



CRISPy-web: An online resource to design sgRNAs for CRISPR applications

Kai Blin^a, Lasse Ebdrup Pedersen^a, Tilmann Weber^{a,*}, Sang Yup Lee^{a,b,**}

^a The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Alle 6, 2970 Hørsholm, Denmark

^b Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea

ARTICLE INFO

Article history:

Received 17 October 2015

Received in revised form 5 January 2016

Accepted 10 January 2016

Available online

Keywords:

CRISPR

Cas9

sgRNA

Guide RNA

Genome editing

Genome engineering

ABSTRACT

CRISPR/Cas9-based genome editing has been one of the major achievements of molecular biology, allowing the targeted engineering of a wide range of genomes. The system originally evolved in prokaryotes as an adaptive immune system against bacteriophage infections. It now sees widespread application in genome engineering workflows, especially using the *Streptococcus pyogenes* endonuclease Cas9. To utilize Cas9, so-called single guide RNAs (sgRNAs) need to be designed for each target gene. While there are many tools available to design sgRNAs for the popular model organisms, only few tools that allow designing sgRNAs for non-model organisms exist. Here, we present CRISPy-web (<http://crispy.secondarymetabolites.org/>), an easy to use web tool based on CRISPy to design sgRNAs for any user-provided microbial genome. CRISPy-web allows researchers to interactively select a region of their genome of interest to scan for possible sgRNAs. After checks for potential off-target matches, the resulting sgRNA sequences are displayed graphically and can be exported to text files. All steps and information are accessible from a web browser without the requirement to install and use command line scripts.

© 2016 The authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

“Clustered regularly interspaced short palindromic repeats” (CRISPR) and their associated RNA-guided endonucleases are bacterial adaptive immune systems protecting the bacteria from infections with bacteriophages.¹ The biotechnological application of this system is currently revolutionizing molecular biology and provides new opportunities for synthetic biology applications.² The CRISPR system allows targeted genome engineering of bacteria,^{3–5} and also eukaryotes including yeast,^{6,7} plants,⁸ human cell lines⁹ and many more. Although several alternative CRISPR systems have recently been described (e.g., References 10–13), most CRISPR systems

for genome engineering are based on the *Streptococcus pyogenes* type II CRISPR/Cas9 system. Cas9 is an RNA-guided endonuclease, which introduces double strand DNA breaks at positions that are complementary to a crRNA sequence that binds to Cas9 in a duplex with a tracrRNA. It has also been demonstrated that Cas9 accepts artificially fused crRNA:tracrRNA-hybrids termed “single guide RNAs” (sgRNAs) to direct it to the target cut sites.¹⁴ Thus CRISPR can be used as an “*in vivo* programmable restriction enzyme,” which cuts the target DNA within an exactly defined target sequence determined by the sgRNA.

One essential step in the design of CRISPR experiments is to identify suitable sgRNA sequences within the target gene(s), which have to fulfill certain conditions: (i) the 20 bp target sequence has to be directly upstream of a “protospacer adjacent motif” (PAM). For the *S. pyogenes* PAM, this motif is “NGG”, (ii) it is desirable that the sequence of the target motif is unique within the genome of the organisms to prevent off-target activity, *i.e.* the cleavage of the chromosome at wrong positions.

Many programs and web-servers exist to assist biologists in finding such CRISPR/Cas9 target sites. However, most of these tools (e.g. CCTop,¹⁵ CHOPCHOP,¹⁶ CRISPR Design,¹⁷ WU-CRISPR,¹⁸ WGE CRISPR Finder,¹⁹ and CRISPy CHO²⁰) are limited to a narrow set of – mostly eukaryotic – target genomes of model organisms. Only very few tools (e.g. sgRNACas9²¹) allow running Cas9 target searches on

* Corresponding author. The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Alle 6, 2970 Hørsholm, Denmark. Tel.: +45 24 89 61 32; fax: +45 45 25 80 01.

E-mail address: tiwe@biosustain.dtu.dk (T. Weber).

** Corresponding author. Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea. Tel.: +82 42 350 3930; fax: +82 42 350 8800.


E-mail address: leesy@kaist.ac.kr (S.Y. Lee).

Peer review under responsibility of KeAi Communications Co., Ltd.

A. Overview sgRNAs in *Streptomyces coelicolor* A3(2) actinorhodin biosynthetic gene cluster

CRISPy web [Go back](#) [help](#) [about](#)

Region: Cluster 11 (5509800 - 5552424)



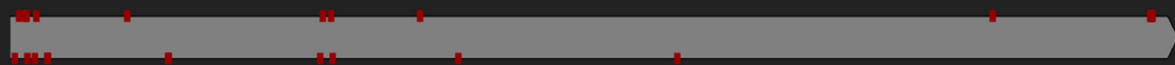
Start	End	Strand	Sequence	PAM	Matches	3bp mismatches	Download
21368	21391	1	CTGCGCGACGCCGAGACCGC	GGG	0	1	Download
1133	1156	-1	CGCCACTTCCCTCCCGGTG	GGG	0	1	Download
36685	36708	-1	CGGTGTCCGGTTGCGGTGA	GGG	0	1	Download
36730	36753	1	CCACGCTCGGAGCGTAAGCG	CGG	0	1	Download
41266	41289	1	AGGTCAGGCGGAAGCCGAGC	GGG	0	1	Download
1860	1883	1	ACCGCGATGACGGCCCTGAG	CGG	0	1	Download
5021	5044	-1	CAGCGCCCTGCGCACCCACGA	CGG	0	1	Download
381	404	-1	TAGGCGTAGGTGGCGGCCCG	CGG	0	1	Download

SCO5087
Show results for this gene only

B. Zoom in on gene *SCO5087* (*actIORF1*; KS α)

CRISPy web [Go back](#) [help](#) [about](#)

Region: (5529800 - 5531204)



Start	End	Strand	Sequence	PAM	2bp mismatches	3bp mismatches	Download
1368	1391	1	CTGCGCGACGCCGAGACCGC	GGG	0	1	Download
1178	1201	1	GATCGGCTCGCTGGAGATCG	CGG	1	1	Download
187	210	-1	GAGCAGTCCAGAACTGCC	GGG	1	1	Download
1367	1390	1	GCTGCGCGACGCCGAGACCG	CGG	1	2	Download
536	559	-1	CCAGGTAGTCGAACATGTGC	CGG	1	2	Download
138	161	1	AGAGTCGTCATCACGGGCGT	CGG	1	2	Download
383	406	1	CGCCGCCAGCGGGCTCGACC	CGG	1	6	Download
16	39	1	CGCCGCTGGATCCGGCATCG	AGG	2	5	Download
373	396	1	GGGAGGCCTTCGCCGCCAGC	GGG	2	7	Download

C. List of selected sgRNAs for export

CRISPy web [Go back](#) [help](#) [about](#)

[Download CSV file](#)

ID	Start	End	Strand	Sequence	PAM	2bp mismatches	3bp mismatches
CY00000001	1368	1391	1	CTGCGCGACGCCGAGACCGC	GGG	0	1
CY00000109	1178	1201	1	GATCGGCTCGCTGGAGATCG	CGG	1	1
CY00000147	187	210	-1	GAGCAGTCCAGAACTGCC	GGG	1	1
CY00000057	1367	1390	1	GCTGCGCGACGCCGAGACCG	CGG	1	2

Fig. 1. Example output of a CRISPy-web run to identify sgRNAs in the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor* A3(2) (NCBI GenBank ID: NC_003888.3). (A) Overview of sgRNAs identified in the complete actinorhodin gene cluster. (B) Zoom view of *SCO5087* (*actIORF1*). (C) List of sgRNAs selected for export.

user-specified genomes. Unfortunately, to the best of our knowledge, none of these generally applicable tools are provided as user-friendly web services, which can be used by non-computer scientists.

Here, we therefore present a web server implementation of the sgRNA prediction software CRISPy (<http://crispy.secondarymetabolites.org/>) that was originally designed for CRISPR-based engineering of Chinese hamster ovary (CHO) cells.²⁰ It is implemented as a stand-alone web application for Cas9 target prediction that can design sgRNA for any gene/region of interest in user-submitted genome data. In addition, CRISPy-web tightly connects with the popular natural product prediction tool antiSMASH^{22–24} to assist researchers working on microbial secondary metabolites in leveraging the CRISPR/Cas9 system. Using antiSMASH or working on natural products of course is not a prerequisite for using CRISPy-web.

2. Design and implementation

CRISPy-web consists of multiple decoupled service components tied together by a message queue. The main services are the CRISPy worker service, the web application programming interface (API) server and the web user interface (UI). Sequences to be scanned for Cas9 target sites can be uploaded directly (in GenBank format) or transferred from antiSMASH^{22–24} by specifying the antiSMASH job ID. In the next version of antiSMASH, a direct link from antiSMASH will also be implemented.

2.1. CRISPy worker service

Two steps are performed by the worker service: preparing a sequence for analysis and then scanning and ranking Cas9 target sites. In the preparation step, the input files are parsed and – if the files were transferred from antiSMASH – annotated with secondary metabolite gene cluster locations using Biopython.²⁵

Based on the CRISPy algorithm initially used in Ronda et al.,²⁰ the scanning step searches for the 'NGG' PAM sequence on both the forward and the reverse strands of the selected region of interest, counting the number of direct off-target hits for sequences with the same 13 bp core sequence upstream of the PAM. It then calculates the number of off-target hits containing 0, 1, 2, or 3 bp mismatches on the whole genome for the 13 bp sequence by considering all possible combinations.

After collecting all possible Cas9 targets, sites that have direct off-target matches, sites with off-target matches with a single base-pair mutation, or sites with more than ten off-target matches with a two base-pair mutation are discarded. All these hits are stored in a GenBank file. To further reduce the number of sites displayed on the user interface, the targets are now sorted by 2 bp mismatch off-targets, 3 bp mismatch off-targets and coordinates on the genome, and only the first 1000 targets are displayed to the user.

2.2. Web application programming interface server

The web API server provides a Representational State Transfer (REST) interface²⁶ that both the CRISPy-web UI and other tools can use to submit jobs to and obtain results from the CRISPy worker service. Data are transferred using the JavaScript Object Notation (JSON) format. The API server is implemented in Python using the Flask web framework and a Redis server for the message queue.

2.3. Web user interface

Built around the AngularJS framework and D3 for cluster visualization, the web UI is the primary method of interacting with CRISPy-web. Communicating with the web API server via asyn-

chronous JavaScript (AJAX) calls, the web UI leads the user through the process of identifying Cas9 target sites in regions of interest.

3. Example usage: scanning for sgRNAs in the actinorhodin gene cluster of *Streptomyces coelicolor*

To scan for sgRNAs in *S. coelicolor*, the user first needs to upload the corresponding genome file in GenBank format. Such a file can easily be obtained from NCBI with the RefSeq ID NC_003888. After a short preliminary scan of the genome, CRISPy-web displays the target selection page. The target selection page gives a short overview of the uploaded genome, such as organism, description, sequence length and the number of annotated genes. Targets can be selected as a nucleotide range or by providing a locus tag, gene name, or protein-ID if annotated in the genome. If available, annotated gene clusters will also be shown in a list and can be selected as regions to scan. On *S. coelicolor*, the actinorhodin cluster is located from 5,509,800 to 5,552,424 bp on the genome, so that is the range to select as target region. After a click on “Find targets”, the scan for sgRNAs starts.

Depending on the size of the selected region and the overall genome size, this sgRNA scan will take several minutes. After the scan is completed, a page graphically displays the scanned region (Fig. 1A). Genes are shown as gray arrows, sgRNAs on the forward strand are displayed as red boxes above the genes. sgRNAs on the reverse strand are displayed below the genes. A table below the graphical overview gives details on the identified sgRNAs, ranked by the least number of potential off-target matches. Hovering over the table row of an sgRNA hit highlights the corresponding sgRNA box in the graphics, so you can quickly locate the hit on the region.

As there are a lot of hits for this region, it is possible to zoom in on a particular gene of interest by clicking on the gene arrow and selecting “show results for this gene only”. Because the whole region already was scanned, this operation is instantaneous (Fig. 1B). On both the overview and on the zoomed view, sgRNAs you want to save for download can be added to the download basket by clicking the corresponding table row. The download basket button on the upper right will track the number of selected sgRNAs.

A click on the download basket icon will display the download overview page (Fig. 1C). There, the selected sgRNAs can be verified again. A click on the “Download CSV file” button will start the download of all selected sgRNAs in comma-separated value format. Files in this format can easily be used in a spreadsheet application, processed for downstream computational analysis or oligonucleotide orders.

Acknowledgment

This work was funded by a grant of the Novo Nordisk Foundation.

References

- Horvath P, Barrangou R. CRISPR/Cas, the immune system of *Bacteria* and *Archaea*. *Science* 2010;**327**:167–70.
- Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014;**346**:1258096.
- Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 2013;**31**:233–9.
- Tong Y, Charusanti P, Zhang L, Weber T, Lee SY. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth Biol* 2015;**4**:1020–9.
- Cobb RE, Wang Y, Zhao H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth Biol* 2015;**4**:723–8.
- Ronda C, Maury J, Jakociunas T, Jacobsen SA, Germann SM, Harrison SJ, et al. CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb Cell Fact* 2015;**14**:97.
- Jakociunas T, Bonde I, Herrgard M, Harrison SJ, Kristensen M, Pedersen LE, et al. Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metab Eng* 2015;**28**:213–22.

8. Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 2013;**31**:686–8.
9. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014;**343**:84–7.
10. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015;**163**(3):759–71.
11. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* 2013;**10**:1116–21.
12. Friedland AE, Baral R, Singhal P, Loveluck K, Shen S, Sanchez M, et al. Characterization of *Staphylococcus aureus* Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol* 2015;**16**:257.
13. Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, et al. Discovery and functional characterization of diverse class 2 CRISPR-cas systems. *Mol Cell* 2015;**60**:385–97.
14. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;**337**:816–21.
15. Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE* 2015;**10**:e0124633.
16. Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* 2014;**42**:W401–7.
17. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013;**31**:827–32.
18. Wong N, Liu W, Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol* 2015;doi:10.1101/026971.
19. Hodgkins A, Farne A, Perera S, Grego T, Parry-Smith DJ, Skarnes WC, et al. WGE: a CRISPR database for genome engineering. *Bioinformatics* 2015;**31**:3078–80.
20. Ronda C, Pedersen LE, Hansen HG, Kallehauge TB, Betenbaugh MJ, Nielsen AT, et al. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol Bioeng* 2014;**111**:1604–16.
21. Xie S, Shen B, Zhang C, Huang X, Zhang Y. sgRNACas9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS ONE* 2014;**9**:e100448.
22. Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R, et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 2015;**43**:W237–43.
23. Blin K, Medema MH, Kazempour D, Fischbach M, Breitling R, Takano E, et al. antiSMASH 2.0 – a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* 2013;**41**:W204–12.
24. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* 2011;**39**:W339–46.
25. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;**25**:1422–3.
26. Fielding RT, Taylor RN. Principled design of the modern Web architecture. *ACM T Internet Techn* 2002;**2**:115–50.