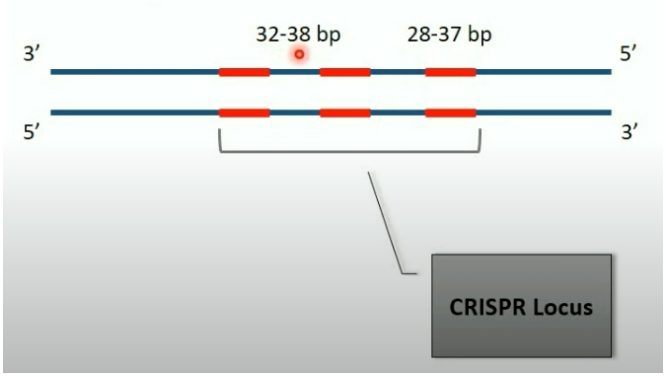
CRISPR/cas9 Gene Editing (April 16th 2021) by Dr Susan Shorter, University of Greenwich

The double helix of a DNA molecule consists of a long chain of paired nucleotide bases. There are only four bases called Adenine (A), Thymine (T), Guanine (G), and Cytosine (C), and in a DNA molecule A always pairs with T, and G with C. Thus if one of the two strands of DNA has a sequence AATCG the other strand will have TTAGC at the same location, these being known as complementary sequences. Conventionally for diagrammatic purposes Thymine is given the colour yellow, Adenine blue, Cytosine red and Guanine green.

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. These were first identified in bacterial cells as short bits of DNA about 28 to 37 base pairs long which read the same whether read left to right or right to left, and which occur at regular intervals along a strand of bacterial DNA. In between there is 'spacer DNA' consisting of 32-38 base pairs. Together these form a CRISPR array.

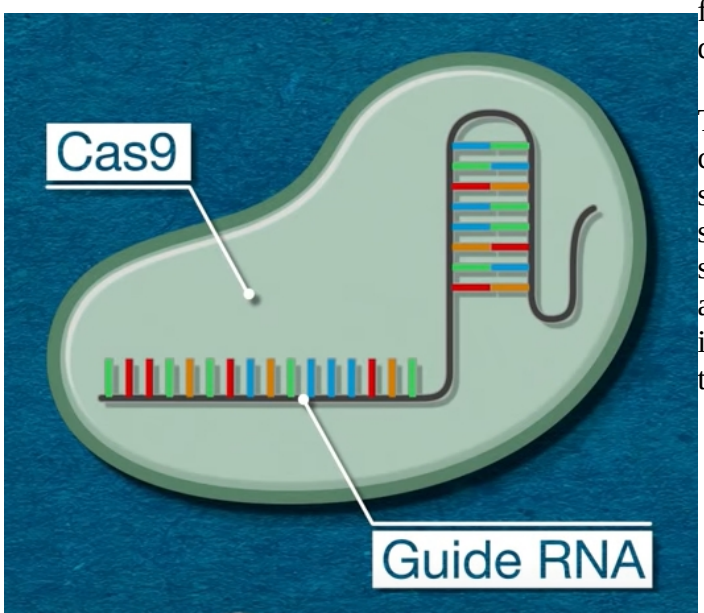


The CRISPR mechanism evolved in bacteria as an immune defence mechanism against bacteriophage viruses. Each bit of spacer DNA in the array is actually a stored part of the DNA of a different bacteriophage virus, so that if the bacteria is invaded by a previously encountered virus it can recognise the viral DNA sequence from the array and deploy a particular protein called cas9 to attach to the stored spacer which can then cut the viral DNA rendering it ineffective.

If a new virus invades the bacterium a snip of the new viral DNA is cut by two proteins called cas1 and cas2 acting in a complex and which act to add a new CRISPR/spacer sequence to the CRISPR array in the bacterial DNA. The cas1/cas 2 complex identifies what is known as a Protospacer Adjacent Motif (PAM) at the end of a viral DNA sequence which always contains NGG, where N is any of the four nucleotides followed by GG, and selects the sequence before the NGG. The PAM is a specific set of between 2 to 6 base pairs with NGG included. Over time a long sequence of bits of viruses that have been met is built up, stored in the array without the NGG sequence.

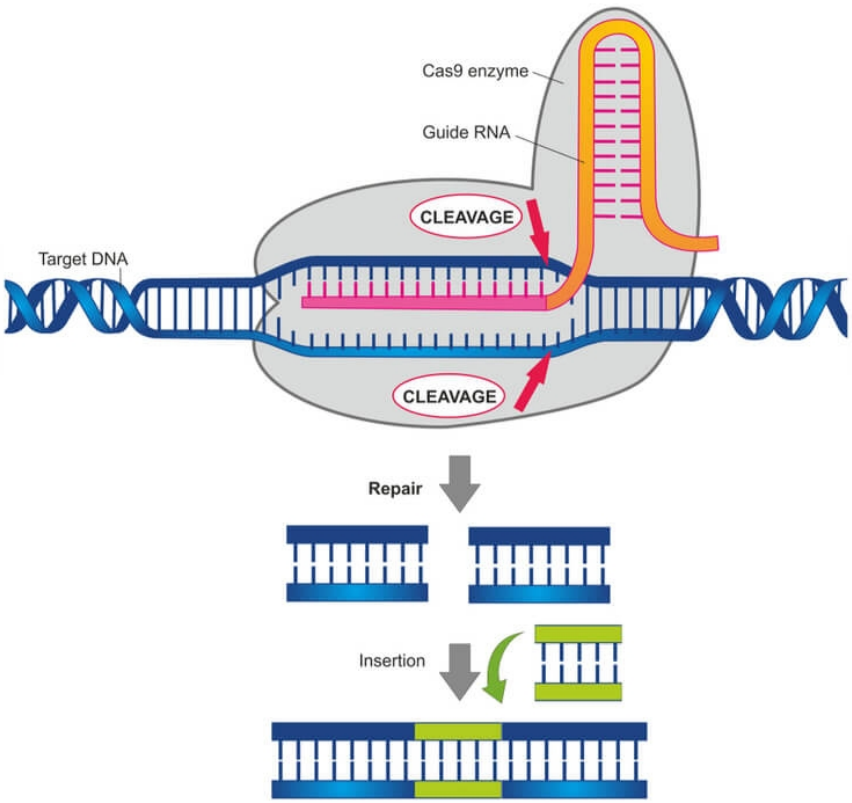
The PAM is what prevents the cas9 protein from attacking the bacteria's own stored bit of DNA in the spacer. The CRISPRcas9 complex attacks only the sequence of stored DNA if it is followed by the PAM.

How is this of use or potential benefit to humans? One idea is to edit specific genes in the human genome that are known to cause genetic diseases. If these genes can be isolated and edited to function as 'normal' genes then a genetic disease might be cured.



To this end CRISPR/cas9 consists of two components, the cas9 protein that as already stated can cut DNA, and a 'guide RNA' sequence that can recognise the particular sequence of DNA that is causing the problem and which it is desired to edit. This complex is introduced into cells and once inside it finds the target sequence, which is the complement

to the guide RNA, and cuts it at the specific location. The sequence matching the guide DNA – or part of it - can then be modified.



In this example the cas9 enzyme cuts the target DNA at the specific location identified by the guide RNA (cleavage) so that the target sequence (the complement to the DNA sequence shown in pink) can be repaired/edited, a new sequence inserted, or even deleted entirely.

Needless to say there are dangers associated with this technique. Editing a piece of DNA may result in unexpected side effects as well as perhaps creating new diseases. The technique also raises ethical issues and there remains the possibility of using the technique as a biological weapon.