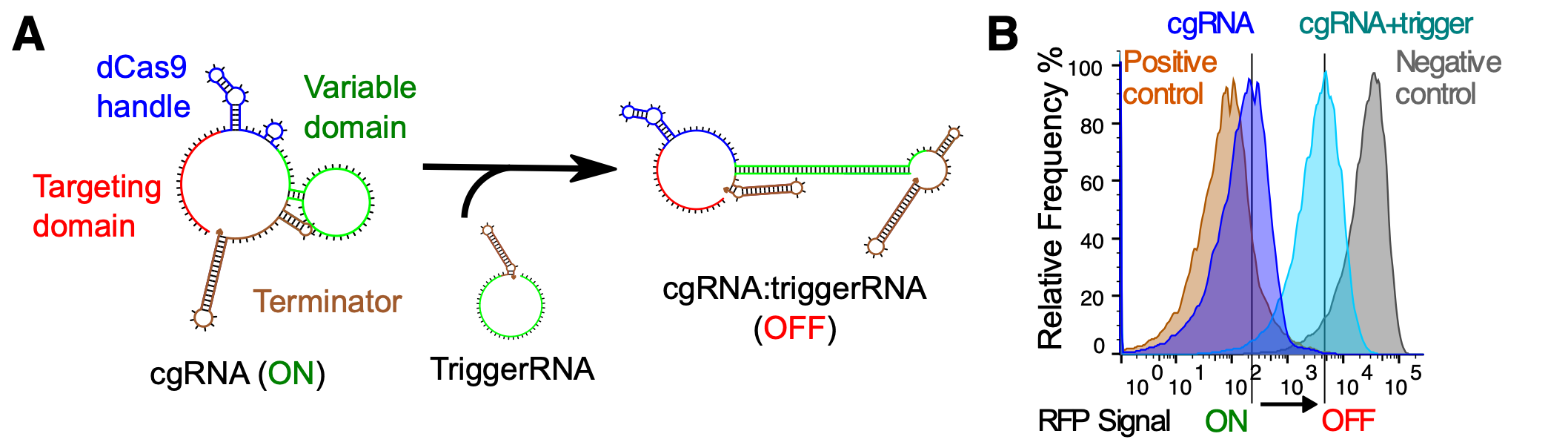
**Introduction**

For my PhD, I recently co-invented a novel biomolecular device dubbed the conditional guide RNA (cgRNA) [1]. cgRNAs are important as they could enable detection of endogenous mRNA and form the basis for large scale control circuits and digital logic inside of a cell. The cgRNA works by undergoing a programmed change in RNA structure when bound to its cognate trigger RNA (see Figure 1). This switches the cgRNA between active and inactive structures (whereby it no longer acts as a good co-factor for Cas9). The guide RNA is important for Cas9 activity because it contain the targeting domain, which specifies where in the chromosome Cas9 should bind.



**Figure 1**: A) Schematic and B) FACS data on cgRNA activity in bacteria

**Motivation**

The design motif was discovered through biochemical screening. Based on previous literature and my own mutational studies, I made a lucky guess and hit upon a previous unknown structure function. The motif above works by conditionally breaking 4 base pairs in the variable domain. Unfortunately, why this should work is still a mystery. This answer is important because not all RNA designs generated by NUPACK, our lab software, are good quality. Although RNA structure defect could be low, the cgRNA activity could be very poor. The interaction with protein is an important missing piece of this puzzle.

**Methodology**

For the class project, I want to validate some of my hypothesis’ on the cgRNA/Cas9 interaction through molecular dynamic (MD) simulations. Although there is ample literature on Cas9 catalytic activity (DNA cutting) [2, 3], there are few MD studies on why gRNAs should have specific affinity for their cognate Cas9 proteins.

My approach would be the following:

1. Find the energies of Cas9/cgRNA variants in their relaxed conformations and see if there is any correlation with biochemical activity
2. See if Cas9:cgRNA:triggerRNA complex may form some plausible inactive transition state (I can explain more with the pdb files)
3. Measure the force of interaction between different nucleotides and amino acid residues. This information could be used to deduce what residues are important for RNA binding. If trained with a generative machine learning model, I would like to be able to distinguish what are good designs or bad designs before I clone them. Taking a step further, I would like the ML model to suggest me better design motifs.

Some complications I anticipate:

1. Cas9 is large and requires some niche MD code (gaussian-accelerated MD, see ref 2) to generate long trajectories (micro to milliseconds) needed to see meaningful conformational change. At the lab level, this is feasible, but it may be too computationally expensive for a 5-week class project.
2. A Cas9:cgRNA:triggerRNA complex might not exist at all. I am still in the process of validating this biochemically. This would make the problem an RNA/protein docking simulation, which could be too computationally expensive to be feasible.

I am open to suggestions and improvements to my proposed approach. Looking to hearing from you guys.

References

1. <https://www.biorxiv.org/content/early/2019/01/21/525857>

2. <https://doi.org/10.1073/pnas.1707645114>

3. <https://www.biorxiv.org/content/10.1101/421537v1>