

# The promoter-search mechanism of *Escherichia coli* RNA polymerase is dominated by three-dimensional diffusion

Feng Wang<sup>1,6</sup>, Sy Redding<sup>2,6</sup>, Ilya J Finkelstein<sup>1,5</sup>, Jason Gorman<sup>3,5</sup>, David R Reichman<sup>2</sup> & Eric C Greene<sup>1,4</sup>

Gene expression, DNA replication and genome maintenance are all initiated by proteins that must recognize specific targets from among a vast excess of nonspecific DNA. For example, to initiate transcription, *Escherichia coli* RNA polymerase (RNAP) must locate promoter sequences, which compose <2% of the bacterial genome. This search problem remains one of the least understood aspects of gene expression, largely owing to the transient nature of search intermediates. Here we visualize RNAP in real time as it searches for promoters, and we develop a theoretical framework for analyzing target searches at the submicroscopic scale on the basis of single-molecule target-association rates. We demonstrate that, contrary to long-held assumptions, the promoter search is dominated by three-dimensional diffusion at both the microscopic and submicroscopic scales *in vitro*, which has direct implications for understanding how promoters are located within physiological settings.

Transcription is the key step of gene expression and regulation, in which the information encoded in genomic DNA is transcribed into RNA<sup>1–3</sup>. A complex network of regulatory features allows precise control over the expression of any given gene. This regulation is achieved through the interplay of promoter DNA sequences that dictate the sites of transcript initiation, along with the effects of a multitude of transcription factors and other regulatory elements that can influence the efficiency of transcript initiation, elongation and/or termination<sup>1–3</sup>. At the heart of this regulatory network is RNA polymerase: the protein machinery directly responsible for RNA synthesis<sup>1–4</sup>.

*Escherichia coli* has ~3,000 promoters, each containing a core sequence of ~35 base pairs (bp) in length, with hexameric consensus sites at the –35 (TTGACA) and –10 (TATAAT) regions<sup>1–6</sup>. Prior to synthesizing a transcript, RNAP must find appropriate promoter sequences. Like all DNA-binding proteins, RNAP is expected to employ some form of diffusion to locate its targets (Supplementary Fig. 1)<sup>7</sup>. There are four potential diffusion-based mechanisms that might contribute to the promoter search: (i) one-dimensional (1D) ‘hopping’, where the protein moves along the same molecule of DNA through a correlated series of submicroscopic dissociation and rebinding events before reequilibrating back into free solution; (ii) 1D sliding, where the protein executes a random walk along the DNA without dissociating; (iii) intersegmental transfer, where the protein moves from one site to another through a looped intermediate; and (iv) three-dimensional (3D) diffusion (or ‘jumping’), where the protein starts out fully equilibrated with free solution (that is, it has no ‘memory’ of whether it has previously visited a DNA site) and then finds its targets through direct 3D collisions from solution

(Supplementary Fig. 1)<sup>7,8</sup>. These mechanisms are not mutually exclusive, and different combinations can, in principle, contribute to site-specific targeting for a given DNA-binding protein. Search mechanisms that employ 1D hopping, sliding or intersegmental transfer are collectively referred to as facilitated diffusion<sup>8–12</sup> because the reduction in dimensionality brought about through use of these mechanisms presents the potential for target-site association rates that exceed the limits imposed by pure 3D diffusion<sup>7,9,12</sup>.

The seminal work of Riggs *et al.* established that, under certain conditions, the *lac* repressor binds its target faster than the 3D diffusion limit<sup>13</sup>. Subsequent theoretical and experimental work verified that target-association rates can be accelerated through facilitated diffusion, and these results are also often used to argue that facilitated diffusion therefore must contribute to target searches<sup>9–12</sup>. However, as noted in the literature<sup>11,14</sup>, there is little evidence to support this generalization on the basis of the findings with the *lac* repressor, and the *lac* repressor itself may be atypical in terms of its DNA-binding and target-search properties. In addition, prior theoretical models demonstrating that the fastest possible searches require both 3D diffusion and 1D sliding assume that the search is being conducted by just one protein molecule<sup>9–11,15–17</sup>. This assumption is reasonable for low-abundance proteins, such as the *lac* repressor (fewer than ten molecules per cell) but is less appropriate when considering proteins present at higher concentrations. Indeed, it has more recently been recognized that facilitated diffusion can in fact slow down target searches by causing proteins to waste too much time surveying nonspecific DNA<sup>10,15,18–20</sup>, which led to the suggestion that this outcome might be avoided, in the case of some proteins,

<sup>1</sup>Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York, USA. <sup>2</sup>Department of Chemistry, Columbia University, New York, New York, USA. <sup>3</sup>Department of Biological Sciences, Columbia University, New York, New York, USA. <sup>4</sup>Howard Hughes Medical Institute, Columbia University, New York, New York, USA. <sup>5</sup>Present addresses: Department of Chemistry and Institute for Cellular & Molecular Biology, University of Texas at Austin, Austin, Texas, USA (I.J.F.) and Vaccine Research Center, National Institutes of Health, Bethesda, Maryland, USA (J.G.). <sup>6</sup>These authors contributed equally to this work. Correspondence should be addressed to E.C.G. (ecg2108@columbia.edu).

Received 13 July; accepted 20 November; published online 23 December 2012; doi:10.1038/nsmb.2472

through a combination of low affinity for nonspecific DNA and increased protein copy number<sup>11</sup>. Nevertheless, on the basis of the work with the *lac* repressor, facilitated diffusion is still commonly assumed to play a part in the majority of cellular target-search processes. Accordingly, a number of studies have reported that RNAP can move long distances along DNA by 1D sliding<sup>21–25</sup>, and as a consequence it is also now widely assumed that RNAP locates promoters through facilitated diffusion involving a 1D search<sup>26</sup>. Despite this, no promoter-association rate exceeding the 3D-diffusion limit ( $\sim 10^8$ – $10^9$  M<sup>-1</sup> s<sup>-1</sup>) has ever been reported<sup>27,28</sup>, and the potential contribution of facilitated diffusion to the promoter search process has been challenged in the literature<sup>29</sup>.

In an effort to help resolve the mechanism of the promoter search, here we used single-molecule optical imaging of nanofabricated DNA curtains to visualize molecules of *E. coli* RNAP as they searched for the native promoters within the phage  $\lambda$  genome. Using this approach, we could identify intermediates consistent with nonspecifically bound proteins, promoter-associated closed complexes, open complexes and actively transcribing RNAP. We also present a theoretical framework for analyzing search mechanisms in the submicroscopic regime, using experimentally measured kinetic parameters obtained from single-molecule observations. Our experimental results and theoretical calculations argue that facilitated diffusion does not contribute to promoter targeting by *E. coli* RNAP at physiologically relevant protein concentrations. We also show that protein concentration has a dominating effect in dictating how proteins find specific target sites, and the potential rate-accelerating benefits of facilitated diffusion can be overcome through increased protein abundance. The concepts derived from our theoretical treatment of the target-search problem are entirely general and can in principle be applied to any site- or structure-specific nucleic acid-binding protein.

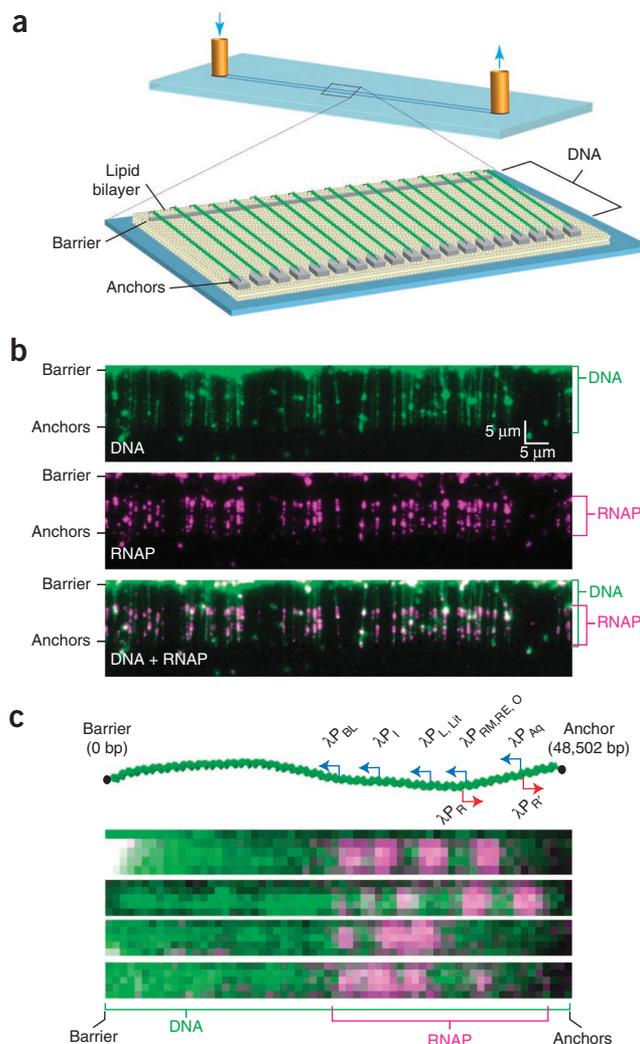
## RESULTS

### Visualizing the promoter search by *E. coli* RNAP

*E. coli* RNAP is among the best-characterized enzymes at the single-molecule level, yet no study has conclusively established how RNAP locates promoters<sup>30</sup>. To distinguish among potential search mechanisms (Supplementary Fig. 1 and Supplementary Table 1)<sup>7,9,11,12</sup>, we used double-tethered DNA curtains to visualize quantum dot-tagged RNAP (QD-RNAP) bound to native promoters within  $\lambda$  DNA (Fig. 1 and Supplementary Table 2). For these assays, we functionalized the  $\lambda$  DNA (48,502 base pairs) at one end with biotin and at the other end with digoxigenin (DIG). We anchored the biotinylated DNA end to a lipid bilayer through a biotin-streptavidin linkage, and the molecules were then aligned along the leading edges of nanofabricated barriers by application of a hydrodynamic force. The DIG-tagged DNA ends were then anchored to anti-DIG antibody-coated pentagons positioned downstream from the linear barriers. This strategy yielded DNA molecules that are all anchored in the same orientation and can be viewed by total internal reflection fluorescence microscopy (TIRFM) in the absence of hydrodynamic flow<sup>31–33</sup>. Using similar assays, we have previously demonstrated that promoter binding by *E. coli* RNAP is  $\sigma^{70}$  dependent and occurs preferentially at physiological ionic strength and that QD-RNAP is active for transcription<sup>34</sup>. We conclude that QD-RNAP faithfully located promoters within the context of our experimental platform.

### Promoter-association assays reveal known intermediates

We next visualized single molecules of RNAP in real time as they searched for native promoters within the  $\lambda$  DNA. To observe the promoter search in real time, QD-RNAP was injected into the flow cell



**Figure 1** Single-molecule DNA-curtain assay for promoter-specific binding by RNA polymerase. (a) Double-tethered DNA-curtain assay for organizing substrates on surfaces of a microfluidic device. (b) Two-color images of YOYO-1-stained DNA (green) bound by QD-RNAP (magenta). (c) Schematic of the  $\lambda$  phage genome (48.5 kb), including relative locations and orientations of promoters aligned with images of QD-RNAP on single DNA molecules (Supplementary Table 2). Most RNAP is shown bound to the promoters, and the left half of the  $\lambda$  DNA that lacks promoters is essentially devoid of bound proteins.

( $\pm$  NTPs), flow was terminated and data were collected at 5, 10 or 100 frames per second (Hz; Fig. 2a,b). These experiments revealed four potential intermediates, for brevity referred to as  $\tau_0$ ,  $\tau_1$ ,  $\tau_2$  and  $\tau_3$  events (Fig. 2a).  $\tau_0$  events were highly transient ( $\tau_0 = 5.58$  (5.46–5.74) milliseconds (ms); range indicates 95% confidence interval;  $R^2 = 0.99$ ) and were also observed with either QDs alone or in the absence of DNA (Supplementary Fig. 2); therefore, we ascribed these events to random diffusion through the detection volume in the absence of any interaction with the DNA, and they were not considered further.  $\tau_1$  events were RNAP dependent, displayed short lifetimes ( $\tau_1 = 29.23$  (24.53–36.18) ms), occurred randomly along the DNA (within our spatial resolution limits of  $\pm 39$  nm), were not observed in the absence of DNA and were uncorrelated with promoter-bound open complexes ( $r = 0.25$ ,  $P = 0.10$ , Pearson correlation analysis; Fig. 2c and Supplementary Fig. 2). RNAP corresponding to  $\tau_2$  dissociated more slowly from the DNA ( $\tau_2 = 3.53$  (2.77–4.8) s;  $R^2 = 0.93$ ) and

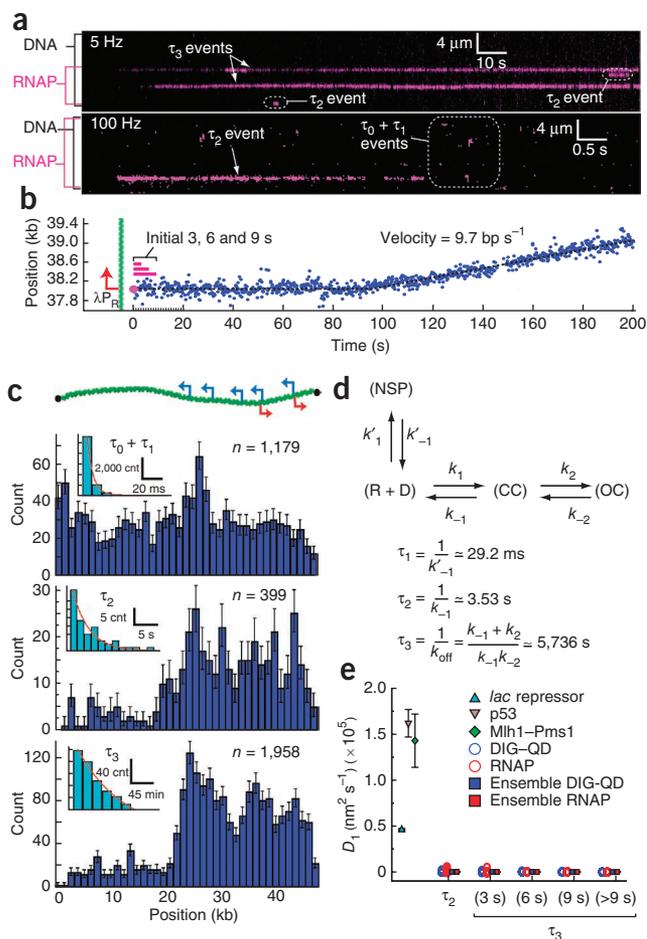
**Figure 2** Visualizing single molecules of RNA polymerase as they search for and engage promoters. **(a)** Kymograms of RNAP binding to  $\lambda$  DNA, showing kinetically distinct intermediates. DNA is unlabeled, and RNAP is magenta. **(b)** Representative example of RNAP binding and initiating transcription from the  $\lambda_{PR}$  promoter; for this assay, RNAP was premixed with all four NTPs immediately before injection into the sample chamber (also **Supplementary Fig. 4**). Initial binding ( $t = 0$  s) is indicated as a magenta dot, and magenta bars highlight the first 3–9 s of the reaction trajectory. **(c)** Binding distributions of kinetically distinct intermediates and corresponding lifetime measurements (insets; also **Supplementary Fig. 2**). A schematic showing the relative promoter location is included. Error bars indicate 70% confidence intervals obtained through bootstrap analysis<sup>34</sup>. Cnt, count. **(d)** Kinetic scheme reflecting observed intermediates. NSP, CC and OC refer to nonspecifically bound, closed complex and open complex, respectively; CC could also represent another intermediate preceding the open complex<sup>3</sup>. Kinetic parameters are not segregated for individual promoters; rather, they are considered collectively, and therefore reported values should be considered an average of all  $\lambda$  promoters. **(e)** Upper bound of observed diffusion coefficients for promoter-bound RNAP compared to immobilized DIG-QDs and other proteins known to undergo 1D diffusion (**Supplementary Figs. 7 and 8** and **Supplementary Table 4**)<sup>32,41,43</sup>. Diffusion coefficients are gamma distributed; therefore, we report the magnitude of the square root of the variance as error bars ( $n \geq 50$  for all data sets).

were strongly correlated with distributions of promoter-bound open complexes ( $r = 0.87$ ,  $P < 7 \times 10^{-14}$ ; **Fig. 2c**)<sup>34</sup>. Molecules classified as  $\tau_3$  exhibited even slower dissociation ( $\tau_3 = 5,736$  (5,150–6,467) s;  $R^2 = 0.99$ ) coincided with known promoters (**Fig. 2c** and **Supplementary Fig. 3**), were resistant to challenge with heparin (a hallmark of open-complex formation) and could initiate transcription (**Fig. 2b** and **Supplementary Fig. 4**), as previously described<sup>34</sup>. These results are consistent with a reaction scheme where  $\tau_1$  corresponds to nonspecifically bound RNAP,  $\tau_2$  to closed complexes and  $\tau_3$  to open complexes (**Fig. 2d**).

Although values for  $\tau_1$  (that is,  $1/k'_{-1}$ ) and  $\tau_2$  (that is,  $1/k_{-1}$ ) have not been previously reported in the literature, our value for  $\tau_2$  was consistent with previously reported equilibrium constants measured in bulk for closed-complex formation ( $k_1/k_{-1}$ ) for several different promoters, assuming a diffusion-limited association rate<sup>35–38</sup>. The value we obtained for  $\tau_3$  (that is,  $1/k_{\text{off}}$ ) was consistent with bulk biochemical data for the lifetime of the open complex<sup>37–39</sup>, providing additional support for our assignment of these events within the reaction scheme. Moreover, within the reaction scheme defined in **Figure 2d**, the ratio of  $\tau_2/\tau_3$  events is equal to  $k_{-1}/k_2$ , yielding a value for  $k_2$  of  $0.056 \text{ s}^{-1}$ , which is also in good agreement with values from the literature<sup>37,38</sup>. We concluded that the promoter-bound intermediates observed in our assay reflected properties consistent with the literature and that the DNA-curtain assay could be used to probe the early stages of RNAP association with promoter DNA that precede the initiation of transcription.

### No microscopically detectable 1D diffusion before promoter binding

Our results demonstrated that QD-tagged RNAP was correctly targeted to promoters in the DNA-curtain assay and that the experimental observables obtained from these assays recapitulated known reaction schemes and kinetic parameters for promoter association and dissociation. Unexpectedly, real-time observations revealed no evidence for microscopic 1D diffusion by RNAP ( $n > 6,000$ ; **Fig. 2a** and **Supplementary Table 2** and described below); only  $<0.5\%$  of proteins exhibited 1D motion, and these rare events did not precede promoter engagement. In addition, the same experiments were conducted over a range of ionic strengths (for example, 0–200 mM KCl, 0–10 mM  $\text{MgCl}_2$ ), including all buffer conditions under which RNAP sliding has



been previously reported (**Supplementary Table 3**). Under no conditions did we find evidence of microscopically detectable 1D diffusion of RNAP along the  $\lambda$  DNA before promoter engagement. Control experiments confirmed that T7 RNAP and *lac* repressor were capable of extensive 1D diffusion in our assays (**Supplementary Fig. 5**)<sup>40–42</sup>, and we have previously shown that the DNA-repair proteins Msh2–Msh6 and Mlh1–Pms1 exhibit 1D diffusion, which indicates that the DNA curtains or QD tags do not prevent protein diffusion along DNA<sup>32,33</sup>. Finally, we readily observed 1D movement for 1.0- $\mu\text{m}$  beads coated with RNAP, which suggests that prior reports of extensive 1D diffusion by *E. coli* RNAP may have been confounded by multivalent aggregates (**Supplementary Fig. 6** and **Supplementary Note**). In summary, we found no direct experimental evidence supporting an extensive contribution of 1D diffusion during the promoter search, which suggests that the promoter search by QD-tagged RNAP within the context of our DNA-curtain assays was dominated by 3D diffusion.

As a further test of the hypothesis that RNAP did not undergo extensive 1D diffusion during the promoter search, we next sought to determine the upper bounds for the observed 1D diffusion coefficients ( $D_{1,\text{obs}}$ ) for QD-RNAP at defined points along the reaction trajectory (**Fig. 2e** and **Supplementary Fig. 7**). Given their transient nature ( $\tau_1 \leq 30$  ms), we could not determine  $D_{1,\text{obs}}$  values for the nonspecifically bound RNAP (described below); however, we did calculate  $D_{1,\text{obs}}$  for intermediates categorized as either closed ( $\tau_2$  events) or open complexes ( $\tau_3$  events) as well as for the first 3–9 s after initial DNA binding for molecules of RNAP that subsequently initiated transcription in the presence of NTPs (**Fig. 2b**). We then compared the resulting  $D_{1,\text{obs}}$  values to published values for several well-characterized proteins

known to undergo 1D diffusion, including the *lac* repressor<sup>41</sup>, p53 (ref. 43) and Mlh1–Pms1 (ref. 32). We also compared the data to immobile QDs coupled to the DNA through a covalent DIG tag (DIG-QD; Fig. 2e, Supplementary Fig. 8 and Supplementary Table 4). The DIG-QD measurements provided an indication of the extent to which the DNA fluctuations contribute to the diffusion-coefficient measurements (Supplementary Fig. 8). The  $D_{1,obs}$  values for RNAP were all several orders of magnitude lower than values reported for the *lac* repressor, p53 and Mlh1–Pms1 (Fig. 2e, Supplementary Fig. 7 and Supplementary Table 4), which further argued against extensive 1D diffusion contributing to the promoter search. The  $D_{1,obs}$  values for RNAP (~15–100 nm<sup>2</sup> s<sup>-1</sup>; Supplementary Table 4) were indistinguishable from values obtained for stationary DIG-QDs. It is important to recognize that the small  $D_{1,obs}$  values obtained for RNAP cannot be interpreted as protein movement along the DNA but rather arise from the underlying diffusive fluctuations of the DNA itself (Supplementary Fig. 8). We concluded that promoter binding by *E. coli* RNAP is not preceded by microscopically detectable 1D diffusion.

### Intersegmental transfer is not essential for the promoter search

As indicated above, the promoter-search mechanism of *E. coli* RNAP appeared to be dominated by 3D random collisions, with no evidence for facilitated diffusion involving 1D sliding over distances along the DNA greater than our current spatial-resolution limits. Facilitated searches can also potentially occur through intersegmental transfer, which would involve RNAP movement from one distal site to another through a looped DNA intermediate (Supplementary Fig. 1). The DNAs used in our experiments were maintained in a stretched configuration, and we anticipate that they would not support intersegmental transfer because the DNAs cannot form the looped intermediates necessary for this mode of facilitated diffusion. Given that RNAP readily bound promoters in the stretched  $\lambda$  DNA, we concluded that intersegmental transfer cannot be an obligatory component of the promoter-search process. However, we emphasize that we could not rule out the possibility that *E. coli* RNAP might use mechanisms involving intersegmental transfer while searching for promoters on the bacterial chromosome *in vivo*.

### Submicroscopic framework for the promoter-search problem

Given the transient nature of encounters between RNAP and non-specific DNA, we were unable to experimentally determine  $D_{1,obs}$  for the protein bound to non-specific sites. Therefore, we next sought to establish a theoretical framework to investigate the promoter search at the submicroscopic scale. Here, ‘submicroscopic’ is defined as any event occurring below existing spatial and temporal resolution limits. Full treatment of the theory is presented in the **Supplementary Note**; for brevity, we highlight key features and results. We began by recognizing that the flux of RNAP onto the promoters is the result of three components: (i) direct binding to promoters from a fully equilibrated solution (that is, 3D diffusion), (ii) promoter binding from solution after dissociation from another region of DNA (that is, hopping) and (iii) promoter binding after undergoing 1D diffusion along the DNA (that is, sliding). As revealed below, the most important of these terms with respect to the promoter search by *E. coli* RNAP was direct binding from solution, which occurs at a rate of

$$k_{\alpha}^{\psi}(t) = \frac{8}{\pi} \psi D_3 C_0 \int_0^{\infty} e^{-D_3 u^2 t} \left[ u \left( J_0^2(u\rho) + Y_0^2(u\rho) \right) \right]^{-1} du,$$

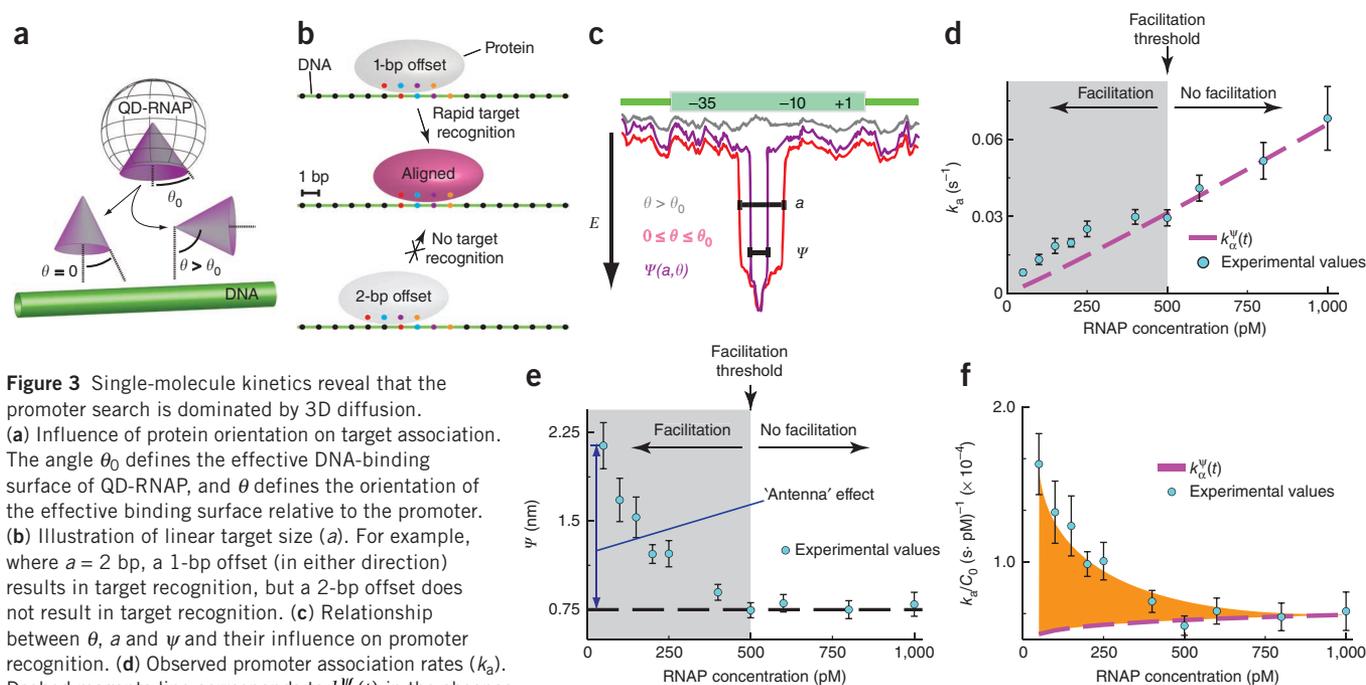
where  $C_0$  is initial protein concentration,  $D_3$  is the 3D-diffusion coefficient of QD-RNAP,  $\psi$  is the effective target size,  $\rho$  is the reaction

radius, and  $J_0$  and  $Y_0$  are Bessel functions of the first and second kind, respectively. The effective target size is a geometric constraint describing the binding surface that transiently samples DNA during the promoter search and is a function of protein orientation ( $\theta$ ) and linear target size ( $a$ ) (Fig. 3a–c)<sup>9,44</sup>. Linear target size should not be confused with promoter length; rather, it describes the range over which a bound protein can be out of register yet still recognize its target (Fig. 3b). An important prediction arising from this formalism is that target-association rates for any protein can become dominated by  $k_{\alpha}^{\psi}(t)$  as  $C_0$  increases, implying that increased protein abundance can obviate any potentially accelerating contributions from facilitated diffusion, regardless of whether the protein in question is capable of hopping and/or sliding along DNA. In simpler terms, what the mathematics revealed was that the probability of directly colliding with a target site increases at higher protein concentration, and any form of search facilitation can be rendered effectively irrelevant by increasing protein abundance because the proteins that reach the target first will do so through 3D diffusion. The question then becomes: at what threshold concentration does 3D diffusion begin to dominate the search? For brevity, we will refer to the concentration at which 3D target binding becomes favored as the facilitation threshold ( $C_{thr}$ ): 3D target binding will be favored when the protein concentration equals or exceeds  $C_{thr}$ , whereas facilitated diffusion will be favored when the concentration is below  $C_{thr}$ . In addition, once facilitated diffusion is removed from the search through increased protein abundance,  $k_{\alpha}^{\psi}(t)$  can be used to recover the effective target size, which in turn provides an estimate of linear target size. These parameters reflect dynamic physical properties of highly transient encounter complexes and cannot be accessed through traditional biochemical analysis of stable or metastable reaction intermediates (closed complexes, open complexes and so forth), nor can they be revealed through structural studies of static protein–nucleic acid complexes. To our knowledge, neither  $\psi$ ,  $a$ , nor  $C_{thr}$  have been experimentally determined for any protein–nucleic acid interaction.

### Single-molecule promoter-search kinetics

By definition, it is not possible to directly visualize submicroscopic events that contribute to target searches. However, we can obtain promoter-association rates from real-time single-molecule measurements (described below). These experimental values can then be compared to the theoretical calculations, allowing us to extract critical features of search mechanisms that are otherwise obscured at the microscopic scale. This provides an independent assessment of the search process that is unhindered by existing spatial or temporal instrument resolution limits. Comparison of the experimental data to values calculated from  $k_{\alpha}^{\psi}(t)$  provides a direct assessment of the promoter-search mechanism, which allows us to determine whether submicroscopic facilitated diffusion contributes to promoter association: if the experimentally observed association rates exceed  $k_{\alpha}^{\psi}(t)$ , then submicroscopic facilitated diffusion must be contributing to the search mechanism; in contrast, if the experimentally observed association rates are equal to  $k_{\alpha}^{\psi}(t)$ , then the search mechanism can be attributed to 3D collisions with no underlying contribution of submicroscopic facilitated diffusion.

We evaluated the potential contribution of submicroscopic facilitated diffusion to the search process by directly measuring promoter-association times over a range of RNAP concentrations within the context of a new DNA-curtain assay designed to separate neighboring molecules by 7  $\mu$ m, thereby eliminating any potential for variation in association kinetics due to differences in local DNA concentration (Fig. 3d and Supplementary Figs. 9–11). Notably, with this



**Figure 3** Single-molecule kinetics reveal that the promoter search is dominated by 3D diffusion. (a) Influence of protein orientation on target association. The angle  $\theta_0$  defines the effective DNA-binding surface of QD-RNAP, and  $\theta$  defines the orientation of the effective binding surface relative to the promoter. (b) Illustration of linear target size ( $a$ ). For example, where  $a = 2$  bp, a 1-bp offset (in either direction) results in target recognition, but a 2-bp offset does not result in target recognition. (c) Relationship between  $\theta$ ,  $a$  and  $\psi$  and their influence on promoter recognition. (d) Observed promoter association rates ( $k_a$ ). Dashed magenta line corresponds to  $k_a^\psi(t)$  in the absence of facilitated diffusion (for  $\psi = 0.75$  nm), and experimental values above this line reflect rate enhancement due to facilitated diffusion. The boundary between the shaded and unshaded regions of the graph represents the facilitation threshold ( $C_{\text{thr}}$ ; as indicated). (e) Effective target size ( $\psi$ ) versus RNAP concentration. The dashed black line highlights the limiting value of  $\psi$ . (f) Rate acceleration ( $k_a/C_0$ ) versus RNAP concentration. The difference between the experimental values and  $k_a^\psi(t)$  reflects facilitated diffusion, and the orange shaded region represents the maximum possible acceleration due to 1D sliding and/or hopping. In d–f, error bars represent s.e.m. ( $n \geq 50$  for each data point).

assay, we were not measuring closed- or open-complex formation; rather, we were measuring the instantaneous time at which single molecules of RNAP are initially detected at a promoter, conditioned upon their subsequent conversion to closed and then open complexes (Supplementary Fig. 11). These measurements were conducted at 100-ms temporal resolution, which is appropriate, given that the slow downstream isomerization steps involved in promoter binding (for example, closed- and open-complex formation) occur on the order of seconds; therefore, any error in determining the initial binding events does not propagate into the measurements of the initial association rates. This experimental format allowed us to precisely define all of the experimental boundary conditions and parameters involved in calculating the predicted promoter association rates from  $k_a^\psi(t)$  (for example, DNA geometry, DNA length, DNA density, number of accessible promoters, protein concentration, solution viscosity, temperature, ionic strength and so forth).

Notably, association rates exceeded  $k_a^\psi(t)$  below 500 pM QD-RNAP, revealing that submicroscopic facilitated diffusion accelerated the promoter search by a factor of three at 50 pM RNAP (Fig. 3d). However, at  $\geq 500$  pM RNAP, association times converged to  $k_a^\psi(t)$ , which indicated that submicroscopic facilitated diffusion did not contribute to the promoter search at higher concentrations (Fig. 3d). Although our results showed that QD-RNAP no longer benefits from facilitated diffusion at concentrations  $\geq 500$  pM, it must be recognized that  $C_{\text{thr}}$  will vary for different proteins and/or different reaction conditions. For example, unlabeled RNAP (hydrodynamic radius  $r = 7.4$  nm)<sup>45</sup> will diffuse more rapidly through solution than QD-RNAP ( $r \approx 13.4$  nm)<sup>46</sup>, so we anticipate that promoter association with unlabeled proteins should converge to  $k_a^\psi(t)$  at an even lower protein concentration, which would be reflected as a reduction in  $C_{\text{thr}}$ . Notably, the physical behavior of RNAP with respect to the search process will not change, regardless of whether the concentration is

above or below  $C_{\text{thr}}$ ; the only thing that changes is the probability of engaging a target through a direct collision ( $P_{3D}$ ) versus the probability of engaging the target after undergoing facilitated diffusion along the DNA ( $P_{FD}$ ).

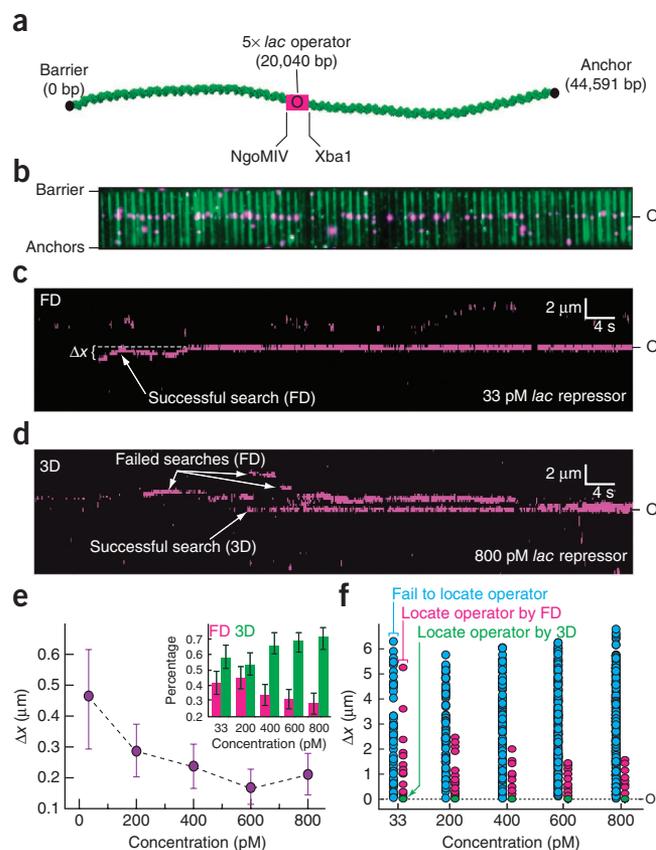
Furthermore, the exact solution of  $k_a^\psi(t)$  yielded an effective target size  $\psi$  of 0.75 nm and an estimated linear target size  $a$  of  $\sim 6$  bp (Fig. 3e)<sup>47</sup>, which indicated that promoters would not be recognized if RNAP is more than  $\pm 3$  bp out of register. The apparent increase in  $\psi$  at low RNAP concentration reflected what is historically referred to as the ‘antenna’ effect<sup>10,22</sup>. At 50 pM RNAP ( $\psi = 2.23$  nm), the antenna was just  $\sim 1.48$  nm (corresponding to  $\sim 6$  bp in our system); the very small size of the antenna indicated the limited contribution that facilitated diffusion (sliding and/or hopping) made to the promoter search even at the lowest RNAP concentrations tested (Fig. 3e,f). An *in vivo* protein concentration of 1 nM corresponds to just 1 protein molecule in a volume the size of an *E. coli* cell<sup>48</sup>; therefore, an *in vivo* concentration of 50 pM would be equivalent to an average of just 1/20<sup>th</sup> of a molecule of RNAP per bacterium, which would not seem physiologically relevant. Taken together, our results demonstrated that although submicroscopic facilitated diffusion can moderately accelerate the promoter search, this acceleration only occurs at exceedingly low RNAP concentrations, whereas at physiologically relevant protein concentrations, the overall promoter search process should be dominated by 3D diffusion.

### Increased protein abundance disfavors facilitated searches

One conclusion arising from our mathematical treatment of target searches is that increased protein abundance will diminish the contribution of facilitated diffusion. This concept is not unique to *E. coli* RNAP and will apply even to proteins that can diffuse long distances along DNA, because the probability of direct collisions ( $P_{3D}$ ) with the target always increases with increasing protein abundance and

**Figure 4** Protein concentration exerts a dominant influence on target searches, even for proteins capable of sliding on DNA. **(a)** DNA schematic showing the location of the 5× *lac* operator (O). **(b)** Two-color image of YOYO-1-stained DNA (green) bound by QD-*lac* repressor (magenta). **(c)** Kymogram showing an example of *lac* repressor binding to nonspecific DNA and then diffusing in 1D to the operator; data were collected at 33 pM *lac* repressor. The distance between the initial binding site and the operator is indicated as  $\Delta x$ . **(d)** Kymogram showing an example of direct operator binding in the absence of any detectable 1D sliding; data were collected at 800 pM *lac* repressor. The successful search through 3D binding is highlighted, as are examples of molecules that searched through facilitated diffusion (FD) but failed to locate the operator. **(e)** Graph showing the mean value of  $\Delta x$  as a function of protein concentration for proteins that successfully engage the operator. Inset, percentage of total operator-binding events that are attributable to facilitated diffusion (magenta) and 3D (green) at each protein concentration. Error bars, s.d. ( $n \geq 54$  for each data point). **(f)** Graph of  $\Delta x$  for all observed proteins. Blue data points correspond to proteins that fail to bind the operator, magenta data points are proteins that bind the operator after undergoing facilitated diffusion, and green data points correspond to 3D binding to the operator. All green data points within each column overlap at zero, but their fractional contribution to operator binding is shown as green bars in the inset of panel **(e)**. These experiments were all conducted in buffer containing 10 mM Tris-HCl, pH 8.0, 1 mM MgCl<sub>2</sub>, 1 mM DTT, and 1 mg ml<sup>-1</sup> BSA.

will eventually exceed the probability of target engagement through facilitated diffusion ( $P_{FD}$ ). As a simple illustration of this point, we used the DNA-curtain assay and a  $\lambda$  DNA-bearing 5-tandem 21-bp symmetric *lac* operator to qualitatively assess target binding by QD-tagged *lac* repressor (Fig. 4a,b)<sup>34</sup>. These experiments were intentionally conducted at low ionic strength, such that nonspecific binding and 1D diffusion were greatly favored, as described<sup>40,41</sup>. At low concentrations, many proteins initially bound to random, nonspecific sites and then diffused thousands of base pairs along the DNA before eventually binding the target; these events were categorized as having occurred through facilitated diffusion (Fig. 4c). Operator binding in the absence of microscopically detectable 1D diffusion was also observed; these events were categorized as 3D (Fig. 4d). For the proteins that successfully engaged the operator, the contribution of facilitated diffusion to the search process is reflected in the distance between the initial binding site and the operator ( $\Delta x$ ) and the change in the ratio of facilitated diffusion to 3D events. As protein concentration increased, the mean value of  $\Delta x$  decreased for the proteins that bound to the operator (Fig. 4e,f), and there was a corresponding increase in the fraction of events categorized as 3D (Fig. 4e, inset). At the highest concentration of *lac* repressor tested (800 pM), ~71% of the total operator-binding events were attributed to 3D diffusion (Fig. 4e, inset). Technical limitations prevented titration to higher protein concentrations, owing to the accompanying increase in background fluorescence, but we anticipate that if the concentration were raised further, eventually all of the operator-binding events would occur through 3D diffusion. This conclusion will even extend into the submicroscopic regime. An in-depth analysis of the facilitation threshold and effective target size for the *lac* repressor (as provided above for RNAP) was beyond the scope of this work; however, the trend in these data clearly illustrated that the contribution of facilitated diffusion diminishes with increased protein abundance, even though the *lac* repressor is capable of sliding great distances on DNA under low-ionic-strength conditions. Notably, at all concentrations tested, many molecules of *lac* repressor bound to random, nonspecific sites all along the length of the  $\lambda$  DNA, and these proteins still exhibited 1D diffusion even when the concentration was raised (Fig. 4d,f); however, as concentration increased, this 1D diffusion



could be considered nonproductive with respect to target association because most of the proteins that bound the operator first did so through 3D collisions (Fig. 4e, inset and Fig. 4d,f).

## DISCUSSION

Our results argue against facilitated diffusion at either the microscopic or submicroscopic scales being a significant contributing component of the *E. coli* RNAP promoter search, and we also show that in general any potential contributions of facilitated diffusion can be overcome through increased protein abundance, even for proteins that can slide long distances on DNA. Facilitated diffusion and 3D collisions can be conceptually considered as two distinct, competing pathways, either of which has the potential to result in target binding, and 3D diffusion will always be favored at protein concentrations equal to or exceeding the facilitation threshold simply because the relative increase in protein abundance increases the probability of a direct collision with the target site (Fig. 5). In other words, just because a protein is physically capable of hopping and/or sliding over long distances along DNA does not mean that these processes will accelerate target binding, because protein concentration can always have a dominating effect on the overall search process. A broader implication of this conclusion is that proteins present at low concentrations in living cells (for example, the *lac* repressor, with fewer than ten molecules per cell) may be more apt to locate targets through facilitated diffusion, whereas those present at higher concentrations (for example, RNAP, ~2,000–3,000 molecules per cell) may be more likely to engage their target sites through 3D diffusion.

Our experimental setting differs substantially from much more complex physiological environments where the promoter search might be influenced by the presence of factors that can assist in the recruitment of RNAP to promoters or by local DNA folding, higher-order chromatin architecture and macromolecular crowding (Fig. 5). Although we cannot yet quantitatively assess the influence of these

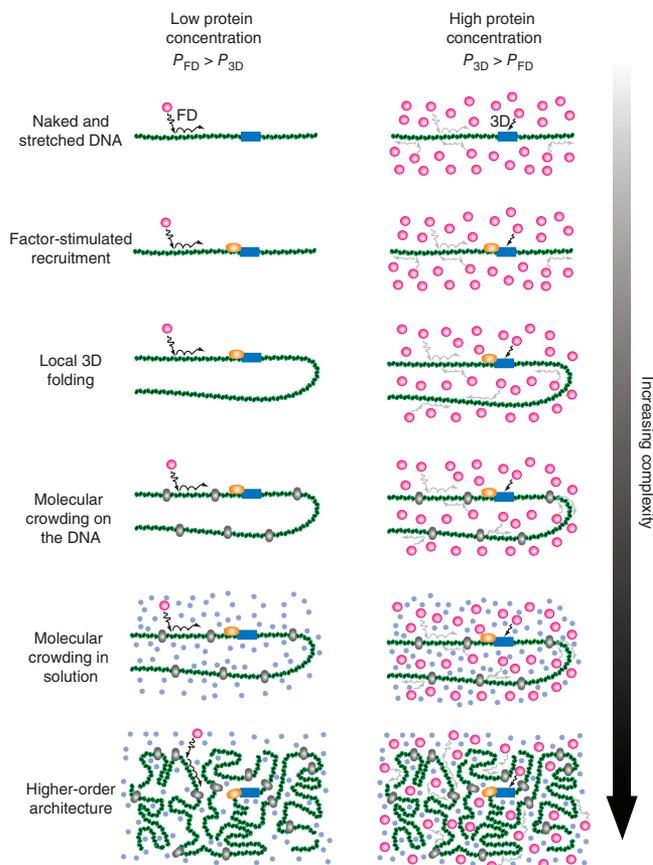
**Figure 5** Increasingly complex environments encountered during *in vivo* searches. Facilitated diffusion (FD) will be favored at concentrations below the facilitation threshold because the initial encounter with the DNA will most often occur at nonspecific sites, so the probability ( $P$ ) of target engagement through facilitated diffusion exceeds the probability of engagement through 3D ( $P_{FD} > P_{3D}$ ). Concentrations equal to or exceeding the facilitation threshold will favor 3D because the relative increase in protein abundance increases the probability of a direct collision with the target site ( $P_{FD} < P_{3D}$ ). Facilitated diffusion–related processes such as sliding and/or hopping can still occur at high protein concentrations, but those proteins undergoing facilitated diffusion are less likely to reach the target site before those that collide directly with the target. Although the facilitation threshold will vary for different proteins and different conditions, higher protein concentrations will still favor 3D collisions, irrespective of the local environment (for example, the presence of recruitment factors, DNA-bound obstacles, macromolecular crowding, local DNA folding) or global DNA architecture.

parameters, we can consider how they might qualitatively affect the promoter search.

Transcriptional activators, such as catabolite activator protein (CAP), are commonly involved in the regulation of gene expression and can exert their effects either by facilitating recruitment of RNAP or by stimulating steps after recruitment (for example, open-complex formation, promoter escape and so forth)<sup>2</sup>. In scenarios involving factor-assisted recruitment, additional protein-protein contacts stabilize interactions between RNAP and the promoter. However, the presence of a transcriptional activator near a promoter should not fundamentally alter the search process by causing RNAP to start sliding and/or hopping along the DNA while executing the search; rather, it would just make the target appear ‘larger’ to RNAP (that is, promoter plus factor, instead of just the promoter), which would in turn reduce the facilitation threshold. Factors that stimulate steps after recruitment would not influence the search because they exert their effects only after the promoter search is complete.

Higher-order organization of DNA *in vivo* has the potential to promote 3D collisions or jumps, but is not expected to favor 1D sliding and/or hopping, both of which can be considered as local events that are not influenced by global DNA architecture<sup>10,16</sup>. In contrast, naked DNA stretched out at low dilution presents the most favorable possible conditions for 1D sliding and/or hopping<sup>14,17</sup>. The fact that we do not detect facilitated diffusion contributing to the promoter search by RNAP under conditions that should otherwise greatly favor hopping and/or sliding suggests these processes are unlikely to occur *in vivo*, simply owing to the more complex 3D DNA environment.

Molecular crowding, either in solution or on the DNA, is a nontrivial issue that can have both positive and negative impacts on DNA binding. Increased nonspecific binding can arise from macromolecular crowding in solution, owing to excluded volume effects<sup>49</sup>, and any increase in nonspecific binding has the potential to promote facilitated diffusion. Although in the case of *E. coli* RNAP, increased nonspecific binding brought about through use of low-ionic-strength conditions still does not lead to microscopically detectable 1D diffusion, which suggests that any increased nonspecific affinity caused by excluded volume effects is unlikely to cause RNAP to start diffusing along DNA. The effects of macromolecular crowding on DNA arise from the presence of other nonspecific DNA-binding proteins, which can reduce nonspecific DNA-binding affinities through competitive inhibition<sup>50</sup> and can also impede 1D diffusion along DNA through steric hindrance<sup>10,15,32</sup>. The net result of the seemingly opposed influences of macromolecular crowding in solution versus molecular crowding on the DNA has yet to be quantitatively explored, although one might anticipate that highly abundant proteins such as Fis and HU (each of which can be present at



concentrations of up to  $\sim 30\text{--}50\ \mu\text{M}$  in *E. coli*) would disfavor facilitated searches by restricting access to nonspecific sites<sup>15</sup>.

In summary, there are at least four reasons why promoter searches in *E. coli* would not benefit from facilitated diffusion. First, there are on the order of  $\sim 2,000\text{--}3,000$  molecules of RNAP in *E. coli*, corresponding to an *in vivo* concentration of  $\sim 2\text{--}3\ \mu\text{M}$ <sup>51</sup>. On the basis of our findings, if even a small fraction of the total RNAP present in a cell were free, then it should still locate promoters through 3D collisions rather than facilitated diffusion. Estimates have suggested that there are on the order of  $\sim 550$  molecules ( $\sim 0.5\ \mu\text{M}$ ) of free  $\sigma^{70}$ -containing RNAP holoenzyme in living bacteria<sup>51</sup>; if these estimates are correct, then the facilitation threshold would have to somehow increase by roughly three orders of magnitude in order for hopping and/or sliding to accelerate the promoter search *in vivo*. In contrast to RNAP, the *lac* repressor, which is thought to employ facilitated diffusion *in vivo* during its target search<sup>41,52</sup>, may need to do so to compensate for its much lower intracellular abundance (fewer than ten molecules per cell) and the corresponding scarcity of its targets (three *lac* operators per genome). Second, long nonspecific lifetimes will lead to slower searches, so RNAP appears to be optimized to avoid wasting time by scanning nonspecific DNA<sup>9–11,14</sup>. Third, other proteins (for example, Fis, HU, IHF, H–NS and so forth) may obstruct 1D diffusion, but such obstacles could be avoided through 3D searches<sup>15</sup>. Fourth, other steps are rate limiting during gene expression (for example, promoter accessibility, promoter escape, elongation and so forth)<sup>29,53–55</sup>, which suggests that there is simply no need for RNAP to locate promoters faster than the 3D-diffusion limit. Finally, despite the much more complicated environments present in physiological settings, our general conclusion regarding the effects of protein abundance on target searches should remain qualitatively true because higher protein concentrations will increase the probability of direct target binding through 3D collisions.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Supplementary information is available in the [online version of the paper](#).

## ACKNOWLEDGMENTS

We thank M. Gottesman, R. Gonzalez and the Greene laboratory for assistance and discussion throughout this work. We thank M. Gottesman, D. Duzdevich and members of our laboratories for carefully reading the manuscript. We thank R. Landick (University of Wisconsin–Madison, Madison, Wisconsin, USA) for providing RNAP expression constructs. This work was supported by US National Institutes of Health grants GM074739 (E.C.G.) and F32GM80864 (I.J.F.) and training grant T32GM00879807 (J.G.), a US National Science Foundation Award (E.C.G.) and by the Howard Hughes Medical Institute.

## AUTHOR CONTRIBUTIONS

F.W. collected the RNAP experimental data, and F.W. and S.R. analyzed the data. S.R. developed the theoretical analysis, conducted theoretical calculations, assisted with the RNAP data collection and collected the data for the *lac* repressor. D.R.R. assisted with development of the theory. I.J.F. assisted in establishing the single-molecule assays for QD-RNAP and QD-*lac* repressor. J.G. developed the substrate for the DIG-QD measurements and collected the corresponding data. E.C.G. supervised the project, and all authors co-wrote the paper.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/nsmb.2472>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Haugen, S.P., Ross, W. & Gourse, R.L. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat. Rev. Microbiol.* **6**, 507–519 (2008).
- Browning, D.F. & Busby, S.J. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**, 57–65 (2004).
- Saecker, R.M., Record, M.T. & Dehaseth, P.L. Mechanism of bacterial transcription initiation: RNA polymerase–promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* **412**, 754–771 (2011).
- Nudler, E. RNA polymerase active center: the molecular engine of transcription. *Annu. Rev. Biochem.* **78**, 335–361 (2009).
- Mendoza-Vargas, A. *et al.* Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE* **4**, e7526 (2009).
- Cho, B.K. *et al.* The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.* **27**, 1043–1049 (2009).
- von Hippel, P.H. & Berg, O.G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
- Gorman, J. & Greene, E.C. Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* **15**, 768–774 (2008).
- 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).
- Mirny, L. *et al.* How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A* **42**, 434013 (2009).
- Halford, S.E. & Marko, J.F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
- 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).
- Riggs, A.D., Bourgeois, S. & Cohn, M. The *lac* repressor–operator interaction. 3. Kinetic studies. *J. Mol. Biol.* **53**, 401–417 (1970).
- Halford, S.E. An end to 40 years of mistakes in DNA–protein association kinetics? *Biochem. Soc. Trans.* **37**, 343–348 (2009).
- Li, G.-W., Berg, O.G. & Elf, J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.* **5**, 294–297 (2009).
- 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).
- Bauer, M. & Metzler, R. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* **102**, 2321–2330 (2012).
- Kolesov, G., Wunderlich, Z., Laikova, O.N., Gelfand, M.S. & Mirny, L.A. How gene order is influenced by the biophysics of transcription regulation. *Proc. Natl. Acad. Sci. USA* **104**, 13948–13953 (2007).
- Wunderlich, Z. & Mirny, L.A. Spatial effects on the speed and reliability of protein–DNA search. *Nucleic Acids Res.* **36**, 3570–3578 (2008).
- Das, R.K. & Kolomeisky, A.B. Facilitated search of proteins on DNA: correlations are important. *Phys. Chem. Chem. Phys.* **12**, 2999–3004 (2010).
- Singer, P. & Wu, C.W. Promoter search by *Escherichia coli* RNA polymerase on a circular DNA template. *J. Biol. Chem.* **262**, 14178–14189 (1987).
- Ricchetti, M., Metzger, W. & Heumann, H. One-dimensional diffusion of *Escherichia coli* DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. *Proc. Natl. Acad. Sci. USA* **85**, 4610–4614 (1988).
- Kabata, H. *et al.* Visualization of single molecules of RNA polymerase sliding along DNA. *Science* **262**, 1561–1563 (1993).
- Guthold, M. *et al.* Direct observation of one-dimensional diffusion and transcription by *Escherichia coli* RNA polymerase. *Biophys. J.* **77**, 2284–2294 (1999).
- Harada, Y. *et al.* Single-molecule imaging of RNA polymerase–DNA interactions in real time. *Biophys. J.* **76**, 709–715 (1999).
- Berg, J., Tymoczko, J. & Stryer, L. *Biochemistry* (W.H. Freeman, 2007).
- Roe, J.H., Burgess, R.R. & Record, M.T. Jr. Kinetics and mechanism of the interaction of *Escherichia coli* RNA polymerase with the lambda PR promoter. *J. Mol. Biol.* **176**, 495–522 (1984).
- Friedman, L.J. & Gelles, J. Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. *Cell* **148**, 679–689 (2012).
- deHaseth, P.L., Zupancic, M. & Record, M.T. Jr. RNA polymerase–promoter interactions: the comings and goings of RNA polymerase. *J. Bacteriol.* **180**, 3019–3025 (1998).
- Herbert, K.M., Greenleaf, W.J. & Block, S.M. Single-molecule studies of RNA polymerase: motoring along. *Annu. Rev. Biochem.* **77**, 149–176 (2008).
- Gorman, J., Fazio, T., Wang, F., Wind, S. & Greene, E.C. Nanofabricated racks of aligned and anchored DNA substrates for single-molecule imaging. *Langmuir* **26**, 1372–1379 (2010).
- Gorman, J., Plys, A.J., Visnapuu, M.L., Alani, E. & Greene, E.C. Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.* **17**, 932–938 (2010).
- Gorman, J. *et al.* Single-molecule imaging reveals target search mechanisms during mismatch repair. *Proc. Natl. Acad. Sci. USA* **109**, E3074–E3083 (2012).
- Finkelstein, I.J., Visnapuu, M.L. & Greene, E.C. Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature* **468**, 983–987 (2010).
- Simons, R.W., Hoopes, B.C., McClure, W.R. & Kleckner, N. Three promoters near the terminus of IS10: pIN, pOUT, and pIII. *Cell* **34**, 673–682 (1983).
- McClure, W.R. Rate-limiting steps in RNA chain initiation. *Proc. Natl. Acad. Sci. USA* **77**, 5634–5638 (1980).
- Hawley, D.K. & McClure, W.R. *In vitro* comparison of initiation properties of bacteriophage lambda wild-type PR and x3 mutant promoters. *Proc. Natl. Acad. Sci. USA* **77**, 6381–6385 (1980).
- Dayton, C.J., Prosen, D.E., Parker, K.L. & Cech, C.L. Kinetic measurements of *Escherichia coli* RNA polymerase association with bacteriophage T7 early promoters. *J. Biol. Chem.* **259**, 1616–1621 (1984).
- Brunner, M. & Bujard, H. Promoter recognition and promoter strength in the *Escherichia coli* system. *EMBO J.* **6**, 3139–3144 (1987).
- Wang, Y.M., Austin, R.H. & Cox, E.C. Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.* **97**, 048302 (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).
- Kim, J.H. & Larson, R.G. Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules. *Nucleic Acids Res.* **35**, 3848–3858 (2007).
- Tafvizi, A., Huang, F., Fersht, A.R., Mirny, L.A. & van Oijen, A.M. A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. USA* **108**, 563–568 (2011).
- Berg, O.G. Orientation constraints in diffusion-limited macromolecular association. The role of surface diffusion as a rate-enhancing mechanism. *Biophys. J.* **47**, 1–14 (1985).
- Austin, R.H., Karohl, J. & Jovin, T.M. Rotational diffusion of *Escherichia coli* RNA polymerase free and bound to deoxyribonucleic acid in nonspecific complexes. *Biochemistry* **22**, 3082–3090 (1983).
- Gorman, J. *et al.* Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2–Msh6. *Mol. Cell* **28**, 359–370 (2007).
- Berg, O.G. & von Hippel, P.H. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Chem.* **14**, 131–160 (1985).
- Moran, U., Phillips, R. & Milo, R. Snapshot: key numbers in biology. *Cell* **141**, 1262–1262.e1 (2010).
- Minton, A.P. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **276**, 10577–10580 (2001).
- Graham, J.S., Johnson, R.C. & Marko, J.F. Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA. *Nucleic Acids Res.* **39**, 2249–2259 (2011).
- Ishihama, A. Functional modulation of *Escherichia coli* RNA polymerase. *Annu. Rev. Microbiol.* **54**, 499–518 (2000).
- Hammar, P. *et al.* The *lac* repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
- McClure, W.R. Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* **54**, 171–204 (1985).
- So, L.H. *et al.* General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet.* **43**, 554–560 (2011).
- Reppas, N.B., Wade, J.T., Church, G.M. & Struhl, K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell* **24**, 747–757 (2006).

## ONLINE METHODS

Single-molecule experiments were conducted on a custom-built total internal reflection microscope and used double-tethered DNA curtains<sup>31,32</sup>. RNAP was expressed with a biotinylation peptide fused to the C terminus of  $\beta'$  and purified as described<sup>34,56</sup>. Biotinylated RNAP holoenzyme was labeled with streptavidin-QDs (Qdot 705,  $r = 12.6$  nm; Invitrogen). Prior to use, flow cells were flushed with transcription buffer (40 mM Tris, pH 8.0, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT and 2 mg ml<sup>-1</sup> BSA) supplemented with 250 pM YOYO-1 and 9  $\mu$ M free biotin to block the surface. QD-tagged RNAP was then diluted into biotin-supplemented transcription buffer ( $\pm$  NTP, 250  $\mu$ M each) to a final concentration of 30–200 pM, and then a 50- $\mu$ l sample was injected into the flow cell at a rate of 0.1 ml min<sup>-1</sup>, and buffer flow was terminated 120 s after beginning the injection. For initial promoter-binding measurements, experiments were conducted in buffer containing 20 mM Tris-HCl, pH 8.0, 25 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT and 2 mg ml<sup>-1</sup> BSA, and images were acquired at 5–100 Hz by using NIS-Elements software (Nikon). For the measurements of promoter-association rates, the fraction of active protein in each preparation was first determined from gel-shift assays using a Cy3-labeled 249-bp DNA fragment containing

the  $\lambda$ P<sub>R</sub> promoter. Gel shifts were conducted under dilute protein conditions equivalent to those used in the single-molecule assays to ensure that ensemble protein activity reflected that in the single-molecule assays. All single-molecule kinetic measurements to determine promoter-association rates used a new type of DNA curtain designed to avoid local DNA concentration effects (**Supplementary Fig. 10**). QD-RNAP (700  $\mu$ l) was then injected into the flow cell at 0.5 ml min<sup>-1</sup> in buffer containing 20 mM Tris-HCl, pH 8.0, 25 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT and 2 mg ml<sup>-1</sup> BSA, by using a standardized sample-injection procedure that eliminated variability in observed association rates due to microfluidic heterogeneities and variations in protein concentration profiles. Images were collected at 10 Hz, and initial promoter-association rates were then obtained by measuring dwell times between successive promoter-binding events for all different DNA molecules within the field of view (**Supplementary Note and Supplementary Fig. 11**).

56. Shaeviz, J.W., Abbondanzieri, E.A., Landick, R. & Block, S.M. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426**, 684–687 (2003).