

Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR *de novo* spacer acquisition

Tao Liu¹, Yingjun Li¹, Xiaodi Wang¹, Qing Ye¹, Huan Li¹, Yunxiang Liang^{1,2}, Qunxin She³ and Nan Peng^{1,2,*}

¹State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, P. R. China, ²Hubei Collaborative Innovation Center for Industrial Fermentation, Wuhan 430070, P. R. China and ³Archaeal Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark

Received November 3, 2014; Revised December 22, 2014; Accepted December 22, 2014

ABSTRACT

Acquisition of *de novo* spacer sequences confers CRISPR-Cas with a memory to defend against invading genetic elements. However, the mechanism of regulation of CRISPR spacer acquisition remains unknown. Here we examine the transcriptional regulation of the conserved spacer acquisition genes in Type I-A of *Sulfolobus islandicus* REY15A. *Csa3a*, a MarR-like transcription factor encoded by the gene located adjacent to *csa1*, *cas1*, *cas2* and *cas4* cluster, but on the reverse strand, was demonstrated to specifically bind to the *csa1* and *cas1* promoters with the imperfect palindromic sequence. Importantly, it was demonstrated that the transcription level of *csa1*, *cas1*, *cas2* and *cas4* was significantly enhanced in a *csa3a*-overexpression strain and, moreover, the Csa1 and Cas1 protein levels were increased in this strain. Furthermore, we demonstrated the hyperactive uptake of unique spacers within both CRISPR loci in the presence of the *csa3a* overexpression vector. The spacer acquisition process is dependent on the CCN PAM sequence and protospacer selection is random and non-directional. These results suggested a regulation mechanism of CRISPR spacer acquisition where a single transcriptional regulator senses the presence of an invading element and then activates spacer acquisition gene expression which leads to *de novo* spacer uptake from the invading element.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR) and their associated genes (*cas*) constitute the CRISPR-Cas immune system, which takes up DNA from invasive genetic elements to serve as a memory source that provides adaptive defense against mobile genetic elements (1,2). The CRISPR-Cas system is widespread and is found in about 90% of archaeal and 40% of bacterial genomes (3). The CRISPR loci consists of identical repeats (typically 20–50 bp in length) interspaced by variable spacer sequences of similar size (4). Repeats are usually conserved in sequence within a given locus and often contain partially palindromic sequences (5). Spacers, in contrast, are highly diverse in sequence and originate from mobile genetic elements (4,6–8), which led to the hypothesis that CRISPRs provide immunity against invading genetic elements (9). An AT-rich leader sequence located upstream of the first repeat promotes transcription of the CRISPR locus (10,11). The *cas* genes encode a diverse family of proteins carrying nuclease, helicase and potential polymerase activities (9,12), and based on the diversity of Cas proteins, CRISPR-Cas systems have been clustered into three basic types: Type I, Type II and Type III, which are further divided into at least 11 subtypes (13). Although most *cas* genes are highly divergent and are only associated with certain CRISPR loci, *cas1* and *cas2* are notably conserved in sequence across the three major types of CRISPR systems (2).

The mechanism by which CRISPR-Cas provides immunity against mobile genetic elements can be dissected in three steps: (i) incorporation of an invading nucleic acid sequence into the CRISPR locus; (ii) CRISPR-RNA (crRNA) transcription and maturation; and (iii) target interference by crRNA and a protein complex of Cas proteins, which is guided by crRNA base-pairing (14,15). Although the processes of crRNA biogenesis and interference are well characterized, spacer acquisition mechanisms re-

*To whom correspondence should be addressed. Tel: +86 27 8728 1267; Fax: +86 27 8728 0670; Email: nanop@mail.hzau.edu.cn

main poorly understood. The first successful demonstration of spacer acquisition under laboratory conditions was in the *Streptococcus thermophilus* Type II-A system. In this study, addition of new repeat-spacer units were found and the new spacer perfectly matched the sequence of the challenging phage (1). Subsequent studies identified more spacer acquisition events in different CRISPR-Cas types, including *Escherichia coli* Type I-E (16–18), *Pseudomonas aeruginosa* Type I-F (19), *Sulfolobus solfataricus* Type I-A (20,21), *Haloarcula hispanica* Type I-B (22,23) and *Pectobacterium atrosepticum* Type I-F (24). A spacer acquisition assay demonstrated that Cas1 and Cas2 are the proteins required for new spacer adaptation at the host CRISPR locus (17,22,25,26). A protospacer-adjacent motif (PAM) of efficient protospacer-mediated spacer acquisition was identified in *E. coli* and *H. hispanica* (18,23). Moreover, other DNA motifs, such as CRISPR leader sequences and priming spacers, play important roles in spacer acquisition (17,23,25,27–29). Recently it was found that the presence of defective phages favors spacer acquisition in *S. thermophilus* (30). It is worthy to know Heler *et al.* have reviewed the current knowledge of the CRISPR spacer acquisition mechanism (31). Although transcriptional regulation of CRISPR array by Cbp1 in *Sulfolobus* (32) and *cas* genes in bacteria by a cyclic AMP receptor protein, heat-stable nucleoid-structuring protein (H-NS) and LeuO, a LysR-type transcription regulator (10,33–36) have been reported, the transcriptional regulation of the most conserved genes *cas1* and *cas2*, which are involved in new spacer acquisition, needs much to be learn.

The archaeon *Sulfolobus islandicus* REY15A encodes one adaptation module and three CRISPR interference modules, one of Type I-A and two of Type III-B (37). The CRISPR-Cas system in this genus has recently been investigated with respect to the dispensability of its *cas* genes for proper CRISPR function (38), dynamic properties of CRISPR loci after plasmid challenge (39) and the mechanism of transcription-dependent DNA interference (37). In this study, we present *in vivo* and *in vitro* evidence for Csa3a activating the transcription of *cas* genes that are involved in new spacer acquisition (*acas* genes), thereby triggering spacer acquisition within both CRISPR loci in *S. islandicus*.

MATERIALS AND METHODS

Strains and growth conditions

Sulfolobus islandicus E233S (Δ *pyrEF* Δ *lacS*) was cultured in SCVy medium at 78°C (40). Electroporation was used to transform *S. islandicus* E233S, and transformants were selected on 2-layer phytal gel plates as described previously (40). *E. coli* DH5 α and Rosetta cells were used for DNA cloning and recombinant protein production. All *E. coli* strains were cultured at 37°C in Luria-Bertani medium and ampicillin or kanamycin was added to reach a 100 μ g/ml or 30 μ g/ml final concentration as required.

General DNA manipulation methods

Restriction and DNA modification enzymes were purchased from New England Biolabs or Fermentas. Plasmid

DNA was extracted from *E. coli* or *Sulfolobus* cells using an AxyPrep Plasmid Miniprep Kit (Wujiang, China). Polymerase chain reaction (PCR) products were purified by using an Axygen PCR Clean-up Kit, and DNA bands fractionated from the agarose gel were extracted by using an Axygen DNA extraction kit. Total DNA was prepared from *Sulfolobus* by using an Axygen Genomic DNA Miniprep Kit. The oligonucleotides were synthesized by Invitrogen (Shanghai, China) and were also used for DNA sequencing.

Protein expression and purification

The *csa3a*, *csa1*, *cas1*, *cas2* and *cas4* genes were amplified from *S. islandicus* REY15A genomic DNA using the primers listed in Supplementary Table S1, and cloned into *E. coli* expression vector pET30a or *Sulfolobus* expression vector pSeSD (41). After amplification in *E. coli* DH5 α , the expression plasmids were transformed into *E. coli* Rosetta cells or *S. islandicus* E233S. For protein expression in *E. coli*, the recombinant proteins were induced at 20°C for 8 h by addition of 0.8 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at the final concentration. After induction, cells were harvested and sonicated. The soluble proteins in the supernatants collected by centrifugation at 12 000 \times g were purified by using a nickel matrix. After dialysis against 10 mM Tris–HCl buffer (pH 8.0), the proteins were either used immediately or stored at –80°C. The protein concentration was determined by using the MicroBCA kit (Pierce; Rockford, IL, USA) according to the manufacturer's instructions.

Electrophoretic mobility shift assay

Electrophoretic mobility shift assay (EMSA) probes were generated by PCR or annealing using the oligonucleotides with one of the primer pair 5'-end biotin-labeled (Supplementary Table S1). Then the products were purified from 6% native polyacrylamide gel electrophoresis (PAGE). The EMSA binding reactions (20 μ l) containing 10 pmol of biotin-labeled probes and different concentrations of Csa3a, as described in the figure legends, were incubated for 20 min at 40°C in the binding buffer (20 mM Tris–HCl, pH 8.0, 50 mM KCl, 5% glycerol, 1 mM ethylenediaminetetraacetic acid, 1 mM dithiothreitol, 5 ng/ μ l poly(dI-dC)). For specific competition, increasing amounts of unlabeled specific probes were added to the reaction mixture. After the reaction, samples were loaded onto a 6% native PAGE gel buffered with 1 \times TBE solution. DNA–protein complexes were separated at 100 V for 90 min and then transferred to a polyvinylidene fluoride membrane (Bio-Rad, Hercules, CA, USA) using the Semi-Dry Electrophoretic Transfer Cell system (Bio-Rad). Bands were visualized by chemiluminescent detection using the clarity Western ECL substrate (Bio-Rad, Hercules, CA, USA) and the MF-Chemibis 3.2 imaging device (DNR; Jerusalem, Israel).

DNase I footprinting assay

A DNase I footprinting assay was performed as described previously (42) with slight modifications. The *csa1* and *cas1*

promoter sequences were amplified from genomic DNA using primer sets *csa1F-csa1R* (−205 to −3, related to translational start codon) and *cas1F-cas1R* (−143 to +97), and subcloned into pMD18T (Takara, Dalian, China). The promoter DNA sequences were PCR-amplified from the pMD18T plasmids carrying *csa1* and *cas1* promoter sequences using primers with either 5′-end 6-HEX-labeled M13F (for the coding strand) or 5′-end FAM-labeled M13R (for the non-coding strand, Supplementary Table S1), and purified by 6% PAGE. In a 60 μl reaction system, 100 ng of a labeled DNA fragment was bound to 100 μg Csa3a (bovine serum albumin was used instead of Csa3a in the control experiment) in the buffer containing 10 mM Tris-HCl, pH7.4, 10 mM MgCl₂, 1 mM CaCl₂, 0.4 mM dithiothreitol, 100 mM KCl and 5% glycerol, and incubated for 20 min at 40°C. After binding, 0.03 U of RNase-free DNase I (Roche, Basel, Switzerland) was added and allowed to react for 5 min at 30°C. The reaction was stopped and precipitated with ethanol. Samples were analyzed in a 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA) and the electropherograms were aligned with GeneMapper v3.5 (Applied Biosystems).

Chromatin immunoprecipitation assay

A chromatin immunoprecipitation (ChIP) assay was performed as described previously (43) using rabbit polyclonal antibody against Csa3a. The DNA from ChIP samples was extracted and resolved in 50 μl TE buffer. Input samples were treated following the same procedure without addition of antiserum and Protein A + G agarose beads (Beyotime, Beijing, China). Recovered DNA was PCR-amplified using primers specific for the *csa1*, *cas1* and *lrs14* (SiRe_0993) promoters (Supplementary Table S1). The recovered DNA was further analyzed by quantitative PCR. Each quantification reaction contained 1 μl recovered DNA, 1 μl (10 μM) gene-specific forward and reverse primers (the same primers as used above) and 10 μl SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA, USA) in a final volume of 20 μl. The reactions were run on a LightCycler 480 (Roche, Basel, Switzerland). The *lrs14* promoter sequence quantification was used as negative control. The qPCR conditions were as follows: 30 s of enzyme activation at 95°C, followed by denaturing at 95°C for 5 s and annealing/extension at 60°C for 20 s. The samples were cooled to 50°C and then heated to 99°C, and the melting curves were determined. Data analysis was performed using the software Roche LightCycler 480 HTC1. The ChIP assay were done in triplicate.

Construction of a *S. islandicus csa3a*-deletion mutant

The gene deletion method recently developed for *S. islandicus* REY15A (44) was employed to knockout the *csa3a* gene in this study. The left sequence arm (L-arm), right sequence arm (R-arm) and target gene arm (G-arm) were amplified from *S. islandicus* REY15A using the primer set shown in Supplementary Table S1. The marker cassette sequence carrying the *pyrEF* and *lacS* genes was amplified from the *Sulfolobus-E. coli* shuttle vector pHZ2lacS(40) using the primers M-F-BamHI and M-R-SalI. All the fragments were subsequently cloned onto vector pUC19, result-

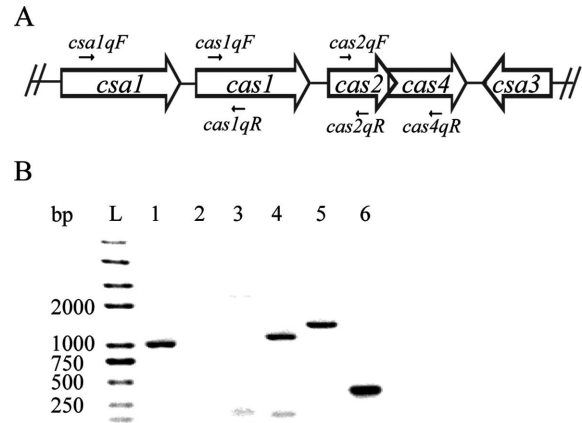


Figure 1. Determination of the co-transcription of the *csa1*, *cas1*, *cas2* and *cas4* genes. (A) Organization of *csa1* (852 bp), *cas1* (873 bp), *cas2* (270 bp) and *cas4* (528 bp) genes is depicted schematically. The locations of primers are indicated that were used for RT-PCR to determine co-transcription. The intergenic regions between *csa1* and *cas1* is 42 bp and between *cas1* and *cas2* is 110 bp; *cas2* and *cas4* are overlapped. (B) Fragments obtained in RT-PCR experiments carried out with the primers *csa1qF* and *cas1qR* (lane 1, product 1016 bp), *csa1qF* and *cas2qR* (lane 2), *csa1qF* and *cas4qR* (lane 3), *cas1qF* and *cas2qR* (lane 4, product 1151 bp), *cas1qF* and *cas4qR* (lane 5, product 1418 bp), and *cas2qF* and *cas4qR* (lane 6, product 376 bp). A DNA ladder (Trans 2K plus II) was run in lane L as a size marker.

ing in the *csa3a* gene-deletion plasmid pDel*csa3a* (Supplementary Figure S1). The selection of *csa3a*-deletion mutants using this plasmid was carried out as described previously (44).

Total RNA preparation

Total RNA was isolated from exponentially growing *Sulfolobus* cultures in SCVy medium or ACVy medium for induction of *csa3a* gene under control of *araS* promoter (OD₆₀₀ = 0.2) as described previously (45). Genomic DNA in the total RNA sample was removed using DNase I (Roche, Basel, Switzerland).

Reverse transcription (RT)-PCR and real-time RT-PCR (RT-qPCR) assay

To determine the co-transcription of the *csa1*, *cas1*, *cas2* and *cas4* genes, three gene-specific reverse primers (*cas1qR*, *cas2qR* and *cas4qR*; Supplementary Table S1, Figure 1) were used to generate the first-strand cDNA from the total RNA sample. Three forward primers (*csa1qF*, *cas1qF* and *cas2qF*; Supplementary Table S1, Figure 1) in combination with the three reverse primers listed above were used to amplify the first-strand cDNAs (Figure 1). The PCR products were separated on a 1% agarose gel. For real-time reverse transcription-PCR (RT-qPCR), first-strand cDNA was synthesized using M-MuLV reverse transcriptase (Promega, Madison, WI, USA) and the specific reverse primers (*csa1qR*, *cas1qR*, *cas2qR* and *cas4qR*; Supplementary Table S1). Each real-time quantification reaction was carried out as described above for the ChIP-qPCR assay using the first-strand cDNAs as template and each forward primer (*csa1qF*, *cas1qF*, *cas2qF*, *cas4qF* or *csa3aqF*) in combination with the reverse primers listed above along

with *csa3a*qR. The transcripts of the *albA* gene were used as the control (45) and the cycle threshold (Ct) values of the control transcript *albA* were used to normalize the Ct values of the *csa1*, *cas1*, *cas2*, *cas4* and *csa3a* transcripts.

Western blotting

Sulfolobus transformants were cultured in ACVy medium to induce expression of *csa3a* gene under control of *araS* promoter. When OD₆₀₀ reached 0.2, cells were harvested and sonicated. Crude protein samples were separated by 12% sodium dodecyl sulphate (SDS)-PAGE, and then transferred to a nylon membrane using the Semi-Dry Electrophoretic Transfer Cell system (Bio-Rad; Hercules, CA, USA). The target proteins were detected by rabbit polyclonal antibodies against Csa1, Cas1 and Csa3a, and then bound by horseradish peroxidase-labeled goat anti-rabbit antibody (Beyotime, Beijing, China). Bands were visualized by chemiluminescent detection using the clarity Western ECL substrate (Bio-Rad; Hercules, CA, USA) and the MF-Chemibis 3.2 imaging device (DNR; Jerusalem, Israel).

PCR amplification, cloning and sequencing of leader proximal CRISPR region

Wild-type *S. islandicus* E233S, E233S carrying an empty expression vector and a *csa3a*-overexpression strain were cultured in 10 ml of ACVy medium at 78°C for 10 days. Samples of each culture (0.1 ml) were taken every 24 h and cells were used for PCR template. The leader proximal regions of two CRISPR loci were amplified by PCR using Taq polymerase and forward primer CRISPR-F and reverse primer CRISPR2S5-R for locus 2; and forward primer CRISPR-F and reverse primer CRISPR1S5-R for locus 1. PCR products were separated on 2% agarose gel and visualized by ethidium bromide staining. Bands larger than those of the wild-type control sample were excised from the gel and purified by Axygen DNA extraction kit. Purified PCR products were cloned into T-vector (Takara, Dalian, China) following the manufacturer's instruction, then the ligation products were transformed in to *E. coli* DH5α. Plasmids from single colonies were purified and sequenced at Invitrogen (Shanghai, China).

RESULTS

Co-transcription of the *csa1*, *cas1*, *cas2* and *cas4* genes in *S. islandicus*

The *S. islandicus* REY15A genome encodes a single gene cassette containing the *csa1* (SiRe.0760), *cas1* (SiRe.0761), *cas2* (SiRe.0762) and *cas4* (SiRe.0763) genes, in which *cas1* and *cas2* have been implicated in spacer acquisition in this archaeon (Figure 1A). To investigate the function of these *cas* genes and the regulation of their gene expression, total RNA was prepared and the co-transcription of these genes was studied. A 1016-bp DNA fragment was amplified using primers *csa1*qF and *cas1*qR from the first-strand cDNA, which was synthesized using *cas1*qR as a primer from total RNA (Figure 1B). This result indicated that the *csa1* and *cas1* genes were co-transcribed under the control

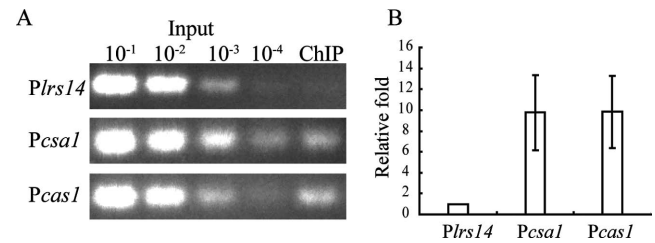


Figure 2. *In vivo* binding of Csa3a to the *csa1* and *cas1* promoters. (A) Chromatin immunoprecipitation (ChIP) assay using antiserum against Csa3a. Input samples were 10⁻¹, 10⁻², 10⁻³ and 10⁻⁴ of the input sample amounts described in the 'Materials and Methods' section. DNA recovered from immunoprecipitates was amplified with primers specific for *csa1* and *cas1* promoters, or the *lrs14* promoter as a control. (B) The ChIP-qPCR result revealed that *csa1* and *cas1* promoter sequences in the recovered DNA were ~10-fold relative to the *lrs14* promoter sequence.

of the *csa1* promoter. However, no PCR products were amplified using *csa1*qF with *cas2*qR or *cas4*qR from the first-strand cDNAs that were synthesized using primers *cas2*qR and *cas4*qR (Figure 1B). This result showed that the *cas2* and *cas4* genes were not co-transcribed with *csa1*. Interestingly, three DNA fragments (1151bp, 1418 bp and 376 bp) were amplified from the first-strand cDNAs using primer pairs *cas1*qF-*cas2*qR, *cas1*qF-*cas4*qR and *cas2*qF-*cas4*qR, which were synthesized using the oligonucleotides *cas2*qR and *cas4*qR from total RNA. These results strongly indicated that *cas1*, *cas2* and *cas4* were cotranscribed. This co-transcription results are consistent with promoters of *csa1* and *cas1* controlling the transcription of this operon. Thus, the operon was named adaptation-associated *cas* operon (*acas* operon) according to the aCas complex annotated by Garrett *et al.* (46).

Csa3a bound both *csa1* and *cas1* promoters *in vivo*

The *csa3a* gene (SiRe.0764), encoding a putative transcriptional regulator, is adjacent to the *acas* operon but in an inverse orientation in the *S. islandicus* REY15A chromosome (Figure 1A). This is consistent with Csa3a functioning as a transcriptional regulator for the *acas* genes. Therefore, a ChIP assay was conducted to test for an interaction between Csa3a and *acas* promoters. The results in Figure 2A demonstrate that Csa3a bound specifically to the *csa1* and *cas1* promoters. Moreover, no PCR product for the promoter of *lrs14*, a transcriptional regulator gene, was found in the negative control experiment. The recovered DNA sample from the ChIP assay was quantified by real-time PCR for the *csa1*, *cas1* and *lrs14* promoters. The recovered DNA contained ~10-fold more DNA of both the *csa1* and *cas1* promoters than of the control *lrs14* promoter (Figure 2B). The results confirmed that Csa3a bound to both *csa1* and *cas1* promoters, suggesting that Csa3a is involved in transcriptional regulation of the *acas* operon.

Localizing the Csa3a-binding sequence in the *csa1* promoter

To study the regulation of the *csa1* promoter by Csa3a, the promoter was used as the probe in an EMSA experiment. In an archaeal promoter, the hexameric TATA box is centered at -26/-27 with respect to the transcription start site

and BRE (transcription factor B recognition element) located immediately upstream of the TATA box. Thus, the promoter fragment P1 (positions -205 to -4 relative to the translational start codon) and two 5'-truncated elements P2 (positions -118 to -4) and P3 (positions -42 to -4) were used in the EMSA experiment to determine the Csa3a-binding site (Figure 3A). Both P1 and P2 DNA fragments produced strong band shifts with Csa3a on the gel, whereas no such retarded band was observed when the P3 fragment was incubated with Csa3a (Figure 3B). We inferred that the binding site for Csa3a is located between positions -118 and -42 on the *cas1* promoter relative to the translational start codon.

Identification of the minimal region of Csa3a binding in the *cas1* promoter

To determine the precise binding region of Csa3a on the *cas1* promoter, a footprinting assay was carried out and the region from -81 to -42 of both strands of the promoter sequence was shown to be protected by Csa3a (Figure 3C). This sequence exhibited an imperfect inverted repeat close to the BRE and TATA box (Supplementary Figure S2). The 25-bp fragment was then used as a probe in the EMSA experiment to test further the DNA binding capacity of Csa3a. The results show that increasing the amount of Csa3a protein enhanced the signal of the retarded bands (Figure 3D) and, moreover, that increasing the amount of the unlabeled specific probe selectively reduced the signal of the retarded bands (Figure 3E). Therefore these results define the minimal region for Csa3a binding on the *cas1* promoter.

Identification of two DNA elements essential for Csa3a binding in the *cas1* promoter and adjacent sequence

Since Csa3a binds both *cas1* and *cas1* promoter sequences *in vivo* and *cas1* and *cas1* are transcribed separately, we decide to analyze the interaction between *cas1* promoter and Csa3a. In the region between positions -35 to $+45$ of the *cas1* promoter (relative to the translational start codon, TTG), and its adjacent sequence, two DNA fragments were found by alignment that were similar in sequence to the minimal sequence required for Csa3a binding on the *cas1* promoter (Figure 4A). Thus, a full-length sequence (P1, -35 to $+45$), an upstream DNA element (P2, -35 to -10) and a downstream element (P3, $+1$ to $+45$) were used as probes in the EMSA experiment to test for binding of Csa3a (Figure 4A). At increasing concentrations of Csa3a, the signals of the P1 retarded bands became stronger, while increasing amounts of the cold probe reduced the signals of the retarded bands (Figure 4B). These results indicated that Csa3a specifically bound to the *cas1* promoter and its adjacent sequence. However, the EMSA results showed that increasing the amount of Csa3a did not retard the P2 fragment (Figure 4C) and that Csa3a was only weakly bound to the P3 fragment (Figure 4D). Csa3a weakly retarded P4 and P5 (P4: P1 with deletion of P2 element; P5: P1 with deletion of P3 element), but did not retard the internal sequence itself (Supplementary Figure S3). Thus, in contrast to the Csa3a-binding to the *cas1* promoter, efficient binding

of Csa3a to the *cas1* promoter requires both P2 and P3 sites and the internal element between P2 and P3.

A footprinting assay using a fragment covering positions -143 to $+97$ of the *cas1* promoter sequence, was performed to determine the precise Csa3a-binding sites. Csa3a protected a large region of the fragment extending from position -35 to $+45$ (Figure 4E). The protected region contained two elements that are highly similar to the Csa3a-protected sequence on the *cas1* promoter (Figure 4E and Supplementary Figure S2). It was revealed that neither site alone was sufficient for Csa3a binding (Figure 4C and D) but that the two regions appeared to act cooperatively in facilitating Csa3a recognition.

Overexpression of *cas3a*-activated transcription of *acas* genes

Although the specific *in vitro* and *in vivo* binding of Csa3a to the *cas1* or *cas1* promoter was confirmed, the specific mechanism of transcriptional regulation of the *acas* operon remained unclear. Therefore, we constructed a *cas3a* gene-deletion mutant ($\Delta cas3a$, Figure 5A) and a *cas3a*-overexpression (OE*cas3a*) strain (Figure 5A). Expression of the *cas3a*, *cas1*, *cas1*, *cas2* and *cas4* genes was then tested in the wild-type (carrying the empty vector pSeSD), $\Delta cas3a$ and OE*cas3a* strains. *Sulfolobus* strains were grown in ACVy medium to $OD_{600} = 0.2$ and then cells were harvested and total RNA and crude protein samples were prepared. The transcripts of the *acas* genes were determined by RT-qPCR using mRNA of the *alba* gene as an internal control(45). The *cas3a* transcription level in OE*cas3a* strain increased by $489(\pm 105)$ -fold compared with the wild-type strain and no cDNA was detected in the $\Delta cas3a$ strain. The transcription levels of *acas* genes were reduced by $1.00(\pm 0.18)$ -, $0.62(\pm 0.08)$ -, $0.84(\pm 0.27)$ - and $0.94(\pm 0.16)$ -fold in the $\Delta cas3a$ strain and increased by $6279(\pm 1583)$ -, $3331(\pm 273)$ -, $1291(\pm 481)$ - and $525(\pm 79)$ -fold in the OE*cas3a* strain, compared with expression in the wild-type strain (Figure 5B). Csa3a was detected neither in wild-type nor in $\Delta cas3a$ strains in a western blot assay, while a strong Csa3a protein band from the OE*cas3a* strain was visualized on the membrane (Figure 5C). Csa1 and Cas1 proteins showed elevated expression in the OE*cas3a* strain and the expression levels in $\Delta cas3a$ strain were similar to that in wild-type strain (Figure 5C). This result revealed that Csa3a acts as a transcriptional activator for the *acas* operon. However, Cas2 and Cas4 proteins were not detected in the wild-type, $\Delta cas3a$ or OE*cas3a* strains by western blotting consistent with low protein levels in the cells. PCNA3 was used as the internal control in western blotting and its expression showed no difference among the wild-type, $\Delta cas3a$ and OE*cas3a* strains. These results clearly indicated that the Csa3a regulator acts as a transcriptional activator for *acas* genes in *S. islandicus*; however, translation of the *acas* mRNA may be controlled by an unknown mechanism.

Activation of *de novo* spacer acquisition by overexpression of *cas3a*

Wild-type (*S. islandicus* E233S), E233S carrying an empty expression vector and *cas3a*-overexpression strains, were cultured in ACVy medium and the latter strain showed

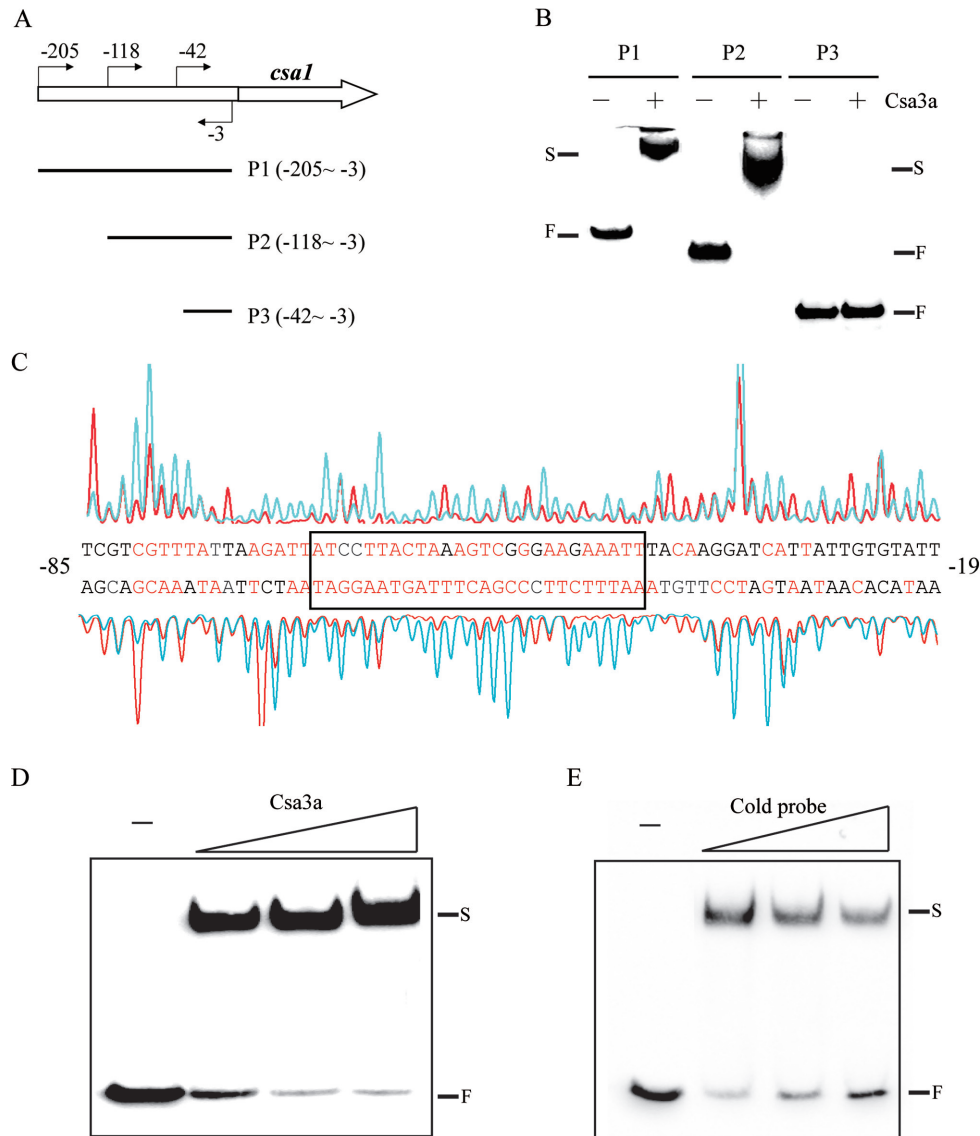


Figure 3. *In vitro* binding of the *csal* promoter by Csa3a. (A) Schematic diagram of the *csal* promoter and the DNA probes used for EMSA. The locations of primers used for amplifying the probes are indicated. (B) EMSA experiments using three truncated probes in the absence or presence of Csa3a (1.6 pmol/ μ l). F: free labeled probe; S: shifted band. (C) DNase I footprinting assay with coding (HEX-labeled, peaks above the promoter sequences) and non-coding (FAM-labeled, peaks below the promoter sequence) strands of the DNA fragments containing the *csal* promoter in the presence (red peaks) and absence (blue peaks) of Csa3a. The protected nucleotides (relative to the translational start codon) are shown in red and the minimal region is boxed. (D) EMSA experiment using the sequence of the boxed region indicated in (C) as a probe with increasing amounts of Csa3a (0.4, 0.8 and 1.6 pmol/ μ l). (E) EMSA experiment in the presence of Csa3a (0.4 pmol/ μ l) using the same probe as in (D) with increasing amounts of unlabeled nucleotides covering the *csal* promoter (molar ratio = 1: 1, 1: 4 and 1: 8).

retarded growth from the beginning of inoculation (data not shown). Larger DNA fragments amplified between the leader sequence and spacer 5 of loci 1 and 2 were observed after 1 day of incubation in the medium (Figure 6A). Two larger fragments were produced by PCR, with the strong band containing one *de novo* spacer and the weak band carrying two spacers for both CRISPR loci (Figure 6A). No expanded bands were detected from E233S and E233S carrying empty vector.

DNA from the larger PCR products was cloned and transformed into *E. coli* and DNA from single colonies was sequenced. This revealed large number of *de novo* spacer se-

quences in each locus (all protospacers were summarized in Supplementary Table S2). For locus 1, 52 unique spacers were acquired with 96.1% of the spacers from the *csa3a* expression plasmid and 3.9% from genomic DNA. For locus 2, 108 unique spacers were acquired with 91.7% of the spacers from the plasmid DNA and the remainder from genomic DNA. Most of the colonies carried only one new acquired unique spacer adjacent to leader sequence and 15 colonies exhibited two and one colony showed four *de novo* spacers (Table 1). The locations of these new spacers did not show significant bias for protein genes or intergenic sequences. However, no spacer was found to originate from

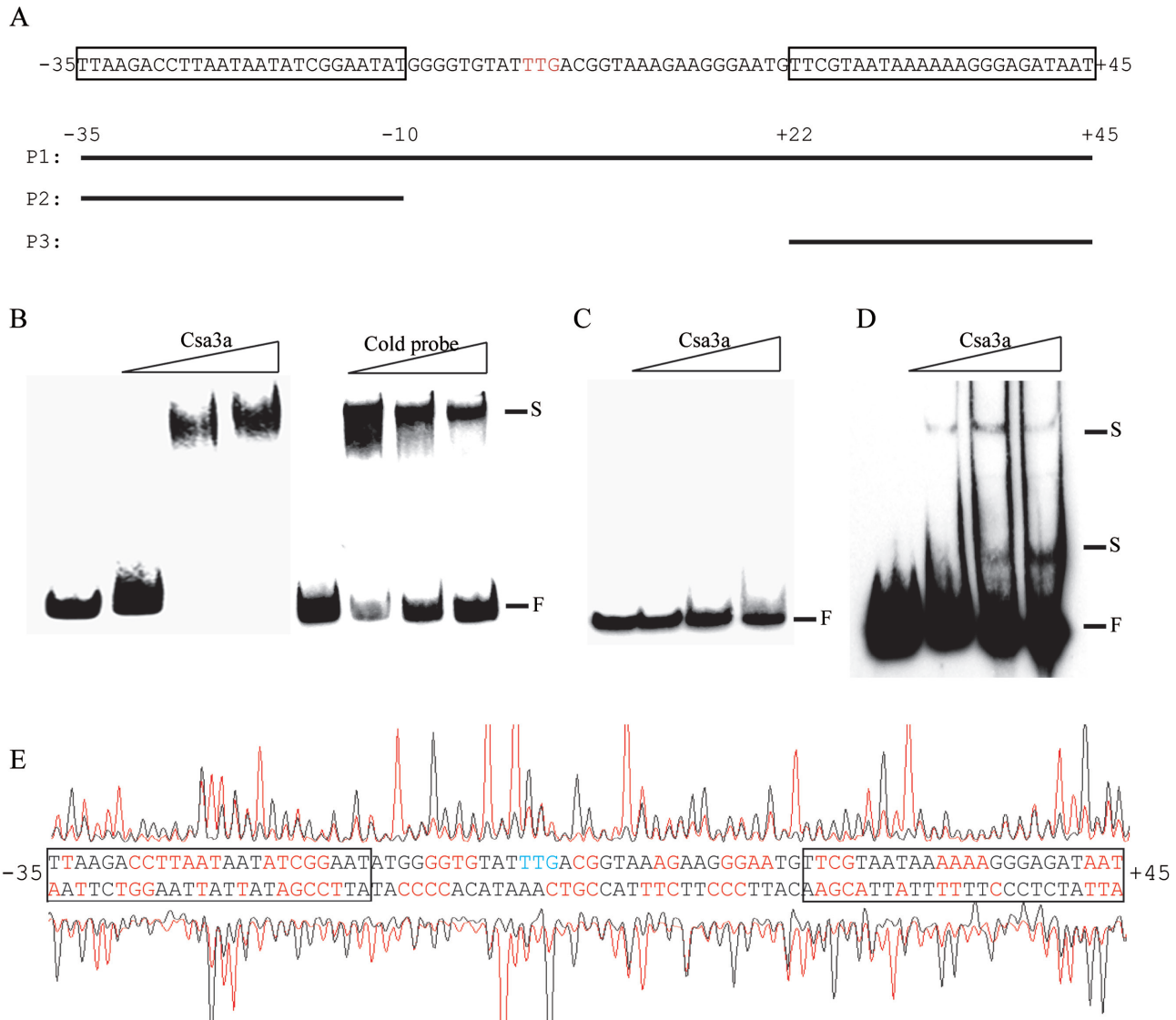


Figure 4. Identification of binding site of Csa3a on the *cas1* promoter and *cas1* coding sequence. (A) Schematic diagram of the *cas1* promoter and the DNA probes used for EMSA. The translational start codon (TTG) is indicated in red. The locations of three probes relative to the translational start codon are indicated. Two regions similar to the minimal binding site of Csa3a on the *cas1* promoter are boxed. (B) EMSA experiments using biotin labeled P1 as a probe with increasing amounts of Csa3a (0.4, 0.8 and 1.6 pmol/ μ l) or an unlabeled P1 cold probe (molar ratio = 1: 1, 1:4 and 1:8). (C and D) EMSA experiments using biotin labeled P2 or P3 as a probe with increasing amounts of Csa3a (2.4, 4.8, 6 pmol/ μ l). F: free labeled probe; S: shifted band. (E) DNase I footprinting assay with coding (HEX-labeled, peaks above the promoter sequences) and non-coding (FAM-labeled, peaks below the promoter sequence) strands of the DNA fragments containing the *cas1* promoter in the presence (red peaks) and absence (black peaks) of Csa3a. The protected nucleotides (relative to the translational start codon) are shown in red and the regions similar to the Csa3a-binding site on the *cas1* promoter are boxed.

the *csa3a* gene on the expression plasmid, which may reflect that strains acquired spacers from *csa3a* will target another copy of this gene on their own genome, thus no survival strains carrying new spacer from *csa3a* gene sequence were detected (Figure 6B and C)(47). Total 11 spacers were found to originate from the host genome. Two protospacers matched to SiRe_0056 gene (inosine/uridine-preferring nucleoside hydrolase gene) and two matched to SiRe_1054 gene (adenylate kinase related protein gene). Each one protospacer matched to SiRe_0560 (major facilitator superfamily MFS_1), SiRe_0654 (4Fe-4S ferredoxin iron-sulfur binding domain protein gene), SiRe_1500, SiRe_1555 (transcription initiation factor TFIIB gene), SiRe_2286 (major facil-

itator superfamily MFS_1), SiRe_2332 (malate dehydrogenase) and SiRe_2393 (mechanosensitive ion channel), respectively. There was no evidence for biased strand directionality for the acquired spacers of both loci (Figure 6B and C; Table 1). Moreover, the PAM sequence CCN was conserved in adjacent to protospacer sequences, and -1 site of PAM was biased to A (Table 1), which was consistent with previous study in *Sulfolobus* (21).

DISCUSSION

In these experiments, we investigated the transcriptional regulation of the *acas* operon, containing *cas1* and *cas2*

Table 1. Analysis of the protospacers

	Locus 1	Locus 2
Protospacer location		
Total	52	108
Plasmid	50 (96.1%)	99 (91.7%)
Genomic DNA	2 (3.9%)	9 (8.3%)
Forward strand	28 (53.8%)	60 (55.6%)
Reverse strand	24 (46.2%)	48 (44.4%)
Within protein gene	40 (76.9%)	72 (66.7%)
Intergenic	12 (23.1%)	36 (33.3%)
PAM sequences		
CCN	35 (67.3%)	80 (74.1%)
CCA	12 (23.1%)	35 (32.4%)
CCT	12 (23.1%)	20 (18.5%)
CCG	6 (11.5%)	15 (13.9%)
CCC	5 (9.6%)	10 (9.3%)
No PAM	17 (32.7%)	28 (25.9%)
Single and multiple spacer insertion		
Single	42	84
Two	3	12
Four	1	

A total of 160 unique spacers inserted into loci 1 and 2 were sequenced and analyzed. The numbers of protospacers matching the expression plasmid and genomic DNA, as well as the protospacers carrying PAM sequences (5'-PAM-protospacer-3'), are given. Percentages are included in brackets.

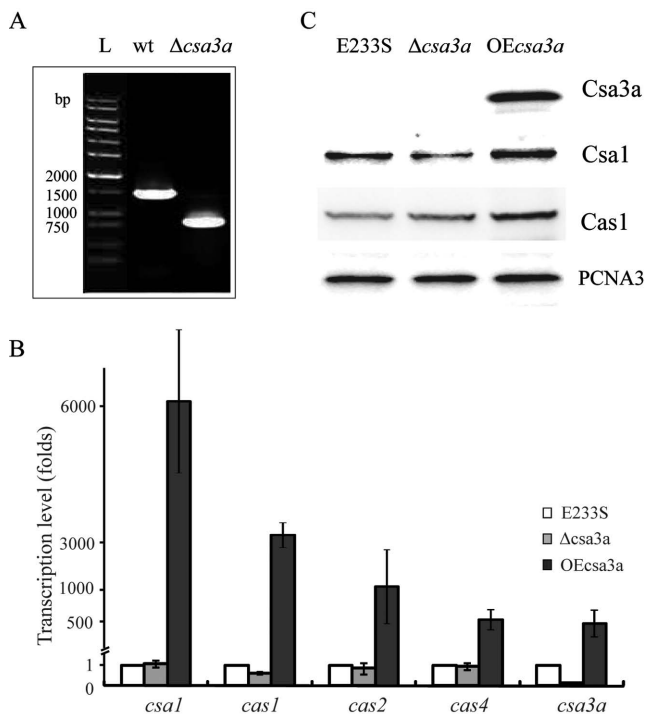


Figure 5. Effect of Csa3a on the expression of *acas* genes. (A) PCR verification of the *csa3a*-deletion mutant. PCR products obtained using specific primers located upstream and downstream of *csa3a* gene from genomic DNAs of *Sulfolobus islandicus* E233S (wild-type) and the *csa3a*-deletion strain (Δ *csa3a*). A DNA ladder was run in lane L as a size marker. (B) The relative transcription levels of *csa3a*, *cas1*, *cas2* and *cas4* in *S. islandicus* *csa3a*-overexpression (OE*csa3a*) and *csa3a*-deletion (Δ *csa3a*) strains after normalization to the level in *S. islandicus* E233S (wild-type) carrying empty vector pSeSD. (C) Western blot analysis of Csa3a, Csa1 and Cas1 protein expression levels in *S. islandicus* E233S (wild-type), *csa3a* overexpression (OE*csa3a*) and *csa3a*-deletion (Δ *csa3a*) strains. Antiserum against PCNA3 was used as a loading control.

genes, by the CRISPR-Cas-specific regulator Csa3a both *in vivo* and *in vitro*. The *csa3* gene is also named as *csx1* and is divided into subtype III-U (2). The *csx1* genes are often found physically linked to archaeal *cmr* gene cassettes, however, it also appears in Type I modules, which suggests a broader role for this gene (37,48). Csa3a protein is predicted to have an N-terminal domain of unknown function fused to a C-terminal MarR-like winged helix-turn-helix domain which suggests that it functions as a transcriptional regulator. The crystal structure of Csa3 from *S. solfatarius* P2 (Mw = 27.8 kDa) which is closely related to Csa3a of *S. islandicus* REY15A (66% identity) has recently been solved (49). Our results demonstrated that overexpression of *csa3a* triggered *de novo* spacer acquisition in the Type I-A CRISPR-Cas system of *Sulfolobus*. Moreover, it was shown that Csa3a bound to an imperfect palindromic sequence (TTCNTAACTAAANANGGNNNGAAA, Supplementary Figure S2A) upstream of the *cas1* promoter or two similar elements covering the BRE and TATA box of the *cas1* promoter and the translational start codon. Overexpression of *csa3a* greatly increased the transcriptional level, and moderately increased translational level, of *acas* genes, demonstrating that Csa3a acted as a transcriptional activator. Csa3a may regulate transcription of other genes, and even CRISPR loci, however, this need to be further confirmed. The Csa3a-binding site on the *cas1* promoter represents a typical activation *cis*-element in archaea, located immediately upstream of core promoter elements (Supplementary Figure S2B) and the binding proteins recruit general transcription factors to activate transcription (50). The two Csa3a-binding sites on the *cas1* promoter covering the TATA box, BRE and translational start codon (Supplementary Figure S2B), presented an atypical mechanism of transcription activation. However, the *in vivo* data suggested that Csa3a activated the transcription of the *cas1* gene (Figure 5B and C). Our results are also consistent with the transcriptomic data of *S. islandicus* LAL 14/1, in which up-regulation of the *acas* operon was accompa-

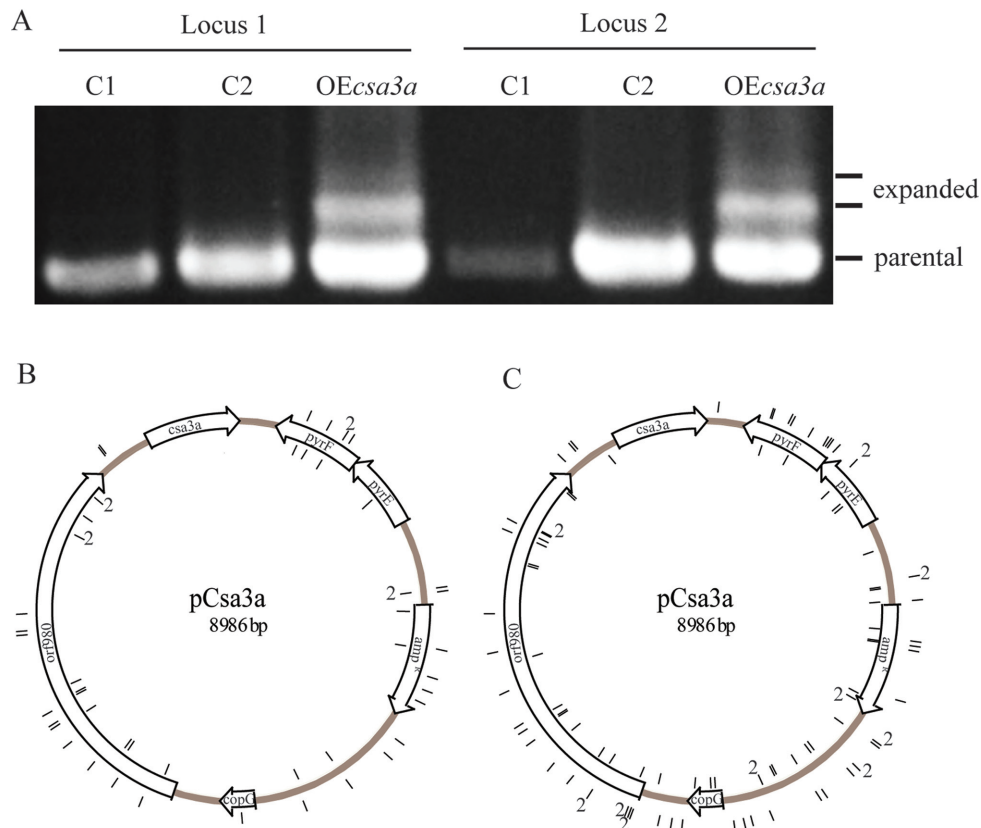


Figure 6. Csa3a triggers CRISPR spacer acquisition in *Sulfolobus*. (A) Spacer acquisition was detected by PCR for both loci in *csa3a* overexpression strain. C1: *S. islandicus* E233S strain, C2: E233S carrying empty vector pSeSD, OE*csa3a*: E233S carrying *csa3a* overexpression plasmid. The *csa3a* expression plasmid is shown with individual genes indicated (B and C). Locations of the protospacers corresponding to newly acquired plasmid-derived spacers on locus 1 (B) and locus 2 (C) are shown in the inner (forward strand) and outer (reverse strand) circles.

nied by increased *csa3a* transcript levels after rudivirus infection (48). These results suggested that *csa3a* was transcribed at a low level before virus infection, while after infection elevated Csa3a levels may activate transcription of *acas* genes. Indeed, Csa3a was not detected in the wild-type strain by western blotting and we also confirmed that the promoter strength of both *cas1* and *cas2* promoters was very low ($\sim 10^{-6}$ -fold) compared with that of the *araS* promoter (50) (Supplementary Figure S2C). Thus we infer that the low transcriptional level of *csa3a* gene caused a low expression activity of *acas* operon.

It was previously reported that virus infection can trigger spacer acquisition. For example the monocaudavirus SMV1 induced spacer acquisition from coinfecting pMGB1 and STSV2 genomes in *Sulfolobus* (20,21) and priming spacer-mediated spacer acquisition was observed in *H. hispanica* (22), *E. coli* (16,25,29) and *P. atrosepticum* (24). It has also been reported that overexpression of Cas1 and Cas2 can facilitate spacer acquisition in *E. coli* (17,25). However, the basic mechanism by which the aCas complex is activated for adaptation remains unknown. In this study, overexpression of the *csa3a* gene on a plasmid strongly activated transcription of the *acas* operon but the translational level of *acas* operon was increased to a lesser degree than the transcriptional level (Figure 5B). Most importantly, we observed that overexpression of *csa3a* gene triggered *de novo*

spacer acquisition efficiently. Thus, we demonstrated that the transcription regulator Csa3a played a crucial role in this pathway and a model is proposed for this process in Figure 7: step 1: invasion of virus or other mobile genetic elements induces expression of the transcriptional regulator gene; step 2: the regulator largely activates transcription of the *acas* operon, including *cas1* and *cas2*; step 3: Cas1, Cas2 and other Cas proteins forming the aCas complex take up new spacers from the invader DNA. The first step has been confirmed in *S. islandicus* LAL 14/1, that *csa3a* gene was induced in the late phase after virus infection. However, *acas* genes were induced slightly in the early phase, which may suggest other factors but not Csa3a were also involved in *acas* induction. Anyhow, no *de novo* spacer acquisition was observed for this strain after virus challenging (48). The increase in *csa3a* and *acas* transcription was much higher in our study. This suggests that an induction threshold value of *csa3a* is required for efficient activation of *acas* genes. Our model suggests further that the Cas1- and Cas2-dependent spacer acquisition process has an efficiency as high as the priming process, as has been reported previously in other archaea or bacterium (22,25). In our model, the *de novo* spacer acquisition process needs to be largely activated by a transcription regulator which may sense invading genetic elements.

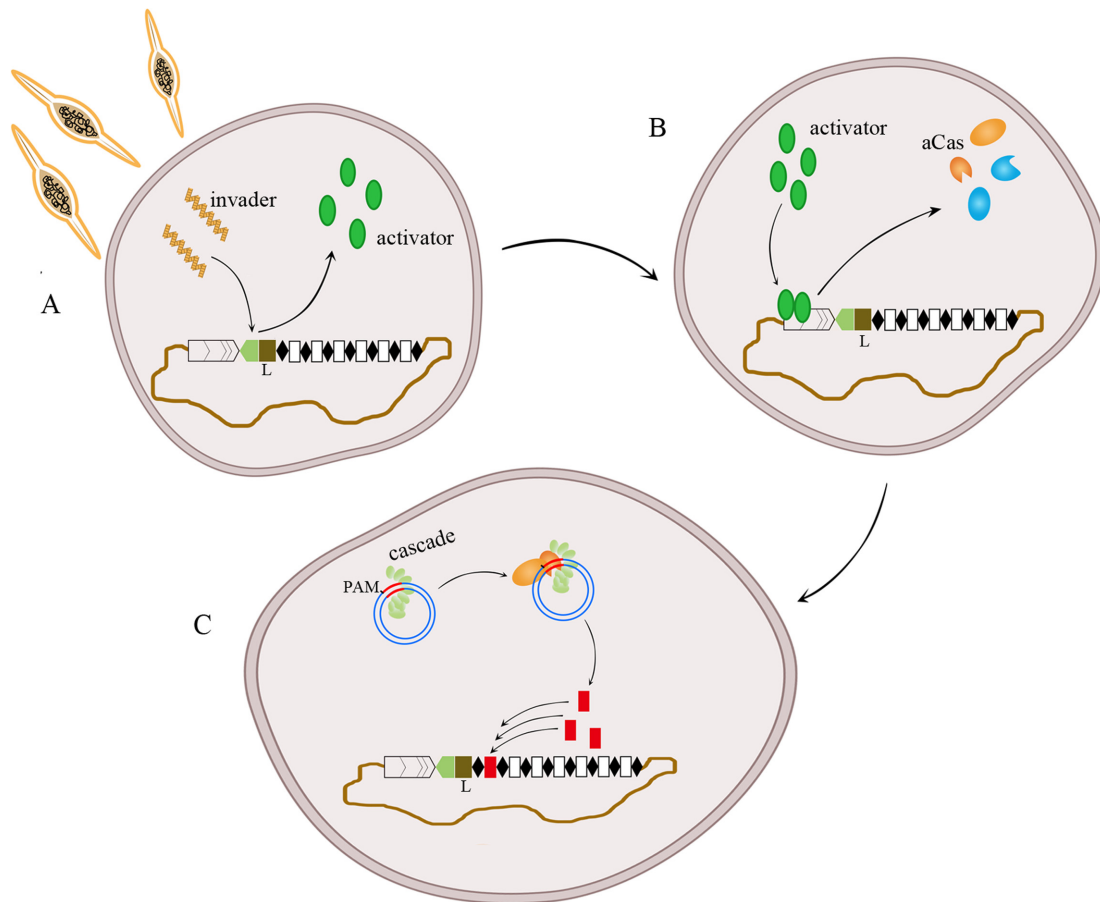


Figure 7. A CRISPR *de novo* spacer acquisition model that is transcriptional regulator-mediated and Cas1- + Cas2-dependent. **(A)** During infection with virus and other genetic mobile elements, a transcriptional regulator senses the invader and its expression is activated. **(B)** The regulator strongly activates the transcription of *acas* genes, including *cas1* and *cas2*. **(C)** Cas1, Cas2 and other Cas proteins (Cascade and Cas3 are possible) together recognize PAM sequences of the protospacer in the invader DNA and *de novo* spacers are inserted into the CRISPR locus adjacent to the leader sequence.

Our results also indicate the greatly induced Cas1- and Cas2-dependent spacer acquisition process is dependent on the PAM sequence of the protospacers (Table 1) but it is not as strict as was observed for the SMV1 virus-induced spacer acquisition (51). Moreover, for an *E. coli* CRISPR-Cas minus strain it was reported that overexpression of *cas1* and *cas2* genes led to *de novo* spacer acquisition from protospacer with less conserved PAM sequence (17). Therefore, we infer that a Cas protein other than Cas1 and Cas2 is responsible for PAM recognition, and that an enhanced level of Cas1 and Cas2 proteins results in spacer acquisition from protospacers with less conserved PAM sequence.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Prof. Roger A Garrett of Danish Archaea Centre in University of Copenhagen for critical reading of the manuscript.

FUNDING

National Natural Science Foundation of China [31100050 to N. P., 31128011 to Q. S. and Y. L.]. Funding for open access charge: National Natural Science Foundation of China [31100050 and 31128011].

Conflict of interest statement. None declared.

REFERENCES

- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **9**, 467–477.
- Sorek, R., Kunin, V. and Hugenoltz, P. (2008) CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.*, **6**, 181–186.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
- Kunin, V., Sorek, R. and Hugenoltz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.*, **8**, R61.

6. Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, **151**, 653–663.
7. Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S.D. (2005) Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**, 2551–2561.
8. Shah, S.A., Hansen, N.R. and Garrett, R.A. (2009) Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem. Soc. Trans.*, **37**, 23–28.
9. Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.*, **1**, 7.
10. Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S. and Shinkai, A. (2010) Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J. Mol. Biol.*, **395**, 270–281.
11. Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L. and Severinov, K. (2010) Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.*, **77**, 1367–1379.
12. Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comp. Biol.*, **1**, e60.
13. Koonin, E.V. and Makarova, K.S. (2013) CRISPR–Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol.*, **10**, 679–686.
14. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
15. Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J. and van der Oost, J. (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu. Rev. Genet.*, **46**, 311–339.
16. Swarts, D.C., Mosterd, C., van Passel, M.W. and Brouns, S.J. (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One*, **7**, e35888.
17. Yosef, I., Goren, M.G. and Qimron, U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.*, **40**, 5569–5576.
18. Yosef, I., Shitrit, D., Goren, M.G., Burstein, D., Pupko, T. and Qimron, U. (2013) DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 14396–14401.
19. Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R. and O'Toole, G.A. (2012) The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J. Bacteriol.*, **194**, 5728–5738.
20. Erdmann, S. and Garrett, R.A. (2012) Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol. Microbiol.*, **85**, 1044–1056.
21. Erdmann, S., Le Moine Bauer, S. and Garrett, R.A. (2014) Inter-viral conflicts that exploit host CRISPR immune systems of *Sulfolobus*. *Mol. Microbiol.*, **91**, 900–917.
22. Li, M., Wang, R., Zhao, D. and Xiang, H. (2014) Adaptation of the *Haloarcula hispanica* CRISPR–Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.*, **42**, 2483–2492.
23. Li, M., Wang, R. and Xiang, H. (2014) *Haloarcula hispanica* CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic Acids Res.*, **42**, 7226–7235.
24. Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H. and Fineran, P.C. (2014) Priming in the Type I-F CRISPR–Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.*, **42**, 8516–8526.
25. Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K. and Semenova, E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.*, **3**, 945.
26. Arslan, Z., Hermanns, V., Wurm, R., Wagner, R. and Pul, U. (2014) Detection and characterization of spacer integration intermediates in type I-E CRISPR–Cas system. *Nucleic Acids Res.*, **42**, 7884–7893.
27. Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H. and Fineran, P.C. (2014) Priming in the Type I-F CRISPR–Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.*, **42**, 8516–8526.
28. Diez-Villasenor, C., Guzman, N.M., Almendros, C., Garcia-Martinez, J. and Mojica, F.J. (2013) CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR–Cas I-E variants of *Escherichia coli*. *RNA Biol.*, **10**, 792–802.
29. Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H. and Brouns, S.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 1629–1638.
30. Hynes, A.P., Villion, M. and Moineau, S. (2014) Adaptation in bacterial CRISPR–Cas immunity can be driven by defective phages. *Nat. Commun.*, **5**, 4399.
31. Heler, R., Marraffini, L.A. and Bikard, D. (2014) Adapting to new threats: the generation of memory by CRISPR–Cas immune systems. *Mol. Microbiol.*, **93**, 1–9.
32. Deng, L., Kenchappa, C.S., Peng, X., She, Q. and Garrett, R.A. (2012) Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in *Sulfolobus*. *Nucleic Acids Res.*, **40**, 2470–2480.
33. Shinkai, A., Kira, S., Nakagawa, N., Kashiwara, A., Kuramitsu, S. and Yokoyama, S. (2007) Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. *J. Bacteriol.*, **189**, 3891–3901.
34. Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N. and Wagner, R. (2010) Identification and characterization of *E. coli* CRISPR–cas promoters and their silencing by H-NS. *Mol. Microbiol.*, **75**, 1495–1512.
35. Westra, E.R., Pul, U., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L. et al. (2010) H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.*, **77**, 1380–1393.
36. Medina-Aparicio, L., Rebollar-Flores, J.E., Gallego-Hernandez, A.L., Vazquez, A., Olvera, L., Gutierrez-Rios, R.M., Calva, E. and Hernandez-Lucas, I. (2011) The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in *Salmonella enterica* serovar Typhi. *J. Bacteriol.*, **193**, 2396–2407.
37. Deng, L., Garrett, R.A., Shah, S.A., Peng, X. and She, Q. (2013) A novel interference mechanism by a type IIIB CRISPR–Cmr module in *Sulfolobus*. *Mol. Microbiol.*, **87**, 1088–1099.
38. Peng, W., Li, H., Hallstrom, S., Peng, N., Liang, Y.X. and She, Q. (2013) Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. *RNA Biol.*, **10**, 738–748.
39. Gudbergdottir, S., Deng, L., Chen, Z., Jensen, J.V., Jensen, L.R., She, Q. and Garrett, R.A. (2011) Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.*, **79**, 35–49.
40. Deng, L., Zhu, H., Chen, Z., Liang, Y.X. and She, Q. (2009) Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. *Extremophiles*, **13**, 735–746.
41. Peng, N., Deng, L., Mei, Y., Jiang, D., Hu, Y., Awaye, M., Liang, Y. and She, Q. (2012) A synthetic arabinose-inducible promoter confers high levels of recombinant protein expression in hyperthermophilic archaeon *Sulfolobus islandicus*. *Appl. Environ. Microbiol.*, **78**, 5630–5637.
42. Abella, M., Rodriguez, S., Paytubi, S., Campoy, S., White, M.F. and Barbe, J. (2007) The *Sulfolobus solfataricus* *radA* paralogue sso0777 is DNA damage inducible and positively regulated by the Stal protein. *Nucleic Acids Res.*, **35**, 6788–6797.
43. Samson, R. Y., Xu, Y., Gadelha, C., Stone, T.A., Faqiri, J.N., Li, D., Qin, N., Pu, F., Liang, Y.X., She, Q. et al. (2013) Specificity and function of archaeal DNA replication initiator proteins. *Cell Rep.*, **3**, 485–496.

44. Zhang,C., Guo,L., Deng,L., Wu,Y., Liang,Y., Huang,L. and She,Q. (2010) Revealing the essentiality of multiple archaeal *pcna* genes using a mutant propagation assay based on an improved knockout method. *Microbiology*, **156**, 3386–3397.
45. Ao,X., Li,Y., Wang,F., Feng,M., Lin,Y., Zhao,S., Liang,Y. and Peng,N. (2013) The *Sulfolobus* initiator element is an important contributor to promoter strength. *J. Bacteriol.*, **195**, 5216–5222.
46. Garrett,R.A., Vestergaard,G. and Shah,S.A. (2011) Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol.*, **19**, 549–556.
47. Manica,A., Zebec,Z., Teichmann,D. and Schleper,C. (2011) *In vivo* activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol. Microbiol.*, **80**, 481–491.
48. Quax,T.E., Voet,M., Sismeiro,O., Dillies,M.A., Jagla,B., Coppee,J.Y., Sezonov,G., Forterre,P., van der Oost,J., Lavigne,R. *et al.* (2013) Massive activation of archaeal defense genes during viral infection. *J. Virol.*, **87**, 8419–8428.
49. Lintner,N.G., Frankel,K.A., Tsutakawa,S.E., Alsbury,D.L., Copie,V., Young,M.J., Tainer,J.A. and Lawrence,C.M. (2011) The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *J. Mol. Biol.*, **405**, 939–955.
50. Peng,N., Xia,Q., Chen,Z., Liang,Y.X. and She,Q. (2009) An upstream activation element exerting differential transcriptional activation on an archaeal promoter. *Mol. Microbiol.*, **74**, 928–939.
51. Shah,S.A., Erdmann,S., Mojica,F.J. and Garrett,R.A. (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol.*, **10**, 891–899.