Kinetics of dCas9 target search in Escherichia coli

Daniel Lawson Jones,* Prune Leroy,* Cecilia Unoson,* David Fange, Vladimir Ćurčić, Michael J. Lawson, Johan Elf†

How fast can a cell locate a specific chromosomal DNA sequence specified by a single-stranded oligonucleotide? To address this question, we investigate the intracellular search processes of the Cas9 protein, which can be programmed by a guide RNA to bind essentially any DNA sequence. This targeting flexibility requires Cas9 to unwind the DNA double helix to test for correct base pairing to the guide RNA. Here we study the search mechanisms of the catalytically inactive Cas9 (dCas9) in living Escherichia coli by combining single-molecule fluorescence microscopy and bulk restriction-protection assays. We find that it takes a single fluorescently labeled dCas9 6 hours to find the correct target sequence, which implies that each potential target is bound for less than 30 milliseconds. Once bound, dCas9 remains associated until replication. To achieve fast targeting, both Cas9 and its guide RNA have to be present at high concentrations.

Cells have evolved two strategies to search their genome for specific information. Transcription factors and restriction enzymes recognize a specific DNA sequence through interactions in double-stranded DNA (dsDNA) grooves, whereas other proteins are dynamically programmed by an RNA or single-stranded DNA (ssDNA) to recognize complementary nucleic acid sequences through base pairing. Examples of the latter are Argonaute (1) and Hfq (2) programmed by small RNAs to target and regulate mRNA, the homologous repair machinery primed by ssDNA to target dsDNA (3), and Cas proteins programmed by guide RNA also to target dsDNA (4–7). How transcription factors search and bind DNA is well understood (8), and in vivo kinetics have been studied for Hfq-mediated targeting by small RNAs (9), but very little is known about how the factors that are dynamically programmed for base pairing specific sequences in dsDNA find their targets in the context of myriad similar sequences.

Cas protein targeting and homologous recombination depend on the unwinding of dsDNA throughout the genome to test for complementarity (Fig. 1) (10, 11). In the case of Cas proteins, the search problem is simplified by requiring a protospacer-adjacent motif (PAM) as a prerequisite for unwinding the dsDNA (4, 7, 12–14). For Streptococcus pyogenes Cas9, which has a GG-dinucleotide PAM, this still implies that every eighth base pair in a genome has to be interrogated (1). Here we investigate how long it takes Cas9 to find a specific target sequence in Escherichia coli and what insight this provides into possible search mechanisms.

To measure the time required for Cas9 to locate a specific target, we fused the catalytically inactive Cas9 (dCas9) to the fluorescent protein YPet (figs. S1 and S6) and expressed it at a few (about five) molecules per cell from the chromosome. At this expression level and a 5-s image acquisition time, nonbound molecules contribute to the diffuse fluorescence background, whereas DNA-bound fluorophores are detectable as individual diffraction-limited spots (15). The dCas9-YPet was programmed by a single-guide RNA (sgRNA-a; see fig. S3C) (5) targeted against the lacO1 operator sequence. This allowed us to trigger the accessibility of the target sequence by isopropyl-β-D-thiogalactopyranoside (IPTG), because IPTG induces dissociation of the LacI repressor from lacO1 (8, 16) (Fig. 2A). In the absence of sgRNA or target sites, very few spots are observed (fig. S2). Throughout the rest of the experiments, all sgRNA were present at saturating concentrations; at these concentrations, every dCas9-YPet is in complex with sgRNA (fig. S3).

We used an array of 36 lacO1 sites cloned in a bacterial artificial chromosome (pSMART, 2.1 copies per cell, fig. S4) as the search target (see table S1 for array sequence). The array was used to speed up the first binding event while maintaining a low copy number of dCas9-YPet, such that individual bound molecules can be detected over the fluorescent background.

We measured the association rate on the basis of the time-dependent appearance of fluorescence spots, corresponding to immobile dCas9-YPet, after making the target sites accessible by IPTG addition (Fig. 2A). Different sets of ~200 cells were imaged at each time point to avoid the complications of photobleaching in time-lapse measurements (15). The rate of dCas9-YPet binding to any target site was determined by an exponential fit to the experimental data for the first binding event per cell (Fig. 2B). When we also accounted for the number of plasmids (fig. S4), that only 50% of the plasmids are accessible to binding (Fig. 2B, left inset; fig. S5);
and supplementary text section 2.2.4), and also the number of fluorescent dCas9-YPet and its uncertainty due to non–full-length bands on the Western blot (6 ± 1, supplementary text section 2.15), we obtained the association rate $2.7 \times 10^{-3} \pm 0.6 \times 10^{-3}$ min$^{-1}$ molecule$^{-1}$ (for details, see supplementary text section 2.1.1). Figure 2B does not plateau at 1 mainly because of cell-to-cell variation in the number of accessible pSMARTs per cell (Fig. 2B, left inset) and the distribution of fluorescent dCas9-YPet per cell (Fig. 2B, right inset). The possible sources of error in the association rate determination due to the maturation of YPet (fig. S7) and potential sliding of dCas9 (fig. S8B) across the array were found to be small (supplementary text sections 2.2.1 and 2.2.2) and are not included in the rate estimate. In summary, an individual sgRNA-programmed dCas9-YPet protein requires, on average, 6 hours to find and bind its target site.

To test how the activity of the fluorescent fusion protein compares to the native protein, we developed a bulk restriction-protection assay (Fig. 2C) where a single lacO1 site, introduced in the intC position of the E. coli chromosome, is targeted by dCas9 (sgRNA-b; see fig. S8C), which is expressed at a 15.1 ± 6.5-fold higher concentration than in the fluorescence assay (fig. S1 and supplementary text section 2.1.2). The lacO1 site contains a cleavage site for the restriction enzyme Bsr BI, and, therefore, binding of dCas9 can be measured as protection from Bsr BI cleavage. After adding IPTG at time zero, which renders the target site accessible for dCas9, we determined the fraction of protected cleavage sites at multiple time points using quantitative polymerase.

**Fig. 2. dCas9 association-rate measurements.** (A) Top panel shows a schematic illustration of the single-molecule assay, where dCas9-YPet is expressed at a low copy number in a strain containing pSMART plasmids, each with 36 lacO1-binding sites. In the absence of IPTG, lacO1 sites are occupied by LacI, preventing dCas9-YPet from binding. LacI dissociates after addition of IPTG, and subsequently, dCas9-YPet binds the lacO1-a target, enabling specific fluorescent spots to be detected using exposure times of 5 s. Bottom panel shows fluorescence images acquired before (left) and 10 min after (right) addition of IPTG (strain PL42F9). Scale bar, 2 μm. (B) Fraction of cells containing at least one spot (y axis) as a function of time after addition of IPTG (x axis). The curve is fitted to a single exponential function. Left inset shows distributions of the number of fluorescent dCas9-YPet (strain PL42F2). Right inset shows the distribution of accessible pSMART (strain PL41D2) per cell, which is based on the pSMART–copy number distribution (fig. S4) corrected for the fact that 50% of the pSMARTs are available for binding (fig. S5). In total, 63,382 cells were analyzed. (C) Top panel shows a schematic illustration of the bulk assay, in which dCas9 or dCas9-YPet is expressed at a high copy number in the presence of only one target site per chromosome. LacI dissociates from lacO1 after addition of IPTG, permitting dCas9-YPet to bind. Bottom panel shows a schematic of the experimental procedure. IPTG is added at time zero, freeing the lacO1 site. At later times, here time 1 or time 2, different batches of cells are fixed, and bound dCas9 are cross-linked to DNA; the chromatin is subsequently purified and digested with Bsr BI. The fraction of cut and uncut DNA is quantified by qPCR with two sets of primers, one upstream and one amplifying across the cut site. (D) Measured bulk association to lacO1-b in the intC locus for dCas9-YPet (strain DJ3F6) is compared to that of dCas9 without YPet (ΔYPet) (strain DJ3I3) (see table S1 for sequences). Note that the concentrations are slightly different in the two strains, which is why the rates should not be directly compared (see main text and supplementary text section 2.1.3.). (E) Measured bulk association to lacO1-a in the intC locus for dCas9-YPet with (strain DJ3F5) or without (strain DJ3F8) a PAM adjacent to the protospacer. For (D) and (E), each data point is derived from three biological replicates (each consisting of three technical replicates). See fig. S6 and table S1 for strains and sequences.
chain reaction (qPCR) (Fig. 2C). The association rate in the dCas9-YPet strain was $1.7 \pm 0.3$ times slower than in the dCas9 strain after we adjusted for the difference in abundance and accounted for uncertainty due to non-full-length bands on the Western blot (Fig. 2D, fig. S1, and supplementary text section 2.1.3). This suggests that the fusion protein is partly impaired, although the absolute in vivo activity of the fusion protein is partly impaired, although the absolute in vivo activity of the fusion protein is partly impaired, although.

We also used the restriction-protection assay with dCas9-YPet for a number of pairwise comparisons: (i) When there was no PAM sequence, we detected no binding (Fig. 2E). (ii) When we targeted the same target sequence introduced at different positions in the chromosome, the rate of binding changed by about 1.8-fold as compared to the intC position, which may be due to a difference in accessibility for dCas9 binding, a difference in the amount of genomic DNA per target site, or a difference in the efficiency in cross-linking (fig. S8A). (iii) When we changed the sgRNA and targeted another part of lacO at the same chromosome position, the rate of binding changed by about 1.6-fold as compared to the lacO1b target (Fig. 2, D and E), which may be due to how the different sgRNA seed regions interact with DNA at nontarget PAM sequences or influence the probability of binding at the target sequence.

Next, we measured how long dCas9-YPet stays bound at the target by reversing the single molecule–association experiment. Thus, cells were initially grown with IPTG to permit dCas9-YPet to bind specifically to the lacO1 array. IPTG was removed at time zero, and dissociation was monitored as the decrease in the mean number of spots per cell (Fig. 3, A and B, and fig. S10). The procedure was repeated under various growth conditions. We observed a strong correlation between generation time and the time taken for all molecules to dissociate (Fig. 3C). This agrees well with the in vitro observation that dCas9 enters an irreversibly bound state once the spacer sequence has hybridized (II) and also with the very slow dissociation measured in eukaryotic cells (II), considering that dCas9 necessarily dissociates at replication.

The dissociation rate measurement offers an opportunity to test the consistency of our association rate estimates based on the steady-state target-site occupancy. We targeted the native lacO1 site in the lacZ promoter with sgRNA-b such that lacZ expression is off when dCas9 is bound (Fig. 3D, inset). The predicted repression ratio (RR) is

$$RR = \frac{rCT}{1 - e^{-rCT}}$$  \hspace{1cm} (1)

where $r$ is the rate of binding per dCas9-YPet, $C$ is the number of dCas9-YPet, and $T$ is the generation time (see supplementary materials and methods section 1.5.2 for derivation).

The bulk association rate assay gives $r = 0.34 \text{ min}^{-1}$ for this strain (fig. S8A). We constructed five variants of the lacZ targeting strains that constitutively express different levels of dCas9-YPet (table S1 and fig. S1). The repression ratio was determined as the ratio of lacZ expression in the absence of the cognate sgRNA to lacZ expression in its presence. In Fig. 3D, we plot the repression ratio against the expression level of dCas9-YPet. The data agree with the prediction (Eq. 1) on the basis of the association rate measurement, the dCas9-YPet expression levels, and the generation time.

Given that it takes dCas9 6 hours to test the $10^9$ PAMs ([(5.4 × 10^7 PAM/ genome) × (∼2 genomes/cell)], it can only spend ~20 ms per PAM unwinding the DNA and testing for hybridization if it binds irreversibly the first time it reaches the cognate target. This is much faster than the 750 ms reported for eukaryotic cells (18) and the ~10 s measured in vitro (II). To investigate whether all PAMs are sampled, we imaged dCas9-YPet at exposure times ranging from 2 ms to 1 s in cells that do not have a specific target site (Fig. 4A). At each exposure time, we selectively observed the fraction of molecules that were immobile, and thus bound, for at least the length of the exposure (Fig. 4B). To translate this into a distribution of nontarget PAM residence times,
we scaled the corresponding density function to account for the different number of binding events at different binding times and renormalized (supplementary materials and methods section 1.7.2.4). This results in a broad distribution with an average of 30 ms (Fig. 4C). This is an upper limit for the average nonspecific residence time, because transiently bound (<5 ms) molecules are indistinguishable from freely diffusing molecules. The <30-ms nonspecific residence time is clearly compatible with a search mechanism that explores all PAMs.

Given the search time of 6 hours per molecule, we may ask if Cas9 could be effective as an adaptive immune system in *S. pyogenes* (19). We determined the abundance of Cas9 in *S. pyogenes* by Western blotting to be almost twice that of the nonfused dCas9 strain (Fig. S1) where the time to bind a single target is 2 min (Fig. 2D), suggesting a search time of 1 min in *S. pyogenes*. Furthermore, the frequency of GG in the *S. pyogenes* genome is two-thirds that of *E. coli*, which can be expected to reduce the search time to 40 s. Thus, the targeting time in *S. pyogenes* could be as little as a few minutes, depending on which fraction of the Cas9 proteins are programmed by the relevant spacer RNA. This suggests that no additional rate-enhancing factors are needed for a Cas9 search in *S. pyogenes*. Overall, dCas9 kinetics is slow because of the flexible targeting mechanism, but association can be made fast for a few selected targets by using high copy numbers.

**REFERENCES AND NOTES**


**SUPPLEMENTARY MATERIALS**

www.sciencemag.org/content/357/6358/1420/suppl/DC1

**REFERENCES**


**ACKNOWLEDGMENTS**

The authors thank K. Abruzzi for supplying the plasmid pAF552 and I. Fontana and E. Charpentier for help with making *S. pyogenes* protein lysates. This work was supported by the European Research Council, the Knut and Alice Wallenberg Foundation, and Vetenskapsrådet. Author contributions are as follows: J.E. and C.U. conceived the project; J.E., C.U., D.L.J., D.F., M.J.L., and P.L. designed the study; D.L.J., C.U., M.J.L., D.F., and P.L. conducted the experiments; D.L.J., P.L., D.F., and V.C. performed the analysis; C.U., P.L., and D.L.J. made the bacterial strains and plasmids; D.L.J., D.F., and J.E. derived the theory; and J.E., D.L.J., P.L., and D.F. wrote the paper.
Flexible association comes at a price

The CRISPR-Cas9 genome editing system is guided to its target DNA sequence by a small RNA that must search the genome to find its target site. To allow the guide RNA to bind, Cas9 must unwind the DNA at each location that it searches. Jones et al. used single-molecule fluorescence methods and bulk biochemistry to show that Cas9 takes 6 hours to find its target sequence, with each potential target bound for less than 30 ms. Overall, the CRISPR-Cas9 system pays for its flexible association mechanism with slow kinetics, but this can be overcome by using high concentrations of Cas9 and guide RNA.

Science, this issue p. 1420