Note: There will now be ~5 assignments throughout the term. Each assignment is worth some proportional fraction of your final grade. Your lowest assignment grade will still be dropped providing a total for the "assignment component" of your final grade as 56%.

Note: Please adhere to the stated expectations: "I expect that you'll interact with your colleagues throughout the course, in discussing any readings, during class, or in considering any assignments. However, your submitted assignments should reflect and be of your own individual work."

Note: When searching online you shouldn't need to request custom pricing or quote information. That is, for questions requiring online work, only use information that is readily available online.

Note: These assignments are both due by 9a Monday March 2nd and will be graded and ready for pickup by office hours that day, so that you can review any materials before the March 5 midterm. Please submit your assignment by email to: endy@stanford.edu

Note: Please send any questions about this assignment by email to endy@stanford.edu. I'll try to respond in real time.

Assignment #2 DNA Engineering

Part 1. Using the standard single letter amino acid abbreviations, what message does the following sequence of DNA encode? (Hint: search online for a DNA translation tool).

CGTGAAGCTCTTCATGCTTGTAAAGAACGTTCTCCTCGTGGTCGTGCTATGGATAATGCT

R E A L H A C K E R S P R G R A M D N A (note the missing "O")

Part 2. Choose any two English words or phrases that you like. Encode both words in a single DNA sequence using the fewest number of DNA bases possible. Each individual word or individual phrase must be at least 7 letters in length. A +5% assignment grade bonus (and bragging rights) will be given to most efficient encoding (i.e., greatest fractional reduction in number of DNA bases used, relative to the total length of the two words or phrases). (Hint: find online or make a good reverse translation tool).

Many different approaches to answering this problem. Some solutions are trivial. For example, if you had two words that overlapped in their letters, such as "photosynthetic" and "synthetic" then you could encode both of these words in a single strand of DNA encoding "photosynthetic" (the DNA sequence would be: CCTCATACTTCTTATAATACTCATGAAACTATTTGT, although note that the letter 'o' is missing here). Here is a neat site that you can use to find words that contain other words (http://www.morewords.com/contains/words/). A trickier (and more biological) puzzle is to encode multiple words in the same DNA sequence, but using different coding frames. Recall that each double stranded sequence of DNA has six coding frames, or reading frames (where a coding/reading frame is defined by starting at any one of the three grouping of triplets, along a sequence of DNA -- there are three reading frames in each direction). For example, GUC codes for Valine (V). The complementary strand of DNA would be CAG. Read backwards in the same frame (GAC) would produce Aspartic Acid (D). You'd likely want to develop a tool to help you do this!

Assignment #3 RNA Engineering

Part 1. You have been hired by Microbesoft, Inc. to develop a RNA-based control system for conditionally regulating the expression of a protein-encoded enzyme. The specific application requires that as little protein as possible be expressed in the absence of the controlling input signal, and that as few cellular resources are wasted as possible. You can choose be-

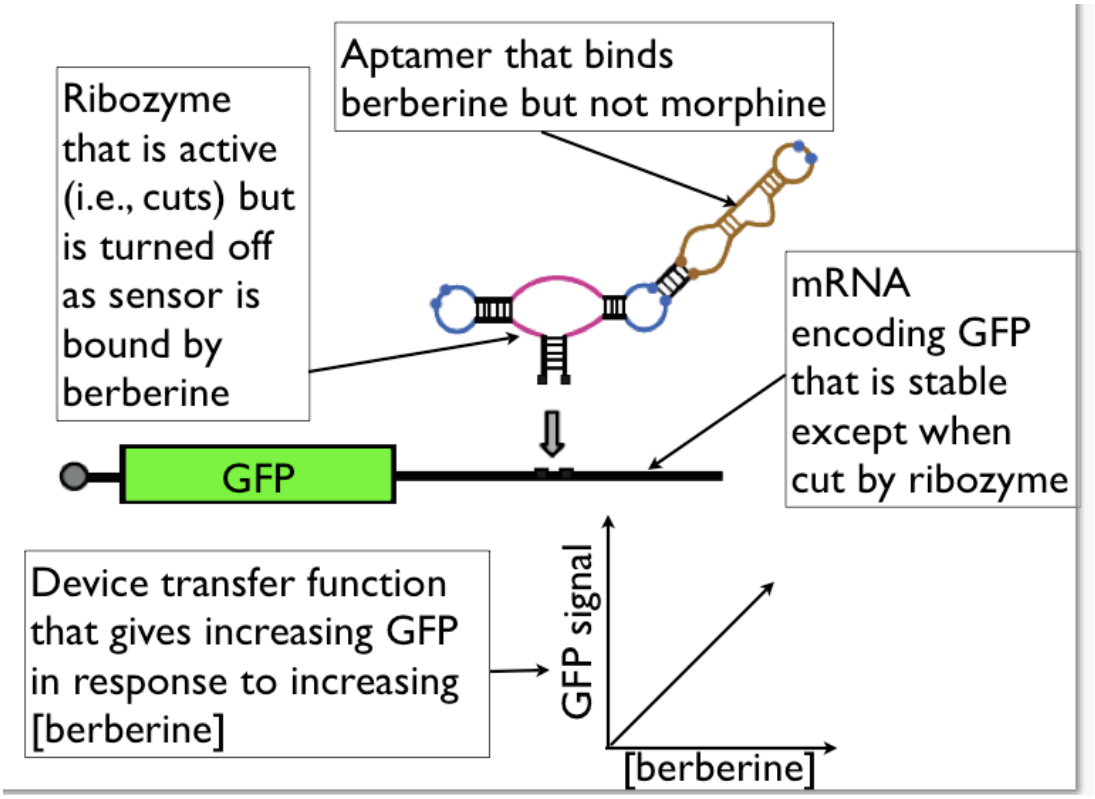
tween (i) a 5'-based RNA element that non-destructively (reversibly) interacts with and occludes ribosome loading on a ribosome binding site, or (ii) a ribozyme-based riboswitch that is integrated into the 3' untranslated region (UTR) and that irreversibly cleaves the transcript in the absence of an input signal, resulting in rapid mRNA degradation. Which control mechanism do you choose? Why (one paragraph or less)?

You could develop an argument for either case. On the one hand, the 5'-based control element might keep the protein from being synthesized at all, which would result in less waste of cellular resources. Also, this controller is non-destructive, which means that you could switch the translation of the RNA "on" or "off" without having to require the cell to remake the mRNA via transcription all the time. However, the 5'-encoded switch might not be able to completely shutdown translation, and could lead to leaky expression. By contrast, the 3'-encoded controller is destructive. So, this might cost the cell more in terms of energy and material resources, but if the mRNA cutting and degradation activities were very high, you could likely produce a very low background signal (i.e., level of protein expression in the absence of the controlling inducer -- or, "off" really means "OFF").

Part 2. BioBricks & Mortar, Inc. would like you to help them debug a metabolic pathway that they have been engineering by developing a new riboswitch that is responsive to the intracellular levels of berberine (http://en.wikipedia.org/wiki/Berberine). Their engineers tell you that they have a hunch that their current version of the pathway might be producing morphine instead (http://en.wikipedia.org/wiki/Morphine). Start by outlining the evolutionary selections that you could use to select an RNA aptamer that can serve as an input sensor for berberine but not morphine. Second, sketch out the component-level architecture for the resulting riboswitch and its integration into a DNA-encoded mRNA transcript that will result in berberine levels being converted to a green fluorescent protein signal.

The trick here is to design a selection that will find RNA apatamers that bind to berberine, but not to morphine. You'd likely want to start by making a population of random RNA molecules. Then, you could select for the subset of individual RNA molecules in your random population that bind to berberine. One approach would be to tether berberine to a surface (e.g., beads), and then flow the RNA aptamers across the beads. Throw away anything that doesn't stick, then elute the tight binders, in order to enrich for berberine-binding aptamers. Now, how to find the subset of berberine-binding aptamers that do *not* bind morphine? Take the berberine binders, and flow them across beads to which morphine has been linked. In this case, only keep the berberine-binding RNA aptamers that do not stick to the morphine-coated beads. In other words, a first selection *for* binding to berberine, followed by a second selection *against* binding to morphine. In theory you could perform these selections in either order (i.e., alternatively, first select for aptamers that don't bind morphine, followed by a selection for binding to berberine).

Next, you'd want to create an "ON" switch that converts increasing berberine levels into increase GFP signal. Use a riboswitch that cuts and destabilizes the GFP-encoding mRNA in the absence of berberine, but whose catalytic core is inactivated as berberine binds its cognate aptamer and disrupts the catalytic core of the ribozyme. See schema on next page.


Aptamer that binds berberine but not morphine
Ribozyme that is active (i.e., cuts) but is turned off as sensor is bound by berberine
mRNA encoding GFP that is stable except when cut by ribozyme
GFP
Device transfer function that gives increasing GFP in response to increasing [berberine]
GFP signal
[berberine]