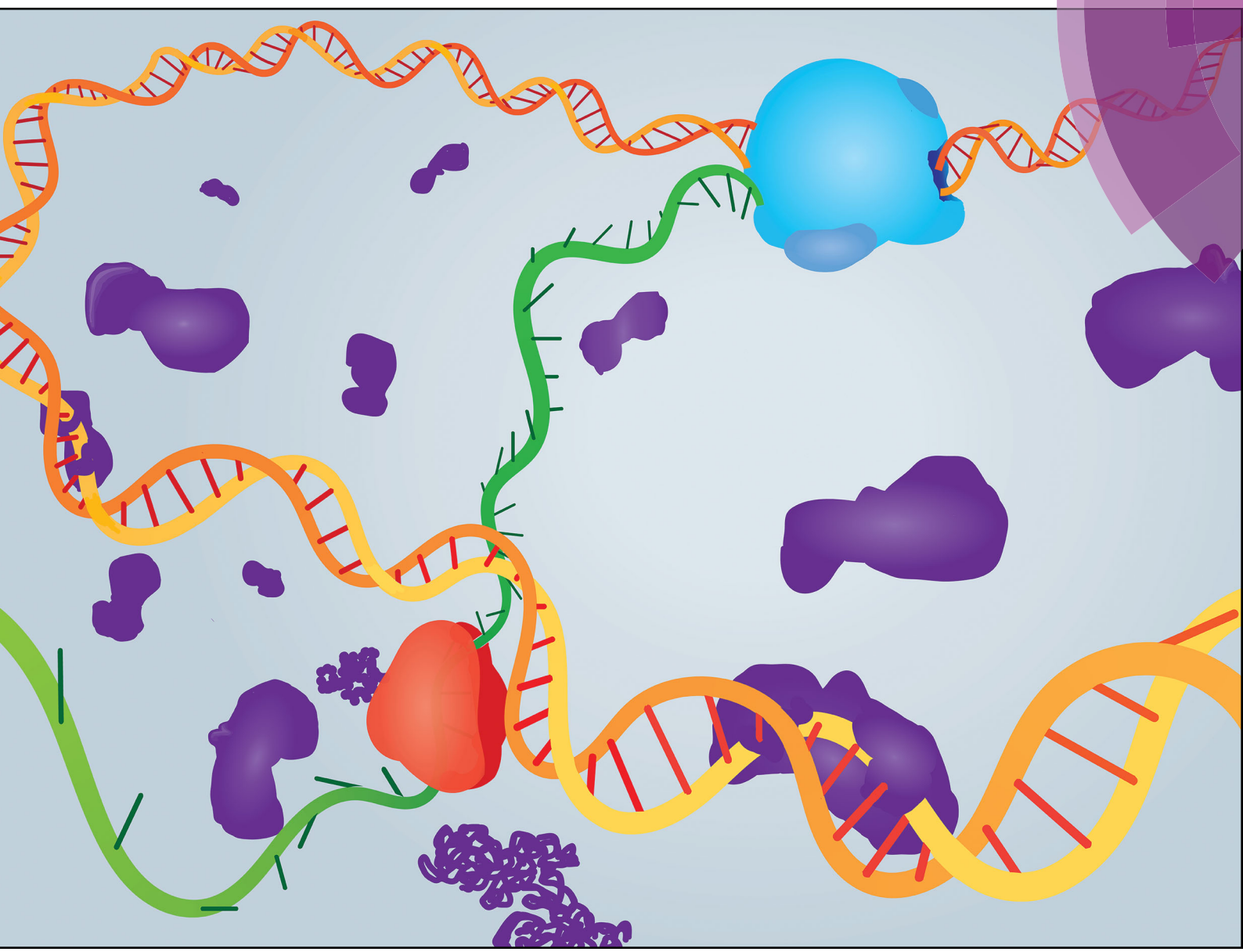
Volume 20 | Number 12 | 28 March 2018 | Pages 7899–8360

PCCP

Physical Chemistry Chemical Physics

rsc.li/pccp



ISSN 1463-9076

ROYAL SOCIETY OF CHEMISTRY

PAPER

Ralf Metzler *et al.*

Acceleration of bursty multiprotein target search kinetics on DNA by colocalisation

PCCP

PAPER

Cite this: *Phys. Chem. Chem. Phys.*, 2018, **20**, 7931

Acceleration of bursty multiprotein target search kinetics on DNA by colocalisation

Prathitha Kar,[ab] Andrey G. Cherstvy [b] and Ralf Metzler *[b]

Received 10th October 2017,
Accepted 20th December 2017

DOI: 10.1039/c7cp06922g

rsc.li/pccp

Proteins are capable of locating specific targets on DNA by employing a facilitated diffusion process with intermittent 1D and 3D search steps. Gene colocalisation and coregulation—*i.e.* the spatial proximity of two communicating genes—is one factor capable of accelerating the target search process along the DNA. We perform Monte Carlo computer simulations and demonstrate the benefits of gene colocalisation for minimising the search time in a model DNA–protein system. We use a simple diffusion model to mimic the search for targets by proteins, produced initially in bursts of multiple proteins and performing the first-passage search on the DNA chain. The behaviour of the mean first-passage times to the target is studied as a function of distance between the initial position of proteins and the DNA target position, as well as *versus* the concentration of proteins. We also examine the properties of bursty target search kinetics for varying physical–chemical protein–DNA binding affinity. Our findings underline the relevance of colocalisation of production and binding sites for protein search inside biological cells.

I. Introduction

A. Facilitated diffusion

Protein–DNA interactions and protein search for target (cognate) sequences on genomic DNA[1–13] control and tune a number of vital biological processes in bacterial and eukaryotic cells. One example of fast intracellular stochastic transport[14] is the search for promoters on DNA by transcription factor (TF) proteins, often required for transcription initiation or repression during gene expression. For instance, the lac repressor (LacI) binding to its operon site on the DNA realises a simple switch controlling the metabolism of lactose in *Escherichia coli* bacteria.[15]

DNA is a long polymeric molecule and TF proteins often have to scan it searching through $\sim 10^{6-9}$ nonspecific sites in order to locate and recognise the required cognate site, see Fig. 1. This search process often involves comparatively fast association rates.[9,13] For instance, for the lac repressor protein (at optimal *in vitro* conditions), the association rates with its target site can be up to $\sim 10^{10}$ M^{-1} s^{-1}.[1] This rate is $\eta \sim 10^{2-3}$ times faster than expected for a pure 3D diffusive motion of such a tracer particle in solution, denoted below as $1/\langle T_{\mathrm{3D\,only}} \rangle$ and obtained by the famed Smoluchowski theory[16,17] (see also the special issue[18] and ref. 10). The speedup with respect to the Smoluchowski limit comes about through switching between intermittent 3D diffusion in the bulk and 1D sliding diffusion along the DNA chain, providing the classical Berg–von Hippel facilitated diffusion mechanism.[2,3] The facilitated diffusion model and its variants quantitatively explain the association rates of several TF proteins with their DNA target sites, see ref. 3, 5, 6, 8, 11, 13 and 19–40.

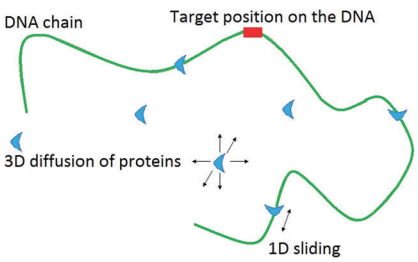


Fig. 1 Schematic view of the protein target search problem: the proteins are the blue symbols and the DNA target is the red rectangle. Two possible types of protein motions—bulk 3D diffusion and 1D sliding along the chain—are indicated by the arrows.

Faster search times are achieved by reducing the dimensionality of the process and by increasing local TF concentrations on the DNA. The double helix provides a nonspecific protein–DNA binding landscape,[6,7,34,41,42] with protein–DNA binding often being triggered by some close-range interactions,[10,11,43–48] see Fig. 2. The TF search and binding process was mapped onto the Michaelis–Menten reaction scheme in ref. 49.

When performing 1D diffusion along a nonspecific DNA fragment with a weaker binding energy E_b, the protein can detach from the double helix and return to 3D bulk diffusion. The protein requires N_R rounds of 3D and 1D diffusion in order

[a] *Dept of Inorganic and Physical Chemistry, Indian Institute of Science, Bengaluru, Karnataka 560012, India*
[b] *Institute for Physics & Astronomy, University of Potsdam, 14476 Potsdam-Golm, Germany. E-mail: rmetzler@uni-potsdam.de*

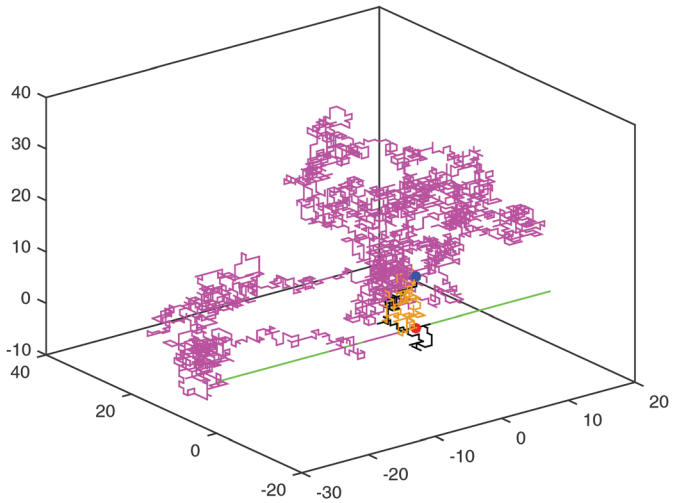


Fig. 2 Trajectories of three proteins (shown in magenta, orange and black) undergoing facilitated diffusion, as obtained from our simulations, before reaching the target on the DNA chain. The initial position of the proteins (0, 0, 10σ) is marked by the blue dot in the simulation box with the volume $V_\sigma = (150\sigma)^3$. The target is the red dot at position (0, 0, 0) on the green fragment denoting the DNA, which extends along the x axis from ($x_{\min} = -74\sigma$, 0, 0) to ($x_{\max} = 74\sigma$, 0, 0). The magenta trace, *e.g.*, meets the DNA chain twice, indicating the events of TF–DNA attachment and detachment before binding to the target site.

to reach the target. The associated mean search time $\langle T_{3D+1D}\rangle$ is given by[9,11,29,50]

$$\langle T_{3D+1D}\rangle = N_R(\langle t_{1D}\rangle + \langle t_{3D}\rangle). \quad (1)$$

Here, $\langle t_{1D}\rangle$ and $\langle t_{3D}\rangle$ are the mean 1D and 3D diffusion times per round of the target search process, respectively.[149] To accelerate the target search process, as follows from eqn (1), either the number of cycles N_R or the durations of 1D or 3D diffusion events need to be reduced. Recent single-molecule tracking experiments in living cells have quantified the time scales of target search of proteins on DNA to be in the range[12,13]

$$\langle \tau \rangle \sim 1\text{–}6 \text{ min} \quad (2)$$

(for one TF protein to find a single DNA binding site in one cell, see also Section III). Once the target site is recognised, some adaptation of the protein structure[48] and DNA deformations often take place, so that protein binding gets stronger to this DNA site.[7,28,29]

B. Transcription bursts and colocalisation

In recent years, the subject of stochasticity and noise in gene expression—both intrinsic and extrinsic[51–56]—and the respective regulation processes has attracted considerable attention from the biophysical community,[51–66] both from the experimental and theoretical perspective. Rapid developments of single-molecule techniques now enable the tracking of individual green fluorescent proteins[67] and monitoring gene expression "one protein molecule at a time"[64] (see ref. 53 for an overview of experimental techniques). A number of (regulatory) proteins have been shown to be produced in the form of transcriptional bursts,[55,56,64,68–72] in which the number of molecules produced in each burst follows an exponential distribution, while the number of bursts per cell cycle is often Poisson-like distributed and burst events in genetically identical cells appear to be uncorrelated, see ref. 53, 56, 62, 63, 64, 70 and 73, but also ref. 61 and 74.

Transcriptional bursting is conserved in all forms of life, from simple prokaryotes[61,64] to complex eukaryotic cells,[72,75,76] but its origin remains somewhat elusive (particularly for eukaryotic cells, with their more complicated transcription regulation mechanisms[61,65,77]). In bacteria, in the course of DNA transcription, positive DNA supercoiling is generated in front and respectively negative DNA supercoiling is built up behind the processive RNA Polymerase complex. This was proposed as one possible mechanism/contribution of the stochastic on-and-off switch for a given gene that can result in transcriptional bursts in bacteria.[61] For human cells, possible effects of chromatin decondensation and post-mitotic transcriptional spikes onto bursting were also suggested, see ref. 72.

In a pioneering study[64] Tsr-Venus fluorescent proteins were expressed in *E. coli* SX4-strain cells *via* replacing the native *lacZ* gene. The protein bursts were shown to occur randomly and were uncorrelated with time, with about four proteins on average per bursting event being translated by the ribosomes from each stochastically-transcribed molecule of mRNA.[64] On average, one burst per cell cycle of *E. coli* occurs, but rather long tails in the distributions of the number of proteins per burst and in the number of bursts per cell cycle exist.[64] This indicates inherent stochasticity and variability in the protein expression,[51–55,61,71,73,78–80] both on the molecular and cellular level; see also ref. 81 for effects of ageing in stochastic gene expression. The average lifetime of degradable bacterial mRNAs (of $\sim$1.5 min, with exponential distribution,[70] see also ref. 56, 77, 78 and 82) defines the typical duration of a protein bursting event. The latter often results in a low copy-number statistic for important proteins in bacterial cells and thus large local concentration fluctuations for transcripts and proteins. Note that the abundances of mRNAs and proteins produced from them are not necessarily strictly correlated.[53,62]

After their production, protein molecules can undergo facilitated diffusion to find their specific DNA target sites.[2,3,6,10,11] Here, a number of factors can contribute to reducing the target search time and thus to more efficient cell functioning. The list includes the actual DNA base pair sequence,[6,7,34,83] electrostatic protein–DNA interactions,[10,28,84] effects of long jumps between distant DNA segments,[10,13,43,85,86] and finally some implications of gene colocalisation (GC).[9,29,87–89] The latter is the main subject of the current study. From the bioinformatic perspective, novel data analysis tools also enable one to uncover multiple aspects of the physical organisation of genomes *via* identifying colocalisation patterns for functionally related genes, see *e.g.* ref. 90. The analysis of significant gene clustering can be performed for multiple species, providing statistical information on the distances between the genes and histograms of cluster sizes.

Since the early studies,[91] the lac repressor binding to its DNA targets has become the canonical system to study the TF-mediated control of gene transcription. In bacteria,

genes producing proteins that later bind to specific DNA targets and the target positions themselves are often colocalised.[9,88] Shorter distances $\mathcal{D}$ from burst position to the target not only reduce the search time, but also increase the efficiency of gene regulation and the precision of response to external stimuli.[89,92] Diffusive encounters and local protein concentration enhancements nonspecifically regulate gene expression in bacteria; for other regulation strategies in pro- and eukaryotes, we refer the reader to ref. 77. For instance, CAP proteins strongly enhance the rate of transcription by RNA polymerase *via* activation, enhancing local protein concentrations in bacteria on the length scale of up to $\sim 10^3$ DNA base pairs.[77] Note here that in a highly compartmentalised eukaryotic cell, the DNA transcription takes place inside the nucleus, where much higher concentrations of TF proteins can locally be established and maintained.[77]

Below, we focus on the effects of GC in the multiple-copy TF target search on DNA, examining first the implications of the initial position of burst proteins from their target on the DNA. Moreover, GC can markedly increase the concentration of TFs produced in bursts near the promoter sites[77] for short distances to their regulated transcription unit,[9,63] thus accelerating the regulation processes. The implications of GC and stochasticity in gene regulation were considered within advanced mathematical models,[63,74] which served as a starting point for the current study. The detailed mathematical analysis of fluctuations in TF–DNA binding[49] and the few-encounter limit for the target search kinetics[93,94] related to strong irreproducibility of first-passage times[95,96] were also proposed recently. Lastly, the models of Berg–Purcell type describing molecular signalling and precision control in the low-concentration limit are to be mentioned here too.[82,97–101]

The paper is organised as follows. We introduce a simple picture of facilitated diffusion of TF proteins in our model cell in Section II. We provide the values of all the parameters involved in the model and establish a connection to the known biological facts and findings. The results of our computer simulations of the facilitated diffusion dynamics are presented in Section III. Most importantly, we examine the effects of protein burst position in relation to the DNA target and higher concentrations of TFs near the target.[77] We discuss the main conclusions and possible future directions of model development in Section IV.

II. Model and approximations

Our model biological cell—or *e.g.*, a compartment occupied by the *E. coli* chromosome[77]—is considered as a discrete cubic lattice with lattice constant σ, with edge length 150σ and reflecting boundaries. The latter mimic the lipid membranes in real biological cells and keep the proteins inside the simulation box. The DNA is modelled as a straight collection of beads, extending along the x-axis and encompassing the entire range of the lattice points, from $(-74\sigma, 0, 0)$ to $(74\sigma, 0, 0)$. The target can be positioned randomly on this chain, see the red dot in Fig. 2. The protein is modelled as a tracer particle occupying a single lattice site. As the concentration of proteins in the box,

$$c_{\mathrm{p}} = N_{\mathrm{p}}/V_{\sigma}, \tag{3}$$

is varied—N_{p} here is the number of proteins and V_{σ} is the cube volume—the average time for the first of them to locate the target on the DNA is calculated as the number of simulation steps n needed for steps in the 3D solution and during sliding events along the 1D DNA chain.

The simulation process starts with the "production" of the first protein at time $t = 0$ (or time step $n = 0$) at a random lattice site (the initial position). Such randomly chosen (burst position)-to-target distances $\mathcal{D}$ are responsible for (biologically expected) random distributions of protein transcripts with respect to the specific DNA site they target. The distance $\mathcal{D}$ is defined *via* the radial separation R of proteins to the DNA and target shift distance along the DNA $x_{\mathrm{p\text{-}t}}$ with respect to the protein production site, which is $\mathcal{D}^2 = R^2 + {x_{\mathrm{p\text{-}t}}}^2$. Further proteins are produced in the simulations at later times in the form of expression bursts. The times of protein production are exponentially distributed,

$$P(n) = 1 - \exp[-n/\delta n], \tag{4}$$

with $\delta n = 4$ steps. This means that in simulations, on average, one protein every four steps is created in the bursts, see Fig. 3. The burst times are thus much shorter than the average protein-target first-passage times in the model, as we confirm below. We expect that altering the burst frequency has a rather small effect on the number of steps required for a protein to reach the DNA target. To the best of our knowledge, there is no clear experimental data on the temporal distribution of proteins in a single burst event, due partly to a fairly slow resolution in the experiments.

The basic quantities presented in the main part are averaged over $M \sim 10^3$ realisations of protein bursts. All bursts are identical in terms of distributions of times at which individual proteins are created. The averaging thus accounts only for stochasticity of the subsequent protein diffusion processes, and does not take into account possible randomness of the burst

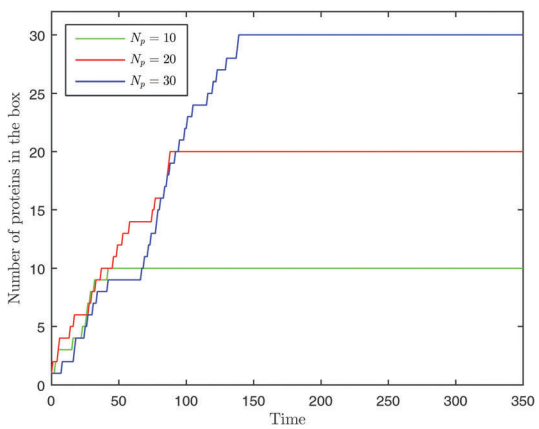


Fig. 3 Time evolution of the number of proteins created in a burst, for the protein number $N_{\mathrm{p}} = 10$, 20, 30 and $\delta n = 4$ in eqn (4).

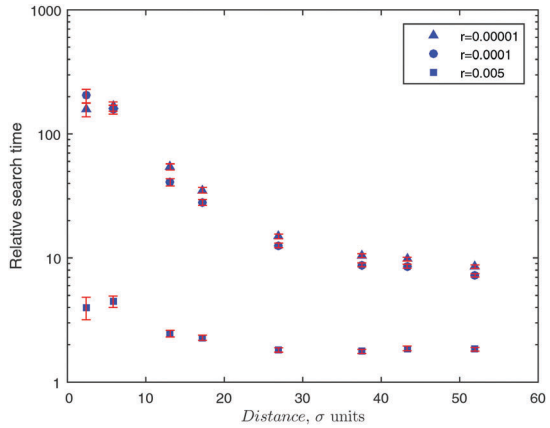


Fig. 4 Ratio η of search times (10) *versus* randomly chosen distance $\mathcal{D}$ of the starting position of protein bursts from the DNA target. The target is fixed at (0, 0, 0) and $N_p = 20$ proteins are simulated. The values of unbinding probabilities r are indicated in the legend, corresponding to $E_b/(k_BT) \approx 11.5$, 9.2, and 5.3, respectively. Error bars are also shown.

structure itself. This simulation procedure therefore neglects (for simplicity) possible intrinsic noise effects in gene expression inside real biological cells and cell-to-cell variability.[51,52,54]

For protein production in bursts—see the results starting from Fig. 4—after the first protein is produced at the start of a burst, the molecules start diluting away from the production site. To initially reach a fixed concentration c_p, a finite number of simulation steps is needed, $n_{burst} \approx N_p \times \delta n$, due to the Poissonian production scenario, see the burst shapes in Fig. 3. For a larger overall number N_p of proteins in the burst, the saturation plateau is reached at longer times. No production or degradation of proteins after this moment occurs (see *e.g.* ref. 68, 74, 78, 79 and 102 for models with degradation of mRNA and protein molecules).

After the first protein in the burst binds to the target upon first encounter, *i.e.* we assume immediate binding, the corresponding first-passage time is enumerated.[150] After this, all proteins are removed from the system and introduced again at the same start position. The clock is set to zero and the proteins are recreated *via* the identical burst. This procedure is repeated M times to compute the mean first-passage time to the target. The protein starting conditions are thus out of equilibrium so that the applications of our findings go beyond standard equilibrium models of facilitated protein diffusion on DNA.[10,11,29]

The starting burst in our simulations initiates the DNA target search process, supported by the fact that the average number of bursts per life cycle in bacteria is often about unity.[64] A given lattice point can only be occupied by a single protein molecule (excluded volume principle) that has equal probabilities of jumping to the six nearest-neighbour sites when performing the 3D diffusion process. If a protein reaches a box boundary, it gets reflected back. When a protein binds to the DNA chain for the first time, it starts a 1D diffusion process along it, modelled as standard Brownian motion as well. Note here that the effects of anomalous bulk diffusion[103–107] of proteins, with the mean squared displacement $\langle \mathbf{r}^2(n) \rangle \simeq n^\alpha$ where $\alpha \neq 1$, on the target search kinetics was recently studied.[50] However, as at this point it is not clear how transient the anomalous diffusion is,[67,108,109] we here use the Brownian scenario with $\alpha = 1$. The protein jumps to one of the two adjacent DNA lattice sites with probability p or unbinds from the chain with detachment probability r. The latter scales as the Boltzmann factor

$$r \simeq \exp[-E_b/(k_BT)] \quad (5)$$

with the energy $E_b > 0$ of nonspecific protein–DNA binding. According to the probability conservation rule, we require $2p + r = 1$. In Fig. 2, the trajectories of three different proteins from our simulations are shown.

While walking along the DNA, the diffusivity of the protein can differ strongly from the value in the bulk solution.[11,50] Due to attractive protein–DNA interactions, a sequence-specific protein–DNA binding landscape,[6,7,28,34,39,41,110] conformational changes in the protein accompanying its DNA binding,[43,48] and helical (rather than straight) protein motions along the DNA,[111] the mobility of TFs and other DNA bound proteins along the 1D DNA is often much slower than for 3D diffusion, that is $D_{1D} \ll D_{3D}$.[10–12,35] To introduce this important aspect into the model, the proteins bound to the DNA are not updated at each step in the simulations, unlike the ones undergoing 3D diffusion. The bound proteins wait n_w steps while other proteins are being updated. The number of waiting steps n_w depends on the ratio of 1D and 3D diffusion constants for a realistic scenario. In Section III we calibrate the value of n_w *versus* D_{1D}/D_{3D}, see Fig. 13. Thus, the conservation law gets modified as $2p + r + w = 1$, where w is the probability that the bound proteins wait and do not move on the DNA.

The value of $r = 10^{-4}$ we use for most of the results presented below corresponds to a protein–DNA binding energy of $E_b \approx 9.2 k_BT$. This is a typical value for a nonspecific TF binding[4,28,29,34,112] to ~10–20 DNA base pairs.[48,84] Weaker binding energies will result in a larger probability of TF unbinding from the DNA track, r. Clearly, different values of the parameters r and n_w will alter the 1D sliding distances of the proteins and affect the target search times. In *in vitro* experiments, r can be systematically controlled by *e.g.* varying the salt concentration in the bulk solution.[1,3,85]

The electrostatic attraction of positive patches on the protein surface to the negatively charged DNA phosphate groups, which is screened by surrounding electrolyte, creates an energetic funnel with a thickness of about one Debye screening length along the DNA.[10,28,113] In the cytoplasm-like electrolyte, the ionic strength is ≈0.15 M rendering the Debye length ≈7 Å. This length is shorter than the DNA radius and can thus be included into an effective binding constant acting upon a direct contact of TFs to the DNA surface. This binding is assumed to be insensitive to the underlying DNA sequence regulating the overall fraction of DNA-bound *versus* free proteins. Our simplistic model also neglects the effects of sequence-specific TF–DNA affinities and the existence of rather broad energetic funnels for 1D sliding of TF proteins along the DNA in the vicinity of their respective binding sites, see the detailed recent

analysis of ref. 114. The corrugated energetic landscape[6,7,34] and localised energetic funnels for protein–DNA binding[28,114] can further facilitate facilitated diffusion of TFs towards their cognate sites, features to be examined in a separate future publication.

We now relate the box dimensions and protein concentrations c_p to the respective values in a bacterial cell. The value of σ is obtained below by equating c_p in the model to that inside a typical *E. coli* cell. The average number of molecules for a specific TF in an *E. coli* cell is $N_p \sim 10$[9,11,77,82,115] corresponding to intracellular concentrations of about 10 nM (or even less[12]) and the cell volume is $V \approx 2\ \mu m^3$.[77,116] We simulate $N_p = 20$ tracers in a box of volume $V_\sigma = (150\sigma)^3$. This yields $\sigma = \left(\frac{20\times 2}{10}\right)^{1/3}\frac{1}{150}\,\mu m \approx 10.6$ nm for the elementary length scale, or about 30 DNA base pairs per bead. Also note that the size of the simulation box sets the effective DNA concentration, which is a key parameter in search problems.[28,85,86,117] In bacteria, the concentration of nonspecific DNA sites is $\sim$0.01 M,[118] while a small TF protein recognises $\sim$10–20 unique DNA base pairs as its cognate site.[28,77,84]

Our goal is to quantify to what extent the phenomenon of GC can accelerate the target search process by multiple proteins created in bursts. The model parameters we vary are the burst-to-target distance $\mathcal{D}$ and the protein concentration c_p, whereas the values of r and n_w are mostly kept constant.

III. Main results

A. Calibration of the model parameters

To calibrate the value of n_w required to fix a particular D_{1D}/D_{3D} ratio close to that in experiments, we perform simulations to evaluate the mean squared displacement (MSD) for free 3D diffusion of the tracer particle in a cubic box with side length $L = 150\sigma$ and free 1D diffusion along the 1D DNA chain. To extract the diffusion constants D_{1D} and D_{3D}, we use

$$\langle \mathbf{r}^2(n)\rangle = 6D_{3D}n \text{ and } \langle x^2(n)\rangle = 2D_{1D}n \tag{6}$$

for 3D and 1D diffusion after n steps. The MSD growth for 3D and 1D diffusion is linear, as expected for Brownian motion, see Fig. 11 and 12 where the particle displacements are shown in terms of σ^2. The values of D_{3D} and D_{1D} follow from fitting the MSD. Namely, from eqn (6) we get $D_{3D} = \text{Slope}_{3D}/6 \approx 0.17$ and from the MSD growth in the 1D case for $n_w = 50$ we get $D_{1D} = \text{Slope}_{1D}/2 \approx 0.01$ (note that D is given in terms of $\sigma^2/\Delta n$, where Δn measures one simulation step). The MSDs of both 1D and 3D free protein diffusion reveal a levelling off for large numbers of simulations steps, because of the finite size of the simulation box, see Fig. 11.[151] The average sliding length of protein molecules[9] on the 1D DNA substrate can then be expressed as

$$\langle l_{sl}\rangle \approx \sqrt{2D_{1D}\langle t_{1D}\rangle}, \tag{7}$$

where $\langle t_{1D}\rangle$ is given by the inverse unbinding rate.

Fig. 12 demonstrates the expected decrease in D_{1D} with increasing n_w value. For calibrating n_w with the ratio D_{1D}/D_{3D}, one can extract from Fig. 13 the relation $D_{1D}/D_{3D} \sim 1/n_w$ expected for Brownian diffusion, see Fig. 12. From this, for a given D_{1D}/D_{3D} ratio, the value of n_w can be predicted. Experimentally, the value of D_{1D}/D_{3D} varies over a broad range, $D_{1D}/D_{3D} \sim 10^{-1}$–$10^{-3}$, depending, *i.e.*, on the binding strength and size of the protein molecules, external solution conditions, and DNA base pair composition.[6,7,11,89,103,114,119] For most of the results presented below, $D_{1D}/D_{3D} \approx 0.06$ corresponding to $n_w = 50$ waiting steps in the simulations of 1D diffusion.

From the experimental evidence, the *in vivo* diffusivities of the lac repressor proteins in living *E. coli* cells were measured to be[12] $D_{3D} \approx 3 \pm 0.3\ \mu m^2\ s^{-1}$ by fluorescence correlation spectroscopy and

$$D_{1D} \approx 0.046 \pm 0.01\ \mu m^2\ s^{-1} \tag{8}$$

by single-molecule tracking. The 1D sliding length at moderate to high salinities inside the cytoplasm of a living *E. coli* cell was quantified to be $\langle l_{sl}\rangle \approx 45 \pm 10$ base pairs,[13] while in the low-salt *in vitro* setups, the lengths of protein sliding can be dramatically longer[3,10] (see also ref. 29 and 89 for the theoretical estimate of $\langle l_{sl}\rangle$ and $D_{3D,1D}$ as well as for their functional dependencies on the parameters). Comparing the protein 1D diffusivities in our computer simulations and in single-particle tracking experiments,[13] using the elementary length scale σ, we find that the unit time step in the simulations corresponds to

$$\delta\tau \sim 0.02\ \text{ms}. \tag{9}$$

We note that further evidence of facilitated association comes from single-stranded DNA binding proteins, such as g32 proteins of the T4 bacteriophage, which are known to destabilise DNA secondary structure.[120] The contribution of the 1D and 3D diffusive terms at varying concentration c_p of these proteins was examined.[23,47,120] Note that the 1D diffusion of g32 proteins with[120] $D_{1D} \sim 0.1$–$1\ \mu m^2\ s^{-1}$ is considerably faster than the 1D motion of the lac repressor, see eqn (8).

B. Facilitated diffusion with protein bursts and colocalisation effects

Having calibrated the parameters, we now study the facilitated search by TFs and quantify GC effects. First, we confirm that the current model results in a facilitated target search due to a combination of 3D and 1D diffusion. In Fig. 4, we show the average target search time $\langle T_{3D\,only}\rangle$—the mean first-passage time of a first protein from a burst of $N_p = 20$ proteins—*via* pure 3D diffusion (the Smoluchowski limit[17]) normalised to the total time of combined 3D and 1D diffusion, namely

$$\eta = \langle T_{3D\,only}\rangle/\langle T_{3D+1D}\rangle. \tag{10}$$

The value of η can be well above unity for a strongly facilitated search. The ratio of search times is particularly large for short distances $\mathcal{D}$ from the DNA, as expected. Namely, when the proteins are produced rather close to the DNA, the probability for one of them to bind to the 1D track after a given number of steps is higher. Thus, the target will be found faster *via* a quick 3D diffusion followed by sliding (the unbinding is very improbable, $r \ll 1$) and the ratio of search times (10) is very large.

In Fig. 4 we present for comparison the results for the acceleration factor η obtained for other unbinding probabilities r, connected *via* eqn (5) to the protein–DNA affinity E_b. We find that for weaker protein–DNA binding (larger r values), the acceleration of the target search times η decreases significantly, as expected. Ultimately, as protein unbinding probability from the DNA reduces even further (the case of $E_b \to 0$), the 1D search route is used only rarely and almost no acceleration in the target search kinetics is detected in the simulations (results not shown). We thus confirm that—for our simulation setup and other model parameters fixed—with a protein–DNA binding strength $E_b \sim 10k_BT$, the facilitated diffusion mechanism accelerates the target search by $\eta \sim 10^2$ times, consistent with experiments.[2,3,11,13] The strength of the protein–DNA binding not only controls the degree η of search speedup, but also the equilibrium fraction of nonspecifically bound proteins.[9–11]

Next, we study the role of the protein concentration c_p in enhancing the target search rate. We keep the initial position of proteins and target constant, while varying c_p. Here, the proteins are created by translational bursts at the initial distances at the start of the simulation, until a certain number of molecules is produced and the required c_p level is reached. For each c_p value and protein starting distance $\mathcal{D}$, the average number of simulation steps to reach the target is then evaluated, as shown in Fig. 5. As expected, the number of simulations steps to reach the target decreases monotonically and, in fact, quite dramatically with the protein concentration. Note that for the relatively small TF concentrations c_p, the protein diffusion is hardly affected by excluded volume effects (the molecules approximately behave as independent walkers).

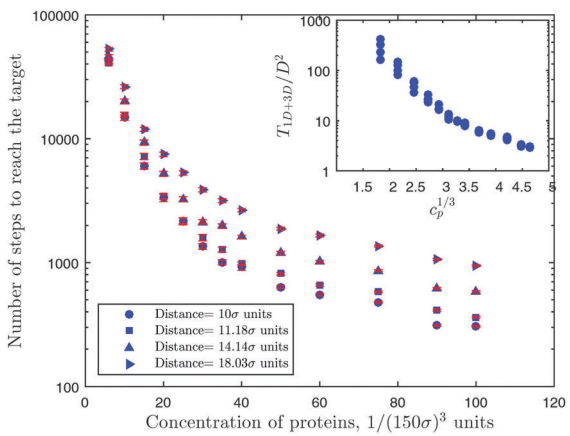


Fig. 5 Number of simulation steps required for the first protein to reach the DNA target $\langle T_{1D+3D}\rangle$ *versus* protein concentration c_p, for different distances $\mathcal{D}$ from the protein production site to the DNA target and for $r = 10^{-4}$. The error bars—shown in red colour, often smaller than the symbol size—are computed from $M \sim 2000$ realisations for each c_p value. The radial distance of protein burst position from the DNA is fixed at $R = 10\sigma$ in this plot, while the total distance $\mathcal{D}$ is varied (see the legend) *via* changing the target position, setting it at ($x_{p\text{-}t}$ = 0, 0, 0), ($x_{p\text{-}t} = 5\sigma$, 0, 0), ($x_{p\text{-}t} = 10\sigma$, 0, 0), ($x_{p\text{-}t} = 15\sigma$, 0, 0), for the curves from bottom to top, correspondingly. The inset shows the same data in rescaled coordinates, $\langle T_{3D+1D}\rangle/\mathcal{D}^2$ *versus* $c_p^{1/3}$ (see text for details).

At high protein concentrations, however, the average number of simulation steps to reach the target (the target search time) starts to saturate, see Fig. 5, partly because the contribution of the 1D chain is reduced and there are always proteins in the solution close to the target.[11,28] In simulations at large c_p, the proteins may visit a large fraction of the lattice sites before one of them hits the target. We refer the reader to a previous study[121] for the theory of multiple walkers on DNA and their space coverage. Thus, even higher concentrations c_p cannot perform better than visiting all sites: the target is therefore not found proportionally faster at higher c_p levels. From Fig. 5, we conclude that keeping the separation $\mathcal{D}$ between the protein production site and target fixed, the effect of the protein concentration itself is sufficient to get a faster search, possibly improving the transcription efficiency.[89]

Notably, we observe a decrease of the search times at fixed protein concentration with decreasing distances $\mathcal{D}$, Fig. 5. Moreover, for smaller protein concentrations, the discrepancy between the target search times computed for different distances $\mathcal{D}$ becomes smaller, as expected. Namely, at low c_p the length of the initial 3D excursion before finding the DNA chain can be very long, so that the effect of different distances $\mathcal{D}$ on the target search time gets reduced. The separations listed in Fig. 5 are the diagonal distances $\mathcal{D}$ from the protein production site to the DNA target. The radial distance of the burst site to the target in Fig. 5 was fixed at $R = 10\sigma$.

The average target search times at different distances $\mathcal{D}$ and varying protein concentrations c_p can be presented in rescaled coordinates $\langle T_{3D+1D}\rangle/\mathcal{D}^2$ *versus* $c_p^{1/3}$ to yield an approximate master curve, as demonstrated in the inset of Fig. 5. The rescaling of search times is based on the functional properties of the first-passage time distribution (see Section 3.2.2. of ref. 122), while for the other axis $\sim 1/c_p^{1/3}$ is the average separation between the proteins at a given c_p. The reader is also referred to a recent study[123] for first-passage time calculations in cylindrical geometry, applicable, *e.g.*, to the problem of reaching the bacterial nucleoid by bio-molecules that first penetrate through the outer cell membrane.

The data of Fig. 5 are shown on a log–log scale in Fig. 6a. We find that in the limit of low protein concentrations, the slope of the search time dependence varies with distance $\mathcal{D}$. The exponent for short $\mathcal{D}$ is found to be close to 2 so that

$$\langle T_{3D+1D}(c_p)\rangle \sim c_p^{-2}. \quad (11)$$

This type of scaling in the low-c_p limit agrees with the theoretical predictions for 1D-dominated search, see the findings of ref. 23. Note that in the volume-dominated search mode, the first-passage time is the inverse search rate that grows linearly with c_p. Under these conditions, due to the fact that a single protein searches a respectively shorter stretch on the DNA, one can write $\langle l_{sl}\rangle^2 \sim (L/N_b)^2 \sim D_{1D}T$, which yields[23]

$$\langle T_{3D+1D}(N_b)\rangle \sim N_b^{-2}. \quad (12)$$

In our case, however, the number of DNA-bound proteins N_b does not scale linearly with c_p, see the inset of Fig. 6a plotted

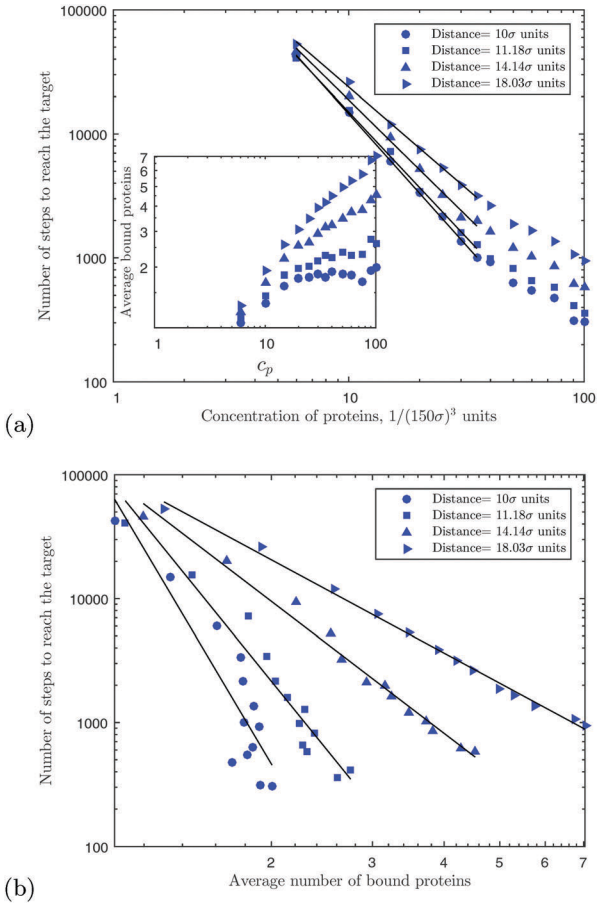


Fig. 6 (a) The data of Fig. 5 on log–log scale ($R = 10\sigma$ = const.). The inset shows the number of bound proteins N_b *versus* the bulk protein concentration. With increasing distance $\mathcal{D}$, as shown in panel (a), the fitting slopes of the relation $\langle T_{3D+1D}\rangle \sim c_p^{\gamma}$ computed at low c_p values are $\gamma \approx -2.1$, -2.0, -1.9, and -1.6, compare to eqn (11). (b) The same data in the log–log scale plotted *versus* the average number N_b of bound proteins. The slopes in the relation $\langle T_{3D+1D}\rangle \sim N_b^{\nu}$ generalising eqn (12) and shown in panel (b) for increasing $\mathcal{D}$ values are $\nu \approx -7.8$, -5.7, -3.5, and -2.5.

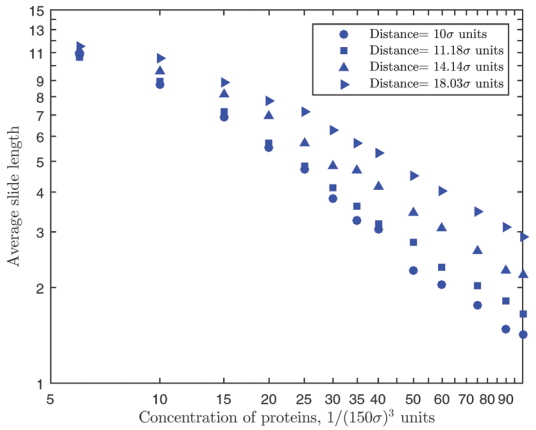


Fig. 7 Variation of the average protein sliding length $\langle l_{sl}\rangle$ (in units of σ) *versus* protein concentration c_p, computed from the data sets of Fig. 5 and 6.

for the fixed radial distance R and varying total separation $\mathcal{D}$. Namely, for shorter burst position to target distances, the number of bound proteins N_b saturates as a function of c_p. In contrast, for larger $\mathcal{D}$ values, the $N_b(c_p)$ dependence is a growing function of c_p, and on average more proteins are associated with the DNA at larger c_p. The reason is that for the burst position just above the target (*e.g.* at $R = \mathcal{D} = 10\sigma$), a rather small number of proteins quickly finds the DNA and the target on it, as Fig. 5 and 6 support. In contrast, when the protein burst site and the target are additionally separated along the DNA chain by a substantial distance $x_{p\text{-}t}$, more proteins bind to the DNA on average, and they spend a considerably longer time in the sliding mode before locating the target. This is an important effect of varying initial separations $\mathcal{D}$ on N_b.

The average number of bound proteins N_b and their average 1D sliding length (presented in Fig. 7) thus both grow in this case, as we show. The exponent of $N_b(c_p)$ growth is however considerably smaller than unity, which is expected for the standard Langmuir-type adsorption isotherm with $N_b \sim c_p$. In Fig. 6b, the search time plotted *versus* N_b also reveals the "saturation effect" at short $\mathcal{D}$, due to a nontrivial $N_b(c_p)$ dependence, so that the equilibrium assumptions of ref. 23 do not apply here.

As compared to our system, finite size and TF number effects modify the expected and measured behaviour for large *in vitro* systems with a large volume and protein number buffer. For the data of Fig. 5, the average sliding length of proteins was evaluated using the times of protein diffusion in 1D, eqn (7). Note that for bursty proteins that bind to and slide along the DNA chain, the times t_{1D} include both the particles that found the target as well as unsuccessful or unfinished protein runs. We remind the reader here that the system of proteins is fully reset after the first protein finds the DNA target, but in the computations of the mean sliding length in Fig. 7, we used the times of all protein paths along the DNA chain encountered in simulations.

Accordingly, at low protein concentrations c_p, the sliding lengths for different burst position to DNA target distances $\mathcal{D}$ do not differ much, see Fig. 7. In contrast, for relatively high c_p, the average sliding distances for larger $\mathcal{D}$ values are measurably longer than for $\mathcal{D} = 10\sigma$ (the protein production site is just on top of the target, at the radial distance of $R = 10\sigma$). The longer mean sliding lengths at longer distances $\mathcal{D}$ are, at least partly, responsible for a larger instantaneous number of DNA-bound proteins, as shown in the inset of Fig. 6a.

In the findings above, we studied the mean target search times for an ensemble of N_p bursty proteins searching for the DNA target *via* diffusion. The distribution of individual first-passage events, $p(\langle T_{3D+1D}\rangle)$, is also interesting to quantify. For different separations $\mathcal{D}$ from the protein production site to the target, the results for the normalised distribution $p(T_{3D+1D})$ are shown in Fig. 8. We find that for relatively short distances $\mathcal{D}$, the distribution has a pronounced power-law tail at intermediate times, in agreement with theoretical predictions.[96] For very long distances $\mathcal{D}$, in contrast, $p(T_{3D+1D})$ is often narrower and it features a quite fast decay of tails for times much shorter and much longer than the most probable search time, at the maximum of the distribution. Physically, for the conditions of a

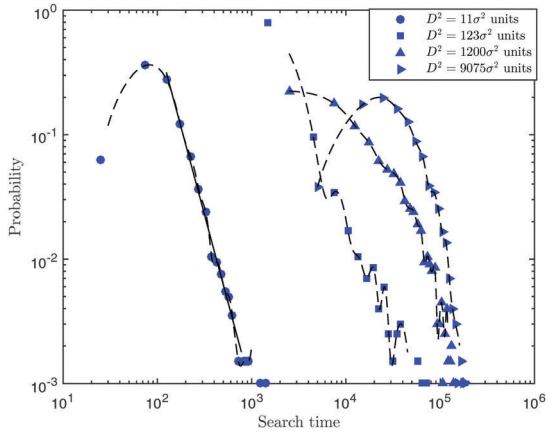


Fig. 8 Distribution of individual target search times, $p(T_{3D+1D})$, evaluated from the data of Fig. 5 for four burst-to-target separations, $\mathcal{D}$, and presented on a log–log scale. The long-time asymptote for the shortest distance $\mathcal{D}$ has the slope of $-(1 + \mu) \approx -2.83$, see eqn (13) in the text.

rather close protein's starting point and DNA target, the direct protein trajectories towards the target are responsible for the maximum of the distribution at relatively short times. The position of the maximum is thus geometry-dominated by the initial burst-to-target diffusion. On the other hand, the particles walking initially away from the target—and locating the DNA chain with the binding site on it after the second, third, *etc.* approach—are responsible for an extended power-law tail of $p(T_{3D+1D})$ function, as we observe in Fig. 8 and Fig. 14.

As we illustrate in Fig. 14, for longer protein start-to-target distances $\mathcal{D}$, the long-time tail of the distribution $p(T_{3D+1D})$ has a distinct exponential shape. These features are consistent with the theoretical findings for the target search kinetics by Brownian walkers in bounded 2D/3D domains with reflecting and absorbing boundaries[96] (for the domain boundary and target, respectively). The functional form proposed there for the first-passage times distribution is[96,124]

$$p(T) \sim \exp[-A/T]1/T^{1+\mu} \exp[-T/B]. \quad (13)$$

The extent of the intermediate power-law region was shown[96] to become larger for closely positioned walker starting positions and targets, while the two exponential functions above describe the fast decays for very short and very long search times. The existence of power-law tails results in pronounced trajectory-to-trajectory fluctuations and, as a consequence, large uncertainties in determining the mean first-passage time.[95,96,124] We do not present here any further quantitative analysis of the first-passage distributions for this diffusive system with multiple protein–DNA binding–unbinding events. We refer the reader to the first-passage time calculations performed for multiple walkers in the presence of reversible target binding kinetics[125] and to the recent reviews on mathematical methods for computing the first-passage characteristics of diffusive processes in cell biology contexts[126] and other out of equilibrium systems.[127]

These results suggest large fluctuations in the target search times that biologically imply a noisy action of proteins on their DNA sites. The reader is referred here to the recent mathematical studies[94,101] for the precision of molecular signalling, universal proximity effects in the target search kinetics, and for analytical forms of the first-passage time density distributions. We note that the maximum of the first-passage time probability distribution function stems from the initial separation of the TF molecules searching for their target and thus witnesses the geometry-control effect.[94]

C. 3D *versus* 1D diffusion

Fig. 9 illustrates an increasing average number of steps required for the proteins to reach the target with separation $\mathcal{D}$ of the protein production site from the DNA target, as intuitively expected. This indeed demonstrates that GC is beneficial for an overall acceleration of the search. The number of steps to reach the target—characterising the total average search time $\langle T_{3D+1D}\rangle$ for N_p proteins—can be decomposed into the total times spent in the 1D ($\langle T_{1D}\rangle$) and 3D ($\langle T_{3D}\rangle$) diffusion modes, namely

$$\langle T_{3D+1D}\rangle = \langle T_{1D}\rangle + \langle T_{3D}\rangle. \quad (14)$$

This decomposition enables one to assess the relative contributions of 1D sliding and 3D bulk excursions by DNA-binding proteins.[11,50]

In Fig. 9, the data for $\langle T_{3D+1D}\rangle$ are shown as blue circles, whereas the average number of steps in 1D and 3D are, respectively, the green squares and black triangles. Note that the dependencies of both $\langle T_{1D}\rangle$ and $\langle T_{3D}\rangle$ on the distance $\mathcal{D}$ reveal a saturation at long distances from protein production site to the target. Physically, for large $\mathcal{D}$ values, the search time grows only very weakly with $\mathcal{D}$ because the target search process under these conditions is fairly slow, so that it takes place effectively for roughly uniformly distributed proteins in the box, at a finite c_p. The effect of colocalisation thus becomes pronouncedly weaker.

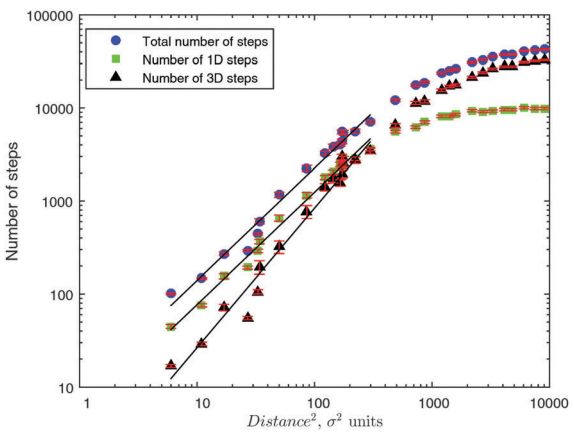


Fig. 9 Total mean numbers of simulation steps to locate the target ($\langle T_{1D+3D}\rangle$, blue circles), total number of all 1D search steps ($\langle T_{1D}\rangle$, green squares), and all 3D search steps ($\langle T_{3D}\rangle$, black triangles) plotted *versus* the diagonal distance $\mathcal{D}$ from a randomly chosen protein production site to the DNA target, which was fixed at (0, 0, 0). The data sets are shown on log–log scale and computed for $N_p = 20$ proteins in the model cell (at $r = 10^{-4}$). The error bars (shown in red in the plot) are evaluated for $M \sim 2000$ realisations.

Note that the random protein production sites in Fig. 9 are sufficiently far away from the DNA chain.

We also mention that for small $\mathcal{D}$ values, the contribution of 1D sliding events prevails, as expected for the case of rather small TF unbinding probability r we consider here. For long separations between the protein production site and the DNA target, the proteins spend significantly more time diffusing in the 3D space. We quantify this stronger 1D contribution for smaller $\mathcal{D}$ values in Fig. 10 below. The scaling exponent in the relation

$$\langle T(\mathcal{D})\rangle \sim \left(\mathcal{D}^2\right)^{\beta} \tag{15}$$

for the data presented in Fig. 9 for the region of short distances $\mathcal{D}$ is $\beta \approx 1.21$ for the total number of steps, $\beta \approx 1.21$ for the 1D diffusion contribution, and $\beta \approx 1.50$ for the 3D search contribution. The slightly sublinear initial growth of $\mathcal{D}^2$ with the number of steps $\langle T\rangle$ for a protein to reach the target (with exponent $1/\beta$) is intuitively expected. Namely, the target search process should be slower than the linear Brownian-like $\beta = 1$ spreading of TF particles in free space.

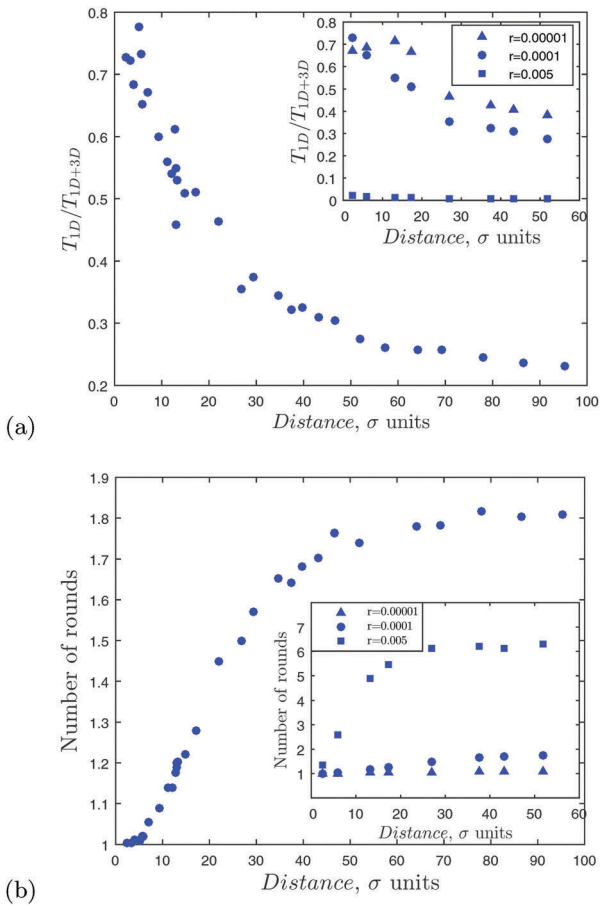


Fig. 10 Variation of the fraction of 1D diffusion time $\langle T_{1D}\rangle/\langle T_{1D+3D}\rangle$ (panel a) and the number of diffusion rounds N_R (panel b) *versus* the separation $\mathcal{D}$ randomly chosen in the simulations, computed for the parameters of Fig. 9. The insets show the results for varying unbinding probability r, for the same values as those in Fig. 4.

Using the elementary time scale $\delta\tau$ estimated in eqn (9), one can compare the search times from our simulations with the experimental values,[13] see eqn (2). For large distances $\mathcal{D}$, in Fig. 9, we obtain mean search times $\langle\tau\rangle \lesssim 1$ s. Note however that for a larger simulation box and thus for longer potentially realisable separations $\mathcal{D}$, naturally, longer search times are expected. Moreover, for conditions of less pronounced facilitated diffusion—for instance, for weaker protein–DNA affinities, in Fig. 4 compare the data for $r = 10^{-4}$ and $r = 0.005$ for the same $N_p = 20$ proteins as in Fig. 9—we would observe a severe slowing down of the search, as compared to $\langle\tau\rangle \lesssim 1$ s for strong protein–DNA binding. These factors, as well as the possibility of multiple TF sliding events over the target without binding,[13] will shift the search times obtained in simulations towards the experimental search times of several minutes, namely $\approx$350 s measured in ref. 12, $\approx$310 s predicted in ref. 31, and tens of seconds in the theoretical model of ref. 63.

Mathematically, 1D diffusion is a highly redundant search process, as argued previously.[9,11,85,89] In particular, for long DNA molecules, this diffusion mode will over-sample the linear substrate, multiply visiting target-free sites. In contrast, the search in 3D is not redundant, but the proteins rarely bind the target directly from the solution[11] (except maybe at very high c_p, which is biologically less relevant). Target localisation typically takes place from the 1D mode and a proper combination of 3D and 1D rounds of diffusion enables the search time in facilitated diffusion to be minimised.[1,8,10,11,29]

We combine the data sets of Fig. 9 and plot in Fig. 10 the relative time the proteins spent in the 1D diffusion mode, $\langle T_{1D}\rangle/\langle T_{1D+3D}\rangle$, as a function of $\mathcal{D}$. At very short distances $\mathcal{D}$, the searchers spend $\sim$80% of the total search time in the 1D diffusion mode. Note that experimentally the lac repressor proteins tracked in living *E. coli* cells were shown[12] to spend $\sim$90% of time in the nonspecifically bound mode, diffusing along the DNA. Naturally, for longer randomly chosen distances $\mathcal{D}$, the duration of 3D searches increases, as compared to diffusion times in 1D, rationalising a decreasing trend for $\langle T_{1D}\rangle/\langle T_{1D+3D}\rangle$ in Fig. 10a. For longer distances $\mathcal{D}$, we find that the proteins spend as little as $\sim$20% in 1D. As in this setup the initial radial protein–DNA distances R are appreciable, for large separations $\mathcal{D}$ the proteins need to wander through more lattice sites before reaching the DNA. This describes a decreasing contribution of 1D search to the data shown in Fig. 10a. At long distances—as compared to the box size—the times of 3D and 1D search reveal a saturation, mirrored in the behaviour of the ratio $\langle T_{1D}\rangle/\langle T_{1D+3D}\rangle$ in Fig. 10a.

The variation of the number of rounds of 3D and 1D diffusion with the distance $\mathcal{D}$ is illustrated in Fig. 10b. The number of rounds $N_R = 1$ corresponds to exactly one 3D and one 1D round of diffusion before the target is located. We find that at first, there is an increase in the number of rounds with distance, after which N_R saturates. As the probability of detachment from the DNA is proportional to the overall time the protein diffuses in the 1D mode, increasing $\langle T_{1D}\rangle$ implies an increasing number of rounds. The number of diffusion rounds we find in Fig. 10b qualitatively agrees with the findings of recent computer

simulations performed for comparable values of nonspecific protein–DNA binding energies E_b, see Fig. 5 in ref. 50.

The main data sets in Fig. 10 are computed for $r = 10^{-4}$. We also find that with increasing protein–DNA binding strength E_b (and thus with decreasing unbinding probability r), the number N_R of rounds of protein diffusion decreases, ultimately approaching $N_R \approx 1$. Concurrently, the fraction of search time that the proteins spend in the 1D mode increases, as expected, see the insets of Fig. 10a. For even smaller unbinding probabilities ($r = 10^{-5}$) the proteins only very rarely leave the DNA chain in the simulations. In contrast, for $r = 0.005$, the tracers spend very little time in the 1D mode, and the number of search cycles for considerable distances $\mathcal{D}$ thus increases from $N_R \approx 1.8$ for $r = 10^{-4}$ to $N_R \approx 6$ found for $r = 5 \times 10^{-3}$, see the insets of Fig. 10b.

The radial distance R of proteins from the DNA can itself be a factor affecting the number of steps to reach the target for the same $\mathcal{D}$ (at least for short R). In Fig. 15, we show the variation of the number of steps with the radial distance from the DNA, while keeping the separation $\mathcal{D}$ constant, $\mathcal{D} \approx 13\sigma$. As one can see, with increasing radial distance R, the proteins spend longer times locating the target, as expected for a more 3D-dominated process at larger R. In contrast, for close radial initial positions, the proteins locate the linear DNA chain quicker and in the facilitated diffusion scheme, they employ more sliding steps.

IV. Discussion and conclusions

The facilitated diffusion model is a paradigm in the description of the protein search process for their cognate binding sites on DNA. Typically, in such studies, either the search by a single protein is exploited, or some equilibrated initial conditions are selected. We here investigated by simulations the target search problem on DNA focusing on the effects of the initial position of bursting proteins with respect to the target position, and protein concentration. We quantified the intuitive expectation that the search time for the DNA target decreases due to increasing local transient protein concentrations and due to shorter distances between the sites of protein production and DNA targets, underlying the rapid search hypothesis.[9] The central result of this study is the physical quantification of GC effects taking place in biological cells achieved *via* computer simulations.

Experimentally, the search time by proteins for a specific DNA site, as quantified by single-molecule detection experiments, is often in the range of several minutes.[12] The mean number of simulation steps $\langle T \rangle$ to reach the target is proportional to the average time $\langle \tau \rangle$ measured experimentally, $\langle \tau \rangle = f\langle n \rangle$, where f is a proportionality factor. Knowing real distances in the cells and using the fact that $\sigma \approx 10.6$ nm in simulations, we can compare the behaviour of the measured target search times $\langle \tau \rangle$ for a particular distance in experiments to our simulation-based findings, while other system parameters, such as E_b and D_{1D}/D_{3D}, are kept constant.

The current model represents a rather simple approach to GC and facilitated diffusion effects, based on a straight DNA fragment and Fickian diffusion of protein molecules in the 3D and 1D propagation modes. Some modifications due to transient anomalous bulk diffusion—indeed detected for spreading of proteins and other macromolecules inside crowded biological cells[35,103,104,128–130]—can be included in the model, thus generalising the canonical Brownian motion picture.[50] It is also important that the implications of multiple targets present on the DNA chain[46,131] are quantified. The limitations of a straight DNA chain—*i.e.* the absence of any DNA coiling[85,86] and conformational dynamics[29,40,89,117,132] in the target search process—can also be relaxed in future simulation-based models.

Another biologically important feature affecting the target search time by the TF proteins is the presence of other proteins bound to the DNA. These may be other TFs, specifically or nonspecifically bound. In addition, these may be strongly DNA bound protein complexes—such as histone-like HU/H-NS structural proteins in the nucleoids of prokaryotes and large nucleosome core particles in the nuclei of eukaryotes[65,133]—that often establish a quasiperiodic and rather dense array of occupied sites on the DNA acting as ''road blocks''. DNA targets (transcription start sites) are positioned often between neighbouring nucleosomes.[134] This can result in dual effects on the target search rate[37,135] and also initiate some cooperative effects in TF–DNA binding.[136] For recent experimental advances, theoretical models and computer simulations on the target search problems in the presence of obstacles, we refer to previous studies.[29,34,36,37,45,137–140]

On the next level of DNA organisation, a number of additional spacial proximity effects of the DNA can emerge for particular 3D structures of genomic DNA molecules compacted inside prokaryotic nucleoids[13,40,92] or *via* a multilevel looping of chromatin fibers ultimately forming eukaryotic chromosomes.[9,61,88,109,141,142] In particular, for the gene-rich Chr19 chromosome of humans, large scale computer simulations have demonstrated spatial proximity features for thousands of coregulated genes,[88] in agreement with the Hi-C experimental evidence available. The possibility of jumps and intersegmental transfers of TF proteins between proximally looped, supercoiled, or chromatin-compacted fragments of genomic DNAs under realistic degrees of DNA compaction inside cells is thus the most biologically important feature to be included in future extensions of this simulation approach.

Going beyond the mean first-passage times, considering the entire distribution of target search and signal transfer times to the target[63,93] will reveal important additional insights. This is vital when the statistics of events is insufficient[94,95] and the first-passage times are strongly spread (*e.g.*, ''fat'' power-law tails). Thus, the few-encounter limit considers the relevance of the most likely binding event. Moreover, statistics of reproducibility of two first-passage events is an important feature to analyse.[94–96]

Lastly, even in simple prokaryotic cells, the spatial distribution of TFs was demonstrated to be strongly heterogeneous.[92] Therefore, new experimental evidence supports the need for a full spatio-temporal modelling of protein–DNA binding kinetics and of local variations in TF concentrations[63] over

thermodynamic approaches[143–147] operating with bulk protein levels (models valid only for large copy-number situations). We believe that the current study will stimulate further experimental and theoretical developments, aiming at a better understanding of physical–chemical effects of GC in gene regulation taking place in living cells.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix

Fig. 11–15.

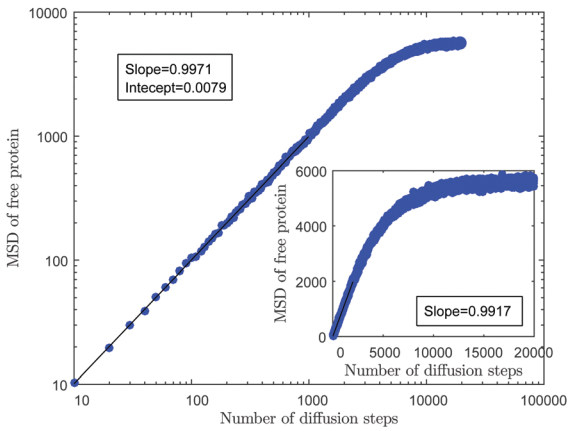


Fig. 11 MSD for 3D tracer diffusion in the simulation box without targets plotted *versus* the number of steps, $\langle \mathbf{r}^2(n)\rangle/\sigma^2$. The results are averaged over $M \sim 10^3$ trajectories and recorded every 10 steps in simulations. The line in this log–log plot is the best fit, in agreement with the linear MSD growth (eqn (6)). The MSD reveals a saturation at long times, as expected for confined Brownian motion, see ref. 106 and 148.

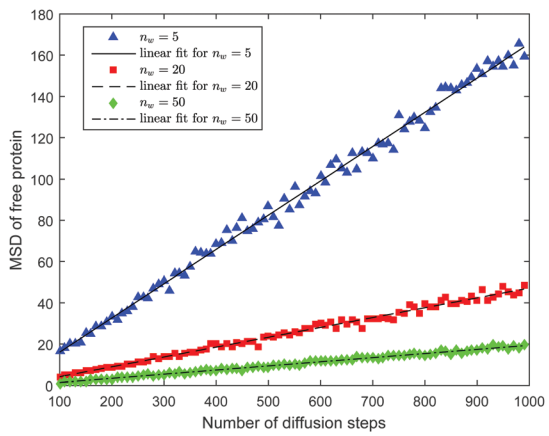


Fig. 12 MSD for target-free 1D protein diffusion, $\langle x^2(n)\rangle/\sigma^2$, computed for varying waiting times n_w given in the legend, with the data shown in the linear–linear scale. The straight lines are fits by eqn (6). For longer simulations times a plateau in the MSD is observed due to the finite model cell size (not shown).

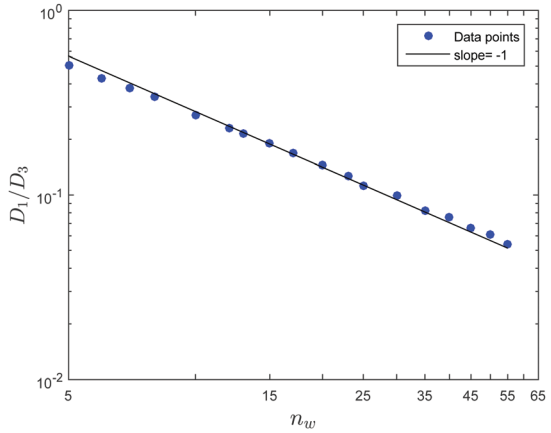


Fig. 13 Calibration curve for D_{1D}/D_{3D} *versus* n_w, shown on log–log scale. The expected slope of −1 is indicated.

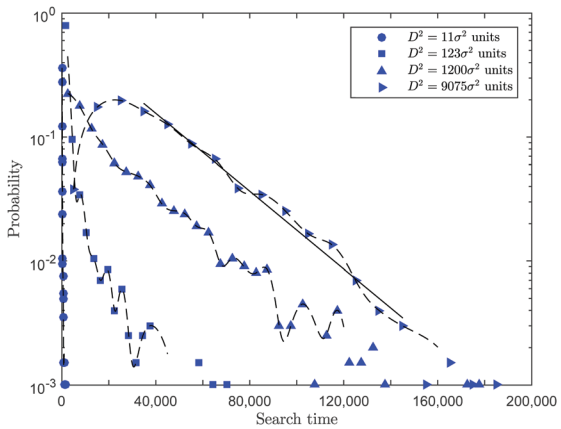


Fig. 14 The data of Fig. 8 plotted on log-linear scale. The straight line indicates an exponential decay of the first-passage time distribution for the *longest* distance $\mathcal{D}$ in the data set, see eqn (13).

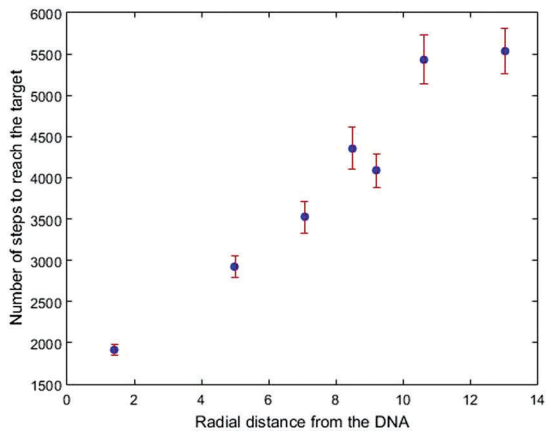


Fig. 15 Average number of steps $\langle n\rangle$ for a protein from the burst of N_p = 20 molecules to reach the target, computed for varying radial distances R of proteins from the DNA, at *fixed* distance $\mathcal{D} \approx 13\sigma$ and $r = 10^{-4}$.

Acknowledgements

P. K. acknowledges financial support *via* the DAAD funding program WISE-2017, Grant 91652046. P. K. thanks Prokash Kumar Kundu for graphical assistance and help with drawings.

References

1 A. D. Riggs, S. Bourgeois and M. Cohn, The lac repressor-operator interaction. 3. Kinetic studies, *J. Mol. Biol.*, 1970, **53**, 401–417.
2 O. G. Berg, R. B. Winter and P. H. von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory, *Biochemistry*, 1981, **20**, 6929–6948.
3 R. B. Winter, O. G. Berg and P. H. von Hippel, Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions, *Biochemistry*, 1981, **20**, 6961–6977.
4 A. Revzin, *The Biology of Nonspecific DNA Protein Interactions*, CRC Press, 1990.
5 P. H. Richter and M. Eigen, Diffusion controlled reaction rates in spheroidal geometry: application to repressor-operator association and membrane bound enzymes, *Biophys. Chem.*, 1974, **2**, 255–263.
6 M. Slutsky and L. A. Mirny, Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential, *Biophys. J.*, 2004, **87**, 4021–4035.
7 M. Slutsky, M. Kardar and L. A. Mirny, Diffusion in correlated random potentials, with applications to DNA, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **69**, 061903.
8 S. E. Halford and J. F. Marko, How do site-specific DNA-binding proteins find their targets?, *Nucleic Acids Res.*, 2004, **32**, 3040–3052.
9 G. Kolesov, Z. Wunderlich, O. N. Laikova, M. S. Gelfand and L. A. Mirny, How gene order is influenced by the biophysics of transcription regulation, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13948–13953.
10 S. E. Halford, An end to 40 years of mistakes in DNA–protein association kinetics?, *Biochem. Soc. Trans.*, 2009, **37**, 343–348.
11 A. B. Kolomeisky, Physics of protein–DNA interactions: mechanisms of facilitated target search, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2088–2095.
12 J. Elf, G.-W. Li and X. S. Xie, Probing transcription factor dynamics at the single-molecule level in a living cell, *Science*, 2007, **316**, 1191–1194.
13 P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg and J. Elf, The lac repressor displays facilitated diffusion in living cells, *Science*, 2012, **336**, 1595–1598.
14 P. C. Bressloff and J. M. Newby, Stochastic models of intracellular transport, *Rev. Mod. Phys.*, 2013, **85**, 135–196.
15 B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, Garland Science, New York, 5th edn, 2008.
16 M. V. Smoluchowski, Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen, *Phys. Z.*, 1916, **17**, 557–585.
17 M. V. Smoluchowski, Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen, *Z. Phys. Chem.*, 1917, **92**, 129–168.
18 E. Gudowska-Nowak, K. Lindenberg and R. Metzler, Preface: Marian Smoluchowski's 1916 paper-a century of inspiration, *J. Phys. A: Math. Gen.*, 2017, **50**, 380301.
19 L. A. Mirny, Biophysics: cell commuters avoid delays, *Nat. Phys.*, 2008, **4**, 93–95.
20 C. Loverdo, O. Bénichou, M. Moreau and R. Voituriez, Enhanced reaction kinetics in biological cells, *Nat. Phys.*, 2008, **4**, 134–137.
21 P. H. von Hippel and O. G. Berg, Facilitated target location in biological systems, *J. Biol. Chem.*, 1989, **264**, 675–678.
22 D. M. Gowers, G. G. Wilson and S. E. Halford, Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15883–15888.
23 I. M. Sokolov, R. Metzler, M. C. Williams and K. Pant, Target search of *N* sliding proteins on a DNA, *Biophys. J.*, 2005, **89**, 895–902.
24 M. Coppey, O. Bénichou, R. Voituriez and M. Moreau, Kinetics of target site localization of a protein on DNA: a stochastic approach, *Biophys. J.*, 2004, **87**, 1640–1649.
25 M. A. Lomholt, T. Ambjörnsson and R. Metzler, Optimal target search on a fast-folding polymer chain with volume exchange, *Phys. Rev. Lett.*, 2005, **95**, 260603.
26 A. Esadze, C. A. Kemme, A. B. Kolomeisky and J. Iwahara, Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: origin of the optimal search at physiological ionic strength, *Nucleic Acids Res.*, 2014, **42**, 7039–7046.
27 L. Hu, A. Y. Grosberg and R. Bruinsma, Are DNA transcription factor proteins maxwellian demons?, *Biophys. J.*, 2008, **95**, 1151–1156.
28 A. G. Cherstvy, A. B. Kolomeisky and A. A. Kornyshev, Protein–DNA interactions: reaching and recognizing the targets, *J. Phys. Chem. B*, 2008, **112**, 4741–4750.
29 L. A. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith and A. Kosmrlj, How a protein searches for its site on DNA: the mechanism of facilitated diffusion, *J. Phys. A: Math. Gen.*, 2009, **42**, 434013.
30 O. Bénichou, C. Loverdo, M. Moreau and R. Voituriez, Intermittent search strategies, *Rev. Mod. Phys.*, 2011, **83**, 81–129.
31 M. Bauer and R. Metzler, *In vivo* facilitated diffusion model, *PLoS One*, 2013, **8**, e53956.
32 M. Bauer and R. Metzler, Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states, *Biophys. J.*, 2012, **102**, 2321–2330.
33 A. Veksler and A. B. Kolomeisky, Speed-selectivity paradox in the protein search for targets on DNA: is it real or not?, *J. Phys. Chem. B*, 2013, **117**, 12695–12701.

34 M. Bauer, E. S. Rasmussen, M. A. Lomholt and R. Metzler, Real sequence effects on the search dynamics of transcription factors on DNA, *Sci. Rep.*, 2015, **5**, 10072.

35 Y. M. Wang, R. H. Austin and E. C. Cox, Single molecule measurements of repressor protein 1D diffusion on DNA, *Phys. Rev. Lett.*, 2006, **97**, 048302.

36 A. A. Shvets and A. B. Kolomeisky, Crowding on DNA in protein search for targets, *J. Phys. Chem. Lett.*, 2016, **7**, 2502–2506.

37 A. A. Shvets, M. Kochugaeva and A. B. Kolomeisky, Role of static and dynamic obstacles in the protein search for targets on DNA, *J. Phys. Chem. B*, 2016, **120**, 5802–5809.

38 K. V. Klenin, H. Merlitz, J. Langowski and C. X. Wu, Facilitated diffusion of DNA-binding proteins, *Phys. Rev. Lett.*, 2006, **96**, 018104.

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

40 E. Chow and J. Skolnick, DNA internal motion likely accelerates protein target search in a packed nucleoid, *Biophys. J.*, 2017, **112**, 2261–2270.

41 I. Goychuk and V. O. Kharchenko, Anomalous features of diffusion in corrugated potentials with spatial correlations: faster than normal, and other surprises, *Phys. Rev. Lett.*, 2014, **113**, 100601.

42 M. Kong and B. Van Houten, Rad4 recognition-at-a-distance: physical basis of conformation-specific anomalous diffusion of DNA repair proteins, *Prog. Biophys. Mol. Biol.*, 2017, **127**, 93–104.

43 R. S. Spolar and M. T. Record, Coupling of local folding to site-specific binding of proteins to DNA, *Science*, 1994, **263**, 777–784.

44 E. F. Koslover, M. de la Rosa, A. Diaz and A. J. Spakowitz, Theoretical and computational modeling of target-site search kinetics *in vitro* and *in vivo*, *Biophys. J.*, 2011, **101**, 856–865.

45 C. A. Brackley, M. E. Cates and D. Marenduzzo, Intracellular facilitated diffusion: searchers, crowders, and blockers, *Phys. Rev. Lett.*, 2013, **111**, 108101.

46 M. Lange, M. Kochugaeva and A. B. Kolomeisky, Protein search for multiple targets on DNA, *J. Chem. Phys.*, 2015, **143**, 105102.

47 K. Pant, R. L. Karpel, I. Rouzina and M. C. Williams, Salt dependent binding of T4 gene 32 protein to single and double-stranded DNA: single molecule force spectroscopy measurements, *J. Mol. Biol.*, 2005, **349**, 317–330.

48 C. G. Kalodimos, N. Biris, A. M. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens and R. Kaptein, Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes, *Science*, 2004, **305**, 386–389.

49 O. Pulkkinen and R. Metzler, Variance-corrected Michaelis-Menten equation predicts transient rates of single-enzyme reactions and response times in bacterial gene-regulation, *Sci. Rep.*, 2015, **5**, 17820.

50 L. Liu, A. G. Cherstvy and R. Metzler, Facilitated diffusion of transcription factor proteins with anomalous bulk diffusion, *J. Phys. Chem. B*, 2017, **121**, 1284–1289.

51 P. S. Swain, M. B. Elowitz and E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12795–12800.

52 M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain, Stochastic gene expression in a single cell, *Science*, 2002, **297**, 1183–1186.

53 B. Munsky, G. Neuert and A. van Oudenaarden, Using gene expression noise to understand gene regulation, *Science*, 2012, **336**, 183–187.

54 N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain and M. B. Elowitz, Gene regulation at the single-cell level, *Science*, 2005, **307**, 1962–1965.

55 E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman and A. van Oudenaarden, Regulation of noise in the expression of a single gene, *Nat. Genet.*, 2002, **31**, 69–73.

56 J. Paulsson, Models of stochastic gene expression, *Phys. Life Rev.*, 2005, **2**, 157–175.

57 D. R. Rigney and W. C. Schieve, Stochastic model of linear, continuous protein synthesis in bacterial populations, *J. Theor. Biol.*, 1977, **69**, 761–766.

58 O. G. Berg, A model for the statistical fluctuations of protein numbers in a microbial population, *J. Theor. Biol.*, 1978, **71**, 587–603.

59 H. H. McAdams and A. Arkin, Stochastic mechanisms in gene expression, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 814–819.

60 J. Paulsson, O. G. Berg and M. Ehrenberg, Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 7148–7153.

61 S. Chong, C. Chen, H. Ge and X. S. Xie, Mechanism of transcriptional bursting in bacteria, *Cell*, 2014, **158**, 314–326.

62 I. Golding, J. Paulsson, S. M. Zawilski and E. C. Cox, Real-time kinetics of gene activity in individual bacteria, *Cell*, 2005, **123**, 1025–1036.

63 O. Pulkkinen and R. Metzler, Distance matters: the impact of gene proximity in bacterial gene regulation, *Phys. Rev. Lett.*, 2013, **110**, 198101.

64 J. Yu, J. Xiao, X. Ren, K. Lao and X. S. Xie, Probing gene expression in live cells, one protein molecule at a time, *Science*, 2006, **311**, 1600–1603.

65 J. M. Raser and E. K. O'Shea, Control of stochasticity in eukaryotic gene expression, *Science*, 2004, **304**, 1811–1814.

66 R. Dessalles, V. Fromion and P. Robert, A stochastic analysis of autoregulation of gene expression, *J. Math. Biol.*, 2017, **75**, 1253–1283.

67 C. Di Rienzo, V. Piazza, E. Gratton, F. Beltram and F. Cardarelli, Probing short-range protein Brownian motion in the cytoplasm of living cells, *Nat. Commun.*, 2014, **5**, 5891.

68 O. G. Berg, J. Paulsson and M. Ehrenberg, Fluctuations in repressor control: thermodynamic constraints on stochastic focusing, *Biophys. J.*, 2000, **79**, 2944–2953.

69 N. Friedman, L. Cai and X. S. Xie, Linking stochastic dynamics to population distribution: an analytical framework of gene expression, *Phys. Rev. Lett.*, 2006, **97**, 168302.

70 L. Cai, N. Friedman and X. S. Xie, Stochastic protein expression in individual cells at the single molecule level, *Nature*, 2006, **440**, 358–362.

71 A. Raj and A. van Oudenaarden, Single-molecule approaches to stochastic gene expression, *Annu. Rev. Biophys.*, 2009, **38**, 255–270.

72 V. V. Hausnerova and C. Lanctot, Transcriptional output transiently spikes upon mitotic exit, *Sci. Rep.*, 2017, **7**, 12607.

73 A. Raj and A. van Oudenaarden, Nature, nurture, or chance: stochastic gene expression and its consequences, *Cell*, 2008, **135**, 216–226.

74 V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17256–17261.

75 D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler and F. Naef, Mammalian genes are transcribed with widely different bursting kinetics, *Science*, 2011, **332**, 472–474.

76 K. B. Halpern, S. Tanami, S. Landen, M. Chapal, L. Szlak, A. Hutzler, A. Nizhberg and S. Itzkovitz, Bursty gene expression in the intact mammalian liver, *Mol. Cell*, 2015, **58**, 147–156.

77 P. Dröge and B. Müller-Hill, High local protein concentrations at promoters: strategies in prokaryotic and eukaryotic cells, *BioEssays*, 2001, **23**, 179–183.

78 M. Thatta and A. van Oudenaarden, Intrinsic noise in gene regulatory networks, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 8614–8619.

79 A. Sanchez and J. Kondev, Transcriptional control of noise in gene expression, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 5081–5086.

80 H. Xu, S. O. Skinner, A. M. Sokac and I. Golding, Stochastic kinetics of nascent RNA, *Phys. Rev. Lett.*, 2016, **117**, 128101.

81 P. Liu, R. Song, G. L. Elison, W. Peng and M. Acar, Noise reduction as an emergent property of single-cell aging, *Nat. Commun.*, 2017, **8**, 680.

82 W. Bialek and S. Setayeshgar, Physical limits to biochemical signaling, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10040–10045.

83 T. Hu and B. I. Shklovskii, How does a protein search for the specific site on DNA: the role of disorder, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**, 021903.

84 A. G. Cherstvy, Positively charged residues in DNA-binding domains of structural proteins follow sequence-specific positions of DNA phosphate groups, *J. Phys. Chem. B*, 2009, **113**, 4242–4247.

85 B. van den Broek, M. A. Lomholt, S.-M. Kalisch, R. Metzler and G. J. L. Wuite, How DNA coiling enhances target localization by proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 15738–15742.

86 M. A. Lomholt, B. van den Broek, S.-M. J. Kalisch, G. J. L. Wuite and R. Metzler, Facilitated diffusion with DNA coiling, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 8204–8208.

87 L. D. Hurst, C. Pal and M. J. Lercher, The evolutionary dynamics of eukaryotic gene order, *Nat. Rev. Genet.*, 2004, **5**, 299–310.

88 M. Di Stefano, A. Rosa, V. Belcastro, D. di Bernardo and C. Micheletti, Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19, *PLoS Comput. Biol.*, 2013, **9**, e1003019.

89 Z. Wunderlich and L. A. Mirny, Spatial effects on the speed and reliability of protein–DNA search, *Nucleic Acids Res.*, 2008, **36**, 3570–3578.

90 R. Aboukhalil, B. Fendler and G. S. Atwal, Kerfuffle: a web tool for multi-species gene colocalization analysis, *BMC Bioinf.*, 2013, **14**, 22.

91 A. B. Pardee, F. Jacob and J. Monod, The genetic control and cytoplasmic expression of ''inducibility'' in the synthesis of β-galactosidase by *E. coli*, *J. Mol. Biol.*, 1959, **1**, 165–178.

92 T. E. Kuhlman and E. C. Cox, Gene location and DNA density determine transcription factor distributions in *Escherichia coli*, *Mol. Syst. Biol.*, 2012, **8**, 610.

93 A. Godec and R. Metzler, First passage time distribution in heterogeneity controlled kinetics: going beyond the mean first passage time, *Sci. Rep.*, 2016, **6**, 20349.

94 A. Godec and R. Metzler, Universal proximity effect in target search kinetics in the few-encounter limit, *Phys. Rev. X*, 2016, **6**, 041037.

95 T. G. Mattos, C. Mejia-Monasterio, R. Metzler and G. Oshanin, First passages in bounded domains: when is the mean first passage time meaningful?, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2012, **86**, 031143.

96 T. G. Mattos, C. Mejia-Monasterio, R. Metzler, G. Oshanin and G. Schehr, *First-passage phenomena and their applications*, World Scientific Publishing, Singapore, 2014.

97 H. C. Berg and E. M. Purcell, Physics of chemoreception, *Biophys. J.*, 1977, **20**, 193–219.

98 W. Bialek and S. Setayeshgar, Cooperativity, sensitivity, and noise in biochemical signaling, *Phys. Rev. Lett.*, 2008, **100**, 258101.

99 B. Hu, D. A. Kessler, W.-J. Rappel and H. Levine, Effects of input noise on a simple biochemical switch, *Phys. Rev. Lett.*, 2011, **107**, 148101.

100 A. Godec and R. Metzler, Signal focusing through active transport, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2015, **92**, 010701(R).

101 A. Godec and R. Metzler, Active transport improves the precision of linear long distance molecular signalling, *J. Phys. A: Math. Gen.*, 2016, **49**, 364001.

102 C.-C. S. Yan, S. R. Chepyala, C.-M. Yen and C.-P. Hsu, Efficient and flexible implementation of Langevin simulation for gene burst production, *Sci. Rep.*, 2017, **7**, 16851.

103 R. Metzler, J.-H. Jeon, A. G. Cherstvy and E. Barkai, Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24128–24164.

104 F. Höfling and T. Franosch, Anomalous transport in the crowded world of biological cells, *Rep. Prog. Phys.*, 2013, **76**, 046602.

105 I. M. Sokolov, Models of anomalous diffusion in crowded environments, *Soft Matter*, 2012, **8**, 9043–9052.

106 S. Burov, J.-H. Jeon, R. Metzler and E. Barkai, Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking, *Phys. Chem. Chem. Phys.*, 2011, **13**, 1800–1812.

107 K. Nørregaard, R. Metzler, C. M. Ritter, K. Berg-Sørensen and L. B. Oddershede, Manipulation and motion of organelles and single molecules in living cells, *Chem. Rev.*, 2017, **117**, 4342–4375.

108 T. Kühn, T. O. Ihalainen, J. Hyväluoma, N. Dross, S. F. Willman, J. Langowski, M. Vihinen-Ranta and J. Timonen, Protein diffusion in mammalian cell cytoplasm, *PLoS One*, 2011, **6**, e22962.

109 C. C. Fritsch and J. Langowski, Anomalous diffusion in the interphase cell nucleus: the effect of spatial correlations of chromatin, *J. Chem. Phys.*, 2010, **133**, 025101.

110 C. A. Brackley, M. E. Cates and D. Marenduzzo, Facilitated diffusion on mobile DNA: configurational traps and sequence heterogeneity, *Phys. Rev. Lett.*, 2012, **109**, 168103.

111 J. M. Schurr, The one-dimensional diffusion coefficient of proteins absorbed on DNA: hydrodynamic considerations, *Biophys. Chem.*, 1979, **9**, 413–414.

112 A. Bakk and R. Metzler, Nonspecific binding of the o_r repressors ci and cro of bacteriophage lambda, *J. Theor. Biol.*, 2004, **231**, 525–533.

113 A. G. Cherstvy, Electrostatic interactions in biological DNA-related systems, *Phys. Chem. Chem. Phys.*, 2011, **13**, 9942–9968.

114 M. Cencini and S. Pigolotti, Energetic funnel facilitates facilitated diffusion, *Nucleic Acids Res.*, 2018, **160**, DOI: 10.1093/nar/gkx1220.

115 G.-W. Li and X. S. Xie, Central dogma at the single-molecule level in living cells, *Nature*, 2011, **475**, 308–315.

116 A. C. S. Yu, J. F. C. Loo, S. Yu, S. K. Kong and T.-F. Chan, Monitoring bacterial growth using tunable resistive pulse sensing with a pore-based technique, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 855–862.

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

118 M. Ptashne, *A genetic switch*, Cambridge, MA, 1992.

119 M. Barbi, C. Place, V. Popkov and M. Salerno, A model of sequence-dependent protein diffusion along DNA, *J. Biol. Phys.*, 2004, **30**, 203–226.

120 K. Pant, R. L. Karpel, I. Rouzina and M. C. Williams, Mechanical measurement of single-molecule binding rates: kinetics of DNA helix-destabilization by T4 gene 32 protein, *J. Mol. Biol.*, 2004, **336**, 851–870.

121 S. B. Yuste and L. Acedo, Number of distinct sites visited by *N* random walkers on a euclidean lattice, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 2000, **61**, 2340–2347.

122 S. Redner, *A guide to first-passage processes*, Cambridge University Press, 2001.

123 D. Grebenkov, R. Metzler and G. Oshanin, Effects of the target aspect ratio and intrinsic reactivity onto diffusive search in bounded domains, *New J. Phys.*, 2017, **19**, 103025.

124 C. Mejia-Monasterio, G. Oshanin and G. Schehr, First passages for a search by a swarm of independent random searchers, *J. Stat. Mech.: Theory Exp.*, 2011, **P06022**, 1–34.

125 D. S. Grebenkov, First passage times for multiple particles with reversible target-binding kinetics, *J. Chem. Phys.*, 2017, **147**, 134112.

126 S. Iyer-Biswas and A. Zilman, First-passage processes in cellular biology, *Adv. Chem. Phys.*, 2016, **160**, 261–306.

127 A. J. Bray, S. N. Majumdar and G. Schehr, Persistence and first-passage properties in nonequilibrium systems, *Adv. Phys.*, 2013, **62**, 225–361.

128 I. Golding and E. C. Cox, Physical nature of bacterial cytoplasm, *Phys. Rev. Lett.*, 2006, **96**, 098102.

129 E. Barkai, Y. Garini and R. Metzler, Strange kinetics of single molecules in living cells, *Phys. Today*, 2012, **65**(8), 29–37.

130 D. Ernst, M. Hellmann, J. Kohler and M. Weiss, Fractional Brownian motion in crowded fluids, *Soft Matter*, 2012, **8**, 4886–4889.

131 V. V. Palyulin, V. N. Mantsevich, R. Klages, R. Metzler and A. V. Chechkin, Comparison of pure and combined search strategies for single and multiple targets, *Eur. Phys. J. E: Soft Matter Biol. Phys.*, 2017, **90**, 170.

132 A. Mondal and A. Bhattacherjee, Understanding the role of DNA topology in target search dynamics of proteins, *J. Phys. Chem. B*, 2017, **121**, 9372–9381.

133 L. Saiz and J. M. G. Vilar, Multilevel deconstruction of the *in vivo* behavior of looped DNA–protein complexes, *PLoS One*, 2007, **2**, e355.

134 D. A. Beshnova, A. G. Cherstvy, Y. Vainshtein and V. B. Teif, Regulation of the nucleosome repeat length *in vivo* by the DNA sequence, protein concentrations and long-range interactions, *PLoS Comput. Biol.*, 2014, **10**, e1003698.

135 L. Liu and K. Luo, DNA-binding protein searches for its target: non-monotonic dependence of the search time on the density of roadblocks bound on the DNA chain, *J. Chem. Phys.*, 2015, **142**, 125101.

136 L. A. Mirny, Nucleosome-mediated cooperativity between transcription factors, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 22534–22539.

137 D. Krepel, D. Gomez, S. Klumpp and Y. Levy, Mechanism of facilitated diffusion during a DNA search in crowded environments, *J. Phys. Chem. B*, 2016, **120**, 11113–11122.

138 D. Krepel and Y. Levy, Protein diffusion along DNA: on the effect of roadblocks and crowders, *J. Phys. A: Math. Gen.*, 2016, **49**, 494003.

139 A. Marcovitz and Y. Levy, Obstacles may facilitate and direct DNA search by proteins, *Biophys. J.*, 2013, **104**, 2042–2050.

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

141 G. Wedemann and J. Langowski, Computer simulation of the 30-nanometer chromatin fiber, *Biophys. J.*, 2002, **82**, 2847–2859.

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

143 A. Bakk and R. Metzler, *In vivo* non-specific binding of λ ci and cro repressors is significant, *FEBS Lett.*, 2004, **563**, 66–68.
144 A. Bakk, R. Metzler and K. Sneppen, Sensitivity of o_r in phage λ, *Biophys. J.*, 2004, **86**, 58–66.
145 L. Saiz and J. M. G. Vilar, Ab initio thermodynamic modeling of distal multisite transcription regulation, *Nucleic Acids Res.*, 2008, **36**, 726–731.
146 J. M. G. Vilar and L. Saiz, Systems biophysics of gene expression, *Biophys. J.*, 2013, **104**, 2574–2585.
147 G. K. Ackers, A. D. Johnson and M. A. Shea, Quantitative model for gene regulation by λ phage repressor, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 1129–1133.
148 S. K. Ghosh, A. G. Cherstvy, D. S. Grebenkov and R. Metzler, Anomalous, non-Gaussian tracer diffusion in crowded two-dimensional environments, *New J. Phys.*, 2016, **18**, 013027.
149 Note that as we always start in the 3D volume here, the definition of NR is different from that in eqn (3) of ref. 50, where the protein starting positions were always on the 1D DNA and the effects of protein–DNA binding affinity onto facilitated diffusion were examined within a (simplistic) model with a straight DNA chain in simulation box at fixed protein abundances.
150 Note however that often the lac repressor protein slides several times over its operator site before the actual binding event takes place[13].
151 Note however that, even at well-controlled in vitro conditions, even simple TFs exhibit a very considerable spread of diffusion coefficients D_{1D} and sliding lengths. The reader is referred here to Fig. 4 in ref. 35 for the lac repressor diffusing on a strongly-stretched 2D-confined fragment of a non-specific DNA. One expects even larger variations and uncertainties of protein diffusivities and sliding lengths under inherently-noisy conditions for TF diffusion inside living cells.

