

Activity 1

Introduction to CRISPR-Cas9 Gene Editing Technology

What is CRISPR-Cas9 gene editing?

In the decades since the discovery of restriction enzymes, researchers have discovered many new molecular tools and techniques that have greatly expanded our genetic engineering capabilities. One of the most exciting recent developments is the CRISPR-Cas9 system (CRISPR). CRISPR derives its name from the system found in nature that allows microbes to defend themselves against viral attack. “Clustered regularly interspaced palindromic repeats” (CRISPR) are sequences in the genomes of some prokaryotes that act as a genomic record of previous viral attack. Along with CRISPR-associated (Cas) proteins, bacteria use the sequences to recognize and disarm future invading viruses. Scientists have adapted this system for genetic engineering purposes.

CRISPR-Cas9 is not the first programmable gene-editing tool, nor is it necessarily the most precise. Other gene-editing tools, like TALENs or zinc-finger nucleases, are also programmable and precise, but they are very expensive and laborious to use. What makes CRISPR-Cas9 so powerful is the combination of its precision and simplicity.

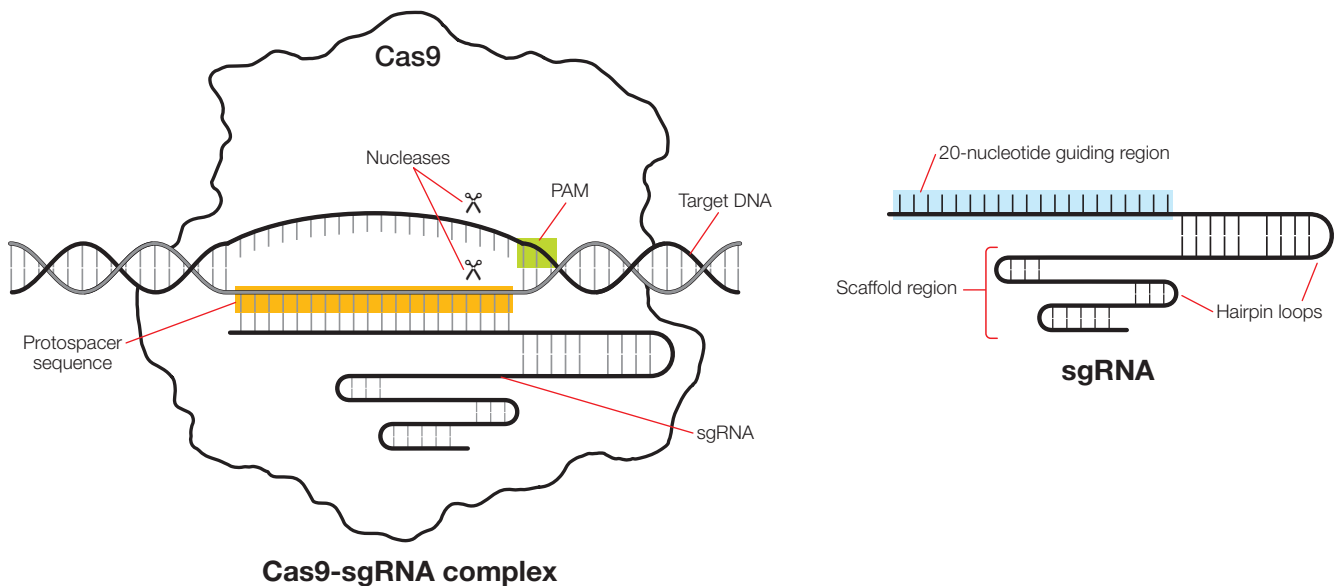


Fig. 1. Anatomy of Cas9 and sgRNA.

The CRISPR-Cas9 system consists of the following components as shown in Figure 1:

- **Cas9 enzyme (Cas9)** — a bacterial endonuclease that forms a double-strand break (cuts) DNA at a specific site within a larger recognition sequence, or target site. The Cas9 recognition sequence includes a 20-nucleotide sequence called the protospacer that is determined by a guide RNA bound to the enzyme
- **Single guide RNA (sgRNA)** — an engineered form of guide RNA that forms a complex with Cas9. The sgRNA is an approximately 100 nucleotide-long fusion of two regions that occur as separate RNAs in nature:
 - Guiding region — part of the CRISPR RNA or crRNA in nature, a typically 20-nucleotide region that is complementary to the target DNA sequence and that defines where Cas9 cuts. Scientists can easily customize this sequence for their own targets
 - Scaffold region — called the transactivating CRISPR RNA or tracrRNA in nature, a region that forms a multi-hairpin loop structure (scaffold) that binds tightly in a crevice of the Cas9 protein. The sequence of this region is typically the same for all sgRNAs
- **Protospacer adjacent motif (PAM)** — a sequence motif immediately downstream of the protospacer sequence in the Cas9 recognition sequence that is required for Cas9 function. Cas9 recognizes the PAM sequence 5'-NGG where N can be any nucleotide (A, T, C, or G). When Cas9 binds the PAM, it separates the DNA strands of the adjacent sequence to allow binding of the sgRNA. If the sgRNA is complementary to that sequence, Cas9 cuts the DNA

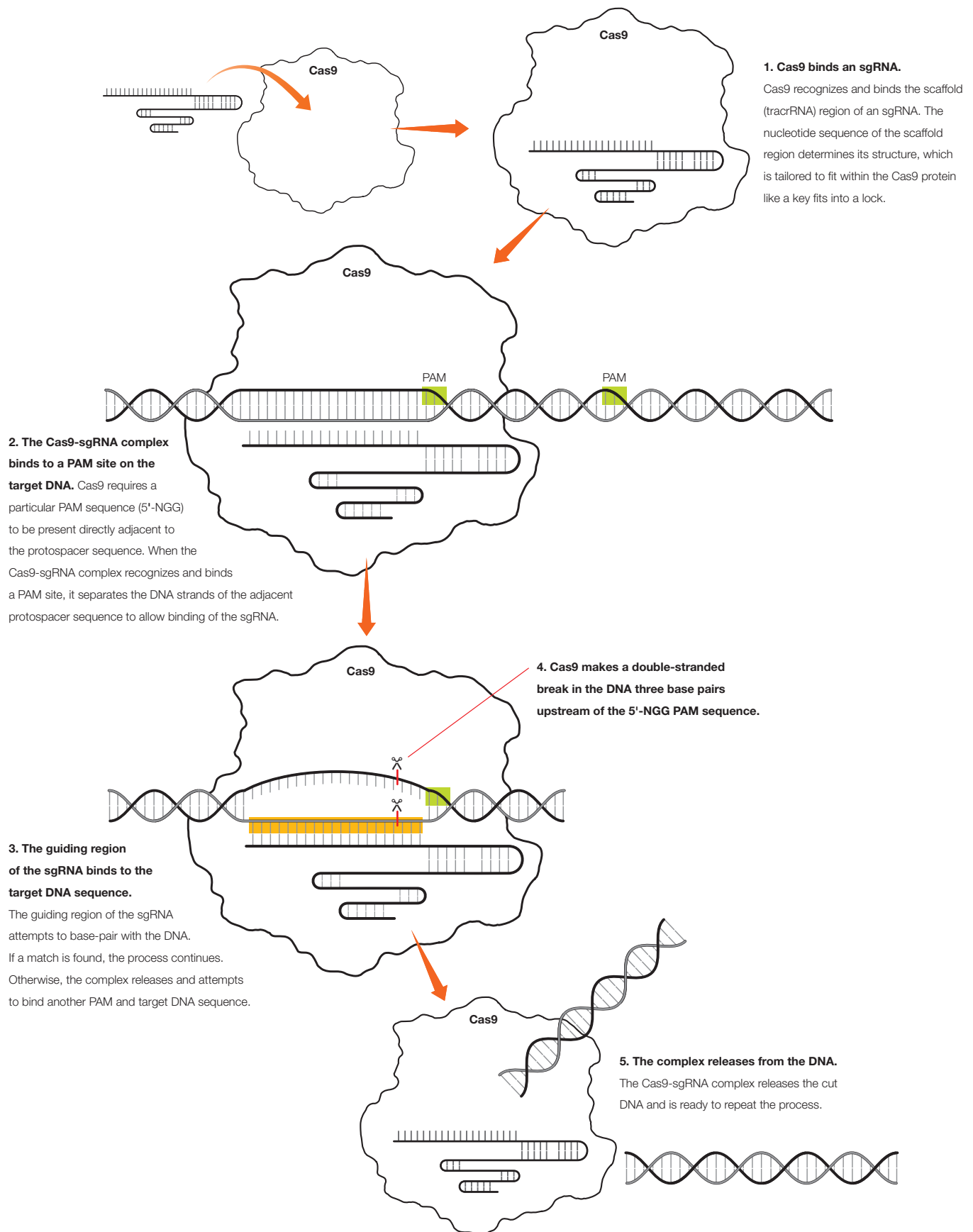


Fig. 2. The steps of Cas9 DNA recognition and cleavage.

Part 1. Simulate the Molecular Mechanism of Cas9 DNA Cleavage

Use the paper model to walk through the steps of CRISPR-Cas9 DNA cleavage using a sequence from the bacterial gene *lacZ* which encodes β -galactosidase. The *lacZ* gene is part of the *lac* operon, a collection of genes that allows bacteria to use lactose, a milk sugar, as a food source. The DNA and sgRNA sequences in the paper model match those used in Activity 2, *lacZ* CRISPR Gene Editing laboratory.

- 1. Cut out the sgRNAs and DNA strips. You may leave the Cas9 protein on its page.**
- 2. Use the steps in Figure 2 as a guide to model the CRISPR-Cas9 mechanism:**
 - a. Cas9 binds an sgRNA: Place sgRNA 1 onto the Cas9 illustration and align it with the dotted lines.
 - b. The Cas9-sgRNA complex binds to a PAM site. Place DNA strip 1 on the stripe across the Cas9 model. Slide the DNA strip until the PAM box on the Cas9 protein matches a PAM (5'-NGG) sequence on the DNA.
 - c. The guiding region of the sgRNA binds to the target DNA sequence. Check whether the DNA sequence is complementary to the sgRNA sequence (U pairs with A, C pairs with G). If they are complementary, continue the process. Otherwise, repeat steps 2.b and 2.c with a new PAM site.
 - d. Cas9 makes a double-stranded break in the DNA: The scissors icons indicate where Cas9 cuts the DNA strands. Use a pencil to draw a vertical line across both strands at this position.
- 3. Verify that you have chosen the correct cut site and then use a pair of scissors to cut DNA strip 1 at that site. Keep the pieces of DNA strip 1 for use in Part 4.**

Focus questions

A. How many nucleotides long is the guiding region of the sgRNA?

B. Does the sgRNA bind to the PAM?

C. Where does Cas9 cut the target DNA relative to the protospacer sequence?

Part 2. Design the Guiding Region of an sgRNA

CRISPR technology is powerful in part because the target DNA sequence is controlled by a customizable sgRNA. In this activity, you will customize the guiding region of the sgRNA to cut a target site on the *lacZ* gene. DNA strip 2 represents a DNA sequence from the *lacZ* gene where you wish to make an edit.

1. **Use sgRNA 2, DNA strip 2, and the steps you followed in Part 1 to determine the sgRNA guiding region sequence required to direct Cas9 to cut DNA strip 2 at the red dashed line.**
2. **Write the nucleotide letters (A, U, C, G) of this sequence into the spaces on sgRNA 2.**
3. **Use the steps of Cas9 DNA cleavage to confirm that the sequence you wrote on sgRNA 2 is correct.**
4. **Record your final sequence below or in a notebook.**

sgRNA guiding region sequence

Focus questions

A. Describe in complete sentences how the requirement of a PAM sequence affects the flexibility of CRISPR-Cas9 gene editing.

B. Describe in complete sentences how you would identify a target DNA cleavage site for CRISPR-Cas9 and design an sgRNA.

Part 3. Compare the Specificity of DNA-Cutting Tools

The flexibility and specificity of CRISPR-Cas9 technology offer a large step forward for gene editing. The first DNA “scissors” were restriction enzymes, which cut DNA at predefined sequences, typically 4–8 base pairs long. For example, EcoRI, a restriction enzyme found in *E. coli*, will cut double-stranded DNA at every GAATTC sequence. If EcoRI were added to a sample that contained the entire human genome, it could cut at every GAATTC sequence.

We can calculate the probability that a particular nucleotide sequence, such as GAATTC, will occur within a larger sequence. Table 1 below shows the calculated probabilities of finding sequences of particular lengths. These calculations are based on the assumption that DNA sequences are entirely random and that every nucleotide position has an equal probability of being A, T, C, or G. Use the table to answer the following questions.

Table 1. Calculated probabilities of finding a specific sequence.

Sequence	Sequence Length	Probability Calculation	Predicted Occurrence in a Sequence the Length of the Human Genome (3,234,830,000 bases)
A	1	$\frac{1}{4} = (1/4)^1 = 0.25$	808,707,500
AC	2	$\frac{1}{4} * \frac{1}{4} = (1/4)^2 = 0.0625$	202,176,875
GAATTC (EcoRI)	6	$\frac{1}{4} * \frac{1}{4} * \frac{1}{4} * \frac{1}{4} * \frac{1}{4} * \frac{1}{4} = (1/4)^6 = 2.44 \times 10^{-4}$	89,753
NNNN...	n	$\frac{1}{4} * \frac{1}{4} * \frac{1}{4} * \dots = (1/4)^n$	$(1/4)^n * 3,234,830,000$

Focus questions

- A. What is the probability that any base in a sequence is an adenine, A? How many times do you expect to find adenine in the human genome?**
- B. What is the probability of finding a particular two-base sequence? How many times do you expect to find that sequence in the human genome?**
- C. How many times would you expect to find an EcoRI cut site in a fragment of DNA 1,000,000 base pairs long?**

- D. How many times would you expect to find a specific 20 base pair sequence in the human genome?**
- E. Write out a complete equation to calculate the predicted occurrence of a sequence of n length within a DNA fragment of X length.**
- F. Using mathematical evidence, explain why CRISPR-Cas9 gene-cutting technology, which uses a target sequence of 20 base pairs, is more specific than classic restriction enzymes.**
- G. Write three different ideas you have about why CRISPR-Cas9 technology could be more useful for gene therapy and/or research than other gene-cutting tools.**
- H. In actuality, the DNA sequence of the human genome is NOT random. Some sequences, including some very large sequences, are repeated many times throughout the human genome. Write two ideas you have for how this fact complicates the use of CRISPR gene-editing technology in humans.**

Part 4. Design Donor Template DNA for DNA Repair

CRISPR-Cas9 can find a specific sequence in a genome billions of base pairs long and then cut at a precise location within that sequence. How do scientists and researchers use the specificity of CRISPR-Cas9 to direct targeted gene editing?

When chromosomal DNA in a bacterial cell is cut, the cell will die unless it's able to repair the cut. Bacteria have evolved processes to repair double-strand DNA breaks that would otherwise lead to cell death. DNA repair can happen in two ways, as shown in Figure 3:

- **Nonhomologous end joining (NHEJ)** — enzymes reconnect the ends of the double-stranded break back together. This process may randomly insert or delete one or more bases and can cause mutations that can disrupt gene function or expression
- **Homology directed repair (HDR)** — enzymes patch the break using donor template DNA. Researchers design the donor template DNA, which may include a desired sequence flanked on both sides by “homology arms” that match the sequence upstream and downstream of the cut. A complementary DNA strand is created during repair

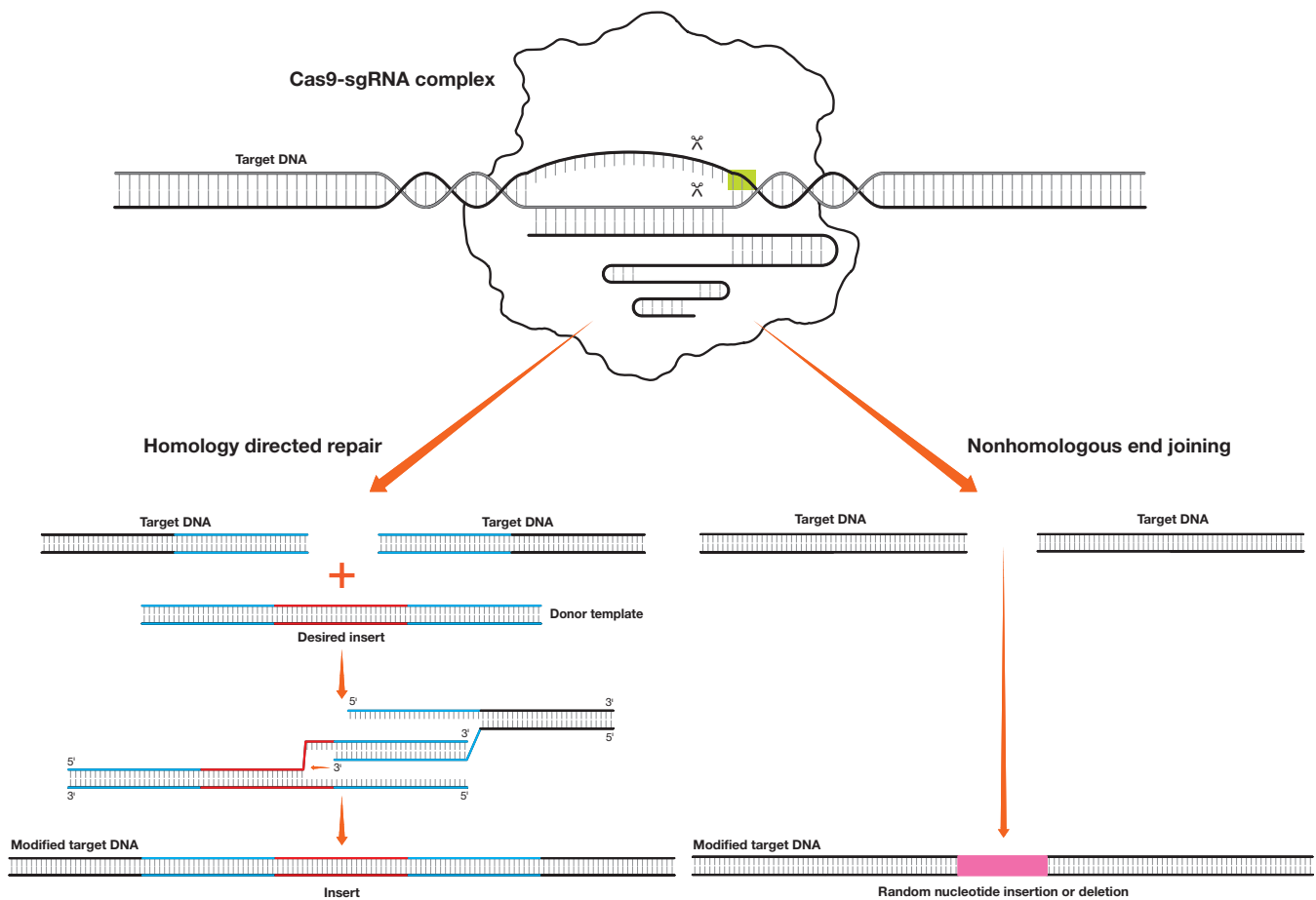


Fig. 3. DNA repair via homology directed repair and non-homologous end joining.

The homology arms used in HDR can be hundreds of base pairs or longer. For simplicity, you will simulate basic HDR in this activity using much shorter, 15 bp homology arms. You will design a donor template DNA sequence that could be used to insert a section of DNA into the cut site you created in DNA strip 1.

1. **Retrieve the two pieces of DNA strip 1. If you have not already used scissors to cut the strip at the cut site, do so now.**
2. **Cut out the donor template DNA strip and one blank DNA strip.**
3. **The shaded region simulates a nonspecific DNA insertion sequence. In the empty boxes on either side of the shaded region and on both strands, write the 15 bp sequences that match the nucleotide sequences on either side of the cut site of DNA strip 1. These 15 bp sequences are your homology arms.**
4. **You now have a complete donor template DNA.**
5. **Use scissors to cut the excess ends of the donor template strip.**
6. **Place the pieces of DNA strip 1 directly on top of the donor template strip nucleotide sequences so that the homology arms are aligned and only the insertion sequence is visible. Tape the pieces together.**
7. **You now have an edited piece of DNA.**

Use the blank DNA strip along with the rest of the paper model pieces to design donor template sequences with 15 bp homology arms that will induce each of the following changes to DNA strip 1 using sgRNA 1. You will need to include any necessary insertion sequences as well as homology arm sequences. Write the final sequences in the table below. Underline the homology arm regions.

Desired Change	Donor Template DNA Sequence
Cause a frameshift	
Insert an EcoRI restriction site	

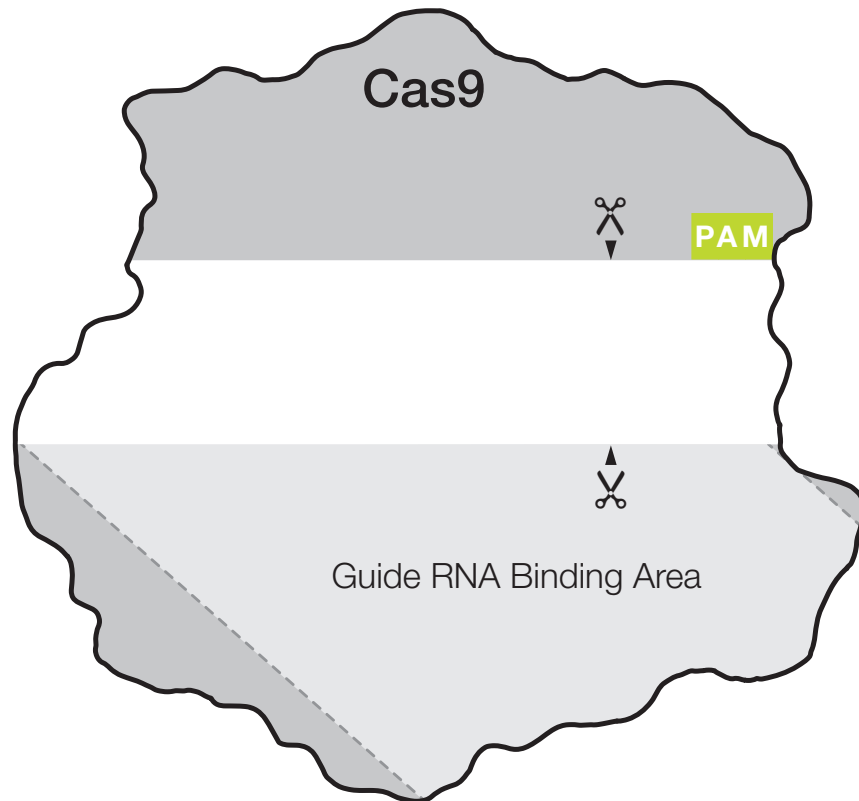
Focus questions

- A. **Describe two possible advantages of using HDR over using NHEJ in a gene editing experiment.**

B. Describe two possible advantages of using NHEJ over using HDR in a gene editing experiment.

C. Explain how CRISPR-Cas9 together with HDR could be used to change a single nucleotide, for instance changing a T to an A.

D. In addition to inserting or exchanging sequences, it is possible to remove short sequences near a cut site using HDR. Think of and describe an idea for how the donor template DNA sequence could be designed to cause such a removal. Use external resources about HDR as needed.



Glossary

Arabinose-inducible promoter — promoter that occurs naturally in bacterial systems and that is used in many expression plasmids to allow regulation of expression of a target gene: expression is induced in the presence of arabinose and repressed in its absence.

Cas9 — CRISPR-associated protein 9 (Cas9), an endonuclease that forms a double-strand break (cuts) in DNA at a specific site within a larger recognition sequence, or target site. It is involved in the natural defense of certain prokaryotes against DNA viruses, and it is heavily utilized in genetic engineering applications to cut DNA at locations specified by a guide RNA (gRNA)

CRISPR — clustered regularly interspaced palindromic repeats (CRISPR) are sequences in the genomes of some prokaryotes that act as a genomic record of previous viral attack. Along with CRISPR-associated (Cas) proteins, bacteria use the sequences to recognize and disarm future invading viruses. Scientists have adapted this system for genetic engineering purposes.

Donor template DNA — engineered sequence of DNA required for homology-directed repair in CRISPR gene editing applications; may include a desired sequence flanked on both sides by “homology arms” that match the sequence upstream and downstream of the cut.

β -galactosidase — encoded by the *lacZ* gene, this enzyme hydrolyzes galactose-containing carbohydrates, including lactose. Conveniently, it also breaks down the colorless compound *X-gal* into two pieces, one of which goes on to form a deep blue pigment.

Guide RNA (gRNA) — non-coding, short RNA sequence that binds to Cas9 and to complementary target DNA sequences, where Cas9 performs its endonuclease activity to cut the target DNA strand.

Guiding region — part of the CRISPR RNA or crRNA in nature, a typically 20-nucleotide region of sgRNA that is complementary to the target DNA sequence and that defines where Cas9 cuts. Scientists can easily customize this sequence for their own targets.

Homology directed repair (HDR) — DNA repair mechanism in which specific proteins patch a double-strand DNA using donor template DNA.

Isopropyl β -d-1-thiogalactopyranoside (IPTG) — a non-metabolizable analog of lactose, which induces transcription of the lac operon

lacZ — part of the *lac* operon in *E. coli*, this gene encodes the enzyme β -galactosidase. For decades, molecular biologists have used the *lacZ* gene as a target site for inserting DNA sequences because the resulting bacterial colony color indicates whether insert was successful.

Non-homologous end joining (NHEJ) — DNA repair mechanism in which specific proteins reconnect the ends of a double-strand DNA break. This process may randomly insert or delete one or more bases that can disrupt gene function or expression.

Protospacer — DNA region targeted for cleavage by the CRISPR system.

Protospacer adjacent motif (PAM) — sequence motif immediately adjacent to the protospacer sequence in the Cas9 recognition sequence that is required for Cas9 function. Cas9 recognizes the PAM sequence 5'-NGG where N can be any nucleotide (A, T, C, or G). When Cas9 binds the PAM, it separates the DNA strands of the adjacent sequence to allow binding of the sgRNA. If the sgRNA is complementary to that sequence, Cas9 cuts the DNA.

Scaffold region — called the trans-activating CRISPR RNA or tracrRNA in nature, a region of sgRNA that forms a multi-hairpin loop structure (scaffold) that binds tightly in a crevice of the Cas9 protein. The sequence of this region is typically the same for all sgRNAs.

Single guide RNA (sgRNA) — engineered form of guide RNA that forms a complex with Cas9; ~100 nucleotide fusion of two regions that occur as separate guide RNAs in nature: the guiding region (crRNA) and the scaffold region (tracrRNA).

X-gal — 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside, a compound consisting of galactose linked to a substituted indole. Its hydrolysis by β -galactosidase yields an insoluble blue pigment and can be used in bacterial cultures to indicate the presence of active β -galactosidase.