2714–2723 *Nucleic Acids Research, 2017, Vol. 45, No. 5* *Published online 29 December 2016*
doi: 10.1093/nar/gkw1309

DNA binding specificities of *Escherichia coli* Cas1–Cas2 integrase drive its recruitment at the CRISPR locus

Clara Moch†, Michel Fromant†, Sylvain Blanquet and Pierre Plateau*

Laboratoire de Biochimie, Ecole polytechnique, CNRS, Université Paris-Saclay, 91128 Palaiseau cedex, France

Received August 04, 2016; Revised November 24, 2016; Editorial Decision December 14, 2016; Accepted December 15, 2016

ABSTRACT

Prokaryotic adaptive immunity relies on the capture of fragments of invader DNA (protospacers) followed by their recombination at a dedicated acceptor DNA locus. This integrative mechanism, called adaptation, needs both Cas1 and Cas2 proteins. Here, we studied *in vitro* the binding of an *Escherichia coli* Cas1–Cas2 complex to various protospacer and acceptor DNA molecules. We show that, to form a long-lived ternary complex containing Cas1–Cas2, the acceptor DNA must carry a CRISPR locus, and the protospacer must not contain 3′-single-stranded overhangs longer than 5 bases. In addition, the acceptor DNA must be supercoiled. Formation of the ternary complex is synergistic, in such that the binding of Cas1–Cas2 to acceptor DNA is reinforced in the presence of a protospacer. Mutagenesis analysis at the CRISPR locus indicates that the presence in the acceptor plasmid of the palindromic motif found in CRISPR repeats drives stable ternary complex formation. Most of the mutations in this motif are deleterious even if they do not prevent cruciform structure formation. The leader sequence of the CRISPR locus is fully dispensable. These DNA binding specificities of the Cas1–Cas2 integrase are likely to play a major role in the recruitment of this enzyme at the CRISPR locus.

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPRs) and the CRISPR-associated (Cas) proteins are at the basis of bacterial and archaeal adaptive and inheritable immune systems that protect microorganisms from invasive phages and plasmids (1–3). CRISPR genomic loci are composed of arrays of direct repeats separated by variable sequences, called spacers, which are generally derived from invader genetic elements (4–6). Repeats usually show some dyad symmetry potentially allowing formation of DNA cruciform structures and hairpin-loop folds in RNA transcripts. Within a locus, repeats and inserts each have conserved size, typically 20–50 bp in length, depending on the considered CRISPR system type. Immediately upstream to this repeat-spacer array is an AT-rich region called leader, which harbors a transcription promoter (7,8). *In vivo*, this region is essential for acquisition of new spacers (9).

Cas proteins are encoded by genes located at the vicinity of the CRISPR loci (10,11). These proteins are involved in the three major steps of the CRISPR–Cas system mode of action: adaptation, expression and interference. During adaptation, protospacer sequences are selected among invader DNA pieces and inserted at the leader-proximal end of a CRISPR array to generate an additional spacer (12). In the expression phase, the repeat-spacer array is transcribed from the leader into a pre-crRNA that is further cleaved within each repeat into individual shorter crRNAs (13,14). During the interference phase, invading nucleic acids matching a crRNA are recognized and degraded by a ribonucleoprotein complex associating the crRNA and dedicated Cas proteins (13,15–16).

To be integrated at a CRISPR locus, protospacers formed from invader DNA need to be flanked by a special adjacent motif of 2–5 nucleotides (17–19). These protospacer adjacent motifs (PAMs) direct the orientation of the newly acquired spacer with respect to the position of the leader sequence. Presence of such PAMs in invader DNA explains the resistance of the cell to auto-immunity (20,21). Indeed, PAMs govern integration at a CRISPR locus during adaptation and, in turn, only DNAs harboring PAMs are cleavable during the interference phase (16,22–23). As a consequence, the interference machinery does not recognize the CRISPR DNA locus because the latter no longer contains PAM sequences adjacent to the spacers. Only the invading DNA pieces, which harbor a PAM, can be attacked.

While the molecular mechanisms involved in the expression and interference phases are now well characterized, some features of the adaptation phase remain to be deep-

*To whom correspondence should be addressed. Tel: +33 1 69 33 48 88; Fax: +33 1 69 33 49 09; Email: pierre.plateau@polytechnique.edu
† These authors contributed equally to this work as first authors.

ened (24,25). Genetic experiments have shown that, in *Escherichia coli* K-12, Cas1 and Cas2 proteins sustain new spacer acquisition (9). These two proteins form a stable complex in which a Cas2 dimer links two Cas1 dimers (26). Structural and biochemical data indicate that, when captured by Cas1–Cas2, a DNA protospacer adopts a dual forked form with a central double-stranded stem flanked by single-stranded overhangs (27,28). Cas2 recognizes the double-stranded region while Cas1 binds the 3′ single-stranded flanks. At this stage, Cas1 may cleave the protospacer at the correct position with respect to its PAM (28) and catalyze its integration as a new spacer at a CRISPR locus (27,28). In *E. coli*, the RecBCD double-stranded DNA break repair complex appears to be involved in the generation of the partially single-stranded DNA fragments that are further captured by Cas1–Cas2 prior to integration (29).

Several features of the integration reaction have already been revealed by *in vitro* experiments. Purified *E. coli* Cas1–Cas2 complex catalyzes a half-site integration, i.e. the covalent attachment of a double stranded oligonucleotide to an acceptor DNA (30). A high activity was found when protospacers possessed a 5-nucleotides 3′-overhang and plasmid acceptor DNA was supercoiled, whatever the presence or absence of a functional CRISPR locus in the plasmid (27–28,30). With plasmids carrying a CRISPR locus, ∼70% of the half-site integration events take place at the borders of all CRISPR repeats. With the control plasmid devoid of CRISPR locus, frequent integration sites were found adjacent to a plasmid inverted repeat sequence having a propensity to form a DNA cruciform fold (30). These observations, together with the requirement of supercoiled acceptor plasmid DNA for integration (30) and the high affinity of Cas1 alone for cruciform DNA structures (31), suggested that recognition by Cas1–Cas2 of a cruciform DNA is an important step in the protospacer integration reaction. *In vivo*, integration takes place at the leader-repeat boundary only. Presence in the leader region of a binding site for integration host factor (IHF) (32), a heterodimeric protein that binds to numerous gene regulation regions in *E. coli* (33), is involved in the specificity of this reaction. One IHF subunit has previously been identified among the proteins copurifying with Cas1 (31). However, whether a direct contact between IHF and Cas1–Cas2 occurs in the course of the recognition of the CRISPR locus remains to be determined.

In vitro studies have also revealed that bases −2, −1 and +1 at the leader-repeat boundary governed the rate of the disintegration reaction (the reverse of the half-site integration reaction) catalyzed by Cas1 *in vitro* (34). Since an enzyme cannot change the equilibrium between the integration and disintegration reactions, a higher disintegration rate actually implies a higher integration rate.

We undertook the present study to go deeper in the dissection of the repeat motif sequence and/or of the leader elements at a CRISPR locus that are involved in Cas1–Cas2 integrase binding. To this end, we implemented electrophoretic mobility shift assays (EMSAs) in order to follow complex formation between Cas1–Cas2, the protospacer and the acceptor plasmid DNA *in vitro*. Results show that a long-lived complex associating the three partners can be formed provided acceptor DNA is supercoiled and displays a CRISPR locus. Systematic assays with 30 mutated CRISPR loci indicate that, in fact, the presence of one palindromic motif of the repeat is enough to obtain the long-lived complex, whatever the presence or absence of a contiguous leader region. Mutagenesis of the bases in this motif shows that most of them are important to drive the interaction.

MATERIALS AND METHODS

Cloning

A plasmid expressing both native Cas1 and a C-terminal His_6-tagged version of Cas2 was constructed by amplifying the *cas1* and *cas2* genes from the chromosomal DNA of *E. coli* K12 (MG1655 (35)) using oligonucleotides OCN504 and OCN483 for *cas1*, and OCN506 and OCN507 for *cas2* (Supplementary Table S1). Amplified *cas1* was inserted between the NcoI and PstI sites of pETDuet-1 expression vector (Novagen) and amplified *cas2* was inserted into the NdeI and KpnI sites of the resulting plasmid, to give plasmid pMFC1+CTc2.

A plasmid only expressing an N-terminal His_6-tagged version of Cas1 was constructed by amplifying the *cas1* gene from the chromosomal DNA of strain MG1655, using oligonucleotides OCN482 and OCN483 (Supplementary Table S1). Amplified *cas1* was inserted between the NcoI and PstI sites of pETDuet-1 to give plasmid pMFCT1.

To clone wild-type and mutant CRISPR loci, plasmids pCOLA-Z0, pBS-Z0 and pBS-Z2 to -Z29 were constructed by annealing two oligonucleotides (Supplementary Table S1) and inserting the resulting fragment into the SacI and KpnI sites of pCOLADuet-1 (Novagen) or pBluescript SK+ (Agilent Technologies), respectively. Plasmid pBS-Z1 was constructed by inserting into SacI-KpnI-cut pBluescript SK+ a SacI-KpnI fragment harboring the last 62 bp of the CRISPR leader region, two repeats and a spacer derived from the cII gene of bacteriophage λ. The sequence of this insert is given in Supplementary Figure S1.

Gene expression and protein purification

Cas1–Cas2 complex was purified from FB810 cells (a *recA*⁻ derivative of BL21(λDE3) (36)) transformed with plasmid pMFC1+CTc2. This strain expressed both native Cas1 and a C-terminally His_6-tagged Cas2. Cells were grown at 37°C in 8 l of 2xTY medium containing 100 μg/ml ampicillin. When the optical density of the culture reached 0.7 at 650 nm, IPTG was added at a final concentration of 0.4 mM, and cells were further incubated at 22°C for 24 h. They were harvested by centrifugation (5000 g, 30 min) and resuspended in 300 ml of buffer A (20 mM Tris–HCl, pH 7.5, 150 mM KCl, 2 mM 2-mercaptoethanol) containing two complete EDTA-free protease inhibitor cocktail tablets (Roche). Cells were disrupted by sonication (10 min, 4°C) and debris was removed by centrifugation (35 000 g, 20 min). The supernatant was applied on a Talon column (5 ml, Clontech) equilibrated in buffer A, then washed with buffer A and eluted with a linear gradient from 0 to 200 mM imidazole in buffer A (1 ml/min, 3.5 mM/min). Fractions containing His_6-tagged Cas1–Cas2 complex were identified

by SDS-PAGE analysis. They were pooled, precipitated by ammonium sulfate (70% saturation) and centrifuged for 20 min at 12 000 g. The pellet was resuspended with 0.6 ml of buffer A and dialyzed against the same buffer. The resulting sample (2 ml) was loaded on a Superdex-75 molecular sieve column (16 × 60 cm, GE Healthcare) equilibrated and eluted with buffer A (0.4 ml/min). Fractions containing the Cas1–Cas2 complex were pooled, dialyzed against buffer B (20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 0.1 mM dithiothreitol (DTT), 60% glycerol), and stored at −20°C. SDS-PAGE analysis indicated that impurities in the Cas1–Cas2 preparation were lower than 10%.

N-terminally His_6-tagged Cas1 was purified from strain BL21(λDE3) (Novagen) transformed with pMFCT1 plasmid. Cells were grown in 1 l of 2× TY medium containing 100 μg/ml ampicillin. Induction by IPTG and disruption of the cells (in 40 ml) were performed as described above for Cas1–Cas2 purification. Cell extract was applied on a 2 ml Talon column which was washed and eluted as described above, except that the flow rate was 0.5 ml/min. Fractions containing tagged Cas1 were pooled, brought to 1.2 M ammonium sulfate and applied on a Widepore HI-Propyl column (0.5 × 5.5 cm, particle size 15 μm, Baker) equilibrated in a 20 mM Tris–HCl buffer (pH 7.5) containing 1.7 M ammonium sulfate and 0.1 mM DTT. Elution was carried out with a linear gradient of 1.7–0 M ammonium sulfate (0.15 ml/min, 1 M/h) in a 20 mM Tris–HCl buffer (pH 7.5) containing 0.1 mM DTT. Fractions containing Cas1 were pooled, dialyzed against buffer A and purified again on a Talon column. The resulting protein sample was concentrated by ammonium sulfate precipitation (70% saturation), dialyzed against buffer B and stored at –20°C. The homogeneity of the purified Cas1 protein was estimated by SDS-PAGE analysis to be >90%.

Electrophoretic mobility shift assays

EMSAs were performed using unlabeled or fluorescent 5′ DY682- or DY782-labeled oligonucleotides (from MWG). Pre-annealed double-stranded protospacer molecules (40–400 nM) were first incubated with Cas1–Cas2 (0-280 nM) for 10–15 min at 22°C in a 20 mM Hepes–KOH buffer (pH 7.5) containing 50 mM KCl, 5 mM EDTA, 1 mM DTT, 0.15% (v/v) Tween 20 and 100 μg/ml bovine serum albumin. Then, target plasmid DNA (7.5 nM) was added and incubation was continued at 22°C for the time indicated in the figures. Resulting samples were separated by electrophoresis in 0.5% (w/w) agarose gels containing 45 mM Tris-borate (pH 8.3) and 1 mM EDTA. Gels were scanned with a Li-Cor Odyssey CLx imaging system before staining with ethidium bromide. poly(dI-dC) was from Sigma. Before use, it was dialyzed against a 1 mM Hepes-KOH buffer (pH 7.5).

To estimate the apparent K_d value governing the interaction between Cas1–Cas2–protospacer complex and acceptor DNA, EMSAs were performed in the presence of 7.5 nM acceptor plasmid DNA, 0–400 nM Cas1–Cas2 and protospacer. Protospacer concentration was systematically in excess over that of Cas1–Cas2 by a factor of 1.2. Iterative nonlinear fits of the theoretical binding equation to the experimental values were performed using the Levenberg–Marquardt algorithm (37) and assuming the binding of one binary complex to acceptor plasmid DNA.

Protospacer half-site integration assays

Complexes between Cas1–Cas2 and DNA were prepared as described above for EMSAs. After the indicated incubation times at 22°C, $MgCl_2$ was added at a final concentration of 10 mM (5 mM in excess over the EDTA concentration). After additional incubation of 5 min at 22°C, the reaction was stopped by addition of EDTA (20 mM final) and SDS (0.5% (w/w) final). Resulting samples were analyzed by agarose gel electrophoresis as described above for EMSAs.

RESULTS

Cas1–Cas2, protospacer and CRISPR-containing plasmid DNA form a stable ternary complex.

The ability of Cas1–Cas2 complex to bind both protospacer and target DNA was studied by agarose gel mobility assays using a 5′ fluorescent-labeled model protospacer and a CRISPR-containing plasmid as the acceptor. Because recently determined structures of Cas1–Cas2–protospacer complexes showed that a privileged ligand of Cas1–Cas2 was a 23-bp duplex flanked by 3′-terminal non-paired overhangs of five nucleotides (27,28), we chose such dual-forked DNAs as potential model protospacers. In a first step, protospacers P1, Q1 and R1 were designed (see Supplementary Table S1 for the corresponding oligonucleotide sequences). The binding results obtained with these three protospacers were very similar. One of them (P1, Figure 1) was adopted for further experiments. Acceptor DNA was the supercoiled pBS-Z0 plasmid, a pBluescript SK+ derivative harboring both one CRISPR repeat and a truncated 62-bp region of the leader sequence that is sufficient to obtain acquisition of new inserts *in vivo* (9). Our incubation buffer contained EDTA to prevent covalent reaction of the protospacer with the plasmid DNA (30).

Analysis of a sample containing Cas1–Cas2 (140 nM), 5′ DY782-protospacer P1 (400 nM) and CRISPR-containing plasmid pBS-Z0 (7.5 nM) revealed a discrete band migrating slightly more slowly than supercoiled pBS-Z0. This band could be efficiently stained with ethidium bromide and was fluorescent upon excitation of the DY782 marker (Figure 1, lane 1). In the absence of either Cas1–Cas2, protospacer or acceptor DNA, the fluorescent slower-migrating band did not appear (Figure 1, lanes 2, 3 and 9). Therefore, this band was assumed to correspond to a ternary complex composed of Cas1–Cas2, fluorescent protospacer and plasmid DNA. This complex was non-covalent since it disappeared upon SDS addition to the sample before electrophoresis (Figure 1, lane 11).

With control plasmid pBluescript SK+, no ternary complex could be detected (Figure 1, lane 6), suggesting that the presence of a CRISPR locus on pBS-Z0 was necessary to the formation of a stable ternary complex. This conclusion was confirmed by introducing the same CRISPR locus region into another plasmid, pCOLADuet-1. Again, a fluorescent band appeared with the sample containing Cas1–Cas2, protospacer and plasmid DNA. It migrated slightly

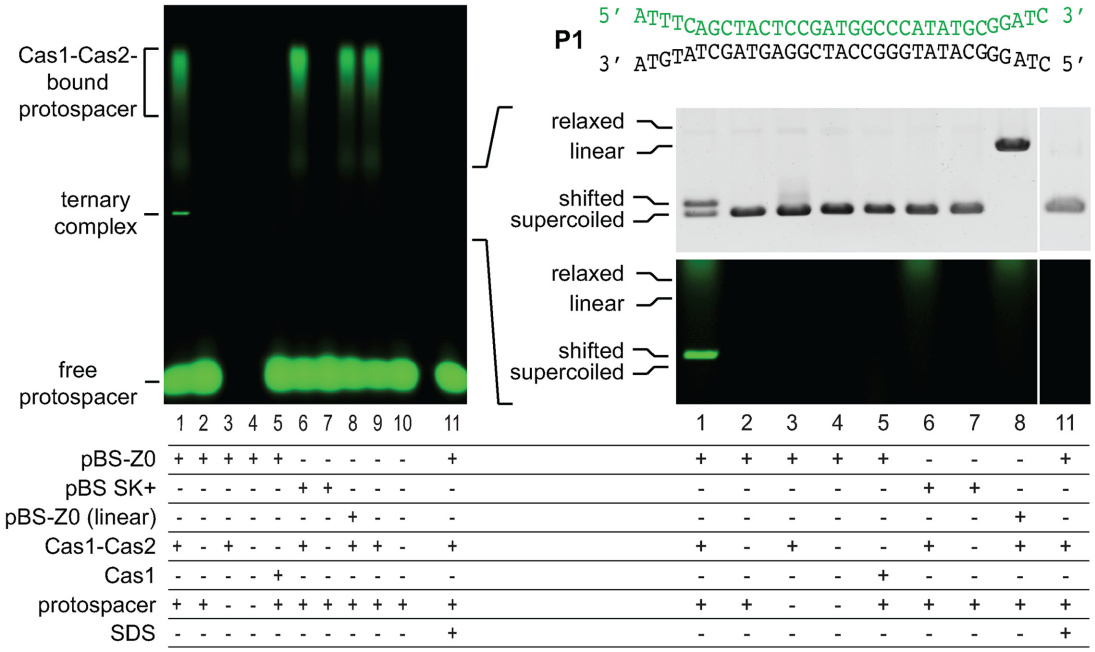

Figure 1. Formation of a stable ternary complex between Cas1–Cas2, protospacer P1 and CRISPR-containing plasmid pBS-Z0. Samples contained combinations of the indicated compounds: 400 nM of the 5′ DY782-labeled protospacer P1 (sequence at the top of the Figure), 140 nM of Cas1 or of Cas1–Cas2 complex and 7.5 nM of either supercoiled pBS-Z0, supercoiled pBluescript SK+ or KpnI-cut pBS-Z0. The sample with Cas1–Cas2, protospacer P1 and supercoiled plasmid pBS-Z0 was also treated with SDS at a final concentration of 0.5% (w/w) prior to electrophoresis (lane 11). All samples were analyzed by EMSA. The gel was scanned for fluorescence of DY782 (in green) and stained with ethidium bromide. The right side of the Figure is a zoom of the central part of the gel.

more slowly than supercoiled DNA (Supplementary Figure S2, lane 2). No band was visible when using the control plasmid pCOLADuet-1. Control experiments were performed in the presence of Cas1 alone instead of Cas1–Cas2. No ternary complex could be detected (Figure 1, lane 5 and Supplementary Figure S2, lane 6), showing that formation of the ternary complex requires the presence of Cas2.

Binding experiments were performed at a given concentration of pBS-Z0 (7.5 nM), in the presence of various concentrations of Cas1–Cas2 (0–400 nM) and of protospacer in excess over Cas1–Cas2 (see Materials and Methods). The gels of Supplementary Figure S3 show that addition of Cas1–Cas2 increased the amount of bound plasmid at the expense of free plasmid. An apparent dissociation constant of 8.7 ± 1.3 nM could be deduced, assuming a 1:1 stoichiometry. With non-CRISPR pBluescript SK+, no association of the Cas1–Cas2–protospacer complex could be detected (gel is shown in Supplementary Figure S3). The same experiments, with pBS-Z0 and pBluescript SK+, were performed in the absence of protospacer. Smears became visible on the gels at the highest concentrations of Cas1–Cas2. In parallel, the intensities of the free plasmids decreased, suggesting weak binding of Cas1–Cas2 to acceptor DNA, likely at non-specific sites. The deduced binding curves indicated decrease in affinity by a factor >10, if compared to the curve obtained with pBZ-Z0 in the presence of protospacer. Altogether, the results show that (i) the affinity for Cas1–Cas2–protospacer complex is much higher when the acceptor plasmid carries a CRISPR locus and (ii) non-specific Cas1–Cas2 binding at other sites is disfavored when the protospacer is present. In brief, protospacer has a synergistic action toward Cas1–Cas2 recruitment at a CRISPR site whereas it is antagonistic towards non-specific binding at non-CRISPR sites.

Supercoiling of the acceptor DNA molecule was previously shown to favor *in vitro* half-site integration of a protospacer catalyzed by Cas1–Cas2. As shown in Figure 1 (lane 8), linearized pBS-Z0 and pCOLA-Z0 plasmids failed to form a stable ternary complex with Cas1–Cas2 and protospacer DNA, evidencing that formation of such a complex indeed needs supercoiled acceptor DNA.

The ternary complex between Cas1–Cas2, protospacer and CRISPR-containing plasmid DNA is long-lived

Chase experiments were performed to estimate the lifetime of the ternary complex described above. A complex between Cas1–Cas2, pBS-Z0 and unlabeled protospacer P1 or protospacer P1 labeled with DY782 (DY682) was left to form for 10 min at 22°C. Next, protospacer P1 carrying DY682 (DY782) was added. The samples were further incubated at 22°C for various times before analysis on an electrophoresis agarose gel. After 4 h of incubation, <2% of the initial protospacer engaged in a binary complex with Cas1–Cas2 or in a ternary complex with Cas1–Cas2 and acceptor DNA had exchanged with the chase protospacer (Figure 2). Similar slow exchanges between free- and protein-bound protospacer were observed whatever the labeling of the chased protospacer (5′-labeled with DY682 or DY782, or unlabeled), excluding that the 5′-fluorescent label could have interfered with the high stability of the binary complex between Cas1–Cas2 and protospacer DNA.

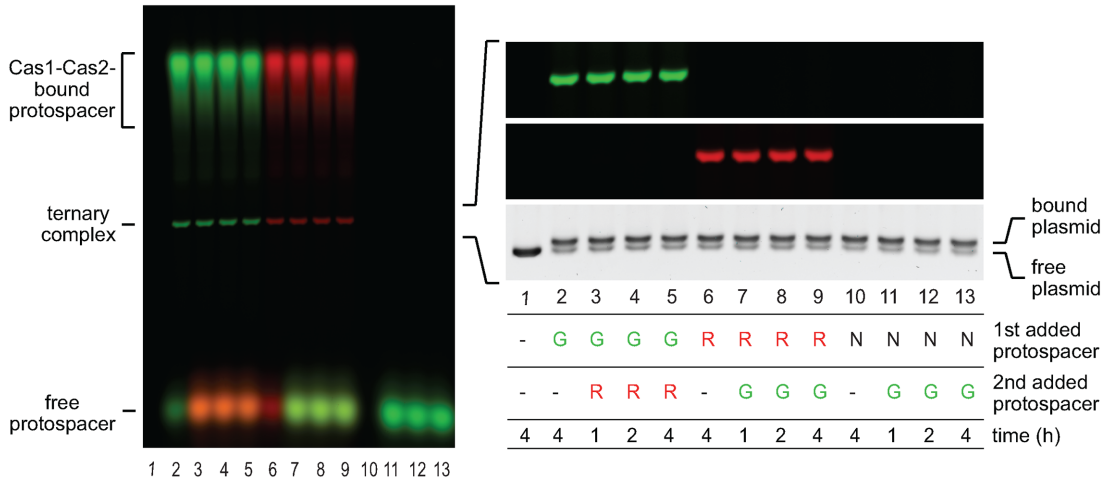

Figure 2. Lifetime of complexes between Cas1–Cas2, protospacer and acceptor DNA: chase with the addition of free protospacer. Samples contained 200 nM of protospacer P1 (Figure 1), 140 nM of Cas1–Cas2 and 7.5 nM of pBS-Z0. Protospacer P1 was either unlabeled (N) or labeled with DY782 (G) or DY682 (R). After 10 min at 22°C, an identical protospacer, but with a different fluorophore (R or G), was added and incubation was continued at 22°C for the indicated times. Resulting samples were analyzed by EMSA. The gel was scanned for fluorescence of DY782 (in green) or DY682 (in red) and stained with ethidium bromide. On the entire gel shown on the left side of the Figure, the DY782 and DY682 images are superimposed. The right side is a zoom of the central part of the gel.

The above experiment demonstrates that protospacer association within the binary or the ternary complex is remarkably long-lived. We next examined whether, once bound to acceptor DNA, the binary Cas1–Cas2–protospacer complex could rapidly exchange with binary complexes free in solution. To address this question, two mixtures were performed. One contained Cas1–Cas2, DY782-labeled protospacer (green) and pBS-Z0 DNA. The other one contained Cas1–Cas2, DY682-labeled protospacer (red) and pCOLA-Z0, a bigger CRISPR-containing plasmid. The two samples were mixed and incubation at 22°C was started. Figure 3 shows that, after 4 h of incubation, <5% of the initial complex had exchanged with the complex of the other color, hence demonstrating that, beyond the stable binary Cas1–Cas2 protospacer complex, the ternary complex including acceptor DNA is also long-lived.

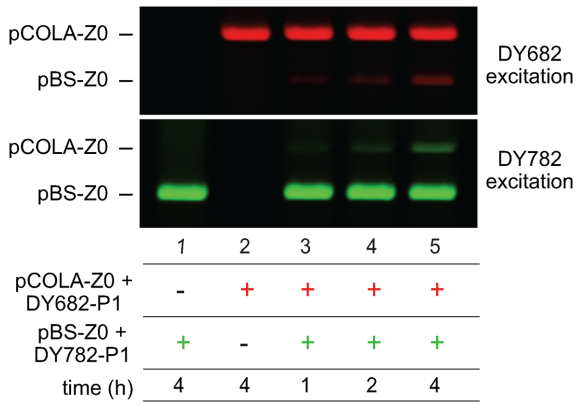

Figure 3. Lifetime of complexes between Cas1–Cas2, protospacer and acceptor DNA: chase with the addition of Cas1–Cas2-bound protospacer. Two samples were prepared. One contained 200 nM of DY782-labeled protospacer P1 (Figure 1), 140 nM of Cas1–Cas2 and 7.5 nM of pBS-Z0. The other one was identical except that protospacer P1 was labeled with DY682 instead of DY782 and that acceptor DNA was pCOLA-Z0 instead of pBS-Z0. After 15 min incubation at 22°C, the two samples were mixed and incubation at 22°C was continued for the indicated times. Resulting samples were analyzed by EMSA. The gel was scanned for fluorescence of DY782 (in green) or DY682 (in red). Shown is the part of the gel where ternary complexes migrate.

Non-CRISPR DNA competes with CRISPR DNA for binding to Cas1–Cas2

The above results show that formation of a stable complex between Cas1–Cas2, protospacer and acceptor DNA requires the presence of a CRISPR locus on the acceptor DNA. However, it was previously shown that, *in vitro*, Cas1–Cas2 may integrate a protospacer at sites different from the CRISPR locus (30). This implied that non-CRISPR DNA sequences can bind Cas1–Cas2. As shown in Figure 1 and Supplementary Figure S3, Cas1–Cas2 association with pBluescript SK+ in the presence of protospacer could not, however, be directly evidenced in EMSA conditions. This may reflect inability of a non-CRISPR DNA ternary complex to resist electrophoresis conditions. To solve this paradoxical question, we measured ternary complex formation with CRISPR-carrying plasmid pCOLA-Z0 in the presence of various concentrations of non-specific competitor DNA. These EMSA experiments were performed at a Cas1–Cas2 concentration (3 nM) lower than that of acceptor DNA (7.5 nM). Various concentrations of competitor DNA were added to the sample prior to pCOLA-Z0 addition. Next, the sample was left to incubate for various times (1-4 h). In the presence of either a non-CRISPR plasmid (pBluescript SK+) or poly(dI-dC) as competitors, ternary complex formation with plasmid pCOLA-Z0 decreased as a function of competitor DNA concentration (Figure 4). At a ratio of 16 between the concentrations (in bp) of non-CRISPR and CRISPR plasmids, the amount of the ternary complex formed with the pCOLA-Z0 CRISPR plasmid was reduced by a factor of

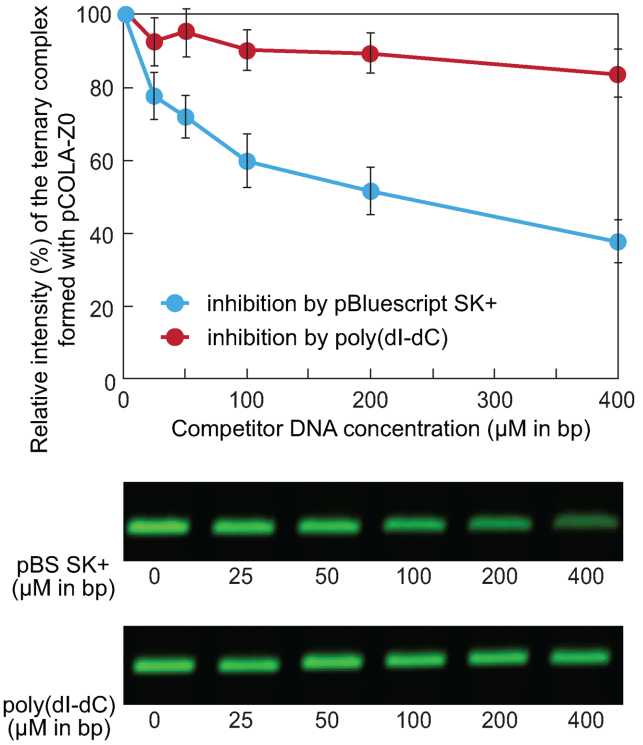

Figure 4. Inhibition by non-specific DNA in the formation of a ternary complex between Cas1–Cas2, protospacer P1 and CRISPR-containing plasmid pCOLA-Z0. Samples were prepared by mixing 40 nM of DY782-labeled protospacer P1 (Figure 1), 3 nM Cas1–Cas2 and various concentrations of either pBluescript SK+ (blue symbols) or poly(dI-dC) (red symbols) as competitor DNA. After 10 min at 22°C, 7.5 nM of target CRISPR-containing plasmid pCOLA-Z0 was added (27 μM in bp) and incubation was continued for 2 h at 22°C. Resulting samples were analyzed by EMSA. The gel was scanned for fluorescence of DY782. In the graph, the intensity of the band of the ternary complex formed with pCOLA-Z0 is calculated as a percentage of the intensity obtained in the absence of competitor DNA. Experiments were made in triplicate. Error bars represent standard deviations (SD). The gels under the graph display a representative example of the experiments. Shown are the parts of the gels where ternary complexes migrate.

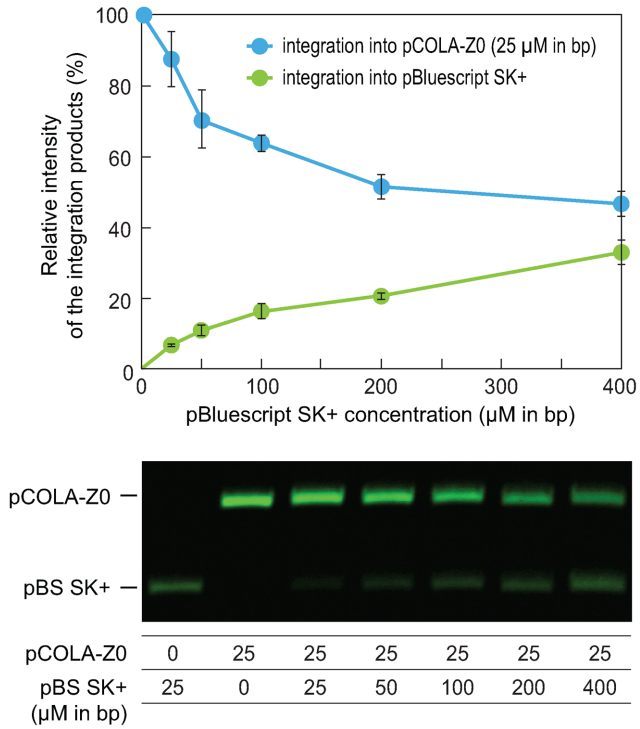

Figure 5. Inhibition by non-specific DNA of the Cas1–Cas2-catalyzed half-site covalent reaction of protospacer P1 with CRISPR-containing plasmid pCOLA-Z0. Samples were prepared by mixing 40 nM DY782-labeled protospacer P1 (Figure 1), 3 nM Cas1–Cas2 and various concentrations of pBluescript SK+ as competitor DNA. After 10 min at 22°C, 7.5 nM of CRISPR-containing plasmid pCOLA-Z0 was added (27 μM in bp) and incubation was continued for 2 h at 22°C. At this step, covalent reaction was started by addition of 10 mM $MgCl_2$ (in excess of EDTA (5 mM) already present). After a further incubation of 5 min at 22°C, reactions were quenched by addition of SDS (0.5%, w/w) and excess EDTA (20 mM). Resulting samples were analyzed on a 0.5% agarose gel, which was scanned for fluorescence of DY782. In the graph, the intensity of the band corresponding to integration of the protospacer into pCOLA-Z0 (blue symbols) or pBluescript SK+ (green symbols) is shown as a percentage of the intensity of the product obtained with pCOLA-Z0 in the absence of competitor. Experiments were made in triplicate. Error bars represent SD. The gel under the graph displays a representative example of the experiments. Shown is the part of the gel where integration products migrate.

~2. At a ratio of 16 between poly(dI-dC) competitor and acceptor DNA bp concentrations, inhibition was smaller, with 17% reduction of the ternary complex intensity. Whatever the pBluescript SK+ concentration in the assay, ternary complex formation with this plasmid could never be detected on the gel. These results indicated that the apparent affinity for CRISPR DNA in ternary complex formation was much higher than that for non-specific DNA. Moreover, the inability of non-specific DNA–Cas1–Cas2–protospacer complex to resist EMSA conditions is confirmed.

Formation of complexes between Cas1–Cas2–protospacer and CRISPR or non-CRISPR DNA was also studied in the presence of Mg^{2+} ions. In these ionic conditions, Cas1–Cas2 was reported to mediate covalent half-site integration of a protospacer into acceptor DNA (30). After incubation (see Materials and Methods), SDS was added to dissociate enzyme from DNA prior to agarose gel electrophoresis. In the presence of either a CRISPR or a non-CRISPR plasmid as acceptor DNA, a fluorescent band migrating close to that of the open-circular plasmid became visible on the gel. This band reflects covalent attachment of the fluorescent protospacer, which is accompanied by the introduction of a nick into the plasmid.

Competition experiments in the covalent reaction of protospacer with CRISPR (pCOLA-Z0) or non-CRISPR (pBluescript SK+) plasmids indicated that the reaction with CRISPR DNA was much favored (Figure 5). Interestingly, the shape of the inhibition curve as a function of non-CRISPR DNA obtained in these experiments (Figure 5) resembled that observed in the inhibition of ternary complex formation (Figure 4). Altogether this set of experiments suggest that the long-lived ternary complex involving CRISPR DNA is proficient in catalyzing grafting of protospacer to acceptor DNA, provided Mg^{2+} ions are present.

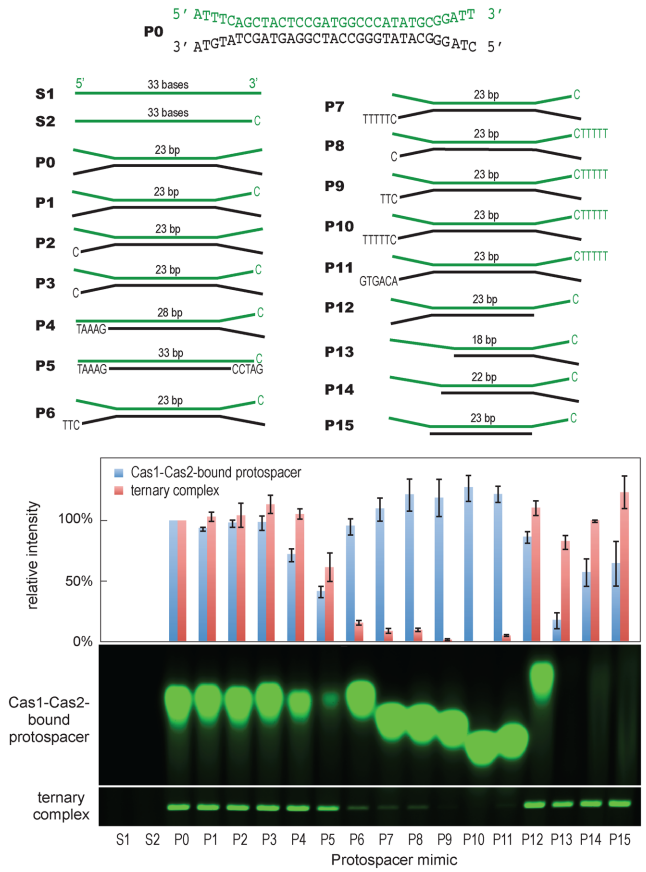

Figure 6. Formation of a ternary complex between Cas1–Cas2, CRISPR-containing plasmid pBS-Z0 and various protospacer mimics. Sequence of the reference protospacer P0 is shown at the top of the Figure. Changes relative to this sequence are indicated beneath. The sequences of all oligonucleotides are given in Supplementary Table S1. Several of the assayed oligonucleotides contained a C at their 3′-end(s). This base corresponds to a base of the PAM that is specifically recognized by Cas1 (28). Samples contained 400 nM of the oligonucleotide under study, 280 nM Cas1–Cas2 and 7.5 nM pBS-Z0. After 2 h incubation at 22°C, samples were analyzed by EMSA. The gel was scanned for fluorescence of DY782. In the histogram, the intensities of the bands obtained with each oligonucleotide are given as percentages of the intensities obtained with the P0 reference protospacer. Experiments were made in triplicate. Error bars represent SD. The gel under the histogram is a representative example of the experiments. The upper panel shows the part of the gel where Cas1–Cas2–protospacer complexes migrate whereas the lower panel shows the region of ternary complexes (see Figure 1 for the image of an entire gel). To improve the visualization of the ternary complexes, the brightness intensity of the lower panel was increased by a factor of 6 as compared to that used in the upper panel.

Various protospacer mimics can form stable complexes in the presence of both Cas1–Cas2 and a CRISPR-containing plasmid

We next investigated the ability of various oligonucleotides to form stable complexes with Cas1–Cas2 and CRISPR-containing plasmid pBS-Z0. Fluorescent oligonucleotides assayed were variants of a reference protospacer P0 (Figure 6), which consists of a base-paired stem of 23 bp with 5-nucleotides 3′- and 5′-overhangs. Some of the oligonucleotides we examined possessed a C at position 5 of the 3′-overhang, which corresponds to a base of the PAM that is believed to specifically interact with Cas1 (28). Some possessed the entire CTT PAM sequence (Figure 6).

Single-stranded oligonucleotides did not form stable complexes either with Cas1–Cas2 alone or with Cas1–Cas2 plus the CRISPR-containing plasmid pBS-Z0 (Figure 6). In contrast, perfectly or non-perfectly matched duplexes all formed detectable binary complexes with Cas1–Cas2. Slight discrepancies between mobilities of these complexes is explained by the lengths of the protospacers. The longer the protospacer, the higher was the mobility in the gel, likely because of a higher negative electric charge. In some cases (oligonucleotides P13 to P15 in Figure 6), the binary complex migrated as a smear rather than as a distinct spot, suggesting partial dissociation of the complex under electrophoresis conditions. Most of the assayed oligonucleotides formed stable ternary complexes with Cas1–Cas2 plus pBS-Z0 (Figure 6). Duplexes with two 3′-overhangs longer than five bases (P9 to P11) failed, however, to associate with the plasmid. Complex formation with oligonucleotides having one 3′-overhang longer than five bases (P6 to P8) was significantly smaller than that observed with the other oligonucleotides. Therefore, exceeding length of the 3′-overhang appears unfavorable in the formation of a stable ternary complex.

The palindromic motif of the CRISPR repeat is determinant in the formation of a stable complex with Cas1–Cas2 and protospacer DNA

To determine which part of the leader-repeat sequence is necessary to drive stable complex formation in the presence of Cas1–Cas2 and protospacer P1, 29 variants of pBS-Z0 were constructed. Deletions were introduced in the leader or in the repeat region of a minimalist CRISPR locus. Stable ternary complexes with Cas1–Cas2 and protospacer occurred with plasmids pBS-Z2 and pBS-Z4, despite the shortened leader region (Figure 7). In contrast, absence of the repeat (pBS-Z3) prevented formation of the ternary complex. Therefore, we concluded that the repeat was necessary and sufficient for the formation of a stable ternary complex. With pBS-Z1, which harbors two repeats, two fluorescent bands were visible on the gel, suggesting formation of plasmids with either one or two bound Cas1–Cas2–protospacer binary complexes (Figure 7).

The wild-type repeat contains a palindromic motif that may adopt a cruciform DNA structure with two seven-base stems and two four-base loops. Mutations outside the motif (pBS-Z20, -Z21, -Z26 and -Z27) had much smaller effects than mutations involving the motif (pBS-Z5 to -Z7, -Z9) (Figure 7). In addition, complete leader deletion of the leader sequence (pBS-Z8, -Z20, -Z26 and -Z27) did not prevent stable ternary complex formation. These results point out the importance of either the palindromic sequence itself or the formation of a hairpin structure. They also show that the leader sequence of PBS-Z0 is fully dispensable.

We noted that pBluescript SK+ and the above-mentioned plasmids contained a palindromic sequence in the ampicillin resistance gene. The same palindromic sequence is carried by pUC19, which allowed Nuñez *et al.* to show that non-CRISPR palindromic motifs may promote integration *in vitro* (30). Lack in our study of long-lived ternary complex

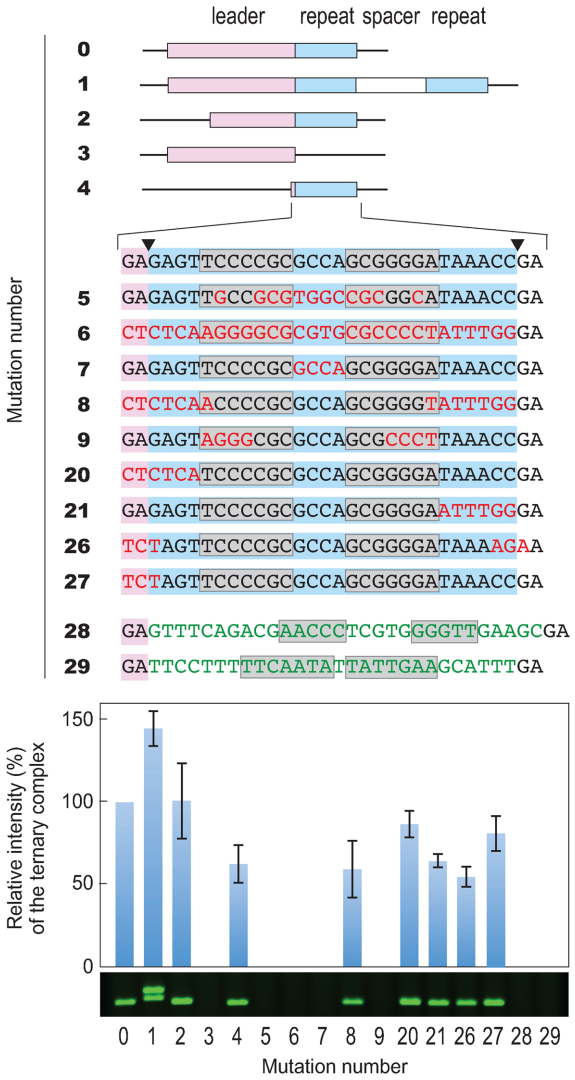

Figure 7. Binding of a Cas1–Cas2–protospacer complex to various mutant CRISPR loci. Mutant CRISPR loci introduced into plasmid pBluescript SK+ are schematized at the top of the Figure. The leader, repeat and spacer regions are shown as pink, blue and white rectangles, respectively. Over the sequence of pBS-Z4, the position of the repeat is delineated by black triangles. The first two bases of the shown sequences (pink background) correspond to bases –2 and –1 of the leader-repeat-junction. Point mutations introduced in the repeat sequence are in red. Palindromic motifs are boxed. Mutation numbers (n) refer to the name of the plasmids (pBS-Zn, see Materials and Methods and Supplementary Table S1). Samples contained 400 nM of protospacer P1 (Figure 1), 280 nM of Cas1–Cas2 and 7.5 nM of the plasmid under study. After 2 h incubation at 22°C, they were analyzed by EMSA. The gel was scanned for fluorescence of DY782. In the histogram, the intensity of the band corresponding to the ternary complex is shown as a percentage of the intensity obtained with the wild-type CRISPR sequence (pBS-Z0). Experiments were made in triplicate. Error bars represent SD. The gel under the histogram displays a representative example of the experiments. Shown is the part of the gel where ternary complexes migrate.

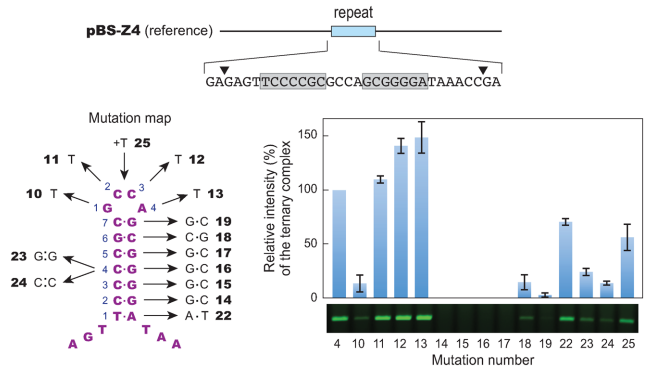

Figure 8. Binding of a Cas1–Cas2–protospacer complex to various mutant CRISPR repeats. Mutations introduced into the CRISPR repeat harbored by plasmid pBS-Z4 (Figure 7) are shown at the left side of the Figure. Mutation numbers (*n*) refer to the name of the plasmids (pBS-Zn, see Materials and Methods and Supplementary Table S1). For the sake of clarity, the palindromic region of the repeat is shown as a hairpin. Base pairs forming the stem are numbered from 1 to 7. Bases forming the loop are numbered from 1 to 4. Formation of such a hairpin on supercoiled plasmid is likely. Note, however, that Cas1–Cas2 has never been demonstrated to associate with such a structure (see text). Samples contained 400 nM of protospacer P1 (Figure 1), 280 nM of Cas1–Cas2 and 7.5 nM of the plasmid under study. After a 2 h incubation at 22°C, they were analyzed by EMSA. The gel was scanned for fluorescence of DY782. In the histogram, the intensity of the band corresponding to the ternary complex is shown as a percentage of the intensity obtained with the wild-type repeat sequence (pBS-Z4). Experiments were made in triplicate. Error bars represent SD. The gel under the histogram displays a representative example of the experiments. Shown is the part of the gel where ternary complexes migrate.

formation with pBluescript SK+ (Figure 1) or several of its derivatives (Figure 7) argued against the idea that the palindrome in the ampicillin resistance gene is enough to obtain stable Cas1–Cas2 recruitment. To probe the hypothesis that the nucleotide sequence composing the palindrome is of primary importance, we constructed pBS-Z28, in which the CRISPR repeat of pBS-Z0 was replaced by the palindromic repeat in the type I-B CRISPR system of *Haloarcula hispanica*. We also constructed pBS-Z29, where the repeat was substituted by the palindromic motif in the ampicillin resistance gene of pUC19. With both plasmids, stable ternary complex formation could not be detected (Figure 7). We concluded that the palindrome nucleotide sequence governs Cas1–Cas2 recruitment.

To map important bases of the palindromic sequence, we mutated bases in the stem and the loop of the putative hairpin (Figure 8). Mutations in pBS-Z11 to -Z13 of positions 2, 3 or 4 of the loop (these positions refer to the strand shown in Figure 8; in the complementary strand, complementary mutations are at positions 3, 2 and 1, respectively), inversion of the base pair at position 1 of the stem (pBS-Z22) or insertion of one nucleotide in the loop (pBS-Z25) did not modify or only slightly changed the intensity of the ternary complex (Figure 8). In contrast, mutation of base 1 in the loop (pBS-Z10), inversion of base pairs at positions 6 or 7 in the stem (pBS-Z18 and -Z19) or mismatches at position 4 in the stem (pBS-Z23 and -Z24) drastically reduced the intensity of the ternary complex. Finally, inversion of base pairs at positions 2, 3, 4 or 5 of the stem (pBS-Z14 to -Z17) fully abolished stable ternary complex formation. The set of results indicates importance of many bases in the palindromic motif region.

DISCUSSION

The integration of a new spacer in a CRISPR locus must be specific because outside integration would be deleterious for the cell. Here, we show that the sequence of the palindromic motif of the repeat is strongly recognized by Cas1–Cas2 when a protospacer is present. As a result, a highly stable ternary complex forms between the Cas1–Cas2–protospacer complex and a plasmid containing this motif. In this condition, it is difficult to understand why, *in vivo*, integration is restricted to the leader-proximal repeat junction and does not occur at all repeats. As mentioned in the Introduction section, two other specificity determinants, an IHF-binding site in the leader region and the bases –2, –1 and +1 at the leader-repeat boundary, were shown to improve the adaptation reaction (32,34). In fact, none of these determinants alone can account for the specificity observed *in vivo*. Many IHF-binding sites are encountered in the *E. coli* genome, and the three-base motif such as that at the leader-repeat boundary is frequent. We may imagine that all three features, IHF binding site, three-base motif and the palindromic motif, have to cooperate to designate the correct integration site *in vivo*.

One may ask why the Cas1–Cas2–protospacer complex has retained the capacity to bind non-CRISPR DNA. Here, we show that the affinity and lifetime of the complex formed between Cas1–Cas2, protospacer and non-specific DNA are much smaller than that of the complex with CRISPR DNA, as shown by the inability of non-specific complexes to resist electrophoresis conditions. Non-specific binding has been documented in the cases of several highly specific DNA-binding proteins (38). Benefit of this behavior was argued to be an acceleration of the rate of target location by the proteins with search by one-dimensional instead of three-dimensional diffusion. This may also apply in the case of Cas1–Cas2.

Binding of a protospacer-like DNA to Cas1–Cas2 enhances binding of the protein to a CRISPR-containing DNA. This synergy is likely to be related to the described large Cas1–Cas2 structural rearrangement that accompanies protospacer association (27,28). Such a structural change may be a prerequisite in the recruitment of Cas1–Cas2 by acceptor DNA.

Formation of a stable complex between Cas1–Cas2 and acceptor DNA occurs in the presence of double-stranded protospacers as well as of various partially matched oligonucleotides. As an exception, duplexes with 3′-overhangs longer than 5 bases bind Cas1–Cas2 with high affinity but do no longer form a stable complex with acceptor DNA, even in the presence of a PAM sequence (CTT) in the 3′-overhangs. The structures of the 1:1 Cas1–Cas2 : protospacer complexes suggest that five unmatched bases at the 3′-end of the protospacer are required for this 3′-end to reach the active site of Cas1 where the attack of acceptor DNA takes place (27,28). Possibly when the 3′-overhang length exceeds five nucleotides, a steric and/or electrostatic clash between protospacer and acceptor DNA prevents binding of the Cas1–Cas2–protospacer complex to acceptor DNA. Interestingly, this raises the possibility that, upon cleavage of the PAM motif by Cas1–Cas2 (28), affinity of the Cas1–Cas2–protospacer complex for acceptor DNA improves. In Figure 6, we compare ternary complex formation in the presence of protospacer P10, which harbors long 3′-overhangs, or P3, which corresponds to mature 3′-shortened P10. The stronger affinity of P3-bound Cas1–Cas2 for CRISPR DNA sustains the above hypothesis.

In our study, formation of a stable ternary complex requires a supercoiled acceptor DNA. This finding is consistent with the observation that supercoiling of acceptor DNA enhances the Cas1–Cas2-mediated covalent attack of this DNA by a protospacer *in vitro* (30). Because Cas1 is also known to recognize four-way DNA junctions (31), it was proposed that the palindromic motif of the repeat adopts a cruciform structure when it interacts with Cas1–Cas2 (30). We show here that many bases forming this putative cruciform structure are important for the interaction with protospacer-bound Cas1–Cas2. In particular, changing base pairs in the stem regions prevents formation of the stable ternary complex. On the other hand, our results do not allow to decide whether Cas1–Cas2 recognizes a linear or a cruciform palindrome DNA structure.

In summary, our study brings to light the importance of the sequence of the palindromic motif present in the *E. coli* CRISPR repeats for the specific interaction of Cas1–Cas2 with a CRISPR locus. Results also uncover a synergistic effect of protospacer in the binding of Cas1–Cas2 to this motif. Recent *in vivo* investigations using the type I-B CRISPR–Cas system of *H. hispanica* and the type I-E system of *E. coli* revealed that accurate duplication of the repeat during adaptation required bases located in and between the palindromic motifs of the repeats (39–41). Our biochemical results suggest that recognition of these bases is in fact achieved by protospacer-bound Cas1–Cas2.

Beyond understanding the specificity rules of the CRISPR adaptation phase, gaining insights into target site recognition by Cas1–Cas2 intregrase may contribute to the emergence of new genome-editing technologies. For instance, *E. coli* Cas1–Cas2 was recently used to record information into genomes, with many potential applications (42). Evolving the target specificity of Cas1–Cas2 and expressing in a same cell integrases of different specificities may help the development of such methods.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Centre National de la Recherche Scientifique (CNRS); École polytechnique. Funding for open access charge: CNRS; École polytechnique.
Conflict of interest statement. None declared.

REFERENCES

1. Barrangou,R. and Marraffini,L.A. (2014) CRISPR–Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell*, **54**, 234–244.
2. Gasiunas,G., Sinkunas,T. and Siksnys,V. (2014) Molecular mechanisms of CRISPR-mediated microbial immunity. *Cell. Mol. Life Sci.*, **71**, 449–465.

3. van der Oost,J, Westra,E.R., Jackson,R.N. and Wiedenheft,B. (2014) Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **12**, 479–492.
4. 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.
5. Mojica,F.J.M., 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. 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. 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.
8. Pul,Ü., 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.
9. 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.
10. Jansen,R., van Embden,J.D.A., Gaastra,W. and Schouls,L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, **43**, 1565–1575.
11. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J., Charpentier,E., Haft,D.H. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
12. 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.
13. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., 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.
14. Carte,J., Wang,R., Li,H., Terns,R.M. and Terns,M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.*, **22**, 3489–3496.
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. Westra,E.R., van Erp,P.B., Künne,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.
17. Deveau,H., Barrangou,R., Garneau,J.E., Labonté,J., Fremaux,C., Boyaval,P., Romero,D.A., Horvath,P. and Moineau,S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.*, **190**, 1390–1400.
18. 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.
19. Leenay,R.T., Maksimchuk,K.R., Slotkowski,R.A., Agrawal,R.N., Gomaa,A.A., Briner,A.E., Barrangou,R. and Beisel,C.L. (2016) Identifying and visualizing functional PAM diversity across CRISPR–Cas systems. *Mol. Cell*, **62**, 137–147.
20. Marraffini,L.A. and Sontheimer,E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, **463**, 568–571.
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. 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.
23. 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.
24. Amitai,G. and Sorek,R. (2016) CRISPR–Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.*, **14**, 67–76.
25. Sternberg,S.H., Richter,H., Charpentier,E. and Qimron,U. (2016) Adaptation in CRISPR–Cas Systems. *Mol. Cell*, **61**, 797–808.
26. Nuñez,J.K., Kranzusch,P.J., Noeske,J., Wright,A.V., Davies,C.W. and Doudna,J.A. (2014) Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.*, **21**, 528–534.
27. Nuñez,J.K., Harrington,L.B., Kranzusch,P.J., Engelman,A.N. and Doudna,J.A. (2015) Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, **527**, 535–538.
28. Wang,J., Li,J., Zhao,H., Sheng,G., Wang,M., Yin,M. and Wang,Y. (2015) Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR–Cas systems. *Cell*, **163**, 840–853.
29. Levy,A., Goren,M.G., Yosef,I., Auster,O., Manor,M., Amitai,G., Edgar,R., Qimron,U. and Sorek,R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, **520**, 505–510.
30. Nuñez,J.K., Lee,A.S.Y., Engelman,A. and Doudna,J.A. (2015) Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, **519**, 193–198.
31. Babu,M., Beloglazova,N., Flick,R., Graham,C., Skarina,T., Nocek,B., Gagarinova,A., Pogoutse,O., Brown,G., Binkowski,A. *et al.* (2011) A dual function of the CRISPR–Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.*, **79**, 484–502.
32. Nuñez,J.K., Bai,L., Harrington,L.B., Hinder,T.L. and Doudna,J.A. (2016) CRISPR immunological memory requires a host factor for specificity. *Mol. Cell*, **62**, 824–833.
33. Dillon,S.C. and Dorman,C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, **8**, 185–195.
34. Rollie,C., Schneider,S., Brinkmann,A.S., Bolt,E.L. and White,M.F. (2015) Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*, **4**, e08716.
35. Blattner,F.R., Plunkett,G. 3rd, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1462.
36. Benson,F.E., Stasiak,A. and West,S.C. (1994) Purification and characterization of the human Rad51 protein, an analogue of *E. coli* RecA. *EMBO J.*, **13**, 5764–5771.
37. Dardel,F. (1994) MC-Fit: using Monte-Carlo methods to get accurate confidence limits on enzyme parameters. *Comput. Appl. Biosci.*, **10**, 273–275.
38. Berg,O.G., Winter,R.B. and von Hippel,P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
39. Arslan,Z., Hermanns,V., Wurm,R., Wagner,R. and Pul,Ü. (2014) Detection and characterization of spacer integration intermediates in type I-E CRISPR–Cas system. *Nucleic Acids Res.*, **42**, 7884–7893.
40. Goren,M.G., Doron,S., Globus,R., Amitai,G., Sorek,R. and Qimron,U. (2016) Repeat size determination by two molecular rulers in the type I-E CRISPR array. *Cell Rep.*, **16**, 2811–2818.
41. Wang,R., Li,M., Gong,L., Hu,S. and Xiang,H. (2016) DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*. *Nucleic Acids Res.*, **44**, 4266–4277.
42. Shipman,S.L., Nivala,J., Macklis,J.D. and Church,G.M. (2016) Molecular recordings by directed CRISPR spacer acquisition. *Science*, **353**, 10.1126/science.aaf1175.