Supplementary Information for

DNA targeting by subtype I-D CRISPR-Cas shows type I and type III features

Jinzhong Lin, Anders Fuglsang, Anders Lynge Kjeldsen, Kaiyan Sun, Yuvaraj Bhoobalan-Chitty & Xu Peng*

*Corresponding author. Email: peng@bio.ku.dk.

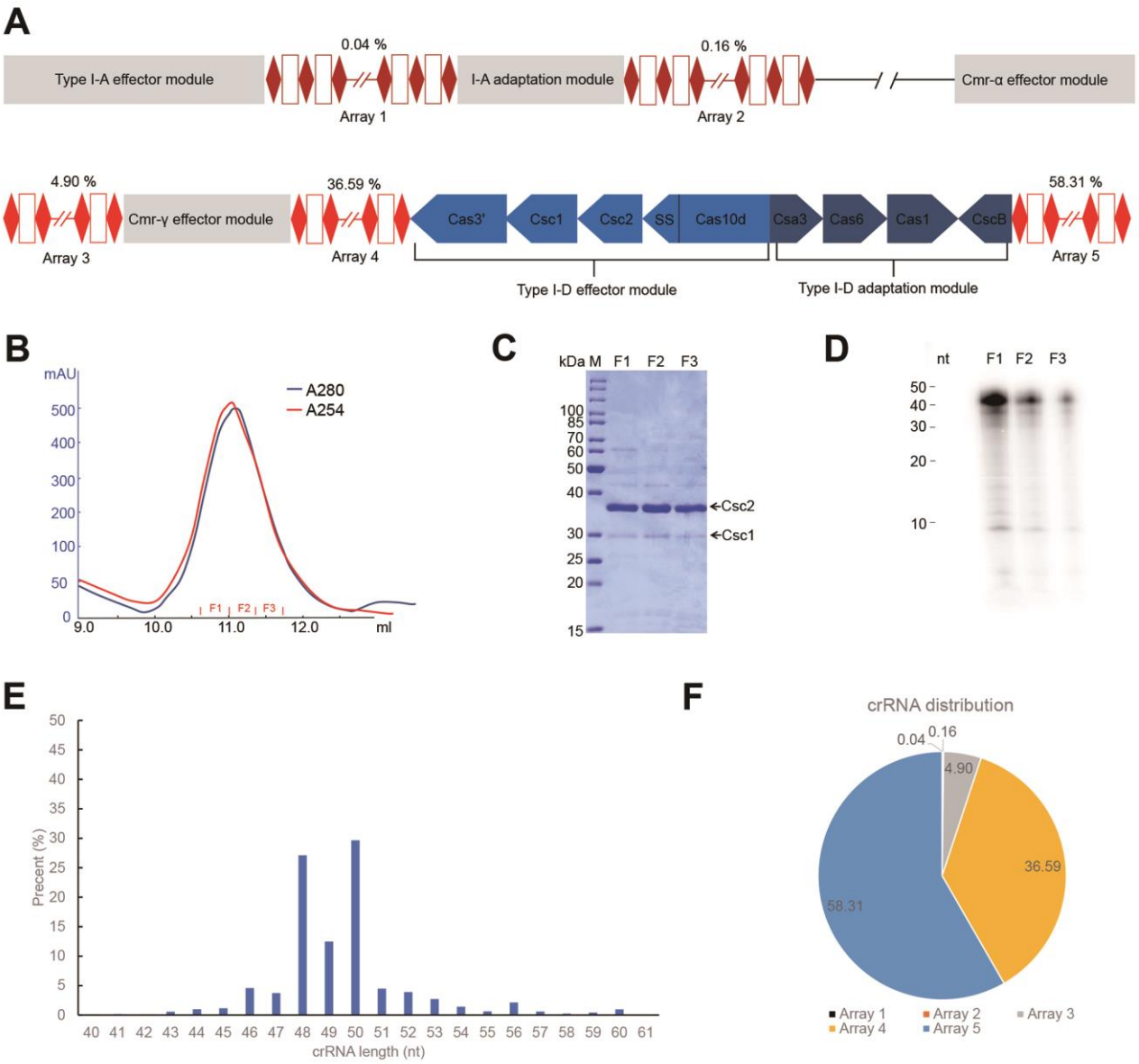

Figure S1. crRNA content of the endogenous backbone complex. **(A)** Overview of the CRISPR-Cas systems in S. islandicus LAL14/1. The proportion of crRNA (%) present in type I-D backbone complex is indicated for each array. **(B)** Size exclusion chromatography (SEC) purification of type I-D backbone, blue: UV absorbance at 280 nm; red: UV absorbance at 254 nm. **(C)** SDS-PAGE of SEC samples collected in the peak region in (B). M: protein mass marker. **(D)** Denaturing gel electrophoresis of 5'-labelled crRNAs from SEC samples collected in the peak region in **(B)**. **(E)** Size-distribution of the type I-D crRNAs. **(F)** Distribution of the type I-D crRNAs over the 5 CRISPR arrays, see details in **Figure S2**.

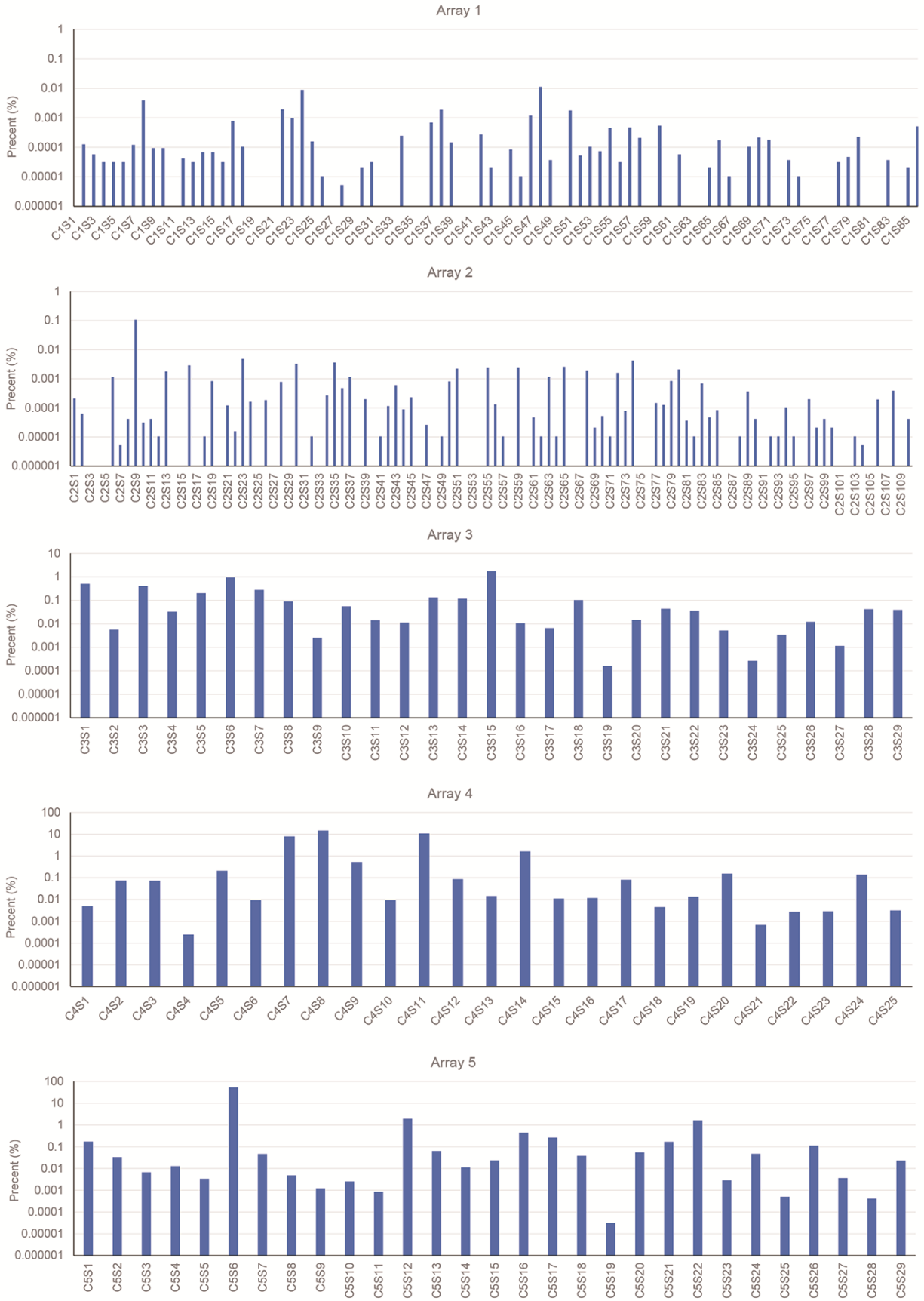

Figure S2. Distribution of the type I-D crRNAs over the 5 CRISPR arrays in detail. CxSy: spacer y in CRISPR array x.

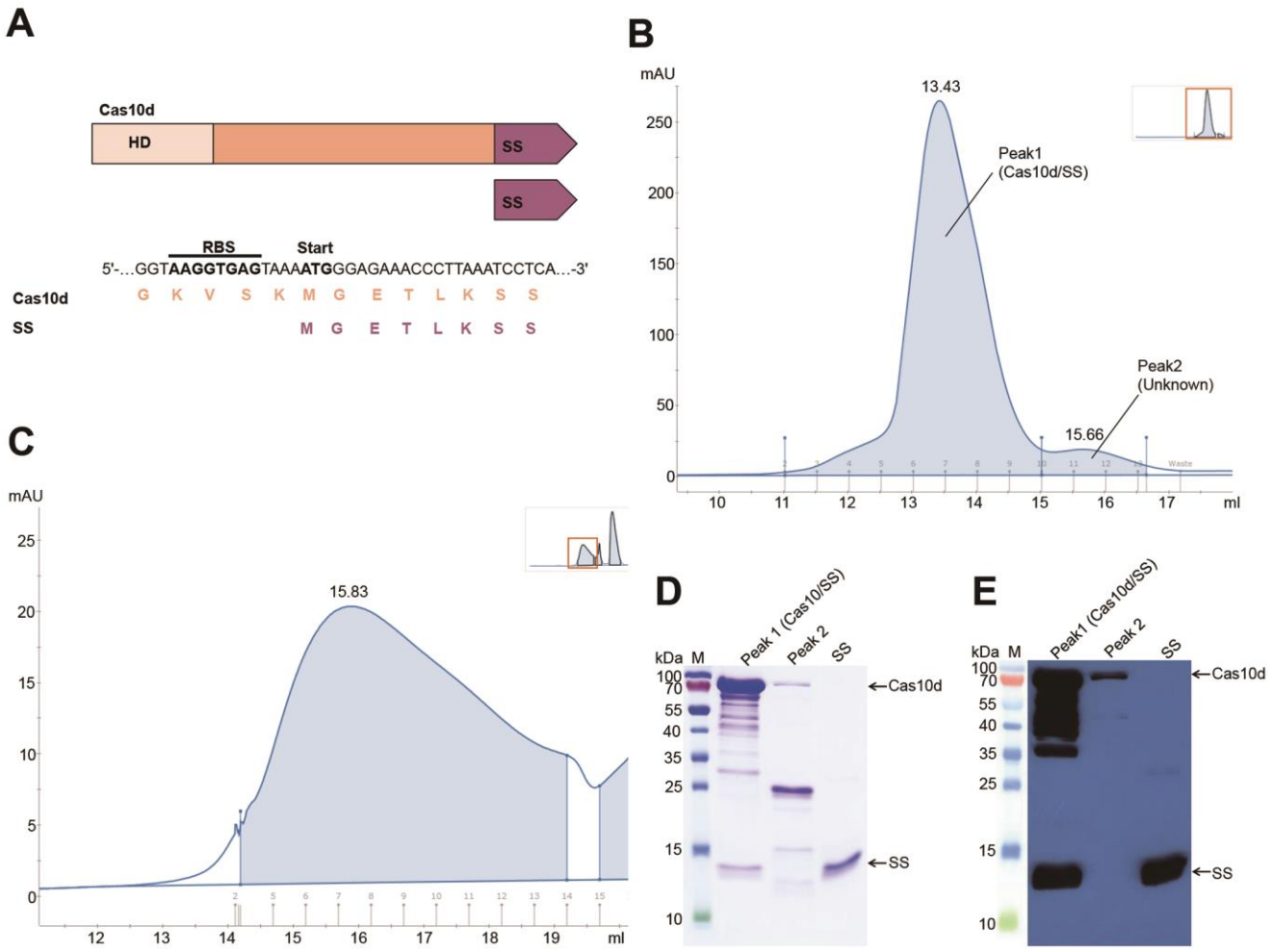

Figure S3. Predicted small subunit is expressed from an internal translational start codon in cas10d. **(A)** Schematic presentation of Cas10d showing the HD domain and the predicted C-terminal small subunit (SS). The DNA sequence around the internal start codon ATG is shown and the upstream ribosome-binding site is indicated. Amino acid sequences for this region are also shown for both Cas10d and SS. **(B)** Size exclusion chromatography (SEC) purification of Cas10d/SS from *E. coli*. Nickel-His tag affinity purified proteins were subjected to SEC and the SEC chromatogram is shown, The complete chromatogram is included in a small scale on top. Peak 1 (elution volume 13.4 ml) contained Cas10d and SS whereas peak 2 (elution volume 15.6 ml) contained unidentified proteins but no SS (see panels **(D)** and **(E)**). **(C)** SEC purification of SS with a C-terminal His-tag from *E. coli*. SS with a C-His tag was expressed in and purified from E. coli separately to examine the SEC elution volume of a free SS. The free SS was eluted in approximately 15.8 ml, but not in 13.43 ml (panle **(B)**), indicating SS must interact with Cas10d (see panel **(B)**). SDS-PAGE **(D)** and Western blotting **(E)** of proteins from peaks 1 and 2 (panel **(B)**) as well as the free SS (panel **(C)**). M: protein mass marker. An anti-His-tag antibody was used in **(E)**. In contrast to Cas10d used in the cleavage assays (**Figure 1B**), Cas10d prepared in this experiment exhibited significant degradation.

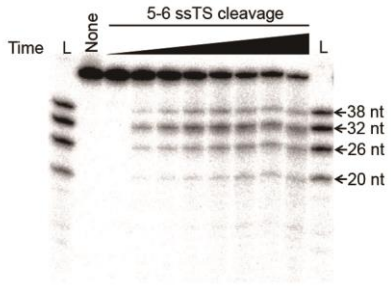

Figure S4. Time course of ssDNA cleavage by type I-D effector. Cleavage products were collected at 0, 20, 40, 60, 80, 100, 120 and 150 min. None: No effector complex was added, L: ssDNA size marker with lengths in nt indicated.

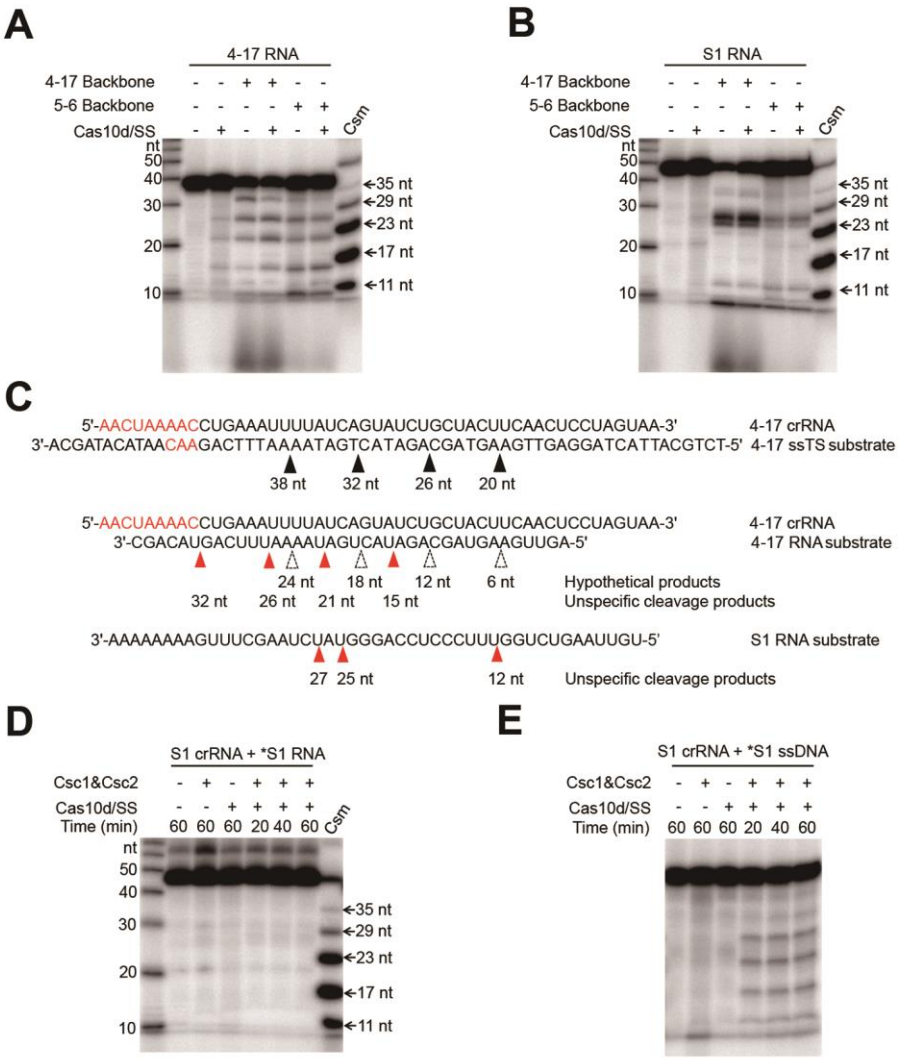

Figure S5. No specific RNA cleavage by type I-D effector. **(A)** and **(B)** Unspecific RNA cleavage by contaminating RNase present in type I-D components. The radiolabelled RNA targets are shown on top of each gel under which the presence (+) or absence (-) of the type I-D components is indicated. S1 RNA cleavage products cleaved by Lactobacillus delbrueckii Type III Csm (1) carrying S1 crRNA are shown in both panels as control. The sizes of S1 cleavage products by Csm are depicted on the right side of each gel. Nearly identical cleavage patterns on the same RNA (4-17 RNA) observed for the I-D effectors carrying different crRNAs (4-17 and 5-6 crRNAs) indicate the cleavage is unspecfic **(A)**. **(C)** Schematic depiction of the cleavage sites on RNA substrates. Hypothetical cleavage sites for a putative specific RNA cleavage are indicated by empty triangles and the observed cleavage site are shown by red triangles. Specific ssDNA cleavage sites (dark triangles) are presented as reference. Both backbone samples purified from Sulfolobus showed cleavage between UA and UG (3'-5'), likely due to a contaminating RNase present in these samples. RNA cleavage **(D)** and ssDNA cleavage **(E)** by type I-D effector. Radiolabled S1 RNA and S1 ssDNA were used and indicated on top. Reactions were performed using Csc1 and Csc2 complex purified from *E. coli* and a purchased S1 crRNA, in the presence (+) or absence (-) of Cas10d/SS.

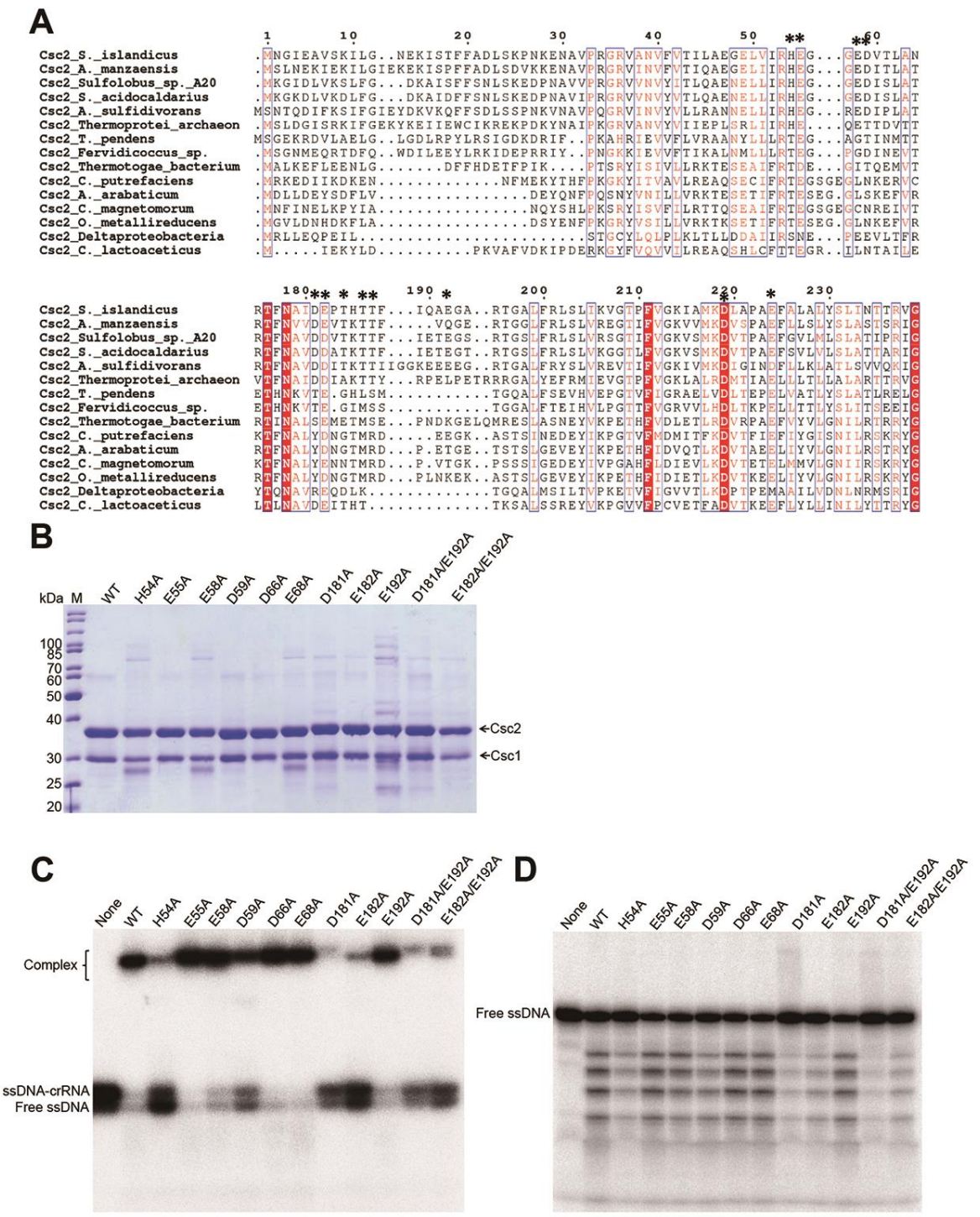

Figure S6. Effect of Csc2 mutations on ssDNA binding, ssDNA and dsDNA cleavage by type I-D effector. **(A)** Conserved residues in Csc2 subunit. Fourteen *Sulfolobus islandicus* Csc2 homologues from: *Acidianus manzaensis*, *Sulfolobus* sp. A20, *Sulfolobus acidocaldarius*, *Acidianus sulfidivorans*, *Thermoprotei archaeon*, *Thermofilum pendens*, *Fervidicoccus* sp., *Thermotogae bacterium*, *Clostridium putrefaciens*, *Acetohalobium arabaticum*, *Candidatus Magnetomorum*, *Orenia metallireducens*, *Deltaproteobacteria bacterium* and *Caldicellulosiruptor lactoaceticus* were selected and aligned using MEGA5 (2) and visualized using ESPript 3 (3). Residues marked with stars were

mutated individually to produce the Csc2 mutants. **(B)** SDS-PAGE of Csc1 and Csc2 (WT and mutants) complex purified from *E. coli*. M: protein mass marker. ssDNA binding **(C)** and ssDNA cleavage **(D)** by type I-D effectors carrying the WT or mutated Csc2 as indicated on top. None: No effector complex was added.

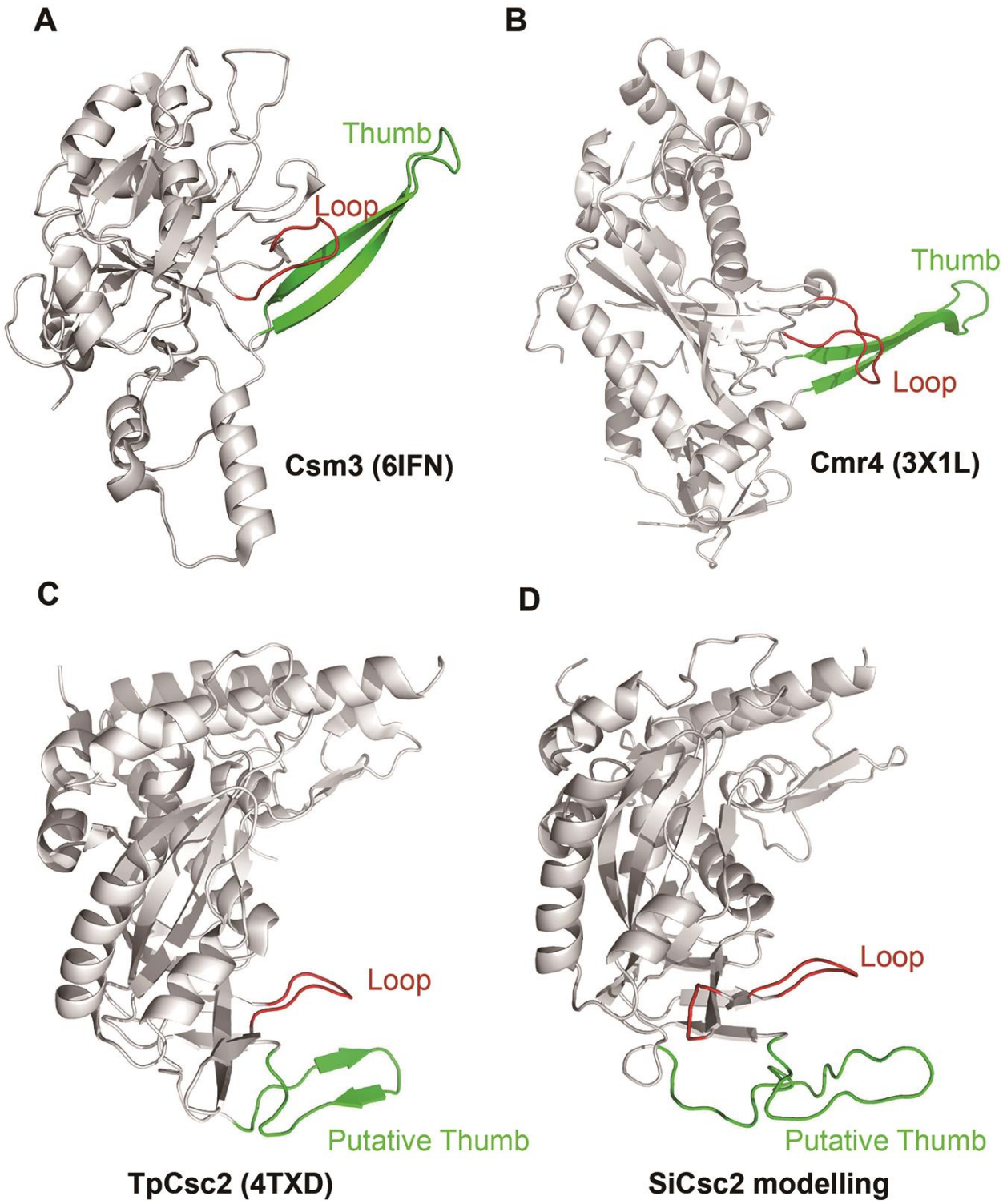

Figure S7. *S. islandicus* LAL14/1 Csc2 (SiCsc2) structure modelling. **(A), (B)** and **(C)** Structures of *Streptococcus thermophilus* Csm3 (4), *Pyrococcus furiosus* Cmr4 (5), and *Thermofilum pendens* Csc2 (6), respectively. **(D)** Structural model of SiCsc2 made by I-TASSER (7) using the solved structure of *T. pendens* Csc2 as reference. Loop and thumb are indicated in each structure.

Table S1 Nucleic acid oligos and substrates used in this study.

ssDNA and RNA	Sequence (5'-3')	Description
oligos		
NdeI-Csc1-F	TAAGAAGGAGAGCACATATGAAGTTGTATAAGGCAAATTTTTAC	For plasmid pEXA2-Csc1 construction
NheI-His-Csc1-R	CTAGGCTAGCTTAATGGTGGTGATGATGATGTTTCTTTTCCTCGCTTTGCC	For plasmid pEXA2-Csc1 construction
Re-5-6-F	GTAACAACACAAAGAAACTAAAACGAAATTTGGAAAGTAGGAGAAAAGAACAA	For plasmid pEXA2-Csc1-5-6 construction
Re-5-6-R	GTTTTAGTTTCTTTGTGTTGTTACTATCTATAGATTGTTCTTTTCTCCTACTT	For plasmid pEXA2-Csc1-5-6 construction
Re-4-17-F	GTAACAACACAAAGAAACTAAAACCTGAAATTTTATCAGTATCTGCTACTTCA	For plasmid pEXA2-Csc1-4-17 construction
Re-4-17-R	GTTTTAGTTTCTTTGTGTTGTTACTTACTAGGAGTTGAAGTAGCAGATACTGA	For plasmid pEXA2-Csc1-4-17 construction
Csc1-F	TAAGAAGGAGAGCACATATGAAGTTGTATAAGGCAAATTTTTAC	For plasmid pETDuet-Csc1-Csc2 construction
His-Csc1-R	TTAATGGTGGTGATGATGATGTTTCTTTTCCTCGCTTTGCC	For plasmid pETDuet-Csc1-Csc2 construction
Csc2-F	ATAAGAAGGAGATATACATATGAATGGGATTGAAGCAGTCTC	For plasmid pETDuet-Csc1-Csc2 construction
Csc2-R	GTTTCTTTACCAGACTCGAGTCACTTTTTAATGAATTCCTCT	For plasmid pETDuet-Csc1-Csc2 construction
Cas3'-F	GAGAACCTCTACTTCCAATCGATGCCTTCCAATAAATTGGC	For plasmid pMAL-TEV-Cas3' construction

Cas3'-R	TTAATGGTGGTGATGATGATGTATAAATAATG CTAAACCTATTAC	For plasmid pMAL-TEV-Cas3' construction
5-6-F	AGGAAAAAATTAACAAAACTACAAATTATGCA GACTGTGGAAATTTGGAA	For 5-6 dsDNA SOE-PCR
5-6-R1	TTGAATATCTATAGATTGTTCTTTTCTCCTACT TTCCAAATTTCCACAGT	For 5-6 dsDNA SOE-PCR
5-6-R	TGTAAATCTATATCTAAAGCTTGTAAAACTATT TGAATATCTATAGATTG	For 5-6 dsDNA SOE-PCR
5-6 NTS-F	AGGAAAAAATTAACAAAAC	For 5-6 dsDNA substrate production
5-6 TS-F	TGTAAATCTATATCTAAAG	For 5-6 dsDNA substrate production
Cas10d-HD-A-F	CCAAGCGGCGAGAAACCCCCCGTAAAACC	For Cas10d HD-A mutation
Cas10d-HD-A-R	TTTCTCGCCGCTTGGAATAAATTGAGTGGAAA AGAGGAGTC	For Cas10d HD-A mutation
Csc2-H54A-F	TAATAAGAGCTGAAGGAGGAGAGGACGTA	For Csc2 H54A mutation
Csc2-H54A-R	CCTCCTTCAGCTCTTATTACCAGTTCTCCTTC	For Csc2 H54A mutation
Csc2-E55A-F	TAAGACATGCAGGAGGAGAGGACGTAACTTT AG	For Csc2 E55A mutation
Csc2-E55A-R	CTCTCCTCCTGCATGTCTTATTACCAGTTCTC	For Csc2 E55A mutation
Csc2-E58A-F	AAGGAGGAGCGGACGTAACTTTAGCAAACA	For Csc2 E58A mutation
Csc2-E58A-R	GTTACGTCCGCTCCTCCTTCATGTCTTATTAC	For Csc2 E58A mutation
Csc2-D59A-F	GGAGGAGAGGCCGTAACTTTAGCAAACATTG	For Csc2 D59A mutation
Csc2-D59A-R	AAAGTTACGGCCTCTCCTCCTTCATGTCTTA	For Csc2 D59A mutation
Csc2-D66A-F	CAAACATTGCCGGAGAAAAATACCCAATGAT	For Csc2 D66A mutation
Csc2-D66A-R	TTTTCTCCGGCAATGTTTGCTAAAGTTACG	For Csc2 D6655A mutation
Csc2-E68A-F	TTGACGGAGCAAAATACCCAATGATATTGCAC	For Csc2 E68A mutation

Csc2-E68A-R	GGGTATTTTGCTCCGTCAATGTTTGCTAAAG	For Csc2 E68A mutation
Csc2-D181A-F	ACGCAATTGCTGAACCTACGCATACTACC	For Csc2 D181A mutation
Csc2-D181A-R	GTAGGTTCAGCAATTGCGTTAAATGTTCTC	For Csc2 D181A mutation
Csc2-E182A-F	GCAATTGATGCACCTACGCATACTACCTTTA	For Csc2 E182A mutation
Csc2-E182A-R	TGCGTAGGTGCATCAATTGCGTTAAATGTTC	For Csc2 E182A mutation
Csc2-E192A-F	TTCAAGCTGCGGGAGCTAGGACTGGTGCAC	For Csc2 E192A mutation
Csc2-E192A-R	CTAGCTCCCGCAGCTTGAATAAAGGTAGT	For Csc2 E192A mutation
Csc2-D181/E182A-F	CGCAATTGCTGCACCTACGCATACTACCTTTATTC	For Csc2 D181A/E182A mutation
Csc2-D181/E182A-R	CGTAGGTGCAGCAATTGCGTTAAATGTTCTC	For Csc2 D181/E182A mutation
Csc2-D181N-F	ACGCAATTAATGAACCTACGCATACTACC	For Csc2 D181N mutation
Csc2-D181N-R	GTAGGTTCATTAATTGCGTTAAATGTTCTCAC	For Csc2 D181N mutation
Csc2-T184A-F	ATGAACCTGCGCATACTACCTTTATTCAAG	For Csc2 T184A mutation
Csc2-T184A-R	GTAGTATGCGCAGGTTCATCAATTGCGTTAA	For Csc2 T184A mutation
Csc2-T186/T187A-F	TACGCATGCTGCCTTTATTCAAGCTGAGGGAGC	For Csc2 T186/T187A mutation
Csc2-T186/T187A-R	GAATAAAGGCAGCATGCGTAGGTTCATCAATTG	For Csc2 T186/T187A mutation
Csc2-E182Q-F	CAATTGATCAACCTACGCATACTACCTTTA	For Csc2 E182Q mutation
Csc2-E182Q-R	TGCGTAGGTTGATCAATTGCGTTAAATGTTC	For Csc2 E182Q mutation
Csc2-D181N/E182Q-F	CGCAATTAATCAACCTACGCATACTACCTTTA	For Csc2 D181N/E182Q mutation
Csc2-D181N/E182Q-R	GCGTAGGTTGATTAATTGCGTTAAATGTTCTCAC	For Csc2 D181N/E182Q mutation

Nucleic acid substrates		
5-6 $ssTS_{CAC}$	ATTTGAATATCTATAGATTGTTCTTTTCTCCTACTTTCCAAATTTCCACAGTCTGCATAA	5-6 ssDNA sequence with CAC in PAM position
5-6 $ssTS_{TTT}$	ATTTGAATATCTATAGATTGTTCTTTTCTCCTACTTTCCAAATTTCTTTAGTCTGCATAA	5-6 ssDNA sequence with TTT in PAM position
4-17 ssTS	TCTGCATTACTAGGAGTTGAAGTAGCAGATACTGATAAAATTTCAGAACAATACATAGCA	4-17 ssDNA sequence
5-6 $dsDNA_{GTG}$ NTS	AGGAAAAAATTAACAAAACTACAAATTATGCAGACTGTGGAAATTTGGAAAGTAGGAGAAAAGAACAATCTATAGATATTCAAATAGTTTTACAAGCTTTAGATATAGATTTACA	5-6 dsDNA NTS sequence with GTG PAM
5-6 $dsDNA_{AAA}$ NTS	AGGAAAAAATTAACAAAACTACAAATTATGCAGACTAAAGAAATTTGGAAAGTAGGAGAAAAGAACAATCTATAGATATTCAAATAGTTTTACAAGCTTTAGATATAGATTTACA	5-6 dsDNA NTS sequence with AAA PAM
5-6 crRNA	AACUAAAACGAAAUUUGGAAAGUAGGAGAAAAGAACAAUCUAUAGAUA	5-6 crRNA sequence
4-17 RNA	AGUUGAAGUAGCAGAUACUGAUAAAAUUUCAGUACAGC	4-17 RNA target sequence
S1 crRNA	AACUAAAACUUCAAAGCUUAGAUACCCUGGAGGGAAACCAGACUUAACA	S1 crRNA sequence
S1 RNA	UGUUAAGUCUGGUUUCCCUCCAGGGUAUCUAAGCUUUGAAAAAAAA	S1 RNA target sequence
S1 ssDNA	ATTTGAATGTTAAGTCTGGTTTCCCTCCAGGGTATCTAAGCTTTGAACACAGTCTGCTAA	S1 ssDNA target sequence

REFERENCES

1. Lin, J., Feng, M., Zhang, H. and She, Q. (2020) Characterization of a novel type III CRISPR-Cas effector provides new insights into the allosteric activation and suppression of the Cas10 DNase. *Cell Discov*, **6**, 29.
2. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, **28**, 2731-2739.
3. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320-324.
4. You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., Xiang, H., Severinov, K., Zhang, X. and Wang, Y. (2019) Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell*, **176**, 239-253 e216.
5. Osawa, T., Inanaga, H., Sato, C. and Numata, T. (2015) Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. *Molecular cell*, **58**, 418-430.
6. Hrle, A., Maier, L.K., Sharma, K., Ebert, J., Basquin, C., Urlaub, H., Marchfelder, A. and Conti, E. (2014) Structural analyses of the CRISPR protein Csc2 reveal the RNA-binding interface of the type I-D Cas7 family. *Rna Biol*, **11**, 1072-1082.
7. Yang, J. and Zhang, Y. (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.*, **43**, W174-181.