Many-body effects on tracer particle diffusion with applications for single-protein dynamics on DNA

Sebastian Ahlberg\textsuperscript{1}, Tobias Ambjörnsson\textsuperscript{2} and Ludvig Lizana\textsuperscript{1}

\textsuperscript{1} Integrated Science Lab, Department of Physics, Umeå University, SE-901 87 Umeå, Sweden
\textsuperscript{2} Department of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, SE-223 62 Lund, Sweden

E-mail: ludvig.lizana@physics.umu.se

Keywords: diffusion, crowding, biophysics

Abstract

30\% of the DNA in \textit{E. coli} bacteria is covered by proteins. Such a high degree of crowding affects the dynamics of generic biological processes (e.g. gene regulation, DNA repair, protein diffusion etc) in ways that are not yet fully understood. In this paper, we theoretically address the diffusion constant of a tracer particle in a one-dimensional system surrounded by impenetrable crowder particles. While the tracer particle always stays on the lattice, crowder particles may unbind to a surrounding bulk and rebind at another, or the same, location. In this scenario we determine how the long time diffusion constant \( \zeta \) (after many unbinding events) depends on (i) the unbinding rate of crowder particles \( k_{\text{off}} \), and (ii) crowder particle line density \( \rho \), from simulations (using the Gillespie algorithm) and analytical calculations. For small \( k_{\text{off}} \), we find \( \zeta \sim k_{\text{off}}^2 \rho \) when crowder particles do not diffuse on the line, and \( \zeta \sim D k_{\text{off}} / \rho \) when they are diffusing; \( D \) is the free particle diffusion constant. For large \( k_{\text{off}} \), we find agreement with mean-field results which do not depend on \( k_{\text{off}} \). From literature values of \( k_{\text{off}} \) and \( D \), we show that the small \( k_{\text{off}} \)-limit is relevant for \textit{in vivo} protein diffusion on crowded DNA. Our results apply to single-molecule tracking experiments.

1. Introduction

There is not much doubt that macromolecular crowding has severe consequences for biological processes inside living cells [1]. In an \textit{E. coli} bacterium the concentration of proteins and RNA is about 300–400 mg ml\(^{-1}\) [2] which is 30–40 times higher than common test tube conditions [3]. There is overwhelming evidence that this level of crowding influences, for example, gene regulation [4], DNA binding constants [5], enzymatic activity [6], protein folding [7, 8], and the diffusion of macromolecules [9]. In order to get a complete picture of \textit{in vivo} dynamics we need a better understanding of the role of crowding, and recent experimental developments provide the means to achieve this.

In recent years, researchers have beaten the diffraction limit and turned optical microscopy into ‘nanoscopy’. Today’s microscopy methods (e.g. STED, STORM and FIONA) [10] not only allow us to image nanometer-sized biological structures, but recent improvements [11, 12] also permit the tracking of fluorescently labelled proteins at the biologically relevant millisecond-scale. This is anticipated to shed new light on biological processes, as well as increase our understanding of particle transport in nano-fluidic systems [13, 14]. In order to properly interpret those experiments \textit{in vivo}, we need new theoretical and computational models in terms of physical properties of the intracellular space, the cytoplasm.

The cytoplasm is cramped with macromolecules and we are interested in how this influences diffusion-controlled processes, a key component in many cellular functions (e.g. gene regulation). While our results are new, aspects of this problem have been studied theoretically before. For example, [15–18] investigate three dimensional diffusion in the cytoplasm and in gels, whereas [19–23] focus on the subdiffusive motion seen in single-molecule experiments. Crowding is also important for DNA search processes where a searcher combines one- and three-dimensional diffusion to quickly find its target, so-called facilitated diffusion [24]. Facilitated
diffusion under crowding is addressed in [4, 25], which resemble this paper, but we ask different questions: rather than focusing on mean target finding times, we calculate the diffusion constant of a tracer particle that is unable to unbind in terms of the key properties of surrounding crowder particles.

Much of the inspiration for this work comes from DNA binding proteins. Of particular interest are repair proteins (MutS and homologs) whose residence time can be very long (~10 min [26]). Once the MutS finds a mismatched base pair it changes conformation (using ATP) into a long-lived sliding clamp that encircles the DNA [27] and diffuses passively along the DNA [28]. We are also inspired by transcription factors, the family of gene regulatory proteins. The yeast regulatory proteins LexA and Gal4 can stay bound to their regulatory sites for several minutes in vitro (LexA ~ 5 min and Gal4 > 30 min) [29], but surprisingly, this number can be reduced by up to 1000 times in vivo. Both classes of proteins have the ability to diffuse along the DNA, unbind to the three dimensional intracellular space, diffuse in space, and rebind to the DNA. We are interested in how the dynamics of those proteins change under crowding.

In order to better understand the role of crowding, we introduce a theoretical model where particles diffuse on a one dimensional lattice where two particles cannot occupy the same site (figure 1). They diffuse at the rate $k_D$ (the same for all particles) and may unbind and rebind to the lattice with the rates $k_{off}$ and $k_{on}$, respectively. These rates are tuned so that the average particle line density is constant at 10–20%, which is not too far from in vitro conditions (30% of the DNA in E. coli is covered by proteins). The unbinding rate for the tracer particle is set to zero, similar to the long-lived protein–DNA complexes described above. Now we ask: What is the long time diffusion constant of a tracer particle in such a crowded quasi-one-dimensional system?

We answer this question numerically using stochastic simulations (Gillespie algorithm), corroborated with analytical results. The main results are shown in figures 4–6 where we show how the diffusion constant changes as a function of our main parameter $k_{off}$. Those results are applicable to single molecule tracking experiments [30].

The unbinding rate $k_{off}$ interpolates between two well studied limits. (i) When $k_{off}$ is large (compared to $k_D$), the tracer’s mobility is only weakly lowered and diffuses almost as if it was free. (ii) When $k_{off} \to 0$, the particles diffuse with unchanged order in a single file. Single-file diffusion is well studied [31–37], where the most famous result is that the mean squared displacement of a tracer particle is proportional to $\sqrt{t}$ rather than $t$ ($t$ is time), which signatures non-Markovian dynamics.

This paper is organised as follows. In section 2, we briefly outline the details of our model. Before showing the results in section 4, we provide analytical estimates of the diffusion constant in section 3, based on a theoretical calculation found in appendix A. In section 3 we also briefly review the dynamics of the model at short, intermediate and long times. We close with a few concluding remarks in section 5.

2. The model

Our model has been used and explained elsewhere [38], but for completeness we summarise it briefly below. Consider a one dimensional lattice on which crowder particles (assumed identical) and the tracer particle diffuse (figure 1). The crowder particles can diffuse, unbind and rebind to the lattice. Rebinding occurs in two ways, either to a random unoccupied lattice site (chosen with uniform probability), or to the exact same location. Both rebinding modes have been used to model transcription factor dynamics on DNA [24, 39], and we will therefore consider both. We, however, do not consider the role of DNA conformation as in [40–42]. The lattice constant is denoted $a$, and the diffusion rate $k_D$ is assumed to be equal in both directions and for all particles. Double occupancy is forbidden and a particle cannot overtake a flanking neighbor (single-file condition). Binding and unbinding dynamics of crowders are characterised by the rates $k_{on}$ and $k_{off}$, which are chosen such that the particle line density is in equilibrium with the bulk, thereby keeping the average filling fraction constant. In our simulations we keep it at 10–20%. We implemented the model using the Gillespie algorithm. (see appendix C).
3. Analytical estimates for the long time diffusion constant

Here we provide analytical estimates to corroborate and better understand the numerical results in the next section. We are mainly interested in the long time diffusion constant \( D \) for the tracer particle, defined through

\[
\langle x^2(t) \rangle \simeq 2Dt
\]  

(1)

where \( \langle x^2(t) \rangle \) is the ensemble averaged mean squared displacement (MSD), and \( t \) is time. Notably, \( D \) is in general not equal to the bare, or free particle, diffusion constant \( D_{\text{free}} = a^2 k_D \).

It is a nontrivial function of \( k_{\text{off}} \) and \( \rho \). To better understand what we mean by long time, we describe in section 3.4 the dynamics leading up to equation (1). But first we summarise our main analytical findings from appendix A which, in figures 4–6, we compare to simulations.

3.1. Long time, small \( k_{\text{off}} \) behaviour

The main idea of our derivation is to calculate the typical length, \( l_0 \), the tracer travels before being hindered by a crowder particle. In terms of \( l_0 \) we have

\[
\tau = \frac{l_0^2}{2\tau},
\]  

(3)

where \( \tau \) is the typical waiting time until a successful jumping event for the tracer. Since the diffusion rate \( k_D \) is fast, the rate limiting step for the tracer particle to move is when a flanking crowder particle unbinds from the lattice. We can therefore envision the tracer’s motions as a single particle diffusion process on a coarse grained lattice with lattice constant \( l_0 \) and jump rate \( 1/\tau \). In short, we choose \( l_0 \) as the standard deviation of the distribution of nearest-neighbour distances, and \( \tau \) as the mean-field corrected inverse unbinding rate of crowder particles \( \rho^{-1}k_D(1 - \rho) \). Appendix A contains the explicit calculations, and the results are below.

To simplify matters, we start by assuming that the crowder particles sit equidistantly on the line with density \( \rho \), unable to diffuse (\( k_D = 0 \)), and rebind to the site from which they unbound. Intuitively, \( l_0 \) should be proportional to \( \rho \), and indeed, a careful calculation shows that \( l_0 = \rho^{-1}\sqrt{3/2} \) which gives

\[
\langle x^2(t) \rangle \simeq -\rho^{-1}\rho\frac{3(1 - \rho)k_{\text{off}}}{4\rho^2}.
\]  

(4)

When crowder particles rebind to a random location rather than to the same site, the distance between two neighbouring particles fluctuate even though the average density is fixed. This leads to a larger effective lattice spacing, \( l_0 = \rho^{-1}\sqrt{(2 - \rho)/2} \), and a larger \( D \)

\[
D \simeq \frac{(2 - \rho)(1 - \rho)}{4\rho^2}k_{\text{off}}.
\]  

(5)

When crowder particles also diffuse (\( k_D \neq 0 \)) the distance between nearest-neighbours becomes difficult to define. We estimate the coarse grained lattice constant as the length the tracer particle explores during a time interval proportional to \( 1/k_{\text{off}} \). This leads to \( l_0 = [4D/(4k_{\text{off}}) \times (1 - \rho)/(\rho^2)]^{1/4} \), and

\[
D \simeq \frac{(1 - \rho)^{3/2}}{\rho} \sqrt{\frac{k_{\text{off}}}{2\pi}}.
\]  

(6)

which has different \( k_{\text{off}} \)-scaling than before. Equations (4)–(6) constitute our main analytical results.

3.2. Long time, large \( k_{\text{off}} \) behaviour

When \( k_D \geq k_{\text{off}} \), crowder particles frequently unbind and rebind to the lattice and the no-passing condition is effectively violated. But crowder particles still hinder the tracer thereby decreasing the diffusion rate. Imagine that the jump rate for a single particle to a neighbouring site on an otherwise empty lattice is \( k_D \), or \( D = a^2 k_D \). Then, when crowder particles are around, some of the jumps are canceled because the target lattice site may be occupied. In that situation the jump rate is reduced by the probability that the target lattice is unoccupied. For very large \( k_{\text{off}} \) this probability is simply \( 1 - \rho \), therefore

\[
D \simeq D(1 - \rho a).
\]  

(7)

This mean field result has been obtained before \([21, 43, 44]\), and as \( k_{\text{off}}/k_D \) is close to or smaller than unity, corrections to this formula become prominent (see figure 5). A better expression for the self-diffusion constant in a system similar to ours valid for \( k_{\text{off}} \approx k_D \) and any density, was found in \([45, 46]\).
3.3. Interpolation formula for $D$
Not all systems fall into the small or large $k_{off}$ category. For those systems that do not, we propose a simple expression, potentially useful for single particle tracking experiments:
\[
\frac{1}{D} = \frac{1}{D_{\text{small } k_{off}}} + \frac{1}{D_{\text{large } k_{off}}},
\]
where $D_{\text{large } k_{off}}$ is equation (7) and $D_{\text{small } k_{off}}$ is one of the equations (4)–(6), depending on the case under study. Equation (8) is constructed so that the small and large $k_{off}$ limits agree with our predictions, but it is a mere phenomenological expression. Nevertheless, we find that it matches simulations surprisingly well (see figure 6).

3.4. How is the long time asymptotics (equation (1)) approached?
Here we clarify the meaning of short, intermediate and long times within our model. To keep the discussion simple, we consider $k_{off}$, $k_{coll}$, $k_D$, and $\rho$ as constant. See also figure 3 which shows $\langle x^2(t) \rangle$ as a function of time, where all relevant regimes are present.

At most we have three regimes of different behaviour. These are separated by the average residence time of the crowder particles $\tau_{off}$, and the average collision time $\tau_{coll}$, which is the time it takes for a particle to diffuse across the average nearest neighbour distance $1/\rho$:
\[
\tau_{coll} = \frac{1}{\rho^2 D}, \quad \tau_{off} = \frac{1}{k_{off}}.
\]
Let us assume that there is a clear separation between these timescales and that $\tau_{coll} \ll \tau_{off}$ and that $k_D$ is the fastest rate in the system, $1/k_D \ll \tau_{off}$, where all relevant regimes are present.

In the third regime, $t \gg \tau_{off}$, particles start unbinding from the lattice which effectively violates the no-passing condition. In this regime we expect diffusive behaviour again $\langle x^2(t) \rangle \sim t$, but with a diffusion constant different from $D$, denoted by $D$ (see equation (1)). This is the one we wish to calculate, in particular in terms of our key parameter $k_{off}$. Note that the second regime can be erased completely if we lower $\tau_{off}$, which is evident from the broad scattering of curves. The particle collisions have taken place but particles diffuse with maintained order since they are unable to pass each other. This is the single-file diffusion regime which is characterised by the memory effects dominate and are the very reason for the subdiffusive behaviour. In the second regime, $t \ll \tau_{coll}$, many particle collisions have taken place but particles diffuse with maintained order.

3.4. How is the long time asymptotics (equation (1)) approached?
Here we clarify the meaning of short, intermediate and long times within our model. To keep the discussion simple, we consider $k_{off}$, $k_{coll}$, $k_D$, and $\rho$ as constant. See also figure 3 which shows $\langle x^2(t) \rangle$ as a function of time, where all relevant regimes are present.

At most we have three regimes of different behaviour. These are separated by the average residence time of the crowder particles $\tau_{off}$, and the average collision time $\tau_{coll}$, which is the time it takes for a particle to diffuse across the average nearest neighbour distance $1/\rho$:
\[
\tau_{coll} = \frac{1}{\rho^2 D}, \quad \tau_{off} = \frac{1}{k_{off}}.
\]
Let us assume that there is a clear separation between these timescales and that $\tau_{coll} \ll \tau_{off}$ and that $k_D$ is the fastest rate in the system, $1/k_D \ll \tau_{off}$, where all relevant regimes are present.

In the third regime, $t \gg \tau_{off}$, particles start unbinding from the lattice which effectively violates the no-passing condition. In this regime we expect diffusive behaviour again $\langle x^2(t) \rangle \sim t$, but with a diffusion constant different from $D$, denoted by $D$ (see equation (1)). This is the one we wish to calculate, in particular in terms of our key parameter $k_{off}$. Note that the second regime can be erased completely if we lower $\tau_{off}$ such that $\tau_{off} \approx \tau_{coll}$ (or smaller). Similarly, the third regime is absent if unbinding is not allowed, i.e. $k_{off} = 0$ (or $\tau_{off} = \infty$). In most of our simulations, diffusion is the fastest process in the system, which is also the likely scenario in a biological cell (see section 5). To sum up,
\[
\langle x^2(t) \rangle \approx \begin{cases} 
2Dt, & t \ll \tau_{coll} \\
(1 - \rho a)/\rho \times \sqrt{4Dt/\pi}, & \tau_{coll} \ll t \ll \tau_{off} \\
2Dt, & t \gg \tau_{off}
\end{cases}
\]

4. Results

In this section we present results from stochastic simulations of the model outlined in section 2, together with our theoretical findings from section 3. The simulation details can be found in appendix C. First, we show the tracer particle’s MSD as a function of time, from which we extract the long time diffusion constant $D$. Second, we investigate $D$ separately for large and small $k_{off}$. Finally, we compare our proposed interpolation formula equation (8) to the full range of $k_{off}$ values.

4.1. Dynamics of the model and extraction of the long time diffusion constant

Figures 2 and 3 show the MSD of the tracer particle as a function of time, for different unbinding rates $k_{off}$. Symbols represent simulated ensemble averages. From such plots we extract the long time diffusion constant $D$ (see appendix B) by fitting a straight line for large times starting from $t = \tau_{off}$ (short vertical dashed lines). The results for $D$ is shown in figures 4–6, but first we discuss some of the features of figures 2 and 3.

Figure 2 shows the MSD when crowder particles do not diffuse but only unbind and rebind. They rebind either always to the same site (upper panel), or to a randomly chosen site (lower panel). To highlight the linear regimes and better see the value of the diffusion constants, we divided the MSD by $2t$. The short time behaviour in both plots is independent of $k_{off}$ and matches well $2Dt$ (upper dashed dark blue line). The long time behaviour is, however, strongly dependent on $k_{off}$, which is evident from the broad scattering of curves. The MSD is still linear in time but the diffusion constant (proportional to the extrapolated intersection with the vertical axis) depends strongly on $k_{off}$. The linear regime sets in when $t \approx \tau_{off}$, as is seen from the shorter vertical
dashed lines. If we increase the particle concentration, the shape of the curves remains the same but the scattering of curves increases, since \( D \propto 1/\rho^2 \) (small \( k_{\text{off}} \)) and \( D \propto 1 - \rho \) (large \( k_{\text{off}} \)).

In Figure 3, crowder particles diffuse and rebind to a randomly chosen site. Rebinding occurs to a randomly chosen site. For shorthand we put \( k_{\text{off}} = k_{\text{off}}/k_D \), \( \tau_{\text{off}} = \tau_{\text{off}} k_D \), and \( \tau_{\text{coll}} = \tau_{\text{coll}} k_D \). Simulation details: lattice constant: \( a = 1 \), diffusion rate: \( k_0 = 1 \) (for all particles including the tracer), filling fraction: \( \rho \approx 0.1 \), number of lattice sites: 501 (\( L = 501a \)), number of simulation runs: 9600.
Figure 4. Long time diffusion constant $D$ as a function the unbinding rate $k_{\text{off}}$, when $k_{\text{off}}$ is small. Symbols represent simulations for the filling fractions $\rho = 0.1$ and $\rho = 0.2$. The dashed lines show our predictions, equations (4)–(6). The panels depict: (top) crowder particles that do not diffuse and rebind to the same location, (middle) crowder particles that do not diffuse and rebind to a random location, (bottom) diffusing crowder particles that rebind to a random location. The data points are extracted from linear fits of figures 2–3 (see appendix B). $R^2$ values from those fits are larger than 0.98.

Figure 5. Long time diffusion constant $D$ as a function of the unbinding rate $k_{\text{off}}$, when $k_{\text{off}}$ is large. Simulation details are the same as in figures 2 and 4.
becomes wider. This is simply because crowder particles have not yet started to unbind from the lattice and therefore diffuse collectively in a single file. We also see that the MSD curves for long times is less scattered than before, indicating that $D$ is less sensitive to $k_{\text{off}}$. This agrees with our theoretical prediction where $D \propto \sqrt{k_{\text{off}}}$ compared to $D \propto k_{\text{off}}$ when crowder particles stand still.

4.2. Small $k_{\text{off}}$ behaviour
In figure 4 we show how $D$ depends on small $k_{\text{off}}$ where each panel depicts: (top) crowder particles that do not diffuse and rebind to the same location, (middle) crowder particles that do not diffuse and rebind to a random location, (bottom) diffusing crowder particles that rebind to a random location. Symbols represent simulation results and dashed lines the small $k_{\text{off}}$ expressions (4)–(6). Each case is plotted for two concentrations, $a\rho = 0.1$ and $a\rho = 0.2$. In order to better compare the three panels with the figures below we scaled the vertical axis with the large $k_{\text{off}}$ limit $D(1 - a\rho)$. In figure 5 we show explicitly how this limit is approached.

Figure 6. Long time diffusion constant $D$ as a function of the unbinding rate $k_{\text{off}}$. Symbols represent simulations for two different filling fractions, $a\rho = 0.1$ and $a\rho = 0.2$. The solid lines show the interpolation formula, equation (8). The panels depict: (top) crowder particles that do not diffuse and rebind to the same location, (middle) crowder particles that do not diffuse and rebind to a random location, (bottom) diffusing crowder particles that rebind to a random location. The data points are a compilation from figures 4 and 5.
The two upper panels in figure 4, where crowder particles do not diffuse are very similar to each other. If both cases were depicted in the same graph, the data points would practically sit on top of each other. For clarity, we therefore separated the data into two figures. We see that the small \( k_{ off} \) behaviour agrees very well with the theoretical results, from equations (4) and (5).

The lower panel depicts when crowder particles diffuse on the lattice. Their movements lead to an overall increase of \( D \) since they no longer act as static road blocks. This also changes the scaling with \( \rho^{2/3} \sim \rho^{1.2} \) from equations (4) and (5), a different \( \rho^{1/2} \approx 0.2 \) when crowder particles stand still, and about 100 times larger when they are moving. We base this estimate on (i) measured one dimensional protein diffusion constants on DNA where \( D \sim 0.1 \mu m^2 s^{-1} \) and (ii) a 30% filling fraction, \( \alpha_p \sim 0.3 \). Considering one base pair as the coarsening of space (\( \alpha_p = 1 \) bp) we find that \( k_{ off} \sim 10^5 s^{-1} \), which when used in the analytical expressions (4)–(6) gives the estimated values for \( D \).

4.3. Large \( k_{ off} \) behaviour

When \( k_{ off} \) is much larger than the diffusion rate \( k_D \), we expect the mean field result, equation (7), to hold. We also expect that corrections to this result become increasingly prominent as \( k_{ off} \) is lowered. Both are confirmed by simulations in figure 5, where we see that \( D/(1 - \alpha_p) \approx 1 \) for \( k_{ off}/k_D \gtrsim 1 \), and \( D/(1 - \alpha_p) < 1 \) for \( k_{ off}/k_D < 1 \). These results validate the mean field argument leading up to equation (7) for our quasi-one-dimensional system. The figure only shows the case where the crowder particles rebind to the same location, since the behaviour at large \( k_{ off} \) is close to identical for all re-binding modes.

4.4. Interpolation formula

In section 3 we proposed equation (8) that ties together the small and large \( k_{ off} \) regimes. The comparison to the full range of \( k_{ off} \) is shown in figure 6 as solid lines (symbols are simulation results). Just as in figure 4, each panel shows: (top) crowder particles that do not diffuse and rebind to the same lattice site, (middle) crowder particles that do not diffuse and rebind to a random lattice site, and (bottom) diffusing crowder particles that rebind to a random site. Overall, equation (8) is a good approximation for the whole range of \( k_{ off} \). The deviations are largest in the transition region, roughly \( 10^{-3} < k_{ off}/k_D < 10^{-4} \) where the maximum relative error for all curves is 79% (top panel, \( \alpha_p = 0.2 \)). The relative error in the small and large \( k_{ off} \) tails is less than 7%.

5. Summary and concluding remarks

We studied the long-time diffusion constant \( D \) of a tracer particle in a one-dimensional crowded many-particle system. We found that \( D \) depends strongly on the unbinding rate \( k_{ off} \) of the surrounding crowder particles and density \( \rho \). For small \( k_{ off} \) we made a simple theoretical model where we deduced that \( D \sim k_{ off}/\rho^3 \) (to first order in \( 1/\rho^2 \)) when crowder particles do not diffuse and only unbind/rebind to the lattice. The prefactor depends on how they rebind, either to the same or to a random site. When they also diffuse we obtain \( D \sim \sqrt{D/\rho^2} \) (to first order in \( 1/\rho \)), a different \( \rho^{1/2} \) scaling than before; \( D \) is the free particle diffusion constant. This means that \( D \) is less sensitive to \( k_{ off} \) and \( \rho \) when crowder particles are diffusing compared to standing still. For large \( k_{ off} \) we found that all cases agreed with the mean field result \( D \approx D(1 - \rho a) \), independent of \( k_{ off} \). Our new expressions showed overall good agreement with the simulations.

It is interesting to see which \( k_{ off}/k_D \) regime we expect to find in the living cell. As mentioned in the introduction, residence times of DNA binding proteins vary from fractions of a second to up to an hour (unspecific binding is even shorter, >5 ms [47]). To get an order of magnitude estimate of \( k_{ off}/k_D \), let us assume that \( k_{ off} \sim 0.1 \mu m s^{-1} \), which lies between the in vivo values for the LexA and Gal4 transcription factors. One dimensional diffusion constants also have a big variation. They are in the range \( D_{1D} \sim 10^{-3} - 10^{-6} (bp)^2 s^{-1} \approx 0.01 - 0.1 \mu m^2 s^{-1} \) [30], which gives \( k_D \approx D_{1D}/(bp)^2 \sim 10^{-5} - 10^{-6} s^{-1} \). This means that \( k_{ off}/k_D \sim 10^{-8} - 10^{-9} \), which clearly indicates that \( k_{ off} \ll k_D \). We should, however, point out that binding energies when proteins are diffusing along the DNA are lower than the values at the specific binding sites. But even for a 1 ms residence time \( k_{ off} \ll k_D \), \( k_{ off}/k_D \sim 10^{-3} - 10^{-5} \).

Here we focused on the ensemble averaged long-time diffusion constant \( D \). Since our system for small \( k_{ off} \) exhibits non-Markovian single-file behaviour (\( MSD \sim \sqrt{t} \)), one may ask whether our results hold also for time averaged MSD. For long times, \( t \gg \tau_{ off} \), \( \tau_{ off} \) is the average residence time of crowder particles, the dynamics is Markovian and the results should be the same. But for \( t \) close to \( \tau_{ off} \), transients may change \( D \). This will be especially problematic when motion occurs in disordered environments where logarithmically slow diffusion emerges [48]. The difference between time and ensemble average deserves attention and we leave it as an open problem.
The model we studied is inspired by protein diffusion on DNA. Our results are simple formulas for the diffusion constant of a tracer particle taking crowding and binding/unbinding dynamics into account. Although a protein is more complex than a hardcore particle, we hope that the simplicity of our results will be useful in a range of settings, in particular, in single-molecule tracking experiments.

Acknowledgments

LL acknowledges the Knut and Alice Wallenberg foundation and the Swedish Research Council (VR), grant no. 2012–4526, for financial support. TA is grateful to VR for funding (grant no. 2009–2924).

Appendix A. Simple model for \( D \) when \( k_{\text{off}} \) is small

In this appendix we outline the derivation for the long time diffusion constant \( D \) in the small \( k_{\text{off}} \) limit that led to equations (4)–(6). The main idea is to calculate the typical length \( l_0 \) that the tracer travels before being hindered by a crowder particle, the typical waiting time until a successful jumping event \( \tau \), and use equation (3).

First we address \( \tau \). Imagine that the tracer particle is flanked by two crowder particles, and the time for any of them to unbind is \( 1/2k_{\text{off}} \). Now, say the that the tracer’s right neighbour unbinds but the tracer tries to jump left anyway. This jump is forbidden, and so is in fact half of all tries the tracer makes. This implies that \( \tau \) is \( 1/k_{\text{off}} \) rather than \( 1/2k_{\text{off}} \). Moreover, we also consider the rebinding of crowder particles, so even if the tracer moves in the direction of the unbound neighbour it may be blocked anyway by another crowder particle. We must therefore correct the jump rate with the probability that the site is vacant, that is \( 1 - a\rho \) (in equilibrium). In summary, we estimate \( \tau \) as

\[
\tau = \frac{1}{k_{\text{off}} (1 - a\rho)},
\]

(A.1)

Secondly, we turn our attention to the coarse-grained lattice distance \( l_0 \). In short, we choose as \( l_0 \) the standard deviation of the distribution of nearest-neighbour distances. Here is how we formally arrive at this result. Since it can happen that nearest and next-to-nearest neighbours are unbound simultaneously, the length that the tracer particle can move, \( z \), can vary. We choose the probability distribution of \( z \), \( \varphi(z) \), to be the probability that there is a separation \( z \) between two nearest neighbour particles. The distribution of \( z \), \( \varphi(z) \), is known (see figure 7), but differs depending on the type of rebinding. If \( z \) can only change in discrete steps of \( \Delta \), that is \( \Delta, 2\Delta, 3\Delta \ldots \), we can define a jump length distribution \( g(l) \) for the tracer particle in the coarse-grained lattice as

\[
g(l) = \sum_{n=-\infty}^{\infty} \delta(l + n\Delta) \varphi(n\Delta),
\]

(A.2)

where \( \delta(x) \) is the Dirac delta function. Now we choose \( l_0 \) as the standard deviation of \( g(l) \), and from equation (A.2) one can show that

\[
l_0 = \Delta \sqrt{\langle n^2 \rangle_{\varphi}}
\]

(A.3)

Figure 7. Distribution of distances between nearest neighbours \( \varphi(z) \) in semi-log scale. Symbols depict simulation results, and dashed curves are theoretical results. The filling fraction is \( a\rho = 0.1 \). The remaining simulation details are the same as in figures 2 and 3.
where \( \langle n^2 \rangle_0 = \sum_{n=1}^{\infty} n^2 \varphi(n) \). In the subsections below, we calculate \( l_0 \) explicitly for the special cases: (1) crowder particles that do not diffuse and placed equidistantly; (2) crowder particles that do not diffuse and placed at a random distance apart from each other; and (3) diffusing crowder particles.

**A.1. Case 1: crowder particles that do not diffuse and placed equidistantly**

In the simulations, the crowder particles sit equidistantly, and unbind and rebind at the rates \( k_{\text{off}} \) and \( k_{\text{on}} \), respectively. In order to make sure that the average density \( \rho \) is constant over time, we choose \( k_{\text{on}} = k_{\text{off}} \), and work with \( 2m \) particles in the whole system (lattice + surrounding bulk). This means that \( m \) particles will, on average, be on the lattice and the density is \( \rho = m/(aN) \), where \( N \) is the number of lattice sites, and \( a \) the lattice constant. The smallest separation between two crowder particles in this setup becomes \( \Delta = 1/(2\rho) \), and increases in discrete steps of \( \Delta \):

\[
\frac{1}{2\rho} - \frac{1}{3\rho}, \frac{1}{2\rho}, \ldots
\]  

(A.4)

Each one of these lengths has a different probability, and the distribution of nearest neighbour distances is

\[
\varphi \left( \frac{n}{2\rho} \right) = \begin{cases} 
\frac{1}{2} \frac{1}{2^n}, & n = \pm 1, \pm 2, \ldots \\
0, & n = 0,
\end{cases}
\]  

(A.5)

which agrees well with simulations (figure 7). Using that \( \langle n^2 \rangle_0 = 6 \) in equation (A.3) gives

\[
l_0 = \frac{1}{\rho} \sqrt{\frac{3}{2}}.
\]  

(A.6)

**A.2. Case 2: crowder particles that do not diffuse and rebound to random locations**

In this case the crowder particles leave the lattice and return to a random vacant lattice site. This means that the smallest separation is the lattice distance of the original lattice, \( \Delta = a \), and distances are in steps of \( a \):

\[
a, 2a, 3a, \ldots
\]  

(A.7)

The inter-particle distance distribution in this case is

\[
\varphi(na) = \begin{cases} 
\frac{1}{2} \frac{1}{2^n} \frac{1 - ap}{1 - ap}, & n = \pm 1, \pm 2, \ldots \\
0, & n = 0
\end{cases}
\]  

(A.8)

which is corroborated by simulations in figure 7. In the continuum limit (small \( a \)), the distribution becomes exponential \( \varphi(na) \sim e^{-|n|a} \). Using that \( \langle n^2 \rangle_0 = (2 - ap)/(2a^2\rho^2) \), we obtain

\[
l_0 = \frac{1}{\rho} \sqrt{\frac{2 - ap}{2}}.
\]  

(A.9)

**A.3. Case 3: diffusing crowder particles with rebinding to random locations**

Here all particles diffuse, which drastically changes the situation. The main difference is that the tracer does not get stuck between two flanking road blocks since they also move. However, we know from simulations that the MSD for the tracer is in the long time limit proportional to \( D \tau \) (figure 5), which is a direct manifestation that the no-passing condition is violated (otherwise we would have had MSD \( \sim \sqrt{t} \)). Altogether, this implies that there is length scale for the coarse-grained lattice and a time scale associated with a jumping event.

For this case we cannot use \( \varphi(z) \) to estimate \( l_0 \) since \( \varphi(z) \) is the same as when crowder particles are not diffusing (see figures 7, \( \bigcirc \) and \( \bigtriangleup \)), and gives the wrong result for \( D \). The reason is that interparticle distances fluctuate at the same rate as the tracer is diffusing, and those fluctuations increase \( D \). In fact, even if \( k_{\text{off}} = 0 \), the tracer particles can still move across the system, although slowly. We estimate \( l_0 \) as the distance the tracer particle explores in a time \( \tau \), that is

\[
l_0 = \sqrt{\left\langle x^2(x = \tau) \right\rangle} = \left( \frac{4D}{k_{\text{off}}} \frac{1 - ap}{\rho^2} \right)^{1/4}.
\]  

(A.10)

Interestingly, \( l_0 \) depends on \( k_{\text{off}} \) and not only \( \rho \) as in the previous cases. This changes the scaling of \( k_{\text{off}} \) in \( D \) from linear in cases 1 and 2 to \( \sqrt{k_{\text{off}}} \) for this case. This can also be understood from the following simple argument. The curve for \( \left\langle x^2(t) \right\rangle \) is continuous for all times, and at some time the dynamics changes behaviour from single-file \( \sim \sqrt{t} \) to regular diffusion \( \sim t \). This occurs at around \( t \approx \tau \), which implies...
This yields $D \propto \sqrt{k_{\text{off}}}$. 

**Appendix B. Extraction of the long time diffusion constant**

The way we determine $D$ from our MSD simulations is illustrated in figure 8. First, $\tau_{\text{off}}$ is the approximate time at which the MSD becomes linear (shown as vertical dashed-dotted lines). Second, we make a linear regression of the MSD curve starting from that point, and obtain the slope which equals $D$. The resulting fits are shown as dashed lines.

**Appendix C. Numerical implementation**

The model (figure 1) is implemented using the Gillespie algorithm [49]. The majority of the details of the implementation have been explained elsewhere [38], but below we point out some key differences.

We keep track of the unbound crowder particles in the bulk in order to have the option to rebind them at the location they detached from. In practice we use two lattices, one of which represents the bulk. The filling fraction is maintained at the level we want by setting $k_{\text{on}} = k_{\text{off}}$, and then letting the systems equilibrate such that half the number of crowder particles sit in the bulk and the other half on the lattice we are interested in. This representation is helpful in investigating all sorts of binding modes especially rebinding to the same location. In [38] the bulk served as an infinite particle reservoir and the concentration on the lattice was tuned via detailed balance (rebinding always occurred to a randomly chosen site). Here, however, the bulk has a finite size and cannot be seen as a strict particle reservoir. However, since we use about 500 particles, fluctuations around the filling fraction $\rho_a$ are so small that we rarely (if ever) deplete the bulk. This means that we have approximately a grand canonical ensemble.

**References**

[1] Zhou H-X, Rivas G and Minton A P 2008 Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences *Annu. Rev. Biophys.* 37 375
[2] Zimmerman S B and Trach S O 1991 Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli* *J. Mol. Biol.* 222 599–620
[3] Ellis R J 2001 Macromolecular crowding: obvious but underappreciated *Trends. Biochem. Sci.* 26 597–604
[4] Li G W, Berg O G and Elf J 2009 Effects of macromolecular crowding and dna looping on gene regulation kinetics *Nat. Phys.* 5 294–7
[5] Morelli M J, Allen R J and Wolde P R T 2011 Effects of macromolecular crowding on genetic networks *Biophys. J.* 101 2882–91
[6] Nakano S-I, Karimata H T, Kitagawa Y and Sugimoto N 2009 Facilitation of rna enzyme activity in the molecular crowding media of cosolutes *J. Am. Chem. Soc.* 131 16881–8
[7] Zhou H-X 2004 Protein folding and binding in confined spaces and in crowded solutions *J. Mol. Recognit.* 17 368–75
[8] Martin J 2002 Requirement for groel/groes-dependent protein folding under nonpermissive conditions of macromolecular crowding *Biochemistry* 41 5035–5
[9] Höfling F and Franosch T 2013 Anomalous transport in the crowded world of biological cells *Rep. Prog. Phys.* 76 046602
[10] Knorr T and Templin M F 2009 Method of the year 2008 Nat. Methods 6 1
[11] Persson F, Barkefor I and Elf J 2013 Single molecule methods with applications in living cells Curr. Opin. Biotechn. 24 737–44
[12] Heller I, Sitters G, Broekmans O D, Farge G, Menges C, Wende W, Hell S W, Peterman E J G and Wuite G J L 2013 Sted nanoscopy combined with optical tweezers reveals protein dynamics on densely covered dna Nat. Methods 10 910–6
[13] Karlsson A, Karlsson R, Karlsson M, Cans A-S, Strömberg A, Ryttäinen F and Orwar O 2001 Molecular engineering: networks of nanotubes and containers Nature 409 130–2
[14] Dekker C 2007 Solid-state nanopores Nat. Nanotechnology 2 209–15
[15] Ando T and Skolnick J 2010 Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion Proc. Natl Acad. Sci. USA 107 18457–62
[16] Dix J A and Verkman A S 2008 Crowding effects on diffusion in solutions and cells. Annu. Rev. Biophys. 37 247–63
[17] Bénichou O, Loverdo C, Moreau M and Voituriez R 2011 Intermittent search strategies
[18] Szymanski J and Weiss M 2009 Elucidating the origin of anomalous diffusion in crowded media of non-inert obstacles J. Chem. Phys. 130 174902
[19] Ghosh S, Cherstvy A G and Metzler R 2015 Non-universal tracer diffusion in crowded media of non-inert obstacles Phys. Chem. Chem. Phys. 17 1847–58
[20] Saxton M J 1996 Anomalous diffusion due to binding: a monte carlo study Biophys. J. 70 1250–62
[21] Szymański J and Weiss M 2009 Elucidating the origin of anomalous diffusion in crowded fluids Phys. Rev. Lett. 103 038102
[22] Bénichou O, Loverdo C, Moreau M and Voituriez R 2011 Intermittent search strategies Rev. Mod. Phys. 83 81
[23] Marcovitz A and Levy Y 2013 Obstacles may facilitate and direct dna search by proteins Proc. Natl Acad. Sci. USA 110 1842–50
[24] Cho W-K, Jeong C, Kim D, Chang M, Song K-M, Hanne J, Ban C, Fishel R and Lee J-B 2013 Atp alters the diffusion mechanics of muts on mismatched dna Structure 21 1264–74
[25] Bénichou O, Cazabat A M, Lemarchand A, Moreau M and Oshanin G 1999 Biased diffusion in a one-dimensional adsorbed monolayer
[26] Bénichou O, Illien P, Oshanin G and Voituriez R 2013 Fluctuations and correlations of a driven tracer in a hard-core lattice gas J. Stat. Phys. 150 351–71
[27] Kunkel T A and Erie D A 2005 Dna mismatch repair Annu. Rev. Biochem. 74 681–710
[28] Sneppen K and Semsey S 2015 Mismatch repair at stop codons is directed independent of gatc methylation on the E.coli chromosome Proc. Natl Acad. Sci. USA 112 17422–7
[29] Luo Y, North J A, Rose S D and Potirier M G 2014 Nucleosomes accelerate transcription factor dissociation Nucleic Acids Res. 42 3017–27
[30] Gorman J and Greene E C 2008 Visualizing one-dimensional diffusion of proteins along dna Nat. Struct. Mol. Biol. 15 768–74
[31] Harris T E 1965 Diffusion with collisions between particles J. Appl. Probab. 2 323–38
[32] Percus J K 1974 Anomalous self-diffusion for one-dimensional hard cores Phys. Rev. A 9 557
[33] Lomholt M A, Lizana L and Ambjörnsson T 2011 Dissimilar bouncy walkers Phys. Rev. Lett. 107 038102
[34] Lomholt M A, Broek B V D, Kalisch S-M J, Wuite G J L and Metzler R 2009 Facilitated diffusion with dna coiling Proc. Natl Acad. Sci. USA 106 8204–8
[35] Haus J W and Kehr K W 1987 Diffusion in regular and disordered lattices Phys. Rep. 150 263–406
[36] Nakazato K and Kitahara K 1980 Site blocking effect in tracer diffusion on a lattice J. Phys. Chem. B 94 3410–6
[37] Bénichou O, Cazabat A M, Lemarchand A, Moreau M and Oshanin G 1999 Biased diffusion in a one-dimensional adsorbed monolayer J. Stat. Phys. 97 351–71
[38] Kolomeisky A B 2011 Physics of protein-dna interactions: mechanisms of facilitated target search Proc. Natl Acad. Sci. USA 108 20830–1
[39] Hu T, Yu Grosberg A and Shklovskii B I 2006 How proteins search for their specific sites on dna: the role of dna conformation Biophys. J. 90 2731–44
[40] Lomholt M A, Ambjörnsson T and Metzler R 2005 Optimal target search on a fast-folding polymer chain with volume exchange Phys. Rev. Lett. 95 260603
[41] Lomholt M A, Brock B V D, Kalisch S-M J, Wuite G J L and Metzler R 2009 Facilitated diffusion with dna coiling Proc. Natl Acad. Sci. USA 106 8204–8
[42] Haus J W and Kehr K W 1987 Diffusion in regular and disordered lattices Phys. Rep. 150 263–406
[43] Nakazato K and Kitahara K 1980 Site blocking effect in tracer diffusion on a lattice J. Phys. Chem. B 94 3410–6
[44] Bénichou O, Cazabat A M, Lemarchand A, Moreau M and Oshanin G 1999 Biased diffusion in a one-dimensional adsorbed monolayer J. Stat. Phys. 97 351–71
[45] Bénichou O, Illien P, Oshanin G and Voituriez R 2013 Fluctuations and correlations of a driven tracer in a hard-core lattice gas Phys. Rev. E 87 032164
[46] Elf J, Li G-W and Xie X S 2007 Probing transcription factor dynamics at the single-molecule level in a living cell Science 316 1191–4
[47] Lomholt M A, Lizana L, Metzler R and Ambjörnsson T 2013 Microscopic origin of the logarithmic time evolution of aging processes in complex systems Phys. Rev. Lett. 110 208301
[48] Gillespie D T 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions J. Comput. Phys. 22 403–34