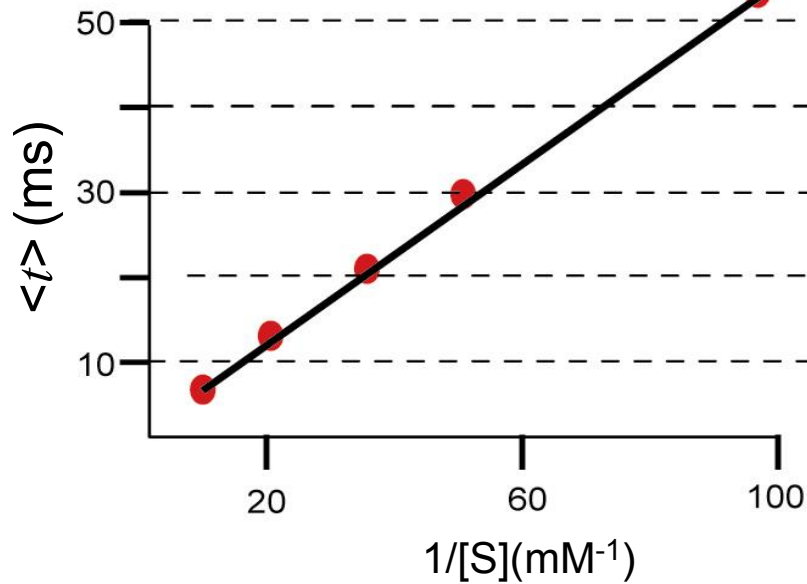
Reaction rate: ↓

$$v = \frac{1}{E(T)} = \frac{k_2[S]}{[S] + K_M}$$

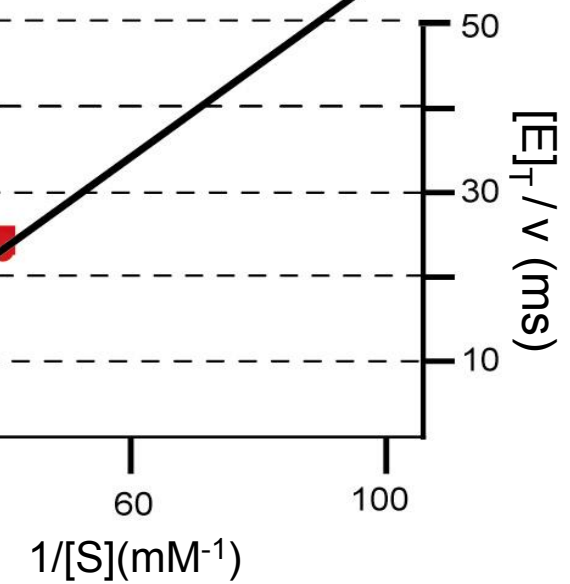
: Still obey the hyperbolic form



Single Molecule



Ensemble



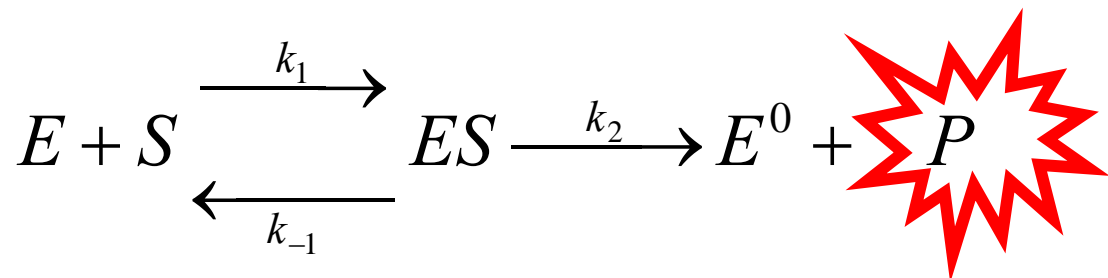
English *et al.*, Nature Chem. Biol., 2, 87 (2006)

β -galactosidase



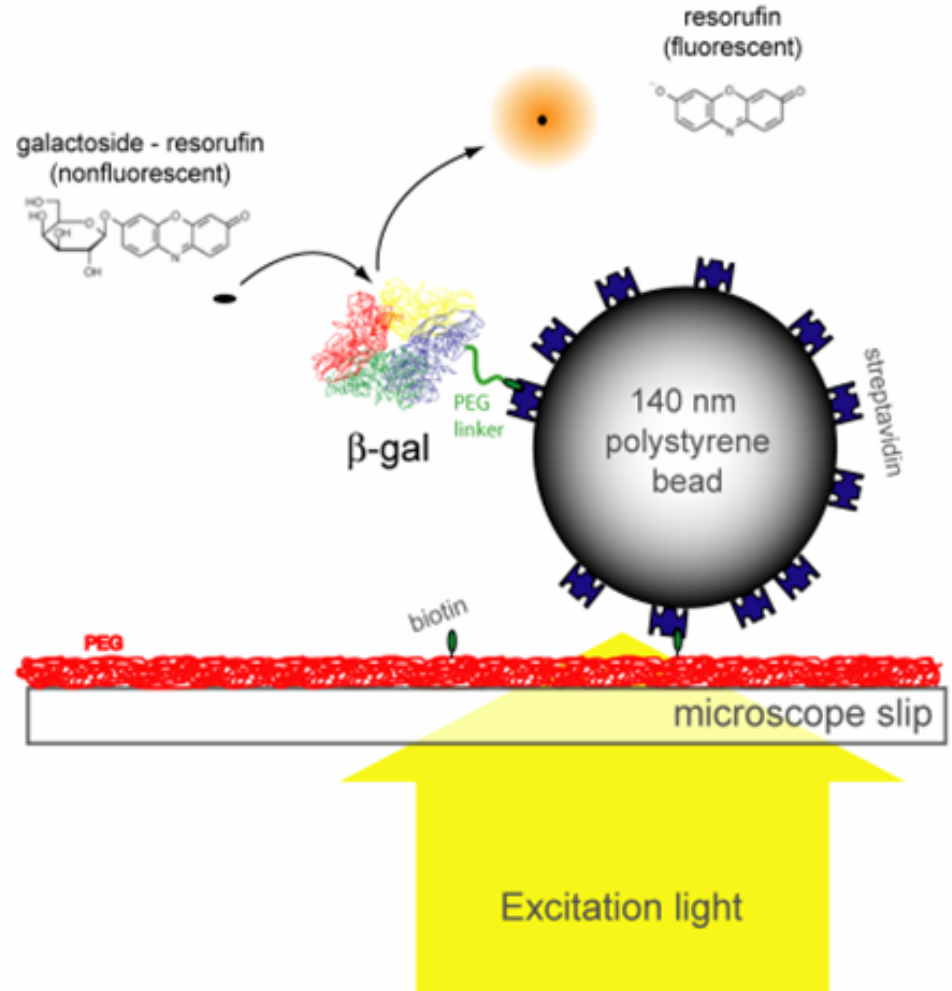
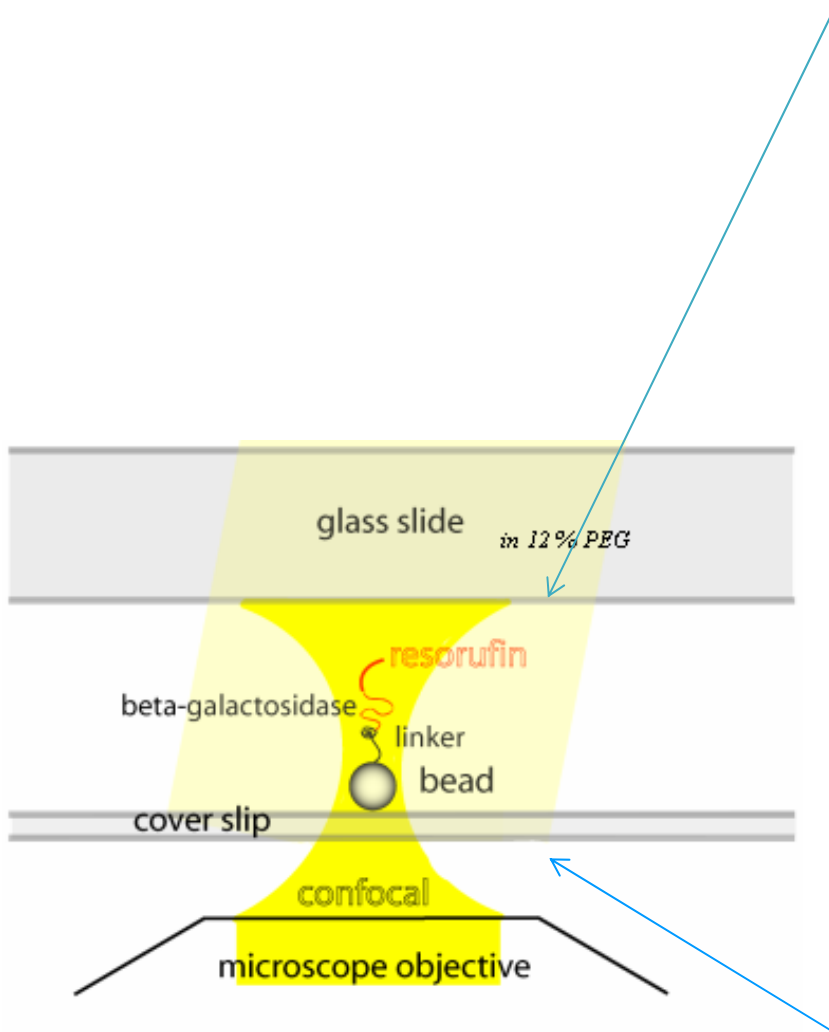
E. coli β -gal catalyzes Hydrolysis of Lactose

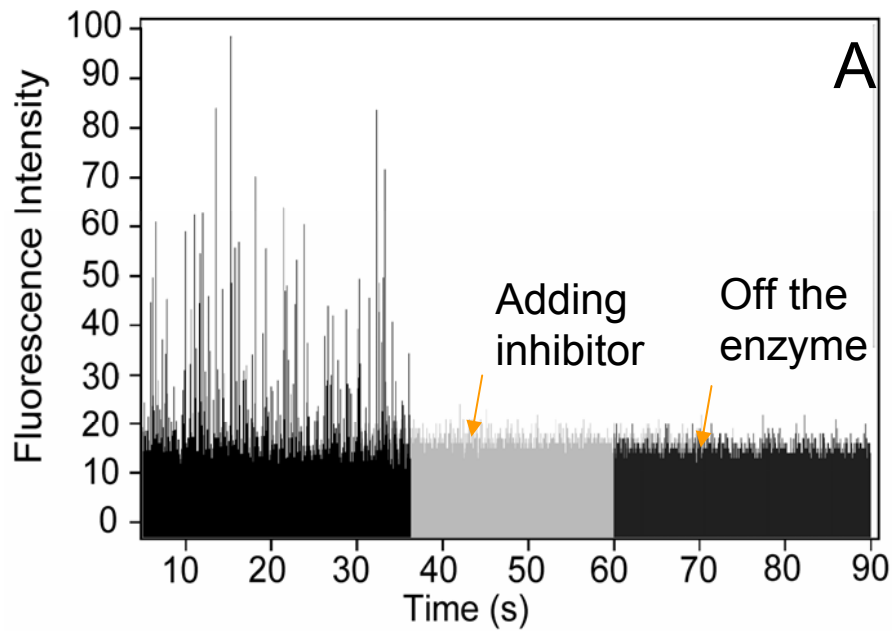
The experiment uses **photogenic** substrate



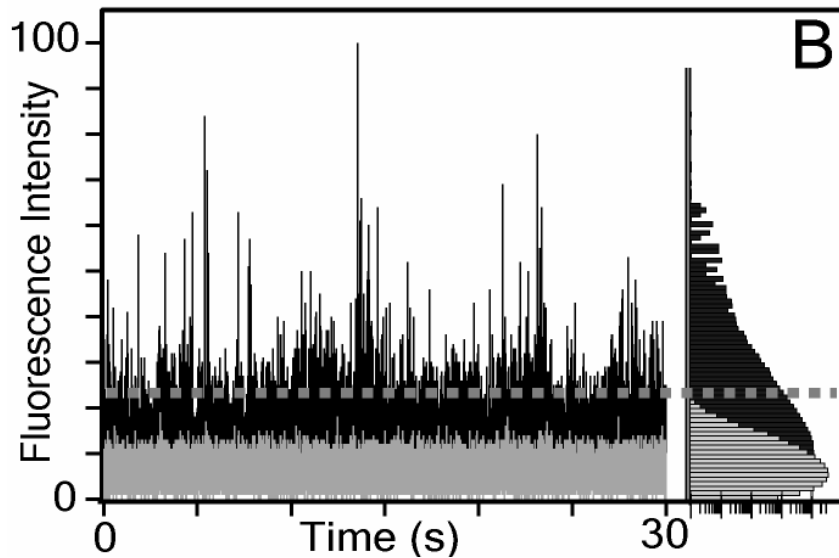
Single Molecule Turnover Experiment of β -galactosidase

Each enzymatic turnover creates a fluorescent burst



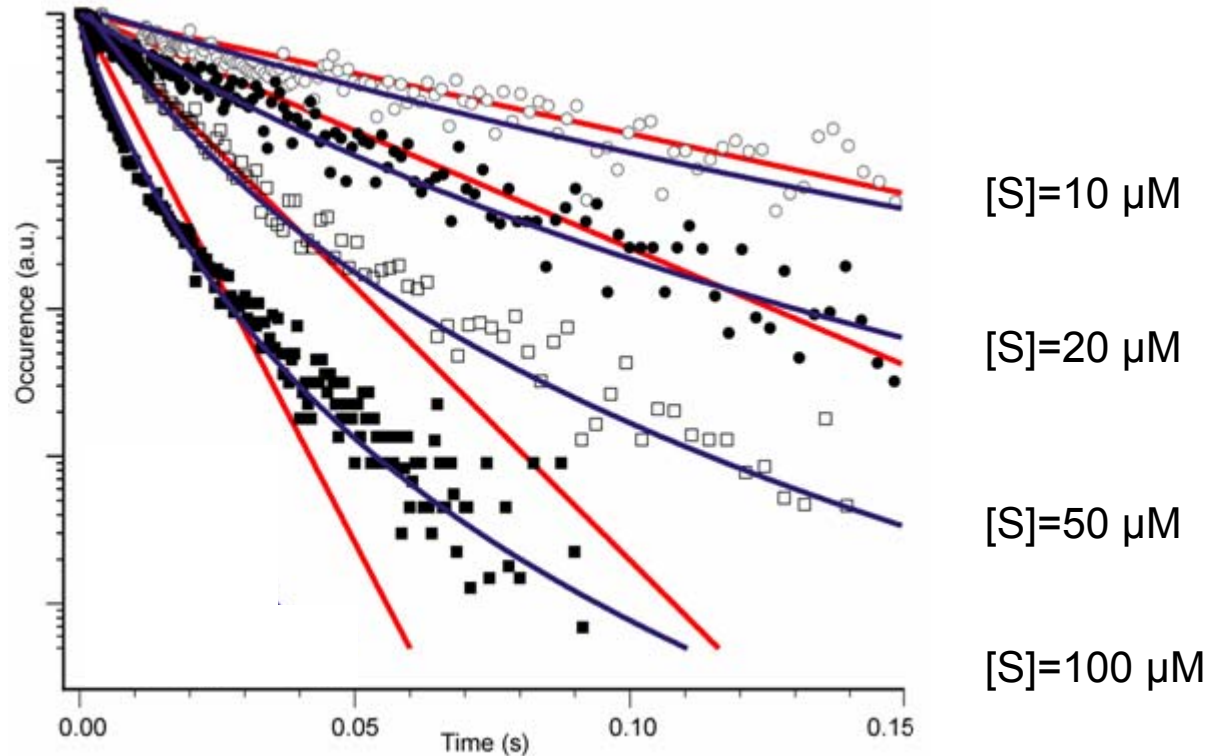


Low Substrate
Concentration
 $20\mu\text{M}$



High Substrate
Concentration
 $100\mu\text{M}$

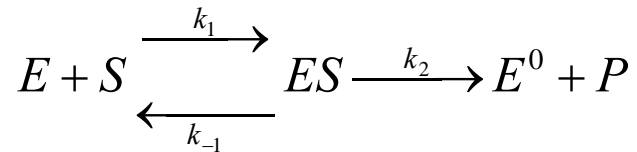
Multi-exponential Distributions of Turnover Times



- Skewed decay at high substrate concentration
- Single exponential decay at low substrate concentration

Memory between successive turnover times

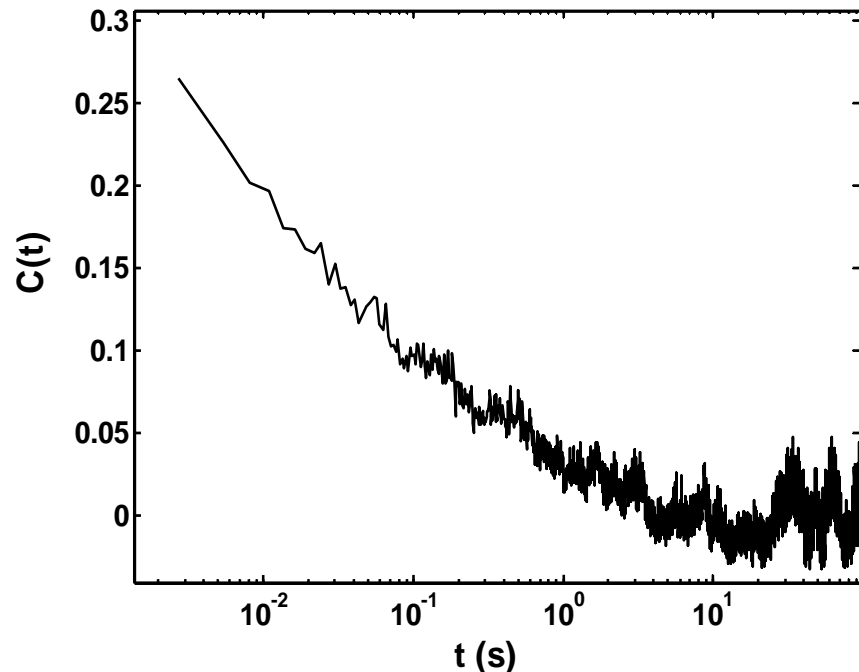
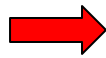
- Under Michaelis-Menten Mechanism

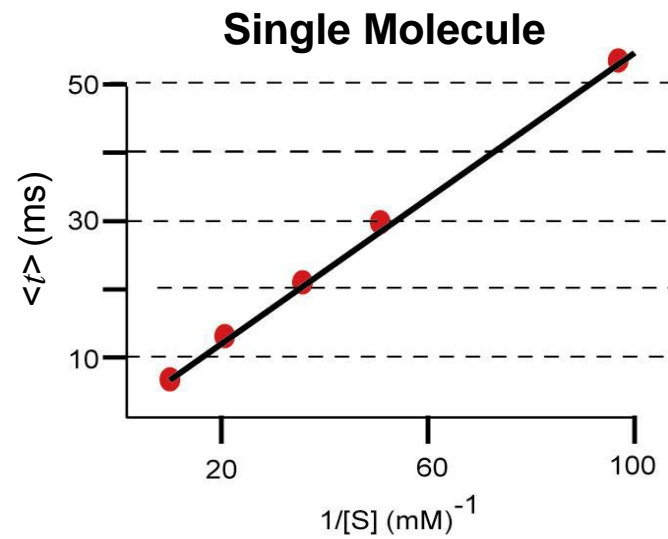
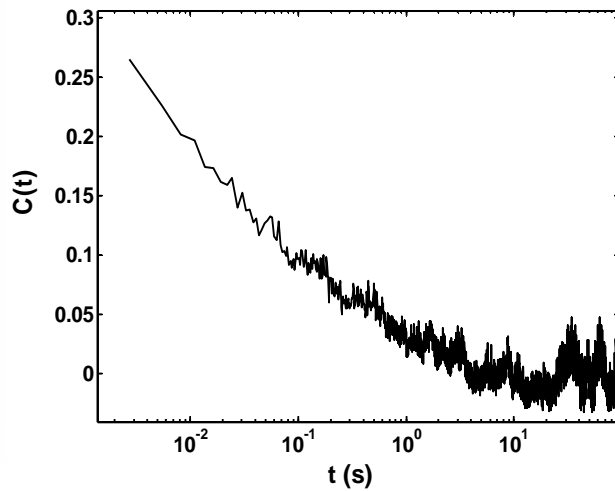
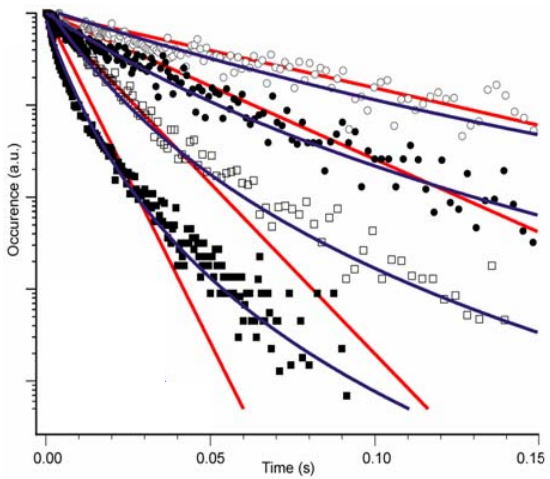


three-state continuous-time **Markov** chain

- Successive turnover times should have **NO** correlation

Experimental data





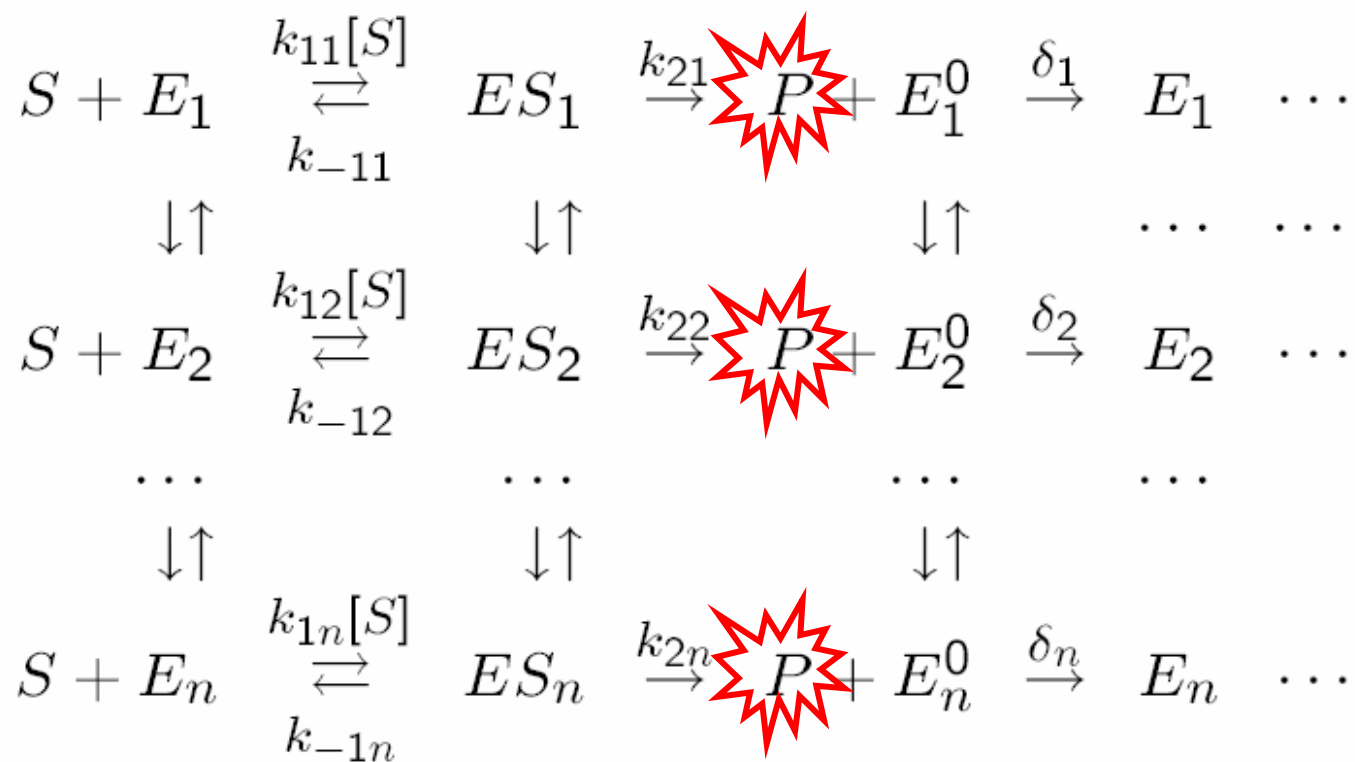


Dynamic disorder

– the fluctuation of enzyme

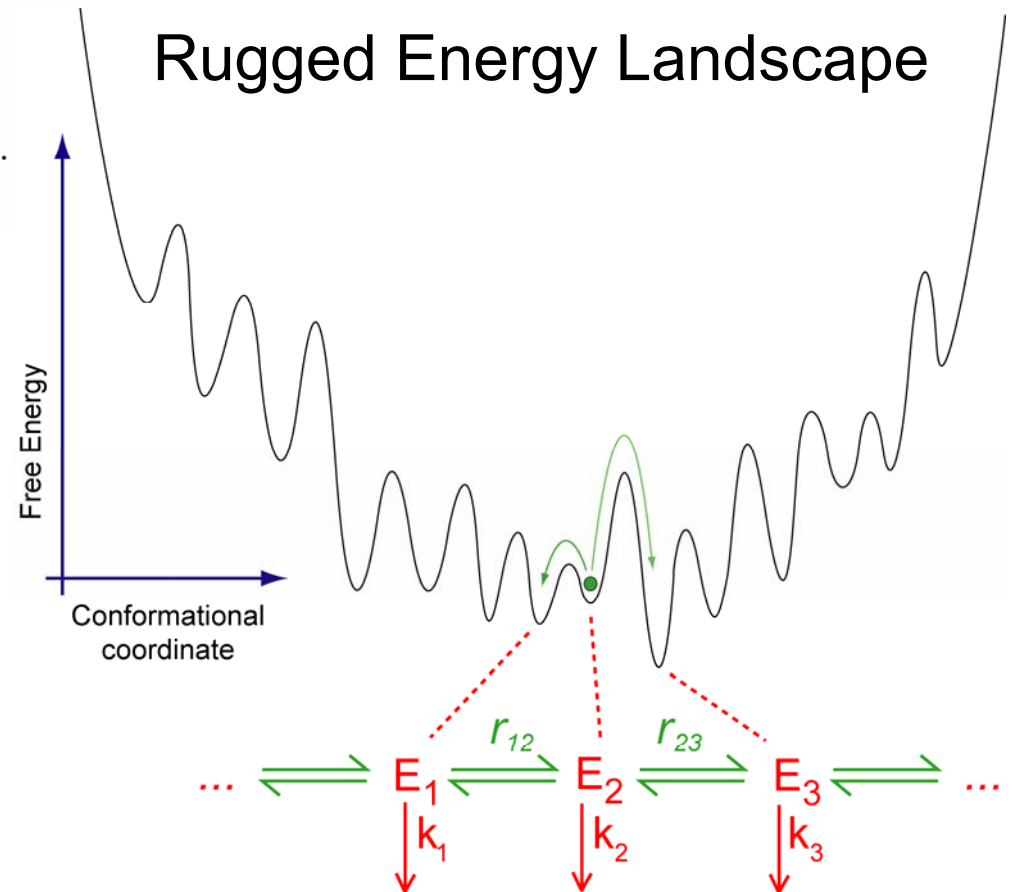
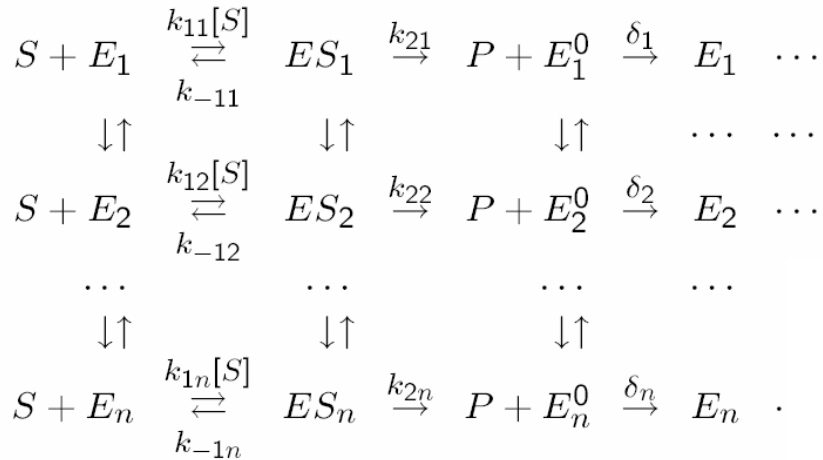
- ❑ An enzyme is a **dynamic** entity with constant spontaneous conformation fluctuation
- ❑ Conformation **fluctuation** occurs on a board range of time scales
- ❑ Different conformation could have **different** enzymatic reaction rate constants.

Dynamic disorder – the fluctuation of enzyme



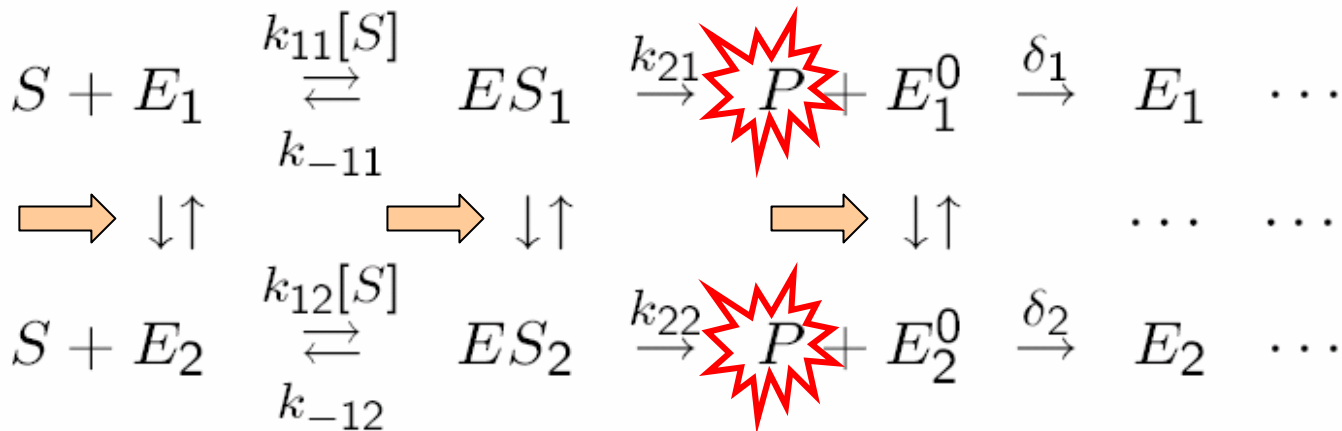
All the E_1, E_2, \dots are experimentally indistinguishable

Dynamic disorder – the fluctuation of enzyme



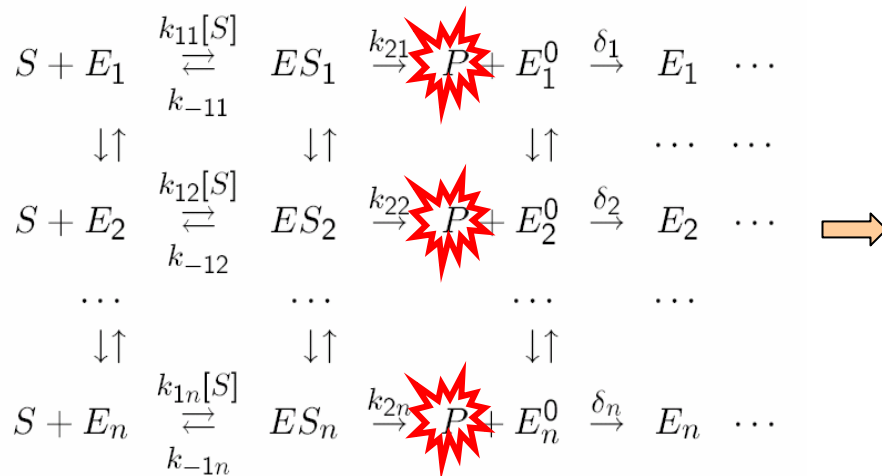
Explain the memory

If parallel transition rates are small



Transitions will stay in one channel for a quite while before going to the next

Naturally give rise to the strong correlation



Turnover time:
first passage time

Can be solved via Laplace
transform and matrix analysis

Under various conditions

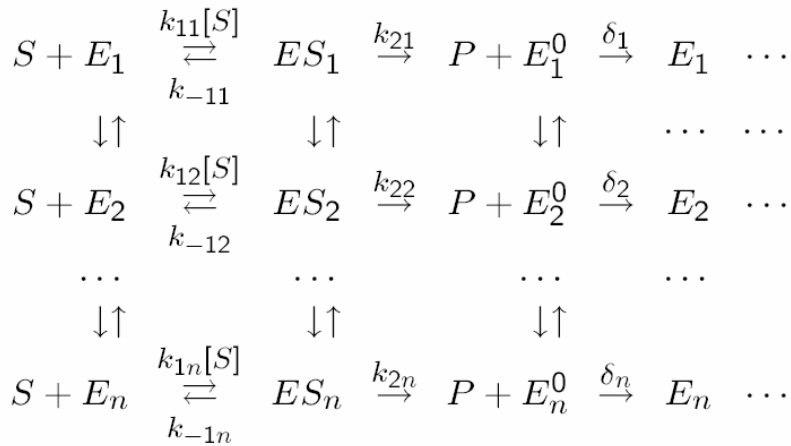
Single Molecule Michaelis-Menten Equation

$$v = \frac{1}{E(T)} = \frac{\gamma_2[S]}{[S] + C_M}$$

$$\gamma_2 = \left[\int_0^\infty \frac{p(k_2)}{k_2} dk_2 \right]^{-1}$$

$$C_M = \frac{\gamma_2 + k_{-1}}{k_1}$$

Weighted harmonic mean



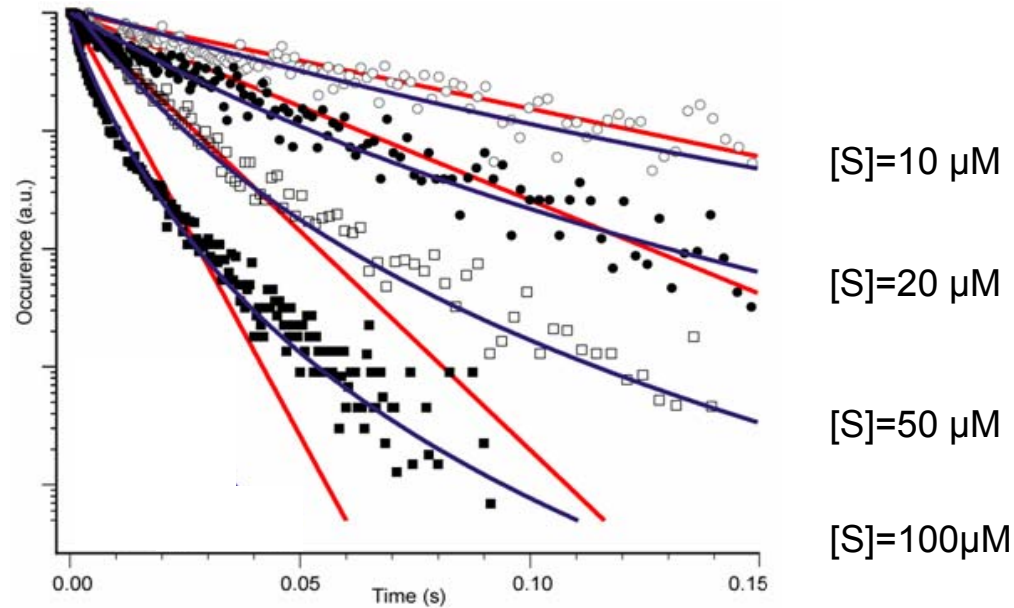
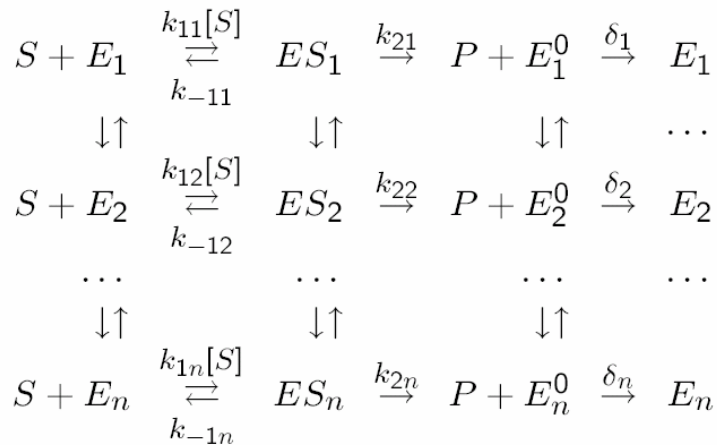
Reaction rate

$$v = \frac{1}{E(T)} = \frac{\gamma_2[S]}{[S] + C_M}$$

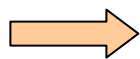
If one of the following conditions holds:

- | | |
|---|----------------------------|
| (a) slow interconversion between E_i | } Quasi static |
| (b) slow interconversion between ES_i | |
| (c) fast interconversion between E_i | |
| (d) fast interconversion between ES_i | |
| (e) $k_{-1i} \gg k_{2i}$ | Quasi equilibrium |
| (f) $k_{2i}/k_{-1i} = \text{const}$ | Conformational equilibrium |

Model fitting



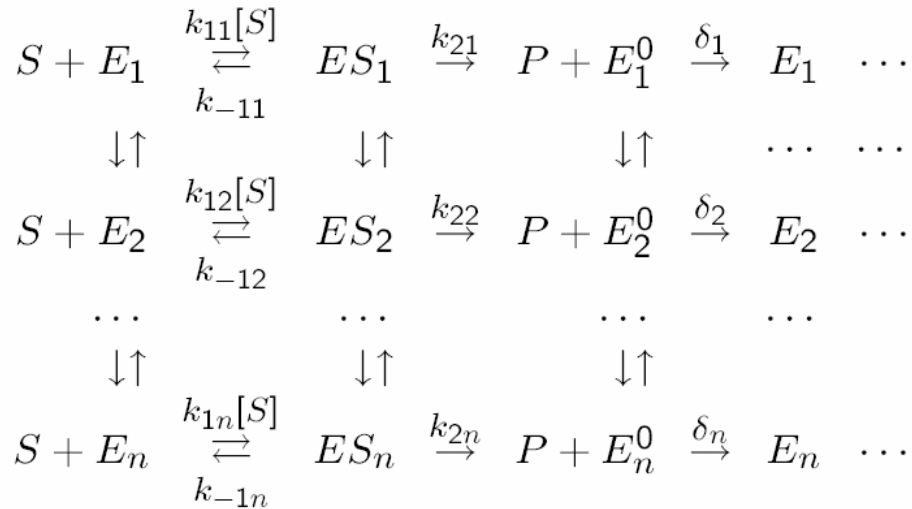
- Skewed decay at high substrate concentration
- Single exponential decay at low concentration



$$k_{11} = k_{12} = \dots = k_{1n} = k_1$$

$$k_{-11} = k_{-12} = \dots = k_{-1n} = k_{-1}$$

Turnover Time distribution for fluctuating enzymes



Quasi Static Limit:

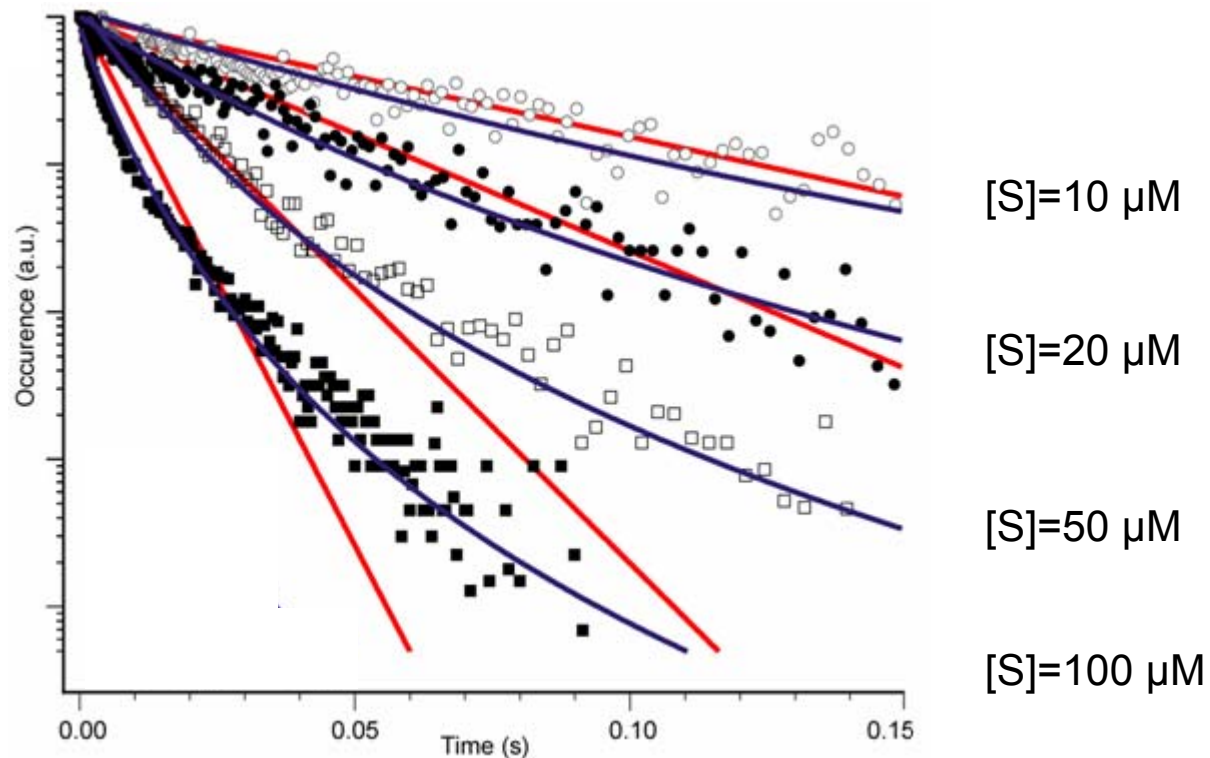
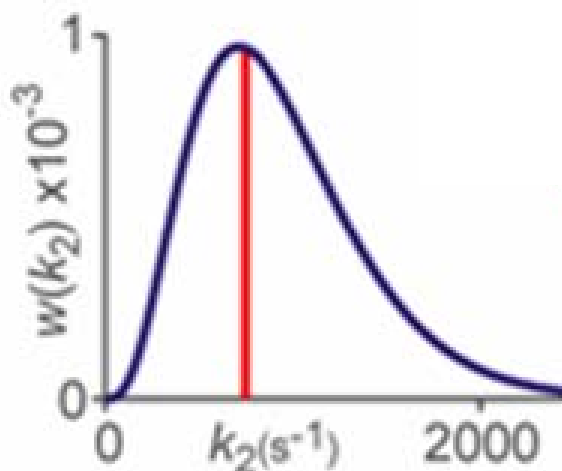
$$f(t) = \int_0^{\infty} dk_2 w(k_2) \frac{k_1 k_2 [S]}{2A} [\exp(A + B)t - \exp(B - A)t]$$

$$A = \sqrt{B^2 - k_1 k_2 [S]}, \quad B = -(k_1 [S] + k_{-1} + k_2) / 2$$

Multi-exponential Distributions of Turnover Times

Gamma Distribution

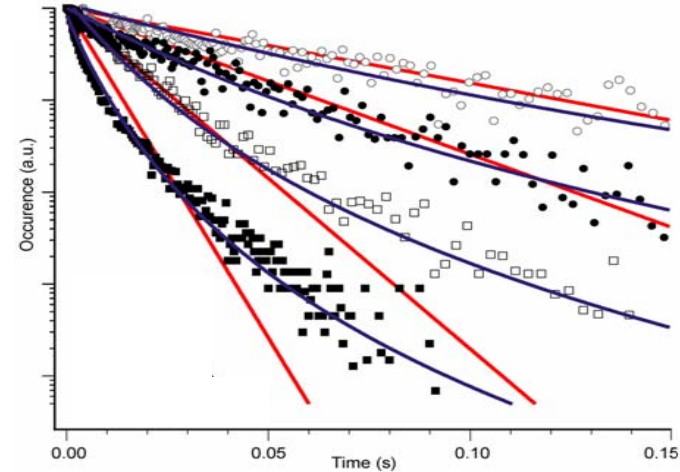
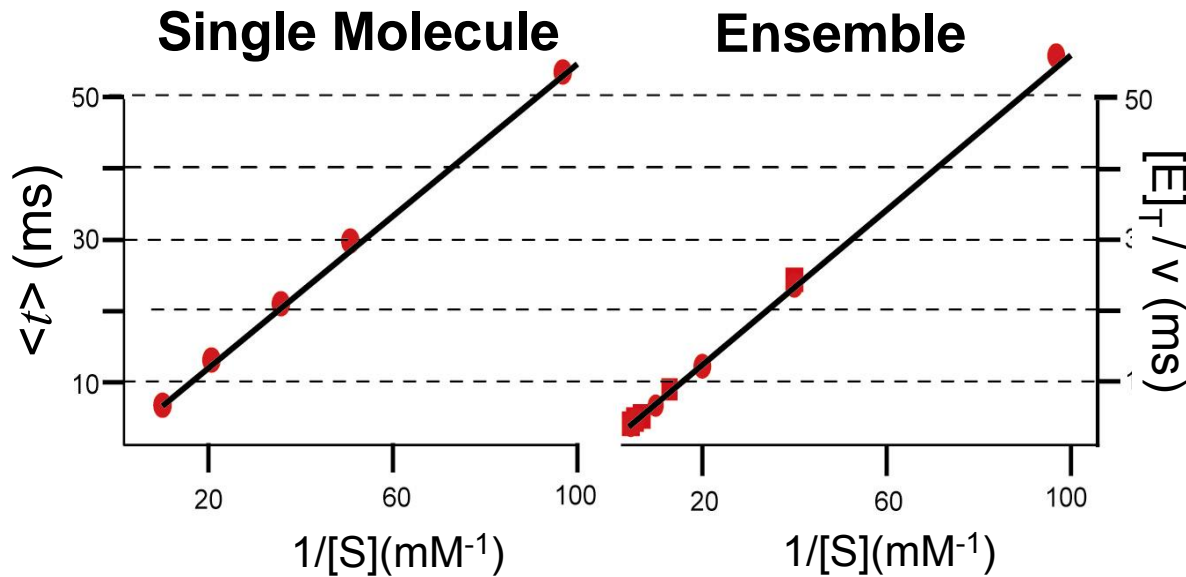
$$w(k_2) = \frac{1}{b^a \Gamma(a)} k_2^{a-1} \exp(-k_2/b)$$



$$k_1 = 5 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$$

$$k_{-1} = 1.83 \times 10^4 \text{ s}^{-1}$$

$$a = 4.2, b = 224 \text{ s}^{-1}$$



Consequences:

- (a) Single Molecule reconciled with ensemble study
- (b) Dynamic disorder might be masked in ensemble studies!
- (c) The apparent k_2 and K_M of the Michaelis-Menten equation are complex functions of the k_2 and K_M of a large distribution of conformers, different from their conventional interpretations.

Summary

- Michaelis-Menten with dynamic disorder
 - Introduce a stochastic network model
 - Explains experimental puzzle
 - Derive single-molecule Michaelis-Menten equation
 - Dynamic disorder might be masked in ensemble studies!
- Generalized Langevin Equation with fGn
 - Explains the observed conformational dynamics
 - One set of parameters fits all
 - Prediction confirmed by experimental data
 - Each model assumption verified from experiments

Summary (Continued)

- The connection
 - Simple underlying picture behind both
 - An enzyme is a dynamic entity with conformational fluctuation on a broad range of time scales.
 - The interconverting conformations have different enzymatic reaction rate constants.



Acknowledgement

- Xie group at Harvard Chemistry & Chemical Biology
- Binny Cherayil
- Attila Szabo
- Hong Qian

- NSF grant DMS-02-04674
- NSF Career Award