

На правах рукописи

Кузнецова Ксения Глебовна

ВЛИЯНИЕ ХИМИЧЕСКОЙ ОБРАБОТКИ ОСТАТКОВ ЦИСТЕИНА В БЕЛКАХ НА
РЕЗУЛЬТАТЫ ПРОТЕОГЕНОМНОГО АНАЛИЗА

03.01.04 – биохимия

АВТОРЕФЕРАТ

диссертации на соискание ученой степени кандидата биологических наук

Москва, 2020

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича».

Научный руководитель: доктор биологических наук, профессор
Мошковский Сергей Александрович

Официальные оппоненты: **Гусев Николай Борисович**,
доктор биологических наук, член-корреспондент
РАН, профессор, заведующий кафедрой биохимии
биологического факультета МГУ им. М.В.
Ломоносова

Зиганшин Рустам Хусманович,
кандидат химических наук, старший научный
сотрудник, руководитель группы масс-спектрометрии
ЦКП «Биоорганика» ИБХ РАН

Ведущая организация: Федеральное государственное учреждение
«Федеральный исследовательский центр
«Фундаментальные основы биотехнологии»
Российской академии наук»
(Институт биохимии им. А.Н. Баха).

Защита состоится «10» декабря 2020 года в 13 часов на заседании Диссертационного совета Д 001.010.01 при Федеральном государственном бюджетном научном учреждении «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича» (ИБМХ) по адресу: 119121, г. Москва, Погодинская ул., д. 10, стр. 8.

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного научного учреждения «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича» и на сайте www.ibmc.msk.ru.

Автореферат разослан «__» _____ 20__ г.

Ученый секретарь диссертационного совета,
кандидат химических наук

Карпова Е.А.

1. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

1.1. Актуальность темы исследования и степень ее разработанности

В последние годы точность и чувствительность панорамного протеомного анализа сильно возросла. Основной причиной такого успеха служит изобретение масс-спектрометра типа Orbitrap, позволяющего за один стандартный эксперимент идентифицировать несколько тысяч белков в одном образце. Разумеется, увеличение глубины анализа поставило ребром вопрос изучения факторов, влияющих на его качество, таких, как химические модификации аминокислот в составе белков. С развитием методов протеогеномики, и вообще, с выходом на передний план так называемых мультиомик, проблема налаживания методов анализа данных встала особо остро. В современной науке ощущается тенденция перехода от обычной протеомики к комплексному исследованию генома, транскриптома и протеома вместе, и соотнесению одного другому.

Протеогеномный метод актуален, в первую очередь, для биомедицинских исследований. Исследуя белковые продукты мутантных генов, возможно обнаружить новые биомаркеры и так называемые неоантигены, специфичные участки белков, характерные для раковых клеток, которые представляют потенциальные мишени для раковых вакцин (Gubin et al. 2014). Поскольку в ходе протеогеномного анализа изучают одноаминокислотные замены, то модификации аминокислотных остатков, даже редкие, могут существенно влиять на результат. За пределами протеогеномики уже описано множество примеров химических модификаций, имитирующих важные, с биологической точки зрения, пост-трансляционные события. Так, Nielsen с соавт. (Nielsen et al. 2008) показали, что при пробоподготовке к протеомному анализу возможна артефактная модификация остатка лизина, молекулярная масса которой равна диглициновым меткам сайтов убиквитинирования. В других исследованиях было показано влияние химической обработки белков на результаты исследования сайтов фосфорилирования (Krüger et al. 2005). Иными словами, изучение влияния методов пробоподготовки на результаты протеомного анализа имеет важное значение для улучшения качества последнего. Особенно это важно в специализированных случаях, таких как исследование пост-трансляционных модификаций или протеогеномный анализ.

Исследования влияния алкилирующих агентов на результаты протеомного анализа ведутся давно, но количество таких работ невелико (Jiang et al. 2013, Müller and Winter 2017). По большому счету, детальная оценка влияния различных алкилирующих агентов на результаты протеомного масс-спектрометрического анализа проведена только в одной работе (Müller and Winter 2017). Мы дополнили эти результаты анализом двух других алкилирующих агентов и отдельно внимательно изучили преобразование метионина в изотреонин, не затронутое в указанной статье.

Целью работы было исследование влияния разных способов искусственной модификации остатков цистеина при пробоподготовке к протеомному анализу для оптимизации метода протеогеномики.

Для достижения этой цели были поставлены следующие **задачи**:

1. Оценка частоты преобразования метионина в изотреонин в протеомных данных и подбор оптимального для протеогеномики метода пробоподготовки, позволяющего избежать маскировки такой модификации под аминокислотную замену метионина на треонин.
2. Проведение сравнительного анализа четырех часто используемых в протеомике алкилирующих агентов, а именно йодацетамида (IAM), 4-винилпиридина (4-VP), метилметантиосульфата (MMTS) и хлорацетамида (CAM), и оценка влияния каждого из них на результат протеомного анализа.
3. Сравнение спектров фрагментации пептидов, содержащих треонин и изотреонин для оценки возможности распознавания этих аминокислотных остатков в пептидах при анализе протеома.

1.2. Научная новизна работы

В работе проведен анализ влияния методов искусственной обработки остатков цистеина на результаты протеомного масс-спектрометрического анализа. Помимо изученных ранее алкилирующих агентов, в анализ были включены два других, не проанализированных в сравнительных работах, но применяемых в протеомике, а именно 4-винилпиридин и метилметантиосульфат. Впервые изучено преобразование метионина в изотреонин в процессе пробоподготовки. Исследована частота данного

явления на нескольких наборах данных и выдвинуты предположения о его причинах. Проведен детальный анализ данных, в том числе методом «открытого» поиска, в ходе которого также впервые продемонстрировано наличие сдвига массы пептидов с метионином, свидетельствующее о его преобразовании в изотреонин.

1.3. Теоретическая и практическая значимость работы

Результаты работы и сделанные выводы помогут исследователям грамотно подбирать методы пробоподготовки к протеомному анализу, учитывая задачу исследования. Адекватно выбранный подход к обработке белков в ходе анализа способствует уменьшению вероятности ложных результатов и артефактов и в целом улучшает результат. Внимательный подход к обработке результатов протеомных исследований помогает обнаружить возможные артефакты и учесть их влияние на результаты анализа.

1.4. Методология и методы диссертационного исследования

В работе проведены как собственные эксперименты, так и обработка общедоступных данных других исследователей, что, в настоящий момент, является распространенной практикой в протеомных исследованиях. Анализ данных проведен по современным методикам с использованием как готового программного обеспечения, так и самостоятельно написанных скриптов на двух языках программирования. Интерпретация масс-спектрометрических данных проведена в соответствии с рекомендациями глобального проекта «Протеом человека» (Deutsch et al. 2016). Экспериментальные данные получены современными методами масс-спектрометрии, их описание приведено в соответствии с требованиями ведущих журналов по протеомике к описанию масс-спектрометрических экспериментов. Данные, полученные в экспериментальной части работы, помещены в репозиторий ProteomeXchange и доступны другим исследователям для проверки и использования, что, также, соответствует требованию большинства научных изданий в области протеомики.

1.5. Положения, выносимые на защиту

1. Анализ преобразования метионина в изотреонин в процессе протеомного эксперимента показал, что метионин преобразуется в изотреонин, в среднем, не более, чем в 1% метионинсодержащих пептидов.

2. При использовании в качестве алкилирующих агентов йодсодержащих веществ, таких как йодацетамид и йодуксусная кислота, количество пептидов с преобразованием метионина в изотреонин может достигать 4% от всех пептидов, содержащих этот аминокислотный остаток. Эта реакция происходит чаще, если образцы проходят электрофоретическое разделение в полиакриламидном геле.

3. Путем оценки влияния соседствующих аминокислотных остатков на частоту преобразования метионина в изотреонин показано, что пролин, следующий за метионином в последовательности белка, увеличивает частоту этой реакции.

4. При сравнительном анализе четырех алкилирующих агентов, а именно йодацетамида, хлорацетамида, 4-винилпиридина и метилметантиосульфоната показано, что из исследованных соединений для протеомного анализа лучше всех подходит хлорацетамид. Он вызывает наименьшее количество побочных реакций и приводит к идентификации наибольшего количества пептидов.

5. Показано, что наиболее часто используемый алкилирующий агент йодацетамид может вызывать карбамидометилирование метионина и, тем самым, препятствовать идентификации пептидов, содержащих этот аминокислотный остаток. Продемонстрировано применение метода «открытого» поиска для учета подобных нежелательных модификаций аминокислотных остатков.

6. Путем сравнения хроматограмм и масс-спектров сходных синтетических пептидов, содержащих треонин или изотреонин, показано, что содержащие треонин и изотреонин пептиды характеризуются разным временем выхода с хроматографической колонки, а также различаются по масс-спектрам фрагментации.

1.6. Личный вклад соискателя

Соискатель проводил поиск и анализ литературы, планирование экспериментов, пробоподготовку к протеомному анализу и анализ данных, интерпретацию результатов и подготовку публикаций. Работа была выполнена в лаборатории медицинской протеомики ИБМХ в период с 2015 по 2020 год.

1.7. Степень достоверности и апробация результатов

Достоверность результатов работы подтверждается согласованностью результатов современных экспериментальных и вычислительных методов протеомного анализа, использованных в работе. Обсуждение результатов проведено с учетом современных исследований, опубликованных в области белковой биохимии, протеомики и протеогеномики. Научные положения и выводы, изложенные в диссертации, обоснованы и подтверждены фактическим материалом.

Основные результаты диссертационной работы были представлены в виде стендовых докладов на конференциях таких, как конгресс Европейской Протеомной Ассоциации (EuPA 2016, Стамбул), конгресс международной организации «Протеом человека» (HUPO 2017, Дублин), конгресс Европейской Протеомной Ассоциации (EuPA 2018, Сантьяго де Компостела) и других.

1.8. Публикации по результатам исследования

По теме диссертации опубликовано 18 работ, из которых 10 статей в рецензируемых научных журналах и 8 публикаций в трудах конференций.

1.9. Объем и структура диссертации

Диссертация изложена на 114 страницах машинописного текста, содержит 6 таблиц и 17 рисунков. Состоит из следующих разделов: Список сокращений, Введение, Обзор литературы, Материалы и методы, Результаты и обсуждение, Заключение, Выводы, Список литературы, а также включает Благодарности, информацию о финансировании и Приложение.

2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Подготовка образцов к масс-спектрометрическому анализу

Клетки культур HeLa и HepG2 были получены из коллекции ИБМХ. Кроме этого были подготовлены образцы бычьего сывороточного альбумина. Клетки лизировали, и полученную смесь белков подвергали ферментативному гидролизу в растворе с использованием реактива ProteaseMAX по методу, описанному ранее (Pirmoradian et al. 2013) с незначительными изменениями. На этапе алкилирования образцы клеточных

культур обрабатывали разными алкилирующими агентами для сравнения их влияния на результаты протеомного анализа. Среди исследуемых агентов были: йодацетамид, 4-винилпиридин, хлорацетамид и метилметантиосульфат. Все образцы клеточных культур были подготовлены в трех повторах: I, II и III. Образцы альбумина готовили с использованием дезоксихолата натрия. Для сравнения масс-спектров пептидов, содержащих треонин и изотреонин были синтезированы пептиды-аналоги, сходные по последовательности, но содержащие оба изомера аминокислотного остатка. Пептиды синтезировали в количестве около 0,5 мкмоль каждого при помощи автоматизированной рабочей станции для обработки жидкостей Hamilton Microlab Star (Hamilton, Германия) с использованием лантерн SynPhas Lanterns (Mimotopes, Австралия) с тритильным линкером и присоединенными аминокислотами Fmoc-L-лизином или Fmoc-L-аргинином.

2.1. Хроматомасс-спектрометрия

Анализ всех образцов проводили хроматомасс-спектрометрическим методом с небольшими различиями. Смеси пептидов разделяли при помощи хроматографической системы UltiMate 3000 RSLC nano-system (Dionex, США). Разделение проводили при помощи высокоэффективной обращенно-фазовой жидкостной хроматографии (ВЖЭХ) в обратном градиенте концентрации ацетонитрила. Сначала пептиды концентрировали и очищали на предколонке, непосредственно соединенной с аналитической колонкой. После разделения, пептиды сразу попадали в масс-спектрометр, соединенный с хроматографической системой «онлайн». Масс-спектры всех образцов получали при помощи масс-спектрометров типа Orbitrap из серии Q Exactive (Thermo Fisher Scientific, США) с использованием метода диссоциации, вызванной столкновением электронов при высокой энергии (higher-energy collisional dissociation – HCD), режима положительно заряженных ионов и зависимо от данных режима (data dependent acquisition – DDA). Все масс-спектры получали в трех технических повторах.

2.2. Анализ данных

В работе помимо собственных экспериментальных данных анализировали данные из трех работ (Moghaddas Gholami et al. 2013, Sharma et al. 2015, Müller and Winter 2017), загруженные из общедоступных источников.

Протеогеомный поиск проводили при помощи поисковой программы X!Tandem (Craig and Beavis 2004) с последующей обработкой результатов при помощи инструмента MPSScore (Ivanov et al. 2014). Для протеогеомного поиска была создана база данных, содержащая варианты белки с одноаминокислотными заменами, характерными для колоректального рака. Количественный анализ пептида с преобразованием метионина в изотреонин проводили при помощи программы MaxQuant. Основной протеомный поиск проводили при помощи программы IdentiPy против баз данных соответствующих организмов из Uniprot. Постобработку также делали при помощи MPSScore. Фиксированные модификации цистеина были следующими: для контрольных образцов: никакие; для образцов, обработанных йодацетамидом и хлорацетамидом – карбамидометилирование (+57,021 Да); для образцов, обработанных 4-винилпиридином – пиридилэтилирование (+105,058 Да), для образцов, обработанных метилметантиосульфатом – сульфенилирование (+45,988 Да).

После протеомного поиска сравнивали общее число идентифицированных пептидов и спектральных идентификаций (PSM), а также абсолютное и относительное число пептидов и PSM с Цис и Мет между всеми образцами и тремя повторами эксперимента. Чтобы получить более воспроизводимые результаты между повторами, мы нормализовали значения в каждом повторе эксперимента (повторы I, II и III) по наивысшему значению среди всех образцов (т. е. методов алкилирования). После этого была проведена оценка критерия Стьюдента попарно среди всех комбинаций образцов. Помимо значений p критерия Стьюдента, вычисляли кратное изменение всех нормализованных значений в парах сравнения. Статистический анализ проводился с использованием скриптов, написанных на R.

Для анализа частоты преобразования метионина в изотреонин в панорамных протеомных данных, во время поиска с использованием IdentiPy, помимо стандартных параметров, в качестве вариабельной модификации была добавлена замена метионина на треонин, по сдвигу массы совпадающая с преобразованием метионина в изотреонин (-29.9928 Да). Затем пептиды с идентифицированной модификацией Мет->Тре были отфильтрованы до группно-специфичного уровня ложноположительных результатов равного 1% согласно (Levitsky et al. 2017) с использованием скриптов, написанных на языке Python на базе библиотеки Pyteomics (Goloborodko et al. 2013). После этого,

сравнивали долю пептидов с Мет, несущих модификацию, между образцами, используя статистический критерий χ^2 . Соответствующие скрипты для статистического анализа были написаны на языке R.

Для того, чтобы убедиться, что наиболее частые сдвиги масс соответствуют ожидаемым модификациям и проверить, не больше ли их, чем ожидалось, мы выполнили открытый поиск (Chick et al. 2015) с последующим анализом результатов с помощью инструмента AA_stat (Bubis et al. 2018). Главной особенностью открытого поиска является то, что при нем используется допустимая ошибка масс прекурсоров ± 500 Да, что гораздо больше, чем при обычном протеомном поиске. Далее из масс пептидов-кандидатов, совпавших с пептидами из базы данных, выбирают достоверные совпадения на основе спектров фрагментации.

Синтетические пептиды, содержащие Тре и изоТре анализировали при помощи MaxQuant. Далее проводили автоматическую классификацию пептидов при помощи машинного обучения. Скрипты для такой классификации были написаны на языке R.

3. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

3.1. Обнаружение преобразования метионина в изотреонин

В процессе протеомного анализа данных девяти клеточных линий из (Moghaddas Gholami et al. 2013) мы обнаружили замену метионина на треонин в одной и той же позиции в белке теплового шока HSC70. Важно отметить, что используемая база данных представляла собой протеомную базу, содержащую участки полиморфизма из злокачественных опухолей толстой и прямой кишки, однако, она не имела точного совпадения с транслированными экзонами какой-либо из исследованных раковых клеточных линий, даже COLO205, имевшей сходное происхождение. Вскоре после того как мы обнаружили данную замену в протеомах клеточных линий NCI60, в открытом доступе появились данные экзотов этих же линий (Abaan et al. 2013). Мы воспользовались этими данными для уточнения присутствия интересующей нас замены в геномах конкретных линий в соответствующей позиции. Результат такой проверки оказался отрицательным, нам не удалось обнаружить замены кодона метионина на кодон треонина в этих данных. Разбираясь в возможных вариантах происхождения обнаруженной модификации метионина на этапе пробоподготовки, мы пришли к выводу,

что преобразование метионина могло произойти не в треонин, а в изотреонин, то есть гомосерин. Известно, что йодуксусная кислота, как и йодацетамид способны алкилировать метионин с образованием карбоксиметилметионина и карбамидометилметионина соответственно (Lundblad 1994). Оба соединения под действием высокой температуры (около 100°C) в нейтральной среде способны преобразовываться в гомосерин и в лактон гомосерина (Gundlach et al. 1959). Поскольку при проведении гель-электрофореза перед трипсинолизом образцы часто нагревают, а именно такой метод был применен в работе (Moghaddas Gholami et al. 2013), мы предположили, что преобразование метионина именно в изотреонин в анализируемых нами данных вполне возможно.

Количественный анализ без использования меток (label free quantification) показал, что относительное содержание модифицированных вариантов пептидов меньше, чем не модифицированных. Средняя интенсивность (LFQ intensity) пика, рассчитанная на основании интенсивности площадей хроматографических пиков в MS¹ (Cox et al. 2014), для модифицированного пептида составляет примерно 2,3% от интенсивности для не модифицированного пептида (стандартное отклонение 1,3%, доверительные интервалы 1,1-4,3%). Анализ образцов бычьего сывороточного альбумина показал, что при проведении алкилирования при повышенных температурах, метионин преобразуется в изотреонин чаще.

3.2. Анализ масс-спектров синтетических пептидов, содержащих треонин и изотреонин

Для того, чтобы оценить возможность на уровне протеома различать метионин, преобразованный в изотреонин, от замены метионина на треонин мы провели анализ синтетических пептидов. Пептидные последовательности выбирали из человеческого протеома и дополнили их модельными последовательностями. Человеческие пептиды включали от 1 до 5 остатков треонина. Модельные последовательности составляли таким образом, чтобы треонин был окружен с каждой стороны двумя одинаковыми аминокислотными остатками и эти остатки во всех пептидах различались. К каждому пептиду с треонином синтезировали его аналог с изотреонином.

На первом этапе нам удалось разделить смесь, содержащую аналогичные пептиды с Тре и изоТре при помощи жидкостной хроматографии. Средняя разница во времени выхода пептидов с Тре и изоТре составила 1,26 минуты, что в данных условиях градиента соответствует разнице концентраций ацетонитрила (ACN) в подвижной фазе, равной 7,37%. Кроме того, мы заметили корреляцию между количеством остатков треонина в пептиде и разницей во времени выхода такого пептида с его аналогом, содержащим изоТре. Иными словами, чем больше остатков треонина или изотреонина в пептиде, тем лучше такие пептиды разделяются при помощи хроматографии.

Главным интересом при анализе синтетических пептидов было сравнить масс-спектры фрагментации пептидов с Тре и изоТре, полученные в условиях типичного протеомного эксперимента. При сравнении мы обращали наибольшее внимание на *b*- и *y*-ионы, аннотированные при помощи программы MaxQuant. Визуально спектры аналогичных пептидов отличались (рис. 1). Данные относительной интенсивности пиков были получены из результатов MaxQuant и использованы для автоматической классификации спектров пептидов.

Для описания свойств спектра были выбраны шесть признаков, представляющих собой относительные интенсивности наиболее часто представленных ионов. Эти шесть ионов имели ненулевую интенсивность в более 80% пептидов. Среди ионов, выбранных в качестве признаков были: b_n , b_{n-2} , y_n , y_{n-1} , y_{n-2} и y_{n+1} , где n – это положение Тре или изоТре в пептиде. По результатам автоматической классификации стало понятно, что по признаку b_n классы различаются достоверно, в то время как по остальным пяти признакам – не достоверно. Этот признак не позволяет определять наличие изотреонина в панорамных протеомных данных без использования контроля. Однако это возможно с применением синтетических стандартов.

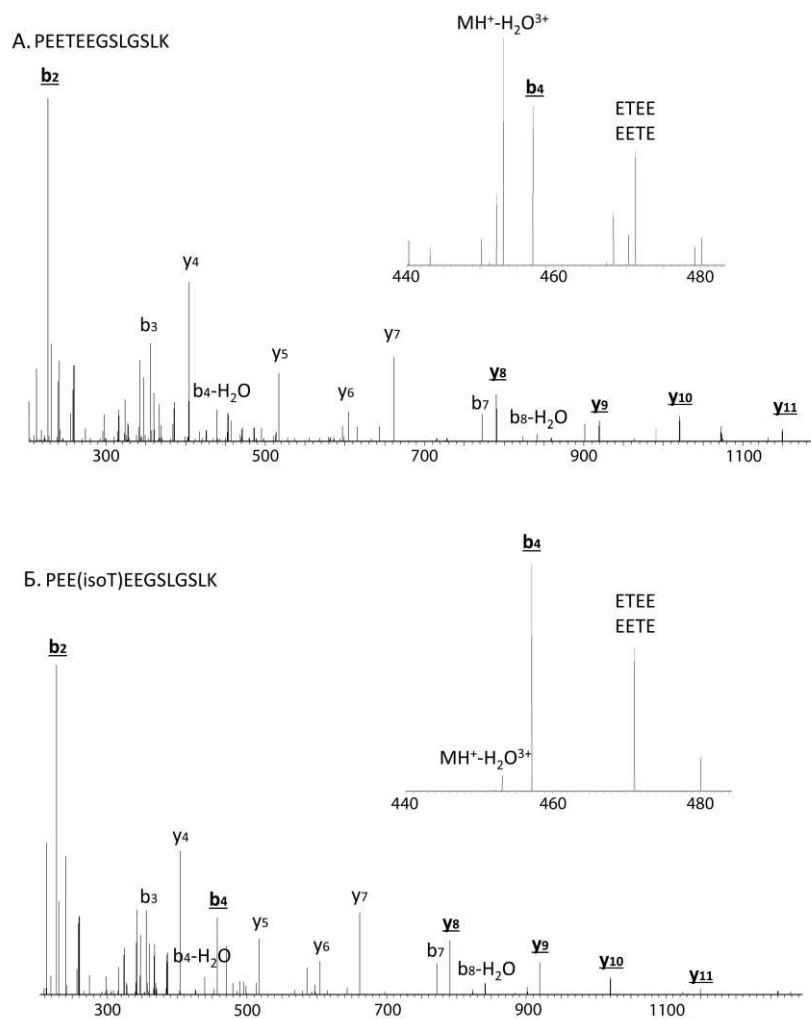


Рис. 1. Масс-спектры модельной пары пептидов PEETEEGSLGSLK (a) и PEE (isoT) EEGSLGSLK (b), показывающие разную относительную интенсивность характеристических ионов: b₄ (b_n), b₂ (b_{n-1}), y₁₀ (y_n), y₉ (y_{n-1}), y₁₁ (y_{n+1}), y₈ (y_{n-2}).

3.3. Сравнение методов алкилирования остатков цистеина при пробоподготовке к протеомному анализу

Одной из задач нашей работы было проведение сравнительного анализа четырех часто используемых в протеомике алкилирующих агентов, а именно йодацетамида (IAM), 4-винилпиридина (4-VP), метилметантиосульфоната (MMTS) и хлорацетамида (CAM) и оценка влияния каждого из них на результат протеомного анализа. Всю процедуру пробоподготовки проводили в 3 повторах (повторы I, II и III), в результате чего получилось три группы по 8 образцов. В каждом повторе один образец был контрольным, то есть не подвергался ни восстановлению ни алкилированию. Остальные образцы

восстанавливали при помощи DTT и алкилировали названными реагентами. Три образца, проалкилированные IAM, 4-VP и SAM были подвергнуты повторной обработке DTT для проверки эффекта гашения избытка алкилирующего агента. Обозначения этим образцам были присвоены: IAMq, 4-VPq, SAMq соответственно.

По общему числу идентифицированных пептидов контрольный образец не отличался от остальных образцов (рис. 2А). Такой эффект, по-видимому, связан с тем, что в образцах, подверженных восстановлению и алкилированию, накапливаются побочные реакции, затрудняющие в результате идентификацию части пептидов. Ожидаемо, в контрольном образце практически отсутствуют идентифицированные пептиды с цистеином (рис. 2Б). Наименьшее количество идентифицированных пептидов наблюдается для образца, алкилированного йодацетимидом (IAM) (рис. 2А), оно в 1,4 раза ниже по сравнению с контрольным образцом ($p < 0,001$). Частично это вызвано тем, что в этом образце очень мало пептидов с Мет (рис. 2В): почти в четыре раза меньше по сравнению с контролем ($p < 0,001$). В отличие от образца, алкилированного с использованием IAM, образец, в котором избыток IAM был погашен добавлением DTT после алкилирования (образец «IAMq»), демонстрирует другой результат. Общее количество идентифицированных пептидов для этого образца показывает небольшое (в 1,33 раза) увеличение по сравнению с образцом IAM ($p < 0,001$). Относительное количество пептидов с Цис для этого образца в 1,3 раза меньше по сравнению с образцом IAM ($p < 0,001$), однако относительное количество пептидов с Мет является высоким и примерно таким же, как в контроле (рис. 2В). Общее количество идентифицированных пептидов для образца, алкилированного 4-винилпиридином (4-VP), в среднем, немного ниже по сравнению с контролем (в 1,1 раза). Тем не менее, отклонение этих значений среди повторов для этого образца очень велико (рис. 2А), особенно по параметру относительного количества пептидов с Цис (рис. 2Б). Гашение избытка 4-VP после реакции (4-VPq) улучшает воспроизводимость с уменьшением общего количества идентифицированных пептидов в 1,32 раза по сравнению с образцом без гашения 4-VP ($p < 0,001$). Тем не менее, использование 4-VP защищает остатки метионина и приводит к высокому относительному количеству метионинсодержащих пептидов (рис. 2В).

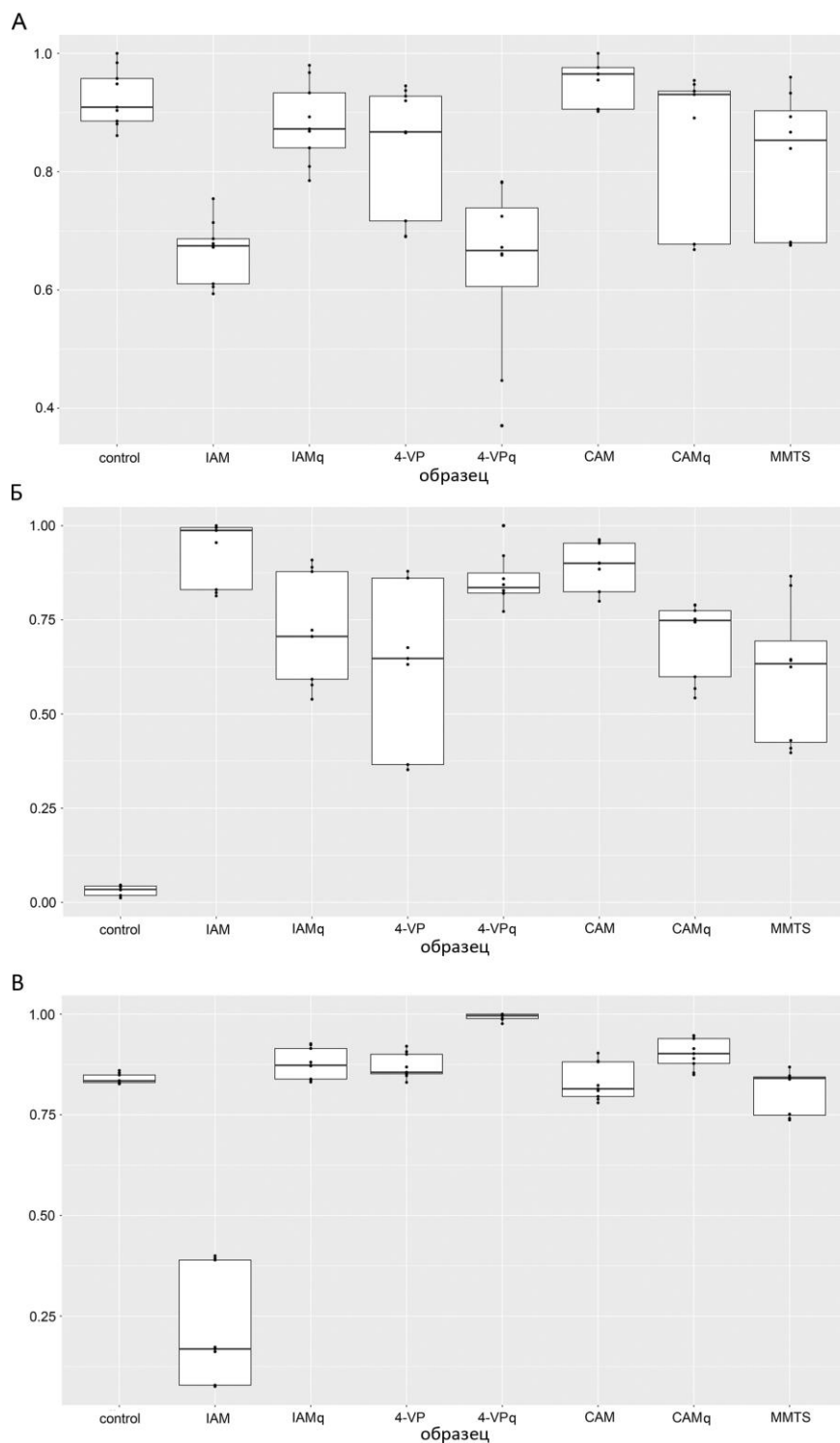


Рис. 2. Диаграммы, показывающие медианы и межквартильные интервалы нормализованных значений числа пептидов, идентифицированных после разных методов алкилирования. q - образцы, в которых избыток алкилирующего агента был погашен добавлением DTT после алкилирования. Значения были нормализованы по наибольшему внутри 3 повторов пробоподготовки, что дало 9 значений, полученных после объединения технических повторов и повторов образца пробоподготовки. А: общее число идентифицированных пептидов; Б: отношение количества идентифицированных пептидов с Цис к количеству всех идентифицированных пептидов; В: отношение количества идентифицированных пептидов с Мет к количеству всех идентифицированных пептидов.

Алкилирование хлорацетамидом (СAМ) дало наибольшее количество идентифицированных пептидов, и этот результат был в высокой степени воспроизводимым (рис. 2А). Соотношения цистеин- и метионинсодержащих пептидов относительно всех идентифицированных пептидов также сравнительно высоки (рис. 2Б, В). Гашение этого реагента с помощью DTT после обработки ухудшает результат с точки зрения общего количества идентифицированных пептидов и, особенно, воспроизводимости. Общее количество пептидов в образце, алкилированном метилметантиосульфатом (ММТS), в 1,23 раза выше ($p = 0,04$) по сравнению с образцом, алкилированным с использованием IАМ (который дал самое низкое количество пептидов, не считая экспериментов с гашением избытка). С другой стороны, оно ниже, чем количество пептидов, идентифицированных для образца, алкилированного с использованием САМ, в 1,17 раза ($p = 0,002$). Относительное количество пептидов с Цис, идентифицированных для образца, алкилированного с использованием ММТS, также в 1,5 раза ниже по сравнению с образцом, алкилированным с использованием САМ ($p < 0,001$). Наблюдаемая картина для спектральных идентификаций (PSM) сходна с таковой для идентифицированных пептидов.

3.4. Анализ частоты преобразования метионина в изотреонин в протеомных данных

Для общей оценки частоты преобразования метионина в изотреонин в протеомных данных мы использовали общедоступные данные из трех работ (Moghaddas Gholami et al. 2013, Sharma et al. 2015, Müller and Winter 2017), а также наши собственные данные, полученные с применением разных алкилирующих агентов. Стоит обратить внимание, что в работе Müller и Winter образцы получали как методом пробоподготовки в геле, так и с использованием центрифужных фильтров (FASP), что в данном случае рассматривается авторами как пробоподготовка в растворе. Данные протеомов клеточных линий панели NCI60 (Moghaddas Gholami et al. 2013) все были получены одинаковым методом, разделением белков электрофорезом с последующим алкилированием при помощи IАМ и трипсинолизом в геле. Кроме этого, мы проанализировали данные протеомов нескольких отделов мозга мыши из работы (Sharma et al. 2015). В данной

работе все образцы были приготовлены в растворе с использованием IAM. Для нашей работы мы использовали данные девяти отделов мозга.

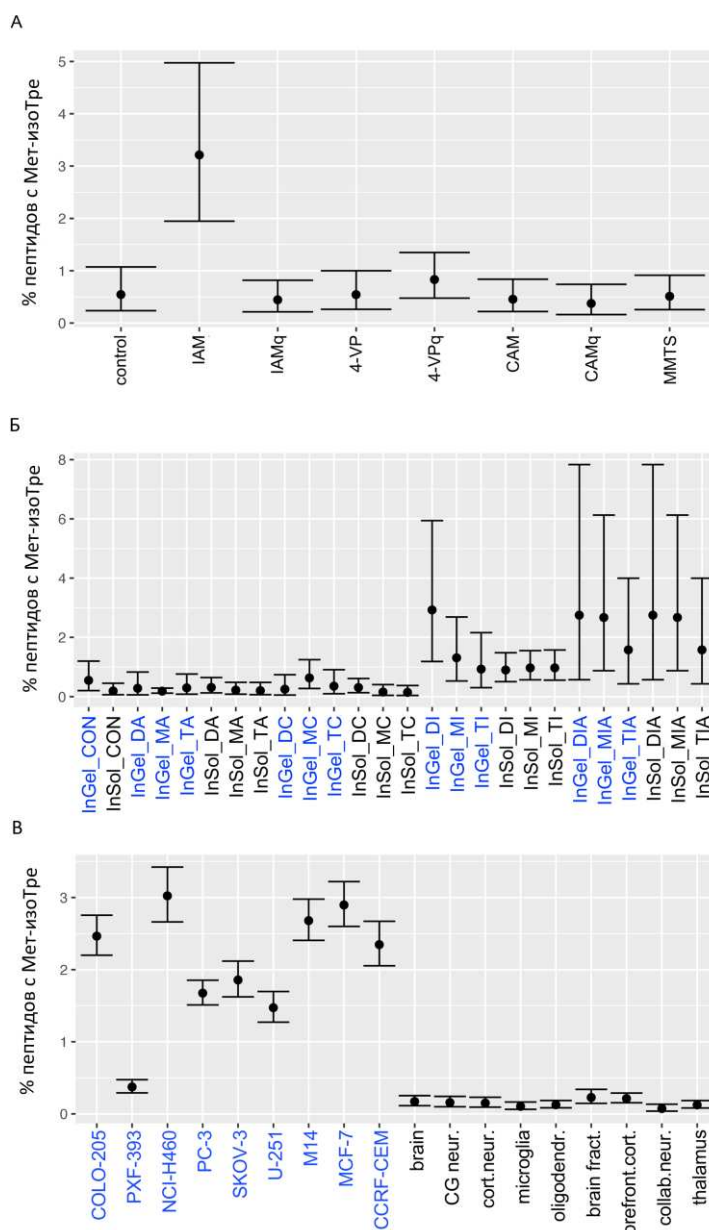


Рис. 3. Отношение числа пептидов с преобразованием Met-изоТре к числу всех идентифицированных пептидов, содержащих метионин. Показаны 95% доверительные интервалы критерия χ^2 . А: Данные из нашей работы; Б: Данные из работы Müller и Winter (Müller and Winter 2017), образцы, отмеченные синим цветом, были приготовлены в геле; В: Данные из работ по клеточным линиям NCI60 (Moghaddas Gholami et al. 2013) (отмечены синим, все образцы приготовлены в геле) и Sharma с соавт. (Sharma et al. 2015).

В данных, полученных в нашей работе, наибольшая доля пептидов, содержащих преобразование метионина в изотреонин наблюдается в образце, алкилированном IAM (рис. 3А). В этом образце содержание соответствующего сдвига массы составляет около

3%. Доля пептидов, несущих модификацию Мет-изоТре в образцах, алкилированных другими агентами, практически одинакова (рис. 3А) и не превышает 1%.

Похожий результат мы получили по данным из работы Müller и Winter. Наибольшая доля преобразований метионина в изотреонин наблюдается в образцах, алкилированных йодацетамидом (IAM) и йодоуксусной кислотой (IAA) (рис. 3Б, все образцы, обозначение которых оканчивается на «I» и «IA») по сравнению с образцами, алкилированными другими агентами. Процент пептидов с превращением Мет в изоТре в образцах IAM и IAA также составлял от 1 до 3%.

В данных по NCI60, для всех клеточных линий, за исключением PXF-393, процент пептидов, содержащих преобразования метионина в изотреонин составил 1,5-3% (рис. 3В). Данные из этой работы представляют собой более «глубокие» протеомы, позволяющие получить большее количество идентифицированных пептидов и белков, чем предыдущие два рассмотренных набора данных. Кроме того, абсолютное количество пептидов с метионином, как и количество пептидов с модификацией Мет-изоТре, в этих данных выше, что делает результат более статистически достоверным. Подобно данным клеточных линий NCI60, данные по мышинному мозгу обеспечивают глубокое покрытие протеома с большим количеством идентификаций пептидов и белков. Тем не менее, доля пептидов с преобразованием Мет в изоТре в них составляет всего 0,1-0,2% (рис. 3В). Возможно, это связано с тем, что образцы в этой работе готовили в растворе.

Воспользовавшись тем фактом, что все образцы NCI60 были подготовлены одинаковым образом, и число преобразований, найденных для этих данных, было самым высоким, мы проанализировали возможное влияние соседствующих аминокислотных остатков на частоту преобразования метионина. Среди аминокислотных остатков, предшествующих метионину не было такого, который бы достоверно влиял на частоту преобразования. Однако остаток пролина, следующий после метионина, по-видимому, увеличивает частоту модификации Мет в изоТре.

3.5. «Открытый» поиск и статистический анализ сдвигов молекулярной массы

Для того, чтобы проверить данные на наличие других модификаций аминокислотных остатков, в том числе метионина, мы провели открытый поиск (Kong et al. 2017) с последующим анализом его результатов с использованием инструмента

AA_stat (Bubis et al. 2018). В результате такого анализа были выявлены наиболее представленные сдвиги молекулярных масс пептидов и предсказана их принадлежность к определенному аминокислотному остатку.

Три сдвига масс было обнаружено во всех анализируемых наборах данных. В наших данных сдвиг массы на $-128,1$ Да присутствует в некоторых образцах, затрагивает 1-2,5% и не зависит от метода обработки остатков цистеина. В данных NCI60 (Moghaddas Gholami et al. 2013) он присутствует во всех клеточных линиях и затрагивает около 0,3% пептидов. По данным (Sharma et al. 2015) он также присутствует во всех образцах и затрагивает 0,5-0,7% пептидов. Согласно Unimod (www.unimod.org) (Creasy and Cottrell 2004), этот сдвиг массы наиболее вероятно соответствует потере С-концевого лизина. По результатам анализа при помощи AA_stat, он не обогащен пептидами, содержащими какую-либо конкретную аминокислоту, и, таким образом, модификация, вероятно, действительно происходит на конце пептида. Второй сдвиг массы, присутствующий во всех трех наборах данных, составлял $+42,01$ Да и также не был приписан AA_stat к какому-либо конкретному остатку (рис. 4А). Он затронул 1–1,5% пептидов в наших данных, от 1,5 до 3,5% пептидов в данных NCI60 (Moghaddas Gholami et al. 2013) и от 0,2 до 0,3% пептидов в данных Sharma и соавт. (Sharma et al. 2015). Этот сдвиг массы наиболее вероятно соответствует ацетилированию N-конца белка, что является обычной посттрансляционной модификацией (Ree et al. 2018).

Наиболее представленный сдвиг молекулярной массы, который также присутствовал во всех трех наборах данных, составил $+57,02$ Да с явным обогащением пептидами, содержащими метионин (рис. 4Б). Стоит обратить внимание, что карбамидометилирование остатков цистеина ($+57,02$ Да) использовали во время поиска в качестве фиксированной модификации. В наших данных этот сдвиг массы наблюдается только для образцов, алкилированных йодацетамидом (IAM), и отсутствует в тех образцах, где избыток IAM был погашен добавлением восстановителя (IAMq). В подгруппе обработки IAM эта модификация затрагивает от 6 до 13% пептидов. В данных NCI60 эта доля варьирует от 0,2 до 3%. В данных по отделах мышинового мозга этот сдвиг массы был обнаружен для 2 из 3 образцов, проанализированных при помощи открытого поиска. В одном из них доля пептидов, содержащих этот сдвиг массы, составляла 0,24%, а в другом она достигала 11,4%. При этом все эти образцы прошли одну и ту же

процедуру алкилирования и гидролиза. Такие различия, по-видимому, представляют собой погрешности пробоподготовки. В образце с повышенным карбамидометилированием метионина, вероятно, оказалась более высокая концентрация IAM.

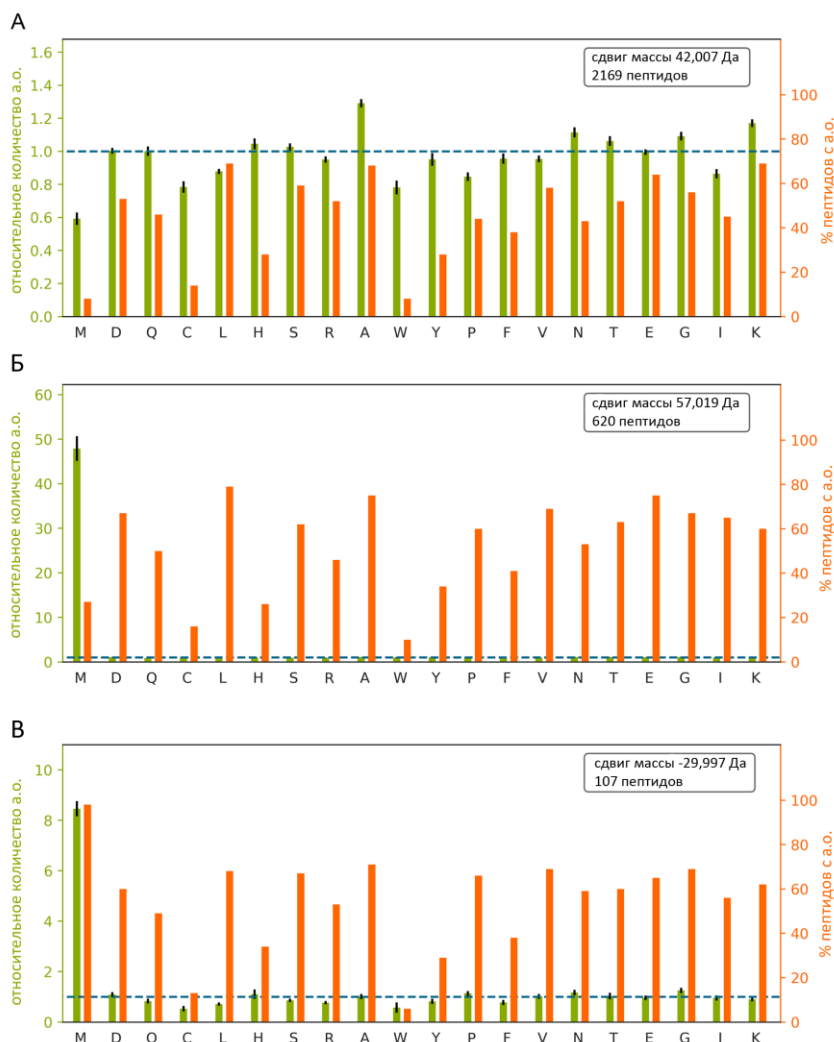


Рис. 4. Примеры результатов AA_stat, показывающих распределение аминокислот в трех бинах сдвигов масс из данных (Moghaddas Gholami et al. 2013). А: Сдвига массы на 42,007 Да соответствует ацетилированию на N-конце пептида; Б: Сдвиг массы на 57,019 Да соответствует карбамидометилированию остатков метионина; В: Сдвиг массы на -29,997 Да соответствует преобразованию метионина в изотреонин.

Мы также обнаружили менее представленный, но интересный для нашей работы, сдвиг массы на -30 Да в 3 образцах из данных по NCI60. Этот сдвиг массы также был обогащен пептидами с метионином (рис. 4В). Доля пептидов, несущих этот сдвиг массы, составляла от 0,15 до 0,17%. Вероятнее всего он соответствует замене метионина на изотреонин. Обнаружение сдвига массы на -30 Да при помощи открытого поиска и

ассоциация этого сдвига к остаткам метионина при помощи AA_stat подтверждают наличие модификации Met-изоТре, обнаруженной нами при обычном протеомном поиске. По результатам открытого поиска мы обнаружили и другие любопытные модификации аминокислотных остатков.

После обнаружения в наших данных широко представленного сдвига массы на +57 Да, насыщенного пептидами, содержащими метионин, в образце, алкилированном с помощью IAM, мы повторили обычный протеомный поиск при помощи IdentiPy с карбамидометилированием метионина в качестве вариabельной модификации. Кроме карбамидометилирования мы включили в список вариabельных модификаций окисление метионина. В образцах, алкилированных с использованием IAM, в зависимости от повтора, доля пептидов, содержащих карбамидометилированный Met, находилась в диапазоне от 40 до 80% по отношению ко всем идентифицированным пептидам, содержащим этот остаток. Доля пептидов с окисленным метионином составила всего лишь от 1 до 5%. В образцах, алкилированных с использованием SAM, напротив, было менее 1% пептидов, содержащих карбамидометилированный Met по отношению ко всем Met-содержащим пептидам. Доля пептидов с окисленным Met в образцах, алкилированных SAM, составляла от 1 до 6%, аналогично образцам, алкилированным IAM.

4. ВЫВОДЫ

1. Методами масс-спектрометрического протеомного анализа показано, что преобразование метионина в изотреонин может затронуть до 3% пептидов, содержащих метионин. Этот эффект имеет большое значение для протеогеномики, поскольку такая модификация может имитировать замену метионина на треонин, закодированную в геноме. Преобразование метионина в изотреонин происходит чаще, если образец подвергается алкилированию йодсодержащими реагентами, то есть йодуксусной кислотой или йодацетамидом, а подготовка образца осуществляется в полиакриламидном геле. Анализ влияния соседствующих аминокислотных остатков на частоту преобразования метионина в изотреонин показал, что пролин, следующий за метионином в последовательности белка, увеличивает частоту модификации.

2. Из четырех проанализированных алкилирующих агентов, а именно, йодацетамида, хлорацетамида, 4-винилпиридина и метилметантиосульфата, хлорацетамид лучше других подходит для подготовки образцов к протеомному анализу, поскольку вызывает меньше побочных реакций и позволяет идентифицировать больше пептидов. Наиболее часто используемый для этого йодацетамид способен вызывать карбамидометилирование метионина, что приводит к уменьшению числа идентифицированных пептидов, содержащих этот остаток.

3. Результаты исследования показали, что пары пептидов, содержащих треонин и изотреонин, могут быть разделены с помощью обращенно-фазовой высокоэффективной жидкостной хроматографии в условиях, обычно используемых при панорамном протеомном анализе. Пептиды, содержащие треонин и изотреонин различаются по времени удержания на хроматографической колонке. Другим обнаруженным различием являются разные спектры фрагментации пептидов, содержащих треонин и изотреонин, в которых присутствие остатка изотреонина приводит к более высокому пику b_n -иона, где n – положение треонина в последовательности. В частности, до 84% всех пептидов, содержащих треонин и изотреонин в своей последовательности, были правильно классифицированы методом машинного обучения с использованием относительной интенсивности характеризующего пика.

5. СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых журналах

1. Polyakova A.A., Kuznetsova K.G., Moshkovskii S.A. Proteogenomics meets cancer immunology: mass spectrometric discovery and analysis of neoantigens // Expert Review of Proteomics. 2015. Vol. 12, № 5. P. 533–541.
2. Chernobrovkin A.L., Kopylov A.T., Zgoda V.G., Moysa A.A., Pyatnitskiy M.A., Kuznetsova K.G., Ilina I.Y., Karpova M.A, Karpov D.S., Veselovsky A.V, Ivanov M.V, Gorshkov M.V, Archakov A.I., Moshkovskii S.A. Methionine to isothreonine conversion as a source of false discovery identifications of genetically encoded variants in proteogenomics // Journal of Proteomics. 2015. Vol. 120. P. 169–178.

3. Ключникова А.А., Кузнецова К.Г. Мошковский С.А. Редактирование матричной РНК аденозиндезаминазами ADAR: анализ на протеомном уровне // Биомедицинская химия. 2016. Т. 62, № 5. С. 510–519.
- 3a. Kliuchnikova A.A., Kuznetsova K.G., Moshkovskii S.A. 2016. ADAR-mediated messenger RNA editing: analysis at the proteome level // Biochemistry (Moscow) Supplement Series B Biomedical Chemistry Vol. 11, № 1. P. 32-42.
4. Lobas A.A., Karpov D.S., Kopylov A.T., Solovyeva E.M., Ivanov M.V, Ilina I.I., Lazarev V.N., Kuznetsova K.G., Ilgisonis E. V, Zgoda V.G., Gorshkov M.V, Moshkovskii S.A. Exome-based proteogenomics of HEK-293 human cell line: Coding genomic variants identified at the level of shotgun proteome // Proteomics. 2016. Vol. 16, № 14. P. 1980–91.
5. Kuznetsova K.G., Trufanov P. V, Moysa A.A., Pyatnitskiy M.A., Zgoda V.G., Gorshkov M.V, Moshkovskii S.A. Threonine versus isothreonine in synthetic peptides analyzed by high-resolution liquid chromatography/tandem mass spectrometry // Rapid Communications in Mass Spectrometry. 2016. Vol. 13, № 23. P. 1323–1331.
6. Kuznetsova K.G., Kliuchnikova A.A., Ilina I.U., Chernobrovkin A.L., Novikova S.E., Farafonova T.E., Karpov D.S., Ivanov M.V, Goncharov A.O., Ilgisonis E.V, Voronko O.E., Nasaev S.S., Zgoda V.G., Zubarev R.A., Gorshkov M.V, Moshkovskii S.A. Proteogenomics of Adenosine-to-Inosine RNA Editing in the Fruit Fly // Journal of Proteome Research. 2018. Vol. 17, №11. P. 3889–3903.
7. Lobas A.A., Pyatnitskiy M.A., Chernobrovkin A.L., Ilina I.Y., Karpov D.S., Solovyeva E.M., Kuznetsova K.G., Ivanov M. V., Lyssuk E.Y., Kliuchnikova A.A., Voronko O.E., Larin S.S., Zubarev R.A., Gorshkov M. V., Moshkovskii S.A. Proteogenomics of Malignant Melanoma Cell Lines: The Effect of Stringency of Exome Data Filtering on Variant Peptide Identification in Shotgun Proteomics // Journal of Proteome Research. 2018. Vol. 17, №5. P. 1801–1811.
8. Мошковский С.А., Иванов М.В., Кузнецова К.Г. Горшков.М.В. Идентификация единичных аминокислотных замен в протеогеномике // Биохимия. 2018. Т. 83, №3. С. 368–378.
- 8a. Moshkovskii S.A., Ivanov M.V., Kuznetsova K.G., Gorshkov M.V. Identification of Single Amino Acid Substitutions in Proteogenomics // Biochemistry (Moscow). 2018. Vol. 83, № 3. P. 250–258.

9. Кузнецова К.Г., Иванов М.В., Пятницкий М.А., Левицкий Л.И., Ильина И.Ю., Чернобровкин А.Л., Зубарев Р.А., Горшков М.В. Мошковский С.А. Протеом головного мозга *Drosophila melanogaster* обогащен ядерными белками // Биохимия. 2019. Т. 84, № 2. С. 252–260.
- 9а. Kuznetsova K.G., Ivanov M.V., Pyatnitskiy M.A., Levitsky L.I., Ilina I.Y., Chernobrovkin A.L., Zubarev R.A., Gorshkov M.V., Moshkovskii S.A., Brain Proteome of *Drosophila melanogaster* Is Enriched with Nuclear Proteins // Biochemistry (Moscow). 2019. Vol. 84. № 2, P. 71–78.
10. Кузнецова К.Г., Соловьева Е.М., Кузиков А.В., Горшков М.В. Мошковский С.А. Модификация остатков цистеина для масс-спектрометрического протеомного анализа: факты и артефакты // Биомедицинская химия. 2020. Т. 66, №1. С. 18-29.
- 10а. Kuznetsova, K.G. Solovyeva, E.M. Kuzikov, A.V. Gorshkov, M.V. Moshkovskii, S.A. Modification of Cysteine Residues for Mass Spectrometry-Based Proteomic Analysis: Facts and Artifacts // Biochemistry (Moscow), Supplement Series B: Biomedical Chemistry. 2020. Vol. 14. С. 204–215.

Материалы трудов конференций

11. Kuznetsova K., Trufanov P., Chernobrovkin A., Moysa A., Ivanov M., Pyatnitskiy M., Zgoda V., Gorshkov M., Moshkovskii S. Methionine to isothreonine conversion as a source of false discovery identifications of genetically encoded variants in proteogenomics // EuPA X Annual Congress, Istanbul, Turkey, 2016. P. 116.
12. Kuznetsova K., Solovieva E., Ivanov M., Karpov D., Zgoda V., Gorshkov M., Mosckovskii S. Investigation of the effects of alkylation during sample preparation on proteomic data // EuBIC Winter School on Proteomics Bioinformatics. Semmering. Austria. 2017. P. 15.
13. Kuznetsova K., Ilina I., Chernobrovkin A., Novikova S., Farafonova T., Karpov D., Ivanov M., Voronko O., Zgoda V., Zubarev R., Gorshkov M., Mosckovskii S. Proteogenomics of adenosine-to-inosine RNA editing in fruit fly // 65th ASMS Conference on Mass Spectrometry and Allied Topics. Indianapolis. USA. 2017. P. 109.
14. Kuznetsova K., Solovieva E., Gorshkov M., Moshkovskii S. Methionine-to-isothreonine conversion during sample preparation and its effects on proteomic data // The HUPO XVI World Congress. Dublin. Ireland. 2017. P. 139.

15. Kuznetsova K., Kliuchnikova A., Ilina I., Chernobrovkin A., Novikova S., Farafonova T., Karpov D., Ivanov M., Goncharov A., Voronko O., Nasaev S., Ilgisonis E., Zubarev R., Gorshkov V., Moshkovskii S. Proteogenomics of adenosine-to-inosine RNA editing in the fruit fly // XII EuPA Congress. Santiago de Compostela. Spain. 2018. P.32.
16. Кузнецова К., Левицкий Л., Пятницкий М., Мошковский С. Преобразование метионина в изотреонин как источник ложноположительного обнаружения одноаминокислотных замен при протеономном анализе // V Международная конференция “Постгеном’2018”. Казань. Россия. 2018. С. 76.
17. Kuznetsova K., Kliuchnikova A., Karpov D., Ivanov M., Mosckovskii S. Proteogenomic method applied to RNA editing investigation // EuBIC Winter School. Zakopane. Poland. 2019. P. 18.
18. Кузнецова К., Бубис Ю., Ильина И., Горшков М., Мошковский С., Открытый поиск и подбор алкилирующих агентов для оптимизации метода панорамного протеомного анализа // Девятый съезд ВМСО. Москва. Россия. 2019. С. 150.

6. БЛАГОДАРНОСТИ

Автор благодарит руководителя и коллег за помощь в проведении работы и подготовки диссертации. В частности, автор благодарен д.б.н. Згоде В.Г. за проведение масс-спектрометрического анализа, к.б.н. Мойсе А.А. за проведение синтеза пептидов, к.б.н. Пятницкому М.А., Труфанову П.А., к.б.н. Чернобровкину А.Л., Левицкому Л.И. и Бубис Ю.А. за консультацию и помощь при обработке данных. Кроме того, автор благодарит к.ф.-м.н. Горшкова М.В., Ильину И.Ю., Карпова Д.С., Карпову М.А., Иванова М.В., Соловьеву Е.М., Кузикова А.В. и других коллег, участвовавших в проекте или помогавших его проведению.