

Molecular search in gene regulation

— Shanghai, 1st July 2018 —

Main protagonist: bacteria cells such as E.coli

Cell size: roughly $2\mu\text{m} \times 1/2\mu\text{m}$

Cell volume: $\approx 1\mu\text{m}^3$

DNA length: 4.7×10^6 base pairs or $\approx 1.6\text{mm}$

Number of proteins in cell: $\approx 2.4 \times 10^6$

Different proteins (# genes): 4,300

Some proteins occur only as few or few tens of copies/cell (nM concentrations)



Main protagonist: bacteria cells such as E.coli

Cell size: roughly $2\mu\text{m} \times 1/2\mu\text{m}$

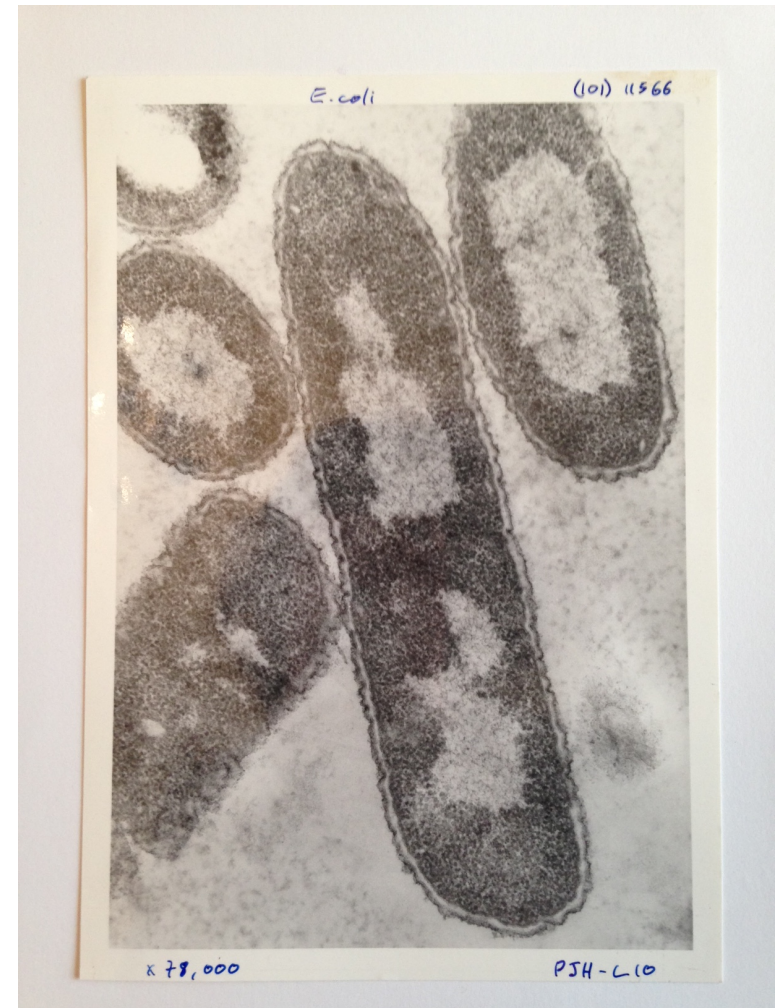
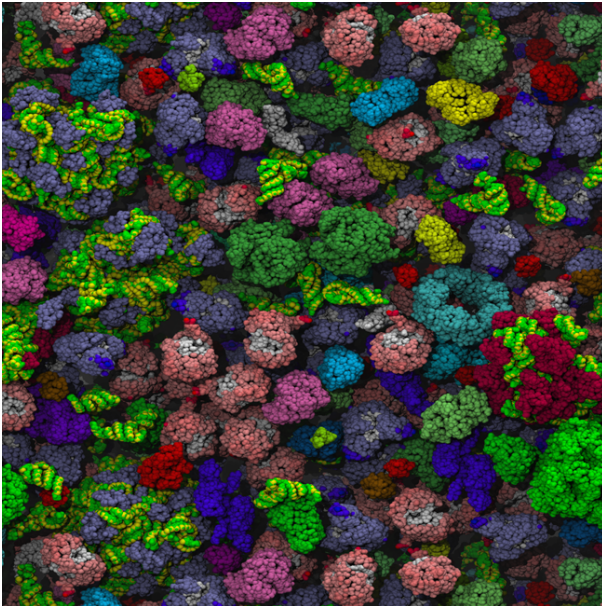
Cell volume: $\approx 1\mu\text{m}^3$

DNA length: 4.7×10^6 base pairs or $\approx 1.6\text{mm}$

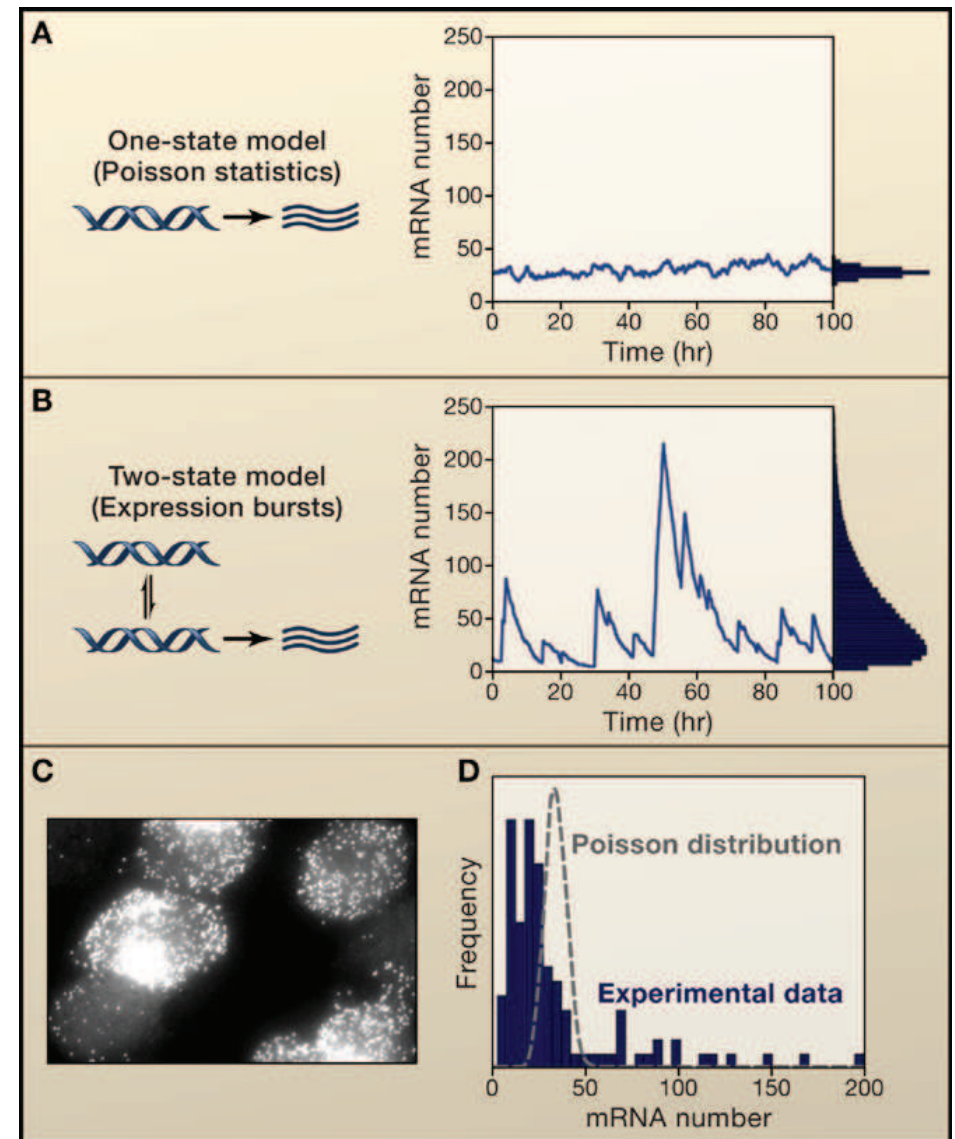
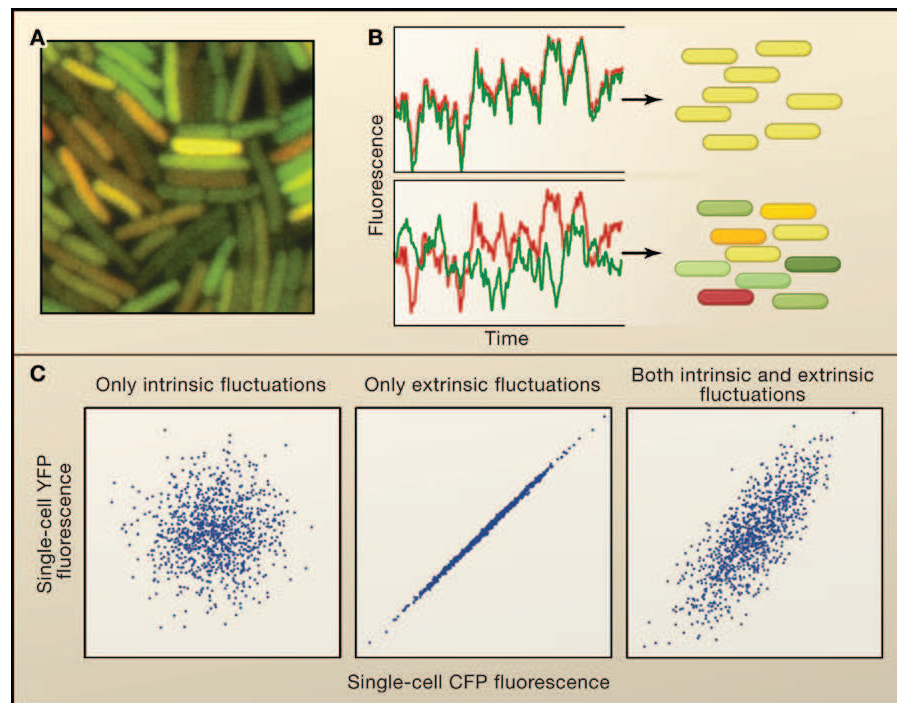
Number of proteins in cell: $\approx 2.4 \times 10^6$

Different proteins (# genes): 4,300

Some proteins occur only as few or few tens of copies/cell (nM concentrations)

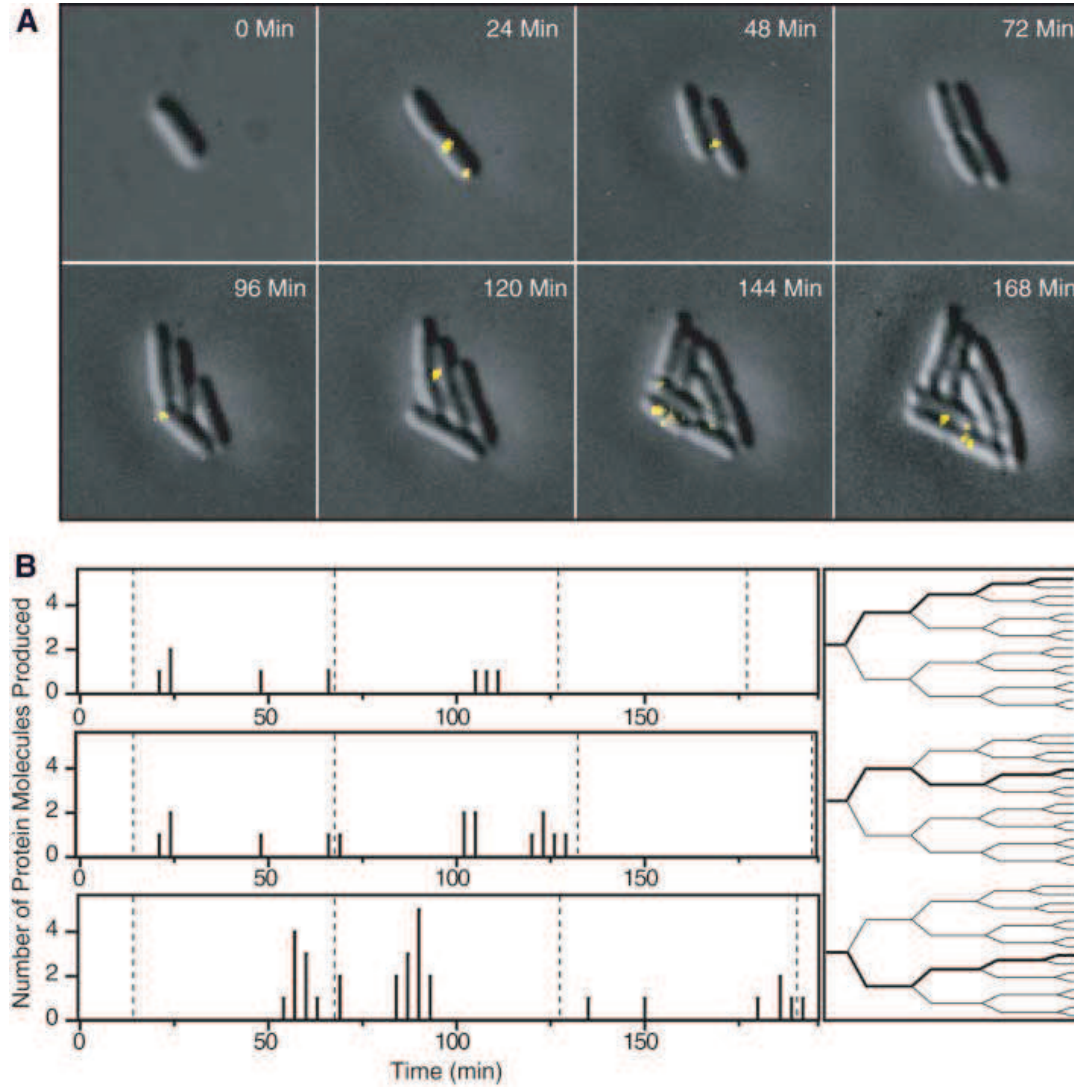


Sources of cellular noisiness: chemical vs physical



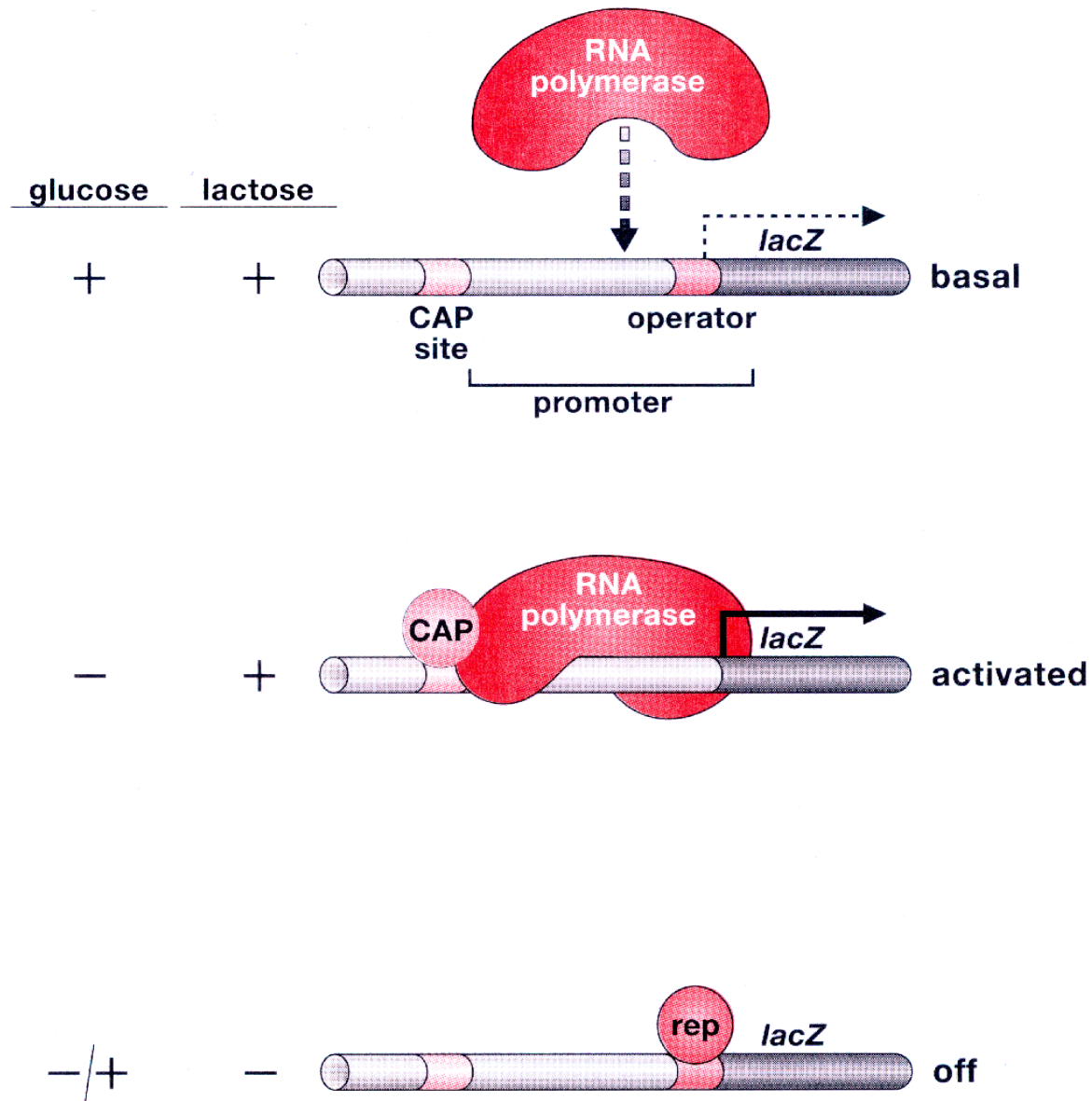
Plus noise due to spatial spreading of TFs!

Gene expression one molecule at a time



synthesised proteins (bursty) along three cell lineages, dashed lines marking cell divisions

Gene regulation by transcription factors: Lac repressor



Smoluchowski search picture

Search rate for a particle with diffusivity D_{3d} to find an immobile target of radius a (assuming immediate binding):

$$k_{\text{on}}^S = 4\pi D_{3d}a$$

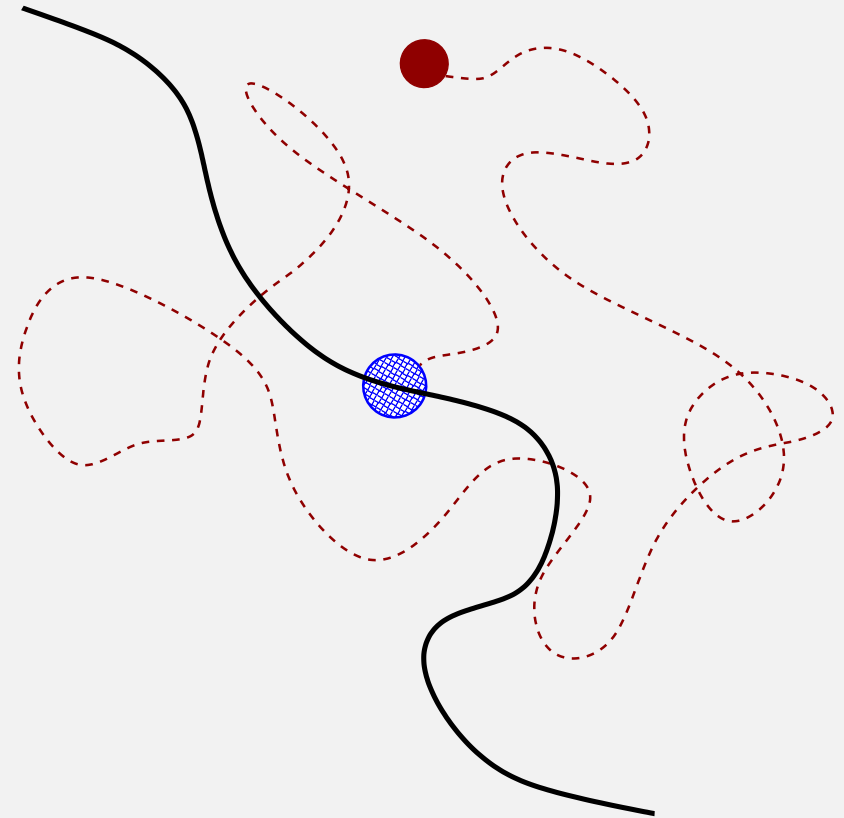
Protein-DNA interaction: $a \approx \{\text{few bp}\} \approx 1\text{nm}$
 $D_{3d} \approx 10\mu\text{m}^2/\text{sec}$ (typically $\varnothing_{\text{TF}} \approx 5\text{nm}$):

$$k_{\text{on}}^S \approx \frac{10^8}{(\text{mol/l}) \times \text{sec}}$$

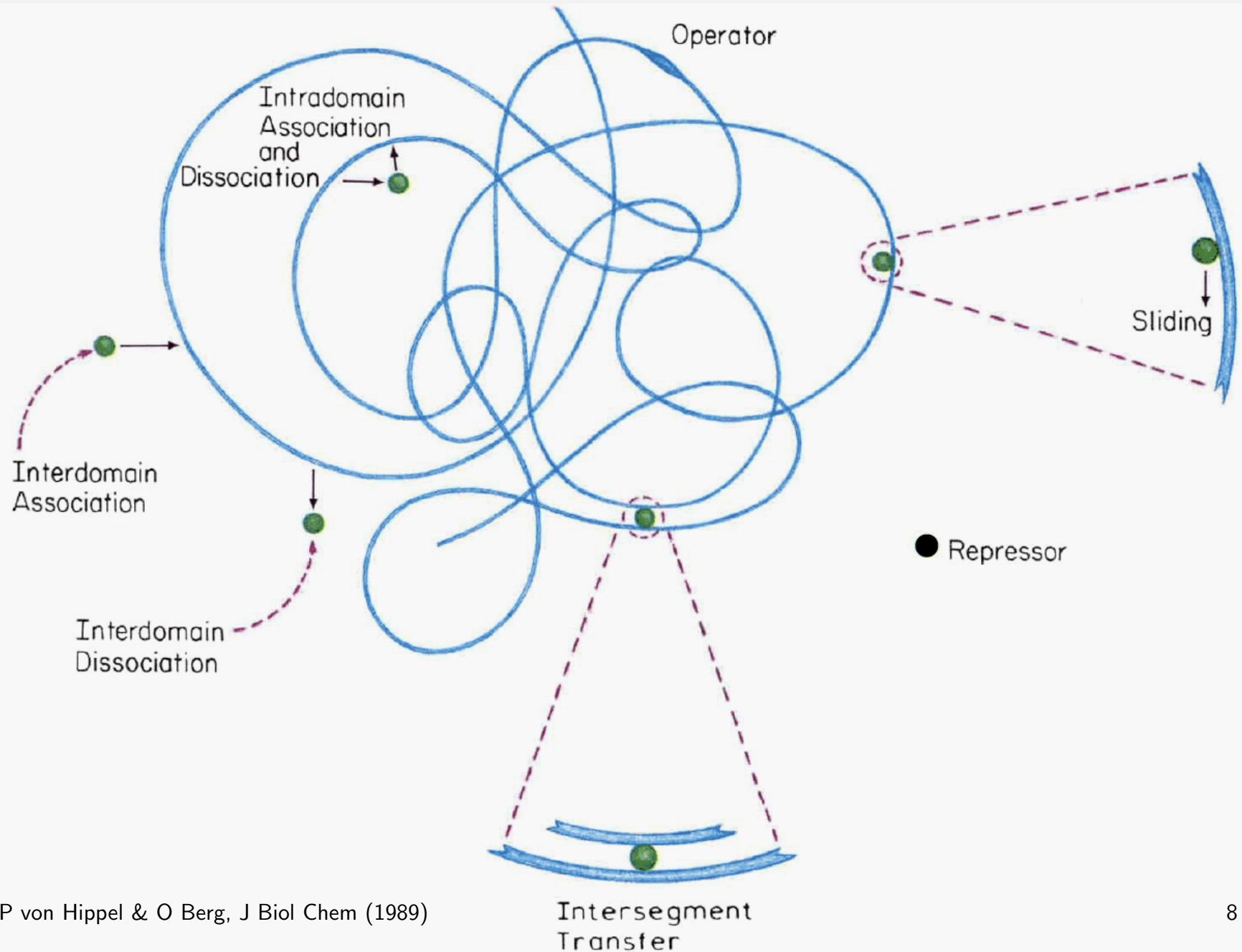
Lac repressor [AD Riggs, S Bourgeois, M Cohn, J Mol Biol 53, 401 (1970)]:

$$k_{\text{on}} \approx \frac{10^{10}}{(\text{mol/l}) \times \text{sec}}$$

→ Facilitated diffusion picture



Facilitated diffusion: the Berg-von Hippel model

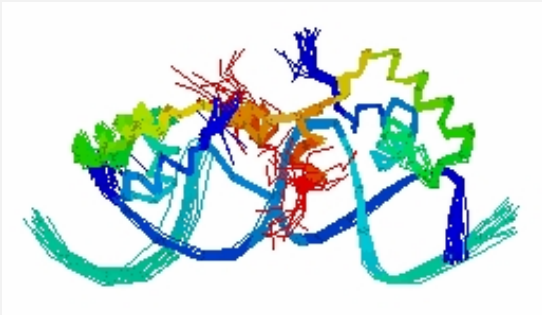


P von Hippel & O Berg, J Biol Chem (1989)

Non-specific binding energy based on *in vivo* data



Lac repressor, nonspecific binding

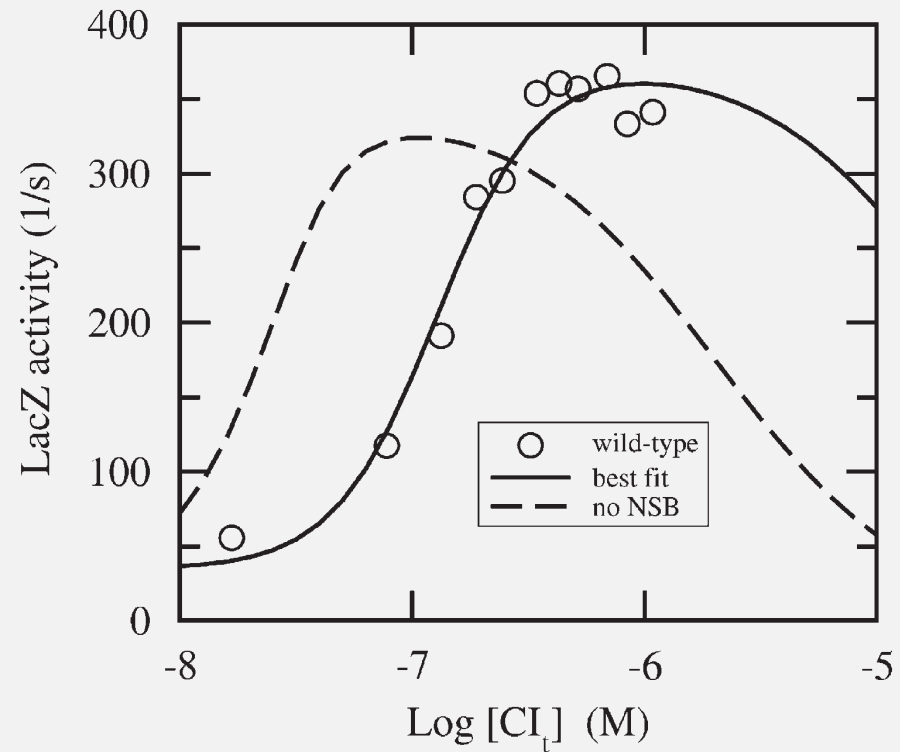


Lac repressor, specific binding

$$[X] = [X_{\text{free}}] + [X_{\text{@O}_P}] + [X_{\text{NSB}}]$$

$$\Delta G_{\text{NSB}}(\text{CI}) = -4.1 \pm 0.9 \text{ kcal/mol},$$

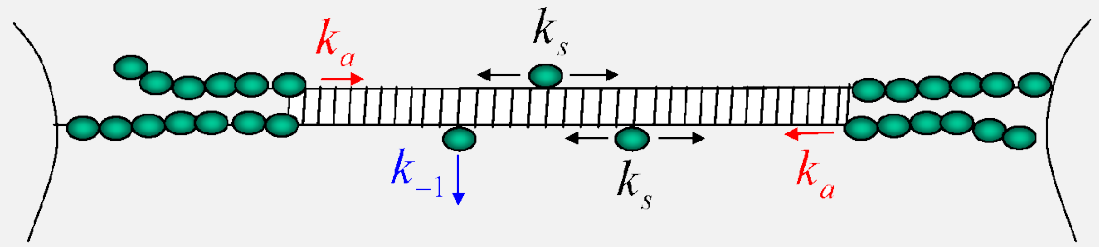
$$\Delta G_{\text{NSB}}(\text{Cro}) = -4.2 \pm 0.8 \text{ kcal/mol}$$



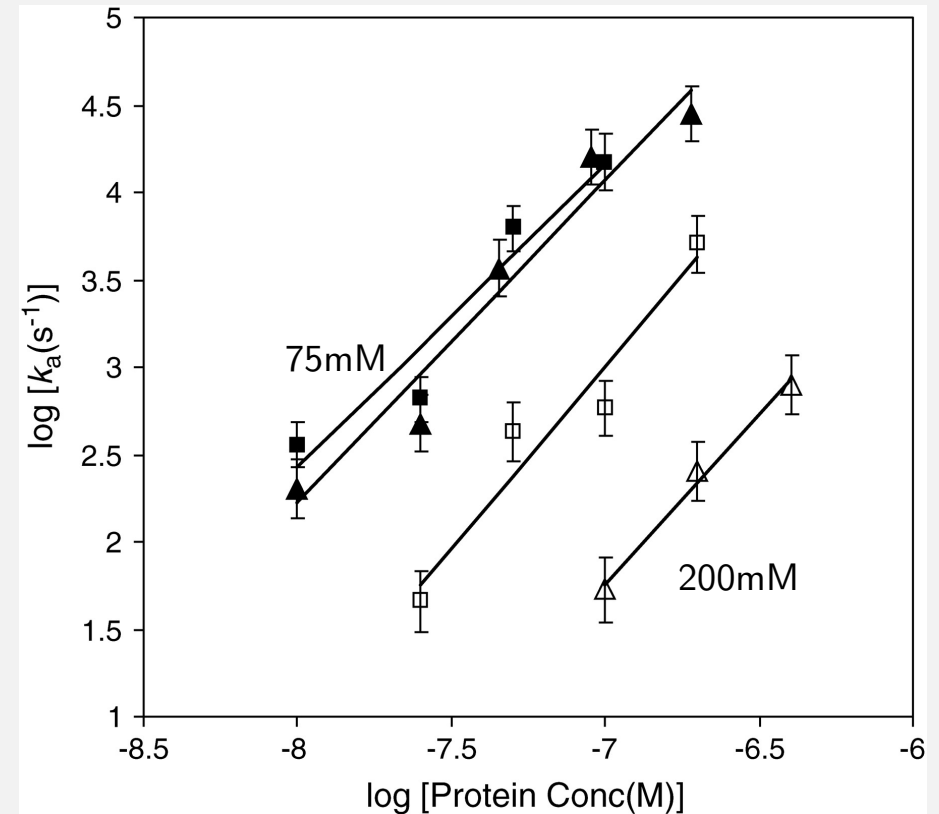
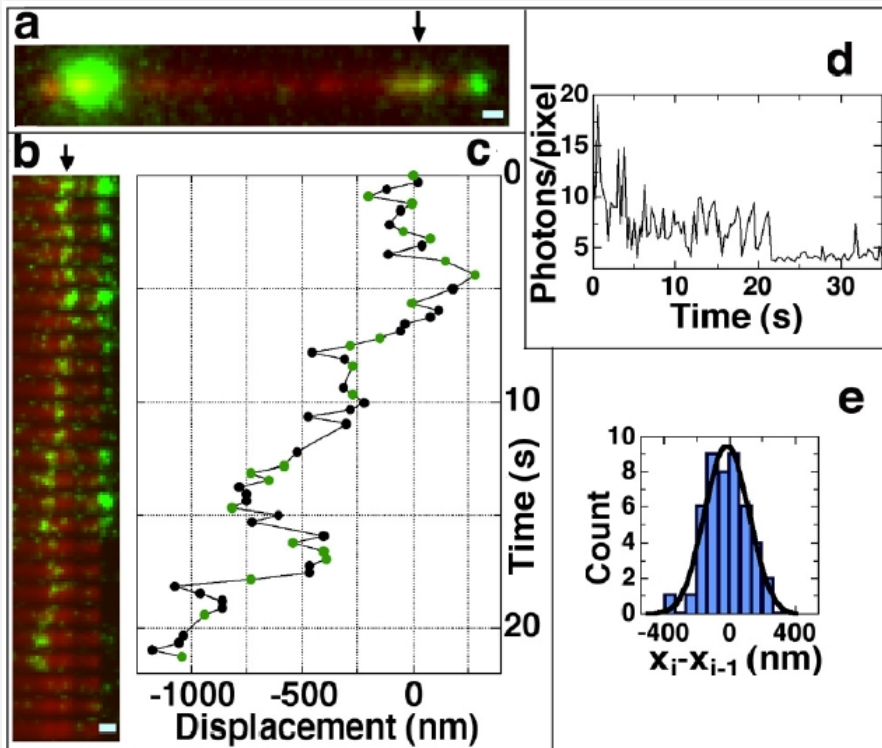
Proof of 1D search mode

McGhee & von Hippel isotherm

$$f = \frac{N\lambda}{L} \simeq K_{\text{ns}}\lambda C, \quad f \ll 1$$



$$k_a \simeq \begin{cases} C, & \text{1D/3D Berg \& von Hippel} \\ C^2, & \text{Pure 1D search} \end{cases}$$



$$\Delta = 1.74 \pm 0.35, 1.85 \pm 0.24, 2.08 \pm 0.39, 1.95 \pm 0.17$$

Calculating facilitated diffusion (our version): manifestation of intermittency

$$\begin{aligned} \frac{\partial n(x, t)}{\partial t} = & \left(D_{1d} \frac{\partial^2}{\partial x^2} - k_{\text{off}} \right) n(x, t) - j(t) \delta(x) + G(x, t) \\ & + k_{\text{off}} \int_{-\infty}^{\infty} dx' \int_0^t dt' W_{\text{bulk}}(x - x', t - t') \end{aligned}$$

n : line density of TFs

x : chemical co-ordinate along DNA

k_{off} : unbinding rate of non-specifically bound TFs

D_{1d} : 1D diffusion constant ($\sim 10^{-2} D_{3d}$)

$j(t)$: flux into target (δ sink @ $x = 0$)

G : virgin flux of previously unbound TFs

W_{bulk} : 3D diffusion propagator

Long chain, fast dynamics: Lévy flights & fractional derivatives

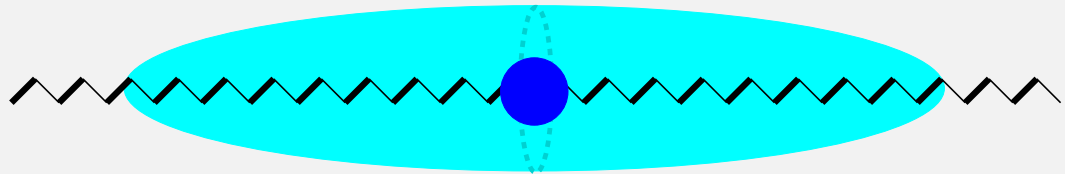
The antenna effect

Target search rate for cylindrical DNA model:

$$k_{\text{on}} \sim 4\pi D_{3d} \ell_{\text{sl}}^{\text{eff}} \times \frac{1}{\sqrt{\ln(\ell_{\text{sl}}^{\text{eff}}/r_{\text{int}})}}$$

Sliding length:

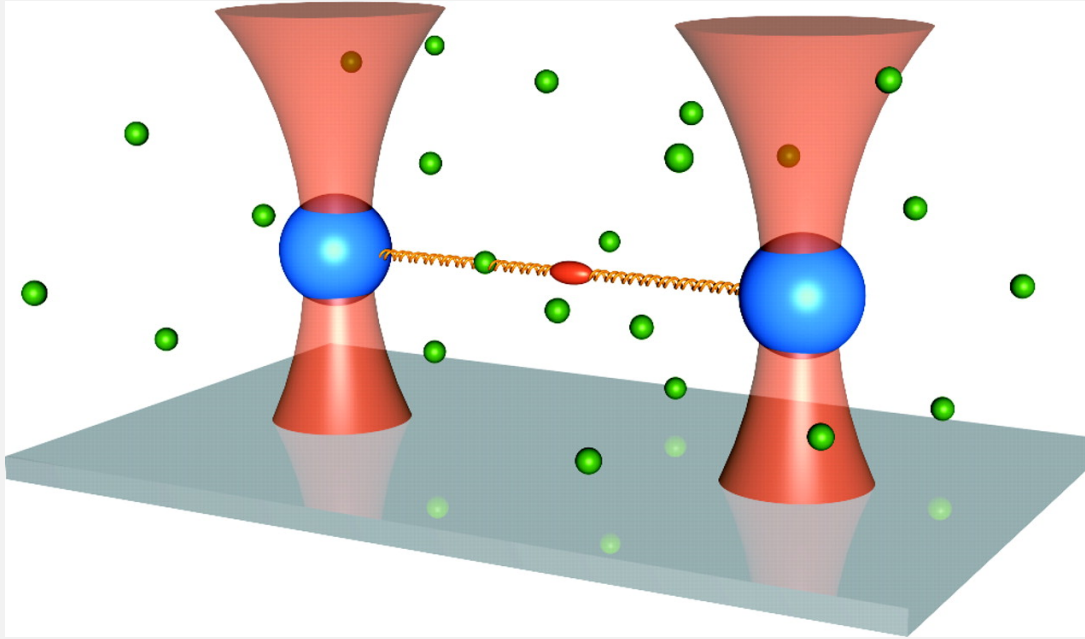
$$\ell_{\text{sl}} = \sqrt{\frac{D_{1d}}{k_{\text{off}}}}$$



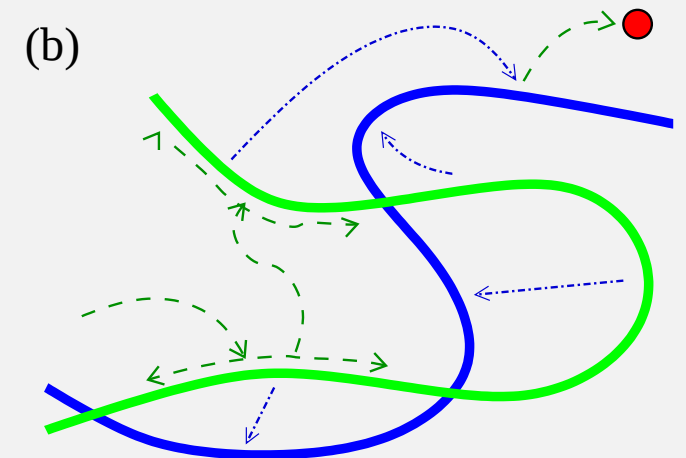
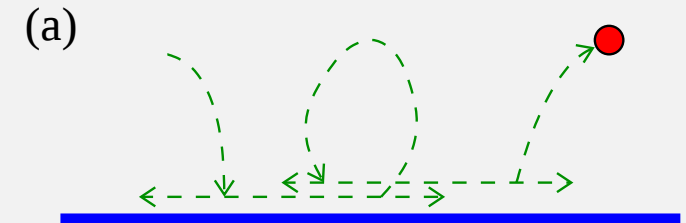
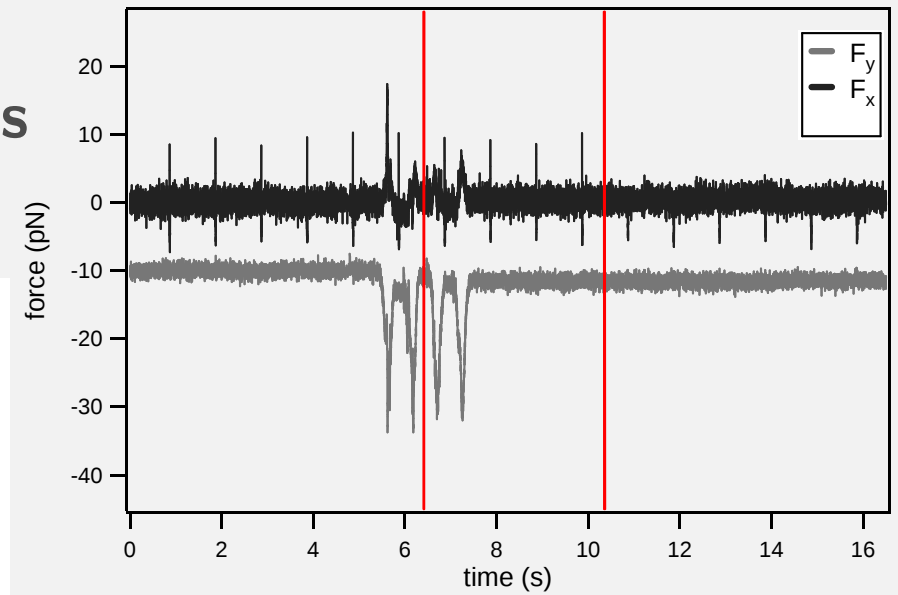
Effective sliding length:

$$\ell_{\text{sl}}^{\text{eff}} = \sqrt{\frac{k_{\text{on}}}{2\pi D_{3d}}} \times \ell_{\text{sl}} \quad \text{microhop correction:} \quad \sqrt{\frac{k_{\text{on}}}{2\pi D_{3d}}}$$

The rôle of DNA conformations



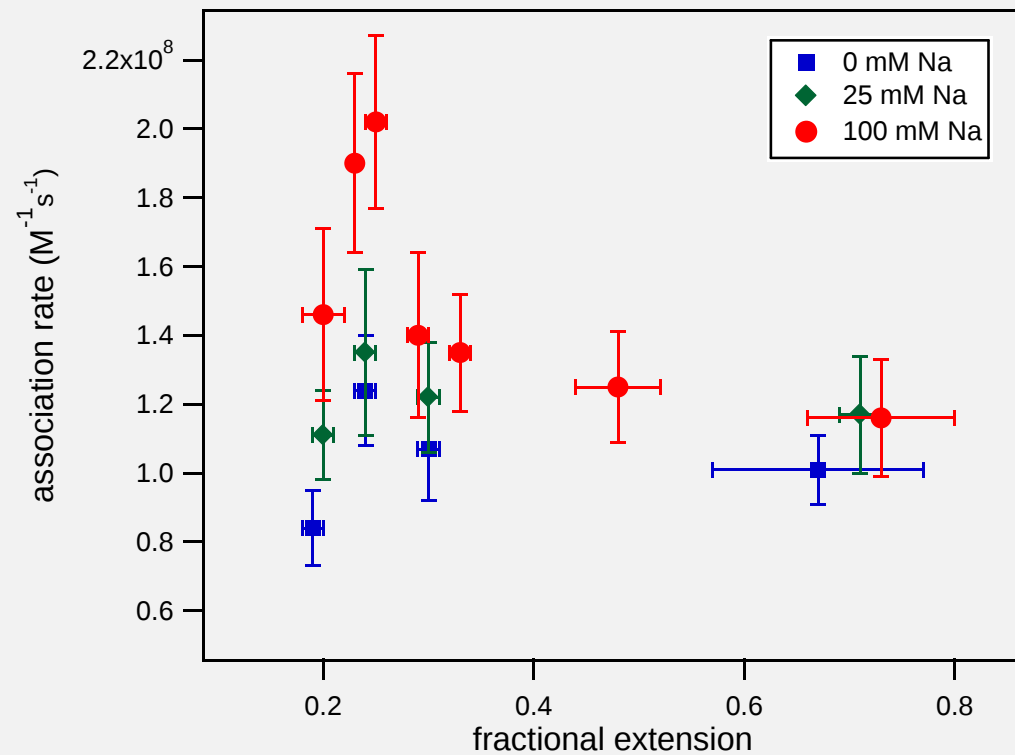
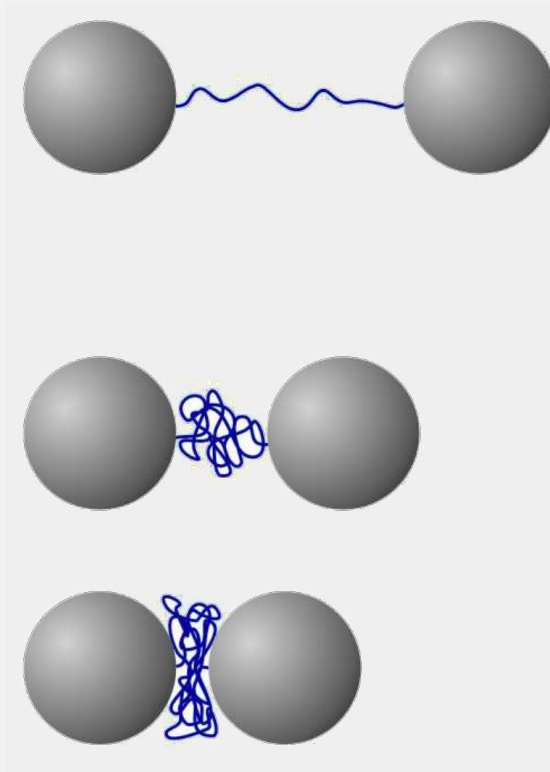
pCco5 plasmid DNA: $6538\text{bp} \approx 2.2\mu\text{m} \approx 45\ell_p$
 [comp λ DNA 48.5kbp]



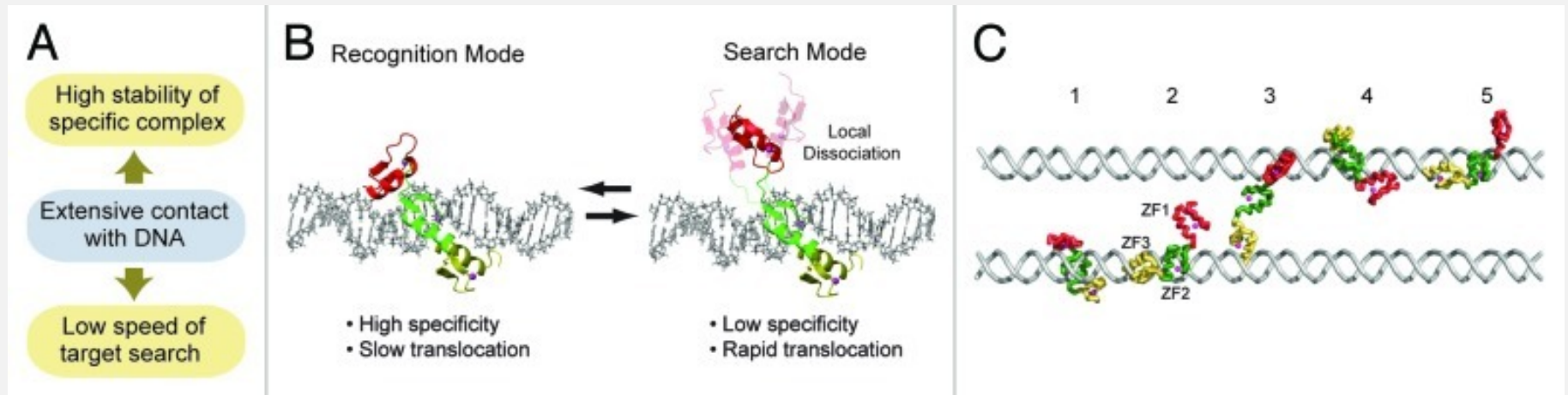
More compact DNA conformations speed up the search

[NaCl]	$k_{\text{on}}^{\text{straight}}$ [Ms]	$l_{\text{sl}}^{\text{eff}}$ [bp]	$1/\sqrt{l_{\text{DNA}}}$ [bp]	ℓ_p [bp]	R_{theory}	R_{measured}
0 mM	0.8×10^8	195	518	188	1.18	1.3 ± 0.2
25 mM	1.0×10^8	250	485	175	1.23	1.1 ± 0.2
100 mM	1.0×10^8	250	150	159	1.67	1.7 ± 0.3
150 mM	0.9×10^9	15.5	120	153	1.15	1.3 ± 0.4

$R = k_{\text{on}}^{\text{max}} / k_{\text{on}}^{\text{straight}}$: enhancement ratio of attachment rates @ max and straight configuration)



Speed-stability paradox in TF search along DNA

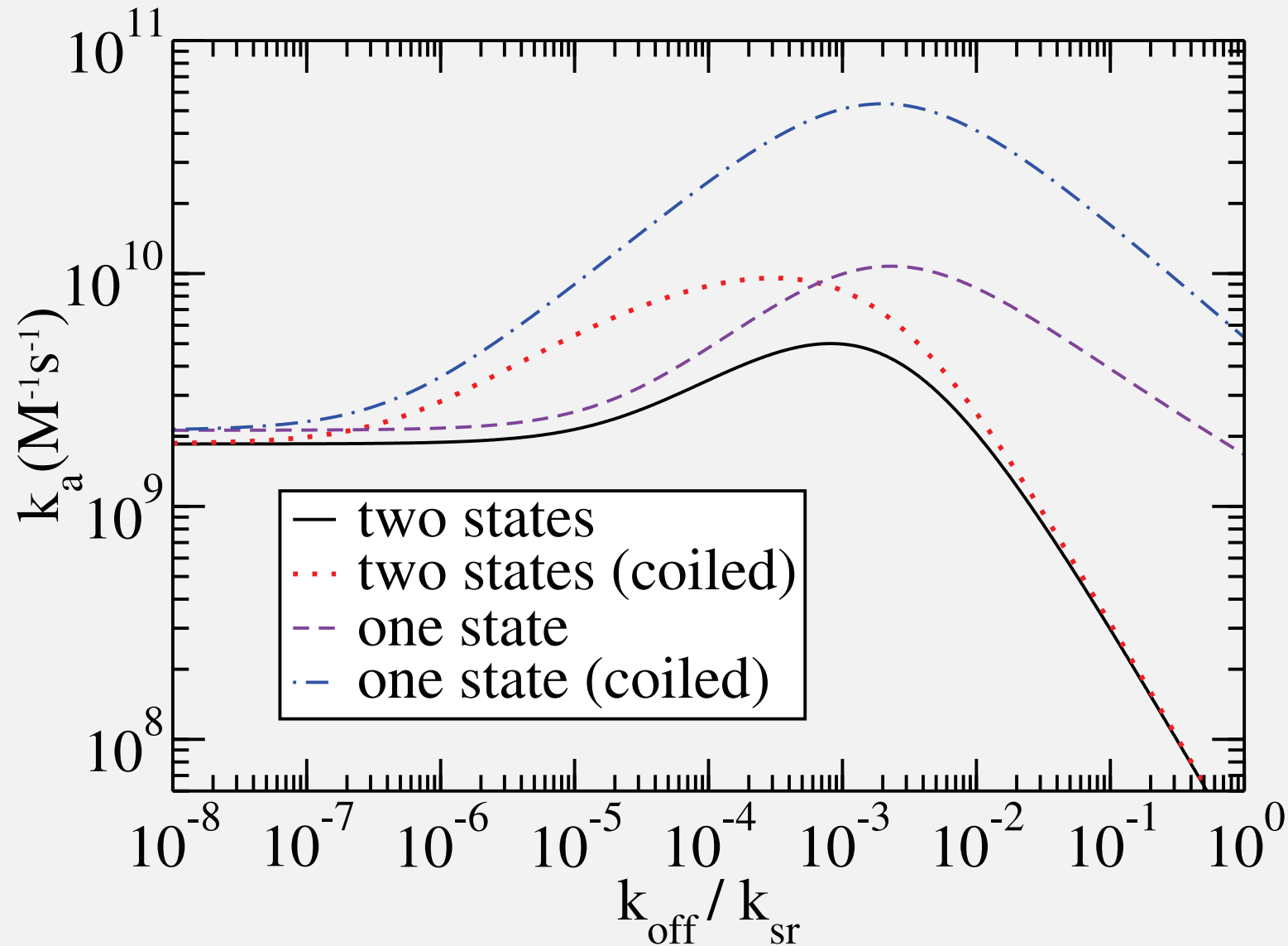


From simulations:

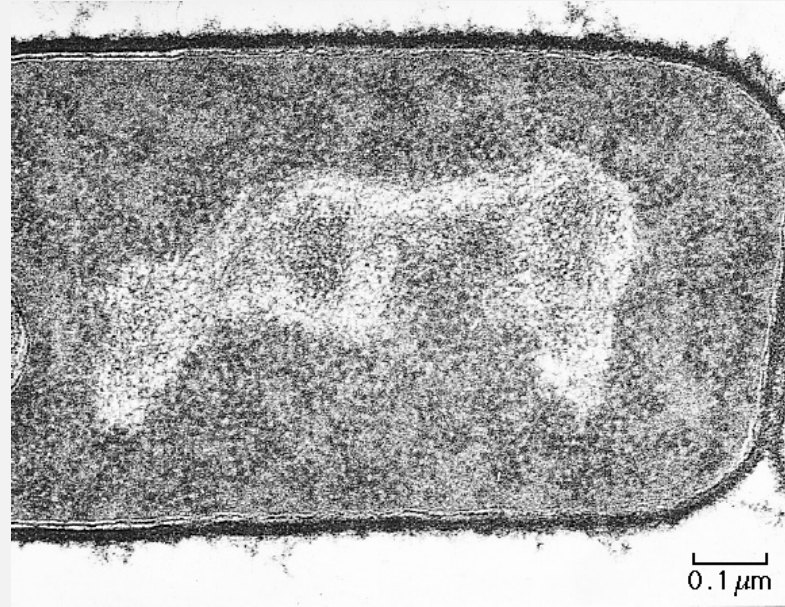
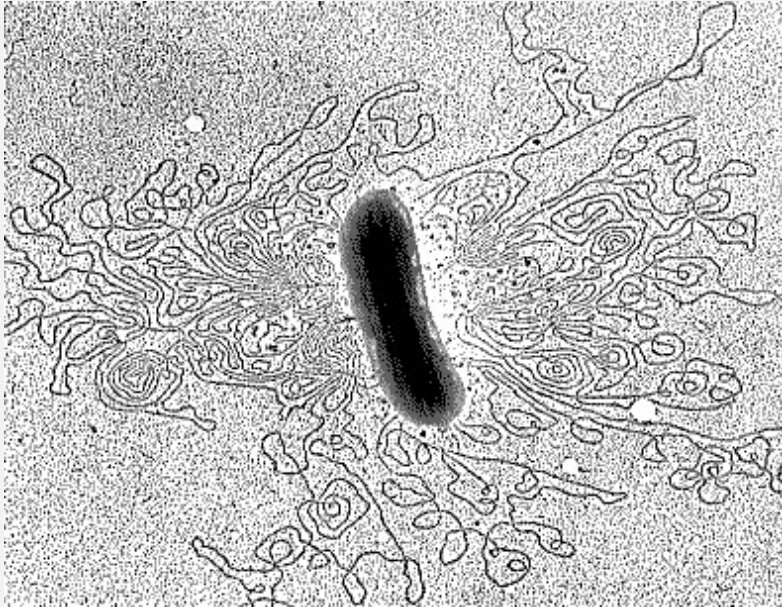
B: Search & recognition modes for a zinc finger protein

C: Intersegmental transfer of the protein

Facilitated diffusion: rate with search & recognition states

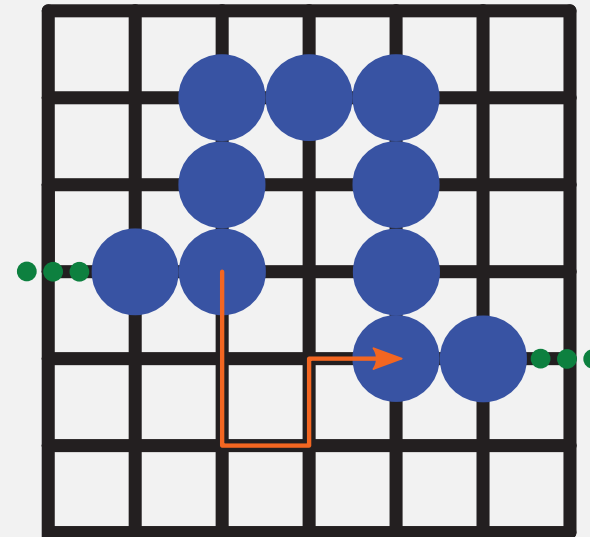


In vivo bacterial gene regulation: E.coli



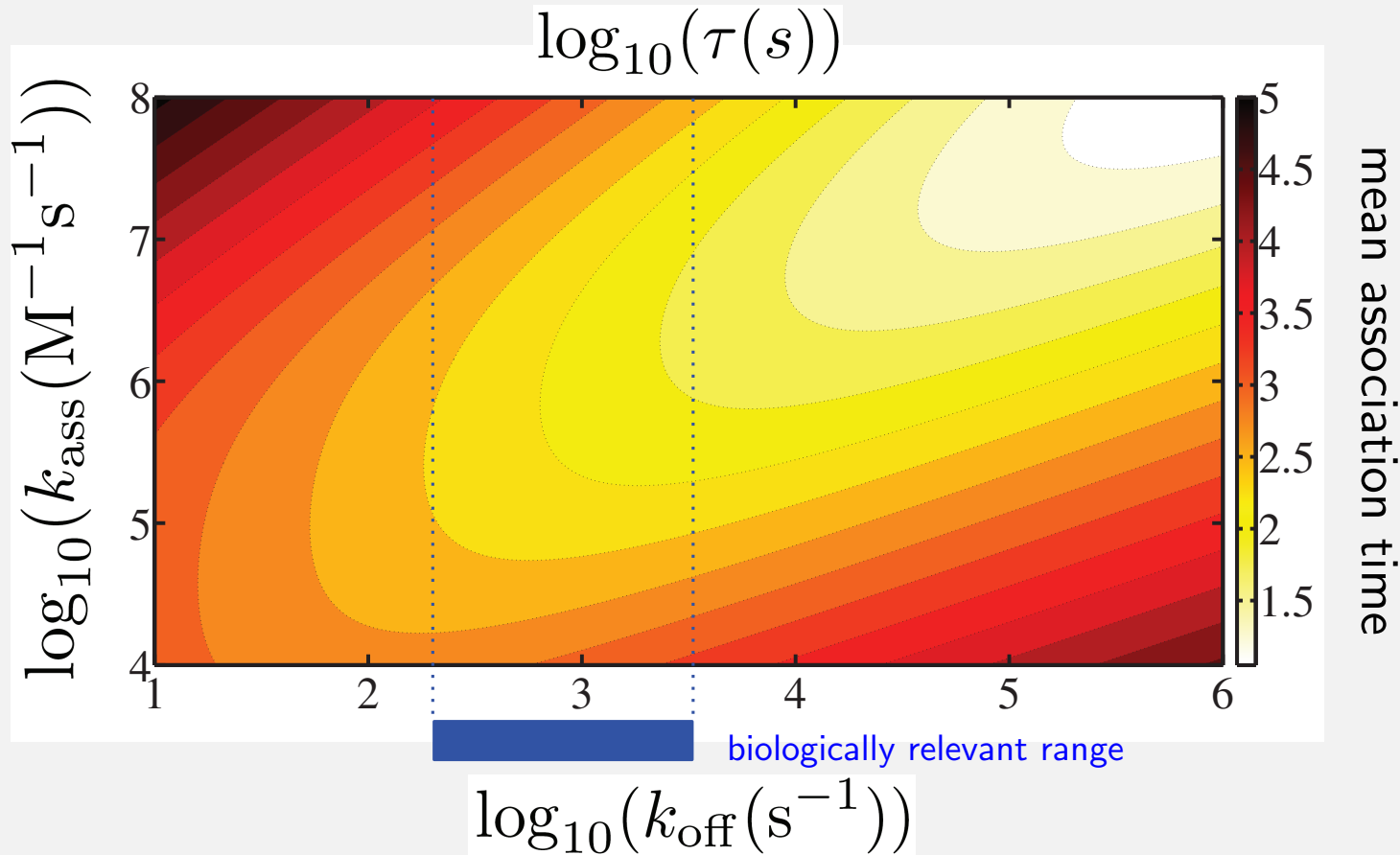
Chromosome is approx an SAW

[M Buenemann & P Lenz, PLoS ONE (2010)]



M Bauer & RM, PLoS ONE (2013)

In vivo gene regulation consistent with facilitated diffusion



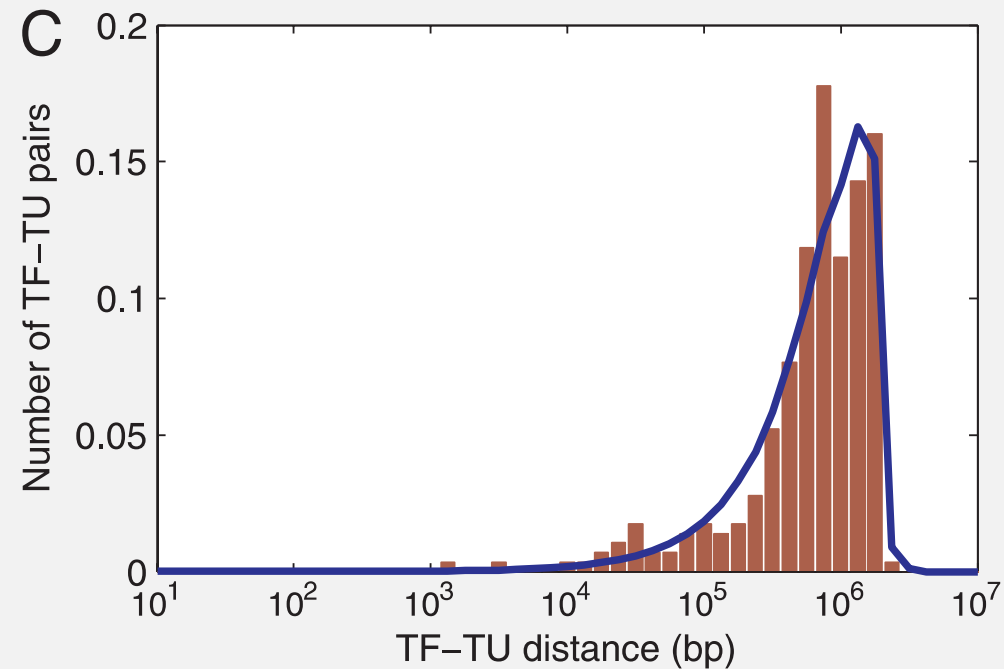
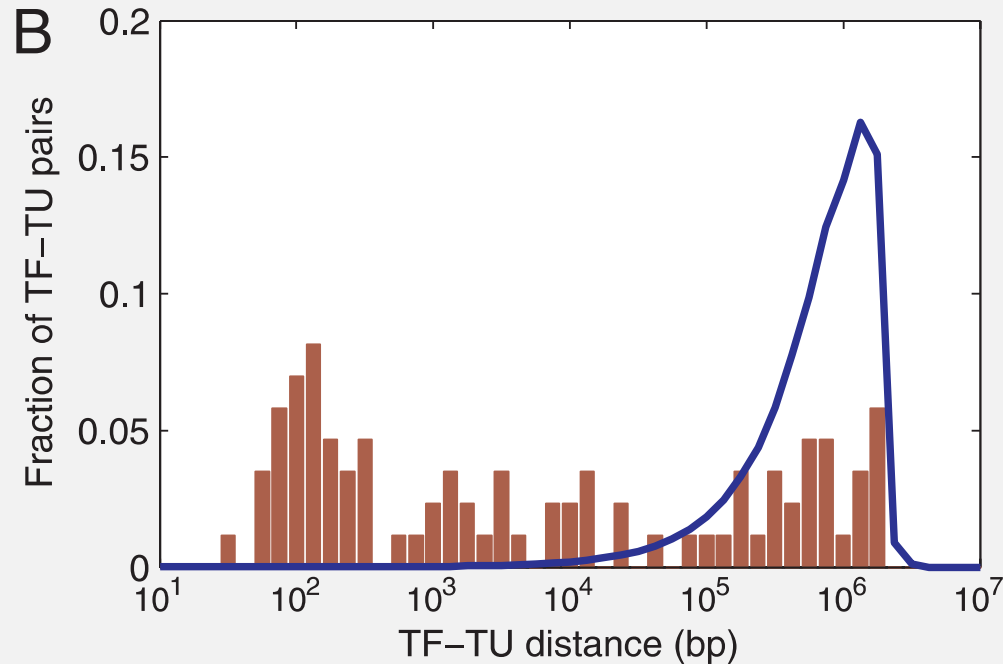
@ optimum the target association time is $\tau \approx 311\text{sec}$ (no fit parameter)

single molecule experiment: $\tau_{\text{exp}} = 354\text{sec}$ [Elf et al, Science (2007)]

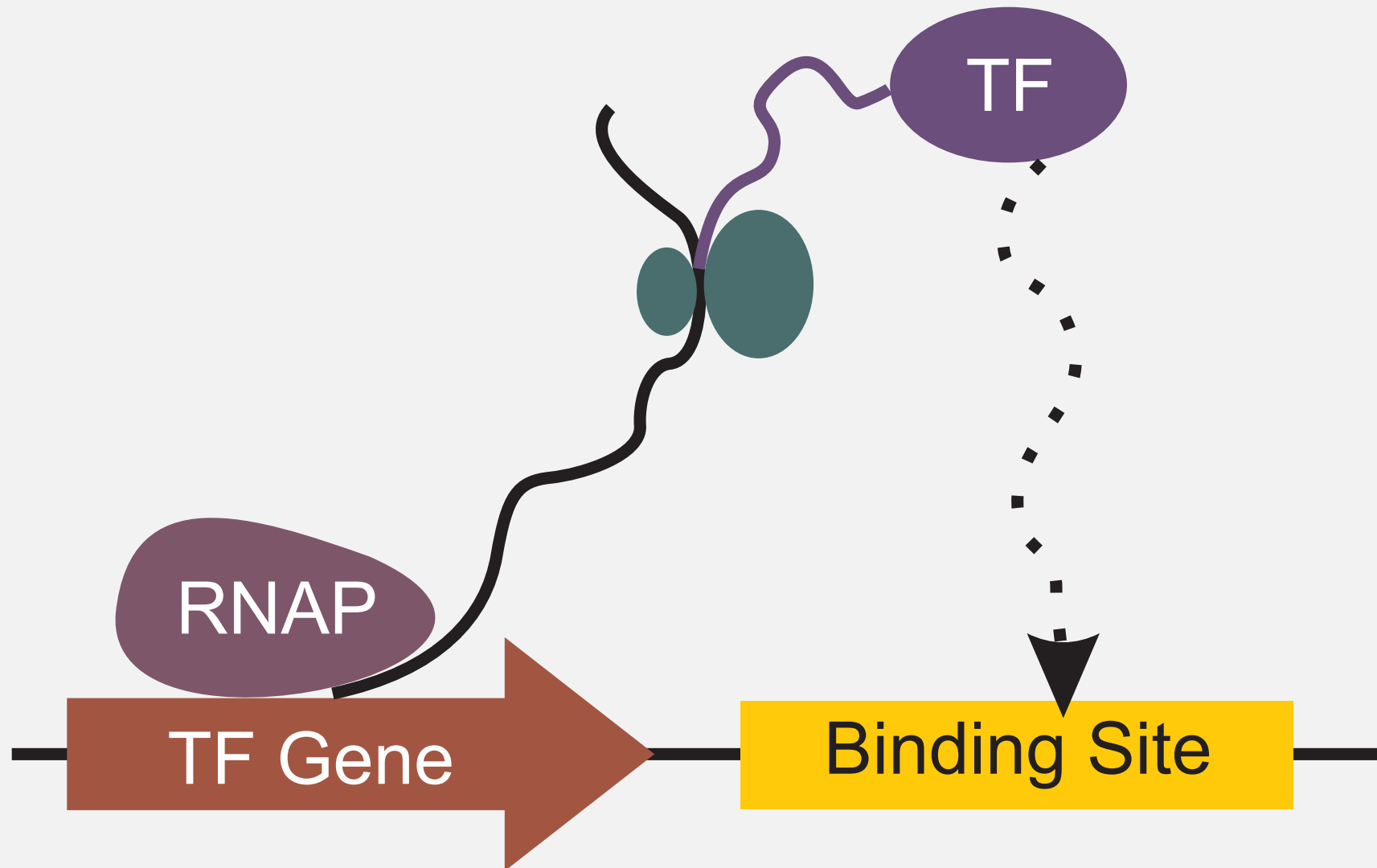
TF regulation effects gene proximity

Does distance between genes interacting via TFs matter?

Gene-gene distance distribution for local TFs (regulate < 4 operons, left) and global (regulate ≥ 4 operons, right). Blue line: random location of genes



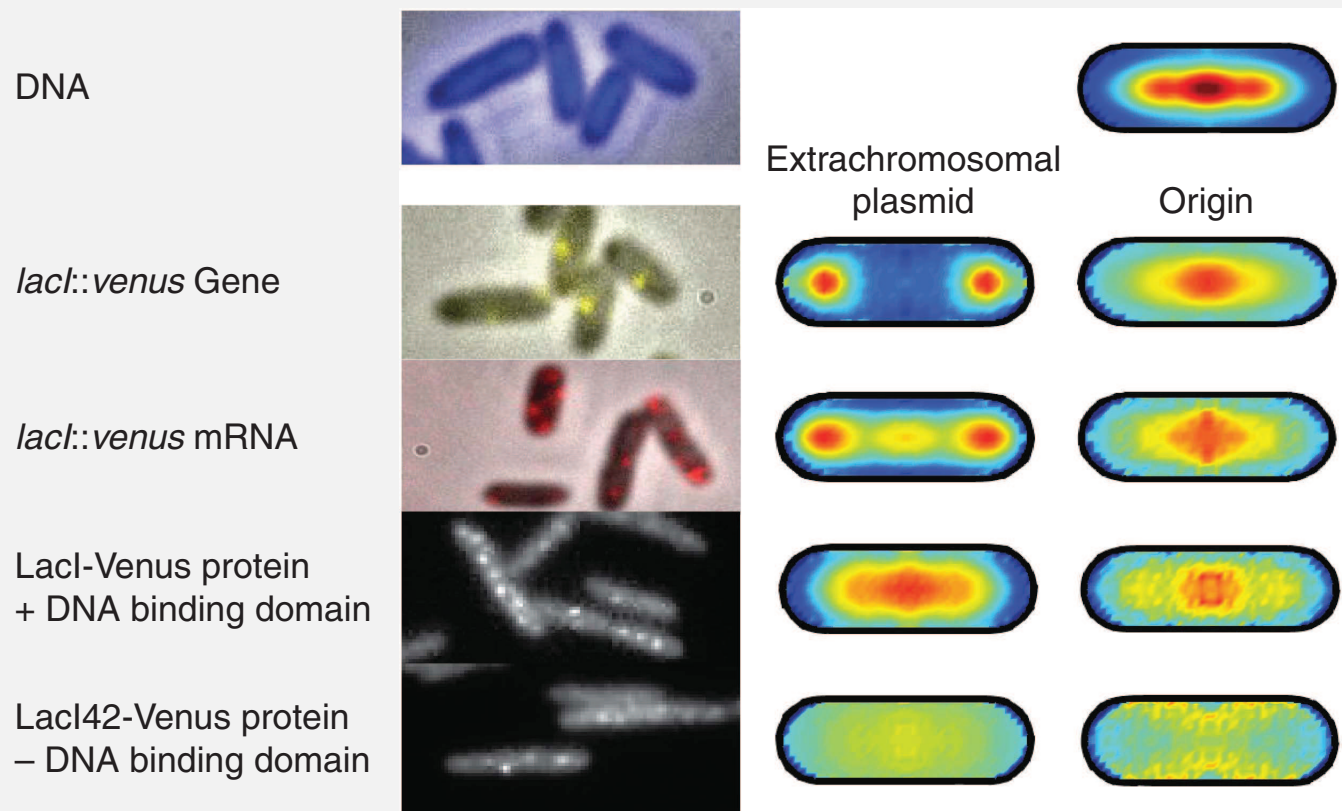
Rapid search hypothesis



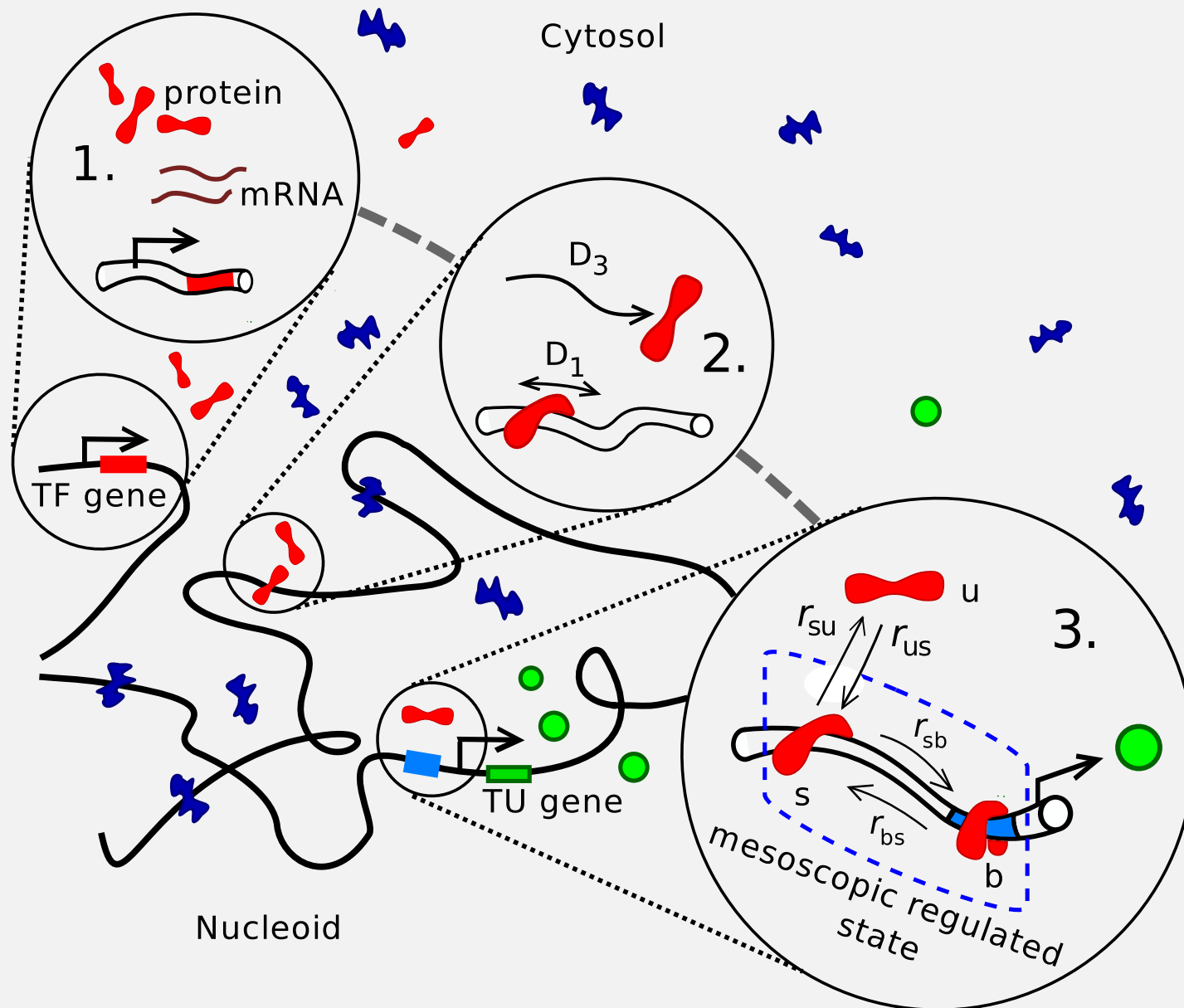
Spatial aspects: do gene locations matter?

Képès: TF targets are typically located next to or at regular distances from the TF gene
→ TF gene-target pairs close in 3D

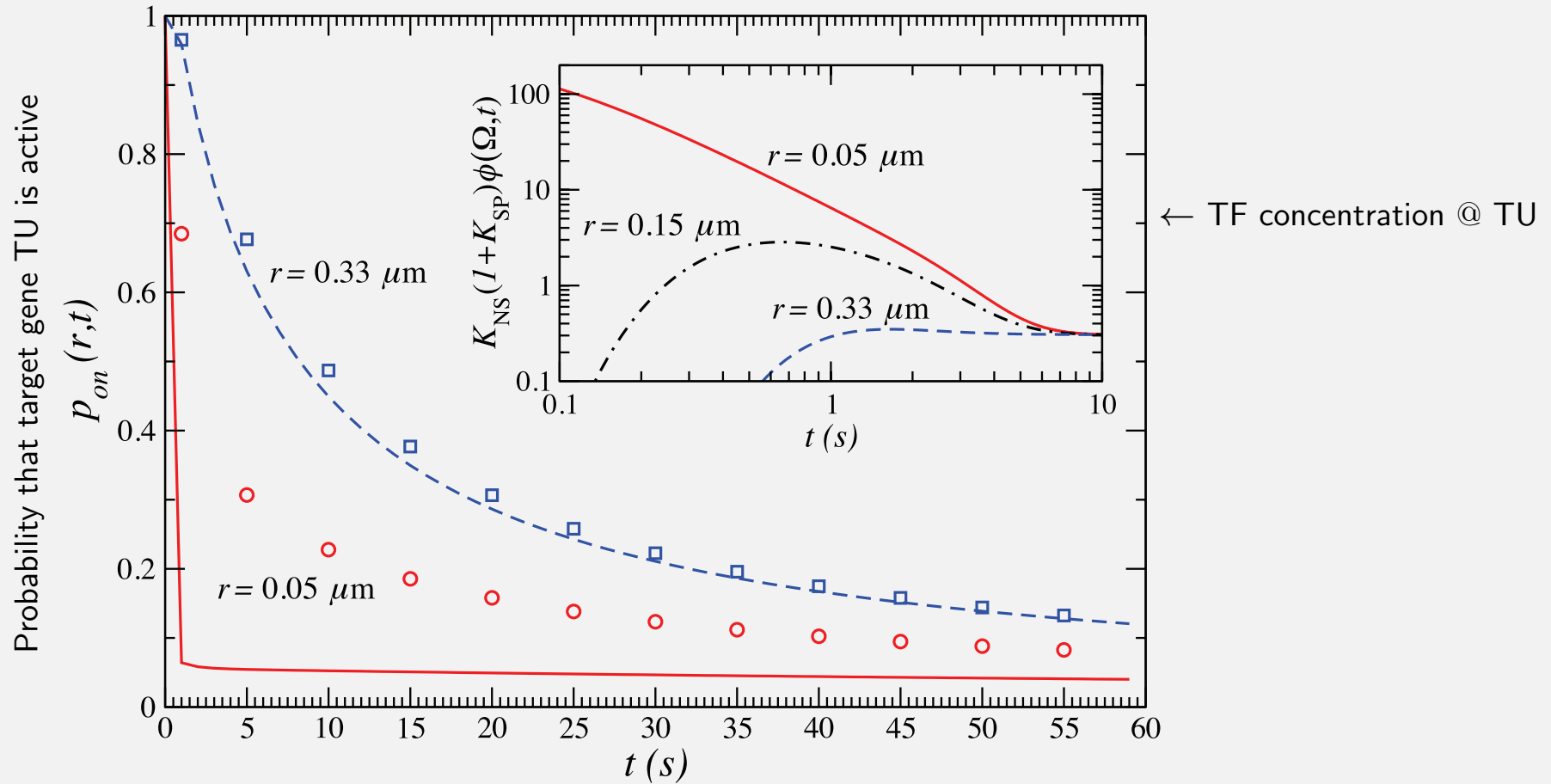
Kuhlman & Cox: • localisation of TF near TF gene • TF distribution highly heterogeneous
• TF gene influences distribution



Transient intracellular signalling is diffusion controlled



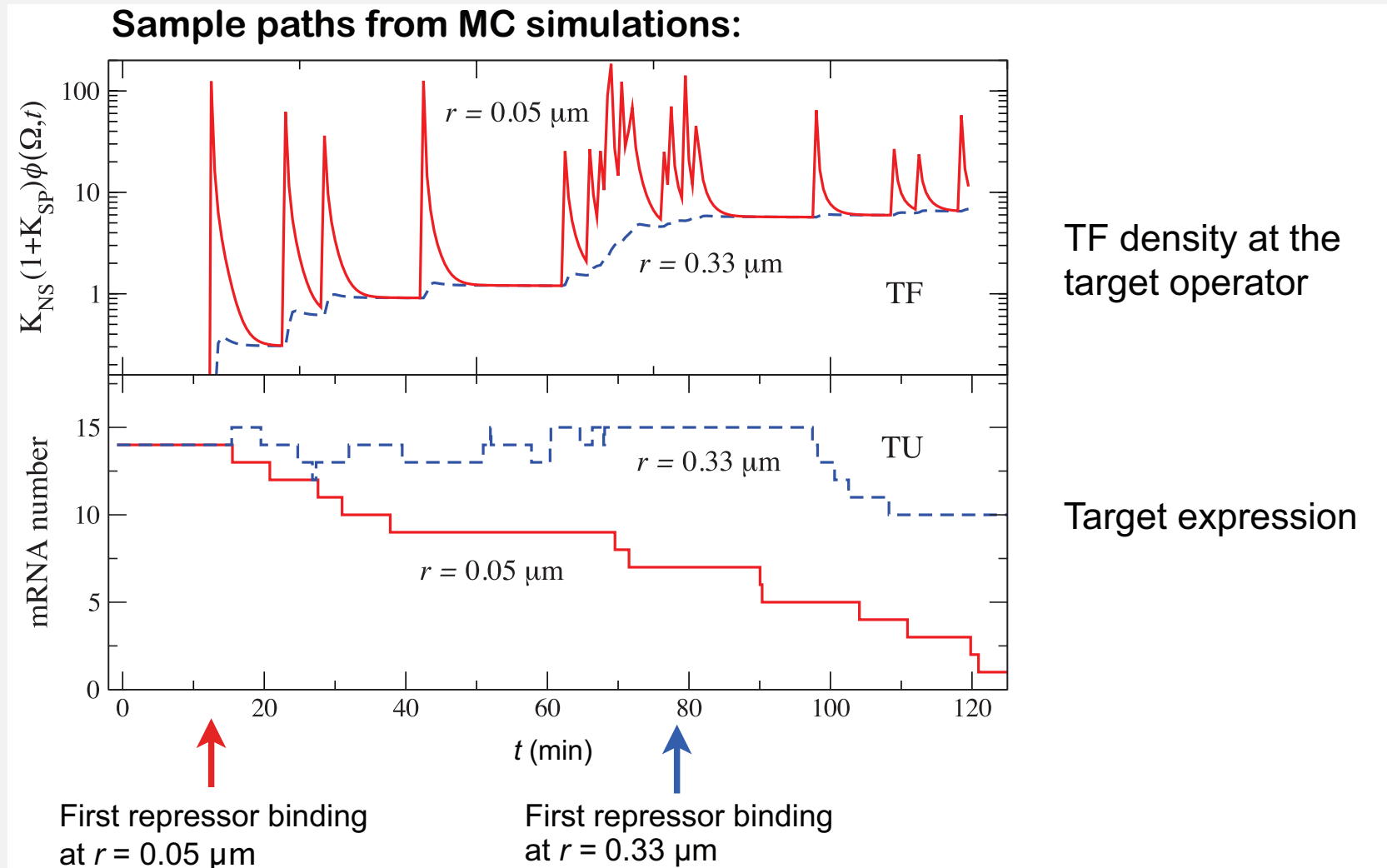
Result 1: transient response to repression



Mean field approximation (full & dashed lines):

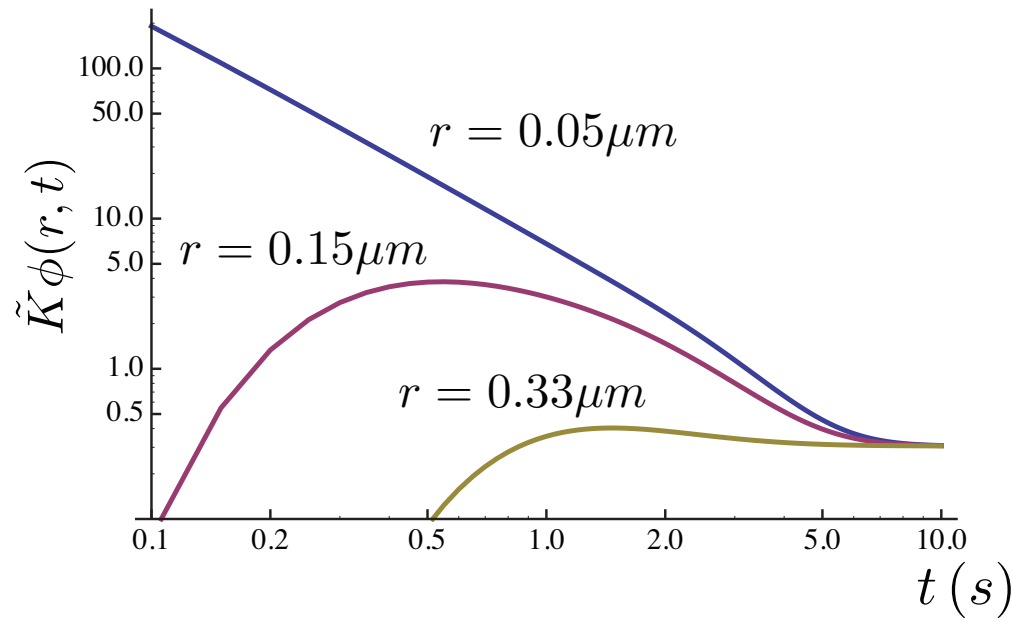
$$p_{on}(r, t) = \left\langle \frac{1 + K_{NS}\rho_{TF}(r, t)}{1 + \tilde{K}\rho_{TF}(r, t)} \right\rangle \approx \frac{1 + K_{NS}\langle\rho_{TF}(r, t)\rangle}{1 + \tilde{K}\langle\rho_{TF}(r, t)\rangle}$$

Result 2: time dependence of gene response

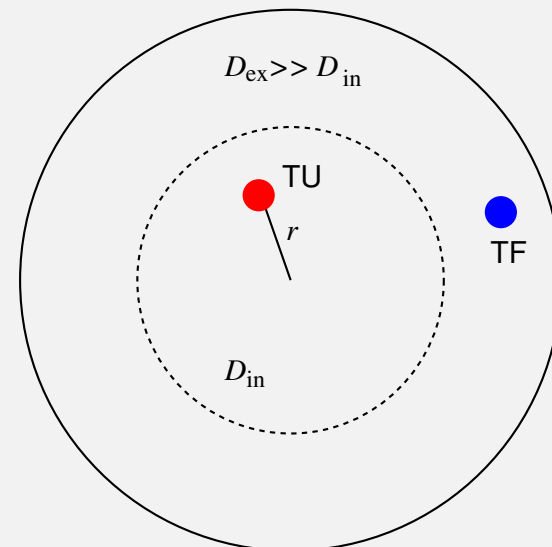
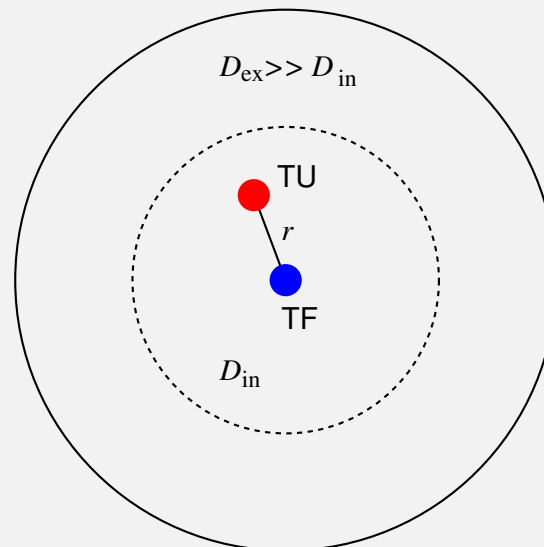
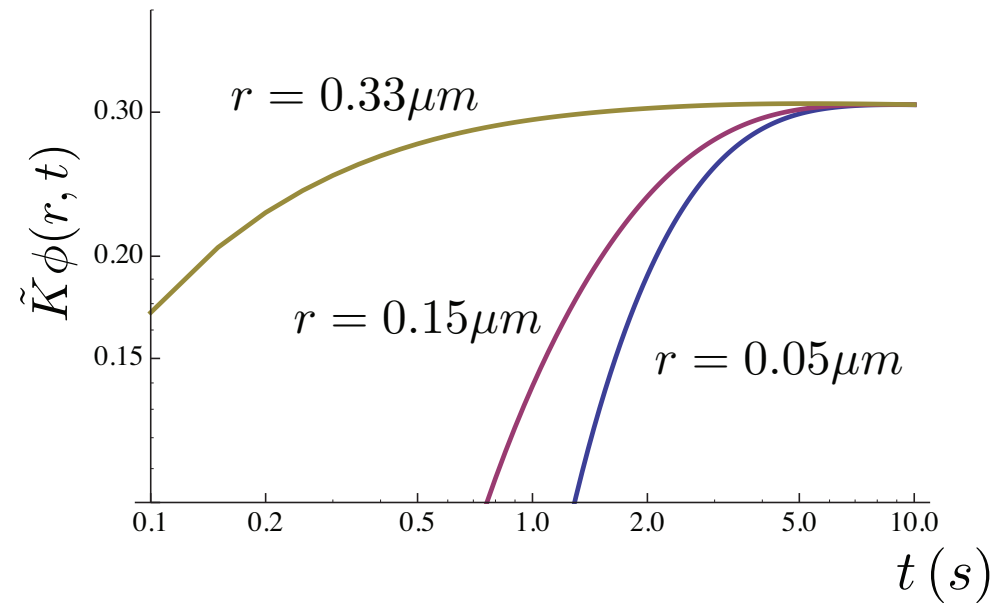


Result 3: gene location matters

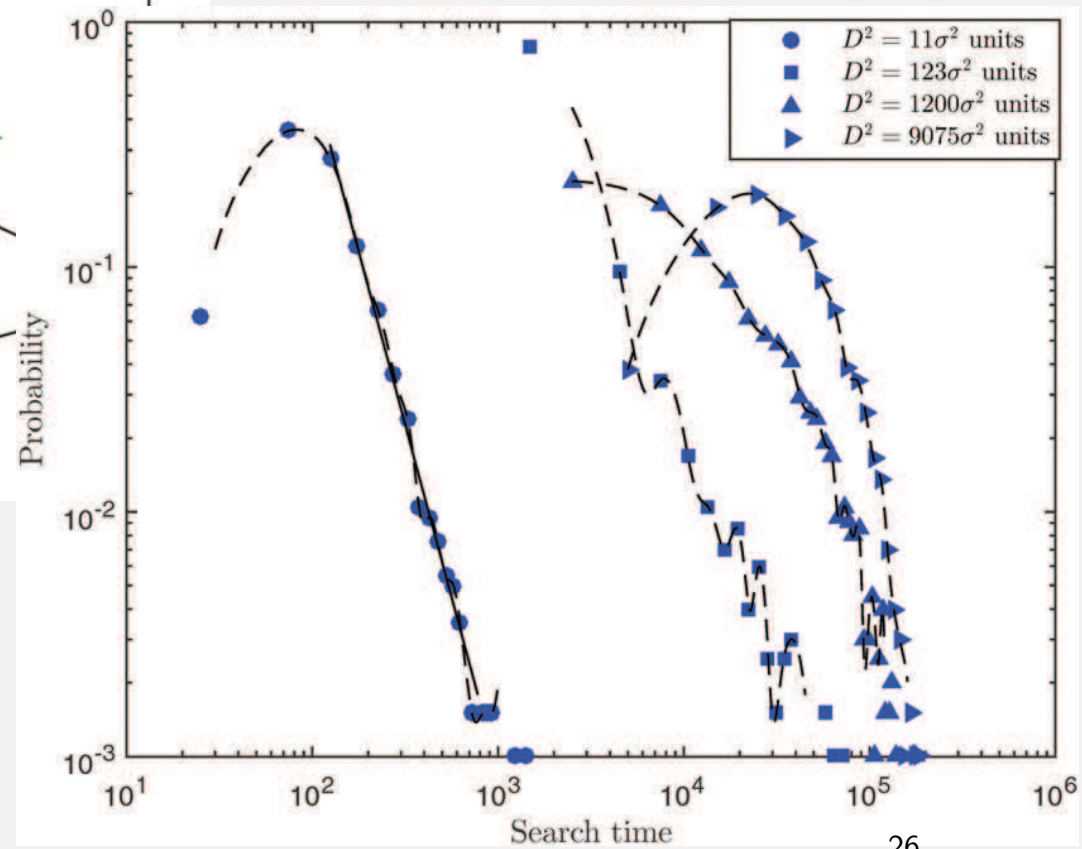
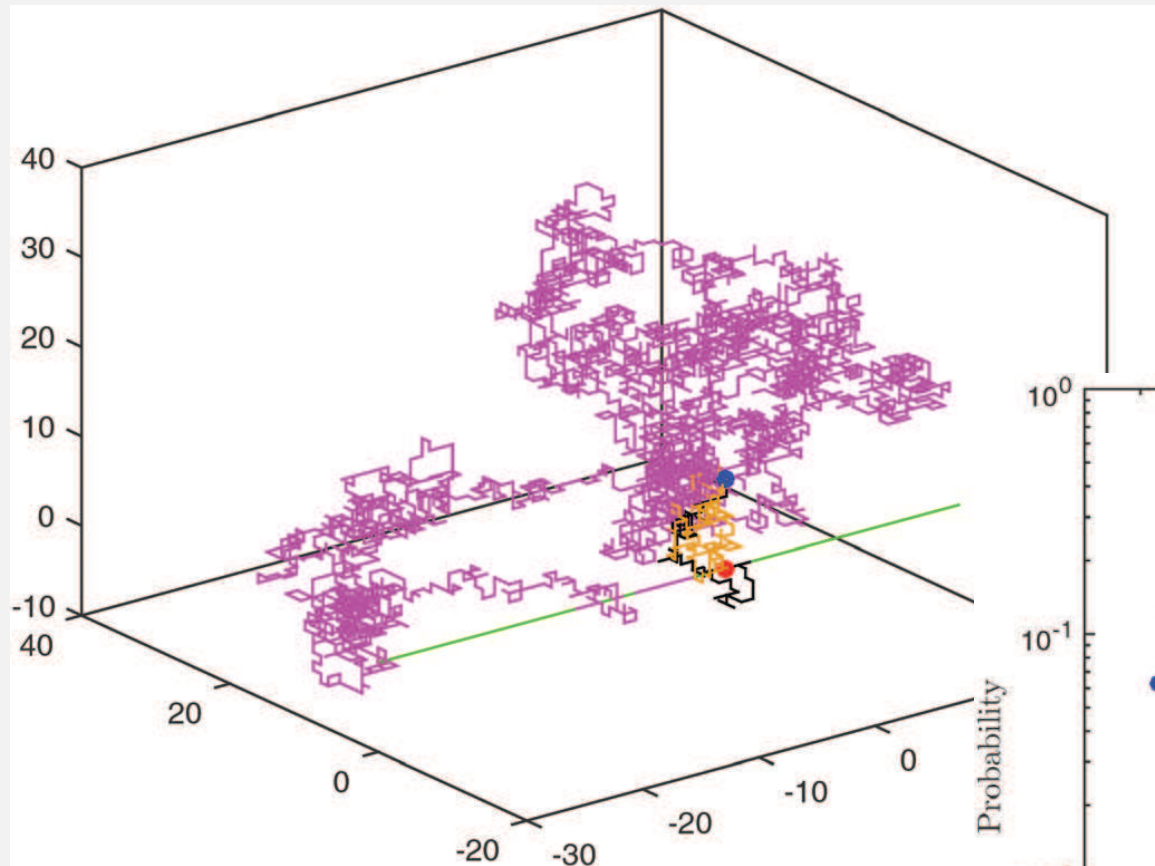
TF gene within the nucleoid



TF gene on a plasmid

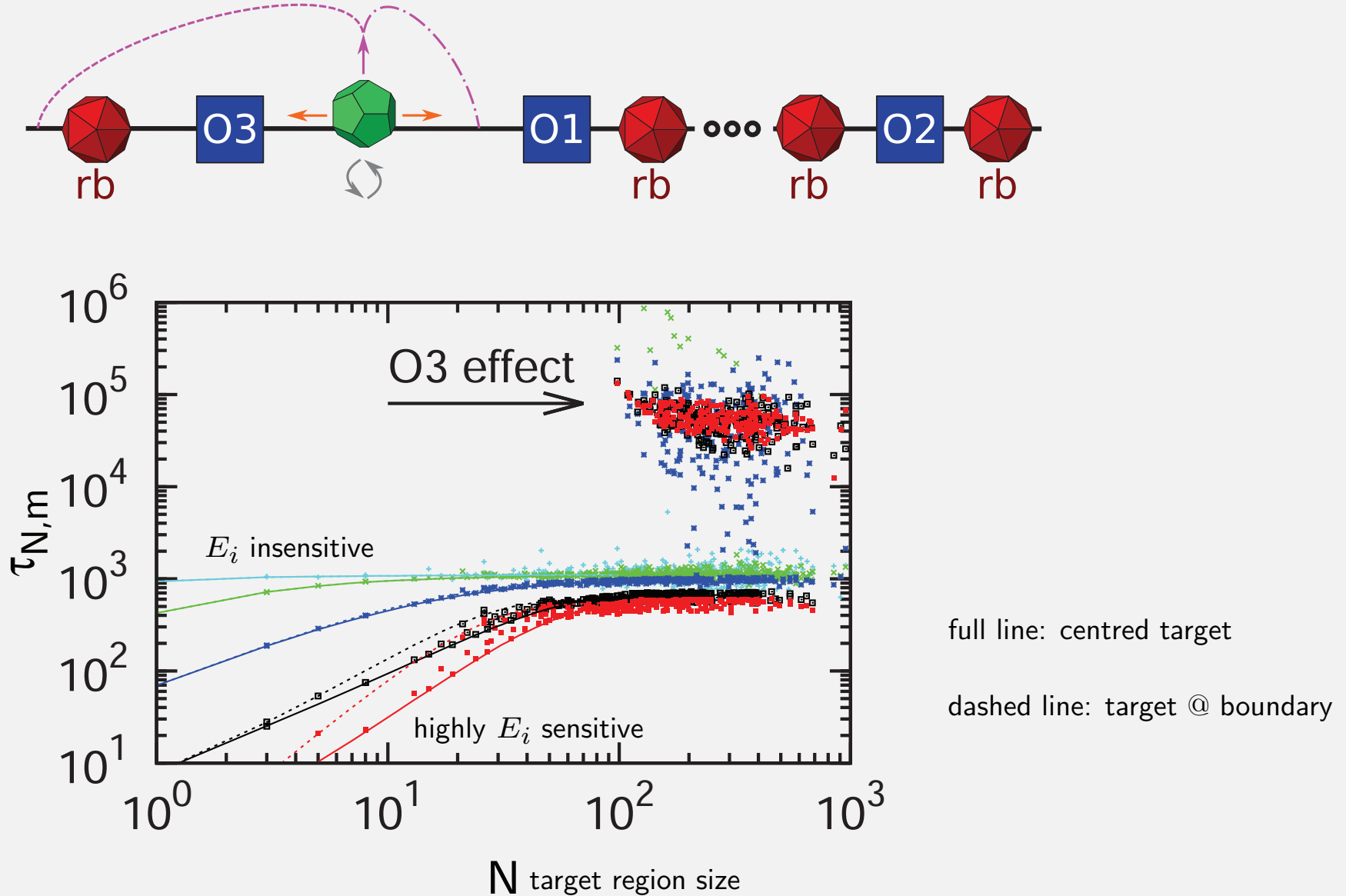


Numerical analysis confirms relevance of proximity effect

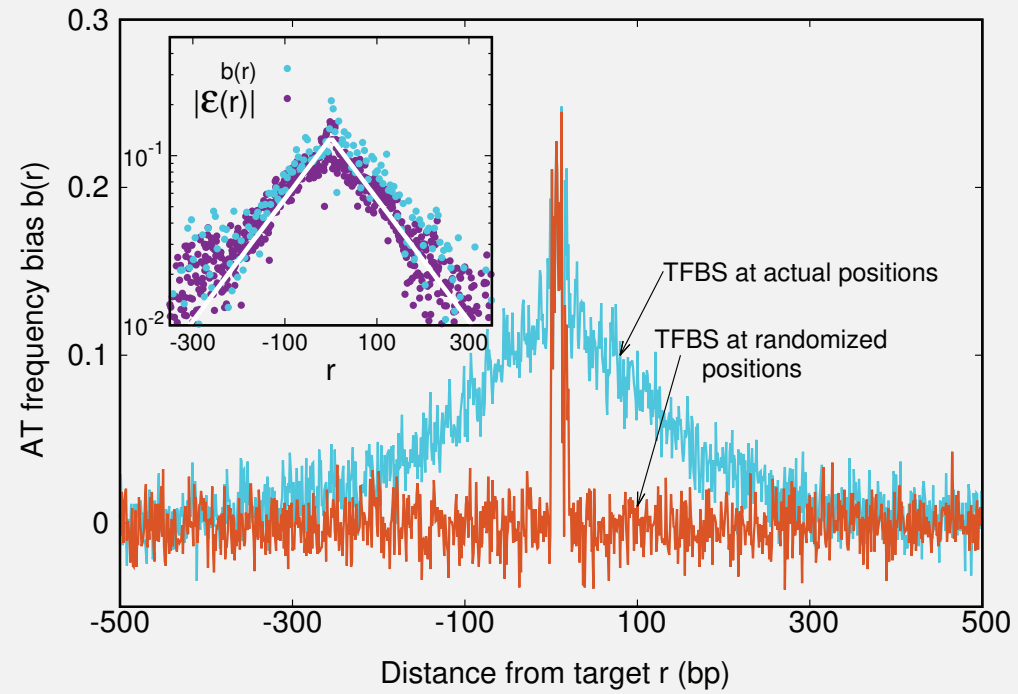
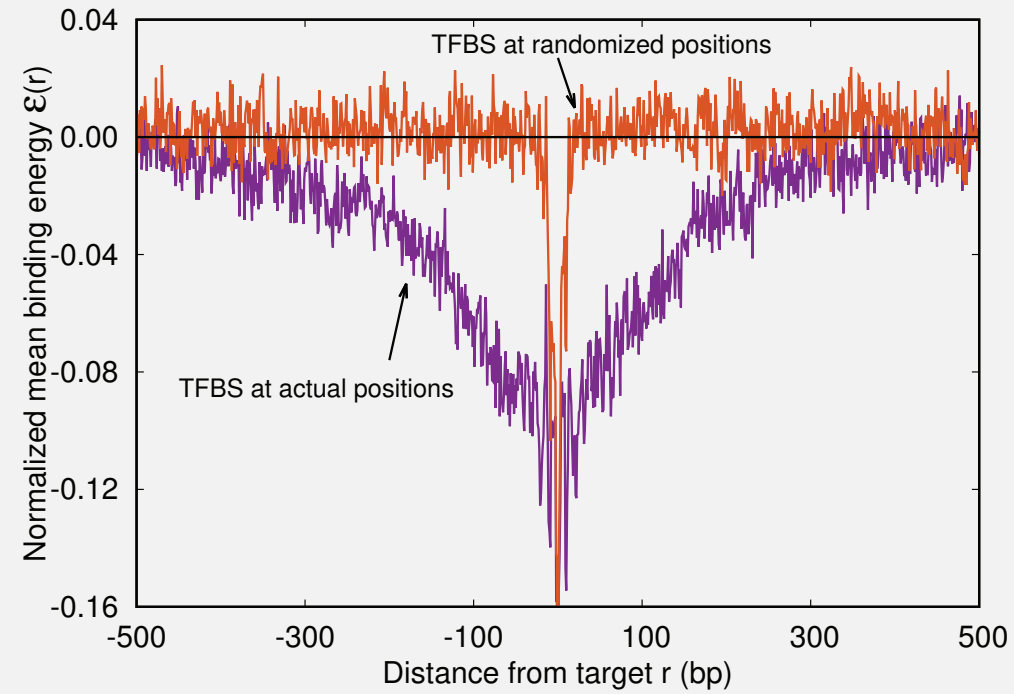


P Kar, AV Cherstvy & RM, PCCP (2018)

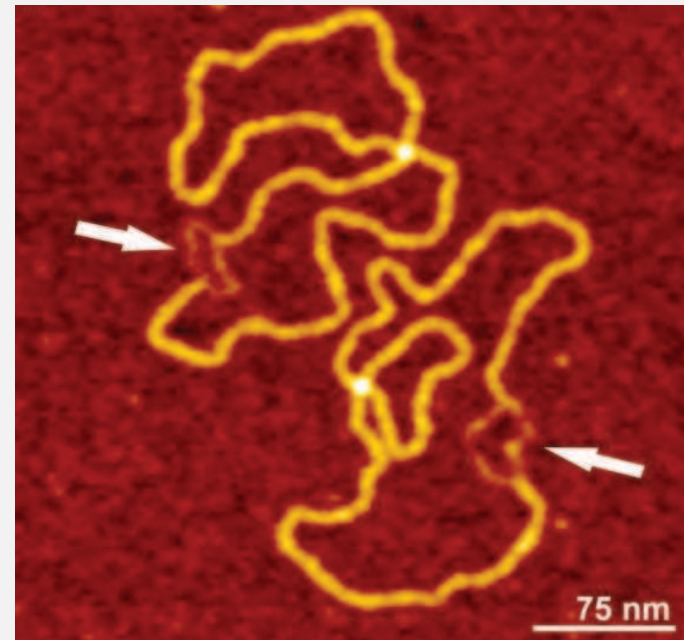
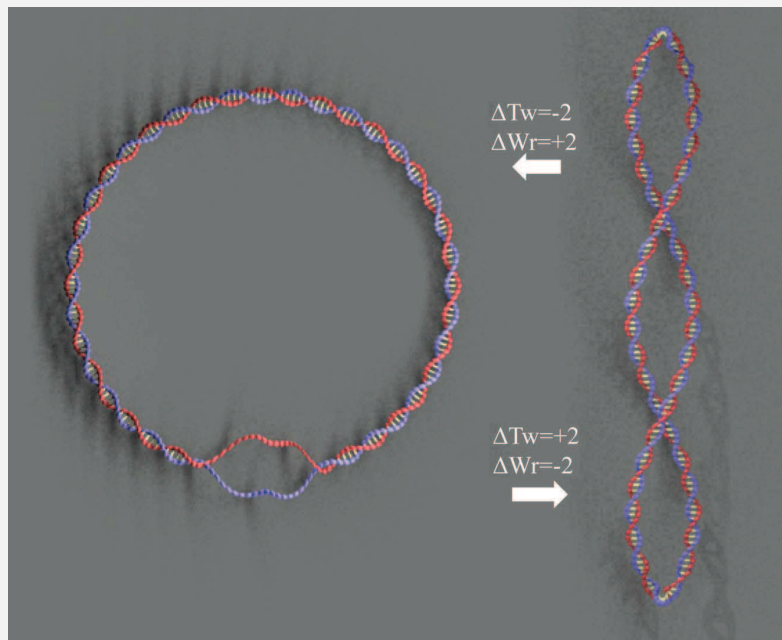
Sequence (binding energy) effects on target search time



Energetic funnel facilitated diffusion

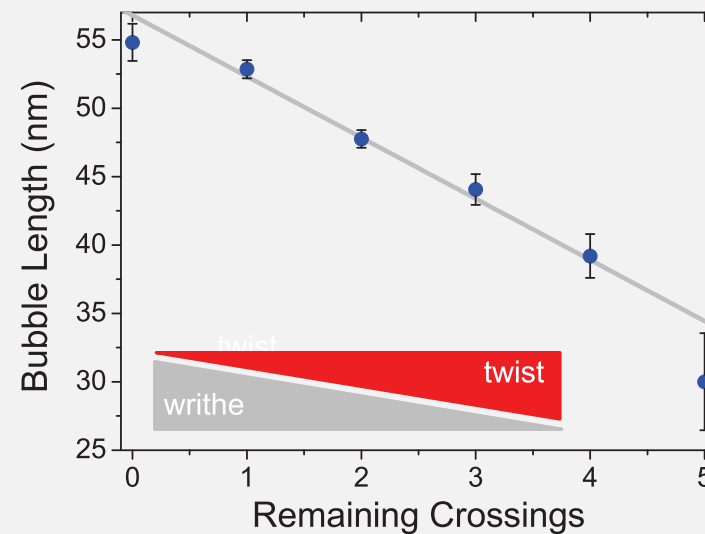


Weak regions at gene starts promote DNA denaturation

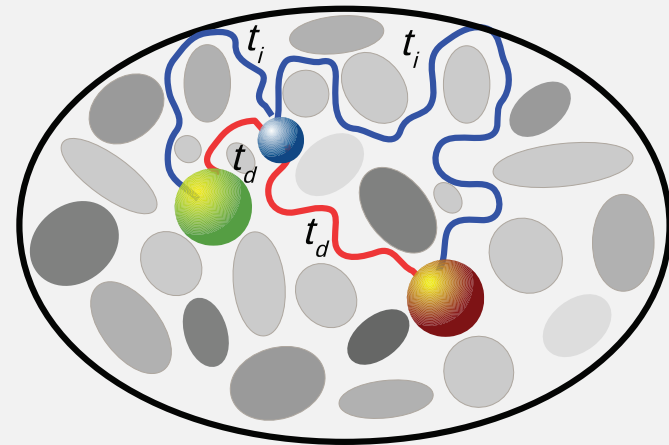
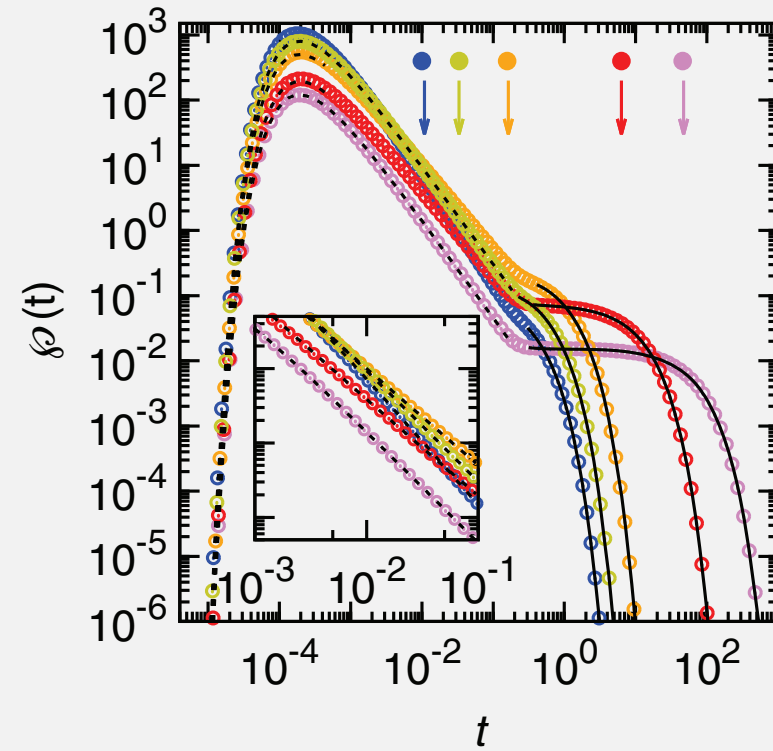
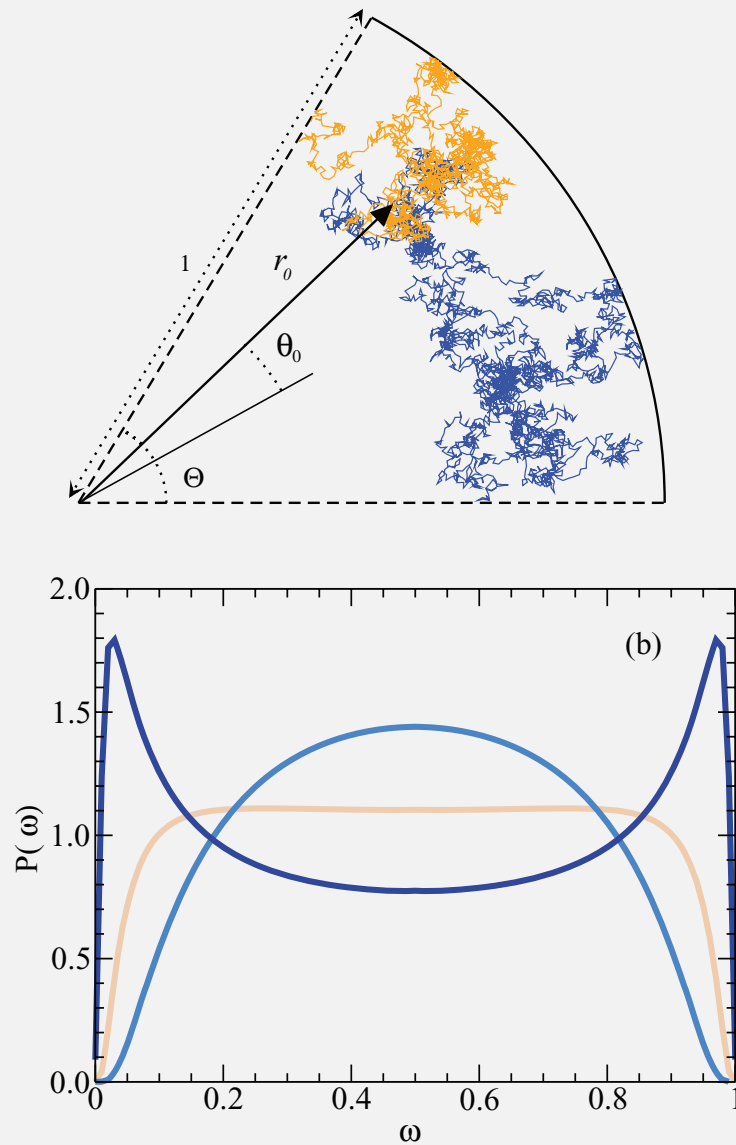


DNA superhelical density:

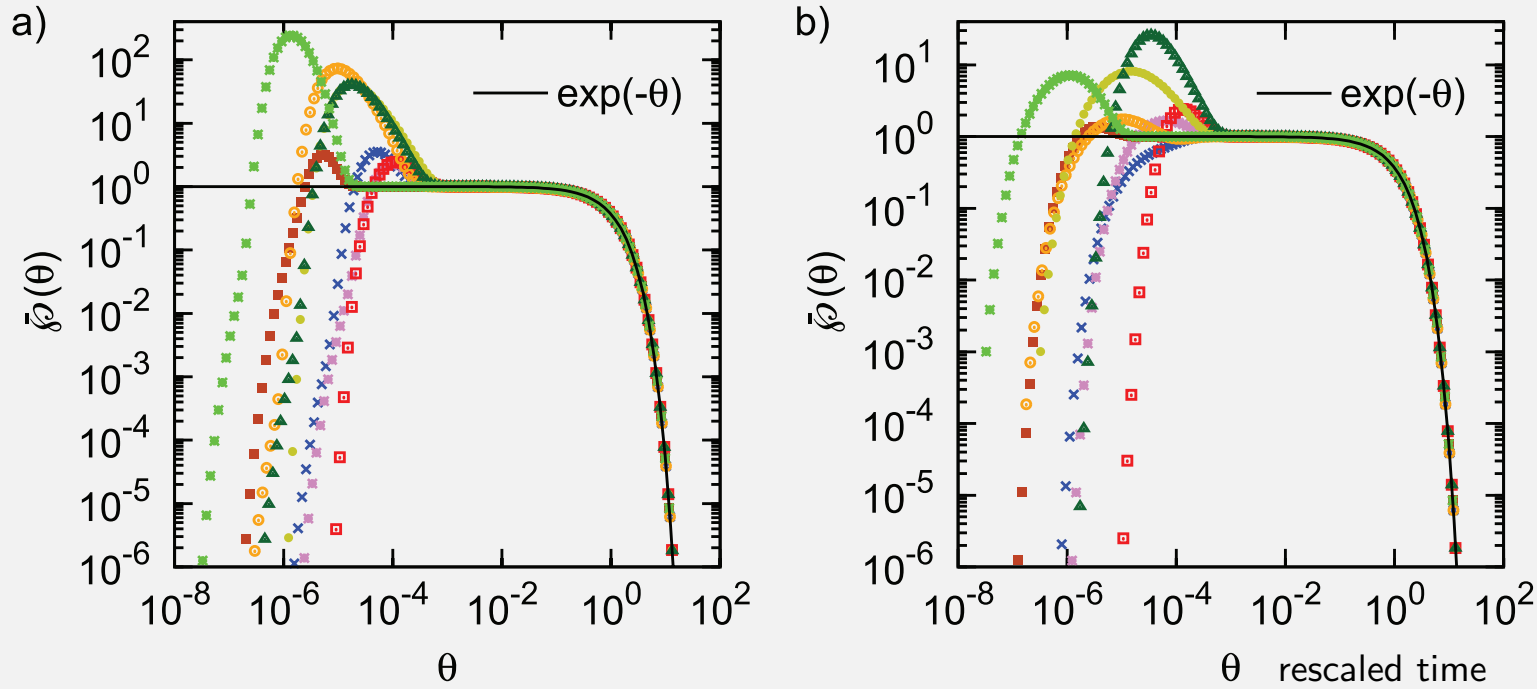
$$\sigma = \frac{Lk - Lk_0}{Lk_0} \approx -0.06$$



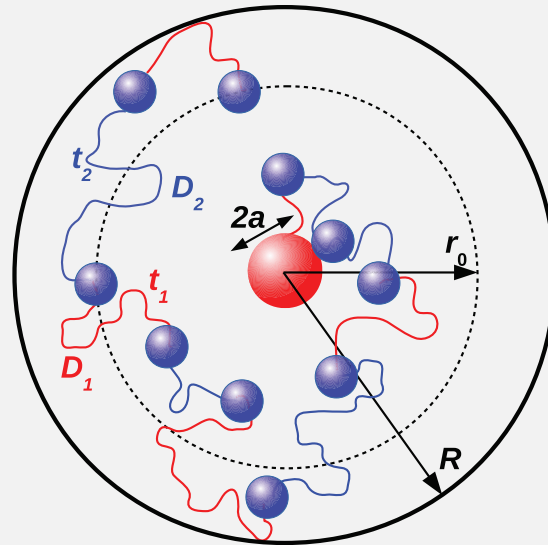
First-past-the-post: few-encounter limit & geometry control



First-past-the-post for 2-channel diffusion



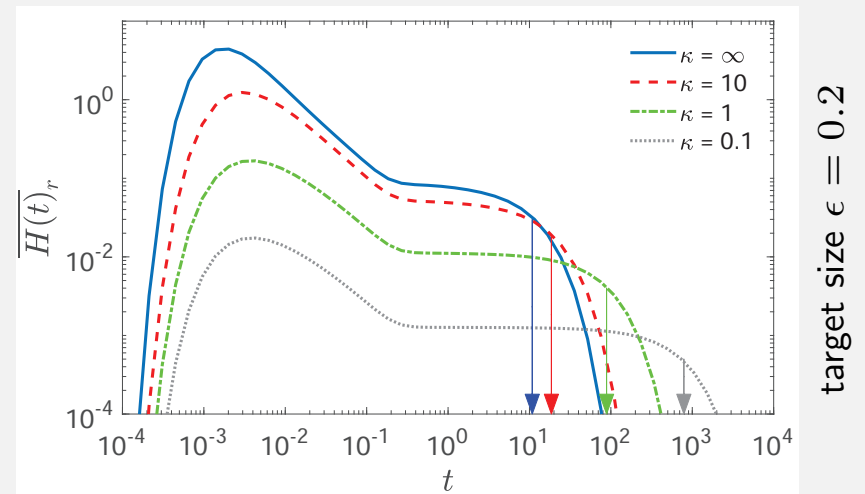
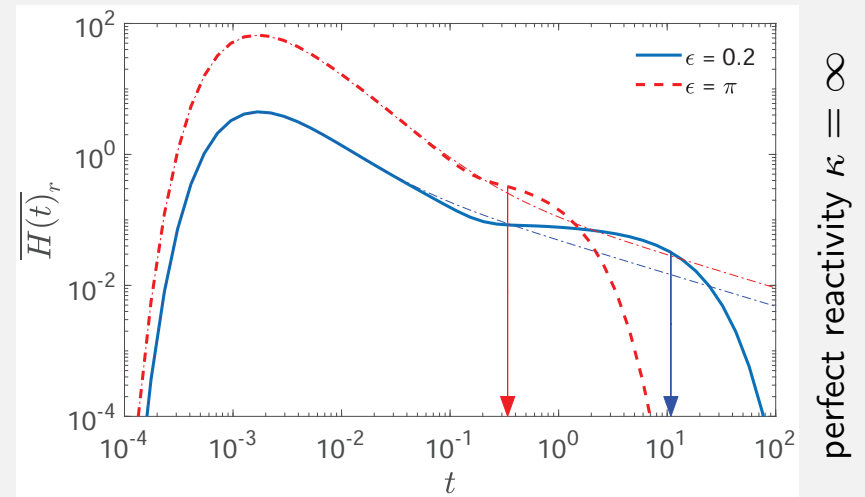
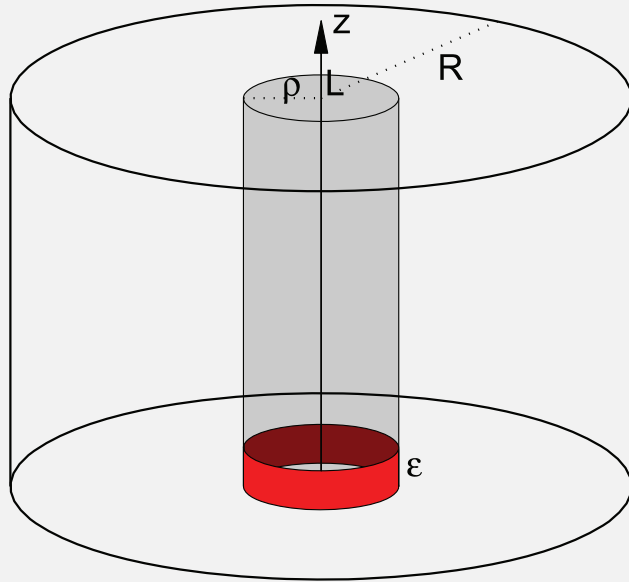
$x_0=0.5$	$k_1=10^1$	$\varphi=10^1$	$z=10^0$	$\times$
$x_0=0.5$	$k_1=10^1$	$\varphi=10^1$	$z=10^1$	$\square$
$x_0=0.5$	$k_1=10^1$	$\varphi=0.5$	$z=10^1$	$\square$
$x_0=0.5$	$k_1=10^2$	$\varphi=10^1$	$z=0.1$	$\square$
$x_0=0.2$	$k_1=10^1$	$\varphi=10^1$	$z=10^0$	$\circ$
$x_0=0.2$	$k_1=10^1$	$\varphi=10^1$	$z=10^1$	$\circ$
$x_0=0.2$	$k_1=10^1$	$\varphi=0.5$	$z=10^1$	$\triangle$
$x_0=0.2$	$k_1=10^2$	$\varphi=10^1$	$z=0.1$	$\triangle$



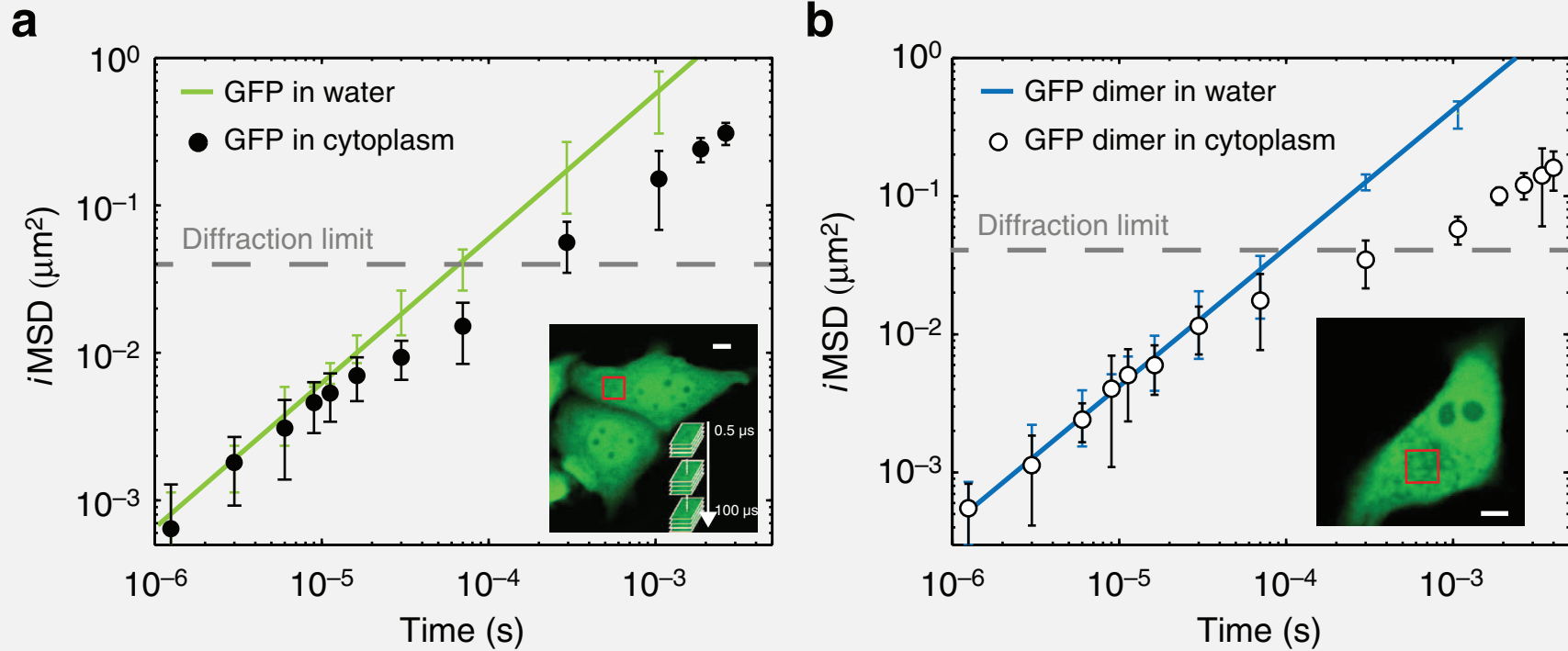
Search mode /w D_1

& recognition mode /w D_2

Few-encounter effect in cylindrical domain /w finite reactivity

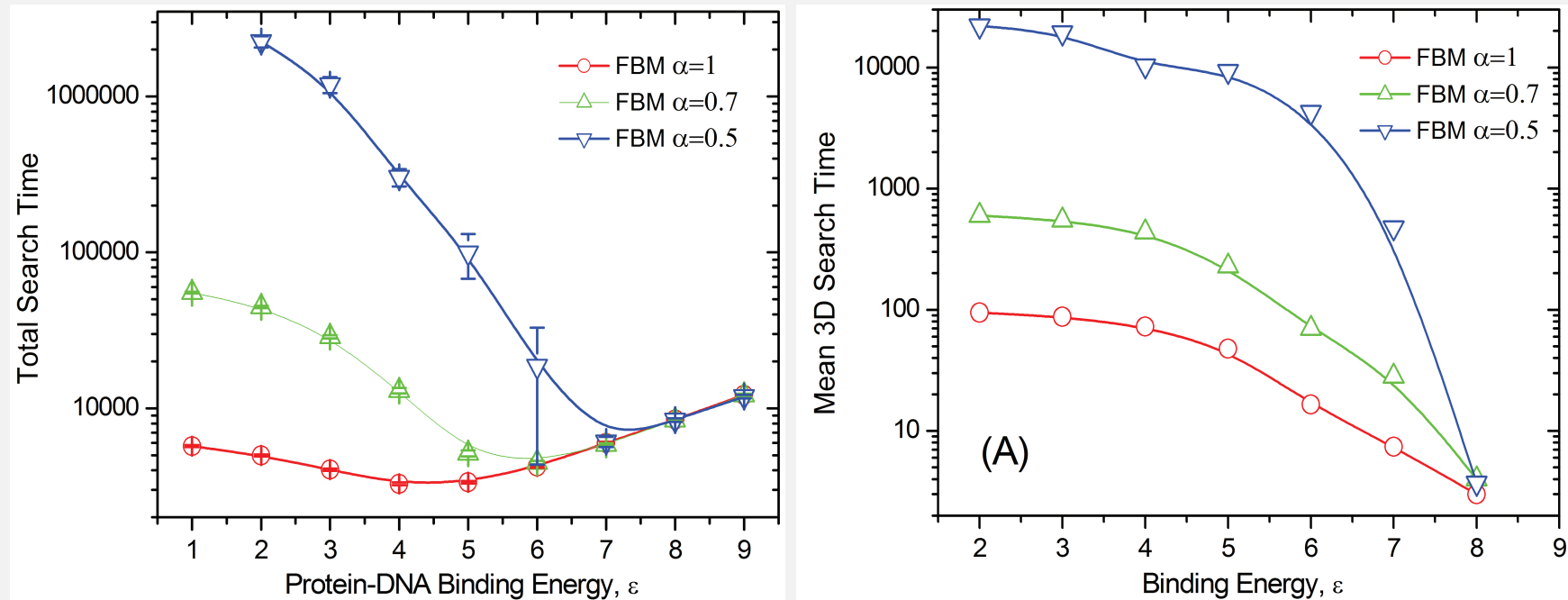


Anomalous diffusion of GFP in cell cytoplasm & nucleus



$$\langle \mathbf{r}^2(t) \rangle \simeq K_\alpha t^\alpha : \text{Subdiffusion when } 0 < \alpha < 1$$

Anomalous facilitated diffusion



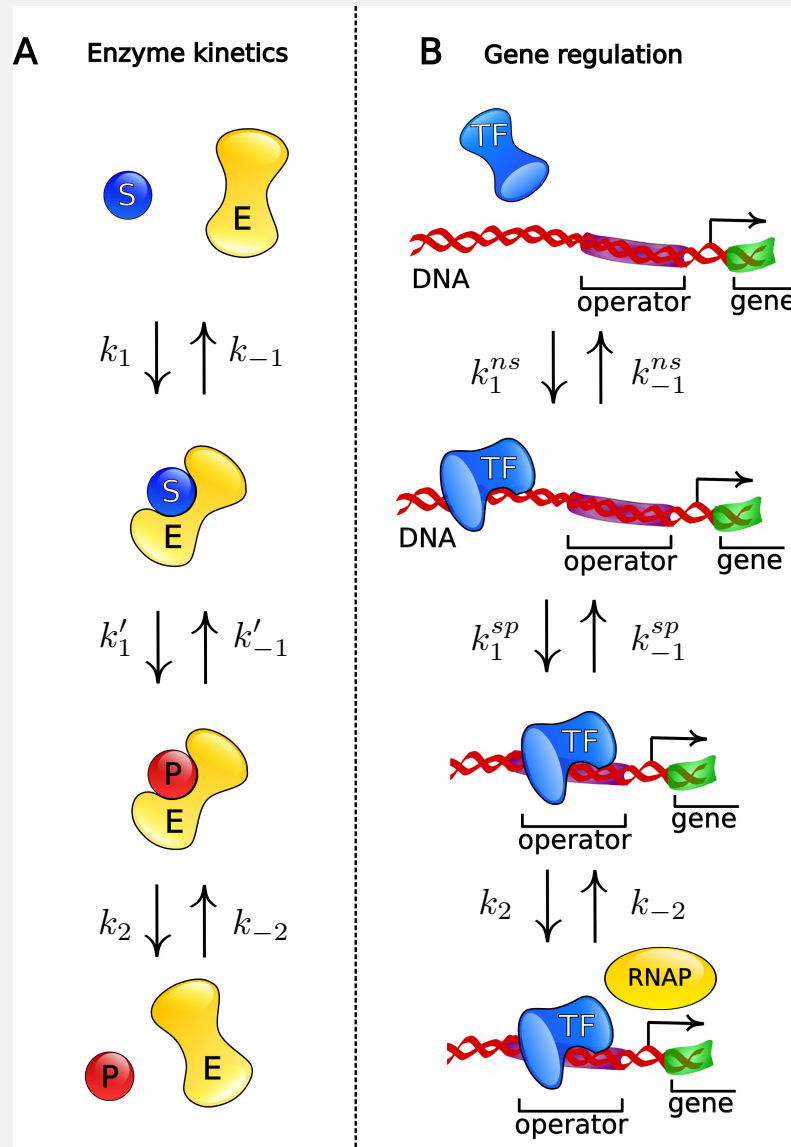
Many unknowns in the modelling:

Physical mechanism of anomalous diffusion & cutoff time of anomalous motion?

Effects of crowders with different sizes: see eg Shin et al, Soft Matter (2015) influencing immediate rebinding?

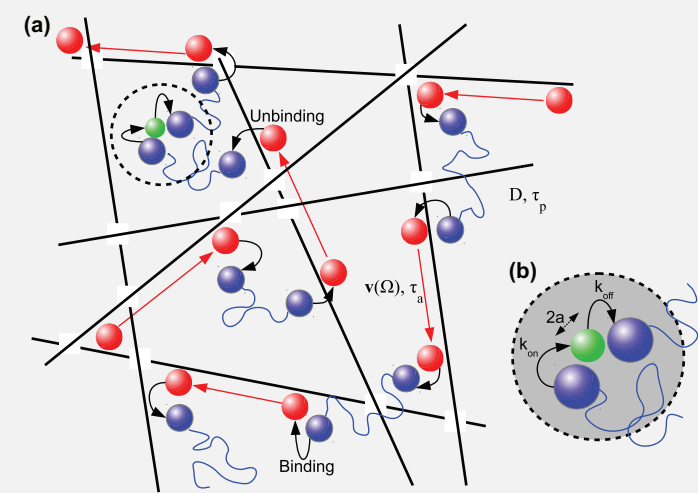
DNA conformations & dynamics due to crowding & active motion: Shin et al, NJP (2015), NJP (2016)

Low-# Michaelis-Menten

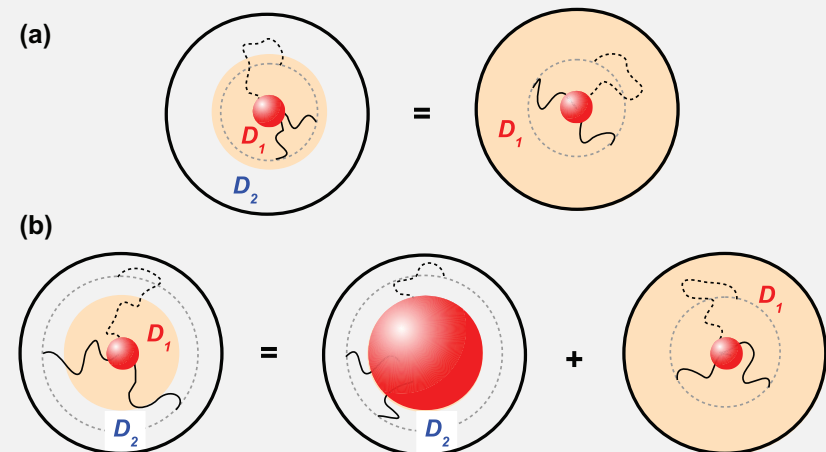


O Pulkkinen & RM, Sci Rep (2015)

Active sensing limit



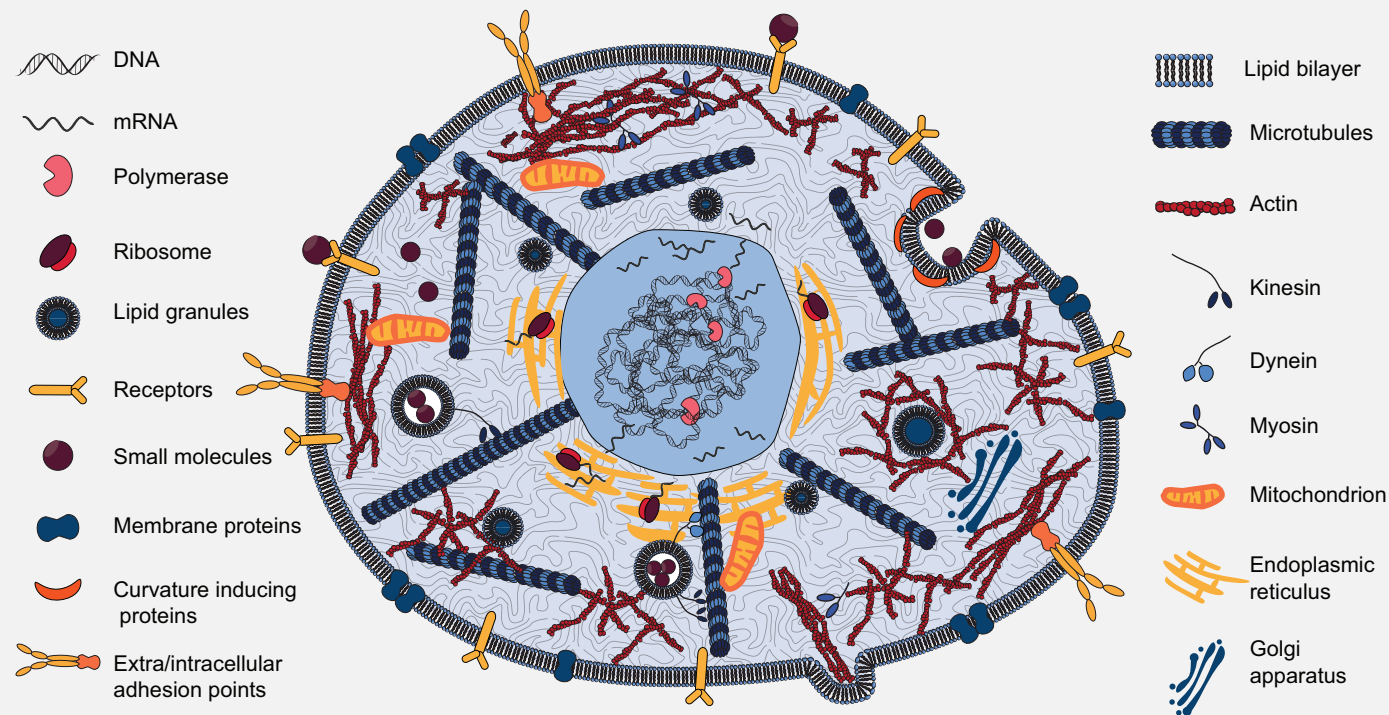
Heterogeneous FPT



New time scale in FP PDF!

A Godec & RM, PRE(R) & PRE (2015), Sci Rep (2016) 35

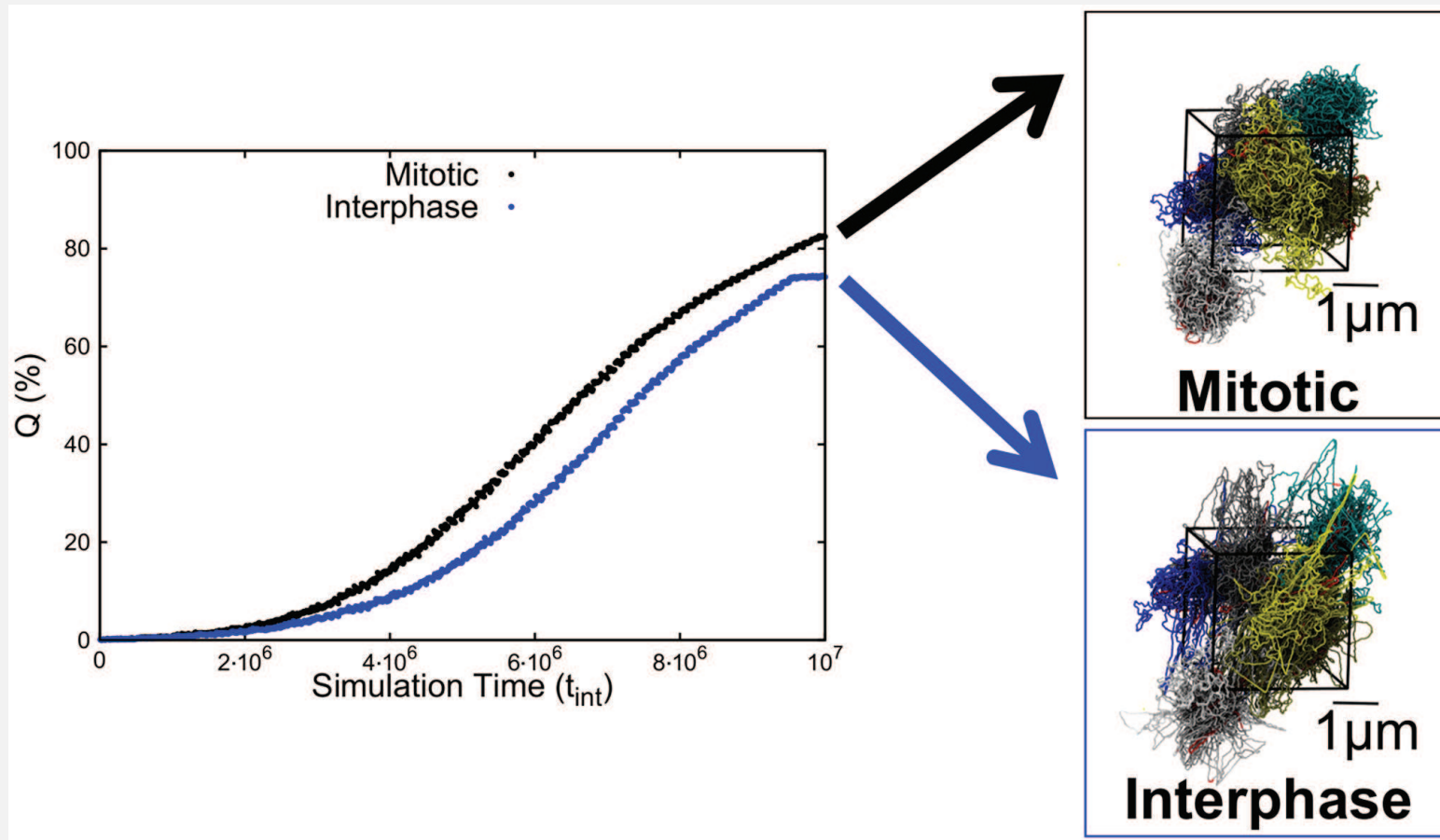
Gene regulation in eukaryotic cells



Exchange versus nucleic membrane, chromosomal dynamics & packaging

Active motion: motor transport, drag, or swirling (cytoplasmic streaming), see, e.g., Seisenberger et al, Science (2001) or Reverey et al, Sci Rep (2015)

Colocalisation still exists in the nucleus



Increase of percentage Q of coregulated pairs of genes in chromosome 19 which colocalise during the MD protocol. Red (???) highlighted regions designate chromosome regions involved in the coregulatory network

Superdiffusion in living *Acanthamoeba castellani*

