Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2016 December 02.

Published in final edited form as:

Nature. 2016 February 25; 530(7591): 499–503. doi:10.1038/nature16995.

Structural basis for promiscuous PAM recognition in Type I-E Cascade from *E. coli*

Robert P. Hayes[1], **Yibei Xiao**[1], **Fran Ding**[1], **Paul B. G. van Erp**[4], **Kanagalaghatta Rajashankar**[2], **Scott Bailey**[3], **Blake Wiedenheft**[4], and **Ailong Ke**[1,*]

[1]Department of Molecular Biology and Genetics, Cornell University, 253 Biotechnology Building, Ithaca, NY 14853, USA

[2]Department of Chemistry and Chemical Biology, Cornell University, NE-CAT, Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA

[3]Department of Biochemistry and Molecular Biology, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD 21205, USA

[4]Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA

Abstract

Clusters of regularly interspaced short palindromic repeats (CRISPRs) and *cas* (CRISPR-associated) operon form an RNA-based adaptive immune system against foreign genetic elements in prokaryotes[1]. Type I account for 95% of CRISPR systems, and have been utilized to control gene expression and cell fate[2,3]. During CRISPR RNA (crRNA)-guided interference, Cascade (CRISPR-associated complex for antiviral defense) facilitates crRNA-guided invasion of double-stranded DNA (dsDNA) for complementary base-pairing with the target DNA strand, while displacing the non-target strand, forming an R-loop[4,5]. Cas3 nuclease/helicase is recruited subsequently to degrade two DNA strands[4,6,7]. Protospacer adjacent motif (PAM) flanking target DNA is crucial for self vs. foreign discrimination[4,8–16]. Here we present a 2.45 Å crystal structure of *E. coli* Cascade bound to a foreign dsDNA target. The 5′-ATG PAM is recognized in double-stranded form, from the minor groove side, by three structural features in Cse1. The promiscuity inherent to minor groove DNA recognition rationalizes the puzzling observation that a single Cascade can respond to several distinct PAM sequences. Optimal PAM recognition coincides with a wedge insertion, initiating the directional target DNA strand unwinding for segmented base-

Reprints and permissions information is available at XXXX.

*Correspondence to ailong.ke@cornell.edu; ak425@cornell.edu.

Supplementary Information

Three supplementary movies.

Author Contributions

R.P.H. and A.K. designed the research, R.P.H. determined the structure and performed biochemical analyses, Y.X., F.D., and P.E. contributed to biochemical analysis, K.R. assisted with diffraction data collection and processing, B.W. and S.B. contributed to assay setup. R.P.H., B.W., and A.K. wrote the manuscript.

The structure factors and coordinates for Cascade/partial R-loop structures with 10-nt and 35-nt non-target protospacer have been deposited in the Protein Data Bank under accession numbers 5H9F and 5H9E, respectively.

The authors declare no competing financial interests.

Takeda EX1127

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

pairing with crRNA. The non-target strand is guided along a parallel path 25 Å apart, and the R-loop structure is further stabilized by locking this strand behind Cse2 dimer. These observations provide the structure basis for understanding the PAM-dependent directional R-loop formation process[17,18].

Differentiating between 'self' and 'non-self' antigens is critical in CRISPR systems, as foreign target sequences (protospacers) are identical to sequences recorded in the host CRISPR locus (spacers). In Type I and II CRISPR systems, foreign DNA detection relies on protein-mediated PAM recognition[4,8–16]. Whereas PAM recognition in Cas9-based Type II systems has been elucidated based on major-groove DNA contact[19,20], it remains unclear as to whether Cascade recognizes PAM from DNA major or minor groove[12], in ss- or ds-form[4,21]. A particularly puzzling observation is the promiscuity in PAM recognition. Five PAM sequences (5′-ATG, AAG, AGG, GAG, and TAG reading from the non-target strand) are capable of triggering robust CRISPR interference via *E. coli* Cascade[4,11,21,22]. Crystal structures of *E. coli* Cascade in free- and ssDNA-bound forms revealed multiple conformational states, and provided valuable insights about the crRNA-guided ssDNA recognition mechanism[21,23,24], however, the mechanisms for dsDNA entry, PAM recognition, and R-loop formation remain poorly defined.

To understand the PAM-dependent foreign DNA recognition mechanism, we determined the 2.45 Å crystal structure of *E. coli* Cascade bound to dsDNA that forms a partial R-loop (Fig. 1a–c, Extended Data Fig. 1); such DNA substrates were efficiently bound by *T. fusca* Type I-E Cascade, and the resulting complex specifically recruited Cas3[25]. This structure agrees well with the cryo-EM reconstruction of the full R-loop/Cascade complex, underlining its validity in explaining the R-loop formation process[21] (Extended Data Fig. 2, Supplementary Video 1). Comparison with high-resolution crystal structures suggests the R-loop formation requires both the sliding of Cse1-CTD (C-terminal domain)-Cse2.1-Cse2.2, as seen in the ssDNA-bound structure, and the engagement of Cse1-NTD (N-terminal domain) as seen in the free-Cascade structure[21,23,24] (Extended Data Fig. 3, Supplementary Videos 2, 3). In addition, a localized conformational change takes place near the putative Cas3 binding site[21], which is only observed in this R-loop bound structure, therefore may play a role in Cas3 recruitment (Extended Data Fig. 4). dsDNA enters Cascade between Cas7.5 and Cas7.6, contacted by the lysine-rich helices[16,26] (Fig. 1d–e, Extended Data Fig. 5). DNA bifurcates underneath PAM. The entire target DNA strand flips to form the segmented DNA-crRNA duplex[27]. The 10-nt non-target strand is guided to a parallel path 25 Å apart by sequence nonspecific contacts, an active mechanism to stabilize R-loop (Fig. 1d–e). Modeling dsDNA beyond PAM projects it across Cse1-CTD without severe steric clashes (Fig. 1f), illustrating a possible PAM-searching scenario[21].

The 5′-ATG PAM sequence is recognized in the double-stranded form, from the minor groove side, by the Cse1 subunit of the Cascade (Fig. 2a–b). This rather surprising mode of recognition strongly biases towards the target DNA strand, which rationalizes the previous observations that mismatched PAMs could be tolerated, provided the target strand sequence was optimal[4,21]. Three structural features in Cse1 are involved in PAM recognition: the glutamine-wedge, the glycine-loop and a lysine-finger (Fig. 2a–b). Only C_{T-1}–G_{NT-1} is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

tolerated at the −1 PAM position (PAM-1) in *E. coli*[5,21], although recent analyses revealed spacer-dependent tolerance of alternative base-pairs at PAM-1[28]. In our structure (Fig. 2a–d), the amide of A355 in the glutamine-wedge donates a H-bond to O2 of C_{T-1}, specifying a pyrimidine in the target strand. The carbonyl of G157 in the glycine-loop accepts a water-mediated H-bond from N2 of G_{NT-1}. G_{NT-1}-to-inosine substitution assayed by Electrophoretic Mobility Shift Assay (EMSA) suggests this contact only provides minor discrimination against A_{NT-1}(Extended Data Fig. 6a). Cascade's affinity for different PAM-1 base-pair combinations corroborate well with the structure observation that a target strand pyrimidine is strongly specified (Extended Data Fig. 6b). The 10-fold differences in Kd, however, do not support an absolute C_{T-1}−G_{NT-1} specification at PAM-1. Indeed, Cas3 was able to cleave alternative PAM-1 targets, provided Cascade concentration is above the corresponding Kd (Extended Data Fig. 6c). These results echo the recent finding in suggesting that PAM-1 readout is further complicated by Cascade/Cas3 expression level and spacer content[28].

Importantly, PAM recognition coincides with the glutamine-wedge insertion into dsDNA underneath PAM (Fig. 2a, d). The tip residue Q354 stacks underneath C_{T-1} and, together with N353, sterically displaces the first two protospacer nucleotides in the target strand, forcing them to rotate outwards. Given its strategic location, it is rather surprising that this wedge is not highly conserved in sequence (Extended Data Fig. 7). Indeed, tip residue substitutions (Q354A, N353A, and Q354A/N353A) had little consequence in DNA-binding. In contrast, trimming this wedge (NNQAS352-356/GG) led to 100-fold DNA binding defect, suggesting the wedge functions through a steric interference mechanism, to nucleate the target strand displacement upon PAM recognition (Fig. 2e–f). A serine-to-phosphate 'lock' is essential in initiating the target strand flipping in Cas9[19]. T125 is in a similar location in our structure but contacts the +1 bridging oxygen instead, and T125A substitution had negligible defect (Fig. 2e).

Recognition at PAM-2 is promiscuous by *E. coli* Cascade. Only G_{T-2}-C_{NT-2} is rejected at this position; the other three combinations lead to efficient interference[22]. Here the glycine-loop residues (159–161) assume a lip-like structure, introduces DNA bending at A_{T-2}-T_{NT-2}, and 'bites' onto PAM-1 base-pair in conjunction with the glutamine-wedge underneath; T_{NT-2} retreats backwards and tilts upwards (Fig. 2b, d). The rim of the glycine-loop explores shape-complementarity to A_{T-2}, and donates a weak H-bond to N3 of A_{T-2}. G160A substitution disrupts the shape complementarity, and reduced Cascade-binding affinity by ~100-fold and Cas3-cleavage to baseline level (Fig. 2e–f). Rejection of G_{T-2}-C_{NT-2} at PAM-2 is rationalized by the fact that N2 amine of G_{T-2} would introduce steric clashes against the glycine loop; whereas T_{T-2}-A_{NT-2} or C_{T-2}-G_{NT-2} would not, based on modeling (Extended Data Fig. 8a). Indeed, removal of this amine in inosine substitution largely rescued the DNA-binding defect, confirming that N2 of G_{T-2} is the anti-determinant for PAM-2 specificity (Extended Data Fig. 8b). An equivalent glycine-rich loop is present in all known Cse1 structures; they likely play a similar minor groove DNA recognition function (Extended Data Fig. 7).

PAM-3 is typically specified as a $pyrimidine_T$-$purine_{NT}$ pair by *E. coli* Cascade[4,11,21]. 5′-TAG PAM also leads to interference, but G_{T-3}-C_{NT-3} containing PAMs fail to[22,28]. Here a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

favorable electrostatic interaction from a lysine-finger (K268) to O2 of T_{T-3} is observed, which rationalizes the strong preference for pyrimidine at this position (Fig. 2b, d). K268A mutation reduced Cascade-binding and Cas3-cleavage by >8-fold and >10-fold, respectively, emphasizing its positive contribution to PAM-3 recognition (Fig. 2e–f). Interestingly, K268A mutant still retained wild-type level discrimination against 5′-CTG-PAM (Extended Data Fig. 8c). G_{T-3}–to-inosine substitution ultimately proved that N2 of G_{T-3} also serves as a strong anti-determinant in the rejection of G_{T-3}–containing PAMs (Extended Data Fig. 8d). Interestingly, K268 makes an electrostatic interaction to C_{T-4} (Fig. 2b, d), implying certain level of sequence discrimination at PAM-4 as well.

Detailed structure dissection also helps rationalizing the self-avoidance mechanism. All spacers in *E. coli* CRISPR loci are 'protected' by a 5′-CCG-PAM. This PAM is the combination of the least-preferred nucleotides at each position (3′-$G_{T-3}G_{T-2}C_{T-1}$), which would strongly disfavor Cascade-mediated R-loop formation, despite a perfect spacer match.

The non-target strand is guided sequence nonspecifically along the Cascade surface, 20–25 Å away from the target strand (Fig. 3a). The sugar-phosphates of nucleotides 1–3 are contacted by K163, G169, and K296 of Cse1-NTD, nucleotides 6–9 by a string of positive charges(R393/K394/K488/H489/R491) across Cse1-CTD (Fig. 3a). The redundant interactions were not disrupted by a point mutation (Y397A). A double mutant (488–489 KH/AA) neutralizing a positive patch, however, led to a 4-fold binding defect (Fig. 3b). The 10th/last nucleotide rests at an intersection between Cse1-CTD and Cse2.1. To investigate whether the following non-target residues travels on the surface or backside of the Cse2 dimer, we further determined a 3.2 Å structure where the non-target protospacer is 22-nt longer. Although most of the additional residues remain unresolved, density clearly reveals that the non-target strand residues take the downward trench route towards the backside of the Cse2.1 dimer, attracted by the favorable electrostatic environment therein (Fig. 3c–e). The non-target strand sequestration likely corresponds to the extra "locking" step after most R-loop forms, as a mechanism to prevent the R-loop collapse[17].

In summary, our structural analysis provides important insights about the PAM-dependent directional R-loop formation process[17] (Fig. 4). Recognition of an interference PAM by Cascade coincides with the wedge-mediated displacement of the first two target strand nucleotides, initiating DNA unwinding. The directional DNA melting ensures the ordered guidance of the non-target DNA strand ~25Å away from the target strand, as a mechanism to stabilize the seed bubble. Further R-loop propagation leads to non-target strand sequestration behind Cse2 dimer, locking the R-loop in place. Conformational changes accompany the process and reorganize the Cse1 surface, paving the way for Cas3 binding. The active guidance of the non-target DNA strand is a theme not observed in Cas9-DNA structures[19,20]. It rationalizes the observation that the Cascade-bound R-loop is significantly more stable than that by Cas9[17]. Besides five interference PAMs, Twenty-one other PAMs stimulate the "primed adaptation" in *E. coli*[22], where Cascade and Cas3 actively recruit the Cas1/Cas2 spacer acquisition complex[29,30]. Priming PAMs may lead to suboptimal Cascade contacts and non-canonical R-loops. Such R-loops are difficult to form and may not be completely unwound[18], requiring higher Cascade concentration[30] and favorable DNA torque[17]. They also fail to directly recruit Cas3[30], which may indicate that the non-target DNA strand is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

misguided, as Cas3 recruitment is contingent upon non-target strand contact as well[25]. Overall, our work provides a framework for future studies to better understand the interference and primed adaptation mechanisms in Type I CRISPR-Cas systems.

METHODS

Expression and purification of Cascade, Cascade-dsDNA complex and Cas3

Sequence information for primers and the synthetic CRISPR expression cassette can be found in Extended Data Table 1. Cascade expression was similar to[26]. Briefly, *cse1* was PCR amplified from *E. coli* K12 genomic DNA and cloned into pRSF-Duet-ORF1 vector (Kan^R), between NcoI and NotI restriction sites (Extended Data Table 1). The *cse2-cas7-cas5e-cas6e* sub-operon was cloned into pET52b (Amp^R) between NcoI and NotI; as a Precission cleavable His_6 fusion at the N-terminal of *cse2*. The pre-crRNA expression cassette was synthesized by Life Technologies and cloned into the pHSG-398 vector (Cam^R) (Extended Data Table 1). *E. col* BL21 (DE3) star cells containing the 3 plasmids were grown in LB medium at 37 °C to O.D. $_{600}$ of 0.6. Cascade expression was induced by the addition of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and cells were further cultured at 20 °C for 12 hr.

The cells ware disrupted by sonication in buffer A (50mM Tris pH 8.0, 20mM imidazole, 300 mM NaCl), loaded to Ni-NTA column, and eluted with buffer A supplemented with 300 mM imidazole. The His_6 tag was cleaved by Precision protease, and back-adsorbed with a second Ni-NTA column binding step. Cascade was concentrated and buffer exchanged into buffer B (20 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT), and further purified on Superdex 200 prep grade column (GE healthcare). Free-Cascade containing fractions were pooled, concentrated to 15 mg/mL, flash-frozen, and stored at −80 °C.

Cascade-dsDNA complex was prepared by mixing free-Cascade with dsDNA R-loop mimicking substrate. DNA substrates were chemically synthesized from IDT (Extended Data Table 1). The non-target strand was annealed with the target strand at a 1.5:1 molar ratio. The resulting R-loop mimicking substrate was mixed with Cascade at a 2:1 molar ratio, incubated at room temperature for 30 minutes, and re-purified on Superdex 200. Cascade-dsDNA complex fractions were pooled and utilized in crystallization trials. The Cascade-dsDNA complex containing the 32-nt non-target strand overhang was also obtained using the above protocol, except the His tag was not cleaved. Cascade mutants were constructed with site-directed mutagenesis, and purified using the same method as free-Cascade, except that the N-terminal His_6 tag was left intact. Cascade integrity was checked using SDS-PAGE (Extended Data Fig. 4b).

The Cas3 gene was amplified from *E. coli* K12 genomic DNA and cloned between BamHI and XhoI into the pET28a-SUMO plasmid. *E. coli* BL21 (DE3) star cells were grown in LB medium at 20 °C to O.D.$_{600}$ of 0.3, induced with 0.2 mM IPTG, and further cultured at 20 °C for 12 hr. The Ni-NTA and SEC purification procedures were similar to the procedure mentioned above. The monomeric SUMO-Cas3 fractions were pooled, concentrated to 2 mg/mL, flash-frozen and stored at -80 °C until usage in biochemical assays.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Crystallization and structure determination of Cascade/partial R-loop complex

Cascade complex crystals were grown using the hanging drop vapor-diffusion method by mixing 2 μL of purified Cascade-ds-DNA complex (15 mg/mL) with 2 μL of mother liquor (1.6 M Na/K Phosphate pH 6.2) at 18 °C. Initial crystals appeared after 2 weeks and grew to full size after ~6 weeks. Crystals were cryoprotected in motherliquor supplemented with 20% ethylene glycol and flash frozen in liquid nitrogen. Diffraction data were collected at Advanced Photon Source – NECAT beamline 24-ID-C and were processed with HKL2000[31]. The structure was solved by molecular replacement with PDB: 4QYZ as the search model. Iterative model building and refinement was conducted with COOT[32] and PHENIX[33]. A summary of the diffraction and refinement statistics can be found in Extended Data Table 2. The Ramachandran plot for the Cascade-dsDNA 10-nt overhang structure indicated 96.65% of residues in the favored region, 3.10% allowed, and 0.25% outliers. The Ramachandran plot for the Cascade ds-DNA 32-nt overhang structure indicated 94.56% of residues in the favored region, 4.81% allowed, and 0.63% outliers. Figures were generated using Pymol[34] and CCP4mg[35].

Electrophoretic mobility shift assay

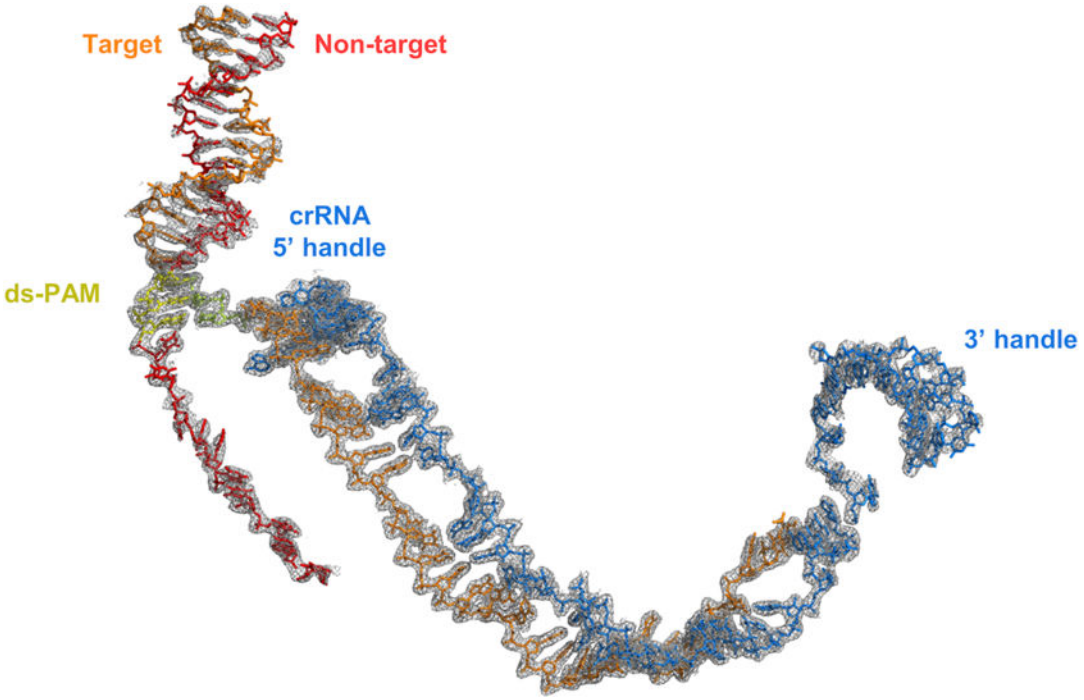
Fluorescent dsDNA target substrates were generated for biochemical assays. The crRNA-matching targets with varied PAMs were cloned into pCDF-Duet between the Pstl and Ncol sites Extended Data Table 1. The dsDNA1 and dsDNA2 substrates were PCR amplified from the plasmid using the indicated fluorescent oligonucleotides (5′ 6-FAM for the non-target strand and 5′ Cy5 for the target strand). The dsDNA3 substrates were prepared by oligonucleotide annealing. All dsDNA substrates were subsequently gel-purified. The dsDNA1 substrate (5′ATG-PAM) was used for all main-text EMSA and Cas3 cleavage assays. The dsDNA2 substrates were used for the experiments shown in Extended Data Figures 6b, 6c, and 8c. The dsDNA3 substrates were used in the experiments shown in Extended Data Figures 6a, 8b, 8d. DNA binding was conducted in 20 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol. The dsDNA substrate concentration was held constant at 3 nM and Cascade concentration was titrated as indicated. The Cascade and dsDNA were incubated at 37 °C for 30 minutes in 20 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol. EMSA was carried out at 4 °C on 2% agarose gels. Fluorescent signals were scanned using a Typhoon 9200 scanner.

Cascade-mediated Cas3 DNA cleavage assay

Cascade-R-loops were pre-formed by mixing 40 nM Cascade or Cascade mutants with 6 nM fluorescent dsDNA target in binding buffer (5 mM HEPES pH 7.5, 60 mM KCl) at 37 °C for 30 minutes. Cascade-R-loops were then mixed with 500 nM SUMO-Cas3 in DNA cleavage RXN buffer (5 mM HEPES pH 7.5, 60 mM KCl, 10 mM $MgCl_2$, 10 μM $CoCl_2$). Either 2 mM ATP or 2 mM AMPPNP was added and the reaction was incubated at 37 °C for 30 minutes. Cy5 and 6-FAM fluorescent signals were recorded by Typhoon 9200 scanner. The wild-type *E. coli* Cascade specifically nicked the non-target DNA strand ~10–12 nt into the R-loop region in the presence of a non-hydrolyzable ATP analog, AMPPNP. Addition of ATP triggered processive degradation of the non-target DNA strand and distributive

degradation of the target strand upstream of the R-loop. These results are consistent with previous studies[11,21].

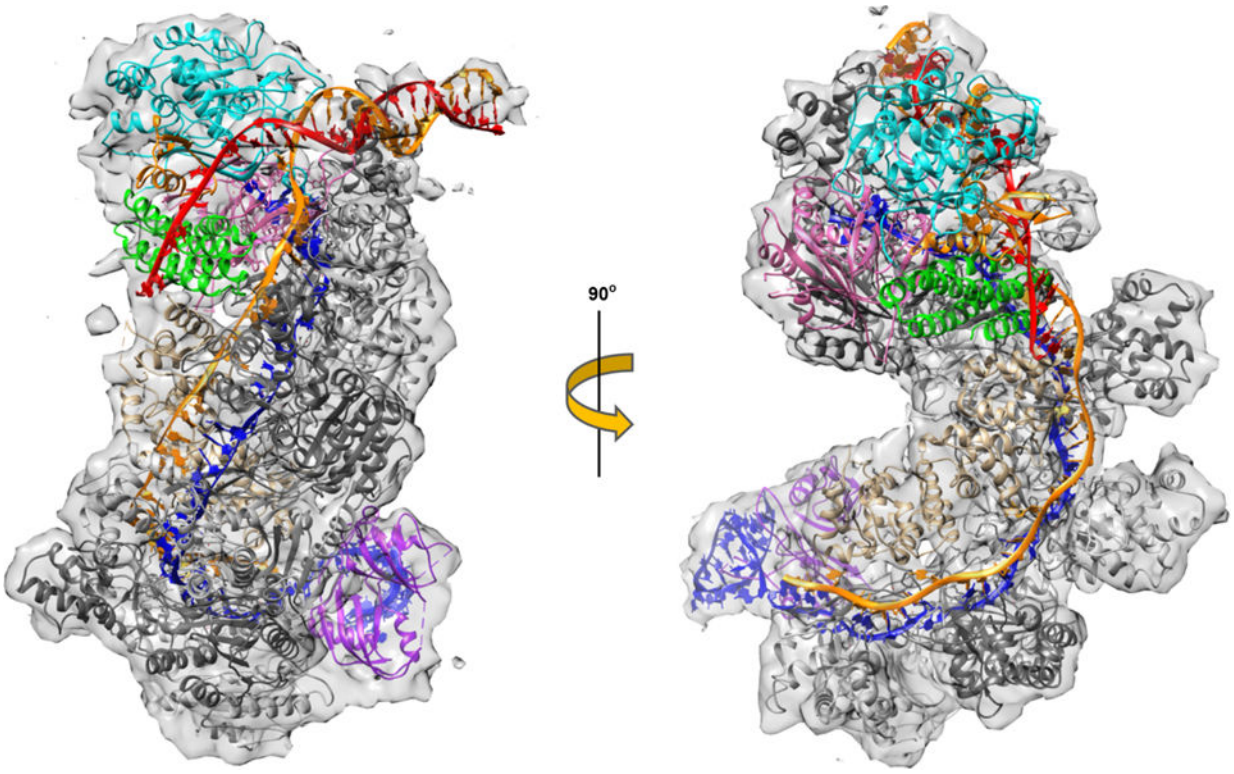
Extended Data



Extended Data Figure 1. Electron density of nucleic acids in Cascade-dsDNA complex structure
$2F_O$-F_C electron density map (1.0 σ). Nucleic acid strands are shown as sticks and colored as previously indicated.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

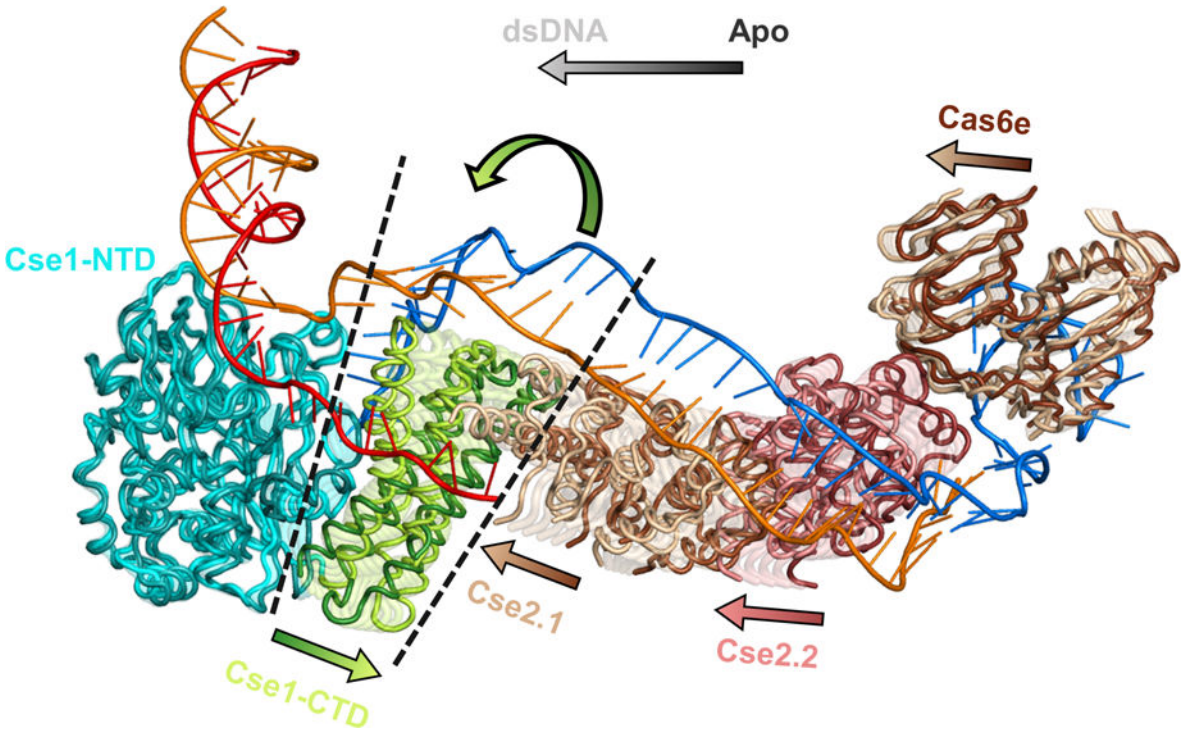
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 2. Comparison between the Cascade-partial R-loop crystal structure and the Cascade-full R-loop EM reconstruction

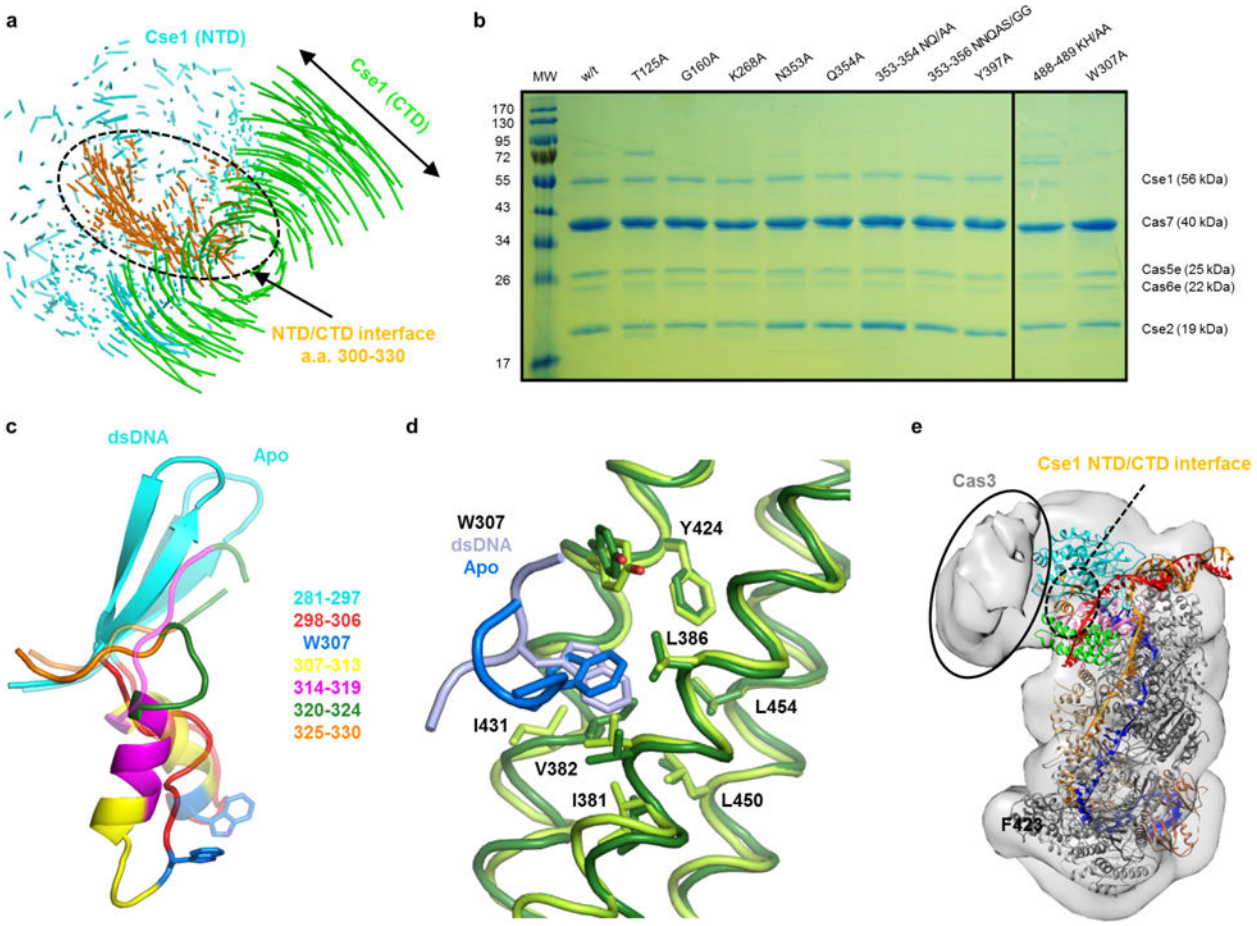
Rigid body docking of the partial R-loop Cascade crystal structure into the EM reconstruction (EMD-5929) of the full R-loop Cascade illustrates a similar overall conformation, with a correlation value of 0.83.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 3. Conformational changes in Cascade upon partial R-loop formation
Comparison of free- (PDB: 4TVX; in a darker shade) and partial R-loop bound Cascade (this study, in lighter shade). Arrows indicate the direction of the movement. The C-terminal domain (CTD) of Cse1 pivots 30° about a hinge at the Cse1 NTD-CTD interface. The amplified motion at the tip of the Cse1-CTD slides with the Cse2 dimer ~12 Å relative to the Cas7 scaffold, and protrudes upwards into the R-loop. Cse1 NTD remains in the docked position, stabilized by the clamping of Cse1 L1 loop onto the exposed tri-nucleotide motif on crRNA 5′-handle.

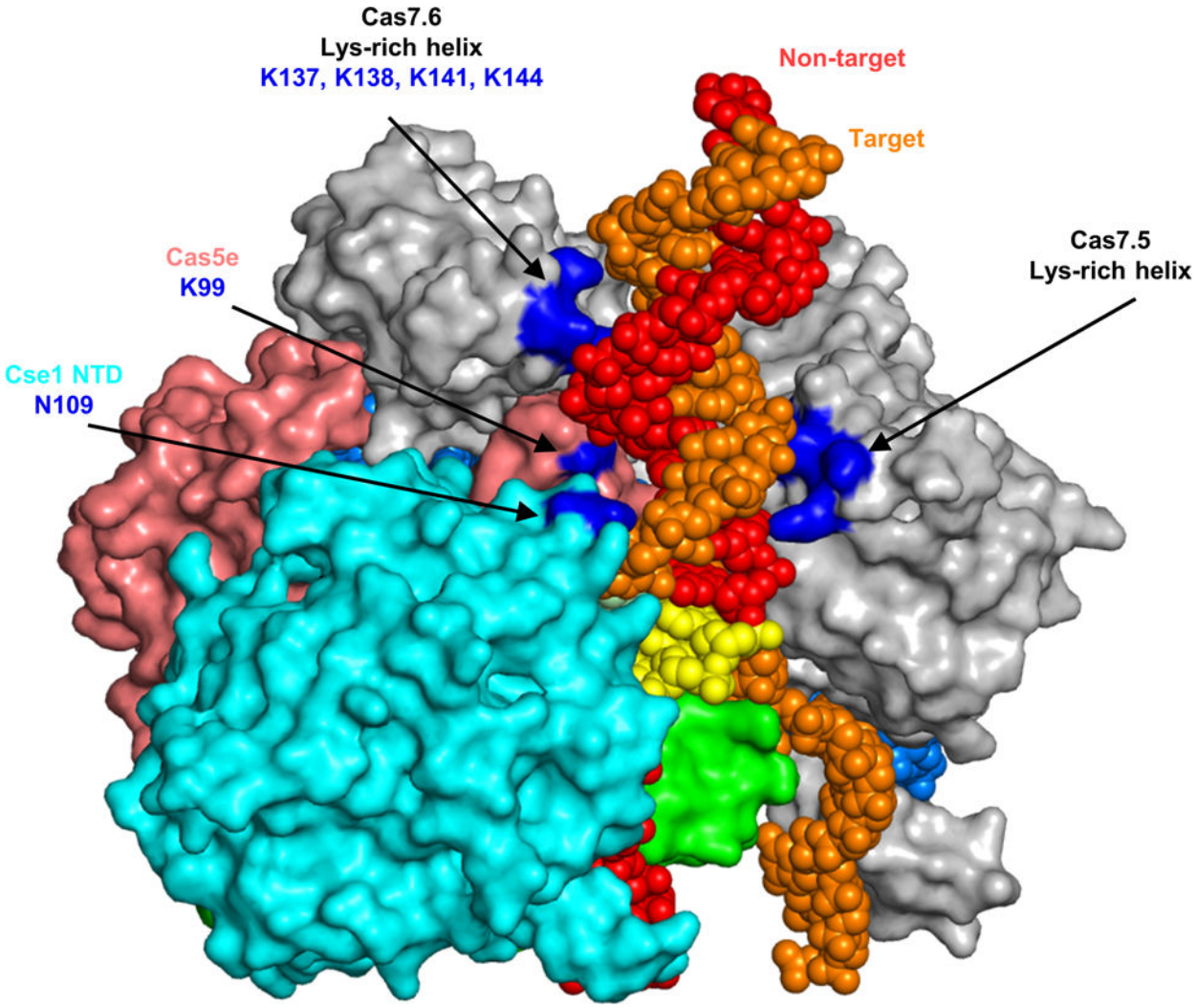
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 4. Local conformational rearrangement in Cse1 at the proposed Cas3 binding site

a, Vector map showing global and local conformational changes in Cse1. Cse1-NTD (in cyan) undergoes moderate movement. Cse1-CTD (in green) swings about a pivoting point (W307) as part of the global conformational changes, The Cse1 NTD/CTD interface (in orange) undergoes significant local conformational rearrangement upon partial R-loop formation. **b**, SDS-PAGE of wild type and mutant Cascades used in mutagenesis. All mutants contained stoichiometric amounts of Cse1, except W307A, in which Cse1 failed to assemble. **c**, Detailed structure rearrangement at Cse1 NTD/CTD interface, involving amino acids 300–326. Different colors are used to differentiate structural elements (amino acid numbers tabulated to the right side). Partial R-loop and free-Cascade structures are rendered in solid and semi-transparent cartoons, respectively. **d**, W307 from Cse1-NTD rotates inside a hydrophobic socket at Cse1-CTD during the conformational change (free-Cascade-CTD rendered in a darker shade of green). Disruption of this interaction (W307A) caused Cse1 dissociation from Cascade. **e**, Docking of our structure into the cryo-EM reconstruction of the crosslinked Cascade-dsDNA-Cas3 complex (EMD-5930). Cas3 density and the location of the local rearrangement in Cse1 are circled to highlight their proximity.

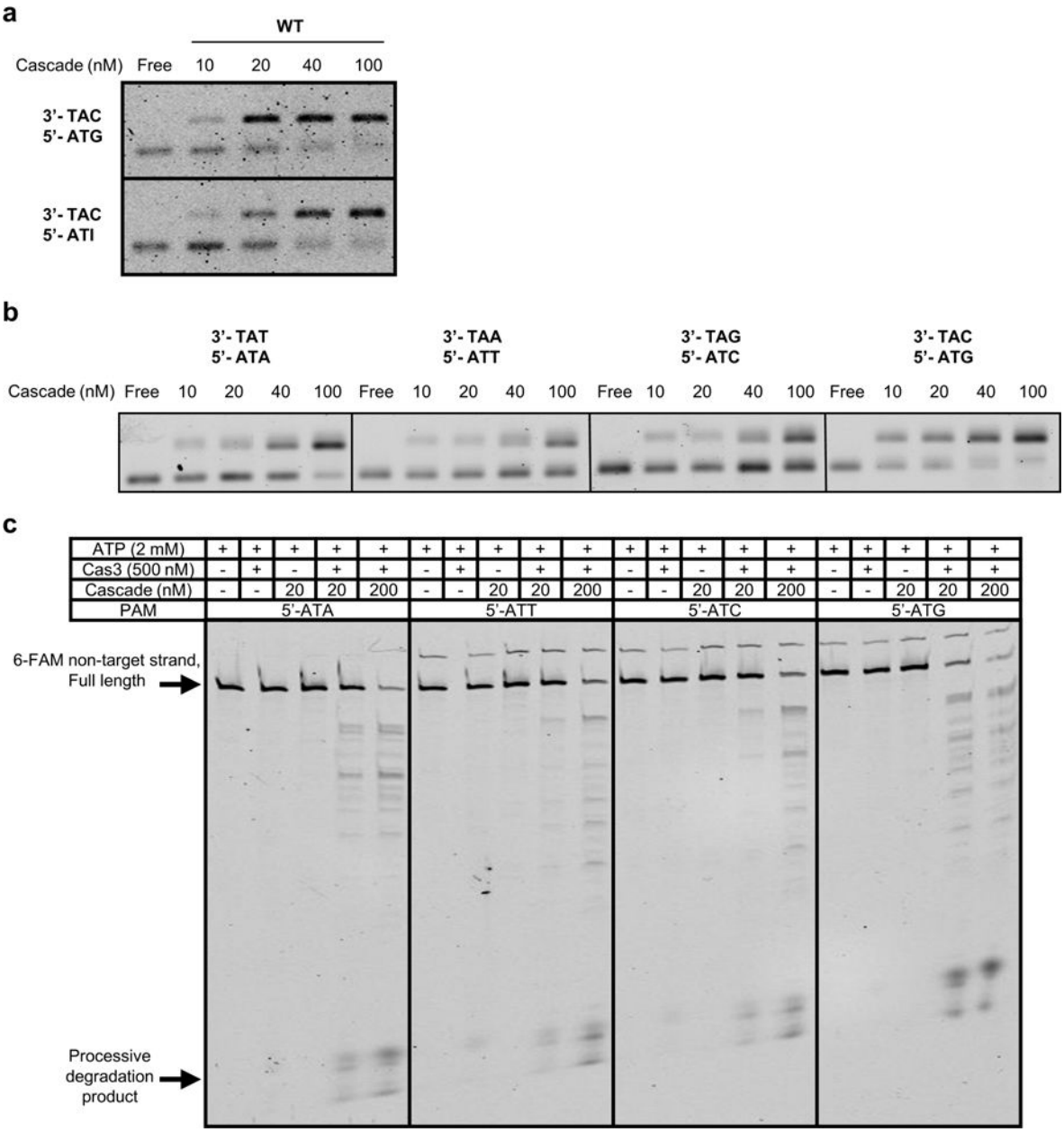
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 5. Guided entry of dsDNA into Cascade

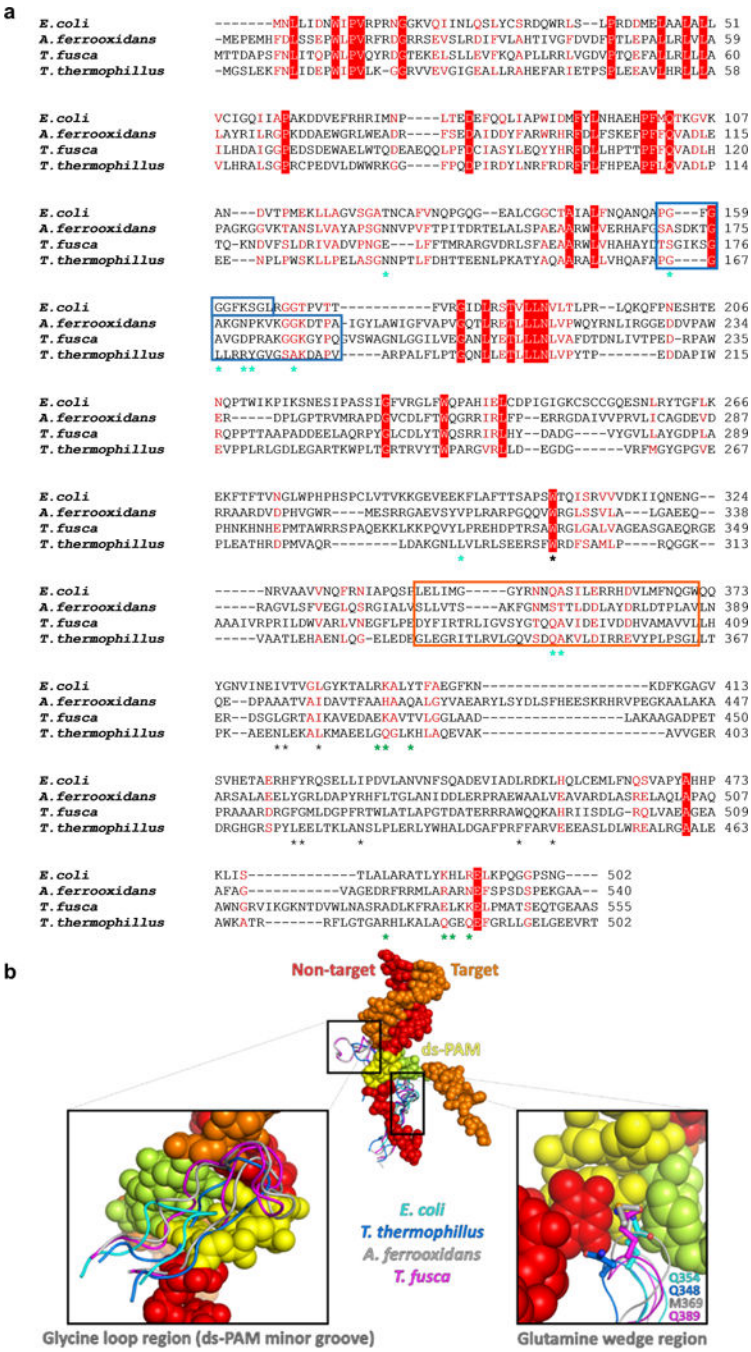
A series of positively charged residues from Cas7.5, Cas7.6, Cas5e and Cse1 (dark blue surfaces) guide the dsDNA into Cascade and towards PAM recognition elements in Cse1-NTD.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Figure 6. Influence of PAM-1 base-pair composition on Cascade binding affinity and Cas3 cleavage efficiency

a, Inosine substitution of G_{NT-1} led to a mild 2-fold binding defect, suggesting the N2 amine of G_{NT-1} is a minor determinant of specificity. **b**, Binding Kd of *E. coli* Cascade for all four base-pair combinations at PAM-1. Kd for 5′-ATG, ATA, ATC, and ATT PAMs were determined to be ~10, 20, 40, and 100 nM, respectively. **c**, Cas3 cleavage efficiency was governed by Cascade's affinity for the corresponding PAM-containing target (5′-ATG>ATA≫ATT≈ATC). Cascade concentration above the Kd led to efficient Cas3 cleavage.

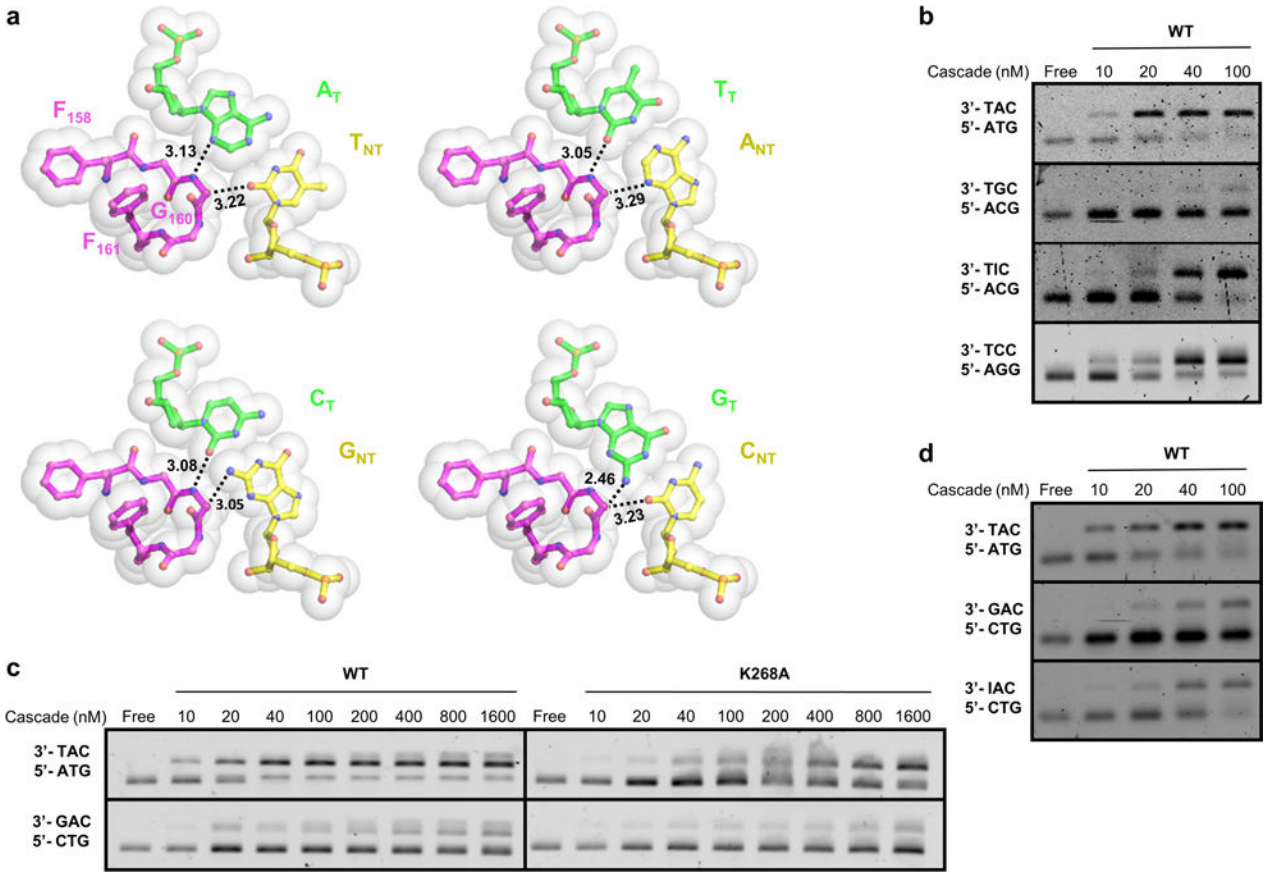


Extended Data Figure 7. Conservation of PAM recognition elements

a, Sequence alignment of Cse1 proteins with known structures. Identical residues are highlighted in red, conserved residues in red text. DNA-contacting residues in the Cse1-NTD and CTD are marked by cyan and green asterisks, respectively. The glycine loop and glutamine wedge are in blue and orange boxes, respectively. Residues mediating the W307 ball-and-socket interaction are marked with black asterisks. **b**, Structural alignment of Cse1 PAM recognition elements. Apo Cse1 from *T. thermophillus* (PDB: 4AN8), *A. ferrooxidans* (PDB: 4H3T) and *T. fusca* (PDB: 3WVO) were superimposed with the Cse1 in our partial

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

R-loop forming *E. coli* Cascade structure. Despite sequence variation, a glycine-rich loop is present in each Cse1 structure, and likely plays a similar function to recognize PAM from the minor groove (left inset). The glutamine-wedge protrusion is highly conserved in 3-D. Each wedge features a long side chain at the tip (right inset), which likely stacks underneath PAM in a similar fashion.



Extended Data Figure 8. Rationalization of PAM-2 and PAM-3 specificity using nucleotide substitution and modeling

a, Modeling of alternative base-pairs at PAM-2 suggests that only the N2 amine of $G_{T\text{-}2}$ would cause steric clashes with C_α of G160 (lower right quadrant), this amine therefore may serve as the anti-determinant for the rejection of G_T-C_{NT} at PAM-2. **b**, EMSAs demonstrating that removal of this amine in inosine substitution rescued the Cascade binding defect. **c**. Whereas K268A contained reduced affinity for the correct PAM, it still possessed strong discrimination against G_T-C_{NT} at PAM-3 (5′-CTG), suggesting the further presence of a mechanism to reject $G_{T\text{-}3}$. **d**, Inosine substitution of $G_{T\text{-}3}$ restored the Cascade binding Kd to ~40 nM, leading to the conclusion that the N2 amine of $G_{T\text{-}3}$ is a minor determinant of specificity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Extended Data Table 1

Oligonucleotides used in this study.

Oligonucleotide	Sequence (5′–3′)
Cse1 (forward)	GCGCGCCATGGCTAATTTGCTTATTGATAACTGGATCC
Cse1 (reverse)	GGCCCGCGGCCGCTCAGCCATTTGATGGCCCTCC
Cse2-Cas7-Cas5e-Cas6e (forward)	GCGCGGGTACCAGATGGCTGATGAAATTGATGCA
Cse2-Cas7-Cas5e-Cas6e (reverse)	CGCGCGCGGCCGCTCACAGTGGAGCCAAAGATAG
Cse2-Cas7-Cas5e-Cas6e (forward)	GCGCGCCATGGGTCATCACCACCATCATCACGGTGCACTTGAAGTCCTCTTTC
R-loop mimic (target strand)	CTGTTGGCAAGCCAGGATCTGAACAATACCGTCATCGAGCACTGCACAGA
R-loop mimic (non-target strand)	TCTGTGCAGTGCTCGATGTTTTATTTAT
SUMO-Cas3 (forward)	GCGCCGCGGATCCATGGAACCTTTTAAATATATATGCCAT
SUMO-Cas3 (reverse)	GCGCCGCCTCGAGTTATTTGGGATTTGCAGGGAT
Target plasmid construction (non-target strand)	CATGG ATG ACGGTATTGTTCAGATCCTGGCTTGCCAACAGCTGCA
Target plasmid construction (target strand)	GCTGTTGGCAAGCCAGGATCTGAACAATACCGT CAT C

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Oligonucleotide	Sequence (5′–3′)
dsDNA1 Fluorescent ATG PAM substrate (forward)	AACTTTAATAAGGAGATATACCATGGATG
dsDNA1 Fluorescent ATG PAM substrate (reverse)	GCGGCCGCAAGCTTGTC
dsDNA2 Fluorescent substrate (forward)	TAATACGACTCACTATAGGG
dsDNA3 (non-target strand)	AGATATACATGG *ATG* ACGGTATTGTTCAGATCCTGGCTTGCCAACAGCTGCAGGTCGACA
dsDNA3 (target strand)	TGTCGACCTGCAGCTGTTGGCAAGCCAGGATCTGAACAATACCGT *CAT* CCATGTATATCT
crRNA expression cassette	GCGCCGGGAATTCCCTGCATTAGG *TAATACGACTCACTATAGG* *GAGTTCCCCGCGCCAGCGGGGATAAACCG* ACGGTATTGTTCAGATCCTGGCTTGCCAACAG *GAGTTCCCCGCGCCAGCGGGGATAAACCG*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Oligonucleotide	Sequence (5′–3′)
	ACGGTATTGTTCAGATCCTGGCTTGCCAACAG GAGTTCCCCGCGCCAGCGGGGATAAACCG ACGGTATTGTTCAGATCCTGGCTTGCCAACAG GAGTTCCCCGCGCCAGCGGGGATAAACCG ACGGTATTGTTCAGATCCTGGCTTGCCAACAG GAGTTCCCCGCGCCAGCGGGGATAAACCG ACGGTATTGTTCAGATCCTGGCTTGCCAACAG GAGTTCCCCGCGCCAGCGGGGATAAACCG ACGGTATTGTTCAGATCCTGGCTTGCCAACAG GAGTTCCCCGCGCCAGCGGGGATTTTGCC CTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTT GGACCGGATCCCCG

Extended Data Table 2

Data collection and refinement statistics.

	Cascade-dsDNA	Cascade-dsDNA (32-nt non-target spacer)
Data collection		
Space group	P2 $2_1 2_1$	P 2 2_1 2_1
Cell dimensions		
a, b, c (Å)	92.98, 150.06,400.55	92.81,149.83,404.10
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	50.0–2.45 (2.49−2.45)	50.0–3.20 (3.26−3.21)*
R_{merge}	0.169 (1.288)	0.265 (1.066)
R_{pim}	0.050 (0.531)	0.143 (0.667)
$I/\sigma I$	12.9 (1.1)	5.8 (1.5)
Completeness (%)	99.7 (97.2)	96.8 (84.1)
Redundancy	12.1 (6.5)	3.6 (3.0)
CC(l/2)	0.997 (0.496)	0.967 (0.422)
Refinement		
Resolution (Å)	50.0–2.45	50.0–3.21
No. reflections	205,381	90,215
R_{work}/R_{free}	0.2058/0.2327	0.2087/0.2470
No. atoms		
Macromolecule	27896	27237
Ligand/ion	1(Zn)	1(Zn)
Water	1051	0
B-factors		
Macromolecule	53.60	62.10
Ligand/ion	51.90	61.50
Water	45.90	N/A
R.m.s. deviations		
Bond lengths (Å)	0.013	0.018
Bond angles (°)	0.840	1.010

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

*One crystal was used for each structure.

*Values in parentheses are for highest-resolution shell.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by NIH grants GM102543 and GM086766 to A.K, GM097330 to S.B. and GM108888 to B.W. NE-CAT beamlines were supported by NIH grants P41 GM103403 and S10 RR029205. We thank Gerald Feigenson and John Mallon for technical help, Ilya Finkelstein and Ian Price for helpful discussions.

References

1. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nature reviews. Microbiology. 2014; 12:479–492. [PubMed: 24909109]
2. Luo ML, Mullis AS, Leenay RT, Beisel CL. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. Nucleic acids research. 2015; 43:674–681. [PubMed: 25326321]
3. Caliando BJ, Voigt CA. Targeted DNA degradation using a CRISPR device stably carried in the host genome. Nature communications. 2015; 6:6989.
4. Westra ER, et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Molecular cell. 2012; 46:595–605. [PubMed: 22521689]
5. Jore MM, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature structural & molecular biology. 2011; 18:529–536.
6. Brouns SJ, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008; 321:960–964. [PubMed: 18703739]
7. Wiedenheft B, et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature. 2011; 477:486–489. [PubMed: 21938068]
8. Westra ER, et al. Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS genetics. 2013; 9:e1003742. [PubMed: 24039596]
9. Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature. 2010; 463:568–571. [PubMed: 20072129]
10. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology. 2009; 155:733–740. [PubMed: 19246744]
11. Mulepati S, Bailey S. In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. The Journal of biological chemistry. 2013; 288:22184–22192. [PubMed: 23760266]
12. Rollins MF, Schuman JT, Paulus K, Bukhari HS, Wiedenheft B. Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. Nucleic acids research. 2015; 43:2216–2222. [PubMed: 25662606]
13. Sashital, Dipali G.; Wiedenheft, B.; Doudna, Jennifer A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. Molecular cell. 2012; 46:606–615. [PubMed: 22521690]
14. Sinkunas T, et al. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. The EMBO journal. 2013; 32:385–394. [PubMed: 23334296]
15. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014; 507:62–67. [PubMed: 24476820]
16. van Erp PB, et al. Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. Nucleic acids research. 2015

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

17. Rutkauskas M, et al. Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. Cell reports. 2015
18. Blosser TR, et al. Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. Molecular cell. 2015; 58:60–70. [PubMed: 25752578]
19. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. 2014; 513:569–573. [PubMed: 25079318]
20. Nishimasu H, et al. Crystal Structure of Staphylococcus aureus Cas9. Cell. 2015; 162:1113–1126. [PubMed: 26317473]
21. Hochstrasser ML, et al. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:6618–6623. [PubMed: 24748111]
22. Fineran PC, et al. Degenerate target sites mediate rapid primed CRISPR adaptation. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:E1629–1638. [PubMed: 24711427]
23. Zhao H, et al. Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. Nature. 2014; 515:147–150. [PubMed: 25118175]
24. Jackson RN, et al. Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science. 2014; 345:1473–1479. [PubMed: 25103409]
25. Huo Y, et al. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nature structural & molecular biology. 2014; 21:771–777.
26. Jackson RN, et al. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science. 2014; 345:1473–1479. [PubMed: 25103409]
27. Mulepati S, Heroux A, Bailey S. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science. 2014; 345:1479–1484. [PubMed: 25123481]
28. Xue C, et al. CRISPR interference and priming varies with individual spacer sequences. Nucleic acids research. 2015
29. Datsenko KA, et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nature communications. 2012; 3:945.
30. Redding S, et al. Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. Cell. 2015
31. Otwinowski Z, Minor W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology. 1997; 276:307–326.
32. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta crystallographica. Section D, Biological crystallography. 2004; 60:2126–2132. [PubMed: 15572765]
33. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, Biological crystallography. 2010; 66:213–221. [PubMed: 20124702]
34. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010
35. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. Acta crystallographica. Section D, Biological crystallography. 1994; 50:760–763. [PubMed: 15299374]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

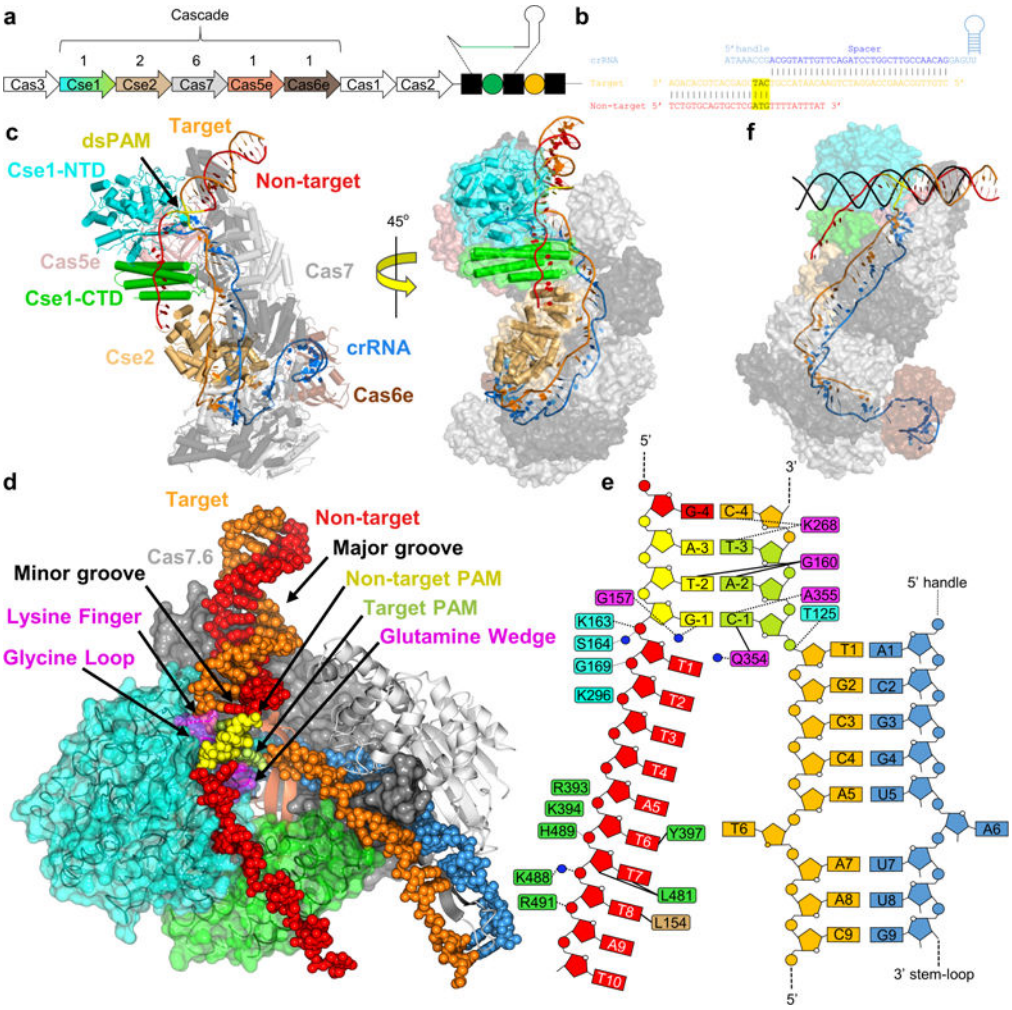


Figure 1. Foreign DNA bound Cascade structure

a, Arrangement of I-E CRISPR-*cas* locus in the *E. coli* K12 genome. The color schemes are preserved throughout all figures. **b**, Nucleic acid sequences and **c**, overall views of foreign DNA-bound Cascade structure. **d**, Entry of dsDNA between Cas7.5 and Cas 7.6, PAM recognition by three structural elements in Cse1-NTD (magenta), and partial R-loop underneath PAM. **e**, Schematic of Cascade-DNA contacts around PAM regions. Hydrogen bonds and electrostatic contacts as dashed lines, hydrophobic interactions as solid lines, and waters as blue circles. **f**, Modeling of Cascade sampling B-form dsDNA for PAM.

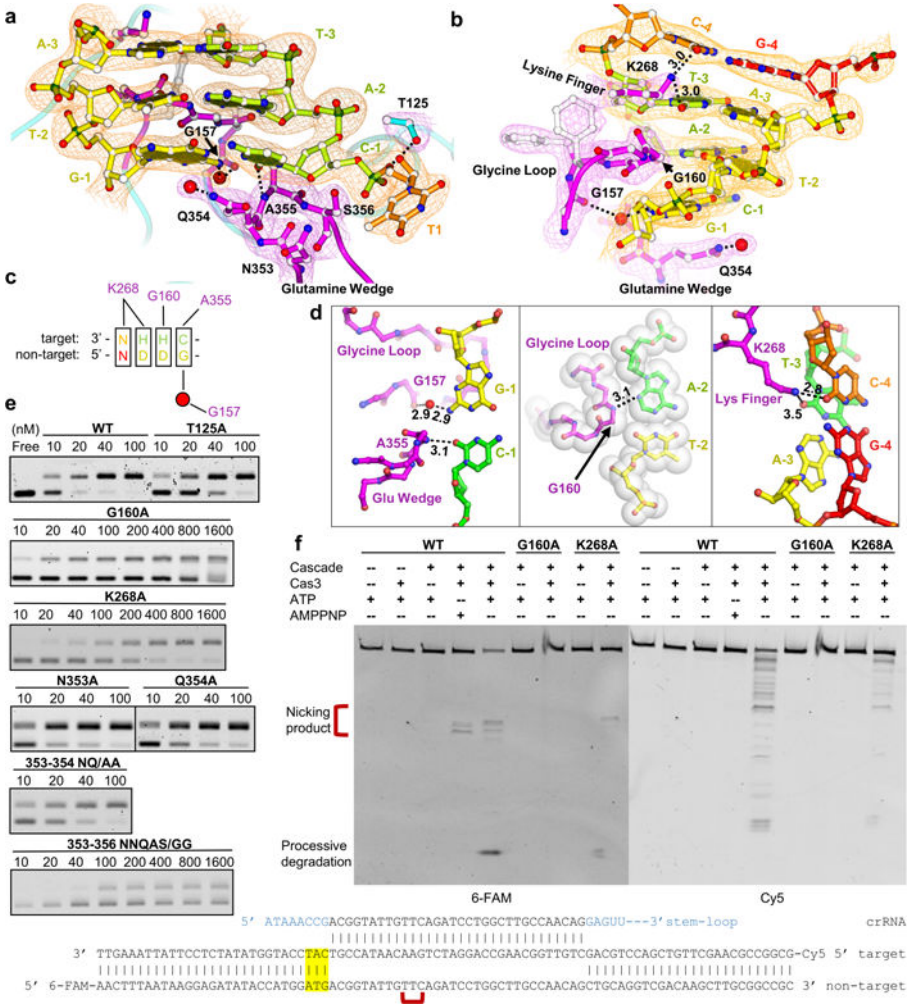


Figure 2. PAM recognition by Cse1 subunit of Cascade

a, b, Detailed views featuring PAM recognition by the glutamine-wedge (and its involvement in target DNA strand displacement), glycine-loop, and lysine-finger of Cse1. The $2F_O$-F_C electron density map is displayed at 1.5 **σ**. **c**, Summarization of the five interference PAMs in *E. coli* Type I-E CRISPR system, with the observed Cascade contacts marked. **d**, Top-down views of PAM recognition at each base-pair. Left: stringent recognition of PAM-1. Middle: shape complementarity to PAM-2. Right: Electrostatic contacts to PAM-3 and -4. **e**, Mutagenesis assayed by Cascade-binding (EMSA) and **f**, Cas3-mediated DNA cleavage to evaluate the observed PAM contacts.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

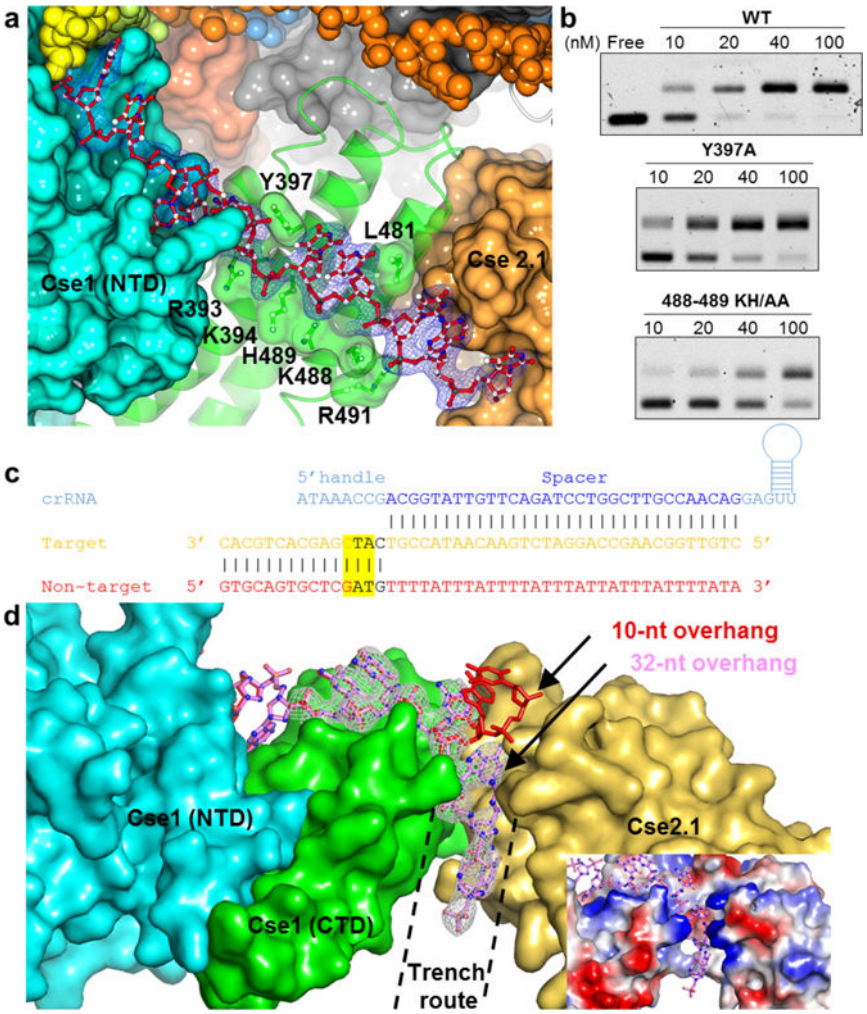


Figure 3. Active guidance of the non-target DNA strand by Cascade

a, Sequence nonspecific electrostatic contacts guide the 10-nt non-target strand (red sticks) protospacer along Cse1-NTD surface (cyan), and across a binding site on Cse1-CTD (green). Y397 and L481 from Cse1-CTD further interdigitate between di-nucleotide stacks. **b**, EMSA evaluating non-target contacts by Cascade. **c, d**, Nucleic acid sequence and Zoom-in view of the 3.2 Å structure of Cascade programmed with a 22-nt longer non-target strand. The $2F_O$-F_C electron density map at 1.0 **σ** clearly reveals that the longer non-target DNA strand takes the trench route. Inset illustrates the favorable electrostatic surface.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

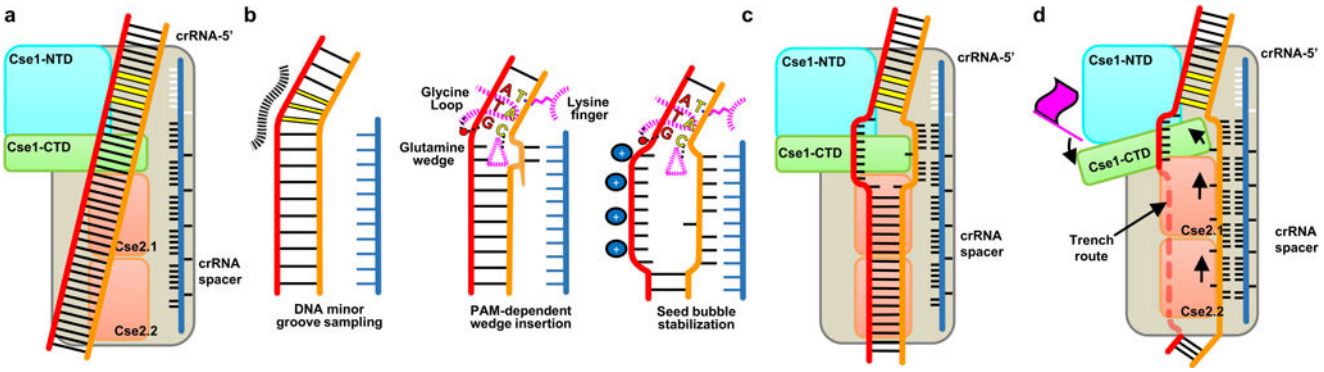


Figure 4. Model for PAM-dependent directional R-loop formation in Type I CRISPR-Cas system

a, Cascade samples dsDNA minor groove for various sequence combinations (yellow sticks). **b**, Left: interference and priming PAMs lead to longer dwell time and local DNA bending. Middle: Interference PAMs allow optimal minor groove interaction, which is coupled with the glutamine wedge insertion and disruption of the first 2-bp of protospacer. Right: Directional DNA melting leads to segmented DNA/crRNA duplex formation at the target strand side, and favorable sugar-phosphate contacts to the non-target side, leading to seed bubble stabilization. **c**, Further DNA unwinding leads to the non-target strand sequestration to the backside of Cse2 dimer, locking R-loop in place. **d**. R-loop formation is accompanied by a pivoting motion in Cse1-CTD and a sliding motion in Cse2 dimer. Local rearrangement occurs in Cse1 (depicted as a flag), licensing Cas3 recruitment.