

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.1.101.01 НА
БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО
УЧРЕЖДЕНИЯ НАУКИ ИНСТИТУТА ПРОБЛЕМ ПЕРЕДАЧИ
ИНФОРМАЦИИ ИМ. А.А. ХАРКЕВИЧА РОССИЙСКОЙ АКАДЕМИИ
НАУК ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
КАНДИДАТА БИОЛОГИЧЕСКИХ НАУК

Аттестационное дело № _____
решение диссертационного совета
от 29 сентября 2025 г., протокол № 9

О присуждении **Матвееву Евгению Викторовичу**,
гражданину Российской Федерации,
ученой степени кандидата биологических наук.

Диссертация «Предсказание сайтов протеолиза на основе информации о пространственной структуре потенциальных субстратов» по специальности 1.5.8. — Математическая биология, биоинформатика, принята к защите 24 июля 2025 года, Протокол № 5, диссертационным советом 24.1.101.01 на базе Федерального бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (127051, г. Москва, Большой Каретный переулок, д.19 стр. 1, приказ о создании №978/нк от 16 декабря 2013 года, номер изменен приказом № 561/нк от 03 июня 2021 года).

Соискатель Матвеев Евгений Викторович, гражданин Российской Федерации, 1998 года рождения, в 2020 году окончил Волгоградский государственный университет по направлению специальности «Биоинженерия и биоинформатика». В 2024 году окончил аспирантуру ИППИ имени А.А. Харкевича РАН по специальности 1.5.8 «Математическая биология, биоинформатика» (направление подготовки 06.06.01 «Биологические науки»).

С 2020 года по настоящее время соискатель работает в Национальном медицинском исследовательском центре детской гематологии, онкологии и иммунологии, с 2024 года в должности младшего научного сотрудника лаборатории цитогенетики и молекулярной генетики.

Диссертация выполнена в Институте проблем передачи информации им А.А. Харкевича Российской академии наук.

Научный руководитель – Казанов Марат Джамалудинович, кандидат технических наук, старший научный сотрудник Центра молекулярной и клеточной биологии Автономной некоммерческой образовательной организации высшего образования «Сколковского института науки и технологий».

Выбор официальных оппонентов и ведущей организации обусловлен их высокой компетентностью в области исследований, посвящённых протеолизу, а также в области математических подходов к анализу биологических данных и методов их интеграции, а также тем, что за последние годы ими было опубликовано большое количество научных работ, посвящённых смежным тематикам исследования.

Официальные оппоненты:

Головин Андрей Викторович, доктор химических наук, профессор и зам. декана факультета биоинженерии и биоинформатики Московского Государственного Университета им. Ломоносова.

Попцова Мария Сергеевна, кандидат физико-математических наук, заведующий международной лабораторией биоинформатики Института искусственного интеллекта и цифровых наук Факультета компьютерных наук Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики».

дали положительные отзывы на диссертацию.

Ведущая организация - Федеральное государственное автономное образовательное учреждение высшего образования «Балтийский федеральный университет имени Иммануила Канта» - дала **положительное заключение**, подписанное кандидатом биологических наук Попадьным Константином Юрьевичем, старшим научным сотрудником Центра геномных исследований БФУ им. Канта, и утвержденное 12 сентября 2025 г. В заключении **подчёркиваются достоинства работы**, к которым можно отнести междисциплинарность работы и сочетание современных методов машинного обучения, структурной биоинформатики и молекулярной биологии, практическую значимость для медицины и биотехнологий, открытые и воспроизводимые результаты (выкладка моделей и кода в открытый доступ), а также **приведены замечания**, касающиеся необходимости дальнейшей интеграции информации о посттрансляционных модификациях и динамике структур белков и дополнительной экспериментальной валидации полученных результатов. При этом в отзыве отмечается, что «высказанные замечания не снижают высокой научной и прикладной значимости проведённого исследования».

В заключении отзыва указано, что «диссертационная работа Матвеева Евгения Викторовича «Предсказание сайтов протеолиза на основе информации о пространственной структуре потенциальных субстратов» представляет собой самостоятельное, завершённое и оригинальное научное исследование, в котором содержится решение научной задачи, имеющей важное значение для современной биоинформатики и молекулярной биологии. Автор заслуживает присуждения учёной степени кандидата биологических наук по специальности 1.5.8. Математическая биология, биоинформатика».

В отзыве оппонента Головина Андрея Викторовича, доктора химических наук, профессора и зам. декана Факультета биоинженерии и биоинформатики МГУ им. Ломоносова, выдвинут ряд дискуссионных вопросов и замечаний:

- 1) Ключевое замечание к работе состоит в том, что автор создаёт 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 несёт 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 предсказаний резко улучшается с посредственных 0.7 до идеальных 0.97. В чём причина такого «рывка» автор не объясняет.

В заключении отзыва отмечено, что «Диссертационная работа Матвеева Евгения Викторовича «Предсказание сайтов протеолиза на основе информации о пространственной структуре потенциальных субстратов» 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 её автор, Матвеев Евгений Викторович, заслуживает присуждения учёной степени кандидата биологических наук по специальности 1.5.8 «Математическая биология, биоинформатика».

В отзыве оппонента Попцовой Марии Сергеевны, кандидата физико-математических наук, заведующей международной лабораторией биоинформатики Института искусственного интеллекта и цифровых наук Факультета компьютерных наук Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики», имеется ряд замечаний, пожеланий и дискуссионных вопросов:

- 1) При сравнении качества моделей машинного обучения, используются метрики, полученные моделями, которые были использованы с дефолтными гиперпараметрами. Вероятно, LDA не остался бы лучшим алгоритмом при тонкой настройке гиперпараметров. Известно, что LDA работает лучше всего на нормально распределённых данных.
- 2) Пробовал ли автор строить модели на большем числе параметров, чем на пяти отобранных? Преимущество алгоритмов машинного обучения как раз и состоит в том, что они могут интегрировать информацию от не сильно коррелирующих признаков. Могло ли бы это улучшить качество предсказаний модели?

- 3) Для сравнения с Procleave был использован очень небольшой тестовый набор (43 события, 3 протеазы). Достаточно ли этого для статистически надёжного вывода о сопоставимости качества? Планировали ли вы валидацию на более масштабных наборах данных?
- 4) Как именно технически вы решали проблему сильного дисбаланса классов (445 позитивных против 68840 негативных примеров) при обучении моделей, кроме подбора соотношений 1:1? Использовались ли методы аугментации данных?
- 5) В диссертации обсуждается, что посттрансляционные модификации (например, гликозилирование) сильно влияют на доступность сайта. Какие Вы видите способы интеграции такой информации в будущем?
- 6) Вопрос дискуссионный: возможно ли построить универсальную модель определения сайтов расщепления с помощью представлений больших языковых белковых моделей (например, ESM-2, ProBERT)? Были ли опубликованы подобные работы?
- 7) Предсказание сайта для катепсина L в позиции K790 является ключевым результатом. Планируется ли его экспериментальная проверка (например, с помощью масс-спектрометрии) для подтверждения?

В отзыве сказано, что «несмотря на имеющиеся замечания и вопросы, считаю, что диссертационная работа Матвеева Евгения Викторовича «Предсказание сайтов протеолиза на основе информации о пространственной структуре потенциальных субстратов» является законченной научно-квалификационной работой и отвечает требованиям п. 9 «Положения о порядке присуждения ученых степеней», утвержденным постановлением Правительства Российской Федерации No 842 от 24 сентября 2013 г., а её автор, Е. В. Матвеев, заслуживает присуждения искомой степени кандидата биологических наук по специальности 1.5.8 – Математическая биология, биоинформатика».

Соискатель имеет три статьи по теме диссертации в рецензируемых журналах из международных библиографических баз данных, включенных в перечень ВАК, в том числе две – в журналах первого квартиля. Кроме того, соискателем опубликовано 3 тезиса по теме диссертации в материалах всероссийских конференций.

Статьи по теме диссертации:

I. Matveev, E. V. Predicting Structural Susceptibility of Proteins to Proteolytic Processing / E. V. Matveev, V. V. Safronov, G. V. Ponomarev, M. D. Kazanov // International Journal of Molecular Sciences. — 2023. — Vol. 24 (13). — P. 10761. — doi: 10.3390/ijms241310761

II. Матвеев, Е. В. Биоинформатический метод идентификации протеаз человека, активных относительно гликопротеинов оболочки вирусов, на примере белка шипа коронавируса SARS-CoV-2 / Е. В. Матвеев, Г. В. Пономарёв, М. Д. Казанов // Молекулярная биология. — 2024. — Т. 58(1). — С. 171–177. — DOI: 10.31857/S0026898424010176

III. Matveev, E. V. Genome-wide bioinformatics analysis of human protease capacity for proteolytic cleavage of the SARS-CoV-2 spike glycoprotein / E. V. Matveev, G. V. Ponomarev, M. D. Kazanov // Microbiology Spectrum. — 2024. — Vol. 12 (2). — P. e03530-23. — doi: 10.1128/spectrum.03530-23

Результаты работы, опубликованные в сборниках тезисов международных конференций:

I. Матвеев Е.В., Сафронов В.В., Казанов М.Д. Биоинформатическая идентификация субстратов протеаз на основе известной трёхмерной структуры белков / XXVII Симпозиум «Биоинформатика и компьютерное конструирование лекарств» // онлайн, 5–7 апреля 2021 г.

II. Matveev E.V., Kazanov M.D. Computational prediction of susceptibility to limited proteolysis for proteins with known 3D structure / XXVIII Symposium on Bioinformatics and Computer-Aided Drug Discovery // Москва, Россия, 24–26 мая 2022 г.

III. Matveev E.V., Kazanov M.D. Bioinformatics screening of human proteases for ability of activating SARS-CoV-2 spike protein / Proceedings of 11th Moscow Conference on Computational Molecular Biology MCCMB'23 // Москва, Россия, 3–6 августа 2023 г.

На автореферат поступили 3 положительных отзыва.

В отзыве на автореферат Гриневича Андрея Анатольевича, доктора физико-математических наук, ведущего научного сотрудника Института биофизики клетки РАН (обособленное подразделение Федерального государственного бюджетного учреждения науки «Федеральный исследовательский центр «Пущинский научный центр биологических исследований Российской академии наук») (ИБК РАН), **отмечено, что** «Автореферат включает все необходимые главы, написан грамотным языком, хорошо проиллюстрирован, что даёт полное представление о содержании работы. Результаты получены с использованием известных вычислительных методов и алгоритмов, апробированы на конференциях и опубликованы в научных журналах, что свидетельствует об их достоверности и надёжности». В отзыве **есть одно замечание** о том, что «вывод 6 не является по сути выводом. Хотелось бы узнать, какой вывод может сделать автор из анализа влияния мутаций, выявленных в S-гликопротеине известных штаммов коронавируса SARS-CoV-2, на эффективность его протеолитического

расщепления». **Несмотря на вышеуказанное замечание**, «анализ автореферата позволяет заключить, что диссертационная работа Евгения Викторовича Матвеева полностью соответствует требованиям ВАК, предъявляемым к диссертациям на соискание учёной степени кандидата биологических наук, а её автор заслуживает присуждения ему учёной степени кандидата биологических наук по специальности 1.5.8 – «математическая биология, биоинформатика».

В отзыве на автореферат Артамоновой Ирины Игоревны, кандидата биологических наук, старшего научного сотрудника и руководителя группы биоинформатики отдела вычислительной системной биологии ИОГен РАН, **указано, что** «обсуждаемое исследование выполнено на высоком научном уровне и соответствует статусу научно-квалификационного труда, результаты исследования оригинальны и, несомненно, представляют интерес для экспериментальной биологии в области исследования протеаз. Автореферат содержит обоснованные выводы и свидетельствует, что обсуждаемая диссертационная работа полностью удовлетворяет требованиям ВАК, предъявляемым к диссертациям на соискание учёной степени кандидата биологических наук. По моему мнению, Евгений Викторович Матвеев заслуживает присуждения учёной степени кандидата биологических наук по специальности 1.5.8 – «Математическая биология, биоинформатика».

В отзыве на автореферат Зеркаленковой Елены Александровны, кандидата биологических наук, старшего научного сотрудника НМИЦ ДГОИ им. Д. Рогачёва, **сформулировано, что** «Работа выполнена на высоком уровне с соблюдением всех требований, предъявляемых к работам на соискание учёной степени кандидата биологических наук. Актуальность, новизна и потенциальная практическая значимость исследования убедительно обоснованы, задачи выполнены, поставленная цель достигнута. Разработанные модели опубликованы на платформе GitHub, три статьи по результатам работы представлены в рецензируемых научных журналах, индексируемых Web of Science и Scopus. Таким образом, Евгений Викторович Матвеев обладает необходимыми навыками и квалификацией для присуждения учёной степени кандидата биологических наук по специальности 1.5.8 – «Математическая биология, биоинформатика».

Диссертационный совет отмечает, что на основании выполненных соискателем исследований впервые:

- Разработана универсальная предсказательная модель, позволяющая оценивать подверженность различных участков белков протеолизу на основании их трёхмерной структуры. Модель применима ко всем типам протеолитических ферментов и не ограничена конкретными классами протеаз;

- Разработаны модели субстратной специфичности для 169 протеаз человека. Функционал метода может быть расширен за счёт добавления моделей специфичности для других протеаз;
- Предсказана позиция сайта расщепления S-белка коронавируса SARS-CoV-2 цистеиновой протеазой катепсин L;
- Получены оценки влияния как мутаций, обнаруженных в штаммах коронавируса SARS-CoV-2, так и теоретически возможных аминокислотных замен, на эффективность расщепления S-белка протеазами человека.

Теоретическая значимость исследования обоснована тем, что разработанный метод идентификации протеолитических субстратов и сайтов расщепления может помочь в выявлении ранее неизвестных элементов сетей протеолитических взаимодействий, что расширяет современные представления об участии протеолитических ферментов в биологических процессах в клетке.

Значение полученных соискателем результатов для практики подтверждается тем, что потенциальные субстраты и продукты протеолитического расщепления могут рассматриваться в качестве лекарственных мишеней, биомаркёров и биосенсоров, продуктов биотехнологической и пищевой отраслей промышленности, инструментов биохимического анализа. Наличие информации о субстратном профиле протеаз может способствовать разработке новых ингибиторов и созданию искусственных протеолитических систем с заданными свойствами. В целом, предложенный биоинформатический подход может служить эффективным дополнением к экспериментальным исследованиям, позволяя сузить область поиска и повысить их направленность, снижая затраты времени и ресурсов.

Свидетельства достоверности результатов исследования:

Результаты, полученные в работе, согласуются с данными, представленными в литературе. Достоверность результатов, научных положений и выводов не вызывает сомнений ввиду корректно выбранных методов анализа и публикации результатов в ведущих научных журналах.

Личный вклад соискателя состоит в поиске и подробном анализе литературы, сборе и обработке данных, построении моделей, визуализации результатов. Интерпретация и обобщение полученных результатов, а также подготовка публикаций и тезисов конференций осуществлялись соискателем совместно с научным руководителем и соавторами.

Результаты, описанные в диссертации, получены автором самостоятельно. В диссертации решены задачи, имеющие значение для развития исследований, направленных на дальнейшее изучение механизмов протеолиза. По своему содержанию диссертация отвечает паспорту специальности «1.5.8. - Математическая биология, биоинформатика».

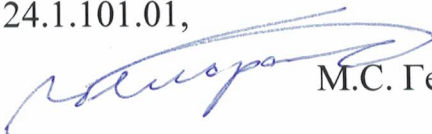
Диссертационный совет пришел к выводу о том, что диссертация представляет собой завершенное научное исследование. По актуальности, новизне, практической значимости диссертация соответствует требованиям, установленным «Положением о порядке присуждения ученых степеней» утвержденным постановлением Правительства Российской Федерации от 24 сентября 2013 года № 842 (с изменениями в редакции Постановления Правительства Российской Федерации от 21 апреля, 2 августа 2016 г., 29 мая, 28 августа 2017 г., 1 октября 2018 г., 20 марта, 11 сентября 2021 г.), предъявляемым к диссертациям на соискание ученой степени кандидата биологических наук.

На заседании 29 сентября 2025 года диссертационный совет принял решение присудить Матвееву Евгению Викторовичу ученую степень кандидата биологических наук по специальности 1.5.8 – «Математическая биология, биоинформатика».

При проведении тайного голосования из 15 человек, входящих в состав диссертационный совет, в заседании участвовало 10 человек, из них 6 докторов наук по специальности рассматриваемой диссертации. Проголосовали: за – 10, против – 0, недействительных бюллетеней – 0.

Председатель диссертационного совета 24.1.101.01,

доктор биологических наук, профессор

 М.С. Гельфанд

Ученый секретарь диссертационного совета 24.1.101.01,

доктор биологических наук

 О.В. Казенников

29 сентября 2025 г.

