

# HW#1 SOLUTIONS

**CSC / BIO 310**

**Bioinformatics**

**Instructor: Dr. Laurie J. Heyer**

**Assignment #1**

**Due Tuesday, Jan 22**

Write a perl program that contains the following line:

```
$DNA = 'AcGTtTAtGcgaTtaGAcGTaGctAtCGaT';
```

Write comments in your file to indicate the solutions to each of the following problems. You should also include the number of each problem in the print statement.

Write a comment at the top of the program containing the five lines at the top of this assignment page, in addition to the names of people in your programming team.

You may not consult with anyone outside of your programming team, other than me.

In addition to the accuracy of your solutions, you will be graded on the readability of your code and your output, and the use of good programming practices as discussed in the text and in class.

1. Print the above DNA sequence as is.

Well, this is simple:

```
print "\n(Question 1) The DNA sequence is $DNA.\n";
```

2. Use the tr command to toggle the case of the DNA sequence, and print the resulting sequence.

```
my $casetoggleDNA = $DNA; #We need to copy the DNA variable, since the transliteration function will overwrite the variable we were working on.
```

```
$casetoggleDNA =~ tr/ATCGatcg/atcgATCG/; #We used the transliteration function to switch lowercase letters to uppercase letters and vice-versa
```

3. Assuming the given sequence is the template strand of a sequence that will be transcribed, use an appropriate perl command to create the corresponding RNA sequence, and print the RNA in upper case.

First, what is a template strand? There is some confusion over this, as described at <http://www.sci.sdsu.edu/~smaloy/MicrobialGenetics/topics/chroms-genes-prots/temp-strand.html>

But the majority of (reliable) web pages that turn up in a Google search do agree with this interpretation. Here is another helpful site:

<http://dwb.unl.edu/Teacher/NSF/C08/C08Links/gregor.rutgers.edu/genetics/Week4/Lecture4-2.html>

Note that the RNA sequence will not be reversed from the order of the template strand.

### Sample Code:

```
# 3. Determine the RNA sequence
$DNA = uc($DNA); # convert all letters to uppercase
# make a copy since the following command will change the
# variable that is searched, and we might want our DNA
# original again
my $RNA = $DNA;
# Complement and change T to U for RNA all in one command
$RNA =~ tr/GCAT/CGUA/;
```

4. Considering the start codon ATG to be the first codon in the coding sequence, print out the fourth codon.

The following solution is especially nice because of the way a variable is used for the desired codon. This is a good example of easily re-usable and generalizable code.

```
# 4. Find the fourth translated codon
my $StartCodon = index($RNA, 'AUG');
my $CodonNumber = 4;
my $Codon = substr($RNA, $StartCodon + 3*($CodonNumber-1),
3);
print "4. Fourth translated codon: $Codon \n\n";
```

5. Use the `tr` command to determine the GC content of this DNA sequence, and print out the result as a percentage. You will need to look up the appropriate (but rather obvious) commands in Chapter 5.

```
#Part 5. Find the GC content of our DNA sequence using the
tr command and print
#First, find the length of the DNA sequence
my $DNALength = length($DNACaps);
#Then, find the number of G's and C's
my $numGC = ($DNACaps =~ tr/GC/GC/);
#Finally, find the percentage of G's and C's in the strand
my $PercentGC = ($numGC / $DNALength) * 100;
print "Part 5. The GC content of our DNA sequence is
$PercentGC%\n\n";
```