

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ
УЧРЕЖДЕНИЕ «НАУЧНО–ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
БИОМЕДИЦИНСКОЙ ХИМИИ ИМЕНИ В.Н. ОРЕХОВИЧА»

На правах рукописи

Константинов Михаил Александрович

**ПРОТЕОЛИТИЧЕСКИЕ СВОЙСТВА СЕКРЕТИРУЕМЫХ ФЕРМЕНТОВ
ПРЕДСТАВИТЕЛЕЙ РОДА *LYSOBACTER***

1.5.4. – Биохимия

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:
кандидат биологических наук
Торопыгин Илья Юрьевич

Научный консультант:
доктор биологических наук
Жданов Дмитрий Дмитриевич

Москва 2025

Оглавление

СЛОВАРЬ ТЕРМИНОВ И СОКРАЩЕНИЙ.....	5
1. ВЕДЕНИЕ, ЦЕЛЬ И ЗАДАЧИ.....	7
1.1. Актуальность проблемы, цель и задачи.....	7
1.2. Научная новизна работы.....	9
1.3. Теоретическая и практическая значимость работы.....	10
1.4. Личный вклад автора	10
1.5. Основные положения, выносимые на защиту.....	11
1.6. Степень достоверности и апробации результатов	11
2. ОБЗОР ЛИТЕРАТУРЫ.....	13
2.1. Протеолитические ферменты и их бактериолитические свойства	13
2.1.1. Ферменты бактерий рода <i>Lysobacter</i>	18
2.2. Современные подходы к изучению специфичности протеаз	24
2.2.1. Биоинформатические методы: машинное обучение, молекулярный докинг, AlphaFold.).....	25
2.3. Методы определения кинетических параметров протеаз.....	33
2.3.1. Спектрофотометрические методы.....	33
2.3.2. Флуориметрические методы	35
2.3.3. Радиометрические методы	36
2.3.4. Электрохимические методы	38
2.4. MALDI и LC-MS/MS как инструмент анализа специфичности и кинетических параметров протеаз	40
2.4.1. История создания масс-спектрометрии, принцип устройства	43
2.4.2. Методы обработки масс-спектрометрических данных для оценки специфичности протеаз	54
2.4.3. Программные инструменты для анализа ферментативной специфичности.....	59
2.5. Методы количественной оценки в масс-спектрометрии с изотопными метками	63
2.5.1. Метод AQUA, использование метки ^{18}O	66
3. МАТЕРИАЛЫ И МЕТОДЫ	70

3.1. Определение специфичности Vlr с использованием MALDI-TOF масс-спектрометрии	70
3.1.1. Гидролиз белков с использованием Vlr	70
3.1.2. Масс-спектрометрия MALDI-TOF	70
3.1.3. Идентификация продуктов гидролиза белков ферментом Vlr	71
3.2. Гидролиз фракций белков лизата печени крысы и сыворотки человека Vlr и L5 LC-MS/MS	72
3.2.1. Получение лизата печени крысы	72
3.2.2. Получение сыворотки человека	72
3.2.3. Разделение и использованием эксклюзионной хроматографии, фракционирование	73
3.2.4. Гидролиз белков для определения специфичности исследуемых ферментов	74
3.2.5. LC-MS/MS анализ	75
3.2.6. Биоинформатрический анализ специфичности Vlr и L5.	77
3.3. Определение кинетических параметров с использованием изотопно-меченых стандартов и MALDI-TOF масс-спектрометрии	79
3.3.1. Получение пептидов для включения изотопной метки	79
3.3.2. Изотопный обмен	80
3.3.3. Гидролиз белков исследуемыми протеазами для определения кинетических параметров	80
3.3.4. MALDI масс-спектрометрия	81
4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ	83
4.1. Идентификация внеклеточных протеаз бактерий рода <i>Lysobacter</i> , и анализ их гомологии с родственными протеазами	83
4.2. Идентификация Vlr и L5. Определение протеолитической специфичности на модельных белках методом MALDI-TOF MS	85
4.3. Разработка и применение программного инструмента CleaveScore для анализа специфичности протеаз	88
4.4. Анализ специфичности протеаз на основе данных LC-MS/MS	105
4.5. Статистический анализ мотивов в окрестности сайтов разреза, сравнительный анализ с трипсином	112
4.6. Сравнительный анализ специфичности ферментов Vlr и L5 по длине идентифицированных пептидов	118

4.7. Определение кинетических параметров с использованием количественной масс-спектрометрии с метками ^{18}O	122
4.7.1. Получение стандартов пептидов, определение концентрации и подтверждение изотопного обмена	123
4.7.2. Расчёт концентрации продуктов гидролиза по концентрации изотопно-меченого стандарта	130
4.7.3. Валидация метода на модельной системе с использованием трипсина	135
4.7.4. Определение кинетических параметров <i>V_{lr}</i> и <i>L5</i>	138
5. ЗАКЛЮЧЕНИЕ	152
6. ВЫВОДЫ	154
8. БЛАГОДАРНОСТИ	156
8. ФИНАНСИРОВАНИЕ	156
9. СПИСОК ЛИТЕРАТУРЫ	157

СЛОВАРЬ ТЕРМИНОВ И СОКРАЩЕНИЙ

a.o. – аминокислотный остаток

ВМВ – внешние мембранные везикулы

K_M – константа Михаэлиса: концентрация субстрата, при которой скорость ферментативной реакции достигает половины максимальной

Alp – α -литическая протеаза

Blp – β -литическая протеаза

FASTA – текстовый формат представления аминокислотных или нуклеотидных последовательностей

FC (Fold change) – кратное изменение: отношение частоты аминокислоты в исследуемом множестве к её частоте в фоновом, характеризующее степень обогащения или подавления.

FDR – (False Discovery Rate) доля ложных отклонений, статистический показатель, отражающий долю ложноположительных результатов среди всех значимых находок

k_{cat} – каталитическая константа: количество молекул субстрата, превращаемых в продукт одной молекулой фермента за единицу времени

LC-MS/MS – жидкостная хроматография в сочетании с тандемной масс-спектрометрией

MALDI-TOF (matrix assisted laser desorption/ionization) – лазерная десорбция/ионизация из объема матрицы; TOF (time of flight) – времяпролетный масс-анализатор

MD – молекулярная динамика

MS – масс-спектрометрия

PD (Percentage difference) – процентная разница: показатель, отражающий разницу в частоте встречаемости аминокислоты между исследуемым и фоновым множеством, выраженную в процентах

ppm (Parts per million) – миллионная доля

PSSM (Position-Specific Scoring Matrix) – позиционно-специфичная матрица оценок: таблица, отражающая частоту или значимость каждой аминокислоты в определённых позициях выровненных последовательностей

SEC (Size-Exclusion Chromatography) – метод разделения молекул по их гидродинамическому размеру, также известный как гель-фильтрация

1. ВЕДЕНИЕ, ЦЕЛЬ И ЗАДАЧИ

1.1. Актуальность проблемы, цель и задачи

Протеолитические ферменты играют ключевую роль в регуляции биологических процессов и находят широкое применение в биотехнологии, фармацевтике, пищевой промышленности, сельском хозяйстве и экологических технологиях. Одним из перспективных направлений является использование бактериальных протеаз, обладающих бактериолитической активностью, то есть способностью разрушать клеточные стенки других микроорганизмов за счёт гидролиза компонентов клеточной стенки. Такие ферменты применимы в качестве средств биоконтроля патогенной микрофлоры, ферментных добавок в кормах, агробиологических препаратов.

Особый интерес представляют внеклеточные ферменты бактерий рода *Lysobacter*, известных своей способностью лизировать широкий спектр микробных организмов, что связано с секрецией внеклеточного комплекса литических ферментов. Эти микроорганизмы продуцируют широкий спектр гидролаз, включая сериновые и металлопротеазы, липазы и глюканазы, которые реализуют внеклеточное пищеварение. Хотя ферменты *Lysobacter* обладают высоким биотехнологическим потенциалом, их молекулярная характеристика, включая аминокислотную специфичность и кинетические параметры, остаётся недостаточно изученной.

Традиционные методы изучения активности протеаз, основанные на использовании хромогенных или флуорогенных субстратов, часто не позволяют охарактеризовать природные ферменты с нестандартной или широкой специфичностью. Аналогичные сложности возникают при определении кинетических параметров, особенно в тех случаях, когда отсутствуют подходящие коммерческие субстраты для фермента. В этих условиях особенно востребованы универсальные подходы, позволяющие исследовать как специфичность, так и

каталитические параметры таких ферментов в условиях, близких к физиологическим.

В настоящей работе объектами исследования были выбраны две внеклеточные протеазы из *Lysobacter capsici* – β -литическая протеаза (Blp) и протеаза L5, отличающиеся по предполагаемому механизму действия и спектру субстратов. Целью исследования являлось определение аминокислотной специфичности и кинетических характеристик этих ферментов с использованием метода жидкостной хроматографии и тандемной масс-спектрометрии (LC-MS/MS) для определения субстратной специфичности и метода количественной масс-спектрометрии с меткой ^{18}O для определения кинетических параметров протеаз. Впервые были представлена полная характеристика сайтов гидролиза Blp и L5 с использованием ПО собственной разработки (CleaveScope), а также проанализированы мотивы аминокислотного окружения сайтов гидролиза. Кроме того, были определены кинетические параметры исследуемых ферментов с использованием собственного метода определения кинетических параметров с использованием изотопно-меченых пептидов и MALDI-TOF масс-спектрометрии. Этот подход позволил получить значения K_M , k_{cat} и k_{cat}/K_M без использования синтетических субстратов.

Цель данной работы – идентификация литически активных ферментов из внеклеточной жидкости представителей группы *Lysobacter*, определение их специфичности и активности для оценки их возможного применения в биотехнологии и биомедицинских исследованиях.

Для достижения указанной цели поставлены следующие **задачи**:

1. Провести идентификацию литически активных протеаз во внеклеточной жидкости бактерий рода *Lysobacter*, обладающих активностью в отношении белковых субстратов.

2. Определить протеолитическую специфичность выделенных ферментов на модельных белках, провести анализ мотивов в окрестности сайтов гидролиза и предположить возможные механизмы действия.

3. Разработать модельную систему для определения кинетических параметров протеаз с использованием масс-спектрометрии MALDI-TOF и изотопно-меченых стандартов.

4. Рассчитать кинетические параметры исследуемых ферментов на выбранных субстратах с использованием предложенного подхода.

1.2. Научная новизна работы

Впервые проведена комплексная идентификация сайтов преимущественного гидролиза белков под действием протеаз B1p и L5, выделенных из внеклеточных везикул *Lysobacter capsici* ВКМ В-2533^Т и культуры *Lysobacter capsici* XL1 соответственно. В отличие от ранее опубликованных данных, основанных на ограниченном числе модельных субстратов, в настоящей работе использован подход, основанный на неселективном картировании продуктов гидролиза большого пула белков с последующей идентификацией пептидов методом LC-MS/MS и статистической обработкой результатов. Это позволило получить достоверную информацию о субстратной специфичности ферментов и выявить воспроизводимые аминокислотные предпочтения в окружении сайтов гидролиза.

Также впервые для указанных ферментов были определены кинетические параметры с использованием количественной масс-спектрометрии MALDI-TOF и стабильной изотопной метки ¹⁸O, внедряемой методом изотопного обмена. Полученные данные позволили количественно охарактеризовать каталитическую активность B1p и L5 по отношению к индивидуальным пептидам, что ранее не проводилось.

1.3. Теоретическая и практическая значимость работы

Полученные данные о специфичности Vlr и L5 позволяют точно определять сайты протеолиза в белках, включая трудногидролизуемые участки, нечувствительные к стандартным протеазам. Это расширяет возможности ферментативной обработки в протеомике, включая анализ рекомбинантных белков без сайтов для трипсина, что показано автором на примере использования Vlr (ASMS 2020, стр. 601). Идентифицированные сайты и кинетические характеристики также имеют прикладное значение для разработки антимикробных препаратов на основе ферментов *Lysobacter*, включая потенциальную замену препарата Лизоамидаза (Патент РФ №2296576), ранее производимого из *L. capsici* XL1. Разработанный метод определения кинетических параметров на основе MALDI-TOF и ^{18}O -меченых пептидов может быть применён для количественной оценки активности различных протеаз без использования флуоресцентных или радиоактивных меток.

1.4. Личный вклад автора

Работа была выполнена в лаборатории межмолекулярных взаимодействий ИБМХ им. В. Н. Ореховича в период с 2019 по 2025 год. Автором осуществлено планирование и выполнение экспериментальных исследований, включая биоинформатическую обработку данных с использованием разработанных им специализированных программных инструментов. Разработан и реализован подход к определению кинетических параметров протеаз на основе количественной MALDI-TOF масс-спектрометрии с изотопными метками ^{18}O . Предложен способ получения и валидации изотопно-меченых стандартов из продуктов гидролиза модельных белков, а также методика расчёта их концентрации с учётом распределения изотопных форм. Все представленные в работе результаты получены лично автором либо при его непосредственном участии.

1.5. Основные положения, выносимые на защиту

1. Идентифицированы и охарактеризованы две внеклеточные протеазы B1p из *Lysobacter capsici* ВКМ В-2533^Т и L5 из *Lysobacter capsici* XL1, обладающие протеолитической активностью в отношении белковых субстратов. Показана гомология с литическими протеазами других представителей рода *Lysobacter*.

2. Субстратная специфичность ферментов B1p и L5 определена на основе данных LC-MS/MS. Для определения специфичности разработан универсальный программный инструмент CleaveScore. Впервые построены профили частот аминокислотных остатков, выявлены характерные предпочтения и мотивы, отражающие различия в механизмах действия ферментов.

3. Разработан универсальный масс-спектрометрический подход к определению кинетических параметров протеаз, основанный на использовании изотопно-меченых (¹⁸O) пептидных стандартов и MALDI-TOF масс-спектрометрии. Метод не требует применения синтетических субстратов и позволяет количественно оценивать активность ферментов в нативных условиях.

4. Впервые рассчитаны кинетические параметры K_M , k_{cat} и k_{cat}/K_M для ферментов B1p и L5 на модельных белковых субстратах. Методика оказалась применимой как для фермента с широкой специфичностью и высокой скоростью реакции (L5), так и для фермента, активность которого не удавалось оценить стандартными методами (B1p), что подтверждает универсальность предложенного подхода.

1.6. Степень достоверности и апробации результатов

Основные положения работы были опубликованы в рецензируемых научных изданиях, а также представлены в виде постерных докладов на 68-й конференции Американского масс-спектрометрического общества ASMS (Virtual ReBoot 2020), VII съезде биохимиков России, X Российском симпозиуме «Белки и пептиды» (Сочи, 2022), устных докладов на VII съезде биохимиков России, X Российском

симпозиуме «Белки и пептиды» (Сочи, 2022), XIX Всероссийской молодежной школе-конференции по актуальным проблемам химии и биологии (Владивосток 2022), VI Международной конференции Постгеном 2024 (Москва, 2024).

По теме диссертации опубликовано 10 печатных работ – 3 статьи в рецензируемых научных журналах, 2 из которых соответствуют рубрикатору ВАК по специальности «Биохимия» и 7 в трудах конференций. Также получено свидетельство о государственной регистрации программы для ЭВМ.

2. ОБЗОР ЛИТЕРАТУРЫ

2.1. Протеолитические ферменты и их бактериолитические свойства

Ферменты представляют собой обширный класс биологических катализаторов, обеспечивающих протекание практически всех биохимических реакций в живых организмах [1–3]. Среди них особое место занимают протеазы (или пептидазы, КФ 3.4, класс гидролаз согласно международной классификации ферментов [4]) – ферменты, катализирующие расщепление пептидных связей в белках [5–8]. Благодаря этой активности протеазы выполняют роль своеобразных “молекулярных ножниц” и участвуют в многих биологических процессах [9].

Протеазы обнаружены во всех царствах жизни – от бактерий и грибов до высших растений и животных; даже вирусы кодируют собственные протеазы, которые расщепляют полипротеиновые прекурсоры на функциональные белки, что необходимо для вирусной репликации [10]. В геноме человека насчитывается по крайней мере 589 генов протеаз, что составляет более 2% от всех генов. Совокупность всех протеолитических ферментов организма обозначается понятием «деградом». Протеазы чрезвычайно разнообразны и по механизму действия подразделяются на пять основных классов: сериновые, цистеиновые, аспартатные, металлопротеазы, а также протеазы с остатком треонина в активном центре (рисунок 1) [11].

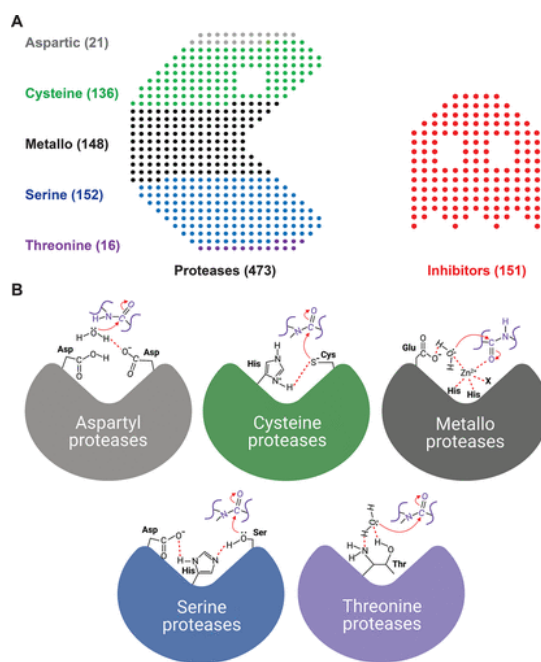


Рисунок 1. Классификация протеаз по каталитическому механизму. Показано общее количество известных протеаз и их деление на пять основных классов: сериновые, металлопротеазы, цистеиновые, аспартатные и треониновые. Справа схематично показаны ингибиторы протеаз. Внизу – схемы механизмов катализа различных типов протеаз. Источник [11]

Кроме того, их классифицируют по локализации действия на эндопептидазы (расщепляющие внутренние связи белковой цепи) и экзопептидазы (отрезающие концевые аминокислоты). Индивидуальные протеазы различаются по специфичности: некоторые строго распознают единственный белок-мишень, тогда как другие способны гидролизовать широкий спектр субстратов [12–14].

Протеолитические ферменты играют центральную роль во многих физиологических процессах. Они участвуют в пищеварении (например, пепсин, трипсин и химотрипсин расщепляют пищевые белки) [15], в каскадах свертывания крови (сериновые протеазы тромбина, плазмина и др.) [16], в активации гормонов и рецепторов [17], в иммунной защите (протеазы комплекса комплемента, гранзимы цитотоксических лимфоцитов) [18,19], а также в системах внутриклеточного протеолиза и контролируемой деградации белков (протеасомы, лизосомальные катепсины и др.) [20,21]. Благодаря ограниченному протеолизу белков-мишеней протеазы способны тонко регулировать сигнальные пути и клеточные события [9].

Активность протеаз строго контролируется на разных уровнях – от транскрипции и синтеза неактивных проферментов до специфических ингибиторов – поскольку дисбаланс протеолиза приводит к тяжелым патологиям [9]. Мутации в генах протеаз ассоциированы как минимум со 135 наследственными заболеваниями у человека, а ненормальная протеолитическая активность вовлечена в развитие множества распространенных болезней – от сердечно-сосудистых и воспалительных расстройств до нейродегенеративных процессов и рака [22,23].

Не случайно протеазы считаются перспективными мишенями для фармакотерапии: эти ферменты признаны отличными молекулярными таргетами для разработки новых ингибиторов (например, антигипертензивные препараты – ингибиторы АПФ [24], антиретровирусные препараты – ингибиторы протеазы ВИЧ) и других лекарственных стратегий [9,25].

В настоящее время в клинической практике используются десятки препаратов, воздействующих на протеазы (ингибиторы сериновой протеазы ДПП-4 при сахарном диабете [26], ингибиторы аспартатной протеазы β -секретазы при болезни Альцгеймера [27], ингибиторы металлопротеаз матрикса при раке [28] и др.), и этот список постоянно пополняется в связи с расширением знаний о роли протеаз в патогенезе заболеваний.

Высока и прикладная значимость протеолитических ферментов: они широко используются в промышленности и биотехнологии благодаря способности разлагать белковые субстраты. Микроорганизмы (бактерии и грибы) служат основным источником протеаз в промышленном производстве за счёт своего огромного биоразнообразия и простоты культивирования. По оценкам, свыше 60% мирового выпуска промышленных ферментов приходится именно на протеазы (преимущественно микробного происхождения), из них около 35% – щелочные протеазы. Эти ферменты находят применение в производстве моющих средств, в пищевой (сыроизготовление, пивоварение, мясопереработка), кожевенной,

фармацевтической и косметической отраслях [29]. Таким образом, исследования протеаз важны не только для медицины, но и для развития биотехнологий.

Особый интерес представляют бактериальные протеазы, обладающие разнообразными функциями. Для самих бактерий эти ферменты выполняют как “хозяйственные” роли – расщепление белков среды до пептидов и аминокислот для питания, устранение поврежденных клеточных белков – так и служат мощными инструментами патогенности и конкуренции. Множество патогенных бактерий вырабатывают внеклеточные протеазы как факторы вирулентности, помогающие колонизации хозяина. Секретируемые протеазы способны разрушать компоненты внеклеточного матрикса и соединительной ткани (коллагеназы, эластазы), инактивировать иммунные белки хозяина (например, иммуноглобулины, антимикробные пептиды) или расщеплять белки свертывающей и комплемент-систем, облегчая распространение инфекции. Такие ферменты, как показано, способствуют инвазии и разрушению тканей хозяина, что приводит к более тяжелому течению инфекций и замедлению заживления повреждений [30]. Например, *Pseudomonas aeruginosa* секретирует эластазу и щелочную протеазу, *Staphylococcus aureus* – сериновые протеазы (стафилопайны) и металлопротеазу ауреолизин, *Streptococcus* вырабатывают цистеиновую протеазу SpeB и т.д., и все они вносят вклад в патогенез соответствующих инфекций.

С другой стороны, бактериальные протеазы могут приносить пользу. В экосистемах они участвуют в круговороте органического вещества, а в биотехнологиях используются для синтеза продуктов и утилизации отходов. Их востребованность обусловлена рядом ценных свойств – многие бактериальные ферменты проявляют высокую стабильность и активность в экстремальных условиях (щелочные или кислотные pH, высокая температура, присутствие детергентов), что делает их удобными для промышленных процессов [29].

Перспективы применения бактериальных протеаз активно изучаются. Одно из направлений – использование ферментов непатогенных бактерий для борьбы с

патогенами. Установлено, например, что некоторые почвенные бактерии-антагонисты синтезируют протеазы, разрушающие биопленки и клеточные стенки других микроорганизмов. Особое внимание привлекает род *Lysobacter* – грамотрицательные бактерии, известные способностью продуцировать широкий набор гидролитических ферментов. *Lysobacter* рассматриваются как факультативные хищники микромира, способные лизировать другие бактерии и грибы с помощью секретируемых ферментов [29,31].

Благодаря такому “вооружению” представители *Lysobacter* проявляют выраженную бактерио- и фунгицидную активность, подавляя патогены растений и даже некоторых устойчивых клинических возбудителей. Эти свойства делают *Lysobacter* перспективным источником новых антимикробных агентов и биопрепаратов. Изучение специфичности литических протеаз и условий их действия может открыть пути для создания на их основе комбинированных терапевтических подходов (например, использование бактериальных протеаз вместе с антибиотиками для повышения эффективности последних).

Протеолитические ферменты – обширный и важнейший класс биокатализаторов, необходимый для жизни организмов. Их значение подтверждается огромным числом физиологических функций и вовлеченностью в заболевания, а также широчайшим применением в биотехнологии. В особенности протеазы бактерий представляют большой научный и прикладной интерес как фактор патогенности и как инструмент, который можно использовать в медицине и сельском хозяйстве. В следующем разделе подробнее рассмотрены протеолитические ферменты рода *Lysobacter* – как пример бактерий, чьи уникальные ферментативные системы открывают новые возможности в биотехнологиях и биомедицине.

2.1.1. Ферменты бактерий рода *Lysobacter*

Представители рода *Lysobacter* известны как продуценты разнообразных антимикробных веществ – как низкомолекулярных антибиотиков, так и бактериолитических ферментов. Представители рода *Lysobacter* являются продуцентами разнообразных антимикробных веществ, таких как низкомолекулярные антибиотики и бактериолитические ферменты [32]. Одними из первых обнаруженных бактериолитических ферментов были α - и β -литические протеазы, выделенные в 1965 г. из грамотрицательных бактерий *Mucobacter* 495 (сейчас классифицируется как *Lysobacter enzymogenes* ATCC 29487) [33,34]. Эти две протеазы получили название благодаря способности вызывать лизис бактериальных клеток. α -литическая протеаза (Alp) и β -литическая протеаза (Blp) стали модельными объектами для изучения бактериолитических ферментов. Показано, что Alp относится к сериновым протеазам (по механизму действия близка к химотрипсину), тогда как Blp является металлопротеазой семейства M23, для активности которой необходим ион Zn^{2+} [35]. Обе протеазы активны в щелочном диапазоне pH и термостабильны, однако Blp *L. capsici* имеет несколько более низкий температурный оптимум и отличается меньшей устойчивостью к высокой ионной силе раствора. Характерно, что β -литическая протеаза проявляет мощное лизирующее действие в отношении стафилококков: например, для Blp *L. capsici* ВКМ В-2533^Т минимальная ингибирующая концентрация (МИК) против метициллин-резистентного *Staphylococcus aureus* (MRSA) составляет всего ~2,85 мкг/мл. При этом Blp практически не действует на грамотрицательные бактерии вследствие экранирующей роли внешней мембраны [32].

Сам фермент Blp синтезируется в форме предшественника – препро-белка, с длиной в 378 аминокислот. При этом длина сигнального пептида составляет 26 аминокислот, а длина пробелка 173. Зрелая же часть белка состоит из 179 аминокислот, с массой 19,1 кДа, и имеющая 2 дисульфидные связи. По классификации базы данных MEROPS, фермент относится к семейству металлоэндопептидаз семейства M23 [36,37], поскольку в его активном центре

содержится мотив НХН, характерный для этого семейства протеаз, также найденный в бета-литических металлоэндопептидазах бактерий *L. enzymogenes* и *Lysobacter sp.*, где молекула цинка присоединена к двум гистидинам [38,39].

Другим важным ферментом, выделенным из *L. enzymogenes*, является эндопротеиназа LysC – сериновая протеаза, избирательно гидролизующая пептидные связи с участием карбоксильной группы остатка лизина [40]. LysC нашла широкое применение в биотехнологиях как фермент для ограниченного протеолиза при идентификации белков. Помимо *L. enzymogenes*, другие виды этого рода также продуцируют бактериолитические протеазы. Так, штамм *Lysobacter sp.* XL1, секретирует в культуральную жидкость по меньшей мере пять разных лизирующих ферментов, обозначенных L1, L2, L3, L4 и L5 [41]. Фермент L1 (ранее называвшийся лизоамидаза) был выделен и охарактеризован первым [42,43]. L1 представляет собой бактериолитическую протеазу, очищенную из культуры *Lysobacter* и способную гидролизовать пептидогликан бактерий; по своей структуре L1 близка к α -литической протеазе [44]. Последовательно были открыты ферменты L2–L4 [45,46], а самым последним обнаруженным оказался фермент L5 – внеклеточная сериновая протеаза, присутствующая в культуральной жидкости *Lysobacter sp.* XL1 преимущественно в составе внешних мембранных везикул (ВМВ) [41,47]. Именно L1 и L5 привлекли особое внимание исследователей благодаря высокой литической активности [48]. L5, будучи секретируемым с участием ВМВ, оказался способен вызывать гибель как грамположительных, так и некоторых грамотрицательных бактерий. Показано, что очищенный фермент L5 эффективно лизирует клетки *S. aureus*, *Micrococcus luteus* и других грамположительных бактерий, а в составе везикул он также способен разрушать и клетки грамотрицательных бактерий (например, фитопатогенной *Erwinia carotovora*) [41]. При этом лизис бактериальных клеток под действием L5 значительно усиливается, когда фермент доставляется к мишени в составе мембранных везикул – в эксперименте ВМВ с L5 вызывали более быстрый и полный лизис тест-культур, чем эквивалентные количества свободного фермента.

Этот факт был одним из первых доказательств бактериолитического действия, опосредованного наружными мембранными везикулами, «начинёнными» гидролитическими ферментами. Высокая активность L5 в отношении широкого круга микроорганизмов позволяет рассматривать его (вместе с содержащими его ВМВ) как перспективный антимикробный препарат нового типа.

Для обнаружения и выделения ферментов типа Вlr и L5 применяется комплекс биохимических и микробиологических методик. Культуральная жидкость *Lysobacter* с бактериолитической активностью подвергается фракционированию: осаждение сульфатом аммония, ионообменная хроматография и гель-фильтрация позволяют получить очищенные препараты целевых ферментов [32]. Так, из супернатанта культуры *L. capsici* ВКМ В-2533^Т было выделено несколько белков с лизирующей активностью, среди которых идентифицирована Вlr. Идентификацию ферментов проводили методом масс-спектрометрического пептидного картирования (MALDI-TOF) с последующей идентификацией белков. Лизирующая (бактериолитическая) активность изучается *in vitro* с использованием тест-культур бактерий. Применяются, в частности, плотностной (турбодиметрический) анализ – измерение уменьшения оптической плотности суспензии убитых или живых бактериальных клеток при добавлении фермента – и метод стандартного агарового диффузионного теста (лунки или пятна на “газоне” (равномерный сплошной слой бактериального роста) из чувствительных бактерий). Эти подходы позволяют выявить спектр действия ферментов и сравнительную активность. Например, литический препарат *L. capsici* вызывал появление зон лизиса у различных штаммов *Staphylococcus*, *Micrococcus* и др., тогда как в отношении грамотрицательных бактерий (*E. coli*, *Pseudomonas spp.*) подобных зон не наблюдалось [35]. Также определяется МИК для ростовых тестов – например, для Вlr *L. capsici* показано полное подавление роста MRSA при 2,85 мкг/мл [32].

Для характеристики протеолитических свойств этих ферментов используются субстратные тесты: исследуется способность протеазы гидролизовать различные модельные белки (казеин, эластин, желатин, коллаген, и

др.), а также хромогенные пептиды. По продуктам гидролиза судят о специфичности фермента – предпочитаемых им связях. Так, β -литическая протеаза *L. capsici* гидролизует широкий набор пептидных связей в субстратах и способна расщеплять сайты, ранее не описанные для других Vlp. В частности, было обнаружено, что Vlp, как и аналогичные ферменты из *Lysobacter enzymogenes*, разрывает гликановый мостик пентаглицина в пептидогликане стафилококков, а также некоторые связи внутри глобулярных белков (например, в казеине и коллагене) – причём преимущественно на поверхности белковой молекулы. Способность лизировать бактерии может быть связана именно со способностью фермента гидролизовать пентаглицин, поскольку такая структура характерна для клеточной стенки бактерий. Это свидетельствует о весьма широком спектре субстратной специфичности Vlp, превосходящем по охвату типов связей даже таковой у лизоцима или лизостафина. Что касается сериновой протеазы L5, то было показано, что она обладает по меньшей мере двумя типами пептидогликан-гидролазной активности: эндопептидазной (расщепляет пептидную связь Gly–Gly в межпептидных мостиках пептидогликана) и амидазной (расщепляет амидную связь между N-ацетилмурамовой кислотой и L-аланином в муреине). Таким образом, L5 одновременно действует как глицин-глициновая эндопептидаза и как N-ацетилмурамоил-L-аланин-амидаза, что весьма необычно для одного фермента. Предполагается, что наличие двух типов активности облегчает L5 разрушение клеточной стенки *Staphylococcus sp.* [48].

Структурно-функциональные исследования проливают свет на молекулярные основы действия бактериолитических протеаз *Lysobacter*. В 1980-е гг. была клонирована и секвенирована генетическая последовательность α -литической протеазы *Lysobacter enzymogenes*, показавшая наличие предшествующего сигнального и про-пептида, необходимых для созревания фермента и правильного фолдинга [49]. Кроме того, в 1985 г. была опубликована кристаллическая структура зрелой формы α -литической протеазы с разрешением 1,7 Å, в которой были подтверждены характерные элементы активного центра и

консервативные структуры, характерные для сериновых хемотрипсиноподобных протеаз [50].

Структура белка L5 была определена с высоким разрешением (1,6 Å). Показано, что по общей конформации L5 очень близок к бактериолитической протеазе L1 того же штамма и к α -литической протеазе *L. enzymogenes*, несмотря на относительно невысокую гомологию аминокислотных последовательностей этих ферментов. В структуре L5 при этом выявлены уникальные участки, отсутствующие у его гомологов, которые, вероятно, обуславливают отличия в свойствах фермента. Интересно, что L5 способен образовывать амилоидоподобные фибриллы при концентрировании и хранении, что может быть связано с особенностями его строения [48].

Как было отмечено выше, бактериолитические ферменты рода *Lysobacter* представляют практический интерес как антибактериальные агенты, однако получение их в достаточных количествах затруднено. Трудности связаны с тем, что экспрессия генов этих ферментов в гетерологичных системах (например, в *E. coli*) часто приводит к образованию нерастворимых включений или токсична для продуцента, а полноценная секреция и созревание фермента требуют специфических секреторных механизмов, присущих *Lysobacter* [34]. Тем не менее, предпринимались успешные попытки гетерологичной экспрессии. В частности, β -литическую протеазу из *L. enzymogenes* удалось экспрессировать в *Bacillus subtilis*, добившись секреции активного фермента во внеклеточную среду. Было показано, что для правильного созревания фермента необходима активность протеаз *B. subtilis*, удаляющих пропептид, а также функционирование тиолдисульфидных оксидоредуктаз, обеспечивающих формирование нативной структуры фермента [51].

Для получения препарата L5 с целью рентгеноструктурного анализа использовали рекомбинантный штамм *Escherichia coli* BL21(DE3)/pLysE, продуцирующий белок L5 в виде телец включения. Штамм был трансформирован

плазмидой pVproZ, кодирующей N-концевой 6xHis-меченый предшественник L5. Предшественник – препробелок, с длиной в 378 аминокислот.

В последних работах предложен альтернативный подход – создание гомологичной системы экспрессии в самом исходном штамме. Так, для *L. capsici* ВКМ В-2533^T разработана плазида на основе вектора pBBR1-MCS5 с сильными промоторами (GroEL(A) и T5), которая была внесена в бактерию и обеспечила многократное (в сотни раз) повышение транскрипции гена Vlp. В результате выход Vlp увеличился в 6,7–8,5 раза по сравнению с диким штаммом, достигнув ~44 мг/л при культивировании в ферментере. Более того, с помощью той же системы была осуществлена сверхэкспрессия гена другого бактериолитического фермента *L. capsici* – сериновой протеазы (обозначенной как Serp), уровень которой повысился почти в 600 раз по сравнению с естественным. Эта работа продемонстрировала принципиальную возможность создания высокоэффективных продуцентов бактериолитических ферментов без использования гетерологичных хозяев. Развитие таких биотехнологических подходов открывает путь к практическому применению ферментов Vlp, L5 и им подобных [34].

Ферменты рода *Lysobacter* имеют значительный потенциал для применения в ветеринарии и пищевой промышленности благодаря антимикробным и протеолитическим свойствам этих бактерий. В сельском хозяйстве виды *Lysobacter* уже рассматриваются как агенты биоконтроля – они могут использоваться для подавления фитопатогенов при обработке растений или почвы, а также в качестве средств защиты растений [52]. Кроме того, фермент LysC и Alp широко используется в протеомике и для первичного анализа белков [39,53,54].

Таким образом, бактериолитические ферменты *Lysobacter* представляют собой ценный ресурс как для фундаментальных исследований, так и для потенциального применения в биомедицине, агротехнологиях и аналитической биохимии. Вместе с тем, для практической реализации такого применения

необходимо детальное изучение каталитических свойств этих ферментов – в первую очередь их субстратной специфичности и кинетических параметров.

2.2. Современные подходы к изучению специфичности протеаз

Протеолитические ферменты катализируют гидролиз амидных связей в белках и пептидах. Специфичность подразумевает селективность фермента по отношению к аминокислоте, образующей гидролизуемую амидную связь с карбоксильной группой. Определение специфичности является необходимым этапом для характеристики любой новой протеазы.

Для описания специфичности широко используется номенклатура Шехтера и Бергера [55,56]: аминокислотные остатки субстрата, находящиеся в направлении N-конца белка от расщепляемой связи, обозначаются как $P_1, P_2, P_3...$ (P_1 – непосредственно перед разрывом), а C-концевые от места расщепления – как $P_1', P_2', P_3'...$; соответствующие им субсайты (или подцентры) фермента именуются $S_1, S_2...$ и S_1', S_2' и так далее. Расщепление происходит между позициями P_1 и P_1' , которые занимают комплементарные субсайты S_1 и S_1' протеазы. Помимо субстратсвязывающих карманов, в активном центре многих протеаз имеется каталитическая триада – три аминокислоты, непосредственно осуществляющие гидролиз пептидной связи (классический пример – Ser-His-Asp у сериновых протеаз). Каталитическая триада обеспечивает сам акт расщепления, тогда как окружение активного центра (субсайты S) обуславливает специфичность фермента к аминокислотам в позициях Р-сайта субстрата. Таким образом, специфичность протеазы можно определить как совокупность предпочтений фермента к аминокислотным остаткам в позициях $P_4-P_3-P_2-P_1 / P_1'-P_2'...$ вокруг расщепляемой связи, обусловленных строением её активного центра и субсайтов.

Изучение специфичности протеаз даёт информацию о структуре и функции активного центра, белок-белковом взаимодействии, регуляции внутриклеточных и внеклеточных путей, а также об эволюции генов протеаз и субстратов [57].

Знание предпочтительных мотивов позволяет предсказывать новые потенциальные субстраты фермента в протеоме. Сравнение известных сайтов расщепления часто используется для вывода консенсусной последовательности, характерной для данной протеазы [58]. На базе такой информации исследователи могут разрабатывать оптимизированные синтетические субстраты.

2.2.1. Биоинформатические методы: машинное обучение, молекулярный докинг, AlphaFold.)

С развитием структурной биологии и компьютерного моделирования стали востребованы *in silico* подходы к предсказанию специфичности протеаз. Так, структура-ориентированные методы на базе машинного обучения позволяют предсказывать и проектировать специфичность протеаз, используя геометрию активного центра и энергетические характеристики взаимодействия с субстратами [59]. Еще одним подходом является структурно-обоснованное моделирование (structure-based prediction), применяющееся для восстановления механизмов селективности и доли позитивного/негативного отбора в специфичности ферментов. В работе С. Хэйр и др. был описан структурно-управляемый алгоритм, позволяющий генерировать пептидные последовательности, потенциально взаимодействующие с протеазами, что облегчает *de novo* предсказание субстратов [60].

Используя структурные данные (из PDB или предсказанные алгоритмом AlphaFold2, AlphaFold3), можно моделировать, как потенциальные субстраты или ингибиторы взаимодействуют с активным центром протеазы. Один из ключевых методов - молекулярный докинг, при котором пептидный субстрат «вкладывается» в активный центр фермента с расчётом энергии комплекса. Такой подход позволяет

быстро оценить прочность связывания заданной аминокислотной последовательности и выявить ключевые взаимодействия между боковыми цепями субстрата и фермента. Существует широкий арсенал инструментов для пептидного докинга, включая AutoDock CrankPer (ADCP), специализированный для гибких пептидов и демонстрирующий высокую точность на тестовых выборках (до 85.7 % успеха на LEADS-PEP) [61]. В частности, ADCP применялся для докинга нативных пептидов к ВИЧ-протеазе, показывая способность успешно предсказывает взаимодействия нативных пептидов, но не распознавая ненативные пептиды.

После первичного этапа молекулярного докинга пептидов к активному центру протеазы важным шагом становится применение молекулярной динамики (MD). В отличие от статического докинга, MD позволяет смоделировать поведение комплекса во времени, учитывая гибкость как пептида, так и фермента. Это обеспечивает более реалистичную оценку стабильности взаимодействия, позволяет отследить сохранение контактов между аминокислотными остатками субстрата (например, в позиции P_1) и субсайтами связывания фермента (например, S_1), а также выявить ключевые водородные связи и гидрофобные взаимодействия, определяющие селективность. Подобный подход был реализован в работе [62], где предложен открытый вычислительный протокол, объединяющий докинг пептидов и MD-симуляции для оценки устойчивости и точности моделей взаимодействия протеазы с потенциальными субстратами. Авторы показали, что включение этапа MD позволяет улучшить отбор биологически релевантных комплексов и повысить предсказательную ценность *in silico* анализа специфичности.

AlphaFold2 также активно используется для структурной характеристики протеаз, особенно когда экспериментальные данные отсутствуют. В работе [63] сообщается, что моделирование 3С-подобных протеиназ невовирусов (3CLpro) с помощью AlphaFold может пролить свет на детерминанты специфичности вирусных протеиназ, структура которых пока неизвестна. Более того, моделирование может оказаться полезным при изучении эволюционного происхождения 3CLpro из многих вирусных семейств, что в будущем поможет

установить субстратную специфичность 3Срго и других вирусных ферментов, имеющих эту гибкую и универсальную структуру.

Несмотря на пользу биоинформатических методов, они имеют ряд существенных ограничений. Алгоритмы машинного обучения часто переобучаются на ограниченных наборах фрагментов субстратов и плохо обобщаются на новые условия. Модели AlphaFold2 дают лишь гипотезы без учета лигандов, ионов, ковалентных модификаций или условий окружающей среды, что было продемонстрировано Тервиллигером и соавторами при сопоставлении с экспериментальными плотностями — лишь при подтверждении через кристаллографию можно утверждать о достоверности модели [64].

В силу этих ограничений, биоинформатические методы служат предварительным этапом исследования структуры и специфичности протеаз, тогда как экспериментальные методы необходимы для верификации *in silico* предсказаний.

2.2.1. Экспериментальные методы: субстратные библиотеки (PS-SCL, HyCoSuL и Phage display)

Классический подход к экспериментальному определению специфичности протеазы — проверка её активности на серии синтетических пептидных субстратов или пептидных библиотек.

Одним из наиболее важных достижений стало создание позиционно-сканирующих синтетических комбинаторных библиотек (PS-SCL — Positional Scanning Synthetic Combinatorial Library) [65]. PS-SCL представляют собой наборы пептидов, в которых одна или несколько позиций последовательно фиксированы на определённой аминокислоте, тогда как остальные позиции содержат эквимолярную смесь всех возможных аминокислотных остатков. Такая организация позволяет систематически оценивать вклад каждой аминокислоты в

конкретной позиции субстрата в эффективность взаимодействия с протеазой. Благодаря применению методов комбинаторной химии подобные библиотеки могут быть быстро синтезированы и протестированы. В зависимости от количества рандомизированных положений (X) и количества аминокислот, включенных в эти положения (Y), в каждой смеси будет YX различных пептидов, каждый из которых будет содержать одну и ту же фиксированную аминокислоту [66]. Ключевым достоинством данного подхода является возможность изолированно оценить влияние конкретной аминокислоты в заданной позиции, минимизируя влияние оставшихся участков пептидной цепи. Одним из первых примеров успешного применения PS-SCL стало исследование, в котором было протестировано более 34 миллионов гексапептидных последовательностей с целью выявления эпитопа, распознаваемого моноклональным антителом [66]. Этот пример наглядно продемонстрировал эффективность метода для быстрого анализа огромного числа пептидных вариантов и выбора наиболее подходящих последовательностей для связывания.

Учитывая высокую эффективность PS-SCL при скрининге разнообразных пептидных последовательностей, данный подход был адаптирован для изучения специфичности протеаз путём включения флуорогенных «уходящих» групп. В первых версиях библиотек для этой цели использовали аминотетилкумарин (АМС), который присоединяли к С-концу пептида (позиции P_1) через амидную связь. При протеолитическом расщеплении этой связи высвобождается свободный АМС, что сопровождается увеличением флуоресценции и служит индикатором активности фермента. Такие библиотеки с флуорогенными метками были успешно применены для изучения специфичности широкого спектра протеаз, включая каспазы 1-9, гранзим В и фермент, активирующий интерлейкин-1 β [67,68].

Использование АМС в качестве флуорогенной метки показало эффективность в изучении специфичности, но имело техническое ограничение: из-за способа связывания со смолой в позиции P_1 можно было использовать лишь аминокислоты с модифицируемыми боковыми цепями (например, аспарагиновая

кислота и лизин). Это вынуждало фиксировать P_1 , ограничивая универсальность метода для протеаз с неизвестной специфичностью [66].

Ограничения АМС-библиотек стимулировали разработку новых флуорогенных субстратов на основе 7-амино-4-карбамоилметилкумарина (АСС) [69]. Благодаря своей бифункциональной структуре, АСС-Fmoc может быть напрямую закреплен на смоле для упрощения синтеза. После присоединения к смоле схемы Fmoc-химии используются для включения любой аминокислоты в начальное положение (P_1), что позволяет варьировать все положения пептида. Эта разработка привела к созданию полных разнообразных библиотек АСС-тетрапептида, которые способны независимо анализировать каждое положение пептидного субстрата. Что наиболее важно, из-за беспрепятственного варьирования аминокислоты в положении P_1 стало возможным анализировать протеазы с неизвестной специфичностью. Кроме того, АСС обладает примерно в три раза большей флуоресценцией по сравнению с АМС, что повышает чувствительность анализа [70].

PS-SCL полезны не только для определения субстратной специфичности протеаз, но и для разработки новых инструментов для их исследования. Однако многие протеазы имеют дополнительную специфичность в основных боковых карманах (prime side pockets), которую не представляется возможным исследовать с помощью данных библиотек. У методов с флуорогенными пептидами есть важные ограничения. Во-первых, присутствие метки (флуорофора, а часто еще и тушителя) изменяет нативную конформацию пептида и может влиять на распознавание. Во-вторых, из-за схемы измерения (отслеживается отщепление метки, привязанной к одному концу пептида) классическая PS-SCL эффективно профилирует сайты гидролиза только с одной стороны от места разреза.

Успешное применение флуорогенной метки АСС в комбинированных библиотеках послужило основой для создания более продвинутых систем [65]. Так, Drag и соавт. разработали позиционно-сканирующие библиотеки, включающие до

110 неприродных аминокислот в позициях P₁–P₄. Эти расширенные библиотеки получили название HyCoSuL (Hybrid Combinatorial Substrate Library) и позволили проводить более глубокий и точный анализ специфичности протеаз. С использованием HyCoSuL был идентифицирован субстрат нейтрофильной эластазы с рекордной каталитической эффективностью [71], а также разработаны селективные субстраты для отдельных каспаз [65]. В дополнение к этому, метод библиотеки контрселекции (CoSeSuL) был применён для создания высокоспецифичных зондов легумаина - важного протеолитического фермента [72]. HyCoSuL-подход продемонстрировал, что введение непредставленных в природе аминокислот позволяет преодолеть проблему “перекрывающейся специфичности” ферментов из одного семейства, добившись уникальных субстратов для каждого из них.

Альтернативной экспериментальной стратегией являются фаговые субстратные библиотеки (метод «substrate phage display») [73]. Здесь потенциал комбинаторики реализуется не химическим синтезом, а биологической селекцией на базе бактериофагов. Пионером в области фаговой технологии принято считать Джорджа П. Смита (1985): он экспрессировал короткие пептиды на поверхности нитчатого бактериофага M13 путём вставки синтетических последовательностей ДНК, кодирующих пептиды, в ген белка оболочки фага gIIIp. При этом жизнеспособность и инфекционность фага сохранялись. Затем случайные пептидные библиотеки были клонированы в фаги или в белки, слитые с gIIIp, расширены до чрезвычайно больших размеров и подвергнуты скринингу *in vitro* на связывание с целевым белком. Такой подход позволил изучить различные типы белок-белковых взаимодействий, включая взаимодействия между эпитопом и антителом (рандомизированные переменные области), рецептором и лигандом (рандомизация области связывания с рецептором) и протеазой и ингибитором (рандомизация остатков, критически важных для взаимодействия с протеазой). Мэтьюз и Уэллс модифицировали метод, чтобы использовать протеолитическую активность протеаз для идентификации новых соответствующих субстратов [74].

Создание библиотеки субстратов для фагового дисплея включает несколько этапов: в геном фага сначала встраиваются нуклеотидные последовательности, которые кодируют все возможные комбинации аминокислот. Далее рекомбинантные фаги экспрессируют химерные поверхностные белки, содержащие вставленный пептид. В зависимости от системы экспрессии могут применяться две стратегии дисплея. Моновалентная система использует плазмиду с фагово-бактериальным началом репликации (фагмиду), несущую ген, кодирующий белок gIII, который в конечном итоге сливается с субстратом. Для инфицирования клеток *E. coli* рекомбинантными фагмидами используется фаг-хелпер (который вносит дополнительные белки, необходимые для получения функционального фага). В этой системе рекомбинантный gIII конкурирует с белком дикого типа на поверхности фага, поэтому экспонируется либо единственная копия субстрата, либо ни одна. Вторая стратегия - поливалентная система, которая использует фаговый вектор, содержащий весь геном фага. Каждый белок gIII (всего 5) слит с субстратом, поэтому в результате получают только рекомбинантные фаги. Недостатком поливалентной системы по сравнению с моновалентной является то, что для селекции фага необходимо расщепление всех 5 белков слияния, а инфекционность поливалентного фага может быть снижена субстратами. С другой стороны, поливалентная система производит только рекомбинантные фаги, поэтому фон фага дикого типа, который является проблемой для моновалентных систем, устраняется (обычно менее 10% фагов обнаруживают субстрат) [74].

Потенциал метода фагового дисплея был наглядно продемонстрирован в работе по идентификации селективных субстратов для двух гомологичных вирусных протеаз – протеазы иммунодефицита человека (ВИЧ-1) и вируса иммунодефицита кошек (ВИК). В исследовании Бэк и соавторов [75] с помощью независимого отбора были получены уникальные пептидные последовательности, отражающие различия в специфичности этих родственных ферментов. Это

подтвердило высокую разрешающую способность метода в различении даже близкородственных протеаз.

В работе Ратникова Б. и соавторов с помощью метода фагового дисплея авторы охарактеризовали протеазную специфичность нескольких матричных металлопротеиназ [76]. Также этот метод был адаптирован к полуавтоматической платформе, которая позволила получить на порядок больше данных для точного сравнения родственных протеаз и определения их функциональных различий.

Объединение фагового дисплея случайного субстрата из 6 аминокислот с высокопроизводительным секвенированием для всестороннего определения специфичности активного центра сериновой протеазы тромбина и металлопротеазы ADAMTS13 (фермент, который действует как регулятор свертываемости крови, расщепляя фактор Виллебранда (VWF)), было продемонстрировано в исследовании [77]. Для тромбина было идентифицировано более 6700 расщеплённых пептидов, и выявленный мотив хорошо совпадал с ранее описанной специфичностью. В отличие от него, протеаза ADAMTS13 расщепила лишь 96 пептидов из более чем 10 миллионов, и при этом не выявлялось чёткого консенсусного мотива. Однако при использовании модифицированной библиотеки, в которой варьировались аминокислоты в интервале P_3 – P_3' пептида VWF73 (естественного субстрата ADAMTS13), удалось определить 1670 расщеплённых последовательностей. Анализ показал, что фермент предпочитает алифатические аминокислоты в большинстве позиций, особенно Leu в P_3 и объёмные гидрофобные остатки в P_1 и P_1' , при этом Arg допустим в P_2 . Эти данные свидетельствуют, что взаимодействие с экзосайтом играет ключевую роль в обеспечении специфичности.

Однако фаговые дисплеи не лишены недостатков. Они требуют иммобилизации фагов и не подходят для исследования протеаз, расщепляющих белки, участвующие в инфекционности фага или несущие аффинные метки, что может привести к потере функциональности или исключению целых библиотек.

Также в методе фагового дисплея фаги с расщепляемыми пептидами отбираются и многократно амплифицируются в *E. coli*, чтобы обогатить библиотеку подходящих субстратов. Однако после нескольких раундов селекции точность метода может снижаться из-за смещений, связанных с репликацией фага, его инфекционностью и экспрессией белков.

2.3. Методы определения кинетических параметров протеаз

Для полноценной характеристики фермента важна не только его специфичность, но и каталитическая эффективность – способность быстро и селективно гидролизовать данные субстраты. Эти параметры описываются с использованием кинетических характеристик (k_{cat} , K_M , k_{cat}/K_M), для определения которых применяются как классические биохимические методы, так и современные подходы.

2.3.1. Спектрофотометрические методы

Спектрофотометрические методы основаны на измерении изменения оптической плотности (абсорбции) при протекании ферментативной реакции. Классический пример – определение активности сериновой протеазы трипсина с помощью синтетического субстрата этилового эфира N α -бензоил-L-аргинина (BAEE). Гидролиз BAEE трипсином сопровождается увеличением поглощения при 253 нм, что позволяет непрерывно регистрировать скорость реакции по приросту абсорбции. Такой непрерывный спектрофотометрический тест (метод Хаммеля) широко применялся для характеристики трипсина; одна единица активности трипсина даже определена как количество фермента, дающее прирост оптической плотности на 0,001 в минуту при 253 нм в указанных условиях [78]. Спектрофотометрические измерения часто удобны благодаря простоте и скорости проведения анализа. Они способны регистрировать относительные изменения

концентраций субстрата или продукта даже при относительно низких концентрациях, особенно если реакция сопровождается заметным сдвигом поглощения [79].

Следует отметить, что спектрофотометрия охватывает как измерения в ультрафиолетовой области, так и видимые (колориметрические) измерения окрашенных продуктов. Колориметрическими методами обычно называют спектрофотометрические тесты, в которых образуется окрашенный продукт, измеряемый в видимой области спектра. Например, активность трипсина часто определяют по гидролизу N- α -бензоил-L-аргинин-4-нитроанилида (BAPNA) – хромогенного субстрата, при расщеплении которого высвобождается окрашенный продукт p-нитроанилин желтого цвета. Увеличение оптической плотности при 405–410 нм пропорционально количеству высвободившегося p-нитроанилина и позволяет вычислить скорость реакции [80]. Такой метод, впервые описанный Эрлангером с соавт. в 1961 г., до сих пор служит стандартным анализом трипсина. В практическом отношении колориметрические тесты просты и подходят для измерения активности во многих образцах; существуют готовые наборы, где интенсивность желтой окраски p-нитроанилина линейно связана с активностью протеазы [81]. Помимо BAPNA, разработан ряд других хромогенных субстратов: например, азоказеин – это казеин, модифицированный азокрасителем, гидролиз которого любыми протеазами высвобождает окрашенные фрагменты. После осаждения нерасщепленного белка трихлоруксусной кислотой, в фильтрате остаются растворимые азопептиды, и их оптическую плотность измеряют при ~440 нм.

Таким образом, спектрофотометрические и колориметрические методы предоставляют относительно простой способ количественно измерить кинетику ферментативного расщепления по изменению оптической плотности, если субстрат или продукт обладают характерным поглощением.

2.3.2. Флуориметрические методы

Флуориметрические (флуоресцентные) методы отличаются высокой чувствительностью и потому широко используются при изучении кинетики протеаз [79].

В этом подходе применяют флуорогенные субстраты – соединения, дающие флуоресцентный сигнал при расщеплении. Как правило, пептидный субстрат мечен флуорофором, который либо освобождается, либо меняет свои свойства после протеолиза. Один из наиболее популярных флуорогенных меток – описанный ранее в контексте PS-SCL 7-амино-4-метилкумарин (АМС). Пептиды с С-концевым АМС слабо флуоресцируют в связанном состоянии, однако после гидролиза высвобождается свободный АМС, дающий яркую флуоресценцию (~440 нм). Определяя скорость нарастания флуоресценции, можно напрямую измерять начальную скорость реакции и, изменяя концентрацию субстрата, определять кинетические параметры фермента. Метод с АМС-субстратами хорошо зарекомендовал себя как для индивидуальных анализов, так и для профилирования специфичности протеаз.

Так, с помощью наборов флуорогенных пептидов на основе АМС и АСС были определены оптимальные сайты расщепления для тромбина, эластазы, папаина, крузаина и многих других протеаз. Две тромбин-чувствительные последовательности с АСС или АМС, показали сопоставимые кинетические константы в качестве субстратов тромбина [70].

Другой мощный подход – использование пар “донор–акцептор” в формате FRET (резонансного переноса энергии флуоресценции). В FRET-субстратах пептидный цепочка метится двумя группами: флуорофором-донором и гасителем (акцептором) флуоресценции. Пока пептид цел, донор и акцептор находятся близко, и флуоресценция донорного красителя подавлена (“затухает”). Гидролиз пептида ферментом разъединяет донор и акцептор, в результате чего флуоресцентный сигнал донора резко возрастает [82].

Например, Матаёши и соавторы синтезировали субстраты для протеазы ВИЧ-1 с органическим флуоресцентным донором, 5-[(2-аминоэтил)амино]нафталин-1-сульфоновой кислотой (EDANS), и гасителем флуоресценции, 4-(4-диметиламинофенилазо)бензойной кислотой (DABCYL), на С- и N-концах пептидных субстратов. Гидролиз пептидного субстрата в Tug-Pro увеличивал флуоресценцию EDANS, и активность протеазы контролировалась по спектру её испускания. Этот подход лежит в основе большинства протеазных анализов, основанных на FRET [83]. FRET-методы позволяют проводить непрерывный мониторинг активности ферментов и весьма чувствительны (пределы обнаружения пикомолярные). В современных вариантах вместо органических красителей применяют наноматериалы – например, квантовые точки или наночастицы золота – как акцепторы или доноры, что повышает стабильность сигнала [82].

Флуориметрические методики хоть и очень чувствительны, но имеют и ограничения. Органические флуорофоры могут подвергаться фотообесцвечиванию (фотостабильность ограничена), а посторонние компоненты в образце способны флуоресцировать сами по себе или гасить сигнал, что требует контролировать чистоту образцов. Тем не менее, благодаря высокому отношению сигнал/шум флуоресцентные субстраты остаются основным инструментом при измерении кинетики протеаз, особенно когда концентрации фермента малы или объемы образцов ограничены.

2.3.3. Радиометрические методы

Радиометрические методы основаны на использовании радиоактивно меченых субстратов и измерении радиоактивности продуктов реакции. Этот подход исторически сыграл большую роль в точном определении кинетических констант ферментов благодаря исключительной чувствительности. В радиометрическом анализе субстрат несет метку, например ^3H или ^{14}C , в

определенном месте молекулы. В ходе ферментативной реакции образуется меченый продукт, который затем количественно отделяют от субстрата (например, хроматографией или осаждением) и измеряют его радиоактивность. Ключевым требованием является наличие простого и быстрого метода разделения субстрата и продукта, а также подходящего меченого субстрата известной специфической активности. Хайданд и соавторы разработали быстрый высокопроизводительный радиометрический анализ протеазы ВИЧ-1 с использованием ионообменной хроматографии, проводимой в 96-луночных фильтрационных планшетах [84]. Анализ позволяет проводить мониторинг активности протеазы ВИЧ-1 на радиоактивно меченной форме гептапептидного субстрата, основанного на сайте расщепления, который был обнаружен в вирусном полипротеиновом субстрате. Специфическое расщепление этого незаряженного гептапептидного субстрата протеазой ВИЧ-1 приводит к высвобождению анионного продукта, который задерживается на миниколонках с анионообменной смолой. Активность протеазы определяется по выделению радиоактивно меченого продукта после элюирования муравьиной кислотой.

Радиометрические измерения обладают уникально высокой чувствительностью (пределы обнаружения очень низкие (фон составляет ~20 распадов в минуту, что часто находится на уровне менее фемтограмма, в зависимости от эффективности маркировки) и специфичностью, поскольку детектируется только метка интересующего атома. Кроме того, на них не влияют оптические свойства образца – например, цветные или мутные растворы, затрудняющие колориметрию. Однако данный подход требует работы с радионуклидами, что связано с радиационной безопасностью и специальными разрешениями. При многостадийном расщеплении метка не позволяет различить, какая из нескольких возможных пептидных связей гидролизована, потому что все меченые фрагменты дадут суммарный сигнал. Кроме того, синтез меченых субстратов является затратным и трудоёмким процессом. Поэтому, несмотря на превосходную чувствительность, радиометрические методы постепенно уступили

место флуоресцентным и другим технологиям в рутинных кинетических исследованиях, оставаясь нишевым инструментом для особо точных измерений.

2.3.4. Электрохимические методы

Электрохимические подходы к измерению активности протеаз – относительно новое и успешно развивающееся направление аналитических методов. Суть обычно заключается в преобразовании акта ферментативного гидролиза в измеримый электрический сигнал, например, ток. Одним из стратегических преимуществ электрохимических датчиков является возможность прямого (без оптических меток) обнаружения событий на поверхности электрода с высокой чувствительностью. Например, разработаны биосенсоры, где на поверхности электрода иммобилизован пептид – специфический субстрат протеазы, содержащий электроактивную метку (ферроцен). В отсутствие фермента метка находится близко к поверхности электрода и дает сильный вольтамперометрический сигнал. При добавлении протеазы пептидный субстрат расщепляется, метка отделяется от поверхности, и зарегистрированный ток ферроцена уменьшается со временем. По кинетике затухания тока можно определить скорость катализа; такой метод был продемонстрирован для цистеиновой протеазы катепсина В, позволив определять концентрации фермента вплоть до 0,3 нМ [85].

В другом подходе используют комбинацию пептидного слоя и наночастиц. Так, для главной протеазы SARS-CoV-2 M_{pro} предложена система, в которой пептид-субстрат закреплен на электроде и способен связывать коллоидные частицы золота. В интактном состоянии пептид способствует сборке наночастиц на поверхности, что резко уменьшает электрохимическое сопротивление за счет высокой проводимости золота. Если же M_{pro} расщепляет пептид, частицы больше не удерживаются, и импеданс (электронное сопротивление слоя) возрастает – этот прирост количественно связан с активностью протеазы. Метод электрохимической

импедансной спектроскопии, реализованный в таком сенсоре, позволил достичь предела обнаружения $\sim 0,1$ пМ для вирусной протеазы [86].

Электрохимические измерения имеют достоинства в виде высокой чувствительности и потенциала миниатюризации (вплоть до микрочипов и одноразовых тест-полосок). Их легко мультиплексировать – например, разместив разные пептидные последовательности на массиве микроэлектродов для одновременного профилирования нескольких ферментов. Тем не менее, такие методы требуют тщательно оптимизированной поверхности электрода и могут страдать от неспецифической адсорбции или биообрастания, что влияет на сигнал.

В заключение следует рассмотреть ограничения описанных подходов. Спектрофотометрические/колориметрические методы просты, но не всегда достаточно чувствительны. Измерение поглощения требует, чтобы субстрат или продукт давали заметный оптический отклик. В случае слабого сигнала или малого коэффициента экстинкции приходится работать с высокими концентрациями субстрата, что может исказить кинетические параметры (например, при гидролизе BAPNA высокие концентрации субстрата тормозят трипсин) [87].

Кроме того, компонентам реальных биологических образцов (сыворотка, клеточные лизаты) присуща собственная мутность или окраска, которая мешает оптическим измерениям. Цветные или мутные растворы могут сильно поглощать свет и влиять на точность колориметрических определений. Требуется проводить контроль фонового поглощения и очищать пробы, либо прибегать к дифференциальным методам. Флуориметрические методы гораздо чувствительнее, однако подвержены влиянию автофлуоресценции компонентов образца (некоторые соединения, витамины, аминокислоты сами флуоресцируют) и наличию примесей. Требуется использовать фильтры и подбирать длины волн возбуждения/излучения, а также иногда применять внутренние контрольные пробы с ингибитором для учета несвязанных источников сигнала. Органические флуорофоры могут выгорать при длительном облучении, что накладывает

ограничение на время считывания или интенсивность источника света. Тем не менее, эти проблемы обычно решаются подбором условий эксперимента, и флуоресцентные анализы остаются одними из наиболее надежных. Радиометрические методы практически свободны от помех со стороны матрицы, но обладают своими существенными недостатками. Работа с радионуклидами требует специальных разрешений и строгого соблюдения техники безопасности; устранение радиоактивных отходов – дорогостоящее и регламентированное мероприятие. Кроме того, необходимо иметь радиометрическое оборудование. В кинетических экспериментах радиометрия, как правило, не позволяет проводить непрерывные измерения – обычно фиксируют точки, останавливая реакцию и отделяя продукты, что замедляет процесс сбора данных по сравнению с прямыми оптическими методами. Электрохимические методы требуют сложной оптимизации поверхности. Сенсоры могут давать дрейф сигналов из-за необратимого адсорбирования компонентов образца на электроде.

Важно подчеркнуть, что выбор подхода определяется поставленной задачей. Спектрофотометрия и колориметрия хорошо подходят для высоких концентраций фермента и быстрого скрининга, флуориметрия – для чувствительных измерений и непрерывного мониторинга, радиометрия – для случаев, когда оптические методы невозможны или требуется предельная точность, а электрохимические сенсоры открывают возможности миниатюризации и *in situ* измерений (например, в аналитических биочипах).

2.4. MALDI и LC-MS/MS как инструмент анализа специфичности и кинетических параметров протеаз

Определение специфичности протеаз долгое время оставалось трудоёмким экспериментальным процессом, что ограничивало исследования анализом лишь небольшого числа субстратов [88]. Использование комбинаторных библиотек позволило частично преодолеть это ограничение за счёт применения больших

наборов синтетических пептидов, обработанных протеазой [70]. Однако интерпретация продуктов расщепления в этих библиотеках нередко требует трудоёмких методов детекции, таких как флуоресцентный анализ или секвенирование.

В этом отношении масс-спектрометрия (МС) представляет собой быстрый и высокоэффективный подход к идентификации широкого спектра субстратов из пептидных библиотек и не только [89]. Родригес и соавторы показали, что на основе МС-анализа больших массивов спектральных данных можно с высокой точностью определить специфичность исследуемого фермента [90]. Важным преимуществом данного подхода является его универсальность: он может применяться даже к данным, полученным для иных целей, и не требует использования дорогостоящих методов маркировки.

Масс-спектрометрические подходы предлагают ряд ключевых преимуществ по сравнению с традиционными биохимическими методами при изучении протеаз.

1) Благодаря высокой чувствительности (позволяет детектировать вещества в фемтомолярных (10^{-15} М) концентрациях) и селективности методов МС возможно обнаруживать даже малые количества продуктов протеолиза с высокой точностью [91];

2) МС-методы обладают высокой скоростью получения данных и возможностью параллельного анализа множества субстратов за один эксперимент. В классических колориметрических или флуориметрических тестах для определения специфичности приходится последовательно исследовать множество отдельных субстратов (например, ~20 экспериментов, чтобы проверить предпочтение каждой аминокислоты в определённой позиции). В случае использования МС весь набор разнообразных субстратов может одновременно подвергаться действию фермента, а продукты их расщепления идентифицируются в одном спектрометрическом анализе, что радикально повышает производительность эксперимента;

3) МС позволяет напрямую идентифицировать продукты протеолиза и установить аминокислотную последовательность в точках разрыва пептидной связи. Анализ масс-спектров дает возможность точно определить, между какими аминокислотными остатками протеаза разрежала субстрат. В то время как оптические методы фиксируют лишь суммарный эффект (например, появление флуоресценции при расщеплении меченого субстрата), с помощью МС можно напрямую определять специфичность фермента через предпочтительные сайты и мотивы расщепления [92];

4) МС-анализ не требует присоединения к субстратам хромофорных или флуоресцентных групп, которые могут мешать нормальному взаимодействию фермента с белком-мишенью. Например, известно, что размещение крупного флуорофора непосредственно рядом с целевой пептидной связью способно исказить фермент-субстратное распознавание;

5) В МС-методах можно использовать либо немеченые синтетические пептиды, либо вообще нативные белковые субстраты, что способствует более физиологичным условиям и достоверному определению специфичности протеазы;

6) МС-методы позволяют исследовать активность в таких сложных образцах, как компоненты крови или клеточные лизаты, без трудоемкой очистки фермента. Добавляя стандартизированную смесь субстратных пептидов в такой сложный образец, можно по появлению специфических пептидов-продуктов судить о наличии и характеристиках активности целевых протеаз. Благодаря тому, что синтетические субстраты подбираются уникальными и не совпадают с естественными белками, их продукты легко отличимы на фоне эндогенных пептидов в пробе.

2.4.1. История создания масс-спектрометрии, принцип устройства

Исторически масс-спектрометрия как аналитический метод возникла в начале XX века и первоначально применялась преимущественно в физике и химии [93]. Первая реализация масс-спектрометра была выполнена Джозефом Джоном Томсоном, который в 1912 году использовал метод для разделения изотопов неона по массе. Впоследствии Фрэнсис Эстон сконструировал более совершенный масс-спектрограф и получил Нобелевскую премию по химии в 1922 году за открытие изотопов с помощью масс-спектрометрии. Однако в течение нескольких десятилетий масс-спектрометрия оставалась методом преимущественно для низкомолекулярных соединений, поскольку существовавшие способы ионизации (например, электронный удар) были слишком «жесткими» и приводили к фрагментации (био-)макромолекул, таких как белки и нуклеиновые кислоты.

Ситуация принципиально изменилась в конце 1980-х годов с разработкой методов «мягкой» ионизации. В 1984 году в СССР был разработан метод ЭРИАД (экстракция растворённых ионов при атмосферном давлении), предложенный М. Л. Александровым и Л. Н. Галль [94]. Метод позволял получать масс-спектры трудноиспаряющихся биоорганических соединений с регулируемой степенью фрагментации. Однако метод не получил широкой международной известности из-за ограниченного доступа к советским публикациям в то время. В это же время в США Джон Б. Фенн разработал метод электроспрейной ионизации (ESI, Electrospray Ionization), в котором ионизация происходила в результате распыления раствора под действием высокого напряжения. Использование ESI в сочетании с предварительным хроматографическим разделением обеспечило высокую чувствительность и широкую применимость в анализе биомолекул.

В 1985–1987 годах Михаэль Карас и Франц Хилленкамп разработали метод матрично-активируемой лазерной десорбции/ионизации (MALDI, Matrix-Assisted Laser Desorption/Ionization). В своей работе 1986 года они показали, что добавление матрицы – низкомолекулярного соединения, поглощающего лазерное излучение, –

позволяет существенно снизить энергию, необходимую для десорбции и ионизации крупных органических молекул, включая пептиды. В 1987 году Коити Танака представил альтернативный подход к мягкой ионизации, используя для этого смесь глицерина и мелкодисперсного кобальта в качестве матрицы, а также азотный лазер с длиной волны 337 нм. Этот метод позволил ему ионизировать белок карбоксипептидазу-А. Несмотря на то, что метод Танаки впоследствии был вытеснен более воспроизводимыми и чувствительными вариантами MALDI, его вклад был высоко оценён. В 2002 году Джон Фенн и Коити Танака были удостоены Нобелевской премии по химии «за разработку методов идентификации и структурного анализа биологических макромолекул».

Принцип устройства масс-спектрометра основан на ионизации молекул аналита с последующим определением их отношения массы к заряду. Любой масс-спектрометр включает три ключевых компонента: источник ионов, масс-анализатор, детектор. На первом этапе исследуемое вещество ионизируется в источнике ионов, после чего ионы поступают в вакуумную систему, где проходят через масс-анализатор (или в случае MALDI, ионизируются уже в вакууме). В зависимости от соотношения массы к заряду ионы разделяются по характеристикам, зависящим от типа анализатора (времени пролёта, траектории или частоте колебаний), и в конечном итоге достигают детектора, регистрирующего масс-спектры анализируемых веществ.

Существует множество методов ионизации, в том числе термические (например, испарение в газовой хроматографии), химические (например, CI – химическая ионизация), жёсткие (электронный удар) и мягкие (MALDI, ESI). Однако в протеомике наибольшее распространение получили два мягких метода ионизации, позволяющих анализировать высокомолекулярные и термолабильные биомолекулы: ESI и MALDI.

Принцип работы ESI основан на создании аэрозоля заряженных капель из раствора аналита [95–97]. Образец, растворённый в летучем растворителе (обычно

смесь воды и органического компонента, например, ацетонитрила или метанола), подаётся через капилляр, к которому приложено высокое напряжение (1–5 кВ). На выходе из капилляра образуется т.н. конус Тейлора, из вершины которого вылетают мелкие заряженные капли. Под действием внешнего электрического поля и испарения растворителя, обычно под действием высокой температуры (150–250° С), капли уменьшаются в размере. Когда сила кулоновского отталкивания становится равной силе поверхностного натяжения капли, происходит её распад (так называемое кулоновское расщепление), в результате которого образуются более мелкие заряженные капли, продолжающие испаряться. Этот процесс повторяется последовательно, приводя к образованию капель с всё большей плотностью заряда на единицу поверхности. При достижении критической плотности заряда ($\sim 10^8$ В/см³), начинается испарение ионов (ion evaporation) – прямое высвобождение ионов из поверхности капли в газовую фазу. Эти ионы затем поступают в вакуумную камеру и направляются в масс-анализатор.

Метод электроспрейной ионизации открыл новые возможности в анализе сложных биологических макромолекул благодаря своей мягкости и совместимости с жидкостной хроматографией. Ионизация в этом методе происходит при атмосферном давлении и комнатной температуре, что позволяет сохранять нативную структуру термоллабильных соединений – таких как пептиды, белки, нуклеотиды и олигосахариды. Одним из ключевых преимуществ ESI является образование многозарядных ионов, что существенно снижает значение отношения массы к заряду (m/z) даже для крупных молекул. Это позволяет использовать анализаторы с ограниченным m/z -диапазоном для детектирования высокомолекулярных соединений. Кроме того, возможность сопряжения ESI с высокоэффективной жидкостной хроматографией (LC) делает этот метод особенно ценным при анализе сложных биологических смесей, таких как клеточные лизаты или фракционированные белковые лизаты [54,98], поскольку дополнительное разделение перед анализом способствует детекции большего количества компонентов, а, следовательно, увеличению чувствительности. Кроме того,

благодаря высокой стабильности ионного тока, ESI стал основой для количественных и полуколичественных методов оценки белков и пептидов, не без использования изотопных меток (таких как безметочное количественное определение, LFQ, label-free quantification).

Тем не менее, у метода ESI есть ряд ограничений. Он требует высокой чистоты растворителей и буферов, поскольку присутствие солей, детергентов или фосфатных компонентов может конкурировать с аналитом за ионизацию. Это приводит к существенному снижению эффективности ионизации и, следовательно, уменьшению интенсивности сигнала. Кроме того, из-за своей чувствительности к матричным эффектам, ESI может давать нестабильный сигнал в условиях сложного химического фона, к примеру при одновременной ионизации большого количества различных молекул аналита одновременно, что требует тщательной оптимизации условий разделения и предварительной подготовки образца. Еще одной особенностью является необходимость использования летучих буферов (например, формиат аммония), что ограничивает выбор условий при работе с биохимическими системами.

Метод MALDI основан на облучении исследуемого образца лазерным импульсом после предварительного смешивания с матрицей – низкомолекулярным органическим соединением, поглощающим энергию лазера [99,100]. Аналит растворяется вместе с избытком матрицы и наносится на металлическую мишень (target), где происходит сокристаллизация аналита с молекулами матрицы. При воздействии лазерного импульса – например, азотного лазера ($\lambda = 337$ нм) или лазера на основе иттриево-алюминиевого граната, легированного неодимом (Nd:YAG, $\lambda = 355$ нм) – матрица поглощает энергию, испаряется, образуя факел (высокотемпературную плазму), которое состоит из молекул аналита и матрицы.

Предполагается, что матрица способствует первичной ионизации аналита, передавая заряд в процессе испарения. Однако основная ионизация происходит за счёт вторичных столкновений, включая перенос протона между возбужденными

молекулами матрицы и аналита (proton transfer reactions) – это и называется вторичной ионизацией. В отличие от метода ESI, при MALDI преимущественно формируются однозарядные ионы, как положительно, так и отрицательно заряженные.

Метод MALDI идеально подходит для быстрой идентификации белков и пептидов в чистом виде или после простейшей подготовки, а также позволяет эффективно анализировать биомолекулы массой до сотен кДа, не вызывая их разрушения. Одним из ключевых достоинств MALDI является его высокая устойчивость к загрязнению, а также способность работать с образцами, содержащими соли и сложные матрицы. Кроме того, ионизация в основном приводит к формированию ионов с одним зарядом, что может быть полезно, к примеру, при изотопном анализе [99]. MALDI-TOF используется для быстрой идентификации белков, методом «отпечатков масс пептидов» (peptide mass fingerprinting) и MALDI-имиджинга биологических тканей.

Однако MALDI имеет и свои ограничения. Прежде всего, из-за характера подготовки образца (сокристаллизация с матрицей) метод невозможно использовать в сочетании с жидкостной хроматографией. Следовательно, ограничивает возможности анализа больших смесей с перекрывающимися массами, и не позволяет проводить анализ сложных смесей. Кроме того, при анализе высокомолекулярных соединений, особенно белков, в спектре часто наблюдается усиление сигналов от загрязняющих компонентов матрицы или фонового шума, что может затруднить детектирование молекул в низкой концентрации. Наконец, подбор подходящей матрицы и условий кристаллизации требует эмпирической оптимизации, особенно при анализе новых классов соединений. В таблице 1 представлены основные соединения, используемые в качестве матрицы в MALDI анализе, и типы молекул, которые могут быть эффективно проанализированы с их использованием.

Таблица 1. Наиболее часто используемые матрицы, применяемые для анализа в масс-спектрометрах с типом ионизации MALDI.

Матрица	Сокращение	Объекты анализа
2,5-дигидрокси-бензойная кислота	DHB	Белки, пептиды, углеводороды, липиды, полярные синтетические полимеры
Синапиновая кислота	SA	Белки, пептиды, гликаны, олигосахариды
α -циано-4-гидроксикоричная кислота	HCCA, CHCA	Белки, пептиды с низкой концентрацией
Дитранол	DIT	Синтетические полимеры
3-гидрокси пиколиновая кислота	3-NPA	Нуклеиновые кислоты

После ионизации молекулы аналита направляются в масс-анализатор, где происходит их разделение по отношению массы к заряду (m/z). Существует несколько типов масс-анализаторов, различающихся по принципу действия, разрешающей способности, точности и диапазону масс.

Основные типы анализаторов, используемые в протеомике, это квадрупольный анализатор (Quadrupole), времяпролетный анализатор (Time-of-Flight, TOF), ионная ловушка (Ion Trap), ионно-циклотронный резонанс (Fourier Transform Ion Cyclotron Resonance, FT-ICR) и орбитальная ловушка (Orbitrap).

Квадрупольный анализатор использует переменное электрическое поле для стабилизации или дестабилизации траекторий ионов. Он позволяет селективно пропускать ионы заданного m/z и широко применяется в tandemных схемах (QqQ, Q-TOF), о которых будет сказано далее, для количественного анализа и селективного фрагментирования (MS/MS).

Времяпролетный анализатор измеряет время, за которое ион достигает детектора. Вакуум и одинаковый импульс ускорения позволяют судить о массе по скорости пролета.

Он основан на простом принципе: квадрат скорости ионов, приобретенной в ускоряющем электрическом поле, обратно пропорционален их массе. Каждой массе, а точнее, отношению массы к заряду соответствует свое время пролета, исходя из измеренного времени пролета вычисляется масса иона по формуле (1)[101]:

$$eU * z = \frac{-mv^2}{2} \quad (1)$$

где m – масса иона; v – скорость полета иона; U – напряжение в электрическом поле, z – число зарядов; e – единичный заряд электрона.

TOF-анализаторы характеризуются высокой скоростью, широким диапазоном масс и хорошим разрешением, особенно при использовании в режиме отражения (reflectron).

Ионные ловушки (3D и линейные) позволяют удерживать ионы во времени и проводить последовательные циклы фрагментации (MS^n). Они особенно полезны при исследовании сложных смесей, однако имеют ограниченное разрешение по сравнению с Orbitrap или TOF. Ионно-циклотронный резонанс и орбитальная ловушка обеспечивают ультравысокое разрешение и точность массы. В этих системах ионы удерживаются во вращающемся или орбитальном электромагнитном поле, и их частота движения регистрируется и преобразуется в массу с использованием преобразования Фурье.

Метод ионизации MALDI, создающий преимущественно однозарядные ионы, чаще всего используется в сочетании с TOF- и TOF/TOF-анализаторами, что обеспечивает высокую скорость анализа и широкий диапазон определяемых масс. Ионизация методом электрораспыления (ESI), напротив, формирует многозарядные ионы, что делает её совместимой с высокоразрешающими анализаторами (Orbitrap, FT-ICR), а также с tandemными конфигурациями (QqQ,

Q-TOF, Ion Trap), обеспечивающими эффективное фрагментирование. В таблице 2 представлены основные типы масс-анализаторов.

Таблица 2. Основные типы анализаторов, принцип их действия и область применения.

Тип анализатора	Принцип действия	Совместимость с ионизацией	Особенности и область применения
Квадруполь (Q)	Селективное пропускание ионов по m/z с помощью переменного электрического поля	ESI	Простота, селективность; широко используется в тандемных конфигурациях (QqQ, Q-TOF) для количественного анализа и MS/MS
Ионная ловушка (Ion Trap)	Удержание ионов во времени в электрическом поле и их последовательное сканирование	ESI	Возможность многократного фрагментирования (MS^n), компактность, умеренное разрешение
TOF (Time-of-Flight)	Измерение времени пролета ионов до детектора	MALDI, ESI	Высокая скорость, широкий диапазон масс, особенно эффективен с MALDI и в режиме TOF/TOF
FT-ICR	Удержание ионов в магнитном поле и измерение их циклотронной частоты	ESI	Ультравысокое разрешение и точность массы; громоздкость, высокая стоимость
Orbitrap	Удержание ионов в орбитальном электрическом поле и регистрация колебаний	ESI	Высокая чувствительность и точность; применяется в «shotgun»-протеомике и анализе сложных смесей

Детекторы, регистрирующие ионы после прохождения масс-анализатора, могут быть различных типов. Наиболее распространён электронный умножитель – устройство, усиливающее сигнал от ионов. В TOF-системах применяются

мультиканальные пластины (MCP), обеспечивающие высокую временную точность. В FT-ICR и Orbitrap ионы не детектируются напрямую, а индуцируют слабые токи, регистрируемые индукционными пластинами (image current detection), после чего сигнал обрабатывается математически.

Все современные масс-спектрометры, обладают возможностью проведения тандемной масс-спектрометрии (MS/MS), которая является ключевым инструментом для идентификации аминокислотной последовательности пептидов. В отличие от обычной масс-спектрометрии (MS, или MS1), определяющей молекулярную массу ионов, тандемная МС включает три последовательных этапа: сначала детектируются ионы-предшественники (прекурсоры), после чего выбранный ион изолируется, фрагментируется, и полученные фрагментные ионы анализируются в масс-анализаторе. Фрагментация, как правило, происходит по пептидной связи, поскольку она является наиболее уязвимой в структуре пептида. Типы фрагментов зависят от ряда факторов, включая первичную структуру, зарядовое состояние и количество внутренней энергии [102,103]. При этом фрагментные ионы должны сохранять хотя бы один заряд, чтобы быть зафиксированными детектором. Классификация таких ионов зависит от участка, на котором удерживается заряд: фрагменты с зарядом на N-конце обозначаются как a, b или c, тогда как с зарядом на C-конце – как x, y или z (рисунок 2). Полученный спектр фрагментации позволяет реконструировать последовательность аминокислот в пептиде.

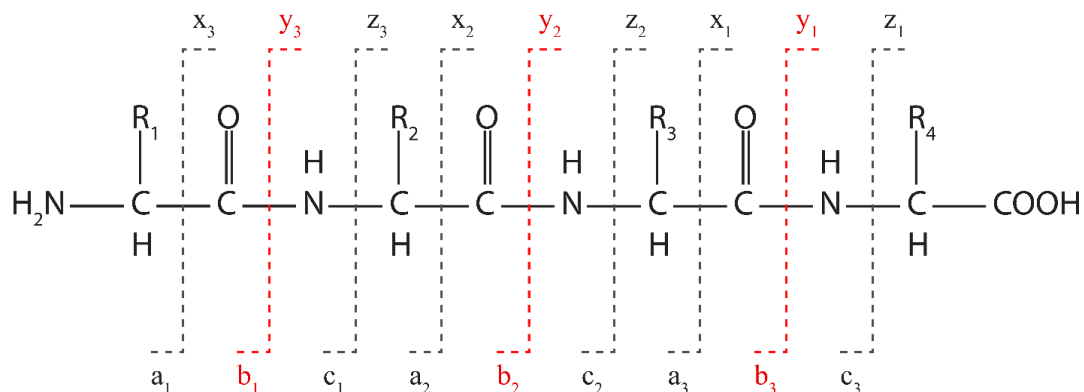


Рисунок 2. Схематическое представление фрагментации пептида в тандемной масс-спектрометрии. Показаны основные типы ионов, образующихся при разрыве пептидной цепи: ионы серии a, b и c формируются при сохранении заряда на N-концевом фрагменте, а ионы серии x, y и z – на C-концевом.

Основными механизмами фрагментации в масс-спектрометрии являются CID, HCD и ETD. Наиболее распространённый из них – CID (collision-induced dissociation, диссоциация, вызванная соударением). При CID ионы-предшественники сталкиваются с молекулами инертного газа (например, аргона или азота), в результате чего их кинетическая энергия преобразуется во внутреннюю. По достижении порогового уровня внутренняя энергия вызывает разрыв наиболее слабых связей, преимущественно пептидных. В результате образуются фрагментные ионы типа b (с зарядом на N-конце) и y (с зарядом на C-конце).

Метод HCD (higher-energy collisional dissociation), по механизму аналогичный CID, реализуется преимущественно в гибридных масс-спектрометрах, таких как квадрупольно-времяпролётные (Q-TOF), квадрупольных [104], соединённых с ионными ловушками и т.д. Он отличается более высокой энергией столкновения, что позволяет получать более полные спектры фрагментов с высоким разрешением. Как и в случае CID, основными продуктами фрагментации являются b- и y-ионы, однако HCD обеспечивает более полное покрытие пептидной последовательности.

ETD (electron transfer dissociation) применяется преимущественно для фрагментации многозарядных катионов. При этом ион-предшественник взаимодействует с донором электрона (обычно анионом), в результате чего происходит фрагментация по другому механизму, отличному от CID. Образуются фрагменты c- и z•- типов, сохраняющие заряд соответственно на N- и C-концах пептидной связи. ETD особенно эффективен для анализа посттрансляционно модифицированных пептидов, так как позволяет сохранить лабильные модификации, включая фосфатные группы.

Все вещества в масс-спектре представлены в виде сигналов, отражающих ионы с определённым соотношением массы к заряду (m/z). Для корректного анализа масс-спектрометрических данных важно понимать, как формируются

сигналы в масс-спектре. В спектрах, полученных на приборах высокого разрешения, каждое органическое соединение представлено не одним пиком, а группой сигналов, образующих изотопный кластер. Образование изотопного кластера обусловлено присутствием стабильных природных изотопов химических элементов – атомов с одинаковым числом протонов, но разным числом нейтронов, отличающихся, соответственно, по массе приблизительно на 1 дальтон. Наиболее распространённые элементы в белках – углерод (C), водород (H), азот (N), кислород (O) и сера (S) – имеют стабильные изотопы, представленные на таблице 3.

Таблица 3. Стабильные изотопы основных элементов, входящих в состав белков, и их среднее природное содержание.

Атом	Изотоп	Процентное содержание
Водород	^1H	99,99%
	^2H	0,01%
Углерод	^{12}C	98,9%
	^{13}C	1,1%
Азот	^{14}N	99,64%
	^{15}N	0,36%
Кислород	^{16}O	99,76%
	^{17}O	0,04%
	^{18}O	0,2%
Сера	^{32}S	95%
	^{33}S	0,76%
	^{34}S	4,22%
	^{36}S	0,02%

Поскольку белки содержат указанные элементы в разных пропорциях, их вклад в изотопные кластеры различен. На основе собственных расчётов, выполненных по полному набору белков из базы Swiss-Prot (база данных от 23 апреля 2025 года [105]), на долю атомов водорода приходится около 52,7% всех атомов, углерода – 26,5%, кислорода – 13,3%, азота – 7,3%, серы – около 0,2%. Наибольший вклад в формирование изотопного кластера вносят атомы углерода за счёт высокой распространённости и наличия стабильного изотопа ^{13}C с долей 1,1%.

Вклад других элементов, таких как серы и азота, обычно ниже ввиду меньшей массовой доли в белках или более низкой доли тяжёлых изотопов.

2.4.2. Методы обработки масс-спектрометрических данных для оценки специфичности протеаз

С развитием масс-спектрометрии стало возможным высокоточное измерение молекулярной массы. Однако, из-за технических особенностей устройства масс-спектрометров, с ростом массы аналита возрастают и погрешности измерений, что ограничивает применение метода для прямой идентификации крупных макромолекул, таких как белки. В связи с этим в протеомике получили распространение косвенные подходы, основанные на идентификации белков по их характеристическим признакам – в частности, по пептидам, уникальным для каждого белка. В современной протеомике используется несколько стратегий: протеомика «снизу вверх» (bottom-up), протеомика «сверху вниз» (top-down) и протеомика «middle-down» [106]. Наиболее широкое распространение получила протеомика «снизу вверх», при которой белки сначала подвергаются гидролизу (чаще всего трипсином), после чего полученная пептидная смесь (гидролизат) анализируется методом жидкостной хроматографии, сопряжённой с tandemной масс-спектрометрией (LC-MS/MS). Полученная смесь пептидов, как правило, находится в оптимальном для масс-спектрометрического детектирования диапазоне масс от 500 до 2500 Да. Жидкостная хроматография обеспечивает предварительное разделение смеси пептидов по времени выхода, уменьшая сложность смеси, поступающей в источник электрораспыления. Это позволяет снизить супрессию ионного сигнала, увеличить число распознаваемых пептидов и, как следствие, расширить покрытие белков.

Важно отметить, что MALDI также может быть использован для анализа продуктов протеолиза, особенно при сравнительно простой белковой смеси. В сочетании с TOF-анализатором MALDI-MS позволяет быстро получить

характерный спектр массы пептидов, сопоставимый с теоретическими массами фрагментов белков из базы данных. Такой метод получил название метод пептидного картирования (peptide mass fingerprinting (PMF)). Однако MALDI имеет существенное ограничение: такой метод не может быть использован для анализа сложных белковых смесей, в том числе продуктов расщепления нескольких белков одновременно, из-за ограниченного динамического диапазона, чувствительности и невозможности подключения к источнику ионов ВЭЖХ, обеспечивающего предварительное разделение. Поэтому MALDI используется преимущественно для анализа небольшого количества белков. Для анализа сложных смесей, состоящих из нескольких белков, часто их предварительно подвергают разделению методами электрофореза или хроматографии [107,108].

Для интерпретации полученных спектров и идентификации белков используются специализированные поисковые системы, такие как Mascot, Sequest, MaxQuant, PEAKS, X! Tandem и другие [109–112]. Принцип их работы основан на сравнении экспериментальных MS/MS спектров с теоретически предсказанными спектрами, сформированными на основе известных последовательностей белков и генов из баз данных, например UniProt (Swiss-Prot) или NCBI nr. Каждый белок из базы данных виртуально расщепляется *in silico* по аминокислотам, специфичным для выбранной протеазы (например, по аргинину и лизину для трипсина). Таким образом формируется набор теоретических пептидов, масса которых сравнивается с экспериментальными значениями в MS-спектре. Для каждого совпадающего пептида далее генерируется теоретический MS/MS-спектр фрагментов (например, ионов типа b, y и т.д.), который сопоставляется с реальным тандемным спектром. Качество совпадения между теоретическим и экспериментальным спектрами оценивается с использованием вероятностных и скоринговых критериев, таких как ion score (Mascot), posterior error probability (MaxQuant), PeptideProphet score, FDR (false discovery rate) и других. Это позволяет ранжировать кандидатные пептиды и определить с высокой вероятностью исходный белок.

Одним из первых алгоритмов оценки совпадений, реализованных в поисковых системах, стал MOWSE (Molecular Weight Search), лежащий в основе метода Peptide Mass Fingerprinting (PMF) [113]. При этом способе идентификации экспериментально измеренные массы пептидов, полученных в результате ферментативного расщепления, сопоставляются с теоретическими значениями, рассчитанными для белков из базы данных. Алгоритм MOWSE учитывает частоту встречаемости пептидов в базе данных [114]. Степень достоверности идентификации выражается через MOWSE Score, который зависит от числа совпавших фрагментов, их статистической значимости и молекулярной массы предполагаемого белка. Этот параметр рассчитывается по формуле [115] (2):

$$Score = \frac{50000}{(P_n * H)} \quad (2)$$

где P_n – произведение оценок частоты распределения для n числа совпадающих фрагментов из исходной базы данных, H – молекулярная масса совпавшего белка.

Поисковые алгоритмы позволяют задавать условия ферментативного расщепления – либо указывать специфичность конкретного фермента с известной специфичностью, либо использовать неспецифический (unspecific) или полуспецифический (semi-specific) режим поиска. В режиме неспецифического поиска, в отличие от поиска с заданной ферментативной специфичностью, виртуальный гидролиз каждого белка из базы данных осуществляется по всем возможным пептидным связям, то есть между любой парой аминокислот. Это приводит к генерации всех возможных пептидов для каждого белка, независимо от их N- или C-концевых остатков. Такой подход особенно полезен при анализе данных, полученных с использованием протеаз с неизвестной или плохо изученной специфичностью.

В режиме полуспецифического поиска создаются виртуальные пептиды, в которых один из концов (N- или C-) соответствует сайту расщепления, характерному для заданного фермента, а второй конец может быть произвольным. Например, при полуспецифическом поиске трипсина допускаются пептиды, у которых либо N-конец, либо C-конец соответствует лизину или аргинину, тогда как на противоположном конце может быть любая аминокислота. Таким образом, белки могут быть успешно идентифицированы даже при отсутствии данных о свойствах используемой протеазы, что делает возможным изучение ферментов с неизвестной специфичностью. При этом продукты гидролиза могут быть достоверно идентифицированы с использованием тандемной масс-спектрометрии, без необходимости в дополнительной модификации субстратов.

Результаты идентификации масс-спектрометрических данных, генерируемые поисковыми системами, как правило, оформляются в виде структурированных отчётов. Эти отчёты содержат перечень идентифицированных белков, соответствующих им пептидных последовательностей, а также параметры достоверности идентификации, (например, ion score, posterior error probability, q-value и др.). Эти данные могут быть экспортированы в различных форматах, таких как .txt, .csv, .xml, .mzIdentML и другие.

Несмотря на полноту представляемой информации и наличие режимов поиска, допускающих отсутствие заранее заданной специфичности фермента, стандартные поисковые системы не выполняют последующий анализ аминокислот, по которым фактически происходило расщепление и были идентифицированы белки. Результаты идентификации не интерпретируются с точки зрения предпочтений фермента. Таким образом, информацию о потенциальной специфичности – то есть, какие аминокислоты статистически преобладают в различных позициях (P_1 , P_1' и т. д.) – из таких отчётов невозможно получить напрямую.

В связи с этим, на этапе постобработки данных требуется применение дополнительных инструментов, чаще всего скриптов. Эти инструменты извлекают аминокислотные последовательности пептидов из результатов поиска, сопоставляют их с полными белковыми последовательностями, определяют точку разреза и формируют выровненные окна фиксированной длины, центрированных относительно предполагаемой точки разреза, (например, $P_6 \dots P_1 \dots P_5'$). Далее на основе этих данных можно провести оценку специфичности фермента, и визуализировать в виде тепловой карты или логотипа последовательностей (sequence logo). Тепловая карта отражает частоту встречаемости каждой аминокислоты в каждой позиции с использованием цветовой градации: чем чаще встречается аминокислота, тем выше интенсивность цвета. Логотип последовательностей представляет собой график, в котором высота каждой буквы пропорциональна частоте встречаемости соответствующей аминокислоты в данной позиции. К примеру, на рисунке 3 представлена тепловая карта и логотип последовательностей для альфа литической протеазы (Alp) *Lysobacter enzymogenes* type 495 [116]. Необходимость детального анализа обусловлена тем, что протеазы могут расщеплять пептидные связи не только по «предпочтительным», но и по атипичным для данного фермента аминокислотам [117]. Кроме того, при работе с большим числом субстратов возрастает вероятность ошибочной идентификации продуктов гидролиза и, как следствие, ошибочному определению сайтов расщепления. В связи с этим требуется статистический подход, основанный на анализе множества выровненных пептидных фрагментов, позволяющий выявлять воспроизводимые закономерности и отличать случайные события от специфических предпочтений фермента.

Однако простой анализ частоты встречаемости аминокислот недостаточен: интерпретация таких данных требует учёта неоднородности аминокислотного состава белков, поскольку аминокислоты изначально встречаются в белках с различной частотой [105]. Поэтому на этапе постобработки данных обычно применяются методы статистической оценки.

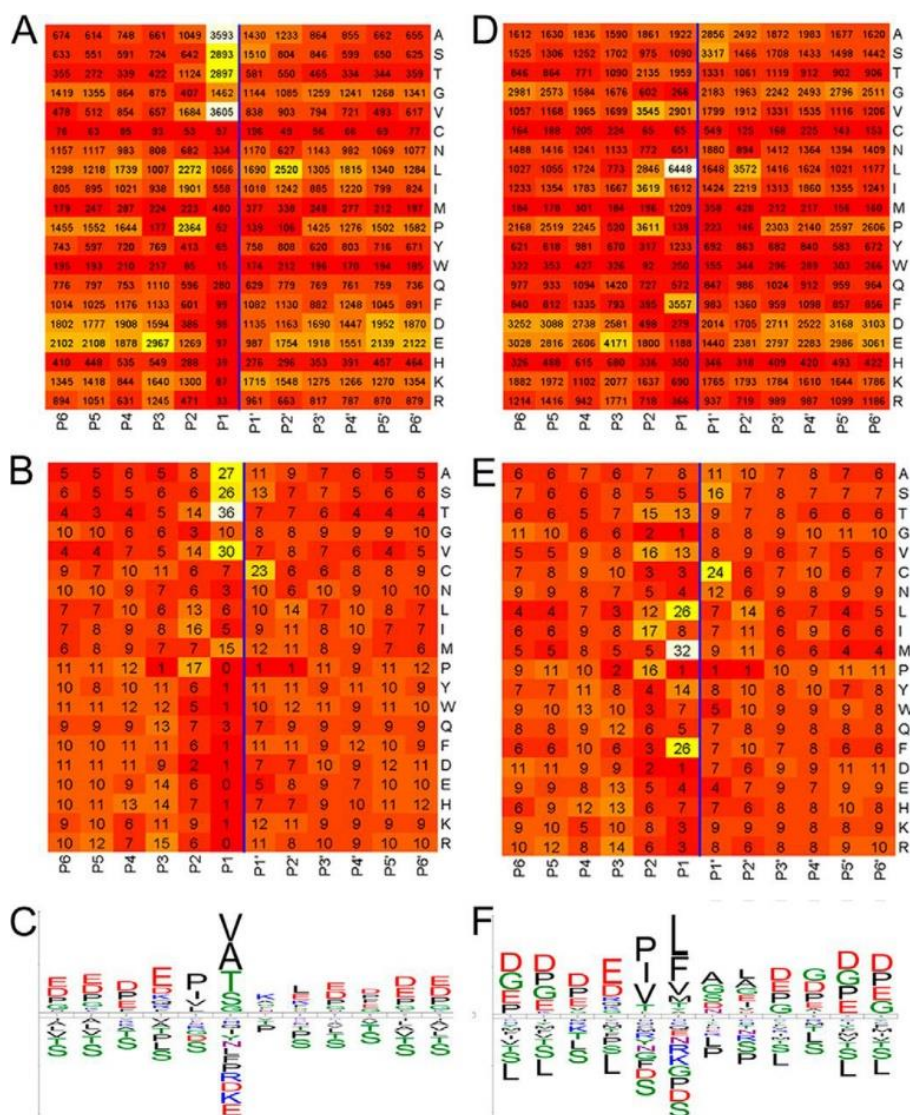


Рисунок 3. Тепловые карты и логотипы последовательностей для α -литической протеазы (Alp) *Lysobacter enzymogenes* type 495. Тепловые карты (A, B, D, E) отражают частоту встречаемости аминокислот в позициях P₈–P₈': чем выше частота, тем интенсивнее цвет. Логотипы (C, F) демонстрируют аминокислотные предпочтения в виде обобщённого мотива, где высота букв соответствует частоте появления остатков. Источник: [116].

2.4.3. Программные инструменты для анализа ферментативной специфичности

Для такой оценки значимости наблюдаемых частот используются специализированные программы визуализации мотивов, такие как iceLogo, WebLogo, pLogo, dagLogo, GibbsCluster и другие [118–122]. Все эти программы

принимают данные в виде списка выровненных фрагментов и решают схожую задачу: они анализируют частоту встречаемости аминокислот в определённых позициях относительно центра, выполняя статистическую обработку и визуализируют полученные данные в виде логотипов последовательностей и их разновидностей, к примеру iceLogo, WebLogo, а также тепловая карта и пр. Различия между этими инструментами заключаются в используемых алгоритмах статистической обработки и нормализации (таблица 4).

Таблица 4. Сравнительная характеристика программных инструментов для анализа протеолитической специфичности.

Программа	Подход к оценке	Учитывает фоновое распределение	Тип визуализации	Особенности
WebLogo	Энтропийный анализ вариабельности	Нет	Логотип последовательностей	Отражает информационную насыщенность (энтропию Шеннона), не учитывает статистическую значимость
iceLogo	«Fold change» и процентная разница	Да	Логотип последовательностей, тепловая карта	Выявляет как обогащённые, так и подавленные аминокислоты относительно фонового распределения
pLogo	Биномиальная модель с поправкой	Да	Логотип последовательностей	Основан на расчёте p-value; строгий статистический подход
dagLogo	Точный тест Фишера	Да	Логотип последовательностей	Поддерживает группировку остатков по химическим свойствам
GibbsCluster	Байесовская кластеризация	Нет	Отсутствует (только кластеры)	Выделяет несколько перекрывающихся мотивов, без визуализации

WebLogo – один из первых и наиболее известных инструментов для построения логотипов последовательностей [119]. Он был разработан как универсальное средство визуализации множественных выравниваний ДНК, РНК и белков. WebLogo до сих пор является своего рода «стандартом» для представления мотивов. Она отображает аминокислотный состав на основе энтропийного подхода, без использования фонового распределения. Этот метод отражает степень вариабельности каждой позиции, но не позволяет выявить статистически значимые отклонения от ожидаемой частоты.

В отличие от него, iceLogo [118] сравнивает частоты аминокислот в исследуемом наборе с фоновым распределением, рассчитывая Z-оценку (Z-score) и позволяя визуализировать как переизбыточные, так и недопредставленные остатки. В качестве фона могут использоваться выровненные аминокислотные фрагменты, белки определённого организма или общее распределение из базы данных.

pLogo реализует вероятностную модель, основанную на биномиальном распределении, с последующей коррекцией на множественное сравнение. Такой подход позволяет строго статистически оценить вероятность наблюдаемого распределения для каждой аминокислоты в каждой позиции. dagLogo использует критерий Фишера и дополнительно позволяет группировать аминокислоты по химическим свойствам, что особенно полезно при выявлении обобщённых закономерностей, таких как обогащение положительно заряженных остатков или гидрофобных групп.

В свою очередь, GibbsCluster не предполагает визуализации, а выполняет кластеризацию выровненных последовательностей с целью выявления нескольких перекрывающихся или конкурирующих мотивов. Результаты кластеризации могут быть затем переданы в другие программы для дальнейшей визуализации.

Каждая из программ визуализации мотивов имеет определённые ограничения. WebLogo не использует фоновое распределение аминокислот и отображает только степень вариабельности (энтропию) аминокислот в каждой

позиции. Такой подход не позволяет оценивать статистическую значимость отклонений от ожидаемого фона. Это особенно затрудняет интерпретацию результатов в случае ферментов, специфично расщепляющих редкие аминокислотные остатки, например, триптофан или цистеин. В отличие от него, iceLogo позволяет проводить статистическую оценку путём сравнения с фоновым распределением аминокислот, однако результаты чувствительны к выбору этого распределения. Кроме того, метод расчета поочерёдно оценивает статистическую значимость отклонения частоты каждой аминокислоты в каждой позиции от фонового распределения. При этом общее число проверяемых гипотез может быть весьма велико: например, при анализе 11 позиций и 20 аминокислот выполняется 220 независимых сравнений. В такой ситуации существенно возрастает вероятность появления ложноположительных результатов – то есть таких аминокислот, которые будут отмечены как статистически значимые не вследствие реальной специфичности фермента, а лишь по причине случайных совпадений.

Программа pLogo использует биномиальную модель, при которой каждая кислота рассматривается независимо, что игнорирует возможную взаимозависимость между позициями и аминокислотами (например, коррелирующие позиции, мотивы), а также имеет ограниченная поддержка альтернативных форматов визуализации. dagLogo, в свою очередь, реализует более гибкий подход, основанный на точном тесте Фишера. Тем не менее, из-за группировки теряется точная информация о конкретных аминокислотах, а настройка параметров требует ручного вмешательства. GibbsCluster не производит визуализацию и не выполняет статистическую оценку, а используется исключительно для кластеризации выровненных последовательностей с целью выделения нескольких возможных мотивов. Полученные результаты требуют последующей передачи в другие программы для интерпретации и визуального представления. Однако, пожалуй, самым главным ограничением является требование к входным данным. Все программы требуют предварительно выровненный список фрагментов и файл фонового распределения, либо в виде

выровненных последовательностей сравнения, либо в формате FASTA. Такие данные невозможно получить непосредственно из поисковых отчётов и требуют использования дополнительных инструментов для предварительной обработки.

2.5. Методы количественной оценки в масс-спектрометрии с изотопными метками

Как следует из раздела 2.3, методы определения кинетических параметров протеаз в основе своей сводятся к измерению накопления продукта или уменьшения концентрации исходного субстрата во времени [123]. В этой связи масс-спектрометрия представляет собой привлекательный инструмент, обладающий высокой чувствительностью, широким динамическим диапазоном, способностью различать близкие по массе соединения, а также возможностью одновременного анализа множества компонентов в сложной смеси.

Тем не менее, МС-детекция по своей природе является полуколичественной [124], поскольку интенсивность сигнала зависит не только от концентрации аналита, но и от его ионизационной эффективности, конкуренции при ионизации и матричных эффектов. Для преодоления этих ограничений и получения достоверных количественных данных, особенно в случае кинетического анализа, широко применяются методы с использованием стабильных изотопных меток. Такие метки позволяют точно сравнивать концентрации аналитов путём сопоставления интенсивностей сигналов изотопно-меченого стандарта и исследуемого соединения, отличающихся массой, но идентичных по химической природе и ионизационному поведению.

Можно выделить четыре группы методов введения изотопной метки в пептиды и белки:

- Метаболическое мечение – культивирование организмов на средах, обогащенных тяжелыми изотопами, к примеру, с аминокислотами, содержащими

тяжелые изотопы (stable-isotope labelling with amino acids in cell culture, SILAC) [125] (рисунок 4А);

- Синтетическое мечение – включение тяжелого изотопа при химическом синтезе пептидов или белков (углеродом или азотом) [126];
- Химическое мечение – модификация пептидов изотопно-различающимися реагентами, например isotope-coded affinity tag, ICAT (рисунок 4Б), включая изобарические метки, которые высвобождают репортерные ионы в MS/MS, и другими коммерчески доступными наборами, как TMT, iTRAQ и IPTL [127,128];
- Изотопный обмен – введение метки в результате изотопного обмена для измерения абсолютной концентрации (AQUA – Absolute QUAntification) [129].

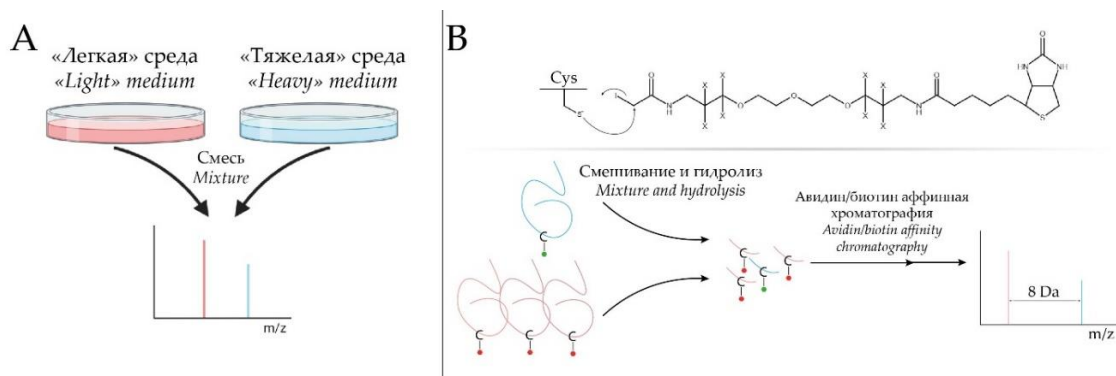


Рисунок 4. Методы изотопного мечения для количественного масс-спектрометрического анализа. А – схема метаболического мечения: синим цветом – клеточная культура на обычной среде («легкая» среда); красным цветом – среда с тяжелыми изотопами («тяжелая» среда). При сравнении двух белков, экстрагированных из клеточных культур, концентрации белков можно определить, сравнивая площади/интенсивности пиков в масс-спектре. В – тиоловые группы цистеинов модифицируются легкой ($X - {}^1\text{H}$) или тяжелой ($X - {}^2\text{H}$) метками (isotope-coded affinity tag, ICAT). Концентрации белков определяются путем сравнения площади/интенсивности пиков в масс-спектре.

Для количественной оценки активности протеаз и анализа их специфичности в протеомике используются различные методы изотопного мечения, позволяющие отслеживать динамику образования пептидов и определять как относительные, так и абсолютные концентрации продуктов гидролиза. Наиболее применяемыми являются подходы на основе стабильных изотопов: SILAC, TMT/iTRAQ, ICAT, IPTL и AQUA.

SILAC (Stable Isotope Labeling by Amino acids in Cell Culture) представляет собой метод метаболического включения изотопно меченых аминокислот (например, ^{13}C -лизина или ^{15}N -аргинина) в белки в процессе их синтеза. Его основное преимущество – высокая точность при сравнении белкового состава клеток, выращенных в разных условиях. Метод активно применяется для оценки влияния ингибиторов, воздействия протеаз, посттрансляционных модификаций. Несмотря на то, что метод SILAC широко применяется для оценки изменений в белковом составе при воздействии протеаз, он не позволяет напрямую исследовать кинетику протеолиза. Это связано с тем, что метод основан на метаболическом включении изотопных аминокислот во время синтеза белков, а не на измерении скорости их деградации.

TMT (Tandem Mass Tags) и iTRAQ (Isobaric Tags for Relative and Absolute Quantification) – методы химического мечения пептидов, основанные на введении изобарических меток, которые идентичны по массе на стадии MS1, но различаются в виде репортёрных ионов на стадии MS2. Это позволяет проводить одновременное сравнение до 16 образцов в одном анализе, минимизируя техническую вариабельность и повышая статистическую достоверность результатов. Эти методы активно применяются для анализа изменений белкового состава в разных физиологических состояниях, а также в задачах оценки относительной активности ферментов. Основные ограничения – высокая стоимость реагентов, сложность пробоподготовки, требование к высокому разрешению масс-спектрометра и возможная интерференция сигналов при анализе сложных образцов (coisolation effect).

ICAT (Isotope-Coded Affinity Tags) позволяет селективно метить цистеинсодержащие пептиды, что упрощает анализ сложных смесей за счёт снижения их химической сложности. Метод включает этап аффинной очистки и позволяет проводить относительное количественное сравнение двух образцов. Он эффективен при анализе систем, где интерес представляют белки с низкой экспрессией или специфические цистеиновые мишени. Недостатки –

невозможность анализа белков, не содержащих цистеин, и дополнительная пробоподготовка, повышающая риск потерь.

IPTL (Isobaric Peptide Termini Labeling) – метод, аналогичный TMT/iTRAQ, но ориентированный на мечение N- или C-концевых групп пептидов. Такой подход делает его потенциально полезным для оценки активности протеаз, поскольку терминальные группы образуются при расщеплении субстрата. Однако метод пока данных метод используется ограниченно.

2.5.1. Метод AQUA, использование метки ^{18}O

Одним из надёжных подходов к абсолютному количественному определению концентраций пептидов и белков в биологических образцах является метод AQUA (Absolute QUAntification, метод абсолютной количественной оценки), основанный на применении изотопно-меченых пептидов в качестве внутренних стандартов [130,131]. В отличие от относительных методов, AQUA позволяет определить абсолютную концентрацию целевых пептидов в молях на единицу объёма, что особенно важно в фармакокинетике, диагностике и ферментологических исследованиях.

В качестве метки применяются стабильные изотопы – ^{13}C , ^{15}N , ^2H или ^{18}O . Изотоп ^{18}O используется особенно часто благодаря простоте включения при ферментативном или химическом гидролизе.

Существует два механизма включения изотопных меток в молекулы аналита. Классический подход мечения пептидов основан на ферментативном гидролизе пептидных связей в воде H_2^{18}O . При протеолизе (например, эндопептидаза типа трипсина) происходит разрыв пептидной связи, и молекула воды с тяжелым изотопом кислорода входит в состав карбоксильного конца образующегося фрагмента. В результате в карбоксильную группу С-конца пептида встраивается один атом ^{18}O (рисунок 5). Повторное связывание в активном центре может

приводить к обмену второго кислорода на ^{18}O в уже образованной свободной карбоксильной группы [132–134].

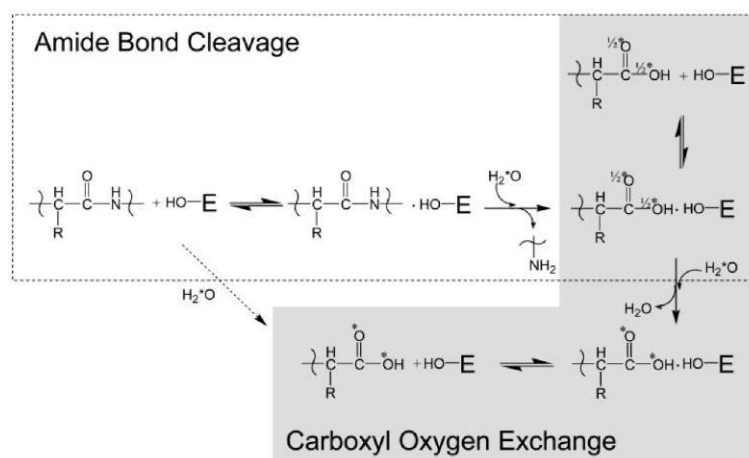


Рисунок 5. Механизм включения атомов ^{18}O в присутствии фермента и без фермента. Схема взята из [135].

Таким образом, на С-конце каждого нового пептида оказываются два атома ^{18}O , что вызывает суммарный сдвиг массы на +4 Да (для однозарядного иона). Первый подобный ферментативный обмен кислорода был описан ещё в 1951 году Спринсоном и Риттенбергом [136], а впоследствии метод был подтверждён масс-спектрометрически. Например, Антонов и соавторы с использованием масс-спектрометрии показали появление пиков со сдвигом массы, соответствующим включению ^{18}O в карбоксильный конец пептидов [137]. Эти данные подтвердили, что сериновые протеазы катализируют гидролиз с участием ацил-ферментного промежуточного комплекса, в ходе которого может происходить обмен карбонильного кислорода с молекулами воды. В ранних работах с помощью ^{18}O также подтвердили наличие тетраэдрического промежуточного комплекса при гидролизе сложных эфиров химотрипсином [134,138].

В отсутствие ферментов включение тяжёлого кислорода в органические молекулы происходит самопроизвольно в результате естественного обмена между водой H_2^{18}O и карбоксильными группами. Главную роль играют карбоксильные

группы – как С-концевой остаток пептида, так и боковые цепи аспарагиновой и глутаминовой кислот. Важная особенность: оба атома карбоксильной группы эквивалентны по способности к обмену, поэтому при достаточном времени оба могут замещаться на ^{18}O . Таким образом, максимальный эффект – внедрение двух ^{18}O на каждую карбоксильную группу [139]. Однако механизмы такого обмена принципиально отличны в кислой и щелочной среде.

При низком pH протонирование карбонильного кислорода активирует карбоксильную группу к нуклеофильной атаке водой. Реакция может идти через образование промежуточного гидратированного состояния с последующим выходом молекулы H_2^{16}O . В результате кислород из растворителя занимает место одного из исходных атомов кислорода кислоты. Многократное повторение этого акта приводит к статистическому распределению изотопов: при наличии достаточного избытка H_2^{18}O каждый из двух атомов в карбоксильной группе со временем с равной вероятностью станет тяжёлым. Экспериментально показано, что при $\text{pH} < 2$ (например, в разбавленной трифторуксусной кислоте) любая карбоксильная группа пептида способна постепенно обменять оба своих атома кислорода с растворителем. Причём это относится не только к С-концу пептида, но и к боковым цепям Asp/Glu: в присутствии H_2^{18}O они присоединяют два изотопа кислорода, что увеличивает массу пептида ещё на +4 Да.

Кислотно-катализируемое включение ^{18}O было предложено как простая и недорогая альтернатива химическому и метаболическому мечению в протеомных исследованиях. Метод не требует ферментов или дорогих реагентов – достаточно обогащённой водой и кислоты. Нильс и др. (2009) продемонстрировали возможность полного мечения пептидов по карбоксильным группам: каждый Asp/Glu остаток приводил к смещению масс, позволяя точно определить число таких остатков по общему сдвигу масс [139]. При этом хроматографическое «поведение» пептидов практически не менялось, поскольку введение ^{18}O не влияет на заряды и гидрофобность молекулы [134]. Такая метка использована для облегчения *de novo*-секвенирования пептидов: например, смесь $^{16}\text{O}/^{18}\text{O}$ (1:1) воды

при ферментативном гидролизе образует характерные двойки фрагментов, упрощающие интерпретацию MS/MS-спектров. Кроме того, кислотное ^{18}O - мечение применяли для получения изотопных стандартов пептидов. Поскольку любой очищенный пептид можно частично или полностью перевести в тяжёлую форму, это даёт удобный способ получения внутреннего стандарта для количественного анализа без химсинтеза меченых аминокислот.

Одним из ключевых применений ^{18}O -мечения является количественная масс-спектрометрия пептидов. Метод $^{16}\text{O}/^{18}\text{O}$ -мечения позволяет сравнивать уровни белков в двух образцах: один подвергается гидролизу в H_2^{18}O , другой – в H_2^{16}O . После объединения образцов соответствующие пептиды регистрируются как пары пиков, разнесённых на +4 Да, и соотношение их интенсивностей отражает относительное содержание белков [130]. Этот подход был реализован (Йао и соавторы) для количественного анализа сотен белков, а также успешно применялся в исследованиях посттрансляционных модификаций [140,141].

Кроме относительного, ^{18}O -мечение применяется для абсолютного измерения концентраций с использованием меченых стандартов. Пример практического применения представлен в исследовании Торопыгина и соавт., где с помощью MALDI-TOF и ^{18}O -стандартов определяли эффективность расщепления пептида A β -(1–16) ангиотензин-превращающим ферментом (ACE) и его отличия при изомеризации Asp⁷ [142]. Такие работы подтверждают высокую чувствительность и применимость метода для анализа специфичности и кинетики протеаз.

Несмотря на простоту и универсальность, метод ^{18}O -мечения обладает рядом ограничений. Основные сложности связаны с неполным включением изотопов, возможностью обратного обмена при последующей обработке, а также перекрытием изотопных форм в спектре, что требует дополнительных вычислительных поправок.

3. МАТЕРИАЛЫ И МЕТОДЫ

3.1. Определение специфичности B1p с использованием MALDI-TOF масс-спектрометрии

3.1.1. Гидролиз белков с использованием B1p

Для оценки протеолитической специфичности фермента B1p были использованы следующие модельные белки: миоглобин (Sigma-Aldrich, США), казеин (Sigma-Aldrich, США), гемоглобин (Sigma-Aldrich, США), овальбумин (Sigma-Aldrich, США), овомукоид (Sigma-Aldrich, США), енолаза 2 (генно-инженерный белок) и бычий сывороточный альбумин (Sigma-Aldrich, США). Перед проведением экспериментов белки очищали от низкомолекулярных примесей с использованием ультрафильтрации через фильтры Microcon YM-3 (Merck Millipore, Германия). Для этого 100 мкл белкового раствора (1 мг/мл) разводили 400 мкл деионизированной воды и наносили на фильтры. Центрифугирование проводили при 14 000 g в течение 100 минут при 25 °C, после чего фильтры переворачивали и дважды центрифугировали при 1 000 g по 3 минуты. Потери белка при этом не превышали 10%. Концентрацию белка восстанавливали до 1 мг/мл с учётом потерь.

Реакционная смесь состояла из 8 мкл 50 mM аммоний-бикарбонатного буфера, pH 7,8 (Sigma-Aldrich, США), 8 мкл белкового раствора (1 мг/мл) и 2 мкл раствора фермента B1p (289 мкг/мл). Инкубацию проводили при температуре 37 °C в термостате BE 400 (Mettler, Германия) в течение 180 минут.

3.1.2. Масс-спектрометрия MALDI-TOF

Продукты гидролиза белков B1p регистрировали с помощью масс-спектрометра MALDI-TOF/TOF Ultraflex II (Bruker, Германия), оснащенного Nd:YAG (иттрий-алюминиевый гранат, легированный неодимом) лазером с длиной волны 355 нм.

В качестве матрицы применяли раствор 2,5-дигидроксibenзойной кислоты (ДНВ) 10 мкг/мл, в 50% ACN и 0,35% TFA. 0,5 мкл образца, содержащего продукты гидролиза белков, смешивался с 0,5 раствора матрицы непосредственно на специализированной мишени масс-спектрометра и высушивали на воздухе при комнатной температуре.

Масс-спектры регистрировали в рефлексорном режиме и режиме положительных ионов при частоте импульсов лазера 66,7 Гц. Ускоряющее напряжение составляло 25 кВ, напряжение на рефлексоре – 26,38 кВ. Тандемные спектры (MS/MS) регистрировали в режиме TOF/TOF с использованием тандемного модуля LIFT при ускоряющем напряжении 8 кВ, напряжении на рефлексоре 29,50 кВ и напряжении LIFT-модуля 19 кВ.

Обработка и анализ спектров проводились с использованием программного обеспечения FlexAnalysis 3.0 (Bruker Daltonics, Германия). Для калибровки использовали два метода: внешнюю и внутреннюю. Внешнюю калибровку выполняли с помощью стандарта Peptide Calibration Standard II (Bruker Daltonics, Германия), содержащего набор пептидов с известными молекулярными массами в диапазоне 700–3500 Да. Внутренняя калибровка осуществлялась по пептидам, образующимся в результате автолиза фермента Vlp, предварительно идентифицированным методом тандемной масс-спектрометрии.

3.1.3. Идентификация продуктов гидролиза белков ферментом Vlp

Идентификация белков проводилась с использованием программы BioTools 3.0 и поисковой системы Mascot. Спектры, полученные в рефлексорном и тандемном режимах, были объединены для комбинированного поиска. Помимо этого, полученные в результате гидролиза массы пептидов сопоставлялись с возможными массами из исходного белка с использованием программы GPMW (LightHous, Дания). Продукты гидролиза идентифицировали путём поиска по базе

данных UniProt без указания специфичности протеазы (none). Допустимая ошибка измерения – 70 ppm. В качестве возможных посттрансляционных модификаций учитывались окисление метионина и пропионамидирование цистеина.

3.2. Гидролиз фракций белков лизата печени крысы и сыворотки человека B1p и L5 LC-MS/MS

3.2.1. Получение лизата печени крысы

Для получения образцов ткани использовали самцов крыс линии Wistar (*Rattus norvegicus*) в возрасте 5 месяцев. Животные содержались при естественном освещении с неограниченным доступом к стандартному корму и воде. Масса животных на момент начала эксперимента составляла 200–240 г. Эвтаназия проводилась методом декапитации под эфирным наркозом в соответствии с *Руководством AVMA по эвтаназии животных: редакция 2020 года*. Образцы печени были собраны, промыты 0,9% раствором NaCl и заморожены в жидком азоте. В эксперимент было включено три животных.

Приблизительно равные по массе фрагменты ткани печени от каждого животного были объединены в один общий образец для минимизации индивидуальных различий. Полученный образец измельчали в стеклянной ступке с пестиком в фосфатно-солевом буфере (PBS; 137 mM NaCl, 2,7 mM KCl, 10 mM Na₂HPO₄, 1,8 mM KH₂PO₄, pH 7,4), содержащем смесь ингибиторов протеаз. Далее лизат ресуспендировали с использованием роторного гомогенизатора Silent Crusher S (Heidolph, Германия).

3.2.2. Получение сыворотки человека

Образец сыворотки крови объёмом 1 мЛ был получен от одного здорового донора (женщина, 30 лет) (поставщик: ООО «АБМ»). Добровольное

информированное согласие и анкета донора были утверждены протоколом №309 независимого этического комитета ООО «Арте Мед Ассистанс» от 25 октября 2023 года. Все эксперименты были выполнены в соответствии с общепринятыми нормами гуманного обращения с лабораторными животными и требованиями Приказа Минздрава РФ от 1 апреля 2016 г. №199н «Об утверждении Правил надлежащей лабораторной практики», а также Директивы 2010/63/EU Европейского парламента и Совета ЕС от 22 сентября 2010 г. «О защите животных, используемых в научных целях».

3.2.3. Разделение и использованием эксклюзионной хроматографии, фракционирование

Разделение белка белкового материала из лизата ткани печени крысы по молекулярной массе методом эксклюзионной хроматографии проводили на хроматографе AKTA Purifier 10 (Cytiva, США) с управлением через программное обеспечение UNICORN v5.31. Образец лизата объёмом 2 мл (около 40 мг общего белка) инкубировали при +4 °C в течение 1 часа для установления равновесия белковых комплексов. Затем образцы центрифугировали при 12 000 g в течение 10 мин при 4 °C для удаления осадка. Фракционирование проводили на колонке HiLoad 16/600 с сорбентом Superdex 200 prep grade (Cytiva, США), предварительно уравновешенной буфером SEC, при скорости потока 800 мкл/мин и температуре 15 °C. Было собрано 6 SEC-фракций для последующей масс-спектрометрической идентификации белков. Концентрация общего белка в фракциях определялась по калибровочной кривой при 280 нм на спектрофотометре QE65000 с источником света DH-2000 (Ocean Optics, США); концентрации белка варьировали от 0,2 до 1 мг/мл.

Фракционирование сыворотки проводили на той же системе AKTA Purifier 10 (GE Healthcare Life Sciences, США) с колонкой Superose 6 10/300 GL (Cytiva, США). Колонку загружали 100 мкл образца сыворотки крови человека.

Хроматографическая система работала при постоянной скорости потока 0,3 мл/мин. Температура в хроматографической камере поддерживалась на уровне 4 °С, элюция белков контролировалась по абсорбции при 280 нм. Сбор фракций начинался после выхода 4 мл элюата; всего было собрано 10 фракций объёмом по 500 мкл.

Концентрация общего белка во фракциях сыворотки и лизата печени определялась по методу Бредфорда с использованием планшетного ридера ClarioStar (BMG, Германия).

3.2.4. Гидролиз белков для определения специфичности исследуемых ферментов

Vlr экспрессирована из *Lysobacter capsici* штамма ВКМ В-2533^Т, а L5 из штамма XL1, согласно протоколу, описанному в работах [34,41,47].

Восстановление и алкилирование остатков цистеина во всех фракциях проводили путём добавления 10 мМ раствора трис(2-карбоксиэтил)фосфина (TCEP) (Thermo Scientific, Waltham, MA, USA) и 40 мМ 2-хлорацетамида (CAA) (Sigma, St. Louis, MO, USA) к каждой фракции, содержащей 10 мкг общего белка, с последующей инкубацией в течение 15 минут при 37 °С. Гидролиз осуществляли добавлением 3 мкл раствора фермента L5 (концентрация 46 мкг/мл) и 50 мМ аммонийбикарбонатного буфера в каждую фракцию. Гидролиз с использованием фермента Vlr проводили в идентичных условиях, за исключением концентрации фермента, составлявшей 289 мкг/мл. Протеолиз трипсином проводили путём добавления 10 мкл раствора трипсина (20 мкг/мл) в 50 мМ аммонийбикарбонатный буфер к каждой фракции, содержащей 10 мкг белка. Гидролиз одиночных белков, использованных в работе, проводился согласно такому же протоколу.

Помимо лизата печени крысы и сыворотки человека, в качестве субстрата также использовали коммерчески доступные, а также генно-инженерные белки:

казеин, BSA, CPOX, FXN, PPOX и SULT2A1. Восстановление и алкилирование остатков цистеина, гидролиз, а также инкубацию проводили согласно протоколу, описанному выше.

Образцы инкубировали в течение ночи (около 16 часов) при 37 °C в термостате BE 400 (MEMMERT, Германия).

3.2.5. LC-MS/MS анализ

Масс-спектрометрический анализ выполнен на приборе micrOTOF-QII (Bruker Daltonik, Бремен, Германия), оснащённом источником ионизации CaptiveSpray и сопряжённым с UHPLC-системой nanoElute (Bruker Daltonik, Бремен, Германия). Перед анализом образцы обессоливали на C18 StageTips (Supelco, Merck, США) согласно протоколу [143]. Высушенные образцы растворяли в 0.1% муравьиной кислоте, и 1–6 мкл каждого образца загружали на колонку-ловушку Acclaim™ RepMap™ C18 (5 мкм, 0.3 мм × 5 мм; Thermo Fisher Scientific, США) при давлении 400 бар. Объём загрузки варьировался в зависимости от выхода пептидов после обессоливания и концентрирования в скоростном вакуумном концентраторе Concentrator plus (Eppendorf, Германия).

Разделение пептидов осуществляли на колонке Aurora Ultimate CSI C18 (1.7 мкм, 120 Å, 75 мкм × 250 мм; IonOpticks, Австралия) при температуре 50 °C и постоянной скорости потока 300 нл/мин, используя линейный градиент от 2% до 85% подвижной фазы В (0.1% муравьиной кислоты в ацетонитриле) относительно фазы А (0.1% муравьиной кислоты в воде), при общей длительности градиента 100 минут, с последующим уравниванием для следующей хроматограммы. Профиль градиента указан в таблице 4.

Таблица 4. Профиль градиента элюирования в ходе хроматографического разделения пептидов.

Время, мин	% фразы В
0	2
1	3
57	17
78	25
92	34
93	85
100	85

Масс-спектрометрический анализ проводили в режиме положительных ионов при напряжении на капилляре 1500 В, температуре осушающего газа (азота) 150 °С и расходе 3 л/мин. MS1-сканирование выполняли в диапазоне m/z 150–2200 с частотой 2 Гц. Спектры MS/MS регистрировали в автоматическом режиме с динамическим выбором прекурсоров с зарядом от 2+ до 4+, временем цикла 3 с и частотой съёма спектров 8–32 Гц. Фрагментацию пептидов осуществляли методом соударения (CID) в режиме DDA (data-dependent acquisition). Энергия соударения автоматически оптимизировалась в зависимости от m/z и зарядового состояния прекурсора.

Обработка полученных спектров выполнялась в программном обеспечении Compass DataAnalysis 5.1 (Bruker Daltonik, Германия). Списки пиков экспортировались в формате Mascot Generic Format (*.mgf) и далее анализировались с использованием поискового сервера Mascot версии 2.3.0 (Matrix Science, Лондон, Великобритания). Поиск осуществлялся с параметрами: фермент – без указания специфичности; переменные модификации – дезамидирование (NQ), окисление (M), карбамидометилирование (C); допуск по массе пептидов – 4.5 ppm; допуск по фрагментам – 0.2 Да; допустимые заряды прекурсоров – 2+, 3+, 4+. Для образцов из печени крысы указывалась таксономия *Rattus norvegicus*, для образцов сыворотки – *Homo sapiens*.

3.2.6. Биоинформатрический анализ специфичности *Vlp* и *L5*.

Валидация результатов идентификации пептидов системой Mascot осуществлялась двумя подходами. Стандартный метод в протеомике основан на расчёте уровня ложных открытий (FDR – false discovery rate) путём сравнения попаданий по основной и decoy (приманке) базе данных, при этом обычно используется порог $\leq 1\%$. Однако корректная оценка FDR требует достаточного количества идентификаций (обычно >100 белков на образец). В настоящей работе в ряде образцов количество идентифицированных белков оказалось менее 100, что объясняется ограниченной протеолитической активностью исследуемых ферментов и низкой чувствительностью используемого масс-спектрометра. В таких условиях расчёт FDR становится нестабильным и склонным к переоценке (например, одно декой-попадание при 50 белках даёт FDR = 2%) [144].

Для преодоления этих ограничений был использован альтернативный подход, основанный на отсечке по значению Ion Score, рассчитываемому в системе Mascot и отражающему степень соответствия теоретического и экспериментального спектров. Для подбора оптимального порога были протестированы значения Ion Score от 1 до 30 с построением тепловых карт для каждого уровня. При значениях ниже 20 наблюдалась высокая вариативность и отсутствие устойчивых закономерностей, вероятно, обусловленная случайными совпадениями. При Ion Score ≥ 20 профили становились стабильными и не менялись при дальнейшем повышении порога. Таким образом, пороговое значение Ion Score = 20 было выбрано для всех образцов как оптимальный компромисс между надёжностью данных и минимизацией ложноположительных совпадений.

Файлы результатов поиска Mascot экспортировались в формате XML с использованием большинства параметров по умолчанию, за исключением: формат – XML, порог значимости – $p < 0,05$, включение аминокислотных последовательностей белков, порог Ion Score установлен на уровне 20.

Для определения конкретных пептидных сайтов, по которым прошел гидролиз, использовалось специально разработанная автором работы программа – CleaveScore, которая подробно будет описана в разделе 4.3. Все файлы результатов идентификации были экспортированы из поисковой системы Mascot формате XML. Для сравнения полученных результатов в программе CleaveScore, были построены логотипы последовательностей в программе IceLogo [118]. В качестве референсного набора использовали фрагменты, полученные из белков сыворотки человека после гидролиза трипсином. Фрагменты формировались методом «скользящего окна» в соответствии с рекомендациями авторов программы, с применением собственного скрипта на Python.

Для поиска возможных повторяющихся мотивов вокруг сайта гидролиза ферментом L5 из каждого уникального пептида, идентифицированного методом LC-MS/MS, извлекались аминокислоты, соответствующие позициям P_3 , P_2 и P_1 . Для обработки данных использовался Python-скрипт. Повторяющиеся пептиды учитывались только один раз во избежание статистических искажений. Частотное распределение пар аминокислот (P_2-P_1) использовалось для построения тепловой карты, отражающей предпочтительные пары. Также анализировались триплеты ($P_3-P_2-P_1$), которые группировались и ранжировались по частоте встречаемости.

Для оценки распределения аминокислот в позициях P_2-P_1 при гидролизе протеазой L5 из набора всех идентифицированных белков извлекали фрагменты в формате P_2-P_1 с использованием программы CleaveScore. Дополнительно из тех же данных выделяли трёхаминокислотные фрагменты в формате $P_3-P_2-P_1$ для анализа повторяющихся мотивов. Все полученные мотивы были сгруппированы и отсортированы по относительной частоте встречаемости. Полученные значения визуализировали в виде тепловой карты [145].

3.3. Определение кинетических параметров с использованием изотопно-меченых стандартов и MALDI-TOF масс-спектрометрии

3.3.1. Получение пептидов для включения изотопной метки

Для валидации предлагаемой методики в качестве модельного фермента, кинетические параметры которого изучены и описаны в литературе, был использован трипсин. Пептид TTMPLW, для введения в него изотопной метки ^{18}O и дальнейшего использования в качестве стандарта, получали путем гидролиза казеина (15,5 мкМ) трипсином (1,84 мкМ) в 100 мМ аммоний бикарбонатном буферном растворе pH 7,8. Пептиды AWYYVPLG, SENSEKTTMPLW и TQYTDAPSFSDIPNPIG – продукты гидролиза казеина B1p – для определения кинетических параметров получали путем гидролиза казеина (18,5 мкМ) B1p (1,69 мкМ) в 100 мМ аммоний бикарбонатном буферном растворе pH 7,8.

Гидролиз проводили в течение 30 мин при 37 °C. Пептидную фракцию полученной смеси отделяли центрифугированием с помощью центрифужных фильтров Microcon YM-3 и далее разделяли методом обращенно-фазовой ВЭЖХ на колонке Luna® C18(2) (5 мкм, 100 Å, 150×4,6 мм, Phenomenex, США, кат. № 00F-4252-E0). Фаза А – 0,1% ТФУ; фаза В – 0,1% ТФУ, растворенная в ацетонитриле; линейный градиент от 4 до 75% фазы В за 40 мин, поток 1 мл/мин, температура колонки 40 °C. Хроматограммы регистрировали на длинах волн 215 и 280 нм. Собранные на коллекторе фракций пептиды идентифицировали с использованием тандемной масс-спектрометрии MALDI TOF/TOF и высушивали на центрифужном концентраторе до остаточного объема 5 мкл. Концентрации пептидов были рассчитаны по площадям хроматографических пиков; спектрофотометрическая ячейка хроматографа предварительно откалибрована по пептидам, содержащим тирозин и/или триптофан.

Пептиды RFFVAPF и RGPFP1V, использованные для определения кинетических параметров L5, были синтезированы.

3.3.2. Изотопный обмен

Введение изотопной метки осуществляли по стандартному протоколу. Пептиды предварительно высушивали в центрифужном вакуумном концентраторе SpeedVac (Eppendorf, Германия), после чего растворяли в 10% растворе ТФУ на воде, обогащённой изотопом ^{18}O (H_2^{18}O , ТОО «НПО Изотоп», Россия), и инкубировали при 60 °С от 20 минут до 12 часов. Степень изотопного обмена контролировали масс-спектрометрически в интервале времени от 20 минут до 12 часов.

3.3.3. Гидролиз белков исследуемыми протеазами для определения кинетических параметров

Для определения кинетических параметров модельного фермента – трипсина – в качестве субстрата использовали казеин. Растворы казеина в концентрациях 5,25; 2,62 и 1,51 мкМ смешивали с раствором трипсина (0,37 мкМ) в 100 мМ аммоний-бикарбонатном буфере, рН 7,8. Реакционную смесь инкубировали в течение 60 минут при температуре 37°С.

Через 5 минут от начала инкубации отбирали от 3 до 10 мкл раствора реакционной смеси и добавляли к ним от 3 до 10 мкл раствора стандарта пептида TTMLPW, меченного ^{18}O (концентрация 6,47 мкМ) в соотношениях 1:2, 1:1, 2:1. Стандарт был растворён в 10% растворе ТФУ в H_2^{18}O , что обеспечивало мгновенную инактивацию фермента. После перемешивания на шейкере смесь немедленно наносили на масс-спектрометрическую мишень. Аналогичные пробы отбирали с интервалом 10 минут до 60-й минуты.

Для определения кинетических параметров V_{1р} использовали растворы казеина в концентрациях от 0,89 до 7,11 мкМ. Растворы белка смешивали с V_{1р} (1,69 мкМ) в 100 мМ аммоний-бикарбонатном буфере. Через 5 минут от начала инкубации отбирали 3 мкл раствора реакционной смеси и добавляли по 3 мкл

стандарта пептида. В качестве изотопно-меченых стандартов использовали пептиды AWYYVPLG, SENSEKTTMPLW и TQYTDAPSFSDIPNPIG. После перемешивания на шейкере смесь наносили на масс-спектрометрическую мишень. Отбор проб повторяли с интервалом 10 минут до 30-й минуты.

Для определения кинетических параметров L5 раствор казеина в концентрациях от 1,28 мкМ до 25,6 мкМ смешивали с раствором L5 (0,15 мкМ) в 100 мМ аммоний-бикарбонатном буфере. Через 5 минут от начала инкубации отбирали 3 мкл раствора реакционной смеси и добавляли по 3 мкл раствора стандартов пептидов RFFVAPF и RGPFPPIV, меченных ^{18}O , в концентрациях согласно таблице 5. После перемешивания смесь немедленно наносили на масс-спектрометрическую мишень. Пробы отбирали каждые 10 минут до 60-й минуты, а также дополнительно через 147 минут.

Таблица 5. Концентрации пептидов стандартов RFFVAPF и RGPFPPIV, которые добавлялись к реакционной смеси.

Концентрация субстрата (мкМ)	Концентрация стандарта RFFVAPF (мкМ)	Концентрация стандарта RFFVAPF (мкМ)
1, 28	$1,18 \pm 0,1$	$1,52 \pm 0,05$
2,56	$1,18 \pm 0,1$	$2,34 \pm 0,07$
5,12	$1,18 \pm 0,1$	$2,34 \pm 0,07$
7,68	$11,3 \pm 0,3$	$36,7 \pm 0,8$
10,24	$11,3 \pm 0,3$	$36,7 \pm 0,8$

3.3.4. MALDI масс-спектрометрия

Подготовку образцов для MALDI-TOF анализа и регистрацию масс-спектров проводили аналогично описанному в разделе 3.1.2. Кратко, в качестве матрицы использовали раствор DHB, образцы наносили непосредственно на мишень масс-спектрометра. Спектры регистрировали в рефлекторном режиме положительных ионов, а тандемные масс-спектры – в режиме TOF/TOF с использованием блока LIFT.

Полученные спектры анализировались программе в FlexAnalysis 3.0. Параметры алгоритма обработки спектров были следующие: метод сглаживания – SavitzkyGolay, ширина пика – 0,2 m/z, количество циклов 1; метод вычитания базовой линии – TopHat. Алгоритм детектирования пика – Centroid, соотношение сигнал/шум – 4.

Полученные данные спектров, а именно m/z и интенсивность пиков экспортировали в Excel для дальнейшей обработки и расчета концентрации пептидов – продуктов реакции.

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

4.1. Идентификация внеклеточных протеаз бактерий рода *Lysobacter*, и анализ их гомологии с родственными протеазами

Очищенные рекомбинантные белки L5 и Vlp были проанализированы методом LC-MS/MS с последующей идентификацией пептидов в поисковой системе Mascot.

Протеаза Vlp синтезируется как профермент длиной 374 аминокислоты и включает сигнальный пептид (1–24), пропептид (25–195) и зрелую каталитическую часть (196–374), состоящую из 179 остатков. Для зрелой молекулы Vlp покрытие аминокислотной последовательности составило 88%. Идентифицированные пептиды перекрывали как структурные, так и функционально значимые области белка, включая предполагаемый сигнальный пептид и каталитический домен, что позволило достоверно подтвердить его идентичность.

Протеаза L5 (длина последовательности 205 аминокислот) была охарактеризована с покрытием 97% аминокислотной цепи. Набор диагностических пептидов охватывал ключевые участки молекулы, включая каталитически активные регионы, что обеспечило высокую достоверность идентификации. Для определения гомологии идентифицированных белков с известными протеазами был проведён BLAST-анализ их аминокислотных последовательностей в базе UniProt.

Анализ протеазы Vlp (178 а.о.) показал полное совпадение (100% идентичности) с β -литической металлопротеазой *Achromobacter lyticus* (E-value $7e^{-136}$). Высокая степень гомологии также выявлена с β -литическими протеазами *L. enzymogenes* (97 %), *L. antibioticus* (83 %). Все ближайшие гомологи принадлежат к семейству металлопротеаз M23, участвующих в гидролизе пептидогликана и обладающих выраженной бактериолитической активностью. Наиболее близким гомологом среди других протеаз является протеаза LasA *Lysobacter capsici* AZ78

(98% идентичности, 99% покрытия, E-value $5e^{-133}$), обладающей бактериолитической активностью (таблица 6). Принадлежность обоих ферментов к семейству M23 и наличие схожих функциональных доменов указывает на возможное сходство механизмов действия [146].

Кроме представителей *Lysobacter*, в числе ближайших гомологов были выявлены металлопротеазы из родов *Rheinheimera*, *Montanilutibacter*, *Marilutibacter*, *Agrilutibacter* и *Kinneretia*, показывающие идентичность в диапазоне от 77 до 84 % при высоком покрытии (от 95 до 98 %). Все они аннотированы как представители семейства M23 и обладают характерными мотивами, типичными для пептидогликан-гидролизующих ферментов.

Таблица 6. Ближайшие гомологи протеазы Vlp по результатам BLAST

Ближайший гомолог	Организм	% идентичности	Покрытие	E-value
Beta-lytic metalloendopeptidase	<i>Achromobacter lyticus</i>	100%	100%	$7e^{-136}$
LasA protease	<i>Lysobacter capsici</i> AZ78	98%	100%	$1e^{-133}$
Beta-lytic metalloendopeptidase	<i>Lysobacter enzymogenes</i>	97%	99%	$5e^{-130}$
Beta-lytic metalloendopeptidase	<i>Lysobacter antibioticus</i>	84%	97%	$1e^{-111}$
Beta-lytic metalloendopeptidase	<i>Lysobacter</i> sp. Root96	83%	97%	$3e^{-110}$
M23 family metallopeptidase	<i>Montanilutibacter psychrotolerans</i>	82%	97%	$3e^{-108}$

Анализ протеазы L5 (205 а.о.) показал наибольшее сходство с сериновыми протеазами семейства S1A. Наиболее близким гомологом оказался стрептогризин С (*Streptomyces griseus*), уровень идентичности с которым составил около 50 %, а покрытие – более 90 % (E-value $7e^{-142}$). Значительное сходство наблюдалось также с α -литическими протеазами *Lysobacter enzymogenes* и другими представителями рода *Lysobacter* (таблица 7). Эти данные позволяют отнести L5 к сериновым эндопептидазам трипсиноподобного типа.

Таблица 7. Ближайшие гомологи протеазы L5 по результатам BLAST

Ближайший гомолог	Организм	% идентичности	Покрытие	E-value
Streptogrisin-C (Serine protease C)	<i>Vulcaniibacterium thermophilum</i> / <i>Cognatilysobacter</i> sp.	~49–51%	>90%	$7e^{-142}$ – $4e^{-86}$
Alpha-lytic protease	<i>Lysobacter enzymogenes</i>	~47%	>85%	$5e^{-75}$
Другие S1-протеазы (Peptidase S1A)	<i>Marilutibacter</i> , <i>Wenzhouxiangella</i> и др.	45–50%	>80%	10^{-61} – 10^{-35}

Таким образом, оба исследованных фермента представляют собой внеклеточные протеазы рода *Lysobacter*, относящиеся к разным семействам: L5 – к сериновым эндопептидазам S1, а B1p – к металлопротеазам M23. Их совокупное присутствие отражает типичную для рода стратегию использования комплекса протеаз различной природы, обеспечивающих как протеолитическую деградацию белковых субстратов, так и бактериолитический эффект за счёт разрушения клеточной стенки.

4.2. Идентификация B1p и L5. Определение протеолитической специфичности на модельных белках методом MALDI-TOF MS

Ранее субстратная специфичность B1p изучалась ограниченном наборе белковых субстратов, включая инсулин и казеин, или другие стандартные субстраты. Эти исследования были выполнены в основном в 1960–1990-х годах и не предусматривали систематического или статистического анализа.

Для оценки субстратной специфичности фермента B1p использовался подход, основанный на гидролизе индивидуальных белков с последующим анализом продуктов реакции методом MALDI-TOF масс-спектрометрии. Данный метод был выбран в качестве первичного инструмента анализа ввиду его высокой чувствительности, скорости анализа, простоты подготовки образцов и совместимости с буферами на основе летучих компонентов, оптимальными для работы фермента.

Для этого анализа в качестве модельных белков были выбраны коммерчески доступные миоглобин, казеин, гемоглобин, овальбумин, овомукоид, енолаза 2 и бычий сывороточный альбумин (BSA). Данный набор белков был сформирован с учетом их аминокислотного состава, поскольку он репрезентативен относительно природного аминокислотного состава, представленного в базе данных Swiss-Prot (рисунок 6). К тому же это хорошо изученные и коммерчески доступные белки.

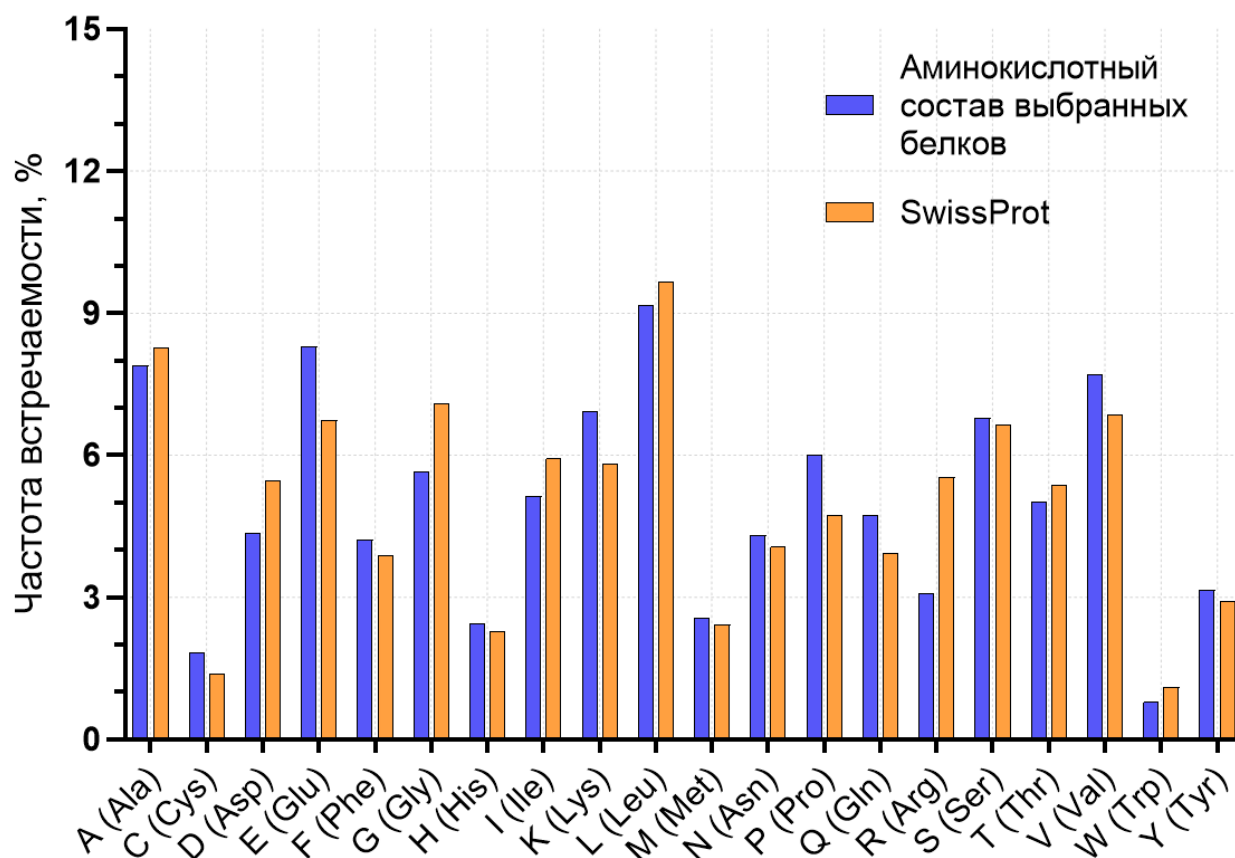


Рисунок 6. Сравнение аминокислотного состава базы Swiss-Prot и набора белков, использованных в качестве субстрата.

Для устранения влияния возможных примесей, часть белков дополнительно подвергалась предварительному разделению методом диск-электрофореза в полиакриламидном геле (ПААГ). Затем проводили гидролиз в геле используя исследуемую протеазу, и полученные продукты анализировались с использованием MALDI-TOF масс-спектрометрии.

В результате анализа полученных спектров были идентифицированы пептиды, содержащие сайты расщепления по аминокислотам, представленным в таблице 8. Учитывались аминокислоты только из тех пептидов, которые были дополнительно идентифицированы методом тандемной масс-спектрометрии.

Таблица 8. Идентифицированные сайты гидролиза B1p по данным, полученным с использованием MALDI-TOF масс-спектрометрии.

Белок	Идентифицированные сайты гидролиза
Миоглобин	Gly, Lys, Ala, Phe, Ser, Asp, Thr
Альфа-S1-казеин	Gly, Lys, Ala
Гемоглобин (A/B)	Gly, His
Овальбумин	Gly, Ala, Lys, Thr
Овомукоид	Gly, Lys, Ala, Val, Glu
Енолаза 2	Gly, Ala
BSA	Gly, Ala, Lys, Thr, Val, Arg, Cys, Ser

Для статистической оценки предпочтений была проведена сводная статистика по аминокислотам, встречающимся в позиции P1. Всего было идентифицировано 67 сайтов гидролиза, среди которых наибольшее число приходилось на глицин – 30 сайтов (44,8%), аланин – 16 (23,8%), лизин – 9 случаев (13,4%), аргинин – 3 (4,5%), и серин – также 3 (4,5%). Остальные аминокислоты, включая треонин, фенилаланин, глутаминовую кислоту, валин и цистеин, встречались в единичных случаях (от 1 до 2 раз) и каждый составлял менее 3% каждого от общего числа. Полученные данные подтверждают выраженное предпочтение фермента к остаткам глицина, но при этом гидролиз может проходить и по другим сайтам.

Таким образом, были показаны новые, ранее не описанные сайты гидролиза B1p. Однако статистическая оценка оказывается затруднительной из-за ограниченного количества анализируемых субстратов. В итоге выборка анализируемых продуктов оказывается статистически маленькой (или нерепрезентативной), что критично при исследовании ферментов с ранее не исследованной специфичностью.

В отличие от Vlp, анализ специфичности протеазы L5, данных о специфичности которой на момент начала представляемой работы опубликовано не было, оказался более затруднительным. L5 проявляет способность расщеплять пептидные связи по нескольким аминокислотным остаткам, без явно выраженного предпочтения к одному или двум из них. Для более точного определения его специфичности необходим анализ с применением методов LC-MS/MS и статистической нормализации.

4.3. Разработка и применение программного инструмента CleaveScore для анализа специфичности протеаз

Такая вариативная специфичность L5 требует более глубокого и количественно обоснованного подхода к анализу. Использование MALDI-TOF MS позволяет наблюдать отдельные продукты гидролиза, но не обеспечивает достаточной глубины покрытия и статистической достоверности при работе с ферментами, специфично гидролизующими по нескольким аминокислотам. Для получения более полного спектра субстратных предпочтений был применён метод LC-MS/MS, позволяющий одновременно анализировать продукты гидролиза сотен белков и идентифицировать тысячи уникальных пептидов. Однако столь высокий объём данных, получаемых при LC-MS/MS, делает невозможным ручной анализ сайтов расщепления. Это обусловило необходимость разработки специализированного программного инструмента, автоматизирующего обработку и визуализацию таких данных. Однако увеличение объема информации делает ручную обработку данных крайне трудоёмкой и нецелесообразной.

Стандартные поисковые системы, такие как Mascot (использованная в настоящей работе), MaxQuant, PEAKS и др., ориентированы преимущественно на идентификацию белков, но не предусматривают специализированного анализа специфичности фермента. Как уже отмечалось в литературном обзоре, программы для визуализации специфичности (например, IceLogo, dagLogo и др.) требуют на

вход заранее подготовленных выровненных фрагментов, содержащих аминокислотные остатки, представленные вокруг предполагаемого сайта разреза. Поэтому полученные из Mascot результаты требуют дополнительной обработки.

Для корректного представления данных о специфичности фермента была реализована процедура фильтрации и нормализации данных, основанная на следующих принципах:

1. Исключение дублирующихся пептидов.

В LC-MS/MS один и тот же пептид может быть идентифицирован многократно в рамках одного анализа. Если такие повторы не исключать, это приводит к завышению частоты гидролиза по соответствующим аминокислотам и искажению статистики. Для получения репрезентативных данных учитывались только уникальные пептиды.

2. Учёт уникальных положений разрезов.

При анализе сайтов гидролиза определяется аминокислота непосредственно до идентифицированного пептида и последняя аминокислота этого же пептида. В случае, если идентифицируется несколько последовательных пептидов одного белка, то С-конец одного идентифицированного пептида, является N-концом следующего идентифицированного пептида. Однако, эти пептиды являются результатом трех актов гидролиза, а не четырех. К тому же, протеазы способны пропускать сайты гидролиза и несколько разных пептида могут быть порезаны по одному аминокислотному остатку, образующему С или N конец пептида. Поэтому учитывалась не только аминокислота до пептида и последняя аминокислота в каждом пептиде, а уникальное положение таких аминокислот в исходной белковой последовательности.

3. Исключение терминальных сайтов.

При анализе С-концевых пептидов, последняя аминокислота не является продуктом гидролиза. А в том случае, если N-концевая аминокислота первого пептида является метионином, важно учитывать, что отщепление иницирующего

метионина (N-мет) представляет собой природную посттрансляционную модификацию, осуществляемую метионинаминопептидазой (МАП), а не результат действия исследуемой протеазы. Такие аминокислоты исключались из анализа.

4. Анализ данных из нескольких независимых экспериментов. В связи с ограниченной пропускной способностью применяемого масс-спектрометра анализ проводился на нескольких независимых выборках, включая фракции печени крысы и фракции сыворотки человека. Для повышения статистической достоверности результаты из всех экспериментов были объединены в единый массив.

Поскольку ни одна из существующих программ не удовлетворяла вышеперечисленным требованиям к фильтрации данных, а также не позволяла обрабатывать выходные данные Mascot, в рамках данной работы была разработана специализированная программа CleaveScope. Этот инструмент обеспечивает полный цикл обработки результатов LC-MS/MS-анализа с целью выявления закономерностей в специфичности протеазы.

Входными данными служат один или несколько файлов отчётов, сгенерированных поисковой системой Mascot в формате XML, содержащих идентифицированные белковые последовательности и соответствующие им пептиды.

На первом этапе для каждого идентифицированного белка извлекается полная аминокислотная последовательность, на основании которой формируется общее распределение аминокислот, используемое впоследствии для нормализации. Далее анализируется каждое вхождение пептида, принадлежащие соответствующему ему белку. Для каждого пептида определяются две аминокислоты: непосредственно предшествующая пептиду и С-концевая аминокислота самого пептида. Затем определяется их уникальное положение в белковой последовательности. Если ранее этот сайт уже был учтён (например, в случае многократной идентификации одного и того же пептида), он исключается

из повторного анализа. Кроме того, предусмотрена возможность исключения N-концевого метионина (рисунок. 7), если он соответствует первому остатку в белке, а также С-концевой аминокислоты, если она принадлежит последнему пептиду.

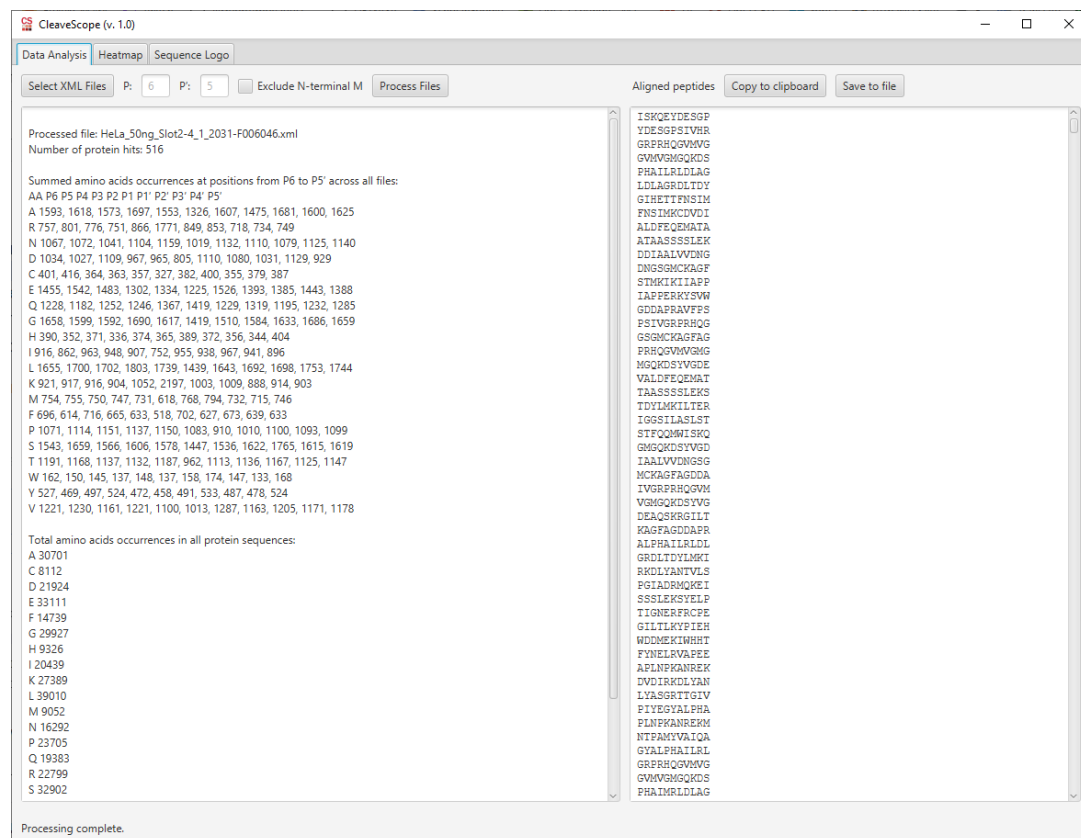


Рисунок. 7. Интерфейс программы CleaveScore. Вкладка ввода данных, выбора диапазона анализируемых сайтов, экспорта полученных выровненных фрагментов.

Вокруг каждого уникального сайта гидролиза извлекается аминокислотный фрагмент заданной длины, выровненный по положению предполагаемого сайта гидролиза. При этом число аминокислотных позиций, подлежащих анализу относительно предполагаемой точки разреза задается вручную в диапазоне от P_{12} до P_{12}' (Рисунок 7). Если количество остатков по одну из сторон оказывается недостаточным (например, сайт находится близко к концу белковой цепи), недостающие позиции заполняются фиктивным символом 'Z', который впоследствии исключается из анализа. Для каждой позиции в этом окне подсчитывается частота появления соответствующих аминокислотных остатков. Результатом обработки является позиционно-специфичная частотная матрица

(PSSM, Position-Specific Scoring Matrix), отражающая распределение аминокислот в заданных позициях относительно предполагаемого сайта протеолиза. Эта матрица служит основой для построения тепловой карты и логотипа последовательностей (Рисунок 8). Помимо этого, формируется список выровненных фрагментов вокруг предполагаемого сайта гидролиза, который возможно экспортировать в виде файла (в формате .txt).

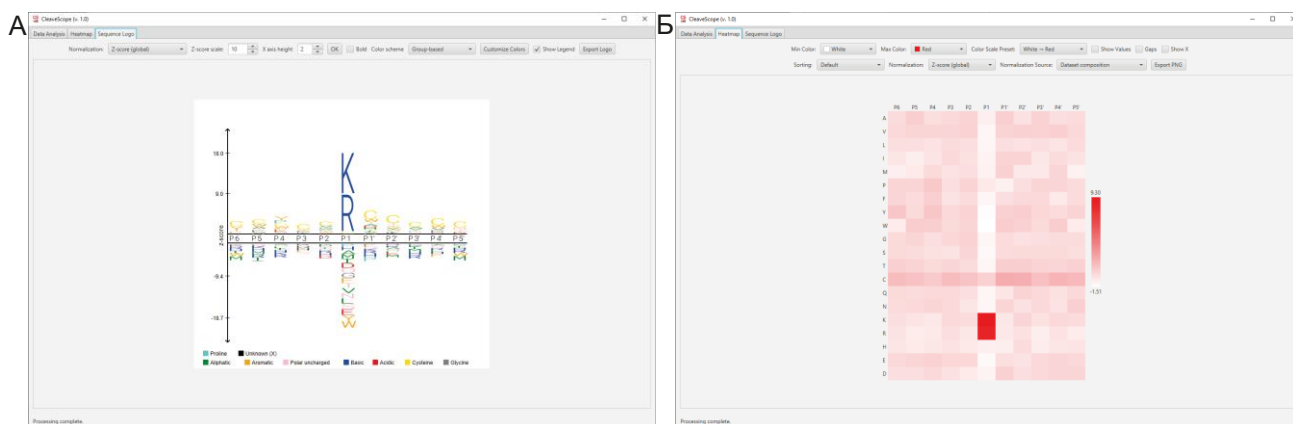


Рисунок 8. А – вкладка построения тепловой карты и выбора метода нормализации программы CleaveScore. Тепловая карта, отображающая абсолютное число вхождений аминокислотных остатков в позиции относительно сайтов гидролиза. Каждая строка соответствует определённой аминокислоте, а столбцы – позициям относительно точки расщепления. Б – вкладка построения логотипа последовательностей.

Однако простого представления в виде частотных распределений недостаточно для корректного анализа специфичности протеазы. Аминокислотные остатки в природе встречаются с различной частотой, и эти различия зависят от вида организма, типа ткани, физиологического состояния и других факторов. Например, согласно базе данных Swiss-Prot (база данных от 23 апреля 2025 года [105]), остатки триптофана в белках встречаются в среднем в 8,7 раза реже, чем остатки лейцина, а остатки метионина – примерно в 4 раза реже, в сравнении с лейцином. Таким образом, фоновые распределения аминокислот в белках являются заведомо неравномерными и биологически обусловленными, что необходимо учитывать при интерпретации данных.

На втором этапе обработки программа выполняет нормализацию полученных значений частоты встречаемости аминокислот с учётом фонового

распределения. Такое распределение может быть задано вручную в виде набора аминокислот с указанием частот, либо загружено из внешнего источника, например, FASTA-файла с белками из выбранного организма. Вместе с тем, как уже упоминалось, по умолчанию программа автоматически рассчитывает общее распределение аминокислот всех идентифицированных белков из загруженных XML-отчётов Mascot. Такой подход позволяет учесть именно те белки, которые присутствовали в образцах и были доступны для протеазы. Нормализация по аминокислотному составу исследуемого набора используется как стандартный режим. На основе нормализованных частот строятся тепловые карты и логотипы последовательностей, что позволяет визуализировать отклонения от фонового распределения и выявить позиционно-специфичные аминокислотные предпочтения фермента.

В рамках настоящей работы были проанализированы различные подходы к нормализации аминокислотных частот, как реализованных в существующих инструментах для визуализации специфичности (например, iceLogo, WebLogo, pLogo), так и основанные на общепринятых статистических методах анализа распределения частот в выборках. Часть из этих методов не предполагает прямого сравнения с фоновым аминокислотным составом, а часть использует способы нормализации, которые могут влиять на интерпретацию специфичности или приводить к смещению оценок частоты встречаемости, о чем будет сказано далее.

В связи с этим, в программе CleaveScore были реализованы четыре метода нормализации, которые, на наш взгляд, представляют собой наиболее логичные и информативные подходы к интерпретации позиционно-специфических частот аминокислот.

Первый метод – прямая нормализация (Direct normalization). Нормализация частоты каждой аминокислоты в позиции относительно её фоновой частоты в выбранном фоновом наборе (либо из текущего распределения аминокислот всех

идентифицированных белков, либо из внешней FASTA-базе), которая рассчитывается по формуле (3):

$$f_{ij}^{\text{norm}} = \frac{f_{ij}}{bg_i} \quad (3)$$

где, f_{ij}^{norm} – нормированное значение, f_{ij} – абсолютное число наблюдений аминокислоты i в позиции j , bg_i – общее количество аминокислоты i в фоновом наборе.

Это базовый подход, который позволяет выявлять аминокислоты, встречающиеся чаще или реже, чем это ожидается из их общей представленности в белковом наборе. Фон в данном случае позиционно-независимый, поскольку берется весь набор аминокислот.

Второй метод нормализации – Z-оценка (глобальная) (Z-score (global)). Двухэтапный подход: полученная после прямой нормализации матрица стандартизируется по всей видимой области (всем позициям и выбранным аминокислотам). Такой метод облегчает визуальное сравнение позиций, выделяя наиболее отклоняющиеся аминокислоты и рассчитывается по формуле (4):

$$Z_{ij}^{\text{global}} = \frac{f_{ij}^{\text{norm}} - \mu}{\sigma} \quad (4)$$

где, Z_{ij}^{global} – Z-оценка аминокислоты i в позиции j , μ – среднее значение f_{ij}^{norm} по всем аминокислотам во всех позициях, σ – стандартное отклонение f_{ij}^{norm} по всем аминокислотам во всех позициях, f_{ij}^{norm} – нормированное значение из прямой нормализации.

Третий метод – Z-оценка (по аминокислотным остаткам) (Z-score (row)). В этом режиме нормализация проводится отдельно для каждой аминокислоты по

всем позициям. Он позволяет выявлять позиционную избирательность внутри остатков одной аминокислоты, независимо от их общей частоты. Однако он не позволяет напрямую сравнивать между собой различные аминокислоты, а показывает, в каких позициях данный остаток встречается чаще или реже, чем в среднем для этой же аминокислоты по всем позициям. Поэтому данный режим может быть полезен как вспомогательный инструмент, но не рекомендуется в качестве основного при интерпретации специфичности протеаз. Метод нормализации рассчитывается по формуле (5), где значения нормализуются по позициям в пределах одной строки (одного остатка):

$$Z_{ij}^{row} = \frac{f_{ij} - \mu_i}{\sigma_i} \quad (5)$$

где, Z_{ij}^{row} – Z-оценка аминокислоты i в позиции j , μ_i – среднее значение f_{ij}^{norm} по ряду аминокислоты i , σ_i – стандартное отклонение f_{ij}^{norm} по ряду аминокислоты i , f_{ij} – абсолютное число наблюдений аминокислоты i в позиции j .

Четвертый метод – процентное соотношение аминокислот (по аминокислотным остаткам) (Percentage (per residue)). Этот подход также не использует фоновое распределение, а позволяет изучить различия внутри одного аминокислотного ряда. В данном режиме абсолютные частоты аминокислот нормализуются так, что сумма всех значений в каждой позиции составляет 100%. Частота встречаемости одной аминокислоты выражается в процентах относительно всех наблюдений для этой аминокислоты в каждой позиции, что позволяет сравнивать распределение остатков внутри позиции и считается по формуле (6):

$$P_{ij} = \frac{f_{ij}^{norm}}{\sum_k f_{ik}^{norm}} \times 100\% \quad (6)$$

где, P_{ij} процентное соотношение аминокислоты i в позиции j , f_{ij}^{norm} – нормированное значение из прямой нормализации, $\sum_i f_{ik}^{\text{norm}}$ – сумма по строке k (всем позициям) аминокислоты i .

Использование собственных алгоритмов нормализации обусловлено недостатками существующих методов, поскольку интерпретация данных сильно зависит от выбранного алгоритма. В качестве примера для сравнения была выбрана программа IceLogo, поскольку в отличие от остальных программ (перечисленных в разделе 2.4.3), основное ее предназначение – анализ протеолитической специфичности. В этой программе для расчёта Z-оценки используется нормальное приближение к биномиальному распределению. При малом числе наблюдений (например, для редких аминокислот или небольших выборок) это приближение некорректно и может давать искаженные значения p-values и Z-оценок. Кроме того, предлагаемый алгоритм «Percentage difference» (PD, разность между наблюдаемой и фоновой частотой, в процентах), может искажать данные ферментативной специфичности при низкой фоновой частоте. К примеру, увеличение частоты встречаемости в исследуемой выборке по сравнению с фоном с 1% до 6% и с 10% до 15% в обоих случаях даст разницу в +5%. Однако для первого случая частота встречаемости («Fold change», FC) увеличится в 6 раз, в сравнении с фоном, а для второго случая в 1,5 раза. Аналогично, использование FC может приводить к занижению различий: например, увеличение с 10% до 25% и с 0,5% до 1% даст $FC = 2,5$ и 2 соответственно. Такую разницу в некоторых случаях оказывается несущественной, и, следовательно, не позволяет увидеть различия между значимыми и незначимыми аминокислотами.

Для оценки работоспособности программы, а также сравнения получаемых результатов с существующими инструментами, она была протестирована на ферменте с хорошо охарактеризованной специфичностью. В качестве модельного фермента был выбран трипсин – один из наиболее изученных протеолитических ферментов, широко применяемый в биохимии и считающийся «золотым

стандартом» в протеомике. Трипсин преимущественно гидролизует пептидные связи после остатков аргинина и лизина [147].

Для анализа были использованы данные поиска в системе Mascot, без указания специфичности фермента, образцов фракций сыворотки человека. Результаты были обработаны с помощью программы CleaveScore, в которой был сформирован список выровненных фрагментов, использованных далее в качестве входных данных для анализа в IceLogo. Поскольку в качестве фонового распределения программа требует список выровненных фрагментов, таковой был получен с использованием собственного скрипта. Следовательно, в качестве фонового распределения были использованы последовательности только тех белков, которые были идентифицированы. В результате были получены логотипы последовательностей, рассчитанные двумя методами: PD и FC (Рисунок 9А и 9Б соответственно).

При использовании PD аминокислоты, соответствующие специфичности трипсина, отчётливо выделяются в предполагаемой позиции гидролиза (позиция 0, что согласуется с известной ферментативной специфичностью. Однако в логотипе, построенном по шкале FC, эти остатки визуальнo практически не доминируют несмотря на то, что их значения достигают максимальных величин среди всех аминокислот в этой позиции (позиция 0). Такое расхождение обусловлено особенностями подсчета, при котором значение суммы FC всех аминокислот в одной позиции могут быть примерно равны значениям суммы FC всех аминокислот в любой другой позиции. В результате даже если в позиции сайта предполагаемого гидролиза наблюдается высокая специфичность (например, только один или два остатка с $FC > 2$), эта позиция может выглядеть менее информативной, чем соседние позиции, в которых множество аминокислот имеют незначительное, но положительное отклонение (например, $FC \approx 0.4-0.8$).

Метод подсчета FC лучше использовать для визуализации с использованием тепловых карт, однако в IceLogo тепловые карты рассчитываются другим методом

с использованием Z-оценки. Такой метод не позволяет отчетливо визуализировать различия в частоте встречаемости аминокислот в различных позициях (Рисунок 9В).

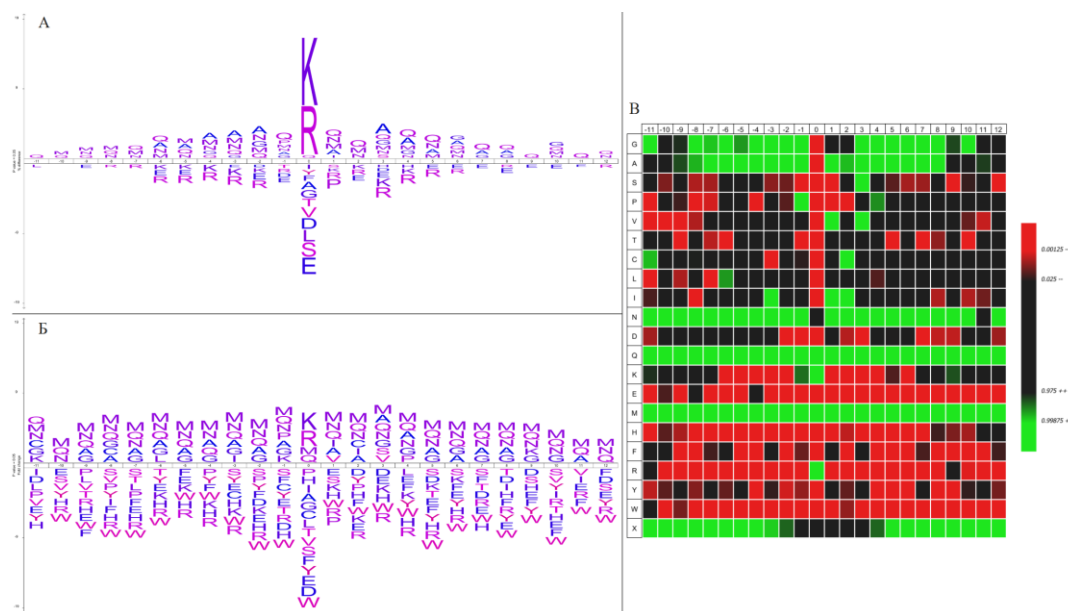


Рисунок 9. А – логотип последовательностей, построенный по экспериментальным данным для трипсина методом «Percentage Difference» в программе IceLogo. Б – логотип последовательностей, построенный по тем же данным методом «Fold Change» в программе IceLogo. В – тепловая карта, построенная по экспериментальным данным для трипсина в программе IceLogo. Позиции выровнены относительно сайта гидролиза: 0 соответствует месту гидролиза, отрицательные значения – остаткам N-концевого фрагмента, положительные – С-концевого; диапазон отображения: от -11 до +11.

При использовании программы CleaveScore для построения логотипов последовательностей использовались два метода: Z-оценка (глобальная) и Z-оценка (по аминокислотным остаткам). В обоих случаях в качестве фона использовался состав аминокислот в белках, идентифицированных в том же эксперименте, как и при анализе с использованием программы IceLogo.

На полученных логотипах последовательностей глобальной Z-оценки (Рисунок 10А) остатки аргинина и лизина отчетливо доминируют в позиции предполагаемого сайта гидролиза (P_1), что полностью соответствует известной специфичности трипсина. При этом остальные позиции (P_2 , P_3 , P_1' и т.д.) не имеют высокой частоты встречаемости, что указывает. Как было сказано ранее, такая

визуализация позволяет оценить значимость отклонения наблюдаемой частоты каждой аминокислоты от среднего значения всех аминокислот. При визуализации с использованием метода Z-оценки по аминокислотным остаткам (Рисунок 10Б), можно наблюдать схожую картину с методом FC в IceLogo, где суммарная высота аминокислот во всех позициях примерно одинаковая. Однако в отличие от IceLogo, такой метод позволяет более точно выявить значимые аминокислоты внутри одного сайта. В частности, позиции P_1 доминируют только две специфичные для трипсина аминокислоты. Кроме того, данный метод позволил подтвердить ограничение, отмеченное в ряде исследований показано: фермент не гидролизует пептидную связь после аргинина или лизина, если в позиции P_2 находится пролин [90]. Такое наблюдение визуализируется отчётливее при использовании второго подхода, в сравнении с первым.

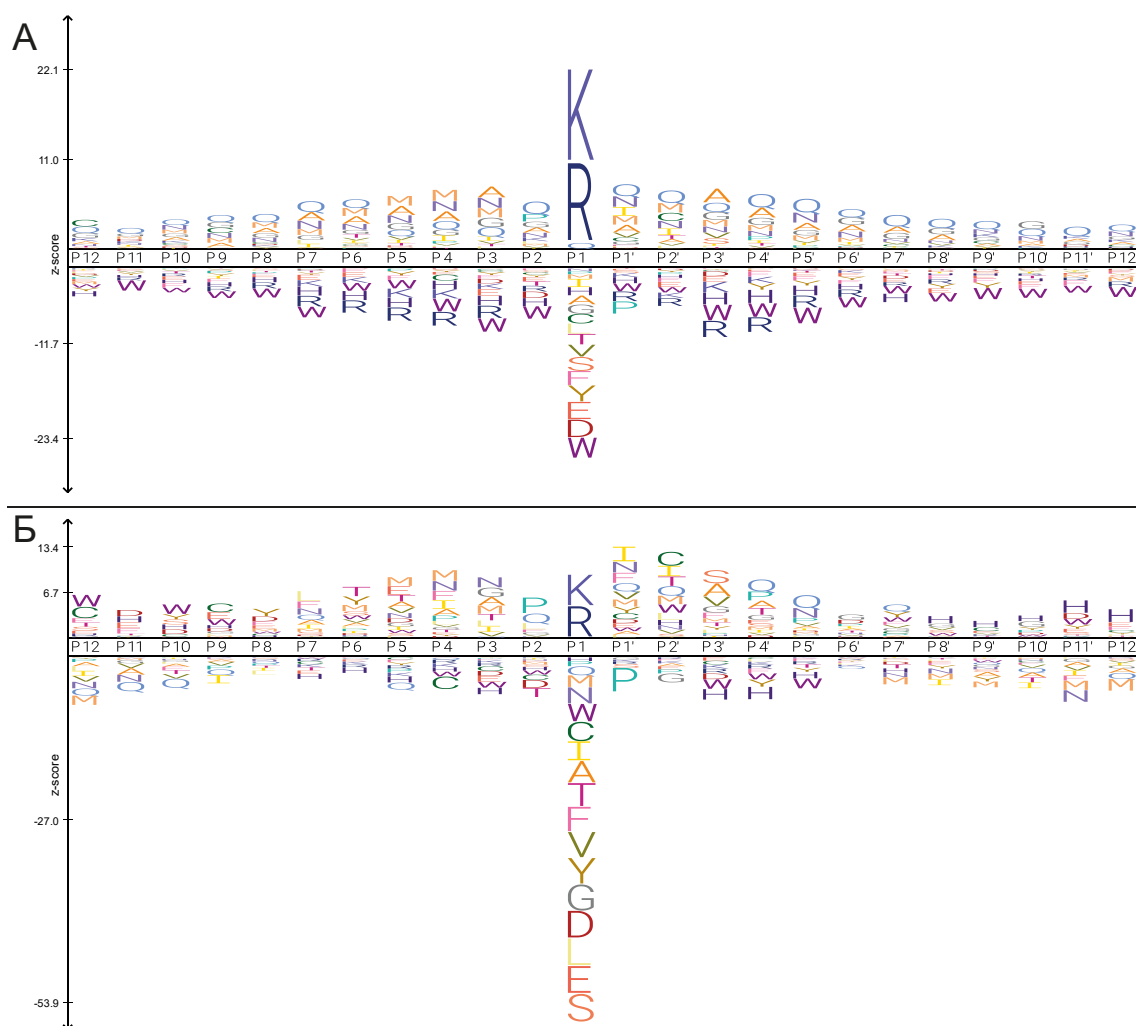


Рисунок 10. Логотип последовательностей, построенный по экспериментальным данным для трипсина с использованием программы CleaveScore. Значения рассчитаны на основе глобальной Z-оценки (А) и с использованием метода Z-оценки по аминокислотным остаткам (Б). Данные нормированы на аминокислотный состав белков, идентифицированных в эксперименте.

Как отмечалось выше, помимо логотипов последовательностей, CleaveScore также позволяет визуализировать результат в виде тепловых карт, отражающих частотное распределение аминокислот в позициях относительно точки разреза (P₁₂–P_{12'}). На рисунке 11 представлены тепловые карты исходных частот (А) и нормализованных значений (Б).

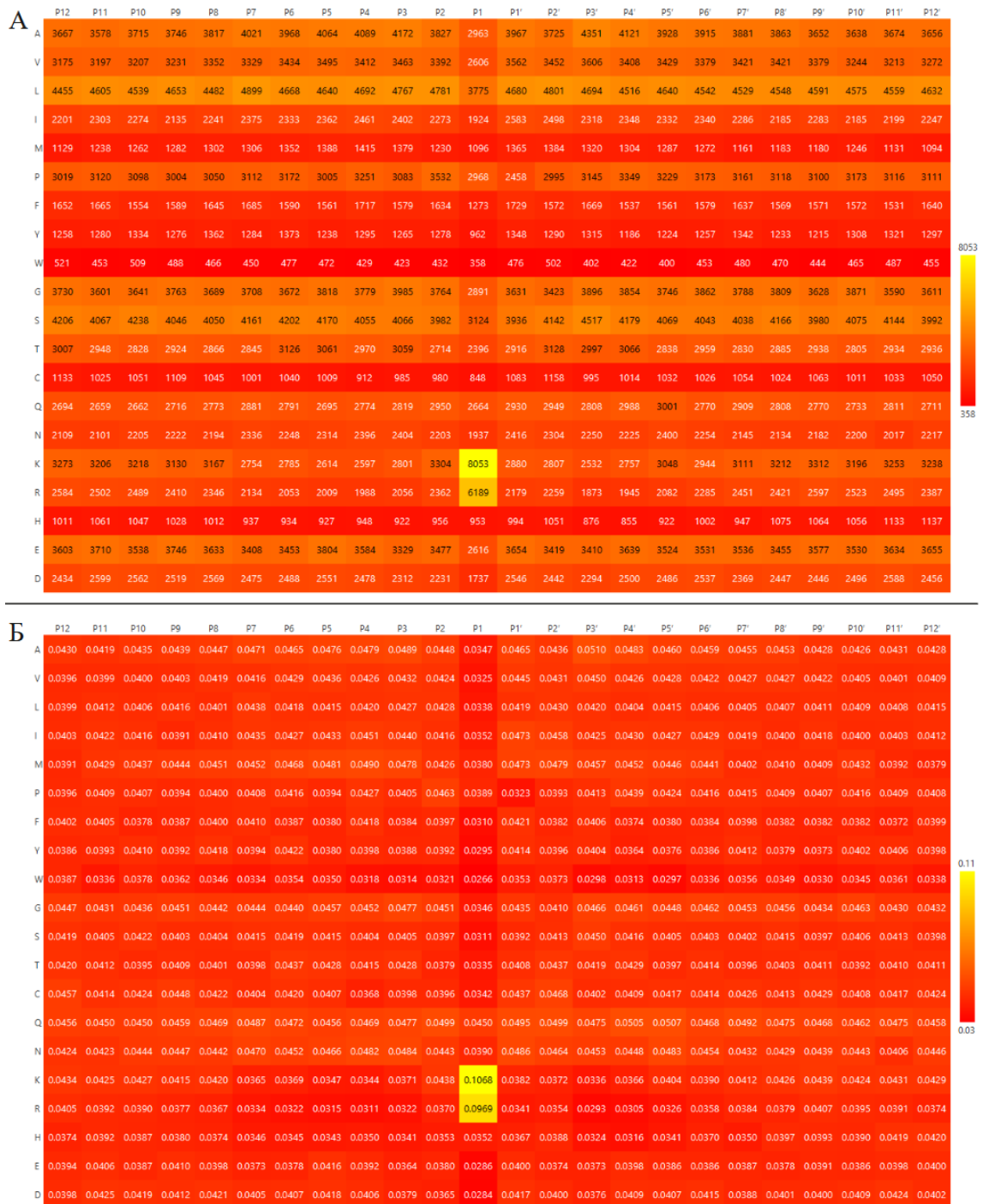


Рисунок 11. Тепловая карта, построенная по экспериментальным данным для трипсина с использованием программы CleaveScore для исходных частот (А) и нормализованных аминокислотный состав белков, идентифицированных в эксперименте (Б).

На тепловой карте исходных данных (то есть единиц аминокислот, наблюдаемых в различных позициях) наибольшие значения наблюдаются для остатков лизина и аргинина. Вместе с тем, например, лейцин имеет высокую частоту во всех позициях, что связано с его общей избыточностью в составе белков

и не отражает специфичность гидролиза трипсином. В отличие от логотипов последовательностей, где подобный эффект не всегда очевиден, именно тепловые карты позволяют сразу увидеть, что целые ряды часто встречающихся аминокислот окрашены значительно интенсивнее, чем ряды редких остатков. Такая визуализация наглядно подчеркивает необходимость нормализации, которая «отделяет» общий фон от специфичных для фермента аминокислот.

После нормализации, когда учитывается общее количество каждой аминокислоты по всем позициям, ясно прослеживается избирательное накопление лизина и аргинина в позиции P_1 . В то же время большинство других аминокислот распределены по позициям относительно равномерно, что указывает на их неспецифический характер. Для улучшения читаемости и облегчения интерпретации данных дополнительно применяется глобальная Z -нормализация, при которой наблюдается аналогичная картина распределения (Рисунок 12А). В данном режиме значения Z -оценка для лизина и аргинина в позиции P_1 достигают 11,56 и 9,83 соответственно, тогда как для остальных аминокислот во всех позициях значения не превышают 2. Все эти данные соответствуют литературным сведениям о специфичности трипсина.

Визуализация процентного соотношения аминокислот (Рисунок 12Б) также подтверждает специфичность трипсина: только в позиции P_1 значения для остатков аргинина и лизина превышают 10%, тогда как для остальных позиций и аминокислот частоты колеблются в пределах 3,5–4,5%. Такой метод также позволяет отчётливо выявить смещение распределения в сторону определённых остатков исключительно в позиции предполагаемого разреза, при одновременной равномерности распределения прочих аминокислот.

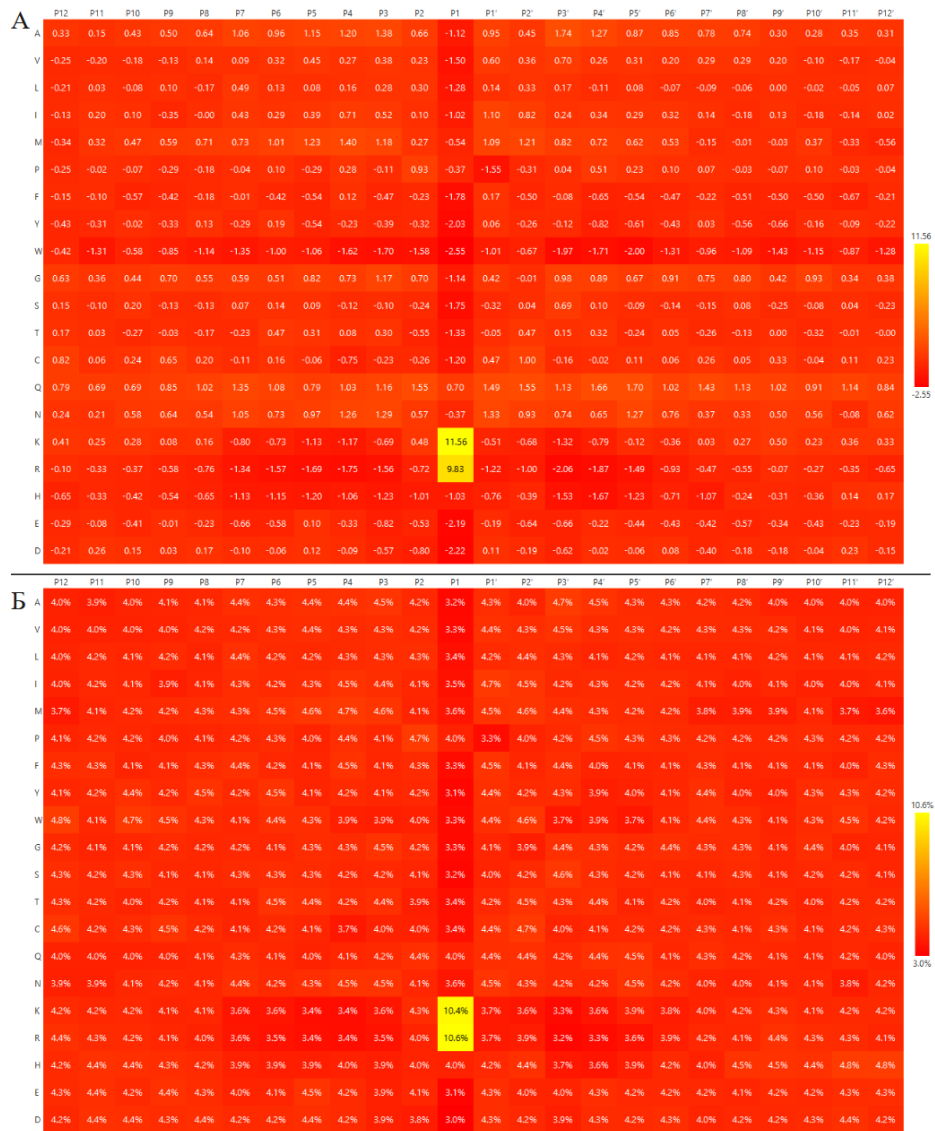


Рисунок 12. Тепловая карта, построенная по экспериментальным данным для трипсина с использованием программы CleaveScore. А – глобальная Z-оценка, Б – процентное соотношение аминокислот, нормированное аминокислотный состав белков, идентифицированных в эксперименте.

Однако такие методы не лишены недостатков: если в наборе исходных (ненормализованных) данных содержатся аминокислоты с низкой фоновый частотой, то даже незначительное их абсолютное количество может приводить к завышенным нормализованным значениям. Так, при анализе образца калибровочного стандарта белков клеточной линии HeLa, гидролизованного трипсином, фоновые частоты аминокислот М, Q и N ниже, чем их частота встречаемости в полученном наборе пептидов. В результате после нормализации значения этих аминокислот во всех позициях могут быть завышенными, что отчётливо проявляется при использовании глобальной Z-оценки (Рисунок 13А, Б),

используя оба метода визуализации. При этом данный эффект носит систематический характер, легко распознаётся визуально и устраняется при применении альтернативных методов нормализации, где подобного смещения не наблюдается (Рисунок 13В).

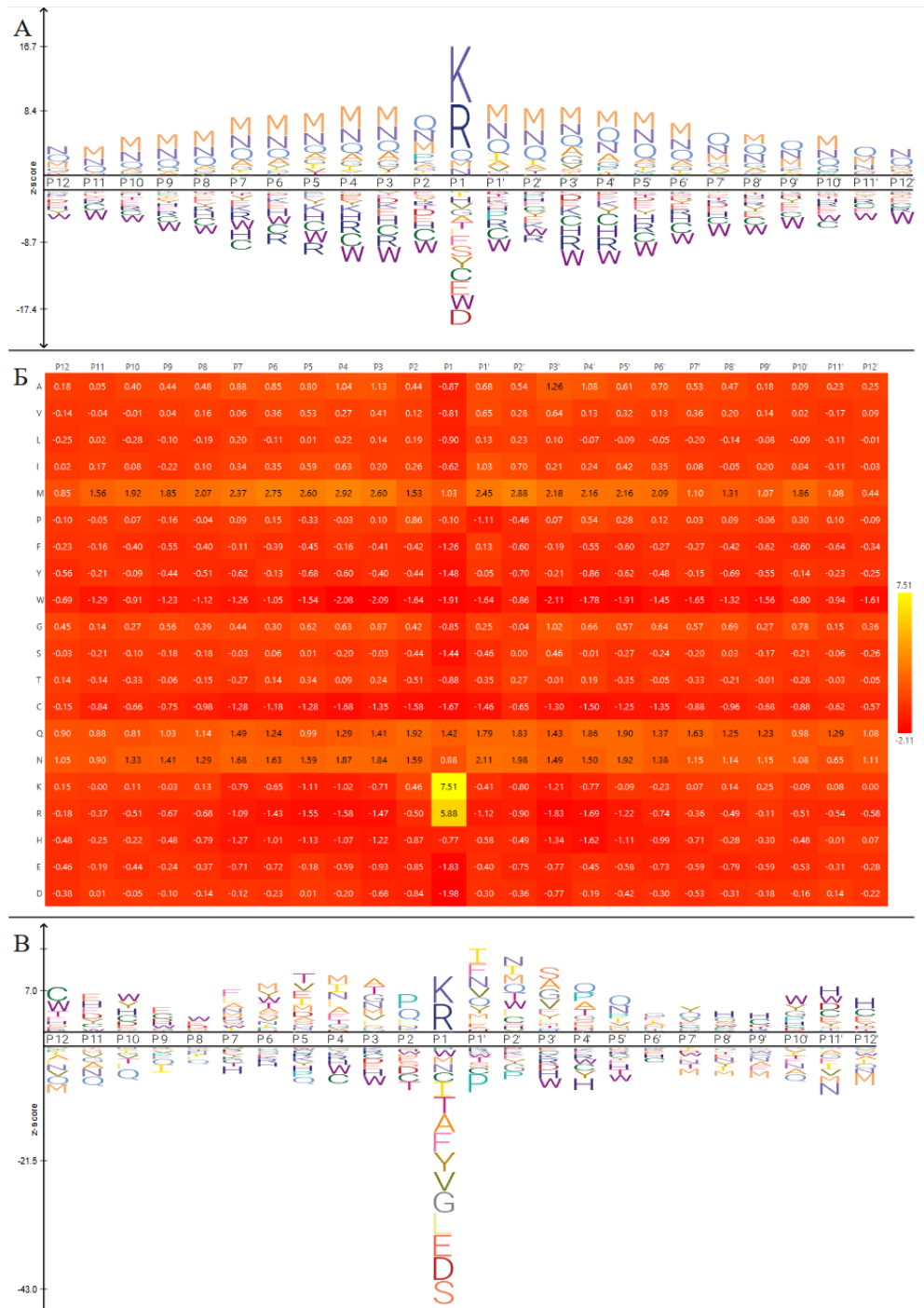


Рисунок 13. А – логотип последовательностей, построенный по экспериментальным данным при анализе образца калибровочного стандарта белков клеточной линии HeLa с использованием метода глобальной Z-оценки, нормированной на аминокислотный состав белков, идентифицированных в эксперименте. Б – анализ тех же данных, но представленный в виде тепловой карты. В – анализ тех же данных, но представленный в виде логотипа последовательностей с использованием метода Z-оценки по аминокислотным остаткам.

Именно по причине различий в результатах, получаемых при использовании разных методов нормализации, в существующих инструментах анализа специфичности, таких как IceLogo, реализовано несколько альтернативных алгоритмов расчёта. Это подчёркивает необходимость комплексного подхода и использования нескольких методов для оценки протеолитической специфичности, а не использование единственного метода.

Несмотря на отдельные ограничения CleaveScore, реализованные методы нормализации и визуализации позволяют наглядно сравнивать частоту каждого остатка в каждой позиции с его фоновой частотой, выявляя специфичные сайты. А использование нескольких подходов к нормализации позволяет компенсировать ограничения каждого из них. Таким образом, разработанная программа обеспечивает автоматизированный, воспроизводимый и статистически обоснованный анализ специфичности протеаз на основе данных, полученных из поисковой системы Mascot, не требующих предварительной обработки.

4.4. Анализ специфичности протеаз на основе данных LC-MS/MS

Для точного определения специфичности протеазы ключевым фактором является разнообразие аминокислотного состава белковых субстратов, тогда как их происхождение играет менее существенную роль. С целью обеспечения максимально широкого спектра потенциальных сайтов расщепления в настоящей работе в качестве источников белков использовались лизат ткани печени крыс и сыворотка крови человека, что позволило сформировать широкий аминокислотный набор. С учётом ограниченного охвата протеома, обусловленного используемой системой LC-MS/MS, образцы печени и сыворотки предварительно фракционировали с помощью эксклюзионной хроматографии (SEC), что обеспечило повышение числа идентифицируемых пептидов и, следовательно, большее количество идентифицированных сайтов гидролиза.

В результате анализа специфичности фермента Vpr было исследовано 8 образцов, из которых идентифицировано 314 белков. Для протеазы L5 анализ включал 9 образцов, в которых идентифицировано 526 белков. Число идентифицированных уникальных сайтов гидролиза для Vpr составило 1212, в то время как для L5 – 3442. При анализе учитывались позиции от P₆ до P₅'.

На рисунке 14 представлены тепловые карты абсолютных (исходные) значений частоты встречаемости аминокислот в позициях P₆–P₅'. Интенсивность цвета отражает относительную частоту соответствующего аминокислотного остатка в данной позиции: более тёмные оттенки бордового соответствуют более высоким значениям.

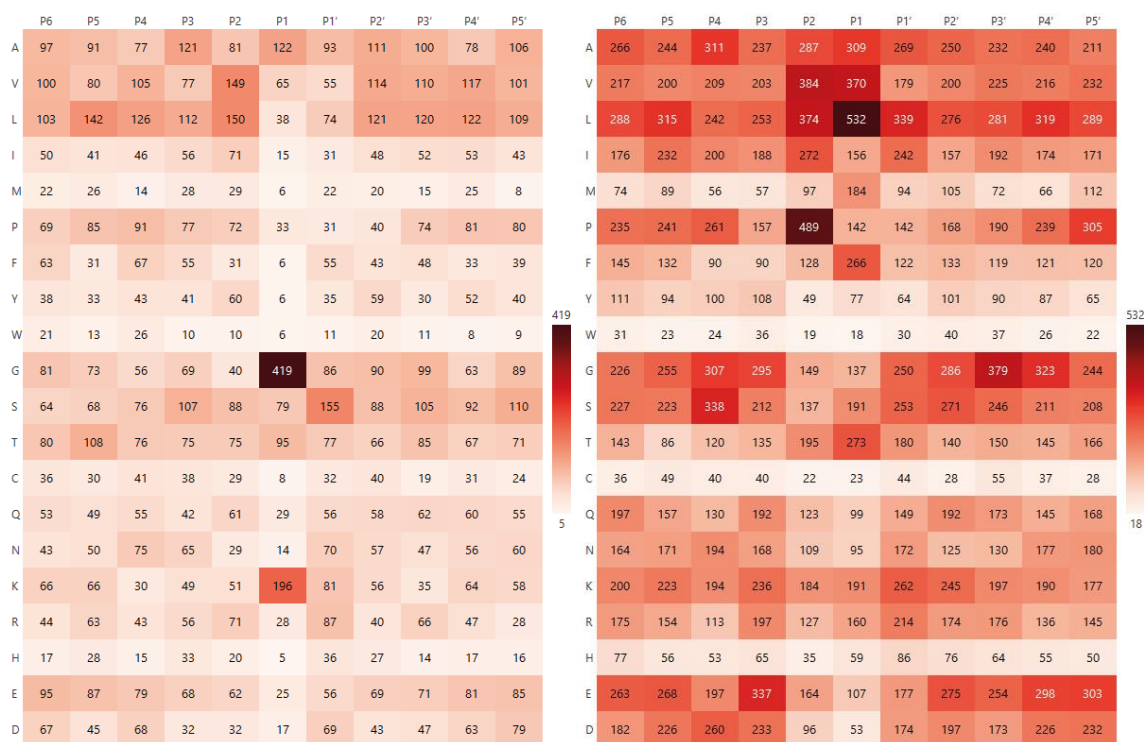


Рисунок 14. Тепловые карты исходных значений частоты встречаемости аминокислот после гидролиза исследуемыми ферментами (Vpr слева, L5 справа).

При анализе исходных данных частот встречаемости аминокислот можно наблюдать, что Vpr проявляет чётко выраженное предпочтение к расщеплению по остаткам глицина – 419 остатков аминокислот (34,5% от всех аминокислот в положении P₁), и лизина – 196 аминокислотных остатков (16,1%). В случае

фермента L5 гидролиз преимущественно осуществляется по остаткам неполярных аминокислот: лейцина – 532 а.о. (14,8% от всех аминокислот в положении P₁), валина – 370 а.о. (10,6%), аланина – 309 а.о. (9%) и фенилаланина 266 а.о. (7,4%), а также по остатку полярной незаряженной аминокислоты – треонина – 273 а.о. (8%).

Как уже обсуждалось ранее, аминокислоты существенно различаются по частоте встречаемости в зависимости от используемой выборки. Поэтому для дальнейшей оценки специфичности фермента была применена нормализация данных с использованием глобальной Z-оценки. В качестве фонового набора данных использовались либо аминокислотный состав всех белков из базы данных Swiss-Prot, либо аминокислотный состав белков, идентифицированных в ходе нашего эксперимента. На рисунках 15 и 16 представлены тепловые карты для ферментов B1p и L5 соответственно.



Рисунок 15. Тепловые карты специфичности фермента B1p, нормированные на аминокислотный состав базы данных Swiss-Prot (слева) и по аминокислотному составу экспериментального набора данных (справа). Данные представлены в виде Z-оценок.

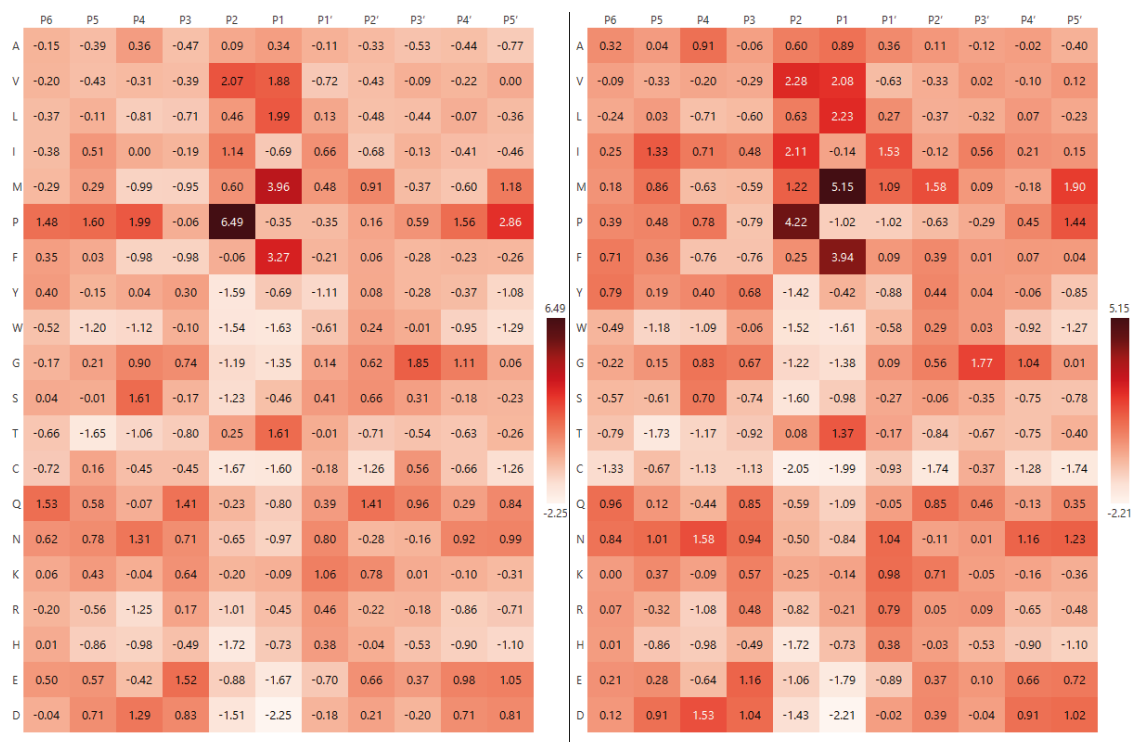


Рисунок 16. Тепловые карты специфичности фермента L5, нормированные на аминокислотный состав базы данных Swiss-Prot (слева) и по аминокислотному составу экспериментального набора данных (справа). Данные представлены в виде Z-оценок.

Как видно на рисунках 15 и 16, существенных различий между нормализацией на базу Swiss-Prot и на экспериментальный набор белков не наблюдается. Этот вывод дополнительно подтверждается результатами корреляционного анализа, в котором сравнивались аминокислотные составы белков из базы Swiss-Prot и таковые из белков, идентифицированных в наших экспериментах. Коэффициент корреляции Спирмена составил 0,91 для B1p и 0,93 для L5, а коэффициент Пирсона – 0,93 и 0,96 соответственно. Такое наблюдение указывает на то, что аминокислотный состав идентифицированных белков в целом хорошо отражает распределение аминокислот в широкой базе данных Swiss-Prot, следовательно, обе выборки можно использовать для оценки специфичности фермента. Тем не менее, нормализация на аминокислотный состав непосредственно тех белков, которые подвергались ферментативному гидролизу, – то есть тех остатков, которые фактически были доступны для действия протеазы, – обеспечивает более точное представление о её специфичности. Поэтому в дальнейшем был использован этот аминокислотный состав.

Также на основе нормализованных данных были построены логотипы последовательностей (рисунок 17). Логотип последовательностей специфичности фермента Vlr отражает выраженное доминирование остатков глицина и лизина в позиции P_1 , что также наглядно подтверждает высокую специфичность к этим аминокислотам. В других позициях выраженных предпочтений не наблюдается.

Для фермента L5 логотип отражает более широкую специфичность: в позиции P_1 наблюдается преимущественное накопление метионина, фенилаланина и лейцина, а также менее выраженное присутствие валина и треонина.

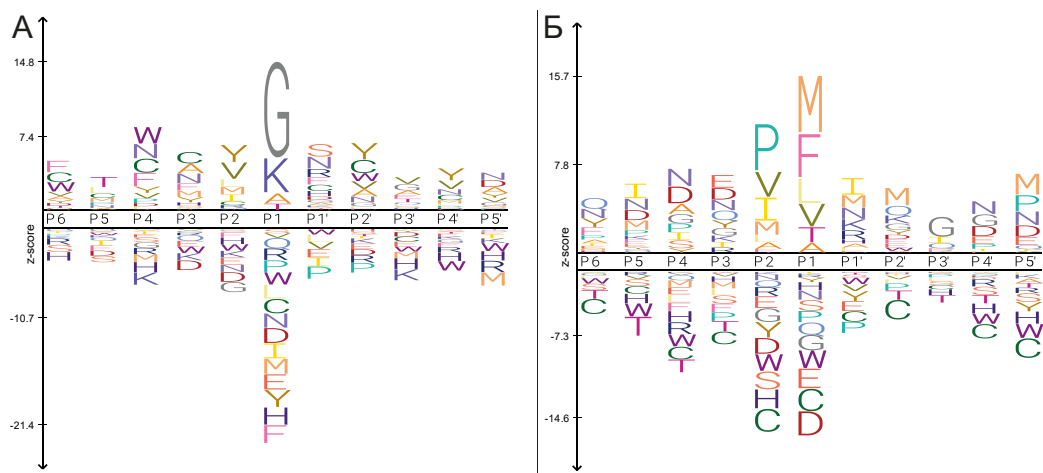


Рисунок 17. Логотипы последовательностей нормализованных данных для исследуемых ферментов: Vlr (А) и L5 (Б).

Как можно заметить, после нормализации на аминокислотный состав идентифицированных белков профиль специфичности Vlr практически не изменился. Основными аминокислотами в положении P_1 остались глицин, со значением Z-оценки 9,66 и лизин – 3,51, при среднем значении остальных аминокислот по всей матрице в диапазоне от -1,83 (низкая представленность) до 1,75 (умеренная представленность).

Полученные данные согласуются с ранее опубликованными результатами других авторов [148,149], в которых специфичность фермента оценивалась с использованием пентаглициновых пептидов. Аналогичный профиль расщепления был также обнаружен в нашем предыдущем исследовании, проведенном с

использованием набора коммерческих белков и MALDI-TOF масс-спектрометрии. Однако подобные подходы ограничены числом доступных субстратов, в то время как для проведения статистически обоснованного анализа требуется не менее 30 различных субстратов [150].

В рамках настоящей работы были идентифицированы сотни сайтов гидролиза, что позволило сформировать значительно более репрезентативный и количественно обоснованный набор данных. Благодаря этому было подтверждено, что фермент Vpr проявляет чёткую и ограниченную специфичность к двум аминокислотам – преимущественно к глицину, а в меньшей степени к лизину – и практически не расщепляет другие остатки в позиции P₁.

В случае фермента L5 нормализация существенно изменила профиль предпочтений к аминокислотным остаткам в сайте гидролиза. Для сравнения нормализованных данных с использованием глобальной Z-оценки и исходных данных частот встречаемости аминокислот, для всех аминокислот в положении P₁ исходных данных («локальная») была рассчитана Z-оценка по формуле (7):

$$Z_{iP1} = \frac{f_{iP1} - \mu}{\sigma} \quad (7)$$

где, Z_{iP1} – Z-оценка аминокислоты i в позиции $P1$, μ – среднее значение частот f_{iP1} по всем аминокислотам в позиции $P1$, σ – стандартное отклонение частот f_{iP1} по всем аминокислотам в позиции $P1$, f_{iP1} – абсолютная частота (или число вхождений) аминокислоты i в позиции $P1$.

Такая нормализация необходима, поскольку исходные частоты представляют собой абсолютные значения (количество а.о.), в то время как нормализованные значения выражаются в единицах стандартного отклонения (Z-оценка). При этом применение «локальной» Z-оценки по аминокислотам не изменяет порядок частоты встречаемости а.о., отражённый в необработанных данных.

Важно отметить, что численные значения глобальной и «локальной» Z-оценки не сопоставимы между собой. Несмотря на то, что оба метода используют стандартное отклонение в качестве меры вариации, они нормализуют данные в разных контекстах – по всем аминокислотам (глобально) или отдельно по каждой (локально). Поэтому величины Z не отражают кратности или абсолютных различий между аминокислотами, а лишь степень их отклонения от соответствующего среднего значения в выбранной модели нормализации. В связи с этим можно делать выводы лишь о том, какие аминокислоты были представлены чаще, а какие реже.

Результаты, полученные с использованием такой «локальной» Z-оценки (на основе частот в позиции P₁), показали наибольшее отклонение у лейцина ($Z = 3,04$) и валина ($Z = 1,70$), тогда как метионин ($Z = 0,15$) имеет умеренное значение (рисунок 18 слева). Напротив, при глобальной нормализации, где учитывается фоновое распределение аминокислот в белках, подвергшихся гидролизу, наиболее значимыми стали метионин ($Z = 5,15$) и фенилаланин ($Z = 3,93$), тогда как частота встречаемости лейцина и валина была менее выражена ($Z = 2,23$; $Z = 2,07$) (рисунок 18 справа). Таким образом, при локальной нормализации доминирующими субстратами для фермента кажутся лейцин, валин и аланин, тогда как при учете аминокислотного распределения наиболее специфичными остатками являются метионин и фенилаланин, лейцин, валин и треонин. Это еще раз подчеркивает, что ферментативная специфичность очень зависит от метода обработки данных.

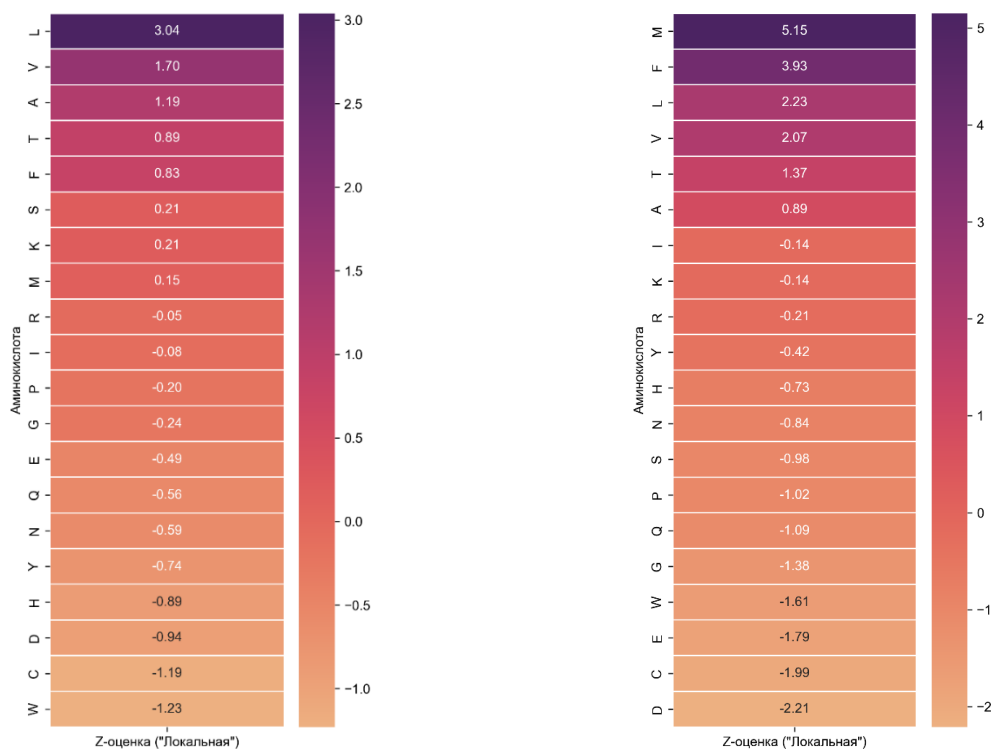


Рисунок 18. Частота встречаемости аминокислот в позиции P_1 при анализе специфичности фермента L5 без учёта (слева) и с учётом (справа) аминокислотного состава набора данных. В обоих случаях значения представлены в виде Z-оценок.

Одним из наиболее неожиданных наблюдений при анализе специфичности фермента L5 стала стабильно высокая частота встречаемости пролина в позиции P_2 , что заметно на тепловой карте (рисунок 16) и логотипе последовательностей (рисунок 17Б). Эта аминокислота имеет один из самых высоких Z-оценок среди всех позиций ($Z=4,22$), что может указывать на его возможную роль в распознавании субстрата. Однако простая визуализация или оценка с использованием вышеупомянутых методов не позволяет оценить вклад пролина, для чего нужен анализ мотивов окружения сайта гидролиза.

4.5. Статистический анализ мотивов в окрестности сайтов разреза, сравнительный анализ с трипсином.

Как уже отмечалось, в ходе анализа специфичности протеазы L5 наблюдалась высокая частота встречаемости пролина в позиции P_2 . В то время как для ряда сериновых протеаз пролин, напротив, известен как ингибитор гидролиза – особенно при локализации в позициях P_1 или P_1' – вследствие своей циклической

структуры, ограничивающей подвижность пептидной связи [151], его расположение в P_2 может выполнять иную роль. На основании этого наблюдения была выдвинута гипотеза о том, что фермент распознает не отдельные аминокислотные остатки, а определённые последовательности – например, дипептидные мотивы, такие как P–M, P–F и т.д.

Следует отметить, что большинство алгоритмов, предназначенных для анализа ферментативной специфичности, оценивают либо избыточность/дефицит отдельных аминокислот относительно их общего содержания в базе данных, либо позиционную предпочтительность конкретных остатков, независимо от соседних позиций. Так, несмотря на то что, например, IceLogo использует позиционно-зависимый фон (reference set), сформированный методом «скользящего окна», данный подход, по сути, представляет собой нормализацию на глобальный аминокислотный состав. Такие методы не позволяют выявлять контекстные закономерности – то есть мотивы, возникающие в результате специфического сочетания остатков в смежных позициях (например, PL, PV, PM и др.).

В частности, для исследуемого фермента в позиции P_2 наблюдается частое присутствие пролина, однако без анализа комбинаций остатков (например, P_2 – P_1) невозможно установить, отражает ли это предпочтение к определённому контексту (например, мотиву PL), либо лишь высокую частоту самого пролина в данной позиции.

С целью проверки данной гипотезы был проанализирован аминокислотный состав в позиции P_1 (непосредственно перед сайтом гидролиза) в двух группах: (I) при наличии пролина в позиции P_2 и (II) при отсутствии пролина, то есть всех остальных случаев. В случае наличия специфичности к дипептидному мотиву типа P–X в первой группе следовало бы ожидать ограниченный набор остатков в позиции P_1 , с доминированием одного или нескольких предпочтительных аминокислот. Однако, как видно из диаграмм (рисунок 19), аминокислотное распределение в позиции P_1 в обеих выборках практически идентично. Пролин в

позиции P_2 сопровождается тем же разнообразием остатков в P_1 (группа I), что и в выборке без пролина (группа II) включая L, V, A, S, T и др., с незначительными колебаниями долей. Это свидетельствует об отсутствии выраженной специфичности типа Р–Х для протеазы L5.

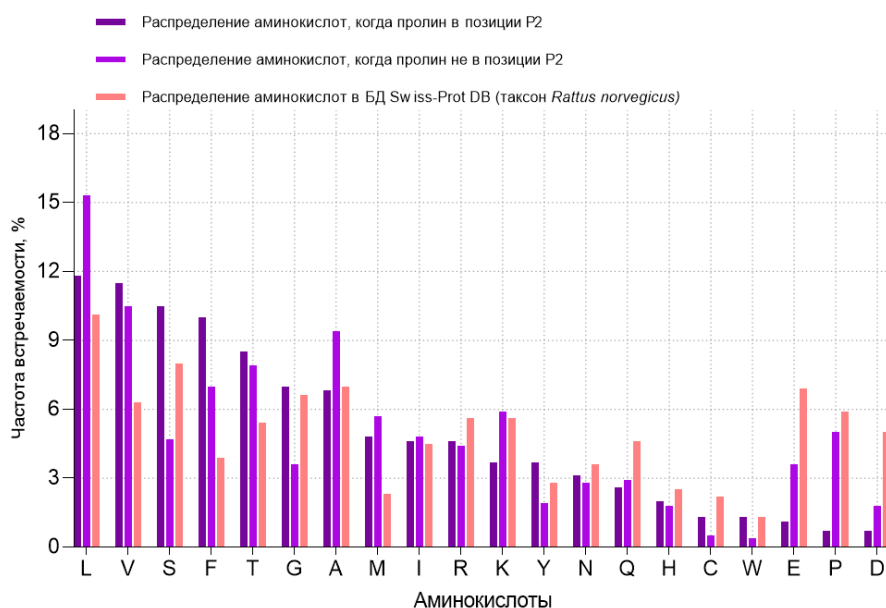


Рисунок 19. Распределение аминокислот в базе данных Swiss-Prot (*Rattus norvegicus*) и в позиции P_1 пептидов, гидролизированных ферментом L5 в двух вариантах: при наличии пролина в позиции P_2 и при его отсутствии. Данные представлены в виде процентной частоты встречаемости аминокислот.

В дополнение был проведён анализ наиболее часто встречающихся пар аминокислот в позициях P_2 и P_1 . На тепловой карте (рисунок 20 сверху) представлены абсолютные частоты сочетаний аминокислотных остатков: по оси Y – аминокислоты в позиции P_2 (один а.о. до сайта расщепления), по оси X – аминокислоты в позиции P_1 (непосредственно в точке расщепления). Несмотря на высокую суммарную частоту пролина в позиции P_2 , он образует сочетания с широким спектром аминокислот в позиции P_1 , включая L, V, F, S и T.

Для учёта аминокислотного состава, частоты пар аминокислот дополнительно нормировали относительно фоновых частот, рассчитанных по всем идентифицированным белкам (рисунок 20 снизу). После нормализации наибольшую относительную частоту показала комбинация I–M (Z-оценка = 8,86),

за которой следовали пары V–M (5,66), P–F (5,31) и P–M (4,9) и т.д. Тем не менее, пролин в позиции P_2 продолжал наблюдаться в сочетании с широким спектром остатков в позиции P_1 , что указывает на отсутствие выраженного предпочтения к какому-либо одному мотиву P_2 – P_1 .

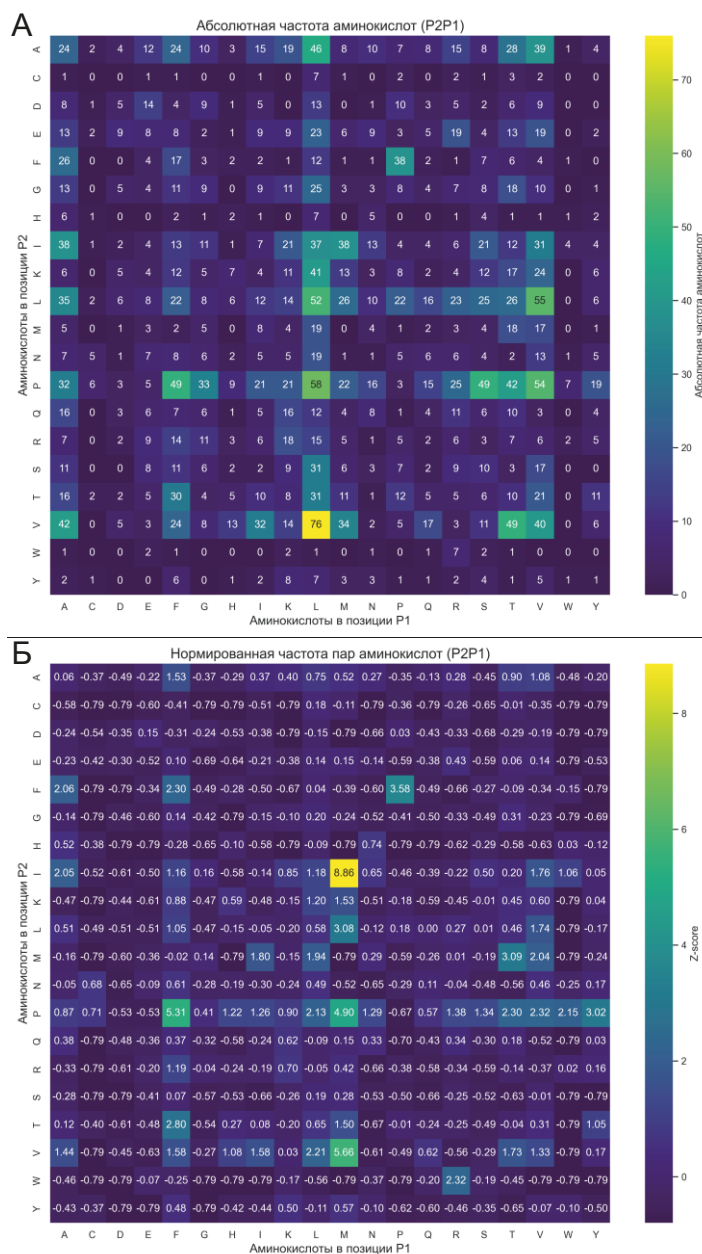


Рисунок 20. А – тепловая карта частоты встречаемости аминокислотных пар P_1 – P_2 , получаемых в результате гидролиза ферментом L5. Цветовая шкала отражает абсолютное число наблюдений каждой пары остатков в соответствующих позициях. Б – нормированная тепловая карта частоты встречаемости аминокислотных пар P_1 – P_2 . Цветовая шкала отражает число наблюдений каждой пары остатков в соответствующих позициях, выраженное в Z-score. По оси абсцисс указаны аминокислоты в позиции P_1 (аминокислота до сайта разреза), по оси ординат – аминокислоты в позиции P_2 (две аминокислоты до сайта разреза).

Протеолитическая активность протеазы L5 ранее была продемонстрирована с использованием казеина и синтетического флуорогенного субстрата Abz-Ala-Ala-Phe-pNA [41,48]. В частности, было показано, что фермент расщепляет пептидную связь после фенилаланина (Phe-pNA), а также гидролизует связи Gly–Gly в пептидогликановых мостиках *S. aureus*. Однако предыдущие исследования специфичности L5 были ограничены узким набором субстратов, что не позволяло провести всестороннюю оценку специфичности фермента.

На основании полученных результатов мы предполагаем, что L5 может стабилизировать субстраты в определённой конформации, способствующей эффективному связыванию с активным центром. Как было показано Ванхуфом и савт. [152], мотивы, содержащие пролин, часто формируют конформации, устойчивые к протеолизу, и расщепление в таких участках встречается относительно редко. В этом контексте высокая частота пролина в позиции P₂, наблюдаемая для L5, может свидетельствовать о способности фермента распознавать или стабилизировать структурно ограниченные мотивы, преодолевая врождённую устойчивость пролин-содержащих сайтов к расщеплению. Таким образом, выраженная представленность пролина в позиции P₂, вероятно, отражает структурные предпочтения фермента.

Эта гипотеза согласуется с нашими наблюдениями, показывающими, что глицин редко встречается в местах расщепления L5, что указывает на отсутствие специфичности к этому остатку. Следовательно, ранее описанное расщепление связей Gly–Gly в пентаглициновых мостиках пептидогликана *S. aureus* может отражать не истинную специфичность L5 к глицину, а скорее структурную доступность этого участка. Гибкая и вытянутая конформация пентапептидного мостика, вероятно, облегчает доступ к амидным связям, позволяя ферменту эффективно связываться и гидролизовать их даже при отсутствии выраженного предпочтения к остаткам глицина.

Подтверждением этой интерпретации служит частое появление мотивов из трех аминокислот G–X–M/F/L/T/V (где X – любой аминокислотный остаток), который был обнаружен 125 раз среди всех уникальных пептидов. Преобладание таких последовательностей также может отражать структурные предпочтения L5 – например, к гибким участкам, содержащим глицин и гидрофобный остаток в позиции расщепления. Тем не менее, для подтверждения значимости этого мотива необходим дополнительный анализ, например, структурное моделирование.

Также возможно, что специфичность L5 определяется в первую очередь распознаванием пролина в позиции P₂ и не зависит от природы остатка в позиции P₁. Если бы это было так, распределение аминокислот в P₁ при наличии пролина в P₂ должно было бы приближаться к фоновому распределению (например, представленному составом белков *Rattus norvegicus* в базе Swiss-Prot). Однако, как показано на рисунке 19, аминокислотные распределения в двух группах различаются.

Кроме того, общее число случаев расщепления пептидной связи при наличии пролина в P₂ (460) значительно уступает числу расщеплений без пролина в этой позиции (2771), что дополнительно свидетельствует о том, что пролин в P₂ не является единственным фактором, определяющим специфичность фермента. Более того, общее число событий, когда при гидролизе в позиции P₂ наблюдался пролин (489) существенно ниже по сравнению с числом событий без пролина в этой позиции (2955), что дополнительно указывает на то, что пролин в P₂ не является единственным фактором, определяющим специфичность L5.

Эти наблюдения позволяют предположить, что специфичность L5 может зависеть не только от взаимодействия с отдельными аминокислотами, но и от позиционных или конформационных особенностей активного центра фермента. Примечательно, что пролин способен стабилизировать изгибы и повороты в полипептидных цепях [153,154], что может способствовать формированию благоприятной для связывания конформации.

4.6. Сравнительный анализ специфичности ферментов B1p и L5 по длине идентифицированных пептидов

В ранних исследованиях специфичность B1p изучалась с использованием отдельных белков и коммерчески доступных синтетических субстратов, содержащих флуоресцентные или хромогенные метки [51,148,149]. В первом случае, как уже было сказано ранее, одного белка недостаточно для полного определения специфичности. То же утверждение справедливо для синтетических субстратов, аминокислотный репертуар которых ограничен. Кроме того, ранее нам не удалось гидролизовать короткие синтетические пептиды, содержащие остатки глицина, что может быть связано с наблюдаемой способностью B1p пропускать потенциальные сайты гидролиза.

Всего было идентифицировано 2031 уникальных пептидов, полученных в результате гидролиза B1p, 1374 уникальных пептидов для L5 и 5660 уникальных пептидов для трипсина.

Анализ пропущенных сайтов гидролиза (missed cleavages) при действии B1p показал, что среди 629 идентифицированных уникальных пептидов 44% (если учитывать только сайты глицина) или 33,5% (если учитывать сайты глицина и лизина) содержали один или более пропущенных сайтов. В совокупности 45,8% всех пептидов имеют один или более пропущенных сайтов гидролиза. Это свидетельствует о том, что даже при наличии узнаваемых остатков глицина, фермент способен не расщеплять соответствующие пептидные связи. Примеры таких пептидов приведены в таблице 9.

Таблица 9. Пептиды, идентифицированные с использованием B1p, содержащие один или более пропущенных сайтов гидролиза.

Белок	Пептиды с пропущенными сайтами гидролиза	Количество пропущенных сайтов гидролиза	Ожидаемые фрагменты, которые не удалось идентифицировать
CPSM_RAT	HPSSVAG <u>G</u> EVV	1	HPSSVAG, EVV
CPSM_RAT	HPSSVAG <u>G</u> EVVFNT	1	HPSSVAG, EVVFNT

CPSM_RAT	EVIKAERPDGLILG	2	EVIK, AERPDG, LILG, EVIKAERPDG, AERPDGLILG
SVEP1_MOUSE	SDPKIQLIFNIT	1	SDPK, IQLIFNIT
CPSM_RAT	ADTIGYPVMIRS	1	ADTIG, YPVMIRS
CPSM_RAT	NMENVDAMGVHTG	1	NMENVDAMG, VHTG
CPSM_RAT	QSVDFVDPNKNQLIAEVSTK	1	QSVDFVDPNK, QNLI AEVSTK
ALDH2_MESAU	STEVGHLIQVA	1	STEVG, HLIQVA
ALDH2_MESAU	SHEDVDKVAFTG	1	SHEDVDK, VAFTG
ALDH2_MESAU	NPFDSRTEQGPQVDETQFK	1	NPFDSRTEQG, PQVDETQFK
BHMT1_RAT	RQVADEGDALVAG	1	RQVADEG, DALVAG
JHD2C_MOUSE	NSQAVGERRL	1	NSQAVG, ERRL
MVP_RAT	AARIEGEGSVLQA	2	AARIEG, EG, SVLQA

Для того, чтобы понять предпочтения фермента в отношении размера продукта, был проведен сравнительный анализ длин всех идентифицированных уникальных пептидов B1p и L5. Средняя длина пептидов, образованных в результате гидролиза B1p, составила $14,85 \pm 7,26$ аминокислот (медиана – 13, мода – 10), тогда как пептиды, гидролизованные ферментом L5, были значительно короче – в среднем $8,5 \pm 2,4$ аминокислоты (медиана и мода – 8). Для сравнения также была проанализирована длина пептидов, полученных в результате гидролиза той же белковой смеси трипсином. В этом случае средняя длина пептидов составила $13,8 \pm 6,7$ аминокислоты (медиана – 12, мода – 8). На рисунке 21 представлена сравнительная оценка специфичности расщепления ферментами B1p и L5 на основе длины идентифицированных пептидов.

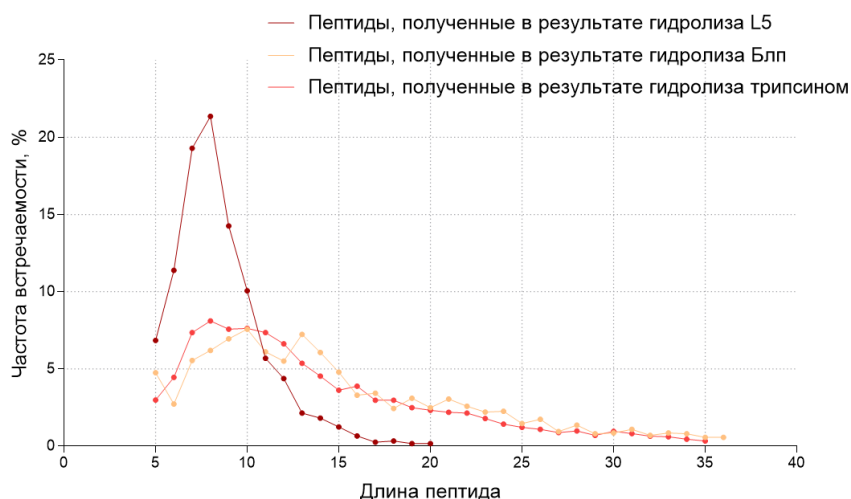


Рисунок 21. График частоты встречаемости уникальных пептидов различной длины, образующихся в результате гидролиза белков протеазами Blp, L5 и трипсином. Частота встречаемости выражена в процентах от общего числа уникальных пептидов, идентифицированных для каждой протеазы. По оси X – длины пептидов, по оси Y – частота встречаемости уникальных пептидов в процентах от общего числа для каждого фермента.

Для облегчения интерпретации данные длины пептидов, образованных каждой протеазой, представлены в виде частотного распределения, а также сгруппированы по диапазонам длины: 5–8, 9–12, 13–16, 17–20 и >21 аминокислота (рисунок 22). Такое представление позволило выявить выраженные различия в характере гидролиза в более явном виде.

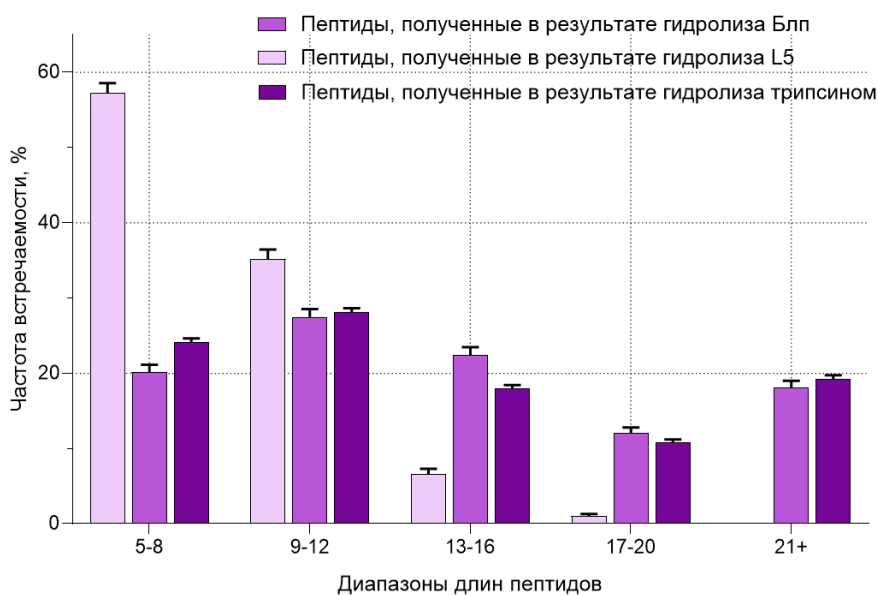


Рисунок 22. Распределение уникальных пептидов по длине для трёх ферментов (Blp, L5, трипсин). Пептиды сгруппированы по диапазонам длин: 5–8, 9–12, 13–16, 17–20 и ≥ 21 аминокислот. По оси X – диапазоны длин пептидов, по оси Y – частота встречаемости уникальных пептидов в процентах от общего числа для каждого фермента.

Протеаза L5 образовывала преимущественно короткие фрагменты: более 57% всех пептидов имеют длину 5–8 остатков, и ещё 35% – длину 9–12. Таким образом, около 92% всех пептидов, образуемых L5, укладывались в диапазон 5–12 аминокислот, при этом количество более длинных фрагментов (13–16 и 17–20 остатков) было минимальным – 6,6% и 1,4% соответственно.

Для сравнения, Vpr имеет более равномерное распределение длины пептидов. Наиболее часто встречающимися были пептиды длиной 9–12 остатков (27%), за которыми следовали диапазоны 13–16 (22%) и 5–8 (20%). Примерно 12% всех фрагментов имеют длину 17–20 аминокислот. Аналогичное распределение было получено и для трипсина, классической протеазы, часто используемой в протеомике: диапазоны 5–8 и 9–12 были доминирующими (24% и 31% соответственно), при этом пептиды длиной 13–16 и 17–20 также составляли существенную долю (18% и 11%).

Различия в длине продуктов расщепления между Vpr и L5 могут отражать особенности механизма действия этих ферментов. Более короткие пептиды, образуемые L5, вероятно, связаны со множественной специфичности, а также, по-видимому, отсутствием стерических ограничений активного центра по длине субстрата, в отличие от Vpr.

Хотя L5 проявляет широкую аминокислотную специфичность и имеет склонность к образованию коротких фрагментов, предыдущие исследования показали, что его бактериолитическая активность существенно ниже в растворимой форме по сравнению с тем, когда он секретируется через везикулы внешней мембраны [47]. Это говорит о том, что L5 может функционировать не как первичный литический фермент, нацеленный на разрушение клеточной стенки, а скорее, как вспомогательная протеаза в бактериальном секрете. Vpr, напротив, обладает более выраженной лизирующей активностью и узкой специфичностью – преимущественно расщепляет пептидные связи после остатков глицина и лизина –

что указывает на его участие в целенаправленном разрушении компонентов клеточной стенки. Таким образом, ферменты L5 и Vlp, вероятно, выполняют взаимодополняющие роли: Vlp выступает в роли основного лизирующего агента, а L5 – как адаптивная протеаза с более широкой, но менее агрессивной специфичностью.

Склонность Vlp к образованию более длинных пептидов может отражать более селективный механизм расщепления. Аналогичный профиль гидролиза, характеризующийся образованием схожих по длине пептидов, был ранее описан для фермента Alp [116]. Интересно, что как WaLP (дикий тип Alp), так и MaLP (мутантный вариант Alp) преимущественно расщепляют пептидные связи после алифатических аминокислот (WaLP – A, V, T и S; MaLP – L, F и V), профиля специфичности, схожего с L5. Однако средняя длина пептидов, образуемых WaLP ($11,7 \pm 4,8$ а.о.) и MaLP ($12,0 \pm 4,5$ а.о.), ближе к Vlp, чем к L5. Более короткие фрагменты, образуемые L5, возможно, являются результатом последовательных множественных расщеплений – свойства, характерного для ферментов, участвующих в бактериолизе.

4.7. Определение кинетических параметров с использованием количественной масс-спектрометрии с метками ^{18}O

В данной работе использовался метод количественной масс-спектрометрии с использованием стабильных изотопных меток (абсолютной количественной оценки концентраций, AQUA). Принцип метода основан на сравнении интенсивностей сигналов двух пептидов в масс-спектре, имеющих идентичную аминокислотную последовательность, но различающихся массой за счёт включения одного или нескольких стабильных изотопов.

В качестве изотопной метки использовался стабильный нерадиоактивный изотоп кислорода ^{18}O , замещающий атомы ^{16}O в пептидах. Обмен происходит

преимущественно в С-концевых карбоксильных группах пептидов, а также в боковых цепях остатков глутаминовой и аспарагиновой кислот. Реакция обмена является самопроизвольной,

Метод определения кинетических параметров заключается в том, что к пробе смеси фермента и с субстрата добавляется известное количество изотопно-меченого стандарта, а далее по соотношению их пиков в спектре рассчитывается концентрация изучаемого продукта. В дальнейшем значения V_0 использовались для построения кинетических кривых и расчёта K_M , K_{cat} и K_{cat}/K_M .

Метод был опробован на модельной системе с использованием трипсина и далее применён к двум исследуемым ферментам – B1p и L5, выделенным из штаммов рода *Lysobacter*. Поскольку эти ферменты обладают различными уровнями активности и разной специфичностью, использование единой методологии позволило объективно сравнить их каталитические характеристики.

4.7.1. Получение стандартов пептидов, определение концентрации и подтверждение изотопного обмена

В настоящей работе пептиды, используемые для обмена, были получены путем гидролиза белков, по скорости образования продуктов которых определялись кинетические параметры, с последующим разделением и фракционированием на ВЭЖХ. Пептиды, использованные для определения кинетических параметров L5, были синтезированы.

Все выбранные пептиды соответствовали ряду критериев, обеспечивающих корректность количественного анализа:

1. Использовались короткие пептиды, так как увеличение числа потенциальных сайтов изотопного обмена повышает разброс масс-спектрометрического сигнала и затрудняет интерпретацию результатов;

2. В диапазоне m/z , соответствующем основному пику выбранного пептида и его изотопным формам с включением ^{18}O (+2 и +4, +6 а.е.м. и т.д.), не должно быть других сигналов, чтобы избежать наложения и интерференции сигналов от других пиков;
3. Выбранный пептид должен быть терминальным фрагментом (N- или C-концевым), если целью является определение скорости гидролиза по одной пептидной связи;
4. Выбранный пептид должен представлять собой конечный продукт гидролиза исследуемой протеазой, не содержащий потенциальных сайтов дальнейшего расщепления, чтобы избежать последующего гидролиза этого продукта.

Важно учитывать, если выбранный пептид (P_1) не является конечным продуктом гидролиза и может подвергаться дальнейшему гидролизу, его концентрация во времени определяется не только скоростью образования (V_1), но и скоростью образования следующего продукта (V_2). Таким образом, наблюдаемая концентрация пептида описывается разностью скоростей двух последовательных стадий (8):

$$\frac{d[P_1]}{dt} = V_1 - V_2 \quad (8)$$

Если V_1 значительно превышает V_2 ($V_1 > V_2$), то на начальном этапе реакции можно считать, что влияние второго этапа (гидролиза продукта) минимально. В этом случае возможно определение начальной скорости накопления продукта (V_0), отражающей кинетику образования промежуточного пептида.

Однако такой подход требует осторожности в интерпретации: полученные параметры (K_M и V_{\max}) будут характеризовать первую стадию реакции ($S \rightarrow P_1$), но не всю последовательность превращений вплоть до конечного продукта. Поэтому в идеале пептид, используемый в качестве количественного маркера, должен

соответствовать критерию 4, описанному выше, – быть конечным стабильным продуктом реакции, не содержащим потенциальных сайтов для дальнейшего гидролиза. Такой пример будет рассмотрен далее, в разделе 4.7.5.

Для оценки степени изотопного обмена атомов ^{16}O на изотопы ^{18}O в пептидах регистрировали масс-спектры в течение трёх часов инкубации. Для всех пептидов через один час наблюдалась частичная замена, при этом основной пик немеченой формы пептида сохранялся. Однако спустя два часа интенсивность первого пика (соответствующего пептиду без включённого ^{18}O) составила менее 1% от интенсивности пятого пика (пептида с двумя атомами ^{18}O) для всех пептидов, использованных в данной работе. Сравнение площадей изотопных сигналов показало, что в 81% молекул произошло включение двух атомов ^{18}O , в 18% – одного, и лишь около 1% молекул сохранили первоначальный изотопный состав. На рисунке 23 один представлены масс-спектры пептида TTMLPW после его инкубации в H_2^{18}O .

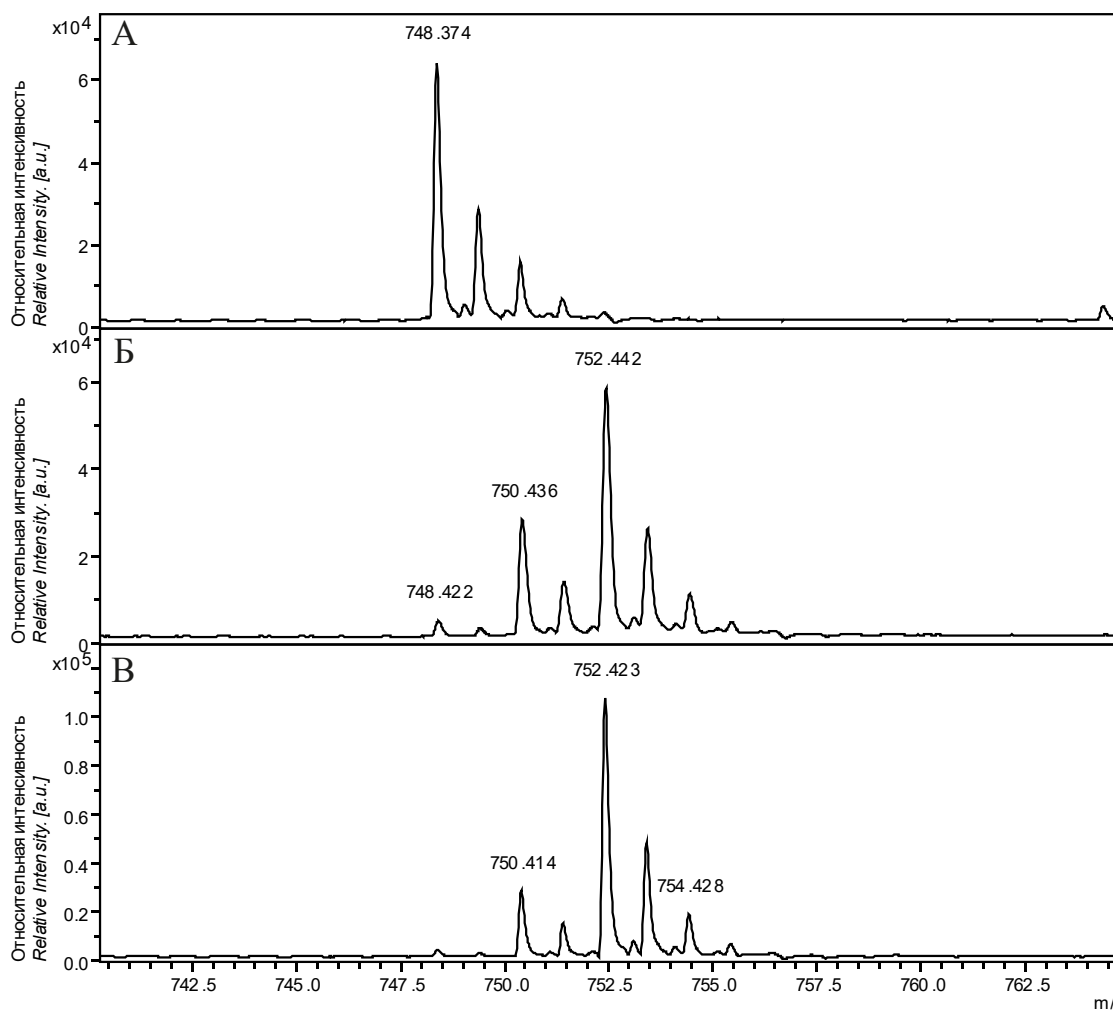


Рисунок 23. Масс-спектры пептида TTMLPW после его перерастворения в 10% растворе трифторуксусной кислоты в H_2^{18}O через 5 мин (А), 60 мин (Б) и 120 мин (В) инкубации.

Методом тандемной масс-спектрометрии подтверждено, что в пептиде TTMLPW два атома ^{16}O в карбоксильной группе С-концевого триптофана были полностью замещены на атомы ^{18}O (рисунок 24). Это подтверждается тем, что N-концевые фрагменты, как у исходного пептида, так и у меченого варианта, совпадают по массам, тогда как массы С-концевых ионов отличаются на 4 Да, что соответствует инкорпорации двух атомов ^{18}O в карбоксильную группу. Несмотря на присутствие в последовательности остатка треонина, содержащего гидроксильную группу, обмен по этому атому кислорода не происходил. Это объясняется тем, что скорости обмена атомов кислорода в различных функциональных группах значительно различаются и для гидроксильных радикалов обмен крайне маловероятен.

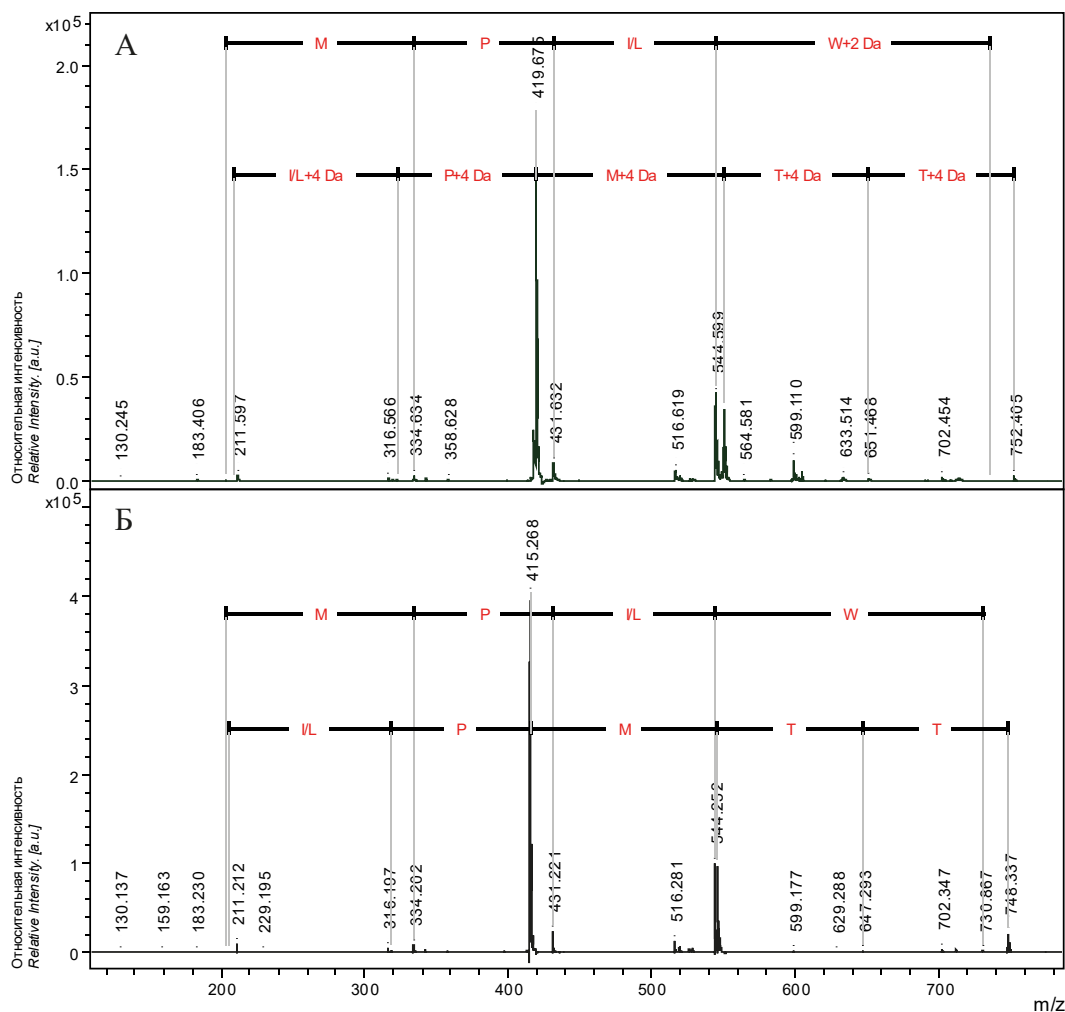


Рисунок 24. Масс-спектры фрагментов пептида TTMLPW после инкубации в воде $H_2^{18}O$ (А) и исходного пептида (Б).

Для других коротких пептидов (AWYYVPLG, RFFVAPF и RGPFPPIV), в изотопном кластере наблюдалось такая же картинка, как и при обмене атомов ^{16}O на ^{18}O в пептиде TTMLPW: основную интенсивность имели пики с разностью 2 Да (Рисунок 25 А-В), что соответствовало полной замене двух атомов ^{16}O на ^{18}O . Между этими пиками присутствовали менее интенсивные пики с шагом в 1 Да, обусловленные включением одного атома ^{18}O или естественным изотопным распределением ^{13}C .

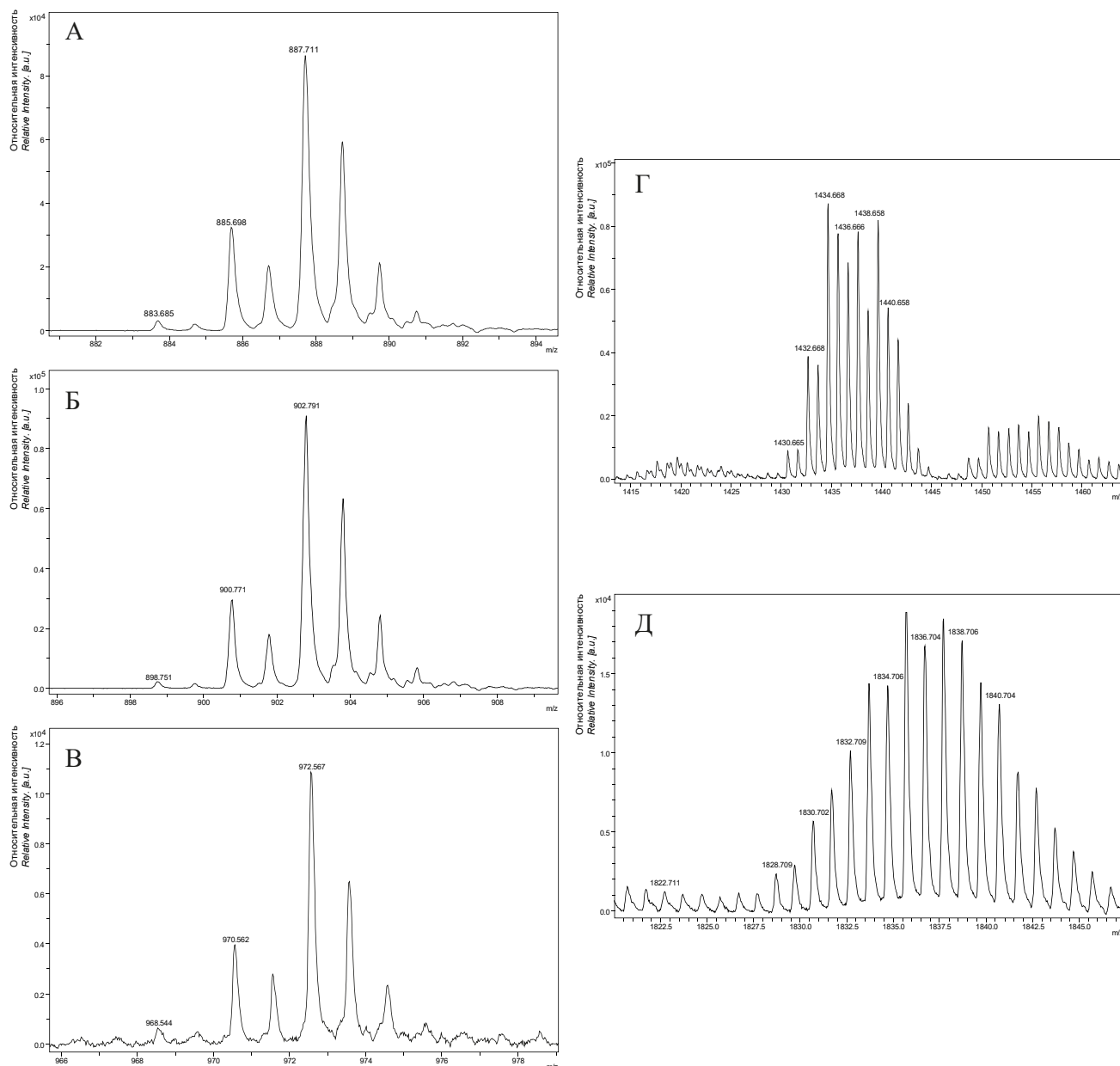


Рисунок 25. Масс-спектры пептидов RFFVAPF (А), RGPFPPIV (Б), AWYYVPLG (В), SENSEKTTMPLW (Г), TQYTDAPSFSDIPNPIG (Д) после полного обмена атомов ^{16}O на ^{18}O .

Однако для пептидов с большей массой (SENSEKTTMPLW и TQYTDAPSFSDIPNPIG) интерпретация изотопного распределения существенно усложняется. Во-первых, по мере увеличения числа атомов углерода в молекуле изменяется изотопное распределение. Вероятность наличия хотя бы одного тяжёлого изотопа (^{13}C , ^{15}N и др.) возрастает, и максимум интенсивности изотопного кластера постепенно смещается на +1 Да вправо относительно моноизотопного пика (то есть пика пептида, у которого все атомы – самые распространённые изотопы) (Рисунок 25 Г, Д). В результате наиболее интенсивным становится не

первый, а второй или третий пик, что приводит к перекрытию сигналов от изотопных пиков немеченой формы и сигналов продуктов обмена.

Во-вторых, большой вклад вносит дезамидирование боковых цепей аспарагина и глутамина, которое является спонтанной природной модификацией. Оно сопровождается увеличением массы пептида приблизительно на +1 Да. Этот процесс протекает быстрее в кислой среде, то есть в условиях, сходных с теми, которые используются для проведения обмена ^{18}O . При этом дезамидированию подвергаются не все молекулы пептида, а лишь определённая доля, зависящая от времени инкубации, pH и температуры. В результате у пептидов наблюдается уменьшение интенсивности первого (моноизотопного) пика относительно второго за счёт появления дезамидированных форм, соответствующих +1 Да. При этом трудно точно оценить количество пептидов, подвергшихся дезамидированию, а значит, корректно рассчитать концентрацию стандарта.

Таким образом, при анализе обмена в длинных пептидах накладываются два эффекта: смещение изотопного распределения и образование дезамидированных форм. В совокупности они приводят к резкому снижению интенсивности моноизотопного пика и увеличению перекрытий между немечеными, модифицированными и мечеными формами. В результате спектр пептида с ^{18}O уже не выглядит как набор нескольких высоких пиков, между которыми располагаются менее интенсивные с шагом 1 Да, а скорее представляет собой смесь перекрывающихся сигналов. Такая картина значительно усложняет интерпретацию и делает количественную оценку степени обмена затруднительной.

Несмотря на описанные ограничения, приближённая оценка степени обмена остаётся возможной. Для этого интенсивность наиболее выраженного пика мечёного пептида принималась за 100% концентрации пептида-стандарта, относительно которого рассчитывалась концентрация пептида-продукта (рисунок 26). Такой подход не обеспечивает высокой точности и не позволяет количественно рассчитать долю обмена, однако он может быть полезен для оценки динамики

процесса. На соответствующих графиках изменение интенсивности меченого пика при увеличении времени инкубации позволяет наблюдать накопление продукта относительно стандарта.

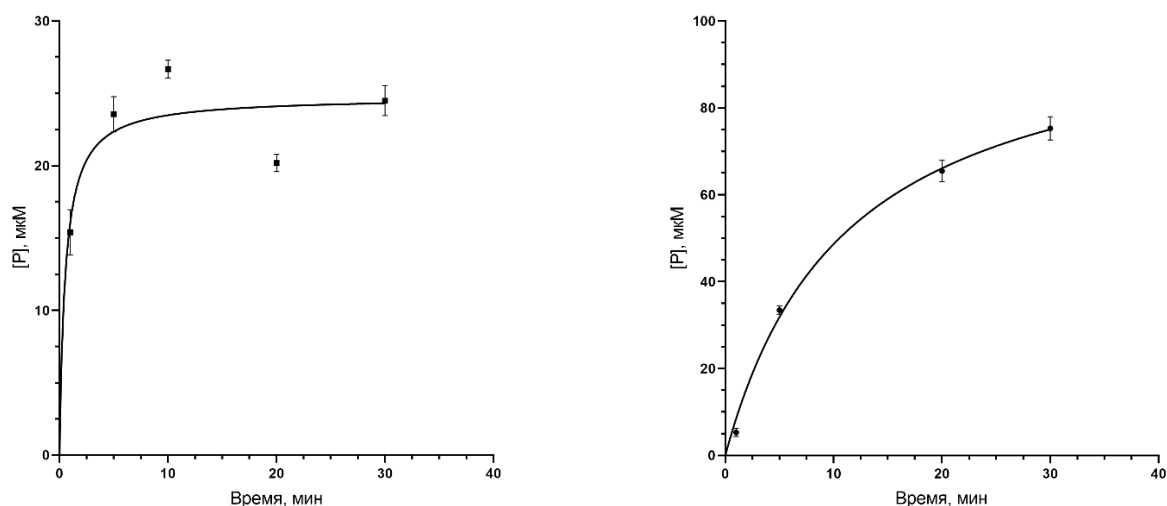


Рисунок 26. Зависимость накопления пептидов-продуктов гидролиза B1p SENSEKTTMPLW (слева) и TQYTDAPSFSDIPNPIG (справа) от времени при различных начальных концентрациях субстрата (казеина). По оси ординат – средняя концентрация продукта (мкМ), по оси абсцисс – время инкубации (мин). Показаны средние значения \pm стандартное отклонение.

Таким образом, как в случае коротких, так и для высокомолекулярных пептидов, интерпретация изотопного кластера требует особой осторожности и, как правило, невозможна без использования дополнительных методов расчёта, учитывающих естественное изотопное распределение, возможные модификации пептидов и наложение изотопных форм.

4.7.2. Расчёт концентрации продуктов гидролиза по концентрации изотопно-меченого стандарта

Для определения концентрации меченого стандарта его смешивали с немеченым пептидом известной концентрации – эталоном, рассчитанной по площади хроматографического пика. Соотношение компонентов в смеси подбиралось таким образом, чтобы интенсивность моноизотопного пика продукта

гидролиза и самого интенсивного пика стандарта (с включёнными атомами ^{18}O) была сопоставимой.

Следует отметить, что при проведении изотопного обмена невозможно добиться строго однородного мечения. В результате в составе стандарта присутствует смесь трёх форм пептида: без включения ^{18}O (рисунок 27 А), с одним атомом ^{18}O (рисунок 27 Б) и с двумя атомами ^{18}O (рисунок 27 В). Эти формы регистрируются в масс-спектре как группы сигналов, смещённых на +0, +2 и +4 m/z относительно немеченого пептида. В частности, частичное включение одного атома ^{18}O приводит к появлению изотопных пиков, смещённых на +2 а.е.м. (обозначено жёлтым цветом на рисунке 27 Б). В молекулах, где произошло включение обоих атомов кислорода, наблюдаются пики сдвига на +4 m/z (обозначено зеленым цветом на рисунке 27 Б). При этом соотношение интенсивностей изотопных пиков этого пептида сохраняется для пептидов без включённых меток, с одним и двумя ^{18}O , однако при их смешении – интенсивности изотопных пиков, имеющих одну массу, суммируются. Такое наложение сигналов требует соответствующего учёта при количественном анализе.

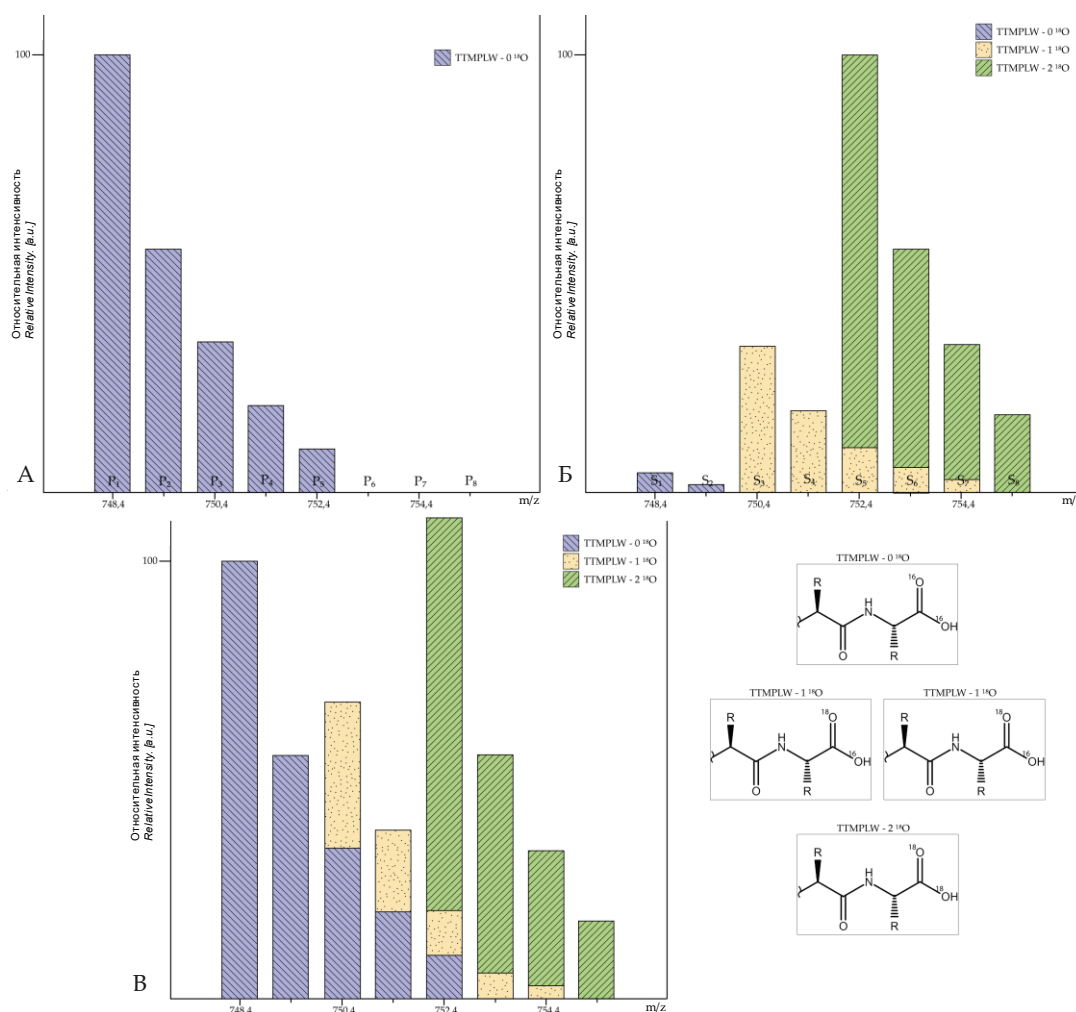


Рисунок 27. Схематическое изображение масс-спектров. А – пептид TTMLPW (продукт); Б – изотопно-меченый стандарт с заменой одного или двух атомов ^{18}O ; В – смеси стандарта и продукта, где наглядно показано наложение относительных интенсивностей пиков стандарта и продукта. Химическими формулами изображены атомы кислорода, способные обмениваться в данном пептиде.

Как было сказано ранее, для коротких пептидов удастся получить стандарт, в котором практически отсутствуют пики пептида без включения ^{18}O . В этом случае отношение концентраций пептида-продукта и стандарта будет равно отношению интенсивности первого пика к сумме третьего и пятого пиков (рисунок 27 В). Это пики стандарта, но на них накладывается третий и пятый пик продукта, а на пятый пик (пептиды с двумя включенными изотопами ^{18}O) – еще и третий пик пептидов (пептиды с одним включенным изотопом ^{18}O). Вклады этих пептидов нужно вычесть [155] согласно формуле (9):

$$\frac{C_{\text{имс}}}{C_{\text{продукт}}} = \frac{q_1}{q_5 - \frac{p_5}{p_1} q_1 - \frac{p_3}{p_1} (q_3 - \frac{p_3}{p_1} q_1) + (q_3 - \frac{p_3}{p_1} q_1)} \quad (9)$$

где q_1 , q_3 и q_5 – интенсивности первого, третьего и пятого сигналов в спектре смеси продукта реакции и стандарта, p_1 , p_3 и p_5 – соответствующие интенсивности сигналов в спектре исходного пептида с неизмененным изотопным составом. Для проведения расчета можно не регистрировать спектр природного пептида, а взять его из другого эксперимента или рассчитать p_1 , p_3 и p_5 теоретически.

Следует отметить, что нет необходимости знать соотношение пиков в спектре стандарта, достаточно только знать, что замена атомов ^{16}O на ^{18}O произошла в большей части пептидов, а число атомов в молекуле пептида, которые могут подвергаться замене, невелико. Однако распределение в экспериментально полученных пептидах часто оказывается отличным от расчетного, что приводит к неправильному расчету p_1 , p_3 , p_5 и увеличению ошибок в расчетах концентрации.

Последние ограничения не исключают возможности точного расчета концентраций: даже если стандарт будет не идеальным, в положении первого пика окажется достаточно сильный сигнал, обмен пройдет не полностью, а подвергаться замене будет большое число атомов, точный расчет концентраций все равно остается возможным – в этом случае нужно будет использовать три спектра:

- спектр пептида с замещенными на ^{18}O атомами кислорода карбоксильной группы (стандарта);
- спектр смеси стандарта и эталонного природного пептида;
- спектр смеси стандарта и пептида, полученного в результате гидролиза (продукта).

Интенсивность каждого сигнала в спектре смеси окажется суммой соответствующих сигналов эталонного пептида и стандарта с коэффициентами t и k (10):

$$\begin{cases} q_1 = kp_1 + ts_1 \\ q_2 = kp_2 + ts_2 \\ \dots \\ q_i = kp_i + ts_i \end{cases} \quad (10)$$

где q_i – интенсивность сигнала пика i в спектре смеси продукта реакции и стандарта; p_i – интенсивность сигнала пика i в спектре эталонного пептида (с неизменным изотопным составом); s_i – интенсивность сигнала пика i в спектре стандарта.

В силу того, что в различных опытах (даже повторах) величины q_i , p_i и s_i могут сильно меняться, рассчитанные значения коэффициентов t и k также будут изменяться. Однако как было сказано ранее, если компоненты смеси химически идентичны, то соотношение пиков таких компонентов не меняется. Для определения концентрации пептида продукта необходимо найти отношение коэффициентов t к k , которое сохраняется для разных спектров. Чтобы найти t и k достаточно двух уравнений; желательно выбрать два наиболее интенсивных сигнала. В нашем случае ими оказались q_1 и q_5 (11):

$$\begin{cases} q_1 = kp_1 + ts_1 \\ q_5 = kp_5 + ts_5 \end{cases} \quad (11)$$

Отношение k к t равняется отношению концентрации продукта к концентрации стандарта и находится путем решения системы линейных алгебраических уравнений согласно (12):

$$\frac{C_{\text{продукт}}}{C_{\text{ИМС}}} = \frac{k}{t} = \frac{(q_1s_5 - q_5s_1)(p_5s_1 - p_1s_5)}{(p_1s_5 - p_5s_1)(p_5q_1 - p_1q_5)} = \frac{q_1s_5 - q_5s_1}{p_5q_1 - p_1q_5} \quad (12)$$

Поскольку пятого пика в спектре эталона мы не наблюдаем ($p_5=0$), а первый пик эталона и пятый пик ИМС нормируются на 1 ($p_1=1$, $s_5=1$), то уравнение принимает вид (13):

$$\frac{C_{\text{продукт}}}{C_{\text{ИМС}}} = \frac{q_1 - q_5s_1}{q_5} \quad (13)$$

Определение концентрации продукта проводили при использовании этого подхода дважды – устанавливая концентрацию стандарта по спектру его смеси с эталоном и затем определяя концентрацию продукта реакции по спектру его смеси с стандарта.

4.7.3. Валидация метода на модельной системе с использованием трипсина

Для валидации разработанного подхода по определению концентрации пептидов был проведён эксперимент с использованием модельной системы – гидролиза казеина трипсином. Определение количества образующегося пептида-продукта TTMPLW проводили после 15-минутной инкубации при трёх различных начальных концентрациях субстрата. Полученные значения затем использовались для расчёта активности фермента.

Для повышения точности масс-спектрометрического измерения концентрации продукта важно, чтобы интенсивности сигналов пептида-продукта и изотопно-меченого стандарта в спектре были сопоставимы. Поскольку при снижении концентрации субстрата уменьшается и количество образующегося продукта, соотношение стандарта к продукту подбиралось индивидуально. Так, для концентрации казеина 7,75 мкМ соотношение субстрата к стандарту составляло 1:1, а для 3,87 и 3,1 мкМ – 1:2,5, соотношение продукт:стандарт составило 2:1.

Хотя концентрацию стандарта можно определить из одного спектра смеси с эталонным пептидом, для повышения точности она определялась отдельно для каждого варианта разведения. Расчёт по спектру смеси ИМС и эталона при соотношении 1:1 дал значение $11 \pm 0,2$ мкМ, а при соотношении 1:2,5 – 3,8 мкМ.

На основании известной концентрации стандарта была определена концентрация пептида-продукта. При начальной концентрации казеина 7,75 мкМ она составила $5 \pm 0,04$ мкМ (0,3 мкМ/мин), при 3,87 мкМ – $3 \pm 0,29$ мкМ (0,2

мкМ/мин), и при 3,1 мкМ – $2,55 \pm 0,25$ мкМ (0,17 мкМ/мин). Минимальная надёжно определяемая концентрация составила $0,5 \pm 0,08$ мкМ.

Предложенный метод обеспечивает воспроизводимое определение низких концентраций пептидов и может быть использован для расчёта кинетических параметров гидролиза. При этом чувствительность MALDI-TOF масс-спектрометрии значительно превышает достигнутый предел обнаружения, что открывает перспективу дальнейшего увеличения чувствительности метода. Стоит отметить, что чувствительность MALDI масс-спектрометрии существенно выше этих значений, что указывает на возможное достижение еще более высокой чувствительности.

На основе полученных данных был построен график в координатах Лайнвивера–Берка ($1/V - 1/[S]$), позволяющий оценить кинетические параметры реакции (Рисунок 28). На графике наблюдается линейная зависимость, что подтверждает соответствие экспериментальных данных модели Михаэлиса–Ментен и позволяет определить значение константы Михаэлиса по оси X и максимальной скорости – по оси Y.

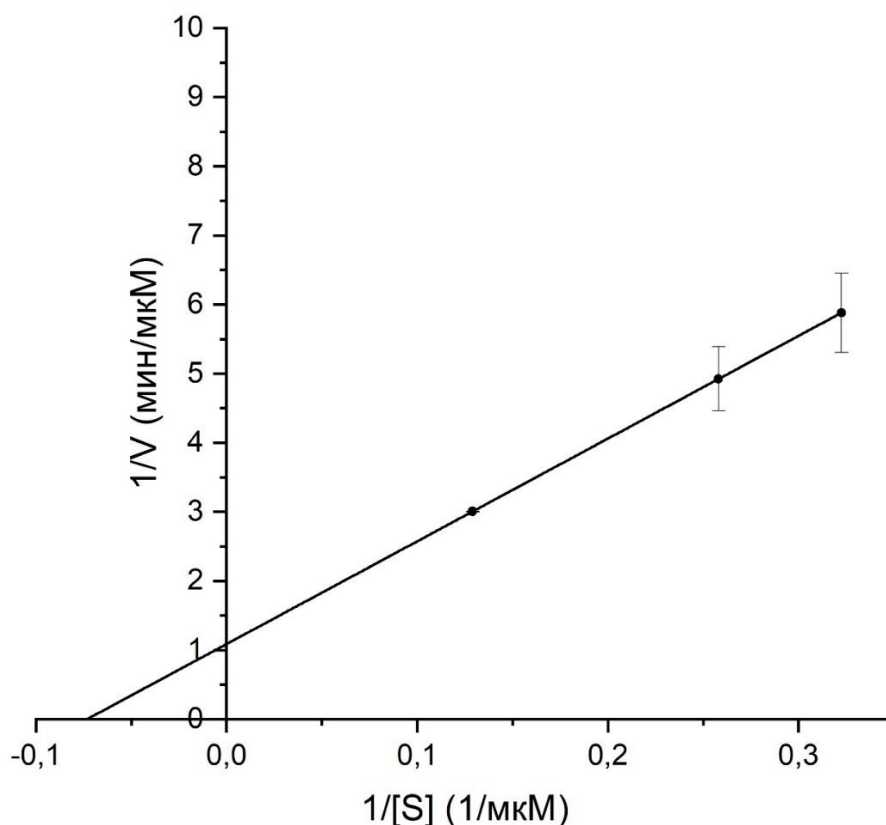


Рисунок 28. Зависимость скорости образования пептида-продукта от концентрации субстрата (казеин) в двойных обратных координатах. По оси Y – $1/\text{скорость образования продукта (мин/мкМ)}$, по оси X – $1/\text{концентрация субстрата (1/мкМ)}$.

Расчёт значения константы Михаэлиса по описанной методике показал, что для трипсина K_M составила $13,65 \pm 0,6$ мкМ. Сравнение с доступными в базе данных BRENDA значениями K_M для казеина при аналогичных температурных и pH-условиях ($K_M = 3,33$ мкМ [156]) показывает разницу примерно в 4 раза. Вероятной причиной этого является различие в используемом субстрате: в приведённой работе применялся казеин, модифицированный меткой FITC, которая ковалентно связывается с первичными аминогруппами белка, в первую очередь – по остаткам лизина. Такая модификация может существенно изменять доступность участков расщепления для трипсина, что, в свою очередь, влияет на кинетические параметры фермента.

4.7.4. Определение кинетических параметров *B1p* и *L5*

Для определения кинетических параметров исследуемых ферментов были выбраны пептиды AWYYVPLG (для *B1p*), RFFVAPF и RGPFPPIV (для *L5*).

Концентрацию пептидов эталонов определяли по площади поглощения пиков на длине волны 280 нм, для пептида AWYYVPLG – 0,59 мкМ ($\pm 0,03$ мкМ).

Концентрацию пептидов RFFVAPF и RGPFPPIV определяли с использованием аминокислотного анализатора, она составила 29,31 мкМ ($\pm 0,4$ мкМ) и 13,04 мкМ ($\pm 0,08$ мкМ) соответственно.

Для определения кинетических параметров протеазы *L5* были выбраны два пептида. Первый пептид – RGPFPPIV – является С-концевым фрагментом β -казеина, образующийся в результате гидролиза по остатку аминокислоты валина, являющегося специфичным для данной протеазы. Второй пептид – RFFVAPF – образуется при гидролизе α S1-казеина по остаткам аминокислот лейцина и фенилаланина, характерных для субстратной специфичности протеазы *L5*.

Для обоих пептидов кинетические параметры определялись в рамках одного эксперимента: стандарты обоих пептидов добавлялись к общей смеси гидролиза, и кинетические параметры рассчитывались по скорости накопления каждого из продуктов одновременно. Такой подход позволил параллельно отслеживать две независимые реакции в идентичных условиях.

Важно отметить, что в первом случае (пептид RGPFPPIV) скорость накопления продукта отражает кинетические параметры протеазы при расщеплении одной пептидной связи. Во втором случае (пептид RFFVAPF) фрагмент образуется в результате последовательного расщепления по двум пептидным связям, поэтому измеряемая скорость отражает суммарный эффект двух стадий, а определяемые кинетические параметры будут характеризовать самую медленную (лимитирующую) стадию гидролиза.

Для оценки кинетических параметров протеазы L5 была проведена серия экспериментов, в которых измеряли концентрацию продукта гидролиза (С, мкМ) в зависимости от времени (t, мин) при различных начальных концентрациях субстрата (казеина). Анализ осуществлялся в условиях квазистационарного режима с использованием MALDI-TOF масс-спектрометрии. На рисунке 29 (А, Б) наиболее интенсивный пик соответствует пептиду, содержащему два атома ^{18}O (100%). Изменение относительной интенсивности первого изотопного пика, не содержащего тяжёлую метку ^{18}O , отслеживалось во времени и использовалось для расчёта скорости накопления продукта.

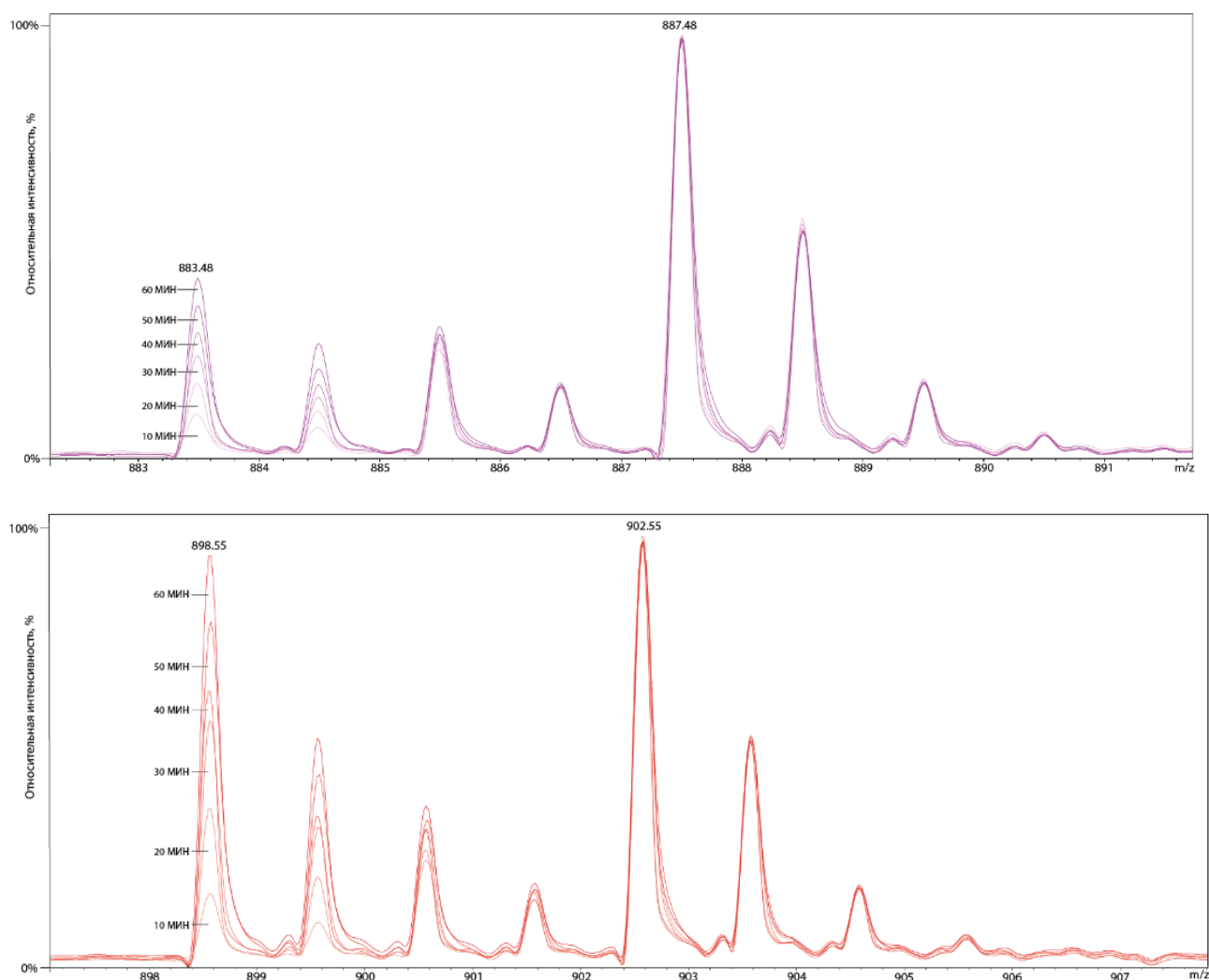


Рисунок 29. Масс-спектры, отражающие накопление пептидов RGPFPPIV (сверху) и RFFVAPF (снизу) под действием протеазы L5. Показаны спектры, полученные на разных этапах инкубации (градиент интенсивности от светлого к тёмному). С нарастанием времени наблюдается рост интенсивностей пиков, соответствующих продуктам гидролиза.

На первом этапе для пептида RGPFP_{IV} были получены кинетические кривые – графики зависимости концентрации продукта от времени для каждой концентрации субстрата (рисунок 30):

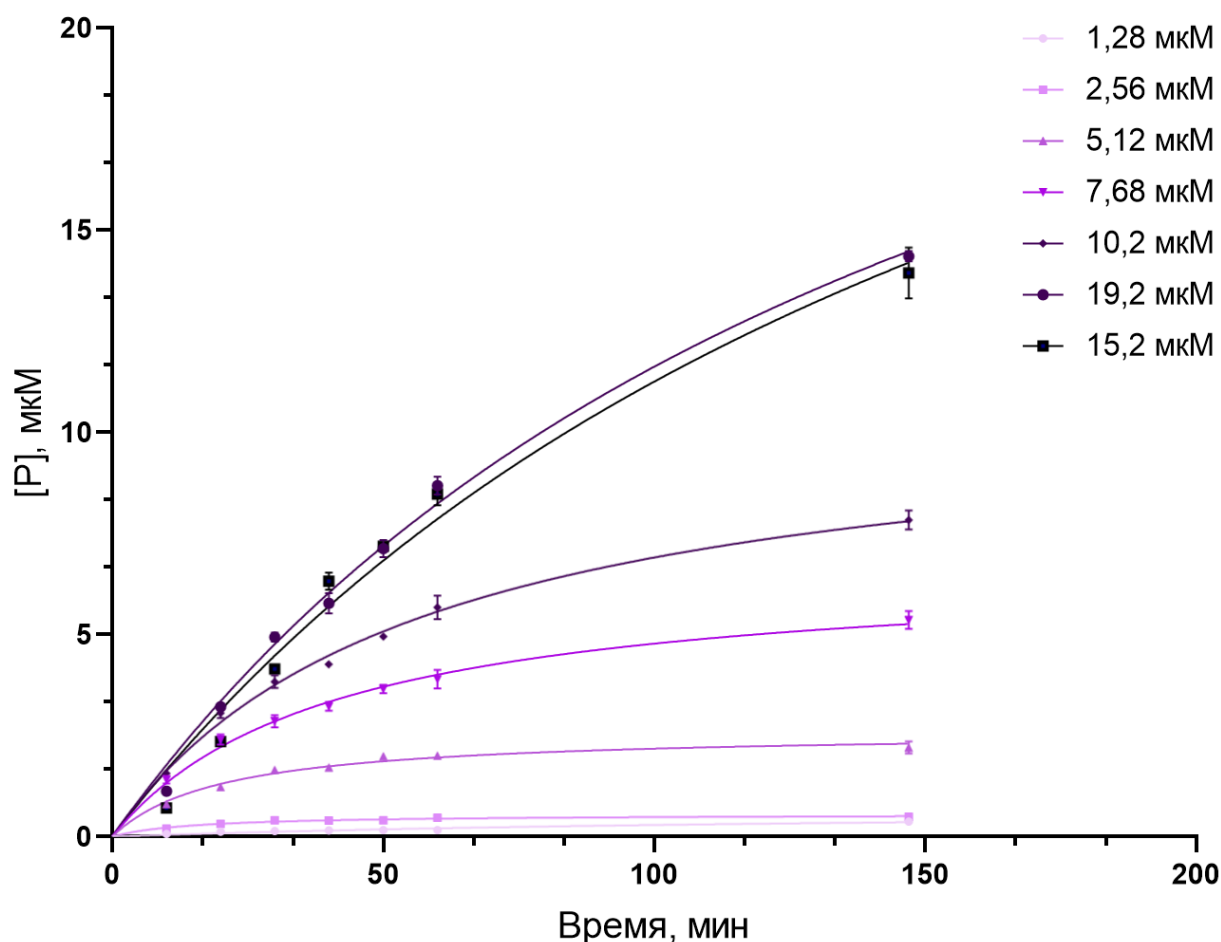


Рисунок 30. Зависимость накопления продукта (RGPFP_{IV}) гидролиза от времени при различных начальных концентрациях субстрата (казеина). Каждая кривая соответствует одной из концентраций субстрата (1,28–19,2 мкМ). По оси ординат – средняя концентрация продукта (мкМ), по оси абсцисс – время инкубации (мин). Показаны средние значения \pm стандартное отклонение. Кривые построены с использованием аппроксимации по методу наименьших квадратов.

Для количественной оценки кинетических параметров фермента L5 был построен график зависимости начальной скорости реакции (V_0 , нМ/мин) от начальной концентрации субстрата ($[S]$, мкМ). Начальная скорость реакции определялась как угловой коэффициент (наклон) касательной к начальной

(линейной) части кривой. Все измерения проводились в трёх и более повторах; на графике представлены средние значения и стандартные отклонения (рисунок 31).

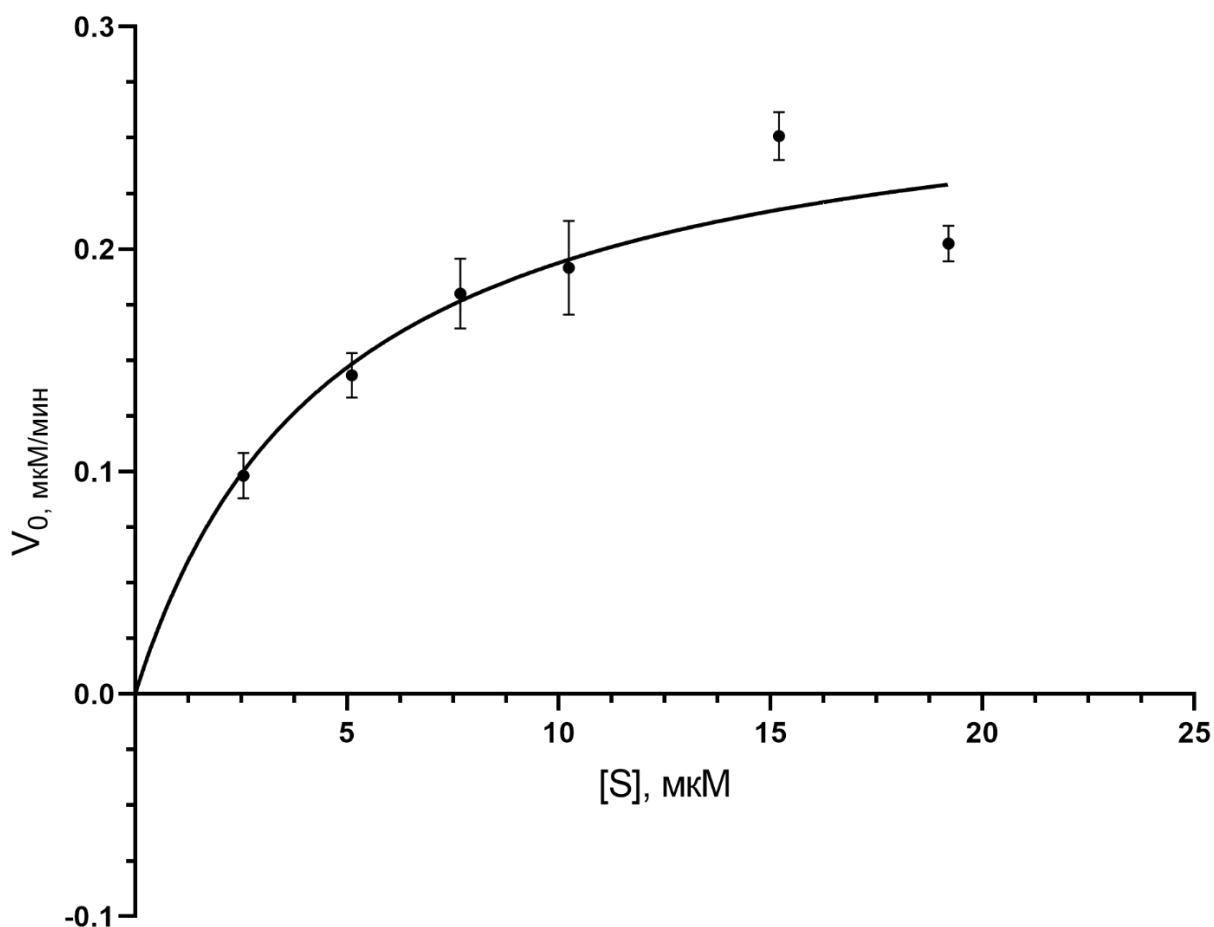


Рисунок 31. Зависимость начальной скорости гидролиза казеина ферментом L5 (V_0 , мМ/мин) от начальной концентрации субстрата ($[S]$, мМ). Скорость определялась по накоплению пептида RGPFPiIV. Кривая аппроксимации построена по уравнению Михаэлиса–Ментен. Представлены средние значения и стандартные отклонения ($n = 3-6$).

Экспериментальные данные были аппроксимированы уравнением Михаэлиса–Ментен (13):

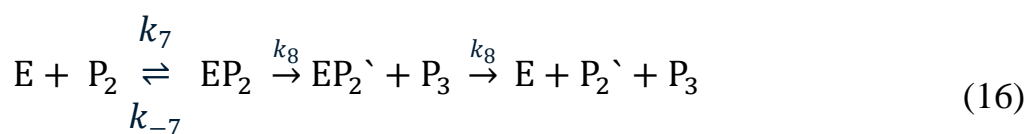
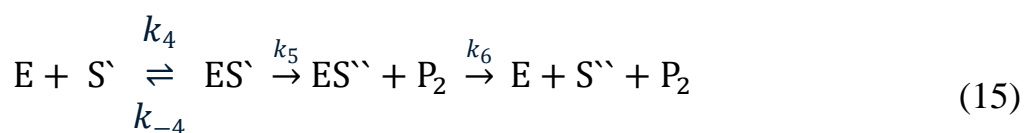
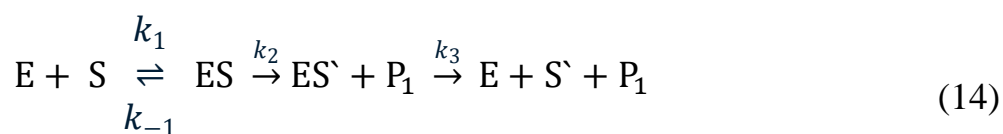
$$V_0 = \frac{V_{\max} \times [S]}{V_{\max} + [S]} \quad (13)$$

Полученная кривая зависимости начальной скорости ферментативной реакции (V_0) от концентрации субстрата была хорошо описана гиперболическим уравнением Михаэлиса–Ментен. Кинетические параметры для фермента L5 по продукту гидролиза RGPFPiIV были рассчитаны на основе аппроксимации

экспериментальных данных по уравнению Михаэлиса-Ментен: максимальная скорость реакции (V_{\max}) составила $0,315 \pm 0,02$ мкМ/мин, а $K_M - 6,55 \pm 0,58$ мкМ.

С учётом концентрации фермента в реакции ($0,147$ мкМ), каталитическая константа (k_{cat}) была рассчитана как $2,14$ мин⁻¹, что отражает число оборотов молекулы фермента в минуту. Каталитическая эффективность, выраженная как отношение k_{cat}/K_M , составила $0,327$ мкМ⁻¹мин⁻¹, что свидетельствует о достаточно эффективном катализе в условиях проведённого эксперимента.

В качестве второго пептида для определения кинетики, как уже упоминалось ранее, выбран пептид RFFVAPF, который является продуктом гидролиза по двум связям. Важно отметить, что данный пептид не является конечным продуктом реакции, а образуется как промежуточный продукт, подвергающийся дальнейшему ферментативному гидролизу. Об этом свидетельствует временная динамика: при низких концентрациях исходного субстрата ($1,28-7,68$ мкМ) концентрация пептида достигает максимума через 30-40 минут, после чего начинает снижаться (рисунок 32А). При высоких концентрациях субстрата ($10,24-25,6$ мкМ) продукт не убывает, что указывает на замедление второй стадии гидролиза. В целом реакция может быть описана тремя последовательными стадиями ферментативного гидролиза (14-16) (рисунок 32Б):



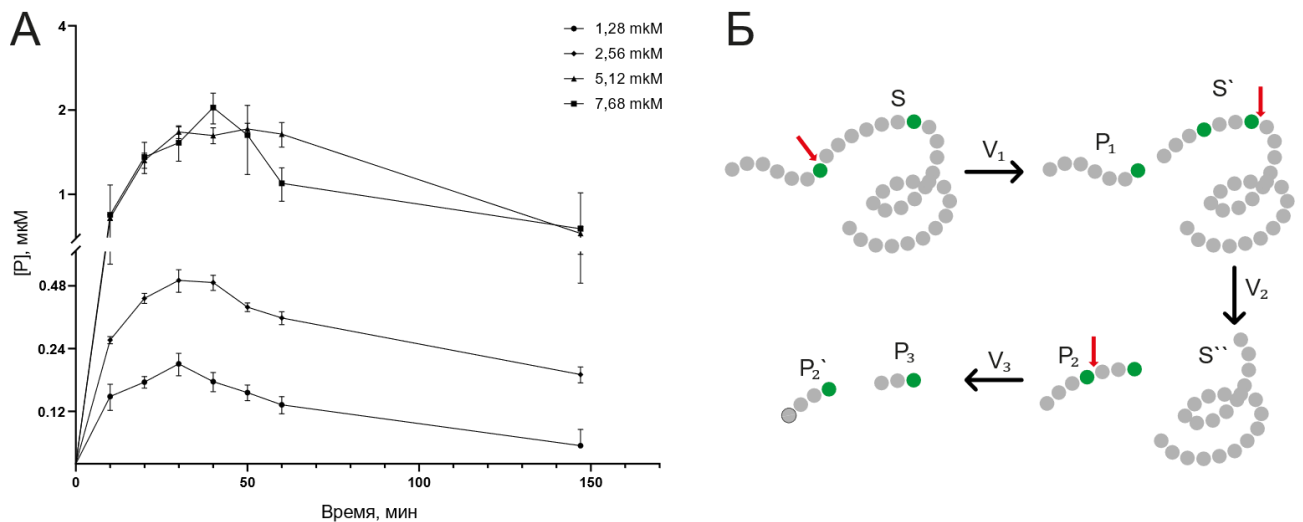


Рисунок 32. А – динамика накопления продукта RFFVAPF гидролиза α s1-казеина ферментом L5 при различных концентрациях субстрата (1,28–7,68 мкМ). По оси Y – концентрация продукта (мкМ) (логарифмическая шкала). Видно, что при низких концентрациях субстрата пептид достигает максимума через 30–40 минут, после чего его концентрация снижается, что указывает на дальнейший ферментативный гидролиз промежуточного продукта. Представлены средние значения \pm стандартное отклонение ($n = 6$). Б – схематическое представление образования пептида RFFVAPF (P_2), по накоплению которого измерялись кинетические параметры фермента. Красным цветом обозначено место разреза.

Применение модели Михаэлиса–Ментен в данной системе обосновано условиями, соблюденными в ходе эксперимента. Во-первых, измерения скорости накопления пептида RFFVAPF проводились на начальной фазе реакции – до начала его убывания. Это позволяет рассматривать образование данного пептида как изолированную первую стадию реакции, минимизируя вклад последующего гидролиза.

Во-вторых, концентрация субстрата на момент начала реакции существенно превышала концентрацию фермента ($[S] \gg [E]$), что удовлетворяет критерию квазистационарного приближения. Это условие необходимо для соблюдения классической схемы образования фермент-субстратного комплекса и расчёта начальной скорости V_0 .

В-третьих, измеряемый продукт (RFFVAPF) в данной системе представляет собой промежуточный продукт, формирующийся в результате последовательного гидролиза по двум пептидным связям. До тех пор, пока он не накоплен, последующие стадии (в том числе образование следующего продукта P_3) не могут

существенно ограничивать реакцию, поскольку субстрат для этих стадий (P_2) ещё не сформирован. Таким образом, на этапе измерения начальной скорости $[P_2] \approx 0$, и влияние третьей стадии реакции (гидролиз $P_2 \rightarrow P_3$) на кинетику пренебрежимо мало.

В-четвёртых, на начальном временном интервале реакция по образованию RFFVAPF может быть охарактеризована как псевдопервого порядка по отношению к субстрату и описана уравнением Михаэлиса–Ментен. Последующий спад концентрации пептида наблюдается только на поздних стадиях (после 30–40 минут) и не влияет на рассчитанные значения начальной скорости.

Таким образом, начальная скорость реакции (V_0) отражает исключительно параметры стадии образования пептида, и её можно использовать для аппроксимации значений K_M и V_{\max} . При расчёте кинетических параметров учитывались только значения начальной скорости до начала убывания продукта (рисунок 33), что обеспечивает корректное описание кинетики первой стадии. Следует подчеркнуть, что полученные параметры отражают именно скорость образования промежуточного продукта, а не полного гидролиза субстрата. Такой подход допустим, если вклад дальнейших стадий в исследуемом временном интервале является минимальным.

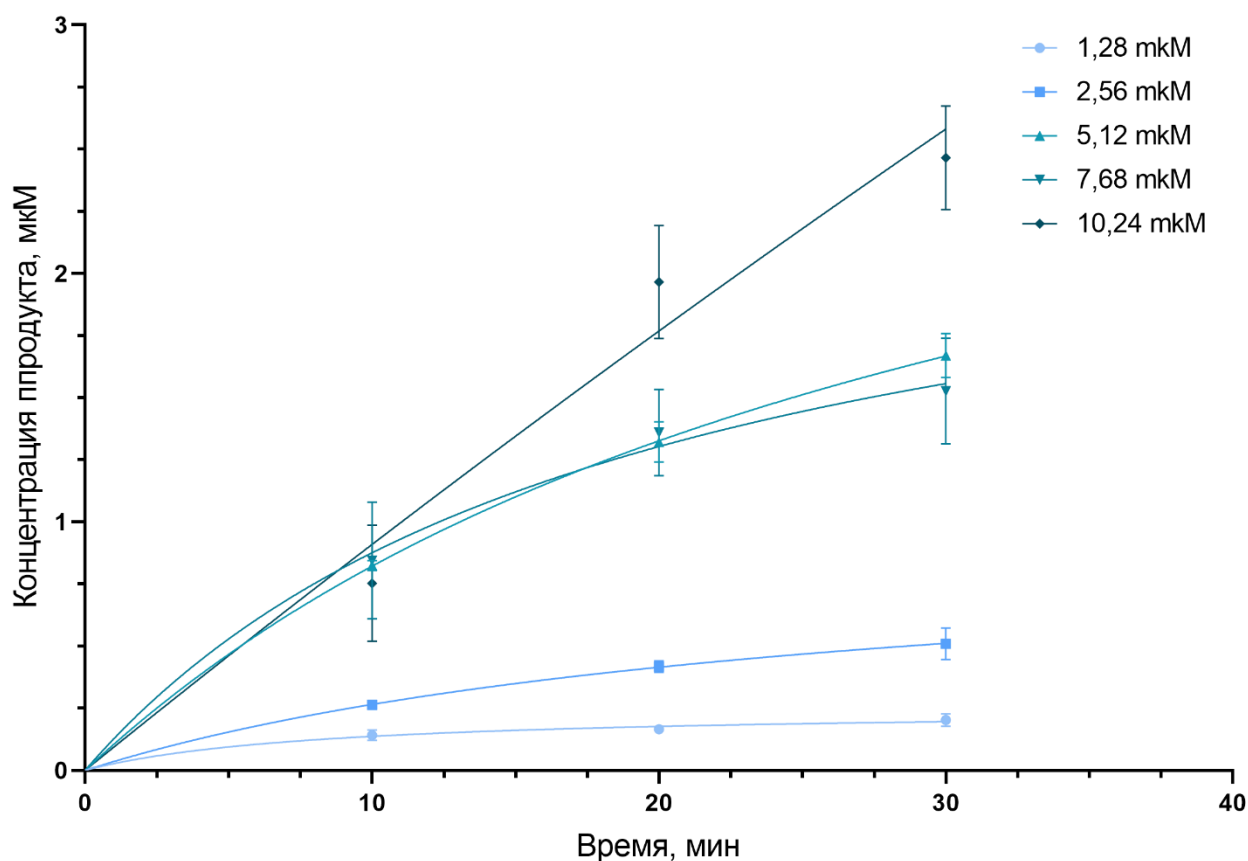


Рисунок 33. Зависимость накопления продукта (RFFVAPF) гидролиза от времени при различных начальных концентрациях субстрата (казеина). Каждая кривая соответствует одной из концентраций субстрата (1,28–10,24 мкМ). По оси ординат – средняя концентрация продукта (мкМ), по оси абсцисс – время инкубации (мин). Показаны средние значения \pm стандартное отклонение. Кривые построены с использованием аппроксимации по методу наименьших квадратов.

Начальная скорость определялась так же, как и для пептида RGPFP_{II}V. По рассчитанным значениям начальной скорости был построен график зависимости начальной скорости реакции (V_0 , мкМ/мин) от исходной концентрации субстрата ($[S]$, мкМ). Все измерения проводились минимум в трёх повторах; на графике представлены средние значения и стандартные отклонения (рисунок 34).

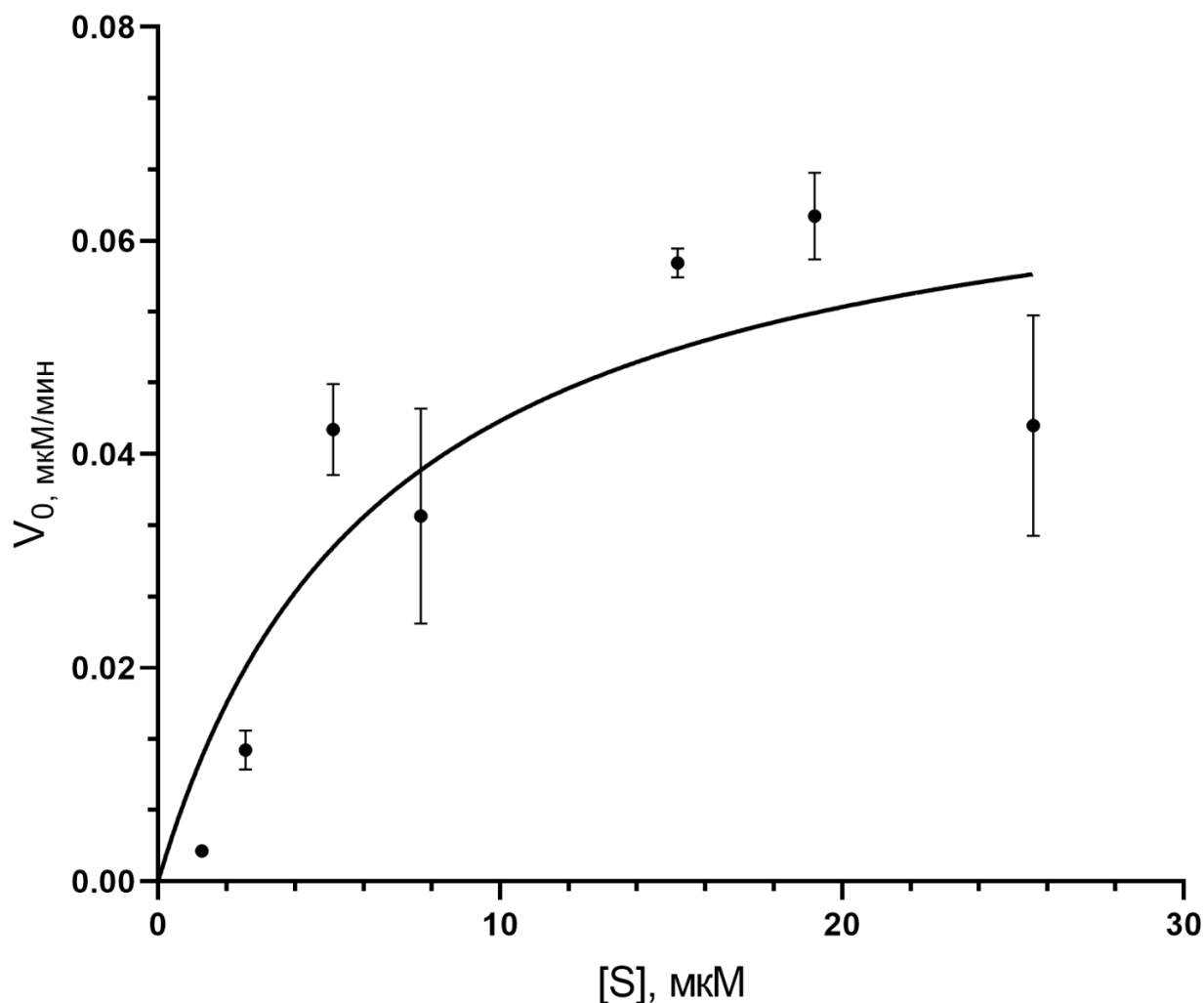


Рисунок 34. Зависимость начальной скорости гидролиза казеина ферментом L5 (V_0 , мкМ/мин) от начальной концентрации субстрата ($[S]$, мкМ). Скорость определялась по накоплению пептида RFFVAPF. Кривая аппроксимации построена по уравнению Михаэлиса–Ментен. Представлены средние значения и стандартные отклонения ($n = 3-6$).

Параметры K_M и V_{\max} для фермента L5 были получены путём аппроксимации зависимости начальной скорости от концентрации субстрата с использованием уравнения Михаэлиса–Ментен. Для пептида RFFVAPF максимальная скорость реакции (V_{\max}) составила $0,07 \pm 0,02$ мкМ/мин, а $K_M - 6,58 \pm 1,72$ мкМ.

С учётом концентрации фермента в реакции (0,147 мкМ), каталитическая константа (k_{cat}) была рассчитана как $0,476 \text{ мин}^{-1}$, что отражает число оборотов молекулы фермента в минуту. Каталитическая эффективность, выраженная как отношение k_{cat}/K_M , составила $0,0723 \text{ мкМ}^{-1}\text{мин}^{-1}$.

Важно учитывать, что пептид RFFVAPF является промежуточным продуктом, образующимся в результате двух последовательных актов гидролиза, и в дальнейшем подвергается гидролизу. Рассчитанные кинетические параметры отражают не полное преобразование субстрата, а только первую стадию последовательной реакции.

В этом контексте значение K_M отражает гидролиз от промежуточного субстрата (S'' , рисунок 32) продукта P_2 , не являющегося конечным. Поскольку данный пептид впоследствии сам подвергается гидролизу, его концентрация в реакции нестабильна, что может приводить к заниженной оценке скорости и, как следствие, недооценке каталитической эффективности. Напротив, пептид RGPFP^{IV}, являющийся стабильным конечным продуктом гидролиза по одной специфической связи, накапливается в системе и позволяет более корректно описать кинетику полного превращения исходного субстрата.

Для определения кинетических параметров V_{pr} использовался пептид AWYYVPLG – продукт гидролиза α S1-казеина по двум остаткам глицина, аминокислоты, к которой данный фермент проявляет специфичность. Поскольку снижение концентрации продукта не наблюдалось, а в составе пептида отсутствуют аминокислоты, подверженные дальнейшему специфическому гидролизу V_{pr} , данный пептид представляет собой конечный продукт реакции.

Для каждого значения начальной концентрации субстрата (в диапазоне 0,89–7,11 мкМ) регистрировалась зависимость концентрации пептида-продукта AWYYVPLG от времени в интервале от 0 до 30 минут. Начальная скорость (V_0) определялась так же, как и для пептидов-продуктов гидролиза ферментом L5. Все измерения проводились в трёх и более повторях; на графике представлены средние значения и стандартные отклонения (рисунок 35).

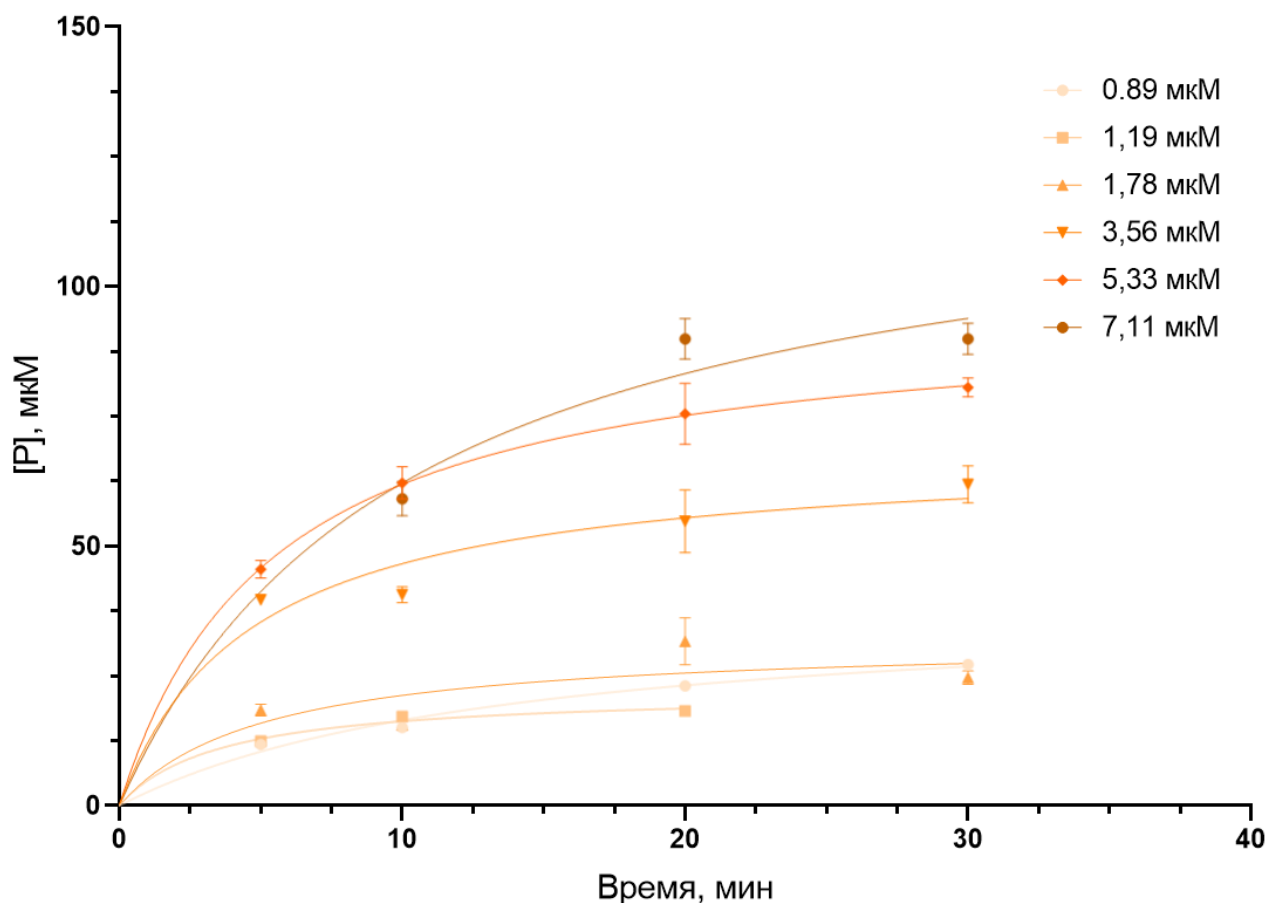


Рисунок 35. График накопления пептида AWYYVPLG, продукта гидролиза α S1-казеина B1p при различных концентрациях субстрата. По оси ординат – средняя концентрация продукта (мкМ), по оси абсцисс – время инкубации (мин). Показаны средние значения \pm стандартное отклонение. Кривые построены с использованием аппроксимации по методу наименьших квадратов.

Зависимость начальной скорости гидролиза от концентрации субстрата была аппроксимирована с использованием уравнения Михаэлиса–Ментен, что позволило рассчитать основные кинетические параметры фермента B1p. По этим значениям был построен график зависимости начальной скорости реакции (V_0 , мкМ/мин) от исходной концентрации субстрата ($[S]$, мкМ) (рисунок 36).

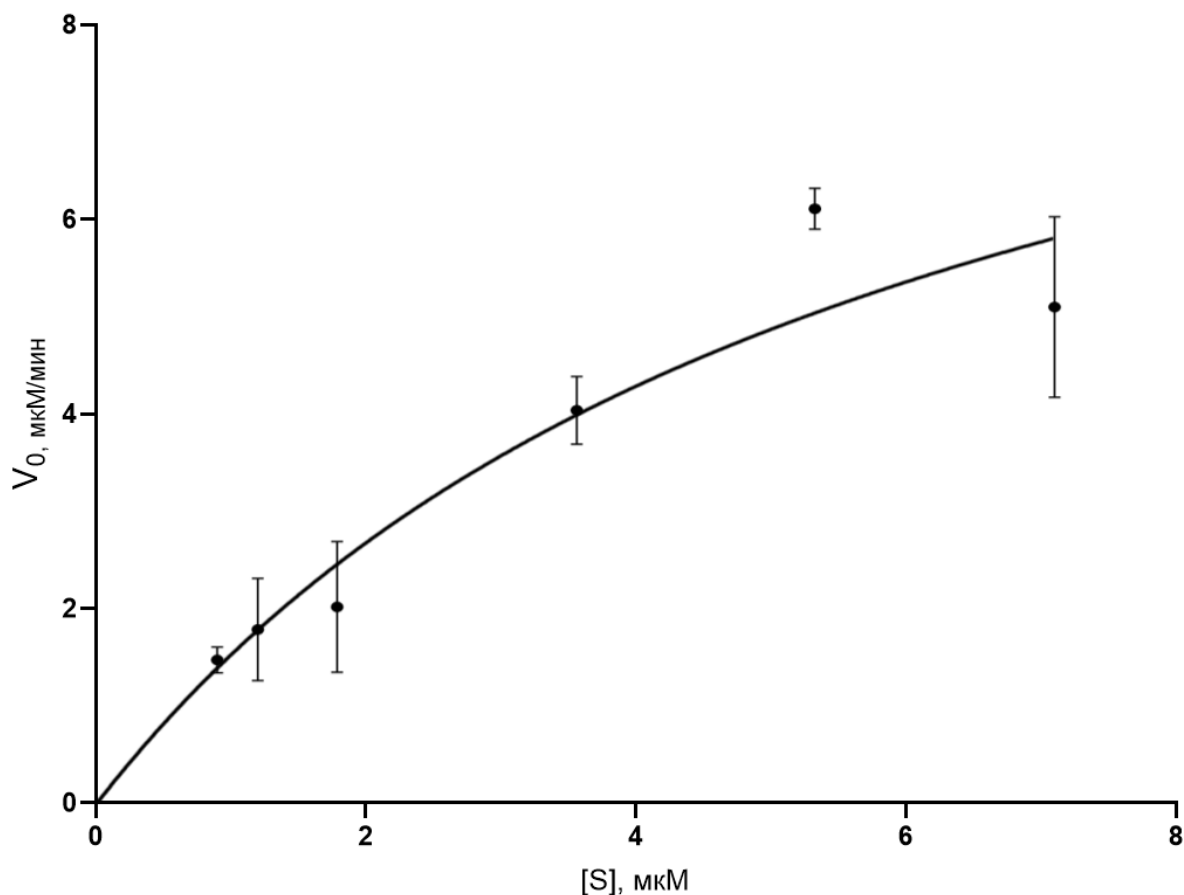


Рисунок 36. Зависимость начальной скорости гидролиза казеина (V_0 , $\mu\text{M}/\text{мин}$) от начальной концентрации субстрата ($[S]$, μM) для фермента Bpr. Скорость определялась по накоплению пептида AWYYVPLG. Кривая аппроксимации построена по уравнению Михаэлиса–Ментен. Представлены средние значения и стандартные отклонения ($n = 3-6$).

Константа Михаэлиса (K_M) составила $6,15 \pm 0,48$ μM , максимальная скорость реакции (V_{\max}) $10,32 \pm 2,6$ $\mu\text{M}/\text{мин}$. С учётом концентрации фермента в реакции (1,69 μM), каталитическая константа (k_{cat}) была рассчитана как 6,11 мин^{-1} . Каталитическая эффективность, выраженная как отношение k_{cat}/K_M , составила 0,993 $\mu\text{M}^{-1}\text{мин}^{-1}$.

Для фермента L5 значения K_M оказались сходными для обоих исследованных пептидов: $6,55 \pm 0,58$ μM для С-концевого фрагмента β -казеина RGPFP_{IV} и $6,58 \pm 1,72$ μM для пептида RFFVAPF. Это указывает на практически одинаковую аффинность фермента к обоим субстратам. Однако, каталитическая активность различалась значительно. Для пептида RGPFP_{IV}, являющегося конечным продуктом гидролиза, значение k_{cat} составило 2,14 мин^{-1} , а каталитическая

эффективность достигала $0,327 \text{ мкМ}^{-1} \cdot \text{мин}^{-1}$, что указывает на эффективное и быстрое превращение субстрата. Напротив, для пептида RFFVAPF, являющегося промежуточным продуктом последовательного гидролиза по двум связям и подверженного дальнейшему расщеплению, k_{cat} ($0,476 \text{ мин}^{-1}$) и каталитическая эффективность ($0,0723 \text{ мкМ}^{-1} \cdot \text{мин}^{-1}$) оказались почти в 4,5 раза ниже.

Низкая каталитическая эффективность в случае RFFVAPF, по-видимому, обусловлена его промежуточной ролью в каскаде гидролитических превращений, в результате чего наблюдаемая скорость характеризует лишь частичную стадию реакции, а не полное превращение субстрата. Таким образом, более высокие значения кинетических параметров для RGPFPiV могут указывать на более высокое предпочтение данного сайта расщепления и его большую кинетическую значимость в условиях эксперимента.

Фермент Vlr проявил несколько иную кинетическую характеристику. Несмотря на сопоставимую с L5 аффинность к субстрату ($K_M = 6,15 \pm 0,48 \text{ мкМ}$), каталитическая константа оказалась значительно выше ($k_{\text{cat}} = 6,11 \text{ мин}^{-1}$), что отражает более активный катализ. Итоговая каталитическая эффективность составила $0,993 \text{ мкМ}^{-1} \cdot \text{мин}^{-1}$ – что существенно превышает значения, полученные для пептидов, образующихся в результате гидролиза ферментом L5 (RFFVAPF и RGPFPiV). Это указывает на высокую способность Vlr быстро и эффективно расщеплять субстрат, при сохранении высокого сродства к субстрату.

Применённый метод определения кинетических параметров ферментов с использованием изотопной метки является универсальным и может быть использован для любой протеазы, независимо от её специфичности и наличия коммерчески доступных субстратов. Поскольку образующиеся в результате гидролиза пептиды имеют как минимум два атома кислорода на С-концевой карбоксильной группе, способные к обмену на изотоп ^{18}O . Также для каждой протеазы можно подобрать пептид, по которому проводить расчет, однако такой пептид должен отвечать ряду требований. Такой метод обеспечивает высокую чувствительность, селективность и точность определения концентрации продукта

реакции даже в сложной белковой матрице, а также позволяет избежать использования флуоресцентных или хромогенных субстратов. В случае промежуточных продуктов расчёт параметров допустим лишь при соблюдении условий минимального вклада последующих стадий в начальный временной интервал, как это выполнялось в данной работе.

С использованием данной методики впервые были получены кинетические параметры для протеазы V1p, для которой ранее не удавалось подобрать подходящие коммерческие субстраты, а также для протеазы L5, обладающей широкой специфичностью. Полученные значения K_M , k_{cat} и k_{cat}/K_M позволили количественно охарактеризовать различия в каталитической эффективности этих ферментов и подтвердили применимость метода для систем с низкой активностью или нестандартными точками приложения.

5. ЗАКЛЮЧЕНИЕ

Протеолитические ферменты бактерий рода *Lysobacter* представляют значительный интерес как с фундаментальной, так и с прикладной точки зрения. Их способность разрушать клеточные стенки других микроорганизмов делает эти ферменты потенциальными инструментами для биотехнологии, биоконтроля и антимикробных разработок. Однако практическое применение таких ферментов ограничено недостаточной изученностью их молекулярной специфичности и каталитических свойств, а также отсутствием универсальных методов для их количественного анализа.

В данной работе выполнено комплексное исследование двух внеклеточных протеаз – Vlr и L5, продуцируемых бактериями *Lysobacter capsici* ВКМ В-2533^T и культуры *Lysobacter capsici* XL1 соответственно. На первом этапе были выделены и идентифицированы активные формы ферментов, установлена их гомология с литическими протеазами других представителей рода *Lysobacter*. Для анализа специфичности был применён подход, основанный на неселективном протеолизе широкого набора белков с последующей идентификацией пептидов методом LC-MS/MS. С целью обработки и визуализации полученных данных разработан специализированный программный инструмент CleaveScope, позволяющий анализировать частоту встречаемости аминокислот, используя различные методы нормировки, а также визуализировать результаты в виде тепловых карт и логотипов последовательностей.

Проведённый анализ позволил впервые охарактеризовать аминокислотные предпочтения ферментов Vlr и L5. Установлено, что Vlr обладает выраженной селективностью к остаткам глицина и лизина в позиции P₁, в то время как фермент L5 проявляет более широкую специфичность и способен расщеплять пептидные связи аминокислот метионина, фенилаланина, лейцина и валина. Различия в предпочтениях подтверждают существование различных механизмов взаимодействия ферментов с субстратом. Помимо этого, было показано, что

фермент имеет высокую частоту встречаемости пролина в позиции P₂. Эти наблюдения позволяют предположить, что специфичность L5 может зависеть не только от взаимодействия с отдельными аминокислотными остатками, но и от позиционных или конформационных особенностей активного центра фермента.

Для количественной оценки каталитических свойств была разработана методика определения кинетических параметров с использованием изотопно-меченых (¹⁸O) пептидных стандартов и MALDI-TOF масс-спектрометрии. Впервые получены значения K_M , k_{cat} и k_{cat}/K_M для фермента Vlp, активность которого не удавалось оценить стандартными методами. Также определены кинетические параметры L5 на двух различных пептидных субстратах, что позволило охарактеризовать его каталитическую эффективность в зависимости от субстрата.

Анализ полученных кинетических параметров показал, что различия в специфичности ферментов Vlp и L5 коррелируют с их функциональными особенностями. Vlp, обладающий узкой селективностью к остаткам глицина и лизина, имеет высокую каталитическую активность и эффективность, что отражает его роль основного лизирующего фермента, способного быстро расщеплять белковые субстраты. Напротив, L5 характеризуется более широкой аминокислотной специфичностью и меньшей скоростью гидролиза промежуточных пептидов, что указывает на его адаптивную, вспомогательную роль в бактериальном секрете. Полученные данные позволяют объяснить их взаимодополняющую функциональную роль.

Таким образом, настоящая работа вносит вклад в изучение природных бактериальных протеаз с литической активностью и предлагает универсальные инструменты и подходы для анализа их специфичности и активности. Полученные результаты расширяют представления о биохимических свойствах ферментов *Lysobacter* и создают основу для их дальнейшего применения в биотехнологии.

6. ВЫВОДЫ

1. Из внеклеточной фракции культуры *Lysobacter capsici* ВКМ В-2533^Т и *Lysobacter capsici* XL1 были выделены и идентифицированы протеолитические ферменты В1р и L5 соответственно, обладающие литической активностью. Проведённый BLAST-анализ аминокислотных последовательностей показал их высокую степень гомологии с известными внеклеточными протеазами бактерий рода *Lysobacter*, включая протеазу LasA. Полученные данные указывают на принадлежность В1р к классу металлопротеаз M23, а L5 – к классу сериновых эндопептидаз.

2. Для обработки большого объёма масс-спектрометрических данных и визуализации аминокислотной специфичности ферментов была разработана специализированная программа CleaveScore. Данный программный инструмент позволил автоматически извлекать сайты гидролиза, строить матрицы частот остатков в позициях P₁₂-P₁₂', формировать тепловые карты и логотипы последовательностей. Программа была предварительно протестирована с использованием трипсина в качестве модельного фермента с известной специфичностью.

3. На основе данных масс-спектрометрического анализа (LC-MS/MS) была определена аминокислотная специфичность двух внеклеточных протеаз – В1р и L5. Фермент В1р имеет предпочтение к гидролизу пептидных связей после остатков глицина и лизина, тогда как фермент L5 обладает более широкой специфичностью гидролизуя связи после метионина, фенилаланина, лейцина и валина, при этом обладает склонностью к образованию более коротких фрагментов. Построенные матрицы частот и логотипы последовательностей позволили выявить аминокислотные предпочтения фермента, а также определить наиболее характерные мотивы. Полученные данные указывают на различия в механизме действия ферментов, предполагая, что В1р обладает более выраженной селективностью, тогда как L5 имеет более широкую субстратную специфичность.

4. Разработана и реализована модельная система для определения кинетических параметров ферментативных реакций с использованием изотопно-меченых пептидных стандартов (^{18}O) и MALDI-TOF масс-спектрометрии. Предложенный подход позволяет оценить кинетические параметры фермента без использования искусственных субстратов и обеспечивает высокую чувствительность, специфичность и универсальность при анализе ферментов.

5. Впервые были определены кинетические параметры протеаз V1p и L5. Для фермента V1p были получены значения $K_M = 6,15 \pm 0,48$ мкМ, $k_{\text{cat}} = 6,11$ мин $^{-1}$, $k_{\text{cat}}/K_M = 0,993$ мкМ $^{-1}$ мин $^{-1}$. Для фермента L5 кинетические параметры рассчитывались по двум различным субстратам. Для пептида RGPFP1V значения составили $K_M = 6,55 \pm 0,58$ мкМ, $k_{\text{cat}} = 2,14$ мин $^{-1}$, $k_{\text{cat}}/K_M = 0,327$ мкМ $^{-1}$ мин $^{-1}$; для пептида RFFVAPF – $K_M = 6,58 \pm 1,72$ мкМ, $k_{\text{cat}} = 0,476$ мин $^{-1}$, $k_{\text{cat}}/K_M = 0,0723$ мкМ $^{-1}$ мин $^{-1}$. Эти данные отражают различия в каталитической активности ферментов и субстратную зависимость их кинетических характеристик, а также демонстрируют практическую применимость разработанного подхода к анализу ферментов с низкой и высокой скоростью реакции.

8. БЛАГОДАРНОСТИ

Автор выражает благодарность научному руководителю И. Ю. Торопыгину, научному консультанту Д. Д. Жданову. Также автор выражает благодарность коллегам из лаборатории биохимии клеточной поверхности микроорганизмов ИБФМ им Г.К. Скрыбина г. Пущино, которые предоставили препараты исследованных в работе ферментов, к.б.н. А. С. Афошину, к.б.н., И. В. Кудряковой и к.б.н. Н. В. Леонтьевской.

Автор выражает благодарность к.б.н. Т. Е. Фарафоновой из лаборатории системной биологии ИБМХ им. В.Н. Ореховича за синтез пептидов RFFVAPF и RGPFPPIV, использованных для определения кинетических параметров L5.

Автор выражает благодарность коллегам из лаборатории межмолекулярных взаимодействий ИБМХ им. В.Н. Ореховича, которые предоставили фракции лизата печени крысы и сыворотки человека, к.б.н. О. В. Гнеденко, к.б.н. Л. А. Калужскому, к.б.н. Ю. В. Мезенцеву, к.б.н. Е. В. Яблокову и д.б.н. профессору А.С. Иванову.

Автор выражает благодарность коллегам из Института биоорганической химии НАН Беларуси (лаборатории молекулярной диагностики и биотехнологии) к.х.н. А. А. Гилепу, А. М. Тумилович, к.х.н. А. В. Свиридау, а также и сотруднику Сколковского института науки и технологии (группа Натальи Струшкевич) к.х.н. Н.В. Струшкевич за предоставленные белки CPOX, PPOX, SULT2A1, FXN, использованные для определения специфичности исследуемых ферментов.

8. ФИНАНСИРОВАНИЕ

Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021 - 2030 годы) (№122030100168-2)

9. СПИСОК ЛИТЕРАТУРЫ

1. Robinson P.K. Enzymes: principles and biotechnological applications // *Essays Biochem.* 2015. Т. 59. С. 1–41.
2. Kermasha S., Eskin M.N.A. Enzymes // *Enzymes.* Elsevier, 2021. С. 15–44.
3. Borkakoti N., Ribeiro A.J.M., Thornton J.M. A structural perspective on enzymes and their catalytic mechanisms // *Curr Opin Struct Biol.* 2025. Т. 92. С. 103040.
4. McDonald A.G., Tipton K.F. Enzyme nomenclature and classification: the state of the art // *FEBS J.* 2023. Т. 290, № 9. С. 2214–2231.
5. Mótyán J., Tóth F., Tózsér J. Research Applications of Proteolytic Enzymes in Molecular Biology // *Biomolecules.* 2013. Т. 3, № 4. С. 923–942.
6. Costa J. de O. и др. Structural and functional comparison of proteolytic enzymes from plant latex and snake venoms // *Biochimie.* 2010. Т. 92, № 12. С. 1760–1765.
7. Bond J.S. Proteases: History, discovery, and roles in health and disease // *Journal of Biological Chemistry.* 2019. Т. 294, № 5. С. 1643–1651.
8. Song P. и др. Microbial proteases and their applications // *Front Microbiol.* 2023. Т. 14.
9. Ćwilichowska N. и др. Diagnostic and therapeutic potential of protease inhibition // *Molecular Aspects of Medicine.* 2022. Т. 88.
10. Tsu B. V. и др. Running With Scissors: Evolutionary Conflicts Between Viral Proteases and the Host Immune System // *Front Immunol.* 2021. Т. 12.
11. Wang L. и др. Biochemical Tools for Tracking Proteolysis // *J Proteome Res.* 2021. Т. 20, № 12. С. 5264–5279.
12. Schauerl M. и др. Characterizing Protease Specificity: How Many Substrates Do We Need? // *PLoS One.* 2015. Т. 10, № 11. С. e0142658.
13. Schilling O., Overall C.M. Proteomic discovery of protease substrates // *Curr Opin Chem Biol.* 2007. Т. 11, № 1. С. 36–45.
14. Kasperkiewicz P. и др. Emerging challenges in the design of selective substrates, inhibitors and activity-based probes for indistinguishable proteases // *FEBS J.* 2017. Т. 284, № 10. С. 1518–1539.
15. Yang M. и др. Digestion of food proteins: the role of pepsin // *Crit Rev Food Sci Nutr.* 2025. С. 1–22.
16. Al-Amer O.M. The role of thrombin in haemostasis // *Blood Coagulation & Fibrinolysis.* 2022. Т. 33, № 3. С. 145–148.
17. Hansen K.K. и др. Proteinases, proteinase-activated receptors (PARs) and the pathophysiology of cancer and diseases of the cardiovascular, musculoskeletal, nervous and gastrointestinal systems // *Naunyn Schmiedebergs Arch Pharmacol.* 2008. Т. 377, № 4–6. С. 377–392.
18. Donado C.A. и др. Granzyme K activates the entire complement cascade // *Nature.* 2025. Т. 641, № 8061. С. 211–221.

19. Turner C.T. Pro-inflammatory granzyme K contributes extracellularly to disease // *Front Immunol.* 2025. T. 16.
20. Knecht E. и др. Intracellular protein degradation in mammalian cells: recent developments // *Cellular and Molecular Life Sciences.* 2009. T. 66, № 15. С. 2427–2443.
21. Rusilowicz-Jones E. V., Urbé S., Clague M.J. Protein degradation on the global scale // *Mol Cell.* 2022. T. 82, № 8. С. 1414–1423.
22. Ketterer S. и др. Inherited diseases caused by mutations in cathepsin protease genes // *FEBS J.* 2017. T. 284, № 10. С. 1437–1454.
23. Crocetti L. и др. Editorial: Serine protease inhibitors and their therapeutic applications // *Front Chem.* 2022. T. 10.
24. Zheng W. и др. Small molecule angiotensin converting enzyme inhibitors: A medicinal chemistry perspective // *Front Pharmacol.* 2022. T. 13.
25. Wang Y., Lv Z., Chu Y. HIV protease inhibitors: a review of molecular selectivity and toxicity // *HIV/AIDS - Research and Palliative Care.* 2015. С. 95.
26. Crepaldi G. и др. Dipeptidyl peptidase 4 (DPP-4) inhibitors and their role in Type 2 diabetes management // *J Endocrinol Invest.* 2007. T. 30, № 7. С. 610–614.
27. Hampel H. и др. The β -Secretase BACE1 in Alzheimer's Disease // *Biol Psychiatry.* 2021. T. 89, № 8. С. 745–756.
28. Raeeszadeh-Sarmazdeh M., Do L., Hritz B. Metalloproteinases and Their Inhibitors: Potential for the Development of New Therapeutics // *Cells.* 2020. T. 9, № 5. С. 1313.
29. Ramkumar A. и др. Production of thermotolerant, detergent stable alkaline protease using the gut waste of *Sardinella longiceps* as a substrate: Optimization and characterization // *Sci Rep.* 2018. T. 8, № 1. С. 12442.
30. Lindsay S., Oates A., Bourdillon K. The detrimental impact of extracellular bacterial proteases on wound healing // *Int Wound J.* 2017. T. 14, № 6. С. 1237–1247.
31. Suaifan G.A.R.Y. и др. Antibiotic–*Lysobacter* enzymogenes proteases combination as a novel virulence attenuating therapy // *PLoS One.* 2023. T. 18, № 3. С. e0282705.
32. Afoshin A.S. и др. Lytic potential of *Lysobacter capsici* VKM B-2533^T: bacteriolytic enzymes and outer membrane vesicles // *Sci Rep.* 2020. T. 10, № 1.
33. Tsai C.S. и др. Lytic enzymes of *Sorangium* sp. Action of the alpha- and beta-lytic proteases on two bacterial mucopeptides. // *Can J Biochem.* 1965. T. 43, № 12. С. 1971–1983.
34. Kudryakova I. и др. The First Homologous Expression System for the β -Lytic Protease of *Lysobacter capsici* VKM B-2533^T, a Promising Antimicrobial Agent // *Int J Mol Sci.* 2022. T. 23, № 10.
35. Afoshin A.S. и др. β -Lytic Protease of *Lysobacter capsici* VKM B-2533^T // *Antibiotics.* 2020. T. 9, № 11. С. 1–14.
36. Rawlings N.D., Barrett A.J. Evolutionary families of metallopeptidases // *Methods Enzymol.* Academic Press, 1995. T. 248, № C. С. 183–228.

37. Rawlings N.D., Morton F.R., Barrett A.J. MEROPS: the peptidase database. // *Nucleic Acids Res.* 2006. Т. 34, № Database issue. С. D270–D272.
38. Spencer J. и др. Crystal structure of the lasA virulence factor from *Pseudomonas aeruginosa*: Substrate specificity and mechanism of M23 metallopeptidases // *J Mol Biol.* 2010. Т. 396, № 4. С. 908–923.
39. Hooper N.M. Families of zinc metalloproteases // *FEBS Lett.* No longer published by Elsevier, 1994. Т. 354, № 1. С. 1–6.
40. Jekel P.A., Weijer W.J., Beintema J.J. Use of endoproteinase Lys-C from *Lysobacter* enzymogenes in protein sequence analysis // *Anal Biochem.* 1983. Т. 134, № 2. С. 347–354.
41. Vasilyeva N. V. и др. Lytic peptidase L5 of *lysobacter* sp. XL1 with broad antimicrobial spectrum // *J Mol Microbiol Biotechnol.* 2014. Т. 24, № 1. С. 59–66.
42. Begunova E.A. и др. Specificity of the Action of Lysoamidase on *Staphylococcus aureus* 209P Cell Walls // *Biochemistry (Moscow).* 2003. Т. 68, № 7. С. 735–739.
43. О.А. Степная Е.А.Б.И.М.Ц.И.С.К. Бактериолитический ферментный препарат лизоамидаза. Выделение и определение некоторых физико-химических свойств внеклеточной мурамидазы бактерии рода *Xanthomonas* // *БИОХИМИЯ.* 1996. Т. 61, № 4. С. 648–655.
44. Т.А. Муранова Л.А.К.И.М.Ц.О.А.С.И.С.К. Структурные исследования и идентификация внеклеточной бактериолитической эндопептидазы Л1 *Lysobacter* sp. XL1 // *БИОХИМИЯ.* 2004. Т. 69, № 5. С. 617–622.
45. Vasilyeva N. V. и др. Outer membrane vesicles of *Lysobacter* sp. // *Dokl Biochem Biophys.* 2009. Т. 426, № 1. С. 139–142.
46. Stepnaya O.A. и др. Isolation and Characterization of a New Extracellular Bacteriolytic Endopeptidase of *Lysobacter* sp. XL1 // *Biochemistry (Moscow).* 2005. Т. 70, № 9. С. 1031–1037.
47. Vasilyeva N. V. и др. Secretion of bacteriolytic endopeptidase L5 of *Lysobacter* sp. XL1 into the medium by means of outer membrane vesicles // *FEBS Journal.* 2008. Т. 275, № 15. С. 3827–3835.
48. Kudryakova I. V. и др. Structural and functional properties of antimicrobial protein L5 of *Lysobacter* sp. XL1 // *Appl Microbiol Biotechnol.* 2018. Т. 102, № 23. С. 10043–10053.
49. Epstein D.M., Wensink P.C. The alpha-lytic protease gene of *Lysobacter* enzymogenes. The nucleotide sequence predicts a large prepro-peptide with homology to pro-peptides of other chymotrypsin-like enzymes. // *Journal of Biological Chemistry.* 1988. Т. 263, № 32. С. 16586–16590.
50. Fujinaga M. и др. Refined structure of α -lytic protease at 1.7 Å resolution analysis of hydrogen bonding and solvent structure // *J Mol Biol.* 1985. Т. 184, № 3. С. 479–502.
51. Hioki T. и др. Heterologous production of active form of beta-lytic protease by *Bacillus subtilis* and improvement of staphylolytic activity by protein engineering // *Microb Cell Fact.* 2021. Т. 20, № 1.

52. Kobayashi D.Y., Yuen G.Y. The potential of *Lysobacter* spp. as bacterial biological control agents for plant diseases. // CABI Reviews. 2007.
53. Tsiatsiani L., Heck A.J.R. Proteomics beyond trypsin // FEBS Journal. 2015. T. 282, № 14. C. 2612–2626.
54. Jiang Y. и др. Comprehensive Overview of Bottom-Up Proteomics Using Mass Spectrometry // ACS Measurement Science Au. 2024. T. 4, № 4. C. 338–417.
55. Schechter I., Berger A. On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain // Biochem Biophys Res Commun. 1968. T. 32, № 5. C. 898–902.
56. Schechter I., Berger A. On the size of the active site in proteases. I. Papain // Biochem Biophys Res Commun. 1967. T. 27, № 2. C. 157–162.
57. Diamond S.L. Methods for mapping protease specificity // Curr Opin Chem Biol. 2007. T. 11, № 1. C. 46–51.
58. Boulware K.T., Daugherty P.S. Protease specificity determination by using cellular libraries of peptide substrates (CLiPS) // Proceedings of the National Academy of Sciences. 2006. T. 103, № 20. C. 7583–7588.
59. Lu C. и др. Prediction and design of protease enzyme specificity using a structure-aware graph convolutional network // Proceedings of the National Academy of Sciences. 2023. T. 120, № 39.
60. Khare S.D., Rubenstein A., Pethe M. Structure-Based Prediction of Protease Multispecificity using Computational Protein Design // Biophys J. 2016. T. 110, № 3. C. 345a.
61. Zhang Y., Sanner M.F. *AutoDock CrankPep*: combining folding and docking to predict protein–peptide complexes // Bioinformatics. 2019. T. 35, № 24. C. 5121–5127.
62. Ochoa R., Santiago Á., Alegría-Arcos M. Open protocols for docking and MD-based scoring of peptide substrates // Artificial Intelligence in the Life Sciences. 2022. T. 2. C. 100044.
63. Sanfaçon H., Skern T. AlphaFold modeling of nepovirus 3C-like proteinases provides new insights into their diverse substrate specificities // Virology. 2024. T. 590. C. 109956.
64. Terwilliger T.C. и др. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination // Nat Methods. 2024. T. 21, № 1. C. 110–116.
65. Poreba M. и др. Unnatural amino acids increase sensitivity and provide for the design of highly selective caspase substrates // Cell Death Differ. 2014. T. 21, № 9. C. 1482–1492.
66. Schneider E.L., Craik C.S. Positional Scanning Synthetic Combinatorial Libraries for Substrate Profiling. 2009. C. 59–78.
67. Rano T.A. и др. A combinatorial approach for determining protease specificities: application to interleukin-1 β converting enzyme (ICE) // Chem Biol. 1997. T. 4, № 2. C. 149–155.
68. Thornberry N.A. и др. A Combinatorial Approach Defines Specificities of Members of the Caspase Family and Granzyme B // Journal of Biological Chemistry. 1997. T. 272, № 29. C. 17907–17911.

69. Maly D.J. и др. Expedient Solid-Phase Synthesis of Fluorogenic Protease Substrates Using the 7-Amino-4-carbamoylmethylcoumarin (ACC) Fluorophore // *J Org Chem*. 2002. Т. 67, № 3. С. 910–915.
70. Harris J.L. и др. Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries // *Proceedings of the National Academy of Sciences*. 2000. Т. 97, № 14. С. 7754–7759.
71. Kasperkiewicz P. и др. Design of ultrasensitive probes for human neutrophil elastase through hybrid combinatorial substrate library profiling // *Proceedings of the National Academy of Sciences*. 2014. Т. 111, № 7. С. 2518–2523.
72. Poreba M. и др. Counter Selection Substrate Library Strategy for Developing Specific Protease Substrates and Probes // *Cell Chem Biol*. 2016. Т. 23, № 8. С. 1023–1035.
73. Boulware K.T., Daugherty P.S. Protease specificity determination by using cellular libraries of peptide substrates (CLiPS) // *Proceedings of the National Academy of Sciences*. 2006. Т. 103, № 20. С. 7583–7588.
74. Deperthes D. Phage Display Substrate: A Blind Method for Determining Protease Specificity // *Biol Chem*. 2002. Т. 383, № 7–8.
75. Beck Z.Q., Lin Y.-C., Elder J.H. Molecular Basis for the Relative Substrate Specificity of Human Immunodeficiency Virus Type 1 and Feline Immunodeficiency Virus Proteases // *J Virol*. 2001. Т. 75, № 19. С. 9458–9469.
76. Ratnikov B., Cieplak P., Smith J.W. High Throughput Substrate Phage Display for Protease Profiling. 2009. С. 93–114.
77. Kretz C.A. и др. High throughput protease profiling comprehensively defines active site specificity for thrombin and ADAMTS13 // *Sci Rep*. 2018. Т. 8, № 1. С. 2788.
78. <https://www.sigmaaldrich.com/US/en/technical-documents/protocol/protein-biology/enzyme-activity-assays/enzymatic-assay-of-trypsin?srltid=AfmBOorUCS7xSBonL2HXOKDCT3W62SOx5IbaNSD01V2RAY8QephNnfAX> [Электронный ресурс].
79. Guilbault G.G. Newer fluorometric methods for the analysis of biologically important compounds // *J Res Natl Bur Stand A Phys Chem*. 1972. Т. 76A, № 6. С. 607.
80. Khangembam B.K., YVR K.S., Chakrabarti R. Purification and characterization of trypsin from the digestive system of carp *Catla catla* (Hamilton) // *Int Aquat Res*. 2012. Т. 4, № 1. С. 9.
81. https://www.sigmaaldrich.com//US/en/product/sigma/mak290?srltid=AfmBOopi8hILWpF2nt2rSralHoySvE2RSnNkjwSO9jEvKUUp_Ggoczxu#:~:text=In%20this%20assay%2C%20trypsin%20cleaves,trypsin%20in%20various%20samples [Электронный ресурс].
82. Ong I.L.H., Yang K.-L. Recent developments in protease activity assays and sensors // *Analyst*. 2017. Т. 142, № 11. С. 1867–1881.
83. Matayoshi E.D. и др. Novel Fluorogenic Substrates for Assaying Retroviral Proteases by Resonance Energy Transfer // *Science* (1979). 1990. Т. 247, № 4945. С. 954–958.

84. Hyland L.J. и др. A radiometric assay for HIV-1 protease // *Anal Biochem.* 1990. Т. 188, № 2. С. 408–415.
85. Song Y. и др. Electrochemical Activity Assay for Protease Analysis Using Carbon Nanofiber Nanoelectrode Arrays // *Anal Chem.* 2019. Т. 91, № 6. С. 3971–3979.
86. Feng Y. и др. Colorimetric and Electrochemical Methods for the Detection of SARS-CoV-2 Main Protease by Peptide-Triggered Assembly of Gold Nanoparticles // *Molecules.* 2022. Т. 27, № 3. С. 615.
87. Liu K. Trypsin Inhibitor Assay: Expressing, Calculating, and Standardizing Inhibitor Activity in Absolute Amounts of Trypsin Inhibited or Trypsin Inhibitors // *J Am Oil Chem Soc.* 2021. Т. 98, № 4. С. 355–373.
88. Keil B. Specificity of Proteolysis. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992.
89. Uliana F. и др. Mapping specificity, cleavage entropy, allosteric changes and substrates of blood proteases in a high-throughput screen // *Nat Commun.* 2021. Т. 12, № 1. С. 1693.
90. Rodriguez J. и др. Does Trypsin Cut Before Proline? // *J Proteome Res.* 2008. Т. 7, № 1. С. 300–305.
91. Yamamoto H. и др. Identification of Protease Specificity Using Biotin-Labeled Substrates // *Open Biochem J.* 2017. Т. 11, № 1. С. 27–35.
92. Lapek J.D. и др. Quantitative Multiplex Substrate Profiling of Peptidases by Mass Spectrometry // *Molecular & Cellular Proteomics.* 2019. Т. 18, № 5. С. 968a–9981.
93. Münzenberg G. Development of mass spectrometers from Thomson and Aston to present // *Int J Mass Spectrom.* 2013. Т. 349–350. С. 9–18.
94. Александров М. Л. Г.Л.Л., К.Н.В., Н.В.И., П.В.А., Ш.В.А.. ЭКСТРАКЦИЯ ИОНОВ ИЗ РАСТВОРОВ ПРИ АТМОСФЕРНОМ ДАВЛЕНИИ - МЕТОД МАСС-СПЕКТРОМЕТРИЧЕСКОГО АНАЛИЗА БИООРГАНИЧЕСКИХ ВЕЩЕСТВ // *МАСС-СПЕКТРОМЕТРИЯ.* 2005. Т. 2, № 1. С. 6–10.
95. Wilm M. Principles of Electrospray Ionization // *Molecular & Cellular Proteomics.* 2011. Т. 10, № 7. С. M111.009407.
96. Banerjee S., Mazumdar S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. // *Int J Anal Chem.* 2012. Т. 2012. С. 282574.
97. Konermann L. и др. Unraveling the Mechanism of Electrospray Ionization // *Anal Chem.* 2013. Т. 85, № 1. С. 2–9.
98. Bian Y., Gao C., Kuster B. On the potential of micro-flow LC-MS/MS in proteomics // *Expert Rev Proteomics.* 2022. Т. 19, № 3. С. 153–164.
99. Knochenmuss R. MALDI Ionization Mechanisms: An Overview // *Electrospray and MALDI Mass Spectrometry: Fundamentals, Instrumentation, Practicalities, and Biological Applications: Second Edition.* John Wiley & Sons, Ltd, 2012. С. 147–183.
100. Molin L. и др. On the Primary Ionization Mechanism(s) in Matrix-Assisted Laser Desorption Ionization // *J Anal Methods Chem.* 2012. Т. 2012. С. 1–8.

101. Wiley W.C., McLaren I.H. Time-of-flight mass spectrometer with improved resolution // Review of Scientific Instruments. 1955. T. 26, № 12. C. 1150–1157.
102. Brodbelt J.S. Ion Activation Methods for Peptides and Proteins // Anal Chem. 2016. T. 88, № 1. C. 30–51.
103. Steen H., Mann M. The abc's (and xyz's) of peptide sequencing // Nat Rev Mol Cell Biol. 2004. T. 5, № 9. C. 699–711.
104. Jedrychowski M.P. и др. Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics // Molecular & Cellular Proteomics. 2011. T. 10, № 12. C. M111.009910.
105. <https://www.uniprot.org/uniprotkb/statistics> [Электронный ресурс].
106. Switzer L., Giera M., Niessen W.M.A. Protein digestion: An overview of the available techniques and recent developments // Journal of Proteome Research. 2013. T. 12, № 3. C. 1067–1077.
107. Gonnet F. и др. MALDI/MS peptide mass fingerprinting for proteome analysis: identification of hydrophobic proteins attached to eucaryote keratinocyte cytoplasmic membrane using different matrices in concert // Proteome Sci. 2003. T. 1, № 1. C. 2.
108. Ritorto M., Borlak J. A simple and reliable protocol for mouse serum proteome profiling studies by use of two-dimensional electrophoresis and MALDI TOF/TOF mass spectrometry // Proteome Sci. 2008. T. 6, № 1. C. 25.
109. Yuan Z.-F. и др. Evaluation of Proteomic Search Engines for the Analysis of Histone Modifications // J Proteome Res. 2014. T. 13, № 10. C. 4470–4478.
110. Brosch M. и др. Comparison of Mascot and X!Tandem Performance for Low and High Accuracy Mass Spectrometry and the Development of an Adjusted Mascot Threshold // Molecular & Cellular Proteomics. 2008. T. 7, № 5. C. 962–970.
111. Cox J. и др. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment // J Proteome Res. 2011. T. 10, № 4. C. 1794–1805.
112. Zhang J. и др. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification // Molecular & Cellular Proteomics. 2012. T. 11, № 4. C. M111.010587.
113. Perkins D.N. и др. Probability-based protein identification by searching sequence databases using mass spectrometry data // Electrophoresis. Germany, 1999. T. 20, № 18. C. 3551–3567.
114. Chamrad D.C. и др. Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data // Proteomics. 2004. T. 4, № 3. C. 619–628.
115. Pappin D.J.C., Hojrup P., Bleasby A.J. Rapid identification of proteins by peptide-mass fingerprinting // Current Biology. 1993. T. 3, № 6. C. 327–332.
116. Meyer J.G. и др. Expanding proteome coverage with orthogonal-specificity α -lytic proteases // Molecular and Cellular Proteomics. 2014. T. 13, № 3. C. 823–835.
117. Siepen J.A. и др. Prediction of Missed Cleavage Sites in Tryptic Peptides Aids Protein Identification in Proteomics // J Proteome Res. 2007. T. 6, № 1. C. 399–408.

118. Colaert N. и др. Improved visualization of protein consensus sequences by iceLogo // *Nat Methods*. 2009. Т. 6, № 11. С. 786–787.
119. Crooks G.E. и др. WebLogo: A Sequence Logo Generator // *Genome Res*. 2004. Т. 14, № 6. С. 1188–1190.
120. Ou J. и др. dagLogo: An R/Bioconductor package for identifying and visualizing differential amino acid group usage in proteomics data // *PLoS One* / под ред. Lisacek F. 2020. Т. 15, № 11. С. e0242030.
121. Andreatta M., Lund O., Nielsen M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach // *Bioinformatics*. 2013. Т. 29, № 1. С. 8–14.
122. O'Shea J.P. и др. pLogo: a probabilistic approach to visualizing sequence motifs // *Nat Methods*. 2013. Т. 10, № 12. С. 1211–1212.
123. Bantscheff M. и др. Quantitative mass spectrometry in proteomics: a critical review // *Anal Bioanal Chem*. 2007. Т. 389, № 4. С. 1017–1031.
124. Liigand J. и др. Quantification for non-targeted LC/MS screening without standard substances // *Sci Rep*. 2020. Т. 10, № 1. С. 5808.
125. Chen X. и др. Quantitative proteomics using SILAC: Principles, applications, and developments // *Proteomics*. 2015. Т. 15, № 18. С. 3175–3192.
126. Vogl D.P., Conibear A.C., Becker C.F.W. Segmental and site-specific isotope labelling strategies for structural analysis of posttranslationally modified proteins // *RSC Chem Biol*. 2021. Т. 2, № 5. С. 1441–1461.
127. Chahrour O., Cobice D., Malone J. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics // *J Pharm Biomed Anal*. 2015. Т. 113. С. 2–20.
128. Petriz B.A., Franco O.L. Application of cutting-edge proteomics technologies for elucidating host-bacteria interactions // *Advances in Protein Chemistry and Structural Biology*. 2014. Т. 95. С. 1–24.
129. Kozmin Y.P. и др. A direct introduction of ^{18}O isotopes into peptides and proteins for quantitative mass spectroscopy analysis // *Russ J Bioorg Chem*. 2011. Т. 37, № 6. С. 719–731.
130. Ye X. и др. ^{18}O Stable Isotope Labeling in MS-based Proteomics // *Brief Funct Genomic Proteomic*. 2009. Т. 8, № 2. С. 136–144.
131. Ye X. и др. Optimized Method for Computing $^{18}\text{O}/^{16}\text{O}$ Ratios of Differentially Stable-Isotope Labeled Peptides in the Context of Postdigestion ^{18}O Exchange/Labeling // *Anal Chem*. 2010. Т. 82, № 13. С. 5878–5886.
132. Yao X. и др. Proteolytic ^{18}O labeling for comparative proteomics: Model studies with two serotypes of adenovirus // *Anal Chem*. 2001. Т. 73, № 13. С. 2836–2842.
133. Fenselau C., Yao X. $^{18}\text{O}_2$ -Labeling in Quantitative Proteomic Strategies: A Status Report // *J Proteome Res*. 2009. Т. 8, № 5. С. 2140–2143.
134. Klingler D., Hardt M. Protease- and Acid-catalyzed Labeling Workflows Employing ^{18}O -enriched Water // *Journal of Visualized Experiments*. 2013. № 72.

135. Yao X., Afonso C., Fenselau C. Dissection of Proteolytic ^{18}O Labeling: Endoprotease-Catalyzed ^{16}O -to- ^{18}O Exchange of Truncated Peptide Substrates // *J Proteome Res.* 2003. T. 2, № 2. С. 147–152.
136. SPRINSON D.B., RITTENBERG D. Nature of the Activation Process in Enzymatic Reactions // *Nature.* 1951. T. 167, № 4247. С. 484–484.
137. ANTONOV V.K. и др. Studies on the Mechanisms of Action of Proteolytic Enzymes Using Heavy Oxygen Exchange // *Eur J Biochem.* 1981. T. 117, № 1. С. 195–200.
138. Bender M.L., Heck H. d'A. Carbonyl Oxygen Exchange in General Base Catalyzed Ester Hydrolysis // *J Am Chem Soc.* 1967. T. 89, № 5. С. 1211–1220.
139. Niles R. и др. Acid-Catalyzed Oxygen-18 Labeling of Peptides // *Anal Chem.* 2009. T. 81, № 7. С. 2804–2809.
140. Mirgorodskaya O.A. и др. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using ^{18}O -labeled internal standards // *Rapid Communications in Mass Spectrometry.* 2000. T. 14, № 14. С. 1226–1232.
141. Stewart I.I., Thomson T., Figeys D. ^{18}O Labeling: a tool for proteomics // *Rapid Communications in Mass Spectrometry.* 2001. T. 15, № 24. С. 2456–2465.
142. Toropygin I.Y. и др. The N-domain of angiotensin-converting enzyme specifically hydrolyzes the Arg-5-His-6 bond of Alzheimer's A β -(1-16) peptide and its isoAsp-7 analogue with different efficiency as evidenced by quantitative matrix-assisted laser desorption/ionization time-o // *Rapid Communications in Mass Spectrometry.* 2008. T. 22, № 2. С. 231–239.
143. Rappsilber J., Mann M., Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips // *Nat Protoc.* 2007. T. 2, № 8. С. 1896–1906.
144. Fu Y., Qian X. Transferred subgroup false discovery rate for rare post-translational modifications detected by mass spectrometry // *Molecular and Cellular Proteomics.* 2014. T. 13, № 5. С. 1359–1368.
145. <https://github.com/MiKonstantinov/AA-Pair-counter> [Электронный ресурс].
146. Razew A. и др. One fold, many functions—M23 family of peptidoglycan hydrolases // *Front Microbiol.* 2022. T. 13.
147. Leiros H.S. и др. Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements // *Protein Science.* 2004. T. 13, № 4. С. 1056–1070.
148. Whitaker D.R. и др. Lytic enzymes of *Sorangium* sp. A comparison of the proteolytic properties of the alpha- and beta-lytic proteases. // *Can J Biochem.* 1965. T. 43, № 12. С. 1961–1970.
149. Ahmed K. и др. Purification, bacteriolytic activity, and specificity of β -lytic protease from *Lysobacter* sp. IB-9374 // *J Biosci Bioeng.* 2003. T. 95, № 1. С. 27–34.
150. Schauperl M. и др. Characterizing protease specificity: How many substrates do we need? // *PLoS One.* 2015. T. 10, № 11.
151. Markert Y. и др. Proline versus charge concept for protein stabilization against proteolytic attack // *Protein Eng.* 2003. T. 16, № 12. С. 1041–1046.

152. Vanhoof G. и др. Proline motifs in peptides and their biological processing // The FASEB Journal. 1995. Т. 9, № 9. С. 736–744.
153. Melnikov S. и др. Molecular insights into protein synthesis with proline residues // EMBO Rep. 2016. Т. 17, № 12. С. 1776–1784.
154. Sahai M.A. и др. First Principle Computational Study on the Full Conformational Space of 1 - Proline Diamides // J Phys Chem A. 2005. Т. 109, № 11. С. 2660–2679.
155. Yao X. и др. Erratum: Proteolytic ^{18}O Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus (Analytical Chemistry (2001) 73 (2386-2382)) // Anal Chem. 2004. Т. 76, № 9. С. 2675.
156. Finehout E.J., Cantor J.R., Lee K.H. Kinetic characterization of sequencing grade modified trypsin // Proteomics. 2005. Т. 5, № 9. С. 2319–2321.