UDC 57.577.2

CRISPR/CAS SYSTEM TO COMBAT PLANT VIRUS

Zharylkassyn G., Iksat N., Omarov R.
gauhar.zharylkassyn@gmail.com
Student of Faculty of Natural Sciences at L.N. Gumilyov Eurasian National University,
Nur-Sultan, Kazakhstan
Scientific Supervisor-Nurgul Iksat

Abstract. CRISPR-Cas and CRISPR-associated systems (CRISPR-Cas) are key immune mechanisms that allow prokaryotic species to ward off RNA and DNA viruses. CRISPR/Cas9 has broad biotechnology and basic research applications, as well as being widely used across eukaryotic species for genome engineering and functional gene analysis. Newly developed CRISPR/Cas13 systems target RNA rather than DNA and have the potential to be used for gene engineering and to combat RNA viruses. This paper provides an overview of the CRISPR/Cas system and discusses the guided RNA. We then describe different types of programs used to detect secondary structure of guided RNA in CRISPR/CAS13.

Introduction. CRISPR/Cas is an immune defense system that is currently used in biotechnology, particularly CRISPR/Cas9 for DNA targeting. CRISPR/Cas technology that targets RNA, on the other hand, is still in the works. The discovery of the type VI CRISPR / Cas system enables the discovery of RNA-guided RNA targeting. CRISPR/Cas13 is a ssRNA-targeting system with enormous potential for RNA virus targeting as well as transcript-level interference.

During the evolution of archaea and bacteria, CRISPR (clustered regularly interspaced short polyadenylic repeats)-Cas (CRISPR-associated protein) were essential for adaptive phage immunity mechanisms. The CRISPR/Cas9 and other CRISPR/Cas systems work by recognizing and binding DNA to RNA for specific nucleic acid cleavage at any desired target site at low cost. In addition to its first applications for genome editing in plants ,CRISPR-Cas is now used for genome editing in a variety of crop species, introducing traits of great value into many of them [1].

They are classified into two major classes and six types, divided further into several subtypes, as a result of their complex classification .A class 1 system (which includes types I, III, and IV) utilizes multisubunit effector complexes, whereas a class 2 system (which includes types II, V, and VI) uses a single, multidomain protein [2].

The class 2 type II CRISPR/Cas9 system of *Streptococcus pyogenes* was the first and most widely used gene-editing tool, and its mechanism is reasonably well understood [3]. Activation of this system requires an additional small RNA, a transactivating

crRNA (tracrRNA), whichis involved in the maturation of crRNA and is transcribed separately from a genomic locus that exists upstream of the CRISPR locus. In tracrRNA, the bases pair with precrRNA repeats, generating dsRNA, which is then cleaved by RNase III into mature crRNA. Further trimming is performed by an unknown enzyme at the 5 end of this crRNA to produce the mature copy [4].

By binding its tracrRNA portion to the Cas9 endonuclease, the mature crRNA:tracrRNA is capable of targeting foreign DNA through a single spacer sequence that complements the foreign DNA in the crRNA portion. It will not attack the foreign nucleic acid unless it contains adjacent protospacer motifs (PAMs) immediately downstream [1].

Relevance of guided RNA in CRISPR/CAS system.

Native CRISPR/Cas9 gRNAs are made up of two RNA molecules: one encoding a sequence that is complementary to the target and the other influencing the interaction between the Cas effector protein and the crRNA [1]. To make CRISPR Cas a programmable tool for DNA editing, the creation of a single gRNA was an important first step. A single gRNA directs enzymatic activity of the Cas protein to a selected DNA target defined by the complementary sequence. In some cases, like Cas13, a single crRNA molecule is used by Class 2 effectors. Natural gRNAs contain a direct repeat (DR) stem loop to mediate interactions with Cas proteins, and a spacer to determine target selectivity. It is sometimes used interchangeably in the literature to refer to gRNA, sgRNA, crRNA, and spacer [5].

Before adapting the system to any application, one must know the details of gRNA design for a given Cas ortholog. At the very least, this would include the spacer length requirements, the sequence, and the structure details of the DR. More detailed information regarding the specificity of the spacer would be ideal. As an example, Cas9 ortholog spacers typically consist of a seed region (usually 8–10 nt at the 3′ end) that tolerates low or no mismatches. It is used for the first target interrogation step [5,6].

Regardless of how well understood the gRNA architecture is, selecting the target-specific spacer sequence may still significantly impact the detection system's performance. There have been several bioinformatics tools developed to help select spacer sequences for Cas9 systems that target DNA. Other toolkits have been developed for non-Cas9 systems and often consider both the gRNA and target RNA secondary structure. For instance, Wessels et al. designed a tool for predicting Cas13d gRNAs for all protein-coding transcripts in the human genome (cas13design.nygenome.org) [7,8].

Nevertheless, a strong consensus exists in the literature that at least two to five spacer sequences must be tested experimentally to determine the most efficient guide for a particular target and working conditions (with other variables including the buffer, temperature, or target binding kinetics) [9].

Tools used for detecting secondary structure of gRNA in CRISPR/Cas13 system.

The very first Cas13 proteins that were characterized in bacteria required a sequence constraint, the PFS, to ensure target cleavage efficiency. This includes *Leptotrichia shahii* Cas13a (LshCas13a), *Bergeyella zoohelcum* Cas13b (BzoCas13b), and *Prevotella buccae* Cas13b (PspCas13b) [10].However, further investigation of PspCas13b in mammalian and plant and other Cas13 orthologs showed high target RNA degradation efficiencies even in the absence of PFS [11].While this gives researchers some flexibility over target site selection, it is necessary to consider the secondary structure of target transcripts, since this negatively affected knock-down efficiency[12].

Nhan Huynh in his work with Drosophila to assess secondary structures used two independent online tools, namely RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) and RNA structure (https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html) [11,13,14].

RNAfold.

RNAfold predicts secondary structures of single stranded RNA or DNA sequences online. As of now, the partition function calculation limit is 7,500 nt and the minimum free energy prediction limit is 10,000 nt. RNAfold computes the minimal free energy (MFE) and backtraces an optimal secondary structure to find an optimal structure. RNAfold can also compute the partition function, the matrix of base pairing probabilities, and the centroid structure using McCaskill's algorithm [15]. RNAfold writes the folded structure and folding energy to the standard output stream as strings. Using the -p option, a PostScript file containing the base pairing probability matrix is also created. In nature, circular RNA sequences are rare, and they are rarely found in practical applications. The circ option handles this case as a post-processing for forward recursion and a pre-processing for backward recursion without compromising the performance of the algorithms for folding linear RNAs [16]. Individual positions can be forced to be paired, unpaired, or paired with specific partners by imposing constraints on the folding algorithms.

RNA structure.

RNAstructure first appeared in 1998 as a package for predicting secondary structures [17]. They offered a method of predicting the lowest free energy structure, as well as a set of low free energy structures .With OligoWalk, bimolecular folding and hybridization thermodynamics were included.

Some of the recent extensions include PARTS, which calculates the ratio of secondary structures common to two sequences and performs stochastic sampling of common structures. In addition, a method for removing pseudoknots that leaves a pseudoknot-free structure with the lowest energy. There are several tutorials on RNA structure [18]. Thus, there are four different prediction and analysis algorithms in the Predict a Secondary Structure server: computing a partition function, predicting a minimum free energy structure, identifying structures with the highest expected accuracy, and predicting pseudoknot structures. Using either DNA or RNA, the server creates a highly probable set of secondary structures, starting with the lowest free energy structure, and including others with varying probabilities of correctness [14].

ViennaRNA package.

Most of the time, structural-seq probes are not available for the RNA in question. If this is the case, computational structure prediction can be used. It has been found that using the ViennaRNA package, RNA structure, and Mfold, which incorporated dynamic programming algorithms, it is possible to derive RNA secondary structure models that are as close to those in vivo as possible[19]. Nevertheless, it is useful to predict RNA secondary structure using complementary algorithms, as no program can completely automate the structures in the cell.

RNA bioinformatics community has used ViennaRNA Package for almost two decades. There are a variety of widely-used software tools and data analysis pipelines built on top of this foundation, either by using the interactive programs or by directly interacting with RNAlib. In machine learning classification, a variety of secondary structure characteristics, such as Gibbs free energy (G), minimal free energy (MFE), or ensemble diversity, or probabilities of MFE structures in an ensemble, have been widely used, such as in detecting microRNA precursors.

CRISPR-RT.

The CRISPR RNA-Targeting (CRISPR-RT) tool helps biologists design crRNAs that target CRISPR/Cas13. Researchers using CRISPR-based RNA editing will be able to set up a wide range of parameters, including the size of the complementarity region and seed region for target RNA, as well as how many mismatches or gaps are tolerated by off targets to maximize flexibility.

Using the input RNA sequence, CRISPR-RT will find potential targets, and then apply a rigorous alignment algorithm to search for both on- and off-target sites within the reference transcriptome. Results are presented using highly interactive graphical interfaces. In order to choose the target candidate with a minimal effect of off targets, users can rank target candidates according to the total number of target sites in the reference transcriptome. Through JBrowse, users can also visualize the annotation of annotated genome and transcript features to validate the on- and off-target sites [20].

The CRISPR-RT system consists of a number of web interfaces and a backend pipeline. PHP and JavaScript are used to implement web interfaces, which accept user inputs and display results interactively(Input data for the backend pipeline is processed by Perl code, which generates multiple result files based on user input [20].

When you set up the proper parameters and click the button, all the PHP code parameters are sent to the main Perl script, which executes specific Perl scripts or commands to perform specific functions (figure1). Abudayyeh et al.'s research was used to select the default parameters [20].

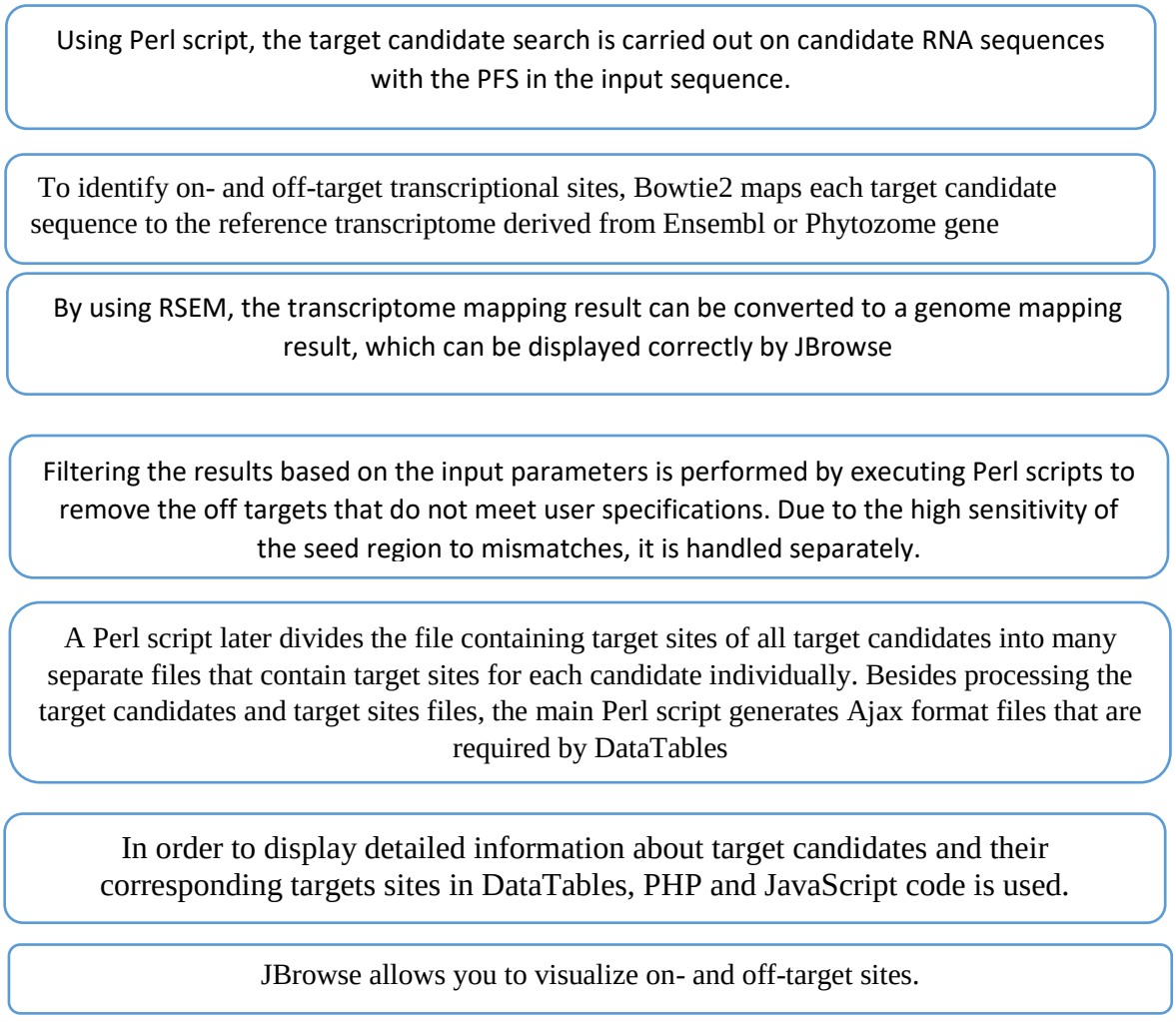

Figure 1 - The CRISPR-RT implementation steps.

Conclusion.

Increasing use of CRISPR technology for genome engineering and related research applications has led to the need for algorithms and online tools that facilitate the prediction of guide RNAs (gRNAs) secondary structure. Free energy minimization has been the most popular method for predicting the secondary structure of RNA for decades. The model is based on a set of

empirical parameters that are derived from experiments using a nearest neighbor model. The prediction is now made using several tools referred to in the review. It should be noted that some of these algorithms not only consider free minimum energy, but also compute the partition function, or the matrix of base pairing probabilities.

Funding. This work was carried out within the framework of the project № AP09258746 "Regulation of CRISPR/Cas13 gene editing system by viral protein expression to confer antiviral plant resistance".

References:

[1] Koonin E V, Makarova K S and Zhang F 2017 Diversity, classification and evolution of CRISPR-Cas systems *Current Opinion in Microbiology* **37** 67–78

[2] Chylinski K, Makarova K S, Charpentier E and Koonin E V 2014 Classification and evolution of type II CRISPR-Cas systems *Nucleic Acids Research* **42** 6091–105

[3] Makarova K S, Wolf Y I, Alkhnbashi O S, Costa F, Shah S A, Saunders S J, Barrangou R, Brouns S J J, Charpentier E, Haft D H, Horvath P, Moineau S, Mojica F J M, Terns R M, Terns M P, White M F, Yakunin A F, Garrett R A, van der Oost J, Backofen R and Koonin E V 2015 An updated evolutionary classification of CRISPR–Cas systems *Nat Rev Microbiol* **13** 722–36

[4] Yan W X, Hunnewell P, Alfonse L E, Carte J M, Keston-Smith E, Sothiselvam S, Garrity A J, Chong S, Makarova K S, Koonin E V, Cheng D R and Scott D A 2019 Functionally diverse type V CRISPR-Cas systems *Science* **363** 88–91

[5] Díaz-Galicia E, Grünberg R and Arold S T 2022 How to Find the Right RNA-Sensing CRISPR-Cas System for an In Vitro Application *Biosensors* **12** 53

[6] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J A and Charpentier E 2012 A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity *Science* **337** 816–21

[7] Xiang X, Corsi G I, Anthon C, Qu K, Pan X, Liang X, Han P, Dong Z, Liu L, Zhong J, Ma T, Wang J, Zhang X, Jiang H, Xu F, Liu X, Xu X, Wang J, Yang H, Bolund L, Church G M, Lin L, Gorodkin J and Luo Y 2021 Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning *Nat Commun* **12** 3238

[8] Anon Massively parallel Cas13 screens reveal principles for guide RNA design | Nature Biotechnology

[9] Fozouni P, Son S, Derby M D de L, Knott G J, Gray C N, D'Ambrosio M V, Zhao C, Switz N A, Kumar G R, Stephens S I, Boehm D, Tsou C-L, Shu J, Bhuiya A, Armstrong M, Harris A R, Chen P-Y, Osterloh J M, Meyer-Franke A, Joehnk B, Walcott K, Sil A, Langelier C, Pollard K S, Crawford E D, Puschnik A S, Phelps M, Kistler A, DeRisi J L, Doudna J A, Fletcher D A and Ott M 2021 Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy *Cell* **184** 323-333.e9

[10] Cox D B T, Gootenberg J S, Abudayyeh O O, Franklin B, Kellner M J, Joung J and Zhang F 2017 RNA editing with CRISPR-Cas13 *Science* **358** 1019–27

[11] Huynh N, Depner N, Larson R and King-Jones K 2020 A versatile toolkit for CRISPR-Cas13-based RNA manipulation in Drosophila *Genome Biology* **21** 279

[12] Abudayyeh O O, Gootenberg J S, Essletzbichler P, Han S, Joung J, Belanto J J, Verdine V, Cox D B T, Kellner M J, Regev A, Lander E S, Voytas D F, Ting A Y and Zhang F 2017 RNA targeting with CRISPR–Cas13 *Nature* **550** 280–4

[13] Gruber A R, Lorenz R, Bernhart S H, Neuböck R and Hofacker I L 2008 The Vienna RNA websuite *Nucleic Acids Res* **36** W70-74

[14] Anon RNA Secondary Structure Analysis Using RNAstructure - Mathews - 2014 - Current Protocols in Bioinformatics - Wiley Online Library

[15] McCaskill J S 1990 The equilibrium partition function and base pair binding probabilities for RNA secondary structure *Biopolymers* **29** 1105–19

[16] Hofacker I L and Stadler P F 2006 Memory efficient folding algorithms for circular RNA secondary structures *Bioinformatics* **22** 1172–6
[17] Zuker M 1989 On finding all suboptimal foldings of an RNA molecule *Science* **244** 48–52
[18] Smit S, Rother K, Heringa J and Knight R 2008 From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal *RNA* **14** 410–6
[19] Bandaru S, Tsuji M H, Shimizu Y, Usami K, Lee S, Takei N K, Yoshitome K, Nishimura Y, Otsuki T and Ito T 2020 Structure-based design of gRNA for Cas13 *Sci Rep* **10** 11610
[20] Zhu H, Richmond E and Liang C 2018 CRISPR-RT: a web application for designing CRISPR-C2c2 crRNA with improved target specificity *Bioinformatics* **34** 117–9