Supplementary Data

Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation

Olga Musharova, Evgeny Klimuk, Kirill A. Datsenko, Anastasia Metlitskaya, Maria Logacheva,

Ekaterina Semenova, Konstantin Severinov, and Ekaterina Savitskaya

Table of content:

Table S1. Oligonucleotides used in the study.

Name	Sequence (5'-3')	Description/site of annealing
PCR		
P4518	AAGGTTGGTGTCTTTTTTAC	CRISPR locus in KD263, KD454
P4581	GTCGCTGCCGTGACGTTATG	CRISPR locus in KD263, KD454
Ec_II for	AACATAATGGATGTGTTGTTTGTG	CRISPR II locus in BW40297
Ec_II rev	GAAATGCTGGTGAGCGTTAATG	CRISPR II locus in BW40297
HS1-check for	CATCATTGGAAAACGTTCTTC	used for amplification of HS1 spacer incorporated inside CRISPR array
HS2 for	GTCATTCTGAGAATAGT	used for amplification of HS2 spacer incorporated inside CRISPR array
qPCR		
Host for	TCGGTTCGGGTGGTGCA	a genomic locus in KD263
Host rev	TGACGTCCAGATACAC	a genomic locus in KD263
Leader for	GGAACTCTCTAAAAGTATACATTTGTTC	amplifies a 138-bp CRISPR leader region, also, together with primer P4581, was used to check for CRISPR array HS1 and HS2 spacers incorporated in CRISPR array (235 bp amplicons)
Leader rev	GGTTGGTGTCTTTTTTACCTGTTTG	amplifies a 138-bp region of CRISPR leader
CS for	GTAACCCACTCGTGCA	amplifies 33-bp CS
Leader_FaiI for	GCTTTAAGAACAAATGT	amplifies 34-bp CRISPR leader region
Leader_FaiI rev	GGAACTCTCTAAAAGTA	amplifies 34-bp CRISPR leader region
CS rev	GCTGAAGATCAGTTGGG	amplifies 33-bp CS
PS for	GTTGTCTTTCGCTGCTG	amplifies 33-bp G8 protospacer
PS rev	GCGGGATCGTCACC	amplifies 33-bp G8 protospacer
HS1 for	GTGCTCATCATTGGAAAA	amplifies 33-bp HS1
HS1 rev	TCGCCCCGAAGAACG	amplifies 33-bp HS1
HS1_long for	GCAGAACTTTAAAAGTGC	amplifies a 61-bp fragment containing HS1; also used as Fr 210 rev
HS1_long rev	ATCCTTGAGAGTTTTCG	amplifies a 61-bp fragment containing HS1
Fr 210 for	TCGCCCTTATTCCCTTT	amplifies a 210-bp fragment of pG8mut or pG8mut_CCG plasmids
HS2 for	GTCATTCTGAGAATAGT	amplifies a 33-bp fragment containing HS2
HS2 rev	CTCGGTCGCCGCATAC	amplifies a 33-bp fragment containing HS2
HS2_long for	GGTGAGTACTCAACCAA	amplifies a 68-bp fragment containing HS2

HS2_long rev	CGGGCAAGAGCAACT	amplifies a 68-bp fragment containing HS2
Gyr for	CGGTCAACATTGAGGAAGAGC	amplifies a 190-bp *gyr* fragment which was used for normalization
Gyr rev	TACGTCACCAACGACACGG	amplifies a 190-bp *gyr* fragment which was used for normalization
generation of model substrates		
HS1_full/ cmp for	GTGCTCATCATTGGAAAACGTTCTTCGGGGCGA	double-stranded 33-bp HS1
HS1_full/ cmp rev	TCGCCCCGAAGAACGTTTTCCAATGATGAGCAC	double-stranded 33-bp HS1
HS1_part/ cmp for	CATCATTGGAAAACGTTCTTCGGGGCGA	splayed HS1 containing 23-bp internal duplex
HS1_part/ cmp rev	CCGAAGAACGTTTTCCAATGATGAGCAC	splayed HS1 containing 23-bp internal duplex
HS2_full/ cmp for	CTCGGTCGCCGCATACACTATTCTCAGAATGAC	double-stranded 33-bp HS2
HS2_full/ cmp rev	GTCATTCTGAGAATAGTGTATGCGGCGACCGAG	double-stranded 33-bp HS2
site-specific mutagenesis		
HS1_CCG for	CCAATGATGAGCACGGTTAAAGTTCTGCTAT	used to generate altered PAM in HS1 of pG8mut_CCG
HS1_CCG rev	ATAGCAGAACTTTAACCGTGCTCATCATTGG	used to generate altered PAM in HS1 of pG8mut_CCG
oligonucleotides used for primer extension		
HS1 for pr/ext	GTCATTCTGAGAATAGT	HS1
HS1 rev pr/ext	ACATCGAACTGGATCTC	HS1
HS2 for pr/ext	CATCCGTAAGATGCTTTTC	HS2
HS2 rev pr/ext	TGGCGCGGTATTATCC	HS2

Table S2. Amplicons produced in qPCR reactions.

Amplicon	Oligonucleotide pair used for amplification	Length, bp	Confirmed by Sanger sequencing YES/NO
HS1 33bp	HS1 for-HS1 rev	33	YES
HS1 47bp	HS1 for-HS1 long_rev	47	YES
HS1 47bp*	HS1 long_for-HS1 rev	47	YES
HS1 61bp	HS1 long_for-HS1 long_rev	61	YES
HS2 33bp	HS2 for-HS2 rev	33	YES
HS2 46bp	HS2 long_for– HS2 rev	46	YES
HS2 50bp	HS2 for – HS2 long_rev	50	YES
HS2 61bp	HS2 long_for– HS2 long_rev	61	YES
Leader 34 bp	Leader *Fai*I for-Leader *Fai*I rev	34	YES
Leader 138 bp	Leader for - Leader rev	138	YES
CS	CS for - CS rev	33	YES
PS	PS for – PS rev	33	YES
Host	Host for – Host rev	33	YES
210 bp fragment	Fr 210 for – HS1_long for	210	YES
Gyr	Gyr for – Gyr rev	190	YES

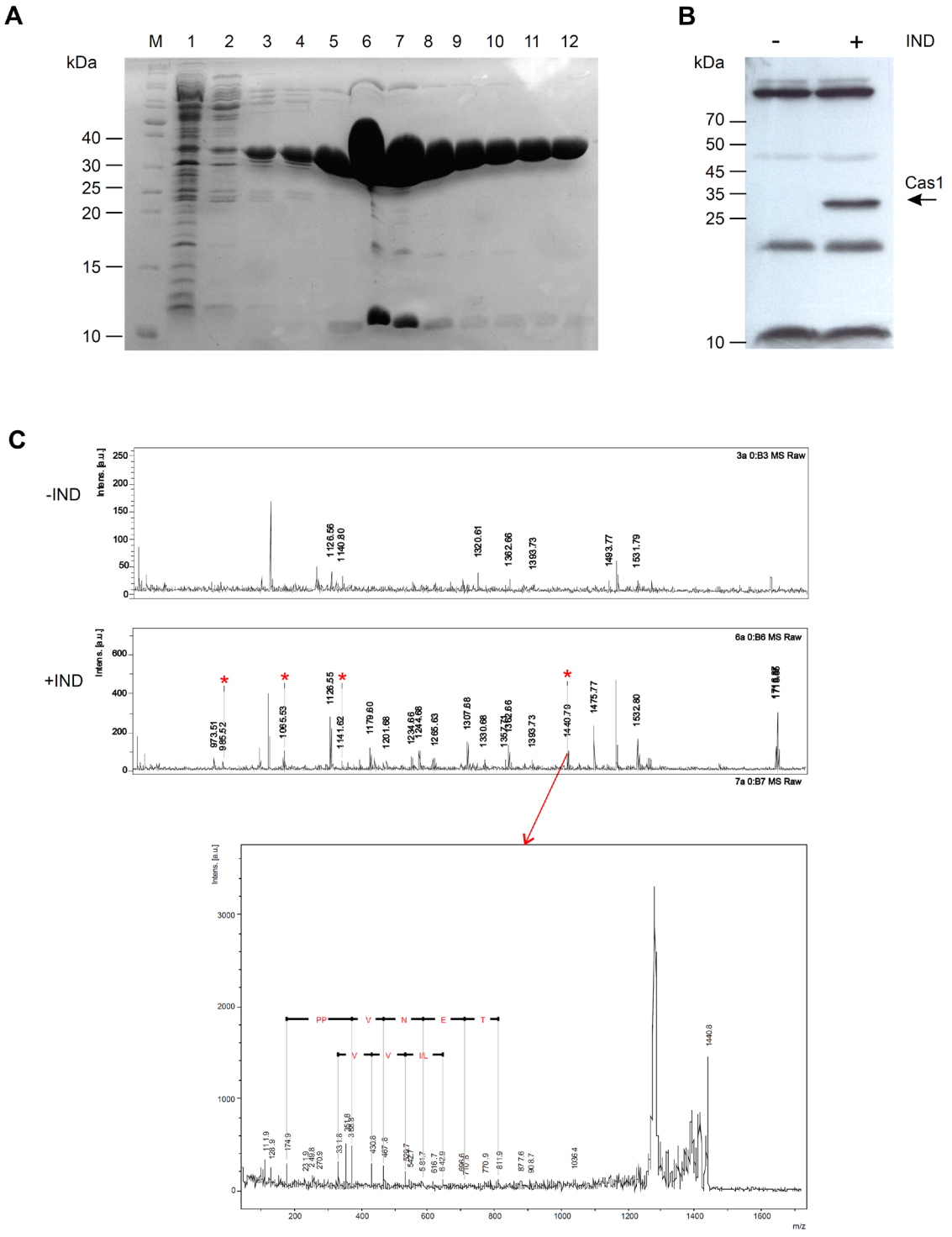

Figure S1.

A. SDS gel showing final stages of affinity purification of hexahistidine-tagged Cas1 and untagged Cas2 from lysates of co-overexpressing cells. Fractions 6 and 7 were pooled and used for immunization.

B. Western blotting of proteins from KD263 cultures in the presence or in the absence of *cas* gene expression inducers with purified anti-Cas1 polyclonal antibody.

C. Purified anti-Cas1 polyclonal antibody was used to precipiate proteins from induced and uninduced *E. coli* cells. Precipiated material was reslolved by SDS PAGE and gel slices containing material with apparent mobility of Cas2 (~11 kDa) were subjected to in-gel tryptic digestion and MALDI-TOF mass-spectrometry. Mass-peaks labeled with red asterisks match tryptic fragments of Cas2. The results of MS-MS analysis of one such peak (m/z=1440.7) are shown below. The deduced sequence matches that of N-terminal fragment of Cas2 with Met^1 removed.

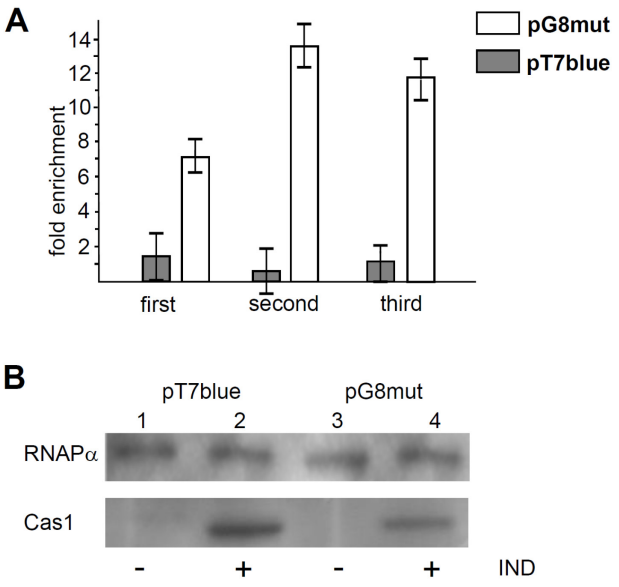

Figure S2.

A. Results of Cas1 ChIP analysis for 138 bp amplicon. Comparison of three biological replicates. Data labeled "first" are also shown in Fig. 1C. Data labeled "second" and "third" are from independent replicates.

B. The results of Western blotting of proteins from KD263 cultures transformed with indicated plasmids and grown in the presence or in the absence of *cas* gene expression inducers with anti-Cas1 polyclonal antibody used for ChIP experiments (below) and with anti-RNA polymerase α subunit antibodies (used as loading control, above) are presented.

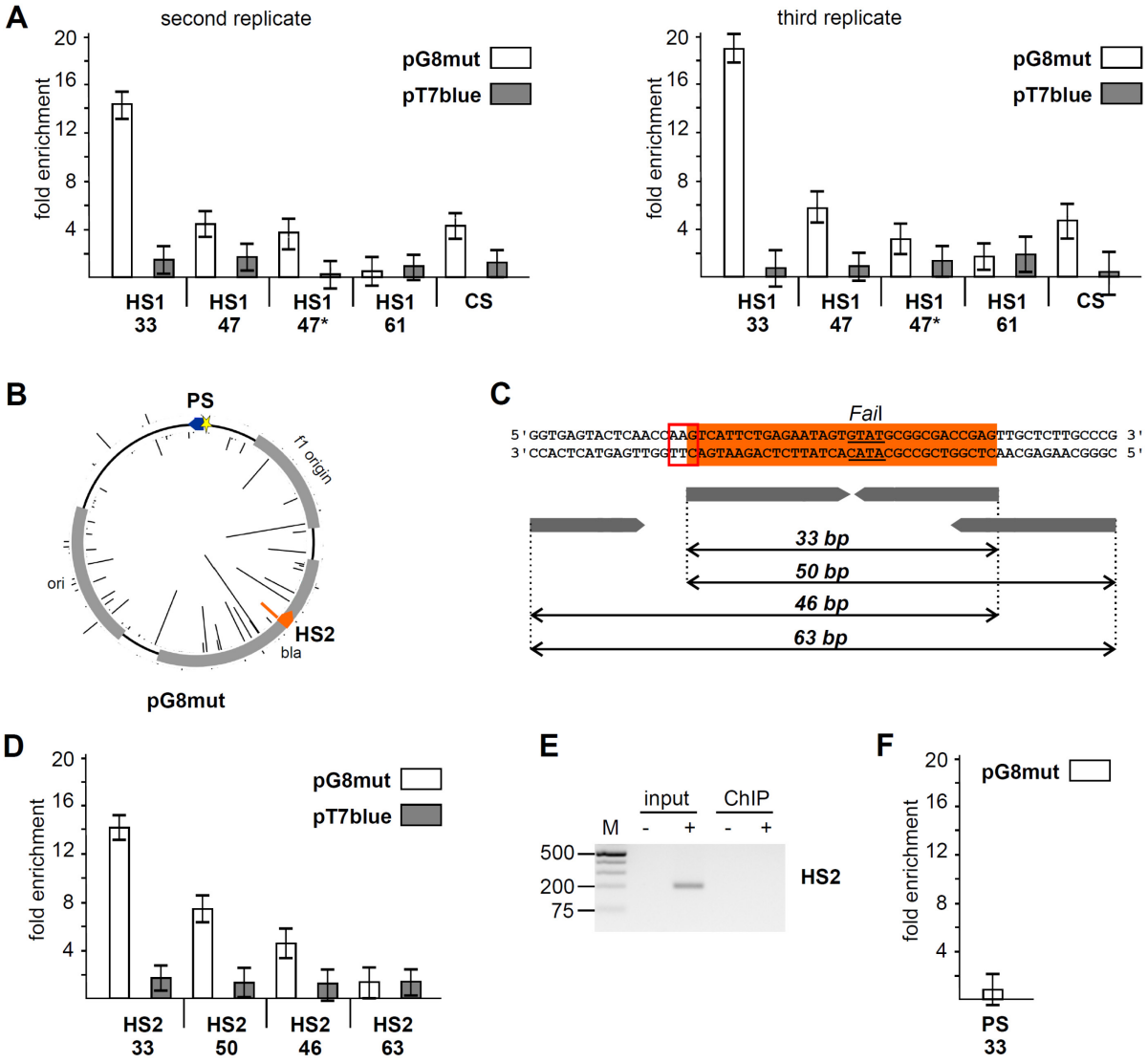

Figure S3.

A. Results of Cas1 ChIP analysis with primer pairs amplifying HS1 amplicons of different lengths and CS 33 bp amplicon are presented for second and third biological replicates (the first replicate results are shown in Fig. 2C and 2E).

B. Spacers acquired by KD263 cells transformed with pG8mut plasmid. The priming G8 protospacer is shown as a blue arrow with an asterisk indicating mutation introducing a mismatch with G8 crRNA spacer. The heights of the bars emanating from circle representing pG8mut plasmid indicate the relative efficiency of protospacer use (number

of times reads with matching acquired spacers were observed). The HS2 hot protospacer is represented by an orange arrow. The bar indicating the efficiency of spacer acquisition from HS2 is highlighted in orange.

C. The HS2 protospacer sequence (orange background) and the flanking plasmid DNA sequences are presented. The *Fai*I recognition site is labeled and underlined. The AAG PAM is highlighted with a red box. Below, primers used to amplify HS2 containing fragments of different lengths (see also Supplementary Table S1) are shown.

D. Results of Cas1 ChIP analysis with primer pairs amplifying HS2 amplicons of different lengths are presented. Fold enrichment corresponds to concentration ratios in induced and uninduced cells containing indicated plasmids. Mean values and standard deviations in triple technical replicates are shown.

E. Results of PCR amplification (40-cycles) of pre- and post- immunoprecipitation material from KD263 cells transformed with pG8mut plasmid. Primers used were complementary to HS2 and a region downstream of CRISPR array.

F. As in **D,** but using a primer pair amplifying priming protospacer region (PS) is shown.

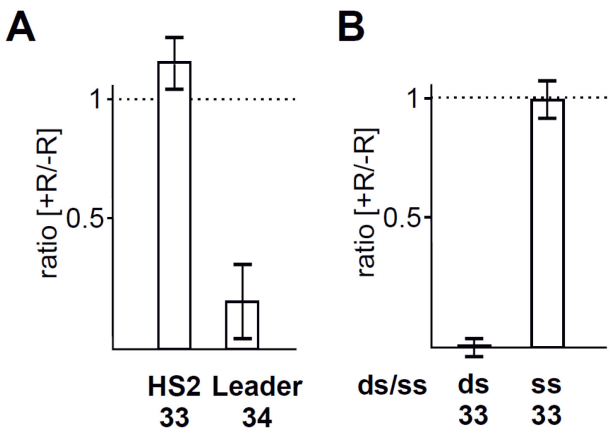

Figure S4. Cas1-associated protospacer DNA is resistant to restriction endonuclease digestion.

A. KD263 cells transformed with pG8mut were processed for Cas1 ChIP. After immunoprecipitation aliquots of Cas1-associated DNA were treated with *Fai*I restriction endonuclease. qPCR analysis with primer pairs amplifying the HS2 protospacer or a short (34 nucleotide) fragment of CRISPR array leader was performed and the ratio of fold enrichment values between *Fai*I-treated and untreated samples was determined.

B. Model single-stranded (ss) or double-stranded (ds) substrates of indicated lengths containing the HS2 protospacer were treated with *Fai*I endonuclease followed by qPCR analysis. For each substrate, the ratio of apparent concentrations (calculated from Ct values) with and without *Fai*I treatment is shown. Mean values and standard deviations for triple replicates are shown.

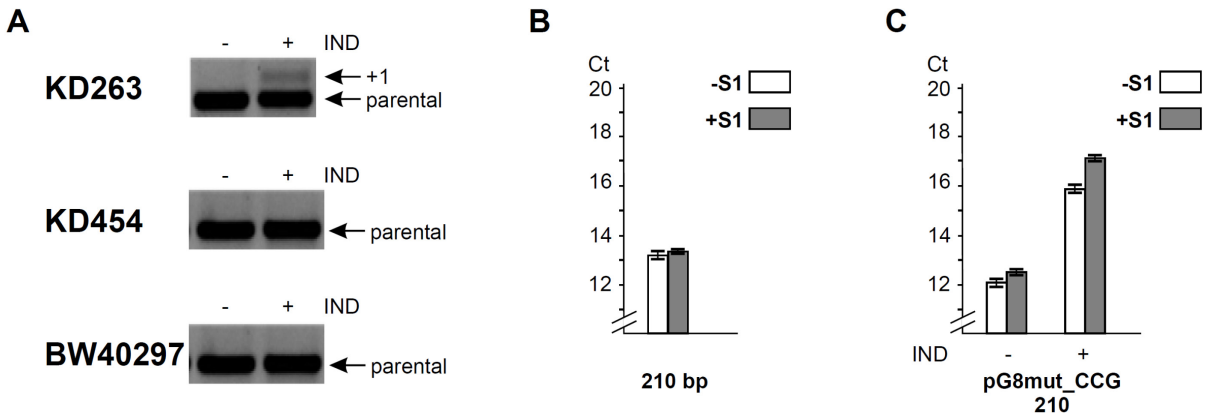

Figure S5. CRISPR interference leads to appearance of single-stranded target DNA *in vivo*.

A. KD263, KD454, or BW40297 cells were transformed with the pG8mut plasmid. Cells were grown with or without induction and PCR with oligonucleotide primers annealing upstream and downstream of the CRISPR array was performed. “Parental” marks a band corresponding to unexpanded, parental CRISPR array. The “+1” band corresponds to CRISPR arrays expanded by one spacer-repeat unit.

B. qPCR analysis (Ct values) of consequences of treatment of a model double stranded 210 bp DNA fragment with S1 nuclease.

C. Total DNA was prepared from KD263 cells transformed with pG8mut_CCG plasmid and grown in the presence or in the absence of *cas* gene expression inducers. qPCR with primer pairs amplifying a 210 bp plasmid fragment was next performed with S1 treated and untreated DNA samples aliquots. Ct values are presented.

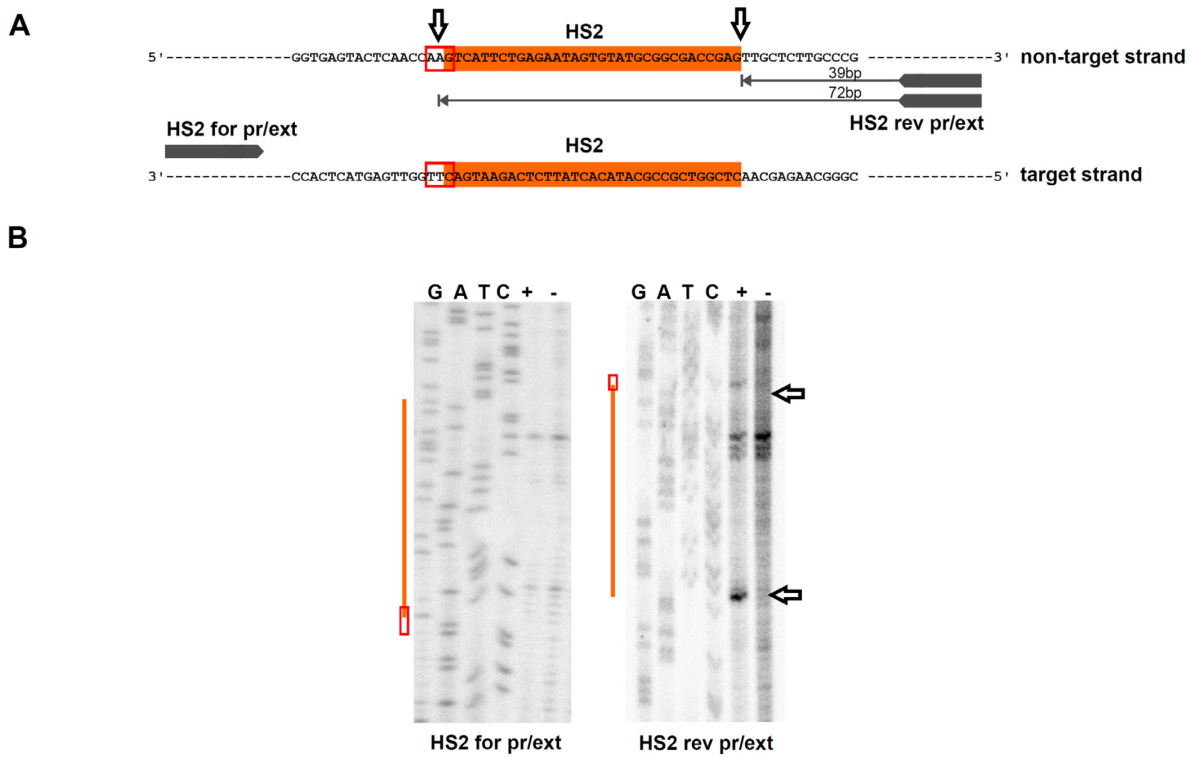

Figure S6. Target DNA is nicked at both sides of HS2 protospacer in cells undergoing primed adaptation.

A. Both strands of a plasmid DNA fragment containing the HS2 protospacer (marked by orange background) is shown and positions of annealing of primers used for primer extension reactions are indicated by solid arrows. For HS2 rev pr/ext primer, leftward-oriented thin arrows indicate the end points of primer extension products detected.

B. Results of primer extension experiment using total DNA purified from KD263 cells transformed with pG8mut and grown with ("+") or without ("-") *cas* gene induction. At the side of each autoradiogram, the position of the HS2 protospacer (orange line) and PAM (red box) is shown. Horizontal arrows indicate the positions of primer extension products migration on the gel.