베이시안 방법론을 통한 바이오 동역학 연구

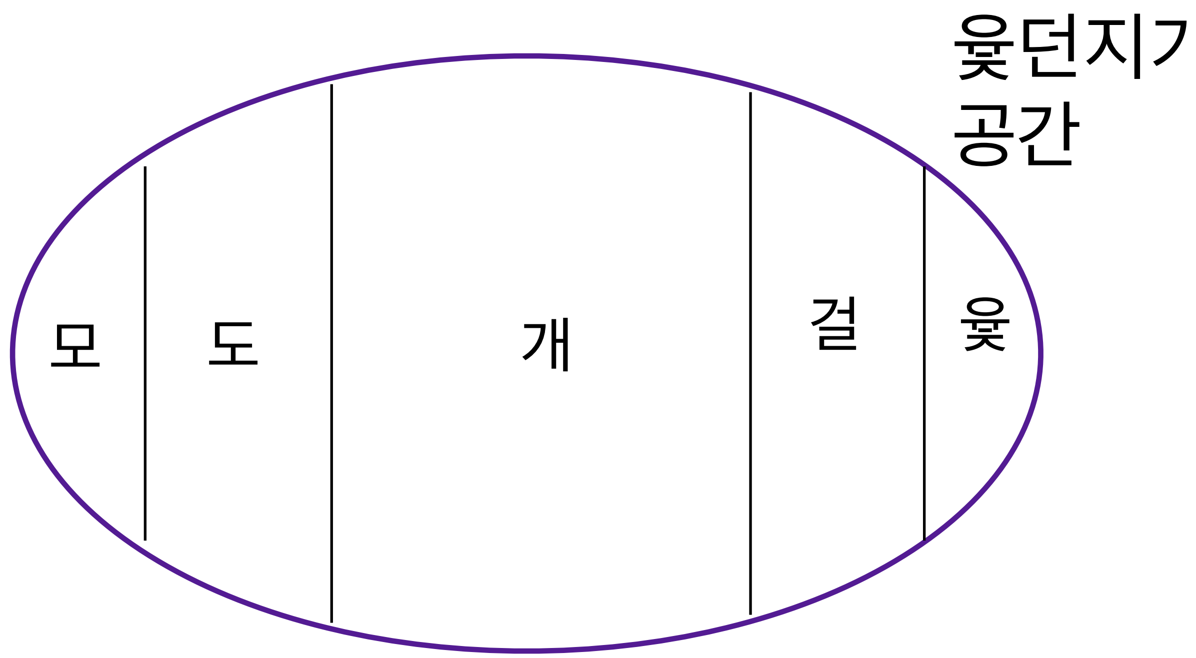
DNA sliding protein의 DNA 위에서의 확산 동역학

Jae-Hyung Jeon (전재형)
Department of Physics, POSTECH
Asia Pacific Center for Theoretical Physics (APCTP)
E-mail: jeonjh@gmail.com

APCTP-POSTECH Biophysics Summer School, July 5th, 2023

Basics of probability theory

Probability $p(\text{도}) = \dfrac{\#\text{ of “도” 나오는 사건 (관측수)}}{\#\text{ of 윷던지기 (실험수)}} = \dfrac{\text{“도” 나오는 영역 (4)}}{\text{윷던지기 공간 (16)}}$



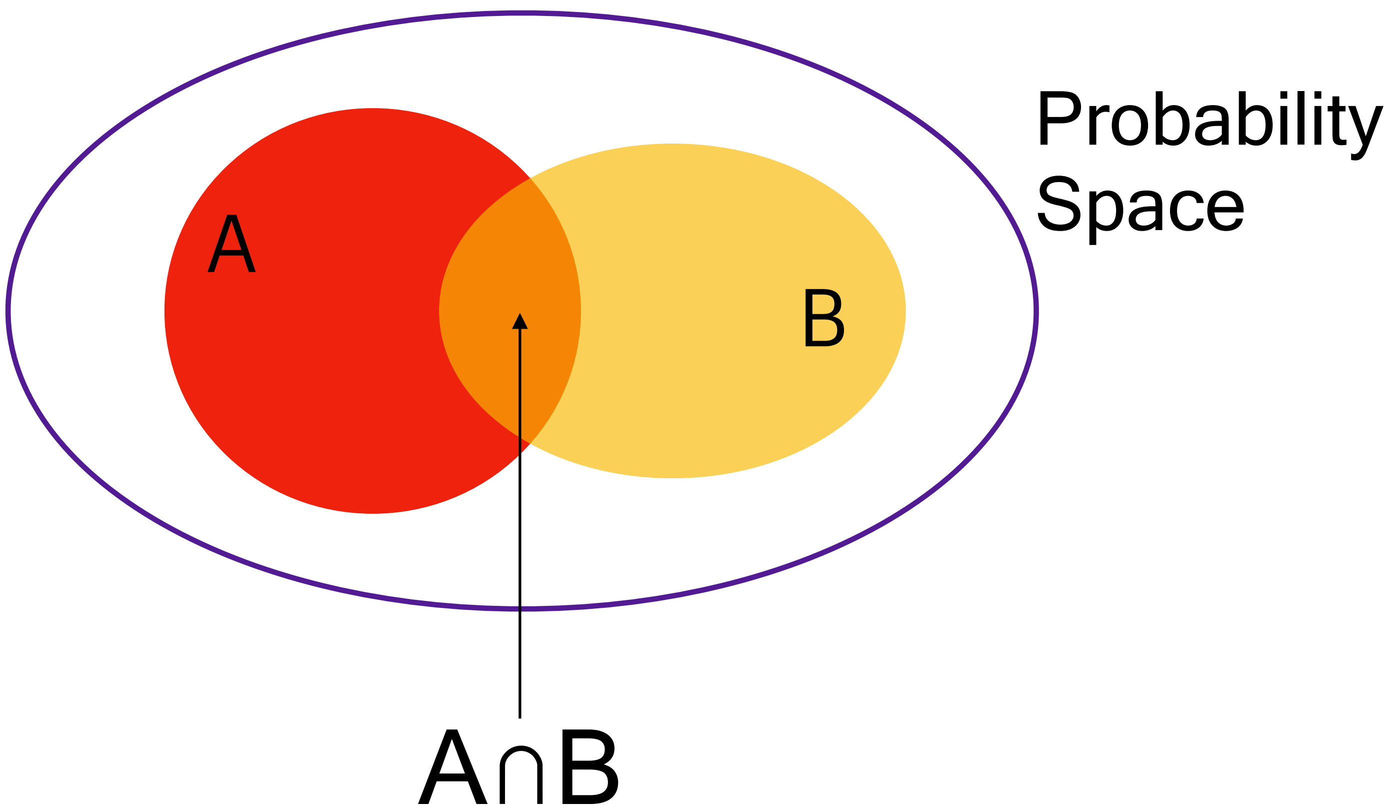
Sum of p over all possible events in the probability space
$= p(\text{모}) + p(\text{도}) + p(\text{개}) + p(\text{걸}) + p(\text{윷}) = 1$

Probability $p(\text{도}^c) = p(\text{모}) + p(\text{개}) + p(\text{걸}) + p(\text{윷}) = 1 - p(\text{도})$

Basics of probability theory (2)

Joint probability: Probability that multiple events occur simultaneously.

p(A, B) = the joint probability that event "A" and event "B" occur simultaneously.



$$p(A, B) = p(A \cap B) = p(B \cap A) = p(B, A)$$

Basics of probability theory (3)

Joint probability: Probability that multiple events occur simultaneously.

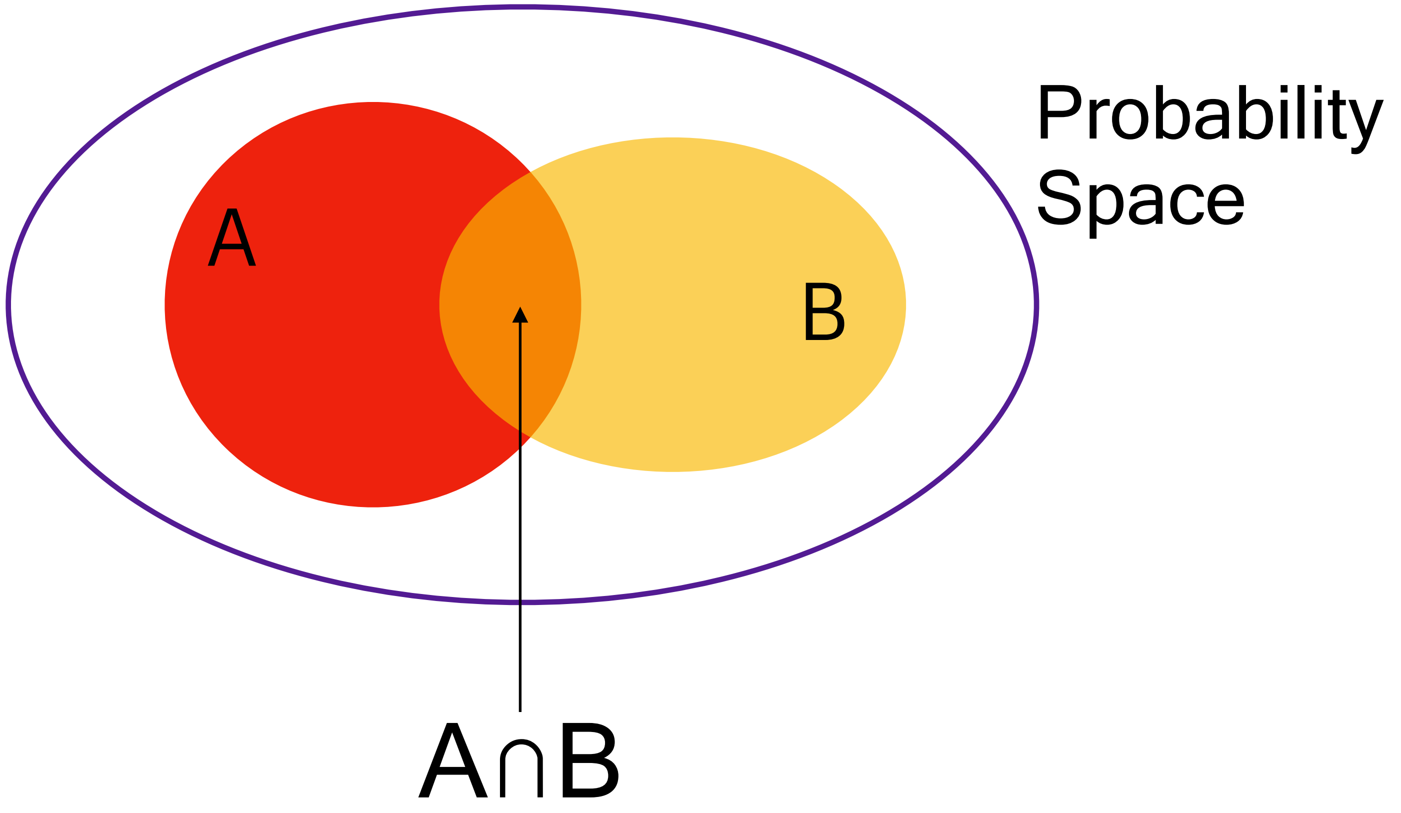
p(try1="윷", try2="모") = 1*1/(16*16)=(1/16)*(1/16) ≃0.004

Basics of probability theory (4)

Conditional probability: Probability of an event, given that other events already occurred.

p(A|B) = the probability that event "A" occurs under the condition that event "B" was given.

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$



NB:

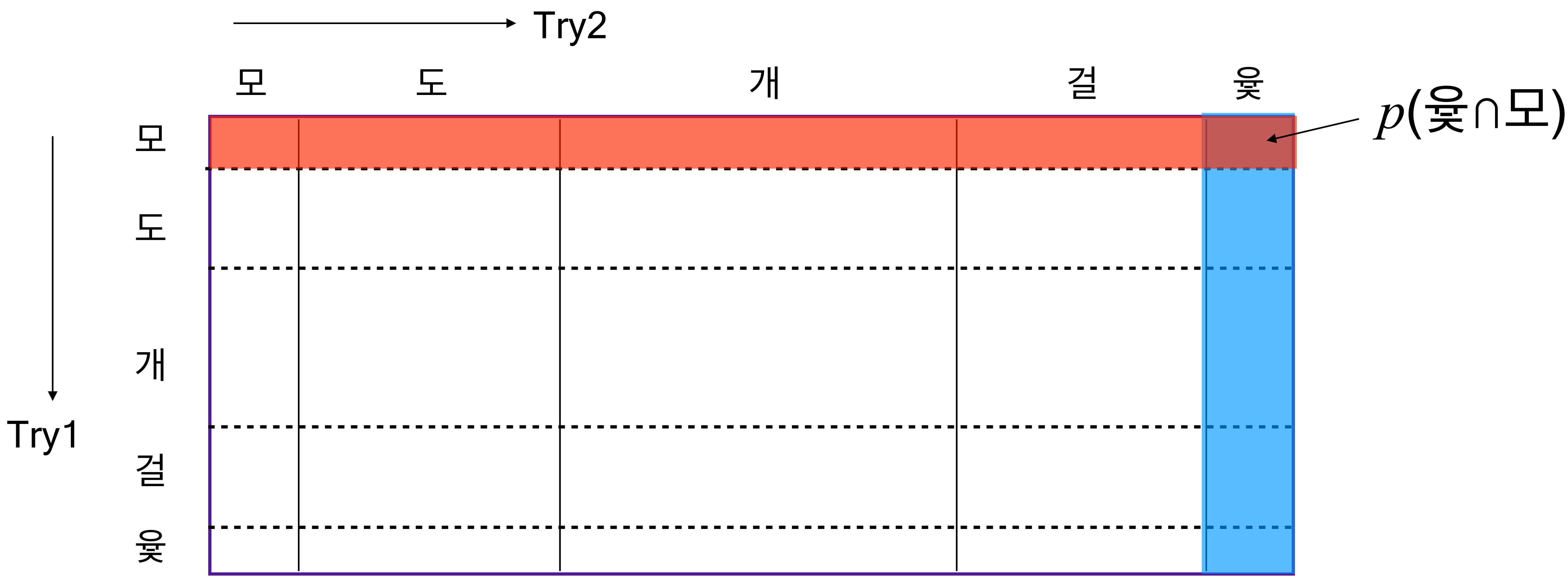
$$p(A|B) \neq p(A, B)$$

$$p(A|B) \neq p(B|A)$$

Basics of probability theory (5)

Conditional probability: Probability of an event, given that other events already occurred.

$$p(\text{try2}=\text{“모”}|\text{try1}=\text{“윷”}) = \frac{p(\text{try1}=\text{“윷”} \cap \text{try2}=\text{“모”})}{p(\text{try1}=\text{“윷”})} = \frac{1/(16*16)}{1/16}$$



Bayes' theorem

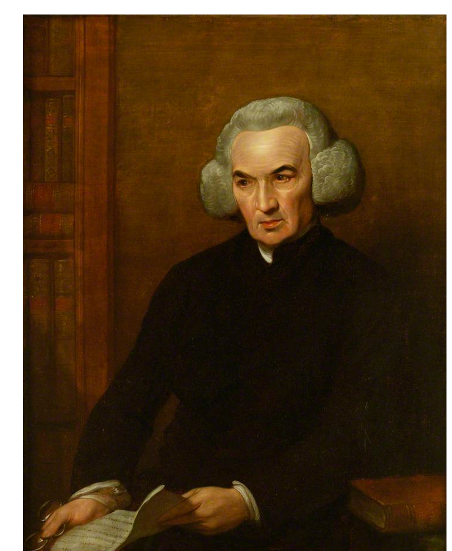
Thomas Bayes
(1747–1760)

Presbyterian (장로교) minister
Statistician & Philosopher

T. Bayes & Richard Prices (1763), An Essay towards solving a problem in the doctrine of chances

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

R. Prices
(1723–1791)



(Pf) Joint prob. p(A, B) = p(A∩B)

= p(A|B) p(B)

= p(B, A)

= p(B|A) p(A)

Bayesianist's interpretation on probability

H: 가설(hypothesis)
D: 데이터(data)

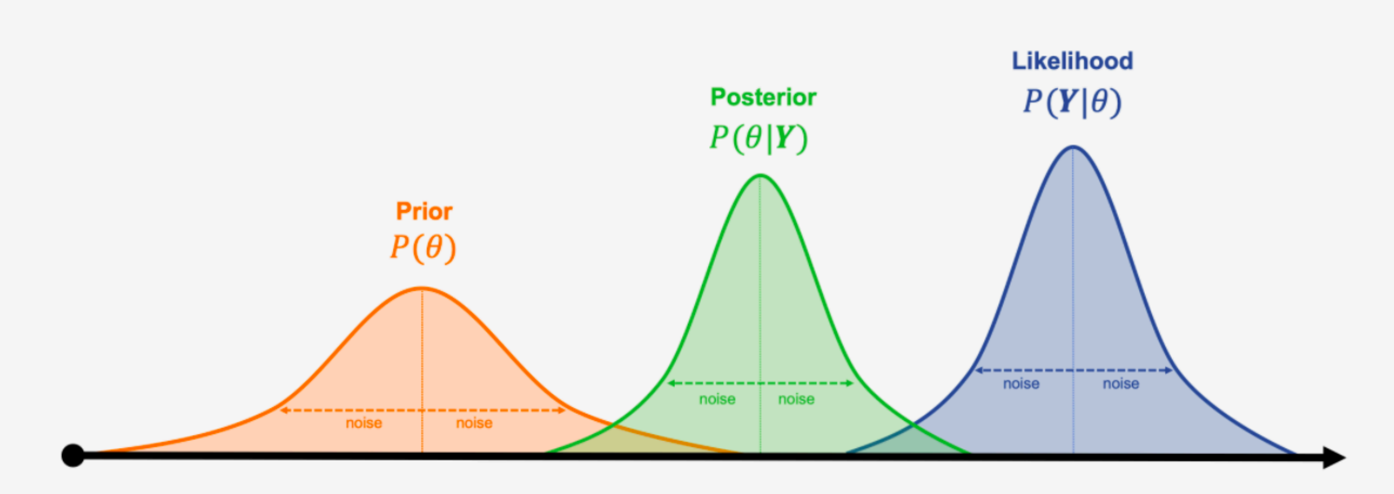
가능성(likelihood) 함수
주어진 가설(H)하에서 데이터(D)를 관측할 가능성(혹은 확률)

사후확률(posterior)
새로운 데이터(D)를 관측한 후 가설(H)에 대한 확률

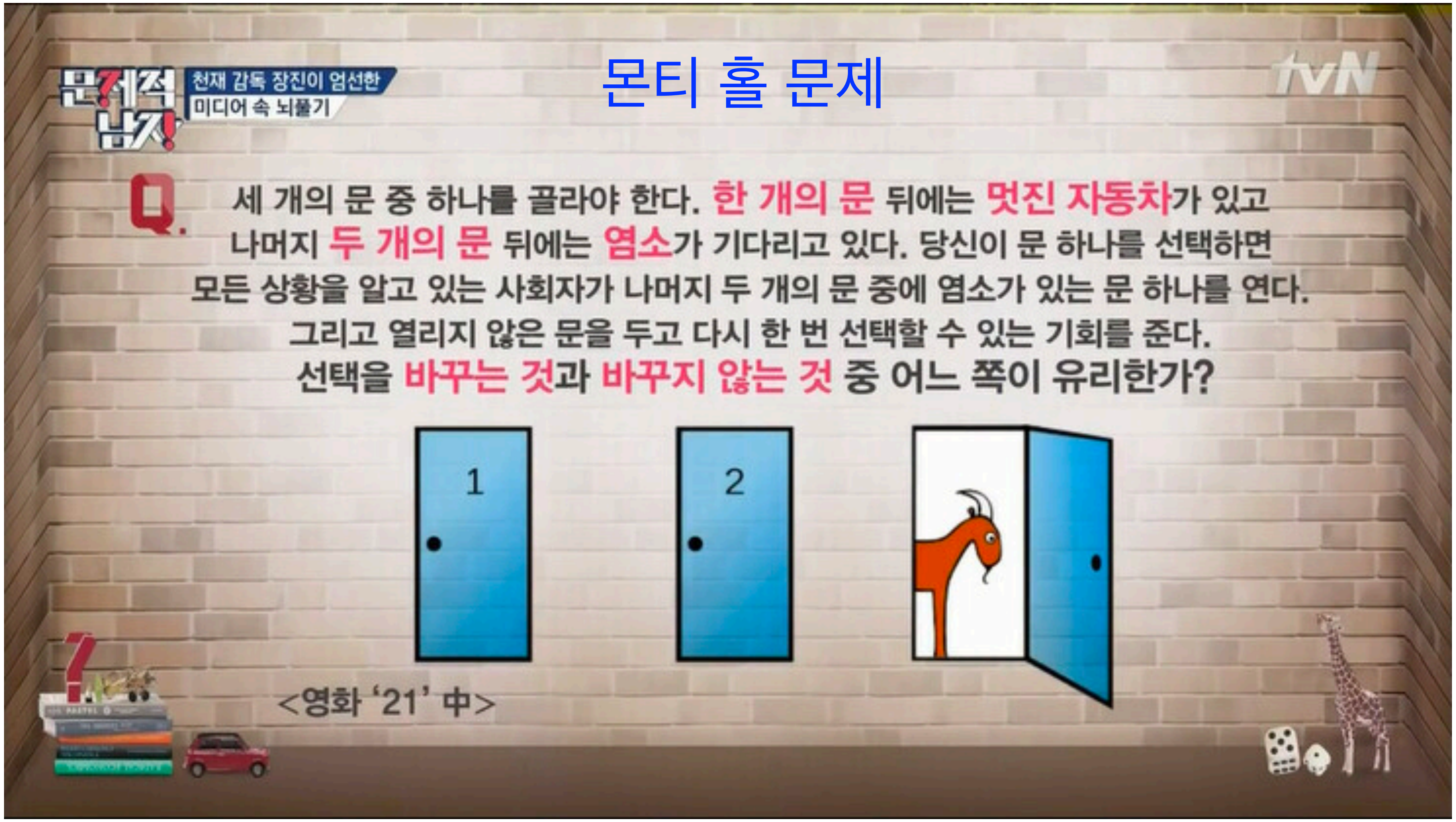
$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

사전확률(prior)
새로운 데이터(D)를 관측하기 전 가설(H)에 대한 확률

전체확률(evidence)
어떤 가설에서든 해당 데이터(D)를 관측할 확률



Bayesianist's interpretation on probability (2)

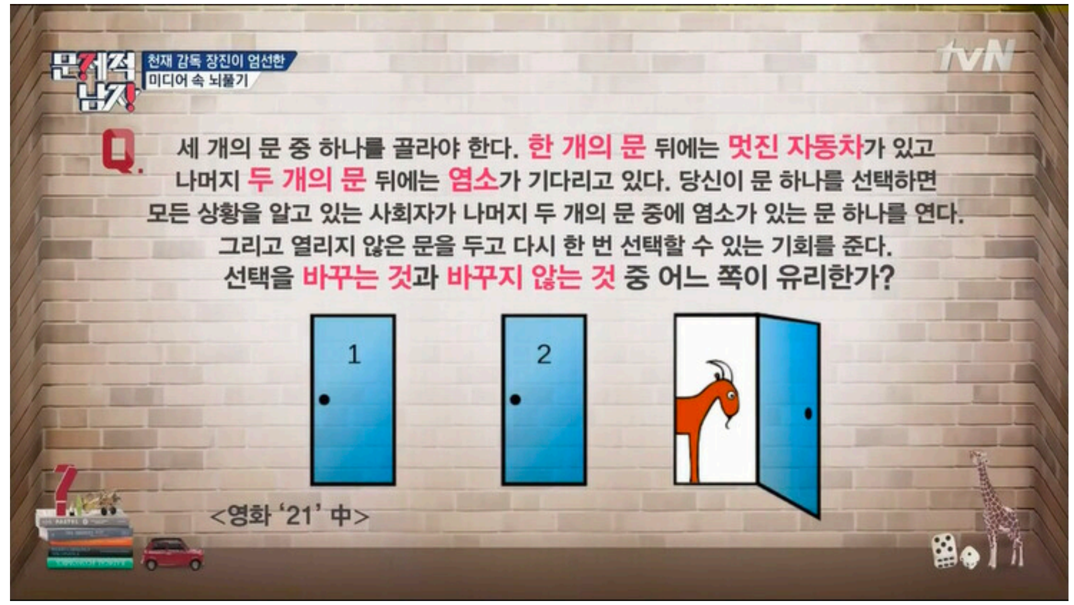


자동차가 1번 문 뒤에 있을 사전확률 (prior)
1/3

→

3번 문 열고난 후 사후확률(posterior)
1/2 (?)

Bayesianist's interpretation on probability (2)



$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

사전확률 $P(H)$: 차가 어떤 문 뒤에 있을 확률 1/3

D(데이터): 3번 문 뒤에 염소 있음

(1) H = 1번 문 뒤에 차가 있음

가능도(likelihood) $P(D|H=1)=1/2$

(2) H = 2번 문 뒤에 차가 있음

가능도(likelihood) $P(D|H=2)=1$

(3) H = 3번 문 뒤에 차가 있음

가능도(likelihood) $P(D|H=3)=0$

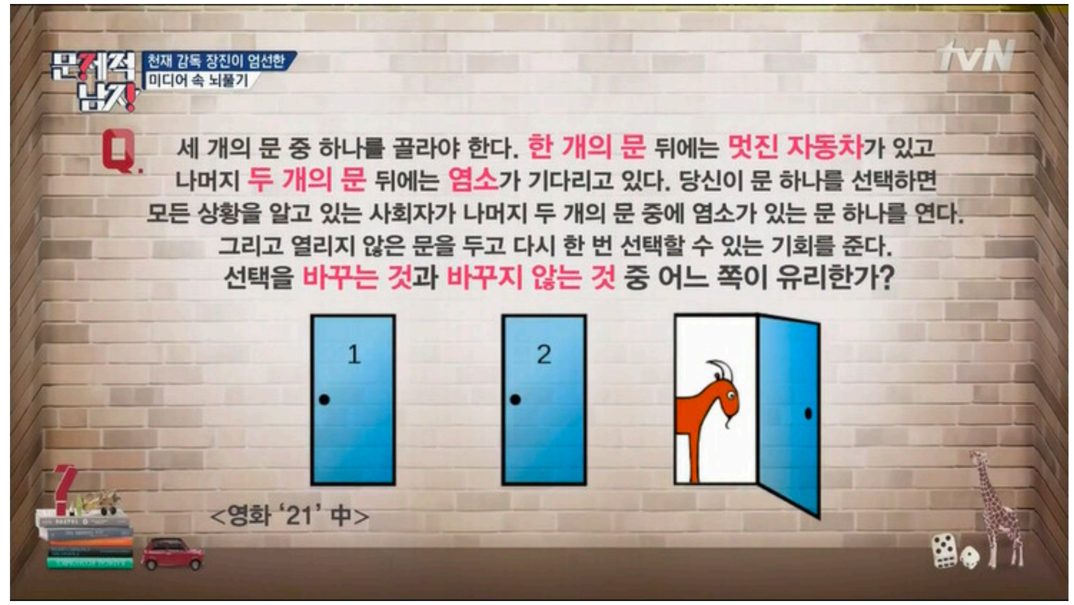
전체확률(evidence) $P(D)=[P(D|H=1)+P(D|H=2)+P(D|H=3)]P(H)=1/2$

사후확률(posterior) $P(H=1|D)=1/3$

사후확률(posterior) $P(H=2|D)=2/3$

사후확률(posterior) $P(H=3|D)=0$

Bayesianist's interpretation on probability (2)



$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

사전확률 $P(H)$: 차가 어떤 문 뒤에 있을 확률 1/3

D(데이터): 3번 문 뒤에 염소 있음

(1) H = 1번 문 뒤에 차가 있음

가능도(likelihood) $P(D|H=1)=1/2$

(2) H = 2번 문 뒤에 차가 있음

가능도(likelihood) $P(D|H=2)=1$

(3) H = 3번 문 뒤에 차가 있음

가능도(likelihood) $P(D|H=3)=0$

전체확률(evidence) $P(D)=[P(D|H=1)+P(D|H=2)+P(D|H=3)]P(H)=1/2$

사후확률(posterior) $P(H=1|D)=1/3$

사후확률(posterior) $P(H=2|D)=2/3$

사후확률(posterior) $P(H=3|D)=0$

차가 2번 문 뒤에 있을 확률이 1번 문 뒤에 있을 확률의 2배!

선택을 바꾸는 게 유리함.

ANNUAL REVIEWS

Annual Review of Biophysics

Bayesian Inference: The Comprehensive Approach to Analyzing Single-Molecule Experiments

Colin D. Kinz-Thompson,[1,2] Korak Kumar Ray,[1] and Ruben L. Gonzalez Jr.[1]

[1]Department of Chemistry, Columbia University, New York, New York 10027, USA; email: rlg2118@columbia.edu

[2]Department of Chemistry, Rutgers University-Newark, Newark, New Jersey 07102, USA

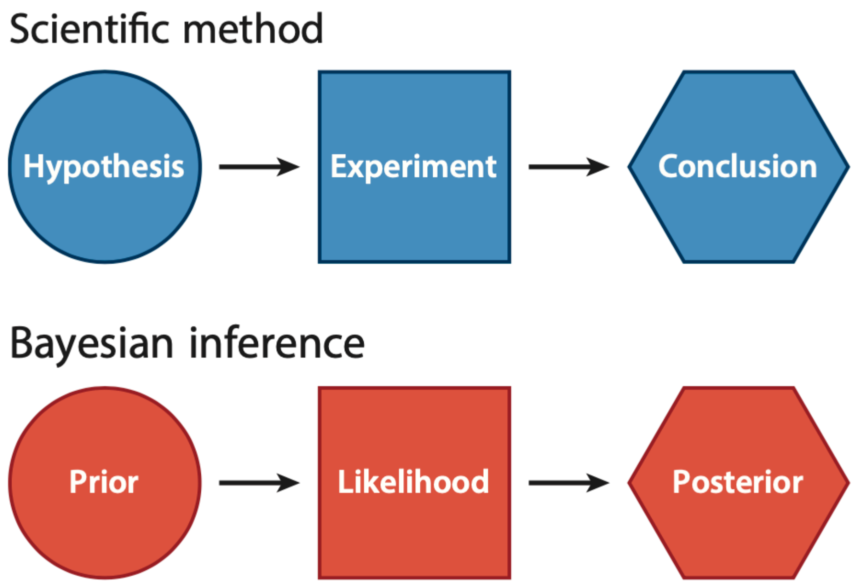
Annu. Rev. Biophys. 2021. 50:191–208

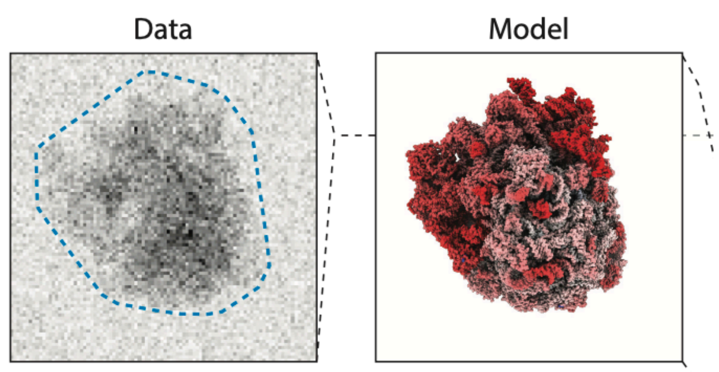
First published as a Review in Advance on February 3, 2021

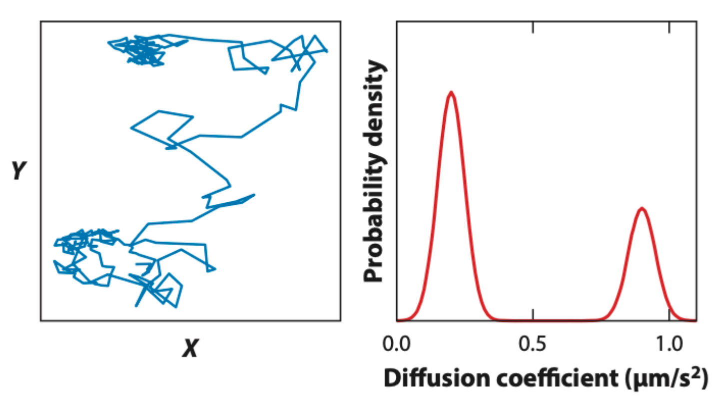
Keywords

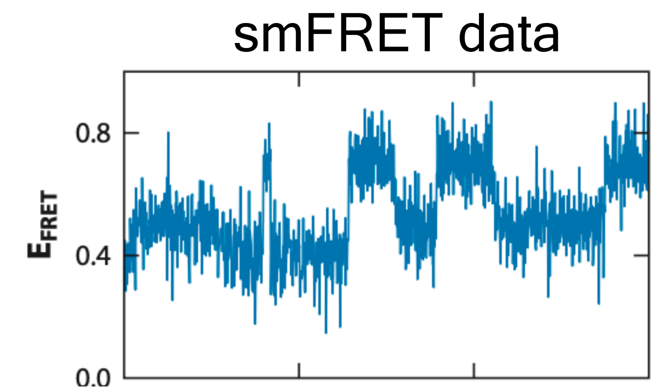
model selection, error propagation, scientific method, probability theory, kinetics, cryo-EM

Bayesian inference & applications in science

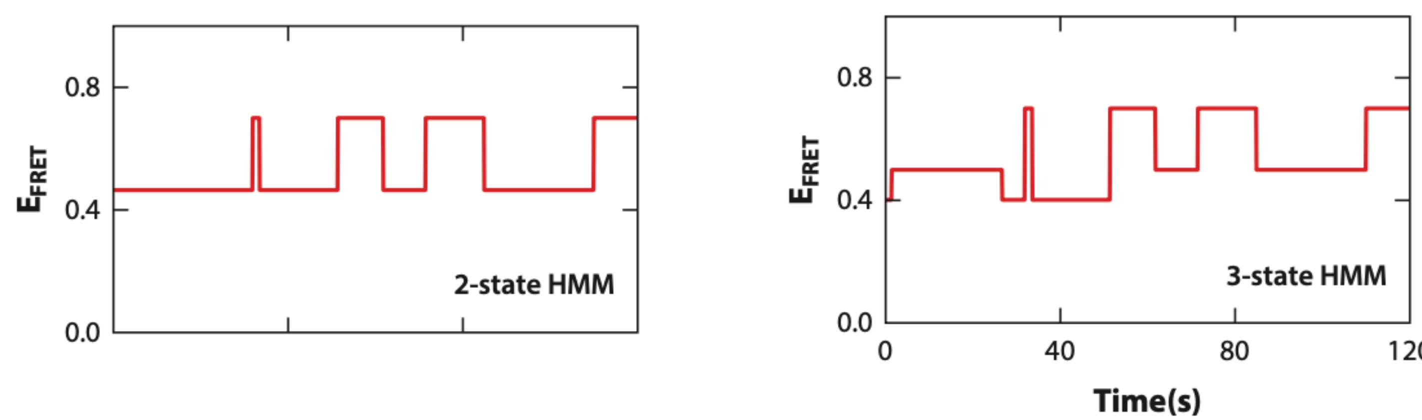


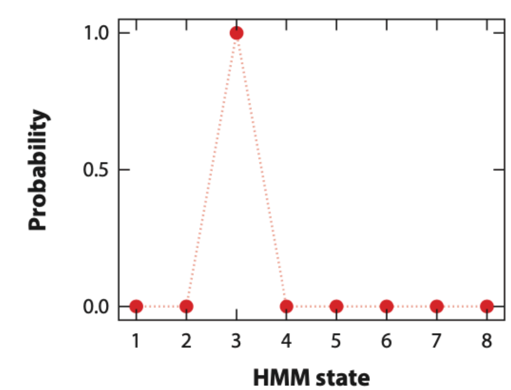






Variational Bayesian inference
(vbFRET)





Bayes' theorem—revisited

Likelihood *Prior prob*

$$P(\{\theta\}|D,M) = \frac{P(D|\{\theta\},M)P(\{\theta\}|M)}{P(D|M)}$$

Posterior prob *Evidence*

The likelihood: The prob of observing a particular value of *D* in the experiment, given that a particular model with $\{\theta\}$ is true.

$$P(D|\{\theta\},M) \equiv \mathcal{L}(\boldsymbol{\theta}|D,M) \text{ or } \mathcal{L}(D|\boldsymbol{\theta},M)$$

Maximum likelihood: $\mathcal{L}(\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}|D,M)$ where $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \mathrm{argmax}_{\theta}\mathcal{L}(\boldsymbol{\theta}|D,M)$

The prior: Knowledge about the model ($\{\theta\}$) before observing data *D*.

$$P(\{\theta_i\}|M) \equiv \pi(\boldsymbol{\theta})$$

Bayes' theorem—revisited

Likelihood *Prior prob*

$$P(\{\theta\}|D,M) = \frac{P(D|\{\theta\},M)P(\{\theta\}|M)}{P(D|M)}$$

Posterior prob *Evidence*

The posterior: The prob that the given model with the parameter $\{\theta\}$ is true under the new observation D.

The evidence (=marginal likelihood): The prob that the observed D comes from the M, regardless of the parameter $\{\theta\}$.

$$P(D|M) = \int_{\Theta} \mathcal{L}(\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}|D,M)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

Bayesian model inference

There are *N* possible distinct models {M_i} that explain the data *D*.

Evidence Prior

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)}$$

The prob that the model M_i is true among *N* possible models

$$\frac{P(M_i|D)}{\sum_j P(M_j|D)} = \frac{P(D|M_i)}{\sum_j P(D|M_j)}$$

if the prior is uniform $P(M_i) = P(M_j)$

Evidence

$$P(D|M_i) = \int P(D|\{\theta\}, M_i)P(\{\theta\}|M_i)d\{\theta\}$$

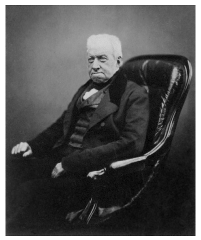
Bayesian inference for model parameters

Find the best parameter set for a given data set *D* if a specific model (M_i) is given.

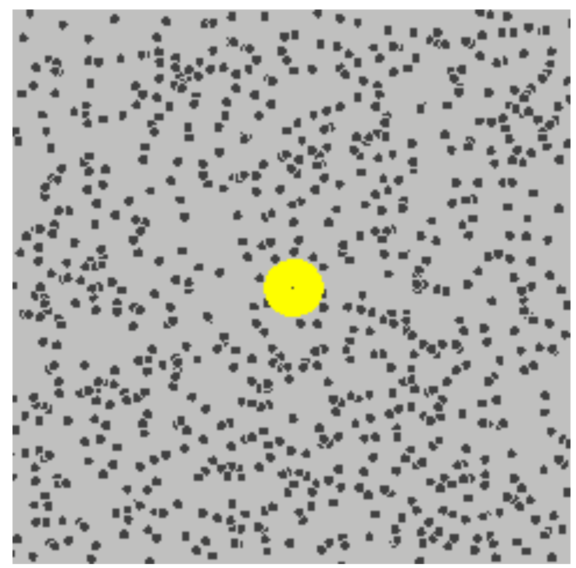
$$P(\{\theta_i\}|D, M_i) = \frac{P(D|\{\theta_i\}, M_i)P(\{\theta_i\}|M_i)}{P(D|M_i)}$$

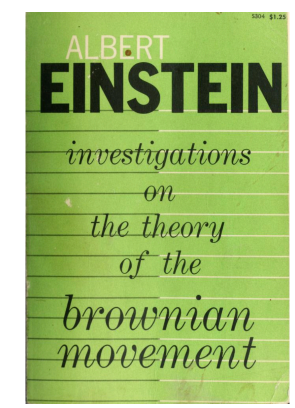
Brownian motion & diffusivity

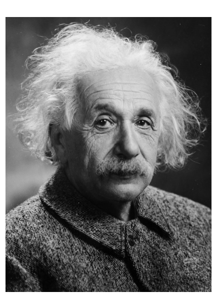
Irregular & incessant motion of pollen grains



R. Brown (1827), A brief account of microscopical observations on the particles contained in the pollen of plants and on the general existence of active molecules in organic and inorganic bodies







A. Einstein (1905), Investigations on the theory of the Brownian movement

On the movement of small particles suspended in a stationary liquid demanded by THE MOLECULAR-KINETIC THEORY OF HEAT

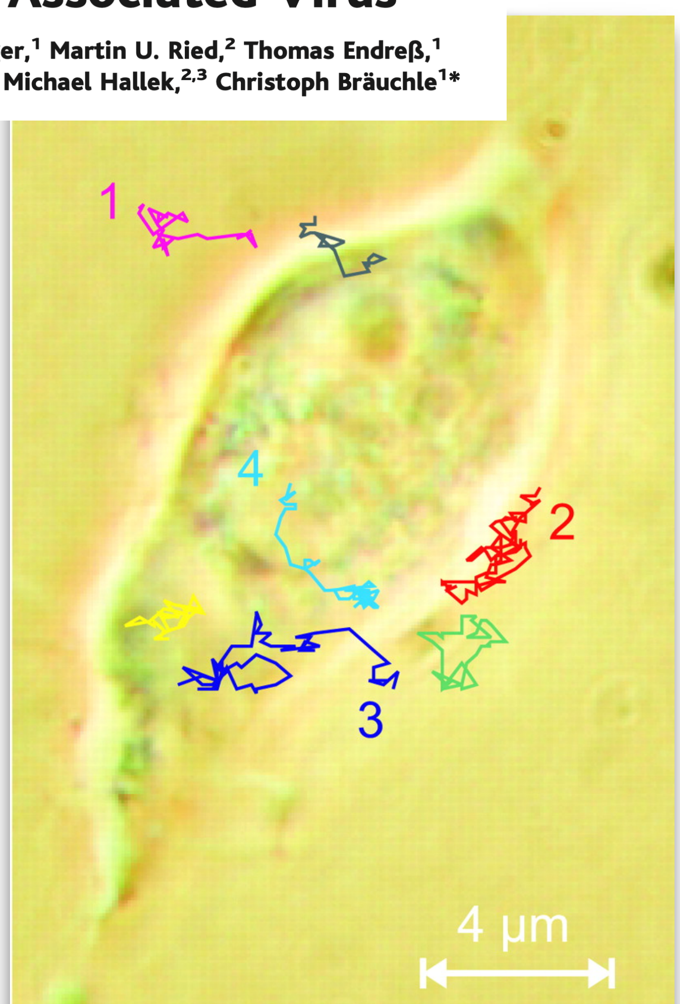
Mean-squared displacement (MSD)

$$\langle x^2(t)\rangle = 2\mathcal{D}t$$

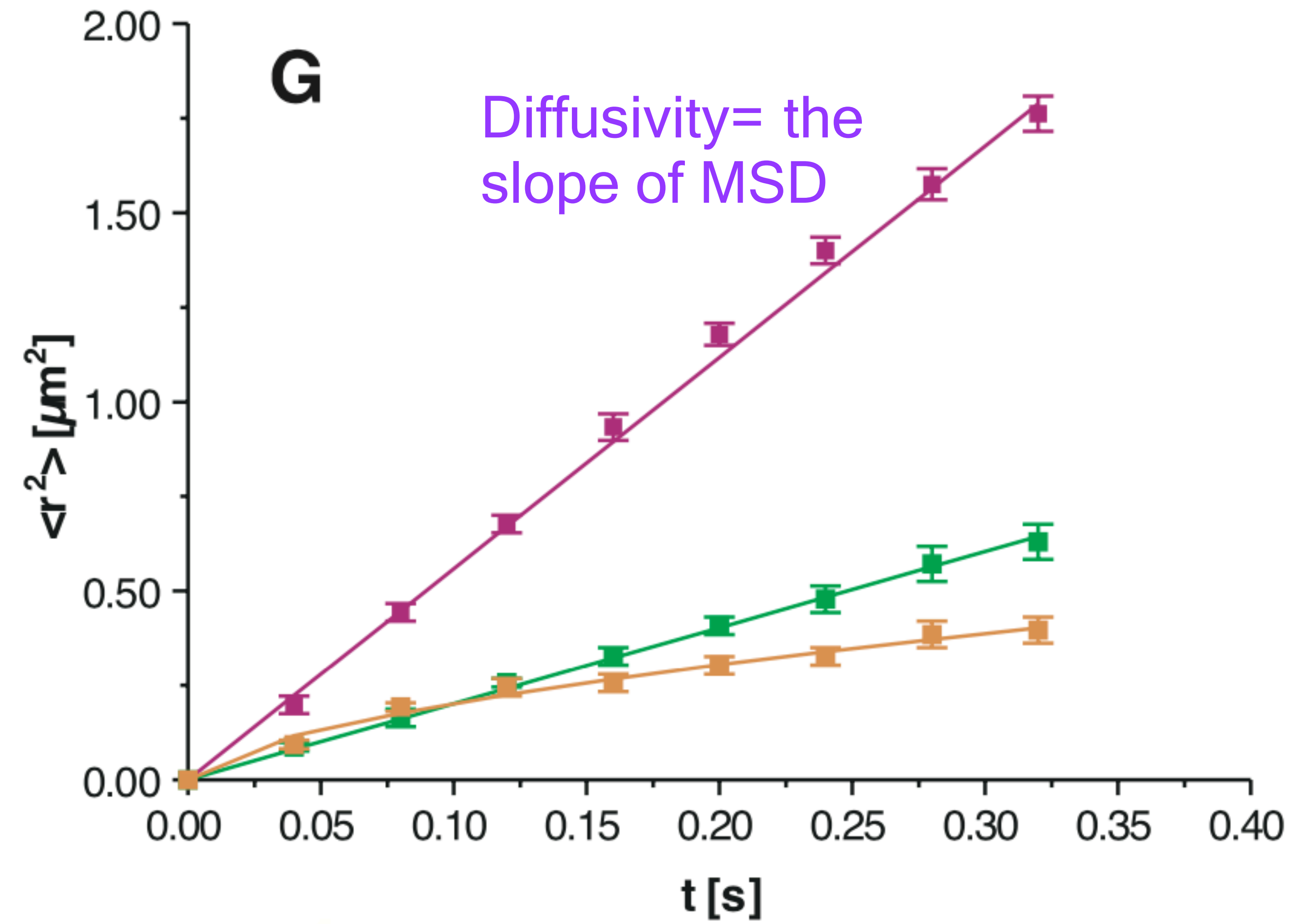
Diffusivity $\mathcal{D} = \dfrac{k_B T}{6\pi\eta r}$

Real-Time Single-Molecule Imaging of the Infection Pathway of an Adeno-Associated Virus

Georg Seisenberger,[1] Martin U. Ried,[2] Thomas Endreß,[1] Hildegard Büning,[2] Michael Hallek,[2,3] Christoph Bräuchle[1]*



Seisenberger et al., Science **294**, 1929 (2001)



<Model> Stochastic theory for Brownian particles

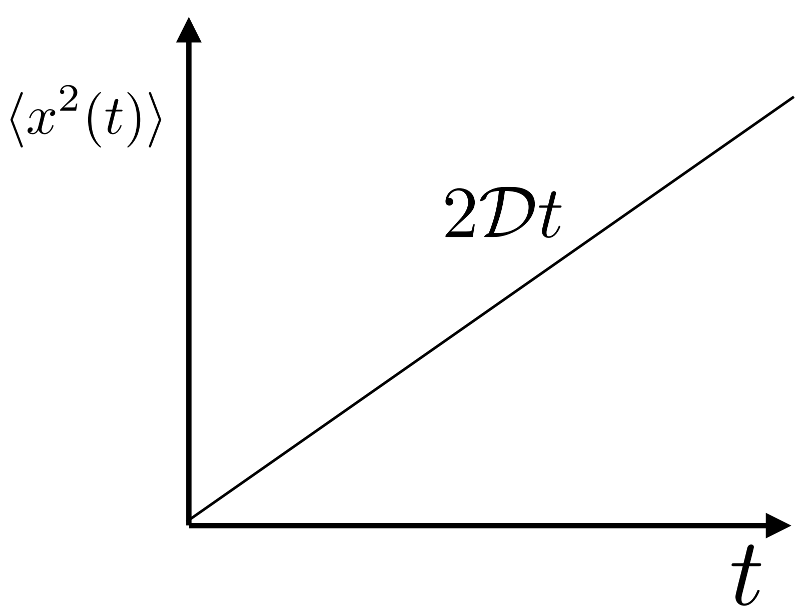
The over-damped Langevin equation for a Brownian particle

$$\gamma \frac{d}{dt} x(t) = \sqrt{2\gamma k_B T}\,\xi(t)$$

frictional force (deterministic)

random force by thermal fluc. (stochastic)

$$\langle x^2(t) \rangle = 2\frac{k_B T}{\gamma} t$$



The Gaussian noise *P*(ξ) having the properties:

$$\langle \xi(t) \rangle = 0$$

$$\langle \xi(t)\xi(t') \rangle = \delta(t - t')$$

Noise has *no memory*.

<Model> Stochastic theory for Brownian particles

The Smoluchowski equation for a Brownian particle

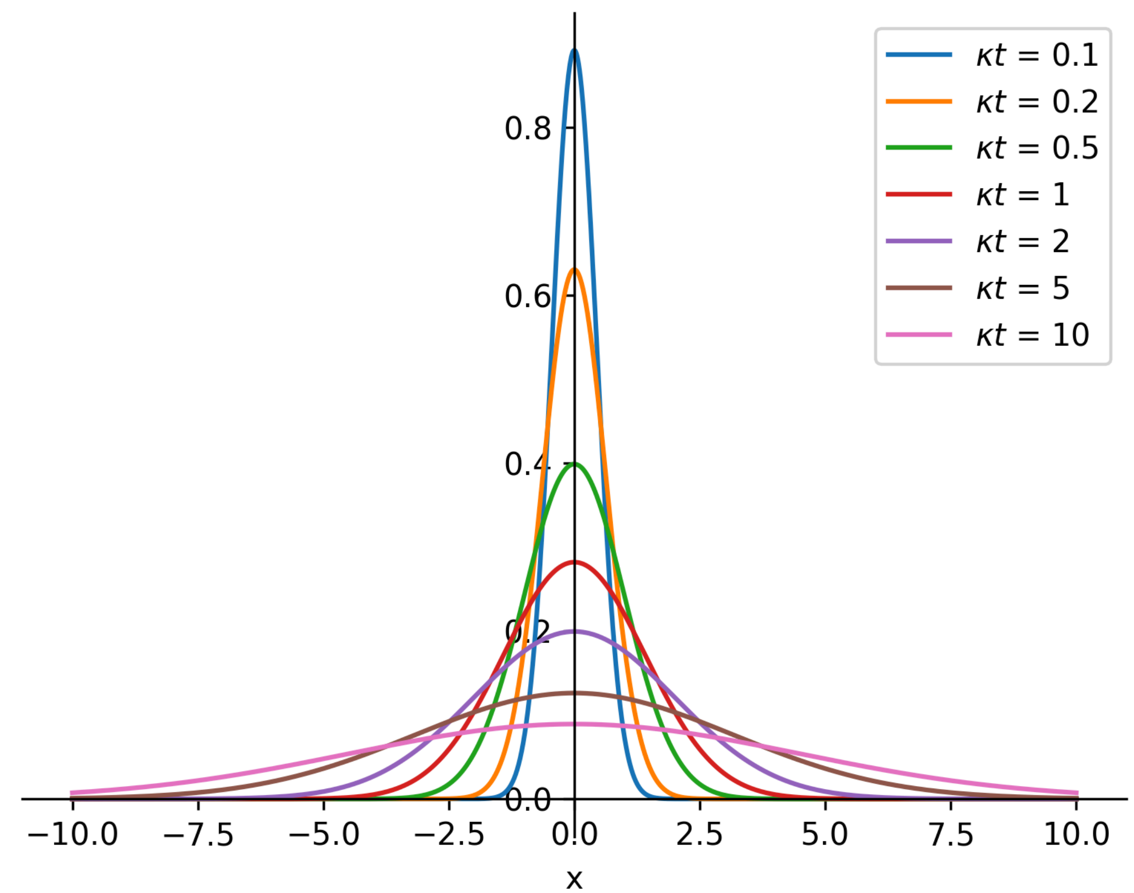
$$\frac{\partial}{\partial t}p(x,t) = \mathcal{D}\frac{\partial}{\partial x}\left(\frac{1}{k_B T}\frac{\partial U(x)}{\partial x} + \frac{\partial}{\partial x}\right)p(x,t)$$

In free space (U=0):

$$p(x,t|x_0) = \frac{1}{\sqrt{4\pi\mathcal{D}t}}e^{-\frac{(x-x_0)^2}{4\mathcal{D}t}}$$

$$= \mathcal{N}(x_0, 2\mathcal{D}t)$$

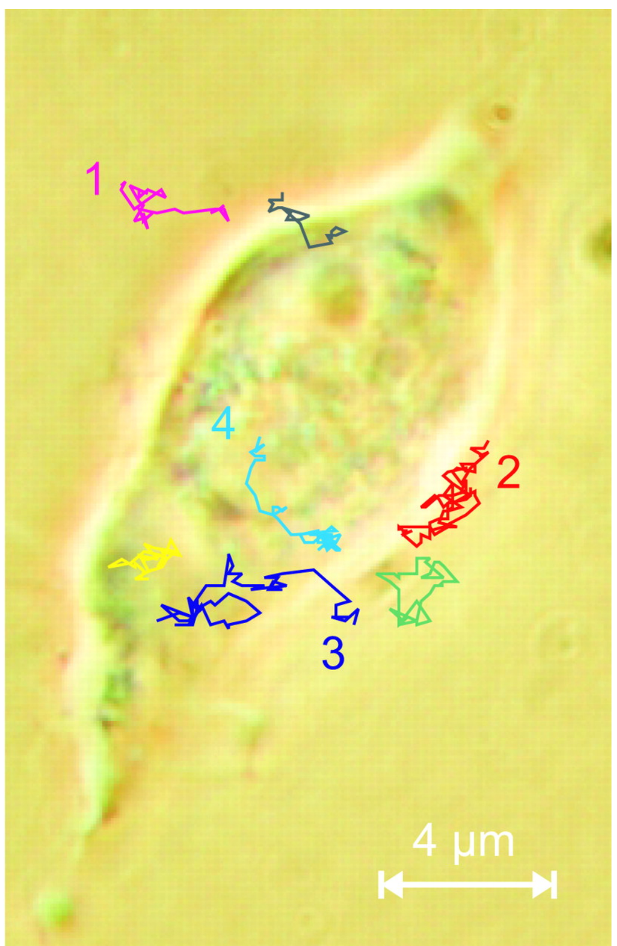
Normal distribution — Mean — Variance



Example: Bayesian inference for diffusivity

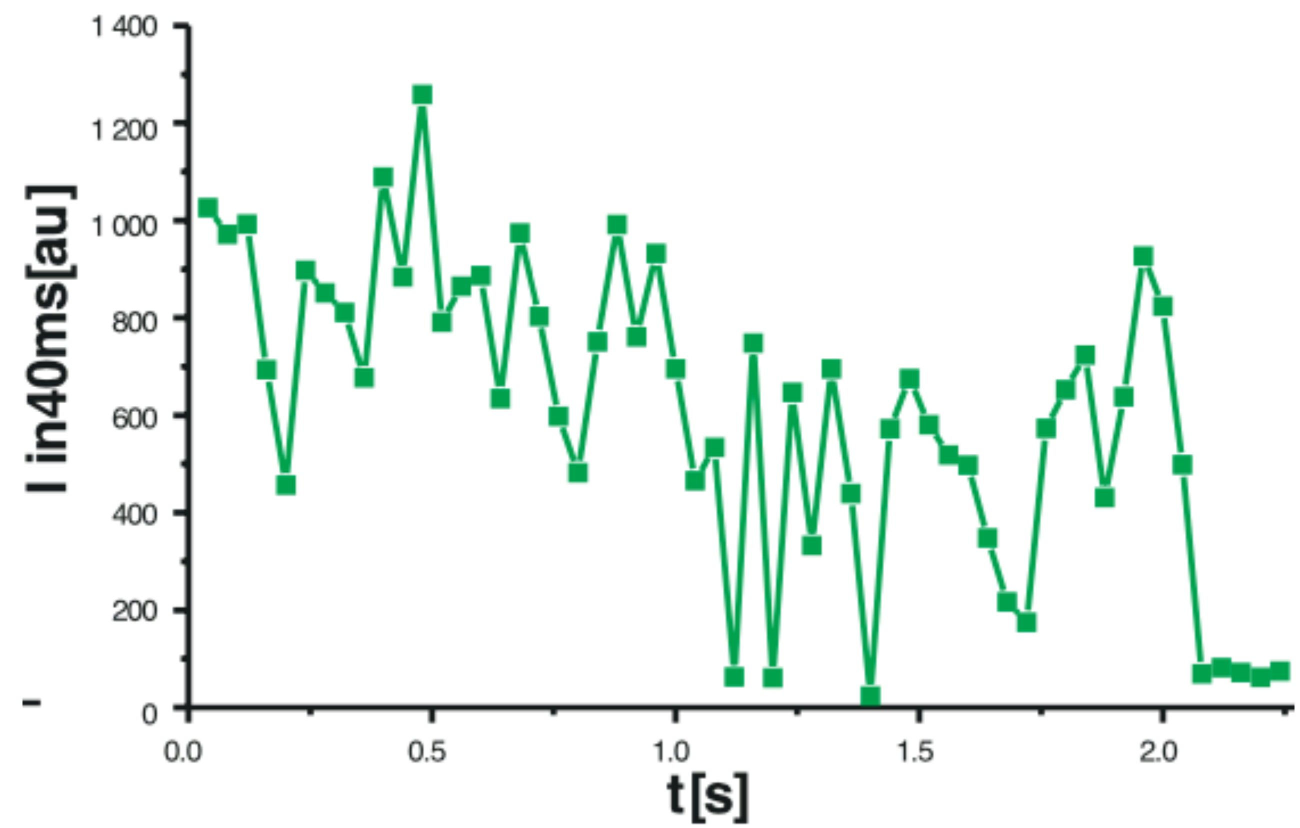
What is the diffusivity of biomolecules (e.g., viruses) exploring the intracellular fluid in an experiment?

Single-particle experiments



Data

Time series: $\{\Delta x_1,\ \Delta x_2, \ldots, \Delta x_N\}$



Example: Bayesian inference for diffusivity

$$P(\mathcal{D}|\{\Delta x_i\}, \mathrm{BM}) = \frac{P(\{\Delta x_i\}|\mathcal{D}, \mathrm{BM})P(\mathcal{D}|\mathrm{BM})}{P(\{\Delta x_i\}|\mathrm{BM})}$$

The most probable diffusivity from the Bayesian inference is the one maximizing the Posterior probability

The likelihood function, a Gaussian weight, is obtained from the Langevin model for Brownian motion

$$P(\{\Delta x_i\}|\mathcal{D}, \mathrm{BM}) = \left(\frac{1}{\sqrt{4\pi\mathcal{D}\Delta t}}\right)^N e^{-\sum_{j=1}^{N}\frac{(\Delta x_j)^2}{4\mathcal{D}\Delta t}}$$

The prior prob: based on our experience, we may look for the diffusivity in an interval

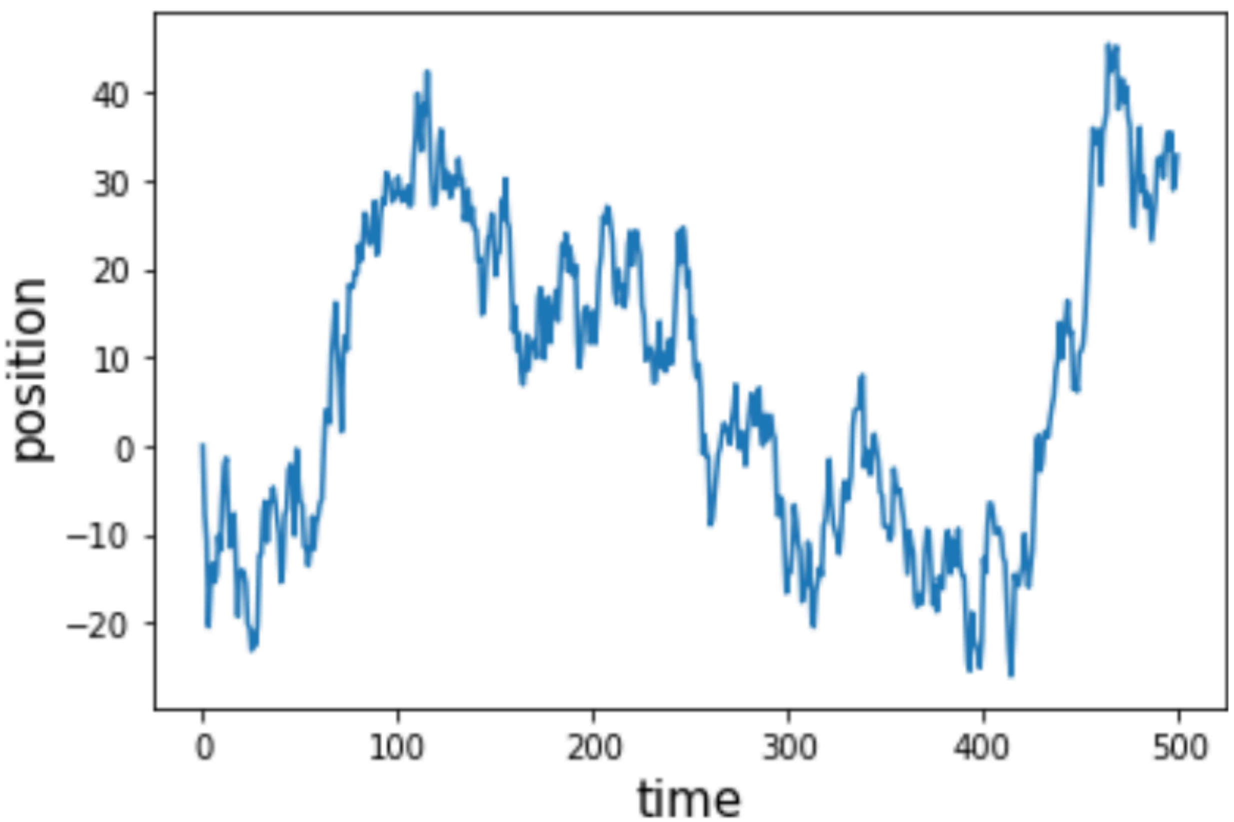
$$P(\mathcal{D}|\mathrm{BM}) = \frac{1}{\mathcal{D}_M - \mathcal{D}_m} \quad \text{in } [\mathcal{D}_m,\ \mathcal{D}_M]$$

The evidence: marginalization of the likelihood function over the prior

$$P(\{\Delta x_i\}|\mathrm{BM}) = \int_0^{\mathcal{D}_M} d\mathcal{D}\left(\frac{1}{\sqrt{4\pi\mathcal{D}\Delta t}}\right)^N e^{-\sum_{j=1}^{N}\frac{(\Delta x_j)^2}{4\mathcal{D}\Delta t}}/\mathcal{D}_M$$

A simulated sample trajectory with a model parameter

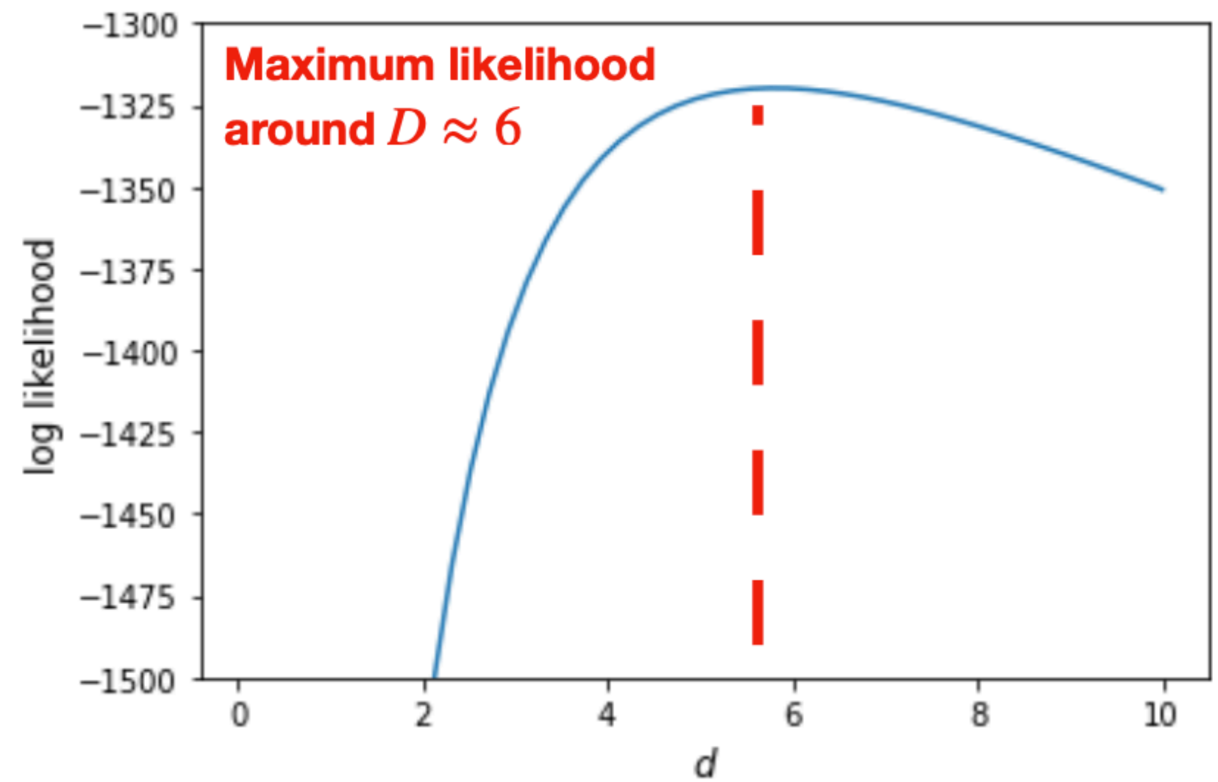
$$x_t \sim N(0, \sqrt{2D}) \quad D = 6$$



Likelihood function

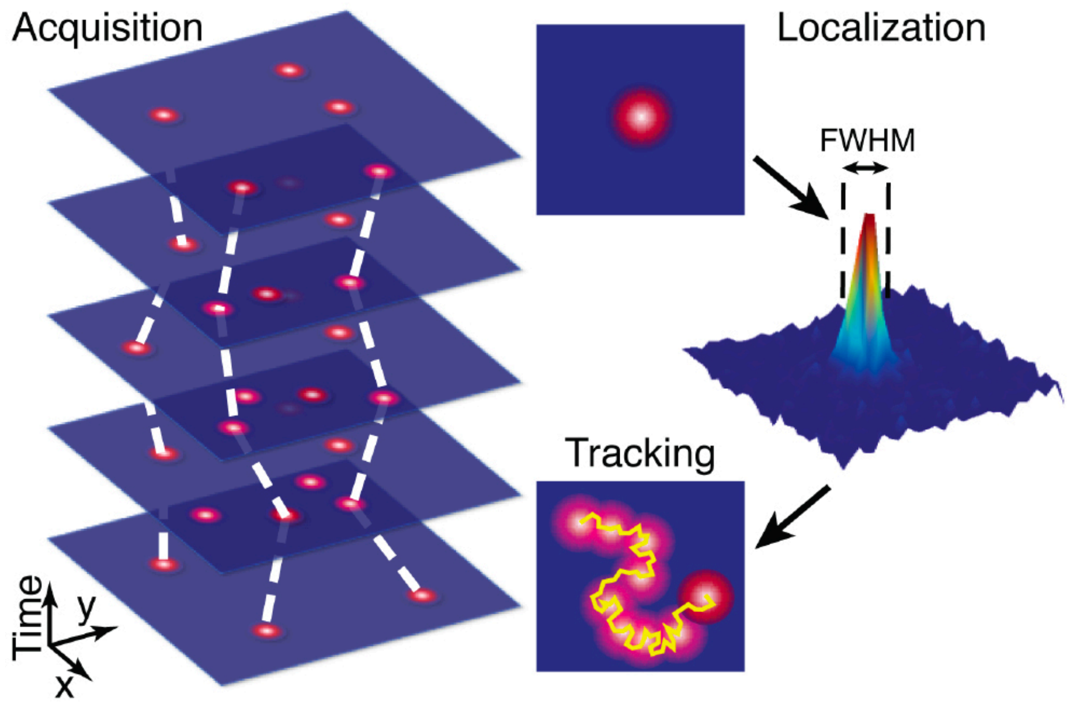
Log likelihood function of one-state BM model

$$\log \mathscr{L}(D = \frac{1}{2}\sigma^2) = \sum_{t=1}^{T} \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{x_t^2}{2\sigma^2} \right]$$



Bayesian inference for particle's diffusivity under finite experimental resolution

Single-particle tracking (SPT) & super-resolution imaging

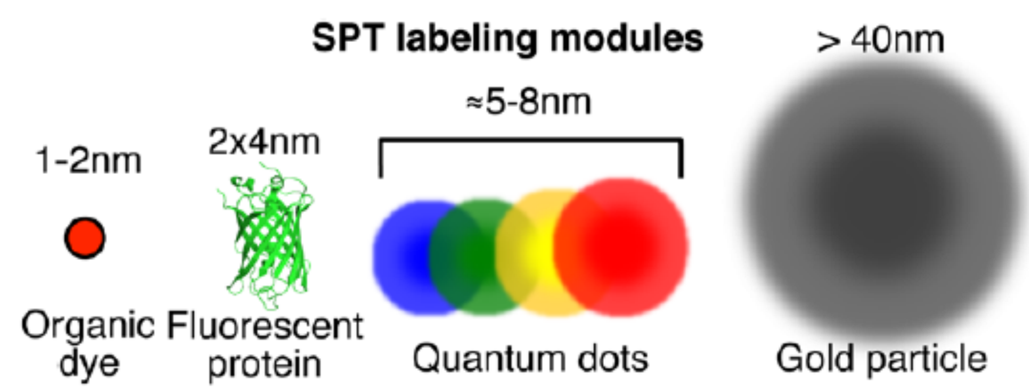


Review: Manzo and Carcia-Parajo, Rep. Prog. Phys. **78**, 124601 (2015)

Trajectory $x(t)$ of individual particles

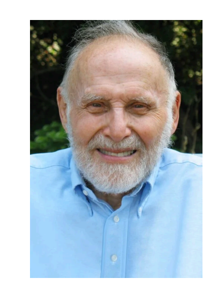
- Time resolution ~ from tens of microseconds to miliseconds

- Spatial resolution ~ tens or hundreds of nanometers



Nobel prize (2014): E. Betzig/S. Hell/W. Moerner for super-resolved fluorescence microscopy
Nobel prize (2018): A. Ashkin for the optical tweezers and their applications to biological systems



Bayesian inference for particle's diffusivity under finite experimental resolution

The likelihood function in the presence of experimental noises

$$\Delta x_i^{(\mathrm{ob})} = \Delta x_i + \eta_i$$

The noise is a Gaussian random variable: $\eta_i \sim \mathcal{N}(0, \sigma^2_{\mathrm{noise}})$

$$\rightarrow \quad \mathcal{L}(\{x_i^{(\mathrm{ob})}\}|\vec{\theta}) = \Pi_j \frac{\exp\left(-\frac{[x_j^{(\mathrm{ob})} - \tilde{x}_j]^2}{2\tilde{\sigma}_j^2}\right)}{\sqrt{2\pi\tilde{\sigma}_j^2}}$$

$\vec{\theta} = \{\sigma^2, \ \sigma^2_{\mathrm{noise}}\}$

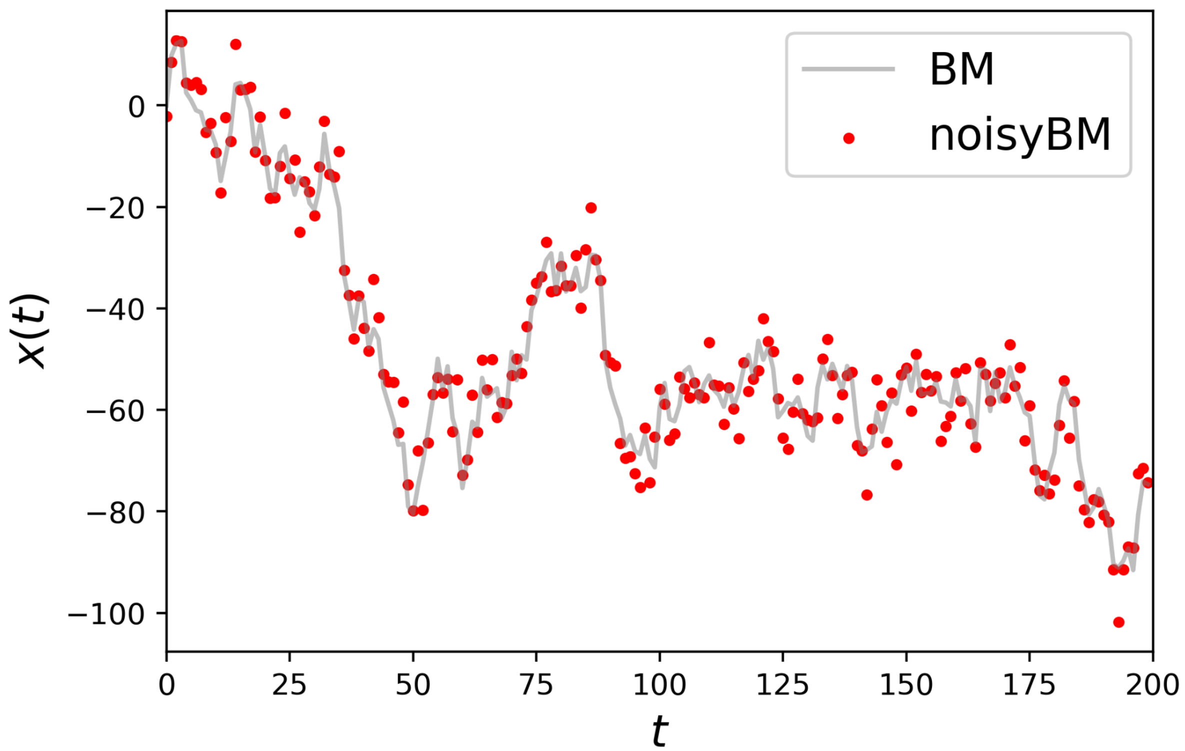
$$\tilde{x}_{j+1} = x_j - \frac{\sigma^2_{\mathrm{noise}}}{\tilde{\sigma}_j}(x_j - \tilde{x}_j)$$

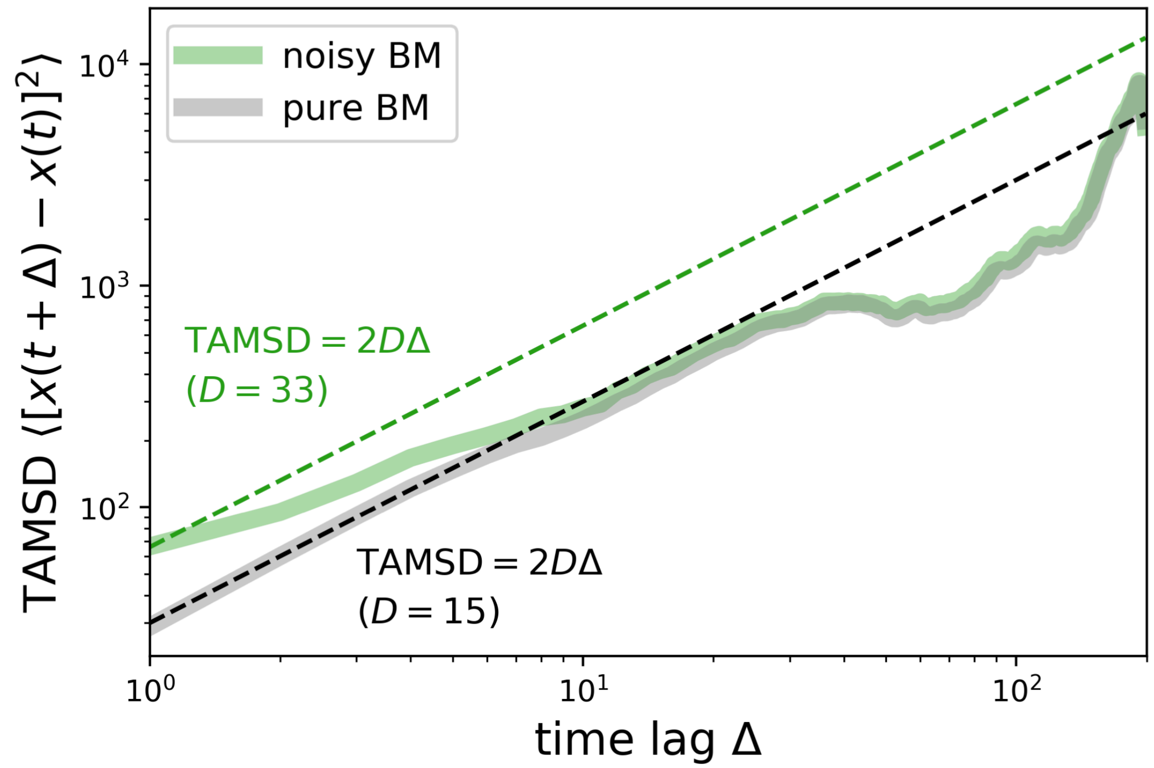
$$\tilde{\sigma}_j^2 = \sigma^2 + \sigma^2_{\mathrm{noise}}\left(2 - \frac{\sigma^2_{\mathrm{noise}}}{\tilde{\sigma}_j^2}\right)$$

Numerical test with noisy Brownian motion

M_0: Brownian motion (BM)

M_1: *Noisy* Brownian motion (NBM)





Simulation parameters: $D = 15, \sigma_{\text{Noise}} = 4$

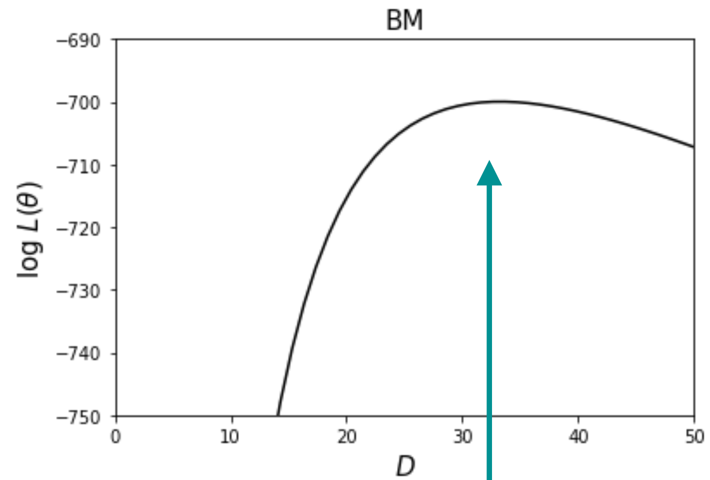
$$\sigma = \sqrt{2Dt_0}$$

$$t_0 = 1$$

Numerical test with noisy Brownian motion (2)

M0

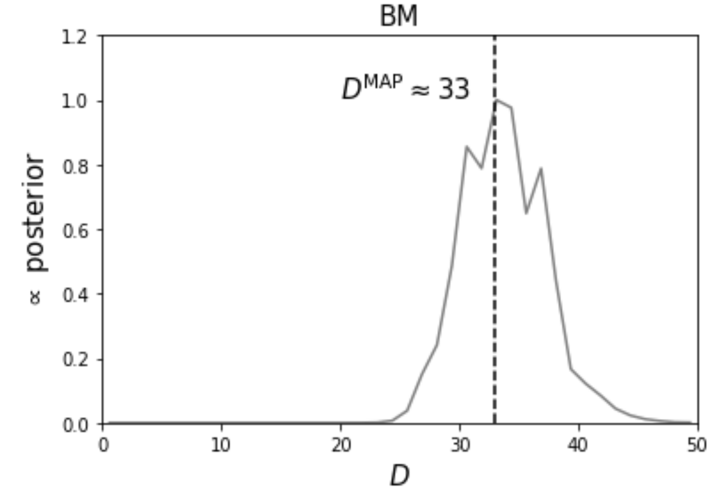
Log likelihood



+ Prior distribution (uniform)

Nested Sampling

Posterior

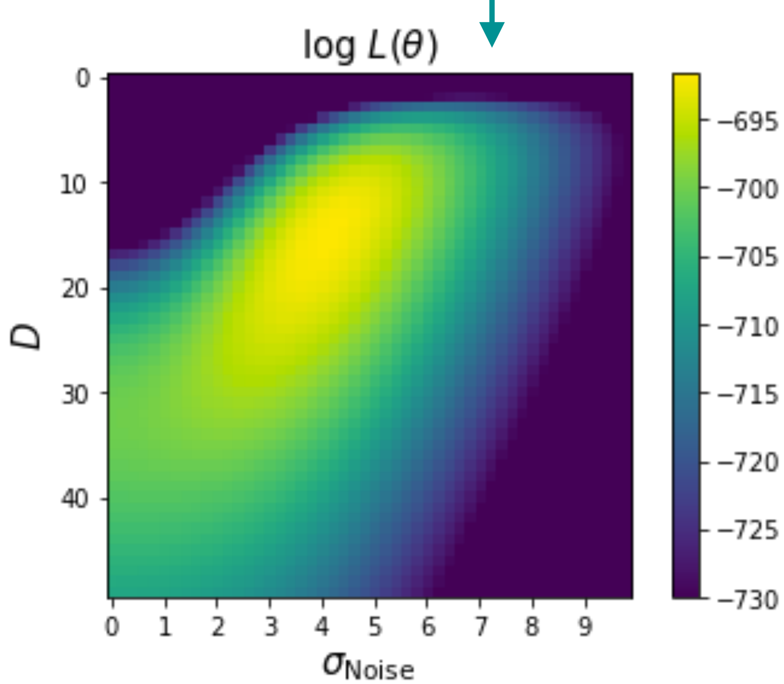


Data: *noisy* BM with D=15 & σ_{Noise}=4

$$\sigma = \sqrt{2Dt_0} \ \& \ t_0 = 1$$

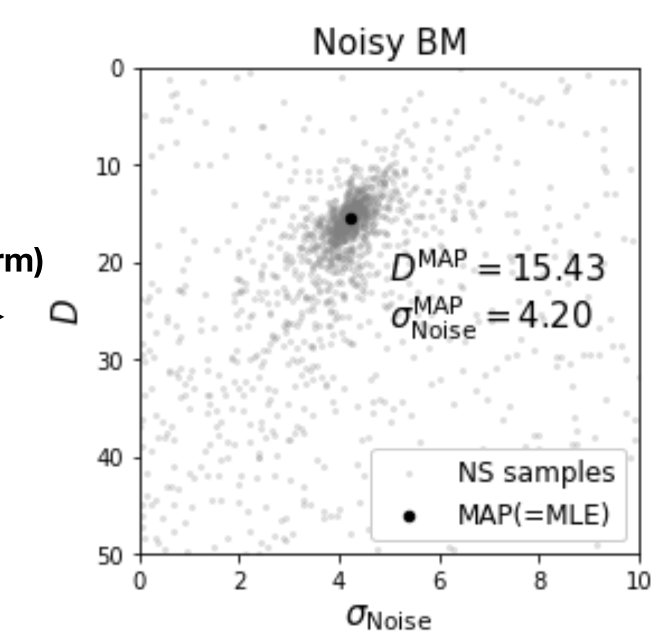
$$D_{\text{eff}}^{\text{Theory}} = \frac{1}{2t_0}(\sigma^2 + 2\sigma_{\text{Noise}}^2) = 31$$

M1



+ Prior distribution (uniform)

Nested Sampling



Model evidence

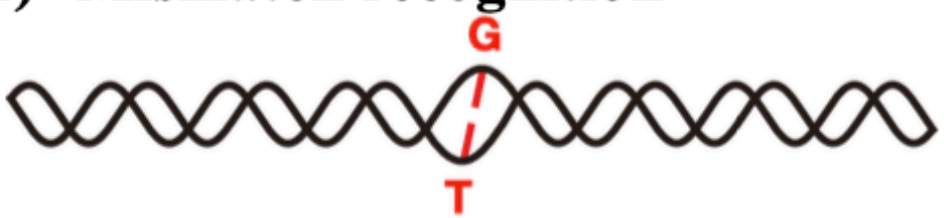
$\log P(\text{Data}\,|\,M_0) = -701.72$

$\log P(\text{Data}\,|\,M_1) = -695.70$

$$\rightarrow \frac{P(M_1\,|\,\text{Data})}{P(M_0\,|\,\text{Data})} = 410.27$$

Example: Identifying hidden diffusion states of MutS homolog proteins in DNA repair processes

(1) Mismatch recognition



- Recognition by MutS homologs (MSH)

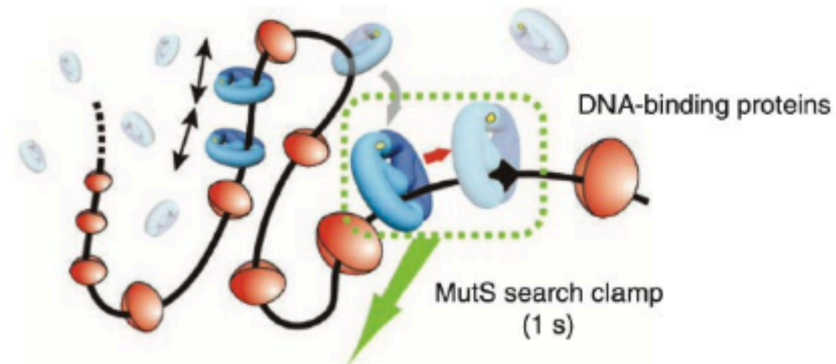
(2) Mismatch removal

- Incision by MutL homologs (MLH)
- Excision by exonuclease

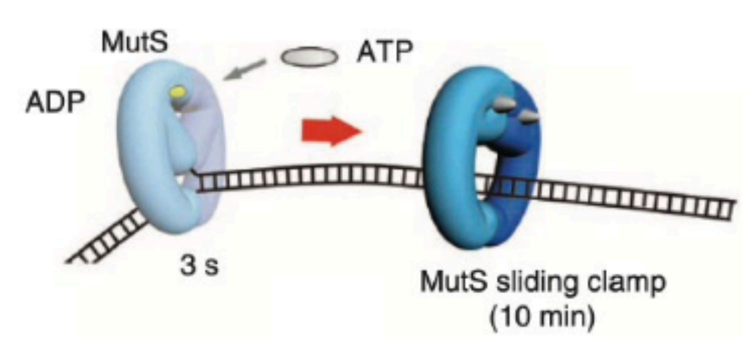
(3) DNA re-synthesis

- Re-synthesis by DNA polymerase
- DNA ligation

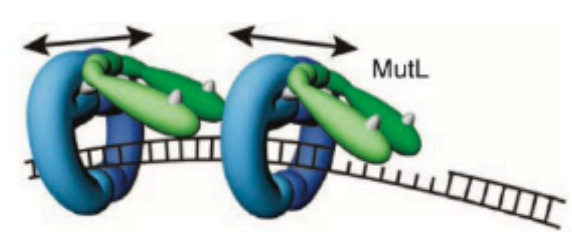
a. MSH searching clamp recognizes a mismatch.



b. ADP → ATP exchange of mismatch-bound MSH (sliding clamp)



c. MSH sliding clamp loads MLH to mediate subsequent MMR process



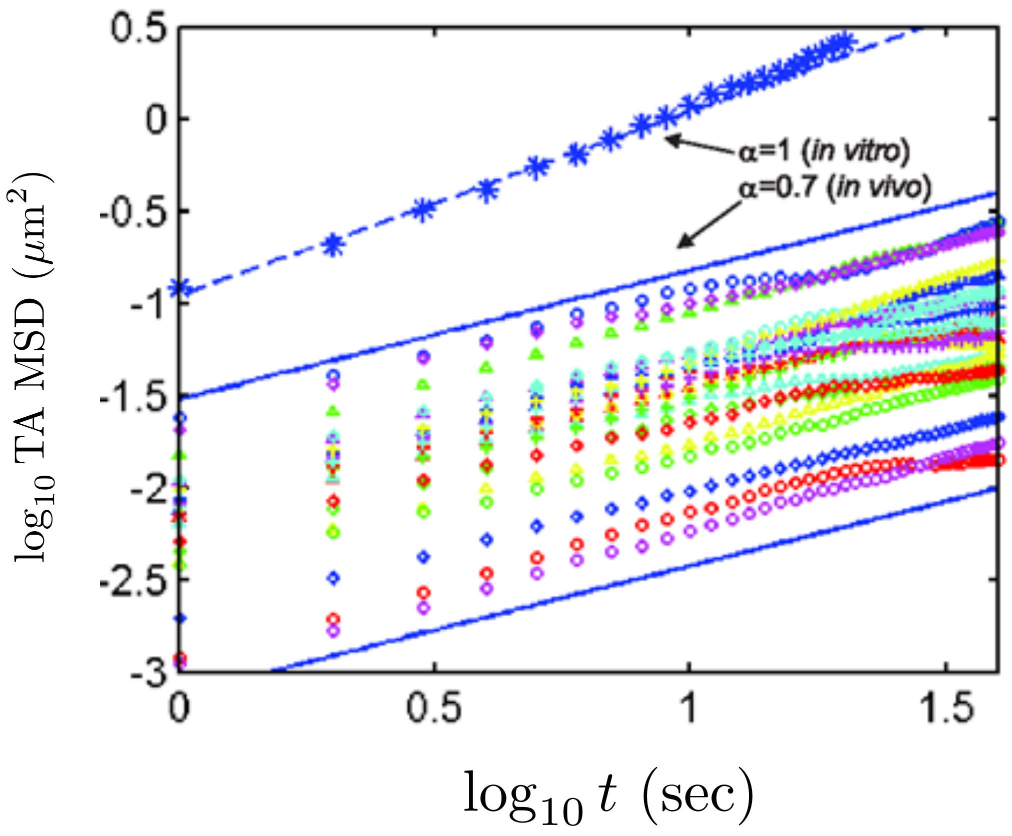
C. Jeong *et al.*, *Nat. Struct. Mol. Biol.* (2009)

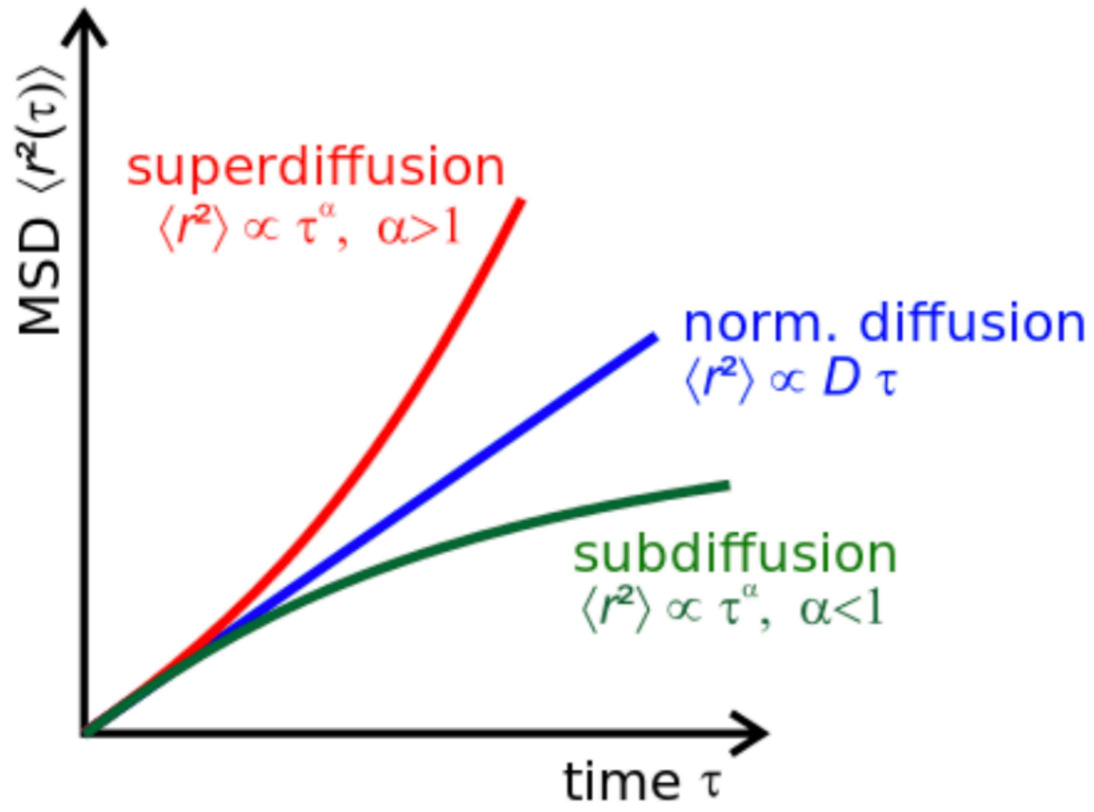
In vivo anomalous diffusion (1)

$$\langle [x(t+\Delta) - x(t)]^2 \rangle_t \simeq 2 D_\alpha \Delta^\alpha$$

Generalized diffusivity

anomaly exponent $0 < \alpha \lesssim 2$





In vivo anomalous diffusion (2)

$$\langle x^2(t)\rangle \propto t^\alpha \ (\alpha \neq 1)$$

Possible physical mechanisms

Molecular crowding

Random obstacles

Fractal space

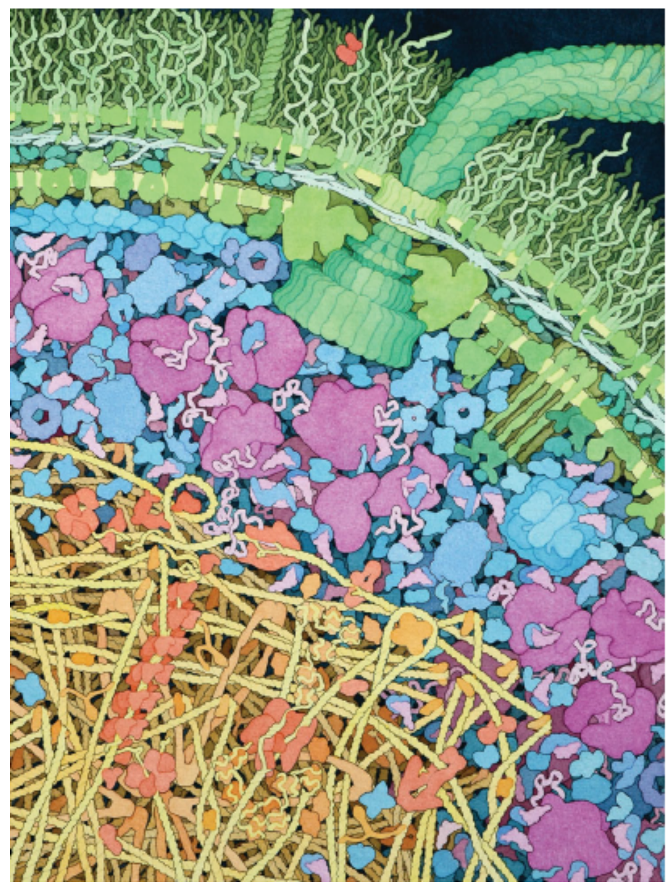
Polymeric environment

Viscoelastic response

Confinement, trap

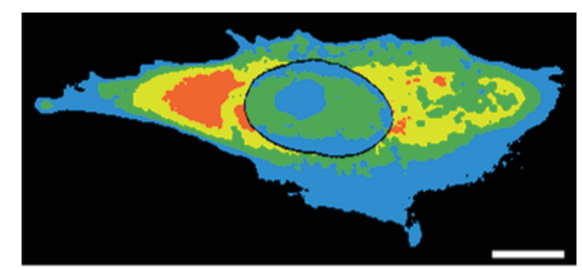
Non-specific interactions

Corrals



Non-equilibrium forces /fluctuations

Motor-driven active motion



Spatiotemporal inhomogeneity

Random energy landscape

Space- or time-dependent diffusivity

Metzler, JHJ, Cherstvy, & Barkai, PCCP (2014); Song, Moon, JHJ, & Park, Nat Commun (2018); JHJ et al., PRX (2016)

Phys. Chem. Chem. Phys.

This journal is © the Owner Societies 2014

PERSPECTIVE

View Article Online
View Journal

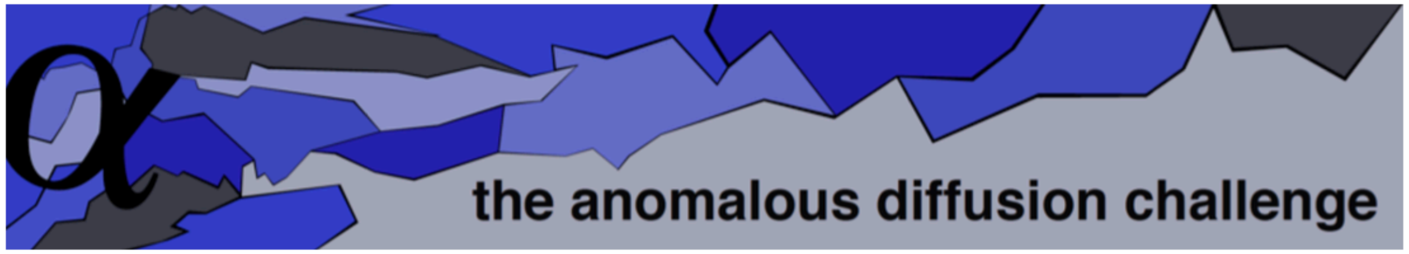
CrossMark
click for updates

Cite this: DOI: 10.1039/c4cp03465a

Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking

Ralf Metzler,*[ab] Jae-Hyung Jeon,[bc] Andrey G. Cherstvy[a] and Eli Barkai[d]

Process	WEB	$\langle x^2(t)\rangle$	$\left\langle\overline{\delta^2(\Delta)}\right\rangle$	Eqn	Ref.
Correlated jump lengths	Yes	$\simeq t^3$	$\simeq \Delta^2 t$	(48) and (49)	120
Lévy walk, $0 < \alpha < 1$	Yes	$\simeq A(\alpha)t^2$	$\simeq \dfrac{A(\alpha)}{1-\alpha}\Delta^2$	(50) and (51)	136 and 137
Lévy walk, $1 < \alpha < 2$	Yes	$\simeq A^*(\alpha)t^{3-\alpha}$	$\simeq \dfrac{A^*(\alpha)}{\alpha-1}\Delta^{3-\alpha}$	(50) and (52)	83, 136 and 138
Lévy flight	Yes	$=\infty[\langle\lvert x\rvert^q\rangle^{2/q} \simeq t^{2/\alpha}]$	$\simeq \Delta t^{2/\alpha-1}$		129, 137 and 275[a]
FBM $0 < \alpha < 2$	No	$\simeq t^\alpha$	$\simeq \Delta^\alpha$	(58)	156, 166, 176 and 276
Brownian motion	No	$\simeq t$	$\simeq \Delta$	(3) and (12)	44, 277 and 278
FLE motion $0 < \alpha < 1$	No	$\simeq t^\alpha$	$\simeq \Delta^\alpha$	(66)	156, 166 and 176
Fractal environment	No	$\simeq t^{2/d_w}$	$\simeq \Delta^{2/d_w}$		50 and 218
HDP $K(x) = K_0\lvert x\rvert^\beta$	Yes	$\simeq t^{2/(2-\beta)}$	$\simeq \Delta t^{2/(2-\beta)-1}$	(90), (91) and (93)	197 and 203
Correlated waiting times	Yes	$\simeq t^{\gamma/(1+\gamma)}$	$\simeq \Delta t^{\gamma/(1+\gamma)-1}$	(8), (46) and (47)	120–122
Subdiffusive CTRW	Yes	$\simeq t^\alpha$	$\simeq \Delta t^{\alpha-1}$	(8) and (20)	44, 63 and 64
Confined subdiffusive CTRW	Yes	$\simeq t^0$	$\simeq (\Delta/t)^{1-\alpha}$	(21)	45, 68 and 70
Quenched trap/patch models	Yes	$\simeq t^\alpha$	$\simeq \Delta t^{\alpha-1}$		198 and 279[b]
Ageing CTRW	Yes	$\simeq \begin{cases} t/t_a^{1-\alpha}, & t \ll t_a, \\ t, & t \gg t_a \end{cases}$	$\simeq \Lambda_\alpha(t_a/t)\Delta t^{\alpha-1}$	(27) and (29)	73
Scaled Brownian motion	Yes	$\simeq t^\alpha$	$\simeq \Delta t^{\alpha-1}$	(8) and (80)	189 and 190
Ultraslow CTRW	Yes	$\simeq \log^\alpha(t)$	$\simeq \log^\alpha(t)\Delta/t$	(43) and (44)	110
Sinai (quenched)	Yes	$\simeq \log^4(t)$	$\simeq \log^4(t)\Delta/t$	(42)	110
CTRW in ageing environment	Yes	$\simeq \log(t)$	$\simeq \log(t)\Delta/t$	(40) and (41)	101
HDP $K(x) = (K_0/2)e^{-2x/x^*}$	Yes	$\simeq \log^2(t)$	$\simeq (\Delta/t)^{1/2}$	(94) and (95)	202



The Anomalous Diffusion Challenge

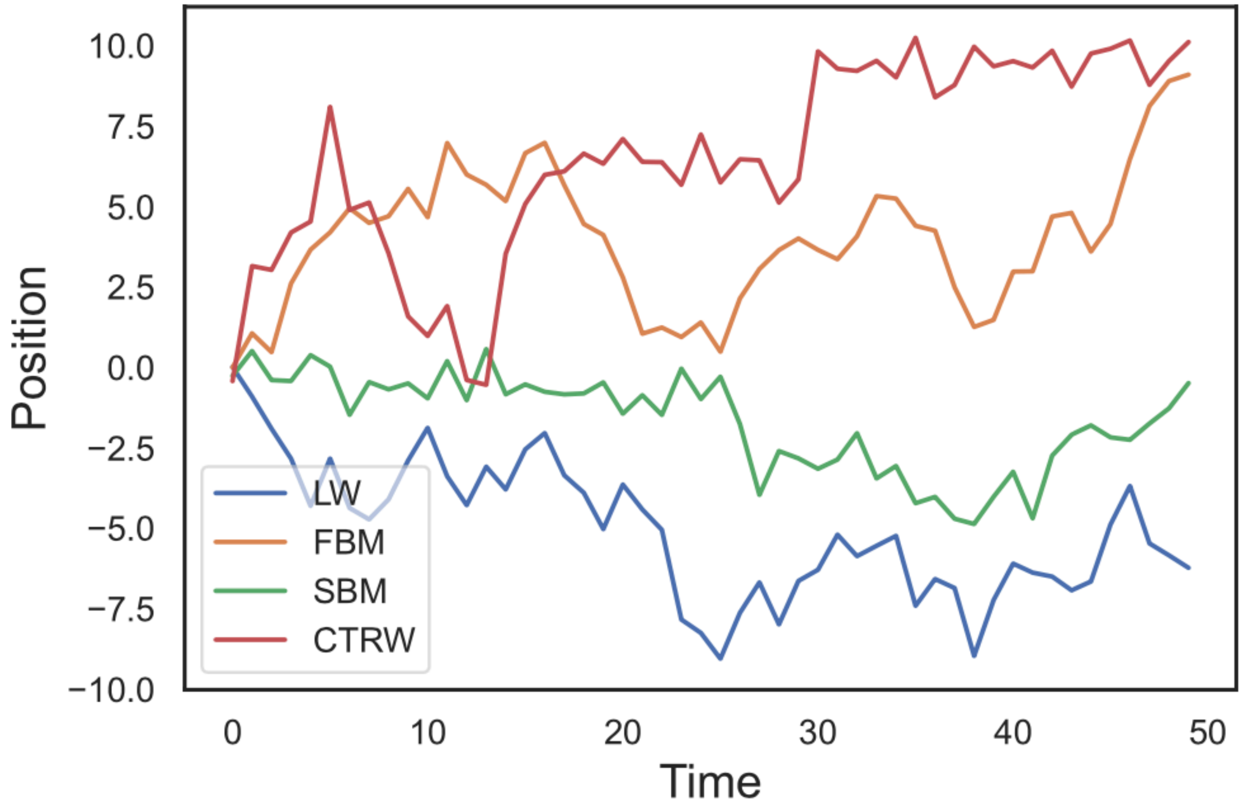
$$\mathrm{MSD}(\Delta) = 2K_\alpha \Delta^\alpha$$

Diffusion model $M_i \in$ {FBM, SBM, LW, CTRW, ATTM}

Task 1. Inference of the anomalous diffusion exponent α
Task 2. Classification of the underlying diffusion model
Task 3. Trajectory segmentation

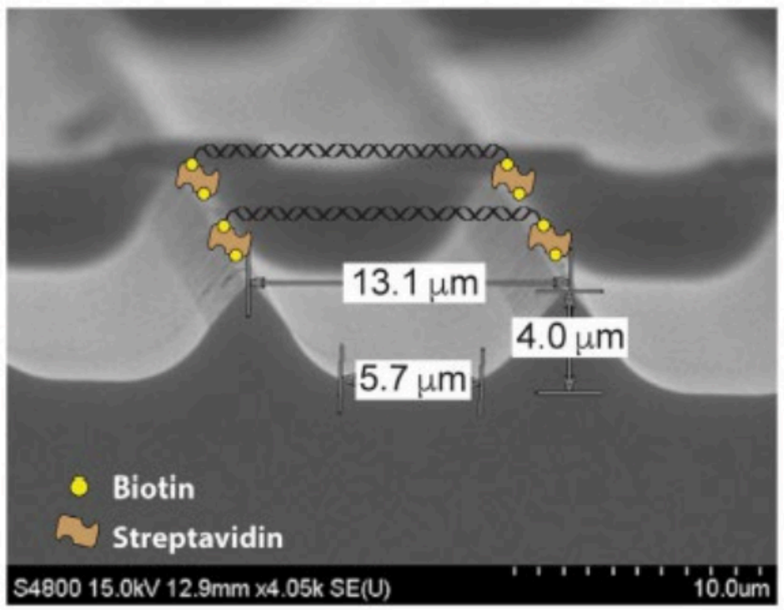
The only Bayesian team

Collaborated with Prof. M. A. Lomholt (Univ. Southern Denmark)
Dr. S. Thapa (Tel Aviv University)

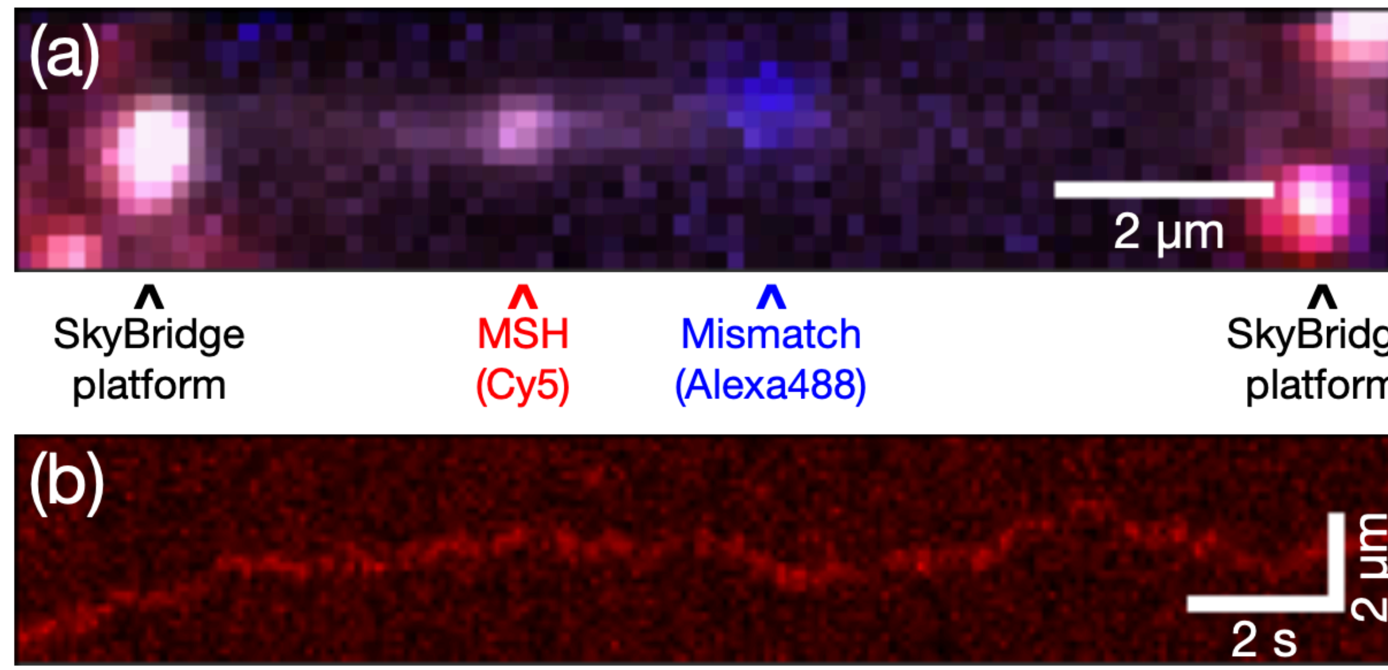


team/software	method
Anomalous Unicorns HYDRAS	Ensemble of CNN and RNN
BIT	Bayesian inference
DecBayComp Gratin	Graph neural network
DeepSPT	ResNet + XGBoost
eduN RANDI	RNN + Dense NN
Erasmus MC FEST	bi-LSTM + Dense NN
FCI	CNN
HNU Just LSTM it	LSTM
NOA	CNN + bi-LSTM

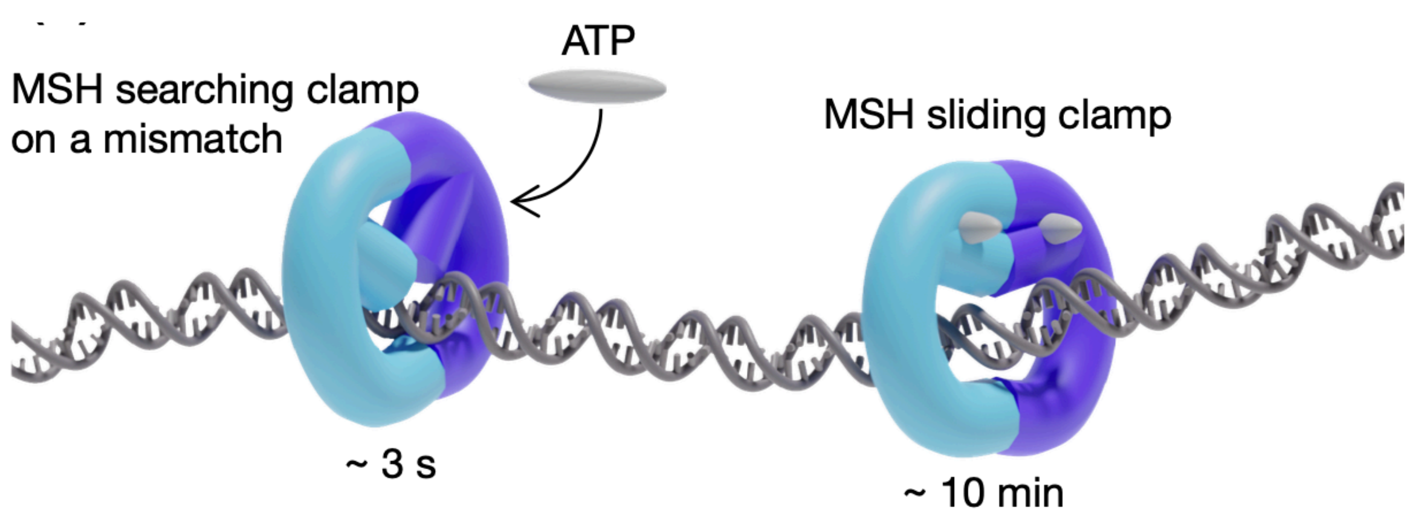
Single-molecule experiments for MutS proteins diffusing along a mismatch-containing DNA



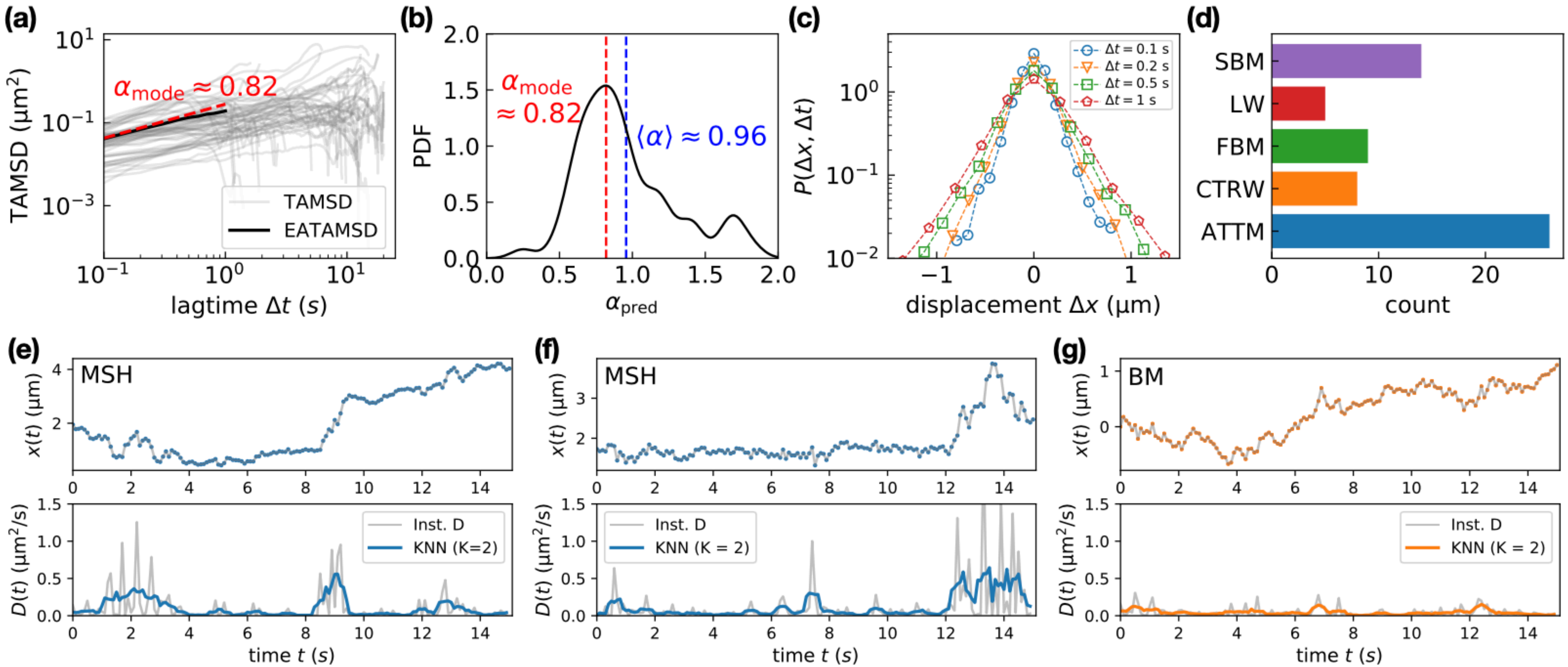
DNA SkyBridge
D. Kim *et al.*, *Nucleic Acids Res.* (2019)



Exp: J-B Lee's group (POSTECH)



Diffusion of MutS proteins is heterogeneous



Bayesian inference:

(1) From trajectories, identify the # of distinct diffusion states

(2) Estimate the diffusivity of each dynamic states

Multi-state Brownian motion model

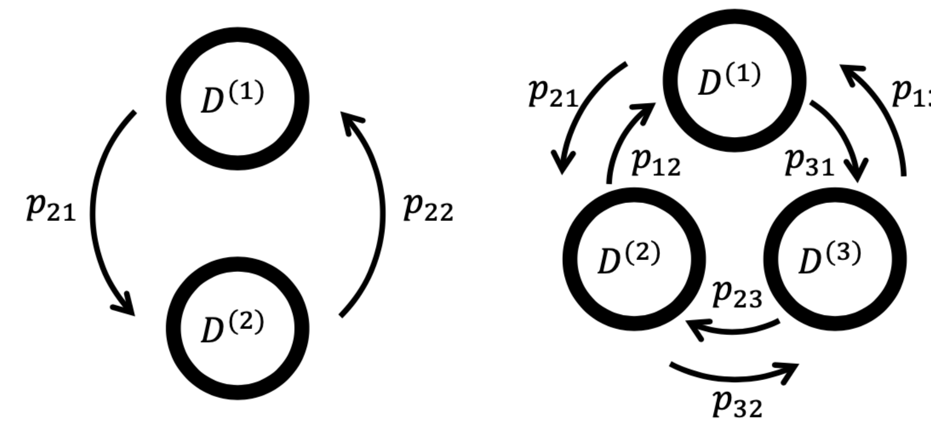
- Multi-state Brownian motion

  $\Delta x_t = \sqrt{2D_t t_0} \cdot \xi_t$: position

  $D_t \in \{D^{(1)}, D^{(2)}, \dots, D^{(N)}\}$: diffusion coefficients

  $\xi_t \sim \mathcal{N}(0,1)$: Gaussian noise

  p_{ij}: transition probability from $D^{(i)}$ to $D^{(j)}$



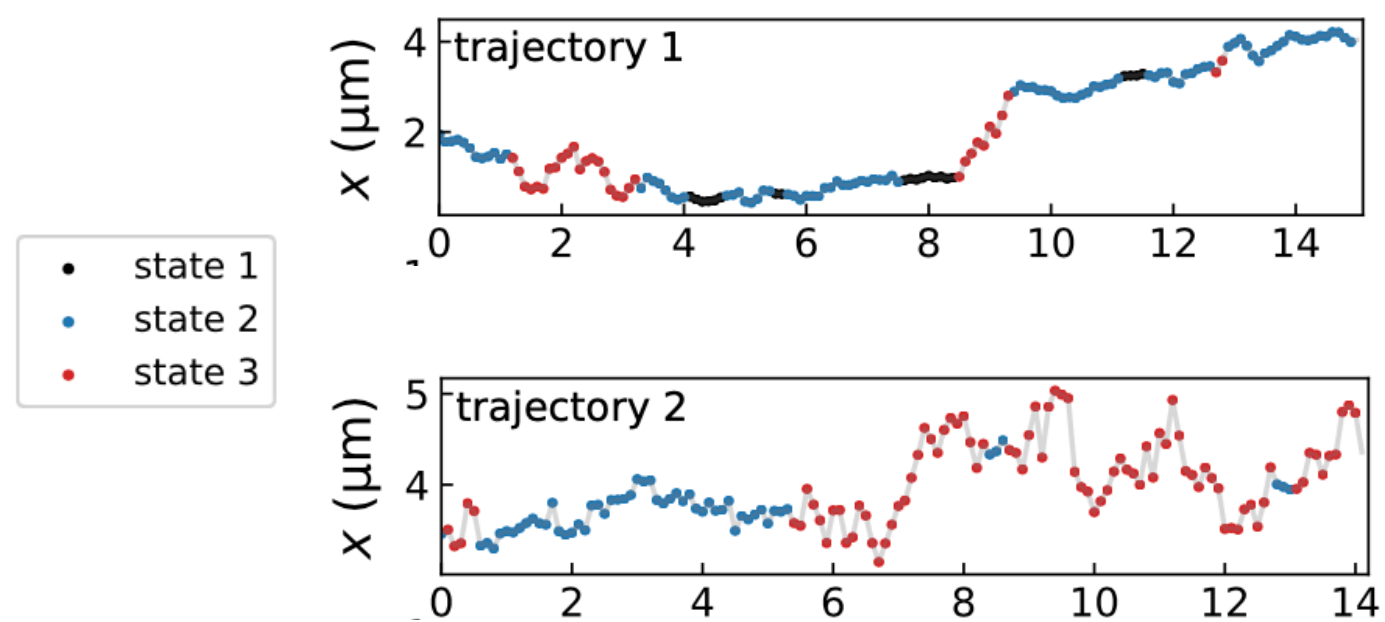
two-state BM **three-state BM**

- Models M_N

  M_1: one-state BM ($N = 1$)

  M_2: two-state BM ($N = 2$)

  M_3: three-state BM ($N = 3$)



Model inference

$$P(\mathrm{M}_N \text{ is true}) = \frac{P(\mathrm{M}_N|\mathrm{Data})}{\sum_{\mathrm{M}_I} P(\mathrm{M}_I|\mathrm{Data})} = \frac{P(\mathrm{Data}|\mathrm{M}_N)}{\sum_{\mathrm{M}_I} P(\mathrm{Data}|\mathrm{M}_I)}$$

Marginalization (evidence function) $\boldsymbol{\theta} = (D^{(1)}, \dots, p_{11}, \dots)$

$$P(\mathrm{Data}|\mathrm{M}_N) = \int_{\Theta} P(\mathrm{Data}|\mathrm{M}_N, \boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathrm{M}_N)\ \mathrm{d}\boldsymbol{\theta}$$

Likelihood

$$\mathcal{L}(\boldsymbol{\theta}|\{\Delta x_i\}, M_N)$$

The prior

$$P(\boldsymbol{\theta}|M_N) = \prod_{i=1}^{N} P(D^{(i)}|M_N) \prod_{i=1}^{N} P(\{p_{1i}, \dots, p_{Ni}\}|M_N)$$

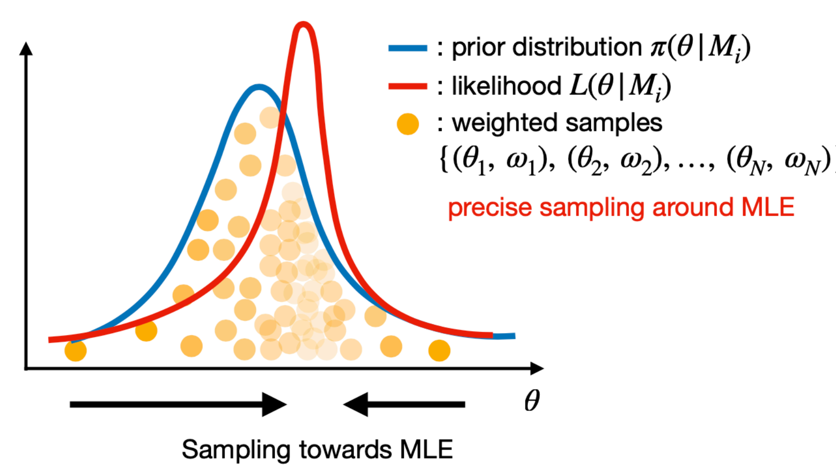
How to calculate: the nested sampling method or Akaike information criterion (AIC)

$$\mathrm{AIC} = 2K_N - 2\log \mathcal{L}(\boldsymbol{\theta}_{\mathrm{MLE}}|\{\Delta x_i\}, M_N)$$

Parameter inference

$$P(\boldsymbol{\theta}|\mathrm{M}_N, \mathrm{Data}) = \frac{P(\mathrm{Data}|\mathrm{M}_N, \boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathrm{M}_N)}{P(\mathrm{Data}|\mathrm{M}_N)}$$

Maximum a posterior (MAP): $\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}$ via the nested sampling method

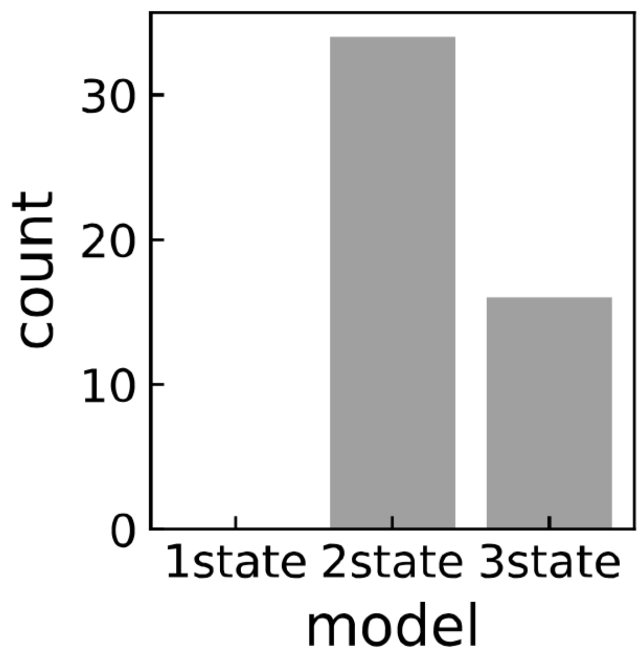


Model and parameter inference for simulated multi-state Brownian motion

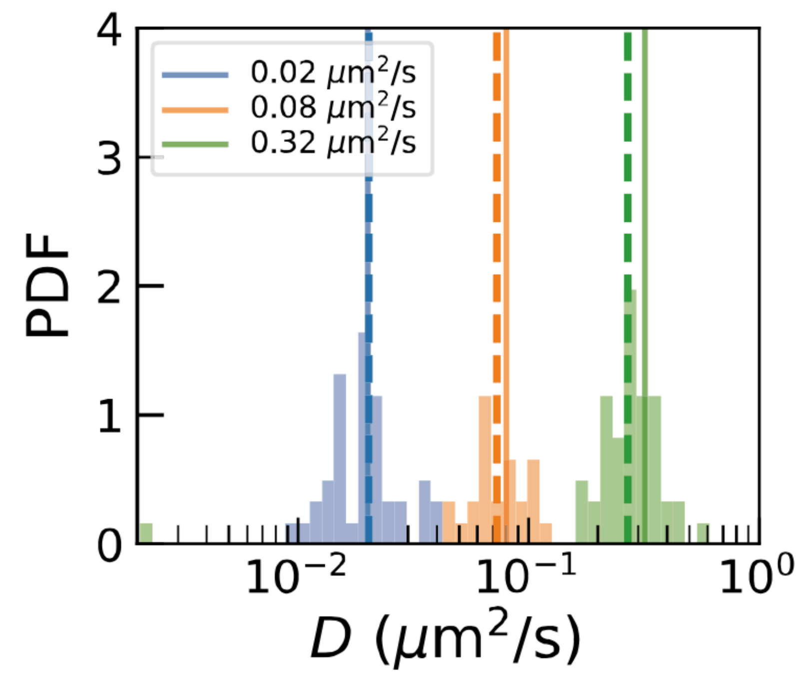
Simulation details: 3-state Brownian motion, 50 trajectories of the length T=200

$D^{(1)}=0.02,\ D^{(2)}=0.08,\ D^{(3)}=0.32$

Histogram of model inference



Histogram of parameter inference



F_1 score

$$F_1 = \frac{2TP}{2TP + FP + FN}$$

TP : True positive
FP : False positive
FN : False negative

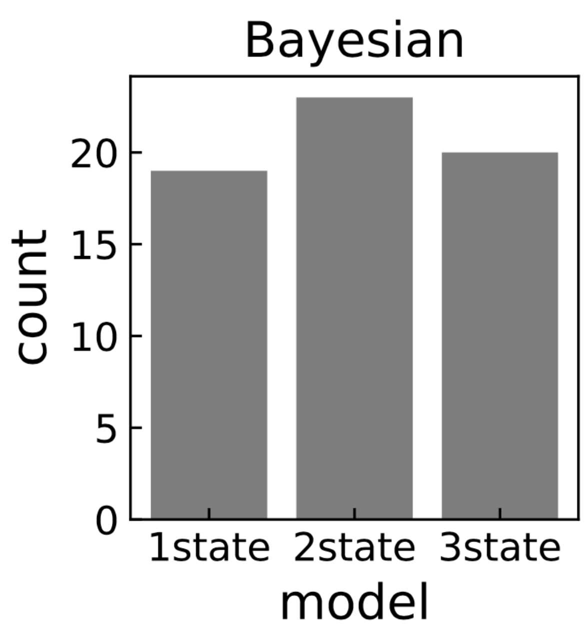
		set1	set2	set3	set4	set5	set6
D ($\mu m^2/s$)	$N=1$	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	$N=2$	(0.02, 0.04)	(0.02, 0.06)	(0.02, 0.08)	(0.02, 0.10)	(0.02, 0.14)	(0.02, 0.18)
	$N=3$	(0.02, 0.04, 0.06)	(0.02, 0.06, 0.10)	(0.02, 0.08, 0.14)	(0.02, 0.10, 0.18)	(0.02, 0.14, 0.26)	(0.02, 0.18, 0.34)
F_1 score	AIC	**0.3667**	**0.5833**	**0.65**	0.6667	**0.7333**	0.6833
	BIC	0.35	0.4333	0.5	0.5833	0.65	0.6667
	Bayesian	**0.3667**	0.5	**0.65**	**0.7167**	**0.7333**	**0.7167**

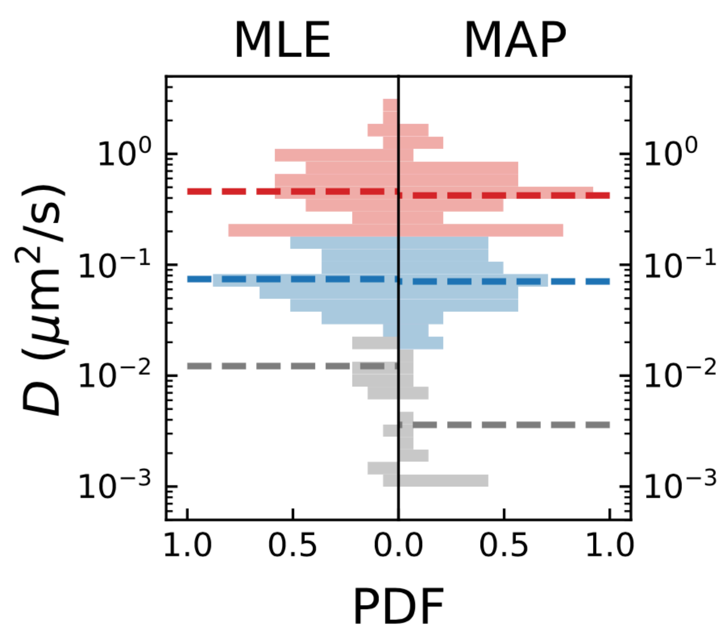
		set7	set8	set9	set10	set11	set12
D ($\mu m^2/s$)	$N=1$	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	$N=2$	(0.02, 0.04)	(0.02, 0.06)	(0.02, 0.08)	(0.02, 0.10)	(0.02, 0.14)	(0.02, 0.18)
	$N=3$	(0.02, 0.04, 0.08)	(0.02, 0.06, 0.18)	(0.02, 0.08, 0.32)	(0.02, 0.10, 0.50)	(0.02, 0.14, 0.98)	(0.02, 0.18, 1.62)
F_1 score	AIC	**0.3833**	**0.6**	0.7	0.8	0.85	0.9333
	BIC	0.35	0.4333	0.5	0.6	0.7	0.7333
	Bayesian	0.3667	0.5333	**0.7167**	**0.8833**	**0.9333**	**0.9667**

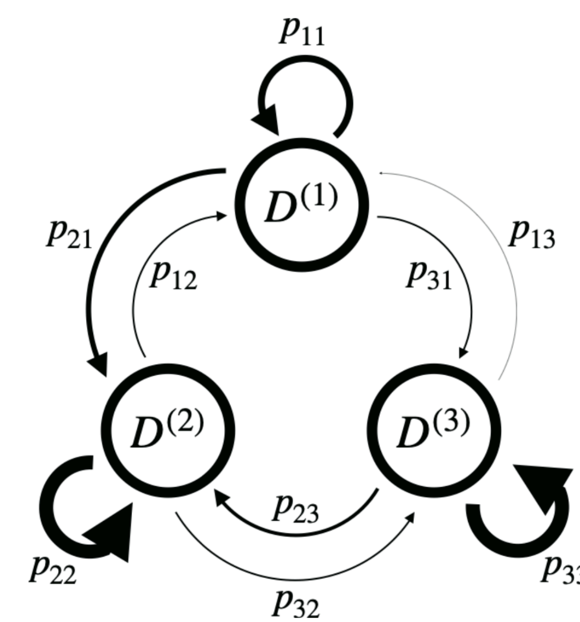
Bayesian analysis for diffusion dynamics of MutS proteins

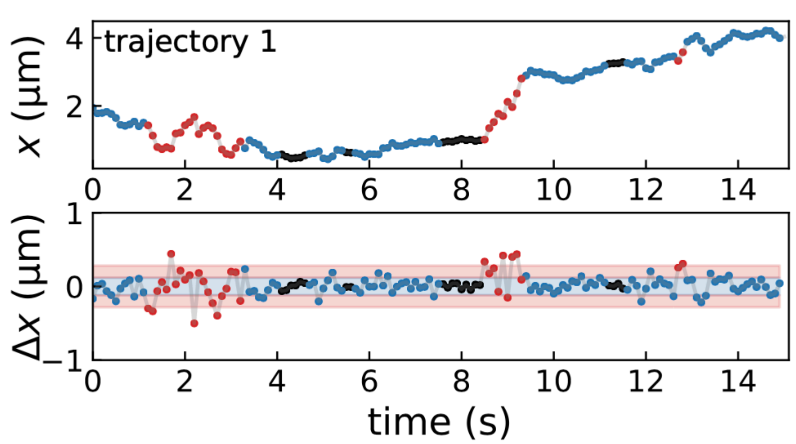
Experiment details: ATP-bound MutS sliding clamps, ~62 trajectories of the length T=150, time-resolution = 0.1 s, spatial resolution = 0.167 μm

ATP-bound MutS sliding clamps have three distinct diffusion states









State 1: Immobile (black)

$$D^{(1)} \approx 7.15 \times 10^{-3} \ \ \mu\text{m}^2/\text{s}.$$

State 2: slow diffusion (blue)

$$D^{(2)} \approx 6.89 \times 10^{-2} \ \ \mu\text{m}^2/\text{s}.$$

State 3: fast diffusion (red)

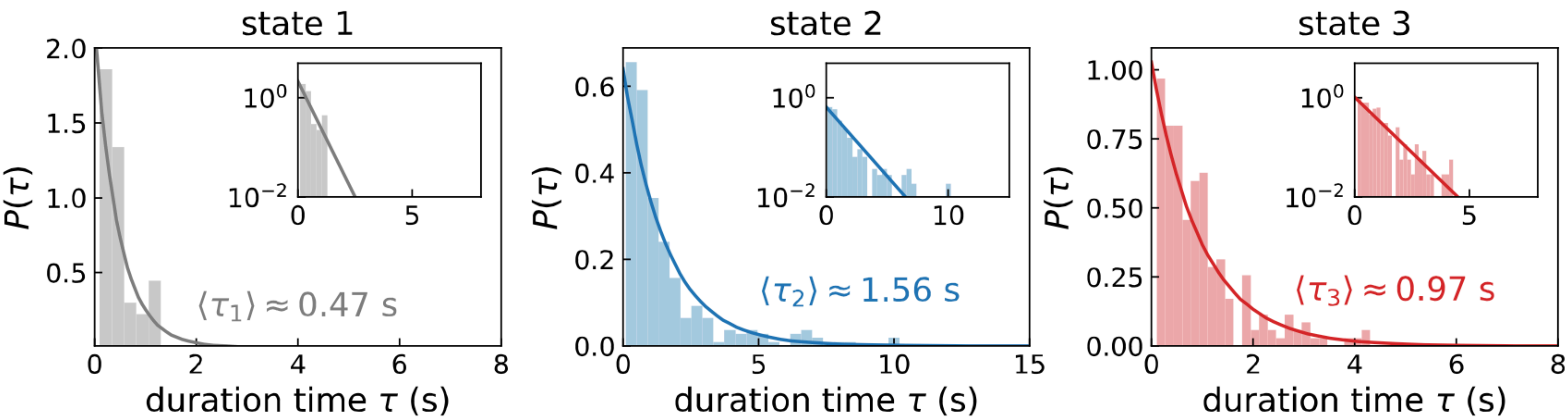
$$D^{(3)} \approx 4.00 \times 10^{-1} \ \mu\text{m}^2/\text{s}.$$

$$\hat{\mathbf{P}}_{\text{MLE}|\hat{\mathbf{D}}} = \begin{pmatrix} \hat{p}_{11} & \hat{p}_{12} & \hat{p}_{13} \\ \hat{p}_{21} & \hat{p}_{22} & \hat{p}_{23} \\ \hat{p}_{31} & \hat{p}_{32} & \hat{p}_{33} \end{pmatrix} \approx \begin{pmatrix} 0.6619 & 0.0460 & 0.0018 \\ 0.3009 & 0.8901 & 0.1177 \\ 0.0372 & 0.0639 & 0.8805 \end{pmatrix}$$

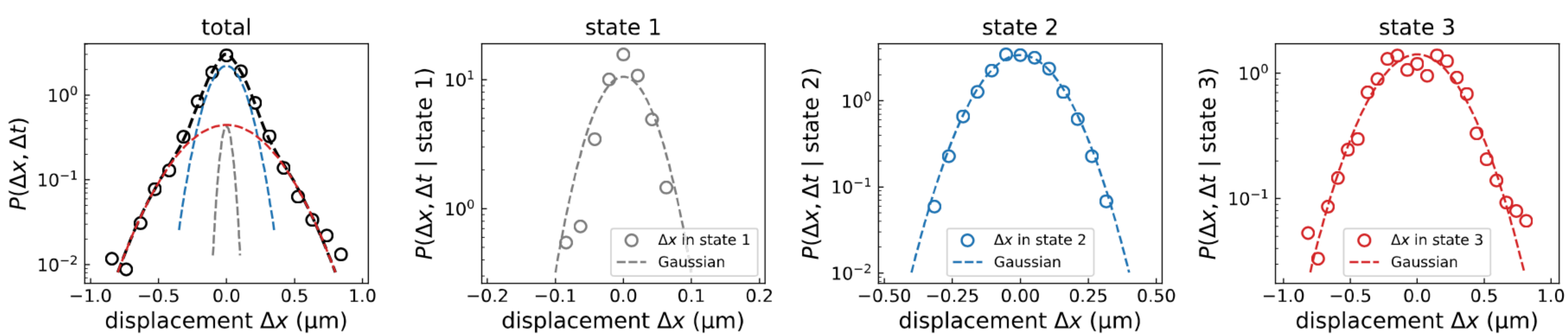
Bayesian analysis for diffusion dynamics of MutS proteins

Residence time distributions of state 1, 2, & 3: exponential law

$$P(\tau) = \frac{1}{\langle \tau_i \rangle} \exp\left(-\frac{1}{\langle \tau_i \rangle}\tau\right)$$



Probability density function for displacements: Gaussian

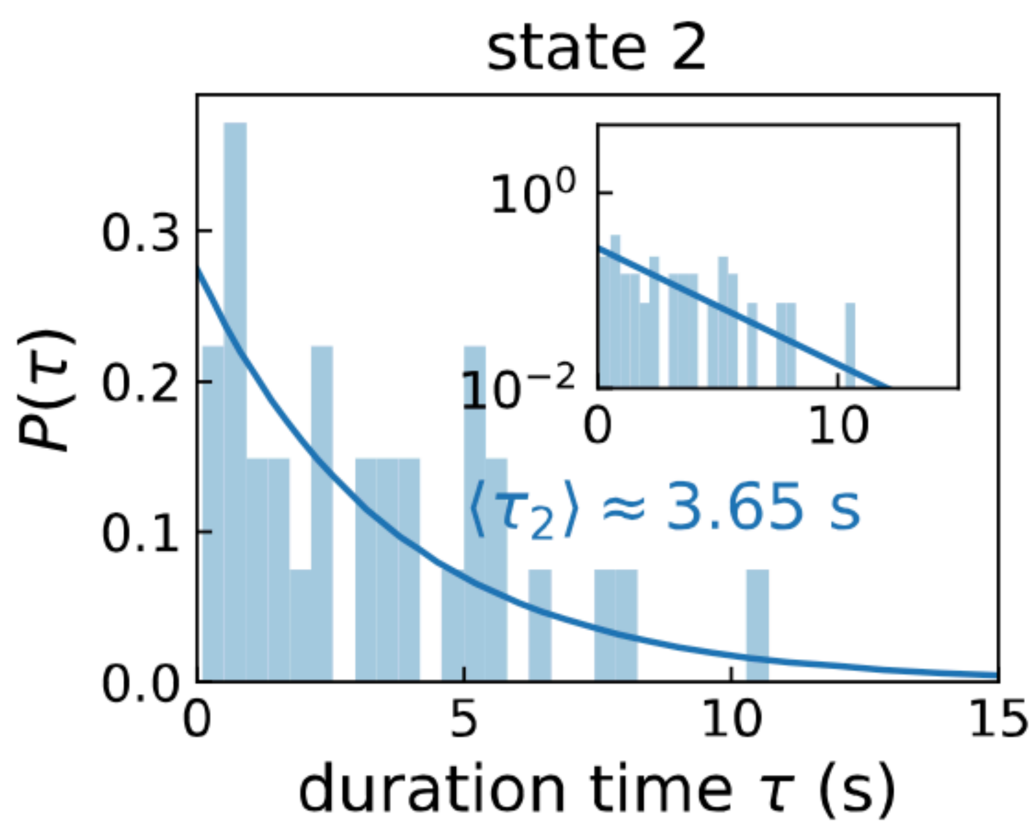


Bayesian analysis for diffusion dynamics of MutS proteins

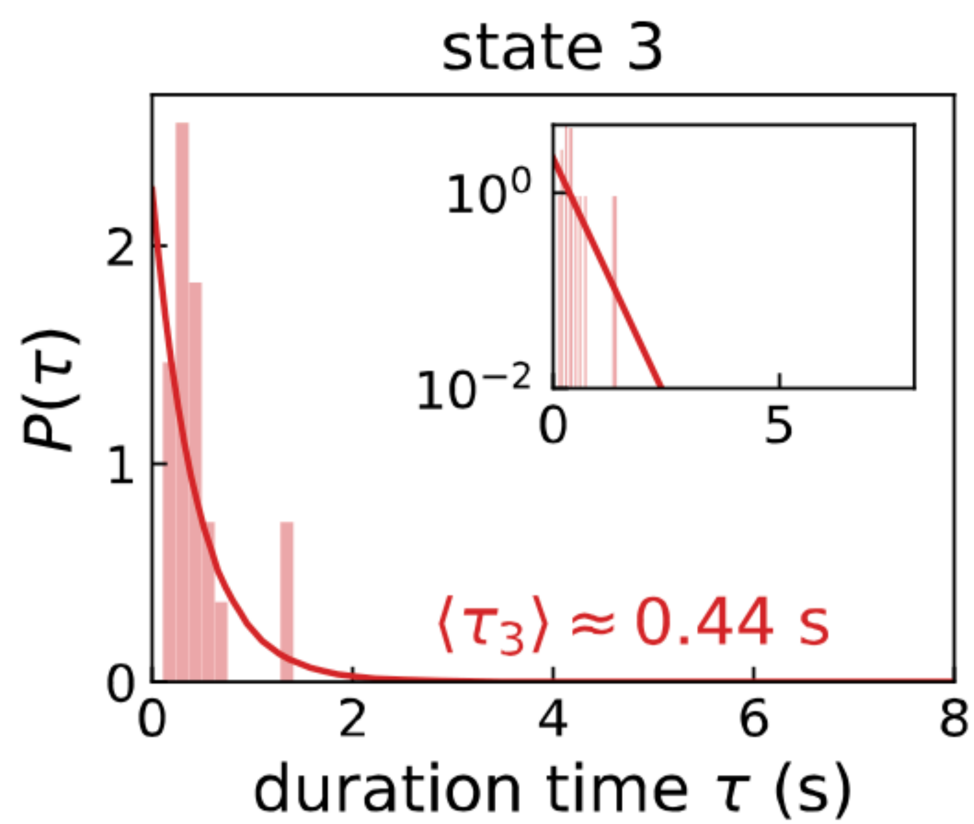
MutS sliding clamps in a ATP-depleted solution (pilot study)

$$\hat{\mathbf{D}} \approx (7.15 \times 10^{-3}, 6.88 \times 10^{-2}, 4.00 \times 10^{-1})\ \mu\text{m}^2/\text{s}$$

$$\hat{\mathbf{P}}_{\text{MLE}|\hat{\mathbf{D}}} \approx \begin{pmatrix} 0.7290 & 0.0083 & 0.0000 \\ 0.2323 & 0.9622 & 0.2451 \\ 0.0387 & 0.0295 & 0.7549 \end{pmatrix}$$



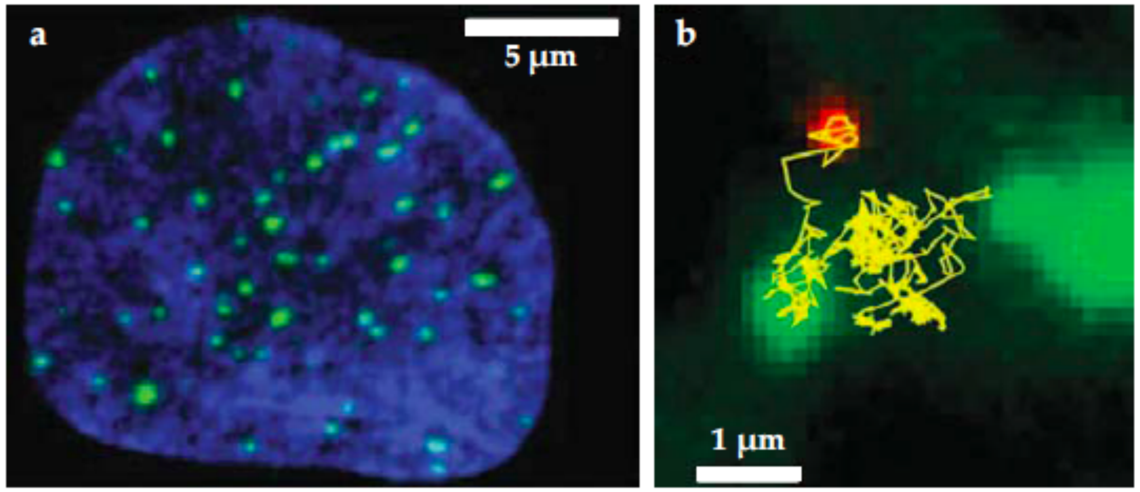
State 2 has an increased residence time



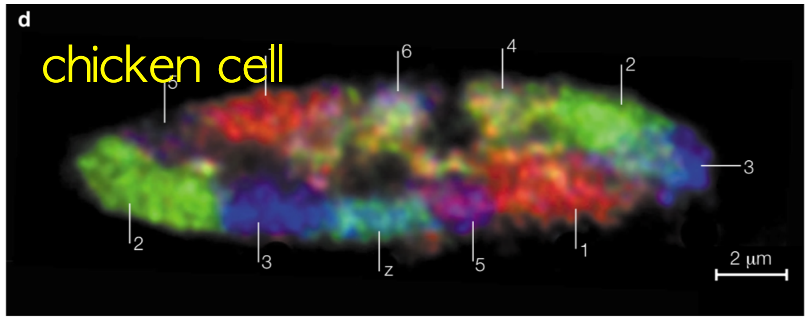
State 3 has an decreased residence time

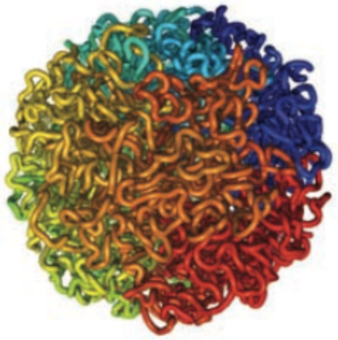
Take-home message: “*Life is physics*”

In vivo anomalous transport in cells

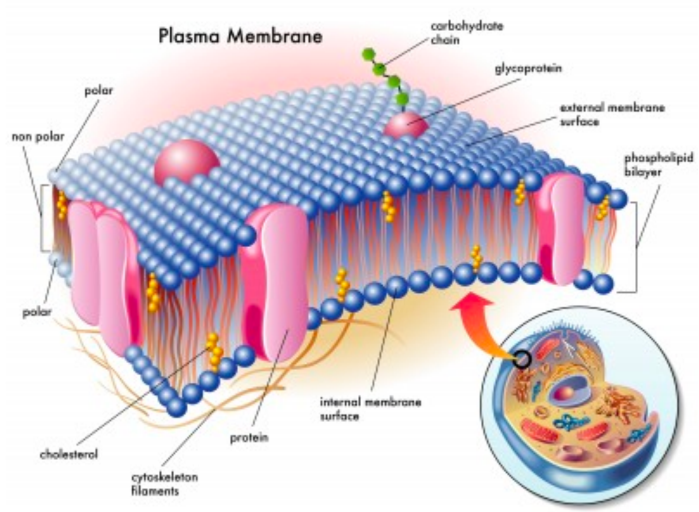


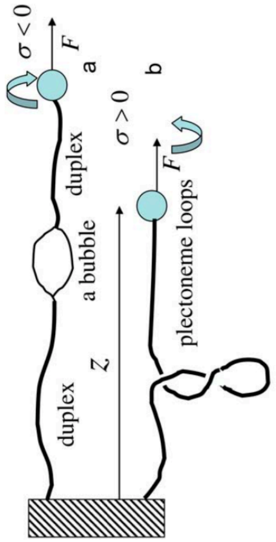
Biophysics of human chromosome



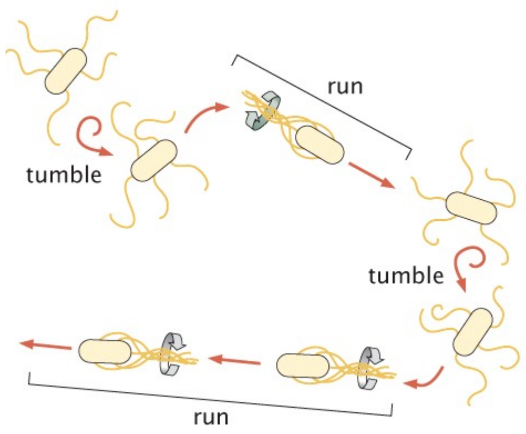


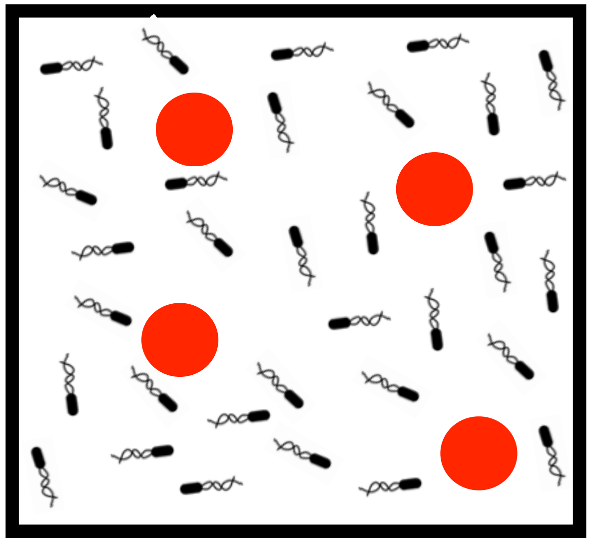
Single-molecule biophysics (DNA, membrane, etc)





Statistical physics of active particle/polymer



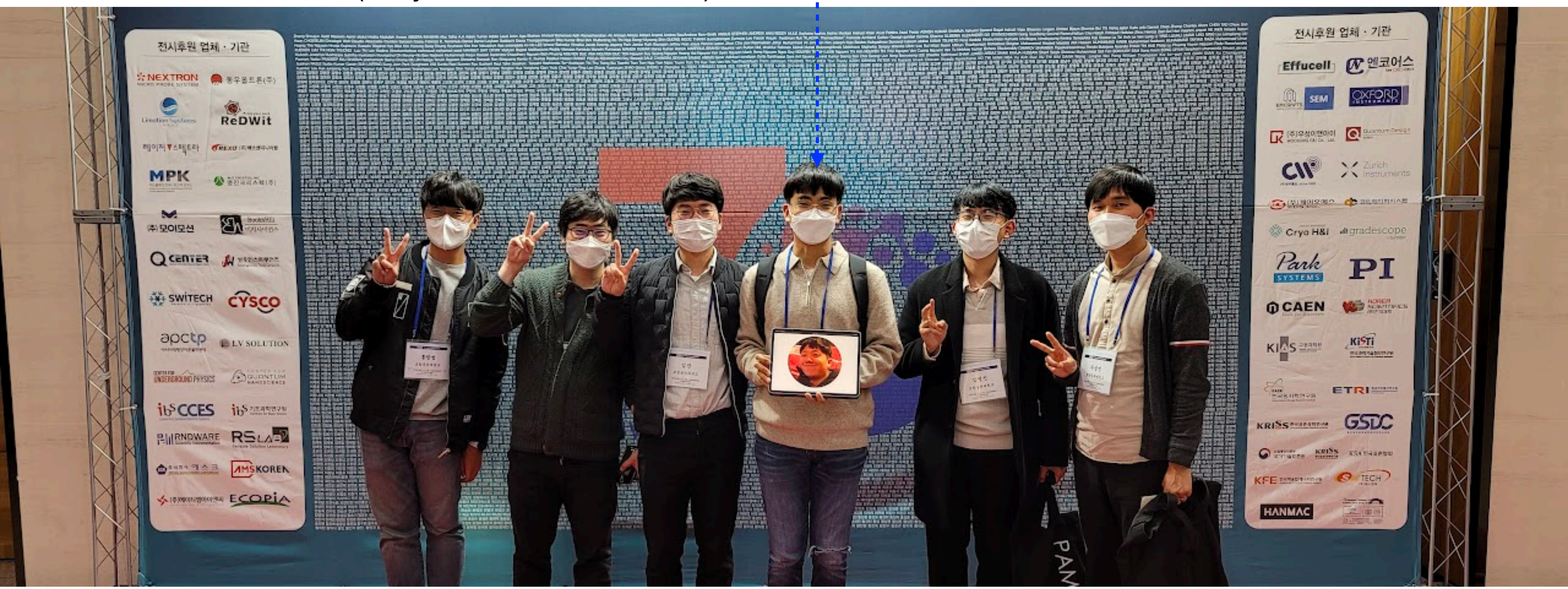


Thanks to

Bayesian inference study of anomalous transport (Levy walks, MutS, PCNA)

Seongyu Park

Target sequence search of DNA proteins in chromosomes



References: S. Park, …, J.-B. Lee, JHJ, *Discovering new diffusion states of MutS homologs from single-molecule trajectories*, manuscript in preparation

S. Thapha, S. Park, Y. Kim, JHJ, …, M. Lomholt, *Bayesian inference of scaled versus fractional Brownian motion*, JPA **55**, 194003 (2022)

S. Park, S. Thapa, …, JHJ, *Bayesian inference of Levy walks via hidden Markov models*, JPA **54**, 484001 (2021)

G. Munoz-Gil, …, S. Park, …, JHJ, …, C. Manzo, *Objective comparison of methods to decode anomalous diffusion*, Nature Commun. **12**, 6253 (2021)