

Практикум 10

Данный практикум посвящен работе с [ColabFold](#).

Выдачу AlphaFold2 можно скачать [здесь](#).

Выбранный кейс: **делеция**, вариант D1.

Дано две последовательности, seqA и seqB. Вторая последовательность – это первая с N-концевой делецией (было удалено 9 аминокислот). Предскажем структуры для обеих независимо.

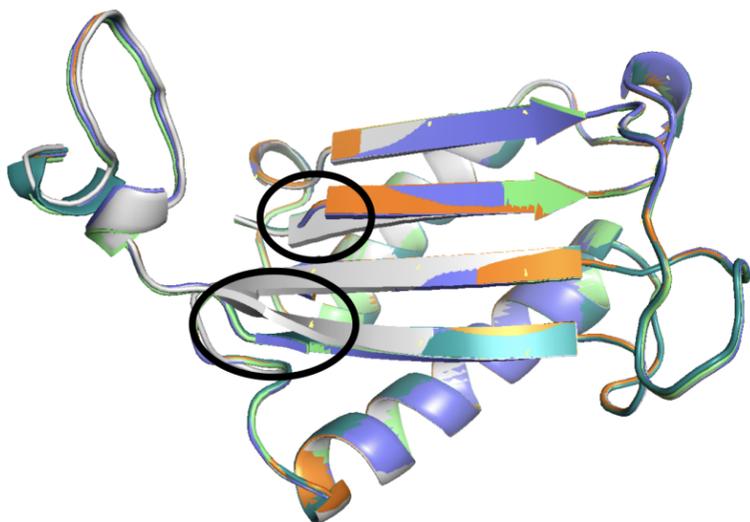
Проверка на наличие ортологов

Поиск по последовательности в ncbi protein blast (против pdb) выдал 7 хитов, 5 хитов с самым высоким процентом идентичности принадлежат D-допахром таутомеразе человека (71% идентичности). Из-за того что процент идентичности не близится к 100%, ожидать полного совпадения структуры мы не будем, но, как мне кажется, удобно иметь некий примерный референс того с чем мы работаем. За референс возьмем структуру [1DPT](#).

Запуск AlphaFold2 с эволюционной информацией

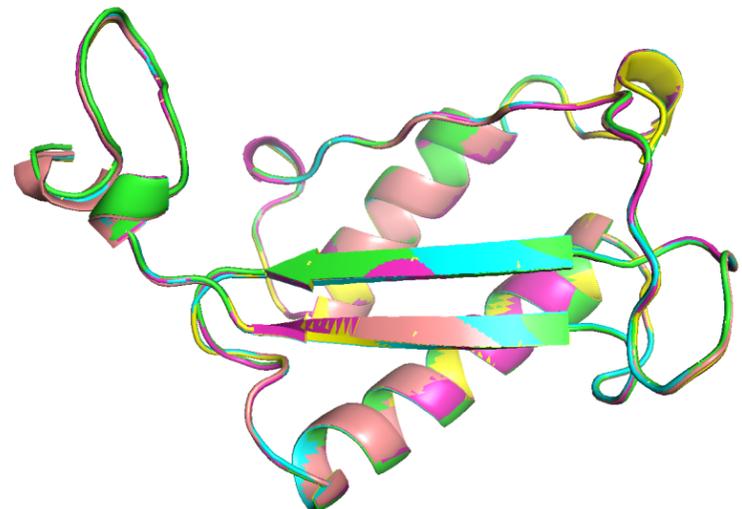
Запуск проводился с параметрами по умолчанию отдельно для последовательностей без и с делецией.

AlphaFold2 выдал предсказания с хорошим качеством, близким к постоянному на протяжении аминокислотной последовательности. На изображениях ниже видно, что пять предсказанных моделей отлично сходятся между собой, разве что для предсказания по полной последовательности есть небольшие расхождения в длине бета-листа с одной из моделей, а также с положением конца бета-листа с другой.



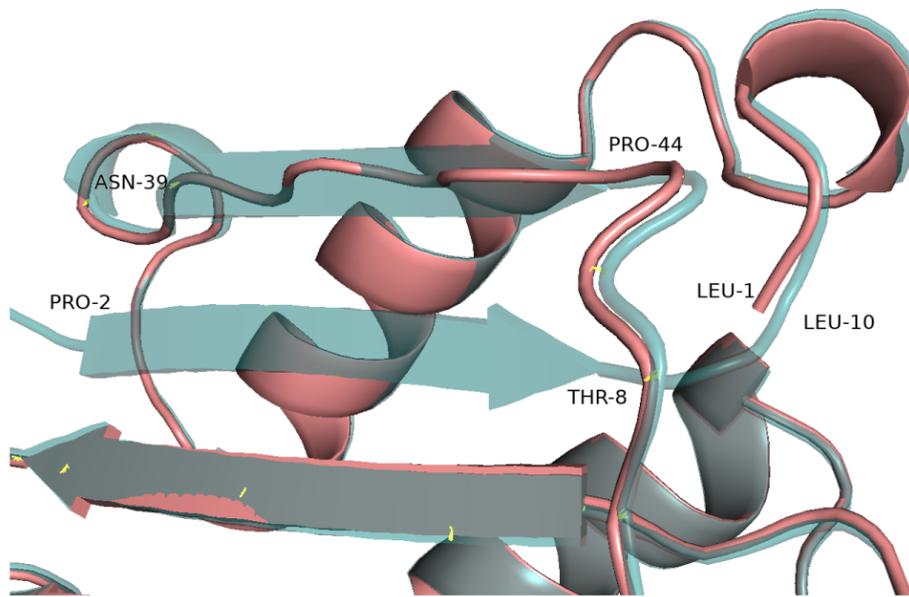
Предсказания для полной последовательности

For Educational Use Only



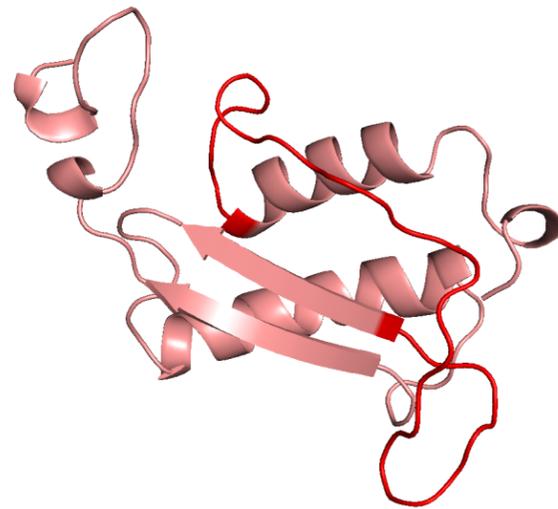
Предсказание для последовательности с делецией

Друг от друга построенные модели различаются двумя бета-листами: N-конец полной последовательности образует с другим участком бета-лист, а при делеции этого участка это “связующее звено” пропадает, и оба листа не формируются. Далее в этом разделе будем визуализировать только модель с лучшим предсказанием.



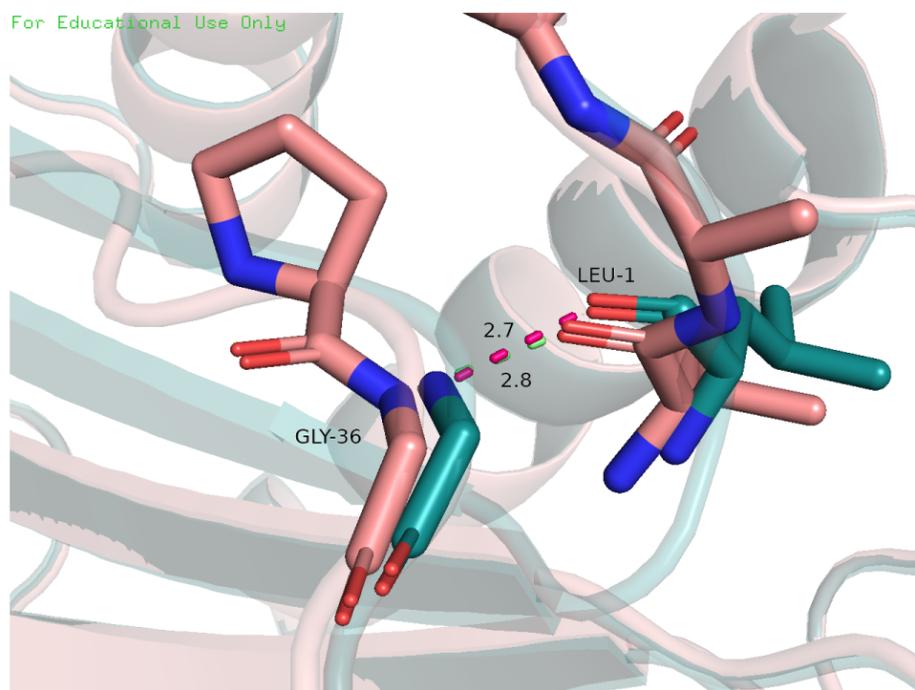
Розовым - последовательность с делецией, зеленым - полная последовательность. Нумерация (кроме 1Leu) для полной последовательности.

Меня смущает то, что образовавшаяся петля вместо бета-слоя в последовательности с делецией (на рисунке выше 39Asn-Pro44 по нумерации для полной последовательности) “заморозилась” в том же положении где до этого был бета-слой. Протяженность всей петли - 24 аминокислоты. Я бы предположила, что этот участок как минимум изменил бы свое положение.



Структура последовательности с делецией, красным выделена петля, о которой идет речь.

Также у меня вызвал подозрение N-конец для случая с делецией: он располагается почти там же, что и этот же участок в структуре полной последовательности. Однако, судя по всему, его положение поддерживается водородной связью.

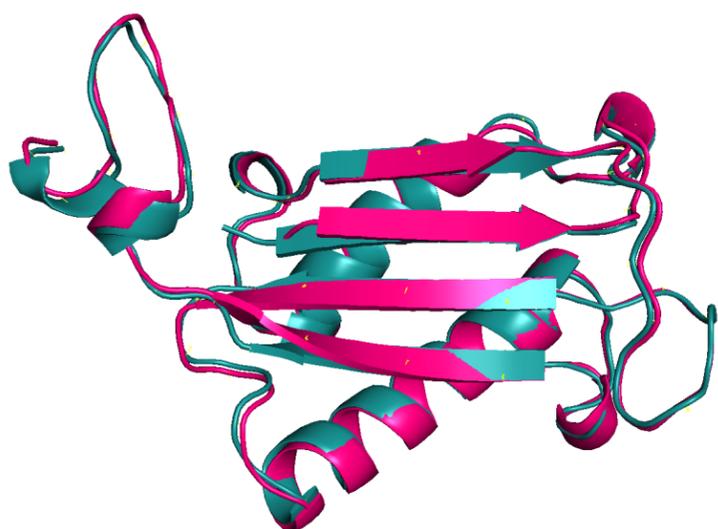


Нумерация остатков по последовательности с делецией.

Сравним полученные структуры с референсом. Для структуры полной последовательности отличия несущественные, заметны расхождения в петлевых участках, а также на С-конце в референсной структуре не

аннотируется элемент альфа-спирали. Отличия с последовательностью с делецией - объединение отличий полной от референсной и полной от делетированной.

For Educational Use Only



For Educational Use Only



Запуск AlphaFold2 без эволюционной информации

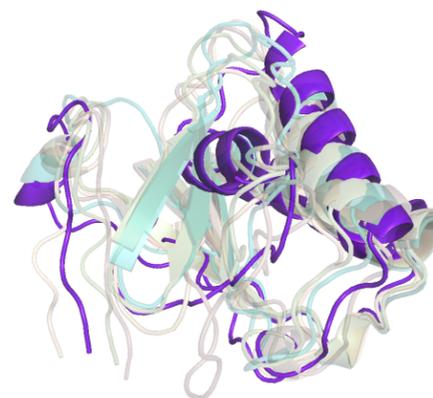
Для запуска AlphaFold2 без учета эволюционной информации (то есть, без построения выравниваний) параметр `msa_mode` был изменен на `single_sequence`. Также `num_recycles` был увеличен до 6.

Модели получились намного более вариабельными как по качеству, так и по структуре.

Для начала рассмотрим модели, построенные для полной последовательности.

Модель с рангом 5 (худшее предсказание) резко отличается от других: хотя одна альфа-спираль угадывается в принципе верно, вместо бета-слоев, которые мы ожидаем увидеть, угадывается альфа-спираль и большая петля. Поэтому мы исключим эту модель из дальнейшего рассмотрения.

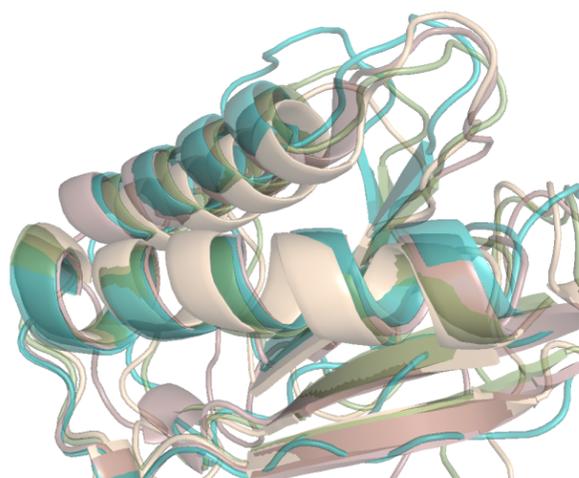
For Educational Use Only



Фиолетовым - модель ранга 5

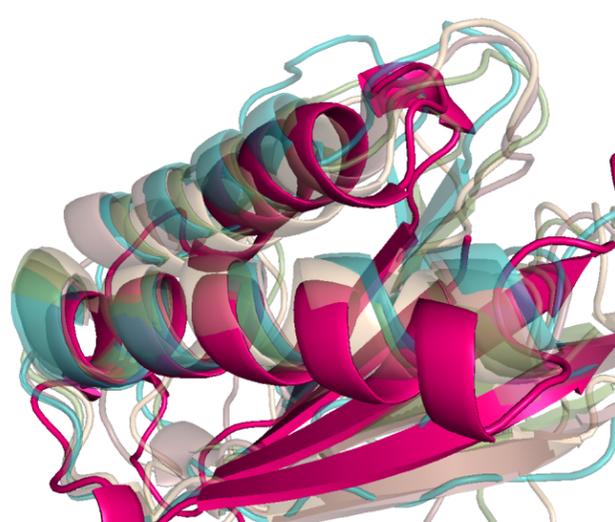
В остальных моделях хорошо выделяются альфа-спирали, которые довольно сопоставимы с положением спиралей в референсном белке.

For Educational Use Only

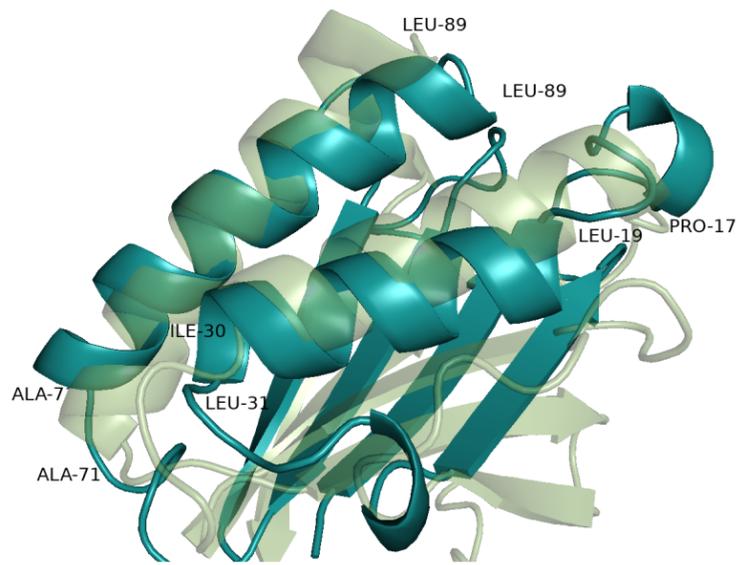


Альфа-спирали в предсказанных структурах

For Educational Use Only



То же, референсная модель покрашена маджентой

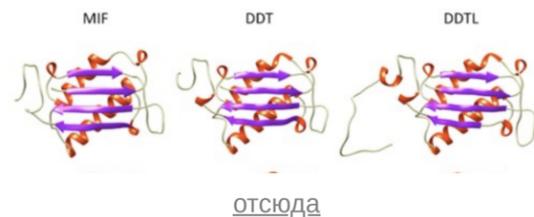


Сопоставление структуры, предсказанной с выравниваниями (темно-зеленый) со структурой, предсказанной без выравниваний (светло-зеленый). Область альфа-спиралей.

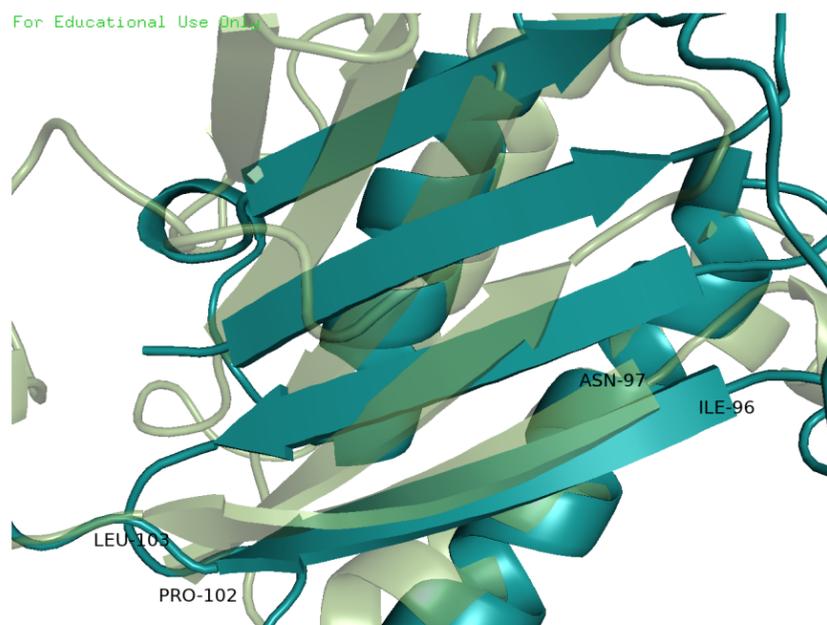
Однако, с бета-листами дело обстоит не так однозначно. В структуре с рангом 1 (лучшее предсказание) число последовательных бета-слоев равно пяти, для двух других структур - трем, а для четвертой по рангу - четырьмя.

Мы ожидали увидеть 4 бета-слоя, т.к.:

- в референсной структуре 4 бета-слоя
- в предсказанной с выравниваниями структуре 4 бета-слоя
- в гомологичных белках 4 бета-слоя



Причем неким обобщением получилась структура с рангом 1, у которой больше всего бета-слоев: бета-слои других структур образуют подмножества от бета-слоев этой структуры (с тем же положением и направлением слоев). Поэтому сравним лучшую модель из предсказания по полной последовательности (с учетом выравниваний) с этой.

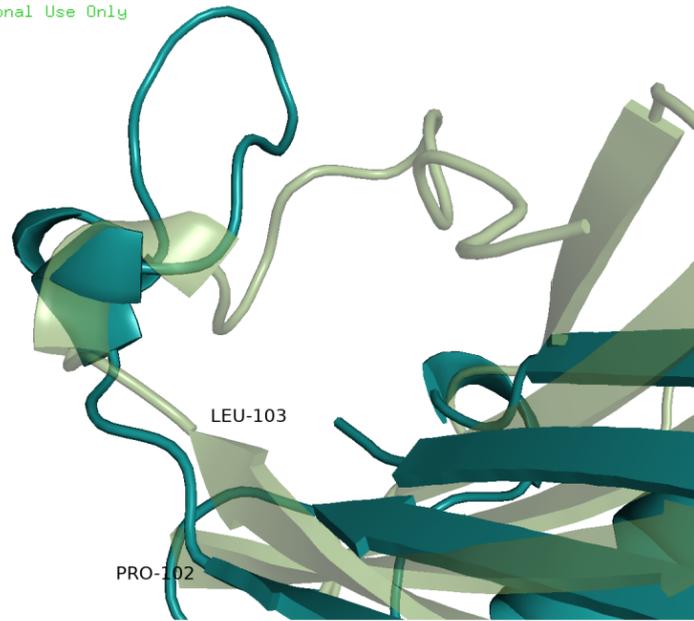


Сопоставление структуры, предсказанной с выравниваниями (темно-зеленый) со структурой, предсказанной без выравниваний (светло-зеленый). Область бета-слоев.

По направлению и расположению почти совпадает только “нижний” бета-лист. Рассмотрим, что пошло не так с остальными.

По “направлению движения” совпавшего бета-слоя находится С-конец белка. Его положение не совпадает между двумя структурами, но это ожидаемо: скорее всего он довольно подвижен.

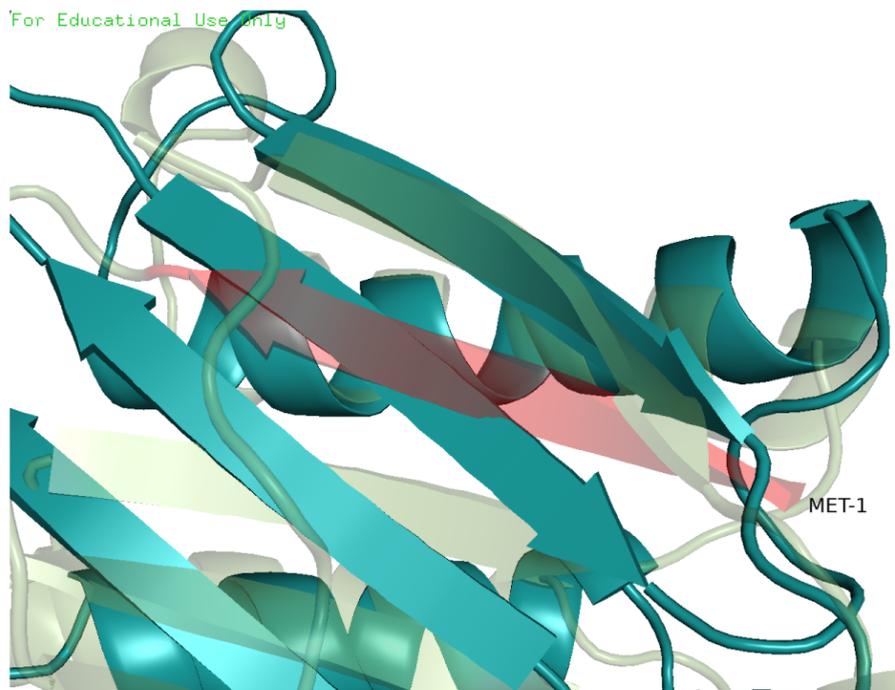
For Educational Use Only



Сопоставление структуры, предсказанной с выравниваниями (темно-зеленый) со структурой, предсказанной без выравниваний (светло-зеленый). Область С-конца.

Другой конец у обеих структур идет на образование альфа-спирали до 71Ala, которые у них обеих довольно сильно совпадают (видно на рисунке выше про альфа-спирали). После альфа-спирали последовательность у обеих структур почти сразу идет на образование бета-слоя, но тут-то и начинается расхождение: если в первоначально предсказанной структуре этот бета-слой образовывал связь с предыдущим, то теперь между ними встраивается бета-слой, образованный из первых семи аминокислотных остатков последовательности.

For Educational Use Only

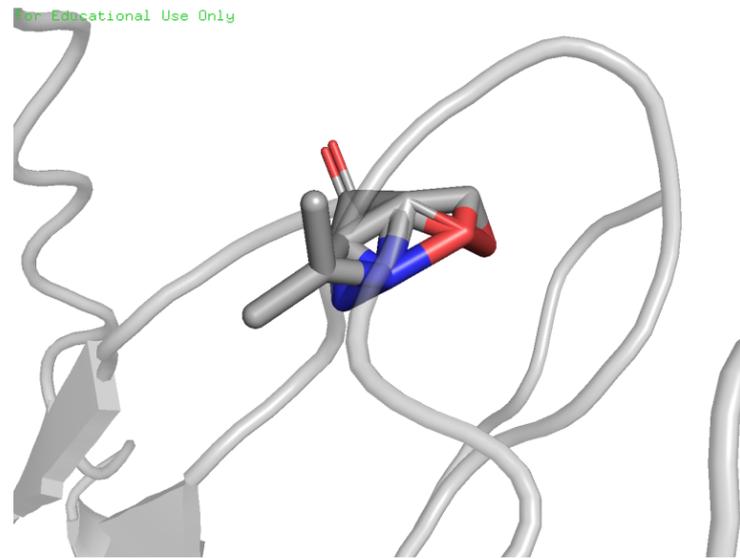


Сопоставление структуры, предсказанной с выравниваниями (темно-зеленый) со структурой, предсказанной без выравниваний (светло-зеленый). Красным выделен бета-слой, образованный началом последовательности.

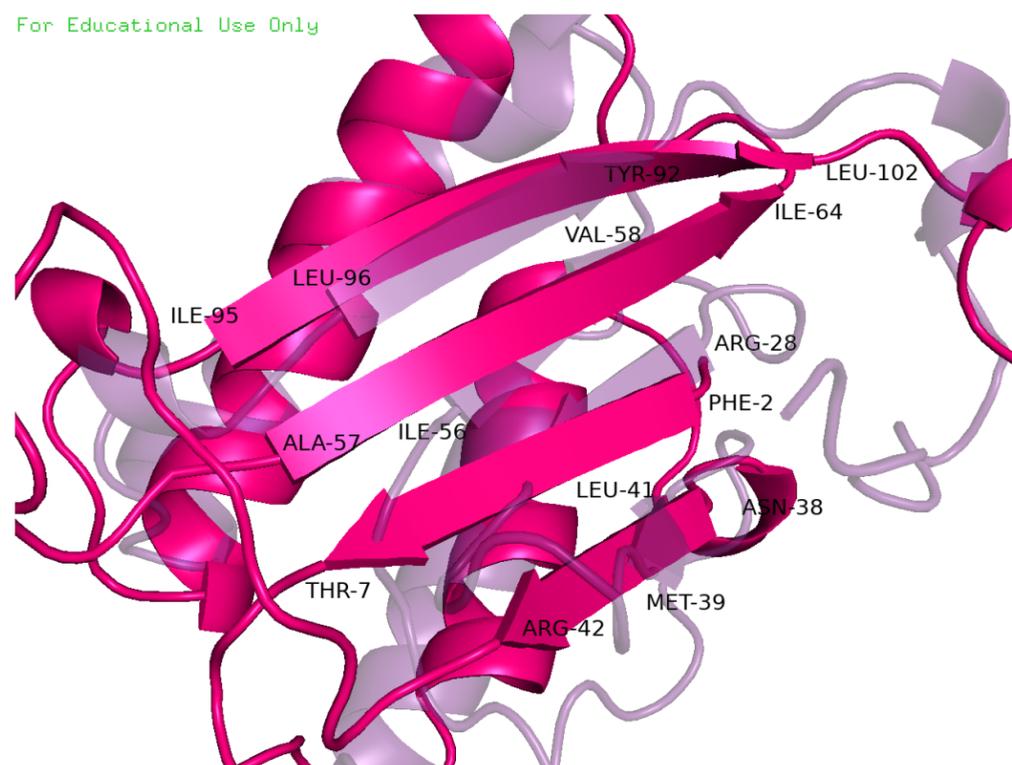
В референсной последовательности N-конец тоже участвует в формировании бета-листа со 2 по 7 аминокислотные остатки, однако он идет следом за рассматриваемым листом. То есть, в предсказании они как бы поменялись местами.

Для последовательности с делецией первых 9 аминокислот предсказание также выдает хорошее совпадение двух альфа-спиралей и с референсной структурой, и со структурами, предсказанными с выравниваниями. Однако, предсказание укладки ожидаемых бета-слоев неоднозначное: у трех структур из пяти (ранги 1, 4, 5) на месте, где мы ожидаем увидеть бета-слои, - сплошной неструктурированный участок. Для двух других было определено 4 (ранг 2) и 3 (ранг 3) бета-слоя.

В попытках исследовать расхождения и проследить их причину я многократно наталкивалась на пересекающиеся петли вроде этой:

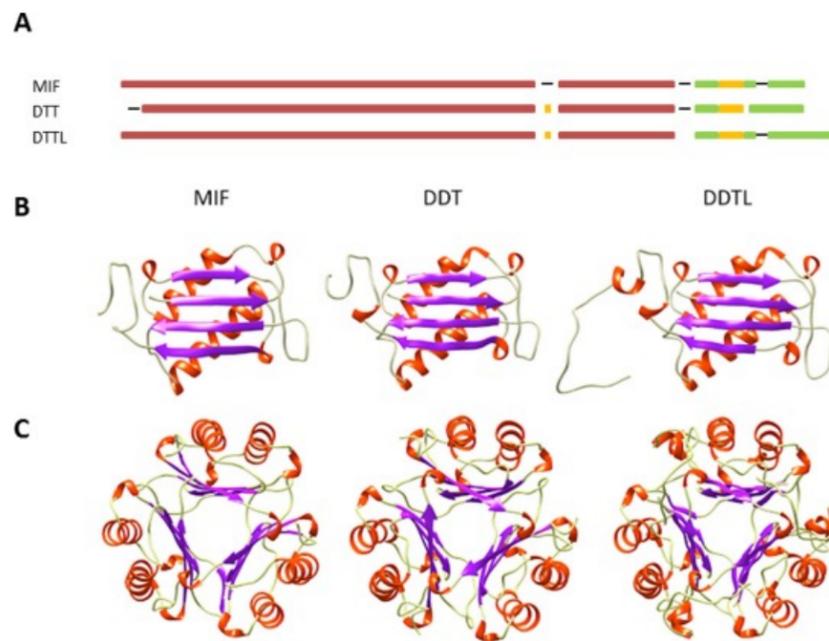


Сопоставление референсной структуры и модели, предсказанной по последовательности с делецией, дает разные результаты в зависимости от выбранной модели. Модели с неструктурированными участками вместо бета-листов я рассматривать не стала. Осталась модель с рангом 2 и 4мя слоями и модель с рангом 3 и 3мя слоями. Для модели ранга два бета-листа более-менее сопоставимы и по последовательности, участвующей в их образовании, и по положению (два верхних листа на рисунке). Третий сверху лист в референсной структуре образуется за счет аминокислот N-конца, делетированных в данном случае. Предсказание выдало неплохое решение: заменить этот слой участком 28Arg-31Val с сохранением направления бета-слоя и образованием четвертого бета-слоя (правда, у него уже противоположное направление).



Тример

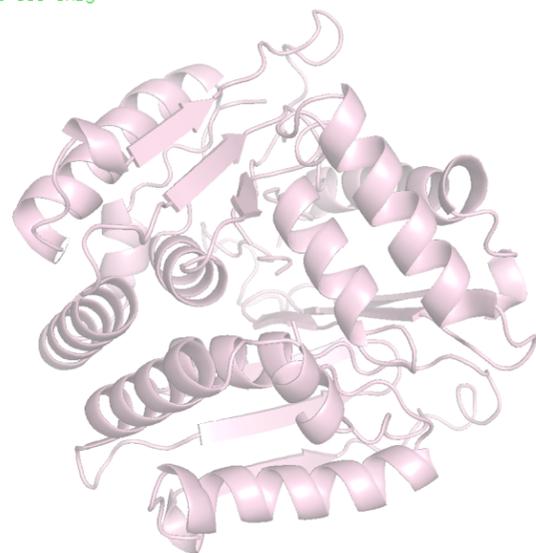
Предсказания выше (без выравниваний) слабо иллюстрируют громкие заголовки об AlphaFold2. Гомологичный белок MIF имеет каталитическую активность в клетке только в качестве тримера, поэтому я решила посмотреть на выдачу AlphaFold2 (без выравниваний) для тримера полной последовательности.



Та же картинка из того же источника

Однако, качество предсказания оставляло желать лучшего. Модель с рангом 1 предсказала 3 альфа-спирали на белок, и в целом структура получилась довольно упорядоченная, но совсем не похожая на то, что я ожидала увидеть.

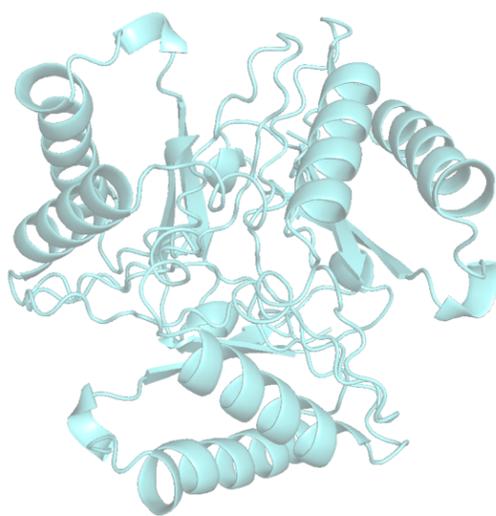
For Educational Use Only



Модель с рангом 1

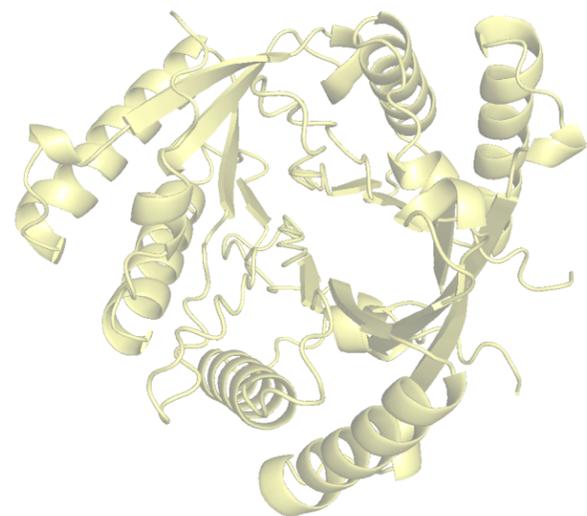
Модели с рангом 2 и 3 имеют в своем составе обращенные вовне альфа-спирали, что мы и ожидаем увидеть, и даже кусочек альфа-спирали на С-конце. Внутри с бета-листами дело обстоит плохо, и предсказанные бета-листы плохо сопоставимы с бета-листами референсной структуры.

For Educational Use Only



Модель с рангом 2

For Educational Use Only



Модель с рангом 3

Модели с рангами 4 и 5 предсказали в качестве тримера стопку из альфа-спиралей, что совсем далеко от истины.

Заключение

В любом случае рассматриваемый фермент с делецией на N-конце не был бы функционален, так как пост-трансляционно N-концевой метионин отрезается, и первый пролин необходим для ферментативной активности белка ([обзор](#)).

Предсказанная AlphaFold2 структура белка с делецией (без построения выравнивания) более-менее правдоподобна, однако сложно доверять ей безоговорочно из-за плохого восстановления полной структуры без выравниваний. С выравниваниями же, судя по всему, AlphaFold2 чересчур опирается на уже известные структуры (петля, образовавшаяся из прежде бета-листа вряд ли продолжила бы занимать в точности то же положение, что и бета-лист). Предсказание структуры тримера не помогло улучшить качество предсказания.

AlphaFold2 - это действительно очень мощный и перспективный инструмент по предсказанию пространственной структуры белка. Однако, в дальнейшем я буду очень осторожно использовать данный инструмент и предварительно по-старинке проверять, есть ли структуры, закодированные ортологичной последовательностью.