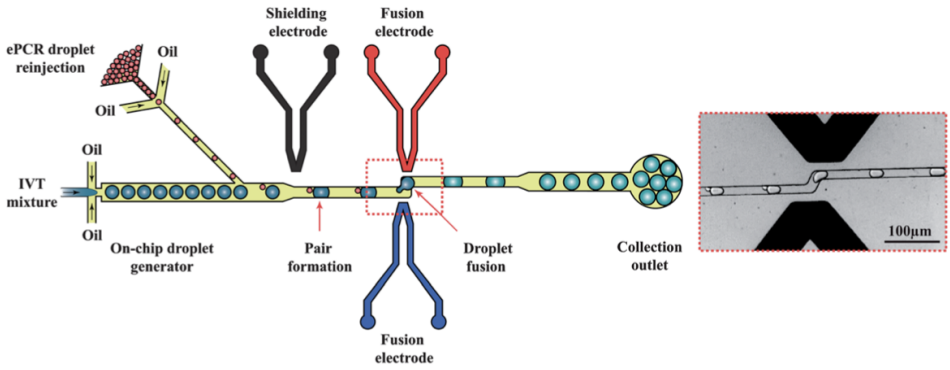
2018 ChemE 355
Problem Set 4
Due Tuesday, June 5th in class

Problem 1: Protein engineering using droplet microfluidics

In class we discussed the benefits and challenges of evolving a new protein in vivo. We also discussed ways to encapsulate proteins and the DNA that encodes them to improve the throughput of in vitro protein engineering. A paper by Griffiths et al (Lab on a Chip, 2012) describes one such approach to in vitro protein engineering using cell free protein synthesis within droplets and illustrates some of the complexities involved in making this work. The overarching goal is to develop an ultrahigh-throughput method using both in vitro expression and assay of a large library of enzyme variants. As a proof of concept experiment, they worked with the enzyme beta-galactosidase LacZ, which can be directly assayed using fluorescence (LacZ can hydrolyze the substrate fluorescein-di-β-D-galactopyranoside to generate the fluorescent product fluorescein). To test if in vitro protein expression and assay would work, they expressed the enzyme lacZ and an inactive variant, ΔlacZ that contained a frame-shift mutation within the middle of the sequence. They then demonstrated that droplets (20 pL each) containing the expressed protein could be sorted by fluorescence to isolate only the active beta-galactosidase and not the inactive variant as an important step towards developing a high throughput in vitro complement to in vivo protein engineering.

Overview of workflow to generate droplets that will then be sorted based on fluorescence: (IVT mixture provides components for transcription and translation and protein assay; this mixture will be fused to droplets containing PCR products as shown)



For the emulsion PCR step, the authors determine that a concentration λ of 0.15 molecules of template DNA per droplet on average is ideal for their system. This results in ~12% of droplets containing one molecule of template, and <1% of droplets containing >1 template molecule.

(a) Why is it important to limit the number of droplets that have >1 template molecule? What would happen if a significant % of droplets contained 2 template molecules?

(b) What would happen to the signal in the beta-galactosidase reaction if one template molecule each for lacZ and ΔlacZ are encapsulated in a droplet? How would this signal be interpreted in terms of protein function?

.

(c) After 32 rounds of PCR, how many total strands of DNA should they theoretically have observed per droplet if they started from one double stranded molecule of template? (Doesn't need to be exact, order of magnitude answer is ok)

(d) In practice, the authors observed 30,000 copies of each gene for droplets containing 1 molecule of DNA template. What is one reason that could account for the discrepancy between this number and your answer to part b?

(e) Does the strategy describe a screen or selection? Please explain in one or two sentences what the difference is between a screen and a selection, and which one better describes this technique.

(f) Why go through the trouble of doing transcription and translation in the droplet? Why not just do transcription/translation in bulk solution and then encapsulate protein molecules? Assume for simplicity that you can get enough signal from one molecule of protein to enable droplet sorting.

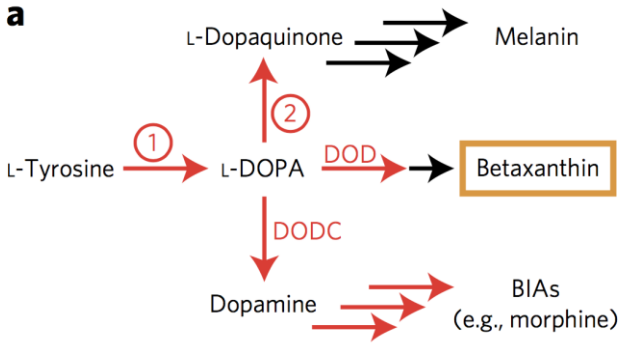
(g) Imagine this approach were used in a paper entitled, “An engineered variant of LacZ that has >1000x the activity of the wild-type enzyme”. In this hypothetical paper, error-prone PCR is used to generate a library of LacZ mutants, which were transcribed and translated to protein in droplets and then sorted. What is the next step after the final round of screening?

(h) In order to produce protein, each droplet containing DNA after PCR is then fused with a second droplet containing transcription and translation reagents (see figure, above). Given 30,000 copies of DNA/droplet, a transcription rate of 48 nucleotides/sec, an extra long transcript half life of 10 min, and a translation rate of 16 AA/sec, what is the frequency of translation initiation in the cell-free system for lacZ (1023 residues)?

(i) For the drops to register a fluorescent signal, they need a concentration of at least 100 uM of fluorescein. Assuming constant linear production of lacZ, how long do the scientists have to incubate the drops to create enough flourescein to trigger the detector given that lacZ has a kcat of 15 flourescein/sec/lacZ (Avogadros # = 6.022e23)

Problem 2: Protein engineering for morphine biosynthesis

In a 2015 paper in Nature Chemical Biology about engineering the early steps of the morphine pathway into yeast, Dueber and coworkers built a library of 200,000 CYP76AD1 mutants in order to enhance enzyme-catalyzed conversion of tyrosine to L-DOPA (the first step in the morphine pathway) and also limit a second oxidation also catalyzed by the WT enzyme that converts L-DOPA to the L-Dopaquinone (en route to melanin – see figure below). The paper is posted on canvas, but is not required to answer the questions.



(a) In order to search for the desired activity of a CYP76AD1 variant, described above, the authors developed an in vivo biosensor that provided a direct, fluorescent readout for L-DOPA production. In addition, the authors chose to develop a second assay that provides an indication of the combined L-DOPA and L-Dopaquinone pools by production of a violet pigment. What additional information is given in this second, less specific assay?

(b) The authors note that 6 of the 17 clones that they characterized with improved L-DOPA production contained the same F309L missense mutation, and that among these, there were two distinct codon changes. Imagine that one has the Phe codon UUU mutated to UUA (Leu), and the other has UUU mutated to CUU (a different Leu codon). What is the significance of this finding? List two reasons why amino acid changes in a given enzyme sequence could lead to accumulation of more product.

(c) DNA shuffling and another round of error-prone PCR is then performed using the six mutants with the largest improvements in fluorescence (indicating the highest yield of DOPA). What is the theory behind DNA shuffling, and why is it superior to the simpler alternative of just another round of error-prone PCR of the mutant sequence with the highest activity?

(d) What is a technique the authors could have used to identify deleterious mutations in a single clone that might have accumulated during the process of error-prone PCR? Describe the major steps involved, and how deleterious mutations would be identified.

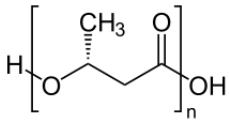
(e) The following table taken from the supplemental information lists the mutations found in the six variants that were used to undergo DNA shuffling in the paper.

	Mutations				
Mutant #1	*87A>C*	L141I			
Mutant #2	*123T>A*	F309L	E465D		
Mutant #3	D2E	*150G>A*	*327C>T*	Y380H	
Mutant #4	*9T>C*	W13L	*1236T>C*	*1281T>C*	
Mutant #5	*180T>A*	F127L	*576G>A*	S232T	*714G>A*
Mutant #6	*684C>T*	F309L			

After performing a round of DNA shuffling with these six mutants, the authors determined (by sequencing active clones) that the combination of mutations F309L and W13L are most beneficial for enzyme activity. If we were to recombine the mutant sequences #1-6 using DNA shuffling (but without the error-prone PCR step), what is the smallest library size needed in order to obtain 20 clones containing these two mutations? Assume equivalent starting amounts of each mutant sequence are used.

(f) Aro4P catalyzes the first step in the shikimate pathway in yeast (recall the shikimate pathway is the source of all aromatic amino acids) and is known to be quite sensitive to tyrosine feedback inhibition. Luckily a variant Aro4P*FBR* had previously been identified that does not suffer from this problem. What is feedback inhibition, and why is it problematic for Aro4P in the context of L-DOPA production in yeast?

Problem 3: Production of the bacterial plastic polyhydroxybutryate



Poly-3-hydroxybutyrate (PHB), structure above, is a biodegradable thermoplastic accumulated intracellularly by many microorganisms under unfavorable growth conditions. *Azotobacter chroococcum* is being investigated for commercial PHB production using cheap soluble starch as the raw material and ammonia as the nitrogen source. Synthesis of PHB is observed to be growth associated with maximum production occurring when the culture is provided with limited oxygen. During steady-state continuous culture of *A. chroococcum*, the concentration of PHB in the cells is 44% w/w and the respiratory coefficient is 1.3. From elemental analysis, *A. chroococcum* biomass without PHB can be represented as $CH_2O_{0.5}N_{0.25}$. The monomeric unit for starch is $C_6H_{10}O_5$; $C_4H_6O_2$ is the monomeric unit for PHB.

(a) Develop an empirical reaction equation for PHB production and cell growth. PHB can be considered a separate product of the culture even though it is not excreted from the biomass.

(b) What is the yield of PHB-containing cells from starch in units of g g^{-1}?

Problem 4: Oxygen demand for production of recombinant protein in *E coli*

Recombinant protein is produced by a genetically engineered strain of *Escherichia coli* during cell growth. The recombinant protein can be considered a product of cell culture even though it is not secreted from the cells; it is synthesized in addition to normal *E. coli* biomass. Ammonia is used as the nitrogen source for aerobic respiration of glucose. The recombinant protein has an overall formula of $CH_{1.55}O_{0.31}N_{0.25}$. The yield of biomass (excluding recombinant protein) from glucose is measured as 0.48 g g^{-1}; the yield of recombinant protein from glucose is about 20% of that for cells.

(a) How much ammonia is required?

(b) What is the oxygen demand?

(c) If the biomass yield remains at 0.48 g g^{-1}, how much different are the ammonia and oxygen requirements for a wild-type strain of *E. coli* that is unable to synthesize recombinant protein?