

Практикум 10. AlphaFold2

Иосиф Финкельберг

1 Введение

Катепсин К (CatK) – это лизосомальная цистеиновая протеаза человека, главная физиологическая функция которой заключается в ресорбции костей. В ходе сигнального пути RANK стимулирующие факторы (в частности, receptor activator of nuclear factor- κ B, RANKL) связываются с рецепторами клеток-предшественниц остеокластов [Jin Hee Park (2017)]. Это приводит к дифференцировке клеток-предшественниц в остеокласты и инициирует в них синтез CatK посредством транскрипционного фактора NFATc1 (nuclear factor of activated T cells). Остеокласты выделяют CatK в так называемую 'sealing zone': богатую актином, закисленную зону ресорбции кости. В 'sealing zone' CatK эффективно разрушает коллаген, остеоонектин и другие белки матрикса кости [Dai et al. (2020)]. Интересно, что CatK, видимо, является единственной протеазой млекопитающих, способной разрезать тройную спираль коллагена сразу в нескольких позициях [Novinec and Lenarčič (2013)]. Одним из характерных для пожилых людей заболеваний является остеопороз, проявляющийся в снижении плотности костей вследствие активности CatK. Важной задачей является разработка ингибиторов катепсина К, так как их можно использовать в качестве лекарств от остеопороза (однако, насколько нам известно, ни один из ингибиторов CatK до сих пор так и не удалось зарегистрировать в FDA).

Как уже было сказано, катепсин К относится к цистеиновым гидролазам. Его активный центр состоит из остатков Cys-25 и His-162 (здесь и далее нумерация дается по зрелой форме белка человека). На рис. 1 приведена характерная для цистеиновых протеаз схема реакции гидролиза пептидной цепи.

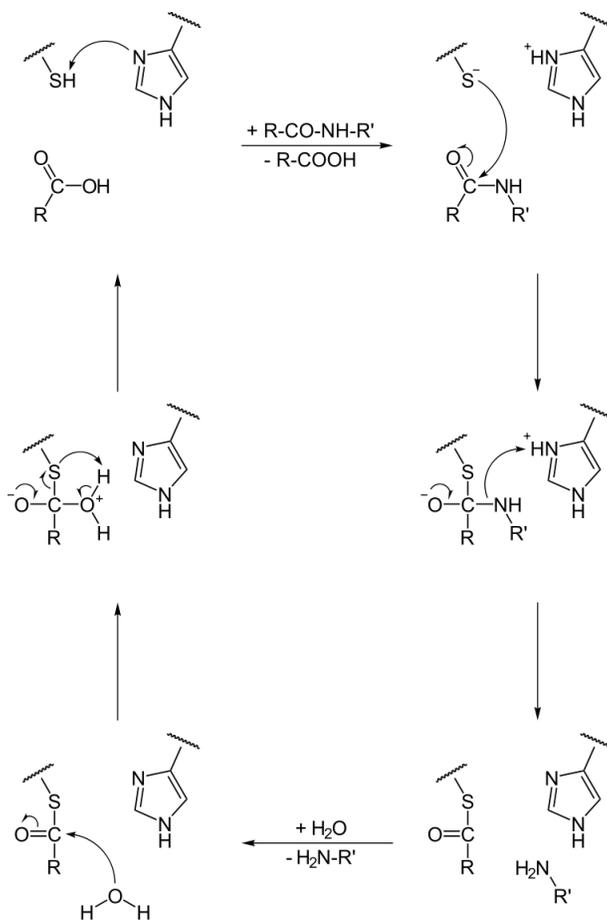


Рис. 1: Схема реакции гидролиза пептидной цепи, катализируемой цистеиновыми протеазами. Схема взята из [википедии](#).

Считается, что специфичность катепсина К определяется аминокислотой субстрата в позиции P2 (по номенклатуре [Schechter and Berger \(1967\)](#), см. рис. 2). Фермент предпочитает, чтобы в позиции P2 находился пролин / гидроксипролин (остатки характерные для тройной спирали коллагена), либо крупный гидрофобный остаток (в первую очередь, Leu, Ile) [[Bossard et al. \(1996\)](#)]. В человеческом белке в сайтах S_2 и S'_2 , отвечающих за специфичность фермента, находятся аминокислоты Tyr-67 и Leu-209. Мутации в этих позициях приводят к потере коллаген-гидролазной активности [[Novinec and Lenarčič \(2013\)](#)]. Литературные данные характерных для каждой позиции P_i субстрата аминокислотах также неплохо подтверждаются базой данных [MEROPS](#).

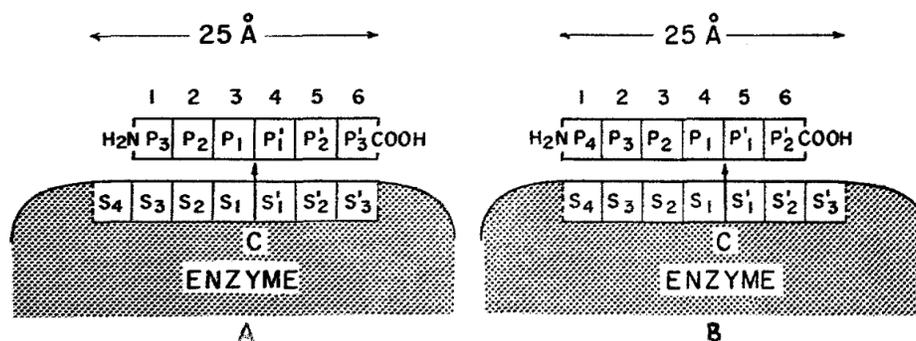


Рис. 2: Схема активного центра цистеиновых протеаз и номенклатур сайтов фермента и субстрата (на примере папаина), предложенная [Schechter and Berger \(1967\)](#).

2 Предсказание структуры с субстратом с помощью AlphaFold2

Мы взяли последовательности катепсина К (из [Uniprot](#)) и его субстрата (согласно базе данных MEROPS): AKLAEQAE. Далее с помощью ColabFold мы предсказали структуру фермент-субстрат. В качестве референса из PDB мы использовали структуру [5tdi](#) катепсина в комплексе с ингибитором, полученную с разрешением 1.4 Å (см. рис 3). Тут важно отметить, что зрелая форма белка значительно отличается от ее предшественника. Так, при созревании первые 114 N-концевых остатков отрываются от белка. В [Uniprot](#) дана последовательность незрелого белка. При изучении структуры катепсина К в комплексе с ингибитором, как правило, используют зрелую форму белка. В частности, в структуре [5tdi](#) приведена уже процессированная, укороченная форма белка. Как показывает построенное **выравнивание**, последовательность [5tdi](#) идентична центральному (по последовательности) участку незрелого белка.

Сперва мы решили дать на вход AlphaFold2 незрелую форму белка вместе с субстратом. Интересно, что AlphaFold2 оказался неуверенным в своих предсказаниях субстрата и N-концевого участка белка (а также в предсказании одной из петель на периферии белка) (см. рис. 3).

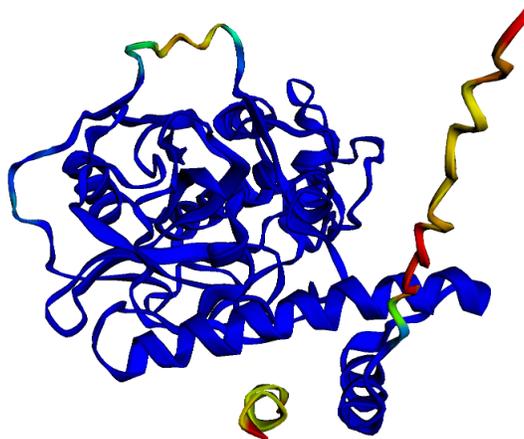


Рис. 3: Структура незрелой формы белка, предсказанная AlphaFold2. Синим показаны остатки, в предсказании которых AlphaFold2 уверен.

Мы сравнили предсказания AlphaFold2 (структуру с рангом 1) со структурой 5tdi (см. рис. 4, 5). Оказалось, что AlphaFold почти что идеально предсказал ту часть структуры, которая соответствовала укороченной, зрелой форме белка. Между тем, расположение субстрата оказалось совершенно неправильным. AlphaFold поместил его в соседстве с α -спиралью из N-концевого участка белка, в далеке от активного центра.

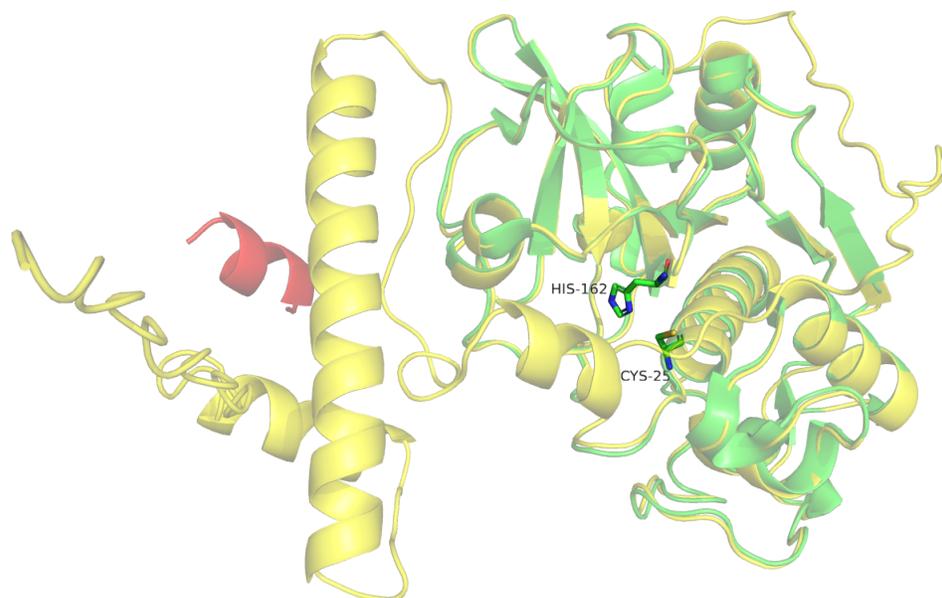


Рис. 4: Выравнивание структуры, предсказанной для последовательности незрелого белка с помощью AlphaFold2 (желтая), и PDB-структуры 5tdi (зеленая). Красны показано предсказанное AlphaFold2 положение субстрата. Показаны остатки каталитической диады His-162 и Cys-25.

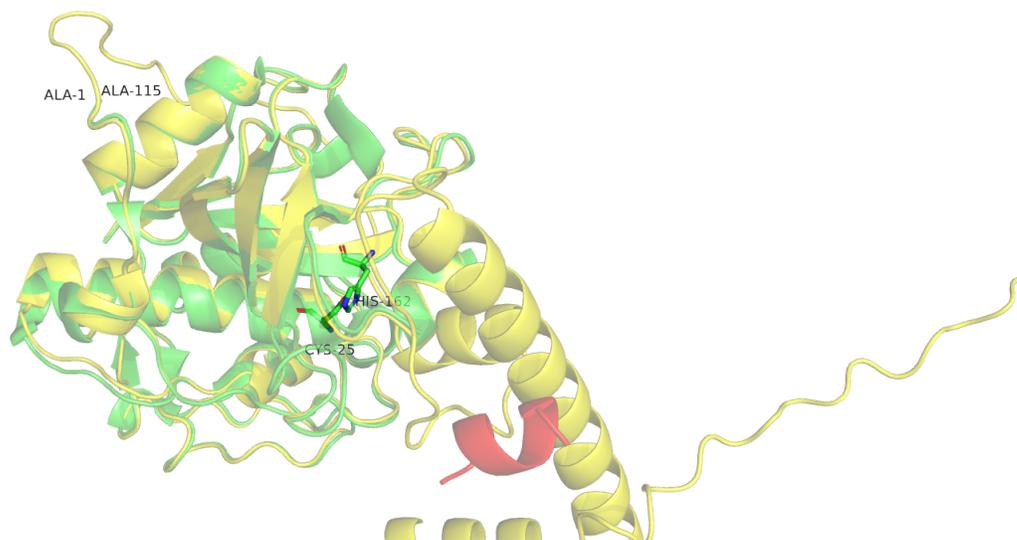


Рис. 5: Выравнивание структуры, предсказанной для последовательности незрелого белка с помощью AlphaFold2 (желтая), и PDB-структуры 5tdi (зеленая). Красны показано предсказанное AlphaFold2 положение субстрата. Подписан N-концевой для 5tdi-белка остаток аланина, за которым следует N-конец незрелой формы из 114 остатков.

Хотя оказывается, что AlphaFold совершенно неверно предсказывает положение субстрата, надо признать, что мы в данном случае даем ему неадекватную задачу. Действительно, реакцию катализирует зрелая форма фермента. Чтобы проверить, сможет ли AlphaFold верно предсказать положение субстрата для зрелой формы фермента, мы дали ему на вход последовательность белка из структуры 5tdi. На этот раз AlphaFold оказался совершенно уверен в своих предсказаниях (см. рис. 6).

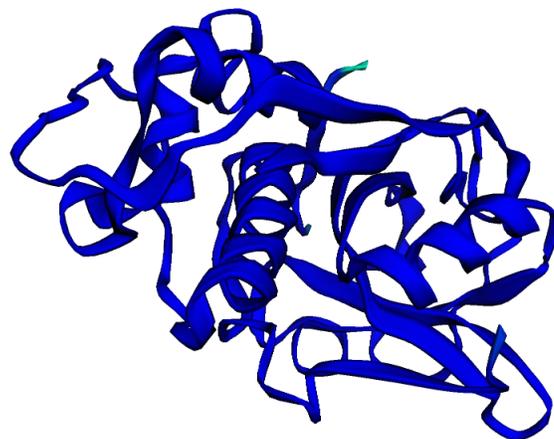


Рис. 6: Структура зрелой формы белка (последовательность взята из 5tdi), предсказанная AlphaFold2. Синим показаны остатки, в предсказании которых AlphaFold2 уверен.

При структурном выравнивании оказалось, что предсказанная AlphaFold2 модель абсолютно идентична 5tdi. Более того, субстрат AlphaFold поместил непосредственно в V-образный активный центр фермента (см. рис. 7)

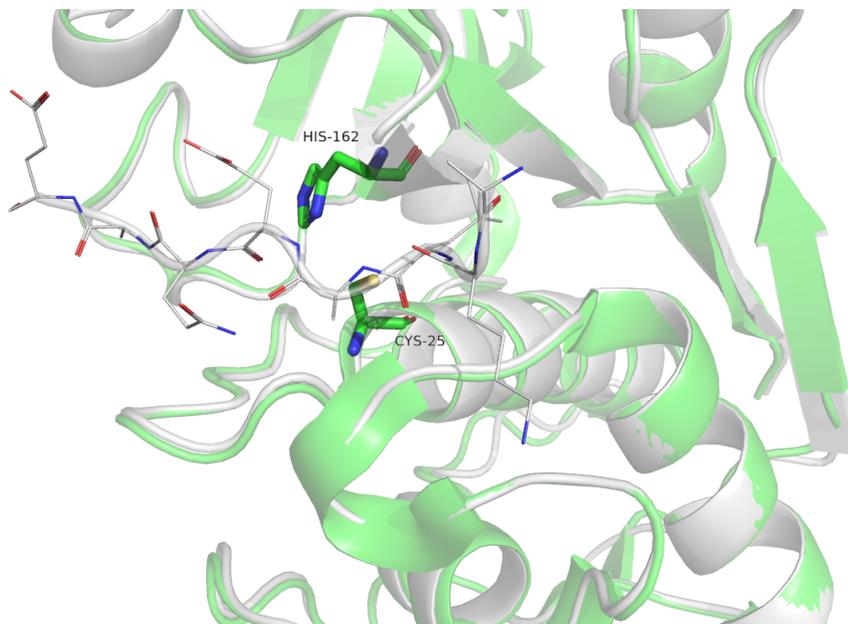


Рис. 7: Структура зрелой формы белка (последовательность взята из 5tdi), предсказанная AlphaFold2 (серая), выравненная с самой структурой 5tdi (зеленая). Показаны остатки активного центра фермента и атомы аминокислот субстрата.

Мы рассмотрели внимательнее предсказанное положение субстрата (см. рис. 8). P2-остаток субстрата (Leu-3) направлен в сторону остатков Leu-209 и Tyr-67, которые, согласно литературным данным ([Novinec and Lenarčič \(2013\)](#)) обосновывают специфичность фермента по отношению к аминокислоте P2 субстрата. В этой позиции должна находиться гидрофобная аминокислота, согласно MEROPS оптимальным остатком является лейцин. В данном случае мы действительно видим, что AlphaFold2 ставит в P2-положение остаток лейцина – Leu-3 субстрата. Хотелось еще отметить, что субстрат идеально фитится в V-образную выемку, образуемую активным центром.

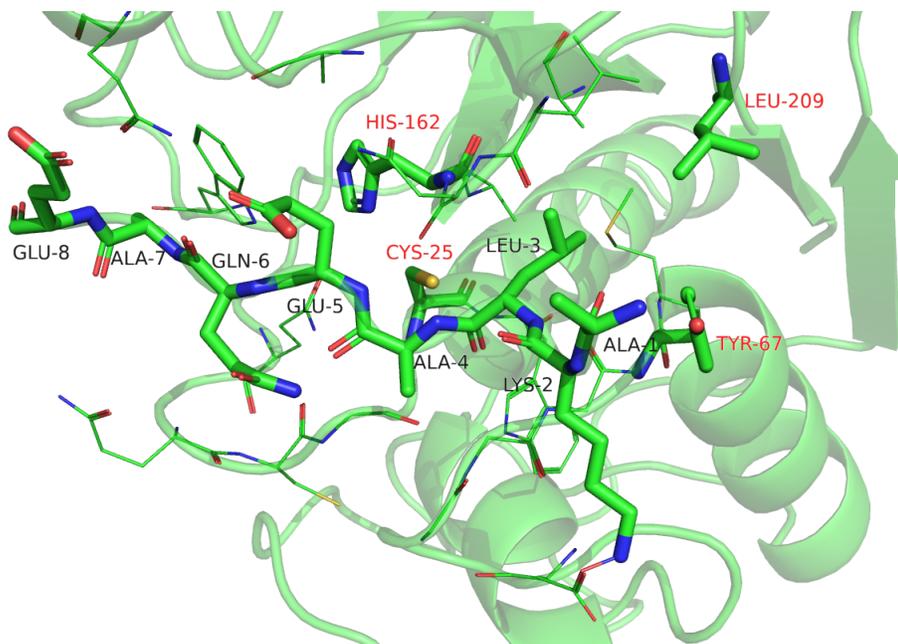


Рис. 8: Активный центр и положение субстрата, предсказанные AlphaFold2. Красным подписаны аминокислоты фермента, черным – субстрата. Остатки His-162, Cys-25 входят в каталитическую диаду. Остатки Leu-209, Tyr-67 обосновывают специфичность фермента по отношению к позиции P2 субстрата.

Нам кажется, что в данном случае, при корректном входе (т.е. правильной, зрелой последовательности белка), AlphaFold2 выдает вполне разумные предсказания. Сложно делать выводы о вообще всех белках на основании одного примера. Но, видимо, для цистеиновых протеаз можно рассчитывать на неплохие предсказания AlphaFold2. В первом приближении это можно проверить по расстоянию между субстратом и активным центром (так, в нашем первом, неудачном запуске мы видели, что субстрат оказывается в совершенно н)

3 Сессии rulumol и предсказания AlphaFold2

https://kodom0.fbb.msu.ru/~osyafinkelberg/term7/prac10/FBB_bioinformatics_Zlobin_prac10.pdf

<https://kodom0.fbb.msu.ru/~osyafinkelberg/term7/prac10/session0.pse>

предсказания для незрелой формы белка

предсказания для зрелой формы белка

Список литературы

- Bossard, M. J., Tomaszek, T. A., Thompson, S. K., Amegadzie, B. Y., Hanning, C. R., Jones, C., Kurdyla, J. T., McNulty, D. E., Drake, F. H., Gowen, M., and Levy, M. A. (1996). Proteolytic Activity of Human Osteoclast Cathepsin K. *Journal of Biological Chemistry*, 271(21):12517–12524.
- Dai, R., Wu, Z., Chu, H. Y., Lu, J., Lyu, A., Liu, J., and Zhang, G. (2020). Cathepsin K: The Action in and Beyond Bone. *Frontiers in Cell and Developmental Biology*, 8:433.
- Jin Hee Park (2017). Current Understanding of RANK Signaling in Osteoclast Differentiation and Maturation. *Molecules and Cells*.
- Novinec, M. and Lenarčič, B. (2013). Cathepsin K: a unique collagenolytic cysteine peptidase. *Biological Chemistry*, 394(9):1163–1179.
- Schechter, I. and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, 27(2):157–162.