На правах рукописи

ЛОБАС Анна Александровна

Комплементарные физико-химические свойства пептидов и их использование в протеомных исследованиях

01.04.17 – химическая физика, горение и взрыв,
физика экстремальных состояний вещества

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва – 2019

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте энергетических проблем химической физики им. В.Л. Тальрозе Российской академии наук

Научный руководитель: **Горшков Михаил Владимирович**
кандидат физико-математических наук
кандидат технических наук
Институт энергетических проблем химической физики им. В.Л. Тальрозе Российской академии наук,
ВРИО директора, заведующий лабораторией

Официальные оппоненты: **Пономаренко Елена Александровна**
доктор биологических наук
Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича, заместитель директора по научной работе, заведующий лабораторией, главный научный сотрудник

Франкевич Владимир Евгеньевич
кандидат физико-математических наук
Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии имени академика В.И. Кулакова Министерства здравоохранения Российской Федерации,
заведующий отделом

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт биохимической физики им. Н.М. Эмануэля Российской академии наук

Защита состоится 05 июня 2019 года в 11 часов 00 минут на заседании диссертационного совета Д.002.012.02 при Федеральном государственном бюджетном учреждении науки Институте химической физики им. Н.Н. Семенова Российской академии наук по адресу: 119991 Москва, ул. Косыгина, д.4, корп. 1.

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института химической физики им. Н.Н. Семенова Российской академии наук.

Автореферат разослан 05 апреля 2019 года.

Автореферат размещен на сайте Высшей аттестационной комиссии Министерства образования и науки Российской Федерации 03 апреля 2019 года.

Ученый секретарь
диссертационного совета Д.002.012.02
доктор физико-математических наук М.Г. Голубков

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Белки, ключевые молекулы любого живого организма, представляют собой набор полипептидных цепей – линейных гетерополимеров, состоящих из аминокислотных остатков. Аминокислотные последовательности белков определяют разнообразие пространственных структур и функций этих биомолекул.

Протеомика – это область науки, объединяющая множество физико-химических методов исследования совокупности белков биологического образца, а также методов анализа данных с получением достоверной биологической информации. Подобные подходы относятся к постгеномным технологиям (поскольку опираются на результаты расшифровки геномов организмов) и представляют интерес, поскольку именно белки являются конечным функциональным продуктом генов, определение белкового состава клетки позволяет сделать выводы о процессах и состоянии клетки. В основе протеомного анализа лежат две технологии: хроматографическое разделение смесей белков и масс-спектрометрия, позволяющая получать информацию о первичной последовательности белковых молекул. Следует отметить, что объектом протеомики являются сложные смеси белков, экстрагированные из биологических проб и характеризующиеся широким диапазоном концентраций (до 10^8) и крайне высокой гетерогенностью, доходящей до миллионов индивидуальных высокомолекулярных соединений. Соответственно, создание новых и развитие существующих методов и подходов к анализу сложных смесей биомолекул является одной из основных задач в протеомике. В частности, дополнение масс-спектрометрии комплементарными физико-химическими методами анализа биомолекул позволяет получить больше количественной и качественной информации об

исследуемых образцах, особенно о низкоконцентрационных компонентах смеси, которые зачастую представляют наибольший интерес как потенциальные маркеры патологических состояний или участники важных биохимических процессов в клетках. К таким комплементарным методам относятся, в первую очередь, высокоэффективная жидкостная хроматография (ВЭЖХ) и тандемная масс-спектрометрия (МС/МС). Они позволяют получать информацию о первичной структуре, основанную на разных физико-химических свойствах биомакромолекул.

Цели и задачи исследования. Цель работы заключалась в выявлении источников комплементарной информации о физико-химических свойствах анализируемых многокомпонентных смесей белковых молекул и разработке теоретических моделей и подходов, позволяющих извлекать эту информацию из результатов хромато-масс-спектрометрического анализа. Конкретными задачами исследования были следующие:

1. Разработка теоретических основ идентификации первичных структур биомакромолекул на основе гидролиза различной специфичности.
2. Определение комплементарных данных о физико-химических свойствах пептидов, получаемых методами масс-спектрометрии в сочетании с высокоэффективной жидкостной хроматографией, оценка их влияния на эффективность идентификации первичной структуры исследуемых веществ.
3. Разработка основ метода идентификации биомакромолекул на основе частичного концевого химического гидролиза и его экспериментальная апробация.
4. Развитие метода критической жидкостной хроматографии биомакромолекул для предсказательного разделения на основе

гидрофильных взаимодействий и получения дополнительной комплементарной информации о первичной структуре анализируемых соединений.

5. Разработка оптимизационных подходов к фракционированию и анализу смесей биомакромолекул методами жидкостной хроматографии в сочетании с масс-спектрометрией.

Научная новизна работы заключается в следующем.

1. Впервые обоснована возможность использования гидролиза биомакромолекул различной специфичности для хромато-масс-спектрометрического анализа сложных смесей.
2. Впервые показана возможность прямой идентификации первичной структуры биомакромолекул методами хромато-масс-спектрометрии с использованием комплементарной физико-химической информации, включая данные частичного концевого химического гидролиза, предсказательной жидкостной хроматографии гидрофильных взаимодействий и изоэлектрического фокусирования.
3. Впервые предложена модель предсказательной критической жидкостной хроматографии биомакромолекул на основе гидрофильных взаимодействий. Экспериментально получены феноменологические параметры модели.
4. Предложен метод оптимизации фракционирования сложных смесей биомакромолекул.

Теоретическая и практическая значимость работы

Полученные результаты могут быть использованы исследователями и специалистами в области разработки физико-химических методов анализа структуры веществ, хроматографии синтетических и

природных биополимеров, разработки постгеномных технологий, протоколов физико-химического анализа биологических и клинических проб, а также методов медицинской диагностики. Предложенные в работе методы, основанные на прямой хромато-масс-спектрометрической идентификации первичной структуры биомакромолекул, позволяют увеличить динамический диапазон анализа протеомов сложных организмов, включая человека, в задачах персонализированной медицины.

Методы исследования

В работе использовались следующие физико-химические методы исследования:

- микро- и нанопотоковая высокоэффективная жидкостная хроматография;
- масс-спектрометрия на основе радиочастотной ионной ловушки;
- времяпролётная масс-спектрометрия высокого разрешения;
- масс-спектрометрия высокого и сверхвысокого разрешения на основе электростатической орбитальной ионной ловушки;
- обращённо-фазовая хроматография на стационарной фазе C_{18} и хроматография гидрофильных взаимодействий HILIC на немодифицированном силикагеле.

Подготовка сложных смесей биомакромолекул проводилась с использованием стандартных биохимических методов.

Теоретические исследования проводились с использованием программных ресурсов, разработанных автором с использованием языка программирования высокого уровня с динамической типизацией Python.

Положения, выносимые на защиту:

1. Использование частичного химического гидролиза и гидролиза различной специфичности для идентификации первичной структуры пептидов;
2. Хромато-масс-спектрометрическая идентификация биомакромолекул на основе комплементарных физико-химических данных;
3. Модель критической жидкостной хроматографии гидрофильных взаимодействий для предсказания времён удерживания биомакромолекул с заданной первичной структурой;
4. Оптимизация условий жидкостной хроматографии для фракционирования сложных смесей на основе предсказательной хроматографии биомакромолекул.

Степень достоверности полученных результатов

Достоверность полученных результатов обеспечена использованием известных апробированных теоретических и экспериментальных подходов. Результаты расчётов подтверждаются экспериментальными данными.

Апробация результатов

Основные результаты диссертации были доложены на конференциях: 9-я конференция Австрийского протеомного общества (г. Вена, Австрия 2011); 6-я конференция по протеомике Центральной и Восточной Европы (г. Будапешт, Венгрия 2012); 6-я и 9-я летние школы по масс-спектрометрии в биологии и медицине (г. Дубровник, Хорватия 2012, 2015); RECOMB-CP 2012 конференция по вычислительной протеомике (г. Сан-Диего, США 2012); 23-й и 24-й Масс-спектрометрический форум (г. Вена, Австрия 2012, 2013); 61-я (г. Миннеаполис, США 2013), 62-я (г. Балтимор, США 2014), 64-я (г. Сан-Антонио, США 2016), 65-я (г.

Индианаполис, США 2017), 66-я (г. Сан-Диего, США 2018) конференции Американского масс-спектрометрического общества; 38-й конгресс Европейских биохимических обществ (г. Санкт-Петербург, Россия 2013); 6-й и 7-й съезды Всероссийского масс-спектрометрического общества (г. Москва, Россия 2013, 2015); 57-я Научная конференция МФТИ (г. Долгопрудный, Россия 2014); 7-й Российский симпозиум "Белки и пептиды"(г. Новосибирск, Россия 2015); 45-й международный симпозиум по высокоэффективным жидкостным методам разделения и связанным методам (г. Прага, Чехия 2017).

Публикации

По материалам диссертации опубликовано 7 печатных работ. Работ, опубликованных в рецензируемых научных изданиях, рекомендованных ВАК – 7.

Личный вклад автора

Результаты, представленные в диссертации, получены при непосредственном участии автора, включая выбор задач исследования, обоснование выбранных методов исследования, выполнение экспериментов, разработку программных ресурсов для обработки результатов экспериментов, а также разработку теоретических моделей.

Объем и структура диссертации

Диссертация состоит из введения, пяти глав, формулировки основных результатов и выводов, списка сокращений и списка цитируемой литературы. Работа изложена на 115 страницах и содержит 44 рисунка, 5 таблиц и библиографию из 173 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обосновывается актуальность исследований, проводимых в рамках представленной диссертационной работы, формулируются цель и задачи, обсуждаются её научная новизна и практическая значимость.

Первая глава содержит обзор литературы. Описываются наиболее распространённые задачи и проблемы анализа биомакромолекул в области протеомики, то есть масштабного исследования белков, и методы их решения, основанные на масс-спектрометрии высокого разрешения, а также тандемной масс-спектрометрии (МС/МС), в сочетании с высокоэффективной хроматографией сложных смесей биомакромолекул и различными физико-химическими методами предварительного фракционирования. Обсуждается ряд ограничений тандемной масс-спектрометрии, таких как низкая скорость анализа, чувствительность и динамический диапазон. В первой главе также представлен обзор моделей предсказания хроматографических времён удерживания пептидов, применяемых в протеомике, в том числе описана модель критической жидкостной хроматографии биомолекул BioLCCC, позволяющая предсказывать времена удерживания пептидов и белков в заданных условиях, таких как неподвижная и подвижная фазы, параметры хроматографической колонки, профиль градиента, на основе их первичной структуры (аминокислотной последовательности).

Во **второй главе** рассмотрен подход к проблеме идентификации биомолекул в сложных смесях, основанный на ферментативном или химическом гидролизе. Поскольку перед хромато-масс-спектрометрическим анализом белков производится гидролиз полипептидной цепи, от специфичности данного гидролиза зависит сложность получаемой смеси пептидов: исследуемую смесь пептидов

можно упростить, если использовать альтернативный наиболее часто используемому трипсину метод гидролиза белковых молекул. Использование гидролизующих агентов различной специфичности и их комбинаций для гидролиза белковых образцов позволяет получить пептидные смеси различной сложности. При гидролизе по более редким аминокислотным остаткам конечная смесь упрощается, но при этом в среднем получаются более длинные пептиды, требующие особых условий хромато-масс-спектрометрического анализа. По этим признакам теоретические протеолитические пептиды были разделены на три группы: 0.6–3 кДа – классическая протеомика коротких пептидов «снизу-вверх», 3–7 кДа – протеомика среднего массового диапазона, характеризующаяся использованием менее удерживающих хроматографических колонок и более высокоэнергетичных методов фрагментации ионов, а также диапазон от 7 до 15 кДа – протеомика длинных пептидов, для которой требуется адаптация методик эксперимента и алгоритмов обработки данных (см. рис. 1).

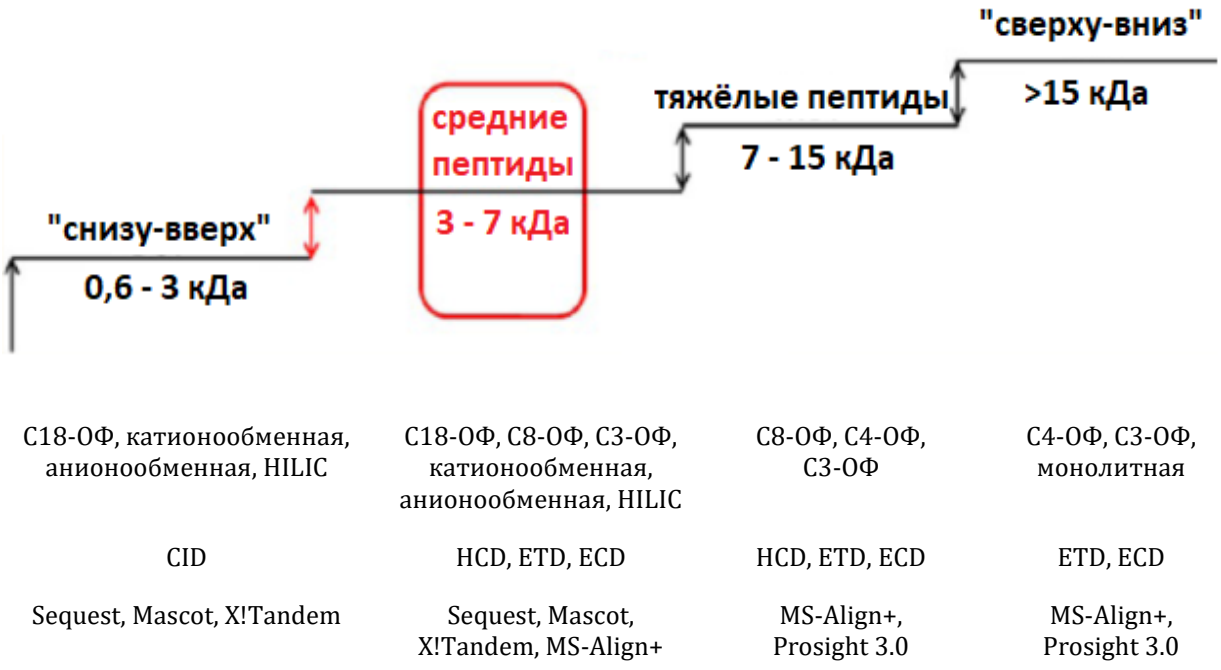

Рисунок 1. Диапазоны масс пептидов, соответствующие различным подходам к хромато-масс-спектрометрическому анализу. Характерные хроматографические и масс-спектрометрические методы, а также программы для идентификации биомолекул, применимые для каждого из приведённых диапазонов.

Кроме того, чем проще получаемая смесь пептидов, тем ниже теоретически достижимое покрытие протеома, то есть количество белков исследуемого организма, имеющих хотя бы один теоретический пептид в детектируемом диапазоне масс. Приведены теоретические расчёты количества пептидов (сложности смеси) и покрытия протеома для случаев гидролиза различной специфичности.

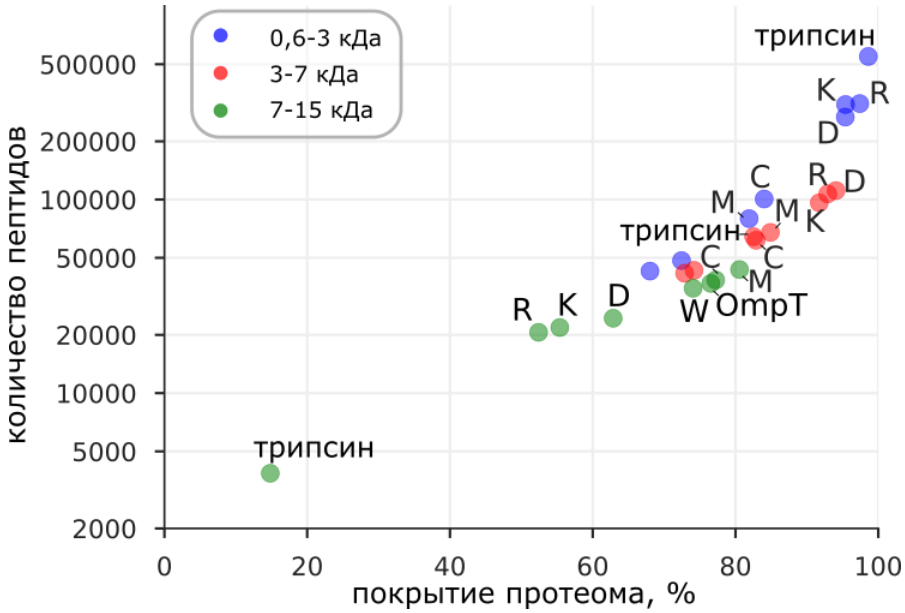

Рисунок 2. Рассчитанные значения покрытия протеома человека (% теоретически идентифицируемых белков) и сложности пептидной смеси (количества пептидов) для гидролизующих агентов различной специфичности и трёх исследуемых диапазонов масс пептидов. Гидролиз по одной аминокислоте обозначен однобуквенным кодом, также приведены результаты для двух ферментов: трипсина (специфичность к остаткам K и R), OmpT (специфичность к участкам KK, RK, RR, KR). Масштаб по вертикальной оси – логарифмический.

Показано, что трипсин даёт максимальное покрытие протеома, однако это происходит за счёт кратного увеличения количества пептидов в смеси (см. рис. 2). Учитывая характерное количество детектируемых в стандартном хромато-масс-спектрометрическом эксперименте пептидов порядка 10^4–10^5, можно сделать вывод о том, что использование трипсина в классической протеомике коротких пептидов, несмотря на удобство и распространённость данного метода, далеко не всегда будет оптимальным выбором. Необходимо разрабатывать надёжные и воспроизводимые методы специфичного

нетриптического гидролиза белковых цепей. Также во второй главе приведены результаты экспериментальной апробации методов химического гидролиза белковых цепей, специфичных к редким аминокислотным остаткам: метионину, цистеину и триптофану.

Рассматривается проблема дискриминации коротких пептидов поисковым алгоритмом Morpheus и предлагается программный модуль GroupFilter для решения этой проблемы путём группировки и раздельной фильтровки результатов.

Третья глава содержит обоснование комплементарных физико-химических подходов для идентификации первичной структуры линейных биомакромолекул с использованием методов, основанных на масс-спектрометрии и жидкостной хроматографии. Приводятся расчёты, подтверждающие применимость комплементарных методов для прямой хромато-масс-спектрометрической идентификации аминокислотных последовательностей пептидов. В качестве источника такой комплементарной информации о первичной структуре пептида могут служить, например, хроматографические и электрофоретические методы разделения сложных смесей биомолекул. Проиллюстрирована ортогональность каждого из приведённых методов стандартным хромато-масс-спектрометрическим измерениям.

Поскольку классические задачи протеомики подразумевают идентификацию первичной структуры пептидов на основе известных баз данных последовательностей, можно теоретически оценить вклад в частоту однозначных идентификаций от использования того или иного источника комплементарной информации о физико-химических свойствах биомолекул. На рис. 3 приведены результаты расчётов, иллюстрирующие эффективность прямой хромато-масс-спектрометрической идентификации пептидов из протеома дрожжей или человека в двумерном пространстве поиска (по точной массе и

хроматографическому времени выхода) и в трёхмерном пространстве при добавлении информации об изоэлектрических точках пептидов. Точность измерения каждого из наблюдаемых физико-химических свойств пептида определяет соответствующий интервал идентификации, то есть тот интервал, в пределах которого экспериментально наблюдаемая величина считается совпавшей с теоретически предсказанной. В приведённых расчётах были заданы следующие точности: для измерения масс – 1 ppm, для времён удерживания пептида на обращённой фазе (ОФ) – от 0.1 до 10 минут, для расчёта изоэлектрической точки – 0.1 или 0.5.

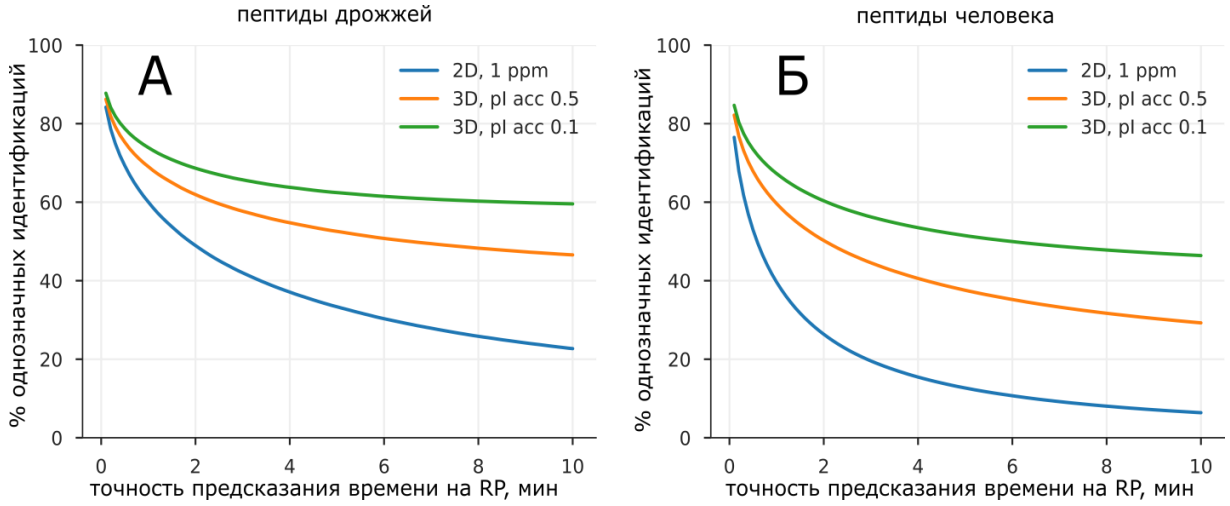

Рисунок 3. Теоретические оценки частоты прямой хромато-масс-спектрометрической идентификации первичной структуры пептидов в двумерном пространстве поиска (масса, хроматографическое время) и в трёхмерном, с добавлением комплементарной информации об изоэлектрических точках (pI), точность измерения 0.1 или 0.5. Оценки проведены для баз данных: (А) пекарских дрожжей *S. cerevisiae*, (Б) белков человека. Во всех расчётах принята точность измерения масс 1 ppm.

Аналогично в качестве комплементарного физико-химического метода в данной главе был рассмотрен частичный концевой химический гидролиз пептида по методу Эдмана. Такой гидролиз позволяет определить концевой аминокислотный остаток, иными словами, дает частичную информацию о последовательности, которая позволяет существенно снизить количество пептидов-кандидатов из базы данных с измеренными значениями точной массы и хроматографического

времени удерживания. В зависимости от точности измерения масс и предсказания хроматографических времен возникает возможность однозначной идентификации пептида. В диссертации произведены оценки количества потенциально идентифицируемых в описанном трёхмерном пространстве поиска (масса, хроматографическое время удерживания, N-концевой аминокислотный остаток) пептидов и белков в зависимости от размера базы данных (для двух наиболее часто исследуемых организмов: пекарских дрожжей *S. cerevisiae* и человека), а также от точностей измерения масс и предсказания хроматографических времён удерживания (см. рис. 4).

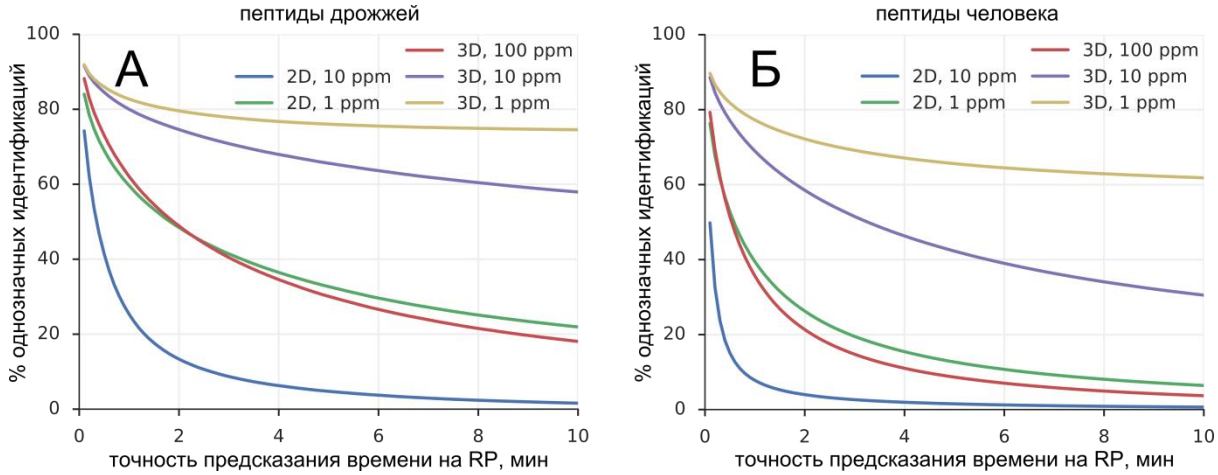

Рисунок 4. Оценка эффективности предложенного трёхмерного метода для однозначной идентификации пептидов. Зависимость частоты (в %) однозначных идентификаций пептидов от значений точности определения масс и хроматографических времён удерживания. (А) База данных пептидов дрожжей, (Б) пептиды человека. Обозначение «2D» соответствует стандартному двумерному поисковому пространству (масса, время удерживания); «3D» – предложенному трёхмерному пространству с учётом информации об N-концевом аминокислотном остатке.

Приведены результаты предварительных экспериментов, подтверждающих возможность проведения реакции N-концевого гидролиза по Эдману за времена, совместимые с потоковой реализацией метода (см. рис. 5), а также идентификации одиночного белка (бычьего цитохрома С) по полученной хромато-масс-спектрометрической информации о его триптических пептидах до и после реакции

частичного концевого гидролиза. Приведённые в данной главе расчёты и экспериментальные результаты подтверждают эффективность и практическую возможность использования следующих комплементарных методов для прямой хромато-масс-спектрометрической идентификации первичной структуры пептидов: изоэлектрического фокусирования, частичного концевого химического гидролиза по методу Эдмана.

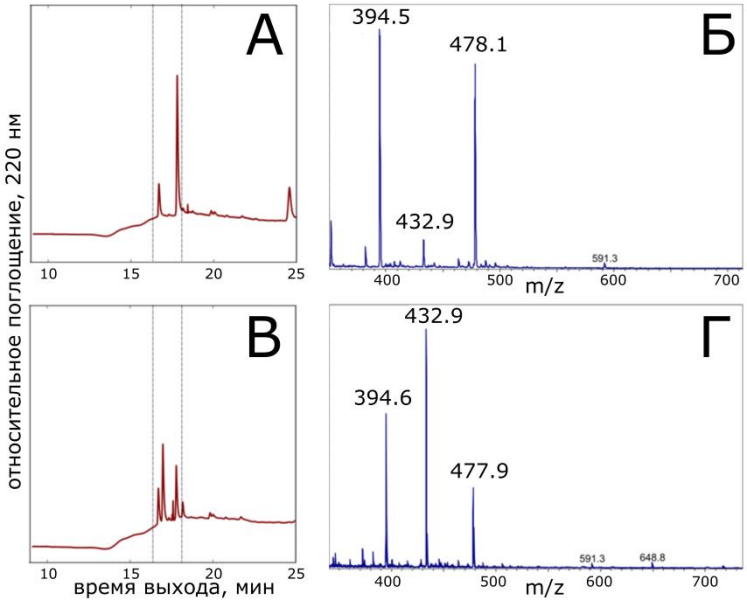

Рисунок 5. Хромато-масс-спектрометрический анализ продуктов деградации по Эдману ангиотензина I при элюировании ступенчатым градиентом. Вертикальными пунктирными линиями обозначены границы собранной фракции. (А) хроматограмма и (Б) масс-спектр для реакционной смеси после проведения реакции Эдмана по стадартному протоколу; (В) хроматограмма и (Г) масс-спектр для реакционной смеси после проведения реакции по укороченному протоколу. Пики в масс-спектре соответствуют трёхзарядным ионам обычного (433 Th), модифицированного (478 Th) и укороченного (394.5 Th) вариантов пептида.

В **<u>четвёртой главе</u>** подробно обсуждается применение хроматографии гидрофильных взаимодействий (HILIC) в задачах анализа сложных протеомных смесей. Гидрофильная хроматография сочетает в себе использование неподвижных фаз, характерных для нормально-фазного разделения, таких как немодифицированный или модифицированный амидными, диольными или другими полярными группами силикагель, и подвижных фаз, характерных для обращённо-

фазовой хроматографии, то есть водно-органических смесей с добавлением кислот или других ион-парных агентов или буферов. Элюирующим компонентом растворителя в гидрофильной хроматографии является вода. Этот метод широко применяется для разделения полярных веществ, в том числе биомакромолекул, включая пептиды, особенно содержащие полярные модификации, такие как гликозилирование и фосфорилирование. Несмотря на кажущуюся противоположность механизмов разделения в гидрофильной и обращённо-фазовой хроматографии, эти методы разделения пептидов являются в достаточной степени ортогональными (см. рис. 6), поэтому могут быть использованы в качестве двух последовательных направлений двумерной хроматографии.

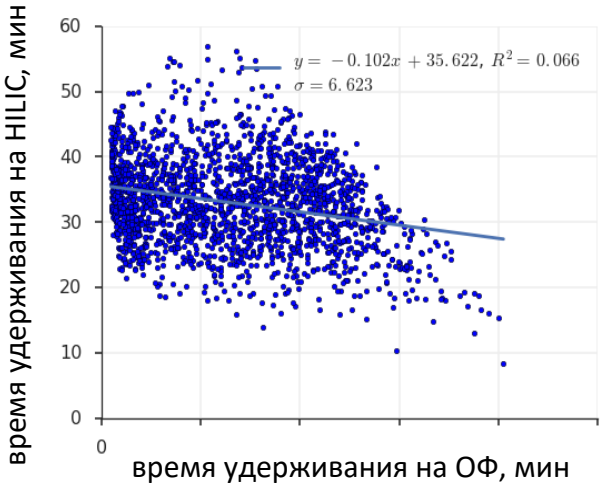

Рисунок 6. Распределение случайной выборки 2000 пептидов из протеома дрожжей S. cerevisiae по теоретическим временам удерживания на обращённой фазе и на гидрофильной фазе HILIC. Низкое значение коэффициента корреляции (R^2=0.066) свидетельствует о том, что данные два хроматографических метода являются ортогональными.

Механизм удерживания биомолекул в хроматографии гидрофильных взаимодействий HILIC до сих пор не выяснен до конца, поскольку не разработана соответствующая физическая модель удерживания пептидов. Однако выделяют следующие основные типы взаимодействий анализируемых молекул с неподвижной фазой: адсорбционное, распределительное между подвижной фазой и адсорбированным на поверхности сорбента слоем воды, электростатическое при наличии на поверхности сорбента заряженных групп. В литературе было показано, что основным из этих

взаимодействий для большинства классически используемых стационарных фаз можно считать адсорбционное. Для проверки этого утверждения, а также для разработки универсального предсказателя времён удерживания пептидов на стационарной фазе HILIC была произведена адаптация модели хроматографии биомакромолекул в критических условиях BioLCCC к описанию удерживания пептидов на гидрофильной фазе. Модель BioLCCC, которая была разработана и откалибрована ранее для наиболее популярной в протеомике обращённой фазы C_{18}, позволяет теоретически рассчитать хроматографические времена выхода пептидов с известной аминокислотной последовательностью для заданных измерений колонки и условий элюирования. В основе модели лежат представления о плоской поре, модель цепи случайных блужданий для описания конформации гетерополимерной молекулы пептида в поре, понятие о линейной элюирующей силе бинарного растворителя (аналог модели Снайдера). Система уравнений модели BioLCCC для расчёта объёма удерживания пептида V_R выглядит следующим образом:

$$\varepsilon_{AB}(V) = \varepsilon_A + \ln\left(N_B(V)e^{\varepsilon_B-\varepsilon_A} - 1 + N_B(V)\right), \quad (1)$$

$$\varepsilon_i(V) = \alpha\left(X_i^0 - \varepsilon_{AB}(V)\right), \quad (2)$$

$$K_d(V) = \frac{1}{D}U^T \prod_i W\left(\varepsilon_i(V)\right)U, \quad (3)$$

$$\int_0^{V_R-V_0} \frac{dV}{V_p(K_d(V)-1)} = 1. \quad (4)$$

Уравнение (1) определяет элюирующую силу бинарного растворителя $\varepsilon_{AB}(V)$, как функцию мольной доли $N_B(V)$ компонента B, соответствующей объёму удерживания V, и сил каждого из растворителей в отдельности, ε_A и ε_B. Уравнение (2) задаёт относительную энергию связывания аминокислотного остатка с поверхностью сорбента $\varepsilon_i(V)$ в каждой точке градиента как разность

эффективной энергии адсорбции данного остатка X_i^0, и десорбции бинарного растворителя, $\varepsilon_{AB}(V)$; свободный параметр α обозначает «активность» адсорбента и отражает различия между хроматографическими колонками, а также эффект старения колонки. В уравнении (3) значение коэффициента распределения $K_d(V)$ определяется множеством допустимых конформаций гетерополимерной цепи в поре через произведение переходных матриц для каждого из аминокислотных остатков в последовательности пептида $W\left(\varepsilon_i(V)\right)$. Единичный вектор U и транспонированный вектор U^T соответствуют равновероятному расположению N-концевого аминокислотного остатка пептида в поре. V_0 обозначает так называемый мёртвый объём хроматографической системы, V_p – объём пор сорбента, D – размер поры в единицах длины мономера.

Адаптация модели критической хроматографии для описания удерживания пептидов на гидрофильной фазе включала в себя определение следующих феноменологических параметров: эффективных энергий взаимодействия всех аминокислотных остатков с поверхностью сорбента, энергии элюирующего растворителя (в случае обращённо-фазовой хроматографии – ацетонитрил, для HILIC – вода), энергий взаимодействия с поверхностью сорбента различных N- и C-концевых химических групп.

Для калибровки модели критической хроматографии гидрофильных взаимодействий пептиды калибровочной смеси (см. таблицу 1) были проанализированы по отдельности в последовательных изократических экспериментах с различным содержанием воды в бинарном растворителе, в качестве ион-парного агента использовалась трифторуксусная кислота, времена удерживания определялись с помощью ультрафиолетового детектора.

Таблица 1. Последовательности пептидов, использованных для калибровки модели критической хроматографии биомолекул BioLCCC на гидрофильной стационарной фазе HILIC в изократическом режиме элюирования.

Имя	**Последовательность**	**Имя**	**Последовательность**
C1	Ac-GVGKGGVGVK-NH_2	C17	Ac-AVGKGAVGVK-NH_2
C2	Ac-VVKGGVGKVGV-NH_2	C18	Ac-YVGKGYVGVK-NH_2
C3	Ac-KGVGKVGGVK-NH_2	C19	Ac-MVGKGMVGVK-NH_2
C4	Ac-VVGVKGGVGK-NH_2	C20	Ac-IVGKGIVGVK-NH_2
C5	Ac-KGVGGKVGVV-NH_2	C21	Ac-LVGKGLVGVK-NH_2
C6	Ac-GVGGVK-NH_2	C22	Ac-FVGKGFVGVK-NH_2
C7	Ac-KGVVKGVGVKGGVKG-NH_2	C23	Ac-WVGKGWVGVK-NH_2
C8	Ac-HGVGHVGGVK-NH_2	C27	H-GVGKGGVGVK-NH_2
C9	Ac-RGVGRVGGVR-NH_2	C28	Ac-GVGKGGVGVK-OH
C12	Ac-NVGKGNVGVK-NH_2	C29	H-VVKGGVGKVGV-NH_2
C13	Ac-SVGKGSVGVK-NH_2	C30	Ac-VVKGGVGKVGV-OH
C14	Ac-QVGKGQVGVK-NH_2	C31	Ac-PVGKGPVGVK-NH_2
C15	Ac-CVGKGCVGVK-NH_2	C15*	Ac-camCVGKGcamCVGVK-NH_2
C16	Ac-TVGKGTVGVK-NH_2	E1	H-WAGGQESGK-OH

Последовательности пептидов, использованных для калибровки модели, подобраны следующим образом: “ядро” составляют 7 пептидов, состоящих только из аминокислотных остатков G, V, K, с концевыми группами Ac- и -NH_2 (пептиды C1-C7 в таблице 1). Варьированием соответствующих эффективных энергий адсорбции, а также энергии элюирующего растворителя, то есть воды, и неизвестного параметра системы – мёртвого объёма V_0 – определялись значения этих параметров, позволяющие минимизировать отклонение рассчитанных на основе модели BioLCCC времён удерживания RT_i^{theor} для семи пептидов от экспериментальных RT_i^{exp} во всех изократических точках:

$$\Delta = \sum_i \left(RT_i^{theor} - RT_i^{exp}\right)^2. \quad (5)$$

В результате минимизации штрафной функции Δ было определено «ядро» модели, после чего использовались изократические данные, полученные для остальных пептидов калибровочной смеси. Энергии, полученные таким образом для всех групп, приведены в таблице 2.

Таблица 2. Энергии (в единицах kT) связывания с колонкой HILIC (немодифицированный силикагель), рассчитанные для элюирующего растворителя (воды), аминокислотных остатков и концевых групп.

H_2O	**Ac-**	**H-**	**-NH_2**	**-OH**	**A**	**C**
1.746	*0.634*	*0.139*	*0.893*	*-0.268*	*1.361*	*1.156*
camC	**D**	**E**	**F**	**G**	**H**	**I**
1.689	*0.614*	*1.444*	*0.610*	*1.581*	*1.509*	*0.746*
K	**L**	**M**	**N**	**P**	**Q**	**R**
1.638	*0.662*	*0.855*	*1.801*	*1.751*	*1.944*	*1.434*
S	**T**	**V**	**W**	**Y**		
1.397	*1.371*	*0.980*	*0.562*	*0.676*		

На рис. 7А показана полученная корреляция между экспериментальными и рассчитанными временами удерживания в изократическом режиме для пептидов калибровочной смеси. В качестве первичной проверки калибровки был проведён эксперимент с градиентным элюированием и масс-спектрометрическим детектированием 25 пептидов из калибровочной смеси (см. рис. 7Б), подтверждающий применимость модели критической хроматографии для описания удерживания биомолекул на гидрофильной фазе как в изократическом, так и в градиентном режиме элюирования.

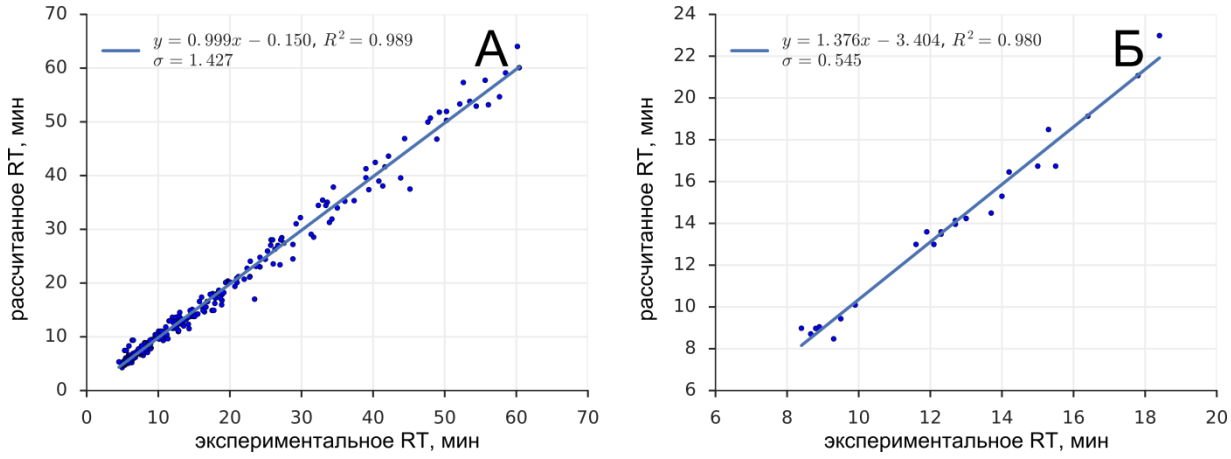

Рисунок 7. Результаты калибровки модели BioLCCC для хроматографии HILIC: корреляция экспериментальных и теоретических времён выхода калибровочных пептидов в (А) изократических и (Б) градиентных условиях.

Проверка и уточнение полученных энергий были проведены на следующих образцах: (1) смесь 20 пептидов с последовательностями H-GGGXGGG-NH_2, где на месте X находится каждый из 20 аминокислотных остатков, (2) гидролизат бычьего сывороточного альбумина (БСА). Результаты приведены на рис. 8. В обоих случаях получена достаточно высокая точность предсказания времён выхода пептидов, что позволяет использовать модель критической хроматографии биомолекул для описания удерживания пептидов на гидрофильной фазе, в том числе для планирования экспериментов *in silico* и фильтровки результатов идентификации.

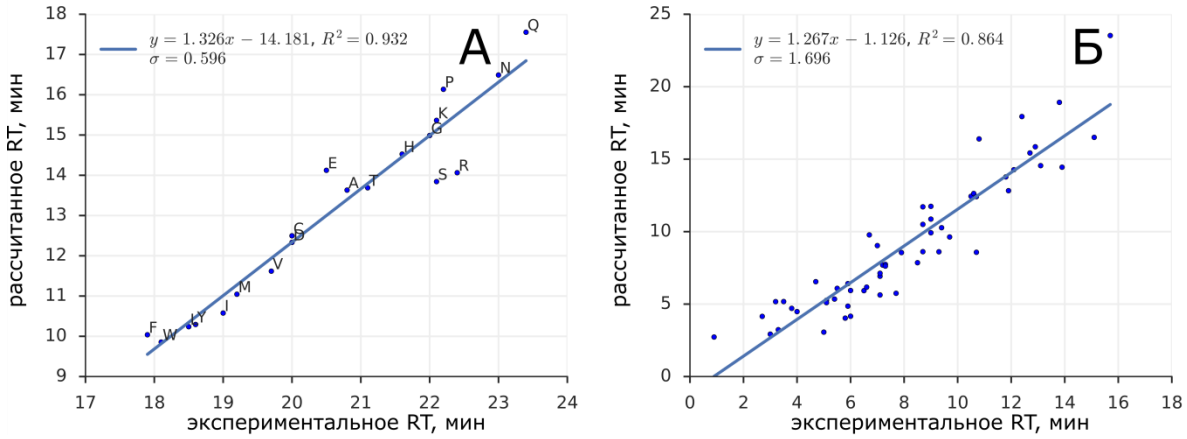

Рисунок 8. Проверка точности калибровки модели BioLCCC на (А) смеси синтетических пептидов вида H-GGGXGGG-NH2, (Б) триптическом гидролизате БСА.

Пятая глава посвящена разработке метода оптимизации фракционирования сложных смесей биомакромолекул, предложенного в качестве одного из возможных применений модели предсказательной хроматографии BioLCCC. Фракционирование образцов перед хромато-масс-спектрометрическим анализом позволяет достичь более глубокого покрытия протеома, то есть получить информацию о большем количестве компонентов смеси с низкими концентрациями. Для фракционирования биомолекул используются такие физико-химические методы разделения, как изоэлектрическое фокусирование

(разделение по значению pI) и различные виды хроматографии: ионообменная, обращённо-фазовая при различных значениях pH, гидрофильная. Как правило, для предварительного фракционирования смеси пептидов используется метод, ортогональный ОФ хроматографии, что позволяет повысить так называемую “пиковую ёмкость” метода, поскольку пептиды, не разделяемые одним методом, могут разделиться другим, ортогональным ему. Однако эффект будет положительным за счёт упрощения анализируемой смеси даже при использовании неортогонального метода (разделение на обращённой фазе и для префракционирования, и для ЖХ-МС/МС анализа каждой из фракций). Важным вопросом при фракционировании смеси биомолекул является распределение компонентов смеси между фракциями, поскольку от количества пептидов в анализируемой фракции зависят оптимальные настройки хромато-масс-спектрометрической системы. Неравномерное распределение компонентов смеси между фракциями приводит к неоптимальному режиму работы масс-спектрометра: при анализе фракций с меньшим количеством пептидов большую часть времени он будет работать вхолостую, а в случае более “богатых” фракций не будет успевать анализировать все пептиды.

Использование предсказательной хроматографии BioLCCC позволяет оптимизировать условия сбора фракций таким образом, чтобы количество компонентов во всех фракциях было одинаковым. Существуют два подхода к “выравниванию” распределения пептидов между фракциями: модификация градиента таким образом, чтобы времена сбора фракций оставались одинаковыми, или подбор некого универсального градиента с варьированием времён сбора фракций. Второй подход оказался предпочтительнее, поскольку предсказание времён удерживания в нелинейном градиенте, необходимое для реализации первого подхода, даёт большие погрешности.

Специально разработанная программа FractionOptimizer позволяет, используя информацию об идентифицированных в предварительном эксперименте без фракционирования пептидах, на основе предсказанных времён удерживания рассчитать оптимальные времена сбора фракций. Кроме того, каждой из фракций соответствует определённый диапазон концентраций органического растворителя, в котором элюируется большая часть пептидов. Особенно ярко выражен этот эффект, если для префракционирования используется обращённо-фазовая хроматография, как и для анализа фракций. Использование предсказательной хроматографии BioLCCC в программе Optimizer позволяет разрабатывать аналитические градиенты для каждой из фракций таким образом, чтобы пептиды выходили на протяжении всего времени анализа и, следовательно, масс-спектрометр работал максимально эффективно. Полная схема анализа сложной смеси пептидов с использованием оптимизации фракционирования приведена на рис. 9.

Рисунок 9. Схема протеомного эксперимента с оптимизацией условий префракционирования и анализа фракций.

В этой главе приведены расчёты для сравнения алгоритма FractionOptimizer с другими подходами к хроматографическому фракционированию смесей протеолитических пептидов. Также результат работы программы Optimizer показан на примере анализа клеточной линии эмбриональной почки человека HEK293 (см. рис. 10). На основе идентификаций пептидов нефракционированного образца были рассчитаны 5 оптимальных фракций, а также аналитические градиенты для каждой из них. Таким образом удалось достичь более равномерного распределения пептидов по фракциям и более глубокого покрытия протеома.

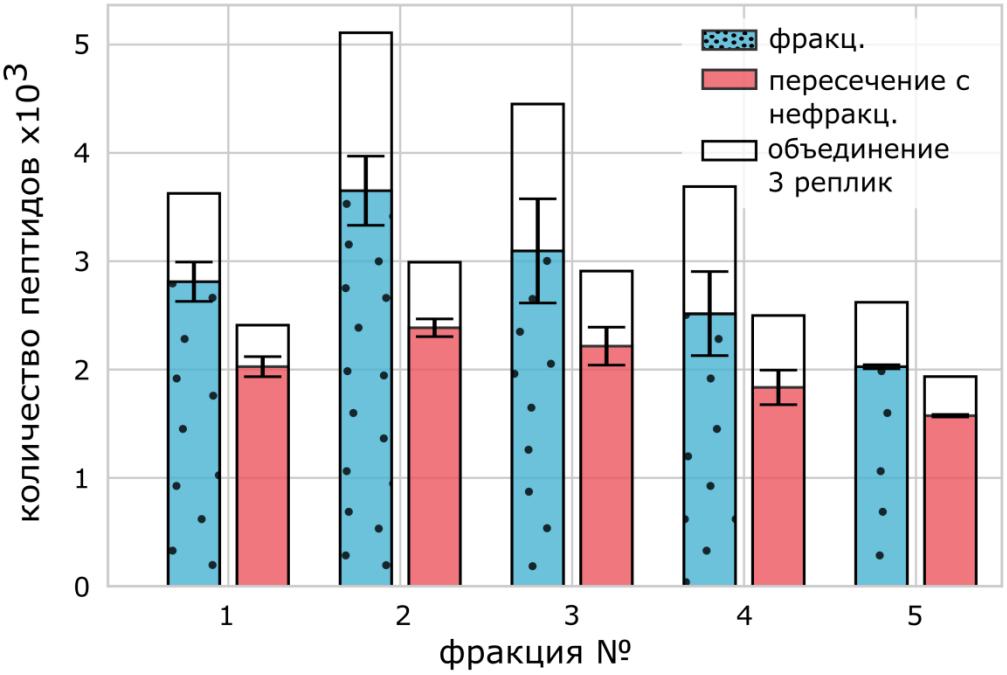

Рисунок 10. Количества пептидов, идентифицированных в каждой из пяти фракций гидролизата клеточной линии HEK293. Синими столбцами показаны все идентифицированные пептиды, красными – те из них, которые были также обнаружены при анализе нефракционированного образца. Кресты ошибок показывают разброс между тремя репликами для каждой из фракций, а белыми столбцами показаны результаты объединения трёх реплик.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Проведена теоретическая оценка возможности использования протеолитического гидролиза различной специфичности для идентификации первичной структуры биомакромолекул, а также экспериментальная апробация химических методов расщепления белковых цепей.

2. Предложены подходы прямой хромато-масс-спектрометрической идентификации белков и пептидов на основе комплементарных физико-химических данных, в том числе изоэлектрических точек и массы N-концевого аминокислотного остатка. Экспериментально показана возможность использования частичного концевого гидролиза по методу Эдмана для определения аминокислотной последовательности белков и пептидов.

3. Предложена модель критической жидкостной хроматографии для предсказания времён удерживания биомакромолекул с заданной первичной структурой на основе гидрофильных взаимодействий. Рассчитаны феноменологические параметры модели.

4. Разработан метод для расчёта оптимальных условий фракционирования сложных смесей протеолитических пептидов на основе предсказательной хроматографии биомакромолекул. Эффективность метода продемонстрирована на примере анализа клеточной линии человека HEK293.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Lobas A.A., Verenchikov A.N., Goloborodko A.A., Levitsky L.I., Gorshkov M.V. Combination of Edman degradation of peptides with liquid chromatography/mass spectrometry workflow for peptide identification in bottom-up proteomics // Rapid Communications in Mass Spectrometry. — 2013. — V. 27. — N. 3. — P. 391–400.

2. Laskay Ü.A., Lobas A.A., Srzentić K., Gorshkov M.V., Tsybin Y.O. Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments // Journal of Proteome Research. — 2013. — V. 12. — N. 12. — P. 5558–5569.

3. Иванов М.В., Левицкий Л.И., Лобас А.А., Горшков М.В. GroupFilter: программное решение проблемы дискриминации идентификаций масс-спектров высокого разрешения при работе поискового алгоритма Morpheus // Масс-спектрометрия. — 2015. — Т. 12. — N. 4. — P. 229–235.

4. Lobas A.A., Karpov D.S., Kopylov A.T., Solovyeva E.M., Ivanov M.V., Ilina I.Y., 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. — V. 16. — N. 14. — P. 1980–1991.

5. Лобас А.А., Левицкий Л.И., Фихтенбаум А., Сурин А.К., Придатченко М.Л., Митулович Г., Горшков А.В., Горшков М.В. Предсказательная жидкостная хроматография пептидов на основе гидрофильных взаимодействий в задачах масс-спектрометрического анализа протеомов // Масс-спектрометрия. — 2017. — Т. 14. — N. 2. — С. 79–88.

6. Solovyeva E.M., Lobas A.A., Kopylov A.T., Ilina I.Y., Levitsky L.I., Moshkovskii S.A., Gorshkov M.V. FractionOptimizer: a method for optimal peptide fractionation in bottom-up proteomics // Analylical and Bioanalytical Chemistry. — 2018. — V. 410. — N. 16. — P. 3827–3833.

7. Srzentić K., Zhurov K.O., Lobas A.A., Nikitin G., Fornelli L., Gorshkov M.V., Tsybin Y.O. Chemical-Mediated Digestion: An Alternative Realm for Middle-down Proteomics? // Journal of Proteome Research. —2018. — V. 17. — N. 6. — P. 2005–2016.