Exercise set 3

1. Write a function that computes the percentage of amino acid residues. It takes two arguments, the protein sequence, and an amino acid residue. Test the function for the given input and output.
   1. MSRSLLLRFLLFLLLLPPLP, M → 5
   2. MSRSLLLRFLLFLLLLPPLP, r → 10
   3. msrslllrfllfllllpplp, L → 50
   4. MSRSLLLRFLLFLLLLPPLP, Y → 0
2. Write a function that computes the percentage of amino acid residues. It takes two arguments, the protein sequence, and, optionally, a list of amino acids. If no list of amino acids is given, the function should compute the percentage for hydrophobic amino acid residues, i.e., A, I, L, M, F, W, Y, and V. Test the function for the given input and output.
   1. MSRSLLLRFLLFLLLLPPLP, [M] → 5
   2. MSRSLLLRFLLFLLLLPPLP, [M, L] → 55
   3. MSRSLLLRFLLFLLLLPPLP, [F, S, L] → 70
   4. MSRSLLLRFLLFLLLLPPLP, → 65
3. The file `data\_6.csv` contains data on a number of genes. Print the gene names for all genes belonging to *Drosophila melanogaster* and *Drosophila smiulans*.
4. The file `data\_6.csv` contains data on a number of genes. Print the gene names for all genes with sequence length between 90 and 110 base pairs.
5. The file `data\_6.csv` contains data on a number of genes. Print the gene names for all genes with AT content of less than 0.5, and an expression level larger than 200.
6. The file `data\_6.csv` contains data on a number of genes. Print the gene names for which the name starts with either 'k', or 'h', except for \*Drosophila melanogaster\*.
7. The file `data\_6.csv` contains data on a number of genes. Print for each gene its name, and whether its AT content is high (i.e., larger than 0.65, low (i.e., less than 0.45), medium otherwise.