How proteins find target sites on DNA?

Figure 1. (A) Schematic representation of the protein–DNA search problem. The protein (yellow) must find its target site (red) on a long DNA molecule confined within the cell nucleoid (in bacteria) or cell nucleus (in eukaryotes). Compare with figure 9 (A) which shows confined DNA. (B) The target site must be recognized with 1 base-pair (0.34 nm) precision, as displacement by 1 bp results in a different sequence and consequently a different site.

Figure 2. (A) The mechanism of facilitated diffusion. The search process consists of alternating rounds of 3D and 1D diffusion, each with average duration $\tau_{3D}$ and $\tau_{1D}$, respectively. (B) The antenna effect. During 1D diffusion (sliding) along DNA, a protein visits on average $\bar{n}$ sites. This allows the protein to associate some distance $\sim \bar{n}$ away from the target site and reach it by sliding, effectively increasing the reaction cross-section from 1bp to $\sim \bar{n}$. The antenna effect is responsible for the speed-up by facilitated diffusion.

1.3. History of the problem: theory

To resolve this discrepancy, one possible mechanism of facilitated diffusion that includes both 3D diffusion and effectively 1D diffusion of protein along DNA (the 1D/3D mechanism) was suggested. This mechanism was first proposed and dismissed by Riggs et al [1] but was soon revived and rigorously studied by Richter and Eigen [3], then further expanded and corrected by Berg and Blomberg [4] and finally developed by Berg et al [5]. The basic idea of the 1D/3D mechanism is that while searching for its target site, the protein repeatedly binds and unbinds DNA and, while bound non-specifically, slides along the DNA, undergoing one-dimensional (1D) Brownian motion or a random walk. Upon dissociation from the DNA, the protein diffuses three dimensionally in solution and binds to the DNA in a different place for the next round of one-dimensional searching (figure 2 (A)).

During 1D sliding the protein is kept on DNA by the binding energy to non-specific DNA. This energy has been measured for several DNA-binding proteins and has a range of $10–15 k_B T$ (at physiological salt concentration), was shown to be driven primarily by screened electrostatic interactions between charged DNA and protein molecules [6].
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.

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

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

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 $n>>1$)

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

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

Average search time

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

$\tau_{1d}$ - characteristic jumping time
$\tau_{3d}$ - characteristic jumping time

$\overline{N_R}$ - average number of rounds

$\overline{t_s}$ - average search time

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

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

\[ \langle N_R \rangle = 1/\langle q \rangle \]

Average search time

\[ \langle t_s \rangle = \langle N_R \rangle (\langle \tau_{1d} \rangle + \tau_{3d}) \]

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

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

\( D_3 \) - diffusion constant in space

\( D_1 \) - diffusion constant along the DNA

\( \tau_{3d} \) - characteristic jumping time
Facilitated diffusion

Average search time

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

Average sliding length

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

Optimal search time

$$\frac{d \langle t_s \rangle}{d \langle \tau_{1d} \rangle} = 0 \quad \Rightarrow \quad \langle t_s \rangle_{\text{opt}} = L \sqrt{\frac{\tau_{3d}}{D_1}}$$

**Search time for jumps alone**

Typical jump time

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

Concentration of non-specific sites

$$[NS] = \frac{L/b}{V}$$

average number of jumps needed to find the target

$$\bar{N}_{\text{jumps}} = \frac{L}{b}$$

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

**Search time for sliding alone**

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

Search time speed up for facilitated diffusion

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

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

$$D_3 \text{ - diffusion constant in space}$$

$$D_1 \text{ - diffusion constant along the DNA}$$
Example: search time for target site in bacteria on DNA with $10^6$ base pairs

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

$D_1 = 0.05 \mu m^2 / s$

$L = 1 \text{mm}$

$b = 0.34 \text{nm}$

search time for jumps alone

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

average search time

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

average sliding length

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

Graph showing the distribution of search time with sliding and jumping events.
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 \( n \)

\[ \int_0^{\infty} dt_s \, t_s \, p_n(t_s) = \frac{\langle t_s \rangle}{n} \]
Statistical mechanics of polymers and filaments
Statistical mechanics of polymers and filaments

molecular dynamics simulation

Note: averaging over time is equivalent to averaging over all possible configurations weighted with Boltzmann weights!

partition function (sum over all possible configurations)

\[ Z = \sum_c e^{-E_c/k_B T} \]

expected value of observables

\[ \langle \mathcal{O} \rangle = \sum_c \mathcal{O}_c \frac{e^{-E_c/k_B T}}{Z} \]

- $E_c$: energy of a given configuration
- $T$: temperature
- $k_B$: Boltzmann constant
- $k_B = 1.38 \times 10^{-23} \text{JK}^{-1}$
Persistence length

correlations between tangents

\[ \langle t(s) \cdot t(s + x) \rangle = e^{-x/\ell_p} \]

tangents become uncorrelated beyond persistence length!

Persistence length

\[ \ell_p = \frac{B}{k_B T} \]

Short filaments remain straight

Long filaments perform self-avoiding random walk

\[ L \ll \ell_p \]

\[ L \gg \ell_p \]

B - filament bending rigidity

T - temperature

L - filament length
Examples: persistence length

polyethylene
\[ \ell_p = 2.6 \text{ nm} \]

Persistence length for polymers is on the order of nm

actin
\[ \ell_p \approx 17 \mu \text{m} \]

microtubule
\[ \ell_p \approx 1.4 \text{ mm} \]

double stranded DNA
\[ \ell_p \approx 50 \text{ nm} \]

single stranded DNA
\[ \ell_p \approx 2 \text{ nm} \]

uncooked spagetti
\[ \ell_p \approx 10^{18} \text{ m} \]
**End-to-end distance**

Short filaments \( L \ll \ell_p \)

Long filaments \( L \gg \ell_p \)

Over time thermal fluctuations reorient filaments in all possible directions!

\[
\langle \vec{R}_{AB} \rangle = 0
\]

\[
\langle \vec{R}_{AB}^2 \rangle \approx L^2
\]

**Exact result**

\[
\langle \vec{R}_{AB}^2 \rangle = 2\ell_p L \left[ 1 - \frac{\ell_p}{L} \left( 1 - e^{-L/\ell_p} \right) \right]
\]

Polymers shrink, when temperature is increased!
Negative thermal expansion of rubber.

\[
\langle \vec{R}_{AB} \rangle = 0
\]

\[
\langle \vec{R}_{AB}^2 \rangle \approx 2\ell_p L = \frac{2BL}{k_BT}
\]
Ideal chain vs worm-like chain

Ideal chain

\(N\) identical unstretchable links (Kuhn segments) of length \(a\) with freely rotating joints

![Diagram of Ideal Chain]

Each configuration \(C\) has zero energy cost.

\[ E_c = 0 \]

Worm-like chain

Continuous unstretchable rod

![Diagram of Worm-like Chain]

Bending energy cost of configuration \(C\):

\[ E_c = \frac{B}{2} \int_0^L ds \left( \frac{d^2 \vec{r}}{ds^2} \right)^2 \]

Each configuration \(C\) appears with probability

\[ p_c \propto e^{-E_c / k_B T} \]

\(L = Na\) - chain length
Ideal chain vs worm-like chain

**Ideal chain**

\[ N \text{ identical unstretchable links (Kuhn segments) of length } a \text{ with freely rotating joints} \]

\[ \langle \vec{R}_{AB}^2 \rangle = Na^2 = aL \]

**Worm-like chain**

Continuous unstretchable rod

\[ \langle \vec{R}_{AB}^2 \rangle \approx 2\ell_p L = \frac{2BL}{k_BT} \]

End-to-end distance fluctuations can be made identical if one chooses the segment length to be

\[ a = 2\ell_p \]

\[ L = Na - \text{chain length} \]
Stretching of ideal freely jointed chain

 Exact result for end-to-end distance

\[ \left\langle x \right\rangle = Na \left( \coth \left[ \frac{Fa}{k_B T} \right] - \frac{k_B T}{Fa} \right) \]

small force \( Fa \ll k_B T \)

\[ \left\langle x \right\rangle \approx \frac{FNa^2}{3k_B T} = \frac{2FLl_p}{3k_B T} \]

large force \( Fa \gg k_B T \)

\[ \left\langle x \right\rangle \approx Na \left( 1 - \frac{k_B T}{Fa} \right) = L \left( 1 - \frac{k_B T}{2Fl_p} \right) \]
Stretching of worm-like chains

Assume long chains \( L \gg \ell_p \)

small force \( F\ell_p \ll k_B T \)

large force \( F\ell_p \gg k_B T \)

\[
\langle x \rangle \approx \frac{2FL\ell_p}{3k_B T} = \frac{F}{k}
\]

entropic spring constant

\[
k = \frac{3k_B T}{2L\ell_p} = \frac{3k_B^2 T^2}{2LB}
\]

\( B \) - filament bending rigidity

Approximate expression that interpolates between both regimes

\[
\frac{F\ell_p}{k_B T} = \frac{1}{4} \left( 1 - \frac{\langle x \rangle}{L} \right)^{-2} - \frac{1}{4} + \frac{\langle x \rangle}{L}
\]

J.F. Marko and E.D. Siggia,
Experimental results for stretching of DNA

\[ L = 32.8 \mu m \]

\[ \langle x \rangle \approx L \left[ 1 - \sqrt{\frac{k_B T}{4F\ell_p}} \right] + \frac{FL}{\gamma} \]

For DNA

\[ \ell_p = 50 \text{nm} \]

\[ \gamma \approx 500k_B T/\text{nm} \approx 2 \text{nN} \]

Improved interpolation formula

\[ \frac{F\ell_p}{k_B T} = \frac{1}{4} \left( 1 - \frac{\langle x \rangle}{L} + \frac{F}{\gamma} \right)^{-2} - \frac{1}{4} \]

\[ + \frac{\langle x \rangle}{L} - \frac{F}{\gamma} \]

Random coil to globule transition in polymers

**random coil**

\[ T > \Theta \]

\[ R \sim \sqrt{L \ell_p} \]

at high temperature
entropic contributions dominate

**compact globule**

\[ T < \Theta \]

\[ R \sim (d^2 L)^{1/3} \]

\( d \)-diameter of polymer chain
at low temperature
attraction between polymer chains dominates

Further reading
Dynamics of actin filaments and microtubules

Actin filament

Microtubule
Cytoskeleton in cells

Cytoskeleton matrix gives the cell shape and mechanical resistance to deformation.

(wikipedia)
Crawling of cells

Immune system: neutrophils chasing bacteria

Migration of skin cells during wound healing

Spread of cancer cells during metastasis of tumors

Amoeba searching for food

$\nu \sim 0.1 \mu m/s$
Movement of bacteria

Listeria monocytogenes
moving in infected cells

Julie Theriot (speeded up 150x)

$v \sim 0.1 - 0.3 \mu m/s$

L. A. Cameron et al.,
Molecular motors

A. Myosin V

B. Myosin II

C. Kinesin-1, Dynein

Transport of large molecules around cells (diffusion too slow)

\[ v \sim 1 \mu \text{m/s} \]

Contraction of muscles


Harvard BioVisions
Cell division

Segregation of chromosomes

- Spindle pole
- Replicated chromosome with sister chromatids
- Kinetochore
- Astral microtubules
- Kinetochore microtubules
- Interpolar microtubules

Contractile ring divides the cell in two

- Contractile ring

Microtubules

Actin
Swimming of sperm cells

Jeff Guasto

$v \sim 50 \mu m/s$

Swimming of Chlamydomonas (green alga)

Jeff Guasto

$v \sim 60 \mu m/s$

Bending is produced by motors walking on neighboring microtubule-like structures
Actin filaments

7nm

actin monomer

Minus end (pointed end)

Plus end (barbed end)

Persistence length \( \ell_p \sim 10 \mu m \)

Typical length \( L \lesssim 10 \mu m \)

Actin treadmilling

Hydrolysis of ATP

Exchange of ATP for ADP

\( \text{ADP-actin} \)

\( \text{ATP-actin} \)
Dynamic filaments

Fig. 11.12 (a) If \([M_c]^+ = [M_c]^−\), both filament ends grow or shrink simultaneously. (b) If \([M_c]^+ \neq [M_c]^−\), there is a region where one end grows while the other shrinks. The vertical line indicates the steady-state concentration \([M_{ss}]\) where the filament length is constant.

The behavior of the filament in the steady-state condition is called treadmilling, as illustrated in Fig. 11.13. Inspection of Table 11.1 tells us that treadmilling should not be observed for microtubules since the critical concentrations at the plus and minus ends of the filament are the same; that is, \([M_c]^+ = [M_c]^−\) and the situation in Fig. 11.12(a) applies. However, \([M_c]^−\) is noticeably larger than \([M_c]^+\) for actin filaments, and treadmilling should occur. If we use the observed rate constants in Table 11.1 for ATP-actin solutions, Eq. (11.7) predicts treadmilling is present at a steady-state actin concentration of 0.17 \(\mu\)M, with considerable uncertainty. A direct measure of the steady-state actin concentration under not dissimilar solution conditions yields 0.16 \(\mu\)M (Wegner, 1982). At treadmilling, the growth rate from Eqs. (11.5) is 

\[
\frac{dn^−}{dt} = \frac{k^−}{k^+_\text{on}} [M] - \frac{k^−}{k^+_\text{off}}
\]

no growth at

\([M]_c^- = \frac{k^−}{k^+_\text{off}}\)

**Steady state regime**

\[
\frac{dn^+}{dt} = -\frac{dn^-}{dt}
\]

\([M]_{ss} = \frac{k^+_\text{off} + k^-}{k^+_\text{on} + k^-} \approx 0.17 \mu\text{M}\)

**front speed**

\[
\frac{dn^+}{dt} = \frac{k^+_\text{on} k^- - k^- k^+_\text{off}}{k^+_\text{on} + k^-} \approx 0.68^{-1}
\]

concentration of free actin monomers

actin shrinks

actin grows

28
Actin filament growing against the barrier

work done against the barrier for insertion of new monomer

\[ W = Fa \]

effective monomer free energy potential without barrier

effective monomer free energy potential with barrier

\[
\begin{align*}
    k_{\text{on}}^+ &\sim 4\pi D_3 a \\
    k_{\text{off}}^+ &\sim e^{-\Delta/k_B T} \\
    k_{\text{on}}^+(F) &\sim k_{\text{on}}^+ e^{-Fa/k_B T} \\
    k_{\text{off}}(F) &\sim k_{\text{off}}^+ 
\end{align*}
\]
Actin filament growing against the barrier

work done against the barrier for insertion of new monomer

\[ W = Fa \]

effective monomer free energy potential with barrier

\[ k_{on}^+ (F) \sim k_{on}^+ e^{-Fa/k_B T} \]

\[ k_{off}^+ (F) \sim k_{off}^+ \]

Growth speed of the tip

\[ v^+(F) = \frac{dn^+(F)}{dt} = k_{on}^+ [M] e^{-Fa/k_B T} - k_{off}^+ \]

Maximal force that can be balanced by growing filament (stall force)

\[ v^+(F_{\text{max}}) = 0 \quad \rightarrow \quad F_{\text{max}} = \frac{k_B T}{a} \ln \left( \frac{k_{on}^+[M]}{k_{off}^+} \right) \]

\[ k_{on}^+ \sim 10 \mu M^{-1} s^{-1} \]

\[ k_{off}^+ \sim 1 s^{-1} \]

\[ [M] \sim 10 \mu M \]

\[ a \approx 2.5 \text{nm} \]

\[ F_{\text{max}} \sim 8 \text{pN} \]
Movement of bacteria

Listeria monocytogenes moving in infected cells

Julie Theriot (speeded up 150x)

\[ v \sim 0.1 - 0.3 \mu m/s \]

Actin polymerization is pushing bacteria

Caps prevent further polymerization in comet tails

Actin is randomly depolymerized in comet tails