# Supplementary Figures

## Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system

Tomas Sinkunas[1], Giedrius Gasiunas[1], Christophe Fremaux[2], Rodolphe Barrangou[3], Philippe Horvath[2], and Virginijus Siksnys[1*]

[1]*Institute of Biotechnology, Vilnius University, Graiciuno 8, LT-02241, Vilnius, Lithuania*

[2] *Danisco France SAS, BP 10, F-86220 Dangé-Saint-Romain, France*

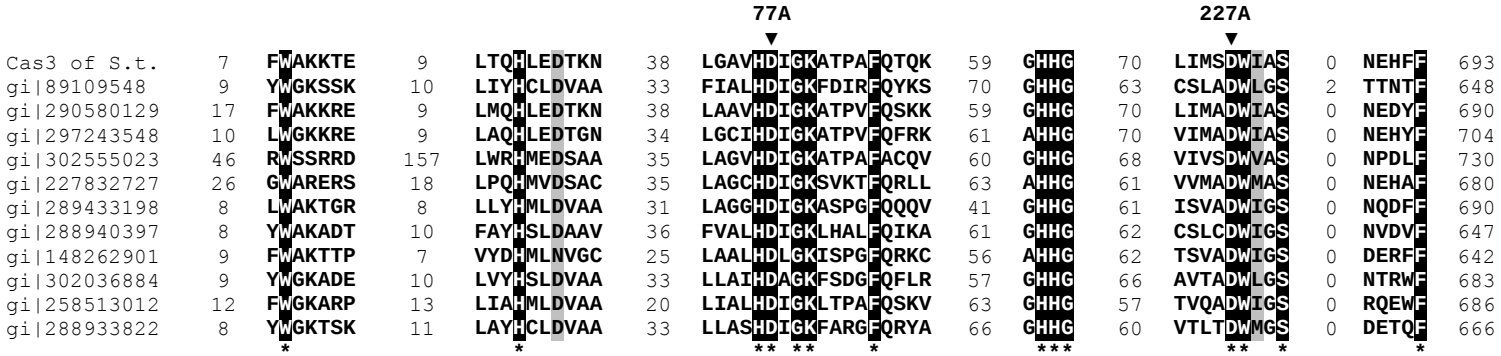
[3]*Danisco USA Inc., 3329 Agriculture Drive, Madison, Wisconsin 53716, USA.*

[*] Corresponding author

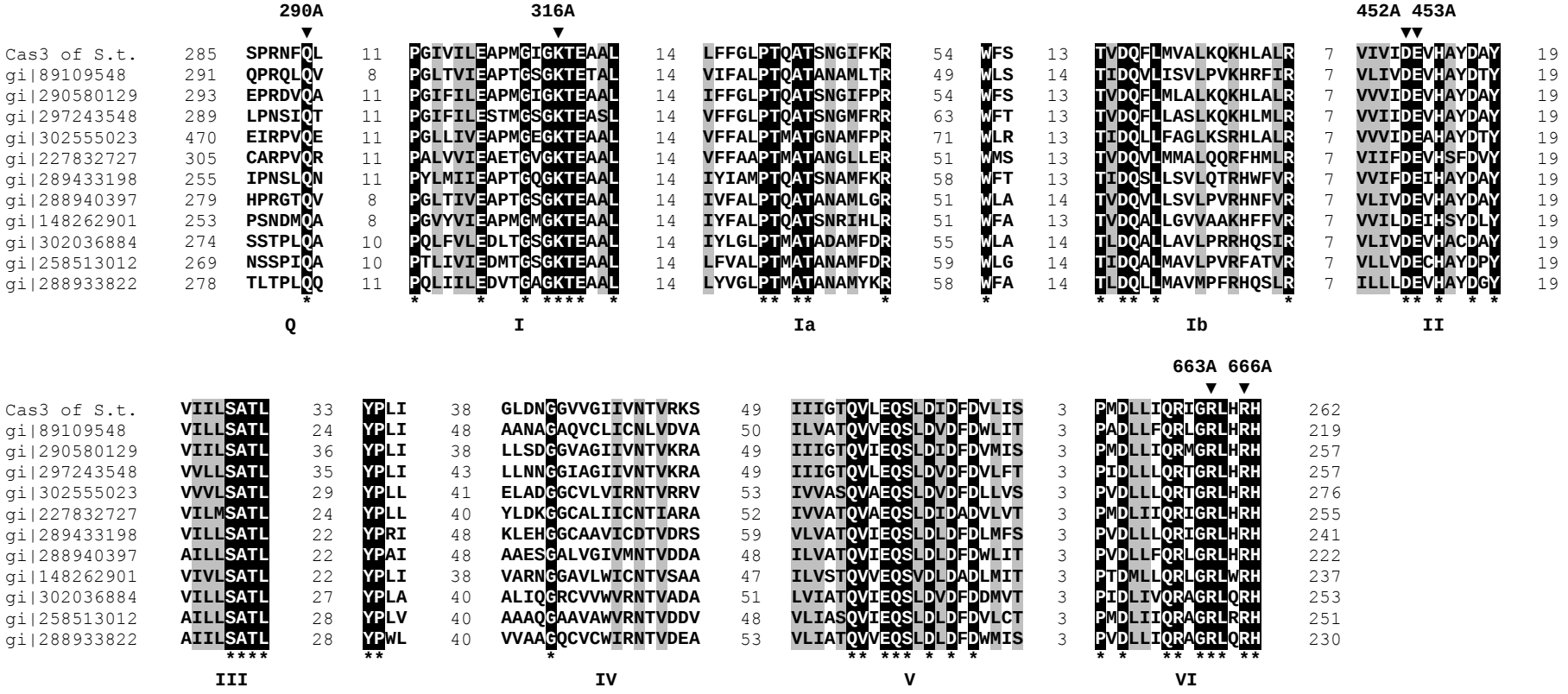
Mail: Institute of Biotechnology, Graiciuno 8, Vilnius LT-02241, Lithuania

Tel: +370-5-2602108; Fax: +370-5-2602116; E-mail: siksnys@ibt.lt

**A**

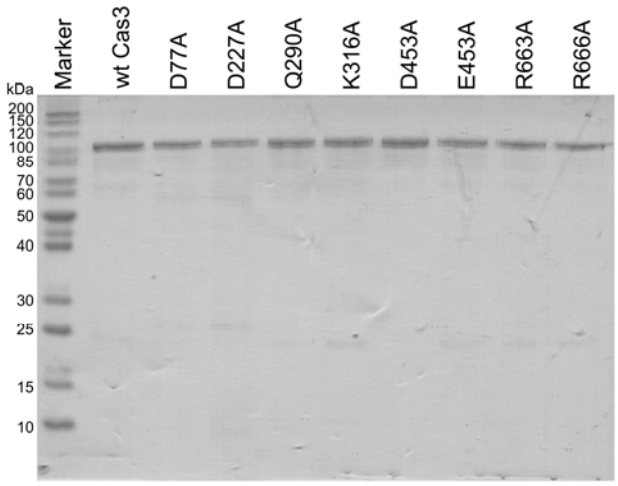

77A
227A
Cas3 of S.t. 7 FWAKKTE 9 LTQHLEDTKN 38 LGAVHDIGKATPAFQTQK 59 GHHG 70 LIMSDWIAS 0 NEHFF 693
gi|89109548 9 YWGKSSK 10 LIYHCLDVAA 33 FIALHDIGKFDIRFQYKS 70 GHHG 63 CSLADWLGS 2 TTNTF 648
gi|290580129 17 FWAKKRE 9 LMQHLEDTKN 38 LAAVHDIGKATPVFQSKK 59 GHHG 70 LIMADWIAS 0 NEDYF 690
gi|297243548 10 LWGKKRE 9 LAQHLEDTGN 34 LGCIHDIGKATPVFQFRK 61 AHHG 70 VIMADWIAS 0 NEHYF 704
gi|302555023 46 RWSSRRD 157 LWRHMEDSAA 35 LAGVHDIGKATPAFACQV 60 GHHG 68 VIVSDWVAS 0 NPDLF 730
gi|227832727 26 GWARERS 18 LPQHMVDSAC 35 LAGCHDIGKSVKTFQRLL 63 AHHG 61 VVMADWMAS 0 NEHAF 680
gi|289433198 8 LWAKTGR 8 LLYHMLDVAA 31 LAGGHDIGKASPGFQQQV 41 GHHG 61 ISVADWIGS 0 NQDFF 690
gi|288940397 8 YWAKADT 10 FAYHSLDAAV 36 FVALHDIGKLHALFQIKA 61 GHHG 62 CSLCDWIGS 0 NVDVF 647
gi|148262901 9 FWAKTTP 7 VYDHMLNVGC 25 LAALHDLGKISPGFQRKC 56 AHHG 62 TSVADWIGS 0 DERFF 642
gi|302036884 9 YWGKADE 10 LVYHSLDVAA 33 LLAIHDAGKFSDGFQFLR 57 GHHG 66 AVTADWLGS 0 NTRWF 683
gi|258513012 12 FWGKARP 13 LIAHMLDVAA 20 LIALHDIGKLTPAFQSKV 63 GHHG 57 TVQADWIGS 0 RQEWF 686
gi|288933822 8 YWGKTSK 11 LAYHCLDVAA 33 LLASHDIGKFARGFQRYA 66 GHHG 60 VTLTDWMGS 0 DETQF 666

**B**

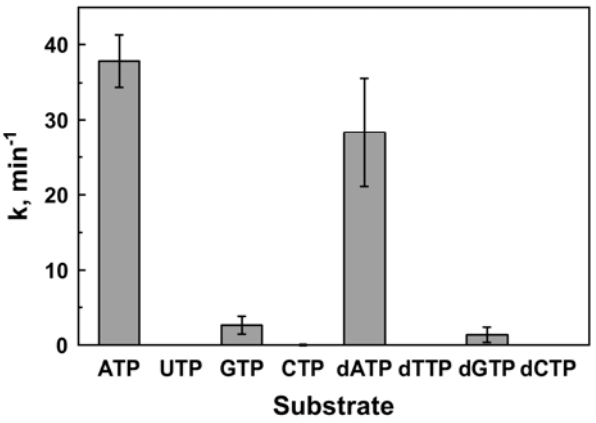

290A
316A
452A 453A
Cas3 of S.t. 285 SPRNFQL 11 PGIVILEAPMGIGKTEAAL 14 LFFGLPTQATSNGIFKR 54 WFS 13 TVDQFLMVALKQKHLALR 7 VIVIDEVHAYDAY 19
gi|89109548 291 QPRQLQV 8 PGLTVIEAPTGSGKTETAL 14 VIFALPTQATANAMLTR 49 WLS 14 TIDQVLISVLPVKHRFIR 7 VLIVDEVHAYDTY 19
gi|290580129 293 EPRDVQA 11 PGIFILEAPMGIGKTEAAL 14 IFFGLPTQATSNGIFPR 54 WFS 13 TVDQFLMLALKQKHLALR 7 VVVIDEVHAYDAY 19
gi|297243548 289 LPNSIQT 11 PGIFILESTMGSGKTEASL 14 VFFGLPTQATSNGMFRR 63 WFT 13 TVDQFLLASLKQKHLMLR 7 VVIIDEVHAYDAY 19
gi|302555023 470 EIRPVQE 11 PGLLIVEAPMGEGKTEAAL 14 VFFALPTMATGNAMFPR 71 WLR 13 TIDQLLFAGLKSRHLALR 7 VVVIDEAHAYDTY 19
gi|227832727 305 CARPVQR 11 PALVVIEAETGVGKTEAAL 14 VFFAAPTMATANGLLER 51 WMS 13 TVDQVLMMALQQRFHMLR 7 VIIFDEVHSFDVY 19
gi|289433198 255 IPNSLQN 11 PYLMIIEAPTGQGKTEAAL 14 IYIAMPTQATSNAMFKR 58 WFT 13 TIDQSLLSVLQTRHWFVR 7 VVIFDEIHAYDAY 19
gi|288940397 279 HPRGTQV 8 PGLTIVEAPTGSGKTEAAL 14 IVFALPTQATANAMLGR 51 WLA 14 TVDQVLLSVLPVRHNFVR 7 VLIVDEVHAYDAY 19
gi|148262901 253 PSNDMQA 8 PGVYVIEAPMGMGKTEAAL 14 IYFALPTQATSNRIHLR 51 WFA 13 TVDQALLGVVAAKHFFVR 7 VVILDEIHSYDLY 19
gi|302036884 274 SSTPLQA 10 PQLFVLEDLTGSGKTEAAL 14 IYLGLPTMATADAMFDR 55 WLA 14 TLDQALLAVLPRRHQSIR 7 VLIVDEVHACDAY 19
gi|258513012 269 NSSPIQA 10 PTLIVIEDMTGSGKTEAAL 14 LFVALPTMATANAMFDR 59 WLG 14 TIDQALMAVLPVRFATVR 7 VLLVDECHAYDPY 19
gi|288933822 278 TLTPLQQ 11 PQLIILEDVTGAGKTEAAL 14 LYVGLPTMATANAMYKR 58 WFA 14 TLDQLLMAVMPFRHQSLR 7 ILLLDEVHAYDGY 19
Q
I
Ia
Ib
II


663A 666A
Cas3 of S.t. VIILSATL 33 YPLI 38 GLDNGGVVGIIVNTVRKS 49 IIIGTQVLEQSLDIDFDVLIS 3 PMDLLIQRIGRLHRH 262
gi|89109548 VILLSATL 24 YPLI 48 AANAGAQVCLICNLVDVA 50 ILVATQVVEQSLDVDFDWLIT 3 PADLLFQRLGRLHRH 219
gi|290580129 VIILSATL 36 YPLI 38 LLSDGGVAGIIVNTVKRA 49 IIIGTQVIEQSLDIDFDVMIS 3 PMDLLIQRMGRLHRH 257
gi|297243548 VVLLSATL 35 YPLI 43 LLNNGGIAGIIVNTVKRA 49 IIIGTQVLEQSLDVDFDVLFT 3 PIDLLLQRTGRLHRH 257
gi|302555023 VVVLSATL 29 YPLL 41 ELADGGCVLVIRNTVRRV 53 IVVASQVAEQSLDVDFDLLVS 3 PVDLLLQRTGRLHRH 276
gi|227832727 VILMSATL 24 YPLL 40 YLDKGGCALIICNTIARA 52 IVVATQVAEQSLDIDADVLVT 3 PMDLIIQRIGRLHRH 255
gi|289433198 VILLSATL 22 YPRI 48 KLEHGGCAAVICDTVDRS 59 VLVATQVIEQSLDLDFDLMFS 3 PVDLLLQRIGRLHRH 241
gi|288940397 AILLSATL 22 YPAI 48 AAESGALVGIVMNTVDDA 48 ILVATQVIEQSLDLDFDWLIT 3 PVDLLFQRLGRLHRH 222
gi|148262901 VIVLSATL 22 YPLI 38 VARNGGAVLWICNTVSAA 47 ILVSTQVVEQSVDLDADLMIT 3 PTDMLLQRLGRLWRH 237
gi|302036884 VILLSATL 27 YPLA 40 ALIQGRCVVWVRNTVADA 51 LVIATQVIEQSLDVDFDDMVT 3 PIDLIVQRAGRLQRH 253
gi|258513012 AILLSATL 28 YPLV 40 AAAQGAAVAWVRNTVDDV 48 VLIASQVIEQSLDLDFDVLCT 3 PMDLIIQRAGRLRRH 251
gi|288933822 AIILSATL 28 YPWL 40 VVAAGQCVCWIRNTVDEA 53 VLIATQVVEQSLDLDFDWMIS 3 PVDLLIQRAGRLQRH 230
III
IV
V
VI

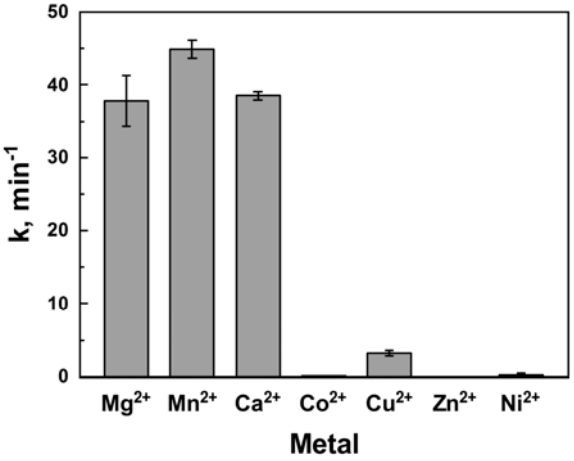
**Supplementary Figure 1.** ***Multiple alignment of the conserved regions in the (A) HD and (B) DExD/H domains of Cas3 protein sequence.*** Alignments were assemled using “COBALT” (Papadopoulos & Agarwala, 2007) tool and visualised with “BioEdit 7.0.5” (Hall, 1999). The numbers preceding and following the alignment indicate the number of amino acid residues between the protein termini and the proximal and distal aligned blocks, respectively; those between individual blocks indicate the number of residues separating them. Homologous sequences with different similarity to Cas3 from *Streptococcus thermophilus* DGCC7710 were used for construction of alignment. GeneBank identification numbers of proteins are given in the left-hand column. Stars below the columns (black-shaded) indicate conservative amino acid residues. Grey-shaded columns indicate similar amino acid recidues. Positions of point mutations to alanin which were made in this work are indicated above the alignment columns. Roman numbers below the alignment columns in the figure B indicate regions of conservative motifs of DExD/H helicases.



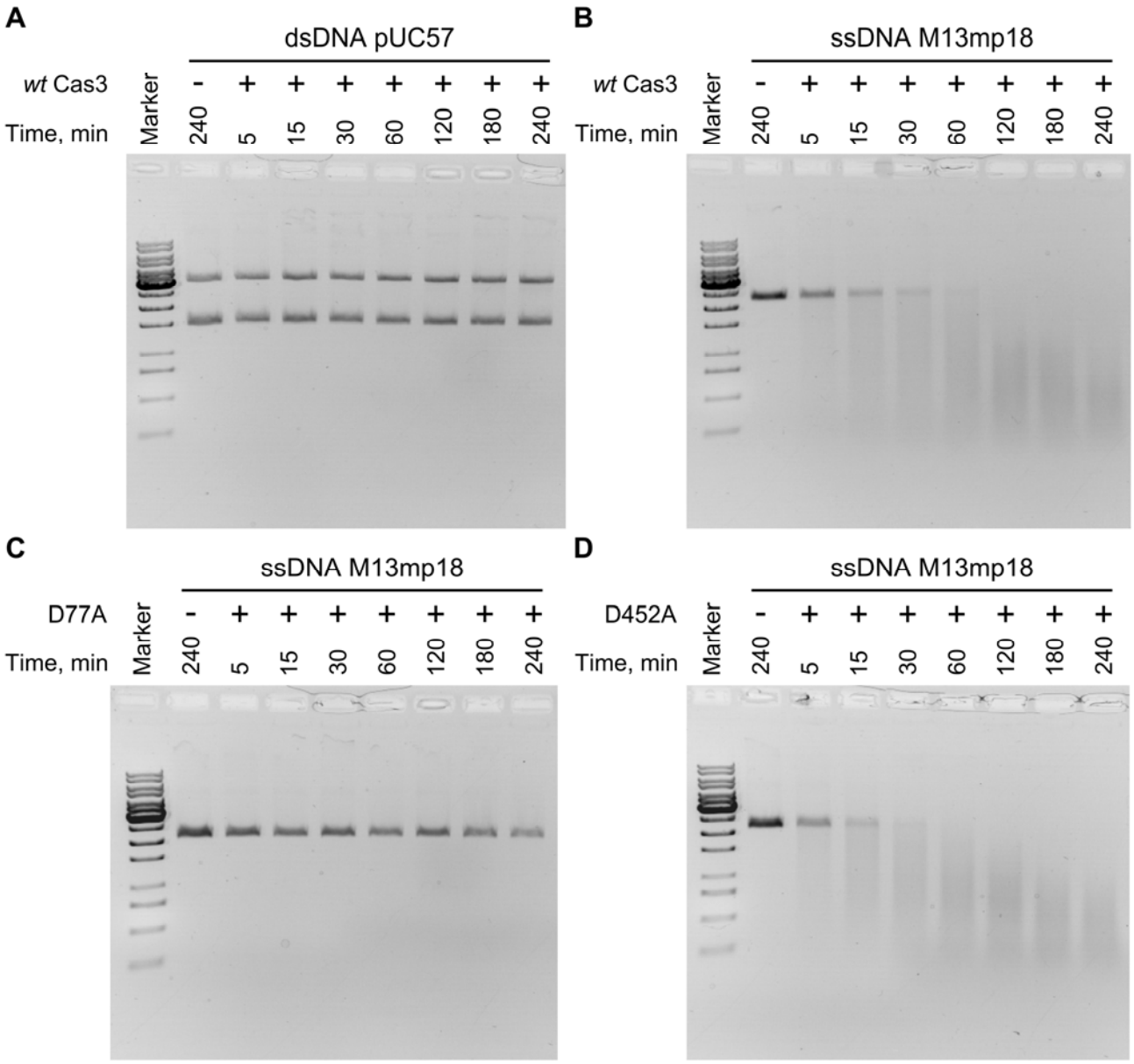

**Supplementary Figure 2.** *SDS-PAGE of the purified Cas3 proteins and its mutants.* Marker – protein molecular marker “PageRuler Unstained Protein Ladder” (Fermentas, Vilnius, Lithuania). The numbers shown on the left indicate markers’ size in kDa. Samples corresponding to 750 ng of each protein were loaded on the gel.




**Supplementary Figure 3.** ***Cas3 ATPase activity dependence on nucleotides and deoxynucleotides.*** ATPase reactions were conducted at 30 °C in a reaction buffer containing 10 mM Tris–HCl (pH 7.5 at 25C), 30 mM KCl, 5% glycerol, 2 mM $MgCl_2$, 0.1 mg/ml of BSA, and 3.0 nM of ssDNA (M13mp18), 250 nM of Cas3 and 500 μM of various nucleotides and deoxynucleotides. Malachite green assay was used to measure ATP hydrolysis through the detection of liberated free phosphate from ATP. Reaction rate constant $k$ ($min^{-1}$) calculated from the linear slopes of respective times courses.




**Supplementary Figure 4.** ***Cas3 ATPase activity dependence on divalent metal ions.*** ATPase reactions were conducted at 30 °C in a reaction buffer containing 10 mM Tris–HCl (pH 7.5 at 25 °C), 30 mM KCl, 5% glycerol, 0.1 mg/ml of BSA, and 3.0 nM of ssDNA (M13mp18), 250 nM of Cas3 and 2 mM of different divalent metal ions. Malachite green assay was used to measure ATP hydrolysis through the detection of liberated free phosphate from ATP. Reaction rate constant $k$ ($min^{-1}$) calculated from the linear slopes of respective times courses.




**Supplementary Figure 5.** *Time courses of nuclease activity of Cas3 and mutant proteins.* Nuclease reactions were conducted at 37 °C in a reaction buffer containing 10 mM Tris–HCl (pH 7.5 at 25 °C), 60 mM KCl, 10% glycerol (v/v), 10 mM $MgCl_2$, 0.1 mg/ml of BSA, and 4 nM of dsDNA supercoiled form of pUC57 plasmid (**A**) or ssDNA of M13mp18 (**C-D**), and 500 nM Cas3 (**A-B**) or mutant proteins (D77A – **C**, D452A – **D**). Time points of the reactions indicated above the figures. Marker – DNA molecular marker "GeneRuler™ 1 kb DNA Ladder" (Fermentas, Vilnius, Lithuania).

**Supplementary references**

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41:** 95-98

Papadopoulos JS, Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**(9)**:** 1073-1079