ANSWER KEY

5.08 Biological Chemistry II (Spring 2016)
Problem Set #5 and #6

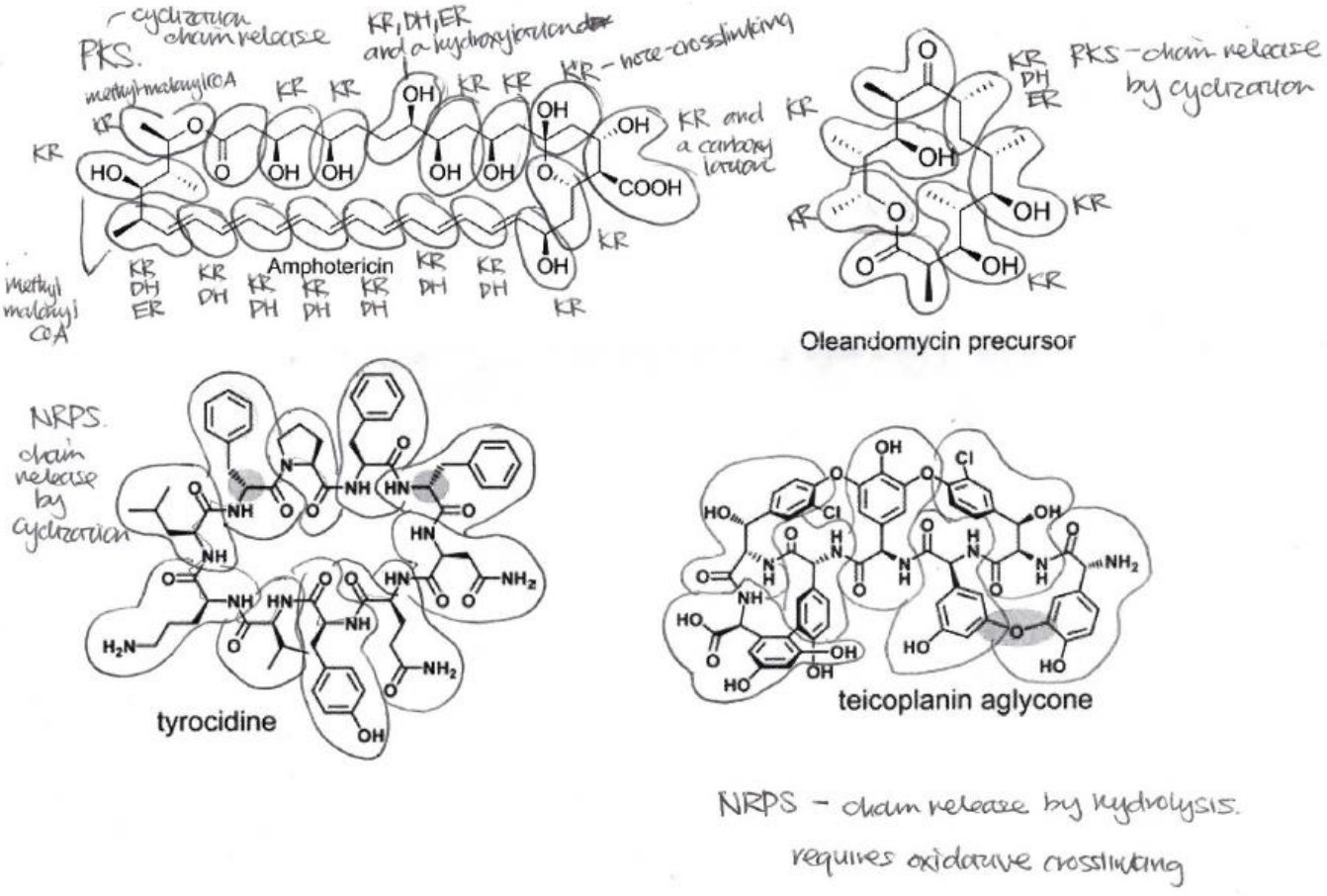
This problem set contains four required problems and six pages.

Problem 1:

We have looked at a number of natural products that are biosynthesized by PKS and NRPS machinery in class. For each natural product shown below, indicate:

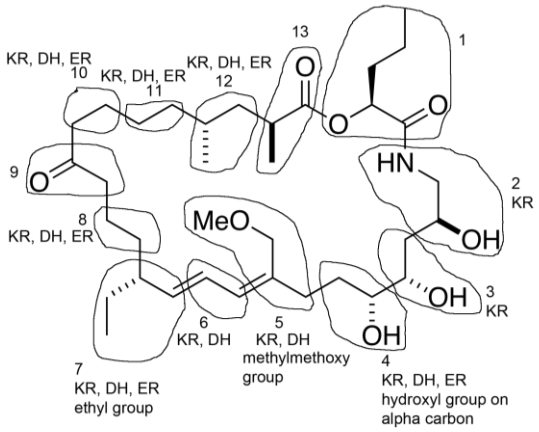
1. What type of assembly line logic is employed in its biosynthesis (i.e. PKS, NRPS, hybrid NRPS-PKS)
2. The monomeric building blocks (circle them)
3. Indicate the likely type of TE-catalyzed chain release

For additional practice: indicate what optional domains, if any, are required.
Note: ignore stereochemistry
Note: ignore the grey circles in some of the figures.

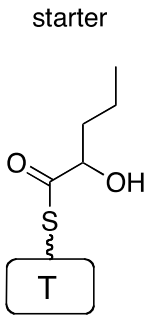


Problem 2:

Myxovirescein A_1 is a small-molecule antibiotic produced by *Myxococcus Xanthus* strain DK1622. Its structure is shown below. This molecule is produced by an assembly line that employs an unusual starting monomer, which is also shown below.



myxovirescein A_1



A) What type of assembly line is responsible for the biosynthesis of myxovirescein A_1?

A mixed PKS/NRPS assembly line.

B) Using the chemical structure above:

i. Circle the monomer units
ii. Number the monomers in the order of chain elongation
iii. If optional domains are required for β-ketone tailoring, specify the optional domains required for each modification

See figure above.

C) How many modules does this assembly line contain?

This assembly line contains 13 modules (loading module and 12 elongating modules).

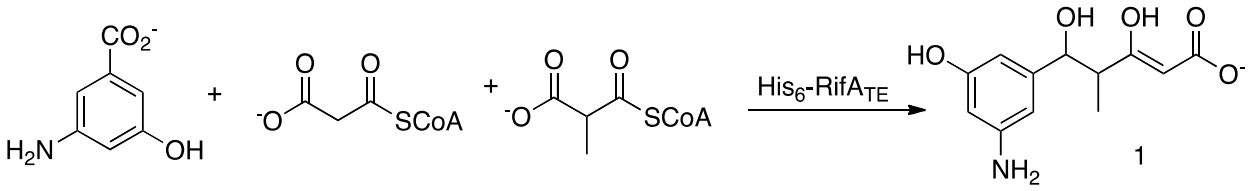
D) In addition to the unusual starter monomer, myxovirescein A_1 has at least two unusual structure features that we did not specifically address in class lectures. What are two of these unusual features? In general terms, what biosynthetic machinery is necessary for the formation of these structural features?

Unusual features are labeled in the structure above, including: hydroxyl group on the α-carbon of monomer 4, methoxymethyl group on monomer 5, and ethyl group on monomer 7. Note that both the methoxymethyl and the ethyl group are installed on what was formerly the carbonyl carbon, so they cannot be generated just by incorporating unusual monomers. All of these structural features are likely incorporated by post-assembly line tailoring or stand-alone/optional domains that modify the growing chain while it is attached to the assembly line.

Problem 3:

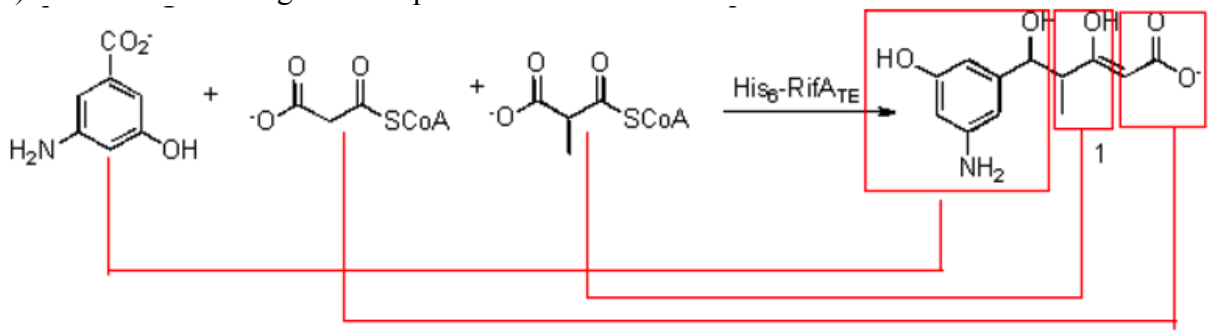
Rifamycin is an antibiotic used to treat tuberculosis. The synthase responsible for rifamycin production is comprised of six multifunctional proteins RifABCDEF. To study the enzymatic activity of RifA, this protein was overexpressed in *E. coli* and purified as an N-terminal His_6 fusion. In this work, the C-terminus of RifA was engineered to have a TE domain (His_6-$RifA_{TE}$).

This engineered His_6-RifATE was incubated with 3-amino-5-hydroxybenzoate, malonyl CoA and methylmalonyl CoA. The following product (**1**) was detected by using HPLC, mass spectrometry and NMR analysis.



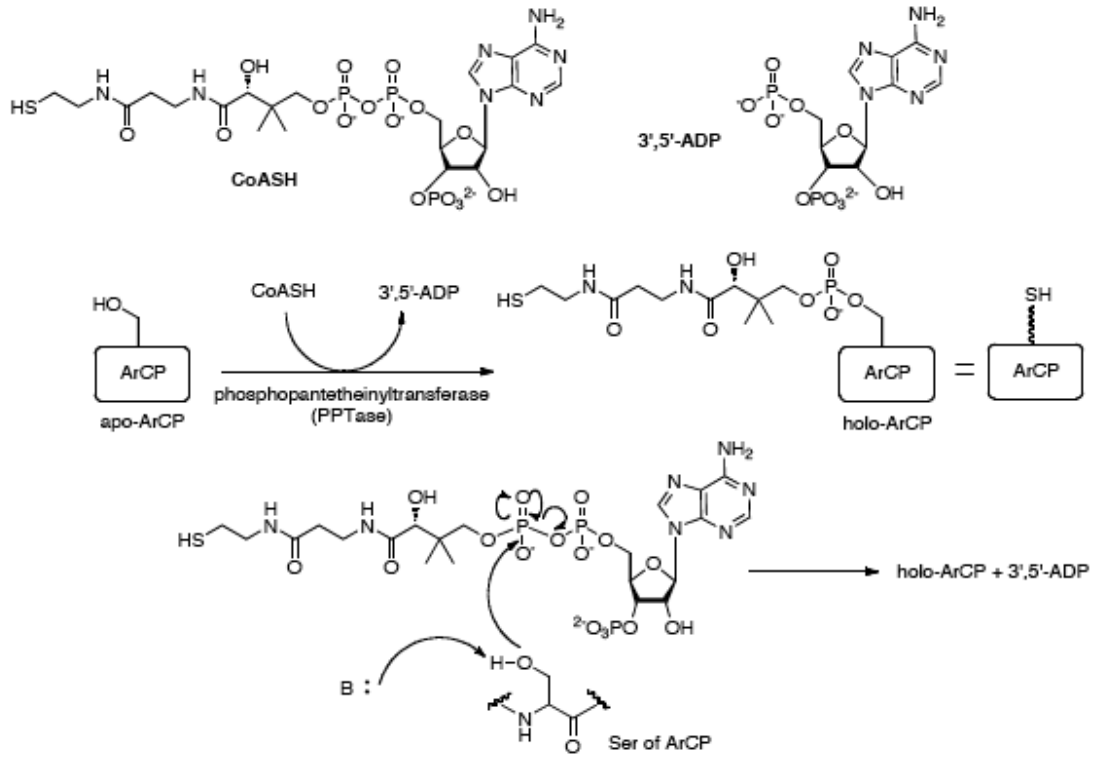
Answer the following questions:

A) Circle each building block in product **1**.



B) Posttranslational modification of domains in NRPS and PKS assembly lines is an essential step in natural product biosynthesis. Which domains must be posttranslationally modified to generate active RifA? Indicate the chemistry of this modification and how many of these domains RifA contains.

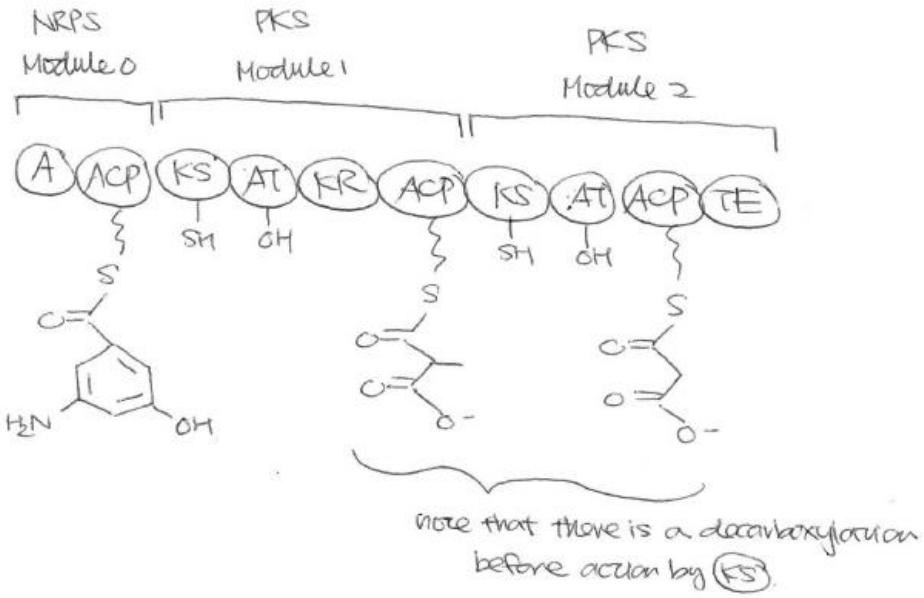
RifA would need to have three of these domains. The ArCP domain for the 3-amino-5-hydroxybenzoate and the ACP domains for the malonyl-CoA and methylmalonyl-CoA need to be posttranslationally modified with the addition of a ppant arm. This is done by a PPTase and CoASH. The chemistry is shown below:



C) What chemical transformation must occur to activate 3-amino-5-hydroxybenzoate?

Adenylation of the 3-amino-5-hydroxybenzoate must occur by an A domain in RifA. Recall that this is an aryl acid (a monomer for an NRPS module), which requires activation by ATP.

D) Draw an assembly line cartoon to indicate the domain and module ordering of RifA required to generate **1**. Indicate what monomer corresponds to each module.



Problem 4:

Cupriachelin, a photoreactive siderophore from *Cupriavidus necator* H16, was recently identified in culture supernatants when the bacterium was grown under iron-limiting conditions. Natural product isolation and a combination of mass spectrometry and NMR spectroscopic studies provided the structure of this molecule, which is shown below.

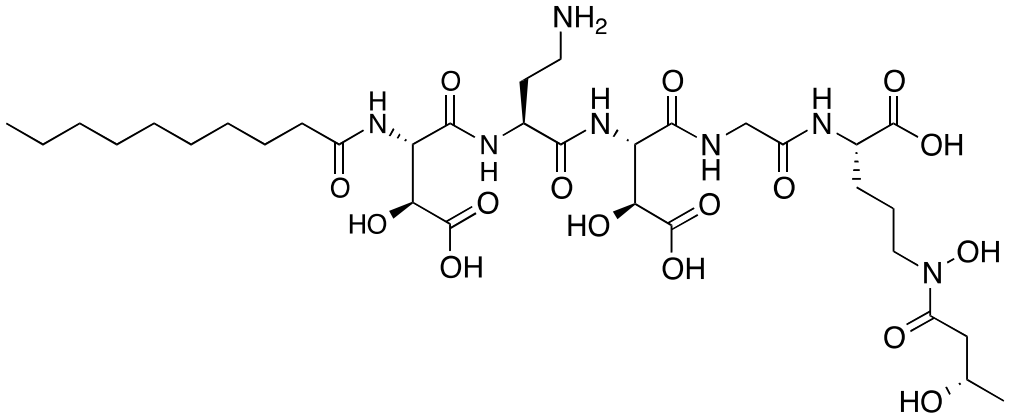


Figure 1. Structure of cupriachelin

Bioinformatic analysis of the *Cupriavidus necator* H16 genome resulted in identification of an NRPS of unknown function and subsequent analysis suggested that it was a likely candidate for cupriachelin biosynthesis. A series of experiments were therefore conducted where genes within this cluster, hereafter named *cuc*, were disrupted (or knocked-out). These mutant strains allowed for the identification of the metabolite produced by the *cuc* cluster. Cupriachelin was determined to be the major metabolite synthesized by the *cuc* NRPS.

The *cuc* NRPS consists of four proteins, CucF, CucG, CucH, and CucJ. Complete core NRPS modules are contained within each protein. In contrast to many of the NRPSs we discussed in class, *cuc* is unusual because it does not contain a loading module.

Bioinformatics analysis afforded the substrate specificity predictions for the A domains, which are given in **Table 1**.

Table 1. Substrate Specificity Predictions for the Adenylation Domains of the NRPSs Encoded in the *cuc* Gene Cluster

A domain	signature sequence	substrate according to prediction
CucF-A_1	DLTKVGHVGK	L-Asp
CucF-A_2	DIWELTADDK	unknown
CucG	DLTKIGHIGK	L-Asp
CucH	DGEGSGGVTK	unknown
CucJ	DILQLGVVWK	Gly

Following bioinformatics analysis, the ATP-PP_i exchange assay (see lecture notes on characterization of aminoacyl tRNA synthetases) was used to ascertain the substrate specificity of the A domain of CucG. First, CucG was overexpressed and purified as a His_6-tagged protein in *E. coli*. **Figure 2** summarizes the results of ATP-PP_i exchange assays conducted with His_6-CucG and a variety of monomers. The y-axis "cpm" (counts per minute, upcoming recitation topic) is a measure of ^{32}P incorporation into ATP.

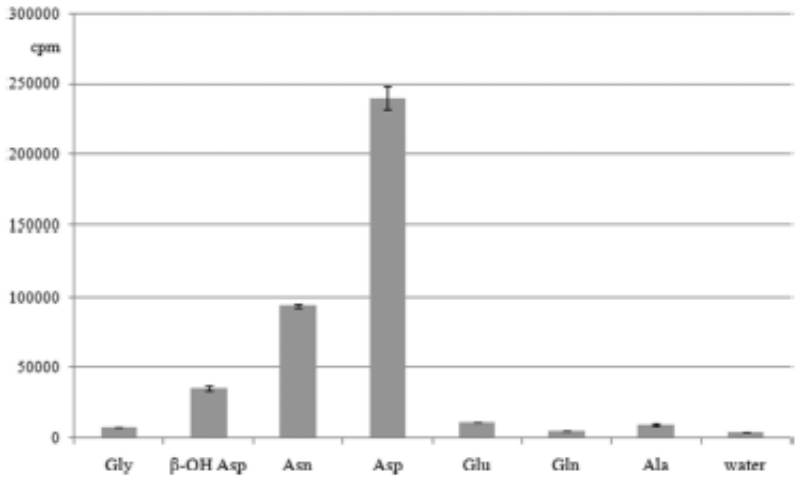


Figure 2. Results from ATP-PP_i exchange assays using His_6-CucG and the substrates listed in the plot. The y-axis "cpm" is a measure of ^{32}P incorporation into ATP. The reactions containing His_6-CucG, ATP, $^{32}PP_i$ and the substrate were incubated in buffer and quenched all at the same timepoint.

Answer the following questions:

A) Circle the monomeric building blocks of cupriachelin.

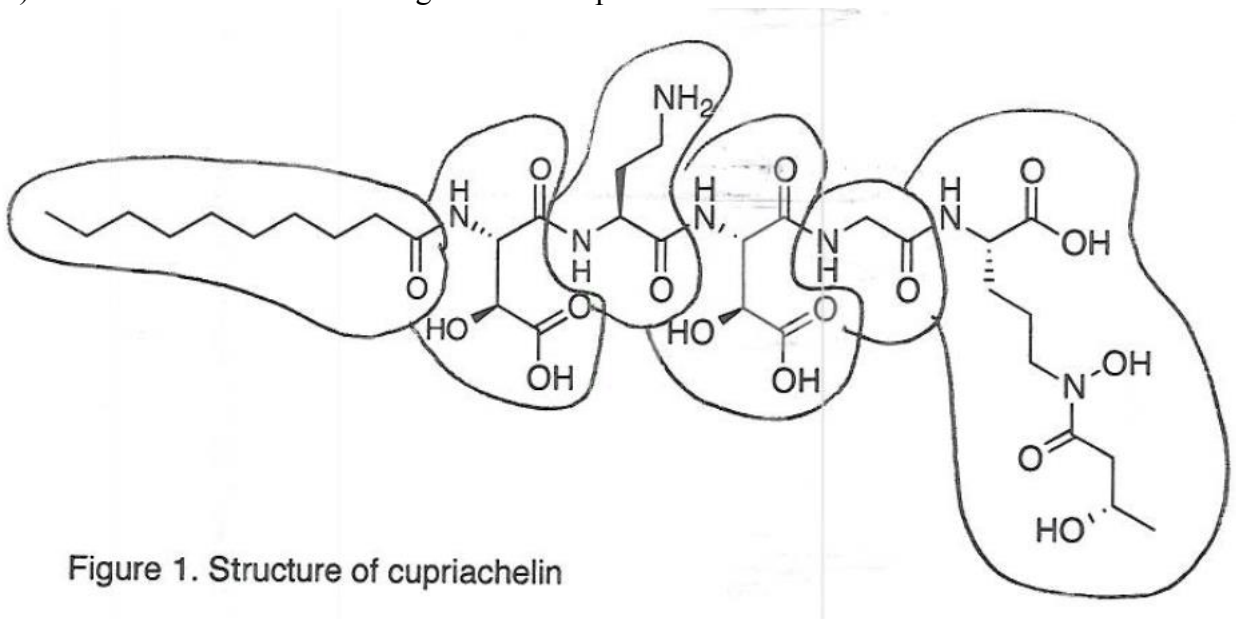
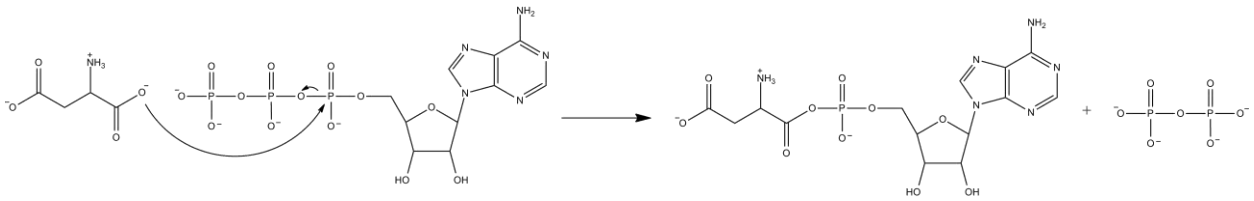


Figure 1. Structure of cupriachelin

B) CucG contains an A domain and is predicted to activate L-Asp. Provided that this prediction is correct, how does CucG activate L-Asp? Draw the mechanism below including all substrates and products.

CucG activates L-Asp using ATP to form the adenylated amino acid. The mechanism is shown below.



The mixed anhydride in the adenylated amino acid can then be attacked by the thiolate of the T domain to form the amino acid-loaded T domain.

C) What do the data in **Figure 2** tell you about the substrate-specificity of CucG? Provide three conclusions that summarize the data completely.

ATP-PP_i exchange only occurs if CucG recognizes and activates the given amino acid.
1) Asp is the preferred substrate of the CucG A-domain
2) The CucG A-domain will also select and activate Asn and β-OH-Asp, but to a lesser degree than Asp
3) There is negligible activity for Gly, Glu, Gln, and Ala (on the same scale as without any amino acid), indicating that the CucG A-domain will not activate these amino acids.

D) Why was ATP/PPi exchange performed with the non-proteinogenic amino acid β-OH-Asp?

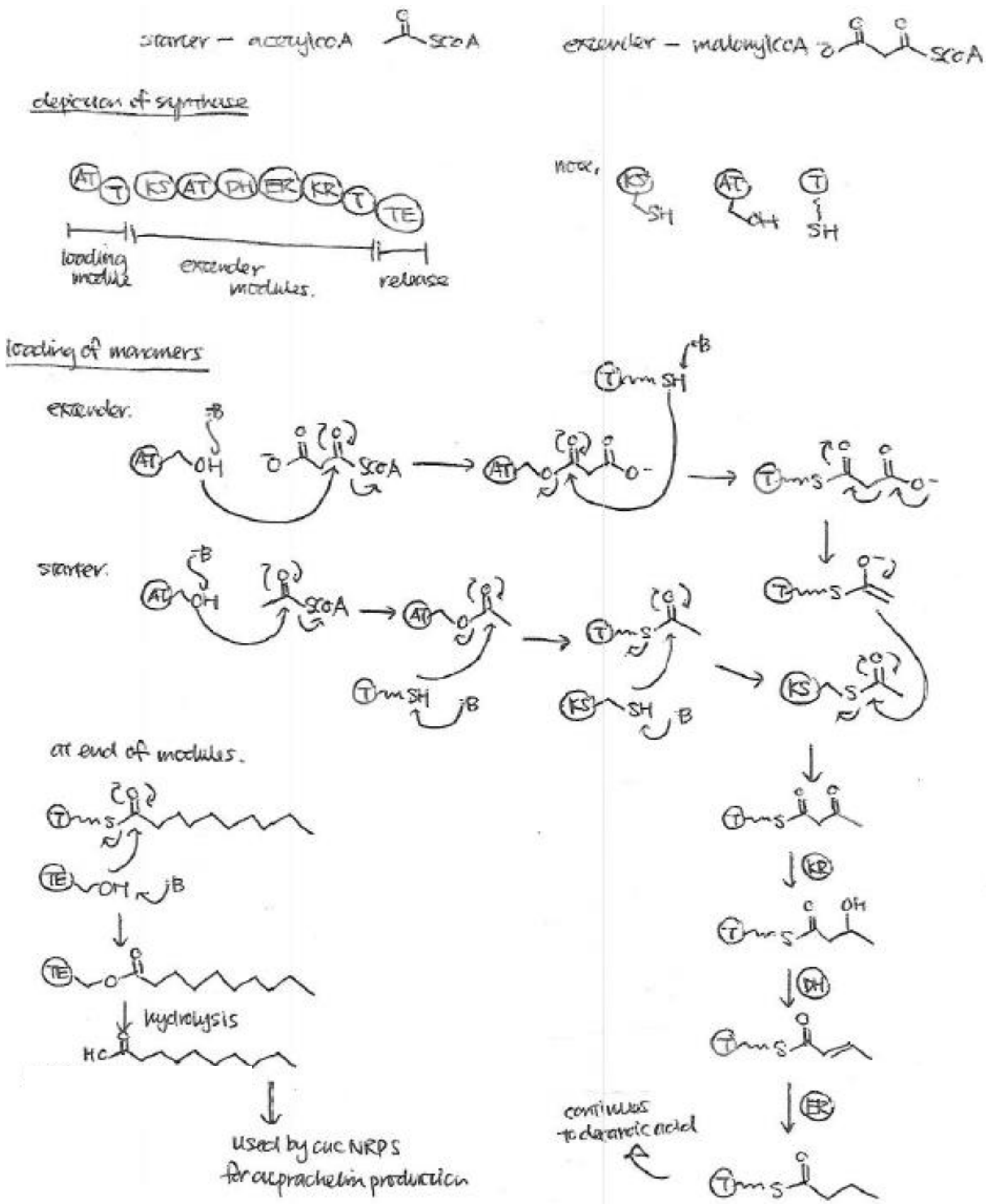
Cupriachelin contains β-OH-Asp. The researchers likely wanted to test if CucG directly inserted β-OH-Asp or if CucG inserted L-Asp with subsequent tailoring to install the hydroxyl group.

E) Considering the structure of cupriachelin and the ATP/PPi exchange data for L -Asp and β-OH-Asp, what do you conclude about the monomer activated by CucG *in vivo* and the overall biosynthesis of cupriachelin?

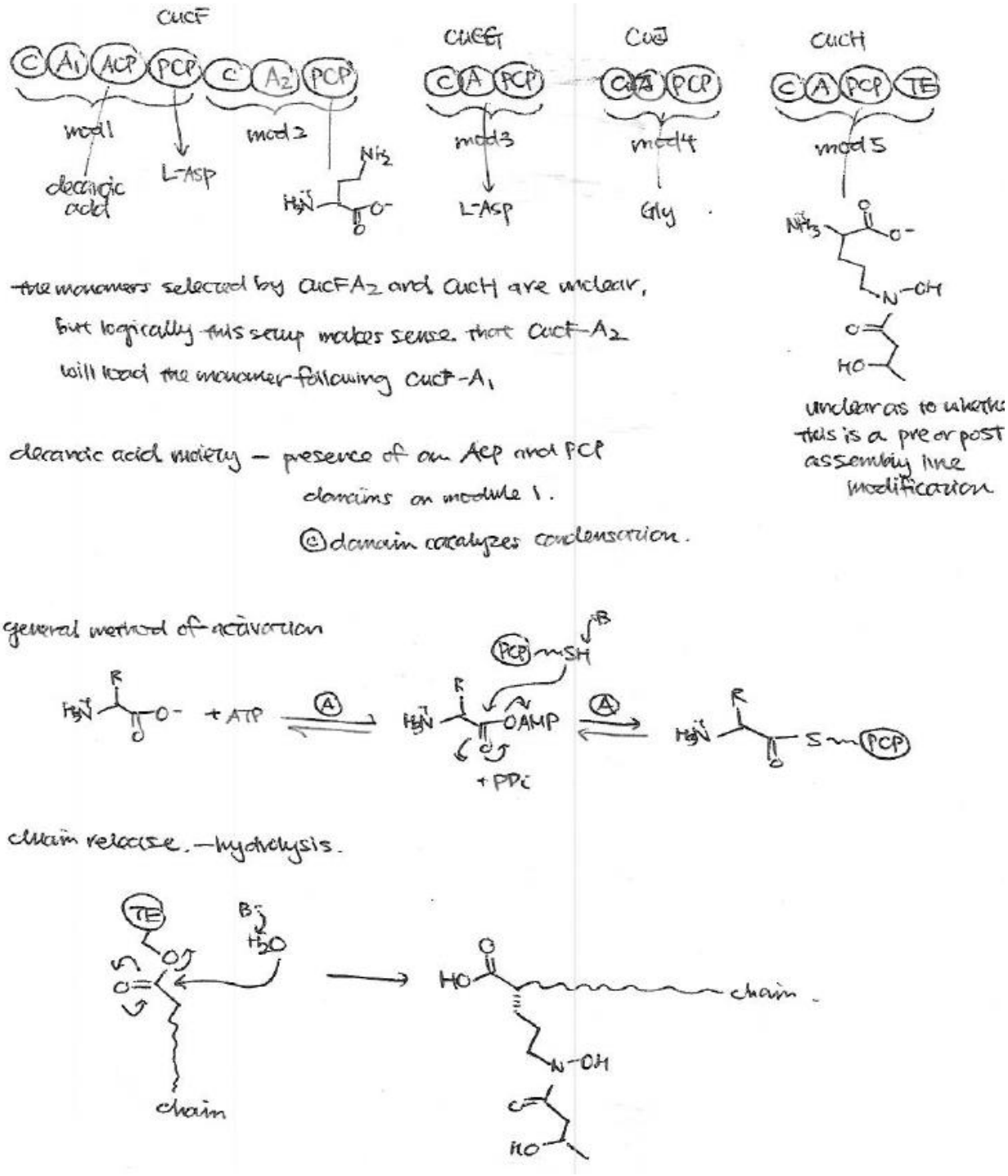
In cupriachelin biosynthesis, the CucG A domain selects and activates L-Asp and transfers this amino acid monomer to the T domain of CucG. Tailoring of L-Asp to provide the β-OH-Asp observed in the final cupriachelin structure occurs separately (optional NRPS domain or independent tailoring enzyme).

F) Decanoic acid has been proposed to act as the biosynthetic starter unit for the *cuc* NRPS. How is decanoic acid biosynthesized? Provide the starter and extender monomers, and a depiction of the synthase including all required chemical steps that lead to decanoic acid production. Specify the mechanism of C—C bond formation. You may abbreviate repeated steps.

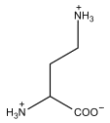
Decanoic acid is biosynthesized by fatty acid synthase.


starter – acetylcoA
SCoA
extender – malonylCoA
SCoA
depiction of synthase
AT
T
KS
AT
DH
ER
KR
T
TE
loading module
extender modules.
release
note,
KS
SH
AT
OH
T
SH
loading of monomers
extender.
starter.
at end of modules.
hydrolysis
used by cuc NRPS
for cuprachelin production
continues
to decanoic acid

G) Using the information provided, propose a biosynthetic pathway line for the assembly of the cupriachelin backbone. Include a cartoon depiction of the NRPS assembly line with protein and domain labels; the monomers selected, activated and covalently attached to the assembly line; a mechanism for introduction of the decanoic acid moiety; and a mechanism for chain release. Assume that the assembly line is already post-translationally modified with ppant arms and disregard stereochemistry.



H) Some additional experiments are required to test your biosynthetic proposal. Using our class discussions of assembly line characterization as a guide, what is the first question you will address and the experiment you will conduct to obtain an answer?

The monomers selected and activated by CucF-A_2 and the CucH A domain must be verified. The first question I would address is: what monomer is selected by the CucF-A_2?
I will purify and express both full-length CucF and CucF-A_2 as was done for CucG. I will then conduct ATP/PP_i exchange assays with both full-length CucF and isolated CucF-A_2 (to eliminate possible interference from CucF-A_1). I will use the putative substrate, all other amino acids derived monomers in cupriachelin, a range of other amino acids, and a water only control. If my biosynthetic hypothesis is correct, I expect to see robust ATP/PP_i exchange with  and no exchange with the other monomers.

If you have trouble with the biosynthetic pathway, you can refer to *J. Am. Chem. Soc.* **2012**, *134*, 5415-5122.

5.08J Biological Chemistry II

Spring 2016