

**Новосибирский государственный университет
Механико-математический факультет
Кафедра программирования**

**Аннотация
выпускной квалификационной работы магистра**

Исследование возможностей моделей глубокого обучения различных архитектур для анализа эпистатических взаимодействий замен в белках

Студент: Коломникова Дарья Юрьевна

Научный руководитель: Штокало Дмитрий Николаевич, Бульонков Михаил Алексеевич

В работе исследуются возможности трансформерных белковых языковых моделей для анализа внутригенного эпистаза – неаддитивных взаимодействий между аминокислотными заменами в белках. Актуальность обусловлена тем, что эффект множественных мутаций определяется сложными нелинейными зависимостями между позициями, что существенно усложняет прогнозирование функциональных свойств белка – ключевую задачу современной биотехнологии, белковой инженерии и разработки лекарственных препаратов.

Цель исследования – оценить применимость современных белковых языковых моделей для предсказания эффектов множественных мутаций и проанализировать, насколько механизм внимания способен выявлять позиции, склонные к эпистатическим взаимодействиям. Основное внимание уделено интерпретации специализации голов внимания в трансформерных архитектурах (ESM, SaProt и аналогичных), обученных на последовательностях белков.

Показано, что головы внимания могут демонстрировать специализацию на выявлении пар позиций, где мутации приводят к значимому эпистазу. Корреляционный анализ карт внимания подтверждает, что трансформерные модели улавливают биологически релевантные дальние зависимости, которые традиционные методы пропускают.

Методы глубокого обучения доказали высокий потенциал для интерпретации и предсказания эпистатических эффектов. Практическая значимость работы заключается в возможности применения полученных результатов при выборе архитектур и стратегий интерпретации моделей для задач направленной эволюции белков, прогнозирования множественных мутаций и ускорения белковой инженерии.