Задачи

1.     Подсчитать GC-content для гена с идентификатором X53950.1 (последовательность взять в GeneBank).

2.     Получить последовательность fasta из pdb файла (файл – frompdb.pdb), каждая цепь (последовательность) записывается отдельно со своим заголовком.

3.     Перевести ДНК последовательность в РНК, а потом в белок (файл dna.txt)

4.     Сделать url запрос в GeneBank по идентификатору гена 7157, распарсить полученный файл, чтобы найти информацию: идентификатор в Uniprot кодируемого этим геном белка.

5.     Поиск мотивов в последовательности с помощью регулярных выражений (dna.txt).

Найти все мотивы (указать индексы начала мотивов в последовательности):

a)    GCNGC, где N означает любой нуклеотид – сайт, который находит BisI restriction enzyme.

b)    Определить эукариотическую мРНК:

ATG – стартовый кодон.

За стартовым кодоном следует последовательность из нуклеотидов (A, T, G или C) длиной от 30 до 1000

Затем следует полиА-хвост (последовательность, состоящая только из А нуклеотидов) длиной от 5 до 10