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.



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. Mg²⁺, Mn²⁺, and Ca²⁺ were determined to be the best cofactors for integration reaction. Mn²⁺ 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 Mg²⁺ and Mn²⁺ 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.

Α	
St_Cas1 GGC998_9STRE D3LV73_9FIRM S4G2D8_GARVA A0A1337XX2_9ACTN F2L2H4 LACAL Q1GA12_LACDA R6BD74_9CLOT A0A1K1Q0B3_9FIRM CAS1_ECOLI	1 MVEKNEAKETSIRGLEKISORVETIYUEHAKINEVOSATTVLESRET-ÜRIFAAKIGVLLLOPOTDISHRAVELLOITOT 1 MVRKSETKETSIRGLEKISORVETIYUEHAKINEVOSATTVLESRET-ÜRIFAAKIGVLLLOPOTDISHRAVELLOITOT 1 MVRKSETARETSIRGLEKISORVETIYUEHAKINEVOSATTVSESRET-ÄRIFAANIGVLLOPOTDISHRAVELLOITOT 1 MKRKSEARKALHOLPRISORVETIYUEHAKINEVOSATTVSERAT-ÄRIFAANIGVLLOPOTDISHRAVELLOITOT 1 MKRKSEARKALHOLPRISORVETIYUEHAKINEVOSATTVSERAT-ÄRIFAANIGVLLLOPOTDISHRAVELLOITOT 1 MKRKSEARKALHOLPRISORVETIYUEHAKINEVOSATTVSERAT 1 MKRKSEARKALGIERSORVETIYUEHAKINEVOSATTVSERAT 2 MKVEASTKEPINELSELERISORVETIYUEHAKINEVOSATTVSEANIT-ÄRIFAANIGVLLLOPOTDISHRAVELLOITOT 1 MKVEASTKEPINELSELERISORVETIYUEHAKINEVOSATTVSEANIT-ÄRIFAANIGVLLLOPOTUHHAMELLOITOT 2 MKVEASTKEPINETIESIERISORVETIYUEHAKINEVOSATTVETORIT-ÜRIFAANIS 2 MKVEASTKEPINETIESIERISORVETIYUEHAKINEVOSATTVEDRIT-ÜRIFAANIS 2 MKVEASTKEPINETIESIERISORVETIYUEHAKINEVOSATTVEDRIT-ÜRIFAANIS 2 MKVEASTKEPINETIESISONNETIYUEHAKINEVOSATTVEDRIT-ÜRIFAANIS 2 MKVEASTKEPINETIESISONNETIYETYETYEN KIKSISONTTVEDRIT-ÜRIFAANIS 2 MKVEASTKEPINETIESISONNETIYETYETYEN KIKSISONTTVENEN 2 MSEITEMISPENALISONNETIYETYETYETYEN KIKSISONNETIYETYEN 2 MSEITEMISPENALISONNETIYETYETYETYEN KIKSISONNETISTI 2 MKVEASTKEPINETIESISONNETIYETYETYETYEN KIKSISONNETISTI 3 MKVEASTKEPINETIESISONNETIYETYETYETYETYEN SONTTVENENTI 3 MKVEASTKEPINENTIYETYETYETYETYETYEN KIKSISONNETIYETYETYETYETYETYETYETYETYETYETYETYETYETY
St_Cas1 G6C898_9STRE D3LV73_9FTRM S4G2D8_CARVA AOA133XX22_9ACTN F2L2H4_LACAL Q1GA12_LACDA G6D74_9CLOT AOA1K1QOB3_9FIRM CAS1_ECOLI	 80 SKUWVGERGVROYAHGRSDAHSTSFLEROAKLVSNSELRUAVARKMYORRFEDEDVSACTMOOLGGE SARVERUTELOS 80 SKUWVGERGVROYAHGRSDAHSTSFLEROALVSNSESRLAVARKMYORRFEDEDVSACTMOOLGGE SARVERUTERMOS 80 SLUWVGERGVROYAHGRADHHERFELEVOALVSNSESRLAVARKMYORRFEDEDVSTTUTMOOLGGE SARVERUTERMOS 80 SLUWVGERGVROYAHGRADHHERFELEVOALVSNTERTUVARKMYORRFEDEDVSTTUMOLGGE SARVERUTERMES 80 SLUWVGERGVROYAHGRADHHERFELEVOALVSNTERTUVARKMYORRFEDEDVSTTUMOLGGE SARVERUTERMES 80 SLUWVGERGVROYAHGRADHHERFELEVOALVSNTERTUVARKMYORRFEDEDVSTTUMOLGGE SARVERUTERMES 80 SLUWVGERGVROYAHGRADHSSELLEOAKUVSNTERTUVARKMYORRFEDEDVSTTUMOLGGE SARVERUTERMES 80 AVWWGENGVROYAHGRADHSSELLEOAKUVSNTERTUSVARMYORRFEDEDVSTTUMOLGGE SARVERUTERS 80 SUMWGENGVROYAHGRADHSSELLEOAKUVSNTERTUSVARMYORRFEDEDVSTUMOLGGE SARVERUTERS 80 SUMWGENGVROYAHGRADHSSELLEOAKUVSNTERTUSVARMYORRFEDEDVSTUMOLGGE SARVERUTERS 80 SUMWGENGVROYAHGRADHSSELLEOAKUVSNTERTUSVARMYORFFEDEDVSTUDOLGGE SARVERUTERS 80 SUMWGENGVROYAHGRADHSSELLEVOAKUVSNTERTUSVARMYORFFEDEDVSTUDOLGGE SARVERUTERS 80 SUMWGENGVROYAHGRADHSSELLEVOXUVSNTERTUSVARMYSTERFEDEDVSTUDOLGGE SARVERUTERS 80 SUMWGENGVROYAHGRADHSSELLEVOXUVSNTERTUSVARMYSTERFEDEDVSTUDOLGGE SARVERUTERS 80 SUMWGENGVROYAHGRADHSSELLEVOXUVSNTERTUSVARMYSTERFEDEDVSTUDOLGGE SARVERUTERS 80 SUMWGENGVROYAHGRADHTHANLLTEVARVENTERSTUDOLGGERAFARTSVELUE 80 SUMWGENGVROYAHSSOFTHANLLTEVARVENTERSTUDOLGGERAFARTSVELUE 80 SUMWGENGVROYAHGRADHTHANLLTEVARVENTERSTUDOLGGERAFARTSVELUE 80 SUMWGENGVROYAHGRADHTHANLTEVARVENTERSTUDOLGGERAFARTSVELUE 80 SUMWGENGVROYAHTANGTUSVARTTUSVARTATISVERUE 80 SUMWGENGVROYANSOFTHANULTEVARVENTERSTUDOLGGERAFARTSVELUE 80 SUMWGENGVROYANSOFTHANULTEVARVENTERSTUDOLGGERAFARTSVELUE 80 SUMWGENGVROYANSOFTHANULTEVARVENTERSTUDOLGGERAFARTSVELUE
St_Casl G6C898_9STRE D3LV73_9F1RM S4G2D8_GARVA A0A133YXX2_9ACTN F2L2H4 LACAL Q1GA12_LACDA R6BD74_9CLOT A0A1K1Q0B3_9F1RM CAS1_ECOLI	160 EKYCUSATKREYNPEDFEGEDIVNOALSAANVALYGUVHSTVIALGASPGLGEVHTY HILSETYDIID YKABLAIPIAF 160 EKYKUPADKREYRPEDFEGETYVNOALSAANVALYGUVHSTVIALGASPGLGEVHTY HILSETYDIID YKABLAIPIAF 160 KLHHENDHREYKVEDFESETYVNOALSAANVALYGUVHSTVIALGASPGLGEVHTY HILSETYDIID YKABLAIPIAF 160 RYNNOANGREYKVEDFESETYVNOALSYENVCLYGUVHSTVIALGASPGLGEVHTY HILSEVYDIID YKABLAIPIAF 160 RYNNOANGREYKVEDFESETYVNOALSYENVCLYGUVHSTVIALGASPGLGEVHTY HILSEVYDYD DYKABLAIFIAF 160 RYNNOANGREYKVETESETYTNAALSYENVCLYGUVHSTVIALGASPGLGEVHTY HILSEVYDYD DYKABLAIFIAF 160 RYNNOANGREYKVETESETYTNAALSYENVCLYGUVHSTVIALGASPGLGEVHTY HILSEVYDYD DYKABLAIFI 160 RYNNOFARGEYDENESETTINAALTEHDHOALSGISYSTVIALGASPGLGEVHTY HILSEVYDYD DYKABLAIFI 160 KKWLEPANGREYDENESETTINAALTEHDHOALSGISYSTVIALGASPGLGEVHTY HILSEVYDYD DYKABLAIFI 160 KKWLEPANGRYDENESETTINAALTEHDHOALSGISYSTVIALGASPGLGEVHTY HILSEVYDYD DYKABLAIFI 160 KKWLEPANGRYDENESETTINAALTEHDHOALSGISYSTVIALGASPGLGEVHTY HILSEVYDYD DYKABLAIFI 160 KKWLEPANGRYDENESETTINAALTEHDHOALSGISYSTVIALGASPGLGEVHTY HILSEVYDYD DYKABLAIFI 160 KKWLEPANGRYDENESETTINAALTEHDAALSGISTVIALGASPGLGEVHTYH HILSEVYDYD DYKABLAIFI 160 KKWLEPANGRYDENESETTINAALTEHDAALSGISTVIALGASPGLGEVHTYH HILSEVYDYD DYKABLAIFI 160 KKWLEPANGRYDENESETTINAALTEHDAALSGISTVIALGASPGLGEVHTYH HILSEVYDYD DYKABLAIFI 160 KKWLEPANGRYDENESETTINAALTEHDAALSGISTAALAITAAL
St_Cas1 GGC998_9STRE D3LV73_9FIRM S4G2D8_GARVA A0A1333XX2_9ACTN F2L2H4_LACAL Q1GAI2_LACAL Q1GAI2_LACAL GBD74 9CLOT A0A1K1Q0B3_9FIRM CAS1_ECOLI	240 SEMANFTEIDDIGKIARQKVRIS VOCKUVRIVGIVGE OY 240 SEMANFTEIDDIGKIARQKVRIS VOCKUVRIVGIVGE OY 240 SEMANFTEIDDIGKIARQKTRIS VOCKUVRIGARUNDIOY 241 SIMARVTEDDIGKIARQKTRIS VOCKUVRIKARUNDIOY 242 SIMARVTEDDIGKIARQKTRIA 243 SIMARVTEDDIGKIARQKTRIA 244 SIMARVTEDDIGKIARQKTRIA 240 SIMARVTEDDIGKIARQKTRIA 241 SIMARVTEDDIGKIARQKTRIA 242 SIMARVTEDDIGKIARQKVRIVAROVARUNDICALLEPDECKIESDSJALNDDVKIKSUVRIKEONKEEVN 240 SIMARVG-VOISDIRLANGRASSCALLONNIKSUVANUKVOILLEP
St_Cas1 G6C898 9STRE D3LV73_9FIRM S4G2D8 GARVA A0A133YXX2_9ACTN F2L2H4_LACAL Q1GAI2_LACDA R6BD74_9CLOT A0A1K1Q0B3_9FIRM CAS1_ECOLT	 316 IVS 315 315 EE- 314
В	
St_Cas2_DnaQ 10Qc67_STROR A0A091EWR1_9ENTE C2EET5_9FIRM E3BSS7_9LACO A0A224WW03 9LACT A0A0R10QP0_9LACO A0A126M6G1_9CLOT A0A1Y4WB69_9FIRM DPO3E_ECOLI	123 PQANSLKEVU D STOGDEKKNTITEIGAVKF-NGQQV-EESNALTAVSEKLEPTTFKLTGISKS DDQERD 124 VSLKPQSSINU D STOGDEKKNTITEIGAVKI-DGGKL-TNGHRUHYJEKLEPTTFKSKLBISQALESON RD 124 PTNEPNYKUU D STEGDYINDRITEIGAVKI-DGGKL-TNGHRUHYJEKEPTDIIKLTGI QDUDENG OP 123 JOTNTERVID STTGDYINDRITEIGAVKI-DGGKL-TNGHRUHYJEKEPTDIIKLTGI QDUDENG OP 123 SIDTNTERVID STTGDYINDRITEIGAVKI-SKENSSINQI IKSEKISNITGI DDEDLKKG 124 SITGIDYINDRITEIGAVKI-SKENSSINQI IKSEKISNITGI DDEDLKKG 125 SIDTNTERVID STTGIDYINDRITEIGAVKI-SKENSSINQI IKSEKISNITGI DDEDLKKG 126 SITGIDSKEDGI EIGAVKI-SGEMR-TEQILISYEKOLEKVIELTGIDEDLKKG 127 DKADSITSIJAND STTGINEKKKIISICAIKYLENDC-KKYKNINYTEPDNIKKI OTNKGUANKEID 139NASISLAND STTGINPSDRITEIGAVKI-SGENEVILV-D000GGRPHEDDISKITGITSEVIPPOL 14KKMAGAVWU DISTGINPSDRITEIGAVKI-SGENEVDOGGE-KTHCDVRGGRPHEDOLSGGETA 1 -MSTAITRCIVI STTGIHPSDRITEIGAVENVNRLTGNNSHVYLEPRIS
St_Cas2_DnaQ I0Qc67 STROR A0A091BwR1_9ENTE C2EBT5_9FIRM E3BSS7_9LACO A0A0224WW03_9LACT A0A0R1GQP_9LACCO A0A01Q6N6G1_9CLOT A0A1Y4WB69_9FIRM DP03E_ECOLI	196 LKEVISE LLEKELTIVGYN HERE GENNKLNKFGLPLLINKTHDIN RYWKDER FEDNYCH GAUKSIGED- 196 LKEALSE LLEKESTIVGYGENED KEINSLNKGGPLINKTHDIKKFYKSEN FLANYCH GAUKENGEE 197 LKESMEEVDELE FEVRYGED DER BUNKYLKDELLIKLSNKPHDIR RYWESKELTSKEK KAUVSGGPD 192 EKNALMEIDIESDIVGYN NO DER INSLKSOEKPKIKNTYDWRYWRDKFERDIESDYKKKAUVSGGFD 194 ESVAK KLICTFLEKALVGYN NO DER INSLKSOEKPKIKNTYDWRYWRDKFERM KEINFERDIESD 195 IKKALLER KLIGHT ON SKRINGER SOLSKSOENFERT KUNTYDWRYWRDKFERM KEINFERDIESD 196 LEKNALMEIDIESDIVGYN DER SELSKSOESNE-SWINSLE SOLSKER KAUKSE VER KENNE 197 EKNALLER KLIGHT ON SKRINGER SOLSKER SOLSKER SOLSKER KENNE DWR KENNE 198 EKNALDER SELVGYN DER SELSKER SOLSKER SOLSKER SOLSKER KENNE DWR KENNE DWR KENNE 199 EKNALDER SELVGYN DER SELSKER SOLSKER SOLS
St_Cas2 DnaQ I0QG67 STROR A0A091EWR1_9ENTE C2EBT5 9FIRM C2EBT5 9FIRM SBSS7 9LACO A0A224WW03_9LACT A0A0R1GQP0_9LACC A0A0R1GQP0_9LACC A0A124WE69_9FIRM DP03E_ECOLI COLAT	271 -SVE REAL FOR ANTIPLE STRVAKFLARWKEKS 271 -VVE REAL OR ANTIPLE STRVAKFLARWKEKS 272 -VVE REAL OR ANTIPLE STRVAKFLARWKEKS 273 -VVE REAL OR ANTIPLE STRVAKFLARWKENS 274 -MVE REAL OR ANTIPLE STRVAKFLARWKENS 275 -MVE REAL OR ANTIPLE STRVAKFLARWKENS

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 wellcharacterized 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.
- 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 F (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.