Homework 3
CS5046

Due by 5:30PM, February 23, 2004
Submit by email to murali@cs.vt.edu

In this homework, you will implement a set of classes that create, store, and process strings of amino-acid sequences and computes statistics about them. When possible, all your methods must ensure that the values of the parameters passed are sensible and print an appropriate error message if they are not. Use the AminoAcidSequence class you implemented in Homework 2. You can reimplement the AminoAcidSequence class, if you like.

Problem 1 Implement a class called AminoAcidSequenceSet that stores multiple amino-acid sequences.

1. (20 points) Implement a constructor that takes no arguments. Implement another constructor that takes two integers as arguments. The first argument \( k \) (your parameter should have a meaningful name) corresponds to the number of amino-acid sequences to store. The second argument \( l \) is the length of each amino-acid sequence. The constructor should create \( k \) random AminoAcidSequences, each of length \( l \).

2. (20 points) Implement an \( \text{add()} \) mutator that takes an amino-acid sequence (a String) as argument and adds it to the list stored.

3. (50 points) Implement a method called \( \text{CountStatistics()} \) that counts the following statistics:
   (a) For each position and for each amino acid, count the number of times the amino acid occurs at that position. The count is over all the stored sequences. The method should take into account the fact that different amino acid sequences can have different lengths.
   (b) For each position and for every sequence of two amino acids, count how many times the sequence appears starting at that position in the entire list of stored sequences. Hint: There are 400 sequences with two amino acids. Do you have to explicitly consider all 400 sequences?

4. (10 points) Implement a method called \( \text{printStatistics()} \) that prints the statistics you computed in \( \text{countStatistics()} \). Hint: What will this method do if the user did not call \( \text{countStatistics()} \) before invoking \( \text{printStatistics()} \)?

5. (Extra credit 20 points) Modify the \( \text{add()} \) mutator so that it does not add the sequence if it is already stored in the AminoAcidSequenceSet. Hint: Your initial implementation of AminoAcidSequenceSet may have to change if your first design of the class does not take this extra credit problem into account.

6. (Extra credit 30 points) Implement a \( \text{remove()} \) mutator that takes an amino acid sequence (a String) as input and removes it from the stored list. Hint: How will this method recognise and deal with a sequence that is not stored?

Notes and Tips

- The hints I have given are meant to help you think about the various issues you have to consider in your implementations. You may find it useful to include comments in your code that detail why you made a particular design decision.
• Do not forget to include full documentation for your source code as described in Chapter 5.10 of Barnes and Kölling.

• Create a folder called <YourName>-Homework3 and use a separate file inside this folder for the source code for each class in each problem.

• Remember to compile your code and to test it in the same way that we have been interacting with classes in BlueJ.

• Submit your homework by emailing the files containing the source code to me. I prefer it if you can zip the entire folder and send the zipped folder to me. If doing so is difficult, just email me the individual files as attachments.