The history and market impact of CRISPR RNA-guided nuclease

Paul BG van Erp1, Gary Bloomer2, Royce Wilkinson1 and Blake Wiedenheft1

The interface between viruses and their hosts’ are hot spots for biological and biotechnological innovation. Bacteria use restriction endonucleases to destroy invading DNA, and industry has exploited these enzymes for molecular cut-and-paste reactions that are central to many recombinant DNA technologies. Today, another class of nucleases central to adaptive immune systems that protect bacteria and archaea from invading viruses and plasmids are blazing a similar path from basic science to profound biomedical and industrial applications.

Addresses
1 Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA
2 Technology Transfer Office, Montana State University, Bozeman, MT 59717, USA

Corresponding author: Wiedenheft, Blake (bwiedenheft@gmail.com)

In retrospect, we probably should have anticipated that bacteria and archaea would have sophisticated immune systems. After all, viruses are the most abundant biological agents on the planet, causing roughly 1023 infections every second [1–3]. The selective pressures imposed by viral predation have resulted in the evolution of numerous phage defense systems, but it was only recently that sophisticated adaptive defense systems were identified in both bacteria and archaea [4–7].

Initial indications that Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) were part of an adaptive defense system came from a series of bioinformatics observations revealing that the short spacer sequences embedded in CRISPRs were sometimes identical to sequences found in phages and plasmids [8–10]. These observations led to the hypothesis that CRISPRs are central components of an adaptive immune system, and in 2007 Barrangou et al., provided the first demonstration of adaptive immunity in bacteria by monitoring CRISPR loci in phage-challenged cultures of Streptococcus thermophilus [11**]. This paper showed that CRISPRs evolve by acquiring short fragments of phage-derived DNA and strains with new spacers are resistant to these phages. It was immediately clear that this paper would serve as a foundation for an emerging team of scientists interested in understanding the mechanisms of adaptive defense systems in bacteria and archaea, but few of us anticipated the broader impacts of these discoveries for new applications in genome engineering.

Building on this initial foundation, a series of mechanistic studies showed that CRISPR loci are transcribed and processed into a library of small CRISPR derived RNAs (crRNAs) that guide dedicated nucleases to complementary nucleic acid targets [5–7,12,13]. In nature, these RNA-guided nucleases provide bacteria and archaea with sequence specific resistance to previously encountered genetic parasites. However, sequence specific nucleases have considerable value in biotechnology and one of these CRISPR-associated nucleases (i.e. Cas9) has recently been co-opted for new applications in biomedical, bioenergy, and agricultural sciences [14–17].

A goldmine for biotechnology

The molecular interface between a parasite and its host is a hot spot for innovation. A resistant host has a competitive advantage over a susceptible host, but an obligatory parasite faces extinction unless it is able to subvert host defense mechanisms. This conflict results in an accelerated rate of evolution that stimulates genetic innovation on both sides of this molecular arms race.

Genes at the interface of a genetic conflict have proven to be a goldmine for enzymes with activities that can be creatively repurposed for applications in biotechnology. In the 1970s, basic research on bacteriophages led to the discovery of DNA restriction endonucleases, which transformed molecular biology by making it possible to cleave specific DNA sequences [18]. The discovery of these enzymes paved the way for the emergence of recombinant DNA technologies, and in 1978 Werner Arber,
Daniel Nathans, and Hamilton Smith shared the Nobel Prize in Physiology or Medicine ‘for the discovery of restriction enzymes and their application to problems of molecular genetics’ [19]. Identification and application of type II restriction enzymes, which are integral to almost every aspect of DNA manipulation, effectively triggered the emergence of a global biotech industry.

Like restriction enzymes, CRISPR systems evolved as components of a prokaryotic defense system. However the mechanisms of sequence recognition by these enzymes are fundamentally different. Unlike DNA restriction enzymes, which typically rely on protein mediated recognition of 4–8 bp; CRISPR-associated nucleases are guided by base pairing between an RNA-guide and a complementary target. The implications of this targeting mechanism have triggered a sea-change in biology and now the historical precedent of nucleases in biotechnology seems poised to repeat itself.

Why all the fuss?
Reverse genetics is a powerful method used to determine the biological function of a specific gene. This approach is used routinely to determine gene function in organisms with simple genomes, but existing techniques are not applicable for high-throughput genetic screens in organisms with large genomes and multiple chromosomes. However, the recently discovered mechanism of DNA cleavage by the CRISPR RNA-guided nuclease Cas9 [20**], has transformed the field of genetics by allowing efficient and precise genetic manipulation of diverse eukaryotic genomes, including human cells [14–17]. To repurpose the Cas9 nuclease for targeted genome editing, the cas9 gene has been codon optimized for expression in eukaryotic systems and tagged with a nuclear localization signal (NLS) [21**,22**,23**]. Cas9 and the guide RNA have been delivered to eukaryotic cells by transient transfections with expression vectors [24] or purified Cas9/sgRNA [25,26], viral transduction using Lentiviruses [27,28*,29*] or Adeno-Associated Virus [30–33], and cytoplasmic or nuclear injections [34–37](Figure 1). In each case, RNA-guided targeting of Cas9 to a complementary DNA target results in a double-stranded DNA break (DSB) at the target site. These lesions are typically repaired by non-homologous end joining (NHEJ), which is an error-prone process that is often accompanied by insertion or deletion of nucleotides (indels) at the targeted site, resulting in a genetic knock-out of the targeted gene due to a frameshift mutation. Alternatively, DSBs are repaired by homology directed repair (HDR). In most systems, HDR is an inefficient process that requires a DNA donor with sequence homology to regions of the genome that flank the DSB [38]. Using Cas9 in combination with a DNA donor provides a method to target cleavage and repair of naturally occurring genetic defects, add foreign DNA encoding genes with new functions to specific locations, or precisely excise defined fragments of DNA [14–16,39].

Cas9 is not the first programmable nuclease developed for engineering eukaryotic genomes [40]. Some of the earliest methods for introducing targeted genome modifications relied on meganucleases (e.g. HO and I-SceI), which are endonucleases that have long recognition sequences (12–40 bp) [41]. The enhanced specificity of these nucleases, as compared to the 4–8 bp recognized by most restriction enzymes, presents an opportunity to target specific locations in large eukaryotic genomes. However, these enzymes rely on protein-mediated recognition of the target DNA, and reprogramming these proteins to target new DNA sequences has been challenging due to the integrated nature of the DNA binding and nuclease domains of these proteins. To address this issue, non-natural chimeric nucleases composed of distinct DNA binding and nuclease domains have been engineered. The most celebrated examples of these are Zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs).

Zinc fingers (ZFs) are sequence specific DNA binding domains found in eukaryotic transcription factors and transcription activator-like effectors (TALENs) are modular DNA binding proteins made by bacterial pathogens that infect plants. The mechanism of DNA recognition by these proteins is well understood and the modular nature of these interactions has been exploited for reprogramming. While the success of these proteins for targeted genome engineering cannot be overstated, they both rely on protein-mediated recognition of the DNA, which means that every new target requires the engineering of new proteins. In contrast to these techniques, Cas9 is an RNA-guided nuclease that relies on complementarity base pairing and protein mediated recognition of an adjacent short sequence motif, commonly referred to as the PAM (protospacer adjacent motif) [42]. PAMs are typically 2–5 bp in length, which means that PAMs occur at high frequency and are rarely a limitation when designing RNA guides to specific target sequences. However, Cas9 proteins from different organisms often recognize different PAM sequences, so in rare instances where a target sequence does not contain a PAM recognized by one Cas9 (i.e. Cas9 from Streptococcus pyogenes recognizes a 5’-NGG PAM), then another Cas9 with a distinct PAM recognition sequence may be used (i.e. Cas9 from Neisseria meningitides recognizes a 5’-NNNNGATT PAM) [43]. The diversity of Cas9 proteins and the simplicity of RNA-guided programming abrogates the need for sophisticated protein engineering, and affords rapid generation of designer nucleases. In fact, guide RNAs that target Cas9 to as many as 20 000 different genes and 1800 microRNAs in the human genome have been generated in a single experiment [28*,29*,44]. These whole-genome knockout techniques are transforming functional genomics and redefining the possibilities of reverse genetics.
Cas9 delivery and repair of the targeted DNA. Cas9 and a single-guide RNA (sgRNA) have been delivered to eukaryotic cells by several methods: transient transfections (expression vectors or purified Cas9/sgRNA complex), cytoplasmic or nuclear injections (expression vectors, mRNA or purified Cas9/sgRNA) and by transduction (Lentiviruses or Adeno-Associated Virus). Cas9 identifies its target by protein mediated PAM (yellow) recognition and base pairing between the sgRNA and the DNA target. Target recognition activates the nuclease sites (red triangles), resulting in double stranded breaks (DSBs) 3-4 nucleotides downstream from the PAM. DSBs can be repaired by non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ results in insertion or deletions (indels), which often results in a frameshift mutation. HDR relies on homologous recombination with a donor DNA molecule. This donor DNA can be used to specifically insert desired sequences.

The emerging market
CRISPR applications span almost every industry that involves biological systems (Table 1). Danisco (DuPont) was an initial pioneer of commercial use of CRISPR technology to enhance viral immunity in bacteria used to make yogurts and cheese, but other markets have been rapidly emerging. Applications in agriculture have lower regulatory hurdles than biomedical application and some of these markets are anticipated to produce earlier returns on investments. Dow Agrosciences has co-developed intellectual property (IP) with Sangamo Biosciences for developing genetically modified crops using Cas9, and Cellectis Plant Sciences is leveraging its relationship with parent Cellectis SA to move the technology into crops. Similarly, Recombinetics Inc. is using TALENs, ZFNs and Cas9 to enhance productivity in the livestock industry [45]. While the USDA has yet to decide on how it will treat genomes edited using Cas9, it has already ruled that ZFNs and TALENs do not fall under their governance [46]. This saves an average 5.5 years and $35 million in related regulatory costs for bringing a product to market [47]. Similar treatment of CRISPR-based genome editing may stimulate economic activity around the development of new agricultural and industrial products.

In addition to Cas9-based applications in the agricultural industry, market segments for Cas9 endonucleases within the human health sectors are experiencing frenetic growth. These markets include: gene-therapy, cell-therapy, and immuno-therapy, fast and efficient development...
of transgenic research animals, drug discovery, as well as target validation and screening. It is difficult to accurately estimate the value of the nascent market for CRISPR RNA-guided nucleases in the biomed sector, but documents from the initial public offering (IPO) of Horizon Discovery Group, plc, which has in-licensed Cas9 IP, indicate a market size of $46 billion [48] and recent private equity financings of Cas9-based genome engineering companies include: Caribou Biosciences (undisclosed venture estimated at $2.9 million from Novartis), CRISPR Therapeutics ($25 million), Recombinetics, Inc. ($5 million), Intellia Therapeutics ($15 million), and Editas Medicine ($43 million). In total, companies with an interest in using Cas9 for applications related to gene therapy have raised over $600 million in venture capital and public markets since the beginning of 2013. The pace of this activity is remarkable given that the first granted patent for the use of CRISPR technology in eukaryotic cells was issued April 15, 2014.

Commercial interest in Cas9 IP has not escaped the interest of big pharmaceuticals. Novartis has partnered with a tier I private equity firm, Atlas Ventures to commit $15 million to kick start Intellia Therapeutics, and Pfizer partner Callectis SA will be using Cas9-based technologies to make T-cells with chimeric antigen receptors. In January of this year, AstraZeneca announced four partnerships with academia around the use of Cas9 nucleases to validate new drug targets.

In addition to the agricultural and biomedical sectors, the research tools market is also embracing CRISPR-based technology. Sigma-Aldrich has in-licensed technology in order to make, use, and distribute tools for the generation of modified plant and animal models, custom cell line creation and for pooled genetic screening. Perhaps the financial activity in each of these market sectors is heightened by over-exuberance common to the early market development. However, given the scope of current applications across multiple industries, we see no limits to research or financial commitments in this space.

Acknowledgements
We are grateful to Jennifer A. Doudna for valuable feedback on this manuscript. Research in the Wiedenheft lab is supported by the National Institutes of Health (F22GM103500 and R01GM108888), the National Science Foundation EPS-13 (EPS-110134), the MJ. Murdoch Charitable Trust, and the Montana State University Agricultural Experimental Station.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:
- of special interest
- of outstanding interest


This paper provides the first direct evidence for adaptive immunity in bacteria.


This paper provides mechanistic insights that explain how Cas9 can be used as a programmable nuclease for genome engineering.


This is one of the first demonstrations that Cas9 can be used for precise multiplexed genome engineering in mouse and human cells.


This is one of the first demonstrations that Cas9 can be used for precise genome engineering in human cells.


This is one of the first demonstrations that Cas9 can be used for precise genome engineering in human cells, including induced pluripotent stem cells.

human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci USA 2013.


