

Структурная биоинформатика

Практикум 10. AlphaFold2.

Поддьяков Иван

Введение	2
Предсказание 1	3
Предсказание 2	7
Источники	9

Введение

В данном практикуме я изучил возможность предсказания структуры белка с помощью AlphaFold2 [1] на платформе ColabFold. Я выбрал вариант задания с предсказанием структуры протеазы папаина и потенциального олигопептидного субстрата. Папаин - цистеиновая эндопептидаза, выделенная из папайи. Данный фермент применяется в косметической промышленности, для обработки кожи, в фармакологии. Я буду работать с последовательностью белка [P00784](#) и последовательностью пептида DKTHTCPP, который указан как физиологически значимый субстрат [2].

[Ссылка](#) на папку с предсказаниями.

Предсказание 1

Сначала был произведен запуск AlphaFold2 на полной последовательности гена и пептида. Я рассмотрел предсказание с рангом 1, которое представлено на Рисунке 1. Для сравнения я выбрал существующую структуру активной формы этого же белка - [3E1Z](#), на которую наложил предсказанную структуру с помощью команды super (RMSD 0.368). Предсказание накладывается на известную структуру очень хорошо, однако пептид не занимает нужного положения в активном сайте, а расположен где-то в стороне. В активном центре же AlphaFold2 прокладывает участок, соответствующий пер-про-папаину.

На Рисунке 2 показаны каталитические остатки фермента. Аспарагин-105 стабилизирует гистидин-159, который, в свою очередь, депротонирует цистеин-25. Последний нуклеофильно атакует карбонильный атом углерода гидролизуемой цепи [3]. Данные остатки в предсказанной структуре хорошо соответствуют таковым из существующей, а напротив цистеина находится как раз карбонильный углерод того остатка, по которому идет разрезание цепи в момент автокатализа [4].

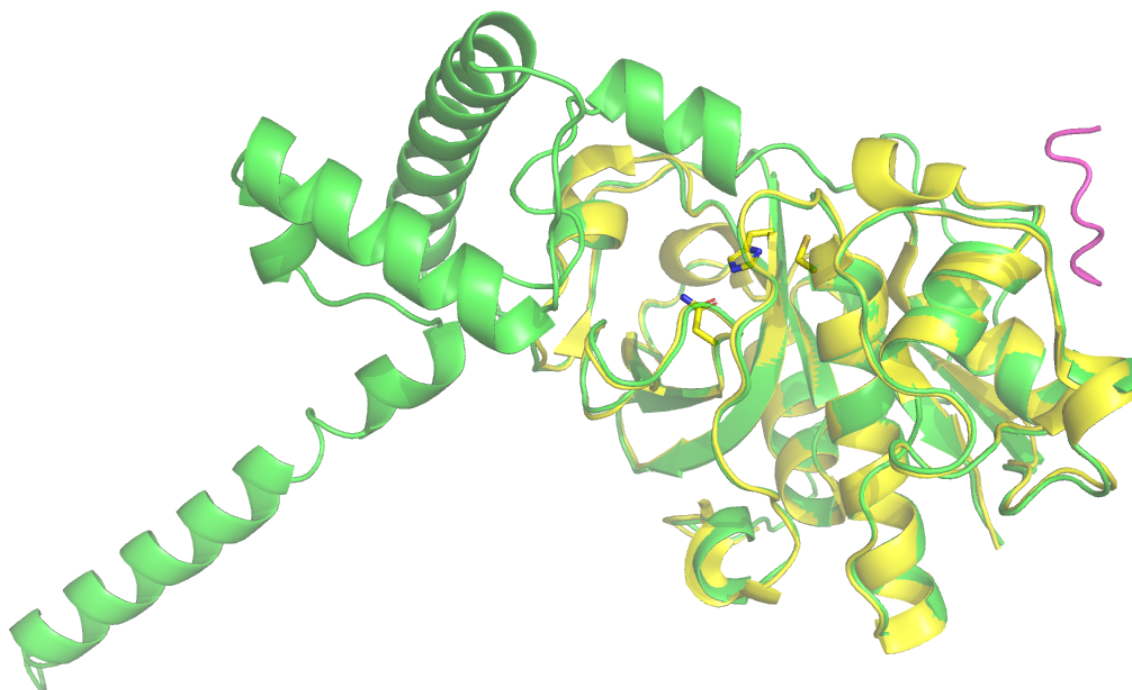


Рисунок 1. Предсказание укладки полной последовательности - зеленым. Известная структура папаина 3E1Z - желтым. Фиолетовым показан олигопептид.

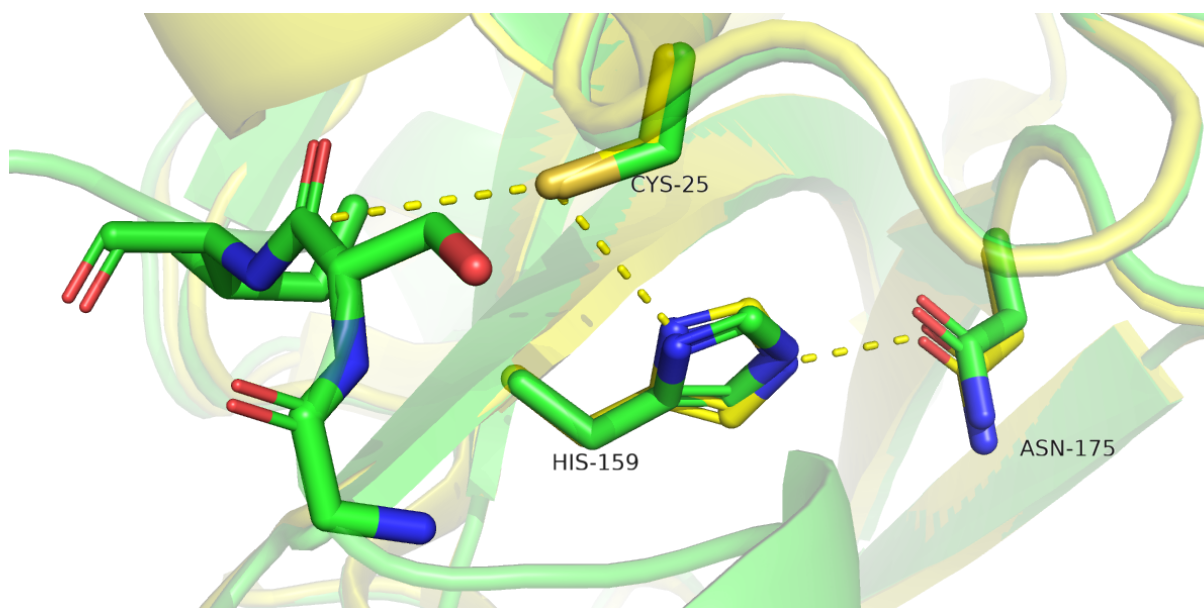


Рисунок 2. Каталитическая триада папаина. Предсказание укладки полной последовательности - зеленым. Известная структура папаина 3E1Z - желтым.

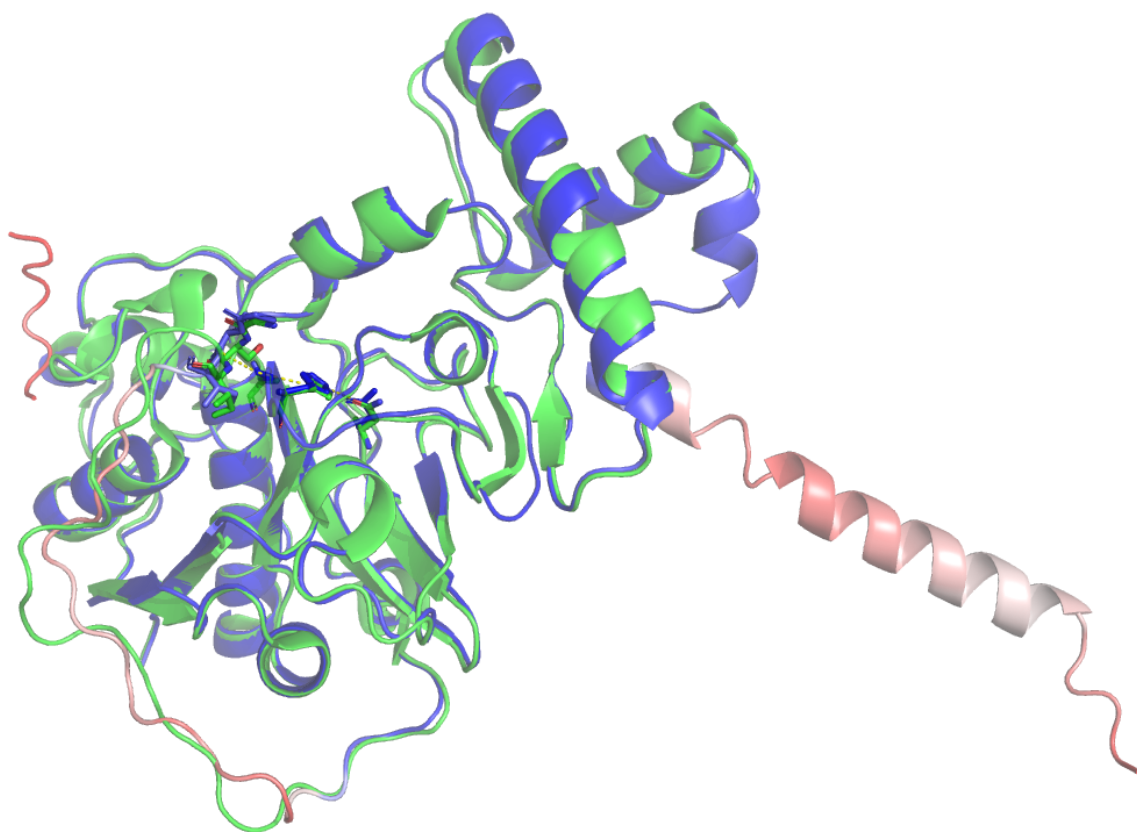


Рисунок 3. Предсказание укладки полной последовательности, покрашено по pLDDT. Известная структура про-папаина 3TNX - зеленым.

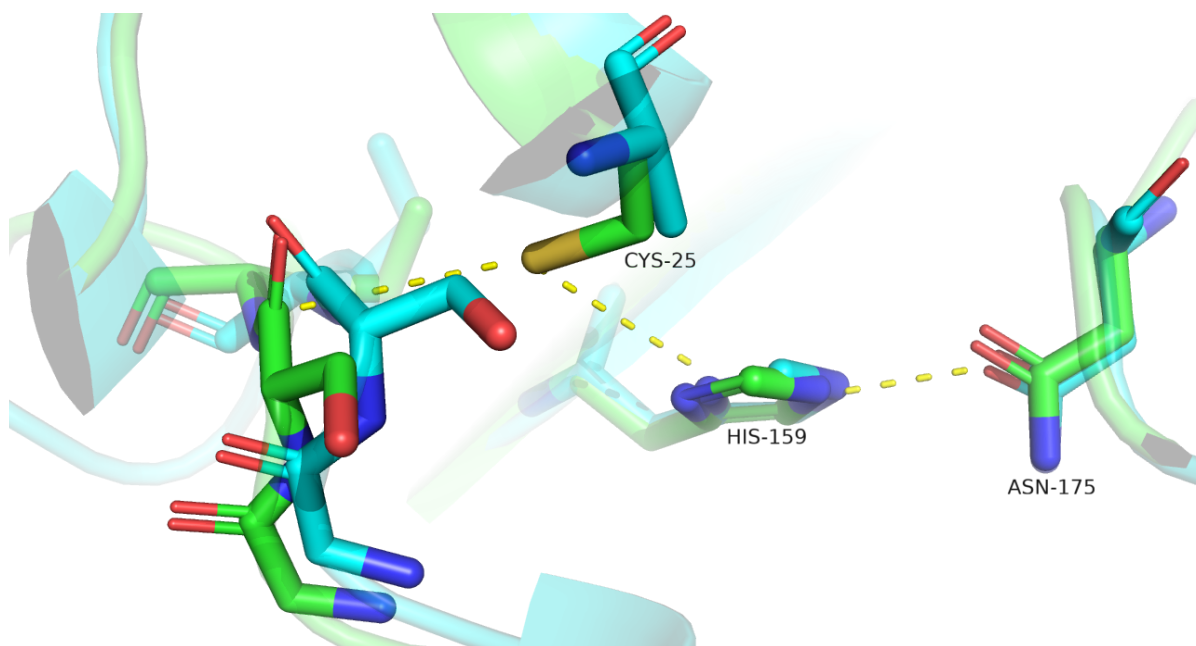


Рисунок 4. Каталитическая триада папаина. Предсказание укладки полной последовательности - зеленым. Известная структура папаина 3TNX - голубым.

Я также совместил известную структуру про-папаина 3TNX, которая включает 27-133 остатка, блокирующие активный центр фермента, с предсказанной AlphaFold2 укладкой (Рисунок 3). На известные остатки предсказанная структура ложится также хорошо, RMSD 0.507, хуже накладывается протяженный участок, находящийся на рисунке слева внизу. На Рисунке 4 показана каталитическая триада фермента (в структуре про-папаина цистеин заменен на аланин, чтобы предотвратить автокатализ), наложение также хорошее. В предсказаниях более низких рангов наложение сравнимо по точности, а пептид не заходит в активный центр.

Рассмотрим метрики качества, которые AlphaFold2 предоставляет по результатам работы. На Рисунке 5 показано распределение количества доступных последовательностей для предсказания каждой позиции в текущей задаче. На Рисунке 3 предсказанная структура покрашена по параметру pLDDT - степень уверенности в правильности помещения остатка (возрастает от красного к синему). Можно заметить, что N-концевой участок предсказан с малой долей уверенности и что для него как раз мало гомологичных последовательностей, по которым можно было бы предсказать структуру. Для участков же внутри глобулы такое покрытие кажется избыточным.

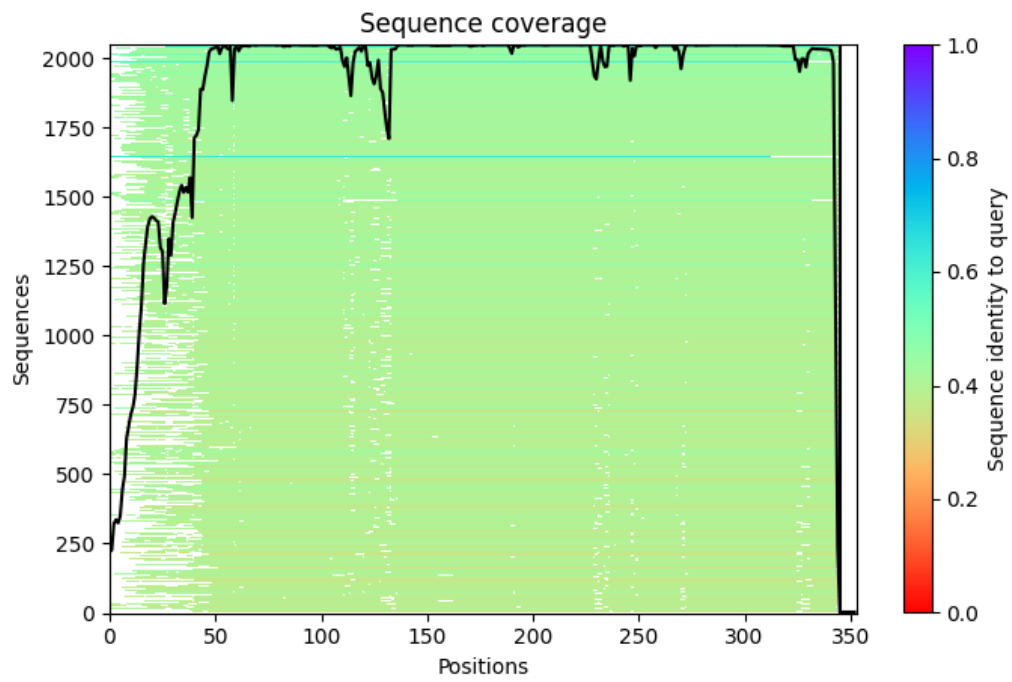


Рисунок 5. Ширина выравнивания в каждой позиции для предсказания.

Предсказание 2

Теперь повторим предсказание из 1 пункта, но возьмем только остатки со 134 (зрелый фермент), чтобы не закрывать активный сайт от субстрата. AlphaFold2 также хорошо предсказывает структуру самого фермента, молекулы субстратов во всех 5 предсказаниях находятся рядом с активным сайтом. Рассмотрим, как субстрат располагается в лучшем предсказании (Рисунок 6). Видно, что нейросеть расположила атомы в субстрате на очень малых расстояниях, а связи - под неестественными углами. Во всех пяти предсказаниях цистеин лиганда “образует” мостик с цистеином-25 активного центра. Для данного предсказания мне не удалось найти сколько-нибудь значимых взаимодействий между субстратом и лигандом. Стоит отметить, однако, что связь His-P1 - Tre-P1', которая должна гидролизываться в данном субстрате, лежит напротив цистеина-25, чего не наблюдается в других предсказаниях.

Субстрат, положение которого предсказывается, является частью тяжелой цепи IgG1 человека, а цистеин, входящий в его состав, в момент разрезания образует мостик с цистеином другой тяжелой цепи [5], поэтому никак не может быть направлен в сторону сайта узнавания протеазы. Таким образом, AlphaFold2 не смог предсказать релевантный комплекс субстрат-фермент в данном случае. Возможно, сказалось отсутствие данных о коэволюции этого пептида и фермента, в естественных условиях они не должны взаимодействовать и полученных структур с взаимодействием папаина и IgG1 нет (в PDB я не нашел записей, содержащих в разных UniProt Molecule Name ‘papain’ и ‘immunoglobulin’).

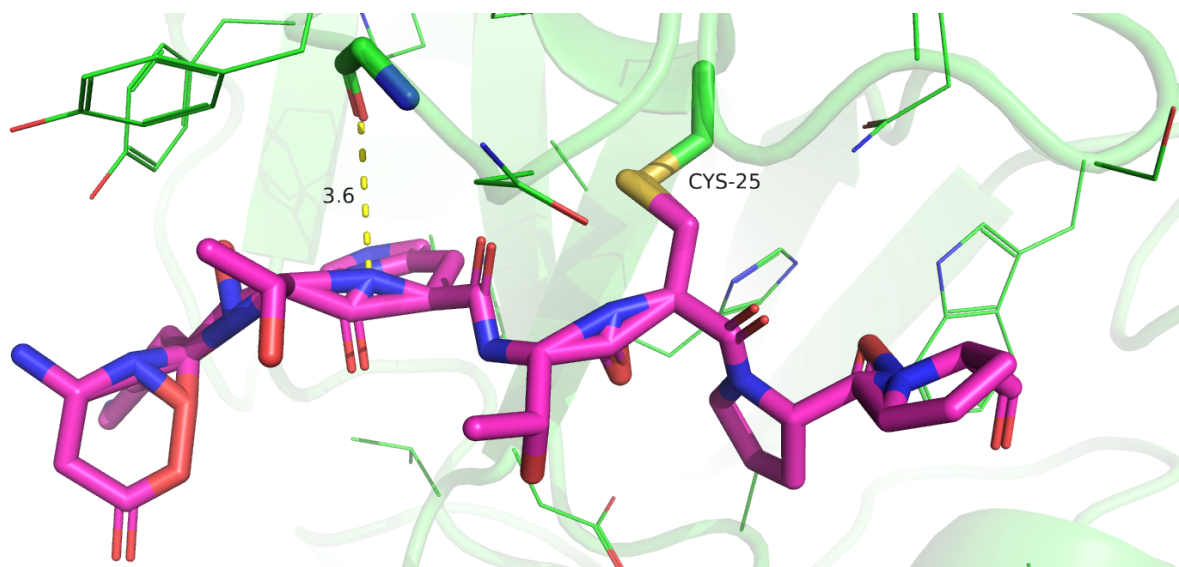


Рисунок 6. Субстрат-пептид (фиолетовым), предсказание 1 ранга. Предсказанный фермент - зеленым.

В литературе мне удалось найти только предположение о способе узнавания субстрата $\text{Ala}_2\text{PheAla}_4$ папаином (Рисунок 7). Вероятно, узнавание конкретного субстрата происходит еще и с учетом его вторичной структуры, что AlphaFold2 не предсказывает для октапептида. Мне кажется, полученный мной вывод о плохом предсказании нельзя обобщать, тем более что существуют примеры корректного предсказания комплексов белок-пептид [6]. Я затрудняюсь сказать, можно ли использовать AlphaFold2 для предсказания комплексов протеаз с субстратами. Для этих целей я бы брал больший участок субстрата, желательно с известной вторичной структурой. Это необходимо, так как данная программа не решает проблему свертки de novo, а учится на выравнивании гомологичных последовательностей. В качестве контроля можно попробовать предсказать, как будет взаимодействовать с ферментом самая часто встречающаяся последовательность субстрата из [MEROPS](#) для данной протеазы. Если и в таком случае предсказание будет плохим, то пользоваться AlphaFold для подобных целей точно не стоит.

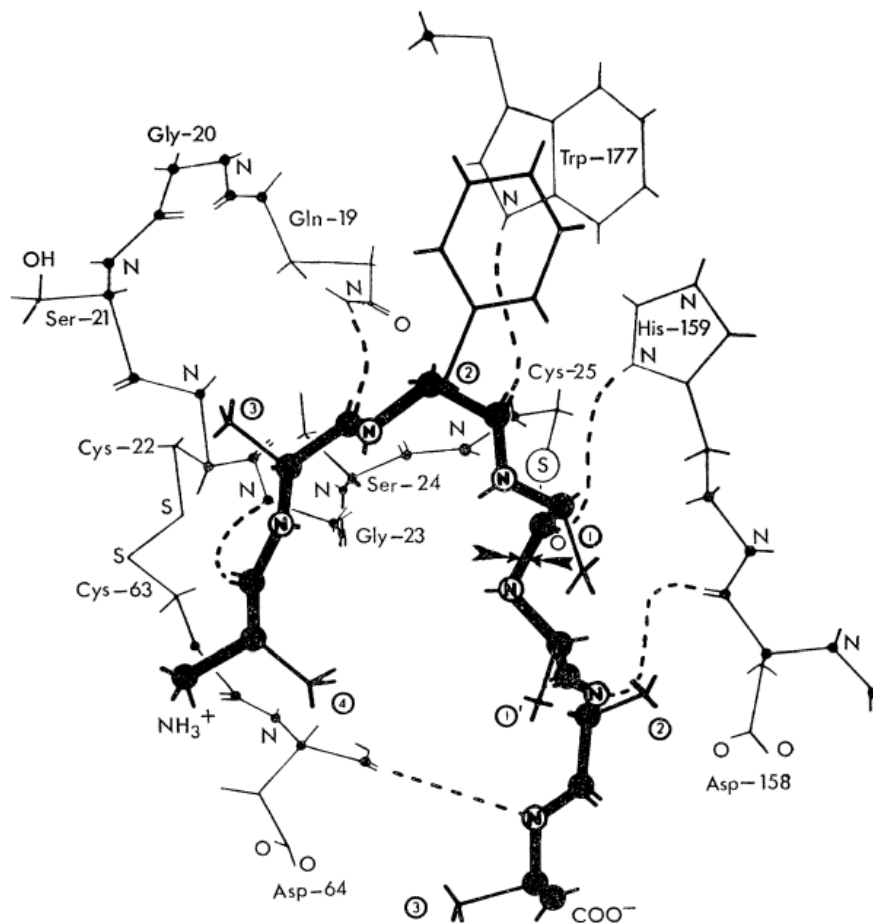


Рисунок 7. Предположительный способ узнавания субстрата $\text{Ala}_2\text{PheAla}_4$ папаином [7].

Источники

1. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596: 583–589.
2. Ryan MH, Petrone D, Nemeth JF, Barnathan E, Björck L, Jordan RE. Proteolysis of purified IgGs by human and bacterial enzymes in vitro and the detection of specific proteolytic fragments of endogenous IgG in rheumatoid synovial fluid. *Mol Immunol*. 2008;45: 1837–1846.
3. Theodorou LG, Lympieropoulos K, Bieth JG, Papamichael EM. Insight into the catalysis of hydrolysis of four newly synthesized substrates by papain: a proton inventory study. *Biochemistry*. 2001;40: 3996–4004.
4. Roy S, Choudhury D, Aich P, Dattagupta JK, Biswas S. The structure of a thermostable mutant of pro-papain reveals its activation mechanism. *Acta Crystallogr D Biol Crystallogr*. 2012;68: 1591–1603.
5. Wang AC, Wang IY. Cleavage sites of human IgG1 immunoglobulin by papain. *Immunochemistry*. 1977;14: 197–200.
6. Ko J, Lee J. Can AlphaFold2 predict protein-peptide complex structures accurately? *bioRxiv*. 2021. p. 2021.07.27.453972. doi:10.1101/2021.07.27.453972
7. Berger A, Schechter I. Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Philos Trans R Soc Lond B Biol Sci*. 1970;257: 249–264.