

ОТЗЫВ

официального оппонента

Головина Андрея Викторовича

на диссертацию Матвеева Евгения Викторовича

“Предсказание сайтов протеолиза на основе информации о пространственной структуре потенциальных субстратов”

представленную на соискание ученой степени кандидата биологических наук по научной специальности 1.5.8 – Математическая биология, биоинформатика

Актуальность темы.

Актуальность исследования протеолиза обусловлена его ключевой ролью в регуляции активности множества молекул и участии в важнейших биологических процессах внутри клетки и за её пределами. Протеолиз регулирует репликацию и транскрипцию ДНК, клеточный цикл, дифференциацию, ремоделирование тканей, ангиогенез, гомеостаз, некроз и апоптоз. Нарушения активности протеаз связаны с развитием многих заболеваний, таких как рак, артрит, сердечно-сосудистые и нейродегенеративные болезни. Поэтому исследование идентификации субстратов протеолитических ферментов и определение сайтов расщепления остаётся актуальной задачей современной протеомики, важной для биомедицинских исследований и разработки новых терапевтических подходов. Экспериментальная идентификация субстратов протеаз сложна, так как протеазы работают в сложных сетях взаимодействий и бывает трудно выделить продукты протеолиза и однозначно определить сайт расщепления. В связи с этим существует высокая потребность в биоинформатических методах прогнозирования субстратов и сайтов протеолиза, которые могут быть комплементарны лабораторным методам, снижая затраты времени и ресурсов. Также протеазы играют важную роль в промышленности, медицине и фармацевтике, что подчёркивает практическую значимость изучения протеолиза. Разработка универсальных методов прогнозирования протеолитических событий, основанных на структурных характеристиках субстратов, актуальна и перспективна, поскольку существующие инструменты часто ограничены специфичностью одной протеазы и не охватывают универсальные закономерности протеолиза. В частности, исследование направлено на создание модели машинного обучения, оценивающей восприимчивость участков белков к протеолизу на основе трёхмерной структуры, и объединение её с протеазо-специфическими моделями, что расширяет возможности прогнозирования и позволяет лучше понять молекулярные механизмы действия протеаз и их субстратов.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации. Диссертационная работа обладает высокой

степенью обоснованности, которая обеспечивается современными вычислительными подходами, использованными в работе, проведении необходимого количества сравнительных анализов и сравнением с литературными данными.

Научная новизна работы определяется тем, что: Впервые разработана универсальная предсказательная модель, способная оценивать восприимчивость различных участков белков к протеолизу на основе их трёхмерной структуры, независимо от специфичности конкретной протеазы. Это отличает модель от большинства существующих инструментов, которые ориентированы на конкретные протеазы или их семейства. Впервые созданы модели субстратной специфичности по аминокислотной последовательности для 169 протеаз человека, что значительно расширяет возможности изучения протеаз и их субстратов. Разработан подход к интеграции универсальной структурной модели с протеазо-специфическими моделями, показывающий качество прогнозирования, сопоставимое с существующими методами, при этом охватывающий более широкий спектр протеаз и обладающий гибкостью и расширяемостью за счёт возможности добавления новых моделей. Предсказана ранее неизвестная позиция сайта расщепления S-белка коронавируса SARS-CoV-2 цистеиновой протеазой катепсином L, а также оценено влияние мутаций в S-белке вируса на эффективность протеолитического расщепления протеазами человека.

Научная и практическая значимость диссертационной работы.

Научная значимость работы заключается в разработке универсальной модели предсказания восприимчивости участков белков к протеолизу на основе их трёхмерной структуры, что расширяет современные представления об участии протеолитических ферментов в биологических процессах. Впервые созданы модели субстратной специфичности для 169 протеаз человека, а также предложен интегративный подход, соединяющий структурную модель и модели специфичности по аминокислотной последовательности, что позволяет прогнозировать протеолитические события более точно и универсально, чем существующие методы. Это способствует выявлению ранее неизвестных элементов протеолитических сетей и пониманию молекулярных механизмов протеолиза.

Практическая значимость исследования выражается в возможности использования полученных данных о потенциальных субстратах и продуктах протеолитического расщепления в качестве лекарственных мишеней, биомаркеров и биосенсоров, а также инструментов для биотехнологической и пищевой промышленности. Информация о субстратном профиле протеаз способствует разработке новых ингибиторов и созданию искусственных протеолитических систем с заданными свойствами. Предложенный биоинформатический подход может существенно облегчить и

повысить направленность экспериментальных исследований, сокращая затраты времени и ресурсов, что ценно для биомедицинских и промышленных применений.

Степень достоверности результатов проведенных исследований.

Высокая степень достоверности обеспечивается выбором современных методов теоретического и вычислительного исследования, что привело к 3 публикациям в рецензируемых журналах во всех из которых Матвеев Е.В. является первым автором.

Публикации основных результатов диссертационной работы. По теме диссертации опубликовано 10 работ, отражающих основные положения исследования, среди которых – 3 работы в рецензируемых журналах, индексируемых в базе Scopus и WoS, 7 – работ в сборниках конференций.

Структура диссертационной работы. Работа состоит из стандартных разделов: введение: где обосновывается выбор темы исследования, отражается её актуальность, научная новизна, практическая и теоретическая значимость, формулируются цель и задачи исследования, методология и методы, даётся характеристика материала, приводятся положения, выносимые на защиту, степень достоверности полученных результатов, апробация и публикации, личный вклад автора, а также структура и объем работы. Основная часть: состоит из нескольких глав, включающих обзор литературы о разнообразии протеаз с фокусом на их роль в жизненном цикле SARS-CoV-2 и анализом известных вычислительных подходов к предсказанию сайтов гидролиз. Глава “Разработка метода прогнозирования протеолитических событий” содержит информацию о построении модели по предсказанию уязвимости участков белка к протеолизу и интеграции этой модели с известными моделями на основе последовательностей. Третья глава содержит анализ активации S-белка коронавируса. Заключение содержит итоги исследования и делаются выводы. Список используемой литературы литературы содержит 142 источника. Объем диссертации — 121 страница, включая 32 иллюстрации, 2 формулы и 142 источника цитирования.

Вопросы по диссертационной работе/ Замечания

1) Ключевое замечание к работе состоит в том, что автор создает модель машинного обучения с использованием структурной информации и пытается, как принято в этой области, на основе статистики оценить качество модели. К сожалению такой подход не несет читателю новой биологической информации. Автор не обсуждает вклад структурного контекста, который использует при обучении, на эффективность гидролиза. Формально такие инструменты есть. Интересно было бы узнать бывает ли расщепление в альфа спиралях и т.д.

2) Автор использует “Температурный фактор”, как я понял для оценки подвижности участка белка на основе PCA данных, но не обсуждает его релевантность в водных

растворах где происходит гидролиз. Также автор не учитывает особенности кристаллической упаковки белков, а именно соседей в ячейках и их влияние на доступность растворителю.

3) В работе не ясно как использовалось значение доступности остатка растворителю. В тексте написано, что это значение бралось из результата работы программы DSSP, моя быстрая оценка этого значения говорит, что это ненормализованное значение для всего остатка. Очевидно, что такое значение будет сильно зависеть от природы аминокислоты и в меньшей степени от структурного контекста. Важно другое, для успешного гидролиза необходима доступность остова белка, а именно пептидной связи. Вероятно, для крупных остатков на поверхности это значение будет высоким, но реальная доступность пептидной связи будет низкой, и наоборот если боковая цепь остатка смотрит внутрь и остов доступен растворителю, это значение будет низким. Это как мне кажется существенное упущение в работе.

4) Меня удивило, что расширяя выборку с помощью AlphaFoldDB автор решил избавиться от температурного фактора и от pLDDT сразу. Вполне можно было использовать структуры построенные AlphaFold вместо данных PCA. Очень вероятно, что отличие предсказаний от данных PCA будет минимальным. Вероятно исключение одного из описаний привело к улучшению ROC AUC метрики.

5) Любопытно отметить, что при сравнении интегративной модели автор использует новый датасет и в этом эксперименте качество предсказания резко улучшается с посредственных 0.7 до идеальных 0.97. В чем причина такого “рывка” автор не объясняет.

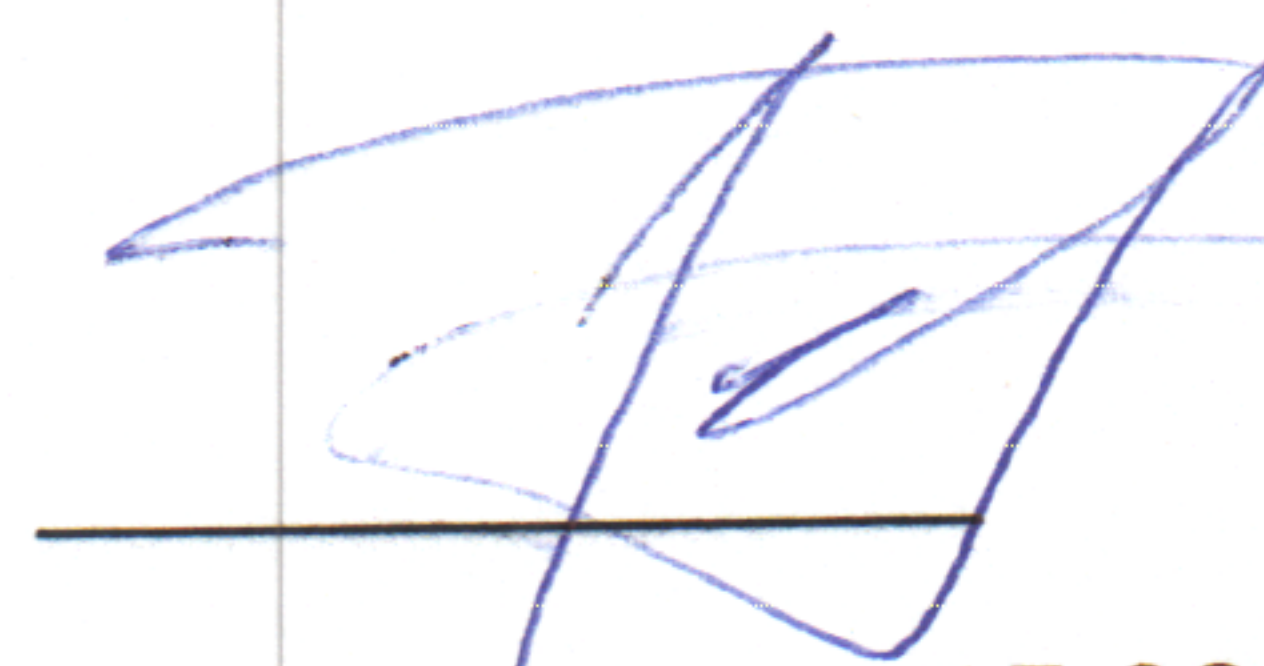
Заключение

Диссертационная работа Матвеева Евгения Викторовича «Структурное аннотирование переменных доменов антител» является самостоятельной, законченной научно-квалификационной работой, выполненной автором на высоком научном уровне. Диссертация соответствует паспорту научной специальности 1.5.8. математическая биология, биоинформатика. Считаю, что диссертационная работа «Предсказание сайтов протеолиза на основе информации о пространственной структуре потенциальных субстратов» по своей новизне, актуальности и значимости результатов полностью соответствует всем требованиям «Положения о присуждении ученых степеней», утвержденного постановлением Правительства РФ № 842 от 24 сентября 2013 года (с актуальными изменениями), в том числе п.9, а ее автор, Матвеев Евгений Викторович, заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8. математическая биология, биоинформатика

Доктор химических наук по специальности 02.00.10 –
«биоорганическая химия»,
зам. декана Факультета биоинженерии и
биоинформатики Московского государственного
университета имени М.В.Ломоносова

Телефон: +7 9647109158

e-mail: golovin@fbb.msu.ru



Головин

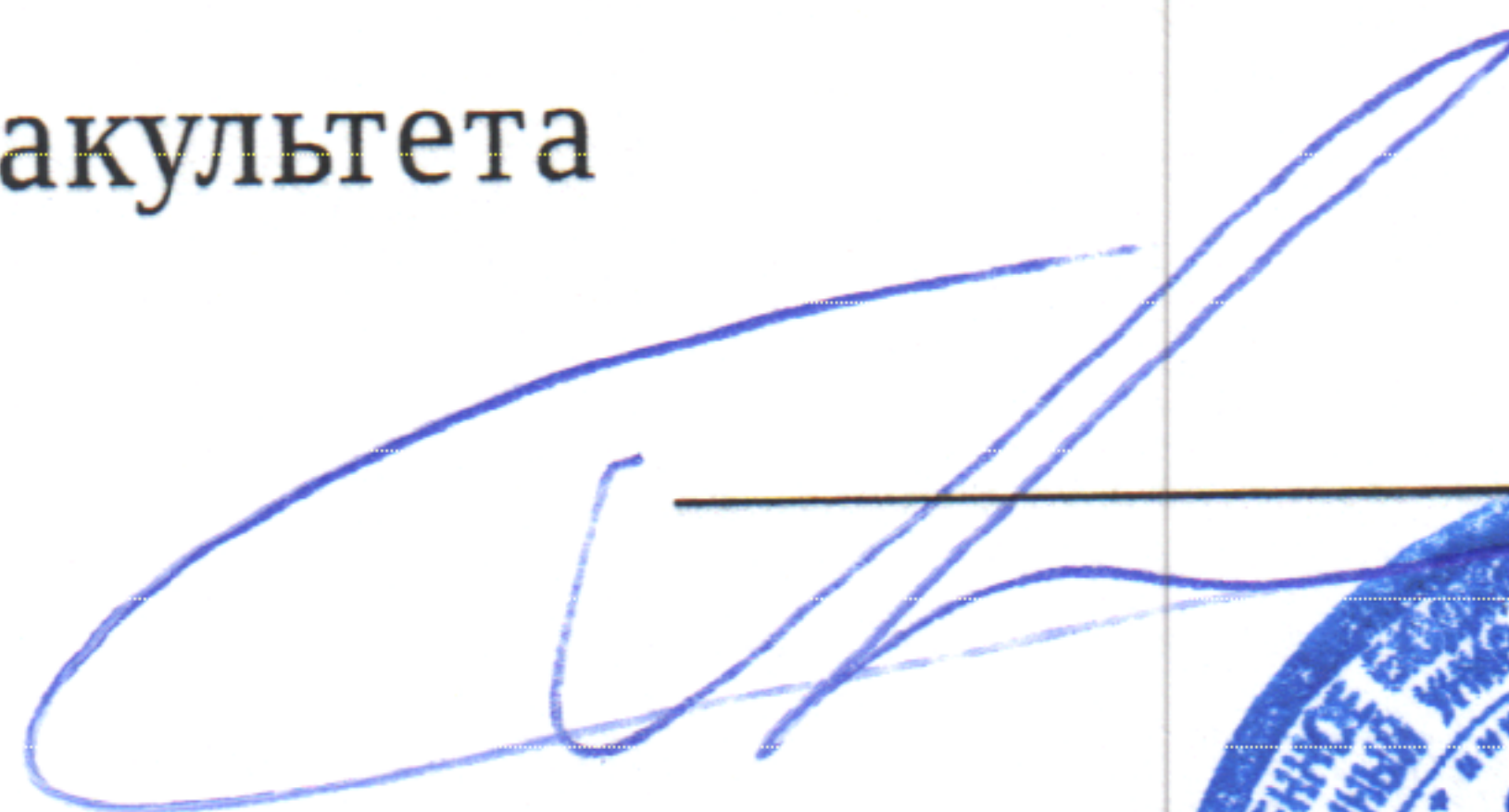
Дата: 15.09.2024 г.

Сведения об организации: Факультет биоинженерии и биоинформатики,
Федеральное государственное бюджетное образовательное учреждение высшего
образования «Московский государственный университет имени М.В.Ломоносова

Адрес организации: 119234 г. Москва, ГСП-1, Ленинские горы МГУ 1, стр. 73,
Факультет биоинженерии и биоинформатики

Подпись Головина А.В. заверяю

Исполняющий обязанности Декана Факультета
биоинженерии и биоинформатики



А.А. Замятнин

