Expanded View Figures

Figure EV1. Cas1:Cas2-DnaQ composition analysis.
Cas1 StrepII-tagged at N-terminus (S-Cas1) was co-expressed with either full-length Cas2-DnaQ protein, Cas2 domain, or DnaQ domain tagged with His6-tag at the C-terminus (Cas2-DnaQ-H, Cas2-H, and DnaQ-H, respectively). Co-elution of S-Cas1 with Cas2-DnaQ-H or separate Cas2-H and DnaQ-H domains from His6-tag (A, D, F) and StrepII-tag (B, E, G) affinity columns was used to probe Cas1 and Cas2 interactions. SDS–PAGE fractions of elution peak are shown in the inset.

A S-Cas1 and full-length Cas2-DnaQ-H isolation on His6-tag affinity column.
B S-Cas1 and full-length Cas2-DnaQ-H isolation on StrepII affinity column.
C S-Cas1 and Cas2-DnaQ-H proteins elute as a single peak from size-exclusion column.
D S-Cas1 and Cas2-H domain isolation on His6-tag affinity column.
E S-Cas1 and Cas2-H isolation on StrepII affinity column.
F When S-Cas1 was co-expressed with DnaQ-H, only DnaQ-H was isolated on His6-tag affinity column.
G When S-Cas1 was co-expressed with DnaQ-H, only S-Cas1 was isolated on StrepII affinity column.

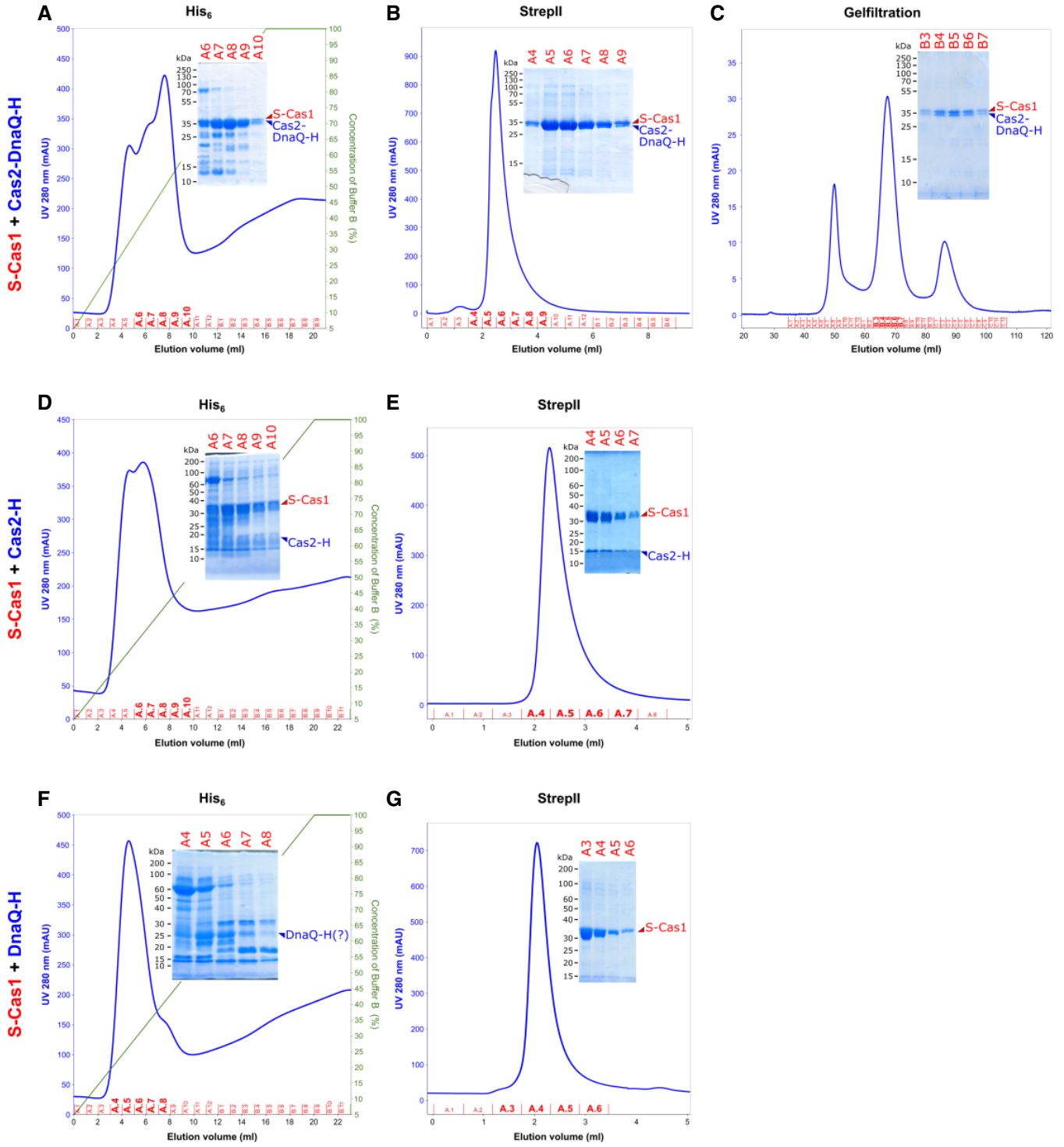
A
His6
S-Cas1 + Cas2-DnaQ-H
B
StrepII
C
Gelfiltration
D
His6
S-Cas1 + Cas2-H
E
StrepII
F
His6
S-Cas1 + DnaQ-H
G
StrepII
UV 280 nm (mAU)
Concentration of Buffer B (%)
Elution volume (ml)
kDa
S-Cas1
Cas2-DnaQ-H
Cas2-H
DnaQ-H(?)

Figure EV1.

Figure EV2. Activity dependence on divalent metal ions and sequence features.

A Integration reactions were performed in the presence of different divalent metal co-factor using Cas1:Cas2-DnaQ complex and 23 bp DNA duplex with 5 nt 3′-overhangs as protospacer. Mg^{2+}, Mn^{2+}, and Ca^{2+} were determined to be the best cofactors for integration reaction. Mn^{2+} activates DnaQ domain so strongly that integration becomes undetectable, due to degradation of spacers. Reactions were incubated for 30 min at 42°C.

B Exonuclease reactions were performed in the presence of different divalent metal co-factor using Cas1:Cas2-DnaQ complex and 72 nt single-stranded DNA oligonucleotide. DNA degradation was most efficient with Mg^{2+} and Mn^{2+} ions.

C Exonuclease rates of Cas1:Cas2-DnaQ complex were probed using polyA, polyT, and polyC single-stranded DNA oligonucleotides.

D Exonuclease rates of Cas1:Cas2-DnaQ complex were probed using polyA oligonucleotide (2′dT), which was blocked at its 3′-end with 2′, 3′-dideoxyadenosine (2′d3′dA) or 3′-deoxyadenosine (3′dA) using terminal deoxynucleotidyl transferase.

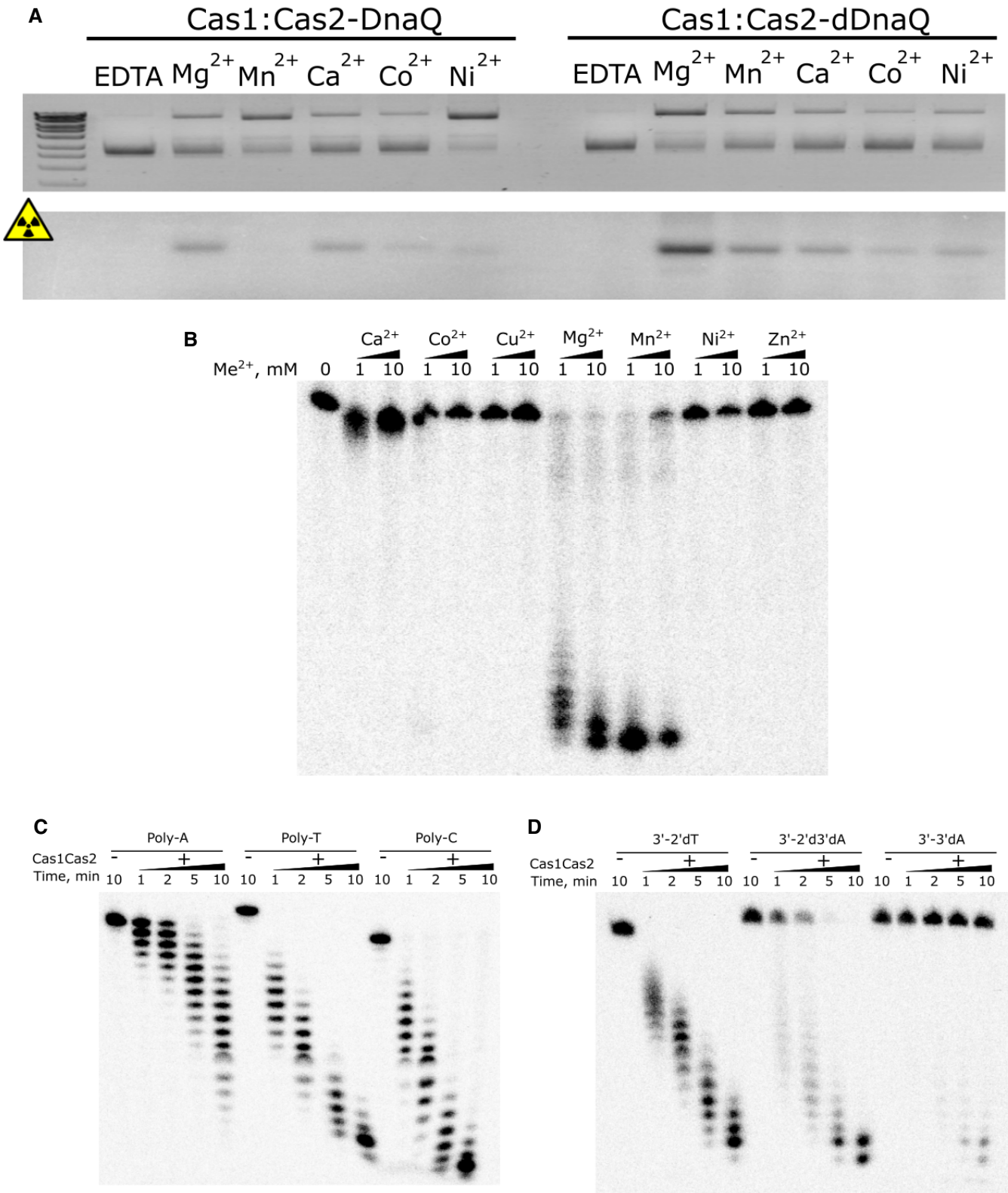

Figure EV2.

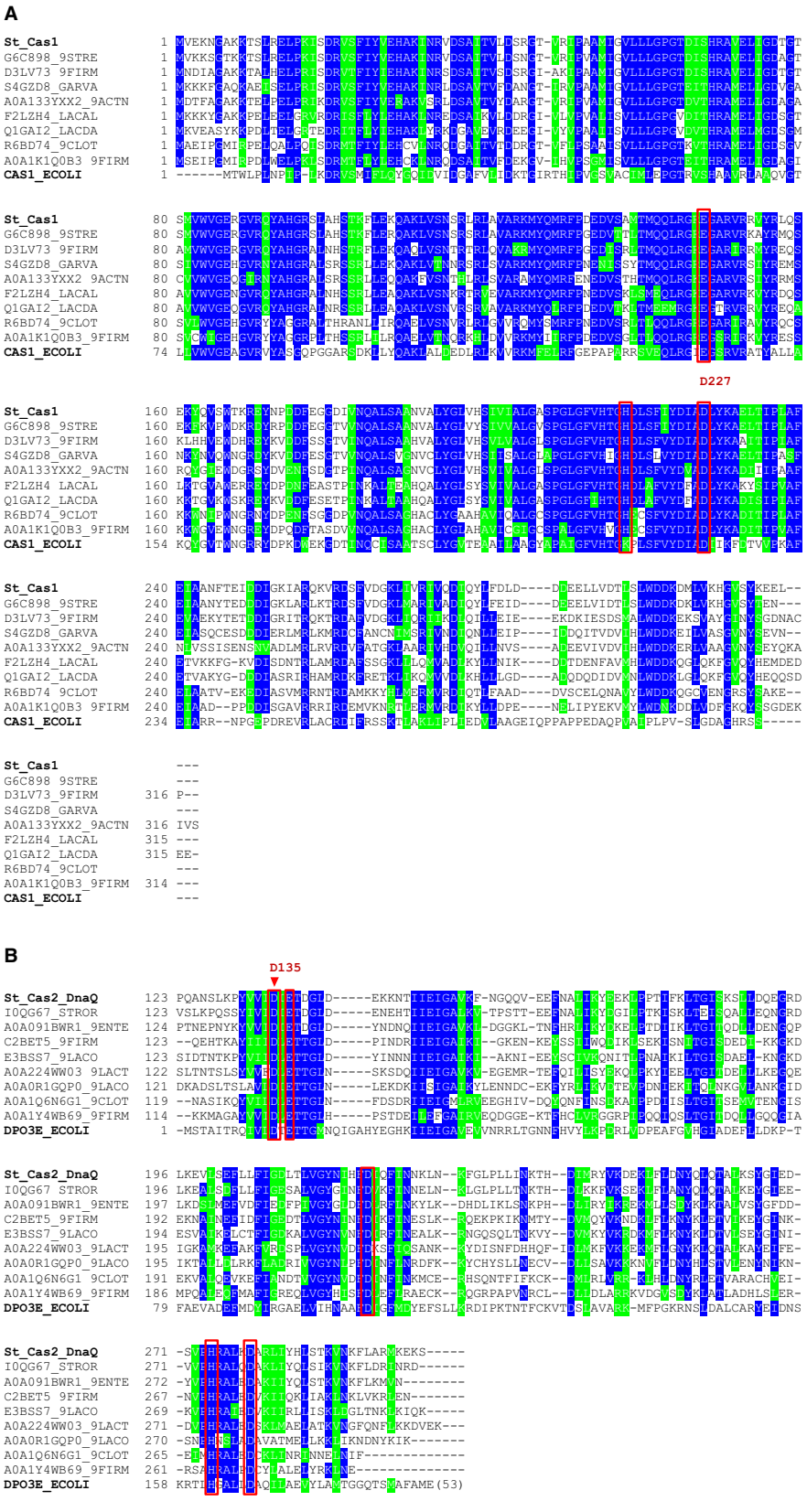

Figure EV3. Multiple sequence alignments used in prediction of active site residues of Cas1 and DnaQ domains.

A Alignment of *Streptococcus thermophilus* Cas1 (St_Cas1) and its homologs including well-characterized *Escherichia coli* Cas1 (CAS1_ECOLI). Positions corresponding to the active site residues of *E. coli* Cas1 are marked with red rectangles. The point mutation in *S. thermophilus* Cas1 is indicated with the red triangle. St_Cas1 homologs are represented by their Uniprot entry names.

B Alignment of *S. thermophilus* Cas2-DnaQ (St_Cas2_DnaQ) and its homologs with *E. coli* PolIII ε-subunit (DPO3E_ECOLI). Positions corresponding to the active site residues of ε-subunit are marked with red rectangles. The amino acid that was mutated in *S. thermophilus* Cas2-DnaQ is labeled with the red triangle. Sequences, except for St_Cas2_DnaQ, are represented by their Uniprot entry names.

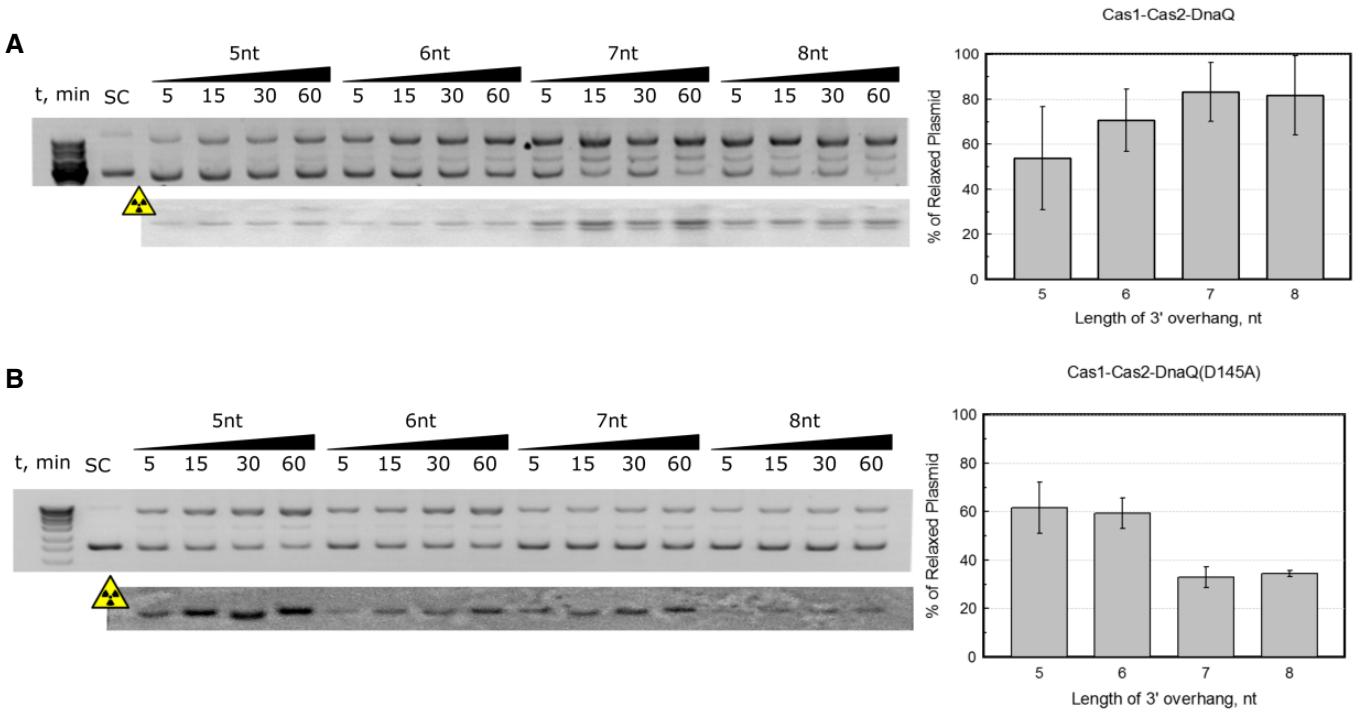

Figure EV4. Integration efficiencies of protospacers with different 3′-overhang lengths.
Numbers represent percentage of relaxed plasmid form after incubating integration reaction for 1 h. Values were normalized as percentage of the largest value, which subsequently was taken as 100%.

A Integration reactions performed using WT Cas1:Cas2-DnaQ complex; $N = 3$; error bars represent standard error of the mean value. This complex integrates duplexes with 7 or 8 nt overhangs better than purported optimal protospacer bearing 5 nt overhangs.

B Integration reactions performed using Cas1:Cas2-DnaQ (D135A) mutant (inactive DnaQ); $N = 3$; error bars represent standard error of the mean value. Conversely, DnaQ mutant integrates optimal protospacer noticeably better than those bearing 7 or 8 nt overhangs.

Figure EV5. PCR-based assay for detection of integration into CRISPR locus.

A Integration products were amplified by PCR using primers binding to the leader sequence and protospacers that were integrated to the CRISPR region.

B PCR products are analyzed on 1.5% agarose gel stained with EtBr. 1st repeat and 2nd repeat denoted bands correspond to integrations after the first or second repeats (as drawn in panel A). Lowest band corresponding to integration event in the vicinity of the first repeat was cut out and purified and subsequently cloned into pJET1.2.

C Length distribution of integrated protospacer 3′-overhangs as determined from Sanger sequencing of separate pJET1.2 clones, $n = 29$.

D Integration products were amplified by PCR using primers binding to the 2nd spacer and protospacers that were integrated into the CRISPR region.

E PCR products are analyzed on 1.5% agarose gel stained with EtBr. 1st repeat and 2nd repeat denoted bands correspond to integrations after the first or second repeats (as drawn in panel A). Upper band, corresponding to integration event in the vicinity of the first repeat was cut out and purified and subsequently cloned into pJET1.2.

F Length distribution of integrated protospacer 3′-overhangs as determined from Sanger sequencing of separate pJET1.2 clones, $n = 16$.

Source data are available online for this figure.

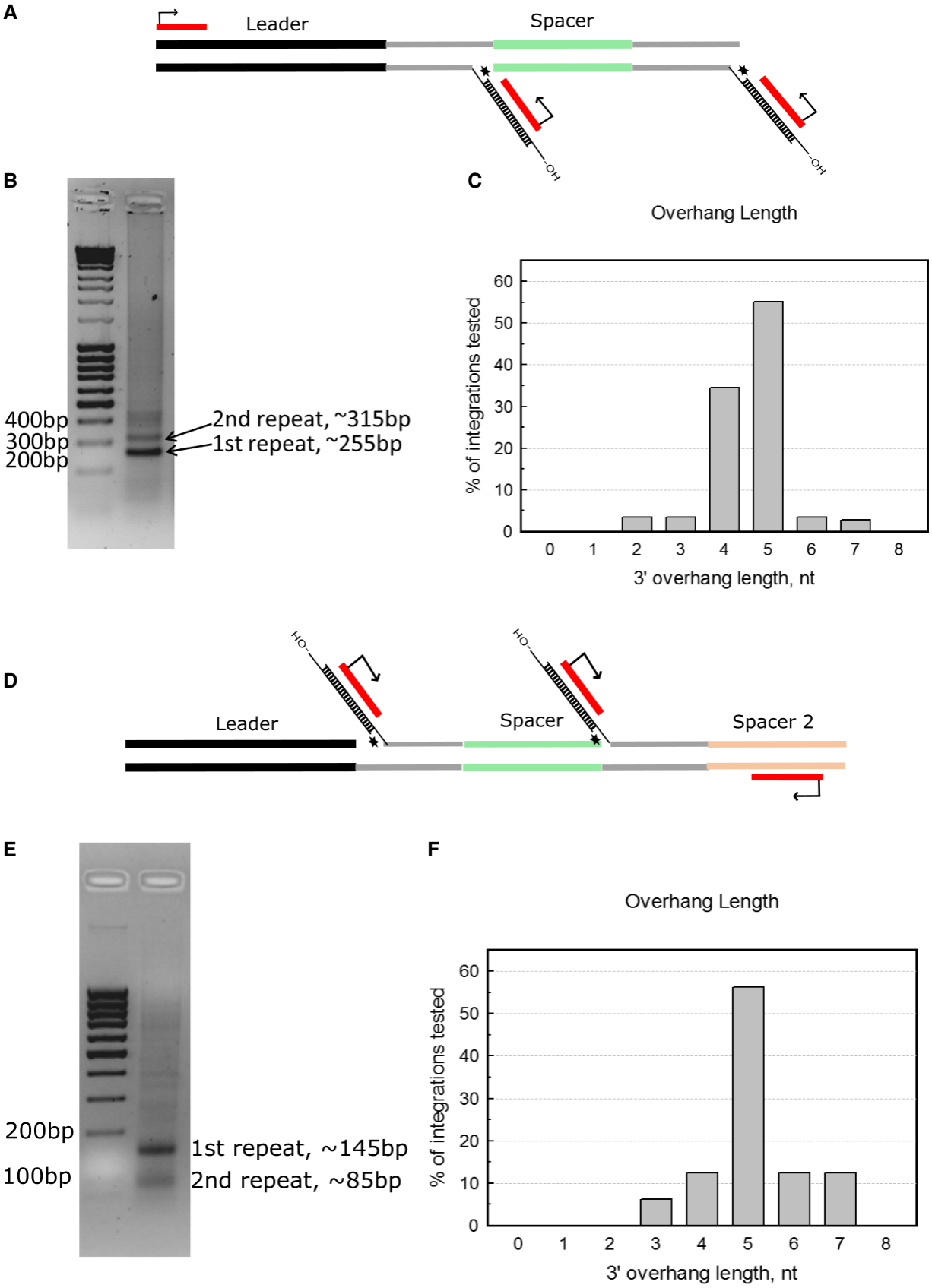

Figure EV5.