CS250 Assignment 1
Hamming Distance

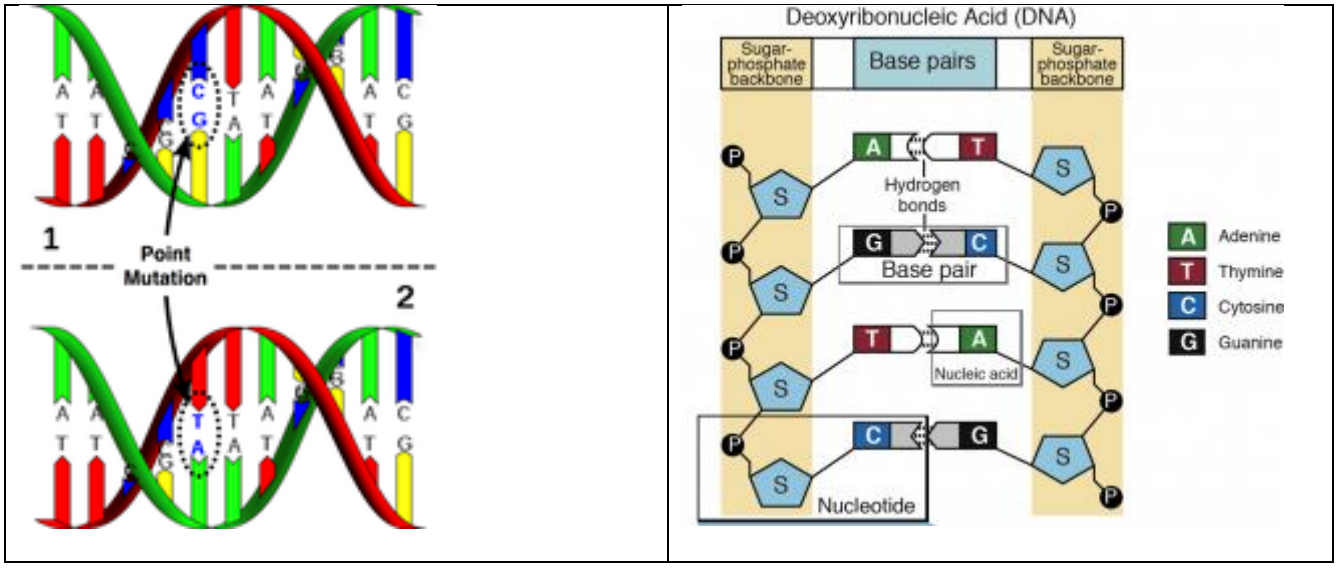
Date assigned: Friday, February 2, 2018
Date due: Friday, February 9, 2018
Points: 25

Goals for Assignment 1

1. Reacquaint yourself with Visual Studio and Grace.
2. Implement a C++ program that uses files, 1D & 2D arrays, and character processing.
3. Break up a program into well-defined functions. This is the first assignment where you choose the functions. It is important that your functions be small and focused on a particular task.

Background

During the creation or copying of a nucleic acid, a mutation can occur. Mutations can be beneficial, harmful, or harmless. A point mutation replaces one base with another at a single nucleotide.



To make a long story short, the number of point mutations can be used to explain evolutionary histories of different organisms or species.

Consider two DNA strings s1 and s2 that are equal in length. The Hamming distance denoted, $d_H(s1, s2)$, is the number of mismatched bases in corresponding locations of s1 and s2.

For example, if

- S1 is: AAAACCCGGT
- S2 is: AAACCCGGGT

then the Hamming distance, denoted d_H (s1,s2) is 2. The differing bases are the 4^{th} and 7^{th} ones.

The Problem

Write a complete C++ program that will read in a collection of at most 20 equal length (at most 60 bases) DNA strings from a file called dnastrings.txt. Next, allow the user the ability to enter a single DNA string. Your program is to output the Hamming distance between the user entered DNA string and each of the DNA strings read in from the file as shown below. Finally, output the entered DNA string and the string with the smallest Hamming distance as shown below. If two or more strings have the same Hamming distance, output the first string encountered that has the smallest Hamming distance.

Input: The file named (dnastrings.txt) contains an unknown number of DNA strings. There will be at most 20 strings and the length of each string will not exceed 60 bases. Each string will be separated by a new line.

```
AAGTACGG
ACGTACGG
AAATACGG
```

Your program is to work and look exactly like the following:

```
********************
  Hamming distance
********************

Enter DNA String: ACGTACGT

Input String: ACGTACGT
         S01: AAGTACGG
          dH: 2

Input String: ACGTACGT
         S02: ACGTACGG
          dH: 1

Input String: ACGTACGT
         S03: AAATACGG
          dH: 3

String S02 has the smallest dH of 1
```

Notes:

1. Your main function is to be mainly variable declarations and function calls.
2. Test your program one function at a time.

To complete this assignment you must submit the following:

1. An electronic copy of your program on Grace

a) Create a new C++ solution in Visual Studio 2017. Your solution must be called **PUNetID-Assignments**. For example, mine would be called ryandj-Assignments. This solution will hold all of your assignment projects for CS 250. You are starting over for CS250, meaning do not include any projects from CS150.

b) Your project for this assignment must be named **01_Hamming**. It is vital that you name your solution and your project correctly!

c) Type your program (fully documented/commented) into the project. You need to follow the coding standards from the CS250 Web page. These coding standards have been modified to include additional C++ language features introduced in CS250, so please be sure to read the new coding standards.

d) Pay attention to the example output. Your program's output must look **exactly** like the sample output. The spacing and newlines in your output must match exactly.

e) Make sure that your program builds without errors & warnings and runs correctly. If you get any errors or warnings, double check that you typed everything correctly. Be aware that C++ is case-sensitive. You will lose 10% if there are any warnings and 40% if your program does not build successfully.

f) Once you are sure that the program works, it is time to submit your program. You do this by logging on to Grace and placing your complete solution folder in the proper **CS250 Drop** folder.

g) The solution must be in the drop folder by the time class starts on the day the assignment is due. Anything submitted after that will be considered late.

2. A hard copy of your program

a) The hard copy must be placed on the instructor's desk by the time class starts on the day that it is due.

b) The hard copy must be printed in color, double-sided, and stapled in the upper left corner if your solution contains multiple pages.

c) Your tab size must be set to 2 and you must not go past column 80 in your output.

Remember, if you have any problems, come to me straight away with your project on a flash drive or on Grace. Good Luck!!!! ☺