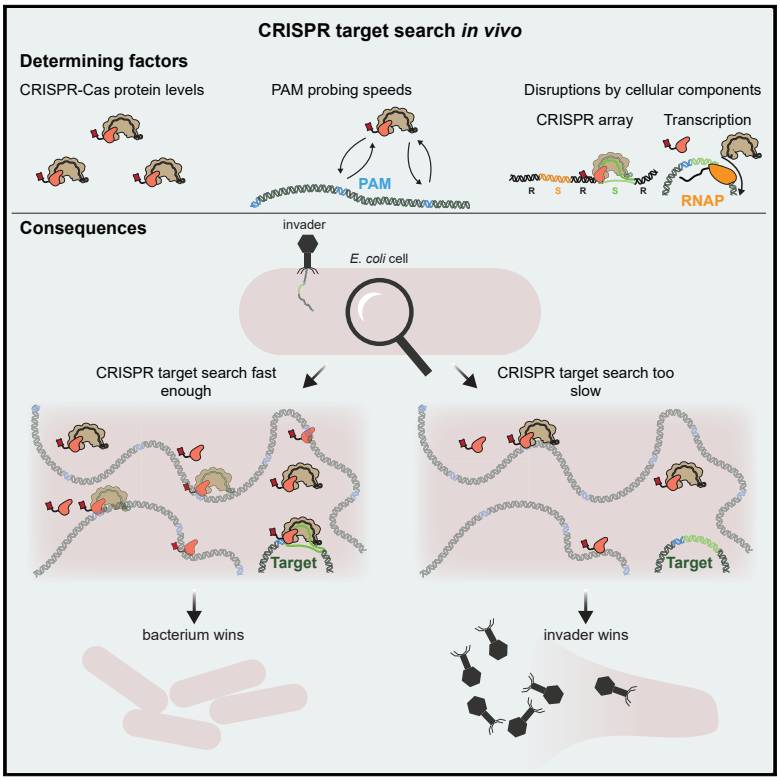
Article

Molecular Cell

# Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism

## Graphical Abstract




## Highlights

- 20 Cascade complexes are required to provide 50% protection
- Cascade spends equal time probing DNA (~30 ms) and diffusing to a next site
- Cas8e dynamically associates with Cascade in cells
- CRISPR target search and invader replication compete in a kinetic "arms race"

## Authors

Jochem N.A. Vink, Koen J.A. Martens, Marnix Vlot, ..., Daan J.W. Brocken, Johannes Hohlbein, Stan J.J. Brouns

## Correspondence

johannes.hohlbein@wur.nl (J.H.), stanbrouns@gmail.com (S.J.J.B.)

## In Brief

Vink et al. tracked single CRISPR RNA-surveillance complexes (Cascade) in the native host cell and determined the influence of Cascade copy numbers, PAM scanning speed, and the presence of CRISPR arrays and transcription on their ability to find and clear invading mobile genetic elements from the cell.

Vink et al., 2020, Molecular Cell 77, 39–50
January 2, 2020 
https://doi.org/10.1016/j.molcel.2019.10.021

CellPress

Molecular Cell
Article

# Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism

Jochem N.A. Vink,[1] Koen J.A. Martens,[2,3] Marnix Vlot,[4,7] Rebecca E. McKenzie,[1] Cristóbal Almendros,[1] Boris Estrada Bonilla,[1] Daan J.W. Brocken,[6,8] Johannes Hohlbein,[2,5,*] and Stan J.J. Brouns[1,9,*]
[1]Kavli Institute of Nanoscience, Department of Bionanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, the Netherlands
[2]Laboratory of Biophysics, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands
[3]Laboratory of Bionanotechnology, Wageningen University & Research, Bornse Weilanden 9, 6708 WG Wageningen, the Netherlands
[4]Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands
[5]Microspectroscopy Research Facility, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands
[6]Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, Leiden, the Netherlands
[7]Present address: TropIQ Health Sciences, Transistorweg 5-C02, 6534 AT Nijmegen, the Netherlands
[8]Present address: Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, 2333 CC Leiden, the Netherlands
[9]Lead Contact
*Correspondence: johannes.hohlbein@wur.nl (J.H.), stanbrouns@gmail.com (S.J.J.B.)
https://doi.org/10.1016/j.molcel.2019.10.021

## SUMMARY

**CRISPR-Cas systems encode RNA-guided surveillance complexes to find and cleave invading DNA elements. While it is thought that invaders are neutralized minutes after cell entry, the mechanism and kinetics of target search and its impact on CRISPR protection levels have remained unknown. Here, we visualize individual Cascade complexes in a native type I CRISPR-Cas system. We uncover an exponential relation between Cascade copy number and CRISPR interference levels, pointing to a time-driven arms race between invader replication and target search, in which 20 Cascade complexes provide 50% protection. Driven by PAM-interacting subunit Cas8e, Cascade spends half its search time rapidly probing DNA (~30 ms) in the nucleoid. We further demonstrate that target DNA transcription and CRISPR arrays affect the integrity of Cascade and affect CRISPR interference. Our work establishes the mechanism of cellular DNA surveillance by Cascade that allows the timely detection of invading DNA in a crowded, DNA-packed environment.**

## INTRODUCTION

RNA-guided CRISPR-Cas surveillance complexes have evolved to specifically and rapidly recognize sequences of previously cataloged mobile genetic elements (MGEs) (Marraffini, 2015). Target DNA recognition depends on CRISPR RNA (crRNA)-DNA complementarity and on the presence of a protospacer adjacent motif (PAM), a short nucleotide sequence flanking the target site (Deveau et al., 2008; Mojica et al., 2009). To work effectively, the complexes need to find their targets fast enough to prevent an MGE from becoming established in the cell, which can occur within minutes upon cell entry (Shao et al., 2015). Target search inside a cell faces a multitude of challenges. The first challenge is that cells are packed with DNA, and crRNA surveillance complexes need to find the "needle in a haystack" before an invading element takes control of the cell. PAM scanning and crRNA-seed interactions with the target have been suggested to speed up the search process by drastically reducing the number of potential target sites in the genome (Gleditzsch et al., 2019; Jones et al., 2017; Martens et al.,2019). Several studies have shown that crRNA-effector complexes spend more time probing PAM-rich regions, which is indicative of its function as the first recognition site (Globyte et al., 2019; Redding et al., 2015; Sternberg et al., 2014). The *Escherichia coli* K12 genome contains 127,081 preferred PAMs (CTT) that are recognized by the crRNA-effector complex Cascade in the type I-E CRISPR-Cas system (Leenay et al., 2016). This large number of PAMs suggests that the interaction with the PAM needs to be sufficiently fast to cover enough sequence space to find an invading DNA sequence in time. A second challenge is posed by the action of other proteins present in the cell such as DNA-binding proteins, DNA or RNA polymerases that may interfere with target search, and the formation of target-bound crRNA complexes (Jones et al., 2017; Vigouroux et al., 2018). Some invading MGEs even use specialized anti-CRISPR proteins to inhibit crRNA-effector complexes and impair the target search process (Bondy-Denomy et al., 2015; Pawluk et al., 2014). A third challenge that microbes face is to produce appropriate levels of Cascade complexes loaded with one particular crRNA to provide protection against a single invading element. While adding more and more spacers to CRISPR arrays will afford the benefit of recognizing many invaders, the tradeoff is that long CRISPR arrays will dilute the number of

Cascade complexes loaded with a particular crRNA, potentially decreasing the CRISPR response against that target. These cellular challenges raise the question of how Cascade can navigate the crowded cell sufficiently fast to find DNA targets, and how many copies of Cascade are required to do so.

Here, we report the visualization of single-molecule type I-E Cascade complexes in a native *Escherichia coli* CRISPR-Cas system *in vivo*. We found that the probability of successful CRISPR protection depends exponentially on Cascade copy numbers, which leads to a time-driven arms race model between Cascade target search and invader replication. The localization of Cascade shows that the complex is enriched inside the nucleoid. We determined that 60% of the Cas8e subunit is incorporated into Cascade complexes and that Cascade DNA probing interactions are very rapid (~30 ms) and are driven by Cas8e. Furthermore, the transcription of targets and CRISPR arrays reduces the number of functional complexes in the cell. Our work sheds light on the target search and dynamical assembly of Cascade complexes in their native cellular environment and describes how these processes affect CRISPR protection levels.

## RESULTS

### Visualizing Cascade Abundance and Target Search at the Single-Molecule Level

To investigate how microbes cope with these challenges at the cellular level, we used intracellular single-particle tracking photoactivated localization microscopy (sptPALM) (English et al., 2011; Manley et al., 2008), a technique that is capable of following with high precision the movement and abundance of individual fluorescently tagged proteins in cells. By genetically fusing a photoactivatable fluorescent protein (PAmCherry2; Subach et al., 2009) to the N terminus of Cascade subunit Cas8e (Figure 1A), which was the only subunit for which labeling had no influence on the CRISPR interference ability of this strain (Figure 1B), we were able to monitor the mobility and abundance of Cascade complexes in *E. coli* cells.

### Twenty Cascade Complexes Provide 50% CRISPR Protection

We wanted to link the copy number of Cascade to a successful target search and established an assay that measures the level of CRISPR protection in cells at the time of cell entry by a MGE. In this assay, all of the Cascade complexes present in the cell must be able to target the incoming MGE and the Cascade target search must be rate limiting. To meet the first requirement, we constructed a high copy plasmid (pTarget; Figure 1B) containing target sites for all 18 spacers found in the genomic arrays of *E. coli* K12, such that all Cascade complexes would be targeting the incoming plasmid. Second, we ensured that Cascade copy numbers were rate limiting (Majsec et al., 2016) by equipping cells with a low copy plasmid expressing the nuclease Cas3 (pCas3, adapted from Westra et al., 2010).

We achieved different expression levels of Cascade in the cell by tuning the expression of the native regulator LeuO (Westra et al., 2010) (Figure 1C). The copy numbers of Cascade under these varying levels of LeuO induction were estimated from the number of fluorescent particles present in the cell, taking complex assembly (see the following section), growth rate (Table S1), and maturation time of PAmCherry into account (Figure 1D; Method Details). We found that the average number of Cascade complexes per cell in the absence of LeuO induction was low (~4 copies) and that copy numbers increased >30-fold for the highest induction level (~130 copies). We measured the interference ability under these conditions by determining the probability that pTarget becomes established in a cell. We observed that establishment of pTarget decreases sharply with the increasing copy numbers of Cascade (Figure 1E). However, even with 130 Cascade complexes present, we still observed a level of pTarget survival (~0.5%).

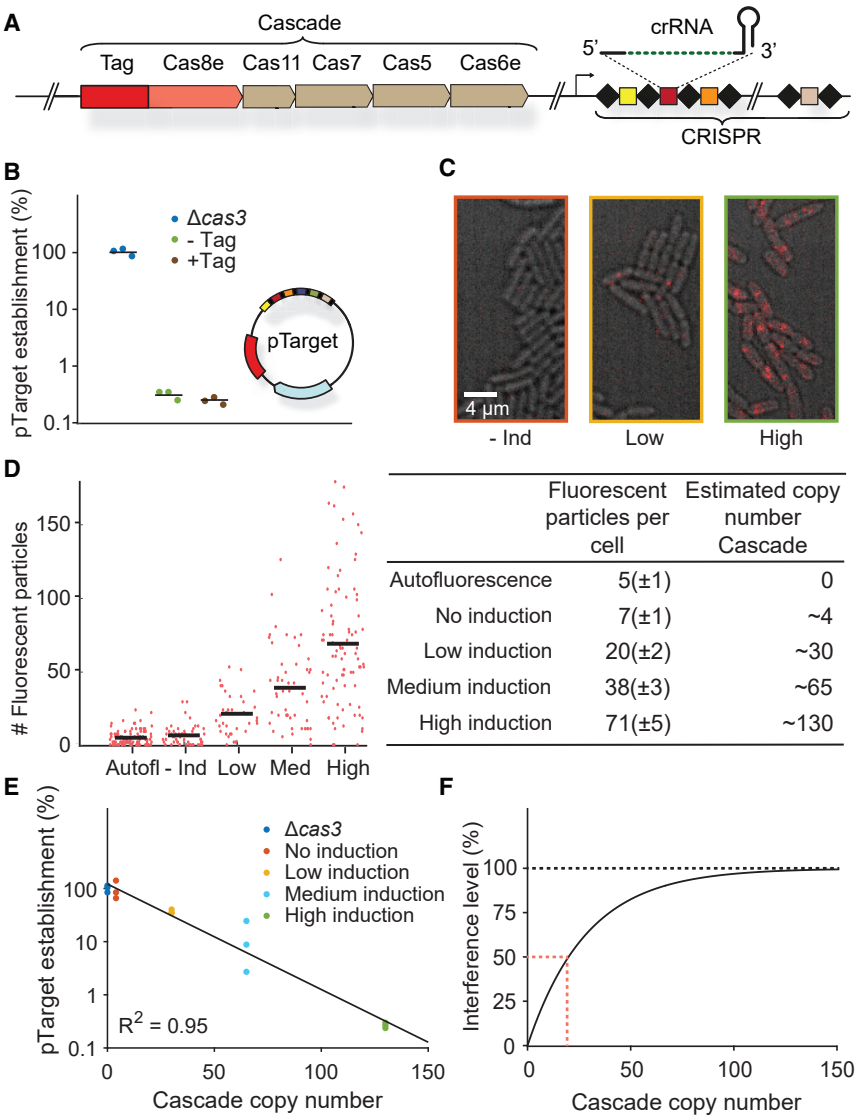
To explain these observations, we modeled the probability that an invading MGE becomes established in the cell, depending on the number of Cascade complexes that target this specific MGE. The model is based on multi-copy plasmids and phage systems, in which the DNA clearance is most likely to occur when an invader enters as a single copy, as the concentration of invading DNA increases over time. Therefore, depending on the invader and the level of CRISPR interference, there will be a critical time point ($t_c$) beyond which the invader is permanently established inside the cell and can no longer be cleared (Severinov et al., 2016). Our model describes the probability that it takes a certain copy number of proteins ($n$), each with an average search time ($\overline{t_s}$), to find the target before $t_c$ is reached.

Our model accurately predicted that pTarget establishment decreases exponentially with increasing copy numbers of Cascade (Figure 1E; Method Details). When we translated these establishment probabilities into interference levels, we could deduce that ~20 Cascade complexes are required to reach a CRISPR interference level of 50% (Figure 1F). The exponential relationship further entails every subsequent 20 complexes half the number of cells not able to achieve interference, which means that 40 Cascade complexes can provide 75% interference and 60 Cascade complexes can provide 87.5% interference.

It becomes very unlikely for the CRISPR system to destroy multiple genetic copies of the MGE if it has failed to destroy the single copy that was present at the start before replication. Therefore, we can approximate $t_c$ using the replication time of the plasmid in the absence of copy number control (~3 min; Olsson et al., 2003), which allows us to retrieve an estimated search time of ~90 min for 1 Cascade complex to find a single target in the cell (see Method Details).

In contrast to pTarget establishment, which decreases exponentially, the average search time decreases linearly with increasing copy numbers of Cascade. Therefore, 10 Cascade complexes require ~9 min to find a single target, while 90 Cascade complexes could achieve this within 1 min.

To summarize, we found a direct relation between the number of Cascade complexes and the establishment probability of an MGE. The native *E. coli* system requires 20 Cascade complexes loaded with a cognate crRNA to obtain 50% CRISPR interference levels. This relation depends on the replication rate of the invading MGE and the average search time of a single complex and demonstrates the importance of rapid target search on CRISPR interference ability.




| | Fluorescent particles per cell | Estimated copy number Cascade |
|---|---|---|
| Autofluorescence | 5(±1) | 0 |
| No induction | 7(±1) | ~4 |
| Low induction | 20(±2) | ~30 |
| Medium induction | 38(±3) | ~65 |
| High induction | 71(±5) | ~130 |

**Figure 1. Cascade Copy Number versus CRISPR Protection**

(A) Chromosomal locus of the Cascade subunits and integration site of the photoactivatable fluorescent protein upstream of *cas8e*.

(B) pTarget establishment, calculated from the ratio of the transformation of pTarget/pGFPuv, is a measure of the interference level of the CRISPR system. To test whether tagged Cascade complexes were able to function normally, we compared the tagged strain to the untagged and the *Δcas3* strain. pTarget (bottom right) contains protospacers for all of the spacers in the K12 genome (colored, not all depicted) and are flanked by a 5′-CTT-3′ PAM (black bars).

(C) Overlay of bright-field image of cells (gray) and single-molecule signal (red) from a single representative frame for different induction levels.

(D) Number of fluorescent particles measured in each cell plotted for different levels of Cascade expression (left). The mean number of fluorescent particles (±SD; table left column) was converted to a Cascade copy number (table right column; see Method Details).

(E) pTarget establishment plotted for different copy numbers of Cascade. The data points were fitted with an exponential decay function. $pTarget\ establishment = e^{-an}$, where $n$ equals Cascade copy number and $a$ is the fitted coefficient. In our model, $a = \overline{t_s}/t_c$.

(F) The fitted exponential decay on the left converted into an interference level ($Interference\ level = 1 - pTarget\ establishment$). Indicated in red (dashed lines) is the amount of Cascade copies required for 50% interference.

## The Majority of Cas8e Assembles into the Cascade Complex

To quantify the dynamics of the target search, we traced the diffusion paths of thousands of individual complexes in the bacterial cell (Figure 2A; Video S1). The apparent diffusion coefficient $D^*$, a measure for mobility, of Cascade was calculated by extracting the displacement of each fluorescent particle for 4 consecutive 10-ms steps, allowing us to investigate the abundance, mobility, and behavior of individual complexes and subunits in the cell. To minimize the influence of spurious autofluorescent particles in *E. coli* (Floc'h et al., 2018), we used expression levels with the highest estimated Cascade copy numbers (~130 copies, high induction; Figure 1D).

To distinguish the diffusion of Cascade complexes from monomeric Cas8e subunits, we measured the diffusion of the tagged Cas8e fusion protein in a strain lacking genes of the other 4 Cascade subunits in the genome (Cas11, Cas7, Cas5, and Cas6e). Based on the role of Cas8e in non-specific DNA binding (Dillard et al., 2018; Jore et al., 2011; Sashital et al., 2012), we expected to find mobile and DNA-bound populations of Cas8e. However, we were unable to describe the data accurately by static two-state models of non-interconverting fractions (Figure S1). We therefore hypothesized that rapid DNA-binding and -unbinding events of Cascade on a timescale similar to the frame rate (~10–40 ms) would lead to time averaging of a mobile state (high $D^*$ values) and a DNA-bound state (low $D^*$ values), giving rise to intermediate $D^*$ values (Figure 2A). We accounted for these events by developing a generally applicable analysis method called analytical diffusion distribution analysis (analytical DDA), which is useful for proteins with fast transitioning kinetics between states with different diffusion coefficients, such as DNA-interacting proteins. The distribution of $D^*$ values is not only affected by the fraction of the time spent bound and freely diffusing but also changes depending on the absolute transition rates (Figure S2). Therefore, this method allows us to extract quantitative information on DNA-binding kinetics and enables the study of fast transition rates previously inaccessible to sptPALM (Method Details).

When we applied the analytical DDA on the Cas8e diffusional data, we retrieved an average residence time of ~30 ms on DNA and a similar average time spent (~30 ms) rapidly diffusing

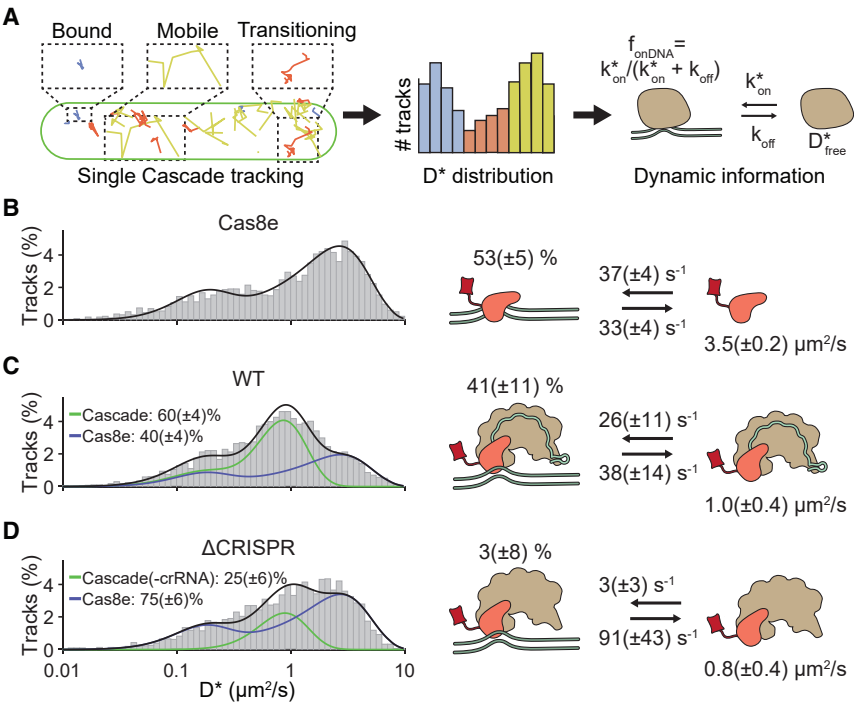



**Figure 2. Diffusion Behavior of Cas8e and Cascade**

(A) Tracks with small (blue), intermediate (orange), and large (yellow) displacements from a single cell of the wild-type (WT) strain (left). The most likely state for three tracks is indicated, although due to limited track length and fast transitions, states cannot be assigned confidently to every individual track. The $D^*$ distribution (center), from a large population of tracks, enables the reliable extraction of DNA interaction kinetic parameters (pseudo-first order on rate ($k^*_{on}$), off rate ($k_{off}$, and the apparent free diffusion coefficient ($D^*_{free}$)) by using analytical diffusion distribution analysis (DDA; right). These parameters further allow the calculation of the fraction DNA bound ($f_{onDNA}$).

(B–D) $D^*$ distributions for (B) Cas8e, (C) Cascade, and (D) Cascade in a ΔCRISPR strain. Total (black), Cas8e (blue), and Cascade (green) fractions fits are indicated by lines. The parameters (right) of Cas8e (B) were used to fit the Cas8e fraction in Cascade (C and D). Error estimation is based on bootstrapping (±SD).

See also Figures S1, S2, and S3 and Video S1.

($D^*$ ~3.5 μm$^2$/s, as expected for a protein of 82 kDa; Method Details), indicating that Cas8e is bound to DNA for ~50% of the time (Figure 2B). The $D^*$ distribution of Cas8e then allowed us to extract the diffusion behavior of the Cascade complex as a whole. We estimated the fraction of free Cas8e and Cascade-containing Cas8e at 40% and 60%, respectively (Figure 2C). This finding suggests that Cas8e is produced in excess (Westra et al., 2010) or is somehow involved in a dynamic interaction with the core Cascade subunits (crRNA, Cas11, Cas7, Cas5, Cas6e) (Jore et al., 2011; Sashital et al., 2012).

We were surprised to find that the DNA-binding kinetics of Cascade were similar to Cas8e alone, indicating that Cas8e is an important driver of DNA probing characteristics of the Cascade complex. Furthermore, the DNA residence times are on average ~30 ms and are thereby considerably shorter than the 0.1–10 s that have been reported for *in vitro* studies previously (Dillard et al., 2018; Redding et al., 2015; Xue et al., 2017). As expected, we found a smaller diffusion coefficient for unbound Cascade complexes (~1.0 μm$^2$/s) (Method Details) due to their larger size. Our analysis shows that >50% of the Cas8e proteins are part of the intact Cascade complexes and that the DNA interacting behavior of Cascade is largely determined by the properties of Cas8e.

The probing kinetics that we measured determine the number of sites that Cascade can scan every minute. The total time that Cascade needs to probe a single site includes the average time that the complex is bound to a DNA site and the average time that it requires to find the next DNA site. The Cascade probing time *in vivo* sums up to ~60 ms ($1/k_{off} + 1/k^*_{on}$), which implies that the complex is able to scan ~1,000 DNA sites per minute. The probing kinetics of single sites are furthermore linked to the distributions of target search times, and with simulations, we could verify that our model of Cascade DNA scanning leads to the expected distribution of interference levels (Figure S3). Using our previous estimate of the overall target search time for a single Cascade of ~90 min, we calculate that the complex scans 90,000 DNA sites in the cell before finding a target (see Method Details).

To investigate the role of crRNAs in the Cascade complex assembly, we deleted all of the CRISPR arrays in the K12 genome (ΔCRISPR). The resulting diffusion behavior can be described by fractions of free Cas8e and with Cascade-like diffusion behavior (Figure 2D) that almost entirely lacks interaction with DNA ($f_{onDNA}$ = 3%). This indicates that although Cascade (sub)complex formation does not strictly require the presence of crRNA (Beloglazova et al., 2015; Brouns et al., 2008), Cascade assembly is greatly enhanced by crRNA. The majority of Cas8e proteins are incorporated in Cascade complexes in the presence of crRNA, and this gives Cascade DNA interacting properties.

## Cascade Is Enriched but Not Exclusively Present in the Nucleoid

Not all potential DNA interaction sites in the host chromosome may be accessible to Cascade. The host DNA is concentrated in the center of the cell in the nucleoid and is very compact, which excludes large complexes such as ribosomes (Mondal et al., 2011). Nucleoid exclusion would reduce the amount of DNA available for scanning and increase the amount of freely diffusing Cascade complexes. To investigate whether the DNA-bound fraction is governed by affinity properties of Cascade for DNA rather than a restricted search space outside the DNA-containing nucleoid region, we studied the spatial distribution of Cascade localizations. Nucleoid-excluded ribosomes are enriched away from the central long axis of the cell

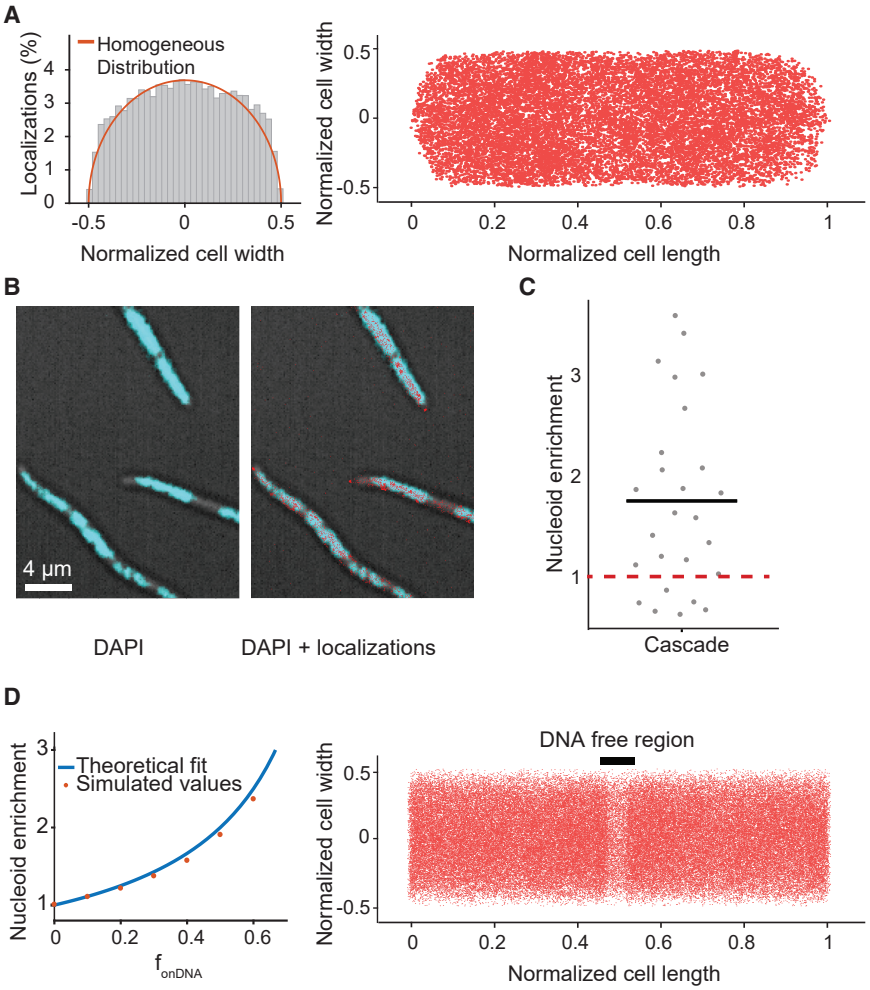



**Figure 3. Cascade Localization Inside the Cell**

(A) Localization of Cascade in the cell. Left: distribution of Cascade over the cell width (n = 33 cells; 15,428 localizations). Orange indicates the expected distribution in the case of a homogeneous localization within the cell. Right: same localizations plotted within the dimensions of single cells in which the cell length and cell width of each cell was normalized.

(B) Overlay of DAPI fluorescence and bright-field image (left) with Cascade localizations (right) in cephalexin-treated cells.

(C) The nucleoid enrichment in the WT strain (27 subregions in 18 cells). The average ratio is indicated with a black bar. The expected nucleoid enrichment of Cascade in the absence of DNA interactions is indicated in red (dashed lines).

(D) Relation between DNA bound fraction and nucleoid enrichment. Left: a theoretical relation between nucleoid enrichment and the DNA bound fraction was derived (Method Details) and compared to simulated values for different amounts of $f_{onDNA}$. Right: localizations of simulated Cascade proteins (n = 50,000) diffusing through part of an elongated cell are plotted on top of long cell axis. A DNA-free region (black bar) is visible due to the enrichment of Cascade binding to DNA in nucleoid regions. Simulations of particles were performed with an off rate of 38 $s^{-1}$ and an on rate of 26 $s^{-1}$ to reach a nucleoid enrichment of 1.8, similar to the average that was found for Cascade.

(Sanamrad et al., 2014). For Cascade, we found a homogeneous spatial distribution throughout the cell (Figure 3A), indicating that Cascade is small enough to freely scan the nucleoid for target sites.

We furthermore used the spatial distribution of Cascade to extract quantitative information on the DNA-bound fraction. To this purpose, we created a DNA-free environment in the cell by adding cephalexin (Reyes-Lamothe et al., 2014). This antibiotic affects cell wall synthesis and causes cells to elongate, forming DNA-free cytoplasmic space between nucleoids without condensing the nucleoid (Figure 3B). The time that Cascade is bound to DNA is inherently linked to the relative amount it spends in DNA-free and DNA-containing regions. Therefore, by calculating the relative amount of localizations in both regions (enrichment factor [*EF*]), we can extract the fraction of time spent on DNA independently from the DDA analysis. Cascade was only moderately enriched (*EF* 1.8- ± 0.2-fold) in the nucleoid regions (Figure 3C), indicating that Cascade spends a considerable amount of time diffusing in the cytoplasm while not associated with DNA. From the *EF*, the fraction of Cascade complexes bound to DNA can be approximated to 45% (Figure 3D; for derivation, see Method Details). This value is consistent with the ~50% value we extracted from the DDA distribution of Cascade (Figure 2C). However, it strongly contrasts other DNA-binding proteins such as Fis and RNA polymerase, which show a much higher nucleoid enrichment (Reyes-Lamothe et al., 2014; Stracy et al., 2015). The above findings indicate that Cascade inherently spends more time freely diffusing the cell and that this is caused by the nature of DNA-Cascade interactions and not by size-based nucleoid exclusion, as is the case for ribosomes (Sanamrad et al., 2014). Therefore, we decided to study the nature of the DNA interactions in more depth.

## Cascade-DNA Interactions Are Not Only PAM Dependent

Next, we assessed how PAM interactions contributed to DNA binding by introducing mutation G160A in the Cas8e subunit, which abolishes the interaction with the PAM (Hayes et al., 2016). This G160A mutation decreased the fraction of DNA-bound Cascade from 41% ± 11% to 28% ± 6% (Figure 4A) without fully inhibiting DNA binding, suggesting that PAM-independent interactions (van Erp et al., 2015; Hayes et al., 2016; Xiao et al., 2017) also play a role in DNA probing. To assess the contribution of these different types of interactions to the average DNA residence time found previously, we measured the persistence of Cascade-DNA interactions by increasing the

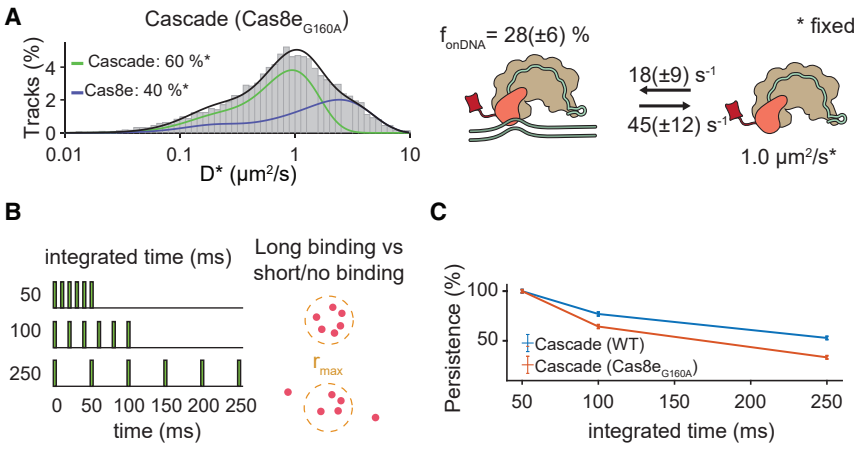



**Figure 4. PAM-Dependent and PAM-Independent DNA Probing**

(A) *D** distributions for Cascade and Cas8e with a mutation (G160A) deficient in PAM binding. To compare kinetic rates, we assumed that the relative Cas8e-Cascade fractions and the diffusion of free Cascade and Cas8e were not altered by the mutation, and those values were fixed.

(B) Depiction of persistence analysis. Increasing the integration time while keeping the exposure time constant and counting the number of localizations within a certain radius allow the calculation of the persistence of binding events.

(C) The relative amount of long binding events (6 consecutive localizations within $r_{max}$: 1 pixel [0.128 μm] of the mean position) for WT and PAM binding mutant Cascade normalized to 50 ms integration time.

Error estimation in (A) and (C) is based on bootstrapping (±SD).

dark time between exposures (Figure 4B). Our data showed that sustained binding events at longer timescales (100–250 ms) were more frequently observed for WT Cascade than for the PAM-binding mutant complex Cascade-Cas8e$_{G160A}$ (Figure 4C). Together with the increased off rate of the mutated complex (Figure 4A), this finding demonstrates that PAM-dependent interactions of Cascade with DNA last longer than PAM-independent interactions.

## Target DNA Binding Is Influenced by the Cellular Environment

After establishing the intrinsic DNA probing characteristics of Cascade, we investigated its diffusion behavior in the presence of targets (Figure 5). To prevent target DNA degradation by Cas3 nucleases, we deleted the *cas3* gene and verified that the deletion did not alter Cascade diffusion behavior (Figure S4). To verify that all Cascade complexes could bind a target, we measured the copy number of pTarget to be ~400 copies per cell (Figure S5). As the native *E. coli* CRISPR arrays contain 18 spacers, this resulted in ~7,000 target sites per cell, which far outnumbers the Cascade copy numbers under our growth conditions (~130; Figure 1D).

Compared to a non-targeted control plasmid (Figure S4), the introduction of pTarget in cells decreased the fraction of free Cascade complexes (from 60% ± 4% to 29% ± 3%) and gave rise to a 34% ± 2% immobile, target-bound Cascade fraction ($D^*_{Cascade(bound)}$ = 0.06 μm²/s) (Figure 5A). As expected, the addition of pTarget increased the persistence of sustained binding events, indicating specific DNA target binding (Figure 5C). The combined information of plasmid copy number and the ratio of probing to target-bound Cascade enabled us to determine a cellular $K_D$ value for the affinity of Cascade for targets of ~180 nM (Figure 5F; Method Details), indicating that the affinity *in vivo* is ~10 times lower than what has been observed *in vitro* (Hayes et al., 2016).

We hypothesized that the transcription of DNA along target sites would be one of the main factors influencing Cascade target DNA binding. To investigate the effects of transcription by host RNA polymerase (RNAP), we introduced a (lac) promoter in front of the pTarget sequence. To our surprise, we observed that the affinity of Cascade for target sites that undergo transcription (~100 nM) was higher than it was for non-transcribed target sites (~180 nM; probing/target bound Cascade from 0.5 ± 0.1 to 0.9 ± 0.1). In addition, we observed an increased fraction of free Cas8e subunits (from 37% ± 2% to 54% ± 2%) in the strain containing transcribed pTarget (Figure 5B). These findings suggest that the transcription of a target DNA sequence somehow facilitates target search and increases the affinity of a target. In addition, it appears that collisions of RNAP with target-bound Cascade result in changes in the Cascade assembly, likely by dissociation of the Cas8e subunit from the complex upon collision with RNAP, which potentially dissociates Cascade from the target.

The relatively dynamic association of Cas8e within the Cascade complex has been observed previously *in vitro* (Jore et al., 2011) and was more recently observed upon binding to the CRISPR array (Jung et al., 2017). We hypothesized that this dynamic behavior may be a functional characteristic and will also occur upon encountering CRISPR arrays inside the cell. To test this hypothesis, we made a variant of pTarget in which all 18 interference PAMs were replaced by the trinucleotide sequence matching the repeats of the CRISPR array (pCRISPR1). Cascade did not show any interaction with the non-transcribed pCRISPR1 plasmid (Figure 5D). However, when we added a promoter sequence in front of the pCRISPR1 array of targets, we observed moderately enhanced levels of free Cas8e (from 40% ± 1% to 56% ± 1%) (Figure 5E), reminiscent of Cas8e expulsion from the complex upon collision with RNAP or from targets with repeat-like PAMs (Jung et al., 2017). Effectively, this shows that transcribed CRISPR arrays may function as target decoys in the cell and can therefore influence the levels of functional Cascade complexes in the cell.

To test whether CRISPR arrays really form decoys in the cell and could affect interference levels, we constructed a compatible high copy number plasmid pCRISPR2 containing a normal CRISPR array (Figure S6). While the introduction of pCRISPR2 into cells containing pTarget only led to a small decrease in the number of Cascade complexes (15% less)

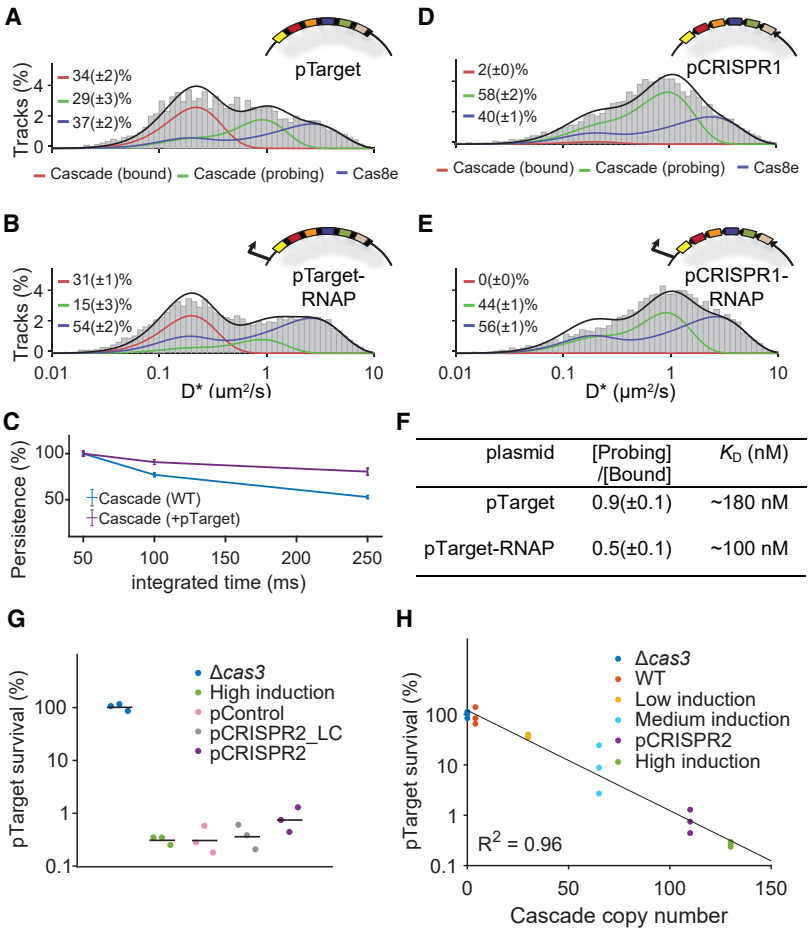



**Figure 5. Cascade-DNA Interactions in the Presence of Targets**

(A and B) $D^*$ distribution for the Δ*cas3* strain carrying pTarget (A) and pTarget-RNAP (B). pTarget contains protospacers for all of the spacers in the K12 genome (colored, not all depicted) and are flanked by a 5′-CTT-3′ PAM (black bars). Cascade (probing) (green) and Cas8e (blue) fractions were fitted with parameters from Figures 1C and 1D, and a new target-bound fraction (Cascade [bound]) was introduced as a single diffusion state ($D^*$ = 0.06 μm²/s (+$\sigma^2$/t); red).

(C) The abundance of sustained binding events as in Figure 3C, but for WT and pTarget-carrying cells.

(D and E) $D^*$ distribution for the *Δcas3* strain carrying pCRISPR1 (D) and pCRISPR1-RNAP (E). pCRISPR1 contains the same protospacers as pTarget that are now flanked by repeat PAMs.

(F) In *vivo* $K_D$ estimates based on the ratio between probing/bound Cascade and the plasmid copy number (Figure S5; Method Details).

(G) pTarget establishment for *Δcas3* (blue), WT (high induction; green), an empty high copy plasmid (pControl; pink), and low or high copy plasmids carrying CRISPR arrays (pCRISPR2_LC/pCRISPR2; gray/purple). Each dot represents an independent biological replicate.

(H) pTarget establishment plotted for different copy numbers of Cascade. The same as Figure 1E, but with the addition of pCRISPR2. The Cascade copy number of the pCRISPR2 strain was estimated from the relative abundance of the Cascade (probing) fraction in the WT (high induction; Figure 2C) and pCRISPR2 (Figure S4) strain. Each dot represents an independent biological replicate.

Error estimation in (A)–(F) is based on bootstrapping (±SD). See also Figures S4, S5, and S6.

(Figure S4), the CRISPR interference levels were reduced by as much as 50% (Figure 5G). This effect was not observed with a low copy variant of pCRISPR2 (pCRISPR2_LC) or with a high copy plasmid lacking CRISPR arrays (pControl), indicating that this effect comes from the presence of a large number of CRISPR arrays in the cell (Figure 5G). We further found that the observed impact of CRISPR arrays on the Cascade copy number and interference level fits well with our previously predicted relation between Cascade copy numbers and the probability of successful MGE establishment (Figure 5H). It furthermore demonstrates how relatively small changes in Cascade copy numbers (15%) can have a significant impact on CRISPR interference levels (50%). Our data indicate that Cascade target search and binding is strongly influenced by the action of RNAP and that CRISPR arrays form target decoys in the cell, which can affect CRISPR interference levels.

## DISCUSSION

How crRNA-effector complexes can achieve the timely detection of incoming MGEs in the crowded environment of the cell is an intriguing aspect of CRISPR biology that remains poorly understood. We provide the first insights into the fundamental kinetics of the surveillance behavior of type I crRNA-effector complexes in their native cellular environment. We determined how many copies of Cascade are required to establish effective immunity and uncovered how Cascade complexes navigate the crowded bacterial cell packed with DNA. Our results indicate that Cascade does not restrict its search space to parts of the cell—for example, the nucleoid-free periphery—but instead is occupied scanning the entire host nucleoid for a match. Depending on the genome size of a microbe and the number of copies of the genome in the cell, the nucleoid size may vary widely. To cover this vast sequence space sufficiently fast, the Cascade complex interrogates DNA sequences by using a combination of PAM-dependent and PAM-independent interactions that on average last only 30 ms. This probing interaction is much faster than the previously reported interaction times determined for type I Cascade complexes by *in vitro* methods, which range between 0.1 and 10 s (Dillard et al., 2018; Redding et al., 2015; Xue et al., 2017). The ability to rapidly probe DNA sequences for potential matches with the crRNA and to move from one place in the nucleoid to the next may explain how a relatively low number of Cascade complexes in *E. coli* may still confer CRISPR immunity. The average probing time of 30 ms for Cascade matches the values found for *Streptococcus pyogenes* dCas9 in *E. coli* (Jones et al., 2017) and *L. lactis* (Martens et al., 2019), suggesting that DNA probing interactions of crRNA-effector

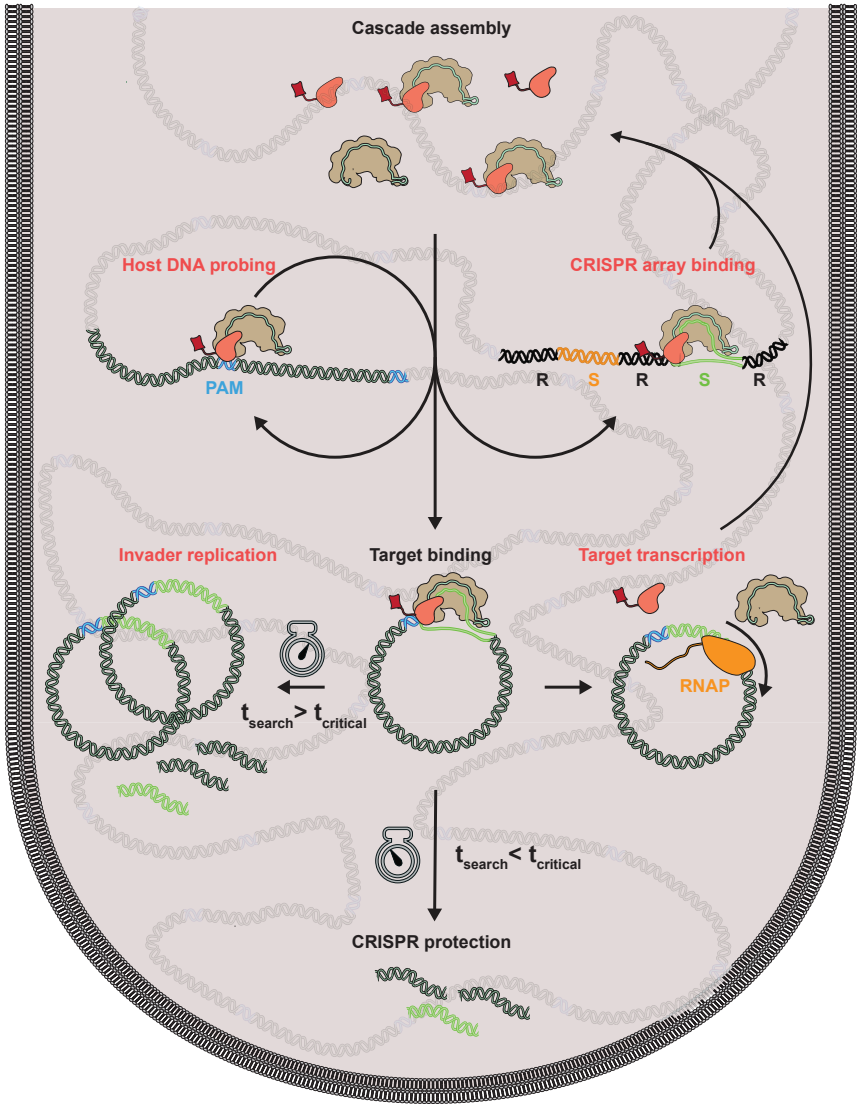



**Figure 6. Model of How Cascade Protects the Cell**

Successful protection against an invader requires Cascade target search to circumvent several potential diversions (red). After Cascade is assembled, the complex probes the host DNA by rapidly binding and dissociating. It uses PAM-dependent and PAM-independent DNA interactions and scans the entire nucleoid region. If it binds to a CRISPR array (S, spacer; R, repeat), then the complex disintegrates. When it has found its target, it depends on the search time ($t_s$) and the critical time ($t_c$) whether the invader is cleared and the cell is protected or the invader can replicate and establish itself in the cell. Moreover, transcription by RNA polymerase (RNAP) can still remove bound complexes, compromising CRISPR protection.

complexes from both class I and II systems may have evolved independently to take place at this timescale.

The probing kinetics that we measured for Cascade will allow the complex to scan 1,000 DNA sites per minute. Given the abundance of PAMs in the host DNA, this interaction time would lead to a search time on the order of hours. This value matches our independently calculated estimate of 1.5 h for a single Cascade to find a single DNA target in the cell, which is 4 times faster than the dCas9 search time estimates of 6 h (Jones et al., 2017). However, our data also indicate that Cascade not only probes PAMs but also the complex spends a considerable amount of time engaged in PAM-independent DNA interactions. These may be constituted by direct crRNA-DNA interactions (Blosser et al., 2015; Xue et al., 2016) or electrostatic interactions of Cascade with the DNA (van Erp et al., 2015; Hochstrasser et al., 2014). This suggests that an even larger DNA sequence space needs to be covered, creating the need for even more efficient and functionally flexible surveillance solutions. This more flexible probing behavior would be required to recognize targets with mutations in the PAM or protospacer to trigger a CRISPR memory update pathway called priming (Datsenko et al., 2012; Jackson et al., 2017), which appears to be unique for type I CRISPR-Cas systems.

One possibility to reconcile Cascade DNA probing characteristics to the overall search time could be that Cascade undergoes facilitated 1-dimensional (1D) DNA sliding, where Cascade probes multiple sites per DNA-binding event. We have shown that Cascade spends 50% of its search time on DNA and the other 50% diffusing to a new site in the cytoplasm. This value may seem low compared to other DNA interacting proteins such as transcription factor LacI, which is DNA bound for 90% of the time (Elf et al., 2007). However, 50% has been theoretically derived as the optimum for a target search process involving 1D DNA sliding and 3D translocation/hopping (Slutsky and Mirny, 2004). Recently, it has been shown *in vitro* that Cascade and Cas9 can slide along the DNA in search of targets (Dillard et al., 2018; Globyte et al., 2019). If this also occurs *in vivo*, then this would be a striking example of a DNA-binding protein having an optimized time division between DNA-bound and freely mobile states to survey the DNA content of the cell.

The relatively high abundance (50%) of freely diffusing Cascade complexes may have benefits as well, as this will lead to more Cascade complexes in the periphery of the cell outside the nucleoid. By surveying these peripheral regions more frequently, Cascade may be able to detect incoming bacteriophages or plasmid DNA more rapidly when these genetic elements enter the cell.

Besides the chromosomal host DNA, other cellular constituents also affect target DNA-binding properties. We found a much higher $K_D$ value *in vivo* (180 nM) than was reported earlier

using *in vitro* methods (20 nM) (Hayes et al., 2016). The discrepancy in binding affinity between *in vivo* and *in vitro* measurements may be caused by an increase in target search time (i.e., a lower on rate) or an increase in target dissociation rate (i.e., a higher off rate) *in vivo*. In any scenario, this discrepancy highlights the strong role of the crowded cellular environment on target binding.

Counterintuitively, we have found that Cascade binds transcribed target sites with higher affinity (100 nM) than non-transcribed target sites (180 nM). Previous studies have shown that negative supercoiling is required for Cascade binding (Westra et al., 2012) and that increased negative supercoiling accelerates the rate of R-loop formation (Szczelkun et al., 2014). As transcribed regions cause more negative supercoiled regions in the DNA (Ma and Wang, 2016), this could explain the increase in the affinity for transcriptionally active sites. The rates of spacer acquisition were also found to be higher for transcriptionally active regions (Staals et al., 2016), so together these effects may influence the abundance and effectivity of spacers in nature.

Next to the positive effect of transcription on target search, we have also found that collisions between RNAP and target-bound Cascade lead to Cascade disassembly, where the Cas8e subunit is expelled from the Cascade core. Furthermore, CRISPR arrays themselves can trigger Cascade disassembly, indicating that they form target decoys in the cell. When present at a high copy number, CRISPR arrays can even affect CRISPR interference levels (Figure 5G). The loose association of Cas8e with the core Cascade complex as observed *in vitro* (Jore et al., 2011) may serve a biological role in cells to recycle Cascade from off-targets, including the CRISPR array, and may prevent Cas3 recruitment and subsequent self-targeting (Xiao et al., 2018).

By measuring cellular copy numbers and accurately measuring CRISPR interference levels, we could uncover an exponential relation between the number of Cascade complexes in the cell and CRISPR interference. This relation is described as every 20 Cascade complexes loaded with one crRNA can provide 50% more protection from an invading DNA element (i.e., 20 copies provide 50%, 40 copies 75% protection). Therefore, at constant Cas protein production and degradation levels, the effective concentrations of Cascade complexes loaded with one type of crRNA will become diluted when CRISPR arrays become longer. The size of the CRISPR array is therefore a trade-off between the higher protection levels of a few spacers and lower protection levels of many spacers. With our findings, we can test the optimality of this trade-off under different conditions and help explain the observed sizes of CRISPR arrays found in nature (Martynov et al., 2017).

The initial entry is the most vulnerable time for the invader, but invading MGEs have the possibility of outrunning CRISPR-Cas immunity by replicating faster than being found. In the native cellular environment, we have found that scanning of host DNA, binding to CRISPR arrays, and encountering transcribing RNA polymerases can prevent Cascade from finding the target before the critical time ($t_c$) is reached and the invader is permanently established (Figure 6). We therefore hypothesize the presence of a kinetic arms race, in which invaders have evolved to replicate increasingly fast upon cell entry, while CRISPR systems have evolved to increase the rate at which they are able to find the target. A recent study has shown that the replication rate of foreign elements affects CRISPR interference levels (Høyland-Kroghsbo et al., 2018). Many bacteriophages use a two-stage injection (Chen et al., 2018; Davison, 2015), which may have evolved to limit the amount of time that their DNA is exposed to intracellular defense mechanisms, while already allowing the production of proteins to replicate phage DNA, control host takeover, or inhibit host defense (e.g., anti-CRISPR proteins) (De Smet et al., 2017). It has been previously shown that the host can counter this strategy by selectively targeting early injected DNA regions, maximizing the time available to look for targets (Modell et al., 2017).

The specificity and kinetics of the CRISPR system inside the crowded cellular environment are remarkable. Our study has observed very rapid scanning of DNA sites by Cascade complexes, and our model predicts the impact of probing kinetics and copy numbers of Cascade on the protection levels of CRISPR-Cas systems. We believe that not only specificity and evasion strategies such as anti-CRISPRs but also target search and infection kinetics have played an important role in the evolution of this immune system. The target search equations established here could be expanded to the population level, allowing us to model how individual variability in Cascade expression levels and replication rates can affect the survival of entire populations. Therefore, our data provide an important framework for further quantitative cellular studies that will address how CRISPR systems optimally deal with the challenges of cost-effective and rapid target search.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cloning
  - Recombination
- METHOD DETAILS
  - Growth Conditions
  - Transformation Assay
  - qPCR
  - Slide Preparation
  - Microscope Set-up
  - Single-molecule Measurements
  - Analysis
  - Determination of Diffusion Coefficients
  - Analytical Diffusion Distribution Analysis (DDA)
  - Copy Number Determination
  - Cascade in DNA-containing/DNA-free Regions
  - Persistence Sustained Binding Events for Different Integrated Times
  - Confinement and Localization Error Simulation
  - Cascade Nucleoid Enrichment Simulation
  - Expected Free Diffusion Coefficients

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2019.10.021.

## ACKNOWLEDGMENTS

The authors thank Dr. A. Košmrlj (Princeton University) for deriving Equation 29 presented in Method Details. We acknowledge S. Creutzberg for supplying plasmid pSC020, M. Siliakus for supplying plasmid pMS011, and all of the members of the Hohlbein and the Brouns groups for input during group discussions. We thank Jaap Keijsers, Fiona Murphy, and Stan van de Wall for providing preliminary measurements and scripts for data analysis. S.J.J.B. is supported by the European Research Council (ERC) Stg grant 639707 and by a Vici grant of the Netherlands Organisation for Scientific Research (NWO; VI.C.182.027). R.E.M. is supported by the Frontiers of Nanoscience (NanoFront) program from NWO/Ministry of Education (OCW).

## AUTHOR CONTRIBUTIONS

S.J.J.B. and J.H. conceived and supervised the project; J.N.A.V., M.V., R.E.M., C.A., D.J.W.B., and B.E.B. performed the experimental work; J.N.A.V. and J.H. derived the theory; J.N.A.V. and K.J.A.M. wrote the analysis scripts; J.N.A.V., K.J.A.M., and J.H. established the microscopy workflow; J.N.A.V., J.H., and S.J.J.B. wrote the manuscript with input from all of the authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 25, 2019
Revised: July 31, 2019
Accepted: October 11, 2019
Published: November 14, 2019

## REFERENCES

Antonik, M., Felekyan, S., Gaiduk, A., and Seidel, C.A.M. (2006). Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. J. Phys. Chem. B *110*, 6970–6978.

Beloglazova, N., Kuznedelov, K., Flick, R., Datsenko, K.A., Brown, G., Popovic, A., Lemak, S., Semenova, E., Severinov, K., and Yakunin, A.F. (2015). CRISPR RNA binding and DNA target recognition by purified Cascade complexes from Escherichia coli. Nucleic Acids Res. *43*, 530–543.

Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Künne, T., Sobota, M., Dekker, C., Brouns, S.J.J., and Joo, C. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. Mol. Cell *58*, 60–70.

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M.F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K.L., and Davidson, A.R. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature *526*, 136–139.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Chandradoss, S.D., Haagsma, A.C., Lee, Y.K., Hwang, J.-H., Nam, J.-M., and Joo, C. (2014). Surface passivation for single-molecule protein studies. J. Vis. Exp. *86*, 4–11.

Chen, Y.J., Wu, D., Gelbart, W., Knobler, C.M., Phillips, R., and Kegel, W.K. (2018). Two-Stage Dynamics of in Vivo Bacteriophage Genome Ejection. Phys. Rev. X *8*, 021029.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. USA *97*, 6640–6645.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. *3*, 945.

Davison, J. (2015). Pre-early functions of bacteriophage T5 and its relatives. Bacteriophage *5*, e1086500.

De Smet, J., Hendrix, H., Blasdel, B.G., Danis-Wlodarczyk, K., and Lavigne, R. (2017). Pseudomonas predators: understanding and exploiting phage-host interactions. Nat. Rev. Microbiol. *15*, 517–530.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. *190*, 1390–1400.

Dillard, K.E., Brown, M.W., Johnson, N.V., Xiao, Y., Dolan, A., Hernandez, E., Dahlhauser, S.D., Kim, Y., Myler, L.R., Anslyn, E.V., et al. (2018). Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. Cell *175*, 934–946.e15.

Durisic, N., Laparra-Cuervo, L., Sandoval-Álvarez, A., Borbely, J.S., and Lakadamyali, M. (2014). Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. Nat. Methods *11*, 156–162.

Edelstein, A., Amodaj, N., Hoover, K., Vale, R., and Stuurman, N. (2010). Computer Control of Microscopes Using μManager. Curr. Protoc Mol. Biol. Unit14.20.

Elf, J., Li, G.W., and Xie, X.S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. Science *316*, 1191–1194.

English, B.P., Hauryliuk, V., Sanamrad, A., Tankov, S., Dekker, N.H., and Elf, J. (2011). Single-molecule investigations of the stringent response machinery in living bacterial cells. Proc. Natl. Acad. Sci. USA *108*, E365–E373.

Farooq, S., and Hohlbein, J. (2015). Camera-based single-molecule FRET detection with improved time resolution. Phys. Chem. Chem. Phys. *17*, 27862–27872.

Floc'h, K., Lacroix, F., Barbieri, L., Servant, P., Galland, R., Butler, C., Sibarita, J.-B., Bourgeois, D., and Timmins, J. (2018). Bacterial cell wall nanoimaging by autoblinking microscopy. Sci. Rep. *8*, 14038.

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., and Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biol. *16*, 504–517.

Globyte, V., Lee, S.H., Bae, T., Kim, J.S., and Joo, C. (2019). CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. EMBO J. *38*, 99466.

Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B.G., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. Nature *530*, 499–503.

Ho, H.N., van Oijen, A.M., and Ghodke, H. (2018). The transcription-repair coupling factor Mfd associates with RNA polymerase in the absence of exogenous damage. Nat. Commun. *9*, 1570.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc. Natl. Acad. Sci. USA *111*, 6618–6623.

Holden, S.J., Uphoff, S., Hohlbein, J., Yadin, D., Le Reste, L., Britton, O.J., and Kapanidis, A.N. (2010). Defining the limits of single-molecule FRET resolution in TIRF microscopy. Biophys. J. *99*, 3102–3111.

Hoogendoorn, E., Crosby, K.C., Leyton-Puig, D., Breedijk, R.M.P., Jalink, K., Gadella, T.W.J., and Postma, M. (2014). The fidelity of stochastic single-molecule super-resolution reconstructions critically depends upon robust background estimation. Sci. Rep. *4*, 3854.

Høyland-Kroghsbo, N.M., Muñoz, K.A., and Bassler, B.L. (2018). Temperature, by Controlling Growth Rate, Regulates CRISPR-Cas Activity in Pseudomonas aeruginosa. MBio. https://doi.org/10.1128/mBio.02184-18.

Huang, B., Wu, H., Bhaya, D., Grossman, A., Granier, S., Kobilka, B.K., and Zare, R.N. (2007). Counting low-copy number proteins in a single cell. Science *315*, 81–84.

Huang, F., Hartwich, T.M.P., Rivera-Molina, F.E., Lin, Y., Duim, W.C., Long, J.J., Uchil, P.D., Myers, J.R., Baird, M.A., Mothes, W., et al. (2013). Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. Nat. Methods *10*, 653–658.

Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C., and Brouns, S.J.J. (2017). CRISPR-Cas: adapting to change. Science *356*, eaal5056.

Jones, D.L., Leroy, P., Unoson, C., Fange, D., Ćurić, V., Lawson, M.J., and Elf, J. (2017). Kinetics of dCas9 target search in *Escherichia coli*. Science *357*, 1420–1424.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. *18*, 529–536.

Jung, C., Hawkins, J.A., Jones, S.K., Jr., Xiao, Y., Rybarski, J.R., Dillard, K.E., Hussmann, J., Saifuddin, F.A., Savran, C.A., Ellington, A.D., et al. (2017). Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. Cell *170*, 35–47.e13.

Kalinin, S., Felekyan, S., Valeri, A., and Seidel, C.A.M. (2008). Characterizing multiple molecular States in single-molecule multiparameter fluorescence detection by probability distribution analysis. J. Phys. Chem. B *112*, 8361–8374.

Knight, S.C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L.B., Bosanac, L., Zhang, E.T., El Beheiry, M., Masson, J.-B., Dahan, M., et al. (2015). Dynamics of CRISPR-Cas9 genome interrogation in living cells. Science *350*, 823–826.

Kumar, M., Mommer, M.S., and Sourjik, V. (2010). Mobility of cytoplasmic, membrane, and DNA-binding proteins in Escherichia coli. Biophys. J. *98*, 552–559.

Lee, S.-H., Shin, J.Y., Lee, A., and Bustamante, C. (2012). Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). Proc. Natl. Acad. Sci. USA *109*, 17436–17441.

Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol. Cell *62*, 137–147.

Ma, J., and Wang, M.D. (2016). DNA supercoiling during transcription. Biophys. Rev. *8* (*Suppl 1*), 75–87.

Majsec, K., Bolt, E.L., and Ivančić-Baće, I. (2016). Cas3 is a limiting factor for CRISPR-Cas immunity in Escherichia coli cells lacking H-NS. BMC Microbiol. *16*, 28.

Manley, S., Gillette, J.M., Patterson, G.H., Shroff, H., Hess, H.F., Betzig, E., and Lippincott-Schwartz, J. (2008). High-density mapping of single-molecule trajectories with photoactivated localization microscopy. Nat. Methods *5*, 155–157.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55–61.

Martens, K.J.A., van Beljouw, S.P.B., van der Els, S., Vink, J.N.A., Baas, S., Vogelaar, G.A., Brouns, S.J.J., van Baarlen, P., Kleerebezem, M., and Hohlbein, J. (2019). Visualisation of dCas9 target search in vivo using an open-microscopy framework. Nat. Commun. *10*, 3552.

Martynov, A., Severinov, K., and Ispolatov, I. (2017). Optimal number of spacers in CRISPR arrays. PLoS Comput. Biol. *13*, e1005891.

McGuigan, J.A.S., Kay, J.W., and Elder, H.Y. (2006). Critical review of the methods used to measure the apparent dissociation constant and ligand purity in Ca2+ and Mg2+ buffer solutions. Prog. Biophys. Mol. Biol. *92*, 333–370.

Michalet, X. (2010). Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. Phys. Rev. E - Stat. Nonlinear. Soft Matter Physiol. *82*, 41914.

Mika, J.T., and Poolman, B. (2011). Macromolecule diffusion and confinement in prokaryotic cells. Curr. Opin. Biotechnol. *22*, 117–126.

Mika, J.T., van den Bogaart, G., Veenhoff, L., Krasnikov, V., and Poolman, B. (2010). Molecular sieving properties of the cytoplasm of Escherichia coli and consequences of osmotic stress. Mol. Microbiol. *77*, 200–207.

Modell, J.W., Jiang, W., and Marraffini, L.A. (2017). CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature *544*, 101–104.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733–740.

Mondal, J., Bratton, B.P., Li, Y., Yethiraj, A., and Weisshaar, J.C. (2011). Entropy-based mechanism of ribosome-nucleoid segregation in E. coli cells. Biophys. J. *100*, 2605–2613.

Nenninger, A., Mastroianni, G., and Mullineaux, C.W. (2010). Size dependence of protein diffusion in the cytoplasm of Escherichia coli. J. Bacteriol. *192*, 4535–4540.

Olsson, J.A., Berg, O.G., Dasgupta, S., and Nordström, K. (2003). Eclipse period during replication of plasmid R1: contributions from structural events and from the copy-number control system. Mol. Microbiol. *50*, 291–301.

Paintdakhi, A., Parry, B., Campos, M., Irnov, I., Elf, J., Surovtsev, I., and Jacobs-Wagner, C. (2016). Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. Mol. Microbiol. *99*, 767–777.

Palo, K., Mets, U., Loorits, V., and Kask, P. (2006). Calculation of photon-count number distributions via master equations. Biophys. J. *90*, 2179–2191.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H.W., Maxwell, K.L., and Davidson, A.R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. MBio *5*, e00896.

Peterson, J., and Phillips, G.J. (2008). New pSC101-derivative cloning vectors with elevated copy numbers. Plasmid *59*, 193–201.

Qian, H., Sheetz, M.P., and Elson, E.L. (1991). Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. Biophys. J. *60*, 910–921.

Redding, S., Sternberg, S.H.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C.C. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. Cell *163*, 854–865.

Reyes-Lamothe, R., Tran, T., Meas, D., Lee, L., Li, A.M., Sherratt, D.J., and Tolmasky, M.E. (2014). High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division. Nucleic Acids Res. *42*, 1042–1051.

Sanamrad, A., Persson, F., Lundius, E.G., Fange, D., Gynnå, A.H., and Elf, J. (2014). Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. Proc. Natl. Acad. Sci. USA *111*, 11413–11418.

Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol. Cell *46*, 606–615.

Severinov, K., Ispolatov, I., and Semenova, E. (2016). The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense. Front. Mol. Biosci. *3*, 45.

Shao, Q., Hawkins, A., and Zeng, L. (2015). Phage DNA dynamics in cells with different fates. Biophys. J. *108*, 2048–2060.

Slutsky, M., and Mirny, L.A. (2004). Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential. Biophys. J. *87*, 4021–4035.

Staals, R.H.J., Jackson, S.A., Biswas, A., Brouns, S.J.J., Brown, C.M., and Fineran, P.C. (2016). Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. Nat. Commun. *7*, 12853.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62–67.

Stracy, M., Lesterlin, C., Garza de Leon, F., Uphoff, S., Zawadzki, P., and Kapanidis, A.N. (2015). Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. Proc. Natl. Acad. Sci. USA *112*, E4390–E4399.

Subach, F.V., Patterson, G.H., Manley, S., Gillette, J.M., Lippincott-Schwartz, J., and Verkhusha, V.V. (2009). Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. Nat. Methods *6*, 153–159.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc. Natl. Acad. Sci. USA *111*, 9798–9803.

Uphoff, S., Reyes-Lamothe, R., Garza de Leon, F., Sherratt, D.J., and Kapanidis, A.N. (2013). Single-molecule DNA repair in live bacteria. Proc. Natl. Acad. Sci. USA *110*, 8063–8068.

van Erp, P.B.G., Jackson, R.N., Carter, J., Golden, S.M., Bailey, S., and Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. Nucleic Acids Res. *43*, 8381–8391.

Van Vliet, L., Sudar, D., and Young, I. (1998). Digital fluorescence imaging using cooled charge-coupled device array cameras. In Cell Biology, 2nd ed., *Vol. III*, J.C. Celis, ed. (Academic Press), pp. 109–120.

Vigouroux, A., Oldewurtel, E., Cui, L., Bikard, D., and van Teeffelen, S. (2018). Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes. Mol. Syst. Biol. *14*, e7899.

Vrljic, M., Nishimura, S.Y., Brasselet, S., Moerner, W.E., and McConnell, H.M. (2002). Translational diffusion of individual class II MHC membrane proteins in cells. Biophys. J. *83*, 2681–2692.

Wallden, M., Fange, D., Lundius, E.G., Baltekin, Ö., and Elf, J. (2016). The Synchronization of Replication and Division Cycles in Individual E. coli Cells. Cell *166*, 729–739.

Westra, E.R., Pul, U., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., et al. (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol. Microbiol. *77*, 1380–1393.

Westra, E.R., van Erp, P.B.G., Künne, T., Wong, S.P., Staals, R.H.J., Seegers, C.L.C., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol. Cell *46*, 595–605.

Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. Cell *170*, 48–60.e11.

Xiao, Y., Luo, M., Dolan, A.E., Liao, M., and Ke, A. (2018). Structure basis for RNA-guided DNA degradation by Cascade and Cas3. Science *361*, eaat0839.

Xue, C., Whitis, N.R., and Sashital, D.G. (2016). Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. Mol. Cell *64*, 826–834.

Xue, C., Zhu, Y., Zhang, X., Shin, Y.K., and Sashital, D.G. (2017). Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. Cell Rep. *21*, 3717–3727.

Zawadzki, P., Stracy, M., Ginda, K., Zawadzka, K., Lesterlin, C., Kapanidis, A.N., and Sherratt, D.J. (2015). The Localization and Action of Topoisomerase IV in Escherichia coli Chromosome Segregation Is Coordinated by the SMC Complex, MukBEF. Cell Rep. *13*, 2587–2596.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Bacterial and Virus Strains | | |
| *E. coli* K12 BW25113 | (Datsenko and Wanner, 2000) | CGSC#7636 |
| *E. coli* K12 BW25113 PAmCherry2-Cas8e | This study | N/A |
| *E. coli* K12 BW25113 PAmCherry2-Cas8e (Δ(*cas11-cas6e*)) | This study | N/A |
| *E. coli* K12 BW25113 (*Δcas3*) | This study | N/A |
| *E. coli* K12 BW25113 (*Δcas3*, *cas8e*(G160A)) | This study | N/A |
| *E. coli* K12 BW25113 (*Δcas3*, ΔCRISPR) | This study | N/A |
| Chemicals, Peptides, and Recombinant Proteins | | |
| IPTG | Gold Biotechnology | Cat#C0378 |
| L-Arabinose | Sigma Aldrich | Cat#A3256 |
| Cephalexin | Sigma Aldrich | Cat#C4895 |
| M9 salts 5x | Sigma Aldrich | Cat#M6030 |
| Glucose | Sigma Aldrich | Cat#G8270 |
| EZ amino acids 5x | Teknova | Cat#M2104 |
| Uracil | Sigma Aldrich | Cat#U0750 |
| $MgSO_4$ | Sigma Aldrich | Cat#230391 |
| $CaCl_2$ | Sigma Aldrich | Cat#223506 |
| DAPI | Sigma Aldrich | Cat#D9542 |
| Kanamycin | Roth | Cat#T832.2 |
| Chloramphenicol | Sigma Aldrich | Cat#C0378 |
| Ampicillin sodium salt | Roth | Cat#K029.4 |
| Streptomycin sulfate salt | Sigma Aldrich | Cat#S6501 |
| Acetone | Sigma Aldrich | Cat#179124 |
| KOH | VWR | Cat#06015 |
| $H_2SO_4$ 95-98% | Sigma Aldrich | Cat#258105 |
| $H_2O_2$ 30% solution | Sigma Aldrich | Cat#H1009 |
| Agarose | Bio-rad | Cat#1613100 |
| Glycerol | Acros organics | Cat#158920010 |
| Critical Commercial Assays | | |
| Genejet Genomic DNA kit | Thermo Fischer Scientific | Cat#K0721 |
| Qubit dsDNA HS Assay kit | Thermo Fischer Scientific | Cat#Q33211 |
| iTaq™ SYBR Green PCR reaction mixture | Bio-rad | Cat#1725120 |
| Deposited Data | | |
| Raw and analyzed data | This paper | https://doi.org/10.17632/w83h66729n.1 |
| Experimental Models: Organisms/Strains | | |
| *E. coli* K12 BW25113 | (Datsenko and Wanner, 2000) | CGSC#7636 |
| Oligonucleotides | | |
| See Table S2 | This study | N/A |
| Recombinant DNA | | |
| See Tables S3 and S4 | This study | N/A |
| Software and Algorithms | | |
| MATLAB | The MathWorks Inc | https://fr.mathworks.com/products/matlab.html |
| Micromanager | (Edelstein et al., 2010) | https://micro-manager.org/ |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| MLE-sCMOS | (Huang et al., 2013) | https://media.nature.com/original/nature-assets/nmeth/journal/v10/n7/extref/nmeth.2488-S3.zip |
| Oufti | (Paintdakhi et al., 2016) | oufti.org |
| Dynamic D* Fitting | This study | Available on request |

## LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled without restrictions by the Lead Contact, Stan Brouns (stanbrouns@gmail.com).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Cloning

The inserts to create pTarget and pCRISPR1 plasmids were purchased as synthetic constructs from Gen9 (pTarget insert and pCRISPR1 insert; Table S3). To increase the copy number of targets in the cell, the constructs were cloned into a pUC19 backbone with XbaI and KpnI restriction sites, yielding pTarget-RNAP and pCRISPR1-RNAP. The lac promoter was removed for both plasmids by digestion with SalI and PciI, creating blunt ends with Klenow Fragment and subsequently religated to yield pTarget and pCRISPR1. CRISPR arrays were amplified from the K12 BW25113 strain (primers BN383 and BN384; BN370 and BN385 for CRISPR array 2.1 and 2.3 respectively) and cloned into pJPC-12 plasmid containing the pSC101 ori with KpnI and SalI sites (for CRISPR array 2.1) and SalI and EcoRV sites (for CRISPR array 2.3). The copy number of the plasmid could be varied by introducing mutations in the *repA* gene with site-directed mutagenesis PCR (BN373-375). The E96R mutation of RepA yields a reported copy number of ~240/cell (pCRISPR2) compared to the WT RepA (pCRISPR2_LC) copy numbers of ~7/cell (Peterson and Phillips, 2008). A plasmid was made from the high copy-variant that did not contain any CRISPR arrays (pControl). All constructs were verified by sequencing.

### Recombination

The strains used in this study were created by using Lambda red recombineering (Datsenko and Wanner, 2000). Strains harboring the pSC020 plasmid that contains both the Lambda red recombinase and Cre-recombinase were grown at 30°C. Before transformation of an insert containing an antibiotic resistance marker, the expression of Red recombinase was induced with 0.2% L-Arabinose. Colonies on the specific antibiotic plate were verified with PCR and sequencing and subsequently Cre recombinase expression was induced with 1 mM IPTG at 37°C to promote plasmid and antibiotic resistance gene loss. The strain was subsequently patch plated to screen for resistance sensitivity due to plasmid loss.

If the scar that is left after lox-site recombination is directly upstream or downstream of a gene it might influence gene transcription/termination. In the design of constructs for *pamcherry2* (Subach et al., 2009) the lox-*cat*-lox sequence was placed upstream of the IGR (Intergenic region) that is present between *cas3* and *cas8e*. To allow for correct termination of *cas3,* a part of the IGR was also added at the 5′ end of the antibiotic resistance marker. The 3′ flank of the constructs overlapped with the *cas8e* gene. The 5′ flank of the constructs matched a sequence upstream and downstream of *cas3* (PAmCherry2 ins; Table S3). Amplification of the constructs with a forward primer matching the downstream region kept *cas3* intact upon insertion (BG7128), whereas a primer matching the upstream region deleted the *cas3* gene allowing measurements in the presence of targets (BG7129). The insert also contained a part of the *cas8e* sequence containing a G160A mutation. This mutation could be introduced into the gene simultaneously with the fluorescent protein, depending on the reverse primer that was used for insert amplification (BG7130 for WT, BG7131 for G160A).

Knockouts of the CRISPR arrays and Cas gene subunits of the K12 strain were made by amplifying a lox-*kan*-lox or lox-*cat*-lox sequence with flanks matching the specific sequences and introducing them into the strain as described above (BG7366+BG7367 for CRISPR array 2.1; BG7368+BG7369 for CRISPR array 2.2+2.3; BG8366+BG8367 for Δ(*cas11-cas6e*)). A full overview of the sequences of these inserts is given in Table S3.

## METHOD DETAILS

### Growth Conditions

To prevent the high-copy target plasmids from influencing the growth rate of the strains and therefore changing the fraction of matured PAmCherry2 complexes we used a rich defined medium with minimal autofluorescence. Strains were grown in M9 minimal medium containing the following supplements: 0.4% glucose, 1x EZ amino acids supplements (M2104 Teknova), 20 μg/ml uracil (Sigma-Aldrich), 1mM $MgSO_4$ (Sigma-Aldrich) and 0.1 mM $CaCl_2$ (Sigma-Aldrich) (further referred to as M9 medium). Strains were inoculated o/n from glycerol stocks and 200x diluted in fresh medium the next day. Cells were always grown with the required antibiotics. The expression level of Cascade for strains carrying the pKEDR13 plasmid could be tuned by different

expression levels of LeuO. The expression level referred to in the text as low induction was achieved by leaky expression of LeuO (no addition of IPTG), medium induction was achieved by addition of low levels of IPTG (0.01 mM), whereas high induction was achieved by addition of 1 mM IPTG upon dilution of the o/n culture. For all sptPALM measurements the high induction condition was used. The cells were grown for ~2.5 hours to an OD of 0.1 before use. For enforced elongation of cells, cephalexin (40 μg/ml) was added 0.5 hour after fresh inoculation and grown for two more hours. When required, DAPI for staining of DNA was added right before imaging (0.5 mg/ml).

### Transformation Assay

Each culture was grown under conditions described above and 30 mL were used to create competent cells. Cells were washed 3 times in ice-cold 10% glycerol solution and the final culture was reduced to 250 μl. The cells were aliquoted and stored at −80°C. A mixture of pTarget (10 pg/μl) and pGFPuv (10pg/μl) was transformed into 40 μl of culture. In case of strong interference levels, the ratio was adjusted to a 100:1 (pTarget (100 pg/μl):pGFPuv (1 pg/μl)). The transformability of strains was linear in these concentration regimes, allowing these different relative concentrations to be used.

Electroporated cells were immediately plated in two dilutions on plates containing ampicillin (100 μg/ml) and glucose (0.4%). Glucose was added to prevent premature expression of GFPuv which would cause a decrease in fitness of cells containing this plasmid. The next day, 96 colonies from each replicate were reinoculated in 96-wells plate with LB containing ampicillin (100 μg/ml) and IPTG (1 mM). After overnight incubation, the 96 well colonies were analyzed in a plate-reader (Synergy H1, Biotek). pTarget establishment was defined as

$$p_{\text{establishment}} = \frac{\#pTarget\ colonies}{\#GFPuv\ colonies}\frac{[pGFPuv\ Transformed]}{[pTarget\ Transformed]} \quad (1)$$

pTarget establishment was further normalized to the interference level of a Δ*cas3* strain.

### qPCR

Each culture grew under conditions described above and 2 mL were used to extract the DNA. DNA was isolated with the Genejet Genomic DNA kit (Thermo Scientific) and concentrations were measured with the Qubit dsDNA HS Assay kit (Thermo Scientific). qPCR was performed with primers that have been used before in plasmid copy determination (BG8677-BG8680) (Reyes-Lamothe et al., 2014). The Ct value of the PCR amplifying the *dxs* gene and the *bla* gene was a measure for the ratio between chromosomal and plasmid DNA. 1 ng of genomic DNA and 0.5 μM of each primer were added to the iTaq™ SYBR Green SYBR Green PCR reaction mixture. A standard curve for the amplification efficiency was made by a dilution series of pMS011, a plasmid containing one copy of the *dxs* and the *bla* gene.

### Slide Preparation

In order to work with very clean slides, an extensive cleaning procedure was used (modified from (Chandradoss et al., 2014)). Slides were burned in the oven at 500°C for two hours, and stored in aluminum foil until the day of usage. Slides were subsequently sonicated in MilliQ, Acetone and KOH, incubated in Piranha Solution (75% $H_2SO_4$, 7.5% $H_2O_2$) and afterward rinsed with MilliQ. 1% Agarose slabs containing the growth medium were hardened between two cleaned glass slides, spaced slightly apart using parafilm. After hardening, a concentrated culture of cells was added in between the slab and one of the slides. The agarose slab was always prepared within 20 minutes of the measurement to prevent desiccation.

### Microscope Set-up

For the acquisition of microscopy data, a home-build TIRF microscope was used, which is described in more detail elsewhere (Martens et al., 2019). Briefly, four lasers with different wavelengths (405, 473, 561 and 642 nm) are situated in a Lighthub laser box (Omicron, Germany), and are transformed in a collimated beam via a reflective collimator and an optical fiber. Stroboscopic illumination was used to allow for 2 ms excitation in the temporal middle of the captured 10 ms long frame (Farooq and Hohlbein, 2015). The excitation laser is focused on the backfocal plane of a 100x oil immersion SR/HP objective (NA = 1.49, Nikon, Japan), and the emission is captured on a Zyla 4.2 plus sCMOS camera (Andor, UK). 2x2 pixel binning was used, resulting in 128x128 nm pixels. Data acquisition was performed using MicroManager (Edelstein et al., 2010). Measurements were performed at room temperature (21°C)

### Single-molecule Measurements

The cells were imaged with a brightfield light and 405 and 561 nm lasers. First brightfield images were taken to find contours of the cells. The 405 nm laser was used to stochastically activate PAmCherry2 and the laser intensity was slowly increased during the measurement up to 10 μW. The laser intensities were measured directly after the reflective collimator. With increasing the laser intensity of the 405 nm laser during the measurements, we aimed at keeping the number of activated molecules relatively constant (~1-10 per FOV). The 561 nm laser was used to excite the fluorescent protein tags (40 mW pulses with 2ms pulse width, leading to average exposure intensity of 8 mW).

To measure Cascade localization in cephalexin-treated cells that were stained with DAPI, we took an alternative approach. To prevent DAPI fluorescence from influencing the fluorescence measurements of the single molecules, we briefly activated a subset of

particles with the 405 nm laser and subsequently tracked Cascade for a couple of frames with 561 nm excitation, repeatedly doing this, until most fluorescent proteins were photobleached.

## Analysis

### *Detection, localization and tracking*

Analysis was done with home-built software, adapted from (Holden et al., 2010; Uphoff et al., 2013). The sCMOS camera we used has pixel dependent offset, gain, and variance, which we took into account to minimize the detection of false positive localizations. We estimated these parameters by measuring 60.000 dark frames and 20.000 homogeneously illuminated frames with increasing levels of intensity (Van Vliet et al., 1998). To further optimize our detection, we implemented a temporal median filter (time window 400 frames) for background estimation (Hoogendoorn et al., 2014). The background estimate was not directly subtracted from the image, but photon statistics were incorporated in a likelihood-ratio test that calculated the probability of a scenario with and without an emitter for each pixel in every frame. Briefly, a raw image was first converted into photon counts by using the camera offset and gain maps. Subsequently for every pixel the intensity ($I_{tot}$) of a potential emitter was estimated by Gaussian-weighted ($\sigma$ = 1 pixel) summation of a 7x7 window to a background subtracted image. Subsequently, potential emitters of more than 50 photons were pre-selected and were further subjected to a ratio test. The ratio test uses the probability defined for pixel *i* to have a transformed value *v* in the 7x7 region around the preselected pixels as previously described (Huang et al., 2013):

$$p_{\mathrm{scmos}}\left(v = \left[(d_i - o_i)/g_i + var_i/g_i^2\right] | \mu_i, var_i, g_i, o_i\right) = \frac{e^{-\left(\mu_i + var_i/g_i^2\right)}\left(\mu_i + var_i/g_i^2\right)^v}{\Gamma(v+1)} \quad (2)$$

Where $d_i$ is the raw image value, $g_i$ is the gain, $var_i$ the variance and $o_i$ the offset for pixel *i*. The ratio test calculates the product of the probability of all pixels in the subregion in case of an emitter $\mu_i = b_i + I_i$, where $b_i$ is the estimated background an $I_i$ is the estimated intensity of the emitter at pixel *i* (which was estimated by a Gaussian from the center of the 7x7 subregion with emitter intensity $I_{tot}$) divided by the product of the probability of all pixels in the subregion in case of absence of an emitter $\mu_i = b_i$.

We set the likelihood to a level that achieved approximately one false positive per frame of 512 × 512 pixels. This method allowed the detection efficiency to be more robust across and between FOVs and independent of manual thresholding for each measurement. Detected particles were subsequently localized with MLE-sCMOS software as previously described (Huang et al., 2013).

The localized particles were subsequently linked. Localizations in subsequent frames which were closer to each other than 6 pixels in length (0.78 μm) were assigned as a track. Particles were allowed to disappear for one frame (due to blinking/moving out of focus), but these steps were not used in the calculation of the apparent diffusion coefficient, $D^*$.

## Determination of Diffusion Coefficients

Several methods were employed to extract diffusion states and their abundances from the analyzed tracks. The distribution of the apparent diffusion coefficients can be fitted to an analytical equation as reported earlier (Stracy et al., 2015; Vrljic et al., 2002). These equations depend on the number of steps that is used to generate the average diffusion coefficients of each particle. We used tracks containing a minimum of four steps and only four steps were used in longer tracks.

For a single diffusion coefficient fitting becomes:

$$f_D(x; D, n) = \frac{\left(\frac{n}{D + \sigma^2/\mathrm{dt}}\right)^n x^{n-1} e^{-\frac{nx}{D + \sigma^2/\mathrm{dt}}}}{(n-1)!} \quad (3)$$

With multiple states this equation becomes:

$$f_D(x; A_i, D_i, n) = \sum_{i=1}^{N} A_i \frac{\left(\frac{n}{D_i + \sigma^2/\mathrm{dt}}\right)^n x^{n-1} e^{-\frac{nx}{D_i + \sigma^2/\mathrm{dt}}}}{(n-1)!} \quad (4)$$

Where $A_i$ are the fractions ($\sum A_i = 1$), $D_i^*$ are the apparent diffusion coefficients of the different states and *n* are the number of steps. The localization error ($\sigma$) was found to be 40 nm, based on the apparent diffusion of the slowest moving fraction in our global dataset and similar to other studies using the same fluorescent protein (Stracy et al., 2015; Uphoff et al., 2013) or set-up (Martens et al., 2019). This equation was fitted to our track distributions with a Maximum Likelihood Estimation algorithm. The uncertainty in the fit was estimated with Bootstrap resampling. The list of $D^*$ values was resampled 20.000 times with replacement to the size of the original dataset. Each resample was then fitted with the same Maximum Likelihood Estimation algorithm.

## Analytical Diffusion Distribution Analysis (DDA)

$D^*$ Distributions have been fitted in numerous studies of DNA binding proteins (see above) (Stracy et al., 2015; Vrljic et al., 2002), making use of distributions developed by Qian et al. (Qian et al., 1991). The goal is to find the distribution of measured $D^*$ values ($x$), for a

certain number of underlying states that each have a probability $A_i$ and a diffusion coefficient $D_i$. It is derived from repeated convolution of the exponential distribution of displacement, resulting in a gamma function for each state. These distributions assume, however, that there is no transitioning occurring between states.

In order to incorporate dynamics of state transitions into our fitting, we incorporated statistics coming from photon distribution analysis (PDA) that is used for single molecule FRET diffusion coefficient distributions (Antonik et al., 2006; Kalinin et al., 2008; Palo et al., 2006). This method, that we term Diffusion Distribution Analysis (DDA), describes the distribution of time spent in each state given a certain $k^*_{on}$, $k_{off}$and the integrated time $t_{int}$. Here we discuss the analytical way to find this distribution.

First, the probability distribution function for time can be calculated by three equations corresponding to 0, an odd and an even number of transitions (Palo et al., 2006):

$$W_{contS1}(t_{S1} = t_{int} \mid k_{off}, t_{int}) = e^{-k_{off}t_{int}} \tag{5}$$

$$W_{oddS1}(t_{S1} \mid k_{off}, k^*_{on}, t_{int}) = k_{off}e^{-k_{off}t_{S1}-k^*_{on}t_{S2}}I_0\left(2\sqrt{k_{off}k^*_{on}t_{S1}t_{S2}}\right) \tag{6}$$

$$W_{evenS1}(t_{S1} \mid |k_{off}, k^*_{on}, t_{int}) = \sqrt{k_{off}k^*_{on}t_{S1}/t_{S2}}e^{-k_{off}t_{S1}-k^*_{on}t_{S2}}I_1\left(2\sqrt{k_{off}k^*_{on}t_{S1}t_{S2}}\right) \tag{7}$$

Where $t_{S1}$ and $t_{S2}$are times spent in state *S1* and state *S2* and $I_0$ and $I_1$ are Bessel functions of order zero and one respectively. Note that $t_{S1} + t_{S2} = t_{int}$. Equations for starting in state 2 ($W_{contS2}$, $W_{oddS2}$ and $W_{evenS2}$), can be found by exchanging $k_{off}$ for $k^*_{on}$ and $t_{S1}$ for $t_{S2}$ and vice versa in Equations 5, 6, and 7.

We can convert the time spent in the mobile state $(t_{S2})$ to the diffusion coefficient by the following equation:

$$D = \frac{D_{free}t_{S2}}{t_{int}} \tag{8}$$

It follows that the probability distribution functions can be converted by:

$$W(D) = W\left(t_{S2} = \frac{Dt_{int}}{D_{free}}\right) \tag{9}$$

Furthermore, the chance that the particle at the start is in state 1 or state 2 is provided by:

$$p_{S1} = \frac{k^*_{on}}{k^*_{on} + k_{off}} \tag{10}$$

$$p_{S2} = \frac{k_{off}}{k^*_{on} + k_{off}} \tag{11}$$

To correctly describe the distribution over a certain number of frames, we first calculated the distribution over a single time frame $t_f$. Within a single frame, a particle started in that state can either end in the same state or in a different state. Therefore, in a two-state system the probability function for four scenarios have to be calculated:

$$W(D \mid k_{off}, k^*_{on}, t_f)_{S1\to S1} = W_{evenS1}(D) + W_{contS1} \tag{12}$$

$$W(D \mid k_{off}, k^*_{on}, t_f)_{S1\to S2} = W_{oddS1}(D) \tag{13}$$

$$W(D \mid k_{off}, k^*_{on}, t_f)_{S2\to S1} = W_{oddS2}(D) \tag{14}$$

$$W(D \mid k_{off}, k^*_{on}, t_f)_{S2\to S2} = W_{evenS2}(D) + W_{contS2} \tag{15}$$

Subsequently the probability to find a certain diffusion coefficient ($x$) for a single time step given the underlying average diffusion coefficient ($D$) is given by $f_D(x \mid D, 1)$ (Equation 3). Then we find the distribution of measured diffusion coefficients for a single frame by:

$$W(x|k_{off}, k^*_{on}, t_f)_{Si\to Sj} = f_D(x \mid D, 1)W(D \mid k_{off}, k^*_{on}, t_f)_{Si\to Sj} \tag{16}$$

i = j = 1,2

Now that we have the distribution for a single time step, we need to find the distribution for the average of multiple frames. For this we use the same method as Qian et al. (Qian et al., 1991), namely repeated convolution of the distribution for a single frame, while keeping track of the start and end state. The probability distributions are therefore:

$$W(x \mid 2t_f)_{S1\rightarrow S1} = \sum_{i=1,2} \left(W(x \mid t_f)_{S1\rightarrow Si} * W(x \mid t_f)_{Si\rightarrow S1}\right) \tag{17}$$

$$W(x \mid 2t_f)_{S1\rightarrow S2} = \sum_{i=1,2} \left(W(x \mid t_f)_{S1\rightarrow Si} * W(x \mid t_f)_{Si\rightarrow S2}\right) \tag{18}$$

$$W(x \mid 2t_f)_{S2\rightarrow S1} = \sum_{i=1,2} \left(W(x \mid t_f)_{S2\rightarrow Si} * W(x \mid t_f)_{Si\rightarrow S1}\right) \tag{19}$$

$$W(x \mid 2t_f)_{S2\rightarrow S2} = \sum_{i=1,2} \left(W(x \mid t_f)_{S2\rightarrow Si} * W(x \mid t_f)_{Si\rightarrow S2}\right) \tag{20}$$

For 4 frames, the distributions found for 2 frames can be convoluted again. The full distribution is then found by summing up each of the partial distributions multiplied by the chance they start in *S1* or *S2*:

$$W_{tot} = p_{S1}\left(W(x \mid 4t_f)_{S1\rightarrow S2} + W(x|4t_f)_{S1\rightarrow S1}\right) \tag{21}$$

We then have to further correct for the broadening of the distribution of immobile particles where the apparent step size comes from localization error (Figure S2). As localization error, in contrast to diffusion, is correlated (Michalet, 2010), the distribution is not described by a gamma distribution, or any other known exact solution. We find very close agreement with simulations when we subtract the fraction of immobile particles after four time steps ($W_{contS1}(t_{S1} = 4t_f)$, Equation 5) multiplied with the distribution of expected $D^*$ for four time steps $f_D(x \mid 0, 4)$ (Equation 3) and replace it with the same fraction of immobilized particles multiplied with the distribution of expected $D^*$ for 2.9 time steps $f_D(x \mid 0, 2.9)$. This value stems from the variance found for correlated MSD values due to localization error (Michalet, 2010).

The assumptions that underlie this model are as follows:

- Each diffusing species can be in two states, namely an immobile and a mobile state.
- The immobile state in our case includes all species bound to chromosomal DNA, including potential 1D sliding events, for which the diffusion is at such a low relative speed that we can consider them as immobile. Our model therefore cannot distinguish between bound and 1D sliding species.
- The immobile state is still perceived as diffusing due to a localization error, $\sigma$, which in our case is 40 nm. As the distribution of sequential localization errors differs from sequential diffusion steps we correct for this (Figure S2).
- The mobile state is defined by the parameter $D_{free}$, which is the diffusion coefficient of a species in the absence of interactions with DNA. All slowing down in the motion because of transient DNA interaction are captured in our model by the introduction of transitions and do not affect the value found for this parameter.
- The transition between the two states for each species is Markovian, meaning that transition rates are independent of past or future states.

For each species that you fit there are four degrees of freedom, namely the abundance of the species in the total population and the three kinetic parameters $k^*_{on}$, $k_{off}$, and $D_{free}$. However, because the sum of all fractions of species is one ($\sum c = 1$) and the sum of the average time spent freely diffusing multiplied by the free diffusion coefficient for each species is equal to the average measured Diffusion Coefficient $\left(\langle D\rangle = \sum c_i \frac{k_{off,i}}{k_{off,i} + k^*_{on,i}} D_{free,i}\right)$, the amount of free fitting parameters is reduced by two. This means that for a single diffusing species (in our case monomeric Cas8e) we only need to fit two parameters and for a two-species distribution (in our case Cascade) for which one is already known (Cas8e) we need to fit three parameters (8 degrees of freedom – 3 already known Cas8e kinetic parameters – 2 from the above described equations). We found that the uncertainty of our fit, determined by bootstrapping and simulations, is reasonable up to three fitting parameters, therefore we designed our experiments in a way, that in the presence of multiple species (such as pTarget (Figure 5)) we already predetermined the kinetic parameters for most species to limit the required fitting parameters to three.

### Copy Number Determination

The copy number of the Cascade complex was determined by generating cell outlines from brightfield images (only well separated cells were chosen). The cell outlines were made with the Oufti software (Paintdakhi et al., 2016). The total number of tracks that were found in the outlined cells generated a copy number (Figure 1D). Because single localization events can partly stem from false positives, the total amount of tracks was estimated based on the distribution of tracks longer than 1 step and subsequently this distribution was fitted with an exponential to calculate the amount of particles that only had a single localization before bleaching. Similarly, as we know the false positive rate was approximately one per frame, we could also subtract the number of frames from the single step tracks and in this way estimate the total number of tracks. This approach yielded comparable results.

The copy number of proteins in cells are hard to quantify (Lee et al., 2012). Currently, protein copy numbers can be estimated either by western blot or by single-molecule fluorescence based methods both of which have specific drawbacks. Although single molecule studies are regarded as the most accurate method, especially at low copy numbers (Huang et al., 2007), there are a lot of variables that can lead to over- or underestimation. Underestimation can originate from maturation time of the protein, misfolded/inactivated protein, false negative detections, overlap of PSFs and linking of two separate molecules in a single track. Overestimation can come from failed linking of tracks, false positive detections and blinking fluorescent proteins.

As has been done in previous studies, we take the underestimations stemming from maturation time (23 min for PAmCherry2 (Subach et al., 2009)), close to growth rate of 31 min) and estimated *in vivo* folding efficiency (50% (Durisic et al., 2014)) into account (Uphoff et al., 2013). We also consider that an estimated 40% of the particles we observed come from Cas8e subunits not active complexes. Taken together, the number of particles we observe are subtracted by the amount of estimated autofluorescent particles and subsequently multiplied by a compensation factor of two to reach our estimated copy number values.

We believe that the assumptions made in this study could maximally lead to over- or underestimating our estimated copy numbers by two to three-fold. We note that the relative amounts we observed between the different expression levels will be independent of these assumptions.

### Cascade in DNA-containing/DNA-free Regions

To get an independent measure of the total time fraction spent probing DNA, Cascade was visualized in cells that were elongated by addition of cephalexin. The drug cephalexin disabled the ability of the cells to divide, creating elongated cells where nucleoids were separated by DNA-free spaces (Reyes-Lamothe et al., 2014). Subregions of cell outlines were manually selected and further refined with the Oufti software (Paintdakhi et al., 2016). The relative amount of localizations of DNA-free and DNA-containing regions was not calculated for entire cells, as differences in illumination intensity between parts of the FOV could also change the amount of localizations detected for different parts of the cell. Each subregion contained one nucleoid free region, flanked by two nucleoid containing regions with a total length of around 4 μm. Segments of 0.1 μm divided along the long axis of the cell are separated into nucleoid or DNA-free segments based on the sum of the DAPI fluorescence within each segment. The average number of localizations of Cascade molecules in nucleoid segments divided by the average number of localizations Cascade molecules in DNA-free segments could be used to infer the DNA bound time fraction (see below, $f_{onDNA}$ from nucleoid enrichment).

### Persistence Sustained Binding Events for Different Integrated Times

To estimate how long binding events last, one could plot the number of particles remaining within a certain radius from the first frame position for different number of steps. However, particles can diffuse away when they are released from DNA or be lost due to photobleaching. To account for bleaching rates, previous studies increased dark time between exposures, while keeping exposure times the same (Ho et al., 2018; Knight et al., 2015). This approach uses the data of all time steps, including only single time steps.

As we are investigating lifetime of binding events on a subsecond timescale this approach fails, as single steps of slow-moving particles, which can be clearly separated from bound particles on larger timescales ($t_{int} > 1$ s), will be counted as bound particles leading to overestimated off-rates. At these timescales, it is more reliable to use tracks of at least 5 steps to distinguish bound from moving particles. As we are interested in how many of these events we observe, depending on the framerate, normalization is required.

For this we cannot use the sum of all tracks observed at each frametime, as a larger amount of fast moving molecules diffuse further than the maximum tracking distance of 0.78 μm between two exposures, and are also more affected by confinement with increasing integrated time. Therefore, the number of moving particles of certain track length is not an accurate normalization when comparing different frame times. However, as we used similar exposure for all frame times, the number of detected localizations per protein is unaffected. Furthermore, bound molecules are not affected by confinement or linking errors with increasing frame rates.

The most robust normalization procedure was therefore to normalize the number of localizations within sustained bound tracks (all localizations within 1 pixel of the mean location of the track) to the total number of localizations, as those do not depend on the length of introduced dark time between exposures. A further increase of the dark time was not possible as on longer timescales the movement of the plasmid ($D^*_{free}$ = 0.06 μm$^2$/s) made plasmid bound particles diffuse further than 1 pixel.

### Confinement and Localization Error Simulation

To verify whether our new transitional $D^*$ analysis yielded accurate parameter predictions and investigate the influence of localization error and confinement on the parameters of the fit, we simulated particles moving and transitioning between bound and free moving

states within the dimensions of an *E. coli* cell, adapted from methodology used in (Sanamrad et al., 2014). At every time step particles were simulated to be either in a bound state *S1* ($D = 0\ \mu m^2/s$), or a mobile state *S2* ($D = D_{free}$). At the starting time point, states were assigned to each particle according to the equilibrium probability $p_{S1}$ and $p_{S2}$ (Equations 10 and 11). Subsequently, at following time steps of 0.1 ms, particles in state *S1* were assigned to *S2* with a probability of $p_{S1\rightarrow S2} = k_{off}t_{step}$ (where $t_{step} = 0.0001\ s$) and particles in state *S2* were assigned to *S1* with a probability of $p_{S2\rightarrow S1} = k^*_{on}t_{step}$. Displacements in three dimensions at each time step were taken from a standard normal distribution multiplied with $\sqrt{2Dt_{step}}$ (where *D* is either 0 for particles in state *S1* or $D_{free}$ for particles in state *S2*). Steps beyond the boundaries of a cell were rejected and new displacements were randomly drawn.

The 2D projection of five localizations at 10 ms time intervals for each molecule was generated as output and was analyzed in our tracking software. Localization error was included in the simulation by addition of a random displacement for each position taken from a Gaussian distribution ($\sigma = 40$ nm). It was found that changes in outcome of the simulation were not sensitive to cell length in the range of our bacteria (3-6 μm), decreasing less than 5% for the smallest size. Most of the confinement effect is caused by the cell width, which was relatively constant between all the cells measured.

### Cascade Nucleoid Enrichment Simulation

The simulation above was adapted to simulate the movement in DNA-free and DNA-containing regions. Particles were simulated to move inside of a cell of 10 μm in length and 1 μm in width consisting of 100 segments without endcaps (0.1 μm per segment). Five segments were modeled as DNA-free segments and the rest of the segments as DNA-containing segments.

Cascade molecules were randomly placed throughout the cell and subsequently were simulating with similar time steps as described above, except that moving particles were only allowed to transition to *S1* (bound state) inside of the nucleoid containing regions. Before recording the position of the simulated particles, the simulation ran for 100.000 time steps (10 s) so that equilibrium was reached. Localization error was added in the same way as described above.

### Expected Free Diffusion Coefficients

The diffusion coefficient of molecules in classic (Newtonian) fluids can generally be estimated by the Stokes-Einstein equation. A study measuring the diffusion of GFP multimers inside the *E. coli* cytoplasm has shown good agreement with the predictions of this equation (Nenninger et al., 2010), whereas a second study found a different relation attributed to the complex nature of the cytoplasmic fluid (Mika and Poolman, 2011). To compare our findings of the apparent free diffusion coefficient of Cas8e (~3.5 $\mu m^2/s$) and Cascade (~1.0 $\mu m^2/s$), we therefore looked for reported free cytoplasmic diffusion coefficient values of proteins of similar size inside *E. coli* cells. For Cas8e, two proteins have been studied with a similar size to PamCherry2-Cas8e (82 kDa), namely CFP-CheR-YFP (86 kDa) (Kumar et al., 2010) and TorA-GFP3 (84 kDa) (Nenninger et al., 2010), which have reported values of 1.7 $\mu m^2/s$ and 6 $\mu m^2/s$. Our estimate for Cas8e lies within the range of these values. For Cascade (430 kDa), the closest reported protein in size is RNA polymerase, for which the $D^*_{free}$ was found to be 1.1 $\mu m^2/s$ (400 kDa core enzyme, 470 kDa holoenzyme) (Stracy et al., 2015). Furthermore larger proteins such β-Gal-$GFP_4$ (582 kDa; 0.6 $\mu m^2/s$) (Mika et al., 2010), and 30S ribosome subunits (900 kDa 0.4 $\mu m^2/s$) (Sanamrad et al., 2014) were reported with lower diffusion coefficients as expected. These findings support the free apparent diffusion value we found for Cascade (~1.0 $\mu m^2/s$).

### $f_{onDNA}$ from Nucleoid Enrichment

The distribution of Cascade in nucleoid-free and nucleoid containing regions depends on the time Cascade spends on DNA. We divided the cell up along the long axis into segments of 100 nm wide. During the time Cascade is bound to DNA it can only be inside of the nucleoid regions whereas, when it is not bound to DNA Cascade can be anywhere within the cell. Therefore, the average number of particles in a DNA-containing segment is given by:

$$\overline{N_{DNA}} = \left(\frac{f_{onDNA}}{sm_{DNA}} + \frac{1 - f_{onDNA}}{sm_{tot}}\right)N_{tot} \tag{22}$$

and the average number of particles in a DNA-free segment is given by

$$\overline{N_{DNA-free}} = \frac{1 - f_{onDNA}}{sm_{tot}}N_{tot} \tag{23}$$

Where $f_{onDNA}$ is the fraction of time bound to DNA, $sm_{DNA}$ and $sm_{tot}$ are the number of DNA segments and the total number of segments respectively and $N_{tot}$ is the total number of particles in a cell. The ratio, which is equal to the *EF*, can then be expressed as:

$$EF = \frac{\overline{N_{DNA}}}{\overline{N_{DNA-free}}} = \left(\frac{f_{onDNA}}{sm_{DNA}} + \frac{(1 - f_{onDNA})}{sm_{tot}}\right) \Big/ \frac{1 - f_{onDNA}}{sm_{tot}} \tag{24}$$

If the number of DNA-free segments is much less than the number of DNA segments $sm_{DNA} \approx sm_{tot}$ the expression above can be simplified to:

$$EF = \frac{1}{1 - f_{onDNA}} \tag{25}$$

This equation allows extraction of $f_{onDNA}$ from EF directly and implies that this value does not depend on the diffusion coefficients of the mobile population.

### *In Vivo* $K_D$ values

The $K_D$ value is a commonly calculated affinity constant used for binding kinetics of proteins and assembly of multicomponent systems (McGuigan et al., 2006), but the $K_D$ has also been used as an estimate for *in vivo* binding affinity (Zawadzki et al., 2015). In the reaction scheme A + B $\rightleftarrows$ AB, the $K_D$ is calculated as

$$K_D = [A][B]/[AB] \tag{26}$$

For Cascade the reaction scheme is as follows: [Cascade (probing)] + [free target sites] $\rightleftarrows$ [Cascade (bound)]. The concentration of a single entity inside of a cell of length 4 μm and width 1 μm with hemispherical endcaps is approximately 0.5 nM. The copy number for pTarget was estimated by qPCR to be approximately 100 plasmids per chromosome. As the number of chromosomes in actively dividing cells is generally higher than one, we used literature values for the number of chromosomes/cell found in (Wallden et al., 2016), providing 4/cell which also used a glucose and amino acid enriched M9 medium as growth medium. This brings the copy number of pTarget to 400/cell, which is equal to 200 nM. For a Cascade complex carrying one of several crRNAs in the cell, the amount of free target sites is equal to the copy number of the plasmid pTarget minus the amount of already occupied target sites of that crRNA, but as the copy number of each target (400) is much higher than the number of Cascade complexes potentially carrying that crRNA (on average 130/18 ≈7), [free targets] ≈[pTarget].The $K_D$ value was then calculated as:

$$K_D = [\text{pTarget}][\text{Cascade(probing)}]/[\text{Cascade(bound)}] \tag{27}$$

### Theoretical Model Interference Level Versus Copy Number

In the case where the interference level is limited by the target search of the proteins, we can model the relation based on the distribution of search times of single proteins. The search time for a single protein, because it is the arrival time of a recurring independent random event, is exponentially distributed and characterized by the average search time, $\langle t_s \rangle$:

$$p_1(t_s) = 1 \Big/ \overline{\langle t_s \rangle} e^{-t_s \Big/ \overline{\langle t_s \rangle}} \tag{28}$$

We have verified given our kinetic model of Cascade with simulations that this is the case (Figure S3). The chance that one of $n$ proteins finds the target at search time $t_s$ while the other proteins have not yet found the target is:

$$p_n(t_s) = np_1(t_s)\left(\int_{t_s}^{\infty} p_1(t)dt\right)^{n-1} = n \Big/ \overline{\langle t_s \rangle} e^{-nt_s \Big/ \overline{\langle t_s \rangle}} \tag{29}$$

We have verified this derivation with simulations (Figure S3). The establishment probability of the plasmid is equal to the likelihood for all search times larger than $t_{critical}$ ($t_c$), the time point at which the cell can no longer clear the invader. Therefore:

$$p_{establishment}(t_c) = \int_{t_c}^{\infty} p_n(t)dt = e^{-nt_c \Big/ \overline{\langle t_s \rangle}} \tag{30}$$

As the chance of targeting after replication is low, we assume in our model that Cascade is only able to clear the foreign DNA before replication. Therefore $t_c$ is equal to the replication time of the plasmid $t_R$.

As we found that 20 copies of Cascade reduce interference level by half, this leads to

$$\ln(0.5) = -20t_R \Big/ \overline{\langle t_s \rangle} \tag{31}$$

or

$$t_R \Big/ \overline{\langle t_s \rangle} = 0.035 \tag{32}$$

Right after transformation, the negative regulators of copy numbers are absent, so replication in that instant is faster than the growth rate of the cell. Replication time of pTarget has not been measured so far, but by using a temperature-dependent ori, Olsson et al. measured a replication time of 3 min for a slightly larger plasmid in the absence of copy number control (Olsson et al., 2003). If we assume pTarget replication occurs on a similar timescale, we get an estimated search time for one Cascade to find a single target of ~90 minutes.

We can further describe the relationship between the average search time $< t_s >$ and the $k_{off}$ and $k^*_{on}$ that were measured for Cascade. This relationship is found by multiplying the amount of time spent for each binding event times the average amount of binding events required to find the target. The amount of time spent for each binding event is equal to the sum of the time spent on binding $(1/k_{off})$ and the time spent on diffusing to the next site $(1/k^*_{on})$. Therefore the average search time is:

$$\overline{\langle t_s \rangle} = \left( \frac{1}{k_{off}} + \frac{1}{k^*_{on}} \right) \frac{\#DNA\ binding\ events}{\#DNA\ target\ sites} \quad (33)$$

We have again verified this description by using our simulations of our kinetic model of Cascade target search (Figure S3).

It must be emphasized that the number of binding events is different from the number of binding sites in the fact that if a single binding event scans multiple sites (during 1D sliding), the number of binding sites probed per event are more than one. Using Equations 30 and 33, the chance of establishment of a single invader in the cell with multiple Cascade copies is therefore as follows:

$$p_{establishment} \sim e^{-n/\#DNA\ binding\ events \left( \frac{1}{k_{off}} + \frac{1}{k^*_{on}} \right)} \quad (34)$$

### Simulation Cascade Search Times

To see whether the above described theoretical model was compatible with our kinetic model of Cascade search, we simulated the search times of Cascade. To do so, we simulated Cascade probing DNA sites as was described above (See Confinement and Localization Error Simulation). Subsequently every time Cascade changed from a mobile state to a bound state we added with a certain probability that the newly probed site is the target (1/90.000). When Cascade located the target the simulation for that particle was stopped and the search time was recorded for each individual Cascade complex. To simulate the search time for 5 Cascades, we grouped the single search times in multiples of five and took the fastest search time of 5 Cascades.

## QUANTIFICATION AND STATISTICAL ANALYSIS

The interference levels were done in biological triplicate, and the copy number and the log-transformed establishment probabilities was fitted with linear regression ($R^2$ indicated in the figure). The standard deviations of the extracted parameters for the diffusion were calculated with bootstrapping and depicted in the figures. Measured growth rates and qPCR were performed in biological triplicate and error reported as the S.E.M.

## DATA AND CODE AVAILABILITY

The following data is available at Mendeley data (https://doi.org/10.17632/w83h66729n.1): The measured copy number of Cascades in the cell, and their interference level measured with transformation assay (Figure 1). The single-molecule localizations, tracks and calculated D* values used to produce the D* histograms (Figures 2, 4, 5, and S4). The counts of single molecules within and outside the nucleoid region in cephalexin elongated cells (Figure 3). qPCR (Figure S5) and growth rate data (Table S1).

The raw microscopy video data and code that support the findings of this study are available upon request to the Lead Contact, Stan Brouns (stanbrouns@gmail.com).