

Tema nr 2

Sucesiunea nucleotidelor într-o catenă de ADN sau ARN este reprezentată în biochimie și în bioinformatică în formatul FASTA. Acesta reprezintă un text conținând literele A (adenină), C (citozină), G (guanină), T (timină) și U (uracil) în ordinea în care nucleotidele conținând bazele azotate corespunzătoare apar în catenă:

```
GGCAGATTCCCCCTAGACCCGCCCGCACCATG
ATAAACAGTGCTGGAGGCTGGCGGGGCAGGCC
ATGAGAGCCCTCACACTCCTCGCCCTATTGGC
...
```

Următoarea funcție se poate folosi pentru citirea succesiunii de nucleotide dintr-un fișier.

```
def citeste_fasta( nume_fisier ):
    catena = ""
    with open( nume_fisier, "r" ) as fisier:
        for linie in fisier:
            catena = catena + linie.strip()
    return catena
```

Alcătuieți un script python cu următoarea funcționalitate:

1. să folosească funcția de mai sus pentru a prelua o succesiune de nucleotide dintr-un fișier text
2. să identifice dacă succesiunea de baze corespunde unei catene de ADN (care conține timină) sau ARN (conține uracil)
3. să calculeze procentul din fiecare nucleotidă catenă (A, C, G și T sau U, după caz)
4. să reprezinte grafic, folosind biblioteca `matplotlib`, procentele de la punctul anterior (grafic de tip bară sau disc)
5. să genereze și opțional să salveze într-un fișier succesiunea de nucleotide corespunzătoare catenei complementare, știind că în cele două catene ale unui acid nucleic perechile de nucleotide complementare sunt G – C și A – T (în ADN) respectiv A – U (în ARN); de exemplu, pentru fragmentul **GATTC**, fragmentul corespunzător în catena complementară ar fi **CTAAG**
6. să compare două catene complementare, citite din câte un fișier text, identificând nepotrivirile dintre bazele nucleotidele corespunzătoare; de exemplu, dacă cele două catene conțin **GATTC** și respectiv **CTTAG**, programul trebuie să identifice eroarea din poziția a 3-a, unde catena a 2-a ar trebui să conțină A în loc de T

Observații: 1. Pentru parcurgerea unui șir de caractere caracter cu caracter se pot folosi aceleași tehnici ca pentru liste; 2. Pentru teste puteți folosi fișierele test atașate