Unraveling the mechanisms of PAMless DNA interrogation by SpRY Cas9

Grace N. Hibshman[1,2]*, Jack P. K. Bravo[1]*†, Hongshan Zhang[3], Tyler L. Dangerfield[1], Ilya J. Finkelstein[3], Kenneth A. Johnson[1], David W. Taylor[1-4]†

[1]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, 78712, USA

[2]Interdiscplinary Life Sciences Graduate Programs, Austin, TX 78712, USA

[3]Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX, 78712, USA

[4]LIVESTRONG Cancer Institutes, Dell Medical School, Austin, TX, 78712, USA

*Co-first authors.

†Correspondence to jack.bravo@austin.utexas.edu (JPKB) and dtaylor@utexas.edu (DWT).

Abstract

CRISPR-Cas9 is a powerful tool for genome editing, but the strict requirement for an "NGG" protospacer-adjacent motif (PAM) sequence immediately adjacent to the DNA target limits the number of editable genes. To overcome the PAM requirement, a recently developed Cas9 variant, called SpRY-Cas9 was engineered to be "PAMless" (1, 2). However, the molecular mechanisms of how SpRY can recognize all potential PAM sequences and still accurately identify DNA targets have not been investigated. Here, we combined enzyme kinetics, cryo-EM, and single-molecule imaging to determine how SpRY interrogates DNA and recognizes target sites for cleavage. Divergent PAM sequences can be accommodated through conformational flexibility within the PAM-interacting region of SpRY, which facilitates tight binding to off-target DNA sequences. Once SpRY correctly identifies a target site, nuclease activation occurs ~1,000-fold slower than for *Streptococcus pyogenes* Cas9, enabling us to directly visualize multiple on-pathway intermediate states. Insights gained from our intermediate structures prompted rationally designed mutants with improved DNA cleavage efficiency. Our findings shed light on the molecular mechanisms of PAMless genome editing with SpRY and provide a framework for the design of future genome editing tools with improved versatility, precision, and efficiency.

Introduction

CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats, CRISPR-associated) is an RNA-guided endonuclease that uses RNA-DNA complementarity to precisely target and cleave double-stranded DNA (3). To find its target, Cas9 must meticulously search double-stranded DNA across the entire genome to identify regions of gRNA complementarity (4, 5). This search is simplified via the protospacer-adjacent motif (PAM) located immediately next to the targeted sequence (6, 7). The PAM acts as a rapid and efficient initial filter for possible target sites, circumventing the slow and energetically unfavorable process to initiate duplex melting. In the case of *Streptococcus pyogenes* Cas9 (henceforth referred to as Cas9), the required PAM sequence is NGG (where N can represent any nucleotide and G signifies guanosine). In its native context, the PAM also serves as a mechanism to distinguish between self and non-self sequences (8–10).

For genome editing applications, the PAM requirement imposes a major limitation on the number of sequences that can be targeted (2). Protein engineering efforts have focused on re-targeting or reducing the PAM specificity (11–16). A PAMless Cas9 variant, termed SpRY, is unique in that it is capable of targeting nearly any site *in vivo* (1). However, since PAM recognition is a prerequisite for efficient DNA melting and R-loop formation, it is unclear how SpRY can identify diverse target sequences.

Promiscuous PAM recognition by SpRY

The Cas9 residues R1333 and R1335 confer specificity for two guanines in the second and third positions of the PAM by forming four hydrogen bonds with the Hoogsteen faces (8). Both residues are mutated in SpRY (R1333P, R1335Q) along with nine additional mutations (A61R, L1111R, D1135L, S1136W, G1218K, E1219Q, N1317R, A1322R, R1333P, R1335Q, and T1337R) that enable PAMless DNA targeting.

To understand how SpRY can recognize diverse PAM sequences, we determined the structures of SpRY bound to NGG, NAC, and NTC PAM-containing DNA substrates with global resolutions of 3.3, 2.8 and 3.7 Å, respectively (**Fig. 1 and Extended Data Fig. 1, Extended Data Table 1**). All three structures were in the product state, with HNH repositioned at the cleaved scissile phosphate of the target strand (TS).

Comparison of these atomic models with a previously determined structure of Cas9 in the product state (PDB 7S4X (17)) shows that most of the complex is unaltered (RMSD of <1Å), except for the PAM-interacting (PI) domain. The mutated residues L111R, G1218K, E1219Q, R1335Q, and T1337R make several electrostatic interactions with the target DNA duplex and increase the overall net charge of the PI domain relative to Cas9 (**Fig. 1d-g**). This change in protein surface potential may stimulate the formation of non-specific electrostatic interactions with DNA to compensate for the loss of specific hydrogen bond contacts with the PAM nucleobases, enabling DNA interrogation.

Structures of the PAM-reprogrammed Cas9 variants VQR and VRQR exhibit a slight shift in the PAM region of the duplex backbone by ~1-2 Å (18, 19). We observed a similar phenomenon in SpRY where the DNA backbone is shifted by up to 6 Å relative to the equivalent position for Cas9. This shift in the DNA backbone is mediated by the bulky, hydrophobic mutations D1135L and S1136W wedged into the minor groove.

In addition to the SpRY mutations that directly engage with the PAM, three more mutations surround the PAM readout region. N1317R and A1322R form electrostatic interactions with the phosphate backbone of the displaced NTS immediately adjacent to the PAM. These interactions likely favor initial unwinding of the DNA duplex. Additionally, A61R forms ionic interactions with E1108 at the base of the bridge helix and U63 of the gRNA scaffold, directly behind the first base of the R-loop, potentially supporting R-loop initiation (**Extended Data Fig. 2**).

Structures of SpRY in complex with NGG, NAC, and NTC PAM sequences are largely similar. Surprisingly, G1218K, N1317R, A1322R, and T1337R adopt different rotamers in the structures to maximize the energetics of binding through the formation of non-specific interactions compatible with the different PAM sequences (**Fig. 1h-j and Extended Data Fig. 2**). This is in stark contrast to Cas9 in which R1333 and R1335 are pre-ordered and poised to quickly recognize a PAM. This observation reveals a mechanism in which SpRY recognizes DNA targets through a mechanism predominantly reliant on electrostatic contacts rather than base-specific contacts.

SpRY DNA cleavage is limited by reduced rates of DNA unwinding

Next, we analyzed DNA cleavage kinetics across different PAM sequences using the same framework we previously employed to directly compare high-fidelity variants to wild type Cas9 (20) (**Fig. 2**). Others have demonstrated essentially PAMless activity *in vitro* with assays using all possible PAM sequences for SpRY (2). We focused our efforts on NGG and two alternative PAM sequences only accessible to SpRY for more detailed kinetic analysis. In this way, we can directly compare SpRY and Cas9, and explore the efficiency of SpRY across diverse sequences.

We first measured the rate of R-loop dependent DNA cleavage by SpRY, by mixing preformed SpRY:gRNA complex with DNA in the presence of MgCl$_2$. We observed that SpRY cleaved the TS at a rate of 0.0025 s^{-1}, and the NTS at a rate of 0.0023 s^{-1}, corresponding to a ~500-fold slower observed rate than that of Cas9 (21). The slow observed rate of cleavage was not unique to the NGG PAM sequence, as SpRY cleaved the TS of NGC and NAC PAM substates with similarly slow rates (**Fig. 2b, Extended Data Figs. 3 & 4**).

To determine whether the slow rate was due to a reduced rate of the chemical step or due to a slow step preceding chemistry, we measured R-loop independent cleavage rates by rapid quench since the reactions were too fast to resolve by hand mixing. In these experiments SpRY:gRNA was pre-incubated with DNA in the absence of magnesium and in the presence of a low concentration of EDTA to chelate any trace metal ions. The reaction was then initiated by adding MgCl$_2$. We observed cleavage rates of 6 s^{-1} for the TS compared to 4.3 s^{-1} for Cas9 (21). Cleavage of the NTS for SpRY was biphasic with rates of 4.3 s^{-1} and 0.54 s^{-1} for the fast and slow phases, respectively, compared to an observed rate of 3.5 s^{-1} for Cas9 (**Fig. 2b**).

This discrepancy between R-loop dependent and independent cleavage strongly suggests that the cleavage defect is due to reduced rates of R-loop formation. To test this, we directly measured the rate of R-loop formation using a stopped-flow assay for DNA unwinding. A DNA substrate containing the fluorescent tricyclic cytosine (tC^o) analog at either position 1 or position 16 on the NTS was used to report on R-loop initiation and completion. As the R-loop propagates, the NTS is displaced resulting in an increase in tC^o fluorescence.

In an experiment mirroring the R-loop dependent cleavage experiment described above but performed with the fluorescent DNA substrate in the stopped flow, SpRY initiated DNA unwinding of NGG, NGC and NAC substrates with nearly identical rates (0.015, 0.022 and 0.046 s^{-1}, respectively). R-loop completion was ~3-8-fold slower (0.004, 0.006, and 0.006 s^{-1}, respectively) (**Fig. 2c**). The difference between the rate of unwinding at these two positions provides direct evidence that SpRY unwinds DNA in a directional manner, from PAM-proximal to -distal, akin to Cas9. Hence, the decreased observed rate of DNA cleavage for SpRY is due to the decreased rate of DNA unwinding.

To fully understand the kinetics and mechanism of SpRY, and how it contrasts with Cas9, all the experiments for SpRY cleaving the NGG PAM substrate were fit globally to a single unified model based on numerical integration of the rate equations in KinTek Explorer (KinTek Corporation; kintekexplorer.com) (**Fig. 2d, Extended Data Fig 5, Extended Data Table 3**). We obtained a global fit to the data that could be explained by our model and a free energy profile was constructed for comparison with Cas9.. The surprising result from this analysis is that the highest barrier relative to the starting material for SpRY, defining the specificity determining step, is R-loop formation. For wild-type Cas9, barrier heights for each step were similar, suggesting that specificity is a function of all steps leading up to the chemical step.

Structural basis of target search and R-loop formation by SpRY

Due to the rapid rate of R-loop formation and DNA cleavage by Cas9, it has only been possible to capture intermediates of this process through covalent crosslinking, DNA substrates with limited complementarity, or catalytically dead enzyme. We reasoned that the significantly reduced enzyme kinetics may make it feasible to determine structures of SpRY during R-loop propagation without the use of non-productive substrates or chemical crosslinkers.

Contrary to typical cryo-EM sample preparation, where samples are optimized for maximal homogeneity and stability, we instead chose to prepare our sample under conditions that enrich for maximal conformational heterogeneity to trap an ensemble of on-pathway intermediate states. Based on our kinetic data, we prepared cryo-EM samples one minute after mixing SpRY:gRNA with NAC PAM DNA. At this time point, ~90% of the DNA is unwound at position 1, less than 50% of the DNA is unwound at position 16, and only ~10% of the DNA is cleaved. This approach captures transient states of SpRY as it initiates, propagates, and ultimately completes R-loop formation to achieve DNA cleavage. This cryo-EM dataset yielded seven high-resolution (2.9-3.4Å) structures of SpRY at distinct states along the directional R-loop propagation pathway: 0 bp, 2 bp, 3bp, 10 bp, 13 bp, 18 bp and 20 bp (product state) (**Fig. 3a and Extended Data Fig. 6**).

Since SpRY lacks a PAM requirement to bind DNA targets, it must instead identify targets through probing for sequence complementarity to the gRNA. The 0 bp structure reveals that SpRY interrogates dsDNA targets by introducing a ~50° kink in the center of the target duplex, distorting the phosphodiester backbone immediately adjacent to the 3' end of the gRNA spacer (**Fig. 3a**). This structure is highly similar to a recently determined structure of Cas9 covalently crosslinked to non-complementary DNA (**Extended Data Fig. 7**). Interestingly, we did not observe an open protein/linear DNA conformation, suggesting that this conformation may be too transient to capture or a product of crosslinking.

The most populated SpRY R-loop intermediate has two bases of the TS base pairing with the gRNA (**Fig. 3a**). Within our 2 bp structure, we resolved a contact between D269 and position 3 of the gRNA spacer. This contact appears to act as a physical barrier for R-loop propagation, tethering REC2 to the seed region of the gRNA and blocking R-loop propagation (**Fig. 3b**). Transition to later R-loop intermediates is accompanied by the displacement of REC2 by ~40Å, creating a channel to accommodate the propagating R-loop (**Fig. 3c**). Mutation of D269 to an alanine residue increased the rate of TS cleavage by ~4-fold for the NGG PAM, confirming that disruption of this contact accelerates R-loop propagation (**Fig. 3d**). This physical barrier may be important for probing initial complementarity by making base-pairing between the gRNA and mismatch-containing duplexes unfavorable. Displacing D269 as the R-loop propagates to 3 bp unlocks REC2 from its binary position and enables further R-loop propagation (**Fig. 3a**).

The next intermediate resolved had 10 bp of R-loop formed. In this structure, the REC2 domain has been displaced, shifting upwards by up to 40 Å and REC3 is poorly resolved due to conformational flexibility. By 13 bp of R-loop formation, REC2 and REC3 are well-resolved and stabilized by R-loop contacts.

The 18 bp structure has a near-complete R-loop and exhibits partial density for the HNH domain repositioned at the scissile phosphate of the TS. When we examined the catalytic sites, we found that cleavage had not occurred at either HNH or RuvC. There is density for the required Mg^{2+} ion in the HNH active site, yet at the RuvC active site, there is only density for one of the two required Mg^{2+} ions. This structure represents a state just prior to cleavage where one of the Mg^{2+} ions is not yet properly coordinated for cleavage.

The distribution of R-loop duplex lengths we captured represent kinetic barriers for R-loop propagation. Large conformational changes are associated with each observed intermediate. While the 0 bp and 2 bp structures largely resemble the Cas9 binary complex (RMSD 1.88 Å), the transition to the 3 bp structure requires the seismic eviction of the REC2 domain. Propagation to 10 bp necessitates movement of the REC3 domain, evidenced by the lack of density in our reconstruction. By 18 bp of heteroduplex formation, HNH begins docking into the active state resulting in DNA cleavage as the R-loop completes 20 bp. Together these structures support a model of R-loop formation distinguished by structural checkpoints, as previously proposed (16, 17, 22, 23).

SpRY tightly binds to non-target DNA sequences

During 2D classification of SpRY in complex with the NAC PAM substrate cryo-EM dataset, we observed a subset of particles that were much larger than anticipated. 3D reconstruction of these particles revealed the presence of a dimeric SpRY complex, with two SpRY gRNPs bound to the same DNA substrate – we refer to this structure as a "SpRYmer" (**Fig. 4b**). Within the SpRYmer structure, the first SpRY is in the product state, with both TS and NTS cleaved. The second SpRY resembles the binary complex, likely due to the lack of substrate complementarity necessary to initiate DNA unwinding. SpRYmer assembly is stabilized by direct linkage through the DNA and by a kissing loop interaction formed between the nexus stem loop of each gRNP (**Extended Fig. 8**). Therefore, we hypothesized that SpRY may have the capacity to stably bind to non-target DNA sequences.

To validate this, we measured the apparent binding affinity ($K_{D(app)}$) of SpRY and Cas9 to an NGG PAM DNA substrate lacking complementarity to the gRNA. As anticipated, Cas9 showed no discernable binding within the concentration regime measured (up to 1000 nM) (**Fig. 4c**). This measurement agrees with previous reports where Cas9 is weakly bound to PAM sequences with an apparent affinity >10 μM (22). Strikingly, SpRY bound to this substrate with a $K_{D(app)}$ of 5.6 nM, $\sim$2000-fold more tightly than Cas9.

To further explore this disparity in non-target DNA binding, we prepared cryo-EM samples of SpRY in complex with the same non-target DNA substrate. It became immediately clear during classification and refinement that this dataset contained a plethora of conformationally heterogeneous structures, which we separated through 3D classification. We resolved four distinct structures from this dataset, corresponding to 1 bp, 6 bp, 8 bp, and 10 bp of R-loop formation, where the 6 bp, 8 bp, and 10 bp structures contain mismatches (**Fig. 4d and Extended Data Fig. 9**).

Close inspection of the DNA sequence revealed a continuous stretch of four nucleotides with complementarity to the gRNA located within the NTS outside of the designated protospacer region. We reasoned that SpRY can associate with either DNA strand during target search, unlike Cas9 which requires the PAM sequence for proper orientation. The mismatch-containing 6, 8, and 10 bp structures align with this region of the DNA sequence where the first four bases of the R-loop are complementary to the gRNA with subsequent mismatches. Like our R-loop intermediates, we observe REC2 and REC3 movement in these structures as the R-loop propagates past that structural checkpoint.

Our structures support a model where the REC2 and REC3 domains must be displaced during R-loop formation to accommodate the propagating gRNA:TS duplex, and then repositioned to facilitate R-loop completion. The lack of PAM specificity enables SpRY to locate partial target sequences from which it cannot readily dissociate, nor achieve nuclease activation.

8

We propose that this high affinity binding to non-target DNA sequences is a direct consequence of the SpRY mutations which create a large, positively charged patch within the PI domain. This patch acts as an anchor for DNA binding, facilitating stable, non-specific electrostatic contacts with duplexes prior to unwinding. This novel property of SpRY compensates for the loss of low affinity, rapid PAM probing employed by Cas9, and supports dwell times long enough to search for complementarity to the gRNA.

Single-molecule visualization of DNA interrogation by SpRY

Next, we used single-molecule DNA curtains to directly visualize DNA interrogation of dSpRY:gRNA on long DNA molecules (PMID:28668123) (**Fig. 5**). dSpRY and dCas9 were fluorescently labeled with anti-FLAG coupled quantum dots, as previously described (PMCID PMC4106473, https://doi.org/10.1101/2022.09.15.508177). We first injected Cas9 gRNP into single-tethered DNA curtains, and then flushed out excess protein with high salt buffer (**Fig. 5a-c**). As expected, dCas9 gRNPs preferentially bound to the target site (PMCID PMC4106473). In contrast, dSpRY gRNPs bound non-specifically along the entire length of the DNA with no significant enrichment at the target site (**Fig. 5a-c**).

We next used a double-tethered DNA curtain to visualize transient interactions with non-target DNA. In these assays, both DNA ends are affixed to microfabricated barriers, allowing direct observation of protein binding and release along the length of the DNA without buffer flow. This enables us to record the locations and corresponding lifetimes of all binding events (**Fig. 5d**). When bound to the target DNA sequence, the lifetime distributions for dSpRY and dCas9 were nearly identical ($t_{1/2}$ of 360±10s and 360±5s, respectively) (**Fig. 5e & f**). In contrast, dCas9 had a relatively short binding lifetime on off-target DNA, and dSpRY remained associated with nearly the same lifetime as at the on-target sequence (**Fig 5g**). These results substantiate that the lack of a PAM requirement drastically slows target search kinetics.

Discussion

While CRISPR-Cas9 has found widespread use as a powerful tool for programmable gene editing, the identification of target sequences by Cas9 represents a formidable challenge, akin to locating a needle within a haystack. This search predicament is simplified by the NGG PAM requirement, which significantly reduces the number of sequences that are probed for gRNA complementarity. Cas9 rapidly probes PAM sequences for an average of ~30 ms, and far less at non-target sequences (5, 24). PAM recognition is a prerequisite for R-loop formation, which is the most significant kinetic bottleneck for DNA cleavage by Cas9 (20, 21). This process relies on the energetically unfavorable disruption of base-pairing and base stacking to expose ssDNA. Once a PAM sequence is located, the adjacent duplex is distorted to promote local base flipping and R-loop nucleation to search for gRNA complementarity (22).

Our work reveals a substantial shift in the energetic landscape governing how SpRY interacts with DNA targets differently than Cas9. As it lacks a PAM specificity, SpRY must initiate DNA melting to explore gRNA complementarity, as exemplified by our 0 bp R-loop structure (**Fig. 3a**). Considering the topological complexity of chromatin (25–27), SpRY is likely to exhibit a preference for targeting regions with low energetic barriers of DNA melting such as bends and active genes as was previously suggested (28, 29). These findings shed light on the mechanism of SpRY targeting and its propensity for specific genomic locations.

The recycling of Cas9 *in vivo* is facilitated by the recruitment of chromatin remodelers and DNA-repair factors after cleavage (5, 30–32). Unlike Cas9, SpRY stably associates with non-target sequences, as revealed by ensemble biochemistry and single-molecule imaging experiments. The accumulation of SpRY at non-target sites would result in a reduction in the effective concentration of active enzyme until it is evicted by cellular factors (33). Although this phenomenon may not drastically increase the occurrence of non-target double-strand breaks due to insufficient gRNA complementarity necessary for HNH realignment and DNA cleavage (17), it does have the potential to induce undesirable gene silencing. In comparison, Cas9 requires a PAM and at least five nucleotides of seed complementarity to stably associate with non-target sequences. This stable association of Cas9 with non-target sequences has been demonstrated to cause cellular toxicity (28, 34). These observations emphasize the importance of carefully considering the implications of SpRY and Cas9 behavior to mitigate off-target effects and enhance the safety of gene editing applications.

In conclusion, despite its limitations in cleavage efficiency, SpRY proves to be an asset in modifying genomic regions that were previously inaccessible to Cas9. Recent successful gene editing in stem cells, plants and model organisms exhibits its effectiveness (35–37). Through our structural analysis, we identified a mutation that accelerates gRNA:TS duplex formation and increases DNA cleavage efficiency (**Fig 3**). Combining this mutation with other hyperactive Cas9 mutations holds the potential to compensate for SpRY's reduced activity (38). The creation of chimeric CRISPR-Cas9

11

variants, where mutants with diverse purposes are combined, could forge the development of highly specific and adaptable gene editing tools that transcend biological constraints. Our discoveries will serve as a valuable foundation for the generation of potent chimeric Cas9 variants advancing the field toward the most effective gene editing tool.

Competing Interests

K.A.J. is the President of KinTek, Corp., which provided the chemical-quench flow and stopped flow instruments and the KinTek Explorer software used in this study.
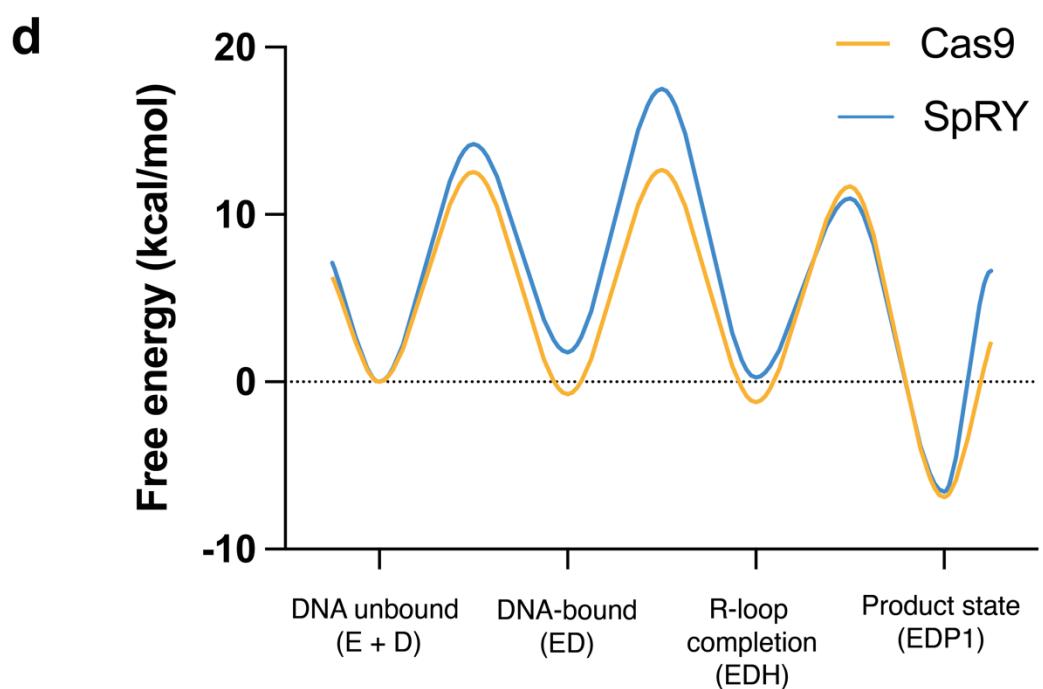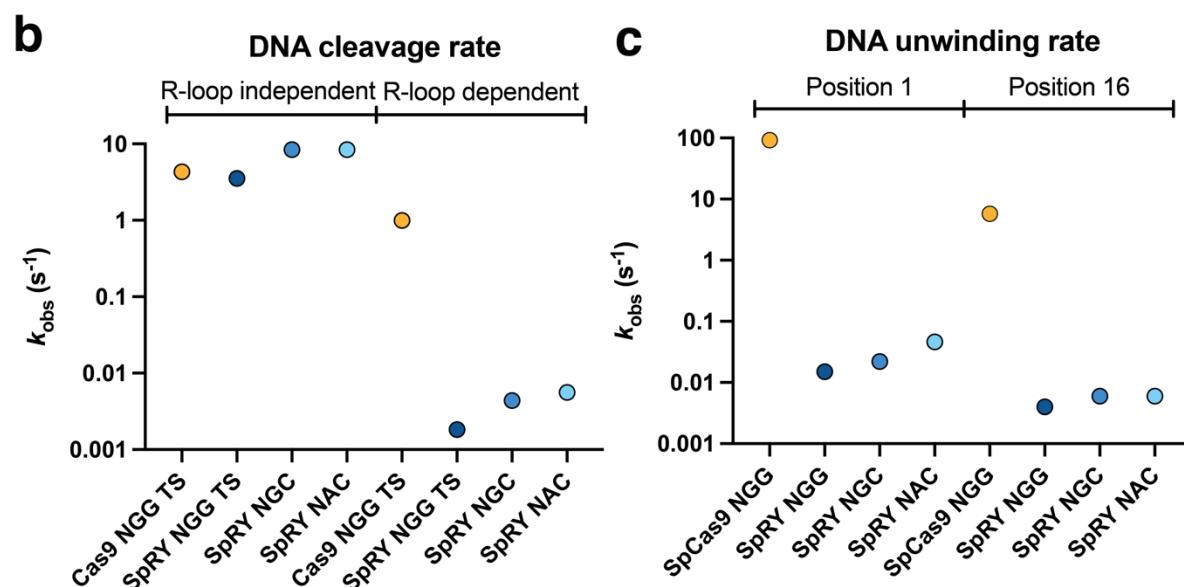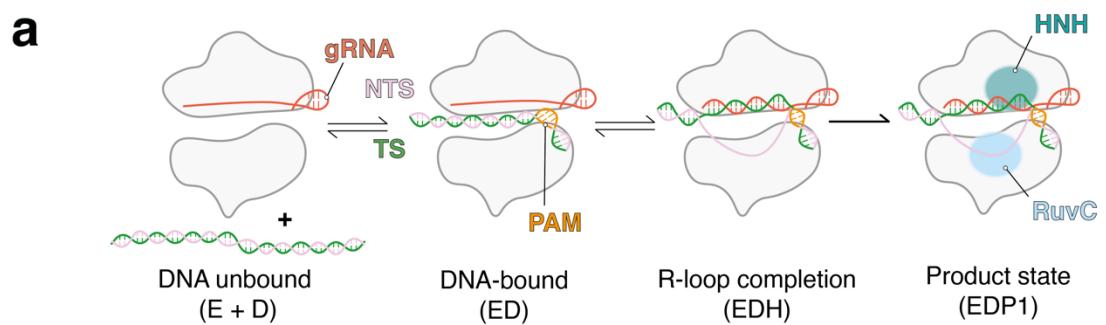
Data Availability

The structures and associated atomic coordinates have been deposited into the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB) with accession codes SpRY NAC PAM 20 bp (EMD-40740 and PDB 8SRS), SpRY NGG PAM (EMD-40681 and PDB 8SPQ), SpRY NTC PAM (EMD-40705 and PDB 8SQH), SpRY NAC PAM 0 bp (EMD-41073 and PDB 8T6O), SpRY NAC PAM 2 bp (EMD-41074 and PDB 8T6P), SpRY NAC PAM 3 bp (EMD-41085 and PDB 8T76), SpRY NAC PAM 10 bp (EMD- 41079 and PDB 8T6S), SpRY NAC PAM 13 bp (EMD- 41080 and PDB 8T6T), SpRY NAC PAM 18 bp (EMD- 41083 and PDB 8T6X), SpRY non-target 1 bp (EMD- 41084 and PDB 8T6Y), SpRY non-target 6* bp (EMD- 41086 and PDB 8T77), SpRY non-target 8* bp (EMD-41087 and PDB 8T78), SpRY non-target 10* bp (EMD- 41088 and PDB 8T79), and SpRYmer (EMD-41093 and PDB 8T7S).
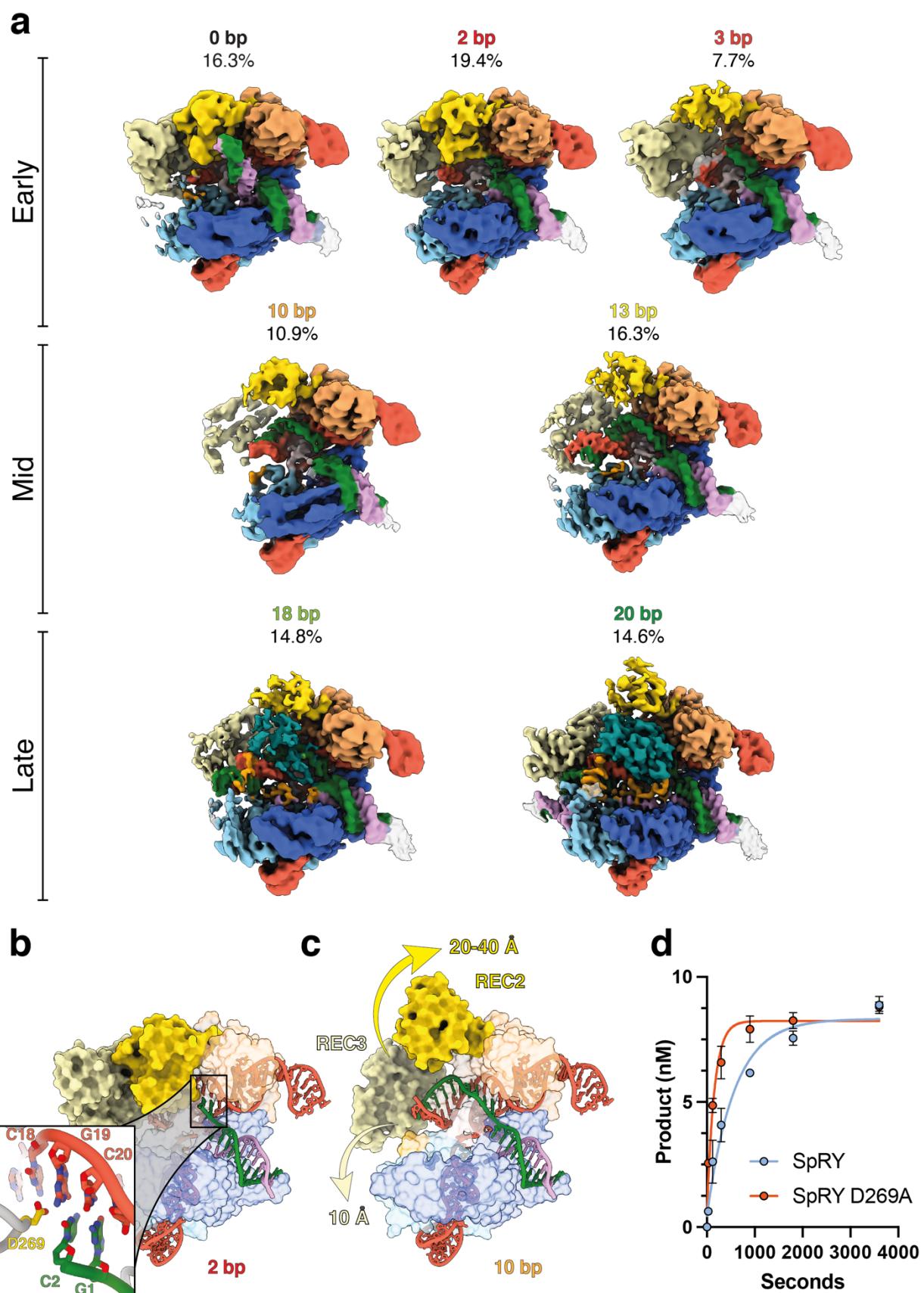
Figure 1: Structural basis of PAMless DNA recognition. a, Cas9 domain organization with SpRY mutations highlighted in light green. **b**, 2.8 Å cryo-EM reconstruction of SpRY bound to target DNA with TAC PAM sequence in the product state. Shadow of the SpRY TAC PAM product state structure with the PAM nucleotides shown as sticks and SpRY amino acid mutations shown as spheres. Molecules are colored: TS green; NTS, pink; gRNA, red; SpRY mutations, light green. **c**, Schematic representation of SpRY PAM site interaction and cleavage centers. The PAM site is highlighted in gold; HNH, dark cyan; RuvC, sky blue; Mg^{2+} in gray. **d**, Surface electrostatic potential map of Cas9. **e**, Detailed Cas9 NGG PAM site interaction (PDB: 7s4x). **f**, Surface electrostatic potential map of

SpRY. **g**, Detailed SpRY NGG PAM site interaction. **h**, View of TAC, **i**, TTC, and **j**, TGG PAM site interaction turned 180° relative to **g**.
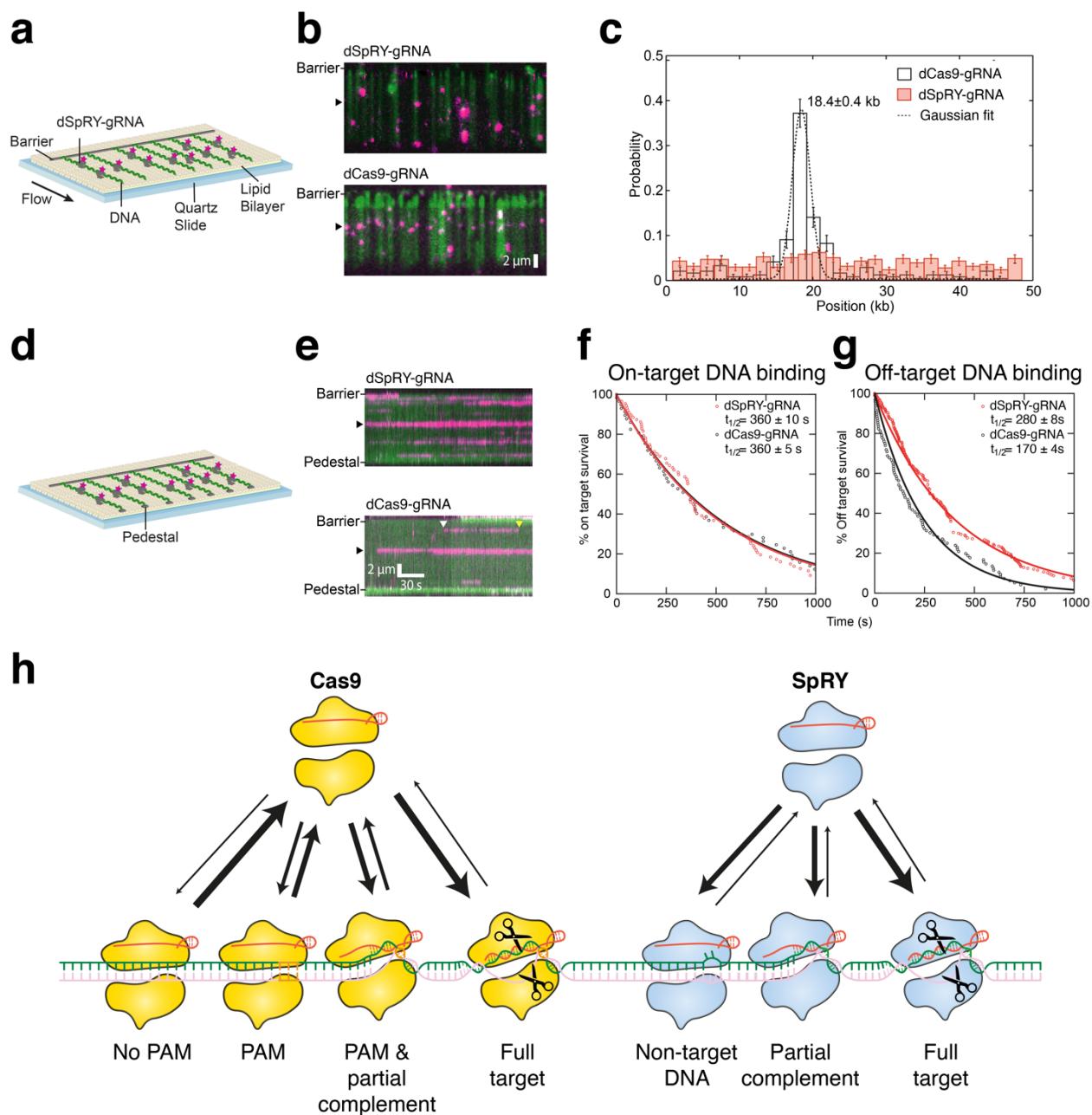
Figure 2: SpRY slowly unwinds DNA to find its target. a, Schematic representation of Cas9 enzymes reaction pathway **b**, SpRY and Cas9 observed R-loop independent (Mg^{2+}-activated) and R-loop dependent (DNA-activated) cleavage rates. **c**, SpRY and Cas9 observed rates of R-loop formation at position 1 (PAM-proximal) and position 16 (PAM-distal). **d**, Free energy profile for SpRY cleavage. Model was truncated to first irreversible step (HNH cleavage of TS). Full scheme and rate constants derived from the global fitting are in **Extended Data Fig 5, Extended Data Table 1**)

Figure 3: Direct visualization of SpRY R-loop formation in real-time. a, Ensemble cryo-EM reconstructions of SpRY during R-loop propagation with the proportion of particles as a percentage. **b & c**, Transition from early- to mid-R-loop formation requires a large conformational change of REC2 domain to create a channel to accommodate the R-loop. **d**, Alanine substitution of D269 increases the TS cleavage rate from 0.0025 s^{-1} to 0.006513 s^{-1}.

Figure 4: SpRY stably binds to off-target sequences. a, Domain organization of active (product) and inactive SpRY models in **b**. **b**, 3.0 Å cryo-EM reconstruction of the SpRYmer with one active SpRY molecule and one inactive SpRY molecule bound to the same DNA substrate. **c**, SpRY tightly binds to non-target DNA but Cas9 does not. SpRY or Cas9 were titrated into 1 nM FAM-labelled non-target DNA substrate. SpRY bound with an apparent K_D of ~5 nM. **d**, Cryo-EM maps of SpRY-gRNA in complex with non-target DNA at different stages of non-productive R-loop formation. *6, 8, and 10 bp structures contain one, three, and five mismatches, respectively. REC2 and REC3 domains exhibit diffuse density in the 6*, 8* and 10* bp structures.

Figure 5: SpRY accumulates at off-target sequences during target search. a, Schematic of a single tether DNA curtain. **b**, YOYO-stained DNA (green) bound by fluorescently labeled dSpRY-gRNA top or dCas9-gRNA bottom (magenta). **c**, Binding distribution of dSpRY-gRNA (n=543) and dCas9-gRNA (n=242); error bars represent 95% confidence intervals. **d**, Schematic of a double tether DNA curtain. **e**, Kymographs depicting distinct binding events for dSpRY-gRNA and dCas9-gRNA. **f**, Lifetime of dSpRY-gRNA and dCas9-gRNA on target and **g,** off-target. The half-lives for each enzyme are indicated. **h**, Schematic model of target search by Cas9 and SpRY.

22

References

1. Walton,R.T., Christie,K.A., Whittaker,M.N. and Kleinstiver,B.P. (2020) Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science (1979)*, **296**, eaba8853.

2. Christie,K.A., Guo,J.A., Silverstein,R.A., Doll,R.M., Mabuchi,M., Stutzman,H.E., Lin,J., Ma,L., Walton,R.T., Pinello,L., *et al.* (2022) Precise DNA cleavage using CRISPR-SpRYgests. *Nat Biotechnol*, 10.1038/s41587-022-01492-y.

3. Wang,J.Y., Pausch,P. and Doudna,J.A. (2022) Structural biology of CRISPR–Cas immunity and genome editing enzymes. *Nature Reviews Microbiology 2022*, 10.1038/s41579-022-00739-4.

4. 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 (1979)*, **337**, 816–821.

5. Jones,D.L., Leroy,P., Unoson,C., Fange,D., Ćurić,V., Lawson,M.J. and Elf,J. (2017) Kinetics of dCas9 target search in Escherichia coli. *Science (1979)*, **357**, 1420–1424.

6. 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.

7. Globyte,V., Lee,S.H., Bae,T., Kim,J.-S. and Joo,C. (2019) CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J*, **38**, e99466.

8. 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.

9. Marraffini,L.A. and Sontheimer,E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature 2010 463:7280*, **463**, 568–571.

10. Heler,R., Samai,P., Modell,J.W., Weiner,C., Goldberg,G.W., Bikard,D. and Marraffini,L.A. (2015) Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature 2015 519:7542*, **519**, 199–202.

11. 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.J., *et al.* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature 2015 523:7561*, **523**, 481–485.

12. Hu,J.H., Miller,S.M., Geurts,M.H., Tang,W., Chen,L., Sun,N., Zeina,C.M., Gao,X., Rees,H.A., Lin,Z., *et al.* (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature 2018 556:7699*, **556**, 57–63.

13. Miller,S.M., Wang,T., Randolph,P.B., Arbab,M., Shen,M.W., Huang,T.P., Matuszek,Z., Newby,G.A., Rees,H.A. and Liu,D.R. (2020) Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nature Biotechnology 2020 38:4*, **38**, 471–481.

14. Kleinstiver,B.P., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Topkar,V. V., Zheng,Z. and Joung,J.K. (2015) Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature Biotechnology 2015 33:12*, **33**, 1293–1298.

15. Chatterjee,P., Jakimo,N., Lee,J., Amrani,N., Rodríguez,T., Koseki,S.R.T., Tysinger,E., Qing,R., Hao,S., Sontheimer,E.J., *et al.* (2020) An engineered ScCas9 with broad PAM range and high specificity and activity. *Nature Biotechnology 2020 38:10*, **38**, 1154–1158.

16. Bravo,J.P., Hibshman,G.N. and Taylor,D.W. (2022) Constructing next-generation CRISPR–Cas tools from structural blueprints. *Curr Opin Biotechnol*, **78**, 102839.

17. Bravo,J.P.K., Liu,M. Sen, Hibshman,G.N., Dangerfield,T.L., Jung,K., McCool,R.S., Johnson,K.A. and Taylor,D.W. (2022) Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature 2022 603:7900*, **603**, 343–347.

18. Anders,C., Bargsten,K. and Jinek,M. (2016) Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol Cell*, **61**, 895–902.

19. Hirano,S., Nishimasu,H., Ishitani,R. and Nureki,O. (2016) Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol Cell*, **61**, 886–894.

20. Liu,M. Sen, Gong,S., Yu,H.H., Jung,K., Johnson,K.A. and Taylor,D.W. (2020) Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat Commun*, **11**, 1–13.

21. Gong,S., Yu,H.H., Johnson,K.A. and Taylor,D.W. (2018) DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep*, **22**, 359–371.

22. Cofsky,J.C., Soczek,K.M., Knott,G.J., Nogales,E. and Doudna,J.A. (2022) CRISPR–Cas9 bends and twists DNA to read its sequence. *Nat Struct Mol Biol*, **29**, 395–402.

23. Pacesa,M., Loeff,L., Querques,I., Muckenfuss,L.M., Sawicka,M. and Jinek,M. (2022) R-loop formation and conformational activation mechanisms of Cas9. *Nature*, **609**, 191–196.

24. Knight,S.C., Xie,L., Deng,W., Guglielmi,B., Witkowsky,L.B., Bosanac,L., Zhang,E.T., Beheiry,M.E., Masson,J.B., Dahan,M., *et al.* (2015) Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science (1979)*, **350**, 823–826.

25. Parmar,J.J. and Padinhateeri,R. (2020) Nucleosome positioning and chromatin organization. *Curr Opin Struct Biol*, **64**, 111–118.

26. Garcia,H.G., Grayson,P., Han,L., Inamdar,M., Kondev,J., Nelson,P.C., Phillips,R., Widom,J. and Wiggins,P.A. (2007) Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers*, **85**, 115–130.

27. Lorch,Y., Davis,B. and Kornberg,R.D. (2005) Chromatin remodeling by DNA bending, not twisting. *Proceedings of the National Academy of Sciences*, **102**, 1329–1332.

28. 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. *Nature Biotechnology 2014 32:7*, **32**, 670–676.

29. Verkuijl,S.A. and Rots,M.G. (2019) The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies. *Curr Opin Biotechnol*, **55**, 68–73.

30. Mao,S. (2020) Very fast CRISPR on demand. *Science (1979)*, **368**, 1203C.

31. Zou,R.S., Marin-Gonzalez,A., Liu,Y., Liu,H.B., Shen,L., Dveirin,R.K., Luo,J.X.J., Kalhor,R. and Ha,T. (2022) Massively parallel genomic perturbations with multi-

target CRISPR interrogates Cas9 activity and DNA repair at endogenous sites. *Nat Cell Biol*, **24**, 1433–1444.

32. Wang,A.S., Chen,L.C., Wu,R.A., Hao,Y., McSwiggen,D.T., Heckert,A.B., Richardson,C.D., Gowen,B.G., Kazane,K.R., Vu,J.T., *et al.* (2020) The Histone Chaperone FACT Induces Cas9 Multi-turnover Behavior and Modifies Genome Manipulation in Human Cells. *Mol Cell*, **79**, 221-233.e5.

33. Hall,P.M., Inman,J.T., Fulbright,R.M., Le,T.T., Brewer,J.J., Lambert,G., Darst,S.A. and Wang,M.D. (2022) Polarity of the CRISPR roadblock to transcription. *Nature Structural & Molecular Biology 2022 29:12*, **29**, 1217–1227.

34. Rostain,W., Grebert,T., Vyhovskyi,D., Pizarro,P.T., Tshinsele-Van Bellingen,G., Cui,L. and Bikard,D. (2023) Cas9 off-target binding to the promoter of bacterial genes leads to silencing and toxicity. *Nucleic Acids Res*, **51**, 3485–3496.

35. Ferrari,S., Valeri,E., Conti,A., Scala,S., Aprile,A., Di Micco,R., Kajaste-Rudnitski,A., Montini,E., Ferrari,G., Aiuti,A., *et al.* (2023) Genetic engineering meets hematopoietic stem cell biology for next-generation gene therapy. *Cell Stem Cell*, **30**, 549–570.

36. Ren,Q., Sretenovic,S., Liu,S., Tang,X., Huang,L., He,Y., Liu,L., Guo,Y., Zhong,Z., Liu,G., *et al.* (2021) PAM-less plant genome editing using a CRISPR–SpRY toolbox. *Nature Plants 2021 7:1*, **7**, 25–33.

37. Vicencio,J., Sánchez-Bolaños,C., Moreno-Sánchez,I., Brena,D., Vejnar,C.E., Kukhtar,D., Ruiz-López,M., Cots-Ponjoan,M., Rubio,A., Melero,N.R., *et al.* (2022) Genome editing in animals with minimal PAM CRISPR-Cas9 enzymes. *Nature Communications 2022 13:1*, **13**, 1–13.

38. Vos,P.D., Rossetti,G., Mantegna,J.L., Siira,S.J., Gandadireja,A.P., Bruce,M., Raven,S.A., Khersonsky,O., Fleishman,S.J., Filipovska,A., *et al.* Computationally designed hyperactive Cas9 enzymes. 10.1038/s41467-022-30598-9.

39. Liu,M., Gong,S., Yu,H.H., Jung,K., Johnson,K.A. and Taylor,D.W. (2020) Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat Commun*, **11**, 1–13.

40. Dangerfield,T.L., Huang,N.Z. and Johnson,K.A. (2021) High throughput quantification of short nucleic acid samples by capillary electrophoresis with automated data processing. *Anal Biochem*, **629**, 114239.

41. Mastronarde,D.N. (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol*, **152**, 36–51.

42. Punjani,A. (2021) Real-time cryo-EM structure determination. *Microscopy and Microanalysis*, **27**, 1156–1157.

43. Croll,T.I. (2018) ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D*, **74**, 519–530.

44. Pettersen,E.F., Goddard,T.D., Huang,C.C., Meng,E.C., Couch,G.S., Croll,T.I., Morris,J.H. and Ferrin,T.E. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, **30**, 70–82.

Methods

Protein expression and purification

Cas9 was expressed and purified as described previously (39). SpRY was expressed from pET28-SpRY-NLS-6xHis (gbr2101) purchased from Addgene (Addgene plasmid # 181743; http://n2t.net/addgene:181743 ; RRID:Addgene_181743) (2). SpRY was then expressed and purified in the same manner as Cas9.

Nucleic acid preparation

55-nt DNA duplexes were prepared from PAGE-purified oligonucleotides synthesized by Integrated DNA Technologies. DNA duplexes used in cleavage assays were prepared by mixing 6-FAM or Cy3 labeled target strands with unlabeled non-target strands at a 1:1.15 molar ratio in annealing buffer (10 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA), heating to 95ºC for 5 min, then cooling to room temperature over the course of 1 hr. The sgRNA was purchased from Synthego and annealed in annealing buffer using the same protocol as for the duplex DNA substrates. The sequences of the synthesized oligonucleotide are listed in **Extended Data Table 1**.

Kinetics

Buffer composition for kinetic reactions

Cleavage reactions were performed in 1X cleavage buffer (20 mM Tris-Cl, pH 7.5, 100 mM KCl, 5% glycerol, 1 mM DTT) at 37 ºC.

DNA cleavage kinetics

R-loop independent DNA cleavage: The reaction of Cas9 with on- and off-target DNA was performed by preincubating Cas9.gRNA (28 nM active-site concentration of Cas9, 100 nM gRNA) with 10 nM DNA with a 6-FAM label on the target strand in the absence of Mg^{2+}, and trace amounts of EDTA to chelate any co-purified cations (0.2 mM). The reaction was initiated by adding Mg^{2+} to 10 mM, then stopped at various times by mixing with 0.3 M EDTA (Extended Data Fig. 1). R-loop dependent DNA cleavage: as above, but Cas9, gRNA and Mg^{2+} were co-incubated, and the reaction was initiated by the addition of 10 nM DNA. Products of the reaction were resolved and quantified using an Applied Biosystems DNA sequencer (ABI 3130xl) equipped with a 36 cm capillary array and nanoPOP6 polymer (MCLab) (40). Data fit to equations were fit using either a single or double-exponential equations shown below:

Single exponetial equation:

$$Y = A_1 e^{-\lambda 1 t} + C$$

Equation (1)

where Y represents concentration of cleavage product, A_1 represents the amplitude, and λ_1 represents the observed decay rate (eigenvalue). The half-life was calculated as $t_{1/2} = \ln(2)/\lambda_1$.

Double exponetial equation:

$$Y = A_1 e^{-\lambda 1 t} \, A_2 e^{-\lambda 2 t} + C$$

Equation (2)

where Y represents concentration of cleavage product, A_1 represents the amplitude and λ_1 represents the observed rate for the first phase. A_2 represents the amplitude and λ_2 represents the observed rate for the second phase.

Stopped-flow kinetic assay

Stopped-flow experiment was performed as previously described (20). Briefly, 250 nM Cas9-gRNA complex (1:1 ratio, active site Cas9 concentration) was mixed with 100 nM tC°-labeled 55/55 nt DNA substrate at 37°C using AutoSF-120 stopped-flow instrument (KinTek Corporation, Austin, TX). Excitation was at 367 nm, and emission was monitored with a 445 nm filter with a 20 nm bandpass (Semrock).

CryoEM sample preparation, data collection, and processing

SpRY bound to different DNA substrates were assembled by mixing SpRY with gRNA in a 1:1.5 molar ratio and incubated at room temperature for 10 min in reaction buffer (20 mM Tris-Cl, pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 5% glycerol, and 5 mM DTT). Each DNA substrate was then added in a 1:1 molar ratio with 10 μM SpRY gRNP. The product state PAM complexes were incubated for 1 hr at room temperature and the non-target DNA complex was incubated for 2 hr at room temperature. Based on our kinetic analysis, R-loop intermediates were prepared by triggering the DNA cleavage reaction using NAC PAM DNA and incubating with SpRY gRNP at 37ºC for 60 s. The reactions were quenched by vitrification. 2.5 μl of sample was applied to glow discharged holey carbon grids (Quantifoil 1.2/1.3), blotted for 6 s with a blot force of 0, and rapidly plunged into liquid nitrogen-cooled ethane using an FEI Vitrobot MarkIV.

All datasets apart from NTC PAM were collected on an FEI Titan Krios cryo-electron microscope equipped with a K3 Summit direct electron detector (Gatan, Pleasanton, CA). Images were recorded with SerialEM(41) with a pixel size of 0.83 Å. Movies were recorded at 13.3 electrons/pixel/second for 6 s (80 frames) to give a total dose of 80 electrons/pixel. The NTC dataset was collected on a FEI Glacios cryo-TEM equipped with a Falcon 4 detector with a pixel size of 0.94 Å, and a total exposure time of 15s resulting in a total accumulated dose of 40 e/Å^2 which was split into 60 EER fractions. All datasets were collected with a defocus range of -1.5 to -2.5 μm. Motion correction, CTF estimation and particle picking was performed on-the-fly using cryoSPARC Live v4.0.0-privatebeta.2(42). Further data processing was performed with cryoSPARC v.3.2. A total of 2,016 movies were collected for the NTC dataset, 4,293 movies for the NGG dataset, 3,546 movies for the NAC dataset, 10,151 movies for the R-loop intermediate dataset, and 8,765 movies for the non-target DNA dataset.

27

The NTC, NGG, and NAC datasets were processed using similar workflows starting with blob picker with a minimum particle diameter set to 100 Å, and a maximum particle diameter set to 180 Å. The particles were then subjected to a single round of 2D classification. The particles selected from 2D classification were processed via ab initio reconstruction, followed by heterogeneous refinement. After multiple rounds of ab initio reconstruction and heterogeneous refinement, the final reconstructions were formed from non-uniform refinement. There appeared to be a second SpRY molecule bound to the same DNA in some of the NAC PAM 2D classes. Particles within these classes were extracted with a larger box size of 512 pixels yielding a dimer structure. Masks for each half of the dimer were created and particle subtraction was performed on each half. The subtracted particles were then subjected to local refinement, and ultimately recombined in ChimeraX producing the SpRYmer structure.

The R-loop intermediate and non-target DNA datasets were processed following similar workflows, again starting with blob picker as in the previous datasets. After the initial round of 2D classification, ab initio reconstruction, and heterogeneous refinement, the particles were processed using 3D classification with the parameters of force hard classification and PCA initialization mode. This method generated classes with a variable number of R-loop base pairs formed. The particles corresponding to a single state were extracted with a 384-pixel box size, globally and locally CTF refined, and fed to non-uniform refinement which yielded the final reconstructions.

Structural model building and refinement

Product state Cas9 (PDB 7S4X) was used as a starting model for the NGG PAM structure. Once built, the NGG PAM structure was used as a starting model for the NAC, NTC, 10 bp, and 18 bp structures. The 18 bp structure was used as a starting model for the 13 bp structure. SpCas9 with 0 bp, closed-protein/bent-DNA conformation (PDB 7S36) was used as a starting model for the 0 bp intermediate. SpCas9 with 3 bp R-loop (PDB 7S38) was used as a starting model for the 2 bp structure, which was then used as the starting model for the 3 bp R-loop intermediate and 1 bp non-target structure. The 3 bp R-loop intermediate was used as a starting model for the 6*bp non-target structure, which was then used as the starting model for the 8*bp non-target structure. The 10 bp R-loop intermediate was used as the starting model for the 10*bp non-target structure. SpCas9 binary complex (PDB 4ZT0) was used as the starting model for the inactive half of the SpRYmer, and the NAC PAM structure created for this manuscript was used as the starting model for the active half. Nucleic acid alterations were made in Coot*, and further modeling was performed using Isolde(43). The models were ultimately subjected to real-space refinement implemented in Phenix.

All structural figures and videos were generated using ChimeraX v1.2 (44).

Single-molecule fluorescence microscopy

Single-molecule fluorescent images were collected using a customized prism TIRF microscopy-based inverted Nikon Ti-E microscope system equipped with a motorized stage (Prio ProScan II H117). The sample was illuminated with a 488 nm laser (Coherent Sapphire) through a quartz prim (Tower Optical Co.). For imaging SYTOX Orange-stained DNA and Anti-FLAG-Qdot705-labled dSpRY or dSpCas9, the 488 nm laser power was adjusted to deliver low power (4 mW) at the front face of the prism using a neutral density filter set (Thorlabs). The imaging was recorded using electron-multiplying charge-coupled device (EMCCD) cameras (Andor iXon DU897). Flowcells used for single-molecule DNA experiments were prepared as previously described (PMDI:32621611). Briefly, a 4-mm-wide, 100-µm-high flow channel was constructed between a glass coverslip (VWR 48393 059) and a custom-made quartz microscope slide using two-sided tape (3M 665). Double-tethered DNA curtains were prepared with 40 µL of liposome stock solution (97.7% DOPC (Avanti #850375P), 2.0% DOPE- mPEG2k (Avanti #880130P), and 0.3% DOPE-biotin (Avanti #870273P) in 960 µL Lipids Buffer (10 mM Tris-HCl, pH 8, 100 mM NaCl) incubated in the flowcell for 30 minutes. Then, 50 µg µL-1 of goat anti-rabbit polyclonal antibody (ICL Labs, #GGHL-15A) diluted in Lipids Buffer was incubated in the flowcell for 10 minutes. The flowcell was washed with BSA Buffer (40 mM Tris–HCl, pH 8, 2 mM MgCl2, 1 mM DTT, 0.2 mg mL- 1 BSA) and 1 µg L-1 of digoxigenin monoclonal antibody (Life Technologies, #700772) diluted in BSA Buffer was injected and incubated for 10 minutes. Streptavidin (0.1 mg mL-1 diluted in BSA Buffer) was injected into the flowcell for another 10 minutes. Finally, ~12.5 ng µL-1 of DNA substrate was injected into the flowcell. To prepare single-tethered DNA curtains, the anti-rabbit antibody and digoxigenin antibody steps were omitted.

For single-tethered DNA experiments, DNA was visualized in the presence of a continuous flow (0.2 mL min^{-1}) in the imaging buffer (40 mM Tris-HCl pH 8.0, 2 mM MgCl2, 0.2 mg mL^{-1} BSA, 50 mM NaCl, 1 mM DTT, 100 nM SYTOX Orange) supplemented with an oxygen scavenging system (3% D-glucose (w/v), 1 mM Trolox, 1500 units catalase, 250 units glucose oxidase). For double-tethered DNA experiments, the flow was turned off after protein entered into flowcell. NIS-Elements software (Nikon) was used to collect the images at 500 ms frame rate with 100 ms exposure time. All images were exported as uncompressed TIFF stacks for further analysis in FIJI (NIH) and MATLAB (The MathWorks). The half lifetime of dSpRY and dSpCas9 were fitted by single-exponential equation.

DNA substate preparation for single-molecule assays

To prepare DNA substates for microscopy, 125 µg of λ-phage DNA was mixed with two oligos (2 µM oligo Lab07 and 2 µM oligo Lab09) in 1× T4 DNA ligase reaction buffer (NEB B0202S) and heated to 70°C for 15 min followed by gradual cooling to 15°C for 2 hours. One oligo will be annealed with the overhand located at the left cohesive end of DNA, and the other oligo will be annealed with the overhand at right cohesive end. After the oligomer hybridization, 2 µL of T4 DNA ligase (NEB M0202S) was added to the mixture and incubated overnight at room temperature to seal nicks on DNA. The ligase was

inactivated with 2 M NaCl, and the reaction was injected to an S-1000 gel filtration column (GE) to remove excess oligonucleotides and proteins.

## dSpRY-gRNA and dSpCas9-gRNA preparation and fluorescent labeling
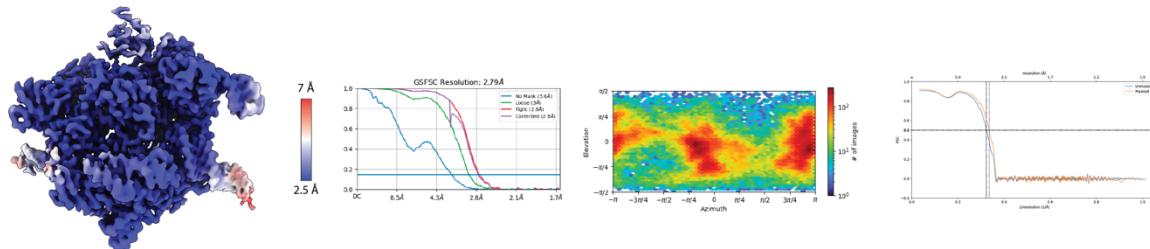
dSpRY and dSpCas9 ribonucleoprotein complexes were reconstituted by incubating a 1:2 molar ratio of apoprotein and sgRNA (sequence: GUG AUA AGU GGA AUG CCA UGG UUU UAG AGC UAG AAA UAG CAA GUU AAA AUA AGG CUA GUC CGU UAU CAA CUU GAA AAA GUG GCA CCG AGU CGG UGC UUU U, λ target-18.1 kb) in incubation buffer (50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 2 mM DTT) followed by incubation at 25 °C for 25 minutes. dSpRY and dSpCas9 were labeled by anti-FLAG antibody coupled with Qdot705, and then diluted to 0.5 nM – 2 nM in imaging buffer before injected into the flowcell. For dSpRY-gRNA and dSpCas9-gRNA binding experiments, single-tethered DNA curtains were assembled and protein was diluted and incubated with sgRNA as described above. Ribonucleoprotein was diluted in imaging buffer to a final concentration of 2 nM. For dSpRY-gRNA and dSpCas9-gRNA binding lifetime experiments, double-tethered DNA curtains were assembled and ribonucleoprotein was diluted in imaging buffer to a final concentration of 0.5 nM.

Fluorescence anisotropy binding assays
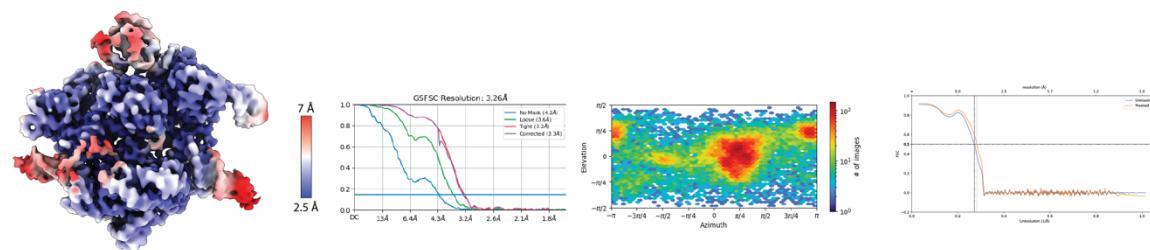
SpRY and Cas9 gRNP complexes were prepared as described above, and serial 2-fold dilutions were incubated with 0.8 nM FAM-labelled scrambled DNA for 3h at 37ºC in 1 x Cas9 buffer supplemented with 0.05 % Tween-20. Data were recorded at 37ºC in a CLARIOstar Plus multi-detection plate reader (BMG Labtech) equipped with a fluorescence polarization optical module (λ_{ex} = 485 nm; λ_{em} = 520 nm). The data were normalised and binding curves were fitted in Graphpad Prism 9 using a hyperbola.

## Extended Data Figures & Tables
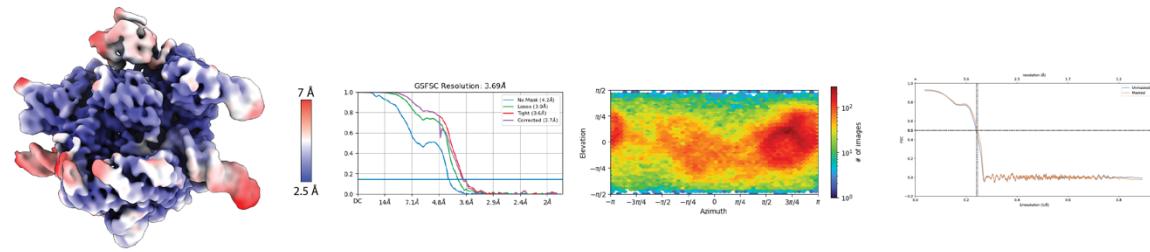
a   SpRY NAC PAM
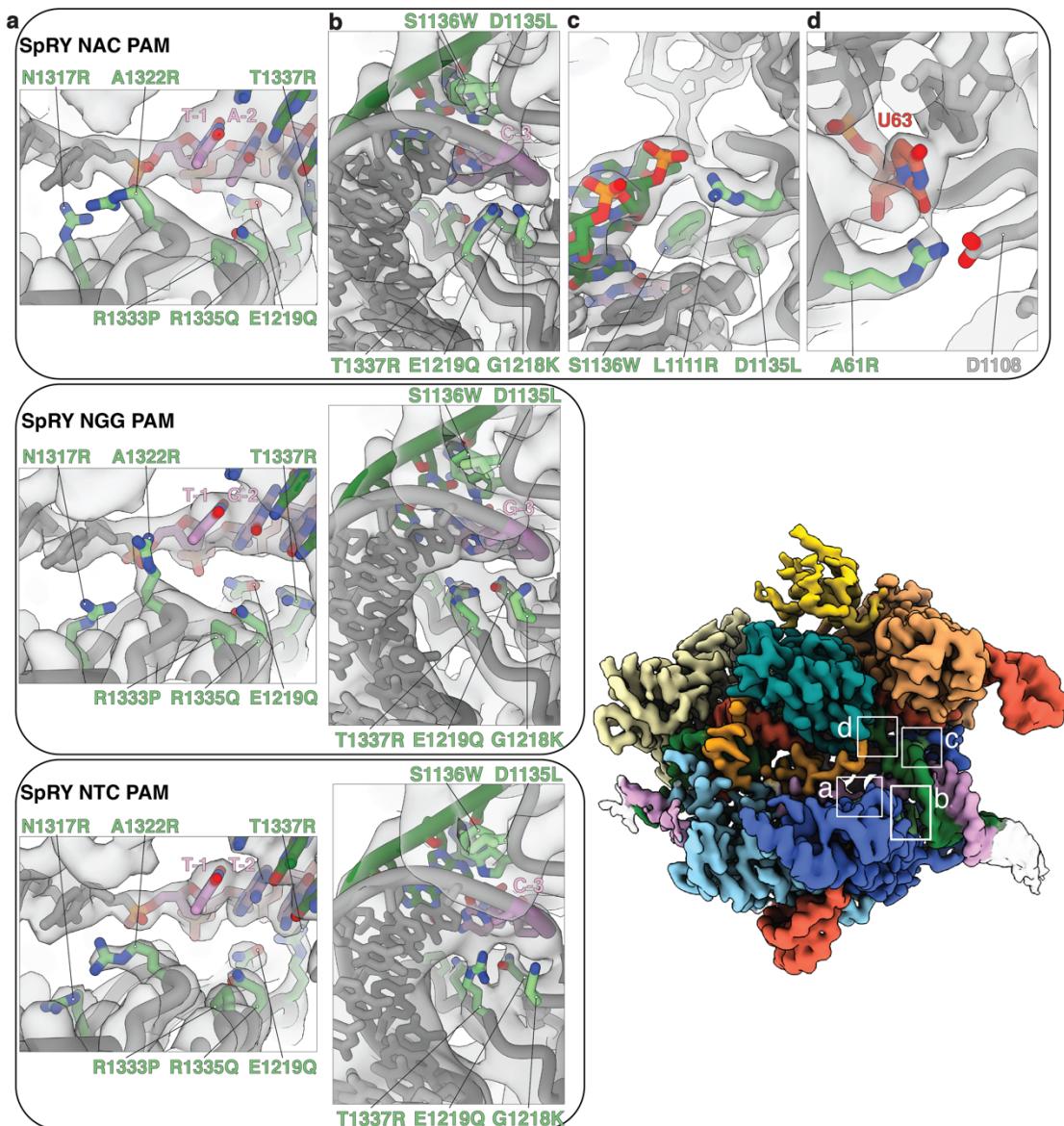
**b**   SpRY NGG PAM
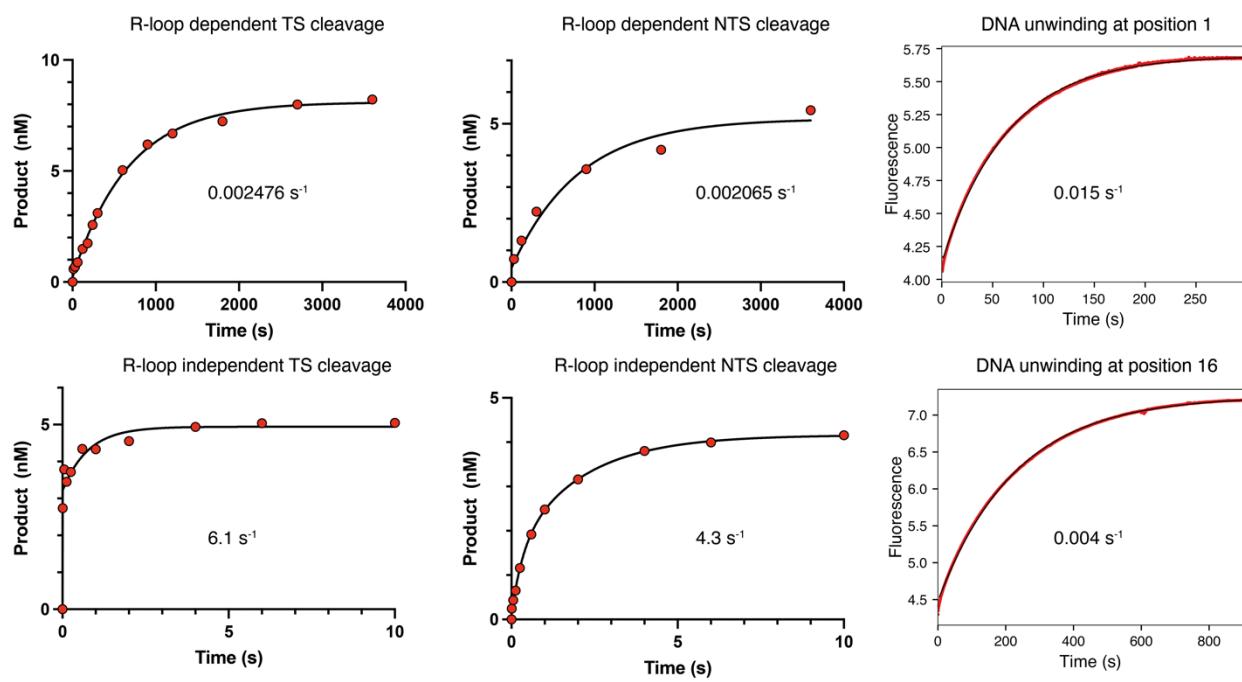
c   SpRY NTC PAM
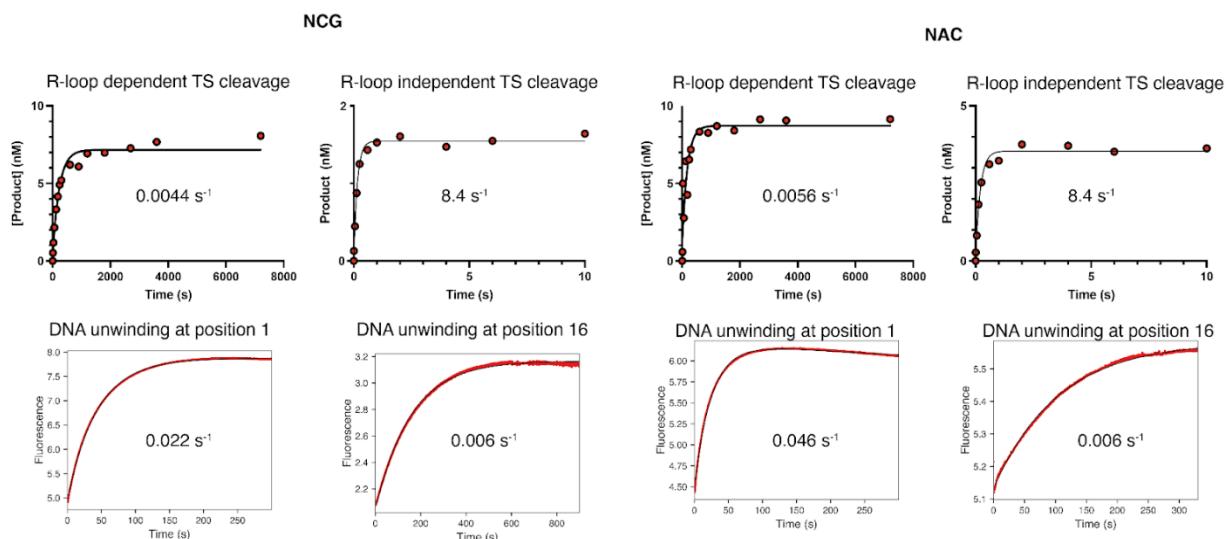
Extended data figure 1. Cryo-EM data analysis of SpRY **a**, NAC PAM DNA **b**, NGG PAM DNA, and **c**, NTC PAM DNA.
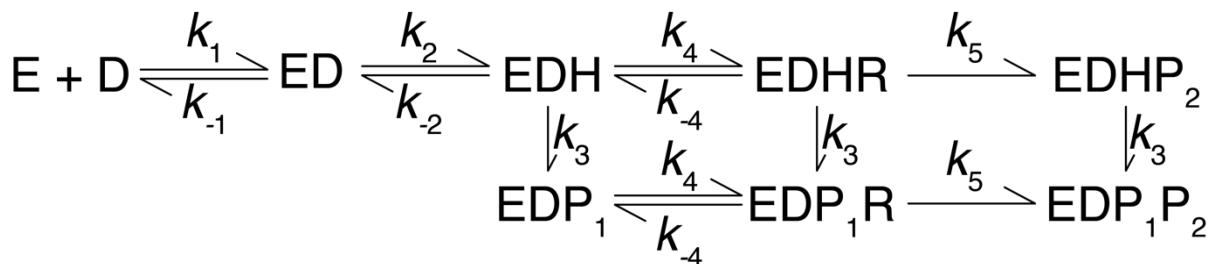
Extended data figure 2. Detailed view of SpRY mutations and the rotamers adopted by different PAM sequences.
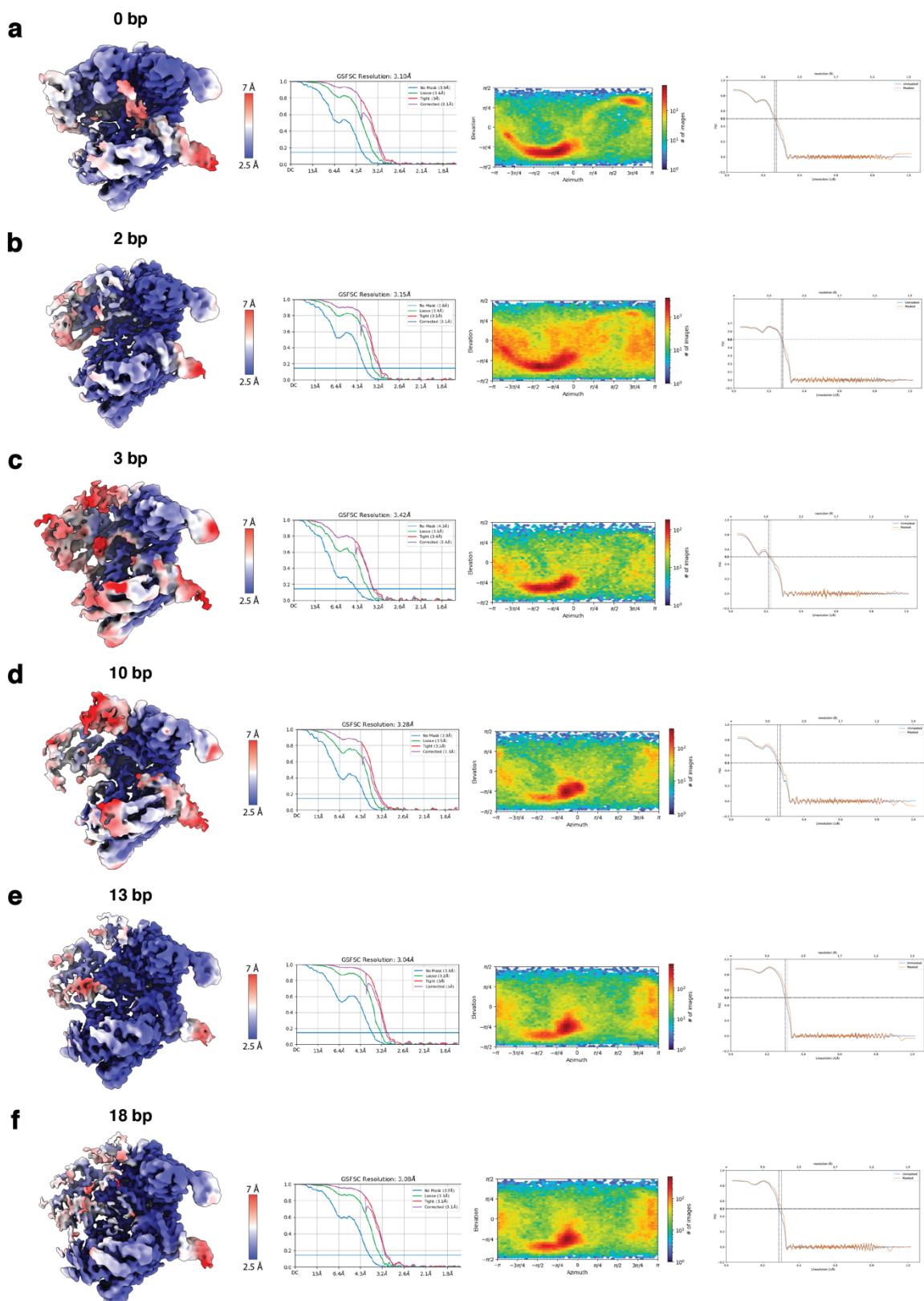
Extended data figure 3. NGG PAM DNA cleavage kinetics.
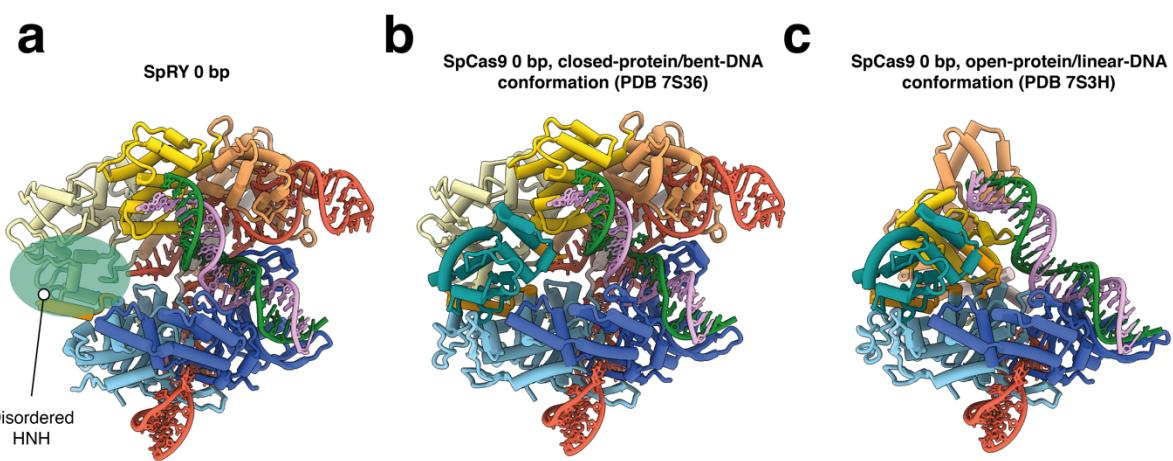
Extended data figure 4. DNA cleavage and unwinding kinetics for NGC and NAC PAM substrates. The low amplitude of product for the Mg^{2+}-initiated reactions may be due to association of SpRY at partially complementary, non-target sites, thus reducing the effective enzyme concentration.
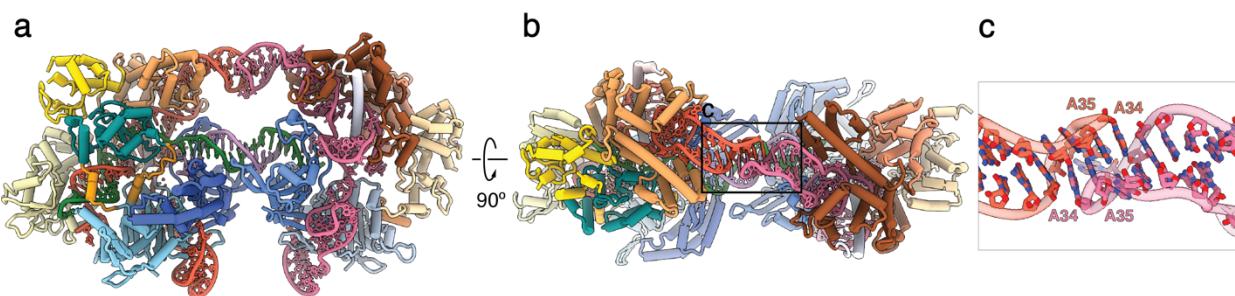
Extended data Figure 5. E is Cas9.gRNA, D is target DNA, ED is Cas9.gRNA.DNA, EDH is R-loop formation with docking of target strand to HNH, EDHR and EDP1R are R-loop formation with docking of non-target strand to RuvC, EDHP2 is RuvC cleavage of non-target strand, EDP_1 is HNH cleavage of the target-strand, EDP_1P_2 is cleavage of both strands, Rates obtained from Fitting in KinTek Explorer are shown in Extended data table 1.

a **0 bp**

b **2 bp**

c **3 bp**
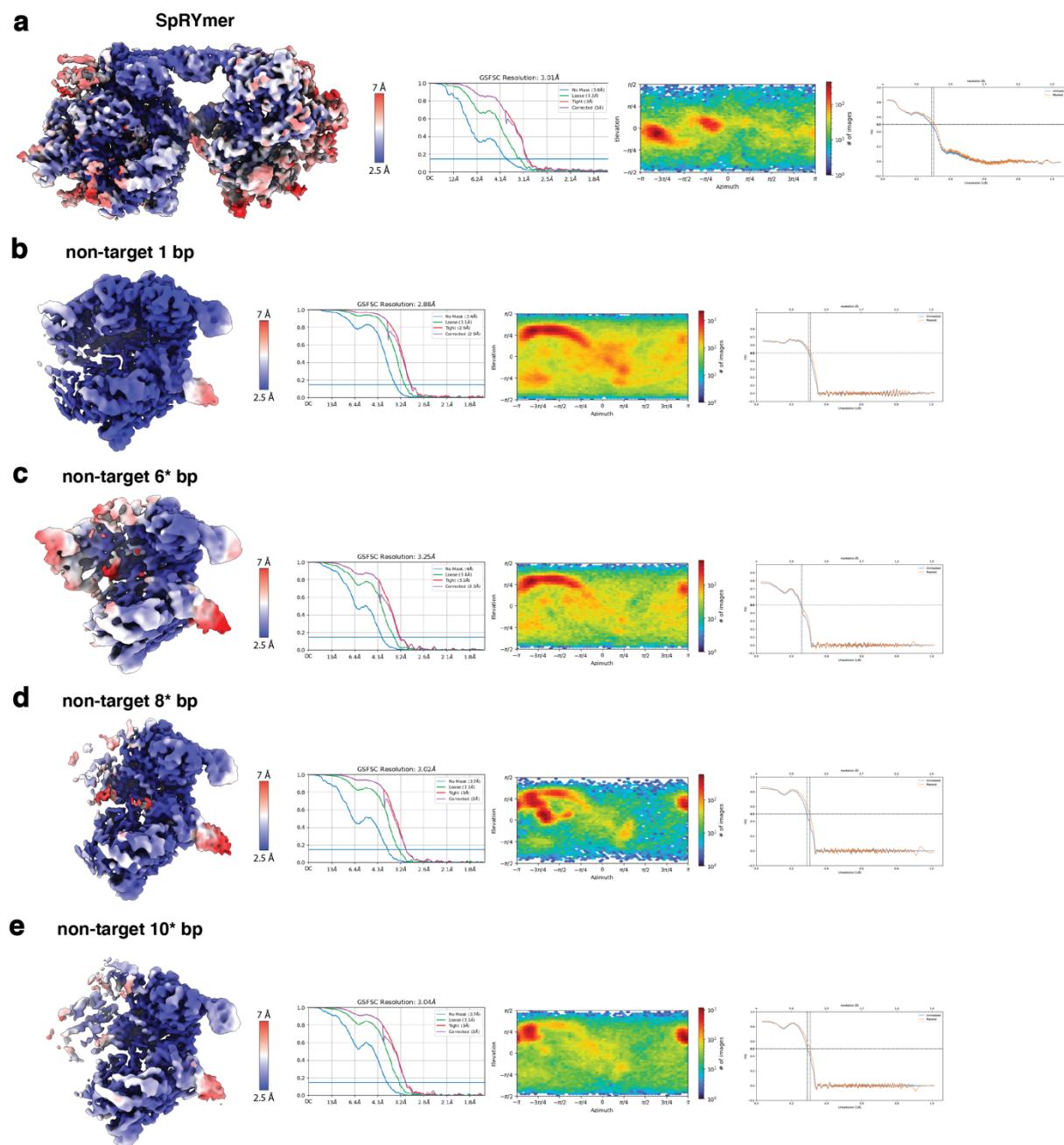
d **10 bp**

e **13 bp**

f **18 bp**

Extended data figure 6. Cryo-EM data analysis of SpRY **a**, 0 bp **b**, 2 bp **c**, 3 bp **d**, 10 bp **e**, 13 bp **f**, 18 bp, and **g**, 20 bp (product state) R-loop intermediates.

a SpRY 0 bp

b SpCas9 0 bp, closed-protein/bent-DNA conformation (PDB 7S36)

c SpCas9 0 bp, open-protein/linear-DNA conformation (PDB 7S3H)

Disordered HNH

Extended Data Figure 7. Comparison of SpRY 0 bp R-loop intermediate structure with previously determined Cas9 0 bp structures.

Extended Data Figure 8. SpRYmer stem loop interaction. **a**, Overview of the SpRYmer atomic model. **b**, Top-down view of the SpRYmer atomic model. **c**, Detailed view of the stem loop kissing interaction.

Extended data figure 9. Cryo-EM data analysis of the **a**, SpRYmer **b**, SpRY bound to non-target DNA with 1 bp **c**, 6 bp **d**, 8 bp, and **e**, 10 bp of R-loop formed. *These structures contain mismatches.

Name	Sequence (5'-3')	Source
TS NGG PAM	/6-FAM/agc tga cgt ttg tac tcc agc gtc tca tct tta tgc gtc agc aga gat ttc tgc t	IDT
NTS NGG PAM	agc aga aat ctc tgc tga cgc ata aag atg aga cgc TGG agt aca aac gtc agc t	IDT
TS NAC PAM	/6-FAM/agctgacgtttgtactGTagcgtctcatctttatgcgtcagcagagatttctgct	IDT
NTS NAC PAM	agcagaaatctctgctgacgcataaagatgagacgctACagtacaaacgtcagct	IDT
TS NTC PAM	/6-FAM/agctgacgtttgtactGAagcgtctcatctttatgcgtcagcagagatttctgct	IDT
NTS NTC PAM	agcagaaatctctgctgacgcataaagatgagacgctTCagtacaaacgtcagct	IDT
TS NGC PAM	/6-FAM/agctgacgtttgtactGCagcgtctcatctttatgcgtcagcagagatttctgct	IDT
NTS NGC PAM	agcagaaatctctgctgacgcataaagatgagacgctGCagtacaaacgtcagct	IDT
NTS NGG PAM Hot	/6-FAM/agc aga aat ctc tgc tga cgc ata aag atg aga cgc TGG agt aca aac gtc agc t	IDT
Scrambled TS	/56-FAM/agctgacgtttgtactccac*g*c*a*gagtagaaatacgcagagcagagatttctgct	IDT
Scrambled NTS	agcagaaatctctgctctgcgtatttct*a*c*t*c*t*g*cgtggagtacaaacgtcagct	IDT
TS NGG cold	agc tga cgt ttg tac tcc agc gtc tca tct tta tgc gtc agc aga gat ttc tgc t	IDT
TS NAC cold	agctgacgtttgtactGTagcgtctcatctttatgcgtcagcagagatttctgct	IDT
TS NGG cold	agctgacgtttgtactGAagcgtctcatctttatgcgtcagcagagatttctgct	IDT
NGG_NTS_tCo_16	agc aga aat ctc tgc tga cg[tCo] ata aag atg aga cgc tgg agt aca aac gtc agc t	Biosynthesis
NGC_NTS_tCo_16	agc aga aat ctc tgc tga cg[tCo] ata aag atg aga cgc tgc agt aca aac gtc agc t	Biosynthesis
NAC_NTS_tCo_16	agc aga aat ctc tgc tga cg[tCo] ata aag atg aga cgc tac agt aca aac gtc agc t	Biosynthesis
NGG_NTS_tCo_1	agc aga aat ctc tgc tga cgc ata aag atg aga cg[tCo] tgg agt aca aac gtc agc t	Biosynthesis
NGC_NTS_tCo_1	agc aga aat ctc tgc tga cgc ata aag atg aga cg[tCo] tgc agt aca aac gtc agc t	Biosynthesis
NAC_NTS_tCo_1	agc aga aat ctc tgc tga cgc ata aag atg aga cg[tCo] tac agt aca aac gtc agc t	Biosynthesis

Extended Data Table 1. List of DNA substrates used in this study. * corresponds to phosphorotioate-modified DNA.

40

Extended Data Table 2. Cryo-EM data processing and modeling.

	SpRY NAC PAM 20 bp (EMD-40740) (PDB 8SRS)	SpRY NGG PAM (EMD-40681) (PDB 8SPQ)	SpRY NTC PAM (EMD-40705) (PDB 8SQH)	SpRY NAC PAM 0 bp (EMD-41073) (PDB 8T6O)	SpRY NAC PAM 2 bp (EMD-41074) (PDB 8T6P)	SpRY NAC PAM 3 bp (EMD-41085) (PDB 8T76)	SpRY NAC PAM 10 bp (EMD-41079) (PDB 8T6S)	SpRY NAC PAM 13 bp (EMD-41080) (PDB 8T6T)	SpRY NAC PAM 18 bp (EMD-41083) (PDB 8T6X)	SpRY non-target 1 bp (EMD-41084) (PDB 8T6Y)	SpRY non-target 6* bp (EMD-41086) (PDB 8T77)	SpRY non-target 8* bp (EMD-41087) (PDB 8T78)	SpRY non-target 10* bp (EMD-41088) (PDB 8T79)	SpRYmer (EMD-41093) (PDB 8T7S)
Data collection and processing														
Voltage (kV)	300 kV	300 kV	200 kV	300 kV	300 kV	300 kV	300 kV	300 kV	300 kV	300 kV	300 kV	300 kV	300 kV	300 kV
Electron exposure (e–/Å²)	80	80	49	80	80	80	80	80	80	80	80	80	80	80
Defocus range (μm)	-1.5 to -2.5 (all)													
Pixel size (Å)	0.8332	0.8332	0.94	0.8332	0.8332	0.8332	0.8332	0.8332	0.8332	0.8332	0.8332	0.8332	0.8332	0.8332
Symmetry imposed	C1 (all)													
Initial particle images (no.)	3,084,834	2,140,357	1,203,543	3,084,834	3,084,834	3,084,834	3,084,834	3,084,834	3,084,834	2,291,944	2,291,944	2,291,944	2,291,944	2,147,844
Final particle images (no.)	89,331	30,690	126,734	99,758	118,550	47,218	67,000	100,084	90,380	371,367	114,471	76,598	123,454	62,283
Map resolution (Å) / FSC threshold	2.79	3.26	3.69	3.10	3.15	3.42	3.28	3.04	3.08	2.88	3.25	3.02	3.04	3.01
Map resolution range (Å)	2.5-7 (all)													
Refinement														
Initial model used (PDB code)	8SPQ	7S4X	8SPQ	7S36	7S38	8T6P	8SPQ	8T6X	8SPQ	8T6P	8T76	8T77	8T6S	8SRS, 4ZT0
Model resolution (Å) / FSC threshold	3.0	3.5	4.1	3.7	3.6	4.0	3.7	3.2	3.3	3.2	3.8	3.3	3.3	3.3
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Map sharpening B factor (Å²)	67.4	67.1	156.8	90.4	102.2	84.1	88.9	91.0	97.8	109.3	97.6	82.4	90.6	49.0
Model composition														
Non-hydrogen atoms	14151	14150	14110	11904	11174	11050	10405	12331	13445	11485	10752	9565	11942	24993
Protein residues	1341	1341	1341	1137	1108	1088	962	1168	1278	1148	1025	863	1146	2469
Nucleotides	149	149	147	131	110	111	119	129	139	109	114	116	119	239
Ligands	4 Mg²⁺, 1 H₂O	4 Mg²⁺	4 Mg²⁺, 1 H₂O	0	0	0	0	0	3 Mg²⁺	0	0	0	0	5 Mg²⁺, 2 H₂O
R.m.s. deviations														
Bond lengths (Å)	0.005	0.005	0.006	0.0009	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.006	0.011
Bond angles (°)	0.85	0.944	1.043	1.141	0.892	1.021	0.994	1.212	0.722	0.983	0.976	0.979	1.281	1.946
Validation														
MolProbity score	1.35	1.35	1.64	1.23	1.33	1.42	1.46	1.45	1.29	1.28	1.42	1.39	1.54	1.44
Clashscore	3.86	4.82	7.09	4.59	4.64	5.81	6.38	5.90	4.44	4.83	6.21	5.40	6.89	4.75
Poor rotamers (%)	0.66	0.33	0.66	0.20	0.43	0.32	0.46	0.48	0.17	0.92	0.33	0.39	0.87	0.75
Ramachandran plot														
Favored (%)	97.00	97.52	96.32	98.23	97.54	97.49	97.45	97.32	97.70	97.89	97.62	97.52	97.08	96.81
Allowed (%)	3.00	2.48	3.68	1.68	2.46	2.51	2.55	2.68	2.30	2.11	2.38	2.48	2.92	3.19
Disallowed (%)	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Parameter	Best fit	
	SpRY	**Cas9**
K_1	32 μM s^{-1}	3*
K_{-1}	0.094 s^{-1}	
K_2	0.0045 s^{-1}	2.5
K_{-2}	0.0003 s^{-1}	1.2
K_3	20 s^{-1}	5.6
K_4	0.03 s^{-1}	1,8
K_{-4}	0.63	2.3
K_5	0.54	4.4

Extended Data Table 3. Rate constants derived from fitting the data in Extended Data Figure 3 to the model shown in **Extended Data Fig 5** are summarized. Rates for Cas9 are taken from (20). The * indicates that rates for DNA binding and dissociation were not defined by the data but were locked at these rate constants to give a K_d of 3 nM.