In Vivo Detection and Signaling of Arbitrary DNA Sequences

Building living medical machines and making microorganisms work for us

Brooks J Rady

Project Goals

The long-term goal of this project is to create a synthetic microorganism capable of breaking down a blood sample and scanning it for known pathogens. An organism like this would be a natural continuation of the trend towards POC* technologies in the field of disease diagnosis.

POC technology has made it possible to move diagnosis out of the lab and into areas of the world that don't have easy access to modern medical equipment and trained personnel. Today, examples of POC technology can be found everywhere (the most idiomatic example is the over-the-counter pregnancy test – a LFIA* that is designed to detect the presence of HCG in urine). LFIA technology, however, is a very blunt tool. It can be invaluable in simple diagnoses – like the pregnancy test example – but more complex assays may require sample preparation by a technician beforehand and, as sensitivity increases, it becomes much more difficult to reproduce results in the presence of changing environmental conditions[1].

In an attempt to overcome these limitations, the medical community is now turning towards microfluidics and lab-on-a-chip technologies which allow for the increased automation and complexity of biological assays[1]. While these devices are becoming exponentially more capable and inexpensive to manufacture, getting these devices to the developing world still requires extensive organization and can cost anywhere from $4,000 to $7,000[2,3].

By engineering an organism to carry out this process *in vivo*, we can save significantly on both cost and hassle. The test doesn't need to be ordered or shipped when supply runs low; rather, because it is a living cell culture, it can just be grown up when more of it is needed. Anywhere that yeast can grow, this organism can provide valuable medical insight to countless people.

As a short-term goal, I've set out to create a proof of concept for the DNA detection portion of the system in *E. Coli*. The goal is to engineer a strain of the bacteria that will change color from tan to red when a particular sequence of DNA is detected inside the cell.

Project Design

In this proof-of-concept experiment, I will be transforming two plasmids into *E. Coli*. The first plasmid, pdCas9-Det, is a modified version of the pdCas9 plasmid and contains all of the genetic circuitry necessary for the detection of an arbitrary DNA sequence. The pdCas9 plasmid on its own expresses the mutant dCas9 for binding the DNA in addition to the tracrRNA* and crRNA* sequences that guide the dCas9 to its target. My version, the pdCas9-Det plasmid, has an additional operon that contains the coding sequence for RFP* as well as a second, different crRNA sequence. The expression of this operon is regulated by LacO*. In order for the LacO repression system to work correctly, I also added the coding sequence for LacI* under a constitutive promoter on the same plasmid.

What makes this system work are the two crRNA sequences that it contains. The first crRNA sequence, which is continually expressed under normal conditions, targets the promoter of the RFP operon. The trick here is that the crRNA does not perfectly match the sequence of the lac promoter. It does, however, perfectly match the sequence of the target DNA. This means that, in the absence of the target DNA, the dCas9 will repress the signaling gene. If, however, the target DNA is introduced, it will compete the dCas9 off of the promoter allowing the RFP gene to be expressed.

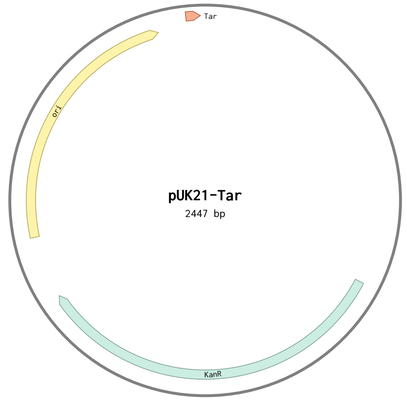
When the signaling gene is activated, in addition to the RFP protein, a second crRNA is expressed. This second crRNA sequence causes the dCas9 to target the expression site of the first crRNA sequence. This is designed to produce a runaway effect – ensuring that, even after all of the target DNA has been used up, the cell will continue to produce RFP. By repressing the first crRNA sequence, we have taken away the information that dCas9 needs to repress the signaling gene.

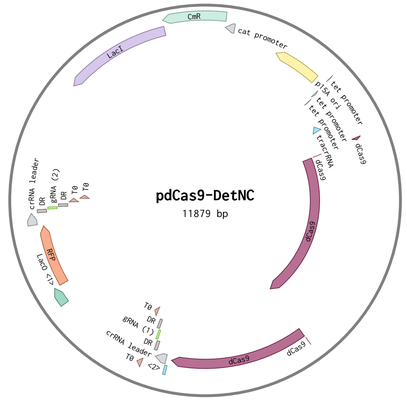
The issue with this system is that, in its current form, there is no way to guarantee that the first crRNA sequence will be expressed first and the second sequence expressed second. If the second sequence happens to be expressed first, then the cell will turn red long before it has the chance to detect any DNA. What we need is a way to ensure that the equilibrium is established in favor of the first sequence so that the cell doesn't change colors prematurely and so that the dCas9 is properly bound to the promoter of the signaling gene.

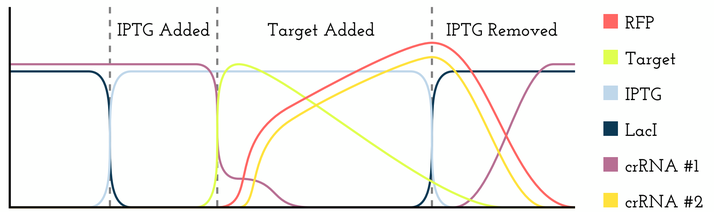
We can control this equilibrium via a secondary repressor. This second repressor, LacI, also represses the signaling gene but unlike dCas9 it is not controlled by crRNA sequences; rather, the LacI will stay bound to the promoter until an Allolactose analog, like IPTG*, is introduced. The cells are first grown up in the absence of IPTG so that the equilibrium is established in favor of the first crRNA and, when the cells are ready to receive the target DNA, the system is armed via the introduction of IPTG. This deactivates the LacI and leaves only the imperfectly-matched and loosely bound dCas9 repressing the RFP.

Finally, I will introduce the target DNA via a second, high copy-number plasmid: pUK21-Tar. This plasmid will duplicate itself hundreds of times and should compete the dCas9 off of the signaling gene, turning the cell red.

In my experiment I will be testing 15 different crRNA sequences containing a variety of mismatches with the promoter sequence. I am looking for a sequence that is easily competed off, but is also strong enough to repress the gene when the target is not present.







Background

CRISPR-Cas9* is an adaptive immune system, originally discovered in bacteria, that has been widely adopted and repurposed by scientists as a flexible gene targeting tool. Cas9*, the protein that carries out the gene targeting work, can be programmed using a standardized tracrRNA* and a target-specific crRNA*. The crRNA sequence is complementary to (matches up with) the sequence of the target DNA and has a separate section that binds to the tracrRNA. The tracrRNA then, in turn, binds to the Cas9 protein; effectively linking the crRNA to the Cas9. The Cas9 protein then looks through the DNA in the cell for a match to the crRNA sequence that is adjacent to a PAM* sequence (usually NGG). The PAM sequence check is necessary for the CRISPR-Cas9 system to avoid binding to itself. If it were not required, then the Cas9 would simply bind to the DNA where the crRNA was originally created and not to the target (as intended). Once the Cas9 has located a match to the crRNA sequence that is adjacent to a PAM, it binds to the DNA and creates a double stranded break – cutting the DNA stand in two. What makes CRISPR-Cas9 such a powerful tool is its programmability. By changing the crRNA sequence, scientists can direct the Cas9 protein anywhere in the genome. Think of Cas9 as a bounty hunter that will track down whatever sequence it is told to target. Not all variants of Cas9 create double stranded breaks in the sequence; for example, "dead" Cas9 or dCas9 is a mutant that will bind to the target site, but won't cut the DNA[4].

Another important concept to understand is that of competition and affinity. These related concepts describe the behavior of systems that involve the bonding of two or more species with a shared and finite substrate. Imagine you are at a table with four other friends – five of you in all. Each of you has a die to roll. You decide to play a game with the dice. The rules are, if you roll anything other than a six, you let the die sit for one second and then re-roll the die. If you roll a six, you let the die sit for 10 seconds before re-rolling. You and all of your friends start rolling your dice at the same time and play by these rules for a minute or two. Then a sixth person walks in the room and looks at the table. What number is he or she most likely to see on each of the five dice on the table? The answer is six. Even though you are far more likely to roll a one, two, three, four, or five, the six has a higher affinity (it stays on the table for longer; 10 seconds as opposed to one) so over time the substrate (the table of dice) tends to contain more sixes than any other number. In biology, molecules are constantly making and breaking bonds in a random fashion. Some materials, however, have a higher affinity for the substrate and bond more strongly to it than others (they stay bonded for longer). This means that after an amount of time, just like the dice, the substrate will be bonded to far more of the high-affinity material than to the low-affinity materials. When this happens, the higher-affinity material is said to have competed the lower-affinity material off of the substrate. This is relevant because Cas9 does not always perfectly match the crRNA sequence when it binds to the DNA. Sometimes there are small mismatches. When there are, however, the affinity of the bond is lowered. This makes the imperfectly-matched Cas9 subject to competition. If a perfectly matching sequence comes along, the Cas9 will abandon the imperfect match for the perfect one, leaving the imperfect match naked without a Cas9 bound to it.

The last major concept that plays a role in my project has to do with how genes are expressed and how proteins are actually made in the cell. Gene expression occurs in two stages: transcription and translation. Transcription is the step during which the DNA is read and an RNA copy is produced. Translation, on the other hand, is the step in which the RNA is actually turned into protein. The step relevant to gene regulation is transcription: the amount of protein produced via the expression of a given gene is proportional to the number of RNA transcripts present in the cell. On a microscopic level, the RNA is formed by a protein called RNAP* which slides along the double helix of DNA like a zipper. Genes can be turned on and off by blocking this zipper with another protein. If another protein is bound to the DNA, it acts as a roadblock and doesn't allow the RNAP to pass[5]. In my case, without anything bound to the DNA, RFP* is expressed and the cell turns red. However, when either LacI* or dCas9 are bound to this gene, the zipper is blocked and no color change takes place.

Glossary

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats. An adaptive immune system discovered in bacteria that has been widely adopted and repurposed for genome engineering.
Cas9 – CRISPR Associated Protein #9. A protein belonging to Type II CRISPR systems that forms a complex with a tracrRNA and crRNA and cuts DNA where matches to the crRNA sequence are found.
tracrRNA – Trans-activating CRISPR RNA. Binds the crRNA and feeds it to the Cas9 protein.
crRNA – CRISPR RNA. An approximately 20bp sequence that instructs Cas9 to target a complementary sequence.
PAM – Protospacer adjacent motif. A short (usually 3bp) long sequence that must follow the targeted sequence for Cas9 to bind and cut the DNA.
LacO – Lac Operator. A section of DNA that binds to LacI.
LacI – Lac Repressor. The protein that binds to LacO and prevents transcription.
RNAP – RNA Polymerase. A protein capable of reading DNA and creating an RNA copy. This is the protein directly responsible for the process of transcription.
RFP – Red Fluorescent Protein. A red colored protein that fluoresces in UV light.
POC – Point-of-care. Medical care and diagnosis that is on-site and doesn't require the transport of samples or patients to obtain results.
HCG – Human Chorionic Gonadotropin. A hormone produced by the body when a fertilized egg implants in the uterine wall. The presence of this hormone indicates the beginning of pregnancy.
LFIA – Lateral Flow ImmunoAssay. A POC assay in which a fluid sample travels across a device and binds to antibodies that indicate the presence of a particular substance within the sample.
IPTG – Isopropyl β-D-1-thiogalactopyranoside. Used to induce genes repressed by LacI.

Conclusion & Future Work

So far this project has just been an exercise in the theoretical design of genetic circuits and the *in silico* design of plasmids. While my system is built upon solid research and scientific principle, it is yet to be proven. Implementation is already underway, but the bulk of the work so far has been in figuring out how to assemble these plasmids and in obtaining the genetic material needed to do so. Currently I'm on track to finish the first round of testing this upcoming summer.

As mentioned previously, this project is a proof-of-concept for a much larger body of work that will enable genetically modified microbes to diagnose diseases and perform basic DNA testing without the need for expensive machinery. Lots of work still needs to be done to improve the sensitivity of the assay (the competition mechanism currently relies on an impractically large amount of the target DNA being present) and the sample preparation *in vivo* will be a project on its own. That being said, this experience has taught me a lot about moving from idea to implementation and I've come away from it swimming in new biological knowledge and laboratory skills. I plan to continue work on this project throughout my time at the University of Sheffield and plan on applying my new-found knowledge to my studies in Bioengineering.

References

1. Sharma S, Zapatero-Rodríguez J, Estrela P, O'Kennedy R. Point-of-Care Diagnostics in Low Resource Settings: Present Status and Future Role of Microfluidics. Newman JD, ed. Biosensors. 2015;5(3):577-601. doi: 10.3390/bios5030577.

2. Gatrad AR, Gatrad S, Gatrad A. Equipment donation to developing countries. Anaesthesia. 2007;62(s1): 90-95. doi:10.1111/j.1365-2044.2007.05309.x.

3. HealthTrust. Giving Medical Supplies to the Developing World-reSOURCEs. Performance Improvement For Healthcare. https://healthtrustpg.com/healthcare-facility-management/supplying-the-developing-world/. Published March 6, 2017. Accessed April 4, 2018.

4. CRISPR Guide. Addgene. https://www.addgene.org/crispr/guide/. Accessed April 5, 2018.

5. Overview of transcription. Khan Academy. https://www.khanacademy.org/science/biology/gene-expression-central-dogma/transcription-of-dna-into-rna/a/overview-of-transcription. Accessed April 5, 2018.