

Published in final edited form as:

*Phys Rev E Stat Nonlin Soft Matter Phys.* 2011 February ; 83(2 0 1): 021907.

## Protein sliding and hopping kinetics on DNA

Michael C. DeSantis<sup>1</sup>, Je-Luen Li<sup>2</sup>, and Y. M. Wang<sup>1</sup>

<sup>1</sup>Department of Physics, Washington University in St. Louis, St. Louis, Missouri 63130, USA

<sup>2</sup>D. E. Shaw Research, New York, New York 10036, USA

### Abstract

Using Monte Carlo simulations, we deconvolved the sliding and hopping kinetics of GFP-LacI proteins on elongated DNA from their experimentally observed seconds-long diffusion trajectories. Our simulations suggest the following results: (i) in each diffusion trajectory, a protein makes on average hundreds of alternating slides and hops with a mean sliding time of several tens of milliseconds; (ii) sliding dominates the root-mean-square displacement of fast diffusion trajectories, whereas hopping dominates slow ones; (iii) flow and variations in salt concentration have limited effects on hopping kinetics, while *in vivo* DNA configuration is not expected to influence sliding kinetics; and (iv) the rate of occurrence for hops longer than 200 nm agrees with experimental data for EcoRV proteins.

### I. INTRODUCTION

Timely target association of DNA-binding (DB) proteins is important for prompt cellular response to external stimuli using mechanisms such as gene regulation, DNA replication, and DNA repair. The target association rates of DB proteins frequently deviate from the diffusion limit due to their interactions with nonspecific DNA via the process of facilitated diffusion [1–3]. Facilitated diffusion mainly consists of two motions: sliding, where a protein diffuses along nonspecific DNA without losing contact, and hopping, where the protein jumps off DNA and undergoes 3D diffusion before reassociating to the same (Fig. 1) or a different segment of DNA (referred to as intersegmental transfer). In this article, we regard events with long hopping distances, usually called jumping, as a form of hopping. A DB protein may slide and hop many times on nonspecific DNA before reaching the target. In order to quantify the effect of facilitated diffusion on DB proteins' target binding rate, how long a protein spends sliding on DNA (mean sliding time  $\langle t_1 \rangle$ ) and how fast it moves along DNA (sliding diffusion coefficient  $D_1$ ) are two critical parameters for all calculations of *in vitro* and *in vivo* DNA geometries [2,4–8].

Single-molecule (SM) fluorescence imaging studies of DB proteins' Brownian diffusion along elongated DNA have obtained effective diffusion coefficients  $D$  for the whole seconds-long diffusions (in this article we define each observed diffusion event between protein association and permanent dissociation to be a diffusion trajectory, and  $t$  is the total time of the diffusion) [3,9–21]. In the past, numerous studies had substituted  $t$  and  $D$  values in the place of  $\langle t_1 \rangle$  and  $D_1$  in target binding rate and protein-nonspecific-DNA binding energy calculations since  $\langle t_1 \rangle$  and  $D_1$  were not experimentally accessible [3,5–8,10,12,17,22]. Since the extent of hopping involvement is unknown, it is dubious to use  $t$  and  $D$  values for  $\langle t_1 \rangle$  and  $D_1$ . Recent evidence suggests that these diffusion trajectories include both sliding and hopping: (i) the sliding time of DB proteins has been estimated to be milliseconds [6,12,22,23]; (ii) the sliding displacement has been estimated to be less than 50 bp [24], shorter than the displacements of whole diffusion trajectories of the reported DB proteins (>100 nm); and (iii) hops longer than 200 nm have been observed [15]. In order to

obtain  $\langle t_1 \rangle$  and  $D_1$  from experimental data, deconvolving sliding and hopping from individual diffusion trajectories is necessary.

## II. SIMULATIONS

Here we deconvolve sliding and hopping in a diffusion trajectory and obtain  $\langle t_1 \rangle$  and  $D_1$  using (i) Monte Carlo simulations, (ii) experimental  $D$  and  $t$  values, and (iii) the following two relations (derived in Ref. [25]):

$$t = N \langle t_1 \rangle + N \langle t_3 \rangle, \quad (1)$$

$$2Dt = 2D_1N \langle t_1 \rangle + 2D_3N \langle t_3 \rangle, \quad (2)$$

where  $N$  is the mean number of sliding and hopping alternations in a diffusion trajectory,  $D_3$  is the 3D diffusion coefficient of the protein, and  $\langle t_3 \rangle$  is the mean hopping time. From hopping simulations we first determine  $N$  and  $\langle t_3 \rangle$ ; then, combining with experimental  $D$  and  $t$  values,  $t_1$  and  $D_1$  are obtained using Eqs. (1) and (2).

For each hopping simulation, a protein was initially positioned at the protein-center-to-DNA-center distance of  $R = r_{\text{DNA}} + r_{\text{protein}} + \Delta r$ , where  $r_{\text{DNA}} = 1$  nm is the DNA radius,  $r_{\text{GFP-LacI}} = 2.68$  nm, and  $\Delta r \approx 0.5$  nm is an estimate of the protein-DNA binding distance (or location of the interaction potential minimum beyond which we consider no protein-DNA interactions) [26,27]. The protein immediately dissociates from DNA and undergoes 3D diffusion until rebinding to DNA, at which time the position was recorded, or until the maximum number of steps of the hopping simulation was reached, in which case the protein was assumed to have permanently dissociated and its diffusion trajectory was not used in subsequent data analysis. Figure 2 describes the criterion for determining whether a hopping protein collided with DNA. For every step, the length of the perpendicular drawn from the center of the DNA to the line connecting the last two protein locations (dashed arrow) was calculated and if less than  $R$ , association occurred. The binding position was chosen to be the midpoint between the two protein locations. We have modeled DNA as an infinite, rigid cylinder assuming 100% probability for association on protein-DNA collision; the distance between the protein binding location and its origin denotes the hopping distance.

The simulation parameters were determined as follows. The hopping simulation step size  $\delta$ , and step time  $\tau$ , are the collision distance and time, respectively [28]. At temperature  $T = 294$  K, the instantaneous velocity of a protein of mass  $m$  in solution is the root-mean-square

(rms) velocity  $\sqrt{\langle v_x^2 \rangle} = \sqrt{k_B T / m} = \delta / \tau = 6.02$  m/s, where  $k_B$  is the Boltzmann constant, and  $m = 67.5$  kDa for a GFP-LacI monomer. Using the Einstein-Stokes relation,  $D_3 = \delta^2 / (2\tau) = k_B T / 6\pi\eta r = 8.03 \times 10^7$  nm<sup>2</sup>/s for GFP-LacI, where the viscosity of water is  $\eta = 10^{-3}$  N s/m<sup>2</sup> and the protein hydrodynamic radius  $r$  is 2.68 nm, assuming a typical protein density of 1.38

g/cm<sup>3</sup>, we obtain  $\delta = 2D_3 / \sqrt{\langle v_x^2 \rangle}$ . Therefore,  $\delta = 0.267$  Å and  $\tau = 4.46$  ps. Each simulation step in the  $x, y, z$  dimensions was drawn from a Gaussian distribution with a mean of zero and a standard deviation of  $\delta$ .

The time limit for simulation of each GFP-LacI hop was  $\approx 1$  ms (or  $2.1 \times 10^8$  steps), selected according to the following two estimations: (i) Since the observed diffusion of proteins on DNA is the combination of sliding and hopping with diffusion coefficients  $D_1$  and  $D_3$ , respectively, the maximum total hopping time of a diffusion trajectory cannot exceed  $Nt_{3,\text{max}} = Dt/D_3$  when  $D_1 \approx 0$ . For GFP-LacI,  $\langle D \rangle \approx 2 \times 10^4$  nm<sup>2</sup>/s [3] which dictates that  $t_{3,\text{max}} \approx 0.25$  ms when  $t$  is on the order of 1 s and using the low bound for  $N$  of one hop

per diffusion trajectory. Therefore, a hopping time limit of  $t_{3,\max} \approx 1$  ms for a single hop should be sufficiently long for all 3D diffusing proteins to return to DNA. (ii) A longer hopping time limit, such as 10 ms per hop (data not shown), results in additional proteins returning to DNA with individual hopping distances longer than  $\sqrt{2\langle D \rangle t} = 200$  nm, a detectable distance in SM measurements that are usually used to separate single diffusion trajectories into segments free of large displacements for accurate  $D$  analysis [3,15].

### III. RESULTS AND DISCUSSION

For  $4 \times 10^5$  GFP-LacI hopping simulations (maximum simulation time  $t_{3,\max} \approx 1$  ms) with  $\delta = 0.267$  Å and  $R = 4.2$  nm, 99.809% of these trials resulted in the protein reassociating to DNA and thus the probability for a simulated hop to return to DNA is  $P = 0.99809$ . The hopping characteristics are shown in Figs. 3(a) and 3(b), in which the mean hopping distance along DNA is 3.37 Å (median, 0.41 Å), the mean hopping height (the maximum radial distance of the protein from DNA) is 4.93 Å (median, 0.45 Å), and the mean number of steps per hop is  $4.97 \times 10^4$  (median, 5), yielding a mean hopping time of  $\langle t_3 \rangle = 0.22$  μs. The mean number of hops in a GFP-LacI diffusion trajectory is  $N = 526$ , obtained by dividing the total number of simulated hops of  $4 \times 10^5$  by the total number of nonreturned hopping events of 763; the distribution for the number of hops per diffusion trajectory is shown in Fig. 4. This set of values have been verified to converge with those from a larger simulation of  $4 \times 10^6$  hops. Specifically,  $N$  values differ by 0.57%. The inset of Fig. 3 shows the distribution of total hopping displacements in a diffusion trajectory with each data point simulated from 526 randomly selected hopping displacements. The rms total hopping displacements per diffusion trajectory is  $127.5$  nm ( $\sqrt{2D_3N\langle t_3 \rangle}$ ), and the mean total hopping time is  $N\langle t_3 \rangle = 115$  μs. Note that although shorter hopping distances, such as ones less than the base pair length of 0.34 Å, do not carry direct biological significance nor do they noticeably disrupt sliding, they are important for correctly assessing rms total hopping displacement statistics in a diffusion trajectory.

We can also compute the “diffusion to capture” probability  $P$  for a protein to return to DNA using a steady-state solution to the diffusion equation, incorporating a cutoff radial distance  $c$  [28]. Proteins released after the initial step at  $b = 4.22$  nm are either adsorbed at the DNA surface ( $R = 4.2$  nm) or escape beyond  $c = R + \sqrt{4D_3t_{3,\max}}$ . The probability is time independent and given by

$$P = \frac{\log(c/b)}{\log(c/R)} = 0.998 \quad 96. \quad (3)$$

Imposing the same cutoff distance  $c = 551.2$  nm in subsequent simulations, we obtained  $P = 0.998$  65, in near agreement with the analytical value above.

Having obtained  $\langle t_3 \rangle$  and  $N$  from simulation, we now solve Eqs. (1) and (2) for  $\langle t_1 \rangle$  and  $D_1$  from the experimentally measured values of  $t$  and  $D$ . With values of  $D$  for GFP-LacI ranging from  $2.3 \times 10^2$  to  $1.3 \times 10^5$  nm<sup>2</sup>/s [3] and  $t = 10.4$  s [Fig. 3(d)],

$$\langle t_1 \rangle = \frac{t}{N} - \langle t_3 \rangle \approx \frac{t}{N} = 19.8 \quad (\text{ms}), \quad (4)$$

$$D_1 \approx D - \frac{D_3N\langle t_3 \rangle}{t} = D - \frac{9309}{10.4} (\text{nm})^2 / (\text{s}). \quad (5)$$

The sliding time is several tens of ms and  $D_1$  ranges from  $\approx 0$  for slow diffusion to  $\approx D$  for fast diffusion. The  $\langle D_1 \rangle$  for GFP-LacI is  $9.1 \times 10^3 \text{ nm}^2/\text{s}$  using  $\langle D \rangle$  of  $10^4 \text{ nm}^2/\text{s}$ . Since  $D_1 > 0$ , Eq. (5) sets the lower bound of  $D$  such that it must be greater than  $D_3 N \langle t_3 \rangle / t \approx 896 \text{ nm}^2/\text{s}$ . The rms total sliding displacement in a diffusion trajectory becomes longer than the rms total hopping displacement when  $D > 2ND_3 t_3 / t \approx 1790 \text{ nm}^2/\text{s}$ .

Since our protein-nonspecific-DNA binding distance is an estimate, we have carried out simulations with  $\Delta r$  ranging from 0.5 to 6.5 nm (corresponding to protein-DNA distances  $R$  of 4.2 and 10.2 nm, respectively). Comparing the  $R = 10.2 \text{ nm}$  results to the  $R = 4.2 \text{ nm}$  results, the distributions for hopping distances [Fig. 3(a)] and hopping times [Fig. 3(b)] are similar, although the mean hopping distance reduces to  $2.82 \text{ \AA}$ , the mean number of steps per hop reduces to  $3.23 \times 10^4$ , and the mean number of hops  $N$  doubles to 1101. Solving for  $\langle t_1 \rangle$  and  $D_1$  at  $R = 10.2 \text{ nm}$ , we found  $\langle t_3 \rangle = 0.14 \mu\text{s}$ ,  $N \langle t_3 \rangle = 154 \mu\text{s}$ ,  $\langle t_1 \rangle = 9.4 \text{ ms}$  (approximately half of the value for  $R = 4.2 \text{ nm}$ ), and  $D_1$  to be similar to the previously calculated value for  $R = 4.2 \text{ nm}$ . Given that the sliding and hopping values at  $R = 4.2$  and  $10.2 \text{ nm}$  are close, our method and results can be safely applied to most DB protein-DNA binding distances.

To investigate hopping distances within a diffusion trajectory, Fig. 3(c) shows the distribution of the number of hops per diffusion trajectory longer than a finite hopping distance, ranging from  $0.25 \text{ \AA}$  to  $800 \text{ nm}$ , for  $R = 4.2$  and  $10.2 \text{ nm}$ . For the  $4.2\text{-nm}$  results, 3.37 hops in a diffusion trajectory were longer than  $5 \text{ nm}$ , and 11% of diffusion trajectories had a hop longer than  $200 \text{ nm}$ . As expected, the results for  $10.2 \text{ nm}$  are approximately twice as large since  $N$  is doubled. The crosses represent EcoRV proteins, which have a comparable hydrodynamic radius of  $2.66 \text{ nm}$  (see Table I), that were experimentally observed in different buffers to have hopped longer than  $200 \text{ nm}$  with reported occurrences ranging from 6 to 16% per diffusion trajectory [15]. These observations are in agreement with our simulations results. Furthermore, for hops longer than  $300$  and  $500 \text{ nm}$ , our observations agree with the reported values in Fig. 4(a) of Ref. [15].

Other DB proteins may differ from GFP-LacI in their sizes and thus  $\delta$  and  $R$ . Table I lists DB proteins that can hop on DNA (instead of proteins that slide only [11]) studied using SM fluorescence tracking methods on elongated DNA. Despite the difference in  $R$  by up to  $1.26 \text{ nm}$ , the  $\delta$  values differ only by less than  $0.07 \text{ \AA}$ . The effect of  $R$  difference is considered in Fig. 5(a), in which the number of hops per diffusion trajectory longer than a finite distance, ranging from  $0.1 \text{ \AA}$  to  $800 \text{ nm}$  for  $\delta = 0.267 \text{ \AA}$  and  $R$  from  $4.2$  to  $10.2 \text{ nm}$  are shown. The number of hops per diffusion trajectory increases with  $R$  moderately for all hopping distances, indicating that our hopping results are applicable to most observed DB proteins.

The step size  $\delta$  in the current approach, based on microscopic Brownian random walk models, can be made larger or smaller for vastly different particle sizes. Figure 5(b) shows distributions of hopping distances for three  $\delta$  values:  $0.267$ ,  $3.4$ , and  $10 \text{ \AA}$  (we used  $R = 4.2 \text{ nm}$  and  $t_{3,\text{max}} \approx 1 \text{ ms}$ ). The distribution curves collapse when protein hopping distances are larger than  $\delta$ , indicating that the tail distribution of protein hopping probability has the same asymptotic form at long distances, in agreement with the solution to the diffusion equation [30]. However, the mean hopping distance [Fig. 5(b, inset); values  $3.37$ ,  $36$ , and  $95 \text{ \AA}$ ], the mean number of hops  $N$  in a trajectory ( $526$ ,  $42$ , and  $14$ ), and  $\langle t_3 \rangle$  ( $0.22$ ,  $3.1$ , and  $9.2 \mu\text{s}$ ) all depend on  $\delta$  sensitively, as short-length scale motions dominate protein-DNA reassociation [Fig. 3(a)]. This regime can not be accessed in the macroscopic theory, i.e., by solving the diffusion equation directly.

When the protein-nonspecific-DNA association probability,  $p$ , is not 100%, e.g., due to rotation of the DNA-binding domain during large hops, hopping statistics and the

subsequent sliding statistics will change. For a low binding probability of  $p = 10\%$ , although on average, 10 consecutive hops would be needed for reassociation, the mean number of association attempts will still be  $N$ . However, the effective mean hopping time  $\langle t'_3 \rangle$ , and the mean hopping distance are expected to increase while the effective number of hops per diffusion trajectory  $N'$  decreases since  $t$  is held constant. The effective total hopping time  $N' \langle t'_3 \rangle$  and the rms total hopping distance per diffusion trajectory should therefore remain constant. The binding probability is thus inversely related to the effective mean sliding time  $\langle t'_1 \rangle$ , according to Eq. (2), which for  $p = 10\%$  results in a 10-fold increase in  $\langle t'_1 \rangle$ .

When salt concentration varies,  $p$  and  $R$  will change, as will  $D_3$  within a few angstroms of the DNA surface. However, since  $t$  remains  $\approx N' \langle t'_1 \rangle$  because  $N' \langle t'_3 \rangle \ll N' \langle t'_1 \rangle$ , the observed changes in  $t$  with salt concentration are likely due to changes in the total sliding time rather than the total hopping time. Consequently, changes in  $t$  as a result of varying salt concentration are not indicative of hopping and should not be used to determine its presence in diffusion trajectories, in disagreement with Refs. [6,10,16,17,21].

Some studies use flow to elongate DNA and/or investigate hopping properties of DB proteins [10,17,20,21,31]. Here we describe the effect of flow on hopping distances using the maximum reported flow rate in SM studies of  $100 \mu\text{m/s}$ . For our mean hopping time of  $\langle t_3 \rangle = 0.22 \mu\text{s}$ , a typical dissociated protein is carried by flow a length  $0.22 \text{ \AA}$  along DNA; this distance is negligible compared to its mean hopping distance of  $3.37 \text{ \AA}$  (the total displacement of the protein from flow alone within a diffusion trajectory consisting of 526 hops will be  $11.6 \text{ nm}$  which is substantially less than the total hopping displacement of  $127.5 \text{ nm}$  observed for GFP-LacI and similarly other proteins, as shown above). On the other hand, for a trajectory that includes a  $1\text{-}\mu\text{m}$ -long hop, which occurs once every 1000 diffusion trajectories, the hopping time is  $6.22 \text{ ms}$  and a protein is flown  $622 \text{ nm}$  along DNA. This distance would be sufficiently large for the protein to be considered dissociated.

Our results suggest that for diffusion trajectories without large hops of longer than of order a few hundred nanometers, a protein is unlikely to have been “washed out” while those that include large hops, the protein may be. However, according to Fig. 3(c), the probability for such an event to occur is approximately 1% of all diffusion trajectories.

Furthermore, sliding kinetics are not expected to be drastically affected by DNA configuration since a protein remains in contact with nonspecific DNA and should not be subject to DNA condensation and coiling either *in vivo* or *in vitro*, contrary to hopping kinetics. The reported values for  $D_1$  and  $t$  can therefore be applied under *in vivo* situations for better estimation of target binding rates.

## IV. CONCLUSION

In summary, this study analyzes DB proteins' hopping on elongated DNA to address sliding kinetics. While we have made several assumptions regarding the nature of protein association and modeling DNA, our study suggests that the observed sliding kinetics is a robust feature. Although hopping kinetics will change according to *in vivo* conditions, the lower bound on  $D$  for a typical DB protein should help future experiments in identifying the presence of hopping in protein diffusion trajectories with greater certainty.

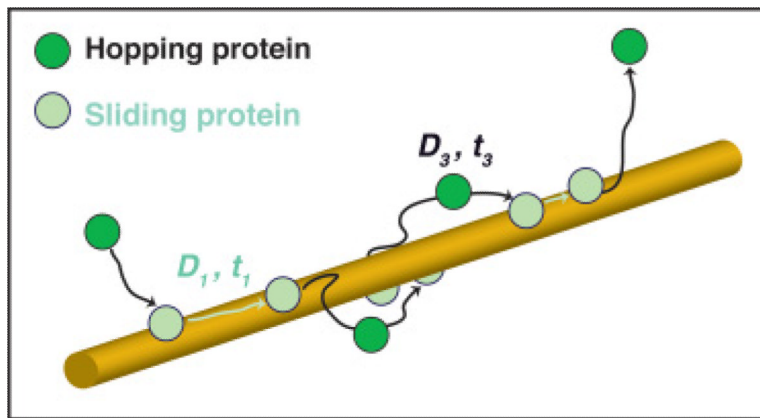
## Acknowledgments

We are grateful to Anders Carlsson for helpful discussions. M.C.D. wishes to thank the National Institutes of Health for Grant No. 5T90 DA022871.

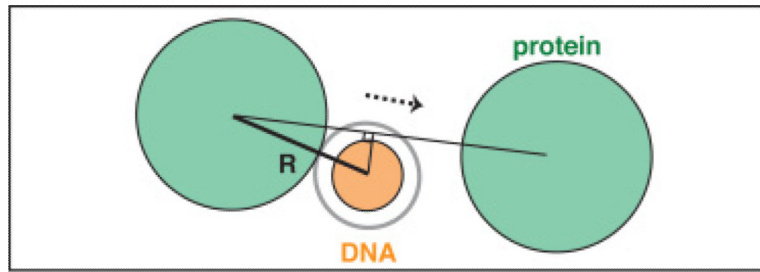
## References

1. Riggs AD, Bougeois S, Cohn M. J. Mol. Biol. 1970; 53:401. [PubMed: 4924006]
2. Halford SE, Marko JF. Nucl. Acids Res. 2004; 32:3040. [PubMed: 15178741]
3. Wang YM, Austin RH, Cox EC. Phys. Rev. Lett. 2006; 97:048302. [PubMed: 16907618]
4. Klenin KV, Merlitz H, Langowski J, Wu CX. Phys. Rev. Lett. 2006; 96:018104. [PubMed: 16486524]
5. Hu L, Grosberg AY, Bruinsma R. Biophys. J. 2008; 95:1151. [PubMed: 18456820]
6. Wunderlich Z, Mirny LA. Nucl. Acids Res. 2008; 36:3570. [PubMed: 18453629]
7. Murugan R. Biophys. J. 2010; 99:353. [PubMed: 20643052]
8. de la Rosa MAD, et al. Biophys. J. 2010; 98:2943. [PubMed: 20550907]
9. Harada Y, et al. Biophys. J. 1999; 76:709. [PubMed: 9929475]
10. Blainey PC, et al. Proc. Natl. Acad. Sci. USA. 2006; 103:5752. [PubMed: 16585517]
11. Granéli A, et al. Proc. Natl. Acad. Sci. USA. 2006; 103:1221. [PubMed: 16432240]
12. Elf J, Li G-W, Xie XS. Science. 2007; 316:1191. [PubMed: 17525339]
13. Kim JH, Larson RG. Nucl. Acids Res. 2007; 35:3848. [PubMed: 17526520]
14. Gorman J, et al. Mol. Cell. 2007; 28:359. [PubMed: 17996701]
15. Bonnet I, et al. Nucl. Acids Res. 2008; 36:4118. [PubMed: 18544605]
16. Komazin-Meredith G, et al. Proc. Natl. Acad. Sci. USA. 2008; 105:10721. [PubMed: 18658237]
17. Tafvizi A, et al. Biophys. J. 2008; 95:L01. [PubMed: 18424488]
18. Porecha RH, Stivers JT. Proc. Natl. Acad. Sci. USA. 2008; 105:10791. [PubMed: 18669665]
19. Biebricher A, et al. Biophys. J. 2009; 96:L50. [PubMed: 19383444]
20. Lin Y, et al. Biophys. J. 2009; 96:1911. [PubMed: 19254550]
21. Etson CM, et al. Proc. Natl. Acad. Sci. USA. 2010; 107:1900. [PubMed: 20080681]
22. Li G-W, Berg OG, Elf J. Nat. Phys. 2009; 5:294.
23. Revzin, A. The Biology of Nonspecific DNA Protein Interactions. CRC Press; London: 1990.
24. Gowers DM, Wilson GG, Halford SE. Proc. Natl. Acad. Sci. USA. 2005; 102:15883. [PubMed: 16243975]
25. The DB protein's displacement on DNA,  $x$ , contains two alternating diffusion displacements: 1D sliding displacement  $x_1$ , and 3D hopping displacement  

$$x = \sum_{i=1}^N x_{1i} + \sum_{j=1}^N x_{3j}, \langle x^2 \rangle = \sum_{i=1}^N \langle x_{1i}^2 \rangle + \sum_{j=1}^N \langle x_{3j}^2 \rangle + \sum_{i,j=1}^N 2 \langle x_{1i} \cdot x_{3j} \rangle = N \langle x_1^2 \rangle + N \langle x_3^2 \rangle = 2D_1 N \langle t_1 \rangle + 2D_3 N \langle t_3 \rangle$$
. This relation has also been verified by simulations.
26. Florescu A-M, Joyeux M. J. Chem. Phys. 2009; 130:015103. [PubMed: 19140636]
27. Dahirel V, Paillusson F, Jardat M, Barbi M, Victor JM. Phys. Rev. Lett. 2009; 102:228101. [PubMed: 19658903]
28. Berg, HC. Random Walks in Biology. Princeton University Press; Princeton, NJ: 1993.
29. Ricchetti M, Metzger W, Heumann H. Proc. Natl. Acad. Sci. USA. 1988; 85:4610. [PubMed: 3290898]
30. Loverdo C, Benichou O, Voituriez R, Biebricher A, Bonnet I, Desbiolles P. Phys. Rev. Lett. 2009; 102:188101. [PubMed: 19518914]
31. Kim JH, et al. Nano. Res. Lett. 2007; 2:185.

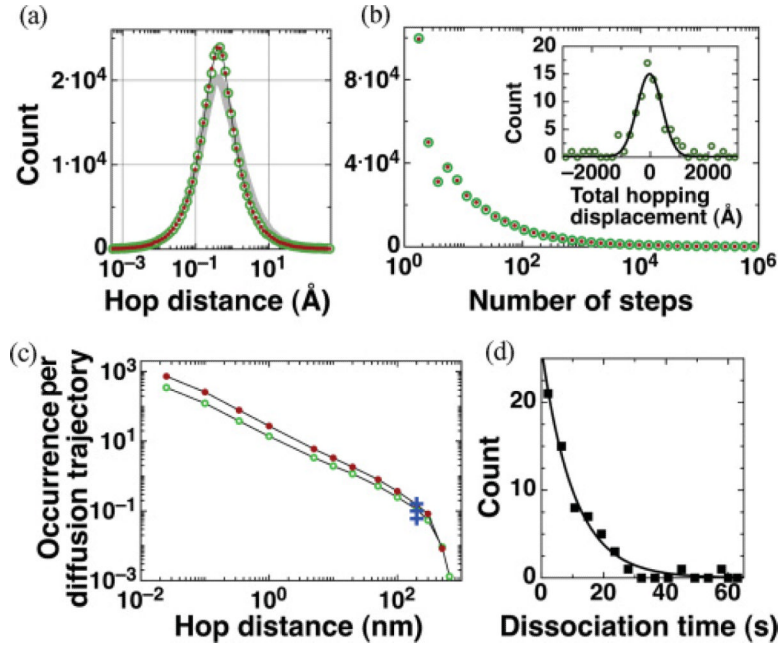


**FIG. 1.** (Color online) Schematics of a diffusion trajectory showing a protein initially binding to DNA, proceeding to slide (light disks) and hop (dark disks), and finally permanently dissociating from DNA. This example diffusion trajectory has two discernible hops.

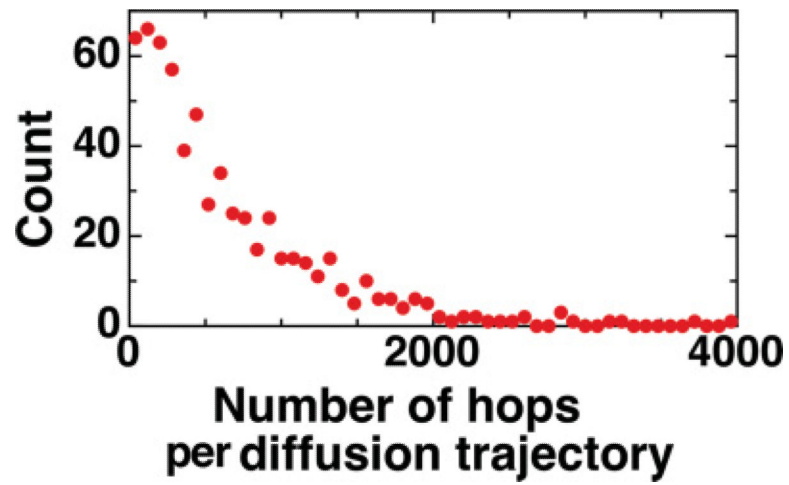


**FIG. 2.** (Color online) Determination of protein-DNA association. The gray (open) circle marks the effective protein-DNA binding distance. The protein moves ballistically between consecutive steps.

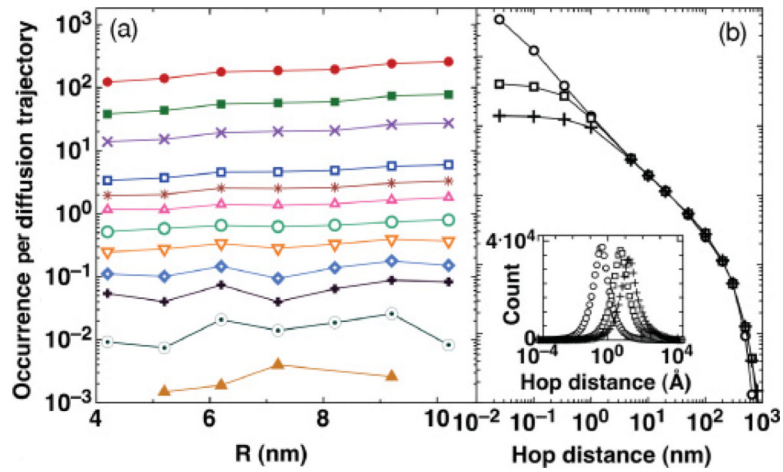


**FIG. 3.**

(Color online) (a) Distributions of hopping distances along DNA for  $\delta = 0.267 \text{ \AA}$  and  $R = 4.2$  (green, open circles) and  $10.2 \text{ nm}$  (red dots), and hopping height for  $R = 4.2 \text{ nm}$  (gray line). (b) Distributions for number of steps per hop for  $R = 4.2$  and  $10.2 \text{ nm}$ . (Inset) Distribution for total hopping displacement per diffusion trajectory and Gaussian fit (solid line). (c) Number of hops per diffusion trajectory longer than  $0.25 \text{ \AA}$ , and up to hops longer than  $800 \text{ nm}$ , for  $R = 4.2$  and  $10.2 \text{ nm}$ . The crosses are experimental data for EcoRV proteins, where the occurrence rate of hops per diffusion trajectory longer than  $200 \text{ nm}$  are  $0.06$ ,  $0.1$ , and  $0.16$  (the  $0.15$  value was omitted for clarity) [15]. (d) GFP-LacI total diffusion time  $t$  distribution (from experimental data in Ref. [3]). The mean of the exponential fit (solid line) is  $10.4 \text{ s}$ .



**FIG. 4.** (Color online) Distribution of number of hops per diffusion trajectory. The results of  $4 \times 10^5$  individual hopping simulations constitute a total of 763 protein diffusion trajectories such that 526 hops occur on average per trajectory.



**FIG. 5.** (Color online) Distributions for number of hops per diffusion trajectory longer than 0.1, 0.34, 1, 5, 10, 20, 50, 100, 200, 300, 500, and 800 nm (top to bottom in a), (a) for  $R$  ranging from 4.2 to 10.2 nm (left to right) and (b) for  $R = 4.2$  nm and  $\delta = 0.267$  (circles), 3.4 (empty squares), and 10.2 Å (crosses). (Inset) Hopping distance distributions for the three  $\delta$  values.

TABLE I

DB protein diffusion properties on elongated DNA.

Protein	$r_{\text{protein}}$ (nm)	$\delta$ (Å)	$D$ (nm <sup>2</sup> /s)
YFP-LacI, 2 <sup>a</sup>	3.13	0.284	$4.6 \times 10^4$ [12]
GFP-LacI	2.68	0.267	$2.3 \times 10^2$ – $1.3 \times 10^5$ [3]
EcoRV, 2	2.66	0.262	$0.9 - 2.5 \times 10^4$ [15]
EcoRV <sup>a</sup>			$3.1 \times 10^3$ [19]
RNAP, 4 <sup>a</sup>			$6.1 \times 10^3 - 4.3 \times 10^5$ [13]
RNAP <sup>a</sup>			$1.3 \times 10^5$ [29], $\sim 10^4$ [9]
hOgg1	2.36	0.247	$5.78 \times 10^5$ [10]
p53	2.34	0.246	$3.01 \times 10^5$ [17]
UL42	2.63	0.261	$5.1 \times 10^3$ – $2.2 \times 10^4$ [16]
T7 gp5, 2	2.86	0.272	$8.0 \times 10^5$ – $1.86 \times 10^6$ [21]
T7 gp5, 2	3.00	0.278	$4.0 \times 10^5$ [21]
C-Ada	1.77	0.214	$1.3 \times 10^6$ [20]

<sup>a</sup>The number 2 indicates a dimer, and 4 indicates a tetramer.

<sup>b</sup>Unknown molecular size due to unspecified/uncertain protein components and/or labels.