Dynamics of chromosomal target search by a membrane-integrated one-component receptor

Linda Martini[1‡], Sophie Brameyer[2‡], Elisabeth Hoyer[2], Kirsten Jung[2*], Ulrich Gerland[1*]

1 Physics of Complex Biosystems, Technical University of Munich, James-Franck-Str. 1, 85748 Garching, Germany
2 Microbiology, Ludwig-Maximilians-University Munich, Großhaderner Str. 2-4, 82152 Martinsried, Germany

‡These authors contributed equally to this work.
* Corresponding authors
E-mail: gerland@tum.de, jung@lmu.de

Abstract

While the classical DNA target search problem of cytoplasmic transcription factors is well studied, it remains unclear how a membrane-integrated protein can kinetically establish contact with its specific binding site on the genome. Here, we address this question for the case of CadC, the membrane-integrated pH-receptor of the acid stress response Cad system in *E. coli*. CadC is a prime example of a one-component signaling protein that directly binds to its cognate target site on the chromosome to regulate transcription. We combined fluorescence microscopy experiments, mathematical analysis, and kinetic Monte Carlo simulations to probe this target search process. Using fluorescently labeled CadC, we measured the time from activation of the receptor until successful binding to the DNA in single cells, exploiting that stable receptor-DNA complexes are visible as fluorescent foci. Our experimental data indicate that CadC is highly mobile in the membrane and finds its target by 2D diffusion. DNA mobility is constrained due to the overall chromosome organization, but a labeled DNA locus in the vicinity of the target site appears sufficiently mobile to randomly come close to the membrane. Relocation of the DNA target site to a distant position on the chromosome had almost no effect on the mean search time, which was between four and five minutes in either case. However, a mutant strain with two binding sites displayed a mean search time that was reduced by about a factor of two. This behavior is consistent with simulations of a coarse-grained lattice model for the coupled dynamics of DNA within a cell volume and proteins on its surface. The model also rationalizes the experimentally determined distribution of search times. Overall our findings reveal that DNA target search does not present a much bigger kinetic challenge for membrane-integrated proteins than for cytoplasmic proteins.

Author summary

Adaptation to changing environments is vital to bacteria and is enabled by sophisticated signal transduction systems. While classical two-component signaling is well studied, the dynamics of one-component signaling, where one protein performs both sensing and response regulation, are insufficiently understood. Specifically, how can a

membrane-integrated protein bind to specific sites on the genome to regulate transcription? Here, we study the kinetics of this process, which involves both protein diffusion within the membrane and conformational fluctuations of the genomic DNA. A well-suited model system for this question is CadC, the signaling protein of the *E. coli* Cad system involved in pH stress response. Fluorescently labeled CadC forms visible foci in single cells upon stable DNA-binding, marking the end of the protein-DNA search process. Moreover, the start of the search is triggered by a medium shift exposing cells to pH stress. We probe the underlying mechanism by varying the number and position of DNA target sites. We combine these experiments with mathematical analysis and kinetic Monte Carlo simulations of lattice models for the search process. Our results suggest that CadC diffusion in the membrane is pivotal for this search, while the DNA target site is just mobile enough to reach the membrane.

Introduction   1

Bacteria are exposed to fluctuating environments with frequent changes in nutrient   2
conditions and communication signals, but also life-threatening conditions such as   3
environmental stresses and antibiotics [1]. To sense and adapt to changing   4
environmental conditions, bacteria have evolved sophisticated signaling schemes,   5
primarily based on one- and two-component systems [2,3].   6

    Two-component signaling systems feature a sensor kinase and a separate response   7
regulator, where the former is typically membrane-integrated while the latter diffuses 8
through the cytoplasm to reach its regulatory target [3]. The majority of response   9
regulators are transcription factors that bind to specific target sites on the genomic 10
DNA to activate or repress transcription. Hence, a key step in the signal transduction 11
pathway of these two-component systems is a DNA target search by a cytoplasmic   12
protein. The target search dynamics of cytoplasmic transcription factors have been   13
thoroughly studied in the past decades, triggered by early *in vitro* experiments   14
indicating that the *Escherichia coli lac* repressor finds its target site faster than the rate 15
limit for three-dimensional (3D) diffusion [4]. Inspired by the idea that a reduction of 16
dimensionality can lead to enhanced reaction rates [5], the experiments were explained 17
by a two-step process, where transcription factors locate their target by alternating   18
periods of 3D diffusion and 1D sliding along the DNA [6]. Compared to pure 3D   19
diffusion, sliding increases the association rate by effectively enlarging the target size 20
("antenna" effect) [7–9]. These dynamics were later probed with single-molecule   21
methods, both *in vitro* [10,11] and *in vivo* [12,13]. For the *lac* repressor as a   22
paradigmatic example of a low copy number cytoplasmic transcription factor, the *in*   23
vivo timescale of the target search was found to be about one minute [12]. Further   24
studies continued to add to the detailed understanding of this target search process, e.g. 25
with respect to effects of DNA conformation [14], DNA dynamics [15], and   26
macromolecular crowding [16].   27

    One-component signaling systems, in contrast, combine sensory function and   28
response regulation within one protein [2]. The subset of one-component systems that   29
are both membrane-integrated sensors and DNA-binding response regulators face an   30
extraordinary DNA target search problem: They must locate and bind to a specific site 31
on the bacterial chromosome from the membrane. This is the case for one-component   32
systems of the ToxR-receptor family, which have a modular structure featuring a   33
periplasmic sensory domain followed by a single transmembrane helix connected via a   34
linker to a cytoplasmic DNA-binding domain [18]. In addition to ToxR, the main   35
regulator for virulence in *Vibrio cholerae*, members of this receptor family include TcpP 36
and TfoS in *V. cholerae* [19], PsaE in *Yersinia tuberculosis* [20] and the pH   37
stress-sensing receptor CadC in *Escherichia coli* [21]. *A priori*, it is not clear how these 38
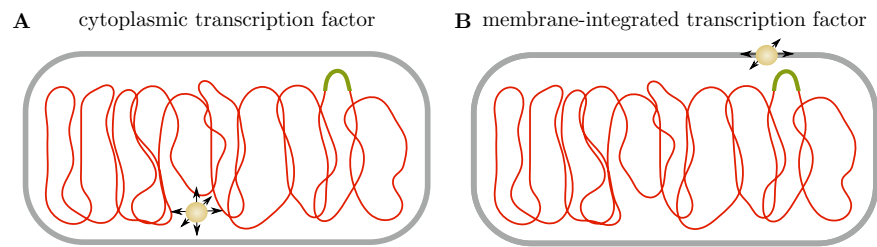
Fig 1. DNA target search of cytoplasmic vs. membrane proteins. (A) A cytoplasmic transcription factor (yellow) locates its specific binding site (green) on the DNA (red) through a combination of 3D diffusion in the cytoplasm and 1D sliding along the polymer. While the DNA itself is not static during this target search, DNA motion does not significantly contribute to its completion. **(B)** In contrast, a membrane-integrated transcription factor can only perform 2D diffusion in the membrane, such that DNA motion becomes essential. At a minimum, the specific DNA-binding site has to move close to the membrane to enable target recognition.

one-component systems establish functional protein-DNA contacts after stimulus perception. One scenario is that the DNA-binding domain is proteolytically cleaved, such that it can search for its specific binding site in the same manner as a regular cytoplasmic transcription factor. Another possible scenario is that simultaneous transcription, translation, and membrane-insertion ("transertion" [22, 23]) tethers the DNA locus of the one-component system to the membrane and, at the same time, places the protein in the vicinity of its binding site on the DNA (which is typically close to the gene encoding the one-component system). The third scenario is a diffusion and capture mechanism [24], whereby the one-component system diffuses within the membrane and captures conformational fluctuations that bring the DNA close to the membrane.

For the case of *E. coli* CadC, a well characterized model system [25], indirect evidence [26, 27] paired with direct observation [28] argues against the proteolytic cleavage and transertion scenarios, and instead supports the diffusion and capture mechanism. For instance, the transertion mechanism should be sensitive to relocating the *cadC* gene to a locus far from its native position, which is close to the CadC target site at the *cadBA* promoter. However, relocation to the distant *lac* operon did not reduce the regulatory output of CadC. One of the observations arguing against proteolytic cleavage was that external signals can rapidly deactivate the CadC response after the original stimulus, whereas cleavage would irreversibly separate the DNA-binding domain from the signaling input [26]. In contrast, the diffusion and capture mechanism appeared consistent with experiments that imaged fluorophore-labeled CadC *in vivo* [28]. These experiments showed that localized CadC spots in fluorescent microscopic images form after cells were shifted to a medium that simultaneously provides acid stress and a lysine-rich environment, the two input signals required to stimulate the CadC response [26]. Formation of these spots was rapidly reversible upon removal of the input signals. Furthermore, the observed number of CadC spots was positively correlated with the number of DNA binding sites, indicating that the spots correspond to CadC-DNA complexes with a much lower mobility than freely diffusing CadC in the membrane.

Taken together, the existing data suggest that the one-component system CadC establishes the protein-DNA contact required for transcription regulation not by a conventional target search akin to cytoplasmic transcription factors (Fig. 1A), but instead by 2D diffusion of the protein in the membrane and fluctuations of the DNA conformation that occasionally bring the DNA region of the target site close enough to the membrane to be captured by the protein (Fig. 1B). Intuitively, a successful diffusion

and capture event seems highly unlikely. However, that such events occur was independently demonstrated in an experiment that artificially tethered the *lac* repressor to the cell membrane [29]. This construct was indeed able to inducibly repress transcription from a chromosomal reporter. Hence, the salient questions are how this type of target search is kinetically feasible and on which timescale. Here, we address these questions using *E. coli* CadC as a model system. We measure the search time of CadC to its target DNA-binding site in single cells by probing the formation of fluorescent CadC spots at different times. This yields experimental search time distributions that we compare against so-called 'first passage time' distributions obtained from stochastic models [30]. We also measure the mobility of a chromosomal locus in our experimental setup using a fluorescent repressor/operator system to inform kinetic Monte Carlo simulations of the target search dynamics. These simulations are able to reproduce the experimental behavior and to elucidate properties of the search process that we cannot obtain experimentally.

Results

Choice of CadC as experimental model system

CadC is a particularly well studied membrane-integrated one-component receptor. It is part of a pH stress response system, which is also dependent on signaling input from the lysine-specific permease LysP [26]. The function of the Cad system is to alleviate acidic stress by activating synthesis of the lysine/cadaverine antiporter CadB and the lysine decarboxylase CadA, which import lysine, convert it to cadaverine, and export cadaverine. CadA and CadB are encoded by the *cadBA* operon, which is transcriptionally upregulated by specific binding of CadC upstream of the promoter [31], see Fig 2A (inset). CadC is known to respond to and integrate three different signals: It is activated by an acidic pH in the presence of external lysine and inhibited by cadaverine. The pH-sensory function as well as the feedback inhibition by cadaverine was assigned to distinct amino acids within the periplasmic sensory domain of CadC [27,32], whereas the availability of external lysine is transduced to CadC via the co-sensor and inhibitor LysP, a lysine-specific transporter [33,34]. High-affinity DNA-binding of CadC requires CadC homodimerization, which is inhibited by LysP via intramembrane and periplasmic contacts under non-inducing conditions [33,34]. A drop in external pH induces dimerization of the periplasmic sensory domain of CadC followed by structural rearrangement of its cytoplasmic linker, permitting the DNA-binding domain of CadC to homodimerize [21,32,35]. The CadC protein number is extremely low (1–3 molecules on average per cell [28]), mainly due to a low translation rate caused by polyproline stalling, which is only partially relieved by elongation factor P [37].

In the present study, CadC serves as an experimental model system to investigate the DNA target search of a membrane-integrated transcription factor (Fig 2A). To probe the kinetics of this search process in individual cells, it is crucial to have a well-defined initial state and clear 'start' and 'stop' events. When cells are initially grown in a medium with neutral pH, CadC is inactivated and homogeneously distributed in the membrane [28]. A sudden medium shift to low pH and rich lysine conditions then serves as a suitable 'start' trigger, which activates CadC via the signal transduction mechanism described above and starts the target search process. However, detecting the successful termination of the target search is challenging. Here, we exploit the previous finding that the formation of stable CadC-DNA complexes is visible as distinct spots in fluorescence microscopy images [28]. The study excluded that spot formation occurs solely due to low pH, using a pH-independent CadC variant that showed spots at both neutral and low pH. The connection between spot formation and

CadC-DNA binding was derived from the observation that no spots were formed when CadC was rendered unable to bind DNA, and the number of spots per cell correlated with the number of DNA binding sites for CadC. Without DNA binding site only 20 % of cells formed spots, possibly due to non-specific DNA binding. We take this into account in our quantitative analysis below.

Experimental measurement of CadC target search times

We used three *E. coli* strains with different binding site conigurations along the chromosome: N-P$_{cadBA}$ (wild-type) with the native DNA-binding site close to *ori* (at 93.9′ [38]), T-P$_{cadBA}$ with the binding site relocated to the terminus, and N+T-P$_{cadBA}$ with both binding sites (Fig 2B). To visualize the temporal and spatial localization of CadC *in vivo*, we transformed each of the strains with plasmid-encoded mCherry-tagged CadC, thereby increasing the average number of CadC molecules per cell to 3–5 [28]. After the medium shift at $t = 0$ min, we took fluorescence and phase contrast microscopy images of cells sampled from the same culture every minute. We used image analysis tools to detect fluorescent spots within the cells, evaluating between 859 and 2506 cells per time step, see 'Materials and methods'. Based on this data, we determined the fraction of cells with at least one fluorescent spot, $\nu(t) = N_{\text{cells with spot}}(t)/N_{\text{cells}}(t)$ at each time t after CadC activation. To take into account the initial fraction of spots attributed to non-specific DNA binding (see above), we defined the response function

$$R(t) = \frac{\nu(t) - \nu(0)}{\nu(\infty) - \nu(0)} , \tag{1}$$

which rises from zero to one. Here, the asymptotic value $\nu(\infty)$ accounts for the fact that fluorescent spots are never detected in all cells (see the raw data in S1 Fig). This is likely due to the heterogeneous distribution of CadC [28]: Given the low average copy number, some cells are expected to have less than the two molecules required for dimerization. Additionally, some spots may have been missed by the spot detection algorithm, in particular for cells that were not perfectly in focus. The time-dependent response $R(t)$ for our three strains is shown in Fig 2C, with examples of fluorescence microscopy images of cells at $t = 0$ min and the last time point in Fig 2D. Before analyzing the experimental response functions, we discuss the description of the target search as a stochastic process and derive theoretical response functions for comparison with the experimental data.

The target search as a stochastic process

On a coarse-grained level, the search of CadC for a specific binding site on the DNA can be described by a stochastic process with a small number of discrete states. Since both CadC and the DNA must move in order to establish a specific protein-DNA contact, it is reasonable to assume a reversible sequential process with an intermediate state,

$$S_1 \underset{k_2^-}{\overset{k_1^+}{\rightleftharpoons}} S_2 \xrightarrow{k_2^+} S_3 . \tag{2}$$

Here, state S_1 corresponds to configurations where the DNA target site is not in direct vicinity of the membrane, and CadC is delocalized on the membrane, unbound to the DNA. State S_3 corresponds to the final state where CadC is bound to a specific target site. The intermediate state S_2 could then correspond to configurations where the DNA segment containing the target site is close to the membrane, but CadC is not bound to this segment. The transition rates between these states are denoted as k_1^+, k_2^+, and k_2^-. We are interested in the so-called 'first-passage time' τ to reach the final state s_3, which
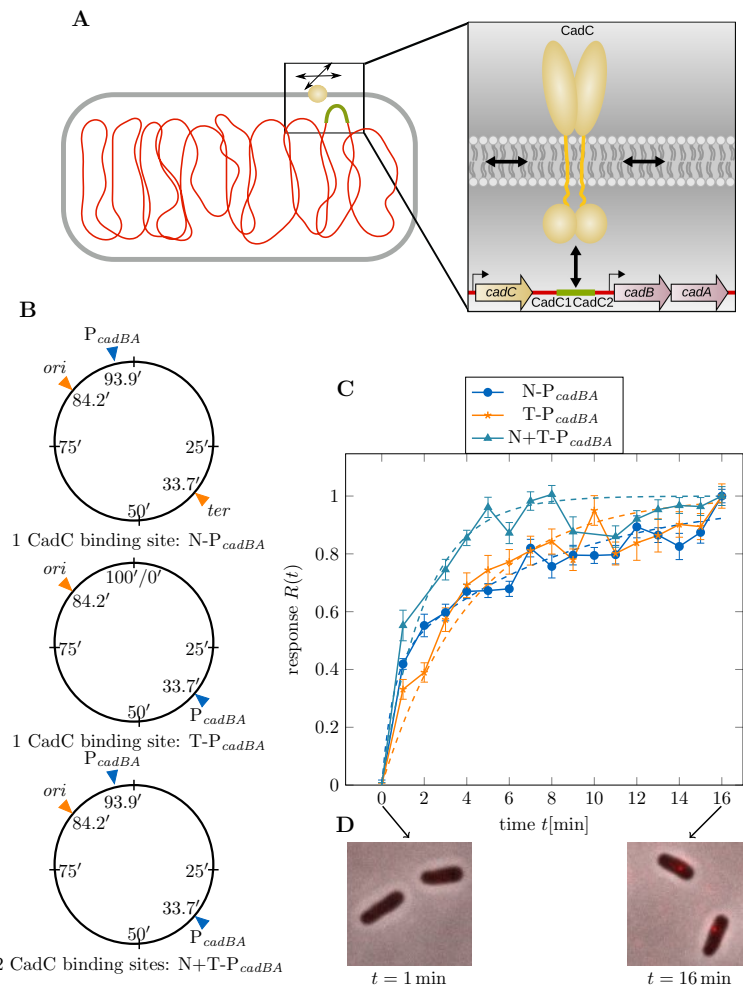
Fig 2. Dynamics of the target search by CadC. (**A**) The target search of membrane-integrated transcription factors is investigated by experimentally measuring the response time of CadC. The molecular model on the right shows CadC in yellow, diffusing in the membrane and forming a dimer to bind to the two CadC binding sites (CadC1 and CadC2), displayed in green. The *cadC* gene is located upstream of the *cadBA* operon. (**B**) Three *E. coli* strains with different positions of the *cadBA* promoter: N-P$_{cadBA}$ with the native DNA-binding site close to *ori*, T-P$_{cadBA}$ with the binding site at the terminus and N+T-P$_{cadBA}$ with both binding sites. (**C**) Experimental results from CadC spot detection. Fluorescent microscopic images were taken every minute after receptor activation and analyzed for CadC foci for the strains defined in panel B. The plot shows the response $R(t)$, defined as the normalized fraction of cells as a function of time t after the medium shift. Error bars correspond to the propagated standard deviation of $\nu(t)$ from averaging over multiple data sets. The dashed lines in the plot show a fit of the response function to the CDF of a sequential reversible two-step model with mixed initial condition for N-P$_{cadBA}$, shown in blue dots and with fixed initial condition for T-P$_{cadBA}$ (orange stars) and N+T-P$_{cadBA}$ (cyan triangles). (**D**) The fluorescence microscopy images demonstrate how fluorescent spots appear after receptor activation.

corresponds to the target search time within this coarse-grained description. The
probability distribution $p(\tau)$ for this time is calculated by making the final state
absorbing, using standard techniques [30]. Assuming that the system is initially in state
s_1, the first passage time distribution is

$$p(\tau) = \frac{e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}}}{\alpha - \beta} \ , \tag{3}$$

where the two timescales α and β of the exponential functions are related to the
transition rates via

$$\alpha \ = \ 2\left(k_1^+ + k_2^+ + k_2^- - \sqrt{(k_1^+ + k_2^+ + k_2^-)^2 - 4k_1^+ k_2^+}\right)^{-1} \ , \tag{4}$$

$$\beta \ = \ 2\left(k_2^- + k_1^+ + k_2^+ + \sqrt{(k_2^- + k_1^+ + k_2^+)^2 - 4k_1^+ k_2^+}\right)^{-1} \ , \tag{5}$$

implying that $\alpha > \beta$. At large times, the distribution $p(\tau)$ decays exponentially (decay
time α), whereas the timescale β corresponds to a delay at short times. Hence,
increasing α leads to a slower decay and increasing β to a longer delay. The average
value of this distribution, referred to as the mean first passage time, is

$$\langle \tau \rangle = \alpha + \beta = \left(1 + \frac{k_2^-}{k_2^+}\right)\frac{1}{k_1^+} + \frac{1}{k_2^+} \ . \tag{6}$$

This corresponds to the average time for the first step $(1/k_1^+)$ multiplied by the average
number of trials needed to reach the second step, plus the average time for the second
reaction step $(1/k_2^+)$. In cases where one of the reaction steps is rate limiting, the mean
first passage time is simply the inverse of the limiting rate, and the process simplifies to
a two-state process with an exponentially distributed first passage time.

To relate the first passage time distribution to our experimental response function,
Eq 1, we consider the cumulative distribution function $\mathrm{CDF}(\tau)$ defined by

$$\mathrm{CDF}(\tau) = \int_0^\tau p(t)\mathrm{d}t \ , \tag{7}$$

which is the probability that the first passage time is less than or equal to τ.
Experimentally, $\mathrm{CDF}(\tau)$ corresponds to the fraction of cells in which the target search
was successful by time τ. The response function in Eq 1 is our best proxy for this
fraction of cells, and hence we identify

$$\mathrm{CDF}(\tau) \mathrel{\hat{=}} R(\tau) \ . \tag{8}$$

The cumulative distribution function for $p(\tau)$ of Eq 3 is

$$\mathrm{CDF}(\tau) = 1 - \frac{\alpha e^{-\frac{\tau}{\alpha}} - \beta e^{-\frac{\tau}{\beta}}}{\alpha - \beta} \ , \tag{9}$$

which we can use as a fit function to describe the experimental data. However, using
Eq 9 amounts to the assumption that all cells are initially in state s_1. If we allow for
the possibility that some cells are in state s_2 when the CadC is activated, we have a
mixed initial condition, where the process starts either from state S_1 or from state S_2.
Denoting the fraction of cells that are initially in state S_2 by x, such that a fraction
$1 - x$ starts in state S_1, we obtain a first passage time distribution and associated

Table 1. Moments of the measured search processes.

Strain	$\langle\tau\rangle$[min]	σ^2[min^2]	s[min^2]	k[min^2]
N-P$_{cadBA}$	4.84 ± 0.19	49.6 ± 8.8	2.537 ± 0.082	11.98 ± 0.53
T-P$_{cadBA}$	4.20 ± 0.15	17.6 ± 2.0	$2.0 \pm 3.2 \times 10^{-16}$	$9.0 \pm 1.5 \times 10^{-15}$
N+T-P$_{cadBA}$	2.02 ± 0.12	4.09 ± 0.66	$2.0 \pm 3.6 \times 10^{-15}$	$9.0 \pm 1.4 \times 10^{-14}$

Results from fitting the experimentally computed CDFs to the sequential reversible model with mixed initial condition (N-P$_{cadBA}$) and fixed initial condition (T-P$_{cadBA}$ and N+T-P$_{cadBA}$). The fit parameters α, β and c were used to compute the mean first passage time $\langle\tau\rangle$, the variance σ^2, the skewness s and the kurtosis k with uncertainties obtained from error propagation using the full covariance matrix.

cumulative distribution function of the form

$$p(\tau) = \frac{e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}} + c\left(\alpha e^{-\frac{\tau}{\beta}} - \beta e^{-\frac{\tau}{\alpha}}\right)}{\alpha - \beta} \, ,$$

$$\mathrm{CDF}(\tau) = 1 - \frac{\alpha e^{-\frac{\tau}{\alpha}} - \beta e^{-\frac{\tau}{\beta}} - \alpha\beta c\left(e^{-\frac{\tau}{\alpha}} - e^{-\frac{\tau}{\beta}}\right)}{\alpha - \beta} \, ,$$

(10)

where the new parameter c corresponds to $c = x k_2^+$. The associated mean first passage time is

$$\langle\tau\rangle = \alpha + \beta - \alpha\beta c \, .$$

(11)

The mean search time is less than 5 min and not affected by relocation of the target site

We first analyzed the experimental response $R(t)$ in the wild-type strain. The wild-type data (N-P$_{cadBA}$) in Fig 2C (blue dots) shows an initially fast increase, followed by a more gradual saturation. This behavior is not well described by Eq 9, but captured by Eq 10 with the mixed initial condition, as can be seen by the fit to the data represented by the blue dashed line in Fig 2C. From the fit parameters α, β, and c, we computed the mean search time according to Eq 11, finding $\langle\tau\rangle \approx 4.80 \pm 0.19$ min. This result is consistent with the transcriptional response of the target gene $cadBA$, which was previously probed by Northern blot analysis, finding that the cell-averaged $cadBA$ mRNA level starts to increase about 5 min after receptor activation [26].

In order to assess whether the position of the DNA-binding site along the chromosome affects the target search, we analyzed the behavior of strain T-P$_{cadBA}$, which has the CadC DNA binding site at the terminus instead of the native position, which is closer to the origin of replication. The corresponding response function shown in Fig 2C (orange stars) features a less pronounced initial increase than observed for N-P$_{cadBA}$, and the dashed orange line shows an adequate fit using the sequential model with fixed initial condition, Eq 9. The calculated mean first passage time of $\langle\tau\rangle \approx 4.20 \pm 0.15$ min is comparable to that for the wild-type strain. The slower initial increase is compensated by a faster increase at later times to yield a slightly smaller mean search time. To characterize the shape of the mean first passage time distributions, we calculated the variance σ^2, the skewness s and the kurtosis k of $p(\tau)$, see Table 1. For instance, the variance is smaller for strain T-P$_{cadBA}$ than for the wild-type.
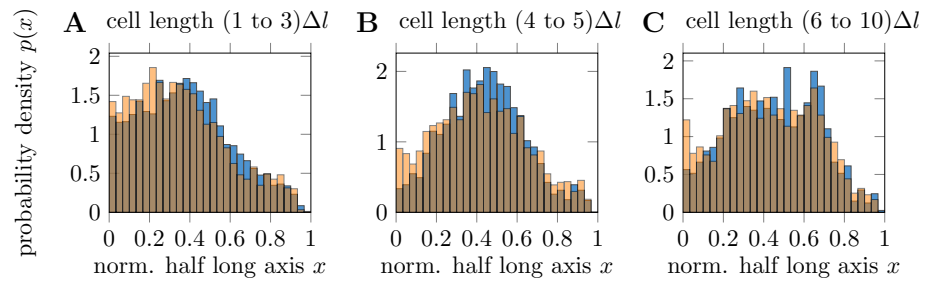
Fig 3. Localization of CadC. Localization probability of CadC foci of N-P$_{cadBA}$ (blue) and T-P$_{cadBA}$ (orange) along the half long axis of the cell. The half long axis is normalized such that mid-cell is at $x = 0$ and the poles are at $x = 1$. Overlaps of the two distributions are shown in darker orange. Cell age is represented by splitting all occurring cell lengths into ten equally spaced steps Δl and categorizing the cells according to their size. From the ten different age classes we observed similar localization probabilities for $l = (1 \text{ to } 3)\Delta l$ (**A**), $l = (4 \text{ to } 5)\Delta l$ (**B**) and $l = (6 \text{ to } 10)\Delta l$ (**C**), which are therefore grouped together in this plot.

Search time is decreased with two chromosomal CadC binding sites

After observing essentially the same mean search time for two very distant locations of CadC target sites on the chromosome, we wondered how a strain harboring both target sites would behave. We therefore repeated the measurements for *E. coli* strain N+T-P$_{cadBA}$, which has the native DNA-binding site and additionally the binding site at the terminus. As shown in Fig 2C (cyan triangles), the response function of this strain saturates much earlier than for the other two strains. Fitting the response data to the sequential model with fixed initial condition (Eq 9), we obtained a mean first passage time of $\langle \tau \rangle \approx 2.02 \pm 0.12\,\text{min}$, which is only around half of the time than for a single chromosomal binding site.

We also wondered whether the fluorescence spots indicating the position of stable CadC-DNA complexes in single cells would show a different spatial distribution between the two strains with a single chromosomal *cadBA* locus. We therefore analyzed the localization of spots along the long axis of the cell in the wild-type and T-P$_{cadBA}$ strains. As the position of chromosomal loci depends on the progression of the cell cycle [39], we grouped cells according to their length into three classes. Fig 3 shows the spatial distribution of the relative spot positions along the half long axis for these three length classes, comparing wild-type cells (blue) to T-P$_{cadBA}$ cells (orange). The data indicate a slightly higher probability to be at mid-cell when the DNA-binding site is at the terminus, but the difference compared to wild type is only minor. Overall, the CadC-DNA complexes are apparently broadly distributed along the cell axis, with some preference for the cell quarters and a profile that depends on the average cell length (and thus age) but only weakly on the position of the *cadBA* promoter on the chromosome. Given our fast growth conditions, we would have expected *cadBA* to be mostly located close to the cell quarters (wild-type) and close to mid-cell (T-P$_{cadBA}$), see S2 Fig). It is currently not clear why our data displays this trend only weakly.

Biophysical model of the target search process

To gain more insight into the dynamics of the target search, we turned to a coarse-grained biophysical model for the coupled dynamics of CadC and the DNA. We simulated the search of a CadC dimer for its target DNA binding site using a lattice

model and a kinetic Monte Carlo approach. As depicted in Fig 4A, the DNA is represented by a path on the 3D lattice (the 'cytoplasm') and the CadC dimer moves on the surface of the lattice (the 'membrane'). Starting from a random initial configuration, the CadC dimer diffuses in the membrane with a rate k_{2D}, binds non-specifically to DNA segments that are at the membrane and close to CadC with rate k_{on} and unbinds (k_{off}) or slides (k_{1D}) along DNA segments at the membrane when bound. The simulation continues until the CadC dimer reaches the specific DNA-binding site to end the search process. As demonstrated in Fig 4B the DNA dynamics was implemented by random displacements of single beads, at a basal rate k_{DNA}, with a kinetic Monte Carlo algorithm to accept moves according to a spring potential between neighboring beads, see 'Materials and methods'. Similarly, CadC dimers move by random displacements constrained either to the membrane or to the polymer. For each set of parameters we run $\geq 10^3$ simulations to compute mean values and search time distributions. 250–262

To compare our simulations to the experiments we defined a parameter set based on known experimental estimates for most parameters, as discussed in 'Materials and methods' and summarized in Table 2. Instead of simulating the full length of an *E. coli* chromosome, we made use of an observed relation between search time and polymer length to reduce computation time. As shown in Fig 4C, the simulated mean search time first increases with the number of DNA segments, but then becomes independent of polymer length and reaches a plateau. We found both the plateau value and the critical polymer length at which it is reached to depend on the size of the simulated cell. Polymer segments far apart along the polymer become uncorrelated in their motion when the polymer in between touches the boundary. Therefore the confinement leads polymer subchains to become independent when the DNA is long enough compared to the simulated cell to frequently encounter the boundary. Hence it is not necessary to simulate the full length of an *E. coli* chromosome, as long as the simulated polymer length lies within the plateau region. For the realistic parameter set used in Fig 4C, simulating a polymer length of $N_{DNA} = 100$ beads is sufficient, which is the minimum value we used in our simulations. To start with the simplest model, non-specific binding and sliding were not included in the realistic parameter set. 263–279

Mobility measurements of the DNA

To complete our parameter set, we estimated the diffusion constant D_0 of a DNA segment in our simulations by experimentally tracking a chromosomal locus. Towards this end we inserted a *parS* gene close to *ori* and let ParB-yGFP bind to it, making the *ori* region visible as a fluorescent spot in microscopy images. For N-P$_{cadBA}$-*parS_ori* and $\Delta cadC$-*parS_ori*, a wild type *E. coli* and an *E. coli* strain lacking *cadC* each containing *parS_ori*, fluorescence and phase contrast microscopy time lapse videos were taken of the same cells every 30 s for activating and inactivating conditions, respectively. 281–287

The mean-square displacement (MSD) was obtained by selecting the closest foci in subsequent image frames and calculating the ensemble-averaged MSD, the result of it being in good agreement with other tracking experiments of chromosomal loci in *E. coli* [41, 42] (see S3 Fig). Previous experiments showed that DNA diffusion in *E. coli* agrees well with the Rouse model [43], reporting diffusion exponents in the range 0.4–0.6 [41, 42, 44]. We therefore fitted the mean-square displacement to $MSD(\tau) = \Gamma \tau^{0.5}$, obtaining $\Gamma = 0.0111 \pm 0.0001 \, \mu m^2 \, s^{-0.5}$ for *E. coli* wild-type under activating conditions and $\Gamma = 0.0091 \pm 0.0001 \, \mu m^2 \, s^{-0.5}$ for $\Delta cadC$ under inactivating conditions. The DNA mobility seems to be independent of the probed conditions, since the two values do not differ significantly. Hence we used the average value $\langle \Gamma \rangle$ to compute the diffusion constant D_0 of a Kuhn segment according to the Rouse model. 288–298

The calculated diffusion constant of $D_0 \approx 0.006 \, \mu m^2 \, s^{-1}$ was used to scale the mean search time in Fig 4C. For the plateau value we obtained $\langle \tau_{plateau} \rangle \approx 3.66 \pm 0.04 \, min$, 299–300
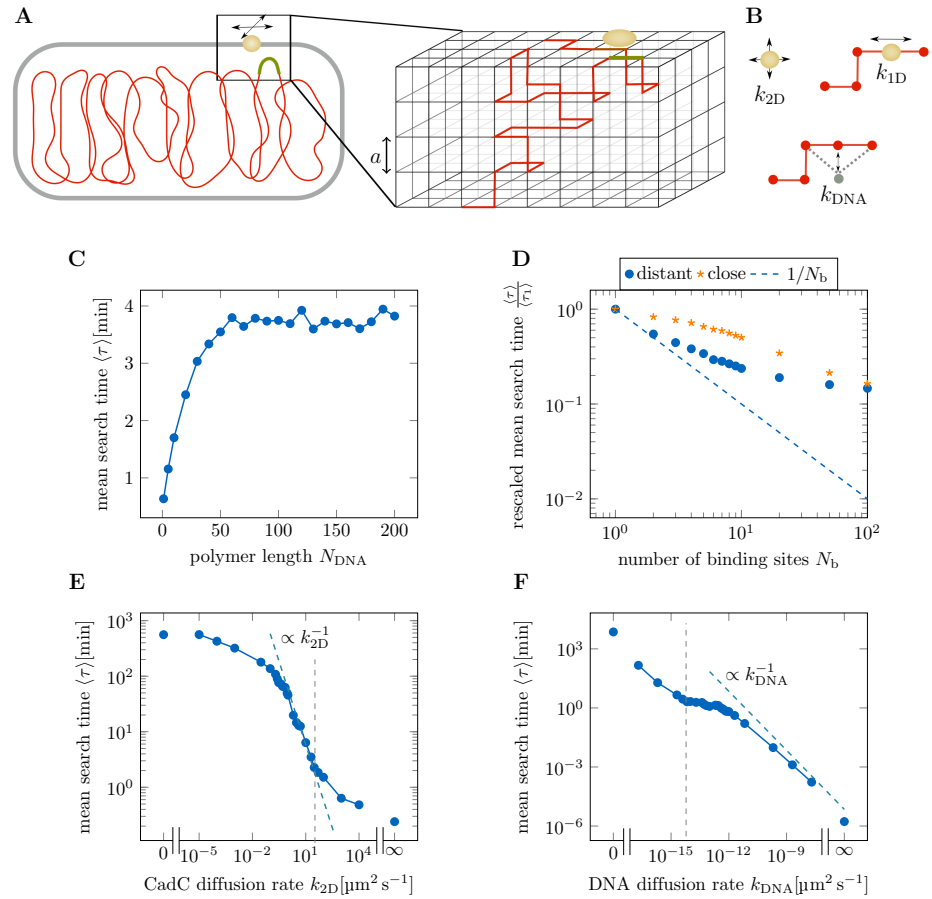
Fig 4. Characterization of the target search from computer simulations.
(**A**) The target search of membrane-integrated transcription factors is studied by kinetic Monte Carlo simulations of a lattice model, shown on the right. The cubic lattice represents a cell with the DNA (red) inside and a CadC dimer (yellow) in the membrane. (**B**) The DNA moves by random displacements of single beads with rate k_{DNA}. CadC dimers diffuse in the membrane with rate $k_{2\mathrm{D}}$, bind and unbind non-specifically to DNA segments at the membrane with rates k_{on}, k_{off} and slide along DNA segments with rate $k_{1\mathrm{D}}$ when bound. (**C**) Mean target search time $\langle\tau\rangle$ as a function of DNA length N_{DNA} using realistic parameters. (**D**) Dependence of the mean search time $\langle\tau\rangle$, normalized by the search time with a single binding site $\langle\tau_1\rangle$, on the number of DNA-binding sites N_{b}. Realistic parameters are used for binding sites placed either uniformly (blue dots) or next to each other (orange stars) along the circular chromosome. (**E**) Mean target search time for different CadC diffusion rates. The data point on the very left corresponds to $k_{2\mathrm{D}} = 0\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, the data point on the very right corresponds to $k_{2\mathrm{D}} \to \infty\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, simulated by making the whole cell surface the target. The dashed gray line marks $k_{2\mathrm{D}} = 0.20\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, used in the realistic parameter set. (**F**) Mean target search time for different DNA diffusion rates. The dashed gray line marks $k_{\mathrm{DNA}} = 0.0060\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, used in the realistic parameter set.

which is surprisingly close to the experimentally measured values, considering the simplicity of our simulations. Even without consideration of non-specific binding and sliding the value only deviates from the experimental measurement by a factor ~ 1.2.

Simulating multiple DNA-binding sites

Based on the experiments with N+T-P$_{cadBA}$ we performed simulations with varying number of DNA-binding sites. In Fig 4D the mean search time $\langle \tau \rangle$ normalized by the mean search time with a single DNA-binding site $\langle \tau_1 \rangle$ is shown as a function of the number of binding sites N_{b}. Placing the binding sites as far apart as possible on the circular chromosome, simulations with realistic parameters show a halving in search time when increasing the number of binding sites from one to two. This is the expected result for two binding sites searching independently of each other due to the decorrelation of polymer subchains in spatial confinement. How far two binding sites have to be placed to behave as independent searchers therefore depends on the size of the simulated cell. The initially linear regime for small N_{b} flattens more and more as the binding sites become closer to one another and more correlated in their movement. Placing the binding sites next to each other (cyan triangles) has an expectedly small effect for small N_{b}, since it only increases the size of the binding site. The curve becomes steeper as the binding site occupies a larger fraction of the polymer.

The two DNA-binding sites in N+T-P$_{cadBA}$ being on opposite sides of the chromosome, we expect them to behave as two independent binding sites. The experimentally measured reduction of the search time by roughly a factor of two is therefore in good agreement with our simulations, predicting a linear dependency between the response time and the number of DNA-binding sites. Also, by construction, the position of the binding site along the chromosome has no effect on the simulated search times.

CadC diffusion strongly affects the target search time

To address the question whether DNA mobility or CadC diffusion is more crucial for the search process, numerical simulations with varying diffusion rates were run. Plotting the mean search time as a function of the CadC diffusion rate k_{2D} in Fig 4E we observe three different regimes. For slow protein diffusion the search time is almost independent of k_{2D}, it follows a range where the search process is entirely dominated by CadC diffusion ($\tau \propto k_{\mathrm{2D}}^{-1}$) and for very fast CadC diffusion it becomes less critical again. The data point for infinitely fast CadC diffusion was obtained by making CadC cover the whole cell surface. The realistic value for the diffusion constant of CadC ($k_{\mathrm{2D}} = 0.20\,\mathrm{\mu m^2\,s^{-1}}$) is marked by a gray dashed line and lies in the regime where the search time strongly depends on k_{2D}. Fig 4F shows the counterpart of this plot for DNA diffusion, marking $k_{\mathrm{DNA}} = 0.0060\,\mathrm{\mu m^2\,s^{-1}}$ with a gray dashed line. While the mean search time is strongly dependent on DNA diffusion for small and large k_{DNA}, it is almost constant in the experimentally relevant intermediate regime.

To further validate this observation, we used our numerical simulations to approximate a target search exclusively due to CadC diffusion in the membrane. We used the realistic parameters but placed the DNA-binding site at the membrane and set the DNA diffusion rate to zero, such that only CadC was moving, yielding a mean search time of $\langle \tau \rangle \approx 18\,\mathrm{s}$. The previously reported [45] approximate formula $\tau \approx \frac{R_{\mathrm{s}}^2}{D}(-2\ln[\frac{r_{\mathrm{t}}}{R_{\mathrm{s}}}] + 2\ln[2] - 1)$ for the mean first encounter time of a particle moving with diffusion constant D on the surface of a sphere with radius R_{s} with a trap of radius r_{t} leads, approximating the cubic cell by a sphere of radius $R_{\mathrm{s}} = 0.69\,\mathrm{\mu m}$ and CadC with $r_{\mathrm{t}} = 0.050\,\mathrm{\mu m}$ to $\tau \approx 16\,\mathrm{s}$. Using the approximation does not only validate

our simulations, but shows that the dependence of the search time on the size of CadC is weak, rendering a correction for overestimating the size of CadC dimers in the simulations unnecessary.

While the DNA is likely to be the less mobile part in the target search, it has to move at least close to the membrane to enable binding to CadC. The data point on the very right of Fig 4E corresponds to the time it takes the DNA-binding site to bind anywhere to the membrane. It corresponds to $\tau \approx 14\,\text{s}$, similar to the time it takes CadC to locate the binding site at the membrane. Therefore, a scenario where the DNA-binding site randomly reaches the membrane and CadC searches the membrane to bind to it seems to lead to reasonable search times according to our simulations.

For comparison we also simulated a target search process where CadC is immobile and DNA diffusion has to account for the whole search. When only the DNA is moving and all other rates are set to zero, with realistic values for DNA length and cell size a mean search time of $\langle \tau \rangle \approx 560\,\text{min}$ was calculated from the simulations, a response time that would not allow *E. coli* cells to survive the transition to acidic environments. Using the approximate formula $\langle \tau \rangle \approx \frac{\Omega}{4 r_t D}$ [46] for the encounter time of a particle with diffusion constant D to locate a target of radius r_t on the surface of a confining spherical domain with volume V we obtained $\tau \approx 14\,\text{min}$ for a monomer with diffusion constant D_0 and $\tau \approx 1400\,\text{min}$ for a monomer with diffusion constant $\frac{D_0}{N_\text{DNA}}$. Finding a value between these two approximations is what we expected, as they do not account for the polymer dynamics.

Quantitative comparison of simulated and experimental search time distributions

As shown above, the numerical simulations of the target search using our experimentally estimated parameter set yield mean search times that are fairly close to what we measured experimentally. To test whether the biophysical model can also quantitatively capture the experimental behavior, we attempted to find simulations with time distributions closely matching the experimentally computed CDFs. In Fig 5A the experimental CDF of N-P$_{cadBA}$ together with the best fit to the sequential model with mixed initial condition is shown in orange. In order to find a simulation with agreeing time distributions, we simulated the target search process using the estimated parameter set but improved the agreement by increasing the simulated cell volume to $V_\text{cell} = 1.3\,\mu\text{m}^3$. As shown by the blue dots in Fig 5A this yields a very good accordance between experimental data and simulations. In Fig 5B the experimental CDF of N+T-P$_{cadBA}$ and the corresponding fit to the sequential model with fixed initial condition is shown in orange. Our attempt to find a matching simulation by using the same parameters as for the simulations in panel A but increasing the DNA-binding sites to $N_\text{b} = 2$ leads to quite a good agreement with the experiment, as shown by the blue dots. Even though we implemented non-specific binding of CadC and sliding along the DNA, this was not necessary to obtain simulations that reproduce the experimental time distributions. Despite the simplicity of the simulations, neglecting in particular constraints of DNA movement due to its specific organization in the cell, the results agree surprisingly well with the experimental findings when using experimentally realistic parameters.

To further investigate the simulated first passage time distributions, we computed them for different parameter settings and fitted to the sequential model with fixed initial condition (Eq 3) and with mixed initial condition (Eq 10) respectively. While the shape of the first passage time distribution is independent of the system size parameters, the dynamic parameters have a big effect. For most parameter settings, including the realistic parmeters, we found the best agreement of simulated CDFs with the sequential
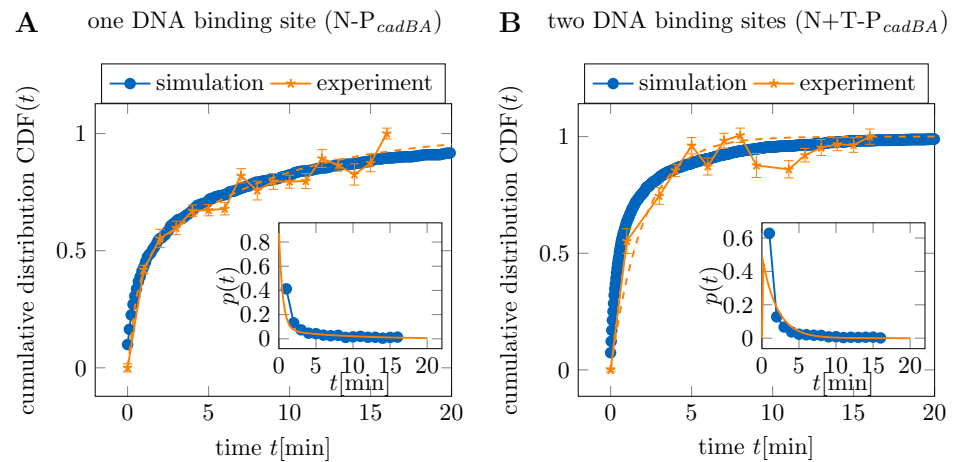
A   one DNA binding site (N-P$_{cadBA}$)   **B**   two DNA binding sites (N+T-P$_{cadBA}$)

Fig 5. Matching simulations to the experimental results. Numerical simulations are directly compared to the experimental distributions. Experimental data are shown in orange stars, the orange dashed line corresponds to the best fit. Blue dots correspond to the simulations that agree most with the experimental data. Insets depict the corresponding first passage time distributions. (**A**) The experimental CDF of N-P$_{cadBA}$ is shown together with the best fit and simulations using the realistic parameter set, summarized in Table 2. To yield better agreement with the experimental CDF the cell volume was increased to $V_{cell} = 1.3\,\mu m^3$. (**B**) The experimental CDF of N+T-P$_{cadBA}$ is shown together with the best fit and simulations using the parameter set matching the experimental values, with cell volume $V_{cell} = 1.3\,\mu m^3$ and two binding sites on the polymer. All simulations were started from a random initial configuration.

model with mixed initial condition. However, for simulations with an immobile polymer or slowly diffusing CadC, the search time distribution agrees rather with the sequential model with fixed initial condition. Since both in the experimental and simulated CDFs the delay β is very small compared to α, a fit to the fixed initial model gives virtually the same result as a single exponential CDF. When either the DNA or CadC accounts for most of the search there is only one limiting rate in the process, which can therefore be described by a single exponential distribution.

Discussion

We combinined fluorescence microscopy experiments, quantitative analysis, and kinetic Monte Carlo simulations to characterize the target search kinetics of membrane-integrated transcription factors for a specific binding site on the chromosomal DNA. We were able to measure the time between the environmental stimulus and stable DNA binding of a membrane-integrated one-component receptor, using the pH stress-sensing receptor CadC in *E. coli* as a model system. The measured mean search time of on average 4.5 min for a single DNA-binding site is consistent with the timescale of the transcriptional response [26]. Surprisingly, the search time is not much slower than the search time of the cytosolic *lac* repressor for its operator, which takes around one minute at a similarly low protein level [12].

As the position of the DNA-binding site along the chromosome has no influence on the mean search time, the process seems to be quite dynamic. Given that the chromosome is highly organized within the cell, we believe this effect to arise mostly due to the mobility of CadC. Computing the localization of CadC foci surprisingly did

not yield a strong dependence on the position of the DNA-binding site, the broadness of the distribution however gives the impression of the localization being not very exact, roughly agreeing with the localization of the average DNA mass [39]. While we do not have new evidence for or against proteolytic processing, this leads us to favor a diffusion and capture mechanism over transertion. As the search time is independent of the position of the DNA-binding site, transertion of CadC is at least not a requirement for fast response, in good agreement with a previous evaluation of the three models [28]. The experiments on mutants with two DNA-binding sites leading to a decrease of the mean search time by a factor two support the idea of a dynamic search. Confirmed by the simulations, this is the expected result for two independent and equally accessible binding sites, where distant parts of the polymer become uncorrelated in their motion due to the confinement in the cell.

Despite the simplicity of our biophysical model of the target search, we obtained search times that match the experimental measurements surprisingly well, even without consideration of non-specific binding and sliding. It therefore does not seem to require a specific strategy to make the target search work. Although the simulations are clearly oversimplifying the biological situation by neglecting chromosomal loci being constrained to move within macrodomains, the agreement between simulations and experimental evidence is striking.

While it is difficult to experimentally answer the question whether CadC or the DNA contribute more to the search process, the consistency between our simulations and experimental findings allows to draw some conclusions from an analysis of the simulations. Evaluating the functional dependence of the mean search time on the diffusion rate of the CadC dimer, the search time was found to depend linearly on the diffusion rate of CadC, but is almost independent of the DNA diffusion rate in the experimentally relevant regime. The search process therefore seems to be greatly limited by the mobility of the transcription factors. The limiting cases we computed for the process also approve that a mechanism, where the binding site only reaches the membrane and CadC searches for it, could be possible.

Beyond mean values the experimental data allowed the computation of the cumulative distribution functions of the target search time. Fitting them to the distribution of a reversible sequential two-step process resulted in good agreement of the wild type data with the model with mixed initial condition, the same being observed for the numerical simulations with most parameter settings. Using a parameter set closely related to the experimental conditions led to a distribution that greatly matches the experimental CDF of *E. coli* wild type. Attempting to find a matching simulation for the mutant with two DNA-binding sites by using the same parameters with two DNA-binding sites also yields a good result. In contrast to the simulations however the experimental CDF of the mutants with two binding sites and the mutants with the binding at *ter* show a better agreement with the sequential model with fixed initial condition, as well as a single exponential CDF. How exactly the position of the DNA-binding site affects the initial condition of the search process can only be speculated. It might be that the native DNA-binding site is closer to the membrane or CadC is naturally located close to *ori*, although both CadC monomers and dimers have been reported to be homogeneously distributed within the membrane [28]. In the simulations, distributions agreeing with the sequential model with fixed initial conditions or a single exponential CDF are obtained when either only CadC is moving, or it is moving very slowly.

Taken together, our experiments and simulations indicate that CadC is highly mobile in the membrane, while the *cadBA* promoter on the *E. coli* chromosome is mobile enough to randomly reach the membrane, enabling CadC to locate the DNA-binding site within about five minutes, independent of its position along the chromosome.

Materials and methods

Model and simulations

Since the target search of CadC for its DNA-binding site(s) has been experimentally shown to succeed within a few minutes, we were aiming at finding the simplest model that can reproduce this fast response. Hence we constrained our simulations to the main components of the target search: the DNA with one or more specific binding sites moving inside an *E. coli* cell and one or more CadC beads diffusing in the cell membrane. CadC dimerization was not simulated explicitly, instead the CadC beads on the cell surface correspond to already formed dimers. Besides diffusion of the DNA and CadC, non-specific binding of CadC to any of the DNA sites, as well as one-dimensional sliding along DNA segments that are aligned along the membrane were included.

To simulate a system as large as the whole genome in an *E. coli* cell for a time span long enough to capture the target search of CadC for its DNA-binding site is computationally challenging. We therefore chose a coarse-grained cubic lattice model in order to reduce the simulation time. In the simple cubic lattice model, the cell is composed of N_{cell}^3 lattice points forming a cuboid. The lattice constant was chosen equal to the Kuhn length $l_{\text{k}} = 2l_{\text{p}}$ of double stranded DNA (persistence length $l_{\text{p}} \approx 50\,\text{nm}$ [47]), such that the polymer could be represented as a freely jointed chain (FJC) on the lattice. It is composed of N_{DNA} beads with coordinates $\boldsymbol{r} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N_{\text{DNA}}})$ placed on the grid points connected by $N_{\text{DNA}} - 1$ bonds, one or more of the beads being specified as the specific binding sites. Each CadC dimer occupies a single lattice point on the cell surface. In analogy to the single-site Bond-Fluctuation model [48] the polymer beads are connected by bonds of variable length, that are free to occupy also diagonal configurations. As we simulate a phantom chain, we replace the upper and lower bound of the bond length intended to ensure excluded volume and prevent chain crossings by a spring potential

$$U\{\boldsymbol{r}_n(t)\} = \sum_{n=1}^{N_{\text{DNA}}-1} k_{\text{s}}(|\boldsymbol{r}_{n+1}(t) - \boldsymbol{r}_n(t)| - b)^2\,, \tag{12}$$

with spring constant k_{s} and an equilibrium bond length $b = l_{\text{k}}$. Following the Bond-Fluctuation model, the polymer is moved by attempting the displacement of a single bead to one of the nearest lattice points, the spring potential being enforced by a Metropolis algorithm [40]. CadC beads are also moved by random displacements but constrained to the cell surface or to the DNA when bound. The kinetic Monte Carlo algorithm simulates the master equation

$$\frac{\mathrm{d}p(\boldsymbol{x}, t)}{\mathrm{d}t} = \sum_{\boldsymbol{x}'}[w_{\boldsymbol{x}'\boldsymbol{x}}p(\boldsymbol{x}', t) - w_{\boldsymbol{x}\boldsymbol{x}'}p(\boldsymbol{x}, t)]\,, \tag{13}$$

with $p(\boldsymbol{x}, t)$ the probability that the system is in state $\boldsymbol{x}$ at time t and transition rates $w_{\boldsymbol{x}'\boldsymbol{x}} = w(\boldsymbol{x}' \to \boldsymbol{x})$ between the states. The rates governing the dynamics are the polymer diffusion rate k_{DNA}, which is set to one unless otherwise stated, the rate for CadC diffusion on the cell surface k_{2D}, the non-specific binding rate k_{on}, the unbinding rate k_{off} and the one-dimensional sliding rate of CadC along the polymer k_{1D}. Unless otherwise stated the simulations were initialized with a random walk configuration of the polymer inside the cell and a randomly placed CadC bead in the membrane. To compute mean values and search time distributions $\geq 10^3$ simulations were run for every choice of parameters, each starting from a different initial configuration.

For quantitative comparison with the experiments, we defined a parameter set that matches the experimental values, as summarized in Table 2. The *E. coli* cell volume of

$\propto 1\,\mu m^3$ is approximated by a simulation box of volume N_{cell}^3 with $N_{\mathrm{cell}} = 10a$ and $a = 100\,\mathrm{nm}$. The *E. coli* chromosome with $4\,639\,221$ bp, which measures $1.58\,\mathrm{mm}$ corresponds to a length of $15773a$. Since our simulations have shown that the target search dynamics are independent of polymer length once it has reached a cell size dependent critical length (see Fig 4C), we save computation time by choosing the polymer length well above this threshold at $100a$. For the CadC dimer diffusion a value of $D_{\mathrm{cadC}} \approx 0.20\,\mu m^2\,s^{-1}$ agrees with both previous measurements of a membrane-integrated protein with two transmembrane domains in *E. coli* [49] and single-molecule measurements of CadC [28]. From our *ori* tracking experiments we calculated the diffusion constant of a DNA bead in our simulation to be $D_0 \approx 0.0060\,\mu m^2\,s^{-1}$, which is the value we chose for the realistic parameter set. To convert the dimensionless search times τ' from the simulations into seconds we computed $\tau' \frac{a^2}{D_0} \approx 1.66\tau'$.

Table 2. Simulation parameters matching the experiments.

Parameter	Symbol	Value
lattice constant	a	$100\,\mathrm{nm}$
cell volume	V_{cell}	$1\,\mu m^3$
DNA length	N_{DNA}	$10\,\mu m$
number of CadC dimers	N_{p}	1
number of binding sites	N_{b}	1
CadC diffusion	$k_{2\mathrm{D}}$	$0.20\,\mu m^2\,s^{-1}$
DNA diffusion	k_{DNA}	$0.0060\,\mu m^2\,s^{-1}$

Parameter set for the kinetic Monte Carlo simulations that were chosen to match the experiments.

Experiments

Construction of strains and plasmids

Molecular methods were carried out according to standard protocols or according to the manufacturer's instructions. Kits for the isolation of plasmids and the purification of PCR products were purchased from Süd-Laborbedarf (SLG; Gauting, Germany). Enzymes were purchased from New England BioLabs (Frankfurt, Germany). Bacterial strains and plasmids used in this study are summarized in table 1.

E. coli strains were cultivated in LB medium ($10\,\mathrm{g\,l^{-1}}$ NaCl, $10\,\mathrm{g\,l^{-1}}$ tryptone, $5\,\mathrm{g\,l^{-1}}$ yeast extract) or in Kim Epstein (KE) medium [50] adjusted to pH 5.8 or pH 7.6, using the corresponding phosphate-buffer. *E. coli* strains were always incubated aerobically in a rotary shaker at $37\,^\circ\mathrm{C}$. KE medium was always supplemented with $0.20\,\%(\mathrm{w/v})$ glucose. Generally, lysine was added to a final concentration of $10\,\mathrm{mmol}$ unless otherwise stated. If necessary, media were supplemented with $100\,\mu g\,ml^{-1}$ ampicillin or $50\,\mu g\,ml^{-1}$ kanamycin sulfate. To allow the growth of the conjugation strain *E. coli* WM3064, we added meso-diamino-pimelic acid (DAP) to a final concentration of $200\,\mu mol$.

In order to gain strain *E. coli* MG1655-*parS_ori*, the *parS* site of *Yersinia pestis* was inserted at the origin of replication (*ori*) at $84.3'$ in *E. coli* MG1655. Briefly, the *parS* region was inserted between *pstS* and *glmS*. Therefore, DNA fragments comprising $650\,\mathrm{bp}$ of *pstS* and *glmS* and the *parS* region were amplified by PCR using MG1655 genomic DNA as template and the plasmid pFH3228, respectively. After purification, these fragments were assembled via Gibson assembly [51] into EcoRV-digested pNPTS138-R6KT plasmid, resulting in the pNTPS138-R6KT-*parS_ori* plasmid. The

resulting plasmid was introduced into *E. coli* MG1655 by conjugative mating using *E. coli* WM3064 as a donor on LB medium containing DAP. Single-crossover integration mutants were selected on LB plates containing kanamycin but lacking DAP. Single colonies were then streaked out on LB plates containing $10\,\%$(wt/vol) sucrose but no NaCl to select for plasmid excision. Kanamycin-sensitive colonies were then checked for targeted insertion by colony PCR and sequencing of the respective PCR fragment. In order to gain strain *E. coli* MG1655-$\Delta cadC$ *parS_ori*, the *parS* site of *Y. pestis* was inserted at the origin of replication (*ori*) at 84.3′ in *E. coli* MG1655-$\Delta cadC$ as described above.

In order to gain strain *E. coli* MG1655_P$_{cadBA}$_terminus, the *cadBA* promoter region was inserted at the terminus (33.7′) in *E. coli* MG1655. Construction of this strain was achieved via double homologous recombination using the pNTPS138-R6KT-P$_{cadBA}$_terminus plasmid [28] as described above. Correct colonies were then checked for targeted insertion by colony PCR and sequencing of the respective PCR fragment.

Details of the strains and plasmids used in this study are summarized in Table 3.

Table 3. Strains and plasmids used in this study.

Strains	Relevant genotype or description	Reference
E. coli DH5αλpir	*endA1 hsdR17 glnV44 (= supE44) thi-1 recA1 gyrA96 relA1 ψ80’lac*Δ *(lacZ)*M15 Δ(*(lacZYA-argF)*U169 *zdg*-232::Tn10 *uidA::pir*⁺	[52]
E. coli WM3064	*thrB*1004 *pro thi rpsL hsdS lacZ*ΔM15 RP4-1360 Δ(*araBAD*)567 Δ*dapA*1341::[*erm pir*]	[53]
E. coli MG1655 (N-P$_{cadBA}$)	K-12 F⁻ λ⁻ *ilvG*⁻ *rfb*-50 *rph*-1	[54]
E. coli MG1655_P$_{cadBA}$_terminus (N+T-P$_{cadBA}$)	Additional *cadBA* promoter region at the terminus (33.7′) in MG1655	This work
E. coli MG1655ΔP$_{cadBA}$-P$_{cadBA}$_terminus (T-P$_{cadBA}$)	Clean deletion of *cadBA* promoter region in MG1655 with relocated *cadBA* promoter region at the terminus (33.7′)	[28]
E. coli MG1655 *parS_ori*	Insertion of the *Yersinia pestis* pMT1*parS* site at the *ori* (84.3′) in MG1655	This work
E. coli MG1655Δ*cadC*	Deletion of *cadC* gene in MG1655, Km^R	[55]
E. coli MG1655Δ*cadC parS_ori*	Insertion of the *Yersinia pestis* pMT1*parS* site at the *ori* (84.3′) in MG1655Δ*cadC*, Km^R	This work
Plasmids	**Relevant genotype or description**	**Reference**
pET-*mCherry-cadC*	N-terminal fusion of *cadC* with *mCherry*, connected with a 22 amino acid long linker containing a 10His tag in pET16b, Amp^R	[28]
pNTPS138-R6KT-P$_{cadBA}$_terminus	pNPTS-138-R6KT-derived suicide plasmid for insertion of *cadBA* promoter region at terminus in *E. coli* MG1655, Km^R	[28]
pNTPS138-R6KT-*parS_ori*	pNPTS-138-R6KT-derived suicide plasmid for insertion of the *Yersinia pestis* pMT1*parS* site at the *ori* in *E. coli* MG1655, Km^R	This work
pFHC2973	N-terminal fusion of *parB* with *ygfp*, Amp^R	[56]
pFH3228	Plasmid carrying the pMT1-*parS* of *Yersinia pestis*, Amp^R	[56]

In vivo fluorescence microscopy

To analyze search response of mCherry-CadC to its binding site(s), overnight cultures of *E. coli* MG1655 (one CadC binding site close to *ori*, N-P$_{cadBA}$), *E. coli* MG1655ΔP$_{cadBA}$-P$_{cadBA}$-terminus (one CadC binding site at terminus, T-P$_{cadBA}$) and *E. coli* MG1655_P$_{cadBA}$_terminus (two CadC binding sites, N+T-P$_{cadBA}$), each carrying pET-*mCherry-cadC*, were prepared in KE medium pH 7.6 and aerobically cultivated at 37 °C. The overnight cultures were used to inoculate day cultures (OD$_{600}$ of 0.1) in fresh medium at pH 7.6. At an OD$_{600}$ of 0.5, cells were gently centrifuged and resuspended, thereby exposing them to low pH (KE medium pH 5.8 + lysine). Then the cultures were aerobically cultivated at 37 °C and every 1 min after the shift to low pH, 2 µl of the culture was spotted on 1 %(w/v) agarose pads (prepared with KE medium pH 5.8 + lysine), placed onto microscope slides and covered with a coverslip. Subsequently, images were taken on a Leica DMi8 inverted microscope equipped with a Leica DFC365 FX camera (Wetzlar, Germany). An excitation wavelength of 546 nm and a 605 nm emission filter with a 75 nm bandwidth was used for mCherry fluorescence with an exposure of 500 ms, gain 5, and 100 % intensity. Before shifting the cells to low pH, 2 µl of the cultures in KE medium pH 7.6 were spotted on 1 %(w/v) agarose pads (prepared with KE medium pH 7.6) and imaged as a control.

To analyze the spatiotemporal localization of a chromosomal locus, the *parS* site was inserted close to the *ori*. The localization of the *parS* site was visualized via the binding of ParB-yGFP [19]. *E. coli* MG1655 *parS_ori* cells carrying plasmid pFH3228 were cultivated in KE medium pH 7.6 as described above. At an OD$_{600}$ of 0.5, 2 µl of the culture were shifted on 1 %(w/v) agarose pads (prepared with KE medium pH 7.6 or pH 5.8 + lysine) and placed onto microscope slides and covered with a coverslip. Subsequently, every 30 s time lapse images of the same cells were taken on a Leica DMi8 inverted microscope equipped with a Leica DFC365 FX camera (Wetzlar, Germany) of the same positions. An excitation wavelength of 485 nm and a 510 nm emission filter with a 75 nm bandwidth were used for ParB-yGFP fluorescence with an exposure of 350 ms, gain 3, and 100 % intensity.

Microscopy image analysis

To analyze the fluorescence microscopy images for CadC or ParB foci within the cells, we used Oufti [57], an open-source software designed for the analysis of microscopy data for cell segmentation of the phase contrast microscopy images. The resulting cell outlines were used in a custom-written software implemented in Matlab and available on request to detect fluorescent foci. Briefly, a graphical user interface (GUI) was implemented that allows testing the parameters in a test mode before running the actual detection. In detection mode a function *SpotDetection.m* is called, that iterates through all frames and all cells. For each cell, from pixels in the fluorescence microscopy images the intensity of which is above a threshold defined by the parameters and dependent on the mean and variance of the fluorescence signal within the cell the connected components are computed. The components are checked for minimum and maximum size and minimum distance to other foci before being added to the list of foci. For further computations, information on all cells and foci were saved for all frames corresponding to a certain time after receptor activation.

Interpretation of CadC foci data

In a custom-written Matlab script the results from the image analysis were used to compute the fraction of cells with foci $\nu(t)$ as a function of time t after receptor activation, which upon normalization corresponds to the CDF of the search time distribution. We fit the data to the CDF of a theoretical model using the *curve_fit*

function of the scipy module in Python, choosing a trust region reflective algorithm, which is an evolution of the Levenberg-Marquardt method that can handle bounds. This algorithm was developed to solve nonlinear least squares problems and combines the gradient descent method and the Gauss-Newton method. It minimizes the sum of the weighted squares of the errors between the measured data y_i and the curve-fit function $\hat{y}(t; \boldsymbol{p})$

$$\chi^2(\boldsymbol{p}) = \sum_{i=1}^{m} \left(\frac{y(t_i) - \hat{y}(t_i; \boldsymbol{p})}{\sigma_{y_i}} \right)^2 ,$$

with an independent variable t and a vector of n parameters $\boldsymbol{p}$ and a set of m data points (t_i, y_i). σ_{y_i} is the measurement error for measurement $y(t_i)$ and $W_{ij} = \frac{1}{\sigma_{y_i}^2}$ is the weighting matrix. [58]

Calculation of the ParB diffusion constant

The results from the image analysis were used to compute trajectories of ParB foci in a custom-written Matlab script by selecting the closest foci in subsequent image frames.

From the trajectories of ParB foci the ensemble-averaged mean-square displacement (MSD) was computed as a function of time lag τ: $\text{MSD}(\tau) = \langle (\vec{r}(t) - \vec{r}(t + \tau))^2 \rangle$, where the mean was taken over different foci and $\tau = n30\,\text{s}$ with $n \in \{1, \dots, N\}$ and the number of time steps N. The Rouse model predicts the MSD in 2D [59]:

$$\begin{aligned}
\text{MSD}(\tau) &= \sqrt{\frac{16 l_{\text{K}}^2 k_{\text{B}} T}{3\pi\gamma}} \tau^{0.5} \\
&= \sqrt{\frac{16 l_{\text{K}}^2}{3\pi}} D_0^{0.5} \tau^{0.5} ,
\end{aligned} \tag{14}$$

with Boltzmann's constant k_{B}, absolute temperature T, Kuhn length l_{K}, friction constant γ and bead diffusion constant D_0, where we have used $D_0 = \frac{k_{\text{B}} T}{\gamma}$ [60]. We fitted the experimentally determined MSD to $\text{MSD}(t) = \Gamma t^{0.5}$ and determined the diffusion constant from $D_0 = \Gamma^2 \frac{3\pi}{16 l_{\text{K}}^2}$.

Supporting information

S1 Fig.   Dynamics of the target search by CadC. Fluorescent microscopy images were taken every minute after receptor activation and analyzed for CadC foci for all three *E. coli* strains. The plot shows the fraction of cells with foci $\nu(t)$ as a function of time t after the medium shift to low pH and lysine

S1 Table.   Fit results. Results from fitting the experimentally computed CDF to the sequential reversible model with mixed initial condition (N-P$_{cadBA}$) and fixed initial condition (T-P$_{cadBA}$ and N+T-P$_{cadBA}$). The fit parameters α, β and c were used to compute the mean first passage time, the variance, the skewness and the kurtosis with uncertainties obtained from error propagation using the full covariance matrix.

S2 Fig.   Localization of chromosomal loci. The localization of *ori* and terminus is shown during the cell cycle of lowly and fast growing *E. coli*.

S3 Fig.   Mean-square displacement of ParB foci. The MSD of ParB foci was calculated by selecting the closest foci in subsequent image frames and calculating the ensemble-averaged mean-square displacement as a function of time lag τ. The dashed lines show the fit to $\Gamma\tau^\alpha$. For each time lag the mean was taken over 234 to 936 values.

Acknowledgments

References

1. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the natural environment to infectious diseases. Nature reviews microbiology. 2004;2(2):95.

2. Ulrich LE, Koonin EV, Zhulin IB. One-component systems dominate signal transduction in prokaryotes. Trends in microbiology. 2005;13(2):52–56.

3. Stock AM, Robinson VL, Goudreau PN. Two-component signal transduction. Annual review of biochemistry. 2000;69(1):183–215.

4. Riggs AD, Bourgeois S, Cohn M. The lac represser-operator interaction: III. Kinetic studies. Journal of molecular biology. 1970;53(3):401–417.

5. Adam G, Delbrück M. Reduction of dimensionality in biological diffusion processes. Structural chemistry and molecular biology. 1968;198:198–215.

6. Richter PH, Eigen M. Diffusion controlled reaction rates in spheroidal geometry: application to repressor-operator association and membrane bound enzymes. Biophysical chemistry. 1974;2(3):255–263.

7. Berg OG, Winter RB, Von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry. 1981;20(24):6929–6948.

8. Slutsky M, Mirny LA. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. Biophysical Journal. 2004;87(6):4021–4035.

9. Hu T, Grosberg AY, Shklovskii B. How proteins search for their specific sites on DNA: the role of DNA conformation. Biophysical journal. 2006;90(8):2731–2744.

10. Bonnet I, Biebricher A, Porte PL, Loverdo C, Bénichou O, Voituriez R, et al. Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. Nucleic acids research. 2008;36(12):4118–4127.

11. Loverdo C, Benichou O, Voituriez R, Biebricher A, Bonnet I, Desbiolles P. Quantifying hopping and jumping in facilitated diffusion of DNA-binding proteins. Physical review letters. 2009;102(18):188101.

12. Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. Science. 2007;316(5828):1191–1194.

13. Hammar P, Leroy P, Mahmutovic A, Marklund EG, Berg OG, Elf J. The lac repressor displays facilitated diffusion in living cells. Science. 2012;336(6088):1595–1598.

14. van den Broek B, Lomholt MA, Kalisch SM, Metzler R, Wuite GJ. How DNA coiling enhances target localization by proteins. Proceedings of the National Academy of Sciences. 2008;105(41):15738–15742.

15. Schötz T, Neher RA, Gerland U. Target search on a dynamic DNA molecule. Physical Review E. 2011;84(5):051911.

16. Li GW, Berg OG, Elf J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. Nature Physics. 2009;5(4):294.

17. Liu L, Cherstvy AG, Metzler R. Facilitated diffusion of transcription factor proteins with anomalous bulk diffusion. The Journal of Physical Chemistry B. 2017;121(6):1284–1289.

18. Miller VL, Taylor RK, Mekalanos JJ. Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. cell. 1987;48(2):271–279.

19. Shaner NC, Lambert GG, Chammas A, Ni Y, Cranfill PJ, Baird MA, et al. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nature methods. 2013;10(5):407.

20. Yang Y, Isberg RR. Transcriptional regulation of the Yersinia pseudotuberculosis pH 6 antigen adhesin by two envelope-associated components. Molecular microbiology. 1997;24(3):499–510.

21. Buchner S, Schlundt A, Lassak J, Sattler M, Jung K. Structural and functional analysis of the signal-transducing linker in the pH-responsive one-component system CadC of Escherichia coli. Journal of molecular biology. 2015;427(15):2548–2561.

22. Woldringh CL. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. Molecular microbiology. 2002;45(1):17–29.

23. Roggiani M, Goulian M. Chromosome-membrane interactions in bacteria. Annual review of genetics. 2015;49:115–129.

24. Rodriguez MAP, Guo X. Biomacromolecular localization in bacterial cells by the diffusion and capture mechanism. Annals of Microbiology. 2013;63(3):825–832.

25. Jung K, Fabiani F, Hoyer E, Lassak J. Bacterial transmembrane signalling systems and their engineering for biosensing. Open biology. 2018;8(4):180023.

26. Fritz G, Koller C, Burdack K, Tetsch L, Haneburger I, Jung K, et al. Induction kinetics of a conditional pH stress response system in Escherichia coli. Journal of molecular biology. 2009;393(2):272–286.

27. Haneburger I, Fritz G, Jurkschat N, Tetsch L, Eichinger A, Skerra A, et al. Deactivation of the E. coli pH stress sensor CadC by cadaverine. Journal of molecular biology. 2012;424(1-2):15–27.

28. Brameyer S, Rösch TC, El Andari J, Hoyer E, Schwarz J, Graumann PL, et al. DNA-binding directs the localization of a membrane-integrated receptor of the ToxR family. Communications biology. 2019;2(1):4.

29. Görke B, Reinhardt J, Rak B. Activity of Lac repressor anchored to the Escherichia coli inner membrane. Nucleic acids research. 2005;33(8):2504–2511.

30. Redner S. A guide to first-passage processes. Cambridge University Press; 2001.

31. Schlundt A, Buchner S, Janowski R, Heydenreich T, Heermann R, Lassak J, et al. Structure-function analysis of the DNA-binding domain of a transmembrane transcriptional activator. Sci Rep. 2017;7(1):1051.

32. Haneburger I, Eichinger A, Skerra A, Jung K. New insights into the signaling mechanism of the pH-responsive, membrane-integrated transcriptional activator CadC of Escherichia coli. Journal of Biological Chemistry. 2011;286(12):10681–10689.

33. Tetsch L, Koller C, Haneburger I, Jung K. The membrane-integrated transcriptional activator CadC of Escherichia coli senses lysine indirectly via the interaction with the lysine permease LysP. Molecular microbiology. 2008;67(3):570–583.

34. Rauschmeier M, Schüppel V, Tetsch L, Jung K. New insights into the interplay between the lysine transporter LysP and the pH sensor CadC in Escherichia coli. Journal of molecular biology. 2014;426(1):215–229.

35. Lindner E, White SH. Topology, dimerization, and stability of the single-span membrane protein CadC. Journal of molecular biology. 2014;426(16):2942–2957.

36. Chang HJ, Mayonove P, Zavala A, De Visch A, Minard P, Cohen-Gonsaud M, et al. A modular receptor platform to expand the sensing repertoire of bacteria. ACS synthetic biology. 2017;7(1):166–175.

37. Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science. 2013;339(6115):82–85.

38. Zhou J, Rudd KE. EcoGene 3.0. Nucleic acids research. 2012;41(D1):D613–D624.

39. Youngren B, Nielsen HJ, Jun S, Austin S. The multifork Escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. Genes & development. 2014;28(1):71–84.

40. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. The journal of chemical physics. 1953;21(6):1087–1092.

41. Weber SC, Spakowitz AJ, Theriot JA. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. Physical review letters. 2010;104(23):238102.

42. Espeli O, Mercier R, Boccard F. DNA dynamics vary according to macrodomain topography in the E. coli chromosome. Molecular microbiology. 2008;68(6):1418–1427.

43. Rouse Jr PE. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. The Journal of Chemical Physics. 1953;21(7):1272–1280.

44. Javer A, Long Z, Nugent E, Grisi M, Siriwatwetchakul K, Dorfman KD, et al. Short-time movement of E. coli chromosomal loci depends on coordinate and subcellular localization. Nature communications. 2013;4:3003.

45. Singer A, Schuss Z, Holcman D, Eisenberg R. narrow escape, part I. Journal of Statistical Physics. 2006;122(3):437–463.

46. Amitai A, Amoruso C, Ziskind A, Holcman D. Encounter dynamics of a small target by a polymer diffusing in a confined domain. The Journal of chemical physics. 2012;137(24):244906.

47. Geggier S, Kotlyar A, Vologodskii A. Temperature dependence of DNA persistence length. Nucleic acids research. 2011;39(4):1419–1426.

48. Carmesin I, Kremer K. The bond fluctuation method: a new effective algorithm for the dynamics of polymers in all spatial dimensions. Macromolecules. 1988;21(9):2819–2823.

49. Oswald F, Varadarajan A, Lill H, Peterman EJ, Bollen YJ. MreB-dependent organization of the E. coli cytoplasmic membrane controls membrane protein diffusion. Biophysical journal. 2016;110(5):1139–1149.

50. Epstein W, Kim BS. Potassium transport loci in Escherichia coli K-12. Journal of Bacteriology. 1971;108(2):639–644.

51. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison III CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods. 2009;6(5):343.

52. Miller VL, Mekalanos JJ. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in Vibrio cholerae requires toxR. Journal of bacteriology. 1988;170(6):2575–2583.

53. Dehio C, Meyer M. Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn5 in Bartonella henselae following conjugal plasmid transfer from Escherichia coli. Journal of bacteriology. 1997;179(2):538–540.

54. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of Escherichia coli K-12. science. 1997;277(5331):1453–1462.

55. Küper C, Jung K. CadC-mediated activation of the cadBA promoter in Escherichia coli. Journal of molecular microbiology and biotechnology. 2005;10(1):26–39.

56. Nielsen HJ, Ottesen JR, Youngren B, Austin SJ, Hansen FG. The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves. Molecular microbiology. 2006;62(2):331–338.

57. Paintdakhi A, Parry B, Campos M, Irnov I, Elf J, Surovtsev I, et al. Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. Molecular microbiology. 2016;99(4):767–777.

58. Seiler MC, Seiler FA. Numerical recipes in C: the art of scientific computing. Risk Analysis. 1989;9(3):415–416.

59. Socol M, Wang R, Jost D, Carrivain P, Vaillant C, Le Cam E, et al. Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. Nucleic acids research. 2019;.

60. Doi M, Edwards SF. The theory of polymer dynamics. vol. 73. oxford university press; 1988.