Lecture and Dialog on Synthetic Biology

Drew Endy (endy@stanford.edu)

Goals for Today

A. Review impact of decoupling design of DNA, from construction of DNA

B. Introduce assembly of "long length" DNA from "gene" length DNA.

C. On the importance of decoupling design & construction (e.g., short or long).

D. Discuss public access to pathogen sequences

**Announcements **

i. Maureen Hillenmeyer will be serving as our course TA. Maureen just finished her PhD at Stanford and is starting a postdoc with Ron Davis at the Stanford Genome Technology Center. Maureen's email address is maureenh@stanford.edu. We are very luck to have her working with us!

ii. We will be establishing weekly office hours. Please send me (endy@stanford.edu) an email today with your top three choices (I'd be grateful if you could provide three blocks of time, so that there's a better chance of overlap with more people).

iii. We are looking into arranging two field trips. One possible field trip would be to DNA2.0, a gene synthesis company in Menlo Park. A second possible field trip would be to Amyris Biotechnology / JBEI a biofuels and therapeutics company in Emeryville. Each field trip would likely take 4-5 hours. Please look at your calendar for the rest of the term and, if you would like to attend either possible field trip, send us three good dates, noting whether AM or PM would be best.

Today

0. (From last time). We can synthesize or assemble DNA, from oligonucleotides, to genes, to genomes. In relative terms, the maturity of DNA fabrication technology decreases with the length of DNA to be synthesized (i.e., oligo synthesis is the most advanced, genome synthesis is the least advanced). However, all areas of synthesis and assembly could be made better.

1. As an aside, DNA synthesis builds on top of existing technologies for manipulating or editing DNA. These earlier technologies are based on restriction enzymes (for cutting DNA) and ligases (for pasting DNA) and polymerase chain reaction (for amplifying and making small numbers of changes to DNA). DNA synthesis doesn't likely replace these earlier technology, but rather builds upon them, expanding the number of people working with DNA, and the scope of possible projects.

A. Review Impact of Decoupling

2. There seem to be at least two big impacts to be derived from getting better at building DNA, thereby allowing for a decoupling of DNA *design* from DNA *fabrication*. (Note: "decoupling" is being used here to represent a separation of process and expertise. As a parallel example of a decoupled process, an architect might design a building, whereas a builder will construct the building. Both the architect and builder can become expert in their work, which can lead to significant advantages and improvements in the overall process of putting up a new building. Of course, the architect and builder also need to learn how to communicate effectively!).

3. The first impact of decoupling is that the designs of natural biological systems may no longer need to be subject to the natural constraints of direct descent and replication with error. I.e., for very simple biological systems, we can now begin to implement systems at the molecular and cellular level that did not themselves derive from any ancestor. This would seem to allow for the design, construction, and testing of disposable biological systems (i.e., living organisms that, as individuals, don't themselves need to be able to reproduce or evolve).

4. The second impact is that we (the scientific and engineering communities) are likely to be struggling / scrambling to figure out what we want to build. By analogy, if you the world's best word processor, but no languages or grammars, what would you say?

5. Note, there will likely be significant secondary impacts to be derived from decoupling, some of which we have already mentioned (e.g,. security issues associated with possible new routes for accessing pathogen genomes) and others still (e.g., what legal framework will support an "iTunes" equivalent for DNA parts?).

6. So, what are we going to say (i.e., write to DNA)? Let's begin by talking about DNA engineering. My notes here are organized into three types of DNA engineering ("function," "architecture," and "fun").

B. Engineering DNA for Function

7. Some biochemical functions are directly encoded by a DNA sequence (i.e., the DNA sequence itself is the primary determinant of biochemical activity, as distinguished from the DNA coding sequence of a protein, which is read out and eventually translated into a protein, but that by itself does not have any biochemical activity). A good example of such a function are "promoters," which initiate transcription (i.e., the production of RNA).

8. Here's the DNA sequence of a classic promoter:

    a. TAATACGACTCACTATAGGGAGA

b. This 23-base pair sequence of DNA is the consensus (i.e., typical, or strongest) promoter for an enzyme known as the T7 RNA polymerase (T7 pol). Only one sequence of the DNA is shown, but the T7 pol binds double stranded DNA (dsDNA) and initiates transcription.

c. You can change the activity of the promoter by changing the sequence!

d. For example, changing the first CGA in the promoter sequence to ACC will create a new promoter (TAATA**ACC**ACTCACTATAGGGAGA) that is recognized by the T3 RNA polymerase (a related but different RNA polymerase). You can read more about this specific hack here: "Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity," by Curtis A. Raskin, George Diaz, Keith Joho, and William T. McAllister, Journal of Molecular Biology Volume 228, Issue 2, 20 November 1992, Pages 506-515. Also, see the image on the first slide from today to see how the polymerase and promoter are though to interact. See if you can find the CGA in the promoter sequence that would be changed to an ACC.

e. Related changes can be made that change the strength of the promoter, so that differing amounts of transcription starts. However, note that the knowing what changes to make isn't a simple thing to figure out (i.e., required significant experimentation).

f. What would it take to get better at using models to directly engineer DNA sequences? Could we ever hope to do without experimentation as a first step?

C. Engineering DNA for Architecture

9. Now that we've reviewed one example of changing DNA function, let's discuss changing DNA architecture.

10. See the second slide from today's set. Note that there's a natural sequence of DNA. This sequence of DNA encodes two proteins (gene 2.8 and gene 3), but that the coding sequences for these genes overlap by ~20 base pairs. Moreover, there's a ribosome binding site (a site that initiates protein synthesis) for gene 3, but that this sequence is completely inside the coding sequence for gene 2.8. What's going on here? (answer: nobody fully knows, as yet).

11. What if we wanted to change this genetic architecture?

12. Well, if we could decide upon new features that we wanted our DNA to encode, then we could try to design such DNA, and if successful in terms of coming up with a design, we could then build and test it! For example, let's say that we wanted to have separated genetic parts that were easy to manipulate using classical approaches based on restriction enzymes... What could we come up with?

13. What other architectural changes to DNA might you imagine? Would putting all the genes of an organism in alphabetical order make sense (i.e., be something worth trying)?

14. Note, architectural changes could be easily automated (such as unstuffing elements, or putting genes in alphabetical order).

D. Engineering DNA for Fun, Poetry, Hidden Messages, et cetera

15. Finally, given that DNA is code, one might imagine that it would be possible to store messages or otherwise inside DNA. Turns out that this is true!

16. In order to encode messages in DNA you first need to settle on a code. The code that most people tend to start with (although this has problems, as we'll see) is the "genetic code," which is the mapping from the 64 triplet codons to the 20 amino acids. The reason many folks start with this code is that the 20 amino acids have single English letter abbreviations, which means that if you can spell out a word as a string of amino acids, you can then reverse translate (i.e., lookup) the DNA sequence that would encode those amino acids. Please see today's last slide for the genetic code, with single letter amino acid abbreviations.

17. For example, Ham Smith is a beloved Nobel Laureate who discovered restriction enzymes thereby helping to give birth to biotechnology as we know it. Ham's name translates pretty easily into the single letter amino acid code: HAMSMITH, where H stands for histidine, A stands for alanine, and so on.

18. A ha! Once you have this mapping, you can then lookup the DNA sequence that would code for HAMSMITH. Since the process of reading from nucleic acids to proteins is called translation, this new process is called "reverse translation" (i.e., going back from amino acids to nucleic acids). There are many tools online that will perform reverse translation for you. Look around.

19. catgcgatgagcatgattacccat is the DNA sequence that will code for HAMSMITH.

20. Now, you might ask a few questions at this point. What if there is a mutation? Or, what else could I write? Or, aren't there better codes?

21. As one example, a single point mutation to Ham's signature, cctgcgatgagcatgattacccat,gives the DNA sequence that will code for PAMSMITH, a well known illustrator or Tarot cards from the early 20th century, according to Wikipedia. Hmm. I'll let you think about the other questions.