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

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

Малов И.А.

гр. 3540904/00202

**Исходные данные**

Используемая сборка генома:

https://www.ncbi.nlm.nih.gov/nuccore/MT123291.2

**Реализация**

Программа была написана на языке scala.

Для входной днк были найдены все открытые рамки считывания (Open Reading Frame) со стартовым кодоном AUG и завершающими UAG, UGA, UAA. Всего было найдено 1078 рамок. Затем каждая ORF была преобразована в потенциальный белок. Для каждого белка была посчитана молекулярная масса.

Все потенциальные белки находятся в файле proteins.txt

Массы белков находятся в файле masses.txt

**Построение модели**

Для построения модели трехмерной структуры был выбран белок:

MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIELCVDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGSLVVRCSFYEDFLEYHDVRVVLDFI

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

<https://swissmodel.expasy.org/repository/uniprot/P0DTC8>

