

Практикум 10

Пункт 0. Введение

AlphaFold2 - это мощный инструмент для предсказания пространственной организации белковой цепи. Он достаточно точно может предсказывать положение остова и с высокой точностью предсказывает ориентации боковых цепей. Однако часто предсказания получаются весьма неточными, противоречивыми и просто не укладывающимися в наши представления о физике. В рамках задачи практикума нам было предложено использовать AlphaFold2 и описать правдоподобность его предсказания.

Мной была выбрана задача Deletion - проверить выдачу сервиса для полноразмерной последовательности и для делетированной на 9 а.к.о. с N-конца.

Пункт 1. Ничто не предвещало беды

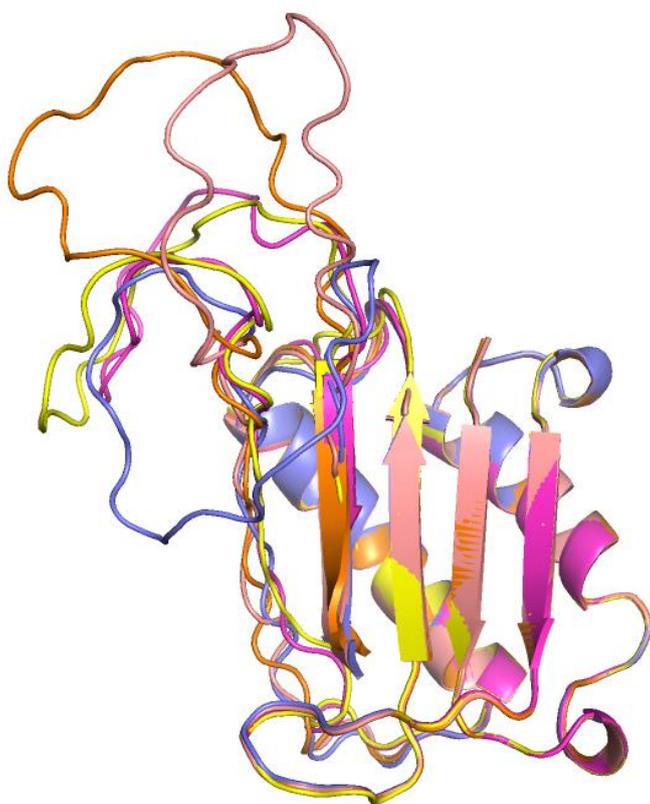


Рис.1. Предсказание структуры полноразмерной последовательности (D1 seqA), rank1-5

Как видно на рисунке выше, AlphaFold2 выдал 5 моделей, очень сильно похожих одна на другую. У них одинаковая топология, они замечательно сходятся на большом протяжении начиная с N-конца. Сильное расхождение наблюдается на C-конце, особенно в области большой петли: здесь все модели идут врозь. Посмотрим теперь более близко на расположение элементов вторичной структуры в структуре ранга 1 - наиболее правдоподобной.

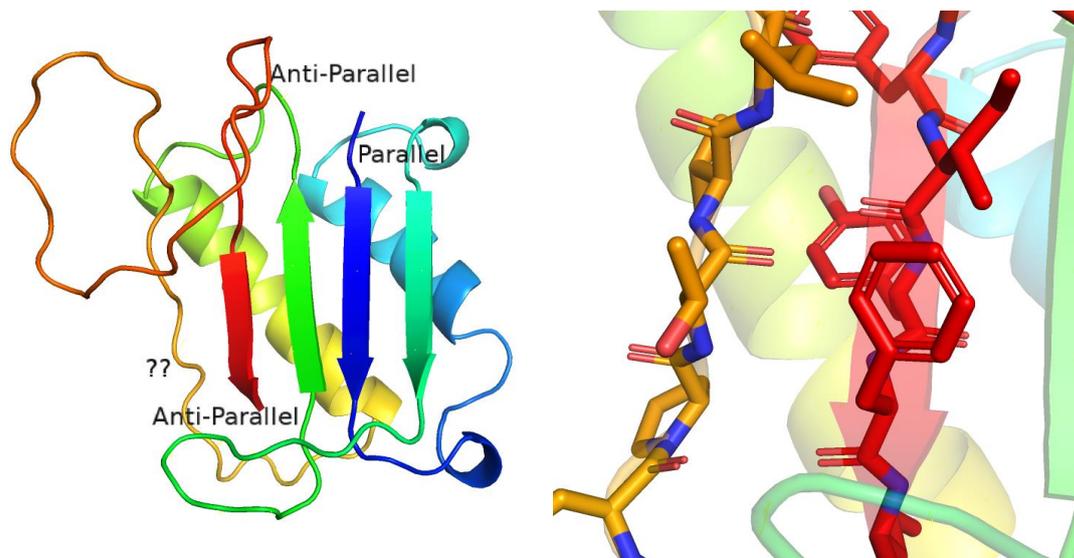


Рис.2. Модель структуры seqA (Rank1) а.Ориентация бета-листов; б.Проверка наличия еще одного бета-листа.

На изображении 2а показана ориентация и топология бета-листов - она одинаковая во всех 5 моделях. Если обозначить листы в порядке последовательности ($\beta 1$ - blue, $\beta 2$ - cyan, $\beta 3$ - green, $\beta 4$ - red), то можно написать, что $\beta 1$ контактирует параллельно с $\beta 2$ и антипараллельно с $\beta 3$, а $\beta 3$, кроме контакта с $\beta 1$, контактирует антипараллельно с $\beta 4$. Оранжевый отрезок, идущий антипараллельно к $\beta 4$, не детектируется как бета-лист (будем звать его $\beta \psi$ - "псевдолист"). При пристальном просмотре водородных контактов между ними не обнаружилось. В таком случае вообще непонятно, что удерживает оранжевый отрезок в таком положении.

Назовем blue-спираль $\alpha 1$, а green-спираль $\alpha 2$. Тогда элементы вторичной структуры (или подобные им) идут в такой последовательности: $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$ - $\beta \psi$ - $\beta 4$. То есть это a+b-белок

Повторюсь, описанная топология верна для всех 5 моделей. Чем же они отличается? Во-первых, очень сильно расходятся петли, наиболее близкие к С-концу (от $\beta \psi$ до $\beta 4$). Мы не можем утверждать, что реальное положение петли соответствует хоть какой-нибудь из 5 моделей (см. рис.1). Во-вторых, петля $\beta 3$ - $\alpha 2$ немного расходится, но общий тренд сохраняется (рис.3). В-третьих, $\beta 4$ имеет разную длину в разных моделях из-за ориентации доноров и акцепторов водородных связей и относительного сдвига цепи (рис.4). Так, в моделях ранга 1 и 5 этот участок очень хорошо сходится и автоматическая разметка показывает $\beta 4$ как лист, немного укороченный с N-конца. В моделях рангов 2-4 этот лист размечается по-разному (лист укорочен в разной мере с С-конца) - рис.5.

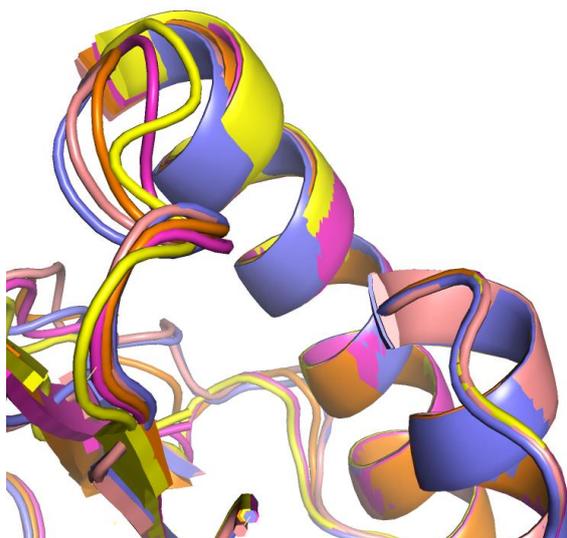


Рис.3. $\beta 3$ - $\alpha 2$ -петля.

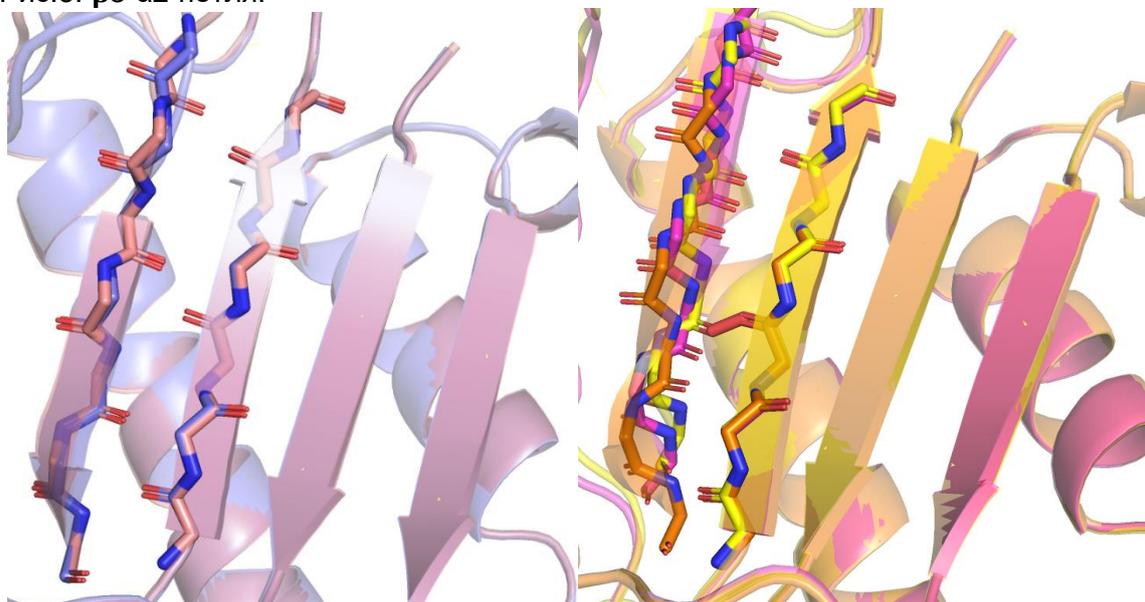


Рис.4. $\beta 4$ -лист в моделях rank1+5 (a) и rank2-4 (b)

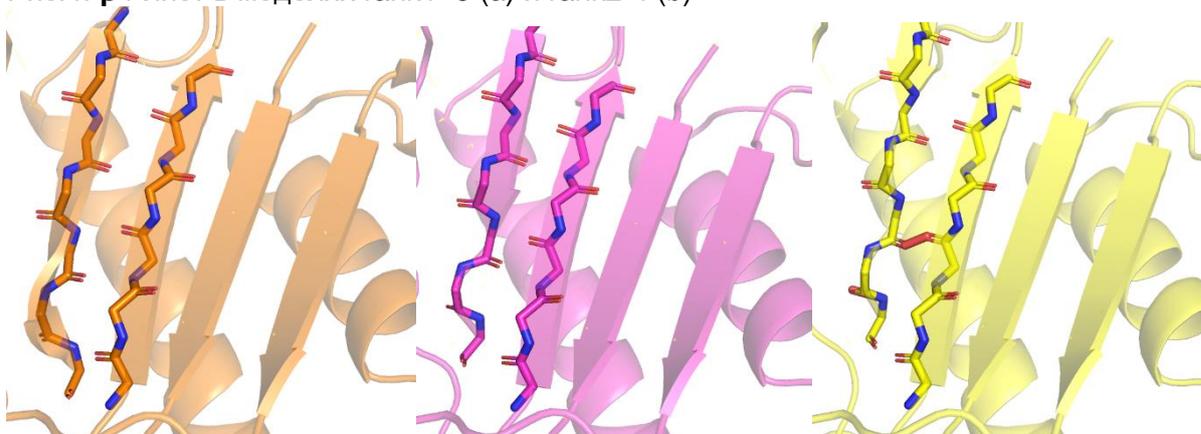


Рис.5. $\beta 4$ -лист в моделях ранга 2, 3 и 4 соответственно (*посмотрите на восхитительный clash в 4 модели*)

Наконец, я решила поискать структуру предложенной последовательности в базах данных (сначала rblast, потом pdb). Последовательность идеально подошла под D-dopachrome decarboxylase-like protein человека (UniProtKB - A6NHG4). Структур непосредственно этого белка в pdb нет, зато была структура D-dopachrome tautomerase

человека (1dpt) - с первого взгляда, очень похожая на то, что я видела в 5 изучаемых моделях, выданных AlphaFold2. И вот тут начались расхождения.

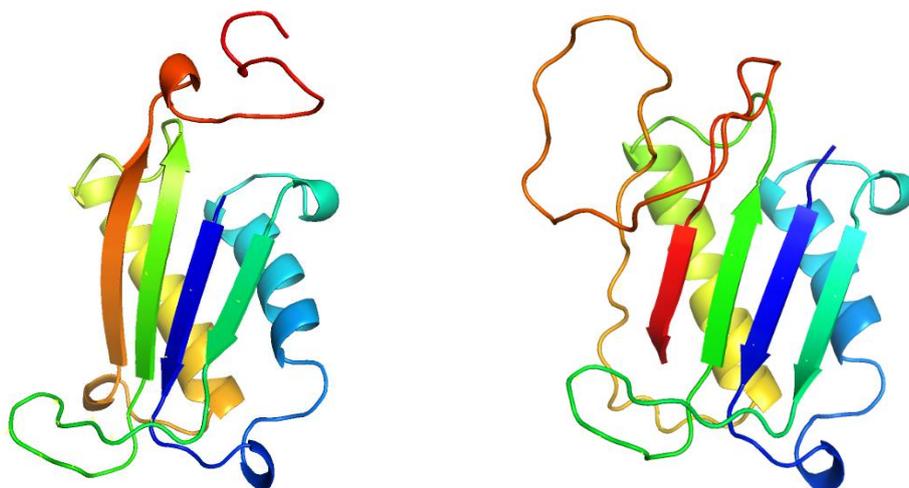


Рис.6. Структуры 1dpt и модель ранга 1 выдачи AlphaFold2

В структуре 1dpt есть параллельный бета-лист в том месте последовательности, где AlphaFold2 показал не структурированный фрагмент $\beta\psi$. Однако не стоит сразу же говорить, что модели неправильные - именно с этого фрагмента до С-конца белковые последовательности различаются. Пока можно допустить, что расположение вторичных элементов может быть именно таким, как в модели.

Ниже приведены графики, выданные AlphaFold2 после работы.

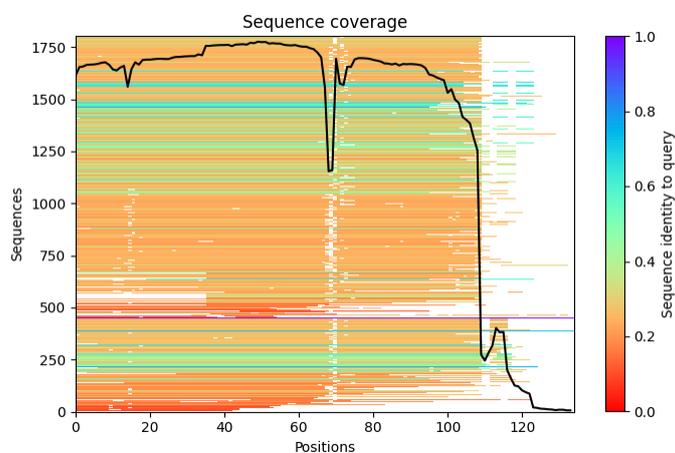
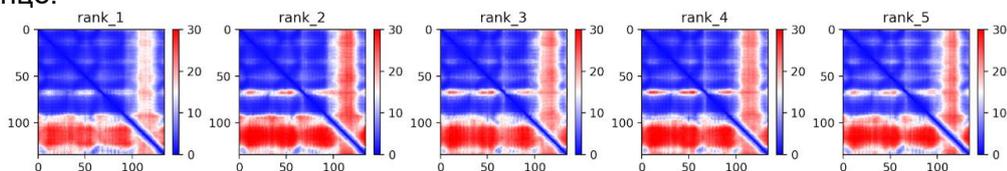


Рис.7. Покрытие позиций последовательностями. Видны провалы в области $\beta3-\alpha2$ -петли и на С-конце.



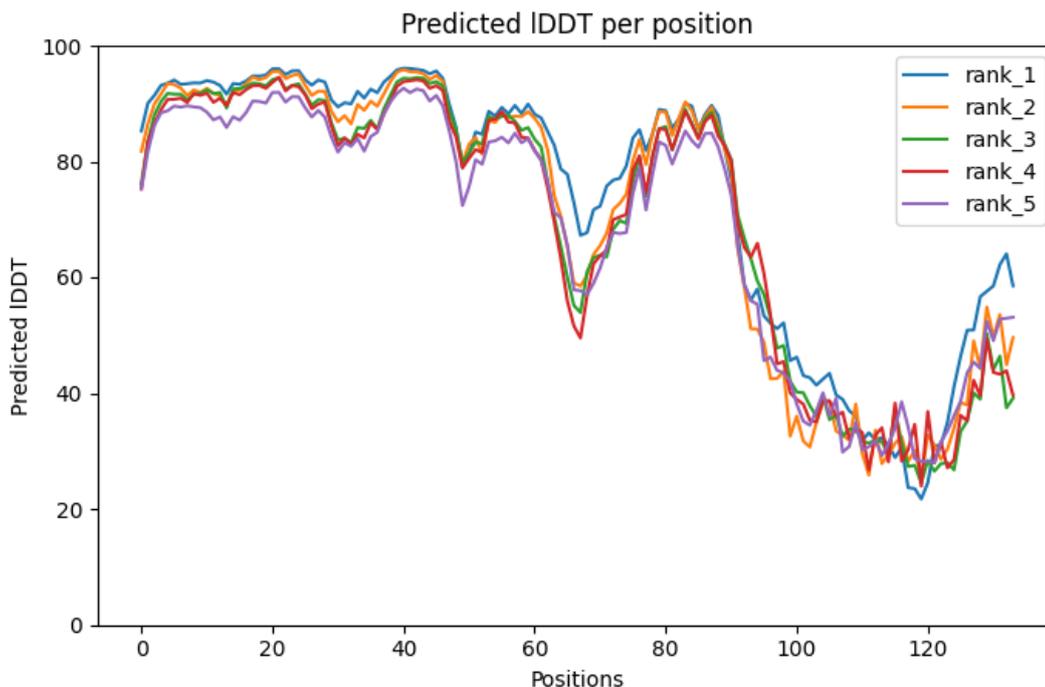


Рис.8. Карта взаимодействий и качество по позициям в каждой из 5 моделей (все примерно одинаковые)

Пункт 2. Все становится интереснее

На этом этапе можно понять, что первые 9 а.к.о. последовательности seqA лежат на β 1-листе (мы в этом почти уверены - так было во всех 5 моделях и в структуре 1dpt). То есть делеция с N-конца удалила бы лист, лежащий не на краю бета-структуры. Можно предположить, что она полностью видоизменила бы структуру и сделала бы ее неузнаваемой.

Потом я запустила AlphaFold2 для укороченной с N-конца на 9 а.к.о. последовательности. Результат можно видеть на рис.9а. На первый взгляд тут есть структура, похожая на предыдущие модели, но не все. Многие фрагменты сходятся во всех 5 моделях, однако наиболее похожи модели рангов 1-3 (рис.9б).

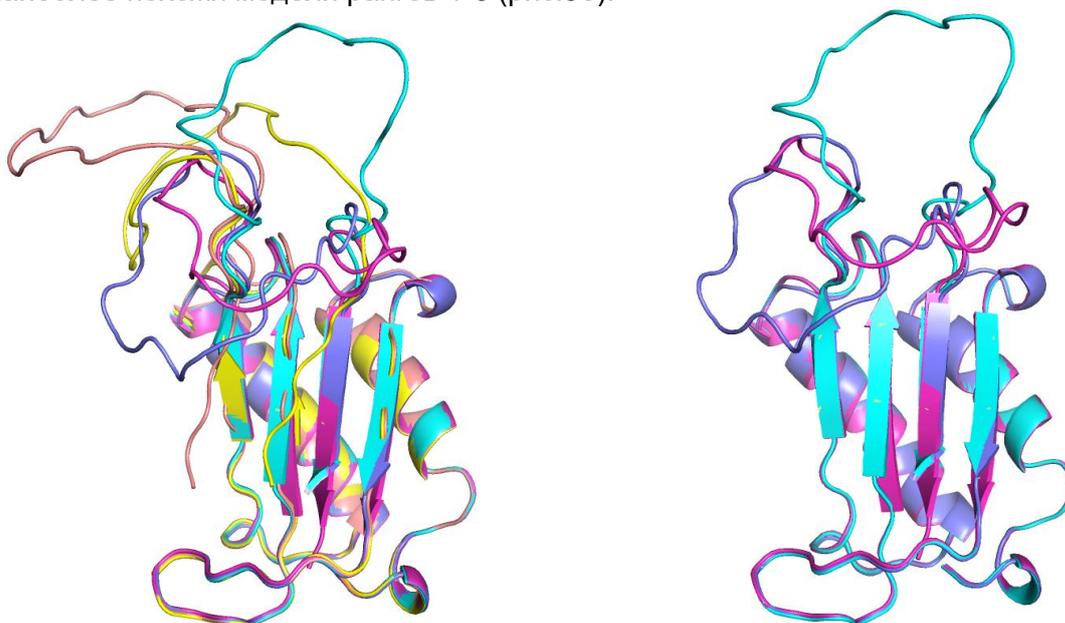


Рис.9. Предсказание структуры укороченной последовательности (D1 seqB), rank1-5 (а) и rank1-3 (б)

Среди моделей ранга 1-3 общие все элементы вторичной структуры - расходитя только та петля на С-конце. На рис.10 показаны модели рангов 4 и 5.

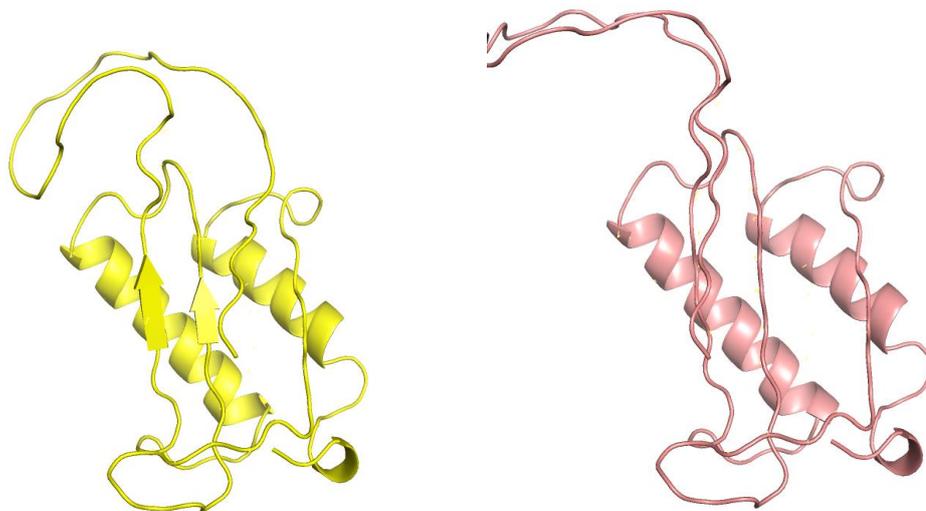


Рис.10. Модели ранга 4 и 5

У них все плохо - бета-листовая структура не идентифицируется из-за ушедшего куда-то в закат С-конца (наверное, это именно то, что я и предполагала увидеть в начале этого пункта). Альфа-спирали при этом определяются правильно. Но самое интересное, что я увидела — это жуткий clash на С-конце в модели ранга 5 (рис.11). На самом деле, clash-ей тут может быть и больше, но этот был самым очевидным.

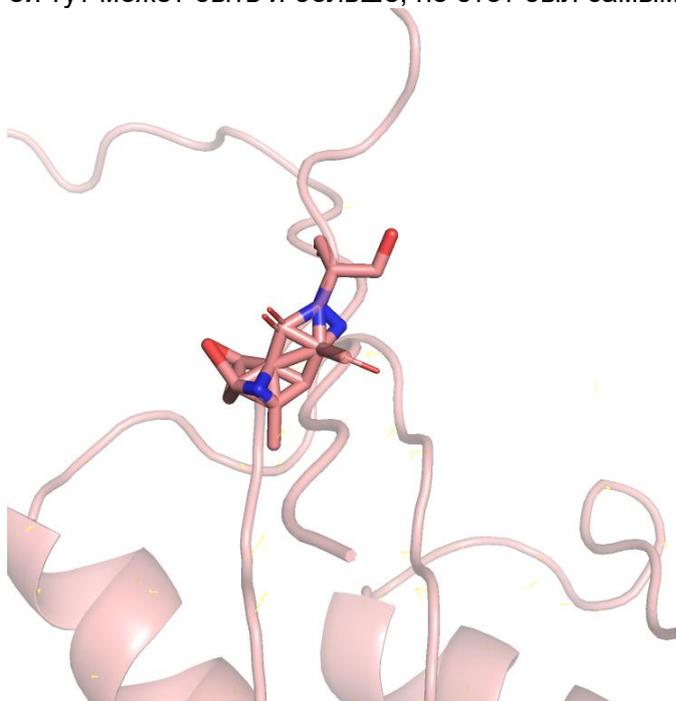


Рис.11. Clash на С-конце модели ранга 5

Интересно все-таки понять, как в моделях ранга 1-3- не потерялась структура несмотря на делецию N-конца. (Рис.12)

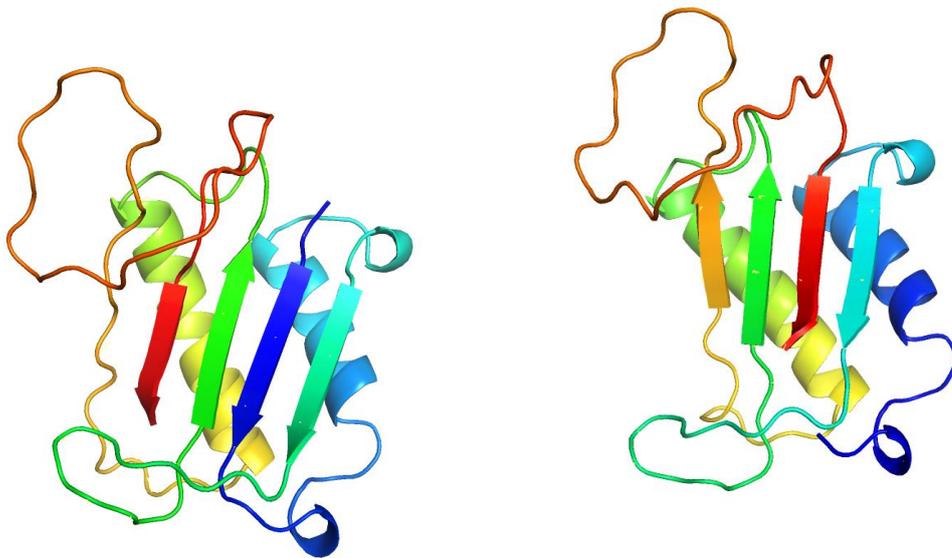


Рис.12. Сравнение топологии лучших моделей структуры seqA (без делеции) и seqB (с делецией)

Как мы видим, структура бета-листа восстановлена за счет изменения топологии и появления $\beta\psi$ -листа (оранжевый). С-концевой $\beta4$ -лист заменил $\beta1$ -лист N-конца. Вместо него с $\beta3$ -листом контактирует $\beta\psi$ -лист, только в параллельной ориентации. Можно попробовать изобразить параллельные и антипараллельные контакты между листами.

	$\beta1$	$\beta2$	$\beta3$	$\beta\psi$	$\beta4$
$\beta1$		+	-		
$\beta2$	+				
$\beta3$	-				-
$\beta\psi$					(-)
$\beta4$			-	(-)	

	$\beta1$	$\beta2$	$\beta3$	$\beta\psi$	$\beta4$
$\beta1$					
$\beta2$					+
$\beta3$				+	-
$\beta\psi$			+		
$\beta4$		+	-		

Рис.13. Сравнение топологии бета-структуры моделей структур seqA и seqB (параллельный контакт +, антипараллельный контакт -)

Непонятно, насколько реалистично и функционально такое изменение топологии? Вообще есть сведения, что делеции с N-концов могут приводить к изменению топологии с сохранением общей бета-структуры. Например, у мышиноного белка ароМ по сравнению с человеческим ароМ есть делеция N-конца [1]. Топология бета-бочек при этом несколько различается. Однако в нашем случае не просто новообразованные элементы структуры заполняют вакантные места, а уже старые, приспособленные под бета-лист фрагменты перескакивают с места на место.

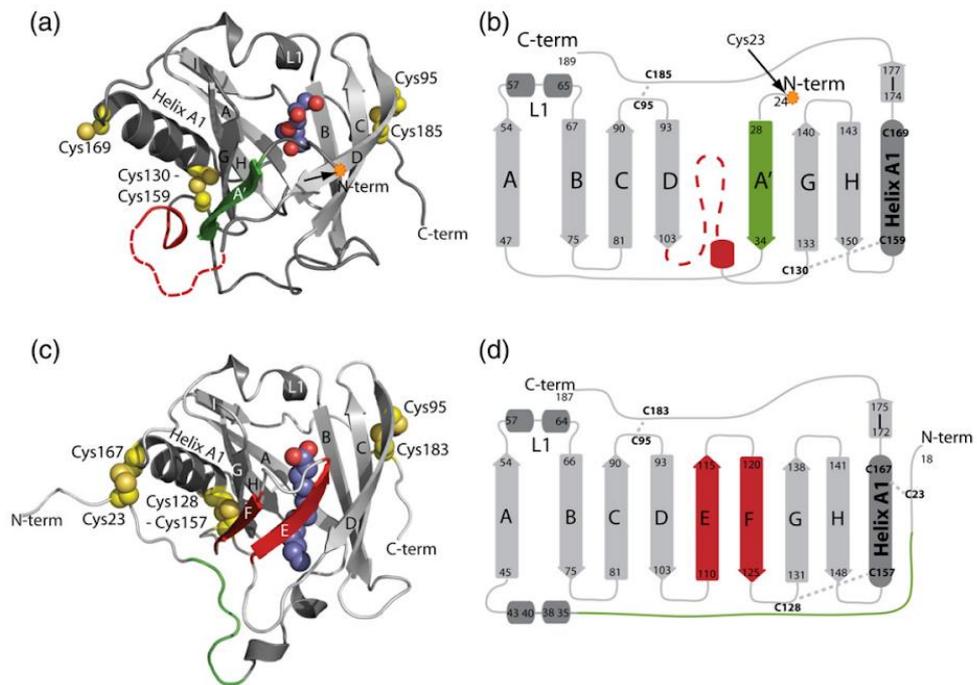
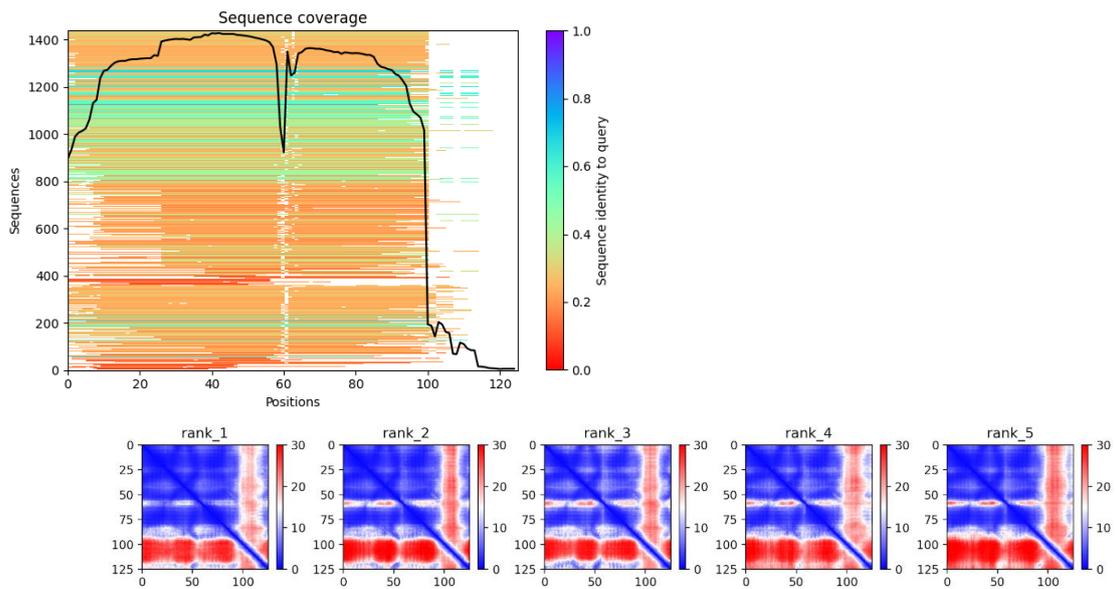


Рис.14 (из статьи [1]). Разная топология m-aroM (a, b) и h-aroM (c, d)

Не забудем вставить картинки о качестве моделей.



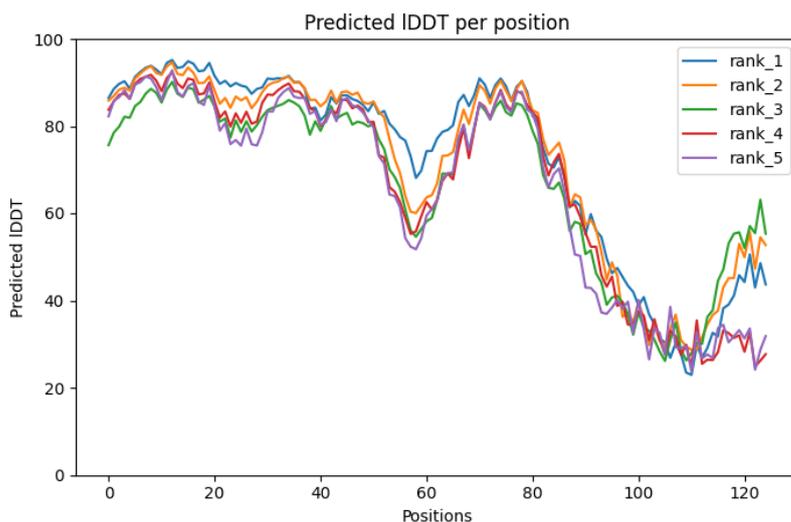


Рис.8. Покрывание позиций последовательностями, карта взаимодействий и качество по позициям в каждой из 5 моделей (видно, что качество моделей рангов 1-3 лучше, чем рангов 4-5 (особенно на С-конце))

Пункт 3. Вывод

AlphaFold2 весьма хорошо справился с нашей задачей. Он предсказал реалистичную структуру неделетированного белка seqA (при этом основные элементы во всех 5 моделях располагались одинаково) и делетированного белка seqB. Однако предсказание для С-конца пострадало в обоих случаях: большая петля сильно расходится, а в случае seqB С-концевой лист нашелся не на всех моделях. Обнаружилось, однако, что делеция структурно важного элемента на N-конце не фатальна и может быть восполнена топологической перестройкой. Это многое может говорить о механизмах эволюции топологии бета-листов: при делеции структурно важного участка белок может незначительно терять в стабильности, поэтому делеция может закрепиться в эволюции.

[1] Sevvana M., K. Kassler, J. Ahnstrom, S. Weiler, B. Dahlback, H. Sticht, Y.A. Muller. **Mouse ApoM displays an unprecedented seven-stranded lipocalin fold: folding decoy or alternative native fold?** *J. Mol. Biol.*, 404 (2010), pp. 363-371