The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2038-20

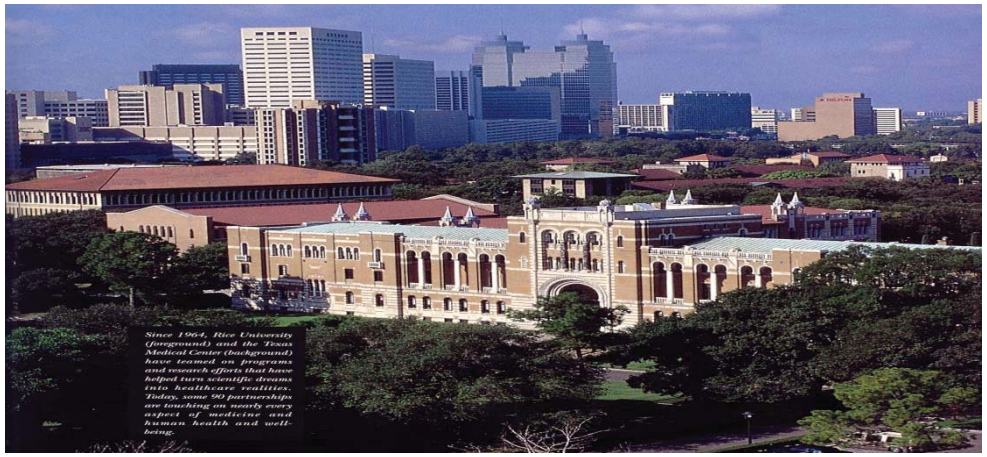
Conference: From DNA-Inspired Physics to Physics-Inspired Biology

1 - 5 June 2009

How Proteins Find Their Targets on DNA

Anatoly B. KOLOMEISKY
Rice University, Department of Chemistry
Houston
TX 77005-1892
USA

RICE UNIVERSITY



Anatoly B. Kolomeisky

Department of Chemistry

How Proteins Find Their Targets on DNA

COLLABORATION:

Prof. A.A. Kornyshev, Department of Chemistry, Faculty of Physical Sciences, Imperial College London, UK.

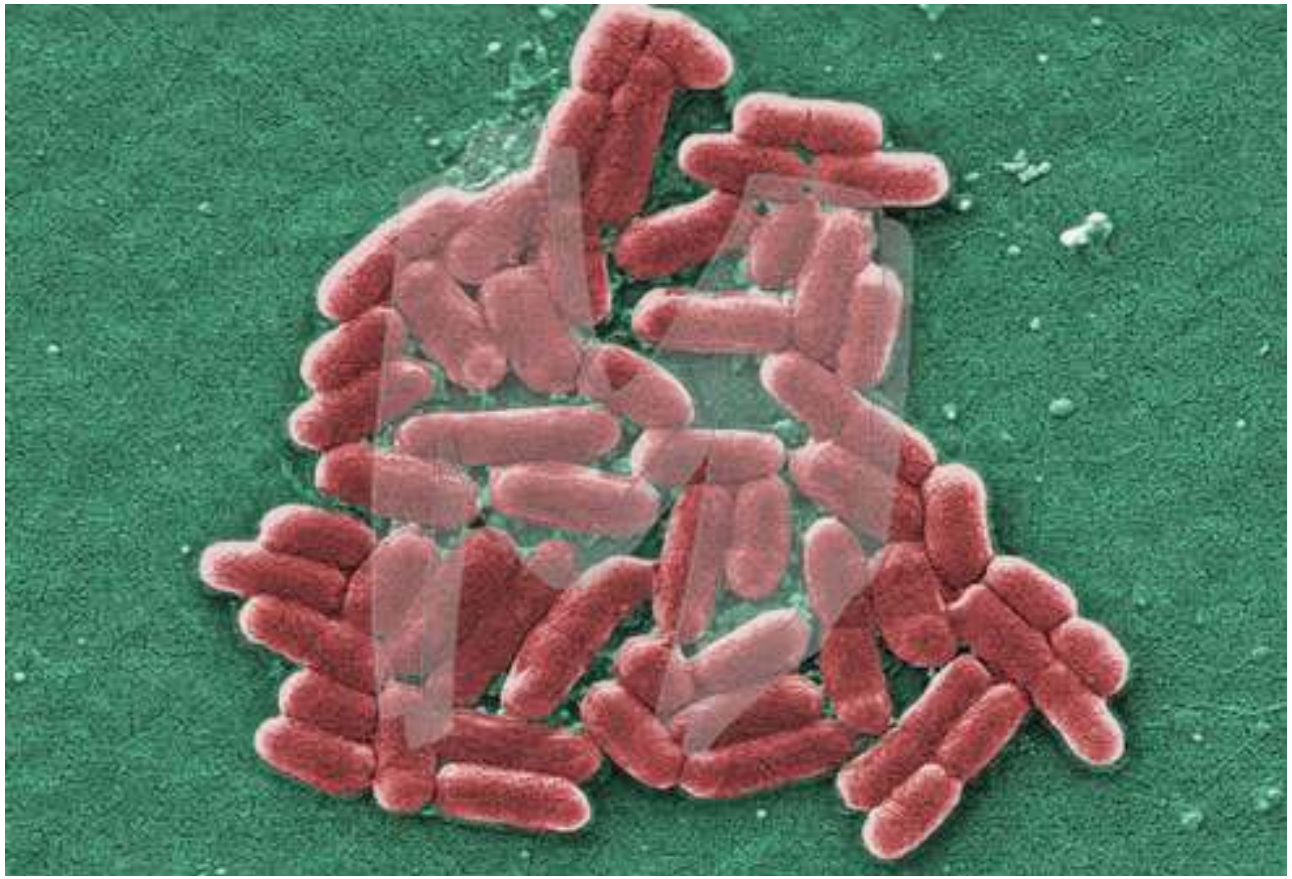
Dr. A.G. Cherstvy, Research Center, Juelich, Germany.

Rahul K. Das, graduate student in my group

Protein-DNA Interactions

Play fundamental role in all biological processes

Example #1: switching genes on and off in *Ecoli* bacteria

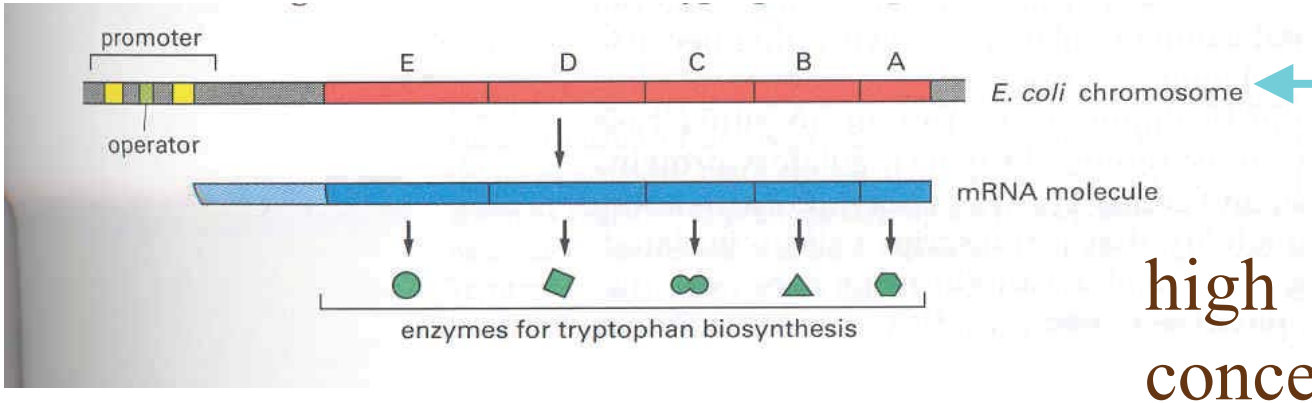


Large numbers of *Ecoli* bacteria live in human intestines

Bacteria needs to produce proteins which are made from aminoacids

Gene Switch in Bacteria

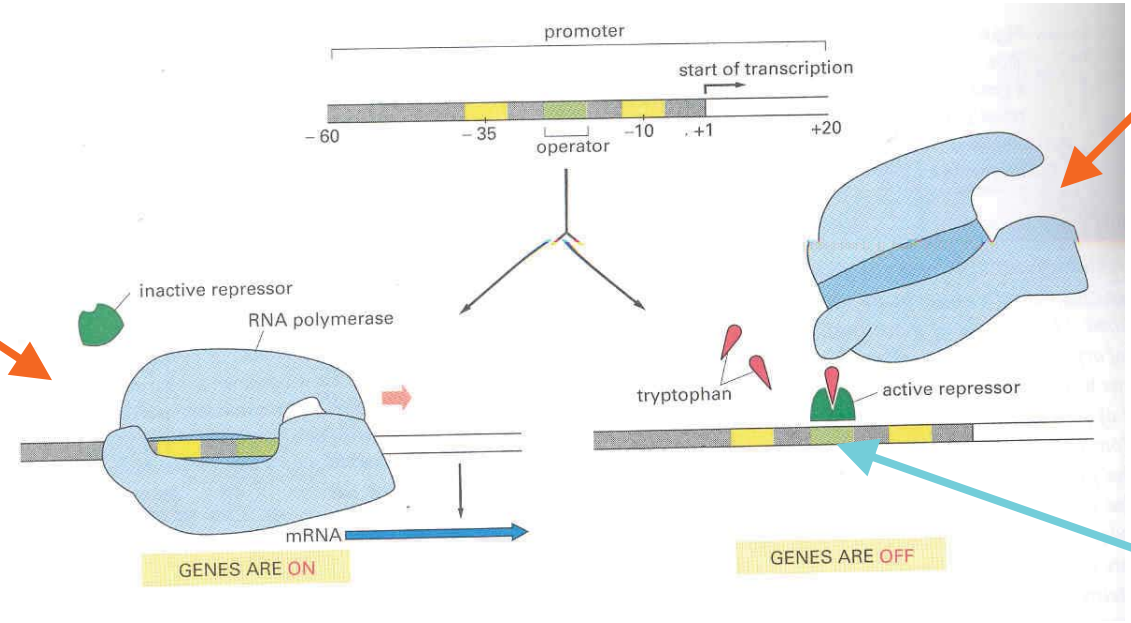
Production of the amino acid tryptophan



DNA

low concentration of tryptophan

high concentration of tryptophan

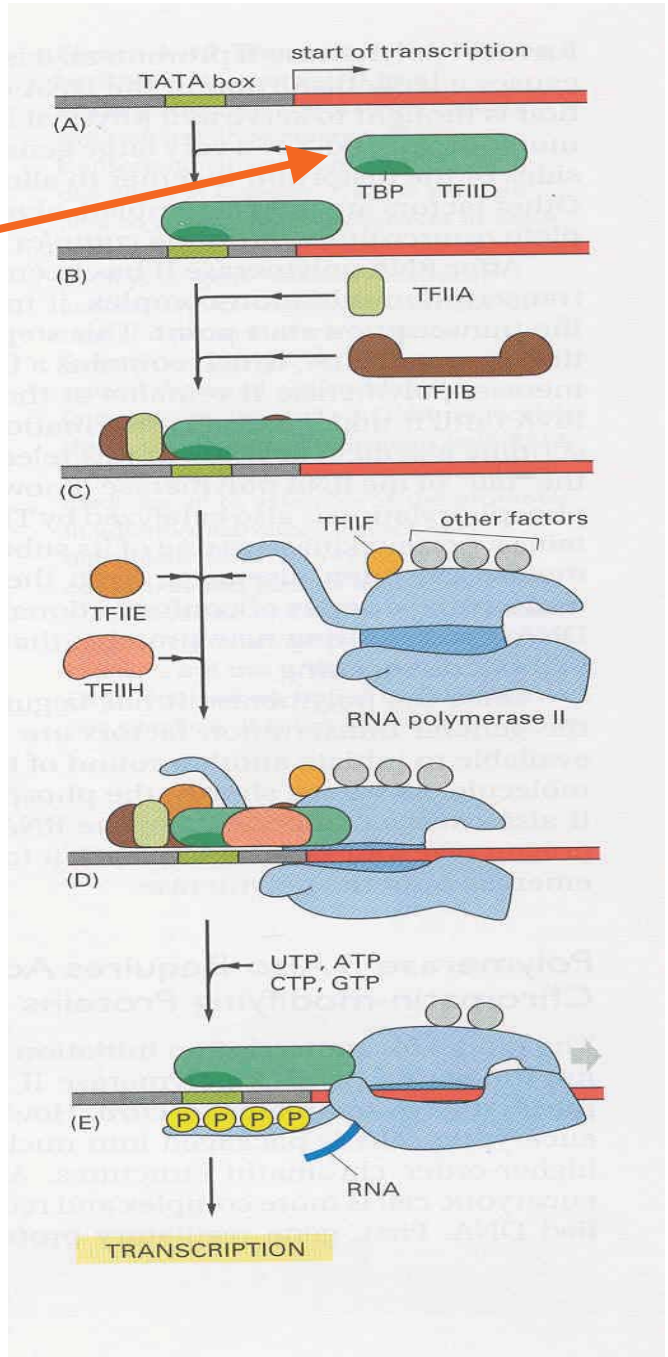


target

Transcription

Example #2:

transcription proteins must find a special sequence (called TATA box), 25 bp from the binding site of RNA Polymerase before transcription starts



Critical step in protein-DNA interactions:

protein finding and recognizing the target on DNA

Protein-DNA Interactions

Experimental Observations: Lac repressor protein finds its target with association rate $k_{ass} \sim 10^{10}\ M^{-1}\ s^{-1}$ (!!!)

Riggler et al., *J. Mol. Biol.* **53** 401-417 (1970)

100-1000 times larger than the maximal rate by three-dimensional diffusion-controlled search (from Debye-Smoluchowski theory)

The phenomenon is called **Facilitated Diffusion**

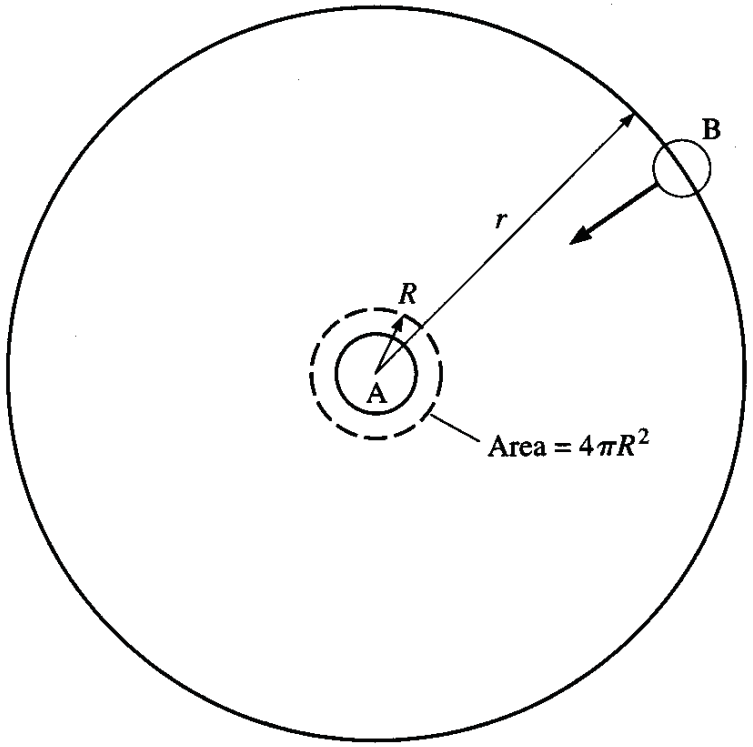
Generated huge controversy – still unresolved

Diffusion-Controlled Rate Constants

Compound B reacts with compound A after approaching it within the radius R

Debye-Smoluchowski

$$k_{\max} = 4\pi(D_A + D_B)R$$



For protein-DNA interactions it is assumed that DNA does not move, i.e., D_A=0, target size is few bases - R~1 nm, for a typical protein $D_B \sim 10^{-11}$ m^2 s^{-1}

$k_{\max} \sim 10^8$ M^{-1} s^{-1}, compare with $k_{exp} \sim 10^{10}$ M^{-1} s^{-1}

Possible Resolution?

transactions

nsoctrans.org

An end to 40 years of mistakes in DNA–protein association kinetics?

Stephen E. Halford[1]

The DNA–Proteins Interaction Unit, Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 4DL, U.K.

Abstract

Proteins that bind to specific sequences in long DNA molecules have to locate their target sites amid myriad alternative sequences, yet they do so at remarkably rapid rates, sometimes approaching 10^{10} $M^{-1} \cdot s^{-1}$. Hence, it has been asserted widely that binding to specific DNA sites can surpass the maximal rate for 3D (three-dimensional) diffusion through solution and that this could only be accounted for by a reduction in the dimensionality of the search for the target in effect by 1D (one-dimensional) diffusion (or 'sliding') along the DNA contour. It will be shown here that there is, in fact, no known example of a protein binding to a specific DNA site at a rate above the diffusion limit, and that the rapidity of these reactions is due primarily to electrostatic interactions between oppositely charged molecules. It will also be shown that, contrary to popular belief, reduced dimensionality does not, in general, increase the rate of target-site location but instead reduces it. Finally, it will be demonstrated that proteins locate their target sites primarily by multiple dissociation/reassociation events to other (nearby or distant) sites within the same DNA molecule, and that 1D diffusion is limited to local searches covering ~50 bp around each landing site.

Biochem Soc. Trans. 2009, **37**, 343-348

Possible Resolution? NO!!!

S.E. Halford, *Biochem Soc. Trans*. 2009, **37**, 343-348

Original experiments have been carried out in low-salt conditions:
10 mM of KCl, Tris/HCl, and magnesium acetate

transactions
nsoctrans.org

An end to 40 years of mistakes in DNA–protein association kinetics?

Stephen E. Halford[1]
The DNA–Proteins Interaction Unit, Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol BS8 4DL, U.K.

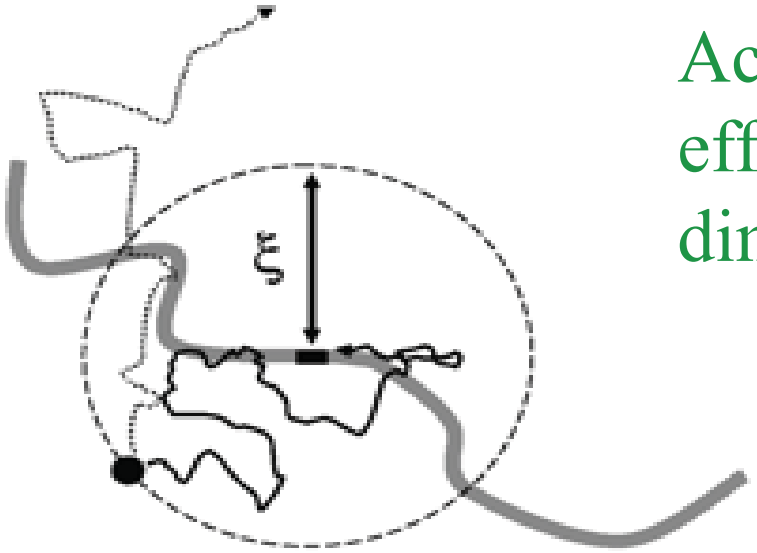
Abstract

Proteins that bind to specific sequences in long DNA molecules have to locate their target sites amid myriad alternative sequences, yet they do so at remarkably rapid rates, sometimes approaching 10^{10} $M^{-1} \cdot s^{-1}$. Hence, it has been asserted widely that binding to specific DNA sites can surpass the maximal rate for 3D (three-dimensional) diffusion through solution and that this could only be accounted for by a reduction in the dimensionality of the search for the target in effect by 1D (one-dimensional) diffusion (or 'sliding') along the DNA contour. It will be shown here that there is, in fact, no known example of a protein binding to a specific DNA site at a rate above the diffusion limit, and that the rapidity of these reactions is due primarily to electrostatic interactions between oppositely charged molecules. It will also be shown that, contrary to popular belief, reduced dimensionality does not, in general, increase the rate of target-site location but instead reduces it. Finally, it will be demonstrated that proteins locate their target sites primarily by multiple dissociation/reassociation events to other (nearby or distant) sites within the same DNA molecule, and that 1D diffusion is limited to local searches covering ~50 bp around each landing site.

Debye length – separates region where electrostatics is important
In this system it is ~1-2 nm, comparable with the target size.
Electrostatics does not play role in the fast search!

Facilitated Diffusion: Current Theoretical Views

Protein search for the target is viewed as a sequence of 3D excursions and 1D sliding along the DNA



Acceleration due to increase in the effective target length or lowering dimensionality; assumed that $D_1 \sim D_3$

$$k_{\max} = 4\pi(D_A + D_B)R$$

S.E. Halford and J.F. Marko, *Nucl. Acids Res.* **32**, 340 (2004)

Experiments support facilitated diffusion picture:

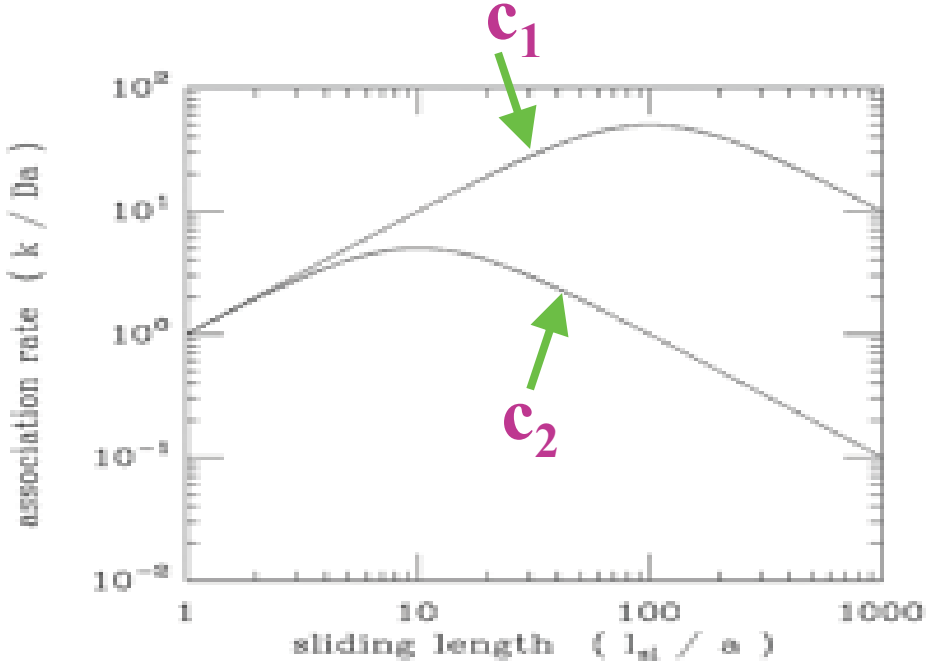
D.M. Gowers, G.G. Wilson and S.E. Halford, *PNAS USA*, **102**, 15883 (2005)

Problems:

Current theories predict: $D_1 \sim D_3 \sim 10^{-11}$ m^2 s^{-1}, and $\tau(1D) \sim \tau(3D)$

Single-molecule experiments: $D_1 \sim 10^{-13}$-10^{-16} m^2 s^{-1},
$\tau(1D)/\tau(3D) > 10$-100

Phys. Rev. Lett., **97**, 048302 (2006);*Science*, **316**, 1191 (2007)



$c_1 < c_2$

What is wrong with this picture?

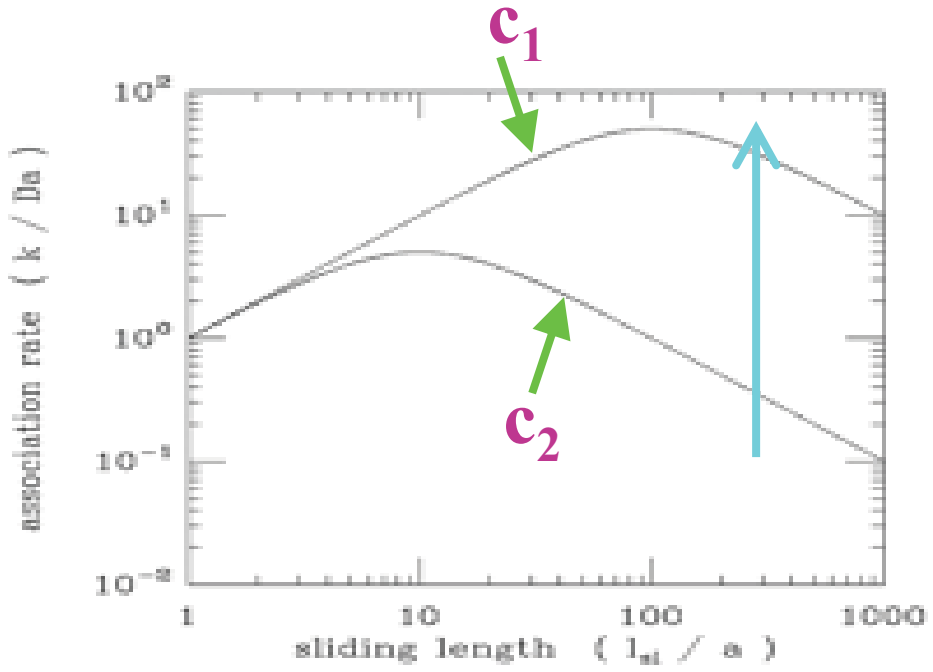
S.E. Halford and J.F. Marko, *Nucl. Acids Res.* **32**, 340 (2004)

Problems:

Current theories predict: $D_1 \sim D_3 \sim 10^{-11}$ m^2 s^{-1}, and $t_1 \sim t_3$

Single-molecule experiments: $D_1 \sim 10^{-13}$-10^{-16} m^2 s^{-1}, $t_1/t_3 > 10$

Phys. Rev. Lett., **97**, 048302 (2006);*Science*, **316**, 1191 (2007)



$c_1 < c_2$

decreasing the concentration of proteins the reaction of association increases!

Contradicts Chemistry!

Unphysical behavior

S.E. Halford and J.F. Marko, *Nucl. Acids Res.* **32**, 340 (2004)

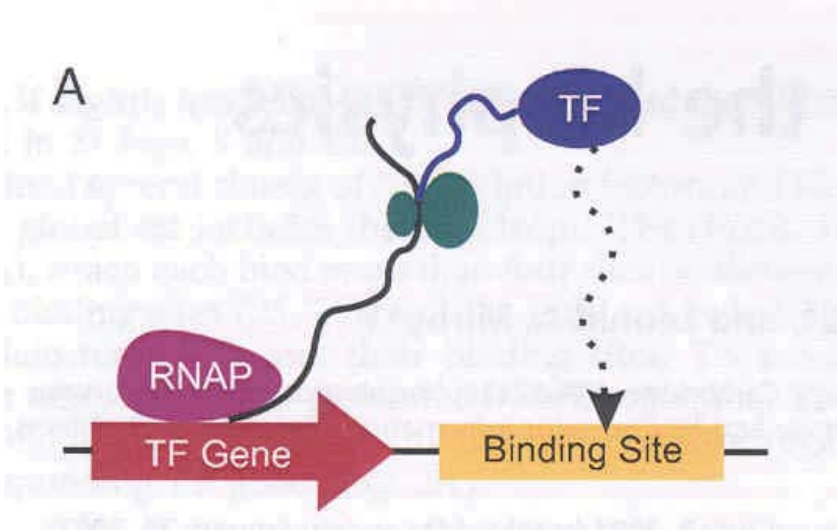
Problems:

G. Kolesov et al., *PNAS USA*, **104**, 13948 (2007)

Search time for transcription factors (TF) proteins from theoretical estimates – 15-500 minutes (!!!), while from experiments ~1min

Colocalization mechanism:

Genes for TF and for their targets are close to enable fast search



However, 1) it does not work for eukaryotes; 2) does not work for proteins with many targets; 3) does not explain *invitro* experiments

Our Goal:

To develop a simple model of protein search and recognition for targets on DNA consistent with experimental observations and basic laws of Chemistry and Physics

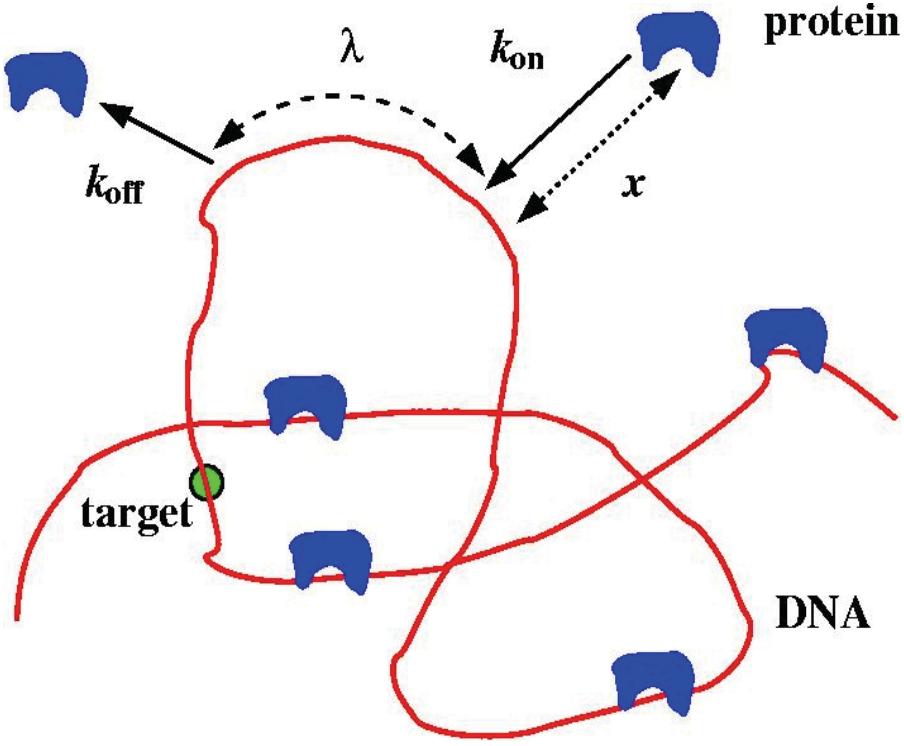
Protein-DNA interactions – **2 steps:**

1) **Finding the target sequence**

2) **Recognizing the target**

Successful theoretical picture must account for both stages of protein-DNA interactions

Our Theoretical Approach



L~1 μm – length of DNA

a~1 nm – target size

1) Reaching the target on DNA is viewed as a sequence of searching events (cycles)

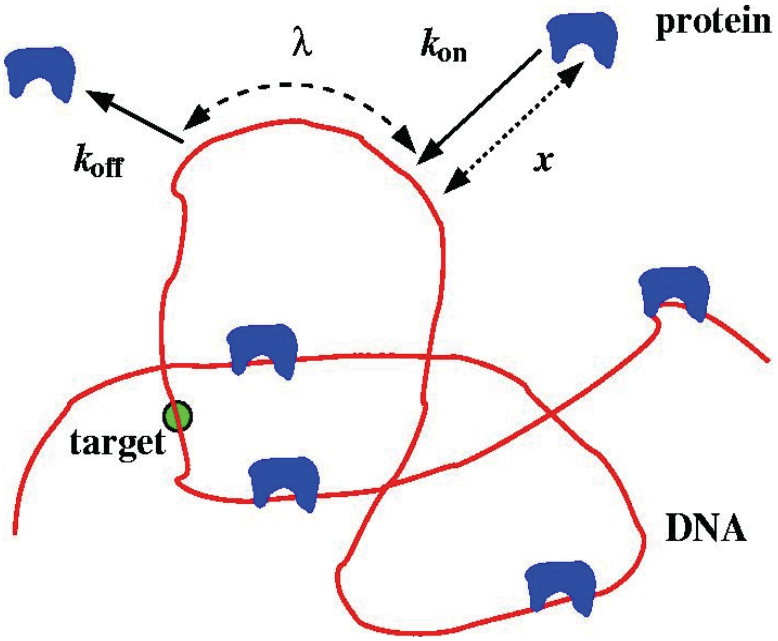
2) Each protein on average binds/unbinds several times before finding the target

3) Each cycles consists of 3D and 1D tracks

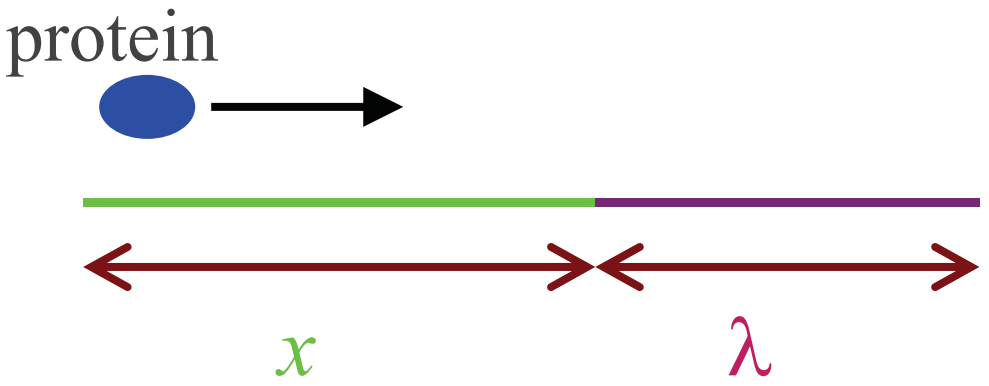
4) Correlations

5) No assumption of equilibrium

Our Theoretical Approach



One search cycle – effective overall 1D motion



3D segment 1D segment

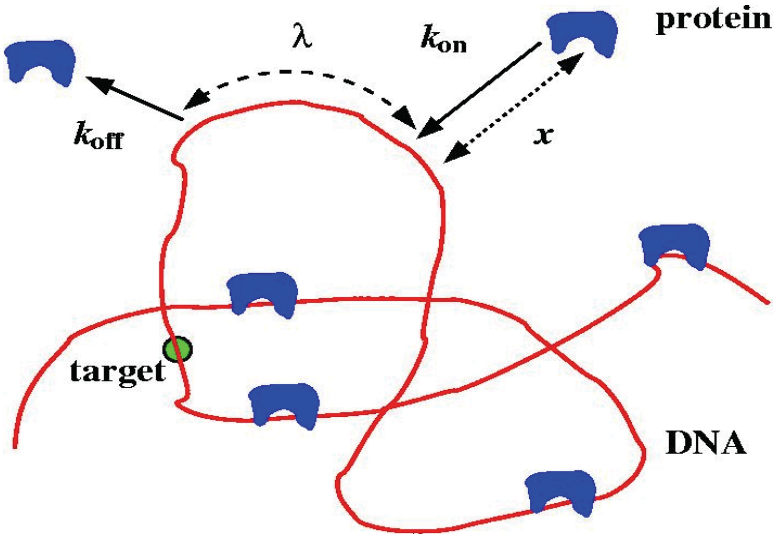
Mean first-passage time for one searching cycle

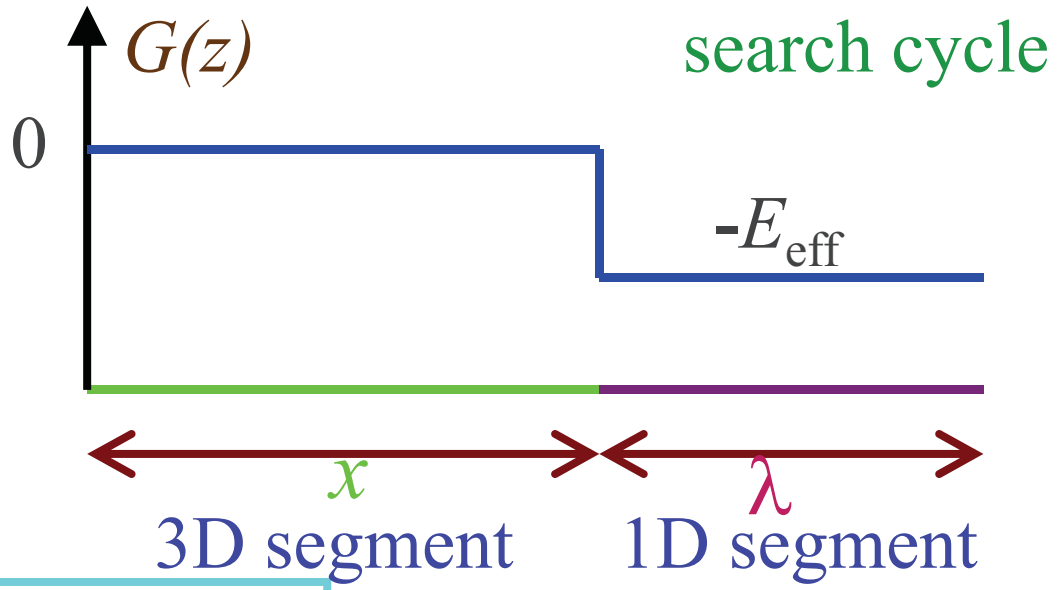
$$\tau_c = \int_0^{x+\lambda} \frac{\exp[\beta G(z)]}{D(z)} dz \int_0^{z} \exp[-\beta G(z')] dz'$$

position-dependent diffusion constant

$$D(z) = \begin{cases} D_3, 0 < z < x \\ D_1, x < z > z + \lambda \end{cases}$$

Our Theoretical Approach





$$\tau_c = \int_0^{x+\lambda} \frac{\exp[\beta G(z)]}{D(z)} dz \int_0^{z} \exp[-\beta G(z')] dz'$$

E_{ads} – non-specific binding energy

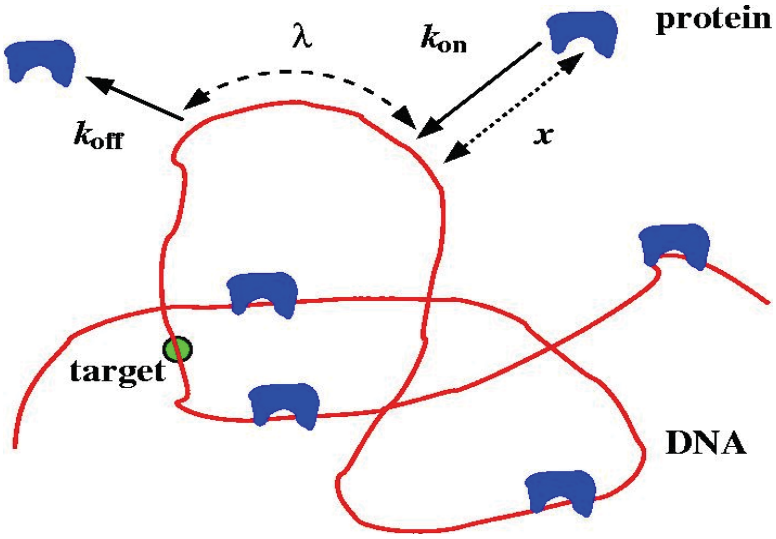
y-equilibrium constant for binding/unbinding

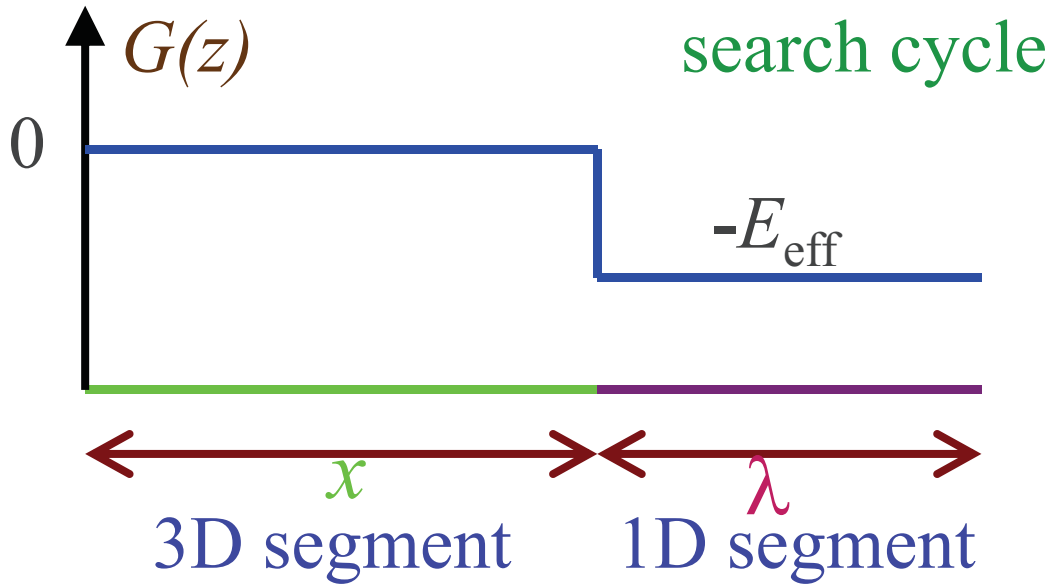
c_p-concentration of proteins in the bulk; c_{ads} – adsorbed to DNA

$$y = \frac{k_{on}}{k_{off}} = \exp\left(\frac{E_{ads}}{k_B T}\right)$$

$$y_{eff} = \frac{k_{on} c_p}{k_{off} c_{ads}} = \exp\left(\frac{E_{eff}}{k_B T}\right)$$

Our Theoretical Approach





Mean first-passage time for one searching cycle

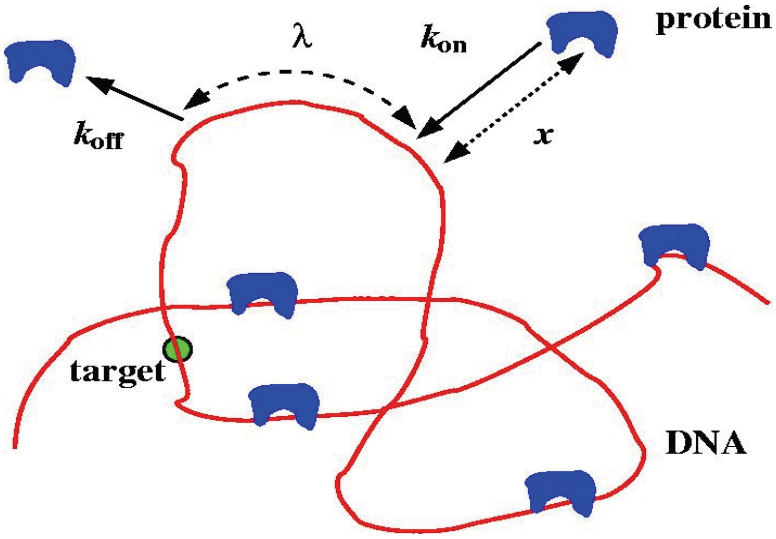
$$\tau_c = \frac{x^2}{2D_3} + \frac{\lambda^2}{2D_1} + \frac{x\lambda}{D_1 y_{eff}}$$

correlation term – reflects forward-backward trajectories

time spent by the protein only on 3D segment

time spent by the protein only on 1D segment

Our Theoretical Approach



Consider 1 DNA molecule

n_{ads}-number of adsorbed proteins per DNA

λn_{ads} – length of DNA scanned during one cycle

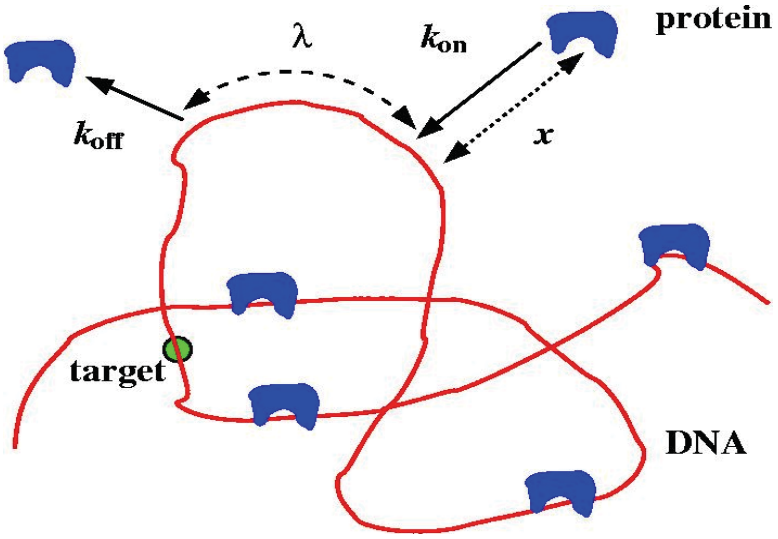
$L/\lambda n_{ads}$ –average number of cycles before finding the target

Total search time is a sum of one-cycle times

$$\tau = \tau_c \left(\frac{L}{\lambda n_{ads}} \right)$$

We assume low concentration of proteins on DNA and no memory in rebinding

Our Theoretical Approach



$$\tau = \tau_c \left(\frac{L}{\lambda n_{ads}} \right)$$

$$S_j = p(1-p)^{j-1}$$

Average number of search cycles

Why average number of search cycles is $L/\lambda n_{ads}$?

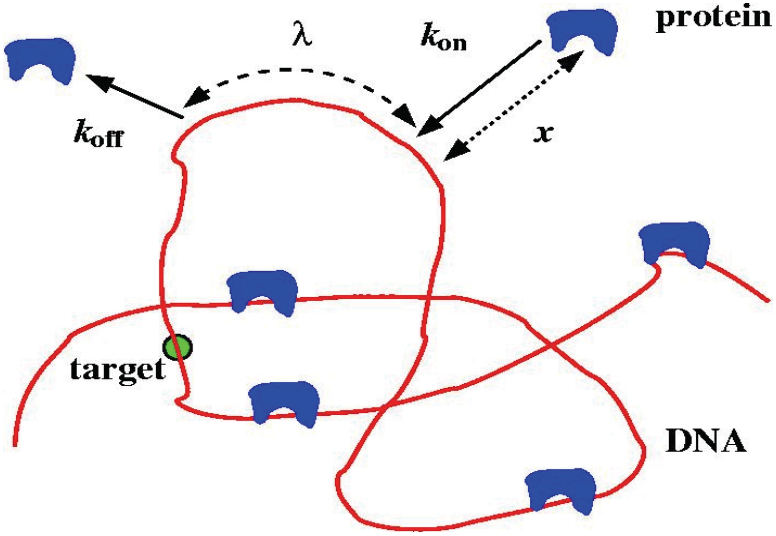
$p=\lambda n_{ads}/L$ - probability to find a target after binding DNA

λn_{ads}-average distance scanned by all proteins on DNA in one cycle

S_j- probability to find the target after j-th cycle

$$\langle j \rangle = \sum_{j-1}^{\infty} jS_j = \frac{1}{p} = \frac{L}{\lambda n_{ads}}$$

Our Theoretical Approach



$$V = Lr^2$$

volume per DNA,

r-effective DNA radius

a-target size

$$c_p = \frac{n_p}{V}, \; c_{ads} = \frac{n_{ads}}{V}$$

Total search time should be compared with Smoluchowski 3D search time

$$\tau_S = \frac{1}{2 D_3 a c_p} = \frac{Lr^2}{2 D_3 a n_p}$$

relative search time:

$$\frac{\tau}{\tau_S} = \frac{a}{r}\left(\frac{1}{\sqrt{n_{ads} y d}} + \frac{n_p \sqrt{y}}{n_{ads}^{3/2}\sqrt{d}} + \frac{2}{\sqrt{n_p y d}}\right)$$

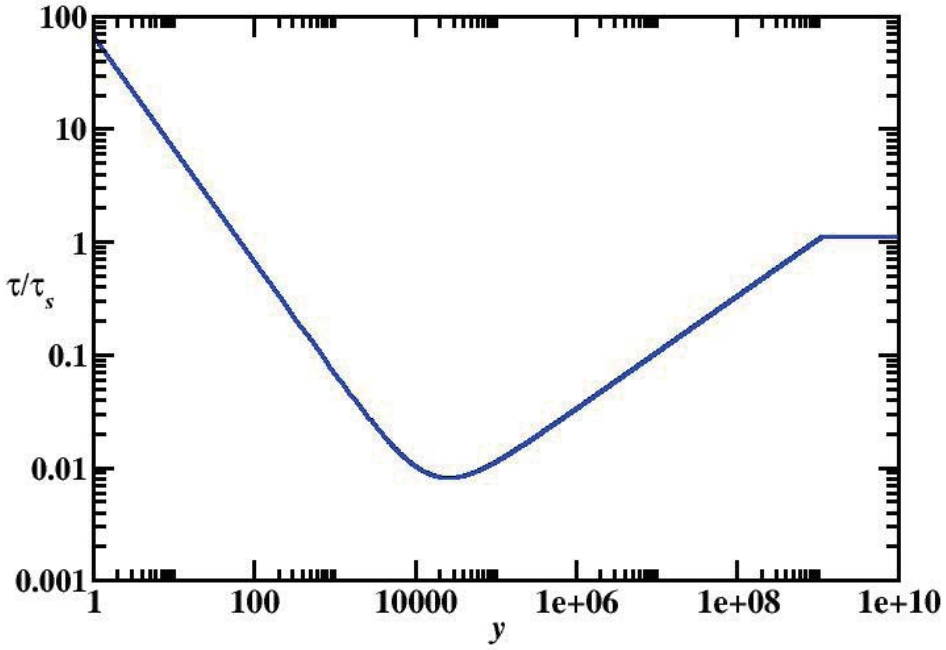
$d=D_1/D_3$-ratio of diffusion constants

y-equilibrium constant

Results

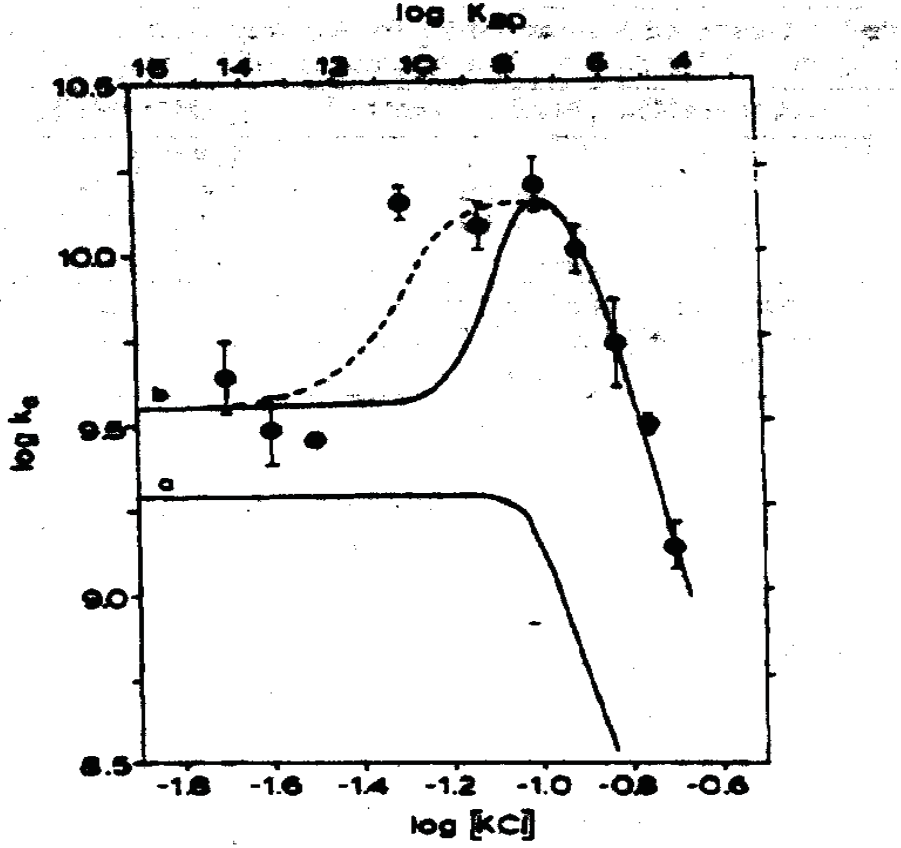
relative search time as a function of the adsorption strength

a=1nm, r=30nm, n_{ads}=1000, n_p=1, d=0.001



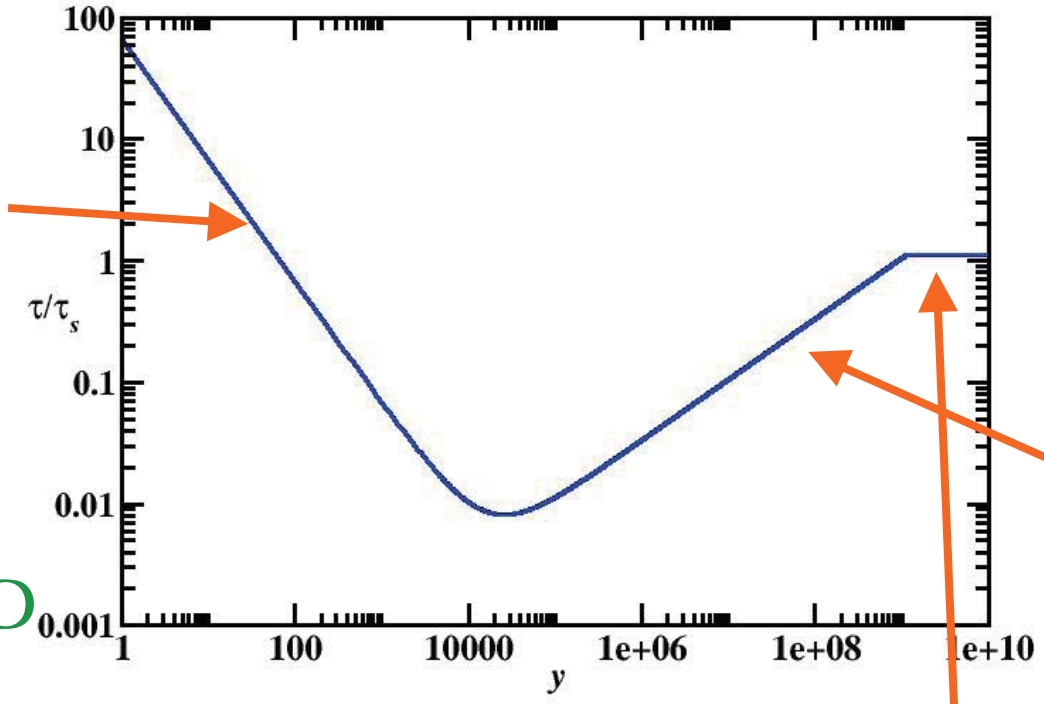
Interactions between proteins and DNA depend on ionic strength

Biochemistry, **20**, 6961 (1981)



Results

at $y<<1$ – strong repulsion between the protein and DNA, the protein is mainly in 3D motion, not enough time to scan DNA



at $y>>1$ – strong attraction between the protein and DNA, the protein is mainly in 1D motion, which is very slow

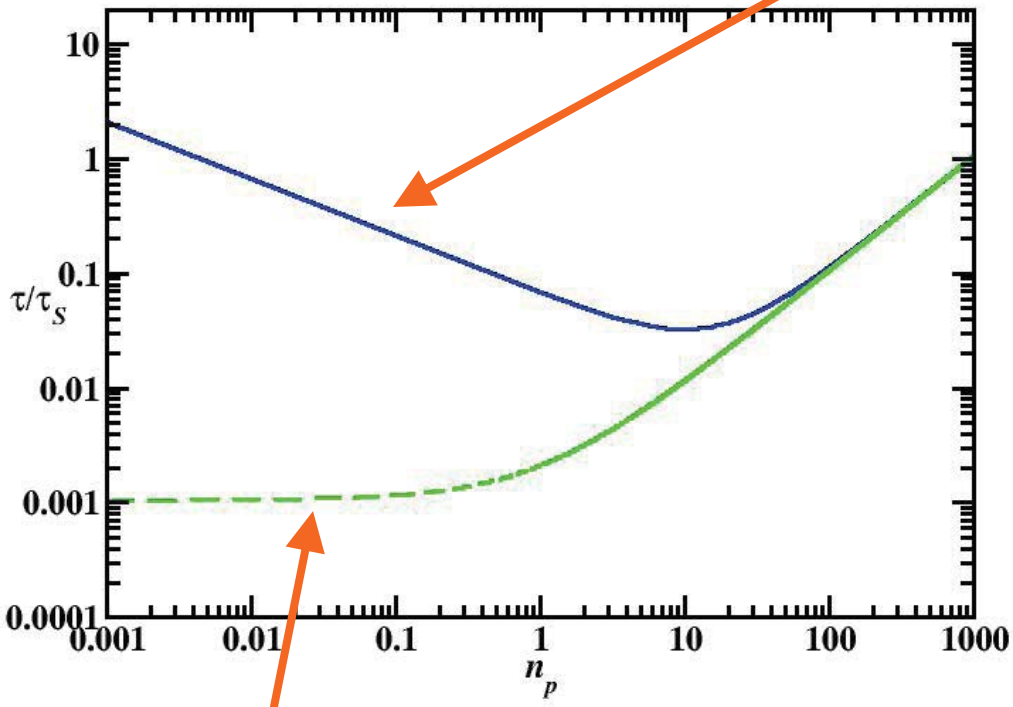
$$y = \frac{k_{on}}{k_{off}} = \exp\left(\frac{E_{ads}}{k_B T}\right)$$

$$\frac{\tau}{\tau_S} = \frac{a}{r}\left(\frac{1}{\sqrt{n_{ads} y d}} + \frac{n_p \sqrt{y}}{n_{ads}^{3/2}\sqrt{d}} + \frac{2}{\sqrt{n_p y d}}\right)$$

sliding length is equal to L

Results

Relative search time as a function of protein concentration



Old theories without correlation term

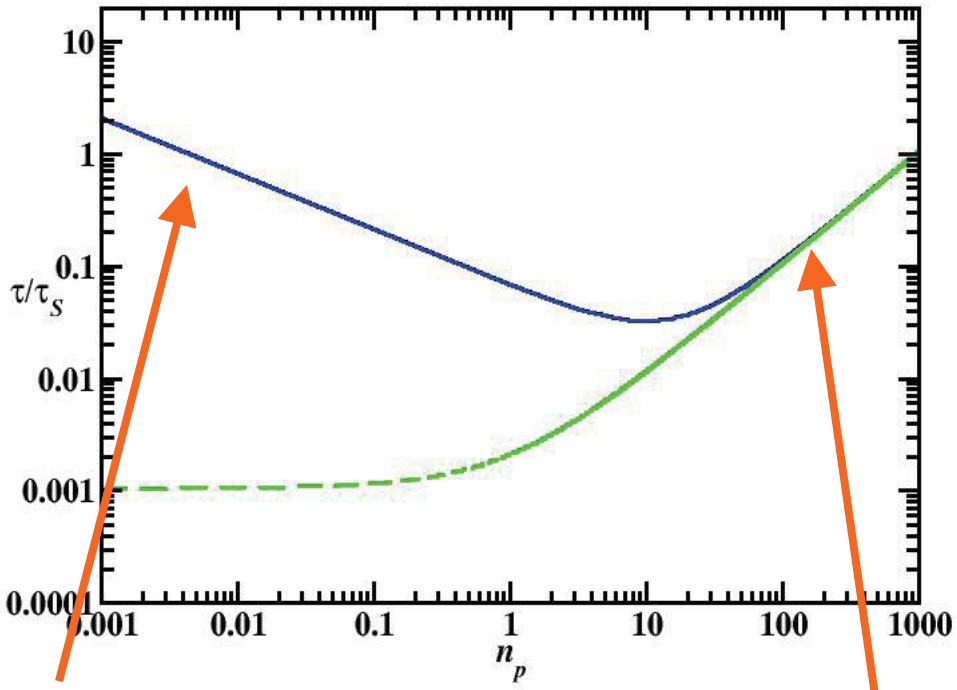
Our theory with correlation term

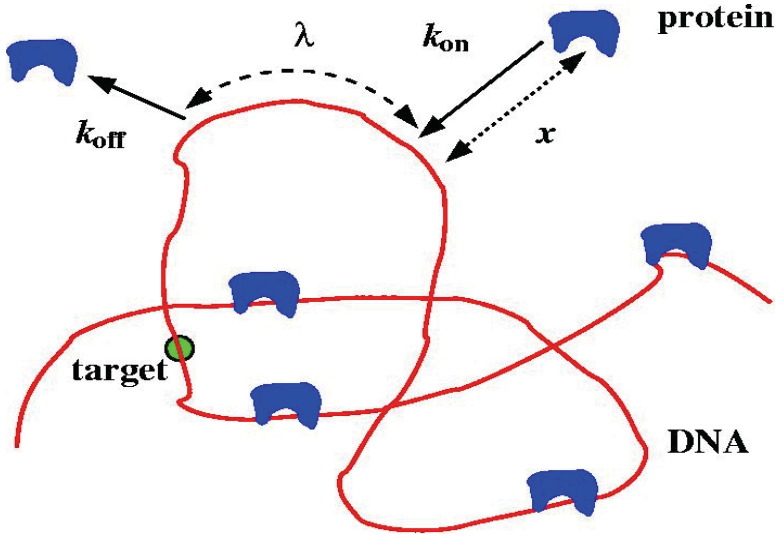
$$\frac{\tau}{\tau_S} = \frac{a}{r}\left(\frac{1}{\sqrt{n_{ads} y d}} + \frac{n_p \sqrt{y}}{n_{ads}^{3/2}\sqrt{d}} - \frac{2}{\sqrt{n_p y d}}\right)$$

Correlation term is critically important!

Results

Relative search time as a function of solution protein concentration:





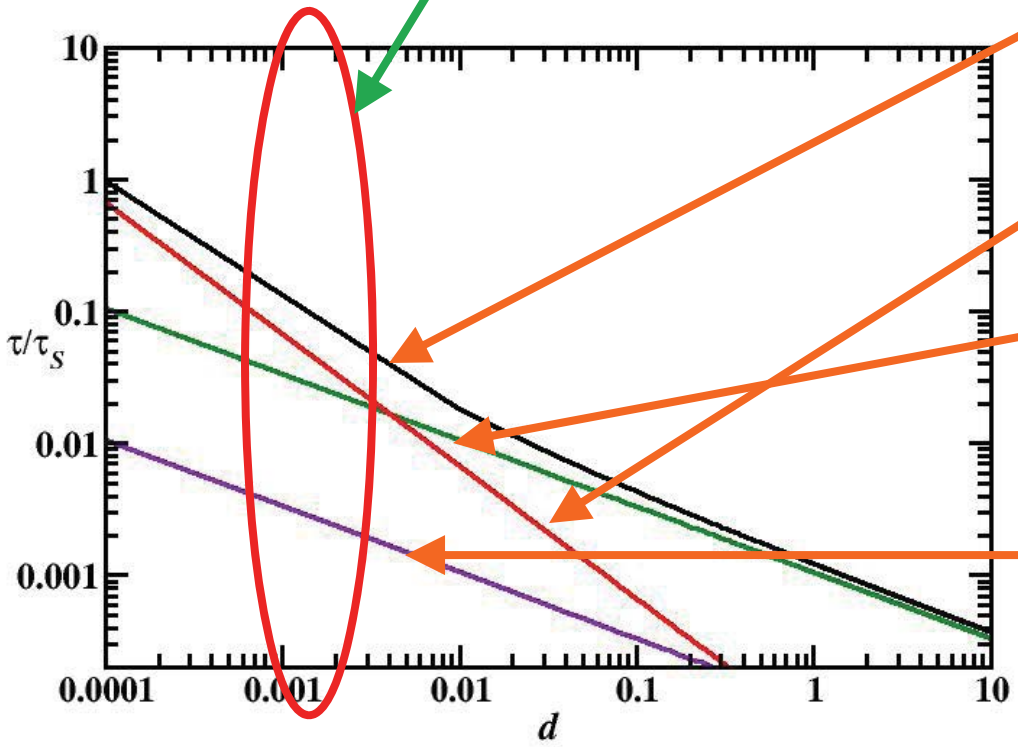
$$\frac{\tau}{\tau_S} = \frac{a}{r}\left(\frac{1}{\sqrt{n_{ads} y d}} + \frac{n_p \sqrt{y}}{n_{ads}^{3/2}\sqrt{d}} + \frac{2}{\sqrt{n_p y d}}\right)$$

Reaching the DNA is a rate-limiting step, protein mostly binds/unbinds

concentration of proteins becomes so large that it is faster to reach the target only by 3D

Results

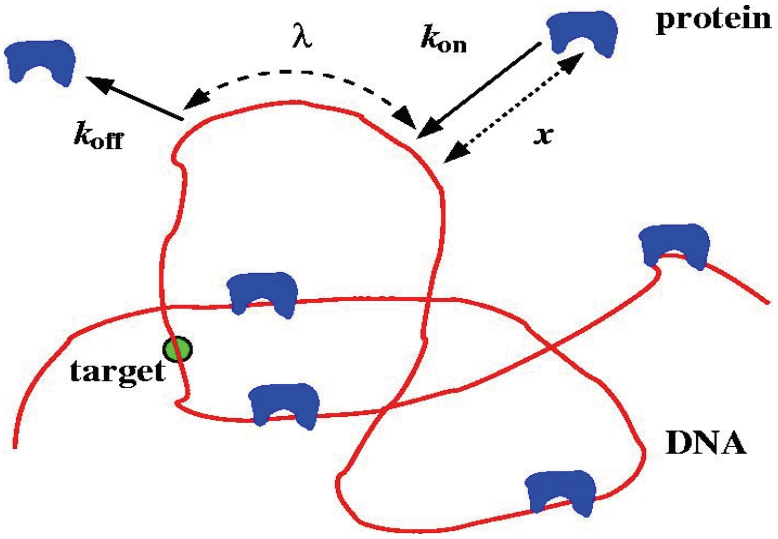
Relative search time as a function of the ratio of diffusion constants:

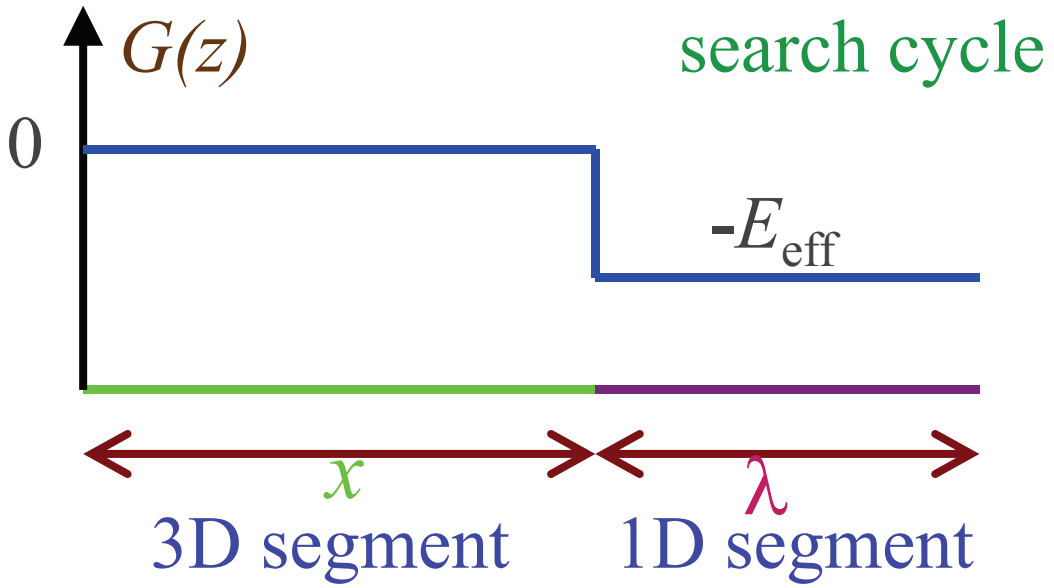


$$\frac{\tau}{\tau_S} = \frac{a}{r}\left(\frac{1}{\sqrt{n_{ads}\, yd}} + \frac{n_p\sqrt{y}}{n_{ads}^{3/2}\sqrt{d}} + \frac{2}{\sqrt{n_p\, yd}}\right)$$

$d=D_1/D_3$-ratio of diffusion constants

Results





Our theory is non-equilibrium, equilibrium is a special case when $E_{eff}=0$

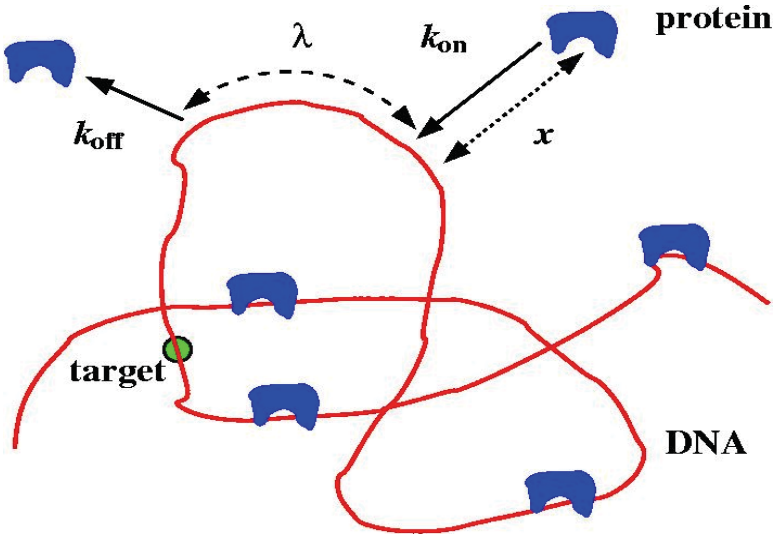
For realistic values a=1nm, r=30nm, y=10, n_p=1,d=0.001

No acceleration in our model

$$\frac{\tau}{\tau_S} = \frac{a}{r}\frac{2}{y\sqrt{n_p}}\frac{\sqrt{d}+1}{d}$$

facilitated diffusion – non-equilibrium process

Our Theoretical Approach



Non-specific binding:

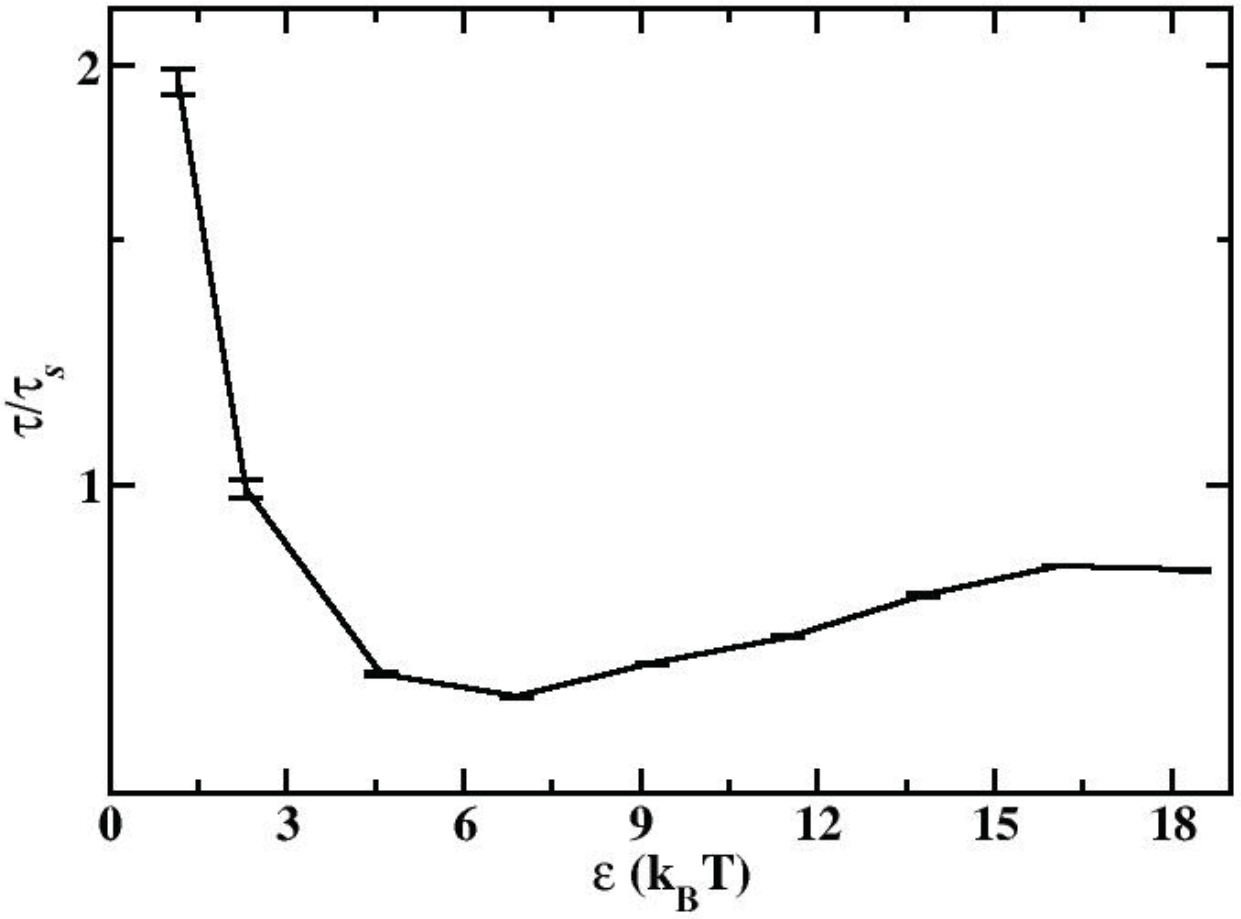
1) Slows down 1D diffusion
2) Increases concentration of proteins on DNA

Our view of facilitated diffusion mechanism:

1) Non-equilibrium process
2) Fast 3D and slow 1D motions
3) Correlations are important
4) Acceleration due to increased local concentration of proteins because of non-specific binding

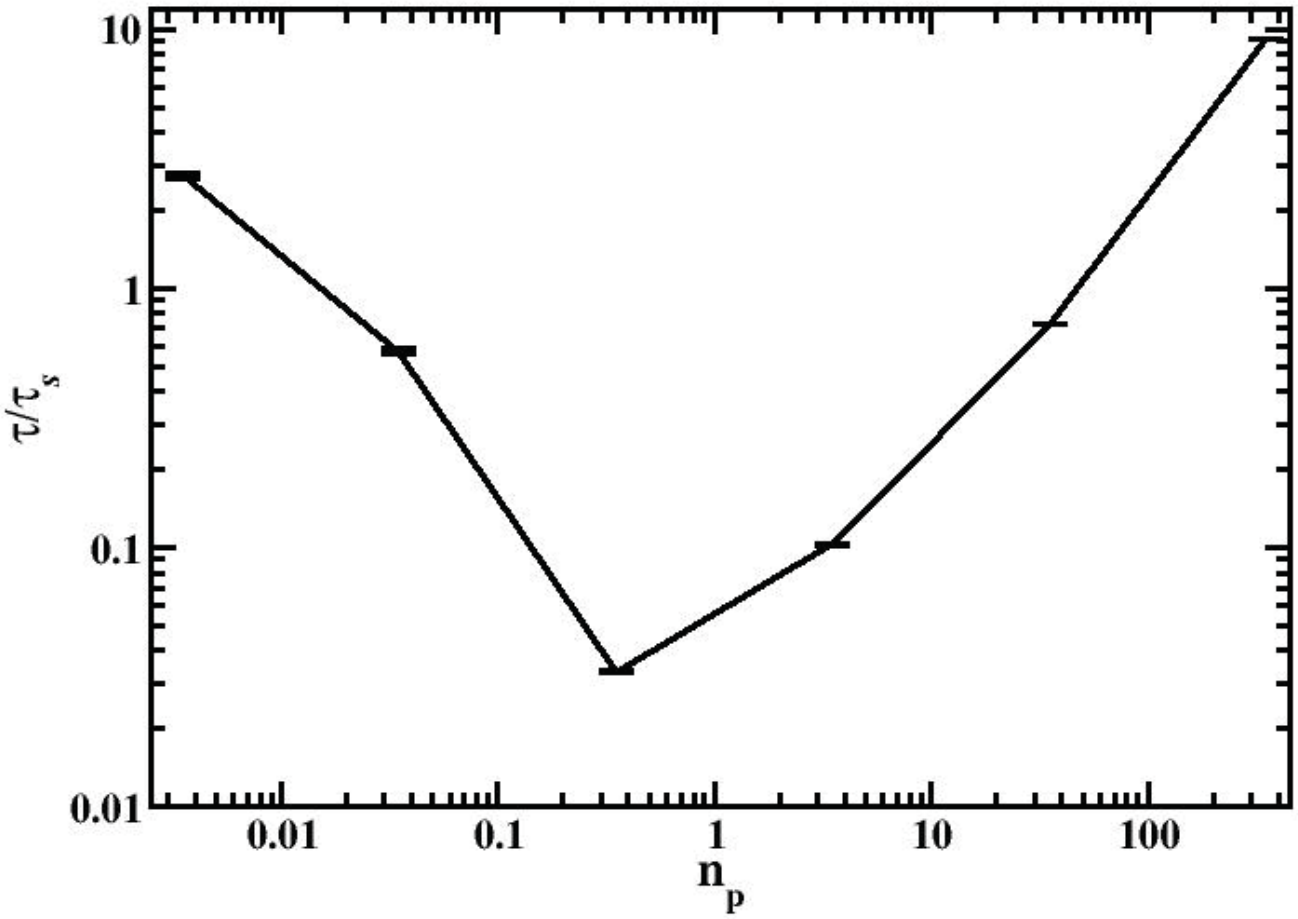
Monte Carlo Simulations

Relative search time as a function of non-specific interaction with DNA



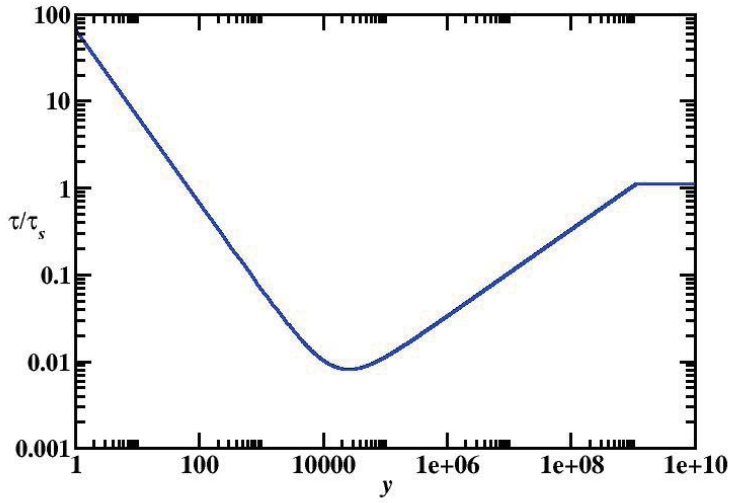
Monte Carlo Simulations

Relative search time as a function of free proteins in solution



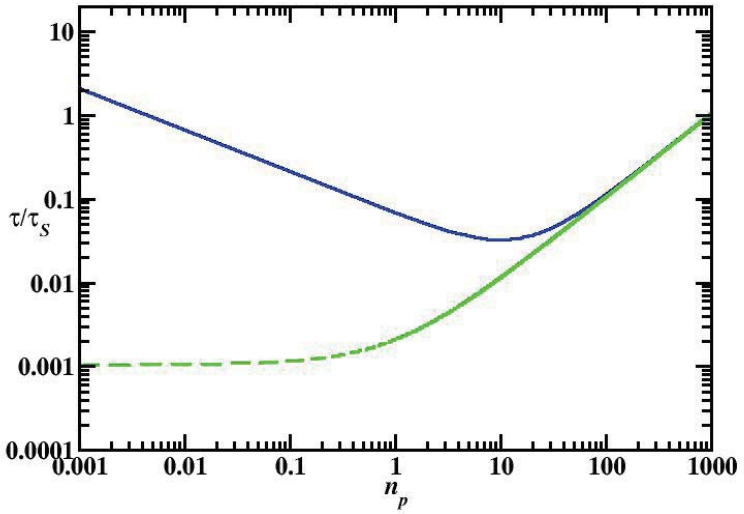
Our Predictions:

1) **Facilitated diffusion mechanism** works for intermediate range of adsorption energies. It can be controlled by changing ionic strength



Agrees with experiments and computer simulations

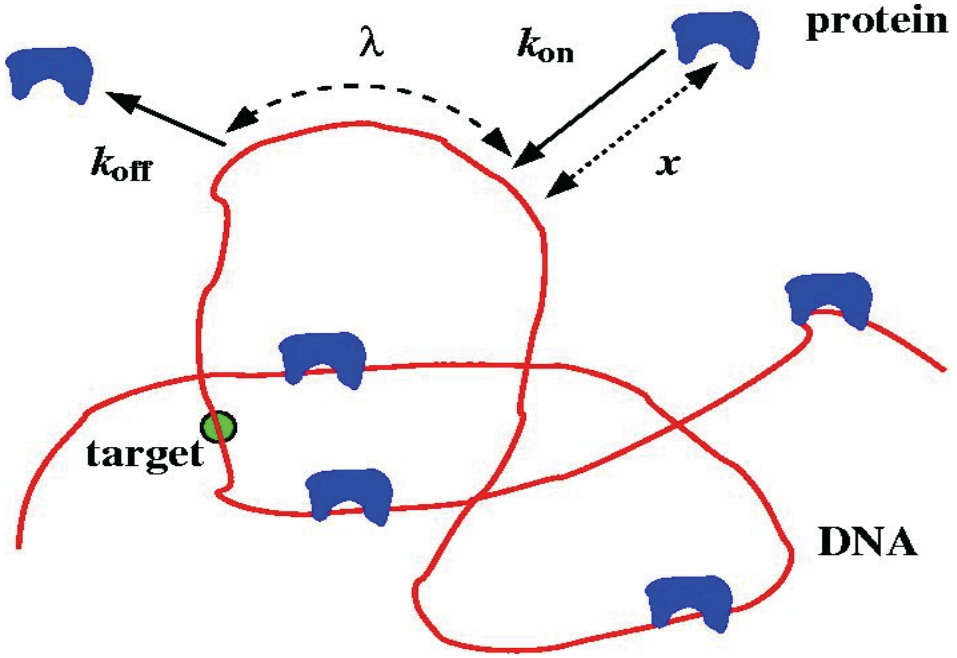
2) **Facilitated diffusion mechanism** works for intermediate concentrations of proteins in the solution



3) τ(1D)/ τ(3D)>>1 – agrees with

Open Questions and Problems:

Our theory – approximate, improvements are needed



1) Correlations between search cycles
2) Correlations in binding positions after dissociation
3) Protein interactions with DNA that have several targets
4) Protein and DNA elasticity
5) DNA conformations
6) Effect of Temperature
7) Sequence dependence

CONCLUSIONS:

- A theoretical approach of how proteins find and recognize their targets on DNA is developed
- **Facilitated diffusion** is a sequence of fast 3D and slow 1D motions
- Acceleration in facilitated diffusion is achieved due to the increased local concentration of adsorbed proteins (from non-specific binding)
- Our results and predictions consistent with basic laws of Chemistry and Physics
- In agreement with all currently available experimental observations

Acknowledgements

- **Financial Support**: National Science Foundation, The Welch Foundation, Hamill Innovation Fund, Humboldt Fellowship
- **Discussions:** Michael E. Fisher, G. Oshanin, A. Berezhkovskii, L. Mirny, B. Shklovskii, A. Grossberg, D. Makarov.

Some of the results are published in

J. Phys. Chem. B **112**, 4741-4750 (2008)

J. Phys. Chem. B **113**, 4242-4247 (2009)