**Проект: поиск специфических последовательностей в ДНК с использованием локальных и сетевых ресурсов.**

Входные данные:

1. строка кода ДНК
2. ДНК последовательность в файле (txt, fasta)
3. ДНК из баз данных (GeneBank, Ensembl)

Функционал:

1. Форматирование последовательности: удвоение не ДНК символов, сообщение о таковых. Возможность генерации рандомных последовательностей с заданными параметрами.
2. Поиск однобуквенных повторов разных длин – заданной пользователем одной длинны или в диапазоне
3. Поиск «скользких» сайтов по известным паттернам
4. Поиск тандемных повторов (типа atgcatgc) с задаваемой длиной повтора, или диапазоном длин сегмента повтора, в возможностью поиска вложенных повторов
5. вычисление GC% как для всей последовательности, так и с разбивкой по сегментам поисках (длина сегментов должна быть регулируема). Также хотелось бы визуализацию в виде графика
6. Вычисления констант энерший и констант диссоциаций для интересующих выбранных коротких участков (по имеющимся формулам – через NN и HMM)
7. Через подключение каким-то образом к онлайн ресурсу найти палиндромы, проверить на разделенные палиндромы, вывести энергии палиндромов и рассчитать константы диссоциации
8. через другой сервис проверить на G-структуры
9. если длинная последовательность (например геном) найти в каком гене находится найденные неканоничные структуры (через blast) если не найдены, то увеличить искомый фрагмент, пока не будет найден ближайший ген/любой аннотированый участок
10. вывести результат поиска в виде csv, текстового отчета, инфорграфики с цветным маркированием найденных последовательностей на искомой последовательности
11. Реализовать в виде работы в коммандной строке, окне, веб-страницы