

# Method for the analysis of contribution of sliding and hopping to a facilitated diffusion of DNA-binding protein: Application to *in vivo* data

Marcin Tabaka,<sup>1</sup> Krzysztof Burdzy,<sup>2</sup> and Robert Hołyst<sup>1,\*</sup>

<sup>1</sup>*Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, 01-224 Warsaw, Poland*

<sup>2</sup>*Department of Mathematics, University of Washington, Box 354350, Seattle, Washington 98195, USA*

(Received 11 May 2015; published 28 August 2015)

DNA-binding protein searches for its target, a specific site on DNA, by means of diffusion. The search process consists of many recurrent steps of one-dimensional diffusion (sliding) along the DNA chain and three-dimensional diffusion (hopping) after dissociation of a protein from the DNA chain. Here we propose a computational method that allows extracting the contribution of sliding and hopping to the search process *in vivo* from the measurements of the kinetics of the target search by the *lac* repressor in *Escherichia coli* [P. Hammar *et al.*, *Science* **336**, 1595 (2012)]. The method combines lattice Monte Carlo simulations with the Brownian excursion theory and includes explicitly steric constraints for hopping due to the helical structure of DNA. The simulation results including all experimental data reveal that the *in vivo* target search is dominated by sliding. The short-range hopping to the same base pair interrupts one-dimensional sliding while long-range hopping does not contribute significantly to the kinetics of the search of the target *in vivo*.

DOI: [10.1103/PhysRevE.92.022721](https://doi.org/10.1103/PhysRevE.92.022721)

PACS number(s): 87.10.Rt, 87.10.Hk, 87.15.Vv

## I. INTRODUCTION

Sequence-specific DNA-binding proteins such as transcription factors (TFs) locate their targets rapidly in cells despite the presence of a large excess of nonspecific DNA [1,2]. A well-established kinetic mechanism, referred to as facilitated diffusion, states that TFs, when bound to DNA, sample many base pairs by one-dimensional diffusion (sliding) [3–16]. Surprisingly, despite over 30 years of theoretical studies, the first experimental proof that sliding occurs in living cells (*Escherichia coli*) was given by Hammar *et al.* [2]. They placed two operators, targets for the *lac* repressor, separated by different distances along DNA [between 25 and 200 base pairs (bp)]. They determined the rate of the target location by the *lac* repressor and compared the results to those obtained for cells having only one operator. The rate  $k$  for two operators was two times higher than the rate for the single operator,  $k_0 = k/2$ , only when the distance was larger than 65 bp. When the distance between two targets along DNA was reduced below 65 bp the rate decreased and approached  $k/2$  for a vanishing distance between them. The common name of this phenomenon is the correlated search process [7]. In principle, sliding [15,17–27] and hopping [28–33] of DNA-binding proteins contribute to the correlated search [7,34–36] [see Fig. 1(a)]. However, Hammar *et al.* [2] interpreted the result as arising solely from sliding over distances of 36 bp, without taking into account hopping in the search process. Other experiments *in vitro* showed that hopping was a non-negligible process in the kinetics of target location. For example, *in buffer* assay [37,38] for cellularlike conditions revealed that the search over 30 bp by a DNA-binding protein involves at least one three-dimensional (3D) hopping step [37]. The recent measurements of the binding kinetics of RNA polymerase to the promoter also show that the primary search mechanism is through 3D diffusion [39,40]. Moreover, theoretical studies [35] suggest that hopping enhances the correlated

search in compacted conformations of DNA. This difference between *in vivo* and *in vitro* results prompted us to ask the following question: What is the contribution of short-range and long-range hopping to the kinetics of target location in living cells?

The direct observation of hopping and sliding of fluorescently labeled DNA-binding proteins on DNA are beyond both spatial and temporal resolution of today's methods [13,37]. Instead, indirect methods using two operators on DNA are applied both *in vitro* [37,38,41] and *in vivo* [2] for describing quantitatively the correlated search process. However, none of these experimental methods can be applied to estimate the contribution of sliding and hopping to the process *in vivo*. This study describes a comprehensive theoretical method of lattice Monte Carlo simulations with the Brownian excursion theory [42]. The method is applied to *in vivo* kinetic binding data of the *lac* repressor [2]. The experiment [2] and molecular dynamics simulations [15] put strong constraints on our mesoscopic simulations. We fix the diffusion coefficients for the *lac* repressor with and without DNA-binding domains [1], include the helical structure of DNA, and set the ratio of the reaction rates  $k/k_0$  as a function of the distance between operators on DNA [2]. We also include the fact that the repressor binding is a diffusion-limited process [15] under conditions of correct mutual orientation of the DNA-protein pair.

## II. MODEL OF SLIDING AND HOPPING

We define sliding as the one-dimensional diffusion along the helical structure of DNA (Fig. 1). Hopping is the 3D diffusion. Hopping starts upon microscopic dissociation of the protein from the DNA strand of predefined length. The protein begins 3D diffusion from a distance  $r_{\text{start}}$  from the DNA axis and continues until the protein returns to the same DNA strand. Short-range and long-range hopping are also defined. The former process consists of dissociation from a base pair, a short excursion in three dimensions, and a return to this base pair or to the nearest-neighbor base pair. The long-range

\*Corresponding author: rholyst@ichf.edu.pl

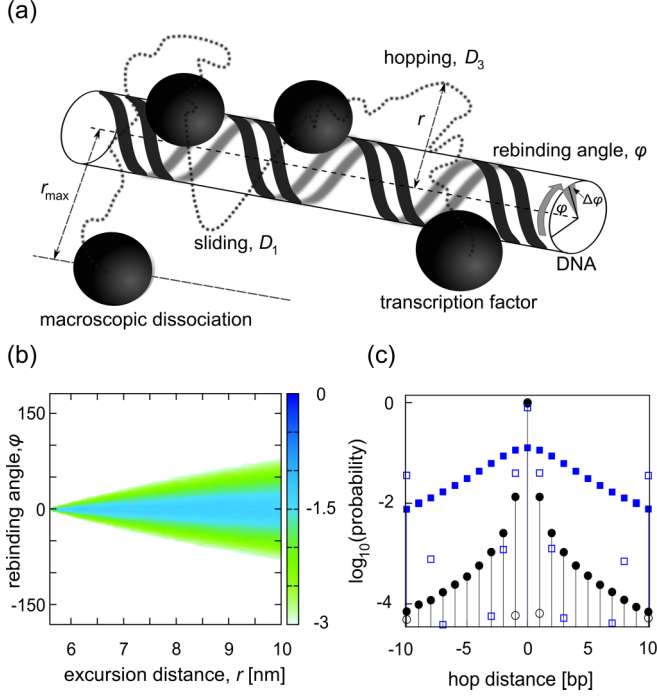


FIG. 1. (Color online) DNA-binding protein reassociation to the DNA chain depends on the helical structure of DNA. (a) The model of facilitated diffusion includes 3D translocations to nearby sites (hopping) interspersed with binding to nonspecific sites with diffusion along contiguous sites (sliding), i.e., along a helical path of DNA. The hopping leads to association only if the protein follows the DNA double helix (according to the angle  $\phi$  between rebinding events). When the protein reaches DNA we check this angle and allow the protein to associate only if the angle is within a given range of angles  $\Delta\phi$  (this range of angles is a parameter in the model). (b) Normalized histograms of the TF rebinding angles as a function of the excursion (hopping) distance  $r$ . The rebinding to the DNA is focused within a narrow range of angles. (c) Histograms of the hop distances, along the axis of the DNA ( $z$  direction), calculated for single excursions starting at  $r_{\text{start}} = 5.51$  nm (black circles) and  $r_{\text{start}} = 6.5$  nm (blue squares) and reaching the point of the DNA given by  $r_{\text{min}} = 5.5$  nm. Closed data points correspond to the case with the whole DNA reactive surface while open data points result from incorporating the helical structure of DNA into the model with the reactive angle range  $\Delta\phi = 36^\circ$ .

hopping includes a longer excursion in three dimensions and a return to any other base pair of the same DNA strand. Finally, macroscopic dissociation is a process consisting of the following steps: dissociation from the DNA and excursion to the predefined distance from the DNA strand ( $r_{\text{max}}$  or  $z_{\text{max}}$ , where  $z_{\text{max}}$  is the distance along the DNA, here taken as the persistence length of DNA) without any association with the DNA strand. By microscopic association we define a process of protein binding to DNA from a distance  $r_{\text{min}}$ .

We model the TF correlated search with the rules describing the probability of the following events during the process.

**Sliding.** A nonspecifically bound TF to DNA performs one of two exclusive steps [2,7]: diffusive transition in either direction to the nearest base pair or microscopic dissociation. The probability of the former is given by  $P = 2D_1/(k_d^{\text{micro}} + 2D_1) = 2s^2/(1 + 2s^2)$  and the latter by

$P_d = 1/(1 + 2s^2)$ , with  $s = \sqrt{D_1/k_d^{\text{micro}}}$ , where  $D_1$  is a one-dimensional diffusion coefficient of the TF and  $k_d^{\text{micro}}$  is a microscopic dissociation rate of the protein from the DNA chain. Microscopic dissociation releases the TF from the DNA to a distance  $r_{\text{start}}$  from which the TF is allowed to diffuse freely in three dimensions [7]. This distance is a parameter in the model. Typically, it is larger than the distance between the DNA axis and the center of mass of the TF [15].

**Hopping.** In the hopping event the TF after dissociation from a given strand of DNA diffuses away from this DNA fragment at a distance  $r$  and then returns to the neighborhood of the same DNA strand or to a different strand uncorrelated with the previous one. Mathematically, we model this event by the probability  $P(r)$  that an object departing at a distance  $r_{\text{start}}$  from the cylinder will return to the cylinder of radius  $r_{\text{min}}$  (we assume that  $r_{\text{min}} = 5.5$  nm for LacI) after making an excursion at a distance equal to or greater than  $r$  [42],

$$P(r) = [\ln(r_{\text{start}}) - \ln(r_{\text{min}})]/[\ln(r) - \ln(r_{\text{min}})]. \quad (1)$$

A long-range excursion from a given fragment of DNA can lead to reassociation of the protein to elsewhere in the nucleoid and loss of the spatial correlation with the chain it has just dissociated [7,43,44] [macroscopic dissociation; see Fig. 1(a)]. The probability given by Eq. (1) allows us to differentiate trajectories that lead to macroscopic dissociations ( $r \geq r_{\text{max}}$ ) from hopping ( $r < r_{\text{max}}$ ). The exact values of  $r_{\text{max}}$  for the *E. coli* nucleoid and LacI repressor are unknown. Berg *et al.* [7] defined TF macroscopic dissociation from a DNA segment as an event in which the TF reaches the average distance between the closest uncorrelated neighboring DNA segments. Following this approach, we get  $r_{\text{max}} = 11$  nm for the predicted nucleoid volume [16]. Nevertheless, we examine different values of  $r_{\text{max}}$  in the fitting procedure assuming  $r_{\text{max}} = 11$  nm as being the minimal distance for macroscopic dissociation [12].

The hopping requires explicit inclusion of the helical nature of DNA into the model of facilitated diffusion, because not every encounter of the TF with DNA leads to reassociation. Only when the TF reaches the surface of DNA close to the main grooves can it associate with the DNA. The main grooves follow the helix along the surface of the DNA and the helical structure of the DNA imposes a certain constraint. For example, protein reassociation by hopping, with a covered distance of  $z = 5$  bp along a DNA chain from the last dissociation site, needs to involve  $\phi = 180^\circ$  of protein excursion around the DNA axis [Fig. 1(a)]. The TF excursion at the distance  $r$  is combined with a translocation along the DNA axis and is characterized by the hopping variables ( $z, \phi, t$ ) describing the distance along the DNA axis, the rebinding angle between consecutive dissociation and association events, and the excursion time, respectively. The analytical distribution of rebinding variables as a function of  $r$  is unknown. Therefore, we perform random walk simulations on the square lattice with mesh size  $r_{\text{step}}$ , in which we count the number of steps  $n$  up to  $r_{\text{min}}$  by the TF from  $r_{\text{start}} = r_{\text{min}} + r_{\text{step}}$ . For every trajectory that reaches  $(r, r + r_{\text{step}})$  we calculated a histogram of steps  $n$  and angles  $\phi$ . The overall excursion time is  $t = n\tau$ , where  $\tau = r_{\text{step}}^2/6D_3$ . The number of steps  $n$  performed in our 2D simulations is 2/3 of all steps in three

dimensions. Therefore, we calculate the distance  $z$  along the DNA axis during each TF excursion according to the Gaussian distribution

$$\begin{aligned} p(z,t) &= \frac{1}{\sqrt{4\pi D_3 t}} \exp\left(-\frac{z^2}{4D_3 t}\right) \\ &= \frac{1}{\sqrt{\pi n r_{\text{step}}^2}} \exp\left(-\frac{z^2}{n r_{\text{step}}^2}\right). \end{aligned} \quad (2)$$

In Fig. 1(b) we show the distributions of rebinding angles  $\phi$  for different excursion distances  $r$ . Three-dimensional diffusion causes reassociation of the TF within a narrow range of angles. Thus, taking into consideration the helical structure of DNA, a single TF translocation by hopping to distinct base pairs is of low probability [Fig. 1(c)], orders of magnitude smaller than rebinding to the same base pair. Therefore, the rebinding is highly peaked at the base pair TF dissociated from, independently of the initial distance  $r_{\text{start}}$ . In our model, the rebinding takes place if  $(z, \phi)$  follows the helical pathway. The rebinding to the consecutive base pair requires  $\phi = \pi/5$  (i.e., 5 bp for  $180^\circ$ ) translocation within the range of reactive angles  $\Delta\phi$ , the parameter in our model.

### III. RESULTS

The simulation algorithm (Fig. 2) counts the number  $M$  of macroscopic association and dissociation events [7], i.e., excursions to  $r \geq r_{\text{max}}$  or excursions along DNA to distances  $|z| \geq z_{\text{max}}$  larger than the persistence length of the DNA [16,31]. The persistence length of the DNA for *in vivo* conditions was calculated to be 23 nm [16] and we use this value herein. Every macroscopic event consists of a recurrent number of sliding and hopping steps that constitute a correlated sampling of DNA sites. Hammar *et al.* [2] studied the *lac* repressor binding rates to two specific binding sites positioned at center-to-center distances of 25, 45, 65, 115, and 203 bp [Fig. 3(d)]. Measured binding rates demonstrated correlations for binding and the results were interpreted as arising from long sliding distances  $s_l = \sqrt{2}s \sim 36$  bp.

In the experiment, the observable was the binding rate  $k$  to one of the operators in the case of the system with double operators and the rate of a single operator system  $k_0$ . The time required for specific association to one of two operators differs only in the number of macroscopic association tries  $M$  (Fig. 2) as compared to the system with a single operator  $M_0$ . We have one binding event to the operator per  $M$  macroscopic tries, thus  $k \sim 1/M(t)$ , where  $\langle t \rangle$  is the average time of the macroscopic dissociation and reassociation event (including hopping and sliding). Thus the ratio of the association rate to one of two operators to the association rate to the DNA with a single operator is given by the exact formula  $k/k_0 = M_0/M$ . The calculated ratio of the macroscopic tries constitutes a fitting curve to experimental data  $k/k_0$  with only one fitting parameter  $s^2$  [Fig. 3(d)].

The measured *in vivo* diffusion constants [1] for the *lac* repressor with ( $D_3^{\text{eff}} = 0.4 \mu\text{m}^2 \text{s}^{-1}$ ) and without ( $D_3 = 3 \mu\text{m}^2 \text{s}^{-1}$ ) DNA-binding domains impose constraints on the parameter values in the fitting procedure. The effective diffusion constant  $D_3^{\text{eff}}$  is lowered by interactions with nonspecific

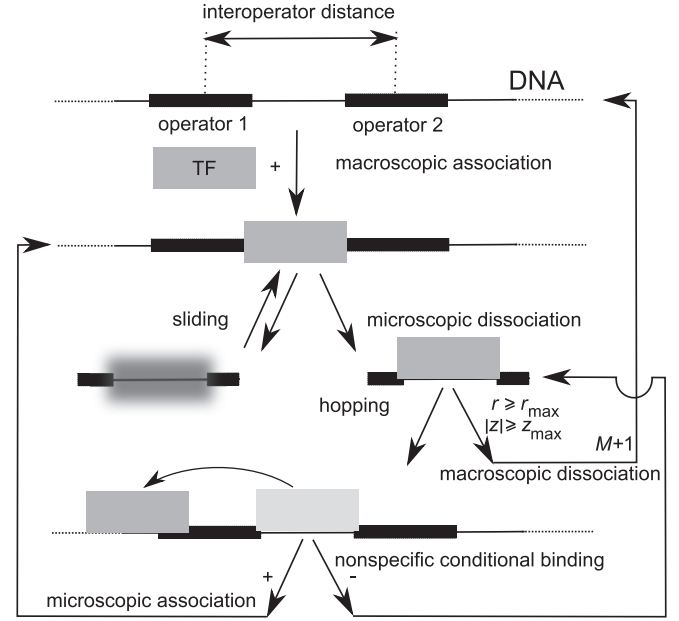


FIG. 2. Outline of the simulation algorithm used in the interpretation of experiments performed for TF binding rates to two specific sites placed at a fixed distance on the DNA. The method requires as a prerequisite the calculation of the histograms of diffusion steps  $n$  and rebinding angles  $\phi$  for the parameters  $r_{\text{start}}$  and  $r_{\text{max}}$  (the maximal distance of excursion after dissociation varied from 11 nm to 40 nm). The event-driven algorithm consists of the following steps: initialize, with  $s^2$ ,  $N$  (number of base pairs), and  $M = 0$ ; macroscopic association, in which a random number  $\xi \in (1, N)$  is generated and the current position of the DNA binding protein  $\mathcal{P}$  is  $\mathcal{P} = \xi$ ; sliding, in which a random number  $x$  is drawn from the uniform distribution in the unit interval and according to  $x$  one of three events is chosen: (i) the position is changed to  $\mathcal{P} + 1$  [with probability  $P_+ = 1/2P = s^2/(1 + 2s^2)$ ], (ii) the position is changed to  $\mathcal{P} - 1$  [with probability  $P_- = s^2/(1 + 2s^2)$ ], or (iii) microscopic dissociation [with probability  $P_d = 1/(1 + 2s^2)$ ] and hopping, in which the excursion distance  $r$  is generated from the distribution  $dP(r)/dr = [\ln(r_{\text{start}}/r_{\text{min}})]/\{r[\ln(r/r_{\text{min}})]^2\}$  by drawing a random number  $x$  from the uniform distribution in the unit interval and taking  $r = r_{\text{start}}^{1/x} r_{\text{min}}^{(1-x)/x}$  is chosen. If  $r \geq r_{\text{max}}$ , then macroscopic dissociation will occur,  $M = M + 1$ . If  $r < r_{\text{max}}$ , the hopping distance  $z$  is calculated along the DNA axis and the rebinding angle  $\phi$  is drawn from the histograms (see the text for details). If the pair  $(z, \phi)$  follows a helical pathway of the DNA a random number  $x$  is drawn from the uniform distribution in the unit interval. If  $x \leq p$  ( $p$  is the probability of microscopic association) the TF binds with the DNA; otherwise hopping is repeated. The final step is specific binding: The procedure is finished when  $\mathcal{P}$  corresponds to the position of the operator or operators.

DNA and is given by

$$D_3^{\text{eff}} = \frac{(1/3)D_1\langle t_1 \rangle + D_3\langle t_3 \rangle}{\langle t_1 \rangle + \langle t_3 \rangle}, \quad (3)$$

where  $\langle t_1 \rangle = 1/k_d^{\text{micro}}$  is the mean time spent by the TF in the nonspecific complex with DNA and  $\langle t_3 \rangle$  describes the mean reassociation time of the TF with DNA during 3D excursions. The value of the 1D diffusion constant for the *lac* repressor is assumed to be equal to [16]  $D_1 = 0.025 \mu\text{m}^2 \text{s}^{-1}$ . Thus,

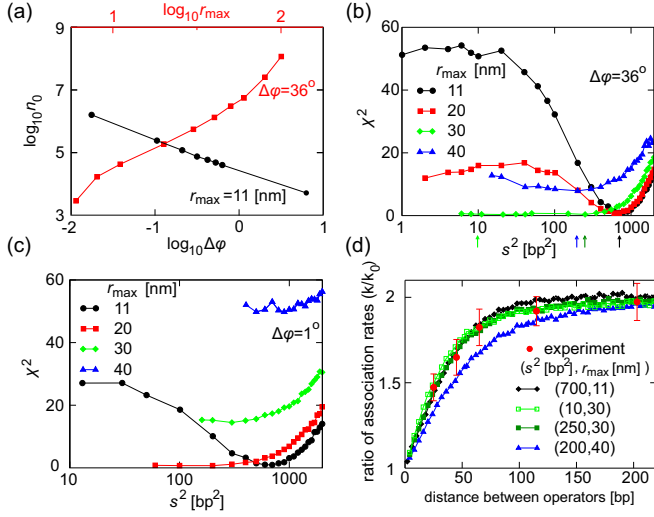


FIG. 3. (Color online) Application of the simulation algorithm to the *in vivo* kinetics data of *lac* repressor binding in *E. coli* [2]. (a)–(d) Results of the simulations for different sets of parameter value pairs  $(\Delta\phi, r_{\max})$  for  $r_{\min} = 5.5$  and  $r_{\text{start}} = 5.51$  nm. (a) Changes in the number of steps  $n_0$  between rebinding events as a function of the range of reactive angle (constant  $r_{\max}$ , black circles) and as a function of the macroscopic dissociation distance (constant  $\Delta\phi$ , red squares). (b) and (c) Residual error ( $\chi^2$ ) of the model predictions and experimental data for *lac* repressor for the range of angles (b)  $\Delta\phi = 36^\circ$  and (c)  $\Delta\phi = 1^\circ$ . The left data points correspond to the diffusion-limited reaction. (d) Comparison of experimental [2] and simulated ratios of association rates for  $\Delta\phi = 36^\circ$  and pairs of parameters  $(s^2, r_{\max})$  marked by arrows in (b).

changing the value of  $s^2$  in the fitting procedure requires a simultaneous change of the value of  $\langle t_3 \rangle$  to keep the ratio given by Eq. (3) constant. Here  $\langle t_3 \rangle$  depends on  $\Delta\phi$  and  $r_{\max}$ . However, the exact value of the parameter  $\Delta\phi$  is unknown for repressor binding to DNA. We test therefore different values of the parameter  $\Delta\phi$ . Similarly, we vary  $r_{\max}$  between 11 and 40 nm. In order to get the correct value of  $\langle t_3 \rangle$  for a given pair of parameters  $(\Delta\phi, r_{\max})$ , which follows constraints for diffusion constants, we have to introduce the probability of microscopic association for the TF within the reactive region  $\Delta\phi$ . We do so in the following way. We run additional random walk simulations on the cubic lattice with the same parameters as during the fitting procedure. The aim of these simulations is to get the average number of diffusive steps of length  $r_{\text{step}}$  that the TF carries out during 3D diffusion. If the TF dissociates macroscopically, we randomize its position at the circumference of radius  $r_{\max}$ . Outside the range  $\Delta\phi$ , reflective boundary conditions are applied. From these simulations we get the number of diffusive steps  $n_0$  for  $p = 1$ , i.e., for the condition in which every reaching of the distance  $r_{\min}$  by TF leads to the microscopic association within the reactive angle range  $\Delta\phi$ . Then the total number of diffusive steps for a given value of the probability of microscopic association  $p$  is given by the scaling  $n_t = n_0/p$ . Figure 3(a) presents values of  $n_0$  obtained from the simulation data as a function of the parameters  $\Delta\phi$  and  $r_{\max}$ . Hence, for a given value  $s^2 = D_1 \langle t_1 \rangle$  we can find the probability of the microscopic association, for which the average time the protein spends

during 3D diffusion ( $t_3$ ) holds the relation between the values of the diffusion constants for the *lac* repressor, as given by Eq. (3). The algorithm described in Fig. 2 is general and can be applied to various DNA-binding proteins. However, only for the *lac* repressor do we have knowledge about its diffusion constants [1] and its binding kinetics to two operators for different distances between them [2].

The results of the simulations are used to fit the kinetic binding data [2] [Figs. 3(b)–3(d)]. The total number of base pairs in the simulations was  $N = 2 \times 10^4$  bp and  $10^5$  simulation realizations were performed to calculate the mean number of macroscopic tries that lead to the specific binding. The residual error  $\chi^2$  of the experimental and simulation data is applied as a measure of the goodness of fit for two values of the parameter  $\Delta\phi = 36^\circ$  and  $1^\circ$ . In both cases a low value of macroscopic dissociation distance  $r_{\max} = 11$  nm gives the best fit for  $s^2 = 700$  bp<sup>2</sup> that corresponds to a sliding length of 37 bp being equal to the one determined in Ref. [2]. The correlated search in this case is due to the sliding where almost every microscopic dissociation leads to macroscopic dissociation. The parameter values  $r_{\max} = 30$  nm,  $\Delta\phi = 36^\circ$  and  $r_{\max} = 20$  nm,  $\Delta\phi = 1^\circ$  equally well explain the experimental results. The striking feature in the first case is a broad, two order of magnitude, range of parameter values  $s^2$  that give good fits to experimental data [Fig. 3(b)]. The sliding events now are interspersed by hopping. The combination of the results shown in Figs. 1(c) and 3 suggests that hopping is mainly to the same base pair TF dissociated from, not contributing to the correlated search of the TF.

Despite the fact that the values of pairs  $(\Delta\phi, r_{\max})$  are unknown *in vivo*, we can envisage plausible mechanisms. Marklund *et al.* [15], using molecular dynamics simulations, have shown that the *lac* repressor during microscopic dissociation does not encounter an energy barrier, noting that the binding of the repressor is diffusion limited within a range  $\Delta\phi$ . Taking this result as the next constraint, we can discard cases that describe well the experimental data but are reaction limited, e.g., for  $\Delta\phi = 36^\circ$  and  $r_{\max} = 11$  nm [Fig. 3(b), black circles]. Hence, two cases,  $\Delta\phi = 36^\circ, r_{\max} = 30$  nm [Fig. 3(b), green diamonds] and  $\Delta\phi = 1^\circ, r_{\max} = 20$  nm [Fig. 3(c), red squares] fulfill all requirements, suggesting that the *in vivo* correlated search consists of short sliding events interspersed with hopping and a macroscopic dissociation distance much larger than 11 nm. During the same simulations that we carry out for generation of data points in Figs. 3(b) and 3(c) we collect information about the number of hopping events per macroscopic dissociation as well as the probability of the hopping that leads to a base pair different from the one TF dissociated from. Thus, we get for  $\Delta\phi = 36^\circ, r_{\max} = 30$  nm that the probability of excursion to a different base pair is 0.01 and the mean number of 140 TF microscopic association events per macroscopic dissociation. Hence, the average number of hopping events contributing to the TF correlated search is 1.4. For the parameter  $\Delta\phi = 1^\circ, r_{\max} = 20$  nm the average number of hops to a different base pair drops to 0.1 per macroscopic dissociation.

Summarizing our results, although the two sets of parameters studied here are consistent with experiments *in vivo*, in none of these cases does long-range hopping of the *lac* repressor contribute to the correlated search along DNA. The

analysis shows that the method explains experimental data when one of the parameters is fixed and the decrease of the range or reactive angle  $\Delta\phi$  is associated with the decrease of the distance of macroscopic dissociation  $r_{\max}$ . The case with an extremely small reactive angle ( $\Delta\phi = 1^\circ$ ) is rather improbable in view of the free rotation of the transcription factor and the large size of major grooves of DNA. Therefore, our simulations give for the set of parameters, consistent with aforementioned experimental constraints  $\Delta\phi = 36^\circ$ ,  $r_{\max} = 30$  nm, and  $s^2 = 6$  bp<sup>2</sup>, the sliding distance of  $s_l = \sqrt{2}s \sim 3$  bp. Despite the fact that the *lac* repressor often dissociates from DNA (140 times before macroscopic dissociation), it does return to the same base pair and therefore effectively stays in close proximity to DNA during many short hopping events. Only one, on average, of all these microscopic reassociations is to a different (usually the closest neighbor) base pair. The effective sliding distance (before macroscopic dissociation) is  $\sqrt{140} \times 3$  bp  $\sim 36$  bp, as in the model of pure sliding by Hammar *et al.* [2]. However, many dissociation events mean that the *lac* repressor does not tightly bind to the nonspecific sites on DNA. The macroscopic dissociation occurs when the repressor diffuses a distance  $r_{\max} = 30$  nm from a DNA strand. The constraints imposed in the simulations are as follows: the diffusion coefficient in three dimensions is  $D_3 = 3 \mu\text{m}^2 \text{s}^{-1}$ , the 1D diffusion coefficient is  $D_1 = 0.025 \mu\text{m}^2 \text{s}^{-1}$ , the effective diffusion constant *in vivo* is  $D_3^{\text{eff}} = 0.4 \mu\text{m}^2 \text{s}^{-1}$ , and  $z_{\max} = 23$  nm. Also, we assume, following the microscopic simulations of Marklund *et al.* [15], that binding of the *lac* repressor to DNA is the diffusion limited reaction without additional barriers, but within the reactive angle  $\Delta\phi = 36^\circ$ .

#### IV. CONCLUSION

We have presented mesoscopic simulations of the LacI repressor searching for an operator on DNA in *E. coli*. The parameters for our simulation algorithm are strongly constrained by the experimental data. The different element in the simulation algorithm is the explicit representation of hopping in the framework of the Brownian excursion theory [42]. Our model also explains two seemingly contradictory statements: that DNA-binding protein performs only long slidings [2] and that in the correlated search with asymmetric dimers for intracellular-like conditions there is at least one hopping step [37]. Densely packed DNA in the bacterial nucleoid imposes that 3D diffusion contributes to the macroscopic events but not to the correlated search by the *lac* repressor in *Escherichia coli*. This result is in contradiction of models showing that 3D excursions contribute to the correlated search [35]. The difference is due to the fact that our model is more detailed and encompasses all available experimental data. From our model it is clear that increasing the macroscopic dissociation distance results in deterioration of the fit to experimental data [Fig. 3(d)]. Long excursions observed *in vitro*, much beyond macroscopic dissociation distances considered in this work, result from an elongated (and not compactly packed) DNA chain [30,31].

The regulation of the expression of the *lac* operon by the LacI repressor serves as a paradigm of gene regulation in prokaryotes. The kinetic modeling of *in vivo* facilitated diffusion of the LacI repressor [2,16,45,46] shows that the

repressor search is close to the optimal conditions for target location. Additionally, the facilitated diffusion contributes to the reduction of noise in gene expression [47] and hence to more precise regulation of gene expression. Thus, the detailed models of facilitated diffusion are crucial to understand the mechanisms of optimization of the target search by transcription factors. The algorithm presented here can support the existing computational models algorithms of facilitated diffusion [48,49]. It can be also extended to include the probability of DNA sequence recognition by a transcription factor [2,17,27,35,49–57], kinetic transitions between different states of a transcription factor [20,58], or the effect of macromolecular crowding on the target search due to the presence of other DNA-binding proteins [16,59–64].

Here we have neglected the effect of interfering of the *lac* repressor motion by other DNA-binding proteins [59,60] and by a densely packed nucleoid. This assumption is justified by the fact that for the fast growth conditions considered here and applied in *in vivo* experiments, the calculated fraction of DNA that is free from DNA-binding proteins is 0.85 [16]. Taking into account this fraction and the model of random positioning of DNA-binding proteins on DNA, we obtained that DNA-binding proteins do not affect significantly the sliding of the *lac* repressor [16]. Also, a lack of specific binding sites for RNA polymerase, cAMP receptor protein, and histonelike nucleoid-structuring protein between *lac* operators does not change the rate of *lac* repressor binding as compared to the rate observed for native interoperator sequences with these binding sites [2]. DNA is less densely packed for fast growth conditions and for example a distribution of the *lac* repressor does not depend on the spatial location of its encoding gene [65]. The repression strength for the *lac* repressor as a function of intergenic distance is also reproduced by a model in which spatial homogeneity of the repressor is assumed [65]. Additionally, the values of mean-square displacement of the LacI dimer [1] show that diffusion is not confined for the repressor. The aforementioned phenomena need to be incorporated into the facilitated diffusion model for slow growth conditions [65].

The model proposed in the present study considers simplified interactions between DNA-binding protein and the helical DNA chain as it includes only orientation constraints for protein-DNA association. Thus it should be considered as a reference to more detailed models of protein hopping and sliding on nonspecific DNA. The combination of molecular dynamics simulations [15,66,67] and detailed Brownian dynamics simulations will give insight into the impact of electrostatic steering [68] and other interactions on the target search at the TF-DNA distances between sliding and unconstrained 3D diffusion. Nevertheless, this model predicts that for all ranges of DNA reactive angles and even for perfect steering approximated by a case in which the whole surface of the DNA transcription factor is reactive, hopping does not contribute to the correlated target search by the *lac* repressor.

The method is applicable also to studies of a TF binding-unbinding kinetics to regulatory regions with multiple TF binding sites. Preliminary studies of the facilitated diffusion of transcription factors have started recently to appear also for eukaryotic cells [69,70]. The analysis presented here

establishes the experimental parameters that are required to gauge the contribution of one- and three-dimensional diffusion to the correlated search process and thus can drive future experiments and support analysis of measured TF binding kinetics.

#### ACKNOWLEDGMENT

Discussions with Johan Elf are greatly acknowledged. The research was supported by the NCN Grant No. DEC-2011/02/A/ST3/00143. K.B. was supported by NSF Grant No. DMS-1206276.

- 
- [1] J. Elf, G. W. Li, and X. S. Xie, *Science* **316**, 1191 (2007).
- [2] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, *Science* **336**, 1595 (2012).
- [3] P. H. Richter and M. Eigen, *Biophys. Chem.* **2**, 255 (1974).
- [4] O. G. Berg and C. Blomberg, *Biophys. Chem.* **4**, 367 (1976).
- [5] O. G. Berg and C. Blomberg, *Biophys. Chem.* **7**, 33 (1977).
- [6] O. G. Berg and C. Blomberg, *Biophys. Chem.* **8**, 271 (1978).
- [7] O. G. Berg, R. B. Winter, and P. H. von Hippel, *Biochemistry* **20**, 6929 (1981).
- [8] R. B. Winter, O. G. Berg, and P. H. von Hippel, *Biochemistry* **20**, 6961 (1981).
- [9] P. H. von Hippel and O. G. Berg, *J. Biol. Chem.* **264**, 675 (1989).
- [10] M. Coppey, O. Bénichou, R. Voituriez, and M. Moreau, *Biophys. J.* **87**, 1640 (2004).
- [11] S. E. Halford and J. F. Marko, *Nucleic Acids Res.* **32**, 3040 (2004).
- [12] Z. Wunderlich and L. A. Mirny, *Nucleic Acids Res.* **36**, 3570 (2008).
- [13] S. E. Halford, *Biochem. Soc. Trans.* **37**, 343 (2009).
- [14] R. K. Das and A. B. Kolomeisky, *Phys. Chem. Chem. Phys.* **12**, 2999 (2010).
- [15] E. G. Marklund, A. Mahmutovic, O. G. Berg, P. Hammar, D. van der Spoel, D. Fange, and J. Elf, *Proc. Natl. Acad. Sci. USA* **110**, 19796 (2013).
- [16] M. Tabaka, T. Kalwarczyk, and R. Hołyst, *Nucleic Acids Res.* **42**, 727 (2014).
- [17] M. Slutsky and L. A. Mirny, *Biophys. J.* **87**, 4021 (2004).
- [18] Y. M. Wang, R. H. Austin, and E. C. Cox, *Phys. Rev. Lett.* **97**, 048302 (2006).
- [19] P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, and X. S. Xie, *Proc. Natl. Acad. Sci. USA* **103**, 5752 (2006).
- [20] L. Hu, A. Y. Grosberg, and R. Bruinsma, *Biophys. J.* **95**, 1151 (2008).
- [21] B. Bagchi, P. C. Blainey, and X. S. Xie, *J. Phys. Chem. B* **112**, 6282 (2008).
- [22] O. Givaty and Y. Levy, *J. Mol. Biol.* **385**, 1087 (2009).
- [23] V. Dahirel, F. Paillusson, M. Jardat, M. Barbi, and J. M. Victor, *Phys. Rev. Lett.* **102**, 228101 (2009).
- [24] P. C. Blainey, G. Luo, S. C. Kou, W. F. Mangel, G. L. Verdine, B. Bagchi, and X. S. Xie, *Nat. Struct. Mol. Biol.* **16**, 1224 (2009).
- [25] J. Dikić, C. Menges, S. Clarke, M. Kokkinidis, A. Pingoud, W. Wende, and P. Desbiolles, *Nucleic Acids Res.* **40**, 4064 (2012).
- [26] J. S. Leith, A. Tafvizi, F. Huang, W. E. Uspal, P. S. Doyle, A. R. Fersht, L. A. Mirny, and A. M. van Oijen, *Proc. Natl. Acad. Sci. USA* **109**, 16552 (2012).
- [27] M. Sheinman, O. Bénichou, Y. Kafri, and R. Voituriez, *Rep. Prog. Phys.* **75**, 026601 (2012).
- [28] N. P. Stanford, M. D. Szczelkun, J. F. Marko, and S. E. Halford, *EMBO J.* **19**, 6546 (2000).
- [29] D. M. Gowers and S. E. Halford, *EMBO J.* **22**, 1410 (2003).
- [30] I. Bonnet, A. Biebricher, P.-L. Porté, C. Loverdo, O. Bénichou, R. Voituriez, C. Escudé, W. Wende, A. Pingoud, and P. Desbiolles, *Nucleic Acids Res.* **36**, 4118 (2008).
- [31] C. Loverdo, O. Benichou, R. Voituriez, A. Biebricher, I. Bonnet, and P. Desbiolles, *Phys. Rev. Lett.* **102**, 188101 (2009).
- [32] M. C. DeSantis, J. L. Li, and Y. M. Wang, *Phys. Rev. E* **83**, 021907 (2011).
- [33] A. Parsaean, M. O. de la Cruz, and J. F. Marko, *Phys. Rev. E* **88**, 040703 (2013).
- [34] S. E. Halford and M. D. Szczelkun, *Eur. Biophys. J.* **31**, 257 (2002).
- [35] T. Hu, A. Y. Grosberg, and B. Shklovskii, *Biophys. J.* **90**, 2731 (2006).
- [36] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, *J. Phys. A: Math. Theor.* **42**, 434013 (2009).
- [37] D. M. Gowers, G. G. Wilson, and S. E. Halford, *Proc. Natl. Acad. Sci. USA* **102**, 15883 (2005).
- [38] R. H. Porecha and J. T. Stivers, *Proc. Natl. Acad. Sci. USA* **105**, 10791 (2008).
- [39] F. Wang, S. Redding, I. J. Finkelstein, J. Gorman, D. R. Reichman, and E. C. Greene, *Nat. Struct. Mol. Biol.* **20**, 174 (2013).
- [40] L. J. Friedman, J. P. Mumm, and J. Gelles, *Proc. Natl. Acad. Sci. USA* **110**, 9740 (2013).
- [41] T. Ruusala and D. M. Crothers, *Proc. Natl. Acad. Sci. USA* **89**, 4903 (1992).
- [42] K. Burdzy, *Multidimensional Brownian Excursions and Potential Theory* (Longman Scientific & Technical, New York, 1987), Vol. 164.
- [43] B. van den Broek, M. A. Lomholt, S. M. Kalisch, R. Metzler, and G. J. Wuite, *Proc. Natl. Acad. Sci. USA* **105**, 15738 (2008).
- [44] M. A. Lomholt, B. van den Broek, S. M. Kalisch, G. J. Wuite, and R. Metzler, *Proc. Natl. Acad. Sci. USA* **106**, 8204 (2009).
- [45] E. F. Koslover, M. A. Díaz de la Rosa, and A. J. Spakowitz, *Biophys. J.* **101**, 856 (2011).
- [46] M. Bauer and R. Metzler, *PLoS ONE* **8**, e53956 (2013).
- [47] J. Paijmans and P. R. ten Wolde, *Phys. Rev. E* **90**, 032708 (2014).
- [48] N. R. Zabet and B. Adryan, *Mol. Biosyst.* **8**, 2815 (2012).
- [49] N. R. Zabet and B. Adryan, *Bioinformatics* **28**, 1287 (2012).
- [50] M. Barbi, C. Place, V. Popkov, and M. Salerno, *Phys. Rev. E* **70**, 041901 (2004).
- [51] A. Tafvizi, L. A. Mirny, and A. M. van Oijen, *Chem. Phys. Chem* **12**, 1481 (2011).
- [52] A. Marcovitz and Y. Levy, *Proc. Natl. Acad. Sci. USA* **108**, 17957 (2011).
- [53] L. Zandarashvili, D. Vuzman, A. Esadze, Y. Takayama, D. Sahu, Y. Levy, and J. Iwahara, *Proc. Natl. Acad. Sci. USA* **109**, E1724 (2012).
- [54] M. Bauer and R. Metzler, *Biophys. J.* **102**, 2321 (2012).
- [55] N. R. Zabet and B. Adryan, *Bioinformatics* **28**, 1517 (2012).

- [56] A. Veksler and A. B. Kolomeisky, *J. Phys. Chem. B* **117**, 12695 (2013).
- [57] A. Marcovitz and Y. Levy, *J. Phys. Chem. B* **117**, 13005 (2013).
- [58] M. Tabaka, O. Cybulski, and R. Hołyst, *J. Mol. Biol.* **377**, 1002 (2008).
- [59] H. Flyvbjerg, S. Keatch, and D. Dryden, *Nucleic Acids Res.* **34**, 2550 (2006).
- [60] G.-W. Li, O. G. Berg, and J. Elf, *Nat. Phys.* **5**, 294 (2009).
- [61] C. A. Brackley, M. E. Cates, and D. Marenduzzo, *Phys. Rev. Lett.* **111**, 108101 (2013).
- [62] T. E. Kuhlman and E. C. Cox, *Phys. Rev. E* **88**, 022701 (2013).
- [63] N. R. Zabet and B. Adryan, *Front. Genet.* **4**, 197 (2013).
- [64] A. Marcovitz and Y. Levy, *Biophys. J.* **104**, 2042 (2013).
- [65] T. E. Kuhlman and E. C. Cox, *Mol. Syst. Biol.* **8**, 610 (2012).
- [66] S. Furini, C. Domene, and S. Cavalcanti, *J. Phys. Chem. B* **114**, 2238 (2010).
- [67] S. Furini, P. Barbini, and C. Domene, *Nucleic Acids Res.* **41**, 3963 (2013).
- [68] D. Tworowski, A. V. Feldman, and M. G. Safro, *J. Mol. Biol.* **350**, 866 (2005).
- [69] J. C. M. Gebhardt, D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis, and X. S. Xie, *Nat. Methods* **10**, 421 (2013).
- [70] J. Chen, Z. Zhang, L. Li, B.-C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig *et al.*, *Cell* **156**, 1274 (2014).