Supporting Information

Gong et al. 10.1073/pnas.1410806111

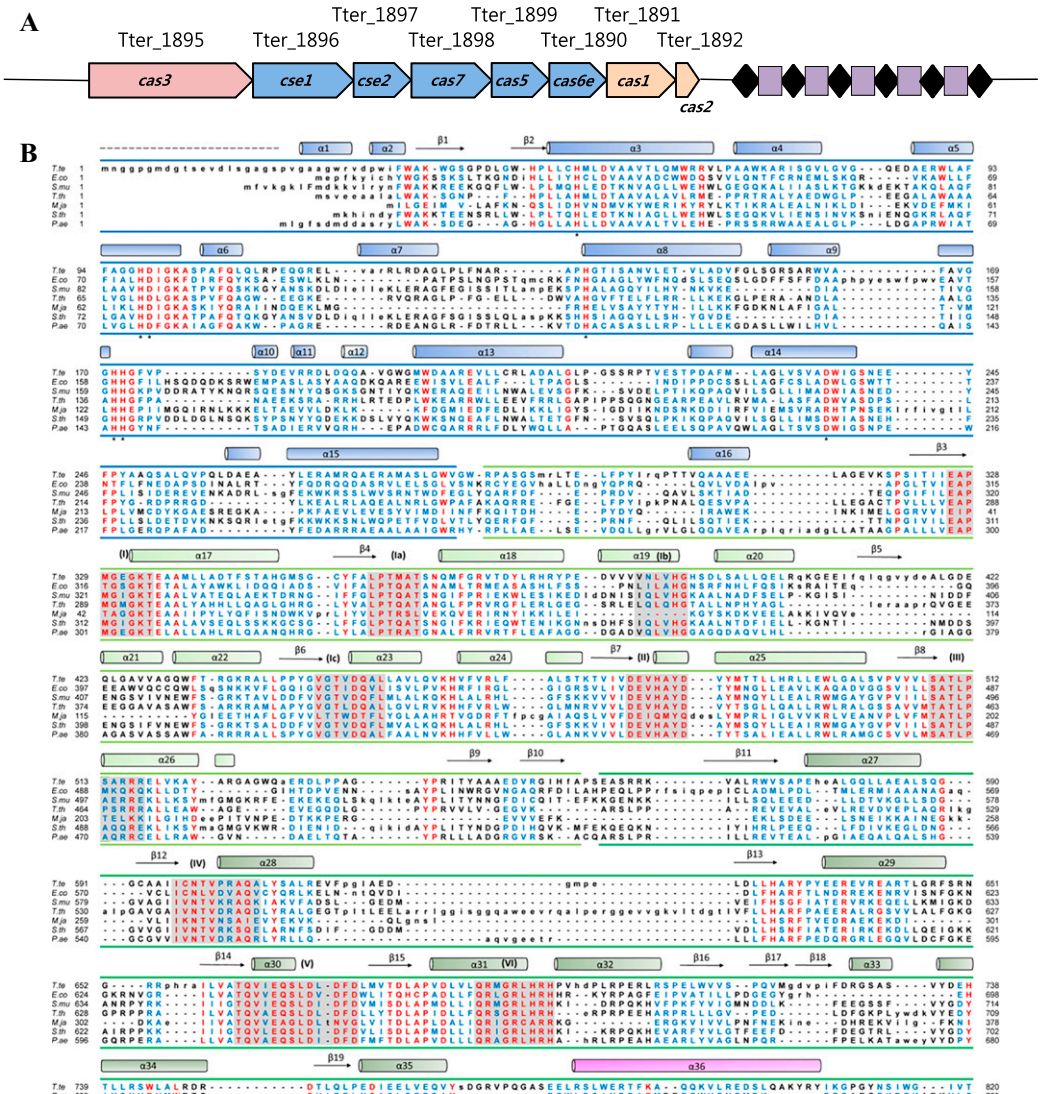


Fig. S1. Type IE Cas3 from *Thermobaculum terrenum*. (*A*) CRISPR-array and type IE *cas* operon on Chromosome 2 of *T. terrenum*. (*B*) Multiple sequence alignment of Cas3 proteins from bacteria and archaea. Each domain is differentiated by alternatively colored bars (corresponding to colors of *Tte*Cas3 domains and RD1-RD2 encompassing α36-helix, used throughout manuscript) above and below the aligned sequences, and strictly conserved residues are highlighted in red columns. Cylinders and arrows above the sequences represent α-helices and β-strands, respectively, and are numbered in order of their appearance in the *Tte*Cas3 structure. Residues that interact with catalytic metal ions and a water molecule are indicated by asterisks, and conserved helicase-forming sequence-motifs are marked (I–VI). Two black-dotted lines above the sequences indicate disordered regions at the N terminus and connecting α36-helix to CTD. *E.co, E. coli* K-12 (gi: 89109548); *M.ja*, *M. jannaschii* DSM 2661 (gi: 1591090 and 1591089); *P.ae*, *Pseudomonas aeruginosa* (gi: 553778515); *S.mu*, *S. mutans NN2025* (gi: 290580129); *S.th*, *S. thermophilus* (gi: 391737982); *T.te, T. terrenum* (gi: 269838931); *T.th*, *T. thermophilus* HB8 (gi: 55978370).

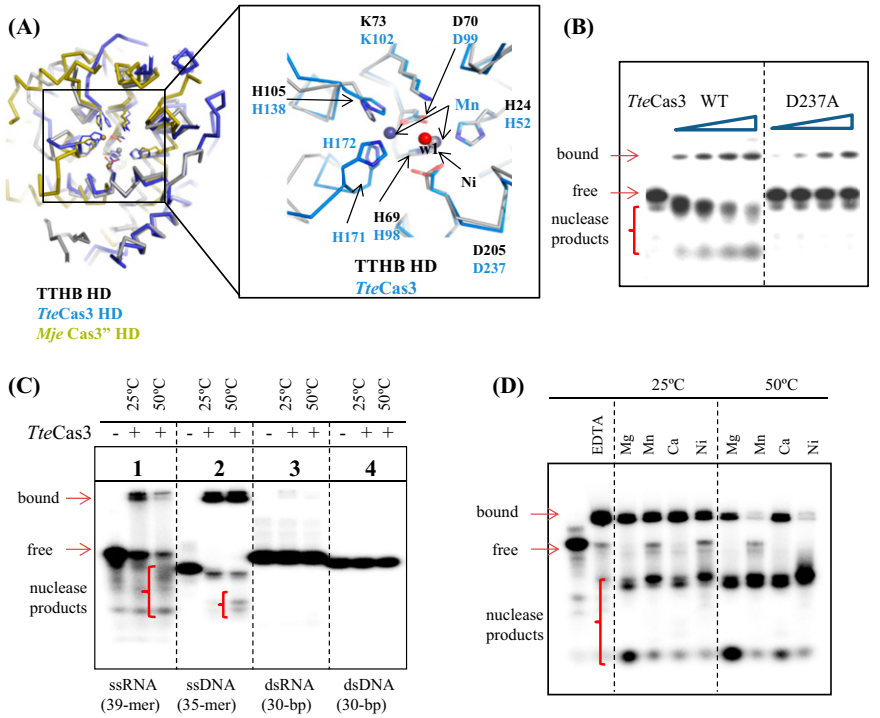


Fig. S2. *Tte*Cas3 exhibits a metal ion-dependent nuclease activity. (*A*) Superposed HD structures. The HD structures of *T. thermophilus* HB8 (TTHB), *T. terrenum* (*Tte*Cas3), and *M. jannachii* are superimposed (*Left*) and their HD cores of *Tte*Cas3 and TTHB are magnified (*Right*). The secondary structures are displayed with coils and differentiated by colors. Conserved residues chelating metal ions and a water molecule, which are displayed by spheres with alternating colors, are depicted as stick models. (*B*) Nuclease degradation of ^{32}P end-labeled 72mer ssDNA substrate (0.1 nM) by *Tte*Cas3 wild type (WT) compared with mutant D237A. Various amounts of proteins (0, 46, 92, 185, and 370 nM) were incubated at 50 °C for 30 min and nuclease products, uncut ssDNA, and Cas3-bound ssDNA were indicated. (*C*) Binding assay with four types of nucleic acids. Wild-type *Tte*Cas3 (370 nM) has been incubated for 30 min at 50 °C with each of the indicated ^{32}P-labeled nucleic acids (0.1 nM). (*D*) ^{32}P-labeled 35-mer ssDNA (0.1 nM) has been incubated with *Tte*Cas3 (370 nM) for 30 min at 25 or 50 °C in the absence or presence of 10 mM $MgCl_2$, $MnCl_2$, $CaCl_2$, $NiCl_2$, or EDTA.

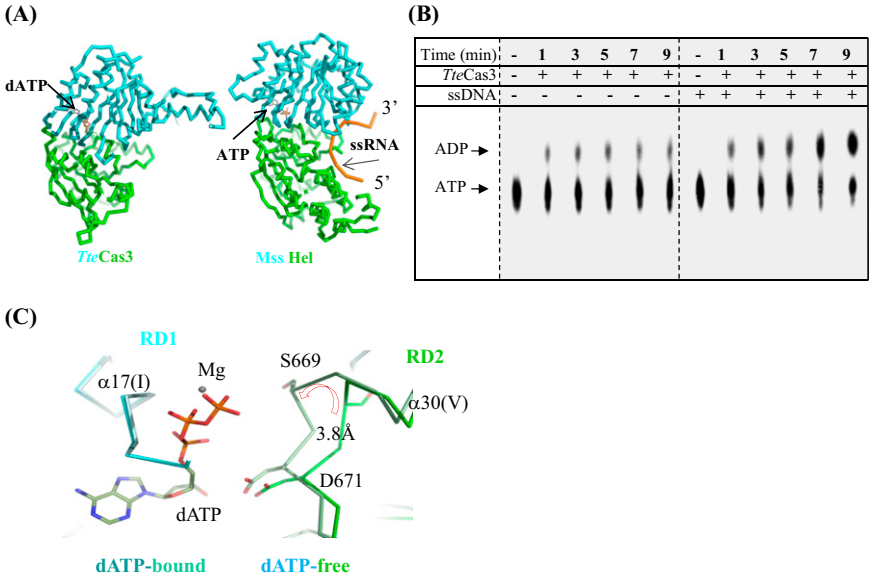
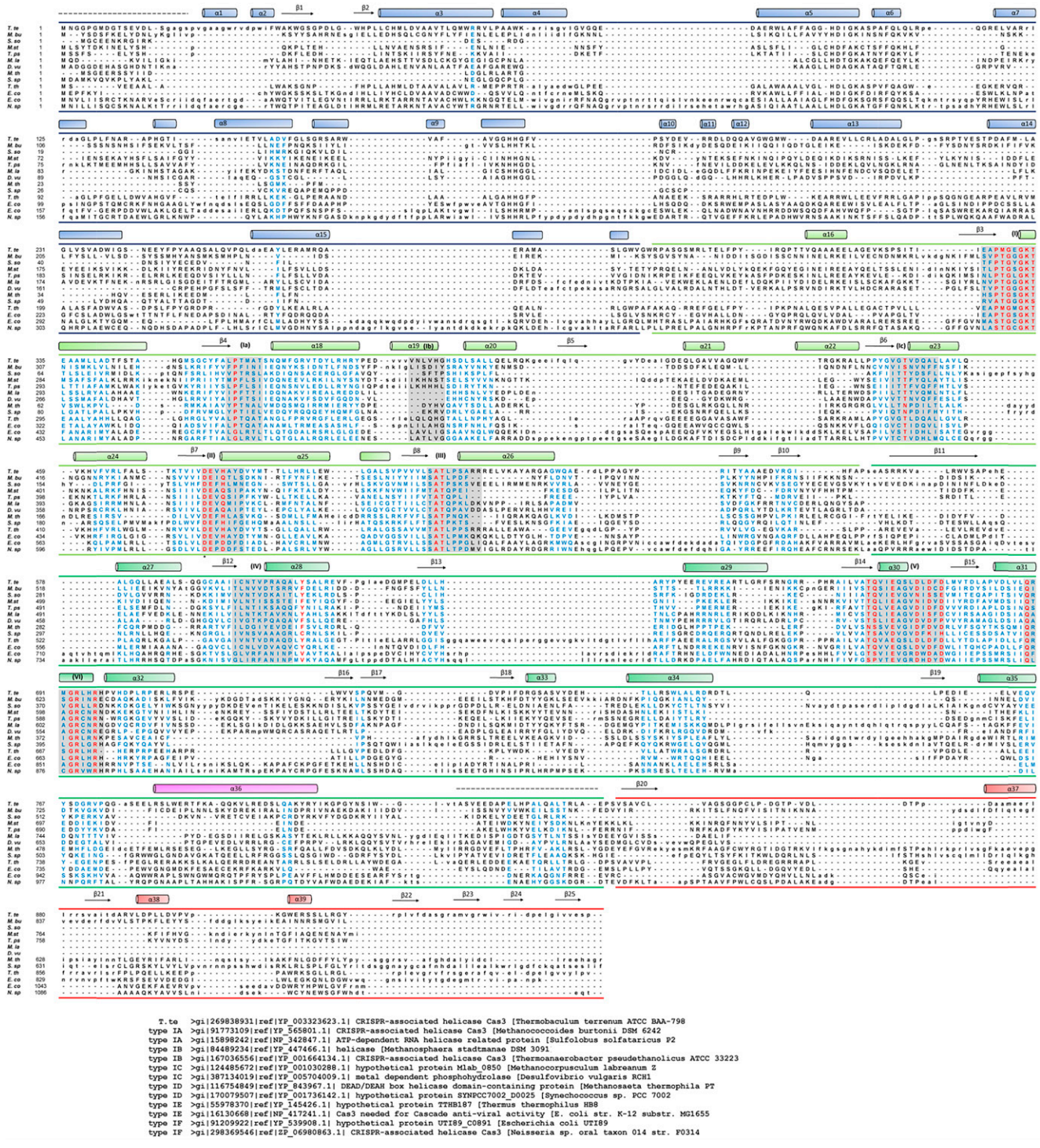


Fig. S3. ATPase site of *Tte*Cas3. (*A*) Structural similarity of *Tte*Cas3 RD domains. *Tte*Cas3 (cyan for RD1 and green for RD2 of *Tte*Cas3; *Left*) superposes well on the yeast Mss166p Helicase (Mss Hel; *Right*). The ssRNA bound to the Mss Hel is shown in a gray ribbon on the right and the ATP bound to Mss Hel is displayed as a stick model. (*B*) Radioactive ATPase assay. For assays, the [a-^{32}P]ATP (0.1 nM) has been incubated with *Tte*Cas3 (46 nM) at 45 °C in the presence of 1 mM $MgCl_2$ alone (*Left*) and with 72-mer ssDNA (*Right*; 2 nM) at the given time. (*C*) Close-up view of the superposed dATP-bound and dATP-free structures. The bound dATP and two residues on the motif V were displayed with stick models. The magnesium ion near the b- and g-phosphate is drawn as a ball.



T.te >gi|269838931|ref|YP_003323623.1| CRISPR-associated helicase Cas3 [Thermobaculum terrenum ATCC BAA-798
type IA >gi|91773109|ref|YP_565801.1| CRISPR-associated helicase Cas3 [Methanococcoides burtonii DSM 6242
type IA >gi|15898242|ref|NP_342847.1| ATP-dependent RNA helicase related protein [Sulfolobus solfataricus P2
type IB >gi|84489234|ref|YP_447466.1| helicase [Methanosphaera stadtmanae DSM 3091
type IB >gi|167036556|ref|YP_001664134.1| CRISPR-associated helicase Cas3 [Thermosinus carboxydivorans Nor1
type IC >gi|224856772|ref|YP_002008308.1| hypothetical protein Mlab_0850 [Methanocorpusculum labreanum Z
type IC >gi|387134019|ref|YP_005704009.1| metal dependent phosphohydrolase [Desulfovibrio vulgaris RCH1
type ID >gi|116754849|ref|YP_843967.1| DEAD/DEAH box helicase domain-containing protein [Methanosaeta thermophila PT
type ID >gi|170079507|ref|YP_001736142.1| hypothetical protein SYNPCC7002_D0025 [Synechococcus sp. PCC 7002
type IF >gi|55978370|ref|YP_145426.1| hypothetical protein TTHB187 [Thermus thermophilus HB8
type IE >gi|16130668|ref|NP_417241.1| Cas3 needed for Cascade anti-viral activity [E. coli str. K-12 substr. MG1655
type IF >gi|91209922|ref|YP_539908.1| hypothetical protein UTI89_C0891 [Escherichia coli UTI89
type IF >gi|298369546|ref|ZP_06980863.1| CRISPR-associated helicase Cas3 [Neisseria sp. oral taxon 014 str. F0314

Fig. S4. Expanded sequence alignment of Cas3 proteins. Two Cas3 proteins from each subtype IA to IF and *Tte*Cas3 were selected and aligned by *Clustal Omega* (www.ebi.ac.uk/Tools/msa/clustalo). The invariant and conserved residues are displayed by red and blue columns, respectively.

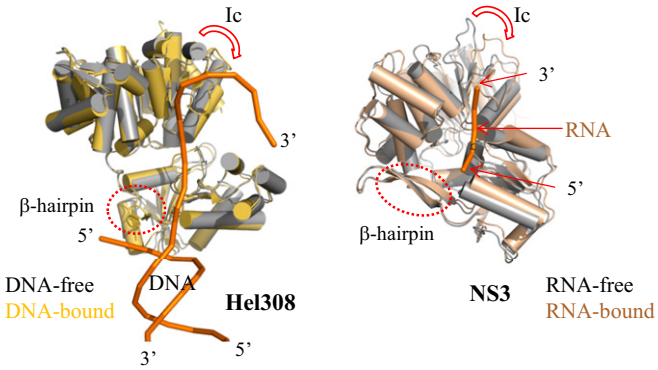


Fig. S5. Rearrangement of helicase RDs upon nucleic acid binding. DNA helicase Hel308 (*Left*) and RNA helicase NS3 (*Right*). The nucleic acid-free and nucleic acid-bound structures were drawn as gray and orange ribbons, respectively.

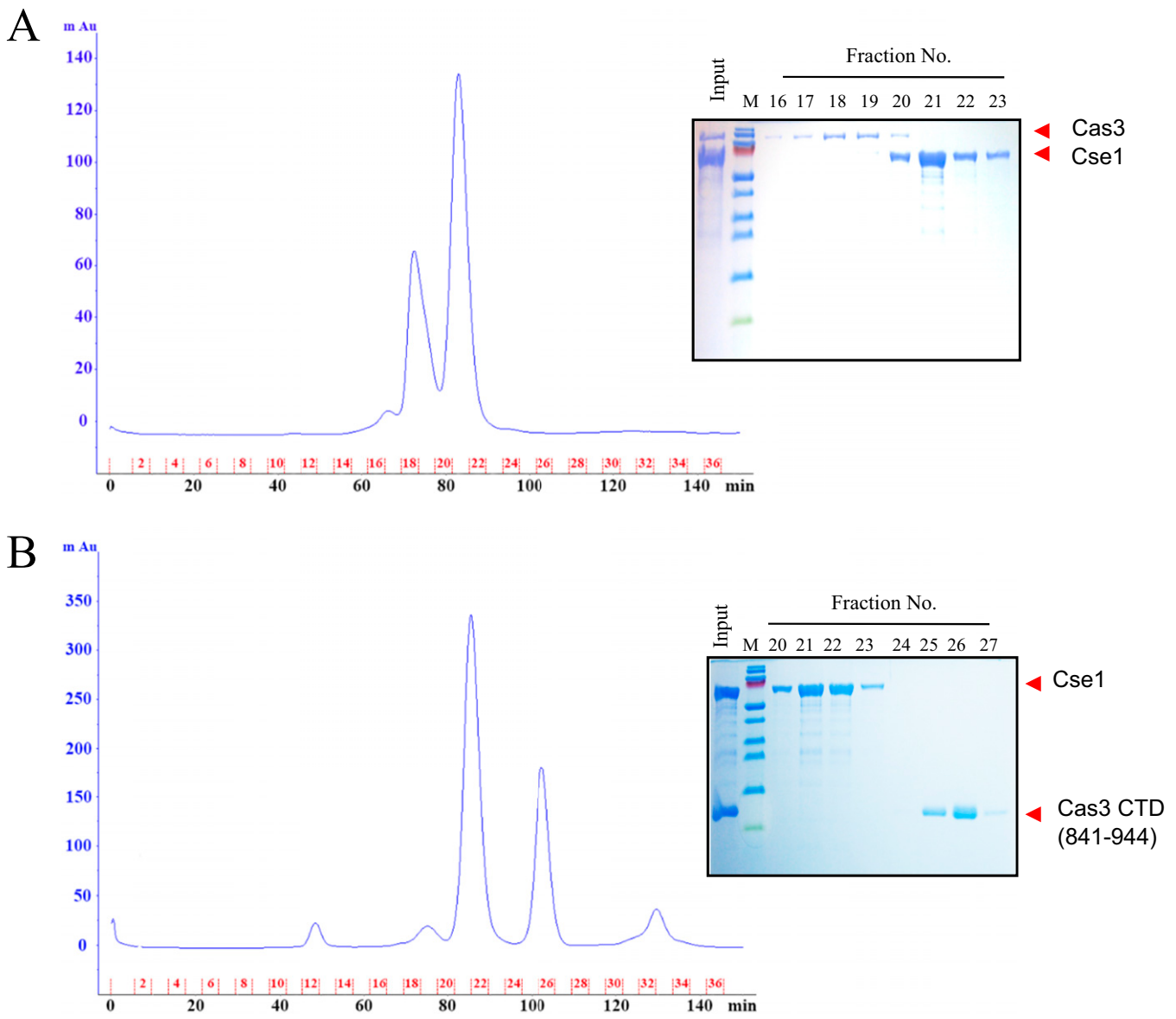


Fig. S6. Analysis of the interaction between *Tte*Cas3 and Cse1 (*A*) and *Tte*Cas3 CTD and Cse1 (*B*) by size exclusion chromatography (SEC). SEC of mixtures was performed with an AKTA Explorer chromatography system by using a Superdex 200 size-exclusion column (26/600; GE Healthcare) in a buffer consisting of 20 mM Tris·HCl at pH 7.5 and 100 mM NaCl at a flow rate of 2.5 mL/min. The chromatograms were obtained by monitoring the absorbance at 280 nm. The fractions were loaded on the 10% (wt/vol) Tricine-SDS/PAGE gel. A set of molecular mass standard markers (Pierce Biotechology) ranging from 10 to 170 kDa was used.

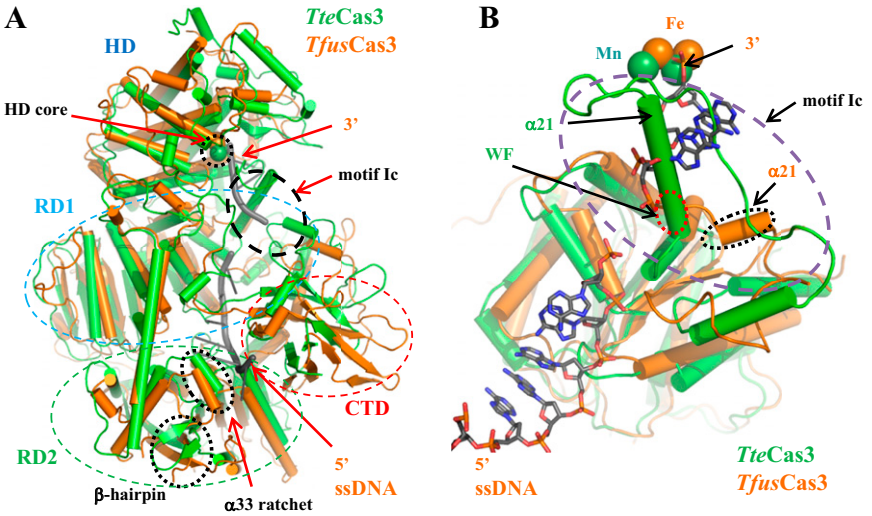


Fig. S7. Comparison of *T. terrenum* Cas3 (*Tte*Cas3) and *Thermobifida fusca* Cas3 (*Tfus*Cas3). (*A*) Overall structure. *Tte*Cas3 (green), *Tfus*Cas3 (orange), and the bound ssDNA (black) to *Tfus*Cas3 are displayed with ribbon diagram. RD1, RD2, and CTD are indicated by dotted circles of alternating colors. Regions involved in catalytic mechanism are indicated by black-dashed circles. Two structures are superposed well with rmsd value of 3.0 Å and noticeable differences are observed in motif Ic of RD1 and orientations of RD2 and CTD domains. (*B*) Magnified view around the Ic motif of RD1. The bound metal ions to each Cas3 protein are displayed with green (*Tte*Cas3) and orange (*Tfus*Cas3), and the bound ssDNA to *Tfus*Cas3 is displayed with stick models. The indicated "a21" of *Tfus*Cas3 is the region that is equivalent to WF-containing helix (a21) of *Tte*Cas3.

Table S1. Oligonucleotide substrates used in this study

Substrate	Sequence
Single-stranded substrate	
ssRNA (39-mer)	5′-AUUGAAAGCA GGAGGGACCG GAAACACACG GUUGAAGGG-3′
ssDNA (35-mer)	5′-GGCGGTGGTG GCGGCATGAA CTACTATCTA TATCC-3′
ssDNA (72-mer)	5′-GATCGGTCTC TACTACATAT CTATAAACGT GCAAGGCAGC GTCAACAAAC TCCATATGAA TATCCTCCTT AG-3′
Duplex substrate	
dsRNA (30-bp)	5′-CUCUACGACA UCGGAUCCGA UGUCGUAGAG-3′
dsDNA (30-bp)	5′-GGCGAGTTTG TTAACCATGC CTTGCACGTT-3′

Table S2. Primer sequences used for site-directed mutagenesis

Mutation	Sequence
D237A	5′– GTC TCC GTG GCG GCC TGG ATA GGC TCG –3′
	5′– CGA GCC TAT CCA GGC CGC CAC GGA GAC –3′
D477A	5′– ACC GTG ATC GTT GCT GAG GTC CAT GCT–3′
	5′– AGC ATG GAC CTC AGC AAC GAT CAC GGT –3′
W432A	5′– GTG GCA GGG CAA GCG TTC ACG CGC GGC – 3′
	5′– GCC GCG CGT GAA CGC TTG CCC TGC CAC – 3′
F433A	5′– GCA GGG CAA TGG GCC ACG CGC GGC AAG – 3′
	5′– CTT GCC GCG CGT GGC CCA TTG CCC TGC – 3′
W432A-F433A	5′– GCA GGG CAA GCG GCC ACG CGC GGC AAG – 3′
	5′– CTT GCC GCG CGT GGC CGC TTG CCC TGC – 3′

Underlines indicate the mutated sequences.