

УТВЕРЖДАЮ:
И.о. ректора федерального
государственного автономного
образовательного учреждения
высшего образования
«Балтийский федеральный университет
имени Иммануила Канта»
кандидат физико-математических наук,
доцент
М.В. Демин
12 сентября 2025 г.



Отзыв ведущей организации на автореферат диссертации Матвеева Евгения Викторовича «Предсказание сайтов протеолиза на основе информации о пространственной структуре потенциальных субстратов», представленной на соискание учёной степени кандидата биологических наук по специальности 1.5.8. Математическая биология, биоинформатика.

Актуальность и новизна исследования

Диссертация Матвеева Е.В. посвящена разработке универсального подхода к предсказанию сайтов протеолиза в белках на основе трёхмерной структуры их потенциальных субстратов. Актуальность работы обусловлена ключевой ролью протеолитических процессов во множестве фундаментальных и прикладных аспектов молекулярной биологии, молекулярной медицины и биотехнологии. Разработка новых биоинформатических инструментов для идентификации сайтов расщепления является важной задачей современной протеомики, существенно дополняя лабораторные методы, которые остаются трудоёмкими и затратными.

Научная новизна исследования определяется тем, что впервые создана универсальная структурная модель для оценки восприимчивости любой пептидной связи к ограниченному протеолизу, не привязанная к конкретному типу протеазы. Кроме того, в рамках работы получены модели специфичности по последовательности (PWM) для 169 протеаз человека, а предложенный интегративный подход сочетания этих моделей позволяет проводить широкий спектр предсказаний, в том числе для ранее не описанных протеолитических событий.

Биологическая основа и научный смысл работы

Работа Матвеева Евгения Викторовича посвящена моделированию одного из фундаментальных биохимических процессов — протеолиза, а

именно ограниченного, функционально значимого расщепления белковых молекул, управляющего огромным спектром процессов в клетке — от развития и дифференцировки до иммунного ответа, гибели клетки и тонкой регуляции межклеточных сигналов. Многие белки существуют в форме неактивных предшественников и превращаются в активные только после строго определённого протеолитического расщепления, осуществляемого специализированными ферментами — протеазами. Более того, именно протеолиз часто запускает каскадные, лавинообразные реакции (например, в системе комплемента, свёртывания крови, апоптозе, регуляции гормональных сигналов), регулирует скорость обновления протеома клетки и способен как активировать, так и выключать биологические функции белка.

Отдельно следует подчеркнуть исключительную роль протеолиза в патогенезе инфекционных, в том числе вирусных процессов. Для многих вирусов — в частности, коронавирусов (SARS-CoV-2) — протеолиз клеточными протеазами является обязательным этапом "созревания" и активации поверхностных белков, необходимых для проникновения в клетку-хозяин. Расщепление S-белка фурином, TMPRSS2 или другими протеазами определяет тканевую и видовую специфичность вируса, его патогенность, восприимчивость к мутациям и реакцию на изменения во внешней среде. Аналогичные явления характерны и для других патогенных вирусов (гриппа, ВИЧ и др.), у которых активное включение протеолиза в "жизненный цикл" формирует перспективные мишени для лекарственного воздействия.

Текущая биологическая проблема в том, что определение сайтов протеолитического расщепления в новых белках исключительно лабораторными методами крайне трудоёмко. Внутри клетки один белок может быть потенциально расщеплён десятками различных протеаз, причём массивное перекрытие специфичности, влияние структуры самой молекулы и множество "шума" (например, побочные, нецелевые гидролизные события при *in vitro* анализе) делают задачу крайне сложной и требует биоинформатического сопровождения. Экспериментально достоверно **определить не только факт, но и точное положение сайта расщепления** в естественных условиях обычно возможно лишь для ограниченного числа белков и условий.

На сегодняшний день, в отличие от "глобальных" катаболических процессов (например, пищеварение), исследовательское внимание сосредоточено на изучении регуляторного, ограниченного протеолиза (*limited/regulated proteolysis*), когда из всего многообразия возможных

пептидных связей в белке разрезу подвергаются только определённые участки. Эти сайты могут быть "скрыты" на трёхмерном уровне, доступны лишь в определённых конформациях, а также зависят от окружающих аминокислот и общего ландшафта белковой поверхности — поэтому чисто линейные, по последовательности ("motif-based") методы несовершенны, особенно если речь идёт о предсказаниях наивысшего уровня: выявлении "неочевидных" сайтов и неизвестных субстратов ферментов.

Сложность задачи состоит, во-первых, в универсальности и множественности ферментов — каждый из которых имеет свою характеристику специфичности, во-вторых, в очень сильном влиянии локальных структурных особенностей субстрата (спрятан ли участок внутри, покрыт ли соседними частями, участвует ли в формировании устойчивых элементов структуры и т.д.). Это и определяет направление работы Матвеева.

Впервые в отечественной и мировой практике создана универсальная **протеазо-независимая** модель, способная предсказывать "уязвимость" пептидных связей любого белка к расщеплению исходя из трёхмерных структурных атрибутов — доступности для растворителя, типов вторичной структуры, биофизических параметров (температурный фактор, протяжённость петлевых областей и др.), не прибегая к информации о самом ферменте. Это приближает возможность глобального анализа потенциальной регуляторной "ломкости" любого белкового участка даже при отсутствии точной информации о его взаимодействиях с конкретной протеазой. Для обучения модели использовались крупные базы экспериментально подтверждённых сайтов (CutDB), а эффективность — строго валидировалась на независимых наборах (реальные сайты, выявленные *in vitro* и *in vivo*).

Сгенерировано 169 протеазо-специфичных PWM-моделей для ферментов человека (по данным MEROPS) — то есть позиционно-весовых матриц, отражающих вероятности появления определённых аминокислот в конкретных позициях вокруг расщепляемой связи для данной протеазы. Эти PWM используются для быстрого "сканирования" белковых последовательностей и оценки, насколько данный отрезок может соответствовать известным предпочтениям протеазы.

Самое ценное — интеграция результатов обеих моделей с помощью методов машинного обучения: одновременно учитываются и общая структурная уязвимость, и конкретные ферментные предпочтения по последовательности. Такой подход позволил выйти на уровень точности, сопоставимый с лучшими мировыми предсказателями (например, Procleave),

но значительно расширить применимость за счёт широты спектра ферментов и "отвязки" от единичных кейсов.

Работа не ограничивается "тестовой" биоинформатикой: продемонстрировано применение подхода для анализа протеолитической активации ключевого патогенетического белка S (шиповидный гликопротеин коронавируса SARS-CoV-2). Впервые произведён биоинформатический скрининг протеаз человека на предмет теоретической способности расщеплять вирусный белок; выявлен ряд сайтов, потенциально важных для понимания инфекции, а также оценено влияние естественных и искусственных мутаций (например, мутации P681H/R, резко повышающие чувствительность сайта S1/S2 к фурину, что коррелирует с известным повышением контагиозности штаммов Альфа, Дельта и Омикрон). Такие результаты открывают путь к молекулярной эпидемиологии, формированию новых терапевтических стратегий (таргетные ингибиторы, вакциндизайн) и уточнению биохимических механизмов вирусной патогенности.

Благодаря сочетанию структурных и последовательностных подходов, работе с большими биологическими данными и строгому машинному анализу, достигнут не только качественный, но и количественный — формализованный и воспроизводимый — прорыв в автоматизации поиска важных регуляторных сайтов протеолиза. Разработанный инструментарий может использоваться в различных биомедицинских, фармацевтических и фундаментальных задачах: от таргетного поиска новых белковых маркёров и мишеней до системной реконструкции протеолитических сетей клеток и тканей. Работа является примером точного переноса сложной биологической проблемы в поле эффективных вычислительных решений, способствует интердисциплинарному диалогу между биологами, биофизиками и специалистами по анализу данных и открывает перспективы для дальнейших проектных исследований в отечественной и мировой науке.

Структура, содержание и основные результаты работы

Структура диссертации традиционна и логична: введение, обзор современной литературы, описание разрабатываемых методов, их апробация и приложение к актуальным биологическим задачам. В работе обосновано использование пространственных и последовательностных характеристик для определения «уязвимых» к протеолизу участков белка; большая часть экспериментов выполнена на крупных массивах данных, что повышает доверие к выводам.

К числу ключевых результатов относятся:

1. Создание универсальной обучаемой модели для оценки восприимчивости участков белка к протеолизу на основе структурных характеристик.

2. Разработка PWM-моделей специфичности для большого числа человеческих протеаз.

3. Построение интегративной системы с сопоставлением качества работы существующим мировым инструментам (Procleave).

4. Демонстрация практической применимости разработанных моделей на примере сайтов расщепления S-белка коронавируса SARS-CoV-2 и прогнозирования эффекта его мутаций.

Достоинства и недостатки работы

Диссертация обладает безусловными достоинствами:

1. Междисциплинарность и сочетание современных методов машинного обучения, структурной биоинформатики и молекулярной биологии.

2. Практическая значимость для медицины и биотехнологии.

3. Открытые и воспроизводимые результаты (выкладка моделей и кода).

4. К замечаниям можно отнести:

5. Необходимость дальнейшей интеграции информации о посттрансляционных модификациях и динамике структур белков, что отмечено и автором.

6. Часть выводов касается прогнозирования вероятности расщепления, что не всегда может быть напрямую подтверждено *in vivo* и требует дополнительной экспериментальной проверки.

Тем не менее, высказанные замечания не снижают высокой научной и прикладной значимости проведённого исследования.

Соответствие паспорту специальности

Работа Матвеева Е.В. в полной мере соответствует специальности 1.5.8. Математическая биология, биоинформатика, и полностью удовлетворяет критериям, предъявляемым к кандидатским диссертациям.

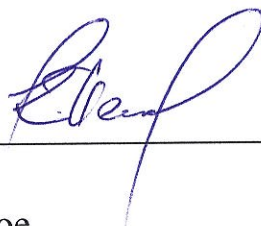
Заключение

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.

Отзыв ведущей организации подготовлен: Попадьным Константином Юрьевичем, к.б.н. Отзыв ведущей организации обсужден в рамках научно-методического семинара Центра геномных исследований, протокол №1 от 2 сентября 2025 года.

Старший научный сотрудник
Центра геномных исследований,
кандидат биологических наук



/ К.Ю. Попадын

Федеральное государственное автономное
образовательное учреждение высшего образования
«Балтийский федеральный университет имени Иммануила Канта»
Адрес: 236016, г. Калининград, ул. А. Невского, д.14.
+7 4012 59 55 96
post@kantiana.ru

«12» 09 2025 г.

