CRISPR 101

Your Guide to Understanding CRISPR
# TABLE OF CONTENTS

- **Introduction** Page 3
- **A Brief History of CRISPR** Page 3
- **Gene Editing Before CRISPR** Page 4
- **What is CRISPR/Cas9?** Page 5
- **CRISPR Applications** Page 7
- **CRISPR Guides** Page 9
- **Developments in Synthetic sgRNA** Page 12
- **Alternatives to *S. pyogenes* Cas9** Page 13
- **Gene Editing is Just the Beginning** Page 14
INTRODUCTION

CRISPR is igniting a revolution. A relatively recent discovery in the timeline of biotechnology, CRISPR is quickly becoming a standard and flexible laboratory tool, and it is well on its way to permeating a large variety of applications. Researchers are deploying CRISPR across a wide range of life science disciplines, from agriculture and medicine to biofuels and industrial fermentation. Read on for a crash course in everything you need to know if you're just getting your first taste of CRISPR.

A BRIEF HISTORY OF CRISPR

The foundational discoveries that led to CRISPR/Cas9 technology can be traced back to 1993, when the genomic regions known as CRISPR loci were first identified. In 2007, after years of studying CRISPR genetic motifs, researchers came to the conclusion that CRISPR’s function is related to microbial cellular immunity.

Throughout the next 5 years, several research groups worked to elucidate the underlying molecular mechanisms behind CRISPR in Prokaryotes. CRISPR works as a form of Prokaryotic immunity that identifies, targets, and eliminates bacteriophage and foreign DNA. By 2012, researchers realized that CRISPR could be adapted for engineering the genomes of microbes, plants, animal, and other varieties of cells.

Today, CRISPR is utilized for countless applications and its adoption continues to increase exponentially in laboratories throughout the world. Due to its adaptability across a wide range of species and its simplicity of use, CRISPR/Cas9 has quickly revolutionized genome engineering.
While CRISPR/Cas9 is revolutionary due to its speed and adaptability, it is not the first technology to enable genome engineering. That distinction belongs to a technology known as zinc-finger nucleases (ZFNs).

The ZFN method of genome editing is accomplished by engineering an enzyme with both a zinc finger DNA-binding motif and a restriction endonuclease domain. The zinc finger domain is designed to target and bind specific sequences of DNA, whereafter the nuclease domain cleaves the DNA at the desired sequence. Although ZFNs represented the first breakthrough in site-specific genome engineering, they proved to have several limitations. In addition to exhibiting off-target effects, ZFNs are expensive and time consuming to engineer, and only allow one genomic edit to be accomplished at a time.

Many years after ZFNs made their debut, a similar method known as transcription activator-like effector nucleases (TALENs) was developed. Like ZFNs, the TALENs method utilizes engineered enzymes containing both a DNA-binding domain and a separate DNA-cleaving domain. TALENs have an advantage over ZFNs because they are more flexible: their DNA-binding motifs can target a wider range of sequences. Although they are easier to design than ZFNs, TALENs are still expensive to produce and can only be used to create one genomic edit at a time.

An additional genome editing technique is derived from the use of engineered restriction enzymes in concert with recombinant adeno-associated viruses (rAAVs). AAV is a non-pathogenic virus that infects mammalian cells at all stages of the cell cycle and integrates into the host genome at predictable sites. The AAV genome can be modified to target specific sequences in the host genome and integrate desired modifications upon infection.

Up until now, the field of genome engineering has provided researchers with few options and many limitations. Not surprisingly, the ease of use and versatility of CRISPR/Cas9 technology has led to its rapid adoption for genome engineering.
WHAT IS CRISPR/Cas9?

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Type II system is a form of prokaryotic immunity that has been adapted for genome engineering. It consists of two components: a specific guide RNA (gRNA) and a non-specific CRISPR-associated endonuclease protein from *Streptococcus pyogenes* (Cas9). In nature, Prokaryotes store small palindromic segments of DNA that are interspaced with other fragments of genetic material. These segments fall between CRISPR loci and correspond to fragments of viral DNA that the cell has previously encountered. After a prokaryotic cell successfully clears a viral infection or encounters a foreign plasmid, it stores fragments of foreign DNA as a way to retain a genetic memory in order to recognize and disable future infections.

Think of CRISPR/Cas9 as a pair of molecular scissors guided by a GPS. The disabling of the invading genetic material is carried out by the Cas9 protein which is a nonspecific but programmable endonuclease that is directed to a specific sequence target by a guide RNA (gRNA). Once located, Cas9 causes a double-stranded break (DSB) in its target loci. The guide RNA is complementary to a segment of the foreign DNA or viral genome; this allows Cas9 to identify and cut DNA with a high degree of specificity.

One of the critical discoveries made about CRISPR/Cas9 was the identification of two distinct segments of RNA that are required for function: CRISPR RNA (crRNA) and transactivating CRISPR RNA (tracrRNA). The crRNA is complementary to the target DNA sequence and will bind to the sequence to be cleaved. The tracrRNA is a small RNA molecule that enables maturation of the crRNA. The crRNA and tracrRNA segments may exist in nature as a duplex or, synthetically, as part of one seamless fusion sequence known as single-guide RNA (sgRNA).
WHAT IS CRISPR/Cas9? (continued)

Under the direction of its corresponding gRNA, the Cas9 enzyme binds DNA at a specific genetic element (known as the protospacer adjacent motif, or PAM) given by one of several trimers with the sequence 5’-NGG-3’. After binding, the Cas9 creates a blunt, double-stranded break in the foreign DNA, rendering it harmless to the cell. As a biotechnological tool, CRISPR/Cas9 operates in a similar way to how it does in nature. However, by providing the Cas9 protein with a gRNA, the nuclease can be programmed to cleave any host organism’s genome at virtually any location, per the design specifications of the gRNA.

After the cut, CRISPR/Cas9 technology relies on the cell’s natural repair mechanisms to attend to the double-stranded break in one of two ways. First, the cell may proceed with non-homologous end joining (NHEJ) of the cleaved fragments. NHEJ binds the double stranded break back together, but in the process may insert or delete a nucleotide and produce a frameshift mutation (Indel). This method typically leads to a frameshift mutation and a knockout of the targeted genetic element’s function. Alternatively, if the object of the experiment is to replace the targeted genetic element with a different sequence (e.g., for gene deletion, single-base editing, etc.), the cell can be directed towards an alternative repair pathway, homology directed repair (HDR). To accomplish this, a cell must be provided with a homologous DNA template containing the desired change in sequence. A certain number of cells will use this template to repair the broken sequence via homologous recombination, thereby incorporating the desired edits into the genome.

As we have seen, the magic of CRISPR is in its ability to force a DSB event. Cells must repair DSBs, or risk dying. Thus, all of the editing that comes from CRISPR is due to the cell’s innate ability to repair itself. Without CRISPR, it is very difficult to cause DSBs in cells or activate these repair pathways.
CRISPR APPLICATIONS

GENOMIC EDITING

CRISPR’s ability to edit genes relies on the fact that the cell’s non-homologous end joining (NHEJ) mechanism for repairing double-stranded breaks is imperfect; the repair enzymes often incorporate insertions or deletions into the site of the cut created by Cas9. Such mutations disrupt the production of the gene’s corresponding protein, allowing researchers to study the impact of the knockout on cellular structure and function. As scientists at Caribou Biosciences recently discovered, the size and nature of the errors made during NHEJ are in fact not random but depend on the target sequence (1).

A CRISPR genomic editing experiment can be orchestrated in several different ways. The essential CRISPR machinery (the gRNA and Cas9 protein) can be delivered in the form of a plasmid, in the form of RNA fragments, or in the form of ribonucleotide particles (RNP). This last RNP-based method—also known as the DNA-free method—has been found to enable the highest editing efficiencies (2). Liang et al. recently achieved editing rates of up to 94% in Jurkat T cells and 87% in induced pluripotent stem cells through the use of CRISPR RNPs, and also found that this approach significantly reduces off-target effects, due to the transient nature of RNPs in the cell (2). In addition, while other non-DNA free CRISPR delivery methods carry a risk of exogenous DNA elements integrating into the genome of the host cell, DNA-free genome editing carries no such risk. For this reason, the method is gaining popularity for human therapeutic methods of editing primary cells and stem cells, as well as in crop engineering.

GENE SILENCING

While CRISPR-induced genetic knockouts/knockdowns can be achieved by editing and disrupting a gene’s constituent sequence, genes can also be silenced using a modified version of CRISPR that suppresses expression without altering the target sequence. By mutating the endonuclease domain of the Cas9 protein, researchers have created a system wherein the CRISPR complex binds to its DNA target but does not cleave it. The binding of the mutated Cas9 interferes with gene expression by preventing the cell’s transcription machinery from accessing the gene, thereby silencing its expression. In a nod to the precursor gene silencing technique known as RNA interference, or RNAi, the technique has been termed CRISPR interference, or CRISPRi (3). Compared to other gene silencing techniques such as RNAi, CRISPRi is associated with higher efficiency, lower off-target effects, greater ease of design, and greater flexibility.
CRISPR APPLICATIONS (continued)

HOMOLOGY DIRECTED REPAIR (HDR)

This method of incorporating specific changes into a cell's DNA sequence is sometimes referred to as a knock-in. CRISPR knock-in applications are achieved by inducing target cells to carry out homology-directed repair (HDR) of the double-stranded break generated by Cas9. To induce HDR, cells must be provided with a strand of DNA reflecting the desired sequence edits, which the cell then uses to repair the severed target sequence via homologous recombination. HDR enables countless genomic re-writing applications, from introducing single point mutations to inserting entire selection markers. While the HDR technique yet requires refining, researchers have already employed the method to correct a genetically-encoded mutation causing cataracts in mice (4), achieving proof of concept for HDR as a method of addressing genetically based diseases.

ACTIVATION OR REPRESSION OF TARGET GENES

While “broken” Cas9 endonucleases can be used to silence gene expression, researchers have begun to further modify the CRISPR apparatus to enable fine-tuned control over the activation or repression of target genes. By complexing additional activator proteins to an inactive Cas9, scientists are now developing systems that toggle the expression of a gene on and off at their signal. Polstein and Gersbach created one such system by fusing light-inducible proteins to muted Cas9, causing gene expression to be activated in the presence of blue light and repressed in its absence (5). Other teams of researchers, such as Zalatan et al., are building more complex systems for multiplexed gene activation and repression at as many as three loci simultaneously (6).

CRISPR SCREENS

Another up-and-coming application of CRISPR technology is in genome-wide functional screening. Until recently, RNAi was the leading approach to perform such screens whereby genes are systematically inhibited across the genome in order to determine their associated function and phenotype. However, as mentioned previously, RNAi is plagued with problems related to low efficiency and high rates of off-target effects. With the advent of CRISPR, genomic screening libraries are now being developed and applied to knockout thousands of genes in a single screen with a high level of effectiveness. Schmidt and colleagues recently interrogated this method by developing a CRISPR screening library based on HEK293T cells that targets nearly every known protein-coding gene in the human genome and that they are making freely available to the research community (7).
CRISPR GUIDES

PLASMID

One route that essential CRISPR components can be delivered to cells is by encoding them into plasmids. This approach requires a researcher to design and clone a plasmid encoding both the Cas9 protein and desired gRNA. Commercial plug-and-play plasmids are now available for this purpose, allowing researchers to insert a gRNA template sequence of their design specifications. First, of course, the DNA template sequence must be designed. This design must consist of the gRNA core sequence, the PAM, and an 17-20bp region in the gene of interest that is both adjacent to a Cas9 PAM and genomically unique in order to minimize off-target effects. Ideally, the 17-20bp target site should correspond to an area near the N-terminus of the protein product to increase the odds of a complete knockout and, if applicable, be part of an exon. A wide selection of software tools and databases are now available to assist in the design of gRNA and the DNA template from which it must be transcribed.

Once the gRNA template sequence has been designed, a corresponding oligo can be ordered and cloned into a plasmid. Engineering a CRISPR plasmid involves all of the usual steps that accompany a typical cloning assay. Once the researcher has obtained both the plasmid and the gRNA template sequence that will be inserted, the insert must be amplified, digested, and ligated to the plasmid before being transformed into cells. After screening for the recombinant plasmid and verifying its sequence, the editing experiment can proceed by delivering the plasmid to the target cell. Overall, the process of preparing a custom CRISPR plasmid in-house consumes 1-2 weeks of time before the editing assay can be undertaken. In addition to the required time investment, the plasmid approach is also limited by the fact that it typically results in higher off-target editing efficiencies compared to other methods, due to the continual presence of Cas9 within the cell. Plasmids also run the risk of integrating into the genome of the host cell, a possibility that is particularly problematic for CRISPR applications related to human medicine and crop plant engineering.
CRISPR GUIDES (continued)

**in vitro Transcription (IVT)**

As in the plasmid approach, the *in vitro* transcription (IVT) method requires that a gRNA DNA template be designed based on the target sequence within the gene of interest. In this case, however, the template must also include an upstream promoter site (typically T7) that is appropriate for the RNA polymerase to use during IVT. There are a number of different ways of synthesizing the sgRNA DNA template; usually these involve PCR amplification using primers designed to incorporate the promoter, the target site, or both. Once the T7-promoted sgRNA DNA template has been obtained, it is then transcribed into a sgRNA product using one of multiple IVT kits available for this purpose. After purification, the sgRNA can be complexed with Cas9 protein and delivered to the target cell as a ribonucleoprotein particle (RNP) or co-transfected into the cell alongside Cas9 mRNA. Unlike the plasmid-based approach, the IVT approach requires only 1-3 days to prepare the gRNA for a CRISPR assay.

Nonetheless, the approach remains labor-intensive and is not scalable for multiple CRISPR targets. In addition, gRNAs that are IVT-derived tend to exhibit highly variable editing efficiencies, due to their impurities; furthermore there is an increase in the possibility of off-target effects, which could result from errors made by the RNA polymerase during the IVT process.
CRISPR GUIDES (continued)

SYNTHETIC GUIDE RNA

Synthetic guide RNA (gRNA) can come in separate crRNA and tracrRNA fragments that must be annealed together, or as seamless single guide RNA (sgRNA) molecules. Of these two forms, sgRNA generally produces a higher editing efficiency due to inefficient annealing of the two pieces and the tendency of tracrRNA fragments to form tetramers that interfere with the Cas9 protein. In addition, synthetic gRNA may be co-transfected with Cas9 mRNA or complexed with Cas9 in the form of an RNP that is delivered to target cells by electroporation, chemical transfection, or microinjection. Of these two delivery formulations, supplying the cells with pre-complexed RNPs generally results in significantly higher editing efficiencies. In fact, by choosing the most optimal combination of the different possibilities—that is, deploying sgRNA in the form of an RNP—researchers can achieve editing efficiencies as high as 90% in their target cell or organism. In addition to the superior efficiency offered by synthetic gRNA, this approach also offers researchers the advantage of being able to incorporate site-specific chemical modifications into gRNA nucleotides. These can provide additional protection against exonucleases within the cell and also help to protect against intracellular immune responses. These can be crucial for the success of experimental applications that require such chemical modifications in order to edit the target DNA, for example, editing of stem cells.

<table>
<thead>
<tr>
<th>Process</th>
<th>Synthetic Guide RNA</th>
<th>Plasmid</th>
<th>IVT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Choose target sequence</td>
<td>1. Choose target sequence</td>
<td>1. Choose target sequence</td>
</tr>
<tr>
<td></td>
<td>2. Order synthetic RNA</td>
<td>2. Design/order DNA primers</td>
<td>2. Design/order DNA primers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. PCR insert</td>
<td>3. Assemble guide by PCR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Ligate into plasmid</td>
<td>4. Perform IVT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Transform into cells</td>
<td>5. Purify guide RNA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Screen cells</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7. Sequence verify plasmid</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8. Purify plasmid DNA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time to Transfection</th>
<th>Ready for transfection</th>
<th>7-14 days</th>
<th>1-3 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfection Labor Time</td>
<td>Minimal</td>
<td>Days of lab work</td>
<td>Full day of lab work</td>
</tr>
<tr>
<td>Off-target Effects</td>
<td>Lowest</td>
<td>Variable</td>
<td>Variable</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Up to 90% efficiency</td>
<td>Variable</td>
<td>Variable</td>
</tr>
<tr>
<td>Consistency</td>
<td>Highest</td>
<td>Variable</td>
<td>Variable</td>
</tr>
</tbody>
</table>
Within the CRISPR field it has become clear that the use of synthetic sgRNA is the most effective option—sometimes the only effective option—for carrying out CRISPR gene editing assays and other applications. In addition, it is also the most efficient method of genome editing, since variations between batches of synthetic sgRNA are very low, especially compared to IVT-derived guides. Not only are plasmid-based and IVT methods of obtaining gRNA time consuming and not scalable, they can also lead to problems with RNA purity, sequence fidelity, off-target effects and DNA integration into the host genome. However, despite these limitations and the clear advantages that synthetic sgRNA offers over them, the adoption of synthetic sgRNA has lagged. Until very recently, the cost associated with producing synthetic sgRNA—even short, oligomeric strands—was high enough to make synthetic sgRNA a prohibitive expense for many laboratories.

However, recent developments in synthetic sgRNA synthesis has made it so that effective CRISPR technology can be accessed affordably and deployed successfully by more labs around the world. One such advancement lies in Synthego’s automated and scalable production methods for full length 100-mer synthetic sgRNA, which allows these molecules to be generated much more rapidly and at a much lower cost than traditional synthesis methods. The high-throughput synthesis technique also results in higher fidelity sgRNA strands compared to IVT-derived guides, enabling more efficient and consistently reproducible targeting of the desired genomic site. In addition, the synthesis platform permits chemical modification of sgRNA ribonucleotides, an option that is a necessity for certain research applications such as stem cells and other difficult to transform cell lines, such as K562 and A549. Improvements such as these are the first steps toward a future where the cost of CRISPR is within everyone’s means and the efficiency and versatility of the technique is maximized.
ALTERNATIVES TO *S. pyogenes* Cas9

The specific Cas9 nuclease that is used in the great majority of CRISPR genome engineering assays is originally sourced from the bacteria *Streptococcus pyogenes*, and is more descriptively termed “SpCas9.” Although SpCas9 is the most widely-used endonuclease, other, similar nucleases are known to exist and many more are thought to be yet undiscovered.

**Cpf1**

One such alternative, known as Cpf1, is the most commonly used nuclease after SpCas9. Cpf1 is an acronym for “CRISPR from Prevotella and Francisella,” indicating the microbial species from which it originates. Although both are Type II CRISPR systems, Cfp1 and SpCas9 function differently in a few key ways. First, Cfp1 recognizes and binds a different PAM than SpCas9: while SpCas9 recognizes the motif 5'-NGG-3', Cpf1 recognizes 5'-TTN-3', prior to the guide RNA, and therefore the latter can be a better nuclease of choice for experiments targeting DNA regions with high AT-content. Cfp1 also creates a staggered double-stranded cut in its target, rather than the blunt-end cut that is generated by its SpCas9 counterpart. This may make Cfp1 a better selection when HDR is the preferred repair outcome. Finally, Cfp1 is a smaller protein than SpCas9 and does not require a tracrRNA. Thus, the gRNA required by Cfp1 is shorter in length and cheaper to generate than the gRNA required by SpCas9.

**S. aureus** Cas9 (saCas9)

Another nuclease that is being used as an alternative to SpCas9 is known as SaCas9, an ortholog from the Cas9 family that originates from the species *Staphylococcus aureus*. SaCas9 recognizes the same PAM as SpCas9. However, like Cfp1, the SaCas9 protein is much smaller than SpCas9 and the sequences that encode them differ by length of about 1kb. Because of this, it is possible to package SaCas9 in an AAV vector for cellular delivery, whereas the DNA sequence of SpCas9 is too large for this approach. As additional nucleases are discovered, we can expect that further variations on CRISPR will become available to researchers, making the technique even more flexible and more powerful than it is today.
GENE EDITING IS JUST THE BEGINNING

The attention that CRISPR has received owes primarily to its ability to edit the genes of living organisms—especially humans. However, while this side of CRISPR occupies the spotlight, researchers have begun tinkering with the technology to unlock its vast potentials that go beyond the applications discussed so far.

Teams of scientists are now using a modified version of CRISPR to explore epigenomics—the genome-wide set of chemical groups that adorn DNA and its associated histone packaging proteins. Previously, researchers have merely been able to catalogue the correlation between epigenetic markers and gene expression in cells. Now, however, scientists have created a CRISPR complex that is capable of acetylating histone proteins at precise locations dictated by the complex’s gRNA (8). With this tool (and others that are under development), researchers will finally be able to study the causal relationship between epigenetic markers and gene expression.

In addition, CRISPR is being explored as a method of fluorescently labeling DNA in live cells. In a method known as tiling, a Cas9-GFP fusion protein is directed and bound to a target sequence in order to label and image a specific genomic region (9).

CRISPR is also enabling the elucidation of the large portions of the human genome—the vast majority, in fact—for which a function is not known. Scientists have long been trying to identify the location and function of non-gene genetic elements that do not code for proteins but are thought to have important regulatory roles in expression. CRISPR is allowing researchers to knock out these previously uncharted regions to study their role in the cell (10).

CRISPR is not only paving the way for us to solve the most difficult of problems in the life sciences, but also enabling the scientific community to explore dimensions of the genome that we’ve been unable to study up until this point. Due to its adaptability across a wide range of species and its simplicity of use, CRISPR/Cas9 has quickly revolutionized genome engineering. The CRISPR/Cas9 technology promises to deliver some truly stunning advances within the coming decades, particularly in relation to human therapeutics, agricultural biology and basic scientific research.
ABOUT SYNTHEGO

Founded by former SpaceX engineers, Synthego is a leading provider of genome engineering solutions. The company’s flagship product, CRISPRevolution, is a portfolio of synthetic RNA designed for CRISPR genome editing and research.

Synthego’s vision is to bring precision and automation to genome engineering, enabling rapid and cost-effective research with consistent results for every scientist.

Headquartered in Silicon Valley, California, Synthego customers include leading institutions around the world.
REFERENCES


