


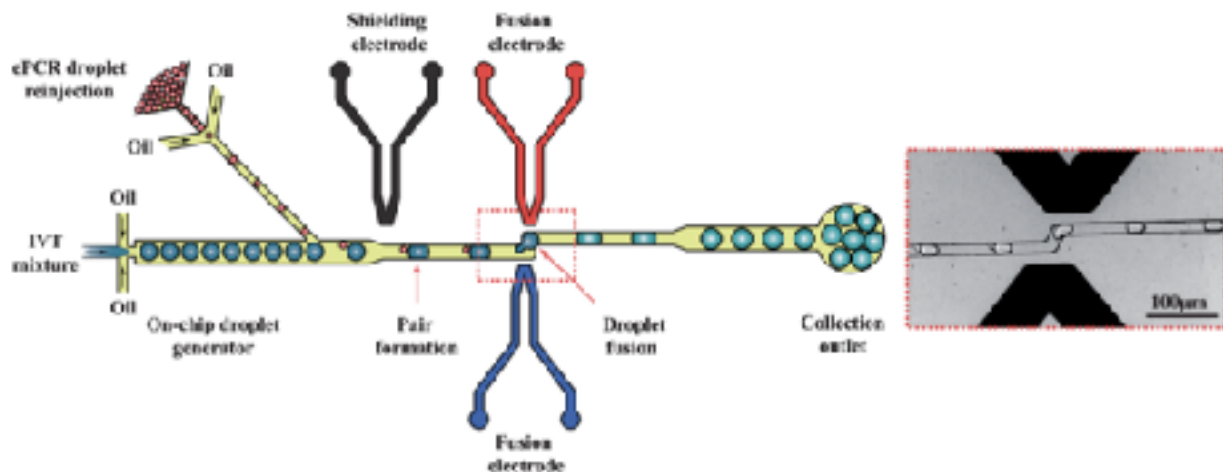
2018 ChemE 355
Problem Set 4
Due Tuesday, June 5th in class

DOUG CHANG 04612354

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--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 [?] 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?

Droplets which contain 2 template molecules have a lower probability of being pure. One template molecule could be the fluorescent one and the other not. You would have 4 different combinations,

- 1) lacZ, lacZ
- 2) lacZ, inactive lacZ
- 3) inactive lacZ, lacZ
- 4) inactive lacZ, inactive lacZ

The first one would fluoresce 2x as bright, the second one and third one normal if there was no deleterious interaction between the 2 enzymes.

An impure droplet with more than one template molecule can cause the following problems:

- 1) when induced, there may be an incomplete reaction because there are 2x the number of template molecules whereas the PCR reagent concentrations were designed for only 1.

- 2) can get strands which are combinations of other inactive genes which we want to remove. There is a 75% chance if all outcomes are equally probable that we get contamination and upon sequencing we get bad/unexpected/inconsistent results.
- 3) have reduced amount of substrate available for the lacZ gene if the lacZ-delta is using up substrate even though there was a early stop codon in the frame shift.

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

One would fluoresce the other wouldn't. Assuming there was enough reagent for both reactions we would get a reduced signal both as a result of less substrate available and additional dilution in the droplet.

(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)

$$2^{32} = 4.3 \times 10^9$$

(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?

More than .15 template molecule in the droplet. Incomplete PCR.

(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.

This is a screen.

Screen: we look at all the clones,

Selection: eliminate some clones using a fitness test.

In vivo: in cell, In vitro: outside cell.

In vivo selection: fastest, but the hardest to setup. Need fitness test; i.e. cells with antibiotic resistance. 10^5 as example of order of magnitude processing rate. Can use flow cytometry here instead of having a human look at each one on a plate.

In vivo screening; need human or robot/tools to manually observe and eliminate. Hundreds/hour.

In vitro screening: most reliable, hardest to do. Most reliable. Limits the size of the library, ability to keep DNA w/protein. Use microfluidics for directed evolution. Display protein from yeast cell on surface, add in reagents and look for fluorescence on droplet.

In vitro selection: outside of an organism. Add reagent like DMF or heat, some selection mechanism which prevents enzyme formation outside of cell. Not killing off organisms here.

(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.

To keep the DNA correlated with the produced protein. Bulk translation/transcription would yield protein after protein purification but it would be very difficult to determine which DNA sequence belonged to a specific protein mutation (AA sequence or folding variation).

(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?

Verify the protein and DNA correlate with each other; the screened DNA really produces the engineered variant of 1000x. We can retransform this DNA sequence into another parent.

(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)?

Assume 1023 AA is length of lacZ.

Transcription DNA->mRNA $\frac{48nt}{sec} * 600s = 28800nt$. Check, transcription half life can transcript ~30k bp DNA strand.

mRNA->codons $28000nt * \frac{1codon}{3nt} = 9600codons$

codons->AA $9600codons * \frac{1AA}{codon} * \frac{1sec}{16AA} = 600s$. Transcription time. check at 10 minutes. Consistent data. Awfully verbose problem. Probably misunderstood the problem again.. so confusing.

From HW announcement on canvas; assume 100nt between ribosomes.

$\frac{16AA}{s} * \frac{1codon}{AA} * \frac{3nt}{codon} = 48nt/translationrate$. mRNA->AA

Translation Initiation Rate = 100nt/48 nt/s = 2.1 s. Translation initiation every 2.1 s

(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 fluorescein to trigger the detector given that lacZ has a kcat of 15 fluorescein/sec/lacZ (Avogadro's # = 6.022e23)

Each drop is 20pL. Each drop needs to have 100uM of fluorescein.

each LacZ produces 15 fluorescein/sec. How many sec to generate enough LacZ for 100uM f in 10pL drops.

$$100\mu M = 100 \frac{10^{-6} \text{ moles}}{L} * \frac{1L}{10^{12} pL} * \frac{6 * 10^{23} \text{ molecules}}{1 \text{ mole}} = \frac{6 * 10^{23} \text{ molecules}}{10^{16} pL} = 6 * 10^7 \frac{\text{molecules}}{pL}$$

For a 20pL drop we need $120 * 10^7$ molecules of fluorescein for the detector to see it which is 20x more than the amount in 1pL.

one mole of lacZ makes 15 moles of fluorescein/sec.

$$120 * 10^7 \text{ molecules fluorescence} * \frac{1 \text{ lacZ}}{15 \text{ fluorescence sec}} = 8 * 10^7 \text{ seconds}$$

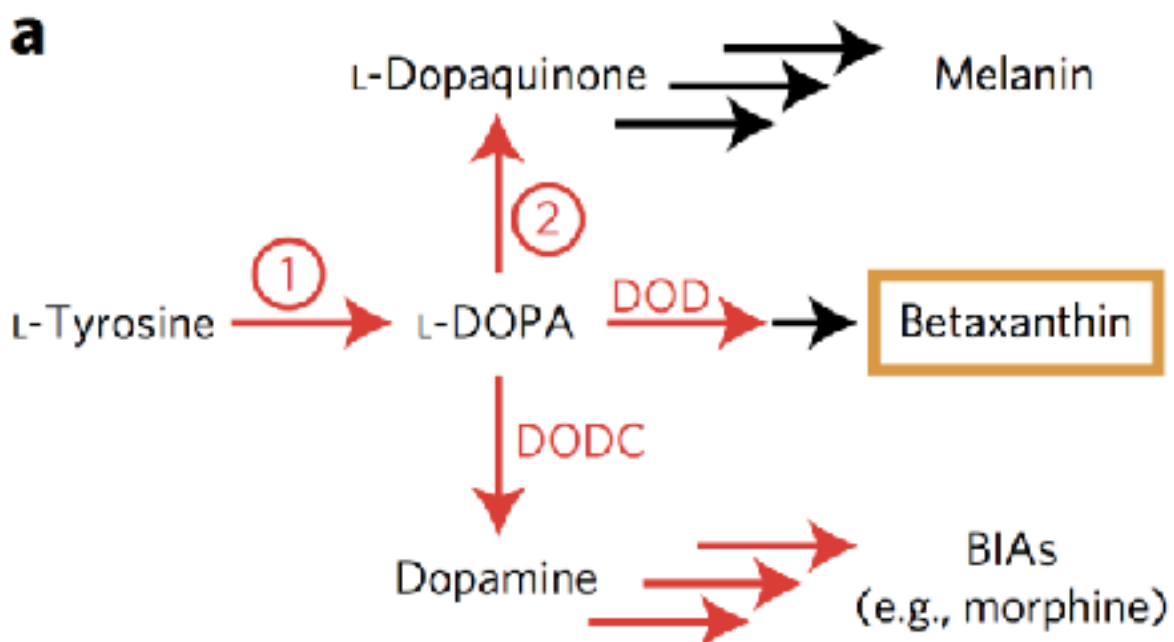
Problem 2: Protein engineering for morphine biosynthesis

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?

The second assay gives details about the the second oxidation from L-DOPA -> L-Dopaquinone. They want to limit this reaction and the second assay gives data on the second oxidation and how to correlate it with the mutants.

(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.



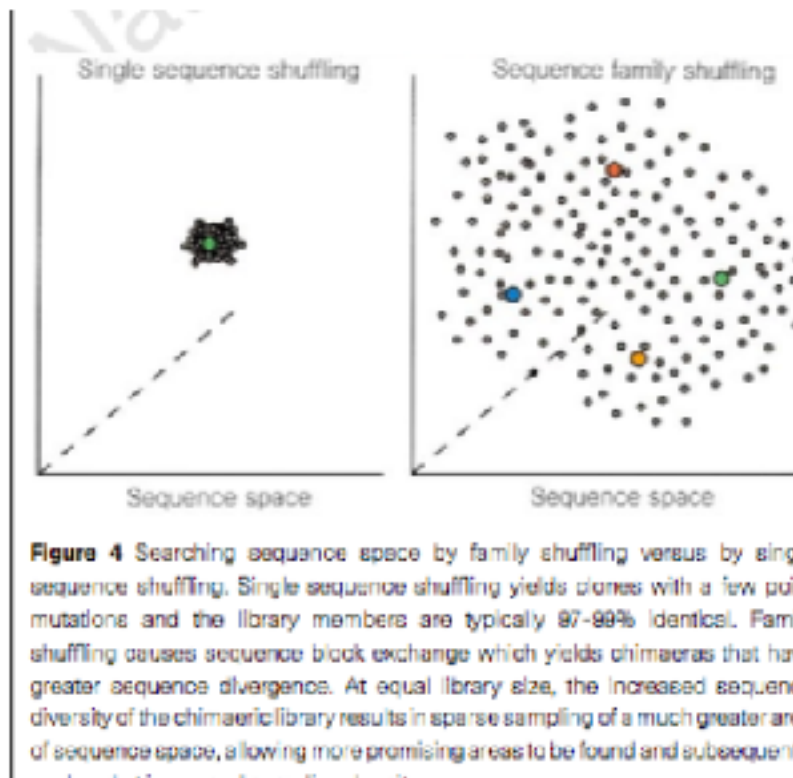
1) Codon optimization; UUU is not being depleted. Allows more production of UUU pathway.

2) DNA shuffling, allowing development of more functions/AA products.

(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?

DNA Shuffling coupled with selection is the process for mutating genes inside a living organism and allowing the mutations to replicate and further mutate. DNA shuffling uses homology to mix related genes from different species.

Select out the variants of interest. DNA shuffling is faster to implement in searching a large search area vs. just performing individual point mutations. Graph showing mutation space vs. point mutation search space. The graph on the right shows the search space for DNA shuffling + point mutations vs. only point mutations on the left.



(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.

Back crossing gives the probability of good error prone PCR random mutations. 1- Probabiltiy of mutations gives deleterous/null mutations.

(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>G	L141I			
Mutant #2	123T>A	F309L	E465D		
Mutant #3	D2E	150G>A	327G>I	Y380H	
Mutant #4	9T>C	W13L	1236T>G	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.

Desire 20 F309L + W13L after second round. Requires 20 each of F309L from Mutant2 and W13L from Mutant4.

There are 20 genes from a pool of 6 mutants. There is a $1/20$ chance of picking one of the 20 genes and a $(1/20)(1/20)$ chance of picking 2 genes from the pool. $1/400$ for 2 genes assuming each is equally probable.

For 20 of these clones, $20(400) = 8000$ required and in that pool of 8k there are 20 copies of any $2/20$ genes. Assuming all 20 are equally probable.

(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?

Feedback inhibition is negative feedback where as the output wants to grow the feedback signal increases further prohibiting growth. This causes a self stabilizing system where the output signal is magnified by some gain.

Would limit the amount of tyrosine avail to L-DOPA production and could become a bottleneck.

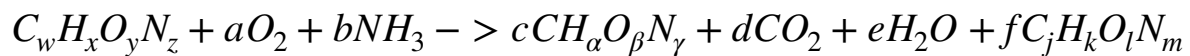
Problem 3: Production of the bacterial plastic polyhydroxybutyrate

Poly-3-hydroxybutyrate (PHB), structure above, is a biodegradable thermoplastic accumulated intracellularly by many microorganisms under unfavorable growth conditions. *Azotobacter chroococcum* is

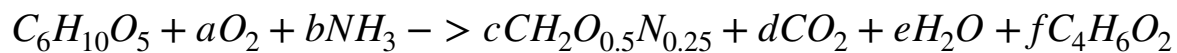
being investigated for commercial PHB production using cheap soluble starch as the raw material and ammonia as the nitrogen source. Synthesis of PHB is observed to be growth associated with maximum production occurring when the culture is provided with limited oxygen. During steady-state continuous culture 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.

From the general form we substitute starch for the first term and the product PHB for the last term.



after substitution w/the starch monomer and PHB monomer:



w=6, x=10, y=5, z=0, alpha=2, beta=0.5, gamma=.25, j=4, k=6, l=2, m=0

$$C: w = c + d + fj$$

$$O: y + 2a = \beta c + 2d + e + fl$$

$$H: x + 3b = \alpha c + 2e + fk$$

$$N: z + b = \gamma c + fm$$

$$\text{Respiration rate} = \frac{\text{mol}CO_2}{\text{mol}O_2} = 1.3 = \frac{d}{a}; 1.3a = d$$

Substitute constants and eliminate d for a:

$$\text{C: } 6 = c + 1.3a + f4$$

$$\text{O: } 5 + 2a = .5c + 2.6a + e + f2$$

$$\text{H: } 10 + 3b = ac + 2e + f6$$

$$\text{N: } b = .25c$$

Before electron balance we have variables a,b,c,e,f or 5 variables and 4 equations.

We need 1 more equation to solve for the variables.

Electron balance

$$\gamma_s : \frac{(4w + x - 2y - 3z)}{w} = (4(6) + 10 - 2(5))/6 = 4 \text{ for glucose}$$

$$\gamma_b : (4 + \alpha - 2\beta - 3\gamma) = 4(1) + 1(1.77) - 2(.49) - 3(.254) = 4.03$$

$$\gamma_p : \frac{(4j + k - 2l - 3m)}{j} = 4(1) + 1(1.55) - 2(.31) - 3(.25) = 4.18$$

$$\text{balance : } \gamma_z - 4a = c\gamma_b + fj\gamma_p$$

$$a = \frac{1}{4}(w\gamma_s - c\gamma_b - fj\gamma_p) = \frac{1}{4}(6 * 4 - 4.03c - (4.18 * 4)f); \text{ we know } j=4, \\ w=6. \text{ Cancel out the 4. } a = 6 - c - 4.18f$$

OK solve for the variables. 5 variables and 5 equations.

$$\text{C: } 6 = c + 1.3a + f4$$

$$\text{O: } 5 + 2a = .5c + 2.6a + e + f2$$

$$\text{H: } 10 + 3b = ac + 2e + f6$$

$$\text{N: } b = .25c$$

$$a = 6 - c - 4.18f$$

1) eliminate b. ; $b = .25c$

$$C: 6 = c + 1.3a + f4$$

$$O: 5 + 2a = .5c + 2.6a + e + f2$$

$$H: 10 + .75c = ac + 2e + f6$$

$$a = 6 - c - 4.18f$$

Simplify:

$$C: 6 = c + 1.3a + f4$$

$$O: 5 = .5c + 6a + e + f2$$

$$H: 10 + .75c = ac + 2e + f6$$

$$a = 6 - c - 4.18f$$

Eliminate a:

$$6 = c + 1.3(6 - c - 4.18f) + f4$$

$$5 = .5c + 6(6 - c - 4.18f) + e + f2$$

$$10 + (.75 - (6 - c - 4.18f))c = 2e + f6$$

Simplify:

$$6 = c + 7.8 - 1.3c - 5.4f + 4f$$

$$5 = .5c + 36 - 6c - 25.08f + e + f2$$

$$10 + (.75 - (6 - c - 4.18f))c = 2e + f6$$

Simplify:

$$-1.8 = .3c - 1.4f$$

$$-31 = -5.5c - 23.08f + e$$

$$10 + (.75 - (6 - c - 4.18f))c = 2e + f6$$

Simplify:

$$f = \frac{1.8 + .3c}{1.4}$$

$$e = -1.3 + 55c$$

$$10 + (.75 - (6 - c - 4.18f))c = 2e + f6$$

last equation simplifies:

$$10 - 5.25c + c^2 + 4.18fc = 2e + 6f$$

$$10 - 5.25c + c^2 + 4.18\left(\frac{1.8 + .3c}{1.4}\right)c = 2(-1.3 + 55c) + 6\left(\frac{1.8 + .3c}{1.4}\right)$$

$$1.9c^2 - 111.6c + 4.9$$

$$c = 58.7, .04$$

$$b = 14.7, .01$$

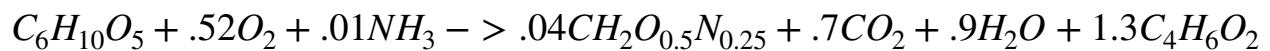
$$f = 13.9, 1.3$$

$$e = 3227.2, .9$$

$$a = -110.8, .52$$

$$d = 1.3a = -144, .7$$

Take second column which is not negative moles.



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

Convert from moles/moles to g/g. PHB containing cells/substrate = f moles/c moles.

Cell weight w/PHB g/ Cell weight wo PHB = .44

Confused what this question is asking and how yield of PHB containing cells is different from $Y_{PS} = \frac{g\ PHB}{g\ substrate\ used} = f \frac{MW_{product}}{MW_{substrate}}$; both definitions are in g/g. Ugh....well I think I can understand the book. All the other sources not so much...

The book uses biomass yield, is this the same as PHB containing cells yield?

$$C_{max} = \frac{w\gamma_s}{\gamma_b} = \frac{6 * 4}{4.03} = 5.95 \frac{mol_{biomass}}{mol_{starch}} * \frac{1 mol_{starch}}{162g} * \frac{25.5g_{biomass}}{1 mole_{biomass}} = .94gg$$

$$MW\ starch = 6(12) + 10 + 5(16) = 162$$

$$MW\ biomass = 12 + 2 + (.5)16 + (.25)14 = 25.5$$

$$MW\ prodduct = 12(4) + 6 + 32 = 86$$

Or is the max product yield the same as PHB containing cell yield? is:

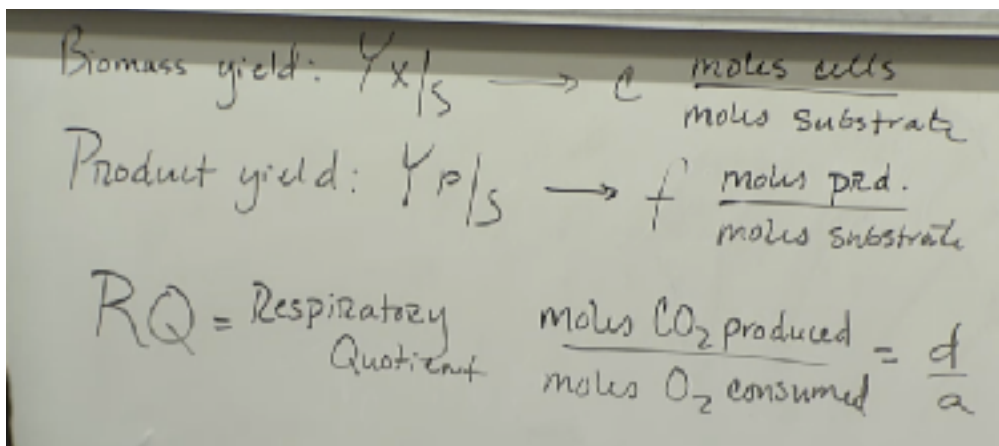
$$f_{max} = \frac{w\gamma_s}{j\gamma_p} = \frac{6 * 4}{4.18} = 5.74 \frac{mol_{prouct}}{mol_{starch}} * \frac{1 mol_{starch}}{162} * \frac{86g_{product}}{1 mole_{product}} = 3.05 \frac{g}{g}$$

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.

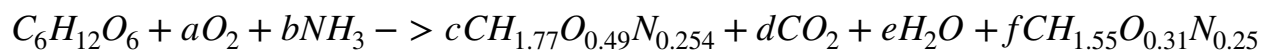
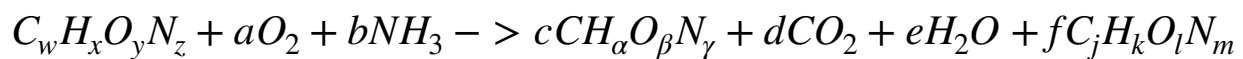
Ammonia is used as the nitrogen source for aerobic respiration of glucose. The recombinant protein has an overall formula of $CH_{1.55}O_{0.31}N_{0.25}$. The yield of biomass (excluding recombinant protein) from

glucose is measured as 0.48 g g^{-1} ; the yield of recombinant protein from glucose is about 20% of that for cells.



Handwritten definitions:

- Biomass yield: $Y_{x/s} \rightarrow c \frac{\text{moles cells}}{\text{moles substrate}}$
- Product yield: $Y_{p/s} \rightarrow f \frac{\text{moles prod.}}{\text{moles substrate}}$
- $RQ = \text{Respiratory Quotient} = \frac{\text{moles } CO_2 \text{ produced}}{\text{moles } O_2 \text{ consumed}} = \frac{d}{a}$



$$Y_{\frac{x}{s}} = .48 \frac{\text{g}}{\text{g}}; Y_{\frac{p}{s}} = (.20) * .48 \frac{\text{g}}{\text{g}} = .1 \frac{\text{g}}{\text{g}}$$

w=6, x=12, y=6, z=0, alpha=1.77, beta=0.49, gamma=.254,
j=1,k=1.55,l=0.31,m=0.25

$$C: w = c + d + fj; 6 = c + d + f$$

$$O: y + 2a = \beta c + 2d + e + fl; 6 + 2a = 0.49c + 2d + e + 0.31f$$

$$H: x + 3b = ac + 2e + fk; 12 + 3b = ac + 2e + 1.55f$$

$$N: z + b = \gamma c + fm; b = .254c + .25f$$

Electron balance

$$\gamma_s : \frac{(4w + x - 2y - 3z)}{w} = (4(6) + 12 - 2(6))/6 = 4 \text{ for glucose}$$

$$\gamma_b : (4 + \alpha - 2\beta - 3\gamma) = 4(1) + 1(1.77) - 2(.49) - 3(.254) = 4.03$$

$$\gamma_p : \frac{(4j + k - 2l - 3m)}{j} = 4(1) + 1(1.55) - 2(.31) - 3(.25) = 4.18$$

$$balance : \gamma_z - 4a = c\gamma_b + fj\gamma_p$$

$$MW \text{ product} = 12 + 1.55 + 16(0.31) + 14(0.25) = 22.01$$

$$MW \text{ biomass} = 12 + 1.77 + 16(0.49) + 14(.254) = 25.17$$

$$MW_{\text{glucose}} = 6 \cdot 12 + 12 + 16 \cdot 6 = 180$$

From the book $Y_{xs} = \text{g cells produced/g substrate produce} = c(\text{MW cells})/\text{MW substrate}$.

Calculate c from biomass yield:

$$Y_{XS} = \frac{.48g \text{ biomass}}{1g \text{ glucose}} * \frac{180g \text{ glucose}}{1mole \text{ glucose}} * \frac{1mole \text{ biomass}}{25.17g \text{ biomass}} = 3.43 \frac{mol \text{ biomass}}{mol \text{ glucose}} = c; \text{ this is too big, do we add MW of protein to the biomass?}$$

No.

Calculate f from product yield by converting from g to moles:

$$Y_{PS} = .1 \frac{g_{product}}{g_{glucose}} * \frac{180g_{glucose}}{1mole_{glucose}} * \frac{1mole_{product}}{22.01g_{product}} = .82 moles_{product}/moles_{glucose}$$

$$= f(22.01)/80$$

$$f = .82(80)/22.01 = 2.98$$

(a) How much ammonia is required?

Find b.

$$b = .254c + .25f$$

$$b = .254 * 3.43 + .25 * 2.98 = 1.62 \text{ moles for 1 mole substrate}$$

(b) What is the oxygen demand?

$$\text{Find a: } a = \frac{1}{4}(w\gamma_s - c\gamma_b - fj\gamma_p) = \frac{1}{4}(6 * 4 - 4.03c - 4.08f) = 6 - c - f$$

From above, $c = 3.43$, $f = 2.98$, $a = -.41$; argh this is wrong.. should be > 0 .

Math error somewhere...

(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?

$$\gamma_b = .48 \frac{\text{g}}{\text{g}} .$$



No f term.

$$Y_{XS} = \frac{.48 \text{ g biomass}}{1 \text{ g glucose}} * \frac{180 \text{ g glucose}}{1 \text{ mole glucose}} * \frac{1 \text{ mole biomass}}{25.17 \text{ g biomass}} =$$

$$3.43 \frac{\text{mol biomass}}{\text{mol glucose}} = c$$

Compare b,a

$$b = .254c = .87$$

$$d = 6 - c = 2.57$$

$$12.61 = 2.57a + 2e$$

$$5 + 2a = 6.82 + e$$

$$2a - 1.82 = e$$

$$12.61 = 2.57a + 2(2a - 1.82)$$

$$a = \frac{12.61 + 2(1.81)}{6.57} = 2.47$$

Compare .87, 2.47 to above. I got negative a, clearly incorrect.

$$w=6, x=10, y=5, z=0, \alpha=1.77, \beta=0.49, \gamma=.254, j=0, k=0, l=0, m=0$$

$$C: w = c + d + fj; 6 = c + d$$

$$O: y + 2a = \beta c + 2d + e + fl; 5 + 2a = 0.49c + 2d + e$$

$$H: x + 3b = \alpha c + 2e + fk; 10 + 3b = \alpha c + 2e$$

$$N: z + b = \gamma c + fm; b = .254c$$

Electron balance

$$\gamma_s : \frac{(4w + x - 2y - 3z)}{w} = (4(6) + 10 - 2(5))/6 = 4 \text{ for glucose}$$

$$\gamma_b : (4 + \alpha - 2\beta - 3\gamma) = 4(1) + 1(1.77) - 2(.49) - 3(.254) = 4.03$$

$$balance : \gamma_s - 4a = c\gamma_b$$

$$MW \text{ product} = 12 + 1.55 + 16(0.31) + 14(0.25) = 22.01$$

$$MW \text{ biomass} = 12 + 1.77 + 16(0.49) + 14(.254) = 25.17$$

$$MW_{\text{glucose}} = 80$$