High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding

Evan A. Boyle[a,1], Johan O. L. Andreasson[a,b,1], Lauren M. Chircus[c,1], Samuel H. Sternberg[d,2], Michelle J. Wu[e], Chantal K. Guegler[d,3], Jennifer A. Doudna[d,f,g,h,i], and William J. Greenleaf[a,j,4]

[a]Department of Genetics, Stanford University, Stanford, CA 94305; [b]Department of Biochemistry, Stanford University, Stanford, CA 94305; [c]Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305; [d]Department of Chemistry, University of California, Berkeley, CA 94720; [e]Biomedical Informatics Training Program, Stanford University, Stanford, CA 94305; [f]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; [g]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720; [h]Innovative Genomics Initiative, University of California, Berkeley, CA 94720; [i]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720; and [j]Department of Applied Physics, Stanford University, Stanford, CA 94305

Edited by Scott Bailey, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, and accepted by Editorial Board Member Kiyoshi Mizuuchi April 13, 2017 (received for review January 11, 2017)

The bacterial adaptive immune system CRISPR–Cas9 has been appropriated as a versatile tool for editing genomes, controlling gene expression, and visualizing genetic loci. To analyze Cas9's ability to bind DNA rapidly and specifically, we generated multiple libraries of potential binding partners for measuring the kinetics of nuclease-dead Cas9 (dCas9) interactions. Using a massively parallel method to quantify protein–DNA interactions on a high-throughput sequencing flow cell, we comprehensively assess the effects of combinatorial mismatches between guide RNA (gRNA) and target nucleotides, both in the seed and in more distal nucleotides, plus disruption of the protospacer adjacent motif (PAM). We report two consequences of PAM-distal mismatches: reversal of dCas9 binding at long time scales, and synergistic changes in association kinetics when other gRNA–target mismatches are present. Together, these observations support a model for Cas9 specificity wherein gRNA–DNA mismatches at PAM-distal bases modulate different biophysical parameters that determine association and dissociation rates. The methods we present decouple aspects of kinetic and thermodynamic properties of the Cas9–DNA interaction and broaden the toolkit for investigating off-target binding behavior.

DNA | molecular biophysics | kinetics | sequencing | CRISPR

CRISPR-associated protein 9 (Cas9) is programmed to bind its target DNA by a guide RNA (gRNA) that, once loaded, forms a ribonucleoprotein (RNP) complex. The *Streptococcus pyogenes* CRISPR system, the most extensively studied and applied system to date, targets a 23-bp DNA sequence containing (*i*) an "NGG" protospacer adjacent motif (PAM) element downstream of the single-guide RNA (sgRNA) target DNA (1) and (*ii*) a 20-bp sequence upstream of the PAM bearing complementarity to the gRNA (2). Genome engineering applications leverage the nuclease activity of the Cas9 RNP, but Cas9 engineered to lack the residues required for cleavage [dCas9 (nuclease-dead Cas9)] has proven valuable by enabling the creation of customizable and programmable DNA binding elements that can activate and repress gene expression with high precision (CRISPRa and CRISPRi) (3).

The biophysical underpinnings of the Cas9 target search have been investigated both by directed biochemical assays (4, 5) and through measurements of off-target Cas9 activity (6–11). These studies have led to a model for binding wherein Cas9 proceeds through a series of steps starting with PAM recognition, followed by DNA melting, RNA strand invasion, and heteroduplex formation dependent on complementarity with a 5–10-bp seed. Structural data have further suggested that conformational changes in the HNH domain reposition catalytic residues and permit allosteric regulation of the RuvC domain. This conformational gating ensures that cleavage occurs only in the context of substantial homology between gRNA and target (12, 13).

The specificity of Cas9 DNA binding is crucial for all potential applications of Cas9's RNA-programmable targeting. Localization of dCas9 using chromatin immunoprecipitation followed by sequencing (ChIP-seq) has indicated that Cas9 stably binds sequences even with multiple mismatches at PAM-distal bases (9, 10, 14); however, analysis of CRISPRi/a screens has suggested that nearly all mismatches across the length of the target contributed to binding specificity (15). Neither of these approaches gauges occupancy over time, which makes direct measurement of biophysical parameters governing dCas9's interactions with target sequences impossible. Thus, there exists an acute need for scalable approaches for exhaustive profiling of off-target binding in vitro that can shed light on the full extent to which sequence controls

Significance

Cas9, a protein derived from the bacterial CRISPR/Cas9 immune system, relies on a programmable single-guide RNA (sgRNA) to bind specific genomic sequences. Cas9 complexed with sgRNA readily binds on-target DNA, but models that can predict the specificity of this process have proven elusive. To investigate this system from a biophysical perspective, we applied a massively parallel method for profiling protein–DNA interactions to quantify nuclease-dead Cas9 (dCas9) binding across thousands of off-target sequences. We observe that mismatches at certain positions of the guide lead to complex dCas9 dissociation patterns, and multiple mismatches between the gRNA and DNA at nonseed bases can produce substantial changes in observed association and dissociation, suggesting the possibility of kinetic and thermodynamic tuning of Cas9 behavior.

Author contributions: E.A.B., J.O.L.A., L.M.C., S.H.S., J.A.D., and W.J.G. designed research; E.A.B., J.O.L.A., L.M.C., S.H.S., M.J.W., and C.K.G. performed research; E.A.B., S.H.S., and J.A.D. contributed new reagents/analytic tools; E.A.B., J.O.L.A., L.M.C., and M.J.W. analyzed data; and E.A.B. and J.O.L.A. wrote the paper.

Conflict of interest statement: S.H.S. is an employee of Caribou Biosciences, Inc. and an inventor on patents and patent applications related to CRISPR-Cas systems and applications thereof. J.A.D. is a cofounder of Editas Medicine, Intellia Therapeutics, and Caribou Biosciences and a scientific advisor to Caribou, Intellia, eFFECTOR Therapeutics, and Driver. J.A.D. receives funding from Roche, Pfizer, the Paul Allen Institute, and the Keck Foundation.

This article is a PNAS Direct Submission. S.B. is a guest editor invited by the Editorial Board.

Data deposition: The sequences reported in this paper have been deposited in Sequence Read Archive database (accession nos. SRP102425 and SRP076741).

[1]E.A.B., J.O.L.A., and L.M.C. contributed equally to this work.

[2]Present address: Caribou Biosciences, Inc., Berkeley, CA 94710.

[3]Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

[4]To whom correspondence should be addressed. Email: wjg@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1700557114/-/DCSupplemental.

dCas9 biophysical binding parameters, allowing for both comprehensive characterization of off-target potential of specific guides and the generation of sufficient data for a predictive model of the physical process underlying dCas9 affinity.

To investigate the sequence determinants of Cas9 binding, we performed a direct, comprehensive survey of dCas9 off-target binding potential. We generated a library of mutant targets on a massively parallel array and assessed binding of fluorescently labeled dCas9–sgRNA complexes in real time (*Materials and Methods*) (16). We chose a well-characterized 20-bp phage λ-target sequence (4, 13) and constructed a library of modified targets with maximal coverage of double substitutions. Flanking Illumina sequencing adapters (Fig. 1*A*) permitted cluster generation and sequencing on an Illumina flow cell. Following sequencing, the GAIIx flow cell comprised a 2D array of clonal, relaxed-state DNA clusters with the template strand tethered to the surface of the flow cell. Each cluster of identical potential DNA binding sites contained anywhere from 0 to 20 substitutions in the 20-nucleotide λ-target sequence plus 3-nucleotide PAM.

After using the high-throughput sequencing data to define the spatial coordinates of sequence clusters in the library, the flow cell was placed into a modified GAIIx instrument (17) for biochemical profiling. Programmed Cy3-labeled RNP complexes were introduced into the flow cell at either 1 nM or 10 nM concentration (*SI Text*) and left to incubate 12 h overnight. Following this incubation, the flow cell was washed with dCas9-free buffer and imaged to track dissociation of dCas9. During association and dissociation experiments, images were collected across all 120 tiles of the flow cell lane with 532-nm excitation (Fig. 1*B*). For each experiment, initial apparent on-rates and off-rates were calculated by fitting fluorescence values for each off-target, which estimate presteady state on-rates and initial observed off-rates (Fig. 1 *C* and *D* and *Materials and Methods*). Because dCas9's strand invasion behavior is not expected to obey simple two-state binding dynamics, we treat these parameters as empirical measurements of binding and unbinding.

Apparent initial association rates were obtained for 84,554 sequences, including all single mutants, 99% of all possible double mutants, and 59% of all possible triple mutants, as well as 64,594 higher order mutants (Fig. 1*E*). Datasets (1 and 10 nM) were merged to evaluate apparent on-rates jointly (*Materials and Methods* and Fig. S1*A*). Single mutants were generally measured across >1,000 clusters. Sequences with four or more mismatches were typically measured across 20 or fewer clusters due to the larger mutational space (Fig. 1*F*). Across single and double mutants, dCas9 initial association rates were highly reproducible ($R^2 = 0.962$) (Fig. 2*A*). Among all potential binding targets, reproducibility was slightly reduced due to lower per-sequence cluster counts ($R^2 = 0.856$; Fig. S1 *B* and *C*).

Stark differences in apparent association rates between targets with intact and disrupted PAM GG dinucleotides agreed with known (d)Cas9 requirements for binding. All off-target DNA with mutations in the PAM GG dinucleotide exhibited approximately equivalent (and slow) association rates. Because most constructs contained at least one GG dinucleotide, either in the barcode or introduced in the λ-target sequence itself, we inferred that these association rates represented slowly accumulating background signal likely related to dCas9's interrogation of PAM elements. Among off-targets lacking such a canonical PAM adjacent to the λ-target positions, we found that targets with no detectable signal on average contained fewer novel GG dinucleotides than those with small but detectable signal, both on the sgRNA sense strand (0.54 vs. 0.71 novel GGs per sequence, $P < 1 \times 10^{-280}$, Wilcoxon rank sums test) and on the strand complementary to the sgRNA (0.48 vs. 0.61, $P < 1 \times 10^{-280}$).

The extent to which dCas9 can recognize PAM sequences aside from GG dinucleotides—known as noncanonical PAMs—has been the subject of conflicting reports (7, 18, 19). At 10 nM dCas9, the initial association rate of NGA or NAG targets was similar to the PAM-scanning behavior we described; however, after equilibration 12 h later, both NGA and NAG PAMs exhibited more signal than other PAM mutants (Fig. S2), suggesting that these are indeed the two most prominent noncanonical PAMs.

We next examined the effect of single mismatches in the bases complementary to the sgRNA (positions −20 to −1) on apparent

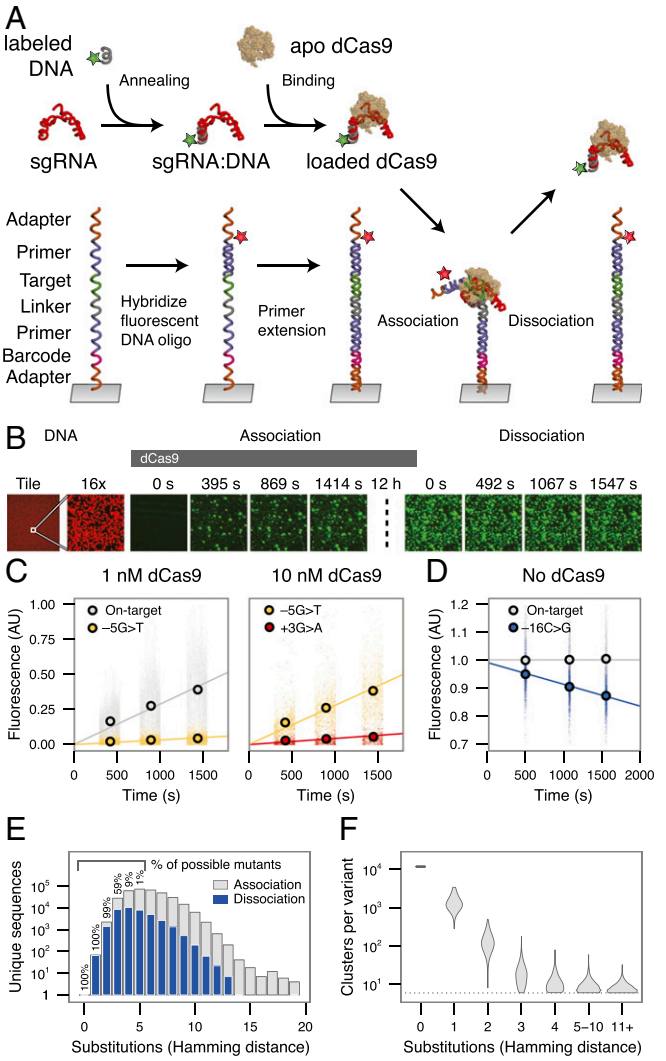

Fig. 1. Quantifying dCas9 binding behavior on a massively parallel array. (*A*) Experimental procedure for high-throughput biochemical profiling. A fluorescent DNA oligo hybridized to the dCas9 sgRNA was loaded into the apo-dCas9. In parallel, an Illumina sequencing-compatible DNA construct was both labeled and made double-stranded by extending a second fluorescent oligo. dCas9 was flowed into the chamber, allowing association with double-stranded DNA. A dissociation experiment was then performed by quantifying the decrease in dCas9 signal upon dilution or chase. (*B*) Example images taken in two channels on the array, Alexa Fluor 647-labeled DNA (red) and Cy3-labeled dCas9 (green). A 12-h incubation, meant to saturate the clusters with dCas9, separates association from dissociation experiments (dotted line). For most clusters, signal accumulated in the on-rate experiment largely remains throughout the dissociation. (Magnification: right nine panels, 16×.) (*C* and *D*) Examples of (*C*) association and (*D*) dissociation lines fit to different targets. The +1 base refers to the first base of the PAM, −1 to the most PAM-proximal base, and −20 to the most PAM-distal base. (*E*) The total number (*y* axis) and percentage (in text) of possible targets profiled for each number of substitutions from the on-target site. Only a fraction of sequences with quantified on-rates are profiled for off-rates (blue) with high confidence. (*F*) Clusters per variant for targets with the given number of substitutions. AU, arbitrary units.

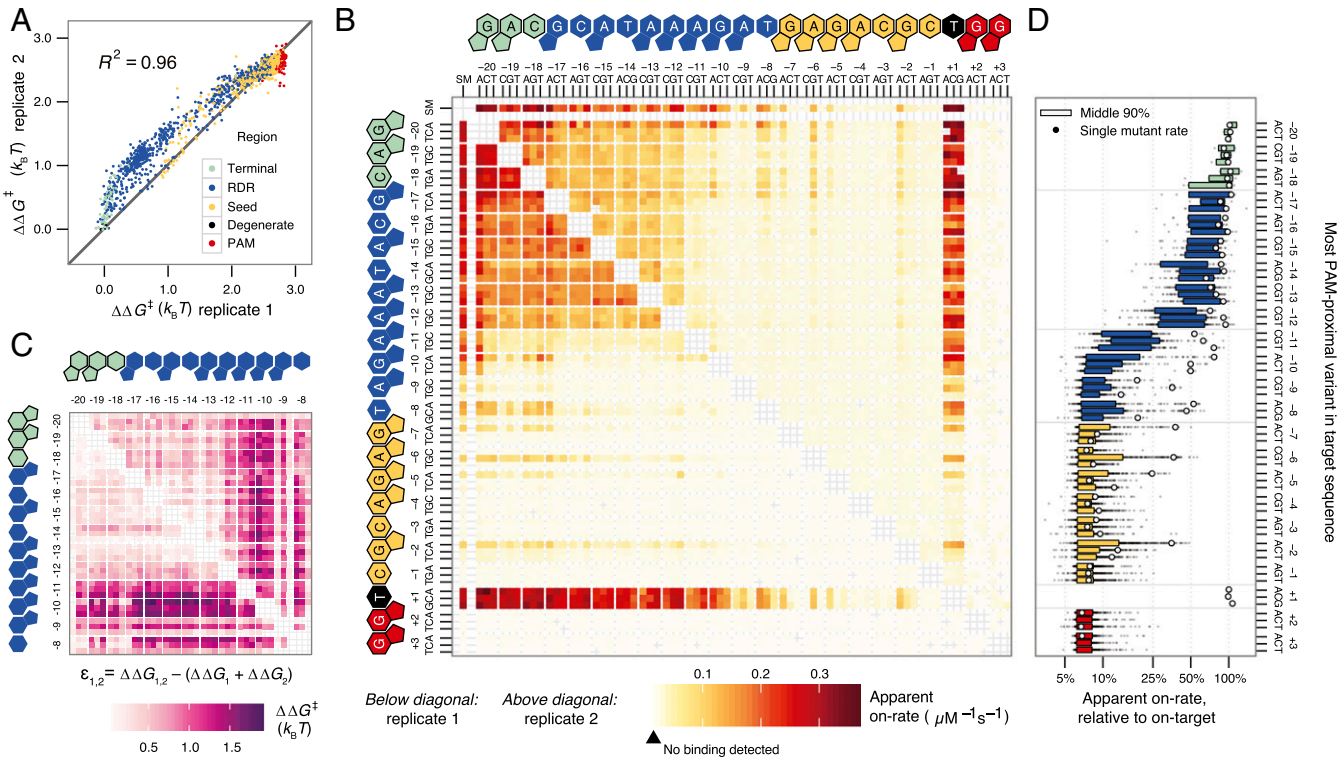

Fig. 2. Deep profiling of dCas9 observed initial association rates across a range of potential off-target sequences. (*A*) dCas9 effective energy barrier reproducibility (natural log of the ratio of observed initial on-rate to on-target observed initial on-rate) for single and double mutants across replicates, calculated relative to the on-target DNA. Points are colored by the more PAM-proximal mutation position, excepting the degenerate base of the NGG PAM. (*B*) Apparent association rates for all single mutants (the series of tiles "SM," horizontal and vertical) and double mutants (all other tiles) across both replicates, shown above and below the diagonal. Heat reflects higher on-rate for off-target sequences with the substitutions indicated on the *x* and *y* axes. Targets with at least six clusters but with no detectable binding were colored the minimum quantified rate. Double mutant cells lacking six clusters are left unfilled. (*C*) Epistasis in energy barriers for double mutants for the PAM-distal nucleotides. Nearly all pairs of mismatches have slower rates than expected by single mismatch estimates. Targets with mismatches in the seed were excluded owing to low variation in rate. (*D*) Distribution of higher order (>2) mutant on-rates summarized by their most PAM-proximal mutation (degenerate base excluded). Single mutants are highlighted in outlined white circles and correspond to the single mutant rate data in *C*.

association rates for canonical dCas9 binding. We observed that mismatches in the ∼7-bp seed region (positions −1 through −7) caused substantial changes to these apparent initial association rates (Fig. 2*B*). Low association rates for seed-mismatched off-targets are due to rapid rejection of these targets by dCas9 following a PAM-dependent initial association. Substitutions outside the seed had more muted effects on apparent initial association rates, leading generally to <twofold changes in apparent on-rates. Although seed mismatches posed greater barriers to dCas9 binding than PAM-distal mismatches (positions −8 to −20), we found that apparent initial association rates were sensitive to both position and base identity of the mismatches between target and sgRNA (Fig. 2*B*).

Next we asked if the effect of multiple substitutions in one off-target could be predicted from the effective energy barriers faced by DNA templates possessing the constitutive single mismatches. We first applied a naive model under which energy barriers to dCas9 association on doubly mismatched DNA templates were additive. For this analysis, PAM and seed mismatches were not assessed, as canonical binding was largely abrogated and presumably reflected PAM-scanning behavior independent of mismatches. The identity of the degenerate NGG PAM base (position +1) had no detectable effect on apparent on-rate kinetics (Kolmogorov–Smirnov test $P > 0.05$), consistent with prior observations (8, 11, 20), and was also excluded from modeling. In the PAM-distal region (positions −8 to −20), we found two milieus of negative epistasis (Fig. 2*C*). For many PAM-distal bases (−12 to −20), double mutants exhibited slightly lower apparent initial association rates than expected under this naive model. The four bases adjacent to the seed (−8 to −11) showed a more exaggerated decline in apparent initial association rates when paired with a second mismatch (Fig. S3).

These patterns were also observed among targets with greater numbers of mismatches (Fig. 2*D*). Curiously, although the three terminal PAM-distal bases (−18 to −20) are considered dispensable for binding (21), and single mismatches in this region produced little change in association rate, we found that the presence of a second mismatch in the four bases adjacent to the seed (−8 to −11) greatly sensitized dCas9 to mismatches in the terminal nucleotides (Fig. 2*C* and Fig. S3*B*). This sensitization was comparable to that observed for other PAM-distal mutants (−12 to −17).

Double substitutions in the seed largely abrogated dCas9 occupancy even after 12 h of binding (Fig. S2). In contrast, the vast majority of single substitutions achieved high levels of occupancy at this time point, even for sequences with slow apparent initial association rates. We also observe considerable variation in the fraction of DNA ultimately bound by dCas9 for double substitutions, suggesting that PAM-distal mismatches that in isolation have little effect on dCas9 association can, in concert with other mismatches, substantially alter dCas9 binding at long time scales.

To obtain a more mechanistic understanding of how mismatches might impair dCas9 association kinetics, we fit a modified version of a previously described kinetic Monte Carlo strand invasion model (22) that could account for mutations throughout the guide sequence (*Materials and Methods* and Fig. S3*A*). This model gave reasonable predictions for single-base

mutations but was unable to recapitulate the nonadditivity we observe in our double mutants (Fig. S3*C*).

To our knowledge, there has been no systematic characterization of how target mismatches at base-pair resolution effect changes in (d)Cas9 off-rates over long time scales. Biophysical data (22, 23) and modeling efforts (24) have suggested that target mismatches may modulate dCas9 off-rates, but methods for probing these features at scale generally lack the temporal resolution needed to probe these comparatively slow off-rates.

In our data, we report substantial variation in apparent initial dissociation rates across off-targets (Fig. 3*A*). In contrast to dCas9 apparent association rates, the apparent dCas9 dissociation rates we estimate are almost exclusively modulated not by the seed but by bases in the PAM-distal region; accordingly, we define a new region, corresponding to positions −8 to −17, as the "reversibility-determining region" or RDR, which modulates both association and dissociation of Cas9 at the time scale of minutes. Although dissociation was immeasurably slow for the on-target λ sequence, consistent with past investigations (4), we found that a single mismatch in the PAM-distal region (−16G) induced near complete dissociation of dCas9 within an hour. To confirm this unexpected behavior, the −16G construct, along with several other test sequences, were assayed for association and dissociation by radioactive filter binding assays (Fig. S4). In general, our data suggest that off-targets with high apparent dissociation rates tend to have lower apparent association rates. The converse, however, is not true. This suggests that the reversible binding we observe relies on subtle modulations of the multistep strand invasion process (Fig. S5*A*). The addition of unlabeled competitor DNA in the high-throughput sequencing flow cell (HiTS-FLIP) experiment yielded systematically higher dCas9 off-rates, but these data were still strongly correlated with passive flow experiments ($R^2 = 0.752$; Fig. S5 *B* and *C*). These observations are broadly consistent with an initial scanning phase of dCas9 binding that is susceptible to interference by competitor DNA. Our report of 38,431 initial observed off-rates for dCas9 thus enables exploration of how different mismatches between target and sgRNA modulate the final state of dCas9 binding.

From these data, it appears that if gRNA strand invasion bypasses seed mismatches in off-target DNA, the final complex is still made highly stable via favorable PAM-distal base pairing. In contrast, R-loop formation over RDR mismatches can jeopardize the long-term stability of dCas9 binding, even with perfect seed complementarity. In further support of this hypothesis, we observe that multiple PAM-distal mismatches, especially in the most distal bases of the RDR (positions −12 to −17), trigger faster dissociation than that of the single mismatches alone (Fig. S5*D*).

To expand upon these results, we developed a label-free method to study protein–DNA interactions, which we term the massively parallel filter-binding assay (*Materials and Methods*). By incubating Cas9 with pools of dsDNA libraries, passing the mixture through protein-binding nitrocellulose membranes at set time points, and sequencing the flow-through, we kinetically resolve Cas9 binding of thousands of species simultaneously by examining the depletion of sequencing counts per species from the pool over time. With this

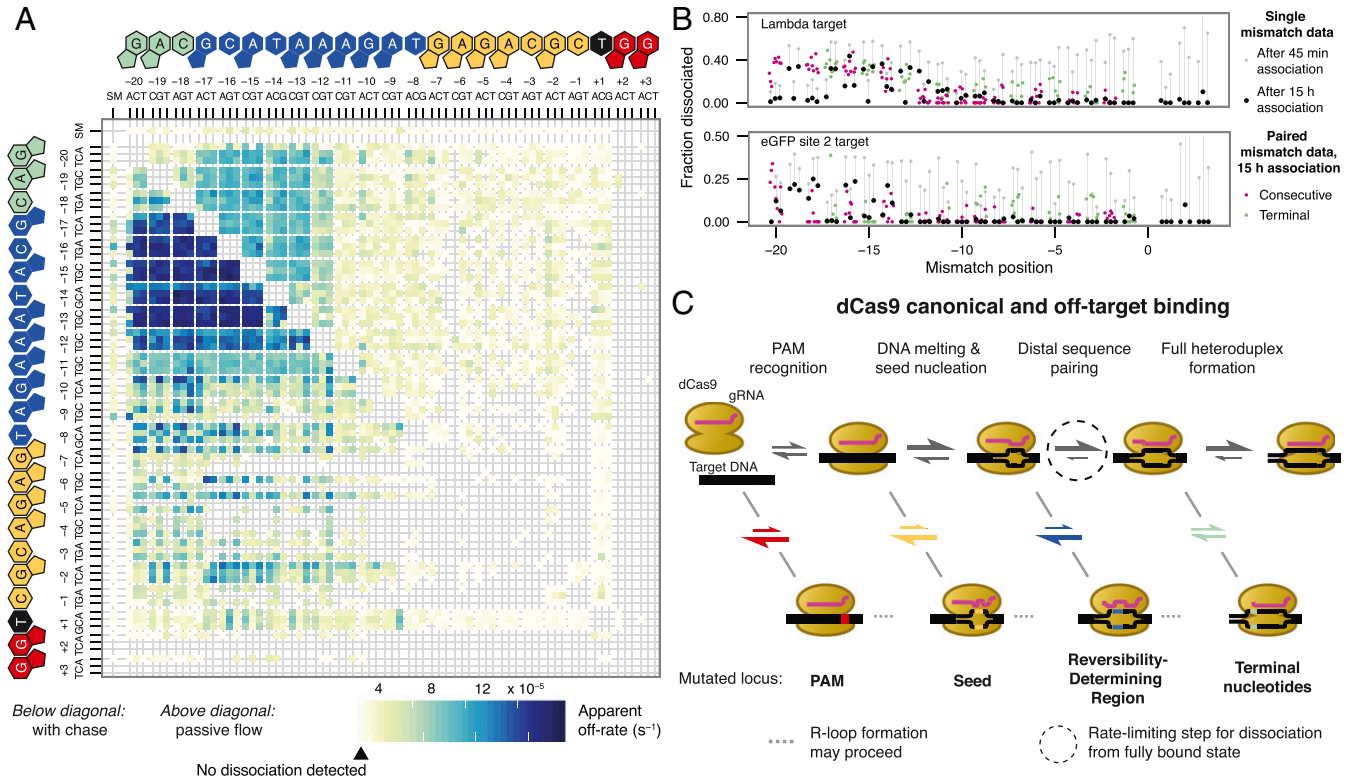

Fig. 3. Variation in dCas9 dissociation rates suggests a model of Cas9 binding behavior. (*A*) dCas9 apparent off-rates for single and double mutants, as in Fig. 2*B*. Apparent off-rates are systematically higher in the presence of an unlabeled competitor dsDNA that prevents rebinding (below diagonal) than without (above diagonal). (*B*) Massively parallel filter-binding results for reversible binding at 10 nM dCas9. At shorter time scales (gray points), dissociation is most rapid for seed mutants for both targets. At longer time scales (all other points), dissociation is almost exclusively controlled by PAM-distal mismatches, either in isolation (black, positions –19 to –14) or when paired with a second mutation (green points). Consecutive mismatches in the seed show minimal dissociation after long association (pink points). (*C*) Model diagram for R-loop formation in different off-target contexts. Rates for protospacer mismatches are color-coded by off-target partition as in *A*. PAM and PAM-proximal mismatches affect early steps in the Cas9 target identification procedure, whereas distal mismatches influence later steps. Dissociation rates in *A* are likely products of the kinetics of unwinding of the R loop across the RDR. Dissociation rates associated with transient binding, as with most PAM mutants that fail to form R-loop structures, do not appreciably bind and thus are not captured in the dissociation experiment.

approach, we estimated the bound fraction for select time points and generated binding curves for every species (Figs. S6–S8 and *SI Text*), including all single mismatches plus several double mismatches, for the original λ-target (HiTS-FLIP $R^2 = 0.746$ for single mismatches; Fig. S9*A*) and two eGFP-derived targets (eGFP site 1 and eGFP site 2) from a study of Cas9 off-target activity (*Materials and Methods*) (25). Importantly, and in contrast to HiTS-FLIP, these measurements cannot be biased by differential photobleaching across the course of the highly dynamic sgRNA strand invasion process we observe.

Consistent with the relative cleavage efficiency data (25), the eGFP site 1 target tolerated numerous mismatches, whereas the eGFP site 2 target proved selective for the on-target sequence (Fig. S9*B*). For the λ-target and the eGFP site 2 target, we confirmed that after 15 h of association, dCas9 binding could be at least partially reversed after 3 h (the eGFP site 1 target was not profiled owing to its slow binding kinetics). This dissociation was contingent on specific PAM-distal mismatches and generally not impacted by the presence of seed mismatches (Fig. 3*B*), thereby confirming the presence of an RDR region across gRNA. However, when 10 nM dCas9 was allowed to associate for only 45 min, the dissociation landscape was radically altered, with PAM disruptions and seed mismatches exhibiting equivalent or greater dissociation than PAM-distal mismatches. Furthermore, for the eGFP site libraries, double-mismatch off-targets diverged from their constitutive single-mismatch off-targets in unpredictable ways (Fig. S9*B*). These results speak to the dynamic and kinetically sensitive nature of the dCas9 strand invasion process. It is clear that although PAM and seed polymorphisms govern initial dCas9 binding kinetics, the contribution of the PAM-distal region cannot be ignored in a full accounting of Cas9 strand invasion. Thus, profiling diverse collections of targets by HiTS-FLIP or massively parallel filter-binding will be crucial for quantifying how base identities and mismatches along the length of the sgRNA modify the kinetics of sgRNA strand invasion across different sequence landscapes.

We also compared dCas9 binding data for the eGFP on-target sequences to in vivo measurements of cleavage efficiency. Although the cleavage data, our simulations, and a model of CRISPRi activity (15) all suggest that the eGFP site 1 target should be highly active, we found that dCas9 bound this target over 100-fold slower than λ-phage target. The cleavage data also suggest that seed mismatches reduce cutting efficiency beyond that anticipated from binding measurements (Fig. S9*C*). This finding is consistent with a Cas9 conformational gating mechanism (12) that enhances the specificity of cleavage over binding; even when binding is robust, cleavage may be impeded by mismatches.

Drawing from these observations, we propose a mechanism for Cas9 binding and dissociation (Fig. 3*C*). PAM mutations act to rapidly release diffusing Cas9 molecules postcollision, whereas seed mismatches impair target melting and nucleation. When DNA melting and seed hybridization is accomplished despite seed mismatches, heteroduplex formation can continue to completion, resulting in effectively irreversibly bound Cas9. Mismatches in the nucleotides adjacent to the proximal RDR modulate the energy barrier to dissociation such that heteroduplexes can be reversed on a shorter time scale, especially when multiple PAM-distal mismatches are present. Finally, mismatches in the terminal nucleotides of gRNA-template pairing have little effect in isolation but still destabilize the full heteroduplex and sensitize Cas9 to any additional mismatches in PAM-distal bases.

Our results reveal the complex effects of combinatorial DNA sequence perturbations on the binding behavior of dCas9 across multiple guide sequences and provide powerful tools to further study complex relationships between parameters of guide sequence, target sequence, binding time, and protein concentration, as they relate to both Cas9-mediated binding and cleavage. We identify altered dissociation kinetics as a functional consequence of PAM-distal mismatches in a set of nucleotides we term the reversibility-determining region, which presents across disparate guide sequences. Furthermore, we observe that modulation of Cas9 off-rate kinetics by targeting specific PAM-distal bases represents a potential area for tuning thermodynamic and kinetic behavior of CRISPR/Cas systems for maximal specificity and may already underlie alternate genome editing approaches including truncated gRNAs and modified Cas9 proteins (26–29). More broadly, these results highlight the challenge of predicting dCas9 off-target kinetics and underscore the need for higher resolution temporal data at off-target sites to develop accurate strand invasion models. Such models comprise a starting point for understanding how Cas9-intrinsic behavior is modulated by other factors such as local chromatin accessibility and superhelical density. We anticipate that our approach, together with other, complementary methodologies, promise to extend an avenue of molecular characterization—high-throughput biochemical profiling—that will facilitate functional dissection of novel nucleic acid-binding molecules, in addition to other members of the CRISPR/Cas family of enzymes, at an unprecedented scale.

Materials and Methods

dCas9 and sgRNA Preparation. dCas9 (the catalytically dead D10A/H840A mutant) was purified as described (20). The sgRNA (*SI Text*) was in vitro transcribed from the BamHI cleavage product of pSHS 256 (https://benchling.com/s/zmUR5HNi/edit) using T7 polymerase. The 3′ end of the sgRNA was extended to permit annealing of the Cy3 probe. Both the 3′ extension and hybridized Cy3 probe were loaded into dCas9 and tested on on-target DNA to ensure no defect in Cas9 binding resulted.

Association and Dissociation HiTS-FLIP Experiments. The 3′ end of the sgRNA was labeled before loading onto dCas9 with a Cy3-labeled oligo (*SI Text*) by incubating 4.95 μM sgRNA with 5 μM of the labeled oligo in hybridization buffer (20 mM Tris•HCl, pH 7.5, 100 mM KCl, 5 mM $MgCl_2$) for 5 min at 95 °C and then slowly cooling to room temperature. For each experiment (1 nM and 10 nM dCas9), the specified concentration of dCas9 was incubated with 50 nM labeled sgRNA at 37 °C for 25 min in binding buffer (20 mM Tris•HCl, pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 0.05 mg/mL heparin, 1 mM DTT, and 0.005% Tween 20) to load the sgRNA onto the dCas9. Each loaded dCas9–sgRNA preparation was placed on ice throughout the course of the experiment.

Before each association and dissociation experiment, any DNA hybridized to the DNA tethered on the flow cell surface in a previous experiment was removed with a 100 mM NaOH solution. Next, an Alexa 647-labeled oligo (*SI Text*) was annealed to a common sequence on the tethered DNA. dsDNA was generated by extending the annealed oligo with Klenow Fragment (3′→ 5′ exo-) (NEB) in buffer, per the manufacturer's recommendations, for 30 min at 37 °C.

HiTS-FLIP Data Processing. Raw images were processed using software previously described (1–3, 7, 16, 17). Time stamps were extracted from image file metadata to assign the exact time the data were recorded. Initial on-rates were calculated by performing linear regression on the quantified fluorescence values across all clusters, constraining the fit to go through the origin. To permit joint analysis of 1 and 10 nM datasets, linear regression was performed on target sequences quantified for both concentrations, and 10 nM slopes absent in the 1 nM dataset due to the limits of detection were inferred from the fit line (Fig. S1*A*).

Initial off-rates were also calculated by linear regression but without constraining the intercept. For both on- and off-rates, SEs and confidence intervals were calculated by bootstrapping the clusters used in linear regression 100 times.

See *SI Text* for extended methods. Fit values are available for both 10-nM (Datasets S1–S4) and 1-nM (Datasets S5–S6) data.

Radioactive Filter-Binding Experiments. DNA targets identical in sequence to the flow cell clusters were selected, using the most common barcode for each off-target. Six targets (on-target, −16G, −16T, −13C, −5T, and +3A) were ordered as gBlocks from IDT (*SI Text*), amplified by PCR, and gel purified. The dsDNA was then 5′ radiolabeled by incubating 150 nM dsDNA, 1x T4 PNK (NEB), 1x PNK buffer (NEB), and 1 μM [γ- 32P]-ATP (PerkinElmer) for 30 min at 37 °C followed by purification with a nucleotide removal kit (Qiagen). The sgRNA was hybridized to a labeled DNA oligo, as described above, and loaded onto the dCas9 by incubating 100 nM dCas9 and 125 nM sgRNA at 37 °C for 25 min and then at <4 °C.

Association rates were measured by incubating radiolabeled DNA targets with loaded dCas9 for different durations in a binding buffer identical to the flow cell experiments (see *Association and Dissociation HiTS-FLIP Experiments*). The total volume was 30 μL, and the concentrations were either 10 nM dCas9 and <240 pM DNA, or 1 nM dCas9 and <150 pM DNA. For dissociation measurements, 10 nM dCas9 and <240 pM target DNA were incubated for 2 h

(on-target, −16G, −16T, −13C, and +3A) or 5 h (−5T) followed by the addition of 6 μL nonradiolabeled cold competitor DNA in binding buffer (final concentration, 83 nM; see Table S1). The quenching step lasted for different durations, and both association and dissociation experiments were timed such that all conditions finished at nearly the same time. The samples were then applied to a 96-well Bio-Dot microfiltration blotting apparatus under low vacuum, passing through a nitrocellulose membrane (Amersham Hybond ECL, GE Healthcare Life Sciences), a nylon membrane (Biodyne B, 0.45 μM, Thermo Scientific), and a filter paper (GE Healthcare Life Sciences) that were all pre-equilibrated with binding buffer. The membranes were allowed to dry, transferred to a phosphor screen overnight, and then measured on a Typhoon imager (GE Healthcare Life Sciences). Images were quantified in TotalLab Quant v12.2, and the dCas9-bound DNA fraction was calculated as the signal from the nitrocellulose membrane divided by the total signal from the both the nitrocellulose and nylon membranes.

EMSA Experiments. Cy5-labeled DNA targets were generated by PCR. To measure association rates, DNA was incubated with loaded dCas9 for varying times followed by a quench step with a high concentration of unlabeled competitor on-target DNA. Sequences and binding buffer were identical to the filter binding experiments (see Table S1 for DNA sequences). Concentrations were 100 pM DNA, 1 nM dCas9, and 80 nM competitor for the on-target target and 400 pM DNA, 10 nM dCas9, and 100 nM competitor for the −5T off-target. Following the quench, samples were resolved by gel electrophoresis on a 10% native polyacrylamide gel (Mini-PROTEAN, Bio-Rad) in TBE running buffer (Bio-Rad) for 30–60 min at 120 V at 4 °C. Gels were imaged on a Typhoon imager (GE Healthcare Life Sciences) and quantified in TotalLab Quant v12.2.

Kinetic Monte Carlo Simulations. Simulations were carried out as in Josephs et al. (22), with additional parameterization. The full strand invasion of dCas9 was modeled as a series of 21 discrete states, where the first state was fully dissociated dCas9, the second state represented PAM binding, and the subsequent 19 states reflected successive strand invasion from 2 to 20 bp of RNA–DNA heteroduplex. The initial on-rate and the free energy of PAM binding were left as free parameters. Mismatches were modeled as increases in the free energy of states following the position of the mismatch, one value for transitions and one for transversions, which was supported by the data. Simulations were compared with data by assuming that HiTS-FLIP measurements corresponded to the fraction of DNA molecules in the bound states (states 2–21). One thousand kinetic Monte Carlo simulations were performed at a time to model clusters on the Illumina flow cell. Simulations were run until 30% (300) of the DNA molecules were in the bound states. Results were robust to choice of threshold. To enable a better fit of the model to the data, a free energy term representing protein conformational change was added at a state that was also set as a free parameter. The six parameters above were optimized by grid search across all single-mutant on-rates from the 1 nM dCas9 HiTS-FLIP experiment. The simulation script and accompanying information are available in Datasets S7–S9.

Massively Parallel Filter-Binding Experiments. Oligos corresponding to every individual position and select pairs of positions were column-synthesized (IDT) for three targets, the original lambda DNA and two eGFP targets. For each oligo's target positions, the three nonreference bases were mixed and incorporated. Oligos were pooled by target followed by PCR to generate dsDNA with extended sequencing adapters. Competitor DNA with the same on-target sequence but with alternate PCR adapters was similarly generated.

Experiments were carried out at 1 nM and 10 nM dCas9 in 450-μL reaction volumes containing 100 pM target pools and 10% excess sgRNA (EnGen sgRNA Synthesis Kit, NEB). dCas9 and targets were incubated at set time points before being loaded into a 1 mL syringe attached to a 0.45 μm pore size, 25 mm diameter nitrocellulose syringe filter (GVS). For dissociation experiments, at the end of the corresponding association experiment, dCas9 was quenched with 20 nM competitor DNA and allowed to sit for 3 h. Flow-through from each time point was amplified by PCR using unique barcodes and sequenced. Counts were normalized both to the starting time point (controlling for DNA input) and to nonbinding DNA in the experiment (representing 0% bound). For association experiments, binding curves were fit as single exponentials using the nls function in R. For dissociation curves, the normalized dissociation signal was compared with the corresponding association data point to calculate the fraction dissociated.

Sequence data are available at SRA accession no. SRP102425.

ACKNOWLEDGMENTS. We thank members of the W.J.G. laboratory for feedback on data visualization. This work was supported by grants from the Beckman Foundation, the Human Frontiers Science Program, and National Institutes of Health Grants 5R01GM111990, 1P50HG00773501, UM1HG009436, and 3P01GM066275 (to W.J.G.). W.J.G acknowledges support as a Chan-Zuckerberg Investigator. E.A.B. acknowledges support from NIH Training Grant 5T32HG000044-19. L.M.C. acknowledges support from NIH Training Grant T32GM067586 and the National Science Foundation graduate research fellowship program (NSF-GRFP). M.J.W. and E.A.B. also acknowledge support from the NSF-GRFP. J.A.D. acknowledges support from the National Science Foundation (Grant MCB-1244557 to J.A.D.). J.A.D. is an investigator of the Howard Hughes Medical Institute.

1. Anders C, Niewoehner O, Duerst A, Jinek M (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513:569–573.
2. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155:733–740.
3. Wang H, La Russa M, Qi LS (2016) CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem* 85:227–264.
4. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507:62–67.
5. Szczelkun MD, et al. (2014) Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci USA* 111:9798–9803.
6. Duan J, et al. (2014) Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Res* 24:1009–1012.
7. Fu BXH, Hansen LL, Artiles KL, Nonet ML, Fire AZ (2014) Landscape of target:guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Res* 42:13778–13787.
8. Pattanayak V, et al. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31:839–843.
9. Wu X, et al. (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* 32:670–676.
10. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* 32:677–683.
11. Hsu PD, et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31:827–832.
12. Sternberg SH, LaFrance B, Kaplan M, Doudna JA (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527:110–113.
13. Jiang F, et al. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 351:867–871.
14. O'Geen H, Henry IM, Bhakta MS, Meckler JF, Segal DJ (2015) A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res* 43:3389–3404.
15. Xu H, et al. (2015) Sequence determinants of improved CRISPR sgRNA design. *Genome Res* 25:1147–1157.
16. Nutiu R, et al. (2011) Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat Biotechnol* 29:659–664.
17. Buenrostro JD, et al. (2014) Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol* 32:562–568.
18. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31:233–239.
19. Zhang Y, et al. (2014) Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci Rep* 4:5405.
20. Jinek M, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821.
21. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32:279–284.
22. Josephs EA, et al. (2016) Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* 44:2474.
23. Singh D, Sternberg SH, Fei J, Doudna JA, Ha T (2016) Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun* 7:12778.
24. Farasat I, Salis HM (2016) A biophysical model of CRISPR/Cas9 activity for rational design of genome editing and gene regulation. *PLOS Comput Biol* 12:e1004724.
25. Fu Y, et al. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31:822–826.
26. Kleinstiver BP, et al. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529:490–495.
27. Oakes BL, et al. (2016) Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat Biotechnol* 34:646–651.
28. Slaymaker IM, et al. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* 351:84–88.
29. Nihongaki Y, Kawano F, Nakajima T, Sato M (2015) Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* 33:755–760.