**Zadanie 1.**

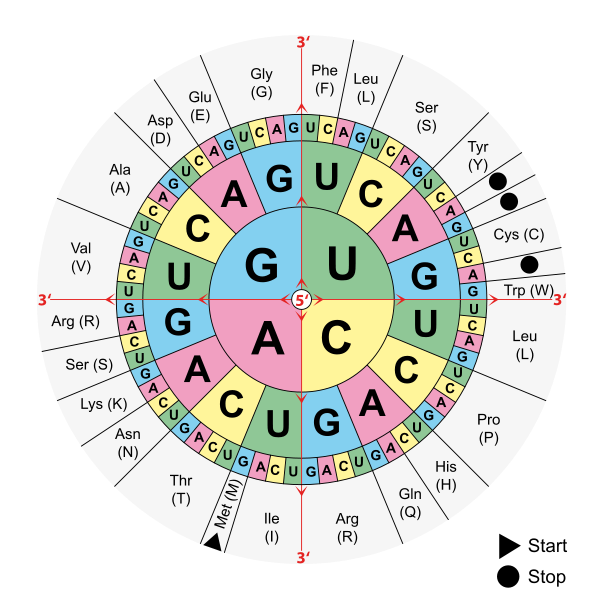
**Problem:**

Pojedynczy aminokwas kodowany jest przez trójkę położonych obok siebie nukleotydów zwanych **kodonem** (tripletem).

**Kodon (triplet)** – jednostka w sekwencji [mRNA](https://pl.wikipedia.org/wiki/MRNA" \o "MRNA) składająca się z trzech [nukleotydów](https://pl.wikipedia.org/wiki/Nukleotydy" \o "Nukleotydy) kodujących określony [aminokwas](https://pl.wikipedia.org/wiki/Aminokwasy" \o "Aminokwasy). Istnieją 64 kodony określające 20 aminokwasów. Wyjątek stanowią cztery kodony: kodon AUG, który inicjuje [translację](https://pl.wikipedia.org/wiki/Translacja_%28genetyka%29" \o "Translacja (genetyka)) ([kodon startowy](https://pl.wikipedia.org/w/index.php?title=Kodon_startowy&action=edit&redlink=1" \o "Kodon startowy (strona nie istnieje))), oraz kodony UAG, UGA   
i UAA, które są znakami do zakończenia translacji ([kodon terminacyjny](https://pl.wikipedia.org/wiki/Kodon_terminacyjny" \o "Kodon terminacyjny)).

Istnienie większej liczby kodonów niż aminokwasów nazywamy [degeneracją kodu genetycznego](https://pl.wikipedia.org/w/index.php?title=Degeneracja_kodu_genetycznego&action=edit&redlink=1" \o "Degeneracja kodu genetycznego (strona nie istnieje)) (kilka kodonów może dyktować ten sam aminokwas). Takie synonimy różnią się między sobą [pozycją tolerancji](https://pl.wikipedia.org/w/index.php?title=Pozycja_tolerancji&action=edit&redlink=1" \o "Pozycja tolerancji (strona nie istnieje)) (trzecią pozycją kodonu).

Każdy aminokwas może być zakodowany na kilka różnych sposobów. W rzeczywistości aminokwasy są kodowane przez jeden kodon jak np. metionina (AUG) i tryptofan (UGG) lub przez kilka kodonów jak leucyna (UUA, UUG, CUU, CUC, CUA, CUG) czy seryna (UCU, UCC, UCA, UCG).

. 

**Dane:** sekwencja DNA

**Szukane:** zestawienie liczby kodonów leucyny (UUA, UUG, CUU, CUC, CUA, CUG) zawartych w sekwencji RNA

**Przykładowy zestaw danych:**

„TCTTGGGGCTTGTCGATTTTCGCTGGAAAGAGTCTTTTTAAGGGCTATTCCTTAACATAGACCCCGTTGCCTAAGCGAAATGATACAATTGTAAACCAAC”

**Przykładowe wyjście:**

‘{"UUA": 2, "UUG": 4, "CUU": 4, "CUC": 0, "CUA": 2, "CUG": 1}’

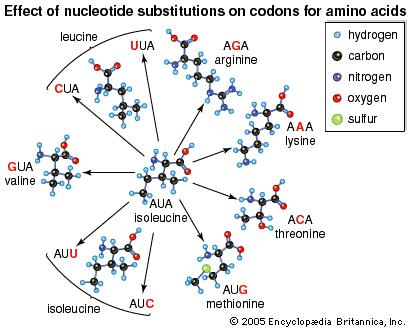
**Wskazówka:**

Zamień sekwencje DNA na RNA, następnie czytaj po 3 znaki naraz.

**Zadanie 2.**

**Problem:**

**Mutacje punktowe** określane też jako **mutacje genowe** dotyczą sekwencji nukleotydowej. Mogą one polegać na zastąpieniu jednej zasady azotowej inną. Tego typu podstawienie nazywa się substytucją lub bardziej dokładnie tranzycją (w przypadku, gdy puryna zastępowana jest inną puryną lub pirymidyna pirymidyną), bądź transwersją (zamiana zasady purynowej na pirymidynową lub odwrotnie). Prowadzą one do zmian w sekwencji aminokwasowej kodowanego białka. Mogą zatem przyczyniać się do upośledzenia jego funkcjonowania oraz zmiany budowy poprzez skrócenie łańcucha (mutacja nonsens) lub jego wydłużenie (zmiany kodonu terminacyjnego). Wyjątkiem jest tzw. mutacja synonimiczna, w efekcie której powstaje trójka nukleotydów kodująca ten sam aminokwas, co pierwotny kodon.   
Jest to możliwe dzięki zdegenerowaniu kodu genetycznego. Mutacje punktowe obejmują również utratę (delecję) oraz wstawienie (insercję) jednego lub kilku nukleotydów. Tego typu zaburzenia powstają głównie w wyniku tzw. poślizgu replikacyjnego fragmentów zawierających krótkie, powtórzone sekwencje i prowadzą do zmiany fazy odczytu kodowanych aminokwasów, czego efektem jest powstawanie zupełnie innego łańcucha polipeptydowego.



**Dane:** struktura RNA oaz trzy kodony stop: UAA, UAG, UGA

**Szukane:** czy istnieje możliwość mutacji punktowej - zmiana ramki odczytu, podanie miejsca zmiany wraz z dokonaną zmianą

**Przykładowy zestaw danych:**

„UGUCUGAGUAACUUGUCGGUCGUAUACAGUGGUAUUGUGCGGAAACAACCGUAGGGAGGUGUUACUGCUGGGAACAGCCUUUCGUUCGCCAGGCAAUAC”

**Przykładowe wyjście:**

'{"UAG": ["UUG", "UCG", "UAC", "UAG", "UCG", "UAC"],

"UAA": ["UAC", "AAA", "CAA", "UAG", "CAA", "UAC"],

"UGA": ["UGU", "GGA"]}'

**Zadanie 3.**

**Problem:**

Struktura drugorzędna RNA jest często definiowana przy użyciu notacji na pasku (ang. Dot-Bracket Notation DBN). Ważnymi strukturami   
w formacie DBN są dobrze zsynchronizowane słowa składające się z kropek '.', Otwierające nawiasy "(" i zamykające ")”. Pozycje kropkowane są niesparowane, podczas gdy dopasowanie do położeń w nawiasach jest reprezentacją nukleotydów wiązania zasad. Ponieważ liczba nukleotydów oddziałujących jest zawsze równa (każdy musi mieć partnera), nawiasy muszą być zrównoważone.  
  
Pseudowęzeł jest strukturą drugorzędową kwasu nukleinowego zawierającą co najmniej dwie struktury pętli macierzystej, w których połowa jednej macierzystej jest sprzężona między dwiema połówkami innego trzonu.. Pseudowęzeł został po raz pierwszy rozpoznany w rzepiku żółtej mozaiki wirusowej w 1982 roku. Pseudowęzeł składają się na trójwymiarowe konformacje w kształcie węzłów, ale nie są prawdziwymi węzłami topologicznymi, takimi jak matematyka. Pseudowęzły są oznaczone przy użyciu alternatywnych [..] lub {..} par kątowych.

|  |  |  |
| --- | --- | --- |
| 3-way junction ((((...((())).....((()))..))))  http://www.cs.put.poznan.pl/tzok/wiki/index.php?n=BioinformatykaStrukturalna2.AlgorytmyPrzewidywaniaStrukturyDrugorz%c4%99dowejRNA?action=download&upname=3way-junction.png | 4-way junction ((((...((())).....((()))..(()).........))))http://www.cs.put.poznan.pl/tzok/wiki/index.php?n=BioinformatykaStrukturalna2.AlgorytmyPrzewidywaniaStrukturyDrugorz%c4%99dowejRNA?action=download&upname=4way-junction.png | Pseudoknot ...[[..((((..]]....)))).. http://www.cs.put.poznan.pl/tzok/wiki/index.php?n=BioinformatykaStrukturalna2.AlgorytmyPrzewidywaniaStrukturyDrugorz%c4%99dowejRNA?action=download&upname=pseudoknot.png |

**Dane:** struktura zapisana w notacji dot-bracket

**Szukane:** maksymalny stopień zagnieżdżenia pseudowęzłów oraz największy występujący w tej strukturze stopień skrzyżowań

**Przykładowy zestaw danych:**

"(....((()...().......()....()...().......)).)Aa(....()...()....().......()....()........()...().......))...)(()...()......()........()....().....()......()..){..}<...>[...](..)((....))"

**Przykładowe wyjście:**

'{"junction\_type": 8, "junction\_count": 2, "pseudoknot\_complexity": "Aa", "pseudoknot\_count": 1}'

**Zadanie 4.**

**Problem:**  
Notacja kropkowo-nawiasowa (ang. dot-bracket) jest wygodnym sposobem reprezentowania struktury drugorzędnej. Każdy znak reprezentuje bazę. Otwarte nawiasy wskazują, że baza jest połączona z inną bazą przed nią. Zamknięte nawiasy wskazują, że baza jest powiązana z inną bazą za nią. Okresy lub kropki wskazują na niesparowaną podstawę. Liczba otwartych i zamkniętych nawiasów zawsze będzie równa. Niektóre bazy danych struktury drugorzędnych obejmują inne znaki ([], {}, <>, a, itd.) W celu reprezentowania parowania w pseudowęzłach.

Drugorzędowa struktura RNA jest tworzona przez parowanie zasad między regionami komplementarnymi.



**Dane:** funkcja (poprawna struktura drugorzędowa (dot-bracket), przewidziana struktura drugorzędowa (dot- bracket))oceniająca poprawność przewidzianej struktury, jako miarę przyjmując procent poprawnie przewidzianych pozycji (nukleotyd niesparowany / otwierający / zamykający parę, wchodzący w skład pseudowęzła).

**Szukane:** zestawienie poprawnej struktury z przewidzianą oraz wartość liczbowa precyzująca procent pozycji identycznych.

**Przykładowy zestaw danych:**

'["(.....(<.>A...a(..().......()........()........()..)

(....([](((....)))(...(((((.....)))){...}(...(",

"(.....(<.>A...a(..().......()........()........()..)

(....([](((....)))(...(((((.....)))){...}(...("]'

**Przykładowe wyjście:**

'{"(": "100.00%", ")": "100.00%", ".": "100.00%", "pseudoknot": "100.00%"}'