Лабораторная работа 2

Вариант 9. Нахождение участков рестрикции

**Цель:** Разработать программу, которая идентифицирует реверсивные палиндромы (участки рестрикции) в заданной последовательности ДНК.

**Задача:** Реализовать алгоритм, который принимает на вход строку ДНК в формате FASTA и возвращает позиции и длины всех реверсивных палиндромов в этой строке, имеющих длину от 4 до 12 символов.

**Инструменты:**

**Язык программирования:** Python (хорош для обработки строк и последовательностей, обладает простым синтаксисом)

**Среда разработки (IDE):** vs code

**Библиотеки:** BioPython - для разбора FASTA файлов (если нужно считывать данные из файла, а не из строки). Если входные данные всегда будут строкой ДНК, то можно обойтись и без BioPython.

**Алгоритмы:**

1. **Создание функции для вычисления обратной комплементарной последовательности:**

Функция должна принимать на вход строку ДНК и возвращать её обратную комплементарную последовательность.

Для этого нужно:

Заменить каждый нуклеотид на его комплементарный: A -> T, T -> A, C -> G, G -> C.

Перевернуть строку.

2.**Создание функции для проверки, является ли строка реверсивным палиндромом:**

Функция должна принимать на вход строку ДНК и возвращать True, если она является реверсивным палиндромом, и False в противном случае.

Для этого нужно:

Вычислить обратную комплементарную последовательность строки.

Сравнить исходную строку с обратной комплементарной. Если они совпадают, строка является реверсивным палиндромом.

3. **Поиск реверсивных палиндромов:**

Перебрать все возможные подстроки в строке ДНК, имеющие длину от 4 до 12 символов.

Для каждой подстроки проверить, является ли она реверсивным палиндромом, используя функцию из шага 3.

Если подстрока является реверсивным палиндромом, вывести её позицию и длину

**Ошибки и способы их предотвращения:**

**Неправильная реализация функции** reverse\_complement**:** Ошибки в замене нуклеотидов или при перевороте строки приведут к неправильным результатам.

**Решение:** Тщательно протестировать функцию на небольших примерах, для которых известен правильный результат.

**Неправильный диапазон длин палиндромов:** Неправильный диапазон длин приведет к тому, что некоторые палиндромы будут пропущены или будут найдены палиндромы неправильной длины.

**Решение:** Убедиться, что цикл for length in range(4, 13) перебирает все нужные длины.

**Выводы:**

Задача нахождения участков рестрикции является важной в биоинформатике, так как ферменты рестрикции широко используются в молекулярной биологии. Алгоритм решения этой задачи достаточно прост и может быть реализован на Python с использованием библиотеки BioPython (или без неё, если входные данные представлены в виде строки).