

Альфафолд2

- Цель данного практикума - с помощью проекта ColabFold познакомиться с искусственным интеллектом, способным предсказывать трехмерные структуры белков.

AlphaFold2 – нейронная сеть, разработанная компанией DeepMind. Известность сеть обрела благодаря соревнованиям по предсказаниям трехмерных структур CASP 2020 года, где команда, представившая искусственный интеллект, победила с огромным перевесом. Как конкретно работает AlphaFold2 неизвестно, предположения о его работе вытекают из первой версии нейронной сети AlphaFold, которую компания создала ранее. На вход сеть получает последовательность, на основе которой выполняется множественное выравнивание с известными белками. Вероятно, если принцип работы сетей схож, на основе данных о корреляциях позиций AlphaFold2 получает распределение вероятностей парных расстояний. Это является ключевым моментом в предсказании модели.

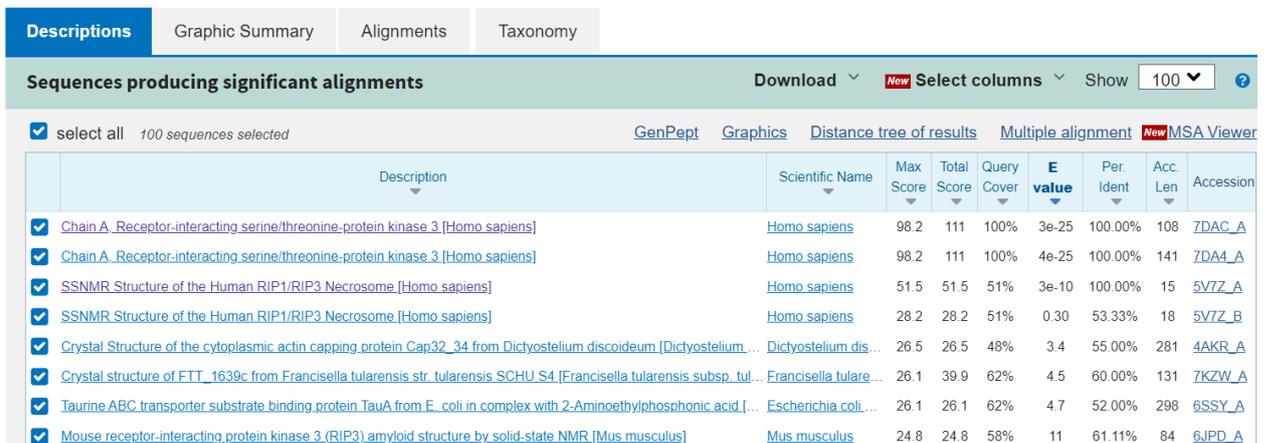
Ознакомиться подробнее можно в обзоре: [AlphaFold2: глубокий разум и его правильное применение / Хабр \(habr.com\)](#)

Амилоиды

Изучению амилоидных фибрилл способствовали белки прионы. После открытия структур, которые способны формировать эти белки, было обнаружено, что в принципе простые белки также могут образовывать амилоиды. Известно, что чаще всего в амилоидных фибриллах цепь уложена в виде сэндвичей (двух- или многослойного) из параллельной бэта-структуры, в которых бэта-тяжи идут перпендикулярно оси, соединенные водородными связями^[2].

Чтобы предсказать структурные особенности амилоидов и проанализировать возможности сети, я выбрала вариант A1 и последовательность:
RPLVNIYNCSGVQVGDNNYLTMQQTTALP.

Я запустила protein Blast по PDB и получила вот такую выдачу:



Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Chain A_Receptor-interacting serine/threonine-protein kinase 3 [Homo sapiens]	Homo sapiens	98.2	111	100%	3e-25	100.00%	108	7DAC_A
<input checked="" type="checkbox"/> Chain A_Receptor-interacting serine/threonine-protein kinase 3 [Homo sapiens]	Homo sapiens	98.2	111	100%	4e-25	100.00%	141	7DA4_A
<input checked="" type="checkbox"/> SSNMR Structure of the Human RIP1/RIP3 Necrosome [Homo sapiens]	Homo sapiens	51.5	51.5	51%	3e-10	100.00%	15	5VTZ_A
<input checked="" type="checkbox"/> SSNMR Structure of the Human RIP1/RIP3 Necrosome [Homo sapiens]	Homo sapiens	28.2	28.2	51%	0.30	53.33%	18	5VTZ_B
<input checked="" type="checkbox"/> Crystal Structure of the cytoplasmic actin capping protein Cap32_34 from Dictyostelium discoideum [Dictyostelium ...]	Dictyostelium dis ...	26.5	26.5	48%	3.4	55.00%	281	4AKR_A
<input checked="" type="checkbox"/> Crystal structure of FTT_1639c from Francisella tularensis str. tularensis SCHU S4 [Francisella tularensis subsp. tu...]	Francisella tulare...	26.1	39.9	62%	4.5	60.00%	131	7KZW_A
<input checked="" type="checkbox"/> Taurine ABC transporter substrate binding protein TauA from E. coli in complex with 2-Aminoethylphosphonic acid [...]	Escherichia coli...	26.1	26.1	62%	4.7	52.00%	298	6SSY_A
<input checked="" type="checkbox"/> Mouse receptor-interacting protein kinase 3 (RIP3) amyloid structure by solid-state NMR [Mus musculus]	Mus musculus	24.8	24.8	58%	11	61.11%	84	6JPD_A

Рис 1. Выдача protein Blast

Подходящей для сравнения результатов предсказаний оказалась PDB структура с ID 7DAC. Она была получена с помощью твердотельной ЯМР-спектроскопии. Выдача Blast, расположенная ниже, относится к тому же энзиму, полученному с помощью криоэлектронной микроскопии:

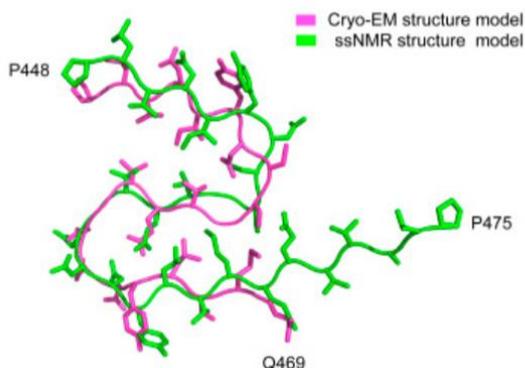


Рис 2. Сравнение структур, полученных двумя методами^[1].

Модели представляют собой взаимодействующую с рецептором серин/треонин-протеинкиназу 3 (RIPK3). Речь пойдет о белке, который преимущественно расположен в цитоплазме. Он является компонентом сигнального комплекса рецептора (TNF)-I и может индуцировать некроптоз за счет взаимодействия с RIPK1 и MLKL в белковом комплексе, называемом некрсомой. Активная полимерная форма белка является амилоидной фибриллой. Она представлена протофиломентом с определенным доменом, который соответствует выданной мне последовательности. Домен образуется тремя бэта-тяжами, которые сложены в S-образную структуру.

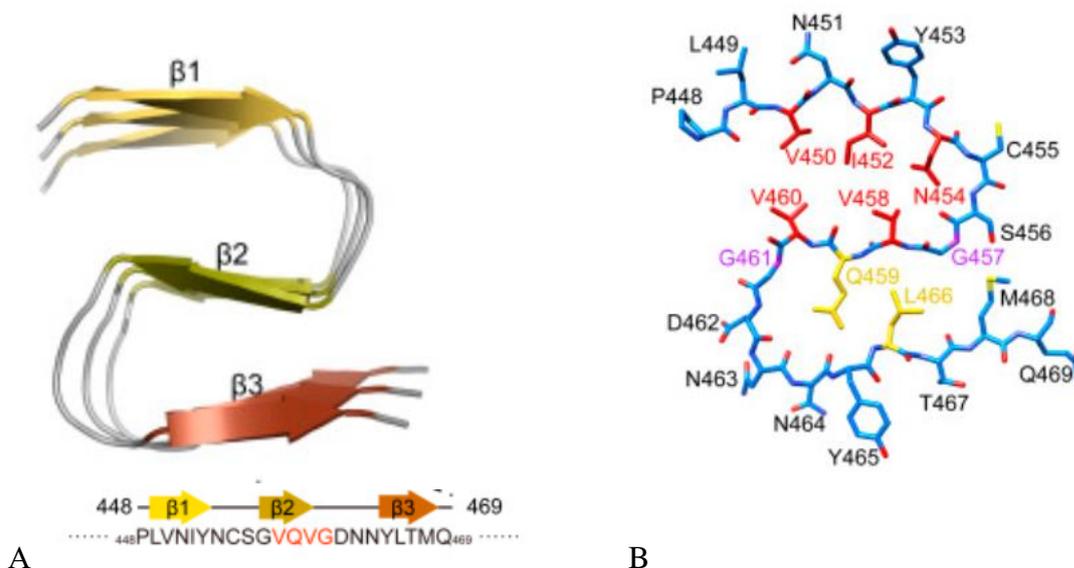


Рис 3. А – пронумерованные бэта-тяжи белка и найденное ядро фибриллы; В – один слой модели белка. Цветами, отличными от синего цвета, отмечены остатки, играющие важную роль в фолдинге^[1].

Для начала ColabFold попытался предсказать структуру мономера. Ниже на рисунке красным отмечена PDB структура 7DAC, а различными цветами предсказания программы. Зеленым цветом указана та структура, которую ColabFold выбрал как лучшее предсказание. Однако, по моему мнению, наиболее близкая модель – это второе предсказание программы, отмеченное бирюзовым цветом. Можно заметить, сравнивая PDB модель и рисунок 3 с предсказаниями, что домен с тремя бэта-тяжами в лучшей модели изменил свое расположение. Это значит, что остатки, играющие важную роль в фолдинге, не сохранили нужную конформацию. Особенно это коснулось тяжей бэта 1 и бэта 2. Между ними расположены гидрофобные остатки, отмеченные на рисунке 3 красным цветом. Эта часть модели хуже всего далась для предсказания, программа слишком сильно разошлась в гипотезах относительно расположения тяжей бэта1 и бэта2. Гидрофобные остатки, отмеченные желтым цветом, были предсказаны в среднем чуть лучше.

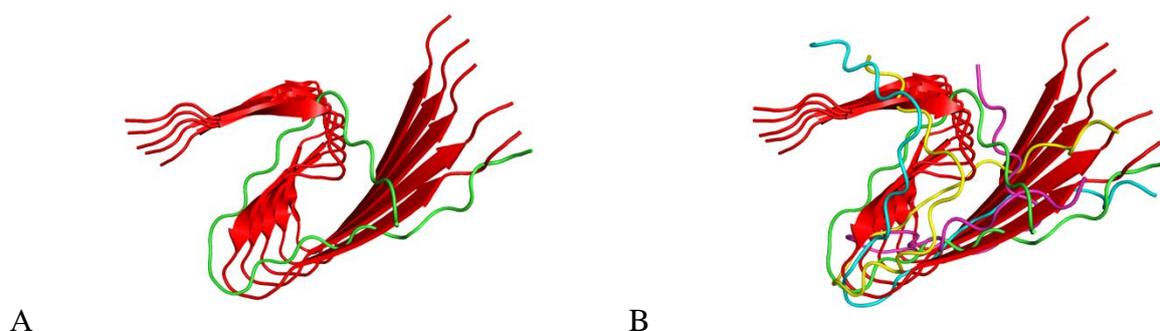
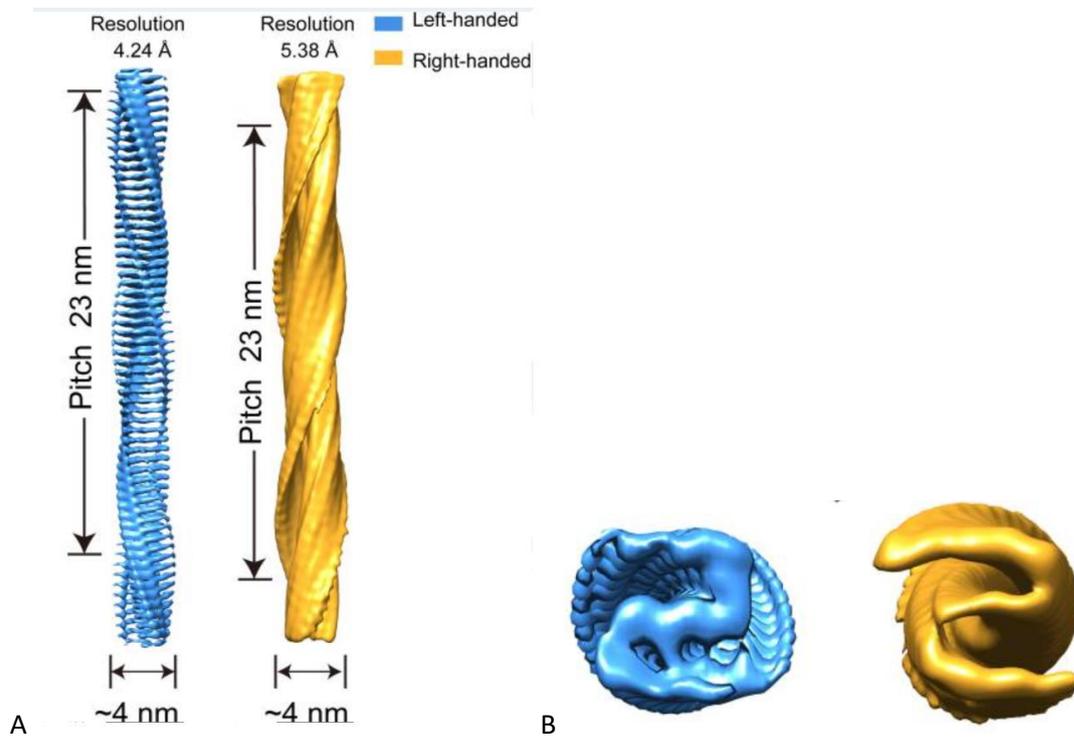


Рис 4. А – лучшее предсказание программы отмечено зеленым цветом, красным отмечена PDB модель 7DAC; В – все предсказания мономера программой ColabFold, красным отмечена модель 7DAC.

С помощью криоэлектронной микроскопии была восстановлена архитектура фибриллы. Примечательно, что фибрилла может существовать в виде правозакрученной и левозакрученной формы. Было установлено в ходе эксперимента, что обе фибриллы имеют ширину 4 нм и шаг 23 нм. На рисунке ниже они представлены с разрешением 4.24 ангстрема, левозакрученная, и 5.38 ангстрем – правозакрученная ^[1].



1) Рис 5. А, В - левозакрученная фибрилла отмечена синим цветом, правозакрученная – желтым^[1].

Ниже на рисунке 6 показано предсказание ColabFold для амилоида, состоящего из пяти повторяющихся молекул. Легко заметить, что полученное лучшее предсказание совершенно не походит на фибриллу, которая характерна для изучаемого белка 7DAC. Здесь отсутствует разметка по вторичной структуре, которая свойственна одному слою фибриллы, отсутствует S-образность, характерная для мономеров. Также надо отметить то, что не очень характерно для амилоидов: бѐта-тяжи идут не параллельно, а антипараллельно. Правая часть модели из двух молекул демонстрирует закономерность в укладке бѐта-тяжей, тем временем левая часть из трех молекул состоит из антипараллельных бѐта-тяжей и хаотично уложенных петель. Это совершенно не похоже на амилоиды.

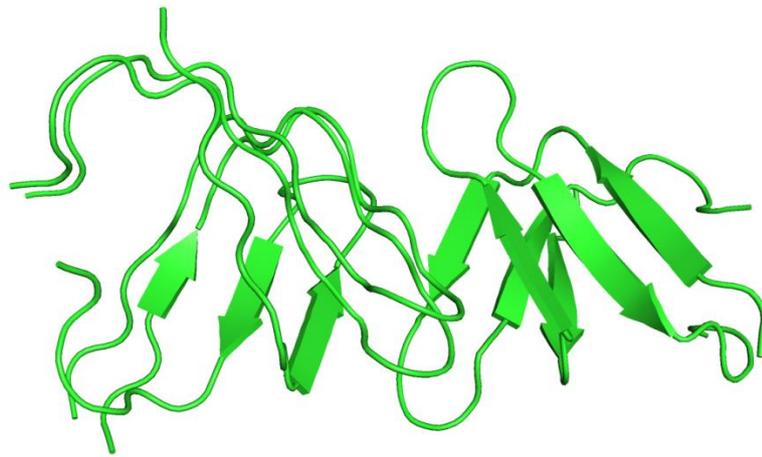


Рис 6. Лучшее предсказание ColabFold для 5 молекул.

Теперь можно рассмотреть результат предсказания программы для структуры из 10 молекул. Как можно увидеть на рисунке ниже, программе удалось отобразить левозакрученность амилоидной фибриллы RIPK3. При наложении лучшего предсказания ColabFold для 10 молекул и модели 7DAC, представленной в PDB, видно, что лучше всего совмещены бэта1-тяжи. Однако есть и не состыковки: тяжи имеют разный размер в двух структурах, тогда как для ядра фибриллы они имеют четкую разметку. Петли между тяжами 2 и 3 плохо совмещаются, так как тяжи бэта 2 структуры 7DAC длиннее и накладываются на петли предсказания. Тяжи бэта 3 обеих структур слишком разные, тут уже речи не идет о хорошем наложении. Однако структурные особенности амилоидов программе показать удалось, к примеру, действительно сохранена параллельность всех тяжей.

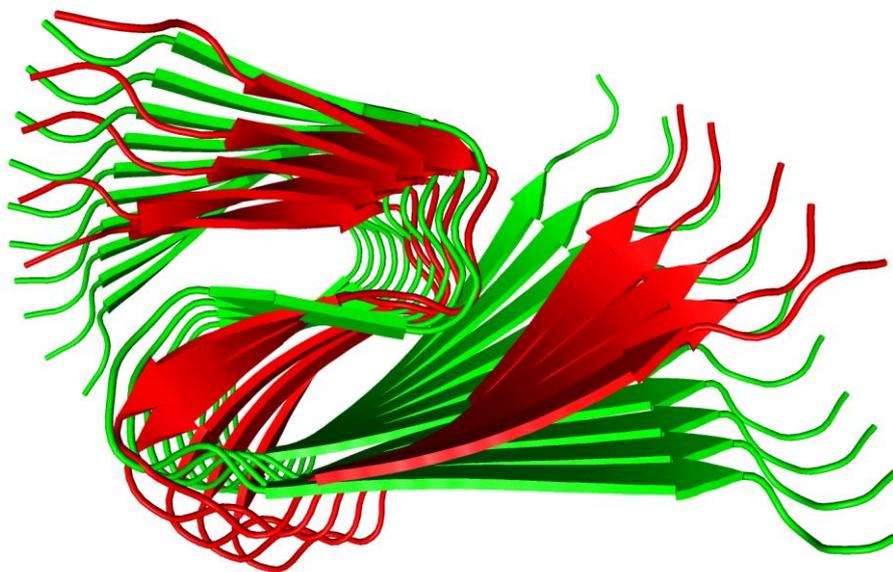


Рис 7. Лучшее предсказание ColabFold из 10 молекул.

Таким образом, если взглянуть на структуры PDB, которые соответствуют выданной мне последовательности, и предсказания, сделанные с помощью нейронной сети, можно сделать вывод, что для амилоидных фибрилл (а конкретно для структуры белка RIPK3), программа еще не может достаточно хорошо предсказать архитектуру. Ни мономер, ни модели из 5 или 10 молекул не совместились в должном виде с полученными методами ЯМР и криоэлектронной микроскопией структурами. Я наблюдала лишь общие правильные тенденции в предсказании, в большей степени для 10 молекул и в меньшей для мономера.

Ссылки на литературу:

- 1) Xialian W., et al. 'The structure of a minimum amyloid fibril core formed by necroptosis-mediating RHIM of human RIPK3'. PNAS. 118 (2021). doi:10.1073/pnas.2022933118
- 2) Лекции А.В. Финкельштейна
- 3) RIPK3 - Wikipedia
- 4) AlphaFold2: глубокий разум и его правильное применение / Хабр (habr.com)