Отчет о выполнении курсовой работы по дисциплине “Программирование биоинформатических приложений на суперкомпьютере”

Получение списка белков, содержащихся в геноме SARS-CoV-2

Агаев А.Ф.

гр. 3540904/00202

**Исходные данные**

Сборка генома, который рассматривался в рамках работы, в формате FASTA доступна по ссылке:  
<https://www.ncbi.nlm.nih.gov/nuccore/MT135041.1?report=fasta> .

**Реализация**

Программа реализована на языке Python 3.

Открытая рамка считывания (Open Reading Frame) — последовательность нуклеотидов в составе ДНК или РНК, потенциально способная кодировать белок. Основным признаком наличия ORF служит отсутствие стоп-кодонов (в случае РНК — обычно UAA, UGA и UAG) на достаточно длинном участке последовательности после стартового кодона (в подавляющем большинстве случаев — AUG).

В рамках работы понятие ORF было использовано для поиска всех потенциальным белков, содержащихся в геноме.

После получения списка потенциальных белков, полученный список сравнили с уже определенными белками для этого генома (<https://www.ncbi.nlm.nih.gov/nuccore/MT135041.1?feature=CDS>). Все известные белки были найдены в результате работы программы.

Результаты приведены в приложении 1.

**Подсчет молекулярной массы**

Для каждого из потенциальных белков была подсчитана молекулярная масса.   
Результаты приведены в приложении 2.

**Построение модели**

Для построение модели трехмерной структуры был выбран белок

{'MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFGWLIVGVALLAVFQSASKIITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLEAPFLYLYALVYFLQSINFVRIIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPYNSVTSSIVITSGDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL': 31084.810649999978}

Модель была получена с помощью инструмента SWISS-MODEL. Результаты приведены ниже и доступны по ссылке:

<https://swissmodel.expasy.org/repository/uniprot/P0DTC3>

