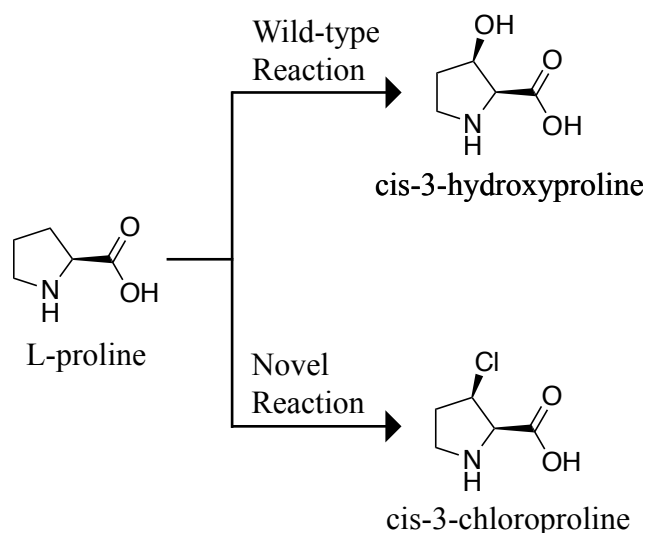


## Challenge 2

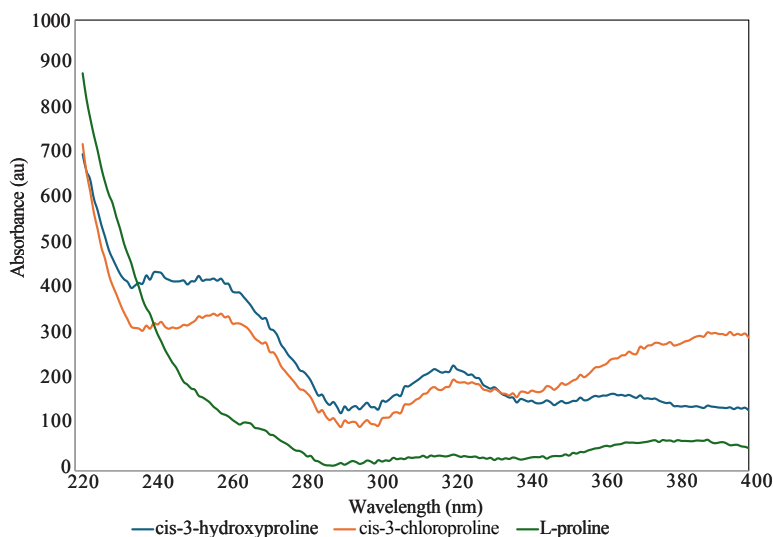
### Enzyme Engineering

**Background:** You are a protein-engineer graduate student in Adv Syn Bio course at UW. A library of 10,000 mutants of an enzyme, proline 3-hydroxylase (P3H) has been generated. The native enzyme hydroxylates L-proline to form cis-3-hydroxyproline. Some of these mutants create a novel product – cis-3-chloroproline.



**Figure 1.** Wild-type and novel reactions from proline-3-hydroxylase and mutants.

The activity of these mutants has been assayed through measuring absorption spectra from the products of these reactions. The resulting absorption spectra can be compared to absorption spectra from standards of L-proline, cis-3-hydroxyproline, and cis-3-chloroproline.



**Figure 2.** Absorbance spectra from cis-3-hydroxyproline, cis-3-chloroproline, and L-proline.

**Goal:** Determine the optimal protein sequence for halogenation and report which amino acid residues are important for what functions. You will need to find which residues are essential for any enzyme activity (1-5 residues), which residues are beneficial to any enzyme activity (1-5 residues), which residues are required for halogenation (1-5 residues), and which residues are beneficial for halogenation (6 - 12 residues).

**Examples:**

Residue 1 must be methionine as that is the only residue found in all enzymes.

Residues 30 to 33 must be glycine, alanine, glycine, and glutamate respectively as those enzymes are conserved in all halogenating compounds.

Residues 100 to 200 do not seem to affect the function of the enzyme if they are mutated.

**Note:** You can report this in other ways as well, such as in a table.

**Sequence Data:** Protein sequences are available in HW2\_mutants.fasta (in fasta format) and in HW2\_Mutant\_absorbance\_data.csv (in columnar format).

**Absorbance Data:** Absorbance data are available in HW2\_Mutant\_absorbance\_data.csv with the header row containing “mutant name”, “sequence”, and absorbance wavelengths (in nanometers). The rows for each mutant contain the name, sequence, then the absorbances for each wavelength (in absorbance units).

**Note:** A given enzyme/mutant could possibly create cis-3-hydroxyproline, cis-3-chloro-proline, both, or neither. No enzyme is guaranteed to consume all of the substrate. Assume that the spectra are approximately linear combinations of only the absorption spectra of the three possible components. Absorption readings are noisy, so you will have to work around this.

**Deliverable:** To receive credit, you will need to provide proof that you did this work (and not copied the answer from another individual). The format of this proof can be a writeup of how you went about solving this problem alongside code you generated.