7226–7235 *Nucleic Acids Research, 2014, Vol. 42, No. 11* *Published online 06 May 2014*
doi: 10.1093/nar/gku389

Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation

Ming Li[1,2], Rui Wang[1,2] and Hua Xiang[1,*]

[1]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China and [2]University of Chinese Academy of Sciences, Beijing 100049, China

Received March 15, 2014; Revised April 20, 2014; Accepted April 22, 2014

ABSTRACT

The prokaryotic immune system CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated genes) adapts to foreign invaders by acquiring their short deoxyribonucleic acid (DNA) fragments as spacers, which guide subsequent interference to foreign nucleic acids based on sequence matching. The adaptation mechanism avoiding acquiring 'self' DNA fragments is poorly understood. In *Haloarcula hispanica*, we previously showed that CRISPR adaptation requires being primed by a pre-existing spacer partially matching the invader DNA. Here, we further demonstrate that flanking a fully-matched target sequence, a functional PAM (protospacer adjacent motif) is still required to prime adaptation. Interestingly, interference utilizes only four PAM sequences, whereas adaptation-priming tolerates as many as 23 PAM sequences. This relaxed PAM selectivity explains how adaptation-priming maximizes its tolerance of PAM mutations (that escape interference) while avoiding mis-targeting the spacer DNA within CRISPR locus. We propose that the primed adaptation, which hitches and cooperates with the interference pathway, distinguishes target from non-target by CRISPR ribonucleic acid guidance and PAM recognition.

INTRODUCTION

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) arrays are present in ∼90% of archaeal and ∼40% of bacterial genomes (1,2). Each array consists of virtually identical repeats that are separated by variable virus- or plasmid-derived sequences, known as spacers (3–5). This special structure is frequently flanked by a gene operon encoding a diverse combination of CRISPR-associated (Cas) proteins (6). These two components together comprise the prokaryotic adaptive immune system against invading genetic elements (7,8). Based on a 'polythetic' criterion, this diverse system has been classified into three major types (9).

In Type I systems, the precursor CRISPR ribonucleic acid (pre-crRNA) is processed by a Cas endoribonuclease, Cas6 in most cases, into mature CRISPR RNA (crRNA) molecules (10–12). Each crRNA contains a spacer guide flanked by two repeat remnants known as 5′- and 3′-handles (10,12). Mature crRNAs are loaded into the Cascade (CRISPR-associated complex for antiviral defence) complex to perform invader deoxyribonucleic acid (DNA) surveillance (10,13,14). The multifunctional Cas3, which possesses ATPase, helicase and nuclease activities (15), is then recruited by the Cascade subunit(s), such as Cse1 from the *Escherichia coli* Type I-E system (16), to destruct the target DNA (10,14). In contrast, the adaptation (or spacer acquisition) pathway, which shapes and updates the CRISPR memory of invader information, has been less characterized since the first report in *Streptococcus thermophilus* (7). Recent studies on the *E. coli* Type I-E system revealed two different adaptation pathways, naïve adaptation and priming adaptation (17,18). Efficient naïve adaptation has only been observed in Cas1- and Cas2-overexpressing *E. coli* cells, in which new spacers were occasionally acquired from the chromosomal DNA (19). During priming adaptation, a pre-existing spacer directs efficient acquisition specifically from the invader DNA carrying a homologous sequence (17,20). The priming pathway allows interference to be restored to escape invaders (17).

Similar to other immune systems, CRISPR requires a discrimination mechanism to tell the 'self' DNA, such as the spacer DNA in the CRISPR cassette, from the 'non-self', such as the protospacer DNA from the invader. Such discrimination should take place during both interference and adaptation stages, otherwise autoimmunity may occur either directly or indirectly. It was recently reported that in Type I-E system, a fully matched target is interfered only when combined with one of four unchangeable PAM (protospacer adjacent motif) sequences (21). Lacking PAM sequences, spacers in the CRISPR locus are automatically defined as a 'non-target' for interference. Therefore, this is termed a 'target versus non-target' discrimination mechanism (21), in contrast to the 'self versus non-self' mecha-

*To whom correspondence should be addressed. Tel: +86 10 6480 7472; Fax: +86 10 6480 7472; Email: xiangh@im.ac.cn

nism described for the *Staphylococcus epidermidis* Type III-A system (22). In the Type III-A system, the spacer DNA is protected by sensing the base pairing between the 5′-handle of the crRNA and the corresponding portion of its preceding repeat, from which the 5′-handle derives. In contrast, the mechanism by which the CRISPR adaptation machinery discriminates the 'self' and 'non-self' sequences is poorly understood. Our recent study of the *Haloarcula hispanica* Type I-B CRISPR provides clues by demonstrating the absence or inactivation of the naïve adaptation pathway in this system, in which a priming process is essentially required (20). This adaptation is strictly restricted to the invader DNA carrying a 'familiar' sequence that could be recognized by the crRNA of a pre-existing spacer, thereby resulting in discriminative adaptation. However, it should be noted that, without an additional self-avoidance mechanism to distinguish the spacer DNA during adaptation-priming, the chromosomal sequences within or around the CRISPR cassette could still be acquired as self-targeting spacers. Previous studies of priming adaptation revealed its insensitivity to PAM mutations flanking a target (17,20), which compromises the possibility of PAM authentication during priming.

Type I-B systems in haloarchaea have been recently investigated so far involving the priming adaptation (20), crRNA maturation (12,23) and interference (24,25) pathways. A plasmid-based invader assay has revealed the important role of PAM during target interference (25). In this study of the *H. hispanica* Type I-B system, we systemically mutated the tri-nucleotide PAM sequence of a fully matched target to determine its role during interference and especially adaptation-priming. Our results revealed that *H. hispanica* Type I-B interference recognizes four specific PAM sequences, and surprisingly, in addition to these four sequences, another 19 PAM variants are differently tolerated to elicit priming adaptation. It was demonstrated that PAM authentication, which strictly recognizes the −1, −2 and −3 nucleotides of a target (spacer-matching) sequence, is common to interference and adaptation-priming processes. Therefore, we propose that both adaptation and interference require the base pairing-independent PAM recognition and the base pairing-dependent crRNA guidance to exclude the spacer DNA and other 'self' sequences.

MATERIALS AND METHODS

Strains and culturing conditions

The *H. hispanica* strains used in this study are listed in Supplementary Table S1. The uracil auxotrophic (*pyrF*-deleted) strain DF60 (26) and its derivatives were cultured at 37°C in AS-168 medium (per litre, 200 g NaCl, 20 g $MgSO_4 \cdot 7H_2O$, 2 g KCl, 3 g trisodium citrate, 1 g sodium glutamate, 50 mg $FeSO_4 \cdot 7H_2O$, 0.36 mg $MnCl_2 \cdot 4H_2O$, 5 g Bacto Casamino Acids, 5 g yeast extract, pH 7.2) with uracil added at a concentration of 50 mg/l. The strains transformed by pWL502 or its derivatives were cultured in yeast extract-subtracted AS-168.

The *E. coli* JM109 used for cloning was cultured in Luria–Bertani medium. When needed, ampicillin was added to a final concentration of 100 mg/l.

Plasmid challenge assay

The target plasmids (listed in Supplementary Table S1) were constructed by cloning a sticky fragment into pWL502 (27) predigested with BamHI and KpnI. The fragment contains a spacer-matching sequence preceded by a designed PAM sequence. In most cases, two different-sized oligonucleotides were annealed to generate this sticky fragment. The DNA fragment of the repeat-flanked target sequence was amplified from the genomic CRISPR DNA with corresponding primers, and digested with BamHI and KpnI before cloning. To construct pR-TCT1 and pR-TTC1, nucleotide substitutions were performed by polymerase chain reaction (PCR) mutagenesis using a pGEM-T vector (pGEM-T Easy, Promega) carrying the wild-type repeat sequence as the template. The corresponding oligonucleotides are listed in Supplementary Table S2.

The plasmid challenge assay was performed by transforming these target plasmids into uracil auxotrophic DF60 cells according to the Halohandbook online protocol (http://www.haloarchaea.com/resources/halohandbook/Halohandbook_2009_v7.2mds.pdf). Individual colonies were screened on yeast extract-subtracted AS-168 agar plates. For each CRISPR-interfered plasmid, three replicates were performed to evaluate the interference effect.

Spacer acquisition assay

Spacer acquisition assay against the target plasmids was performed as previously described (20) with a few modifications. Briefly, for each target plasmid, at least three transformant colonies were separately inoculated into yeast extract-subtracted AS-168 medium and cultured for at least 5 days to allow sufficient interaction between the CRISPR system and the target plasmid. The liquid cultures were centrifuged at 10 000 rpm for 1 min to collect the cells, which were then lysed in distilled water. For these samples, CRISPR expansion was monitored by PCR using primer pairs amplifying the leader-proximal end (ExTest-CAS2 and ExTest-SP1, which locate within *cas2* and spacer1, respectively) (Supplementary Table S2).

CRISPR mutant construction

To construct the CRISPR mutants $S1^{C\text{-}1A}$ and $S1^{C\text{-}1A}$, a 1294-bp CRISPR structure containing only one spacer (spacer1) was first generated by bridge PCR. This structure was then cloned into the pGEM-T vector and subjected to PCR mutagenesis. The mutated CRISPRs were subsequently cloned into the suicide plasmid pHAR and used to replace the wild-type CRISPR through the pop-in-pop-out gene knockout strategy (26).

RESULTS

H. hispanica CRISPR recognizes four of 64 PAM variants for interference

Recently, we reported the adaptation of *H. hispanica* Type I-B CRISPR to an invading virus or plasmid, in which the priming match between a pre-existing spacer and the invader DNA was strictly required (20). CRISPR interference

was not observed due to escape mutations within the spacer-matched sequence, i.e. protospacer. To exclusively investigate the role of PAM, we constructed a series of plasmids using a target sequence (protospacer1) that is fully matched by spacer1 (Figure 1A). Two oligonucleotides were annealed to form a dsDNA fragment containing protospacer1 and a preceding tri-nucleotide (at positions −1, −2 and −3) as the PAM sequence. The fragment with two sticky ends was cloned into restricted pWL502 (27) (carrying the selection marker gene *pyrF*) to generate the target plasmid. The possible base composition of the PAM sequence was sampled exhaustively, yielding a total of 64 (4 × 4 × 4) different plasmids, each named in the format pNNN1, where 'NNN' represents the distinct PAM sequence and '1' represents the common protospacer1. These plasmids were transformed into uracil auxotrophic *H. hispanica* DF60 (Δ*pyrF*) cells (26) under selection pressure. Given a *bona fide* PAM, the Cascade complex loaded with the crRNA of spacer1 (s1 crRNA) should recognize the target DNA, form an R-loop structure (28) (Figure 1A), and recruit Cas3 for target interference (10), which will cause reduced plasmid transformation efficiency.

Although each plasmid carries the fully-targeted protospacer1, varying interference effects were observed (Figure 1B), suggesting that interference activity is tightly regulated by a PAM sensing event. Interestingly, a TT or CC dinucleotide at the −3 and −2 positions appeared necessary for interference, and evidently reduced transformation efficiency was observed only for target plasmids with TTC, TTT, TTG or CCC as the PAM sequence. This suggests that *H. hispanica* interference only recognizes these four PAM sequences, which we described as TIP (target interference permissive). Actually, the four TIP sequences are not equally favoured by *H. hispanica* CRISPR, because the cells showed almost absolute resistance to pTTC1, slightly compromised resistance to pTTG1 and more compromised to pTTT1 and pCCC1 (Figure 1C). Correspondingly, in our previous study, we observed numerous new spacers acquired from foreign sequences conservatively preceded by TTC (20). Notably, these TIP sequences are not only different from the four PAMs recognized by the Type I-E interference machinery (21), but also different from those adopted by the *Haloferax volcanii* Type I-B system (25), suggesting the divergently-evolved PAM selectivity between subtypes and/or organisms.

Adaptation-priming tolerates as many as 23 PAM sequences

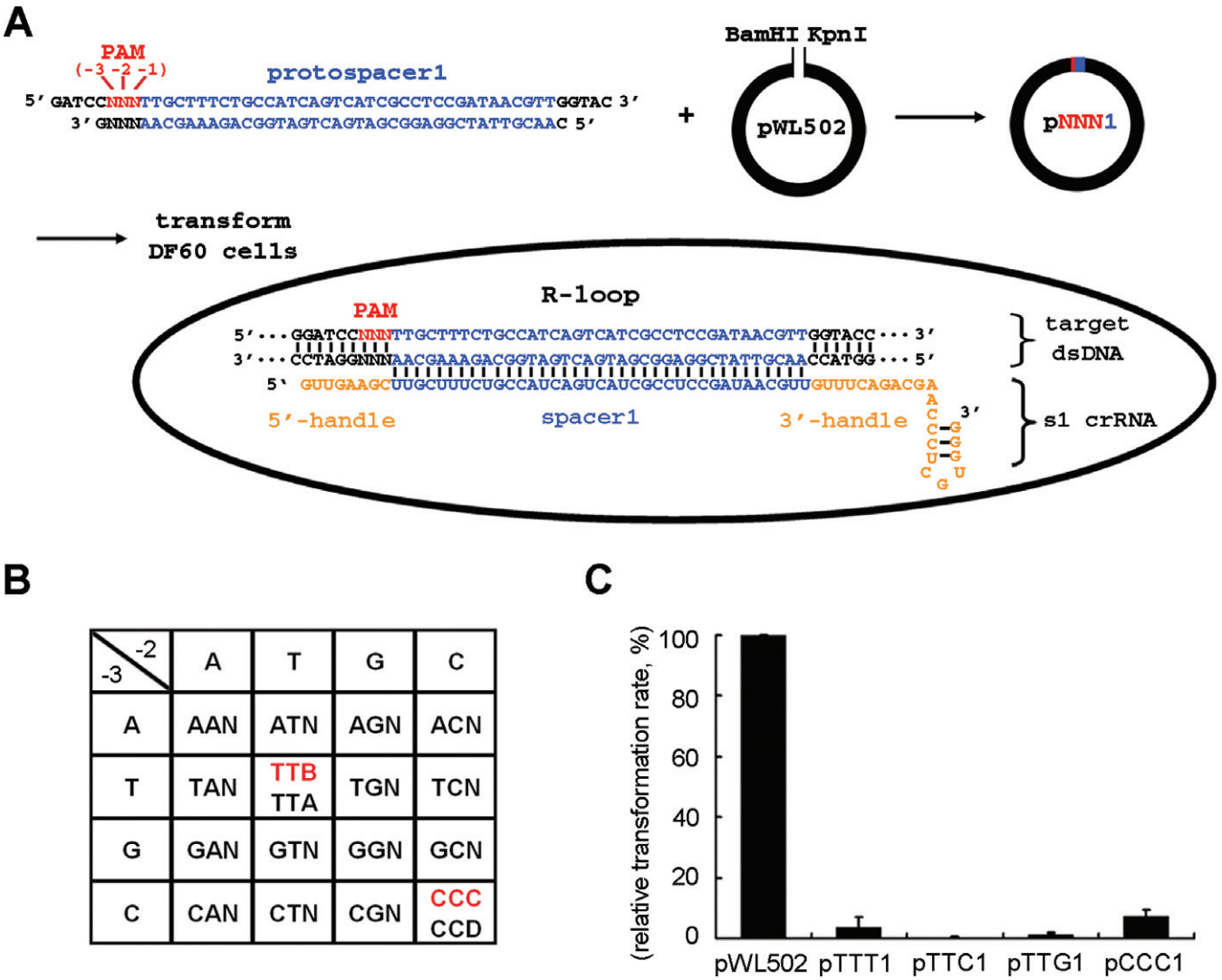
The majority of PAM mutations should block target interference because of the strict PAM selectivity. However, priming adaptation can counter these escape mutations by acquiring new spacers from the target-bearing DNA (17). Whether all types of PAM mutations can be tolerated to prime adaptation remains unknown. Therefore, for the 60 plasmids that escaped CRISPR interference, a spacer acquisition assay was subsequently performed against their transformants after a 5-day cultivation. Specific primers were used to amplify the CRISPR leader end, and arrays with new spacers incorporated were expected to produce larger-sized PCR products (20). Strikingly, CRISPR adaptation was observed for nearly one third of these escape plasmids (Figure 2A), revealing 19 different PAM sequences that are recognized to prime adaptation, which we described as PAP (priming adaptation permissive). Correspondingly, the remaining 41 tri-nucleotides were referred as PAIN (priming adaptation and interference non-permissive) sequences. Notably, PAP sequences are not all equally favoured, because faint expanded bands were observed for pTAA1, pCAG1, pCCG1 and pCGC1, whereas for most of the other plasmids, evident CRISPR expansion was observed (Figure 2A).

The TIP sequences TTC, TTT, TTG and CCC are also PAP. For example, CRISPR expansion was similarly detected for pCCC1 and pTTT1 transformants (Figure 2A). However, for pTTC1 and pTTG1, transformed colonies were rarely observed due to the extreme interference effect, and even when some colonies were observed, the CRISPR activity may have been inactivated to survive the selection pressure. For example, spacer1 deletion was observed for some pTTC1 and pTTG1 transformants (Supplementary Figure S1). To circumvent this barrier, we replaced protospacer1 with a sequence that is partially matched by spacer13 (Figure 2B), which we designated 'protospacer13v' for its derivation from the halovirus HHPV-2 (20). As expected, interference was not observed to the modified plasmids pTTC13v and pTTG13v (data not shown), whereas adaptation was readily detected for their transformants (Figure 2C), indicating that TTC and TTG are also PAP sequences. This result also suggests that compared to interference, priming adaptation tolerates more crRNA-protospacer mismatches. Adaptation to the protospacer13v target combined with another TIP sequence (TTT), two PAP sequences (TCC and CTC) (20) and two PAIN sequences (AGC and ACC), was also respectively tested, which showed similar results to the protospacer1-based assay (Figure 2C).

It should be noted that neither interference nor adaptation was observed to the empty pWL502 (data not shown), and the engineered pNNN1 target plasmids are completely the same except the designed PAM preceding protospacer1, hence their different performance in interference and adaptation assays clearly demonstrates the PAM selectivity of the interference and priming processes (summarized in Figure 3). Interestingly, most (15 of 19) PAP sequences could result from one of the four TIP sequences through a single point mutation, suggesting that adaptation-priming has shaped its PAM selectivity to tolerate these point mutations that escape interference. However, sequences with a purine (A or G) at the −3 position are consistently PAIN, therefore a purine (A or G) mutation at this position can cause escape from both interference and priming adaptation.

PAM authentication prevents interference and priming adaptation occurring to the CRISPR locus

Within the *H. hispanica* CRISPR cassette, the three repeat nucleotides immediately preceding each spacer are conserved AGC, which was identified as a PAIN sequence (Figure 3). Hence both interference and priming adaptation are prevented from occurring to the spacer DNA. However, it could also be attributed to additional base pairing formed at the −1, −2 and −3 positions according to the 'self versus



-3 \ -2	A	T	G	C
A	AAN	ATN	AGN	ACN
T	TAN	TTB TTA	TGN	TCN
G	GAN	GTN	GGN	GCN
C	CAN	CTN	CGN	CCC CCD

Figure 1. Identification of functional PAM sequences for CRISPR-interference. (**A**) Diagram depicting the plasmid challenge assay to determine the effects of various PAM sequences on CRISPR-interference. Each target plasmid carries a sequence that is fully matched by spacer1 (protospacer1). The tri-nucleotide PAM sequence located at the −1 to −3 positions of protospacer1 was exhaustively varied to generate 64 different sequences. Two oligonucleotides were annealed to form a sticky fragment containing protospacer1 and the PAM sequence, which was inserted into BamHI- and KpnI-digested pWL502. The pWL502 plasmid carries a *pyrF* gene that is required for *Haloarcula hispanica* DF60 cells to grow under uracil-free selection pressure. Cascade loaded with the crRNA of spacer1 (s1 crRNA) is expected to recognize protospacer1 with a *bona fide* PAM and form an R-loop structure to initiate target destruction. The 5′- and 3′-handles of the s1 crRNA are derived from the spacer1-flanking repeats. (**B**) Interference was observed to targets flanked by four PAM sequences (TTT, TTC, TTG and CCC) but not to those by the other 60 tri-nucleotides. The 64 tri-nucleotide sequences are arranged according the −2 and −3 nucleotides. At the −1 position, N stands for A, T, G or C; B stands for nucleotide T, G or C; whereas D means nucleotide A, T or G. (**C**) Interference effects to target plasmids carrying TTT, TTC, TTG or CCC as the PAM sequence. Three replicates were performed for each plasmid, and the relative transformation rate was calculated against the control pWL502.

non-self' theory of the Type III-A system (22). By analysing the crRNA-PAM base pairing pattern for each possible PAM sequence (Figure 3), interference and adaptation-priming showed no dependence on this extended base pairing. Therefore, *H. hispanica* CRISPR discriminates target from non-target by authenticating the PAM sequence instead of by sensing the crRNA-PAM base pairing.

A recent study reported a repeat binding protein that specifically binds to the CRISPR direct repeats (29), which may impede priming adaptation on the CRISPR DNA. Therefore, the spacer-flanking repeat sequences potentially provide additional self-protective mechanisms during priming adaptation. To test this possibility, we constructed plasmids containing a repeat sequence immediately preceding or following the protospacer1 target (Figure 4). As expected, neither interference nor adaptation was observed to pR1, because the preceding repeat provides AGC as a PAIN 'PAM'. When AGC were mutated to a TIP sequence (TTC) in pR-TTC1 and mutated to a PAP sequence (TCT) in pR-TCT1, interference and priming adaptation were observed, respectively (Figure 4). On the other hand, compared to pAGC1 and pTCT1, addition of a downstream repeat in pAGC1-R and pTCT1-R did not affect their performance in spacer acquisition assay (Figure 4). We previ-

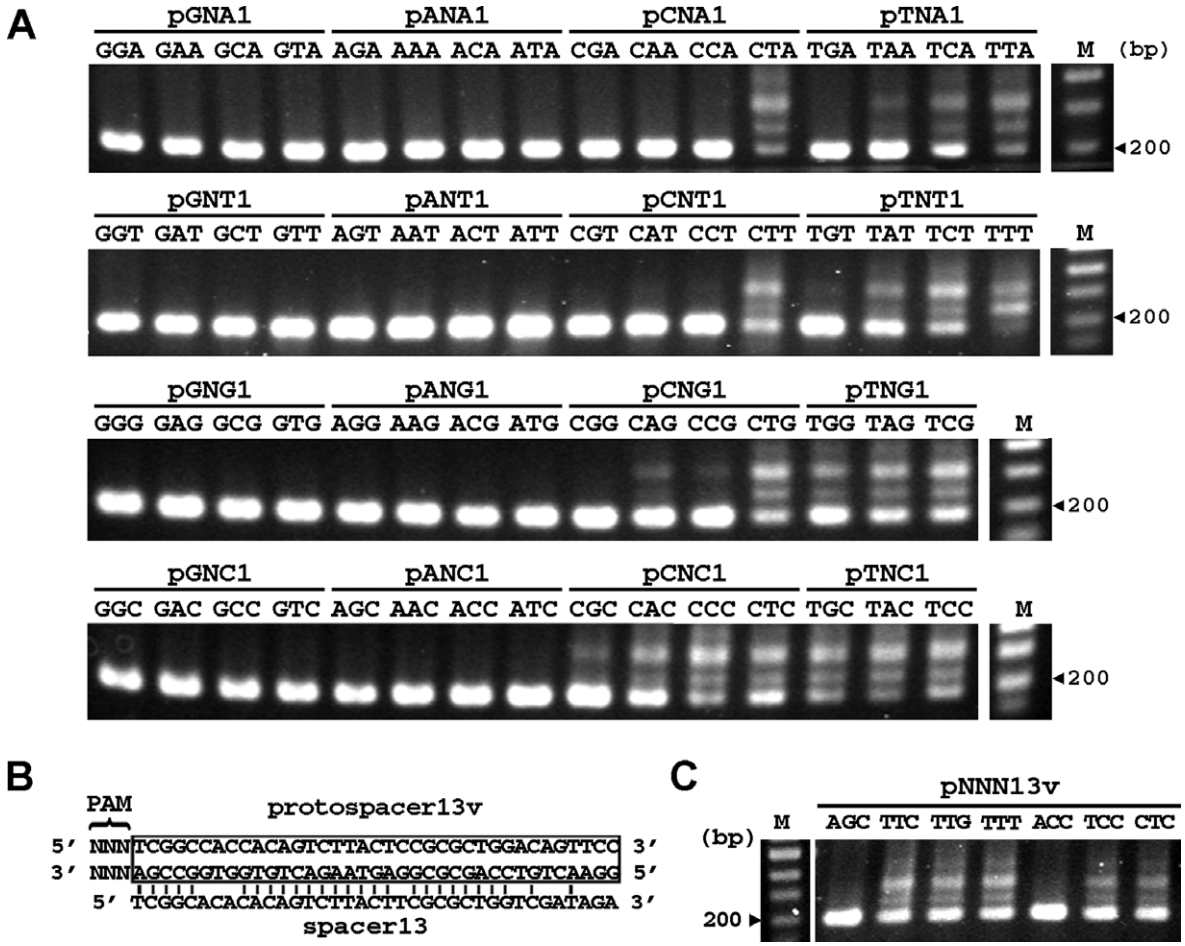


Figure 2. Adaptation priming tolerates 23 PAM sequences. (**A**) Spacer acquisition assay performed to target plasmids with protospacer1 preceded by 62 different PAM sequences (TTC and TTG not included). For each plasmid, three independent transformant colonies were tested, and a representative result is shown. The wild-type CRISPR generates a ∼200-bp PCR product, and larger-sized PCR products indicate that new spacers have been acquired causing expanded CRISPRs. (**B**) Scheme showing the provirus-derived sequence (protospacer13v, framed) that is partially matched by spacer13. (**C**) Spacer acquisition assay performed to target plasmids (pNNN13v) containing protospacer13v that is preceded by seven different PAM sequences, including TTC and TTG.

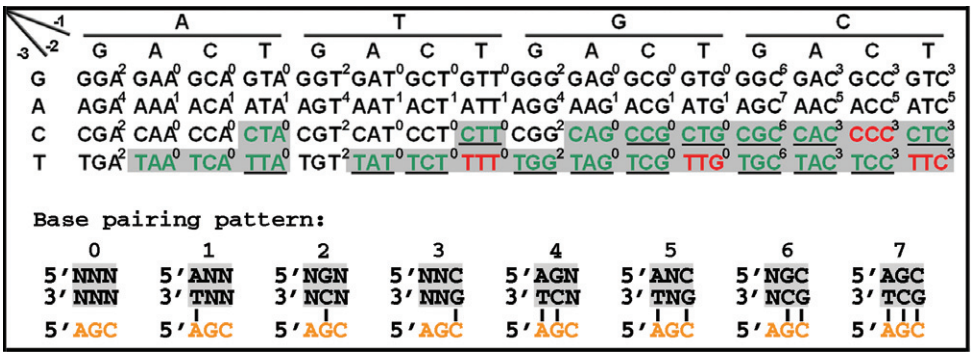


Figure 3. Summary of the potentials of the 64 tri-nucleotides to serve as a functional PAM during interference and/or adaptation-priming. The 23 tri-nucleotides shown against grey are permissive for priming adaptation (PAP), of which TTC, TTG, TTT and CCC are also permissive for interference (TIP). The other 41 tri-nucleotides are permissive for neither interference nor priming adaptation (PAIN). The underlined PAP variants could result from one of the TIP sequences through a single point mutation. Each tri-nucleotide is labelled with a number varying from 0 to 7, and these numbers indicate the different base pairing patterns (shown at the bottom) which potentially occur between each PAM sequence and the crRNA 5′-handle nucleotides AGC. N in the crRNA-sense strand or the crRNA-complementary strand signifies a nucleotide that is not same to or not complementary to the corresponding 5′-handle nucleotide.

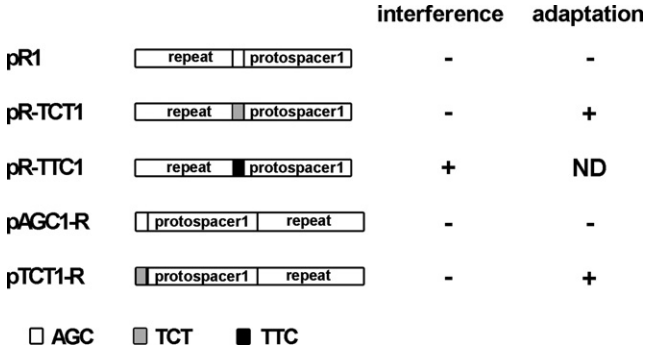


Figure 4. Upstream repeat protects the spacer DNA by providing a PAIN 'PAM' sequence (AGC). The plasmid pR1 has an intact CRISPR repeat sequence preceding protosapcer1 and the last three repeat nucleotides AGC were mutated to TCT or TTC to generate pR-TCT1 or pR-TTC1. For pAGC1-R and pTCT1-R, the protospacer1 target is preceded by AGC and TCT, respectively, and followed by an intact repeat. According to Figure 3, AGC, TCT and TTC are PAIN (priming adaptation and interference non-permissive), PAP (priming adaptation permissive) and TIP (target interference permissive) PAM sequences, respectively. Interference/adaptation to the target plasmid was observed (+) or not observed (−). ND, not determined.

ously revealed that DNA sequences upstream and downstream of the priming protospacer can both be acquired as new spacers, albeit with different strand bias and efficiency (20). By randomly selecting individual colonies showing an expanded CRISPR, we collected 32 new spacers from 22 pTCT1-R colonies and 39 new spacers from 28 pR-TCT1 colonies (Supplementary Table S3). Protospacers, from which these new spacers derived, could locate both upstream and downstream of the priming protospacer1 (Supplementary Figure S2), and showed a preference pattern similar to that observed for HHPV-2 and pVS (20). It was suggested that given a PAP PAM sequence, spacer acquisition from either side of the priming protospacer could not be impeded by a flanking repeat. It appears that for the spacer DNA, PAM authentication that recognizes its upstream repeat nucleotides as a PAIN signal serves as the only self-protective mechanism.

PAM authentication strictly recognizes nucleotides −1, −2 and −3 of the target sequence

The striking finding that adaptation-priming tolerates more than 20 PAM variants made us doubt whether this tolerance actually derives from the relaxed PAM selectivity at the theoretical −1, −2 and −3 positions, or nucleotides next to these positions have been misrecognized as a portion of PAM. Given the latter possibility, some results in Figure 2A may be false positives. Because the canonical PAM of this system proves to be TTC (20), we designed target plasmids pGTT5, pATT5, pCTT5 and pTTT5, with protospacer5 that is fully matched by spacer5. If misrecognition could occur to the PAM-3′-side nucleotide(s), the TTC nucleotides at the −2, −1 and +1 positions may be misrecognized as a permissive signal for both interference and priming adaptation (Figure 5A). However, consistent with the prostospacer1-based assay, CRISPR interference was only observed to pTTT5 (Figure 5C), and adaptation observed for pCTT5 but not for pATT5 or pGTT5 (Figure 5A), indicating the +1 nucleotide could not be misrecognized as a portion of PAM. We hypothesized that crRNA base pairing at the +1 position may have prevented this misrecognition. Therefore, we further designed plasmids pGTT4ms, pATT4ms, pCTT4ms and pTTT4ms ('ms' represents mismatch) with a modified protospacer4, whose first nucleotide A was substituted by C to introduce a mismatch (to the s4 crRNA) at the +1 position (Figure 5B). The crRNA matching within the seed region (positions +1 to +10) has been reported essentially required for Type I-B interference (24), and consistently, pTTT4ms escaped CRISPR interference (Figure 5C). Adaptation to pATT4ms and pGTT4ms was consistently not observed (Figure 5B), indicating even without crRNA matching at +1 position, the +1 nucleotide can still not be misrecognized for PAM authentication. Curiously, priming adaptation to pCTT4ms was blocked, hence the crRNA matching at position +1 appears to be important; however, adaptation to pTTT4ms was not affected (Figure 5B). From Figure 2A, we can see that TTT seemed a more favoured PAM for adaptation-priming, which may have compensated the +1 mismatch.

By ruling out misrecognition at position +1, CCC should be a reliable PAP (and TIP) PAM during the pCCC1 challenge assay, because the immediate 5′ upstream sequence of this PAM is a designed BamHI restriction site (5′-GGATCC-3′), and no matter the −4 or even −5 nucleotide could be misrecognized as a portion of PAM, authentication consistently occurred to a CCC tri-nucleotide (Figure 6A). Accordingly, the failure to recognize CCA and CCT as PAP (Figure 2A) suggests that misrecognition of the −4 nucleotide diagrammed in Figure 6A could not happen. We noted that crRNA-PAM matching at the −1 position occurred for pCCC1 but not for pCCA1 or pCCT1, so we constructed two mutant CRISPRs, $S1^{C\text{-}1A}$ and $S1^{C\text{-}1T}$, respectively carrying a C-to-A and a C-to-T repeat mutation at the −1 position of the spacer1 DNA (Supplementary Figure S3). These CRISPR mutants retained the adaptation phenotype to pCCC1 (Figure 6C), indicating that crRNA biogenesis was not affected. Then we challenged the $S1^{C\text{-}1A}$ and $S1^{C\text{-}1T}$ cells with pCCA1 and pCCT1, respectively, in which the additional −1 base pairing was introduced by these repeat mutations (Figure 6B). However, adaptation was still not observed (Figure 6C), indicating that the PAM-5′-side nucleotide(s) can not be misrecognized for PAM authentication, with or without a −1 base pairing.

From above, we conclude that for a priming protospacer, PAM authentication strictly recognizes its −1, −2 and −3 nucleotides, and the nearby crRNA-target base pairing does not affect this recognition. This is consistent with the very recent finding that the *Streptococcus pyogenes* Cas9 protein recognizes PAM prior to R-loop formation (30). After all, identification of these 23 PAP sequences proved to be convincing.

DISCUSSION

The ability to discriminate 'self' from 'non-self' is essential for every immune system. CRISPR-Cas serves as the only adaptive defence line in prokaryotes. Guided by crRNA molecules, CRISPR interference is directed to homologous foreign DNA. However, the crRNA-encoding DNA

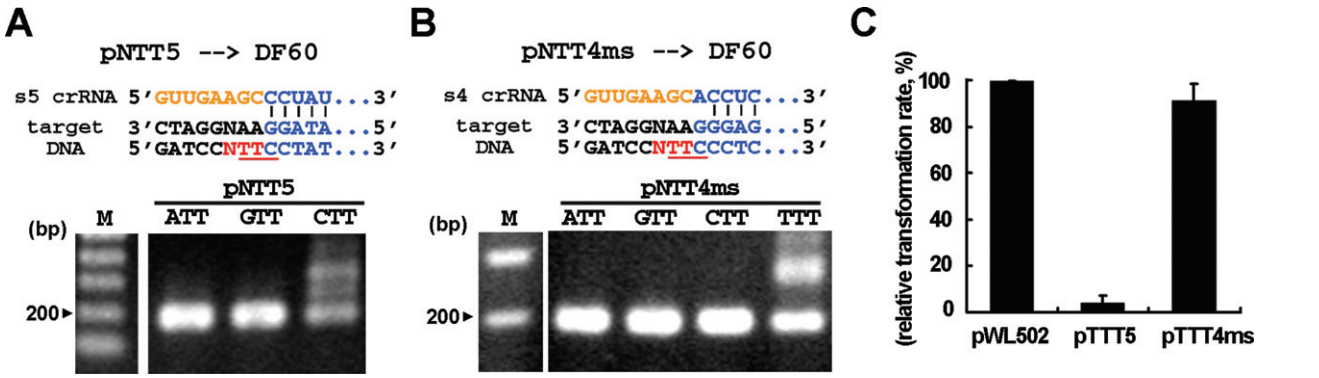


Figure 5. The +1 nucleotide was not recognized as a portion of PAM with (**A**) or without (**B**) a crRNA-target base pairing at this position. The pNTT5 and pNTT4ms ('ms' stands for mismatch) plasmids contain the protospacer sequence of spacer5 and spacer4, respectively, and their protospacer is preceded by a PAM sequence of ATT, GTT, CTT or TTT. The A-to-C substitution at the +1 position of protospacer4 introduces a mismatch here while generating a TTC sequence at positions −2, −1 and +1. The crRNA of spacer4 or spacer5 is denoted as s4 or s5 crRNA. The 5′-handle nucleotides are shown in orange, the spacer and protospacer sequences are in blue and the designed PAM sequence is in red. If the +1 nucleotide could be recognized as a portion of PAM, the underlined sequences may be misrecognized for PAM authentication. Lane Ms, dsDNA size markers. In panel (**C**), different interference effects to pTTT5 and pTTT4ms in DF60 cells are shown. Three replicates were performed for each plasmid, and the relative transformation rate was calculated against the control pWL502.

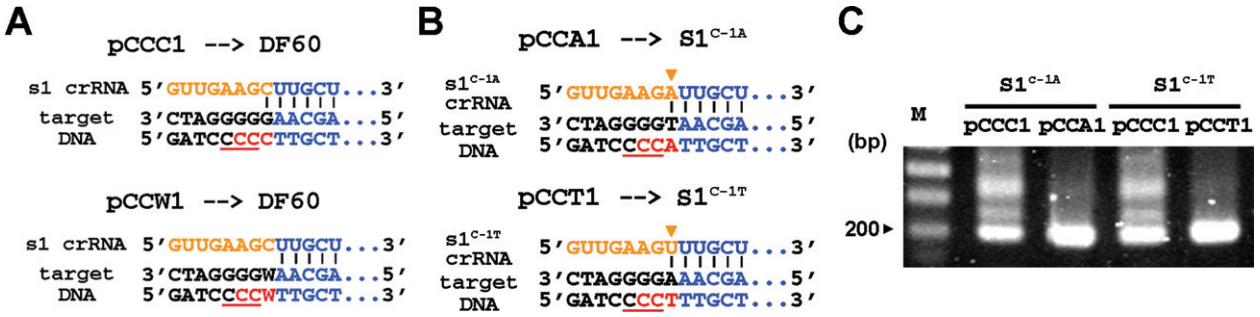


Figure 6. The −4 nucleotide was not recognized as a portion of PAM without (**A**) or with (**B**) an additional −1 base pairing. Priming adaptation was observed to pCCC1, but not to pCCW1 (see Figure 2A) (W is an A or T). DF60 cells encode a wild-type crRNA of spacer1 (s1 crRNA), whereas the $S1^{C-1A}$ and $S1^{C-1T}$ mutant cells express variant s1 crRNA molecules ($s1^{C-1A}$ and $s1^{C-1T}$ crRNA), in which the 5′-handle carries a C-to-A or C-to-U mutation at the −1 position (indicated by an orange arrow). The 5′-handle nucleotides are shown in orange, the spacer and protospacer sequences are in blue and the designed PAM sequence is in red. If the −4 nucleotide could be recognized as a portion of PAM, the underlined sequences may be misrecognized for PAM authentication. Panel (**C**) shows $S1^{C-1A}$ and $S1^{C-1T}$ CRISPRs could not adapt to pCCA1 and pCCT1, respectively. Lane M, dsDNA size marker.

(i.e. the spacer DNA) in the chromosome must be discriminatively protected. For this discrimination, two different mechanisms have been proposed, the 'self versus non-self' mechanism for the Type III-A system (22), and the 'target versus non-target' mechanism for the Type I-E system (21). Our data demonstrate that similar to Type I-E, the Type I-B interference machinery in *H. hispanica* cells also adopts a 'target versus non-target' mechanism based on PAM authentication. The tri-nucleotide sequences TTC, TTG, TTT and CCC can separately serve as a functional PAM during interference, which we termed 'target interference permissive' or TIP sequences. In *E. coli*, PAM recognition occurs to the nucleotides on the crRNA-complementary strand (termed target interference motif or TIM) (16,31), whereas the corresponding mechanism for the Type I-B system remains to be further investigated, given their different Cas components. The *E. coli* Cascade consists of five different Cas proteins (Cse1, Cse2, Cas7, Cas5 and Cas6e) in an uneven stoichiometry (1:2:6:1:1) (32). The Cse1 subunit is believed essential for PAM sensing, because direct interaction was observed between its conserved L1 loop and the PAM sequence (16). However, Cse1 is not common to other Type I subtypes. For example, the Type I-B Cascade comprises Cas5, Cas6, Cas7 and probably the specialized Cas8b (12,20,23). Interestingly, the Type I-E Cse1 and Type I-B Cas8b are both large proteins and predicted to share similar domain organization (33). Moreover, previous studies on the *Haloferax* Type I-B systems revealed that Cas8b is not required for the *in vivo* stability of crRNA (12,23), suggesting that Cascade lacking the Cas8b subunit exists stably, consistent with the observed Cse1 disassociation from *E. coli* Cascade at low concentrations (14). Thus, we infer that Cas8b probably monitors the PAM sequence in Type I-B systems. Interestingly, a previous study of the *H. volcanii* Type I-B system reported that ACT, TAA, TAT, TAG and CAC could serve as a functional PAM for interference (25). However, in our assay, the *H. hispanica* CRISPR does not interfere plasmids carrying these PAM sequences. These two haloarchaeal Type I-B systems carry nearly identical repeat sequences, whereas their Cas proteins are less conserved (Supplementary Figure S4), particularly Cas8b

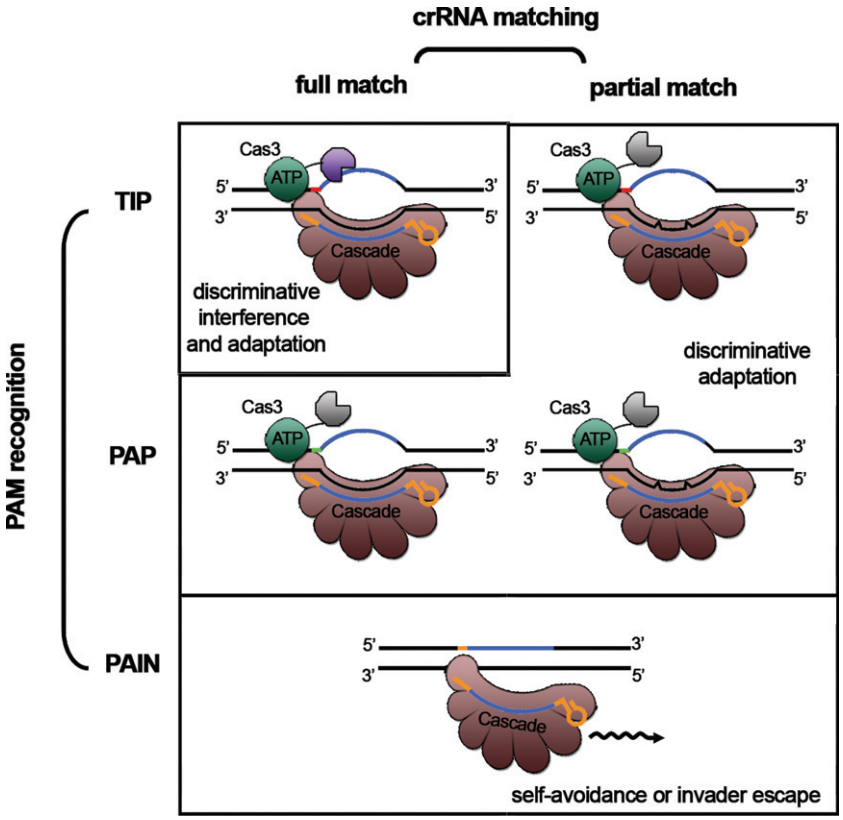


Figure 7. An integrated target discrimination model for CRISPR interference and adaptation-priming. The Cascade complex may quickly scan large DNA molecules for a TIP (target interference permissive) or PAP (priming adaptation permissive) PAM sequence. This PAM recognition process would ignore sequences preceded by a PAIN (priming adaptation and interference non-permissive) tri-nucleotide, including the spacer DNA and some PAM-mutated invader targets. When a TIP or PAP PAM sequence is recognized, Cascade may utilize its RNA component (i.e. crRNA) to further examine the PAM-following sequence (i.e. protospacer) while forming the R-loop structure. PAM interaction and the crRNA-protospacer matching potential both affect the Cascade-target affinity, which may regulate the nuclease activity of the subsequently recruited Cas3. In the case of a fully-matched protospacer combined with a TIP PAM, interference and priming adaptation may both occur. When the PAM is mutated to a PAP sequence, or when some mismatches are introduced within protospacer, the Cas3 nuclease activity may be downregulated and only priming adaptation occurs.

with an identity of 22.6%, which may underlie their different PAM selectivity.

The interference target is predetermined by the invader-derived spacer sequences, which have been integrated into CRISPR arrays during adaptation. This indicates that indiscriminative adaptation would lead to self-targeting spacers, similar to those observed during naïve adaptation in Cas1- and Cas2-overexpressing *E. coli* cells (19). Therefore, the CRISPR adaptation machinery also requires a discrimination mechanism which has been elusive for years. Our recent study demonstrates that in the *H. hispanica* Type I-B system, a priming crRNA partially matching the invader DNA is essentially required for adaptation (20), suggesting that discriminative adaptation to foreign DNA may be achieved by this priming requirement. However, similar to the crRNA-guided interference, the crRNA-primed adaptation also has to preclude the host spacer DNA. Although mutations in the PAM sequence have previously been shown tolerated during adaptation-priming (17,20), here, our data demonstrate that PAM authentication does occur, but with relaxed stringency. This authentication tolerates as many as 23 PAM sequences, which we described as 'priming adaptation permissive' or PAP. Moreover, we confirmed this relaxed PAM selectivity by showing that PAM authentication is precisely positioned to the −1, −2 and −3 nucleotides of the priming protospacer. From Figure 3, the rules for a PAP PAM for the *H. hispanica* Type I-B CRISPR can be concluded as: (a) purines are not allowed at the −3 position; (b) T is favoured at the −2 and −3 positions; and (c) C is favoured at the −1 position. Within the CRISPR cassette, repeat nucleotides immediately preceding each spacer are consistently AGC, which is 'priming adaptation and interference non-permissive' or PAIN according to rule (a). This PAIN sequence will be ignored during PAM recognition, thereby protecting the spacer DNA from interference and adaptation-priming. Our data demonstrate that these three repeat nucleotides preceding a spacer DNA serve as its only protective determinant, suggesting mutations at these positions, particularly −3 and −2, can cause priming adaptation or even interference to the CRISPR DNA itself, and in this case, CRISPR immunity must be inactivated. Consistently, when the *E. coli* crRNA was mutated at these positions, interference to a TIP target was not observed for unknown reasons (21).

A previous study has proposed a model for Cascade-mediated target DNA recognition (16). Based on that model and our data here, we propose an integrated target-recognition model for CRISPR interference and

adaptation-priming in Figure 7. We speculate that the Cascade complex may utilize its protein subunit(s) to scan DNA molecules for a permissive PAM (TIP or PAP), and this process could preclude 'self' sequences preceded by a PAIN sequence, such as the spacer DNA. Once a TIP or PAP PAM is detected, Cascade may utilize its RNA component (i.e. crRNA) to further examine the spacer-matching potential of the PAM-following sequence (i.e. protospacer). By these two mechanisms, interference and priming adaptation could be discriminatively directed to the target-bearing invader DNA, but not to the crRNA-encoding 'self' DNA or other sequences. It should be noted that compared to interference, the adaptation-priming process seems to tolerate more PAM variations and more protospacer mutations. It has been reported that for a PAM- or protospacer-mutated target that escapes interference, the *E. coli* Cascade binds with decreased affinity (14), suggesting Cascade possibly requires stronger target affinity to elicit interference than to prime adaptation, which may explain their different tolerance.

Combining our previous finding that the *H. hispanica* system strictly requires a priming process for adaptation (20), we propose that the base pairing-independent PAM recognition and base pairing-dependent crRNA guidance together provide reliable 'target versus non-target' discriminations for CRISPR interference and adaptation pathways in this system. The relaxed PAM selectivity during the adaptation-priming process explains how this process maximizes its tolerance of PAM mutations of a target (that escape interference), while avoiding mis-targeting the spacer DNA. Though details may differ, these discrimination mechanisms may function similarly for other CRISPR systems where adaptation strictly requires being primed.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online including [1–6].

FUNDING

National Natural Science Foundation of China [31271334, 31330001]. Funding for open access charge: National Natural Science Foundation of China [31271334].
Conflict of interest statement. None declared.

REFERENCES

1. Sorek,R., Kunin,V. and Hugenholtz,P. (2008) CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.*, **6**, 181–186.
2. Horvath,P. and Barrangou,R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167–170.
3. Bolotin,A., Quinquis,B., Sorokin,A. and Ehrlich,S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**, 2551–2561.
4. 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.
5. Mojica,F.J., Díez-Villaseñor,C., García-Martínez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
6. Jansen,R., Embden,J.D., Gaastra,W. and Schouls,L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, **43**, 1565–1575.
7. 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.
8. Marraffini,L.A. and Sontheimer,E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, **322**, 1843–1845.
9. 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.
10. 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.
11. Haurwitz,R.E., Jinek,M., Wiedenheft,B., Zhou,K. and Doudna,J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355–1358.
12. Li,M., Liu,H., Han,J., Liu,J., Wang,R., Zhao,D., Zhou,J. and Xiang,H. (2013) Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon *Haloferax mediterranei*. *J. Bacteriol.*, **195**, 867–875.
13. Semenova,E., Jore,M.M., Datsenko,K.A., Semenova,A., Westra,E.R., Wanner,B., van der Oost,J., Brouns,S.J. and Severinov,K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10098–10103.
14. Westra,E.R., van Erp,P.B., Kunne,T., Wong,S.P., Staals,R.H., Seegers,C.L., Bollen,S., Jore,M.M., Semenova,E., Severinov,K. *et al.* (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, **46**, 595–605.
15. Sinkunas,T., Gasiunas,G., Fremaux,C., Barrangou,R., Horvath,P. and Siksnys,V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.*, **30**, 1335–1342.
16. Sashital,D.G., Wiedenheft,B. and Doudna,J.A. (2012) Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell*, **46**, 606–615.
17. 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.
18. 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.
19. 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.
20. 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.
21. Westra,E.R., Semenova,E., Datsenko,K.A., Jackson,R.N., Wiedenheft,B., Severinov,K. and Brouns,S.J. (2013) Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.*, **9**, e1003742.
22. Marraffini,L.A. and Sontheimer,E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, **463**, 568–571.
23. Brendel,J., Stoll,B., Lange,S.J., Sharma,K., Lenz,C., Stachler,A.E., Maier,L.K., Richter,H., Nickel,L., Schmitz,R.A. *et al.* (2014) A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of crRNAs in *Haloferax volcanii*. *J. Biol. Chem.*, **289**, 7164–7177.
24. Maier,L.K., Lange,S.J., Stoll,B., Haas,K.A., Fischer,S., Fischer,E., Duchardt-Ferner,E., Wöhnert,J., Backofen,R. and Marchfelder,A. (2013) Essential requirements for the detection and degradation of invaders by the *Haloferax volcanii* CRISPR/Cas system I-B. *RNA Biol.*, **10**, 865–874.

25. Fischer,S., Maier,L.K., Stoll,B., Brendel,J., Fischer,E., Pfeiffer,F., Dyall-Smith,M. and Marchfelder,A. (2012) An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J. Biol. Chem.*, **287**, 33351–33363.
26. Liu,H., Han,J., Liu,X., Zhou,J. and Xiang,H. (2011) Development of *pyrF*-based gene knockout systems for genome-wide manipulation of the archaea *Haloferax mediterranei* and *Haloarcula hispanica*. *J. Genet. Genomics*, **38**, 261–269.
27. Cai,S., Cai,L., Liu,H., Liu,X., Han,J., Zhou,J. and Xiang,H. (2012) Identification of the haloarchaeal phasin (PhaP) that functions in polyhydroxyalkanoate accumulation and granule formation in *Haloferax mediterranei*. *Appl. Environ. Microbiol.*, **78**, 1946–1952.
28. Jore,M.M., Lundgren,M., van Duijn,E., Bultema,J.B., Westra,E.R., Waghmare,S.P., Wiedenheft,B., Pul,U., Wurm,R., Wagner,R. *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, **18**, 529–536.
29. 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.
30. Sternberg,S.H., Redding,S., Jinek,M., Greene,E.C. and Doudna,J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62–67.
31. 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.
32. Wiedenheft,B., Lander,G.C., Zhou,K., Jore,M.M., Brouns,S.J., van der Oost,J., Doudna,J.A. and Nogales,E. (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, **477**, 486–489.
33. Makarova,K.S., Aravind,L., Wolf,Y.I. and Koonin,E.V. (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct*, **6**, 38