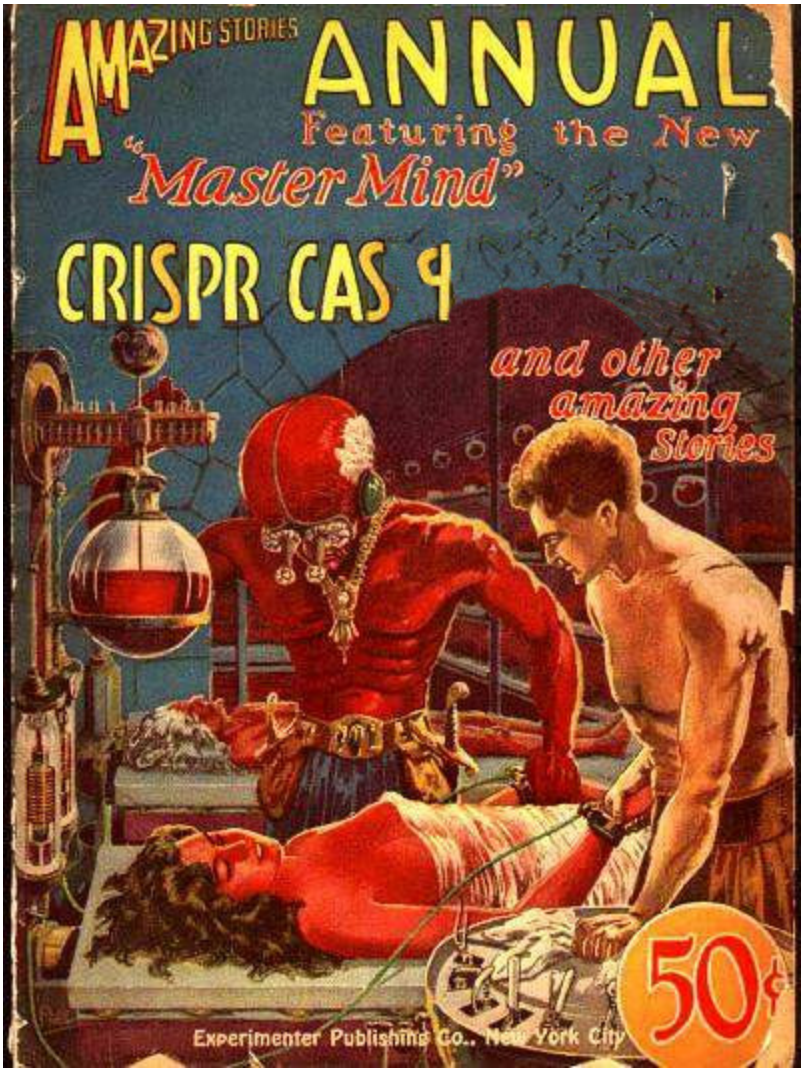
The Scinder

The structure behind the simplicity of CRISPR/Cas9

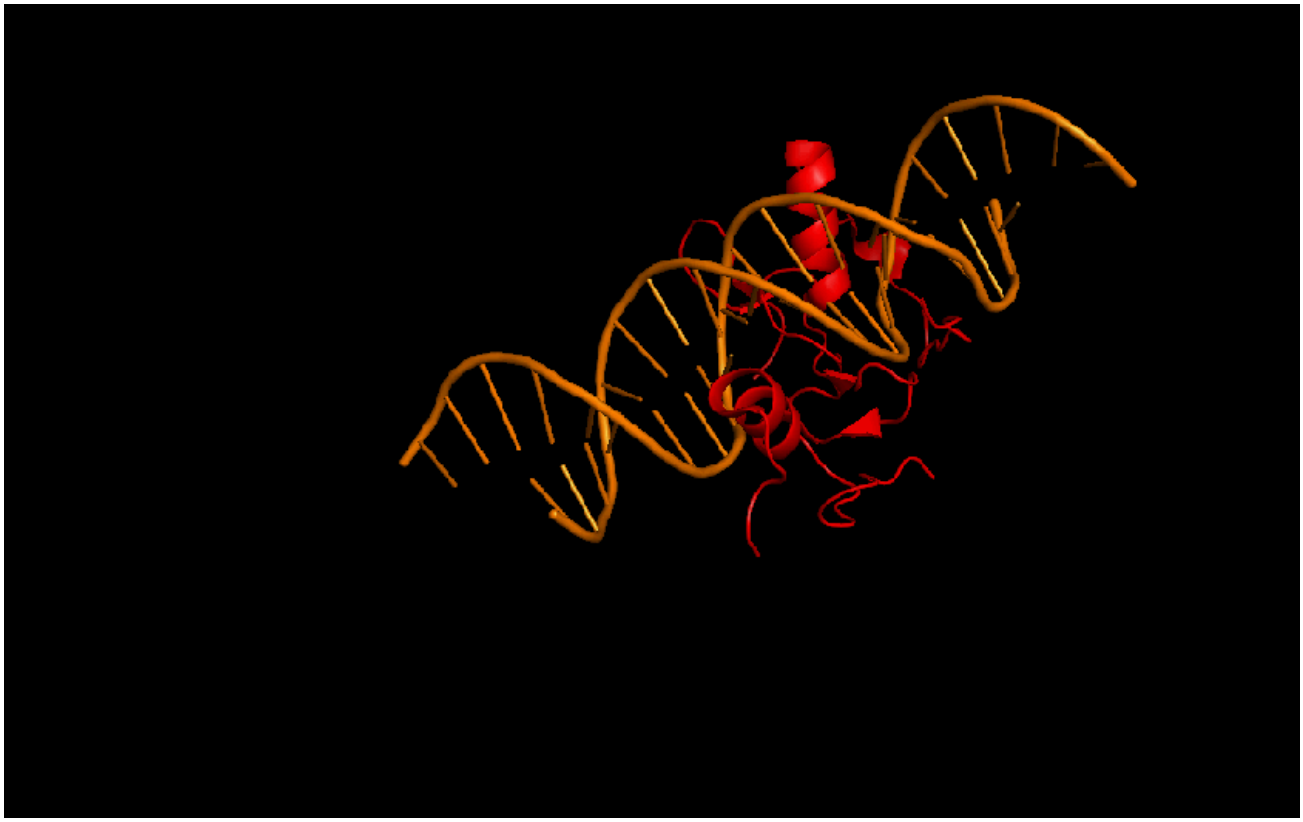


The International Summit on Human Gene Editing took place in Washington D.C. a few weeks ago, underlining the critical attention continuing to follow CRISPR/Cas9 and its applications to genome editing. Recently I compared published protocols for CRISPR/Cas9 and a competing technique based on Zn-finger nucleases. Comparing the protocols suggests editing with CRISPR/Cas9 is vaguely simpler than using Zn-fingers, but didn't discuss the biomolecular

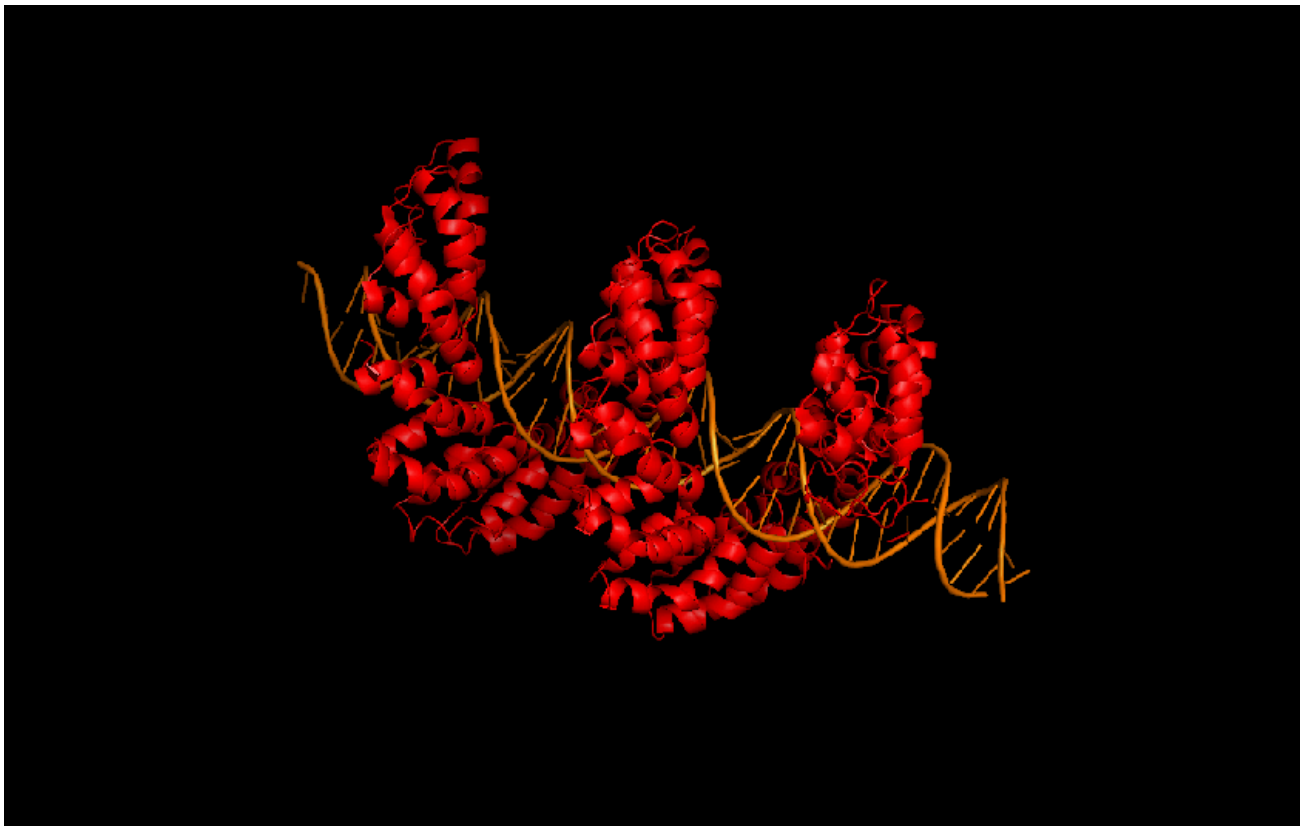
mechanisms underlying the increased ease of use. Here I'll illustrate the fundamental difference between genome editing with Cas9 in simple terms, using relevant protein structures from the Protein Data Bank.

Each of the techniques I'll mention here have the same end-goal: break double stranded DNA in a specific location. Once a DNA strand undergoes this type of damage, a cell's own repair mechanisms take over to put it back together. It is possible to introduce a replacement strand and encourage the cell to incorporate this DNA into the break, instead of the original sequence.

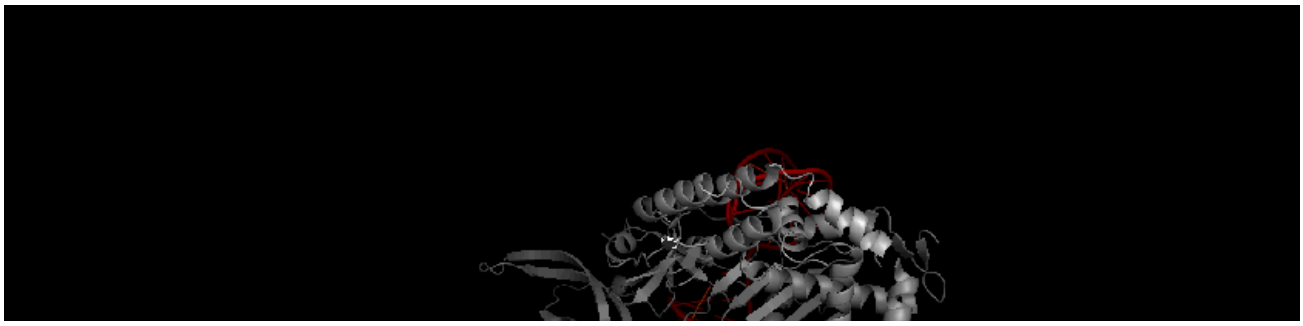
The only fundamental difference in the main techniques used for genome editing is the way they are targeted. Cas9, Zn-finger, and Transcription Activator Like (TAL) nucleases all aim to make a targeted break in DNA. Other challenges, such as getting the system into cells in the first place, are shared alike by all three systems.

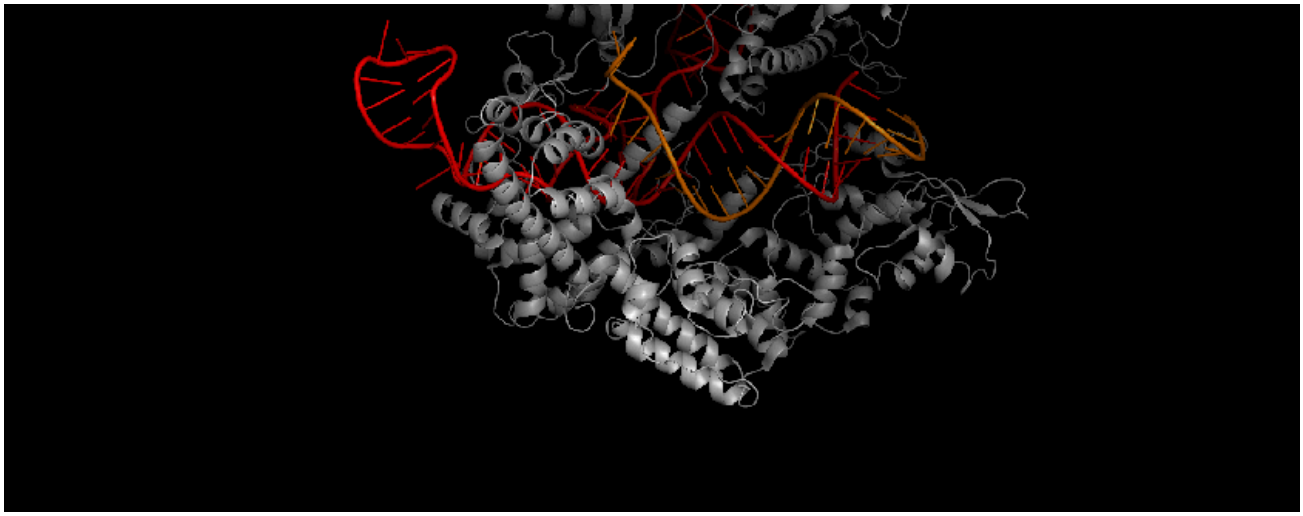


Zinc Fingers (red) bound to target DNA (orange). A sufficient number of fingers like these could be combined with a nuclease to specifically cut a target DNA sequence.



Transcription Activator Like (TAL) region bound to target DNA. Combined with a nuclease, TAL regions can also effect a break in a specific DNA location.





Cas9 protein (grey) with guide RNA (gRNA, red) and target DNA sequence (orange). The guide RNA is the component of this machine that does the targeting. This makes the guide RNA the only part that needs to be designed to target a specific sequence in an organism. The same Cas9 protein, combined with different gRNA strands, can target different locations on a genome.

Targeting a DNA sequence with an RNA sequence is simple. RNA and DNA are both chains of nucleotides, and the rules for binding are the same as for reading out or copying DNA: A binds with T, U binds with A, C binds with G, and G binds with C [1]. Targeting a DNA sequence with protein motifs is much more complicated. Unlike with nucleotide-nucleotide pairing, I can't fully explain how these residues are targeted, let alone in a single sentence. This has consequences in the initial design of the gRNA as well as the efficacy of the system and the overall success rate.

So the comparative ease-of-application stems from the differences in protein engineering vs. sequence design. Protein engineering is hard, but designing a gRNA sequence is easy.

How easy is it really?

Say that New Year's Eve is coming up, and we want to replace an under-functioning Acetaldehyde Dehydrogenase [2] with a functional version. First we would need a ~20 nucleotide sequence from the target DNA, like this one from just

upstream of the ALDH1B gene:

5′-AAC GAC ATG AGC ACA GCA GG -3′

You can write out the base-pairings by hand or use an online calculator to determine the complementary RNA sequence:

5′-AAC GAC ATG AGC ACA GCA GG-3′
3′-**UUG CUG UAC UCG UGU CGU CC**-5′

To associate the guide RNA to the Cas9 nuclease, the targeting sequence has to be combined with a scaffold RNA which the protein recognises.

Scaffold RNA:
5′-*GUU UUA GAG CUA GAA AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA ACU UGA AAA AGU GGC ACC GAG UGG UGC UUU UUU*-3′

Target Complement:
5′-**CCU GCU GUG CUC AUG UCG UU**-3′

Target complement + scaffold = guide RNA:
5′-**CCU GCU GUG CUC AUG UCG UU***G UUU UAG AGC UAG AAA UAG CAA GUU AAA AUA AGG CUA GUC CGU UAU CAA CUU GAA AAA GUG GCA CCG AGU GGU GCU UUU UU*-3′

With that sequence we could target the Cas9 nuclease to the acetaldehyde dehydrogenase (ALDH1B) gene, inducing a break and leaving it open to replacement. The scaffold sequence above turns back on itself at the end, sinking into the proper pocket in Cas9, while the target complement sequence coordinates the DNA target, bringing it close to the cutting parts of Cas9. If we introduce a fully functional version of the acetaldehyde dehydrogenase gene at the same time, then we surely deserve a toast as the target organism no longer suffers from an abnormal build-up of toxic acetaldehyde. Practical points remain to actually prepare the gRNA, make the Cas9 protein, and introduce the replacement sequence, but from an informatic design point of view that is, indeed, the gist.

That's the basics of targeting Cas9 in 1,063 words. I invite you to try and explain the intricacies of TAL effector nuclease protein engineering with fewer words.

Notes:

[1] That's C for cytosine, G for guanine, U for uracil, and A for adenine. In DNA, the uracil is replace with thymine (T).

[2] Acetaldehyde is an intermediate produced during alcohol metabolism, thought to be largely responsible for hangovers. A mutation in one or both copies of the gene can lead to the so-called "Asian Flush".

Sources for structures:

I rendered all of the structures using PyMol. The data come from the following publications:

PDB structure: 3VEK (Zn-finger)

Wilkinson-White, L.E., Ripin, N., Jacques, D.A., Guss, J.M., Matthews, J.M. DNA recognition by GATA1 double finger.To Be Published

PDB structure: 3ugm (TAL)

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., Stoddard, B.L. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. (2012) Science 335: 716-719

PDB structure: 4008 (Cas9)
Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Crystal structure of Cas9 in complex with guide RNA and target DNA. (2014) Cell(Cambridge,Mass.) 156: 935-949

Comic cover original source:

"Amazing Stories Annual 1927" by Frank R. Paul – Scanned cover of pulp magazine. Licensed under Public Domain via Wikimedia Commons – https://commons.wikimedia.org/wiki/File:Amazing_Stories_Annual_1927.jpg#/media/File:Amazing_Stories_Annual_1927.jpg

theScinder / December 23, 2015 / Video / Essays / 3ugm, 3vek, 4oo8, Cas9, CRISPR, CRISPR/Cas9, DNA, Genome editing, human gene editing, Nuclease, PDB, protein structures, RNA, science, TAL, TALEN, Zn-Finger, Zn-finger nuclease

5 thoughts on "The structure behind the simplicity of CRISPR/Cas9"

Vanan

March 6, 2016 at 09:04

Hey 'The Scinder'! I really liked your example on the Acetaldehyde Dehydrogenase gene replacement. I was wondering if I could cite this example in a review paper I am writing on CRISPR/Cas9. If thats okay with you can I have more details so that I can fully credit you? Thanks!

★ Like

theScinder

March 17, 2020 at 23:42

Hi, sorry that I never replied to your comment until now. I hope my silence didn't deter you from using the ADH example or coming up with something even better, and I'd be interested to know how your review paper turned out.

★ Like

Zee

November 2, 2018 at 04:29

Hello The Scinder! Thank you! This is really helpful. I am wondering how did you build nucleic acid sequences in PyMol? Do you have an article or a protocol on how to build protein-nucleic acid structures? I have to do something similar for one of my assignments. I appreciate your help.

★ Like

theScinder

March 17, 2020 at 23:39

Sorry this is too late for your assignment, but perhaps the info will be useful for others.

I find structures of interest using the RCSB protein databank at
https://www.rcsb.org/
These files include the amino acid sequence, but if you want the sequence of the mRNA transcript/cDNA you'll want to search pubmed e.g.
https://www.ncbi.nlm.nih.gov/nuccore/?term=nsp1+viral
Pubmed will usually be able to find the gene and amino acid sequence together

for a given protein if they are available.
If all you have is a sequence and you want to make some educated guesses about the structure, a good place to start is SWISS-MODEL (https://swissmodel.expasy.org/) where you can use your sequence's homology to sequences with known structures (or structural elements) to find motifs that may be present. The workflow is pretty well automated but they have some documentation and tutorials to get started (https://swissmodel.expasy.org/docs/examples)

★ Like

Anonymous

March 17, 2020 at 23:44

Test comment ;)

★ Like