CRISPR/Cas paper model: sickle cell gene therapy

CRISPR/Cas9: a revolutionary genome editing tool

Modifying the DNA in a living cell is called genome editing. Recent scientific advances have made it possible to alter the DNA in all sorts of cells, even human cells. While scientists previously had a limited set of tools to modify DNA, the CRISPR/Cas system revolutionized genome editing because it allows researchers to target regions of the genome far more precisely than ever before.

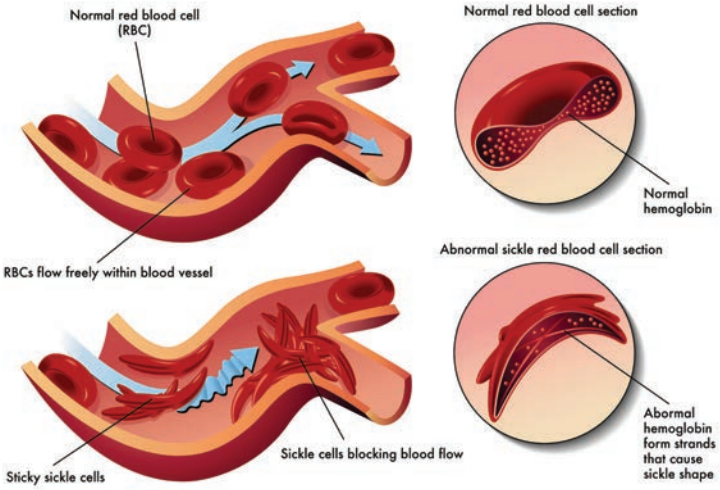


Figure 1. Sickle cell disease. Red blood cells are flexible and can squeeze through tiny capillaries (top left), but sickle shaped red blood cells clump together and block blood flow (bottom left). Normally, hemoglobin protein floats freely in red blood cells (top right). In sickle cell disease, hemoglobin protein forms long strands that distort the red blood cell into a sickle shape (bottom right).

Illustration courtesy of Diana Grib under Creative Commons license.

Sickle cell disease

Sickle cell disease is a hereditary condition that affects red blood cells. Red blood cells need to be flexible to squeeze through the tiny capillaries in our bodies. Sickled red blood cells are rigid and get stuck in tiny blood vessels (Figure 1). Depending on where the blockage occurs, this can cause a range of symptoms, from pain to infections or even death.

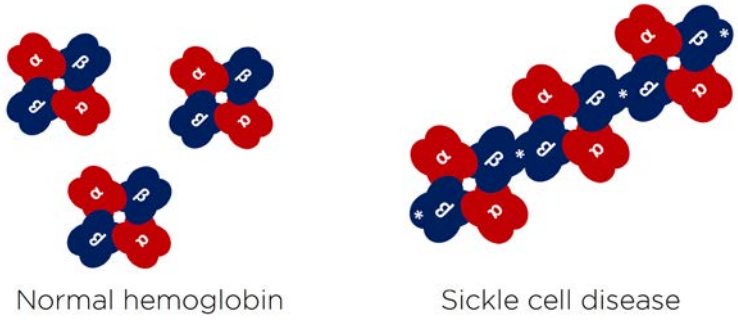


Figure 2. Hemoglobin protein clumps into long strands in sickle cell disease. Hemoglobin protein is composed of two α-globin subunits and two β-globin subunits. Normal hemoglobin protein floats freely inside the red blood cell (left). Sickle cell disease is caused by a mutation in β-globin (indicated by an asterisk). The sickle cell mutation makes β-globin from one hemoglobin protein stick to β-globin from another hemoglobin protein (right). The resulting strands of mutant hemoglobin distort the red blood cell into a sickle shape.

Sickle cell disease is caused by a change in a single DNA base that leads to a small structural change in the protein hemoglobin. Hemoglobin is found in red blood cells, and is responsible for transporting oxygen to all the cells of your body. A complete hemoglobin protein is actually a complex of four smaller protein subunits that are encoded by two separate genes. Hemoglobin has two alpha-globin (α-globin) subunits and two beta-globin (β-globin) subunits (Figure 2). Normally, hemoglobin protein floats freely inside red blood cells. In sickle cell disease, a mutation in the β-globin gene causes the hemoglobin proteins to clump together into long strands (Figure 2), which distorts the red blood cells into a sickle shape (Figure 1).

While we have understood the cause of sickle cell disease for decades, most treatments can only manage a patient's symptoms; however, advances in genome editing have made curing sickle cell disease a possibility. One option would be to correct the mutation in the β-globin gene. While research is underway to do just this, the first successful attempt to use genome editing to treat sickle cell disease took a different approach.

Fetal hemoglobin

The hemoglobin we have been discussing so far is called adult hemoglobin. This is because during development, humans produce a different form of hemoglobin called fetal hemoglobin. Like adult hemoglobin, fetal hemoglobin contains two α-globin subunits, but instead of using β-globin, it contains two gamma-globin (γ-globin) subunits (Figure 3). Around the time of birth, a genetic switch flips, and our cells stop using the γ-globin gene. Instead, cells start making the adult form of hemoglobin with two α-globin and two β-globin subunits. Importantly, even though γ-globin stops being produced shortly after birth, it is equally functional in adults.

Scientists reasoned that if they could increase γ-globin production in adult red blood cells, the presence of functional fetal hemoglobin protein might decrease the clumping of adult hemoglobin (Figure 4). The advent of CRISPR/Cas9 genome editing provided scientists with a way to accomplish this goal.

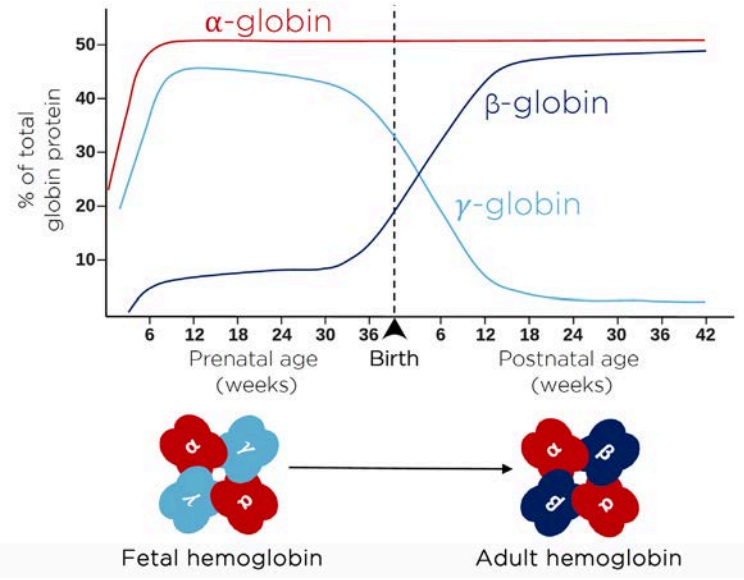


Figure 3. Transition from fetal hemoglobin to adult hemoglobin. Before birth, the majority of hemoglobin protein is fetal hemoglobin, which contains two α-globin subunits and two γ-globin subunits. After birth, the majority of hemoglobin protein is adult hemoglobin, which contains two α-globin subunits and two β-globin subunits. The transition occurs when γ-globin production decreases and β-globin production increases.

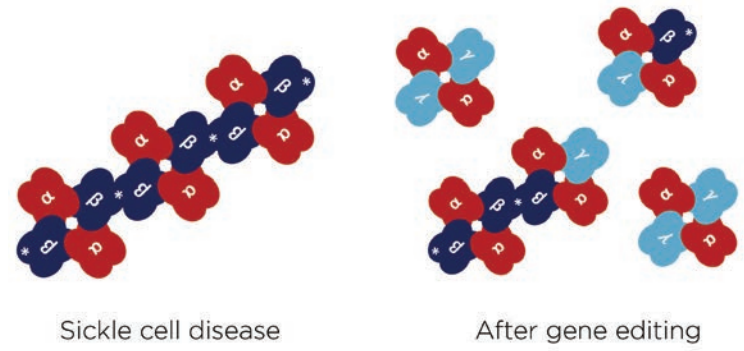


Figure 4. A possible treatment for sickle cell disease. In sickle cell disease, a mutation in β-globin causes adult hemoglobin to clump together (left). If red blood cells continued to produce γ-globin, the presence of functional fetal hemoglobin could reduce clumping of adult hemoglobin with the sickle cell mutation (right).

Using CRISPR/Cas9 to treat sickle cell disease

Before we get into the specifics of the genome editing strategy that scientists used to treat sickle cell disease, we need to discuss how CRISPR/Cas genome editing works. The CRISPR/Cas system involves two main components: a Cas nuclease and a guide RNA. Nucleases are enzymes that cut nucleic acids like molecular scissors. There are several Cas nucleases, but Cas9 is most commonly used in genome editing. Cas9 is a powerful tool because it can be programmed to cut nearly any DNA sequence.

Where Cas9 will cut is determined by a short RNA molecule called a guide RNA that binds to the Cas protein (Figure 5). After the guide RNA binds to Cas9, the complex scans the genome for a three base sequence called the PAM. The Cas9 PAM sequence is 5' NGG 3', where N can be any base. When Cas9 encounters the PAM sequence, it unzips the DNA, separating it into single strands. Cas9 then uses the guide RNA to establish whether or not to cut the DNA. At one end of the guide RNA are ~20 bases that determine which DNA sequence Cas9 will cut. If this ~20 base sequence in the guide RNA is complementary to the DNA, then Cas9 will cut both strands of the DNA. If the guide RNA does not match the DNA, then the complex will move on to the next PAM site, and the double helix will re-zip into double stranded form. The trick to using Cas9 as a gene editing tool is that scientists can customize this ~20 base sequence to target Cas9 to a specific region of DNA, basically allowing them to program where Cas9 will cut.

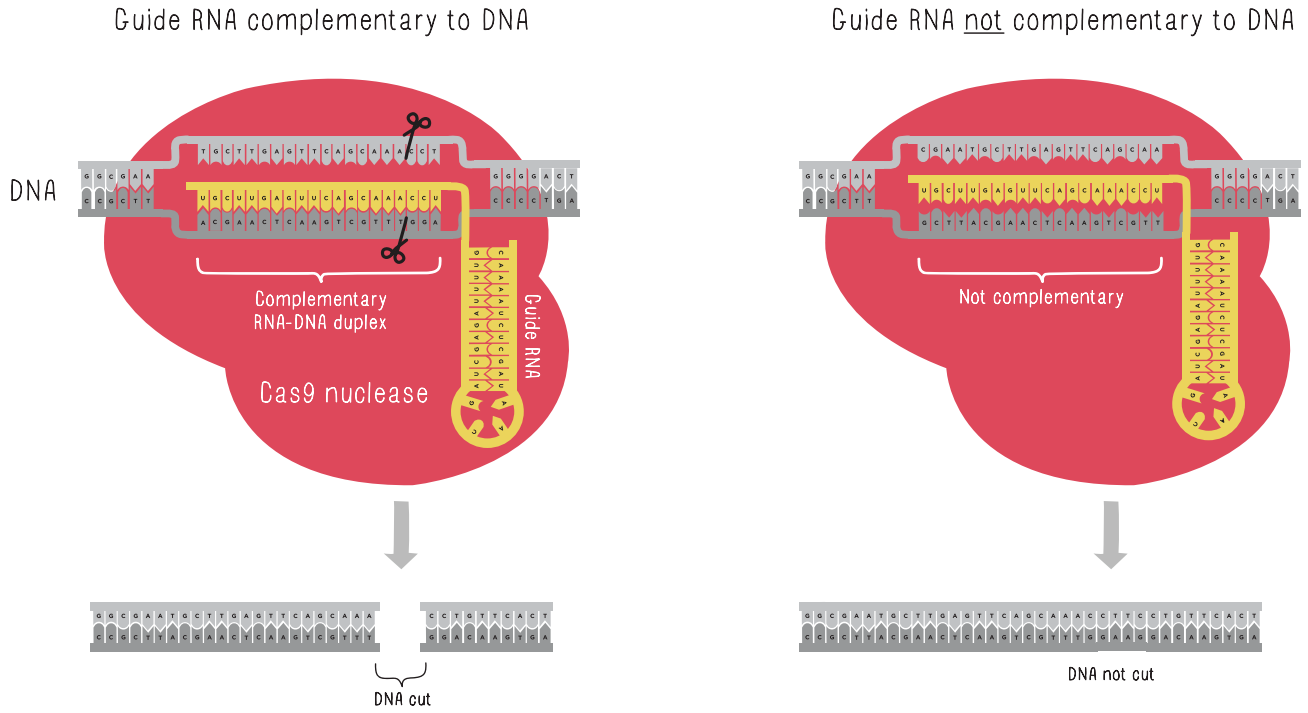


Figure 5. Cas9 cuts DNA as instructed by a guide RNA. When the guide RNA encounters a complementary DNA sequence, Cas9 will cut the DNA (left). When the guide RNA is not complementary to the DNA, Cas9 will not cut (right).

In eukaryotic cells, once Cas9 cuts the target DNA, the cell will try to repair the break. A common fix is for the cell to reattach the broken strands of DNA through a process called non-homologous end joining (NHEJ). When the cell does this, it often ends up adding or removing a few DNA bases. These act much like typos in the DNA. These typos, or mutations, can disrupt a gene or other DNA sequence. This is the case with a proposed genome editing treatment for sickle cell disease.

Continued on the next page

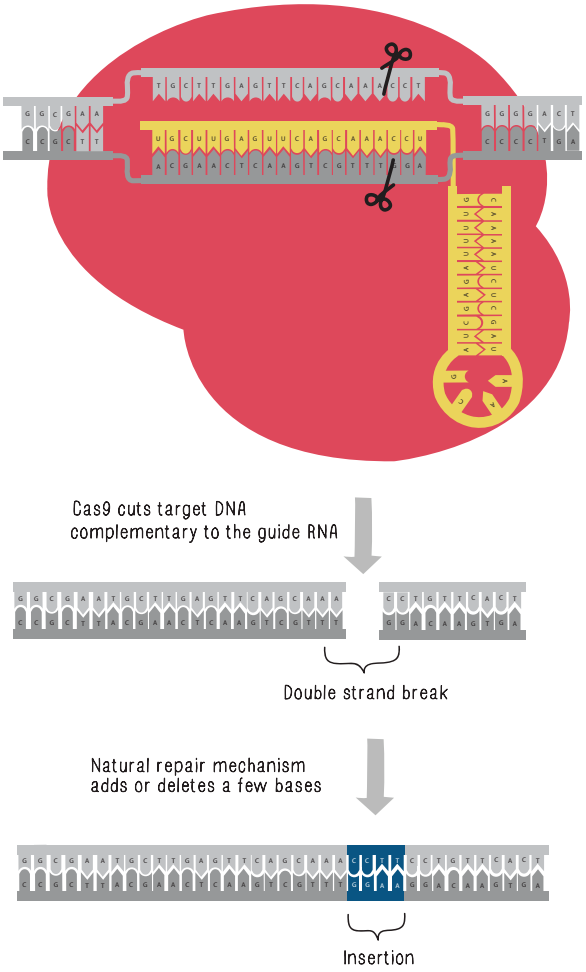


Figure 6. DNA breaks can lead to mutations. Breaks in the DNA can be repaired by two different mechanisms in eukaryotic cells. The more common repair mechanism, non-homologous end joining, reattaches the broken strands of DNA, but in the process often adds or removes a few random DNA bases. These insertions and deletions often disable the gene.

Increasing fetal hemoglobin in adult red blood cells

Scientists recently discovered the nature of the genetic switch that turns off the production of γ-globin shortly after birth. Another protein called BCL11A turns off γ-globin production (Figure 7). To increase fetal hemoglobin, scientists tried to dial down the production of BCL11A protein in red blood cells. They set out to accomplish this by mutating a regulatory DNA sequence that influences how much BCL11A protein is made in red blood cells (Figure 7). With less BCL11A protein, adult red blood cells should produce fetal hemoglobin again. Scientists reasons that in patients with sickle cell disease, the production of fetal hemoglobin would lead to fewer sickled red blood cells. The first human patients whose blood cells were edited to contain a mutation in the *BCL11A* regulatory DNA are still being monitored closely, but early clinical results are promising.

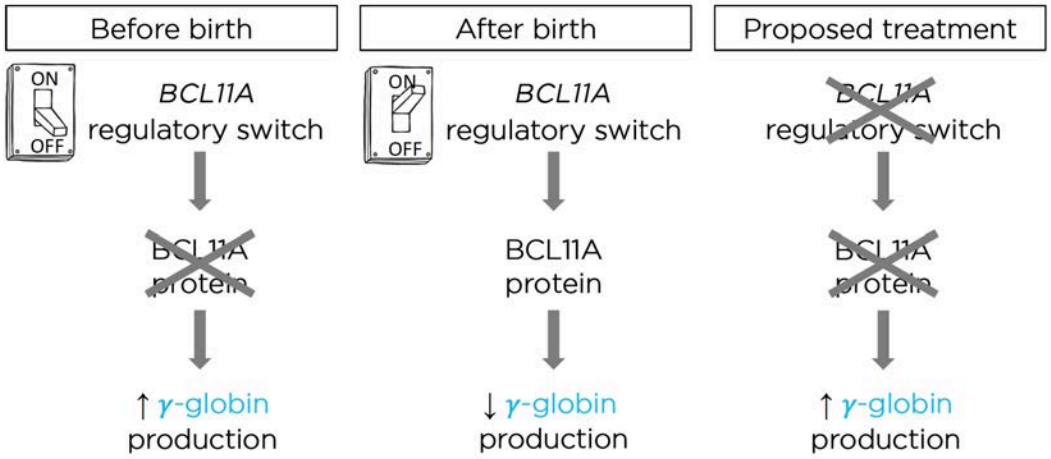


Figure 7. Increasing fetal hemoglobin to treat sickle cell disease. During fetal development, the BCL11A protein is not produced, so γ-globin protein is made at high levels (left). After birth, BCL11A protein is produced, which leads to low levels of γ-globin production (middle). To try to increase γ-globin production in patients with sickle cell disease, scientists used CRISPR/Cas genome editing to mutate the regulatory DNA responsible for allowing BCL11A protein to be produced. In the absence of BCL11A protein, γ-globin protein can be produced in adult red blood cells (right).

Reading questions

1. Sickle cell disease is characterized by the presence of sickle shaped red blood cells. According to the text, what health problems can sickle shaped red blood cells cause?

2. Which globin subunits make up adult hemoglobin protein? Select all that apply.
   a. α-globin
   b. β-globin
   c. γ-globin

3. Which globin subunits make up fetal hemoglobin protein? Select all that apply.
   a. α-globin
   b. β-globin
   c. γ-globin

4. Which gene is mutated in sickle cell disease?
   a. α-*globin*
   b. β-*globin*
   c. γ-*globin*

5. What protein switches off the production of fetal hemoglobin after birth?

6. Complete the following sentences by circling the correct term in parentheses:
   When BCL11A protein production is high, then fetal hemoglobin production is **(high/low)**.
   When BCL11A protein production is low, then fetal hemoglobin production is **(high/low)**.
   Therefore, to increase fetal hemoglobin production after birth, scientists want to **(increase/decrease)** the production of BCL11A protein in red blood cells.

7. Which DNA sequence is being targeted for genome editing in the first clinical trial using CRISPR/Cas to cure sickle cell disease?
   a. β-*globin* gene
   b. γ-*globin* gene
   c. *BCL11A* regulatory DNA

8. Critical thinking - Why would the production of fetal hemoglobin be expected to decrease clumping of adult hemoglobin in patients with sickle cell disease? Hint: think about the different globin subunits contained in each form of hemoglobin.

Modeling CRISPR/Cas

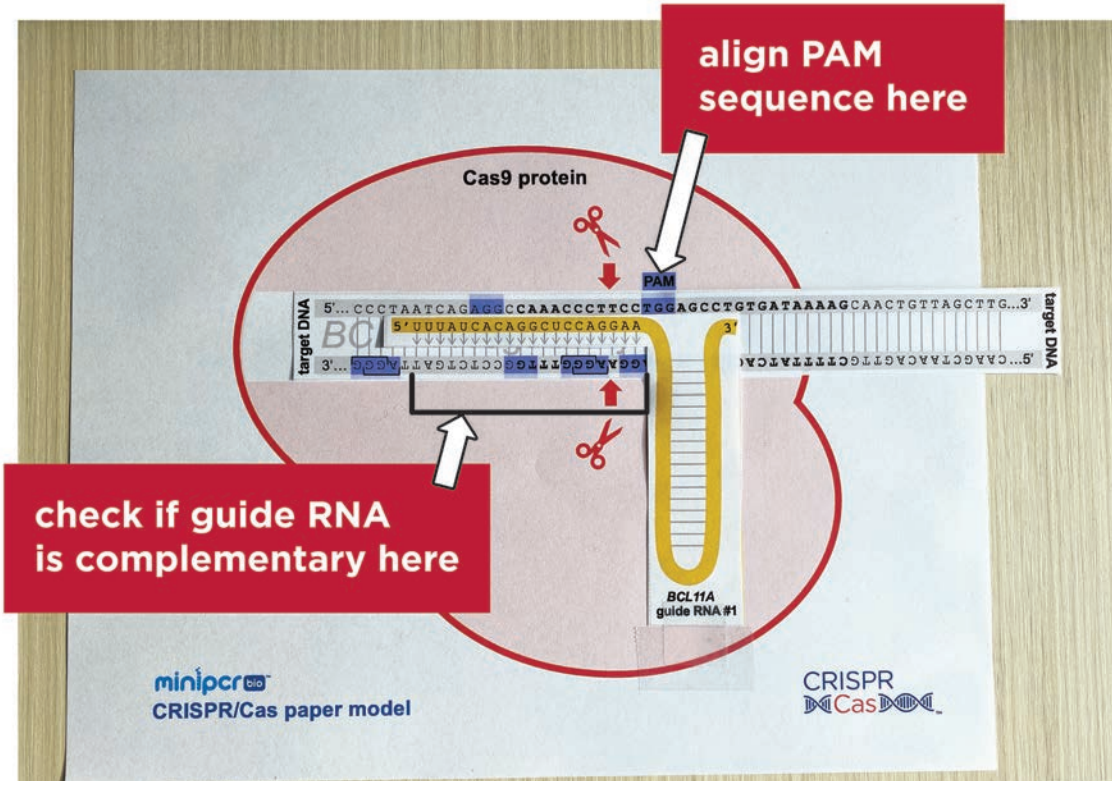
Now you will use a paper model to illustrate the genome editing process used in sickle cell patients!

Supplies needed: tape, scissors, single-sided print out of CRISPR/Cas9 paper model (pages 14-15)

1. Cut out the two long strips of DNA labeled "target DNA" (page 15). This sequence includes a section of the *BCL11A* regulatory DNA. The part of the sequence marked in bold has been called an "Achilles' heel" because disrupting this small area of DNA results in decreased BCL11A production, leading to substantial increases in γ-globin production.
2. Cut out Guide RNA #1 (page 15). This guide has been designed to target Cas9 to disrupt this "Achilles' heel" region of the *BCL11A* regulatory DNA.
3. Cut out Guide RNA #2 (page 15). This guide is blank. You will fill in a sequence that you design to mutate the *BCL11A* regulatory DNA.

Scientists start with blood stem cells that have been removed from a patient's bone marrow. Use the paper model to walk through what happens in the patient's cells after scientists add Cas9 and the guide RNA. Remember that the goal is to mutate the *BCL11A* regulatory DNA.

4. First, the guide RNA associates with Cas9 to form a complex.
    - Place the ***BCL11A* guide RNA #1** onto the yellow zone in the Cas9 protein.
    - Be sure to line up the RNA carefully.
    - Place a piece of tape across the guide RNA where it says "guide RNA" to hold it in place.



5. Then the guide RNA/Cas9 complex scans the genome for a PAM sequence.
   - Place one copy of the ***BCL11A* regulatory DNA** on the white stripe on the Cas9 protein. The target DNA should be under the guide RNA so that you can still read the sequence on both strands of the DNA (see picture above).
   - It doesn't matter which way you put the DNA in since Cas9 will check each strand of DNA for PAM sequences in turn.
   - Slide the target DNA until the purple PAM box on the Cas9 protein aligns with a purple PAM sequence on the target DNA.
   - Potential PAM sequences are highlighted in purple. When multiple PAM sequences are in a row or overlapping, they are also marked with brackets.

6. Once Cas9 encounters a PAM sequence, the guide RNA will have a chance to pair with the DNA sequence (see picture above, U pairs with A and C pairs with G).
   - If the guide RNA and the target DNA are not complementary, slide the target DNA to the next PAM site and repeat this step.
   - Remember: The PAM sequence is three nucleotides long. Sometimes, however, there can be multiple PAM sequences in a row. These are indicated on the DNA with brackets.
   - When you find a complementary sequence, circle the PAM site on the DNA that is aligned with the PAM box on Cas9, and then advance to step 7.
   - If you find no matches on one strand of DNA, take the DNA out and turn it around to try the other strand.

7. If the guide RNA and the target DNA are complementary, Cas9 will cut the DNA.
    - Mark both strands of the target DNA where the pink arrows point.
    - Take the DNA out and use scissors to cut both strands where you marked the DNA.

Questions

A. What is the Cas9 PAM sequence?
5’ __ __ __ 3’

B. How does the requirement that Cas9 bind to a PAM sequence affect the ability of scientists to target Cas9 to the exact DNA sequence they are interested in?

You have successfully cut the *BCL11A* regulatory DNA, but it isn’t mutated yet...

8. When the cell’s DNA repair machinery detects that the DNA has been cut, it will move in to repair the break. The most common repair mechanism (NHEJ) adds or removes some nucleotides from the cut ends of the DNA, then re-attaches the two ends.
    - Cut off between 0-5 bases from each end of the DNA at the cut site. You get to pick how much DNA to cut off—there is no right or wrong answer here.
    - Tape the two ends of the DNA back together where you cut them.

Continued on the next page

The next step is to sequence the cell's DNA to see what mutations were generated. Below is a section of the original (wild-type) *BCL11A* regulatory sequence and two different mutant sequences that might be generated after Cas9 cuts the DNA. The PAM is marked in purple. If there is a "-" it means those bases were removed from the sequence, much like you did with scissors.

Wild-type *BCL11A*	5' CAAGCTAACAGTTGCTTTTATCACAGGCTCCAGGAAGGGTTTGGCCTCTGATTAGGG 3'
Mutant 1	5' CAAGCTAACAGTTGCTTTTATCACAGGCT-------GGGTTTGGCCTCTGATTAGGG 3'
Mutant 2	5' CAAGCTAACAGTTGCTTTTATCACAGGCTCC----AGGGTTTGGCCTCTGATTAGGG 3'

Questions

C. Compare the repaired DNA sequence from your paper model to the wild-type sequence. On the "your result" sequence below (indicated by the arrow), cross out the DNA bases that were deleted when you repaired the DNA.

Wild-type *BCLA11A*	5' CAAGCTAACAGTTGCTTTTATCACAGGCTCCAGGAAGGGTTTGGCCTCTGATTAGGG 3'
Your result	5' CAAGCTAACAGTTGCTTTTATCACAGGCTCCAGGAAGGGTTTGGCCTCTGATTAGGG 3'

D. Is your resulting sequence the same as either of the example mutant sequences?

E. Compare your mutation to those made by other students. Did any groups create the same mutants? Using just this one guide RNA, how many different mutations were created?

F. NHEJ introduces random mutations. Considering this, why do you think it is important for scientists to sequence the cell's DNA after CRISPR/Cas genome editing?

Mission accomplished! Now you have used the CRISPR/Cas system to mutate the *BCL11A* regulatory DNA in a patient's blood cells. After you transplant these cells back into the patient, there should be an increase in the production of fetal hemoglobin protein and a decrease in the number of sickled red blood cells.

When scientists want to use CRISPR/Cas to disrupt a target DNA sequence, they typically test multiple guide RNAs. Now that you've got some experience with the CRISPR/Cas system, try to design your own guide RNA to mutate the *BCL11A* regulatory DNA!

9. Design another guide RNA that could be used to disrupt the *BCL11A* regulatory sequence in the "Achilles' heel zone" marked in bold. There is more than one acceptable answer to this question.
    - Remove guide RNA #1 and place the blank ***BCL11A* guide RNA #2** onto the Cas9 protein. Be sure to line up the RNA carefully. Place a piece of tape across the guide RNA where it says "*BCL11A* guide RNA" to hold it in place.
    - Place the uncut ***BCL11A* regulatory DNA** strip on the white stripe on the Cas9 protein. The target DNA should be under the guide RNA, but placed so that you can still read the sequence on both strands of the DNA.
    - Slide the target DNA so that a purple PAM sequence aligns with the purple PAM box on the Cas9 protein.
    - Where the pink arrows align on the DNA is where Cas9 will cut if this PAM is used. Remember that your goal is to disrupt the "Achilles' heel zone" using Cas9.
    - Once you have chosen an appropriate PAM sequence, use the *BCL11A* sequence to fill in the matching guide RNA sequence in the empty boxes. The guide RNA should be complementary to the bottom strand of DNA in this model. Remember that RNA uses U instead of T.
    - Also, record your guide sequence here:
      5' __ 3'

10. Remove the DNA and give your completed ***BCL11A* guide RNA #2** to another group and have them test it to see if it will lead to Cas9 cutting the *BCL11A* regulatory DNA in the "Achilles' heel" zone.

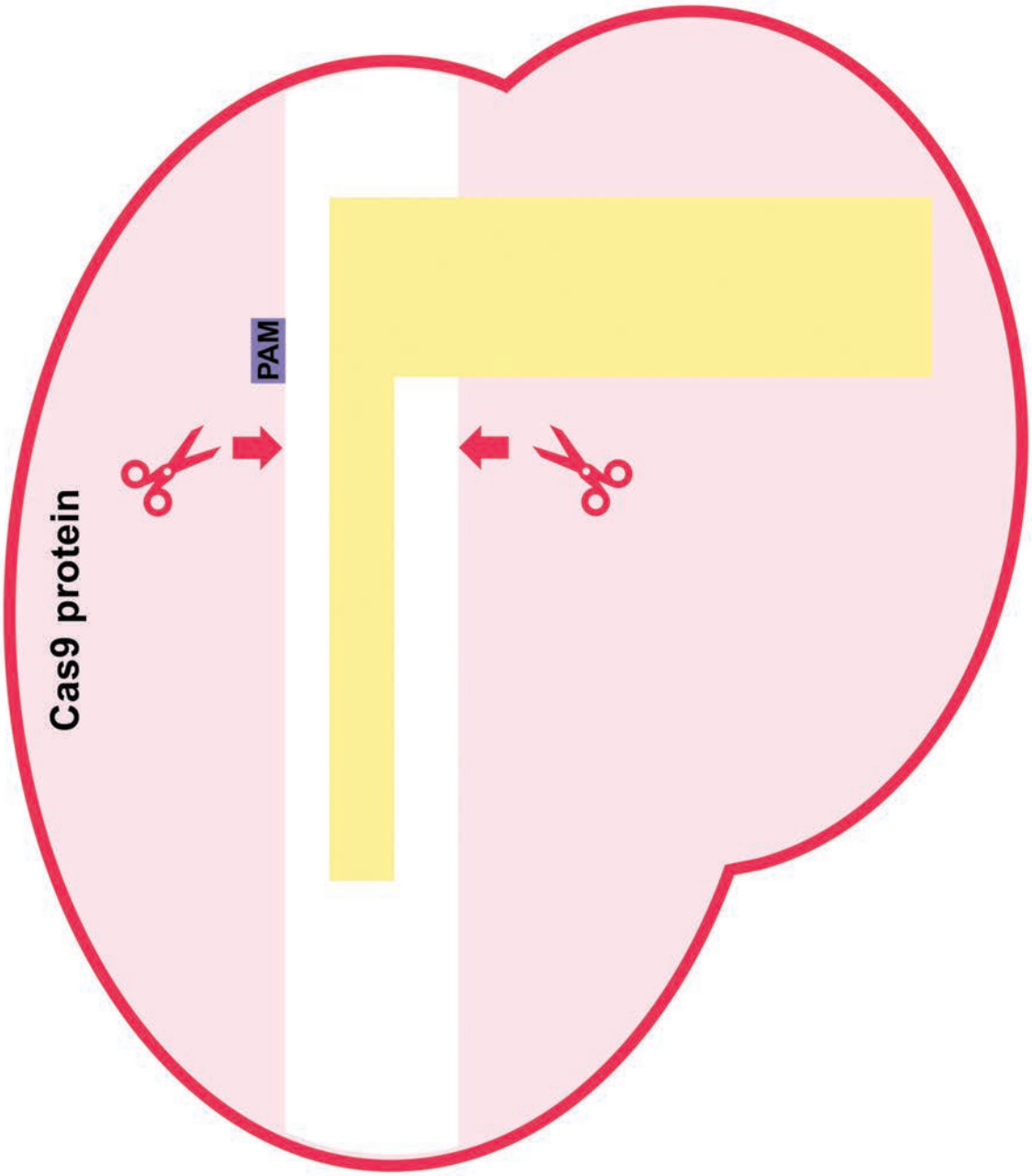
Just like that, you've designed a guide RNA that could be used to cure sickle cell disease! This exercise shows one of the main advantages of CRISPR/Cas for genome editing: that knowledge of the simple base-pairing rules is pretty much all you need to target a specific DNA sequence for modification.

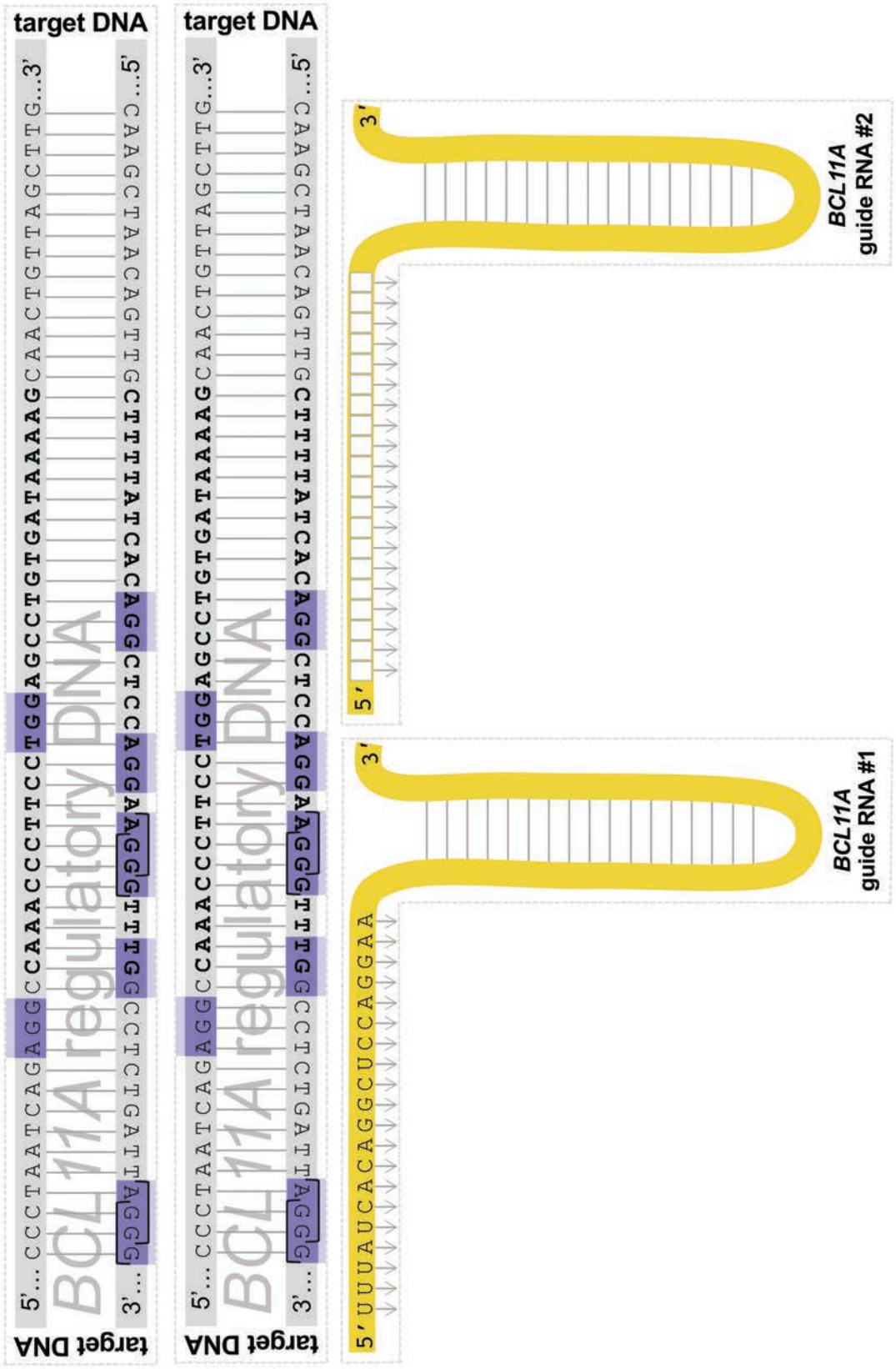
Questions

G. Compare your guide RNA sequence with your classmates'. How many different guide RNAs did your class design to disrupt just this short region of the *BCL11A* regulatory DNA?

H. Why is it helpful for scientists that the Cas9 PAM sequence is relatively common in the genome?

I. DNA is said to have a universal genetic code. Cas9 is a bacterial protein. Explain why a universal genetic code allows us to use any organism, including humans, to express the *cas9* gene to make a protein.


PAM
Cas9 protein


target DNA
5'... CCCTAATCAGAGGCCAAACCCTTCCTGGAGCCTGTGATAAAAGCAACTGTTAGCTTG...3'
3'...GGGATTAGTCTCCGGTTTGGGAAGGTCCTCCAGGTGACACTATTTTAGCTTCAA...5'
BCL11A regulatory DNA
target DNA
5'... CCCTAATCAGAGGCCAAACCCTTCCTGGAGCCTGTGATAAAAGCAACTGTTAGCTTG...3'
3'...GGGATTAGTCTCCGGTTTGGGAAGGTCCTCCAGGTGACACTATTTTAGCTTCAA...5'
BCL11A regulatory DNA
5' UUUAUCACAGGCUCCAGGAA 3'
BCL11A guide RNA #1
5' 3'
BCL11A guide RNA #2