

МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ ИДЕНТИФИКАЦИИ ТКАНЕСПЕЦИФИЧНЫХ БЕЛКОВ И ПЕПТИДОВ, ФОРМИРУЮЩИХ КОРРИГИРУЮЩИЕ СВОЙСТВА ИННОВАЦИОННЫХ МЯСНЫХ ПРОДУКТОВ

Вострикова Н.Л.^{1,*}, Чернуха И.М.¹, Хвостов Д.В.^{1,2}

¹ Федеральный научный центр пищевых систем им. В.М. Горбатова РАН, Москва, Россия

² Научный центр биомедицинских технологий Федерального медико-биологического агентства России, Московская область, Россия

Ключевые слова: протеомика, двумерный электрофорез, биомаркеры, масс-спектрометрия

Аннотация

Одним из направлений решения стоящих перед отраслью вопросов качества питания является разработка стандартизированных и аттестованных методик, связанных с проведением глубоких исследований биохимических показателей качества и безопасности мяса и мясных продуктов. Мировая лабораторная практика в области качества и безопасности продуктов питания свидетельствуют о постоянном расширении списка контролируемых показателей пищевого сырья и продукции. Важной особенностью современного периода в развитии биомедицинских и биотехнологических исследований является внедрение целого комплекса постгеномных технологий, в основе которых лежит системный подход к изучению особенностей функционирования протеома млекопитающего при различных физиологических и патологических состояниях, в том числе при формировании и развитии алиментарно-зависимых патологий. В связи с этим проблема многостороннего изучения пищевых продуктов, в частности их идентификация, является наиболее актуальной, поскольку современная технология их производства претерпела существенные изменения и требует разработки «щадящих» режимов обработки. Они касаются исходного сырья и вспомогательных материалов, используемых на всех этапах производства продукции. Это и новые технологии производства белковых продуктов из растительного сырья, а также введение в продовольственное сырье и продукты питания пищевых добавок искусственного происхождения и избыточное внесение добавок растительного и животного может стать причиной фальсификации продуктов, а также сказаться на здоровье потребителя. Оценка качества пищевых продуктов включает контроль компонентов готовых изделий. Наиболее сложно определить долю мышечнотканного белка в многокомпонентных мясных продуктах, прошедших термическую обработку. Поэтому в практике при оценке качества пищевых продуктов возникает необходимость идентификации его реального состава в соответствии с декларированным нормативным документам. В настоящее время перспективным направлением научных исследований в области определения состава готовых пищевых продуктов является выделение биомаркеров различных компонентов. В связи с этим актуальным является разработка методологии идентификации биохимических изменений пищевого сырья под действием технологических факторов с использованием современных методов исследования. В данной работе представлен обзор методологии анализа белков и пептидов, включающий последние технологии, которые приобретают всю большую актуальность.

Review paper

METHODOLOGICAL ASPECTS OF IDENTIFICATION OF TISSUE-SPECIFIC PROTEINS AND PEPTIDES FORMING THE CORRECTIVE PROPERTIES OF INNOVATIVE MEAT PRODUCTS

Natal'ya L. Vostrikova^{1,*}, Irina M. Chernukha¹, Daniil V. Khvostov²

¹ V.M. Gorbatov Federal Research Center for Food Systems of Russian Academy of Sciences, Moscow, Russia

² Scientific Center of Biomedical Technology of the Federal Medical-Biological Agency of Russia, Moscow region, Russia

Key words: proteomics, two-dimensional electrophoresis, biomarkers, mass spectrometry

Abstract

One of the ways to address the food quality issues facing the industry is the development of standardized and certified methods related to the conduct of in-depth studies of biochemical indicators of quality and safety of meat and meat products. The world laboratory practice in the field of food quality and safety shows a constant expansion of the list of controlled indicators of food raw materials and products. An important feature of the modern period in the development of biomedical and biotechnological research is the introduction of a whole complex of postgenomic technologies, which are based on a systematic approach to the study

ДЛЯ ЦИТИРОВАНИЯ: Вострикова Н.Л., Чернуха И.М., Хвостов Д.В. Методологические аспекты идентификации тканеспецифичных белков и пептидов, формирующих корригирующие свойства инновационных мясных продуктов. Теория и практика переработки мяса. 2018;3(3): 36-55. DOI 10.21323/2414-438X-2018-3-3-36-55

FOR CITATION: Vostrikova N.L., Chernukha I.M., Khvostov D.V. Methodological aspects of identification of tissue-specific proteins and peptides forming the corrective properties of innovative meat products. Theory and practice of meat processing. 2018;3(3): 36-55 (In Russ.). DOI 10.21323/2414-438X-2018-3-3-36-55

of the functioning of the mammalian proteome in various physiological and pathological conditions, including the formation and development of alimentary-dependent pathologies. In this regard, the problem of multilateral study of food products, in particular their identification, is the most relevant, because the modern technology of their production has undergone significant changes and requires the development of “gentle” processing modes. They concern raw materials and auxiliary materials used at all stages of production. This and new technologies of production of protein products from plant raw materials, as well as the introduction of food raw materials and food additives of artificial origin and the excess introduction of additives of plant and animal origin can cause falsification of products, as well as affect the health of the consumer. Food quality assessment includes the control of components of finished products. It is most difficult to determine the proportion of muscle protein in multi-component meat products that have undergone heat treatment. Therefore, in practice, when assessing the quality of food products, there is a need to identify its real composition in accordance with the declared normative documents. Currently, a promising area of research in the field of determining the composition of finished food is the selection of biomarkers of various components. Therefore, it is important to develop a methodology for the identification of biochemical changes in food raw materials under the influence of technological factors using modern research methods. This paper provides an overview of the protein and peptide analysis methodology, including the latest technologies that are becoming increasingly important.

Введение

Белки представляют собой крупные биомолекулы или макромолекулы, состоящие из одной или нескольких длинных цепей аминокислотных остатков. Белки выполняют широкий спектр функций внутри организмов, включая катализирующие метаболические реакции, репликацию ДНК, реакцию на раздражители и перенос молекул из одного места в другое. Белки отличаются друг от друга в первую очередь последовательностью аминокислот, которая продиктована нуклеотидной последовательностью их генов. Как правило, это приводит к сгибанию белка в специфическую трехмерную структуру, которая определяет его активность [1]. Линейная цепь аминокислотных остатков называется полипептидом. Белок содержит, по меньшей мере, один длинный полипептид. Короткие полипептиды, содержащие менее 20–30 остатков, редко считаются белками и обычно называются пептидами или иногда олигопептидами. Отдельные аминокислотные остатки соединены вместе пептидными связями и соседними аминокислотными остатками. Последовательность аминокислотных остатков в белке определяется последовательностью гена, который кодируется в генетическом коде. В общем, генетический код указывает 20 стандартных аминокислот; однако в некоторых организмах генетический код может включать селеноцистеин и в некоторых пирролизин. Вскоре после или даже во время синтеза часто происходят посттрансляционные модификации остатков в белке, которые изменяют физические и химические свойства, складывание, стабильность, активность и, в конечном счете, функцию белков [2]. Иногда белки содержат не пептидные группы, которые можно назвать протезными группами или кофакторами. Белки могут также работать вместе для достижения определенной функции, и они часто ассоциируются с образованием стабильных белковых комплексов [3,4].

После образования белки существуют только в течение определенного периода времени, а затем деградируют и перерабатываются аппаратом клетки через процесс белкового цикла. Продолжительность жизни

белка измеряется периодом его полураспада и охватывает широкий диапазон. Они могут существовать в течение нескольких минут или лет со средней продолжительностью жизни 1–2 месяца в клетках млекопитающих. Аномальные или несогласованные белки деградируют быстрее благодаря тому, что они подвергаются разрушению, либо из-за неустойчивости [5].

Как и другие биологические макромолекулы, такие как полисахариды и нуклеиновые кислоты, белки являются важными частями организмов и участвуют практически во всех процессах в клетках. Многие белки являются ферментами, которые катализируют биохимические реакции и являются жизненно важными для метаболизма. Белки также имеют структурные или механические функции: актин и миозин в мышцах, белки в цитоскелете, которые образуют систему микрофиламентов, поддерживающих форму клеток. Другие белки важны в клеточной передаче сигналов, иммунных реакциях, клеточной адгезии и клеточном цикле. У животных в рационе необходимы белки, для обеспечения незаменимыми аминокислотами, которые невозможно синтезировать. Пищеварение разрушает белки для использования в метаболизме [6].

Чтобы выполнить свою функцию, белки взаимодействуют с другими веществами с другими веществами на молекулярной или ионном уровне или другими белками. Белки взаимодействуют в различных контекстах и с разными результатами. Некоторые белки активируют или дезактивируют другие белки путем связывания с ними или путем (де-) фосфорилирования их. В процессе фосфорилирования, фосфатную группу добавляют (или удаляют) из белка, что активирует или дезактивирует белок. Другие белки связываются друг с другом, создавая так называемые белковые комплексы. Они имеют важные роли во всей клетке, например, в репликации ДНК. Другой класс белков связывается друг с другом для создания структурных комплексов, которые передают клетке свою трехмерную структуру. Все эти взаимодействия, являются центральной частью функционирования клетки. Понимание этих основных взаимодействий имеет жизненно важное

значение для понимания внутренней работы жизнедеятельности организма.

Тканеспецифическая экспрессия генов может приводить к наличию или отсутствию определенных белковых взаимодействий и комплексов, что приводит к глубоким функциональным различиям биологических процессов в тканях [7].

Понимание характера этих взаимодействий важно для изучения того, как и почему работают клетки. Взаимодействия между всеми белками в клетке составляют так называемое белково-белковое взаимодействие (Protein-protein interactions-PPI) сети. Не все белки присутствуют во всех клетках и типах тканей, поэтому белковые взаимодействия ограничены клеточными и тканевыми типами, где оба взаимодействующих белка присутствуют. Эти тканезависимые взаимодействия образуют тканеспецифические PPI (TSPPI) [8].

С этой целью изучают методологические аспекты анализа и выбирают, или разрабатывают эффективные алгоритмы, которые показывают наиболее достоверные результаты. Кроме того, изучают основные свойства TSPPI, чтобы получить представление о структуре взаимодействий, а также свойства определенных групп белков. Наконец, используют алгоритмы кластеризации с целью выявления тканеспецифических функциональных модулей в рамках TSPPIs [9].

Значительный интерес представляет изучение механизмов образования веществ белковой и пептидной природы, обуславливающих биокорректирующие и качественные характеристики путей их биосинтеза [10].

Все больше внимания уделяется биоинформатике как инструменту изучения протеома, с точки зрения гипотетического наличия в нем, тех или иных биологически активных пептидов и белков-маркеров. Согласно мировым базам данных в мышечных белках продуктивных животных содержатся аминокислотные последовательности, обладающие различными биологическими свойствами [11].

Основная часть

Подходы к изучению белков

Активность и структуры белков можно исследовать *in vitro*, *in vivo* и *in silico*. Исследования *in vitro* очищенных белков в контролируемых средах полезны для изучения того, как белок выполняет свою функцию: например, исследования кинетики ферментов, выявляют химический механизм каталитической активности фермента и его относительное сродство к различным возможным молекулам субстрата. Напротив, эксперименты *in vivo* могут предоставить информацию о физиологической роли белка в контексте клетки или даже всего организма. В исследованиях *in silico* используются вычислительные методы для изучения белков.

Изучение белков *in vivo* часто связано с синтезом и локализацией белка в клетке. Очень часто не ясна специфичность того, как белки нацелены на определенные органеллы или клеточные структуры, хотя известно о многих синтезированных в цитоплазме внутриклеточных белках, связанных с мембраной или секретируемых белках в эндоплазматическом ретикулуме [4]. Полезный метод оценки клеточной локализации использует генную инженерию для экспрессии в клетке слитого белка или химеры, состоящей из природного белка, представляющего интерес, связанного с «репортером», таким как зеленый флуоресцентный белок (GFP) [12]. Положение слитого белка в клетке может быть чисто и эффективно визуализировано с использованием микроскопии [13].

Существуют экспериментальные методы, например, для обнаружения, выделения и очистки белков, а также для расшифровки структуры и функций, часто требуется очистки исходного белка. Вычислительные методы обычно используют компьютерные программы для анализа белков. Однако многие экспериментальные методы (например, масс-спектрометрия) требуют вычислительного анализа необработанных данных.

Генетические методы

Экспериментальный анализ обычно требует экспрессии и очистки белков. Выражение достигается путем манипулирования ДНК, которая кодирует, представляющий интерес белок. Следовательно, анализ белка обычно требует методов ДНК, особенно клонирования. Некоторые белки никогда не были непосредственно секвенированы, однако, путем перевода кодонов из известных последовательностей мРНК в аминокислоты с помощью метода, известного как «концептуальная трансляция», было проанализировано большинство белковых молекул. Сайт-направленный мутагенез избирательно вводит мутации, которые изменяют структуру белка. Функция частей белка может быть лучше понята путем изучения изменения фенотипа в результате этого воздействия. Белковые спайки получают путем вставки протеиновых меток, таких как His-tag, для получения модифицированного белка, который легче отслеживать. Примером этого может служить GFP-Snf2H меченых комплексов (хроматин-ремоделирование), который состоит из белка, связанного с зеленой флуоресцентной меткой [14].

Анализ ДНК-аллелей может быть идентифицирован как связанный с болезненными состояниями, например, при вычислении оценки LOD, разработанной Ньютоном Мортонем. Она представляет собой статистический тест, который часто используется для анализа сцепления в популяциях человека, животных и растений. Оценка LOD сравнивает вероятность получения тестовых данных, если эти два локуса действительно связаны, с вероятностью наблюдения

случайных событий. Позитивные оценки LOD благоприятствуют наличию связи, тогда как отрицательные оценки LOD указывают на то, что привязка менее вероятна. Компьютеризированный анализ оценки LOD — это простой способ анализа сложных семейных родословных, чтобы определить связь между менделевскими чертами (или между признаком и маркером, или двумя маркерами) [15].

Существуют и другие возможности. Например, иммуногистохимия обычно использует антитело к одному или нескольким представляющим интерес белкам, которые конъюгированы с ферментами, давая либо люминесцентные, либо хромогенные сигналы, которые можно сравнить между образцами, что позволяет получить информацию о локализации. Другим применимым методом является кофракционирование в градиентах сахарозы (или другого материала) с использованием изопикнического центрифугирования [16].

Через другое применение в области генной инженерии, известное как сайт-направленный мутагенез, исследователи могут изменять последовательность белка и, следовательно, его структуру, клеточную локализацию и восприимчивость к регуляции. Этот метод даже позволяет включить неестественные аминокислоты в белке с использованием модифицированных тРНК [17] и рациональный дизайн новых белков с новыми свойствами [18].

Протеомика

К основным экспериментальным методам протеомики относятся двумерный электрофорез [19], который дает возможность разделять большое количество белков, масс-спектрометрию, что позволяет быстро идентифицировать белки и секвенировать пептиды с высокой пропускной способностью (чаще всего после переваривания в геле), белковые микрочипы, которые позволяют обнаруживать относительные уровни большого количества белков, присутствующих в клетке, и двухгибридный скрининг, который позволяет систематически исследовать белок-белковые взаимодействия. Систематическая попытка определить структуры белков, известна как структурная протеомика [20,21].

Биоинформатика

Для анализа структуры, функций и эволюции белков был разработан широкий спектр вычислительных методов.

Развитие таких инструментов было обусловлено большим количеством геномных и протеомных данных, доступных для различных организмов, включая человеческий геном. Экспериментально изучать все белки просто невозможно, поэтому лишь немногие из них подвергаются лабораторным экспериментам, а вычислительные инструменты используются для экстраполяции аналогичных белков. Такие гомоло-

гичные белки могут быть эффективно идентифицированы в отдаленных родственниках путем выравнивания последовательностей. Геномные и генные последовательности могут быть различными инструментами для определенных свойств. Инструменты профилирования последовательности могут найти сайты ферментов рестрикции и предсказать вторичные структуры. Филогенетические «деревья» могут быть построены, и эволюционные гипотезы развиваются с использованием специального программного обеспечения, такого как ClustalW, относительно происхождения современных организмов и выражаемых ими генов. Поле биоинформатики теперь незаменимо для анализа генов и белков.

Определение структуры

Обнаружение третичной структуры белка или четвертичной структуры его комплексов может дать важные сведения о том, как белок выполняет свою функцию. Общие экспериментальные методы определения структуры включают рентгеновскую кристаллографию и ЯМР-спектроскопию, которые могут генерировать информацию при атомном разрешении. Тем не менее, эксперименты ЯМР могут предоставить информацию, из которой можно оценить подмножество расстояний между парами атомов, а окончательные возможные конформации белка определяются путем решения проблемы геометрии расстояния. Двойная поляризационная интерферометрия является количественным аналитическим методом для измерения общей конформации белка и конформационных изменений из-за взаимодействий или другого стимула. Циркулярный дихроизм — еще один лабораторный метод определения внутреннего β -листового/ α -спирального состава белков. Криоэлектронная микроскопия используется для получения структурной информации меньшего разрешения об очень больших белковых комплексах, включая собранные вирусы, вариант, известный как электронная кристаллография, также может давать информацию с высоким разрешением в некоторых случаях, особенно для двумерных кристаллов мембранных белков. Разрешенные структуры обычно депонируются в Банке данных о белках (PDB), доступном ресурсе, из которого можно получить структурные данные о тысячах белков в виде декартовых координат для каждого атома в белке.

Методы прогнозирования структуры белка пытаются обеспечить способ создания правдоподобной структуры для белков, структуры которых экспериментально не определены [22,23].

Хроматографические методы

Наиболее универсальным методом определения чистоты пептидов и белков является обращенно-фазовая хроматография (ОФ ВЭЖХ), основанная на использовании широкого спектра гидрофобных не-

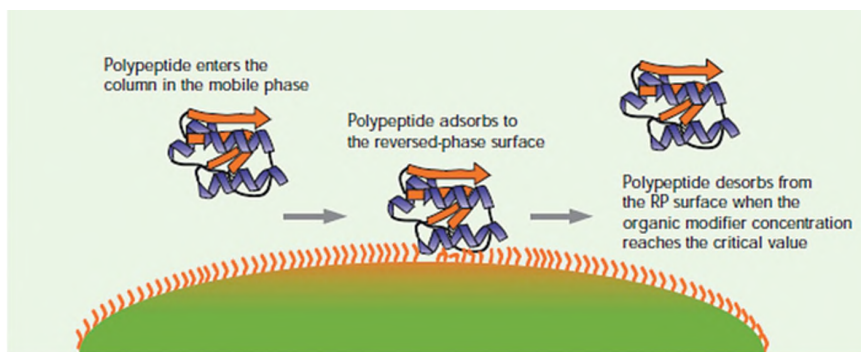


Рис. 1. «Гидрофобная нога» полипептида адсорбируется на гидрофобную поверхность материала с обращенной фазой, где он остается до тех пор, пока концентрация органического модификатора не повысится до критической концентрации [27].

подвижных фаз [24]. Для разделения компонентов лекарственных средств пептидной и белковой природы чаще остальных используются обращенные фазы с использованием силикагеля с привитыми алкильными группами C8 и C18, со средними размерами пор 300 Å. Элюирование лекарственных средств пептидной или белковой структуры или их фрагментов, полученных в результате ферментативного расщепления, в ОФ ВЭЖХ, в основном проводится в режиме градиентного элюирования при pH=2–3 при комнатной температуре, поскольку при повышении температуры возможна частичная или полная денатурация белка, что также влияет на хроматографическую подвижность [25,26].

Понимание механизма взаимодействия полипептидов с поверхностью с обращенной фазой важно для понимания разделения их RP-HPLC. Разделение малых молекул связано с непрерывным разделением их между подвижной и гидрофобной неподвижной фазами. Однако полипептиды слишком велики для разделения на гидрофобной фазе; они адсорбируются на гидрофобной поверхности после входа в колонку и остаются адсорбированными до тех пор, пока концентрация органического растворителя не достигнет критической концентрации, необходимой для десорбции (Рис. 1). Затем они десорбируются и, слегка взаимодействуя с поверхностью, сходят с колонки [27].

Полипептиды можно рассматривать как «сидящие» на неподвижной фазе, причем большая часть молекулы подвергается воздействию подвижной фазы и только часть молекулы — «гидрофобный хвост» — в контакте с обращенной фазой. ОФ-ВЭЖХ отделяет полипептиды на основе тонких различий в «гидрофобном хвосте» отделяемого полипептида. Различия в «гидрофобном хвосте» обусловлены различиями в аминокислотных последовательностях и различиями в конформации [27].

Методы масс-спектрометрии

В настоящее время высокоэффективная жидкостная хроматография в сочетании с масс-спектрометрией (МС) является одним из самых распространенных методов в исследованиях пептидов и белков [28,29].

Недавние успехи иллюстрируют роль протеомики на основе масс-спектрометрии как незаменимого инструмента для систем молекулярной и клеточной биологии и для новых областей. К ним относятся изучение белково-белковых взаимодействий с помощью аффинно-зависимых изоляций в малом и широко распространенном масштабе, картирование многочисленных органелл, параллельное описание генома паразита малярии и формирование количественных профилей белков у разных видов. Можно ожидать, что способность масс-спектрометрии идентифицировать и, все чаще, количественно определять тысячи белков из сложных образцов, может широко влиять на биологию и медицину.

Масс-спектрометрические измерения проводятся в газовой фазе на заряженных ионах. По определению, масс-спектрометр состоит из источника ионов, масс-анализатора, который измеряет отношение массы к заряду (m/z) ионизированных аналитов и детектор, который регистрирует количество ионов при каждом значении m/z .

В настоящее время в исследованиях протеомики используются четыре основных типа массового анализатора. Это анализаторы ионной ловушки, времени пролета (TOF), квадрупольного и Фурье-ионного циклотронного (FT-MS). Они отличаются друг от друга по дизайну и производительности, каждый из которых обладает собственной силой и слабостью. Эти анализаторы могут быть автономными или, в некоторых случаях, объединены в тандеме, чтобы воспользоваться преимуществами каждого из них [30].

Для достижения разделения масс могут применяться три разных принципа: разделение на основе времени пролета (TOF MS), разделение квадрупольными электрическими полями, генерируемыми металлическими стержнями (квадрупольная МС), или разделение путем селективного выброса ионов из трехмерного поля захвата (ионная ловушка или электронно-циклотронного резонанса Фурье). Для структурного анализа, такого как пептидное секвенирование, проводят две стадии масс-спектрометрии в тандеме (тандемная масс-спектрометрия или MS/MS), что может быть выполнено с использованием одного и того же принципа

разделения дважды или путем объединения двух разных принципов разделения MS. И MALDI, и электрораспыление могут быть связаны с любым из этих трех методов разделения. Тот факт, что MALDI производит короткие всплески ионов в вакууме, а электрораспыление создает непрерывный пучок ионов в атмосфере, как правило, приводит к соединению MALDI с TOF MS и электрораспыление с квадруполем и ионно-захватывающей ловушкой [29].

Идентификация пептидов и белков

Одной из важнейших задач определения качества лекарственных средств пептидной и белковой структуры является идентификация белков. Белки идентифицируются главным образом по их аминокислотной последовательности. При масс-спектрометрическом анализе пептидов и белков различают три подхода.

Наиболее широко применяемым методом по-прежнему является протеомика «снизу вверх», где белок переваривается в пептиды, которые могут быть эффективно проанализированы с использованием широкого спектра инструментов LC-MS или MALDI-TOF-MS. Подготовка образца для экспериментов по методу «снизу вверх» требует нескольких трудоемких этапов перехода от белка к пептидному уровню. Наиболее важным шагом в таких подходах является переваривание белка, которое часто является узким местом, так как занимает много времени. Следовательно, значительное увеличение пропускной способности может быть достигнуто путем ускорения процесса пищеварения. Современные методы позволяют сократить время переваривания от ночной инкубации (~ 15 часов) до минут или даже секунд. Это продвижение также делает возможным интеграцию в онлайн-системы, тем самым уменьшая количество утомительных шагов обработки образцов и риск потери образца [31]. В этом разделе дается обзор имеющихся в настоящее время стратегий пищеварения и последних изменений в ускорении процесса пищеварения.

Переваривание белка, как ферментативно, так и неферментативно, является важным и (почти) незаменимым инструментом идентификации, характеристики и количественной оценки протеина с помощью стратегий протеомики [32]. Протеомика играет жизненно важную роль в основных областях исследований, включая обнаружение биомаркеров болезни и биосистем, и, как таковая, значительно способствует пониманию биологических процессов, которые необходимы для жизни [33].

Выбор подхода протеомики должен основываться на типе вопроса, на который нужно ответить. Глобальная протеомика, такая как нахождение биомаркера в очень сложном образце, касается идентификации часто малочисленного неизвестного белка, который присутствует в сложном образце с широким динамическим диапазоном белков. Для этого требуется совсем другой подход, чем целевой анализ белка, например, характеристика состояний посттрансляционной модификации (ПТМ), где требуется подробное исследование и полное картирование последовательности известного белка для локализации, возможно, малообеспеченных ПТМ. Кроме того, (классы) белков со специальными характеристиками, могут требовать особой подготовки для них.

Подходы в протеомике можно различать по уровню, на котором проводится анализ (Рис. 2). Достижения в области масс-спектрометрии (МС) теперь позволяют напрямую анализировать белки. В таком эксперименте «сверху вниз» очищенные белки детектируются интактными и после фрагментации с использованием ударно-активированной диссоциации (CAD), диссоциации электрон-захвата (ECD) или диссоциации переноса электронов (ETD), обеспечивая информацию об интактной массе белка и аминокислотной последовательности [34,35]. Анализ нижних нитей интактных белков сводит к минимуму подготовку проб и сохраняет информацию, которая иногда теряется в других стратегиях протеомики,

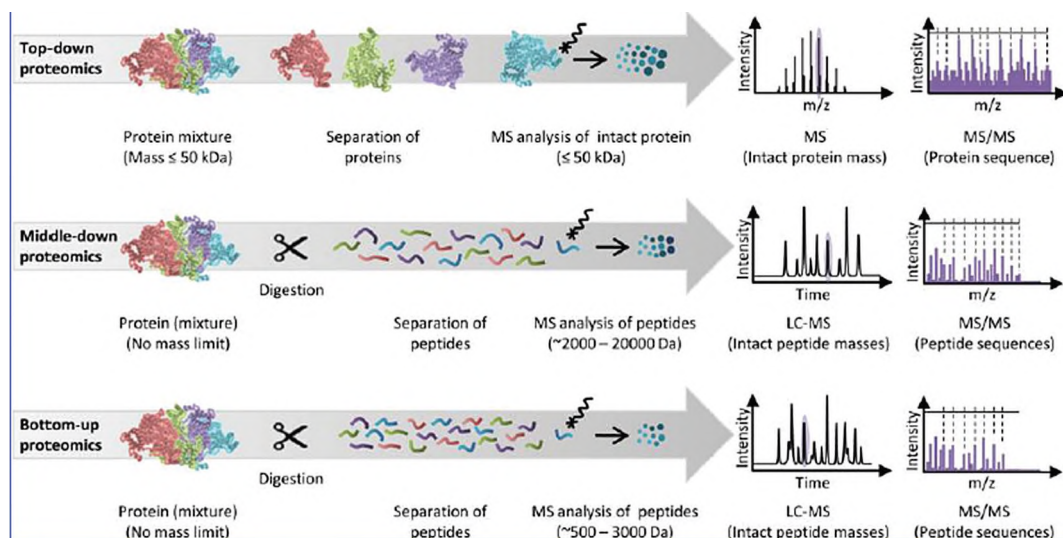


Рис. 2. Обзор подходов протеомики [31]

таких как связывание нескольких ПТМ, но относительно нечувствительна [36]. Из-за большого размера аналитов требования к приборам MS с точки зрения разрешения и массовой точности осуществляются только высокопроизводительными масс-спектрометрами, такими как ионно-циклотронный резонанс Фурье MS (FTICR-MS) и Orbitrap MS [37]. Тем не менее, интактный анализ белков также проводился с использованием более доступных инструментов, таких как тандемный квадруполь и квадрупольный времяпролетный масс детектор (QTOF) MS. На практике массовый диапазон белков, который может быть проанализирован с использованием протеомики «сверху вниз», ограничен до ~ 50 кДа, таким образом, примерно 500 аминокислот [38]. В противном случае секвенируются только C- и N-концы [39,40]. Несмотря на очевидные преимущества протеомики «сверху вниз», необходима дальнейшая разработка инструментария MS, прежде чем она станет основной технологией.

подавляющее большинство экспериментов с протеомикой предполагает процедуру переваривания белка в пептиды перед анализом MS. Анализ пептидов имеет несколько преимуществ перед белками, включая более эффективное разделение жидкостной хроматографией (LC), более низкую молекулярную массу и меньшее количество зарядовых состояний, что приводит к улучшению чувствительности [41]. В зависимости от размера полученных пептидов подход может называться «снизу вверх», либо «с середины вниз». В стратегии «снизу вверх» белок гидролизуются в пептиды в пределах ~ 500–3000 Да. Эти пептиды затем анализируют с помощью жидкостной хроматографии с электрораспылительной ионизацией MS (LC-ESI-MS) или методом лазерной десорбционной ионизации с времяпролетным MS (MALDI-TOF-MS). Идентификация белка выполняется на основе анализа отпечатков пальцев в пептидной массе или пептидной последовательности [34]. Секвенирование пептидов может быть эффективно достигнуто путем диссоциации, вызванной столкновением (CID), в более широкодоступных масс-спектрометрах ESI-ион-ловушки и ESI-QTOF или с использованием MALDI-TOF / TOF приборов [30].

Переваривание сложного образца белка, например целого протеома с «снизу вверх» подходом, дает огромное количество пептидов, и даже больше, чем может анализировать даже самый эффективный инструмент. Сложность набора пептидов может быть уменьшена без ущерба для информационного содержания за счет получения меньшего количества, но больших по размеру пептидов. Это, так называемый, средний уровень протеомного подхода, сочетает в себе лучшее из нисходящего и восходящего, используя преимущества усовершенствованной измерительной аппаратуры MS и доступность методов фрагментации на основе

электронов [42], сохраняя при этом уровень чувствительности, связанный с анализ пептидов [37]. Пептиды среднего уровня (~ 3000–20000 Да) показывают улучшенное разделение LC, и после ESI несут большее количество зарядов, что увеличивает фрагментацию CID [43], ETD [44] или ECD [45] в Orbitrap MS, квадрупольный линейной ион ловушки (QTrap) MS и квадрупольные приборы FTICR-MS. По сравнению с меньшими пептидами получают более уверенную идентификацию пептидов, что приводит к улучшению охвата последовательности белка и идентификации ПТМ [46].

Предсказуемая природа фрагментации пептидов позволяет сопоставлять экспериментальные спектры MS / MS с предсказанными спектрами при *in silico* известных белковых последовательностей для установления идентичности белка [47]. В этом отношении передовые средства биоинформатики играют ключевую роль в любых стратегиях протеомики. Однако эти средства основаны на предположении, что процесс переваривания (включая восстановление и алкилирование дисульфидных мостиков) является оптимальным. Если это не так, например, когда пептиды связаны через неповрежденный дисульфидный мостик, генерируются непредсказуемые пептиды и / или спектры сложной фрагментации, которые не будут идентифицироваться в автоматическом поиске базы данных. Переваривание белков может привести к потере информации, например, наличию и связности ПТМ [39] или способности отличать близкородственные белки из-за сбоя в обнаружении определенных частей белковой последовательности из-за неадекватного размера или неблагоприятных ионизационных свойств генерируемых пептидов. Наконец, качество полученных идентификаторов и модификаций белка можно контролировать с помощью ложных скоростей обнаружения, пороговых значений белка и пептида, но для этого требуется критический обзор полученных результатов.

Переваривание белка является важным шагом в стратегии протеомики «снизу вверх» и «с середины вниз» и оказывает значительное влияние на качество идентификации белка [48]. На протяжении многих лет переваривание белка улучшалось за счет разработки новых методов для увеличения пропускной способности и воспроизводимости [31].

Базы данных белков

На практике пептиды идентифицируют, как правило, используя системы поиска по базам данных белков, например Mascot [49], SEQUEST [50], X! Tandem [51], Inspect [52], OMSSA [53], MassMatrix [54], Crux [55], MyriMatch [56], MS-GFDB [57] и др.

Наиболее распространенными поисковыми системами являются первые три из перечисленного выше списка.

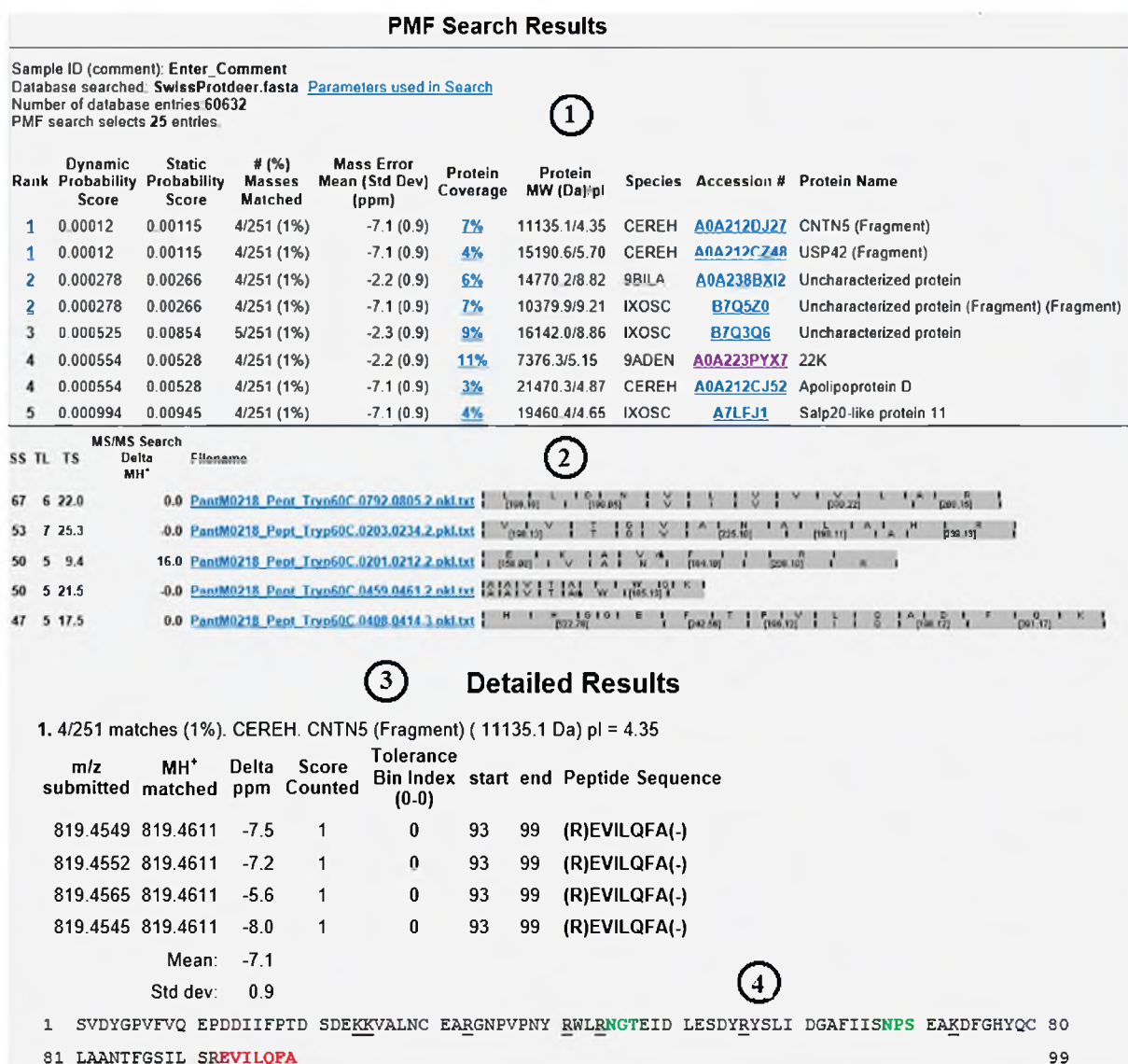


Рис. 3. Идентификация белка по международной базе данных Национального центра биотехнологической информации США (NCBI) (программа Mascot«MatrixScience», США) [59]. (1 — список потенциальных белков-кандидатов; 2 — расшифровка триптических пептидов с определением числа совпадений; 3 — распределение выявленных пептидов по аминокислотной последовательности; 4 — итоговый результат)

Поисковая система Mascot основана на алгоритме MOWSE (MOlecular Weight Search), предложенном в 1993 году Paprin с соавт.[58].

Данный алгоритм использует поиск по массовым «отпечаткам пальцев» пептидов. Вначале проводится сравнение масс пептидов из базы данных с экспериментальными данными масс пептидов с учетом заданной погрешности. Затем для каждого совпадения рассчитывается величина Score (так называемая величина уровня достоверности) в соответствии с формулой:

$$\text{Score} = \frac{5000}{M_{\text{prot}} \times \prod_n m_{i,j}}$$

где M_{prot} — молекулярная масса каждого совпавшего белка, \prod — произведение, которое рассчитывается из Mowse-матрицы весов, $m_{i,j}$ — для каждого совпадения экспериментальных данных и масс пептидов, рассчитанных из записей в геномной базе данных.

Данный алгоритм можно применять для поиска МС/МС. В этом случае в формуле для Score роль белка выполняет пептид, а роль пептида — фрагмент. Сумма Score пептидов дает Score для белка. Также, для каждого из кандидатов, указаны видовая принадлежность, что может стать решающим при интерпретации, и ссылки на персональные страницы (итоговый результат), содержащие исчерпывающую информацию о потенциальном белке (значения его молекулярной массы и изоэлектрической точки, расшифровка последовательности триптических пептидов, число совпадений,% покрытия полной аминокислотной последовательности белка выявленными пептидами и т.д.) [60,61]. Данный алгоритм может быть применен для МС/МС-поиска.

Алгоритм оценки основан на вероятности, который имеет ряд преимуществ: (а) Простое правило может использоваться для оценки значимости результата или нет. Это особенно полезно для защиты

от ложных срабатываний. (b) Оценки можно сравнить с результатами других видов поиска, такими как гомология последовательности. (c) Параметры поиска могут быть легко оптимизированы путем итерации [62].

Поисковая система Sequest основана на отдельной идентификации каждого масс-спектра [50]. Ознакомиться с ресурсом можно на сайте [63,64]. В этом методе нуклеотидные базы данных транслируются в шести кадрах считывания, и полученные аминокислотные последовательности обыскиваются «на ходу» для идентификации и подгонки линейных последовательностей к шаблонам фрагментации, наблюдаемым в тандемных масс-спектрах пептидов. Затем используется функция взаимной корреляции для обеспечения измерения сходства между отношениями массы к заряду для ионов фрагментов, предсказанных аминокислотными последовательностями, переведенными из базы нуклеотидов, и ионами фрагментов, наблюдаемыми в тандемном спектре масс. В общем случае разница между 0,1 и 2 между нормированными функциями взаимной корреляции для результатов поиска первого и второго ранжирования указывает на успешное совпадение между последовательностью и спектром [45].

Поисковая система X! Tandem наиболее развита, так как является программным обеспечением с открытым исходным кодом. Ознакомиться с ресурсом можно на сайте [65].

TANDEM был написан для запуска из командной строки с именем входного XML-файла в качестве единственного параметра для командной строки. Код был создан с использованием набора классов, которые выполняют следующие задачи:

- (1) чтение файлов параметров ввода XML;
- (2) считывание последовательности белка из файлов FASTA;
- (3) чтение спектров MS / MS в общих форматах ASCII (DTA, PKL и Matrix Science);
- (4) состояние MS / MS спектров для удаления шума и общих артефактов;
- (5) обработка пептидных последовательностей с реагентами расщепления, посттрансляционными и химическими модификациями;
- (6) оценка последовательности пептидов;
- (7) создание файла вывода XML, в котором будут отображены лучшие скоринговые последовательности и некоторые статистические распределения, относящиеся к процессу подсчета очков [51].

Описанные алгоритмы не лишены недостатков. Например, Mascot присваивает высокие значения индекса коротким пептидам, которые не всегда являются уникальными для исследуемого белка, также при анализе протеолитических смесей пептидов в списке кандидатов присутствует множество последовательностей с близкими значениями Score, но программа окончательном рапорте оставляет только одну последовательность.

Недостатком алгоритма Sequest является высокая сложность вычислений и обобщений спектров. А X! Tandem, как вероятностный подход, в отдельных случаях не позволяет провести достоверную идентификацию. Краткое описание более 100 различных алгоритмов и пакетов программ обработки масс-спектрометрических данных по пептидам и белкам представлены на сайтах http://en.wikipedia.org/wiki/Mass_spectrometry_software и <http://www.ms-utils.org>. Использование баз данных для идентификации белков и пептидов позволяет расшифровывать масс-спектры сложных смесей за короткое время [66]. Почти все известные в настоящий момент аминокислотные последовательности белков и пептидов объединены в базы данных, которые находятся в открытом доступе в сети Интернет. Каждая из них имеет свой формат хранения данных, различную степень избыточности, взаимосвязи с родственными или аналогичными базами данных. Все базы данных можно разделить на пять типов. Первый тип — это архивные базы данных, в которых информация добавляется исследователями. К таким базам данных относятся GenBank, EMBL, PDB. Второй тип — это курируемые базы данных — содержание записей курируется специалистами, к ним относится, например, Swiss-Prot. Третий тип — автоматические базы данных, в них записи генерируются компьютерными программами, к ним относится, например, TrEMBL. Следующий тип — производные базы данных, которые пополняются за счет обработки данных из баз данных первых двух типов. Сюда входят SCOP, PFAM, GO и др. Пятый тип — интегрированные базы данных, которые объединяют информацию из различных баз. К такому типу относится ENTREZ [67]. Определение аминокислотной последовательности пептидов и белков без использования поисковых программ и баз данных называется de novo секвенированием. Такой подход применяется для идентификации не описанных ранее белков, при наличии неисследованных мутаций, пост- трансляционных модификаций и т.д. Применяемые алгоритмы de novo 36 секвенирования основаны на различных математических методах. Первые алгоритмы определения аминокислотной последовательности [68,69] представляли собой перебор всех возможных комбинаций аминокислот, составляющих массу родительского иона, фрагментацию которых сравнивали с экспериментальным масс-спектром. Очевидно, что погрешность измерения массы родительского иона влечет за собой увеличение количества соответствующих ему комбинаций.

Заключение

В настоящее время появляются новые лекарственные препараты белковой и пептидной природы, методы контроля конкретных белков в пищевой отрасли. Для определения сложных, многокомпонентных изменений, которые происходят в мясных продуктах, необ-

ходимо использовать методологические подходы, при которых возможно регистрировать сотни и тысячи белков. Одним из таких подходов является протеомика, которая позволяет идентифицировать и выявлять количественные и качественные изменения в белковом составе клеток и тканей. Достижения протеомики помогают исследователям решать задачи посттрансляционных модификаций, клеточной сигнализации, функциональной и структурной гомологии белков, определения уровня экспрессии генов. Наиболее значимым инструментом протеомики является изучение белковых карт тканей человека и животных.

Протеомные технологии представляются весьма перспективными и эффективными для выявления в мясных продуктах биохимических изменений, таких как изменения термостойчивых и видоспецифичных белков, способных стать соответствующими биомаркерами.

Introduction

Proteins are large biomolecules or macromolecules that consist of one or several long chains of amino acid residues. Proteins have a wide spectrum of functions inside organisms, including catalyzing metabolic reactions, DNA replication, a response to stimuli and transport of molecules from one place to another. Proteins are different from each other, primarily, by the amino acid sequence, which is determined by the nucleotide sequence of their genes. As a rule, this leads to protein folding into a specific three-dimensional structure, which determines its activity [1]. A linear chain of amino acid residues is called a polypeptide. A protein contains, at least, one long polypeptide. Short polypeptides that contain less than 20–30 residues are seldom recognized as proteins and usually called peptides or sometimes oligopeptides. The individual amino acid residues are linked together by the peptide bonds and adjacent amino acid residues. The sequence of amino acid residues in a protein is determined by a sequence of a gene that is encoded in the genetic code. In general, the genetic code specifies 20 standard amino acids; however, in some organisms the genetic code can include selenocysteine and in some organisms pyrrolysine. Shortly after or even during synthesis, post-translational modifications of the protein residues often occur, which change the physical and chemical properties, folding, stability, activity and, finally, the protein function [2]. Sometimes proteins contain non-peptide groups, which can be called prosthetic groups or co-factors. Proteins can also act together achieving a certain function and they are always associated with the development of stable protein complexes [3,4].

After formation, proteins exist only for a particular period of time, then they are degraded and processed by a cellular apparatus through a process of the protein cycle. The lifespan of proteins is measured by the period of their half-decay and comprises a wide range. They can exist for

В последние годы протеомика стала широко применяться в области биотехнологии. В работе, проделанной в ФГБНУ «ФНЦ пищевых систем им. В.М. Горбатова» РАН, с помощью протеомных технологий были разработаны методические подходы по идентификации белкового профиля мясных продуктов и в исследуемых образцах мяса и в специально выработанных колбасных изделиях, были определены тканеспецифичные белки, которые могут быть использованы как индивидуальные биомаркеры при контроле мясных изделий на соответствие заявленному составу. Также были зарегистрированы соевый и куриный белки, что является маркером фальсификации [70,71].

Выражение признательности

Исследование выполнено за счет гранта Российского научного фонда (проект № 16–16–10073).

several minutes or years with an average lifespan of 1–2 months in the mammalian cells. Abnormal or misfolded proteins degrade faster because of destruction or instability [5].

Like other biological macromolecules, such as polysaccharides and nucleic acids, proteins are crucial parts of organisms and participate almost in all processes in cells. Many proteins are enzymes that catalyze biochemical reactions and are essential for metabolism. Proteins also have the structural and mechanical functions: actin and myosin in muscles, proteins in the cytoskeleton, which form the system of microfilaments for maintenance of the cell shape. Other proteins play an important role in cell signaling, immune reactions, cell adhesion and the cell cycle. Proteins are necessary in animals' diets to provide the essential amino acids, which cannot be synthesized. Digestion disintegrates proteins to use in metabolism [6].

To perform their function, proteins interact with other substances on the molecular or ionic level, or with other proteins. Proteins interact in different contexts and with different results. Several proteins activate or deactivate other proteins by bonding with them or by their (de) phosphorylation. In the process of phosphorylation, the phosphate group is added (or removed) from a protein, which activates or deactivates a protein. Other proteins are linked with each other forming the so-called protein complexes. They play important roles in the whole cell, for example, in DNA replication. Other class of proteins is linked with each other to form structural complexes, which give a cell their three-dimensional structure. The understanding of these main interactions has a vital importance for the understanding of the internal processes of the vital activity of an organism.

Tissue-specific gene expression can lead to existence or absence of certain protein interactions and complexes resulting in deep functional differences of biological processes in tissues [7].

The understanding of the character of these interactions is important for studying how and why cells function. The interactions between all proteins in a cell comprise the so-called protein-protein interactions (PPIs) of a network. Not all proteins are present in all cells and tissue types; therefore, protein interactions are limited by cellular and tissue types, where both interacting proteins exist. These tissue-dependent interactions form tissue-specific PPIs (TSPPIs) [8].

To this end, the methodological aspects of analysis are studied and effective algorithms, which show the most significant results, are chosen or developed. In addition, the main properties of TSPPIs are studied to gain an insight about the structure of interactions and properties of particular protein groups. Finally, the algorithms of clusterization are used to reveal tissue-specific functional modules in the framework of TSPPIs [9].

It is particularly interesting to study the formation mechanisms of substances of protein and peptide nature that determine bio-corrective and quality characteristics of the pathways of their biosynthesis [10].

More and more attention is paid to bio-informatics as an instrument for studying proteome from the viewpoint of the hypothetical presence of certain biologically active peptides and marker proteins. According to the world databases, the muscle proteins of productive animals contain amino acid sequences that have various biological properties [11].

Main part

Approaches to studying proteins

The protein activity and structures can be studied in vitro, in vivo and in silico. In vitro investigations of purified proteins in the controlled environment are useful for studying how a protein executes its function: for example, studies of enzyme kinetics reveal the chemical mechanism of the catalytic activity of an enzyme and its relative affinity to different possible substrate molecules. On the contrary, in vivo experiments can give information about the physiological role of proteins in the context of a cell or even the whole organism. When using in silico investigations, the computational methods are used to study proteins.

In vivo investigations of proteins are often connected with protein synthesis and localization in a cell. The specifics of how proteins are targeted to specific organelles or cellular structures is often unclear, although it is known about many intracellular proteins synthesized in the cytoplasm and membrane-bound or secreted proteins in the endoplasmic reticulum [4]. A useful method for assessing cellular localization applies genetic engineering for expression in a cell of a fused protein or chimera that consists of the natural protein of interest linked with a «reporter», such as the green fluorescent protein (GFP) [12]. The fused protein location in a cell can be clearly and effectively visualized by microscopy [13].

There are experimental methods (for example, for detection, extraction and purification of proteins, as well as for deciphering the protein structure and functions, purification of the initial protein is often required). Computational methods usually use computer programs for protein analysis. However, many experimental methods (for example, mass-spectrometry) require computational analysis of raw data.

Genetic methods

Experimental analysis usually requires protein expression and purification. Expression is achieved by manipulating DNA, which encodes a protein of interest. Therefore, protein analysis usually requires the DNA methods especially cloning. Several proteins have never been directly sequenced; however, by translating codons from the known mRNA sequences into amino acids by the method known as «conceptual translation», the majority of protein molecules were analyzed. Site-directed mutagenesis differentially introduces mutations that change the protein structure. A function of protein parts can be better understood by analyzing changes in a phenotype as a result of this impact. Fusion proteins are obtained by inserting protein tags, such as His-tag to obtain a modified protein, which is easier to trace. An example of this can be Snf2H-GFP labeled complexes (chromatin-remodeling), which consist of protein linked with the green fluorescent label [14].

Analysis of DNA alleles can be identified as associated with diseases, for example, when calculating a LOD score developed by Newton Morton. It is a statistical test which is used for linkage analysis in human, animal and plant populations. The LOD score compares the likelihood of obtaining test data, if the two loci are really linked, with the likelihood of observing random events. Positive LOD scores favor the presence of linkage, while negative LOD scores indicate that linkage is less probable. Computerized analysis of a LOD score is a simple method for analysis of complex family pedigrees to detect a linkage between Mendelian traits (or between a trait and a marker, or two markers) [15].

There are also other possibilities. For example, immunohistochemistry usually uses an antibody to one or several proteins of interest, which are conjugated with enzymes giving either luminescent or chromogenic signals that can be compared between samples allowing obtaining information about localization. Another useful method is co-fractionation in gradients of sucrose (or other substance) with the use of isopycnic centrifugation [16].

Through another application in the field of genetic engineering, known as site-directed mutagenesis, the researchers can change the protein sequence and, therefore, its structure, cellular localization and susceptibility to regulation. This method even allows incorporation of non-natural amino acids into proteins using modified tRNA [17] and rational design of new proteins with new properties [18].

Proteomics

The main experimental methods of proteomics are two-dimensional electrophoresis [19], which makes it possible to separate a great quantity of proteins, mass-spectrometry, which allows rapid protein identification and high-throughput peptide sequencing (most often after digestion in a gel), protein microarrays, which enable detection of relative levels of many proteins that are present in a cell, and two-hybrid screening, which allows a systematic study of protein-protein interactions. The systematic endeavor to detect the protein structure is known as structural proteomics [20,21].

Bioinformatics

For analysis of the structure, functions and evolutions of proteins, a wide spectrum of computational methods was created.

The development of such instruments was conditioned by a large number of genomic and proteomic data available for different organisms including the human genome. It is impossible to study all proteins experimentally; thus, only several of them are subjected to laboratory experiments, and computational tools are used for extrapolation of similar proteins. These homologous proteins can be effectively identified in distant relatives by a sequence alignment. Genomic and gene sequences can be different tools for certain traits. Sequence profiling tools can find sites of restriction enzymes and predict secondary structures. Phylogenetic trees can be built, and evolution hypotheses are developed using the special software, such as ClustalW, regarding the origin of modern organisms and their expressed genes. The bioinformatics field is now irreplaceable for the gene and protein analysis.

Structure determination

Determination of the protein tertiary structure or the quaternary structure of its complexes can give important information about how a protein executes its function. Common experimental methods for structure determination include X-ray crystallography and NMR spectroscopy, which can generate information at atomic resolution. Nevertheless, NMR experiments can give information that allows assessing a subset of distances between atom pairs, and

the final possible protein conformations are determined by solving the distance geometry problem. Dual-polarization interferometry is a quantitative analytical method for measuring general conformation of proteins and conformational changes due to interactions or another stimulus. Circular dichroism is the other laboratory method for detection of the internal β -sheet/ α -helix protein composition. Electron cryomicroscopy is used for gaining structural information of lower resolution about very large protein complexes including assembled viruses, a variant known as electron crystallography can also give information with high resolution in some cases, especially, for two-dimensional crystals of membrane proteins. Determined protein structures are usually deposited in the ProteinData Bank (PDB), a freely accessible resource, from which structural data about thousands of proteins can be obtained as Cartesian coordinates for each atom in a protein.

Methods for prediction of a protein structure thrive to provide a way of creating credible structure for proteins, which structures are not determined experimentally [22,23].

Chromatographic methods

The most universal method for detection of peptide and protein purity is reversed-phase chromatography (RP-HPLC) based on the use of a broad spectrum of hydrophobic stationary phases [24]. To separate the components of the pharmaceutical preparations of peptide and protein nature, the reversed phases are most often employed with the use of the silica gel bonded to C8 and C18 alkyl groups with the average pore size of 300 Å. In RP-HPLC, elution of the pharmaceutical preparations of the peptide and protein nature or their fragments obtained as a result of the enzymatic degradation is usually performed in the mode of gradient elution at pH = 2 – 3 and a room temperature as partial or complete protein denaturation is possible when a temperature is increased, which also affects the chromatographic mobility [25,26].

The understanding of the mechanism of the polypeptide interaction with the reversed phase surface is important for the understanding of their separation in RP-HPLC. Separation of small molecules is connected with their continual partitioning between the mobile phase and the hy-

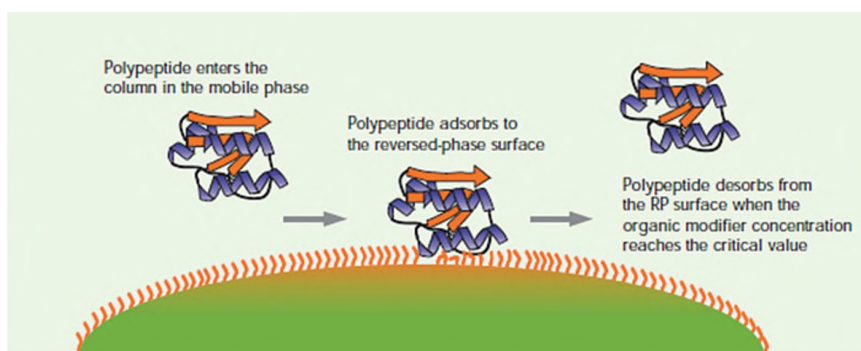


Figure 1. The «hydrophobic foot» of the polypeptide adsorbs to the hydrophobic surface of the reversed-phase material where it remains until the concentration of the organic modifier increases to the critical concentration [27]

drophobic stationary phase. However, polypeptides are too large for partition into the hydrophobic phase; they adsorb to the hydrophobic surface after they enter the column and remain adsorbed until the concentration of the organic modifier reaches the critical concentration necessary for desorption (Figure 1). Then, they desorb and, slightly interacting with the surface, elute down the column [27].

Polypeptides can be regarded as «sitting» on the stationary phase; with that, most of the molecule is exposed to the mobile phase and only a part of the molecule (the «hydrophobic foot») is in contact with the reversed-phase surface. RP-HPLC separates polypeptides based on the slight differences in the «hydrophobic foot» of the polypeptide being separated. The differences in the «hydrophobic foot» are conditioned by the differences in amino acid sequences as well as in conformation [27].

Mass spectrometry methods

At present, high performance liquid chromatography coupled with mass spectrometry (MS) is one of the most common methods for protein and peptide analysis [28,29].

Recent advances illustrate the role of proteomics based on the mass spectrometry as an irreplaceable instrument for systems of molecular and cellular biology, as well as for new fields. They include analysis of the protein-protein interactions through affinity-based isolations on a small and widespread scale, mapping of multiple organelles, simultaneous description of the genome of the malaria parasite and generation of the quantitative protein profiles in different species. It can be expected that the ability of mass spectrometry to identify and increasingly often to quantify thousands of proteins in complex samples can profoundly affect biology and medicine.

Mass-spectrometric measurements are carried out in the gas phase on the charged ions. By definition, a mass spectrometer consists of a ion source, a mass-analyzer, which measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector, which records the number of ions at each m/z value.

At present, in proteomic investigations, four main types of mass analyzers are used. These are the ion-trap, time-of-flight (TOF), quadrupole and Fourier transform ion cyclotron (FT-MS) analyzers. They differ from each other in design and performance and each of them has its own strengths and weaknesses. These analyzers can be automatic or, in some cases, combined in tandem to take advantage of each of them [30].

To achieve mass separation, three different principles can be used: separation based on time of flight (TOF MS), separation by quadrupole electric fields generated by metal rods (quadrupole MS) or separation by selective ejection of ions from a three-dimensional trapping field (the ion trap or Fourier transform ion cyclotron resonance). For structural analysis, such as peptide sequencing, two stages of mass spectrometry are carried out in tandem (tandem mass spectrometry or MS / MS), which can be achieved

using the same principle of separation twice or by combination of two different principles of MS separation. Both MALDI and electrospray can be coupled to any of these three methods of separation. The fact that MALDI produces short bursts of ions in the vacuum, and electrospray creates a continuous beam of ions in atmosphere, as a rule, leads to coupling MALDI to TOF MS and electrospray to quadrupole and ion trap [29].

Identification of peptides and proteins

One of the main tasks of quality assessment of pharmaceutical preparations having the peptide and protein structure is protein identification. Proteins are identified, mostly, by their amino acid sequence. There are three approaches to mass spectrometric analysis of peptides and proteins.

The most common method is still «bottom-up» proteomics, when a protein is digested into peptides, which can be effectively analyzed using a broad spectrum of LC-MS or MALDI-TOF-MS instruments. Sample preparation for an experiment by the «bottom-up» method requires several laborious stages of transition of proteins to the peptide level. The most important step in these approaches is protein digestion, which is always a bottleneck as it requires a lot of time. Therefore, a significant increase in the throughput can be achieved by acceleration of the digestion process. Modern methods allow reducing digestion duration from overnight incubation (~ 15 hours) to minutes or even seconds. This achievement also makes possible integration into on-line systems, thereby, reducing the number of tiresome steps for sample processing and a risk of sample loss [31]. This section gives an overview of the available digestion strategies and recent changes in acceleration of the digestion process.

Protein digestion (both enzymatic and non-enzymatic) is an important and (almost) irreplaceable tool for identification, characterization and quantitative assessment of proteins using proteomics strategy [32].

Proteomics plays a crucial role in the main fields of investigations, including detection of disease biomarkers and biosystems and, as such, significantly contributes to the understanding of the vital biological processes [33].

The choice of proteomics approach should be based on a type of the question, which is to be answered. Global proteomics, for example, finding a biomarker in a very complex sample, deals with the identification of an often low-abundant unknown protein, which is present in a complex sample with a broad dynamic range of proteins. To this end, it is necessary to use a completely different approach than targeted protein analysis, such as, characterization of the post-translational modification (PTM) states, which requires a detailed study and complete mapping of a known protein sequence for localization of the possibly low-abundant post-translational modifications (PTMs). In addition, protein (classes) with certain characteristics may need special preparation.

Approaches in proteomics can be distinguished by the level of the analysis (Fig.2). Achievements in the field of mass spectrometry now allow direct protein analysis. In this «top-down» experiment purified proteins are detected intact and after fragmentation with the use of collisional-activated dissociation (CAD), electron capture dissociation (ECD) and electron transfer dissociation (ETD) giving information about the intact protein mass and amino acid sequence [34,35]. Analysis of the intact proteins by the «top-down» method minimizes sample preparation and saves information which can be lost in other proteomics strategies, such as connectivity of several PTMs, but it is comparatively insensitive [36]. Since analytes have large sizes, the requirements to the MS equipment in terms of resolution and mass accuracy are met only by high throughput mass spectrometers, such as Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) and Orbitrap MS [37]. Nevertheless, intact protein analysis was also carried out using more available equipment such as tandem quadrupole and quadrupole time-of-flight mass spectrometer (QTOF MS). In practice, the protein mass range that can be analyzed using «top-down» proteomics is limited to ~ 50 kDa, therefore, about 500 amino acids [38]. Otherwise, only the C-terminal and N-terminal ends are sequenced [39,40]. Despite the obvious advantages of «top-down» proteomics, the further development of MS equipment is necessary before it will become the main technology.

The overwhelming majority of proteomic experiments envisage a procedure of protein digestion into peptides before MS analysis. Analysis of peptides has several advantages compared to proteins, including more effective separation by liquid chromatography (LC), lower molecular mass and lower number of charge states resulting in an increase in sensitivity [41]. Depending on the size of the obtained peptides, an approach can be called «bottom-up» or «middle-down». In the «bottom-up» strategy a protein is hydrolyzed to peptides in a range of ~ 500–3000 Da. Then, these peptides are analyzed by liquid

chromatography coupled to electrospray ionization MS (LC-ESI-MS) or matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry (MALDI-TOF MS). Protein identification is carried out based on peptide mass fingerprints or peptide sequences [34]. Peptide sequencing can be successfully performed by collision-induced dissociation (CID) in more available ESI-ion trap and ESI-QTOF mass-spectrometry or MALDI-TOF/TOF equipment [30].

Digestion of the complex protein sample, for example, the whole proteome using the «bottom-up» approach gives a large number of peptides and even more that can be analyzed by the most powerful instrument. The complexity of the peptide set can be reduced without loss of the information content due to production of lower quantity of larger peptides. This is the so-called «middle-down» proteomic approach, which combines the best of the «top-down» and «bottom-down» approaches using the advantages of the improved MS measuring equipment and availability of the fragmentation methods based on electrons [42], maintaining the same sensitivity level associated with the peptide analysis [37]. Peptides of the middle range (~ 3000–20000 Da) show improved separation by LC and after ESI carry a higher number of charges, which enhances fragmentation by CID [43], ETD [44] or ECD [45] in Orbitrap MS, quadrupole-linear ion trap (QTrap) MS and quadrupole-FTICR MS instruments. Compared to smaller peptides, more reliable peptide identification is obtained resulting in an improvement of protein sequence coverage and PTM identification [46].

The predictable nature of peptide fragmentation allows comparing the experimental MS / MS spectra with predicted spectra in *in silico* of known protein sequences for protein identification [47]. In this regard, the advanced tools of bioinformatics play a key role in all proteomics strategies. However, these tools are based on an assumption that the digestion process (including reduction and alkylation of disulfide bridges) is optimal. If it is not true, for example, when peptides are bond through the intact

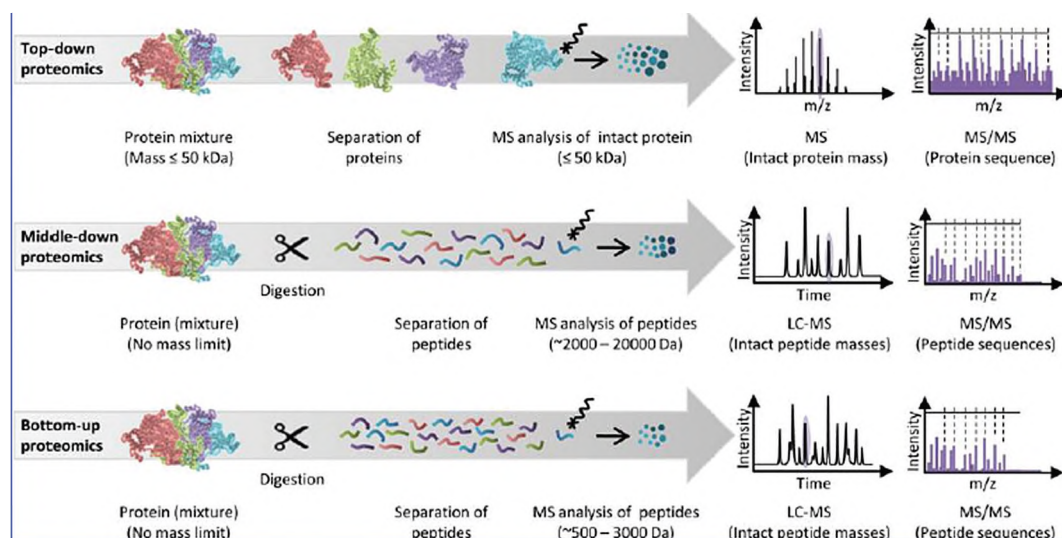


Figure 2. Overview of proteomics approaches [31]

disulfide bridge, the unpredictable peptides and/or complex fragmentation spectra are generated, which will not be identified in an automated database search. Protein digestion can lead to loss of information, for example, the PTM presence and connectivity [39] or an ability to distinguish closely related proteins due to a failure in detection of certain parts of a protein sequence because of the inadequate size or unfavorable ionization properties of the generated peptides. Finally, the quality of the obtained protein identifications and modifications can be controlled by false detection rates, protein and peptide score threshold; however, a critical review of the obtained results is necessary.

Protein digestion is an important step in the «bottom-up» and «middle-down» proteomics strategy and significantly affects protein identification quality [48]. For many years, protein digestion has been improved due to the development of new methods for increasing throughput and reproducibility [31].

Protein databases

In practice, peptides, as a rule, are identified by searching protein databases using the search systems, for example, Mascot [49], SEQUEST [50], X! Tandem [51], Inspect [52], OMSSA [53], MassMatrix [54], Crux [55], MyriMatch [56], MS-GFDB [57] and others.

The most common search systems are the first three of the above-mentioned list.

The search system Mascot is based on the MOWSE algorithm (MOlecular Weight Search), proposed by Pappin et al. in 1993 [58].

This algorithm uses the search by mass fingerprints of peptides. At first, the peptide masses from a database are compared with the experimental data on peptide masses with consideration for the specified error. Then, for each match, the Score value (the so-called confidence level) is calculated according to the equation:

$$\text{Score} = \frac{5000}{M_{\text{prot}} \times \prod_{i,j} m_{i,j}}$$

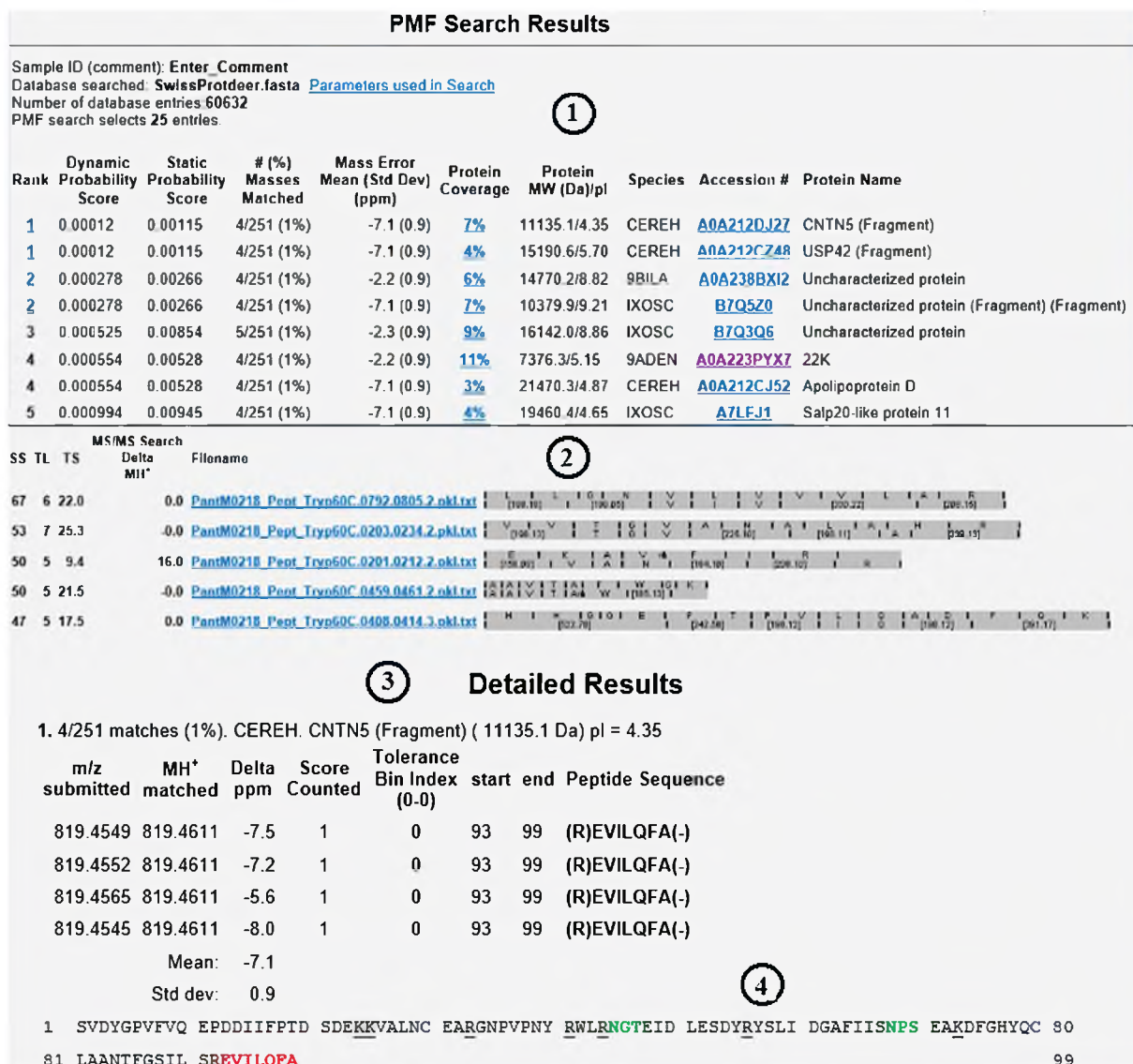


Figure 3. Protein identification by the international database of the the National Center for Biotechnology Information (NCBI) (USA) (software Mascot «MatrixScience», (USA) [59]). (1 — list of potential candidate proteins; 2 — decryption of tryptic peptides with a certain number of matches; 3 — distribution of the revealed peptides along the amino acid sequence; 4 — final result)

where M_{prot} — molecular mass of each matching protein,

Π_n — is a product, which is calculated from the Mowse-weight matrix, $m_{i,j}$ — for each matching between the experimental data and peptide masses calculated from the entries in the genomic database.

This algorithm can be used for an MS/MS search. In this case, in the equation for Score, a peptide plays a role of a protein, and a fragment plays a role of a peptide. The sum of the peptide scores gives the Protein Score. Also, for each candidate, the species origin is given, which can become decisive for interpretation, and references to the personal pages (final result) that contain the comprehensive information about a potential protein (values of its molecular mass and isoelectric point, decryption of a tryptic peptide, a number of matches, % of coverage of the complete amino acid sequence of a protein by peptides and so on) [60,61]. This algorithm can be used for an MS/MS search.

The assessment algorithm is based on probability, which has several advantages: (a) a simple rule can be used for assessment whether a result is significant or not. This is especially useful for protection from false responses. (b) Scores can be compared with the results of other types of search, such as sequence homology. (c) Search parameters can be easily optimized by iteration [62].

The Sequest search system is based on the individual identification of each mass spectrum [50]. This resource can be found on the site [63,64]. In this method, the nucleotide databases are translated in six reading frames and obtained amino acid sequences are searched during the process to identify and fit the linear sequences to the fragmentation templates found in the tandem mass-spectra of peptides. Then the mutual correlation function is used for measuring the similarity between the mass-to-charge ratio for the ions of fragments predicted by the amino acid sequences translated from the nucleotide database and the ions of fragments observed in the tandem mass spectrum. Generally, the difference of 0.1 to 2 between normed mutual correlation functions for the results of the search for the first and second ranking indicates a successful matching between a sequence and a spectrum [45].

The X! Tandem searching system is the most developed, as it is a software with an open source code. The information about the resource can be found on the site [65].

TANDEM was created for running from the command line with the name of the input XML file as a single parameter for the command line. The code was developed with the use of a set of classes that can accomplish the following tasks:

- (1) reading XML input parameter files;
- (2) reading protein sequences from FASTA files;
- (3) reading MS / MS spectra in common ASCII formats (DTA, PKL and Matrix Science);
- (4) conditioning MS / MS spectra to remove noise and common artifacts;
- (5) processing peptide sequences with the cleavage reagents, post-translational and chemical modifications;

(6) scoring peptide sequences;

(7) creating an XML output file that will reflect the best scoring sequences and several statistical distributions relevant to the scoring process [51].

The described algorithms are not without drawbacks. For example, Mascot assigns high values of the index to short peptides, which are not always unique for the studied protein; also, in analysis of the proteolytic peptide mixtures many sequences with close score values are present in the list of candidates; however, the program in the final report leaves only one sequence.

A disadvantage of the Sequest algorithm is the high complexity of calculations and generalizations of spectra. X! Tandem, as a probabilistic approach, in certain cases does not allow performing reliable identification. A brief description of more than 100 different algorithms and software packages for processing of mass spectrometry data on peptides and proteins are presented on the sites http://en.wikipedia.org/wiki/Mass_spectrometry_software и <http://www.ms-utils.org>.

The use of the databases for protein and peptide identification allows deciphering mass-spectra of complex mixtures over a short period of time [66]. Almost all known amino acid sequences of proteins and peptides are combined in web open-access databases. Each of them has its own format for data storage, various degree of redundancy, interaction with related or similar databases. All databases can be divided into five types. The first type is archive databases, where information is entered by users. These databases include GenBank, EMBL, PDB. The second type is curated databases, which content is curated by specialists, such as Swiss-Prot. The third type is automated databases, where entries are generated by computer programs; they include, for example, TrEMBL. The next type is derivative databases which are supplemented by processing data from databases of the first two types. They include SCOP, PFAM, GO and others. The fifth type is integrated databases, which combines information from different databases, such as ENTREZ [67]. Detection of an amino acid sequence of peptides and proteins without using search programs and databases is called de novo sequencing. This approach is used for identification of proteins that were not described earlier, in case of the presence of unstudied mutations, post-translational modifications and so on. The used algorithms for de novo 36 sequencing are based on different mathematical methods. The first algorithms for determination of amino acid sequence [68,69] represented a search of all possible combinations of amino acids comprising the mass of the parent ion, which fragmentation was compared to the experimental mass spectrum. It is obvious that an error in measuring the mass of the parent ion leads to an increase in the number of corresponding combinations.

Conclusion

At present, new pharmaceutical preparations of protein and peptide nature, the methods for control of specific proteins emerge in food sector. To detect complex, multi-com-

ponent changes occurring in meat products, it is necessary to use methodological approaches, which allow registering hundreds and thousands of proteins. One of these approaches is proteomics, which makes it possible to identify and reveal quantitative and qualitative changes in the protein composition of cells and tissues. The advances of proteomics help researchers to solve tasks of post-translational modifications, cell signaling, functional and structural homology of proteins, detection of a gene expression level. The most significant instrument of proteomics is investigation of the protein maps of human and animal tissues.

Proteomic technologies are considered quite promising and effective for detection of the biochemical changes in meat products, for example, the changes in the thermally stable and species specific proteins, which are capable of becoming corresponding biomarkers.

Recently, proteomics became widely used in the field of biotechnology. Using the proteomic technologies, the Gorbатов Research Center for Food Systems has developed the methodological approaches for identification of the protein profile of meat products, experimental meat samples and specially produced sausage products, determined tissue-specific proteins, which can be used as individual biomarkers upon controlling meat products for correspondence to the stated composition. Also, soya and chicken proteins were registered, which are the marker of falsification [70,71].

Acknowledgments

The study was funded by the grant of the Russian Scientific Foundation (project No. 16-16-10073).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК:

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts K., Walter P. (2007). *Molecular Biology of the Cell*. Garland Science, 318.
2. Kent, S.B. (2009). Total chemical synthesis of proteins. *Chemical Society Reviews*, 38(2), 338–351.
3. Pickel, B., Schaller, A. (2013). Dirigent proteins: molecular characteristics and potential biotechnological applications. *Applied Microbiology and Biotechnology*, 97(19), 8427–8438.
4. Protein. (2018). [Электронный ресурс: <https://en.wikipedia.org/wiki/Protein> Дата обращения 01.05.2018]
5. Wickner, S. (1999). Posttranslational quality control: folding, refolding and degrading proteins. *Science*, 286, 1888–1893.
6. Марри, Р., Греннер, Д., Мейес, П., Родуэл, В. (1993). Биохимия человека: В 2-х томах. Т. 2. (Перевод с английского канд. физ.-мат. наук В. В. Борисова и канд. физ.-мат. наук Е. В. Дайниченко, под редакцией д-ра хим. наук А. М. Гинодмана). М., Мир. — 415 с. ISBN: 5-03-001774-7
7. Emig, D., Albrecht, M. (2011). Tissue-specific proteins and functional implications. *Journal of Proteome Research*, 10(4), 1893–1903.
8. Chen, G., Wang, J. (2012). Identifying functional modules in tissue specific protein interaction network. *Proceedings — 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops*, BIBMW 6470204, 581–586.
9. Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M. (2013). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, 4(9), 117.
10. Picard, B., Lebret, B., Cassar-Malek, I., Liaubet, L., Berri, C., Nihan-Duval le, B., Hocquette, J.F., Renand, G. (2015). Recent advances in omic technologies for meat quality management. *Meat Science*, 109, 18–26
11. Вострикова, Н.Л., Чернуха, И.М. (2017). Биоинформатика — инструмент интерпретации протеомных профилей белков мяса. *Теория и практика переработки мяса*, 2(1), 4–17.
12. Stepanenko, O.V., Verkhusha, V.V., Kuznetsova, I.M., Uversky, V.N., Turoverov, K.K. (2008). Fluorescent proteins as biomarkers and biosensors: throwing color lights on molecular and cellular processes. *Current Protein and Peptide Science*, 9(4), 338–369.
13. Yuste, R (2005). Fluorescence microscopy today. *Nature Methods*, 2(12), 902–904.
14. Collins, N., Poot, R.A., Kukimoto, I., Garcia-Jimenez, C., Delaire, G., Varga-Weisz, P.D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nature Genetics*, 32, 627–632.
15. Morton N.E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics*, 7 (3), 277–318.
16. Walker, J.H., Wilson, K. (2000). *Principles and Techniques of Practical Biochemistry*. Cambridge, UK: Cambridge University Press, 287–289.
17. Hohsaka, T., Sisido, M. (2002). Incorporation of non-natural amino acids into proteins. *Current Opinion in Chemical Biology*, 6(6), 809–815.
18. Cedrone, F., Ménez, A., Quéméneur, E. (2000). Tailoring new enzyme functions by rational redesign. *Current Opinion in Structural Biology*, 10(4), 405–410.
19. Görg, A., Weiss, W., Dunn, M.J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics*, 4(12), 3665–3685.
20. Koegl, M., Uetz, P. (2007). Improving yeast two-hybrid screening systems. *Briefings in Functional Genomics and Proteomics*, 6(4), 302–312.
21. Conrotto, P., Souchelnytskyi, S. (2008). Proteomic approaches in biological and medical sciences: principles and applications. *Experimental Oncology*, 30(3), 171–180.
22. Zhang, C., Kim, S.H. (2003). Overview of structural genomics: from structure to function. *Current Opinion in Chemical Biology*, 7(1), 28–32.
23. Standley, D.M., Kinjo, A.R., Kinoshita, K., Nakamura, H. (2008). Protein structure databases with new web services for structural biology and biomedical research. *Briefings in Bioinformatics*, 9(4), 276–285.
24. Harvey, D. (2000). *Modern analytical chemistry*. USA: The McGraw-Hill Companies. — 816 p.
25. Kroeff, E., Owens, R., Campbell, E., Johnson, R., Marks, H. (1989). Production scale purification of biosynthetic human insulin by reversed-phase high-performance liquid chromatography. *Journal of Chromatography A*, 461, 45–61.
26. Seino, S., Funakoshi, A., Fu, Z., Vinik, A. (1985). Identification of insulin variants in patients with hyperinsulinemia by reversed-phase high-performance liquid chromatography. *Diabetes*, 34(1), 1–7.
27. Grace, V. (2002). Technical Support Group. *The Handbook of Analysis and Purification of Peptides and Proteins by Reversed-Phase HPLC*. Third Edition, 4–5.
28. Aebersold, R., Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198–207.
29. Mann, M., Hendrickson, R., Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry*, 70, 437–473.
30. Aebersold, R., Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198–207.
31. Switzer, L., Giera, M., Niessen, W. (2013). Protein digestion: an overview of the available techniques and recent developments. *Journal of Proteome Research*, 12(3), 1067–1077.
32. Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., Smith, R. D. (2012). Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews*, 41(10), 3912–3928.
33. Marko-Varga, G., Fehniger, T. E. (2004). Proteomics and disease—the challenges for technology and discovery. *Journal of Proteome Research*, 3(2), 167–178.
34. Calligaris, D., Villard, C., Lafitte, D. (2011). Advances in top-down proteomics for disease biomarker discovery. *Journal of Proteomics*, 74(7), 920–934.
35. Chait, B. T. Chemistry. (2006). Mass spectrometry: bottom-up or topdown. *Science*, 314(5796), 65–66.
36. Wu, S. — L., Hühner, A. F. R., Hao, Z., Karger, B. L. (2007). On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an iso-

lated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *Journal of Proteome Research*, 6(11), 4230–4244.

37. McLafferty, F. W., Breuker, K., Jin, M., Han, X., Infusini, G., Jiang, H., Kong, X., Begley, T. P. (2007). Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS Journal*, 274(24), 6256–6268.

38. Lanucara, F.; Evers, C. E. (2013). Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. *Mass Spectrometry Reviews*, 32(1), 27–42.

39. Han, X., Jin, M., Breuker, K., McLafferty, F. W. (2006). Extending top-down mass spectrometry to proteins with masses great than 200 kilodaltons. *Science*, 314(5796), 109–112.

40. Chi, A., Bai, D. L., Geer, L. Y., Shabanowitz, J., Hunt, D. F. (2007). Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. *International Journal of Mass Spectrometry*, 259(1–3), 197–203.

41. Compton, P. D., Zamdborg, L., Thomas, P. M., Kelleher, N. L. (2011). On the scalability and requirements of whole protein mass spectrometry. *Analytical Chemistry*, 83(17), 6868–6874.

42. Boyne, M. T., Garcia, B. A., Li, M., Zamdborg, L., Wenger, C. D., Babal, S., Kelleher, N. L. (2008). Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *Journal of Proteome Research*, 8(1), 374–379.

43. Cannon, J., Lohnes, K., Wynne, C., Wang, Y., Edwards, N., Fenselau, C. (2010). High-throughput middle-down analysis using an orbitrap. *Journal of Proteome Research*, 9(8), 3886–3890.

44. Hauser, N. J., Han, H., McLuckey, S. A., Basile, F. (2008). Electron transfer dissociation of peptides generated by microwave D-cleavage digestion of proteins. *Journal of Proteome Research*, 7(5), 1867–1872.

45. Kalli, A., Hakansson, K. (2010). Electron capture dissociation of highly charged proteolytic peptides from Lys N, Lys C and Glu C digestion. *Molecular Biosystems*, 6(9), 1668–1681.

46. Wu, S. — L., Kim, J., Hancock, W. S., Karger, B. (2005). Extended Range Proteomic Analysis (ERPA): A new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and Epidermal Growth Factor Receptor (EGFR). *Journal of Proteome Research*, 4(4), 1155–1170.

47. Deutsch, E. W., Lam, H., Aebersold, R. (2008). Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiological Genomics*, 33(1), 18–25.

48. Brownridge, P., Beynon, R. J. (2011). The importance of the digest: proteolysis and absolute quantification in proteomics. *Methods*, 54(4), 351–360.

49. Perkins, D., Pappin, D., Creasy, D., Cottrell, J. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18), 3551–3567.

50. Eng, J., McCormack, A., Yates, J. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of The American Society for Mass Spectrometry*, 5(11), 976–989.

51. Craig, R., Beavis, R. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, 20(9), 1466–1467.

52. Tanner, S., Shu, H., Frank, A., Wang, L., Zandi, E., Mumby, M., Pevzner, P., Bafna, V. (2005). InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Analytical Chemistry*, 77(14), 4626–4639.

53. Geer, L., Markey, S., Kowalak, J., Wagner, L., Xu, M., Maynard, D., Yang, X., Shi, W., Bryant, S. (2004). Open mass spectrometry search algorithm. *Journal of Proteome Research*, 3(5), 958–964.

54. Xu, H., Freitas, M. (2009). MassMatrix: a database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. *Proteomics*, 9(6), 1548–1555.

55. Park, C., Klammer, A., Kall, L., MacCoss, M., Noble, W. (2008). Rapid and accurate peptide identification from tandem mass spectra. *Journal of Proteome Research*, 7(7), 624–627.

56. Tabb, D., Fernando, C., Chambers, M. (2007). MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *Journal of Proteome Research*, 6(2), 654–661.

57. Kim, S., Mischerikow, N., Bandeira, N., Navarro, J., Wich, L., Mohammed, S., Heck, A., Pevzner, P. (2010). The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Molecular and Cellular Proteomics*, 9(12), 2840–2852.

58. Pappin, D., Hojrup, P., Bleasby, A. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*, 3(6), 327–332.

59. Mascot software, the benchmark for identification, characterisation and quantitation of proteins using mass spectrometry data. [Электронный ресурс: <http://www.matrixscience.com>. Дата обращения 07.07.2018]

60. Perkins, D., Pappin, D., Creasy, D., Cottrell, J. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18), 3551–3567.

61. Автономов, Д., Агрон, И., Кононихин, А., Николаев, Е. (2009). Создание базы данных точных массово-временных меток для качественного и количественного подхода в исследовании протеома мочи человека с использованием изотопного мечения. *Труды МФТИ*, 1(1), 24–29 с.

62. Scoring Schemes. [Электронный ресурс: https://edwardslab.bmcb.georgetown.edu/mascot/help/scoring_help.html. Дата обращения 08.06.2018]

63. Protein Digestion. [Электронный ресурс: <http://proteomicsresource.washington.edu/protocols06/sequest.php>. Дата обращения 13.07.2018]

64. SEQUEST correlates uninterpreted tandem mass spectra of peptides with amino acid sequences from protein and nucleotide databases. [Электронный ресурс: <http://fields.scripps.edu/sequest/index.html>. Дата обращения 03.04.2018]

65. X! Tandem open source is software that can match tandem mass spectra with peptide sequences, in a process that has come to be known as protein identification. [Электронный ресурс: <http://www.thegpm.org/tandem>. Дата обращения 11.04.2018]

66. Sparkman, D. (2012). Informatics and mass-spectral databases in the evaluation of environmental mass spectral data. Saint Albans: ILMPublications.—528 p.

67. Объединенный центр вычислительной биологии и информатики. [Электронный ресурс: <http://www.jcabi.ru/index.html>. Дата обращения 10.05.2018]

68. Hamm, C., Wilson, W., Harvan, D. (1986). Peptide sequencing program. *Computer Applications in the Biosciences*, 2(2), 115–118.

69. Беризовская, Е.И., Ихалайнен, А.А., Антохин, А.М., Таранченко, В.Ф., Гончаров, В.М., Митрофанов Д.А., Аксенов, А.В., Родин, И.А., Шпигун, О.А. (2014). Возможности масс-спектрометрии высокого разрешения с ионизацией электрораспылением в определении моноизоотопных молекулярных масс рекомбинантного инсулина человека и его аналогов. *Масс-спектрометрия*, 11(4), 231–238.

70. Vostrikova, N.L., Chernukha, I.M., Kullkovskiy, A.V., Shishkin, S.O. (2016). Study and identification of main proteins and peptides to determine the content of muscle protein in structureless cooked products by the method of two-dimensional electrophoresis followed by the time-of-flight mass spectrometry identification. *Foods and Raw Materials*, 4(2), 136–147.

71. Вострикова, Н.Л., Куликовский, А.В., Чернуха, И.М., Ковалев, Л.И., Савчук, С.А. (2017). Определение белков мышечной ткани методами 2D электрофореза и времяпролетной масс-спектрометрии. *Журнал аналитической химии*, 72(10), 932–943.

REFERENCES

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts K., Walter P. (2007). *Molecular Biology of the Cell*. Garland Science, 318.
2. Kent, S.B. (2009). Total chemical synthesis of proteins. *Chemical Society Reviews*, 38(2), 338–51.
3. Pickel, B., Schaller, A. (2013). Dirigent proteins: molecular characteristics and potential biotechnological applications. *Applied Microbiology and Biotechnology*, 97(19), 8427–8438.
4. Protein (2018). [Electronic resource: <https://en.wikipedia.org/wiki/Protein> Date of circulation 01.05.2018]
5. Wickner, S. (1999). Posttranslational quality control: folding, refolding and degrading proteins. *Science*, 286, 1888–1893.

6. Marri, R., Grenner, D., Meyyes, P., Roduelli, W. (1993). *Human Biochemistry: In 2 volumes. T. 2.* (Translated from English by V.V. Borisova and Candidate of Physical and Mathematical Sciences E.V. Dainichenko, edited by L.M. Ginodman, Doctor of Chemical Sciences). M: World. — 415 p. ISBN: 5–03–001774–7. (In Russian)
7. Emig, D., Albrecht, M. (2011). Tissue-specific proteins and functional implications. *Journal of Proteome Research*, 10(4), 1893–1903.
8. Chen, G., Wang, J. (2012). Identifying functional modules in tissue specific protein interaction network. *Proceedings IEEE In-*

- ternational Conference on Bioinformatics and Biomedicine Workshops, BIBMW 6470204, 581–586.
9. Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M. (2013). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, 4(9), 117.
 10. Picard, B., Lebret, B., Cassar-Malek, I., Liaubet, L., Berri, C., Nihan-Duval le, B., Hocquette, J.F., Renand, G. (2015). Recent advances in omic technologies for meat quality management. *Meat Science*, 109, 18–26.
 11. Vostrikova, N.L., Chernukha, I.M. (2017). Bioinformatics – instrument interpretation proteomic profiles of meat protein. *Theory and practice of meat processing*, 2(1), 4–17
 12. Stepanenko, O.V., Verkhusha, V.V., Kuznetsova, I.M., Uversky, V.N., Turoverov, K.K. (2008). Fluorescent proteins as biomarkers and biosensors: throwing color lights on molecular and cellular processes. *Current Protein and Peptide Science*, 9(4), 338–69.
 13. Yuste, R. (2005). Fluorescence microscopy today. *Nature Methods*, 2(12), 902–904.
 14. Collins, N., Poot, R.A., Kukimoto, I., Garcia-Jimenez, C., Delaire, G., Varga-Weisz, P.D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nature Genetics*, 32, 627–632.
 15. Khwannimit, B. (2008). Serial evaluation of the MODS, SOFA and LOD scores to predict ICU mortality in mixed critically ill patients. *Journal of the Medical Association of Thailand*, 91(9), 1336–1342.
 16. Walker, J.H., Wilson, K. (2000). Principles and Techniques of Practical Biochemistry. Cambridge University Press, 287–289.
 17. Hohsaka, T., Sisido, M. (2002). Incorporation of non-natural amino acids into proteins. *Current Opinion in Chemical Biology*, 6(6), 809–815.
 18. Cedrone, F., Ménez, A., Quéméneur, E. (2000). Tailoring new enzyme functions by rational redesign. *Current Opinion in Structural Biology*, 10(4), 405–410.
 19. Görg, A., Weiss, W., Dunn, M.J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics*, 4(12), 3665–3685.
 20. Koegl, M., Uetz, P. (2007). Improving yeast two-hybrid screening systems. *Briefings in Functional Genomics and Proteomics*, 6(4), 302–312.
 21. Conrotto, P., Souchelnytskyi, S. (2008). Proteomic approaches in biological and medical sciences: principles and applications. *Experimental Oncology*, 30(3), 171–180.
 22. Zhang, C., Kim, S.H. (2003). Overview of structural genomics: from structure to function. *Current Opinion in Chemical Biology*, 7(1), 28–32.
 23. Standley, D.M., Kinjo, A.R., Kinoshita, K., Nakamura, H. (2008). Protein structure databases with new web services for structural biology and biomedical research. *Briefings in Bioinformatics*, 9(4), 276–285.
 24. Harvey, D. (2000). Modern analytical chemistry. USA: The McGraw-Hill Companies. – 816 p.
 25. Kroeff, E., Owens, R., Campbell, E., Johnson, R., Marks, H. (1989). Production scale purification of biosynthetic human insulin by reversed-phase high-performance liquid chromatography. *Journal of Chromatography A*, 461, 45–61.
 26. Selino, S., Funakoshi, A., Fu, Z., Vinik, A. (1985). Identification of insulin variants in patients with hyperinsulinemia by reversed-phase high-performance liquid chromatography. *Diabetes*, 34(1), 1–7.
 27. Grace, V. (2002). Technical Support Group. The Handbook of Analysis and Purification of Peptides and Proteins by Reversed-Phase HPLC. Third Edition, 4–5.
 28. Aebersold, R., Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198–207.
 29. Mann, M., Hendrickson, R., Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry*, 70, 437–473.
 30. Aebersold, R., Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198–207.
 31. Switzar, L., Giera, M., Niessen, W. (2013). Protein digestion: an overview of the available techniques and recent developments. *Journal of Proteome Research*, 12(3), 1067–1077.
 32. Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., Smith, R. D. (2012). Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews*, 41(10), 3912–3928.
 33. Marko-Varga, G., Fehniger, T. E. (2004). Proteomics and disease—the challenges for technology and discovery. *Journal of Proteome Research*, 3(2), 167–178.
 34. Calligaris, D., Villard, C., Lafitte, D. (2011). Advances in top-down proteomics for disease biomarker discovery. *Journal of Proteomics*, 74(7), 920–934.
 35. Chait, B. T. Chemistry. (2006). Mass spectrometry: bottom-up or topdown. *Science*, 314(5796), 65–66.
 36. Wu, S. — L., Hühner, A. F. R., Hao, Z., Karger, B. L. (2007). On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *Journal of Proteome Research*, 6(11), 4230–4244.
 37. McLafferty, F. W., Breuker, K., Jin, M., Han, X., Infusini, G., Jiang, H., Kong, X., Begley, T. P. (2007). Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS Journal*, 274(24), 6256–6268.
 38. Lanucara, F.; Eysers, C. E. (2013). Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. *Mass Spectrometry Reviews*, 32(1), 27–42.
 39. Han, X., Jin, M., Breuker, K., McLafferty, F. W. (2006). Extending top-down mass spectrometry to proteins with masses great than 200 kilodaltons. *Science*, 314(5796), 109–112.
 40. Chi, A., Bai, D. L., Geer, L. Y., Shabanowitz, J., Hunt, D. F. (2007). Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. *International Journal of Mass Spectrometry*, 259(1–3), 197–203.
 41. Compton, P. D., Zamdborg, L., Thomas, P. M., Kelleher, N. L. (2011). On the scalability and requirements of whole protein mass spectrometry. *Analytical Chemistry*, 83(17), 6868–6874.
 42. Boyne, M. T., Garcia, B. A., Li, M., Zamdborg, L., Wenger, C. D., Babai, S., Kelleher, N. L. (2008). Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *Journal of Proteome Research*, 8(1), 374–379.
 43. Cannon, J., Lohnes, K., Wynne, C., Wang, Y., Edwards, N., Fenselau, C. (2010). High-throughput middle-down analysis using an orbitrap. *Journal of Proteome Research*, 9(8), 3886–3890.
 44. Hauser, N. J., Han, H., McLuckey, S. A., Basile, F. (2008). Electron transfer dissociation of peptides generated by microwave D-cleavage digestion of proteins. *Journal of Proteome Research*, 7(5), 1867–1872.
 45. Kalli, A., Hakansson, K. (2010). Electron capture dissociation of highly charged proteolytic peptides from Lys N, Lys C and Glu C digestion. *Molecular Biosystems*, 6(9), 1668–1681.
 46. Wu, S. — L., Kim, J., Hancock, W. S., Karger, B. (2005). Extended Range Proteomic Analysis (ERPA): A new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and Epidermal Growth Factor Receptor (EGFR). *Journal of Proteome Research*, 4(4), 1155–1170.
 47. Deutsch, E. W., Lam, H., Aebersold, R. (2008). Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiological Genomics*, 33(1), 18–25.
 48. Brownridge, P., Beynon, R. J. (2011). The importance of the digest: proteolysis and absolute quantification in proteomics. *Methods*, 54(4), 351–360.
 49. Perkins, D., Pappin, D., Creasy, D., Cottrell, J. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18), 3551–3567.
 50. Eng, J., McCormack, A., Yates, J. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of The American Society for Mass Spectrometry*, 5(11), 976–989.
 51. Craig, R., Beavis, R. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, 20(9), 1466–1467.
 52. Tanner, S., Shu, H., Frank, A., Wang, L., Zandi, E., Mumby, M., Pevzner, P., Bafna, V. (2005). InsPecT: Identification of posttranslationally modified peptides from tandem mass spectra. *Analytical Chemistry*, 77(14), 4626–4639.
 53. Geer, L., Markey, S., Kowalak, J., Wagner, L., Xu, M., Maynard, D., Yang, X., Shi, W., Bryant, S. (2004). Open mass spectrometry search algorithm. *Journal of Proteome Research*, 3(5), 958–964.
 54. Xu, H., Freitas, M. (2009). MassMatrix: a database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. *Proteomics*, 9(6), 1548–1555.
 55. Park, C., Klammer, A., Kall, L., MacCoss, M., Noble, W. (2008). Rapid and accurate peptide identification from tandem mass spectra. *Journal of Proteome Research*, 7(7), 624–627.
 56. Tabb, D., Fernando, C., Chambers, M. (2007). MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *Journal of Proteome Research*, 6(2), 654–661.
 57. Kim, S., Mischerikow, N., Bandeira, N., Navarro, J., Wich, L., Mohammed, S., Heck, A., Pevzner, P. (2010). The generating

function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Molecular and Cellular Proteomics*, 9(12), 2840–2852.

58. Pappin, D., Hojrup, P., Bleasby, A. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*, 3(6), 327–332.

59. Mascot software, the benchmark for identification, characterisation and quantitation of proteins using mass spectrometry data. [Electronic resource: <http://www.matrixscience.com>. Date of circulation 07.07.2018]

60. Perkins, D., Pappin, D., Creasy, D., Cottrell, J. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18), 3551–3567.

61. Avtonomov, D., Agron, I., Kononikhin, A., Nikolayev, Ye. (2009). Creation of a database of accurate mass-time marks for a qualitative and quantitative approach in the study of a human urine proteome using isotope labeling. *Proceedings of MFTI*, 1(1), 24–29 p. (In Russian)

62. Scoring Schemes. [Electronic resource: https://edwardsiab.bmcb.georgetown.edu/mascot/help/scoring_help.html. Date of circulation 08.06.2018]

63. Protein Digestion. [Electronic resource: <http://proteomicsresource.washington.edu/protocols06/sequest.php>. Date of circulation 13.07.2018]

64. SEQUEST correlates uninterpreted tandem mass spectra of peptides with amino acid sequences from protein and nucleotide databases. [Electronic resource: <http://fields.scripps.edu/sequest/index.html>. Date of circulation 03.04.2018]

65. X! Tandem open source is software that can match tandem mass spectra with peptide sequences, in a process that has come to be known as protein identification. [Electronic resource: <http://www.thegpm.org/tandem>. Date of circulation 11.04.2018]

66. Sparkman, D. (2012). Informatics and mass-spectral databases in the evaluation of environmental mass spectral data. Saint Albans: ILMPublications.—528 p.

67. Joint Center for Computational Biology and Informatics. [Electronic resource: <http://www.jcibi.ru/index.html>. Date of circulation 10.05.2018]

68. Hamm, C., Wilson, W., Harvan, D. (1986). Peptide sequencing program. *Computer Applications in the Biosciences*, 2(2), 115–118.

69. Berizovskaya, E.I., Ihalaynen, A.A., Antohin, A.M., Taranchenko, V.F., Goncharov, V.M., Mitrofanov, D.A., Aksenov, A.V., Rodin, I.A., Shpigun, O.A. (2014). Capability of high resolution electrospray ionization mass spectrometry in the determination of monoisotopic molecular masses of the recombinant human insulin and its analogues *Mass spectrometry*, 11 (4), 231–238. (In Russian)

70. Vostrikova, N.L., Chernukha, I.M., Kullkovskiy, A.V., Shishkin, S.S. (2016). Study and identification of main proteins and peptides to determine the content of muscle protein in structureless cooked products by the method of two-dimensional electrophoresis followed by the time-of-flight mass spectrometry identification. *Foods and Raw Materials*, 4(2), 136–147.

71. Vostrikova, N.L., Kullkovskiy, A.V., Chernukha, I.M., Kovalev, L.I., Savchuk, S.A. (2017). Determination of muscle proteins by 2D methods of electrophoresis and time-of-flight mass spectrometry. *Journal of Analytical Chemistry*, 72(10), 1102–1112. (In Russian)

СВЕДЕНИЯ ОБ АВТОРАХ

Принадлежность к организации

Вострикова Наталья Леонидовна — кандидат технических наук, заведующий лабораторией «Научно-методические работы, биологические и аналитические исследования», Федеральный научный центр пищевых систем им. В.М. Горбатова РАН

109316, г. Москва, ул. Талалихина, 26

Тел.: +7-495-676-79-81

E-mail: n.vostrikova@fnpcs.ru

*автор для переписки

Чернуха Ирина Михайловна — доктор технических наук, профессор, главный научный сотрудник Экспериментальной клиники-лаборатории биологически активных веществ животного происхождения, Всероссийский научно-исследовательский институт мясной промышленности имени В.М. Горбатова,

109316, г. Москва, ул. Талалихина, 26

Тел.: +7-495-676-63-21

E-mail: imcher@inbox.ru

Хвостов Даниил Владиславович — младший научный сотрудник лаборатории биоаналитических исследований, Научный центр биомедицинских технологий Федерального медико-биологического агентства

143442, Московская область, Красногорский район, пос. Светлые Горы, владение 1.

Тел. +7-495-561-5264

E-mail: scbmt@yandex.ru

Критерии авторства

Ответственность за работу и предоставленные сведения несут все авторы.

Все авторы в равной степени участвовали в этой работе.

Авторы в равных долях имеют отношение к написанию рукописи и одинаково несут ответственность за плагиат

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов

Поступила 16.07.2018

AUTHOR INFORMATION

Affiliation

Natalya L. Vostrikova — candidate of technical sciences, head of laboratory «Scientific and methodical work, biological and analytical research», V.M. Gorbatov Federal Research Center for Food Systems of Russian Academy of Sciences

109316, Moscow, Talalikhina str., 26

Tel.: +7-495-676-79-81

E-mail: n.vostrikova@fnpcs.ru

*corresponding author

Irina M. Chernukha — doctor of technical sciences, professor, leading research scientist of Experimental clinic — laboratory «Biologically active substances of an animal origin», V.M. Gorbatov Federal Research Center for Food Systems of Russian Academy of Sciences

109316, Moscow, ul. Talalikhina, 26

Tel.: +7-495-676-63-21

E-mail: imcher@inbox.ru

Daniil V. Khvostov — junior researcher of the Laboratory of Bioanalytical Research, Scientific Center for Biomedical Technologies of the Federal Medical-Biological Agency

143442, Moscow region, Krasnogorsk district, Svetlye gory village, 1

Tel. +7-495-561-5264

E-mail: scbmt@yandex.ru

Contribution

All authors bear responsibility for the work and presented data.

All authors made an equal contribution to the work.

The authors were equally involved in writing the manuscript and bear the equal responsibility for plagiarism.

Conflict of interest

The authors declare no conflict of interest

Received 16.07.2018