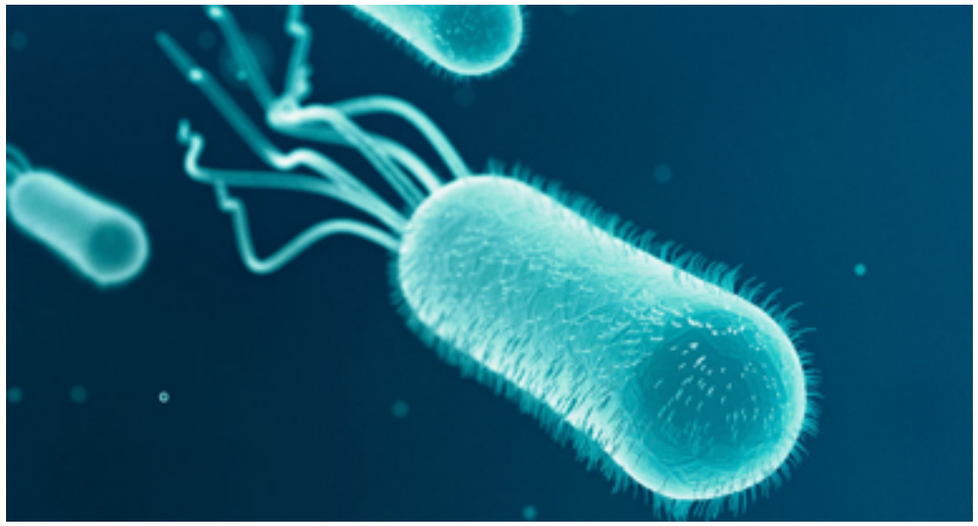
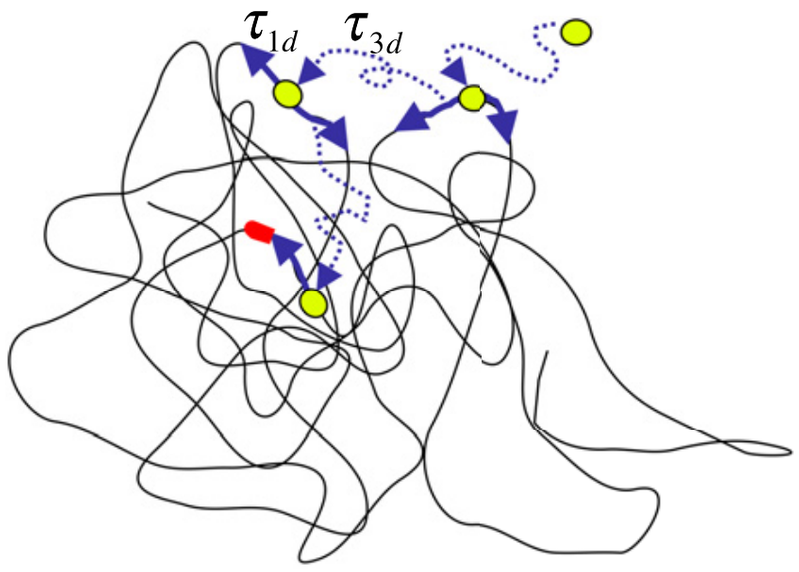
MAE 545: Lecture 3 (9/24)

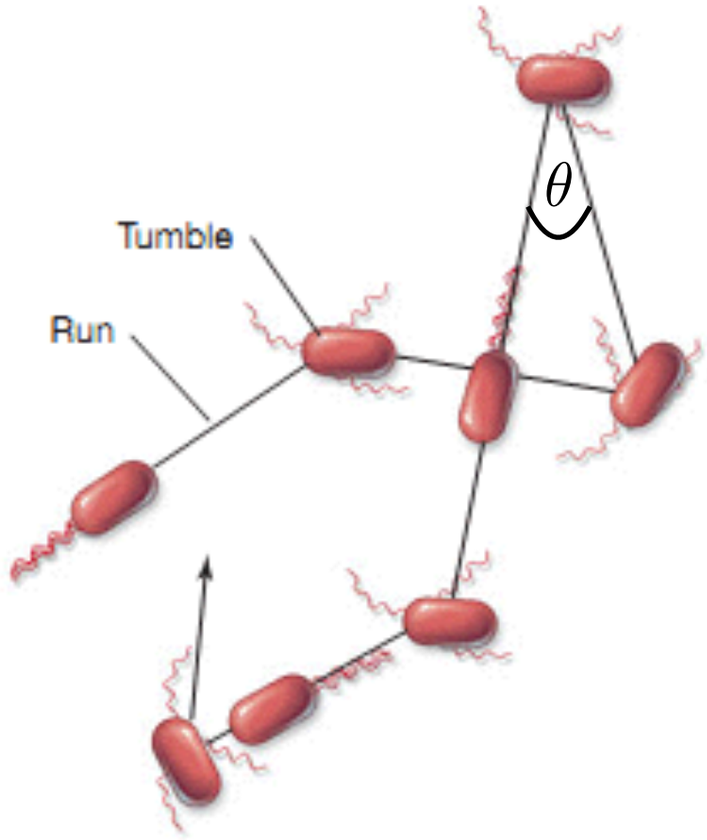
E. coli chemotaxis (continued)

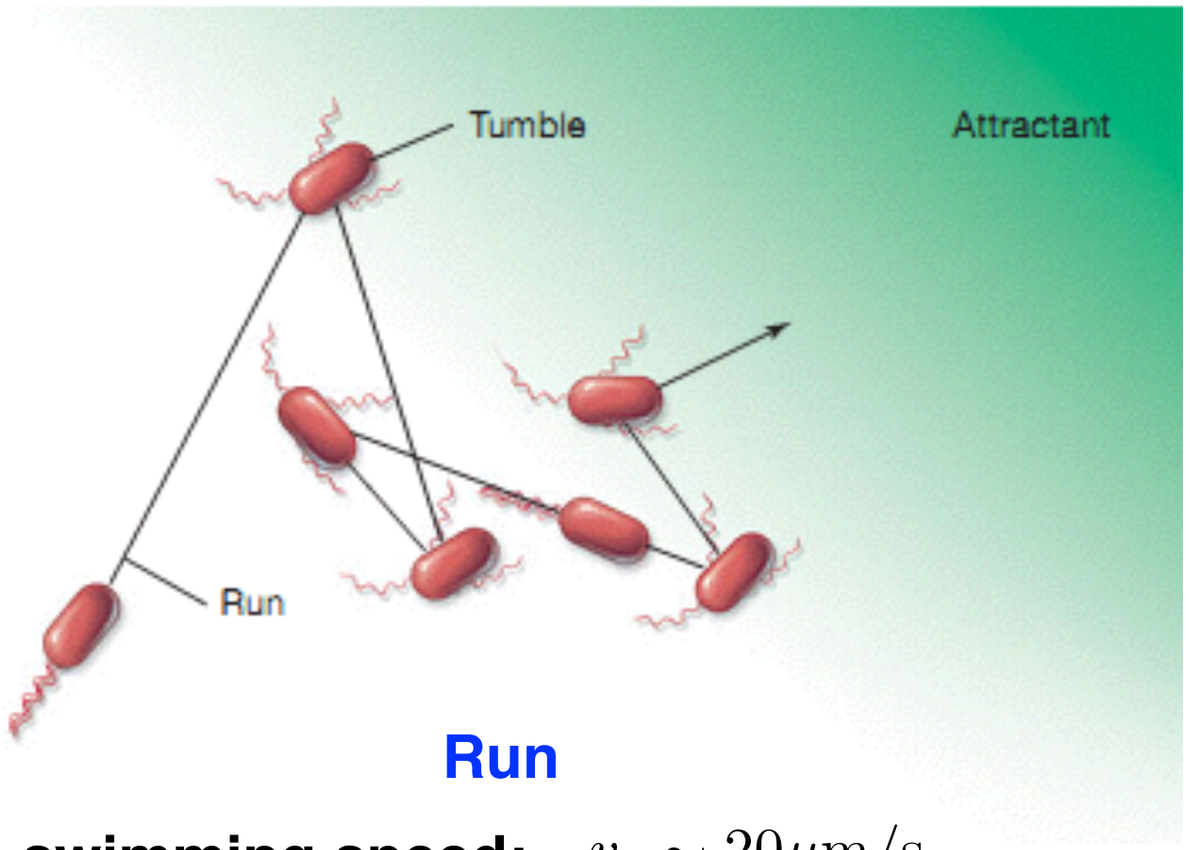


How proteins find target sites on DNA?



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.

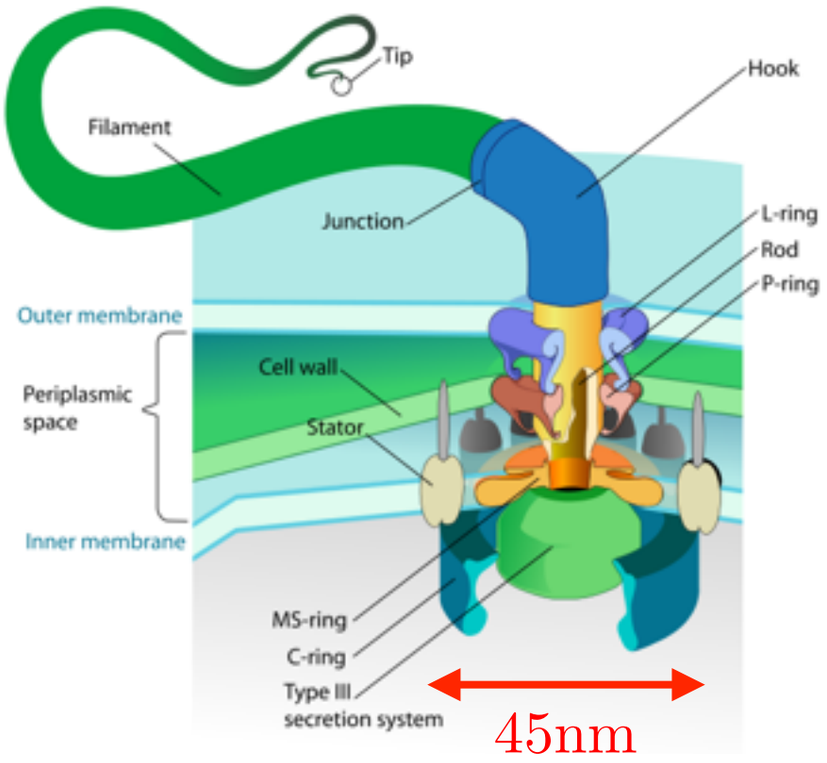
Tumble

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

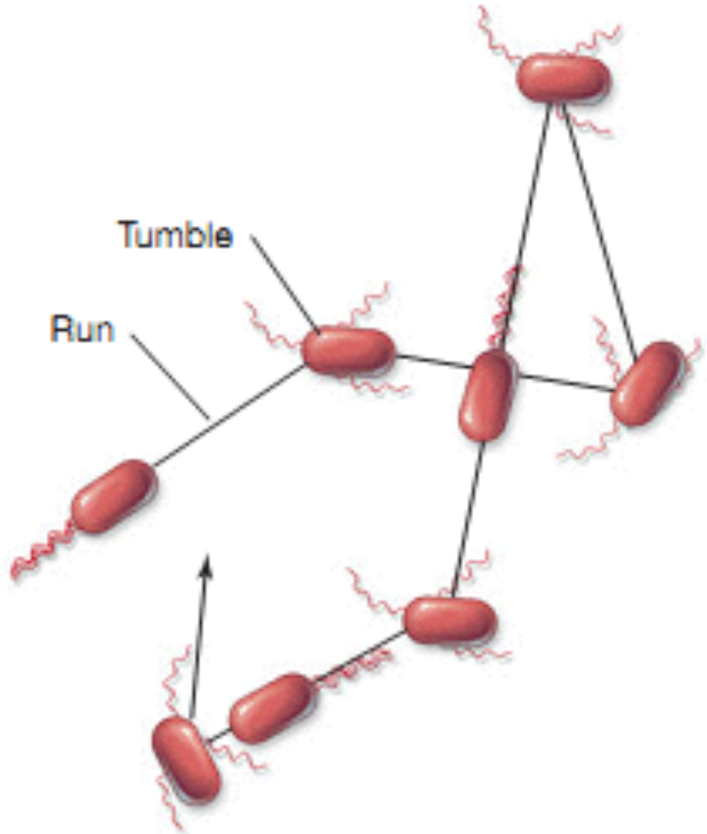
typical duration: $t_t \sim 0.1\mathrm{s}$

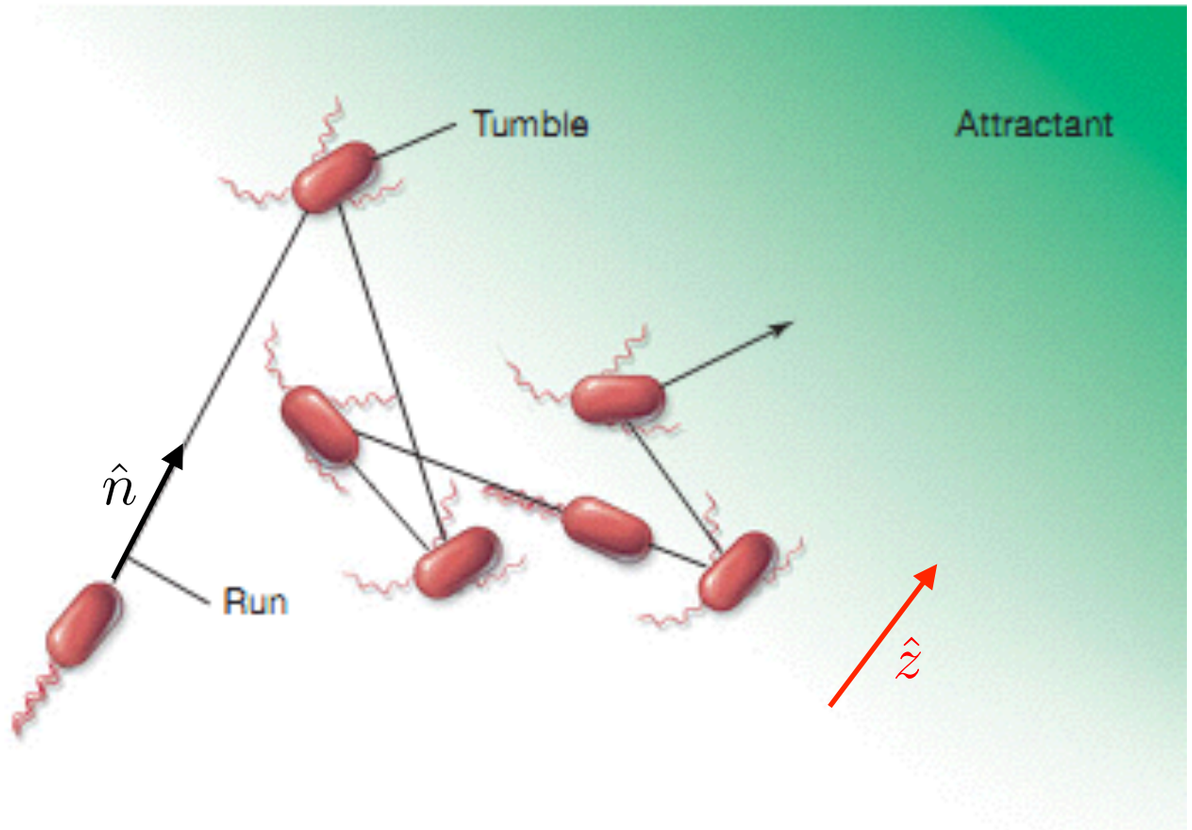
one or more motors turning clockwise

Rotary motor



E. coli chemotaxis





Homogeneous environment

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

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

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

drift velocity

$$v_d = 0$$

effective diffusion

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

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

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

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

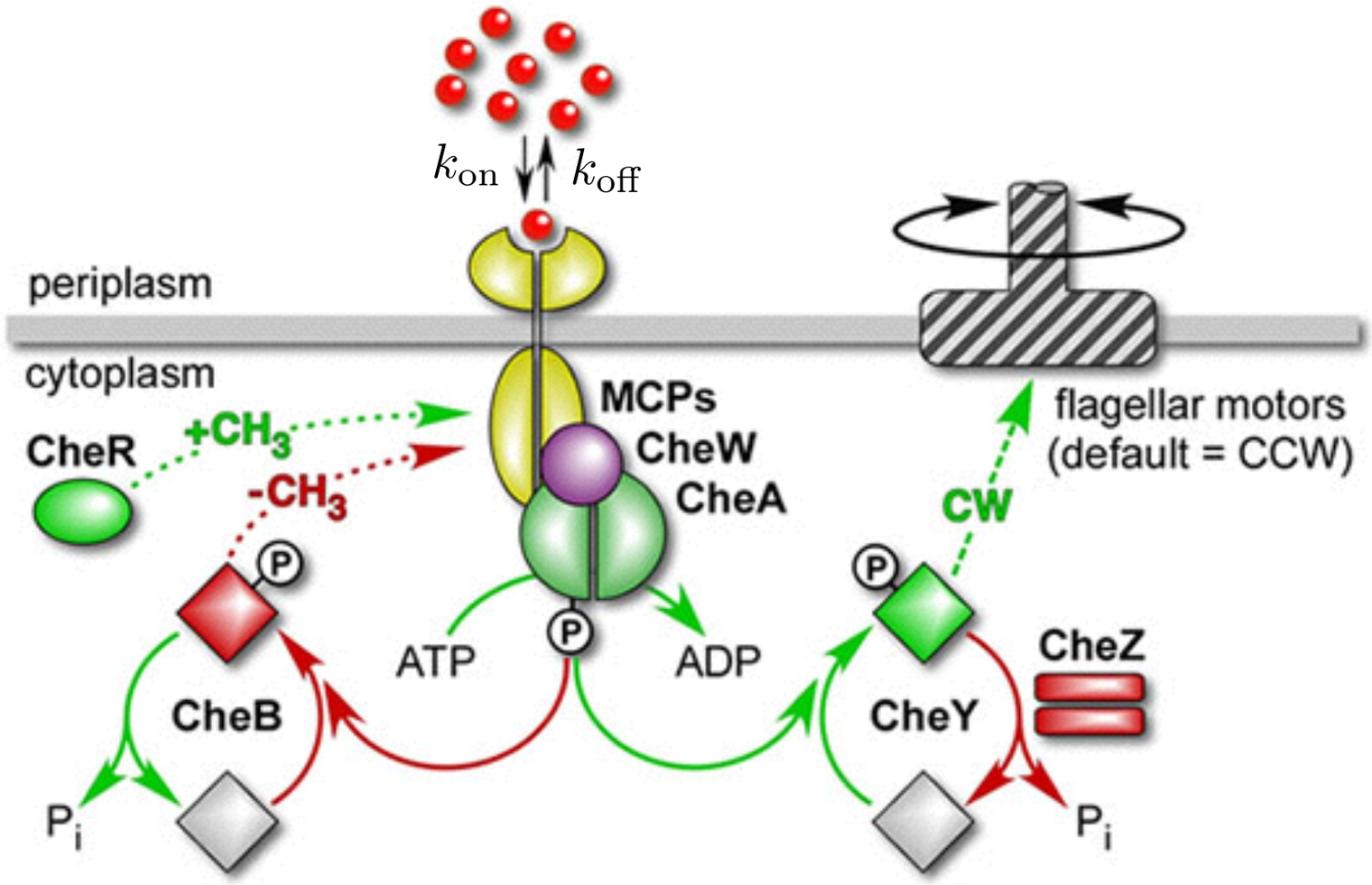
Sensing of environment

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

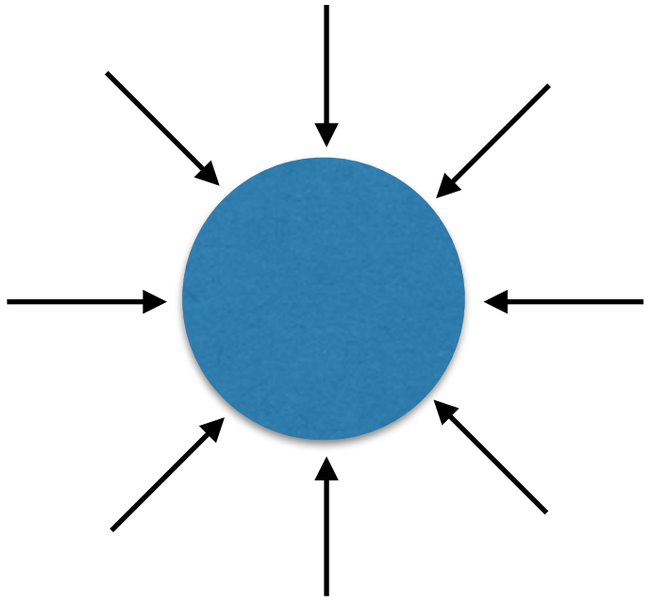
Average fraction of bound receptors p_B is related to concentration c of molecules.

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

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



Diffusion limited flux of molecules to E. coli



absorbing sphere

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

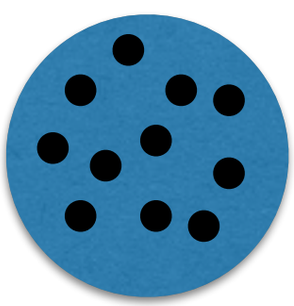
$$J(r) = -D\frac{\partial c(r)}{\partial r} = -\frac{Dc_\infty R}{r^2}$$

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* absorbing disks of radius *s

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

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?

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

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

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

Error in measurement

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

E.coli can be more precise by counting molecules for longer time *t*. However, they need to wait some time *t0* in order for the original molecules to diffuse away to prevent double counting of the same molecules!

$$t_0 \sim R^2/D \sim 10^{-3} s \qquad \overline{N} \sim R^3 ct/t_0 \sim DRct \qquad \text{Err} \sim (DRct)^{-1/2}$$

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

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

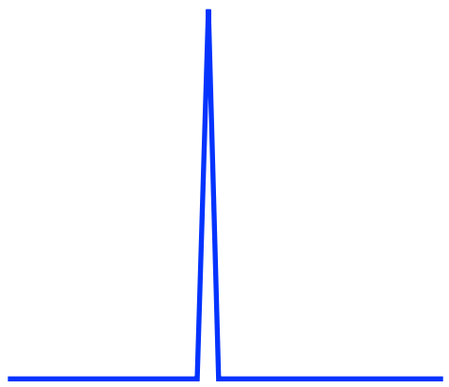
Molar concentration

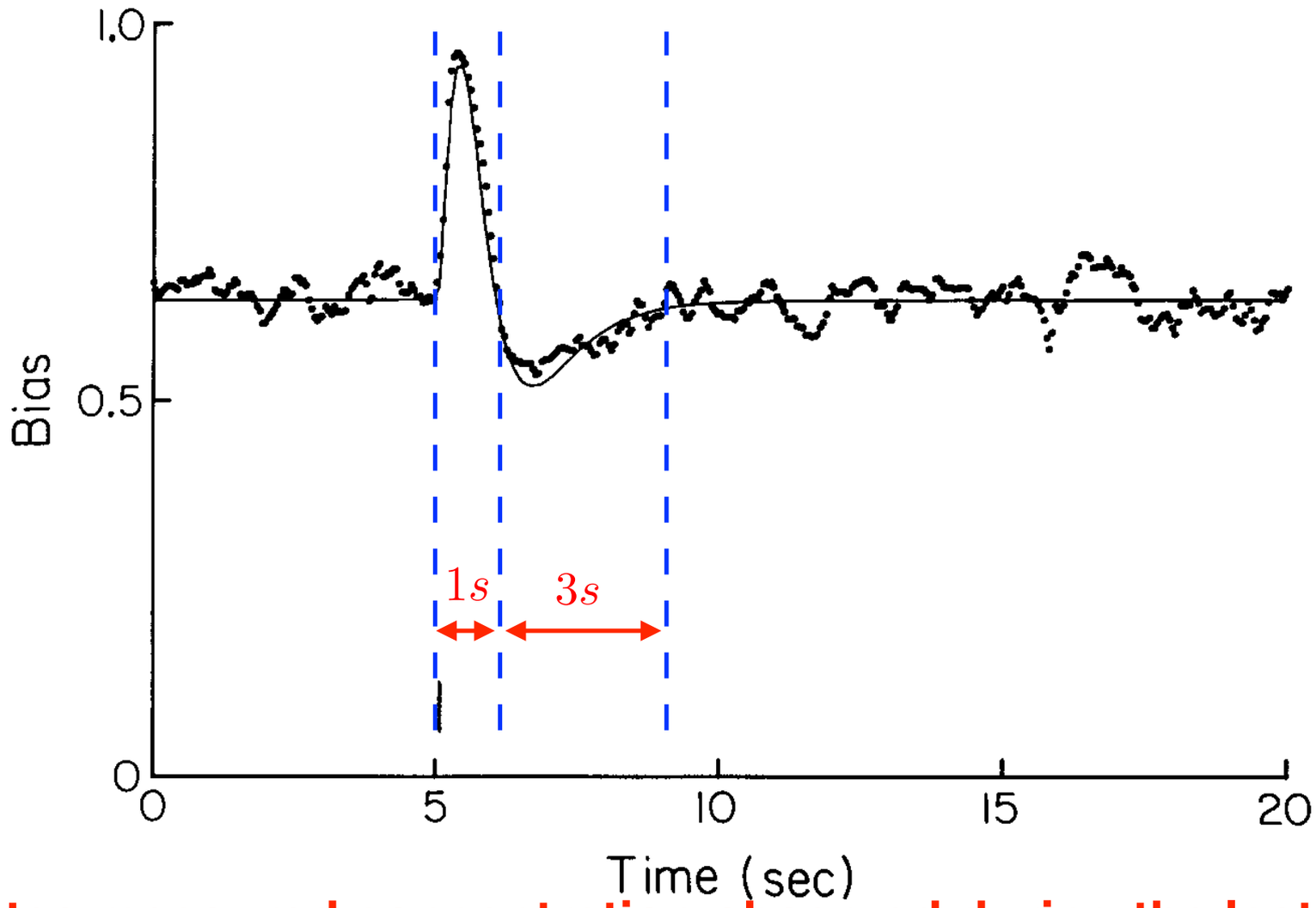
$$1M = 6 \times 10^{26}\text{m}^{-3}$$

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

Input concentration



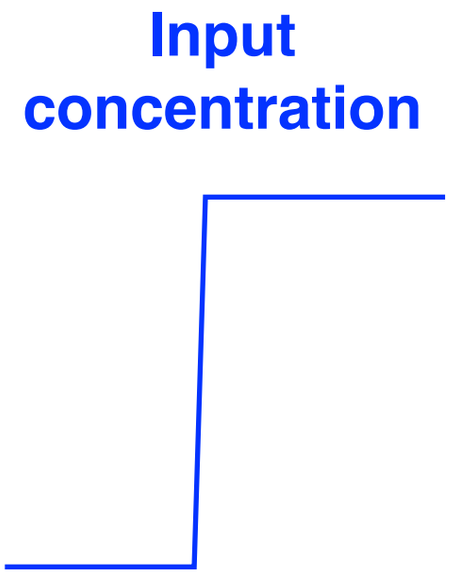


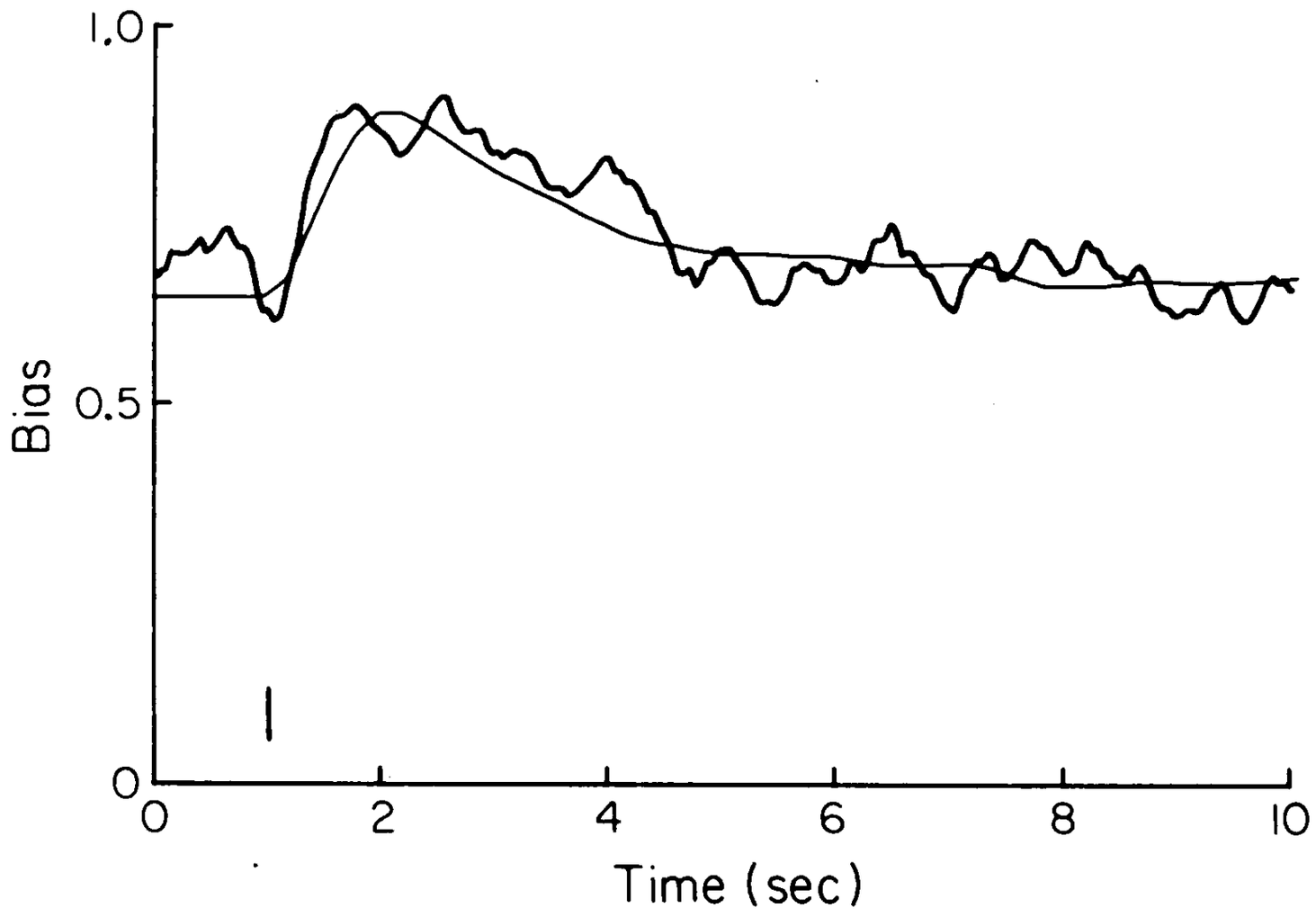
E. coli integrates measured concentration observed during the last second and compare this with measured concentration during the previous 3 seconds. If difference is positive then increase the probability of runs, otherwise increase the probability of tumbles.

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





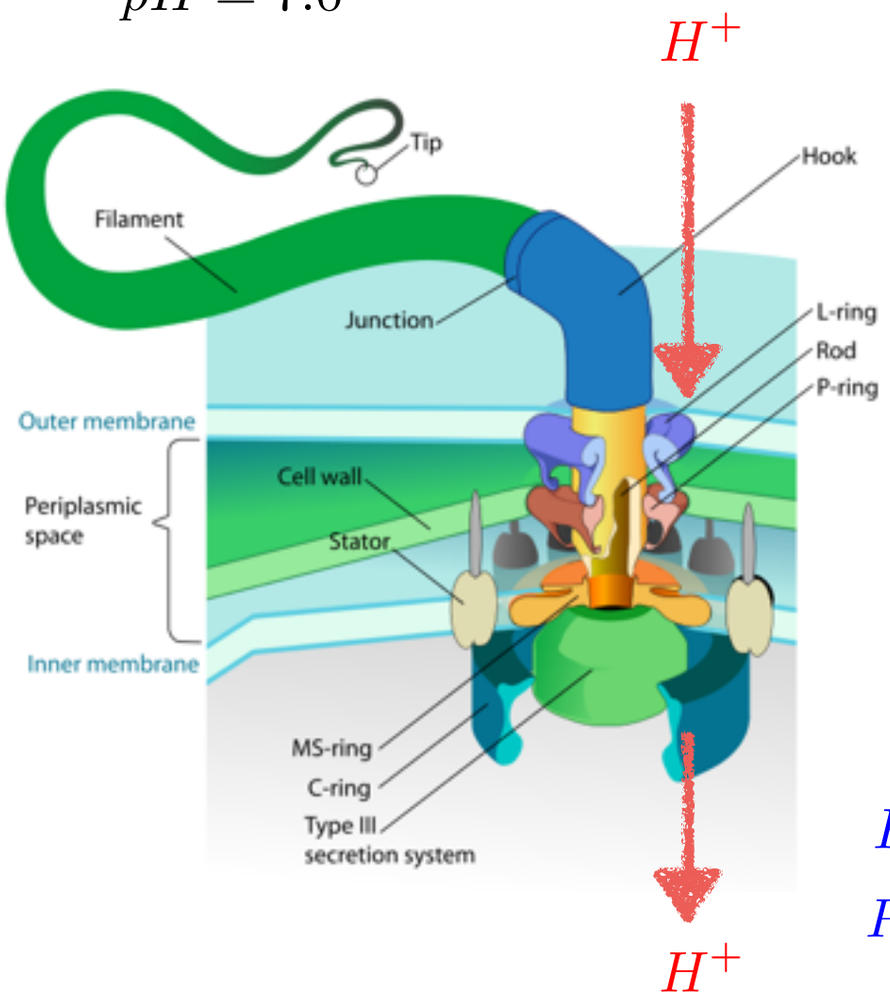
E. coli adapts to the new level of concentration in about 4 seconds. This enables E. coli to be very sensitive to changes in concentration over a very broad range of concentrations!

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

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

Change in pH

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

Total protonmotive force

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

Need 1200 protons per one revolution

Input power

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

Power loss due to stokes drag

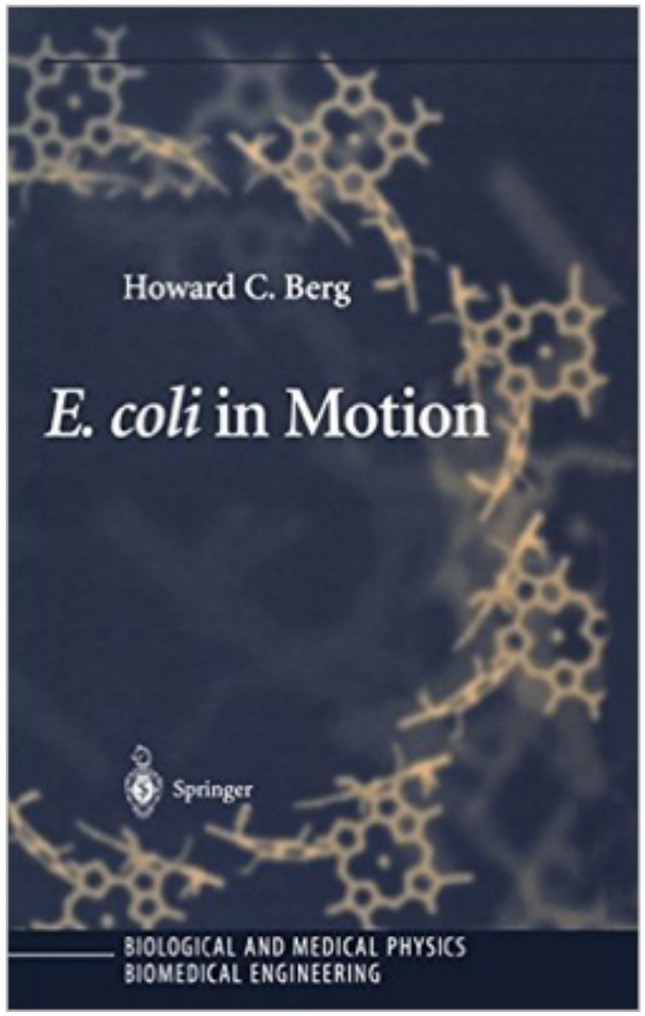
$$P_{\text{rot}} = N \times (2\pi f) \approx 4600\text{pN nm} \times (20\pi\text{Hz}) \approx 2.9 \times 10^5\text{pN nm/s}$$

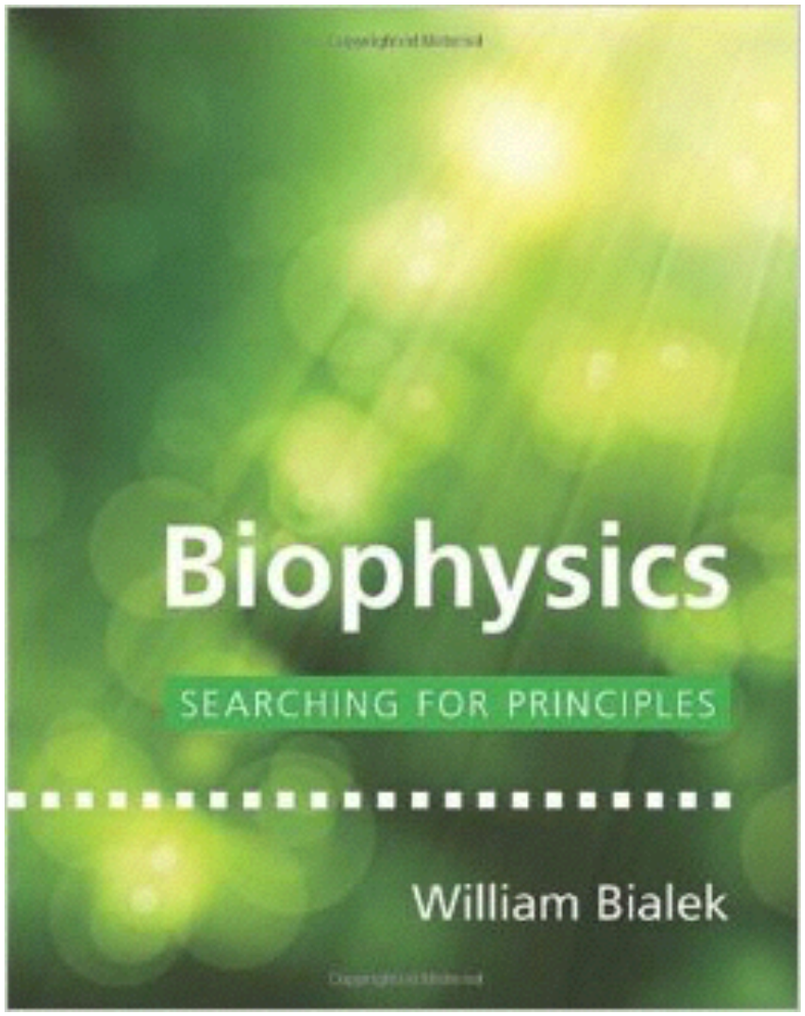
$$P_{\text{trans}} = F \times v \approx 0.4\text{pN} \times 20000\text{nm/s} \approx 8 \times 10^3\text{pN nm/s}$$

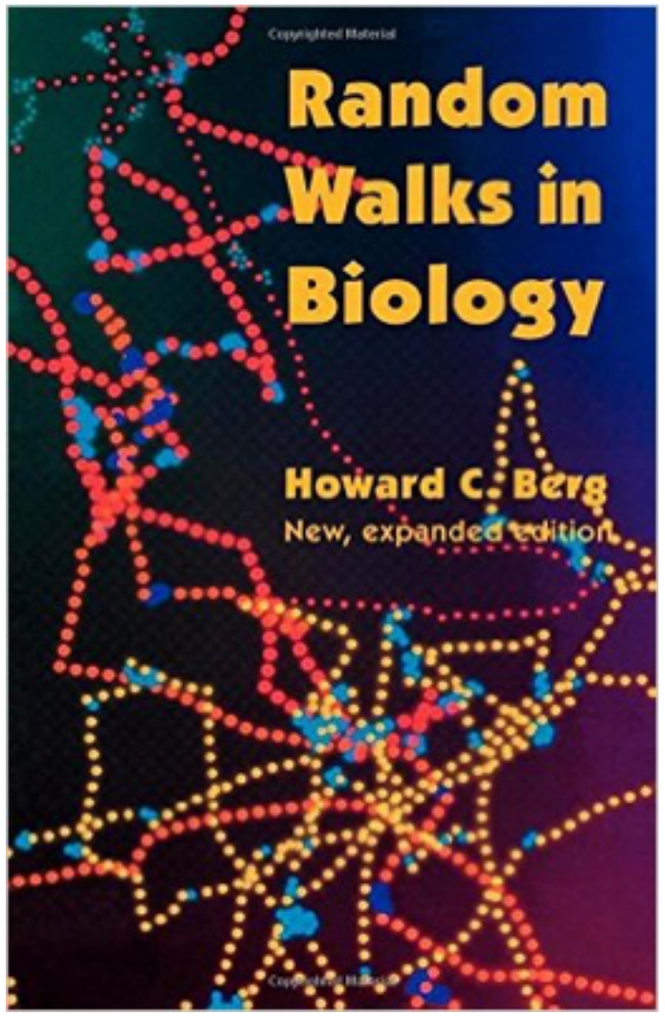
Motor efficiency

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

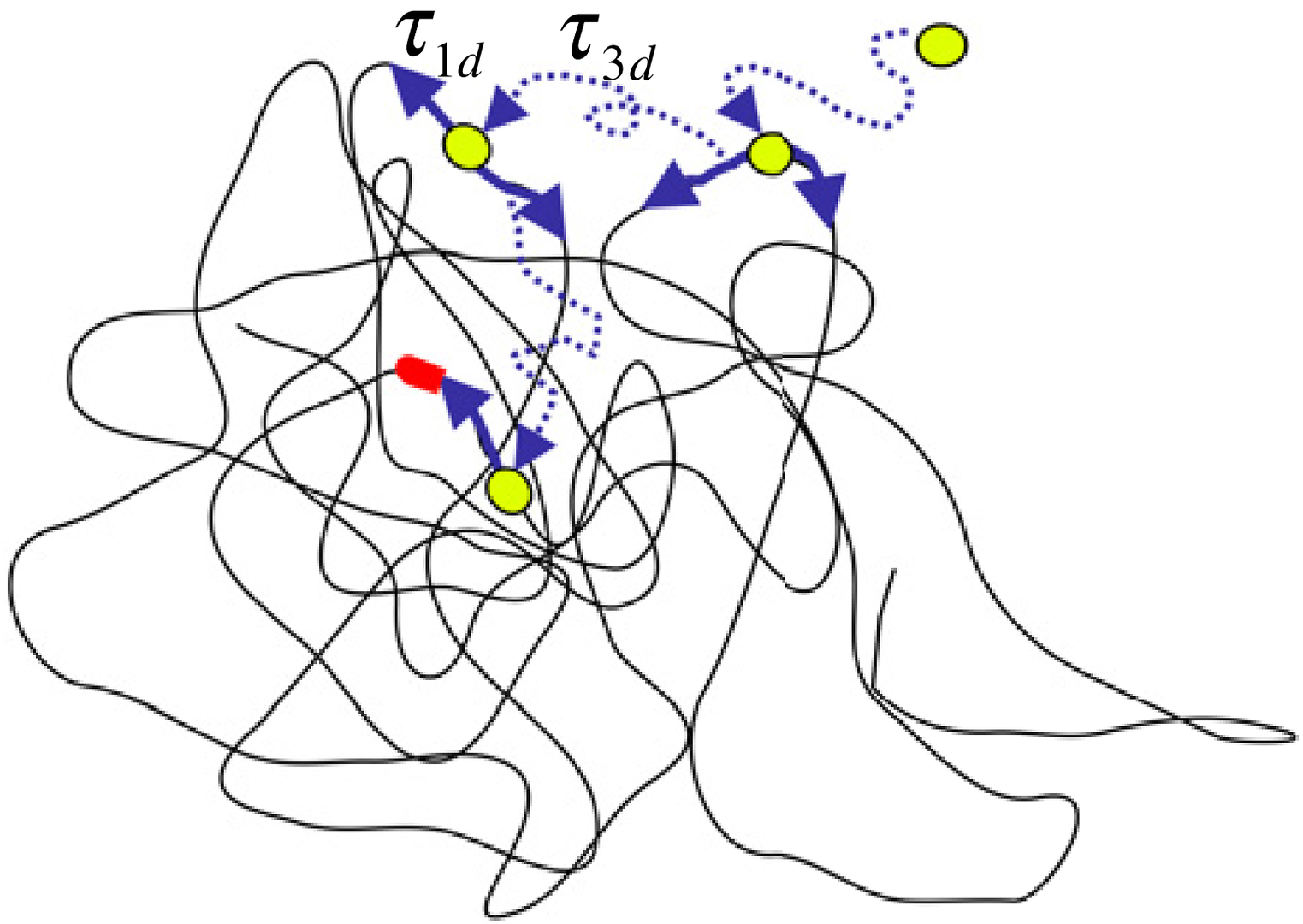
Further reading





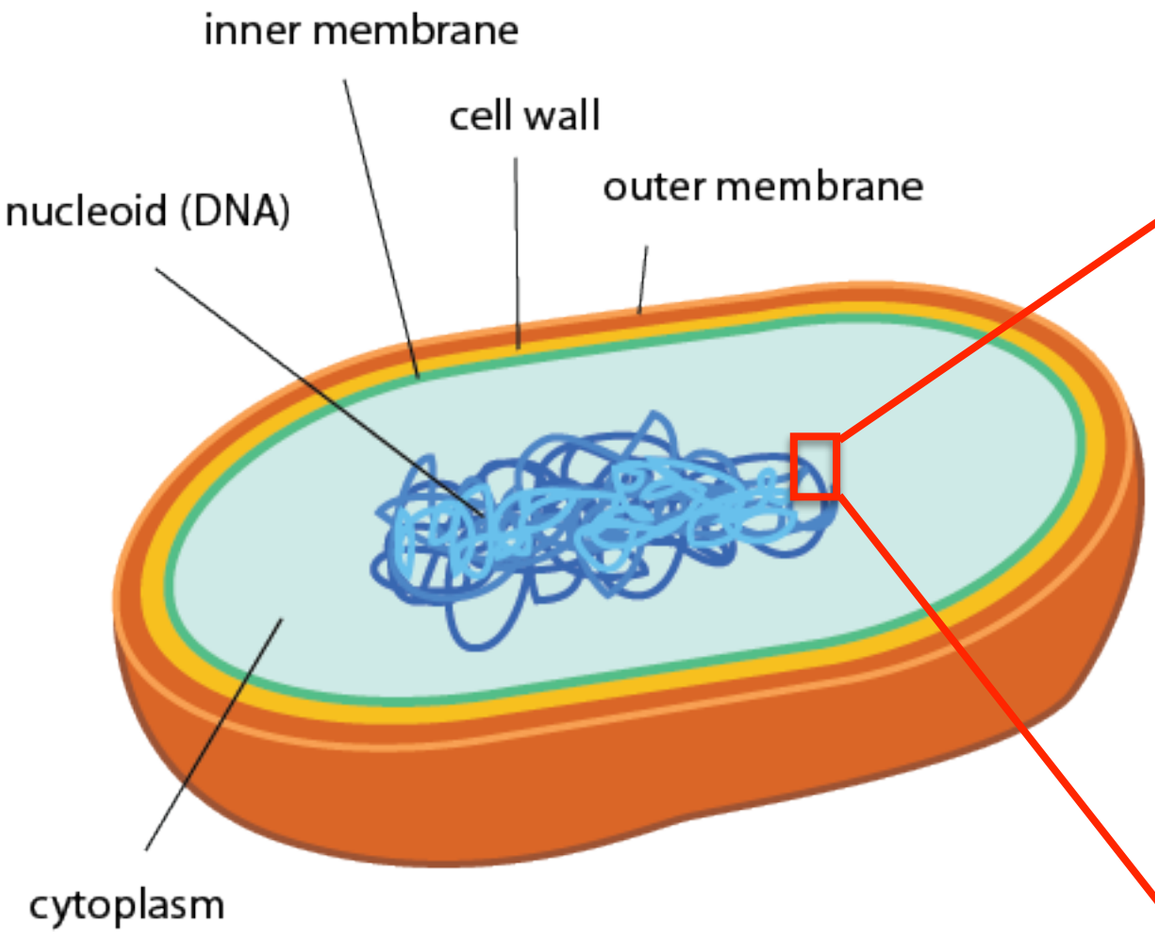


How proteins find target sites on DNA?

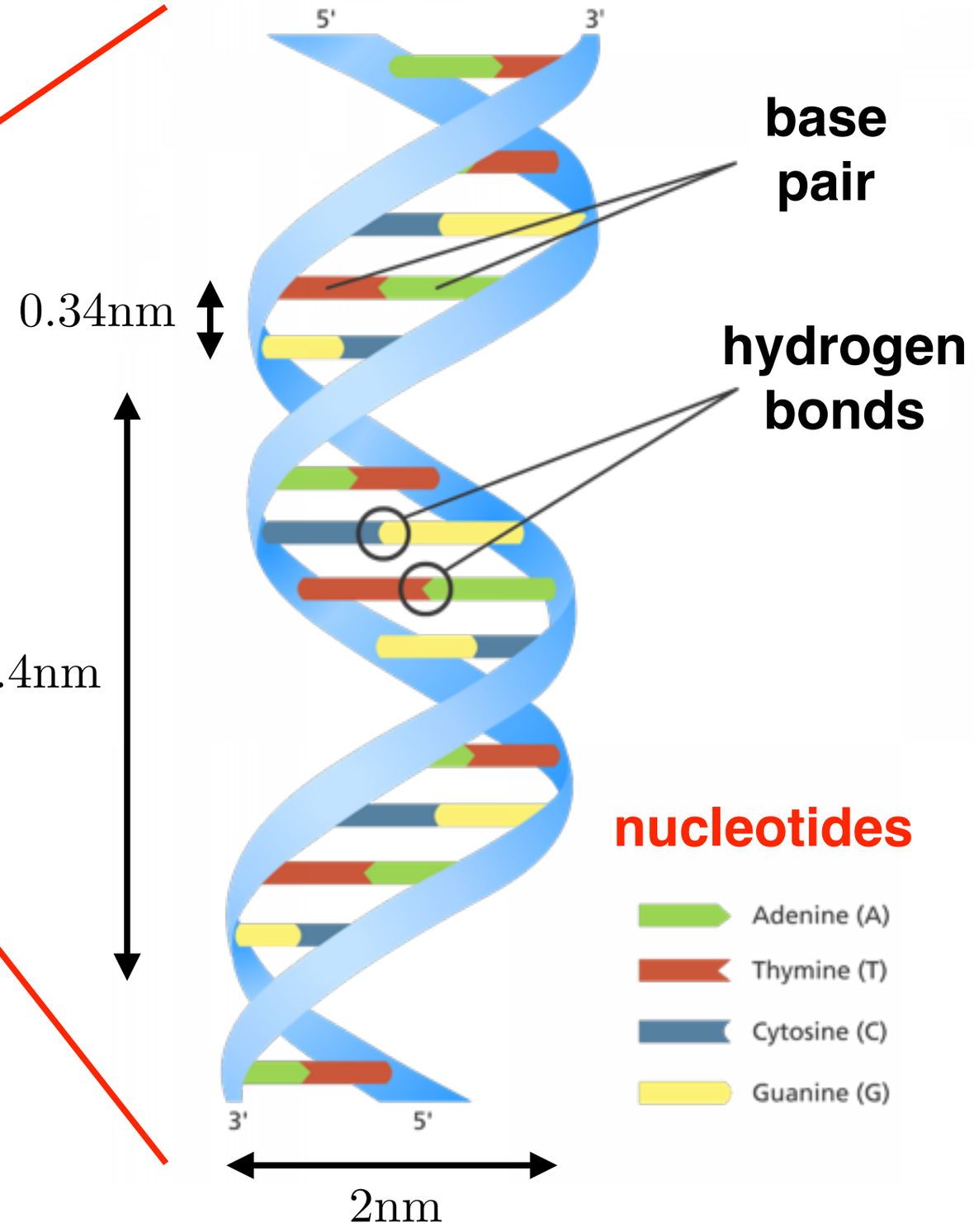


DNA

prokaryotic cell (bacteria, archaea)

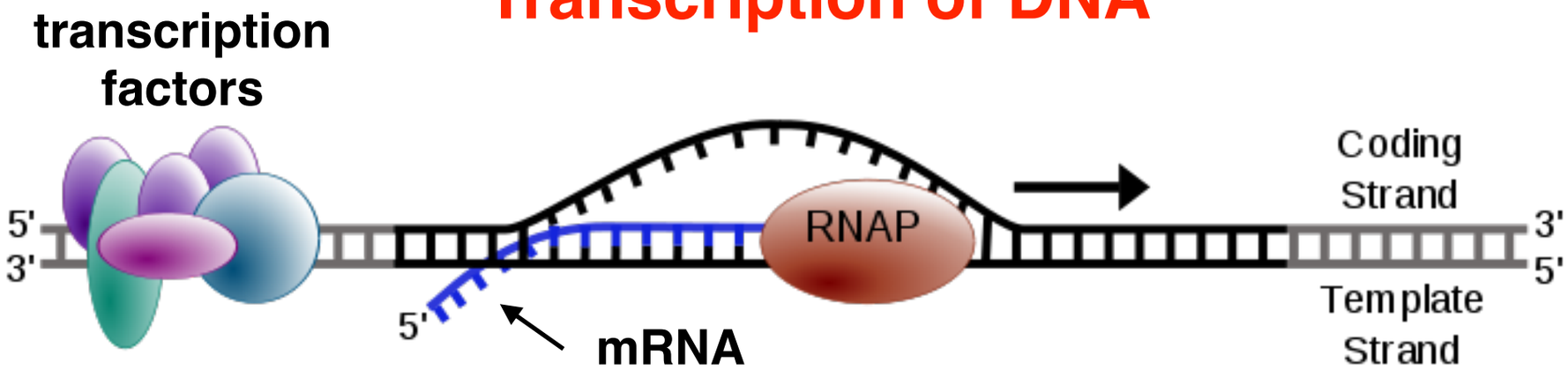


DNA stores genetic information encoded with sequence of bases



Production of new proteins

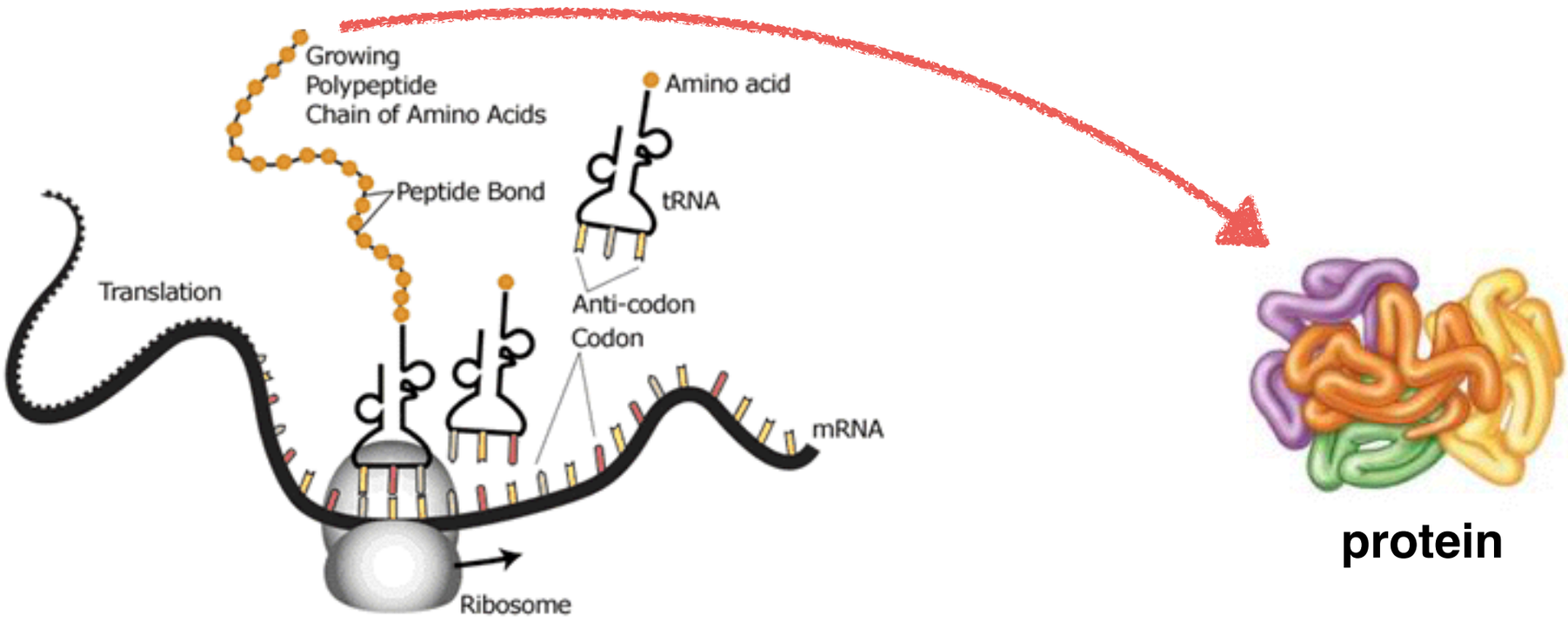
Transcription of DNA



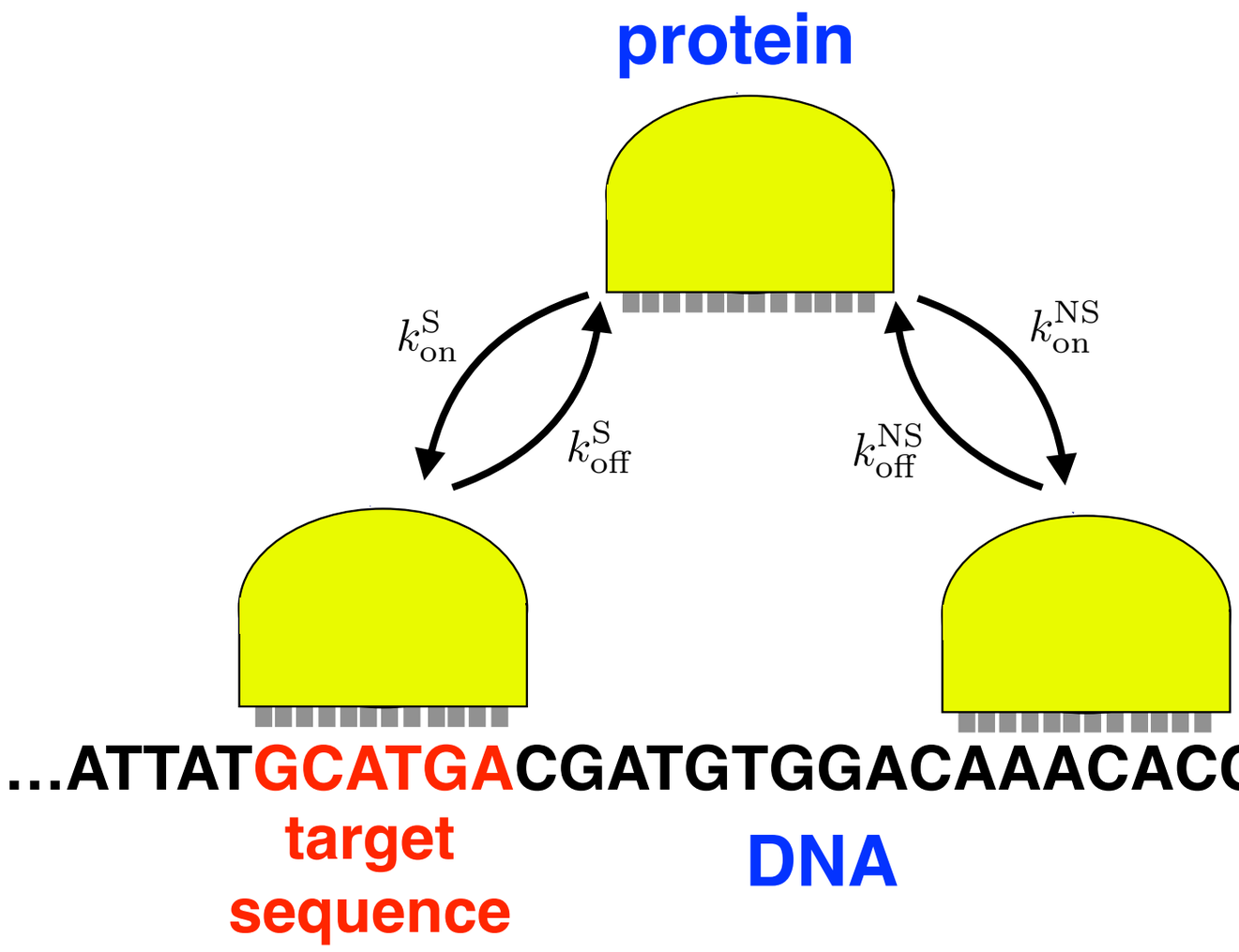
Transcription factors are proteins, which bind to specific locations on DNA, and they help recruiting RNA polymerase (RNAP) that makes a messenger RNA (mRNA) copy of certain DNA segment.

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.

---

on rates are diffusion limited

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

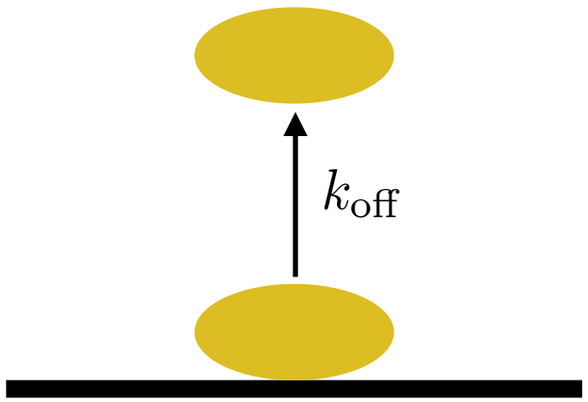
(see slide 5)

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 $t+\Delta t$:

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

$\Delta t \to 0$

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

Average binding time

$$\langle t \rangle = k_{\rm off}^{-1}$$

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

How many target sites are occupied?

1. Ignore non-specific sites

$$[\mathrm{P}] + [\mathrm{S}] \underset{k^{\mathrm{S}}_{\mathrm{off}}}{\overset{k^{\mathrm{S}}_{\mathrm{on}}}{\rightleftharpoons}} [\mathrm{P\text{-}S}]$$

[P] concentration of free proteins

[S] concentration of empty target sites

[P-S] concentration of proteins bound to target sites

Kinetics

$$\frac{d[\mathrm{P\text{-}S}]}{dt} = k^{\mathrm{S}}_{\mathrm{on}}[\mathrm{P}][\mathrm{S}] - k^{\mathrm{S}}_{\mathrm{off}}[\mathrm{P\text{-}S}]$$

Equilibrium

$$K^{\mathrm{S}}_{\mathrm{eq}} = \frac{[\mathrm{P}][\mathrm{S}]}{[\mathrm{P\text{-}S}]} = \frac{k^{\mathrm{S}}_{\mathrm{off}}}{k^{\mathrm{S}}_{\mathrm{on}}}$$

Probability that protein is bound to target site

$$p_B = \frac{[\mathrm{P\text{-}S}]}{[\mathrm{P\text{-}S}] + [\mathrm{S}]} = \frac{[\mathrm{P}]}{[\mathrm{P}] + K^{\mathrm{S}}_{\mathrm{eq}}}$$

How many target sites are occupied?

2. Include non-specific sites

$$[\mathrm{P}] + [\mathrm{S}] \underset{k^{\mathrm{S}}_{\mathrm{off}}}{\overset{k^{\mathrm{S}}_{\mathrm{on}}}{\rightleftharpoons}} [\mathrm{P\text{-}S}]$$

$$[\mathrm{P}] + [\mathrm{NS}] \underset{k^{\mathrm{NS}}_{\mathrm{off}}}{\overset{k^{\mathrm{NS}}_{\mathrm{on}}}{\rightleftharpoons}} [\mathrm{P\text{-}NS}]$$

[P] concentration of free proteins

[S] concentration of empty target sites on DNA

[P-S] concentration of proteins bound to target sites

[NS] concentration of empty non-specific sites on DNA

[P-NS] concentration of proteins bound to non-specific sites

Note that the total number of non-specific sites on DNA (~10^6) is much larger than the total number of relevant proteins and target sites.

Total concentration of proteins

$$[\mathrm{P_{tot}}] = [\mathrm{P}] + [\mathrm{P\text{-}NS}] + [\mathrm{P\text{-}S}] \approx [\mathrm{P}] + [\mathrm{P\text{-}NS}]$$

Equilibrium

$$K^{\mathrm{S}}_{\mathrm{eq}} = \frac{[\mathrm{P}][\mathrm{S}]}{[\mathrm{P\text{-}S}]} = \frac{k^{\mathrm{S}}_{\mathrm{off}}}{k^{\mathrm{S}}_{\mathrm{on}}} \qquad K^{\mathrm{NS}}_{\mathrm{eq}} = \frac{k^{\mathrm{NS}}_{\mathrm{off}}}{k^{\mathrm{NS}}_{\mathrm{on}}} = \frac{[\mathrm{P}][\mathrm{NS}]}{[\mathrm{P\text{-}NS}]} = \frac{[\mathrm{P}][\mathrm{NS}]}{([\mathrm{P_{tot}}] - [\mathrm{P}])} \rightarrow [\mathrm{P}] = \frac{[\mathrm{P_{tot}}]}{\left(1 + [\mathrm{NS}]/K^{\mathrm{NS}}_{\mathrm{eq}}\right)}$$

Probability that protein is bound to target site

$$p_B = \frac{[\mathrm{P\text{-}S}]}{[\mathrm{P\text{-}S}] + [\mathrm{S}]} = \frac{[\mathrm{P}]}{[\mathrm{P}] + K^{\mathrm{S}}_{\mathrm{eq}}} = \frac{[\mathrm{P_{tot}}]}{[\mathrm{P_{tot}}] + K^{\mathrm{S}}_{\mathrm{eq}}\left(1 + [\mathrm{NS}]/K^{\mathrm{NS}}_{\mathrm{eq}}\right)}$$

Example of lac repressor in E. coli

E. coli helps us metabolize lactose that is present in milk. When lactose is absent lac repressor binds to a specific site on DNA to stop the production of relevant enzymes in order to save the energy that is needed for the enzyme production.

1. Lactose absent

10 lac repressors in E. coli, $V \approx 1\mu\mathrm{m}^3$

$$[\mathrm{P_{tot}}] = \frac{10}{V} \sim 10^{-8} M \qquad K_{\mathrm{eq}}^{\mathrm{S}} \sim 10^{-12} M$$

$$[\mathrm{NS}] \sim \frac{10^6}{V} \sim 10^{-3} M \qquad K_{\mathrm{eq}}^{\mathrm{NS}} \sim 10^{-6} M$$

$$[\mathrm{P}] = \frac{[\mathrm{P_{tot}}]}{\left(1 + [\mathrm{NS}]/K_{\mathrm{eq}}^{\mathrm{NS}}\right)} \sim [\mathrm{P_{tot}}] \times 10^{-3}$$

all lac repressors are bound to DNA

$$p_B = \frac{[\mathrm{P_{tot}}]}{[\mathrm{P_{tot}}] + K_{\mathrm{eq}}^{\mathrm{S}}\left(1 + [\mathrm{NS}]/K_{\mathrm{eq}}^{\mathrm{NS}}\right)} \sim 0.9$$

target site is occupied most of the time

2. Lactose present

$$K_{\mathrm{eq}}^{\mathrm{S}} \sim 10^{-9} M \longrightarrow p_B \sim 0.01$$

target site is empty most of the time

Lactose binds to lac repressor and modifies biding free energies

Molar concentration

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

How quickly proteins find target sites on DNA?

1917 Smoluchowski theory

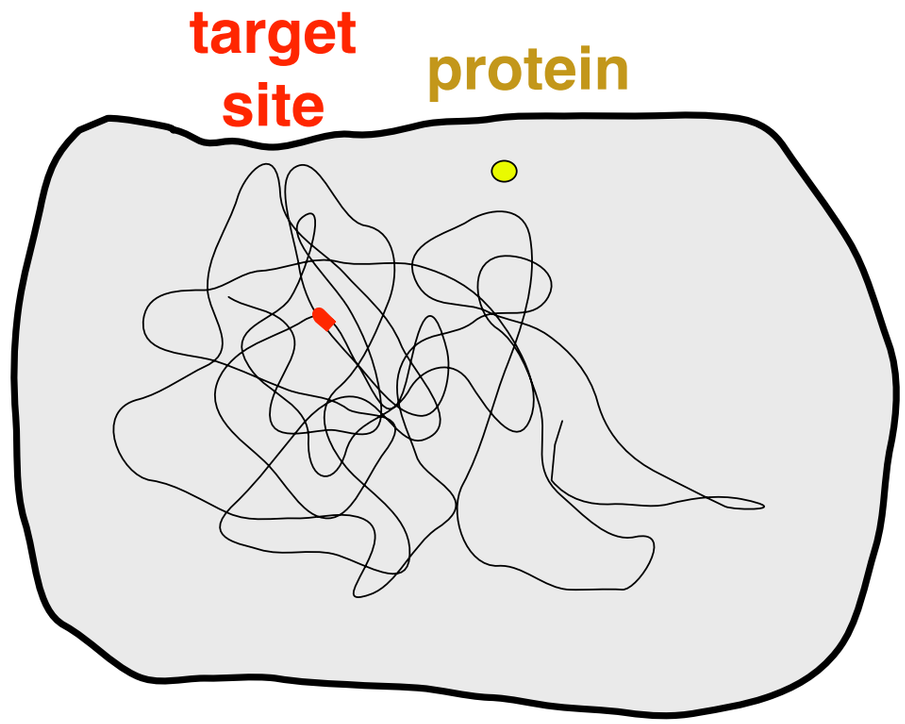
Fick's law

$$\frac{\partial c}{\partial t} = D_3 \nabla^2 c = D_3 \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial c}{\partial r} \right)$$

boundary conditions

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

$$c(b) = 0$$



steady state

$$c(r) = [P] \left(1 - \frac{b}{r} \right)$$

flux density of molecules

$$J(r) = -D_3 \frac{\partial c(r)}{\partial r} = -\frac{Db[\mathrm{P}]}{r^2}$$

rate of absorbing molecules

$$I(r) = J(r) \times 4\pi r^2 = -4\pi D_3 b[\mathrm{P}] = -k_{\mathrm{on}}[\mathrm{P}]$$

short time binding kinetics for initially empty target sites

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

characteristic search time

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

example lac repressor in E. coli

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

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

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

in vitro experiments (1970)

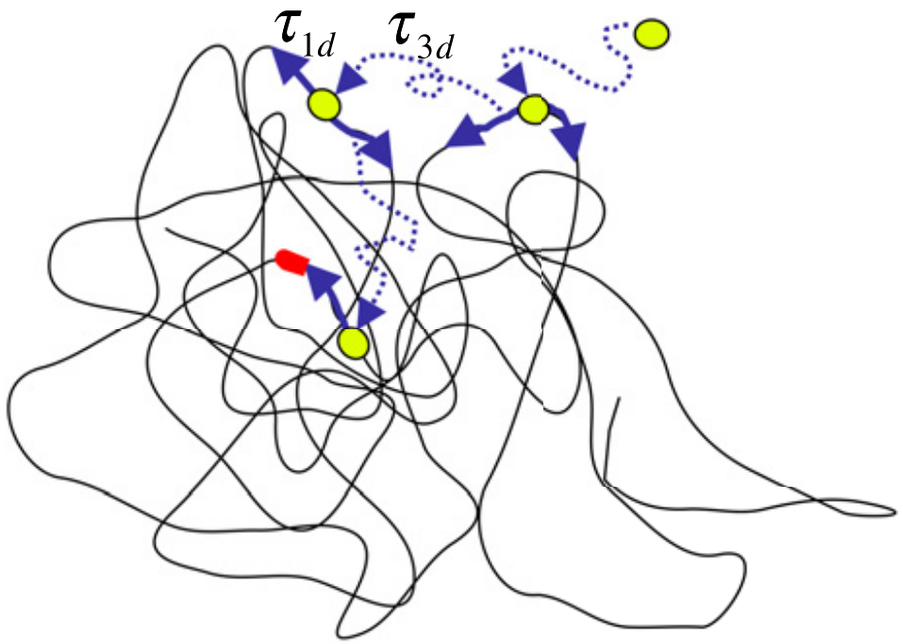
$$k_{\mathrm{on}}^{\mathrm{exp}} \sim 10^{10} M^{-1} s^{-1} \qquad t_s \sim 0.1s$$

Why is experimentally observed rate 100 times larger?

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

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

First assume fixed sliding time τ_{1D}

Number of distinct sites visited during sliding

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

(valid for *n*>>1)

Probability that target site is found during a sliding event

$$q = nb/L$$

Probability that target site is found after N_R rounds

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

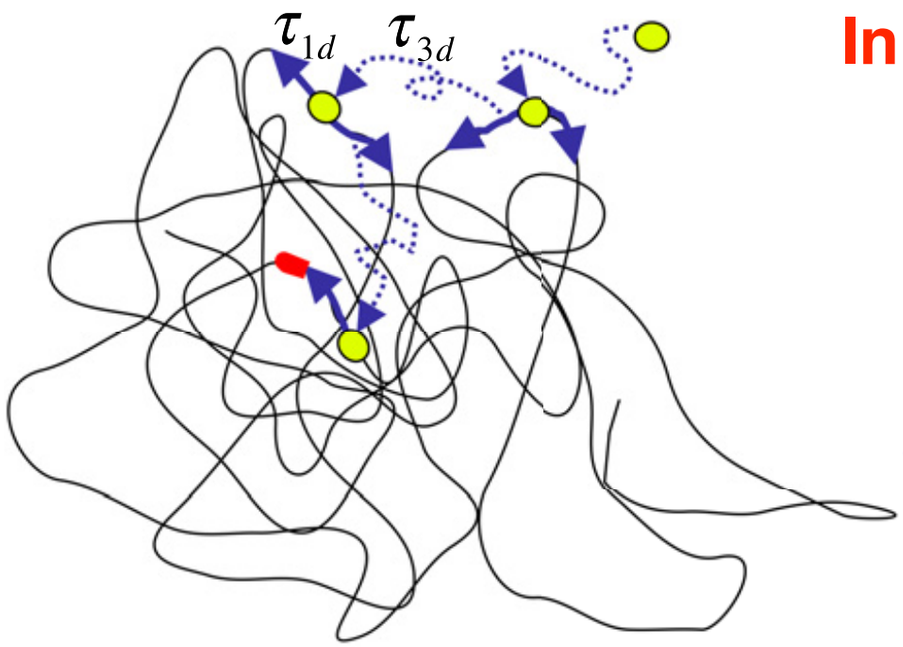
$$\overline{N}_R = 1/q$$

Average search time

$$\bar{t}_s = \overline{N}_R(\tau_{1d} + \tau_{3d})$$

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

Facilitated diffusion



In reality sliding times are exponentially distributed

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

Average number of distinct sites visited during sliding

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

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

Average probability that target site is found during a sliding event

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

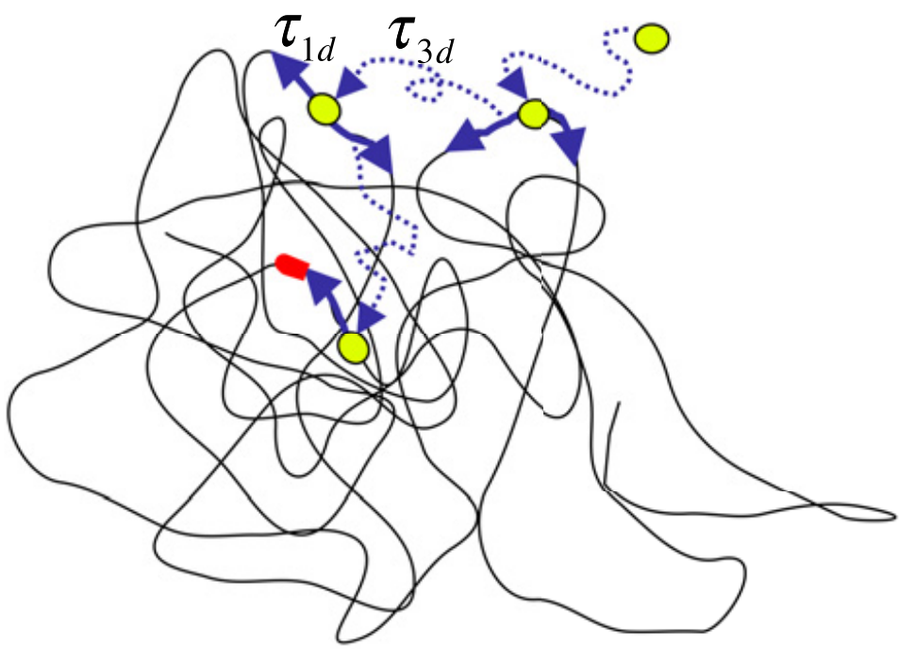
Average number of rounds N_R needed to find the target site

$$\langle \overline{N}_R\rangle = 1/\langle q\rangle$$

Average search time

$$\langle \bar{t}_s\rangle = \langle \overline{N}_R\rangle\,(\langle\tau_{1d}\rangle + \tau_{3d}) = \frac{L}{2\sqrt{D_1\,\langle\tau_{1d}\rangle}}\,(\langle\tau_{1d}\rangle + \tau_{3d})$$

Facilitated diffusion



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

Average search time

$$\langle \bar{t}_s \rangle = \frac{L}{\langle \ell_{\text{sl}} \rangle} \left(\langle \tau_{1d} \rangle + \tau_{3d} \right)$$

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

Optimal search time

$$\frac{d \langle \bar{t}_s \rangle}{d \langle \tau_{1d} \rangle} = 0 \longrightarrow \langle \tau_{1d} \rangle_{\text{opt}} = \tau_{3D}$$

$$\langle \bar{t}_s \rangle_{\text{opt}} = L\sqrt{\frac{\tau_{3d}}{D_1}}$$

Typical jump time

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

Search time for sliding alone

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

Search time for jumps alone

$$\langle t_s \rangle_{\text{jumps}} = \frac{L}{b}\tau_{3d} = \frac{V}{4\pi D_3 b}$$

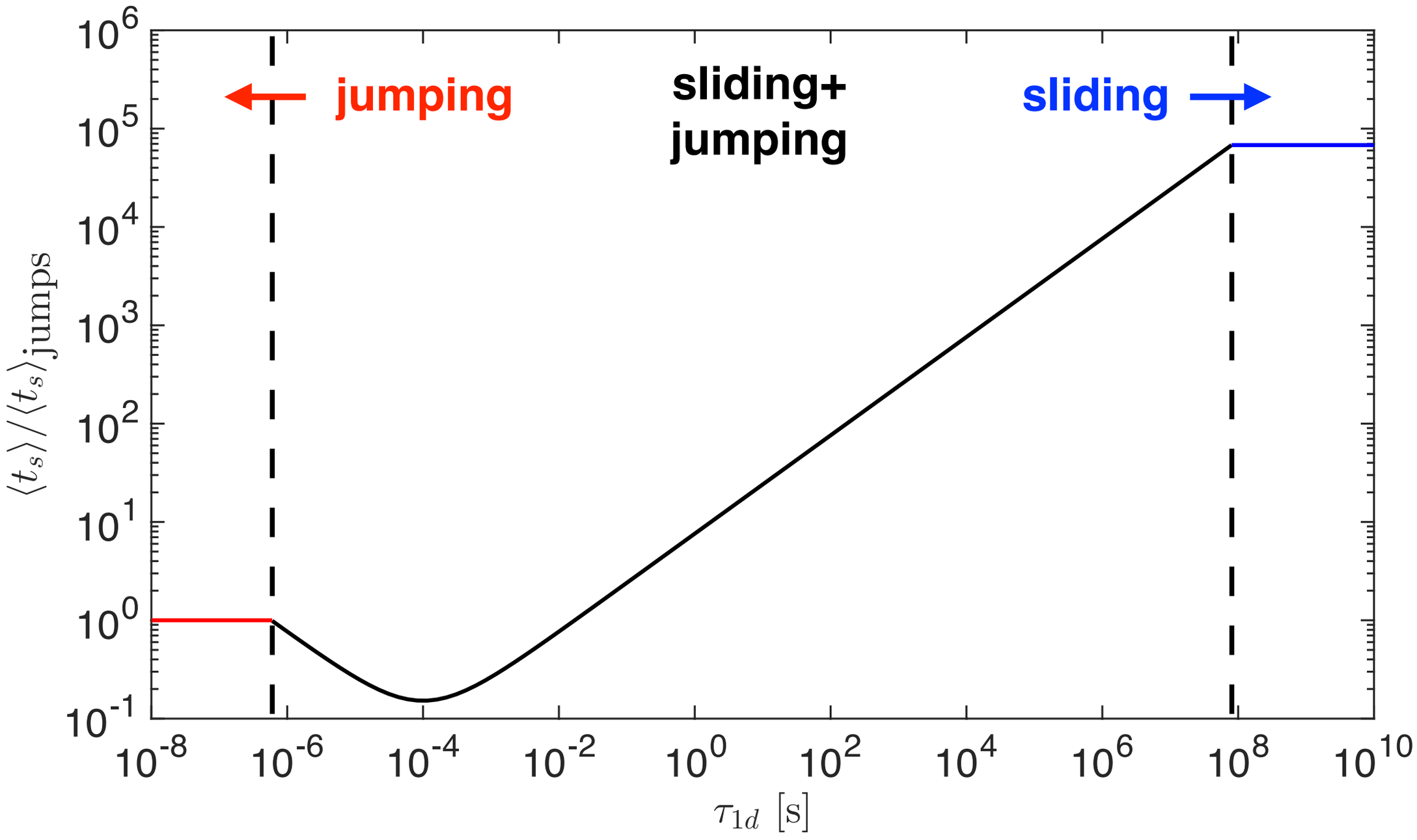
(= Smoluchowski result for [T]=1/V)

Search time speed up

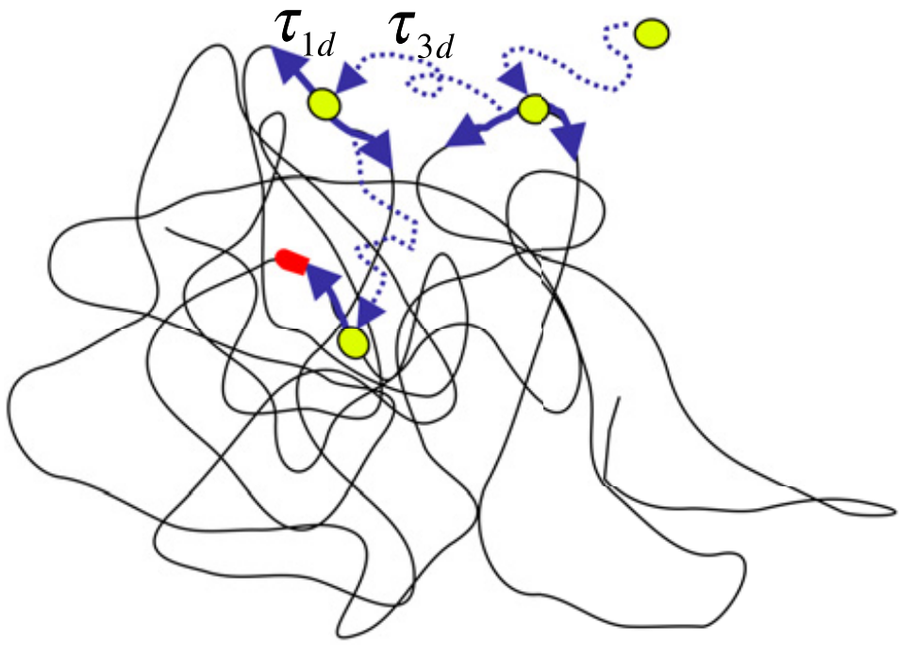
$$\frac{\langle t_s \rangle_{\text{jumps}}}{\langle \bar{t}_s \rangle} = \frac{\langle \ell_{\text{sl}} \rangle}{b} \frac{\tau_{3d}}{\left(\langle \tau_{1d} \rangle + \tau_{3d} \right)}$$

Facilitated diffusion

$$\tau_{3d} = 10^{-4}\mathrm{s} \qquad L = 1\mathrm{mm} \qquad D_1 = 0.05\mu\mathrm{m}^2/s$$



Facilitated diffusion



Is biology operating at the optimum?

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

No! We demonstrated that most of the time proteins are non-specifically bound to DNA and are thus sliding most of the time. For some proteins this actually results in slower search time then by pure 3D diffusion!

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

Lac repressor example

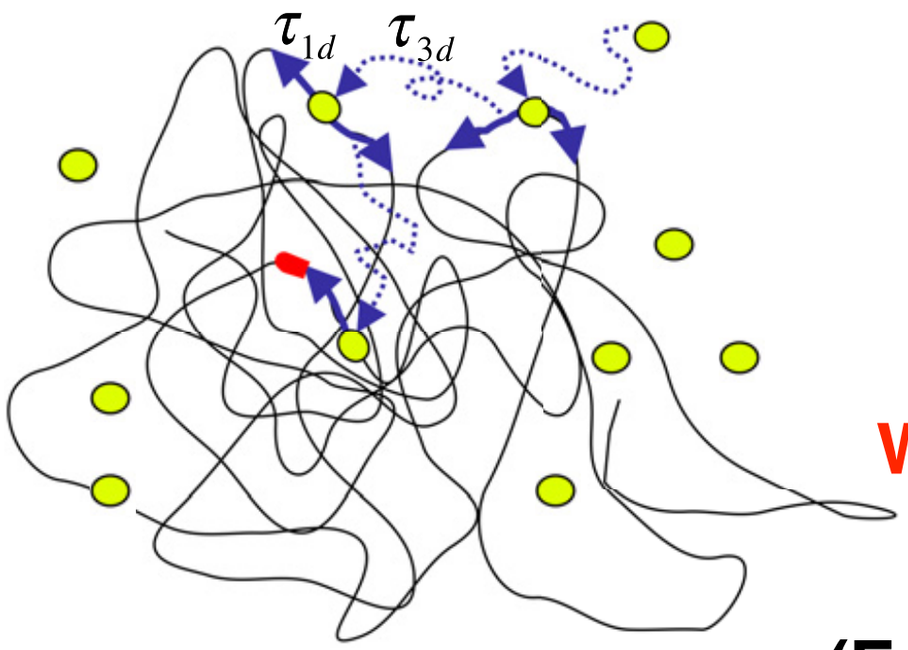
$$L \approx 1\text{mm} \qquad V \approx 1\mu\text{m}^3$$

$$D_1 \approx 0.05\mu\text{m}^2/s \qquad D_3 \approx 30\mu\text{m}^2/s$$

$$\langle \tau_{1d} \rangle \sim 1\text{ms} \qquad \tau_{3d} = \frac{V}{4\pi D_3 L} \sim 3\mu s$$

$$\langle \ell_{\rm sl} \rangle \sim 10\text{nm} \qquad \langle \bar{t}_s \rangle = \frac{L}{\langle \ell_{\rm sl} \rangle} \left(\langle \tau_{1d} \rangle + \tau_{3d} \right) \sim 10 - 100\text{s}$$

Simultaneous search by multiple proteins



Individual search times are exponentially distributed

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

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

(Extreme value distributions)

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

$$p_n(t_s) = \frac{n}{\langle \bar{t}_s \rangle} e^{-nt/\langle \bar{t}_s \rangle}$$

Average search time is reduced by factor *n*

$$\frac{\langle \bar{t}_s \rangle}{n}$$