# Supplementary material for

**Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli***

Paul B. G. van Erp[a1], Ryan N. Jackson[a1,] Joshua Carter[a1], Sarah M. Golden[a], Scott Bailey[b], Blake Wiedenheft[a2]

**Fig. S1.**

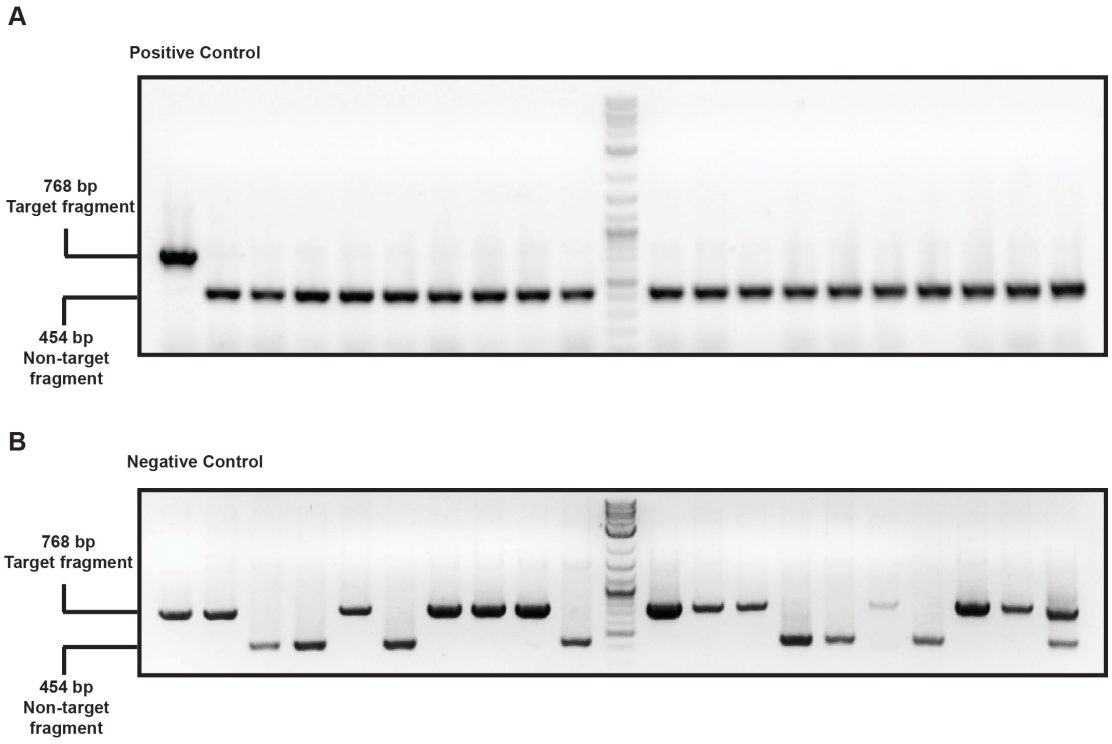


**Fig. S1. Representative agarose gels used to screen colonies in plasmid curing assays.** **(A)** Agarose gel of PCR products from 20 colonies of *E. coli* that contain a Type IE CRISPR-mediated programed to target a high-copy number (pUC19) plasmid (positive control). Only one colony contained the target plasmid (lane1, 768-bp fragment), while 19 colonies contained the non-target plasmid (454-bp fragment). **(B)** Agarose gel of PCR products from 20 colonies of *E. coli* that contain a Type IE CRISPR-mediated programed to target a sequence that is not represented in the plasmid (negative control). Twelve colonies contained the target plasmid, seven colonies contained the non-target plasmid, and one colony contained both plasmids (lane 20).

**Fig. S2.**

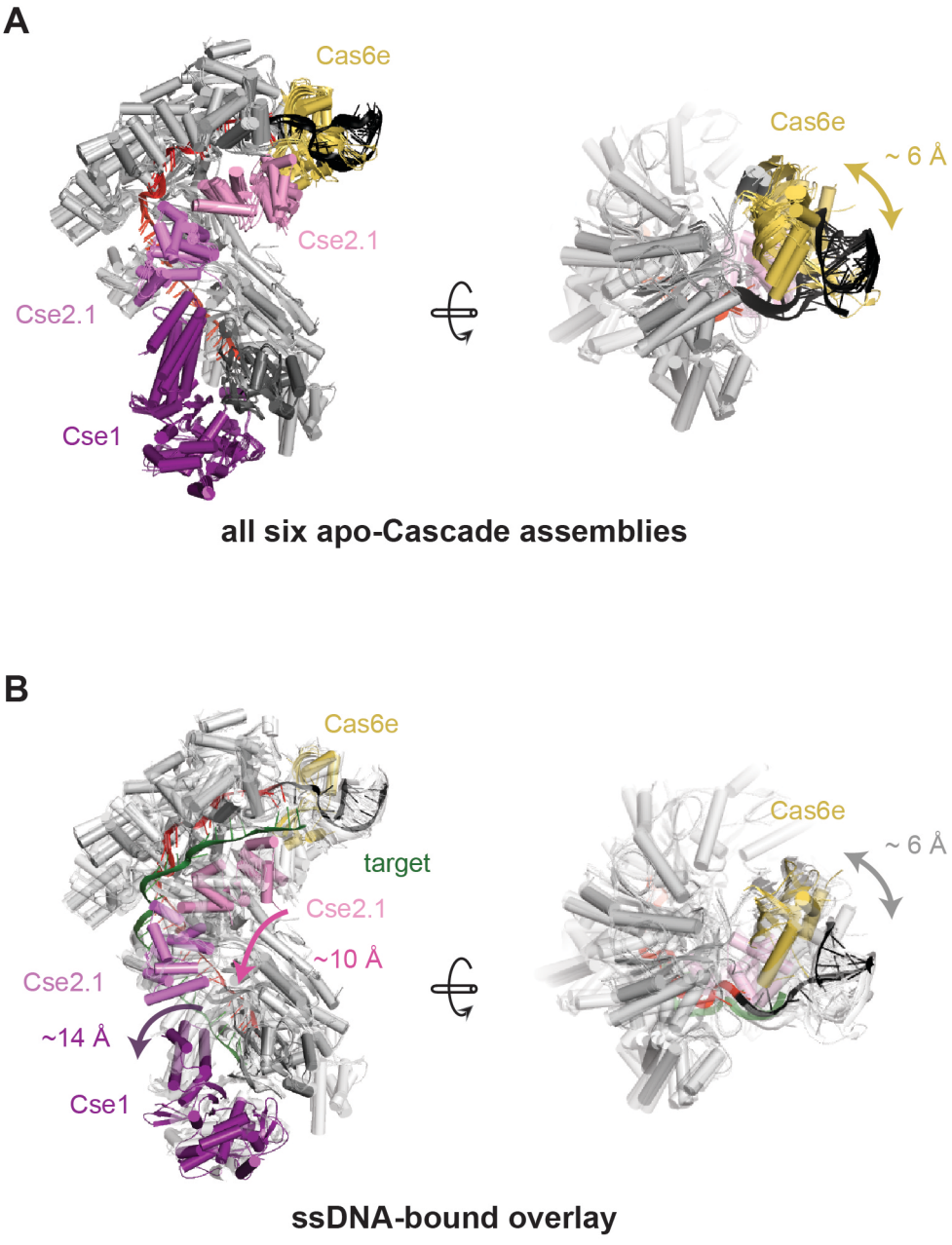


**Fig. S2. Superposition of Cascade structures. (A)** All six Cascade assemblies from the three available structures are aligned. Cas6e head adopts different conformations that vary by ~6 Å. **(B)** Overlay of the ssDNA-bound Cascade structure (colored) onto all the apo structures (transparent white). Cas6e of the ssDNA-bound structure adopts a conformation similar to that observed in apo Crystal assemblies, but target binding causes a conformational rearrangement in the Cse2 and Cse1 subunits, shown as colored arrows. The grey arrow indicates the variable position of Cas6e in the apo structures.

**Fig. S3.**

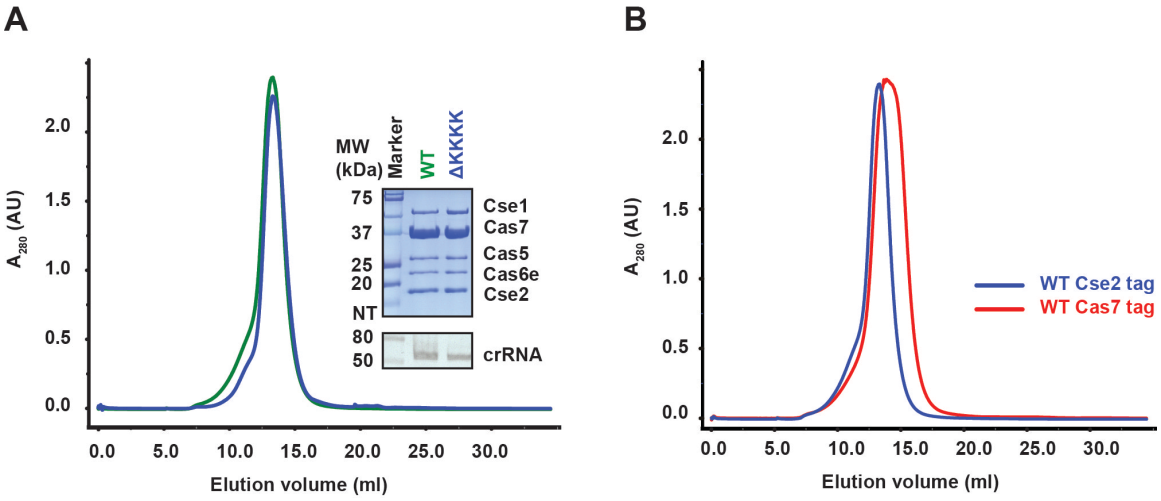


**Fig. S3. The ΔKKKK mutant expresses and purifies similar to the wild type Cascade complex. (A)** Elution profile of WT Cse2-tagged Cascade (green) and Cse2-tagged ΔKKKK (blue). The insert shows a Commassie blue-stained SDS-PAGE gel (top) and a denaturing polyacrylamide gel of phenol-extracted crRNA isolated for the Cascade complexes (bottom). ΔKKKK expresses and purifies as WT Cascade. **(B)** Elution profile of Cse2-tagged Cascade (blue) and Cas7-tagged Cascade (red). The broader peak of the Cas7-tagged Cascade consists of WT Cascade and CascadeΔCse2 complexes.

**Fig. S4.**

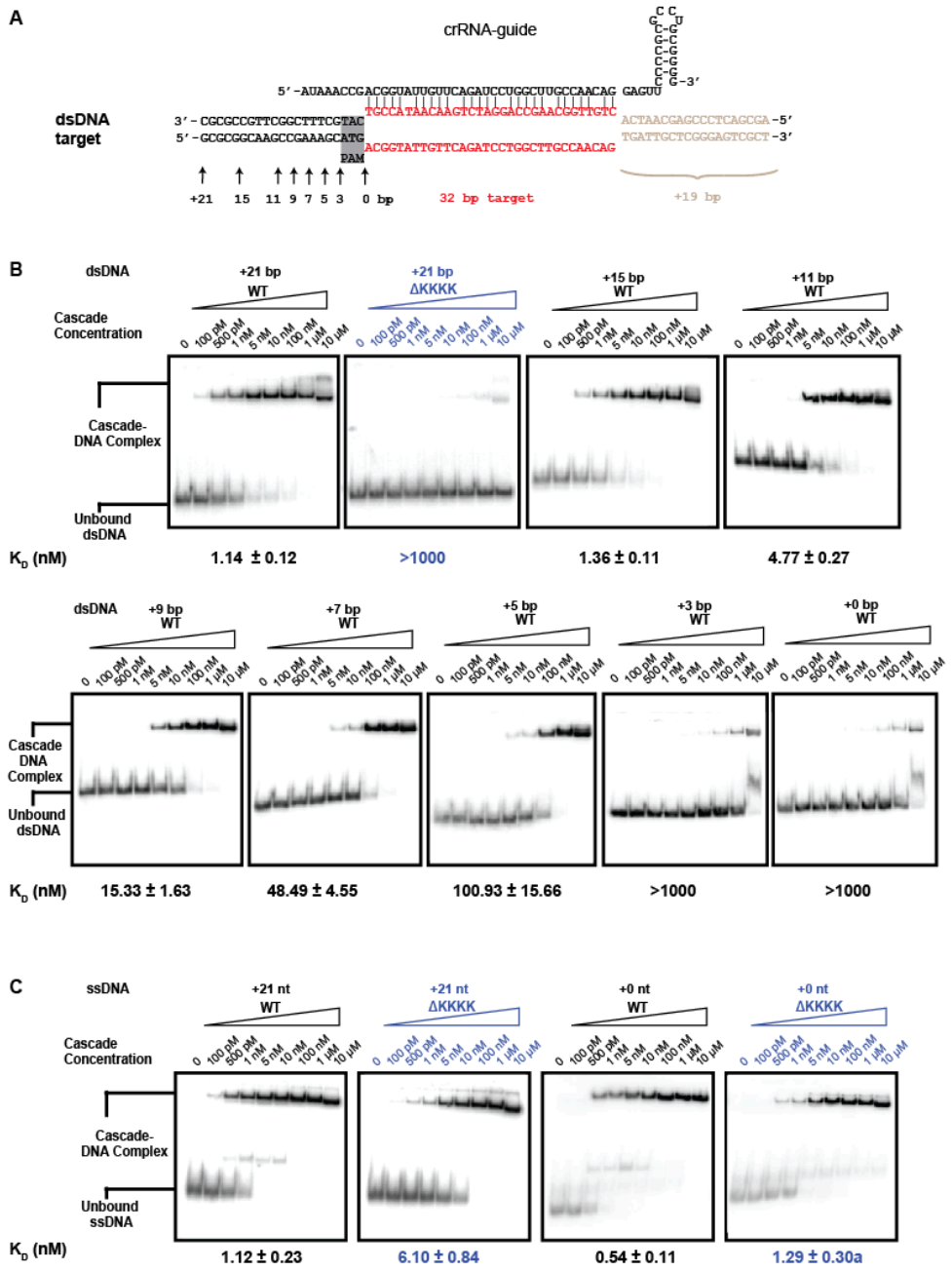


**Fig. S4. Cascade relies on a lysine-rich vise for binding dsDNA targets. (A)** Schematic of the crRNA and a 72-bp dsDNA target. The protospacer (red) is flanked by 19-bps (brown) on the 5'-end and 0 to 21-bps on the 3'-end. **(B)** Electrophoretic mobility shift assays (EMSA) performed using Cascade or ΔKKKK-Cascade, with dsDNA substrates that have different 3'-ends. Only dsDNA targets that reach the lysine-rich vise (K-rich vise) are bound with high-affinity. Deletion of the lysine vise (ΔKKKK) results in a binding defect. **(C)** EMSA performed with Cascade or ΔKKKK-Cascade, using either 32-nt or 72-nt ssDNA targets. Mutation of the K-rich vise (ΔKKKK-Cascade) has little effect on ssDNA binding. Imagines of gels used for Electrophoretic Mobility Shift Assays were saved as JPGs and the brightness and contrast have been altered to enhance visibility.

**Fig. S5.**

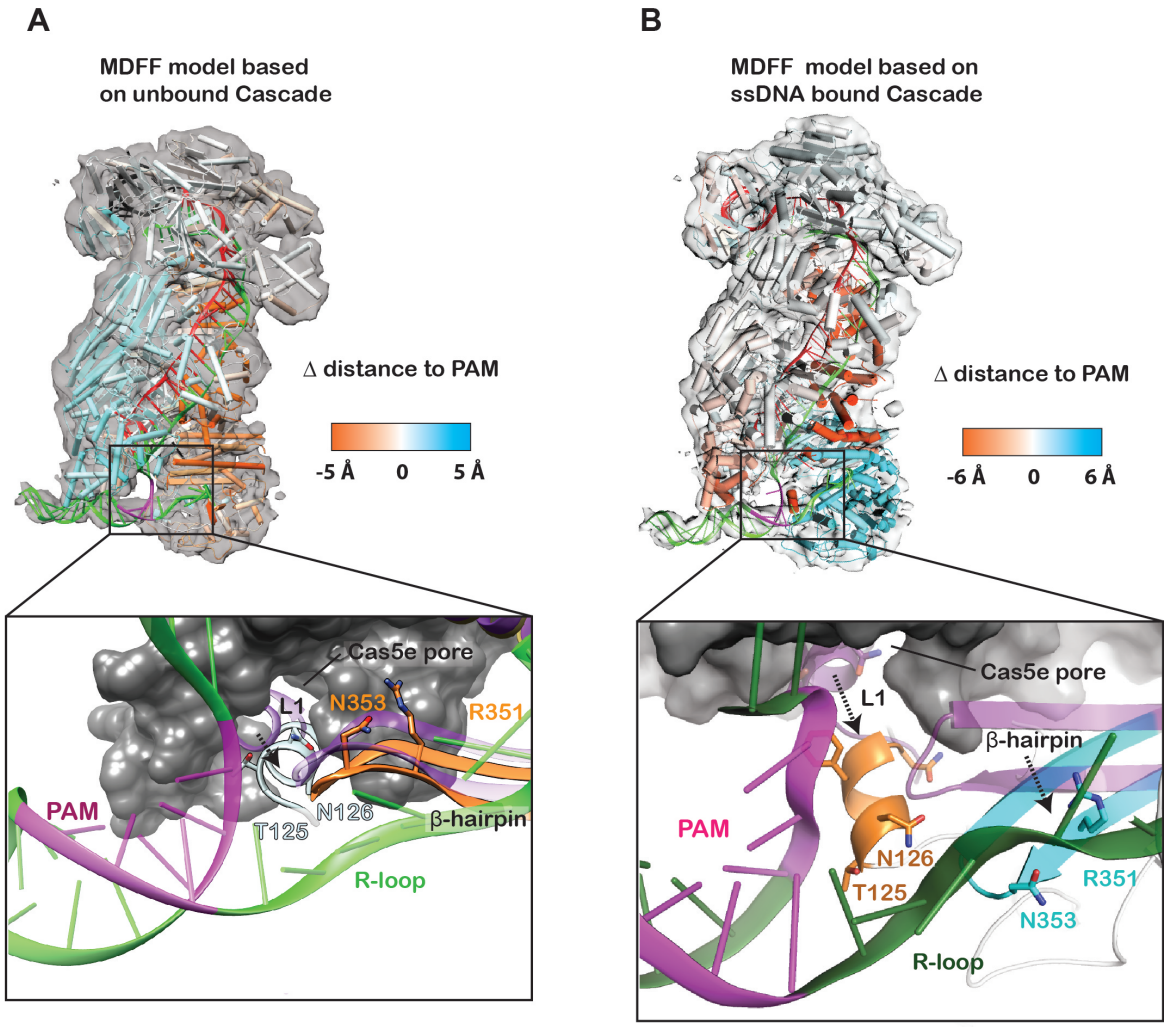


**Fig. S5. Structural models generated using Molecular Dynamics Flexible Fitting. (A)** Atomic coordinates of Cascade were modeled into the cryo-EM density of Cascade bound to a dsDNA target with a PAM (magenta) using Molecular Dynamics Flexible Fitting (MDFF). The model is colored according to changes in distance relative to the PAM. Atoms that move towards the PAM are colored orange and away from the PAM are blue. **(B)** Same as in panel A, only the atomic coordinates of the ssDNA bound structure of Cascade were used as the starting model. In this simulation the L1-helix moves out of the Cas5e pore, and is repositioned next to the PAM.

**Fig. S6.**

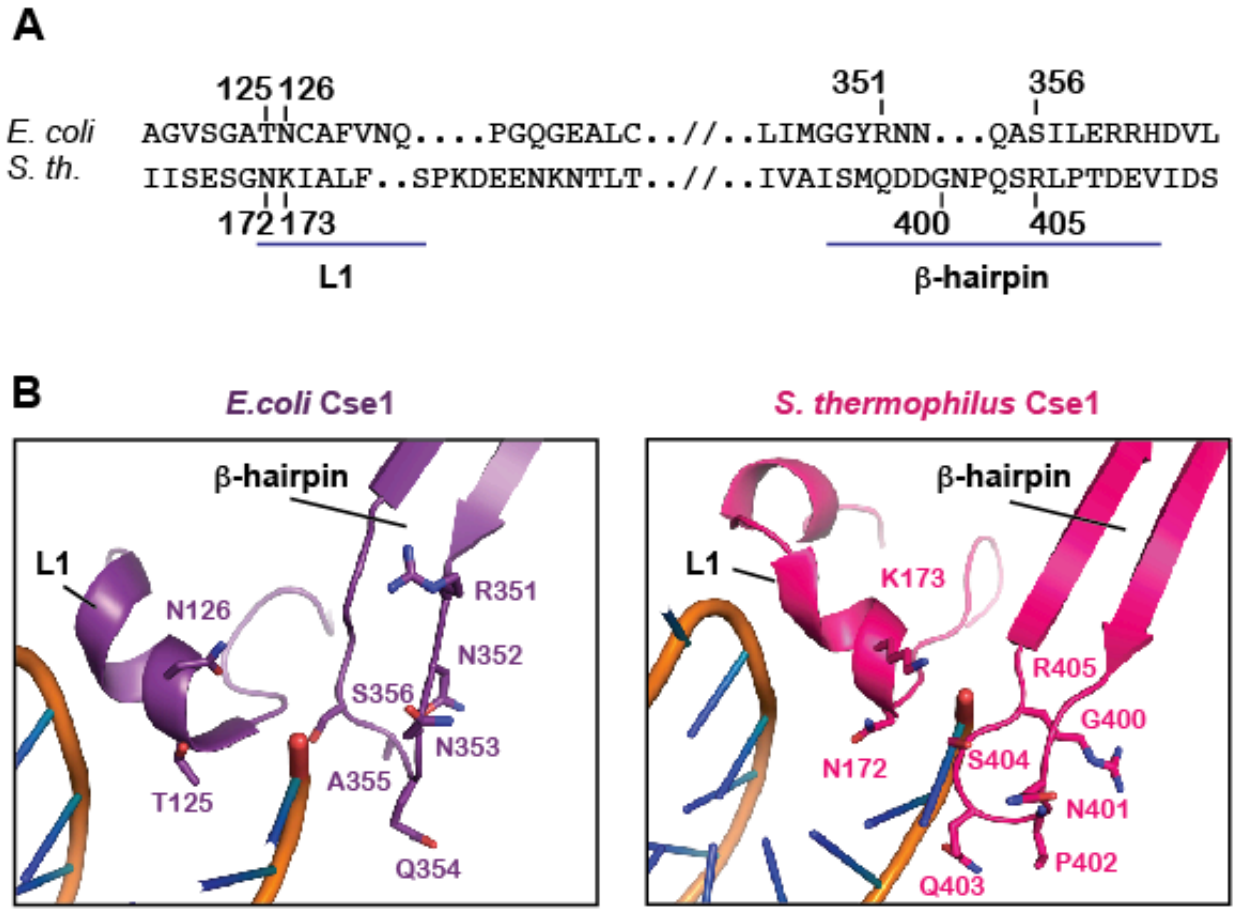


**Fig. S6. Sequence and structural alignment of PAM sensing features from *E. coli* and *S. thermophilus*. (A)** Amino acid sequence alignment of the L1 and β-hairpin features of Cse1 from *E. coli* and *S. thermophilus*. **(B)** Structural comparison of the L1 and β-hairpin features of Cse1 from *E. coli* (left) and a homology model of Cse1 from *S. thermophilus* (right).

**Fig. S7.**

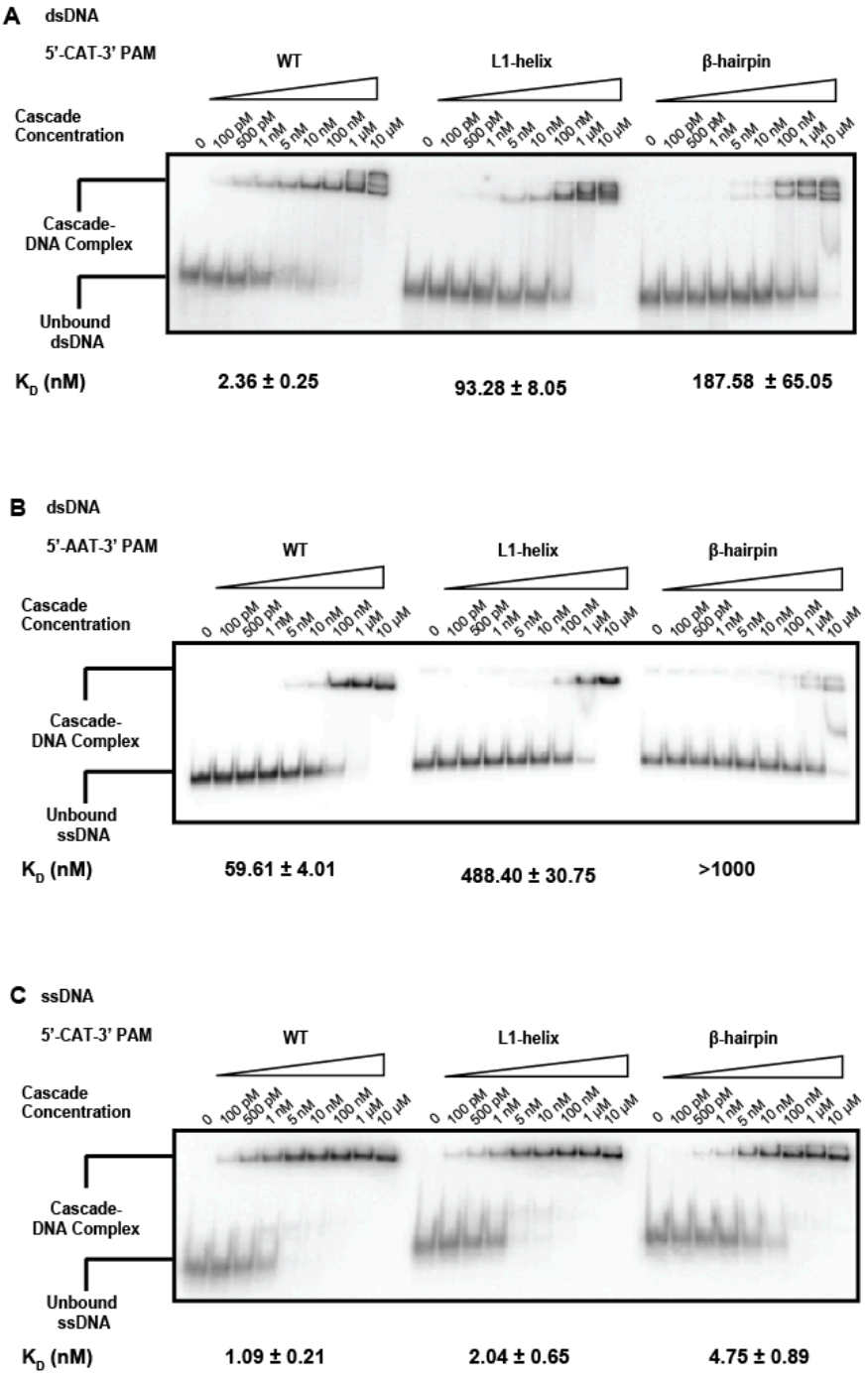


**Fig. S7. The L1-helix and β-hairpin are required for high-affinity binding to dsDNA targets.** **(A)** EMSAs of WT Cascade, the L1-helix mutant (Cse1 T125N/N126K), and the β-hairpin mutant (Cse1 R351G/N353P/A355S/S356R) performed using a 72-bp dsDNA target. **(B)** Mutation of the L1-helix or the β-hairpin has little effect on ssDNA binding. EMSAs of Cascade, the L1-helix mutant, and the β-hairpin mutant, were performed using 72-nt ssDNA target.

**Fig. S8.**

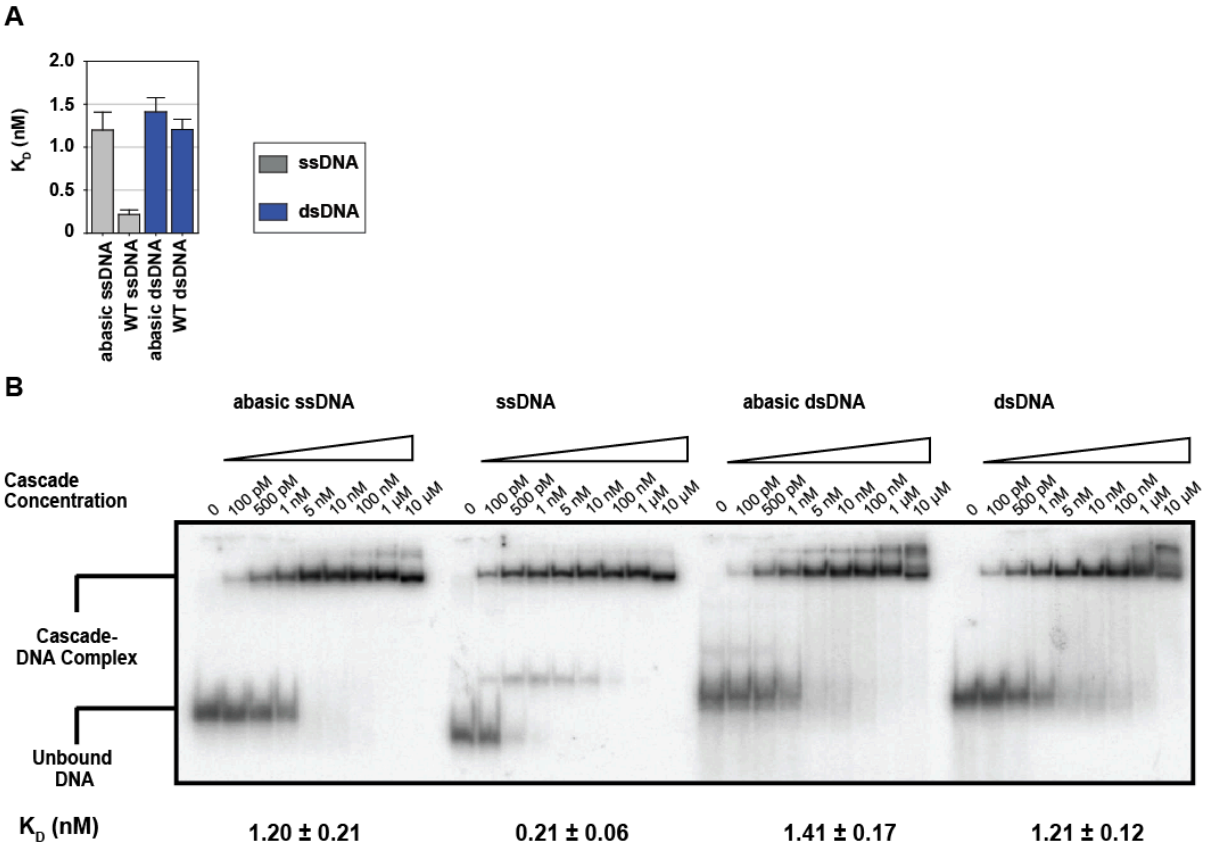


**Fig. S8. Conserved arginines stabilize flipped-out bases on the DNA target. (A)** Equilibrium dissociation constants of Cascade for 72-bp targets and 72-bp targets that are abasic at the positions corresponding to the flipped-out bases in the target strand. **(B)** Representative, EMSA performed using Cascade and ssDNA targets with or without abasic nucleotides a every 6$^{th}$ position of the protospacer. The absence of a nucleobase results in a modest but reproducible binding defect.

**Fig. S9.**

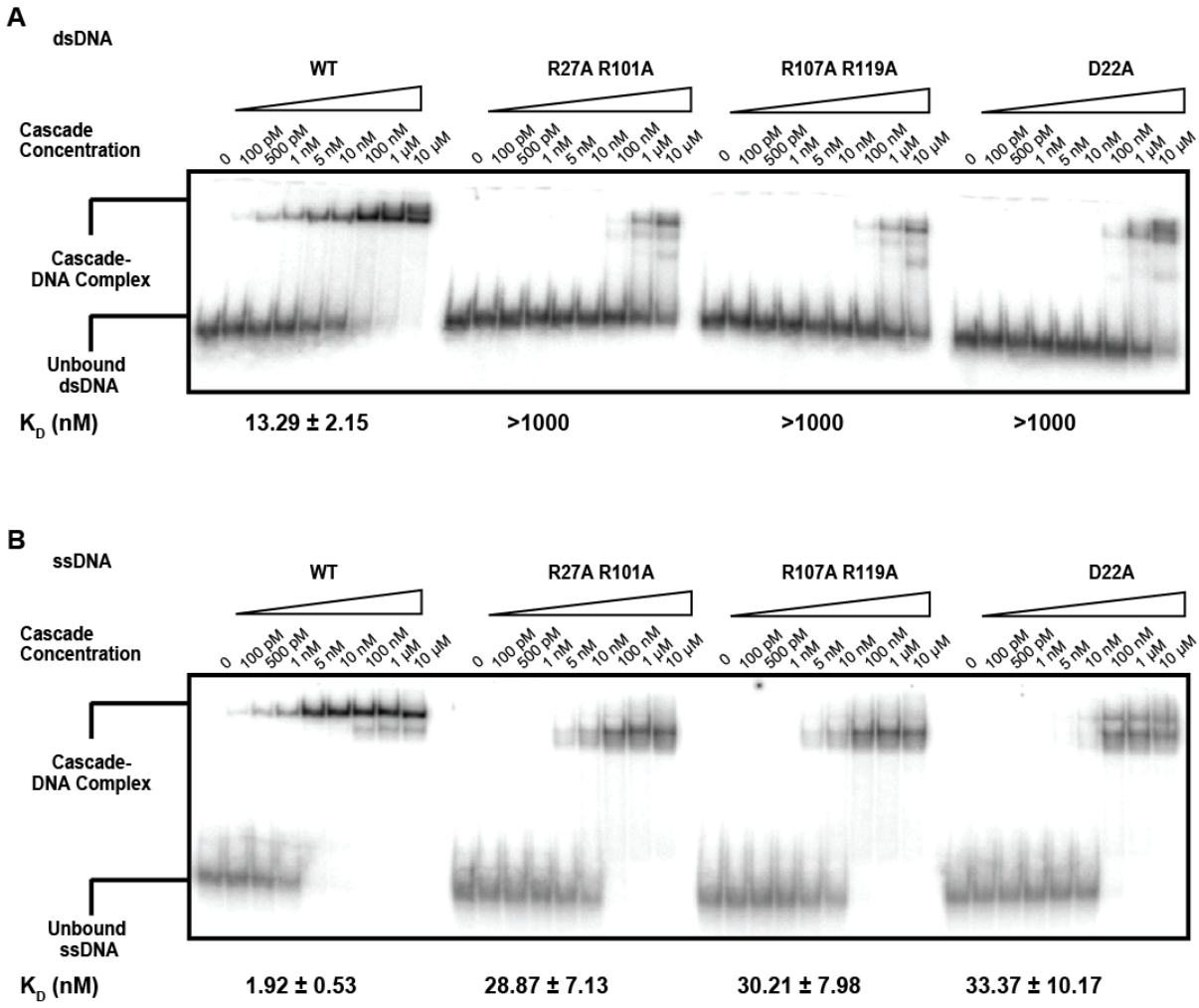


**Fig. S9. Mutations that perturb the arginine relay result in a dsDNA binding defect. (A)** EMSAs performed using Cascade or Cascade mutants (Cse2R27A/R101A or Cse2R107A/R119A) with 72-bp dsDNA target. Absence of the Cse2 subunits in the mutants leads to a severe dsDNA-binding defect. Cas7-tagged Cascade has a higher $K_D$ (13.3 nM) than Cse2-tagged Cascade because it contains a subpopulation of Cascades lacking Cse2 (Fig. S3). **(B)** EMSA of Cascade and arginine or aspartic acid mutants with 72-nt ssDNA target. These mutations result in a 15-fold binding defect. Visible is the subpopulation of CascadeΔCse2 binding ssDNA in the WT Cascade lanes.

**Fig. S10.**

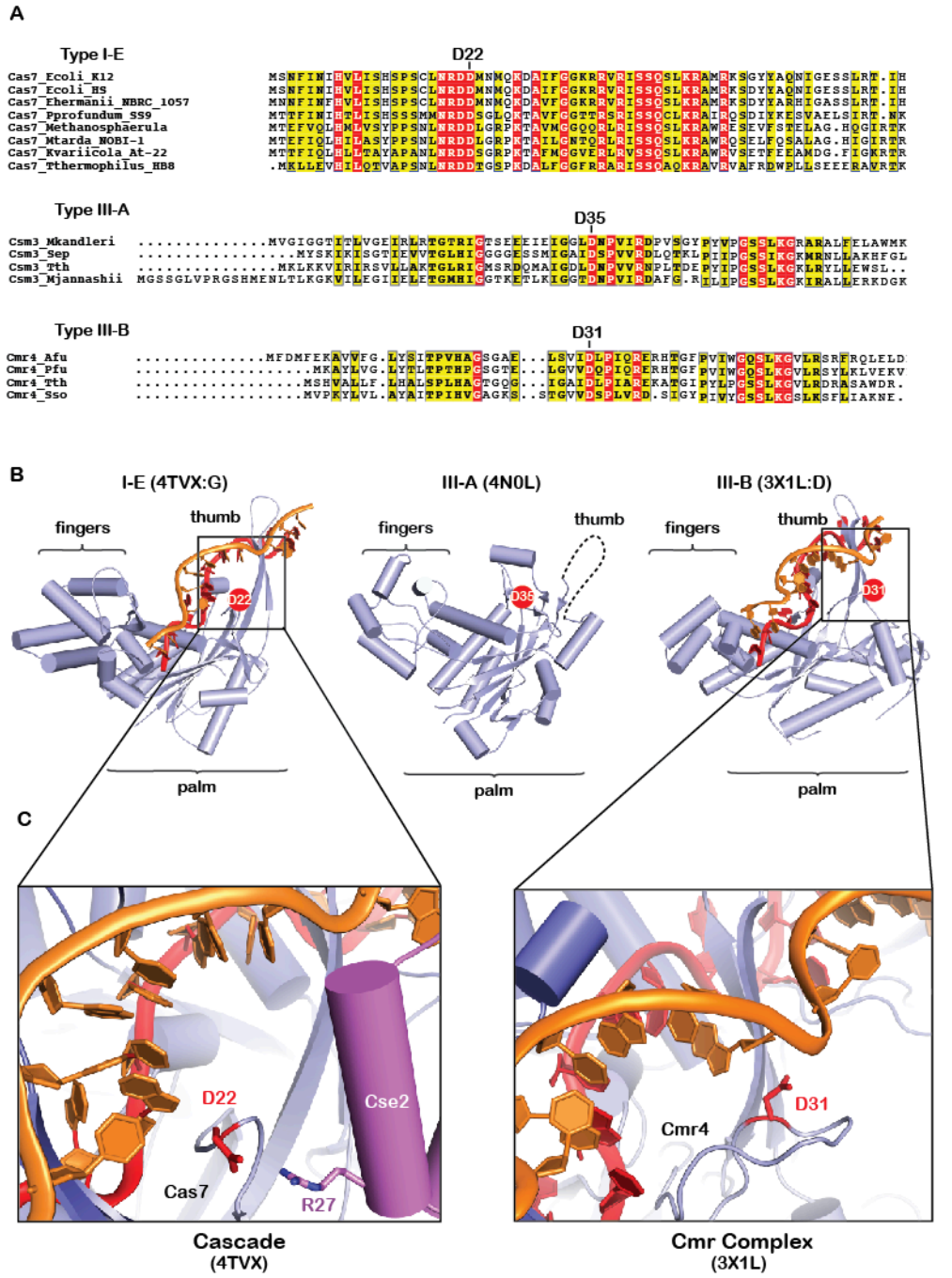


**Fig. S10. Aspartate residue conservation in Cas7 proteins from Type I and Type III CRISPR systems. (A)** Amino acid sequence alignments of the Cas7 subunit from Type I-E (Cas7), Type III-A (Csm3), and Type III-B (Cmr4). The conserved aspartate is indicated on each alignment. **(B)** Cas7 subunits are shaped like a right hand with palm, fingers, and thumb domains. The conserved aspartate is located between the palm and thumb domains. PDB codes are indicated. **(C)** Views of the conserved aspartate in context of the Cascade complex (left) and Cmr Complex (right). In Cascade, D22 forms a salt bridge with R27, whereas in the Cmr Complex D31 forms part of the RNase active site.

## Table S1. Plasmids used

| Plasmid | Description | source |
|---|---|---|
| pWUR397 | *Cas3* in pRSF-1b | Brouns et al. (1) |
| pWUR400 | *cse1-cse2-cas7-cas5-cas6e* in pCDF-1b | Brouns et al. (1) |
| pWUR407 | *cse1-cse2* in pRSF-1b | Brouns et al. (1) |
| pWUR408 | *cse1* in pRSF-1b | Brouns et al. (1) |
| pWUR515 | *cas7* with Strep-tag II (N-term)-*cas5-cas6e* in pET52b | Brouns et al. (2) |
| pWUR547 | *E. coli* R44 CRISPR, 7x spacer P7 in pACYCDuet-1 | Jore et al. (3) |
| pWUR610 | pUC-J3; pUC19 containing J3-protospacer on a 350 bp phage λ amplicon | Semenova et al. (4) |
| pWUR613 | pUC-P7-CGT; pUC19 containing R44-protospacer with CGT PAM on a 350 bp phage P7 amplicon | Jore et al. (3) |
| pWUR630 | *E. coli* CRISPR, 4x spacer J3 in pACYCDuet-1 | Semenova et al. (4) |
| pWUR656 | *cse2* with Strep-tag II (N-term)-*cas7-cas5-cas6e* in pCDF-1b | Westra et al. (5) |
| pUC-P7-CAT | pUC19 containing R44-protospacer with CAT PAM on a 350 bp phage P7 amplicon | This study |

**Table S2. Data collection and refinement statistics**

| Crystal morphology | Shard |
|---|---|
| Data collection | |
| Beamline | APS 23-ID |
| Space group | $P2_12_12_1$ |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 105.99, 244.80, 426.74 |
| α, β, γ (deg) | 90.0, 90.0, 90.0 |
| $CC_{1/2}$ | 0.984 (0.566) |
| Completeness (%) | 100 (100.0) |
| Redundancy | 7.3 (6.9) |
| $R_{merge}$(%) | 0.304 (1.369) |
| *I* / σI | 8.0 (2.1) |
| | |
| Refinement | |
| Resolution† (Å) | 49.74-3.20 (3.25-3.20)* |
| No. reflections | 1347007 |
| $R_{work}/R_{free}$ (%) | 25.0/27.6 |
| | |
| RMSD | |
| Bond lengths (Å) | 0.005 |
| Bond angles (deg) | 0.839 |
| Ramachandran | |
| Favored (%) | 94.6 |
| Outliers (%) | 0.58 |
| Clashscore | 4.3 |

*Values in parentheses are for highest-resolution shell.

†Resolution limits use the criterion of I/σI > 2.0

## Table S3. DNA substrates used for band shift assays

| Description | Sequence<br>target strand (top) and non-target strand (bottom); PAM in black and spacer in blue |
|---|---|
| 72 bp dsDNA target (+21) | 3′-CGCGCCGTTCGGCTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-GCGCGGCAAGCCGAAAGGATGACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 66 bp dsDNA target (+15) | 3′-GTTCGGCTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-CAAGCCGAAAGGATGACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 62 bp dsDNA target (+11) | 3′-GGCTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-CCGAAAGGATGACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 60 bp dsDNA target (+9) | 3′-CTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-GAAAGGATGACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 58 bp dsDNA target (+7) | 3′-TTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-AAGGATGACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 56 bp dsDNA target (+5) | 3′-CGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-GCATGACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 54 bp dsDNA target (+3) | 3′-TACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-ATGACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 51 bp dsDNA target (+0) | 3′-TGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-ACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 72 bp a-basic dsDNA target | 3′-CGCGCCGTTCGGCTTTCGTACTGCCA-AACAA-TCTAG-ACCGA-CGGTT-TCACTAACGAGCCCTCAGCGA-5′<br>5′-GCGCGGCAAGCCGAAAGGATGACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 72 bp dsDNA target 5′-AAT-3′ PAM | 3′-CGCGCCGTTCGGCTTTCGTAATGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′<br>5′-GCGCGGCAAGCCGAAAGGATTACGGTATTGTTCAGATCCTGGCTTGCCAACAGTGATTGCTCGGGAGTCGCT-3′ |
| 72 nt ssDNA target (+21) | 3′-CGCGCCGTTCGGCTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |
| 66 nt ssDNA target (+15) | 3′-GTTCGGCTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |
| 62 nt ssDNA target (+11) | 3′-GGCTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |
| 60 nt ssDNA target (+9) | 3′-CTTTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |
| 58 nt ssDNA target (+7) | 3′-TTCGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |
| 56 nt ssDNA target (+5) | 3′-CGTACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |
| 54 nt ssDNA target (+3) | 3′-TACTGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |
| 51 nt ssDNA target (+0) | 3′-TGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |
| 72 nt a-basic ssDNA target | 3′-CGCGCCGTTCGGCTTTCGTACTGCCA-AACAA-TCTAG-ACCGA-CGGTT-TCACTAACGAGCCCTCAGCGA-5′ |
| 72 nt ssDNA target 5′-AAT-3′ PAM | 3′-CGCGCCGTTCGGCTTTCGTAATGCCATAACAAGTCTAGGACCGAACGGTTGTCACTAACGAGCCCTCAGCGA-5′ |

## Table S4. Primers used for site directed mutagenesis

| Primer | Sequence (5`-3`) | description |
|---|---|---|
| PvE 80 | gctgttcttgctGAAGATATTGCCGCCATAC | *cas7* K137A K138A K141A K144A (fw) (ΔKKKK) |
| PvE 81 | gagcagagcagcATCATCCAGATTATCAGCC | *cas7* K137A K138A K141A K144A (rv) (ΔKKKK) |
| PvE 106 | cagagccgcATTCTTGAACGGCGTCATG | *cse1* R351G N353P A355S S356R (fw) (β-hairpin) |
| PvE 114 | tggatttccATATCCCCCCATAATCAATTC | *cse1* R351G N353P A355S S356R (rv) (β-hairpin) |
| PvE 132 | CatgACGGTATTGTTCAGATCCTGGCTTGCCAACAG | pUC-p7-CAT (fw) |
| PvE 135 | CTTTCGAGCTTGCCGATCAGCTTGG | pUC-P7-CAT (rv) |
| PvE 148 | GCAAATTAGAgctGTTTCAGAACCTGATGAATTAC | *cse2* R27A(fw) |
| PvE 149 | GCACATGATCCATTATCC | *cse2* R27A(rv) |
| PvE 150 | TAACGAGCGCgctATCTTTCAATTAATTCGG | *cse2* R101A(fw) |
| PvE 151 | ATTCTTCCACTATTGGCTAAAG | *cse2* R101A(rv) |
| PvE 152 | TCAATTAATTgctGCTGACAGAACAGCCGATATGGTC | *cse2* R107A(fw) |
| PvE 153 | AAGATACGGCGCTCGTTA | *cse2* R107A(rv) |
| PvE 154 | TAACGAGCGCgctATCTTTCAATTAATT | *cse2* R101A R107A(fw) |
| PvE 166 | AAGCGGCGCGaacaagTGTGCATTTGTCAATCAAC | *cse1* T125N N126K (fw) (L1) |
| PvE 167 | ACCCCAGCCAACAGTTTTTC | *cse1* T125N N126K (rv) (L1) |
| PvE 176 | CCAGTTACGTgctTTACTTACTCACGC | *cse2* R119A(fw) |
| PvE 177 | ACCATATCGGCTGTTCTG | *cse2* R119A(rv) |

**Movie S1. Mechanism of dsDNA binding by Cascade.** Cascade is a seahorse-shaped ribonucleoprotein complex consisting of 11 protein subunits and a crRNA guide molecule (red). X-ray crystal structures of Cascade determined before and after binding a complementary DNA target reveal a conformational change that is primarily restricted to the belly (Cse2) and tail (Cse1) subunits. To understand the mechanism of DNA binding, the unbound structure of Cascade was docked into the 9 Å cryo-EM density and a piece of B-form dsDNA (green) with PAM sequence (pink) was modeled into density near the tail. Lysine rich helices (blue) are present on the Cas7 backbone subunits. The lysine rich helices of Cas7.5 and Cas7.6 form a vise, which interacts with the phosphate backbone of the DNA target. Molecular dynamics flexible fitting (MDFF) was used to model atomic coordinates of the unbound Cascade structure into the cryo-EM density. The simulation shows that upon dsDNA binding the L1-helix slightly out of a pore formed by Cas5e. In contrast to the modest movement in L1, a large β-hairpin is repositioned in between the complementary and displaced strands of the DNA target. Residues that move towards the PAM are colored orange in the simulation. In addition to conformational changes in the tail, target binding triggers an “arginine relay” in which specific salt bridges between arginine residues on Cse2 (R27 and R101) and aspartic acid residues (D22) on the Cas7 subunits, are broken and compensatory salt bridges are formed (R107 R119) with acidic residues further down the Cas7 backbone. The arginines R27 and R101 interact with the flipped-out bases of the target DNA strand.

## Supplemental References


1. 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.
2. Brouns, unpublished
3. 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.
4. Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A*, **108**, 10098-10103.
5. Westra, E.R., van Erp, P.B.G., Künne, T., Wong, S.P., Staals, R.H.J., Seegers, C.L.C., 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.