

# Effects of macromolecular crowding and DNA looping on gene regulation kinetics

Gene-Wei Li<sup>1</sup>, Otto G. Berg<sup>2</sup> and Johan Elf<sup>3\*</sup>

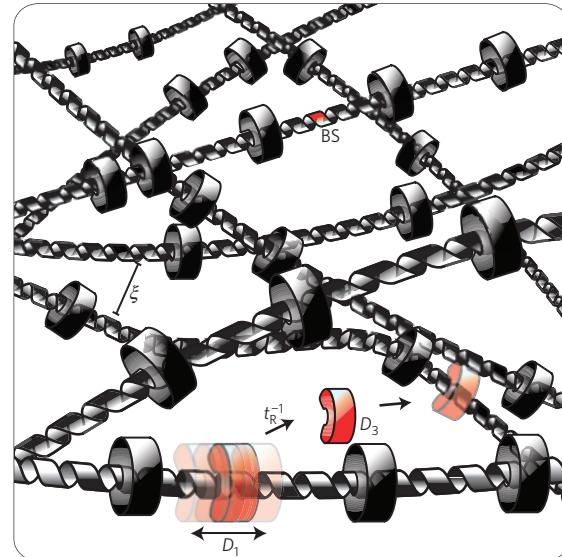
**DNA-binding proteins control how genomes function.** The theory of facilitated diffusion<sup>1</sup> explains how DNA-binding proteins can find targets apparently faster than the diffusion limit by using reduced dimensionality<sup>2,3</sup>—combining three-dimensional (3D) diffusion through cytoplasm with 1D sliding along DNA (refs 3–15). However, it does not include a description of macromolecular crowding on DNA as observed in living cells. Here, we show that such a physical constraint to sliding greatly reduces the search speed, in agreement with single-molecule measurements. Interestingly, the generalized theory also reveals significant insights into the design principles of biology. First, it places a hard constraint on the total number of DNA-binding proteins per cell. Remarkably, the number measured for *Escherichia coli* fits within the optimal range. Secondly, it defines a new role for DNA looping, a ubiquitous topological motif in genomes. DNA looping can speed up the search process by bypassing proteins that block the sliding track close to the target.

In living cells, DNA-binding proteins search for their specific binding sites across millions of chemically similar DNA sequences (Fig. 1). Unlike the widely studied settings in test tubes, the search problem is further complicated by the vast amount of other proteins, both cytoplasmic and DNA-bound. Taking the bacterium *Escherichia coli* as an example, its macromolecule density is close to that of a protein crystal, which slows down the three-dimensional (3D) diffusion by at least one order of magnitude compared with that in water. Furthermore, about 30% of its DNA is associated with proteins, mostly bound with no sequence specificity<sup>16,17</sup>. As a result, the sliding track of bacterial DNA is truncated into short and mostly uniformly distributed DNA fragments<sup>18</sup>. The crowding on a 1D track has drastically different effects from 3D crowding. Whereas the latter can be modelled as increased viscosity, macromolecules in 1D block the sliding motion of other proteins along DNA and will therefore be referred to as ‘roadblocks.’ Although roadblocks are omnipresent in biological cells, little attention has been given to their effects on the search kinetics, which we will show to be important.

To account for the effects of crowding, we build on the conventional formalism that deals with DNA-binding proteins with diffusion constants  $D_3$  in cytoplasm and  $D_1$  along DNA. The modified Smoluchowski bi-molecular association rate constant between the protein and the specific site is<sup>3</sup>:

$$k_a = \frac{2\pi}{\ln(\xi/2b)} (\alpha D_3) \langle L \rangle \quad (1)$$

In essence, the rate constant is the product of an effective diffusion constant ( $\alpha D_3$ ) and an effective target size ( $\langle L \rangle$ ),

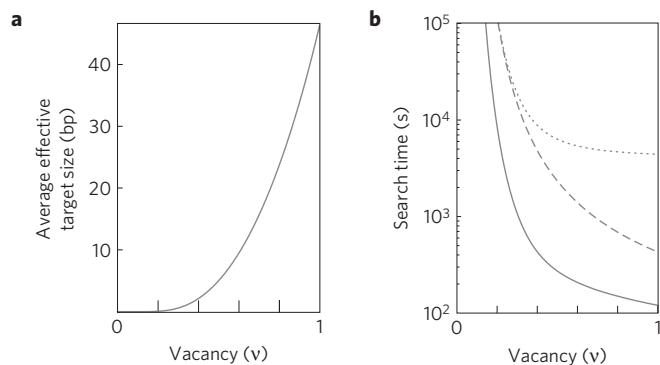


**Figure 1 | Facilitated diffusion.** In the model investigated by Berg *et al.*<sup>3</sup>, the DNA-binding protein (red cylinders) searches for its binding site (BS) by combining 1D diffusion ( $D_1$ ) along nonspecific DNA sequences and 3D diffusion ( $D_3$ ) in cytoplasm. The nonspecific residence time ( $t_R^{-1}$ ) is the average time to dissociate from and diffuse away a distance  $\xi/2$  from DNA. In a living cell, with  $10^6$ – $10^9$  nonspecific sites, many rounds of 1D and 3D searches are needed before locating the target. In this work, we study the effects of other DNA-binding proteins (black cylinders) that are present in a vast amount in cells.

both of which depend on the intracellular crowding (below). The logarithmic factor corrects for short-ranged microscopic dissociation/association events (hopping) on the same DNA segment<sup>19</sup> (see Supplementary Information). The ratio  $\xi/2b$  compares the length scales of macroscopic versus microscopic dissociation. In the current context,  $b$  is the DNA radius (1 nm) and  $\xi$  is the concentration-dependent correlation length<sup>20</sup> of the DNA polymer. The search time, which is defined as the average time for one DNA-binding protein to find one specific site, is related to the rate constant as  $\tau = V k_a^{-1}$  for a system of volume  $V$ .

In *E. coli*, macromolecular crowding in 3D increases the viscosity by at least tenfold, and the motion remains normal diffusive<sup>13</sup>. In addition, the effective diffusion constant is reduced by the transient binding to nonspecific sites, which are subject to competition with other DNA-binding proteins. Define the vacancy,  $v$ , as the fraction of roadblock-free DNA. Under the assumption that the average length occupied by a roadblock on

<sup>1</sup>Department of Physics and Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA, <sup>2</sup>Department of Molecular Evolution, Uppsala University, Uppsala SE-752 36, Sweden, <sup>3</sup>Linnaeus Center for Bioinformatics, Uppsala University, Uppsala SE-751 24, Sweden.  
\*e-mail: johan.elf@lcb.uu.se



**Figure 2 | Kinetic effects of roadblocks.** **a**, The effective target size,  $\langle L(v) \rangle$ . Lower chromosomal vacancy means higher density of roadblocks on DNA, and hence smaller effective target size. The parameters used in this plot are derived from previous experiments as described in the Methods section. **b**, Facilitation by sliding along nonspecific DNA. The search time as a function of vacancy is plotted in three scenarios: searching through combined 1D and 3D diffusion (solid line), 3D diffusion and transient nonspecific binding without 1D diffusion (dotted line) and 3D alone (dashed line). Among all cases, sliding along DNA yields the shortest search time despite the presence of roadblocks.

DNA and the length of the specific binding sites are the same,  $d \sim 20$  base pairs (bp), the effective diffusion constant becomes (see Supplementary Information)

$$\alpha D_3 \approx \frac{D_3}{1 + K_{RD} c_{ns} v e^{1-1/v}} \quad (2)$$

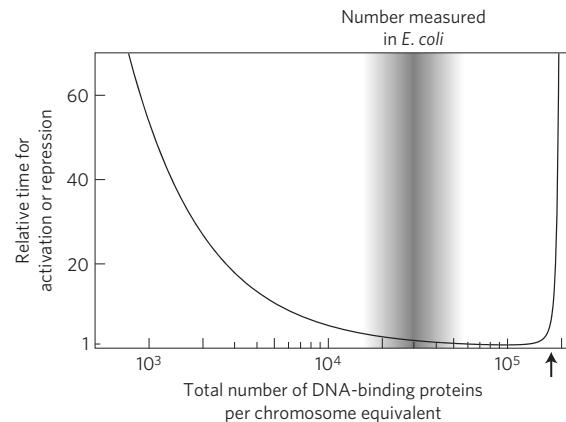
where  $c_{ns}$  is the concentration of nonspecific DNA and  $K_{RD}$  is the binding constant between the protein and an unoccupied nonspecific site. Here,  $v e^{1-1/v}$  is the fraction of accessible DNA for a DNA-binding protein with size  $d$  (see Supplementary Information).

The effective target size  $\langle L \rangle$  measures the length of target-flanking DNA from which the searcher can slide into the target before dissociation. In the presence of roadblocks on the track, such length is reduced. Consider identical non-passing particles diffusing in 1D. The effective target size can be obtained by analysing how far a particular particle can travel within the residence time  $t_R$  of nonspecific binding. Such single-file diffusion<sup>21,22</sup> has been shown to exhibit subdiffusion at long times. Using a first-passage-time approach (see the Methods section), the size is approximately

$$\langle L(v) \rangle \approx v e^{1-\frac{1}{v}} \left( l + 2 \sqrt{\frac{D_1 t_R}{1 + (t_R/t_x)^{1/2}}} \right) \quad (3)$$

where  $t_x = (dv/(1-v))^2/\pi D_1$  is the diffusion time across the average gap between roadblocks<sup>23</sup> and  $l$  is the size of one base pair. The effective target size for the prototypical *lac* repressor in *E. coli* is plotted in Fig. 2a. Combining equations (2) and (3) into equation (1), the search time as a function of vacancy can be obtained (Fig. 2b). It is worth noting that when nonspecific binding is dominant ( $K_{RD} c_{ns} v e^{1-1/v} \gg 1$ ), the factor  $v e^{1-1/v}$  in the denominator cancels with that in  $\langle L(v) \rangle$ . In other words, the enhancement in global diffusion cancels out the effect of partially blocked targets. This leaves only the vacancy dependence in the restricted sliding length. Meanwhile, the classical results  $\langle L(v) \rangle = (l + 2\sqrt{D_1 t_R})$  and  $\alpha D_3 = D_3 / (1 + K_{RD} c_{ns})$  can be recovered in the limit  $v \rightarrow 1$ .

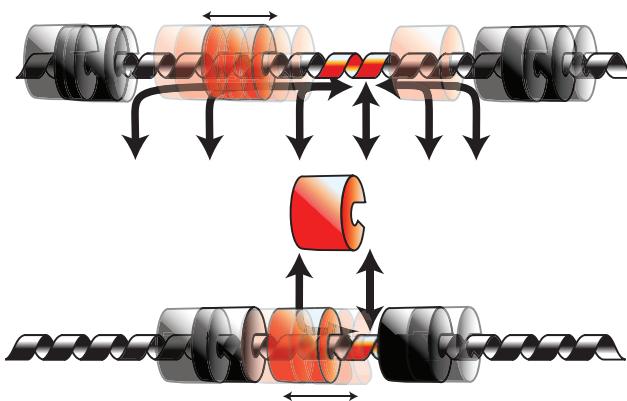
The search time measurement *in vivo* for a transcription factor became possible only recently with live-cell single-molecule imaging techniques<sup>13</sup>. With fluorescently labelled *lac* repressor



**Figure 3 | Optimal range of total DNA-binding proteins.** In the *E. coli* genome, we consider a single site controlled by a specific DNA-binding protein of which the concentration is at a fixed fraction of all DNA-binding proteins. The amount of time it takes for the site to be occupied, normalized to the lowest point, is plotted against the total number of all DNA-binding proteins (solid line). Increasing the overall number would speed up the search process only when the number is small (<10<sup>4</sup>). However, for the total number close to that measured in *E. coli* (shaded), further increasing the copy number has a marginal effect in reducing the reaction time. At  $\sim 1.6 \times 10^5$  per chromosome ( $v \sim 0.3$ , arrow), the response time increases catastrophically.

dimers in *E. coli*, we probed the search time by removing the inducers from the growth media and watching how soon the rebinding occurs. It takes at most 60 s for three *lac* repressors to find any of the two strong operators used in the experimental strain<sup>13</sup>. The search time for a single transcription factor to find a single operator site is therefore less than 360 s. This is in agreement with the predicted facilitated target location on a crowded DNA with the biologically relevant vacancy ( $v \sim 0.7$ ), shown in Fig. 2b (solid line). The measurement is however not compatible with models without sliding ( $D_1 = 0$ ) or without nonspecific binding ( $K_{RD} = 0$  and  $D_1 = 0$ ) as plotted with dotted and dashed lines, respectively, in Fig. 2b. In the wild-type *E. coli*, there are 5–10 tetrameric *lac* repressors<sup>24</sup> searching independently. The time to reach a single operator for the first repressor is therefore 35–70 s.

It may seem that the speed of gene regulation could be increased indefinitely by increasing the concentration of the transcription factor. However, the kinetic effect of crowding (equation (1)) on DNA immediately sets a hard constraint on the maximum number of DNA-binding proteins for a given genome. Although increasing the number of a particular transcription factor species can speed up the corresponding gene regulation, a global increase in all transcription factor concentrations could instead lead to increased search times owing to increased crowding. We applied the generalized model (equations (1)–(3)) to examine this effect. Assuming the number of a specific transcription factor ( $n$ ) represents a fixed fraction of the DNA-binding protein ( $N$ ),  $n = fN$ , the response time of a given transcription factor is  $t = V k_a^{-1} / fN$ . For a chromosome of  $M$  base pairs, the vacancy is related to  $N$  by  $v = 1 - Nd/M$ . When using the parameters for *E. coli*, a minimum response time is achieved at  $N = 10^4$ – $10^5$  total proteins per chromosome (Fig. 3) independent of the small parameter  $f$ . The actual number of these proteins in *E. coli* was reported to be  $\sim 30,000$  during exponential growth<sup>16,17</sup>. It is thus evident that further increasing the overall concentration from the current value has little effect in shortening the reaction time. The new kinetic constraint applies to the total concentration of DNA-binding proteins and implies that the concentrations of individual DNA-binding proteins are subject to competition during the evolution of the proteome.



**Figure 4 | Blocked association and dissociation pathways.** In addition to randomly placed roadblocks, specific sequences can be used to position roadblocks (black cylinders) at a precise location away from the target site (red helix). When they are placed shorter than  $\sqrt{D_1\tau_R}$  away from the target, both association and dissociation rates are reduced. In the case of association, the effective target size is reduced. In the case of dissociation, the chance of returning to the binding site after dissociation is increased because of the reflective boundary.

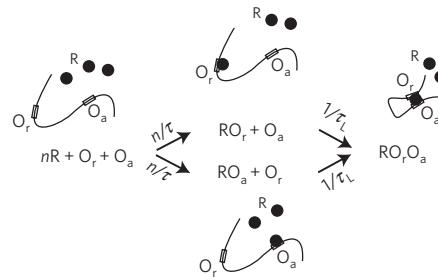
Slow association due to blocked binding pathways directly implies slow dissociation (Fig. 4). Because the equilibrium binding constant for a fixed vacancy is independent of roadblock position, detailed balance requires that association and dissociation kinetics are changed in the same way, that is,  $k_d = k_a K_{eq}^{-1}$ . We therefore suggest that it is possible that cells regulate binding kinetics of specific proteins by positioning site-specific roadblocks close to the target. This makes it possible to tune the kinetics more gradually than what is possible by the on/off-like regulation using competitive binding to a specific site.

Beyond the original facilitated diffusion with 1D and 3D pathways, our extension to crowded DNA also shows that DNA looping between specific binding sites emerges as an alternative strategy for speeding up target location. DNA looping is a ubiquitous topological motif in bacterial genomes<sup>25,26</sup>. It is often established by a single protein complex that binds to two specific chromosomal binding sites, of which only one is responsible for transcription regulation and the other is considered an auxiliary site. Whereas the two sites are generally separated by 50 bp or more, the dense roadblocks make sliding distances shorter than 20 bp at biologically relevant vacancy ( $v \sim 0.7$ , Fig. 2a). The two binding sites therefore become independent targets in the search process (Fig. 5) and the auxiliary site constitutes an alternative path to the regulatory site by looping. The mean time to reach the regulatory site is

$$t = \frac{\tau}{n} \left( \frac{1 + \tau_L n / \tau}{2 + \tau_L n / \tau} \right) \quad (4)$$

where  $\tau = V k_a^{-1}$  is again the search time,  $n$  is the number of transcription factor molecules and  $\tau_L$  is the mean loop formation time, which is much faster than the dissociation time from the auxiliary site. Equation (4) implies that the regulatory site can therefore be found twice as fast if  $n\tau_L \ll \tau$ .

Indeed, the loop formation time for a freely jointed DNA chain is predicted to be shorter than seconds for typical loop sizes (150–1,500 bp; refs 27–29), whereas the initial search process is orders of magnitude slower as analysed above. It therefore implies that DNA looping can facilitate target location by using an auxiliary binding site from which rapid intersite transfer is possible. The effect should be most prominent for transcription factors with low copy number ( $n \ll \tau / \tau_L$ ). It should be noted



**Figure 5 | DNA looping and search kinetics.** Association to binding sites that are farther apart than the average sliding distance occurs independently at rate  $n/\tau$ , and the rate for the first association to any one site is  $2n/\tau$ . When the auxiliary site  $O_a$  is accurately positioned relative to the regulatory site,  $O_r$ , rapid transfer through looping is imminent after binding the auxiliary site. As long as the loop formation time ( $\tau_L$ ) is much shorter than  $\tau/n$ , the time to go from the free repressor state ( $nR + O_r + O_a$ ) to the occupied regulatory site state ( $RO_r + O_a$  or  $RO_r O_a$ ) is halved when compared with no DNA looping.

that this new role of looping in speeding up target location is different from the hypothetical mechanism of intersegment transfer between nonspecific sites, as transfer efficiency from a random site is much lower than that from an auxiliary site. It should also be noted that the mechanism does not hold when the loop is not formed by a single protein complex, but by two transcription factors binding individually on their sites, such as in the case of the lambda repressor. The new mechanism is not limited to just one auxiliary site. The classic *lac* operon in *E. coli* does actually have two auxiliary sites. The reason for having two auxiliary sites has remained puzzling, as one auxiliary site would be sufficient to form a strong loop<sup>30</sup>. However, the topological enhancement introduced here suggests that the two auxiliary sites can reduce the search time by a factor of 3, if looping from both auxiliary sites is fast. Increasing the number of auxiliary sites instead of the numbers of transcription factors may therefore be a strategy to maintain the response time in gene regulation without contributing to crowding on the chromosome.

## Methods

**Effective target size for diffusing roadblock.** Single-file diffusion<sup>21,22</sup> has been shown to exhibit subdiffusion at long times. For time longer than the diffusion time across the average gap size, the mean-square displacement scales as the time to the one-half power:  $\langle x^2(t) \rangle = 2Ft^{1/2}$ , where  $F = dv(1-v)^{-1}\sqrt{D_1/\pi}$ . As we are interested in the scaling for all vacancy levels, a complete knowledge of  $\langle x^2(t) \rangle$  across all timescales is needed. A phenomenological approximation has been shown to be valid on all relevant time scales<sup>23</sup>. Lin *et al.* have demonstrated experimentally that the single-particle Green's function retains a Gaussian shape at all times, with the variance well described by  $\langle x^2(t) \rangle = 2D_1 t / (1 + (t/t_x)^{1/2})$ , where  $t_x = (F/D_1)^2$  is the diffusion time across the average gap size<sup>23</sup>. The effective target size can be expressed as a product of the probability to land on DNA (binding to a roadblock-free nonspecific site) on encounter and the length of DNA along which the protein can diffuse into the target without dissociation. A simple statistical calculation (see Supplementary Information) shows that the probability for a free segment longer than the protein footprint ( $d$ ) at a particular position is  $v e^{-1/v}$ . This is also the probability that the protein can bind to DNA when diffusing from cytoplasm. Here, we assume the roadblocks stay bound during each sliding event for physiological relevance (see Supplementary Information).

To calculate the average length of DNA from which the protein can slide into the target before dissociation, we use a first-passage-time-based method. For a protein initially bound at position  $x$  on nonspecific DNA, the probability to reach the target is the time integral of the first-passage-time density,  $f(x, t)$ , multiplied by the likelihood of staying bound nonspecifically at a given time. The average DNA length is thus an integral along the DNA, weighted by this probability of reaching the target. If the nonspecific residence time is exponentially distributed with average  $t_R$ , the effective target size is (including the factor from the previous paragraph):

$$\langle L(v) \rangle = v e^{-1/v} \int_{-\infty}^{\infty} \left( \int_0^{\infty} f(x, t) e^{-t/t_R} dt \right) dx \quad (5)$$

As the free propagator in single-file diffusion is Gaussian-shaped with variance of  $2D_1 t / (1 + (t/t_x)^{1/2})$ , we can obtain the first-passage-time density, for  $x > 0$ , with absorbing boundary condition at  $x = 0$ , using the method of images.

$$f(x, t) = -\frac{\partial}{\partial t} \int_0^\infty (G(x', t; x) - G(x', t; -x)) dx' \quad (6)$$

where  $G(x', t; x)$  is the free propagator from  $(x, 0)$  to  $(x', t)$ .

$$G(x', t; x) = \sqrt{\frac{1 + (t/t_x)^{1/2}}{4\pi D_1 t}} \exp\left(-(x' - x)^2 \frac{1 + (t/t_x)^{1/2}}{4D_1 t}\right) \quad (7)$$

From equations (5)–(7), we obtain a complicated expression for the effective target size:

$$\langle L(v) \rangle = ve^{1-\frac{1}{v}} \frac{\sqrt{D_1 t_R}}{\sqrt{\pi}} \begin{bmatrix} \left(\frac{t_x}{t_R}\right)^{\frac{1}{4}} \Gamma\left(\frac{1}{4}\right) {}_1F_1\left(\frac{5}{4}, \frac{1}{2}, -\frac{t_x}{t_R}\right) \\ + 6\left(\frac{t_x}{t_R}\right)^{\frac{3}{4}} \Gamma\left(\frac{3}{4}\right) {}_1F_1\left(\frac{7}{4}, \frac{3}{2}, -\frac{t_x}{t_R}\right) \\ - 8\left(\frac{t_x}{t_R}\right)^{\frac{5}{4}} \Gamma\left(\frac{3}{4}\right) {}_1F_1\left(\frac{3}{4}, \frac{1}{2}, -\frac{t_x}{t_R}\right) \\ + 4\left(\frac{t_x}{t_R}\right)^{\frac{1}{4}} \Gamma\left(\frac{1}{4}\right) {}_1F_1\left(\frac{5}{4}, \frac{3}{2}, -\frac{t_x}{t_R}\right) \\ + 8\left(\frac{t_x}{t_R}\right)^{\frac{1}{2}} {}_2F_2\left(\frac{1}{2}, 1; \frac{1}{4}, -\frac{t_x}{t_R}\right) \\ - 8\left(\frac{t_x}{t_R}\right)^{\frac{1}{2}} {}_2F_2\left(1, \frac{3}{2}; \frac{5}{4}, -\frac{t_x}{t_R}\right) \end{bmatrix} \quad (8)$$

where  ${}_pF_q$  is the generalized hypergeometric function and  $\Gamma$  is the gamma function. This expression can be approximated as

$$\langle L(v) \rangle = ve^{1-\frac{1}{v}} \left[ 2 \sqrt{\frac{D_1 t_R}{1 + (t_R/t_x)^{1/2}}} \right] \quad (9)$$

We plot the exact solution, equation (8), and the approximated function, equation (9), in Supplementary Fig. S1. The approximation holds for six decades in  $t_x$ . Equation (9) is the effective target size for binding through nonspecific sites. Direct binding is accounted for in equation (3) by adding an extra term ( $l$ ), the size of the specific site.

**Kinetics parameters implied from experiments.** When applying the theory to real biological systems (specifically, the lac repressor protein in *E. coli*), we used the recent single-molecule measurements<sup>13</sup> to estimate the relevant parameter values. First, by tracking single lac repressors in *E. coli*, we determined the effective diffusion constant  $D \approx 0.4 \mu\text{m}^2 \text{s}^{-1}$ , whereas the 3D diffusion constant in cytoplasm is  $D_3 \approx 3 \mu\text{m}^2 \text{s}^{-1}$ . Using equation (2) and  $c_{ns} = 4.6 \times 10^6 \mu\text{m}^{-3}$ , the nonspecific equilibrium binding constant can be estimated as  $K_{RD} \approx 0.85 \text{ mM}^{-1}$  if  $v = 1$ , or  $K_{RD} \approx 1.9 \text{ mM}^{-1}$  if  $v = 0.7$ . With the knowledge of the equilibrium binding constant, the nonspecific residence time  $t_R$  can be obtained under the condition of diffusion-controlled association ( $t_R = K_{RD}/k_a^{ns}$ ). The diffusion-controlled macroscopic association rate to a nonspecific site is<sup>3</sup>  $k_a^{ns} = 2\pi D_3 l / \ln(\xi/2b)$ . Here,  $l$  is the size of one base pair ( $l \approx 0.34 \text{ nm}$ ),  $b$  is the DNA radius (1 nm) and  $\xi/2$  can be defined as half of the average distance between nearest DNA segments,  $\pi(\xi/2)^2 Ml = \text{nucleoid volume}$ , where  $M$  is the total number of base pairs. For an *E. coli* chromosome ( $M = 4.6 \times 10^6$ ) with 1  $\mu\text{m}^3$  of nucleoid volume,  $\xi/2 = 14 \text{ nm}$ . This gives  $k_a^{ns} \approx 1.5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ . As a result, the macroscopic dissociation constant is  $t_R \approx 0.6 \text{ ms}$  if the diffusion constants above were measured at  $v = 1$ , and  $t_R \approx 1.3 \text{ ms}$  if those were measured at  $v = 0.7$ .

Received 3 October 2008; accepted 11 February 2009;  
published online 15 March 2009

## References

- von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
- Richter, P. H. & Eigen, M. Diffusion controlled reaction rates in spheroidal geometry. Application to repressor-operator association and membrane bound enzymes. *Biophys. Chem.* **2**, 255–263 (1974).
- Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
- Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli* repressor-operator interaction: Equilibrium measurements. *Biochemistry* **20**, 6948–6960 (1981).
- Winter, R. B., Berg, O. G. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: Kinetic measurements and conclusions. *Biochemistry* **20**, 6961–6977 (1981).
- Barkley, M. D. Salt dependence of the kinetics of the lac repressor-operator interaction: Role of nonoperator deoxyribonucleic acid in the association reaction. *Biochemistry* **20**, 3833–3842 (1981).
- Gowers, D. M. & Halford, S. E. Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.* **22**, 1410–1418 (2003).
- Slutsky, M. & Mirny, L. A. Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
- Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucl. Acids Res.* **32**, 3040–3052 (2004).
- Zhou, H. X. & Szabo, A. Enhancement of association rates by nonspecific binding to DNA and cell membranes. *Phys. Rev. Lett.* **93**, 178101 (2004).
- Klenin, K. V., Merlitz, H., Langowski, J. & Wu, C. X. Facilitated diffusion of DNA-binding proteins. *Phys. Rev. Lett.* **96**, 018104 (2006).
- Hu, T., Grosberg, A. Y. & Shklovskii, B. I. How proteins search for their specific sites on DNA: The role of DNA conformation. *Biophys. J.* **90**, 2731–2744 (2006).
- Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
- Eliazar, I., Koren, T. & Klafter, J. Searching circular DNA strands. *J. Phys. Condens. Matter* **19**, 065140 (2007).
- Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA* **103**, 5752–5757 (2006).
- Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* **181**, 6361–6370 (1999).
- Johnson, R. C., Johnson, L. M., Schmidt, J. W. & Gardner, J. F. in *The Bacterial Chromosome* (ed. Higgins, N. P.) (ASM Press, 2005).
- Varshavsky, A. J., Nedospasov, S. A., Bakayev, V. V., Bakayeva, T. G. & Georgiev, G. P. Histone-like proteins in the purified *Escherichia coli* deoxyribonucleoprotein. *Nucl. Acids Res.* **4**, 2725–2745 (1977).
- Berg, O. G. & Blomberg, C. Association kinetics with coupled diffusional flows. Special application to the lac repressor-operator system. *Biophys. Chem.* **4**, 367–381 (1976).
- Rubinstein, M. & Colby, R. H. *Polymer Physics* (Oxford Univ. Press, 2003).
- Harris, T. E. Diffusion with ‘collisions’ between particles. *J. Appl. Probab.* **2**, 323–338 (1965).
- Wei, Q., Bechinger, C. & Leiderer, P. Single-file diffusion of colloids in one-dimensional channels. *Science* **287**, 625–627 (2000).
- Lin, B., Meron, M., Cui, B., Rice, S. A. & Diamant, H. From random walk to single-file diffusion. *Phys. Rev. Lett.* **94**, 216001 (2005).
- Gilbert, W. & Muller-Hill, B. Isolation of the lac repressor. *Proc. Natl Acad. Sci. USA* **56**, 1891–1898 (1966).
- Schleif, R. DNA looping. *Annu. Rev. Biochem.* **61**, 199–223 (1992).
- Halford, S. E., Welsh, A. J. & Szczelkun, M. D. Enzyme-mediated DNA looping. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 1–24 (2004).
- Jun, S., Bechhoefer, J. & Ha, B.-Y. Diffusion-limited loop formation of semiflexible polymers: Kramers theory and the intertwined time scales of chain relaxation and closing. *Europhys. Lett.* **64**, 420–426 (2003).
- Hyeon, C. & Thirumalai, D. Kinetics of interior loop formation in semiflexible chains. *J. Chem. Phys.* **124**, 104905 (2006).
- Merlitz, H., Rippe, K., Klenin, K. V. & Langowski, J. Looping dynamics of linear DNA molecules and the effect of DNA curvature: A study by Brownian dynamics simulation. *Biophys. J.* **74**, 773–779 (1998).
- Oehler, S., Eismann, E. R., Kramer, H. & Muller-Hill, B. The three operators of the lac operon cooperate in repression. *EMBO J.* **9**, 973–979 (1990).

## Acknowledgements

We are grateful for helpful discussions and comments from P. Choi, X. Sunney Xie, N. Dekker, M. Ehrenberg, B. English and J. Yu. This work was supported by the European Research Council, the Swedish Foundation for Strategic Research, KAW, URRC and NSF Graduate Research Fellowship (G.-W.L.).

## Author contributions

G.-W.L. and J.E. designed research; G.-W.L., O.G.B. and J.E. carried out research; G.-W.L. and J.E. wrote the paper.

## Additional information

Supplementary Information accompanies this paper on www.nature.com/naturephysics. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions. Correspondence and requests for materials should be addressed to J.E.