**Chemical Engineering 355**
**Advanced Biochemical Engineering**
**Spring 2018**

*FINAL Exam*

**NAME:** _____

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 11$^{th}$ 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.**

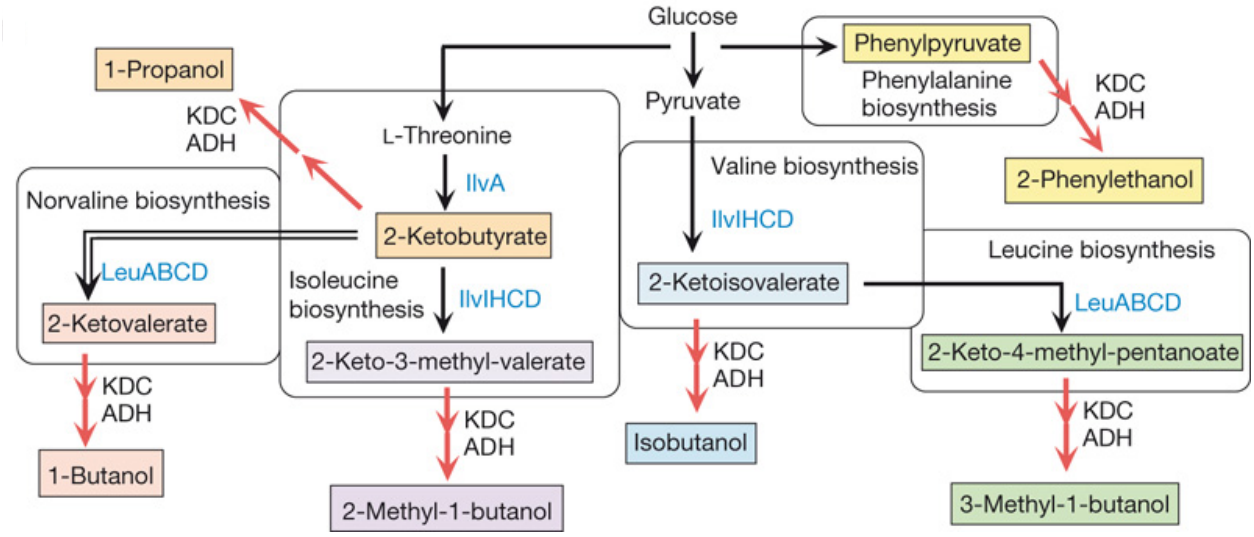
**You can use a calculator and refer to any in-class or other personally written notes. A computer or laptop can be used to access the online course text book and all class materials on Canvas.**

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

| | | |
|---|---|---|
| **1. Metabolic Engineering, Mass Balance, Fermentation Stoichiometry** | **22 pts** | _____ |
| **2. Optogenetic Regulation for Optimized Biofuels Production from Microbes** | **23 pts** | _____ |
| **3. Gene Regulation and Gene editing** | **10 pts** | _____ |
| **4. Modeling Output of Gene Expression and Feedback Regulation** | **8 pts** | _____ |
| **5. Production of Silk in *E. coli*** | **15 pts** | _____ |
| **Total** | **78 pts** | _____ |

## 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)**

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.

i) Calculate the amount of biomass produced from the culture in g/(L*hr). **(4 points)**

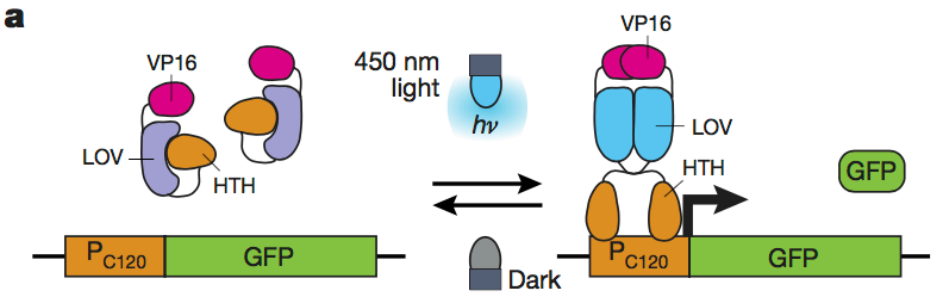
ii) If we find our yield of 1-butanol to be 3 g/(L*hr), what is the respiratory quotient for the system? **(8 points)**

## 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)**




*LOV:*

*HTH:*

*VP16:*

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)**

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

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

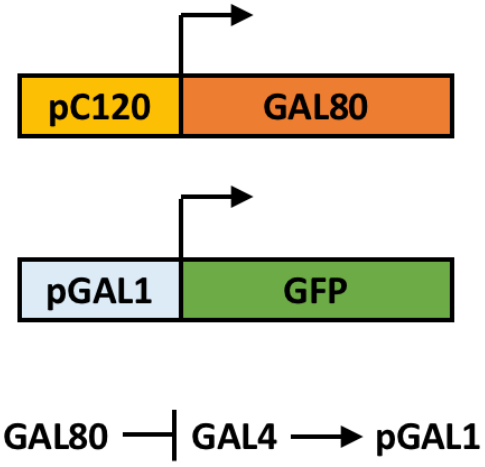
b) The authors describe a yeast strain YEZ139 in which a fusion protein VP16-EL222 is produced.

i. What do the authors mean by a “fusion” protein? **(1 point)**

ii. Why is the addition of a VP16 domain necessary? **(1 point)**

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

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)**




| | GAL80 expression | GFP expression |
|---|---|---|
| **Light OFF, GAL4 OFF** | | |
| **Light OFF, GAL4 ON** | | |
| **Light ON, GAL4 OFF** | | |
| **Light ON, GAL4 ON** | | |

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

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)**

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)**

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)**

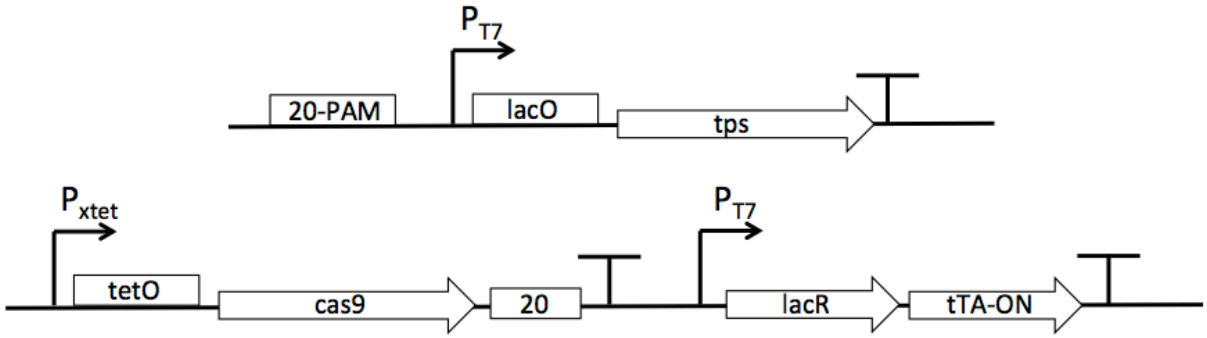
h) In their PDC-induction strategy, how were the authors also able to overcome the possible pitfall (from part a above) of using light as an inducer? **(1 point)**

## 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 |

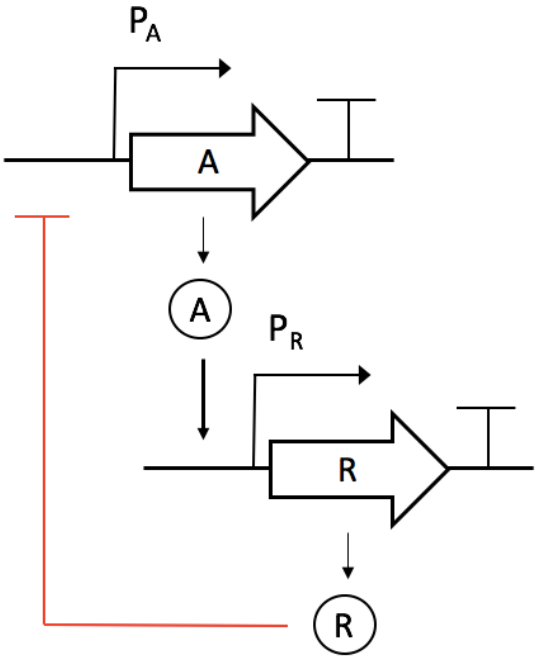
a) Add RBSs where needed in the diagram above using asterisks. **(2 points)**

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?** |
| **+** | **–** | | |
| **+** | **+** | | |
| **–** | **+** | | |
| **–** | **–** | | |

## 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)**



## 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.

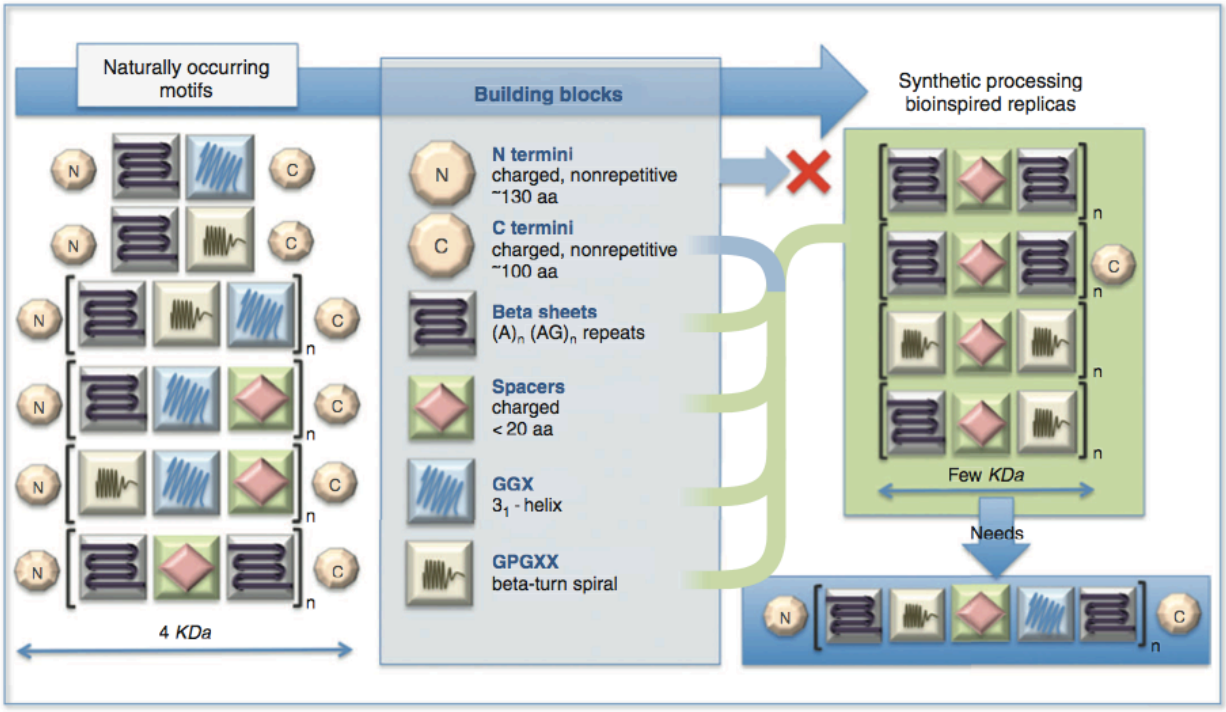



**Fig. 2.** Modular designs of silk proteins. Silks are fibrous proteins and are characterized by modular units linked together to form high molecular weight, highly repetitive proteins. These modular units or domains, each with specific amino acid sequences and chemistries, provide specific functions. 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.

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:

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)**

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

c) What challenges might a repetitive sequence introduce into the processes of:
   i. in vitro gene synthesis **(2 points)**

   ii. in vivo gene stability **(2 points)**

   iii. mRNA translation **(2 points)**

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

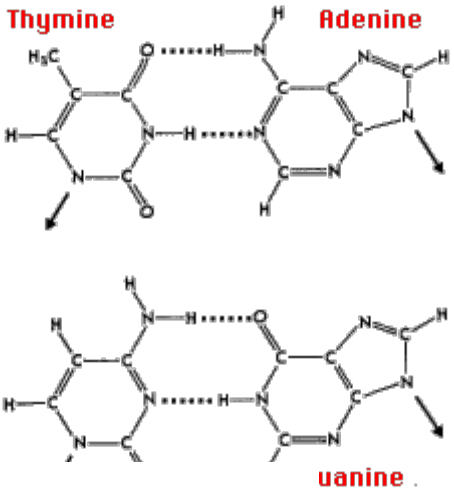
e) In this effort, due to the high stability of silk, researchers were able to isolate it directly by extraction and precipitation. What are two other ways we discussed this quarter that can be used to purify a protein out of a crude extract? **(2 points)**

## Appendix

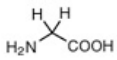
The standard genetic code:

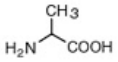
| | U | C | A | G | |
|---|---|---|---|---|---|
| U | UUU, UUC } Phe - F | UCU, UCC, UCA, UCG } Ser - S | UAU, UAC } Tyr - Y | UGU, UGC } Cys - C | U |
| | UUA, UUG } Leu - L | | UAA stop | UGA stop | C |
| | | | UAG stop | UGG } Trp - W | A |
| | | | | | G |
| C | CUU, CUC, CUA, CUG } Leu - L | CCU, CCC, CCA, CCG } Pro - P | CAU, CAC } His - H | CGU, CGC, CGA, CGG } Arg - R | U |
| | | | CAA, CAG } Gln - Q | | C |
| | | | | | A |
| | | | | | G |
| A | AUU, AUC, AUA } Ile - I | ACU, ACC, ACA } Thr - T | AAU, AAC } Asn - N | AGU, AGC } Ser - S | U |
| | AUG Met - M start | | AAA | AGA | C |
| G | GUU, GUC, GUA, GUG } Val - V | | | | |

DNA base pairing:

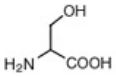


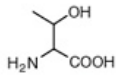

**Small**

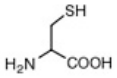

Glycine (Gly, G)
MW: 57.05


Alanine (Ala, A)
MW: 71.09

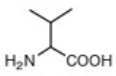
**Nucleophilic**

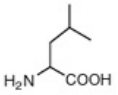

Serine (Ser, S)
MW: 87.08, $pK_a$ ~ 16

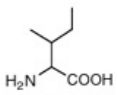

Threonine (Thr, T)
MW: 101.11, $pK_a$ ~ 16

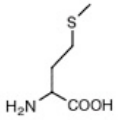

Cysteine (Cys, C)
MW: 103.15, $pK_a$ = 8.35

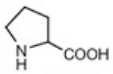
**Hydrophobic**


Valine (Val, V)
MW: 99.14

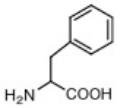

Leucine (Leu, L)
MW: 113.16

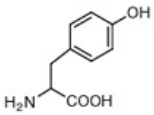

Isoleucine (Ile, I)
MW: 113.16

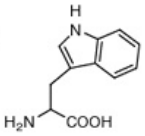

Methionine (Met, M)
MW: 131.19


Proline (Pro, P)
MW: 97.12

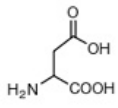
**Aromatic**

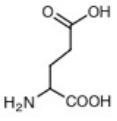

Phenylalanine (Phe, F)
MW: 147.18


Tyrosine (Tyr, Y)
MW: 163.18

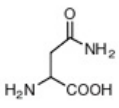

Tryptophan (Trp, W)
MW: 186.21

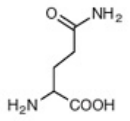
**Acidic**


Aspartic Acid (Asp, D)
MW: 115.09, $pK_a$ = 3.9

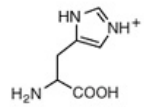

Glutamic Acid (Glu, E)
MW: 129.12, $pK_a$ = 4.07

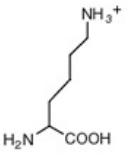
**Amide**

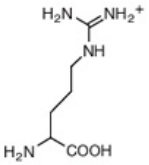

Asparagine (Asn, N)
MW: 114.11


Glutamine (Gln, Q)
MW: 128.14

**Basic**


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

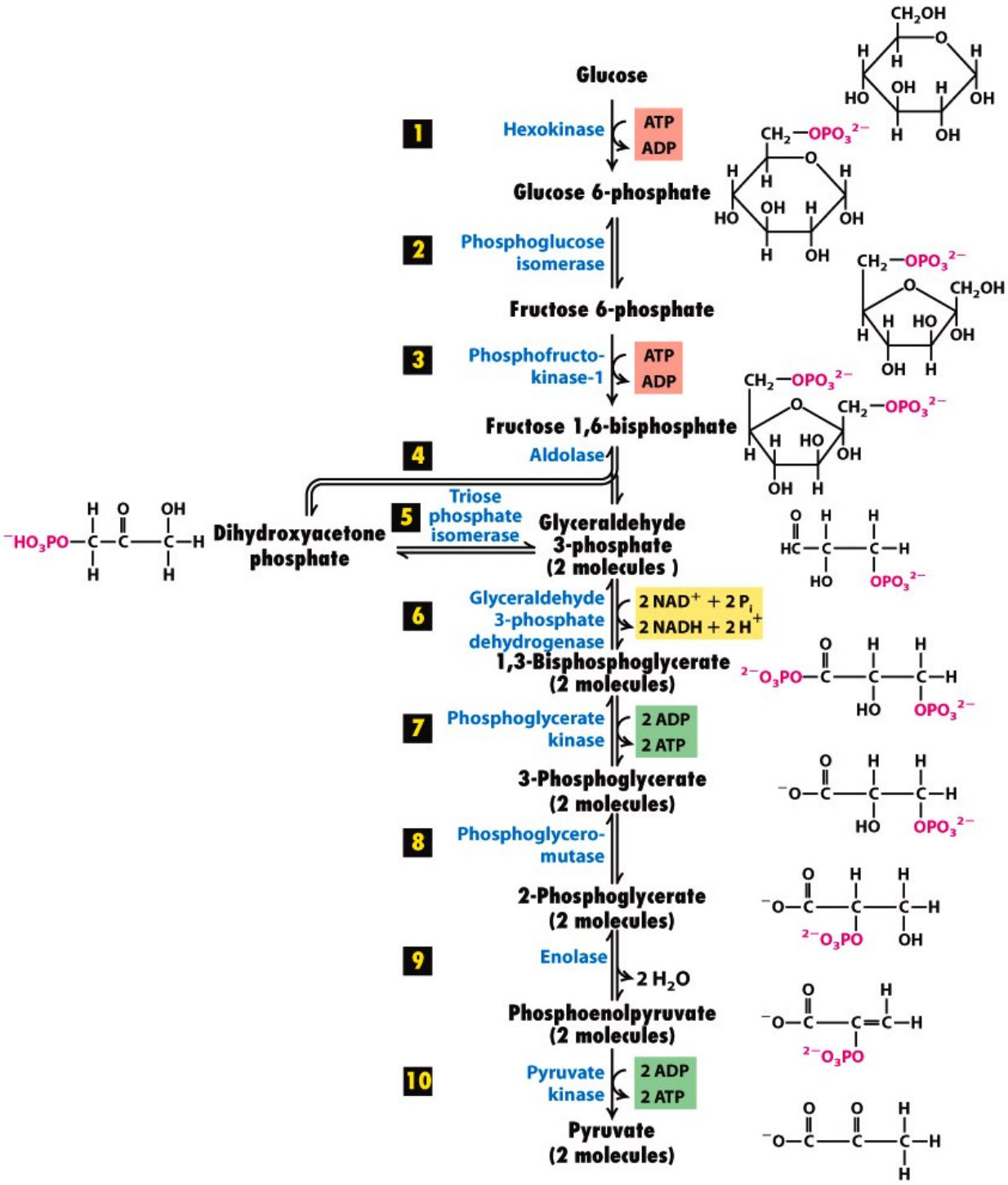
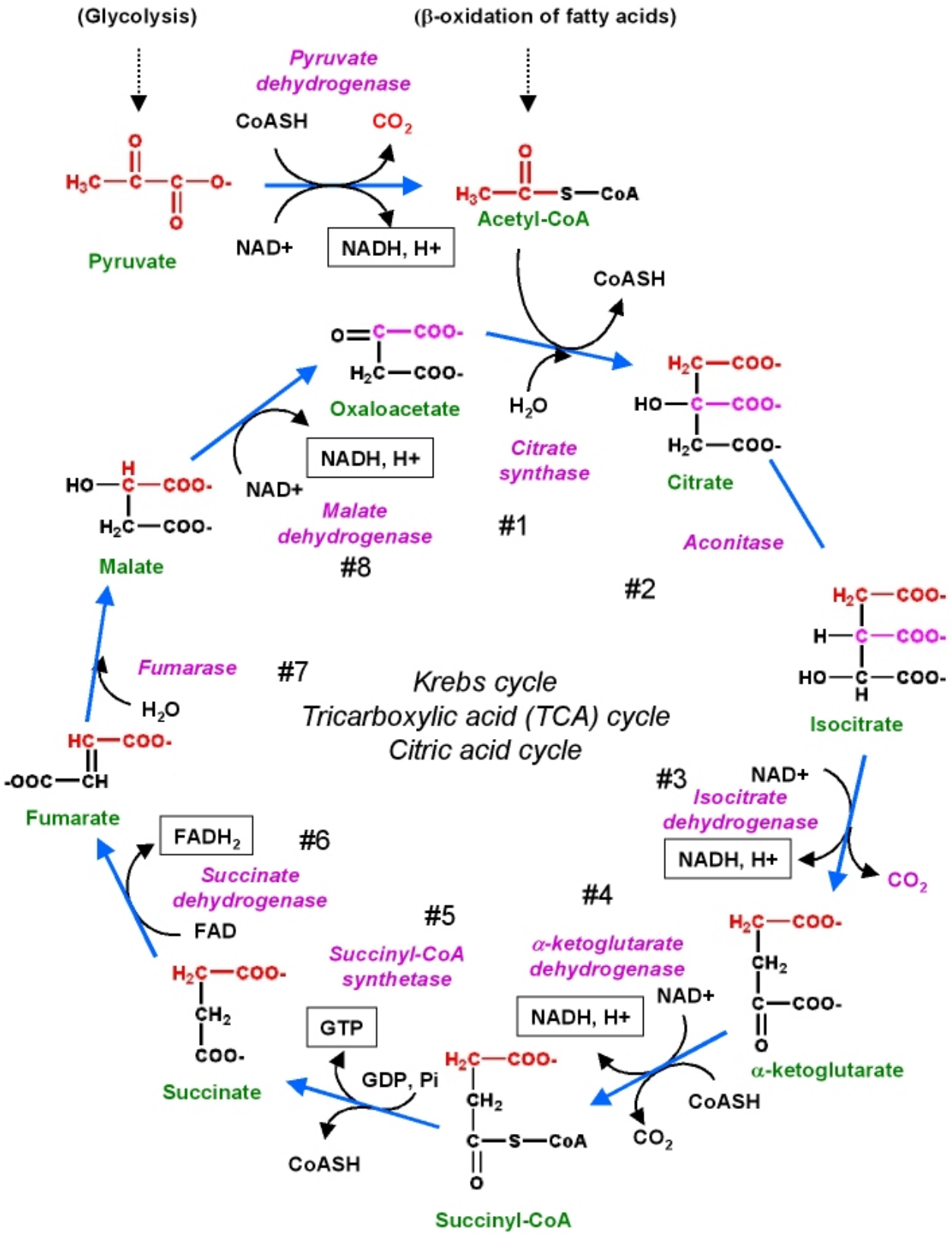



**Figure 12-3**
***Molecular Cell Biology, Sixth Edition***


(Glycolysis)
(β-oxidation of fatty acids)
Pyruvate dehydrogenase
CoASH
$CO_2$
NAD+
NADH, H+
Pyruvate
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
Succinyl-CoA synthetase
α-ketoglutarate dehydrogenase
NAD+
NADH, H+
GTP
GDP, Pi
Succinate
CoASH
$CO_2$
α-ketoglutarate
CoASH
Succinyl-CoA