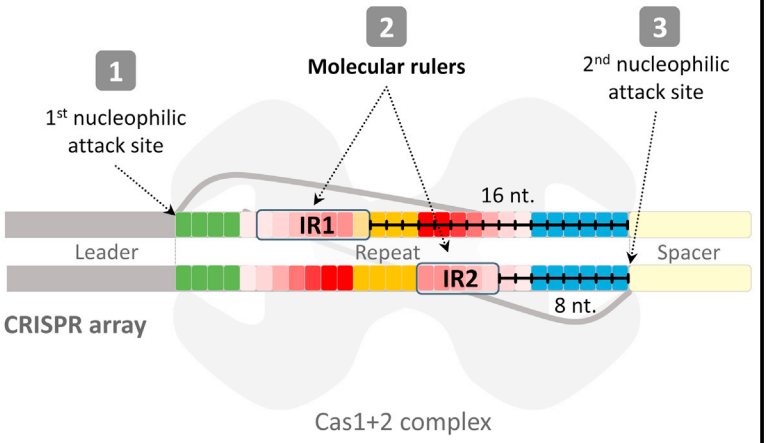
Cell Reports

Report

Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array

Graphical Abstract



Two molecular rulers determine the nucleophilic attack site at the leader-distal repeat-end

Highlights

- Inverted repeats in the type I-E CRISPR-Cas system are essential for adaptation
- Each inverted repeat encodes a motif serving as an anchor site for a molecular ruler
- These molecular rulers determine the spacer insertion site regardless of the sequence
- The findings support a model considering all known steps in spacer adaptation

Authors

Moran G. Goren, Shany Doron, Rea Globus, Gil Amitai, Rotem Sorek, Udi Qimron

Correspondence

rotem.sorek@weizmann.ac.il (R.S.), ehudq@post.tau.ac.il (U.Q.)

In Brief

Goren et al. map elements that are essential for adaptation in the *E. coli* CRISPR-Cas type I-E repeat. Two elements were identified as anchor sites for two molecular rulers that maintain a constant repeat size. Their findings support a comprehensive model for spacer adaptation.

Goren et al., 2016, Cell Reports *16*, 2811–2818
September 13, 2016 © 2016 The Authors.
http://dx.doi.org/10.1016/j.celrep.2016.08.043

CellPress

Cell Reports
Report

OPEN ACCESS
CellPress

Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array

Moran G. Goren,[1,3] Shany Doron,[2,3] Rea Globus,[1] Gil Amitai,[2] Rotem Sorek,[2,*] and Udi Qimron[1,4,*]
[1]Department of Clinical Microbiology and Immunology, Sackler School of Medicine, Tel Aviv University, 69978 Tel Aviv, Israel
[2]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel
[3]Co-first author
[4]Lead Contact
*Correspondence: rotem.sorek@weizmann.ac.il (R.S.), ehudq@post.tau.ac.il (U.Q.)
http://dx.doi.org/10.1016/j.celrep.2016.08.043

SUMMARY

Prokaryotic adaptive immune systems are composed of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins. These systems adapt to new threats by integrating short nucleic acids, termed spacers, into the CRISPR array. The functional motifs in the repeat and the mechanism by which a constant repeat size is maintained are still elusive. Here, through a series of mutations within the repeat of the CRISPR-Cas type I-E, we identify motifs that are crucial for adaptation and show that they serve as anchor sites for two molecular rulers determining the size of the new repeat. Adaptation products from various repeat mutants support a model in which two motifs in the repeat bind to two different sites in the adaptation complex that are 8 and 16 bp away from the active site. This model significantly extends our understanding of the adaptation process and broadens the scope of its applications.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins have been identified as central components of prokaryotic immune systems (Barrangou et al., 2007). These intriguing systems are found in up to 90% of archaeal genomes and in ~50% of bacterial genomes (Sorek et al., 2013) and are analogous to the mammalian immune system (Abedon, 2012; Goren et al., 2012b). Various types of CRISPR-Cas systems (Makarova et al., 2011, 2015) defend prokaryotes against viruses and horizontally transferred DNA (Barrangou et al., 2007; Brouns et al., 2008; Marraffini and Sontheimer, 2008) and RNA (Abudayyeh et al., 2016; Hale et al., 2009; Staals et al., 2014). The genetic loci of all systems include a CRISPR array—short repeated sequences, called "repeats," that flank similarly sized sequences, called "spacers." The spacers are acquired from DNA sequences termed "protospacers." Their incorporation into the bacterial CRISPR array, termed "adaptation," enhances the spacer repertoire of the array against foreign elements. The CRISPR array is usually preceded by a "leader" DNA sequence that is located near a cluster of *cas* genes (Deveau et al., 2010; Marraffini and Sontheimer, 2010; Sorek et al., 2008). RNA transcribed from the CRISPR array (crRNA) is processed by Cas proteins into RNA-based spacers flanked by partial repeats. These RNA spacers specifically direct Cas interference proteins to target and cleave nucleic acids encoding matching protospacers. Thus, the system can adaptively and specifically target invaders.

The adaptation process has been thoroughly characterized for the type I-E CRISPR-Cas system in the model organism *Escherichia coli* (Sternberg et al., 2016). In vivo, two proteins, Cas1 and Cas2, are both necessary and sufficient for acquiring new spacers in this system (Yosef et al., 2012). Expression of these two proteins from a plasmid results in significant spacer adaptation into CRISPR arrays. Adaptation requiring only Cas1 and Cas2 proteins is termed "naive," as opposed to "primed" adaptation, which requires additional Cas proteins guided by a targeting spacer (Datsenko et al., 2012). These in vivo findings have been supported in vitro by naive adaptation experiments comprising Cas1, Cas2, a CRISPR array, and a donor spacer (Nuñez et al., 2015b; Rollie et al., 2015).

A single repeat of 28 bp is both necessary and sufficient for adaptation (Goren et al., 2012a; Yosef et al., 2012). The repeat encodes a heptameric palindrome composed of two inverted repeats (IRs) interspaced by 4 nt that can form a stem-loop structure in a single-strand nucleic acid. This secondary structure in the mature crRNA is thought to serve as a "molecular handle" for the interference proteins (Gesner et al., 2011; Sashital et al., 2011). Bioinformatics analysis has shown that the IRs are the most conserved sequences in the type I-E repeat, whereas the sequence connecting the IRs is the least conserved (Kunin et al., 2007). Although the IRs have been shown to be required for the adaptation step (Arslan et al., 2014), a thorough characterization of the entire repeat element has not been reported for the type I-E system.

A recent report did thoroughly characterize the repeat element in *Haloarcula hispanica* (Wang et al., 2016). However, that study focused solely on primed adaptation, because there is no system for studying naive adaptation in that archaeon (Li et al., 2014). The study showed that certain substitutions in the leader-proximal end of the repeat significantly reduce adaptation efficiency. The leader-proximal IR (IR1) was important for

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

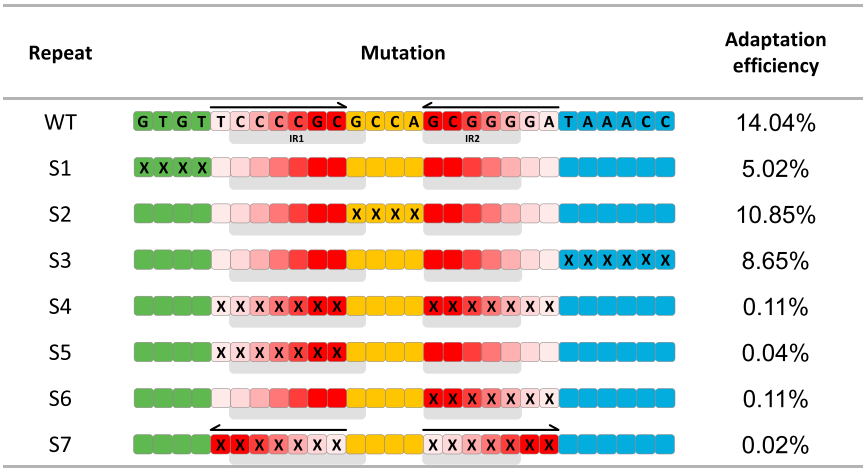


Figure 1. Determination of Essential Elements in the Repeat

Adaptation assays were carried out for the different repeat variants as described in Experimental Procedures. Each box represents a nucleotide of the indicated repeat as follows: green, leader-proximal region; red, IRs; orange, region between IRs; blue, leader-distal region; X, substitution mutation. Percentage of adaptation efficiency was determined by analyzing high-throughput DNA sequencing products as described in Experimental Procedures.

adaptation, as mutating it reduced adaptation significantly. Interestingly, the leader-distal IR (IR2) could be mutated without detrimental effect on adaptation. A second motif between these two IRs was found to be important for adaptation, as mutating it reduced adaptation significantly. This motif was suggested to serve as an anchor site for a molecular ruler that measures a specific distance from which the spacer is inserted. This putative molecular ruler inserted the spacer 8 nt downstream of the end of this motif regardless of the sequence of the downstream nucleotides. Overall, the study showed that in a primed type I-B adaptation system, the adaptation machinery probably recognizes two sites (Wang et al., 2016). One site, at the leader-repeat junction, serves as a docking site for the protein complex, and the other probably serves as an anchor for a molecular ruler that measures a specific length, regardless of the downstream sequence.

In the present study, we searched for motifs in the repeat that are essential for spacer adaptation in the *E. coli* type I-E system and possibly determine the fidelity of the process and the maintenance of a constant repeat size. We found that the IRs, as well as their orientation, are essential for efficient adaptation, whereas other elements are not. Most significantly, we found that motifs in these IRs are anchor sites from which a constant distance is measured to initiate the leader-distal nucleophilic attack. The differences and similarities between type I-E and I-B systems are discussed and highlight a mechanism for "quality control" of size determination in the adaptation process that ensures a constant repeat size is maintained.

RESULTS

Experimental Setup

To identify motifs in the repeat affecting the efficiency and fidelity of spacer acquisition, we used a plasmid-based adaptation assay. The plasmid encoded a leader-repeat sequence as well as Cas1 and Cas2 expressed from an inducible promoter (Figure S1). The plasmid was transformed into *E. coli* BL21-AI, lacking the interference *cas* genes (Brouns et al., 2008) and deleted for one of its endogenous CRISPR arrays. Following Cas1 and Cas2 expression, DNA from a sample of the culture was used as a template for PCR amplification of the adapted region. To determine the extent of adaptation and the size and sequence of the newly inserted repeats, the obtained products were analyzed by high-throughput DNA sequencing, as elaborated in Experimental Procedures. This system thus allowed us to efficiently monitor adaptation under naive conditions.

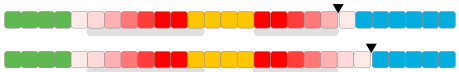
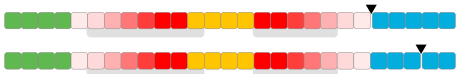
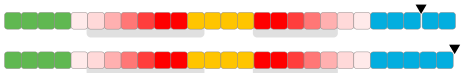
Determining Motifs Required for Adaptation

Analysis of spacer acquisition into a plasmid encoding a wild-type (WT) repeat indicated that ~14% of the templates contain new spacer-repeat insertions (Figure 1, WT). We tested repeats having substitution mutations in the leader-proximal end, the region between the IRs, the leader-distal end, and the IRs. Spacer acquisition in all mutants outside the IRs was only up to 3-fold reduced compared to spacer acquisition into the WT repeat (Figure 1, repeats S1–S3). Conversely, mutations in the IRs reduced the adaptation efficiency ~100-fold compared to the WT repeat (Figure 1, repeat S4). This reduction was also detected when each IR was individually mutated (Figure 1, repeats S5 and S6). Moreover, maintaining the IR sequences but reversing their orientation also significantly reduced adaptation efficiency (Figure 1, repeat S7). These results demonstrated that the IR sequences, as well as their orientation, are major determinants of adaptation, whereas other regions of the repeat are less important for adaptation efficiency.

Identifying an Anchor Motif for a Molecular Ruler

We speculated that at least one of these IRs serves as a docking site for the Cas1–Cas2 integrase, as shown by Xiang and colleagues for the type I-B system (Wang et al., 2016). We therefore generated single-nucleotide deletions across the repeat sequence as shown in Figure 2A. We expected that a deletion upstream of such a docking site would simply be duplicated, resulting in a 27-bp repeat, which is shorter than the WT repeat due to the deleted nucleotide. On the other hand, deletion downstream of the docking site would result in a regular-sized 28-bp repeat, as the molecular ruler would measure a defined distance downstream to this site, regardless of the deletion. Therefore, in these cases, a single nucleotide from the sequence immediately downstream of the repeat would be added to the repeat. Indeed, deletions from the leader-proximal end of the repeat up to the leader-distal IR (IR2) resulted, in over 97% of the cases, in a duplicated

A

Repeat	Mutation	Adaptation efficiency	New repeat length compared to original repeat		
			Identical	1 bp longer	Other
WT		12.37%	98.56%	1.31%	0.12%
D1		2.00%	98.10%	1.75%	0.15%
D2		1.52%	98.02%	1.87%	0.11%
D3		1.61%	97.21%	2.60%	0.19%
D4		3.43%	98.42%	1.44%	0.13%
D5		7.66%	98.43%	1.31%	0.26%
D6		4.70%	98.96%	0.87%	0.17%
D7		4.54%	0.91%	98.50%	0.59%
D8		14.07%	0.26%	99.72%	0.02%
D9		17.12%	0.26%	99.72%	0.02%
D10		15.24%	0.27%	99.72%	0.01%

B

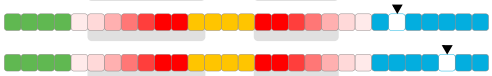
Repeat	Mutation	Adaptation efficiency	New repeat length compared to original repeat		
			Identical	1 bp shorter	Other
WT		12.37%	98.56%	0.12%	1.31%
I1		2.32%	76.73%	1.08%	22.19%
I2		2.08%	78.47%	1.02%	20.51%
I3		4.37%	0.79%	99.10%	0.10%
I4		5.55%	0.64%	98.69%	0.68%
I5		2.41%	2.20%	96.87%	0.94%
I6		11.33%	0.42%	99.52%	0.06%
I7		13.52%	0.42%	99.52%	0.06%
I8		16.72%	0.47%	99.49%	0.04%

Figure 2. Determination of Anchor Sites for a Molecular Ruler

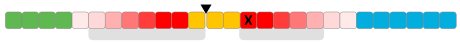
Experiment and representation as described for Figure 1 except that arrowheads represent location of a single nucleotide deletion (A) or insertion (B). Percentage of new repeat length was determined as described in Experimental Procedures.

27-bp repeat (Figure 2A, repeats D1–D6). Deletions of nucleotides downstream of a motif in the IR2 resulted almost exclusively (>98%) in a repeat that was extended by a single G nucleotide, the nucleotide found immediately adjacent to the repeat (Figure 2A, repeats D7–D10). These results indicated that the "GCGGG" motif (or some part thereof) in IR2 serves as an anchor site for a molecular ruler. The spacer is inserted 8 nucleotides downstream of the end of this motif, probably as a result of a molecular ruler that dictates nucleophilic attack of the spacer 8 nt downstream of the end of this motif (see also Figure 4 for graphical illustrations of these results).

Identifying an Additional Anchor Motif for a Molecular Ruler

We speculated that single-nucleotide insertions upstream of the newly identified docking site would result in duplicated 29-bp repeats (longer than the 28-bp WT repeats due to the nucleotide insertion), whereas nucleotide insertions downstream of this docking site would result in a regular-sized repeat due to the molecular ruler that measures 8 nucleotides downstream of the motif. Indeed, insertion downstream of the motif resulted in a regular-sized repeat (Figure 2B, repeats I6–I8). Surprisingly however, several nucleotide insertions upstream of this motif also resulted in regular-sized repeats (Figure 2B, repeats I3–I5). These insertions were located between IR2 and the leader-proximal IR1. Insertions upstream of IR1 resulted in mostly (>76%) duplicated 29 bp repeats (Figure 2B, repeats I1 and I2). These results suggested that IR1 also encodes a "CCCCGCG" motif (or some part thereof) that serves as a docking site for another molecular ruler. The distance measured from the end of this motif to the spacer-insertion site was 16 nucleotides. Apparently, the measurement of 16 nt from this motif was masked in the repeat deletion mutants shown in Figure 2A. Thus, the "ruler measurement" of IR1 is masked by the ruler activity of IR2 in cases where the repeat is being lengthened to the regular length but is revealed when the repeat is being shortened to the regular length. The

A

Repeat	Mutation	Adaptation efficiency	New repeat length compared to original repeat		
			Identical	1 bp longer	Other
D5		7.66%	98.43%	1.31%	0.26%
D5S1		4.94%	97.99%	1.64%	0.37%
D5S2		2.67%	31.13%	45.76%	23.10%
D5S3		5.10%	95.22%	4.51%	0.26%
D5S4		2.11%	97.73%	2.16%	0.11%
D5S5		3.28%	92.68%	6.69%	0.63%
D5S6		1.20%	3.53%	91.02%	5.45%

B

Repeat	Mutation	Adaptation efficiency	New repeat length compared to original repeat		
			Identical	1 bp shorter	Other
I4		5.55%	0.64%	98.69%	0.68%
I4S1		0.10%	6.89%	91.78%	1.33%

Figure 3. Validation of Anchor Sites
Experiment and representation as described for Figure 2.

dominant ruler is the one that measures the shorter distance in each case. These results revealed an additional anchor site for a molecular ruler.

Validation of Both Anchor Sites

We hypothesized that a nucleotide deletion between IR1 and IR2 would be processed to a regular-sized repeat if the IR2 molecular ruler is disabled. In this case, IR1 would be the only anchor site for a molecular ruler, and would be unmasked by the molecular ruler of IR2. We therefore constructed a series of substitution mutants in the IR2 motif in repeats having a deletion between IR1 and IR2, and monitored repeat length in the products. As speculated, we observed that some of the mutations disrupted the suspected motif in IR2 and have led to repeats lengthened by 1 bp. Specifically, a single A substitution in the second base of IR2 led to elevation of the percentage of lengthened repeats from 1.31% in the parental repeat to 45.76% in the mutant (Figure 3A, D5S2), suggesting that this base is central to the anchor motif in IR2. In accordance, a mutation that included the same A substitution, in addition to two A substitutions in its flanking bases, resulted in >91% of the new repeats being longer (Figure 3A, D5S6), further suggesting that these mutations disrupted the IR2 anchor motif. These results clearly indicated that, indeed, IR1 is an anchor for a molecular ruler that is masked by the presence of the IR2 molecular ruler.

We further speculated that mutating the IR1 motif in repeats having an insertion between IR1 and IR2 would eliminate the shortening of the repeat to a regular-sized one. This would result because IR1 is the only anchor site for a molecular ruler that is found upstream of this insertion, whereas the downstream IR2 does not "measure" such an insertion. We therefore constructed a mutant having an insertion between IR1 and IR2, in addition to a 3-A substitution in IR1. This mutant significantly increased the proportion of repeats that remained long, as predicted. The percentage of long repeats increased >10-fold from 0.64% in the parental repeat (Figure 3B, repeat I4) to 6.89% in these settings (Figure 3B, repeat I4S1). We hypothesize that the introduced substitutions did not entirely disable the activity of the IR1 ruler, and thus repeat shortening remained relatively high. The importance of this anchor site in adaptation was reflected in the low acquisition detected from this 3-bp-substitution mutant. This result further indicated that motifs in both IR1 and IR2 function as anchors for molecular rulers. Taken together, these results demonstrate the presence of two independent anchor sites for two molecular rulers measuring distinct distances.

DISCUSSION

We studied the requirement for each element in the repeat sequence for efficiency and fidelity of adaptation. We found that the only essential elements in the repeats are the IRs; the other elements could be individually mutated. The most important finding of this study was that both of the IRs encode motifs that serve as anchors for two distinct molecular rulers that determine the distance of the nucleophilic attack on the leader-distal end. These motifs maintain a constant-sized repeat by lengthening or shortening an irregular-sized repeat to the regular size.

Xiang and colleagues characterized the type I-B repeats by substitution, deletion, and insertion mutations in the repeat's elements followed by Sanger sequencing of several products.

Interestingly, they found that the leader-proximal end of the repeat is essential for adaptation. Certain substitutions in this region completely abrogated adaptation, whereas others only mildly impaired it (Wang et al., 2016). IR1 was essential for adaptation and possibly served as a docking site for the protein complex, whereas IR2 was not. In addition, mutating the region between the two IRs resulted in significantly reduced adaptation. Thus, a major difference between type I-B and type I-E is that in the latter, the nucleotides between the IRs are dispensable, whereas both of the IRs are essential. There are also differences in the proteins and spacers required for proper adaptation: currently, type I-B can only be studied in a primed state (Li et al., 2014), whereas type I-E can be studied in both states (Datsenko et al., 2012; Swarts et al., 2012; Yosef et al., 2012). Thus, type I-B strictly requires the presence of the interference proteins and a targeting spacer for adaptation of a new spacer (most likely for generation of spacers, but not for the spacer integration step). In addition, the type I-B adaptation complex probably requires the non-interference protein Cas4 for adaptation, whereas type I-E does not (Li et al., 2014). These differences indicate that systems of the same type exhibit different mechanisms for spacer adaptation.

The most significant difference is that type I-E has two anchor sites for molecular rulers within the IRs, whereas in type I-B, only a single putative site was found between the two IRs (Wang et al., 2016). In all cases in the latter study, insertions and deletions upstream of this type I-B motif resulted in the expected repeat duplication. In addition, in most cases, deletions and insertions downstream of it resulted in lengthened or shortened repeats, respectively, as expected (Wang et al., 2016).

In this work, we found two anchor sites for molecular rulers, each measuring a different distance and eventually coordinating a nucleophilic attack at the repeat-spacer junction in the WT repeat. The existence of two docking sites for molecular rulers may serve as "quality control" for size determination and to maintain a constant repeat size. The first nucleophilic attack is dictated by the position of the leader-repeat junction (Nuñez et al., 2015b; Rollie et al., 2015), and therefore, the rulers determine the distance to the second nucleophilic attack, at the leader-spacer junction. In the WT repeat, both rulers deliver the repeat to the same site, where nucleophilic attack of the spacer takes place (Figure 4A). In repeats encoding a deletion or insertion upstream of both IRs, the two molecular rulers both "miss" the length correction but both deliver the repeat at the same site for spacer nucleophilic attack (Figure 4B). In repeats encoding a deletion or insertion downstream of both IRs, both molecular rulers correct the size by either extending the repeat length in the case of a deletion or shortening it in the case of an insertion (Figure 4C). In repeats encoding a deletion or insertion between the two molecular rulers, the molecular ruler that shortens the repeat dominates (Figure 4D). We speculate that this is because the repeat cannot be delivered for nucleophilic attack at a distance of more than 8 or 16 nt from the anchor site due to tight docking. It is only when one of these docking sites is mutated that the other molecular ruler can take over (Figure 4E). We suppose that in this case, the mutated motif does not dock the repeat, and therefore, the other site is allowed to deliver the repeat for nucleophilic attack at its programmed distance. Mutating the IR2 allows total domination of IR1, whereas mutating IR1 allows increased domination of IR2 rather than total domination. One may speculate that IR1 and the leader-repeat docking site are on the same protein subunit in the complex and thus less flexibility is allowed in IR1 substitution, whereas IR2 is on a different protein subunit in the complex, thus allowing more flexibility in its substitution. Another possibility is that the IR1 anchor helps define the first integration site at the leader-repeat junction, and therefore, its absence is more pronounced on adaptation.

The integration complex has intrinsic symmetry in its structure (Nuñez et al., 2015a; Wang et al., 2015). Despite this symmetry, which suggests that the mechanism of integration would be symmetrical as well, directional adaptation is observed with regard to the protospacer adjacent motif (PAM) sequence (Nuñez et al., 2015a; Wang et al., 2015; Yosef et al., 2012). The asymmetrical mechanism is also reflected in the two molecular rulers measuring two distinct distances in the same direction, rather than similar distance in opposite directions. This directionality is probably dictated by the docking site that is recognized at the leader-repeat junction, dictating that the measurement would be to the other end lacking a distinct docking site. Thus, the asymmetry of the docking sites flanking the repeat probably dictate the asymmetrical mechanism operating in the integration machinery.

What is the mechanism maintaining constant repeat size in other CRISPR-Cas systems? As shown for type I-B, a ruler that is not located in IR1 or IR2, but rather between these two motifs serves as a molecular ruler. Other CRISPR-Cas systems, such as the type II-A in *Streptococcus pyogenes*, have their IRs in the extreme ends of the repeat. We speculate that in that case, the IRs would serve as direct attachment sites to the integration complex and consequently would be directly involved in the nucleophilic attack. This mechanism would preserve the repeat size without molecular rulers. In cases where there are no IRs in the repeat, we speculate that the leader-repeat junction may serve as one docking site, and another site will serve as a docking site for a molecular ruler, as is the case in the type I-B system. Altogether, we believe that various types of CRISPR-Cas evolved distinct mechanisms but with similar principles to maintain the periodicity of the array.

Our observations of adaptation into mutated repeats are explained by the function of these two motifs, serving as anchor sites, and substantiate the model shown in Figure 4F. This model is based on recent in vitro work and structural studies (Nuñez et al., 2015a, 2015b; Rollie et al., 2015; Wang et al., 2015). Those studies revealed that the adaptation proteins Cas1 and Cas2 bind the 5′ and 3′ ends of the newly inserted spacer during the integration process, enabling nucleophilic attack at the integration sites. The structure of these proteins with the repeat and spacer DNAs has not yet been published. Thus, our study is important for elucidating one of the remaining unresolved stages of the adaptation process: the delivery of the repeat to the spacer for nucleophilic attack. A crystal structure containing the repeat and spacer should reveal the exact contact residues of the proteins with the repeats. Rational modification of these residues may produce proteins generating repeats of various lengths, which in turn will extend the repertoire of adaptation products

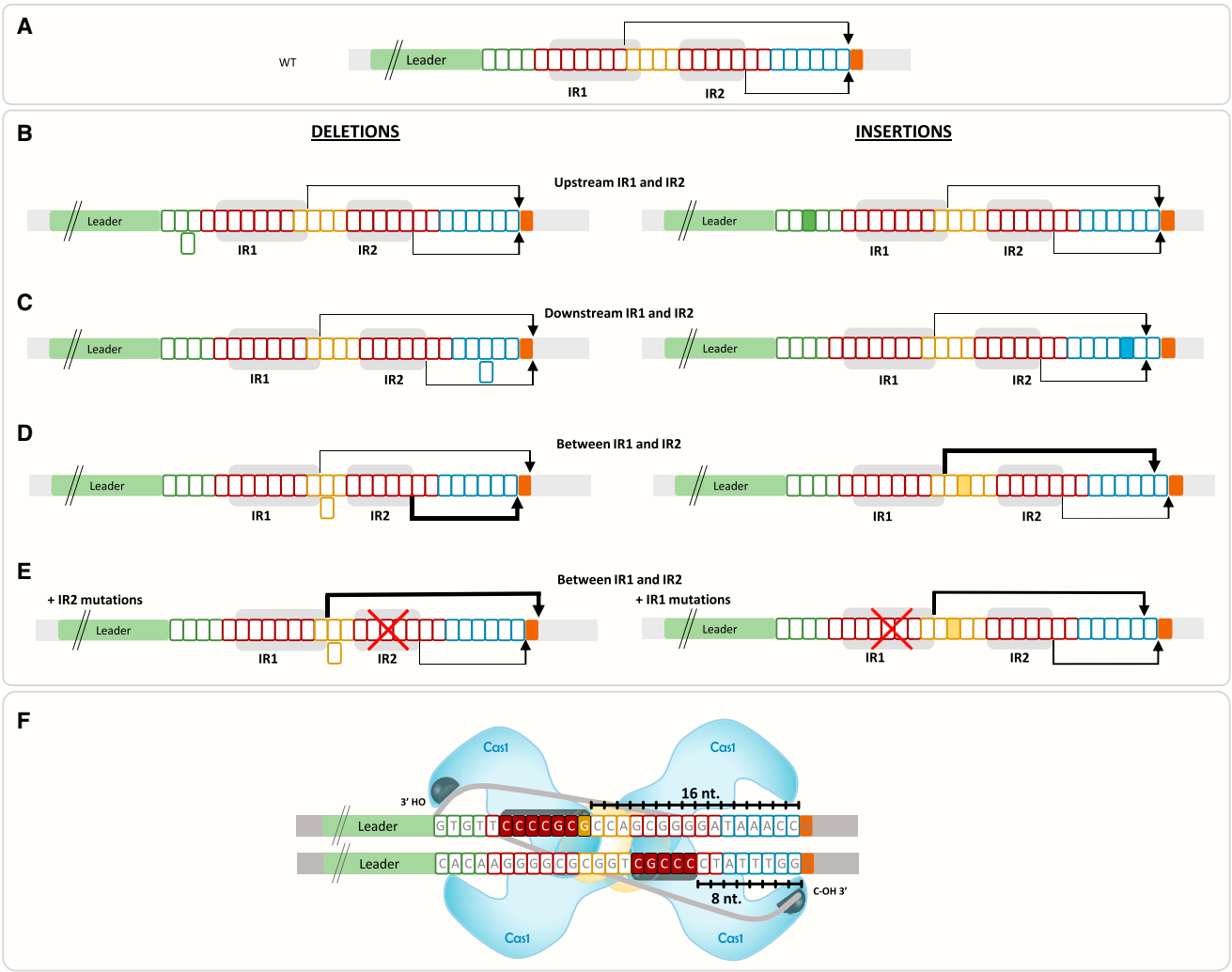


Figure 4. Schematic Summary of the Results Leading to a Proposed Model

(A–C) Schematics of the repeats and spacer-insertion sites. Arrows point to 16 and 8 bp from the end of the IR1 and IR2 motifs, respectively, where most spacer insertions were observed.

(D and E) Same as (A)–(C). Thickness of arrows correlates schematically with spacer insertions at the indicated site (not to scale).

(F) A model depicting the two anchor sites (fully colored boxes) for two molecular rulers (black-marked lines). In accordance with the structure, four Cas1 (light blue) and two Cas2 (yellow) form a heteromeric complex binding a protospacer substrate (gray strands). We envision that the complex also binds the repeat (boxed letters) at the two identified anchor sites. The ends of these sites are positioned 8 and 16 bp away from the active site (dark gray half circle) of nucleophilic attack on the repeat.

used for different applications (e.g., Shipman et al., 2016). Our results also contribute to such applications as they map the elements that are permutable and can thus significantly extend the repertoire of barcoded functional repeats.

EXPERIMENTAL PROCEDURES

Reagents, Strains, Plasmids, and Plasmid Constructions

The above are described in Supplemental Experimental Procedures.

Adaptation Assay

A single colony from each IYB5283 strain harboring the different pCas1+2R mutant plasmids was inoculated in lysogeny-broth (LB) medium containing 50 μg/ml streptomycin and aerated at 37°C for 16 hr. Each of the overnight cultures was then diluted 1:300 in LB medium containing 50 μg/ml streptomycin with 0.2% (w/v) L-arabinose + 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown for an additional 10–16 hr at 37°C. This procedure was repeated twice more. A sample from each culture was used as the template in PCR1 (see Supplemental Experimental Procedures).

PCR Products for Deep Sequencing

PCR1 and PCR2 production are described in Supplemental Experimental Procedures.

Determination of Adaptation Efficiency and New Repeat Length

Illumina sequencing libraries were prepared using the PCR1 and PCR2 products. The libraries were sequenced using the Illumina Miseq or NextSeq500 platforms generating 150-bp reads. Sequenced reads were demultiplexed and mapped to the *E. coli* "BL21-Gold(DE3)pLysS AG" genome (NC_012947.1)

and pCas1+2R plasmid using blastn (with parameters: -e 1e-10 -F F). Adaptation efficiency was determined as previously described (Levy et al., 2015), except that new acquisition events were inferred if the read alignment spanned the old repeat and the sequence downstream of it but did not include the leader upstream, meaning that a new sequence had been inserted between the old repeat and the leader. New repeat length was determined using the PCR2 library reads. Reads were identified as representing an acquisition event if they contained two alignments to the repeat sequence, with a sequence in between that maps elsewhere in the genome or the plasmid (the potential spacer). The repeat length was initially determined according to the alignment. Spacers recorded as 34 or 35 nt in length, with the first nucleotide being a G that was not aligned to the genome, were considered as 33- or 34-nt spacers, respectively, derived from a repeat that was 1 bp longer. Spacers that were 32 nt in length with an upstream C in the genome were considered to be 33-nt spacers (with C as their first nucleotide) with a repeat that was shorter by 1 bp. The percentage of new repeat length presented in Figures 2 and 3 is based on acquisitions of 33-nt spacers.

ACCESSION NUMBERS

The accession number for the deep-sequencing data reported in this paper is ENA: PRJEB15054.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, one figure, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.08.043.

AUTHOR CONTRIBUTIONS

M.G.G., S.D., R.G., G.A., R.S., and U.Q. conceived and designed the experiments and analyzed the data. M.G.G., R.G., and G.A. performed experiments. U.Q. wrote the manuscript with feedback from all authors. R.S. and U.Q. secured funding.

ACKNOWLEDGMENTS

We thank Camille Vainstein for professional language editing and Oren Auster for providing the pCas1+2R plasmid. The study was supported by the European Research Council StG and CoG programs (grant 336079 to U.Q. and grant 681203 to R.S.), the Israel Science Foundation (grant 268/14 to U.Q., grant 1303/12 to R.S., and I-CORE grant 1796 to R.S.) and the Israeli Ministry of Health (grant 9988-3 to U.Q.).

Received: June 30, 2016
Revised: August 1, 2016
Accepted: August 12, 2016
Published: September 13, 2016

REFERENCES

Abedon, S.T. (2012). Bacterial 'immunity' against bacteriophages. Bacteriophage *2*, 50–54.

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

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.

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.

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.

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.

Deveau, H., Garneau, J.E., and Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. Annu. Rev. Microbiol. *64*, 475–493.

Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M., and Macmillan, A.M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nat. Struct. Mol. Biol. *18*, 688–692.

Goren, M.G., Yosef, I., Auster, O., and Qimron, U. (2012a). Experimental definition of a clustered regularly interspaced short palindromic duplicon in Escherichia coli. J. Mol. Biol. *423*, 14–16.

Goren, M., Yosef, I., Edgar, R., and Qimron, U. (2012b). The bacterial CRISPR/Cas system as analog of the mammalian adaptive immune system. RNA Biol. *9*, 549–554.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945–956.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. *8*, R61.

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.

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.

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.

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.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843–1845.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. *11*, 181–190.

Nuñez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015a). Foreign DNA capture during CRISPR-Cas adaptive immunity. Nature *527*, 535–538.

Nuñez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193–198.

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.

Sashital, D.G., Jinek, M., and Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat. Struct. Mol. Biol. *18*, 680–687.

Shipman, S.L., Nivala, J., Macklis, J.D., and Church, G.M. (2016). Molecular recordings by directed CRISPR spacer acquisition. Science *353*, aaf1175.

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.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu. Rev. Biochem. *82*, 237–266.

Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., et al. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. Mol. Cell *56*, 518–530.

Sternberg, S.H., Richter, H., Charpentier, E., and Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. Mol. Cell *61*, 797–808.

Swarts, D.C., Mosterd, C., van Passel, M.W.J., and Brouns, S.J.J. (2012). CRISPR interference directs strand specific spacer acquisition. PLoS ONE *7*, e35888.

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.

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.

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.

Cell Reports, Volume *16*

Supplemental Information

Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array

Moran G. Goren, Shany Doron, Rea Globus, Gil Amitai, Rotem Sorek, and Udi Qimron

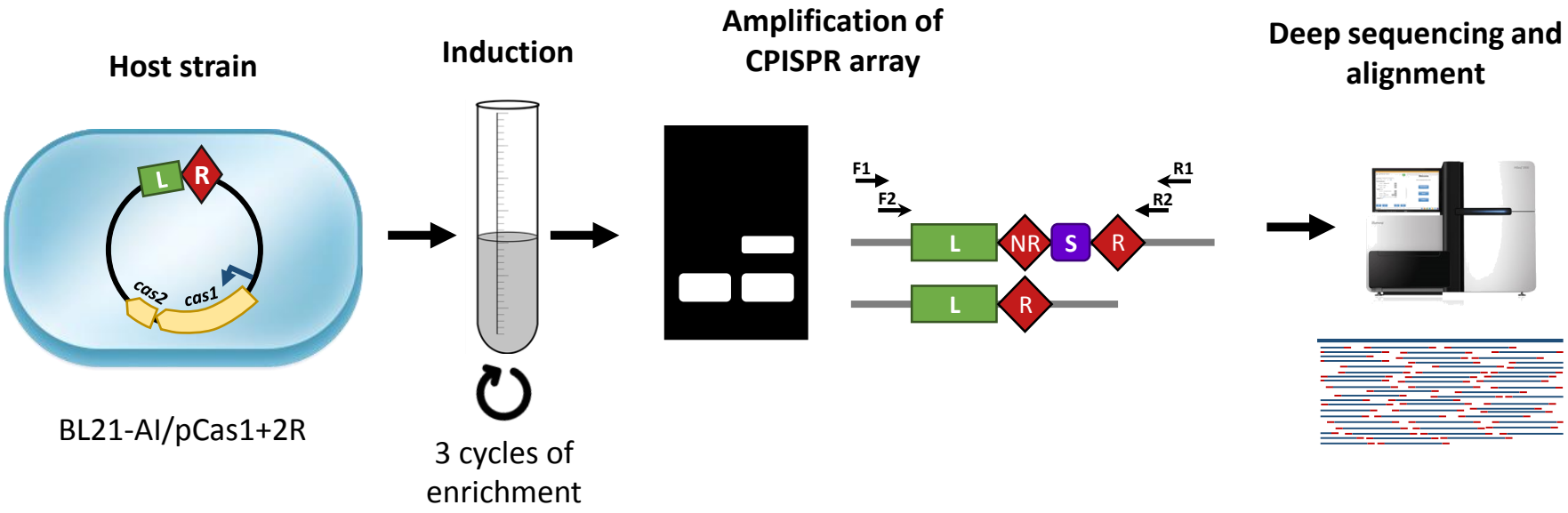


Figure S1; Related to Figures 1, 2, and 3. Assay for monitoring the effects of repeat modification on adaptation. A plasmid encoding Cas1–Cas2 and a leader–repeat sequence, occasionally with modifications to the WT sequence, was transformed into *E. coli* BL21-AI. Bacteria were induced to express Cas1–Cas2, grown, diluted, and grown again, three times in total. A sample of the culture was then taken for PCR1 using primers F1–R1. PCR2 was generated following gel extraction of the adapted band using primers F2–R2. Both products were analyzed using high-throughput DNA sequencing.

Table S1. Bacterial strains, plasmids and oligonucleotides used in this study. Related to Experimental Procedures.

Bacterial strains	**Description/sequence**	**Source or reference**
NEB5α	F$^-$ ϕ80*lacZ*ΔM15Δ(*lacZYA-argF*) U169 *deoR recA1 endA1 hsdR17* (r_K^-, m_K^+) *gal*$^-$ *phoA supE44* λ$^-$ *thi*$^-$*1 gyrA96 relA1*	New England Biolabs
IYB5283	BL21-AI with no repeats in CRISPR I, kan^r, tet^r	(Yosef et al., 2012)[a]
Plasmids		
pCas1+2	pCDF-1b (Novagen) cloned with *cas1,2* under T7 promoter, str^r	(Yosef et al., 2012)[a]
pCas1+2R (WT)	pCDF-1b (Novagen) cloned with *cas1,2* under T7 promoter, minimal leader and single repeat of CRISPR 1 array, str^r	This study
The following plasmids are Identical to pCas1+2R except for the repeat sequence specified below:		
S1	TGTGTCCCCGCGCCAGCGGGGATAAACC	This study
S2	GTGTTCCCCGCTAACGCGGGGATAAACC	This study
S3	GTGTTCCCCGCGCCAGCGGGGAGCCCAA	This study
S4	GTGTGAAAATAGCCATATTTTCTAAACC	This study
S5	GTGTGAAAATAGCCAGCGGGGATAAACC	This study
S6	GTGTTCCCCGCGCCATATTTTCTAAACC	This study
S7	GTGTAGGGGCGGCCACGCCCCTTAAACC	This study
D1	G_GTTCCCCGCGCCAGCGGGGATAAACC	This study
D2	GTGT_CCCCGCGCCAGCGGGGATAAACC	This study
D3	GTGTTCC_CGCGCCAGCGGGGATAAACC	This study
D4	GTGTTCCCCGC_CCAGCGGGGATAAACC	This study
D5	GTGTTCCCCGCG_CAGCGGGGATAAACC	This study
D6	GTGTTCCCCGCGCC_GCGGGGATAAACC	This study
D7	GTGTTCCCCGCGCCAGCGGG_ATAAACC	This study
D8	GTGTTCCCCGCGCCAGCGGGGA_AAACC	This study
D9	GTGTTCCCCGCGCCAGCGGGGATAA_CC	This study
D10	GTGTTCCCCGCGCCAGCGGGGATAAAC_	This study
I1	GTGTTTCCCCGCGCCAGCGGGGATAAACC	This study
I2	GTGTTCCCCCGCGCCAGCGGGGATAAACC	This study
I3	GTGTTCCCCGCGGCCAGCGGGGATAAACC	This study
I4	GTGTTCCCCGCGCCCAGCGGGGATAAACC	This study
I5	GTGTTCCCCGCGCCAAGCGGGGATAAACC	This study
I6	GTGTTCCCCGCGCCAGCGGGGGATAAACC	This study
I7	GTGTTCCCCGCGCCAGCGGGGATTAAACC	This study
I8	GTGTTCCCCGCGCCAGCGGGGATAAAACC	This study
D5S1	GTGTTCCCCGCGCAACGGGGATAAACC	This study
D5S2	GTGTTCCCCGCGCAGAGGGGATAAACC	This study
D5S3	GTGTTCCCCGCGCAGCAGGGATAAACC	This study
D5S4	GTGTTCCCCGCGCAGCGAGGATAAACC	This study
D5S5	GTGTTCCCCGCGCAGCGGAGATAAACC	This study
D5S6	GTGTTCCCCGCGCAGAAAGGATAAACC	This study
I4S1	GTGTTCCAAACGCCCAGCGGGGATAAACC	This study
Oligonucleotides	**5'→3'**	
OA2F	CCTTTGATCTTTTCTACTGA	
OA2R	ATGGGGCTGACTTCAGGTGC	
RE10RD	NNNNTGGATGTGTTGTTTGTG	
IY230R1	NNNNAAATGAGCGATGATATTTGTGCT	
MG132F	GTTATGTTTAGATGTGTCCCCGCGCCAGCGG	
MG132R	CCGCTGGCGCGGGGACACATCTAAACATAAC	
MG82F	GCGGGGATAAACCGAGCACA	
MG86R	GTTAGCGGGGAACACTCTAAACATAACCTATTAT	
MG126F	GCCCAAGAGCACAAATATCATCGCTC	
MG126R	TCCCCGCTGGCGCGGGGAACACTC	
MG85F	TATTTTCTAAACCGAGCACAAATATCA	
MG85R	TGGCTATTTTCACACTCTAAACATAACCTATTAT	
MG197F	GAAAATAGCCAGCGGGGATAAACCGAG	
MG197R	ACACTCTAAACATAACCTAT	
MG198F	TATTTTCTAAACCGAGCACAAATATCA	

MG198R	TGGCGCGGGGAACACTCTAA
MG88F	CGCCCCTTAAACCGAGCACAAA
MG196R	TGGCCGCCCCTACACTCTAAACAT
MG199F	GTTCCCCGCGCCAGCGGGGA
MG199R	CTCTAAACATAACCTATTAT
MG200F	TCCCCGCGCCAGCGGGGATA
MG200R	CACTCTAAACATAACCTATT
MG218F	TGTTTAGAGTGTTCCCGCGCCAGCGGGGAT
MG218R	ATCCCCGCTGGCGCGGGAACACTCTAAACA
MG202F	CCAGCGGGGATAAACCGAGC
MG202R	GCGGGGAACACTCTAAACAT
MG203F	CAGCGGGGATAAACCGAGCAC
MG203R	CGCGGGGAACACTCTAAAC
MG204F	GCGGGGATAAACCGAGCACA
MG204R	GGCGCGGGGAACACTCTAAA
MG205F	GGATAAACCGAGCACAAATA
MG215R	CCCGCTGGCGCGGGGAACACT
MG206F	AAACCGAGCACAAATATCAT
MG216R	AATCCCCGCTGGCGCGGGGAA
MG219F	GCGCCAGCGGGGATAACCGAGCACAAATAT
MG219R	ATATTTGTGCTCGGTTATCCCCGCTGGCGC
MG208F	CGAGCACAAATATCATCGCT
MG208R	TTTATCCCCGCTGGCGCGGG
MG200F	TCCCCGCGCCAGCGGGGATA
MG210R	AACACTCTAAACATAACCTAT
MG201F	CGCGCCAGCGGGGATAAACC
MG211R	GGGGAACACTCTAAACATAAC
MG202F	CCAGCGGGGATAAACCGAGC
MG212R	CCGCGGGGAACACTCTAAACA
MG203F	CAGCGGGGATAAACCGAGCAC
MG213R	GGCGCGGGGAACACTCTAAAC
MG204F	GCGGGGATAAACCGAGCACA
MG214R	TTGGCGCGGGGAACACTCTAA
MG205F	GGATAAACCGAGCACAAATA
MG215R	CCCGCTGGCGCGGGGAACACT
MG206F	AAACCGAGCACAAATATCAT
MG216R	AATCCCCGCTGGCGCGGGGAA
MG220F	CGCCAGCGGGGATAAAACCGAGCACAAATAT
MG220R	ATATTTGTGCTCGGTTTTATCCCCGCTGGCG
MG243F	AGTGTTCCCCGCGCAACGGGGATAAACCGAG
MG243R	CTCGGTTTATCCCCGTTGCGCGGGGAACACT
MG244F	GTGTTCCCCGCGCAGAGGGGATAAACCGAGC
MG244R	GCTCGGTTTATCCCCTCTGCGCGGGGAACAC
MG245F	TGTTCCCCGCGCAGCAGGGATAAACCGAGCA
MG245R	TGCTCGGTTTATCCCTGCTGCGCGGGGAACA
MG246F	GTTCCCCGCGCAGCGAGGATAAACCGAGCAC
MG246R	GTGCTCGGTTTATCCTCGCTGCGCGGGGAAC
MG247F	TTCCCCGCGCAGCGGAGATAAACCGAGCACA
MG247R	TGTGCTCGGTTTATCTCCGCTGCGCGGGGAA
MG248F	GTGTTCCCCGCGCAGAAAGGATAAACCGAGCAC
MG248R	GTGCTCGGTTTATCCTTTCTGCGCGGGGAACAC
MG239F	TGTTTAGAGTGTTCCAAACGCCCAGCGGGGATA
MG239R	TATCCCCGCTGGGCGTTTGGAACACTCTAAACA

[a]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.

Table S2. Primers and templates used for plasmid construction. Related to Experimental Procedures.

Plasmid	Primers for PCR	DNA template
S1	MG132F, MG132R	pCas1+2R
S2	MG82F, MG86R	"
S3	MG126F, MG126R	"
S4	MG85F, MG85R	"
S5	MG197F, MG197R	"
S6	MG198F, MG198R	"
S7	MG88F, MG196R	"
D1	MG199F, MG199R	"
D2	MG200F, MG200R	"
D3	MG218F, MG218R	"
D4	MG202F, MG202R	"
D5	MG203F, MG203R	"
D6	MG204F, MG204R	"
D7	MG205F, MG215R	"
D8	MG206F, MG216R	"
D9	MG219F, MG219R	"
D10	MG208F, MG208R	"
I1	MG200F, MG210R	"
I2	MG201F, MG211R	"
I3	MG202F, MG212R	"
I4	MG203F, MG213R	"
I5	MG204F, MG214R	"
I6	MG205F, MG215R	"
I7	MG206F, MG216R	"
I8	MG220F, MG220R	"
D5S1	MG243F, MG243R	"
D5S2	MG244F, MG244R	"
D5S3	MG245F, MG245R	"
D5S4	MG246F, MG246R	"
D5S5	MG247F, MG247R	"
D5S6	MG248F, MG248R	"
I4S1	MG239F, MG239R	"

Supplemental Experimental Procedures. Related to Experimental Procedures.

Reagents, strains and plasmids

Lysogeny broth (LB) medium (1% w/v tryptone, 0.5% w/v yeast extract, 0.5% w/v NaCl) and agar were from Acumedia. Antibiotics and L-arabinose were from Calbiochem. Isopropyl-β-D-thiogalactopyranoside (IPTG) was from Bio-Lab. Restriction enzymes, T4 polynucleotide kinase, Antarctic phosphatase and Quick Ligation Kit were from New England Biolabs. KAPA HiFi HotStart Ready Mix was from Kapa Biosystems. Taq DNA polymerase was from LAMDA Biotech. NucleoSpin Gel and PCR Clean-Up Kit were from Geneaid. The bacterial strains, plasmids and oligonucleotides used in this study are listed in Supplementary Table 1.

Plasmid construction

Plasmids were constructed using standard molecular biology techniques according to the manufacturers' instructions.

pCas1+2R plasmid encodes Cas1, Cas2 and a type I-E CRISPR array of a leader and a single repeat. pCas1+2R was constructed by amplifying the leader and repeat sequences of array I from the BL21-AI genome using oligonucleotides OA1F and OA1R (Supplementary Table 1). The amplified DNA was digested by *XbaI* and *SpeI* and ligated to *XbaI*-linearized pCas1+2 (Supplementary Table 1). The ligation yielded pCas1+2R plasmid that was further sequenced to exclude mutations introduced during cloning. The various mutant repeat plasmids were constructed using bidirectional PCR (repeats S2-S7, D1, D2, D4-D8, D10, and I1-I7) or site-directed mutagenesis (repeats S1, D3, D9, I8, D5S1-D5S6, and I4S1) methods. Plasmids constructed by bidirectional PCR were amplified using oligonucleotide pairs facing opposite directions followed by phosphorylation and self-ligation. Plasmids constructed by site-directed mutagenesis utilized complementary oligonucleotide pairs, each carrying the desired mutation with 15 bases of homologous sequence on both sides. Supplementary Table 2 lists the oligonucleotide combinations used to construct the various plasmids. Newly constructed plasmids were introduced into *E. coli* strain NEB5α by electroporation and sequenced to verify that the desired mutation was obtained. Once verified, the plasmids were purified from strain NEB5α and introduced into *E. coli* strain IYB5283 (Supplementary Table 1).

PCR products for deep sequencing

DNA of bacterial cultures subjected to acquisition assay was amplified in two consecutive PCRs termed PCR1 and PCR2. In PCR1, the reaction contained 25 μL Taq 2X Master Mix, 1.25 μL of 10 mM OA2F and OA2R primers (Supplementary Table 1), 5 μL bacterial culture and 16.5 μL double-distilled water. The PCR started with 3 min at 95°C followed by 35 cycles of 20 s at 95°C, 20 s at 55°C and 20 s at 72°C. The final extension step at 72°C was performed for 5 min. Part of the PCR1 content (20 μL) was purified using the DNA Clean-Up Kit and used for standard library preparation procedures followed by deep sequencing (MiSeq), while the remainder (30 μL) was loaded on a 1.8% (w/v) agarose gel and electrophoresed for 120 min at 120 V. Following electrophoresis, the expanded band was excised from the gel and purified using the DNA Clean-Up Kit. The extracted band served as the template for PCR2 aimed at amplifying the expanded CRISPR-array products. PCR2 contained 15 μL Taq 2X Master Mix, 0.5 μL of 10 mM RE10RD and IY230R1 primers (Supplementary Table 1), 2 ng of the gel-extracted DNA from PCR1 and double-distilled water to 20 μL. The PCR2 cycling program was identical to that of PCR1. The entire PCR2 content was loaded on a 1.8% agarose gel, electrophoresed, excised and purified from the gel under the same conditions as in PCR1, and used for standard library preparation procedures followed by deep sequencing (NextSeq500).