Assignment 2: Handling Reverse Complement
Computational Genomics

Objectives: Practice integrating your knowledge of computer science techniques to develop efficient solutions to genomics problems.

Background:

Consider creating a database designed to search for specific sequences within a given segment of double-stranded DNA. Here's the DNA segment for reference:

```
5' ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAA 3'
3' TAATTTCCAAATATGGAAGGGTCCATTGTTTGGTTGGTTGAAAGCTAGAGAACATCTAGACAAGAGATTT 5'
```

This database can determine whether a particular sequence is present or absent in this DNA segment. For instance, if you search for the sequence TTAAAGGT, the database will confirm its presence (return true). However, if you search for GGATCCT, the database will indicate that this sequence is not present (return false).

It's important to note that each position (locus) in the DNA can be represented by two sequences - one from the forward strand and one from the reverse strand. Therefore, a query for AATTTCCA will also return true, as it corresponds to a sequence present on the reverse strand of the same DNA segment as TTAAAGGT.

To enable queries on both DNA strands, we have two options: store both the forward and reverse strands in the database, or compute the reverse complement for each query and check for the presence of either the original query or its reverse complement

Questions:
1. Is it possible to achieve accurate results by storing and querying only one sequence?

Submit your answer in a PDF to Canvas.