

**ОЦЕНКА ВЛИЯНИЯ СТАБИЛИЗИРУЮЩИХ МУТАЦИЙ В ТЕТРАМЕРЕ TTR  
ПРИ ОБУЧЕНИИ МОДЕЛИ SDAM-TTR**

Степанов Я. А. (ИТМО)

Научный руководитель – Пьянков И. А. (ИТМО)

**Введение.** В настоящее время существует необходимость определения ассоциации мутаций с амилоидозом *in silico* при использовании данных геномного секвенирования. Ранее была разработана модель машинного обучения SDAM-TTR, которая позволяет определить эффект мутации в структурной области белка TTR, связанного с кардиомиопатии ATTR (транстиретиновый амилоидоз), на основе данных о влиянии мутаций на стабильность структуры тетрамера, амилоидогенность и физико-химические свойства белка [1]. Один из способов для повышения точности работы модели - добавление в обучение мутаций, которые создают новые связи, что может привести к стабилизации белка. Целью данной работы является оценка влияния стабилизирующих мутаций при обучении модели машинного обучения для прогнозирования патогенности мутаций SDAM-TTR за счет включения таких вариантов в обучающий набор данных.

**Основная часть.** В настоящей работе рассматривается влияние стабилизирующих мутаций на качество работы модели машинного обучения при добавлении их в обучающий набор. При отборе стабилизирующих вариантов рассматривались позиции, в которых аминокислоты положительно или отрицательно заряжены и находятся на поверхности в структуре тетрамера. В этих позициях проводилась замена соседних аминокислот на противоположно заряженные, что способствовало формированию новых ионных связей. В результате был сформирован набор из 20 мутаций, которые способны образовывать ионные связи, за счет которых, предположительно, стабилизируется структура белка. Для полученного набора вариантов были получены данные о их влиянии на стабильность структуры тетрамера, амилоидогенность и физико-химические свойства белка. Для определения данных параметров использовались следующие инструменты: TAPASS и ArchCandy, а также значения  $\Delta\Delta G$ , рассчитанные с помощью FoldX и mCSM [2,3,4].

Расчеты  $\Delta\Delta G$  позволяют оценить влияние полученных мутаций на стабильность белка, результаты подтверждают их стабилизирующий эффект. Данные по стабилизирующим вариантам были добавлены в обучающий набор, который использовался при обучении SDAM-TTR. При создании модели использовался Extra Trees Classifier библиотеки Scikit-learn на Python [5]. На основе расширенного набора данных была получена улучшенная версия модели. Для оценки ее производительности использовались метрики: F1, Accuracy, AUROC и MCC, их средние значения были получены из 75 различных групп обучения. Результаты показали, что добавление стабилизирующих мутаций повышает точность прогнозирования модели: F1 увеличился с 0.8227 до 0.8373, Accuracy повысился с 0.8277 до 0.8455, а AUROC вырос с 0.8728 до 0.8814.

**Выводы.** В ходе работы была получена улучшенная версия модели SDAM-TTR. При сравнении моделей видно, что добавление мутаций, которые могут образовывать новые ионные связи, в процесс обучения позволяет повысить точность модели, что подтверждается улучшением метрик F1, Accuracy и AUROC. Таким образом, предложенный подход демонстрирует перспективность использования вариантов, стабилизирующих структуру белка, для повышения точности прогнозирования эффекта мутаций, связанных с ATTR.

**Список использованных источников:**

1. Pyankov I. A. et al. A computational approach to predict the effects of missense mutations on protein amyloidogenicity: A case study in hereditary transthyretin cardiomyopathy // Journal of

Structural Biology. – 2025. – Vol. 217. – № 1. – P. 108176. DOI: doi.org/10.1016/j.jsb.2025.108176.

2. Ahmed A. B. et al. A structure-based approach to predict predisposition to amyloidosis // Alzheimer's & Dementia. – 2015. – Vol. 11. – № 7. – P. 681–690. DOI: doi.org/10.1016/j.jalz.2014.06.007.

3. Schymkowitz J. et al. The FoldX web server: an online force field // Nucleic Acids Research. – 2005. – Vol. 33. – Web Server issue. – P. W382–W388. DOI: doi.org/10.1093/nar/gki387.

4. Pires D. E. V., Ascher D. B., Blundell T. L. mCSM: predicting the effects of mutations in proteins using graph-based signatures // Bioinformatics. – 2014. – Vol. 30. – № 3. – C. 335–342. DOI: doi.org/10.1093/bioinformatics/btt691.

5. Pedregosa F. et al. Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research. – 2011. – Vol. 12. – P. 2825–2830. URL: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>.