

CRISPR-Cas9 Activities with Truncated 16-Nucleotide RNA Guides Are Tuned by Target Duplex Stability Beyond the RNA/DNA Hybrid

Yue Li, Brendon H. Cooper, Yukang Liu, Difei Wu, Xiaojun Zhang, Remo Rohs, and Peter Z. Qin*



Cite This: *Biochemistry* 2023, 62, 2541–2548



Read Online

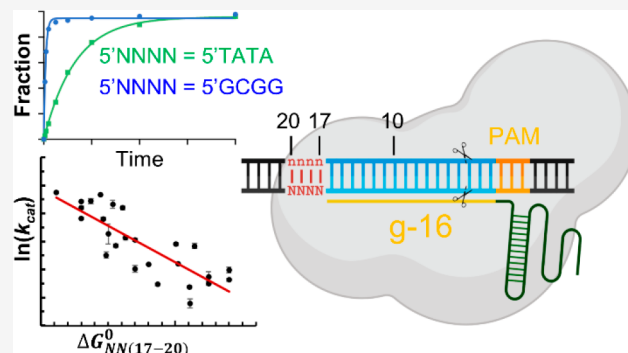
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: CRISPR-Cas9 has been adapted as a readily programmable genome manipulation agent, and continuing technological advances rely on an in-depth mechanistic understanding of Cas9 target discrimination. Cas9 interrogates a target by unwinding the DNA duplex to form an R-loop, where the RNA guide hybridizes with one of the DNA strands. It has been shown that RNA guides shorter than the normal length of 20-nucleotide (-nt) support Cas9 cleavage activity by enabling partial unwinding beyond the RNA/DNA hybrid. To investigate whether DNA segment beyond the RNA/DNA hybrid can impact Cas9 target discrimination with truncated guides, Cas9 double-stranded DNA cleavage rates (k_{cat}) were measured with 16-nt guides on targets with varying sequences at +17 to +20 positions distal to the protospacer-adjacent-motif (PAM). The data reveal a log-linear inverse correlation between k_{cat} and the PAM+(17–20) DNA duplex dissociation free energy ($\Delta G_{\text{NN}(17-20)}^0$), with sequences having smaller $\Delta G_{\text{NN}(17-20)}^0$ showing faster cleavage and a higher degree of unwinding. The results indicate that, with a 16-nt guide, “peripheral” DNA sequences beyond the RNA/DNA hybrid contribute to target discrimination by tuning the cleavage reaction transition state through the modulation of PAM-distal unwinding. The finding provides mechanistic insights for the further development of strategies that use RNA guide truncation to enhance Cas9 specificity.



CRISPR-Cas9 has been a revolutionary tool for genome engineering and manipulation, and its power stems from mechanistic understanding on how Cas9 specifically acquires a target within a complex genome:^{1–4} (i) base-pairing between a segment (the “guide” or “guide sequence”) of the single-guide RNA (sgRNA) and a segment of DNA designated as the protospacer and (ii) protein–DNA interactions with a short protospacer-adjacent-motif (PAM). However, Cas9 has been shown to bind and cleave genomic targets containing mismatches with the RNA guide, resulting in undesired off-target effects that are detrimental for applications.^{5–7} Advances in understanding Cas9 targeting mechanisms have led to protein and sgRNA engineering that greatly enhances specificity^{8–13} and remain the key for further development of Cas9-based applications.

Studies have revealed that Cas9 discriminates a target through unwinding of the DNA duplex to form a three-stranded R-loop, in which the DNA target-strand (TS) forms a hybrid with the RNA guide.^{14–16} While the guide of an sgRNA is generally 20-nucleotide (nt) long, formation of a stable RNA/DNA hybrid at the PAM-proximal 8–12 base-pair (bp) segment is sufficient to attain complete binding affinity by the Cas9 ribonucleoprotein complex.^{17,18} Subsequently, the structure and dynamics of the R-loop at the PAM-distal segment coordinates Cas9 nuclease conformational changes, which serve as checkpoint(s) in preventing cleavage of

incorrect targets.^{16,19–22} It has also been shown that RNA guides shorter than 20-nt can support Cas9 cleavage activities.^{21,23–25} Compared to the 20-nt guide, using truncated guides leads to a shorter RNA/DNA hybrid with a smaller number of RNA/DNA pairings, which would likely amplify the role of RNA-guide/DNA-target mismatch(es), thus enhancing Cas9 specificity.^{23,24}

Recently, we reported that, for RNA guides that are longer than 15-nt, the Cas9 ribonucleoprotein complex partially unwinds PAM-distal DNA beyond the RNA/DNA hybrid, and the rate of double-stranded DNA cleavage correlates with the degree of PAM-distal unwinding.²⁵ This would imply that, with truncated guides, intrinsic properties of the DNA protospacer beyond the RNA/DNA hybrid could modulate Cas9 cleavage rates by tuning the PAM-distal unwinding. Specifically, we hypothesize that the duplex dissociation free energy, which measures the stand-alone DNA duplex stability, can serve as an indicator of the ability of DNA to unwind upon interacting

Received: May 12, 2023

Revised: July 24, 2023

Published: August 8, 2023



with Cas9, and it would thus correlate with the observed Cas9 double-strand cleavage rate.

In this study, we tested this hypothesis by examining cleavage activities of *Streptococcus pyogenes* Cas9 (referred to as “Cas9”) ribonucleoprotein complexes containing 16-nt RNA guides on DNA substrates with varying PAM+(17–20) sequences. The measured cleavage rates at saturation (k_{cat}) show a good log–linear inverse correlation with $\Delta G_{\text{NN}(17-20)}^0$, the PAM+(17–20) segment DNA dissociation free energy. Sequences with smaller $\Delta G_{\text{NN}(17-20)}^0$ (less stable) were shown to unwind to a larger degree by a 2-amino-purine (2AP) fluorescence assay and indeed yielded larger k_{cat} , i.e., faster cleavage. In addition, k_{cat} was found to be most impacted by protospacer base-pairings at PAM+17 and +18. The data demonstrate that with a 16-nt guide, “peripheral” DNA sequences beyond the RNA/DNA hybrid impact target discrimination by tuning the transition state through modulation of PAM-distal unwinding. The finding provides a mechanistic basis for further development of strategies that use RNA guide truncation to reduce “off-target” effects.

MATERIALS AND METHODS

Protein Expression and Purification. Plasmids encoding *Streptococcus pyogenes* Cas9 (pMJ806) and catalytically inactive dCas9 (pMJ841, containing D10A/H840A mutations) were obtained from Addgene (<http://www.addgene.org/>), and the proteins were expressed and purified following previously described procedures.^{25,26} Purified proteins were stored at -80°C in a “Storage Buffer” with 20 mM Tris, pH 7.5, 250 mM NaCl, 5% glycerol, 5 mM MgCl_2 , and 0.5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP).

sgRNA Preparation. sgRNAs with the same 80-nt core but varying in guide sequence and length were synthesized by T7 in vitro transcription. The double-stranded DNA templates for transcription were prepared by overlapping PCR as previously described.²⁵ Transcriptions and sgRNA purifications were carried out following previously reported procedures.²⁶ Purified sgRNAs were dissolved in ME buffer (10 mM MOPS, pH 6.5, and 1 mM EDTA) and stored at -20°C .

DNA Cleavage Assays. Detailed information for Cas9 target sequences is presented in Supporting Information (SI, sect. S1.1). The Cas9 target was embedded in either a *ScaI* linearized pUC19 plasmid (Figure S1, Table S1) or a fragment (~ 700 bp) obtained by PCR amplification of exon 3 of the human EMX1 gene. Control experiments showed that differences in the peripheral DNA beyond the 20-nt protospacer had no impact on the measured cleavage rates (SI, sect. S1.2). For a given Cas9 cleavage reaction, the Cas9:sgRNA ribonucleoprotein (RNP) complex at a desired concentration was assembled with a protein:sgRNA molar ratio of 1:1.2 by incubation in the reaction buffer (20 mM Tris, pH 7.5, 100 mM KCl, 5 mM MgCl_2) for 10 min at 37°C . A given DNA substrate was then added to a final concentration of 5 nM. The reaction was allowed to proceed at 37°C for the desired time and then stopped by addition of denaturing loading dye equal to 1/5th of the reaction sample volume (New England BioLabs cat. no. B7024S supplemented with 0.48% SDS and 60 mM EDTA). The stopped reactions were loaded on a 1% agarose ethidium bromide gel to resolve the uncleaved precursors and cleaved products. Gels were imaged using a Gel Doc XR+ Gel Documentation System (Bio-Rad). DNA bands were selected and quantified by ImageLab 6.0.1, with background correction done by using baseline adjust-

ment. The fraction of products, f_{pro} , at each time point of a reaction was calculated as

$$f_{\text{pro}} = \frac{I_{\text{C1}} + I_{\text{C2}}}{I_{\text{U}} + I_{\text{C1}} + I_{\text{C2}}} \quad (1)$$

with I_{U} being the uncleaved “precursor” band intensity and I_{C1} and I_{C2} being intensities of cleaved products 1 (top band) and 2 (bottom band), respectively. It has been established that the active Cas9 concentrations used in this study were at saturating conditions.²⁵ The reaction rate constant, which represents k_{cat} , was determined by fitting the time dependence of f_{pro} to a single-exponential using Origin 2018:

$$f_{\text{pro}} = a \cdot (1 - e^{-k_{\text{cat}} \cdot t}) \quad (2)$$

with a being the active fraction of the target DNA. Multiple measurements were carried out to obtain the average and standard deviation for each reported k_{cat} value.

2-Amino-Purine (2AP) Fluorescence Assay for Assessing Cas9-Induced DNA Unwinding. DNA strands (unmodified and 2AP-substituted, SI, sect. S3, Table S3) were obtained by solid-phase chemical synthesis from a commercial source (Integrated DNA Technologies, Coralville, IA). To form a target DNA duplex (unmodified or with a 2AP label), the desired target-strand (TS) and nontarget-strand (NTS) were mixed in a 1:1 molar ratio in an “Annealing Buffer” (50 mM Tris, pH 7.5, 100 mM NaCl) and then incubated at room temperature overnight. The mixture was purified by size exclusion chromatography (SEC) using a Superdex 200 increase 10/300 GL column (GE Healthcare) in Annealing Buffer. The purified DNA duplexes were stored at -20°C for future use.

The catalytically inactive dCas9/sgRNA mixture was first incubated at room temperature for 10 min, followed by addition of the desired DNA duplex in Reaction Buffer, with dCas9:sgRNA:DNA in a molar ratio of 1:1.25:1.25. The ternary mixture was then incubated at 37°C for 30 min and purified using SEC in a buffer composed of 20 mM Tris, pH 7.5, 5% glycerol, 5 mM MgCl_2 , and 100 mM KCl. Homogenous complex fractions were pooled (≈ 2 mL) and used directly for measurements.

For a given 2AP-containing sample, fluorescence emission was measured using a Fluorolog Spectrofluorometer (Horiba Jobin Yvon, Edison, NJ) with excitation set at 320 nm to minimize dCas9 emission interference.²⁵ Absorbance was obtained immediately after the fluorescence measurement on a LAMBDA UV/vis/NIR Spectrophotometer (PerkinElmer). The normalized 2AP quantum yield of a tested sample with respect to the corresponding DNA duplex, $\text{ratio}(\varphi)$, was obtained using the concentration-normalized background-corrected 2AP emission (F/A) as described in our previous study.²⁵

$$\text{ratio}(\varphi) = \frac{\varphi}{\varphi_{\text{duplex}}} = \frac{\varepsilon}{\varepsilon_{\text{duplex}}} \cdot \frac{(F_{370}/A_{260})}{(F_{370}/A_{260})_{\text{duplex}}} \quad (3)$$

where ε and $\varepsilon_{\text{duplex}}$ are the extinction coefficients at 260 nm for the sample and the duplex, respectively; F_{370} is the background-corrected 2AP emission of the sample at 370 nm; A_{260} is the absorbance of the same sample at 260 nm; $(F_{370}/A_{260})_{\text{duplex}}$ is the F/A value for the corresponding duplex.

Multiple Linear Regression Analysis. To dissect the correlation between Cas9 cleavage rates and the intrinsic

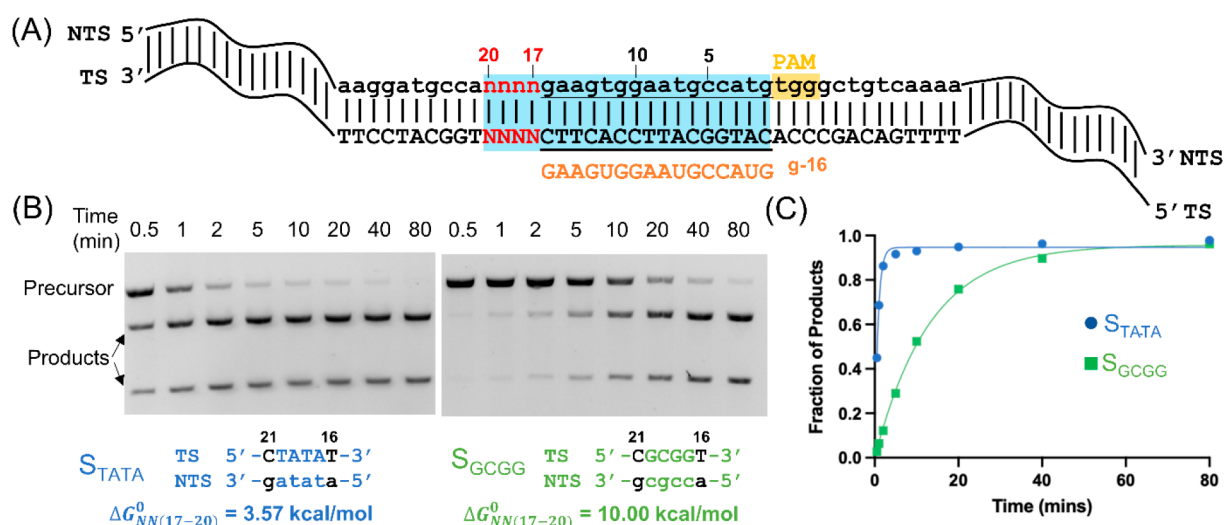


Figure 1. PAM-distal sequence beyond the RNA/DNA hybrid modulates double-stranded DNA cleavage by Cas9 with a 16-nt guide. (A) Schematic representation of DNA substrates. Cas9 target sequence embedded in a linearized pUC19 plasmid was shown by letters, with the PAM highlighted in yellow, the 20-nt protospacer highlighted in blue, the TS sequence represented by uppercase letters, the NTS sequence represented by lowercase letters, and the 16-nt RNA guide indicated with orange uppercase letters. The variable sequence at PAM+(17–20) was indicated by red letters. (B) Double-stranded DNA cleavage results for substrates S_{TATA} (left) and S_{GCGG} (right). Representative electrophoresis images of the time course are shown on top, and the corresponding PAM-distal variable sequences and $\Delta G_{NN(17-20)}^0$ values are shown at the bottom, with nucleotides varying from positions PAM+17 through +20 (colored) but fixed at +16 and +21 (black). $\Delta G_{NN(17-20)}^0$ values were calculated based on the nearest-neighbor parameters as described in the [Materials and Methods](#). (C) Single-exponential fit (eq 2) to derived cleavage rate constants for data sets shown in (B). The examples shown here gave $k_{cat} = 1.28 \text{ min}^{-1}$ and $a = 0.95$ for S_{TATA} and $k_{cat} = 0.08 \text{ min}^{-1}$ and $a = 0.96$ for S_{GCGG} .

dissociation propensity of a given DNA duplex sequence, the dissociation free energy (ΔG_{NN}^0) of a designated segment of a DNA/DNA duplex or DNA/RNA hybrid was calculated based on the nearest-neighbor model. Nearest-neighbor parameters reported for DNA/DNA duplex²⁷ and RNA/DNA hybrid²⁸ were used to compute the sequence-dependent standard-state dissociation enthalpy (ΔH^0) and entropy (ΔS^0) from which the standard-state dissociation free energy, ΔG_{NN}^0 , was calculated as

$$\Delta G_{NN}^0 = \Delta H^0 - T\Delta S^0 \quad (4)$$

where $T = 310.15 \text{ K}$.

Multiple linear regression (MLR) models were trained to predict the measured k_{cat} using the ΔG_{NN}^0 values of various input sequences as features ($\vec{x}$). Training determines the set of parameters ($\vec{\beta}$) that will lead to the best predictions of $\ln(k_{cat})$, as shown below.

$$\ln(k_{cat}) = \beta_0 + x_1\beta_1 + x_2\beta_2 + \dots \quad (5)$$

Model performance is measured using the adjusted coefficient of determination, R_{adj}^2 , which accounts for the number of features included in modeling to avoid overfitting. ΔG_{NN}^0 can be presented to the model as individual features associated with each dinucleotide step or as a single feature covering multiple steps. For example, if the entire PAM+(17–20) segment is considered as a whole, the sum of ΔG_{NN}^0 of five dinucleotides spanning the segment, $\Delta G_{NN(17-20)}^0$, is the only feature:

$$\begin{aligned} \Delta G_{NN(17-20)}^0 &= \Delta G_{NN(16/17)}^0 + \Delta G_{NN(17/18)}^0 \\ &+ \Delta G_{NN(18/19)}^0 + \Delta G_{NN(19/20)}^0 + \Delta G_{NN(20/21)}^0 \end{aligned} \quad (6a)$$

and the model is fit to the following:

$$\ln(k_{cat}) = \beta_0 + \Delta G_{NN(17-20)}^0 \cdot \beta_1 \quad (6b)$$

Alternatively, a model considering all five dinucleotides separately contains five features:

$$\begin{aligned} \ln(k_{cat}) &= \beta_0 + \Delta G_{NN(16/17)}^0 \cdot \beta_1 + \Delta G_{NN(17/18)}^0 \cdot \beta_2 \\ &+ \Delta G_{NN(18/19)}^0 \cdot \beta_3 + \Delta G_{NN(19/20)}^0 \cdot \beta_4 + \Delta G_{NN(20/21)}^0 \cdot \beta_5 \end{aligned} \quad (6c)$$

The total number of features can then be reduced by removing features that do not appear to be sufficiently informative of k_{cat} as described. For example, a model considering only dinucleotides 17/18 and 18/19 contains two features:

$$\ln(k_{cat}) = \beta_0 + \Delta G_{NN(17/18)}^0 \cdot \beta_1 + \Delta G_{NN(18/19)}^0 \cdot \beta_2 \quad (6d)$$

RESULTS

Cas9 Cleavage with a 16-Nucleotide Guide Is Highly Correlated with DNA Duplex Stability beyond the RNA/DNA Hybrid. To assess the impact of the DNA segment beyond the RNA/DNA hybrid on Cas9 activity, we first measured double-strand cleavage rates of Cas9 ribonucleoprotein complexes containing a 16-nt guide (Figure 1A, designated as g-16). The DNA substrates have the same PAM+(1–16) sequences that fully match the RNA guide but vary in their PAM+(17–20) segment (Figure 1A, SI, sect. S1). Each DNA substrate was designated as S_{NNNN} , with the 4-letter subscript indicating PAM+(17–20) nucleotides on the TS in the 5' to 3' direction. For example, S_{TATA} represents a DNA duplex sequence with the PAM+(17–20) positions being (5'-TATA-3') on the TS and (3'-atat-5') on the NTS, while S_{GCGG} represents a sequence containing TS (5'-GCGG-3') and NTS (3'-cgcc-5') at the PAM+(17–20) positions (Figure 1B, Table S1). Note that varying the PAM+(17–20) positions could give rise to $\Delta G_{NN(17-20)}^0$ ranging from 3.57 to 10.07 kcal/

mol. In this work, 26 substrates were studied that uniformly cover the full range of the possible $\Delta G_{\text{NN}(17-20)}^0$ (Table S2).

Figure 1 shows representative examples of a time course analysis of cleavage on S_{TATA} and S_{GCCG} . With both substrates, the linearized duplexed precursor was cleaved completely into two shorter fragments (Figure 1B), and the time course of product generation was fit nicely with a single-exponential model (eq 2) to yield the cleavage reaction rate (Figure 1C), which represents k_{cat} of the reaction.²⁵ The measurements yielded $k_{\text{cat}} = (0.08 \pm 0.01 \text{ min}^{-1})$ for S_{GCCG} , which is ~ 16 times slower than that of S_{TATA} ($1.29 \pm 0.05 \text{ min}^{-1}$) (Figure 1C, Table S2). Note that $\Delta G_{\text{NN}(17-20)}^0$ is 10.00 kcal/mol for S_{GCCG} and 3.57 kcal/mol for S_{TATA} (Figure 1B, Table S2). As such, the data clearly show that a difference in $\Delta G_{\text{NN}(17-20)}^0$ results in a difference in Cas9 cleavage, with the more stable duplex (i.e., S_{GCCG} with a higher $\Delta G_{\text{NN}(17-20)}^0$) cleaving more slowly.

The correlation between k_{cat} and $\Delta G_{\text{NN}(17-20)}^0$ was found to hold for the 26 substrates studied (Figure 2, Table S2). Linear

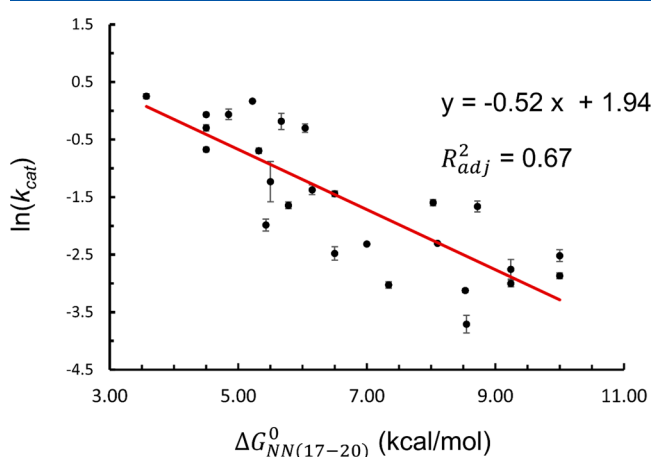


Figure 2. Linear regression analysis of $\ln(k_{\text{cat}})$ vs $\Delta G_{\text{NN}(17-20)}^0$. Black dots represent data obtained on the 26 substrates, and the red line represents the best linear fit. See Table S2 for further information.

regression analysis between $\ln(k_{\text{cat}})$ and $\Delta G_{\text{NN}(17-20)}^0$ (eq 6b, Materials and Methods) yielded a relationship of $\ln(k_{\text{cat}}) = -0.52 \times \Delta G_{\text{NN}(17-20)}^0 + 1.94$, with R_{adj}^2 being 0.67 (Figure 2). R_{adj}^2 indicates a high degree of linear dependence of $\ln(k_{\text{cat}})$ on $\Delta G_{\text{NN}(17-20)}^0$. The negative slope of -0.52 indicates that a sequence with a higher $\Delta G_{\text{NN}(17-20)}^0$, which is expected to form a more stable duplex, would have a smaller k_{cat} . Analysis of the 26 substrates clearly supports the notion that the DNA duplex stability at PAM+(17–20) modulates Cas9 cleavage when using a 16-nt guide.

While R_{adj}^2 of 0.67 indicates a high degree of correlation between cleavage (i.e., k_{cat}) and duplex stability beyond the RNA/DNA hybrid ($\Delta G_{\text{NN}(17-20)}^0$), we do note that in some cases similar $\Delta G_{\text{NN}(17-20)}^0$ yields quite different k_{cat} (Figure 2, Table S2). This suggests additional sequence- and/or position-dependent factors may contribute to modulating Cas9 cleavage. Additional analysis indicates that the stacking energy of the PAM-distal NTS also contributes to modulating Cas9 cleavage rate (SI, sect. S2.2). However, more work is needed to fully elucidate the role of these factors (see also Discussion).

PAM-Distal Sequences beyond the RNA/DNA Hybrid Tunes Cleavage Activity by Modulating Cas9-Induced DNA Unwinding. Our working hypothesis is that the duplex

segment beyond the RNA/DNA hybrid tunes cleavage with the 16-nt RNA guide (Figures 1 and 2) by affecting the degree of Cas9-induced PAM-distal unwinding. To test this, a previously developed 2-amino-purine (2AP) assay²⁵ was applied to assess unwinding of two DNA substrates, S_{TATA} and S_{GATA} (Figure 3, SI, sect. S3). In these measurements, the adenine of the TS PAM+18 position was substituted with 2AP (Figure 3A). The concentration-normalized 2AP fluorescence, which is proportional to the 2AP emission quantum yield, was measured for the DNA duplex and the corresponding ternary complex with the matching g-16 RNA guide. The ratio between the quantum yield of the ternary complex and that of the duplex, $\text{ratio}(\varphi)$ (eq 3), reports the degree of DNA unwinding in the Cas9 ternary complex, with higher $\text{ratio}(\varphi)$ indicating more unwinding.²⁵ The results show that S_{TATA} , which has a $\Delta G_{\text{NN}(17-20)}^0$ of 3.57 kcal/mol, gives a $\text{ratio}(\varphi)$ of 7.55 ± 0.55 , which is significantly higher than that of 4.58 ± 0.53 obtained for S_{GATA} , which has a $\Delta G_{\text{NN}(17-20)}^0$ of 5.50 kcal/mol (Figure 3A,B). This indicates that S_{TATA} , which has a lower duplex stability, as indicated by the smaller $\Delta G_{\text{NN}(17-20)}^0$, allows a larger degree of DNA unwinding. Indeed, cleavage of S_{TATA} was faster than that of S_{GATA} (Figure 3C), consistent with the notion that a larger degree of unwinding allows for higher activity.

Furthermore, with the 16-nt guide, studies show that the measured k_{cat} has a measurable dependence on the concentration of KCl in the reaction buffer, with k_{cat} clearly decreasing with increasing [KCl] (SI, sect. S4). This is consistent with the notion that increasing KCl concentration enhances DNA duplex stability, subsequently reducing DNA unwinding and impeding Cas9 cleavage. Together, the data strongly support the hypothesis that PAM-distal DNA sequences modulate unwinding and, in turn, tune Cas9 cleavage activity.

PAM+(17–20) DNA Tunes In Vitro Cleavage Activity with Multiple 16-nt RNA Guide Sequences. To evaluate if the guide sequence also plays a role in the modulation of Cas9 cleavage rates by the PAM+(17–20) positions, studies were carried out on samples containing different 16-nt RNA guide sequences, which vary in the stability between the TS and NTS strands ($\Delta G_{\text{dd}(1-16)}^0$) as well as that between the RNA-guide and the DNA TS ($\Delta G_{\text{RD}(1-16)}^0$) (Figure 4). Analysis showed that R_{adj}^2 between $\ln(k_{\text{cat}})$ and $\Delta G_{\text{dd}(1-16)}^0$ as well as $\Delta G_{\text{RD}(1-16)}^0$ were negative and that between $\ln(k_{\text{cat}})$ and $\Delta \Delta G_{(1-16)}^0 = \Delta G_{\text{RD}(1-16)}^0 - \Delta G_{\text{dd}(1-16)}^0$ was 0.14 (Figure 4B). These results indicate that there is no strong correlation between the observed in vitro cleavage activity and the varying sequences at the PAM+(1–16) segment involved in RNA/DNA interactions. On the other hand, an R_{adj}^2 of 0.85 was obtained between the measured $\ln(k_{\text{cat}})$ values and the corresponding $\Delta G_{\text{NN}(17-20)}^0$, the DNA duplex dissociation free energy for the PAM+(17–20) segment beyond the RNA/DNA hybrid (Figure 4B). Furthermore, the additional data points fit nicely to the linear regression model originally obtained from analyzing the 26 S_{NNNN} sequences using the original 16-nt RNA guide (SI, sect. S5). Together, the results indicate that PAM+(17–20) positions are predictive of Cas9 activity with multiple 16-nt RNA guides.

PAM + 17 and 18 Nucleotides Are Most Influential in Determining k_{cat} when Using g-16 RNA Guides. The analysis reported in Figure 2 considered the stability over the entire segment of PAM+(17–20) (eqs 6a and 6b, Materials and Methods). However, the five dinucleotides within this

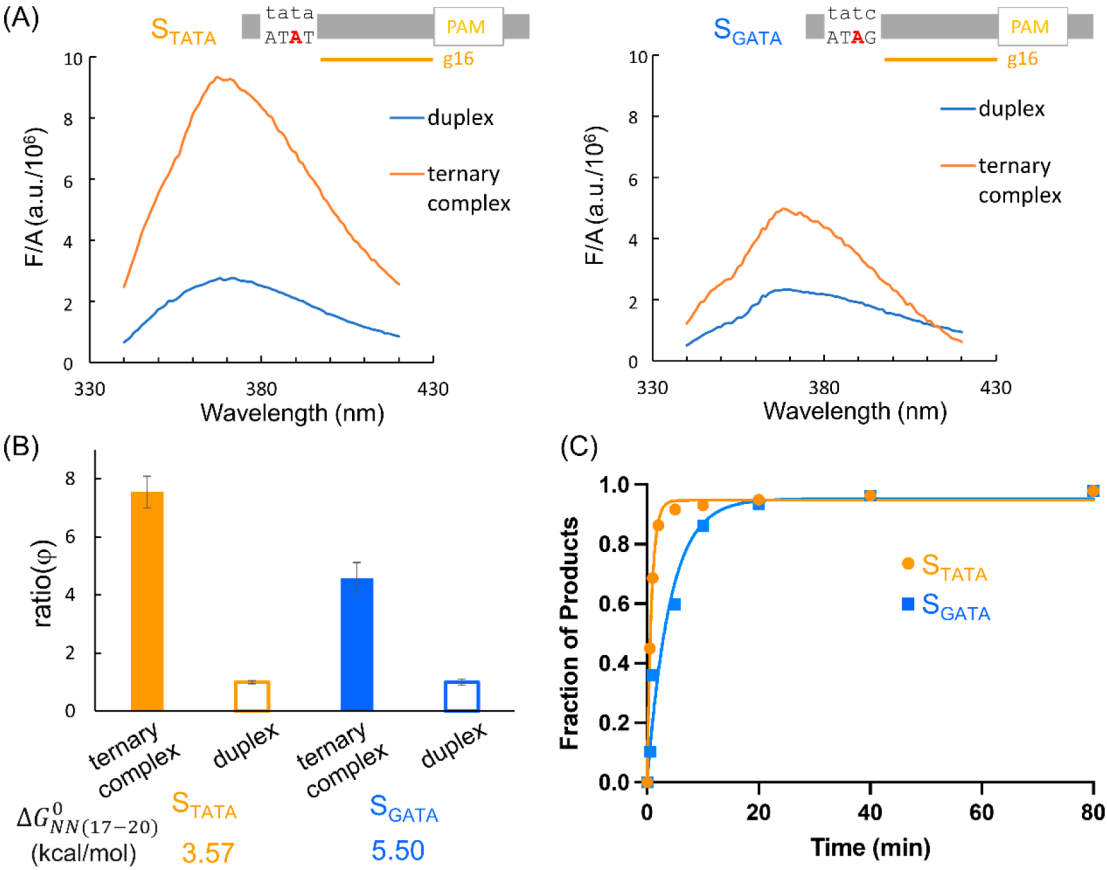
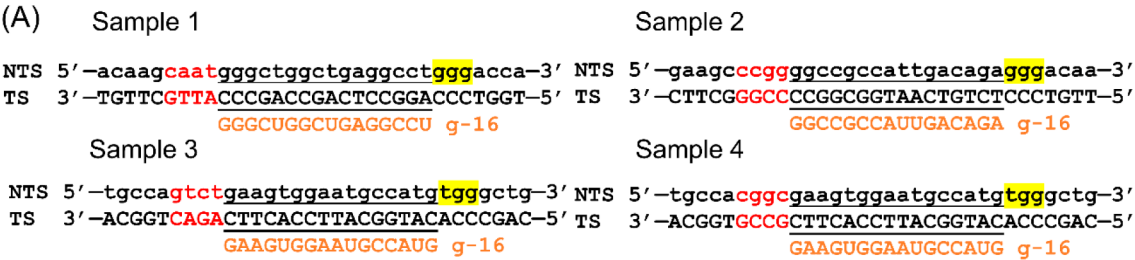


Figure 3. Sequence-dependent PAM-distal unwinding assessed by the 2AP fluorescence assay. (A) Representative concentration-normalized 2AP emission spectra (F/A) for substrates S_{TATA} (left) and S_{GATA} . (B) Measured average ratio(ϕ) values for substrates S_{TATA} (left) and S_{GATA} . See SI, sect. S3 for more details. (C) k_{cat} measurements for S_{TATA} and S_{GATA} . The examples shown here gave $k_{cat} = 1.28 \text{ min}^{-1}$ and $a = 0.95$ for S_{TATA} and $k_{cat} = 0.24 \text{ min}^{-1}$ and $a = 0.95$ for S_{GATA} .



sample	$\ln(k_{cat})$	$\Delta G_{NN(17-20)}^0$ (kcal/mol)	$\Delta G_{dd(1-16)}^0$ (kcal/mol)	$\Delta G_{RD(1-16)}^0$ (kcal/mol)	$\Delta \Delta G_{(1-16)}^0$ (kcal/mol)
1	-1.10	6.86	24.83	24.11	-0.72
2	-3.22	9.63	23.71	21.18	-2.53
3	-1.44	6.50	20.53	19.95	-0.58
4	-2.84	10.00	20.53	19.95	-0.58
R_{adj}^2		0.85	-0.46	-0.13	0.14

Figure 4. Sequences of the 16-nt RNA guide do not significantly influence cleavage activity. (A) Detailed sequence of four samples with three different 16-nt guide sequences. (B) Analysis of the corresponding linear correlation between $\ln(k_{cat})$ and different aspects of the thermodynamic stability of the samples.

segment are not equal in their positioning within the Cas9 RNP, particularly with respect to the RNA/DNA hybrid. To examine whether the contribution to Cas9 activation from individual positions varies, we further carried out MLR analysis on models predicting $\ln(k_{\text{cat}})$ of 26 sequences, considering the stability of every dinucleotide step along the PAM+(17–20) DNA as individual features (e.g., eq 6c, Materials and Methods).

Table 1 summarizes the R_{adj}^2 values that are obtained for a given model, with the row number indicating the starting

Table 1. R_{adj}^2 Values Obtained from MLR Analysis of $\ln(k_{\text{cat}})$ vs $\Delta G_{\text{NN}(16/17)}^0$ of All Possible Contiguous Stretches of the PAM+(17–20) DNA

start/end	17	18	19	20	21
16	0.29	0.77	0.80	0.79	0.78
17		0.73	0.78	0.78	0.77
18			0.55	0.54	0.54
19				0.26	0.26
20					0.17

position of the first dinucleotide and the column number indicating the final position of the last dinucleotide considered. For example, the cell corresponding to row 17 column 18 indicates that a model with one feature, the dissociation free energy of dinucleotide 17/18 (i.e., $\Delta G_{\text{NN}(17/18)}^0$), gave an R_{adj}^2 value of 0.73 (Table 1, row 17, column 18). This is higher than that obtained with the PAM+(17–20) segment considered as a whole ($R_{\text{adj}}^2 = 0.67$, Figure 2), indicating that positions PAM+(17–18) alone can better predict the measured activities. Expanding to a two-feature model by adding dinucleotide 18/19 (i.e., eq 6d) returned an even higher R_{adj}^2 of 0.78 (Table 1, row 17, column 19), while further including 19/20 and 20/21 dinucleotides did not improve R_{adj}^2 (Table 1, row 17, columns 20 and 21).

The conclusion that positions PAM+(17–18) are the most influential on the measured Cas9 activity can be drawn similarly from data shown in row 16 of Table 1, as models including positions PAM+(17–18) gave high R_{adj}^2 values (0.77 and 0.80, Table 1, row 16, columns 18 and 19), while including additional positions did not result in model improvement. Furthermore, models including only the PAM+(19–20) positions gave $R_{\text{adj}}^2 < 0.26$ (Table 1, rows 19 and row 20), indicating these positions have little influence on Cas9 activity.

Taken together, the analysis shows that, with the 16-nt guide, the contribution to Cas9 activation from individual positions varies, with the PAM+(17–18) positions playing the dominant role.

DISCUSSION

Data presented clearly demonstrate that, for Cas9 enzymes containing 16-nt RNA guides, DNA protospacer sequences beyond the RNA/DNA hybrid tune Cas9 cleavage rates (i.e., k_{cat}) (Figures 1 and 2). This is linked to variable capacities for DNA unwinding (Figure 3), which can be predicted to a high degree by the computed intrinsic duplex stability at the DNA protospacer segment beyond the RNA/DNA hybrid (i.e., $\Delta G_{\text{NN}(17-20)}^0$). As the transition state free energy barrier of Cas9 cleavage ($\Delta G^\ddagger$) is proportional to $\ln(k_{\text{cat}})$, the observed linear correlation between $\ln(k_{\text{cat}})$ and $\Delta G_{\text{NN}(17-20)}^0$ (Figure 2) therefore indicates that $\Delta G^\ddagger$ is proportional to $\Delta G_{\text{NN}(17-20)}^0$. This reveals that DNA unwinding at the PAM+(17–20)

positions is the major factor that tunes the transition state for the Cas9 enzyme with the 16-nt guide. This further expands the prior finding that truncated guides support partial protospacer unwinding beyond the RNA/DNA hybrid, and the degree of unwinding is correlated with Cas9 cleavage activity.²⁵

Our analysis also revealed that the PAM+(17–18) positions are the most influential in determining k_{cat} , while positions PAM+(19–20) show little impact (Table 1). An atomic force microscopy study has indicated that PAM+(14–17) RNA/DNA hybrid stability, which is directly linked to unwinding of DNA, serves as a gate to control Cas9 activation.¹⁹ A recent study reported that Cas9 twists and bends the PAM-proximal DNA duplex to enable RNA/DNA hybrid formation and propagation.²⁹ We have previously proposed that PAM-distal unwinding beyond RNA/DNA hybrid occurs as part of the sequential mechanism for R-loop formation and that, in order to propagate the R-loop, Cas9 is able to “pre-unwind” PAM-distal DNA base-pair(s) ahead of the RNA/DNA hybrid to expose the target strand and assess pairing with the RNA guide.²⁵ With the 16-nt guide, the RNA/DNA hybrid formally terminates at PAM+16, and Cas9’s “pre-unwinding” ability is likely responsible for unwinding the PAM+(17–20) segment. Given that PAM+(17–18) is adjacent to the terminus of the RNA/DNA hybrid, their DNA stability (i.e., high $\Delta G_{\text{NN}(16/17)}^0$ and/or $\Delta G_{\text{NN}(17/18)}^0$) likely impacts the ability of Cas9 to “pre-unwind”, in particular unwinding at PAM+17/18 that influences the “gate”, thus accounting for their significant roles in tuning cleavage rates (Table 1).

Note that duplex stability is one aspect of DNA shape, which is defined as the intrinsic physical properties of a DNA duplex that are collectively determined by its sequence.³⁰ Our results expand on prior reports that DNA shape contributes to target discrimination in Cas9³¹ and Cas12a.³² While Cas9 activity significantly correlates with intrinsic DNA duplex stability (Figures 1 and 2), additional DNA shape or sequence-dependent features may influence Cas9 function with 16-nt guides. One aspect may be intrinsic single-stranded properties of the DNA (SI, sect. S2.2), although much more work is needed in this area. Furthermore, variations in protein–DNA interactions may play a role. With 20-nt guides, contacts between the protein and the PAM-distal segment of the RNA/DNA hybrid play a key role in facilitating conformational changes to achieve the catalytically competent state.^{16,20–22,33} With 16-nt guides, RNA nucleotides at PAM+(17–20) are completely missing. The resulting loss of RNA/protein and RNA/DNA interactions may render the protein to directly interrogate the DNA duplex, thus elevating the role of intrinsic shape features of the DNA duplex, including the PAM+(17–20) stability, as reported here (Figures 1 and 2). On the other hand, the lack of PAM+(17–20) nucleotides in the RNA guide results in loss of negative charges, steric bulkiness, and functional groups for hydrogen bonding and ion coordination, all of which may impact the protein–RNA–DNA interactions that drive the system to the catalytically active conformation.^{16,20–22,33} As such, it remains to be learned whether intrinsic DNA duplex stability plays a similar role between the 16-nt guides studied here and other ones, such as 20-nt guides with PAM+(17–20) mismatches to target DNA or longer truncated guides (i.e., 17-nt or 18-nt guides) that extend the RNA/DNA hybrid.

RNA guide truncation has been used as one of the strategies to enhance Cas9 specificity, and it has been reported that 18-nt

and 17-nt guides can significantly reduce off-target effects in human genome editing.^{23,24} However, guides shorter than 17-nt do not appear to exhibit cleavage activity in cells.^{23,24} Data presented here indicate that, with a 16-nt guide, Cas9 cleavage could occur in vitro and could differ by more than 50-fold depending on the PAM+(17–20) DNA sequence (Figure 2, Table S2). This suggests that, while selecting genomic targets for 16-nt guides, consideration of the “peripheral” PAM+(17–20) positions may provide additional information that promotes successful editing. In addition, we note that rules for defining an optimal target matching a 16-nt guide may differ from those matching a 20-nt guide or other truncated guides (e.g., 17-nt or 18-nt guides), especially considering the role of the PAM-distal RNA/DNA hybrid in enabling Cas9 activation.^{16,20–22,33} Much remains to be learned about Cas9 activation with 16-nt and other truncated guides.

CONCLUSION

In conclusion, the data presented show that, with 16-nt guides, intrinsic DNA duplex stability at the PAM+(17–20) protospacer segment modulates Cas9-induced PAM-distal unwinding, which in turn tunes the transition state of the cleavage reaction, thus impacting Cas9 activity. This clearly demonstrates that, with a 16-nt guide, “peripheral” DNA sequences beyond the DNA/RNA hybrid play a role in Cas9 target discrimination. The finding provides a mechanistic understanding for further developing strategies based on RNA guide truncation to enhance Cas9 specificity and reduce “off-target” effects.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.3c00250>.

Additional information on DNA substrates, thermodynamic parameters of DNA duplex segment, 2-amino-purine fluorescence data, dependence of activities on KCl concentration, and additional information on impact of varying sequence of the 16-nt RNA guide (PDF)

Accession Codes

UniProt protein ID: Q99ZW2.

AUTHOR INFORMATION

Corresponding Author

Peter Z. Qin – Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States; orcid.org/0000-0003-3967-366X; Phone: (213) 821-2461; Email: pzq@usc.edu; Fax: (213) 740-2701

Authors

Yue Li – Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States; Present Address: FutureGen Biopharmaceutical Co., Ltd., Beijing 102629, China

Brendon H. Cooper – Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California 90089, United States

Yukang Liu – Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States

Difei Wu – Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States

XiaoJun Zhang – Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States

Remo Rohs – Department of Chemistry and Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California 90089, United States; orcid.org/0000-0003-1752-1884

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.biochem.3c00250>

Author Contributions

Y. Li, X.Z., and P.Z.Q. designed the overall studies; Y. Li and D.W. carried out the enzyme kinetics measurements and analysis; Y. Liu carried out fluorescence measurements and analysis; B.H.C. and R.R. designed and carried out the linear regression analysis. Y. Li, X.Z., and P.Z.Q. contributed to manuscript writing, and all authors participated in manuscript editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation (MCB-1818107 to P.Z.Q.) and the National Institute of Health (R35GM145341 to P.Z.Q. and R35GM130376 to R.R.). The authors would like to thank Yun Fang (Department of Chemistry, University of Southern California) for assistance on protein purification and performing gel electrophoresis.

REFERENCES

- (1) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816–821.
- (2) Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas9-CrRNA Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria. *P Natl. Acad. Sci. USA* **2012**, *109* (39), E2579–E2586.
- (3) Cong, L.; Ran, F. A.; Cox, D.; Lin, S. L.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X. B.; Jiang, W. Y.; Marraffini, L. A.; Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339* (6121), 819–823.
- (4) Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science* **2013**, *339* (6121), 823–826.
- (5) Fu, Y.; Foden, J. A.; Khayter, C.; Maeder, M. L.; Reyon, D.; Joung, J. K.; Sander, J. D. High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells. *Nat. Biotechnol.* **2013**, *31* (9), 822–826.
- (6) Lin, Y. N.; Cradick, T. J.; Brown, M. T.; Deshmukh, H.; Ranjan, P.; Sarode, N.; Wile, B. M.; Vertino, P. M.; Stewart, F. J.; Bao, G. CRISPR/Cas9 Systems Have off-Target Activity with Insertions or Deletions between Target DNA and Guide RNA Sequences. *Nucleic Acids Res.* **2014**, *42* (11), 7473–7485.
- (7) Yang, L. H.; Grishin, D.; Wang, G.; Aach, J.; Zhang, C. Z.; Chari, R.; Homys, J.; Cai, X. Y.; Zhao, Y.; Fan, J. B.; Seidman, C.; Seidman, J.; Pu, W.; Church, G. Targeted and Genome-Wide Sequencing Reveal Single Nucleotide Variations Impacting Specificity of Cas9 in Human Stem Cells. *Nat. Commun.* **2014**, *5*, 5507.
- (8) Kim, D.; Luk, K.; Wolfe, S. A.; Kim, J.-S. Evaluating and Enhancing Target Specificity of Gene-Editing Nucleases and Deaminases. *Annu. Rev. Biochem.* **2019**, *88*, 191–220.

- (9) Moon, S. B.; Kim, D. Y.; Ko, J.-H.; Kim, J.-S.; Kim, Y.-S. Improving CRISPR Genome Editing by Engineering Guide RNAs. *Trends Biotechnol* **2019**, *37* (8), 870–881.
- (10) Galizi, R.; Jaramillo, A. Engineering CRISPR Guide RNA Riboswitches for in Vivo Applications. *Curr. Opin Biotechnol* **2019**, *55*, 103–113.
- (11) Anzalone, A. V.; Koblan, L. W.; Liu, D. R. Genome Editing with CRISPR–Cas Nucleases, Base Editors, Transposases and Prime Editors. *Nat. Biotechnol.* **2020**, *38* (7), 824–844.
- (12) Doudna, J. A. The Promise and Challenge of Therapeutic Genome Editing. *Nature* **2020**, *578* (7794), 229–236.
- (13) Huang, X.; Yang, D.; Zhang, J.; Xu, J.; Chen, Y. E. Recent Advances in Improving Gene-Editing Specificity through CRISPR–Cas9 Nuclease Engineering. *Cells* **2022**, *11* (14), 2186.
- (14) Jiang, F.; Doudna, J. A. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **2017**, *46* (1), 505–529.
- (15) Zuo, Z.; Liu, J. Allosteric Regulation of CRISPR–Cas9 for DNA-Targeting and Cleavage. *Curr. Opin. Struct. Biol.* **2020**, *62*, 166–174.
- (16) Pacesa, M.; Loeff, L.; Querques, I.; Muckenfuss, L. M.; Sawicka, M.; Jinek, M. R-Loop Formation and Conformational Activation Mechanisms of Cas9. *Nature* **2022**, *609* (7925), 191–196.
- (17) Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A. DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9. *Nature* **2014**, *507* (7490), 62–67.
- (18) Boyle, E. A.; Andreasson, J. O. L.; Chircus, L. M.; Sternberg, S. H.; Wu, M. J.; Guegler, C. K.; Doudna, J. A.; Greenleaf, W. J. High-Throughput Biochemical Profiling Reveals Sequence Determinants of DCas9 off-Target Binding and Unbinding. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (21), 5461–5466.
- (19) Josephs, E. A.; Kocak, D. D.; Fitzgibbon, C. J.; McMenemy, J.; Gersbach, C. A.; Marszalek, P. E. Structure and Specificity of the RNA-Guided Endonuclease Cas9 during DNA Interrogation, Target Binding and Cleavage. *Nucleic Acids Res.* **2015**, *43* (18), 8924–8941.
- (20) Chen, J. S.; Dagdas, Y. S.; Kleinstiver, B. P.; Welch, M. M.; Sousa, A. A.; Harrington, L. B.; Sternberg, S. H.; Joung, J. K.; Yildiz, A.; Doudna, J. A. Enhanced Proofreading Governs CRISPR–Cas9 Targeting Accuracy. *Nature* **2017**, *550* (7676), 407–410.
- (21) Dagdas, Y. S.; Chen, J. S.; Sternberg, S. H.; Doudna, J. A.; Yildiz, A. A Conformational Checkpoint between DNA Binding and Cleavage by CRISPR–Cas9. *Science Advances* **2017**, *3* (8), No. eaao0027.
- (22) Pacesa, M.; Lin, C.-H.; Cléry, A.; Saha, A.; Arantes, P. R.; Bargsten, K.; Irby, M. J.; Allain, F. H.-T.; Palermo, G.; Cameron, P.; Donohoue, P. D.; Jinek, M. Structural Basis for Cas9 Off-Target Activity. *Cell* **2022**, *185* (22), 4067–4081.e21.
- (23) Fu, Y. F.; Sander, J. D.; Reyon, D.; Cascio, V. M.; Joung, J. K. Improving CRISPR–Cas Nuclease Specificity Using Truncated Guide RNAs. *Nat. Biotechnol.* **2014**, *32* (3), 279–284.
- (24) Zhang, J.-P.; Li, X.-L.; Neises, A.; Chen, W.; Hu, L.-P.; Ji, G.-Z.; Yu, J.-Y.; Xu, J.; Yuan, W.-P.; Cheng, T.; Zhang, X.-B. Different Effects of SgRNA Length on CRISPR-Mediated Gene Knockout Efficiency. *Sci. Rep.* **2016**, *6* (1), 28566.
- (25) Li, Y.; Liu, Y.; Singh, J.; Tangprasertchai, N. S.; Trivedi, R.; Fang, Y.; Qin, P. Z. Site-Specific Labeling Reveals Cas9 Induces Partial Unwinding Without RNA/DNA Pairing in Sequences Distal to the PAM. *CRISPR J.* **2022**, *5* (2), 341–352.
- (26) Tangprasertchai, N. S.; Di Felice, R.; Zhang, X.; Slaymaker, I. M.; Vazquez Reyes, C.; Jiang, W.; Rohs, R.; Qin, P. Z. CRISPR–Cas9 Mediated DNA Unwinding Detected Using Site-Directed Spin Labeling. *ACS Chem. Biol.* **2017**, *12* (6), 1489–1493.
- (27) SantaLucia, J., Jr.; Allawi, H. T.; Seneviratne, P. A. Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability. *Biochemistry* **1996**, *35* (11), 3555–3562.
- (28) Banerjee, D.; Tateishi-Karimata, H.; Ohyama, T.; Ghosh, S.; Endoh, T.; Takahashi, S.; Sugimoto, N. Improved Nearest-Neighbor Parameters for the Stability of RNA/DNA Hybrids under a Physiological Condition. *Nucleic Acids Res.* **2020**, *48* (21), 12042–12054.
- (29) Cofsky, J. C.; Soczek, K. M.; Knott, G. J.; Nogales, E.; Doudna, J. A. CRISPR–Cas9 Bends and Twists DNA to Read Its Sequence. *Nat. Struct. Mol. Biol.* **2022**, *29* (4), 395–402.
- (30) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of Specificity in Protein–DNA Recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233–269.
- (31) Newton, M. D.; Taylor, B. J.; Driessen, R. P. C.; Roos, L.; Cvetsi, N.; Allyjaun, S.; Lenhard, B.; Cuomo, M. E.; Rueda, D. S. DNA Stretching Induces Cas9 Off-Target Activity. *Nat. Struct. Mol. Biol.* **2019**, *26* (3), 185–192.
- (32) Jiang, W.; Singh, J.; Allen, A.; Li, Y.; Kathiresan, V.; Qureshi, O.; Tangprasertchai, N.; Zhang, X.; Parameshwaran, H. P.; Rajan, R.; Qin, P. Z. CRISPR–Cas12a Nucleases Bind Flexible DNA Duplexes without RNA/DNA Complementarity. *ACS Omega* **2019**, *4* (17), 17140–17147.
- (33) Bravo, J. P. K.; Liu, M.-S.; Hibshman, G. N.; Dangerfield, T. L.; Jung, K.; McCool, R. S.; Johnson, K. A.; Taylor, D. W. Structural Basis for Mismatch Surveillance by CRISPR–Cas9. *Nature* **2022**, *603* (7900), 343–347.

Recommended by ACS

Engineering a *Streptococcus* Cas9 Ortholog with an RxQ PAM-Binding Motif for PAM-Free Gene Control in Bacteria

Yuxi Teng, Yajun Yan, et al.

AUGUST 29, 2023
ACS SYNTHETIC BIOLOGY

READ

Effect of Cas9 Protein on the Seed-Target Base Pair of the sgRNA/DNA Hybrid Duplex

Ting Yu, Wenbing Zhang, et al.

MAY 27, 2023
THE JOURNAL OF PHYSICAL CHEMISTRY B

READ

Active CRISPR–Cas12a on Hydrophilic Metal–Organic Frameworks: A Nanobiocomposite with High Stability and Activity for Nucleic Acid Detection

Zhirun Ji, Bingzhi Li, et al.

JULY 06, 2023
ANALYTICAL CHEMISTRY

READ

Development of CRISPR/Cas Delivery Systems for In Vivo Precision Genome Editing

Yuxuan Chen and Yuan Ping

AUGUST 01, 2023
ACCOUNTS OF CHEMICAL RESEARCH

READ

Get More Suggestions >

Supporting Information

CRISPR-Cas9 Activities with Truncated 16-Nucleotide RNA Guides are Tuned by Target Duplex Stability Beyond the RNA/DNA Hybrid

Yue Li,^{†,§} Brendon H. Cooper,[‡] Yukang Liu,[†] Difei Wu,[†] Xiaojun Zhang,[†] Remo Rohs,^{†,‡} Peter Z. Qin^{†,}*

[†]Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA

[‡]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California 90089, USA

[§]Present address: FutureGen Biopharmaceutical Co., Ltd, Beijing 102629, China

*Corresponding author: Peter Z. Qin, TRF 119, 3430 S. Vermont Ave., Los Angeles, California 90089, USA; Tel: (213) 821-2461; Fax: (213) 740-2701; Email: pzq@usc.edu

Table of Contents

S1. Additional information on DNA substrates

S1.1. Additional information on target DNA sequences

S1.2. Peripheral DNA sequences beyond the protospacer have no impact on measured Cas9 cleavage rate

S2. Additional information on correlating activities of 16-nucleotide guides with DNA sequences beyond the RNA/DNA hybrid

S2.1. Thermodynamic parameters of the PAM+(17-20) DNA duplex segment

S2.2. Impacts of single-stranded DNA property at the PAM-distal NTS segment on Cas9 cleavage rate

S3. Additional information on 2-amino-purine fluorescence measurements

S4. Dependence of activities of 16-nucleotide guides on KCl concentration

S5. Additional information on impact of varying sequence of the 16-nt RNA guide

S1. Additional information on DNA substrates

S1.1: Additional information on target DNA sequences

Figures S1 shows a schematic of the DNA target substrates cloned into the pUC19 plasmid. Additional details on primers used to generate these clones are listed in Table S1.

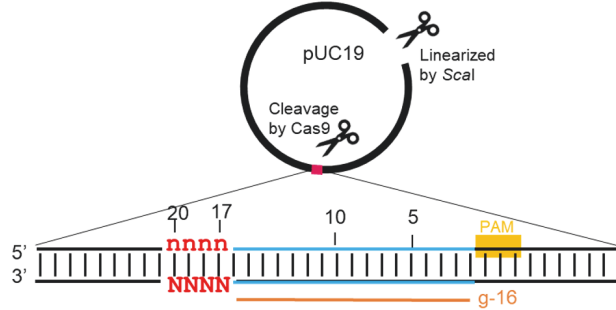


Figure S1. Schematic representation of the substrate DNA cloned into pUC19 vector between the *HindIII* and *BamHI* restriction sites.

Table S1. Primers for generating Cas9 substrates by mutagenesis using pUC19

Target	PAM+(16-21) sequence ^(a)	Primer	Sequence (5'→3') ^(b)
1	5' -CTATAT-3' 3' -gatata-5'	P5-1_a_For	CCACATGGCATTCCACTTCTATATGGCATCCTTCCACTC
		P5-1_a_Rev	GAGTGAAGGATGCCATATAGAAGTGGAAATGCCATGTGGGC
2	5' -CTTATT-3' 3' -gaataa-5'	P5-1-10_For	CCACATGGCATTCCACTTCTTATTGGCATCCTTCCACTC
		P5-1-10_Rev	GAGTGAAGGATGCCAATAAGAAGTGGAAATGCCATGTGG
3	5' -CTAAAT-3' 3' -gattta-5'	P5-1-8_For	CCACATGGCATTCCACTTCTAAATGGCATCCTTCCACTC
		P5-1-8_Rev	GAGTGAAGGATGCCATTTAGAAGTGGAAATGCCATGTGG
4	5' -CTAATT-3' 3' -gattta-5'	P5-1-9_For	CCACATGGCATTCCACTTCTAATTGGCATCCTTCCACTC
		P5-1-9_Rev	GAGTGAAGGATGCCAATTAGAAGTGGAAATGCCATGTGG
5	5' -CTTTT-3' 3' -gaaaaa-5'	P5-1-d_For	CCACATGGCATTCCACTTCTTTTTGGCATCCTTCCACTC
		P5-1-d_Rev	GAGTGAAGGATGCCAAAAAGAAGTGGAAATGCCATGTGGGC
6	5' -CTAGAT-3' 3' -gatcta-5'	P5-1-11_For	CCACATGGCATTCCACTTCTAGATGGCATCCTTCCACTC
		P5-1-11_Rev	GAGTGAAGGATGCCATCTAGAAGTGGAAATGCCATGTGG
7	5' -CAAAAT-3' 3' -gtttta-5'	P5-1-e_For	CCACATGGCATTCCACTTCAAAATGGCATCCTTCCACTC
		P5-1-e_Rev	GAGTGAAGGATGCCATTTTGAAGTGGAAATGCCATGTGGGC
8	5' -CTATGT-3' 3' -gataca-5'	P5-1-14_For	CCACATGGCATTCCACTTCTATGTGGCATCCTTCCACTC
		P5-1-14_Rev	GAGTGAAGGATGCCACATAGAAGTGGAAATGCCATGTGG
9	5' -CGATAT-3' 3' -gctata-5'	P5-1-13_For	CCACATGGCATTCCACTTCGATATGGCATCCTTCCACTC
		P5-1-13_Rev	GAGTGAAGGATGCCATATCGAAGTGGAAATGCCATGTGG
10	5' -CTTCTT-3' 3' -gaagaa-5'	P5-1-18_For	CCACATGGCATTCCACTTCTTCTTGGCATCCTTCCACTC
		P5-1-18_Rev	GAGTGAAGGATGCCAAGAAGAAGTGGAAATGCCATGTGG
11	5' -CAAGAT-3' 3' -gttcta-5'	P5-1-12_For	CCACATGGCATTCCACTTCAAGATGGCATCCTTCCACTC
		P5-1-12_Rev	GAGTGAAGGATGCCATCTTGAAGTGGAAATGCCATGTGG

12	<i>S_{TTGT}</i>	5'-CTTGT-3' 3'-gaaca-5'	P5-1-17_For	CCACATGGCATTCCACTTCTTGTGGCATCCTTCCACTC
			P5-1-17_Rev	GAGTGGAAGGATGCCACAAGAAGTGGAATGCCATGTGG
13	<i>S_{ATTG}</i>	5'-CATTGT-3' 3'-gtaaca-5'	P5-1-16_For	CCACATGGCATTCCACTTCATTGTGGCATCCTTCCACTC
			P5-1-16_Rev	GAGTGGAAGGATGCCACAATGAAGTGGAATGCCATGTGG
14	<i>S_{TCAC}</i>	5'-CTCACT-3' 3'-gagtga-5'	P5-1_For	CCACATGGCATTCCACTTCTCACTGGCATCCTTCC
			P5-1_Rev	GAGTGGAAGGATGCCAGTGAGAAGTGGAATGCCATG
15	<i>S_{AGAC}</i>	5'-CAGACT-3' 3'-gtctga-5'	P5-1-20_For	CCACATGGCATTCCACTTCAGACTGGCATCCTTCCACTC
			P5-1-20_Rev	GAGTGGAAGGATGCCAGTCTGAAGTGGAATGCCATGTGG
16	<i>S_{AGCT}</i>	5'-CAGCTT-3' 3'-gtcga-5'	P5-1-21_For	CCACATGGCATTCCACTTCAGCTTGGCATCCTTCCACTC
			P5-1-21_Rev	GAGTGGAAGGATGCCAAGCTGAAGTGGAATGCCATGTGG
17	<i>S_{CCTC}</i>	5'-CCCTCT-3' 3'-gggaga-5'	P5-1-26_For	CCACATGGCATTCCACTTCCCTCTGGCATCCTTCCACTC
			P5-1-26_Rev	GAGTGGAAGGATGCCAGAGGAAGTGGAATGCCATGTGG
18	<i>S_{GCTC}</i>	5'-CGGTCT-3' 3'-gccaga-5'	P5-1-25_For	CCACATGGCATTCCACTTCGGTCTGGCATCCTTCCACTC
			P5-1-25_Rev	GAGTGGAAGGATGCCAGACGAAGTGGAATGCCATGTGG
19	<i>S_{TCGC}</i>	5'-CTCGCT-3' 3'-gagcga-5'	P5-1-27_For	CCACATGGCATTCCACTTCTCGCTGGCATCCTTCCACTC
			P5-1-27_Rev	GAGTGGAAGGATGCCAGCGGAAGTGGAATGCCATGTGG
20	<i>S_{CGCT}</i>	5'-CCGCTT-3' 3'-gcgga-5'	P5-1-29_For	CCACATGGCATTCCACTTCCGCTTGGCATCCTTCCACTC
			P5-1-29_Rev	GAGTGGAAGGATGCCAAGCGGAAGTGGAATGCCATGTGG
21	<i>S_{GTCG}</i>	5'-CGTCGT-3' 3'-gcagca-5'	P5-1-28_For	CCACATGGCATTCCACTTCGTCGTGGCATCCTTCCACTC
			P5-1-28_Rev	GAGTGGAAGGATGCCACGACGAAGTGGAATGCCATGTGG
22	<i>S_{CCCC}</i>	5'-CCCCCT-3' 3'-gggga-5'	P5-1-f_For	CCACATGGCATTCCACTTCCCCCTGGCATCCTTCCACTC
			P5-1-f_Rev	GAGTGGAAGGATGCCAGGGGAAGTGGAATGCCATGTGGC
23	<i>S_{GGGG}</i>	5'-CGGGGT-3' 3'-gcccc-5'	P5-1-g_For	CCACATGGCATTCCACTTCGGGGTGGCATCCTTCCACTC
			P5-1-g_Rev	GAGTGGAAGGATGCCACCCGAAGTGGAATGCCATGTGGC
24	<i>S_{CGGG}</i>	5'-CCGGGT-3' 3'-gcccc-5'	P5-1-30_For	CCACATGGCATTCCACTTCCGGGTGGCATCCTTCCACTC
			P5-1-30_Rev	GAGTGGAAGGATGCCACCCGAAGTGGAATGCCATGTGG
25	<i>S_{GCCG}</i>	5'-CGCCGT-3' 3'-gcgga-5'	P5-1-32_For	CCACATGGCATTCCACTTCGCCGTGGCATCCTTCCACTC
			P5-1-32_Rev	GAGTGGAAGGATGCCACGGCGAAGTGGAATGCCATGTGG
26	<i>S_{GCGG}</i>	5'-CGCGGT-3' 3'-gcgga-5'	P5-1_b_For	CCACATGGCATTCCACTTCGCGGTGGCATCCTTCCACTC
			P5-1_b_Rev	GAGTGGAAGGATGCCACGCGAAGTGGAATGCCATGTGGC

(a) TS sequence is shown on top with upper-case letters, and NTS sequence is shown at the bottom with lower-case letters. Variable sequences are indicated in red.

(b) Sequences corresponding to the PAM+(16-21) segment are highlighted in yellow.

S1.2. Peripheral DNA sequences beyond the protospacer have no impact on measured Cas9 cleavage rate

Figure S2 shows cleavage data on DNA target embedded either in a *ScaI* linearized pUC19 plasmid (Figure S1, Table S1), or a ~700 base-pair fragment obtained by PCR amplification of exon 3 of the human EMX1 gene (PCR products). The data show the measured cleavage rates are not impacted by the peripheral DNA sequences beyond the PAM and the 20-nt protospacer.

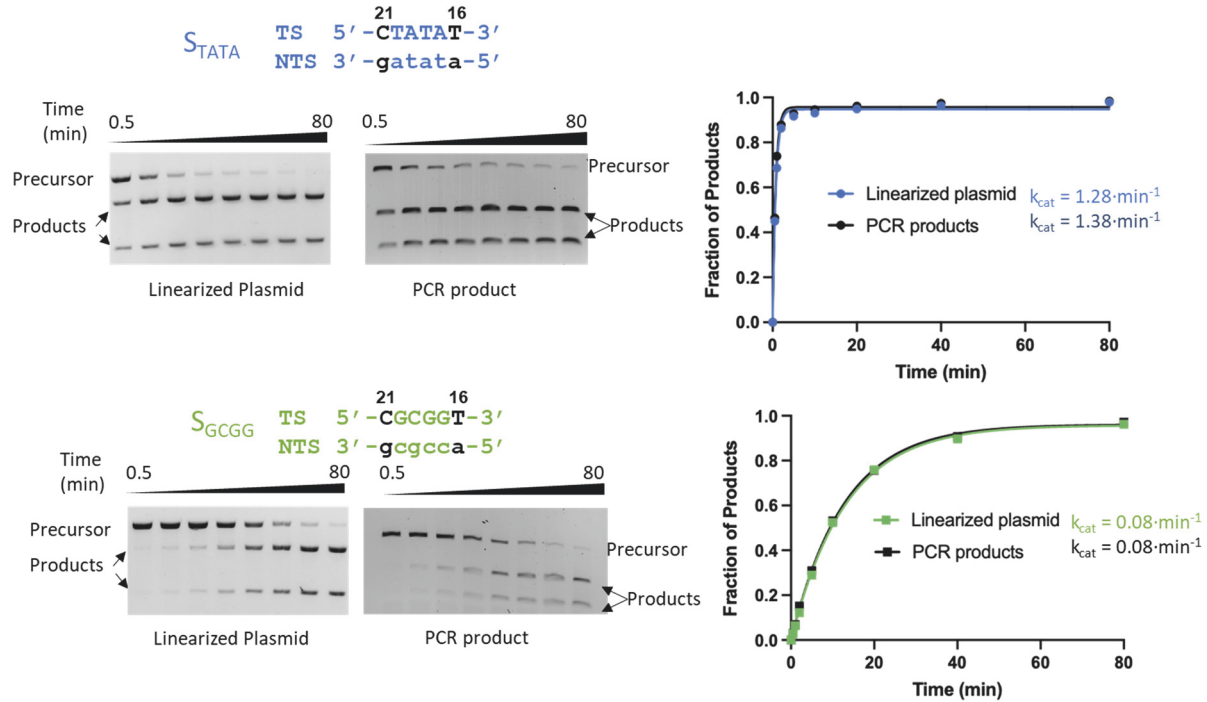


Figure S2. Assessing impact on measured cleavage rates by peripheral DNA sequences beyond the 20-nucleotide protospacer. Two sequences were tested, one with small $\Delta G_{NN(17-20)}^0$ (top panels), and the other with large $\Delta G_{NN(17-20)}^0$ (bottom panels). For both sequences, the cleavage rate determined using linearized plasmid and using PCR products are very similar.

S2. Additional information on correlating activities of 16-nucleotide guides with DNA sequences beyond the RNA/DNA hybrid

S2.1. Thermodynamic parameters of the PAM+(17-20) DNA duplex segment

Table S2 lists the thermodynamic parameters of the PAM+(17-20) DNA segment for the 26 target DNA substrates (Table S1). Note that ΔH^0 and ΔS^0 values were calculated according to the nearest-neighbor parameters obtained in 1M NaCl, while the Cas9 cleavage reactions were carried out in a buffer containing 100 mM KCl and 5 mM MgCl₂ (Main Text, Methods). However, such differences in the salt concentration result in an offset of ΔG_{NN}^0 values that is constant for all the 26 substrates,¹ which is not expected to impact the linear regression analysis on the correlation between $\ln(k_{cat})$ and ΔG_{NN}^0 .

Table S2: Thermodynamic parameters of the PAM+(17-20) DNA segment

Target	PAM+(16-21) sequence ^(a)	k_{cat} ^(b) (min ⁻¹)	Duplex Stability			NTS Flexibility $\Delta G_{nts(17-20)}^0$ ^(e) (kcal/mol)
			$\Delta H_{(17-20)}^0$ ^(c) (kcal/mol)	$\Delta S_{(17-20)}^0$ ^(c) (cal/mol/K)	$\Delta G_{NN(17-20)}^0$ ^(d) (kcal/mol)	
1	S_{TATA} 5'-CTATAT-3' 3'-gatata-5'	1.29±0.05	31.7	90.7	3.57	7.4
2	S_{TTAT} 5'-CTTATT-3' 3'-gaataa-5'	0.51±0.02	35.7	100.6	4.50	7.4
3	S_{TAAA} 5'-CTAAAT-3' 3'-gattta-5'	0.74±0.04	35.7	100.6	4.50	6.6
4	S_{TAAT} 5'-CTAATT-3' 3'-gattaa-5'	0.94±0.03	35.7	100.6	4.50	7.0
5	S_{TTTT} 5'-CTTTT-3' 3'-gaaaaa-5'	0.14±0.01	39.7	110.5	5.43	7.4
6	S_{TAGA} 5'-CTAGAT-3' 3'-gatcta-5'	0.94±0.09	32.7	89.8	4.85	7.4
7	S_{AAAA} 5'-CAAAAT-3' 3'-gtttta-5'	0.50±0.02	39.1	108.9	5.32	6.1
8	S_{TATG} 5'-CTATGT-3' 3'-gataca-5'	1.18±0.03	34.9	95.7	5.22	7.6
9	S_{GATA} 5'-CGATAT-3' 3'-gctata-5'	0.29±0.10	37.1	101.9	5.50	7.1
10	S_{TTCT} 5'-CTTCTT-3' 3'-gaagaa-5'	0.19±0.01	36.7	99.7	5.78	8.2
11	S_{AAGA} 5'-CAAGAT-3' 3'-gttcta-5'	0.83±0.12	36.1	98.1	5.67	6.9
12	S_{TTGT} 5'-CTTGTT-3' 3'-gaacaa-5'	0.25±0.02	38.9	105.6	6.15	7.6
13	S_{ATTG} 5'-CATTGT-3' 3'-gtaaca-5'	0.74±0.05	38.3	104.0	6.04	7.5
14	S_{TCAC} 5'-CTCACT-3' 3'-gagtga-5'	0.08±0.01	35.9	94.8	6.50	8.8
15	S_{AGAC} 5'-CAGACT-3' 3'-gtctga-5'	0.24±0.01	35.9	94.8	6.50	8.0
16	S_{AGCT} 5'-CAGCTT-3' 3'-gtcgaa-5'	0.10±0.00	39.1	103.5	7.00	7.8
17	S_{CCTC} 5'-CCCTCT-3' 3'-gggaga-5'	0.05±0.00	33.3	83.7	7.34	9.0
18	S_{GGTC} 5'-CGGTCT-3' 3'-gccaga-5'	0.20±0.01	39.2	100.5	8.03	8.0
19	S_{TCGC} 5'-CTCGCT-3' 3'-gagcga-5'	0.10±0.00	41.1	106.4	8.10	8.4
20	S_{CGCT} 5'-CCGCTT-3' 3'-ggcgaa-5'	0.04±0.00	42.4	109.2	8.53	8.0

21	S_{GTCC}	5'-C GT CGT-3' 3'-g cagc a-5'	0.19±0.02	45.1	117.3	8.72	7.8
22	S_{CCCC}	5'-C CCCC T-3' 3'-g gggg a-5'	0.02±0.00	32.9	78.5	8.55	9.0
23	S_{GGGG}	5'-C GGGG T-3' 3'-g cccc a-5'	0.05±0.00	38.8	95.3	9.24	7.2
24	S_{CGGG}	5'-C CGGG T-3' 3'-g ggcc a-5'	0.06±0.01	38.8	95.3	9.24	7.6
25	S_{GCCG}	5'-C GCCG T-3' 3'-g cggc a-5'	0.06±0.00	46.6	118.0	10.00	7.8
26	S_{GCGG}	5'-C GCGG T-3' 3'-g cgcc a-5'	0.08±0.01	46.6	118.0	10.00	7.4

(a) TS sequence is shown on top with upper-case letters, and NTS sequence is shown at the bottom with lower-case letters. Variable sequences are indicated in red.

(b) Reported as (average ± s.d.), with the standard-deviation (s.d.) obtained from repeated measurements.

(c) ΔH^0 and ΔS^0 values are calculated according to the nearest-neighbor model using parameters reported in ref.² Contributions of five nearest-neighbor base-pairs were included, i.e., PAM+(16-17), (17-18), (18-19), (19-20), and (20-21).

(d) $\Delta G_{NN(17-20)}^0 = \Delta H^0 - T\Delta S^0$ with T set at 310.15 K. Values reported here are used for Main Text Figure 2.

(e) Flexibility of the single-stranded NTS PAM+(17-20) segment was represented using the free energy of base stacking computed according to the nearest-neighbor model using parameters reported in ref.³

S2.2. Impacts of single-stranded DNA property at the PAM-distal NTS segment on Cas9 cleavage rate

Note that reported a Cas9 ternary structure revealed that the unwound non-target-strand (NTS) has its PAM-distal segment squeezed out of the protein surface (e.g., pdb id:5Y36⁴). We therefore hypothesized that, in addition to DNA duplex stability beyond the RNA/DNA hybrid (Table S2, $\Delta G_{NN(17-20)}^0$), sequence-dependent flexibility at the single-stranded PAM-distal NTS segment may also influence Cas9 cleavage rate. As a first attempt, we represented the flexibility of the single-stranded NTS PAM+(17-20) segment as the free energy of base stacking computed according to the nearest-neighbor model using parameters reported in ref.³ (Table S2, $\Delta G_{nts(17-20)}^0$). For the 26 samples shown in Table S2, regression analysis gave an R_{adj}^2 value of 0.36 between $\ln(k_{cat})$ and $\Delta G_{nts(17-20)}^0$ (Figure S3A), indicating that $\Delta G_{nts(17-20)}^0$ by itself does not strongly influence on Cas9 activity. However, analysis on a model including $\Delta G_{NN(17-20)}^0$ and $\Delta G_{nts(17-20)}^0$ yielded an R_{adj}^2 of 0.76 (Figure S3B). This is higher than that obtained using just $\Delta G_{NN(17-20)}^0$ (R_{adj}^2 of 0.67, Main Text, Figure 2), indicating that the single-stranded NTS segment plays a role in modulating Cas9 activity with the 16-nucleotide guide. However, it remains unclear how well $\Delta G_{nts(17-20)}^0$ represents flexibility of the single-stranded DNA segment within the Cas9 complex. In addition, the R_{adj}^2 of 0.76 may indicate additional sequence-dependent factor(s) also play some roles in tuning Cas9 activity.

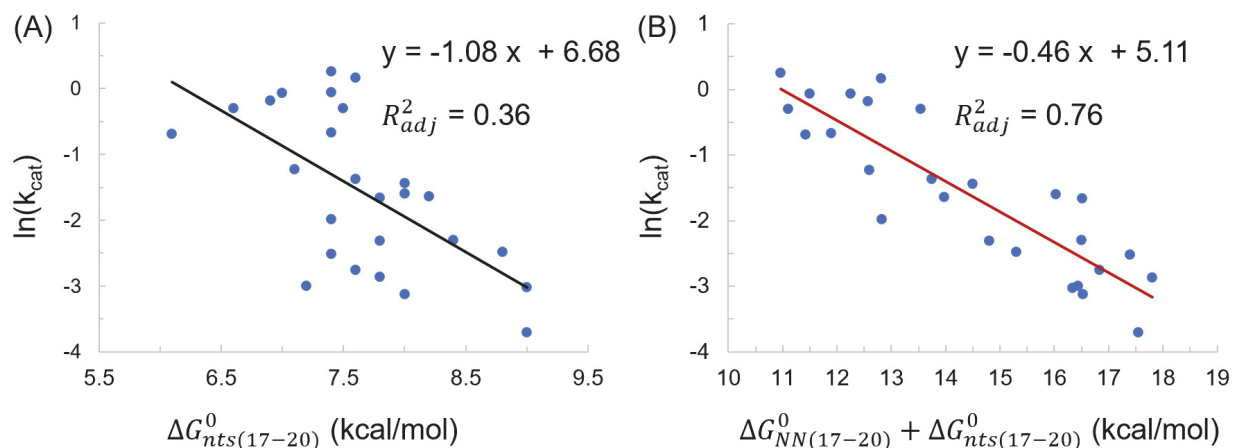


Figure S3. PAM-distal DNA NTS stacking free energy also contributes to Cas9 cleavage activity. (A) Linear regression analysis of $\ln(k_{cat})$ vs. $\Delta G_{nts(17-20)}^0$. (B). Linear regression of $\ln(k_{cat})$ vs. the sum of $\Delta G_{NN(17-20)}^0$ and $\Delta G_{nts(17-20)}^0$.

S3. Additional information on 2-amino-purine fluorescence measurements

Table S3 shows sequences of DNAs used for 2AP measurements. Table S4 lists fluorescence and absorbance data for 2AP related samples.

Table S3. Sequence of DNA strands used for 2AP measurements.

Name		Sequence (5' → 3') ^(a)
S_{TATA}	TS	GCTCAATTTTGACAGCCCA <u>CATGGCATTCCACTTCT</u> /i2AmPr/TATGGCATCCTTCCACTC
	NTS	GAGTGGAAGGATGCCATATAGAAAGTGGAAATGCCATGTGGGCTGTCAAAATTGAGC
S_{GATA}	TS	GCTCAATTTTGACAGCCCA <u>CATGGCATTCCACTTCG</u> /i2AmPr/TATGGCATCCTTCCACTC
	NTS	GAGTGGAAGGATGCCATATCGAAGTGGAAATGCCATGTGGGCTGTCAAAATTGAGC

(a) PAM+(1-20) positions are underlined, and the 2AP substitution is indicated with “i2AmPr /”

Table S4: F_{370}/A_{260} and $ratio(\varphi)$ for 2AP related samples.

Sample		F_{370}/A_{260} Ave $\pm$ SD ($\times 10^6$) ^(a)	Extinction coefficient ($L \cdot mol^{-1} \cdot cm^{-1}$)	$ratio(\varphi)$ Ave $\pm$ SD ^(b)
S_{TATA}	Duplex ^(c)	2.67 ± 0.14	893,507	1.00 ± 0.05
	ternary complex	9.25 ± 0.67	1,945,970	7.55 ± 0.55
S_{GATA}	Duplex ^(c)	2.29 ± 0.24	895,161	1.00 ± 0.10
	ternary complex	4.82 ± 0.56	1,947,624	4.58 ± 0.53

(a) Average and standard deviation obtained from three repeats.

(b) Standard deviation (SD) obtained from propagation of the standard deviations (i.e., “errors”) listed in the (F_{370}/A_{260}) column.

(c) Duplex sequences of S_{TATA} and S_{GATA} are shown in Table S3. The normalized quantum yields of the duplexes are different (2.67 ± 0.14 vs 2.29 ± 0.24), mainly due to 2-amino-purine having different surrounding nucleotides (T vs G at the 5' side, respectively).

S4. Dependence of activities of 16-nucleotide guides on KCl concentration

It is well known that a DNA duplex becomes more stable and harder to unwind with increasing monovalent salt concentrations, and extensive studies have established duplex free energy of formation depends on monovalent salt concentrations in a log-linear fashion, (i.e., $\Delta G_{NN}^0 \propto \ln([monovalent\ salt])$).¹ If duplex stability at the PAM+(17-20) segment indeed plays a significant role in modulating the Cas9 cleavage rates (i.e., $\ln(k_{cat}) \propto \Delta G_{NN}^0$, main text, Figure 2), one would then expect a linear correlation between $\ln(k_{cat})$ and $\ln([monovalent\ salt])$. To test this hypothesis, k_{cat} was measured for three target substrates, S_{TATA} , S_{TCAC} , and S_{GCGG} (Table S2) at three KCl concentrations (42 mM, 100 mM, and 200 mM). As shown in Figure S4, for each substrate, k_{cat} decreases as [KCl] increases, and $\ln(k_{cat})$ correlates with $\ln([KCl])$ in a highly linear fashion, with the adjusted R^2 all being greater than 0.90. Also note that while the three substrates have rather different PAM+(17-20) stability (i.e., different $\Delta G_{NN(17-20)}^0$, Table S2), the slope for the corresponding $\ln(k_{cat})$ vs. $\ln([KCl])$ plots are rather similar (Figure S4), further supporting the notion that the observed k_{cat} variations arise from a common factors, i.e., salt-dependent variations in DNA duplex stability. Overall, the observed KCl dependence of k_{cat} is in complete agreement with the conclusion that duplex stability at the PAM+(17-20) segment tunes the degree of DNA unwinding, subsequently modulates Cas9 cleavage rates.

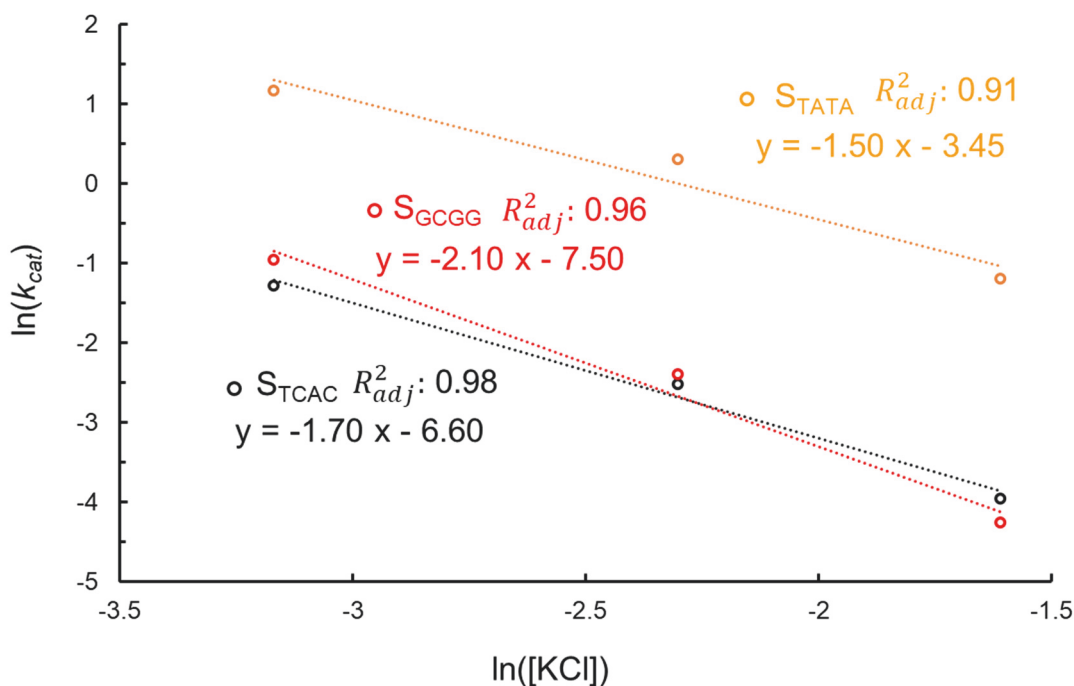


Figure S4: KCl dependence of measured k_{cat} for three target substrates.

S5. Additional information on impact of varying sequence of the 16-nt RNA guide

Figure S5 shows that the four additional data points presented in Figure 4 of the Main Text fit nicely to the linear regression model originally obtained from analyzing the 26 S_{NNNN} sequences. The results support the notion that PAM+(17-20) positions modulate Cas9 activity independent of the 16-nt guide sequence (i.e., the PAM+(1-16) segment involving in RNA/DNA interactions).

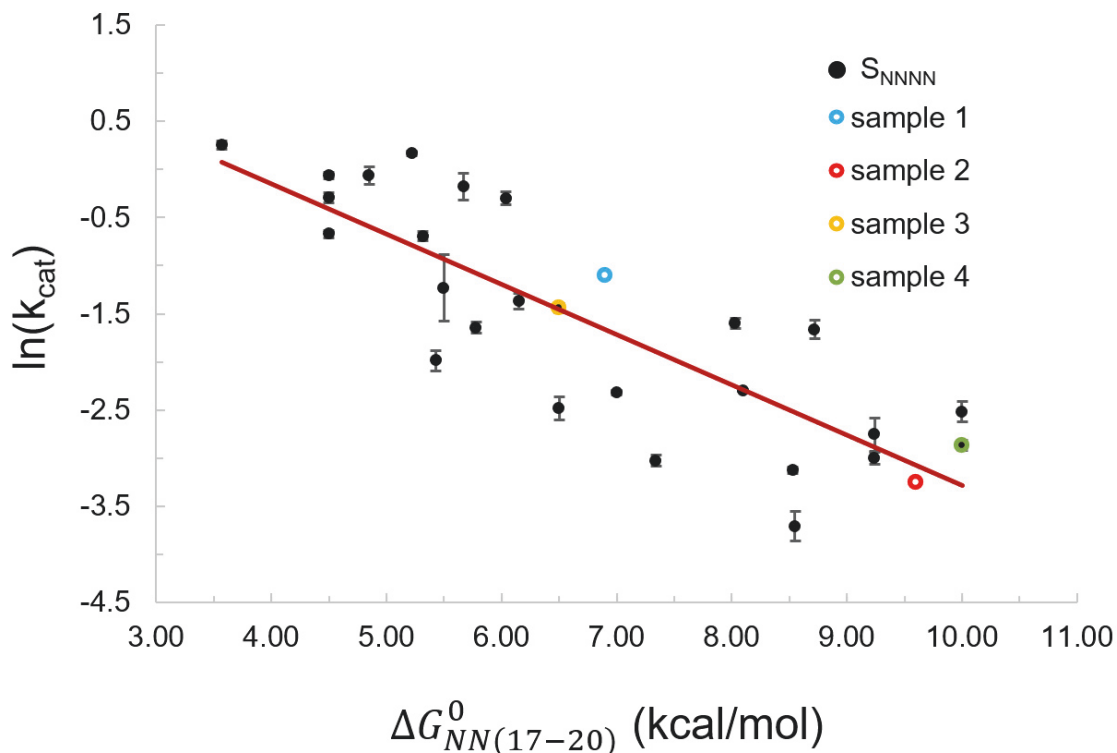


Figure S5: Correlation between $\ln(k_{cat})$ vs. $\Delta G^0_{NN(1-17)}$ for the expanded samples. Colored dots indicate data obtained for the four additional samples shown in Figure 4 of the Main Text. Black dots show the 26 data points presented in Figure 2 of the Main Text, with the corresponding linear fit shown as the red line.

References:

- (1) SantaLucia, J., Jr. A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proc. Natl. Acad. Sci. USA* **1998**, 95 (4), 1460–1465.
- (2) SantaLucia, J., Jr.; Allawi, H. T.; Seneviratne, P. A. Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability. *Biochemistry* **1996**, 35 (11), 3555–3562.
- (3) Pal, A.; Levy, Y. Structure, Stability and Specificity of the Binding of SsDNA and SsRNA with Proteins. *PLoS Comput. Biol.* **2019**, 15 (4), e1006768.
- (4) Huai, C.; Li, G.; Yao, R.; Zhang, Y.; Cao, M.; Kong, L.; Jia, C.; Yuan, H.; Chen, H.; Lu, D.; Huang, Q. Structural Insights into DNA Cleavage Activation of CRISPR-Cas9 System. *Nat. Commun.* **2017**, 8 (1), 1375.