

Практикум 7.

Задание 3.

Объект исследования в этом задании – папаин, протеолитический фермент, содержащийся в плодах дынного дерева (папайи). Активный фермент состоит из 212 аминокислот, при этом его предшественник включает 345 и содержит на N-конце сигнальный и активационный пептиды [1]. Его активный центр представлен каталитической триадой Cys25, His159 и Asn175, которая локализована между двумя доменами. Механизм реакции включает, как и для всех протеаз, образование ацильного интермедиата за счет формирования ковалентной связи между атомом серы цистеина и карбонильным углеродом субстрата [2]. Характеризуется широкой субстратной специфичностью, при этом имеет сильное сродство к олигопептидам, содержащим аминокислотные остатки с объемными гидрофобными боковыми группами за 2 позиции с N-конца до разрываемой связи. Структура папаина из базы данных PDB представлена на рисунке 1.

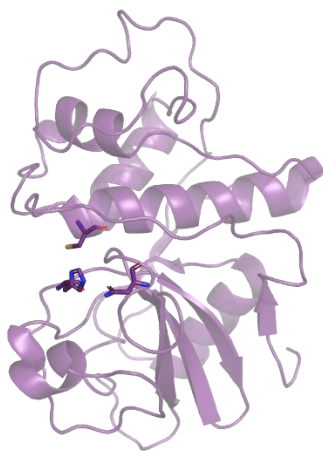


Рисунок 1. Структура папаина (PDB: 1PAD) с выделенным активным центром. Отчетливо видны разделение структуры на два домена и щелевидный карман между ними. [PyMol-сессия](#)

Папаин характеризуется высокой степенью гомологии с другими цистеиновыми гидролазами, в том числе гидролазами млекопитающих, повышенный уровень протеолитической активности которых ассоциирован с определенными заболеваниями. В связи с этим он часто выступает в роли модели для дизайна ингибиторов. В ходе одного из таких исследований была получена структура 1PAD – результат блокирования папаина предполагаемым ингибитором за счет образования ковалентно связанного фермент-субстратного комплекса [3]. Так как в данном случае реакция прошла, можно предполагать, что субстрат в активном центре стабилизирован и расположен оптимальным образом. Авторы отмечают, что связывание ингибитора не нарушает нативную структуру. На эту структуру я буду ориентироваться для оценки результатов AlphaFold. Структура ингибитора представлена на рисунке 2.

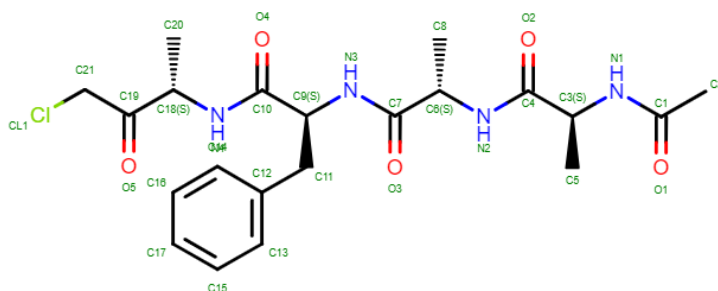


Рисунок 2. Структура ингибитора, имитирующего субстрат папаина

На рисунке 3 показано положение ингибитора в активном центре папаина и стабилизирующие его взаимодействия, в которые вовлечены остатки каталитической триады и остатки, образующие карман связывания. На рисунке отмечены каталитическая триада (Cys25, His159, Asn175), оксианионный центр (Ser24, Gln19) и остатки, связывающие пептид: Val157 и Val133 стабилизируют гидрофобный фенилаланин (расположен как раз за два остатка до разрываемой связи), Tyr67 образует водородные связи.

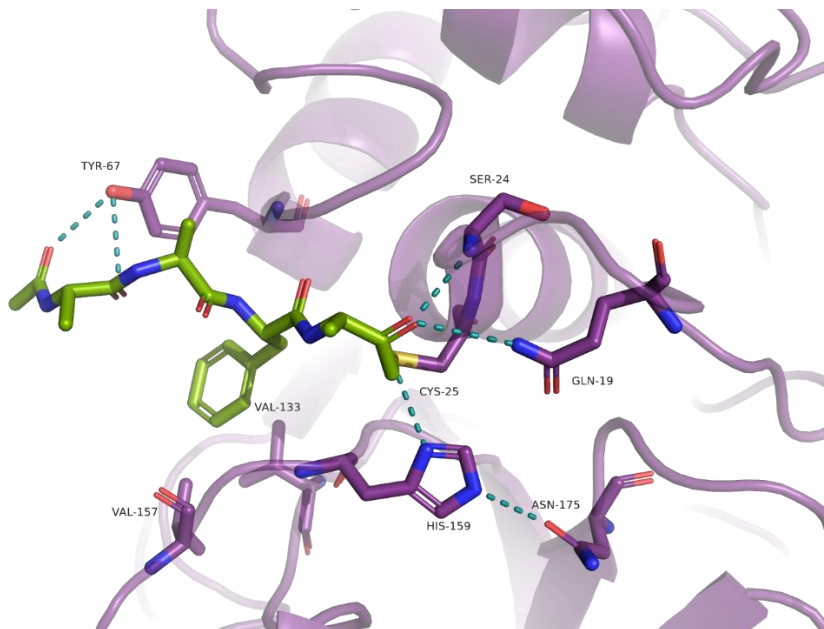


Рисунок 3. Связывание аналога субстрата в активном центре папаина. [PyMol-сессия](#)

Теперь попробуем предсказать эту структуру методом с помощью AlphaFold. Я запустила скрипт, добавив в параметрах оптимизацию 5 полученных структур с помощью AMBER и увеличив число шагов оптимизации до 2000, чтобы повысить качество итоговых моделей. Так как предстоит моделировать связывание субстрата, я взяла только последовательность активированного фермента, не учитывая 1-133 остатки оригинальной последовательности. Стоит отметить, что качество предсказания для всех полученных моделей распределено относительно аминокислотных остатков приблизительно одинаково. На рисунке 4 можно видеть соответствующий график. Метрика IDDT характеризует «уверенность» модели в позиции каждого остатка.

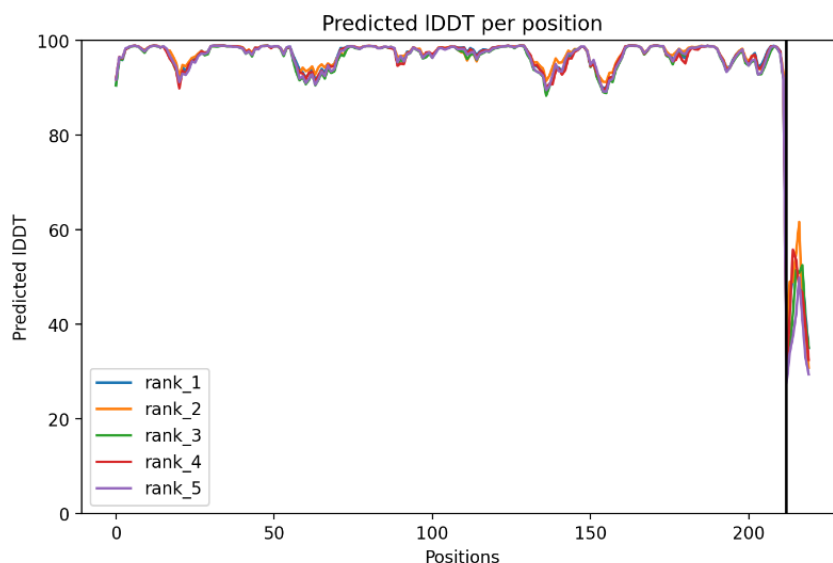


Рисунок 4. Качество предсказаний для каждой позиции

Как видно из графика, AlphaFold не смог точно определить положение в пространстве для остатков 213-221, что соответствует пептиду, следовательно, его положение ни в одной из моделей не заслуживает доверия по версии самого же AlphaFold. В то же время структура самого белка предсказана с достаточно убедительным качеством. На рисунке 5 показаны структуры папаина по версии AlphaFold. Серым цветом показана настоящая структура папаина, и видно, что полученные модели хорошо ей соответствуют.

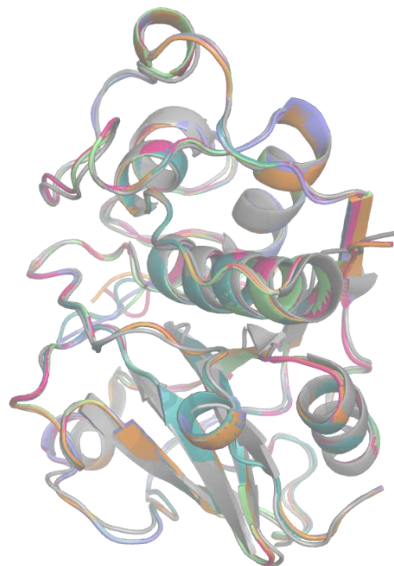


Рисунок 5. Структуры папаина из AlphaFold. [PyMol-сессия](#)

Далее рассмотрим положение пептида. Во всех пяти структурах пептид занял нужный карман, но из-за моей невнимательности я только на этапе анализа результатов поняла, что в нем отсутствует остаток с объемной боковой цепью, следовательно, пептид может не связаться в кармане не из-за ошибки AlphaFold, а из-за физических свойств поверхности кармана. Вероятно, это и происходит, так как сближение карбонильного углерода с атомом серы цистеина не достигается ни в одном из пяти вариантов. Более того, в двух случаях вместо ацильного интермедиата, по-видимому, образуется дисульфидный мостик. Отсутствие остатка с объемной боковой цепью также создает трудности с определением того, какая связь должна расщепляться ферментом. На рисунках 6А и 6В показаны структуры с рангами 1 и 2, которые иллюстрируют качество помещения субстрата в активный центр. Как мы видим, в обоих случаях алгоритм стремится поместить треонины в гидрофобный карман, так как у них есть неполярная метильная группа, тогда как остальные остатки на N-конце вообще являются заряженными. В структуре с рангом 1 произошло образование дисульфидного мостика, так как цистеин пептида слишком сблизился с цистеином активного центра. В обоих случаях цепи пептидов слишком далеко от активного центра, чтобы могла произойти нуклеофильная атака.

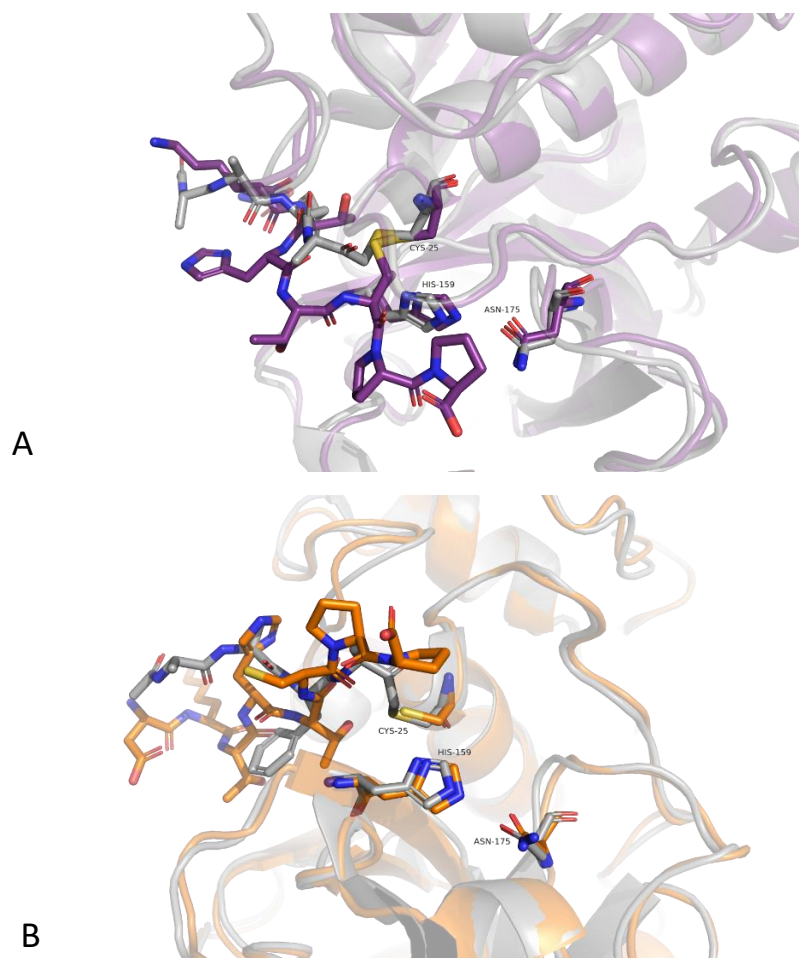


Рисунок 6. Положение субстрата в активном центре. А - структура ранга 1 ([сессия](#)), В - структура ранга 2 ([сессия](#)). Серым цветом показана структура из базы PDB

Я решила проверить свою гипотезу: заменить в пептиде гистидин на фенилаланин и повторить моделирование. Новый график качества представлен на рисунке 7, из него видно, что хотя качество предсказания структуры пептида остается ниже, чем для остальной структуры, уверенность модели несколько возросла по сравнению с предыдущим запуском.

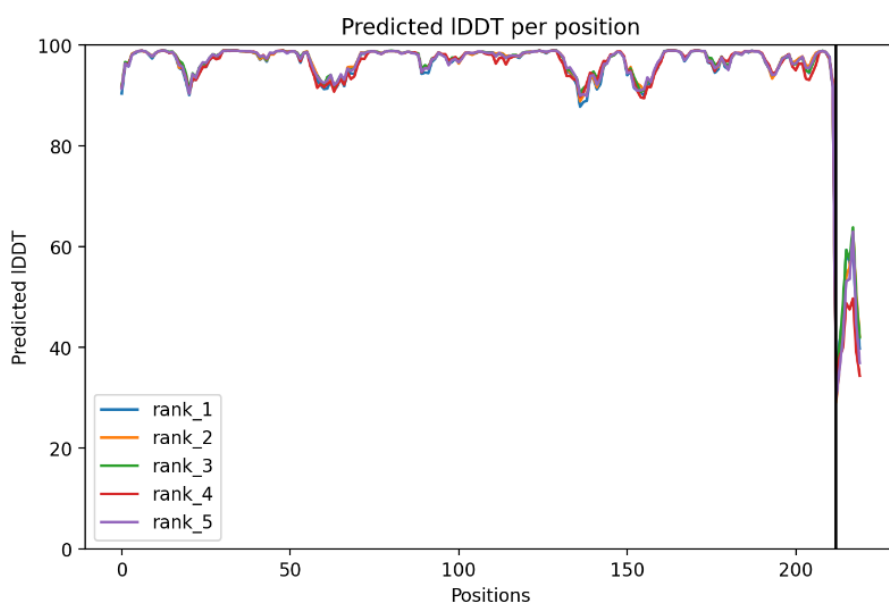


Рисунок 7. Качество предсказаний для каждой позиции

При сравнении положения пептида в кармане активного центра различия очень существенные. При предыдущем запуске во всех пяти структурах с наивысшим рангом положение субстрата было разным, менялось даже положение С и N-концов. Теперь же мы видим, что все вариации положения пептида довольно хорошо совпадают друг с другом, что свидетельствует о более высоком качестве предсказания, так как алгоритм раз за разом находит одну и ту же конформацию. На рисунке 8 представлена первые две по рангу модели. В том, что положение пептида в остальных предсказанных моделях практически совпадает с продемонстрированным на этом рисунке, можно убедиться, просмотрев PyMol-сессию по указанной ссылке.

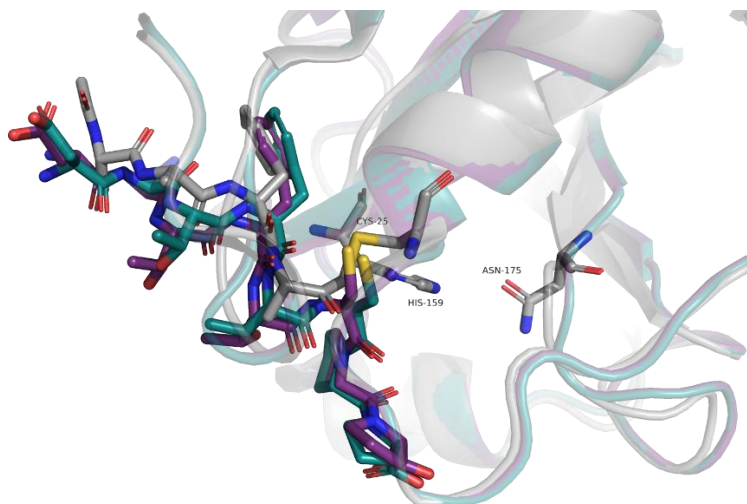


Рисунок 8. Положение субстрата в структурах с рангами 1 (зеленый) и 2 (фиолетовый). Серым показана структура из базы PDB. [PyMol-сессия](#)

Как видно из рисунка, фенилаланин занимает предназначенное для него положение. Исходя из субстратной специфичности фермента, в ходе реакции должна быть разорвана связь между треонином и цистеином. На рисунке 9 на примере структуры ранга 1 сферами показаны карбонильный атом углерода, который должен подвергнуться нуклеофильной атаке со стороны Cys25, и атакующий атом серы. Расстояние между ними составляет 5.1 ангстрем, что слишком далеко для нуклеофильной атаки (оптимальное значение 3 ангстрема и меньше), при этом снова происходит образование дисульфидного мостика. О попадании атома кислорода около разрываемой связи в оксианионный центр речи также не идет. Таким образом, AlphaFold не смог поместить субстрат в активный центр настолько хорошо, чтобы можно было говорить о моделировании предреакционного состояния, но в целом его положение было определено достаточно качественно, а при втором запуске даже однозначно. Правда, эта однозначность, вероятно, обоснована и тем, что все получившиеся структуры содержат ковалентную связь между цистеинами пептида и активного центра, которой на самом деле быть не должно, если речь идет о реакции гидролиза данного пептида. Эта дополнительная связь, которая стабилизирует данное положение субстрата, в нашем случае скорее мешает оптимизировать структуру. Сближение атома серы активного центра именно с -SH-группой субстрата вместо карбонильного углерода основной цепи можно объяснить тем, что AlphaFold делает предсказание структуры на основе уже известных, а в случае папаина структуры из базы данных PDB содержат в качестве субстратов ингибиторы, которые связываются с активным центром иным способом в сравнении с нативным субстратом. Структур же с нативным субстратом расшифровано не было. Как можно видеть на рисунке 8, пептид подгоняется по форме под ингибитор, но так как связывание цистеином ингибитора происходит не в середине его цепи, а на конце, рассматриваемый пептид также стремится связаться концом, в качестве которого выступает реакционноспособный боковой радикал цистеина. Неудача в данном случае, как мне кажется, на самом деле связан не только с безупречной работой AlphaFold, но и со спецификой субстрата. Вероятно, если бы на месте цистеина был другой остаток, пептид мог бы глубже войти в карман. С другой стороны, первый

запуск AlphaFold показал, что не подходящий ферменту субстрат будет смоделирован в активном центре с гораздо меньшей «уверенностью», его положение будет сильно различаться в разных структурах. Следовательно, AlphaFold по крайней мере способен определить, свяжет ли фермент данный субстрат, и в первом случае мы можем говорить о том, что связать субстрат не удалось, так как не было зафиксировано его однозначного положения. Этот случай по сути является контролем, который можно было бы предложить для валидации предсказаний AlphaFold: попытка поместить в активный центр пептид, который не является предпочтительным субстратом данного фермента. Свидетельством хорошей работы алгоритма в данном случае будет снижение качества предсказания.

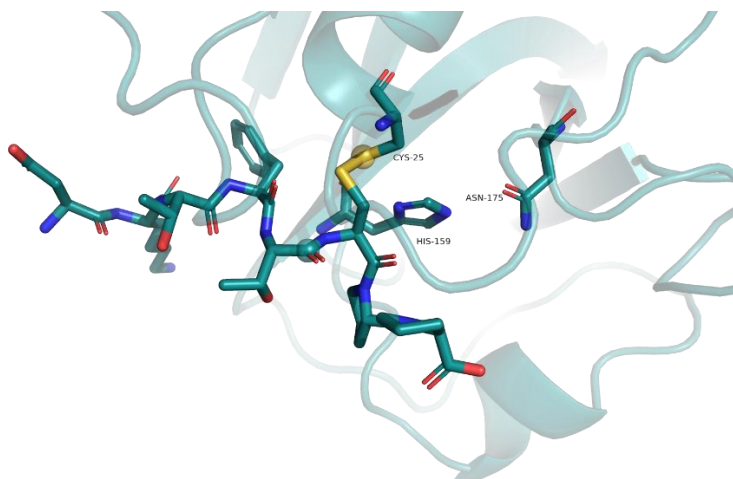


Рисунок 9. Структура ранга 1 с выделенными сферами реагирующими атомами. [PyMol-сессия](#)

Я решила проверить, получится ли сблизить реагирующие атомы, если заменить цистеин на глицин, устранив таким образом препятствия в виде бокового радикала. Для этой замены у меня уже нет биологических предпосылок, и я руководствуюсь исключительно любопытством.

Замена цистеина на глицин еще больше повысила уверенность предсказаний (рисунок 10). Вероятно, это связано с тем, что глицин может образовывать меньше взаимодействий с соседними остатками и не имеет различных конформаций боковой цепи.

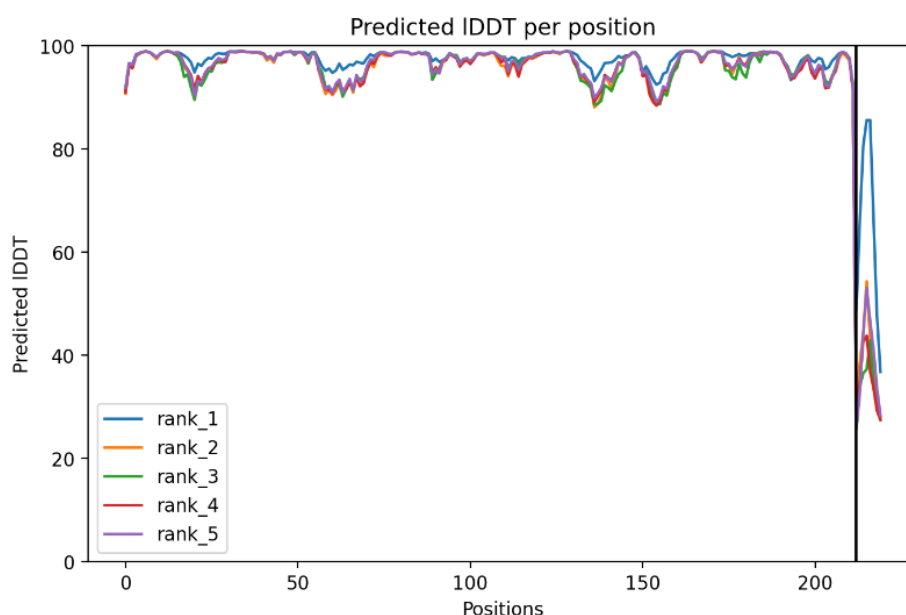


Рисунок 10. Качество предсказаний для каждой позиции

Предположение относительно роли дисульфидного мостика в стабилизации субстрата подтвердилось, так как после удаления цистеина положение субстрата снова стало меняться от структуры к структуре. В то же время эта замена позволила приблизить карбонильный углерод разрываемой связи к нуклеофильной сере цистеина (рисунок 11). Я полагаю, это можно считать успехом AlphaFold, хотя и стоит отметить, что выбранная мной как оптимальная структура ранга 2 не является лучшей по мнению алгоритма, который намного больше уверен в структуре ранга 1. Таким образом, можно сделать вывод, что AlphaFold может справляться с задачей помещения субстрата в активный центр, но перед использованием в работе его модели должны по возможности валидироваться исследователем.

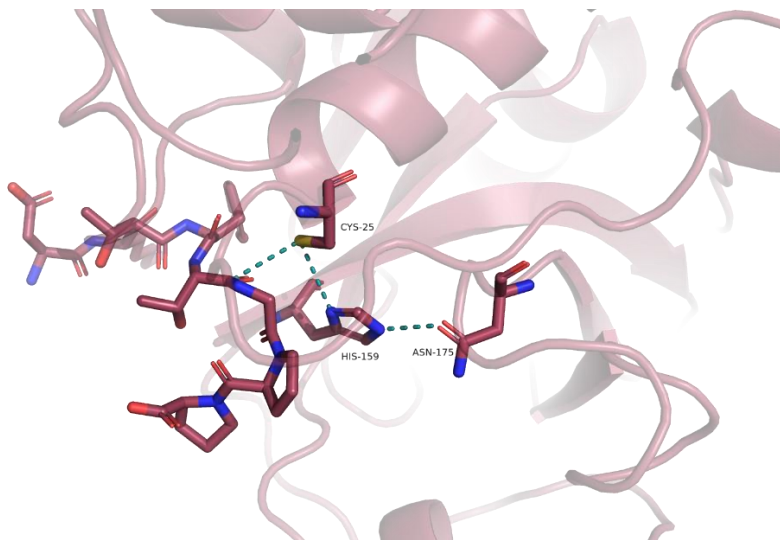


Рисунок 11. Предсказанная структура (ранг 2), близкая к состоянию нуклеофильной атаки. Расстояние между реагирующими атомами составляет 3.5 ангстрем. [PyMol-сессия](#)

Задание 4.

Далее я попробовала смоделировать N-концевую делецию. По результатам запроса в BLAST, доставшийся мне белок – D-допахрома таутомераза (декарбоксилаза), фермент, катализирующий таутомеризацию D-допахрома с образованием 5,6-дигидроксииндола и отщеплением CO₂. Этот процесс играет важную роль в метаболизме меланина. Одна биологическая единица представляет собой гомотример (рисунок 12), доменов в составе отдельных субъединиц не выделено [4]. Согласно исследованиям, такая структура является эволюционно консервативной. Пространство, ограниченное бета-листами трех доменов, представляет собой канал растворителя, который важен для каталитического механизма. Например, было показано, что объем канала прямо пропорционально соотносится со скоростью реакции, так как в катализ вступают молекулы воды. Диаметр канала определяется размером боковых радикалов аминокислот, которые образуют ворота [5].

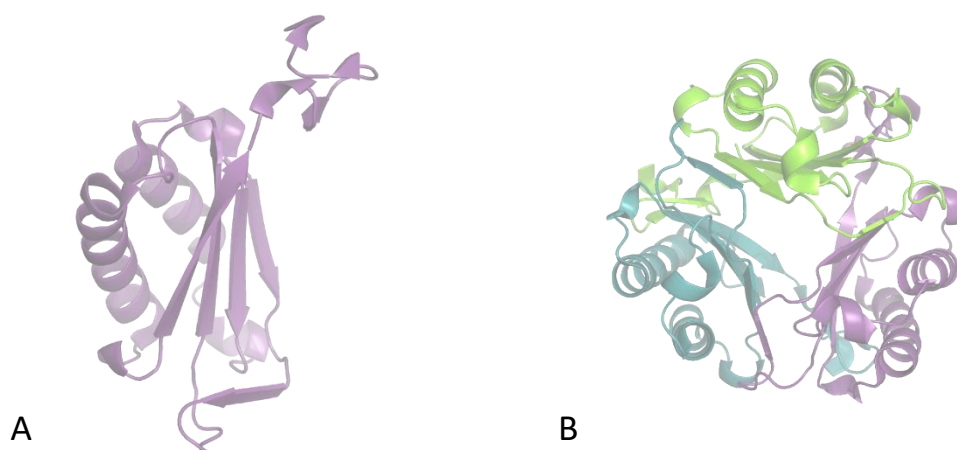


Рисунок 12. Структура D-допахром декарбоксилазы (PDB: 3KER). А - мономер ([сессия](#)), В - гомотример, биологическая единица ([сессия](#))

В рамках задания алгоритму AlphaFold были переданы сначала полная последовательность белка, затем белок с N-концевой делецией. Результат предсказания структуры полного белка представлен на рисунке 13. Как видно из рисунка, упорядоченные элементы структуры предсказываются сравнительно хорошо: обе альфа-спирали присутствуют во всех пяти моделях, также в них присутствуют три бета-цепи из четырех. Четвертая бета-цепь частично обнаружена в одной из моделей (на рисунке показана фиолетовым). По-видимому, данная последовательность белка длиннее большинства из базы данных, поэтому структуру С-конца достоверно предсказать не удастся. Однако наблюдается тенденция к формированию стабильной структуры – пятой бета-цепи, которая должна войти в состав имеющегося бета-листа. На рисунке 14 (модель ранга 1) видно, что на самом деле в структуре в связи с увеличенной длиной имеет место перестройка: аминокислотные остатки, которые в структуре из базы данных формировали четвертую бета-цепь, в данном случае формируют пятую, а потенциальная четвертая образуется С-концом, который, сформировав неупорядоченный участок, возвращается обратно.

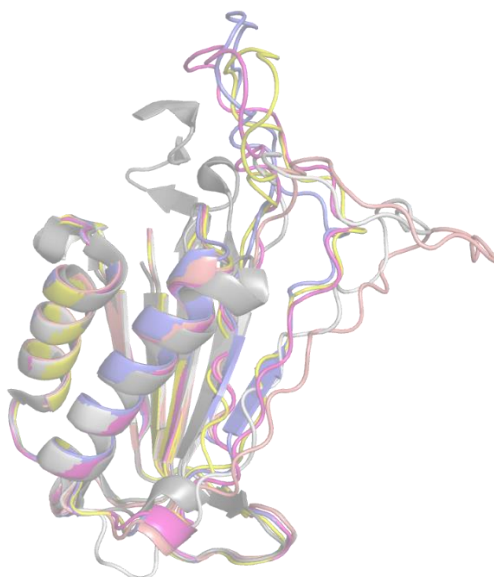


Рисунок 13. Предсказание структуры полного белка. Серым цветом показана структура из базы PDB. [PyMol-сессия](#)

На рисунке 14 цветом выделен участок, который для дальнейших расчетов был удален из последовательности. Он включает одну бета-цепь и часть альфа-спирали.



Рисунок 14. Модель ранга 1 и делетируемый участок (показан зеленым)

Предсказание структуры после делеции пошло по двум разным путям, которые показаны на рисунке 15. С-конец по-прежнему предсказывается плохо, поэтому не акцентируем внимания на нем и рассмотрим только N-конец.

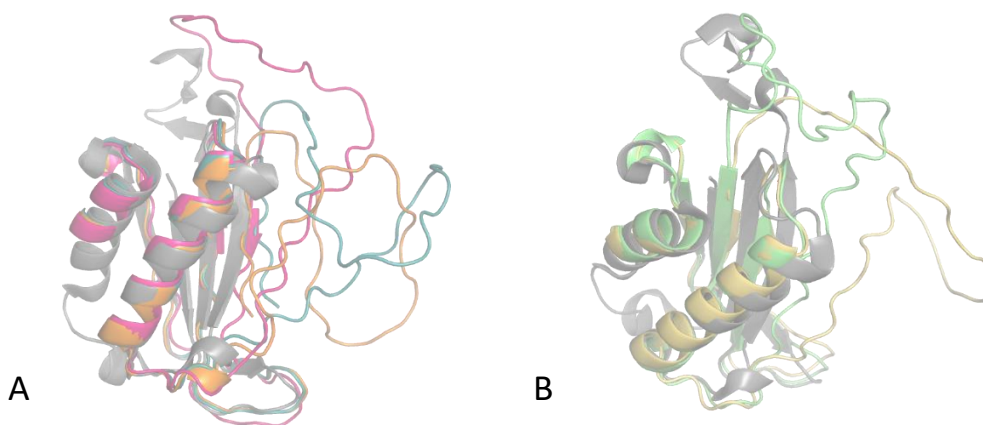


Рисунок 15. Предсказание структуры белка с N-концевой делецией. Серым цветом показана структура из базы PDB. [PyMol-сессия](#)

Вариант на рисунке 15A предполагает, что теряемая при делеции бета-цепь из центра бета-листа не заменяется, за счет чего практически весь бета-лист разрушается, оставляя только маленький участок 3 и 4 цепей. И хотя водородных связей между бывшими частями бета-листа уже нет вследствие увеличения расстояния, это не приводит к перестройкам, в белке просто появляется больше неупорядоченных участков. Такой расклад невыгоден с точки зрения физики белка, и в отсутствие водородных связей между соседними цепями они не будут оставаться в том же положении, в котором они были в бета-листе. Следовательно, такая гипотеза скорее всего ошибочна. На рисунке 15B представлен другой вариант: «избыточный» С-конец встает между 1 и 3 бета-цепями, замещая потерянную вторую. Это позволяет сохранить стабильность бета-листа и фолд белка. Если такие преобразования возможны в живых системах, вероятно, такое свойство могло бы сохранить белок функциональным в случае утраты N-конца. Вариант 15B показывает, что AlphaFold способен менять положение участков последовательности в пространстве для лучшей их стабилизации, и это его плюс.

А в конце я решила попробовать предсказать с помощью AlphaFold структуру гомотримера. Рисунок 16 свидетельствует о том, что, не считая неупорядоченных С-концевых аминокислот, которым не находится аналога в базах данных структур, алгоритм достаточно качественно предсказал строение биологической единицы, и это тоже очередная победа AlphaFold.

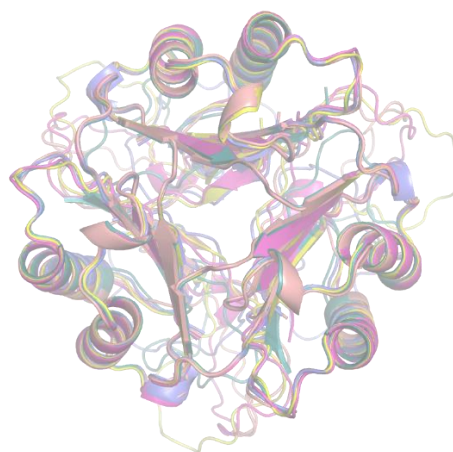


Рисунок 16. Предсказание структуры гомотримера. [PyMol-сессия](#)

Список источников:

1. <https://www.uniprot.org/uniprotkb/P00784/entry>
2. Judith M. LaLonde, Baoguang Zhao et al., Use of Papain as a Model for the Structure-Based Design of Cathepsin K Inhibitors: Crystal Structures of Two Papain–Inhibitor Complexes Demonstrate Binding to S'-Subsites. Journal of Medicinal Chemistry 1998 41 (23), 4567-4576
3. J. Drenth, K. H. Kalk, and H. M. Swen, Binding of chloromethyl ketone substrate analogs to crystalline papain. Biochemistry 1976 15 (17), 3731-3738
4. <https://www.uniprot.org/uniprotkb/P30046/entry>
5. Chen, E., Reiss, K., et al., A structurally preserved allosteric site in the MIF superfamily affects enzymatic activity and CD74 activation in D-dopachrome tautomerase. J Biol Chem 297: 101061-101061 (2021)