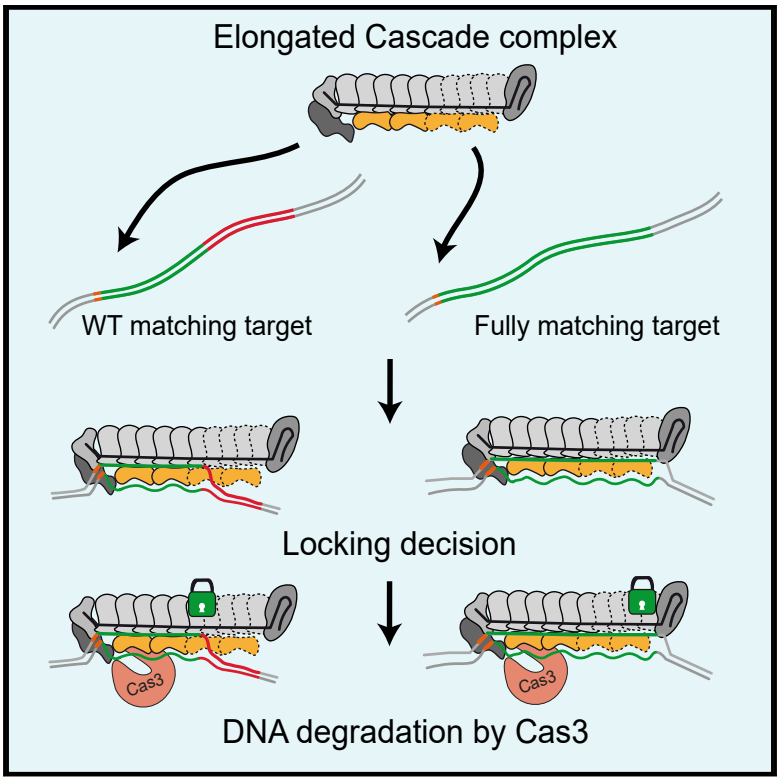
Article

Cell Reports

Decision-Making in Cascade Complexes Harboring crRNAs of Altered Length

Graphical Abstract



Authors

Inga Songailiene, Marius Rutkauskas, Tomas Sinkunas, ..., Carla Schmidt, Virginijus Siksnys, Ralf Seidel

Correspondence

siksnys@ibt.lt (V.S.), ralf.seidel@physik.uni-leipzig.de (R.S.)

In Brief

Songailiene et al. show that engineered versions of Cascade with elongated crRNAs can base-pair with target DNA over 57 bp. However, target recognition only requires wild-type-sized base-pairing of ~30-bp length. These findings define constraints for improving the specificity of these complexes in biotechnological applications.

Highlights

- Cascades with elongated crRNA spacers support the formation of up to 57-bp-long R-loops
- Target recognition and DNA degradation require only wild-type-sized, ~30-bp-long R-loops
- Complexes with elongated crRNAs bind targets with significantly increased affinity

Songailiene et al., 2019, Cell Reports *28*, 3157–3166
September 17, 2019
https://doi.org/10.1016/j.celrep.2019.08.033

CellPress

Cell Reports
Article

Decision-Making in Cascade Complexes Harboring crRNAs of Altered Length

Inga Songailiene,[1,4] Marius Rutkauskas,[2,4] Tomas Sinkunas,[1] Elena Manakova,[1] Sabine Wittig,[3] Carla Schmidt,[3] Virginijus Siksnys,[1,*] and Ralf Seidel[2,5,*]
[1]Institute of Biotechnology, Vilnius University, Vilnius 10257, Lithuania
[2]Molecular Biophysics Group, Peter Debye Institute for Soft Matter Physics, Universität Leipzig, Leipzig 04103, Germany
[3]HALOmem, Charles Tanford Protein Centre, Martin Luther University Halle-Wittenberg, Halle 06120, Germany
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: siksnys@ibt.lt (V.S.), ralf.seidel@physik.uni-leipzig.de (R.S.)
https://doi.org/10.1016/j.celrep.2019.08.033

SUMMARY

The multi-subunit type I CRISPR-Cas surveillance complex Cascade uses its crRNA to recognize dsDNA targets. Recognition involves DNA unwinding and base-pairing between the crRNA spacer region and a complementary DNA strand, resulting in formation of an R-loop structure. The modular Cascade architecture allows assembly of complexes containing crRNAs with altered spacer lengths that promise increased target specificity in emerging biotechnological applications. Here we produce type I-E Cascade complexes containing crRNAs with up to 57-nt-long spacers. We show that these complexes form R-loops corresponding to the designed target length, even for the longest spacers tested. Furthermore, the complexes can bind their targets with much higher affinity compared with the wild-type form. However, target recognition and the subsequent Cas3-mediated DNA cleavage do not require extended R-loops but already occur for wild-type-sized R-loops. These findings set important limits for specificity improvements of type I CRISPR-Cas systems.

INTRODUCTION

CRISPR and Cas genes constitute adaptive immune systems against foreign nucleic acid elements in ~50% of bacteria and ~90% of archaea (Grissa et al., 2007). Defense against invaders is promoted by effector complexes that are guided by CRISPR RNA (crRNA) to recognize complementary protospacer sequences within the foreign invader nucleic acids (Barrangou et al., 2007; Brouns et al., 2008; Jore et al., 2011; Sinkunas et al., 2011, 2013; Westra et al., 2012). Successful base-pairing along the spacer region of the crRNA triggers a conformational checkpoint, followed by cleavage of the invader element (Redding et al., 2015; Sternberg et al., 2015; Szczelkun et al., 2014).

The fact that sequence specificity is mostly determined by the spacer sequence of the crRNA promoted wide-spread application of CRISPR-Cas systems as convenient and cost-effective molecular tools (Hatoum-Aslan, 2018; Kiro et al., 2014; Luo et al., 2015). Most applications are currently based on class 2 CRISPR systems, in particular the effector complexes Cas9 and Cas12a, because they only comprise a single protein subunit. Successful applications include gene activation, gene silencing, gene knockdown and knockin, base editing, epigenetic editing, and others (Hilton et al., 2015; Roy et al., 2018; Thakore et al., 2015;Shimatani et al., 2017; for a recent review, see Adli, 2018). However, the use of CRISPR-Cas systems also has significant challenges. Because of the limited length of the crRNA spacers, recognition of off-targets remains problematic in applications requiring high on-target specificity, despite elegant engineering approaches (Lee et al., 2018; Slaymaker et al., 2016; Tsai and Joung, 2016). Furthermore, the additional requirement for a short protospacer-adjacent motif (PAM) that is recognized by a protein domain of the effector complex limits the available sequence space (Gleditzsch et al., 2019; Karvelis et al., 2017; Nishimasu et al., 2018).

A promising alternative to class 2 systems in biotechnological applications may be the Cascade surveillance complexes of type I CRISPR-Cas systems (Hatoum-Aslan 2018; Kiro et al., 2014). They form large multimeric complexes of Cas proteins of different composition. For example, the subunit stoichiometry of the type I-E Cascade complexes from *Escherichia coli* (EcCascade) and *Streptococcus thermophilus* (StCascade) is $Cse1_1Cse2_2Cas7_6Cas5_1Cas6e_1$, in which the hexameric Cas7 forms a central filament that positions the single 61-nt crRNA (Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014; Sinkunas et al., 2013). These complexes recognize their double-stranded DNA targets by searching for a suitable PAM (Hayes et al., 2016; Sinkunas et al., 2013; Westra et al., 2012, 2013). PAM binding nucleates PAM-adjacent base-pairing between the crRNA and the DNA target strand while expelling the non-target strand, forming a so-called R-loop (Szczelkun et al., 2014). Driven by thermal fluctuations, this unstable triple-strand structure can expand farther toward the PAM-distal end of the protospacer (Rutkauskas et al., 2015), as also seen for class 2 systems (Josephs et al., 2015; Lim et al., 2016; Singh et al., 2018a, 2018b). When the R-loop extends up to the final 5 bp from the PAM-distal spacer end, a conformational change of the Cse1-Cse2 filament of Cascade "locks" the R-loop in a

stable conformation (van Erp et al., 2015; Szczelkun et al., 2014). This causes bulge formation of the non-target DNA strand, which triggers recruitment of the helicase-nuclease Cas3, which carries out DNA degradation (Xiao et al., 2017, 2018). Thus, R-loop locking represents the molecular decision after which a target is considered to be recognized (Rutkauskas et al., 2015).

Type I CRISPR systems have recently been used in various biotechnological applications, including gene silencing, programmed DNA degradation, regulation of metabolic pathways, and engineering of phage genomes (Chang et al., 2016; Kiro et al., 2014; Li et al., 2016; Rath et al., 2015; Yosef et al., 2015). Most recently, a natural Cascade transposon system has been employed for highly specific, genome-wide DNA integration that obviates the requirement for double-stranded breaks (Klompe et al., 2019). Advantages of using multimeric Cascade complexes compared with monomeric class 2 effector complexes are (1) an extended range of available PAM sequence motifs (Jung et al., 2017; Karvelis et al., 2017; Leenay et al., 2016) that can be limited to only a single A/T nucleotide (e.g., in the case of StCascade; Sinkunas et al., 2013) and (2) significantly longer spacer sequences compared with class 2 systems that may allow improved specificity.

Remarkably, engineered Cascade complexes that harbor crRNAs with even longer spacers promise further increased sequence specificity. Such complexes with shortened or elongated crRNA spacers have been reported for the type I-F Cascade from *Shewanella putrefaciens* (Gleditzsch et al., 2016) (with spacer length changes between −18 nt and +18 nt) and the type I-E EcCascade (with spacer length changes between −18 to +24 nt) (Kuznedelov et al., 2016; Luo et al., 2016). Mass spectrometry analysis showed that crRNAs with altered length were incorporated by a modular alteration of the Cas7 backbone, with one Cas7 subunit included or omitted for every 6 nt of altered crRNA length (Kuznedelov et al., 2016; Luo et al., 2016). The number of Cse2 subunits increased or decreased by one when the spacer length changed by 12 nt (Luo et al., 2016). Cascades with altered spacer lengths partially retained their ability for DNA cleavage when the spacers were elongated but not shortened by 6 bp or more (Kuznedelov et al., 2016). *In vivo*, in some cases, EcCascade complexes with elongated crRNAs exhibited improved gene silencing ability (Luo et al., 2016). However, the molecular details of this altered performance remain unclear.

Here we explore the mechanistic consequences of the remarkable structural plasticity of the Cascade complex. In particular, we investigate target recognition by StCascade complexes harboring crRNA spacers ranging from 15–57 nt. Using single-molecule magnetic tweezers and bulk solution experiments, we show that the produced complexes can form R-loops, as dictated by their spacer length. Target recognition (R-loop locking) does, however, not require fully formed R-loops, but it already occurs for R-loop lengths at which the wild-type (WT) complex undergoes locking. When locked, Cas3-mediated target degradation occurs at similar rates independent of the formed R-loop length. Direct mechanical measurements reveal that locked StCascade complexes with extended spacers can have considerably increased affinities for DNA. Altogether, our measurements provide detailed mechanistic insight into target recognition and decision-making by StCascade complexes with modified spacer lengths and reveal the limits and possibilities of specificity improvements using these systems.

RESULTS

The Composition and Stability of StCascade Depend on the Spacer Length

The WT StCascade complex is assembled along a crRNA backbone composed of a 33-nt spacer flanked by 7-nt 5′ and 21-nt 3′ repeat handles (Sinkunas et al., 2013). To evaluate whether StCascade could accommodate crRNAs of altered length, we co-expressed its Cas proteins together with crRNAs whose spacer length was either decreased or increased in steps of 6 nt compared with the WT spacer so that the periodicity of the Cas7 filament was matched. Spacer length modifications of the crRNA included −18, −12, −6, +6, +12, +18, and +24 nt (hereafter used to denote the corresponding StCascade complex). The total crRNA guide spacer lengths thus ranged from 15–57 nt (Figure 1A). SDS-PAGE analysis of isolated complexes confirmed the presence of all expected Cas proteins for all produced complexes (Figure 1B). Denaturing urea-PAGE revealed the presence of a crRNA of the expected length for each complex (Figure 1B). Additionally, an RNA originating from the pre-crRNA 5′ end was detected to various extents (Figure S1) that was likely part of a non-functional pre-complex containing Cas7, Cse1, and Cas5 (Sinkunas et al., 2013).

Given the periodicity of Cas7 in the EcCascade filament (Mulepati et al., 2014; Zhao et al., 2014), one would expect that the number of Cas7 subunits changes by one for every spacer length alteration of ± 6 nt (Kuznedelov et al., 2016; Luo et al., 2016). In contrast, an added or removed Cse2 subunit would be expected when the spacer length is altered by ± 12 nt. To test this, the intact StCascade complexes were analyzed using mass spectrometry modified for transmission of high-mass complexes (also called "native" mass spectrometry) (Figure 1D). Measured masses of the dominant species of the −12, +12, and +24 complexes confirmed the expected changes in stoichiometry (Figures 1D and 1E; see Table S1 for molecular weights); i.e., the +24 complex could indeed incorporate four additional Cas7 and two additional Cse2 subunits. In contrast, the largest detectable species for the −18, −6, +6, and +18 complexes were considerably smaller than expected (see black dots in Figure 1E). The mass spectra for these complexes revealed the presence of different alternative species (Figure S2). We attribute this to severe assembly defects of these complexes, where the periodicity of the extending Cse2 filament no longer matches the complex length.

Additionally, we probed the overall architecture of the WT and +12 complexes using small angle X-ray scattering (SAXS) (Figures 1C and S3; Table S2). For WT StCascade, the shape of the particle obtained from SAXS data modeling was very similar to the crystal structure of EcCascade (van Erp et al., 2015; Figure S3). The shape of the +12 complex was found to be considerably elongated compared with WT StCascade. This agrees with incorporation of two additional Cas7 and one additional Cse2 subunit found by mass spectrometry (Figure 1C; Figure S3).

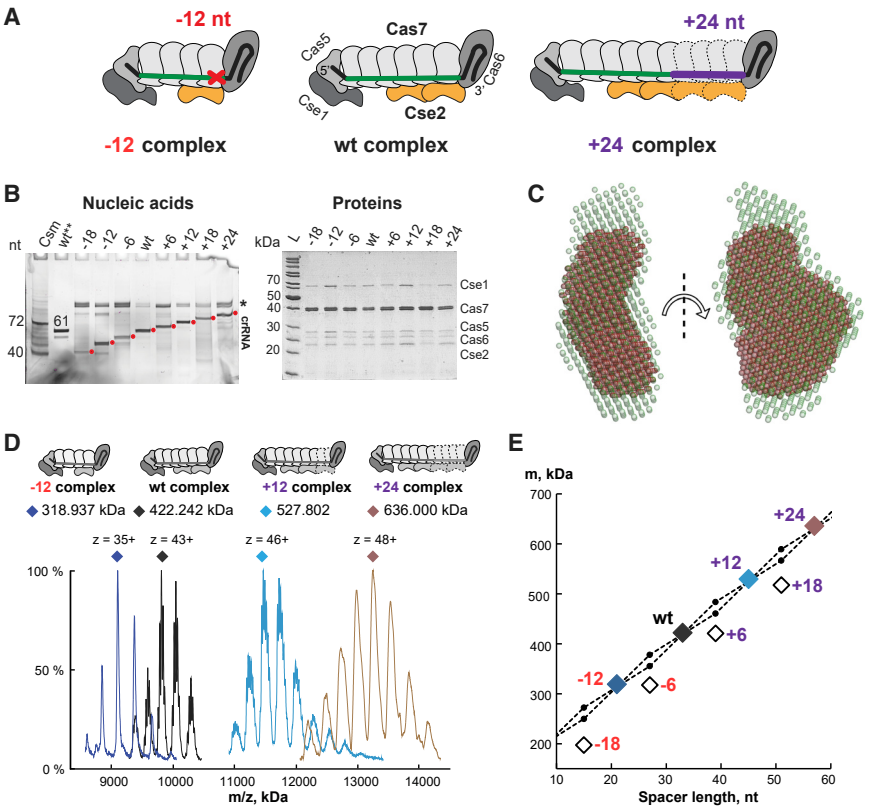


Figure 1. Composition and Structure of StCascade Complexes with crRNA Spacer Lengths Ranging from −18 to +24 nt Compared with the WT

(A) Scheme of Cascade complexes containing a WT spacer (center), a 12-nt shortened spacer (left), and a 24-nt elongated spacer (right). Incorporation of spacers with altered length is accomplished by adapting the number of Cas7 and Cse2 subunits (proposed additional subunits are shown with dashed borders).

(B) PAGE analysis of the RNA (left) and protein (right) content of the different purified complexes. As size markers, RNAs of known length from the type III *S. thermophilus* Csm complex (Tamulaitis et al., 2014) and a previously purified StCascade complex (WT**) (Sinkunas et al., 2013) were used. The samples contained a crRNA of the expected length (labeled with red dots) and an additional RNA (asterisk; see Figure S1). All complexes contained the expected Cas proteins.

(C) GASBOR model obtained from SAXS analysis of the WT and the +12 Cascade complex, shown as small red and transparent green spheres, respectively.

(D) Native mass spectra for the −12, WT, +12, and +24 complexes. The charge state (*z*) of the largest peak of each complex (marked with a diamond) is given. Schemes show the most probable composition of the complexes and the mass of the most abundant and largest species.

(E) Mass of the largest detected species of the different protein samples. Black dots connected by black dashed lines show the expected mass when two Cas7 and one Cse2 subunits are added or omitted for every spacer length alteration of 12 nt. The two dots for altered spacer lengths of odd multiples of 6 nt reflect the ambiguity in the Cse2 stoichiometry of these complexes. The largest detectable species in theses samples were found to be significantly smaller than predicted.

Dependence of R-loop Size and Stability on Spacer Length

We next probed whether the StCascade complexes containing spacers of altered length were able to form R-loops. For this purpose, we employed a magnetic tweezer assay that can monitor the formation and extent of the actual R-loop structure on single DNA molecules (Rutkauskas et al., 2017). 2.1-kbp-long DNA molecules containing a single target sequence were bound at one end to a magnetic bead and on the other end to the surface of the fluidic cell of the setup (Figure 2A, left panel). A pair of magnets was used to generate a pulling force onto the beads and to stretch the DNA with a defined tension. Video microscopy was used to monitor the bead position and, thus, the DNA lengths (Huhle et al., 2015). Rotation of the magnets allowed introduction of DNA supercoiling. At sufficiently low forces, this leads to a reduction of the DNA length for both negative and positive supercoiling because the DNA is writhing so that a so-called plectonemic superhelix is formed (Figure 2A, left panel; Brutzer et al., 2010). When StCascade binds to the target and forms a stable R-loop, the DNA is locally unwound; i.e., negative supercoils are absorbed by the R-loop structure. Therefore, the characteristic supercoiling curve becomes shifted toward negative supercoiling by the helical turns that were unwound by the R-loop (Figure 2A, right panel; Rutkauskas et al., 2017). Locked R-loops of StCascade are stable under moderate positive supercoiling (Szczelkun et al., 2014). Therefore, the supercoiling curve appears to be shifted on both the left and the right side toward negative turns. Unlocked R-loops can still form and are stabilized with the help of negative supercoiling (Rutkauskas et al., 2015). As soon as supercoiling vanishes or positive supercoiling is applied, unlocked R-loops readily dissociate so that the corresponding supercoiling curves are only shifted on the left but not on the right side (Figure 2B).

Using a target that was fully complementary to the crRNA even for the complex with the longest spacer, R-loops were formed for the different complexes at negative supercoiling, and supercoiling curves were recorded by positively twisting the DNA. R-loop formation was observed for all complexes (Figure 2B). However, for the −18, −6, +6, and +18 complexes, an ~100-fold higher concentration compared with the other samples was required to obtain R-loop formation on a 1-min timescale. This indicated that these samples contained only a minor fraction of functional (i.e., fully assembled) complexes, in agreement with the mass spectrometry data (Figure S2; Table S1). StCascade complexes with shorter spacers than the WT (−6, −12, and −18) only formed unlocked R-loops that dissociated immediately after positive torsion was applied (Figure 2B). WT StCascade and StCascades with elongated spacers formed exclusively locked R-loops (Figures 2B and 2C).

The DNA untwisting caused by the R-loops was obtained from the shift of the left side of the supercoiling curve and was found to be strongly correlated with spacer length (Figure 2C, green

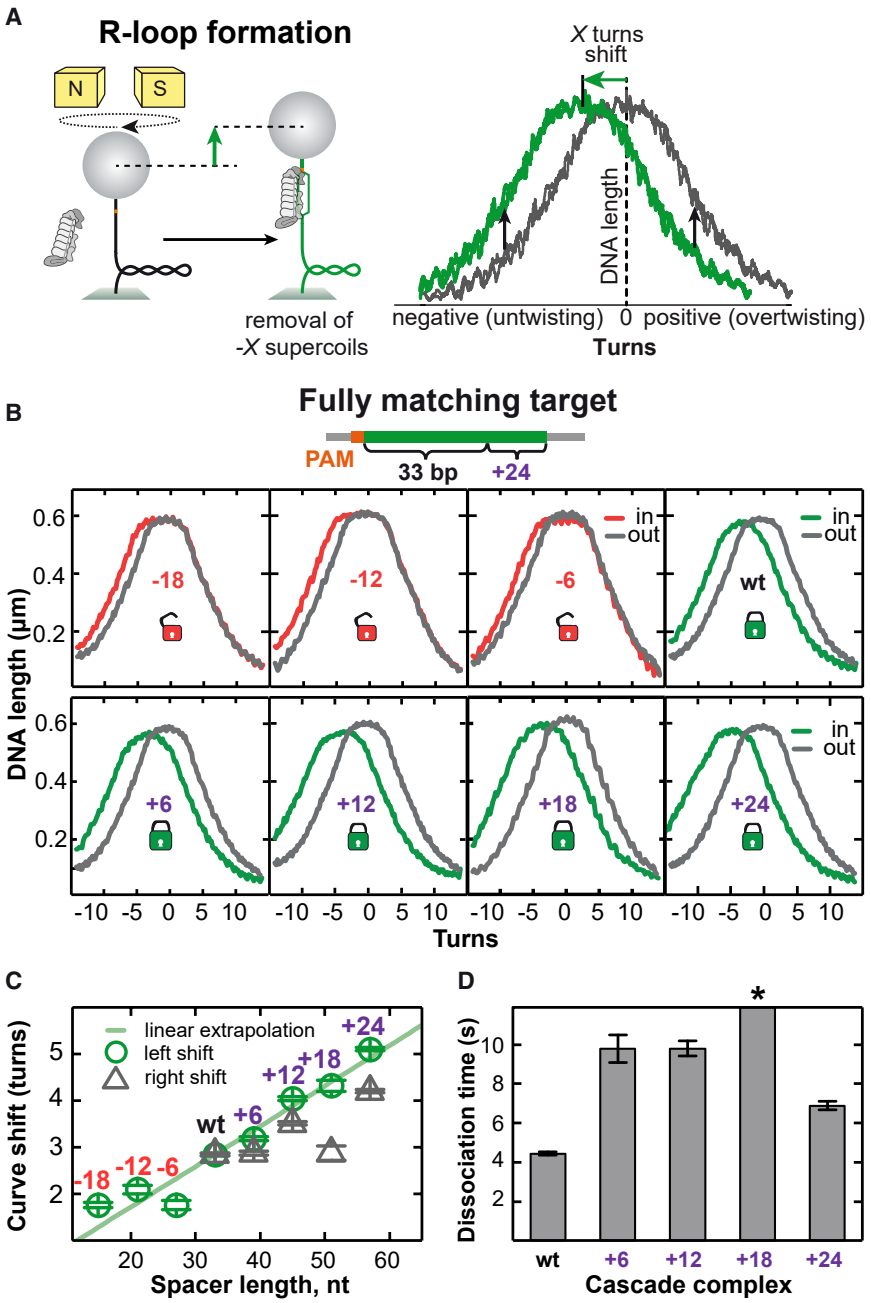


Figure 2. R-loop Formation and Locking on Fully Matching Targets

(A) Scheme of the magnetic tweezer assay, in which a magnetic-bead-tethered DNA is stretched and twisted using the magnetic field of a magnet pair (left). DNA twisting leads to length reduction because of formation of a DNA superhelix so that a typical bell-shaped supercoiling curve is observed (gray curve on the right). R-loop formation causes local helix unwinding and, thus, global overwinding (positive supercoiling) of the entire molecule. In the presence of an R-loop, the supercoiling curve (green curve on the right) is therefore shifted leftward by the number of untwisted helical turns in the R-loop structure.

(B) Supercoiling curves obtained for the complexes containing spacers of different lengths at a force of 0.4 pN. Gray lines represent supercoiling curves in the absence of an R-loop. Green (locked R-loop) and red (unlocked R-loop) lines represent rotation curves in the presence of an R-loop. Curve recording was carried out from negative to positive turns. A shift of the left side of the rotation curve indicates the presence of an R-loop at negative supercoiling, and a shift on the right side, correspondingly, the presence of a stable, i.e. locked, R-loop that resists positive supercoiling. R-loops formed at negative supercoiling for all complexes. Only R-loops for StCascades with WT, +6, +12, +18, and +24 spacers were found to be locked (indicated by closed versus open lock symbols).

(C) R-loop-mediated DNA unwinding as a function of spacer length. The unwound DNA turns were determined from the shift of the supercoiling curve either on the left side (green circles) or the right side (gray triangles). The green solid line shows the expected untwisting in the case of a proportional dependence on spacer length, using the WT spacer data as a reference.

(D) Mean dissociation times of R-loops recorded at defined positive supercoiling (torque of +22.7 pN nm, F = 3 pN) for complexes exhibiting R-loop locking. The asterisk for the +18 spacer indicates that no R-loop dissociation was obtained within the measurement time of 1,800 s.

All single-molecule experiments for a particular target sequence were performed on at least two different DNA molecules during different days. Error bars of all correspond to SEM.

circles; r = 0.98). The DNA untwisting measured for the engineered complexes matched the expectation when proportionally extrapolating the untwisting of the WT StCascade over the spacer length (green line in Figure 2B). Thus, these complexes can promote the formation of R-loops that extend over the whole crRNA spacer; i.e., up to 57-nt length. The shift of the right side of the supercoiling curves provided the DNA untwisting when locked R-loops were challenged with moderate positive supercoiling. For all complexes with elongated spacers, the untwisting at positive supercoiling was significantly smaller than at negative supercoiling ($p < 0.05$), whereas, for the WT complex, equal values were obtained (Figure 2C, gray triangles). This suggests that, in case of extended spacers, the PAM-distal end of the R-loop is not fully stabilized by locking.

Locked R-loops can be dissociated with positive supercoiling and elevated stretching force (Figure S4A) that translates into increased torsion (Maffeo et al., 2010; Szczelkun et al., 2014). Quantifying the mean dissociation time at a given torsion (Figure S4) provides a measure of R-loop stability. R-loops for the +18 complex could not be dissociated at the given torsion of 22.7 pN nm over the timescale of the experiment of approximately 30 min. Similar strong locking has been observed previously for

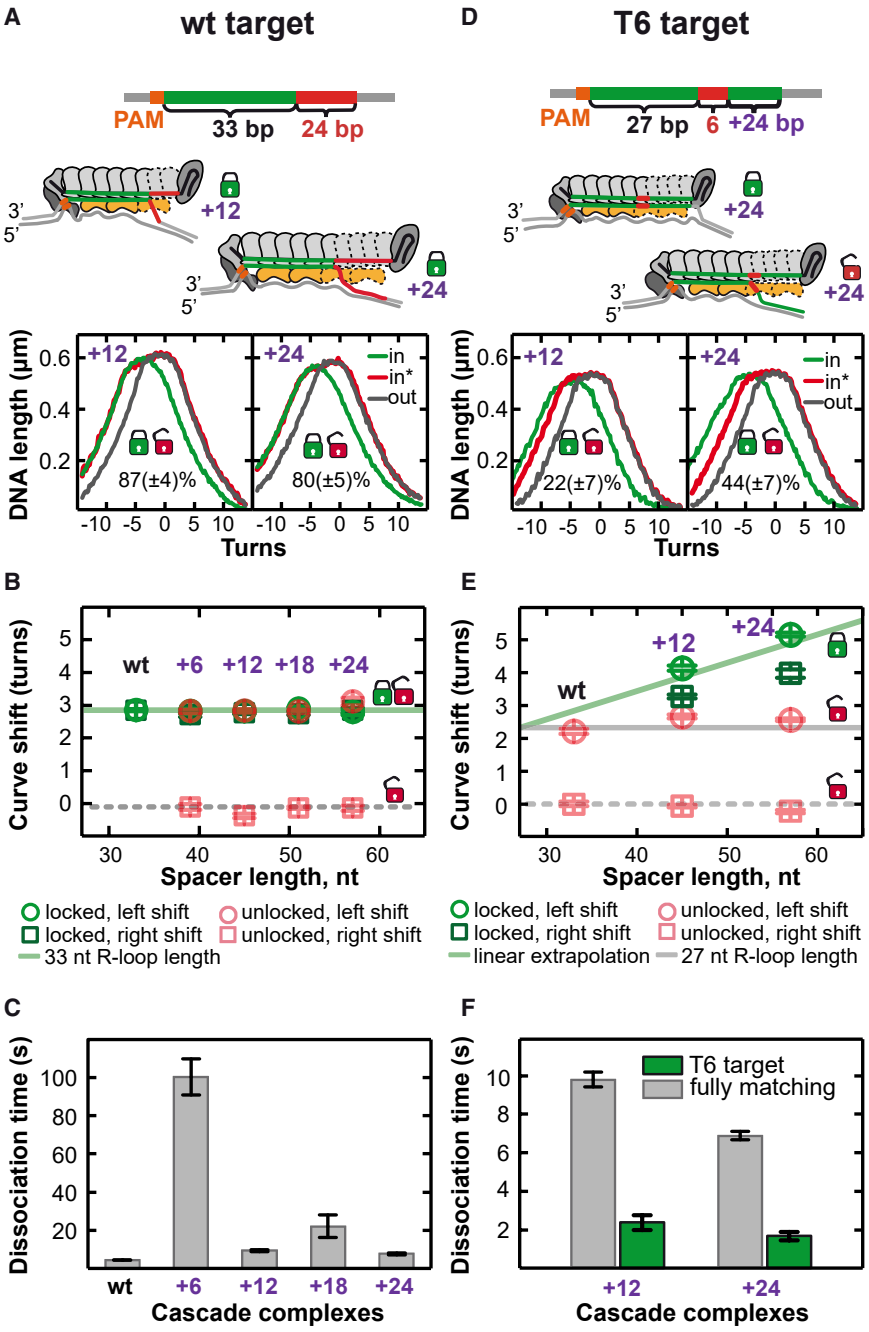


Figure 3. R-loop Formation and Locking at Partially Matching Targets

(A and D) Supercoiling curves for the +12 and +24 complexes recorded using the WT target and the T6 target. Gray curves were taken in the absence of an R-loop and green and red curves in the presence of an R-loop from negative to positive turns. For curves marked in green, the right curve side was shifted, indicating stable R-loop locking, whereas for the curves shown in red, no shift was seen, indicating the absence of locking. For the T6 target, supercoiling curves of unlocked complexes (shown in red) exhibited reduced untwisting at negative supercoiling compared with supercoiling curves of locked complexes (shown in green). The percentage of the fraction of locked R-loops is indicated below the lock symbol. The WT target was matching over the length of the WT spacer but was mismatched at the extended spacer part (scheme and cartoons in A). The T6 target was mismatched over 6 nt at the PAM-distal end of the WT spacer part (red part in scheme in D).

(B and E) R-loop-mediated DNA unwinding as a function of spacer length for the WT target (B) and the T6 target (E). DNA unwinding was obtained both from the shift of the left (circles) and the right (squares) side of the supercoiling curve. Data points shown in green were obtained for locked R-loops, whereas data points shown in red were obtained for unlocked R-loops.

(C and F) R-loop dissociation times measured at defined positive supercoiling (torque of 22.7 pN nm) for the WT target (C) and the T6 target (F). For the T6 target, the corresponding dissociation times measured on the fully matching target are shown for comparison (gray bars).

All single-molecule experiments for a particular target sequence were performed on at least two different DNA molecules during different days. Error bars of all correspond to SEM.

EcCascade (Krivoy et al., 2018). For all other StCascade complexes with elongated spacers, the R-loops were also found to be significantly more stable than those of the WT ($p < 0.05$) but lacked any obvious correlation with spacer length (Figure 2D).

R-loop Locking Requires WT-Sized R-loops Independent of Spacer Length

Six or more mismatches at the PAM-distal end abolish locking for the WT StCascade complex (Krivoy et al., 2018; Szczelkun et al., 2014); i.e., a minimum R-loop length of 28 bp is required to support locking. In agreement with this, all complexes with shortened spacers (−6, −12, and −18) did not support R-loop locking (Figure 2B). To obtain a minimum R-loop length for the StCascade complexes with elongated spacers, we first tested whether locking occurred for R-loops of the WT length. For this purpose, a target was used that allowed base-pairing over the length of the WT spacer but not with the extended part of the spacer. Similar to the WT complex (Figure 2B), all StCascade complexes with an extended spacer could form stably locked R-loops (supercoiling curves for the +12 and +24 complexes in Figure 3A). In contrast to the WT complex, 13% to 20% of the R-loops formed by extended spacers remained unlocked and dissociated at positive supercoiling. Experimentally, this was observed by a corresponding fraction of supercoiling curves (red curve in Figure 3A) that lacked a shift on the right side of the curve (see above). This indicated either slower R-loop locking or structural

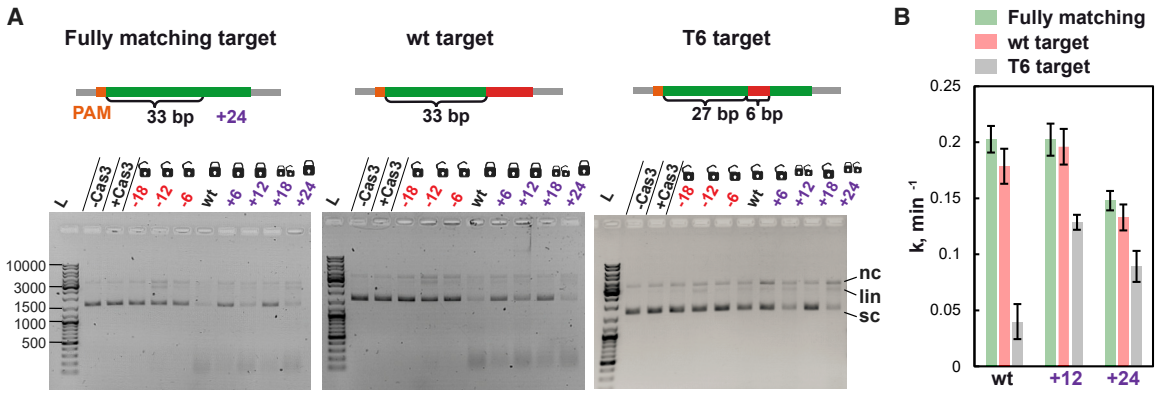


Figure 4. DNA Cleavage Mediated by StCascade Complexes with Altered Spacer Lengths

(A) Cleavage of supercoiled plasmid DNA by Cas3 containing either a fully matching, a 33-bp (WT), or a T6 target. Cleavage was tested after allowing the StCascade complexes to pre-bind the targets, and then the mixture was incubated for 20 min at 37°C. DNA cleavage was seen as the disappearance of the initial supercoiled species (sc) and the appearance of nicked (nc) and linear (lin) intermediates as well as of short degradation products, visible as fuzzy bands at low molecular weight. The fully matching target and the WT-sized target are efficiently degraded by complexes with WT, +12, and +24 spacers and less efficiently by complexes with +6 and +18 spacers. The control sample (−Cas3) contained DNA and Cascade but no Cas3, whereas the (+Cas3) control contained DNA and Cas3 but no Cascade.

(B) DNA cleavage rates of linearized plasmid DNA by the WT, +12, and +24 complexes, measured for the fully matching, WT-sized, and T6 target. Error bars correspond to SD of the mean from at least three repeated measurements.

heterogeneity between individual StCascade complexes. The measured DNA untwisting from R-loop formation (STAR Methods) was, within error, equal to the WT complex for negative and positive supercoiling (Figure 3B). Thus, the complexes with extended spacers formed R-loops of approximately 33-bp length on this target. Locking of WT-sized R-loops was additionally confirmed by an electrophoretic mobility shift assay (EMSA) that revealed similar high affinities of StCascade complexes with extended spacers on targets that either support or restrict base-pairing with the extended part of the spacer (Figure S5A).

The WT-sized R-loops that were formed by the complexes with elongated spacers also exhibited significantly increased stability when challenged with high positive torsion ($p < 0.05$; Figure 3C). There was no obvious correlation between the stability of these R-loops and spacer length.

R-loop Locking Involves Contacts at the Extended Spacer Part

To further define the minimum R-loop length that is required for locking by the StCascade complexes with elongated spacers, we carried out R-loop formation experiments using a target containing mismatches from base pair position 28–33 (with respect to the PAM) but no mismatches in the extended spacer part (called T6 target) (Figure 3D). Such a target is known to abolish R-loop locking for the WT complex (Krivoy et al., 2018; Szczelkun et al., 2014) because the R-loop cannot form over more than 27 bp. For the WT complex, the corresponding DNA untwisting was 2.2 ± 0.1 turns (Figure 3E), in agreement with an R-loop of 27-bp length. When testing the +12 and +24 complexes on this substrate, two different populations of formed R-loops were observed. The majority of R-loops (78% and 56% for the +12 and +24 complex, respectively) were found to be unlocked (Figure 3D, red supercoiling curve without a shift on its right side) so that they dissociated immediately when positive supercoiling was applied. The untwisting for these R-loops was ~2.6 ± 0.1 turns (Figure 3E), in agreement with a length of 27 bp as found for the WT complex on this target. Additionally, locked R-loops were detected for the +12 and +24 complexes. The corresponding DNA untwisting at negative supercoiling was 4.1 ± 0.1 turns and 5.2 ± 0.1 turns, equaling, within error, the untwisting on the full-length targets (Figure 2B). This indicates that the locked R-loops extended over the full spacer length; i.e., that they were able to bypass 5 consecutive mismatches. As seen before on the full-length targets, the untwisting was lower when the R-loops were subjected to positive supercoiling (Figure 3E). The presence of a fraction of locked R-loops for the +12 and +24 complexes is in agreement with EMSA experiments that revealed, for the T6 target, an intermediate affinity between the non-specific and full-length target (Figure S5A).

The locked R-loops found on the T6 target were considerably less stable than the R-loops formed by the +12 and +24 complexes on the fully matching and the WT-sized target ($p < 0.05$; Figures 2D and 3F). Their stability was even lower than for R-loops formed by the WT complex on a fully matching target (see Figure 2D for comparison). This indicates that base-pairing between the target strand and crRNA at base pair positions 28–33 is a major determinant for R-loop locking. However, additional contacts made to the crRNA-target-strand hybrid in the extended spacer region also contribute to locking in the case of engineered complexes with an extended crRNA guide.

DNA Cleavage Occurs at WT-like Rates Independent of Spacer Length

R-loop locking is considered to be a pre-requisite for Cascade-mediated target cleavage by Cas3 (van Erp et al., 2015; Rutkauskas et al., 2015). To test whether StCascade complexes with

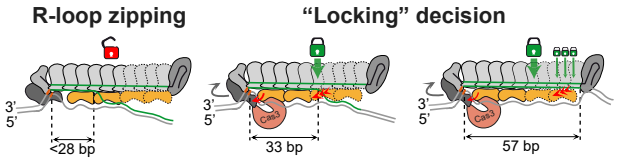


Figure 5. Model of the Decision-Making Process by Cascade Complexes with Elongated spacers

The Cascade complexes form R-loops in a zipper-like fashion, starting from the PAM. The R-loops are unlocked, (i.e., reversible) when they are shorter than 28 bp (left). R-loops that exceed 28 bp in length become locked independent of the spacer length (center). For complexes with elongated spacers, R-loops can extend over the full spacer length. Protein contacts to the R-loop at the PAM-distal end provide additional R-loop stabilization (right).

altered spacer length support the degradation of bound targets, we carried out DNA cleavage experiments. To this end, StCascade complexes were allowed to pre-bind to supercoiled or linearized DNA substrates carrying a single target site. Subsequently, Cas3 was added. None of the complexes with shortened spacers (−18, −12, and −6) supported DNA cleavage of any of the targets used in this study (Figure 4A). This is in agreement with the inability of these complexes to form locked R-loops (Figure 2B). The WT, +12, and +24 complexes exhibited efficient DNA cleavage. Less efficient cleavage was seen for the +6 and +18 complexes, in agreement with a reduced fraction of functional complexes in these samples (Figure 4A; Table S1).

Additionally, we recorded the cleavage kinetics of linearized target substrates for the WT, +12, and +24 complexes (Figure 4B; Figure S5B). The complexes cleaved the fully matching and the WT-sized target at comparable rates of 0.20 min^{-1}, 0.20 min^{-1}, and 0.15 min^{-1} for the WT, +12, and +24 complexes, respectively. The cleavage rate of the T6 target was 5-fold lower (0.04 min^{-1}) for WT StCascade compared with the fully matching target and 1.6- and 1.7-fold lower for the +12 (0.13 min^{-1}) and +24 (0.09 min^{-1}) complexes, respectively. This additionally confirmed that only part of the complexes with elongated spacers could undergo locking on this substrate (Figure 4B; Figure S5B). Altogether, these data confirm that DNA cleavage is obtained for all of the engineered complexes as soon as the targets support R-loop locking.

DISCUSSION

In this work, we produced StCascade complexes containing crRNA spacers of altered length ranging from 15–57 nt. These complexes could successfully form R-loops that extended over the full spacer, representing a more than 70% increase in R-loop length for the longest tested spacer. Complexes for which the spacer length differed by even multiples of 6 nt from the WT (i.e., −12, +12, and +24) were found to be stable and homogeneous (Figure 1D; Table S1). In contrast to reports for EcCascade (Kuznedelov et al., 2016; Luo et al., 2016), complexes for which the spacer length differed by odd multiples of 6 nt (i.e., −18, −6, +6, and +18) were found to be unstable, with the dominating population being of significantly reduced molecular weight (Table S1). The different complex stabilities and, thus, activities need to be considered in applications for these proteins. Stability differences can be explained by the periodicities of the Cas7 backbone and the Cse2 filament with respect to the crRNA (Jackson et al., 2014; Mulepati et al.,2014; Zhao et al., 2014). Cas7 exhibits a periodicity of 6 nt so that each alteration of the spacer length by 6 nt changes the number of Cas7 units by one. Cse2 exhibits a periodicity of 12 nt so that its number changes by one for every spacer length alteration of 12 nt, in agreement with the mass spectrometry data (Figure 1E; Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014). In this case, the ends of the Cas7 backbone (including Cas6) and the Cse2 filaments remain approximately aligned and can most likely establish protein-protein contacts as in the WT complex. This differs when the spacer length changes by an odd multiple of 6 nt. Then the ends of the Cas7 backbone and the Cse2 filament are misaligned by approximately half the length of a Cse2 subunit. Missing protein-protein contacts (e.g., between Cse2 and the Cas6 or Cas7 subunits; Zhao et al., 2014) can thus destabilize the final complex sufficiently so that predominantly subcomplexes are observed (Table S1). Nonetheless, these types of complexes could form R-loops (Figure 2B) that covered the whole length of the spacer. Most likely, these R-loops originate from a minority fraction of full complexes that were not observed by native mass spectrometry.

The central result of this study was to elucidate the decision-making by these complexes. Uniformly, R-loop locking occurred when an R-loop could extend over more than 27 bp independent of the spacer length of the complex (Figure 5). Consistently, the same R-loop length was sufficient to trigger Cas3-mediated DNA cleavage. This sets considerable limits for potential specificity improvements by the elongated complexes with respect to off-targeting. In fact, we do not expect greatly improved specificity with respect to DNA binding and interference by Cas3, which, however, needs to be more thoroughly tested by assessing the effect of mismatches and other target alterations. An exception may be applications that require precise positioning of functional domains, such as transcriptional activators, dimerizing DNA nucleases, or base-editing enzymes (Guilinger et al., 2014; Tang and Fu, 2018; Shimatani et al., 2017). When fused to the PAM-distal end of Cascade, the required positioning of the domain with respect to the DNA should only be achieved for R-loops that extend over the full spacer. For locked R-loops that are not fully extending on elongated spacers, the DNA would exit the Cascade complex at a significant distance from the complex end so that the positioning would be lost.

Despite the common minimum R-loop length that triggered locking for all complexes, R-loop stability was found to be considerably increased for elongated spacers compared with the WT spacer (Figures 2D and 3C). This is in agreement with observations *in vivo* that extended spacers can improve Cascade-mediated gene silencing (Luo et al., 2016). However, no clear correlation of R-loop stability with the R-loop or the spacer length was observed. Unlocked R-loops dissociated immediately after low positive torsion was applied, independent of R-loop length. This provides evidence that RNA:DNA base-pairing is not the major determinant of R-loop stabilization. Rather, stabilization is achieved by the contacts of the protein complex to the R-loop. Possibly, interactions at the exit point of the DNA from the complex influence stability, which would explain the absent

correlation with R-loop length. An influence of the DNA exit point on R-loop stability may also be responsible for the considerable effect of mismatches in the extended spacer region on transcriptional regulation and explain the much lower effect of these mismatches on plasmid interference (Luo et al., 2016). In this case, the apparent increased specificity of the complexes with elongated spacers would arise from the different capacities of the complexes to form a road block against transcription.

Interestingly, a minor population of the complexes with elongated spacers supported R-loop locking even in the case of mismatches in the ''locking-sensitive'' region from positions 28–33 bp. The tight interactions between the R-loop and the Cas7-Cse2 filament in this region may help the R-loop to propagate through the mismatched region. Furthermore, this finding demonstrates that not only contacts between the R-loop and the Cse2 filament in this region but also in the extended spacer region are important for locking. The combination of these interactions should cause the increased stability of the R-loops for the engineered Cascade variants.

Overall, our data reveal that the WT StCascade complex is optimally adapted to recognize a protospacer length of 33 bp despite its modular architecture, which may explain why longer spacers would not be preferred naturally. Elongated spacers neither increase the specificity nor the efficiency of target recognition. This optimization may be the result of coevolution of Cascade alongside the evolution of spacer selection by Cas1 and Cas2 (Koonin and Makarova 2017). Extending the recognition region for Cascades with elongated spacers to the whole spacer region will therefore most likely require significant protein engineering efforts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - DNA and proteins
  - Native mass spectrometry
  - Small angle X-ray scattering experiments
  - Single-molecule experiments
  - EMSA experiments
  - Cas3-mediated DNA cleavage experiments
  - Northern blot hybridization assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.08.033.

ACKNOWLEDGMENTS

This work was supported by an ERC consolidator grant (GA 724863) and grant SE 1646/9-1 within Priority Program SPP 2141 of the Deutsche Forschungsgemeinschaft (DFG) to R.S., as well as an intramural Vilnius University grant to V.S. C.S. acknowledges funding from the Federal Ministry for Education and Research (BMBF, ZIK Program, 03Z22HN22), the European Regional Development Funds (EFRE, ZS/2016/04/78115), and the MLU Halle-Wittenberg. I.S. acknowledges funding of a research internship to Leipzig University from the Research Council of Lithuania. The SAXS experiment at the P12 EMBL beamline at the PETRA III storage ring of DESY synchrotron, Hamburg, Germany, was supported by iNEXT project number 653706, funded by the Horizon 2020 Program of the European Union. We thank dr. G. Tamulaitis for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

I.S., M.R., T.S., C.S., V.S., and R.S. designed the research. I.S., M.R., T.S., E.M., and S.W. performed the research. I.S., M.R., and R.S. wrote the paper. All authors contributed to data analysis.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 6, 2019
Revised: June 21, 2019
Accepted: August 9, 2019
Published: September 17, 2019

REFERENCES

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat. Commun. *9*, 1911.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Blanchet, C.E., Spilotros, A., Schwemmer, F., Graewert, M.A., Kikhney, A., Jeffries, C.M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., et al. (2015). Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). J. Appl. Cryst. *48*, 431–443.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Brutzer, H., Luzzietti, N., Klaue, D., and Seidel, R. (2010). Energetics at the DNA supercoiling transition. Biophys. J. *98*, 1267–1276.

Chang, Y., Su, T., Qi, Q., and Liang, Q. (2016). Easy regulation of metabolic flux in Escherichia coli using an endogenous type I-E CRISPR-Cas system. Microb. Cell Fact. *15*, 195.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., and Svergun, D.I. (2017). *ATSAS 2.8*: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J. Appl. Cryst. *50*, 1212–1225.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. USA *109*, E2579–E2586.

Gleditzsch, D., Müller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., Bange, G., and Randau, L. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. Nucleic Acids Res. *44*, 5872–5882.

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., and Randau, L. (2019). PAM Identification by CRISPR-Cas Effector Complexes: Diversified Mechanisms and Structures. RNA Biol. *16*, 504–517.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 172.

Guilinger, J.P., Thompson, D.B., and Liu, D.R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. *32*, 577–582.

Hatoum-Aslan, A. (2018). Phage Genetic Engineering Using CRISPR–Cas Systems. Viruses *10*, 335.

Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. Nature *530*, 499–503.

Hernández, H., and Robinson, C.V. (2007). Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. Nat. Protoc. *2*, 715–726.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Huhle, A., Klaue, D., Brutzer, H., Daldrop, P., Joo, S., Otto, O., Keyser, U.F., and Seidel, R. (2015). Camera-based three-dimensional real-time particle tracking at kHz rates and Ångström accuracy. Nat. Commun. *6*, 5885.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science *345*, 1473–1479.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. *18*, 529–536.

Josephs, E.A., Kocak, D.D., Fitzgibbon, C.J., McMenemy, J., Gersbach, C.A., and Marszalek, P.E. (2015). Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. Nucleic Acids Res. *43*, 8924–8941.

Jung, C., Hawkins, J.A., Jones, S.K., Jr., Xiao, Y., Rybarski, J.R., Dillard, K.E., Hussmann, J., Saifuddin, F.A., Savran, C.A., Ellington, A.D., et al. (2017). Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. Cell *170*, 35–47.e13.

Karvelis, T., Gasiunas, G., and Siksnys, V. (2017). Harnessing the natural diversity and in vitro evolution of Cas9 to expand the genome editing toolbox. Curr. Opin. Microbiol. *37*, 88–94.

Kiro, R., Shitrit, D., and Qimron, U. (2014). Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. RNA Biol. *11*, 42–44.

Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S., and Sternberg, S.H. (2019). Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature *571*, 219–225.

Koonin, E.V., and Makarova, K.S. (2017). Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. Genome Biol. Evol. *9*, 2812–2825.

Kozin, M.B., and Svergun, D.I. (2001). Automated Matching of High- and Low-Resolution Structural Models. J. Appl. Cryst. *34*, 33–41.

Krivoy, A., Rutkauskas, M., Kuznedelov, K., Musharova, O., Rouillon, C., Severinov, K., and Seidel, R. (2018). Primed CRISPR adaptation in Escherichia coli cells does not depend on conformational changes in the Cascade effector complex detected in Vitro. Nucleic Acids Res. *46*, 4087–4098.

Kuznedelov, K., Mekler, V., Lemak, S., Tokmina-Lukaszewska, M., Datsenko, K.A., Jain, I., Savitskaya, E., Mallon, J., Shmakov, S., Bothner, B., et al. (2016). Altered stoichiometry Escherichia coli Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation. Nucleic Acids Res. *44*, 10849–10861.

Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.H., Lee, K., Jung, I., Kim, D., Kim, S., and Kim, J.S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. Nat. Commun. *9*, 3048.

Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol. Cell *62*, 137–147.

Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., Chen, L., Liang, Y.X., and She, Q. (2016). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Res. *44*, e34.

Lim, Y., Bak, S.Y., Sung, K., Jeong, E., Lee, S.H., Kim, J.S., Bae, S., and Kim, S.K. (2016). Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. Nat. Commun. *7*, 13350.

Luo, M.L., Mullis, A.S., Leenay, R.T., and Beisel, C.L. (2015). Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic Acids Res. *43*, 674–681.

Luo, M.L., Jackson, R.N., Denny, S.R., Tokmina-Lukaszewska, M., Maksimchuk, K.R., Lin, W., Bothner, B., Wiedenheft, B., and Beisel, C.L. (2016). The CRISPR RNA-guided surveillance complex in Escherichia coli accommodates extended RNA spacers. Nucleic Acids Res. *44*, 7385–7394.

Maffeo, C., Schöpflin, R., Brutzer, H., Stehr, R., Aksimentiev, A., Wedemann, G., and Seidel, R. (2010). DNA-DNA interactions in tight supercoils are described by a small effective charge density. Phys. Rev. Lett. *105*, 158101.

Morgner, N., and Robinson, C.V. (2012). Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. Anal. Chem. *84*, 2939–2948.

Mulepati, S., Héroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science *345*, 1479–1484.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science *361*, 1259–1262.

Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P.R., and Lundgren, M. (2015). Efficient programmable gene silencing by Cascade. Nucleic Acids Res. *43*, 237–246.

Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. Cell *163*, 854–865.

Roy, B., Zhao, J., Yang, C., Luo, W., Xiong, T., Li, Y., Fang, X., Gao, G., Singh, C.O., Madsen, L., et al. (2018). CRISPR/Cascade 9-Mediated Genome Editing-Challenges and Opportunities. Front. Genet. *9*, 240.

Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V., and Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. Cell Rep. *10*, 1534–1543.

Rutkauskas, M., Krivoy, A., Szczelkun, M.D., Rouillon, C., and Seidel, R. (2017). Single-Molecule Insight Into Target Recognition by CRISPR-Cas Complexes. Methods Enzymol. *582*, 239–273.

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., Teramura, H., Yamamoto, T., Komatsu, H., Miura, K., et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nat. Biotechnol. *35*, 441–443.

Singh, D., Mallon, J., Poddar, A., Wang, Y., Tippana, R., Yang, O., Bailey, S., and Ha, T. (2018a). Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). Proc. Natl. Acad. Sci. USA *115*, 5444–5449.

Singh, D., Wang, Y., Mallon, J., Yang, O., Fei, J., Poddar, A., Ceylan, D., Bailey, S., and Ha, T. (2018b). Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. Nat. Struct. Mol. Biol. *25*, 347–354.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J. *30*, 1335–1342.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J. *32*, 385–394.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88.

Sobott, F., Hernández, H., McCammon, M.G., Tito, M.A., and Robinson, C.V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. Anal. Chem. *74*, 1402–1407.

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.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Cryst. *25*, 495–503.

Svergun, D.I., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J. Appl. Cryst. *28*, 768–773.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H.J. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophys. J. *80*, 2946–2953.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc. Natl. Acad. Sci. USA *111*, 9798–9803.

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, Č., Nwokeoji, A.O., Dickman, M.J., Horvath, P., and Siksnys, V. (2014). Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus. Mol. Cell *56*, 506–517.

Tang, Y., and Fu, Y. (2018). Class 2 CRISPR/Cas: an expanding biotechnology toolbox for and beyond genome editing. Cell Biosci. *8*, 59.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*, 1143–1149.

Tsai, S.Q., and Joung, J.K. (2016). Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. Nat. Rev. Genet. *17*, 300–312.

van Erp, P.B.G., Jackson, R.N., Carter, J., Golden, S.M., Bailey, S., and Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. Nucleic Acids Res. *43*, 8381–8391.

Volkov, V., and Svergun, D. (2003). Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Cryst. *36*, 860–864.

Westra, E.R., van Erp, P.B., Künne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol. Cell *46*, 595–605.

Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS Genet. *9*, e1003742.

Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. Cell *170*, 48–60.e11.

Xiao, Y., Luo, M., Dolan, A., Liao, M., and Ke, A. (2018). Structure basis for RNA-guided DNA degradation by Cascade and Cas3. Science *361*, eaat0839.

Yosef, I., Manor, M., Kiro, R., and Qimron, U. (2015). Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. Proc. Natl. Acad. Sci. USA *112*, 7267–7272.

Yoshioka, K. (2002). KyPlot – A User-oriented Tool for Statistical Data Analysis and Visualization. Comput. Stat. *17*, 425.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. Nature *515*, 147–150.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Escherichia coli BL21 (DE3)	New England Biolabs	#C2527
Escherichia coli DH5α	Thermo Fisher Scientific	#18265017
Chemicals, Peptides, and Recombinant Proteins		
SYBR Gold Nucleic Acid Gel Strain	Thermo Fisher Scientific	S11494
[γ^{33}P]-adenosine-5′-triphospate	Hartmann Analytic	FF-10130
[γ^{32}P]-adenosine-5′-triphospate	Hartmann Analytic	FP-1013
T4 polynucleotide kinase	Thermo Fisher Scientific	EK0031
FastDigest SapI	Thermo Fisher Scientific	FD1934
FastDigest Eco31I	Thermo Fisher Scientific	FD0293
XmaI	Thermo Fisher Scientific	ER0171
FastDigest PstI	Thermo Fisher Scientific	FD0614
Ribolock RNase inhibitor	Thermo Fisher Scientific	EO0382
Cascade complex (wt)	Sinkunas et al., 2013	N/A
Csm complex	Tamulaitis et al., 2014	N/A
Cas3	Sinkunas et al., 2013	N/A
Critical Commercial Assays		
TranscriptAid T7 High Yield Transcription Kit	Thermo Fisher Scientific	K0441
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific	K0502
GeneJET PCR Purification Kit	Thermo Fisher Scientific	K0701
Oligonucleotides		
Probe 1 5′-ccccgcgtgtgcgggaaaaacactaaaag	Metabion	N/A
Probe 2 5′-gcgtgggaggccattgatataggtatagg	Metabion	N/A
Probe 3 5′-gcagcctaggttaatatatgacctaaatttgt aggatcac	Metabion	N/A
Recombinant DNA		
pCDF-Cascade	Sinkunas et al., 2013	N/A
pBAD-Cas7-C-His	Sinkunas et al., 2013	N/A
pACYC-Duet1	Merck	#71147
pACYC-minCR (from −12 to +24)	This manuscript	N/A
pACYC-minCR-Eco31I/SapI	This manuscript	N/A
pUC19	New England Biolabs	N3041
pUC-T6-target	This manuscript	N/A
pUC-wt-target	This manuscript	N/A
pUC-TS132/133	Sinkunas et al., 2013	N/A
pUC-TS140/141	Sinkunas et al., 2013	N/A
Software and Algorithms		
MassLynx v4.0	Waters Inc.	https://www.waters.com/waters/en_US/MassLynx-Mass-Spectrometry-Software-/
Massign	Morgner and Robinson, 2012	http://massign.chem.ox.ac.uk/
Kyplot 2.0	Yoshioka, 2002	https://kyplot.software.informer.com/2.0/
OptiQuant v.3	PerkinElmer	N/A
ATSAS	Franke et al., 2017	https://www.embl-hamburg.de/biosaxs/software.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GNOM 5.0	Svergun, 1992	https://www.embl-hamburg.de/biosaxs/gnom.html
GASBOR	Svergun et al., 2001	https://www.embl-hamburg.de/biosaxs/gasbor.html
DAMAVER	Volkov and Svergun, 2003	https://www.embl-hamburg.de/biosaxs/damaver.html
CRYSOL	Svergun et al., 1995	https://www.embl-hamburg.de/biosaxs/crysol.html
SUPCOMB	Kozin and Svergun, 2001	https://www.embl-hamburg.de/biosaxs/supcomb.html
Other		
Dynabeads MyOne Streptavidin C1	Thermo Fisher Scientific	65001

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ralf Seidel (ralf.seidel@physik.uni-leipzig.de). Reagents (such as plasmids or proteins) generated in the study are freely available under material transfer agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

StCascade complexes with altered spacer lengths were expressed in *E.coli* BL21 (DE3) cells at 37°C in LB broth ("Thermo Fisher") using the pACYC-minCR derivatives (Cm^R). 1/20 volume of overnight culture were inoculated in fresh LB medium supplemented with 25 μg/ml streptomycin, 25 μg/ml ampicillin, 30 μg/ml chloramphenicol. The expression was induced at mid-log phase (OD_{600nm} 0.5-0.8) by adding 1 mM IPTG and 0.2% L(+)-arabinose and the cells were further incubated at 37°C for 3–4 h, proteins were purified as described (Sinkunas et al., 2013). *E. coli* DH5α strain was used for cloning and plasmid amplification, it was grown at 37°C on liquid LB borth or LB agar ("Thermo Fisher").

METHOD DETAILS

DNA and proteins

For the production of crRNAs with different spacer lengths a sequence containing two repeats with a SapI cleavage site after the first repeat and an Eco31I cleavage site before the second repeat was cloned into the pACYC-Duet1 vector (resulting vector pACYC-minCR-Eco31I/SapI). All spacer variants were introduced into the produced vector through the SapI and Eco31I sites using synthetic oligonucleotides with corresponding single stranded overhangs (Table S5). The chosen spacer sequences were based on spacer 1 of the CRISPR4 Cascade system from strain DGCC7710. StCascade complexes with altered spacer lengths were expressed in *E.coli* BL21 (DE3) cells and purified as described (Sinkunas et al., 2013) using the pACYC-minCR derivatives (Cm^R) instead of pACYC plasmid with homogenious CRISPR region pCRh. Cas3 protein was purified as previously described (Sinkunas et al., 2011). DNA substrates for single-molecule and bulk cleavage experiments were derived from vector pUC19 by cloning desired target sequences with a permissive AA dinucleotide PAM as blunt ended synthetic DNA duplexes into the XmaI site of the vector (Rutkauskas et al., 2015; Szczelkun et al., 2014; Table S3 for target sequences).

Native mass spectrometry

For native mass spectrometry approximately 2 μl containing ~5 μM StCascade complexes were used per experiment. First, the storage buffer was exchanged to 200 mM ammonium acetate, pH 7.0 using size exclusion spin columns (Micro Bio-Spin 6 columns, Bio-Rad). Spectra were then acquired on an Ultima quadrupole time-of-flight (Q-ToF) mass spectrometer modified for transmission of high masses (Sobott et al., 2002) using in-house prepared gold coated glass capillaries (Hernández and Robinson, 2007). Typical instrument conditions were: capillary voltage, 1.50 kV; cone voltage, 80 V; RF lens 1 energy, 80 V; collision energy, 20 V; aperture 3, 13.6 V. At least 100 scans of each complex were acquired and combined. Spectra were analyzed using MassLynx v4.0 and Massign (Morgner and Robinson, 2012) software.

Theoretical and experimentally determined molecular weights of the complexes are listed in Table S1.

Small angle X-ray scattering experiments

The scattering data for the wt StCascade complex was collected at the X33 EMBL beamline on the DORIS storage ring and the data for the +12 complex was collected at the P12 EMBL beamline at the PETRA III storage ring (Blanchet et al., 2015) (DESY, Hamburg, Germany).

The complexes were transferred into buffer containing 20 mM Tris-HCl pH 7.0, 0.5 M NaCl, 0.1 mM EDTA and 0.1 mM DTT. Concentrations were measured spectrophotometrically using absorption of 1 mg/ml solution (wt: 1.14 and +12: 1.6 AU280). Proteins were concentrated before measurements by ultrafiltration and 5-15 min centrifuged at the maximum speed before transferring into the sample changer. Data collection parameters and main structural parameters of complexes are given in Table S2. Scattering data (Figure S3) were analyzed using ATSAS software (Franke et al., 2017).The indirect Fourier transform and pair distance distribution functions were calculated by GNOM (Version 5.0) (Svergun, 1992). The complex shapes in solution were reconstituted out of SAXS data as pseudo-chain of dummy atoms using GASBOR (Svergun et al., 2001). The individual runs of the modeling were compared and averaged by DAMAVER (Volkov and Svergun, 2003). The theoretical scattering intensity of the EcCascade complex (PDB ID: 5CD4) (van Erp et al., 2015) was calculated and compared with the experimental SAXS data of the wt StCascade complex using CRYSOL (Svergun et al., 1995). Models were superimposed using SUPCOMB (Kozin and Svergun, 2001).

Single-molecule experiments

R-loop formation experiments were carried out in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 mg/ml BSA at a StCascade concentration of 150 pM (wt and complexes with spacers alterations by even multiples of 6 nt) or 1-15 nM (complexes with spacers alterations by odd multiples of 6 nt) using a home-built magnetic tweezers set-up (Huhle et al., 2015; Rutkauskas et al., 2017). The protein concentration was adapted such that clearly discernable R-loop formation events occurred in average ~10 s after the DNA became negatively supercoiled. Target-sequence containing DNA molecules were bound to 1 μm magnetic beads (MyOne, Invitrogen) on one end and to the flow cell on the other end. After lowering the magnets to stretch tethered molecules, DNA length measurements were started. Videomicroscopy was used to track the height of the DNA-tethered bead and a surface-bound reference particle (Huhle et al., 2015). The DNA length was calculated from the difference between the two bead positions. Applied rotations (in turns) of the magnet pair above the fluidic cell allowed to introduce DNA supercoiling. Data were recorded at an acquisition frequency of 120 Hz and smoothed to 2 Hz using a sliding average filter (only smoothed data are presented). After initial DNA characterization, proteins were added and changes in DNA length observed as a function of applied DNA twist.

The experiments were carried out in consecutive R-loop association-dissociation cycles, i.e., the DNA was held at negative supercoiling until an R-loop formed. Subsequently, the R-loop was challenged by applying positive supercoiling (Szczelkun et al., 2014). R-loops that dissociated as soon as positive supercoiling was applied (no visible shift of supercoiling curve at right side) were considered to be unlocked, stable R-loops with clearly discernable dissociation events were considered to be locked. Mean R-loop dissociation times were determined from exponential fits to cumulative distributions of the data (Figure S4B). Each mean-time was calculated from > 25 events. Average supercoiling curves were obtained from raw traces of at least 30 individual curves that were grouped into different categories (no-R-loop, unlocked R-loop and locked R-loop). Supercoiling changes that were associated to DNA untwisting during R-loop formation were analyzed by measuring the shift of either the left or the right side of the supercoiling curve as described before (Szczelkun et al., 2014). All single molecule experiments for a particular target sequence were performed on at least two different DNA molecules during different days. Error bars of all data shown correspond to standard errors of the mean. Statistical significances (p values) were calculated using a two sample t-Test.

EMSA experiments

222 bp PCR products (Table S5) bearing a single target sequence were obtained from the pUC19 based target plasmids pUC-fully-matching (also called pUC-TS132/133), pUC-wt-target, pUC-TS140/141 (non-specific target), pUC-T6-target (Table S4) and oligonucleotides pUC-dir, pUC-rev labeled at their 5′ ends using T4 polynucleotide kinase (Thermo Scientific) and [γ^{33}P]ATP (Hartmann Analytic). Labeled PCR products (0.1 nM) was incubated at 37°C for 10 min with increasing concentrations of the StCascade complexes in binding buffer 40 mM Tris, 20 mM acetic acid, 1 mM EDTA (pH 8.0), 150 mM NaCl, 0.1 mg/ml BSA, 10% glycerol. The samples were analyzed using 8% (w/v) polyacrylamide gels at 5 V/cm and visualized using a FLA-5100 phosphorimager (Fujilm).

Cas3-mediated DNA cleavage experiments

Cleavage reactions were performed with linear and supercoiled pUC plasmids, containing fully matching, wt, T6 and non-specific targets (Table S4). Reactions were carried out at 37°C in buffer containing 10 mM Tris–HCl (pH 8.0), 100 mM NaCl, 50 mM KCl, 5% (v/v) glycerol, 1.5 mM $MgCl_2$, 0.1 mM $NiCl_2$, 2 mM ATP and 0.1 mg/ml BSA. First, 2 nM supercoiled or PstI-linearized target plasmid was pre-incubated for 5 min with saturating concentrations of StCascade that were estimated based on EMSA experiments for StCascade complexes in absence of $MgCl_2$, $NiCl_2$ and ATP to allow the binding of the complexes to the targets. Subsequently, 100 nM Cas3 together with $MgCl_2$, $NiCl_2$ and ATP were added. The control reactions contained reaction buffer instead of the missing components. Reactions were stopped at different time points by adding 1/3 of the reaction volume of stop solution containing 67.5 mM EDTA, 27% (v/v) glycerol, 0.3% (w/v) SDS and 0.1% w/v Orange G. Reaction products were analyzed using 0.8% agarose gels and a voltage of 3 V/cm followed by visualization using ethidium bromide staining. The fraction of uncleaved plasmid as function of time was quantified from the gel images using OptiQuant (Perkin Elmer). The first order DNA cleavage rates of the reactions were

determined by fitting a single exponential decay to the data using KyPlot. Error bars provided in Figure 4B correspond to standard deviation (SD) of the mean from at least three repeated-measurements.

Northern blot hybridization assays

Northern blot analysis was performed as described previously (Gasiunas et al., 2012). RNA from the different StCascade complexes was purified using phenol/chloroform extraction. The RNA was detected using ^{32}P-labeled oligodeoxynucleotide probes (Table S4). An RNA used as a positive control was prepared by *in vitro* transcription with T7 polymerase (Thermo Fisher Scientific) using 336 bp PCR product of the pACYC-minCR-wt plasmid that included the wt-sized target site. This RNA represents the pre-crRNA that is formed in *E.coli* cells *in vivo* prior to crRNA maturation. The T7 promoter and the termination sequence were present in PCR product to check hybridization of the probe.

QUANTIFICATION AND STATISTICAL ANALYSIS

For statistical analysis of single molecule experimental data Origin software was used. Statistical details of experiments can be found in the main text of the publication and in the uploaded file "Single molecule data" in Mendeley data https://doi.org/10.17632/479324n2d5.1. All single molecule experiments for a particular target sequence were performed on at least two different DNA molecules during different days. Error bars of all data shown correspond to standard errors of the mean. Statistical significances (p values) were calculated using a two-sample t-Test. For DNA cleavage experiments the fraction of uncleaved plasmid as function of time was quantified from the gel images using OptiQuant software (Perkin Elmer). The first order DNA cleavage rates of the reactions were determined by fitting a single exponential decay to the data using KyPlot. Error bars provided in Figure 4B correspond to SD of the mean from at least three repeated-measurements.

DATA AND CODE AVAILABILITY

Original single molecule, SAXS and quantification data of plasmid cleavage is available in Mendeley data https://doi.org/10.17632/479324n2d5.1