Intracellular Facilitated Diffusion: Searchers, Crowders, and Blockers

C. A. Brackley, M. E. Cates, and D. Marenduzzo

SUPA, School of Physics and Astronomy, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom

(Received 9 May 2013; published 3 September 2013)

In bacteria, regulatory proteins search for a specific DNA-binding target via "facilitated diffusion": a series of rounds of three-dimensional diffusion in the cytoplasm, and one-dimensional (1D) linear diffusion along the DNA contour. Using large scale Brownian dynamics simulations we find that each of these steps is affected differently by crowding proteins, which can either be bound to the DNA acting as a road block to the 1D diffusion, or freely diffusing in the cytoplasm. Macromolecular crowding can strongly affect mechanistic features such as the balance between three-dimensional and 1D diffusion, but leads to surprising robustness of the total search time.

DOI: 10.1103/PhysRevLett.111.108101 PACS numbers: 87.10.Mn, 87.15.ap

Many cellular processes involve the binding of a protein to a specific target base pair (bp) sequence of DNA [1]. As targets are typically 10–20 bp in size, it is remarkable that proteins can quickly and accurately locate them on a bacterial chromosome which is over 10^6 bp long. Interest in the mechanism of the target search first arose in the 1970s, when experiments by Riggs *et al.* [2] measured the *in vitro* rate at which the *lac* repressor binds with its promoter; early interpretations of these results found that rate to be much greater than would be expected for a search via simple diffusion in three dimensions [3]. This prompted a series of papers by Berg and von Hippel [6,7], in which they developed an analytical model of a protein-DNA target search known as *facilitated diffusion*. Their premise was that portions of the search involve events where, via a nonspecific (sequence independent) interaction, the searching protein "slides" along the DNA backbone—effectively performing diffusion in one dimension. Following these seminal works, much theoretical and computational effort [8–20] has been spent addressing different facets of this important problem.

However, the process is still not fully understood, despite recent advances in single molecule imaging techniques that have given new insight into facilitated diffusion mechanisms [21–23]. Most importantly, the bacterial cytoplasm within which the search is performed is very different from the test tube of *in vitro* experiments: it is a very crowded environment, with about a million proteins per cell [1,24], many of which bind to the DNA and perform functions such as transcription, gene regulation, replication and repair, and chromosome structuring [1,25]. So far, only limited theoretical [26,27] or simulation work explicitly addresses this issue. Furthermore, colloidal physics arguments suggest that crowding should give rise to depletion [28] and other entropic interactions between the constituents of the system: these are important for the thermodynamics of many intracellular processes [29,30], but their impact on facilitated diffusion has not yet been explored.

Our goal in this Letter is to investigate the effects of macromolecular crowding, both in the cytosol and along the DNA, on the target search process. Our most interesting results are those for diffusing crowder proteins. These strongly bias the search towards one-dimensional (1D) diffusion along the genome, as found *in vivo* [21], and through an intriguing combination of effects lead, at physiological crowder densities, to a search time which is robust to changes in protein-DNA affinity, as might be beneficial within a living cell. We also show that blockers (proteins tightly bound to, and diffusing along, the DNA) do not greatly hinder facilitated diffusion, even at surprisingly large densities. To gain our results, we use coarse grained Brownian dynamics (BD) simulations. These allow us to explicitly include the (often disregarded) conformational dynamics of DNA, provide us with a detailed treatment of the three-dimensional (3D) component of the search, and avoid the need to make the *a priori* assumptions on the relative time scales of the kinetics in the search process which are necessary in most analytical approaches [6,13,31].

For the simulations we coarse grain DNA as a bead-and-spring polymer. The level of coarse graining we use (the size of a polymeric bead equals the hydration thickness of DNA, 2.5 nm) allows many repeat simulations of large systems to be performed, at the cost of losing atomistic details in the resolution of the genome and protein structures. As in Ref. [18], we model searcher proteins as a sphere with a small patch; the latter is sticky and acts as a DNA-binding site, interacting with an energy ϵ. For simplicity, we neglect sequence heterogeneity—this can potentially lead to more complicated 1D dynamics which have been studied in Ref. [18] in the absence of macromolecular crowding [32]. We introduce two further protein types which we shall refer to as *crowders* and *blockers* [Fig. 1]. Crowders are proteins which diffuse freely in the space around the DNA, and are modeled as simple spheres. Blockers are modeled as larger proteins which bind strongly to the DNA (but can still diffuse along it),

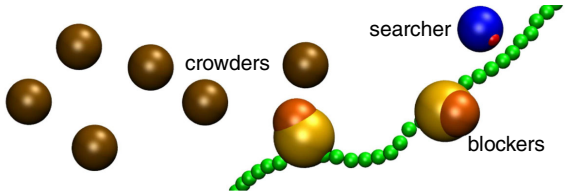

FIG. 1 (color online). A bead-and-spring DNA with crowder, blocker, and searcher proteins.

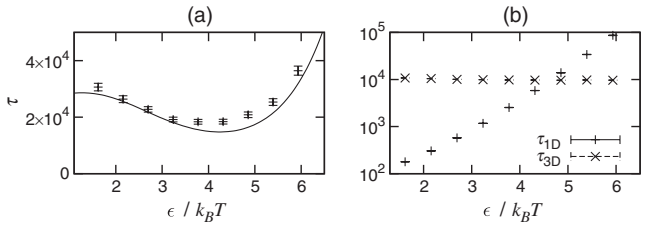

FIG. 2. Facilitated diffusion in the absence of macromolecular crowding. (a) Plot showing how the total search time depends on the searcher-DNA interaction strength ϵ, for a system with a single searcher protein, and no blockers or crowders. The solid line shows a fit to Eq. (1). (b) Plot showing the mean duration of 1D and 3D search episodes from the same simulations.

blocking the path of the searchers from all sides. The simulations are performed using the LAMMPS code [33]. We use a system size which reproduces a typical bacterial DNA volume fraction of 1%; periodic boundary conditions are employed, so that the genome is under compaction (its radius of gyration is larger than the system size), similar to the situation *in vivo*. Full details of the simulation scheme are given in Ref. [34].

We analyze our results with reference to the analytic model of Halford and Marko [31]: while more detailed models exist [4,15,16,35], this framework allows a simple explanation of how different parameters affect the search. According to Ref. [31], if during each 1D sliding episode the protein searches on average a length l_s, then it will require on average $N_s = L/l_s$ rounds of 3D and 1D diffusion to find the target on a DNA molecule of length L. This relationship between N_s and l_s depends on the assumption that there is no correlation between the regions of the DNA which are searched during successive slides (some recent theories have sought to address this [4,15,35]). Scaling arguments then give the relations $\tau_{3D} \sim V/LD_3$ and $\tau_{1D} \sim l_s^2/D_1$ for the mean duration of each 1D or 3D search event, where D_3 and D_1 are 3D and 1D diffusion constants, and V is the system volume. This leads to an equation $\tau = N_s(\tau_{1D} + \tau_{3D})$, and finally

$$\tau = A\frac{Ll_s}{D_1} + B\frac{V}{D_3 l_s}, \qquad (1)$$

for the mean total search time. Here the dimensionless prefactors A, B cannot be inferred from simple scaling.

In Ref. [18] we showed that Eq. (1) only gives a good fit to the BD results for an *unstructured* DNA molecule. Figure 2(a) shows the mean total search time as a function of the protein-DNA interaction ϵ for such a system, and Fig. 2(b) shows individually the 1D and 3D search times. In the theory of Ref. [31], l_s is taken as a parameter, whereas in the BD model we can only directly vary ϵ. Now l_s is a function of ϵ, as is D_1 (which was not the case in [31], but might be expected for a protein sliding through a rugged potential due to a sequence of discrete monomers). This means that there is an optimal value of ϵ where the search time is minimized. From Eq. (1) at the minimum, the fraction of time spent sliding $f = \tau_{1D}/(\tau_{1D} + \tau_{3D}) = 1/2$. (This holds only approximately in Fig. 2 due to the additional dependence of D_1 on ϵ in our simulations.)

A natural starting point to address the effects of crowding would be to consider a system containing *multiple searcher proteins* and a DNA molecule with a single target. If the search for the target were a simple Poisson process, one would expect the mean search time to scale inversely with the number of searchers; i.e., the search time for M searchers should obey $\tau(M) = \tau(1)M^{-1}$. This relationship gives a very good approximation to our simulation results in most cases, and we provide full details in [34].

Much more interesting consequences arise when considering the search process in the presence of many *crowder proteins*, whose volume fraction we denote by ϕ. The presence of the crowders has a large, but opposite effect on the 3D and 1D components of the search [Fig. 3(a)]: an increase in crowder density from zero to $\phi = 0.25$ leads to a roughly twofold decrease in τ_{3D}, but to an increase by a similar factor in τ_{1D}. Both results can be attributed to a depletionlike interaction [28,29] between the searcher and the DNA: when the searcher is close to the DNA, there is a region between it and the polymer from which crowders are excluded due to steric interactions. The resulting osmotic pressure due to the crowders acts to effectively increase the searcher-DNA attraction when the protein is bound (a known effect of macromolecular crowding [36]), and an increased likelihood of the searcher immediately returning to the DNA after is has escaped. This manifests as the progressive deviation away from a simple exponential for the probability distribution of 3D excursion times [Fig. 3(b), top], in favor of a distribution suggestive of biphasic or stretched exponential behavior.

Interestingly, the effect of increasing crowder density on the number of search rounds N_s is also different depending on the value of ϵ [Fig. 3(c)]. There are two competing effects here. Because of the increase of τ_{1D}, we expect an increase of l_s with ϕ which should lead to a decrease in N_s (assuming $N_s = L/l_s$). However, the diffusion of both the DNA and the searcher [37] is hindered by the crowders and this leads to an increased likelihood of repeatedly searching the same DNA region, causing a breakdown of the $N_s = L/l_s$ relation. This is demonstrated by the fact that the product $N_s l_s$ increases with ϕ for all ϵ (see [34]).

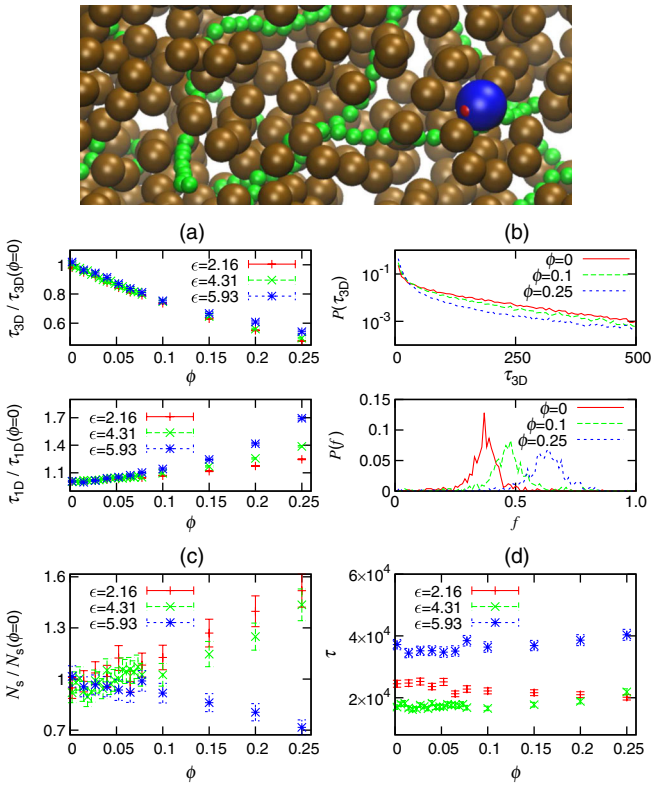

FIG. 3 (color online). Simulation results for systems containing crowder proteins, and a single searcher. Top: Snapshot of the DNA, crowders and the searcher during the simulation. (a) The mean 3D and 1D search times scaled by the time for $\phi = 0$. (b) The probability distribution functions for the duration of 3D excursions (top), and for the fraction of time spent sliding, f (bottom), for $\epsilon = 4.31 k_B T$. (c) The mean number of search rounds required to find the target, scaled by the value for $\phi = 0$, and (d) the mean total search time.

N_s increases with ϕ for small ϵ, but for the $\epsilon = 5.93 k_B T$ case shown in the figure, the increase in l_s dominates, and N_s *decreases* with ϕ.

The various dependencies on ϕ of N_s, τ_{1D} and τ_{3D} at different ϵ combine to give remarkable stability of the total mean search time τ [Fig. 3(d)]. For small ϵ, τ decreases with ϕ whereas for large ϵ it increases. However, despite the large changes in τ_{1D} and τ_{3D}, the overall change in τ remains small (see [34]). Moreover, at the physiologically relevant $\phi \approx 0.2$, τ shows little variation between the $\epsilon \sim 2 k_B T$ and $4 k_B T$ cases (a plot showing τ as a function of ϵ is given in [34]).

The depletion attraction thus enhances the robustness of the search time. According to Eq. (1) the cell must precisely tune the protein-DNA interaction to optimize the search—which is difficult since ϵ will vary from protein to protein, depend on DNA sequence, and be dependent on conditions such as salt concentration. Thus, our finding that τ is relatively insensitive to ϵ near its minimum suggests that the facilitated diffusion search mechanism is better adapted to the subcellular conditions in which it operates than to a hypothetical, uncrowded environment. Another important result is that there is not a clear link between the fraction of time spent sliding, f (recently measured *in vivo* [21]) and the search time. From Eq. (1), an optimal search leads to $f = 0.5$, whereas Fig. 3(b) shows that the probability distribution of f gets shifted to sliding-dominated searches as ϕ increases, independent of ϵ, with f increasing twofold as ϕ increases from 0 to 0.25.

We finally consider *blocker proteins*, which bind to, and diffuse along, the DNA rather than diffuse within the cytosol. For transparency we consider their effects in isolation, addressing a system with no crowders and a single searcher. Similar "road-block" proteins have previously been considered analytically in [26,27]. The theory by Li *et al.* [26] predicts an overall increase in the search time τ due to the reduced l_s, and the chance that the target is covered by a blocker. Our results are qualitatively in agreement with this, showing an increase in τ with an average blocker coverage density ρ which is almost independent of ϵ [Fig. 4(a)].

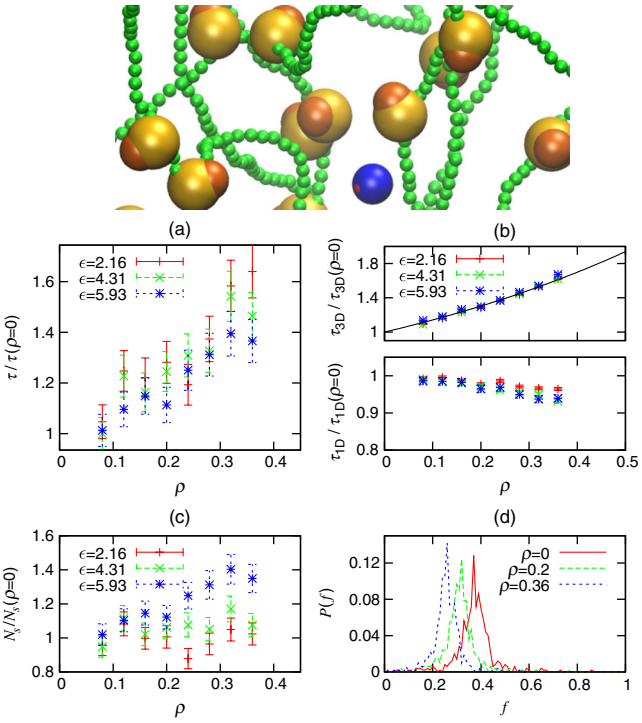

FIG. 4 (color online). Simulation results for systems containing blocker proteins, and a single searcher. ρ is defined as the average fraction of DNA covered by blockers during the simulation. Top: Snapshot of the DNA, blockers, and the searcher during the simulation. (a) The mean search time as a function of ρ, scaled by the search time for the $\rho = 0$ case. (b) The mean 3D and 1D search times as a function of ρ. The black line shows an exponential fit to the $\epsilon = 4.31 k_B T$ results. (c) The mean number of search rounds required to find the target, scaled by the value for $\rho = 0$. (d) The probability distribution function for the fraction of time spent sliding, f, for $\epsilon = 4.31 k_B T$.

Quantitatively, however, this increase is not dramatic; this is perhaps another well-adapted feature of facilitated diffusion within a cell where a large number of DNA-binding proteins must be present at any time [1]. We find that the duration of each 3D excursion (τ_{3D}) increases with ρ since it takes longer for the searcher to encounter DNA which is free from blockers; Fig. 4(b) (top) shows a good fit to an exponential function, which is expected since the probability of finding an uncovered region of a given length decreases exponentially with coverage density [26,38]. In contrast to Ref. [26], we also observe a small decrease in τ_{1D} as ρ increases [Fig. 4(b), bottom]. Presumably, collisions between the searcher and the blockers can lead to the less strongly bound searcher being "knocked off" of the DNA. We note that in our model, searchers can only bypass blockers by transiently detaching from the DNA. While this is a good assumption for large blocking proteins or protein clusters (such as polymerases, replication, or transcription factories), a searcher may in practice hop over smaller proteins without leaving the genome, as discussed in Ref. [39]—this effect may further diminish the slowdown of the search process imparted by blockers.

Figure 4(c) shows that the number of search rounds also increase with ρ. This stems from the reduction in l_s (due to reduction in τ_{1D}), and of the increased likelihood of the target being covered by a blocker. Because of the effects in Fig. 4(b), the presence of blockers also affects the fraction of time spent sliding, f: now the target-finding mechanism becomes more and more dominated by 3D excursion as the DNA gets increasingly obstructed [Fig. 4(d)].

To summarize, in this work we have studied the effect on facilitated diffusion of macromolecular crowding caused by the simultaneous presence of many proteins in the cellular environment. Crowding proteins diffusing in the cytosol lead to depletion interactions which effectively increase the searcher-DNA affinity, and reduce the 3D search time, at the same time increasing the chance of repeat-pass searches through the same DNA region. This leads to a decrease in search time for small ϵ but an increase for large ϵ meaning that at physiological volume fractions of crowders the curve shown in Fig. 2(a) becomes flatter, with τ increasing only at large ϵ. That τ is independent of ϵ near its minimum may be advantageous for efficient searching, since in reality the nonspecific searcher-DNA interaction must vary both among searchers and along the DNA sequence. Robustness to changes in ϵ has previously been shown to result from intersegmental transfers [40,41], in that case at high DNA-protein affinity; this process relies on DNA strands passing close to each other in space and so depends on the compaction of the DNA. Because of their geometry, intersegmental transfers occur relatively infrequently with our searchers; concurrent investigation of both of these effects would be an interesting future study.

For DNA-binding *blocker* proteins, we have confirmed previous analytical work which predicted an exponential increase in the 3D search time due to impaired access to the DNA [26]. Our simulations also suggest that the blocker-induced increase in the total search time is modest, which is again advantageous in the biological context where a large number of DNA-binding proteins must be present simultaneously for the proper functioning of both bacterial and eukaryotic chromosomes. Finally, we have shown that if multiple searcher proteins are present then the search time scales inversely with the copy number, except at very high copy numbers (see [34]).

Our work suggests that a proper account of the crowded cellular environment is crucial for a full understanding of the protein-DNA target search. This is becoming an increasingly important issue as direct experimental probes of the process *in vivo* become a reality [21,23]. We have found that, at least in terms of the robustness of the total search time τ, the facilitated diffusion mechanism is well adapted to the crowded subcellular conditions in which it takes place. This work also lays the foundations for future studies of the combined effect of crowders and blockers, the role of architectural DNA-binding proteins, the effect of genome confinement [42,43], and the possible influence of hydrodynamic interactions [44].

We acknowledge EPSRC Grant No. EP/I034661/1 for funding. M. E. C. is funded by the Royal Society.

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell* (Garland Science, New York, 2002).
[2] A. Riggs, S. Bourgeois, and M. Cohn, J. Mol. Biol. **53**, 401 (1970).
[3] The Riggs experiment [2] and many subsequent works were performed at low ionic strengths, and it is still a matter of debate in the literature [4,5] as to whether the measured rates do exceed the 3D limit if this is taken into account. Nevertheless, it is still impressive that rates can reach anywhere near this limit given the inevitability that proteins will interact nonspecifically with the DNA.
[4] A. B. Kolomeisky, Phys. Chem. Chem. Phys. **13**, 2088 (2011).
[5] S. E. Halford, Biochem. Soc. Trans. **37**, 343 (2009).
[6] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry **20**, 6929 (1981).
[7] P. von Hippel and O. Berg, J. Biol. Chem. **264**, 675 (1989).
[8] I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Biophys. J. **89**, 895 (2005).
[9] I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Phys. Rev. E **72**, 041102 (2005).
[10] C. Loverdo, O. Bénichou, R. Voituriez, A. Biebricher, I. Bonnet, and P. Desbiolles, Phys. Rev. Lett. **102**, 188101 (2009).
[11] N. R. Zabet and B. Adryan, Bioinformatics **28**, 1517 (2012).

[12] G. D. Stormo and D. S. Fields, Trends Biochem. Sci. **23**, 109 (1998).
[13] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).
[14] U. Gerland, J. D. Moroz, and T. Hwa, Proc. Natl. Acad. Sci. U.S.A. **99**, 12015 (2002).
[15] O. Bénichou, Y. Kafri, M. Sheinman, and R. Voituriez, Phys. Rev. Lett. **103**, 138102 (2009).
[16] M. Sheinman, O. Bénichou, Y. Kafri, and R. Voituriez, Rep. Prog. Phys. **75**, 026601 (2012).
[17] A. Marcovitz and Y. Levy, Proc. Natl. Acad. Sci. U.S.A. **108**, 17957 (2011).
[18] C. A. Brackley, M. E. Cates, and D. Marenduzzo, Phys. Rev. Lett. **109**, 168103 (2012).
[19] A.-M. Florescu and M. Joyeux, J. Chem. Phys. **130**, 015103 (2009).
[20] A.-M. Florescu and M. Joyeux, J. Phys. Chem. A **114**, 9662 (2010).
[21] J. Elf, G.-W. Li, and X. S. Xie, Science **316**, 1191 (2007).
[22] Y. M. Wang, R. H. Austin, and E. C. Cox, Phys. Rev. Lett. **97**, 048302 (2006).
[23] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, Science **336**, 1595 (2012).
[24] J. Pelletier, K. Halvorsen, H. Bae-Yun, R. Paparcone, S. J. Sandler, C. L. Woldringh, W. P. Wong, and S. Jun, Proc. Natl. Acad. Sci. U.S.A. **109**, E2649 (2012).
[25] M. Nicodemi and A. Prisco, Biophys. J. **96**, 2168 (2009).
[26] G.-W. Li, O. G. Berg, and J. Elf, Nat. Phys. **5**, 294 (2009).
[27] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, J. Phys. A **42**, 434013 (2009).
[28] S. Asakura and F. Oosawa, J. Polym. Sci. **33**, 183 (1958).
[29] D. Marenduzzo, K. Finan, and P. R. Cook, J. Cell Biol. **175**, 681 (2006).
[30] J. A. Dix and A. S. Verkman, Annu. Rev. Biophys. **37**, 247 (2008).
[31] S. E. Halford and J. F. Marko, Nucleic Acids Res. **32**, 3040 (2004).
[32] Recent models have proposed that most DNA-binding proteins can interact with the DNA in two different modes: a nonspecific search mode and a recognition mode; in the latter the protein moves through a disordered potential landscape due to the sequence [13,14]. If the energy barrier for switching between the two modes is large enough for all but the target sequence, the search may be performed entirely in the nonspecific mode—this has been recently suggested to be the case for some proteins at least [15,16]. In this situation, sequence disorder would be irrelevant.
[33] S. Plimpton, J. Comp. Physiol. **117**, 1 (1995).
[34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.111.108101 for full details of the simulation scheme, results for the simulations with multiple searcher proteins, and some additional plots relating to the data presented in Figs. 3 and 4.
[35] A. G. Cherstvy, A. B. Kolomeisky, and A. A. Kornyshev, J. Phys. Chem. B **112**, 4741 (2008).
[36] S. V. Zimmerman and B. Harrison, Proc. Natl. Acad. Sci. U.S.A. **84**, 1871 (1987).
[37] The motion of the searcher when it is not bound to the DNA remains diffusive for all values of ϕ investigated, although its diffusion coefficient decreases with ϕ [30], as does that of the DNA beads.
[38] J. D. McGhee and P. H. von Hippel, J. Mol. Biol. **86**, 469 (1974).
[39] A. Marcovitz and Y. Levy, Biophys. J. **104**, 2042 (2013).
[40] An intersegmental transfer is an event where a searcher moves from one DNA strand directly to another—which is close in space, but may be distant along the contour—by transiently binding with both.
[41] M. Sheinmann and Y. Kafri, Phys. Biol. **6**, 016003 (2009).
[42] G. Foffano, D. Marenduzzo, and E. Orlandini, Phys. Rev. E **85**, 021919 (2012).
[43] M. Bauer and R. Metzler, PLoS One **8**, e53956 (2013).
[44] B. Dünweg and A. J. C. Ladd, *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, Advances in Polymer Science Vol. 221 (Springer, Berlin, Heidelberg, 2009), pp. 89–166.