Lecture and Dialog on Synthetic Biology

Drew Endy (endy@stanford.edu)

Goals for Today

A. Understand assembly of genes from oligos, including error filtration.

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

0. (From last time). 4-step cycle of adding a new base to a growing oligonucleotide is not 100% perfect (i.e., the coupling efficiency is < 100%). Thus, you can't synthesize at reasonable yields oligos above some length. You can estimate the final yield via a simple probability analysis: yield = (coupling efficiency)^N, where N is the number of bases comprising the oligo. So, w/ a 99% coupling efficiency, you'd get yields of 99% (N=1), 90% (N=10), 77% (N=25), 60% (N=50), 36% (N=100), <1% (N=500).

A. Building Genes From Oligos

1. Thus, we need some way to build longer fragments, such as genes, from oligos.

2. Pim Stemmer figured out one interesting way to assemble longer fragments of DNA from oligonucleotides. His method is called "PCR Assembly." Via this method, overlapping oligonucleotides are combined in a single multi-stage PCR reaction, allowing for a longer fragment of DNA to be assembled quickly. (Stemmer et al, "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides." Gene. 1995 Oct 16;164(1):49-53.)

3. We'll walk through PCR slowly via the black board. But, some basic issues with PCR assembly include how long the individual oligos are, how much overlap there is, as the oligos pair together. Also, given that some oligos, during their synthesis, may "cap" (or terminate early), it becomes worth considering that de novo oligo synthesis chemistry builds DNA in the 3->5' direction, but that enzyme mediated DNA replication occurs in the 5->3' direction.

4. Also, PCR assembly can work well, but also can run into big problems. For example, "cross talk" in the hybridization of oligos. As a second example, errors introduced by the polymerases used in the PCR process. Practically, today, it's very difficult to get PCR assembly to work reliably (i.e., on any DNA sequence) beyond ~2,500 bp, and even at that length you can run into problems with sequences that have a high proportion of C or G based (aka high "GC content"), or that have repeated sequences. There are other methods of assembling oligos into genes, that are more akin to a serial ligation process, and these can work with higher reliability.

5. However, methods are available to detect and filter errors in the assembly of genes from oligos. For example, point mutations (single base pair errors, arising either in the oligo synthesis or gene assembly stages) can be detected and filtered out. The way this works is for single stranded DNA to be hybridized against sequences that should be the same, but that may have errors (and thus will be slightly different). These single base mutations can be detected by natural proteins (e.g., MutS) that bind to single base mismatches in double stranded DNA. If the MutS protein is attached to a bead or can be otherwise removed from solution, errors in individual DNA molecules can be removed from a population of DNA molecules. Note, if the error rate in individual oligos is too high, then an error filtration process based on "yanking out" mistmatched oligos will remove everything. That is, at least some of your individual oligo molecules need to be error free.

6. Now, as would-be engineers of biology, we face an interesting and important question. Do we want to become experts on the synthesis of oligos and genes? Perhaps making improvements to the basic chemistry or assembly process? Or, do we want to be consumers of oligo synthesis and gene assembly tools and services? (Note that the same question will hold for longer length DNA, such as genomes). There are important opportunities in both areas, and the future of biotechnology will depend on revolutionaries and experts in both areas, but what should you do? If you choose to be an expert on synthesis and construction, then maybe people with brilliant ideas about what to make will seek you out? If you choose to be an expert on "content" then you might be able to choose from many possible builders! (but you better have something important or interesting to make!)

B. <u>Building Genomes From "Genes"</u>

7. Please note that my use of the word "gene" here is really just a placeholder, for a fragment of DNA somewhere between 500 to ~15,000 base pair long. There are many DNA sequences that fall within this length range, from single genes to entire genome. However, most genomes are > 15,000 base pairs. For example, bacteriophage lambda's genome is 45,802 base pairs long. The E.coli genome is about 4 million base pairs long, and so on. So, how could we build a longer piece of DNA, a "genome length" fragment?

8. Building genome length DNA is a relatively esoteric area of research (this doesn't mean that it's unimportant, quite the opposite)!. So far, I've seen three classes of approaches to the puzzle. What's interesting is that these three approaches can be distinguished from one another in terms of the "order" of the assembly process. The word "order" here is being used in an algorithmic sense, and has to do with how many steps are involved in carrying out the process.

9. For example, Itaya and colleagues from Mitsubishi built a ~7,700,000 base pair genome. To do this they invented a process which they called "Inch Worm Elongation" (or IWE for short). IWE as you might imagine turns out to be a very slow process. As a result, it took them ~7 years to build their genome. This is, in part, due to the fact that IWE is a "linear" or "serial" process, in which many individual fragments are

added one after the other. Thus, you can't start working on adding the last piece until you've added every other piece. If you had to put 8 fragments together it would require 7 steps. Hmmm...You can read more about their work here: "Combining two genomes in one cell: stable cloning of the Synechocystis PCC6803 genome in the Bacillus subtilis 168 genome." by Itaya M, Tsuge K, Koizumi M, Fujita K., Proc Natl Acad Sci U S A. 2005 Nov 1;102(44):15971-6. Epub 2005 Oct 18.PMID: 16236728

10. As a second example, Dan Gibson built most of a smaller genome, ~600,000 base pairs long, but he did it via a "logarithmic" method. Starting from many individual "gene" length fragments of DNA, he could combine small numbers of the fragments in parallel, making a smaller number of fragments of intermediate length that could also be combined, and repeat. Thus, rather than having to go through N-1 steps, where N is the number of fragments to combine (as in a linear process) Dan's method requires ~log2(N) steps. Thus, if you had to put 8 fragments together it would take only 3 steps! Logarithmic assembly methods have been around for awhile. For example, Tom Knight's BioBrick Assembly Standard #1 is a good early example. Read more about it online here: http://dspace.mit.edu/handle/1721.1/21168

11. As a final example, Dan Gibson rebuilt the same genome, but via a parallel process. He did this by adding all the fragments at once into yeast, and then relying on yeast recombination to figure out how to bring everything together. So, if you had 8 fragments to combine, this would only take 1 single step. Wow! (Note that this is similar to PCR assembly, in that all oligos are added at once). You can read more about Dan's most recent work here: http://www.pnas.org/content/105/51/20404.short (It will be interesting and important to determine how reliably this process is, and how far it can be extended, in terms of the number of fragments that can be combined at once, how big or small the fragments need to be, et cetera)

C. Impact of DNA Construction

12. (let's quickly check what's available online, in terms of DNA synthesis and assembly)

13. Remember #6 above: will you be a designer or builder of DNA? Both paths are important are important and need improving, but will lead to different experiences, and shape the markets and the world in different ways. Should you combine to two somehow?

14. What else will DNA synthesis impact, if you can go from genetic information to genetic material?

D. For Next Time

15. Please complete and turn in Assignment #1 by the beginning of Thursday's class.