Chemical Engineering 355 Advanced Biochemical Engineering Spring 2018

FINAL Exam

NAME: ____Doug Chang 04612354 I understand and follow the Stanford Honor Code **SIGNATURE:**

_____

The Honor Code is an undertaking of the students, individually and collectively:

1. that they will not give or receive aid in examinations; that they will not give or receive unpermitted aid in class work, in the preparation of reports, or in any other work that is to be used by the instructor as the basis of grading;

2. that they will do their share and take an active part in seeing to it that others as well as themselves uphold the spirit and letter of the Honor Code.

The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid, as far as practicable, academic procedures that create temptations to violate the Honor Code.

While the faculty alone has the right and obligation to set academic requirements, the students and faculty will work together to establish optimal conditions for honorable academic work.

The final exam will be due Monday, June 11th at 12 PM PST outside of Beth's office, Shriram 271. Please submit a hard copy if you are an on-campus student. All answers should be recorded on the exam – if you need additional space, please attach extra sheets of paper and clearly indicate which problem they belong to. A copy of the genetic code and metabolic pathways are provided in the appendix.

If you have any questions during the exam, please include Nikita,

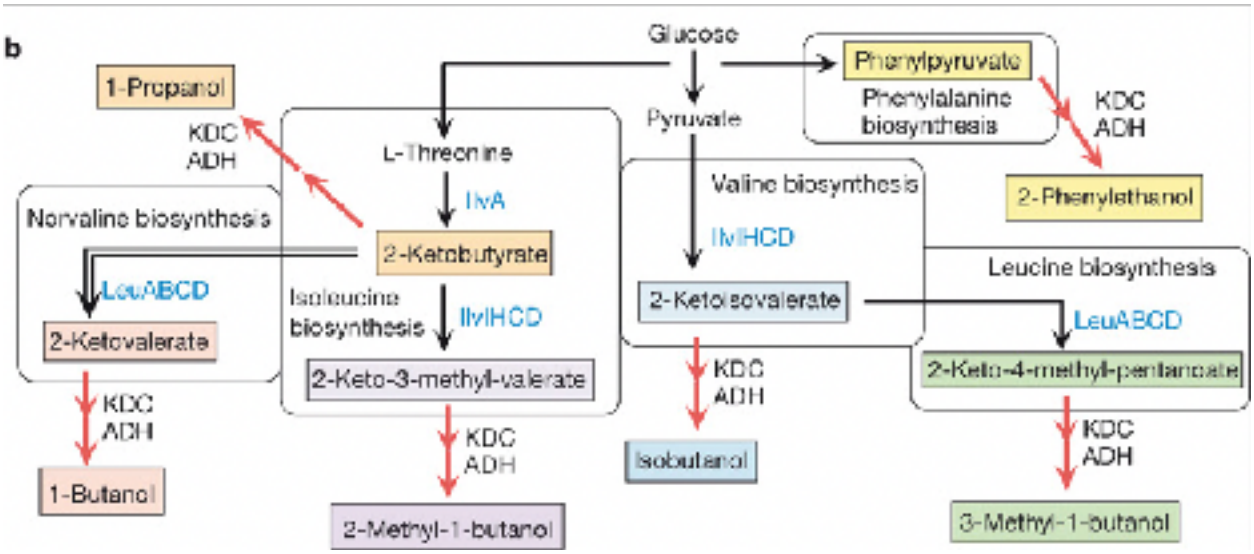
Osman, and Beth on all emails.

You can use a calculator and refer to any in-class or other personally written notes. A computer or laptop can be used to access the online course text book and all class materials on Canvas.

*** Not allowed: use of the internet to access any other resources**

I. Metabolic Engineering, Mass Balance, Fermentation Stoichiometry (22 points)

In class we discussed how *E. coli* can be engineered to produce a number of higher order biofuels. In one effort led by James Liao (paper posted on canvas, but not required to answer questions, below), the production of different alcohols from amino acid biosynthetic pathways was achieved, as illustrated below. Here we would like to engineer *E. coli* strains that produce 2-methyl-1-butanol and 1-butanol.



a) We'd like to carry out a 10 L fed batch fermentation of an *E. coli* strain that produces 2-methyl-1-butanol. We start by inoculating the media with 100 mL of an OD_{600} = 0.21 *E. coli* culture (OD_{600} of 1 implies 0.47 g cell/L) and feed it glucose and ammonia. After 4 hours of growth, we determine that 200 g of glucose was consumed and 15.0 g of 2-methyl-1-butanol ($C_5H_{12}O$) was produced. Assuming the *E. coli* grew exponentially with a respiratory quotient of 1.40, what is the growth rate constant in hr^{-1}? (Reminder of exponential growth equation: $M_{cells,t} = M_{cells,initial} * e^{kt}$ where *k* is the growth rate constant). **(10 points)**

We have to adjust the moles for Glucose after we find the coefficients for 1 mole glucose

$$C_wH_xO_yN_z + aO_2 + bNH_3 -> cCH_\alpha O_\beta N_\gamma + dCO_2 + eH_2O + fC_jH_kO_lN_m$$

$$C_6H_{12}O_6 + aO_2 + bNH_3 -> cCH_{1.77}O_{0.49}N_{0.24} + dCO_2 + eH_2O + fC_5H_{12}O$$

w=6, x=12, y=6, z=0

$\alpha = 1.77, \beta = 0.49, \gamma = 0.24$

j=5, k=12, l=1, m=0

C: $w = c + d + fj$; 6 = c + d + 5f

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

H: $x + 3b = \alpha c + 2e + fk$; 12+3b=1.77c + 2e + 12f

N: $z + b = \gamma c + fm$; b =.24c

Respiration rate = $\frac{molCO_2}{molO_2}$=1.4 = $\frac{d}{a}$; 1.4a = d

6 Unknowns, c, d, f, e, a, b; need one more for solution.

Substrate = 200g produes 15g of product $Y_{PS} = 15g/200g = .075gg^{-1}$

$MW_{glucose} = 180g/mol$

$MW_{product} = 60 + 12 + 16 = 88g/mol$

$MW_{ecoli} = 12 + 1.77 + 16(.49) + 14(.24) = 25$

200g/180g/mol = 1.11 moles glucose

15g/88g/mol = .17 moles 2-methyl-1-butanol

$$f = Y_{PS} * \frac{MWsubstrate}{MW_{product}} = 0.075gg^{-1} * \frac{180g/mol}{88g/mol} = .153$$

c/f = 1.11/.17 = 6.52 after 4 h.

Rearrange for matrix solution:

C: 6 = 0a + 0b + c + d +0e + 5f

O: 6 = - 2a+ 0b + 0.49c+2d+e+f

H: 12= 0a - 3b + 1.77c + 0d+ 2e + 12f

N: 0 = 0a - b + .24c + 0d + 0e + 0f

RR: 0 = -1.4a + 0b + 0c + d + 0e + 0f

Yps: .153 = 0a + 0b + 0c + 0d + 0e + f

```
x = np.array([[0,0,1,1,0,5],[-2, 0,.49,2,1,1],[0,-3,1.77,0,2,12],
              [0,-1,.24,0,0,0],[-1.4,0,0,1,0,0],[0,0,0,0,0,1]])
x.shape
```

```
(6, 6)
```

```
y = np.array([[6],[6],[12],[0],[0],[.153]])
y.shape

```

```
(6, 1)
```

```
c = np.matmul(np.linalg.inv(x),y)
c.shape
```

```
(6, 1)
```

```
c
```

```
array([[1.11687279],
       [0.88113074],
       [3.67137809],
       [1.56362191],
       [3.1545265 ],
       [0.153     ]])
```

a=1.1

b=.88

c=3.7

d=1.6

e=3.2

f=.153

$$C_6H_{12}O_6 + 1.1O_2 + .88NH_3 -> 3.7CH_{1.77}O_{0.49}N_{0.24} + 1.6CO_2 + 3.2H_2O + .153C_5H_{12}O$$

Now we have to adjust for moles of glucose.

1.11 moles glucose. Multiply by 1.11

a=1.1*1.11=1.2

b=.88*1.11=.98

c=3.7*1.11=4.1

d=1.6*1.11=1.8

e=3.2*1.11=3.6

f=.153=.17

$$1.11C_6H_{12}O_6 + 1.1O_2 + .98NH_3 -> 4.1CH_{1.77}O_{0.49}N_{0.24} + 1.8CO_2 + 3.6H_2O + .17C_5H_{12}O$$

$$M_{cells,t} = M_{cells,initial} * e^{kt}$$

Moles cells initial, how many moles initial ecoli? Assume this Mcells,t is for moles, is this the same as cell mass? OD = cell density.

$(.21).47\frac{gecoli}{L} * 10^{-3}\frac{L}{mL} * 10^{2}ml$=.01g/100ml of ecoli. .01g *1mol/25g=.0004 moles ecoli for initial.

After 4h we used 1.11 moles of glucose so we produced 4.1 moles ecoli.

$$.0004moles = 4.1molese^{k4h}$$

ln(.0004/4.1)=k4

k=ln(.0004/4.1)(.25)=-2.38

b) In a different *E.coli* fermentation attempt,we use a continuous culture system with a feed of 25 g/(L*hr) glucose to produce 1-butanol ($C_4H_{10}O$). From previous experiments, we have determined that even though all the glucose is consumed, only 35% of the energy from glucose goes towards creating biomass. Assume the glucose is degraded aerobically through glycolysis and the TCA cycle. Also, assume each molecule of NADH gives 3 molecules of ATP, each molecule of $FADH_2$ gives 2 molecules of ATP, the $Y_{X/ATP}$ = 4.63 g biomass/mol ATP, and one molecule of GTP is equivalent to one molecule of ATP.

Lets keep the standard mass balance equation then use a separate set of ATP balance equations.

MW 1-butanol ($C_4H_{10}O$) = 48+10+16 = 74 g/mol

$$C_wH_xO_yN_z + aO_2 + bNH_3 -> cCH_\alpha O_\beta N_\gamma + dCO_2 + eH_2O + fC_jH_kO_lN_m$$

$$C_6H_{12}O_6 + aO_2 + bNH_3 -> cCH_{1.77}O_{0.49}N_{0.24} + dCO_2 + eH_2O + fC_4H_{10}O$$

w=6, x=12, y=6, z=0

$\alpha = 1.77, \beta = 0.49, \gamma = 0.24$

j=4, k=20, l=1, m=0

C: $w = c + d + fj$; 6 = c + d + 4f

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

H: $x + 3b = \alpha c + 2e + fk$; 12+3b=1.77c + 2e + 20f

N: $z + b = \gamma c + fm$; b =.24c

6 unknowns, 4 equations. We can probably do this w/electron balance.

Glycolysis produces 38ATP from lecture. From the appendix 4 NADH, 1 GTP, 1 FADH = 15ATP, we get 15ATP form glycolysis. 15/38=.4; ATP is used for both producing biomass and product.

$Y_{XATP} = c * MW(biomass)$ because Yxatp is in units of g/ moles.

c = Yxatp/MW(biomass) = 4.63g/mol/25 g/mol = .19

Redo for matrix:

C: 6 = 0a + 0b + c + d+ 0e + 4f

O: 6= -2a + 0b+ 0.49c+2d+e+f

H: 12= 0a -3b + 1.77c + 0d+ 2e + 20f

N: 0 =0a - b + .24c + 0d + 0e + 0f

.19 = 0a + 0b + c + 0d + 0e + 0f

Still missing 1 equation. Add in part ii of problem.

i) Calculate the amount of biomass produced from the culture in g/(L*hr).

(4 points)

Solve for c using ATP equaions, Convert to g/L*hr from glucose input. Convert glucose input to moles, use c to figure out moles of biomass/moles of glucose, convert back to g/L*hr from moles/L*hr

Glucose $\frac{25g}{L-hr} * \frac{1mole}{180g} = \frac{.14mole}{Lhr}$ input

For 1 mole of glucose c = .19 from above ATP equation balance.

For .14 m/L*hr, biomass moles = .14m/L*hr*.19 = .026m/L*hr

Convert to g/L*hr; .026m/L*hr * 25g/mole = .65 g/L*hr

ii) If we find our yield of 1-butanol to be 3 g/(L*hr), what is the respiratory quotient for the system? **(8 points)**

Yps=3g/L*hr. Convert to moles/L*hr. Odd units. This isn't really a ration with the substrate like the hw because of the units. This looks like f after conerted to moles/L*hr

f = 3g/L*hr * 1mole/74g/mole = .041 moles/L*hr

Solve for d/a.

C: $6 = 0a + 0b + c + d + 0e + 4f$

O: $6 = -2a + 0b + 0.49c + 2d + e + f$

H: $12 = 0a - 3b + 1.77c + 0d + 2e + 20f$

N: $0 = 0a - b + .24c + 0d + 0e + 0f$

$.041 = 0a + 0b + 0c + 0d + 0e + f$

$.026 = 0a + 0b + c + 0d + 0e + 0f$

```
Out[21]: array([[1.11687279],
                [0.88113074],
                [3.67137809],
                [1.56362191],
                [3.1545265 ],
                [0.153     ]])
```

```
In [4]:  1 x = np.array([[0,0,1,1,0,4],[-2, 0,.49,2,1,1],[0,-3,1.77,0,2,20],
         2                [0,-1,.24,0,0,0],[0,0,0,0,0,1],[0,0,1,0,0,0]])
         3 x.shape
```

```
Out[4]: (6, 6)
```

```
In [5]:  1 y = np.array([[6],[6],[12],[0],[.041],[.026]])
         2 y.shape
```

```
Out[5]: (6, 1)
```

```
In [6]:  1 c = np.matmul(np.linalg.inv(x),y)
         2 c.shape
```

```
Out[6]: (6, 1)
```

```
In [7]:  1 c
```

```
Out[7]: array([[5.625045],
               [0.00624 ],
               [0.026   ],
               [5.81    ],
               [5.57635 ],
               [0.041   ]])
```

a = 5.62, d = 5.81

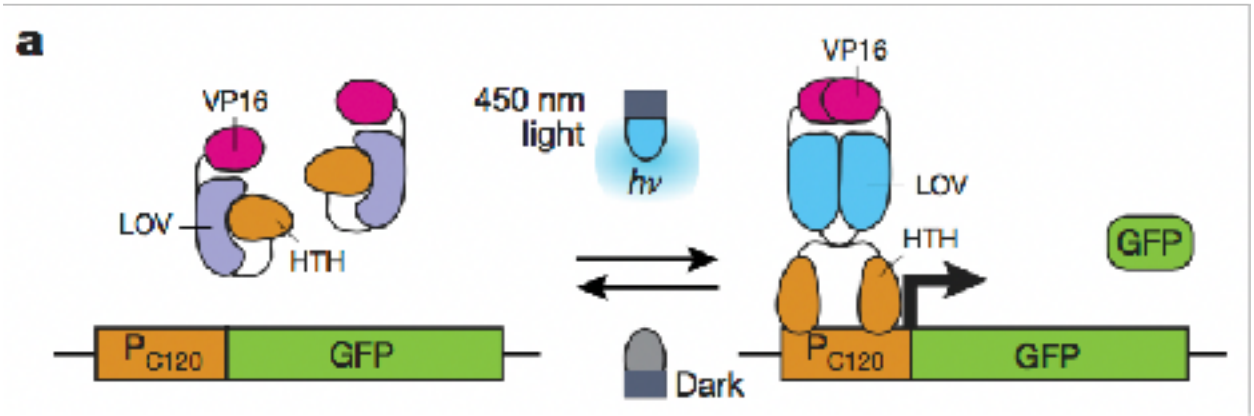
RR = d/a = 5.81/5.62 = 1.033

II. Optogenetic Regulation for Optimized Biofuels Production (23 points)

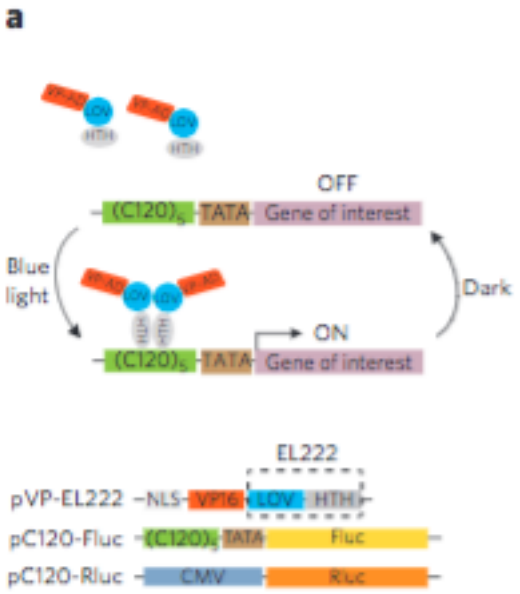
A recent paper from our ChemE colloquium speaker last Monday, Jose Avalos, and his group at Princeton University, describes the use of light to optimize the production of valuable chemicals in a microbial culture (paper posted in canvas). Specifically, the authors use an optogenetic circuit to control the mitochondrial pathways for isobutanol and 2-methyl-1-butanol (2-MBOH) in order to maximize carbon economy by limiting ethanol as a byproduct.

a) Even though light has been used to regulate gene expression in microbial cultures in the past, this paper represents a major breakthrough in the application of this technology to control biofuel production. Avalos *et al*. call their genetic switch to induce biofuel production "OptoEXP", which is illustrated in Figure 1a of their manuscript.

i. Provide a brief description below explaining the role for each of the components of the OptoEXP genetic switch shown in Figure 1a. You are welcome to use the Nature Chemical Biology paper by Motta-Mena *et al.* (posted on Canvas along with the Avalos *et al.* paper). **(3 points)**



From Motta et al:



The Avalos diagram doesn't contain the NLS sequence. Both show dimerization and expression when VP-EL222 is attached to DNA. GOI in Avalos is GFP.

VP-EL222 = VP_AD + (LOV + HTH) where LOV and HTH are part of EL222 from bacteria. Add VP16 and NLS to N-terminal.

LOV:

Blue Light @450nm causes a protein-flavin adduct in LOV domain allowing EL222 to dimerize(2 of these pair up) and attach to DNA. This turns on transcription of the GOI. When blue light is removed the LOV interacts w/HTH disconnecting from the DNA and separating into 2 separate molecules. Time to activate in light <10s and reverse is 50s.

HTH: Helix Turn Helix DNA binding domain. Binds to DNA when LOV reacts to presence of light. Reversible process when light not there.

VP16: Protein added to EL222 to allow EL222 to work in eukaryotes, transcriptional activation domain causes VP16-E222 to be distributed in nucleus and cytoplasm in eukaryotes. Transcription occurs in nucleus in eukaryotes. Add a NLS sequence before the VP16 gene when creating plasmid.

ii. Briefly describe a different induction system that relies on something besides light as an inducer (e.g. one that we discussed in class or in the course reading). **(1 point)**

An environmental signal impacts a transcription factor which is a repressor or activator. Types of environmental include small molecules or heat which cause conformational changes in a protein or small molecule. Example: cI protein changes w/heat from reading. Tet repressor example binds to operator, when Tetracycline is added, binds to Tet Repressor and allows RNA Polymerase to induce gene.

iii. Describe 2 main advantages of using light compared to the example you chose in ii. **(2 points)**

Advantages: nontoxic, inexpensive, compatible w/any C or N source, can be removed/added instantaneously.

iv. What might be a potential problem in using light as an inducer in the context of a microbial culture? **(1 point)**

Toxicity, low transcriptional activation(<20x from dark to light),long deactivation times (>2h). The VP16-EL222 is leaky, so growing multiple generations where you are trying to screen for a population; you may never see what you are screening for if there is a subpopulation which is toxic with a higher cell death rate because this will be overwhelemed by the nontoxic population since they can multiply faster.

b)The authors describe a yeast strain YEZ139 in which a fusion protein VP16-EL222 is produced.

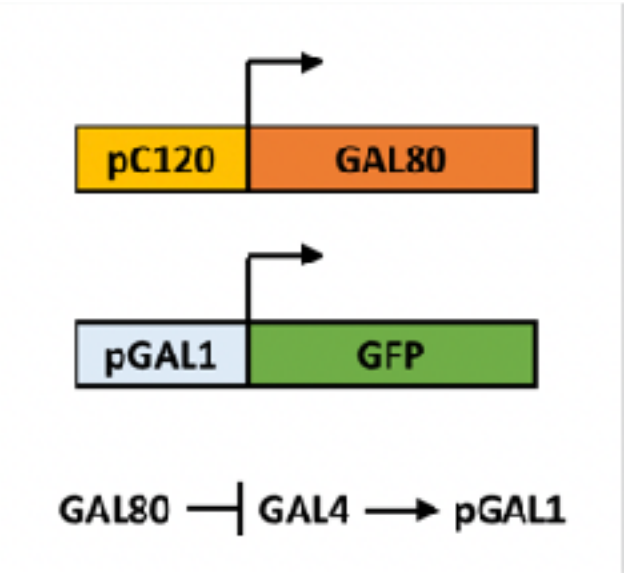
i. What do the authors mean by a "fusion" protein? **(1 point) Combination of 2 proteins VP16 and ELL222 to combine functionality of 2 proteins into 1 by concatenating the DNA sequences for the 2 genes and removing the stop codon for the first gene/protein. Common example is to add a reporter gene like GFP to the end of a gene/ protein. The VP16-EL222 genes were synthesized and bought as a gBlock from IDT.**

ii. Why is the addition of a VP16 domain necessary? **(1 point) The VP16 gene/domain added to get ELL222 to work in yeast cell nucleus where transcription occurs. ELL222 from prokaryote/bacteria.**

iii. The VP16-EL222 fusion protein also includes a nuclear localization signal. Why is this required for proper function of the circuit? **(1 point)**

NLS distributes VP16-EL222 into nucleus and cytoplasm. Transcription in Eularyotes/Yeast works in cell nucleus.

c) The optogenetic circuit, as constructed in yeast, relies on the GAL regulatory network, a well-characterized system commonly used in yeast to induce protein expression through the addition of galactose. In this paper, the authors manipulate GAL4 and GAL80; GAL4 is a transcription factor that drives protein expression from a GAL1 promoter, while GAL80 is a transcription factor that represses expression of GAL4. The authors regulate their circuit by regulating the expression of GAL80 by the C120 promoter as shown schematically below. Please fill in the table below to describe (with “ON” or “OFF”) whether expression will be observed. **(4 points)**


pC120
GAL80
pGAL1
GFP
GAL80 ⊣ GAL4 → pGAL1

	GAL 80 Expression	GFP Expression
Light OFF GAL4 OFF	**NO**	**NO**
Light OFF GAL4 ON	**NO**	**YES**
Light ON GAL4 OFF	**YES**	**NO**
Light ON GAL4 ON	**YES**	**NO**

There are 2 circuits; the OPTOExp and NOT of OPTOExp. When one is on the other is off. When one is off the other is on. The GAL4 circuit is OptoEXPINV.

Initial condition: light ON, no GAL4:

no GAL80, no GFP.

Add light, produces GAL80. No GAL4 so no GFP, GAL80 yes, no GFP

Add GAL4, light ON

light produces GAL80, GAL80 represses GAL4, no PGAL1, no GFP

Light OFF, Add GAL4.

Light off no GAL80, GAL4 makes PGAL1 makes GFP if GAL4 present

Light OFF, no GAL4. light off no GAL80. no GAL4, no pGAL1, no GFP.

d) The authors mention the yeast strain has a deletion in GAL80. Why is this necessary for the circuit to function properly? (1 point)

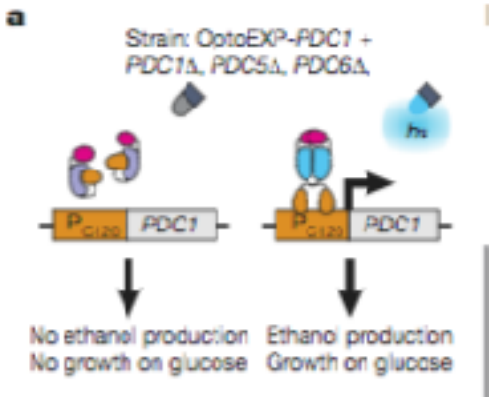
Remove the original wild type GAL80 and GAL4 genes and use the different variants of GAL4 promotor strength to test better light dependent repression.

- e) The GAL4 used in the optogenetic circuit was fused to a photosensitive degron (PSD) domain, which causes proteases to more actively target the fusion protein. How might this function to induce faster and more complete light-dependent repression of gene expression? Why is it useful for the broader biofuels objectives of the paper? **(2 points)**

Increases protein stability. Better repression makes gene less leaky and increases dynamic range of switch, allows higher production when metabolism is switched between growth and production.

f) To apply their optogenetic circuit to biofuels production, the authors place a pyruvate decarboxylase (PDC) gene under the control of the C120 promoter. Briefly explain how this allowed greater production of isobutanol and 2-MBOH relative to ethanol. Please include why deleting the PDC genes is not a viable strategy to minimize ethanol production by cells, and how the authors used the optogenetic circuit to circumvent this issue. **(3 points)**

Switch between cell growth and production of product. Deleting the PDC gene does not allow the cell to grow. The PDC genes are needed for NAD+ recycling for glycolysis. They put the PDC genes under control of light, the Pc120 promoter and the VPP16-EFF as a repressor when there is no light.



g) One key finding of the paper is that periodic pulses of light following the growth phase greatly improved biofuel yields. Why did these pulses of light help cell productivity? How much higher was their titers of isobutanol and 2-MBOH, respectively, than previous literature? **(2 points)**

At a yield of 122+-.11g/L of isobutanol, which is toxic to the cell, the host was unable to consme all the glucose indicating stalled fermentation and stalled metabolic arrest. The theory was NAD+ depletion; change to light pulses(30 minutes blue light in duty cycle of 25s on and 65s off) instead of continuous light every 10h to get 3.37g/L isobutanol and 433+-69mg/L 2-MBOH over 80h fermentation cycle. This prevents the toxicity of products causing the metabolic arrest.

Additional yield obtained by BAT1 deletion which is a competing metabolic pathway for alpha-ketoisovalerate precursor.

Isobutane improved 5x, 2-MBOH 20x

h) In their PDC-induction strategy, how were the authors also able to overcome the possible pitfall (from part a above) of using light as an inducer? **(1 point)**

Problem of low yield attributable to multiple factors including metabolic arrest from depletion of NAD+. Tested different light schedules to turn on PDC1 expression to increase NAD+.

III. Gene Regulation and Gene editing (10 points)

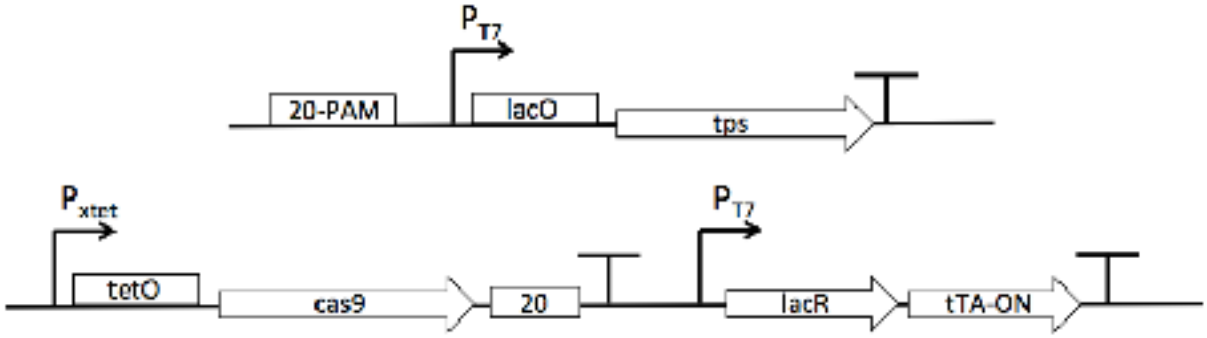
We spent a lot of time this quarter thinking about how to get DNA into an organism. However, release of genetically engineered microbes (and engineered DNA) into the environment is a major concern and several groups are working to develop systems that "self-deconstruct", e.g. after an E coli fermentation to produce a drug is complete. The following excerpt from a recent paper out of the Voigt lab (Nature Communications 2015; paper posted in canvas but not required to answer the question) provides motivation for triggered ways to degrade DNA:

"The ability to programme cells to eliminate engineered DNA at a defined time point or change in environments would benefit many applications in biotechnology. For example, after bio-manufacturing a chemical, cells could be programmed to degrade their DNA at the end of the process or when they are removed from a defined medium. This would aid the protection of sequence information as a trade secret, make it easier to remove DNA contamination from a product, reduce the cost of biomass disposal and decrease the amount of DNA in the environment after an accidental release. There are similar needs for 'out of the bioreactor' applications, such as using engineered cells as living therapeutics (for example, in the gut microbiome) or in forming associations with crop plants in a field[1]. In these cases, it is impossible to recollect cells for disposal, so they need to be programmed to degrade their own DNA when they leave a defined environment or after a defined time period.

Various genetic switches have been developed that induce cell death[2]... Inducing cell death does not address the problem of the release of DNA, which persists after cells die or are killed. Indeed, waste streams from fermenters are rich with recombinant DNA, even when the cells have been inactivated by heat, pH and

antibiotics[10] and, in fact, these methods of rapid cell death exacerbate the release of extracellular DNA[11]. The waste biomass of engineered microbes is often used as agricultural fertilizer (for example, NovoGro), and this has been shown to contain significant amounts of DNA[10]. Furthermore, DNA molecules are stable and, once introduced, plasmid and genomic DNA can be recovered from environmental samples via PCR for 1–5 months[11, 12] and is likely to be detectable longer with advances in deep sequencing[13]. In one study, 35% of plasmid DNA molecules that were exposed to the extreme heat and pressures of atmospheric re-entry on the surface of a rocket still retained their biological function[14].”

The approach described in this paper involves the use of CRISPR to degrade plasmid DNA after use. Using the list of genetic parts below we have designed the following constructs to put in E coli to test this system:



Element	Function
Genes	
tps: Terpene synthase	Codes for an enzyme that produces the terpene product
cas9: Nuclease	Codes for the nuclease that cleaves DNA recognized by guide RNA
lacR: Lac repressor gene	Lac repressor binds to the lac operator sequence (*lacO*), preventing transcription from any upstream promoter; IPTG binding causes lac repressor to dissociate from the lac operator
20-PAM: Recognition sequence	DNA sequence that is targeted by guide RNA for CRISPR-catalyzed cleavage
20: Guide sequence	Codes for guide RNA that anneals to target DNA sequence for CRISPR-catalyzed cleavage
tTA-On: Tet-On activator gene	Tet-On activator binds to the tet operator sequence (*tetO*), activating transcription from P_{xtet}; however, Tet- On binds to the operator sequence *only if* bound to aTc
Promoters, terminator, and RBS	

P_{T7}: T7 promoter	Constitutive promoter; always “on,” unless activity blocked by a repressor protein bound downstream
P_{xtet}: Hybrid tet promoter	Is “on” only if a Tet activator is bound to the *tetO* operator sequence; can also be blocked by repressor
T: Terminator	Generic terminator that terminates transcription for any gene or operon
RBS: Ribosome binding site	Required to initiate translation

Operators	
lacO: Lac operator	Bound by lac repressor
tetO: Tet operator	Bound by either Tet activator (Tet-On or Tet-Off)
Small molecules	
IPTG: Isopropyl β-D-1-thiogalactopyranoside	Binds to Lac repressor, causing it to dissociate from the lac operator; diffuses freely into the cell, but is not degraded
aTc: Anhydrotetracycline	Binds to both Tet-On and Tet-Off activator genes; diffuses freely into the cell but is not degraded

a) Add RBSs where needed in the diagram above using asterisks. **(2points)**

RBSs are only present after the gene is transcripted to mRNA then translated to an AA sequence. For translation to protein from mRNA.

*** in front of all the genes which code for translation: tps, cas9, lacR, tTA-ON.**

PAM isn't a gene. 20 creates the guide RNA assume the guide RNA gene there only fro transcription. But do you need a RBS to get the gRNA to bind to CAS9? I say no for now.

b) Fill in the following table with the inputs required to get the desired result.Assume that CRISPR-catalyzed cleavage at *PAM-20* results in the degradation of the *tps* gene as well. **(8 points)**

Input		**Output**	
ATc	**IPTG**	**Terpene present?**	***tps* gene present?**
+	**–**	**NO**	**NO**
+	**+**	**(NO), starts out as yes before tps gene removed.**	**NO**
–	**+**	**YES**	**YES**
–	**–**	**NO**	**YES**

Initial condition No ATc, No IPTG. T7 is always on, produces lacR which binds to lacO. Tet-On present in system. Does not bind to tetO since no ATc. No terpene because lacR attached to lacO there is no IPTG. No terpene, tps gene present.

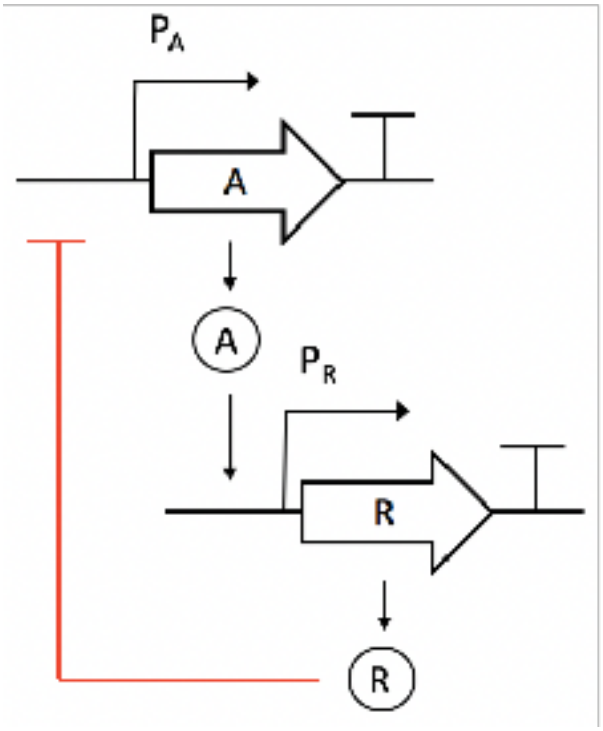
Add IPTG: removes lacR from lacO allowing production of terpene. Still no ATc. Crispr is off. IPGT present, no Atc, terpene and tps YES.

Add Atc to path above which is producing terpene. Atc turns binds to Tet-On binding to tetO which tursn on Cripsr which removes tps. Was producing terpene but removes tps which then removes terpene. ATC and IPTG has terpene going from Yes->No and TPS being removed.

Alternate path from Initial condition: add ATc, don't add IPTG. No terpene bc lacR is on lacO. ATc attaches to Tet-On which attaches to Atc and attaches to tetO which turns on cas9 removing tps gene. No terpene, not tps gene. Have ATC, no IPTG

IV. Modeling Output of Gene Expression and Feedback Regulation (8 points)

In biological systems, sometimes the protein product of a gene can inhibit that gene's expression (directly or indirectly); this process is known as negative feedback. You can imagine when such a system might be useful; for example, to avoid overproduction of a particular metabolite. Such a system is represented by the illustration shown below, where the protein R inhibits production of transcription factor A, which in turn regulates the production of protein R. Write a pair of differential equations that describe how the concentration of A changes over time, and how the concentration of R changes with time. Please provide a clear rational for each term in your equations. **(8 points)**



First lets consider both A(t) and R(t) as concentrations of A and R being produced as an output given by exponential equations
$A(t) = k_a e^{\alpha t}; and R(t) = k_r e^{\beta t}$ if both were independent genes.

In the case of feedback A(t) is connected to the input of R(t) and the output of R(t) is connected to the input of A(t).

Consider the effect of repressor protein/transcription factor in A(t) which would decrease the concentration of A as the concentration of the Repressor got higher. Make the repressor a constant level first.

$A(t) = k_a e^{\alpha t} - Repressor$; the repressor is not a constant but a growing amount starting from t=0; and is represented as the output of a gene.

We can replace A(t), R(t) w derivatives because they include constants

$$\frac{dA(t)}{dt} = k_a e^{\alpha t} - k_r e^{\beta t}$$

There is no feedback on the second gene R(t)

$$\frac{R(t)}{dt} = k_r e^{\beta t}$$

V. Production of Silk in *Escherichia coli* (15 points)

Spider silk is a protein made up of modules of repeating amino acids (see figure below adapted from an informative review in *Science*, 2010). Silk has excellent material properties, but the supply is limited since, at present, the only source is the cultivation of silkworms. Several groups have been working to engineer the production of silk protein (or at least some of the silk protein modules) into alternative production hosts.

Please read the paper posted on Canvas describing efforts to produce dragline silk in *E. coli* (Sang Yup Lee et al, PNAS 2010, 14059), and answer the following questions about this engineered biological system:

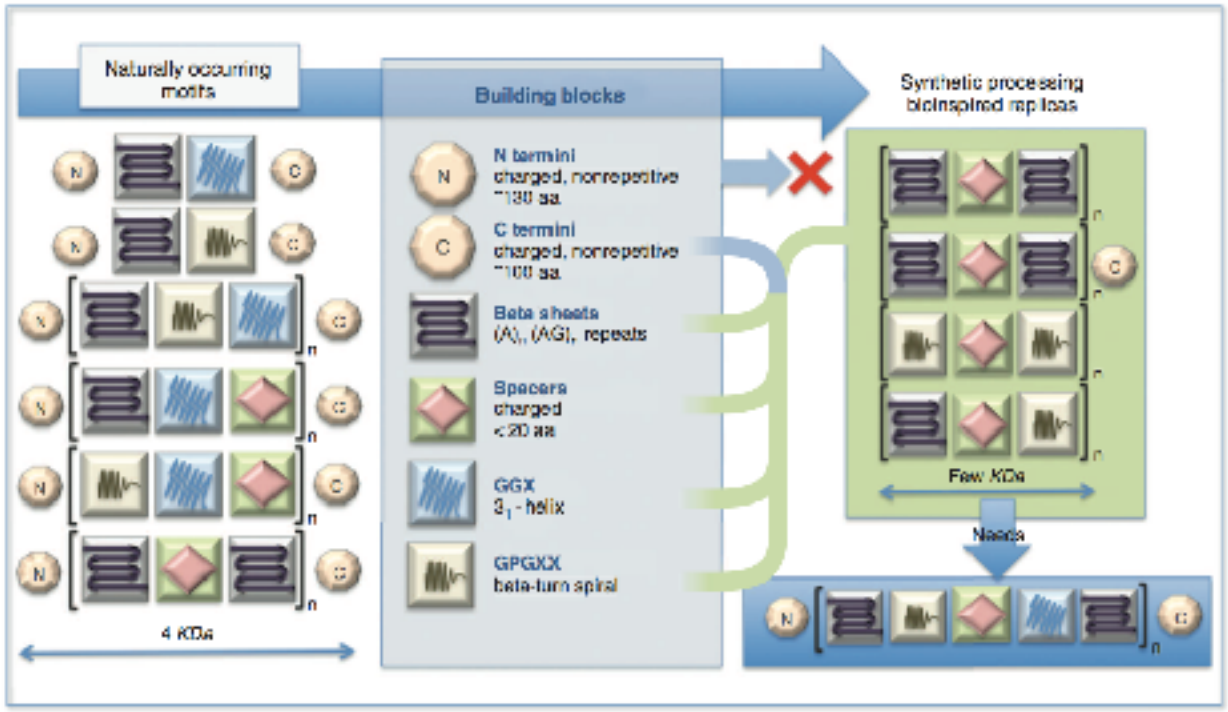


Fig. 2. Modular designs of silk proteins. Silks are fibrous proteins and are characterized by modular units linked together to form high molecular weight, highly repetitive proteins. These modular units or domains, each with specific amino acid sequences and chemistries, provide specific functions. In particular, sequence motifs such as polyalanine (polyA) and poly alanine-glycine (polyAG) (β sheet–forming), GXX (31-helix), GXG (stiffness), and GPGXX (β spiral) are key components in different silks whose relative positioning and arrangement are intimately tied with the end material properties. These domains are linked together to generate high molecular weights and also include characteristic and highly conserved N and C termini. Charged amino acids are strategically located at the chain ends and in spacers to optimize water interactions related to processing and assembly. Current modes of expression of silks in heterologous hosts are generally confined to a limited number of these modular units and lack the inclusion of both N- and C-terminal domains. The incorporation of all of these key domains, along with issues of size (number of modular units), will facilitate improved material properties from silks generated by recombinant DNA techniques.

a) In addition to the obvious advantage of using an easy-to-cultivate host to make silk, how could a heterologous host be used to enhance the properties of silk products? **(2 points)**

Can change the cDNA sequence for additonal properties. i.e. make longer molecular weights than what is possible in nature which correspond to longer chains and may give better product than what is availalbe w/wild type DNA. The paper replicates the cDNA n times. Can adjust n to higher values than what is observed in wildtype MaSp1/MaSP2.

b) Where do the authors get the DNA that encodes silk variants that they use to engineer *E. coli*? **(1 point)**

cDNA they obtained from 3rd party publication (3) in paper references. The cDNA was sequenced and probably publicly available. cDNA is reverse transcripted from spider silk protein of Nephila clavipes. Spider drag silk= protein MaSP1 + MaSP2. They synthetically generated genes which matched the cDNA sequence. They repeated the gene n times and expressed it in a plasmid.

c) What challenges might a repetitive sequence introduced into the processes of:

i. in vitro gene synthesis **(2 points)**
Difficult to fold. Longer chains are even more difficult to fold.

ii. in vivo gene stability **(2 points)**
Foreign gene, could be toxic to host. Difficult to fold; they added His Tag and S Tag near the start sequence vs. the end; probably as a result of folding dynamics. Normal protocol is to insert tags near terminator.

iii. mRNA translation **(2 points)**
A repetitive sequence uses the same codons and you have chance of depleting the codons since you are using the same ones in a repetitive sequecne.

d) What engineering strategies were used to address the challenges associated with translating a repetitive sequence? **(2 points)**

Codon optimization, solved for codon shortage problem by creating tRNAs via glyA overexpression.

e) In this effort, due to the high stability of silk, researchers were able to isolate it directly by extraction and precipitation. What are two other ways we discussed this quarter that can be used to purify a protein out of a crude extract? **(2 points)**

1) FPLC
2) Affinity Chromatography, prob this is not a separate answer. Chromatography in general counts as 1 method.

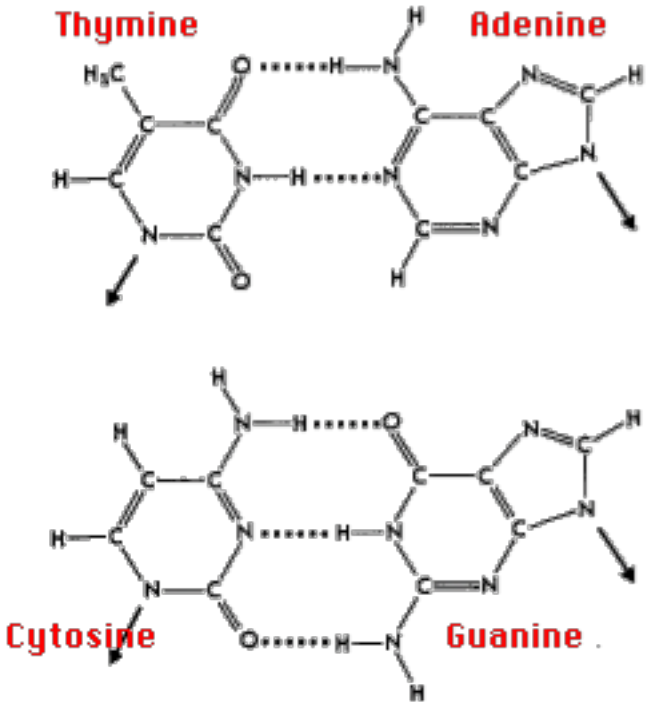
2) Gel separating protein sizes then remove the gel block which contains the size of interest. Add SDS/detergent. Assumes you can separate by size.

Antibody separation.

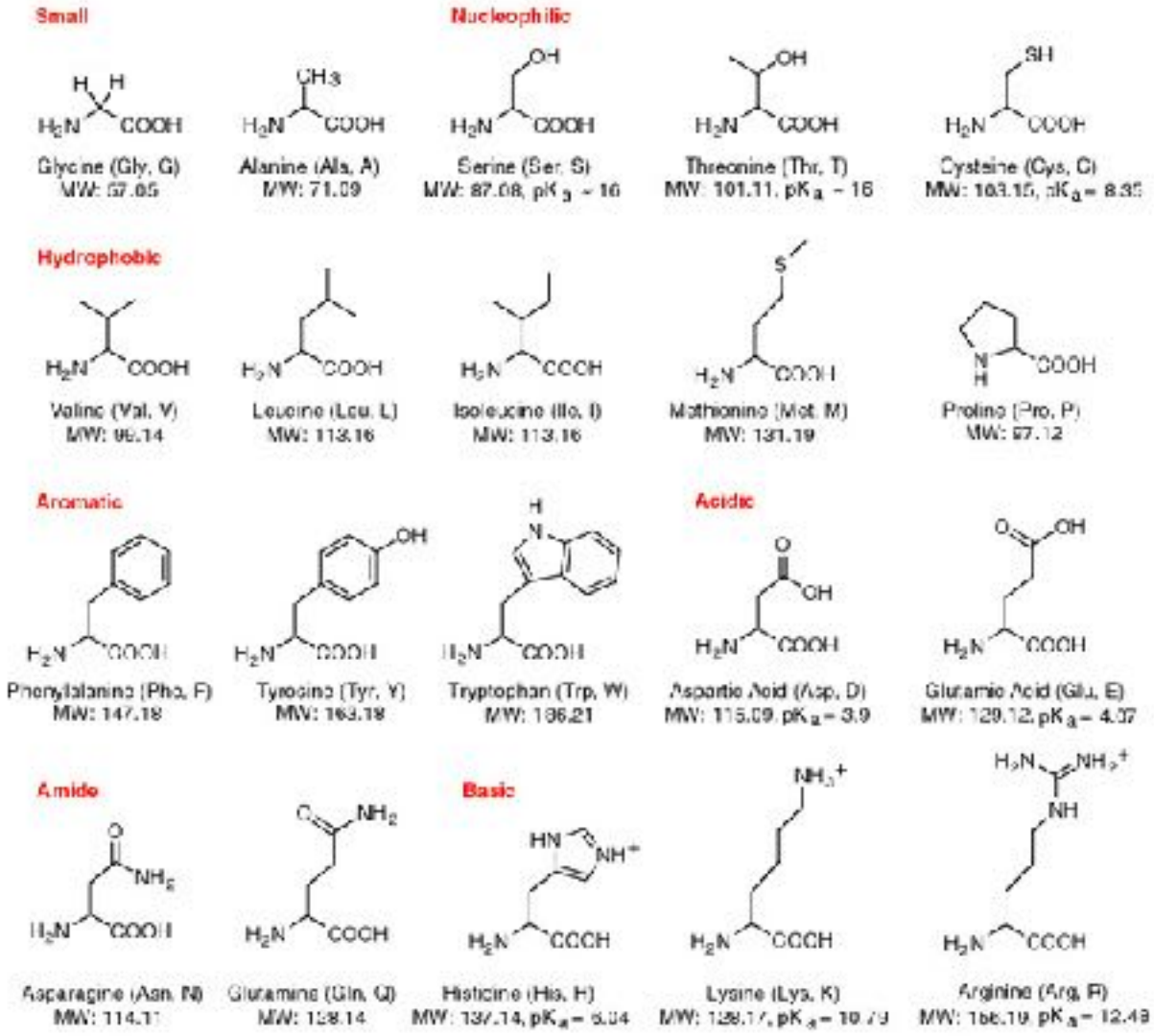
Appendix

The standard genetic code: DNA base pairing: DNA base pairing:

	U	C	A	G	
U	UUU, UUC: Phe - F UUA, UUG: Leu - L	UCU, UCC, UCA, UCG: Ser - S	UAU, UAC: Tyr - Y UAA: stop UAG: stop	UGU, UGC: Cys - C UGA: stop UGG: Trp - W	U C A G
C	CUU, CUC, CUA, CUG: Leu - L	CCU, CCC, CCA, CCG: Pro - P	CAU, CAC: His - H CAA, CAG: Gln - Q	CGU, CGC, CGA, CGG: Arg - R	U C A G
A	AUU, AUC, AUA: Ile - I AUG: Met - M start	ACU, ACC, ACA, ACG: Thr - T	AAU, AAC: Asn - N AAA, AAG: Lys - K	AGU, AGC: Ser - S AGA, AGG: Arg - R	U C A G
G	GUU, GUC, GUA, GUG: Val - V	GCU, GCC, GCA, GCG: Ala - A	GAU, GAC: Asp - D GAA, GAG: Glu - E	GGU, GGC, GGA, GGG: Gly - G	U C A G



Amino acid structures:


Small
Nucleophilic
Glycine (Gly, G)
MW: 57.05
Alanine (Ala, A)
MW: 71.09
Serine (Ser, S)
MW: 87.08, pK a ~ 16
Threonine (Thr, T)
MW: 101.11, pK a ~ 16
Cysteine (Cys, C)
MW: 103.15, pK a = 8.35
Hydrophobic
Valine (Val, V)
MW: 99.14
Leucine (Leu, L)
MW: 113.16
Isoleucine (Ile, I)
MW: 113.16
Methionine (Met, M)
MW: 131.19
Proline (Pro, P)
MW: 97.12
Aromatic
Acidic
Phenylalanine (Phe, F)
MW: 147.18
Tyrosine (Tyr, Y)
MW: 163.18
Tryptophan (Trp, W)
MW: 186.21
Aspartic Acid (Asp, D)
MW: 115.09, pK a = 3.9
Glutamic Acid (Glu, E)
MW: 129.12, pK a = 4.07
Amide
Basic
Asparagine (Asn, N)
MW: 114.11
Glutamine (Gln, Q)
MW: 128.14
Histidine (His, H)
MW: 137.14, pK a = 6.04
Lysine (Lys, K)
MW: 128.17, pK a = 10.79
Arginine (Arg, R)
MW: 156.19, pK a = 12.48

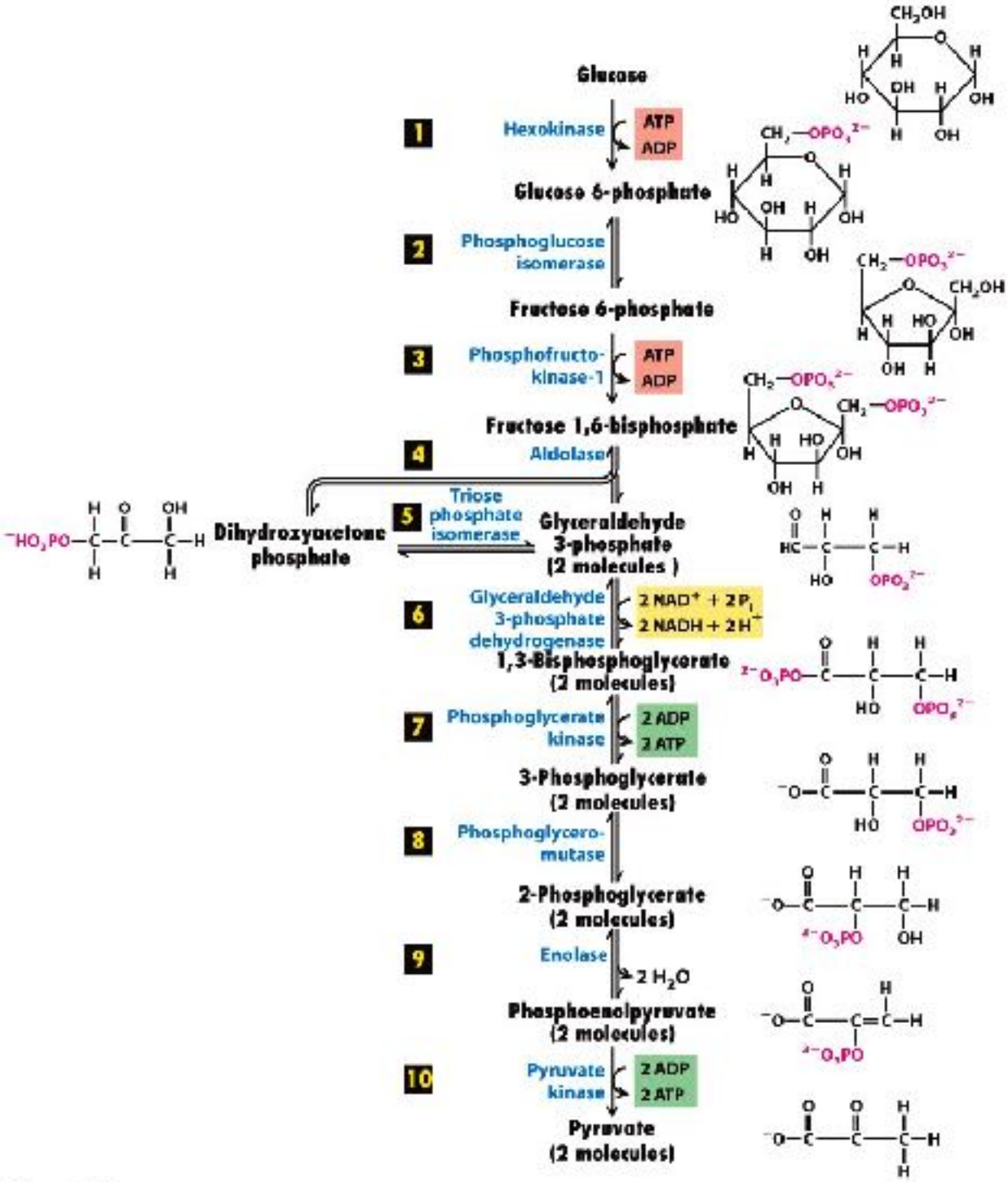
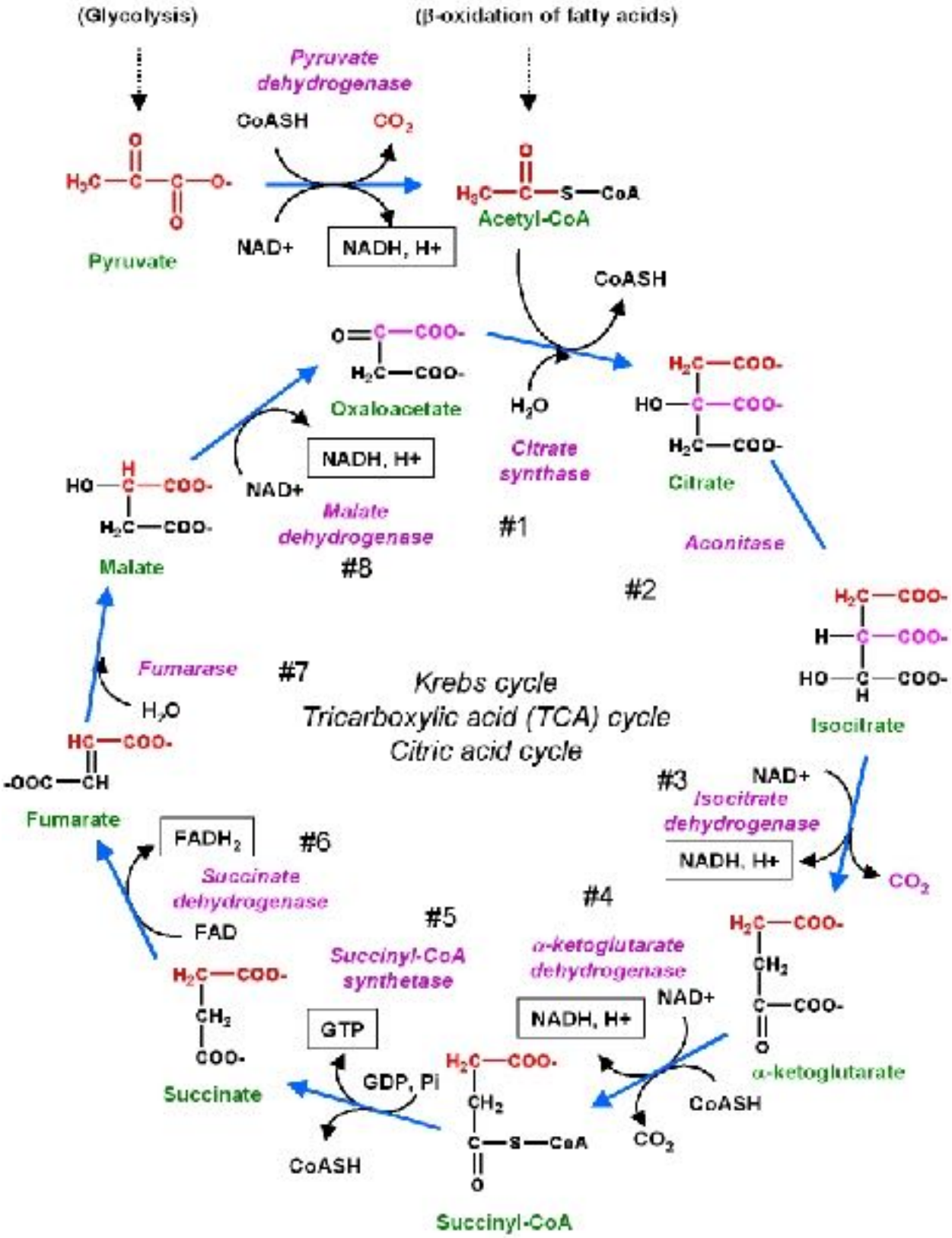


Figure 12-3
Molecular Cell Biology, Sixth Edition


(Glycolysis)
(β-oxidation of fatty acids)
Pyruvate dehydrogenase
CoASH
CO_2
Pyruvate
NAD+
NADH, H+
Acetyl-CoA
CoASH
Oxaloacetate
H_2O
Citrate synthase
Citrate
NADH, H+
NAD+
Malate dehydrogenase
#1
Aconitase
Malate
#8
#2
Isocitrate
Fumarase
#7
H_2O
Krebs cycle
Tricarboxylic acid (TCA) cycle
Citric acid cycle
#3
NAD+
Isocitrate dehydrogenase
NADH, H+
CO_2
Fumarate
$FADH_2$
#6
Succinate dehydrogenase
FAD
#4
#5
α-ketoglutarate dehydrogenase
Succinyl-CoA synthetase
NAD+
NADH, H+
GTP
GDP, Pi
Succinate
α-ketoglutarate
CoASH
CO_2
CoASH
Succinyl-CoA