

Problem1_5 :

Description of functions used:

Function of transcription for DNA-RNA conversion takes dna sequence as an input and iterates each character by concatenating it in rna(another string variable) finding if any of them is T. If found it is replaced by U and then concatenated.

Hence we get the final RNA sequence.

Function for generating protein sequence iteratively considers sequence of 3 increments i by 3 from 0 to len(m_rna)-3, finds the corresponding sequence in codon_table.tsv .As soon as it is found it appends the respective protein character in the fasta file

Function for finding GC content count the occurrences of 'GC' in the DNA sequence and calculates its composition by using the formula : $\text{count}/\text{length}(\text{DNA}) \times 100$

Genbank is a collection of all publicly available DNA sequences which is maintained by NCBI. The format contains source, references and authors who are the contributors for the data in the gb file, Version of the file, Organisms list whose DNA has been included in the file all under the tags such as : LOCUS, DEFINITION, VERSION, DBLINK, KEYWORDS, SOURCE, ORGANISMS, REFERENCES, AUTHOR, etc.,. The most important thing to note is the features tag which further contains CDS. All CDS features must have a product qualifier (protein name). CDS region is the coding sequence which contains start and end location of the sequence it refers to, gene, locus-tag with each having a unique protein ID. In the end of the file it contains all the DNA sequences with location as the start of each line for better search.

Central Dogma of Life : It was first described by Francis Crick. It is a model describing flow of genetic information in living cells from a language of nucleic acid to that of proteins. It consists of three steps:

(i) Replication: formation of 2 daughter DNA

(ii) Transcription: formation of mRNA

(iii) Translation: formation of proteins from mRNA

There are a lot many proteins helping for this process to occur such as helicase (separate two DNA strands during replication), DNA polymerase, DNA ligase, etc.,

Diagrammatically it can be represented as:



