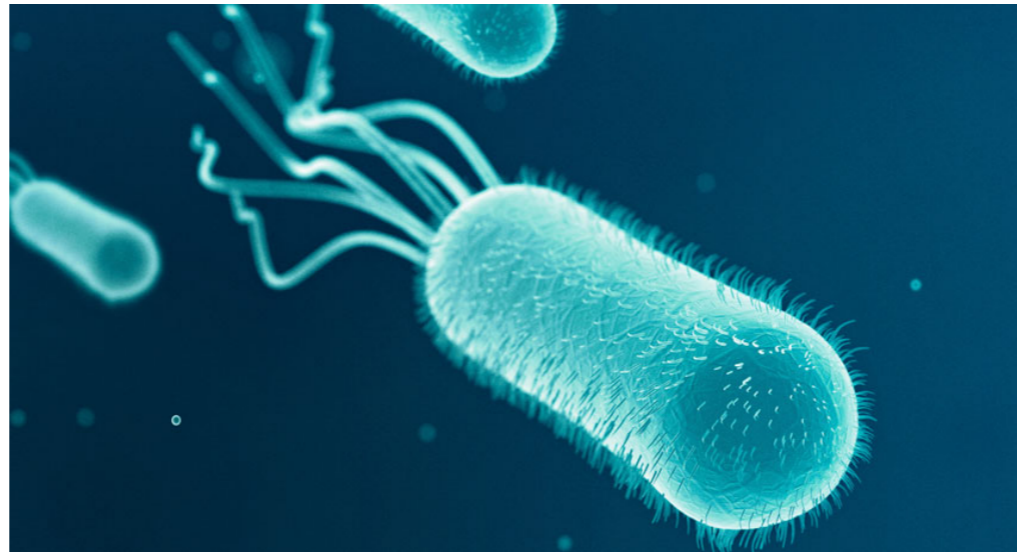
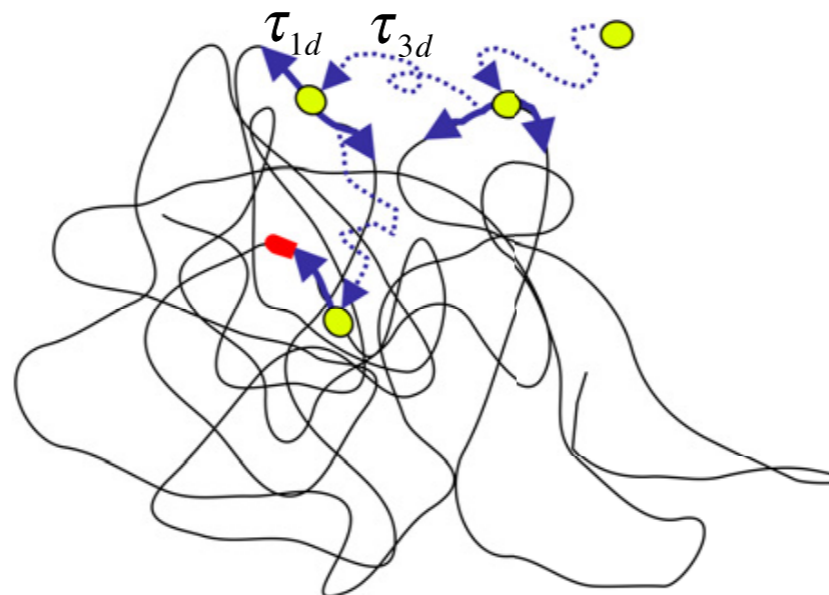


## MAE 545: Lecture 20 (4/19)

### **E. coli chemotaxis (continued)**

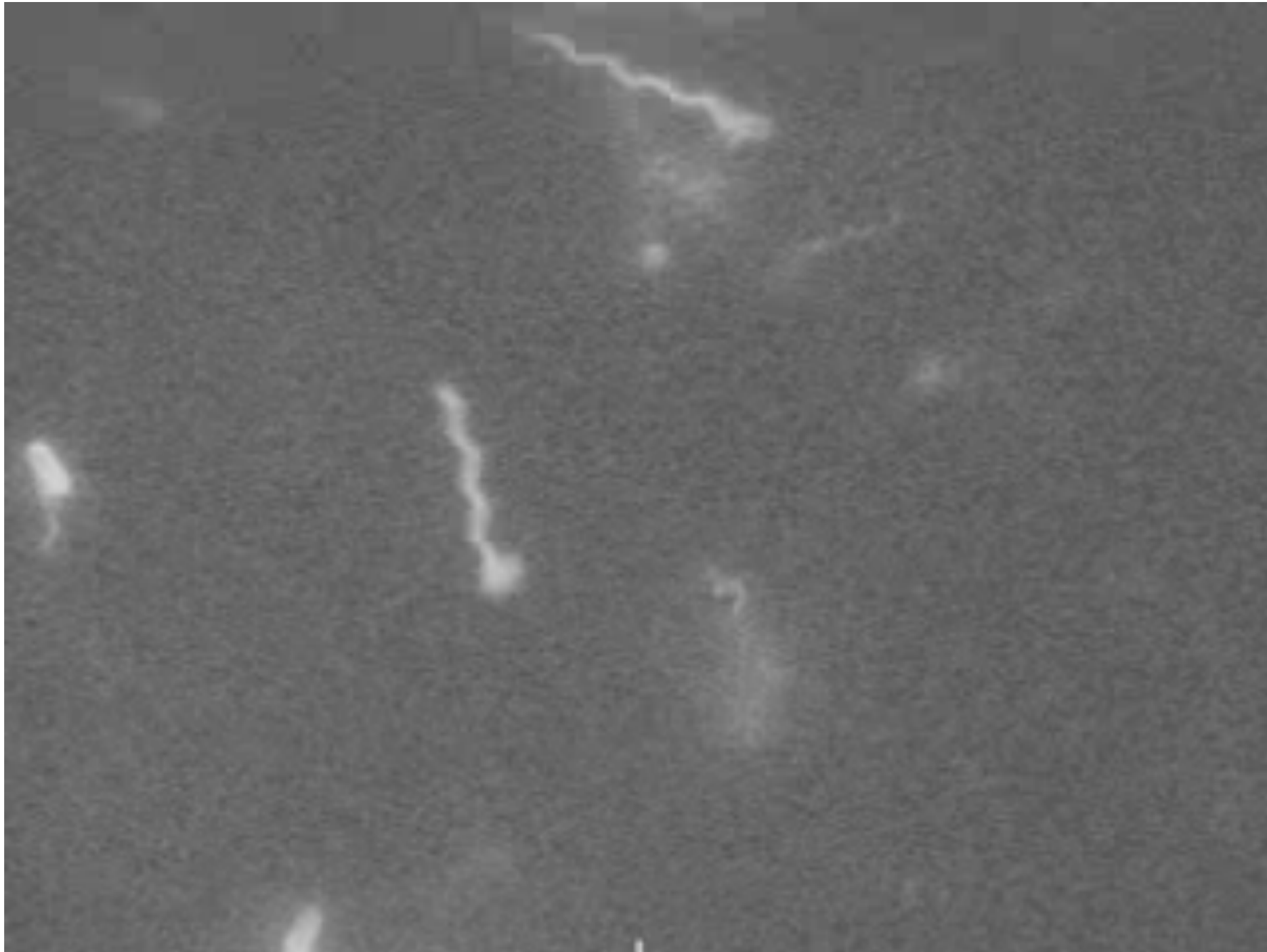


**How proteins find target sites on DNA?**



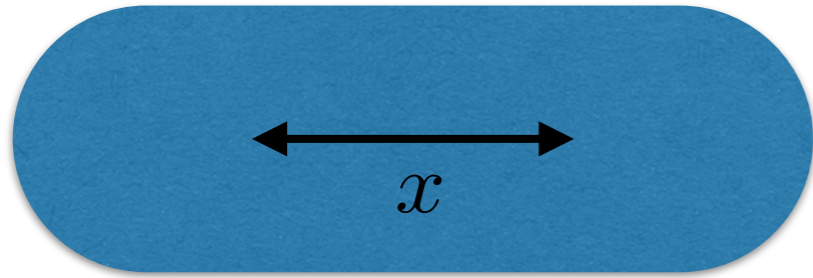
# E. coli chemotaxis

10 $\mu$ m



L. Turner, W.S. Ryu, H.C. Berg, J. Bacteriol. **182**, 2793-2801 (2000)

# Translational and rotational diffusion of E. coli



$$\langle x^2 \rangle = 2D_T t$$

**Einstein - Stokes  
relation**

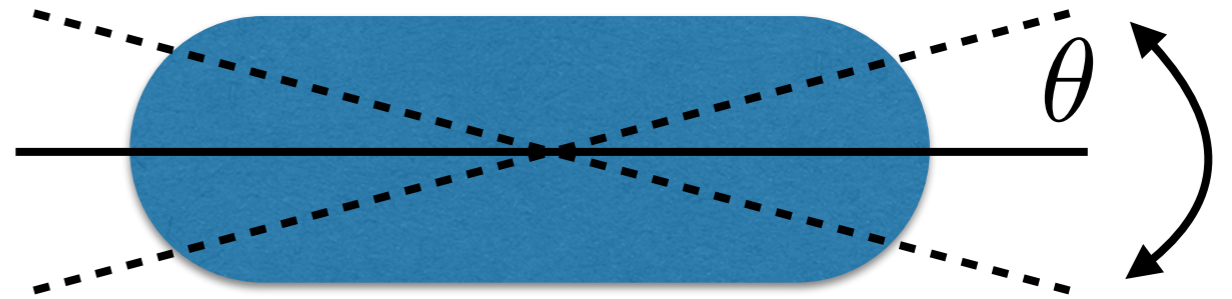
$$D_T \approx \frac{k_B T}{6\pi\eta R} \approx 0.2 \mu\text{m}^2/\text{s}$$

**size of E. coli**  $R \approx 1 \mu\text{m}$

**water viscosity**  $\eta \approx 10^{-3} \text{kg m}^{-1} \text{s}^{-1}$

**Boltzmann constant**  $k_B = 1.38 \times 10^{-23} \text{J/K}$

**temperature**  $T = 300\text{K}$



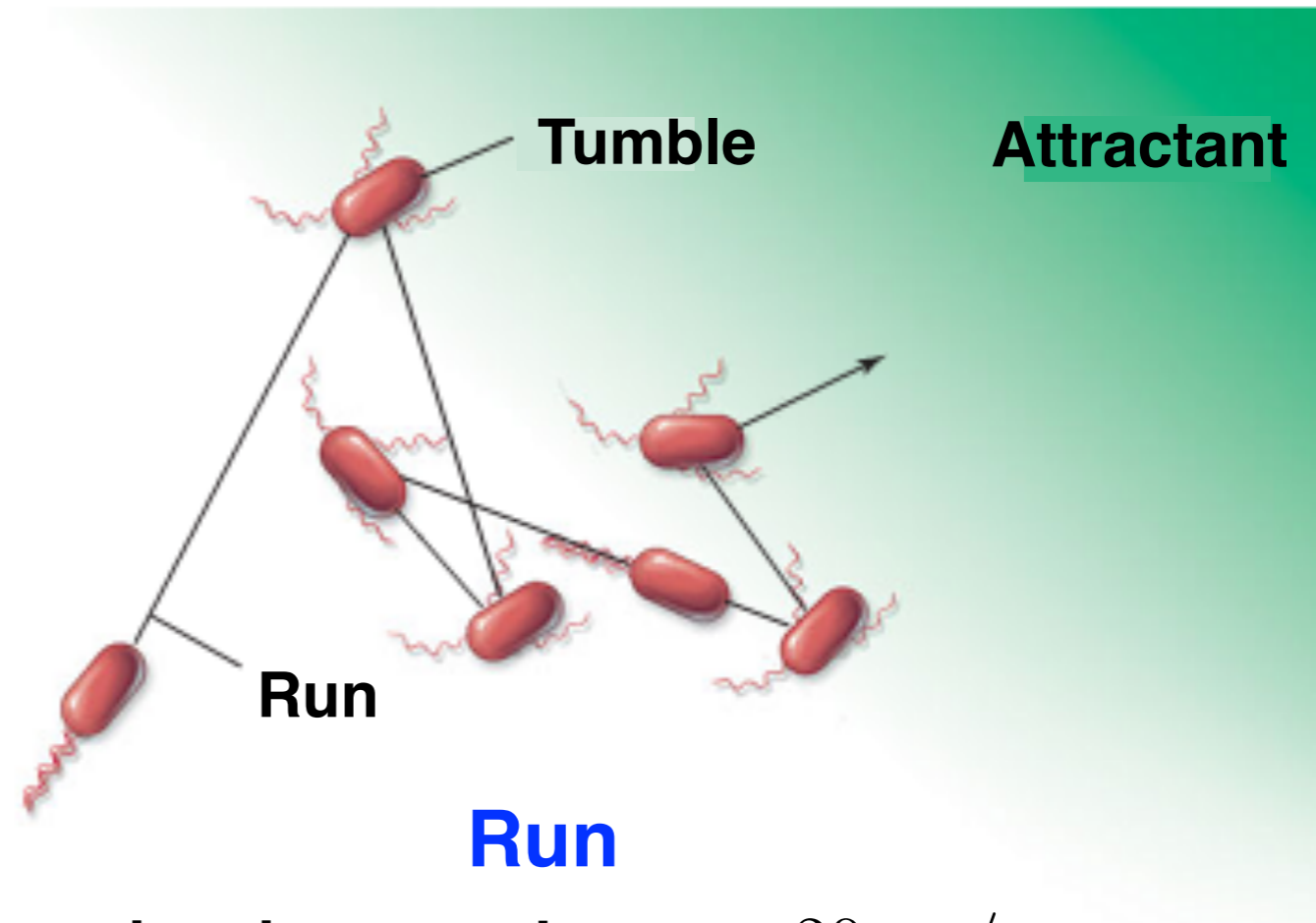
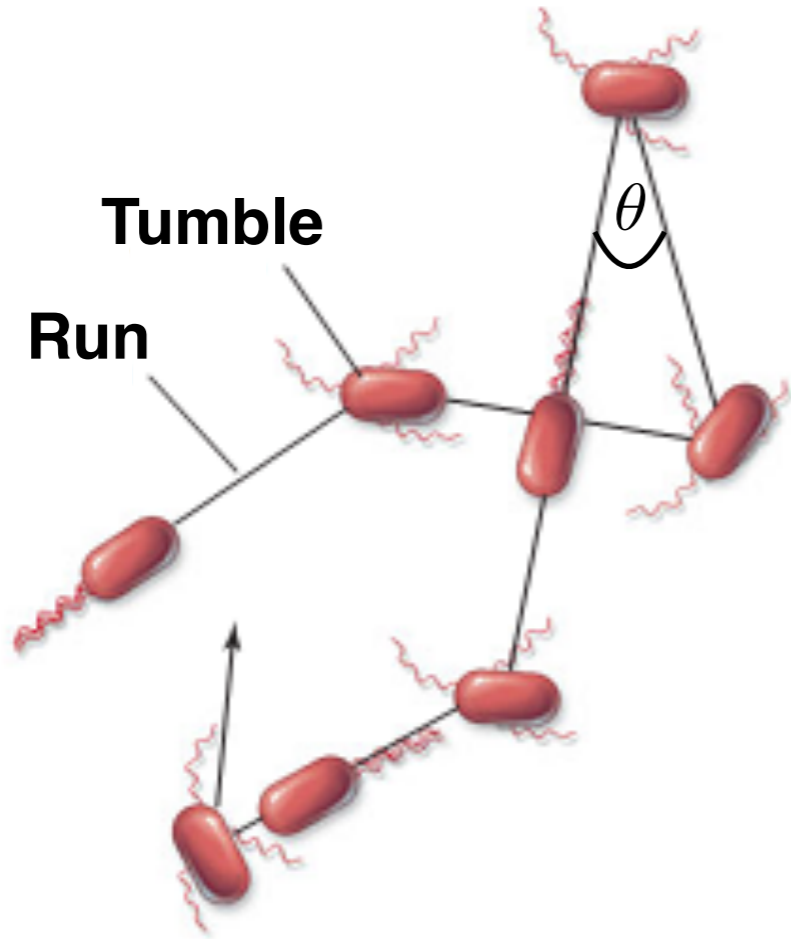
$$\langle \theta^2 \rangle = 2D_R t$$

**Einstein - Stokes  
relation**

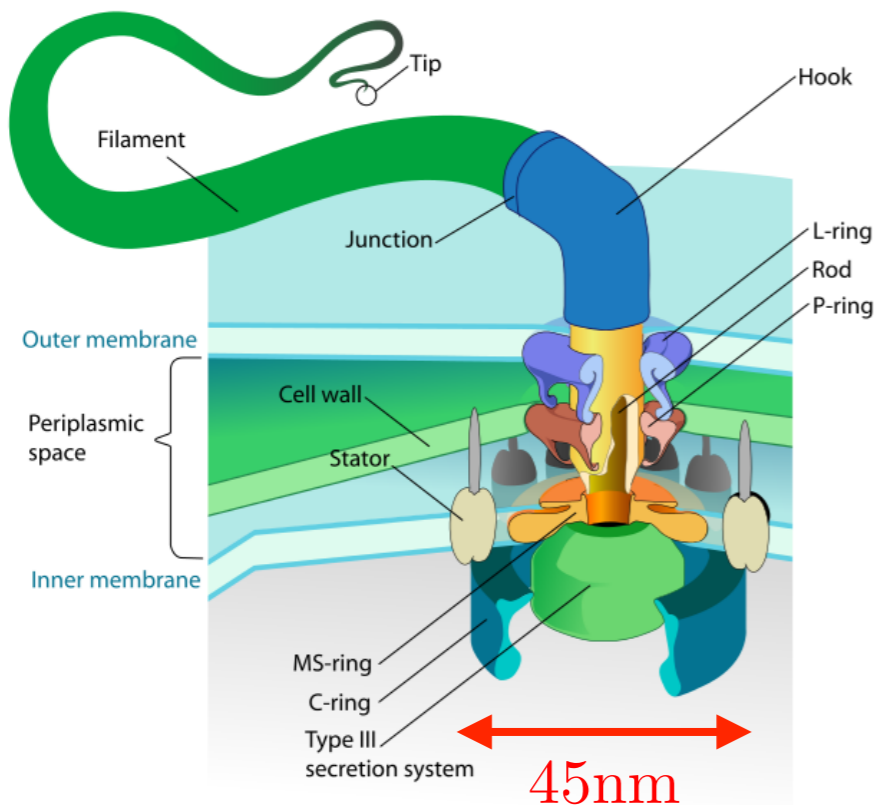
$$D_R \approx \frac{k_B T}{8\pi\eta R^3} \sim 0.2 \text{rad}^2/\text{s}$$

**After ~10s the orientation of  
E. coli changes by 90° due  
to the Brownian motion!**

# E. coli chemotaxis



**Rotary motor**



swimming speed:  $v_s \sim 20\mu\text{m/s}$

typical duration:  $t_r \sim 1\text{s}$

all motors turning counter clockwise

**Increase (Decrease) run durations, when swimming towards good (harmful) environment.**

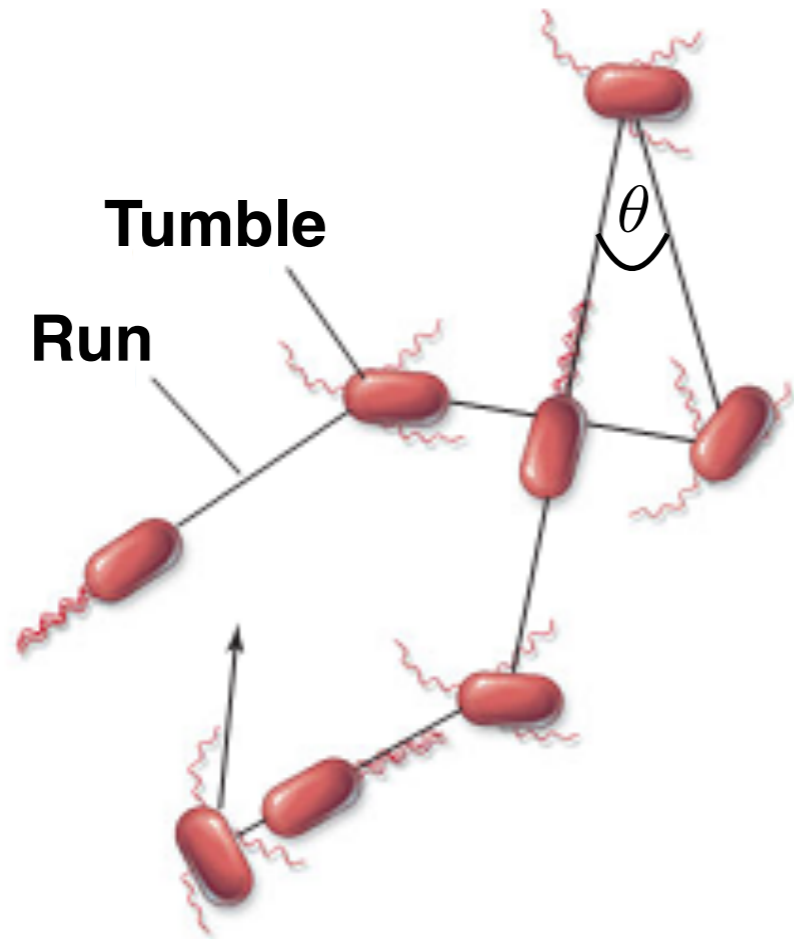
**Tumble**

random change in orientation  $\langle \theta \rangle = 68^\circ$

typical duration:  $t_t \sim 0.1\text{s}$

one or more motors turning clockwise

# E. coli chemotaxis



## Homogeneous environment

run duration:  $t_r \sim 1\text{s}$   
 tumble duration:  $t_t \sim 0.1\text{s}$   
 swimming speed:  $v_s \sim 20\mu\text{m/s}$

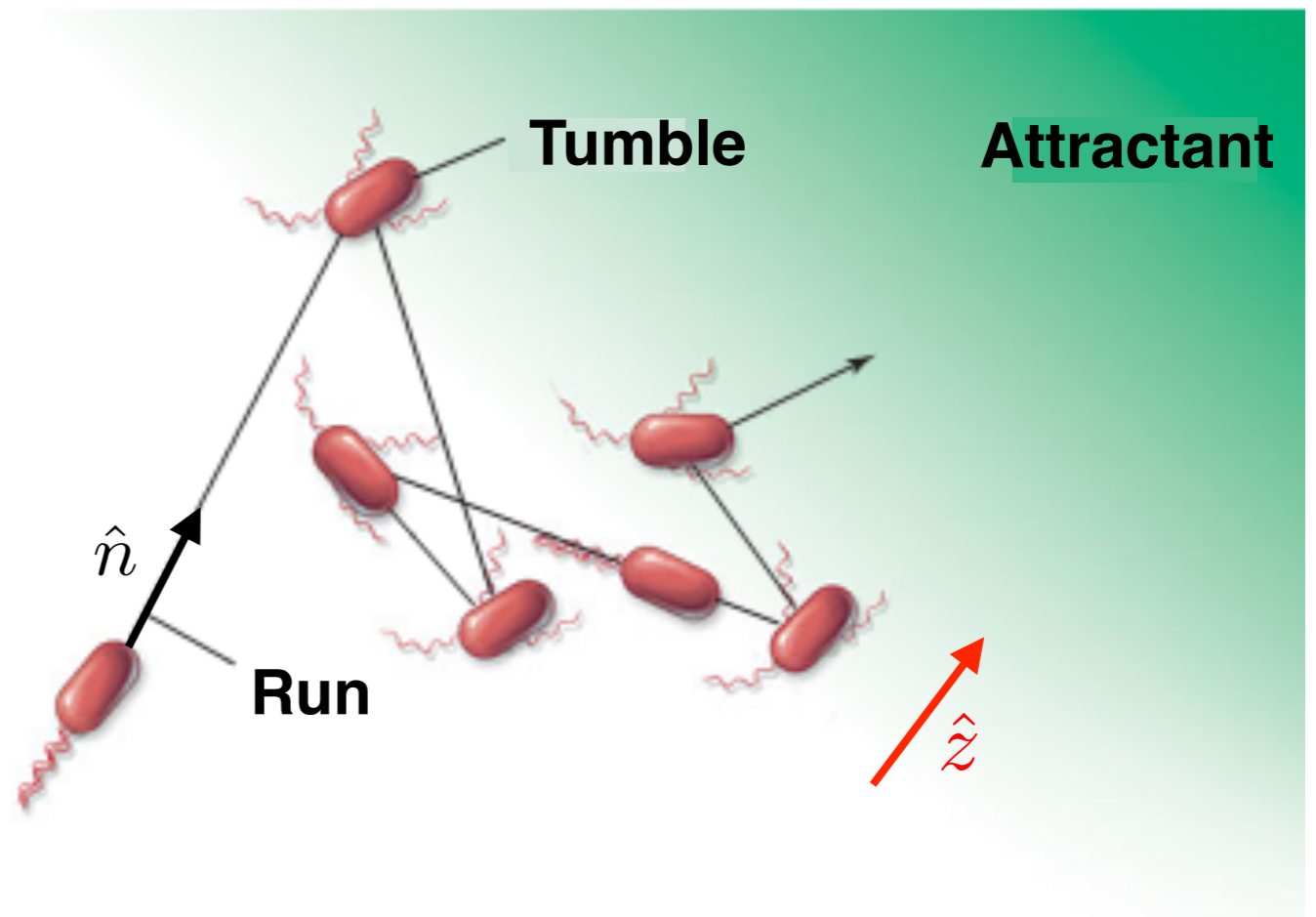
**drift velocity**

$$v_d = 0$$

**effective diffusion**

$$D_{\text{eff}} = \frac{\langle \Delta l^2 \rangle}{6 \langle \Delta t \rangle}$$

$$D_{\text{eff}} \approx \frac{v_s^2 t_r^2}{6(t_r + t_t)} \sim 60\mu\text{m}^2/\text{s}$$



## Gradient in "food" concentration

run duration increases (decreases) when swimming towards (away) from "food"

$$t_r(\hat{n}) = \bar{t}_r + \alpha(\hat{n} \cdot \hat{z})(\partial c / \partial z)$$

**drift velocity**

$$v_d = \frac{\langle \Delta z \rangle}{\langle \Delta t \rangle} \approx \frac{v_s \alpha (\partial c / \partial z)}{3(\bar{t}_r + t_t)}$$

$$\langle \Delta z \rangle = \langle v_z(\hat{n}) t_r(\hat{n}) \rangle = \langle v_s (\hat{n} \cdot \hat{z}) t_r(\hat{n}) \rangle$$

# Sensing of environment

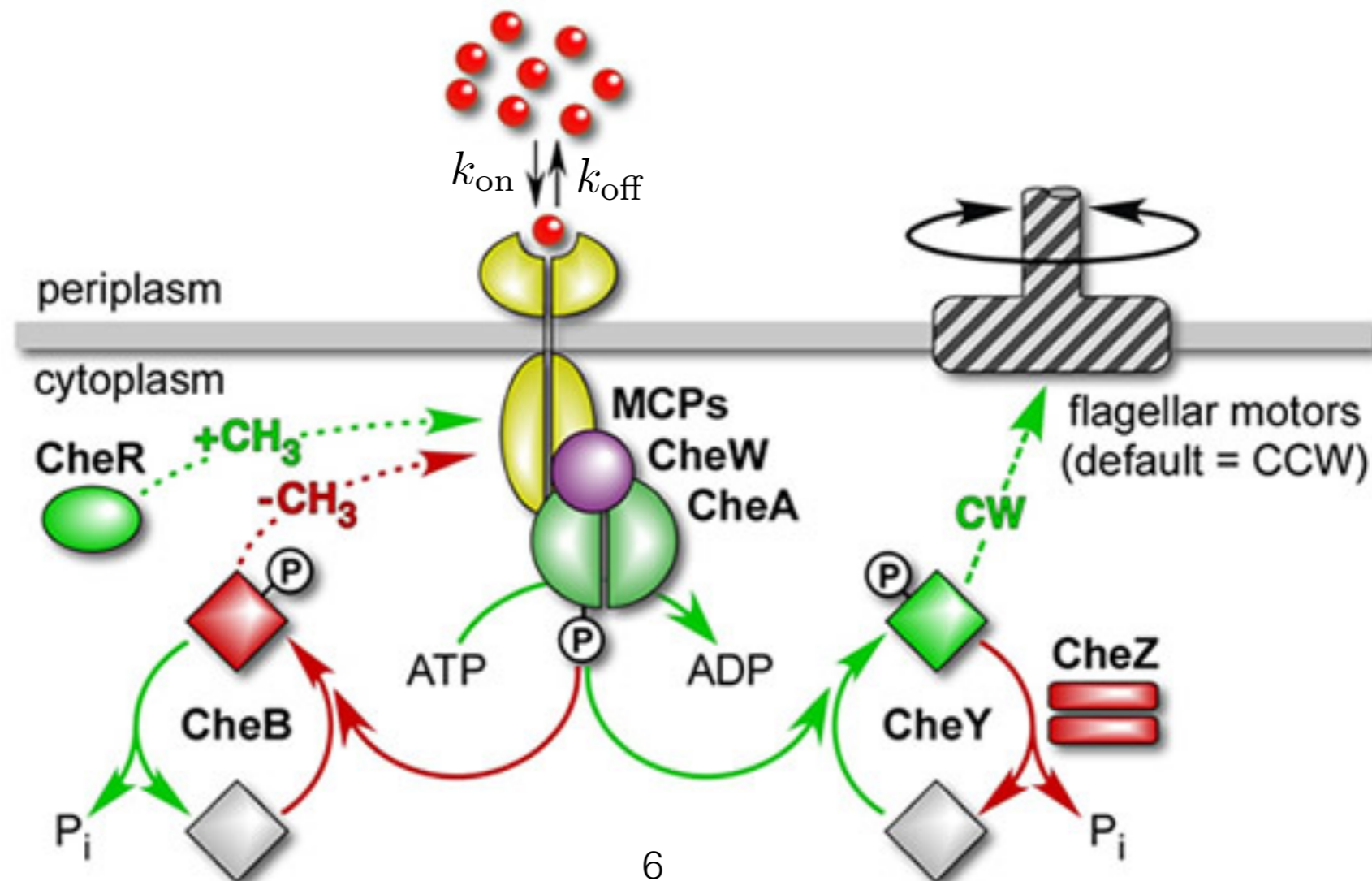
**E. coli surface is covered with receptors, which can bind specific molecules.**

**Average fraction of bound receptors  $p_B$  is related to concentration  $c$  of molecules.**

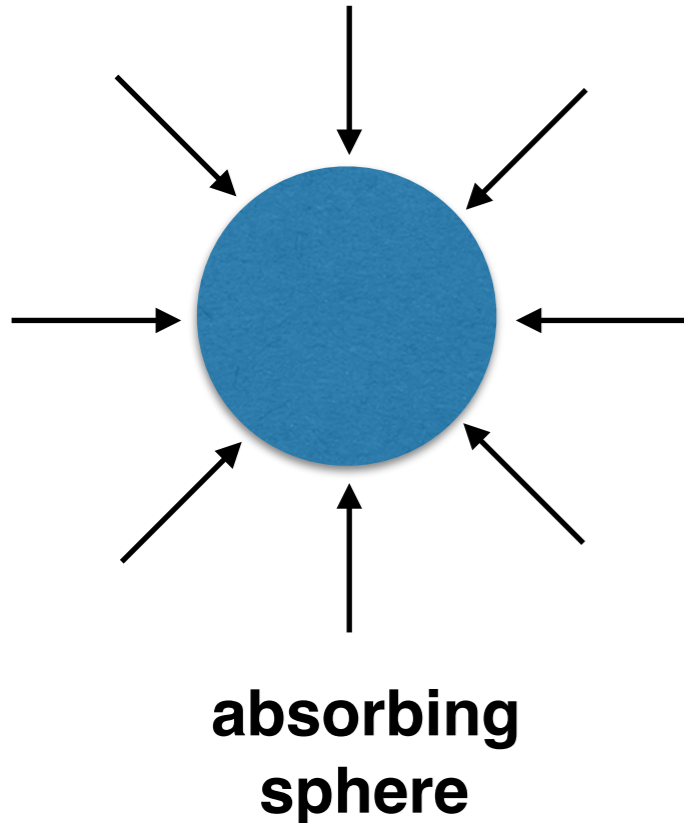
$$p_B = \frac{c}{c + c_0}$$

$$c_0 = \frac{k_{\text{off}}}{k_{\text{on}}}$$

**Chemical signaling network inside E. coli analyzes state of receptors and gives direction to rotary motor.**



# Diffusion limited flux of molecules to E. coli



## Fick's law

$$\frac{\partial c}{\partial t} = D \nabla^2 c = D \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial c}{\partial r} \right)$$

## boundary conditions

$$c(r \rightarrow \infty) = c_\infty$$

$$c(R) = 0$$

## steady state

$$c(r) = c_\infty \left[ 1 - \frac{R}{r} \right]$$

## flux density of molecules

$$J(r) = -D \frac{\partial c(r)}{\partial r} = -\frac{D c_\infty R}{r^2}$$

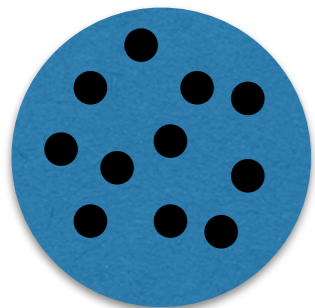
## rate of absorbing molecules

$$I(r) = J(r) \times 4\pi r^2 = -4\pi D R c_\infty = I_0 = -k_{\text{on}} c_\infty$$

diffusion constant for small molecules

$$D \approx 10^3 \mu\text{m}^2/\text{s}$$

$$k_{\text{on}} = 4\pi D R \sim 10^4 \mu\text{m}^3/\text{s}$$



$N$  absorbing disks of radius  $s$

$$I = \frac{I_0}{1 + \pi R / N s}$$

**example**  $R \sim 1 \mu\text{m}$   $s \sim 1 \text{nm}$

**flux drops by factor 2 for**

$$N = \pi R / s \sim 3000$$

**fractional area covered by these receptors**

$$(N \pi s^2) / (4\pi R^2) \sim 10^{-3}$$



**E. coli can use many types of receptors specific for different molecules, without significantly affecting the diffusive flux**

# Accuracy of concentration measurement

How many molecules do we expect inside a volume occupied by E. coli?

$$\bar{N} \sim R^3 c$$

Probability  $p(N)$  that cell measures  $N$  molecules follows Poisson distribution

$$p(N) = \frac{\bar{N}^N E^{-\bar{N}}}{N!} \quad \text{mean } \bar{N} \quad \text{standard deviation } \sigma_N = \sqrt{\bar{N}}$$

## Error in measurement

$$\text{Err} \sim \frac{\sigma_N}{\bar{N}} \sim (R^3 c)^{-1/2} \quad \text{for } c = 1\mu\text{M} = 6 \times 10^{20} \text{m}^{-3} \Rightarrow \text{Err} \sim 4\%$$

**E.coli can be more precise by counting molecules for longer time  $t$ .  
However, they need to wait some time  $t_0$  in order for the original molecules to diffuse away to prevent double counting of the same molecules!**

$$t_0 \sim R^2/D \sim 10^{-3} \text{s} \quad \bar{N} \sim R^3 ct/t_0 \sim DRct$$
$$\text{Err} \sim (DRct)^{-1/2} \quad \text{for } t=1\text{s, precision improves to Err} \sim 0.1\%$$

**When E. coli is swimming, it wants to swim faster than the diffusion of small molecules**

$$v_s t \gtrsim (Dt)^{1/2} \Rightarrow t \gtrsim D/v_s^2 \sim 1\text{s}$$

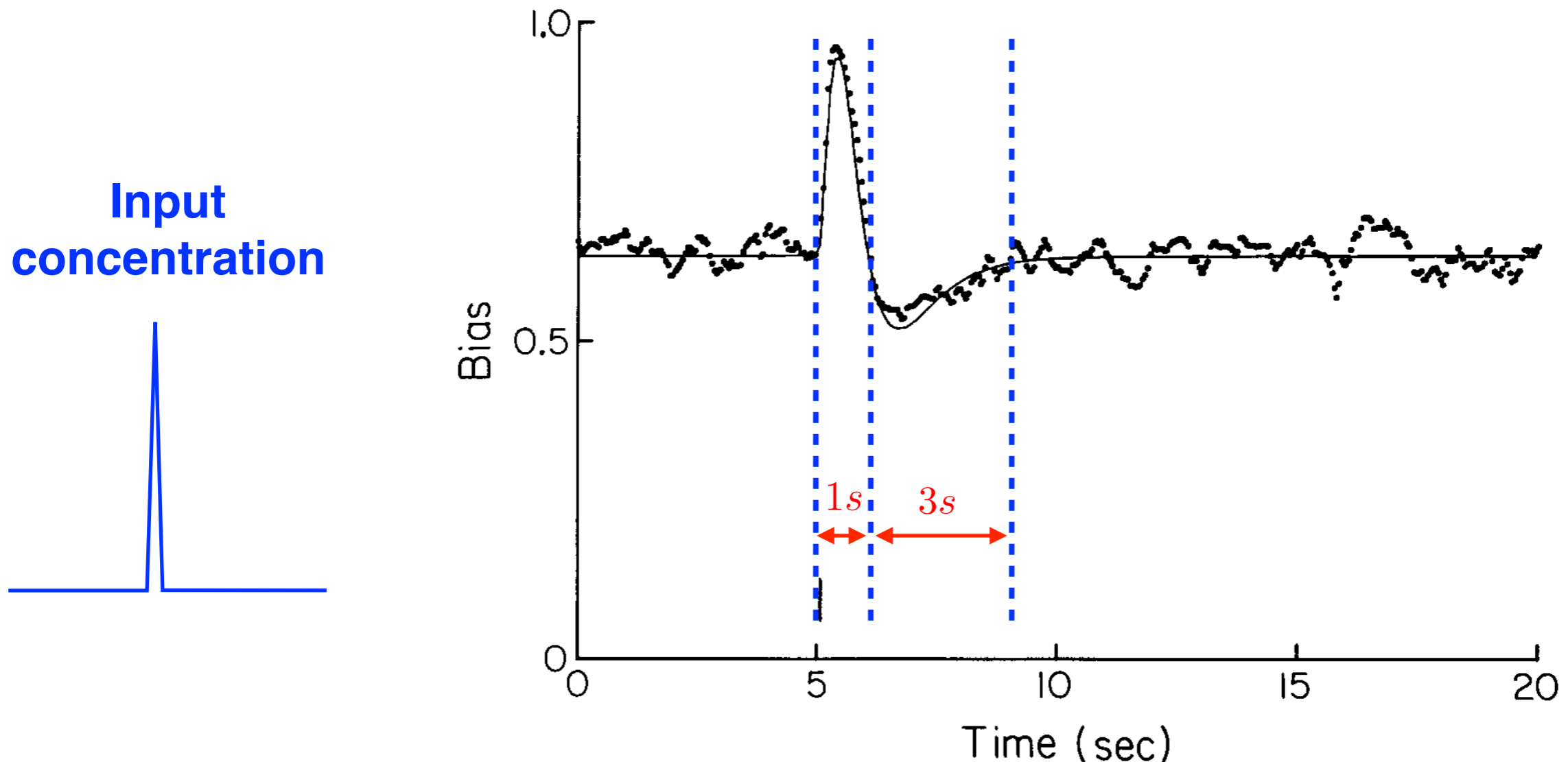
**Molar concentration**

$$1M = 6 \times 10^{26} \text{m}^{-3}$$



# How *E. coli* actually measures concentration?

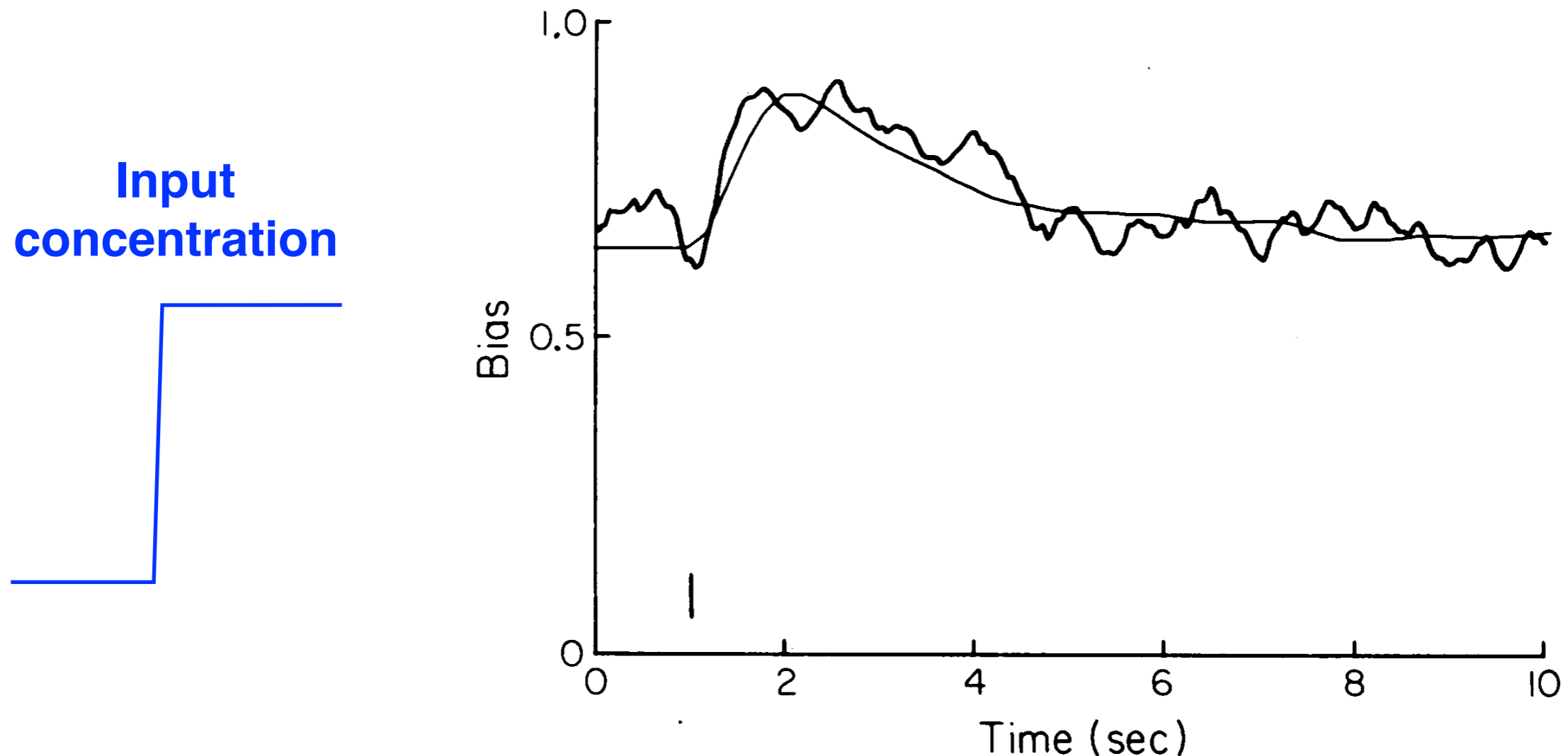
Probability for motor to rotate in CCW direction (runs) as a function of time in response to short pulse in external molecular concentration



***E. coli* integrates measured concentration observed during the last second and compare this with measured concentration during the previous 3 seconds. If difference is positive then increase the probability of runs, otherwise increase the probability of tumbles.**

# Adaptation

Probability for motor to rotate in CCW direction (runs) as a function of time in response to a sudden increase in external molecular concentration



**E. coli adapts to the new level of concentration in about 4 seconds.  
This enables E. coli to be very sensitive to changes in  
concentration over a very broad range of concentrations!**

J. E. Segall, S. M. Block, and H. C. Berg,  
PNAS **83**, 8987–8991 (1986)

# How efficient is motor of *E. coli*?

Energy source for rotary motor are charged protons

Each proton gains energy due to Transmembrane electric potential difference

$$\delta\psi \approx -120\text{mV}$$

Change in pH

$$\delta U = (-2.3k_B T/e)\Delta pH \approx -50\text{mV}$$

Total protonmotive force

$$\Delta p = \delta\psi + \delta U \approx -170\text{mV}$$

Need 1200 protons per one body revolution

Input power

$$P_{\text{in}} = n \times e\Delta p \times f = 1200 \times 0.17\text{eV} \times 10\text{Hz} \approx 3.2 \times 10^5 \text{pN nm/s}$$

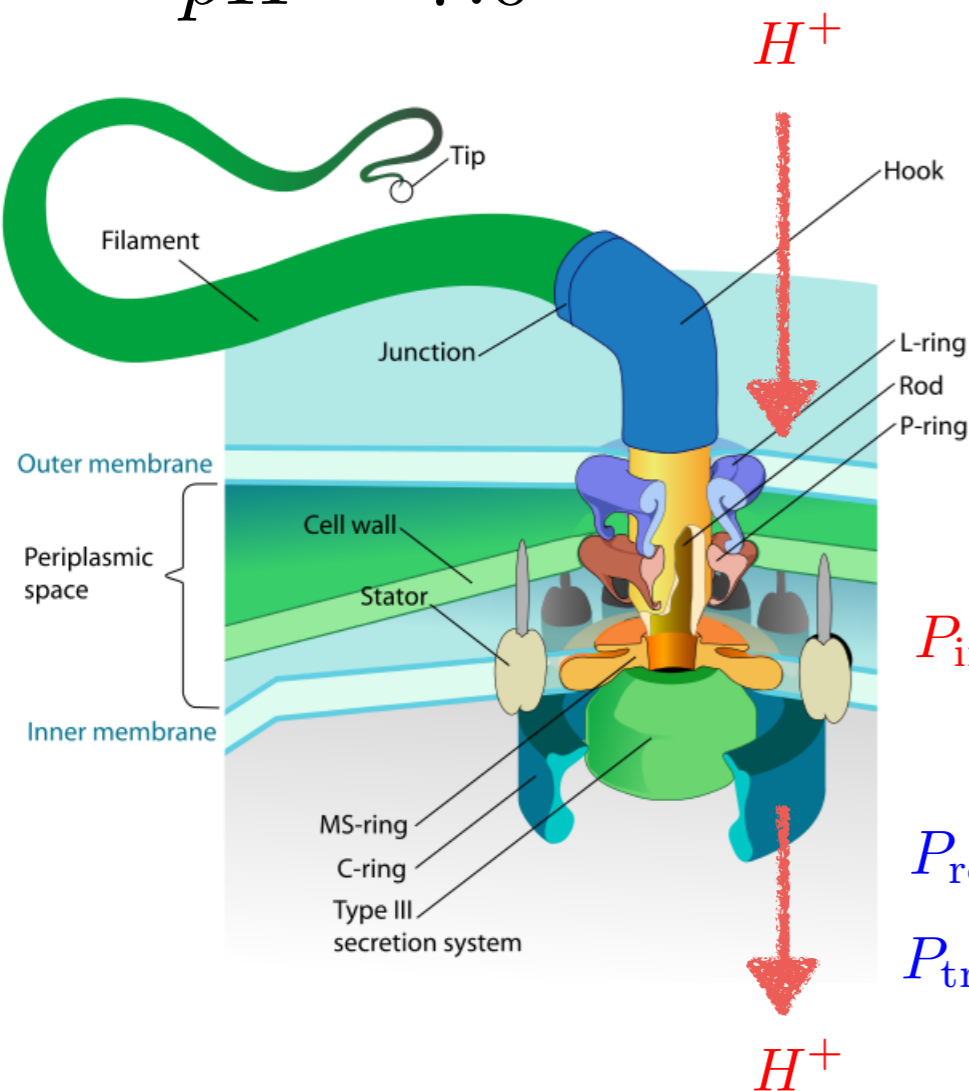
Power loss due to Stokes drag

$$P_{\text{rot}} = N \times (2\pi f) \approx 4600\text{pN nm} \times (20\pi\text{Hz}) \approx 2.9 \times 10^5 \text{pN nm/s}$$

$$P_{\text{trans}} = F \times v \approx 0.4\text{pN} \times 20000\text{nm/s} \approx 8 \times 10^3 \text{pN nm/s}$$

$$\frac{P_{\text{trans}} + P_{\text{rot}}}{P_{\text{in}}} \approx 90\%$$

$$pH = 7.0$$



$$pH \approx 7.8$$

Motor efficiency

# pH value of solutions

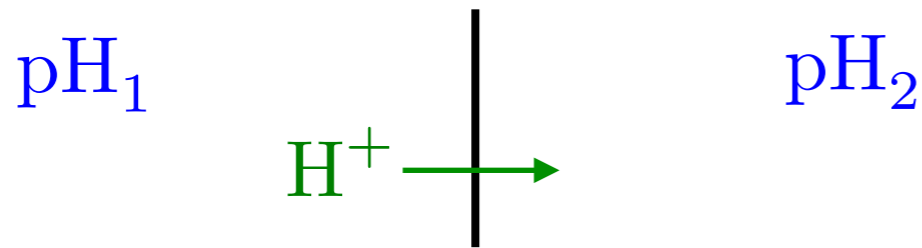
$$\frac{[\text{H}^+][\text{OH}^-]}{c_0^2} = \frac{[\text{H}_2\text{O}]K_{\text{eq}}(T, p)}{c_0^2} \approx 10^{-14}$$

$c_0 = 1\text{M}$  **at room temperature**

$$\text{pH} = -\log_{10}([\text{H}^+]/c_0)$$

$$\text{pOH} = -\log_{10}([\text{OH}^-]/c_0) \approx 14 - \text{pH}$$

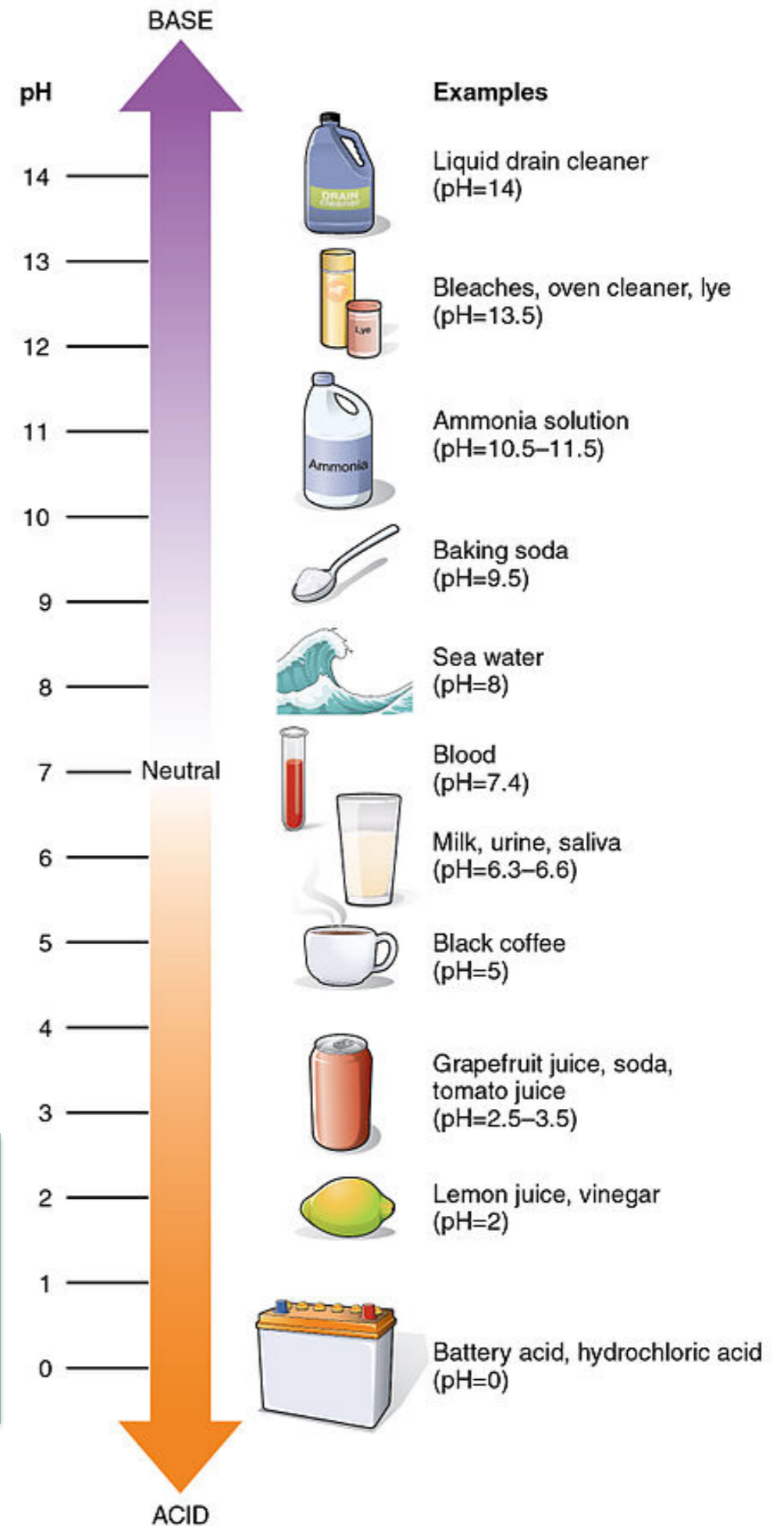
**How much free energy is changed when H<sup>+</sup> goes to environment with different pH?**



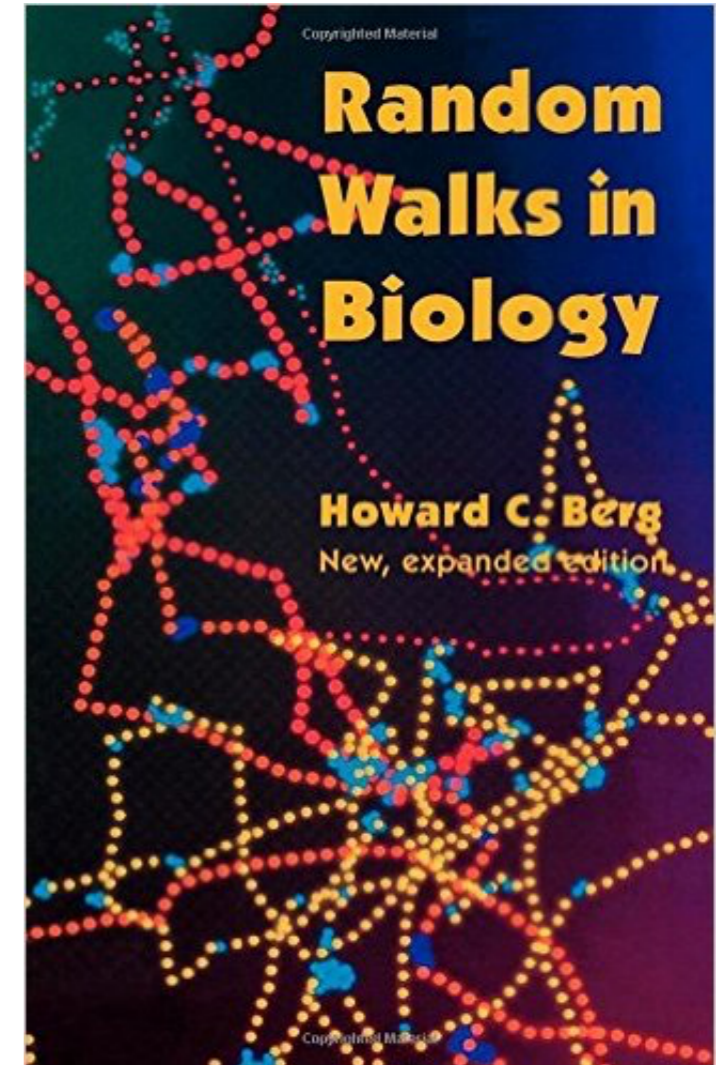
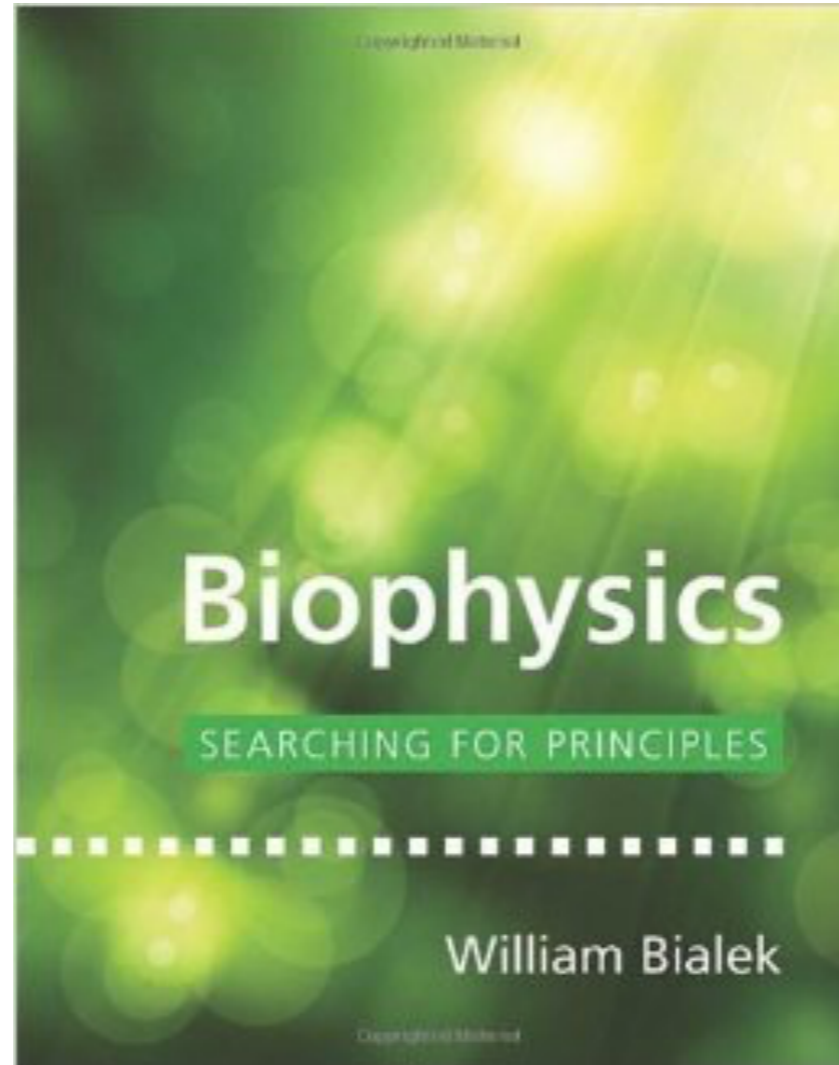
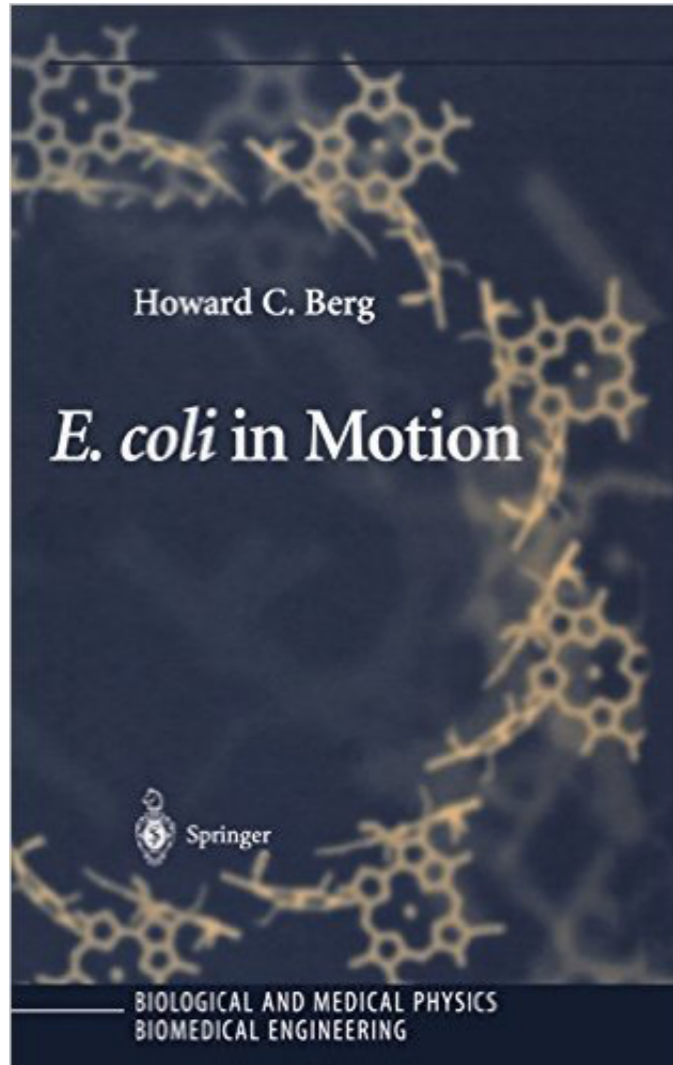
$$\mu_2 - \mu_1 = k_B T \ln([\text{H}^+]_2/[\text{H}^+]_1)$$

$$E = \frac{\mu_2 - \mu_1}{e_0} \approx -\frac{2.3026 k_B T}{e_0} (\text{pH}_2 - \text{pH}_1)$$

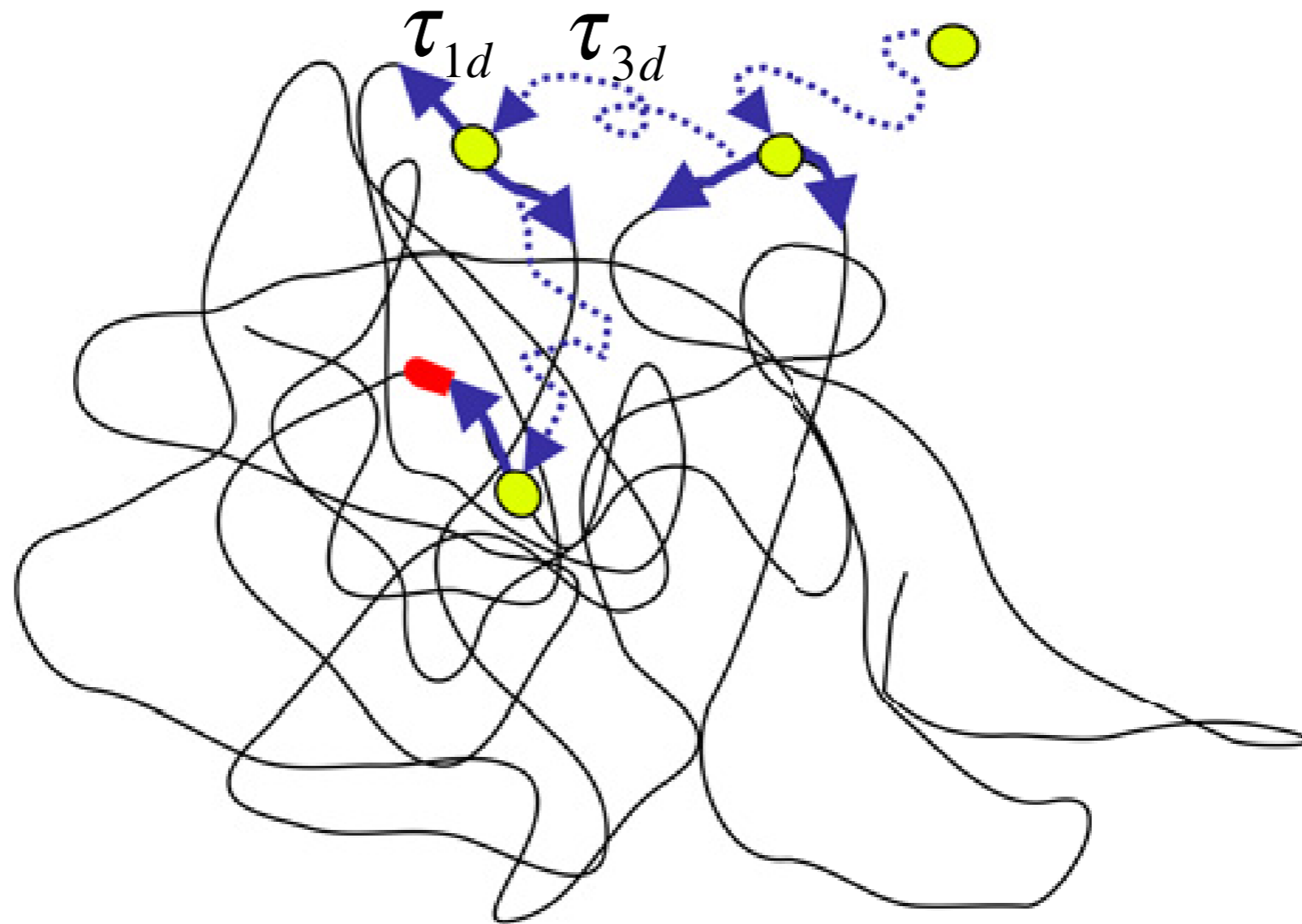
**Nernst electric potential  $E$**



# Further reading

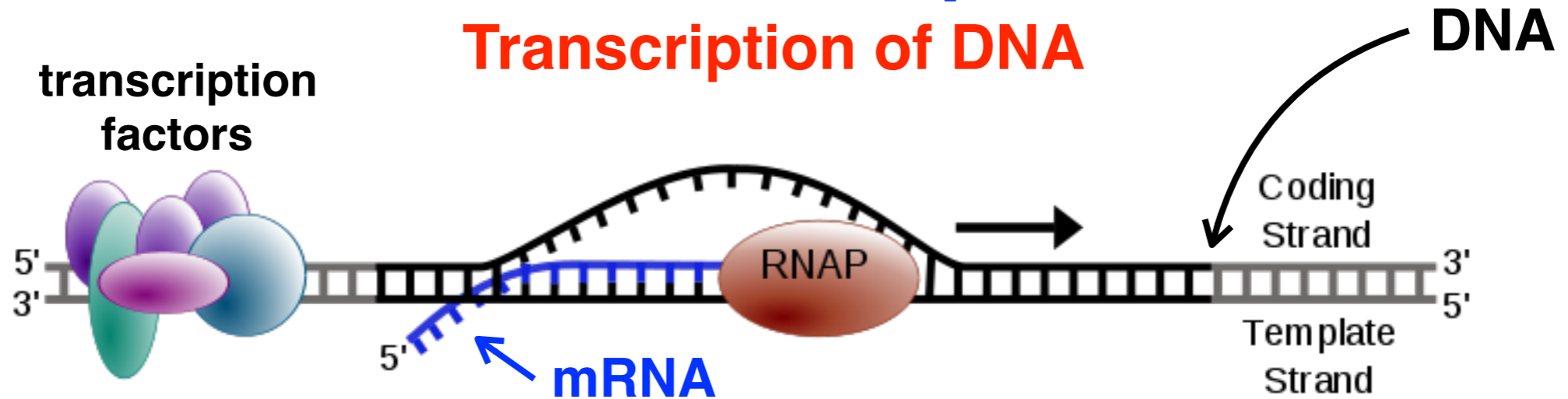


# How proteins find target sites on DNA?



# Production of new proteins

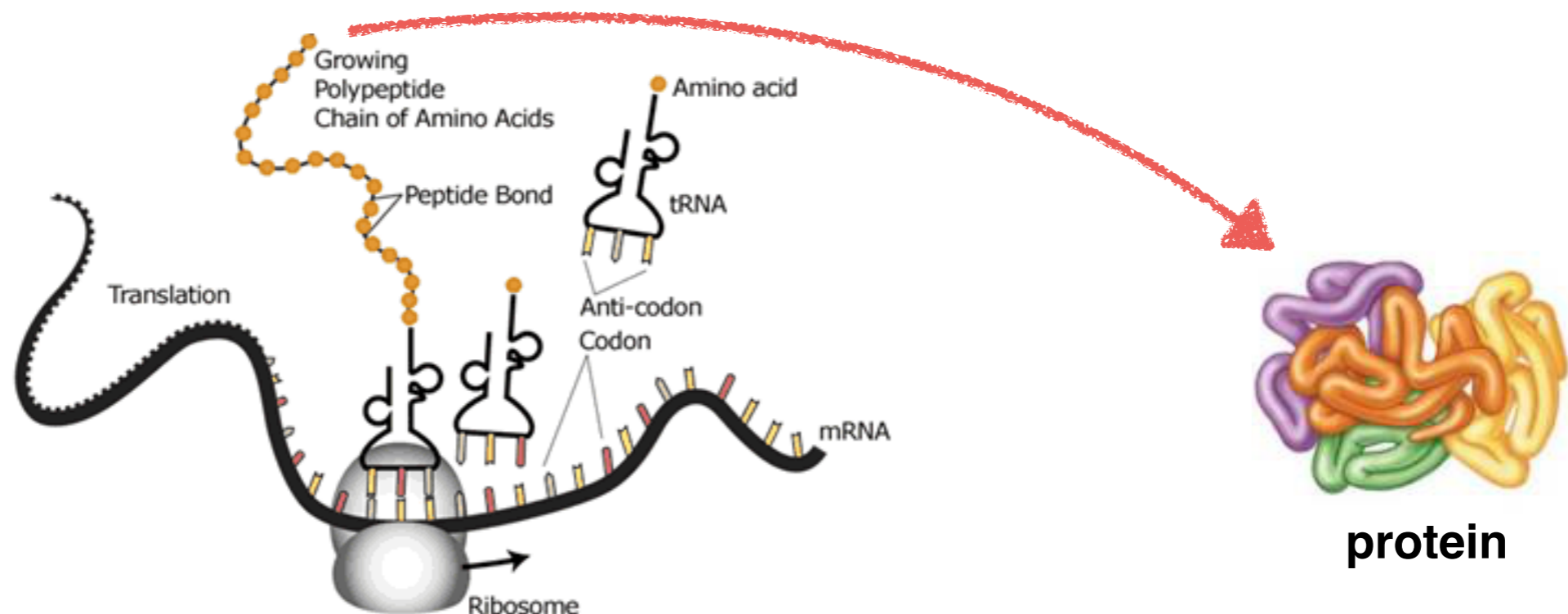
## Transcription of DNA



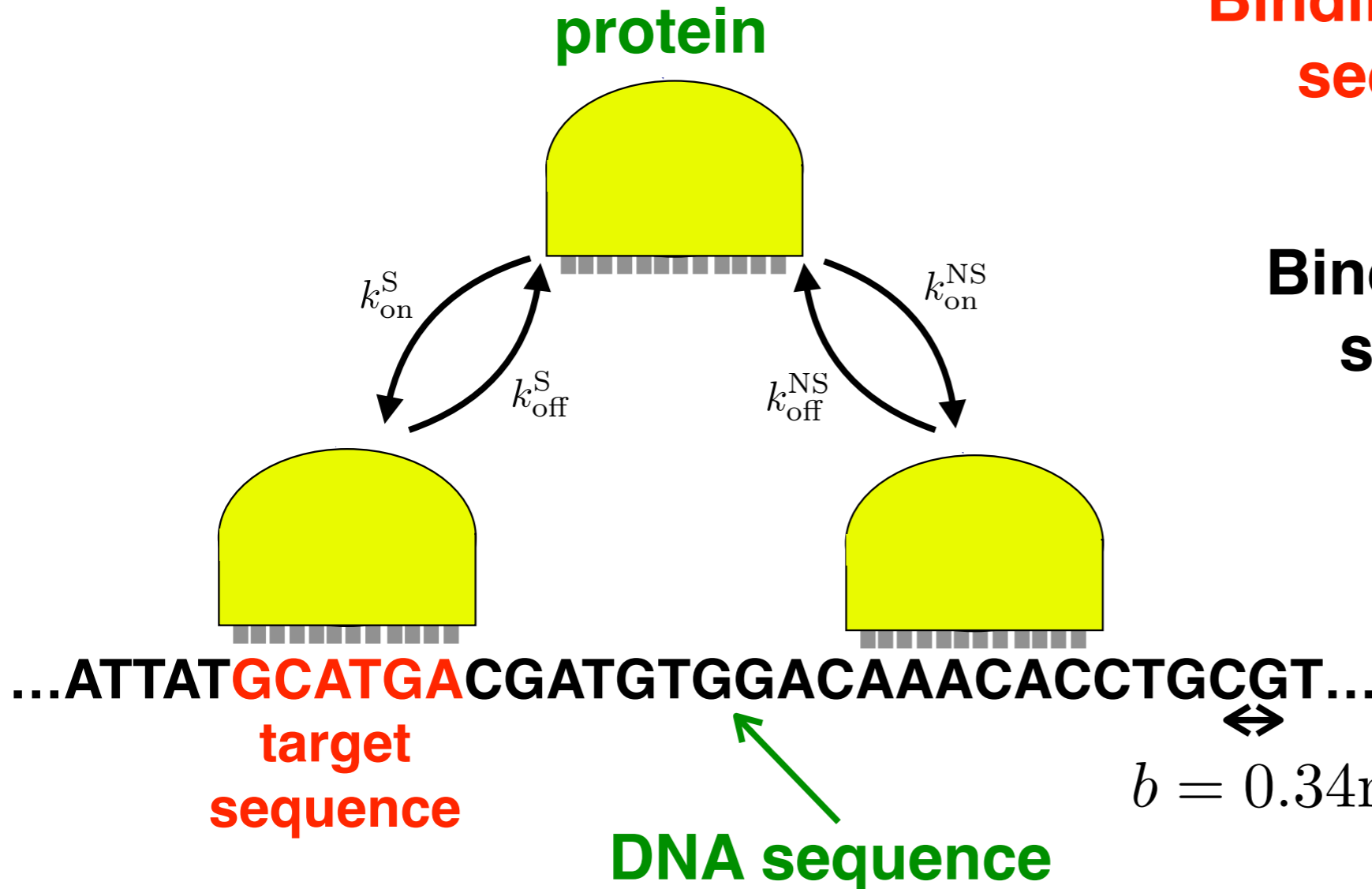
Transcription factors are proteins, which bind to specific locations on DNA, and they help recruiting RNA polymerase (RNAP) that makes a messenger RNA (mRNA) copy of certain DNA segment.

Note: some transcription factors (repressors) also prevent transcription.

## Translation of mRNA



# Protein-DNA interactions



**Binding to specific target sequence is strong**

$$\Delta G^S \sim 20 - 25k_B T$$

**Binding to nonspecific sequence is weak**

$$\Delta G^{\text{NS}} \sim 5 - 10k_B T$$

**(Binding free energies can be modified by changing salt concentration, etc.)**

**on rates are diffusion limited**

$$k_{\text{on}}^S \approx k_{\text{on}}^{\text{NS}} \approx 4\pi D_3 b$$

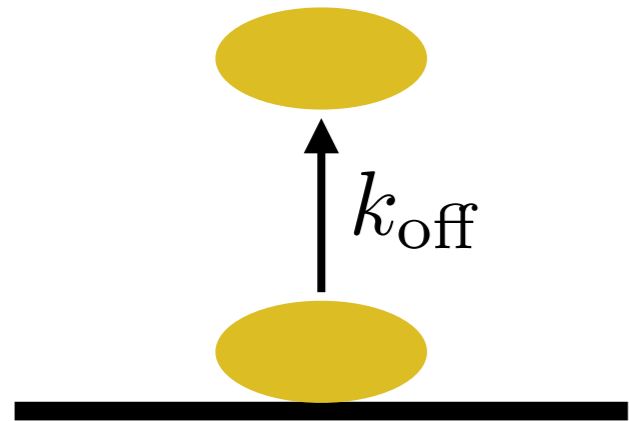
**off rates depend on binding strengths**

$$k_{\text{off}}^S = A_s e^{-\Delta G^S / k_B T} \ll k_{\text{off}}^{\text{NS}} = A_s e^{-\Delta G^{\text{NS}} / k_B T}$$

$$\frac{k_{\text{off}}^S}{k_{\text{off}}^{\text{NS}}} \sim 10^{-6}$$



# How long proteins remain bound on DNA?



Probability that protein unbinds in a small time interval  $\Delta t$  :

$$k_{\text{off}}\Delta t$$

Probability that protein remains bound for time  $t$  and then it unbinds between time  $t$  and  $t + \Delta t$  :

$$k_{\text{off}}\Delta t \times (1 - k_{\text{off}}\Delta t)^{t/\Delta t}$$

limit  $\Delta t \rightarrow 0$

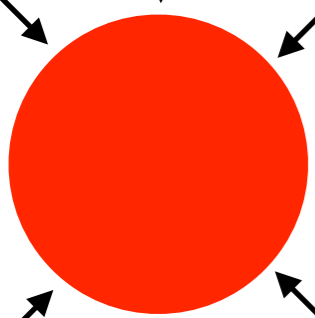
$$p(t) = k_{\text{off}}e^{-k_{\text{off}}t}$$

Average binding time  $\langle t \rangle = \int_0^{\infty} t p(t) dt = \frac{1}{k_{\text{off}}}$

Proteins remain bound to specific target sites for minutes to hours, while they unbind from nonspecific sites after milliseconds to seconds.

# How quickly proteins find target sites on DNA?

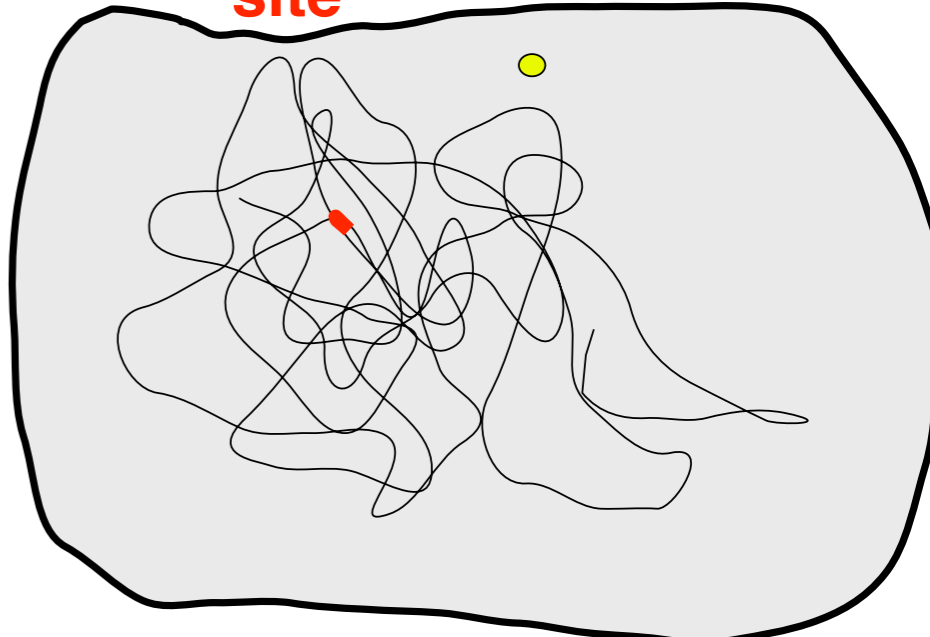
## Characteristic search time via 3D diffusion



Approximate target site as absorbing sphere of radius  $b = 0.34\text{nm}$

concentration of free proteins far from the target site  $c(r \rightarrow \infty) = [P]$

rate of absorption (see slide 6)

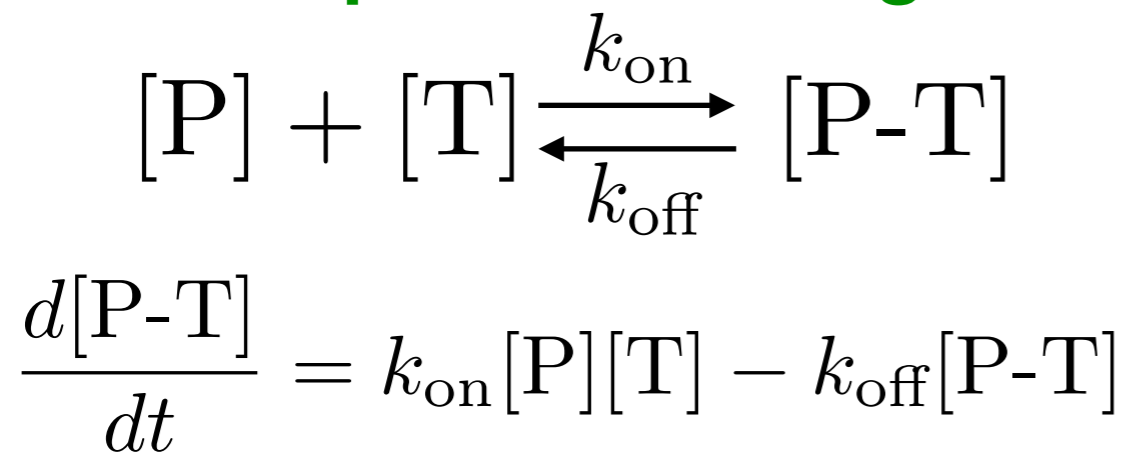
$$I_0 = 4\pi D_3 b [P] \equiv k_{\text{on}} [P]$$


target site

protein

$$k_{\text{on}} = 4\pi D_3 b$$

## Kinetics of protein binding/unbinding



short time binding kinetics for initially empty target sites  $[P-T]=0$

$$\frac{d[P-T]}{dt} = (k_{\text{on}} [T]) [P] \equiv \frac{[P]}{t_s}$$

characteristic search time

$$t_s = (k_{\text{on}} [T])^{-1}$$

$[P-T]$  concentration of proteins bound to target sites

$[P]$  concentration of free proteins

$[T]$  concentration of empty target sites

# How quickly proteins find target sites on DNA?

## Characteristic search time via 3D diffusion

$$k_{\text{on}} = 4\pi D_3 b \quad t_s = (k_{\text{on}} [\text{T}])^{-1}$$

### 1917 Smoluchowski theory

### Example: characteristic search time for lac repressor protein in *E. coli*

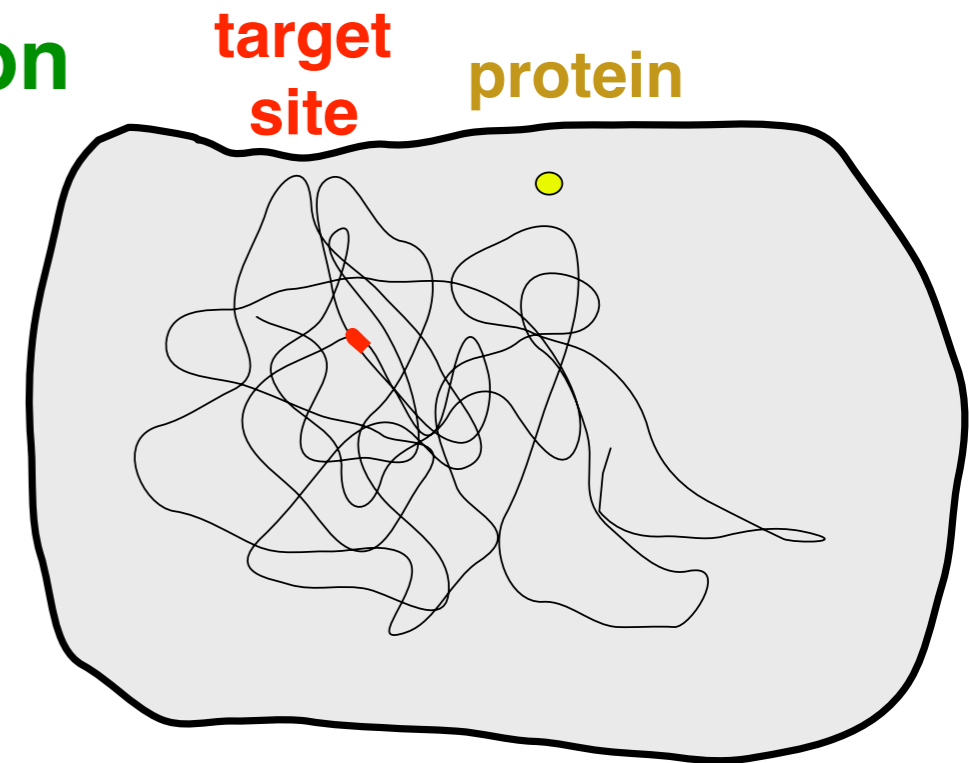
$$b \approx 0.34 \text{ nm} \quad D_3 \approx 30 \mu\text{m}^2/\text{s}$$

$$[\text{T}] \sim 1 \text{ per cell} \sim 10^{-9} \text{ M}$$

$$k_{\text{on}} \sim 10^8 \text{ M}^{-1} \text{ s}^{-1} \quad t_s \sim 10 \text{ s}$$

### in vitro experiments (1970)

$$k_{\text{on}}^{\text{exp}} \sim 10^{10} \text{ M}^{-1} \text{ s}^{-1} \quad t_s \sim 0.1 \text{ s}$$



### Molar concentration

$$1 \text{ M} = 6 \times 10^{26} \text{ m}^{-3}$$

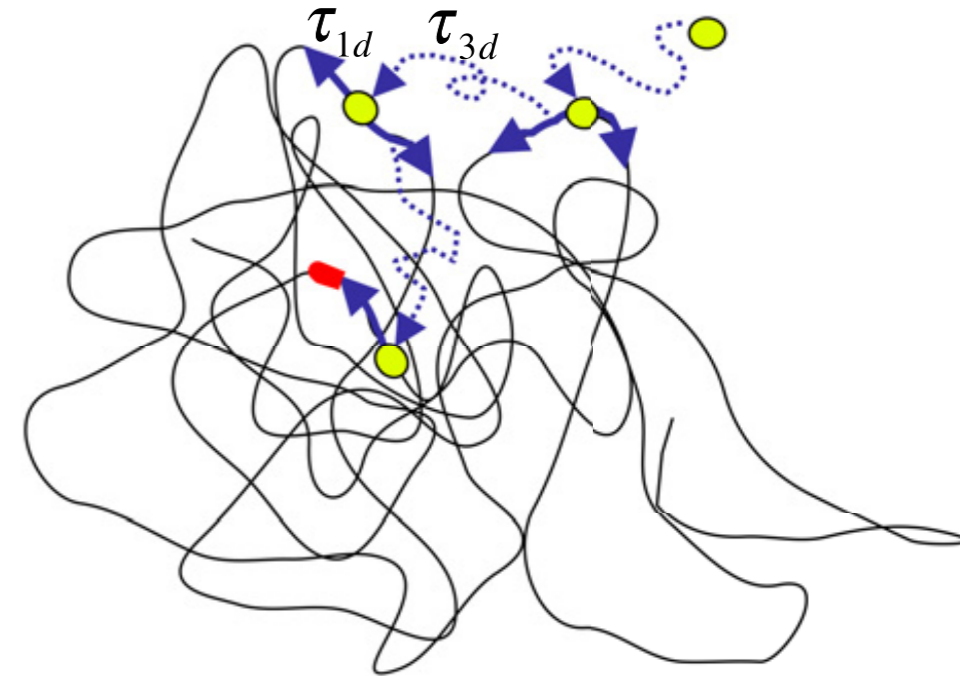
A.D. Riggs *et al.*,  
J. Mol. Biol. **53**, 401-417 (1970)

**Why is experimentally observed rate 100 times larger?**

# Berg - von Hippel theory (1980s)

## (facilitated diffusion)

1. Proteins diffuse in space and non-specifically bind to a random location on DNA.
2. Proteins slide (diffuse) along the DNA.
3. Proteins jump (diffuse) to another random location on DNA and continue this sliding/jumping process until the target site is found.



$b = 0.34\text{nm}$   $L$  - DNA length

$D_3$  - diffusion constant in space

$D_1$  - diffusion constant along the DNA

**How long that is it take to find a target site in this process?**

# Berg - von Hippel theory (1980s)

First assume fixed sliding time  $\tau_{1d}$

Number of distinct sites visited during each sliding event

$$n = \sqrt{16D_1\tau_{1d}/(\pi b^2)}$$

(valid for  $n \gg 1$ )

Probability that target site is found during a sliding event

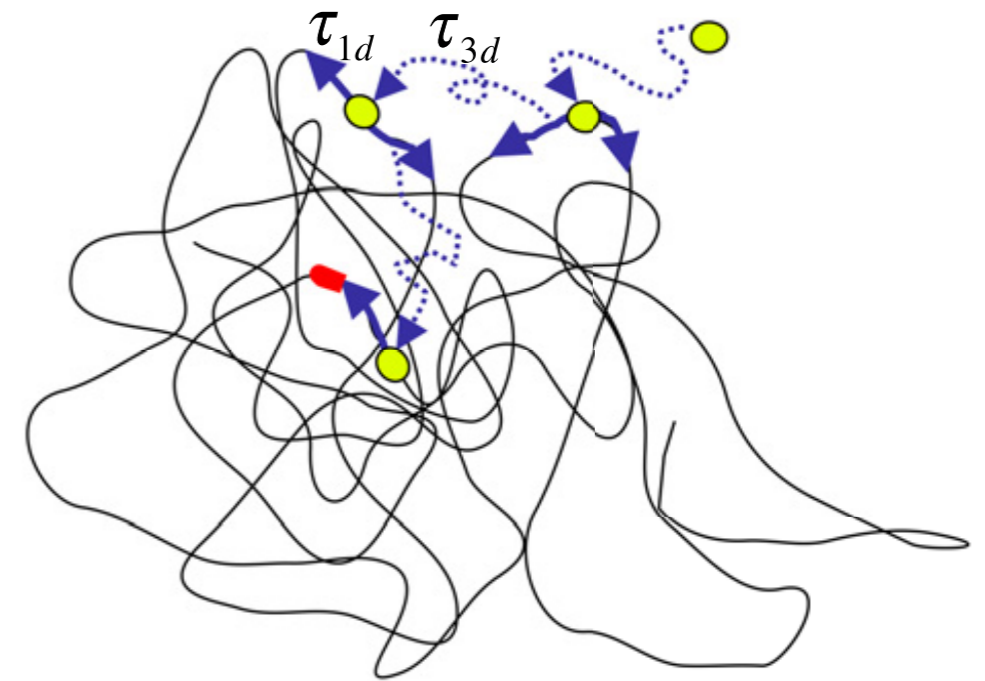
$$q = nb/L$$

Probability that target site is found exactly after  $N_R$  rounds

$$p(N_R) = q(1 - q)^{N_R - 1}$$

Average number of rounds needed to find the target

$$\overline{N_R} = \sum_{N_R=1}^{\infty} N_R p(N_R) = 1/q$$



$b = 0.34\text{nm}$   $L$  - DNA length

$D_3$  - diffusion constant in space

$D_1$  - diffusion constant along the DNA

$\tau_{3d}$  - characteristic jumping time

**Average search time**

$$\overline{t_s} = \overline{N_R} (\tau_{1d} + \tau_{3d})$$

O.G.Berg et al.,

Biochemistry **20**, 6929-48 (1981)

# Facilitated diffusion

**In reality sliding times are exponentially distributed**

$$p(\tau_{1d}) = k_{\text{off}}^{\text{NS}} e^{-k_{\text{off}}^{\text{NS}} \tau_{1d}}$$

$$\langle \tau_{1d} \rangle = \int_0^{\infty} d\tau_{1d} \tau_{1d} p(\tau_{1d}) = 1/k_{\text{off}}^{\text{NS}}$$

**Average number of distinct sites visited during each sliding**

$$\langle n \rangle = \int_0^{\infty} d\tau_{1d} p(\tau_{1d}) \sqrt{16D_1\tau_{1d}/(\pi b^2)}$$

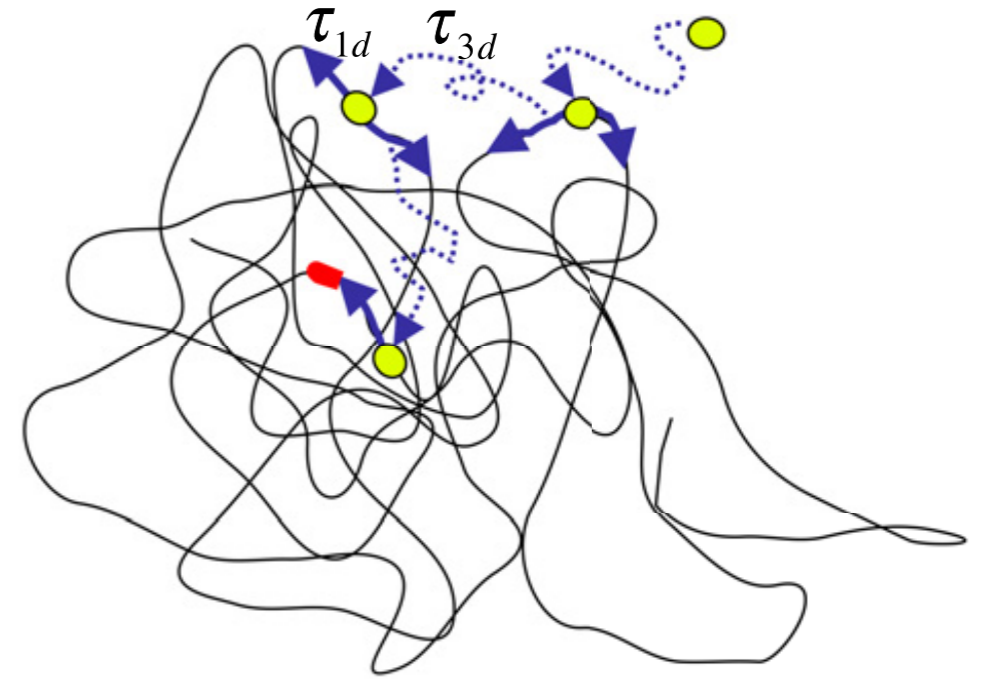
$$\langle n \rangle = 2\sqrt{D_1 \langle \tau_{1d} \rangle / (b^2)}$$

**Average probability that target site is found during a sliding event**

$$\langle q \rangle = \langle n \rangle b/L$$

**Average number of rounds  $N_R$  needed to find the target site**

$$\overline{\langle N_R \rangle} = 1/\langle q \rangle$$



$b = 0.34\text{nm}$   $L$  - DNA length

$D_3$  - diffusion constant in space

$D_1$  - diffusion constant along the DNA

$\tau_{3d}$  - characteristic jumping time

**Average search time**

$$\overline{\langle t_s \rangle} = \overline{\langle N_R \rangle} (\langle \tau_{1d} \rangle + \tau_{3d})$$

$$\overline{\langle t_s \rangle} = \frac{L}{2\sqrt{D_1 \langle \tau_{1d} \rangle}} (\langle \tau_{1d} \rangle + \tau_{3d})$$

# Facilitated diffusion

**Average search time**  $\overline{\langle t_s \rangle} = \frac{L}{\langle \ell_{sl} \rangle} (\langle \tau_{1d} \rangle + \tau_{3d})$

**Average sliding length**  $\langle \ell_{sl} \rangle = 2\sqrt{D_1 \langle \tau_{1d} \rangle}$

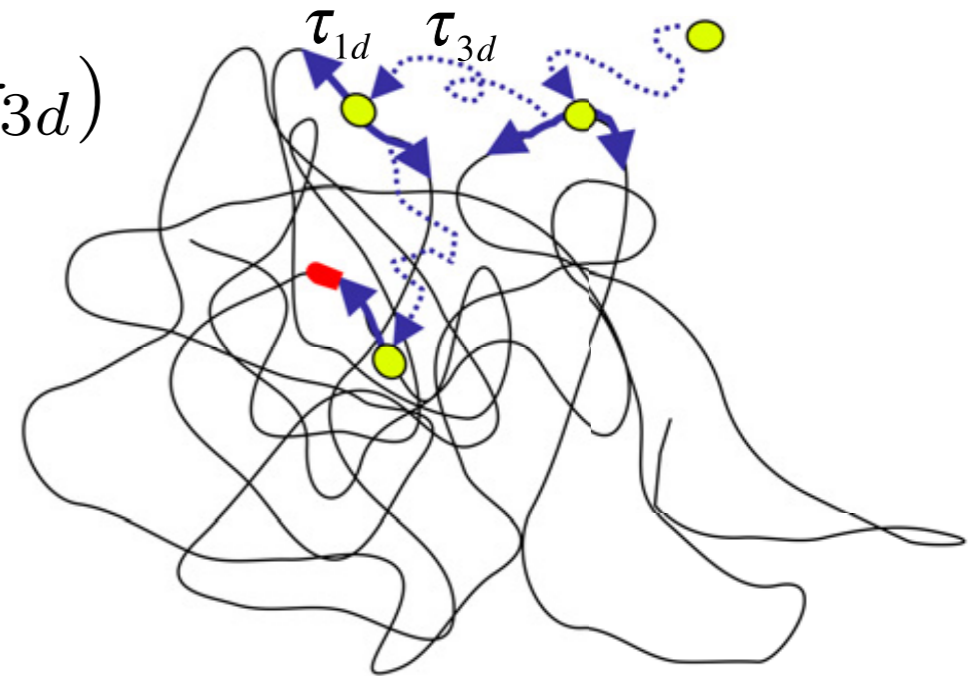
## Optimal search time

$$\frac{d\overline{\langle t_s \rangle}}{d\langle \tau_{1d} \rangle} = 0$$



$$\langle \tau_{1d} \rangle_{\text{opt}} = \tau_{3d}$$

$$\overline{\langle t_s \rangle}_{\text{opt}} = L \sqrt{\frac{\tau_{3d}}{D_1}}$$



$b = 0.34\text{nm}$   $L$  - DNA length  
 $D_3$  - diffusion constant in space  
 $D_1$  - diffusion constant along the DNA

## Search time for jumps alone

**Typical jump time**  $\tau_{3d} = \frac{1}{k_{\text{on}} [\text{NS}]} = \frac{V}{4\pi D_3 L}$

**Concentration of non-specific sites**  $[\text{NS}] = \frac{L/b}{V}$

**average number of jumps needed to find the target**  $\overline{N}_{\text{jumps}} = \frac{L}{b}$

$$\overline{t_{s,\text{jumps}}} = \overline{N}_{\text{jumps}} \tau_{3d} = \frac{V}{4\pi D_3 b}$$

## Search time for sliding alone

$$\langle t_s \rangle_{\text{sliding}} \sim \frac{L^2}{D_1}$$

**Search time speed up for facilitated diffusion**

$$\frac{\overline{t_{s,\text{jumps}}}}{\overline{\langle t_s \rangle}} = \frac{\langle \ell_{sl} \rangle}{b} \frac{\tau_{3d}}{(\langle \tau_{1d} \rangle + \tau_{3d})}$$

# Example: search time for target site in bacteria on DNA with $10^6$ base pairs

$$\tau_{3d} = 10^{-4} \text{ s}$$

$$D_1 = 0.05 \mu\text{m}^2/\text{s}$$

$$L = 1 \text{ mm}$$

$$b = 0.34 \text{ nm}$$

**search time for jumps alone**

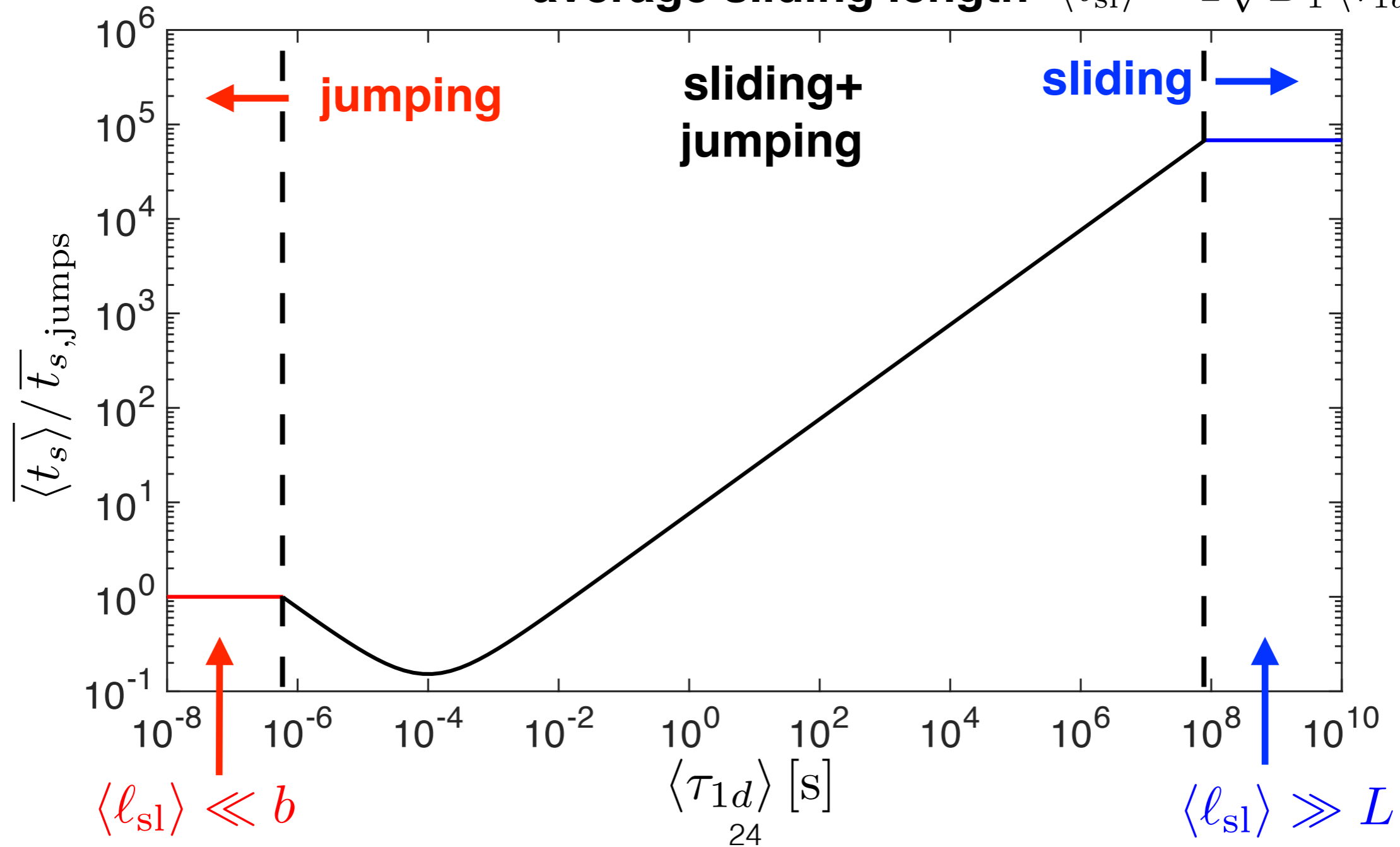
$$\overline{t_{s,\text{jumps}}} = (L/b)\tau_{3d} \approx 300 \text{ s}$$

**average search time**

$$\langle t_s \rangle = \frac{L}{\langle \ell_{sl} \rangle} (\langle \tau_{1d} \rangle + \tau_{3d})$$

**average sliding length**

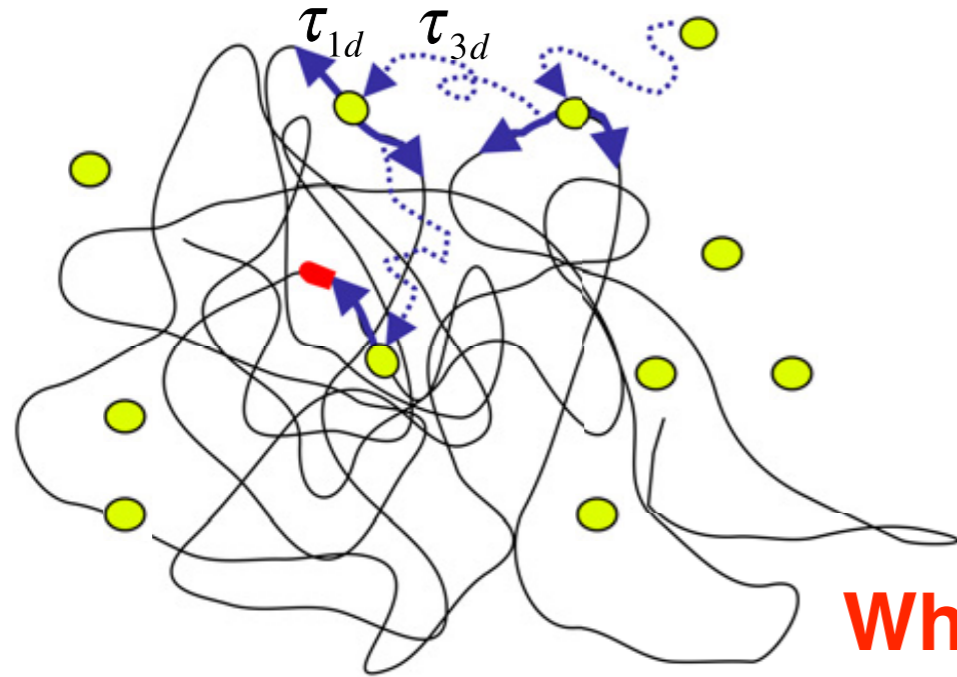
$$\langle \ell_{sl} \rangle = 2\sqrt{D_1 \langle \tau_{1d} \rangle}$$





# Simultaneous search for target site by multiple proteins

Interactions and collisions between proteins are ignored



Search times for target site by individual proteins are exponentially distributed

$$p_1(t_s) = \frac{1}{\langle t_s \rangle} e^{-t_s / \langle t_s \rangle}$$

What is the typical search time for the fastest of  $n$  independently searching proteins?

(Extreme value distributions)

$$p_n(t_s) = n \times p_1(t_s) \times \left( \int_{t_s}^{\infty} dt' p_1(t') \right)^{n-1} = \frac{n}{\langle t_s \rangle} e^{-nt_s / \langle t_s \rangle}$$

probability that one of  $n$  proteins finds the target site at time  $t_s$

probability that other  $n-1$  proteins take longer time to find the target site

Average search time is reduced by factor  $n$

$$\int_0^{\infty} dt_s t_s p_n(t_s) = \frac{\langle t_s \rangle}{n}$$