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

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.

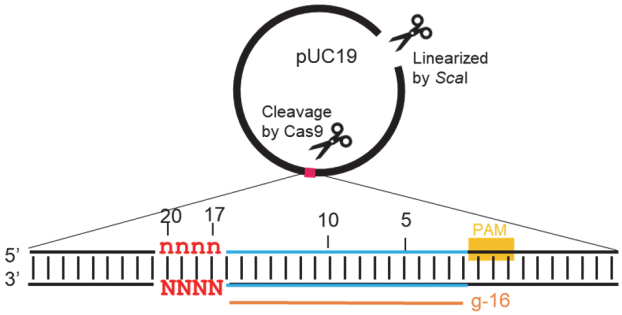

Figure S1. Schematic representation of the substrate DNA cloned into pUC19 vector between the *HindIII* and *BamHI* restriction sites.

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

Target		PAM+(16-21) sequence [a]	Primer	Sequence (5'→3') [b]
1	*S_{TATA}*	5′-CTATAT-3′ 3′-gatata-5′	P5-1_a_For	CCACATGGCATTCCACTTCTATATGGCATCCTTCCACTC
			P5-1_a_Rev	GAGTGGAAGGATGCCATATAGAAGTGGAATGCCATGTGGGC
2	*S_{TTAT}*	5′-CTTATT-3′ 3′-gaataa-5′	P5-1-10_For	CCACATGGCATTCCACTTCTTATTGGCATCCTTCCACTC
			P5-1-10_Rev	GAGTGGAAGGATGCCAATAAGAAGTGGAATGCCATGTGG
3	*S_{TAAA}*	5′-CTAAAT-3′ 3′-gattta-5′	P5-1-8_For	CCACATGGCATTCCACTTCTAAATGGCATCCTTCCACTC
			P5-1-8_Rev	GAGTGGAAGGATGCCATTTAGAAGTGGAATGCCATGTGG
4	*S_{TAAT}*	5′-CTAATT-3′ 3′-gattaa-5′	P5-1-9_For	CCACATGGCATTCCACTTCTAATTGGCATCCTTCCACTC
			P5-1-9_Rev	GAGTGGAAGGATGCCAATTAGAAGTGGAATGCCATGTGG
5	*S_{TTTT}*	5′-CTTTTT-3′ 3′-gaaaaa-5′	P5-1-d_For	CCACATGGCATTCCACTTCTTTTTGGCATCCTTCCACTC
			P5-1-d_Rev	GAGTGGAAGGATGCCAAAAAGAAGTGGAATGCCATGTGGGC
6	*S_{TAGA}*	5′-CTAGAT-3′ 3′-gatcta-5′	P5-1-11_For	CCACATGGCATTCCACTTCTAGATGGCATCCTTCCACTC
			P5-1-11_Rev	GAGTGGAAGGATGCCATCTAGAAGTGGAATGCCATGTGG
7	*S_{AAAA}*	5′-CAAAAT-3′ 3′-gtttta-5′	P5-1-e_For	CCACATGGCATTCCACTTCAAAATGGCATCCTTCCACTC
			P5-1-e_Rev	GAGTGGAAGGATGCCATTTTGAAGTGGAATGCCATGTGGGC
8	*S_{TATG}*	5′-CTATGT-3′ 3′-gataca-5′	P5-1-14_For	CCACATGGCATTCCACTTCTATGTGGCATCCTTCCACTC
			P5-1-14_Rev	GAGTGGAAGGATGCCACATAGAAGTGGAATGCCATGTGG
9	*S_{GATA}*	5′-CGATAT-3′ 3′-gctata-5′	P5-1-13_For	CCACATGGCATTCCACTTCGATATGGCATCCTTCCACTC
			P5-1-13_Rev	GAGTGGAAGGATGCCATATCGAAGTGGAATGCCATGTGG
10	*S_{TTCT}*	5′-CTTCTT-3′ 3′-gaagaa-5′	P5-1-18_For	CCACATGGCATTCCACTTCTTCTTGGCATCCTTCCACTC
			P5-1-18_Rev	GAGTGGAAGGATGCCAAGAAGAAGTGGAATGCCATGTGG
11	*S_{AAGA}*	5′-CAAGAT-3′ 3′-gttcta-5′	P5-1-12_For	CCACATGGCATTCCACTTCAAGATGGCATCCTTCCACTC
			P5-1-12_Rev	GAGTGGAAGGATGCCATCTTGAAGTGGAATGCCATGTGG

12	S_{TTGT}	5′-CTTGTT-3′ 3′-gaacaa-5′	P5-1-17_For	CCACATGGCATTCCACTTCTTGTTGGCATCCTTCCACTC
			P5-1-17_Rev	GAGTGGAAGGATGCCAACAAGAAGTGGAATGCCATGTGG
13	S_{ATTG}	5′-CATTGT-3′ 3′-gtaaca-5′	P5-1-16_For	CCACATGGCATTCCACTTCATTGTGGCATCCTTCCACTC
			P5-1-16_Rev	GAGTGGAAGGATGCCACAATGAAGTGGAATGCCATGTGG
14	S_{TCAC}	5′-CTCACT-3′ 3′-gagtga-5′	P5-1_For	CCACATGGCATTCCACTTCTCACTGGCATCCTTCC
			P5-1_Rev	GAGTGGAAGGATGCCAGTGAGAAGTGGAATGCCATG
15	S_{AGAC}	5′-CAGACT-3′ 3′-gtctga-5′	P5-1-20_For	CCACATGGCATTCCACTTCAGACTGGCATCCTTCCACTC
			P5-1-20_Rev	GAGTGGAAGGATGCCAGTCTGAAGTGGAATGCCATGTGG
16	S_{AGCT}	5′-CAGCTT-3′ 3′-gtcgaa-5′	P5-1-21_For	CCACATGGCATTCCACTTCAGCTTGGCATCCTTCCACTC
			P5-1-21_Rev	GAGTGGAAGGATGCCAAGCTGAAGTGGAATGCCATGTGG
17	S_{CCTC}	5′-CCCTCT-3′ 3′-gggaga-5′	P5-1-26_For	CCACATGGCATTCCACTTCCCTCTGGCATCCTTCCACTC
			P5-1-26_Rev	GAGTGGAAGGATGCCAGAGGGAAGTGGAATGCCATGTGG
18	S_{GGTC}	5′-CGGTCT-3′ 3′-gccaga-5′	P5-1-25_For	CCACATGGCATTCCACTTCGGTCTGGCATCCTTCCACTC
			P5-1-25_Rev	GAGTGGAAGGATGCCAGACCGAAGTGGAATGCCATGTGG
19	S_{TCGC}	5′-CTCGCT-3′ 3′-gagcga-5′	P5-1-27_For	CCACATGGCATTCCACTTCTCGCTGGCATCCTTCCACTC
			P5-1-27_Rev	GAGTGGAAGGATGCCAGCGAGAAGTGGAATGCCATGTGG
20	S_{CGCT}	5′-CCGCTT-3′ 3′-ggcgaa-5′	P5-1-29_For	CCACATGGCATTCCACTTCCGCTTGGCATCCTTCCACTC
			P5-1-29_Rev	GAGTGGAAGGATGCCAAGCGGAAGTGGAATGCCATGTGG
21	S_{GTCG}	5′-CGTCGT-3′ 3′-gcagca-5′	P5-1-28_For	CCACATGGCATTCCACTTCGTCGTGGCATCCTTCCACTC
			P5-1-28_Rev	GAGTGGAAGGATGCCACGACGAAGTGGAATGCCATGTGG
22	S_{CCCC}	5′-CCCCCT-3′ 3′-gggggga-5′	P5-1-f_For	CCACATGGCATTCCACTTCCCCCTGGCATCCTTCCACTC
			P5-1-f_Rev	GAGTGGAAGGATGCCAGGGGGGAAGTGGAATGCCATGTGGGC
23	S_{GGGG}	5′-CGGGGT-3′ 3′-gcccca-5′	P5-1-g_For	CCACATGGCATTCCACTTCGGGGTGGCATCCTTCCACTC
			P5-1-g_Rev	GAGTGGAAGGATGCCACCCCGAAGTGGAATGCCATGTGGGC
24	S_{CGGG}	5′-CGGGGT-3′ 3′-ggccca-5′	P5-1-30_For	CCACATGGCATTCCACTTCCGGGTGGCATCCTTCCACTC
			P5-1-30_Rev	GAGTGGAAGGATGCCACCCGGAAGTGGAATGCCATGTGG
25	S_{GCCG}	5′-CGCCGT-3′ 3′-gcggca-5′	P5-1-32_For	CCACATGGCATTCCACTTCGCCGTGGCATCCTTCCACTC
			P5-1-32_Rev	GAGTGGAAGGATGCCACGGCGAAGTGGAATGCCATGTGG
26	S_{GCGG}	5′-CGCGGT-3′ 3′-gcgcca-5′	P5-1_b_For	CCACATGGCATTCCACTTCGCGGTGGCATCCTTCCACTC
			P5-1_b_Rev	GAGTGGAAGGATGCCACCGCGAAGTGGAATGCCATGTGGGC

(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.

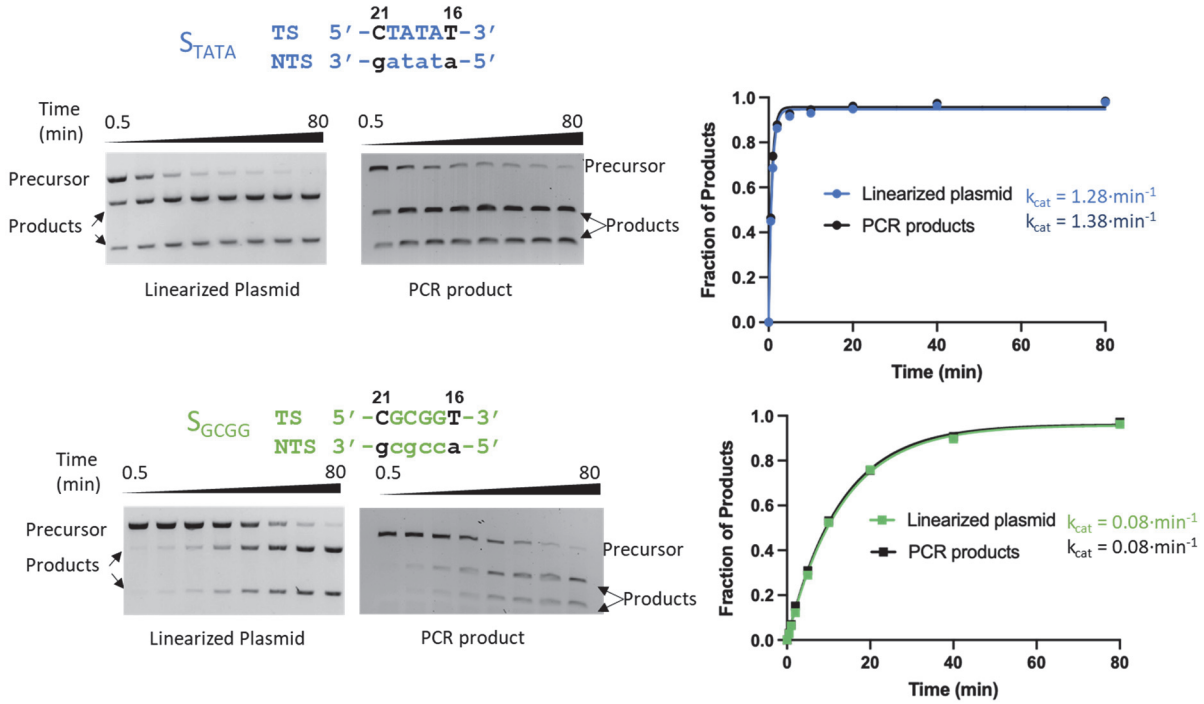

Figure S2. Assessing impact on measured cleavage rates by peripheral DNA sequences beyond the 20-nucleotide protospacer. Two sequences were tested, one with small $\Delta G^0_{NN(17-20)}$ (top panels), and the other with large $\Delta G^0_{NN(17-20)}$ (bottom panels). For both sequences, the cleavage 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 $MgCl_2$ (Main Text, Methods). However, such differences in the salt concentration result in an offset of ΔG^0_{NN} 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^0_{NN}.

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
				$\Delta H^0_{(17-20)}$ [c] (kcal/mol)	$\Delta S^0_{(17-20)}$ [c] (cal/mol/K)	$\Delta G^0_{NN(17-20)}$ [d] (kcal/mol)	$\Delta G^0_{nts(17-20)}$ [e] (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′-gatta-5′	0.94±0.03	35.7	100.6	4.50	7.0
5	S_{TTTT}	5′-CTTTTT-3′ 3′-gaaaa-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′-gaaga-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′-gaaca-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′-gtcga-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′-ggcga-5′	0.04±0.00	42.4	109.2	8.53	8.0

21	S_{GTCG}	5′-CGTCGT-3′ 3′-gcagca-5′	0.19±0.02	45.1	117.3	8.72	7.8
22	S_{CCCC}	5′-CCCCCT-3′ 3′-ggggga-5′	0.02±0.00	32.9	78.5	8.55	9.0
23	S_{GGGG}	5′-CGGGGT-3′ 3′-gcccca-5′	0.05±0.00	38.8	95.3	9.24	7.2
24	S_{CGGG}	5′-CCGGGT-3′ 3′-ggccca-5′	0.06±0.01	38.8	95.3	9.24	7.6
25	S_{GCCG}	5′-CGCCGT-3′ 3′-gcggca-5′	0.06±0.00	46.6	118.0	10.00	7.8
26	S_{GCGG}	5′-CGCGGT-3′ 3′-gcgcca-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) $\Delta G^0_{NN(17-20)}$ = ΔH^0 - TΔS^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 unwounded non-target-strand (NTS) has its PAM-distal segment squeezed out of the protein surface (e.g., pdb id:5Y36[4]). We therefore hypothesized that, in addition to DNA duplex stability beyond the RNA/DNA hybrid (Table S2, $\Delta G^0_{NN(17-20)}$), 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, $\Delta G^0_{nts(17-20)}$). For the 26 samples shown in Table S2, regression analysis gave an R^2_{adj} value of 0.36 between ln(k_{cat}) and $\Delta G^0_{nts(17-20)}$ (Figure S3A), indicating that $\Delta G^0_{nts(17-20)}$ by itself does not strongly influence on Cas9 activity. However, analysis on a model including $\Delta G^0_{NN(17-20)}$ and $\Delta G^0_{nts(17-20)}$ yielded an R^2_{adj} of 0.76 (Figure S3B). This is higher than that obtained using just $\Delta G^0_{NN(17-20)}$ (R^2_{adj} 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 $\Delta G^0_{nts(17-20)}$ represents flexibility of the single-stranded DNA segment within the Cas9 complex. In addition, the R^2_{adj} of 0.76 may indicate additional sequence-dependent factor(s) also play some roles in tuning Cas9 activity.

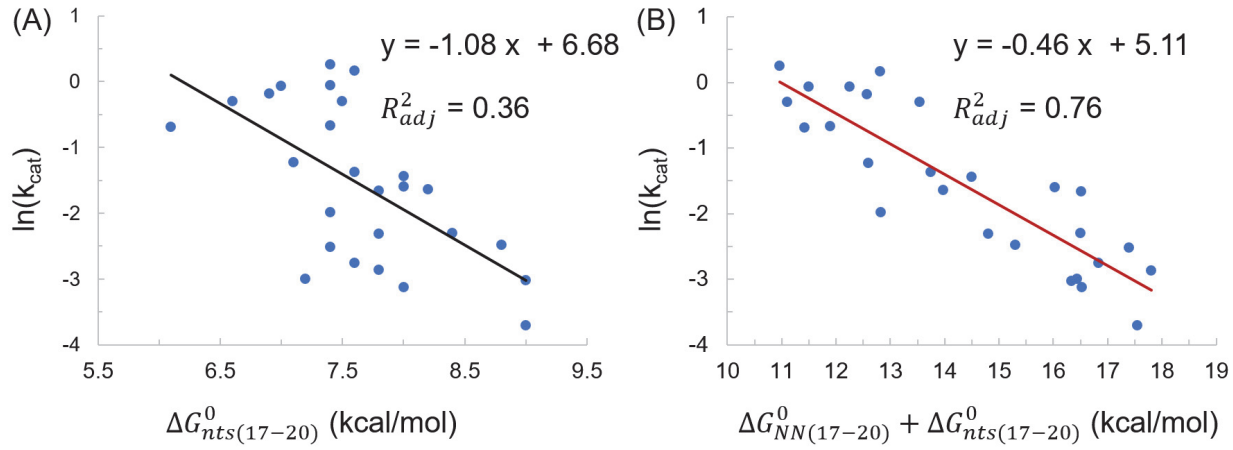

Figure S3. PAM-distal DNA NTS stacking free energy also contributes to Cas9 cleavage activity. (A) Linear regression analysis of ln(k_{cat}) vs. $\Delta G^0_{nts(17-20)}$. (B). Linear regression of ln(k_{cat}) vs. the sum of $\Delta G^0_{NN(17-20)}$ and $\Delta G^0_{nts(17-20)}$.

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	GCTCAATTTTGACAGCCCACATGGCATTCCACTTCT/i2AmPr/TATGGCATCCTTCCACTC
	NTS	GAGTGGAAGGATGCCATATAGAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGC
S_{GATA}	TS	GCTCAATTTTGACAGCCCACATGGCATTCCACTTCG/i2AmPr/TATGGCATCCTTCCACTC
	NTS	GAGTGGAAGGATGCCATATCGAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGC

(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 ± SD (x 10^6)[(a)]	Extinction coefficient (L·mol^{-1}·cm^{-1})	$ratio(\varphi)$ Ave ± 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., $\Delta G^0_{NN} \propto \ln([monovalent\ salt])$.[1] If duplex stability at the PAM+(17-20) segment indeed plays a significant role in modulating the Cas9 cleavage rates (i.e., $\ln(k_{cat}) \propto \Delta G^0_{NN}$, 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 $\Delta G^0_{NN(17-20)}$, 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.

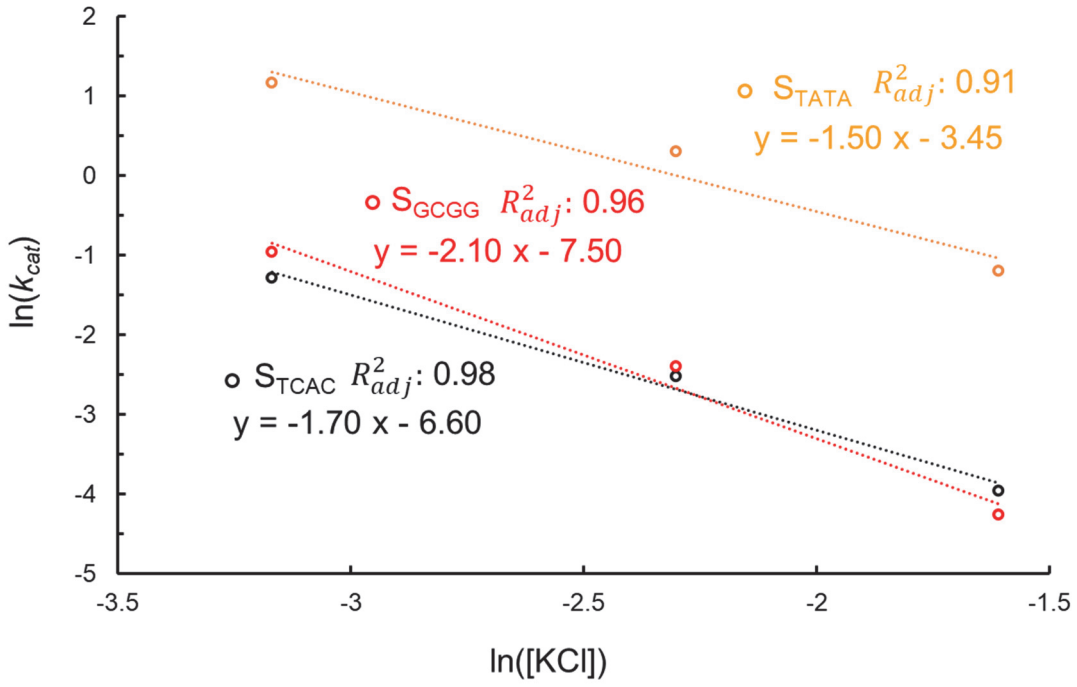

Figure S4: KCl dependence of measured k_{cat} for three target substrates.

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).

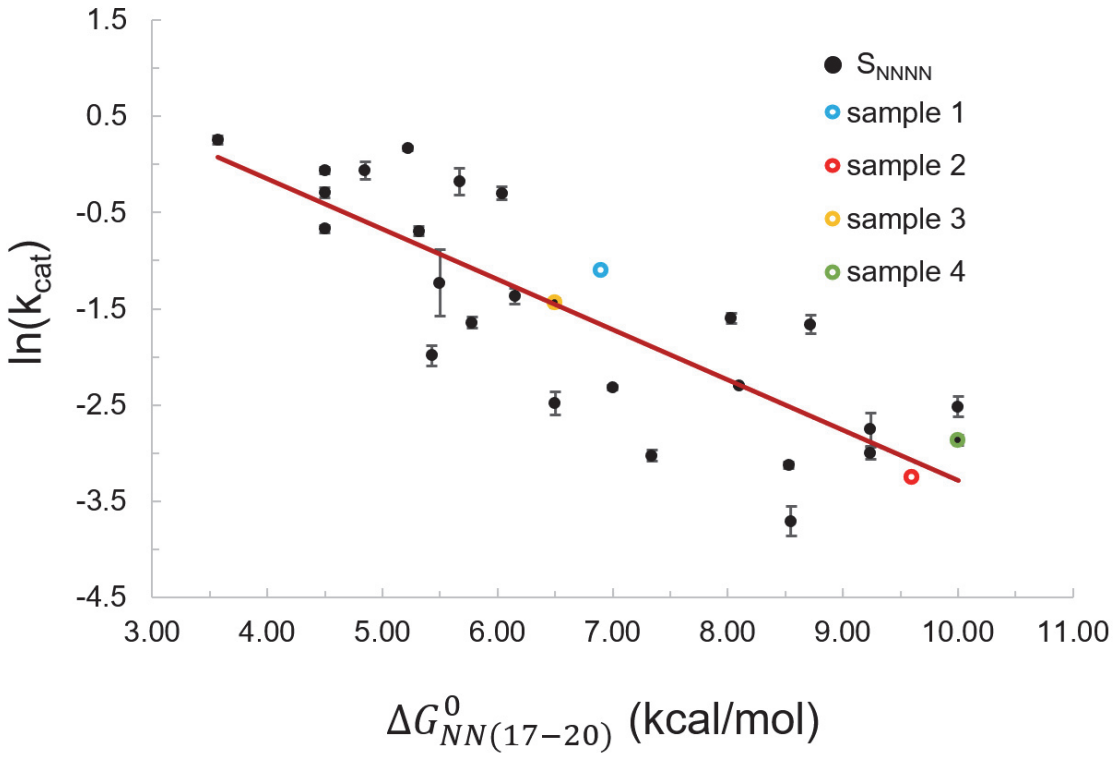

Figure S5: Correlation between ln(k_{cat}) vs. $\Delta G^0_{NN(1-17)}$) for the expanded samples. Colored dots indicate data obtained for the four additional samples shown in Figure 4 of the Main Text. Black dots show the 26 data points presented in Figure 2 of the Main Text, with the corresponding linear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