

ДОМАШНЕЕ ЗАДАНИЕ 3 - РАБОТА С COLABFOLD И АЛГОРИТМОМ DOMAK

НЕВЕРОВА-СИМЧИТ ЕЛЕНА, 402 ГРУППА

ЗАДАНИЕ 1 – РАБОТА В COLABFOLD

Для выполнения выбрала задачу с амилоидами. Амилоид A9 с последовательностью:

SVQIVYKPVDSLKVTSKCGSLGNIHNKPGGGQVEVKSEKLDKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFR

Выполнение задачи начала с того чтобы загрузить мономер в [программу](#) ColabFold. Далее загружала гомодимеры из 5 и 10 молекул на основе этой же последовательности, разделив их “:”.

АРХИВЫ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

- [Мономер](#)
- [Структура из 5 молекул](#)
- [Структура из 10 молекул](#)

ОПИСАНИЕ РЕЗУЛЬТАТОВ

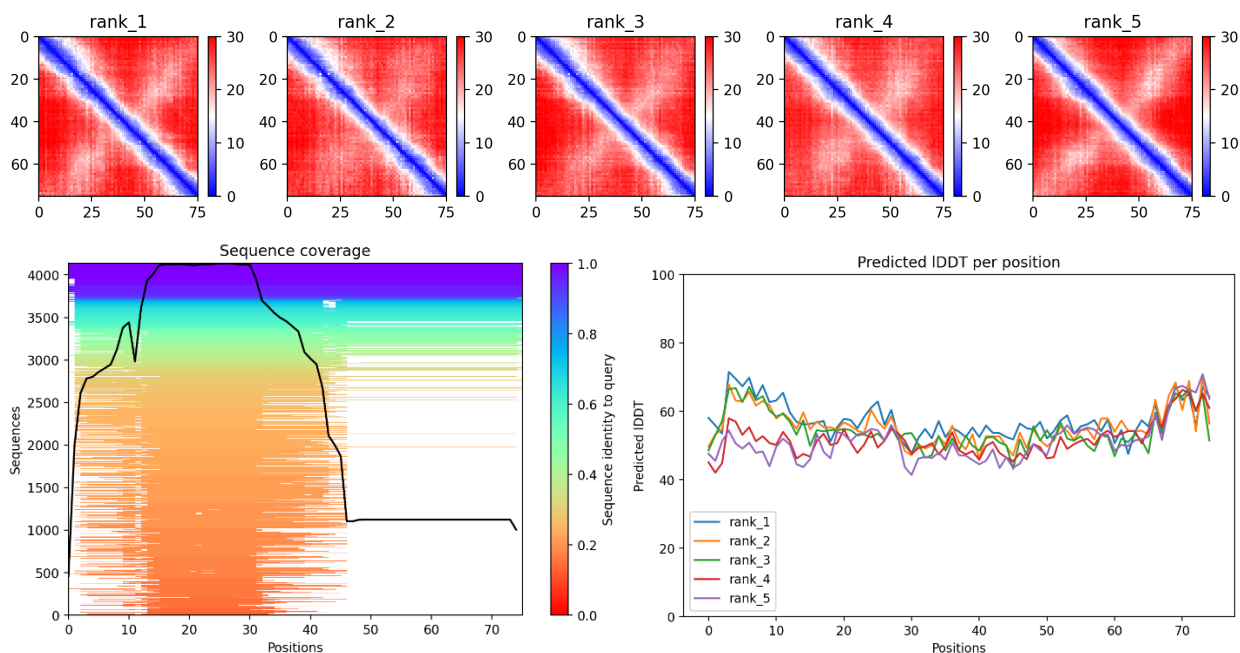
Далее представлено описание полученных результатов и обсуждение того, получилось ли у AlphaFold предсказать структурные особенности амилоидов.

СТРУКТУРА МОНОМЕРА

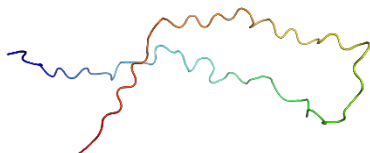
РЕЗУЛЬТАТЫ

Программа выдала нам три графика:

- Верхний – Тепловая карта показывает ожидаемую ошибку выравнивания (в ангстремах, Å) для каждой пары остатков в предсказанной модели. Видно, что большинство областей – красные (>15Å). То есть взаимное расположение цепей предсказано с низкой уверенностью. Но есть и достаточное количество пар остатков со средней точностью предсказаний (белые).
- Нижний слева – График, показывающий количество неидентичных последовательностей в множественном выравнивании, которые были использованы для предсказания, для каждой позиции в последовательности. Половина остатков показывают высокое количество найденных для них гомологов и, по сути, предсказание должно быть надежным. Вторая часть белка и особенно концы – не столь «поддерживаемые» эволюционными данными позиции.
- Нижний справа – pLDDT (predicted Local Distance Difference Test) показывает насколько мы «уверены» в положении а.о. Значения варьируются от 40 до 80 для мономера, что соответствует либо тому, что все совсем плохо и мы вторичную структуру не верим, либо для лучших значений – верим только в расположение остова. Также видно, что pLDDT для структур rank_4 и rank_5 значительно меньше.



РАССМОТРЕНИЕ СТРУКТУРЫ



Слева представлена структура, которая считается лучшим предсказанием (окраска chainbows, представление cartoon). Она представлена в виде случайного клубка без стабильных вторичных структур (альфа-спиралей или бета-тяжей). Но на самом деле, подобное предсказание вполне логично, так как амилоидные свойства (поперечно-бета-слоистая структура) проявляются только при сборке в олигомеры и фибриллы, и их отсутствие у мономера — ожидаемый и правильный результат.

ВЫВОД

Результаты предсказания для мономера корректны и ожидаемы. AlphaFold предсказал белок в неагрегированном, преимущественно неупорядоченном состоянии (случайный клубок), что соответствует его растворимой форме *in vivo*. Низкие значения pLDDT (40-80) и высокие ошибки на PAE-карте для мономера отражают присущую ему структурную гибкость и отсутствие жесткого глобулярного ядра. Это не ошибка алгоритма, а верное предсказание нативного состояния пептида, которое предшествует амилоидогенезу. Таким образом, для мономера не удалось (и не ожидалось) предсказать амилоидные черты, что полностью согласуется с биологией процесса.

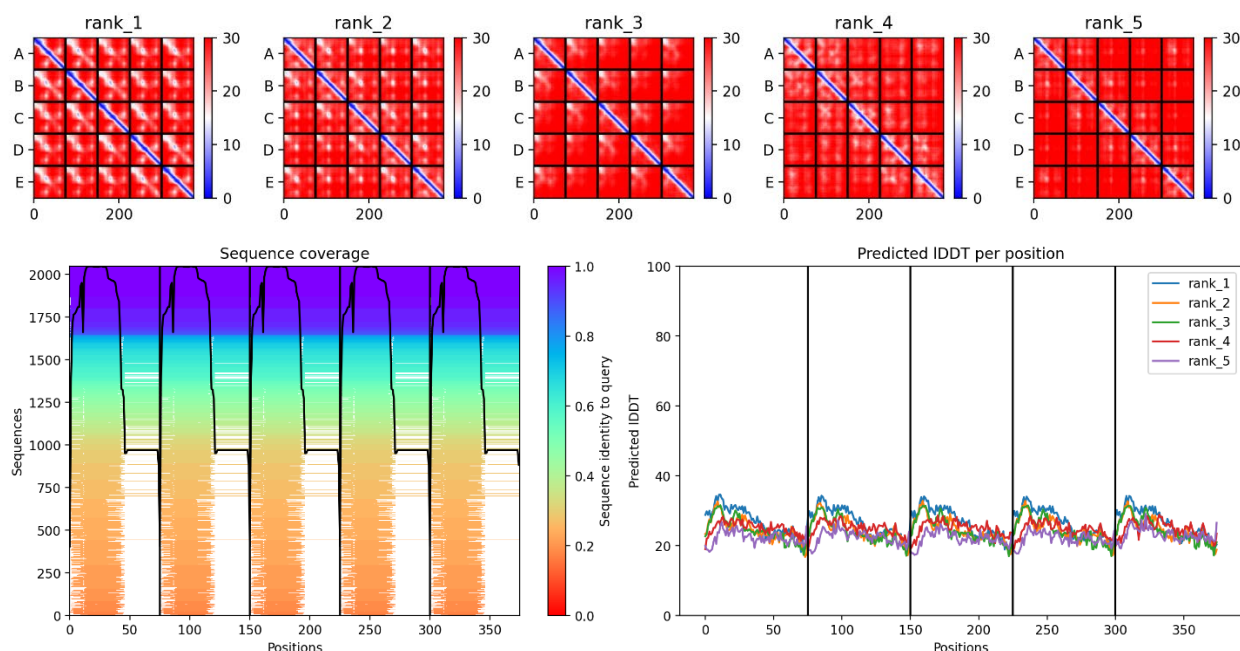
СТРУКТУРА ИЗ 5 МОЛЕКУЛ

РЕЗУЛЬТАТЫ

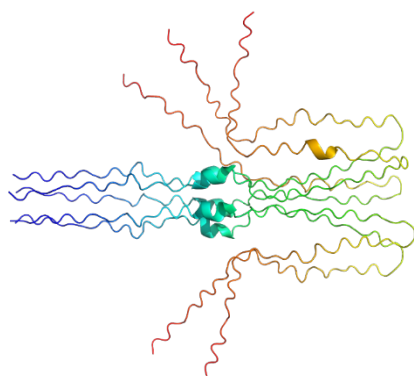
У нас снова есть три графика:

- Верхний — сохраняется высокое содержание красных участков на тепловой карте, при этом нужно учитывать, что теперь у нас не мономер, а пентамер, то есть должна наблюдаться корреляция между разными доменами — наилучшим образом это видно на примере самого качественного предсказания — rank_1 (там появляются белые участки).
- Нижний слева — тенденции уверенности в отдельных остатках сохраняются в пределах единичного домена, при этом абсолютное количество последовательностей, гомология с которыми обнаруживалась, упало в 2 раза.

- Нижний справа – pLDDT (predicted Local Distance Difference Test) упал, если раньше у нас были значения выше 70, которые обеспечивали нам «уверенность» в остове, то теперь «сомнительность» приобрела вся структура: pLDDT не выше 40.



РАССМОТРЕНИЕ СТРУКТУРЫ



Здесь уже появляются первые признаки организации. В структуре можно наблюдать участки, показанные cartoon в виде альфа-спиралей. Окраска chainbows позволяет убедиться, что при этом есть некоторая симметрия в распределении одинаковых частей пентамера. AlphaFold уловил тенденцию к агрегации, но не смог предсказать стабильную, высокоупорядоченную амилоидную укладку для олигомера из 5 цепей. Возможно, модель представляет собой нестабильный промежуточный агрегат или одну из множества возможных конформаций, что отражает полиморфизм и сложность процесса нуклеации амилоидов.

ВЫВОД

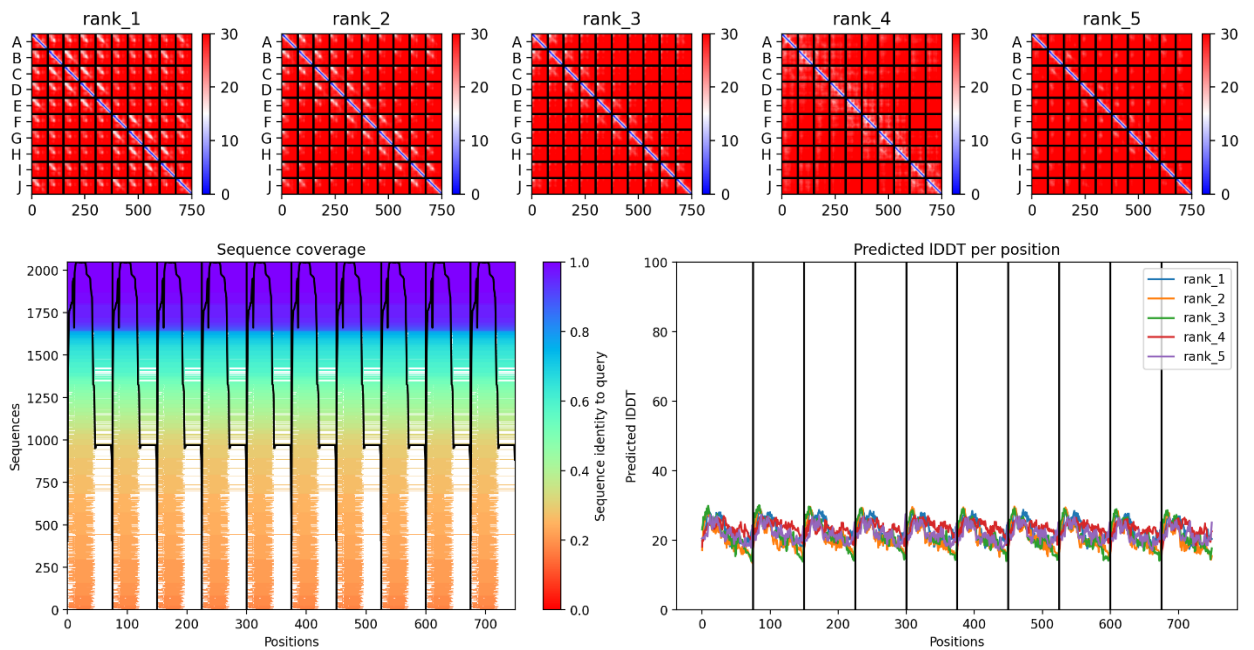
Предсказание для пентамера демонстрирует переходное состояние. Алгоритм фиксирует факт агрегации (появление симметрии и мультимерных блоков на PAE), но не может выявить единую, энергетически выгодную укладку, характерную для зрелого амилоида. Резкое падение общей уверенности (pLDDT < 40) и преобладание красных зон на PAE указывают на высокую неопределенность модели.

СТРУКТУРА ИЗ 10 МОЛЕКУЛ

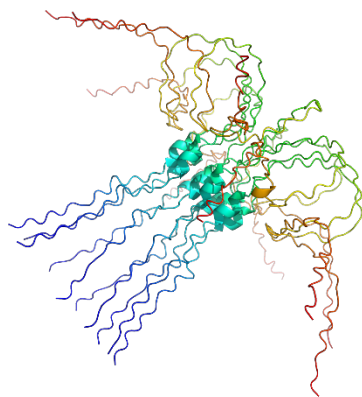
РЕЗУЛЬТАТЫ

Еще раз анализируем наши три графика:

- Верхний – общая картина распространяется не на 5 одинаковых частей, а 10. При этом наибольшая корреляция наблюдается именно между соседними единицами гомодимера.
- Нижний слева – тенденции сохраняются и даже не происходит падения количества гомологичных последовательностей.
- Нижний справа – наша «уверенность» по pLDDT падает на еще несколько единиц.



РАССМОТРЕНИЕ СТРУКТУРЫ



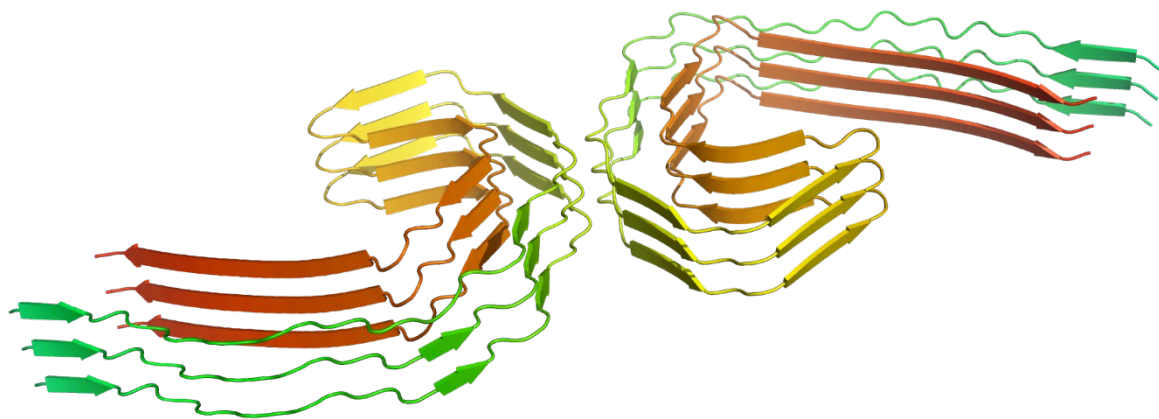
Структура напоминает пентамер, сохраняется идентичность распределения отдельных мономеров гомодимера. Количество участков вторичной структуры не увеличилось. К сожалению, формирование ядра из параллельных бета-листов не наблюдается, то есть особенность амилоидов в виде архитектуры cross-beta отсутствует.

ВЫВОД

Увеличение размера сборки до 10 субъединиц не привело к предсказанию классической амилоидной архитектуры. Вопреки ожиданиям, модель не показала формирования протяженного, плотно упакованного ядра из параллельных бета-листов (cross-beta). Вместо этого наблюдается простое «масштабирование» неуверенной укладки, предсказанной для пентамера, с дальнейшим снижением pLDDT. Это позволяет сделать ключевой вывод: Для данной конкретной амилоидогенной последовательности (A9) AlphaFold/ColabFold не смог достоверно предсказать стабильную структуру амилоидной фибриллы даже для крупной сборки.

PDB ДАННЫЕ

Для сравнения была рассмотрена официально представленная структура моего пептида: [6HRE](#). Здесь уже наблюдается характерная структура амилоидов: у нас наблюдаются продольные бета-тяги на разных участках последовательности. Некоторые бета-листы сцеплены между собой. Есть сэндвич-подобные участки.



AlphaFold обучался преимущественно на глобулярных белках и стабильных комплексах, и его архитектура (особенно механизм MSA и эволюционного сдерживания) может быть недостаточной для однозначного предсказания амилоидного полиморфизма – способности одной последовательности укладываться в несколько различных, но стабильных, кристаллографических форм. Кроме того, в экспериментальных условиях фибрилла стабилизируется в одной конкретной упаковке под влиянием факторов среды (pH, ионная сила, поверхность раздела), которые модель не учитывает. Для изучения амилоидов необходима осторожная интерпретация результатов AlphaFold, обязательно с последующей валидацией экспериментальными методами.

ЗАДАНИЕ 2 – АЛГОРИТМ ДОМАК

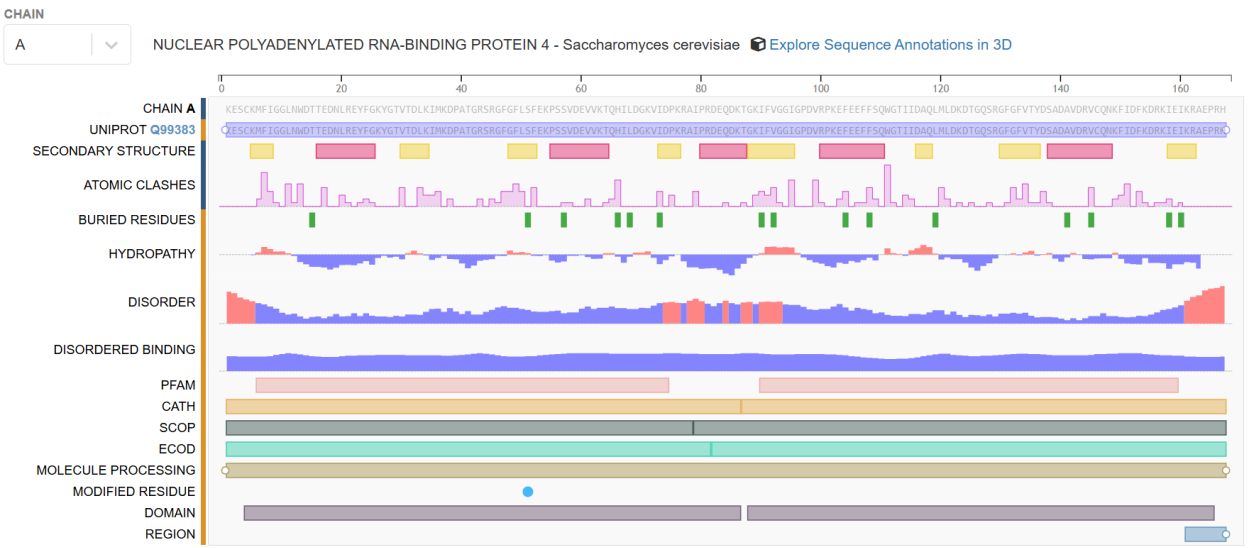


Мой [колаб](#). Тут возникли некоторые сложности при дописании кода, но в итоге все получилось. Разделение моего белка на домены алгоритмом ДОМАК [2С1К](#) было получено. Только нумерация остатков почему-то не с 1, а со 156. Найденные пики представлены в таблице ниже. Распределение представлено на графике.

Пик	Позиция по коду	Реальная позиция	Значение split_value
1	160	5	14.3200
2	170	15	1.6548
3	183	28	2.4480
4	208	53	17.2117
5	219	64	29.8761
6	233	78	928.0000
7	265	110	4.1200
8	279	124	2.1966

9	292	137	11.2996
10	302	147	7.5611
11	319	164	246.8889

ЗАДАНИЕ 3 – SCOP И CATH



Согласно базам данных CATH и SCOP белок делится также на 2 домена, но сами БД не совпадают в оценке остатка, по которому происходит разделение. Разметки на домены представлены в таблице ниже.

	Домен 1	Домен 2
CATH	1-86	87-167
SCOP	1-78	79-167
DOMAK	1-78	79-167

Получается, предсказание алгоритма DOMAK в моем случае совпадает с данными о структурных доменах от БД SCOP.