Published in final edited form as:

Trends Genet. 2010 August ; 26(8): 335–340. doi:10.1016/j.tig.2010.05.008.

Self-targeting by CRISPR: gene regulation or autoimmunity?

Adi Stern^{1,†}, Leeat Keren^{2,†}, Omri Wurtzel¹, Gil Amitai¹, and Rotem Sorek^{1,*}

¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

²Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel

Abstract

CRISPR/Cas is a recently discovered prokaryotic immune system, which is based on small RNAs (“spacers”) that restrict phage and plasmid infection. It has been hypothesized that CRISPRs can also regulate self gene expression by utilizing spacers that target self genes. By analyzing CRISPRs from 330 organisms we found that one in every 250 spacers is self targeting, and that such self-targeting occurs in 18% of all CRISPR-bearing organisms. However, complete lack of conservation across species, combined with abundance of degraded repeats near self-targeting spacers, suggests that self-targeting is a consequence of autoimmunity rather than gene regulation. We propose that accidental incorporation of self nucleic-acids by CRISPR can incur an autoimmune fitness cost, which may explain the abundance of degraded CRISPR systems across prokaryotes.

CRISPR/Cas, an acquired anti-viral system in prokaryotes

Clustered regularly interspaced short palindromic repeats (CRISPR) loci are found in nearly all of archaeal and about 40% of sequenced bacterial genomes. CRISPR loci, together with their associated *cas* genes, have recently been shown to constitute a defense system that restricts propagation of incoming viruses and plasmids [1,2]. CRISPR arrays are composed of short repeat sequences separated by similarly sized hyper-variable “spacer” sequences, flanked on one side by an AT-rich sequence called the leader. The discovery that CRISPR spacers often match DNA from foreign elements led to the realization that they represent a “memory of past genetic aggressions” [3-5]. Step by step, the mechanism underlying CRISPR defense has begun to unravel, yet our understanding of this system is far from complete. It has been revealed that the CRISPR locus is transcribed into a single RNA transcript, which is then further cleaved by the Cas proteins to generate smaller CRISPR RNA (crRNA) units, each including one targeting spacer [6]. These units then interfere with the incoming foreign genetic material by complementary base-pairing with the foreign nucleic acids [7-10]. CRISPR systems have been divided into different clusters based on their repeat sequences [11], which correlate with different subtypes of *cas* genes [12]. *cas* subtypes *mtube*, *ecoli* and *nmeni* were shown to likely target DNA [2,5,6,13], whereas the *cas* module *ramp* was shown to target RNA [14].

© 2010 Elsevier Ltd. All rights reserved.

*Corresponding author: rotem.sorek@weizmann.ac.il.

†These authors contributed equally

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Although CRISPR/Cas was initially prophesized to be analogous to eukaryotic RNA interference (RNAi) [15], it is now becoming clear that there are key differences between RNAi and CRISPR [10]. Nevertheless, the conceptual similarities between these two systems allow us to use our broader understanding of RNAi to guide the study of the CRISPR system [8]. Eukaryotic RNAi systems are divided into two branches: the antiviral branch that targets viruses and transposons for degradation, and the regulatory branch, that utilizes microRNAs (miRNAs) for translational repression of target mRNA molecules via partial base pairing [16]. Previous limited searches have revealed CRISPR spacers targeting chromosomal genes [4,5,17-19], and thus, based on the conceptual analogy between RNAi and CRISPR, it has been hypothesized that the CRISPR system in prokaryotes may also participate in gene regulation [9].

Here, we explored this possibility by studying self-targeting CRISPR spacers from all known CRISPR arrays [20] in all currently sequenced prokaryotic genomes (Table S1). Unexpectedly, our results point to a different explanation for self-targeting by CRISPR: leaky incorporation of self nucleic-acids leading to autoimmunity. We further explore this new concept of CRISPR-based autoimmunity from an evolutionary angle, as well as its consequences and fitness costs on CRISPR-bearing organisms.

Self-targeting CRISPR spacers

To identify potential self-targeting spacers, 23 550 spacers from 330 CRISPR encoding organisms were scanned for an exact full match between the spacer and a portion of the endogenous genomic sequence that is not part of a CRISPR array (termed target, or self proto-spacer). Our results reveal that 100 of 23 550 spacers (0.4%) are self-targeting (Table S2). However, encoding a self-targeting spacer is not a rare phenomenon: 59 out of 330 (18%) CRISPR-encoding organisms possess at least one array with at least one self-targeting spacer. These spacers are widely distributed over diverse phylogenetic lineages (Fig. S1), and are dispersed throughout different arrays in each organism.

Is CRISPR/Cas a regulatory system?

One of the basic postulates of evolutionary theory is that functional elements undergo purifying selection, leading to their conservation across different organisms [21]. Returning to the superficial analogy with eukaryotic RNAi, miRNAs are one of the most highly conserved non-coding elements in mammalian genomes [22]. Hence, an essential requirement of CRISPR functioning as an established regulatory system is the evolutionary conservation of the self-targeting spacers across several species. To test for such conservation, we compared the sequences of all CRISPR self-targeting spacers. However, we did not find a conserved self-targeting spacer in even one single case.

Considering the possibility that CRISPR regulation might occur via partial base pairing (as in eukaryotic miRNAs or in the *ramp* CRISPR-associated module [14]), we also examined spacers with partial or inexact matches to endogenous DNA. Once again, both partial and fully matching endogenous spacers showed no signs of conservation, i.e., were present in only one organism (apart from rare cases where self-targeting spacers were present in very closely related strains; see supplementary material). To summarize, our results showed that each pair of self-targeting spacer and target exists only in one organism. This lack of conservation casts doubts on the hypothesis that the self-targeting spacers we detected regulate self genes: had the initial insertion of a self-targeting spacer conferred an evolutionary advantage to the organism, and had it acquired a functional role in gene regulation, purifying selection would have led to its perpetuation.

Self-targeting spacers frequently target non-mobilome genes

We returned to examine the targets of the 100 full-match self-targeting spacers. About half of the self-targets were found to reside within elements of putative exogenous origin such as proviral sequences, transposon sequences, and established native plasmids. The existence of an exact match to a proviral sequence indicates that this virus once infected the organism, yet managed to escape CRISPR degradation. However, it is also possible that CRISPR has a role in preventing the induction of these latent viruses, and in general has a role in preventing the expansion of mobile elements.

Nevertheless, this role cannot explain all the self-targeting spacers detected: 53 spacers from 39 different organisms were found to target genes that are unlikely to be from a mobile origin, based on their putative gene function and on their gene neighborhood. Examples include spacers targeting 16S RNA, DNA polymerase I, tRNA synthetases, and others (Table 1). If so, what may explain the existence of a CRISPR spacer against a cellular gene?

Negative effects of self-targeting spacers

One possible explanation for the acquisition of self-targeting spacers is that they represent accidents of the CRISPR insertion mechanism, potentially leading to deleterious effects on the cell. Although the average size of a CRISPR array with self-targeting spacers is 30 spacers, we found that 37% of all self-targeting spacers are located at the first or second positions in the array (near the leader sequence), which is a four-fold enrichment compared to all spacers in our dataset ($P < 10^{-13}$; Fisher exact test; Table S1). Since addition of new spacers occurs in a polarized fashion proximal to the leader end of the CRISPR array, it appears that self-targeting spacers represent recent acquisition events by the CRISPR array. This implies that self-targeting spacers survive only a short time, and are thus not selectively neutral, but may rather be deleterious to the organism.

Based on these results, we hypothesized that following the integration of a self-targeting spacer, the CRISPR/Cas system must become inactivated in order to survive. For instance, *Lactobacillus acidophilus* NCFM harbors a self-targeting spacer against 16S ribosomal RNA, which could have a highly negative cost if functional. However, this organism appears to have lost all *cas* genes. Thus, the negative effect of autoimmune self-targeting spacers might explain the abundance of highly degraded CRISPR systems that contain *cas* pseudogenes [12,23] (Table 1).

We next sought to determine whether the self-targeting spacer itself could become inactivated without affecting the entire array. Studies have shown that the repeats are target sites for multiple Cas proteins, and participate in crRNA maturation and functioning [6,14,24]. We found that the two repeats flanking the self-targeting spacer are twice as likely to harbor mutations from the consensus repeat sequence, as compared to a background of all CRISPR spacers ($P < 0.005$; Fisher exact test; Table 1). Such mutations could potentially affect the maturation of the self-targeting spacer, while leaving the rest of the array functional. Self-targeting spacers flanked by mutated repeats were found throughout the array (and not specifically at the beginning of the array), suggesting that such mutations allow the self-targeting spacer to perpetuate without any negative effect on the organism (Fig. S2).

It was recently shown that CRISPR has a unique mechanism that enables it to avoid targeting the locus encoding the CRISPR itself [25]. Accordingly, base-pairing between three specific bases of the upstream repeat sequence and the crRNA results in protection from CRISPR degradation. We set out to test whether targets are protected in such a manner from their cognate spacers. Since it is possible that different CRISPR systems have slightly

different types of protection, we used a sliding window to scan whether three bases upstream or downstream from a non-mobile target match those of the repeat sequence. Our results showed that 14% and 17% of the targets match three base-pairs upstream or downstream of the repeat, respectively. Notably, these numbers do not deviate significantly from the expected number under a random binomial distribution (see supplementary material), and such base-pairing may therefore be a random property of the targets.

We next tested whether targets display proto-spacer adjacent motifs (PAMs). These motifs were first identified experimentally in *Streptococcus thermophilus* [18,26], and later identified in a widespread computational analysis as recurring sequences adjacent to the target/proto-spacer [27]. PAMs were suggested to take part in the acquisition and/or the interference stages for some CRISPR/Cas subtypes, since mutation at these sequences allowed viral escape [18,26,28]. Interestingly, PAMs appear to unimportant for CRISPR interference in arrays associated with subtype *mtube* [25] and with the Cas module *ramp* [14]. Furthermore, PAMs may also be less important for CRISPR interference with subtype *ecoli*, based on the ability of spacers lacking a matching PAM to restrict phage infection [6]. It is, however, important to note that the role of PAM in CRISPR interference has not yet been determined conclusively for each of the *cas* subtypes.

When testing whether targets display a PAM, we observed an intriguing pattern. First, for the targets associated with *cas* subtypes where PAM currently appears to play less of a role in interference (as defined above), 14 of 21 (67%) targets were putatively protected by flanking mutated repeats, by base-pairing with the repeat, or by loss of the *cas* operon. By contrast, in CRISPR types that require PAM for interference, such putative protection from autoimmunity was only observed in 2 of 17 (12%) targets ($P < 0.001$; Fisher exact test). Notably, in all these targets the PAM sequence was absent from the target. To summarize, our results tentatively suggest that when PAM is necessary for CRISPR interference, absence of a PAM sequence can protect the organism from autoimmunity, while when PAM is unnecessary for interference, mutations in other elements must occur to protect from the deleterious effects of CRISPR autoimmunity.

Autoimmunity in bacteria?

All in all, our analysis shows that the self-targeting CRISPR spacers are not evolutionary conserved, and that their occurrence is frequently associated with partial or full degradation of CRISPR activity. We thus conclude that the self-targeting spacers have most probably not been selected to take part in non-transient endogenous gene regulation. We propose a model whereby CRISPR self-targeting spacers result from leaky incorporation of self-nucleic-acids into CRISPR arrays (Figure 1), which may lead to a negative fitness cost to the organism. Our results suggest that some CRISPR subtypes are more prone to such leaky incorporation than others (see supplementary material). The rate of incorporation of self-targeting spacers is at least 0.2% (Table S2). This is probably an underestimate, as the calculation does not take into account self-targeting spacers that were highly deleterious and immediately cleared by purifying selection, led to immediate death of the cell, or were counter-selected at the population level.

One may envisage several different mechanisms by which such leaky incorporation may occur: viruses, plasmids or transposons may harbor genes from previous rounds of infection (as occurs during lateral gene transfer) [29], and this may lead to CRISPR recognizing these genes as foreign DNA. Alternatively, faulty incorporation of self-nucleic-acids may occur simply because of CRISPR “errors”. Notably, no matter what the mechanism of acquisition is, it is expected that in the absence of protection, harboring a self-targeting spacer will incur a fitness cost to the CRISPR-bearing organism. This cost may be high or low, depending

among other factors on the level of transcription of the self-targeting CRISPR spacer, on the mode of operation of the CRISPR array (i.e., targeting of DNA or RNA), and on the identity of the targeted gene. If this cost is relatively high, the self-targeting spacer, the targeted gene, or even the entire CRISPR/*cas* locus is prone to be lost, inactivated, or mutated. This scenario may become more likely in a virus-free environment. Interestingly, a recent study in strains of *Escherichia coli* supports the notion of such a fitness cost, showing that in strains where a spacer targeted one of the endogenous *cas* genes, the *cas* operon was lost [30].

Although the autoimmunity model is yet to be experimentally substantiated, this model may explain why a system that is so valuable in combating foreign invaders is present in only ~40% of bacteria. Together with lateral gene transfer [12,31], the model also explains the checkered pattern of existence or non-existence of CRISPR among closely related species, and the abundant occurrences of degraded CRISPR arrays (Figure 1). Nonetheless, we cannot rule out that other models exist to explain the existence of self-targeting spacers, where the spacers perform some type of transient regulation. For instance, an alternative model is that CRISPR targets endogenous host genes that contribute to virus replication. Experimental validation for this model would include deleting the self-targeting spacer and testing the fitness cost to the organism with and without viral infection.

We note that our results do not preclude the notion of alternative forms of CRISPR taking part in gene regulation (see also supplementary material), since our analysis focused on all identified ‘typical’ CRISPR structures that have multiple tandem repeats and spacers. If CRISPR had indeed evolved to perform gene regulation, the structure of the system would possibly have been altered and may thus differ from the CRISPR system as we know it today. Such a system may be composed of only one spacer flanked by partial repeats. Such altered, regulatory CRISPR systems are yet to be discovered.

Concluding remarks

Until today, the CRISPR system was heralded as an exceptional form of defense against foreign invaders, with apparently no fitness cost to the host [10]. However, we detected CRISPR spacers that match cellular genes having important housekeeping roles. This targeting is completely non-conserved and thus it is proposed here to be a flaw of the CRISPR mechanism. If indeed this self-targeting induces autoimmunity, this is a striking example of the Achilles’ heel of the CRISPR system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Uri Gophna, Debbi Lindell, Itay Mayrose, Sarit Edelheit, Hila Shimon, and Hila Sberro for stimulating discussions. R.S. is an EMBO Young Investigator. He was supported, in part, by the Israel Science Foundation Focal Initiatives in Research in Science and Technology (FIRST) program (grant 1615/09), NIH R01AI082376-01, the Wolfson Family Trust miRNA research program, the Minerva foundation, and the Yeda-Sela Center for basic research. A.S. was supported by the Clore postdoctoral fellowship. O.W. was supported by the Kahn Center for Systems Biology of the Human Cell, and is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

References

1. Lillestol RK, et al. A putative viral defence mechanism in archaeal cells. *Archaea* 2006;2:59–72. [PubMed: 16877322]

2. Barrangou R, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007;315:1709–1712. [PubMed: 17379808]
3. Pourcel C, et al. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 2005;151:653–663. [PubMed: 15758212]
4. Mojica FJ, et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 2005;60:174–182. [PubMed: 15791728]
5. Bolotin A, et al. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 2005;151:2551–2561. [PubMed: 16079334]
6. Brouns SJ, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008;321:960–964. [PubMed: 18703739]
7. van der Oost J, et al. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* 2009;34:401–407. [PubMed: 19646880]
8. Karginov FV, Hannon GJ. The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. *Molecular Cell* 2010;37:7–19. [PubMed: 20129051]
9. Sorek R, et al. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* 2008;6:181–186. [PubMed: 18157154]
10. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 2010;327:167–170. [PubMed: 20056882]
11. Kunin V, et al. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 2007;8:R61. [PubMed: 17442114]
12. Haft DH, et al. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 2005;1:e60. [PubMed: 16292354]
13. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 2008;322:1843. [PubMed: 19095942]
14. Hale CR, et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 2009;139:945–956. [PubMed: 19945378]
15. Makarova KS, et al. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* 2006;1:7. [PubMed: 16545108]
16. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 2004;5:522–531. [PubMed: 15211354]
17. Horvath P, et al. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol* 2009;131:62–70. [PubMed: 18635282]
18. Horvath P, et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 2008;190:1401–1412. [PubMed: 18065539]
19. Shah SA, et al. Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochem Soc Trans* 2009;37:23–28. [PubMed: 19143596]
20. Grissa I, et al. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 2007;8:172. [PubMed: 17521438]
21. Kimura, M. Neutral theory of molecular evolution. Cambridge University Press; 1983.
22. Pang KC, et al. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet* 2006;22:1–5. [PubMed: 16290135]
23. van der Ploeg JR. Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* 2009;155:1966–1976. [PubMed: 19383692]
24. Hale C, et al. Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *Rna* 2008;14:2572–2579. [PubMed: 18971321]
25. Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 2010;463:568–571. [PubMed: 20072129]
26. Deveau H, et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 2008;190:1390–1400. [PubMed: 18065545]

27. Mojica FJ, et al. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 2009;155:733–740. [PubMed: 19246744]
28. Semenova E, et al. Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS microbiology letters* 2009;296:110–116. [PubMed: 19459963]
29. Partridge SR, et al. Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol Rev* 2009;33:757–784. [PubMed: 19416365]
30. Diez-Villasenor C, et al. Diversity of CRISPR loci in *Escherichia coli*. *Microbiology*. 2010 Epub ahead of print.
31. Godde JS, Bickerton A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* 2006;62:718–729. [PubMed: 16612537]

Legend

- Spacer
- Repeat
- *cas* genes
- Phage DNA/RNA
- Host DNA/RNA
- * Novel spacer

Figure 1. A model for CRISPR autoimmunity: leaky incorporation of self-DNA/RNA and its possible outcomes

During infection, genetic material in the cell will include both viral and host nucleic acids. CRISPR activity leads to insertion of new spacers, most likely derived from foreign genetic material. Virus-resistant colonies (boxed in purple) will include one or more spacers against the virus, but may also include spacers against the endogenous host DNA. Some of these colonies will not survive due to autoimmunity. However, if the fitness cost of autoimmunity is low enough, these colonies may thrive due to the beneficial existence of the anti-viral spacer. To prevent the negative effects associated with autoimmunity, inactivation of the CRISPR/*cas* locus can occur in a variety of ways, including mutation or loss of the *cas* genes, or mutations at the leader sequence. Inactivation of the self-targeting spacer may also occur by several mechanisms: mutation of the adjacent repeats, mutation of the spacer sequence, mutation of the targeted gene, or mutation of the PAM.

Table 1

A list of organisms bearing self-targeting spacers against non-mobile elements.

Organism	Target protein description	Putative protection from autoimmunity ^a	Associated <i>cas</i> subtype
<i>Actinobacillus pleuropneumoniae</i> sv 3 JL03	Hydrogenase maturation factor	4	ypest
<i>Bifidobacterium adolescentis</i> ATCC 15703	3-phosphoshikimate 1-carboxyvinyltransferase	4	dvulg
<i>Bifidobacterium longum</i> DJO10A	<i>fhu</i> operon transcription regulator	4	nmeni
<i>Campylobacter hominis</i> ATCC BAA-381	Preprotein translocase subunit SecA		unknown
<i>Chlorobium limicola</i> DSM 245	Succinyl-CoA synthetase, beta subunit		csx.ramp
<i>Chlorobium phaeobacteroides</i> DSM 266	Amidophosphoribosyltransferase		unknown
<i>Clostridium botulinum</i> E3 Alaska E43	Amino acid carrier protein AlsT	1	ramp ^b
<i>Clostridium tetani</i> E88	Putative S-layer		hmari
<i>Clostridium tetani</i> E88	Stage IV sporulation protein A	4	csx.tneap
<i>Clostridium tetani</i> E88	DNA mismatch repair protein	4	csx.tneap
<i>Clostridium tetani</i> E88	Phosphoribosylformylglycinamide synthase II	4	csx.tneap
<i>Clostridium tetani</i> E88	Putative sporulation sigma-E factor processing	4	csx.tneap
<i>Enterobacter</i> sp. 638	FdrA family protein	4	ypest
<i>Erwinia carotovora</i> atroseptica SCRI1043	Putative plasmid transfer protein	2,4	ypest
<i>Flavobacterium psychrophilum</i> JIP02/86	Hypothetical protein		nmeni
<i>Francisella philomiragia</i> <i>philomiragia</i> ATCC 25017	Hypothetical protein		unknown
<i>Frankia alni</i> ACN14a	Putative acyl-CoA dehydrogenase	4	ecoli
<i>Frankia</i> sp. Cc13	Amino acid adenylation		dvulg
<i>Geobacter uraniumreducens</i> Rf4	2-methylisocitrate lyase	2	unknown
<i>Lactobacillus acidophilus</i> NCFM	16S ribosomal RNA	1,2,4	ecoli ^b
<i>Lactobacillus acidophilus</i> NCFM	Peptide binding protein	1,2,4	ecoli ^b
<i>Lactobacillus acidophilus</i> NCFM	Dipeptidase	1,4	ecoli ^b
<i>Lactobacillus delbrueckii</i> <i>bulgaricus</i>	Hypothetical protein	4	ecoli
<i>Methanospirillum hungatei</i> JF1	Dihydroorotate dehydrogenase 1B		csc.apern
<i>Methylobacillus flagellatus</i> KT	Glycosyl transferase, group 1	4	dvulg
<i>Mycoplasma arthritidis</i> 158L31	Phenylalanyl-tRNA synthetase beta subunit		unknown
<i>Mycoplasma gallisepticum</i> R	Hypothetical protein		nmeni
<i>Mycoplasma gallisepticum</i> R	ParC/GyrA; Topoisomerase IV subunit A		nmeni
<i>Mycoplasma mobile</i> 163K	Putative glycosyltransferase		nmeni
<i>Myxococcus xanthus</i> DK 1622	Putative lipoprotein	2,4	dvulg
<i>Nocardia farcinica</i> IFM 10152	Putative non-ribosomal peptide synthetase	2,4	ecoli
<i>Nocardia farcinica</i> IFM 10152	Putative transcriptional regulator	4	ecoli
<i>Pelobacter carbinolicus</i> DSM 2380	Histidyl-tRNA synthetase	2	ecoli

Organism	Target protein description	Putative protection from autoimmunity ^a	Associated <i>cas</i> subtype
<i>Pelobacter propionicus</i> DSM 2379	Hydrophobe/amphiphile efflux-1 (HAE1) family	2,4	ecoli
<i>Pelobacter propionicus</i> DSM 2379	DNA topoisomerase I	2,4	ecoli
<i>Pelobacter propionicus</i> DSM 2379	RND efflux system outer membrane lipoprotein	2,4	ecoli
<i>Porphyromonas gingivalis</i> W83	Saccharopine dehydrogenase	4	unknown
<i>Porphyromonas gingivalis</i> W83	Saccharopine dehydrogenase	4	unknown
<i>Prosthecochloris aestuarii</i> SK413	Chromosome segregation protein SMC	2,4	ecoli
<i>Pyrococcus horikoshii</i> OT3	Hypothetical protein		unknown
<i>Roseiflexus castenholzii</i> DSM 13941	PAS domain-containing protein	4	mtube
<i>Salinispora arenicola</i> CNS-205	Hypothetical protein	4	ecoli
<i>Streptococcus agalactiae</i> 2603V/R	Uracil permease	4	nmeni
<i>Streptococcus agalactiae</i> A909	DNA polymerase I	4	nmeni
<i>Streptomyces avermitilis</i> MA-4680	Multidrug resistance efflux protein	2,4	ecoli
<i>Streptomyces griseus griseus</i> NBRC 13350	PII uridylyl-transferase	2	ecoli
<i>Sulfolobus tokodaii</i> 7	Hypothetical protein	3	apern
<i>Treponema denticola</i> ATCC 35405	Hypothetical protein	4	nmeni
<i>Yersinia pestis</i> Antiqua	Phosphotransferase enzyme II, A component	4	ypest
<i>Clostridium tetani</i> E88	Intergenic region	3	hmari
<i>Granulibacter bethesdensis</i> CGDNIH1	Intergenic region	4	ecoli
<i>Lactobacillus acidophilus</i> NCFM	Intergenic region	1,2	ecoli ^b
<i>Porphyromonas gingivalis</i> W83	Intergenic region	3	ramp

^aEvidence suggesting that the target is protected from autoimmunity is marked by 1-4: (1) Lack of *cas* genes, (2) mutated adjacent repeats, (3) extended base-pairing with the upstream flanking repeat, and (4) absence of PAM as defined by Mojica et al. [27] for some of the repeat clusters. Full details on all self-targeting spacers can be found in Supplementary Table S1.

^bFor organisms with no *cas* genes, the subtype identity of the inferred lost *cas* operon was deduced based on the homologous array in a related strain, and based on the identity of the repeats.