Bacterial CRISPR: accomplishments and prospects

Jason M Peters¹, Melanie R Silvis¹,², Dehua Zhao⁴,⁵,⁶, John S Hawkins¹,³, Carol A Gross¹ and Lei S Qi⁴,⁵,⁶

In this review we briefly describe the development of CRISPR tools for genome editing and control of transcription in bacteria. We focus on the Type II CRISPR/Cas9 system, provide specific examples for use of the system, and highlight the advantages and disadvantages of CRISPR versus other techniques. We suggest potential strategies for combining CRISPR tools with high-throughput approaches to elucidate gene function in bacteria.

Addresses
¹ Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA 94158, USA
² Tetrad Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA
³ Biophysics Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA
⁴ Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
⁵ Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA
⁶ Chemistry, Engineering & Medicine for Human Health (ChEM-H), Stanford University, Stanford, CA 94305, USA

Corresponding authors: Gross, Carol A (cgrossuscsl@gmail.com) and Qi, Lei S (stanley.qi@stanford.edu)

Current Opinion in Microbiology 2015, 27:121–126
This review comes from a themed issue on Microbial systems biology
Edited by Eric Brown and Nassos Typas

http://dx.doi.org/10.1016/j.mib.2015.08.007
1369-5274/© 2015 Elsevier Ltd. All rights reserved.

Introduction

Bacteria exist in a sea of foreign DNA that is internalized via phage infection or various DNA transfer and uptake systems. About 40% of bacterial species use CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) as a genetic adaptive immune system to defend against invading DNA [1,2]. CRISPR consists of CRISPR (cr) RNAs that target specific foreign DNA sequences, primarily via RNA–DNA binding, and their associated Cas (CRISPR-associated) proteins that then cleave the targeted DNA [3,4]. The sequences of crRNAs that dictate DNA targeting by CRISPR are a direct result of acquisition of foreign DNA into the CRISPR array during phage or plasmid exposure; in this way CRISPR records previous encounters and allows specific restriction of remembered ‘attackers.’ Furthermore, there is increasing evidence that CRISPR arrays and Cas proteins often carry additional roles in endogenous gene regulation and genome rearrangements; sometimes by distinct mechanisms than those employed in immunity (reviewed in [5,6]). Recent discoveries of diverse CRISPR systems that function in immunity by alternative mechanisms as well as anti-CRISPR systems deployed by bacteriophage as part of an evolutionary arms race (reviewed in [5,7]) suggest that we have only begun to appreciate the fascinating biology and technological potential of CRISPR.

There are three major types of CRISPRs, some of which use multiple Cas proteins to target and degrade DNA, but the Type II system was the starting point for genome engineering due to its simplicity [8,9]. Most work has focused on the Streptococcus pyogenes Type II CRISPR/Cas9 system in which the natural two RNA duplex has been further simplified to a chimeric single guide RNA (sgRNA) which targets Cas9 to specified DNA sequences. The endonucleolytic activity of Cas9 then causes a double-strand (ds) break in the target DNA [8,9]. Cas9 has HNH- and RuvC-like nuclease domains, both of which are required for cutting dsDNA [9,10], and an alphahelical lobe that makes contacts with the sgRNA (other Cas9 structural features are reviewed in [11]). In addition to the 20 nt sequence specified in the sgRNA, Cas9 requires a short protoscaler adjacent motif (PAM; NGG in S. pyogenes) to recognize DNA. Once bound to DNA, the Cas9-sgRNA complex is extremely stable; in vitro binding kinetics indicate nearly irreversible binding over long timescales [12]. The most important feature of CRISPR is its programmability; Cas9 can be directed to any PAM-adjacent sequence in the genome by modifying the basepairing region of the sgRNA. A second important feature is its versatility; whereas Cas9 is used for precise genome editing, a catalytically dead Cas9 (dCas9) unable to cleave DNA is used to precisely activate and repress gene expression [13**,14•]. CRISPR/Cas9 is revolutionizing genome engineering, and many reviews have detailed its use in eukaryotes [15–18,19•]. Here, we briefly review CRISPR/Cas9 systems for editing and gene regulation in bacteria.

CRISPR editing

CRISPR/Cas9 genome editing, used as a tool for site-directed mutagenesis in vivo, has radically increased the ease and throughput of genetic studies in eukaryotic and
especially mammalian systems, but has been applied only sporadically to bacteria so far (reviewed in [20*]). CRISPR/Cas9 editing studies in γ proteobacteria [21**,22,23] represent improvements to existing genome editing technology, while those in less studied Firmi- cutes [24,25], and Actinobacteria [26,27] mark the transfer of a highly portable technology into bacterial species that previously lacked sufficient genetic tools.

The basic protocol for editing relies on the fact that double-strand DNA breaks caused by Cas9 are fatal events in most bacterial genomes. Cas9 is targeted in a sequence-specific manner to un-edited genomes, thereby selecting for recombinants that are designed to lack the targeting or nearby PAM sequence (Figure 1). This CRISPR-based negative selection in combination with traditional recombineering is the primary methodology for Cas9 genome engineering in bacteria [20*]. In bacteria with a low intrinsic frequency of recombination, expression of recombinases — such as the lambda-red recombinase [28] — in addition to the active CRISPR editing system enhances recovery of bacteria that have undergone the desired recombination event [21**].

Thus far, bacterial CRISPR editing is plasmid-based and introduced to recipients via direct transformation (e.g., electroporation) [25,23,24], or conjugation [27,26]. Plasmid-based approaches allow high-throughput assembly of large numbers of gene-specific sgRNAs by pooled cloning. Bacteriophage offer another route of delivering the system for editing, as they have been shown to efficiently deliver CRISPR/Cas9 as a sequence-specific antimicrobial [29].

Existing systems for expression of the active CRISPR system in recipient bacteria may require optimization. Cas9 toxicity makes it important to limit exposure of cells to the editing system by placing Cas9 under tight control of an inducible promoter and employing a curable plasmid (e.g., with a temperature-sensitive origin of replication). An inducible cas9 gene has two main benefits: it reduces the opportunity for potential off target effects and

---

**Figure 1**

CRISPR genome editing in bacteria, sgRNAs are designed to target the ‘unedited’ genome. The Cas9-sgRNA complex cuts unedited target DNA, resulting in a lethal double-strand break; this is a negative selection. DNA edited with the assistance of recombinases (e.g., λ Red in E. coli) is immune to cleavage by the Cas9-sgRNA complex that targets unedited DNA. Exo is an exonuclease that processes double stranded DNA to single stranded ends, while β drives recombination of single stranded DNA.
it lessens the likelihood of suppressor mutations that inactivate Cas9 or sgRNA. These suppressor mutations may result in ‘escaper’ colonies that are initially indistinguishable from correct recombinants [21**,20*]. Codon optimization of Cas9 may also be necessary. *S. pyogenes* cas9 has low GC content (35% GC), and optimizing Cas9 codons may be necessary in high GC organisms, as has been found in Streptomyces [27,26]. Problems with optimizing and regulating Cas9 expression have been circumvented in eukaryotic systems by electroporation of active Cas9-sgRNA complexes directly into eukaryotic cells [30,31]; this possibility remains unexplored in bacteria.

**CRISPR interference (CRISPRi)**

Bacterial CRISPRi has been extensively characterized in *E. coli* [13**,14*,32]. In this organism, the dCas9-sgRNA complex represses transcription either by occluding RNA polymerase (RNAP) binding to promoter DNA [13**,14*] or by causing a steric block to transcription elongation, as demonstrated by deep sequencing of nascent transcripts (Figure 2A) [13**,33]. An elongation block targeting the non-template strand is most effective (up to 1000-fold repression). The repression is highly specific, as determined by RNA-seq [13**]. CRISPRi has been used to predictably control genetic circuits in model bacteria [13**], modulate essential gene expression in pathogenic bacteria [34], and alter flux through metabolic pathways [32]. Recent studies have established a chromosomal CRISPRi system in the model Gram-positive bacterium *Bacillus subtilis* (Peters et al., unpublished), the important human microbiome commensal *Bacteroides thetaiotaomicron* [52], and in *Actinomyces* that produce medically-relevant secondary metabolites [53].

CRISPRi has important advantages compared to traditional techniques for regulating gene expression. First, it is easy to repress new targets simply by inserting a new 20 nt base-pairing region into the sgRNA (methods for sgRNA construction are described in [35,36]). Second, CRISPRi is scalable; thousands of sgRNAs can be constructed using pooled cloning strategies with complex oligo libraries [37,38,39*,40]. Third, CRISPRi is inducible, which enables manipulation of essential genes by partial repression [34]. Further, inducible CRISPRi libraries of non-essential genes are likely more genetically stable (i.e., less likely to accumulate suppressor mutations) during passaging and amplification than transposon or gene deletion libraries in which gene products are constitutively inactivated. Finally, CRISPRi can be multiplexed; i.e., several genes can be simultaneously repressed in the same
cell using multiple sgRNAs [13**,32]. Multiplexing is important for genetic and synthetic biology applications that require modulating the expression of many genes; this is particularly useful for bacteria with extremely long doubling times (e.g., 24 hours for Mycobacterium tuberculosis [34]), as making multiple deletion mutations by standard techniques in such strains is very time consuming. The major disadvantage of CRISPRi is that it reduces expression of downstream genes in operons (i.e., polarity) due to the dCas9 block to elongating RNAP. In some cases, we have also observed reduced expression of upstream genes in operons (i.e., ‘reverse polarity’); these effects may be transcript or organism-specific (Peters et al., unpublished). The fact that operons often contain genes of related function somewhat mitigates this issue. As is the case with CRISPR editing, the strong selective pressure associated with CRISPRi knockdown of genes required for growth (i.e., essential genes) may result in mutations that inactivate the CRISPR system.

**CRISPR activation (CRISPRa)**

CRISPRa enhances transcription of target genes by using a modified dCas9-sgRNA complex containing activator domains to recruit RNA polymerase to promoter DNA. In the existing E. coli CRISPRa system, dCas9 is fused to the ω subunit of RNAP (dCas9:ω). In the presence of an appropriate sgRNA, RNAP is recruited to a position upstream of the promoter, thereby activating transcription (Figure 2B) [14*]. Initial characterization based on four promoters indicates that there may be a narrow window for efficient activation: repression occurs if dCas9:ω is too close, and activation fails when dCas9:ω is too far upstream. Activation works best for weak promoters (the weakest promoter tested was upregulated by ~28-fold [14*]). This system requires further characterization and optimization to be broadly applicable.

An optimized bacterial CRISPRa system could be used to perform genome-scale gain-of-function screens (without the need to clone thousands of genes), to mitigate polar effects of CRISPRi (by targeting the many weak promoter-like sequences in intergenic regions), or to create an ‘allelic series’ of gene overexpression and depletion strains using CRISPRa and CRISPRi, respectively (as demonstrated in mammalian cells [39*]). The last application is a powerful way to interrogate both gene function and mechanisms of antibiotic action. To realize this potential, we may need alternative strategies for bacterial CRISPRa that are less dependent on precise spacing from the promoter. dCas9 variants that recruit activators via long, flexible linkers, analogous to those used in mammalian CRISPRa systems [41,39*] could reduce the spacing dependency. Additionally, the published E. coli CRISPRa system requires a strain that lacks the native copy of ω [14*], which may alter complex regulatory pathways such as the stringent response [42].

**sgRNA design**

One of the most attractive features of CRISPR is the ability to flexibly and precisely target Cas9 to essentially any genomic location. DNA targeting requires recognition of a short PAM sequence by Cas9 [8,12], and base-pairing of the 20-nt target sequence with the spacer region of the sgRNA (Figure 2C). Thus, sgRNA designs must meet two criteria: specific targeting to a region with a PAM (e.g., 3 nt PAM +20 nt sgRNA for S. pyogenes) and minimized off-target binding. The strongest sgRNAs (i.e., those resulting in highest fold repression by dCas9) have a perfect match for a PAM-adjacent target, but targets with imperfect matches may still be bound by the dCas9-sgRNA complex. A single mismatch in the 12 nucleotides at the 3’ end of the sgRNA or in the mandatory GG bases of the PAM sequence will greatly curtail the repressive effect of dCas9-sgRNA [13**,14*]. Multiple mismatches between the sgRNA and a potential target will further reduce the repressive effect [13**,14*,39*]. Repression can also be tuned by targeting different positions within the gene: locating the sgRNA target in the 3’ end or template strand of a gene reduces repression [13**,34].

Although imperfectly matched sgRNA may have lower efficacy, they may still result in off-target cutting by Cas9 or mitigate CRISPRi repression by titrating dCas9. Computational tools like Bowtie [43] can be used to compare candidate sgRNA sequences to all other PAM-adjacent off-target sites, using a weighted threshold function to discard sgRNA designs with potential off target effects [39*]. More elaborate algorithms incorporating data from assays such as ChIP-Seq have also been developed [44,45], although transient off-target association of dCas9-sgRNA with DNA detected by ChIP-seq may not be functionally relevant, especially in the case of repression.

**Future directions**

There are no published high-throughput CRISPR screens in bacteria, likely because there are outstanding issues to be resolved before such screens can be performed. Eukaryotic CRISPR editing screens rely on non-homologous end joining (NHEJ) pathways to mitigate Cas9 killing by repairing double strand DNA breaks (DSBs) in DNA [16]; this mutagenic process leads to loss of function insertion/deletion mutations. Many bacteria lack a NHEJ pathway [46]; e.g., E. coli K-12), or express it conditionally (e.g., B. subtilis [47]). The simplest bacterial NHEJ systems consist of just two proteins, Ku and LigD [48], suggesting that CRISPR editing screens may be possible by inducing NHEJ concomitant with Cas9 cutting. However, even in bacterial strains with active NHEJ (e.g., Mycobacterium smegmatis), DSBs cause significant killing [49].

Genome-scale CRISPRi screens also require optimization. Mammalian CRISPRi utilizes a chromatin silencer
We also will have to solve the question of CRISPRi delivery to a broad range of bacteria. Ji et al. took an initial step using a matable plasmid to transfer a CRISPRi system between two _E. coli_ strains [51] (Figure 2D). However, it is likely that there will be no ‘one size fits all’ CRISPRi system that functions in all bacteria; instead, a modular approach in which components (e.g., regulatable promoters) can be swapped in and out of the matable system has more versatility, and, thus, more potential to be robust to species specific issues.

Numerous powerful tools originating from bacteria and phages have transformed the landscape of biological research. The CRISPR-Cas9 system may revolutionize modern genetics in ways comparable in magnitude to the advent of recombinant DNA technologies. Bacterial CRISPR-Cas9 holds enormous promise to accelerate the pace of gene function discovery by radically increasing the scale of genetic screens and by providing novel genetic tools for previously intractable organisms.

Acknowledgements

We would like to thank members of the Gross and Qi labs for critically reading the manuscript. L.S.Q. was supported by an NIH Director’s Early Independence Award (OD017887) and NIH R01 (DA036858); J.M.P. is a Ruth L. Kirchstein Fellow (F32 GM108222).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


The authors demonstrate the efficacy of CRISPR-based gene silencing in _bacteria_ using only dCas9 and a chimeric sgRNA. They show that transcriptional repression by dCas9-sgRNA can result from a steric block to the progress of RNA polymerase during elongation.


The authors demonstrate CRISPRi activation in _bacteria_ by directing a dCas9-sgRNA translational fusion upstream of weak promoters that control expression of reporter genes.


This review places CRISPR in the context of other genome engineering tools, provides an excellent description of the history of CRISPR function and mode of action and briefly discusses its applications for editing and gene regulation in eukaryotes.


An excellent, detailed review devoted entirely to bacterial genome editing using CRISPR-Cas9.


The first report of marker-free mutations in bacteria assisted by Cas9-sgRNA sequence-specific counter-selection in _E. coli_ and _S. pneumoniae_. In combination with recombining, the authors demonstrate that the majority of cells surviving Cas9 cutting have edited genomes, as well as simultaneous editing of multiple loci. They also establish guidelines for abolishing Cas9-sgRNA recognition in the desired editing template, including mutations in the PAM and seed sequence.


