Homework Assignment 6: *Synthetic Biology*

Your colleague, Dr. Q. Werty, has been boasting to you that he has designed a protein, which if expressed, will prevent the spread of cancerous cells in a mouse. However, the peptide sequence he specified does not exist in nature, and he has thus far been unable to prove his claims. Synthetic gene assembly sounds like a snap to you, so you've offered to assemble his protein for him, if he'll fund it. Dr. Werty is skeptical, and requests that you assemble the first domain before he's willing to support the entire project. The domain has the amino acid sequence DERTMISMHELWPWILVGKGM.

Knowing that to assemble the full gene you will need overlapping oligos, you set out to determine if you can design three 21 bp oligos (which will code for DERTMIS, MHELWPW, and ILVGKGM, respectively) from this peptide, which you can then combine into two overlapping 42 bp oligos (corresponding to DERTMISMHELWPW and MHELWPWILVGKGM) for assembly. To make PCR easier, all 21 bp oligos need to have a melting temperature within **0.5 degrees of 58.0 degrees Celsius**. (Remember, a good approximation for melting temperature is MT = 64.9 + (41.0 * (#G + #C - 16.4) / Length of oligo).) Additionally, to facilitate cloning later in the experimental process, you want to avoid including the restriction sites *NdeI, XhoI, TaqI,* and *BfaI* in your 21 bp oligos.

To translate from amino acid sequence to DNA sequence, you will want to look at all possible codons for each amino acid. The easiest way to do this will be to set up a recursive function which will split for each amino acid with multiple codons. Another colleague, Dr. Polk, has provided you with code located in /home/assignments/assignment6/Polk.py which his lab has been using to print all possible DNA sequences for a given polypeptide.

The usage of the script is currently:

$ python3 Polk.py <AA sequence>

The usage of the script will be:

$ python3 Polk.py Any number of <AA sequence> Any number of <melting temps>

Please modify and comment the provided script so that it will take as input an amino acid string(s) and melting temperature(s). For each sequence first print a line that says the sequence and melting temperature used. Then print a tab delimited list of each DNA sequence and its respective calculated melting temperature with an appropriate header. As a hint, there should be multiple possible DNA sequences for each AA string.

What to turn in (submission folder)

- A commented and modified Polk.py script that can take in a sequence or sequences and a melting point temperature/temperatures.
- A README.txt file with how to run your code followed by output from the three AA strings.