

## Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules

Klein, Misha; Eslami Mosallam, Behrouz; Gonzalez Arroyo, D.W.; Depken, Martin

**DOI**

[10.1016/j.celrep.2018.01.045](https://doi.org/10.1016/j.celrep.2018.01.045)

**Publication date**

2018

**Document Version**

Final published version

**Published in**

Cell Reports

**Citation (APA)**

Klein, M., Eslami Mosallam, B., Gonzalez Arroyo, D. W., & Depken, M. (2018). Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. *Cell Reports*, 22(6), 1413-1423. <https://doi.org/10.1016/j.celrep.2018.01.045>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

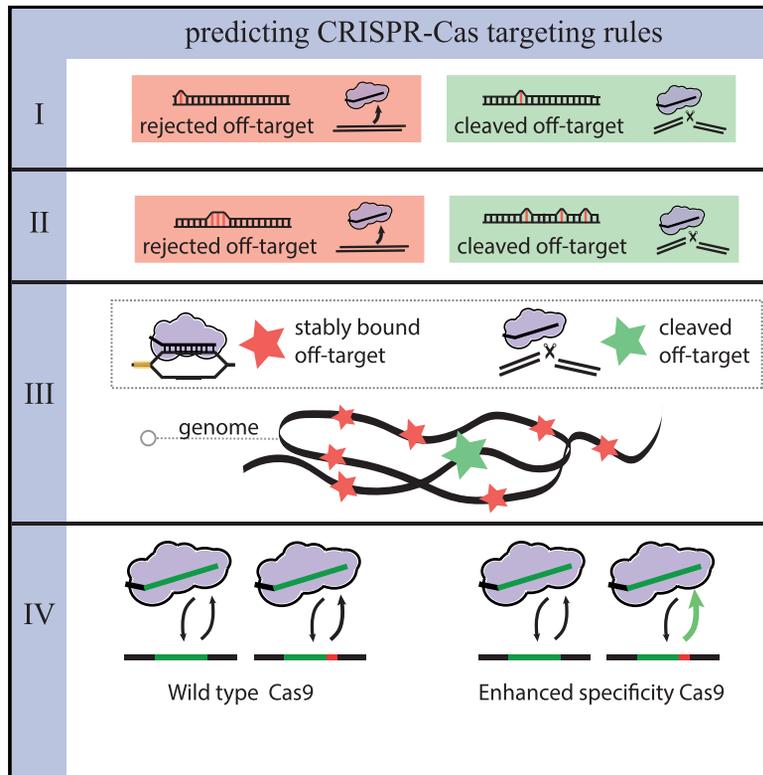
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

## Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules

### Graphical Abstract



### Authors

Misha Klein, Behrouz Eslami-Mossallam, Dylan Gonzalez Arroyo, Martin Depken

### Correspondence

s.m.depken@tudelft.nl

### In Brief

Predicting potential off targets forms an important part of any gene-editing technology intended for therapeutic use. By constructing a kinetic model for target recognition by RNA-guided nucleases, Klein et al. explain experimentally observed off-targeting rules as a direct consequence of guide-target hybridization dynamics.

### Highlights

- Physical model shows CRISPR/Argonaute off-targeting rules to be kinetic in origin
- Seed region and mismatch-pattern dependence is due to the kinetics of hybridization
- Binding is more promiscuous than cleavage due to kinetically stalled hybridization
- Engineered systems can increase specificity without losing on-target efficiency



# Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules

Misha Klein,<sup>1</sup> Behrouz Eslami-Mossallam,<sup>1</sup> Dylan Gonzalez Arroyo,<sup>1,2</sup> and Martin Depken<sup>1,3,\*</sup>

<sup>1</sup>Kavli Institute of NanoScience and Department of BioNanoScience, Delft University of Technology, Delft 2629HZ, the Netherlands

<sup>2</sup>Present address: Mathematical Institute, Leiden University, 2333CA Leiden, the Netherlands

<sup>3</sup>Lead Contact

\*Correspondence: [s.m.depken@tudelft.nl](mailto:s.m.depken@tudelft.nl)

<https://doi.org/10.1016/j.celrep.2018.01.045>

## SUMMARY

Due to their specificity, efficiency, and ease of programming, CRISPR-associated nucleases are popular tools for genome editing. On the genomic scale, these nucleases still show considerable off-target activity though, posing a serious obstacle to the development of therapies. Off targeting is often minimized by choosing especially high-specificity guide sequences, based on algorithms that codify empirically determined off-targeting rules. A lack of mechanistic understanding of these rules has so far necessitated their *ad hoc* implementation, likely contributing to the limited precision of present algorithms. To understand the targeting rules, we kinetically model the physics of guide-target hybrid formation. Using only four parameters, our model elucidates the kinetic origin of the experimentally observed off-targeting rules, thereby rationalizing the results from both binding and cleavage assays. We favorably compare our model to published data from CRISPR-Cas9, CRISPR-Cpf1, CRISPR-Cascade, as well as the human Argonaute 2 system.

## INTRODUCTION

RNA-guided nucleases (RGNs) target nucleic acid sequences based on complementarity to any guide RNA (gRNA) loaded into the complex. This versatility, together with the ability to design synthetic gRNAs complementary to any target of choice, holds great promise for gene-editing and gene-silencing applications (Cox et al., 2015; Tycko et al., 2016). Among the known RGNs, the Cas nucleases Cas9 (Cong et al., 2013; Gasiunas et al., 2012; Jinek et al., 2012; Mali et al., 2013) and Cpf1 (Zetsche et al., 2015) are of special interest, as they are comparatively simple single-subunit enzymes.

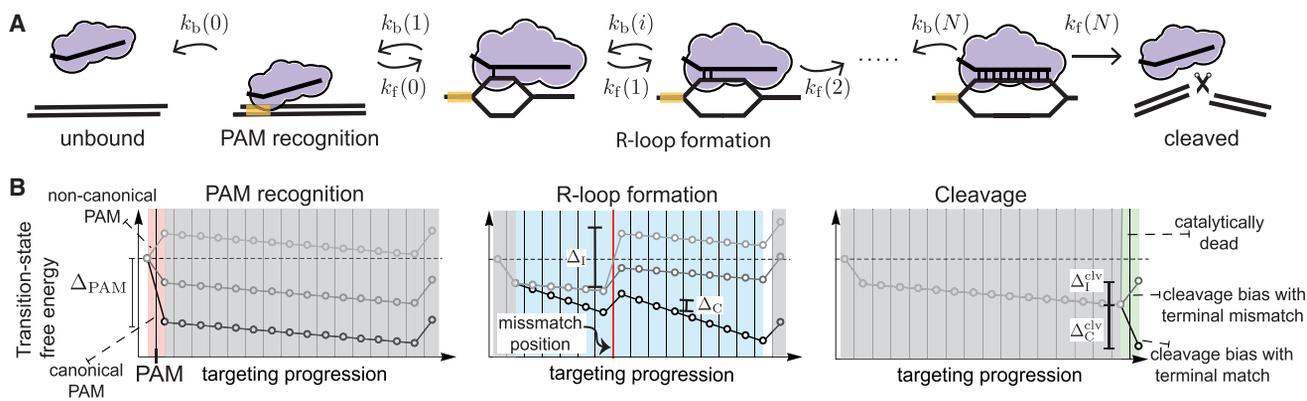
Cas nucleases originate from the CRISPR-Cas adaptive immune system, which many prokaryotes use to fight off foreign genetic elements. *In vivo*, the Cas protein (complex) is programmed by loading RNA transcribed from a CRISPR locus in the host genome. The transcribed sequence includes sections referred to as spacers, which were acquired during past encounters with foreign genetic elements (Wiedenheft et al., 2012).

Once programmed, the Cas nuclease is able to target and degrade genetic elements with the same sequence as the stored spacer and so offers protection against repeat invasions. An autoimmune response to sequences stored at the CRISPR locus is prevented through the additional requirement of a protein-mediated recognition of a short protospacer-adjacent motif (PAM) sequence present in the foreign genome, but not incorporated into the CRISPR locus with the spacer (Anders et al., 2014; Jinek et al., 2012).

As viruses evolve in response to the selective pressure induced by the CRISPR-Cas immune system, the host is in turn under pressure to attack slightly mutated target sequences in addition to the target. It is therefore not surprising that Cas nucleases exhibit considerable off-target activity on sequences similar to the intended target (Anderson et al., 2015; Fu et al., 2013, 2014a, 2016; Hsu et al., 2013; Kim et al., 2016; Kleinstiver et al., 2016a; Kuscu et al., 2014; O'Geen et al., 2015; Pattanayak et al., 2013; Wu et al., 2014). Such off targeting presents a severe problem for therapeutics, as DNA breaks introduced at the wrong site could lead to loss-of-function mutations in a well-functioning gene or the improper repair of a disease-causing gene (Cox et al., 2015).

To shed light on the determinants of off-target activity, a recent flurry of experiments has probed the level of binding and/or cleavage on mutated target sequences: high-throughput screens of large libraries of off targets (Doench et al., 2016; Fu et al., 2013, 2014a, 2016; Hsu et al., 2013; Pattanayak et al., 2013); genome-wide identification (Cameron et al., 2017; Frock et al., 2015; Kim et al., 2015; Kleinstiver et al., 2016a; Kuscu et al., 2014; Ran et al., 2015; Tsai et al., 2015, 2017; Wu et al., 2014); systematic biochemical studies (Anderson et al., 2015; Cong et al., 2013; Doench et al., 2016; Jinek et al., 2012; Kim et al., 2016; Kleinstiver et al., 2016a; Lin et al., 2014; Ran et al., 2015; Semenova et al., 2011); structural studies (Anders et al., 2014; Jiang et al., 2015, 2016; Jinek et al., 2014; Nishimasu et al., 2014; Xiao et al., 2017; Zhao et al., 2014); and single-molecule biophysical studies (Jo et al., 2015; Josephs et al., 2015; Rutkauskas et al., 2015; Salomon et al., 2015; Singh et al., 2016; Sternberg et al., 2014; Szczelkun et al., 2014), providing insights into the mechanics of targeting. To date, a number of rather peculiar targeting rules have been empirically established for Cas nucleases: (1) seed region: single mismatches within a PAM proximal seed region can completely disrupt interference (Künne et al., 2014; Semenova et al., 2011), whereas PAM distal mismatches have much less of an effect (Anderson et al., 2015; Cong et al., 2013; Doench et al., 2016; Fu et al., 2013, 2014a,





**Figure 1. Kinetic Model of RGN Target Recognition**

(A) The RGN initially binds its substrate at the PAM site, from which it can either unbind with rate  $k_b(0)$  or initiate R-loop formation with rate  $k_f(0)$ . A partially formed R-loop of length  $n$  grows to length  $n + 1$  with rate  $k_f(n)$  or shrinks to length  $n - 1$  with rate  $k_b(n)$ . Eventually, the RGN will either cleave its substrate with rate  $k_f(N)$  or reject the substrate and unbind with rate  $k_b(0)$ . In the special case of a RGN that does not utilize PAM binding, it is assumed to bind straight into the initial state of R-loop formation.

(B) The transition landscape of our minimal model. In the left panel, we illustrate a PAM bound enzyme kinetically biased toward R-loop formation by different amounts (black, gray, and light gray curves). The kinetic bias for the canonical PAM is shown as  $\Delta_{\text{PAM}}$ . In the middle panel, we illustrate two kinetic biases toward R-loop extension (black and gray curves), with the larger bias indicated as  $\Delta_{\text{C}}$ . In the same panel, we further illustrate two kinetic biases against R-loop extension (gray and light gray curves) at mismatches (red vertical lines), with the largest bias shown as  $\Delta_{\text{I}}$ . Once the complete R-loop is formed, the system is kinetically biased against cleavage by  $\Delta_{\text{C}}^{\text{clv}} = \Delta_{\text{clv}} \mp \Delta_{\text{C}}/l$ , as dictated by the nature of the terminal base pairing. See Figure S1 for complete energy landscapes.

2016; Hsu et al., 2013; Jinek et al., 2012; Kim et al., 2016; Kleinstiver et al., 2016a; Kuscu et al., 2014; O’Geen et al., 2015; Patanayak et al., 2013; Rutkauskas et al., 2015; Semenova et al., 2011; Sternberg et al., 2014; Szczelkun et al., 2014; Wu et al., 2014); (2) mismatch spread: when mismatches are outside the seed region, off targets with spread out mismatches are targeted most strongly (Boyle et al., 2017; Fu et al., 2013, 2014a; Hsu et al., 2013); (3) differential binding versus differential cleavage: binding is more tolerant to mismatches than cleavage (Bikard et al., 2013; Dahlman et al., 2015; Duan et al., 2014; Kuscu et al., 2014; O’Geen et al., 2015; Tsai et al., 2015; Wu et al., 2014); (4) specificity-efficiency decoupling: weakened protein-DNA interactions can improve target selectivity while still maintaining efficiency (Kleinstiver et al., 2015a, 2015b, 2016b; Slaymaker et al., 2016). Although these experimental observations have already aided the development of strategies to improve the specificity of the CRISPR-Cas9 system (Fu et al., 2014b; Kleinstiver et al., 2015a, 2016b; Ran et al., 2013; Slaymaker et al., 2016), an understanding of the mechanistic origin behind target selectivity is still lacking, and our ability to predict off targets remains limited (Cameron et al., 2017; Haeussler et al., 2016; Tsai et al., 2015; Tycko et al., 2016).

Current off-target prediction algorithms are often based on sequence alignment with the target and discard potential targets if they have more than some (user-defined) threshold number of mismatches (Bae et al., 2014; Haeussler et al., 2016; Heigwer et al., 2014; Labun et al., 2016). To recover the mismatch-position dependence observed as seed regions (rule [1]) and their cooperativity (rule [2]), such scoring schemes must be supplemented with *ad hoc* rules that penalize seed and closely spaced mismatches more than non-seed mismatches (Doench et al., 2016; Hsu et al., 2013). To move beyond *ad hoc* scoring schemes, we here use biophysical modeling to incorporate

knowledge of the underlying targeting process. With this aim, it would be attractive to assume that the binding dynamics has had time to equilibrate before DNA degradation (Farasat and Salis, 2016; Khorshid et al., 2013), as this would allow us to use simple binding/hybridization energetics to predict cleavage activity. Though attractive, this approach has recently been questioned by Bisaria et al. (2017) by noting that off rates are generally not found to be much faster than cleavage rates, as would be required for establishing a binding equilibrium before cleavage. In addition, the authors show how abandoning the equilibration assumption directly explains the specificity increase observed with shortened gRNA (Fu et al., 2014b).

Inspired by these observations, we go beyond binding energetics to build a biophysical model capturing the kinetics of guide-target hybrid formation. We show that the targeting rules (1)–(4) can be seen as simple consequences of kinetics. The targeting rules are captured by four parameters that pertain to transition barriers between metastable states of the nuclease-guide-target complex, and we translate these into four experimentally observable quantities: the length of the seed region; the width of the transition region from seed to non-seed; the maximum amount of cleavage on single-mismatch off targets; and the minimal distance between mismatches outside the seed region that allows for the cleavage of targets with multiple mismatches. By tying microscopic properties to biological and technological function, we here open the door to refined and rational reengineering of the CRISPR-Cas system to further its use in therapeutic applications.

Though we frame our considerations in terms of the well-studied and technologically important Cas9, our approach applies to any RGN that displays a progressive matching between guide and target before cleavage (Figure 1A). To demonstrate the generality and power of our approach, we present fits to targeting data from Argonaute 2 (hAgo2), as well as type I, II, and V CRISPR systems.

## RESULTS

At the start of target recognition, Cas nucleases bind to double-stranded DNA (dsDNA) from solution. The subsequent recognition of a PAM sequence triggers the DNA duplex to open up (Figure 1A), exposing the PAM proximal nucleotides to base pairing interactions with the guide (Anders et al., 2014; Jiang et al., 2016). From here, an R-loop is formed, expanding the guide-target hybrid in the PAM distal direction (Josephs et al., 2015; Rutkauskas et al., 2015; Semenova et al., 2011; Singh et al., 2016; Szczelkun et al., 2014; Xiao et al., 2017). If the target and guide reach (near-) full pairing, cleavage of the two DNA strands is triggered (Sternberg et al., 2015).

To establish the determinants of off- versus on-target cleavage, we construct a biophysical model of sequential target recognition in the unsaturated binding regime (Experimental Procedures). Using this model, we can calculate the rate of cleavage for off targets, given the guide. To incorporate the mechanics of hybrid formation, we envision the changing extension of the R-loop as a diffusion through a free-energy landscape, eventually ending in either unbinding from, or degradation of, the targeted sequence (Figures S1A and S1B). Our model is parameterized by the free energy of transition states surrounding the metastable states of PAM binding and the different progressions of R-loop formation (Experimental Procedures and Supplemental Experimental Procedures 1). When in a metastable state, the RGN will be biased toward transitioning to the neighboring state with the lowest intervening barrier. The differences in heights of the surrounding barriers thus encode the directions in which the system is most likely to progress, and we therefore refer to these differences as kinetic biases (Figure S1C). The balance between eventual unbinding or cleavage can be calculated with reference to kinetic biases alone and visualized by a “transition landscape” tracing out the transition states (Figures 1B and S1; Experimental Procedures). In such a landscape, the R-loop typically grows whenever the forward barrier is lower than the backward barrier; that is, whenever the transition landscape tilts downward. To facilitate the discussion of our exact results, we appropriate a rule of thumb from the limit of large biases (Experimental Procedures): after binding the PAM, Cas9 is most likely to unbind before cleavage if the highest barrier to cleavage is greater than the highest barrier to unbinding and vice versa (Figures S1A and S1B).

Though we treat the general scenario in the Experimental Procedures section, we here further limit ourselves to a minimal description with only four effective microscopic parameters, pertaining to the average kinetic bias for R-loop initiation after PAM binding ( $\Delta_{\text{PAM}}$ ), R-loop extension past a correctly matched ( $\Delta_{\text{C}}$ ) and mismatched ( $\Delta_{\text{I}}$ ) base pairs, and additional bias against cleavage once the R-loop is fully formed ( $\Delta_{\text{clv}}$ ; for definitions, see Figure 1B and Experimental Procedures). The parameter  $\Delta_{\text{clv}}$  is chosen such that the forward barrier after R-loop completion is independent of the nature of the terminal base (Experimental Procedures), setting the final bias against cleavage to  $\Delta_{\text{C/I}}^{\text{clv}} = \Delta_{\text{C/I}} \mp \Delta_{\text{clv}}$  (Figure 1B). Using this approach, we investigate to what extent our minimal model explains the four empirical targeting rules deduced from experiments.

### Rule (1): Seed Region

Following PAM binding, base pairing between guide and target is attempted (Figure 1B, middle panel). To establish whether the above-mentioned dependence of the cleavage propensity on the position of mismatches within the guide-target hybrid could originate from the kinetics of the targeting process, we calculate the relative cleavage probability on a sequence with a single mismatch at position  $n$  compared to the cleavage probability on the target sequence. In Supplemental Experimental Procedures 2, we show that this relative cleavage probability is in general sigmoidal

$$p_{\text{clv}}(n) = \frac{p_{\text{max}}}{1 + \exp[-(n - n_{\text{seed}})\Delta_{\text{C}}]}, \quad (\text{Equation 1})$$

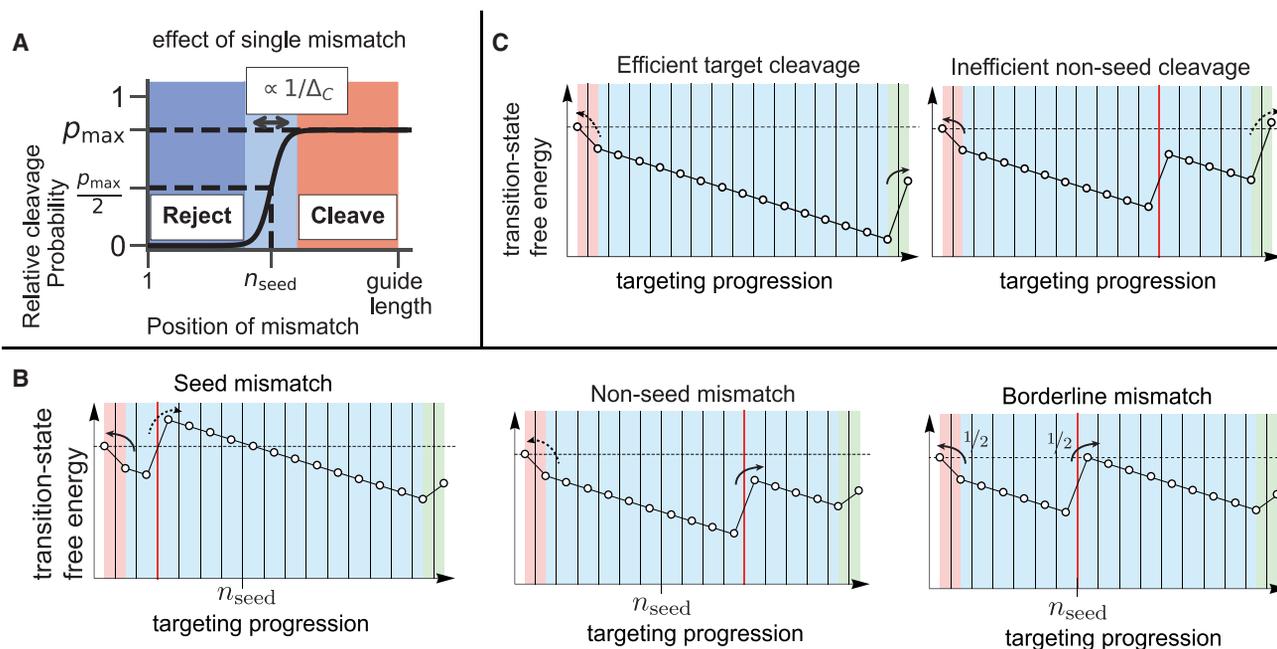
with  $n_{\text{seed}}$  giving the position where the cleavage probability is half that of its maximum  $p_{\text{max}}$  (Figure 2A) and the biases are measured in units of  $k_{\text{B}}T$ . We identify  $n_{\text{seed}}$  as the length of the kinetic seed region, beyond which a mismatch will no longer strongly suppress cleavage (Figure 2A). From Equation 1, we see that the width of the transition from seed to non-seed region directly reports on the (average) correct-match bias ( $\Delta_{\text{C}}$ ; Supplemental Experimental Procedures 2), becoming narrower as the bias increases (Figures 2A and S2A).

The emergence of a seed-like region can be understood from considering the rule of thumb that the fate of the enzyme is dictated by the largest barrier: when a mismatch is placed at  $n_{\text{seed}}$  (Figure 2B, right panel), the highest barrier to cleavage matches the barrier toward unbinding, guaranteeing a near equal probability for cleavage and unbinding. Placing the mismatch closer to the PAM increases the highest barrier toward cleavage (compare highest node to first node in Figure 2B, left panel), increasing the probability of rejecting such off targets. Moving the mismatch distally from the PAM will gradually lower the highest barrier toward cleavage (Figure 2B, middle panel), increasing the probability of accepting such off targets. Though the exact form of the parameters of Equation 1 are given in the Supplemental Information, it is informative to here give the kinetic seed length in the large-bias limit (Experimental Procedures; Supplemental Experimental Procedures 2),

$$n_{\text{seed}} \approx \frac{\Delta_{\text{I}} - \Delta_{\text{PAM}}}{\Delta_{\text{C}}} + 1. \quad (\text{Equation 2})$$

From this, we see that PAM bias and the base pairing biases all contribute to setting the extent of the seed region (Figures 2A and S2B). Weakening the PAM or correct-match bias extends the seed region, whereas weakening the bias for incorrect matches shrinks it.

After PAM recognition and R-loop formation, cleavage completes a successful targeting process (Figure 1B, right panel). Tuning the final transition state allows us to toggle between different regimes of minimal single-mutation specificity. Targets with a PAM distal mismatch get cleaved with near unity probability ( $p_{\text{max}} \approx 1$ ) only if all transition states toward cleavage (including the cleavage step) lie well below the transition state to unbinding (Figures 2C, left panel, and S2C). For slow enough enzymatic activity, the final barrier toward cleavage might not go



**Figure 2. Rule (1)—Seed Region**

(A) The relative-to-WT cleavage probability of a target with a single mismatch. Our model predicts a sigmoidal curve, with maximum off-target activity  $p_{\max}$ , seed length  $n_{\text{seed}}$ , and width of the seed to non-seed transition  $\sim 1/\Delta_C$ . See Figure S2 for parametric sweeps.

(B) Transition landscapes illustrating that the placement of a single mismatch (fltr: before, exactly at, beyond the seed's border) influences the cleavage probability.

(C) Increasing the kinetic bias against cleavage can suppress cleavage of off targets with a PAM-distal mismatch (compare right panel to right panel in B) while still maintaining a high on-target activity (left panel).

far below the barrier to unbinding, limiting the maximal cleavage compared to the perfect match ( $p_{\max} < 1$ ; Figure 2C, right panel). Consequently, there can be a noticeable effect on off-target activity also when the mismatch is outside the seed region (Figures 2A and S2C). Reversing this logic implies that  $p_{\max} < 1$  is indicative of a relatively slow cleavage reaction.

### Rule (2): Mismatch Spread

Considering more complex mismatch patterns, we start by addressing all possible dinucleotide mismatches (Figures 3A and 4B). The overall cleavage and binding patterns obtained strongly resemble experimental observations (Boyle et al., 2017; Fu et al., 2013, 2014a; Hsu et al., 2013). As expected, placing both mismatches within the seed disrupts cleavage (Figure 3A). However, moving the mismatches outside the seed does not necessarily restore cleavage activity. With the first mismatch outside the seed region, a second mismatch only abolishes cleavage if it is situated before  $n_{\text{seed}} + n_{\text{pair}}$  (Figure 3B), with

$$n_{\text{pair}} \approx \frac{\Delta_1}{\Delta_C} + 1, \quad (\text{Equation 3})$$

in the large-bias limit (Experimental Procedures; Supplemental Experimental Procedures 2). The general form of the two-mismatch seed region is shown in Figure 3B, where only off targets in the red region lead to cleavage. In the dark blue region, off targets are rejected due to the first mismatch, and

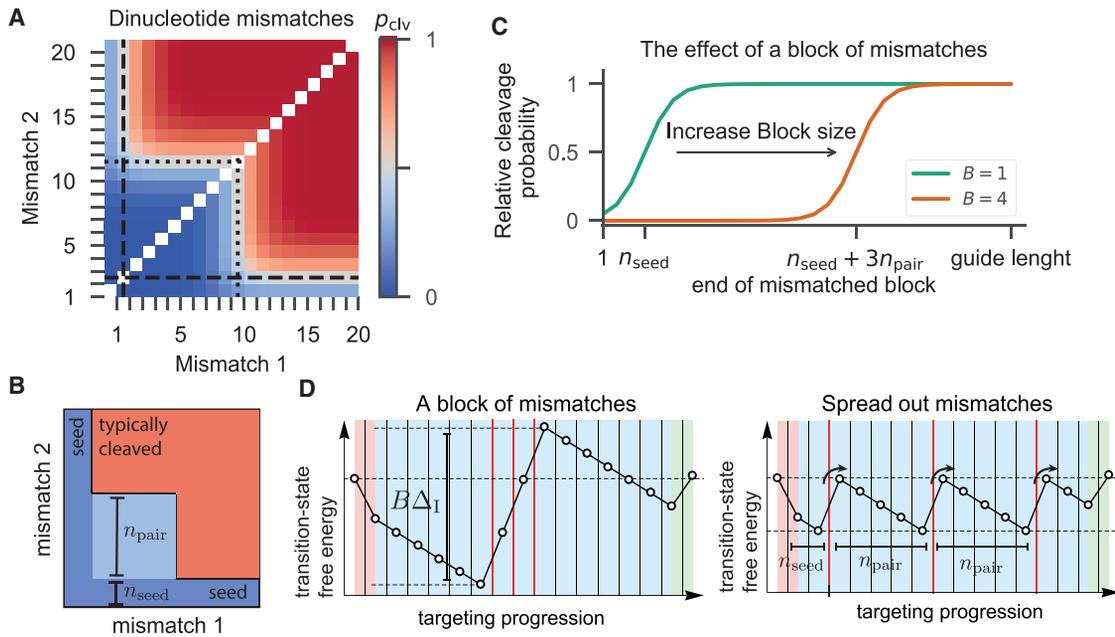
in the light blue region, they are rejected due to the second mismatch.

The single- and double-mismatch rules can now be unified and generalized (Figure 3D, right panel) into a single rule for any number of mismatches: off targets will typically be rejected if any mismatch, say the  $m$ :th mismatch, is positioned closer than  $n_{\text{seed}} + (m - 1)n_{\text{pair}}$  to the PAM.

Note that, for systems not requiring PAM recognition,  $n_{\text{seed}} = n_{\text{pair}}$ . The above rule also captures the extreme case of a “block” of  $B$  consecutive mismatches, which has also been investigated experimentally (Fu et al., 2013; Hsu et al., 2013; Jo et al., 2015; Singh et al., 2016). Placing such a block effectively acts as placing a single mismatch with the bias  $\Delta_1$  scaled by the size of the block (Figures 3C, 3D, and S3), giving a block-seed region of size  $n_{\text{seed}} + (B - 1)n_{\text{pair}}$ . Hence, a block of mismatches leads to less off targeting compared to spread out mismatches (Figures 3C and 3D). Given the correspondence of these predictions with literature, our model seems to automatically and correctly capture the non-multiplicative cleavage suppression by multiple mismatches, in sharp contrast to the *ad hoc* scoring schemes employed in current prediction algorithms (Haeussler et al., 2016).

### Rule (3): Differential Binding versus Differential Cleavage

Catalytically dead systems (for example, dCas9 [Jinek et al., 2012] or Cascade without Cas3) bind strongly to sites that their



**Figure 3. Rule (2) – Mismatch Spread**

- (A) The relative-to-WT probability to cleave a target with two mismatches for a system with  $\Delta_{\text{PAM}} = 3.5 k_B T$ ,  $\Delta_I = 4 k_B T$ ,  $\Delta_C = 1 k_B T$ , and  $\Delta_{\text{civ}} = 1 k_B T$ . The seed length  $n_{\text{seed}}$  is indicated with dashed lines, and  $n_{\text{seed}} + n_{\text{pair}}$  is indicated with dotted lines.
- (B) Schematic of the probability to cleave a target with two mismatches. The target is typically rejected in both blue regions and rejected in the red.
- (C) Probability to cleave a target with a block of  $B$  mismatches as a function of the location of the last mismatch. Also see Figure S3.
- (D) Spreading out blocked mismatches (left panel) around their average position significantly lessens the barrier to cleavage (right panel).

catalytically active counterparts do not cleave (Boyle et al., 2017; Duan et al., 2014; Kuscus et al., 2014; O’Geen et al., 2015; Tsai et al., 2015; Wu et al., 2014). In order to explain this effect, we model inactive systems with a very large cleavage barrier (gray in Figure 1B, right panel; Experimental Procedures). In agreement with experimental observations (Semenova et al., 2011), our model predicts a dissociation constant that is higher when a mismatch is placed closer to the PAM (Figures 4B and S4).

Similar to the cleavage efficiency in the kinetic regime, the dissociation constant takes on a sigmoidal form (Supplemental Experimental Procedures 3). However, this equilibrated seed length (Figure S4) is different from the kinetic counterpart discussed above (Supplemental Experimental Procedures 3). Binding affinities therefore do not need to report on cleavage activity. In general, the gene-editing (Cas9) and gene-silencing (dCas9) capabilities should be seen as two related but separate properties of the RGN. For example, the most stable configuration of the RGN on the mismatched target shown in the right panel of Figure 4A is a bound state with a partial R-loop (purple). However, a catalytic active variant will most likely eventually reject this off target (gray), as the barrier to cleavage is higher than to unbinding. Hence, even though cleavage sites are strong binders (Figure 4A, left panel), observing a long binding time on an off-target site should not be taken to imply that this site will also display substantial off-target cleavage (Figure 4A, right panel).

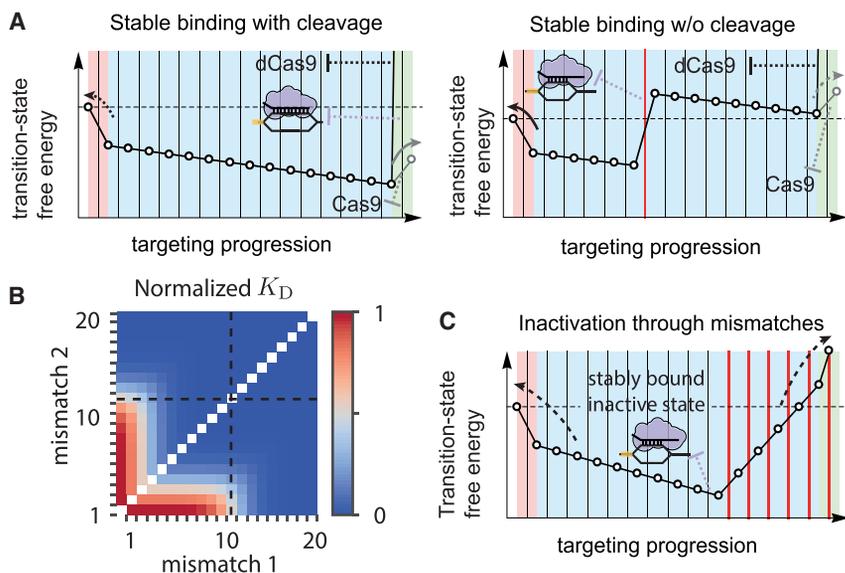
Active Cas9 variants also strongly bind to sites they are incapable of cleaving, especially those containing multiple

PAM-distal mismatches (Bikard et al., 2013; Dahlman et al., 2015). Such a series of mismatches induces a large barrier that opposes, and thereby likely prevents, cleavage (Figure 4C). Although we are yet to extract temporal information from our model, it is clear that the state right before the first mismatch (purple) might be stably bound over experimental timescales.

#### Rule (4): Specificity-Efficiency Decoupling PAM Recognition

R-loop formation is preceded by PAM recognition. Although PAM mismatches often completely abolish interactions with the target (Hsu et al., 2013; Semenova et al., 2011; Sternberg et al., 2014), binding to (and interference with) targets flanked by non-canonical PAM sequences has been observed (Leenay et al., 2016). Because PAM mismatches will shift the entire free-energy landscape upward from the bound PAM state onward (Figure 1B, left panel), these always increase the highest barrier to cleavage, thereby reducing the cleavage efficiency on any sequence. For increased specificity, we thus need the cleavage efficiency for the off targets to be reduced more than for the target itself.

Protein reengineering approaches most easily affect the overall strength of PAM interactions, influencing the kinetic bias for both the correct PAM ( $\Delta_{\text{PAM}}$ ) and incorrect PAM ( $\Delta'_{\text{PAM}}$ ). In Figure 5A, we show the relative cleavage efficiency between protospacers flanked by incorrect and correct PAMs, and in Figure 5B, we show the cleavage efficiency



**Figure 4. Rule (3)—Differential Binding versus Differential Cleavage**

(A) Transition landscapes illustrating the difference between active Cas9 (gray curves) and dCas9 (black curves) when encountering either the cognate site (left panel) or an off target with a mismatch within the seed (right panel).

(B) The dissociation constant for targets with any combination of two mismatches for energetic biases  $\delta_{PAM} = 7.5 k_B T$ ,  $\delta_C = 1 k_B T$ , and  $\delta_i = 8 k_B T$ . The end of the seed region is indicated with dashed lines. See Figure S4 for single-mismatched off targets.

(C) Transition landscape for an active Cas9 bound to an off-target possessing a block of mismatches placed at the PAM-distal end. Even though cleavage is unlikely, unbinding takes a long time.

with the correct PAM—both as functions of the average kinetic bias ( $(\Delta_{PAM} + \Delta'_{PAM})/2$ ) and the kinetic bias difference ( $\Delta_{PAM} - \Delta'_{PAM}$ ). As long as the system operates in region A (Figure 5A), it is possible to increase the specificity by lowering the average kinetic bias toward R-loop initiation without changing the kinetic-bias difference (Supplemental Experimental Procedures 2). Outside this region, the system either does not discriminate between PAMs (region C) or is insensitive to the average kinetic bias (region B). Interestingly, it is only in region B that lowering the average bias also leads to a lower on-target efficiency (Figure 5B), and consequently, the wild-type (WT) nuclease can only be improved if brought into region A, where it is possible to engineer specificity increases with limited costs in the on-target efficiency. The transition-state diagrams shown in the top panel of Figure 5C show a situation where the barrier to cleavage (rightmost node) is substantially lower than the barrier to unbinding (leftmost node) for two different PAM biases, both resulting in near unit probability to cleave, and corresponding to region C in Figure 5A. Re-engineering the nuclease to have overall weaker PAM binding (Figure 5C, bottom panel) brings the system into region B, where the cleavage probability for the correct PAM (black) remains close to unity, whereas the probability of cleaving with the incorrect PAM (gray) is drastically lowered. The above scenario might explain how PAM mutant Cas9s are able to outperform their WT counterparts (Kleinstiver et al., 2015a, 2015b) on specificity without significant loss in efficiency.

### Sequence Recognition

Another approach to gain specificity is to weaken the protein-DNA interactions affecting the bias for R-loop extension (Kleinstiver et al., 2016b; Slaymaker et al., 2016). In Figure 5D, we show how engineering the PAM-bound nuclease in this way, inducing a lower gain for correct base pairing, can render previously cleaved off targets (gray line in top panel) rejected (gray line in bottom panel). We further see how we can retain on-target

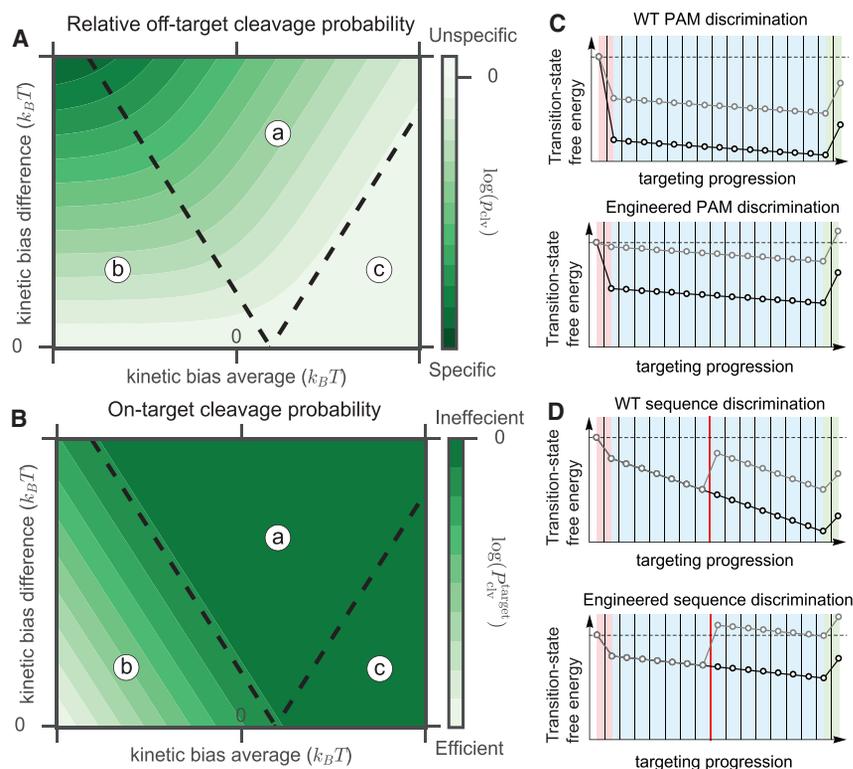
specificity if the highest transition state toward cleavage (rightmost node of black line) remains substantially lower than the transition state to unbinding (leftmost

node of black line). The above scenario might explain how mutant Cas9s could have an extended seed while having negligible reduction in on-target cleavage activity (Kleinstiver et al., 2016b; Slaymaker et al., 2016).

### Comparison to Experimental Data for a Broad Class of RNA-Guided Nucleases

To test our model, we acquired published datasets from different RGN systems and fitted Equation 1 to singly mismatched targets and blocks of mismatches. The fitted sigmoid has only three effective fit parameters ( $p_{max}$  or  $K_{D,max}$ ,  $n_{seed}$ , and  $\Delta_C$ ), so we can unfortunately not get an estimate for all microscopic parameters from the single-mismatch datasets (Supplemental Experimental Procedures 2 and 3)—for this, further experiments are required, as outlined below. Details of the fitting procedure and additional fits can be found in Supplemental Experimental Procedures 4.

Perhaps the best-characterized RGN system is the type II CRISPR-associated *Streptococcus Pyogenes* Cas9 (spCas9). Among the systems we estimate parameters for, the dataset from Anderson et al. (2015) traces out the sigmoidal trend particularly well. For this dataset, we fit out a kinetic seed of about 11.3 [11.0,11.4] nt (68% confidence interval between 11.0 and 11.4) and an average bias per correct base pair of about  $\Delta_C = 1.70$  [1.15, 4.0]  $k_B T$  (Figure 6A). This positive bias indicates that association with the RGN stabilizes the hybrid, which is in line with recent studies demonstrating that the protein has a strong contribution to the energetics of the resulting bound complex (Kleinstiver et al., 2016b; Salomon et al., 2015; Slaymaker et al., 2016). The relative cleavage probability levels off around  $p_{max} = 0.74$  [0.72,0.77], indicating that spCas9 retains some specificity even against errors that are outside the seed. We performed additional fits using a second target site from the dataset of Anderson et al. and data obtained from Pattanayak et al. (2013), which produced results that do not significantly differ (Figures S5A–S5C).



**Figure 5. Rule (4) – Specificity-Efficiency Decoupling**

(A) The cleavage probability on a fully cognate target but with a mismatched PAM, compared to one with the correct PAM, as a function of the average and difference in the kinetic bias of the correct and incorrect PAM. Independent of the sequence following both PAMs, one can identify three regimes (Supplemental Information). Only in regime a is the RGN's specificity improved through a decrease in the average PAM bias toward R-loop initiation. (B) On-target efficiency for the target with the correct PAM. In regime a, the RGN's efficiency is not compromised, allowing for simultaneous maintenance of on-target efficiency and specificity. (C) The cognate protospacer flanked by either a canonical PAM (black) or incorrect PAM sequence (gray) is bound by a WT (top panel) or engineered RGN (panel). (D) A matched/mismatched protospacer (black/gray) bound by WT/engineered RGN (top/bottom panel).

ingly, the energetic gain for a match again suggests a large contribution of the protein to the overall stability (energetic bias  $\delta_C \approx 3.7 k_B T$ ). Structurally, subunits of the Cascade complex bind to nucleotides 6, 12, 18, 24, and 30 of the

Recently, the type V CRISPR-associated enzyme Cpf1 has been characterized as another single-subunit RGN (Zetsche et al., 2015). Kleinstiver et al. (2016a) performed *in vivo* (human cells) cleavage assays using two different variants named LbCpf1 (Figure 6B) and AsCpf1 (Figure 6C). Both variants exhibit quantitatively similar off targeting, both with seed lengths ( $n_{\text{seed}} \approx 18.9$  [18.5, 19.2] nt for LbCpf1 versus 19.1 [18.7, 19.3] nt for AsCpf1) and maximum off-target activity ( $p_{\text{max}} \approx 0.84$  [0.66, 1.0] nt for LbCpf1 versus 0.83 [0.71, 1.0] for AsCpf1). Compared to spCas9, the Cpf1s are much more specific as the seed region is significantly larger.

Single-molecule fluorescence resonance energy transfer (FRET) experiments done with hAgo2 (Jo et al., 2015) utilized targets with two consecutive mismatches. Given that hybrid formation is not preceded by PAM recognition, and that consecutive mismatches impose a combined penalty (Figures 3C and 3D), the estimated half-saturation point is approximately twice the kinetic seed length for a single mismatch ( $n_{\text{seed}} \approx 10$  [9.5, 9.9] nt). The hAgo2 data thus suggest a similar seed length as that of spCas9 (Figure 6D), consistent with the observation that hAgo2 and spCas9 display structural similarities within their respective seed regions (Jiang et al., 2016). Our fits further reveal that hAgo2 likely exhibits a substantially lower gain per correctly formed base pair ( $\Delta_C \approx 0.77$  [0.63, 0.92]  $k_B T$ ).

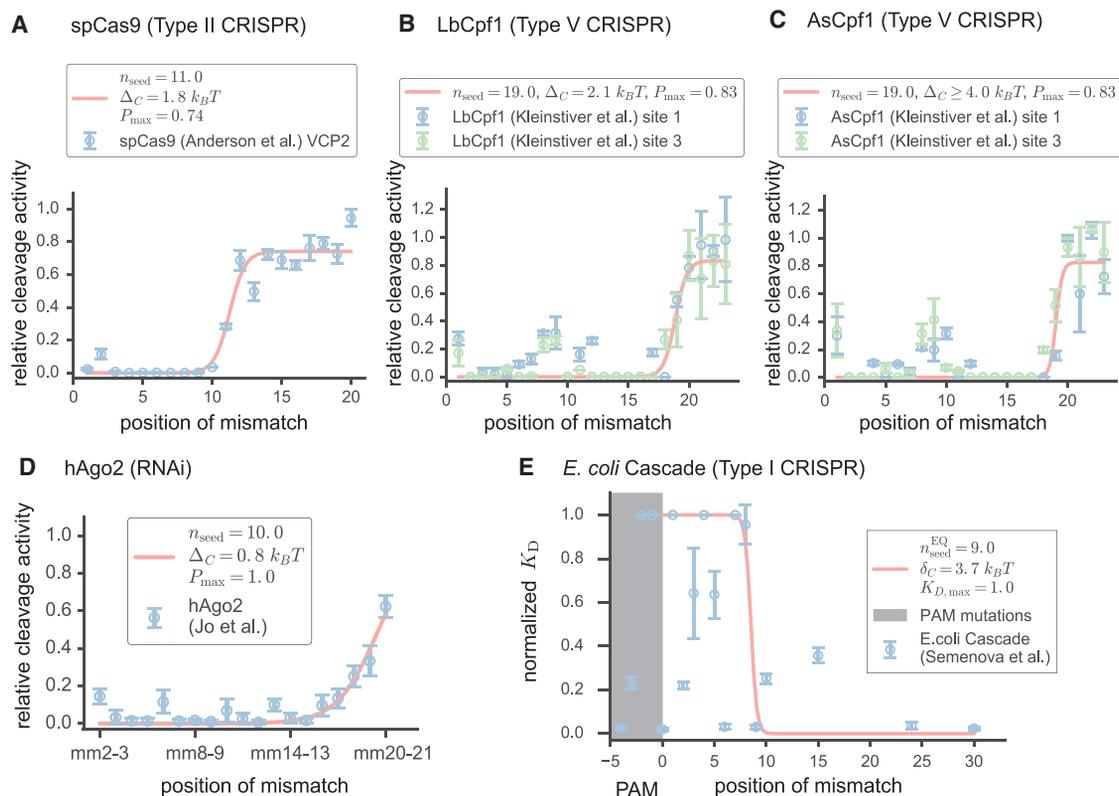
Unlike the aforementioned RGNs, the type I CRISPR uses a multi-subunit protein complex, termed Cascade, to target invaders (Brouns et al., 2008). Semanova et al. (2011) measured the dissociation constant *in vitro* of the *E. coli* subtype I-E Cascade. Fitting their data, we find that mismatches within the first 9 nt of the guide lead to rapid rejection (Figure 6E). Interest-

ingly, the energetic gain for a match again suggests a large contribution of the protein to the overall stability (energetic bias  $\delta_C \approx 3.7 k_B T$ ). Structurally, subunits of the Cascade complex bind to nucleotides 6, 12, 18, 24, and 30 of the guide (Zhao et al., 2014). To model this property, we assume that incorporating matches or mismatches at the Cascade-guide binding positions does not affect affinity. Including this effect mainly reduced the estimated energetic gain for matches ( $\delta_C \approx 1.9 k_B T$ ; Supplemental Experimental Procedures 4; Figure S5D), a value more in line to those obtained for the other CRISPR systems.

## DISCUSSION

We have presented a general description of target recognition by RGNs with a progressive matching between guide and target (Figure 1A), applicable to both CRISPR and Argonaute systems. In its simplest form, our model contains only two parameters to describe the R-loop formation process: an average kinetic bias toward incorporation beyond a match ( $\Delta_C$ ) and an average kinetic bias against extending the R-loop beyond a mismatch ( $\Delta_I$ ; Figure 1B, middle panel). Despite the simplifications going into this minimal model, we can qualitatively understand the targeting rules for these RGNs as resulting from kinetics, as illustrated graphically for seed region (Figure 2B), mismatch spread (Figure 3D), the poor match between cleavage propensity and binding propensity (Figure 4A), and the specificity-efficiency decoupling (Figures 5C and 5D). Based on our model, we have been able to establish a general targeting rule: off targets will typically be rejected if any mismatch, say the  $m$ :th mismatch, is positioned closer than  $n_{\text{seed}} + (m - 1)n_{\text{pair}}$  to the PAM.

Although Figure 6 shows that our model can already describe experimental data from various RGNs, the number of microscopic



**Figure 6. Comparison to Experimental Data**

(A–E) Fit of sigmoid (Equation 1) to experimental data from (A) spCas9 (Anderson et al., 2015), (B) LbCpf1 (Kleinstiver et al., 2016a), (C) AsCpf1 (Kleinstiver et al., 2016a), (D) human Argonaute 2 (Jo et al., 2015), and (E) *E. coli* Cascade complex (Semenova et al., 2011). Values reported in (A)–(D) correspond to the median of 1,000 bootstrap replicates, and the confidence intervals in the text correspond to 68%. See Figure S5 for additional fits. All experimental data shown corresponds to mean  $\pm$  SD.

parameters in the physical model ( $\Delta_{\text{PAM}}$ ,  $\Delta_C$ ,  $\Delta_I$ , and  $\Delta_{\text{clv}}$ ; Figure 1B) exceeds the number of fit parameters available from single-mismatch experiments ( $\Delta_C$ ,  $p_{\text{max}}$ , and  $n_{\text{seed}}$ ). It is therefore not possible to determine all the microscopic parameters from single-mismatch experiments alone. However, Figure 3B shows that, with two mismatches, we could also fit out  $n_{\text{pair}}$  and so determine all the microscopic parameters. It should be possible to directly extract all four microscopic parameters once such extended datasets become available.

One should recognize that our minimal model does not capture all the physics of the targeting process. Nucleic acid interactions are explicitly sequence dependent, RGNs are known to undergo conformational changes prior to cleavage (Klein et al., 2017; Sternberg et al., 2015; Xue et al., 2016), and the  $\Delta_C$  we fit out in Figure 6 technically only reports the matching bias at the end of the seed, allowing for variable biases along the R-loop. Although these are all topics that need to be explored for future improved quantitative predictions, such extensions are not needed to explain the observed targeting rules and will not qualitatively alter the trends predicted by our model. An exception might be the data from Cpf1 (Figures 6B and 6C), because it shows an increased tolerance to mismatches of nucleotides 1, 2, 8, and 9 compared to our minimal model, with a second independent study showing the same behavior (Kim

et al., 2016). Similarly, deviations from the sigmoidal trend are observed for Cascade (Figure 6E). Such features could be explained either through a sequence or position dependence of the kinetic biases.

In conclusion, our model is capable of explaining the observed off-targeting rules of CRISPR and Argonaute systems in simple kinetic terms. After having established the general utility of this approach, the next step will be to move beyond our minimal model and gradually allow for conformational control and sequence effects by letting our parameters depend on the nature of matches/mismatches as well as their positions. Fitting such a generalized model against training data would likely improve on present target prediction algorithms by limiting overfitting, as it captures the basic targeting rules deduced from experiments while using only a minimal set of physically meaningful parameters.

## EXPERIMENTAL PROCEDURES

### A General Model for RGNs with Progressive R-Loop Formation followed by Cleavage

Given the observed dependence of cleavage activity on Cas9 concentration (Cameron et al., 2017; Fu et al., 2013; Kuscu et al., 2014; O'Geen et al., 2015; Pattanayak et al., 2013), we here limit ourselves to the regime where nuclease concentrations are low enough that all binding sites are unsaturated.

The unsaturated regime is also the regime with the highest specificity and should therefore be of particular interest in gene-editing applications.

We define the cleavage efficiency  $P_{\text{cliv}}(s|g)$  as the fraction of binding events to sequence  $s$  that result in cleavage, given the RGN is loaded with guide sequence  $g$ . If we in the unsaturated regime assume the binding rate to be independent of sequence, we can express the relative rate of non-target versus target cleavage as

$$P_{\text{cliv}}(s|g) = \frac{P_{\text{cliv}}(s|g)}{P_{\text{cliv}}(g|g)} \quad (\text{Equation 4})$$

This relative efficiency is a direct measure of *specificity*, approaching unity for non-specific targeting ( $P_{\text{cliv}}(s|g) \approx P_{\text{cliv}}(g|g)$ ) and zero for specific targeting ( $P_{\text{cliv}}(s|g) \ll P_{\text{cliv}}(g|g)$ ).

In our model, we denote the PAM-bound state as 0 and the subsequent R-loop states by the number of base pairs that are formed in the hybrid. Each of the states  $n = 1, \dots, N$  are taken to transition to state  $n - 1/n + 1$  with backward/forward hopping rate  $k_b(n)/k_f(n)$  (Figure 1A). The ratio between forward and backward rates sets the relative probability of going forward and backward from any state and can be parametrized in terms of  $\Delta(n)$ , the difference in the free-energy barrier between going backward and forward from state  $n$  (Figure S1A),

$$\frac{k_f(n)}{k_b(n)} = e^{\Delta(n)}. \quad (\text{Equation 5})$$

Here, we measure energy in units of  $k_B T$  for notational convenience, and we will refer to  $\Delta(n)$  as the bias toward cleavage. The model (Figure 1A) is known as a birth-death process (Nowak, 2006), and the cleavage efficiency is given by the expression (Supplemental Experimental Procedures 1)

$$P_{\text{cliv}}(s|g) = \frac{1}{1 + \sum_{n=0}^{N-1} e^{-\Delta T(n)}}, \quad \Delta T(n) = \sum_{m=0}^n \Delta(m). \quad (\text{Equation 6})$$

Here,  $\Delta T(n)$  represents the free-energy difference between the transition state to solution and the forward transition state from position  $n$  (Figures S1A–S1C). For systems like hAgo2, there is no initial PAM binding (Bartel, 2009; Klein et al., 2017), and the sums in Equation 6 should omit the PAM state ( $n, m = 0$ ).

### Building Intuition by Using the Transition Landscape (Large Bias Limit)

Though we will use the exact results of Equation 7 for all calculations, it is useful to build intuition for the system by considering the case of large biases. In this limit, the term (say  $n = n^*$ ) with the highest transition state dominates the sum in Equations 6 and 7 (Figures S1A and S1B), and the cleavage efficiency can be approximated as

$$P_{\text{cliv}}(s|g) \approx \frac{1}{1 + e^{-\Delta T(n^*)}}. \quad (\text{Equation 7})$$

Based on this, we deduce the rule of thumb that cleavage dominates ( $P_{\text{cliv}} > 1/2$ ) if the first state of the transition landscape is the highest ( $\Delta T(n^*) > 0$ ; Figure S1A). Conversely, a potential target is likely rejected ( $P_{\text{cliv}} < 1/2$ ) if any of the other transition states lies above the first ( $\Delta T(n^*) < 0$ ; Figure S1B).

### A Minimal Model for RGNs with Progressive R-Loop Formation followed by Cleavage

Given that the defining feature of RGNs is their ability to target any sequence, we expect the major targeting mechanisms to depend more strongly on mismatch position than on the precise nature of the mismatches. With this in mind, we consider a sequence-independent model with the aim of finding a description that captures the gross, sequence-averaged features with a minimal number of parameters.

Focusing first on how PAM binding effects the system (Figure 1B, left panel), we see that  $\Delta(0) = \Delta_{\text{PAM}}$  controls the kinetic bias between initiating R-loop for-

mation and unbinding. A canonical PAM (black) promotes R-loop initiation, whereas a non-canonical PAM lessens (darker gray) or reverses (lighter gray) the bias toward R-loop formation. Note that PAM-independent systems omit this initial step.

Turning to the bias of R-loop progression, we represent the guide-target hybrid as a sequence of matches ( $C$ , correct base pairing) and mismatches ( $I$ , incorrect base pairing). Defining the average kinetic bias *toward/against* extending the R-loop by one *correct/incorrect* base pair as  $\Delta_C/\Delta_I$  (Figure 1B, middle panel), we take  $\Delta(n) = \Delta_C$  or  $\Delta(n) = -\Delta_I$ , depending on whether the base pairing is correct or incorrect (Supplemental Experimental Procedures 2). In the middle panel of Figure 1B, we show a transition landscape with moderate gains for correct base pairings and moderate costs for incorrect base pairings (dark gray). The black transition landscape corresponds to an increased gain for matches, whereas the light gray corresponds to an increased penalty for mismatches.

Lastly, considering the bias between cleavage and unwinding of the R-loop, we assume that an incorrect base pair at the terminal position adds the same change in bias as it did in the interior of the R-loop. Therefore, introducing the cleavage bias  $\Delta_{\text{cliv}}$ , we take  $\Delta(N) = \Delta_C^{\text{cliv}}$  for a correct match and  $\Delta(N) = -\Delta_I^{\text{cliv}}$  for a mismatch, with  $\Delta_C^{\text{cliv}} = \Delta_{C/I} \mp \Delta_{\text{cliv}}$  as bias *against* cleavage from the fully hybridized state (Figure 1B, right panel). In the right panel of Figure 1B, we show examples where the terminal bias  $\Delta_{C/I}^{\text{cliv}}$  corresponds to a terminal match (black), terminal mismatch (dark gray), and for a catalytically dead nuclease (light gray).

### Dissociation Constant for Catalytically Dead Nucleases

Apart from examining cleavage propensity, many experiments have focused on the binding of catalytically dead Cas9 (dCas9) or other catalytically dead RGNs (Boyle et al., 2017; Josephs et al., 2015; Kucsu et al., 2014; O'Geen et al., 2015; Ran et al., 2015; Semenova et al., 2011; Wu et al., 2014). To be able to relate pure binding experiments to cleavage experiments, we also calculate the dissociation constant  $K_D$  for our minimal model when describing a catalytically dead system ( $\Delta_{\text{cliv}} \approx \infty$ ; Figure S1D) through

$$P_{\text{bound}} = \frac{[\text{RGN}]}{[\text{RGN}] + K_D}. \quad (\text{Equation 8})$$

Here,  $P_{\text{bound}}$  equals the probability to bind a substrate in any of the ( $M$ ) possible R-loop configurations and follows from Equation 7 (Supplemental Experimental Procedures 3). Further,  $[\text{RGN}]$  denotes the concentration of effector complex. Differences in stability of the bound states now parameterize our model (Figure S1D).

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.01.045>.

### ACKNOWLEDGMENTS

First and foremost, M.D. would like to thank Ralph Seidel for introducing him to the problem of sequential target recognition during a visit to Delft. We would also like to thank Aafke van den Berg, Michiel Bongaerts, and Kristian Blom for fruitful discussion regarding the theoretical modeling. We further acknowledge discussions we have had with all members of the CRISPR/microRNA community in Delft: Viktorija Globytė, Luuk Loeff, Tao Ju (Thijs) Cui, Stanley Chandradoss, Sungchul Kim, Seung Hwan Lee, Chirlmin Joo, Rebecca McKenzie, Jochem Vink, Sebastian Kieper, and Stan Brouns. M.K. also thanks Iason Katechis for introducing him to the Python fitting library (lmfit) and Rebecca, Thijs, and Chirlmin for carefully reading and commenting on the manuscript. We gratefully thank Emily Anderson, Benjamin Kleinstiver, and Keith Joung; Sungchul Hohng, Soochul Shin, and Myung Hyun Jo; Ekaterina Semenova; and Vikram Pattanayak for sharing their data and answering all our questions. This work (M.K.) was supported by the Netherlands Organisation for Scientific Research, as part of the Frontiers in Nanoscience program.

M.D. acknowledges financial support from a TU Delft startup grant. This work (B.E.-M.) forms part of the research program “Crowd management: the physics of genome processing in complex environments,” which is (partly) financed by the Netherlands Organisation for Scientific Research.

#### AUTHOR CONTRIBUTIONS

M.K. and M.D. designed the research. M.K., B.E.-M., and D.G.A. performed the research and, together with M.D., interpreted the data. M.K., M.D., and B.E.-M. wrote the manuscript.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 7, 2017

Revised: December 7, 2017

Accepted: January 17, 2018

Published: February 6, 2018

#### REFERENCES

- Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* *513*, 569–573.
- Anderson, E.M., Haupt, A., Schiel, J.A., Chou, E., Machado, H.B., Strezoska, Z., Lenger, S., McClelland, S., Birmingham, A., Vermeulen, A., and Smith, A.V. (2015). Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. *J. Biotechnol.* *211*, 56–65.
- Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* *30*, 1473–1475.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* *136*, 215–233.
- Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* *41*, 7429–7437.
- Bisaria, N., Jarmoskaite, I., and Herschlag, D. (2017). Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-based genome editing. *Cell Syst.* *4*, 21–29.
- Boyle, E.A., Andreasson, J.O.L., Chircus, L.M., Sternberg, S.H., Wu, M.J., Guegler, C.K., Doudna, J.A., and Greenleaf, W.J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc. Natl. Acad. Sci. USA* *114*, 5461–5466.
- 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.
- Cameron, P., Fuller, C.K., Donohoue, P.D., Jones, B.N., Thompson, M.S., Carter, M.M., Gradia, S., Vidal, B., Garner, E., Slorach, E.M., et al. (2017). Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods* *14*, 600–606.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* *339*, 819–823.
- Cox, D.B.T., Platt, R.J., and Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. *Nat. Med.* *21*, 121–131.
- Dahlman, J.E., Abudayyeh, O.O., Joung, J., Gootenberg, J.S., Zhang, F., and Konermann, S. (2015). Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* *33*, 1159–1161.
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* *34*, 184–191.
- Duan, J., Lu, G., Xie, Z., Lou, M., Luo, J., Guo, L., and Zhang, Y. (2014). Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Res.* *24*, 1009–1012.
- Farasat, I., and Salis, H.M. (2016). A biophysical model of CRISPR/Cas9 activity for rational design of genome editing and gene regulation. *PLoS Comput. Biol.* *12*, e1004724.
- Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.-J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* *33*, 179–186.
- Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* *31*, 822–826.
- Fu, B.X.H., Hansen, L.L., Artilles, K.L., Nonet, M.L., and Fire, A.Z. (2014a). Landscape of target:guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Res.* *42*, 13778–13787.
- Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014b). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* *32*, 279–284.
- Fu, B.X.H., St Onge, R.P., Fire, A.Z., and Smith, J.D. (2016). Distinct patterns of Cas9 mismatch tolerance *in vitro* and *in vivo*. *Nucleic Acids Res.* *44*, 5365–5377.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* *109*, E2579–E2586.
- Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* *17*, 148.
- Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nat. Methods* *11*, 122–123.
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* *31*, 827–832.
- Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). Structural biology. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* *348*, 1477–1481.
- Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* *351*, 867–871.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* *337*, 816–821.
- Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* *343*, 1247997.
- Jo, M.H., Shin, S., Jung, S.R., Kim, E., Song, J.J., and Hohng, S. (2015). Human Argonaute 2 has diverse reaction pathways on target RNAs. *Mol. Cell* *59*, 117–124.
- Josephs, E.A., Kocak, D.D., Fitzgibbon, C.J., McMenemy, J., Gersbach, C.A., and Marszalek, P.E. (2015). Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res.* *43*, 8924–8941.
- Khorshid, M., Hausser, J., Zavalan, M., and van Nimwegen, E. (2013). A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. *Nat. Methods* *10*, 253–255.
- Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.I., and Kim, J.S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* *12*, 237–243.
- Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H., and Kim, J.S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* *34*, 863–868.

- Klein, M., Chandradoss, S.D., Depken, M., and Joo, C. (2017). Why Argonaute is needed to make microRNA target search fast and reliable. *Semin. Cell Dev. Biol.* 65, 20–28.
- Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P.W., Li, Z., Peterson, R.T., Yeh, J.-R., et al. (2015a). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485.
- Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z., and Joung, J.K. (2015b). Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* 33, 1293–1298.
- Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J., and Joung, J.K. (2016a). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* 34, 869–874.
- Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016b). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495.
- Künne, T., Swarts, D.C., and Brouns, S.J.J. (2014). Planting the seed: target recognition of short guide RNAs. *Trends Microbiol.* 22, 74–83.
- Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* 32, 677–683.
- Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B., and Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* 44 (W1), W272–W276.
- Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gooma, 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.
- Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., and Bao, G. (2014). CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* 42, 7473–7485.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.
- Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949.
- Nowak, M.A. (2006). *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press).
- O’Geen, H., Henry, I.M., Bhakta, M.S., Meckler, J.F., and Segal, D.J. (2015). A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* 43, 3389–3404.
- Pattanayak, V., Lin, S., Guillinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 31, 839–843.
- Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389.
- Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186–191.
- Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V., and Seidel, R. (2015). Directional R-loop formation by the CRISPR-cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.* 10, 1534–1543.
- Salomon, W.E., Jolly, S.M., Moore, M.J., Zamore, P.D., and Serebrov, V. (2015). Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. *Cell* 162, 84–95.
- Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. USA* 108, 10098–10103.
- Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A., and Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* 7, 12778.
- Slymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88.
- 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.
- Sternberg, S.H., LaFrance, B., Kaplan, M., and Doudna, J.A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527, 110–113.
- 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.
- Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197.
- Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* 14, 607–614.
- Tycko, J., Myer, V.E., and Hsu, P.D. (2016). Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell* 63, 355–370.
- Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338.
- Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 32, 670–676.
- 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 hand-over mechanisms in type I CRISPR-Cas system. *Cell* 170, 48–60.e11.
- 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.
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771.
- Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* 515, 147–150.