Physics of protein–DNA interactions: mechanisms of facilitated target search

Anatoly B. Kolomeisky

Received 28th September 2010, Accepted 2nd November 2010
DOI: 10.1039/c0cp01966f

One of the most critical aspects of protein–DNA interactions is the ability of protein molecules to quickly find and recognize specific target sequences on DNA. Experimental measurements indicate that the corresponding association rates to few specific sites among large number of non-specific sites are typically large. For some proteins they might be even larger than maximal allowed three-dimensional diffusion rates. Although significant progress in understanding protein search and recognition of targets on DNA has been achieved, detailed mechanisms of these processes are still strongly debated. Here we present a critical review of current theoretical approaches and some experimental observations in this area. Specifically, the role of lowering dimensionality, non-specific interactions, diffusion along the DNA molecules, protein and target sites concentrations, and electrostatic effects are critically analyzed. Possible future directions and outstanding problems are also presented and discussed.

I. Introduction

A starting point of many biological processes is a protein binding to specific target sequences on DNA molecules.[1,2] This process is one of the ways for transferring genetic information contained in DNA. As an example, let us consider a fundamentally important process of transcription, when RNA polymerase (RNAP) enzyme moves along the DNA molecule and synthesizes the RNA molecule which is a corresponding copy of the sequence of bases on the DNA.[1,2] The initial point of transcription is determined by a special sequence of 4 bases, known as TATA box, that is positioned 25 base pairs (bp) ahead of the actual starting position. RNA polymerase itself cannot find this starting point, and it relies on the action of proteins known as transcription factors that search and recognize the TATA box sequence and recruit other proteins to create a special activated complex. After this the RNAP binds to this complex and the transcription starts. The important observation here is the fact that the transcription factor had to find a small target (size of order of 1 nm) along a large DNA chain (typically of order 10^6–10^9 base pairs) fast in order for the transcription and all following biological processes to proceed correctly.

Protein–DNA interactions phenomena have been extensively studied by various experimental techniques in the last 40 years.[3–21] Early kinetic measurements have yielded a very unexpected observation that the association rate for the Lac repressor protein to bind to its target sequence on DNA is close to $k_{exp} \simeq 10^{10}$ M^{-1} s^{-1}.[3] This value is approximately 100–1000 times faster than the maximal solution diffusion rate as specified by a Debye-Smoluchowski theory,[4,5,7,22,23] although it should be mentioned that such very high rates have been observed only for low salt concentrations in the solutions.[3°C5,18] The phenomenon of fast protein search on DNA is called a *facilitated diffusion*.[25] Several other experimental methods beyond classical chemical kinetics methods have been utilized in studies of the protein search for targets on DNA.[9–21] Specifically, electrophoresis and chromatography have been used to analyze products of reactions between proteins and DNA molecules with two target sites and controlled distance between them.[18,21] Recent advances in single-molecule spectroscopy allowed to visualize and quantify with a high precision the motion of fluorescently labeled protein molecules along DNA chains.[12–17,19,20] Most experimental investigations have been performed for *in vitro* conditions with a few studies addressing protein–DNA interactions in living cells.[14]

Surprising experimental results have stimulated serious theoretical efforts to understand physical and chemical aspects

Department of Chemistry, Rice University, Houston, TX 77005, USA

Anatoly B. Kolomeisky

Anatoly Kolomeisky is Professor of Chemistry and Chemical and Biomolecular Engineering at Rice University in Houston, TX, USA (http://python.rice.edu/kolomeisky/). He graduated with a PhD in Chemistry from Cornell University in 1998. Trained as a Theoretical Physical Chemist he is renowned for his work on modelling complex biological and chemical processes using methods of Statistical Mechanics. An author of more than 80 original papers and review articles and several book chapters, he was a recipient of Dreyfus New faculty Award in 2000, NSF CAREER Award in 2002, Sloan Fellowship Award in 2004, Hamill Innovation Award in 2006 and Humboldt Research Fellowship in 2008.

of protein–DNA interactions.[9,18,20,24,26–52] However, although some progress in deciphering microscopic picture of protein search and recognition has been achieved, fundamental mechanisms that govern these processes are still not well understood. Furthermore, there are strong debates in evaluating experimental studies and different theoretical ideas of the facilitated diffusion. This is one of the most fundamental problems in biology that is still not fully resolved. It is interesting to note that some protein–protein association rates, measured again at low salt conditions, are also very large and violate the predictions from Smoluchowski theory of diffusion-controlled reaction rates.[53]

The most intriguing observation in studies of protein–DNA interactions was the fact that the experimentally measured association rate for lac repressor violated three-dimensional diffusional limits.[3] According to a theoretical analysis developed by Debye and Smoluchowski,[22,23] the maximal association rate of two particles A and B is limited by the rate of the reciprocal diffusional approach, *i.e.*, the reaction cannot happen until two molecules collide with each other at least once. The corresponding expression for the maximal rate in the system without long-range interactions is given by

$$k_{\max} = 4\pi(D_A + D_B)R, \qquad (1)$$

where D_A and D_B are diffusion constants for molecules A and B, respectively, and R is the contact distance between the centers of the molecules, the sum of two radii. In normal biological cells DNA molecules are typically much less mobile, suggesting that $D_A \simeq 0$. Since target sites are small (a few base pairs) for Lac repressor proteins one can estimate the contact distance to be of the order of few nanometers and the protein diffusion constant is $D_B \simeq 10^{-11}$ m^2 s^{-1}. Substituting these values into eqn (1) yields the maximal possible association rate $k_{\max} \simeq 10^8$ M^{-1} s^{-1}, which is two orders of magnitude smaller than the rates measured in experiments.[3] There are several other proteins that bind to their targets on DNA faster than allowed by the bulk solution diffusion, and many other proteins have association rates very close to the three-dimensional diffusional limit.[9] However, even these observations suggest that mechanisms of the search is complex and it cannot be explained by the simple 3D diffusion. It is not possible that any collision between the protein molecule and DNA always leads to finding the correct target sequence.

Experimental observations of facilitated diffusion have produced a paradigmal shift in explaining mechanisms of protein–DNA interactions. It is critically important to understand protein search for targets on DNA from microscopic point of view. In this paper I will critically review several existing theoretical ideas and approaches to describe protein search phenomena. Although this review is the result of multiple discussions and collaborations with many colleagues, it mainly represents my subjective personal view of the topic. In addition, only physical aspects of protein–DNA interactions during the search will be addressed, although the biological side of this problem is very rich and mostly unknown.

II. Current theoretical approach

A BWH model

To explain major discrepancies between experimentally measured association rates and expected from bulk solution diffusion estimates for proteins searching for specific targets on DNA Berg, Winter and von Hippel in several seminal papers developed a theoretical model (BWH model) of the process that is currently the most known and widely utilized.[4,5,7] The main underlying idea of this approach is that the search process is a combination of three-dimensional motions in the solution and one-dimensional hoppings along the DNA chain. The facilitated target search is schematically shown in Fig. 1. It is often argued that the main idea of the BWH model is the concept of *lowering of dimensionality* that leads to an acceleration in the search process since the protein bound to the DNA molecule has a higher probability to move in the direction of the target sequence as compared with the situation in the bulk solution.[9,20,29]

The search process in this model is viewed as a sequence of following events. The protein molecule binds non-specifically to the DNA molecule, moves along the DNA contour scanning for the correct sequence, and unbinds if no target is found. After dissociation the protein molecule can bind with equal probability to any site on DNA. The protein performs several such searching cycles before finally attaching to the target. For a single protein the expression for the total search time can be written then as[9,20,29]

$$\tau_s = \frac{L}{\lambda}(\tau_{1D} + \tau_{3D}), \qquad (2)$$

where L is the total contour length of DNA, λ is the average length of DNA that the protein molecule scans during each searching round, and τ_{1D} and τ_{3D} are times spent by the protein molecule on one-dimensional and three-dimensional segments of the search cycle, correspondingly. In eqn (2) τ_s is the average search time provided that every cycle is statistically independent from previous binding cycles. Assuming that the protein molecule moves with diffusion constants D_1 along the

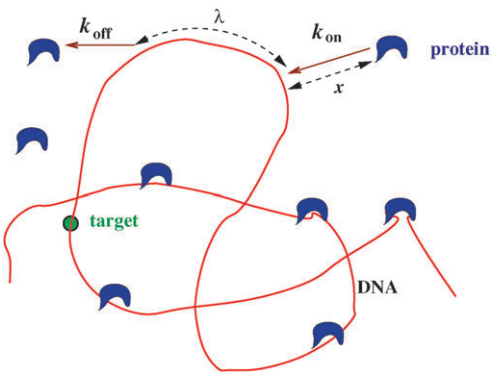

Fig. 1 A general schematic picture of the protein search for the target on DNA. Parameters k_{on} and k_{off} are adsorption and dissociation rates constants for protein molecules; x is the average distance of a protein in the solution from DNA; and λ is the length that each protein at average moves along DNA.

DNA and with D_3 in the bulk solution, the scanning time on the DNA molecule can be simply written as

$$\tau_{1D} = \lambda^2/2D_1. \quad (3)$$

Then the search time in eqn (2) can be presented as a function of τ_{1D},

$$\tau_s = \frac{L}{\sqrt{2D_1}}(\sqrt{\tau_{1D}} + \tau_{3D}/\sqrt{\tau_{1D}}). \quad (4)$$

This leads to an interesting observation that the most optimal search time is reached when $\tau_{1D} = \tau_{3D}$, and it is important to note that this result is valid for general values of 1D and 3D diffusion constants, *i.e.*, when $D_1 \neq D_3$. Then the acceleration in the search for the target of size a by utilizing the combined 3D/1D search mechanism over purely three-dimensional diffusion is equal to $\lambda/2a$. If the scanning length of the protein non-specifically bound to DNA is several hundreds base pairs[9,20] then the association rates to the target sequence can be larger by two orders of magnitude in comparison with Debye-Smoluchowski limiting results, in agreement with kinetic experiments.[3]

This theoretical picture has become a very popular because it is rather simple and physically reasonable, and it was also supported by several experimental observations.[5,9,20] The strongest proof that the protein search is a combination of 3D and 1D motions came from recent single-molecule experiments.[10,15] Using BbvCI restriction enzyme that cuts DNA molecules in the specific sites it was shown by analyzing products of the cutting reaction that the protein does slide along the DNA molecule searching for the appropriate target sequence. It was also shown that the sliding length is not large and it depends on the ionic strength of the surrounding solution.[10] The success of the BWH model of the combined 3D/1D search stimulated many theoretical efforts to extend the approach by including the effects of DNA conformations,[29] protein flexibility[20,26,27] and sequence dependence.[20,26,27,30]

B Critical evaluation

The BWH model and related theoretical approaches present physically reasonable and appealing possible mechanisms of the facilitated diffusion phenomena that were also supported by several experimental investigations[5,10] and some computer simulations.[28,44] However, although this theoretical method allowed to explain many properties of the fast protein search on DNA,[9,20,29] there are many serious problems and contradictions associated with the utilization of this approach to real biological systems.

In most applications of the method it is assumed that diffusion constants for one-dimensional and three-dimensional motions are similar, $D_1 \simeq D_3$.[9] But this assumption is rather unrealistic. It is hard to imagine that the protein molecule moving along the DNA chain (most probably following the helical pathway) and strongly interacting with charged groups in the nucleic acid has the same mobility as a free protein molecule in the solution. Indeed, recent theoretical calculations[54] and single-molecule experimental measurements[13,14,17] suggest that 1D diffusion coefficients are of the order $D_1 \simeq 10^{-13}$–10^{-16} m^2 s^{-1}, which are 100–10 000 smaller than the diffusion constant for the same proteins in the bulk solutions estimated using the Stokes–Einstein relation. In addition, these experiments have shown that the partitioning of the times that the protein spends on DNA and in the solution, $\tau_{1D}/\tau_{3D} > 10$–100, is very different from the most optimal conditions ($\tau_{1D} = \tau_{3D}$) predicted by the BWH model. The protein molecule most of the time stays bound to the DNA where it moves very slowly scanning for the target sequence. It is important to note that under experimentally observed conditions the current theoretical approach does not predict acceleration at all![18,20,35,49] It leads to a surprising conclusion that the 1D/3D combination mechanism is actually slowing down the search significantly, and it cannot explain fast association rates for the protein search on DNA. In addition, these models yield clearly unphysical results in some limiting cases. For example, the BWH model predicts for small concentrations of target sites and/or small concentrations of proteins in the solution that the decrease in the concentration increase the association rate,[9,29] in violation of expectations from basic laws of chemical kinetics where rates are proportional to concentrations (*e.g.*, see Fig. 6 in ref. 9). Thus the current theoretical approach fails to describe correctly protein search for targets on DNA, and new ideas on mechanisms of protein–DNA interactions are needed.

III. Alternative theoretical approaches

The contradictions and serious problems revealed in the application of the original BWH model and related approaches to protein search phenomena have been realized by several research groups,[18,20,33,35,49] and it stimulated a development of alternative theoretical methods for description of facilitated diffusion processes. Several of them will be critically discussed below.

A Electrostatic mechanism

Recently it was suggested[18,44] that observed fast association rates are the result of electrostatic interactions between oppositely charged molecules and they *do not* violate the 3D diffusion limit as widely accepted. This view is based on the fact that the maximal association rate in the Debye-Smoluchowski theory when the reacting molecules have long-range interactions with each other is different from eqn (1),[23]

$$k_{\max} = 4\pi(D_A + D_B)R\beta, \quad (5)$$

where

$$\beta = \frac{1}{\int_R^\infty \frac{e^{U(r)/k_B T}}{r^2}\mathrm{d}r}, \quad (6)$$

and $U(r)$ is the intermolecular potential of interactions. For non-interacting particles $U(r) = 0$ for $r > R$ and the original expression (1) is recovered. The long-range attractive interactions can significantly speed up the diffusional fluxes, yielding larger collisional rates and increasing the diffusion limit. It was argued[18] that the original kinetic experiments[3] have been performed at the "low-salt" conditions and electrostatic effects had to be properly taken into account in

estimating the diffusion limit. To test the validity of this argument one could evaluate the Debye length which determines the region beyond which the electrostatic interactions are screened and can be neglected. The Lac repressor association rates measurements have been performed in a buffer solution containing 10 mM concentrations of KCl, 10 mM magnesium acetate and 10 mM Tris/HCL. In this case it can be shown that the Debye length is of the order $\lambda_D \simeq 1$–2 nm,[35] which is comparable to the target size on DNA. It leads to the conclusion that the protein and DNA start to feel electrostatics only when they collide with each other. Then one might conclude that electrostatic forces, although critically important for protein–DNA interactions,[4,5,33] do not modify the diffusion limit and they cannot explain fast protein search dynamics as proposed in ref. 18. Explanation of the facilitated diffusion most probably requires a different mechanism.

B Colocalization mechanism

In the current theoretical view of the facilitated diffusion the slow search is the result of many cycles of 3D wondering and 1D scanning. The search could be accelerated significantly if the number of such cycles is reduced. This is the main idea of the colocalization mechanism proposed by L. Mirny and coworkers.[20,40] It is argued that in bacteria genes responsible for producing specific proteins are found close to targets to which these proteins must bind. This colocalization mechanism suggests that proteins are produced near their binding sites on DNA, and it allows to significantly accelerate the search by eliminating repeating search cycles. The proteins are already produced spatially close to target sequences on DNA. Using simulations and genomic analysis the possibility of colocalization mechanism was argued for transcription factor proteins in bacteria.[20,40]

The colocalization mechanism provides a very elegant way of explaining fast protein search processes. However, it might not work for eukaryote organisms where transcription and translation processes are separated in time and space.[1,2] In addition, it cannot explain processes where proteins have multiple targets on DNA. Furthermore, the original kinetic experiments on Lac repressors[3] have been done *in vitro* which probably rules out the application of the colocalization mechanism here. Thus the colocalization mechanism does not allow to explain the fundamental origins of the facilitated diffusion in all biological systems.

C Correlation mechanism

The original BWH model can be considered as uncorrelated mechanism of facilitated diffusion since in this picture 3D and 1D motions are totally uncoupled: the protein after the dissociation from DNA can bind with equal probability to any site on the DNA. Given a complex structure of biological systems this might not be the adequate way to properly describe protein search phenomena. Recently a different theoretical approach, that takes into account the correlations between one-dimensional scanning and three-dimensional motion and the effect of non-specific interactions,[31–35] has been introduced. Related ideas of correlated re-associations have also been discussed by L. Mirny and colleagues.[20]

Similarly to the BWH model, reaching the target on DNA is viewed here as a sequence of searching cycles: see Fig. 1. Each protein on average binds and unbinds several times before finding the target. The search cycles for each protein molecule consists of 3D and 1D segments. No assumption of equilibrium with respect to protein binding/unbinding is made.[33,35] Also, in contrast to the BWH model, the correlations are explicitly included in this approach by making an assumption that the motion in the three-dimensional segment of length x can be viewed as an effective one-dimensional motion with properly rescaled diffusion constant. This approximation is critical since it reflects the tendency of the just dissociated protein to return back to the same position on DNA because of non-specific interactions. Then the problem of finding the search time for one cycle reduces to finding a mean first-passage time for a one-dimensional system that consists of 2 sequential segments with different particle diffusivities,[33,55,56]

$$\tau_{1C} = \int_0^{x+\lambda} \frac{\exp[\beta G(z)]}{D(z)} \mathrm{d}z \int_0^{z} \exp[-\beta G(z')] dz'. \qquad (7)$$

In this equation we have $\beta = 1/k_B T$, $G(z)$ as a free energy of the protein at the position z, and $D(z)$ as the diffusion constant for the protein molecule that depends on the spatial position of the particle,

$$D(z) = \begin{cases} D_3, & 0 < z < x, \\ D_1, & x < z < x + \lambda, \end{cases} \qquad (8)$$

where D_3 and D_1 are 3D and 1D protein diffusion constants, respectively.

The strength of non-specific interaction between DNA and the protein is given by energy E_{ads}, that also determines the equilibrium constant for adsorption with binding and unbinding rates,[33]

$$y = \frac{k_{on}}{k_{off}} = \exp\left(\frac{E_{ads}}{k_B T}\right). \qquad (9)$$

It is assumed then that the free energy of the protein molecule in the solution is zero, while the protein molecule non-specifically bound to DNA has a free energy ($-E_{eff}$) given by

$$y_{eff} = \frac{k_{on} c_p}{k_{off} c_{ads}} = \exp\left(\frac{E_{eff}}{k_B T}\right), \qquad (10)$$

with c_p and c_{ads} describing the concentration of proteins in the solution and adsorbed to DNA, respectively. It is important to note the difference between E_{ads} and E_{eff}. E_{ads} describes the standard free energy difference for the protein to be associated to DNA or to be found free in the solution, while E_{eff} presents a real free energy difference that depends on concentrations of protein molecules in different states. The equilibrium is reached when $E_{eff} = 0$.

Combining all arguments presented above the search time for one cycle can be easily evaluated using eqn (7),

$$\tau_{1C} = \frac{x^2}{2D_3} + \frac{\lambda^2}{2D_1} + \frac{x\lambda}{D_1 y_{eff}}. \qquad (11)$$

This equation has a clear physical meaning. The first two terms describe times spent by the protein molecule only in the solution or only on DNA. The last term reflects the

stochasticity in the motion of the particle by including temporal contributions from trajectories when the protein goes back and force between 3D and 1D segments before finally reaching the end.[33,35] It effectively couples two dynamical modes of the system. Another way to understand the correlation term is to recall that the search time for one cycle is an average quantity over many trajectories, and there is a significant portion of them include several forward-backward crossings between 3D and 1D segments, effectively increasing the time to reach the end of the cycle. The BWH model and related approaches completely neglect this contribution and it leads to unphysical behavior discussed above.

To simplify calculations the volume V around one DNA molecule of length L and radius r with one target site on it is considered. It allows us to present proteins concentrations *via*

$$c_p = n_p/V,\ c_{\text{ads}} = n_{\text{ads}}/V, \tag{12}$$

where n_p is the number of free proteins in the solution around the DNA molecule and n_{ads} represents the number of proteins adsorbed on the DNA. Eqn (11) provides an explicit expression for the time of one search cycle. To estimate the average number of cycles before the protein reaches the target it can be argued that the length scanned by *all* adsorbed proteins during one cycle is equal to λn_{ads}. One can introduce the probability p that the target will be found after binding to DNA of length L, and it is given by $p = \lambda n_{\text{ads}}/L$. Then the probability that the target is reached after te jth cycle is

$$S_j = p(1-p)^{j-1}. \tag{13}$$

The average number of cycles can be easily evaluated,

$$\langle j\rangle = \sum_{j=1}^{\infty} jS_j = 1/p = \frac{L}{\lambda n_{\text{ads}}}. \tag{14}$$

This results suggests that on average the whole length of DNA must be scanned before successfully reaching for the target. It produces the expression for the total search time,

$$\tau = \tfrac{L}{\lambda n_{\text{ads}}}\tau_{1C}$$

which can be explicitly written as,[33]

$$\tau = \frac{Lr}{2D_3 n_p}\left(\frac{r}{\lambda n_{\text{ads}}} + \frac{\lambda n_p}{r n_{\text{ads}} d} + \frac{2}{yd\sqrt{n_p}}\right), \tag{15}$$

where $d = D_1/D_3$ is a dimensionless ratio of diffusion coefficients. To test if this correlation mechanism allows for acceleration in the search in comparison with simple three-dimensional search the relative search time is analyzed,

$$\frac{\tau}{\tau_S} = \frac{a}{r}\left(\frac{1}{\sqrt{n_{\text{ads}} yd}} + \frac{n_p\sqrt{y}}{n_{\text{ads}}^{3/2}\sqrt{d}} + \frac{2}{yd\sqrt{n_p}}\right), \tag{16}$$

with τ_S corresponding to a purely three-dimensional search given by the Debye-Smoluchowski theory,

$$\tau_S = \frac{1}{2D_3 a c_p} = \frac{Lr^2}{2D_3 a n_p}. \tag{17}$$

This theoretical model predicts that the acceleration in the protein search could be achieved for some ranges of parameters, as shown in Fig. 2 and 3. The effectiveness of

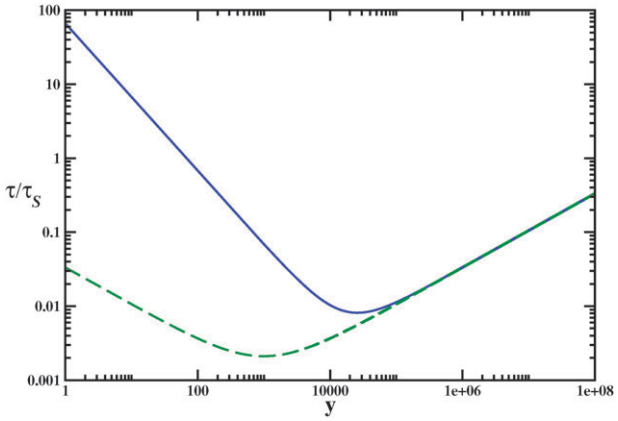

Fig. 2 Relative search time as a function of the adsorption equilibrium constant y that measures the non-specific interaction strength for $a = 1$ nm, $r = 30$ nm, $L = 1$ μm, $n_p = 1$, $n_{\text{ads}} = 1000$ and $d = 0.001$. Solid curve corresponds to the correlated mechanism while the dashed line describes the uncorrelated mechanism.

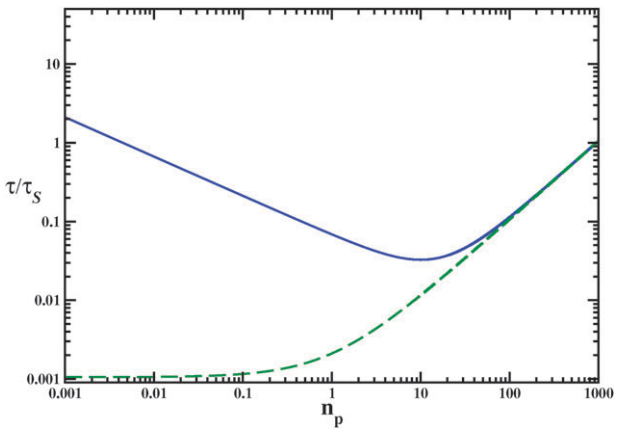

Fig. 3 Relative search time as a function of the concentration of free proteins in the solution for $n_{\text{ads}} = 1000$, $y = 1000$ and $d = 0.001$. Solid curve corresponds to the correlated mechanism while the dashed line describes the uncorrelated mechanism.

the search process strongly depends on the strength of non-specific interaction. Typical cellular conditions, when E_{ads} is between 3 and 8 k_BT, provide a range of parameters when the facilitated diffusion is faster than the ordinary Debye-Smoluchowski three-dimensional search. The non-monotonous dependence of the relative search time, as shown in Fig. 2, can be qualitatively explained using the following arguments. Small values of y correspond to weak attraction between the protein molecule and the DNA chain. Then the protein does not spend enough time scanning DNA for the target and many cycles are needed before the target can be reached. In this case 3D search seems to be more effective. For $y \gg 1$ the attraction to the DNA is strong and the protein molecule spends most of the search time here. But the motion on DNA is very slow, and it makes the Debye-Smoluchowski mechanism faster. Only for intermediate values of equilibrium constant y the facilitated mechanism wins over the simple 3D search. It should be noted that both correlated and uncorrelated mechanisms predict qualitatively similar pictures (see Fig. 2),

although the uncorrelated model predicts larger accelerations. As expected, both theoretical methods become almost the same for large interactions strength where the effect of correlations is negligible. Predictions from both theoretical approaches agree, at least qualitatively, with experimental observations in which the association rates have been modified by changing the ionic strength of the buffer solutions.[5]

The facilitated diffusion also depends on the number of free proteins in the solution as presented in Fig. 3. The correlated mechanism predicts that the search can be optimal only for intermediate range of concentrations. This prediction is very different from the uncorrelated mechanism result, as shown in Fig. 3, where a monotonous increase in the search times is always observed. For low concentrations the uncorrelated mechanism suggests that fastest search is achieved when there are no proteins in the solution at all which is rather unphysical. The correlated mechanism views the low concentration regime differently. Here reaching the DNA molecule is a rate-limiting step, and in the Debye-Smoluchowski mechanisms it has to be done only once, while in the correlated mechanism many cycles should be performed. It makes the search process that combines 3D and 1D motions slower. At large protein concentrations both uncorrelated and correlated mechanisms agree, and dynamics here is explained by the fact that the concentration of proteins becomes so large that it is faster to reach the target directly *via* 3D motion. The increase in the concentration of free proteins lowers the effect of correlations because the free energy difference between two states for proteins is larger.

To better understand correlation mechanism of the facilitated search one can analyze contributions of different terms in the overall search times as functions of relevant parameters as presented in Fig. 4 and 5. As shown in Fig. 4, the absolute value of the ratio of diffusion constants $d = D_1/D_3$ determines the fate of correlations: for smaller (more realistic) values of d protein is found mostly in the mode of binding and unbinding events, while for large d the effect of correlation is getting smaller. Similar effect is observed for the strength of non-specific interactions: see Fig. 5. For small values of equilibrium constant y the correlation term dominates, and the role of correlations is negligible at strong non-specific interactions. This is because once the protein goes to DNA it is energetically unfavorable to dissociate, limiting binding/unbinding events.

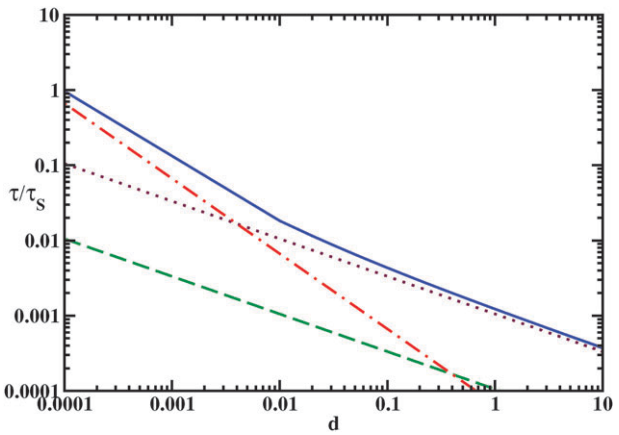

Fig. 4 Relative search time as a function of the ratio of diffusion constants and contributions of different terms in the overall search for $n_{ads} = 100$, $n_p = 1$ and $y = 1000$. Solid curve is the total search time, the dashed lined is for the time in 3D, the dotted line corresponds to the time in 1D, and the dash-dotted line describes the correlation term.

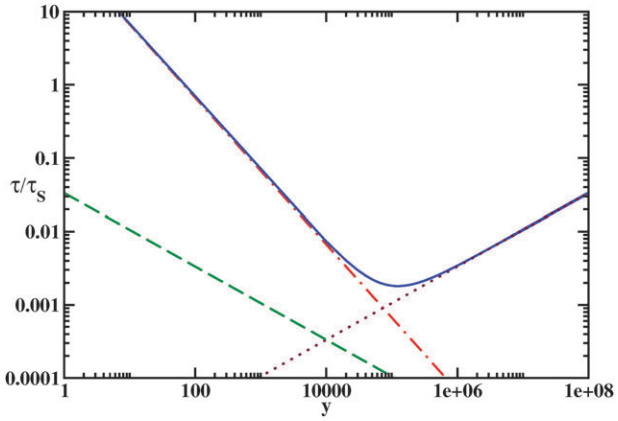

Fig. 5 Relative contributions to overall search time as a function of the adsorption equilibrium constant y for $n_{ads} = 1000$, $n_p = 1$ and $d = 0.001$. Solid curve is the total search time, the dashed lined is for the time in 3D, the dotted line corresponds to the time in 1D, and the dash-dotted line describes the correlation term.

The correlation mechanism also predict that the average length for each protein is a complex function of system parameters,[33]

$$\lambda = \frac{r\sqrt{yd}}{\sqrt{n_{ads}}}. \qquad (18)$$

The dependence of this scanning length on the strength of non-specific interactions is shown in Fig. 6. As expected, the stronger these interactions the longer distances the protein scans along the DNA molecule. One can easily see (comparing with Fig. 5) that at the most optimal conditions for given set of parameters we have $\lambda \simeq 10$–20 nm, which is smaller than the sliding length of order of 500 nm measured for 1D lac repressor diffusion in ref. 13, but it is much closer to the value of $\simeq 15$ nm (50 bp) measured in single-molecule experiments

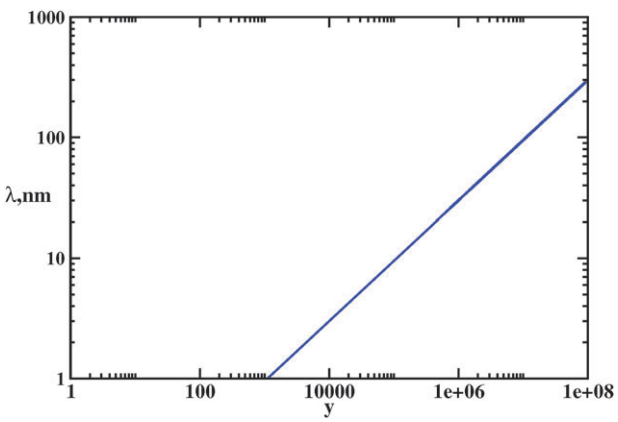

Fig. 6 Average scanning length λ as a function of the adsorption equilibrium constant y for $n_{ads} = 1000$, $d = 0.001$ and $r = 30$ nm.

on protein translocation along DNA by BbvCI restriction enzymes.[10]

One of the most important predictions from the correlated mechanism is that the protein spends most of the time on DNA, and still the acceleration in the search can be achieved for some sets of parameters, including the most optimal search conditions (see Fig. 5).[33] It agrees well with all available single-molecule measurements,[13,14] as well as with extensive Monte Carlo computer simulations.[35] Another important observation from the correlated mechanism is that the facilitated diffusion is most probably a non-equilibrium phenomenon, and the fast search cannot be achieved for equilibrium conditions with respect to non-specific binding/unbinding processes.[33,35] This might explain how nature accelerates search processes in real living systems.

The correlated approach provides a comprehensive description of the facilitated diffusion that is also consistent with all available experimental observations and relevant Monte Carlo computer simulations. It suggests that the proteins search is the non-equilibrium process where correlations between fast 3D and slow 1D motions are very important. The main idea of the correlated mechanism is the critical role of non-specific interactions for the facilitated diffusion. Due to non-specific binding proteins move slower along the DNA and it slows down the search. However, non-specific interactions make more proteins bound to DNA and because all of them search in parallel it accelerates the search. The increase in the local concentration of bound proteins sometimes might win over the slower one-dimensional motion in the protein search for target sequences on DNA.

Although the correlated mechanism seems to be the most successful in explaining complex processes associated with fast protein search on DNA, as judged by comparison with experimental observations and numerical simulations, it has also several problems. First of all, in theoretical calculations[33] the correlations between 3D and 1D motions are made artificially much stronger that one would expect in reality by substituting three-dimensional excursions with effective 1D segments. Second, the model implicitly assumes that the search takes place faster than the relaxation to equilibrium for binding to DNA. In other words, this theoretical approach is working when the protein finds the target faster than protein concentrations change in the system. It remains an open question to test the validity of these assumptions in experiments and in computer simulations.

In the parameters utilized to illustrate different mechanisms of protein search on DNA, as shown in Fig. 3–6, it is assumed that the length of available DNA L is of the order of 1 μm, the energy of non-specific protein–DNA binding is between 3 and 8 k_BT and the concentration of proteins is in the micromolar–millimolar range. Based on our knowledge of cellular systems they provide a reasonable and realistic description. However, it should be noted also that most theoretical models discussed in this work are approximate scaling approaches.

IV. Summary and conclusions

Experimental observations that association rates for some proteins searching for targets on DNA are larger than allowed by three-dimensional diffusion limit have stimulated strong discussions on mechanisms of the facilitated diffusion. We have analyzed critically theoretical ideas utilized in the analysis of fast protein search processes on DNA. The approach that is currently widely used and which is based on the BWH model argues that the search is a combination of three-dimensional motion and one-dimensional hoppings. It assumes no correlation between two dynamical modes in the search process. According to these models the main reason for acceleration is lowering of the dimensionality in the protein motion. Although this theoretical approach was able to explain some qualitative features of the facilitated diffusion and even predict a sufficient speed of search for some range of parameters, the detailed theoretical analysis indicates that it does not predict any acceleration in the search for real biological systems unless the parallel scanning of many proteins is taken into account. In addition, it shows unphysical behavior at some range of parameters. These problems and contradictions stimulated several alternative theoretical models for the facilitated diffusion. One approach is arguing that there is no violation of the diffusion limit because electrostatic interactions between proteins and DNA have not been properly taken into account. However, the calculation of the Debye length, which specifies the region in which the electrostatic forces are important, for relevant experimental conditions indicates that it is rather small and it is similar to the size of the target sequence, eliminating electrostatic forces as a possible source of acceleration. Another proposed theoretical method is a colocalization mechanism which suggests that proteins are made near the target sites, and this increases the search speed by not making many search cycles. But this mechanism might not work for eukariotic organisms where processes of making proteins and searching for the targets are spatially and temporally separated. It also cannot explain protein search processes on DNA with many targets. A different alternative approach that is based on taking explicitly into account the effect of correlations between 3D and 1D motions has also been proposed. The method argues that non-specific interactions are the most critical part of the protein search on DNA. This theoretical approach allows accelerations for the search for realistic biological parameters as observed in experiments, it explains all available experimental observations, it agrees with Monte Carlo computer simulations and it also does not make any unphysical predictions. Although the correlated mechanism seems to be the most successful theoretical approach in explaining the facilitated diffusion, it also has several questionable assumptions and approximations (such as strong correlations and non-equilibrium nature of the involved processes) that should be tested more carefully in experimental studies and in numerical simulations.

The comparison of different theoretical ideas related to the protein search on DNA suggest that, although a significant progress in our understanding of this aspect of protein–DNA interactions has been achieved, we still do not understand fully the complex phenomena associated with the facilitated diffusion. New ideas and new methods of investigations are needed. It seems that computer simulations might provide critical information on mechanisms of protein search at the

atomistic level.[41,42,51] There are many open questions and problems that should be addressed in the future. For example, how does the facilitated diffusion mechanism change in the crowded cellular environment where it is not realistic to describe 3D diffusion as a free bulk solution process?[20] How will search dynamics be modified in real biological systems with varying DNA density? Another important problem is to understand the mechanism of primary recognition, *i.e.* when the protein approaches to the right sequence how does it distinguish it from other sequences? Recently it was suggested that the complementarity of the charge patterns on a target DNA sequence and on the protein might result in electrostatic recognition.[33,34] It is also important to investigate the role of multi-particle cooperativity in the facilitated diffusion. There are indications that it might lead to directionality in the search process.[50] It is reasonable to suggest that to better understand protein–DNA interactions it will be critically important to test different theoretical ideas regarding facilitated diffusion with single-molecule experiments and extensive computer simulations.

Acknowledgements

The author would like to acknowledge the support from the Welch Foundation (Grant No. C-1559), the U.S. National Science Foundation (Grant No. ECCS-0708765) and the U.S. National Institute of Health (Grant No. R01GM094489). The author also would like to thank A.A. Kornyshev, G.T. Barkema, A.G. Cherstvy, M.E. Fisher, G. Oshanin, L. Mirny, B. Shklovskii, A. Grossberg, D. Makarov, R. K. Das and A. Dinner for collaboration, useful discussions and technical help.

References

1 B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, Garland Science, 4th edn, 2002.
2 H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore and J. Darnell, *Molecular Cell Biology*, W.H. Freeman and Company, New York, 4th edn, 2002.
3 A. Riggs, S. Bourgeois and M. Cohn, *J. Mol. Biol.*, 1970, **53**, 401.
4 O. G. Berg, R. B. Winter and P. H. von Hippel, *Biochemistry*, 1981, **20**, 6929.
5 R. B. Winter and P. H. von Hippel, *Biochemistry*, 1981, **20**, 6948.
6 M. D. Barkley, *Biochemistry*, 1981, **20**, 3833.
7 O. G. Berg and P. H. von Hippel, *Annu. Rev. Biophys. Biophys. Chem.*, 1985, **14**, 131.
8 N. P. Stanford, M. D. Szczelkun, J. F. Marko and S. E. Halford, *EMBO J.*, 2000, **19**, 6546.
9 S. Halford and J. Marko, *Nucleic Acids Res.*, 2004, **32**, 3040.
10 D. M. Gowers, G. G. Wilson and S. E. Halford, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15883.
11 J. Iwahara, M. Zweckstetter and G. M. Clore, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15062.
12 P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine and X. S. Xie, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 5752.
13 Y. M. Wang, R. H. Austin and E. C. Cox, *Phys. Rev. Lett.*, 2006, **97**, 048302.
14 J. Elf, G.-W. Li and X. S. Xie, *Science*, 2007, **316**, 1191.
15 J. Gorman and E. C. Greene, *Nat. Struct. Mol. Biol.*, 2008, **15**, 768.
16 B. van der Broek, M. A. Lomholt, S.-M. J. Kalisch, R. Metzler and G. J. L. Wuite, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **105**, 157388.
17 P. C. Blainey, G. Luo, S. C. Kou, W. F. Mangel, G. L. Verdine, B. Bagchi and X. S. Xie, *Nat. Struct. Mol. Biol.*, 2009, **16**, 1224.
18 S. E. Halford, *Biochem. Soc. Trans.*, 2009, **37**, 343.
19 A. Tafvizi, F. Huang, J. S. Leith, A. R. Fersht, L. A. Mirny and A. M. van Oijen, *Biophys. J.*, 2008, **95**, L01.
20 L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith and A. Kosmrlj, *J. Phys. A: Math. Theor.*, 2009, **42**, 434013.
21 D. C. Rau and N. Y. Sidorova, *J. Mol. Biol.*, 2010, **395**, 408.
22 H. Gutfreund, *Kinetics for the Life Sciences*, Cambridge University Press, 1995.
23 P. L. Houston, *Chemical Kinetics and Reaction Dynamics*, McGraw Hill, New York, 2001.
24 R. F. Bruinsma, *Physica A*, 2002, **313**, 211.
25 P. H. von Hippel and O. G. Berg, *J. Biol. Chem.*, 1989, **264**, 675.
26 M. Slutsky and L. Mirny, *Biophys. J.*, 2004, **87**, 4021.
27 M. Slutsky, M. Kardar and L. Mirny, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **69**, 061903.
28 K. Klenin, H. Merlitz, J. Langowski and C.-X. Wu, *Phys. Rev. Lett.*, 2006, **96**, 018104.
29 T. Hu, A. Grosberg and B. Shklovskii, *Biophys. J.*, 2006, **90**, 2731.
30 T. Hu and B. I. Shklovskii, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**, 021903.
31 H.-X. Zhou and A. Szabo, *Phys. Rev. Lett.*, 2004, **93**, 178101.
32 H.-X. Zhou, *Biophys. J.*, 2005, **88**, 1608.
33 A. G. Cherstvy, A. B. Kolomeisky and A. A. Kornyshev, *J. Phys. Chem. B*, 2008, **112**, 4741.
34 A. G. Cherstvy, *J. Phys. Chem. B*, 2009, **113**, 4242.
35 R. K. Das and A. B. Kolomeisky, *Phys. Chem. Chem. Phys.*, 2010, **12**, 2999.
36 U. Gerland, J. D. Moroz and T. Hwa, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12015.
37 M. Coppey, O. Benichou, R. Voituriez and M. Moreau, *Biophys. J.*, 2004, **87**, 1640.
38 C. Loverdo, O. Benichou, R. Voituriez, A. Biebricher, I. Bonnet and P. Desbiolles, *Phys. Rev. Lett.*, 2009, **102**, 188101.
39 O. Benichou, Y. Kafri, M. Sheinman and R. Voituriez, *Phys. Rev. Lett.*, 2009, **103**, 138102.
40 G. Kolesov, Z. Wunderlich, O. N. Laikova, M. S. Gelfand and L. A. Mirny, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13948.
41 O. Givaty and Y. Levy, *J. Mol. Biol.*, 2009, **385**, 1087.
42 D. Vuzman, M. Polonsky and Y. Levy, *Biophys. J.*, 2010, **99**, 1202.
43 M. A. D. de la Rosa, E. F. Koslover, P. J. Mulligan and A. J. Spakowitz, *Biophys. J.*, 2010, **98**, 2943.
44 A.-M. Florescu and M. Joyeux, *J. Phys. Chem. A*, 2010, **114**, 9662.
45 I. Eliazar, T. Koren and J. Klafter, *J. Phys.: Condens. Matter*, 2007, **19**, 065140.
46 J. D. Eaves and D. R. Reichman, *J. Phys. Chem. B*, 2008, **112**, 4283.
47 I. M. Sokolov, R. Metzler, K. Pant and M. C. Williams, *Biophys. J.*, 2005, **89**, 895.
48 G. Guigas and M. Weiss, *Biophys. J.*, 2008, **94**, 90.
49 P.-W. Fok, C.-L. Guo and T. Chou, *J. Chem. Phys.*, 2008, **129**, 235101.
50 T. Zhao and A. R. Dinner, *Biophys. J.*, 2008, **94**, 47.
51 J. Hu, A. Ma and A. R. Dinner, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4615.
52 Y. Lin, T. Zhao, X. Jian, Z. Farooqui, X. Qu, C. He, A. R. Dinner and N. F. Scherer, *Biophys. J.*, 2009, **96**, 1911.
53 G. Schreiber, G. Haran and H.-X. Zhou, *Chem. Rev.*, 2009, **109**, 839.
54 B. Bagchi, P. C. Blainey and X. S. Xie, *J. Phys. Chem. B*, 2008, **112**, 6282.
55 S. Redner, *A Guide to First-Passage Processes*, Cambridge University Press, 2001.
56 N. G. van Kampen, *Stochastic Processes in Chemistry and Physics*, North Holland, Amsterdam, 1992.