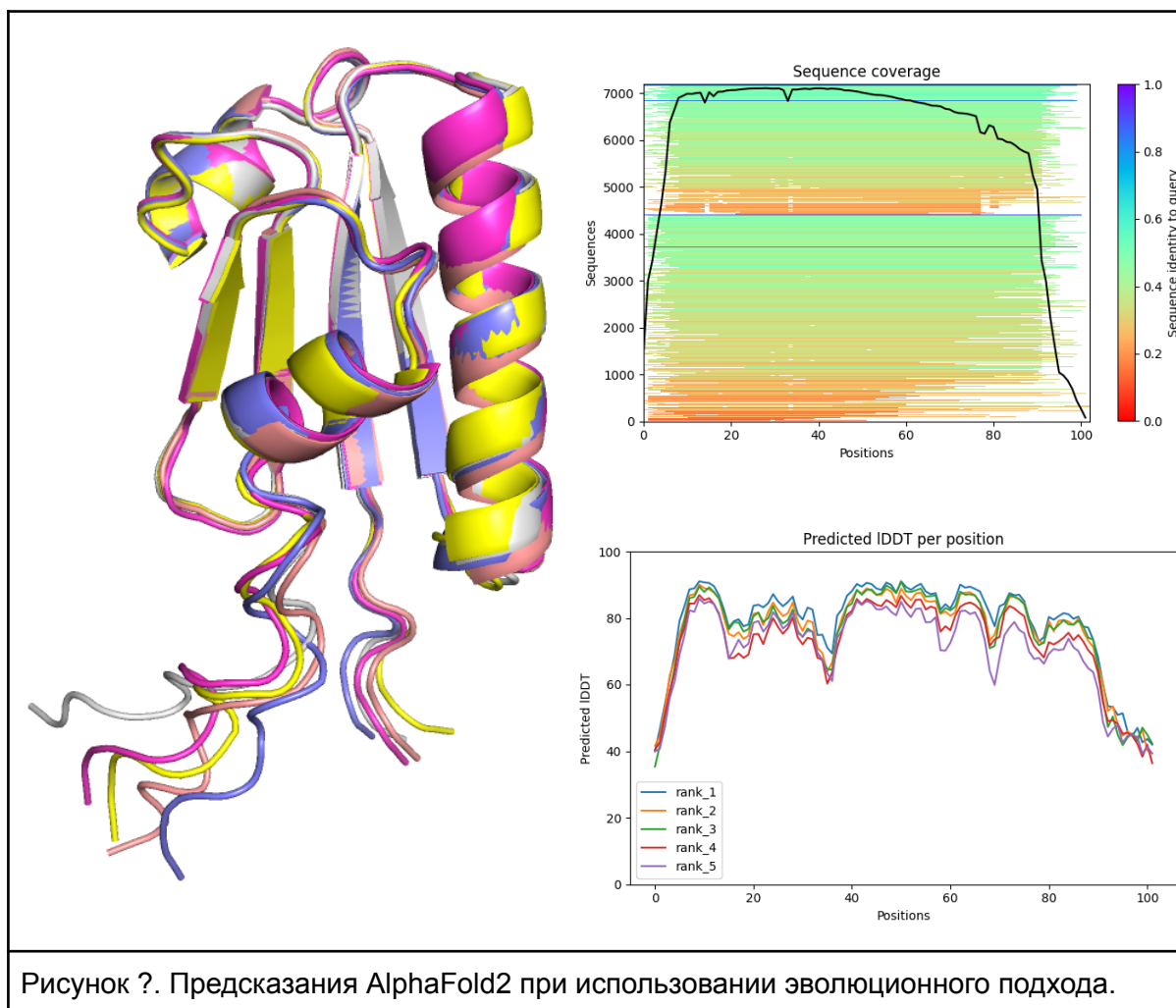


Введение

В рамках данной задачи необходимо протестировать способность AlphaFold2 предсказывать структуру метаморфных белков, которые в условиях физиологической нормы могут принимать несколько структур, находящихся в динамическом равновесии в системе. Оценка качества работы сервиса будет проведена на примере белка KaiB, который является белком циркадного ритма цианобактерий (конкретная референсная последовательность принадлежит белку *Synechococcus elongatus* PCC 7942).

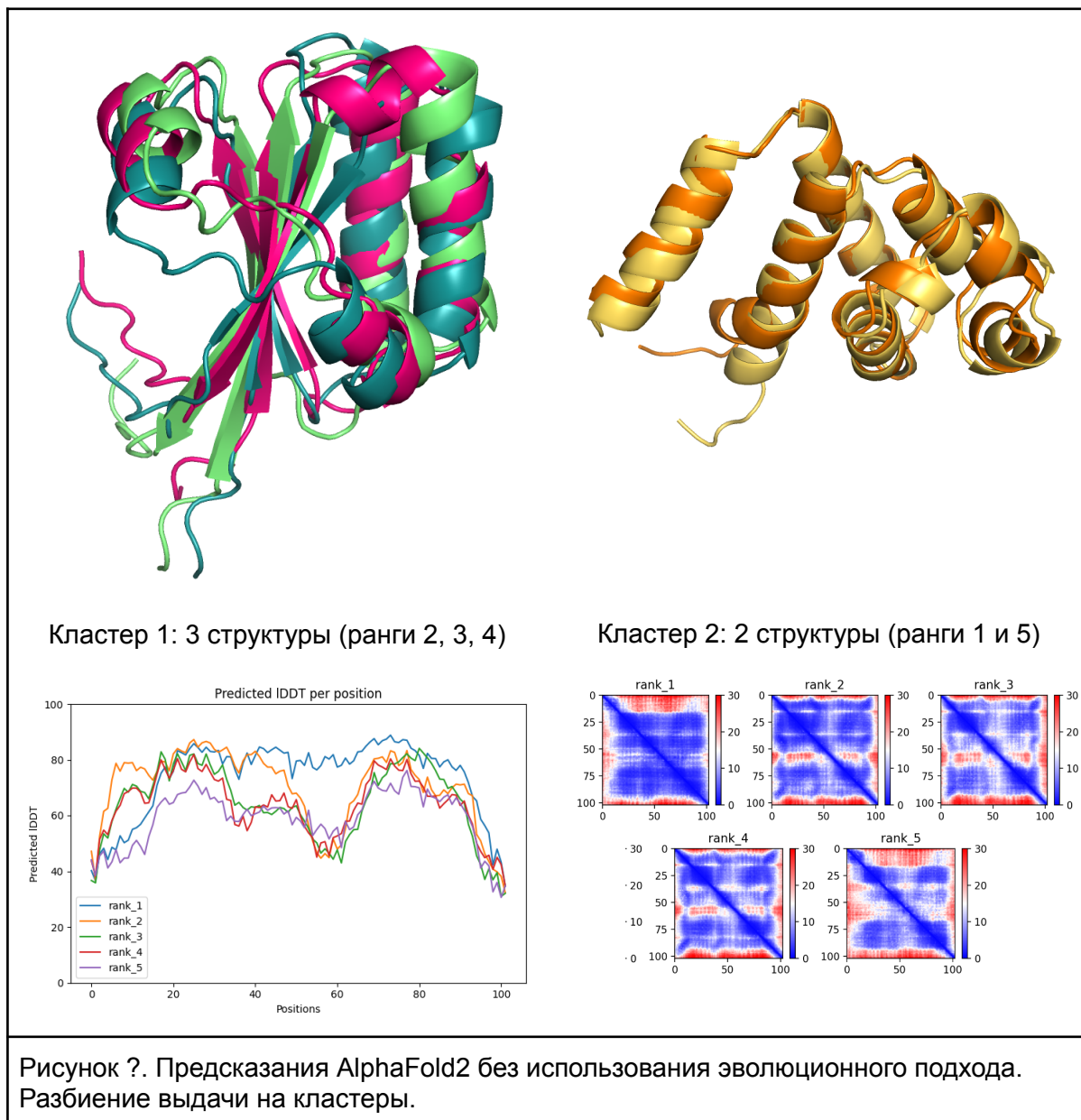
Результаты работы AlphaFold2

Запустим алгоритм и воспользуемся эволюционным подходом, позволив пакету построить выравнивания и взять за основу моделирования те или иные структуры, достаточно гомологичные исходной по аминокислотной последовательности.



Мы получили пять структурных предсказаний, которые в значительной степени похожи друг на друга и явно отражают всего лишь одно из возможных состояний метаморфного белка. Структура предсказана с достаточно неплохим качеством, которые снижается лишь для концевых участков полипептидной цепи, где, вероятно, располагаются в целом подвижные петли.

Повторим запуск, но не будем использовать эволюционный подход. Предскажем структуру, основываясь только на аминокислотной последовательности. Стоит отметить, что в данном случае алгоритм на выходе выдает на порядок большее разнообразие структур, чем в первом варианте. В целом, их можно разбить на два кластера, ни один из которых не соответствует моделям, полученным при использовании эволюционного подхода.



Стоит отметить, что структуры также интересно кластеризуются: модели с рангами 1 и 5 похожи между собой (хотя ожидалось скорее разделение по качеству моделей). В данном случае общее качество предсказаний значительно падает. В целом, стоит отметить, что использование именно такого подхода позволяет получить более “разные” структуры, что и может отражать конформационные переходы метаморфных белков.

Похожи ли полученные модели между двумя подходами?

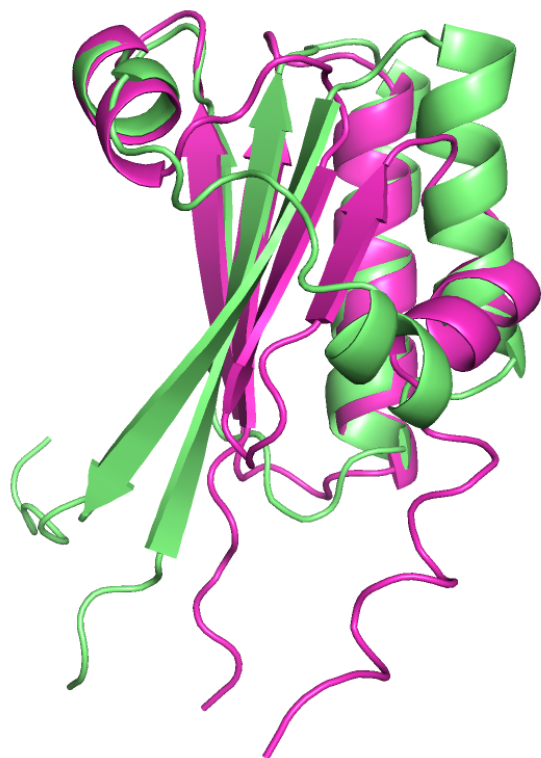


Рисунок ?. (Не)сходство моделей, полученных разными подходами. Зеленым – структура кластера 1, розовым – одна из структур, предсказанных ранее.

Модели из кластера 2 очевидным образом не имеют ничего общего со структурами-результатами первого запуска, а вот модели второго похожи гораздо сильнее. Наложим структуры друг на друга и увидим, что, несмотря на общее сходство, структуры отличаются друг от друга. Так, в неэволюционной структуре мы видим удлинение β -листа в одном направлении и укорочение его же в другом, появление достаточно длинных α -спиралей и исчезновение коротких участков с той же вторичной структурой.

Какие выводы можно сделать из данных запусков?

AlphaFold2 достаточно хорошо предсказывает структуры белков, для которых есть структуры каких-либо гомологов с достаточно хорошей степенью идентичности последовательностей, однако в данном случае задача превращается в задачу гомологичного моделирования, которая, очевидно, проще, чем моделирования de-novo. Однако и в случае моделирования de-novo мы получаем достаточно внятные модели, которые действительно могут отражать нативную структуру белков.

Сравнение с референсными структурами

Оценим предсказательную силу инструмента с помощью сравнения смоделированных структур с реальными. Для начала с помощью BLAST по базе данных PDB подберем структуры, имеющие схожую (или одинаковую) аминокислотную последовательность и проанализируем их структуры.

ID структуры	Организм	Тип	Метод получения
4KSO	Synechococcus elongatus PCC 7942	1	X-ray, 2.62 Å

1R5P	Nostoc sp. PCC 7120	1	X-ray, 2.2 Å
5JWQ	Thermosynechococcus vestitus BP-1	2	X-ray, 3.87 Å
5JYT	Thermosynechococcus vestitus BP-1	1	NMR

Рассмотри две формы для данного белка:

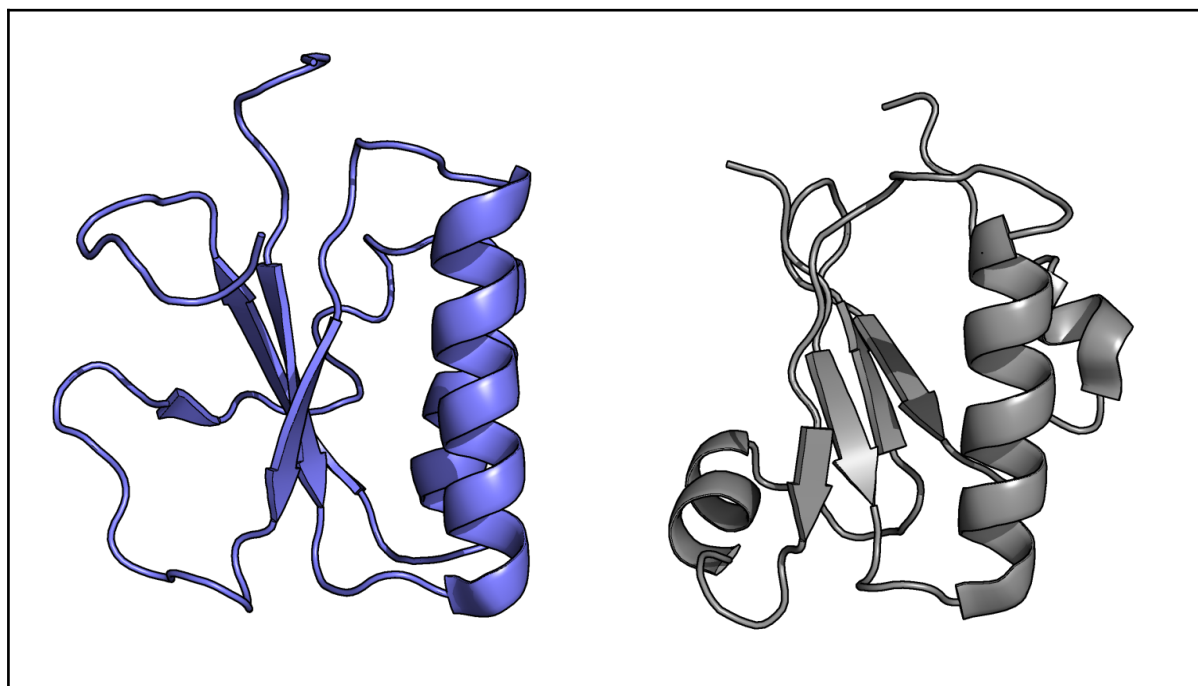


Рисунок ?. Две формы метаморфного белка.

Мы видим явные схожести с теми структурами, которые смоделировал нам AlphaFold2. Наложим модели на структуру и посмотрим, насколько сильно они отличаются друг от друга.

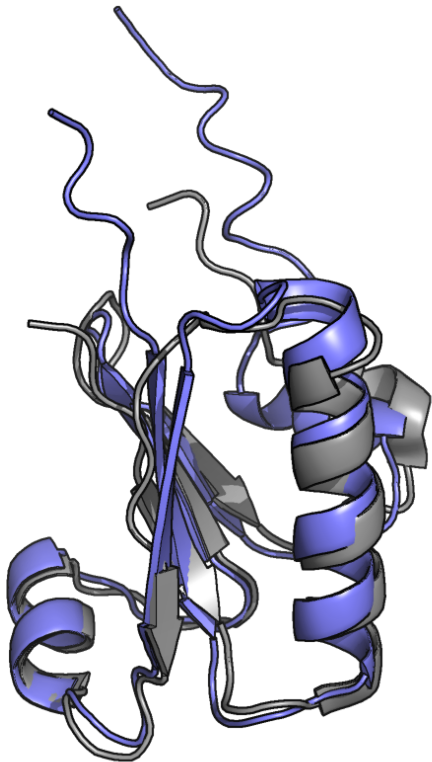
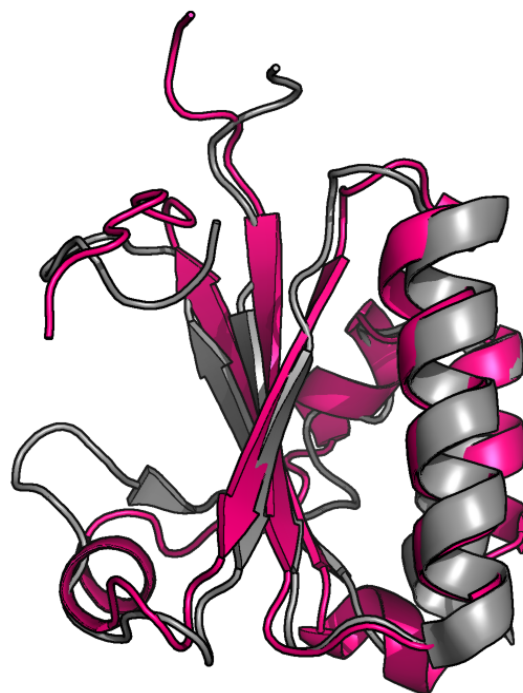


Рисунок ?. Выравнивание одной из форм метаморфного белка с лучшей моделью, построенной AlphaFold2 с использованием эволюционных методов и выравнивания. Серым – структура из PDB, синим – AlphaFold2-модель.

Стоит отметить, что структуры достаточно хорошо выравниваются друг на друга, алгоритм справился с построением основных элементов белка, однако не идеально. В некоторых местах мы видим более протяженные β -листы и α -спирали. Здесь стоит отметить, что структура получена методом рентгеноструктурного анализа, а в водном растворе ее структура может отличаться от того, что мы видим в кристалле.

Рисунок ?. Выравнивание одной из форм метаморфного белка с лучшей моделью, построенной AlphaFold2 без использования эволюционных методов. Серым – структура из PDB, синим – AlphaFold2-модель из кластера 1.

Модели, построенные AlphaFold2 без использования эволюционных методов, напротив, больше похожи структурно на вторую форму белка. Структуры также имеют некоторые расхождения, но в целом, модель также отражает основные особенности структурной организации белка.



Найти примеры структуры, полученной как кластер 2 при втором запуске, мне не удалось. Вероятно, это является артефактом моделирования.

Заключение

В целом, стоит отметить, что AlphaFold2 действительно является одним из наиболее перспективных инструментов, разработанных в последнее время. Алгоритм подходит для выполнения большого спектра практических задач, однако он, как и любой другой метод, может ошибаться, поэтому все результаты моделирования должны быть проанализированы на предмет ошибок и неточностей.

Дополнение: AlphaFold2 как советчик в генно-инженерных манипуляциях

В прекрасном будущем наша научная группа хочет заниматься исследованиями ионо-специфичности АТФ-синтаз и определить, какие аминокислотные остатки отвечают за тип переносимого через мембрану катиона (протон или катион натрия). Для этого, логично, нам требуются две генно-инженерные конструкции: с ферментом, переносящим протон, и с ферментом, переносящим натрий. Протонная АТФ-синтаза имеется у нас в распоряжении, а вот натриевую необходимо будет собрать, подняв нужные гены с генома с помощью ПЦР.

Общая длина оперона составляет приблизительно 7000 п.о., поэтому поднять такую конструкцию с помощью одного ПЦР не получится, необходимо провести несколько последовательных ПЦР, кусочки которых будут позже соединены вместе с помощью последовательных рестрикций и лигирований.

В перспективе, нам будет необходимо работать с очищенным препаратом белка, поэтому его необходимо будет выделять методом аффинной хроматографии. Для выделения необходимо прикрепить гексагистидиновый тег на одну из субъединиц, для этих целей обычно используют субъединицу β . Однако наша схема клонирования предполагает клонирования между следующими субъединицами δ и α , а после α и γ , т.е. в данной схеме клонирования нам было бы удобно вставить тег на N-конец субъединицы α . Можем ли мы это сделать?

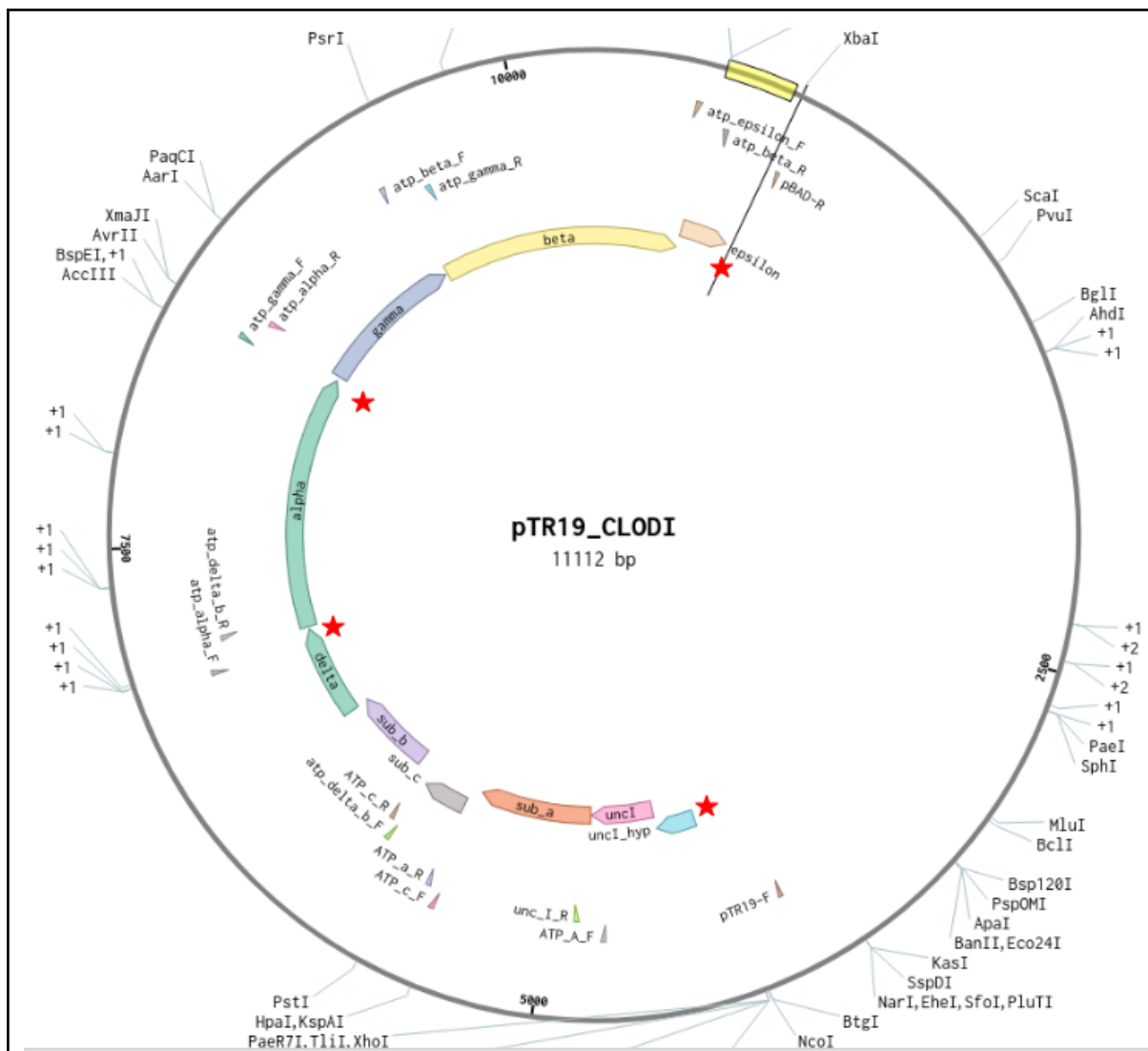


Рисунок ?. Предполагаемая схема конечного оперона. Звездочками отмечены места сшивок внутри оперона и между опероном и вектором.

Данный подход вызывает опасение, так как как минимум для некоторых структур АТФ-синтаз показано, что N-конец субъединицы α “ложиться” на субъединицу δ , поэтому введение тега может привести к тому, что фермент просто не соберется.



Рисунок ?. Структура фермента *Bacillus sp.* PS3, зеленым – субъединица α , фиолетовым – δ , голубым – β .

Промоделируем AlphaFold2 данные субъединицы, выровняем их на референсную структуру *Bacillus sp.* PS3 и посмотрим, может ли там появиться эта концевая α -спираль.

И хоть AlphaFold2 не укладывает спираль на субъединицу δ , он все равно ее моделирует, причем довольно уверен в том, что смоделировал ее верно. Скорее всего эти результаты не ошибочны, и фермент *Clostridioides difficile* имеет терминальную α -спираль, контактирующую с субъединицей δ . От такого варианта клонирования мы отказались.

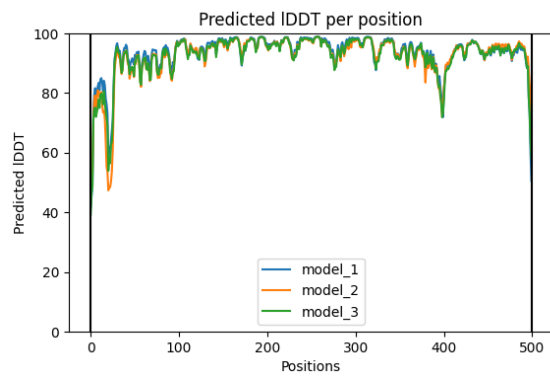
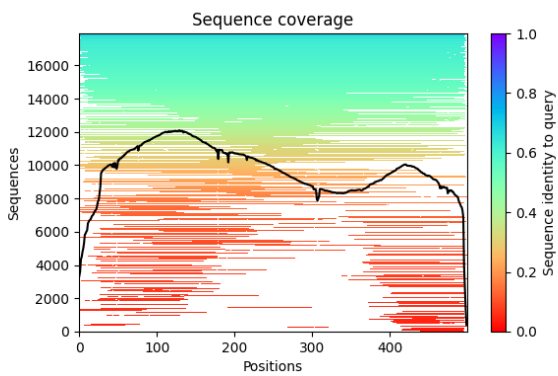
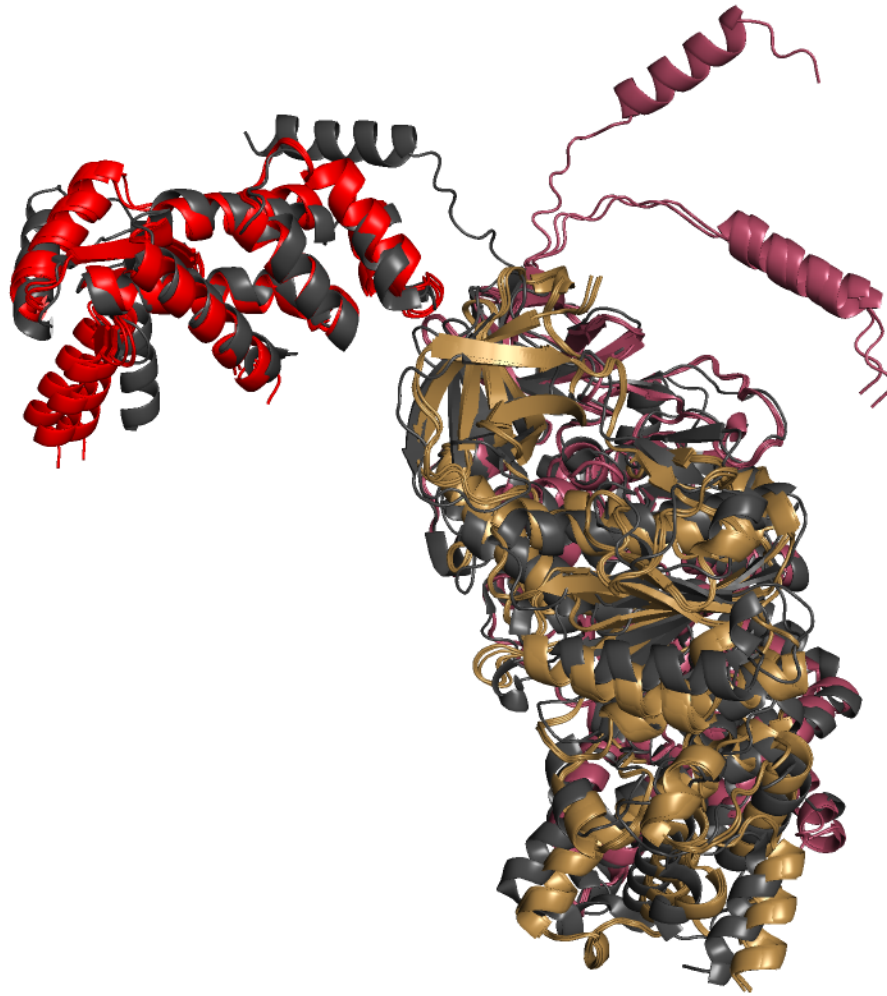


Рисунок ?. Предсказания AlphaFold2 для трех субъединиц клостридии. Серым – референсная структура для структурного выравнивания, винным – субъединица α , красным – δ , песочным – β .