Session 1: CRISPR Crash Course

Learning Goals:

- Learn about the history and uses of CRISPR-Cas tools
- Use the Protein Database to visualize protein structures in 3-dimensional space

Background:

Many people are casually familiar with the term “CRISPR” as a promising and popular tool for DNA editing. Since its initial discovery in 2012, CRISPR systems have been used across branches of life for a variety of applications in medicine, agriculture, and biotechnology. In 2020, scientists Jennifer Doudna and Emmanuelle Charpentier became the first all-woman team to win a Nobel Prize in Chemistry for their discovery of CRISPR. But where does CRISPR actually come from, and how does it work?

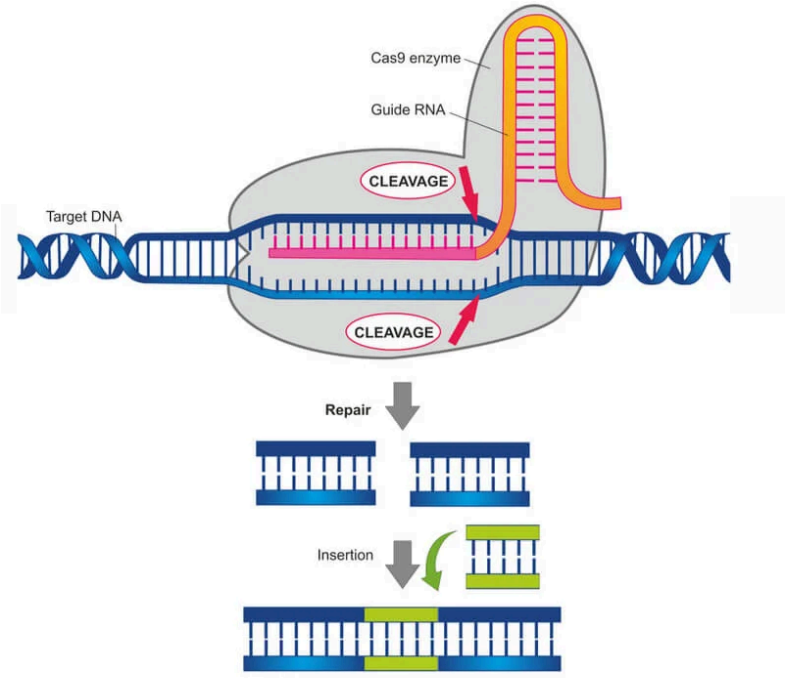
What are CRISPR-Cas systems?

CRISPR, or CRISPR-Cas systems, are a broad class of enzymes (*Cas proteins*) that are able to identify and introduce double-stranded breaks (*cleave*) into specific DNA or RNA sequences. CRISPR actually originally evolved in bacteria as a type of immune response, called *adaptive immunity*, to protect against viruses (*bacteriophages*). The CRISPR-Cas system uses a guide RNA that recognizes DNA or RNA sequences specific to the invading bacteriophage. The guide RNA leads the Cas system to the target DNA or RNA, where it is able to selectively cleave the nucleic acid and prevent the virus from replicating.

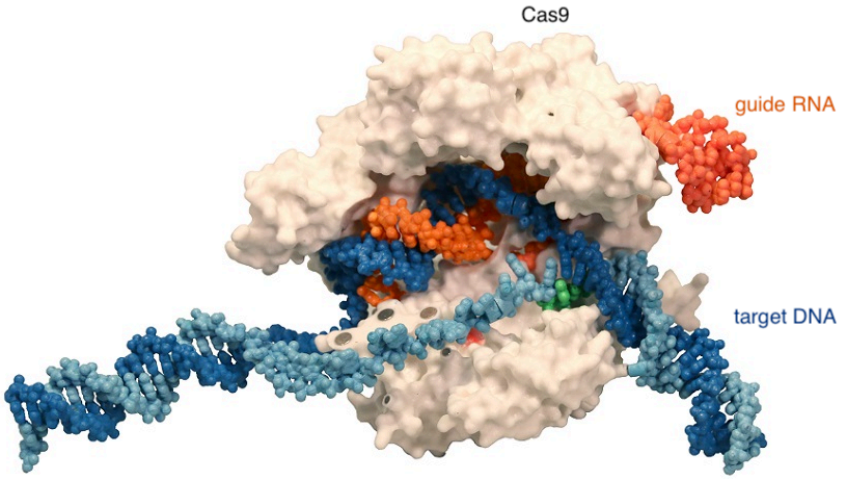
Uses of CRISPR-Cas systems

The most widely used CRISPR system is Cas9, which recognizes DNA targets for cleavage. Through DNA cleavage, CRISPR systems are able to easily introduce *gene knockouts,* where a critical piece of genetic code is lost, ablating the function of the gene. By providing short DNA or RNA templates with matching 5’ and 3’ ends to the cleavage site, gene insertions and edits can also be made using Cas9.

Since these initial applications, applications of CRISPR-Cas tools have expanded. Many applications rely on a deactivated form of Cas9 (*dCas9*) that is still able to recognize DNA targets, but lacks the ability to cleave DNA. Using dCas9, researchers have shown the ability to introduce large-scale *gene* knock-ins and targeted base edits, as well as repressing and upregulating specific genes within the genome.



https://www.labiotech.eu/in-depth/crispr-cas9-review-gene-editing-tool/



https://3dmoleculardesigns.com/product/crispr-cas9-mini-model/

In this course, we will use dCas9 to abolish the function of a gene in the model species *E. coli* using two different approaches. In one approach, we will directly repress the expression of the DNA (*transcription*, more on this in **Session 3**), while leaving the DNA sequence unchanged from the original sequence. In the other approach, we will edit the

underlying DNA sequence, making it so the cell is unable to properly read the DNA to make the correct gene.

First, in today's session, we will learn how to use computational tools to visualize the dCas9 structure in complex with its gRNA to gain a biophysical understanding of the enzyme's function.

Lab Tasks:

1. Install PyMOL, a python-based, open-source molecular visualization tool.
    a. If you are unable or don't want to install PyMOL on your personal computer, please follow along with another participant.
2. Navigate to the online Protein Database: RCSB PDB and type "6K57" into the search bar.
    a. This should return an entry for a protein structure of dCas9 in complex with its guide RNA (sgRNA) and target DNA.
    b. If you are interested in learning more about how 3D protein structures are generated, you can read more at Analyzing Protein Structure and Function - Molecular Biology of the Cell and PDB-101: Learn: Guide to Understanding PDB Data: Methods for Determining Structure
3. Explore a bit on the PDB page! See if you can identify for this structure:

Number of amino acids in dCas9 (*sequence length*)	
sgRNA length	

4. Open PyMOL. In the command line, type "`fetch 6K57`". The structure from the PDB website should auto-load in the PyMOL viewer. Spend some time navigating the structure: try zooming in and out and rotating the view. Can you tell what is the protein, the DNA, and the RNA?
5. Under *Display*, check *Sequence.* Under *Display* > *Sequence Mode*, select *Residue Names.* This will display amino acids and nucleotides by their three and single letter codes. For example, you can find the serine residue at position 6 by scrolling to "SER" under the number 6.
6. This structure depicts dCas9, its guide RNA, and the target DNA. There are nucleic acids shown in the sequence at both the beginning and end - can you tell which corresponds to DNA and which to RNA?
7. Select the RNA nucleotides in the sequence menu. See if you can change the color to make it stand out more.
    a. Select nucleotides > right click > 'color' > pick a color you like!

   b. Repeat with the DNA nucleotides.
8. Looking at the RNA structure, see if you can identify which nucleotides form the “hairpin” structure that binds to the Cas9 enzyme.
9. Select the RNA nucleotides by using ctrl+clicking from the start to the end of the sequence. Then right click, go to “find” → “polar contacts” → “within selection”. This will highlight all the interactions between RNA bases. Can you identify where the hydrogen bonds between RNA bases are?

If you finish early, here are some other fun protein structures you can play around with!

- 6MDR
- 3LJ5

Resources:

There are a massive amount of informative & accessible resources about CRISPR tools online - here are some of our favorites:

1. What is CRISPR: The Ultimate Guide | Synthego
   https://www.synthego.com/learn/crispr

2. What is CRISPR? - Explained by Jennifer Doudna & IGI Experts
   https://innovativegenomics.org/education/digital-resources/what-is-crispr/

3. Addgene: CRISPR Guide
   https://www.addgene.org/guides/crispr/#overview