DNA Sequence II Project Plan

This is the second iteration of my DNA sequence translator, and is meant to improve upon the first version; namely, by using more advanced programming techniques such as OOP.

The purpose of this programme is to take a sequence of DNA in the FASTA format, and convert the sequence into the complementary amino acid sequence. On top of the basic purpose of the programme, I want it to have some additional features such as writing to a file the DNA sequence, and the amino acid sequence with included title and date of processing.

Step by step plan of the project:

1. Make a CSV file which includes headers of all the amino acids, and below each header is the complementary amino acid sequence. This is so that the programme can access this file, and read it to identify the amino acid sequence. There are two potential ways of doing so: either the file is read each time a new triplet is read; or a new list is made in the programme to load in all of the amino acids and their complementary sequences. I think the latter option is better for two reasons: firstly, to include all of the amino acid sequences does not take a huge amount of data so loading it in at once will be fast. Secondly, if the file must be read each time a triplet is identified, it is much slower than loading everything into a list and reading from there.
2. Ask the user to enter the DNA sequence in the FASTA format, and store as a variable.
3. Segment the DNA sequence into three nucleotide sequences (triplets) which can be processed.
4. Compare the triplet sequence, and identify the complementary amino acid. The amino acid should be appended to a list, which builds up as the translation continues.
   1. There should be an option to write the amino acid sequence as individual names, 3-letter names, or 1-letter names
5. The amino acid sequence is written to a file, alongside the FASTA DNA sequence, the time of translation, and the user’s desired title.