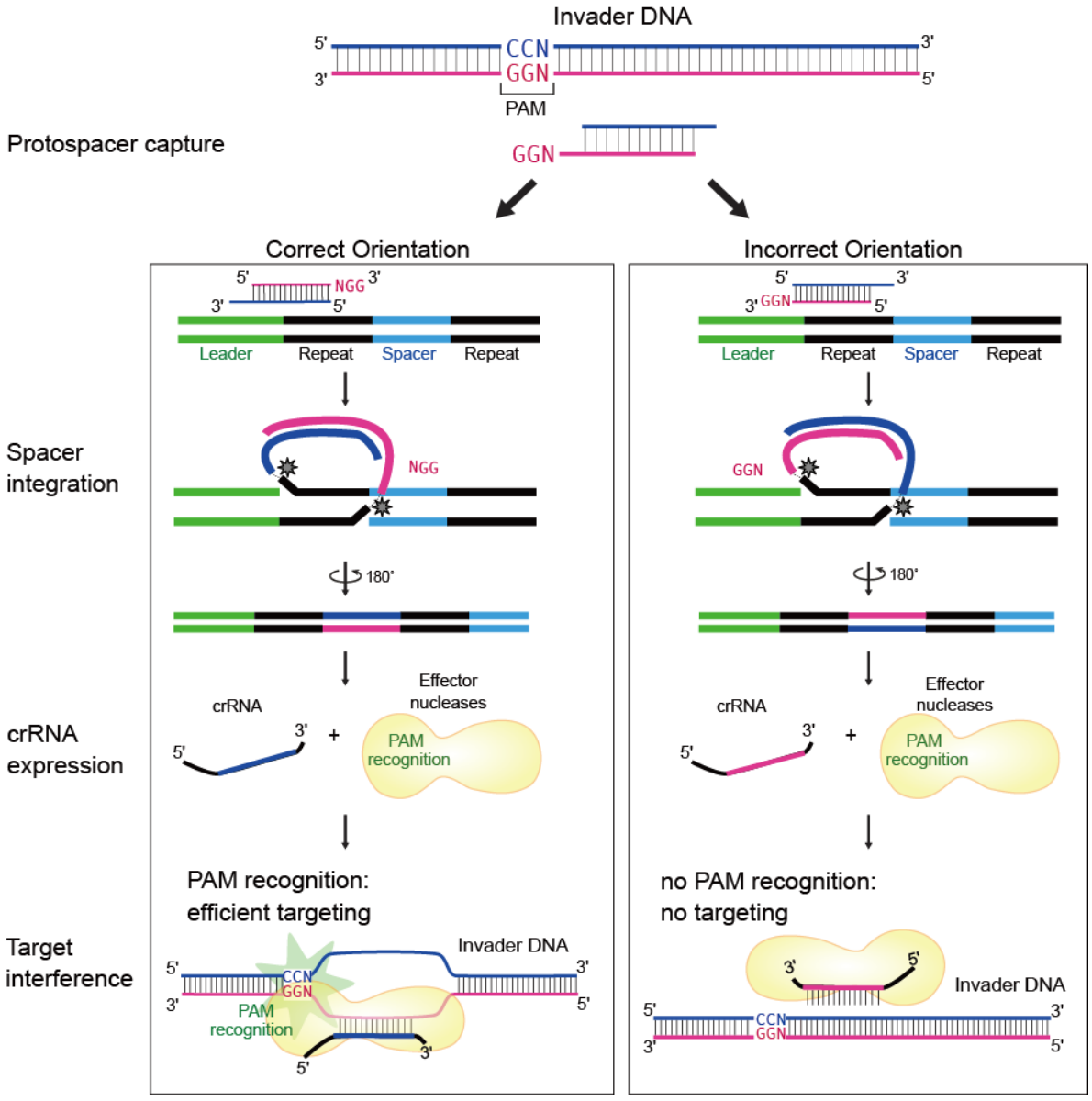

Invader DNA
5'
CCN
3'
3'
GGN
5'
PAM
Protospacer capture
GGN
Correct Orientation
Incorrect Orientation
5'
NGG
3'
3'
5'
3'
GGN
5'
Leader
Repeat
Spacer
Repeat
Spacer integration
NGG
GGN
180°
crRNA expression
crRNA
Effector nucleases
5'
3'
+
PAM recognition
PAM recognition: efficient targeting
no PAM recognition: no targeting
Target interference
Invader DNA
CCN
GGN
PAM recognition

Figure S1. Related to Figure 2; CRISPR-Cas immunity requires that the spacer is integrated into the CRISPR array in a polarized manner. Protospacers are integrated into the leader-proximal repeat of the CRISPR array. When spacer integration is in the correct orientation (left panel), the PAM adjacent side of the spacer points towards the leader. After crRNA expression, target interference is initiated by strand-specific recognition of the 5'-NGG-3' PAM on the target strand by immune effector nucleases. When spacers are integrated in the incorrect orientation (right panel), the PAM adjacent side of the spacer points away from the leader and the resulting crRNA cannot initiate target interference because effector crRNA-Cas protein complexes cannot recognize the NGG PAM on the DNA target strand (complementary to crRNA). Thus, target interference requires that a spacer is integrated in the correct orientation.

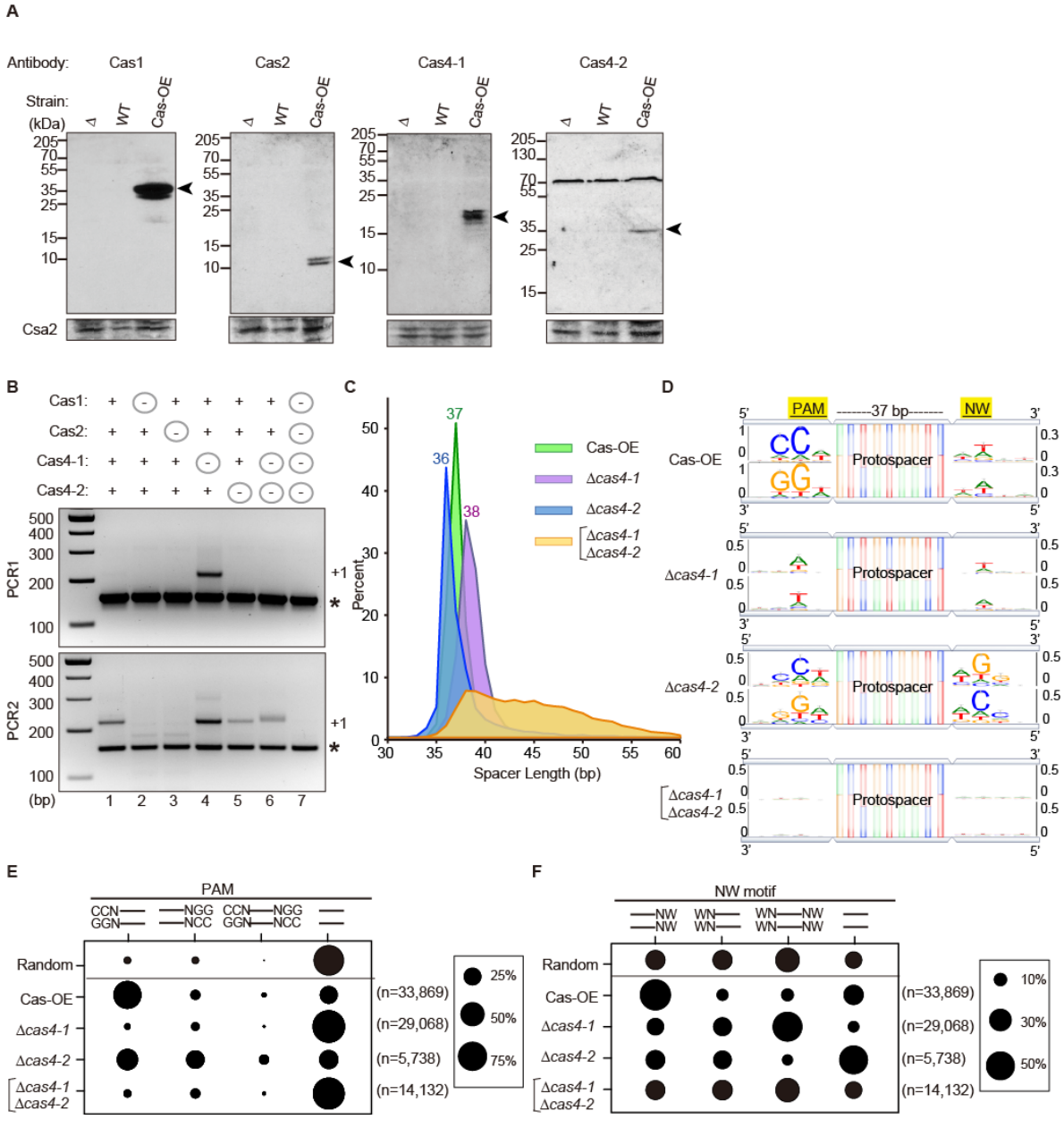

A
Antibody:
Cas1
Cas2
Cas4-1
Cas4-2
Strain:
Δ
WT
Cas-OE
(kDa)
Csa2
B
Cas1:
Cas2:
Cas4-1:
Cas4-2:
PCR1
PCR2
+1
(bp)
C
Percent
Spacer Length (bp)
Cas-OE
Δcas4-1
Δcas4-2
Δcas4-1
Δcas4-2
D
PAM
37 bp
NW
Protospacer
E
PAM
CCN
GGN
NGG
NCC
Random
(n=33,869)
(n=29,068)
(n=5,738)
(n=14,132)
25%
50%
75%
F
NW motif
NW
WN
10%
30%
50%

Figure S2. Related to Figure 2; Cas4 deletion mutants in Cas overexpression background. (A) Western blot with Cas1, Cas2, Cas4-1, Cas4-2 and Csa2 antibodies for *P. furiosus* strains: Δ (Cas1, Cas2, Cas4-1 and Cas4-2 deletion strain), wild type (WT) and Cas-OE (Cas1, Cas2, Cas4-1 and Cas4-2 overexpression strain). Csa2 is used as a loading control. The indicated antibodies detect a major band (shown by arrowhead) of the expected molecular weight for each antigen. The predicted molecular weight of each protein are as follows: Cas1 (37.5 kDa), Cas2 (10.0 kDa), Cas4-1 (20.2 kDa), Cas4-2 (25.1 kDa) and Csa2 (37.4 kDa). (B) Analysis of adaptation in deletion strains created in a background strain overexpressing Cas1, Cas2, Cas4-1, and Cas4-2 proteins. The leader/first spacer region of CRISPR7 was amplified with primers indicated in Figure 1D. The PCR product corresponding to the parental array with no new spacer is indicated with an asterisk and the product corresponding to the array with a new repeat-spacer unit is indicated with a +1. (C) Line graph showing the length distributions of new spacers acquired into the CRISPR5 and CRISPR7 arrays. The X-axis indicates spacer length, and the Y-axis indicates % of spacers observed. Pooled data from four experiments are presented. (D) Newly-acquired spacers in CRISPR7 arrays were aligned to the genome and plasmids to identify the corresponding protospacers, and upstream and downstream sequences were extracted and used to generate consensus motifs on both strands of DNA. Four bp of flanking sequence on each side of the protospacers is shown. (E and F) Percentage of protospacers with/without upstream and/or downstream motifs. (E) CCN/NGG motif. (F) NW/WN motif. "Random" indicates the percentage at which each scenario (upstream, downstream, both, neither) would be observed if 37 bp spacers were selected randomly from the *P. furiosus* genome. Representations show data that were pooled from between 4 experiments (CRISPR5 and CRISPR7 arrays, 2 replicates each, see Supplemental Table 2 for raw values).

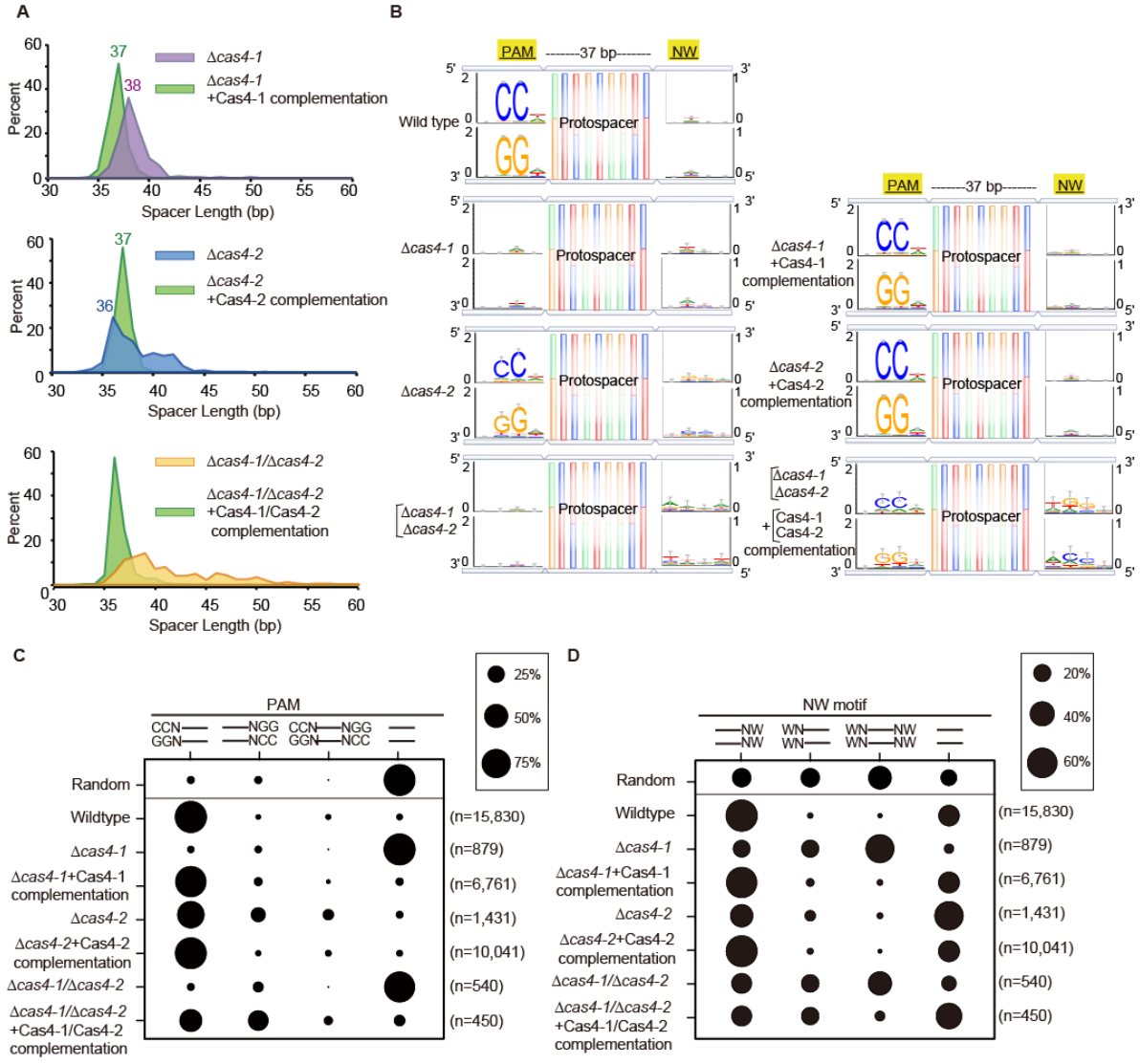

A
B
C
D
Percent
Spacer Length (bp)
37
38
36
Δcas4-1
Δcas4-1 +Cas4-1 complementation
Δcas4-2
Δcas4-2 +Cas4-2 complementation
Δcas4-1/Δcas4-2
Δcas4-1/Δcas4-2 +Cas4-1/Cas4-2 complementation
PAM
-------37 bp-------
NW
Wild type
Protospacer
Δcas4-1 +Cas4-1 complementation
Δcas4-2 +Cas4-2 complementation
Δcas4-1 Δcas4-2
+ Cas4-1 Cas4-2 complementation
CCN GGN NGG NCC
Random
Wildtype
Δcas4-1
Δcas4-1+Cas4-1 complementation
Δcas4-2
Δcas4-2+Cas4-2 complementation
Δcas4-1/Δcas4-2
Δcas4-1/Δcas4-2 +Cas4-1/Cas4-2 complementation
(n=15,830)
(n=879)
(n=6,761)
(n=1,431)
(n=10,041)
(n=540)
(n=450)
25%
50%
75%
NW motif
WN NW
20%
40%
60%

Figure S3. Related to Figure 2; Complementation of Cas4 deletion strains. (A) Line graph showing the length distributions of new spacers acquired into the CRISPR5 and CRISPR7 arrays. The X-axis indicates spacer length, and the Y-axis indicates % of spacers observed. Pooled data from eight experiments for deletion strains and four experiments for complemented strains are presented. (B) Newly-acquired spacers in CRISPR5 and CRISPR7 arrays were aligned to the genome and plasmids to identify the corresponding protospacers, and upstream and downstream sequences were extracted and used to generate consensus motifs on both strands of DNA. Four bp of flanking sequence on each side of the protospacers is shown. (C, D) Percentage of protospacers with/without upstream and/or downstream motifs. (C) CCN/NGG motif. (D) NW/WN motif. "Random" indicates the percentage at which each scenario (upstream, downstream, both, neither) would be observed if 37 bp spacers were selected randomly from the *P. furiosus* genome.

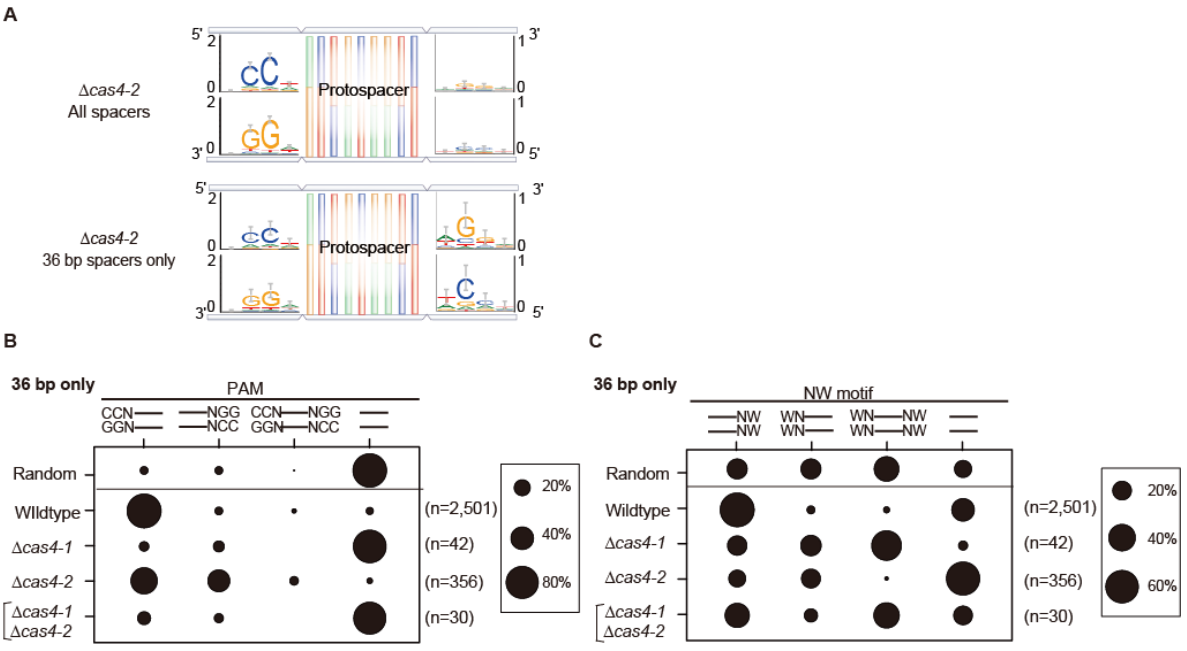


Figure S4. Related to Figure 2; Cas4-1 and Cas4-2 define the correct orientation. (A) Newly-acquired spacers in the CRISPR5 and CRISPR7 arrays in Δ*Cas4-2* were aligned to the genome and plasmids to identify the corresponding protospacers, and upstream and downstream sequences were extracted and used to generate consensus motifs on both strands of DNA. Four bp of flanking sequence on each side of the protospacers is shown. (B, C) Percentage of 36 bp protospacers with/without upstream and/or downstream motifs. Data from 8 replicates were pooled. "Random" indicates the percentage at which each scenario (upstream, downstream, both, neither) would be observed if 37 bp spacers were selected randomly from the *P. furiosus* genome. (B) CCN/NGG motif. (C) NW/WN motif.

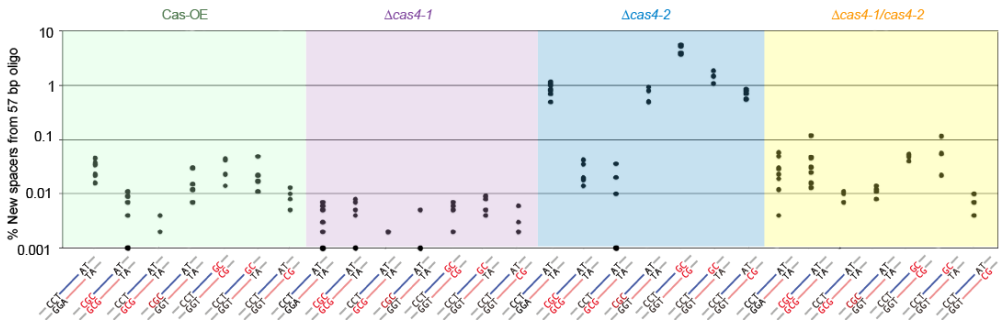


Figure S5. Related to Figure 3 and Figure 4; Cas4-1 protects DNA fragments with NGG PAM. A scatter plot shows the relative abundance of integrated 57 bp duplexed DNA oligonucleotide detected in a genome-wide adaptation assay. Most new spacers captured in the assay were derived from genome or plasmid DNA; the percent of new spacers corresponding to processed 57 bp oligo is indicated on the Y-axis (out of total spacers). The version of 57 bp oligo used is indicated along the X-axis. Above, the strain in which the experiment was conducted is shown (*cas-OE*, Δ*cas4-1*, Δ*cas4-2*, Δ*cas4-1/*Δ*cas4-2*). Pooled data from eight experiments are presented for wildtype oligo and mutated PAM oligo, and data from four experiments are presented for other samples. In this assay, the 57 bp duplexed DNA oligonucleotide was added at the beginning of the experiment. Its abundance, in proportion to cellular DNA with continuous turnover, is presumed to reflect a combination of both its stability and integration efficiency.

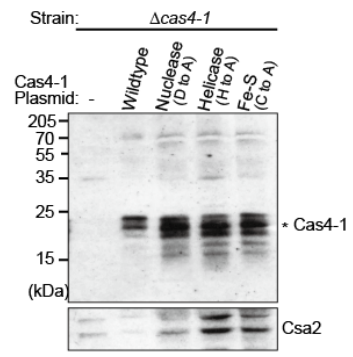
A

```
Cas4-1 PF1119     1 ME-----------------------------------------EYALEELFVRGIEINYLYVCPTKLWFFTHGIT---MEQESEWVDLGKFLH 49
Cas4-2 PF1793     1 ------------------------------------------------MIEFYASEALICPRRVWFRIKGFPEKWPEIAIPRLNKGVKTH 42
       SSO0001    1 M----------ITE--------------------FLLKK-KLEEHLSHVKEENTIYVTDLVRCPRRVRYESEYKELAISQVYAPSAILGDILH 62
       SSO1449    1 M-------------------------------------------------VSVTDLKDYSLCKVIPWIRKKMG---WKEPITNSQRIAKNVN 40
       SSO1392    1 M---------RVIAC--------------------FIRFV-----TLNVLMTISGITIKHFAYCPQIVRIESMGF----TERVSEAMIEGEQVE 56
       PAE1763    1 MELYLPRPLCTVVNCKDLEHLDVDTALTEMKKEQEVFRLLPDVYAYRYDFKRLSPSVINDYEYCPRLLWTQYKLGLKLFSIDSAVAIVKGRLLH 94
       Pisl_1772  1 MELLRPRPICRLINCDDLENLDAEGALEELKKEQEIFRLFPDIYAYRYDFKRISPSVINDFEYCPRLLWVQHKLRLKLFTKGSAVSLIRGRLLH 94
       Pcal0546   1 MELLSPKPLCSVVNCEDLEKLDHVSALNELRREQEIFKLLPGIYAHRYDFRRVSPSIINDFEYCPRLLWVQHKLGLKLLSEKSVVSIIRGRILH 94
                                                                                 Fe-S cluster

                                                  Nuclease
                                                  (D to A)
                                                      *                                  Helicase
                                                                                         (H to A)
                                                                                             *
       PF1119     EQRYENE-------------EKEVLIG--------PIRIDFISRG--DVIEVHEVKLGKR----MERAHEMQALYYLYYLKKHGIK-AKAILHY 115
       PF1793     EVMGKIL-RDEFGF----ELEKHMILRFPRLGFEIHGRMDAYK------EFPIEIKGKSILPRFPLEYHLAQLNIYLRWAEAE-----RGFIYY 120
       SSO0001    LGLESVL-KGNFNAETEVETLREINVGGKV--YKIKGRADAIIRNDNGKSIVIEIKTSRSDKGLPLIHHKMQLQIYLWLFSAE-----KGILVY 148
       SSO1449    L--KEIV-RDFPE-----PKYYEVHLRDKD--SGLNGIVDVLTE-----DSVIEIKAFYRR---FYNHFRIQLLAYAYLAERNGFR-VRNTVLL 115
       SSO1392    KDKVMNFLYATLK-----PLNVVGKPIFRY--GDLIGSPDYVLIYP-NHWVPLDVKSGRK-----RYDHKLQMKYYLYLMDMNGINVKEGLLYY 137
       PAE1763    ERYERAV-SIFDN--------VIAEYKIEI--GDLVGVVDIVMRKG-NEYIPVEIKTGLA----SREAHRRQLQIYIHMLKAR-----YGYLVY 167
       Pisl_1772  ERYERAV-SVFDN--------VIIEYKVEI--GDLVGVIDVVVKKG-RNLIPVEIKAGTA----TREAHRRQLQIYIHLLNAK-----YGYLVY 167
       Pcal0546   ERYERLL-SQYEN--------VVAEYKVEI--GDLVGVVDLVIKRG-GEYIPVEIKTGF-----SKEAHKTQLQIYISMLKAR-----FGYLVY 166
                                                  RecB like motif-II          motif-III          QhxxY motif

                                                                                                 Fe-S
                                                                                                 (C to A)
                                                                                                     *
       PF1119     PKLN----------------------EKREVTLEGRE-REVEEAIVEVEVIKSSSKPPKPRK-SKICRKCAYKELCWG---------- 169
       PF1793     VKLHDNPERIVESIDFRGFPRIKGKHLRVFEVPYDGTLFKETVKFFYELKLYIKEDSLPPAWK-GPSCKFCPYNYLCSQSFLS--FYSEKTL 209
       SSO0001    ITPD----------------------RIAEYEINEPL--DEATIVRLAEDTIMLQNSPRF---NWECKYCIFSVICPAKLT---------- 202
       SSO1449    MNKE----------------------RKINIEVRKEHVEYIGKVVNELVESLEDDSPPVVNPSPLLCKACQYRRVCSVSVLL---------- 175
       SSO1392    VSLK----------------------EMVRLEYNYAEKRYVEKVLSKIREAINGKVRVVQD--ARKCYNCGFFVYCKPKIKGNLAYVD--- 201
       PAE1763    RNRV----------------------EKVERD------SKAVEMLYKIKSVLQMDKPP-----NVNCKSCPFKYICKNVL---------- 214
       Pisl_1772  RNRV----------------------EIVERD------DNALKLLKEIRQILESDKP----PENKKCHSCVFDIICKNILGEI-------- 218
       Pcal0546   RNHV----------------------EVVHRN------DAALDVLKKIREILSAREAP-----PAKCNSCIFKPICKNLL---------- 213
                                                                                 Fe-S cluster
```

B



C

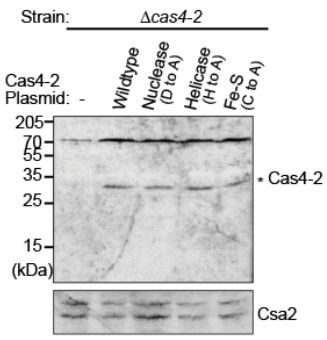


Figure S6. Related to Figure 6; Sequence alignment of Cas4 proteins from *P. furiosus* and other organisms. (A) The Rec-B like motif residues (II: E; III: EhK, and QhXXY) and conserved histidine predicted to support helicase activity(Lemak et al., 2013; Lemak et al., 2014; Zhang et al., 2012) are highlighted in yellow. We note that mutation of this conserved histidine residue in *Sulfolobus solfataricus* does not completely inactivate helicase activity (Lemak et al., 2013). The conserved cysteine residues involved in the coordination of the Fe-S cluster (Lemak et al., 2013; Lemak et al., 2014; Zhang et al., 2012) are highlighted in blue. Residues mutated to alanine are marked with asterisks. The proteins compared are: PF1119 (UniProt ID Q8U1T6), PF1793 (Q8U027) from *Pyrococcus furiosus*, SSO0001 (Q97TX9), SSO1449 (Q97Y86), SSO1392 (Q97YD3) from *Sulfolobus solfataricus,* PAE1763 (Q8ZWJ5) from *Pyrobaculum aerophilum*, Pisl_1772 (A1RV91) from *Pyrobaculum islandicum*, Pcal_0546 (A3MTK6) from *Pyrobaculum calidifontis*. (B, C) Western blot of *P. furiosus*. Csa2 is used as a loading control. The indicated antibodies detect a major band (indicated with asterisk) of the expected molecular weight for each antigen. (B) Δ*cas4-1* strains with Cas4-1 and Csa2 antibodies. The predicted molecular weight of Cas4-1 is 20.2 kDa. (C) Δ*cas4-2* strains with Cas4-2 and Csa2 antibodies. The predicted molecular weight of Cas4-2 is 25.1 kDa.

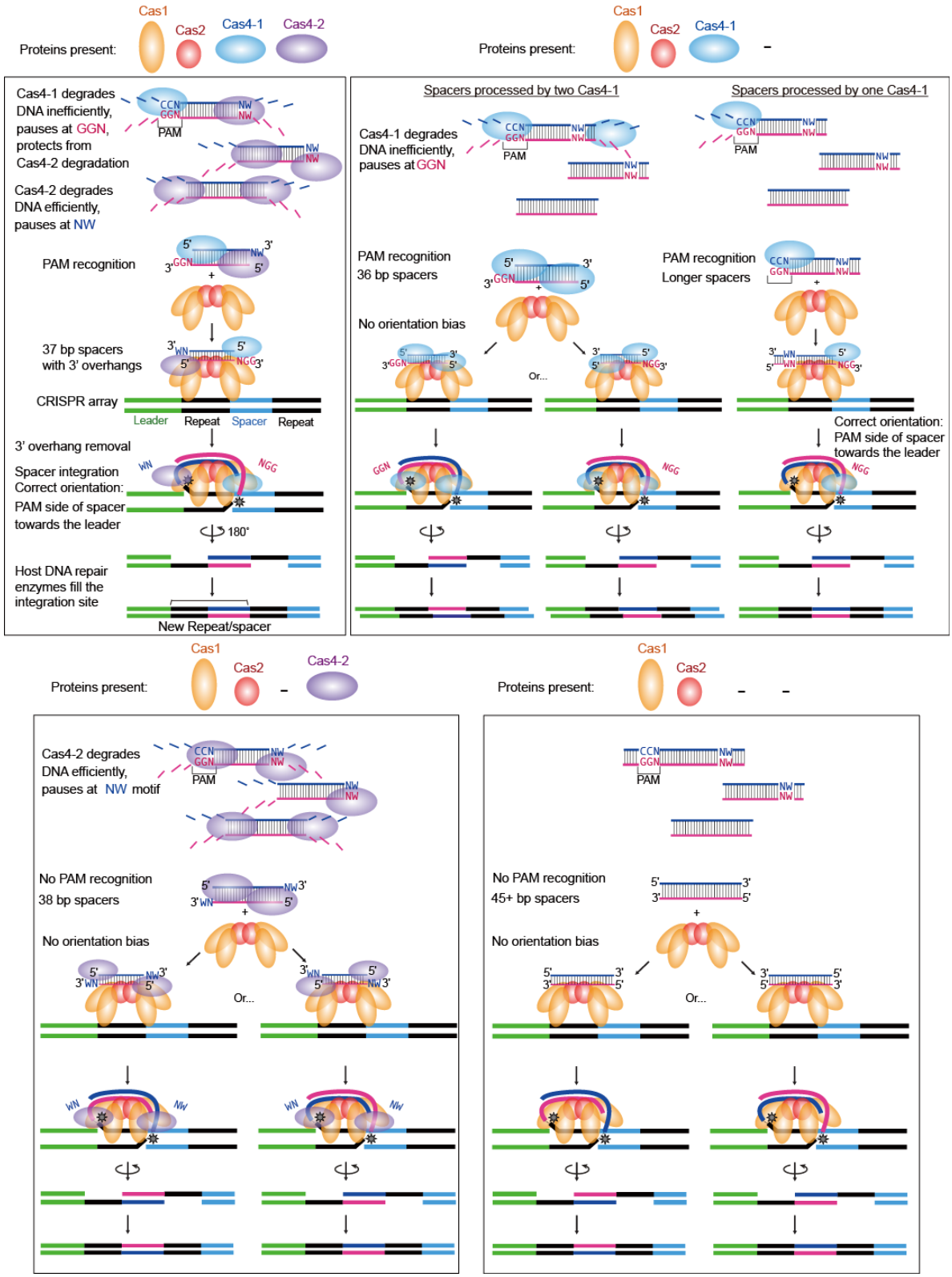

Proteins present:
Cas1
Cas2
Cas4-1
Cas4-2
Cas4-1 degrades DNA inefficiently, pauses at GGN, protects from Cas4-2 degradation
Cas4-2 degrades DNA efficiently, pauses at NW
PAM recognition
37 bp spacers with 3' overhangs
CRISPR array
Leader
Repeat
Spacer
Repeat
3' overhang removal
Spacer integration
Correct orientation:
PAM side of spacer towards the leader
180°
Host DNA repair enzymes fill the integration site
New Repeat/spacer
Spacers processed by two Cas4-1
Spacers processed by one Cas4-1
Cas4-1 degrades DNA inefficiently, pauses at GGN
PAM recognition
36 bp spacers
No orientation bias
Or...
PAM recognition
Longer spacers
Correct orientation:
PAM side of spacer towards the leader
Cas4-2 degrades DNA efficiently, pauses at NW motif
No PAM recognition
38 bp spacers
No orientation bias
Or...
No PAM recognition
45+ bp spacers
No orientation bias
Or...

Figure S7. Related to Figure 7; Model for Cas4 functions in *P. furiosus* spacer acquisition. Top left panel: Cas1, Cas2, Cas4-1 and Cas4-2 can capture and trim spacers with both a PAM and a NW motif and then integrate them into the CRISPR array in the correct orientation. The resulting new spacers can produce crRNAs capable of initiating target interference. Cas4-1 nuclease recognizes the upstream PAM and produces an appropriate 3' DNA overhang. Cas4-2 is an efficient nuclease (see Figure S5) and degrades DNA but spares pre-spacers when they are bound by Cas4-1. Top right panel: Cas1, Cas2 and Cas4-1 can coordinate to capture and trim protospacers with a PAM and integrate spacers in the correct orientation if spacers are processed by one Cas4-1. However, they integrate spacers into the CRISPR array in both incorrect and correct orientations when spacers are processed by two Cas4-1. Bottom left panel: Cas1, Cas2 and Cas4-2 can coordinate to capture and trim spacers with a NW motif. Protospacers are not flanked by a PAM. Bottom right panel: Cas1 and Cas2 alone lead to indiscriminate spacer acquisition with respect to size and PAMs, and resultant spacers cannot generate crRNAs capable of target DNA recognition and interference.