

MODULE1 PRACTICAL THINKING - BIOINFORMATICS

You have been given a gene sequence, which is in RAW file format. Perform the following tasks by visually inspecting the sequence using information you were provided during lectures for Module 1, Topics 2 and 3.

```
CCCTGGTTATTTATTGACAGGGGTATATAGTCCAGTGGTAGGCTGGATTA
TATAATTTATTGAAATTTTCGATAAATTCGGTTATAATAATATTTAAATCT
GAAAATATTTAGTAAAAGGATGTGAGATATGTGAAGCTAAAACGCTTGTC
TTTCTTCATGTTTGTACCTTACTGGTATTTATATCTGTTTTTTCCTGTTT
ATGCAAACGATTTGAAAAACATGGTACATATTATGAAATTTTTGTCAGG
TCTTTTTATGACTCTGATGGTGACGGAATAGGGGATTTGAAAGGTATAAT
AGAAAACTGGATTATCTTAATGATGGAGACCCTGAAACCATTGCTGATC
TGGGGGTTAATGGTATCTGGTTAATGCCTATCTTTAAATCTCCCTCCTAT
CATGGCTATGATGTAACCGATTATTATAAGATTAATCCTGACTACGGGAC
TCTGGAAGACTTCCATAAGCTTGTGAGGCTGCCCATCAAAGGGGAATCA
AGGTTATAATTGATTTACCCATCAATCATAACAGTGAAAGACATCCCTGG
TTTCTCAAGGCTTCCCAGGATAAGAATAGTGAATACAGGGATTATTATGT
CTGGGCTGGCCCCGATACCGATACCAAAGAAACCAAGTTAGATGGAGGCC
GGGTCTGGCATTATTCCCCGACCGGCATGTATTATGGGTATTTCTGGAGT
GGCATGCCTGATTTAAACTATAATAACCCTGAAGTTCAGGAAAAGGTTAT
TGGGATAGCAAATACTGGTTAAAACAGGGGGTTGATGGTTTCAGGCTTG
ATGGAGCCATGCATATCTTCCACCGGCCAGTATGATAAAAACCTTTACC
TGGTGGGAGAAGTTCCGTCAGGAAATGAAGAGGTAAAACCCGTTTACCT
GGTGGGTGAGGTCTGGGATATTTCCGAAACGGTAGCTCCTTACTTCAAAT
ATGGTTTTGATTCTACCTTTAACTTTAACTGGCAGAGGCAGTTATCGCT
ACGGCTAAAGCTGGATTTCCCTTTGGTTTTAATAAAAAGGCAAAACATAT
TTACGGGGTATATGACAGGGAGGTTGGATTTGGGAATTATATCGATGCTC
CCTTCTGACCAACCATGATCAGAACCGGATTTTGGACCAGCTTGGGCAG
GATCGTAATAAGGCCAGGGTTGCTGCCAGTATTTATTTGACCTTGCCTGG
TAATCCCTTTATTTACTATGGTGAAGAAATCGGTATGAGGGGGCAGGGGC
CCCATGAAGTTATCAGGGAGCCCTTCCAGTGGTATAATGGATCCGGGGAG
GGAGAAACATACTGGGAGCCAGCCATGTATAATGATGGCTTTACTTCTGT
TGAACAGGAAGAAAAGAATCTCGATTCCCTCTTAAATCACTACAGGAGGT
TAATCCATTTCCGGAACGAAAATCCTGTCTTTTATACCGGTAAGATTGAG
ATTATAAATGGAGGATTAATGTAGTTGCATTTAGAAGATATAATGATAA
GAGGGATTTATATGTCTACCATAACCTGGTAAACAGACCGGTTAAAATAA
AAGTGGCTAGTGGTAACTGGACCTTATTGTTTAAATTCAGGTGATAAGGAA
ATTACCCCTGTTGAAGATAATAATAAACTTATGTATACTATCCCTGCTTA
TACTACCATTGTTCTGGAAAAGGAGTAAAGGGAGAGGGTGAGTAATAATT
ATGAAAAGGAGAGGGTAAATCCCTCTCCTTTTTTTGTTAAA
```

- Identify the possible Open Reading Frame (ORF). Hint: search for the START and STOP codons.
- How many amino acids will the gene code for?
- Translate the sequence and deduce the amino acid sequence for the putative protein.
- Can you identify a signal peptide? If yes, identify it?
- Sometimes, it is not possible to identify a gene using the method you used in question (a) above. One of the reasons is that the gene has not been sequenced completely – this happens during random sequencing of clones from bacterial genomic libraries. Perform a 6 frame translation of the given gene. Can you see the difference between the amino acid composition of all 6 translations.

- (f) Determine the G+C content of the coding region of the gene. Change the G+C content of the gene without changing the amino acid composition.