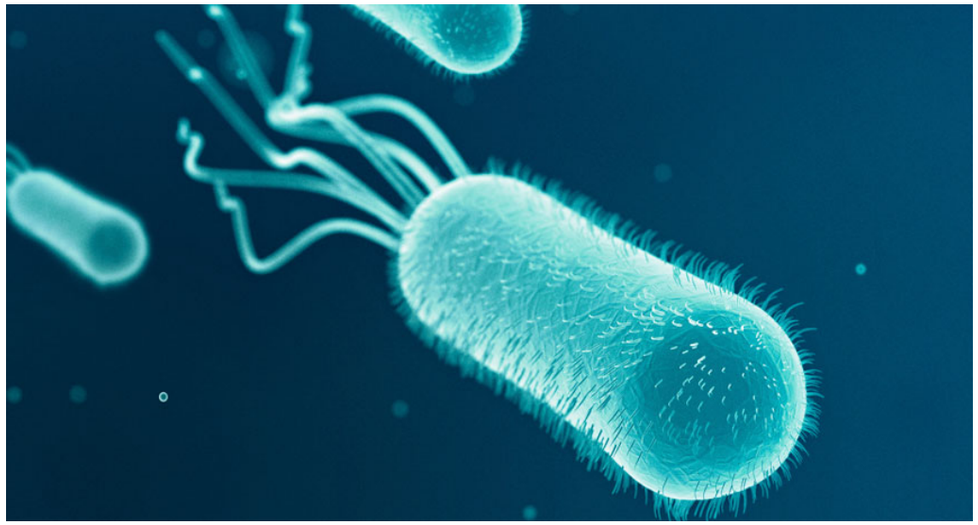
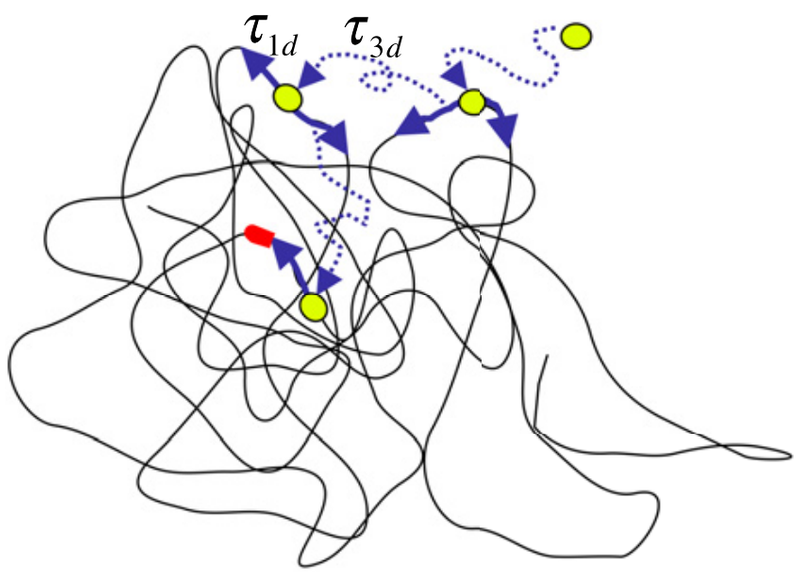
MAE 545: Lecture 18 (4/25)

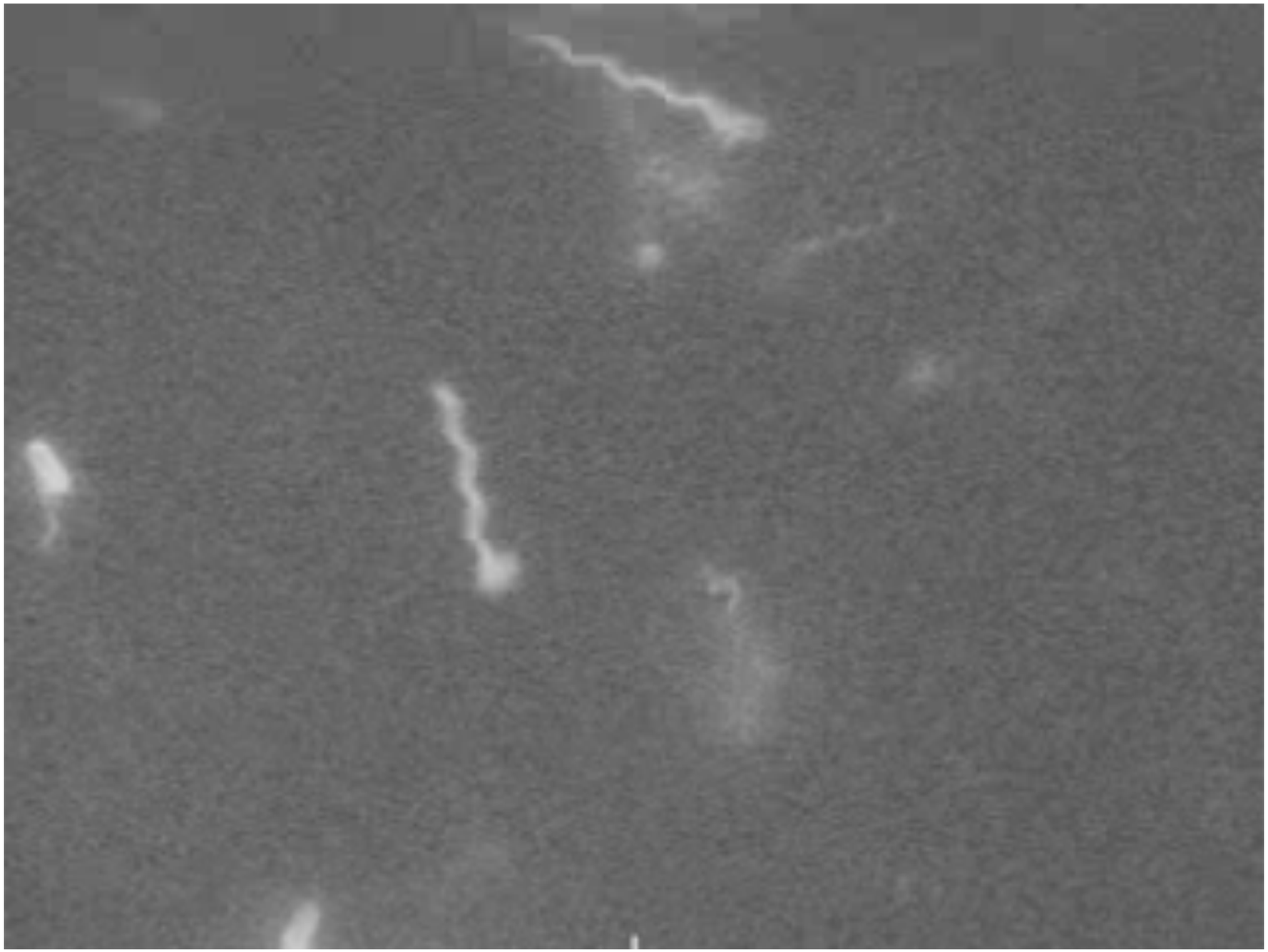
E. coli chemotaxis (continued)



How proteins find target sites on DNA?

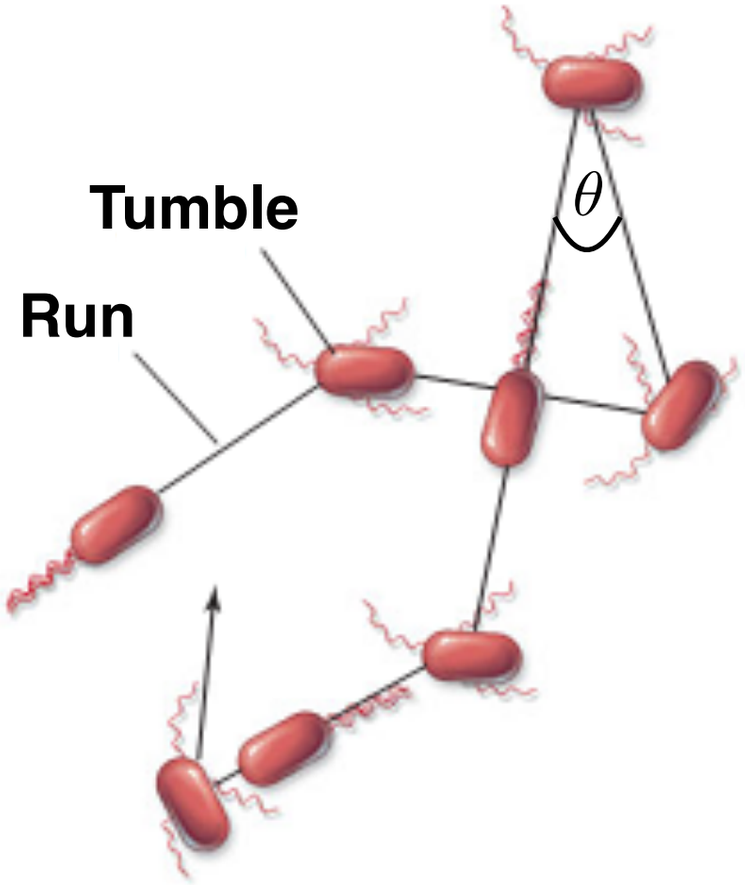


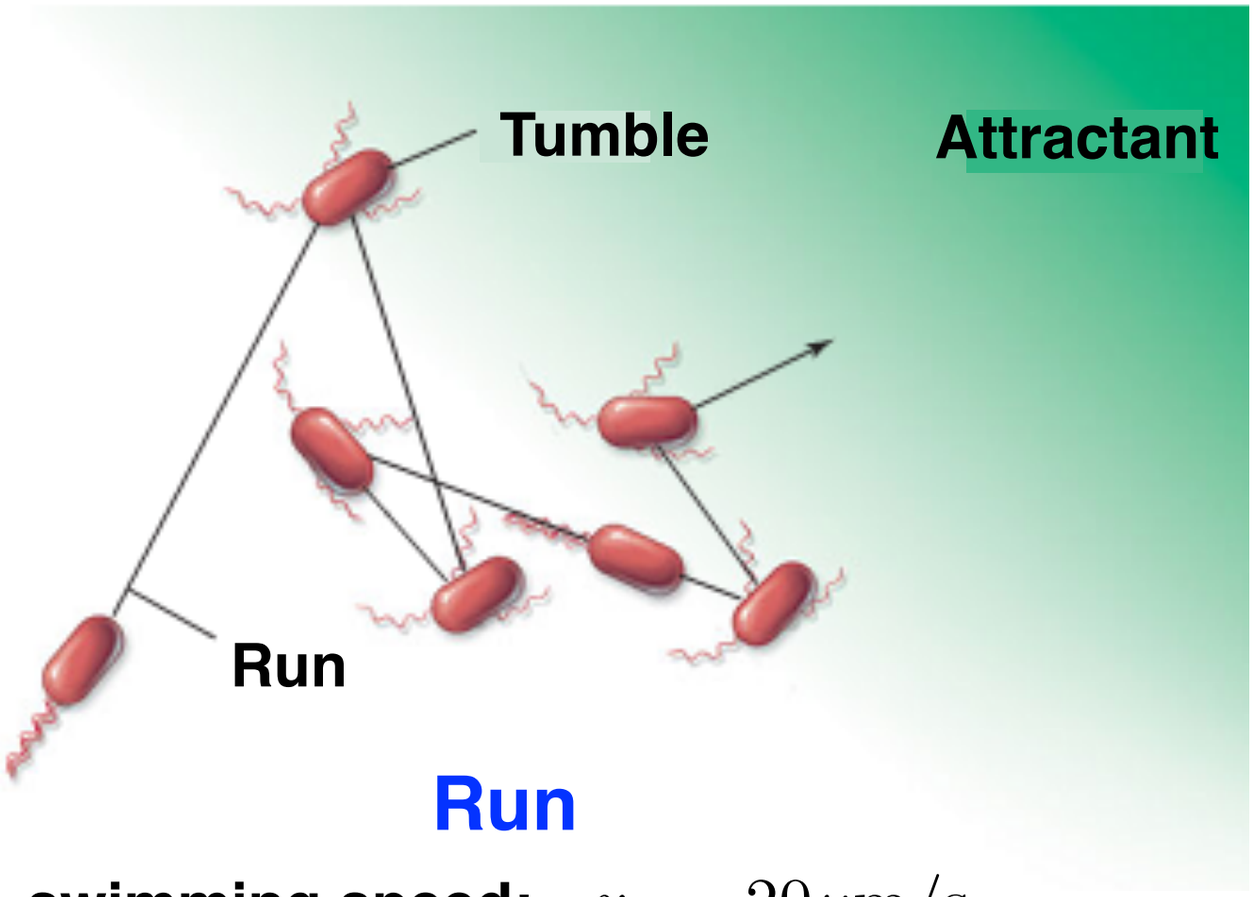
E. coli chemotaxis



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

E. coli chemotaxis





Run

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

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

all motors turning counter clockwise

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

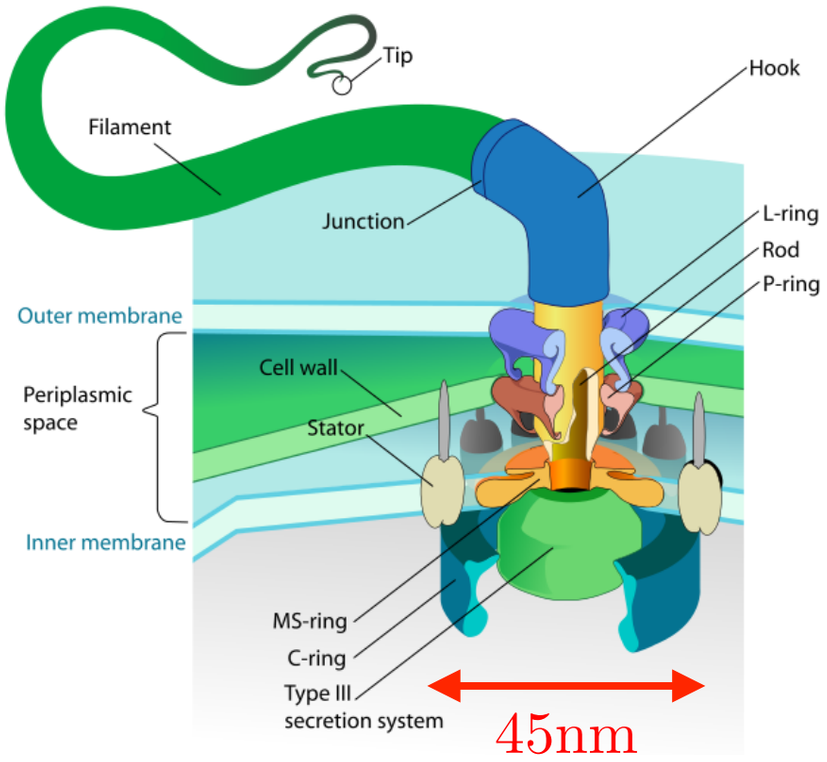
Tumble

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

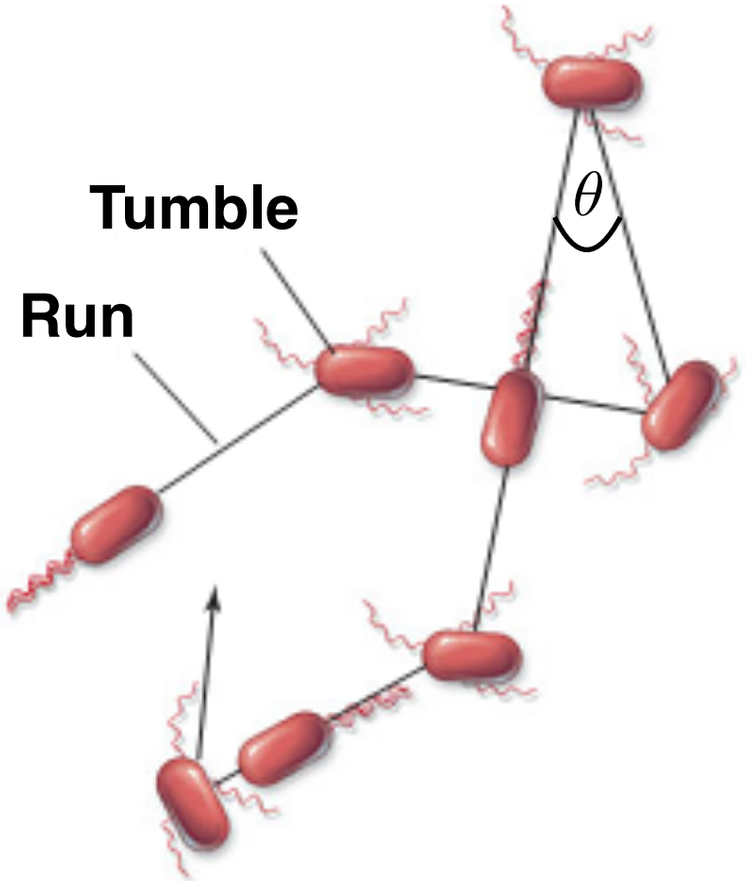
typical duration: $t_t \sim 0.1\mathrm{s}$

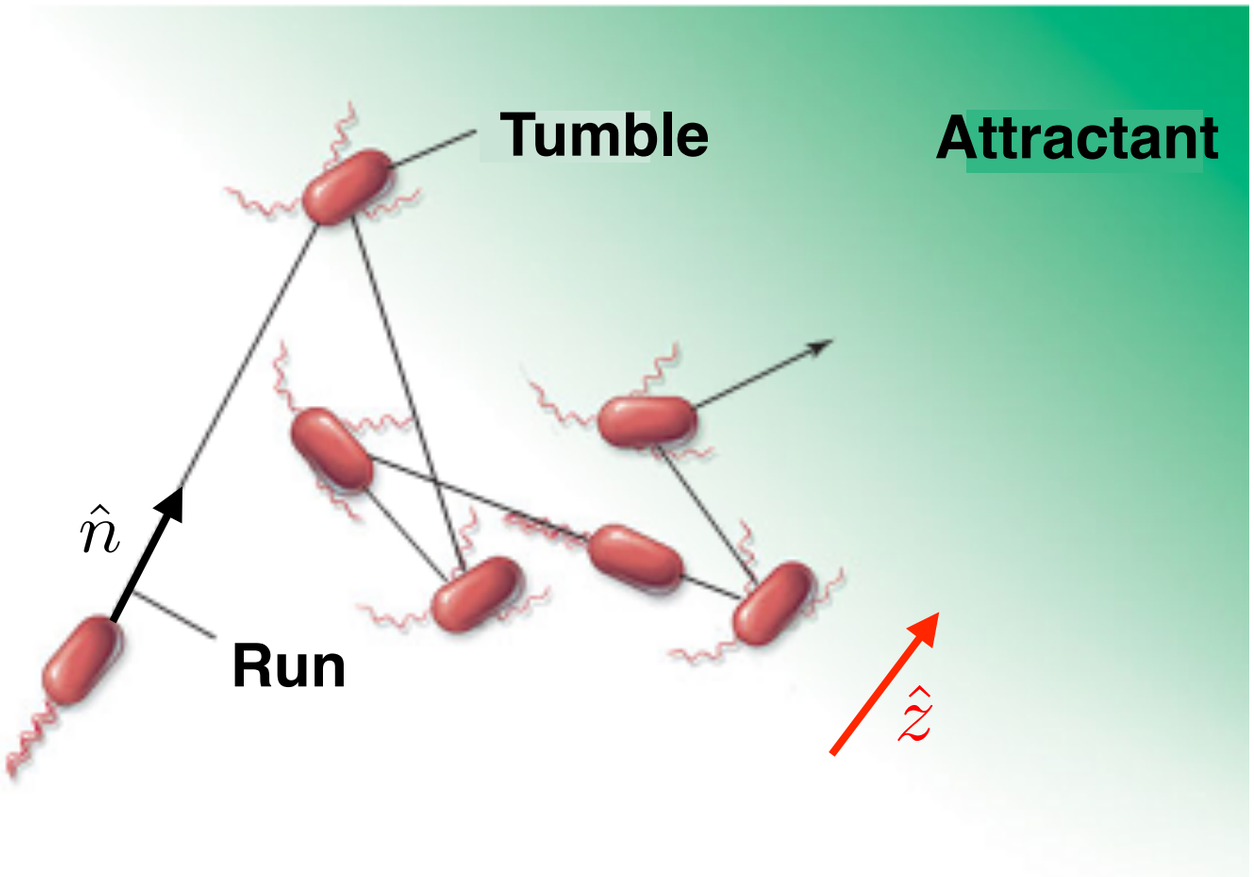
one or more motors turning clockwise

Rotary motor



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

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

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

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

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

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

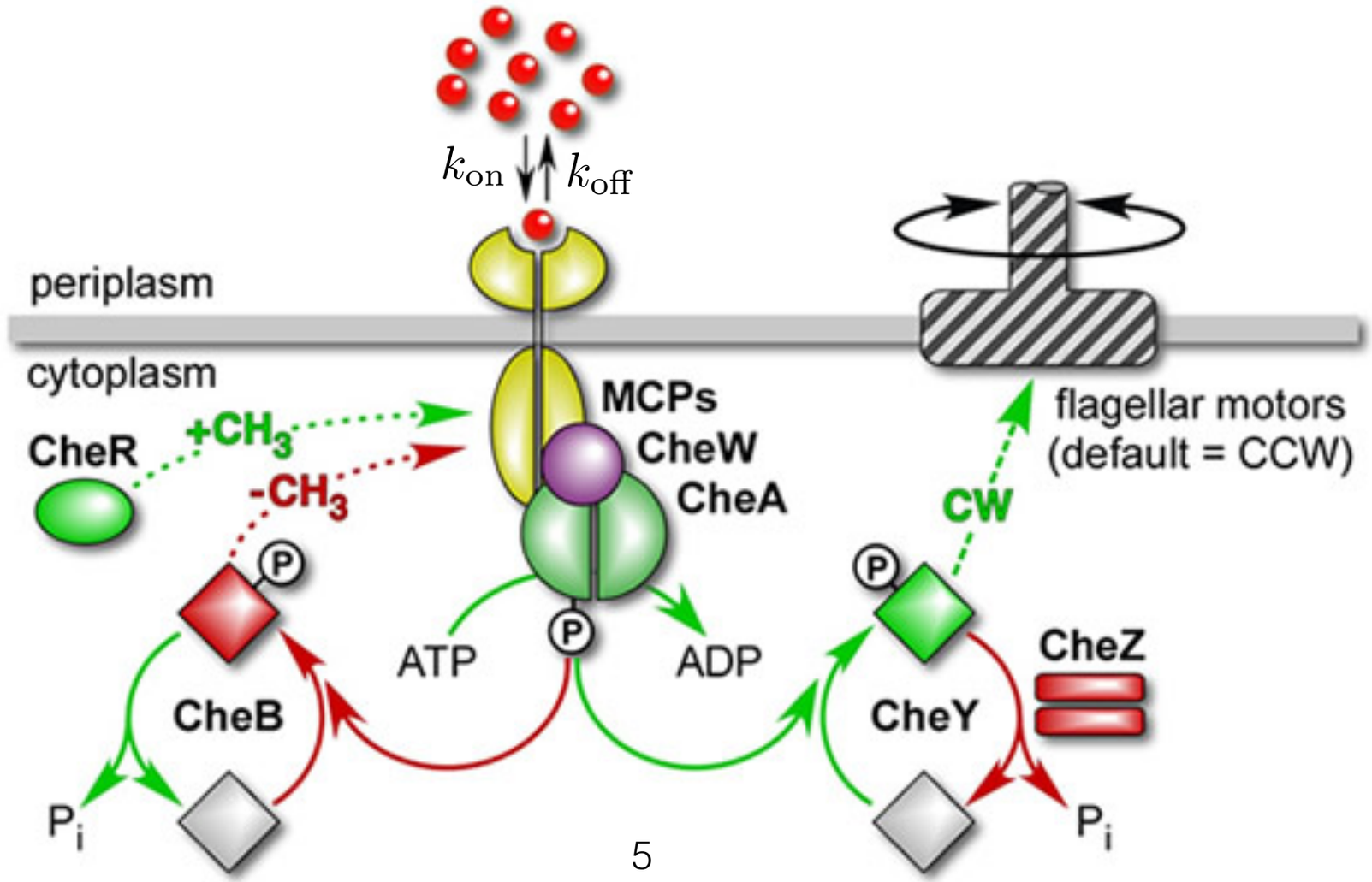
Sensing of environment

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

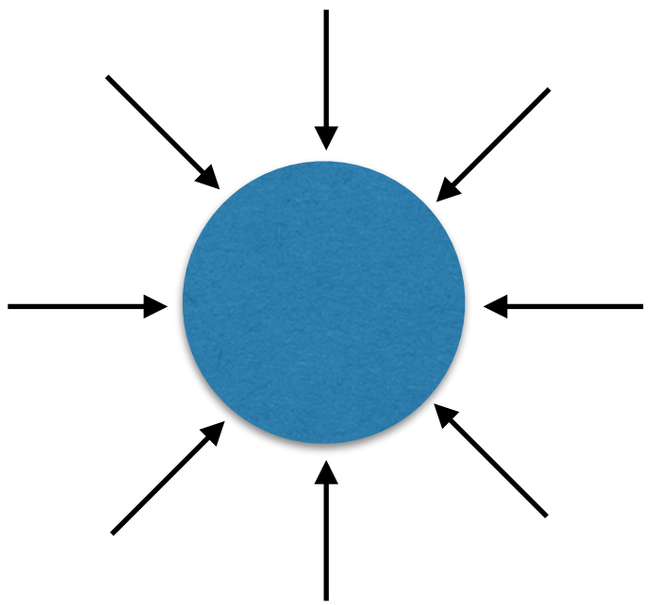
Average fraction of bound receptors p_B is related to concentration c of molecules.

$$p_B = \frac{c}{c + c_0} \qquad c_0 = \frac{k_{\text{off}}}{k_{\text{on}}}$$

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



Diffusion limited flux of molecules to E. coli



absorbing sphere

1917 Smoluchowski theory

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

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

$$c(R) = 0$$

steady state

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

flux density of molecules

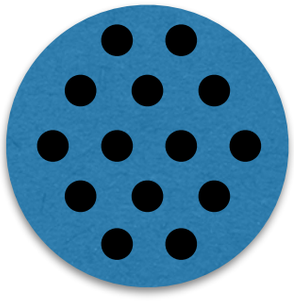
$$J(r) = -D\frac{\partial c(r)}{\partial r} = -\frac{Dc_\infty R}{r^2}$$

rate of absorbing molecules

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

diffusion constant for small molecules $D \approx 10^3 \mu\text{m}^2/s$

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



N* equally spaced absorbing disks of radius *s

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

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

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?

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

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

$$p(N) = \frac{\overline{N}^N E^{-\overline{N}}}{N!}$$

mean $\overline{N}$

standard deviation $\sigma_N = \sqrt{\overline{N}}$

Error in measurement

$$\mathrm{Err} \sim \frac{\sigma_N}{\overline{N}} \sim (R^3 c)^{-1/2}$$

for $c = 1\mu\mathrm{M} = 6 \times 10^{20}\mathrm{m}^{-3} \Rightarrow \mathrm{Err} \sim 4\%$

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!

$$t_0 \sim R^2/D \sim 10^{-3}s$$

$$\overline{N} \sim R^3 ct/t_0 \sim DRct$$

$$\mathrm{Err} \sim (DRct)^{-1/2}$$

for *t*=1s, precision improves to Err~0.1%

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

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

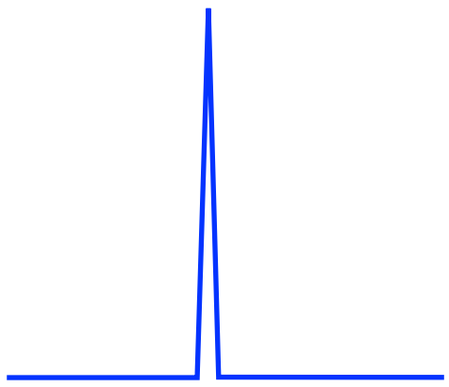
Molar concentration

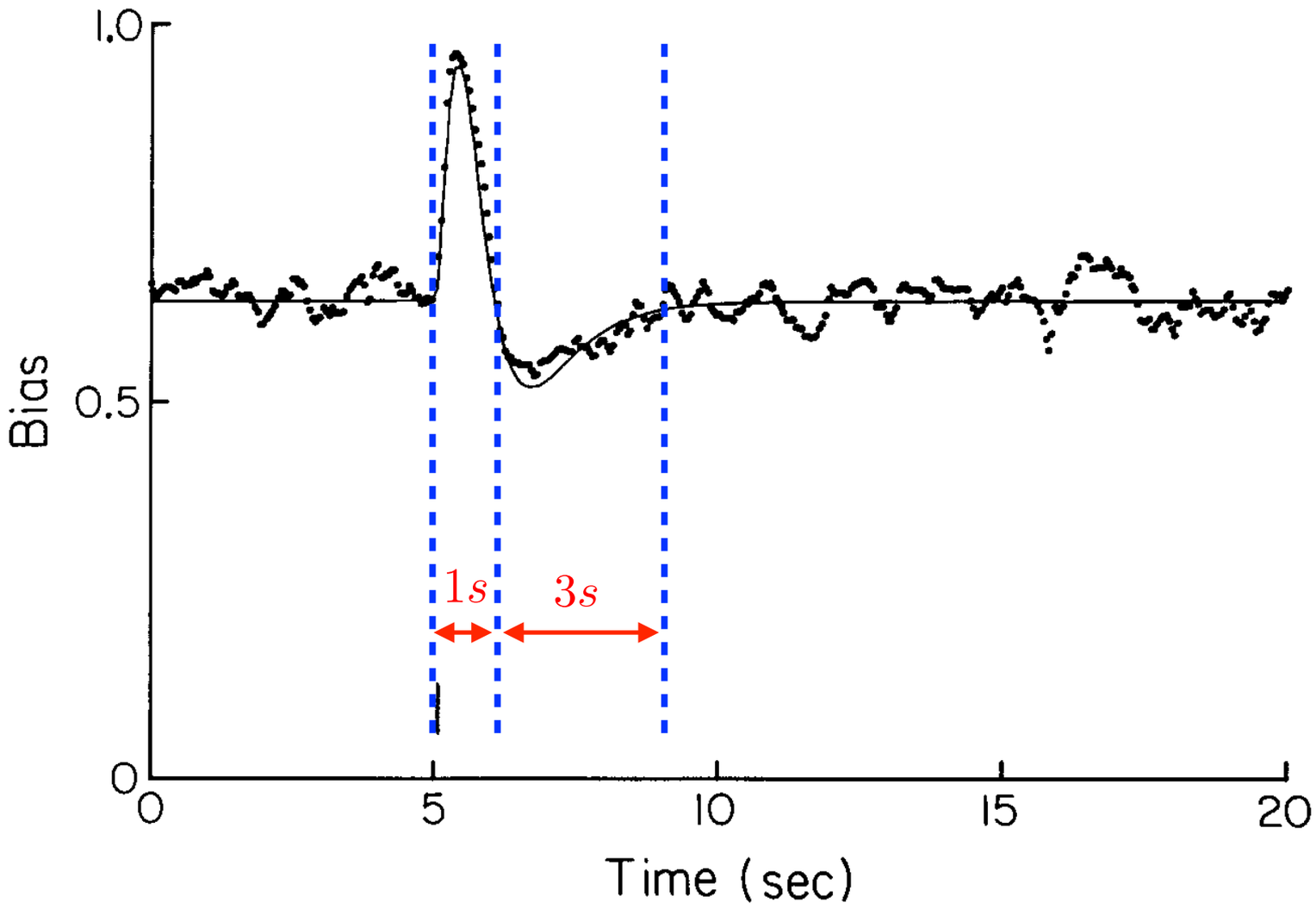
$$1M = 6 \times 10^{26}\mathrm{m}^{-3}$$

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

Input concentration





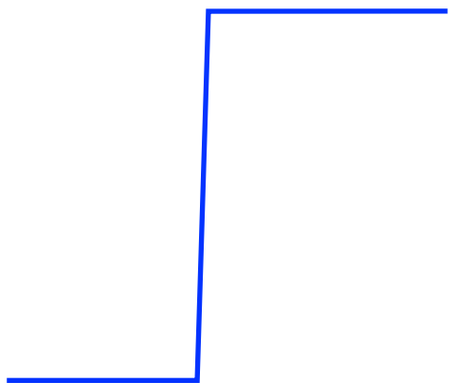
E. coli integrates measured concentration observed during the last second and compare this with measured concentration during the previous 3 seconds. If difference is positive then increase the probability of runs, otherwise increase the probability of tumbles.

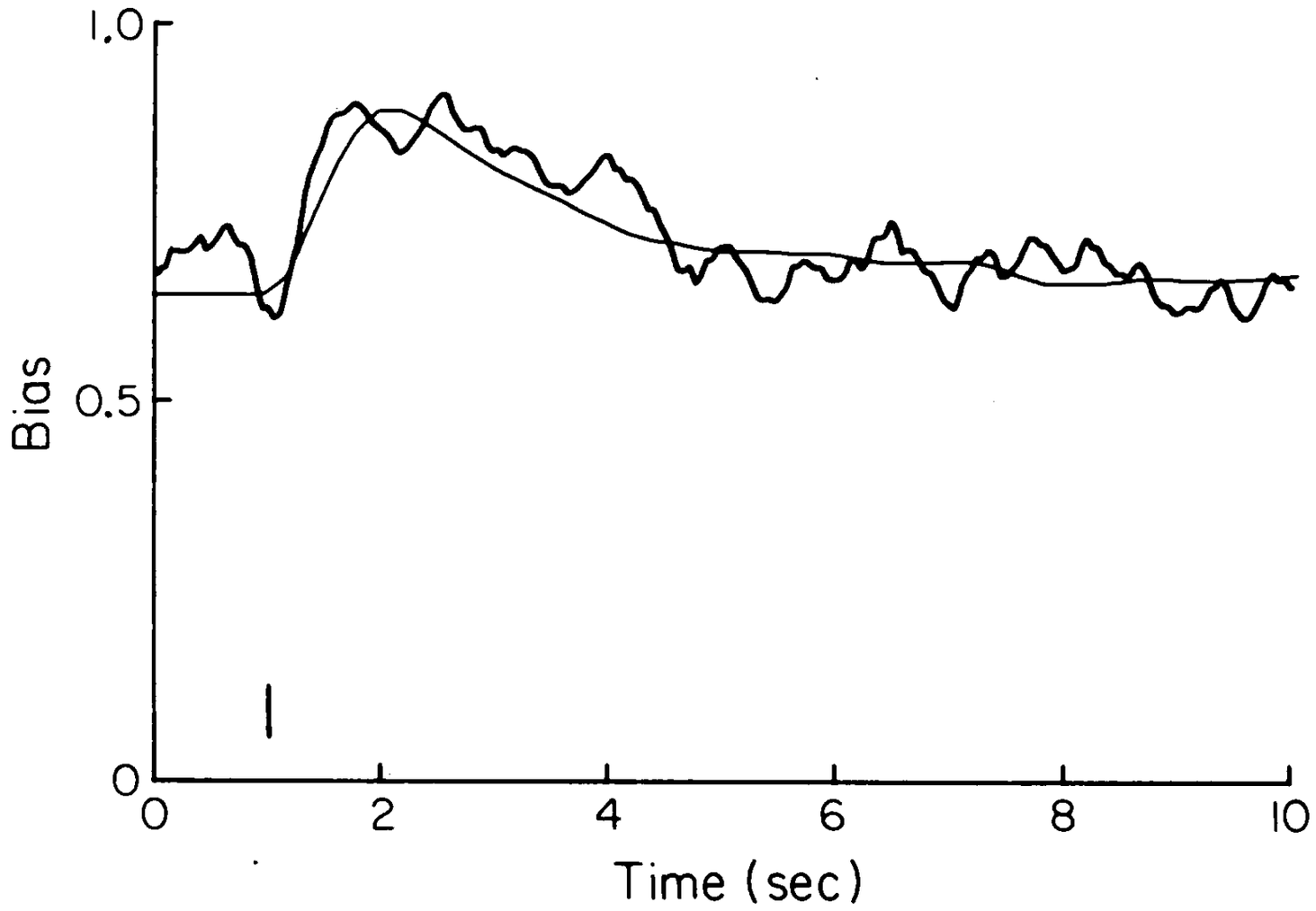
J. E. Segall, S. M. Block, and H. C. Berg,
PNAS **83**, 8987–8991 (1986)

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

Input concentration





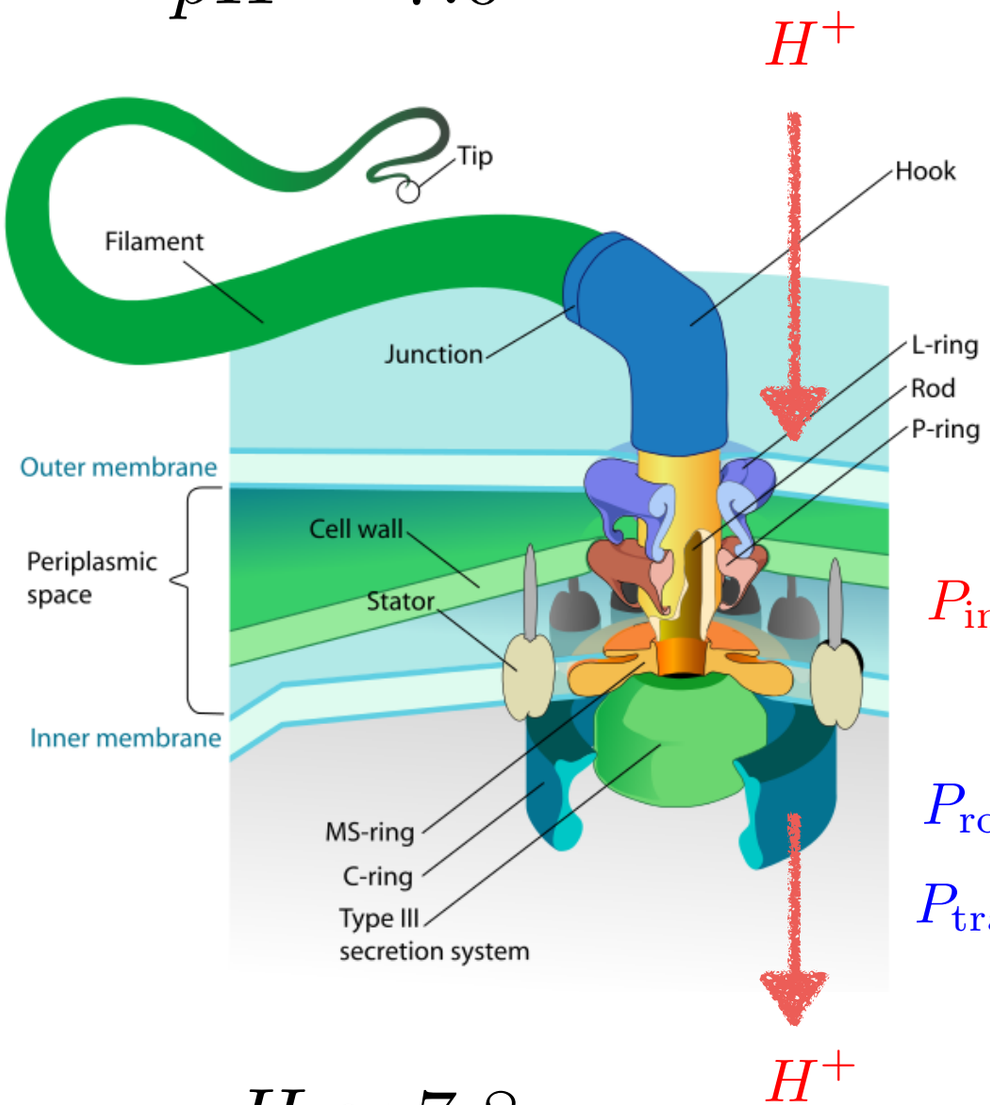
E. coli adapts to the new level of concentration in about 4 seconds. This enables E. coli to be very sensitive to changes in concentration over a very broad range of concentrations!

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

$$pH = 7.0$$



$$pH \approx 7.8$$

Each proton gains energy due to

Transmembrane electric potential difference

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

Change in pH

$$\delta U = (-2.3k_BT/e)\Delta pH \approx -50mV$$

Total protonmotive force

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

Need 1200 protons per one body revolution

Input power

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

Power loss due to Stokes drag

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

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

Motor efficiency

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

pH value of solutions

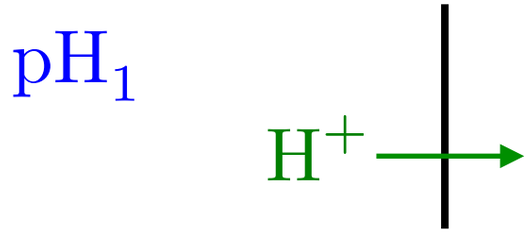
$$\frac{[\mathrm{H}^+][\mathrm{OH}^-]}{c_0^2} = \frac{[\mathrm{H_2O}]K_{\mathrm{eq}}(T,p)}{c_0^2} \approx 10^{-14}$$

$$c_0 = 1\mathrm{M}$$

at room temperature

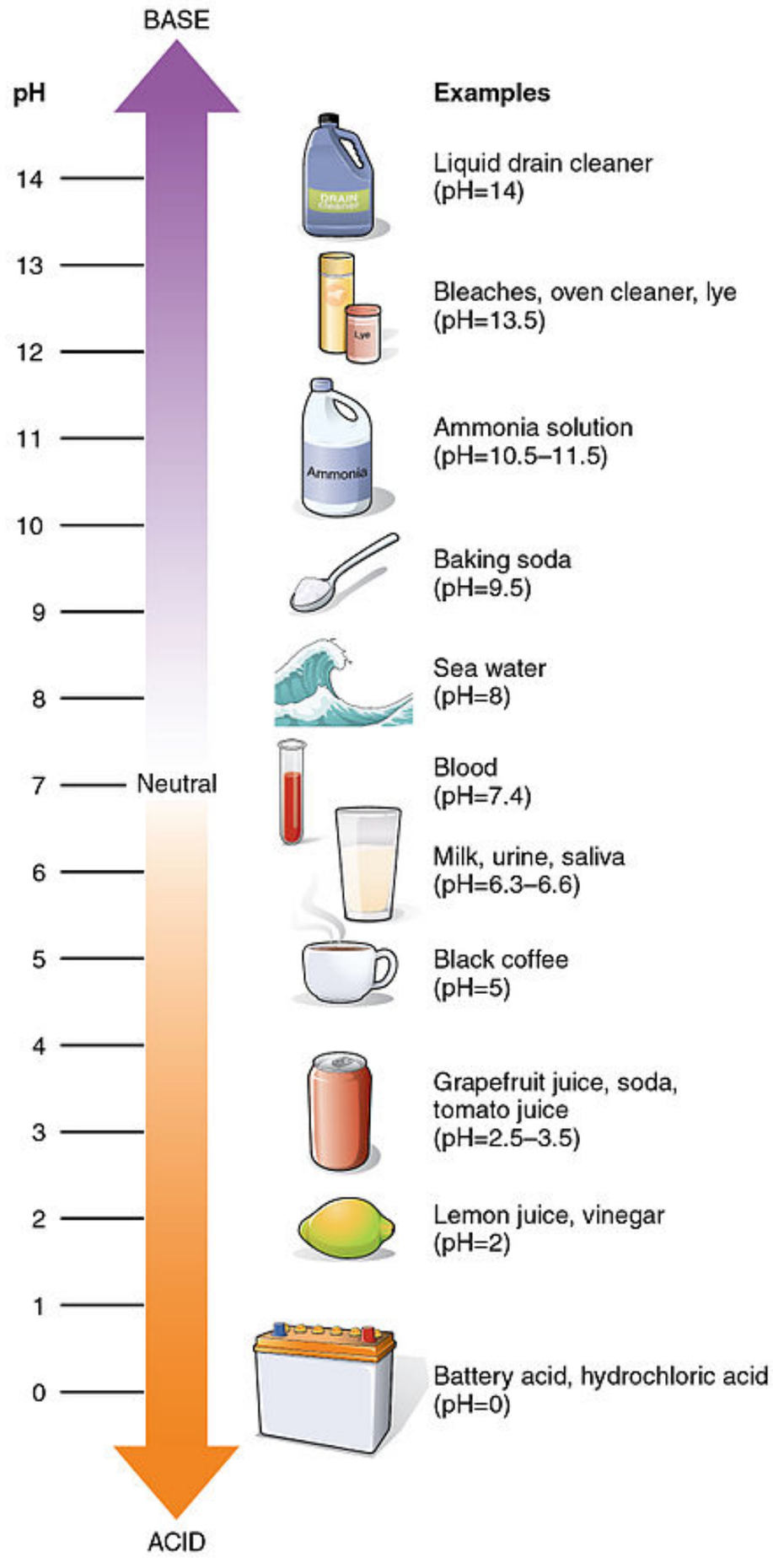
$$\mathrm{pH} = -\log_{10}\left([\mathrm{H}^+]/c_0\right)$$
$$\mathrm{pOH} = -\log_{10}\left([\mathrm{OH}^-]/c_0\right) \approx 14 - \mathrm{pH}$$

How much free energy is changed when H+ goes to environment with different pH?

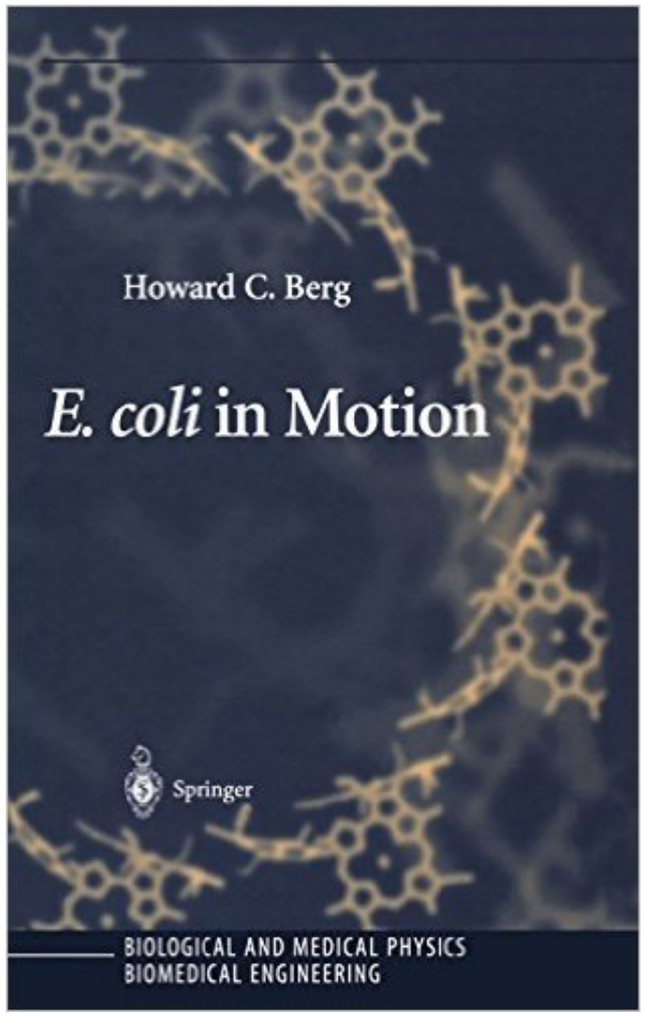


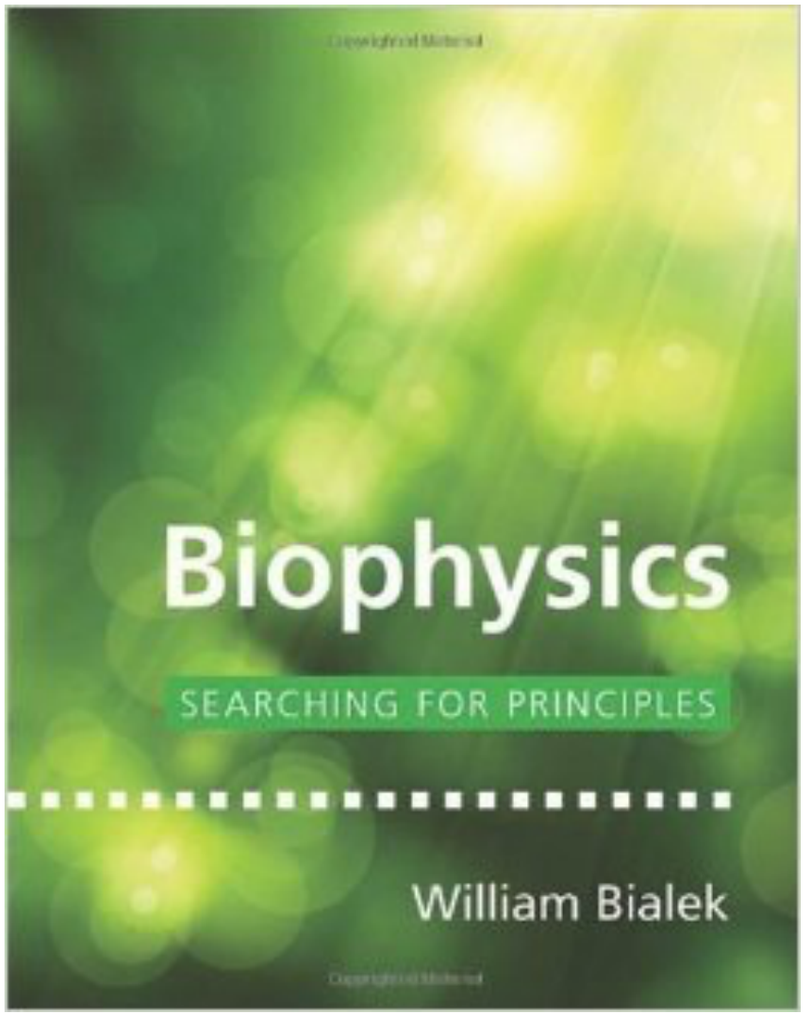
$$\mu_2 - \mu_1 = k_B T \ln\left([\mathrm{H}^+]_2/[\mathrm{H}^+]_1\right)$$
$$E = \frac{\mu_2 - \mu_1}{e_0} \approx -\frac{2.3026\, k_B T}{e_0}\left(\mathrm{pH}_2 - \mathrm{pH}_1\right)$$

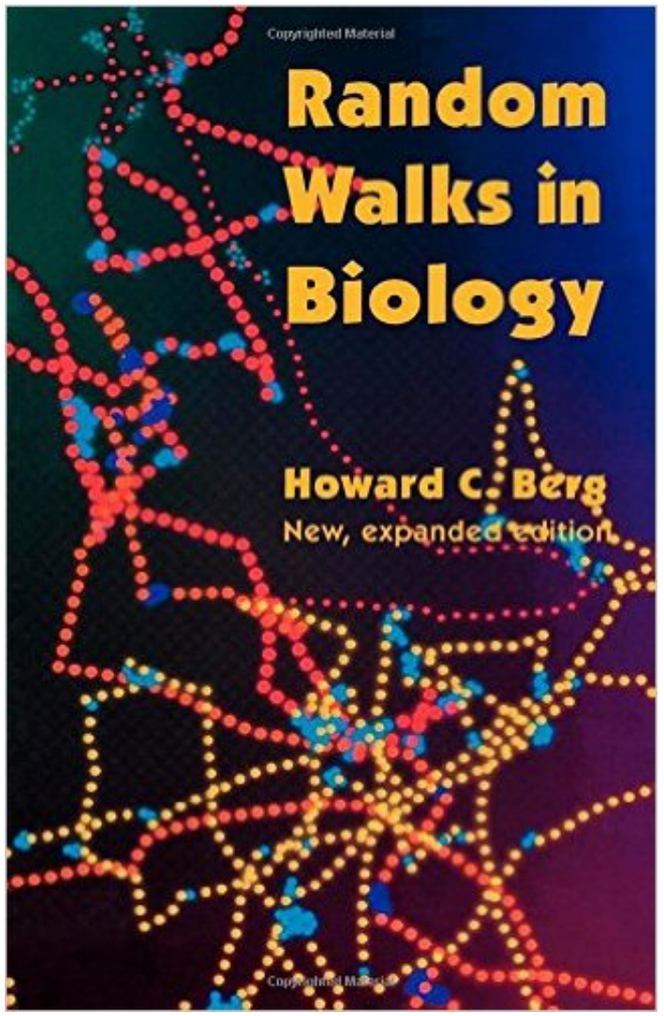
Nernst electric potential *E*



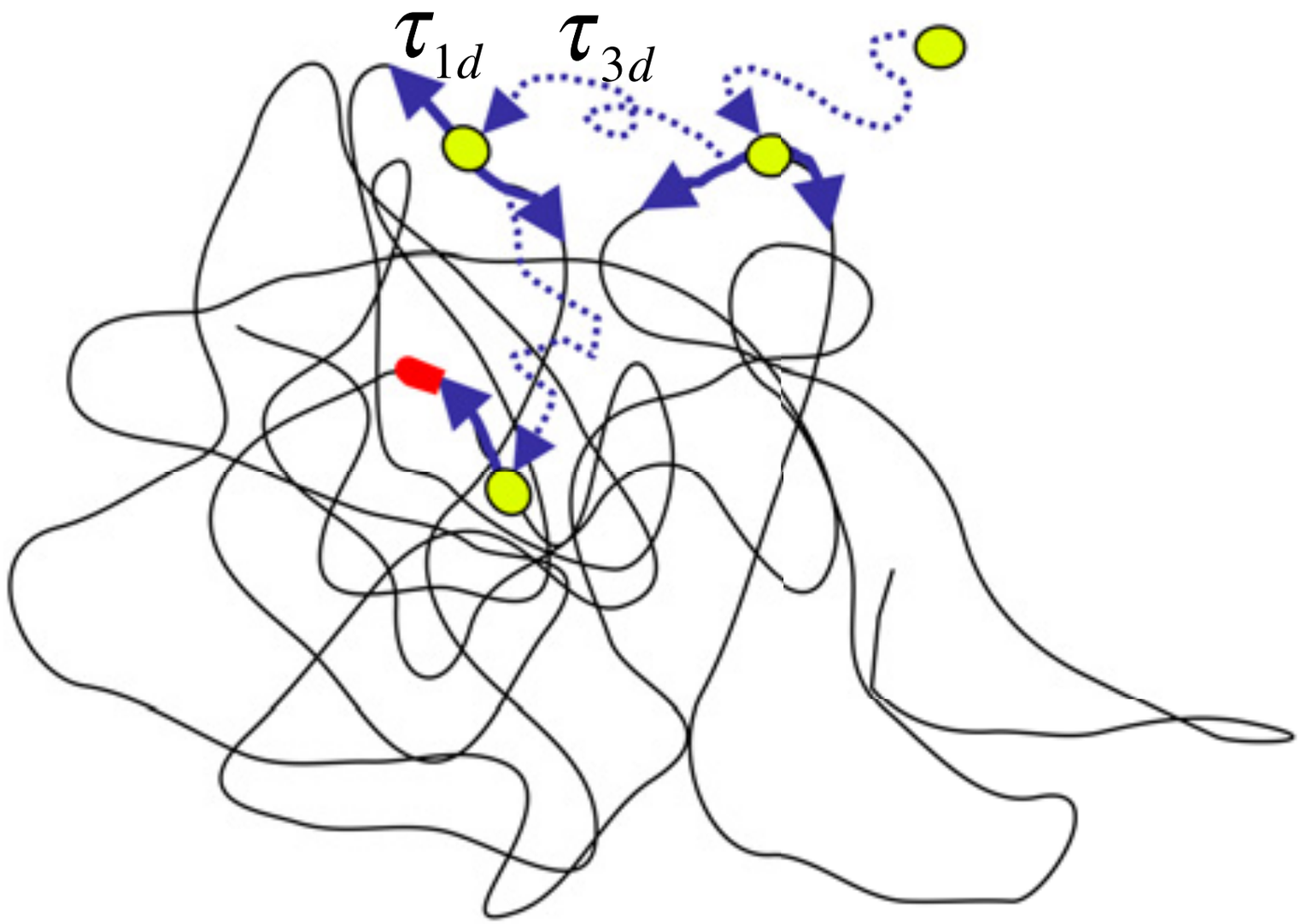
Further reading





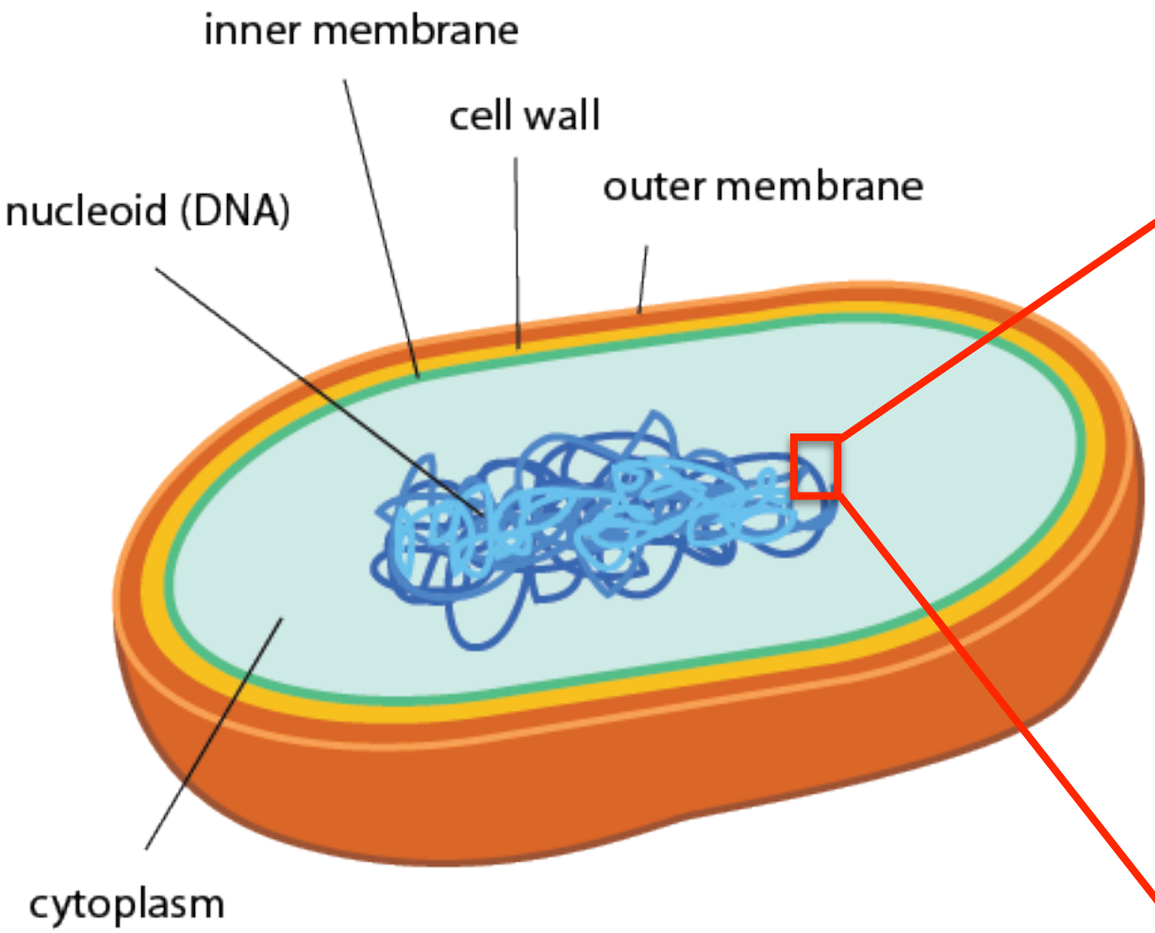


How proteins find target sites on DNA?

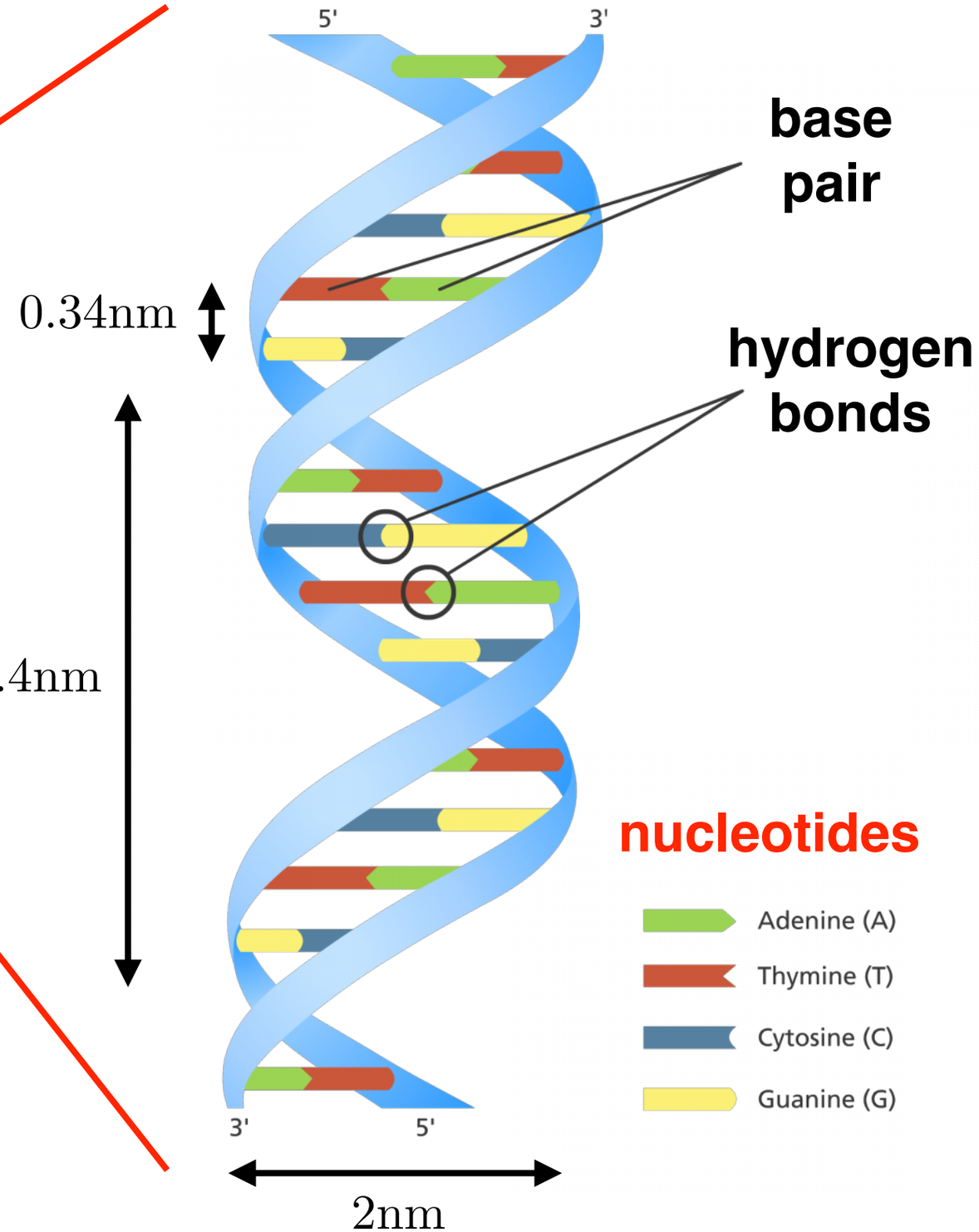


DNA

prokaryotic cell (bacteria, archaea)

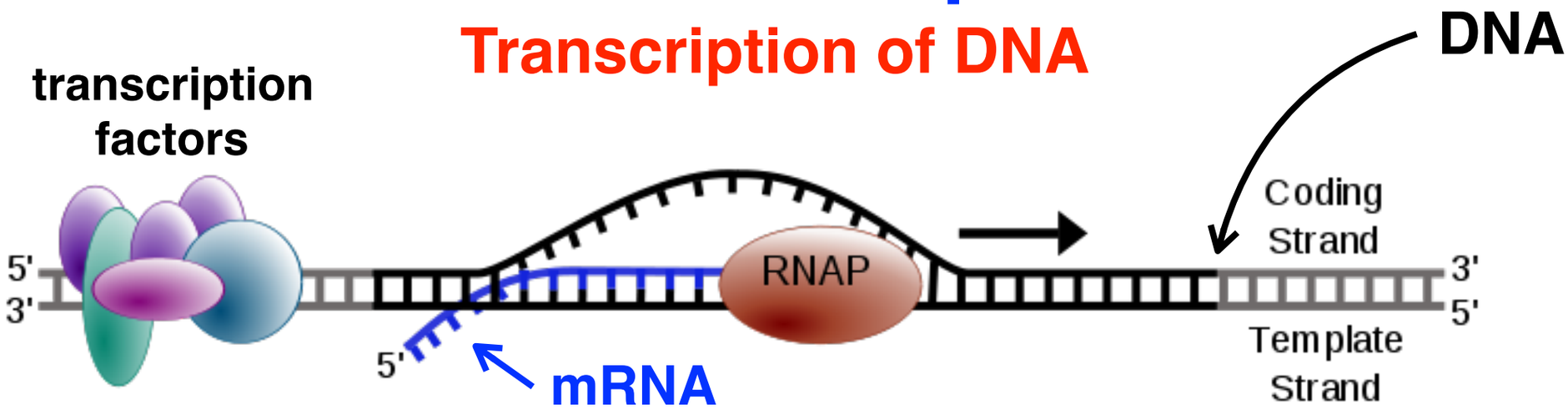


DNA stores genetic information encoded with sequence of bases



Production of new proteins

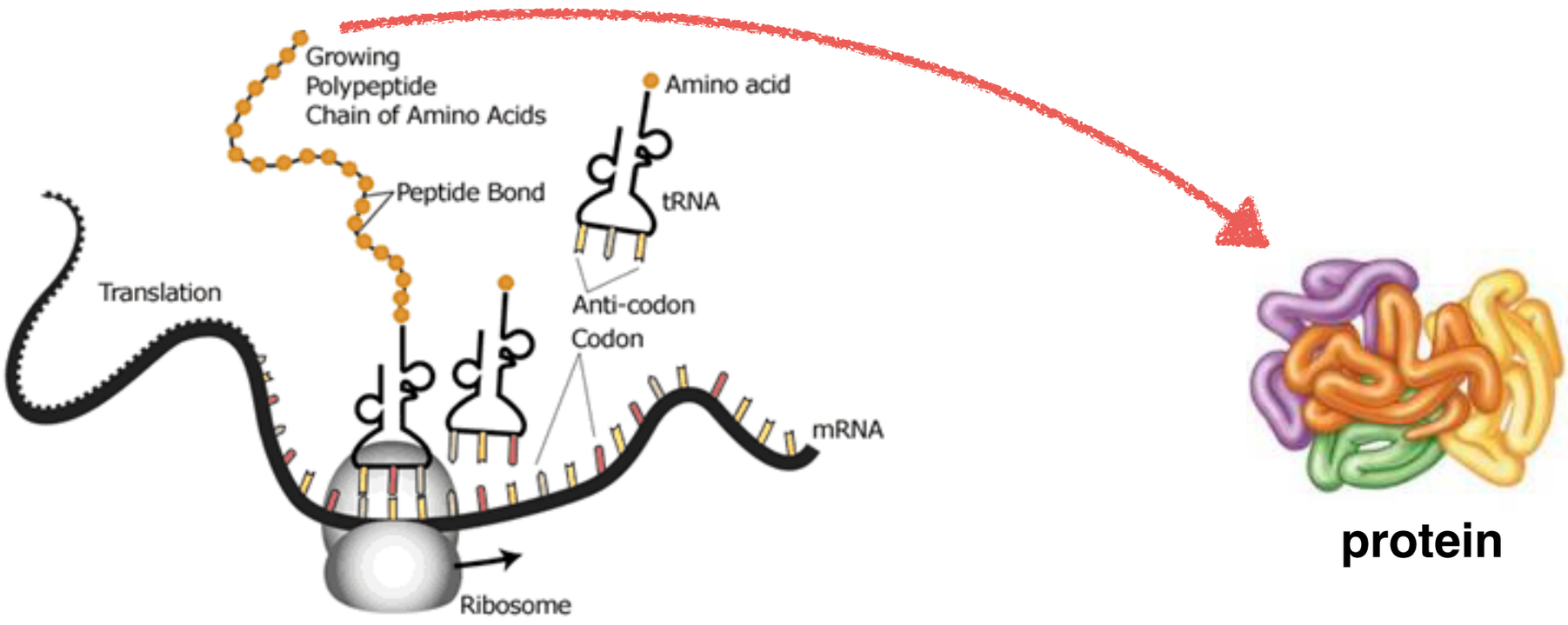
Transcription of DNA



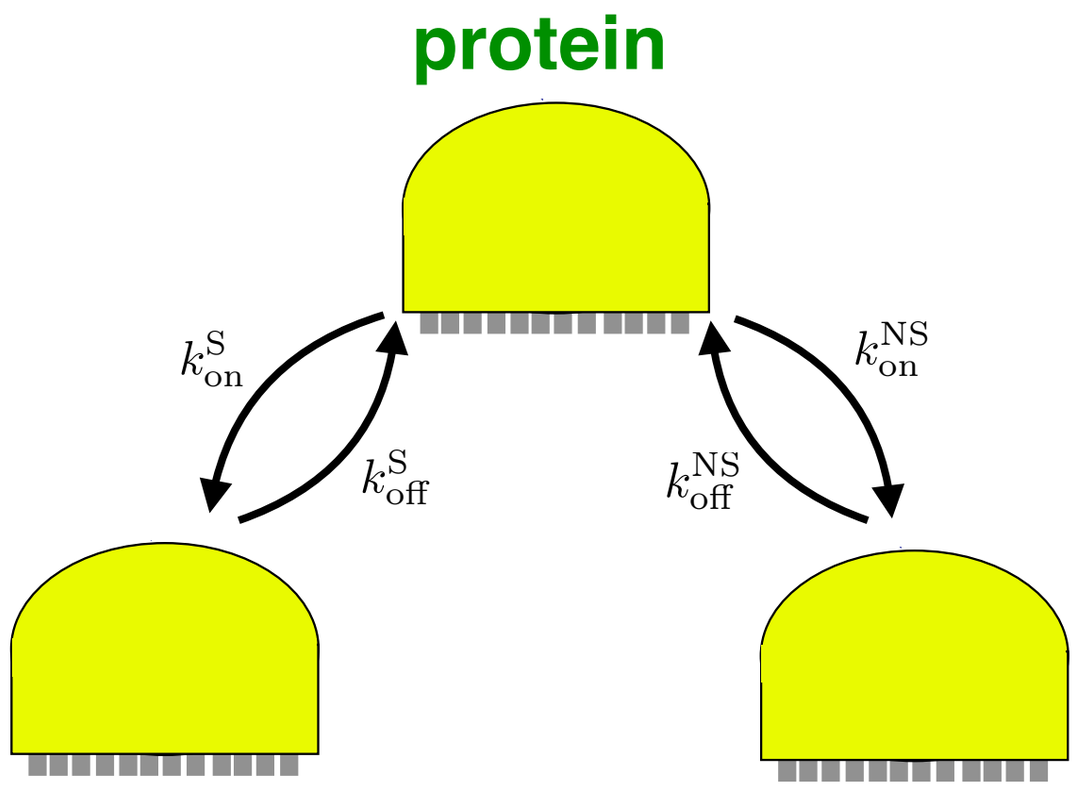
Transcription factors are proteins, which bind to specific locations on DNA, and they help recruiting RNA polymerase (RNAP) that makes a messenger RNA (mRNA) copy of certain DNA segment.

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

Translation of mRNA



Protein-DNA interactions



Binding to specific target sequence is strong

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

Binding to nonspecific sequence is weak

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

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

…ATTATGCATGACGATGTGGACAAACACCTGCGT…

target sequence

DNA sequence

$\Leftrightarrow$

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

---

on rates are diffusion limited

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

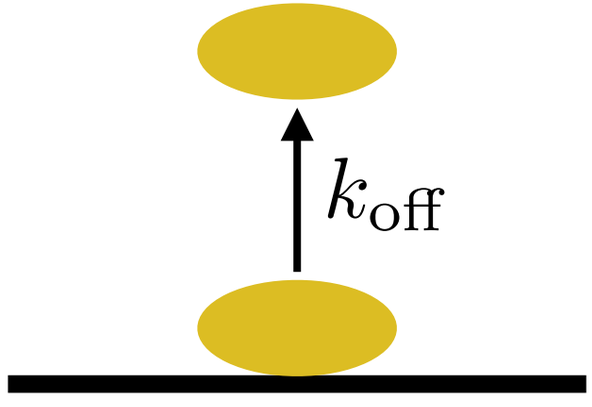
(see slide 6)

off rates depend on binding strengths

$$k_{\rm off}^{\rm S} = A_{\rm s} e^{-\Delta G^{\rm S}/k_B T} \ll k_{\rm off}^{\rm NS} = A_{\rm s} e^{-\Delta G^{\rm NS}/k_B T}$$

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

How long proteins remain bound on DNA?



Probability that protein unbinds in a small time interval Δt :

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

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

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

limit $\Delta t \to 0$

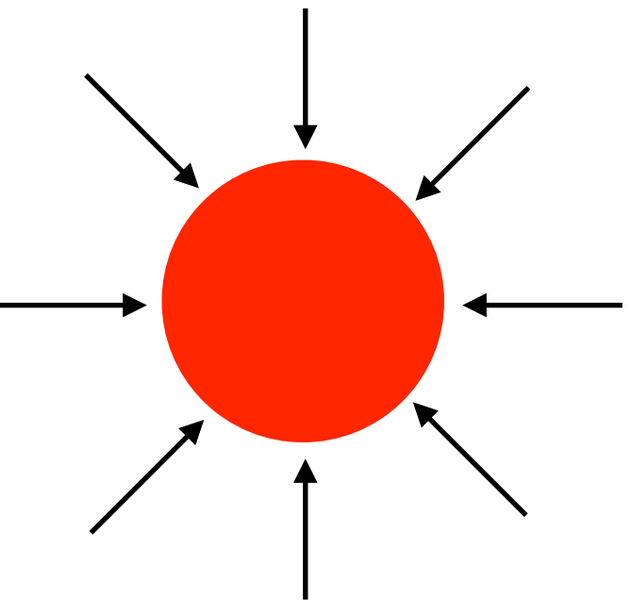
$$\boxed{p(t) = k_{\rm off} e^{-k_{\rm off} t}}$$

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

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

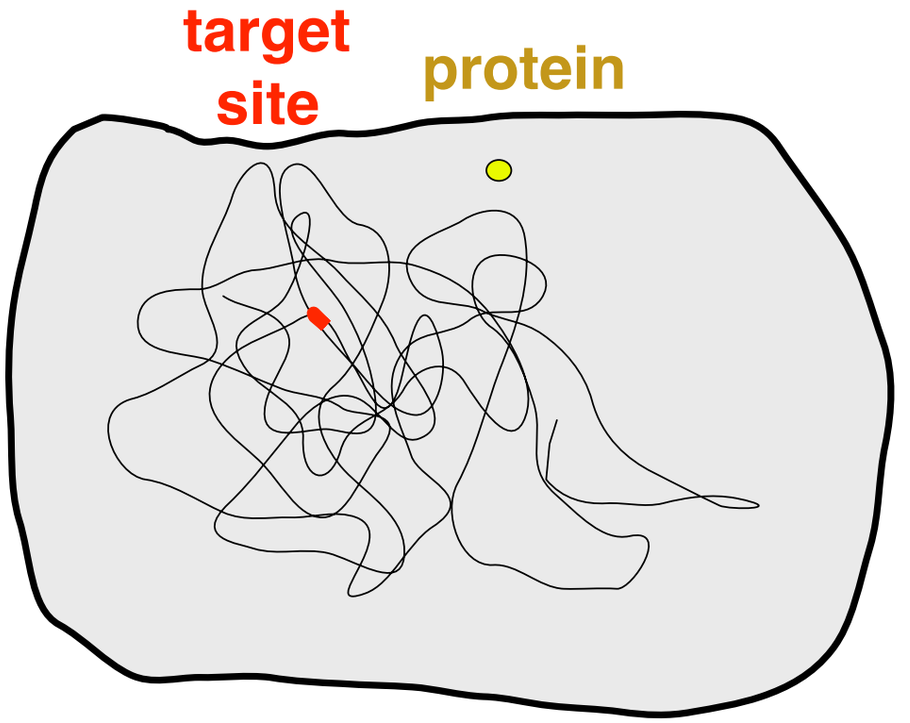


Approximate target site as absorbing sphere of radius

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

concentration of free proteins far from the target site

$$c(r \to \infty) = [P]$$



rate of absorption (see slide 6)

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

Kinetics of protein binding/unbinding

$$[\mathrm{P}] + [\mathrm{T}] \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} [\mathrm{P\text{-}T}]$$

$$\frac{d[\mathrm{P\text{-}T}]}{dt} = k_{\mathrm{on}}[\mathrm{P}][\mathrm{T}] - k_{\mathrm{off}}[\mathrm{P\text{-}T}]$$

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

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

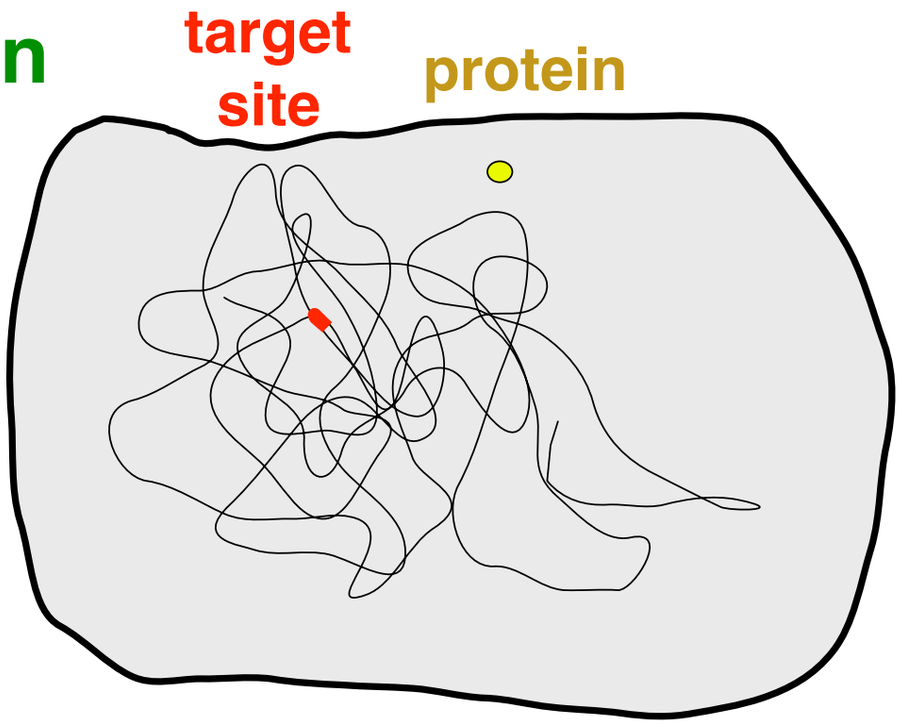
$$\frac{d[\mathrm{P\text{-}T}]}{dt} = (k_{\mathrm{on}}[\mathrm{T}])[\mathrm{P}] \equiv \frac{[\mathrm{P}]}{t_s}$$

characteristic search time

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

How quickly proteins find target sites on DNA?

Characteristic search time via 3D diffusion



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

1917 Smoluchowski theory

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

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

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

$$k_{\rm on} \sim 10^8 M^{-1} s^{-1} \qquad t_s \sim 10 s$$

Molar concentration

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

in vitro experiments (1970)

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

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

Why is experimentally observed rate 100 times larger?

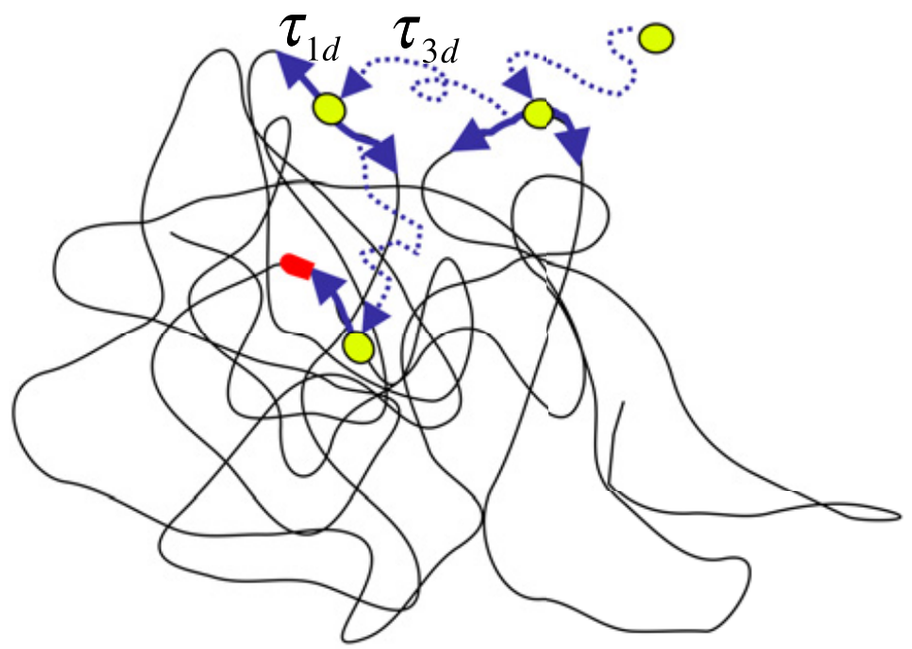
Berg - von Hippel theory (1980s)

(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?

O.G.Berg et al.,
Biochemistry **20**, 6929-48 (1981)

Berg - von Hippel theory (1980s)

First assume fixed sliding time τ_{1d}

Number of distinct sites visited during each sliding event

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

(valid for *n*>>1)

Probability that target site is found during a sliding event

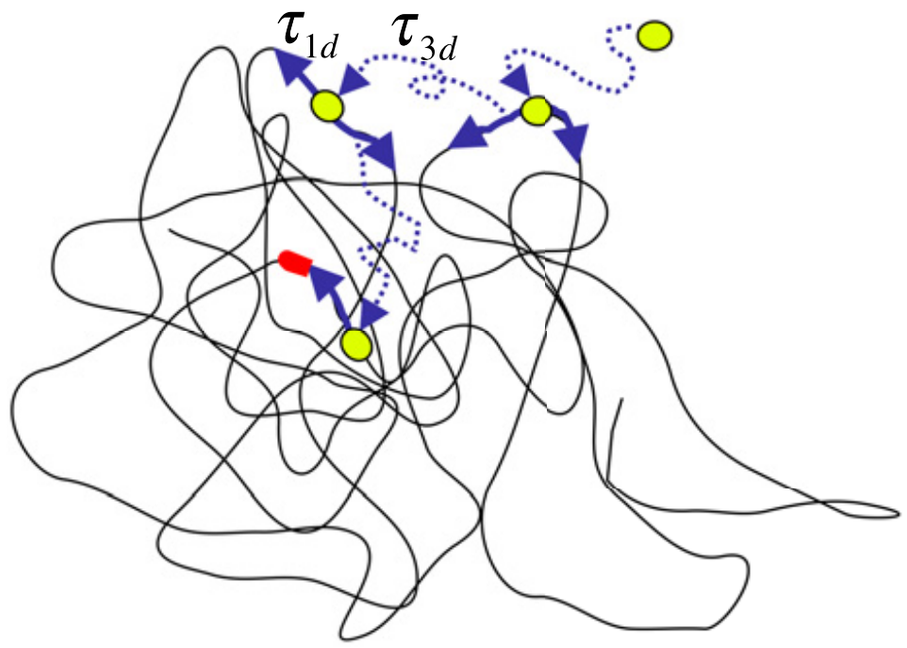
$$q = nb/L$$

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**

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

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

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

Average number of distinct sites visited during each sliding

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

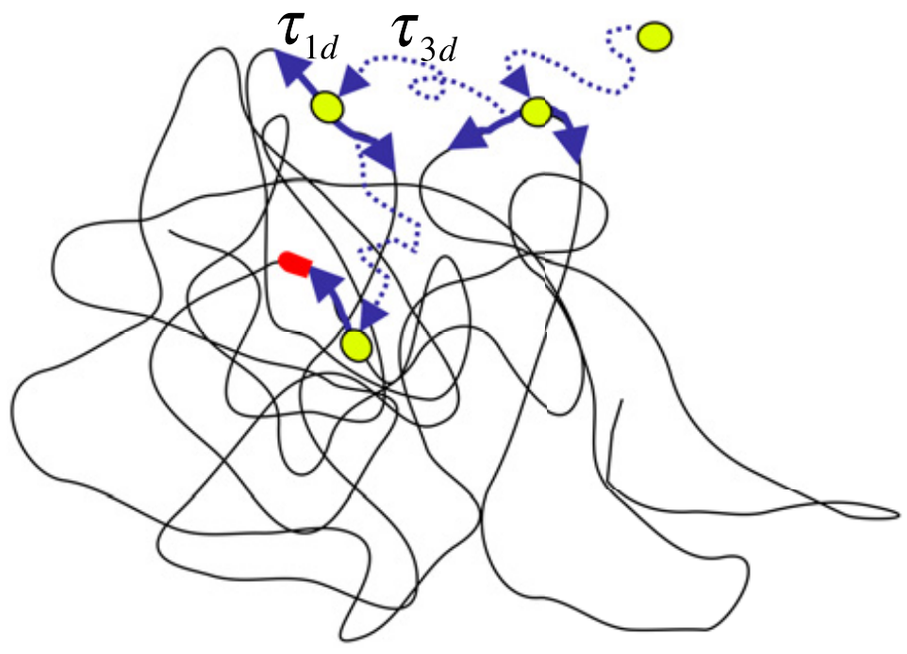
$$\langle n \rangle = 2\sqrt{D_1 \langle \tau_{1d} \rangle / (b^2)}$$

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**

Average search time

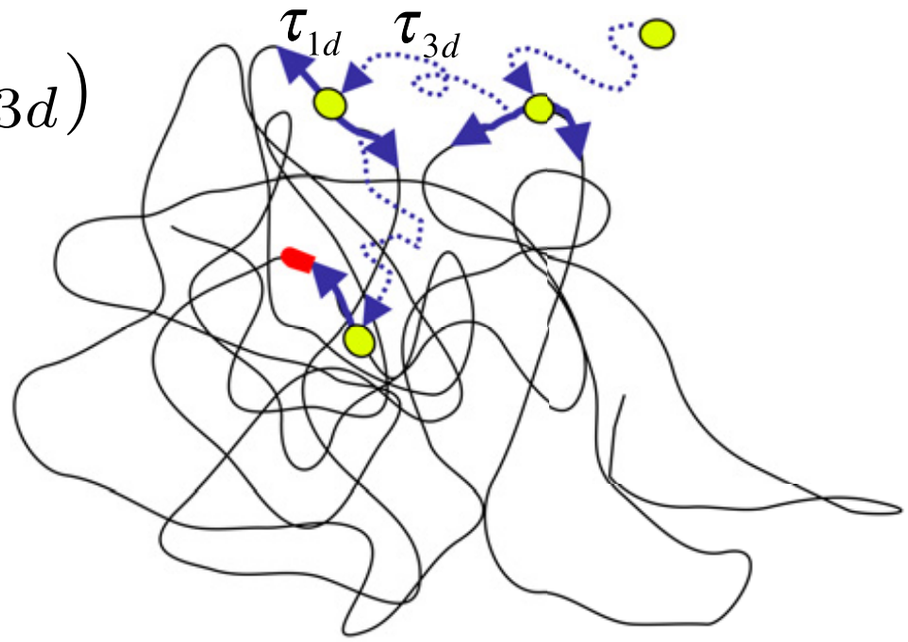
$$\overline{\langle t_s \rangle} = \overline{\langle N_R \rangle} \left(\langle \tau_{1d} \rangle + \tau_{3d} \right)$$

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

Facilitated diffusion

Average search time $\overline{\langle t_s \rangle} = \dfrac{L}{\langle \ell_{\rm sl} \rangle} \left(\langle \tau_{1d} \rangle + \tau_{3d} \right)$

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



Optimal search time

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

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

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

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

D_3 **- diffusion constant in space**

D_1 **- diffusion constant along the DNA**

Search time for jumps alone

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

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

average number of jumps needed to find the target $\overline{N}_{\rm jumps} = \dfrac{L}{b}$

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

Search time for sliding alone

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

Search time speed up for facilitated diffusion

$$\frac{\overline{t}_{s,\rm jumps}}{\overline{\langle t_s \rangle}} = \frac{\langle \ell_{\rm sl} \rangle}{b} \frac{\tau_{3d}}{\left(\langle \tau_{1d} \rangle + \tau_{3d} \right)}$$

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

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

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

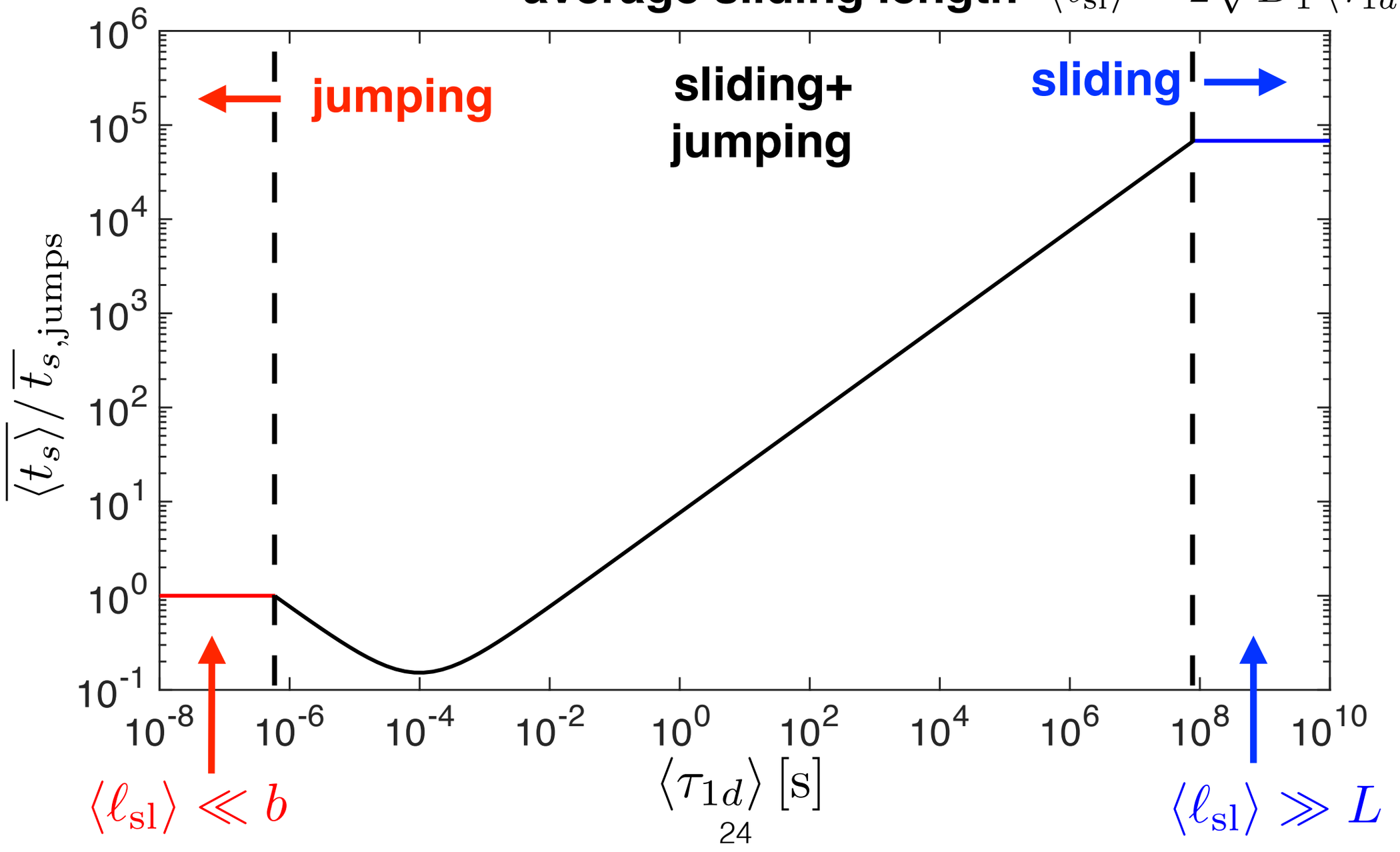
$L = 1\mathrm{mm}$

$b = 0.34\mathrm{nm}$

search time for jumps alone $\overline{t_s}_{,\mathrm{jumps}} = (L/b)\tau_{3d} \approx 300s$

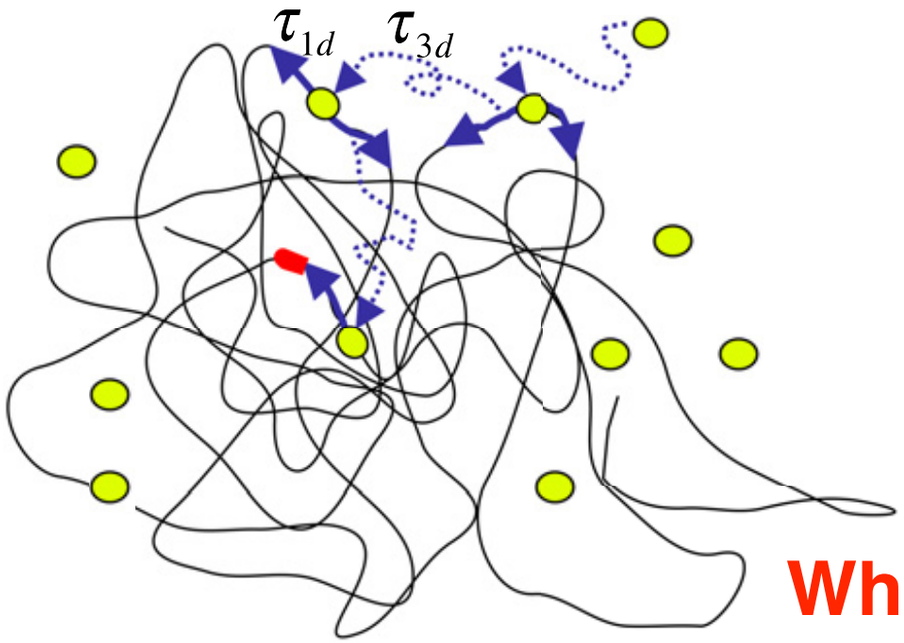
average search time $\overline{\langle t_s \rangle} = \dfrac{L}{\langle \ell_{\mathrm{sl}} \rangle} \left(\langle \tau_{1d} \rangle + \tau_{3d} \right)$

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



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)

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

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*

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