Obstacles May Facilitate and Direct DNA Search by Proteins

Amir Marcovitz and Yaakov Levy*
Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

ABSTRACT   DNA recognition by DNA-binding proteins (DBPs), which is a pivotal event in most gene regulatory processes, is often preceded by an extensive search for the correct site. A facilitated diffusion process in which a DBP combines three-dimensional diffusion in solution with one-dimensional sliding along DNA has been suggested to explain how proteins can locate their target sites on DNA much faster than predicted by three-dimensional diffusion alone. Although experimental and theoretical studies have recently advanced understanding of the biophysical principles underlying the search mechanism, the process under in vivo cellular conditions is poorly understood. In this study, we used various computational approaches to explore how the presence of obstacle proteins on the DNA influences search efficiency. At a low obstacle occupancy (i.e., when few obstacles occupy sites on the DNA), sliding by the searching DBP may be confined, which may impair search efficiency. The obstacles, however, can be bypassed during hopping events, and the number of bypasses is larger for higher obstacle occupancies. Dynamism on the part of the obstacles may even further facilitate search kinetics. Our study shows that the nature and efficiency of the search process may be governed not only by the intrinsic properties of the DBP and the salt concentration of the medium, but also by the in vivo association of DNA with other macromolecular obstacles, their location, and occupancy.

INTRODUCTION

DNA binding proteins (DBPs) mediate cellular gene expression and regulation by rapidly and specifically binding to DNA sites in the genome (1). DNA target localization by proteins may exceed the diffusion limit by two orders of magnitude, as was observed in vitro for the lac repressor (2). To explain this phenomenon, the theory of facilitated diffusion (3,4) suggested that DBPs use a reduced dimensionality search process involving three-dimensional (3D) diffusion in solution and one-dimensional (1D) sliding along DNA. Such 1D movement of proteins along DNA is enabled through electrostatic attraction between the positively charged patches common in DBPs (5) and the negatively charged backbone of DNA. This attraction typically results in a transient nonspecific protein-DNA association, with affinities that strongly depend on the salt concentration and that can be several orders of magnitude lower than the affinities to the specific DNA target sites (6).

Early evidence from ensemble studies that focused on endonuclease activity provided important indirect evidence for the ability of proteins to move one-dimensionally along DNA (7,8). More recently, NMR (9), single-molecule studies (10), and in vivo evidence (11,12) have demonstrated that a combination of 1D and 3D motions is indeed common in many DBPs and have further elucidated the mechanism and kinetics of 1D motion along DNA. In particular, it is now accepted that proteins may track the DNA backbone when they helically slide along the major groove (13,14) or may perform microscopic hopping from one site to another.

Despite considerable theoretical (15–21) and experimental efforts, researchers have yet to obtain a detailed description and understanding of the possible effects of macromolecular crowding and of the presence of obstacles on the properties and quality of 1D movement along DNA, and the overall efficiency of DNA search by proteins is still poorly understood (22–24). A particularly puzzling issue is how proteins that slide along DNA stretch in a rotation-coupled manner (13) to overcome obstacles, such as statically bound nucleosomes in eukaryotes (25–27), nucleoid proteins in bacteria (28), or other nonspecifically bound DBPs that may diffuse independently along the DNA (22,29). DNA coverage by DBPs is high—indeed, they are estimated to cover ~30% of genomic prokaryotic DNA (28,30) and a much higher percentage of eukaryotic DNA (31). In the yeast genome, the presence of obstacles results in the average nucleosomal spacer region length between obstacles ~10–40 bp (25). The presence of obstacles may interfere with the 1D motion of DBPs along DNA. Although several theoretical studies suggested that increasing the number of obstacles occupying sites on the DNA increases the search time for a specific target (22,32), several experiments demonstrated that obstacles could be bypassed under certain conditions. Specifically, proteins that employ more hopping in their 1D DNA search can bypass obstacles as large as nucleosomes (33) or other DBPs (34,35). By contrast, proteins that slide can be blocked by nucleosomes or even by DNA deformations (33,35).

Here, our purpose is to understand the consequences that DNA coverage by obstacles has for the mechanism and kinetics of 1D DBP searches of DNA. Using coarse-grained molecular dynamic (MD) simulations for nonspecific protein-DNA interactions, which have been demonstrated to capture many of the search features observed

experimentally (36–38), we provide several new, to our knowledge, insights regarding the effects of the obstacles on the search.

MATERIALS AND METHODS

Coarse-grained simulation model

We used a coarse-grained model (5,14,36) in which the protein is represented by a single bead per residue centered at the C_α position and the DNA is modeled by three beads per nucleotide (representing phosphate, sugar, and base) that are positioned at the geometric center of the represented group. We performed MD simulations of a human DNA-binding protein Sap-1 (PDB code: 1bc8), member of the Ets transcription factors in the presence of 100 bp double-stranded DNA. Sap-1, a 93 amino-acids globular protein with a total of 15 and 6 positively and negatively charged residues, respectively, uses a winged-helix DNA-binding domain to activate transcription. The protein was simulated by a native topology-based model that excludes nonnative interactions and uses Lennard-Jones potential to represent native contact interactions. Electrostatic interactions acting between all the charged beads in the system were modeled by the Debye-Hückel potential (see below). For all simulations, we used a 100 bp double-stranded DNA that remains rigid and static in its canonical B-form and that was centered on and aligned with the Z-axis. Obstacles were equally spaced along the DNA. Each obstacle occupied 2 bp (see Fig. S1 in the Supporting Material) and was modeled by either changing the charge of the phosphate beads from -1 to $+1$ or by neutralizing the charge of the phosphates and increasing the exclusion volume repulsion radius (C_{ex}) by a factor of 2 or 4 relative to the radius of DNA beads in nonoccupied sites (C_{ex}^0).

The dynamics of the protein along the DNA was simulated using the Langevin dynamics (36,39). The simulations were performed at constant temperature below the unfolding temperature of the protein Sap-1 (i.e., $T_{simulation} \approx T_f$, where T_f is the equilibrium proteins folding/unfolding temperature), and was analyzed in terms of sliding and hopping (together termed 1D diffusion) and 3D diffusion. All runs were simulated for 200×10^6 time steps that allow extensive DNA sampling by the protein and transitions between sliding, hopping, and 3D search modes. To differentiate in silico between protein sliding, hopping, and free 3D diffusion, we used the definitions defined by Givaty and Levy (36). 1D diffusion along the DNA was used to calculate the mean square displacement profiles along the Z-axis. We accounted for obstacle bypasses in the simulations as hopping events that result in the protein jumping from one side of an obstacle to another and calculated the kinetics for the obstacle bypassing as $k_{bypass} = 1/\tau_{bypass}$, where τ_{bypass} is the median protein residence time between two adjacent obstacles in a 1D search mode. We point out that due to coarse-graining of the model, one cannot easily convert the time steps to realistic time. Nevertheless, our model reproduces the ratio $D_3/D_1 \approx 100$ as characterized experimentally as well as other kinetic and mechanistic features such as the dependence of the linear diffusion on salt concentration.

Modeling nonspecific protein-DNA interaction using Debye-Hückel potential

Because of the elusive nature of nonspecific interactions between protein and DNA, which are central to the DNA search process by proteins, we used a simplified representation for nonspecific DNA binding in which the interactions between proteins and DNA are governed solely by electrostatic forces. Beads representing the charged amino acids (Arg, Lys, Asp, and Glu) and the DNA phosphate groups were charged in the model. We employed the Debye-Hückel potential to describe the electrostatic interactions between the protein and DNA. Although the Debye-Hückel model is powerful in introducing the salt effect of screening electrostatic interactions into the Coulomb potential, one should be aware of its approximations. The

model is valid for relatively dilute conditions, as it approximates that the potential energy of an ion is determined by pairwise interactions with other neighboring ions. The detailed effects of higher salt concentrations and of ion condensation on DNA have to be studied using the nonlinearized Poisson-Boltzmann equation, as well with atomistic simulations that can elucidate the dynamics of the ionic layer during sliding on the DNA.

Because the model is coarse-grained, the distances between the charged beads of the protein and the charged DNA beads are longer compared to fully atomistic models (the charges are placed at the phosphate and C_α beads instead of their actual atoms). Consequently, the salt concentrations reported in this work using the coarse-grained model (10–50 mM) are smaller than typical physiological salt conditions (~100–150 mM). The range for the salt condition selected in this work is however a measure of strength within the model that successfully recapitulates many important characteristics of protein search modes on DNA (36,40–43). We used a dielectric constant of 70–80 (the typical range in water), because the protein-DNA interface is much more hydrated in the nonspecific complex than in the specific complex (17).

Estimation of the search efficiency of the DNA

To address the efficiency of the DNA search, we employed the positions probed (PP) measure (36), which indicates the number of new DNA sites visited by a protein during a simulation when the helical recognition site is located at the DNA major groove. This measure is calculated by dividing the DNA into sections having a length of 3.3 Å (the length of a single DNA bp). Any frame in which the protein is situated at the DNA major groove is added to the PP measure, unless this position has already been visited by the sliding mode. When the protein dissociates from the DNA, all marked positions are wiped, and the number of probed positions is left unchanged. Thus, the next time the DNA is probed, all positions will be unmarked. Wiping the marked positions after each dissociation event is performed to mimic the fact that, upon dissociation, the probability of the protein returning to the same DNA section is negligible; thus, upon reassociation, the protein would be probing an unprobed region of the DNA. Because our model does not include the specific site, the PP measure serves as an estimate for the ease with which the protein locates the specific site.

Monte Carlo simulations for a random walk on a 1D lattice with Gaussian obstacles

We simulated the movement of a random walker on a 1D lattice with uniform potential. For each set of simulations, we imposed on the 1D lattice equally distributed Gaussian potentials at spacing ranging from 5 to 50 lattice sites (Gaussian height equals 10, and standard deviation is $\sigma = 7$, 10, or 15). The protein moves between adjacent sites with the moves being accepted or rejected according to the Metropolis criteria.

Lattice simulations of a random walker with mobile obstacles

We account for a scenario in which some obstacles leave the DNA, whereas others randomly associate with it at the same rate such that constant equilibrium obstacle occupancy is maintained. The obstacles may cover the target site and hide it for a searcher protein. The model we use is a circular 1D lattice representing a DNA with $N = 200$ sites. The protein takes random single move walks in each step where $k_{Prot,1D}$ is the rate constant for protein 1D movements. The dynamics of the obstacles is incorporated through the factor $\mu = k_{Obst}/k_{Prot,1D}$, where k_{Obst} is the rate at which each obstacle is evacuated from its current site and randomly binds at another nonoccupied site). Salt effects are introduced through an additional pair of rate constants for the protein: k_{1D-3D} for on- to off-lattice transitions and k_{3D-1D} for off- to on-lattice sites. For simplicity, we represent the salt

condition by an exponential measure $\rho = k_{\text{1D-3D}}/k_{\text{3D-1D}}$. A low value of ρ corresponds to a low salt concentration at which the protein has a higher probability for 1D than 3D diffusion. We studied five values of ρ (see Table S1).

RESULTS AND DISCUSSION

DNA search efficiency depends on obstacle occupancy and salt concentration

A typical DBP often alternates between 3D diffusion in solution and 1D excursion along DNA that may involve sliding or microscopic hopping events, depending on salt concentration (36). In addition, the protein may transfer between two distant DNA segments through a looped DNA region (44) (Fig. 1 A). To explore the effects of obstacles on the 1D motion of a DBP along DNA, we used coarse-grained (MD) simulations of a globular transcription factor (Sap-1, PDB code: 1bc8, see Materials and Methods) interacting with a DNA molecule that is 100 bp long. In our model, only nonspecific electrostatic interactions, which are incorporated through the Debye-Hückel potential, exist between the protein and the DNA, and protein motion is driven by thermal diffusion. Obstacles were represented by perturbations: electrostatic repulsions or strong excluded volume effects (see Materials and Methods) introduced onto the canonical B-DNA to create a nonuniform backbone incompatible with continuous 1D sliding by a DBP along DNA (Fig. 1 A). In each simulation, various numbers (0–15) of obstacles (each occupying 2 bp) were positioned at equal intervals along the DNA. The efficiency with which the protein sampled the DNA was monitored using the PP measure (see Materials and Methods), which records the number of DNA sites that are newly visited by a DBP using the sliding search mode (36).

At a low salt concentration (0.01 M) where electrostatic screening is relatively weak, the protein is mostly bound to the DNA (i.e., performs sliding) and is likely to encounter the obstacle quite often. Fig. 1 B shows a decrease in the PP

at a moderate occupancy of 3–5 obstacles per 100 bp (corresponding to $0.1 < S < 0.2$, where S is the obstacle occupancy—a unitless ratio of the protein diameter R to the distance between two adjacent obstacles l; see inset). As the obstacle occupancy on DNA increases ($S > \sim 0.35$), the search efficiency recovers. Naturally, on a fully loaded DNA, the DNA backbone is not accessible for sliding and consequently the previous rapid rise in the number of newly visited DNA sites tapers off. At a higher salt concentration of 0.03 M, a similar trend is observed, though the overall search efficiency is higher and, when the DNA is fully loaded with obstacles, the search efficiency is strongly diminished. When further increasing the salt concentration to 0.05 M, moderate DNA occupancies (3–5 obstacles, $S \sim 0.2$) have less effect on DNA sampling, presumably because the sliding events are much more transient and obstacles are encountered less frequently. Marked changes in PP values are evident only at high obstacle occupancies ($S > \sim 0.4$) at which the DNA has limited accessibility to the protein. Similar effects on search efficiency are obtained for charged and bulky obstacles.

To highlight interference by the obstacle in the kinetics and mechanism of the search performed by the DBP, we calculated the mean-squared displacement (MSD) profiles for 1D dynamics of the protein along the DNA occupied by 0 ($S = 0.04$), 4 ($S = 0.2$), and 7 ($S = 0.34$) obstacles, under salt concentrations of 0.01, 0.03, and 0.05 M. Fig. S2 shows MSD profiles for 10 representative trajectories. For a moderate occupancy of $S = 0.2$, the overall slope of the profiles is lower than that obtained from naked DNA simulations, therefore implying a decrease in the 1D diffusion coefficient of the protein. As the occupancy increases ($S = 0.34$), so do the slopes of the MSD profiles, implying that the 1D diffusion coefficient has recovered, in agreement with the increase in DNA sampling efficiencies observed in Fig. 1 B for 0.01 and 0.03 M at high values of S. Similar to the effect of salt concentration on the usage of sliding, hopping, and 3D diffusion (36), we find that increasing obstacle
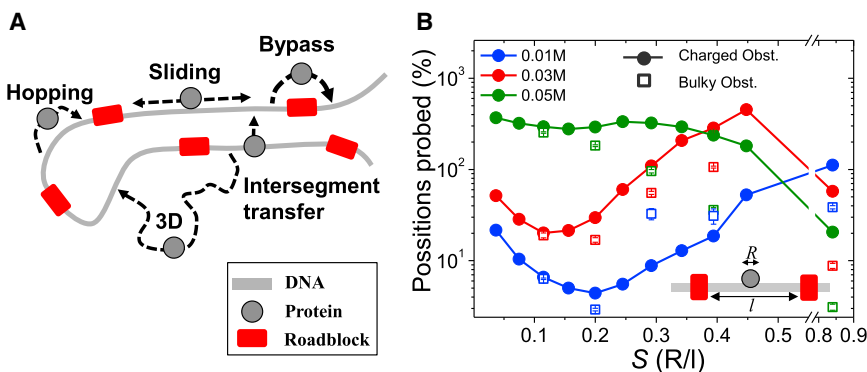

FIGURE 1 Protein search mechanism on obstacle-covered DNA. (A) A schematic view of protein search mechanisms on DNA in the presence of obstacles (red rectangles). The protein may encounter an obstacle during sliding or may perform microscopic hops, which may assist bypassing the obstacles. DNA loops may also assist obstacle bypassing as the protein may occasionally use intersegmental transfers to traverse between two DNA regions that are far apart in sequence but are spatially close. (B) PP as a function of obstacle occupancy S (calculated as R/l, where R is the diameter of the searching protein and l is the length of spacer DNA between two adjacent obstacles, see inset). Obstacles repel the searching protein via electrostatic repulsions (DNA phosphate bead charge is modified from -1 to $+1$ at the location of the obstacles—shown by solid circles) or via excluded volume interactions (bulky obstacles with an excluded volume four times that of the protein, $C_{\text{Ex}}/C_{\text{Ex}}^0 = 4$—shown by empty squares). Reported results reflect averages over 40–50 simulations under each salt condition and obstacle settings. Error bars reflect standard errors.

occupancy may interfere with the population of these search mechanisms. At a low salt concentration of 0.01 M, the protein performs mostly sliding. However, under higher salt conditions, increasing the load of obstacles on the DNA results in more hopping and 3D diffusion events at the expense of sliding (Fig. S3). We next aim to understand the source for the modified sampling efficiency under varying obstacles occupancies, and to elucidate how the preference for a particular combination of sliding, hopping, and 3D diffusion may enable the protein to efficiently probe loaded DNA.

Moderate obstacle occupancies confine 1D search to the linker region

To test whether obstacles block the linear diffusion path for the searcher protein as it moves along DNA, we focused on 1D excursion by the DBP. For each 1D event, we measured d_z, which is the distance between the minimal and maximal positions of the DBP on the Z-axis (with which the DNA is aligned) and MSD_z, which is the sum of all projections on the Z-axis during a single 1D excursion (14) (see Fig. 2 A). A higher value for MSD_z/d_z, reflects greater redundancy in 1D search, with the DBP revisiting previously searched DNA sites. Fig. 2 A shows how the average redundancy changes as a function of DNA obstacle occupancy, with a clear hierarchy between different salt concentrations. At 0.01 M, redundancy in sampling the DNA sites is greatest for $S \sim 0.2$ (3–5 obstacles per 100 bp). Repeatedly revisiting DNA sites reduces the overall search efficiency, and indeed search redundancy is significantly negatively correlated with PP at a salt concentration of 0.01 M (Fig. 1 B) (Pearson $= -0.74$, $p < 0.01$). A high degree of 1D sampling redundancy for $S \sim 0.1–0.2$ is also observed under 0.03 M salt conditions, but to a lesser extent under high salt conditions of 0.05 M. Fig. S4 illustrates that, at low to moderate obstacle occupancies ($S < \sim 0.2$), the mean value of d_z decreases and is confined by obstacle spacing length. For a salt concentration of <0.05 M, the values are highly correlated with the length of the linker region between two adjacent obstacles. These observations imply that moderate occupancies confine the protein to 1D motion in regions restricted between two adjacent obstacles and therefore reduce the overall sampling efficiency of the DNA.

Obstacle bypass is enabled by enriched hopping events

The negative effect of obstacles on overall DNA search efficiency that arises from confinement of the DBP to a restricted region is more pronounced under conditions that commit DBP sliding to a helical motion along the DNA major groove. This can be seen in both the PP (Fig. 1 B) and search redundancy (Fig. 2 A) data at relatively low salt concentrations and for low obstacle occupancies.
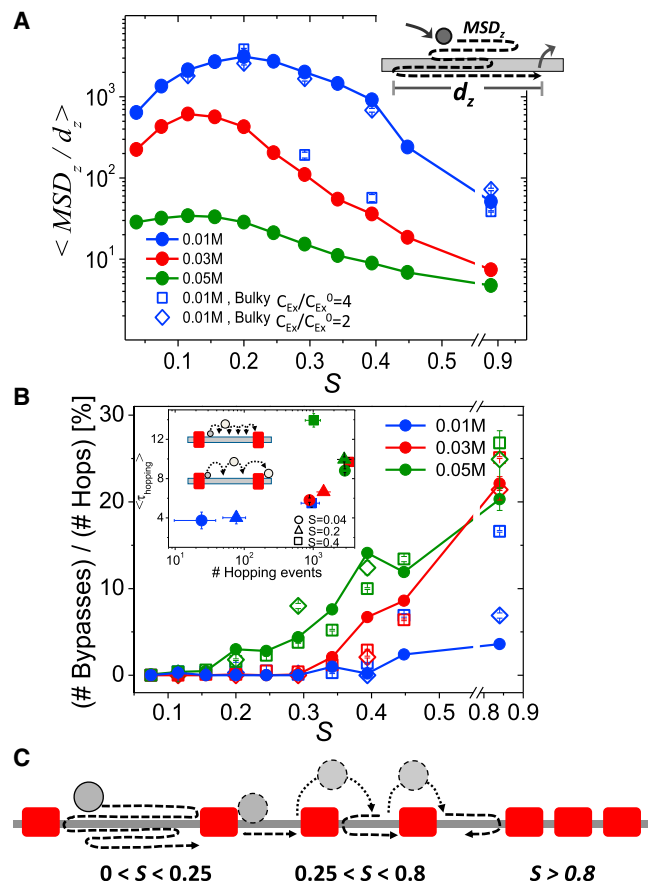

FIGURE 2 Protein confinement and obstacle bypassing. (A) Each 1D event (sliding and/or hopping dynamics), is characterized by two measures: d_z, the distance between the minimal and maximal positions of the protein along the Z-axis and MSD_z, the total mean square displacement of the protein along the Z-axis (diagrammed in the inset). Redundancy in 1D searches of DNA (MSD_z/d_z) is shown as a function of obstacle occupancy S for various salt concentrations for charged obstacles (solid circles) and for bulky obstacles (squares and diamonds) characterized by different ratios of the excluded volume repulsion radius (C_{ex}) to the radius of DNA beads in nonoccupied sites (C_{ex}^0). (B) Productive hopping events versus S for the same salt concentrations and obstacle types as in (A). The inset shows the average duration of all hopping events $\langle \tau_{hopping} \rangle$ (in units of simulation time steps) against the average number of hopping events per simulation for 0, 4, and 8 obstacles (corresponding to $S = 0.04$ (circles), 0.2 (triangles), or 0.4 (squares), respectively) for transient (unproductive) hops (upper diagram) and successful bypasses (lower diagram). (C) Protein sampling of DNA with a nonuniform obstacle occupancy including tightly packed inaccessible regions ($S > \sim 0.8$), low occupancy oversampled regions ($0 < S < \sim 0.25$), and moderately occupied efficiently sampled regions ($\sim 0.25 < S < \sim 0.8$).

However, we found that as the obstacle load on DNA increases, the protein may perform more hopping events (see Fig. S3). Because the protein is more detached from the DNA backbone during hopping than sliding, these events may enable it to bypass obstacles. Fig. 2 B shows the percentage of productive hopping events that result in bypassing an obstacle (measured as the number of hopping events from one side of an obstacle to the other side, normalized by the total number of hopping events) as a function of

S. At a low salt concentration of 0.01 M, the percentage of productive hopping events at $S \sim 0.45$ does not exceed $\sim 8\%$ (for a charged obstacle). At higher salt concentrations, hopping events are more productive and $\sim 14\%$ of hops result in crossing an obstacle at occupancies of $S \sim 0.45$.

The effect of salt concentration on the ability of a protein to bypass obstacles arises from its effects on both the total number of hopping events and the duration of each such microscopic jump (the product of the total number of hopping events and their durations, $\tau_{hopping}$, is the fraction of hopping used in the search, Fig. S3). The inset of Fig. 2 *B* shows the average duration of hopping as a function of the number of hopping events for 0, 4, and 8 obstacles on the DNA ($S = 0.04$, 0.2, and 0.4, respectively) under three different salt conditions. At a low salt concentration of 0.01 M, the number of hops dramatically increases with obstacle occupancy but their duration remains relatively short, and rarely enables obstacle bypassing. At a salt concentration of 0.03 M, the number of jumps does not increase dramatically with *S*, but their duration becomes longer and they may be more productive in bypassing obstacles (see *inset* of Fig. 2 *B*). This is consistent with the observation that, as the length of DNA flanking an obstacle decreases, the protein is able to bypass the obstacle more frequently (see Fig. S5). When further increasing the salt concentration to 0.05 M, the number of jumps at large *S* values decreases because of the dominance of 3D diffusion over sliding and hopping (Fig. S3).

Collectively, the observations of protein confinement in a redundant 1D scanning of the linker region under low obstacle occupancy (Fig. 2 *A*) and of enhanced bypassing through hopping as the obstacle occupancy increases (Fig. 2 *B*) may outline a predictive scenario for scanning an inhomogeneous genome bearing a nonhomogenous distributions of obstacles (Fig. 2 *C*). Although regions that are fully packed with obstacles ($S > \sim 0.8$) restrict protein accessibility for nonspecific DNA binding, low occupancy regions ($0 < S < \sim 0.25$) might be oversampled by a searcher DBP, and regions with moderate occupancy ($\sim 0.25 < S < \sim 0.8$) can facilitate protein bypass from one linker region to another and thus increase the number of sites to be visited by a protein. Such a mechanism may enable nucleosomal repositioning, for example, to guide regulatory proteins to key genomic regions and prevent their access to other regions, or may cause the proteins to diffuse rapidly among different genomic regions that occupy internucleosomal DNA. Fig. 3 shows scenarios of DNA sampling and obstacles bypassing by the simulated protein Sap-1 as it interacts with DNA occupied uniformly by 4 obstacles ($S = 0.2$). The protein intimately, but nonspecifically, interacts with the DNA major groove (i.e., sliding dynamics) via its recognition helix in the middle of spacer regions where the obstacles have no effects. Bypassing of the obstacles by the DBP is enabled through perturbing the rotation-coupled translation dynamics of the DBP from the helical trace of the DNA.

Protein-DNA energy landscape in the presence of obstacles: effect on bypass kinetics

Fig. 4 *A* shows three trajectories of the protein movement along the DNA axis under two salt conditions and for three different values of obstacle occupancy. The random walk nature of motion observed for the protein when no obstacles present, is restricted by the presence of the obstacles (4 ($S = 0.2$) and 6 ($S = 0.29$)). The number of transitions from one spacer region to another through obstacle bypass increases with the number of obstacles. In the absence of obstacles ($S = 0$), the motion of the DBP along the DNA is random and not restricted. Fig. 4 *B* shows that the obstacle bypass rate achieved by the protein at low and medium salt concentrations (0.01 and 0.03 M, respectively) increases with obstacle occupancy on DNA ($0.2 < S < 0.9$). The manner in which the rate increases and saturates depends on the type of obstacle and the salt conditions (e.g., the bypass rate is higher at 0.03 M compared with 0.01 M because of the greater number of hopping events, see Fig. 2 *B*), although qualitative agreement between the trends is observed.

The origin for the enhanced kinetics can be determined by plotting the mean protein-DNA energy (the sum of the electrostatic and exclusion volume energies) at each position
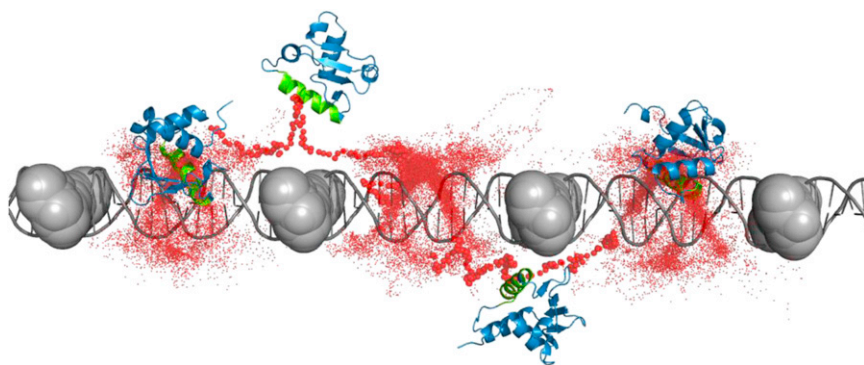

FIGURE 3 A typical trajectory of Sap-1 search on DNA associated with obstacles, under salt concentration of 0.03 M. For illustrative purposes, the obstacles (4 equally spaced charged obstacles) are shown as bulky spheres. Red dots mark the C_α position of residue 61 (from the center of the protein DNA recognition helix that is shown in *green*) during the simulation. The left and right conformations depict simulation snapshots where the protein propagates through helical sliding in which its recognition helix is situated at the DNA major groove. The two protein conformations in the middle depict bypass snapshots. Obstacle bypassing is enabled as the protein undergoes hopping and its recognition helix is excluded from the DNA.
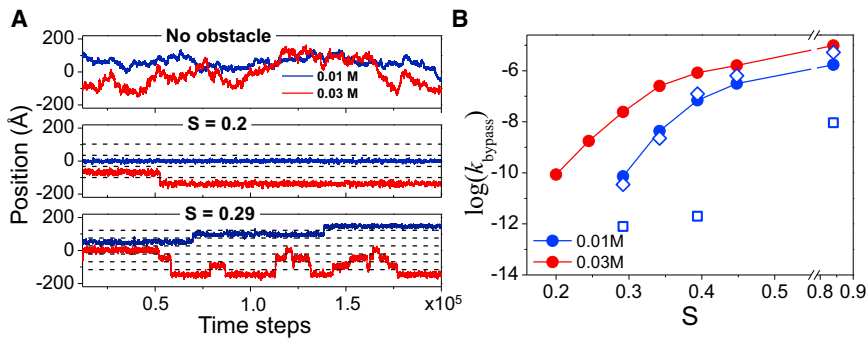
FIGURE 4 Obstacle bypass kinetics. (A) Raw trajectories data of the displacement of the protein along the DNA axis (Z-axis) under salt concentrations of 0.01 M (*blue*) and 0.03 M (*red*). Upper panel: No obstacles ($S = 0$), middle panel: 4 obstacles ($S = 0.2$), low panel: 6 obstacles ($S = 0.29$). Dashed black lines denote the positions of obstacles. (B) Dependence of obstacle bypass kinetics (k_{bypass}) on obstacle occupancy S at various salt concentrations for charged (*circles*) and bulky obstacles with C_{ex}/C_{ex}^0 equals 2 (*diamonds*) or 4 (*squares*).

along the DNA. Fig. 5 A shows the energy of the DBP on DNA that is loaded with 4 obstacles (*top*, $S = 0.2$) and 6 obstacles (*bottom*, $S = 0.29$). Notably, the energetic barrier for crossing the obstacles is lower for the more crowded DNA. In addition, the bypass barrier is lower for the higher salt concentration.

To rationalize the linkage between the energetic barrier for bypassing obstacles and their density on the DNA, we performed simplified Monte Carlo simulations in which we measured the bypass kinetics for a random walker on a 1D lattice that is uniformly loaded with obstacles. In this model, the obstacles were represented by Gaussian potentials (defined by their width σ) and were positioned at equal spacings L along the 1D lattice. Under crowded conditions, the Gaussian tails of the obstacles overlap, therefore elevating the potential energy between the Gaussian peaks and reducing the energy barrier for crossing the obstacles (Fig. 5 B and Fig. S6).

Dynamic obstacles impair DNA search less than stationary ones

In vivo, a searching DBP may encounter other DNA-bound macromolecules, such as other DBPs, that search for their own target sites on the DNA. Although some obstacles, such as nucleosomes, are relatively stationary, others, such as enzymes or transcription factors, may be more mobile. To quantify the effects of obstacle dynamics on search mechanism, we simulated a random walker (representing the DBP) on a 1D lattice (representing DNA) at constant

obstacle occupancies in the range of 0–95% (each obstacle occupies a lattice site that is inaccessible to the walker). The walker moves one-dimensionally from one site to another with a characteristic rate of $k_{Prot,1D}$, as each obstacle is evacuated from its current site and immediately binds randomly at another nonoccupied site at a rate of k_{Obst} (Fig. S7). Thus, the dynamics of the obstacles is incorporated into the model through μ, which is the ratio of the two characteristic rate constants ($\mu = k_{Obst}/k_{Prot,1D}$). In addition, we introduced a salt dependency that came into play in the ratio of rate constants, $\rho = k_{1D\text{-}3D}/k_{3D\text{-}1D}$, where $k_{1D\text{-}3D}$ and $k_{3D\text{-}1D}$ are the rate constants for protein transition from the DNA into the bulk and vice versa, respectively. At a low ρ value corresponding to a low salt concentration, the protein is predominantly bound to the 1D lattice, whereas at higher ρ values the protein dissociates from the lattice and rebinds randomly to an unoccupied site. We studied five values of ρ (see Table S1).

We measured search efficiency (calculated as the ratio of the number of visited sites to the total number of sites in the lattice) for various conditions of μ, ρ, and percentage of obstacle occupancy. Beyond the pronounced, but expected, improvement in search efficiency at higher salt concentrations (at which dissociation of the walker from the lattice is followed by reassociation at a distant site), the results clearly show that increasing the obstacle dynamics (increasing μ) enables the random walker to maintain lattice sampling efficiency with no interruptions even at high occupancies of ~75% (Fig. 6, A and B). Fig. 6 C illustrates the sampling efficiency for a random walker
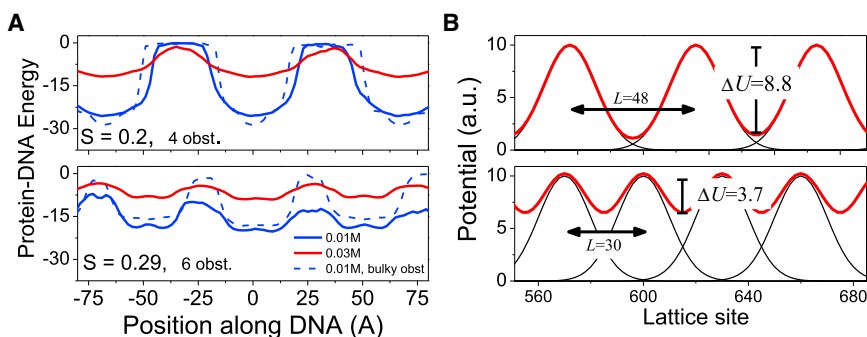

FIGURE 5 Protein-DNA energy landscape. (A) Protein-DNA energy landscape at two salt concentrations and under two obstacle occupancy conditions on DNA loaded with charged (*smooth line*) or bulky (*dotted line*) obstacles. (B) Monte Carlo simulation landscapes for a random walker on a 1D lattice showing a reduced energy barrier for bypassing under increasing obstacle occupancy.
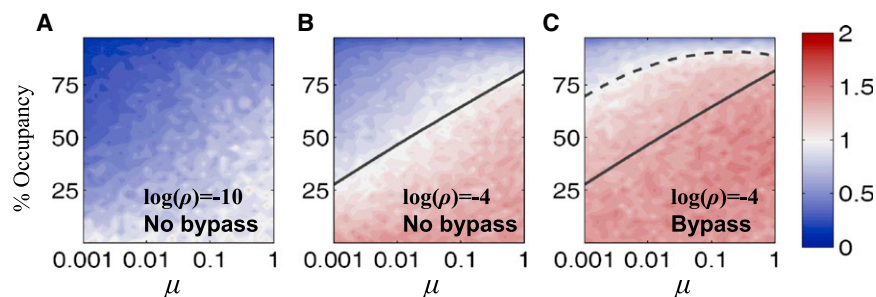
FIGURE 6 Effect of mobile obstacles on DNA sampling efficiency. Lattice simulations of a protein (*random walker*) moving at a rate $k_{Prot,1D}$ on a 1D lattice containing mobile obstacles. The protein may dissociate to the bulk at rate k_{1D-3D} or reassociate at rate k_{3D-1D}. The obstacles dissociate from their sites at a rate k_{Obst}. Immobile obstacles have a small μ value, dynamic obstacles have a large μ value, and ρ indicates the tendency of the protein to dissociate from the DNA and so corresponds to the salt concentration. (*A* and *B*) Lattice sampling efficiency (average of 10 simulations, normalized by the mean at 0% occupancy) under low salt (*A*, $\rho = 10^{-10}$) and higher salt (*B*, $\rho = 10^{-4}$) conditions. (*C*) As in *C*, but here the protein may hop over single obstacles with a probability of 0.5. The black lines in *B* and *C*, correspond to a sampling of obstacle-covered DNA with sampling efficiency similar to that achieved on naked DNA when obstacle bypassing are excluded (*solid*) or included (*dashed*).

that can bypass a single obstacle with probability of 0.5. The ability of the protein to hop enables it to overcome immobile obstacles and scan the DNA efficiently at occupancies as high as 50%. Our observations imply that smaller proteins that diffuse relatively fast will experience obstacles as static objects, whereas large proteins experience them as dynamic objects (i.e., for a given DNA environment, $\mu_{large-protein} > \mu_{small-protein}$). We expect therefore, that small DBPs will be more affected by crowding on DNA in comparison with larger DBPs that search for their regulatory sites.

CONCLUSIONS

We demonstrated that, as the DBP performs more sliding along the DNA, its search efficiency in the presence of obstacles may be reduced as its dynamics confine it to redundant 1D scanning of the spacer region between two adjacent obstacles. This finding may imply that an inhomogeneous positioning of obstacles along DNA can sequester sampling to within certain regions while reducing the possibility of searching other regions, as is the scenario in yeast genomes, for example, where most regulatory regions exhibit low nucleosomal occupancies (26). The sequestration effect by the obstacle may result in a slower search when the target site is not located in the accessible region that is defined by the obstacles.

Although the obstacles may restrict the DNA region that is accessible for searching, they may be bypassed during hopping events. The number of hops and their duration, which could differ between different DBPs and can be modulated by the salt concentration, might determine the possibility of jumping over obstacles. More importantly, at a relatively high obstacle density, the nonspecific interaction between the DBP and the DNA is weaker compared with less crowded DNA and the protein therefore performs more hopping at the expense of sliding. This observation serves as another example of the importance of hopping in searching DNA (34,45–48). The enhanced hopping dynamics may be viewed also as an outcome of the

more frequent collisions of the DBP with obstacles as the density of the latter increases. Alternatively, the hopping may be viewed as two-dimensional sliding, which was proposed as a mechanism for bypassing obstacles (35). We anticipate that protein hopping in looped DNA regions may facilitate intersegmental transfers between DNA regions far apart in sequence (Fig. 1 *A*), which further assist obstacle bypassing. Increasing obstacle occupancy thus reduces the energetic barrier for bypassing an obstacle. By contrast, on less crowded DNA, obstacle effects become more localized and the protein may be trapped and perform a redundant and inefficient sampling of the linker region (49,50). In causing the protein to linger in a particular region, the obstacles may mask other DNA regions and thus bias the search to the linker region and reduce the search time (51). On the other hand, causing the protein to pause in a particular region may be advantageous if, by increasing the search time, it enables the search to be better synchronized with other cellular processes (22).

Our lattice simulations imply that dynamic obstacles, such as other DBPs that search for their own target DNA sites, are far less of an impediment to the search process than immobile obstacles. Consequently, a protein may be able to efficiently sample DNA even if it is relatively densely loaded with obstacles. Thus, the synergism between obstacle occupancy, the intrinsic dynamics of obstacles, and the salt concentration may facilitate the search and contribute to the ability of proteins to navigate through the complex DNA organization to find their regulatory binding sites.

SUPPORTING MATERIAL

One table and seven figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)00369-X.

REFERENCES

1. Zakrzewska, K., and R. Lavery. 2012. Towards a molecular view of transcriptional control. *Curr. Opin. Struct. Biol.* 22:160–167.

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

3. Berg, O. G., R. B. Winter, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry.* 20:6929–6948.

4. von Hippel, P. H., and O. G. Berg. 1989. Facilitated target location in biological systems. *J. Biol. Chem.* 264:675–678.

5. Marcovitz, A., and Y. Levy. 2011. Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc. Natl. Acad. Sci. USA.* 108:17957–17962.

6. Record, Jr., M. T., J. H. Ha, and M. A. Fisher. 1991. Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol.* 208:291–343.

7. Jack, W. E., B. J. Terry, and P. Modrich. 1982. Involvement of outside DNA sequences in the major kinetic path by which *Eco*RI endonuclease locates and leaves its recognition sequence. *Proc. Natl. Acad. Sci. USA.* 79:4010–4014.

8. Gowers, D. M., G. G. Wilson, and S. E. Halford. 2005. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. USA.* 102:15883–15888.

9. Iwahara, J., M. Zweckstetter, and G. M. Clore. 2006. NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. USA.* 103:15062–15067.

10. Gorman, J., and E. C. Greene. 2008. Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* 15:768–774.

11. Elf, J., G. W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science.* 316:1191–1194.

12. Hammar, P., P. Leroy, …, J. Elf. 2012. The lac repressor displays facilitated diffusion in living cells. *Science.* 336:1595–1598.

13. Blainey, P. C., G. Luo, …, X. S. Xie. 2009. Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.* 16:1224–1229.

14. Marcovitz, A., and Y. Levy. 2012. Sliding Dynamics Along DNA: A Molecular Perspective. RSC, United Kingdom.

15. Slutsky, M., and L. A. Mirny. 2004. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* 87:4021–4035.

16. Kolomeisky, A. B. 2011. Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* 13:2088–2095.

17. Dahirel, V., F. Paillusson, …, J. M. Victor. 2009. Nonspecific DNA-protein interaction: why proteins can diffuse along DNA. *Phys. Rev. Lett.* 102:228101.

18. Sheinman, M., O. Bénichou, …, R. Voituriez. 2012. Classes of fast and specific search mechanisms for proteins on DNA. *Rep. Prog. Phys.* 75:026601.

19. Bénichou, O., Y. Kafri, …, R. Voituriez. 2009. Searching fast for a target on DNA without falling to traps. *Phys. Rev. Lett.* 103:138102.

20. Hu, T., A. Y. Grosberg, and B. I. Shklovskii. 2006. How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys. J.* 90:2731–2744.

21. Klenin, K. V., H. Merlitz, …, C. X. Wu. 2006. Facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.* 96:018104.

22. Li, G., O. Berg, and J. Elf. 2009. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* 5:294–297.

23. Bénichou, O., C. Chevalier, …, R. Voituriez. 2011. Facilitated diffusion of proteins on chromatin. *Phys. Rev. Lett.* 106:038102.

24. Lomholt, M. A., B. van den Broek, …, R. Metzler. 2009. Facilitated diffusion with DNA coiling. *Proc. Natl. Acad. Sci. USA.* 106:8204–8208.

25. Shivaswamy, S., A. Bhinge, …, V. R. Iyer. 2008. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* 6:e65.

26. Kaplan, N., I. K. Moore, …, E. Segal. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature.* 458:362–366.

27. Poirier, M. G., M. Bussiek, …, J. Widom. 2008. Spontaneous access to DNA target sites in folded chromatin fibers. *J. Mol. Biol.* 379:772–786.

28. Ali Azam, T., A. Iwata, …, A. Ishihama. 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* 181:6361–6370.

29. Tafvizi, A., L. A. Mirny, and A. M. van Oijen. 2011. Dancing on DNA: kinetic aspects of search processes on DNA. *ChemPhysChem.* 12:1481–1489.

30. Flyvbjerg, H., S. A. Keatch, and D. T. Dryden. 2006. Strong physical constraints on sequence-specific target location by proteins on DNA molecules. *Nucleic Acids Res.* 34:2550–2557.

31. Sabo, P. J., M. S. Kuehn, …, J. A. Stamatoyannopoulos. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat. Methods.* 3:511–518.

32. Zhou, H. X. 2005. A model for the mediation of processivity of DNA-targeting proteins by nonspecific binding: dependence on DNA length and presence of obstacles. *Biophys. J.* 88:1608–1615.

33. Gorman, J., A. J. Plys, …, E. C. Greene. 2010. Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.* 17:932–938.

34. Hedglin, M., and P. J. O'Brien. 2010. Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. *ACS Chem. Biol.* 5:427–436.

35. Kampmann, M. 2004. Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding. *J. Biol. Chem.* 279:38715–38720.

36. Givaty, O., and Y. Levy. 2009. Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.* 385:1087–1097.

37. Zandarashvili, L., D. Vuzman, …, J. Iwahara. 2012. Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. USA.* 109:E1724–E1732.

38. Terakawa, T., H. Kenzaki, and S. Takada. 2012. p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains. *J. Am. Chem. Soc.* 134:14555–14562.

39. Mor, A., G. Ziv, and Y. Levy. 2008. Simulations of proteins with inhomogeneous degrees of freedom: the effect of thermostats. *J. Comput. Chem.* 29:1992–1998.

40. Vuzman, D., and Y. Levy. 2010. DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc. Natl. Acad. Sci. USA.* 107:21004–21009.

41. Vuzman, D., A. Azia, and Y. Levy. 2010. Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *J. Mol. Biol.* 396:674–684.

42. Vuzman, D., M. Polonsky, and Y. Levy. 2010. Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. *Biophys. J.* 99:1202–1211.

43. Khazanov, N., and Y. Levy. 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.

44. Sheinman, M., and Y. Kafri. 2009. The effects of intersegmental transfers on target location by proteins. *Phys. Biol.* 6:016003.

45. Halford, S. E., and J. F. Marko. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040–3052.

46. Halford, S. E. 2009. An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.* 37:343–348.

47. Schonhoft, J. D., and J. T. Stivers. 2012. Timing facilitated site transfer of an enzyme on DNA. *Nat. Chem. Biol.* 8:205–210.

48. Bonnet, I., A. Biebricher, …, P. Desbiolles. 2008. Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.* 36:4118–4127.

49. Jeltsch, A., J. Alves, …, A. Pingoud. 1994. Pausing of the restriction endonuclease *Eco*RI during linear diffusion on DNA. *Biochemistry.* 33:10215–10219.

50. Pluciennik, A., and P. Modrich. 2007. Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair. *Proc. Natl. Acad. Sci. USA.* 104:12709–12713.

51. Mirny, L., M. Slutsky, …, A. Kosmrlj. 2009. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.* 42:434013.

Obstacles may facilitate and direct DNA search by proteins

Amir Marcovitz and Yaakov Levy

Supporting Information

Salt condition	1	2	3	4	5
k_{1D-3D} (τ_s^{-1})	10^{-10}	10^{-6}	$5 \cdot 10^{-6}$	$5 \cdot 10^{-6}$	$5 \cdot 10^{-4}$
k_{3D-1D} (τ_s^{-1})	1	10^{-2}	$5 \cdot 10^{-4}$	$5 \cdot 10^{-6}$	$5 \cdot 10^{-6}$
ρ	10^{-10}	10^{-4}	10^{-2}	1	10^{2}

Table S1. Salt (ρ) effects are introduced into the lattice simulations. Each salt has two rate constants for on-off and off-on lattice transitions. At a low salt concentration ($\rho = 10^{-10}$) where $k_{1D-3D} << k_{3D-1D}$ the protein remains in the on-lattice state essentially all the simulation. As we increase the salt concentration, the partition between the 1D and off lattice modes gradually changes in favor of the latter.
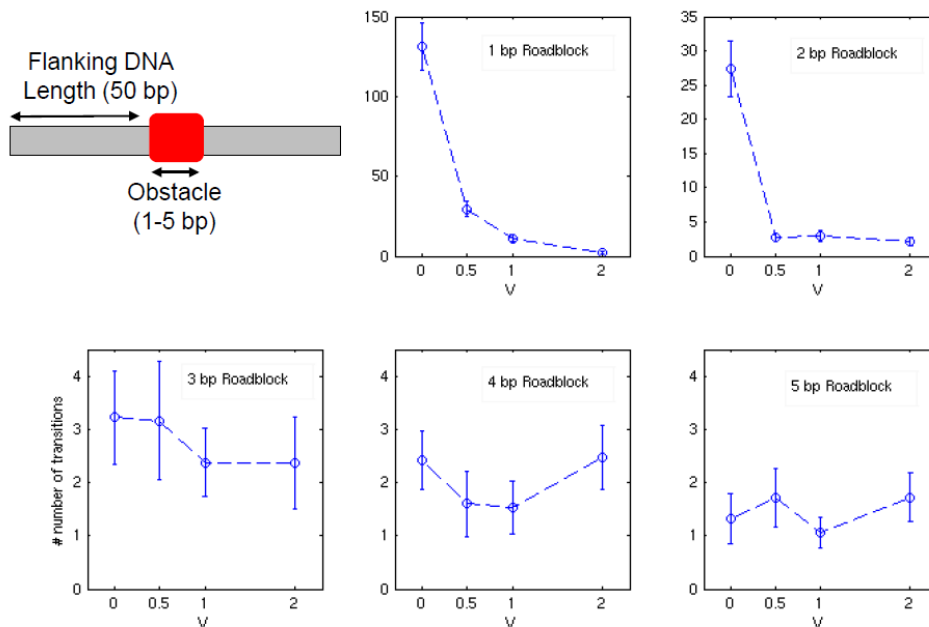
Figure S1

Effects of roadblock size on protein's ability to take a bypass from one side of an obstacle to another. Charged obstacles were tried with a repulsion factor, V, of 0, 0.5, 1 and 2 (the magnitude of the electrostatic repulsion). Obstacles that occupy more than ~3-4 bp are restrictive to protein bypass, even at lower repulsion. An obstacle occupying 2bp was extensively simulated in this study, due to its ability to block DBP bypass as it is electrostaticaly repel the DBP. For comparison, the number of transitions on a naked DNA under the same condition is about 2000.
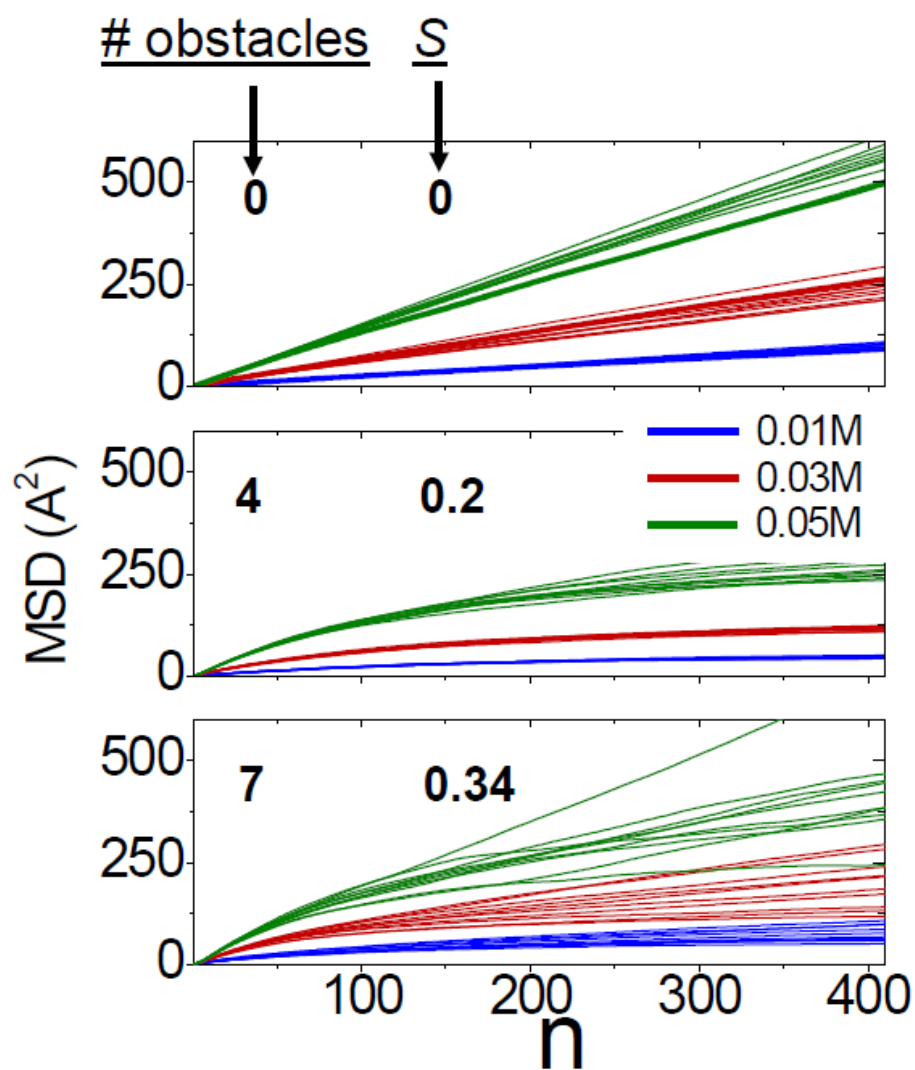
Figure S2

Mean square displacement (MSD) profiles for 10 trajectories simulated at three different salt concentrations (0.01 M, 0.03 M and 0.05 M), and three different obstacle occupancies, S. The slopes of the MSD profiles reflect 1D diffusion coefficient D_1. The minimal 1D diffusion is apparent under low occupancy of 4 roadblocks per 100bp dsDNA (S=0.2)
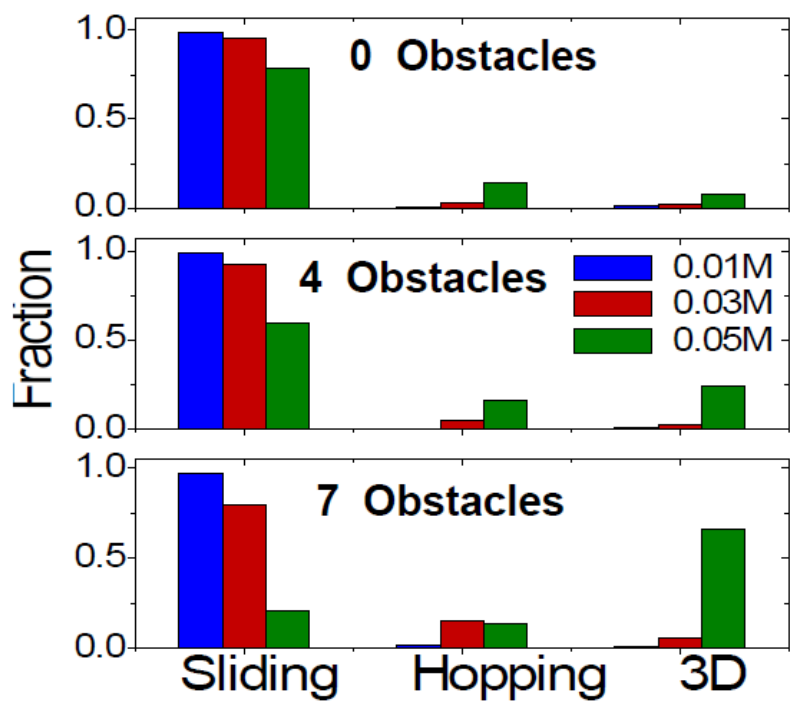
Figure S3

The fraction of sliding, hopping, and 3D diffusion for each salt concentration and *S*. Each snapshot of the simulation is classified into one of these three search mechanisms.
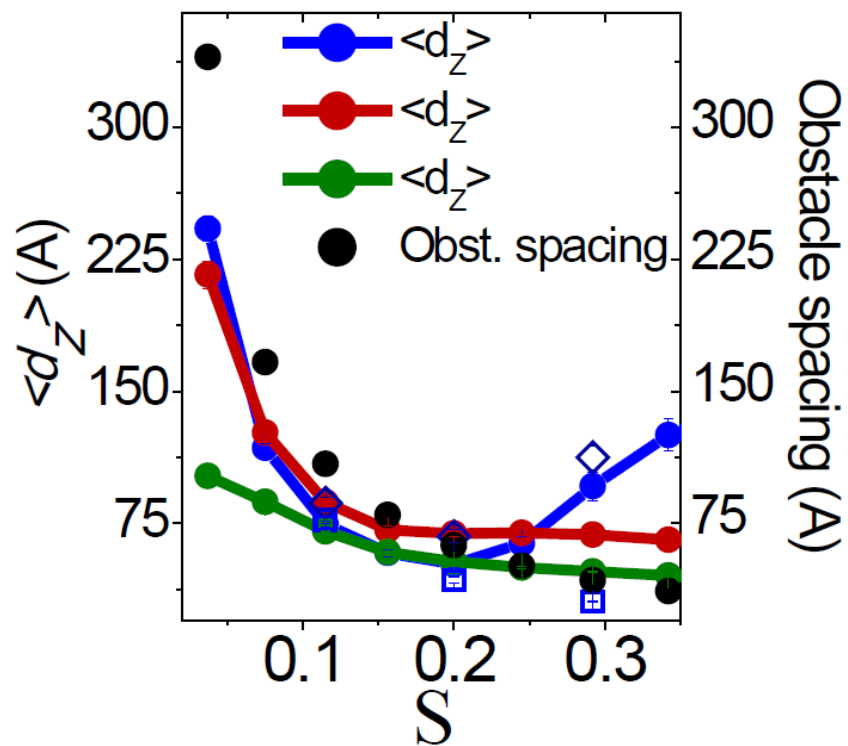
Figure S4

Mean value of d_z as a function of S. The black dots indicate the length of the spacing region between two adjacent obstacles. At a low salt concentration (<0.05 M) and for small S values (<0.25), d_z is similar to the spacing distance between two adjacent obstacles, so illustrating the confinement of the protein and the high degree of redundancy in the search.
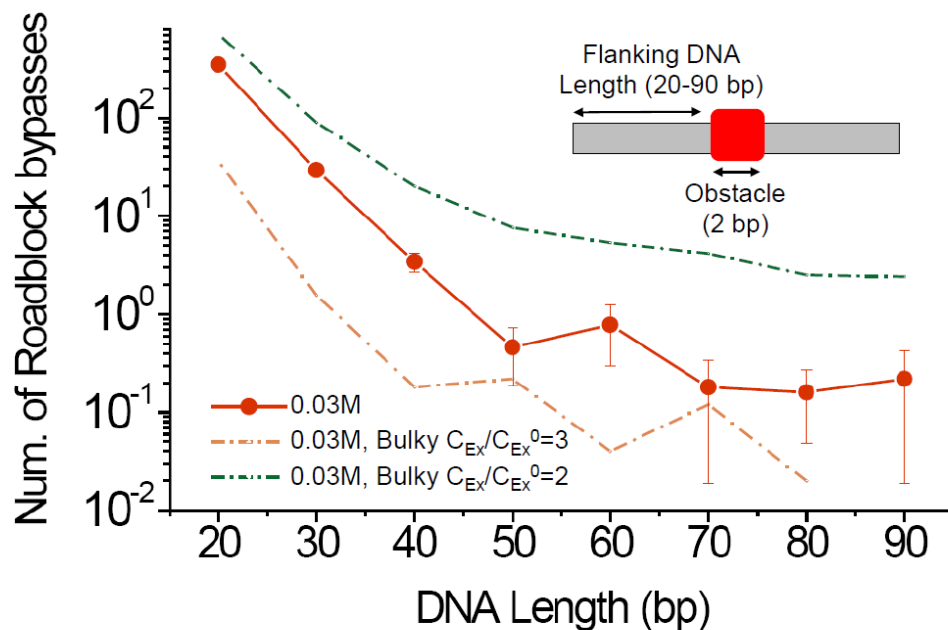
Figure S5

Number of bypasses over a single roadblock in the middle of a DNA molecule of varying lengths (20-90 bp), under salt concentration of 0.03M. As the length of DNA flanking the roadblock is shorter, the number of bypasses increases. Several obstacle type were tested (a charged obstacle and two types of bulky obstacles characterized by different ratios of the excluded volume repulsion radius (C_{ex}) to the radius of DNA beads in non-occupied sites (C_{ex}^0)).
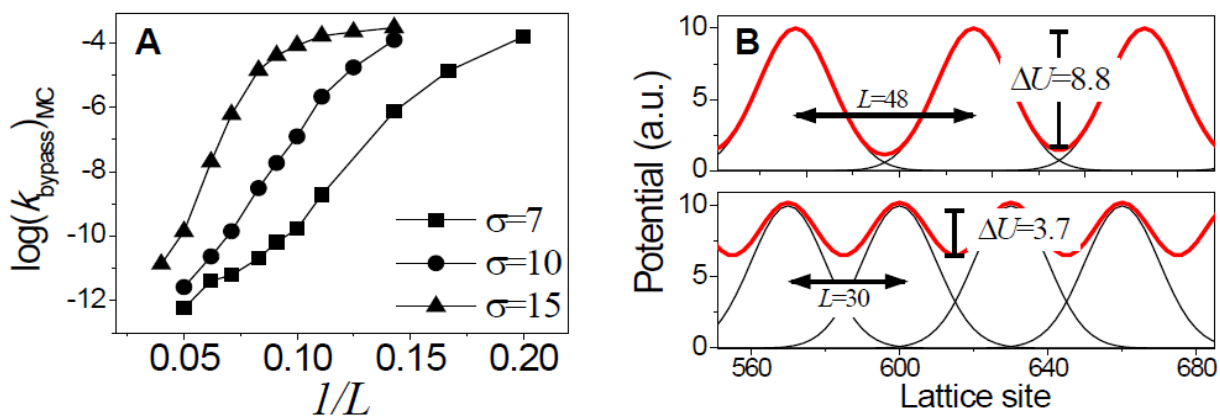
Figure S6

Obstacle and energy landscape kinetics. a) Kinetics of a random walker bypassing Gaussian obstacles on a 1D lattice in Monte-Carlo simulations. Gaussians with different dimensionalities (tuned by the width of the Gaussian, σ) were tested while varying their density ($1/L$) on the lattice. b) Decomposition of the lattice potential into its individual Gaussian potentials for two lattices with differently spaced Gaussians.
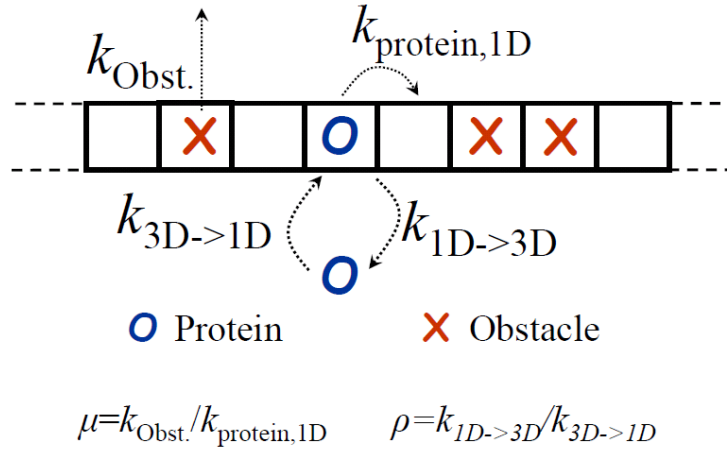
Figure S7

Effect of mobile obstacles on DNA sampling efficiency. Lattice simulations of a protein (random-walker) moving at a rate $k_{\text{Prot,1D}}$ on a 1D lattice containing mobile obstacles. The protein may dissociate to the bulk at rate $k_{\text{1D-3D}}$ or reassociate at rate $k_{\text{3D-1D}}$. The obstacles dissociate from their sites at a rate k_{Obst}. Immobile obstacles have a small μ value, dynamic obstacles have a large μ value, and ρ indicates the tendency of the protein to dissociate from the DNA and so corresponds to salt concentration.