Prediction and Validation of Native and Engineered Cas9 Guide Sequences

Alexandra E. Briner,[1] Emily D. Henriksen,[1] and Rodolphe Barrangou[1,2]

[1]Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, North Carolina 27695

Cas9-based technologies rely on native elements of Type II CRISPR–Cas bacterial immune systems, including the *trans*-activating CRISPR RNA (tracrRNA), CRISPR RNA (crRNA), Cas9 protein, and protospacer-adjacent motif (PAM). The tracrRNA and crRNA form an RNA duplex that guides the Cas9 endonuclease to complementary nucleic acid sequences. Mechanistically, Cas9 initiates interactions by binding to the target PAM sequence and interrogating the target DNA in a 3′-to-5′ manner. Complementarity between the guide RNA and the target DNA is key. In natural systems, precise cleavage occurs when the target DNA sequence contains a PAM flanking a sequence homologous to the crRNA spacer sequence. Currently, the majority of commercial Cas9-based genome-editing tools are derived from the Type II CRISPR–Cas system of *Streptococcus pyogenes*. However, a diverse set of Type II CRISPR–Cas systems exist in nature that are potentially valuable for genome engineering applications. Exploitation of these systems requires prediction and validation of both native and engineered dual and single guide RNAs to drive Cas9 functionality. Here, we discuss how to identify the elements of these immune systems to develop next-generation Cas9-based genome-editing tools. We first discuss how to predict tracrRNA sequences and suggest a method for designing single guide RNAs containing only critical structural modules. We then outline how to predict the PAM sequence, which is crucial for determining potential targets for Cas9. Finally, validation of the system elements through transcriptome analysis and interference assays is essential for developing next-generation Cas9-based genome-editing tools.

MATERIALS

It is essential that you consult the appropriate Material Safety Data Sheets and your institution's Environmental Health and Safety Office for proper handling of equipment and hazardous materials used in this protocol.

Reagents

Direct-zol RNA MiniPrep Kit (Zymo Research)

The Direct-zol RNA MiniPrep Kit from Zymo Research is highly recommended for extraction of RNA as it allows purification of RNA molecules as small as 17 nt, thereby retaining the smaller crRNAs that many traditional kits discard.

Next-generation sequencing reagents (HiSeq or MiSeq; Illumina)
Sample for RNA extraction
TruSeq Small RNA Sample Preparation Kit (Illumina)

---

[2]Correspondence: rbarran@ncsu.edu

Equipment

Basic Local Alignment Search Tool (BLAST) and Conserved Domain Database (CCD) (National Center for Biotechnology Information [NCBI])

CRISPR identification program

CRISPRfinder (http://crispr.u-psud.fr/Server/) is a web-based CRISPR identification platform that allows users to upload their own sequences of interest as well as browse genomes annotated directly from the NCBI database (Grissa et al. 2007). CRISPR Recognition Tool (CRT) (http://www.room220.com/crt/ [Bland et al. 2007]) is available as a command-line tool or plug-in for several commercially available graphical user interface (GUI) bioinformatics platforms (e.g., Geneious9 [Biomatters]). Both programs identify potential CRISPR spacers and repeats, based on a repeat-finding algorithm.

Genome sequence of interest

Quality control software for RNA sequencing data (e.g., FastQC)

RNA folding prediction software (e.g., NUPACK; http://www.nupack.org/ [Zadeh et al. 2011])

Sequence motif identification program (e.g., WebLogo [Crooks et al. 2004])

Short read sequence alignment tool (e.g., Bowtie 2 [Langmead and Salzberg 2012])

METHOD

Predicting CRISPR Repeats and *cas* Genes In Silico

Type II CRISPR–Cas systems have been identified in only 5% of bacterial genomes and are yet to be detected in archaeal genomes (Makarova et al. 2011; Chylinski et al. 2013). These are the least widely distributed CRISPR–Cas systems overall and are enriched in Firmicutes and Actinobacteria. Notably, lactic acid bacteria and bacteria closely related with humans have been shown to contain Type II systems more frequently.

1. Upload the genome sequence of interest to a CRISPR identification program.

2. Identify the CRISPR repeat consensus sequence (Fig. 1A).

   Although documented CRISPR repeats have been between 24 and 47 nt in length, overall most CRISPR repeats are between 29 and 36 nt. For Type II systems in particular, CRISPR repeats are generally 36 nt in length (Makarova et al. 2011; Chylinski et al. 2013).

3. Search the region flanking the repeat-spacer array for *cas* genes, looking specifically for the universal *cas1* and the signature Type II gene, *cas9*, to ensure you are investigating a Type II CRISPR–Cas system (Fig. 1B). Confirm the *cas* gene annotation using BLAST or CDD.

   For the Type II subtypes, the prototypical Cas9 proteins are from the CRISPR–Cas systems found in Strepto-coccus thermophilus *CNRZ 1066,* Legionella pneumophila *str.* Paris, *and* Neisseria lactamica *020-06, representing the II-A, II-B, and II-C subtypes, respectively (Makarova et al. 2011). Cas9 should contain the two nickase domains RuvC and HNH (Sapranauskas et al. 2011).*

4. Determine the orientation of the array by identifying the leader sequence and the terminal repeat (Fig. 1C).

   The leader sequence is adjacent to the first repeat and initiates transcription of the repeat-spacer array. It is often AT-rich. The terminal end of the array can be determined by identifying CRISPR repeats with single point mutations. The terminal repeat often differs from the consensus repeat sequence by several mutations, frequently at the 3′ end. The correct orientation of the array should start at the leader on the 5′ side and end with the terminal repeat on the 3′ side.

Predicting tracrRNA In Silico

5. To search for the antirepeat portion of the tracrRNA, perform a local BLAST nucleotide alignment (Altschul et al. 1997) between the consensus CRISPR repeat sequence and DNA sequences within one of four noncoding regions: within 500 nt upstream of Cas9, between *cas9* and *cas1*, between *csn2/cas4* and the repeat-spacer array, or within 500 nt downstream from the repeat-spacer array (Chylinkski et al. 2014) (Fig. 1D). Use the following BLAST parameters.
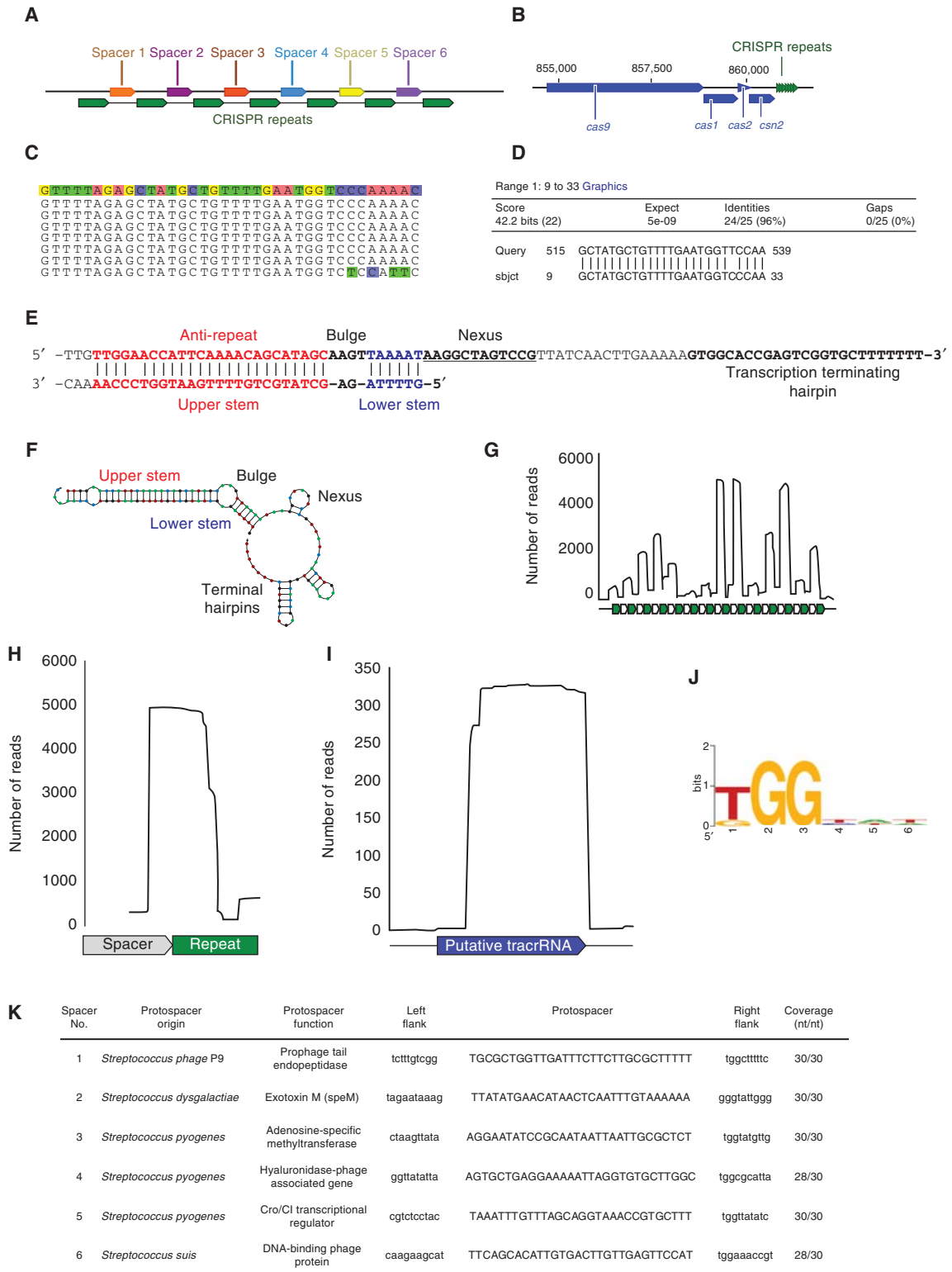
A Spacer 1 Spacer 2 Spacer 3 Spacer 4 Spacer 5 Spacer 6

CRISPR repeats

B CRISPR repeats

855,000 857,500 860,000

cas9 cas1 cas2 csn2

C
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC
GTTTTAGAGCTATGCTGTTTTGAATGGTCTCCATTC

D
Range 1: 9 to 33 Graphics

Score	Expect	Identities	Gaps
42.2 bits (22)	5e-09	24/25 (96%)	0/25 (0%)

Query 515 GCTATGCTGTTTTGAATGGTTCCAA 539
            |||||||||||||||||||| ||||
sbjct   9 GCTATGCTGTTTTGAATGGTCCCAA 33

E
Anti-repeat Bulge Nexus

5′ -TTGTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3′
3′ -CAAAACCCTGGTAAGTTTTGTCGTATCG-AG-ATTTTG-5′

Transcription terminating hairpin

Upper stem Lower stem

F Upper stem Bulge Nexus
Lower stem
Terminal hairpins

G Number of reads

H Number of reads
Spacer Repeat

I Number of reads
Putative tracrRNA

J bits

K

Spacer No.	Protospacer origin	Protospacer function	Left flank	Protospacer	Right flank	Coverage (nt/nt)
1	*Streptococcus phage* P9	Prophage tail endopeptidase	tctttgtcgg	TGCGCTGGTTGATTTCTTCTTGCGCTTTTT	tggctttttc	30/30
2	*Streptococcus dysgalactiae*	Exotoxin M (speM)	tagaataaag	TTATATGAACATAACTCAATTTGTAAAAAA	gggtattggg	30/30
3	*Streptococcus pyogenes*	Adenosine-specific methyltransferase	ctaagttata	AGGAATATCCGCAATAATTAATTGCGCTCT	tggtatgttg	30/30
4	*Streptococcus pyogenes*	Hyaluronidase-phage associated gene	ggttatatta	AGTGCTGAGGAAAAATTAGGTGTGCTTGGC	tggcgcatta	28/30
5	*Streptococcus pyogenes*	Cro/CI transcriptional regulator	cgtctcctac	TAAATTTGTTTAGCAGGTAAACCGTGCTTT	tggttatatc	30/30
6	*Streptococcus suis*	DNA-binding phage protein	caagaagcat	TTCAGCACATTGTGACTTGTTGAGTTCCAT	tggaaaccgt	28/30

FIGURE 1. Annotation and validation of Type II CRISPR–Cas9 system elements. (*A*) Visualization output (in the Geneious7 graphical user interface) of the CRISPR Recognition Tool (CRT) (Bland et al. 2007), which is used to identify CRISPR repeats and spacers. Variable spacers (colored arrows) are flanked by conserved CRISPR repeats (green arrows). (*B*) Identification and confirmation of *cas* genes adjacent to the CRISPR repeat-spacer array. Predicted *cas* genes should be confirmed using NCBI's BLAST or CDD to ensure correct annotation of the locus. (*C*) The terminal repeat often contains nucleotide mutations that can be used to determine the directionality of the CRISPR array. (*Legend continues on following page*.)

Algorithm	Somewhat similar sequences (blastn)
Word size	7
Match/mismatch scores	1, –2
Gap costs	Existence: 1, Extension: 2

See Troubleshooting.

6. From the alignment results, identify "potential tracrRNA" sequence(s) (Fig. 1E).

   The alignment should cover between 8 nt (usually for Type II-B and II-C systems) and three-quarters of the length of the CRISPR repeat (~36 nt). For example, if you have a 36-nt repeat, look for an antirepeat in which the alignment spans 8–27 nt of the repeat sequence.

7. In the genome, locate the antirepeat of the tracrRNA (Fig. 1E). Extend the tracrRNA search in the 3′ direction until you find a sequence that resembles a Rho-independent transcription terminator (i.e., a GC-rich hairpin followed by a string of Ts). If there is not an obvious transcriptional terminator, extend the tracrRNA search in the 5′ direction, looking for a similar structure flanked by an A-rich string. If the tracrRNA is encoded in the 3′- to-5′ direction, use the reverse complement of the sequence during further analyses.

   Typically, tracrRNA sequences are at least 50 nt and <150 nt (Chylinski et al. 2013; Briner et al. 2014). Predicting directionality of the antirepeat and tracrRNA can be difficult. The directionality of the CRISPR repeat sequence can help determine the orientation of the tracrRNA. The 5′ end of the CRISPR repeat generally starts at a G that forms a G-U wobble with the tracrRNA. Additionally, the CRISPR repeat will have ~5–7 nt that base-pair with the tracrRNA to form the lower stem, followed by an unpaired segment on the tracrRNA that forms the bulge. When looking for antirepeat sequences, only the upper stem is usually identified because of the long segment of complementary base-pairing between the crRNA repeat and the tracrRNA antirepeat followed by nonpairing nucleotides in the bulge (Briner et al. 2014).

8. Using an RNA folding prediction software like NUPACK, predict the secondary structures of the crRNA and tracrRNA duplex (Fig. 1F). Use a folding algorithm that will allow for G-U base

---

FIGURE 1. (*Continued*) The consensus CRISPR repeat sequence is shown on the top line highlighted in colored boxes. The final CRISPR repeat has four mutated nucleotides at the 3′ end (highlighted); thus, it is likely to be the terminal repeat. (*D*) The alignment output from BLAST local alignment identifies the antirepeat portion of the tracrRNA that forms the upper stem (Briner et al. 2014). The antirepeat covers 1/5 to 2/3 of the entire repeat sequence. The "Query" sequence is the genomic region upstream of the *cas9* gene in *Streptococcus pyogenes*. The "sbjct" (Subject) sequence is the consensus CRISPR repeat sequence from *Streptococcus pyogenes*. The antirepeat detected by the local alignment had 96% identity to a 25-nt stretch of the CRISPR repeat with zero gaps. (*E*) The tracrRNA in the CRISPR locus is identified by locating the antirepeat segment from the local alignment. The antirepeat forms the upper stem (red) portion of the crRNA:tracrRNA duplex. Extending through the CRISPR repeat, there are several unpaired nucleotides that form the bulge (bold), followed by reestablished complementarity to form the lower stem module (blue). Adjacent to the lower stem is the nexus module (underlined) that forms a small hairpin structure in the RNA secondary structure prediction (Briner et al. 2014). (*F*) The predicted tracrRNA search is extended through a Rho-independent transcriptional terminator, such as the final bolded nucleotides in the *Streptococcus pyogenes* tracrRNA sequence. The RNA structure prediction from NUPACK (Zadeh et al. 2011) contains the five functional modules formed when the crRNA and tracrRNA form a duplex molecule (Briner et al. 2014). (*G*) RNA sequencing reads that map to the CRISPR repeat-spacer array can be used to determine the processing boundaries for the crRNAs. The conserved repeats (green arrows) flank variable spacer sequences (gray arrows). The crRNAs can be some of the most highly transcribed small RNAs in the cell. (*H*) The 5′ boundary of the crRNA will be in the spacer sequence as a result of cellular nuclease activity. The 3′ boundary of the crRNA in the CRISPR repeat is matured by RNase III processing activity when the pre-crRNA is complexed with the tracrRNA (Deltcheva et al. 2011; Karvelis et al. 2013). (*I*) The predicted tracrRNA can be confirmed through RNA sequencing analyses. The 5′ end of the tracrRNA is matured by RNase III activity when the tracrRNA is interacting with the pre-crRNA. The 3′ end of tracrRNA either is the transcript terminator or is processed by cellular nucleases. The predicted tracrRNA is often longer than the RNA sequencing-confirmed tracrRNA, as the predicted sequence contains a portion that is removed during crRNA biogenesis (Deltcheva et al. 2011; Karvelis et al. 2013). (*J*) A motif detection program, like WebLogo (Crooks et al. 2004), is used to identify the conserved PAM in the region downstream from the protospacer. The height of each nucleotide correlates to its conservation at each position. (*K*) The table of protospacer hits shows that the spacer sequences match phage and streptococci sequences in publicly available data. The best matches (>90% identity over the entire spacer sequence) can be used to extract flanking regions and predict the PAM.

wobbles. For NUPACK, ensure that tracrRNA and crRNA sequences are entered in the 5′ to 3′ direction. Use the following options (required).

Nucleic acid type	RNA
Number of strand species	2
Maximum complex size	2
Concentration of strand1	1 μM
Concentration of strand2	1 μM

See Troubleshooting.

9. Identify the secondary structures established by Briner et al. (2014), including the upper stem, bulge, lower stem, nexus, and hairpins (Fig. 1F).

    See Troubleshooting.

Confirming crRNA and tracrRNA Boundaries

After in silico determination of the putative crRNA and tracrRNA boundaries, validation of sequences by RNA sequencing is strongly recommended. Steps 12 and 13 can be performed by various programs available through open source or commercial platforms. A program such as FastQC is recommended to assess the quality and adapter content of the reads both before and after processing.

10. Extract RNA using the Direct-zol RNA MiniPrep Kit.

11. Size-select for small RNAs (between 17 and 200 nt) and perform deep sequencing using next-generation sequencing.

    We recommend using the TruSeq Small RNA Sample Preparation Kit followed by HiSeq or MiSeq Illumina sequencing with single-end 150-nt read lengths.

12. After demultiplexing the samples, trim and filter the sequencing reads to remove adapters and poor quality bases. First, remove adapters specific to the type of sequencing performed. Next, trim to remove poor quality bases using an error probability limit of at least 0.01 (Phred 20).

    More stringent trimming to 0.001 (Phred 30) is recommended.

13. After trimming, filter reads to remove sequences shorter than 15 nt, as they can map indiscriminately to the reference sequence.

14. Using a short read alignment algorithm like Bowtie 2, map the trimmed and filtered reads to the reference genome.

    A coverage map of the CRISPR–Cas locus (or individual components thereof) can be used to determine the boundaries of the crRNAs and tracrRNA as depicted in Figure 1H,I.

Designing Single Guide RNA from Chimeric crRNA:tracrRNA

15. Create a chimeric, single guide RNA by linking the 3′ end of the crRNA to the 5′ end of the tracrRNA sequences in the upper stem portion of the crRNA:tracrRNA duplex using an artificial nucleotide tetraloop composed of noncomplementary nucleotides (Jinek et al. 2012).

    • If you confirmed the boundaries of the tracrRNA and crRNAs using RNA sequencing, use the RNase III processing as the artificial linker point (Deltcheva et al. 2011).

    • If you did not confirm the boundaries, join the two molecules between 2 and 6 nt above the bulge in the upper stem.

Predicting PAM Sequences In Silico

16. Identify the protospacer sequence that each spacer was derived from by extracting the spacer sequences from the CRISPR array and searching for homologous sequence in publically available databases (NCBI) (Fig. 1K). Use only protospacers that show 90% identity over the entire spacer length for further analyses. Look for hits in plasmids, phages, and prophage regions of the chromosome.

If using BLASTn to identify protospacers, the following databases are recommended:
- *Nucleotide collection (nr/nt)*
- *Whole-genome shotgun contigs (wgs)*
- *Organism: Enter the genus of your organism (*Streptococcus *[taxid:1301]).*
- *WGS Project: Select metagenomes that would contain your organism.*

When looking for protospacers, manual curation of BLAST results is the key to finding high-quality matches. Ensure that perfect matches are to actual protospacers and are not, in fact, matches to identical spacer sequences in other strains of your select species (Deveau et al. 2008; Mojica et al. 2009; Shah et al. 2013). See Troubleshooting.

17. Extract 10 nucleotide-flanking regions from both edges of the protospacer. If the BLAST result did not cover the entire spacer sequence, extend the protospacer region to cover the entire spacer length and then extract the flanking regions.

    Typically, PAMs flank the 3′ end of the protospacers in Type II systems.

18. Using a motif-identifying program like WebLogo, identify the conservation of nucleotides at each position within the flanking regions (Fig. 1J). If using WebLogo, under the Advanced Logo Options, select DNA/RNA for Sequence Type.

    Ensure that you have the correct directionality of the spacer and protospacer sequences.

TROUBLESHOOTING

Problem (Step 5): An antirepeat cannot be found in any of the regions searched.

Solution: First, try broadening your search window. Search within the flanking 1000 nt and within the *cas* genes. Additionally, try using a less stringent match/mismatch and gap cost matrix that will not penalize mismatches as harshly. Finally, if the genome is in draft status, the rest of the CRISPR locus and tracrRNA may be on a separate contig. Search the rest of the genome for additional parts of the repeat-spacer array that did not assemble well, and search the flanking regions for a tracrRNA.

Problem (Steps 8 and 9): The RNA folding prediction does not form a crRNA:tracrRNA duplex or does not contain the five modules established by Briner et al. (2014) and one RNA strand forms secondary structures with itself.

Solution: Ensure that both sequences are entered in the 5′-to-3′ direction. If the program does not predict any strand1–strand2 interactions, try entering the reverse complement of the crRNA sequence. If the folding prediction does not contain the modules established by Briner et al. (2014), try adjusting the length of the crRNA and tracrRNA sequences to decrease the amount of self-binding from the RNA strands. If this still does not yield a crRNA:tracrRNA duplex, you may not have the correct directionality of the tracrRNA. Try extending the tracrRNA search to the opposite side of the identified antirepeat (i.e., if you extended the tracrRNA on the 3′ side but did not form a crRNA:tracrRNA complex, extend the tracrRNA on the 5′ side).

Problem (Step 16): There were not enough positive, high-quality protospacer hits to infer the PAM sequence.

Solution: Oftentimes, the same CRISPR–Cas system can be found in other strains of the same organism. Identify other strains that contain an identical system (with identical repeats and highly similar Cas proteins as determined through an alignment). Use the spacer sequences from these strains to complement your PAM search.

DISCUSSION

The activity of predicted Type II CRISPR–Cas system elements should be validated through further analyses. In vivo interference testing of native systems in their bacterial backgrounds demonstrates the

A.E. Briner et al.

system's native ability to target foreign DNA. Additionally, biochemical in vitro testing of Cas9 and guide sequences can help confirm dual nickase activity (for genesis of double-stranded breaks) and possibly determine efficiency. Finally, editing with engineered CRISPR–Cas systems in vivo can help determine efficiency and specificity (Cong et al. 2013; Mali et al. 2013; Doudna and Charpentier 2014; Hsu et al. 2014; Sander and Joung 2014). Although off-target effects have been exaggerated, they must be somewhat quantified before in vivo implementation. Regarding efficiency, several items should be considered when designing guide RNAs and selecting targets, including avoiding PAM redundancies in the spacer sequence (especially the seed sequence), avoiding homopolymeric runs in target sequences, avoiding hairpin-forming sequences within the spacer, and ensuring the spacer sequence does not compromise the overall guide structure. For concurrent use of various Cas9 proteins and their corresponding guides, it is critical to ensure that these systems are orthogonal and do not cross-react (Cong et al. 2013; Esvelt et al. 2013; Fonfara et al. 2014). Accordingly, users should ensure that various guides with incompatible structures, nexus sequences, and PAMs are selected appropriately (Briner et al. 2014). We anticipate the development of novel CRISPR–Cas9 systems will expand the toolbox and open new avenues for diverse genetic engineering applications including genome editing, transcriptional control, imaging, epigenetics, and remodeling of chromosomes.

REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. 2007. CRISPR Recognition Tool (CRT): A tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8:** 209.

Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, Barrangou R. 2014. Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* **53:** 333–339.

Chylinski K, Le Rhun A, Charpentier E. 2013. The tracrRNA and Cas9 families of type II CRISPR–Cas immunity systems. *RNA Biol* **10:** 726–737.

Chylinkski K, Makarova KS, Charpentier E, Koonin EV. 2014. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* **42:** 6091–6105.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:** 819–823.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* **14:** 1188–1190.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471:** 602–607.

Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190:** 1390–1400.

Doudna JA, Charpentier E. 2014. Genome editing. The new frontier of genome engineering with CRISPR–Cas9. *Science* **346:** 1258096.

Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10:** 1116–1121.

Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lécrivain AL, Bzdrenga J, Koonin EV, Charpentier E. 2014. Phylogeny of Cas9 determines func-

tional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR–Cas systems. *Nucleic Acids Res* **42:** 2577–2590.

Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8:** 172.

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR–Cas9 for genome engineering. *Cell* **157:** 1262–1278.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337:** 816–821.

Karvelis T, Gasiunas G, Miksys A, Barrangou R, Horvath P, Siksnys V. 2013. crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol* **10:** 841–851.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, et al. 2011. Evolution and classification of the CRISPR–Cas systems. *Nat Rev Microbiol* **9:** 467–477.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339:** 823–826.

Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155:** 733–740.

Sander JD, Joung JK. 2014. CRISPR–Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32:** 347–355.

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* **39:** 9275–9282.

Shah SA, Erdmann S, Mojica FJ, Garrett RA. 2013. Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biol* **10:** 891–899.

Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, Dirks RM, Pierce NA. 2011. NUPACK: Analysis and design of nucleic acid systems. *J Comput Chem* **32:** 170–173.

Cite this protocol as *Cold Spring Harb Protoc;* doi:10.1101/pdb.prot086785

Cold Spring Harbor Protocols

Prediction and Validation of Native and Engineered Cas9 Guide Sequences

Alexandra E. Briner, Emily D. Henriksen and Rodolphe Barrangou

Email Alerting Service	Receive free email alerts when new articles cite this article - **click here.**
Subject Categories	Browse articles on similar topics from *Cold Spring Harbor Protocols*. Bioinformatics/Genomics, general (181 articles) Computational Biology (100 articles) Genome Analysis (165 articles) Molecular Biology, general (1139 articles) Prediction of RNA Secondary Structure (2 articles) RNA (263 articles) RNA, general (223 articles) Sequence Database Searching (38 articles)