

# Quantifying the two-state facilitated diffusion model of protein–DNA interactions

Itai Leven and Yaakov Levy<sup>ID\*</sup>

Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Received March 16, 2019; Revised April 13, 2019; Editorial Decision April 15, 2019; Accepted April 17, 2019

## ABSTRACT

**The current report extends the facilitated diffusion model to account for conflict between the search and recognition binding modes adopted by DNA-binding proteins (DBPs) as they search DNA and subsequently recognize and bind to their specific binding site. The speed of the search dynamics is governed by the energetic ruggedness of the protein–DNA landscape, whereas the rate for the recognition process is mostly dictated by the free energy barrier for the transition between the DBP’s search and recognition binding modes. We show that these two modes are negatively coupled, such that fast 1D sliding and rapid target site recognition probabilities are unlikely to coexist. Thus, a tradeoff occurs between optimizing the timescales for finding and binding the target site. We find that these two kinetic properties can be balanced to produce a fast timescale for the total target search and recognition process by optimizing frustration. Quantification of the facilitated diffusion model by including a frustration term enables it to explain several experimental observations concerning search and recognition speeds. The extended model captures experimental estimate of the energetic ruggedness of the protein–DNA landscape and predicts how various molecular properties of protein–DNA binding affect recognition kinetics. Particularly, point mutations may change the frustration and so affect protein association with DNA, thus providing a means to modulate protein–DNA affinity by manipulating the protein’s association or dissociation reactions.**

## INTRODUCTION

DNA-binding proteins (DBPs) possess a remarkably efficient ability to search and recognize their specific binding sites embedded within the genomic DNA. Early experiments showed that DBPs can find their target site at a rate two-orders of magnitude faster than the 3D diffusion limit

(1). Following these results, it was suggested that DBPs accelerate the search for their target site via a mechanism of facilitated diffusion in which 1D diffusion alternates with 3D diffusion during the search (2,3). 1D diffusion itself comprises two distinct search modes, sliding and hopping, which differ in the degree to which translocation and rotation along the DNA are coupled as well as in the dependency of their corresponding diffusion coefficients on salt concentration (4–6). The understanding that DBP search of DNA proceeds through a combination of diffusions in different dimensional spaces is referred to as the facilitated diffusion model. This model is well supported by numerous experimental and theoretical studies and is able to explain the high target association rates observed *in vitro* and in the cell (7–13).

The biophysical characteristics of DNA search may depend on the molecular properties of the searching proteins. For example, the dimensions of the protein, its oligomeric state, electrostatic potential and degree of flexibility may affect the relative usage of the different search modes (14–19). The DNA sequence may also affect search speed by producing different DNA geometries and consequently affecting the protein’s ability to interact with the DNA major groove and thus its ability to perform coupled rotation–translation diffusion (20,21). The energy landscape for 1D diffusion along DNA may also be affected by the DNA sequence, which can affect the ruggedness of the potential energy landscape and, consequently, friction in protein–DNA interactions (9,22,23).

The 1D diffusion coefficient of proteins on DNA has been expressed as  $D = G(T, R, \eta) \cdot F(\sigma)$ , where  $G$  is a term that depends on the temperature,  $T$ , the size of the protein,  $R$ , and the viscosity of the solution,  $\eta$  (10).  $F$  is a term that represents the ruggedness of the potential energy landscape. Assuming that the ruggedness of the potential energy follows a Gaussian distribution, then  $F(\epsilon) = \exp[-(\sigma/k_B T)^2]$ , where  $\sigma$  denotes the variance in the protein–DNA sliding potential and is related to the average energetic barrier for sliding (9,10,23). It was estimated that the ruggedness of the protein–DNA landscape must be low ( $\sigma < 2k_B T$ ) to achieve reasonable association rates (9).

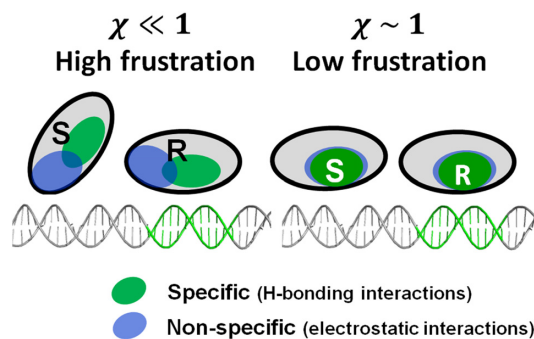
Although many studies aimed to biophysically characterize the mechanism of facilitated diffusion that proteins

\*To whom correspondence should be addressed. Tel: +972 8 9344587; Email: koby.levy@weizmann.ac.il

adopt when searching for their target sites (2,7,24–31), it is clear that the kinetics of many protein–DNA recognition interactions does not depend solely on the search speed. Recognition requires not only finding the target site, but also specifically binding to it. The time-scale for specific binding may affect the kinetics of protein–DNA recognition because finding the site does not guarantee immediate binding. The discrepancy between finding and binding the target site is related to the different types of interactions used for the search and recognition modes. In the search mode (designated here as the *S* state), a DBP interacts with DNA non-specifically through mostly electrostatic interactions between positively charged protein residues and negatively charged phosphates on the DNA backbone (32,33). Specific binding upon reaching the target site requires the protein to switch to its recognition mode (designated here as *R* state) by forming sequence specific contacts with the DNA, supported by hydrogen bonding between the protein residues and the DNA bases (34). Thus, the full search kinetics involves not only searching in the non-specific binding mode, *S*, but also switching to the specific binding mode, *R*, upon recognition of the cognate binding site (11,23,35–39). It was argued that the existence of the *S* and *R* binding modes is necessary to solve an apparent conflict between speed and stability whereby the conditions for fast search (which requires  $\sigma < 2k_B T$ ) are incompatible with stable protein–DNA interactions (which requires  $\sigma > 5k_B T$ ) (9,23). This two-state model is supported by various experimental approaches (including X-ray crystallography, NMR and single-molecule techniques) indicating that numerous DBPs adopt different conformations for their specific versus non-specific interactions with DNA (8,14,23,40–44). The existence of two states can also be inferred from the low roughness for sliding that was found for various proteins (10,45), which is unlikely for the *R* states.

The existence of the *S* and *R* binding modes suggests the presence of an energetic barrier governing the transition between them and therefore a separation of time scale between finding and binding the target site. This implication is supported by single-molecule experiments conducted in living cells on the lac repressor, which displays facilitated diffusion characterized by a low sliding energy barrier of  $\varepsilon \sim 1.0 k_B T$  and a scanning length of  $45 \pm 10$  base pairs for each 1D search round (8). Surprisingly, the lac repressor was found to slide numerous times over its promoter before target recognition is achieved, which reflects the existence of a barrier that needs to be overcome for specific recognition to occur.

Despite being insightful, the two-state model for protein–DNA recognition is still not quantitative. In particular, the interplay between the speed of sliding (via the *S* mode) and the recognition rate (transition from the *S* to *R* mode) is unclear. While some crystal structures of protein–DNA complexes reveal conformational changes associated with the transition from non-specific to specific binding (32,33,46), other structures indicate high similarity between the non-specific and the specific binding modes (47,48). The latter scenario describes reactions with much smaller barriers for recognition in comparison with the former scenario. One may therefore ask: what are the kinetic consequences for sliding dynamics and what is the overall kinetics when the



**Figure 1.** Schematic representation of the search and recognition modes of a DNA-binding protein on DNA. The degree of similarity between the search and recognition binding modes (designated as *S* and *R*, respectively) is governed by the extent of overlap between the surface patches the DNA-binding protein uses to interact with non-specific and specific DNA residues. Non-specific binding is dictated by electrostatic interactions (indicated by the blue ellipse), whereas specific binding is dictated by a set of hydrogen bonds that can be identified from the X-ray structures (indicated by the green ellipse). The overlap between the *S* and *R* binding modes is quantified by the similarity index,  $\chi$ . A smaller overlap between the non-specific and specific patches (i.e.  $\chi \ll 1$ ) corresponds to high frustration and suggests that the protein needs to undergo a conformational change (which potentially involves a change in the DNA conformation as well) to make the *S*  $\rightarrow$  *R* transition and so bind its DNA target site (left). A high value of  $\chi$  ( $\sim 1$ ) reflects less frustration between the *S* and *R* states and suggests that these two states are very similar (right).

barriers between the *S* and *R* binding modes are small? Furthermore, is the rate limiting step in protein–DNA recognition the search kinetics or the transition barrier from *S* to *R*?

In recent studies, we introduced the concept of molecular frustration denoted  $\chi$  between the non-specific and specific protein–DNA binding modes (40,41,49). Frustration is quantified as the degree of overlap between the protein surface patches that are used for the *S* and *R* binding modes. For a given protein, the *S* mode is represented by the largest positively charged patch on the protein surface and the *R* mode is given by its X-ray structure. Greater frustration corresponds to a smaller overlap between the *S* and *R* binding modes (Figure 1). Frustration between the protein residues forming specific versus non-specific contacts with the DNA creates means that the two binding modes represent different energetic and conformational states, so creating the two-state model (40,41). It was shown that numerous DBPs, and particularly enzymes, have a medium to high degree of frustration between their *S* and *R* binding modes. Furthermore, coarse-grained molecular dynamics simulations qualitatively showed that frustration strongly influences the kinetics of the protein–DNA search process, such that high (low) frustration is associated with fast (slow) sliding but a poor (high) target recognition probability (41).

In this paper, we show how frustration links together two of the most prominent aspects of the protein–DNA recognition process, namely, the kinetics of finding the target site via the facilitated diffusion mechanism and the kinetics of binding the target, which depend on the target recognition probability. We extend Slutsky and Mirny’s (9) facilitated diffusion theory by introducing the concept of frustration between the non-specific and specific binding modes and its

effect on DNA recognition kinetics. Utilizing the new theoretical model, we elucidate and enable quantification of various factors influencing protein–DNA recognition kinetics.

## MATERIALS AND METHODS

### Theoretical model

We study the kinetics of protein–DNA recognition by building on Slutsky and Mirny (9,50) facilitated diffusion theory by introducing explicit terms for the  $S$  to  $R$  transition. In the framework of this theory, a single DBP searches for its target site on a long DNA of  $M$  bps through rounds of 1D and 3D diffusion. The mean durations of the 1D and 3D diffusion are  $\tau_{1D}$  and  $\tau_{3D}$ , respectively. The number of 1D and 3D rounds needed to find the target site can be estimated as  $M/\langle n \rangle$ , where  $\langle n \rangle$  is the mean number of sites scanned in each 1D round (i.e.  $\langle n \rangle = 2\sqrt{D_{1D}\tau_{1D}}$ , where  $D_{1D}$  is the diffusion coefficient of sliding). The mean total search time is then:

$$\tau_s = \frac{M}{\langle n \rangle} (\tau_{1D} + \tau_{3D}) \quad (1)$$

The expression in Equation (1) assumes that the target is recognized (i.e.  $S$  to  $R$  transition occurs) in the first search round that visits the cognate site. However, it is possible for the DPB to miss the target site and then additional 1D and 3D rounds might be needed. Accordingly, the total search time,  $\tau_s$ , must be multiplied by  $1/P_f$ , where  $P_f$  is the probability of recognizing the target site and, thus, the efficiency of the search. Given that, in a single round of 1D diffusion, the protein covers  $\sim n$  sites and makes  $n^2$  steps, each site is revisited  $\sim n$  times. Thus, the overall probability of the DPB locating the target site, once the protein associates inside a region of size  $\sim n$  that contains the site, is  $P_{loc} = \min[1, \langle n \rangle \cdot P_f]$ . The total recognition time is therefore estimated by:

$$\tau_R = \tau_s / P_{loc} \quad (2)$$

The value of  $P_f$  depends on the rate,  $k_{res}$ , to move a single step while searching using the  $S$  state (i.e. high  $k_{res}$  implicates small residence time and will reduce the probability for  $S \rightarrow R$  transition) and the transition rate,  $k_{S \rightarrow R}$ , for switching from the  $S$  to the  $R$  state.  $P_f$  thus can be expressed by:

$$P_f = \frac{k_{S \rightarrow R}}{k_{S \rightarrow R} + k_{res}} \quad (3)$$

The transition rate,  $k_{S \rightarrow R}$ , is dictated by the energy barrier  $\Delta G_{S \rightarrow R}^\ddagger$  between the two binding modes (Figure 2):

$$k_{S \rightarrow R} = \frac{1}{\tau_0} \cdot \exp\left(-\frac{\Delta G_{S \rightarrow R}^\ddagger}{k_B T}\right) \quad (4)$$

Where  $\tau_0$  represents the characteristic time to move between non-specific site. It is suggested that for very small energy barrier (i.e.  $\Delta G_{S \rightarrow R}^\ddagger \sim 0$ ), the kinetics of target recognition will be dictated by the diffusion speed. We can define  $k_{res}$  as the inverse of the average time the protein spends on a given DNA site during 1D sliding (i.e.  $k_{res} = \tau_{res}^{-1}$ , being the inverse of the residence time in the  $S$  state and is obtained by having  $n = 1$  in the relation  $\langle n \rangle = 2\sqrt{D_{1D}\tau_{1D}}$ )

(Figure 2B):

$$k_{res} = 4D_{1D}/BP^2 \quad (5)$$

The diffusion coefficient for protein sliding on a rugged DNA potential energy surface is expressed by:

$$D_{1D} = \frac{BP^2}{\tau_0} \exp\left(-\left(\frac{\sigma}{k_B T}\right)^2\right) \quad (6)$$

Where  $\tau_0$  expresses the typical time it takes the protein to hop to a neighboring site. For a protein undergoing spin coupled diffusion (51):

$$\tau_0 = \frac{\left(6\pi\eta R + \left(\frac{2\pi}{10BP}\right)^2 (8\pi\eta R^3 + 6\pi\eta R R_{OC}^2)\right) BP^2}{k_B T} \quad (7)$$

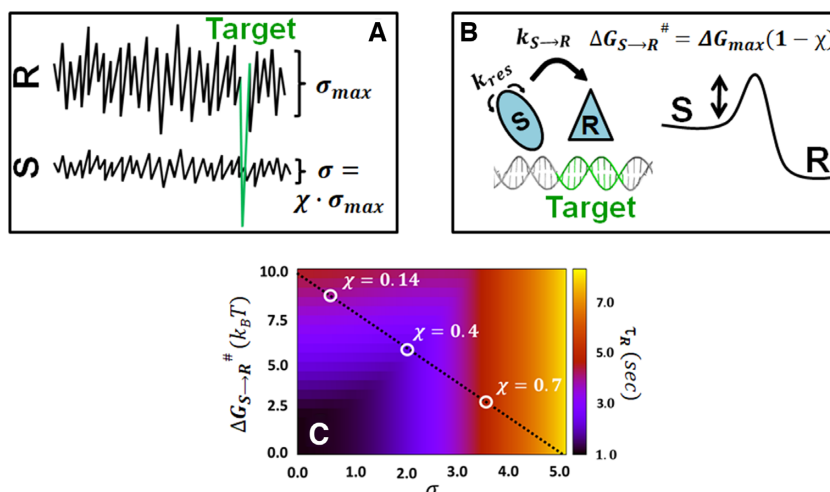
Where  $R$  is the protein's radius,  $R_{OC}$  is the distance between the center of mass of the protein and the DNA,  $BP$  is the distance between two base pairs along the DNA axis and  $\eta$  is the solution viscosity. In our model, we assumed  $R_{OC} = R$  and a protein radius of  $R = 3$  nm. The average amount of time the protein stays bound to DNA for each 1D search round is dependent on the non-specific protein–DNA binding strength, which is given by the inverse of the dissociation rate,  $k_{off}$ , of the protein from non-specific DNA site obtaining:

$$k_{off}^{-1} = \tau_{1D} = \tau_0 \exp\left(\frac{E_{ns}}{k_B T} + \frac{1}{2}\left(\frac{\sigma}{k_B T}\right)^2\right) \quad (8)$$

Where  $E_{ns}$  denotes the non-specific binding energy and is mostly governed by the electrostatic interactions between the positively charged residues and the negatively charged backbone of the DNA. The energetic contribution of state  $S$  may include a contribution from the formation of semi-specific interactions between the protein and sequences that share some similarity to the target site. The formation of such occasional hydrogen bonds that depend on the DNA sequence is captured by the roughness of the protein–DNA energy landscape,  $\sigma$  (Figure 2A). Combining the above equations with the diffusion law,  $\bar{n} = 2\sqrt{D_{1D}\tau_{1D}}$ , the  $\tau_R$  term can be expressed in terms of  $\sigma$ ,  $E_{ns}$ ,  $\Delta G_{S \rightarrow R}^\ddagger$ ,  $\tau_0$  and  $T$ . We note that in our model, the transition  $S \rightarrow R$  can take place at any DNA site although it is likely that some unknown DNA features may support the transition state of the  $S \rightarrow R$  transition.

In this study, we argue that the roughness of the protein–DNA energy landscape for sliding,  $\sigma$ , and the energetic barrier for the transition between the search and recognition modes,  $\Delta G_{S \rightarrow R}^\ddagger$ , are linked. For example, a protein having a high probability of interacting with semi-specific sites while in the  $S$  state is expected to transition faster from state  $S$  to  $R$ , simply because the  $S$  state already shares some similarity to the  $R$  state even prior to the transition. Accordingly,  $\sigma$  and  $\Delta G_{S \rightarrow R}^\ddagger$  are expected to be anti-correlated. Namely, the  $S$  and  $R$  states are linked and the relationship between them can be quantified by the degree of frustration. The molecular frustration between the  $S$  and  $R$  binding modes can be quantified through the degree of overlap between the non-specific and specific binding patches on the DBP (Figures





**Figure 2.** A schematic representation of the presented model for a DBP finding and binding its DNA target site. (A) DBP search of the DNA via 1D diffusion, during which the protein is in the *S* binding mode, is governed by the ruggedness of the protein–DNA energy landscape,  $\sigma$ . The energetic ruggedness is related to the DNA sequence and to the ability of the protein to form specific hydrogen bonds; thus, it is correlated with the similarity index,  $\chi$ . The ruggedness and frustration are linked by the relation  $\sigma = \chi \cdot \sigma_{max}$  where  $\sigma_{max}$  is the variation of the sliding energy potential in the recognition mode. The target site (marked green) is illustrated by a deep energy well in the recognition model. (B) Schematic presentation of the transition from the search (*S*) to the recognition (*R*) mode at the target site with a typical transition rate of  $k_{S \rightarrow R}$  while having to cross an energy barrier of  $\Delta G_{S \rightarrow R}^\ddagger$ . The term  $\tau_{res}$  ( $= k_{res}^{-1}$ ) is the average time the protein spends on one base pair during sliding. The free energy barrier  $\Delta G_{S \rightarrow R}^\ddagger$  depends on the similarity index through the relation  $\Delta G_{S \rightarrow R}^\ddagger = \Delta G_{max}(1 - \chi)$ , where  $\Delta G_{max}$  represents a maximal barrier for highly frustrated protein–DNA binding. (C) The total time to recognize the target site,  $\tau_R$ , is affected by the time scale of both 1D diffusion and the transition from *S* to *R*. These two process have opposite dependencies on the similarity index. Accordingly,  $\sigma$  and  $\Delta G_{S \rightarrow R}^\ddagger$  are negatively correlated. The dotted line expresses the values of  $\Delta G_{S \rightarrow R}^\ddagger$  and of  $\sigma$  when they are linked through frustration according to the proposed model, and various  $\chi$  values are marked. The total recognition times for different values of  $\chi$  and  $\Delta G_{S \rightarrow R}^\ddagger$  are indicated by the color bar. This plot was calculated for  $\Delta G_{max}$ ,  $\sigma_{max}$  and  $E_{ns}$  of 10, 5 and  $9k_B T$ , respectively.

1 and 2) (40,41). As the degree of overlap between the *S* and *R* binding modes (also known as the similarity index,  $\chi$ ) increases, additional residues may interact with DNA; thus  $\sigma$  increases. To model the relationship between  $\Delta G_{S \rightarrow R}^\ddagger$  and  $\chi$ , we use a coarse-grained model to simulate the rate of binding a target site for a set of six DBPs having varying  $\chi$  values. The kinetics of recognizing the binding site can be estimated to depend exponentially on  $\chi$  (see Supplementary Data). We therefore define linear relationships for  $\sigma$  and  $\Delta G_{S \rightarrow R}^\ddagger$  with the overlap or similarity index  $\chi$  and obtain:

$$\sigma = \chi \cdot \sigma_{max} \quad (9)$$

$$\Delta G_{S \rightarrow R}^\ddagger = \Delta G_{max}(1 - \chi) \quad (10)$$

Where  $\sigma_{max}$  and  $\Delta G_{max}$  are the maximal ruggedness and transition energy barrier, respectively. A schematic representation of the proposed model is shown in Figure 2. The maximal sliding roughness,  $\sigma_{max}$ , corresponds to the roughness in the recognition mode, *R*, in which all the protein residues participate in specific contacts with the DNA. Sufficient stability at the target site can be achieved for  $\sigma_{max} \sim 5k_B T$  and therefore this value is used throughout this study (9). An estimate of the value of  $\Delta G_{max}$  is, however, less straightforward as it can be influenced by various factors, including loss of non-specific contacts and strain, such as bending and deformation upon formation of the DNA–protein complex. We therefore evaluate the influence of various  $\Delta G_{max}$  values on the target recognition rate. We note that the search time of the target site by DBPs may be in-

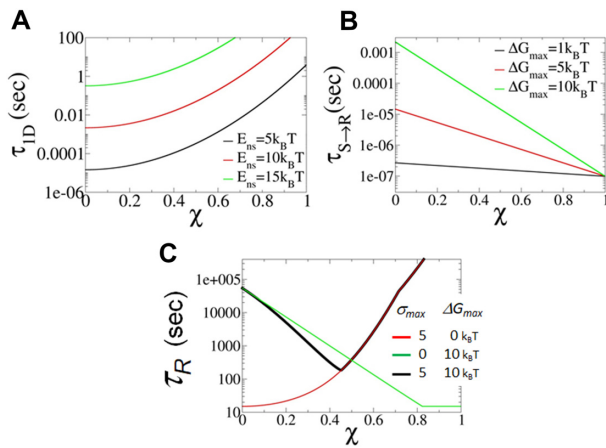
fluenced by other factors not included in the current study such as roadblocks and crowders (12,13,52), the geometric properties of the searched DNA and the locations of the target genes (25,53–55), DBP’s concentration and sequence effects of the DNA (20,39,56). Furthermore, the search kinetics might be characterized not only by the mean search time but also by the full distribution of reaction times (57,58). Despite the simplicity of the model, it quantitatively highlights the tradeoff between the search and recognition kinetics.

### Calculation of the frustration between the two-state binding modes

The similarity between the *S* and *R* states is estimated as the overlap between the binding modes. The residues that interact with DNA in the *R* mode can be obtained from crystal structures and those that interact in the *S* mode are linked to a positively charged patch on the protein surface. To estimate the overlap between *S* and *R*, we first define  $\chi_i$  for each protein residue forming specific contacts with DNA in the complex according to the equation:

$$\chi_i = \frac{\left( \sum_j q_j \cdot \exp\left(-a \frac{r_{ij}}{r_c}\right) \right)}{\left( \sum_j \exp\left(-a \frac{r_{ij}}{r_c}\right) \right)} \quad (11)$$

Where  $j$  denotes any protein residue closer than a cutoff distance of  $r_c = 8 \text{ \AA}$  to residue  $i$ ;  $q_j$  is the point charge of residue  $j$  and takes values of  $-1, 0$  or  $1$ ;  $r_{ij}$  is the cutoff distance between residues  $i$  and  $j$ ; and  $a = 5$  is an exponential



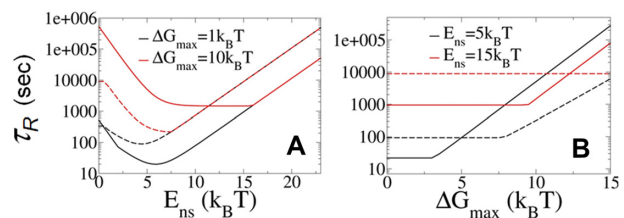
**Figure 3.** Contribution of the quantified two-state model to understanding the kinetics of DNA recognition. The existence of two states,  $S$  and  $R$ , are manifested by the similarity index,  $\chi$ . The larger the value of  $\chi$  (i.e. the lower the frustration) the greater the similarity of the  $S$  and  $R$  states. (A) The time scale of 1D diffusion,  $\tau_{1D}$ , is plotted as a function of the similarity index for three different values of non-specific DNA binding energy,  $E_{ns}$ , using a value for the free energy barrier for recognition of  $\Delta G_{max} = 5k_B T$ . (B) Transition time from  $S \rightarrow R$  as a function of  $\chi$  for various values of  $\Delta G_{max}$  ( $\tau_{S \rightarrow R}$  is estimated by  $k_{S \rightarrow R}^{-1}$ ). (C) Total recognition time as a function of  $\chi$  for different values of  $\Delta G_{max}$  and  $\sigma_{max}$  ( $E_{ns} = 5k_B T$  and  $M$  is  $5 \times 10^6$ ).

decay constant. We then average all  $\chi_i$  values of the protein to obtain its similarity index  $\chi$ . The values of  $\chi$  lie between  $(-1)$  and  $(+1)$  and depend on the parameters  $a$  and  $r_c$ . The larger the value of  $\chi$  (i.e. the lower the frustration) the greater the similarity between the  $S$  and  $R$  states. The similarity indices can be evaluated by considering all the positively charged residues in the structure or by selecting those that support the  $S$  state. The latter subset can be elucidated from coarse-grained simulations that were applied to study the sliding of proteins along DNA (4,5). We find strong correlation between the values of  $\chi$  calculated using these two approaches.

## RESULTS

### Trade-off between sliding rate and recognition rate

We first address how frustration (which impedes a protein's ability to switch from its searching mode to its recognition and binding mode) influences the different components of target recognition kinetics; namely, 1D diffusion and  $S \rightarrow R$  transition. Such frustration is associated with a positive effect on the search kinetics by reducing the energetic roughness for sliding,  $\sigma$ , and maximizing the number of BPs scanned in each 1D search round. Figure 3A plots the elapsed time for a single 1D search round as a function of the similarity index (being negatively correlated with frustration) for three different non-specific binding energies. Overall, increased frustration (i.e. decreased  $\chi$ ; Figure 3A) reduces the amount of time spent in each 1D search round by reducing the number of energetic traps for the searching protein on the DNA (i.e.  $\sigma$  is smaller for more frustrated interfaces, Equations 6 and 9). In addition, the 1D search time is highly sensitive to the magnitude of the non-specific



**Figure 4.** The effect of non-specific binding energy and transition barrier height on the total recognition time. (A) Total recognition time,  $\tau_R$ , as a function of non-specific binding energy,  $E_{ns}$ , for two values of  $\Delta G_{max}$  plotted for frustration indices  $\chi = 0.2$  (solid lines, high frustration) and  $\chi = 0.4$  (dashed lines, low frustration). (B) Total recognition time as a function of the maximal energy barrier for transition,  $\Delta G_{max}$ , for two  $E_{ns}$  values plotted using  $\chi = 0.2$  (solid lines) and  $\chi = 0.4$  (dashed lines).

binding energy, showing a four orders of magnitude difference when raising  $E_{ns}$  from 5 to 15  $k_B T$ .

Frustration also has a negative effect on the search kinetics by increasing the transition time from the  $S$  to  $R$  binding mode and thereby decreasing the probability of the protein recognizing the target DNA sequence (Equation 10). The overall transition rate for a given protein also depends on the value of  $\Delta G_{max}$  (Figure 3B), which accounts for interaction variability in different protein–DNA complexes. For a moderate similarity index value of  $\chi = 0.3$ , a two-order of magnitude increase in transition time is observed upon increasing the free energy barrier for recognition,  $\Delta G_{max}$ , from  $5k_B T$  (red line) to  $10k_B T$  (green line).

Figure 3C plots the total mean search time as a function of similarity index under three conditions: the two extreme search kinetics settings for a smooth sliding landscape (i.e.  $\sigma_{max} = 0$ ) and an immediate transition  $S \rightarrow R$  (i.e.  $\Delta G_{max} = 0$ ). The figure shows that, for cases with low transition barriers (Figure 3C, red line), the DBP favors high frustration and exhibits fast 1D diffusion. At the other extreme, where the protein lacks sliding roughness but the transition barrier is high, the protein favors low frustration (Figure 3C, green line). In the presence of both an energy barrier and sliding roughness (Figure 3C, black line), there exists an optimal frustration value at which both sliding and target recognition proceed at an adequate speed. The optimal frustration value between the  $S$  and  $R$  states may vary for different proteins depending on their electrostatic energy to bind DNA (i.e.  $E_{ns}$ ) and the free energy barrier for recognition (i.e.  $\Delta G_{max}$ ).

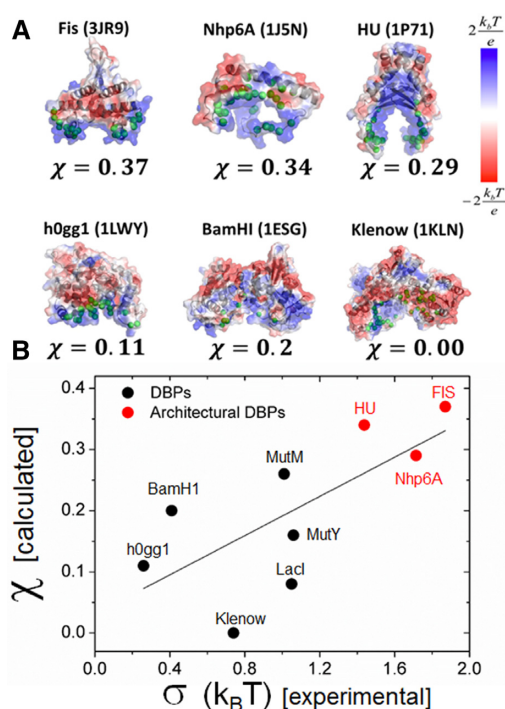
To understand better how the molecular properties of the recognition process affect its kinetics, we study the relationship between the non-specific (electrostatic) binding energy ( $E_{ns}$ ), the magnitude of the transition energy barrier ( $\Delta G_{max}$ ) and the total recognition time ( $\tau_R$ ) for two different similarity index values ( $\chi$ ) as a protein shifts between the  $S$  and  $R$  states. Figure 4A plots  $\tau_R$  as a function of  $E_{ns}$ . Changing the value of  $E_{ns}$  can be viewed as equivalent to changing the salt concentration, which is well known to tune the strength of non-specific protein–DNA binding. At high non-specific binding energies, the total recognition time is long (Figure 4A). This is due to the sluggish 1D dynamics at high values of  $E_{ns}$  (Equation 8). At low values of  $E_{ns}$ , 1D scanning in each round of 1D and 3D search is less efficient, as the time spent in the sliding mode is shorter and

consequently the number of scanned sites in each round is smaller as  $\langle n \rangle = 2\sqrt{D_{1D}\tau_{1D}}$ . As  $\langle n \rangle$  decreases, the total recognition time,  $\tau_R$ , becomes longer (Equation 1). The dependency of the total recognition time on  $E_{ns}$ , as obtained from our theory, therefore supports one of the hallmarks of the facilitated diffusion model regarding the existence of an optimal combination of 1D and 3D search modes (2,4,9,24). It can be seen that proteins with a high  $S$  to  $R$  transition barrier (Figure 4A, red curves) prefer greater frustration, that is, a lower similarity index (Figure 4A, solid versus dashed lines) in order to reduce the sliding barrier and maximize the number of target binding attempts required in each 1D search round. However, proteins with lower transition barriers (black line) do not require multiple target recognition attempts and therefore prefer lower frustration, that is, a higher similarity index (Figure 4A, dashed versus solid lines). The effect of  $E_{ns}$  on the recognition rate has been shown experimentally; increasing the salt concentration markedly affects the association rate while the dissociation rate is hardly affected (59).

Figure 4B plots the total recognition time as a function of the maximal transition energy barrier,  $\Delta G_{max}$ . The figure shows that, for a given non-specific binding energy,  $E_{ns}$ , the recognition timescale reduces as  $\Delta G_{max}$  decreases until the value of  $\Delta G_{max}$  is low and the probability for  $S \rightarrow R$  transition is very high (i.e.  $P_f = 1$ ). Figure 4B also shows that faster recognition is achieved for more frustrated proteins (lower similarity index; solid compared to dashed lines).

### Frustration is linked with the diffusion coefficient for sliding

Our group has previously calculated molecular frustration for a dataset of 125 DBPs on the basis of the crystal structure of their complexes with DNA and with the goal of estimating the degree of frustration for various protein structures and functions (40,41). Here, we quantify the frustration indices for nine DBPs (Equation 11) whose linear diffusion coefficient,  $D_{1D}$ , was measured experimentally and whose crystal structures with DNA are known. Figure 5A shows pictorially the overlap between the  $R$  state (represented by the green spheres) and the  $S$  state (represented by the blue surface and corresponding to an electrostatically positive patch). As can be seen in Figure 5A, proteins with low frustration (top row) obtain high overlap (i.e. high  $\chi$  values) and vice versa for proteins with high frustration (bottom row). The similarity index can explain the linear diffusion coefficient of different DBPs, because  $\chi$  is linked to the roughness of the protein–DNA energy landscape,  $\sigma$ . Recently, experimental data have shown architectural binding proteins to be associated with relatively high energetic roughness compared with other DBPs (60). According to our model, we expect these proteins to be characterized by low frustration (i.e. high  $\chi$  values). Figure 5B plots the frustration indices of these nine DBPs and the experimental  $\sigma$  values, which were derived from their corresponding  $D_{1D}$  values. The correlation between  $\chi$  and  $\sigma$ , shown in Figure 5B, strongly supports our model regarding the linkage between frustration and sliding (Equation 9). It is shown that, indeed, architectural binding proteins are associated with high  $\chi$  values. We therefore hypothesize that, while transcription factors are optimized for fast search and therefore



**Figure 5.** Relationship between the ruggedness of the protein–DNA energy landscape and the frustration between the  $S$  and  $R$  states. (A) Schematic illustration of the electrostatic potential surface (color bar on the right) and residues that are found to participate in specific contacts with the DNA (green beads) for several proteins whose linear diffusion coefficients were measured experimentally. The calculated similarity index,  $\chi$ , is shown below each structure. (B) Correlation between the calculated similarity index and the experimentally estimated roughness of the energy landscape ( $\sigma$ ) obtained from the linear diffusion coefficient (60). The three DNA-binding proteins (DBPs) that are marked in red were classified as architectural binding proteins. The correlation between  $\chi$  and  $\sigma$  suggests that greater energetic barriers to sliding are related to low frustration (high  $\chi$ ) and also points to the molecular difference between architectural and non-architectural DNA-binding proteins.

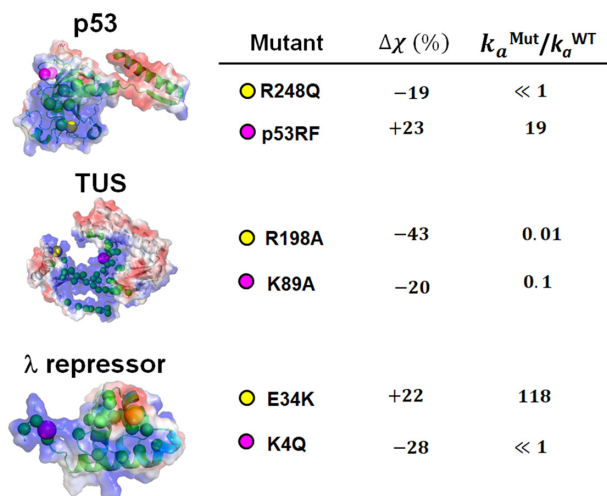
high frustration (low  $\chi$  value), architectural binding proteins are less frustrated and diffuse more slowly, as required for their function.

### Frustration can explain the effect of mutations on recognition kinetics

The effect of point mutations on the kinetics of the  $S \rightarrow R$  transition can be predicted by the frustration between these two states. Experiments measuring the association and dissociation rates of mutated proteins are routinely performed to shed light on the mechanism of protein–DNA kinetics (59,61,62). Generally, these experiments measure the binding kinetics of DBPs with short DNA oligonucleotides containing the target site and therefore they do not involve significant 1D sliding on DNA. Accordingly, we postulate that mutations that change the level of frustration will also modify the protein–DNA association rates by shifting the transition energy barrier,  $\Delta G_{S \rightarrow R}$ . The transition barrier is expected to increase for larger frustration (i.e. for lower  $\chi$ ; Equation 10).

To test our hypothesis, we calculate the change in the similarity index for six different mutants of the p53, TUS and  $\lambda$





**Figure 6.** The effect of mutations on the kinetics of DNA recognition can be explained by a change in the frustration. The kinetics of binding to a specific site on DNA was analyzed for six mutants of three different proteins (P53, TUS and  $\lambda$  repressor proteins). The electrostatic potential of each of the three proteins is shown. In each protein, the residues that participate in specific contacts with DNA (green beads) as well as the mutated sites (yellow and magenta beads) are shown. The mutants are analyzed in terms of the change in the similarity index,  $\Delta\chi$ , in comparison with the wild-type protein and with respect to the association rate relative to the wild-type protein,  $k_a^{\text{Mut}}/k_a^{\text{WT}}$ . Association rate values are taken from (59,66,68). The PDB codes used to calculate the similarity index: P53RF (4MZRa), P53wt (3Q05), TUS (1ECCR) and  $\lambda$  repressor (1LMB).

repressor proteins and for the corresponding wild-type proteins. Then, the experimentally measured association rates are compared to the change in  $\chi$ . The results are summarized in Figure 6. For the p53 protein, we studied frustration in the R248Q mutant, which has been shown to be inactivate (63,64), and for the p53FG mutant, which contains two substitutions, S121F and V122G in the Loop L1 and is associated with high activation rates (65,66). The similarity index of the wild-type p53 is  $\chi = 0.26$  whereas that of R248Q is lower (indicating increased frustration) at  $\chi = 0.21$  and that of the p53FG mutants is higher at  $\chi = 0.32$ . These results support our hypothesis by suggesting that inactivation of the R248Q protein results from an increased energetic barrier for the  $S \rightarrow R$  transition, which is manifested in a 19% decrease in the similarity index. The increased activation rates of the p53FG mutant can be attributed to a decreased barrier, which is characterized by a 23% increase in the similarity index.

Similar results are obtained for the TUS and  $\lambda$  repressor proteins. In the case of the TUS protein, both the K89A and the R198A mutants are characterized by decreased association rates (59) and accordingly show frustration indices decreased by 20% and 43%, respectively. The  $\lambda$  repressor K4Q mutant is characterized by a significant reduction in activity (67,68) and a decrease in the similarity index of 28%, accordingly. In contrast, the E34K mutant, which is a secondary mutant that also contains the K4Q mutation, is characterized by a substantial increase in the similarity index from 0.13 for the K4Q mutant to 0.22 for the E34K mutant, which corresponds to a 22% increase with respect to the wild-type and is accordingly associated with an in-

creased association rate. In all of the above cases, apart from the p53FG mutant, the mutations are associated with substitutions of charged residues that alter frustration by influencing the degree of overlap between specific protein–DNA contacts and the positively charged residues. In the case of the p53FG mutant, the variation in frustration is a result of modification to specific protein–DNA contacts.

## CONCLUSIONS

In this study, the facilitated diffusion model of DNA search by proteins for their target sites is revisited in order to consider explicitly not only the timescales for finding the target sites, but also the timescales for recognizing that site after it was identified. We show that these two processes of finding and binding the target sites are coupled via protein frustration between the non-specific and specific binding modes (i.e. the search,  $S$ , and recognition,  $R$ , states). Low frustration means that the non-specific and specific binding modes of the protein to the DNA are very similar, so recognition can occur relatively easily after the target is identified. Nonetheless, low frustration suggests hindered 1D diffusion because semi-specific protein–DNA interactions can be formed with higher probabilities. Accordingly, low frustration defines a low barrier for recognition but a more rugged protein–DNA energy landscape. High frustration, on the other hand, implies that a conformational change will be required during the transition from non-specific association with the specific complex. Indeed, in many protein–DNA complexes, either or both the protein and the DNA change conformation to affect specific binding (32–34). Our theory predicts that the magnitude of the conformational change depends on the similarity index. Consistently with this perspective, an NMR study of homeodomain HoxD9 that reported high similarity between the non-specific and specific binding modes (47,69) can be explained by the high similarity measure (i.e. low frustration) of this system ( $\chi = 0.37$ ). The experimentally estimated  $D_{1D}$  of HoxD9 (70) corresponds to an energetic ruggedness of  $\sim 2k_B T$ , again in agreement with low frustration. Lac repressor, on the other hand, is characterized by a lower similarity index ( $\chi = 0.08$ ) that explains the lower energetic ruggedness of  $< 1 k_B T$  for its 1D diffusion as was deduced experimentally (10). This high frustration may result not only in faster 1D diffusion but also in high transition barrier that may cause to unsuccessful recognition after finding the target site, as was indeed concluded experimentally for the lac repressor (8).

The concept of frustration between the  $S$  and  $R$  binding modes is utilized to successfully predict several experimentally measured phenomena. We show that the degree of frustration explains differences in the linear diffusion coefficients for sliding along DNA and thus provides a molecular interpretation for the experimentally reported barrier to sliding (10,60). In particular, we argue that architectural proteins slide along DNA with a smaller diffusion coefficient due to lower frustration. Furthermore, the concept of frustration has practical implications in predicting variations in recognition kinetics, particularly the effect of point mutations on recognition rates. Although mutations that change specific hydrogen bonds with the target DNA sites

often affect the dissociation rates (i.e.  $k_{\text{off}}$ ), some mutations that involve positively charged residues are found to affect the association rate (i.e.  $k_{\text{on}}$ ). The latter can be rationalized as a change in the molecular frustration between the *S* and *R* states upon mutation. This also implies that the protein–DNA binding affinity can be modulated via  $k_{\text{on}}$  and not only via  $k_{\text{off}}$ , as is often found. Our study explains the origin of the experimentally observed changes in association rate with salt concentration while the dissociation rates are hardly changed (59). Changing the energy of the specific complex (for example, by mutating the DNA target site) is expected to have a much greater effect on the dissociation rates than on the association effects, as was reported earlier for several proteins using comprehensive kinetic measurements (61).

Our study illustrates that there is a trade-off between the speed of sliding diffusion during the search process and the kinetics of the transition from the search mode to the recognition mode. Accordingly, our model suggests that the speed of sliding dynamics and the speed of recognition are tightly coupled. Most importantly, this trade-off is modulated by frustration between the non-specific and specific protein–DNA interactions and therefore there is an optimal degree of frustration that minimizes the total time for recognition (i.e. search plus recognition). The concept of frustration suggests that it is not only the search process that dictates the overall recognition time but also the energetic barrier for the specific complex. The magnitude of the trade-off depends on various parameters of the protein–DNA system such as the nature of the conformational change and the protein structures (e.g. existence of multi domains (14)). Finally, we show that our model is supported by experimental results, which find a strong correlation between the similarity index and the kinetic properties of protein–DNA recognition. Our model advances understanding and quantifies the relationship between protein characteristics and the facilitated diffusion mechanism and thus can be utilized to assist in the design of mutations to engineer and control the kinetic search process of DBPs.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

This work was supported by the Kimmelman Center for Macromolecular Assemblies and the Minerva Foundation. Y.L. holds The Morton and Gladys Pickman professional chair in Structural Biology.

## FUNDING

Federal German Ministry for Education and Research. Israel Science Foundation grant number 1583/17. Funding for open access charge: Internal funding.

Conflict of interest statement. None declared.

## REFERENCES

- Riggs, A.D., Bourgeois, S. and Cohn, M. (1970) The lac repressor-operator interaction. 3. Kinetic studies. *J. Mol. Biol.*, **53**, 401–417.

- Berg, O.G., Winter, R.B. and von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
- von Hippel, P.H. and Berg, O.G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.*, **264**, 675–678.
- Givaty, O. and Levy, Y. (2009) Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.*, **385**, 1087–1097.
- Bhattacharjee, A., Krepel, D. and Levy, Y. (2016) Coarse-grained models for studying protein diffusion along DNA. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **6**, 515–531.
- Marcovitz, A. and Levy, Y. (2012) Sliding dynamics along DNA: a molecular perspective. In: *RSC Biomolecular Sciences No. 24, Innovations in Biomolecular Modeling and Simulations*. Vol. **2**, pp. 236–262.
- Bauer, M. and Metzler, R. (2012) Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.*, **102**, 2321–2330.
- Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E.G., Berg, O.G. and Elf, J. (2012) The lac repressor displays facilitated diffusion in living cells. *Science*, **336**, 1595–1598.
- Slutsky, M. and Mirny, L.A. (2004) Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.*, **87**, 4021–4035.
- Blainey, P.C., Luo, G., Kou, S.C., Mangel, W.F., Verdine, G.L., Bagchi, B. and Xie, X.S. (2009) Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.*, **16**, 1224–1229.
- Zhou, H.X. (2011) Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 8651–8656.
- Krepel, D. and Levy, Y. (2016) Protein diffusion along DNA: on the effect of roadblocks and crowders. *J. Phys. A-Math Theor.*, **49**, 494003.
- Krepel, D., Gomez, D., Klumpp, S. and Levy, Y. (2016) Mechanism of facilitated diffusion during a DNA search in crowded environments. *J. Phys. Chem. B*, **120**, 11113–11122.
- Zandarashvili, L., Esadze, A., Vuzman, D., Kemme, C.A., Levy, Y. and Iwahara, J. (2015) Balancing between affinity and speed in target DNA search by zinc-finger proteins via modulation of dynamic conformational ensemble. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E5142–E5149.
- Vuzman, D. and Levy, Y. (2014) The “Monkey-Bar” mechanism for searching for the DNA target site: the molecular determinants. *Israel J. Chem.*, **54**, 1374–1381.
- Vuzman, D. and Levy, Y. (2012) Intrinsically disordered regions as affinity tuners in protein–DNA interactions. *Mol. BioSyst.*, **8**, 45–57.
- Vuzman, D., Polonsky, M. and Levy, Y. (2010) Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. *Biophys. J.*, **99**, 1202–1211.
- Khazanov, N., Marcovitz, A. and Levy, Y. (2013) Asymmetric DNA-search dynamics by symmetric dimeric proteins. *Biochemistry*, **52**, 5335–5344.
- Khazanov, N. and Levy, Y. (2011) Sliding of p53 along DNA can be modulated by its oligomeric state and by cross-talks between its constituent domains. *J. Mol. Biol.*, **408**, 335–355.
- Bhattacharjee, A. and Levy, Y. (2014) Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. *Nucleic Acids Res.*, **42**, 12404–12414.
- Bhattacharjee, A. and Levy, Y. (2014) Search by proteins for their DNA target site: 2. The effect of DNA conformation on the dynamics of multidomain proteins. *Nucleic Acids Res.*, **42**, 12415–12424.
- Tafvizi, A., Mirny, L.A. and van Oijen, A.M. (2011) Dancing on DNA: kinetic aspects of search processes on DNA. *Chemphyschem.*, **12**, 1481–1489.
- Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J. and Kosmrlj, A. (2009) How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A Math. Theor.*, **42**, 434013.
- Halford, S.E. and Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
- Wunderlich, Z. and Mirny, L.A. (2008) Spatial effects on the speed and reliability of protein–DNA search. *Nucleic Acids Res.*, **36**, 3570–3578.
- Halford, S.E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.*, **37**, 343–348.



27. Kolomeisky, A. (2011) Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.*, **13**, 2088–2095.
28. Shvets, A.A., Kochugaeva, M.P. and Kolomeisky, A.B. (2018) Mechanisms of protein search for targets on DNA: theoretical insights. *Molecules*, **23**, 2106.
29. Sokolov, I.M., Metzler, R., Pant, K. and Williams, M.C. (2005) Target search of N sliding proteins on a DNA. *Biophys. J.*, **89**, 895–902.
30. Brackley, C.A., Cates, M.E. and Marenduzzo, D. (2013) Intracellular facilitated diffusion: searchers, crowders, and blockers. *Phys. Rev. Lett.*, **111**, 108101.
31. Hu, T., Grosberg, A.Y. and Shklovskii, B.I. (2006) How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys. J.*, **90**, 2731–2744.
32. Viadiu, H. and Aggarwal, A.K. (2000) Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell*, **5**, 889–895.
33. Kalodimos, C.G., Biris, N., Bonvin, A.M., Levandoski, M.M., Guennegues, M., Boelens, R. and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science*, **305**, 386–389.
34. von Hippel, P.H. (2007) From 'simple' DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 79–105.
35. Kochugaeva, M.P., Shvets, A.A. and Kolomeisky, A.B. (2016) How conformational dynamics influences the protein search for targets on DNA. *J. Phys. A Math. Theor.*, **49**, 444004.
36. Niranjani, G. and Murugan, R. (2016) Generalized theory on the mechanism of site-specific DNA-protein interactions. *J. Stat Mech-Theory E*, **2016**, 053501.
37. Murugan, R. (2007) Generalized theory of site-specific DNA-protein interactions. *Phys. Rev. E*, **76**, 011901.
38. Hu, L.H., Grosberg, A.Y. and Bruinsma, R. (2008) Are DNA transcription factor proteins Maxwellian Demons? *Biophys. J.*, **95**, 1151–1156.
39. Bauer, M., Rasmussen, E.S., Lomholt, M.A. and Metzler, R. (2015) Real sequence effects on the search dynamics of transcription factors on DNA (vol 5, 10072, 2015). *Sci. Rep.*, **5**, 10072.
40. Marcovitz, A. and Levy, Y. (2011) Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 17957–17962.
41. Marcovitz, A. and Levy, Y. (2013) Weak frustration regulates sliding and binding kinetics on rugged protein-DNA landscapes. *J. Phys. Chem. B*, **117**, 13005–13014.
42. Tafvizi, A., Huang, F., Fersht, A.R., Mirny, L.A. and van Oijen, A.M. (2011) A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 563–568.
43. Zandarashvili, L., Vuzman, D., Esadze, A., Takayama, Y., Sahu, D., Levy, Y. and Iwahara, J. (2012) Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E1724–E1732.
44. Cuculis, L., Abil, Z., Zhao, H.M. and Schroeder, C.M. (2015) Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.*, **6**, 7277.
45. Cuculis, L., Abil, Z., Zhao, H.M. and Schroeder, C.M. (2016) TALE proteins search DNA using a rotationally decoupled mechanism. *Nat. Chem. Biol.*, **12**, 831–837.
46. von Hippel, P.H. (2004) Biochemistry. Completing the view of transcriptional regulation. *Science*, **305**, 350–352.
47. Iwahara, J. and Clore, G.M. (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*, **440**, 1227–1230.
48. Iwahara, J., Zweckstetter, M. and Clore, G.M. (2006) NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15062–15067.
49. Ferreira, D.U., Komives, E.A. and Wolynes, P.G. (2014) Frustration in biomolecules. *Quarter. Rev. Biophys.*, **47**, 285–363.
50. Slutsky, M. (2005) Protein-DNA interaction, random walks and polymer statistics MIT, June 2005. <https://dspace.mit.edu/handle/1721.1/32295?show=full>.
51. Bagchi, B., Blainey, P.C. and Xie, X.S. (2008) Diffusion constant of a nonspecifically bound protein undergoing curvilinear motion along DNA. *J. Phys. Chem. B*, **112**, 6282–6284.
52. Marcovitz, A. and Levy, Y. (2013) Obstacles may facilitate and direct DNA search by proteins. *Biophys. J.*, **104**, 2042–2050.
53. Benichou, O., Chevalier, C., Klafter, J., Meyer, B. and Voituriez, R. (2010) Geometry-controlled kinetics. *Nat. Chem.*, **2**, 472–477.
54. Pulkkinen, O. and Metzler, R. (2013) Distance matters: the impact of gene proximity in bacterial gene regulation. *Phys Rev Lett*, **110**, 198101.
55. Di Stefano, M., Rosa, A., Belcastro, V., di Bernardo, D. and Micheletti, C. (2013) Colocalization of Coregulated Genes: A Steered Molecular Dynamics Study of Human Chromosome 19. *PLoS Comput. Biol.*, **9**, e1003019.
56. Cencini, M. and Pigolotti, S. (2018) Energetic funnel facilitates facilitated diffusion. *Nucleic Acids Res.*, **46**, 558–567.
57. Grebenkov, D., Metzler, R. and Oshanin, G. (2018) Strong defocusing of molecular reaction times results from an interplay of geometry and reaction control. *Commun. Chem.*, **1**, 96.
58. Godec, A. and Metzler, R. (2016) Universal proximity effect in target search kinetics in the few-encounter limit. *Phys. Rev. X*, **6**, 041037.
59. Neylon, C., Brown, S.E., Kralicek, A.V., Miles, C.S., Love, C.A. and Dixon, N.E. (2000) Interaction of the Escherichia coli replication terminator protein (Tus) with DNA: A model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry*, **39**, 11989–11999.
60. Kamagata, K., Mano, E., Ouchi, K., Kanbayashi, S. and Johnson, R.C. (2018) High free-energy barrier of 1D diffusion along DNA by architectural DNA-binding proteins. *J. Mol. Biol.*, **430**, 655–667.
61. Geertz, M., Shore, D. and Maerkl, S.J. (2012) Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16540–16545.
62. Maerkl, S.J. and Quake, S.R. (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, **315**, 233–237.
63. Itoh, Y., Murata, A., Sakamoto, S., Nanatani, K., Wada, T., Takahashi, S. and Kamagata, K. (2016) Activation of p53 facilitates the target search in DNA by enhancing the target recognition probability. *J. Mol. Biol.*, **428**, 2916–2930.
64. Ory, K., Legros, Y., Auguin, C. and Soussi, T. (1994) Analysis of the most representative tumor-derived p53 mutants reveals that changes in protein conformation are not correlated with loss of transactivation or inhibition of cell-proliferation. *EMBO J.*, **13**, 3496–3504.
65. Emamzadah, S., Tropia, L., Vincenti, I., Falquet, B. and Halazonetis, T.D. (2014) Reversal of the DNA-Binding-Induced Loop L1 Conformational Switch in an Engineered Human p53 Protein. *J. Mol. Biol.*, **426**, 936–944.
66. Petty, T.J., Emamzadah, S., Costantino, L., Petkova, I., Stavridi, E.S., Saven, J.G., Vauthey, E. and Halazonetis, T.D. (2011) An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity. *EMBO J.*, **30**, 2167–2176.
67. Hecht, M.H., Nelson, H.C.M. and Sauer, R.T. (1983) Mutations in lambda-repressors amino-terminal domain - implications for protein stability and DNA-binding. *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 2676–2680.
68. Nelson, H.C.M. and Sauer, R.T. (1985) Lambda repressor mutations that increase the affinity and specificity of operator binding. *Cell*, **42**, 549–558.
69. Iwahara, J. and Clore, G.M. (2006) Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *J. Am. Chem. Soc.*, **128**, 404–405.
70. Sahu, D. and Iwahara, J. (2017) Discrete-state kinetics model for NMR-based analysis of protein translocation on DNA at equilibrium. *J. Phys. Chem. B*, **121**, 9548–9556.