Discovery and Biochemical Characterization of Cas13b, a Type VI-B CRISPR-Associated RNA-Guided RNase

by

Neena Kenton Pyzocha

B.S. Biology, Rensselaer Polytechnic Institute (2011)

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

©2017 Massachusetts Institute of Technology. All rights reserved.
Discovery and Biochemical Characterization of Cas13b, a Type VI-B CRISPR-Associated RNA-Guided RNase

by

Neena Pyzocha

Submitted to the Department of Biology, on May 18, 2017 in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

Targeted methods to perturb nucleic acid sequences or epigenetic modifications enable better understanding of nucleic acid function or dysfunction in disease. Particular CRISPR-associated (Cas) proteins are used for targeted DNA editing because they are easy to use, precise, and specific. CRISPR-Cas systems function in microbes as adaptive immune systems and are composed of endonuclease protein(s) with targeting guide RNAs that together provide interference against foreign nucleic acids. CRISPR-Cas systems exhibit wide diversity, with the ability to target DNA or RNA, and some large single effector Class 2 Cas proteins have been harnessed for genome engineering. There is an interest to find and characterize additional Class 2 Cas proteins, especially ones that target RNA, to enable further targeting capabilities. A large computational database search was conducted to find a large uncharacterized protein within 10 kilobases of a CRISPR array and from this search cas13b was identified. The cas13b loci may contain the genes csx27 or csx28 and one or two CRISPR arrays with canonical or long direct repeats. Experimentally, purified Cas13b processes its own CRISPR array(s) into crRNA(s) and Cas13b with either crRNA architecture targets single stranded RNA cleavage using its HEPN domains. SsRNA cleavage occurs at pyrimidine residues and is constrained by 5’ and 3’ protospacer-flanking sequences on the target. Cas13b also displays the collateral effect, a non-specific cleavage of ssRNAs after targeted cleavage. Cas13b with mutated HEPN domains lose ssRNA cleavage activity but maintain strong, targeted binding capacity. Genetically, Cas13b-mediated RNA interference occurs with Cas13b alone and is repressed by Csx27 or enhanced by Csx28. Cas13b is characterized as a Type VI-B CRISPR-Cas system and represents the second Class 2 Cas protein to target RNA, the other being Cas13a (C2c2). Future studies of Cas13b are warranted to better understand its functional mechanisms, specificity, role of small proteins, and acquisition. Cas13b could be developed into a suite of tools for transcriptome engineering to mediate RNA translation, splicing, or deposition of epitranscriptomic marks. Cas13b could also be utilized for RNA diagnostic or RNA imaging assays. Identifying and characterizing novel CRISPR-Cas systems opens new opportunities for utilizing Cas enzymes for biomedical advances.

Thesis Supervisor: Feng Zhang
Title: Associate Professor of Biological Engineering and Brain and Cognitive Sciences
ACKNOWLEDGEMENTS

There are many people I would like to acknowledge for their scientific or personal support who have helped me throughout graduate school.

Scientifically, I would like to thank my PI Feng Zhang for everything he has helped me with over the past five years as a member of his lab. It has truly been an honor to work with such a forward-thinking and energetic young principal investigator who has assembled a truly amazing group of laboratory personnel. I’d next like to thank my committee members: Steve Bell for being my committee chair, providing thoughtful suggestions for my work, and early support during graduate school; Mary Gehring for being on my committee my whole graduate career and helpful scientific and personal advice, and lastly Kevin Eggan for serving as my external committee member. I’d next like to thank those students and post-docs who co-authored papers with me. First, I’ll thank David Cox for his hard work, critical evaluation of science, and great friendship. I’d also like to thank Aaron Smargon for his enthusiasm and interest for embarking on the discovery of Cas13b with me. Lastly, I will thank Kaijie Zheng, Ian M. Slaymaker, Jonathan S. Gootenberg, Omar A. Abudayyeh, Patrick Essletzbichler, Fei Ann Ran, and Patrick D. Hsu for their help with publishing my first author papers. I also want to thank Sergey Shmakov, Kira S. Makarova, and Eugene V. Koonin, our helpful collaborators. Also to thank is the whole Zhang lab. The members of the Z-lab are the most persistent, dedicated, hard working, intelligent, and talented group of people I’ve ever met. It has been an honor to work amongst them, albeit the most humbling 5 years of my life. I’d also like to thank Michael Yim, Rich Belliveau, Erin Blackwell, and Rhiannon Macrae, the best support staff a lab could have. In addition to those who directly contributed to my published work, I’d like to thank several others who helped me academically. First, thank you to all my former teachers, especially Mr. Richard Parent, whose class in high school was one big contributor in me pursuing a Ph.D. and doing research. I will also thank Eric Rutledge, George Makhatadze, George Edick, Erin Piazza, Mandana Sassanfar, Krishanu Saha, and Rudolf Jaenisch for support during undergraduate or summer research. I’d also like to thank the NSF GFRP for selecting me as a fellow and for financial support. I’d also like to thank scholarship funds that allowed me to go to RPI.

There are many other people who provided personal support during my graduate tenure at MIT. First, I would like to thank my family: Mom, Dad, Natasha, and Nicholas for all the love and support from the day I was born and throughout graduate school. Also the Pyzocha family dogs Maddie, Meeka, Milo, Mixie, and Max (and Ray the cat) just always make me smile. I would next like to thank the person who made the greatest positive difference in my life during graduate school, and that is my fiancé Colin M. Smith who provided daily love, support, and encouragement. I’d also like to thank my extended family: grandparents Nana, Papa, RIP Babci, as well as uncles, aunts, cousins, and soon to be family-in-law including Bruce and Diane Smith. I’d also like to thank my closest friends Lindsay, Jen, Laura K., Brian, Andy, Chris M., Krissy, Chris W., Mike, Cait, and roommates Steph and Chris B. I’d also like to thank supportive classmates including Laura F., Shuyu, and Xinchen. I’d also like to thank particular lab members including Naomi Habib for her support, care, and willingness to discuss life and family as a female scientist; Sara Jones for her support, humor, and an ear to always listen; David Cox for his friendship, humor, and support; Ann Ran for her help and care in the early stages of my graduate career; Bernd Zetsche for his humor and care; Michael Yim for his friendship and willingness to listen and Sourav Choudry for career and personal support. I’d also like to thank Frank Solomon and Anne Welsh.
Undergraduate Education:
2007-2011: B.S. (Biology), Rensselaer Polytechnic Institute, Troy, NY.
Cumulative GPA: 3.93/4.0. Graduated Summa cum laude

Awards:
2011: NSF Graduate Research Fellow
2013: Teresa Keng Graduate Teaching Prize (Biology Department, MIT)

Publications:
Smargon, A*; Cox, D*; Pyszocha, N*; Zheng, K; Slaymaker, I; Gootenberg, J; Abudayyeh, O;
Essletzbichler, P; Shmakov, S; Makarova, K; Koonin, E; Zhang, F. Cas13b is a Type VI-B CRISPR-associate
RNA-Guided RNAse differentially regulated by accessory proteins Csx27 and Csx28. Molecular Cell
2017, 65, 1-13. (* Co-first authors)

Shmakov, S; Smargon, A; Scott, D; Cox, D; Pyszocha, N; Yan, W; Abudayyeh, O; Gootenberg, J;
Makarova, K; Wolf, Y; Severinov, K; Zhang, F; Koonin, E. Diversity and evolution of Class 2 CRISPR-
Cas systems Nature Reviews Microbiology 2017, 15, 169-182.

Pyszocha, N; Ran, F; Hsu, P; Zhang, F. RNA-guided Genome Editing of Mammalian Cells. Methods in
molecular biology 2013, 1114, 269-277.

Saha, K; Mei, Y; Reisterer, C; Pyszocha, N; Yang, J; Muffat, J; Davies, M; Alexander, M; Langer, R;
Anderson, D; Jaenisch, R. Surface-engineered substrates for improved human pluripotent stem cell
culture under fully defined conditions. Proceedings of the National Academy of Sciences of the United

Mei, Y; Saha, K; Bogatyrev, S; Yang, J; Hook, A; Kalcioğlu, Z; Cho, S; Mitalipova, M; Pyszocha, N; Rojas,
F; Van Vliet, K; Davies, M; Alexander, M; Langer, R; Jaenisch, R; Anderson, D. Combinatorial
development of biomaterials for clonal growth of human pluripotent stem cells. Nature
Materials 2010, 9, 768–78.

Tsamaloukas, A; Pyszocha, N; Makhatadze, G. "Pressure Perturbation Calorimetry of Unfolded

Teaching Experience and Certifications:
2016: Contributing author for Introductory Biology online textbook secretoflife.org. Wrote, edited,
and composed figures for several subchapters.
Spring 2015: Teaching Assistant, Introduction to Experimental Biology and
Communication (7.02), MIT
Spring 2015: Volunteer Science Fair Judge at the Massachusetts State Science & Engineering
Fair, Pioneer Charter School of Science in Everett, MA, and Boston Latin Academy in
Boston, MA.
Spring 2014: Kaufman Teaching Certificate Program, MIT
Spring 2013: Teaching Assistant, Introductory Biology (7.013), MIT
2011-2013: Volunteer tutoring, Tutoring Plus, Cambridge, MA
TABLE OF CONTENTS

Chapter 1 ------------------------------------------------- Page 7
Introduction: CRISPR-Cas Systems And Their Use In Genome And Transcriptome Engineering

Chapter 2 ------------------------------------------------- Page 33
Thesis Work: Cas13b Is A Type VI-B CRISPR-Associated RNA-Guided Rnase Differentially Regulated By Accessory Proteins Csx27 And Csx28

Chapter 3 ------------------------------------------------- Page 71
Future Directions: Basic Biology And Engineering Applications Of Cas13b
CHAPTER 1
INTRODUCTION
CRISPR-CAS SYSTEMS AND THEIR USE IN GENOME AND TRANSCRIPTOME ENGINEERING
CRISPR-Cas editing technologies have revolutionized modern molecular biology by making targeted DNA edits simple and scalable for use in mammalian genome engineering and other biological assays. CRISPR-Cas technologies are harnessed from naturally occurring microbial systems that display wide diversity for targeted nucleic acid cleavage. Several CRISPR-Cas single effector enzymes have been characterized and engineered for use in mammalian cells, and their unique properties can make a critical difference in experimental use or targeting specificity. Currently, there exist tools to efficiently target DNA, and tools to easily modify RNA illustrate an unmet need. Increasing our understanding of novel CRISPR-Cas systems will expand the opportunities for utilizing them for biomedical advances. Therefore there exists an interest to continue to identify, characterize, and engineer CRISPR-Cas enzymes and further develop them for genome or transcriptome editing.

Part I: CRISPR Origins, Function, and Categorization

The evolutionary web of life represents a complex and competitive environment in which organisms must continually evolve for better adaptation and survival. The microbial world displays this concept notably: phage (viruses that infect bacteria) can outnumber bacteria 10:1 in some environments (Suttle, 2005) and parasitic mobile genetic elements such as plasmids cause unnecessary energetic or genomic burdens (Doolittle and Sapienza, 1980). To combat these forces, bacteria and archaea have evolved several different forms of resistance to foreign genetic material. A collective list of immune systems used by bacteria and archaea include physical blockage, restriction enzymes, abortive infection, bacteriophage exclusion, and CRISPR-Cas systems, but vary in their representation in specific bacterial or archaeal species. Physical blockage mechanisms result in physically altered cell surfaces to reduce phage entry. Two specific physical blockage mechanisms include adsorption resistance, where surface receptors mutate to inhibit bacteriophage interaction, and penetration blocks where surface receptors are physically obstructed from bacteriophages, for example using extracellular polymeric substances (Labrie et al., 2010). Alternatively, numerous restriction modification system enzymes have sequence specific endonuclease activity to cleave unmethylated phage DNA leaving methylated host DNA protected (Meselson and Yuan, 1968). Bacteriophage exclusion immunity mechanisms specifically inhibit phage DNA replication through unknown mechanisms and protect host cell DNA specifically based on host cell DNA methylation (Goldfarb et al., 2015). Abortive infection systems are a suite of multiple different enzymatic approaches to eliminate phage infection by targeting phage
replication, transcription, or translation machinery but results in phage and the organisms' own
death to prevent bacteriophage infection of the greater bacterial community (Snyder, 1995). Lastly,
CRISPR-Cas adaptive immune systems protect bacteria and archaea against foreign invading
genetic elements such as viruses or plasmids (Marraffini and Sontheimer, 2008) by encoding those
foreign DNA fragments into the host genome and later target the foreign nucleic acid for cleavage
based on complementary association with the encoded sequence (Barrangou et al., 2007).

Basic Components of CRISPR-Cas Systems

The acronym CRISPR-Cas describes the basic components in the microorganisms' genomic
locus necessary for the adaptive immunity process. CRISPR stands for Clustered Regularly
Interspaced Short Palindromic Repeats (Jansen et al., 2002), which describes the repetitive CRISPR
array targeting moiety, and Cas is short for CRISPR-associated, which describes the diverse cas
genes whose proteins perform various functions required for CRISPR-Cas based immunity (Jansen
et al., 2002). The composition of the CRISPR array and function of the Cas proteins are described in
further detail below and illustrated in figure 1.

The CRISPR array has a regulatory leader sequence upstream of the alternating short
variable protospacer (also known as spacer) sequences and short direct repeat (DR) sequences.
The A-T rich regulatory leader sequence directly upstream of the CRISPR array contains the
promoters for CRISPR array transcription (Pougach et al., 2010). Unique spacers are sequences
identical to sequences from invading foreign genetic elements, which can originate from multiple
different phages or mobile genetic elements such as plasmids (Bozolin et al., 2005; Mojica et al.,
2005; Pourcel et al., 2005). The partially palindromic direct repeats within a single CRISPR-Cas
system usually have the same sequence and when transcribed, this direct repeat individually, or in
some cases, complexed to another unique trans-activating RNA (tracrRNA) sequence (Deltcheva et
al., 2011), forms a unique hairpin-like structure that is recognized by the Cas protein(s) from its
locus (Brouns et al., 2008). The CRISPR array is not translated and its final functional form is
processed individual CRISPR RNA (crRNA) units consisting of a spacer and direct repeat.
Figure 1: Schematic of a CRISPR-Cas locus. Although the loci will vary depending on the CRISPR-Cas type, the Type II CRISPR-Cas9 locus from the bacteria *Streptococcus pyogenes* is shown here as a representative. Figure adapted from Cong et al., 2013.

The Cas genes are organized in an operon expression system and have diverse roles to result in adaptive immunity. There are four distinct functional modules that classify Cas protein function including spacer acquisition, crRNA processing, target cleavage, and ancillary roles. These proteins play critical roles in the three phases of CRISPR adaptive immunity: spacer acquisition proteins mediate adaptation, crRNA processing proteins are involved in creating the CRISPR-Cas functional complex, and target cleavage proteins lead to interference of foreign nucleic acid propagation and immunity. The three phases of CRISPR adaptive immunity are described in further detail in the next section and are summarized in Figure 2.

Figure 2: Schematic summarizing the three stages of CRISPR-Cas adaptive immunity. Image from http://rna.berkeley.edu/crispr.html.
The three phases of CRISPR adaptive immunity

Adaptation is the first phase of CRISPR adaptive immunity and is the process by which foreign nucleic acids are encoded into the CRISPR array (Barrangou et al., 2007). Acquisition of DNA protospacers requires the Cas1 and Cas2 proteins (Datsenko et al., 2012; Nunez et al., 2014; Yosef et al., 2012), which are conserved across different CRISPR-Cas systems (Makarova et al., 2011). Cas1 functions as an integrase, contains the catalytic residues responsible for inserting new spacers into the CRISPR array (Nunez et al., 2014), and is the most highly conserved Cas protein in all CRISPR systems (Takeuchi et al., 2012). The Cas2 protein forms a complex with the Cas1 protein (Nunez et al., 2014), but its specific function is less well understood. Other proteins may also form a complex with Cas1 and Cas2 to assist in acquisition in certain CRISPR-Cas types, for example Cas9 in Type II systems, which is the key protein involved in target cleavage (Heler et al., 2015). In addition to protein requirements, many systems have known target sequence requirements that must be present to insert a DNA fragment into the CRISPR array. This sequence, known as the protospacer adjacent motif (PAM), is adjacent to the spacer sequence originating from the phage and not inserted into the CRISPR array (Deveau et al., 2008; Mojica et al., 2009). It is the preferred recognition site for Cas proteins involved in acquisition (Heler et al., 2015; Yosef et al., 2012). Spacers are also added in a directional fashion, with the newest spacers inserted in between the leader sequence and the first repeat concurrently with direct repeat duplication (Yosef et al., 2012). Acquisition is critical because it allows bacteria to adapt to new foreign genetic insults.

Expression is the second phase of CRISPR adaptive immunity, during which the components for targeted nucleic acid cleavage are synthesized, processed, and complexed to form the RNA-guided endonuclease complex. In the expression phase, cas genes in the operon are transcribed and translated into numerous Cas proteins. The CRISPR array is transcribed as a single, long, immature pre-CRISPR RNA (pre-crRNA) transcript containing multiple spacers and direct repeats. It is processed into its final form, the crRNA, which consists of one spacer and one direct repeat, by either an endonuclease subunit of a Cas protein (Carte et al., 2008) or other cellular RNases, such as bacterial RNase III (Deltcheva et al., 2011). The individual crRNAs may be further processed to remove some RNA bases into a shortened mature form by nuclease activity (Deltcheva et al., 2011). The targeting RNAs (crRNA only or crRNA+tracrRNA) interact with Cas protein(s) to form a targeting nuclease complex (Brouns et al., 2008).

Interference is the final phase of CRISPR adaptive immunity and results in cleavage of the targeted foreign nucleic acid. Interference mechanisms vary between different CRISPR types but in all systems, targeted nucleic acid cleavage depends on the proper base pairing between the crRNA
and the target sequence in the protospacer and the protospacer flanking region(s). In systems that utilize a PAM for adaptation, the presence of the correct PAM on the target DNA is a requirement for interference (Mojica et al., 2009). The CRISPR-Cas RNA-guided endonuclease complex first recognizes and binds the PAM sequence on the target DNA (Sashital et al., 2012; Sternberg et al., 2014). Subsequently, the target DNA and the crRNA protospacer sequence form Watson-Crick base pairs at the PAM proximal or “seed region” and continue outward to the PAM distal portion of the targeted DNA (Sternberg et al., 2014; Wiedenheft et al., 2011). With sufficient pairing, the cognate nucleic acid is cleaved (Gasiunas et al., 2012; Semenova et al., 2011), resulting in interference. The PAM sequence serves an important role during interference and because the sequence requirement is outside the protospacer sequence present in the CRISPR array, it ensures that the CRISPR-Cas system will not cleave its own genomic DNA. In systems that do not utilize a PAM for adaptation, target cleavage is only achieved in the absence of base pairing proximal to the protospacer (Marraffini and Sontheimer, 2010). Perfect base pairing proximal to the protospacer would only occur if the RNA-targeting nuclease complex were targeting its own DNA (Marraffini and Sontheimer, 2010; Samai et al., 2015). The sequence binding requirements for interference are summarized in Figure 3.

**Figure 3: Base pairing between protospacer adjacent sequences during interference.** Left: For CRISPR-Cas systems that require a PAM (such as Cas9 shown here), the PAM serves as an initial binding site, which leads to DNA melting and cleavage with extensive cleavage. Image from Wu et al., 2014. Right: For CRISPR-Cas systems that do not utilize a PAM, this schematic illustrates how crRNA binding helps the enzyme discriminate between self and non-self. Image from Marraffini and Sontheimer 2010.

A majority of the known Cas proteins function in one of the three phases of CRISPR adaptive immunity, but some CRISPR systems have ancillary proteins, that conduct various but generally less
well characterized roles. These rare ancillary proteins are believed to provide regulatory, modulating, or additional functions. For example, Cas4 has known functions including 5' to 3' ssDNA exonuclease activity and involvement with acquisition processes (Zhang et al., 2012), but it is also hypothesized to have roles in CRISPR-Cas-coupled programmed cell death. (Makarova et al., 2014) Also, the sl7009 gene has been implicated as a negative regulator in a specific CRISPR-Cas system in Synechocystis sp (Hein et al., 2013). Some ancillary Cas proteins have a CRISPR-associated Rossman fold (CARF) domain, which is predicted to bind nucleotides. Some Cas proteins with CARF domains also exhibit nucleic acid cleavage activity (Makarova et al., 2014). Two Cas proteins with CARF domains, Csx1 and Csm6, play key roles in certain RNA-targeting CRISPR-Cas systems. Further analysis of ancillary proteins could help elucidate specific regulation or function of certain CRISPR-Cas systems.

**CRISPR Nomenclature**

CRISPR-Cas systems in bacteria or archaea exhibit notable diversity and are organized according to a specific classification scheme. Approximately 47% of analyzed bacterial genomes and 87% of analyzed archaea genomes have at least one CRISPR system and many contain multiple CRISPR-Cas systems (Makarova et al., 2015). CRISPR-Cas protein classification is based on phylogenetic, comparative genomic, and protein structural analyses (Makarova et al., 2015; Shmakov et al., 2017). Once cas genes are examined to determine the homology to other known cas genes, categorization is determined by the presence of a specific signature protein directly involved in target nucleic acid cleavage and further classified into subgroups based on the presence of additional signature genes or specific gene arrangements. Currently, there are two classes of CRISPR-Cas systems composed of six major types, which are further divided into nineteen subtypes (Makarova et al., 2015; Shmakov et al., 2017). It is important to note however, that the naming of signature proteins and CRISPR-Cas types is based on the timeline of characterization or experimental validation and therefore does not have a sequential naming scheme based on characterization alone (Makarova et al., 2011; Makarova et al., 2015; Shmakov et al., 2015; Shmakov et al., 2017).

CRISPR-Cas systems are most broadly characterized as either Class 1 or Class 2 (Makarova et al., 2015). Class 1 systems require multiple Cas proteins to come together in a complex to mediate interference against foreign genetic elements. Class 1 systems occur in both bacteria and archaea and are the most common CRISPR-Cas systems utilized in microbial adaptive immunity (Makarova et al., 2015). Class 1 systems are further divided into three CRISPR-Cas types based on
the presence of a specific signature protein: Type I contains Cas3, Type III contains Cas10, and the putative Type IV contains CsfI. Type I mediates interference through the multisubunit Cascade complex containing Cas3. Type III systems are characterized based on the presence of Cas10 but notably form complexes with additional Cmr or Csm proteins. Type IV is putative due to its computational characterization, but lacks experimental evidence confirming targeted nucleic acid cleavage. In contrast to Class 1 systems, Class 2 systems utilize a large single Cas enzyme to mediate interference. Class 2 systems are less common than Class 1 systems and occur almost exclusively in the bacterial domain of life. Sequencing studies have found only four species of archaea with Class 2 systems (Burstein et al., 2017; Shmakov et al., 2017). Class 2 systems are further divided into three CRISPR-Cas types based on the presence of other specific signature proteins: Type II contains Cas9, Type V contains Cas12 (also named Cpf1, C2c1, and C2c3), and Type VI contains Cas13 (also named C2c2). Figure 4 shows a schematic of the two classes and six major subtypes of CRISPR-Cas systems with their associated Cas proteins involved with expression, interference, adaptation, and ancillary functions.

Each CRISPR system cleaves specific nucleic acids:

CRISPR-Cas systems display a range of targeted nuclease activity; they either cleave DNA, RNA, or both RNA and DNA based on the presence of specific nucleic acid cleavage domains observed on the signature Cas proteins. Although there are several different signature Cas proteins,
there are a few recurrent functional catalytic nuclease domains. In this section, the nucleic acid
target and its associated active nuclease domain(s) will be discussed for each type of CRISPR-Cas
system. Table 1 summarizes these conclusions.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>CRISPR type</th>
<th>Protein(s) with nuclease activity</th>
<th>Nuclease Domain(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dsDNA</td>
<td>Type I</td>
<td>Cas3</td>
<td>HD</td>
</tr>
<tr>
<td></td>
<td>Type II</td>
<td>Cas9</td>
<td>RuvC and HNH</td>
</tr>
<tr>
<td></td>
<td>Type V</td>
<td>Cas12 (Cpf1, C2c1, C2c3)</td>
<td>RuvC and Nuc (analogous to HNH)</td>
</tr>
<tr>
<td>ssRNA</td>
<td>Type VI</td>
<td>Cas13a (C2c2)</td>
<td>HEPN</td>
</tr>
<tr>
<td>ssDNA &amp;</td>
<td>Type III</td>
<td>Cas10</td>
<td>HD</td>
</tr>
<tr>
<td>ssRNA</td>
<td></td>
<td>Csm3/Cmr4</td>
<td>HEPN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Csm6/Csx1</td>
<td>HEPN</td>
</tr>
</tbody>
</table>

Table 1: A summary of CRISPR-Cas targeted substrates and associated CRISPR types, nuclease proteins, and associated nuclease domains. With multiple different CRISPR types and unique proteins with nuclease activity, certain nuclease domains are common and dictate substrate specificity.

The CRISPR-Cas types that exclusively cleave dsDNA are Type I, II, and V, and the DNase
domains utilized vary with respect to whether the system is a Class 1 or Class 2 CRISPR system. The
Type I signature target cleaving Cas protein, Cas3, cleaves DNA residues using the Histidine-aspartate (HD) nuclease (Huo et al., 2014; Sinkunas et al., 2011). Mutagenesis of the Cas3 HD
nuclease catalytic domains eliminates DNase activity (Sinkunas et al., 2011). On the other hand,
Type II and V CRISPR-systems are Class 2 large single effector enzymes and both of these DNA
nucleases utilize two separate nuclease domains. The two DNA nuclease domains include RuvC-like and either HNH-like, present in Type II systems, or Nuc, present in Type V systems and functionally analogous to HNH. Crystal structure evidence illustrates the presence of these nuclease domains in
Cas9 (Jinek et al., 2014; Nishimasu et al., 2015; Nishimasu et al., 2014) and Cas12 (specifically Cpf1 and C2c1) (Dong et al., 2016; Liu et al., 2017a; Yamano et al., 2016; Yang et al., 2016). The RuvC and
HNH nuclease domains are also catalytic domains present in other DNases: the RuvC domain has a
similar structure and function to the RuvC holiday junction resolvase (Gorecka et al., 2013;
Nishimasu et al., 2014) and the HNH nuclease has a ββα-metal fold tertiary structure (Nishimasu et al., 2014) that is present in phage T4 endonuclease VII (Biertumpfel et al., 2007). Genetic evidence
has also identified specific residues involved in DNA cleavage activity. For example in Cas9, both
nuclease domains must be mutated to create a catalytically inactive protein because each nuclease
domain cleaves one of the two strands of DNA to result in a DSB (Gasiunas et al., 2012). Mutating either HNH or RuvC alone creates a nickase enzyme that cleaves just one of the two strands of DNA (Gasiunas et al., 2012; Jinek et al., 2012).

The C2c2 protein of the Type VI CRISPR-Cas system exclusively cleaves ssRNA using HEPN domains (Abudayyeh et al., 2016). HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) (Grynberg et al., 2003) domains exhibit wide diversity of amino acid sequences and occur in proteins involved in several forms of bacterial immunity including abortive infection, restriction modification, and CRISPR-Cas systems (Anantharaman et al., 2013). HEPN domains are the catalytic domains of many different RNases including RNase LS and Irel (Anantharaman et al., 2013; Lee et al., 2008; Otsuka and Yonesaki, 2005). HEPN domains generally consist of α-helical structures and have four commonly observed amino acids: E upstream of an R(X4-6)H motif where the amino acid after R is typically polar (often N, D, or H) (Anantharaman et al., 2013). Mutagenesis studies in the HEPN domain of Irel demonstrated that arginine and histidine are essential for nuclease activity (Lee et al., 2008) and mutagenesis of arginine in either HEPN domain of C2c2 resulted in a catalytically dead enzyme (Abudayyeh et al., 2016). To function as an active RNase, two HEPN domains need to associate together. Two HEPN domains from the same protein can come together, as is the case for C2c2 (Liu et al., 2017b), or a single HEPN domain can be present but require two individual proteins to dimerize for activity (Lee et al., 2008).

The Type III CRISPR-Cas systems cleave both RNA and DNA co-transcriptionally (Peng et al., 2015; Samai et al., 2015) using Cas proteins with RNase and DNase domains previously described. The Cas10 family of proteins (also known as Csm1 or Cmr2) are characteristic for Type III CRISPR systems and usually have a HD nuclease domain, similar to the Type I systems, which are responsible for cleaving DNA (Elmore et al., 2016). As a Class 1 CRISPR-system, Type III systems consist of many proteins, and other Cas proteins within the locus have a HEPN domain that cleaves RNA. Notably, there are two different mechanisms by which RNA is cleaved in Type III systems: crRNA-guided RNase activity or crRNA-independent RNase activity. In guided RNase activity, Csm3 in III-A systems (Samai et al., 2015; Staals et al., 2014; Tamulaitis et al., 2014) or Cmr4 in III-B systems (Benda et al., 2014; Ramia et al., 2014) associates with the target crRNA and cleaves the RNA transcript. RNase activity independent of the crRNA is mediated through Csm6 in Type III-A (Jiang et al., 2016; Niewoehner and Jinek, 2016) and Csx1, which is a Csm6 ortholog in Type III-B (Kim et al., 2013; Lintner et al., 2011; Sheppard et al., 2016), and no stable complex formation with the crRNA or crRNA binding proteins is observed. Both targeted and non-targeted RNases help mediate interference against foreign nucleic acids although the mechanism by which trans acting
systems help mediate interference is still under investigation. There are a few other notable features of Type III systems relating to its activity as an adaptive immune system. Many Type III systems do not have a cas1 gene within its locus (Makarova et al., 2015) and targeting of RNA is often PAM-independent (Hale et al., 2009). As mentioned before, target cleavage in PAM-independent systems is achieved due to the absence of base pairing proximal to the protospacer. Lastly, since Cas1 is the only known protein required for adaptation, it is not known whether some Type III systems lack adaptation capabilities, utilize Cas1 in trans, or utilize other unknown mechanisms.

**Part II: Genome Engineering using Class 2 CRISPR-Cas Systems**

In the previous section, the basic biology of several different CRISPR-Cas systems was introduced to provide a background understanding of CRISPR-Cas origins, function, and categorization. In this part, CRISPR-Cas systems will be discussed with respect to their utility as genome or transcriptome editing tools. The ability of CRISPR-Cas systems to undergo targeted nucleic acid cleavage within bacteria or archaea represents a system that can be repurposed for targeted editing of nucleic acids in other organisms. Harnessed CRISPR-Cas systems can provide a simple system for genome or transcriptome modulation and can also be applied for multiple uses in other biological assays. Class 2 CRISPR-Cas systems are employed for genome engineering or assay development simply because there are fewer components to engineer compared to Class 1 CRISPR-Cas systems. The targeting realm of useful genome or transcriptome editing CRISPR-Cas technologies and related assays are dependent on the diversity of characterized Class 2 CRISPR-Cas systems.

**Genome editing and predecessors to CRISPR-Cas systems**

Traditional genome editing occurs by targeting DNA double stranded breaks (DSBs) and utilizing natural repair processes within cells to insert, delete, or alter the genetic code. Although genome editing can happen spontaneously by simply providing an edited template (Smithies et al., 1985; Thomas et al., 1986), this process is exceedingly rare and can even result in higher incorporation of the fragment in other parts of the genome (Lin et al., 1985). Introduction of a DSB near the site of interest increases targeted genome editing efficiency (Bibikova et al., 2001; Bibikova et al., 2002) because it stimulates efficient cellular DSB repair mechanisms known as non-homologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ is the predominant
repair pathway often guided by microhomology of the DNA (Roth and Wilson, 1986), but is an error-prone process and frequently results in small insertions or deletions (indels) (Roth et al., 1985). Indels occurring within a protein-coding gene have a 2/3 chance of resulting in a frameshift mutation leading to a premature stop codon and gene knockout. Alternatively, HDR leads to precise gene editing. However, HDR requires a homology template and does not function in non-dividing cells (Symington and Gautier, 2011).

Genome engineering has origins in three other systems: homing endonucleases (HEs) (Smith et al., 2006; Stoddard, 2005), Zinc Finger Nucleases (ZFs) (Kim et al., 1996; Miller et al., 2007; Porteus and Baltimore, 2003), and transcription activator-like effector nucleases (TALENs) (Christian et al., 2010; Miller et al., 2011; Zhang et al., 2011). HEs are sequence specific endonucleases that originate from microbial mobile genetic elements (Smith et al., 2006). ZFs are derived from modular eukaryotic transcription factor DNA binding domains and fused to the non-specific DNA cleavage domain of the FokI endonuclease (Kim et al., 1996). TALENs are modular DNA binding domains derived from a family of proteins called the transcription activator-like effectors (TALE) from the plant pathogen Xanthomonas genus, and, like ZFs, the TALE binding module is fused to the FokI endonuclease domain to create the TALEN (Christian et al., 2010). Although these systems have some limitations with respect to either specificity for HEs (Smith et al., 2006) or module interference for ZFs (Maeder et al., 2008) and TALENs (Juillerat et al., 2014), a key distinction is that these three methods employ protein-based DNA targeting modalities in which sequence specificity is conferred through direct protein-DNA interactions. Therefore, a unique protein must be cloned, expressed, and tested for each specific targeted DNA sequence, which requires a lengthy cloning process that is not scalable. CRISPR-Cas systems offer a notable advantage over protein-based targeting modalities because sequence specificity is determined by RNA-DNA (or in the case of RNA-targeting CRISPR-Cas systems RNA-RNA) interactions and guide RNA synthesis is simple and easy to complete compared to engineering a whole protein. CRISPR-Cas systems also offer a promising potential to develop RNA-based transcriptome engineering tools that can mediate RNA modification other than knockdown which can be accomplished using existing siRNA or miRNA technologies.

Currently developed and promising CRISPR-Cas systems in genome and transcriptome engineering

To use CRISPR-Cas systems in heterologous genome or transcriptome editing contexts, the system must be identified, characterized, harnessed to function in the organism of choice, and analyzed for genome-wide targeted efficiency or developed into molecular tools. The Class 2
enzymes described earlier fall along this spectrum: some have limited characterization whereas others have undergone several targeted cleavage analyses. For an illustration of Class 2 CRISPR-Cas systems and their development for molecular tools, see figure 5. Currently, there are only a few CRISPR-Cas systems harnessed for mammalian genome editing and they exclusively target DNA. In this section, the Class 2 CRISPR-Cas enzymes with respect to their functional understanding and repurposing as a genome-editing tool will be described.

**Figure 5: Functional characterization of Class 2 CRISPR-Cas systems for use as genome editing or transcriptome editing tools.**

The first known CRISPR-Cas systems were discovered empirically due to their high abundance in nature whereas other more rare CRISPR-Cas systems were discovered through large computational searches. The first empirically discovered CRISPR-Cas systems were Types I, II, and III which are relatively common (Makarova et al., 2011). Next, while analyzing the genomic DNA of the intracellular pathogen *Francisella tularensis*, the Type V *cpf1* gene was first discovered as an uncharacterized gene next to *cas1, cas2, cas4*, and a CRISPR array (Schunder et al., 2013) and later found in other bacterial species (Makarova et al., 2015). With increasing diversity of CRISPR-Cas systems and successful repurposing of Cas9 as a genome-engineering tool, an initial comprehensive search for uncharacterized Class 2 CRISPR proteins was seeded on the presence of *cas1* (Shmakov et al., 2015). It was seeded on *cas1* because it encodes the most highly conserved Cas protein (Takeuchi et al., 2012) and is present in most CRISPR-Cas loci (Makarova et al., 2011; Makarova et al., 2015). From this search, the three large single effector Cas enzymes C2c1, C2c2, and C2c3 were found (Shmakov et al., 2015) and later categorized as Types V or VI. After computational discovery, the system can be further characterized to ascertain predicted functional nuclease domains, tested for expression as well as processing of the CRISPR array, screened for protospacer adjacent
sequence requirements (such as a PAM) for targeting, and illustrate targeted nucleic acid cleavage biochemically or genetically (Abudayyeh et al., 2016; Shmakov et al., 2015; Zetsche et al., 2015).

To utilize CRISPR-Cas systems for targeted genome engineering in heterologous organisms such as mammals, both the Cas proteins as well as the guide RNAs must undergo specific modifications for proper expression and localization in the cell type of interest. Guide RNAs need to be expressed by a promoter recognized by endogenous transcription machinery. In mammalian cells, the RNA polymerase III U6 promoter is used because it is short and able to transcribe short RNAs (Miyagishi and Taira, 2002). Second, if a tracrRNA is required, such as for Cas9, a chimeric crRNA-tracrRNA hybrid RNA can be used to reduce the number of expressed RNAs required for targeted genome editing (Jinek et al., 2012). Lastly, the spacer sequence of the guide RNA must be chosen in the genomic area of interest with the appropriate flanking PAM sequence.

Several modifications are also required for Cas protein function in a mammalian cell. Different orthologs of CRISPR-Cas nucleases display a range of catalytic activity as assayed by indel frequency and PAM sequence requirements (Ran et al., 2015; Zetsche et al., 2015). Orthologs in mammalian cells that express well and display robust as well as consistent levels of targeted DNA cleavage have traditionally been selected for genome editing purposes. PAM sequence requirements can also vary across multiple orthologs (Cong et al., 2013; Ran et al., 2015) and may play a role in the selection of a particular Cas protein for genome editing. After an active and specific Cas ortholog is identified, the cas gene is codon optimized for efficient translation in the specific organism and nuclear localization sequence tags are added so that the Cas9 protein localizes to the nucleus (Cong et al., 2013). Varying promoters can be used to allow for tissue specific or ubiquitous constitutive expression.

The first CRISPR associated single effector enzyme to be repurposed as a mammalian genome-editing tool was Cas9 (Cong et al., 2013; Mali et al., 2013b). The first and most well characterized Cas9 ortholog comes from the species Streptococcus pyogenes SF370, termed SpCas9, which is a large 1368aa protein with a short, permissive NGG PAM (Figure 1). Several other orthologs of Cas9 have been harnessed for mammalian genome editing including Staphylococcus aureus, a notable smaller ortholog of 1053aa but a longer PAM of NNGRR(T) (Ran et al., 2015). Orthologs of Cas9 are widely used to mediate genome editing within in vitro or in vivo environments and have undergone several genome-wide off-target cleavage analyses in mammalian cells to better understand their on-target and off-target cleavage patterns (Kim et al., 2015; Tsai et al., 2015).
In addition to its use as a way to edit the genome, its specific activity, ease of targeting, and modularity have enabled Cas9 to be repurposed for many other assays, summarized in figure 6. Cas9 can be used in genome-wide knockout screens in which a pooled library of guide RNAs can be delivered to a cell population through lentiviral vectors and then cells can undergo positive or negative selection (Shalem et al., 2014). Cas9 based screening functions have similar applications to other screening methodologies such as RNA interference (RNAi), but results in protein knockout instead of knockdown and can have improved specificity. Alternatively, catalytically inactive Cas9, or dCas9, can be used as a programmable DNA binding protein. For targeted transcription initiation, dCas9 is either directly fused to a transcriptional activator such as VP64 (Maeder et al., 2013; Mali et al., 2013a; Perez-Pinera et al., 2013) and/or the guide RNA can undergo structure-guided alterations to direct localization of VP64. A MS2 RNA stem loop, from the MS2 bacteriophage, can be added to a permissive area of the guide RNA and co-transfected with an engineered MS2 coat protein fused to the VP64 protein, in which the MS2 coat protein strongly and specifically binds to the MS2 RNA stem loop (Konermann et al., 2015; Zalatan et al., 2015). Also, dCas9 has also been used for targeted DNA imaging (Tanenbaum et al., 2014) or to alter epigenetic genetic markers such as targeted histone demethylase LSD1 (Kearns et al., 2015).
**Figure 6: Applications of CRISPR-Cas9.** Figure from (Hsu et al., 2014). A. Cas9 can be used for targeted genomic edits as a nuclease, nickase or B. double nickase. C. All components for Cas9 can be delivered on a plasmid or D. direct injection of purified Cas9 protein and guide RNA. E. Viral vectors can be used to deliver CRISPR reagents for *in vivo* targeting. F. Cas9 is useful in genome-wide screening assays, G. activating endogenous transcription, mediating epigenetic modulation, and H. live-imaging labeling of DNA loci. I. Cas9 proteins have been engineered in various ways, including light-inducible regulation.

Cpf1 has also been studied and harnessed for mammalian genome editing (Zetsche et al., 2015). Cpf1 has unique differences compared to Cas9, which can be exploited and used to complement Cas9. Unlike Cas9 enzymes, which have G-rich PAM sequences, Cpf1 enzymes have a T-rich PAM as illustrated by the ortholog *Francisella novicida U112* with a 5'-TTN PAM. Cpf1 is also naturally targeted to the cognate DNA using only the single crRNA which is simpler than the crRNA-tracrRNA duplex used by Cas9 and shorter than the chimeric Cas9 guide RNA. Cpf1 is also unique in that the enzyme itself cleaves its own CRISPR array into individual crRNA units which is especially promising for multiplexed genome editing because only the CRISPR array containing the targeting
spacers is required for Cpf1. Lastly, Cpf1 cleavage results in a staggered cut unlike the blunt cut of Cas9, which may aid in directing homologous recombination. Although Cpf1 exhibits notable differences compared to Cas9, it could be just as useful in many of the same applications as Cas9 such as screening or imaging.

There exist more opportunities to develop more DNA editing tools beyond the characterized orthologs of Cas9 and Cpf1. First, uncharacterized orthologs or engineered versions of characterized enzymes could be explored. Characterizing other orthologs has produced genome engineering tools with reduced DNA sizes such as SaCas9 (Ran et al., 2015) that are ideal for viral delivery methods with packaging limits. Structure-guided engineering has produced genome engineering tools with altered PAM specificity (Klein et al., 2017) and engineered Cas9 enzymes with enhanced cutting specificity (Klein et al., 2016; Slaymaker et al., 2016). Alternatively, there are other Type V and Type V-like Class 2 enzymes that could be adapted for genome or transcriptome editing including C2c1, C2c3, CasX, and CasY. C2c1 and C2c3 are formally categorized as Type V whereas CasX and CasY have similar predicted domain organization. Available data suggests that these four unique enzymes may behave similarly to Cpf1. For example, solved crystal structures show that C2c1 is structurally similar to Cpf1 and biochemically known to generate staggered DNA cuts distal to its T-rich PAM (Liu et al., 2017a; Shmakov et al., 2015; Yang et al., 2016). CasY has local amino acid alignment similarities to C2c3, and cuts DNA with a TA PAM. CasX also cuts DNA with a T-rich TTC PAM (Burstein et al., 2017). Although known differences include that C2c1 uses a tracrRNA and CasX is a small 980 amino acid protein, identifying additional unique properties about each protein could lead to different applications.

Repurposing the first characterized single effector CRISPR-associated RNase C2c2 would enable an entirely new area of RNA or transcriptome engineering. No existing CRISPR-Cas system can simply target endogenous mammalian RNA; for example, Cas9 is capable of targeting RNA but requires an unnatural dsDNA at the PAM (O’Connell et al., 2014) RNA-targeting CRISPR-systems could be used for endogenous RNA knockdown, translation control, splicing control, localization studies or in other applications for targeted RNA-binding assays. C2c2 has two individual HEPN domains that come together to form the RNase active site (Liu et al., 2017b) and validating in vitro cleavage data with purified protein shows cleavage of ssRNA at the RNA base Uracil (Abudayyeh et al., 2016). C2c2 also cleaves its own CRISPR array (East-Seletsky et al., 2016) similar to Cpf1, which could aid in multiplexed targeting of multiple RNA sequences. C2c2 also exhibits a cleavage phenomenon called the collateral effect, which is promiscuous RNase cleavage activity after initial targeted RNA cleavage (Abudayyeh et al., 2016). Although Type III systems also have non-targeted
RNA cleavage activity, C2c2 is unique in that the protein itself first undergoes targeted and then nonspecific cleavage. This unique property has been proposed to be useful in development for RNA detection assays (East-Seletsky et al., 2016) where the RNA for detection is mixed with the C2c2 protein, targeting crRNA, and non-targeted fluorescent-quenched RNAs such that when C2c2 cleaves its target, it also cleaves the fluorescent-quenched RNAs to enable detection. Although C2c2 is a promising prospect for targeted RNA modulation or detection, there are only 30 known orthologs, a rarity compared to the 3,822 Cas9 orthologs (Shmakov et al., 2017), and fewer orthologs reduces the exploitable diversity offered by the system. Identifying additional C2c2 orthologs or other unique RNA-targeting single effector Cas enzymes would increase the diversity of the RNA targeting CRISPR-Cas toolbox.

Based on the utility of characterizing and harnessing unique Cas proteins and orthologs, there exists a desire to conduct an exhaustive search to find additional Class 2 CRISPR associated enzymes for genome or transcriptome engineering purposes. We hypothesized that there may be additional uncharacterized CRISPR-Cas systems and that additional Cas single effector enzymes could be found by simply seeding a search on the presence of a large protein near a CRISPR array. As seen in some Type III systems, cas1 is sometimes not present in active CRISPR-Cas loci and therefore are not required for CRISPR-Cas interference activity. Any uncharacterized protein identified through a search can be computationally analyzed to identify any conserved catalytic domains and then biochemically or genetically characterized in the lab.
References:


CHAPTER 2

CAS13B IS A TYPE VI-B CRISPR-ASSOCIATED RNA-GUIDED RNASE DIFFERENTIALLY REGULATED BY ACCESSORY PROTEINS CSX27 AND CSX28

This chapter is adapted from the following article:


Contributions: A.A.S., D.B.C, and N.K.P. are co-first authors; A.A.S., D.B.C, N.K.P. designed and performed all experiments and analyzed data under the guidance of F.Z; A.A.S. led the design and implementation of the CRISPR discovery pipeline with help from K.Z and N.K.P; A.A.S. and D.B.T.C. performed screening experiments and computational analysis. D.B.C. led the microbiological assays for characterizing in vivo function of Cas13b, Csx27, and Csx28 with help from J.S.G., O.A.A., and P.E; N.K.P. purified proteins with help from I.M.S; N.K.P. completed the biochemical characterizations of Cas13b activity. S.S., K.S.M., and E.V.K provided input on annotation, classification, and naming of Cas13b, Csx27, and Csx28. A.A.S., D.B.C, N.K.P., and F.Z. wrote the manuscript with input from all authors.
CRISPR-Cas adaptive immune systems defend microbes against foreign nucleic acids via RNA-guided endonucleases. Using a computational sequence database mining approach, we identify two Class 2 CRISPR-Cas systems (subtype VI-B) that lack Cas1 and Cas2 and encompass a single large effector protein, Cas13b, along with one of two previously uncharacterized associated proteins, Csx27 or Csx28. We establish that these CRISPR-Cas systems can achieve RNA interference when heterologously expressed. Through a combination of biochemical and genetic experiments, we show that Cas13b processes its own CRISPR array with short and long direct repeats, cleaves target RNA, and exhibits collateral RNase activity. Using an *E. coli* essential gene screen, we demonstrate that Cas13b has a double-sided protospacer-flanking sequence. We also find that Csx27 represses, whereas Csx28 enhances, Cas13b-mediated RNA interference. Characterization of these CRISPR systems creates opportunities to develop tools to manipulate and monitor cellular transcripts.

**Introduction**

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins) systems are divided into two classes, Class 1 systems, which utilize multiple Cas proteins and CRISPR RNA (crRNA) to form an effector complex, and the more compact Class 2 systems, which employ a large, single effector with crRNA to mediate interference (Makarova et al., 2015). CRISPR-Cas systems display a wide evolutionary diversity, involving distinct protein complexes and different modes of operation, including the ability to target RNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Hale et al., 2009; Jiang et al., 2016; Staals et al., 2013; Staals et al., 2014; Tamulaitis et al., 2014).

Computational sequence database mining for diverse CRISPR-Cas systems has been carried out by searching microbial genomic sequences for loci harboring the *cas1* gene, the most highly conserved *cas* gene involved in the adaptation phase of CRISPR immunity (Marraffini, 2015). Among other findings, this approach led to the discovery of the Class 2 subtype VI-A system with its signature effector Cas13a (previously known as C2c2), which targets RNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Shmakov et al., 2015). Since distinct variants of functional Class 1 CRISPR systems have been characterized that lack *cas1* (Makarova et al., 2015), we sought to identify Class 2 CRISPR-Cas systems lacking *cas1* by modifying the computational discovery pipeline so that it is not seeded on Cas1. Here we report the characterization of a Class 2 subtype, VI-B, which was
discovered through this computational approach, and demonstrate that the VI-B effector, Cas13b, is an RNA-guided RNase.

**Results:**

**Computational discovery of Class 2 subtype VI-B CRISPR systems**

We designed a computational pipeline to search specifically for putative CRISPR-Cas loci lacking Cas1 and Cas2 (Figure 1A). Fully assembled microbial genomes were searched for all proteins within 10kb of CRISPR arrays (Edgar, 2007; Yates et al., 2016). The list of identified loci was further narrowed down using the following criteria: no more than one neighboring protein larger than 700aa (to eliminate Class 1 system false positives), presence of a putative single effector of size 900aa to 1800aa (informed by the size distribution of previously classified Class 2 effectors), and absence of cas1 and cas2 genes within 10kb of the CRISPR array (Method Details). Candidate effectors were grouped into families according to homology (Camacho et al., 2009; Hildebrand et al., 2009; Remmert et al., 2012), and discarded if they matched previously identified CRISPR-Cas systems (Makarova et al., 2015). To focus on likely functional CRISPR loci, we limited the candidate list to families of at least 10 non-redundant effectors in which the putative effector was near a CRISPR array for at least 50% of the members.

Among the candidates, we identified two genetically diverse putative Class 2 CRISPR-Cas systems (105 genomic loci, 81 containing a unique entry Cas13b in the non-redundant NCBI protein database, and 71 of these 81 containing an annotated CRISPR array) represented in Gram-negative bacteria (Figure 2A). For some genera, in particular *Porphyromonas* and *Prevotella*, Cas13b proteins are encoded in several unique sequenced loci, and, occasionally, in the same sequenced genome. These systems often co-occur with other CRISPR-Cas systems. Of the 81 type VI-B loci found across complete and incomplete bacterial genomes, 62 also possess at least one other CRISPR-Cas locus that includes the key adaptation endonuclease, Cas1. However, three complete genomes carrying the type VI-B locus (Flavobacterium_branchiophilum_FL_15_GCA_000253275.1, Paludibacter_propionicigenes_WB4_GCA_000183135.1, and Porphyromonas_gingivalis_AJW4_GCA_001274615.1) may lack Cas1 altogether.
Figure 1: Discovery of two Class 2 CRISPR-Cas systems, subtype VI-B1 and VI-B2, containing Cas13b. (A) Bioinformatic pipeline to discover putative Class 2 CRISPR loci lacking Cas1 and Cas2. (B) A schematic phylogenetic tree of the subtype VI-B loci. Loci with Csx27 (brown) comprise variant VI-B1; loci with Csx28 (gold) comprise variant VI-B2. Loci may contain either Csx27, Csx28, or neither protein in its locus.

All these loci encode a large (~1100aa) candidate effector protein and, in about 80% of the cases, one different additional small (~200aa) protein (Figures 1B and Figure 2A). The putative effector proteins contain two predicted HEPN domains (Anantharaman et al., 2013) at their N- and C-termini (Figure 2B), similar to the domain architecture of the large effector of subtype VI-A (Cas13a) (Shmakov et al., 2015). Beyond the occurrence of two HEPN domains, however, there is no significant sequence similarity between the predicted effector and Cas13a. These systems were also identified by a generalized version of the pipeline described above as part of a comprehensive analysis of Class 2 CRISPR-Cas systems, and were classified into subtype VI-B, with predicted effector protein Cas13b (Shmakov et al., 2017).
Figure 2: Phylogenetic tree of Cas13b bifurcates into two variants of subtype VI-B CRISPR loci. (A) A phylogenetic tree (alignment generated by BLOSUM62) of non-redundant Cas13b effectors, with the full type VI-B locus depicted in every instance. Accession numbers for genome, Cas13b (blue), and Csx27 (brown)/Csx28 (gold) are included, as well as number of nearby spacers detected by PILER-CR, the presence of Cas1 in the sequenced genome, and the size of Cas13b. (B) Two HEPN sequences identified via multiple sequence alignment (BLOSUM62) of putative non-redundant Cas13b proteins. (C) Divergent HEPN sequence identified via multiple sequence alignment (BLOSUM62) of putative non-redundant Csx28 proteins.
CRISPR-Cas13b loci contain small accessory proteins

The identity of the putative accessory protein correlates with the two distinct branches in the phylogenetic tree of Cas13b (Figures 1B and 2A) (Henikoff and Henikoff, 1992), indicative of the existence of two variant systems, which we denote VI-B1 (accessory protein referred to as Csx27) and VI-B2 (accessory protein referred to as Csx28). While subtype VI-B2 systems almost invariably contain csx28, csx27 is less consistently represented in VI-B1 loci. The protein sequences of Csx27 and Csx28 show no significant similarity to any previously identified Cas proteins. Both putative accessory proteins were predicted to contain one or more transmembrane segments (Figure 3A) (Moller et al., 2001). However, Csx27 of Bergeyella zoohelcum and Csx28 of Prevotella buccae tagged with RFP at either the N- or C-terminus did not show membrane localization when expressed in E. coli (Figure 3B). In addition to the predicted hydrophobic domains, analysis of the multiple sequence alignment of Csx28 proteins indicated the presence of a divergent HEPN domain (Figure 2C).

Figure 3 | Predicted transmembrane domains of Csx27 and Csx28 not validated experimentally. (A) Transmembrane domain prediction in Csx27 of B. zoohelcum and Csx28 of P. buccae using TMHMM v2. (B) N- and C-terminally fused RFP imaging of Csx27 of B. zoohelcum and Csx28 of P. buccae.

Cas13b-associated CRISPR arrays display unique features

In contrast to their differing putative accessory proteins, both variants of subtype VI-B systems show distinct, conserved features in the CRISPR arrays. The direct repeats in the CRISPR arrays are conserved in size, sequence, and structure, with a length of 36 nt, a poly-U stretch in the
**Figure 4:** Predicted sequence and secondary structure of type VI-B direct repeats; predicted protospacer flanking sequences. (A) Predicted secondary structure folds of structurally unique CRISPR class 2 type VI-B1 direct repeats (Vienna RNAfold). (B) Predicted secondary structure folds of structurally unique CRISPR Class 2 type VI-B2 direct repeats. (C) Weblogo of all unique VI-B direct repeat sequences of length 36 nt, taken as the same transcriptional orientation as Cas13b. (D) Weblogo of all unique VI-B protospacer flanking sequences from CRISPRTarget mapping of protospacers to phage databases.
open loop region, and complementary sequences 5'-GUUG and CAAC-3' at the ends of the repeat predicted to yield a defined secondary structure mediated by intramolecular base-pairing (Figure 4) (Lorenz et al., 2011). Our analysis revealed 36 Cas13b spacers mapped with greater than 80% homology to unique protospacers in phage genomes. Twenty-seven of the identified Cas13b spacers targeted the coding strand of phage mRNA, while seven spacers targeted the noncoding strand and two spacers targeted regions of the phage genome without predicted transcripts. Although the composite of these imperfect mappings revealed no consensus flanking region sequence (Figure 4D) (Biswas et al., 2013), the well-conserved protospacer length of 30 nt, combined with the conserved direct repeat sequence and length, suggests that the nucleic acid targeting rules may be similar among different VI-B loci.

RNA sequencing of the total RNA from B. zoohelcum (subtype VI-B1) showed processing of the pre-crRNA into a 66-nt mature crRNA, with the full 30-nt 5' spacer followed by the full 36-nt 3' direct repeat (Figure 5A) (Heidrich et al., 2015; Li and Durbin, 2009; Shmakov et al., 2015). A longer 118-nt crRNA, distal to the 36-nt crRNAs in the CRISPR array and with a direct repeat consisting of 5' and 3' fragments of the 36-nt direct repeat sequence interrupted by an intervening repeat sequence, was also processed. This phenomenon was computationally predicted to occur in additional VI-B loci, such as those from Capnocytophaga canimorsus, Myroides odoratimimus, and Riemerella anatipestifer. Other CRISPR Class 2 effectors are known to process their arrays without involvement of additional RNases (East-Seletsky et al., 2016; Zetsche et al., 2015). Similarly, we find that purified BzCas13b is capable of cleaving its CRISPR array, generating mature crRNAs with short or long direct repeats, an activity which is not affected by mutation of the predicted catalytic residues of the HEPN domain (Figures 5B and 6). This suggests that an independent RNase domain cleaves the CRISPR array into individual crRNA units and does not involve the HEPN domains. Additionally, spacers in the crRNA are not further processed beyond 30 nt, which is the spacer length in the full CRISPR array.
**Figure 5: Cas13b from the VI-B1 locus processes a CRISPR array with two direct repeat variants.** (A) RNA-sequencing of the native VI-B1 locus from *Bergeyella zoohelcum* ATCC 43767. (B) Denaturing gel showing cleavage products of *in vitro* synthesized short-DR containing or long-DR containing CRISPR arrays from the *B. zoohelcum* genome by either wildtype or HEPN mutant *BzCas13b* (D1, R116A/H121A; D2, R1177A/H1182A; Q, R116A/H121A/R1177A/H1182A). The schematic shows fragment lengths of a cleaved CRISPR array.
Figure 6: Protein gels of purified WT BzCas13b and three mutant BzCas13b proteins.

An E. coli essential gene screen reveals targeting rules for BzCas13b

To validate the expected interference activity of the VI-B system and to determine the targeting rules for the VI-B1 locus from B. zoohelcum, we developed an E. coli essential gene screen (Figure 7A). For this negative selection screen, we generated a library of 54,600 unique spacers tiled with single-nucleotide resolution over the coding region of 45 monocistronic essential genes (Baba et al., 2006; Gerdes et al., 2003), plus 60 nt into the 5' and 3' UTRs. We also included 1100 randomly generated non-targeting spacers to establish baseline activity. We then transformed this library with plasmids carrying bzcas13b (cas13b gene from B. zoohelcum) and bzcxs27, just bzcas13b, or a control empty vector. After quality-control-filtering of all screened spacers, we found a statistically significant depletion of targeting spacers over non-targeting spacers, indicating that Cas13b, alone or with Csx27, can achieve nucleic acid interference (Figure 7B).
Figure 7: Heterologous expression of Cas13b mediates knockdown of E. coli essential genes by a double-sided PFS. (A) Design of E. coli essential gene screen to determine targeting rules of nucleic acid interference. (B) Manhattan plots of mean spacer depletions mapped over 45 genes and aggregated across normalized gene distance for either the full B. zoohelcum VI-B1 locus (left) or cas13b alone (right), with non-targeting spacers in gray, safely depleted spacers (>50% above mean depletion of non-targeting spacers) above blue line, and strongly depleted spacers (top 1% depleted) above red line. For the full locus, 36,142 targeting spacers and 630 non-targeting spacers passed QC filter. Of the targeting, 367 are strongly depleted and 1672 are safely depleted. For cas13b alone, 35,272 targeting spacers and 633 non-targeting spacers passed QC filter. Of the targeting, 359 are strongly depleted and 6374 are safely depleted. (C) Weblogo of sequence motifs of strongly depleted B. zoohelcum spacers. (D) Normalized PFS score matrix, where each score is the ratio of number of safely depleted B. zoohelcum spacers to total number of spacers for a given
PFS, scaled so that maximum PFS score is 1. (E) Spacers targeting kanamycin to validate PFS targeting rules of 5’ PFS (D) and 3’ PFS (NAN or NNA). (F) Schematic of kanamycin validation screen for *B. zoohelcum cas13b* in *E. coli*. (G) Results from kanamycin validation screen; spacer abundances versus control for individual *B. zoohelcum* spacers, with abundances colored by type of spacer.

To assess the targeting rules for Cas13b, we established two spacer depletion levels: strongly depleted (top 1% of depleted spacers) and safely depleted (spacers depleted 5σ above the mean depletion of the filtered non-targeting spacers). From spacers passing the strongly depleted cutoff we derived sequence motifs qualitatively identifying a double-sided protospacer flanking sequence (PFS) (Figure 7C) (Crooks et al., 2004). Because each position in a sequence motif is assumed to be independent, we developed a more quantitative, base-dependent PFS score defined as the ratio of the number of safely depleted spacers to the number of all spacers with a given PFS, normalized across all PFS scores (Figure 7D).

The normalized PFS scores revealed a 5’ PFS of D (A, U, or G) and 3’ PFS of NAN or NNA, consistent for Cas13b with Csx27, as well as for Cas13b alone. To validate these sequence-targeting rules, we performed an orthogonal depletion screen with Cas13b alone, targeting the Kanamycin resistance gene (Figures 7E and 7F). Four classes of spacers were created: non-targeting, targeting with both 5’ and 3’ PFS rules, targeting with only the 5’ or 3’ PFS rule, and targeting with neither rule. Consistent with our findings from the *E. coli* essential gene screen, the combined 5’ and 3’ PFS spacers resulted in the highest Kanamycin sensitivity (Figure 7G). The kanamycin screen also provides an additional control compared to the *E. coli* essential gene screen because it targets the same transcript with consistent expression levels and consistent local RNA secondary structure.

*BzCas13b cleaves single-stranded RNA and exhibits collateral activity in vitro*

Based on the presence of the computationally predicted HEPN domains that function as RNases in other CRISPR-Cas systems, including VI-A and some Class 1 systems (Abudayyeh et al., 2016; Kim et al., 2013; Sheppard et al., 2016; Staals et al., 2014), we anticipated that Cas13b interferes with RNA. We confirmed this by demonstrating that purified Cas13b exclusively cleaves single-stranded RNA with both direct repeat architectures (Figures 8A and 9A). We then validated the PFS targeting rules biochemically, showing that a 5’ PFS of C greatly inhibits single-stranded RNA cleavage (Figure 8B), whereas a 3’ PFS of NAN or NNA enhances this activity (Figure 8C).
Figure 8: Cas13b is a programmable single-stranded RNase. (A) Schematic showing the RNA secondary structure of the cleavage target in complex with a targeting 30-nt spacer connected to short direct repeat (top). Denaturing gel demonstrating short direct repeat and long direct repeat crRNA-mediated ssRNA cleavage (bottom). Reactions were incubated for 10 minutes. The ssRNA target is 5' labeled with IRDye 800. Three cleavage sites are observed. (B) Schematic showing three numbered protospacers for each colored 5' PFS on a body-labeled ssRNA target (top). Denaturing gel showing crRNA-guided ssRNA cleavage activity demonstrating the requirement for a D 5' PFS (not C) (bottom). Reactions were incubated for 60 minutes. crRNAs correspond to protospacer numbered from the 5' to the 3' end of the target. Gel lane containing RNA ladder not shown. (C) Schematic of a body-labeled ssRNA substrate being targeted by a crRNA (top). The protospacer region is highlighted in blue, and the orange bars indicate the 5' PFS and 3' PFS sequences. The orange letters represent the altered sequences in the experiment. Denaturing gel showing crRNA-guided ssRNA cleavage activity after 60 minutes of incubation, with the 5' PFS tested as A, and the 3' PFS tested as ANN (bottom). The orange 3' PFS letters represent the RNA bases at the second and third 3' PFS position within each target ssRNA. Gel lane containing RNA ladder not shown. Dashed line indicates two separate gels shown side by side. (D) Schematic showing the secondary structure of the body labeled ssRNA targets used in the denaturing gel. The variable loop of the schematic (represented as N) is substituted with five monomers of the variable loop base in the gel (top). Denaturing gel showing cleavage bands of the homopolymer variable loop base (bottom). The targets were incubated for 30 minutes. Dashed line indicates where the image was stitched together to remove U/A heteropolymer RNA lanes (shown in Figure 9B). Gel lane containing RNA ladder not shown.
Figure 9: Cas13b cleaves and binds to single-stranded RNA. (A) Denaturing gels demonstrating no cleavage of dsRNA, ssDNA, or dsDNA by BzCas13b with either the short DR or long DR. Reactions were incubated for 10 minutes, the same amount of time which results in robust ssRNA cleavage for this target and crRNA pair. The ssDNA and top strand of the dsDNA target is 5' labeled with IRDye 800. The dsRNA target is body labeled. Gel lane containing RNA ladder not shown. (B) Denaturing gel showing cleavage bands from the variable loop target as shown in Figure 8D. The U/A
A heteropolymer consists of the N5 variable loop of alternating U and A residues (5' AUAUA 3'). (C) ssRNA cleavage requires BzCas13b and a targeting crRNA, and this cleavage activity is abolished by addition of EDTA. Gel lane containing RNA ladder not shown. (D) Denaturing gel showing PbCas13b cleavage activity of an ssRNA targeted substrate. The ssRNA is 5' labeled with IRDye 800 and incubated for 30 minutes. Gel lane containing RNA ladder not shown. (E) EMSA gels that were used to quantify the Kd of the WT and mutant BzCas13b proteins, using an on-target crRNA complementary to the targeted ssRNA. (F) EMSA gel of WT BzCas13b with an off-target crRNA. The off-target crRNA is non-complementary to the targeted ssRNA.

Other HEPN domain-containing CRISPR-Cas RNA-targeting systems, such as Csx1 from the Type III-B CRISPR-Cas systems, preferentially cleave targets containing specific single-stranded nucleotides (Sheppard et al., 2016). To determine if Cas13b exhibits such a preference, we tested an RNA substrate with a variable homopolymer loop outside of the spacer:protospacer duplex region (Figure 8D). A heteropolymer loop consisting of alternating A then U was also tested (Figure 9B). We observed cleavage at pyrimidine residues, with a strong preference for uracil. This activity is abolished in the presence of EDTA (Figure 9C), suggesting a divalent metal ion-dependent mechanism for RNA cleavage akin to that of a similar HEPN-containing, Class 2 effector protein, Cas13a (Abudayyeh et al., 2016; East-Seletsky et al., 2016).

Given that Cas13a has also been reported to cleave RNA non-specifically once activated by interaction with the target ("collateral effect") (Abudayyeh et al., 2016; East-Seletsky et al., 2016), we sought to test the ability of Cas13b to cleave a second, non-specific substrate following target cleavage. Using an in vitro assay similar to the one we previously used with Cas13a (Abudayyeh et al., 2016), we incubated Cas13b-crRNA complexes with both a target and non-target RNA substrate. We observed collateral cleavage of the non-targeted RNA, but only in the presence of the target RNA (Figure 10).
**Figure 10: Cas13b exhibits collateral activity.** Denaturing gel showing BzCas13b collateral cleavage activity after 30 minutes of incubation, with schematic of cleavage experiment to the right. Two crRNAs (A and B) target substrate 1 (1A and 1B) or substrate 2 (2A and 2B). Gel lane containing RNA ladder not shown. Dashed line indicates two separate gels shown side by side.

Cas13b shows robust HEPN-dependent interference and is repressed by Csx27 activity

To validate RNA interference *in vivo*, we assayed interference against the lytic, single-stranded RNA bacteriophage MS2, whose life cycle contains no DNA intermediates. We performed an MS2 drop plaque assay at serial dilutions of phage for both *bzcas13b* with *bzcsx27* and *bzcas13b* alone with three spacers targeting the MS2 genome, two at the *lys-rep* interface and one in *rep*, as well as one non-targeting spacer (Figure 11A). We observed substantial reduction in plaque formation for all targeting spacers compared to the non-targeting spacer, confirming sequence-specific RNA targeting by VI-B1 systems. (Figures 11A and 12B). Notably, the presence of *bzcsx27* weakened RNA interference by *bzcas13b* for all three targeting spacers.

To confirm the lack of DNA interference *in vivo*, we modified an existing plasmid interference assay with a protospacer placed either in-frame at the 5’ end of the *bla* ampicillin-resistance gene (transcribed target) or upstream of the *bla* gene promoter on the opposite strand (non-transcribed target). Bacteria co-transformed with *bzcas13b* and spacer as well as the non-transcribed target plasmid survived at a comparable rate to co-transformation of the same target with the empty vector on dual antibiotic selection. For bacteria co-transformed with the transcribed target, the colony forming unit rate under dual antibiotic selection was reduced by ~2 orders of magnitude in the presence of *bzcas13b*, corroborating that Cas13b exclusively targets RNA *in vivo* (Figure 11B).
Figure 11: HEPN domains mediate RNA cleavage by Cas13b, whose activity is repressed by Csx27. (A) Protospacer design for MS2 phage plaque drop assay to test RNA interference (left). Plaque drop assay for full B. zoohelcum VI-B1 locus (center) and bzcas13b (right). (B) DNA interference assay schematic (top) and results (bottom). A target sequence is placed in frame at the start of the transcribed bla gene that confers ampicillin resistance or in a non-transcribed region on the opposite strand of the same target plasmid. Target plasmids were co-transformed with bzcas13b plasmid or empty vectors conferring chloramphenicol resistance and plated on double selection antibiotic plates. (C) Schematic (top) and denaturing gel (bottom) showing ssRNA cleavage activity of WT and HEPN mutant BzCas13b. The protein and targeting crRNA complexes were incubated for 10 minutes. Gel lane containing RNA ladder not shown. (D) Electrophoretic Mobility Shift Assay (EMSA) graph showing the affinity of BzCas13b proteins and targeting crRNA complex to a 5' end labeled ssRNA. The EMSA assay was performed with supplemental EDTA to reduce any cleavage activity. (E) Quantification of MS2 phage plaque drop assay with B. zoohelcum wildtype and Q (R116A/H121A/R1177A/H1182A) mutant Cas13b.
**Figure 12: Targeting rule validation of BzCas13b and MS2 interference assay of BzCas13b and PbCas13b.** (A) Spacers targeting kanamycin to validate PFS targeting rules of 5' PFS (D) and 3' PFS (NAN or NNA) (left). Second kanamycin validation screen bioreplicate of spacer abundances versus control for individual *B. zoohelcum* spacers, with abundances colored by type of spacer (right). (B) Plaque drop assay with bioreplicates for *B. zoohelcum* VI-B1 locus and cas13b, for *P. buccae* VI-B2 locus and cas13b, and for *P. buccae* cas13b with pUC19, *B. zoohelcum* csx27, and *P. buccae* csx28.
We next tested if predicted catalytic residues in the HEPN domains were responsible for RNA cleavage by Cas13b. Three HEPN mutants were obtained by replacing the conserved catalytic arginines and histidines in the two HEPN domains with alanines (R116A/H121A, termed domain 1 (D1); R1177A/H1182A, termed domain 2 (D2); and R116A/H121A/R1177A/H1182A, termed quadruple (Q)) (Figure 6). All mutants lacked observable cleavage activity (Figure 11C), yet retained RNA binding capacity in vitro (Figures 11D and 9E). The wildtype and all three HEPN mutant Cas13b proteins showed comparable binding affinities for a single-stranded target RNA substrate, with K₀ values ranging from 27nM to 42nM (Figures 11D and 9E). The K₀ for off-target binding was found to be greater than 188nM (Figure 9F).

We confirmed the involvement of the HEPN domains in RNA interference in vivo, finding ~5.5 orders of magnitude decrease in resistance to MS2 phage in the quadruple HEPN mutants versus wildtype Cas13b (Figures 11E and 12B). Interestingly, quadruple mutant Cas13b with spacers 2 and 3 still showed weak phage resistance, potentially due to catalytically inactive Cas13b binding to phage genomic RNA, leading to reduced phage replication.

CRISPR-Cas13b effectors are differentially regulated by Csx27 and Csx28

To determine if the established RNA targeting rules generalize across the subtype VI-B systems from diverse bacteria, we characterized the subtype VI-B2 locus from P. buccae. RNA sequencing of the CRISPR array revealed processing effectively identical to that of B. zoohelcum, excluding the long crRNA (Figure 13A). The E. coli essential gene screen with pbcas13b and pbcxs28 or pbcas13 alone led to the identification of a PFS matrix similar to that of B. zoohelcum, with certain PFS's disfavored (Figures 14A, 13B, and 13C). Similar to BzCas13b, PbCas13b was found to cleave targeted single-stranded RNA in vitro (Figure 9D). As with bzcxs27, the presence of pbcxs28 did not appreciably alter the PFS. We further explored the apparent reduced activity of pbcas13b alone relative to the respective full CRISPR-Cas locus using the MS2 phage plaque drop assay and found that pbcxs28 enhances MS2 phage interference by up to four orders of magnitude (Figures 14B and 12B). The differential ability of csx27 to repress and csx28 to enhance cas13b activity generalizes across thousands of spacers in the E. coli essential gene screen (Figure 14C), highlighting the distinctive regulatory modes of the two variants of subtype VI-B CRISPR-Cas systems.
Figure 13: RNA-targeting of *P. buccae* VI-B2 CRISPR locus. (A) RNA-Sequencing of heterologously expressed VI-B2 locus from *P. buccae* ATCC 33574 in *E. coli*. (B) Manhattan plots of spacer depletions mapped over 45 genes and aggregated across normalized gene distance for full *P. buccae* VI-B2 locus (left) and *cas13b* (right), with non-targeting spacers in gray, safely depleted (>5σ above mean depletion of non-targeting spacers) spacers above blue line, and strongly depleted (top 1% depleted) spacers above red line. For the full locus, 36,141 targeting spacers and 859 non-targeting spacers passed QC filter. Of the targeting, 370 are strongly depleted and 8065 are safely depleted. For *cas13b* alone, 41,126 targeting spacers and 824 non-targeting spacers passed QC filter. Of the targeting, 419 are strongly depleted and 3295 are safely depleted. (C) Sequence weblogos of strongly depleted *P. buccae* spacers, revealing double-sided PFS (protospacer flanking sequence).

To further explore the ability of the small accessory proteins to modulate Cas13b activity, we tested if Csx27 can also repress PbCas13b using the MS2 drop plaque assay. Cells co-transformed with *pbcas13b* and *bzcxs27* expression plasmids exhibited a 10⁵ fold reduction in interference activity relative to *pbcas13b* expression plasmid and pUC19 empty vector, indicating that Csx27 exerts an inhibitory effect on PbCas13b (Figures 14D and 12B). The ability of Csx27 to modulate the interference activity of BzCas13b and PbCas13b suggests that it is a modular protein that can function across multiple VI-B loci.
Figure 14: Class 2 type VI-B systems are differentially regulated across two loci by Csx27 and Csx28. (A) Normalized PFS matrix, for *P. buccae* VI-B2 locus (left) and *pbcasl3b* (right). (B) MS2 Plaque drop assay for full *P. buccae* VI-B2 locus (left) and *pbcasl3b* (right). (C) Spacer depletions of *bzcasl3b* with and without *bzcsx27* (brown), as compared to *pbcasl3b* with and without *pbcsx28* (gold). (D) Fold resistance to MS2 infection for cells co-transformed with *pbcasl3b* and the indicated *csx* expression plasmid.

Discussion

Here we describe two RNA-targeting CRISPR Class 2 systems of subtype VI-B (VI-B1 and VI-B2), containing the computationally discovered RNA-guided RNase Cas13b. Type VI-B systems show several notable similarities to the recently characterized VI-A system. The single protein effectors of both systems cleave single-stranded RNA via HEPN domains, process their CRISPR arrays independent of the HEPN domains, and exhibit collateral RNase activity. Cas13b proteins, however, show only limited sequence similarity to Cas13a, and the common ancestry of the two type VI subtypes remains uncertain. Furthermore, the type VI-B systems differ from VI-A in several other ways, including the absence of both *cas1* and *cas2*, which are involved in spacer acquisition in other CRISPR-Cas systems (Mohanraju et al., 2016). The VI-B CRISPR arrays contain multiple spacers that differ among closely related bacterial strains, suggesting that acquisition does occur, either autonomously or possibly *in trans*, by recruiting Cas1 and Cas2 encoded in other CRISPR-Cas...
loci from the same genome. In trans utilization of adaptation modules of other CRISPR-Cas systems is compatible with the finding that the great majority of type VI-B systems co-occur in the same bacterial genome as other CRISPR-Cas loci that include cas1 and cas2 genes; conceivably, the three VI-B-carrying genomes that lack adaptation modules have lost them recently. Additionally, VI-B systems differ from VI-A systems by the presence of the small accessory proteins Csx27 (VI-B1 systems) and Csx28 (VI-B2 systems), which exhibit opposing regulatory effects on Cas13b activity.

Repression of Cas13b by Csx27 in VI-B1 systems could be part of an important regulatory mechanism of phage interference. The ability of Csx27 to repress Cas13b activity may be a general property, as we found that it can also repress PbCas13b (subtype VI-B2). In the case of type VI-B2 systems, Csx28 might enhance the collateral activity of Cas13b to inactivate numerous transcripts of invading bacteriophages or to promote programmed cell death. Both Csx27 and Csx28 contain predicted long, hydrophobic α-helices that might enable them to interact physically with Cas13b, but this remains to be determined. We did not find homologs of Csx27 or Csx28 encoded in any CRISPR-Cas loci other than type VI-B loci, suggesting that, at least in the CRISPR-Cas context, these proteins might function in tight association with Cas13b.

As with previously characterized Class 2 CRISPR-Cas effectors, such as Cas9 and Cpf1, there is enormous potential to harness Cas13b for use as a molecular tool (Cong et al., 2013; Mali et al., 2013; Wright et al., 2016). A holistic understanding of the factors that affect target selection is essential to the success of any such tools, particularly those that target RNA, where secondary structure will likely impact activity. We therefore developed an E. coli essential gene screen to explore the targeting rules of Cas13b more fully. This E. coli screen offers several advantages by increasing the number of guides testable in a single experiment to explore how diverse spacer and flanking sequences may affect Cas13b activity. This screen revealed a double-sided PFS in VI-B systems, which may give insight into Cas13b protein-RNA interactions, and could help improve specificity by expanding sequence targeting constraints (Ran et al., 2015).

The characterization of Cas13b and other RNA-targeting CRISPR systems raises the prospect of a suite of precise and robust in vivo RNA manipulation tools for studying a wide range of biological processes (Abil and Zhao, 2015; Filipovska and Rackham, 2011; Mackay et al., 2011). The ability of Cas13b to process its own CRISPR array could be extended to multiplex transcriptome engineering. In addition, the VI-B functional long direct repeats could be altered to incorporate stem loops akin to the Cas9-SAM system (Konermann et al., 2015). Like Cas9 and Cpf1, Cas13a and Cas13b may be utilized for complementary applications in science and technology.
Acknowledgements

We would like to thank R. Belliveau for overall research support; R. Macrae for critical reading of the manuscript; and the entire Zhang laboratory for support and advice. D.B.T.C. is supported by award number T32GM007753 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the NIH. I.M.S. is supported by the Simons Center for the Social Brain. J.S.G. is supported by a D.O.E. Computational Science Graduate Fellowship. O.A.A. is supported by a Paul and Daisy Soros Fellowship, a Friends of the McGovern Institute Fellowship, and the Poitras Center for Affective Disorders. K.S.M. and E.V.K. and S.S. are supported by the intramural program of the US Department of Health and Human services (to the National Library of Medicine). S.S. is also supported by a Skoltech-MIT Next Generation Program grant to K. Severinov and F.Z. F.Z. is supported by the NIH through NIMH (5DP1-MH100706 and 1R01-MH110049); NSF; the New York Stem Cell Foundation; the Allen Distinguished Investigator Program, through The Paul G. Allen Frontiers Group; the Simons and Vallee Foundations; the Howard Hughes Medical Institute; the Skoltech-MIT Next Generation Program; James and Patricia Poitras and the Poitras Center for Affective Disorders; Robert Metcalfe; and David Cheng. F.Z. is a New York Stem Cell Foundation-Robertson Investigator. A patent application has been filed related to this work and the authors plan to make the reagents widely available to the academic community through Addgene and to provide protocols and software tools via the Zhang lab website (www.genome-engineering.org) and GitHub (www.github.com/fengzhanglab).
References:


Experimental Model And Subject Details

*P. buccae* (Holdeman et al.) Shah and Collins: *P. buccae* was not grown in this study.

*B. zoohelcum* (Holmes et al.) Vandamme et al.: *B. zoohelcum* ATCC 43767 was grown in ATCC medium 44 (Brain Heart Infusion broth) at 37°C at 250 rpm overnight.

*E. coli* (Migula) Castellani and Chalmers (C3000): *E. coli* was grown in LB at 37°C at 250 rpm overnight.

**One Shot Stbl3™** *E. coli*: *E. coli* was grown in LB at 37°C at 250 rpm overnight.

**NEB® 10-beta Competent E. coli (High Efficiency)**: NEB® 10-beta Competent *E. coli* was transformed on LB agar at 37°C overnight.

**MegaX DH10B™ T1R Electrocomp™ Cells**: NEB® 10-beta Competent *E. coli* was transformed on LB agar at 37°C overnight.

**One Shot® BL21(DE3)pLysE Chemically Competent E. coli**: The BzCas13b expression construct was transformed into One Shot® BL21(DE3)pLysE (Invitrogen) cells. 25 mL of 6hr growing culture were inoculated into 2 liters of Terrific Broth 4 growth media (12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L K2HPO4, 2.2 g/L KH2PO4, Sigma). Cells were then grown at 37°C to a cell density of 0.6 OD600, and then SUMO-BzCas13b expression was induced by supplementing with IPTG to a final concentration of 500 μM. Induced culture was grown for 16-18 hours before harvesting cell paste, which was stored at -80°C until subsequent purification. For each BzCas13b mutant, 1 L of Terrific Broth was used to generate cell paste and all other reagents were scaled down accordingly. Protein purification was performed using the same protocol as wild-type Cas13b. PbCas13b was cloned into the same pET based vector and purified using a similar protocol as BzCas13b with the following differences: cells were grown at 21°C for 18 hours.
Method Details

Computational Sequence Analysis

From complete compiled Ensembl Release 27 genomes (Yates et al., 2016), CRISPR repeats were identified using PILER-CR (Edgar, 2007). Proteins within 10kb of identified CRISPR arrays were clustered into loci, with loci rejected if more than one protein of size 700 amino acids or larger or if either Cas1 or Cas2 were present. For candidate Class 2 effectors, only proteins in these remaining loci of size 900aa to 1800aa were selected. These candidate effectors were subjected to the BLASTP (Camacho et al., 2009) search against the NCBI non-redundant (NR) protein sequence database with an E-value cutoff of 1e-7. All discovered proteins were then grouped into putative families via a nearest-neighbor grouping with the same E-value cutoff. Only putative families with at least ten candidate effectors and more than 50% of candidate effectors within 10kb of CRISPR arrays were considered. HHpred (Remmert et al., 2012) and existing CRISPR locus classification rules (Makarova et al., 2015) were used to classify each family, leaving Cas13b as the only unclassified family. Additional Cas13b proteins in the family were found through a nearest-neighbor search of previously discovered Csx27/Csx28 against the NCBI non-redundant (NR) protein sequence database with an E-value cutoff of 1e-7, and then by searching in genomes within 1kb of any newly discovered Csx27/Csx28. Within this Cas13b family, truncated or suspected partially sequenced effectors were discarded, leaving 105 loci, and 81 with a unique protein accession number in the NCBI non-redundant (NR) protein sequence database. Multiple sequence alignments on these 81 proteins (as well as the accessory Csx27 and Csx28 proteins) were performed using BLOSUM62 (Henikoff and Henikoff, 1992) to identify the HEPN domains and to sort the loci into phylogenetic trees. Loci represented in the tree of 81 non-redundant proteins were selected first for annotated Csx27/Csx28 within 1kb of Cas13b, and next for annotated CRISPR array within 10kb of Cas13b. Vienna RNAfold (Lorenz et al., 2011) was used to predict the secondary structure of each direct repeat, whose transcriptional orientation was chosen as identical to that of Cas13b in its locus. CRISPRTraget (Biswas et al., 2013) was used to search the spacers in each locus against NCBI phage and plasmid genomes. Weblogos were generated for all unique direct repeats and protospacer flanking sequences (Crooks et al., 2004). TMHMM Server v. 2.0 (Moller et al., 2001) was used to predict the transmembrane helices in Csx27 and Csx28.
Nucleic Acid Preparation

For *in vitro* synthesis of RNA, a T7 DNA fragment must be generated. To create T7 DNA fragments for crRNAs, top and bottom strand DNA oligos were synthesized by IDT. The top DNA oligo consisted of the T7 promoter, followed by the bases GGG to promote transcription, the 30-nt target and then direct repeat. Oligos were annealed together using annealing buffer (30 mM HEPES pH 7.4, 100 mM potassium acetate, and 2 mM magnesium acetate). Annealing was performed by incubating the mixture for 1 minute at 95°C followed by a -1°C/minute ramp down to 23°C. To create ssRNA targets, short targets (Trunc2, 3, 4) were synthesized as top and bottom strand oligos containing the T7 promoter. For long ssRNA targets (E1, E2, S and L CRISPR Arrays), DNA primers with a T7 handle on the forward primer were ordered and the DNA fragment was amplified using PCR. T7 DNA constructs for RNA generation without body labeling were incubated with T7 polymerase overnight (10-14 hours) at 30°C using the HiScribe T7 Quick High Yield RNA Synthesis kit (New England Biolabs). Body-labeled constructs were incubated with Cyanine 5-UTP (Perkin Elmer) and incubated with T7 polymerase overnight at 30°C using the HiScribe T7 High Yield RNA Synthesis kit (New England Biolabs). For a complete list of crRNAs and target ssRNAs used in this study, visit http://www.sciencedirect.com/science/article/pii/S1097276516308668. 5’ end labeling was accomplished using the 5’ oligonucleotide kit (VectorLabs) and with a maleimide-IR800 probe (LI-COR Biosciences). 3’ end labeling was performed using a 3’ oligonucleotide labeling kit (Roche) and Cyanine 5-ddUTP (Perkin Elmer). RNAs were purified using RNA Clean and Concentrator columns™-5 (Zymo Research). Body-labeled dsRNA substrates were prepared by T7 DNA fragments for the bottom and top RNA strand. After synthesis, 1.3-fold excess of non-labeled bottom strand ssRNA was added and re-annealed to ensure the top strand would be annealed to a bottom strand by incubating the mixture for 1 minute at 95°C followed by a -1°C/minute ramp down to 23°C.

BzCas13b Protein Purification

The mammalian codon-optimized gene for Cas13b (*B. zoohelcum*) was synthesized (GenScript) and inserted into a bacterial expression vector (6x His/Twin Strep SUMO, a pET based vector received as a gift from Ilya Finkelstein) after cleaving the plasmid with the BamHI and NotI restriction enzymes and cloning in the gene using Gibson Assembly® Master Mix (New England Biolabs). The BzCas13b expression construct was transformed into One Shot® BL21(DE3)pLysE (Invitrogen) cells. 25 mL of 6hr growing culture were inoculated into 2 liters of Terrific Broth 4 growth media (12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L K2HPO, 2.2 g/L KH2PO4, Sigma). Cells were then
grown at 37°C to a cell density of 0.6 OD600, and then SUMO-BzCas13b expression was induced by supplementing with IPTG to a final concentration of 500 μM. Induced culture was grown for 16-18 hours before harvesting cell paste, which was stored at -80°C until subsequent purification. Frozen cell paste was crushed and resuspended via stirring at 4°C in 500 mL of Lysis Buffer (50mM NaH₂PO₄ pH 7.8, 400mM NaCl) supplemented with protease inhibitors (cOmplete, EDTA-free, Roche Diagnostics Corporation) and 1250U of benzonase (Invitrogen). The resuspended cell paste was lysed by a LM20 microfluidizer at 18,000 psi (Microfluidics). Lysate was cleared by centrifugation at 10,000g for 1 hour. Filtered lysate was incubated with StrepTactin Sepharose High Performance (GE Healthcare Life Sciences) at 4°C for 1 hour with gentle agitation, and then applied to an Econo-column chromatography column (Bio-Rad Laboratories). Resin was washed with Lysis Buffer for 10 column volumes. One column volume of fresh Lysis Buffer was added to the column and mixed with 10 units of SUMO protease (Invitrogen) and incubated overnight. The eluate was removed from the column, SUMO cleavage was confirmed by SDS-PAGE and BlueFast protein staining (Eton Bioscience), and the sample was concentrated via Centrifugal Filter Unit to 2 mL. Concentrated sample was loaded onto a HiTrap Heparin HP column (GE Healthcare Life Sciences) via FPLC (AKTA Pure, GE Healthcare Life Sciences) and eluted over a gradient with an elution buffer with salt concentration of 1.2 M. The resulting fractions were tested for presence of BzCas13b protein by SDS-PAGE; fractions containing BzCas13b were pooled, and concentrated via Centrifugal Filter Unit to 1 mL. Concentrated sample was loaded a gel filtration column (HiLoad 16/600 Superdex 200, GE Healthcare Life Sciences) via FPLC (AKTA Pure, GE Healthcare Life Sciences) with buffer 500 mM NaCl, 50 mM Tris-HCl pH 7.5, 1 mM DTT.

**BzCas13b HEPN Mutant Protein Purification**

Alanine mutants at each of the HEPN catalytic residues were generated using the Q5® site-directed mutagenesis kit (New England Biolabs) and transformed into One Shot® BL21(DE3)pLysE cells (Invitrogen). For each mutant, 1 L of Terrific Broth was used to generate cell paste and all other reagents were scaled down accordingly. Protein purification was performed using the same protocol as wild-type Cas13b.

**PbCas13b Protein Purification**

PbCas13b (*Prevotella buccae*) was cloned into the same pET based vector and purified using a similar protocol as BzCas13b. However, these bacterial cells were grown at 21°C for 18 hours. The frozen cell paste was resuspended into 500 mM NaCl, 50 mM HEPES 7.5 and 2 mM DTT prior to
breaking cells in the microfluidizer. The Superdex 200 column was run in 500 mM NaCl, 10 mM HEPES 7.0, and 2 mM DTT.

**Nuclease Assay**

Nuclease assays were performed with equimolar amounts of end-labeled or body-labeled ssRNA target, purified protein, and crRNA, for targeted ssRNA cleavage. For CRISPR array cleavage, protein was supplied in a four times molar excess of the CRISPR array. Reactions were incubated in nuclease assay buffer (10 mM TrisHCl pH 7.5, 50 mM NaCl, 0.5 mM MgCl₂, 20U SUPERase In™ (ThermoFisher Scientific), 0.1% BSA). Reactions were allowed to proceed at 37°C for times specified in the figure legends. After incubation, samples were then quenched with 0.8U of Proteinase K (New England Biolabs) for 15 minutes at 25°C. The reactions were mixed with equal parts of RNA loading dye (New England Biolabs) and denatured at 95°C for 5 minutes and then cooled on ice for 2 minutes. Samples were analyzed by denaturing gel electrophoresis on 10% PAGE TBE-Urea (Invitrogen) run at 45°C. Gels were imaged using an Odyssey scanner (LI-COR Biosciences).

**EMSA Assay**

For the Electrophoretic Mobility Shift Assay (EMSA), binding experiments were performed with a series of half-log complex dilutions (crRNA and BzCas13b) from .594 to 594 nM. Binding assays were performed in nuclease assay buffer (without MgCl₂) supplemented with 10 mM EDTA to prevent cutting, 5% glycerol, and 5μg/mL heparin in order to avoid non-specific interactions of the complex with target RNA. Protein was supplied at two times the molar amount of crRNA. Protein and crRNA were preincubated at 37°C for 15 minutes, after which the 5'-labeled target was added. Reactions were then incubated at 37°C for 10 minutes and then resolved on 6% PAGE TBE gels (Invitrogen) at 4°C (using 0.5X TBE buffer). Gels were imaged using an Odyssey scanner (LI-COR Biosciences). Gel shift of the RNA targets was quantified from an EMSA gel using ImageJ (Wayne Rasband, NIH) and plotted in GraphPad Prism version 7 (GraphPad Software, La Jolla California USA). Line regression was performed in Prism 7 using nonlinear fit with one-site binding hyperbola. Kᵦ values are calculated by GraphPad Prism based on regression analysis of data.
RFP-Tagged Protein Fluorescent Imaging

One Shot Stbl3 Chemically Competent *E. coli* were transformed with plasmids containing RFP (negative control) or RFP fused to the N- or C-terminus of Csx27 of *B. zoohelcum* or Csx28 of *P. buccae*. Clones were cultured up in 5mL of antibiotic LB overnight, then spun down at 5000g and resuspended in PBS with 1% methanol-free formaldehyde. After 30 minutes fixation, cells were washed once with PBS and then diluted 1:2 in PBS. 5uL of sample was pipetted onto a silane-coated slide, which was covered with a coverslip. Fluorescent imaging was performed in a 63x objective microscope with oil immersion.

Bacterial RNA-Sequencing

RNA was isolated and prepared for sequencing using a modification of a previously described protocol (Heidrich et al., 2015; Shmakov et al., 2015). RNA was isolated from 5 mL of stationary phase of bacterial cultures by resuspending pelleted cells in 1mL of TRIzol (ThermoFisher Scientific) and then homogenizing with 300 uL zirconia/silica beads (BioSpec Products) in a BeadBeater (BioSpec Products) for 7 1-minute cycles. 200 uL of chloroform was added to the homogenized sample and then samples were centrifuged for 15 min. (12000xg, 4°C). The aqueous phase was then used for input into the Direct-Zol RNA miniprep kit (Zymo). Purified RNA was DNase treated with TURBO DNase (Life Technologies) and 3’ dephosphorylated/5’ phosphorylated with T4 Polynucleotide Kinase (New England Biolabs). rRNA was eliminated using the bacterial Ribo-Zero rRNA removal kit (Illumina). Next, RNA was treated with RNA 5’ polyphosphatase (Epicentre Bio) to convert 5’-triphosphates to 5’-monophosphates for adapter ligation. Samples were then polyA tailed with *E. coli* Poly(A) polymerase (New England Biolabs), and a 5’ RNA Illumina sequencing adapter ligated to cellular RNA using T4 RNA Ligase 1 (ssRNA ligase) (New England Biolabs). RNA was reverse transcribed using AffinityScript cDNA synthesis kit (Agilent Technologies) and an oligo-dT primer. cDNA was amplified with Herculase II polymerase (Agilent Technologies) and barcoded primers. The prepared cDNA libraries were sequenced on a MiSeq (Illumina)

For RNA sequencing of native *B. zoohelcum* ATCC 43767, we repeated the experiment with a modified protocol, omitting RNA 5’ polyphosphatase prior to 5’ adapter ligation, to promote enrichment of processed transcripts originating from the CRISPR array. For heterologous *P. buccae* ATCC 33574 RNA sequencing in *E. coli*, we cloned the locus into pACYC184. Reads from each sample were identified on the basis of their associated barcode and aligned to the appropriate RefSeq reference genome using BWA (Li and Durbin, 2009). Paired-end alignments were used to extract
entire transcript sequences using Galaxy (https://usegalaxy.org), and these sequences were analyzed using Geneious 8.1.8.

**E. coli Essential Gene Screen Experiment**

The intersection of two *E. coli* DH10B strain essential gene studies (Baba et al., 2006; Gerdes et al., 2003) was taken, and further pared down to 45 genes by only selecting genes exclusive to their respective operons. Over these 45 genes 54,600 spacers were designed to tile at single resolution across the coding region, as well as to extend 60 nt into the 5’ UTR and 3’ UTR. In addition, 1100 non-targeting, pseudorandomly generated spacers with no precise match to the *E. coli* DH10B strain genome were added to the library as a non-targeting negative control. The library of spacers was cloned into a *B. zoohelicum* or *P. buccae* direct repeat-spacer-direct repeat backbone containing a chloramphenicol resistance gene using Golden Gate Assembly (NEB) with 100 cycles, and then transformed over five 22.7cm x 22.7cm chloramphenicol LB Agar plates. Libraries of transformants were scraped from plates and DNA was extracted using the Macherey-Nagel Nucleobond Xtra Midiprep Kit (Macherey-Nagel). 50ng of library plasmid and equimolar gene plasmid containing an ampicillin resistance gene (*bzcas13b, bzcas13b & bzcpx27, pbcas13b, pbcas13b & pbcsx28*, empty vector pBR322) were transformed into MegaX DH10B™ T1R Electrocomp™ Cells (ThermoFisher) according to manufacturer’s protocol, with four separate 22.7cm x 22.7cm carbenicillin-chloramphenicol LB Agar plates per bioreplicate, and three bioreplicates per condition (twelve transformations total per condition). Eleven hours post-transformation, libraries of transformants were scraped from plates and DNA extracted using the Macherey-Nagel Nucleobond Xtra Maxiprep Kit (Macherey-Nagel).

**E. coli Essential Gene Screen Analysis**

Prepared DNA libraries were sequenced on a NextSeq (Illumina), with reads mapped to the input library of spacers. Spacer depletions were calculated as the read abundance of a spacer in the empty vector condition divided by read abundance in each gene plasmid condition. Mean depletions over three bioreplicates were calculated. We imposed a two-step quality-control filter on the data: a maximum coefficient of variation of 0.2 for depletion over three bioreplicates, and a minimum spacer read abundance of 1/3*N* in each bioreplicate, where *N* = 55,700. Weblogos of the strongly depleted (top 1% depleted) spacers were generated (Crooks et al., 2004), and from each identified PFS, heatmaps of the ratio of moderately depleted (referred to as safely depleted) (>5σ above mean depletion of non-targeting spacers) spacers to all spacers in the screen were generated.
For spatial analysis via empirical cumulative distribution functions, safely depleted spacers were aggregated across the first or last 250 nt of genes.

For secondary structure analysis, we utilized the RNA accessibility model from Vienna RNAplfold (Bernhart et al., 2006).

**Kanamycin Validation Screen Experiment**
A total of 160 kanamycin-targeting spacers was selected, 42 of which contain both PFS rules, 47 of which contain one rule, and 71 of which contain no rules, to which 162 non-targeting control spacers were added. The library of spacers was cloned into either a **bzcas13b** and *B. zoohelcum* direct repeat-spacer-direct repeat backbone or simply a *B. zoohelcum* direct repeat-spacer-direct repeat backbone containing a chloramphenicol resistance gene using Golden Gate Assembly (NEB) with 100 cycles, and then transformed over one 22.7cm x 22.7cm carbenicillin LB Agar plate. The two cloned library plasmids were then re-transformed with over a 22.7cm x 22.7cm chloramphenicol LB Agar plate or a 22.7cm x 22.7cm kanamycin-chloramphenicol LB Agar plate. Libraries of transformants were scraped from plates and DNA extracted using the Qiagen Plasmid Plus Maxi Kit (Qiagen). 100 ng of library DNA and 100 ng of pMAX-GFP (Lonza), containing a kanamycin resistance gene were added to 50 ul of chemically competent 10-beta cells (NEB) and transformed according to the manufacturer’s protocol.

**Kanamycin Validation Screen Analysis**
Prepared DNA libraries were sequenced on a NextSeq (Illumina), with reads mapped to the input library of spacers. For normalizing the abundance of spacers of two separate clonings, the corrected experimental read abundance of a given spacer was calculated as the read abundance of that spacer in the **bzcas13b** plasmid (kanamycin-chloramphenicol transformation) multiplied by the ratio of the read abundance ratio of that spacer in the non-**bzcas13b** plasmid (chloramphenicol-only transformation) to the read abundance ratio of that spacer in the **bzcas13b** plasmid (chloramphenicol-only transformation).

**MS2 Phage Drop Plaque Assay**
Individual spacers for bacteriophage MS2 interference were ordered as complementary oligonucleotides containing overhangs allowing for directional cloning in between two direct repeat sequences in vectors containing *cas13b*. 10 uM of each complementary oligo were annealed
in 10X PNK Buffer (NEB), supplemented with 10mM ATP and 5 units of T4PNK (NEB). Oligos were incubated at 37°C for 30 min., followed by heating to 95°C for 5 min. and then annealed by cooling to 4°C. Annealed oligos were then diluted 1:100 and incubated with 25 ng of Eco31I digested cas13b vector in the presence of Rapid Ligation Buffer and T7 DNA ligase (Enzymatics). Individual plasmids were prepared using the QIAPrep Spin Miniprep Kit (Qiagen), sequence confirmed and then transformed into C3000 (ATCC 15597) cells made competent using the Mix & Go E. coli Transformation Kit (Zymo). In the case of experiments using csx27 or csx28, C3000 cells harboring csx plasmids were made competent and then transformed with cas13b direct repeat-spacer-direct repeat plasmids. Following transformation, individual clones were picked and grown overnight at 37°C in LB containing the appropriate antibiotics. The following morning, cultures were diluted 1:100 and grown to an OD_{600} of 2.0 by shaking at 37°C with 5% CO2 at 250 rpm, then mixed with 4mL of antibiotic containing Top Agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride, 5 g/L agar) and poured on to LB-antibiotic base plates. 10 fold serial-dilutions of MS2 phage (ATCC 15597-B1) were made in LB and then spotted onto hardened top agar with a multi-channel pipette. Plaque formation was assessed after overnight incubation of the spotted plates at 37°C. For assessing interference levels, samples were blinded using a key and the lowest dilution of phage at which plaque formation occurred was compared to a pACYC condition by eye, where the lowest dilution of MS2 that formed plaques on pACYC was set to 1. The lowest dilution of phage used was 1.05*10^{8} pfu.

**DNA Interference Assay**

A 34-nt target sequence consisting of a 30-nt protospacer and a permissive PFS (5'-G, 3'-AAA) was cloned into pUC19 in two locations. For the transcribed target, the target sequence was cloned into the coding strand of the bla gene, in frame immediately after the start codon, with the G of the start codon serving as the 5’ PFS. For the non-transcribed target the identical target sequence (protospacer and PFS) were cloned into the AatII site of pUC19, so that the protospacer appears on the non-transcribed strand with respect to the pBla and pLac promoters. To determine interference, 25 ng of the ampicillin resistant target plasmid and 25 ng of the chloramphenicol resistant bzcas13b or empty vector (pACYC) were added to 5 µL of NovaBlue GigaSingle cells (Novagen). The cells were incubated for 30 minutes on ice, heatshocked for 30 seconds at 42°C and incubated on ice for 2 minutes. Then, 95 µL of SOC was added to cells and they were incubated with shaking at 37°C for 90 minutes, before plating the entire outgrowth (100 µL) on plates containing both chloramphenicol and ampicillin.
Quantification And Statistical Analysis

MS2 Interference Assay-HEPN Mutants
Three bioreplicates of the MS2 interference assay were performed for the fold resistance quantification. For assessing interference levels, samples were blinded using a key and the highest dilution of phage at which plaque formation occurred was compared to a vector only condition by eye, where the highest dilution of MS2 that formed plaques on pACYC was set to 1. The error bars are the standard deviation of the fold-resistance for each condition.

DNA Interference Assay
Three bioreplicates of the DNA interference assay were performed for the colony forming unit quantification. The mean values were taken from the mean of number of colony forming units from a standard colony forming unit count, and the standard deviation values accordingly from the same standard count.

E. coli Essential Gene Screen
Spacer depletions from the screen were calculated as the read abundance of a spacer in the empty vector condition divided by read abundance in each gene plasmid condition. Mean depletions over three bioreplicates were calculated. We imposed a two-step quality-control filter on the data: a maximum coefficient of variation of 0.2 for depletion over three bioreplicates, and a minimum spacer read abundance of $1/3N$ in each bioreplicate, where $N = 55,700$. This reduced the number of guides represented from $N$ to approximately 30,000-40,000.

K_D Calculations
Gel shift of the RNA targets was quantified from an EMSA gel using ImageJ (Wayne Rasband, NIH) and plotted in GraphPad Prism version 7 (GraphPad Software, La Jolla California USA). Line regression was performed in Prism 7 using nonlinear fit with one-site binding hyperbola. K_D values are calculated by GraphPad Prism based on analysis of regression data.

Data And Software Availability

Data Resources
Data have been deposited in the following resources:


Table S1 | All crRNAs, nucleic acid targets, and primers used in biochemical experiments.

Table S2 | All Cas13b plasmids used in this study.

Table S3 | *E. coli* essential genes represented in *E. coli* essential gene screen library of spacers.

Table S4 | Spacers from *E. coli* essential gene screen.

Table S5 | Spacers from kanamycin validation screen.

Table S6 | Spacers targeting MS2 and pBLA plasmids.

Table S7 | EMSA raw data.
CHAPTER 3

FUTURE DIRECTIONS

BASIC BIOLOGY TO INVESTIGATE AND POTENTIAL ENGINEERING APPLICATIONS OF CAS13B
The identification and characterization of Cas13b increases the known diversity of CRISPR-Cas adaptive immune systems and provides an additional single effector Cas enzyme that may be harnessed for tool development. Cas13b represents the second unique single effector RNA-guided RNAse that has both similar and unique properties compared to the other single effector Cas RNase C2c2. The discovery of Cas13b following characterization of C2c2 is analogous to the discovery of Cpf1 after Cas9 characterization. Both Cpf1 and Cas9 have been harnessed as effective genome editing tools that can be used for similar or complementary applications and Cas13b and C2c2 may later exhibit a similar relationship. In order to harness Cas13b for transcriptome engineering purposes, it is important to better understand the basic biological mechanisms of Cas13b activity and then use this knowledge to develop useful RNA tools.

Basic Biology to Investigate Related to Cas13b

There are many fundamental properties of Cas13b that are currently unknown and there are many intriguing questions that could be investigated to better understand Cas13b function. What are the mechanisms of Cas13b binding, Cas13b targeted cleavage activity, and Cas13b collateral effect? How specific is Cas13b? How do the small proteins Csx27 and Csx28 modulate Cas13b interference? Lastly, what is the mechanism of acquisition that occurs in Cas13b loci? Thoughts on these questions will be elaborated on below.

The mechanism by which Cas13b first binds to its target RNA, cuts its target, and then undergoes collateral nonspecific RNAse activity is currently unknown. It is possible that the Cas13b binding mechanism operates similarly to other Cas enzymes. For example, Cas9 has a known process to first promiscuously bind DNA at the PAM sequence, unwind the DNA from the seed region next to the PAM to the end of guide RNA sequence, and with sufficient pairing, cleavage occurs (Sternberg et al., 2014; Wu et al., 2014). Based on biochemical studies shown in Chapter 2, Cas13b must be able to specifically recognize RNA versus DNA, search the sequence, and initiate binding to its RNA target. This is specifically supported by the EMSA data: WT Cas13b exhibits nonspecific RNA binding affinity (188-1000nM \(K_D\)) and higher affinity binding to RNA with its target sequence (33nM \(K_D\)). Analogous to the Cas9 PAM binding, it is possible that the 5' and/or 3' PFS sequences have roles in the initial Cas13b binding process. Cas13b may recognize 5'D or 3'NAN/NNA first and then direct RNA base pairing between the crRNA and the target RNA proximal to the 5' or 3' end. However, data from C2c2 studies suggests that the seed region is in the central part of the guide (Abudayyeh et al., 2016). Based on the number of functional similarities between
C2c2 and Cas13b, it is possible that Cas13b may also have a central seed region. If binding nucleation is independent from the PFS sequences, it is interesting to speculate on the role of the PFS sequences. It is hypothesized that the non-permissive PFS base across from the DR sequence in the crRNA is determined by the protospacer flanking base that is able to form a complementary base-pair with the first base of the DR (see Figure 1). For C2c2, its 3' PFS is H (or not G) and the protospacer adjacent DR base in the crRNA is C (the crRNA orientation is 5'-DR-spacer-3')(Abudayyeh et al., 2016) whereas the Cas13b 5'PFS is D (or not C) and the protospacer adjacent DR base in the crRNA is G. This hypothesis is supported by the C2c2 crystal structure, which illustrates a kink in this region that would be disrupted if base pairing between the DR and PFS would occur (Liu et al., 2017). Cas13b has also been identified to have a 3'PFS preference of NAN/NNA. One hypothesis is that a flexible A in either position 2 or 3 from the protospacer target participates in a stabilizing protein:RNA interaction with Cas13b to promote cleavage activity.

**Figure 1: PFS sequence requirements opposite the first DR base.** This figure illustrates the hypothesis, which suggests that the PFS opposite the first DR base is based on the DR base. The figure shows a RNA target and the protospacer (blue) targeted either by Cas13b or C2c2. The PFS opposite the DR is shown in orange.

After Cas13b binds to its target RNA, it would be ideal to understand how targeted RNA cleavage is accomplished and the specificity of Cas13b targeted RNA cleavage. C2c2, like all other single effector Class 2 Cas proteins, is a bilobed structure in which its HEPN domains associate together when bound to its target (Liu et al., 2017). It is probable that Cas13b also undergoes a conformational shift to bring together its HEPN domains, perhaps like joining one's index finger to thumb. In addition to conformational changes, Cas13b cleavage specificity is key to understand. Cas9 is known to tolerate some mismatches and indels (Kim et al., 2015; Tsai et al., 2015) and WT Cas13b may also tolerate mutations. To test specificity, single synthesized targets with specific mismatches could be tested for biochemical cleavage at a low throughput levels. Although pooled higher throughput screening assays would be more efficient, the presence of the collateral effect would likely confound results.

The mechanism by which Cas13b undergoes the collateral effect is worthwhile to investigate and structural analyses could provide insight. The C2c2 crystal structure illustrates that its catalytic active site is exposed to the surface, hypothesized to interact with RNAs with a variety...
of secondary structures (Liu et al., 2017). The openness of the catalytic residues may explain nonspecific RNAse activity because after targeted cleavage, the targeted RNA may dissociate and the active catalytic site may remain exposed to other RNAs. Cas13b may have a similar exposed or otherwise accessible active site. Crystal structures of Cas13b with and without a crRNA and target RNA could show the structure of the active site before and after targeted RNA cleavage. Alternatively, protein structure assays such as FRET could be employed to analyze domain reorganization. Although the phenomenon of the collateral effect itself is interesting, it is just as interesting to speculate why a targeted RNAse would also have subsequent activity as a nonspecific RNAse. The collateral effect shows parallels to other bacterial immunity systems such as abortive infection, which eliminates all RNAs in the system to reduce the chance of bacteriophage survival. Additionally, the collateral effect parallels the activity of the non-targeted RNase enzymes Csm6 or Csx1 involved in Type III CRISPR-Cas systems to mediate interference. Perhaps a generalized RNase activity during viral infection provides a critical survival advantage to the population of bacterial hosts.

Small proteins Csx27 and Csx28

One key distinction between Cas13b and C2c2 is that the small proteins Csx27 or Csx28 can modulate Cas13b interference activity, but the mechanism by which these proteins modulate interference is unknown. One key piece of information that is unknown is whether or not Csx27 or Csx28 directly interact with either Cas13b or the crRNA. To test for direct interaction between Cas13b and either Csx27 or Csx28, a protein complex immunoprecipitation assay could be completed. Since the small Csx27 or Csx28 proteins are predicted to be membrane associated, and if stable direct interactions occur between Csx27 or Csx28 and Cas13b, these small membrane proteins may recruit Cas13b to the membrane, which is the first area of entry of bacteriophage nucleic acid. Alternatively, it is possible that Csx27 and Csx28 are independent interference mediators or that Csx27 and Csx28 interact with the crRNA only to mediate interference. If Csx27 and Csx28 are independent interference mediators, then expressing these proteins alone should mediate some level of interference without expressing Cas13b or the crRNA. It is also possible that Csx27 and Csx28 interact specifically with the crRNA to mediate interference. Binding assays between these proteins and the crRNA would be critical to illustrate this effect. Lastly, Csx28 may participate in RNA cleavage if it contains a bonafide active HEPN motif, perhaps dimerizing with a second Csx28 protein.
Cas13b Adaptation

Cas13b was not identified in previous computational searches for novel Class 2 CRISPR-Cas systems because it is lacking Cas1, the only known protein required for adaptation. Without Cas1 and Cas2 in Cas13b loci, there are two main possibilities to explain acquisition: acquisition either occurs using Cas1 proteins from other CRISPR-Cas loci in the genome or it has a unique form of adaption without Cas1, perhaps involving Csx27 or Csx28. A recent report by Laanto et al. provides evidence supporting an acquisition mechanism that utilizes Cas1 proteins in trans expressed by other non Type VI CRISPR-Cas loci in the genome. In brief, Laanto et al. conducted a long-term field sampling study with the bacterial species *Flavobacterium columnare*, which has a Type II locus and Type VI-B locus. The study illustrated that both CRISPR loci acquired spacers over time matching sequenced phages in the environment. Notably, all fifteen spacers in the VI-B CRISPR locus targeted ORFs on the phage coding strand, and the authors provided some, but in my opinion weaker, evidence of sequence similarity in the predicted PAM sequence for *F. columnare* Cas9. Overall, this paper provides evidence that Type VI loci can undergo acquisition specifically for RNA coding viral sequences and suggests that trans expressed acquisition machinery assist in this process. This data is also supported by our findings that most type VI loci have other diverse CRISPR-Cas loci in their genome. Although we found three exceptions of species of bacteria with exclusively a Cas13b Type VI system, the lack of a second CRISPR-Cas locus could be explained by an incompletely sequenced genome, misannotation, or loss of acquisition machinery or CRISPR-Cas system at some point in the species history. Lastly, since the *F. columnare* VI-B locus does not have either csx27 or csx28 genes in its locus, this study provides evidence that Csx27 and Csx28 are not involved in the adaptation process.

Engineering applications of Cas13b

Like other Class 2 single effector Cas enzymes, Cas13b has the potential to direct cellular perturbations or development of other biological assays. Since Cas13b acts on RNA, it may be developed for transcriptome modulation of cells through RNA knockdown, RNA editing, or epitranscriptomics and can be used to develop assays including RNA imaging or RNA diagnostics. The properties of Cas13b itself could also be modulated through structure-guided mutagenesis to engineer new or altered functions. This section will further describe the engineering applications of Cas13b.
Targeting cellular RNA for transcriptome modulation represents an unrealized potential in genome engineering, but utilizing CRISPR-Cas enzymes is not the only proposed method for targeted RNA modulation. In total, three systems have the potential for targeted transcriptome engineering and they include the use of Pumilio proteins, argonaute proteins, or CRISPR-Cas systems. Pumilio/FBF (also known as PUF or Pum), function similarly to TAL effectors or zinc fingers and rely on protein engineering of modular domains that bind to specific RNA bases (Crittenden et al., 2002; Zamore et al., 1997). As mentioned before, protein engineering is more difficult than using a programmable RNA, but Pum systems have an additional key disadvantage. Reprogramming Pum domains for targeted RNA editing exhibits low target specificity due to flexible nature of RNA (Wang et al., 2013). Therefore, these systems have not become a widespread or simple method to modulate RNA. Besides Pum or CRISPR-Cas systems there exists another system, the Natronobacterium gregoryi Argonaute (NgAgo), which has a potential for targeted RNA modulation. NgAgo undergoes DNA-dependent RNA cleavage using a DNA guided endonuclease protein and has no PAM requirement (Sunghyeok et al.). This is a relatively new finding and has potential, but additional studies are required to better understand the mechanism and specificity. Additionally, NgAgo has some confounding conflicting studies (Lee et al., 2016) suggesting that NgAgo cleaves dsDNA (Gao et al., 2016) and further studies are needed to better understand its true target.

While Cas13b represents a new area of technology with promise for RNA modulation, it is worthwhile to note existing RNA modulating technologies and where Cas13b can make an impact. RNA knockdown can already be achieved in mammalian cells using RNAi pathways. However siRNA, shRNA, or morpholino antisense oligo technologies exhibit a wide variety of specificity with some that are very specific and some that are less specific (Cullen, 2006). While some shRNAs have been quite useful, one unmet need is directing RNA knockdown to the nucleus. SiRNAs only interact with cytoplasmic RNAs and by targeting RNAs in the nucleus, one could directly probe nuclear lincRNAs, nascent transcripts, or probe allele specific RNA expression. A table comparing RNAi with Cas13b is shown in Table 1.
RNAi | Cas13b
---|---
**Specificity** | Range of non-specific or specific depending on target | Unknown
**Delivery** | Viruses, liposomes, cationic polymers, siRNA conjugates, direct injection (Kanasty et al., 2013) | Likely plasmid based using transfection reagents or viral vectors, possibly purified protein and RNA
**RNA cleavage mechanism** | Association of siRNA with RNA-induced silencing complex and argonaute 2 | crRNA targeted cleavage by Cas13b enzyme and later collateral cleavage

| **Potential or current Applications** | Knockdown only | Knockdown, splicing regulation, translational regulation, localization analysis
| **Cellular location of RNA modulation** | Cytoplasm only | Cytoplasm or nucleus

**Table 1: Comparison of RNAi and Cas13b with respect to known or predicted specificity, delivery, RNA cleavage mechanism, applications or cellular location of RNA modulation.**

There are many potential applications for targeting mammalian RNA transcripts using Cas13b. A summary of mammalian RNA targeting applications is illustrated in Figure 2. Cas13b can meet an unmet need in directing RNA binding proteins to modulate translation, dictate alternative splicing, assess or alter transcript localization, target transcripts for degradation, serve as a binding protein to prevent binding with other RNA transcripts, or add epitranscriptomic marks. Each one of these tools could be promising to develop and used to elucidate the basic biology of different cellular functions or diseases. To create an RNA binding protein with RNA editing modules, structure guided engineering of Cas13b can be completed similar to Cas9 engineering used for transcriptional activation. Specifically, Cas13b can be directly fused to RNA editing proteins or alternatively, MS2 or PP7 hairpins can be added onto permissive areas of the crRNA and MS2 or PP7 coat protein fusions with RNA editing proteins can be utilized. Although various regions of Cas13b and the crRNA can be empirically tested for optimal placement of a fusion protein or stem loop addition, it is generally easiest to use structure-guided placement, which would require the Cas13b crystal structure.
Figure 2: Applications for targeted RNA modulation in mammalian cells. The yellow ovals represent RNA binding proteins, in this case Cas13b. Modular RNA effector domains are illustrated by various linked shapes. Figure from (Mackay et al., 2011).

Selecting functional and specific modular domains to mediate targeted translation, alternative splicing, RNA editing, or RNA epitranscriptomics is critical for development of an RNA editing tool. To direct translation of a specific mRNA, Cas13b with the eukaryotic translation initiation factor eIF4G can be targeted near the start codon within the 5’ UTR. For targeted translation, it is ideal to target an early and rate-limiting step, which for translation is formation of the initiation complex. The initiation of translation requires many different components. The protein eIF4G serves as a scaffold protein, which binds to many other proteins including eIF3 and plays a role in recruiting the 40S ribosomal subunit (Bhat et al., 2015). Critically, the eIF4G subunit
was shown to be sufficient in driving translation of downstream mRNAs in HeLa cells using a direct fusion of the C-terminal end of eIF4G to the protein IRP-1 (De Gregorio et al., 1999). Cas13b fusions with eIF4G could be a tool for targeted upregulation of translation.

Cas13b could also modulate alternative splicing by blocking splice sites or targeting splicing events. For RNA splicing to occur, proteins of the spliceosome interact with critical sequences at the 5’ splice site, A branch point, polypyrimidine tract within the intron, and 3’ splice site (Matlin et al., 2005). To target alternative splicing, Cas13b can be fused to an arginine and serine-rich (RS) protein domain to facilitate binding of the U1 small nuclear ribonucleoprotein to the 5’ splice site. A proof of concept example of this has been illustrated in one study that fused the arginine and serine-rich domain of the protein SF2 to an RNA targeting PUM protein to enhance RNA-splicing at a particular site (Wang et al., 2009). Alternatively, Cas13b can be targeted to a splice site and physically block it from interacting with the spliceosome.

Cas13b could also be used for targeted RNA editing or alteration of the primary RNA sequence when used in conjunction with RNA base editing enzymes such as ADAR or cytidine deaminase (Gerber and Keller, 2001). Unlike DNA’s repair mechanisms NHEJ or HR that fix cut DNA, cut RNA does not undergo repair mechanisms and thus RNA base editing enzymes could be used for targeted single base substitutions. ADAR is an enzyme that mediates A to I transitions on dsRNA, via adenosine deamination where I can be recognized as G base. Alternatively cytidine deaminase could be used, which mediates C to U transitions (Nishikura, 2010). If successful, RNA editing enzymes could be used to better understand RNA editing in a natural context. For example, C to U transitions in the 3’UTR of an mRNA may alter miRNA targeting specificity (Gu et al., 2012) and with a targeted editing tool, miRNA binding on endogenous transcripts could be studied. Alternatively, RNA editors could in theory have therapeutic applications, especially for targeting single base pair mutation diseases such as hemophilia or tyrosinemas.

Lastly, targeted RNA binding proteins could be useful to direct and better understand RNA epitranscriptomics. RNA is known to acquire numerous post-transcriptional small chemical modifications (Frye et al., 2016), for example, the addition of a methyl group on adenosine residues to create N^6- methyladenosine (m^6A), a well-studied mRNA modification. The presence of m^6A in the 5’ UTR of an mRNA has been correlated with cap-independent translation (Meyer et al., 2015) and this modification is also enriched proximal to stop codons and at large internal exons (Dominissini et al., 2012; Meyer et al., 2015). Proteins known to deposit (large complex with METTL3 protein) (Bokar et al., 1997) or remove (FTO and ALKBH5) (Jia et al., 2011; Zheng et al., 2013) m^6A modifications have also been identified. A tool to target RNA modifications could help
link a phenotype or mechanism directly to an RNA modification. Although creating tools for targeted epigenomic modifications has potential, additional basic biology studies are needed to better understand the various function or mechanism of less understood RNA modifications and additional mapping technologies are needed to identify the location of these modifications.

Two major points of concern for any RNA targeting tool that modulates cellular transcripts include specificity and temporary versus long-lasting effects. Specificity for any genome or transcriptome engineering tool is critical. It is important that the tool functions on the targeted transcript with little to no effects on non-targeted transcripts. At this point, the initial targeting specificity of Cas13b is unknown and will require further studies. With Cas13b tools requiring a modular domain, it is also important to ensure that these domains are specific as well. Another concern about RNA targeting tools is the inherent temporary nature of RNA. For example, mRNAs are transcribed and readily degraded within the cell and thus any edit to RNA would only have a temporary effect. To make any lasting or permanent effect, RNA would need to be continually edited. However, editing RNA would not produce any permanent off-target mutations that are possible with DNA targeting enzymes.

Besides modulating RNA within a mammalian cell, Cas13b could be developed for use in a RNA diagnostic assay. Cas13b used as a specific RNA detection tool relies on the collateral effect: an activating target RNA, mixed with quenched-fluorescent RNAs, undergoes targeted RNA cleavage and then the activated enzyme cleaves the fluorescent quenched RNAs in the sample, resulting in fluorescence. This proof of concept experiment was described for C2c2 and illustrated in Figure 3 (East-Seletsky et al., 2016). A sensitive RNA diagnostic assay could be used to detect clinically relevant RNA viruses in patients such as HIV or Zika virus. However, detection would require an increase in accuracy, sensitivity, and speed (Gootenberg et al., 2017), potentially by developing rapid paper-based diagnostics (Pardee et al., 2014).
Cas13b also shows promise for RNA based imaging applications. RNA localization is of interest to study because cellular localization of mRNA can illustrate spatial specific mRNA expression or regulation, especially in cells with well-defined subcellular compartments such as budding yeast, Drosophila embryos, or neurons (Martin and Ephrussi, 2009). Catalytically inactive Cas13b with a tag such as FLAG, HA, or even GFP can bind to its target and be visualized using simple traditional immunostaining methods or potentially live cell imaging. If efficacious, RNA localization could enable simple and rapid mRNA localization studies compared to other techniques such as fluorescence in situ hybridization (FISH) which are time consuming and require conjugation of a visible marker to each unique RNA probe (Long et al., 1995). Additionally, Cas13b localization can work on endogenously expressed mRNAs unlike other studies that transfected artificial mRNAs containing a MS2 hairpin loop and assayed for its localization using a MS2 coat protein GFP fusion (Bertrand et al., 1998). Lastly for a Cas13b localization tool, it is critical to ensure that the targeted protein binding does not alter RNA expression levels, localization, or stability, all of which would need to be validated through orthogonal assays such as qPCR or FISH.

Conclusion

CRISPR-Cas adaptive immune systems in bacteria and archaea naturally exhibit wide diversity in protein composition to target foreign nucleic acids for degradation. Class 2 CRISPR-Cas systems employ a large single effector nuclease to mediate interference and can be harnessed for genome engineering. The targeting capabilities for engineered single effector Cas proteins is broadened by identifying, characterizing, harnessing, and analyzing the diverse pool of naturally
available enzymes and thus there is an interest to continue to identify uncharacterized CRISPR-Cas systems. In a large computational sequence database mining approach, two novel Class 2 CRISPR-Cas systems containing the signature protein Cas13b were discovered. Cas13b is classified as a Type VI CRISPR-Cas system targeting single stranded RNA for cleavage using its two HEPN domains and also exhibits cleavage of its own CRISPR array(s) and non-specific collateral cleavage activity. Cas13b loci are devoid of the acquisition genes cas1 and cas2 but may harbor either csx27 or csx28, whose proteins repress or enhance interference activity, respectively. Cas13b represents the second unique single effector Cas protein with targeted RNA cleavage or binding activity that may be repurposed as a transcriptome modifying tool or to develop RNA diagnostic or imaging assays. Harnessing Cas13b to modify the transcriptome would enable targeted modifications to RNA translation, splicing, or deposition of epitranscriptomic marks, all of which are RNA modifications that are not easily manipulated using current technologies. Before harnessing Cas13b, it is critical to elucidate its basic biology including functional mechanisms, specificity, role of small proteins, or even how it undergoes acquisition. Genome modulating technologies have the promise to revolutionize biological understanding of endogenous nucleic acids or epigenetic markers in any organism through targeted perturbation of the genome. CRISPR-Cas systems provide an ideal platform for genome editing because they are active, specific, easy to target, and modular. Development of these CRISPR-Cas genome or transcriptome engineering tools could help elucidate the function of causal genetic variations and even offer a potential therapeutic avenue to fix rare genetic diseases.
References:


