

УЛЬТРАБЫСТРАЯ ПРОТЕОМИКА

Мини-обзор

© 2024 И.И. Федоров^{1,2}, С.А. Протасов^{1,2}, И.А. Тарасова², М.В. Горшков^{2*}

¹ Московский физико-технический институт (национальный исследовательский университет),
141700 Долгопрудный, Московская обл., Россия

² Федеральный исследовательский центр химической физики имени Н.Н. Семенова РАН,
Институт энергетических проблем химической физики имени В.Л. Тальрозе,
119334 Москва, Россия; электронная почта: mike.gorshkov@gmail.com

Поступила в редакцию 23.05.2024

После доработки 21.06.2024

Принята к публикации 24.06.2024

Современный этап развития протеомных исследований в области биологии, медицины, разработки новых лекарств, популяционного скрининга или персонализированных подходов к терапии диктует необходимость анализа больших выборок образцов с наименьшими затратами экспериментального времени. До недавнего времени технологии масс-спектрометрических измерений в протеомике относились к уникальным для идентификации и количественного определения белкового состава, но также были низкопроизводительными, с многочасовым анализом одного образца. Это входило в противоречие с динамикой изменений биологических систем на уровне всего клеточного протеома под воздействием внешних и внутренних факторов. Так, низкая скорость полнопротеомного анализа является основным фактором, ограничивающим развитие функциональной протеомики, где требуется аннотировать участие белков во внутриклеточных процессах не только в широком диапазоне условий жизнедеятельности клетки, но и на протяжении длительного времени. Огромный уровень гетерогенности клеток тканей или новообразований даже одного типа диктует необходимость анализа биологической системы на уровне индивидуальных клеток. В этих исследованиях речь идёт о получении молекулярных характеристик уже для десятков, если не сотен тысяч индивидуальных клеток, в том числе их полного протеомного профиля. Развитие технологий масс-спектрометрии высокого разрешения и точности измерения масс, предсказательной хроматографии, новых методов разделения ионов пептидов в газовой фазе и обработки протеомных данных на основе алгоритмов искусственного интеллекта открыли перспективу существенно, если не радикально, увеличить производительность полнопротеомного анализа и привели к реализации концепции ультракороткой протеомики. В работах буквально последних нескольких лет были продемонстрированы производительности полнопротеомного анализа на уровне нескольких сотен образцов в сутки при глубине анализа в несколько тысяч белков, что было немыслимо ещё буквально три-четыре года назад. В обзоре рассматриваются предпосылки развития, а также основные современные методы и подходы в реализации ультракороткого полнопротеомного анализа.

КЛЮЧЕВЫЕ СЛОВА: протеомика, масс-спектрометрия, пептиды, белки, ультракороткий анализ, количественная протеомика.

DOI: 10.31857/S0320972524080011 EDN: KEERFQ

ВВЕДЕНИЕ

В настоящее время полнопротеомный анализ широко используется в различных областях биологии и медицины [1, 2]. Основным методом такого анализа является масс-спектрометрия,

позволяющая получать количественную информацию об изменении белкового состава клеток, подверженных тому или иному воздействию. Среди основных начальных этапов становления количественной протеомики следует отметить реализацию концепций поиска по базам данных и/или спектральным библиотекам [3] и идентификации белков по уникальному набору масс

* Адресат для корреспонденции.

их протеолитических (как правило, триптических) пептидов (пептидные отпечатки белков – Peptide Mass Fingerprint, PMF [4]). Если говорить об используемых типах масс-анализаторов, то первые протеомы модельных организмов были идентифицированы на радиочастотных квадрупольных ионных ловушках и времяпролётных масс-спектрометрах [5]. В этих результатах, как правило, отсутствовал контроль доли ложноположительных идентификаций (FDR, False Discovery Rate) [6, 7], концепция которого появилась в протеомике в 2007 г. с реализацией подхода на основе ложных или «бессмысленных» (декойных) последовательностей [7]. Ранее, начиная с работ Смита и др., появилось понимание необходимости использования для полнопротеомного анализа масс-спектрометрии высокого разрешения, представленной на начальных этапах исключительно масс-анализаторами ионного циклотронного резонанса с преобразованием Фурье (ИЦР ПФ) в сочетании с нанопотоковыми режимами разделения и ионизации пептидов [8, 9]. Появление более производительных по сравнению с ИЦР ПФ масс-анализаторов высокого разрешения Орбитрэп (Orbitrap) позволило к 2015 г. выйти на уровень идентификации до половины и более всех белков простых организмов в рамках одночасового, «однопрогонного» эксперимента ВЭЖХ–МС/МС (МС/МС – тандемная масс-спектрометрия) [10]. В настоящее время для протеомов клеток человека идентифицируется 5000–6000 белков в однопрогонном двухчасовом эксперименте [11]. Дальнейшее увеличение глубины полнопротеомного анализа достигается либо дополнительным фракционированием проб на уровне белков или пептидов [12–14], либо увеличением длительности градиента ВЭЖХ до нескольких часов и использованием длинных хроматографических колонок [12, 15]. Так, например, комбинация фракционирования протеолитических смесей и многочасовых градиентов ВЭЖХ-разделения на капиллярных колонках длиной 40 см и более позволила идентифицировать более половины кодируемого протеома человека [16]. Хотя достижение такого уровня покрытия протеома представляет значительный интерес, полное затраченное инструментальное время в цитируемых выше экспериментах составило 288 часов, что делает такой анализ уникальным, но нецелесообразным для многих

задач химической, клинической или популяционной протеомики, предполагающих рутинный поток из сотен образцов в сутки. Одновременный анализ нескольких меченых образцов, объединённых в один пул (мультиплексинг образцов) [17, 18], в настоящее время реализуемый с использованием ТМТ-(Tandem Mass Tag) меток [19], частично решает проблему инструментальных затрат на полнопротеомный количественный анализ в пересчёте на один образец. Однако связанные с его использованием проблемы увеличения аналитической сложности образцов и необходимости использования фракционирования не переводят его в разряд методов ультракороткой протеомики, которую можно определить как анализ 200 и более образцов в сутки. Действительно, как показали недавние исследования на примере клеточных линий глиобластомы, обработанных интерфероном, количественный 40-минутный анализ с использованием 10-плексного набора ТМТ-меток (что примерно соответствует производительности в 200 протеомных анализов в сутки) даёт довольно скудную картину по количеству идентификаций интерферон-регулируемых белков [20].

После значительного промежутка времени, прошедшего с момента первых демонстраций количественного полнопротеомного анализа на основе меток точных масс и времён элюирования пептидов (ATM, Accurate Mass and Time tags) в минутном диапазоне градиентов разделения, в последние годы возобновился интерес к этому направлению, которое можно условно назвать «ультрабыстрой протеомикой». В её основе лежит реализация новых методов масс-спектрометрии высокого разрешения в сочетании с ультракоротким разделением смесей пептидов (включая разделение ионов пептидов в газовой фазе), таких как информационно независимый полнопротеомный анализ (DIA, Data Independent Acquisition) [13] и прямая хроматомасс-спектрометрическая идентификация белков (DirectMS1) [20]. Эти методы позволяют полуколичественно анализировать протеомы с производительностью более 200 образцов в сутки.

В данном обзоре рассматриваются развиваемые в последние несколько лет новые подходы ультрабыстрой протеомики и кратко обсуждаются перспективы их дальнейшего развития.

Принятые сокращения: AMT tags – метки точных масс и времён элюирования пептидов; DDA – метод информационно зависимого полнопротеомного анализа; DIA – метод информационно независимого полнопротеомного анализа; DirectMS1 – метод прямой хроматомасс-спектрометрической идентификации белков; DISPA – метод протеомного анализа на основе прямой подачи протеолитической смеси в источник ионизации, минуя стадию жидкостно-хроматографического разделения; МС/МС – тандемная масс-спектрометрия, МС1 – масс-спектры первого уровня ионов-предшественников; PMF – пептидные отпечатки белков.

ПРЕДЫСТОРИЯ УЛЬТРАКОРОТКОЙ ПРОТЕОМИКИ

Одной из первых реализаций идеи ультракороткого протеомного анализа являлся подход на основе пептидных отпечатков белков (PMF) [4, 21, 22]. Этот подход заключается в предварительном разделении белков с помощью гель-электрофореза или жидкостной хроматографии, специфическом гидролизе белковых фракций на протеолитические пептиды (обычно используется фермент трипсин), что образует уникальный для каждого белка набор масс его пептидных фрагментов, и регистрации масс-спектров ионов пептидов. Для идентификации белков экспериментально полученные массы ионов пептидов сравниваются с теоретическими, полученными на основе имеющихся баз данных последовательностей белков протеома соответствующего организма, что схематично представлено на рис. 1 [23].

Масс-спектрометрия на основе матрично-активированной лазерной десорбции/ионизации (MALDI-MS, Matrix-Assisted Laser Desorption and Ionization Mass Spectrometry) является наиболее часто используемым методом для реализации PMF-подхода [4, 24]. Понятно, что такой подход не является в буквальном смысле ультракоротким, поскольку он требует предварительного разделения белков на большое количество фракций – как правило, с использованием электрофореза SDS-PAGE (Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis) [21] – каждая из которых подвергается гидролизу и анализу. Также довольно быстро пришло понимание того, что метод PMF неэффективен при анализе сложных смесей [25], в которых представлены гидролизаты десятков белков, не говоря уже о тех, которые содержат тысячи белков протеома. В настоящее время он используется практически исключительно в ана-

лизе и подтверждении индивидуальных, как правило, предварительно выделенных белков.

С прогрессом в разработке масс-анализаторов высокого разрешения, таких как масс-спектрометры ионного циклотронного резонанса, идея идентификации белков на основе измерения масс-спектров ионов пептидов как способ быстрого полнопротеомного анализа была воплощена в методе точных массовых меток (Accurate Mass Tags, AMT) [26]. Метод заключается в предварительном МС/МС-анализе объединённого образца (пула) с идентификацией и составлением списка пептидов, потенциально присутствующих в анализируемых пробах. Последующая идентификация белков в каждом образце пула осуществляется на основе сопоставления этого «пул»-специфичного списка масс идентифицированных пептидов с массами пептидов, регистрируемых в отдельном анализе каждого из образцов. Базовая идея подхода состоит в том, что если молекулярную массу одного пептида можно измерить с достаточно высокой точностью, так, чтобы его масса была уникальной среди всех возможных пептидов, предсказанных на основе полногеномного секвенирования, то её можно было бы использовать в качестве «точной массовой метки» для идентификации белков. Соответственно, генерация списка таких AMT позволяет анализировать продукты гидролиза белков всего протеома (например, в результате расщепления трипсином) с большей скоростью и чувствительностью. Более того, последующий анализ можно проводить без использования фрагментации пептидов, т.е. подход потенциально является бесфрагментационным и, соответственно, совместимым с короткими градиентами разделения протеолитических смесей.

Довольно быстро стало понятно, что использование дополнительных к точной массе комплементарных данных, таких как времена

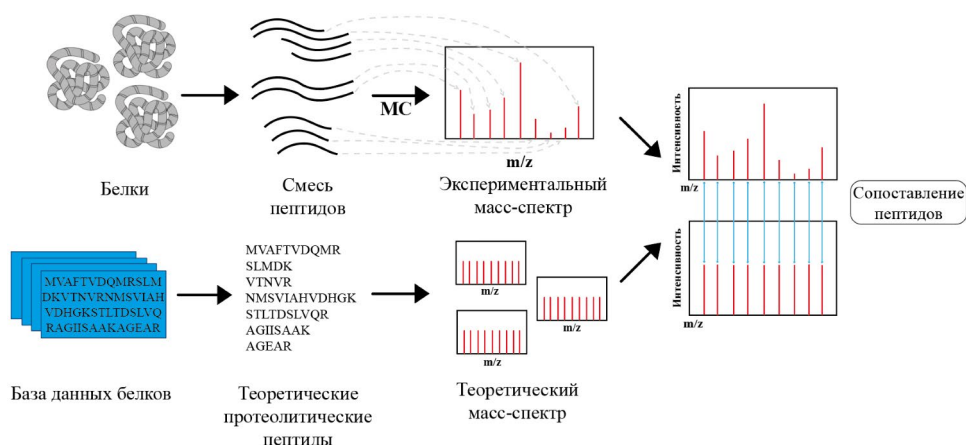


Рис. 1. Схема реализации метода PMF. Поскольку массы пептидов не специфичны относительно аминокислотной последовательности, то в сложных смесях на одну и ту же измеренную массу или набор масс может попадать несколько, в некоторых случаях десятки возможных белков, что затрудняет их идентификацию

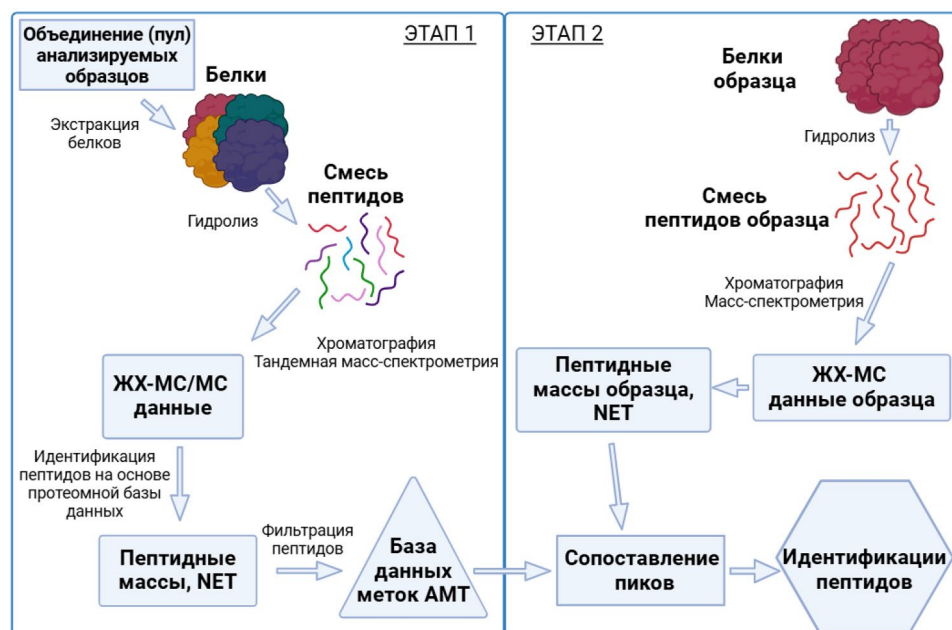


Рис. 2. Схема реализации метода АМТ на основе точных масс ионов пептидов и нормализованных времён элюирования, NET. Нормализация времён осуществляется в диапазон [0, 1]. Предсказанные NET рассчитываются для используемых условий разделения с помощью либо простых линейных преобразований, либо на основе нейронных сетей [27]

элюирования пептидов, существенно повышает уникальность такой комбинации для аминокислотной последовательности пептидов. Соответственно, дальнейшим развитием этого подхода явилось добавление нормализованных времён элюирования пептидов, что превратило его в метод меток точных масс и времён (AMT-tags) [26]. Уже в первых демонстрациях метода AMT-tags была продемонстрирована возможность полнопротеомного анализа для относительно небольших протеомов, в частности *Deinococcus radiodurans* [27]. Более того, поскольку метод не требует фрагментации пептидов для их идентификации (за исключением стадии составления списка АМТ-меток), такой полнопротеомный анализ был впервые продемонстрирован в минутном диапазоне времён.

Стандартная реализация метода АМТ содержит 2 основных этапа (рис. 2): (1) создание «пул»-специфичной базы данных меток АМТ пептидов для пулов образцов контрольной и тестовой групп с использованием глубокого (как правило, на основе фракционирования) полнопротеомного МС/МС-анализа; и (2) быстрый анализ ВЭЖХ-МС1 образцов (МС1 – масс-спектры первого уровня ионов-предшественников) с идентификацией белков на основе созданной на первом этапе «пул»-специфичной базы данных меток АМТ. На первом этапе с каждым из идентифицированных пептидов сопоставляется его точная масса в пределах ошибки измерения и нормализованное время выхода (NET, Normalized Elution Time). Затем сле-

дует этап ВЭЖХ-МС1-анализа большой когорты нефракционированных образцов, результатом которого является получение списка экспериментальных точных масс и зарядовых состояний пептидов, а также времён элюирования. Последние приводятся к нормализованной шкале времён, в наиболее простом исполнении – линейным преобразованием. Также могут использоваться NET, предсказанные на основе нейронных сетей и ранее полученных экспериментальных наборов данных идентифицированных пептидов [27]. Идентификация пептидов проанализированного образца осуществляется на основе сравнения экспериментальных наборов данных с базой меток АМТ, после чего осуществляется идентификация белков и их количественный анализ. Количественные характеристики идентифицированных белков определяются на основе интенсивностей ионов пептидов в масс-спектрах.

В то время как метод АМТ продемонстрировал возможность ультракороткого количественного анализа протеомов ряда организмов [28, 29], его более широкое применение в протеомике затруднено отсутствием контроля FDR. Также проблемами являются выравнивание времён элюирования пептидов для расчёта NET между различными экспериментами и различными хроматографическими условиями. В первую очередь речь идёт об условиях, при которых создавалась база меток АМТ, и условиях, при которых проводился последующий быстрый протеомный анализ [30]. Одним из решений последней из перечисленных

проблем стало использование различных моделей предсказания времён элюирования пептидов [31, 32] и создание стандартизированных и/или универсальных баз данных времён элюирования пептидов для АМТ-меток на их основе [33].

По мере внедрения полнопротеомного анализа с использованием информационно зависимого подхода (DDA, Data Dependent Acquisition) на основе гибридных масс-спектрометров с ионной ловушкой высокого разрешения Орбитрэп [34–36] в лабораторную практику метод АМТ перестал широко использоваться. При этом сама концепция DDA, в котором ионы пептидов, детектируемые в спектрах MS1, последовательно изолируются в радиочастотной ионной ловушке гибридного масс-спектрометра и накапливаются до количества, достаточного для получения полноценного спектра фрагментации, подразумевает использование длинных градиентов ВЭЖХ. Даже в случае многочасовых разделений гидролизатов протеомов, достигающих до 10 часов в отдельных экспериментах [37], идентифицируется только малая часть регистрируемых в MS1 и доступных для анализа пептидов [38–40]. Тем не менее DDA стал основным методом количественного полнопротеомного анализа в последние годы с глубиной покрытия протеома, достигающей в ряде работ 10 000 и более белковых идентификаций [16, 37, 41, 42]. Несмотря на очевидную важность получения как можно большей глубины анализа протеома, нельзя также не указать и на очевидную проблему: огромные затраты инструментального времени для анализа одного образца, особенно в случае использования интенсивного префракционирования анализируемых смесей [41, 43–47].

МЕТОД ИНФОРМАЦИОННО НЕЗАВИСИМОГО АНАЛИЗА В УЛЬТРАБЫСТРОЙ ПРОТЕОМИКЕ

Одним из наиболее очевидных методов ультракороткой протеомики стал метод информационно независимого анализа (DIA) [13]). В отличие от DDA, в этом методе отсутствует стадия последовательного отбора ионов-предшественников по точной измеренной массе MS1 для последующей изоляции, накопления и фрагментации, которая является одной из основных причин использования длинных градиентов разделения. Вместо этого в DIA накопление и фрагментация ионов осуществляются в широком окне масс с последующим перестроением накопительного устройства на соседнее окно и т.д. (рис. 3). В результате практически все прекурсорные ионы, присутствующие в MS1, фрагментируются в серии таких окон (стандартно, размером в 20–25 Th), покрыва-

ющих весь диапазон m/z детектирования ионов пептидов. Понятно, что спектры фрагментации в таких окнах являются смешанными (или, как ещё говорят, обладают высоким уровнем мультиплексности), с одновременным присутствием фрагментов от десятков ионов пептидов, что ставит перед экспериментаторами непростую задачу их интерпретации (деконволюции). Каждой серии таких окон соответствует предварительно зарегистрированный масс-спектр MS1 и время элюирования. Последнее является также одним из ключевых параметров для последующей деконволюции спектров фрагментации и идентификации пептидов. Размер окон и, соответственно, эффективность деконволюции спектров фрагментации определяется характеристиками масс-анализатора. Так, комбинация масс-анализатора Орбитрэп и анализатора Astral (ASymmetric TRAck Lossless) позволила уменьшить окна фрагментации до 2 Th, что фактически стирает границу между методами DIA и DDA в полнопротеомном анализе [48, 49].

Описанная выше схема DIA является его наиболее широко используемой реализацией, называемой SWATH-MS (Sequential Window Acquisition of All Theoretical Mass Spectra [50]). Основным преимуществом этого метода является преодоление проблемы стохастичности данных в стандартном подходе на основе DDA, связанной с выбором ограниченного количества наиболее интенсивных в данном масс-спектре ионов-предшественников для селективного накопления и фрагментации. Результатом является существенно меньший уровень пропущенных значений интенсивностей пептидов, что делает DIA альтернативой DDA в количественной протеомике [51]. Одновременно, поскольку в DIA происходит фрагментация всех ионов прекурсоров в ограниченном числе окон, этот метод позволяет работать с более короткими градиентами ВЭЖХ [52]. Дальнейшая оптимизация окон изоляции позволила увеличить глубину анализа протеомов с использованием ультракоротких градиентов разделения [53, 54]. Дополнительное разделение ионов пептидов по ионной подвижности продемонстрировало возможность идентифицировать в 5 нг гидролизата HeLa более 1000 белков в режиме DIA и 5-минутных градиентов ВЭЖХ [55].

Одним из сдерживающих факторов в развитии DIA в качестве рутинно используемого метода ультрабыстрого полнопротеомного анализа являлся высокий уровень мультиплексности масс-спектров фрагментации, что требует сложных алгоритмов обработки данных и деконволюции спектров. Стандартным решением этой проблемы являлось использование библиотек спектров фрагментации для анализируемого пула образцов. Такие библиотеки создавались с использованием

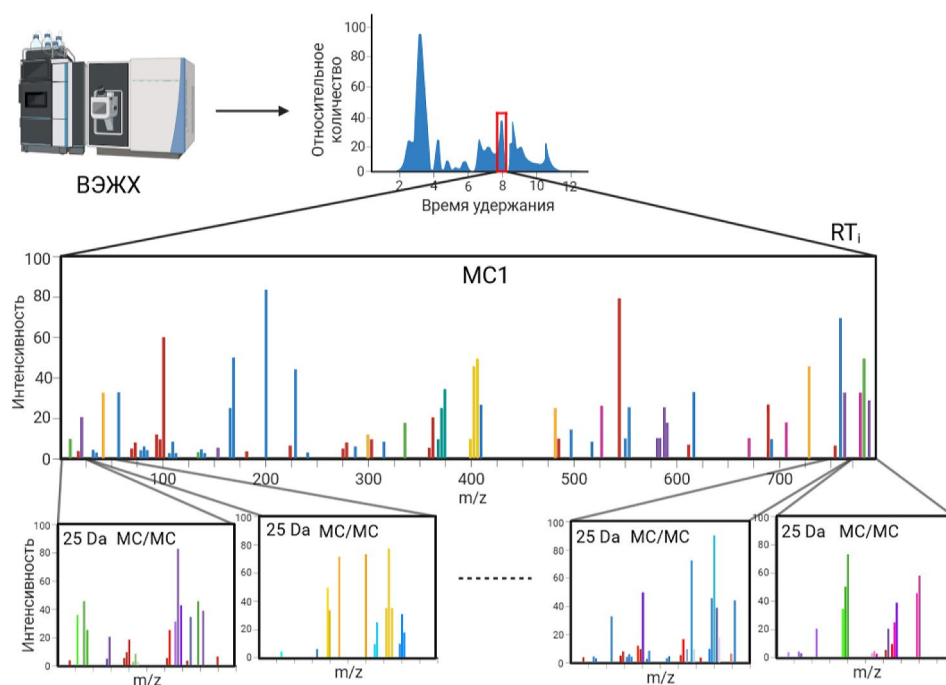


Рис. 3. Схема работы метода DIA. Вместо изолирования для последующей фрагментации отдельных ионов-предшественников, регистрируемых в масс-спектрах первого уровня, MS1, весь диапазон масс делится на широкие окна, в которых осуществляется накопление и фрагментация всех присутствующих в них ионов. Таким образом, DIA позволяет получить спектры фрагментации для всех ионов-предшественников, потенциально присутствующих в образце

глубокого полнопротеомного анализа методом DDA, что, в свою очередь, делало DIA не в полной мере информационно независимым. Помимо очевидных затрат инструментального времени на получение таких библиотек, что превращает DIA в условно быстрый метод протеомного анализа, использование экспериментальных библиотек существенно сдерживало использование DIA в межлабораторных и клинических исследованиях. Наконец, остаётся базовая проблема такого подхода – невозможность идентифицировать пептиды, спектры фрагментации которых в библиотеке отсутствуют. Прогресс в разработке алгоритмов на основе машинного обучения для предсказания спектров фрагментации и времён удерживания пептидов *in silico* позволил решить эту проблему [56–58]. Тем не менее крайне высокий уровень мультиплексности спектров в результате интерференции спектров фрагментации от разных одновременно элюируемых ионов-предшественников существенно усиливается в случае использования коротких градиентов разделения. До недавнего времени это делало невозможным извлечение сколь-либо значимого количества идентификаций в таких спектрах и ограничивало использование DIA в приложениях, требующих анализа больших выборок образцов. Эти проблемы были учтены в недавней разработке алгоритма DIA-NN, основанного на использовании нейронных сетей для различения сигналов ионов фрагментов и

шума в масс-спектрах, а также использующего новые стратегии для извлечения количественной информации и выравнивания хроматограмм по идентифицируемым пептидам образца [59]. В ходе работы алгоритма DIA-NN каждый пик элюирования ионов-предшественников описывается набором индексов, и через процедуру итераций на основе линейного классификатора определяется наилучший кандидат на пик элюирования того или иного иона. Ключевым этапом работы алгоритма является присвоение статистической значимости (*q*-value) идентифицированным предшественникам, которая рассчитывается для целевых и ложных кандидатов на основе характеристик соответствующих пиков элюирования с использованием глубокой нейронной сети. Возможности алгоритма DIA-NN для использования в ультракоротком полнопротеомном анализе в полной мере были продемонстрированы на примере реализации метода ScanningSWATH [60, 61]. Технически этот метод заключается в замене последовательного выбора окон изоляции пептидов, в которых осуществляется фрагментация, непрерывным сканированием в широком окне *m/z* первого изолирующего ионы радиочастотного квадрупольного масс-спектрометра по всему диапазону масс с одновременной фрагментацией поступающих из него ионов-предшественников в столкновительном радиочастотном квадруполе. Таким образом, создаётся дополнительная размер-

ность для сопоставления спектров фрагментации и кандидатов-предшественников в последующей деконволюции сильно интерферирующих тандемных масс-спектров и идентификации пептидов алгоритмом DIA-NN. В недавней совместной работе создателей метода ScanningSWATH и разработчиков алгоритма DIA-NN была продемонстрирована ранее недостижимая возможность его использования в режиме ультракоротких, порядка 0,5–5 мин, градиентов ВЭЖХ-разделения протеолитических смесей пептидов при глубине анализа протеомов клеточных линий человека, достигающей нескольких тысяч белков [61]. Следует отметить, что для работы метода в режиме ультракоротких градиентов с целью сохранения разрешающей способности хроматографического разделения сложных смесей необходимо использование микропоточковой ВЭЖХ (порядка нескольких сотен мкл/мин), что, в свою очередь, приводит к большим затратам образца (до нескольких мкг гидролизата протеомов клеточных линий человека).

Реализация метода DIA в режиме ультракороткого разделения наиболее естественна для времяпролётных масс-анализаторов, которые в настоящее время позволяют получать масс-спектры в широком диапазоне m/z с разрешением пиков в спектрах 50 000 и выше и частотой сканирования порядка 100 Гц, что на порядок превосходит производительность масс-анализаторов высокого разрешения на основе ионных ловушек. Одним из примеров использования времяпролётного масс-анализатора высокого разрешения для ультракороткого полнопротеомного анализа является комбинация дополнительного разделения ионов-предшественников по ионной подвижности в режиме удерживания ионов в градиенте электрического поля (TIMS, Trapped Ion Mobility Spectrometry [62, 63]) с методом параллельного накопления и последовательной фрагментации ионов пептидов (PASEF, Parallel Accumulation SErial Fragmentation [64]). В TIMS ионы пептидов, элюируемых с колонки ВЭЖХ и ионизируемых в источнике ионизации, поступают в дрейфовую камеру, в которой удерживаются в радиальном направлении постоянным электрическим полем, компенсирующим их дрейф в столкновениях с молекулами газа-носителя, и, соответственно, разделяются по ионной подвижности. Вместо выбора массы одного иона-предшественника для фрагментации реализуется синхронизированное с работой камер ионной подвижности изменение параметров радиочастотного квадрупольного поля для изолирования и фрагментации ионов с выделенными значениями или диапазоном m/z . Один шаг изменения напряжённости удерживающего электрического поля в камере TIMS длительностью

50 мс позволяет получить спектры фрагментации нескольких ионов пептидов. PASEF многократно увеличивает скорость наработки спектров фрагментации без потери чувствительности анализа [65]. Реализация комбинации TIMS-TOP/PASEF в режиме DIA-анализа (dia-PASEF), а также использование алгоритма DIA-NN для обработки данных продемонстрировало возможность осуществления полнопротеомного анализа с глубиной в несколько тысяч идентифицированных белков клеточных линий человека со скоростью до 400 образцов в сутки инструментального времени (3 минуты градиента ВЭЖХ) [66].

МЕТОД ПРЯМОГО ВВОДА ОБРАЗЦА DISPA

Одним из логичных шагов в развитии методов ультракороткой протеомики и упрощения инструментальной составляющей самого анализа является отказ от хроматографического разделения протеолитических смесей в реальном времени. Сам подход, конечно, не является чем-то уникальным и использовался в практике анализа протеомов более полутора десятков лет [67]. Однако его ранние реализации осуществлялись на масс-анализаторах низкого разрешения и точности измерения масс при отсутствии как развитых инструментов поиска идентификаций, так и каких-либо возможностей дополнительного разделения ионов, например, по ионной подвижности. Несколько лет назад с развитием масс-спектрометрических технологий, появлением масс-анализаторов высокого разрешения, а также методов быстрого разделения ионов в газовой фазе по ионной подвижности концепция прямой подачи протеолитической смеси в источник ионизации без предварительного хроматографического разделения приобрела новое звучание в методе скорострельного протеомного анализа на основе прямого ввода (DISPA, Direct Infusion Shotgun Proteomic Analysis) [68]. Технически реализация метода довольно проста: протеолитическая смесь подаётся в нанопотоковом режиме напрямую со шприца с пробой в источник ионизации масс-спектрометра. В качестве дополнительной размерности используется разделение ионов по ионной подвижности. Сам анализ осуществляется методом DIA. Понятно, что мультиплексность спектров фрагментации в этом случае более чем на порядок выше, чем в случае разделения ВЭЖХ, что, соответственно, ограничивает достижимую глубину анализа протеома. Так, в указанной работе удалось достичь глубины порядка 500 белков для протеома клеточной линии человека. Однако эта глубина была достигнута за несколько минут экспериментального времени, что позволило

авторам проанализировать 132 образца за 4,5 часа (3 минуты на образец) с количественной идентификацией более 300 белков. Отсутствие возможности «привязки» пиков фрагментов к хроматографическим временам с целью эффективной деконволюции получаемых тандемных спектров с крайне высоким уровнем мультиплексности существенно ограничивало возможности метода DISPA при использовании стандартных, ВЭЖХ-ориентированных алгоритмов обработки данных. Чтобы преодолеть эти ограничения, было разработано программное обеспечение на основе алгоритма CsoDIAq (Cosine similarity optimization for DIA qualitative and Quantitative analysis [69]). Использование этого алгоритма для работы с данными DISPA позволило продемонстрировать глубину анализа протеомов клеточных линий человека (HeLa и 293T) порядка 2000 белков в одном прогоне образца с суммарными затратами экспериментального времени порядка нескольких минут [70]. Также в этой работе были продемонстрированы возможности метода для количественного анализа больших выборок образцов. В частности, 96 образцов протеома клеточных линий человека, обработанных лекарственным соединением, были проанализированы за 8 часов с глубиной около 1000 количественно идентифицированных белков. Следует отметить, что метод DISPA является интересной альтернативой стандартным подходам к ультракороткому полнопротеомному анализу на основе ВЭЖХ, однако находится в начальной стадии своего развития. Ограничения метода вытекают из крайней сложности протеолитических смесей (если речь идёт о полнопротеомном анализе), в которых одновременно могут присутствовать миллионы индивидуальных пептидных последовательностей в динамическом диапазоне концентраций, достигающем нескольких порядков (более 10 порядков в случае, например, протеома плазмы крови человека). Эти ограничения влияют как на общее число идентификаций, так и на точность количественного анализа. Высокий динамический диапазон концентраций пептидов приводит к сильному эффекту подавления ионов низкоконцентрационных пептидов в источнике ионизации и в накопительной ионной ловушке. К ограничениям метода следует отнести низкое покрытие последовательностей идентифицируемых белков из-за низкой эффективности деконволюции смешанных тандемных масс-спектров, а также сложности контроля уровня ложноположительных идентификаций в результатах анализа. Тем не менее метод DISPA продолжает демонстрировать свою потенциально высокую эффективность как метод ультрабыстрого протеомного анализа. Так, в совсем недавней работе была продемонстрирована производительность в

1000 образцов в сутки (1,4 минуты на образец) на примере профилирования белков плазмы крови в короне наночастиц с идентификацией 280 белков, из которых 44 являлись подтверждёнными биомаркерами различных патологий [71].

МЕТОД ПРЯМОЙ ИДЕНТИФИКАЦИИ БЕЛКОВ

Как уже было отмечено выше, одной из причин длительности классического полнопротеомного анализа с использованием метода DDA является необходимость получения спектров фрагментации от как можно большего количества ионов пептидов, элюируемых из хроматографической колонки. Доминирующим методом фрагментации является диссоциация пептидов в столкновениях с молекулами газа-носителя по связям -CO-NH- между остатками аминокислотной последовательности с образованием преимущественно *у*- и *б*-серий фрагментов. Процесс столкновительной диссоциации пептидов является эргодическим, требующим времени для разрыва связей. Кроме того, для повышения информативности спектров фрагментации и, соответственно, большей точности идентификации последовательности пептида желательнее получить как можно большее количество ионов фрагментов, что также требует времени на накопление изолируемых для этой цели ионов-предшественников. Таким образом, в процессе анализа протеолитической смеси на каждый MS1 в широком диапазоне *m/z*, регистрируемый для всех ионов пептидов, элюируемых в данный конкретный момент времени, в методе DDA идёт последовательный отбор ограниченного количества наиболее интенсивных пиков ионов-предшественников, их накопление и фрагментация. В случае протеомов сложных организмов анализируемая протеолитическая смесь может содержать миллионы индивидуальных пептидных последовательностей. Соответственно, чтобы увеличить количество отбираемых на фрагментацию ионов пептидов и, следовательно, глубину протеомного анализа, а также учесть временные ограничения, накладываемые скоростями накопления ионов и получения тандемных масс-спектров, используются многочасовые градиенты ВЭЖХ. Таким образом, для решения проблемы повышения производительности полнопротеомного анализа была создана концепция прямой масс-спектрометрической идентификации белков, в которой отсутствует стадия фрагментации ионов пептидов. Белки идентифицируются напрямую из MS1 на основе точно измеряемых масс всех представленных в них ионов с учётом специфичности гидролиза. При этом распределение интенсивностей

регистрируемых пиков иона пептида в масс-спектре, соответствующих разному содержанию в последовательности изотопа ^{13}C (спектральный профиль иона пептида), даёт информацию, специфичную по отношению к атомному составу пептида [72]. Также времена хроматографического элюирования являются специфичными по отношению к аминокислотным последовательностям пептидов [73–75], включая пептиды с модификациями остатков [76, 77]. Таким образом, MS1 позволяют извлекать набор комплементарных данных о последовательности и элементном составе пептидов. Понятно, что отказ от фрагментации пептидов снижает возможности их идентификации в силу существенно более низкой специфичности хроматографических времён к аминокислотной последовательности по сравнению с тандемными масс-спектрами. В то же время отказ от стадии фрагментации позволяет существенно сократить время анализа за счёт возможности использования коротких градиентов ВЭЖХ.

Концепция прямой масс-спектрометрической идентификации белков была реализована в методе DirectMS1, который уже в первых работах продемонстрировал глубину полнопротеомного анализа более 1000 белков при использовании 5-минутных градиентов разделения [78]. Схема реализации метода представлена на рис. 4. Масс-анализатор работает в режиме непрерывной регистрации MS1 в течение всего времени гради-

ентного элюирования протеолитической смеси. Скорость получения MS1 зависит от типа масс-анализатора и требований к разрешению масс в спектрах и точности их измерения. Эти требования высоки, и метод работает только с масс-анализаторами высокого разрешения: не менее 100 000 по массам и менее 1 ppm по точности их измерения. Важным условием эффективности метода является также производительность масс-анализатора. Так, современные масс-анализаторы на основе ионной ловушки Орбитреп позволяют получать до 10 спектров в секунду с указанными выше масс-спектральными характеристиками. Ещё одним ключевым фактором достижения высокой эффективности метода является точность предсказания времён элюирования пептидов, которая сравнивается с экспериментальными временами и используется в алгоритме обработки данных DirectMS1 для различения правильных и ложных идентификаций. До недавнего времени существовало несколько моделей и алгоритмов предсказания времён элюирования пептидов с точностью предсказания (по корреляции экспериментальных и предсказанных времён) $R^2 \sim 0,96$ [32]. В последние годы с развитием алгоритмов машинного обучения появились модели предсказания времён элюирования пептидов на их основе с существенно более высокой точностью. В частности, в алгоритме поиска DirectMS1 используется модель предсказания DeepLC,

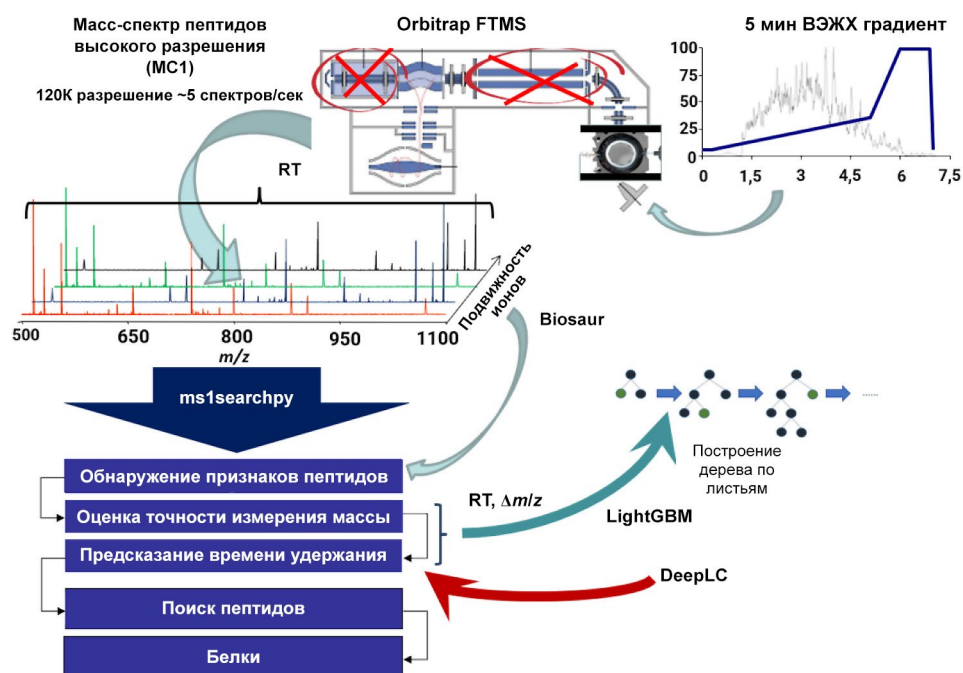


Рис. 4. Схема реализации метода DirectMS1. Ключевым фактором эффективности метода является использование алгоритмов машинного обучения для предсказания времён элюирования пептидов и классификации правильных и ложных идентификаций на основе совокупности комплементарных данных, таких как спектральный профиль иона пептида по изотопу углерода ^{13}C , времена элюирования, ионная подвижность (при использовании дополнительного разделения по ионной подвижности) и точные массы пептидов

которая позволила существенно увеличить глубину анализа протеома до более 2000 при использовании 5-минутных градиентов ВЭЖХ и 7,5 минут общих затрат времени на один экспериментальный прогон [79]. Дополнительное увеличение количества идентифицируемых белков даёт разделение пептидов по ионной подвижности. Реализация метода DirectMS1 не требует каких-либо существенных изменений в инструментальной составляющей, за исключением необходимости использования более высоких потоков ВЭЖХ, до 1 мкл/мин и более, с целью поддержания хроматографического разрешения в условиях ультракоротких градиентов. Ключевыми элементами метода являются программные средства для обработки масс-спектров ионов пептидов, ранжирование идентификаций, соотнесение их с белками в соответствующих базах данных и определение уровня достоверности. Эту работу выполняют алгоритмы определения спектральных профилей ионов пептидов в масс-спектрах первого уровня, такие как Biosaur [80] и алгоритм прямой идентификации белков ms1searchpy [81], основанный на использовании алгоритмов машинного обучения и интегрированный с моделями предсказания времён элюирования пептидов. Следует отметить, что недостатком метода прямой масс-спектрометрической идентификации белков DirectMS1 является отсутствие контроля FDR на пептидном уровне. По оценкам авторов метода уровень ложноположительных идентификаций пептидов может достигать 30% [78]. Важно, что в режиме ультракоротких градиентов, в отличие от подходов на основе МС/МС, метод DirectMS1 позволяет идентифицировать белки с существенно (почти на порядок) большим покрытием их аминокислотных последовательностей. Это, в свою очередь, даёт более точные измерения относительного содержания белков. В частности, было показано, что, несмотря на меньшую глубину покрытия протеома, метод DirectMS1 позволяет получать результаты количественной протеомики ультракороткого анализа, включая идентификацию дифференциально экспрессированных белков в задачах поиска специфического отклика клеток на химическое или лекарственное воздействие [82]. Эти результаты сравнимы с полученными методами DIA и DDA на основе длинных градиентов ВЭЖХ [20].

ЗАКЛЮЧЕНИЕ

На данный момент происходит активное развитие полнопротеомного масс-спектрометрического анализа и его использование во многих

областях постгеномных исследований. При этом производительность анализа, который в типичном случае занимает часы экспериментального времени на количественное профилирование одной протеолитической смеси, является одним из основных сдерживающих факторов использования протеомики в большом круге современных областей биомедицины, включая разработку новых и перетаргетирование существующих лекарственных препаратов, персонализированную медицину, популяционную и клиническую протеомику, протеомику единичных клеток и др. При этом развитие технологий высокопроизводительной масс-спектрометрии высокого разрешения и точности измерения масс, а также новых методов обработки масс-спектрометрических данных, основанных на алгоритмах машинного обучения, позволило вывести производительность на уровень нескольких сотен полнопротеомных анализов в сутки. Эти возможности в протеомике были реализованы буквально за несколько последних лет, и в настоящее время методы ультракороткой протеомики испытывают период бурного развития и становятся доминирующими подходами в решении многих из перечисленных выше задач. Такие методы, как DIA, DirectMS1 и DISPA, позволяют не только сократить время анализа более чем на порядок, но и увеличить его глубину до невозможных ещё десятилетие назад 2000–5000 белков, идентифицируемых за 3–5 минут суммарных затрат экспериментального времени. Дальнейшее развитие технологий и методов ультракороткого протеомного анализа позволит с меньшими затратами времени проводить масштабные исследования на больших выборках образцов, что позволяет более точно определять механизмы взаимодействия белков между собой и изменения в клетках на протеомном уровне в результате патологического процесса, химиотерапевтического или внешнего воздействия.

Вклад авторов. И.И. Федоров, С.А. Протасов – подбор и анализ литературы по теме обзора, написание текста обзора; И.А. Тарасова – обсуждение темы и структуры обзора; М.В. Горшков – руководство работой над обзором, написание и редактирование текста.

Финансирование. Работа выполнена при финансовой поддержке Российского научного фонда, грант № 20-14-00229-П.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Соблюдение этических норм. Настоящая статья не содержит описания каких-либо исследований с участием людей или животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Wang, H., Dey, K. K., Chen, P. C., Li, Y., Niu, M., Cho, J. H., Wang, X., Bai, B., Jiao, Y., Chepyala, S. R., Haroutunian, V., Zhang, B., Beach, T. G., and Peng, J. (2020) Integrated analysis of ultra-deep proteomes in cortex, cerebrospinal fluid and serum reveals a mitochondrial signature in Alzheimer's disease, *Mol. Neurodegener.*, **15**, 43, <https://doi.org/10.1186/s13024-020-00384-6>.
2. Radovanovic, M., Day, R. O., Jones, G. D. R., Galettis, P., and Norris, R. L. G. (2022) LC-MS/MS method for simultaneous quantification of ten antibiotics in human plasma for routine therapeutic drug monitoring, *J. Mass Spectrom. Adv. Clin. Lab.*, **26**, 48-59, <https://doi.org/10.1016/j.jmsacl.2022.11.001>.
3. Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S. N., and Aebersold, R. (2006) The PeptideAtlas project, *Nucleic Acids Res.*, **34**, D655-D658, <https://doi.org/10.1093/nar/gkj040>.
4. Henzel, W. J., Watanabe, C., and Stults, J. T. (2003) Protein identification: the origins of peptide mass fingerprinting, *J. Am. Soc. Mass Spectrom.*, **14**, 931-942, [https://doi.org/10.1016/S1044-0305\(03\)00214-9](https://doi.org/10.1016/S1044-0305(03)00214-9).
5. Lay, J. O., Jr. (2001) MALDI-TOF mass spectrometry of bacteria, *Mass Spectrom. Rev.*, **20**, 172-194, <https://doi.org/10.1002/mas.10003>.
6. Aggarwal, S., and Yadav, A. K. (2016) False Discovery Rate Estimation in Proteomics, in *Statistical Analysis in Proteomics* (Jung, K., ed) Springer, New York, NY, pp. 119-128.
7. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, *Nat. Methods*, **4**, 207-214, <https://doi.org/10.1038/nmeth1019>.
8. Shen, Y., Tolić, N., Masselon, C., Pasa-Tolić, L., Camp, D. G., 2nd, Hixson, K. K., Zhao, R., Anderson, G. A., and Smith, R. D. (2004) Ultrasensitive proteomics using high-efficiency on-line micro-SPE-nanoLC-nanoESI MS and MS/MS, *Anal. Chem.*, **76**, 144-154, <https://doi.org/10.1021/ac030096q>.
9. Shen, Y., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., and Smith, R. D. (2002) High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics, *Anal. Chem.*, **74**, 4235-4249, <https://doi.org/10.1021/ac0202280>.
10. Richards, A. L., Hebert, A. S., Ulbrich, A., Bailey, D. J., Coughlin, E. E., Westphall, M. S., and Coon, J. J. (2015) One-hour proteome analysis in yeast, *Nat. Protoc.*, **10**, 701-714, <https://doi.org/10.1038/nprot.2015.040>.
11. Hebert, A. S., Prasad, S., Belford, M. W., Bailey, D. J., McAlister, G. C., Abbatiello, S. E., Huguet, R., Wouters, E. R., Dunyach, J. J., Brademan, D. R., Westphall, M. S., and Coon, J. J. (2018) Comprehensive single-shot proteomics with FAIMS on a hybrid orbitrap mass spectrometer, *Anal. Chem.*, **90**, 9529-9537, <https://doi.org/10.1021/acs.analchem.8b02233>.
12. Zubarev, R. A. (2013) The challenge of the proteome dynamic range and its implications for in-depth proteomics, *Proteomics*, **13**, 723-726, <https://doi.org/10.1002/pmic.201200451>.
13. Doerr, A. (2015) DIA mass spectrometry, *Nat. Methods*, **12**, 35-35, <https://doi.org/10.1038/nmeth.3234>.
14. Geyer, P. E., Kulak, N. A., Pichler, G., Holdt, L. M., Teupser, D., and Mann, M. (2016) Plasma proteome profiling to assess human health and disease, *Cell Syst.*, **2**, 185-195, <https://doi.org/10.1016/j.cels.2016.02.015>.
15. Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., and Zubarev, R. A. (2013) Rapid and deep human proteome analysis by single-dimension shotgun proteomics, *Mol. Cell Proteomics*, **12**, 3330-3338, <https://doi.org/10.1074/mcp.O113.028787>.
16. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line, *Mol. Syst. Biol.*, **7**, 548, <https://doi.org/10.1038/msb.2011.81>.
17. Masselon, C., Anderson, G. A., Harkewicz, R., Bruce, J. E., Pasa-Tolic, L., and Smith, R. D. (2000) Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures, *Anal. Chem.*, **72**, 1918-1924, <https://doi.org/10.1021/ac991133+>.
18. Chapman, J. D., Goodlett, D. R., and Masselon, C. D. (2014) Multiplexed and data-independent tandem mass spectrometry for global proteome profiling, *Mass Spectrom. Rev.*, **33**, 452-470, <https://doi.org/10.1002/mas.21400>.
19. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K., and Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS, *Anal. Chem.*, **75**, 1895-1904, <https://doi.org/10.1021/ac0262560>.
20. Ivanov, M. V., Bubis, J. A., Gorshkov, V., Tarasova, I. A., Levitsky, L. I., Solovyeva, E. M., Lipatova, A. V., Kjeldsen, F., and Gorshkov, M. V. (2022) DirectMS1Quant: ultrafast quantitative proteomics with MS/MS-free mass spectrometry, *Anal. Chem.*, **94**, 13068-13075, <https://doi.org/10.1021/acs.analchem.2c02255>.
21. Aebersold, R., and Goodlett, D. R. (2001) Mass spectrometry in proteomics, *Chem. Rev.*, **101**, 269-295, <https://doi.org/10.1021/cr990076h>.

22. Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting, *Curr. Biol.*, **3**, 327-332, [https://doi.org/10.1016/0960-9822\(93\)90195-t](https://doi.org/10.1016/0960-9822(93)90195-t).
23. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics, *Nature*, **422**, 198-207, <https://doi.org/10.1038/nature01511>.
24. Gevaert, K., and Vandekerckhove, J. (2000) Protein identification methods in proteomics, *Electrophoresis*, **21**, 1145-1154, [https://doi.org/10.1002/\(SICI\)1522-2683\(20000401\)21:6<1145::AID-ELPS1145>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1522-2683(20000401)21:6<1145::AID-ELPS1145>3.0.CO;2-Z).
25. He, Z., Yang, C., Yang, C., Qi, R. Z., Tam, J. P., and Yu, W. (2010) Optimization-based peptide mass fingerprinting for protein mixture identification, *J. Comput. Biol.*, **17**, 221-235, <https://doi.org/10.1089/cmb.2009.0160>.
26. Conrads, T. P., Anderson, G. A., Veenstra, T. D., Pasa-Tolić, L., and Smith, R. D. (2000) Utility of accurate mass tags for proteome-wide protein identification, *Anal. Chem.*, **72**, 3349-3354, <https://doi.org/10.1021/ac0002386>.
27. Norbeck, A. D., Monroe, M. E., Adkins, J. N., Anderson, K. K., Daly, D. S., and Smith, R. D. (2005) The utility of accurate mass and LC elution time information in the analysis of complex proteomes, *J. Am. Soc. Mass Spectrom.*, **16**, 1239-1249, <https://doi.org/10.1016/j.jasms.2005.05.009>.
28. Shen, Y., Strittmatter, E. F., Zhang, R., Metz, T. O., Moore, R. J., Li, F., Udseth, H. R., Smith, R. D., Unger, K. K., Kumar, D., and Lubda, D. (2005) Making broad proteome protein measurements in 1-5 min using high-speed RPLC separations and high-accuracy mass measurements, *Anal. Chem.*, **77**, 7763-7773, <https://doi.org/10.1021/ac051257o>.
29. Shen, Y., Smith, R. D., Unger, K. K., Kumar, D., and Lubda, D. (2005) Ultrahigh-throughput proteomics using fast RPLC separations with ESI-MS/MS, *Anal. Chem.*, **77**, 6692-6701, <https://doi.org/10.1021/ac050876u>.
30. Lange, E., Tautenhahn, R., Neumann, S., and Gröpl, C. (2008) Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements, *BMC Bioinformatics*, **9**, 375, <https://doi.org/10.1186/1471-2105-9-375>.
31. Petritis, K., Kangas, L. J., Yan, B., Monroe, M. E., Strittmatter, E. F., Qian, W. J., Adkins, J. N., Moore, R. J., Xu, Y., Lipton, M. S., Camp, D. G., 2nd, and Smith, R. D. (2006) Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information, *Anal. Chem.*, **78**, 5026-5039, <https://doi.org/10.1021/ac060143p>.
32. Tarasova, I. A., Masselon, C. D., Gorshkov, A. V., and Gorshkov, M. V. (2016) Predictive chromatography of peptides and proteins as a complementary tool for proteomics, *Analyst*, **141**, 4816-4832, <https://doi.org/10.1039/c6an00919k>.
33. Tarasova, I. A., Guryca, V., Pridatchenko, M. L., Gorshkov, A. V., Kieffer-Jaquinod, S., Evreinov, V. V., Masselon, C. D., and Gorshkov, M. V. (2009) Standardization of retention time data for AMT tag proteomics database generation, *J. Chromatogr. B*, **877**, 433-440, <https://doi.org/10.1016/j.jchromb.2008.12.047>.
34. Han, X., Aslanian, A., and Yates, J. R. (2008) Mass spectrometry for proteomics, *Curr. Opin. Chem. Biol.*, **12**, 483-490, <https://doi.org/10.1016/j.cbpa.2008.07.024>.
35. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., and Yates, J. R., 3rd (2013) Protein analysis by shotgun/bottom-up proteomics, *Chem. Rev.*, **113**, 2343-2394, <https://doi.org/10.1021/cr3003533>.
36. Miller, R. M., and Smith, L. M. (2023) Overview and considerations in bottom-up proteomics, *Analyst*, **148**, 475-486, <https://doi.org/10.1039/d2an01246d>.
37. Wang, H., Yang, Y., Li, Y., Bai, B., Wang, X., Tan, H., Liu, T., Beach, T. G., Peng, J., and Wu, Z. (2015) Systematic optimization of long gradient chromatography mass spectrometry for deep analysis of brain proteome, *J. Proteome Res.*, **14**, 829-838, <https://doi.org/10.1021/pr500882h>.
38. Michalski, A., Cox, J., and Mann, M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS, *J. Proteome Res.*, **10**, 1785-1793, <https://doi.org/10.1021/pr101060v>.
39. Kreimer, S., Belov, M. E., Danielson, W. F., Levitsky, L. I., Gorshkov, M. V., Karger, B. L., and Ivanov, A. R. (2016) Advanced precursor ion selection algorithms for increased depth of bottom-up proteomic profiling, *J. Proteome Res.*, **15**, 3563-3573, <https://doi.org/10.1021/acs.jproteome.6b00312>.
40. Hebert, A. S., Thöing, C., Riley, N. M., Kwiecien, N. W., Shiskova, E., Huguet, R., Cardasis, H. L., Kuehn, A., Eliuk, S., Zabrouskov, V., Westphall, M. S., McAlister, G. C., and Coon, J. J. (2018) Improved precursor characterization for data-dependent mass spectrometry, *Anal. Chem.*, **90**, 2333-2340, <https://doi.org/10.1021/acs.analchem.7b04808>.
41. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells, *Nat. Methods*, **11**, 319-324, <https://doi.org/10.1038/nmeth.2834>.
42. Zhang, B., Pirmoradian, M., Chernobrovkin, A., and Zubarev, R. A. (2014) DeMix workflow for efficient identification of cofragmented peptides in high resolution data-dependent tandem mass spectrometry, *Mol. Cell Proteomics*, **13**, 3211-3223, <https://doi.org/10.1074/mcp.O114.038877>.

43. Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., and Smith, R. D. (2012) Mass spectrometry-based proteomics: existing capabilities and future directions, *Chem. Soc. Rev.*, **41**, 3912-3928, <https://doi.org/10.1039/c2cs15331a>.
44. Stanley, J. R., Adkins, J. N., Slys, G. W., Monroe, M. E., Purvine, S. O., Karpievitch, Y. V., Anderson, G. A., Smith, R. D., and Dabney, A. R. (2011) A statistical method for assessing peptide identification confidence in accurate mass and time tag proteomics, *Anal. Chem.*, **83**, 6135-6140, <https://doi.org/10.1021/ac2009806>.
45. Zimmer, J. S. D., Monroe, M. E., Qian, W.-J., and Smith, R. D. (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach, *Mass Spectrom. Rev.*, **25**, 450-482, <https://doi.org/10.1002/mas.20071>.
46. Dos Santos, A., Court, M., Thiers, V., Sar, S., Guettier, C., Samuel, D., Bréchet, C., Garin, J., Demaugre, F., and Masselon, C. D. (2010) Identification of cellular targets in human intrahepatic cholangiocarcinoma using laser microdissection and accurate mass and time tag proteomics, *Mol. Cell Proteomics*, **9**, 1991-2004, <https://doi.org/10.1074/mcp.M110.000026>.
47. Agron, I. A., Avtonomov, D. M., Kononikhin, A. S., Popov, I. A., Moshkovskii, S. A., and Nikolaev, E. N. (2010) Accurate mass tag retention time database for urine proteome analysis by chromatography-mass spectrometry, *Biochemistry (Moscow)*, **75**, 636-641, <https://doi.org/10.1134/s0006297910050147>.
48. Heil, L. R., Damoc, E., Arrey, T. N., Pashkova, A., Denisov, E., Petzoldt, J., Peterson, A. C., Hsu, C., Searle, B. C., Shulman, N., Riffle, M., Connolly, B., MacLean, B. X., Remes, P. M., Senko, M. W., Stewart, H. I., Hock, C., Makarov, A. A., Hermanson, D., Zabrouskov, V., Wu, C. C., and MacCoss, M. J. (2023) Evaluating the performance of the astral mass analyzer for quantitative proteomics using data-independent acquisition, *J. Proteome Res.*, **22**, 3290-3300, <https://doi.org/10.1021/acs.jproteome.3c00357>.
49. Guzman, U. H., Martinez-Val, A., Ye, Z., Damoc, E., Arrey, T. N., Pashkova, A., Renuse, S., Denisov, E., Petzoldt, J., Peterson, A. C., Harking, F., Østergaard, O., Rydbirk, R., Aznar, S., Stewart, H., Xuan, Y., Hermanson, D., Horning, S., Hock, C., Makarov, A., Zabrouskov, V., and Olsen, J. V. (2024) Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition, *Nat. Biotechnol.*, <https://doi.org/10.1038/s41587-023-02099-7>.
50. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, *Mol. Cell Proteomics*, **11**, O111.016717, <https://doi.org/10.1074/mcp.O111.016717>.
51. Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L. Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., and Reiter, L. (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues, *Mol. Cell Proteomics*, **14**, 1400-1410, <https://doi.org/10.1074/mcp.M114.044305>.
52. Kelstrup, C. D., Bekker-Jensen, D. B., Arrey, T. N., Hogrebe, A., Harder, A., and Olsen, J. V. (2018) Performance evaluation of the Q Exactive HF-X for shotgun proteomics, *J. Proteome Res.*, **17**, 727-738, <https://doi.org/10.1021/acs.jproteome.7b00602>.
53. Doellinger, J., Blumenschein, C., Schneider, A., and Lasch, P. (2020) Isolation window optimization of data-independent acquisition using predicted libraries for deep and accurate proteome profiling, *Anal. Chem.*, **92**, 12185-12192, <https://doi.org/10.1021/acs.analchem.0c00994>.
54. Doellinger, J., Blumenschein, C., Schneider, A., and Lasch, P. (2023) Increasing proteome depth while maintaining quantitative precision in short-gradient data-independent acquisition proteomics, *J. Proteome Res.*, **22**, 2131-2140, <https://doi.org/10.1021/acs.jproteome.3c00078>.
55. Bekker-Jensen, D. B., Martínez-Val, A., Steigerwald, S., Rütger, P., Fort, K. L., Arrey, T. N., Harder, A., Makarov, A., and Olsen, J. V. (2020) A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients, *Mol. Cell Proteomics*, **19**, 716-729, <https://doi.org/10.1074/mcp.TIR119.001906>.
56. Yang, Y., Liu, X., Shen, C., Lin, Y., Yang, P., and Qiao, L. (2020) In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics, *Nat. Commun.*, **11**, 146, <https://doi.org/10.1038/s41467-019-13866-z>.
57. Lou, R., Liu, W., Li, R., Li, S., He, X., and Shui, W. (2021) DeepPhospho accelerates DIA phosphoproteome profiling through in silico library generation, *Nat. Commun.*, **12**, 6685, <https://doi.org/10.1038/s41467-021-26979-1>.
58. Van Puyvelde, B., Willems, S., Gabriels, R., Daled, S., De Clerck, L., Vande Casteele, S., Staes, A., Impens, F., Deforce, D., Martens, L., Degroove, S., and Dhaenens, M. (2020) Removing the hidden data dependency of DIA with predicted spectral libraries, *Proteomics*, **20**, e1900306, <https://doi.org/10.1002/pmic.201900306>.
59. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., and Ralser, M. (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput, *Nat. Methods*, **17**, 41-44, <https://doi.org/10.1038/s41592-019-0638-x>.
60. Moseley, M. A., Hughes, C. J., Juvvadi, P. R., Soderblom, E. J., Lennon, S., Perkins, S. R., Thompson, J. W., Steinbach, W. J., Geromanos, S. J., Wildgoose, J., Langridge, J. I., Richardson, K., and Vissers, J. P. C. (2018) Scanning

- quadrupole data-independent acquisition, part A: qualitative and quantitative characterization, *J. Proteome Res.*, **17**, 770-779, <https://doi.org/10.1021/acs.jproteome.7b00464>.
61. Messner, C. B., Demichev, V., Bloomfield, N., Yu, J. S. L., White, M., Kreidl, M., Egger, A. S., Freiwald, A., Ivosev, G., Wasim, F., Zelezniak, A., Jürgens, L., Suttorp, N., Sander, L. E., Kurth, F., Lilley, K. S., Mülleder, M., Tate, S., and Ralser, M. (2021) Ultra-fast proteomics with Scanning SWATH, *Nat. Biotechnol.*, **39**, 846-854, <https://doi.org/10.1038/s41587-021-00860-4>.
62. Fernandez-Lima, F., Kaplan, D. A., Suetering, J., and Park, M. A. (2011) Gas-phase separation using a trapped ion mobility spectrometer, *Int. J. Ion Mobil. Spectrom.*, **14**, <https://doi.org/10.1007/s12127-011-0067-8>.
63. Ridgeway, M. E., Lubeck, M., Jordens, J., Mattias, M., and Melvin, A. P. (2018) Trapped ion mobility spectrometry: a short review, *Int. J. Mass Spectrom.*, **425**, 22-35, <https://doi.org/10.1016/j.ijms.2018.01.006>.
64. Meier, F., Beck, S., Grassl, N., Lubeck, M., Park, M. A., Raether, O., and Mann, M. (2015) Parallel Accumulation-Serial Fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device, *J. Proteome Res.*, **14**, 5378-5387, <https://doi.org/10.1021/acs.jproteome.5b00932>.
65. Meier, F., Brunner, A. D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N., Decker, J., Kosinski, T., Park, M. A., Bache, N., Hoerning, O., Cox, J., Räther, O., and Mann, M. (2018) Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer, *Mol. Cell Proteomics*, **17**, 2534-2545, <https://doi.org/10.1074/mcp.TIR118.000900>.
66. Szyrwiell, L., Gille, C., Mülleder, M., Demichev, V., and Ralser, M. (2024) Fast proteomics with dia-PASEF and analytical flow-rate chromatography, *Proteomics*, **24**, e2300100, <https://doi.org/10.1002/pmic.202300100>.
67. Chen, S. (2006) Rapid protein identification using direct infusion nanoelectrospray ionization mass spectrometry, *Proteomics*, **6**, 16-25, <https://doi.org/10.1002/pmic.200500043>.
68. Meyer, J. G., Niemi, N. M., Pagliarini, D. J., and Coon, J. J. (2020) Quantitative shotgun proteome analysis by direct infusion, *Nat. Methods*, **17**, 1222-1228, <https://doi.org/10.1038/s41592-020-00999-z>.
69. Cranney, C. W., and Meyer, J. G. (2021) CsoDIAq software for direct infusion shotgun proteome analysis, *Anal. Chem.*, **93**, 12312-12319, <https://doi.org/10.1021/acs.analchem.1c02021>.
70. Jiang, Y., Hutton, A., Cranney, C. W., and Meyer, J. G. (2023) Label-free quantification from Direct Infusion Shotgun Proteome Analysis (DISPA-LFQ) with CsoDIAq software, *Anal. Chem.*, **95**, 677-685, <https://doi.org/10.1021/acs.analchem.2c02249>.
71. Jiang, Y., and Meyer, J. G. (2024) 1.4 min plasma proteome profiling via nanoparticle protein corona and direct infusion mass spectrometry, *bioRxiv*, 2024.02.06.579213, <https://doi.org/10.1101/2024.02.06.579213>.
72. Miladinović, S. M., Kozhinov, A. N., Gorshkov, M. V., and Tsybin, Y. O. (2012) On the utility of isotopic fine structure mass spectrometry in protein identification, *Anal. Chem.*, **84**, 4042-4051, <https://doi.org/10.1021/ac2034584>.
73. Krokhnin, O. V. (2006) Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents, *Anal. Chem.*, **78**, 7785-7795, <https://doi.org/10.1021/ac060777w>.
74. Gorshkov, A. V., Tarasova, I. A., Evreinov, V. V., Savitski, M. M., Nielsen, M. L., Zubarev, R. A., and Gorshkov, M. V. (2006) Liquid chromatography at critical conditions: comprehensive approach to sequence-dependent retention time prediction, *Anal. Chem.*, **78**, 7770-7777, <https://doi.org/10.1021/ac060913x>.
75. Pridatchenko, M. L., Perlova, T. Y., Ben Hamidane, H., Goloborodko, A. A., Tarasova, I. A., Gorshkov, A. V., Evreinov, V. V., Tsybin, Y. O., and Gorshkov, M. V. (2012) On the utility of predictive chromatography to complement mass spectrometry based intact protein identification, *Anal. Bioanal. Chem.*, **402**, 2521-2529, <https://doi.org/10.1007/s00216-011-5350-3>.
76. Perlova, T. Y., Goloborodko, A. A., Margolin, Y., Pridatchenko, M. L., Tarasova, I. A., Gorshkov, A. V., Moskovets, E., Ivanov, A. R., and Gorshkov, M. V. (2010) Retention time prediction using the model of liquid chromatography of biomacromolecules at critical conditions in LC-MS phosphopeptide analysis, *Proteomics*, **10**, 3458-3468, <https://doi.org/10.1002/pmic.200900837>.
77. Sargaeva, N. P., Goloborodko, A. A., O'Connor, P. B., Moskovets, E., and Gorshkov, M. V. (2011) Sequence-specific predictive chromatography to assist mass spectrometric analysis of asparagine deamidation and aspartate isomerization in peptides, *Electrophoresis*, **32**, 1962-1969, <https://doi.org/10.1002/elps.201000507>.
78. Ivanov, M. V., Bubis, J. A., Gorshkov, V., Tarasova, I. A., Levitsky, L. I., Lobas, A. A., Solovyeva, E. M., Pridatchenko, M. L., Kjeldsen, F., and Gorshkov, M. V. (2020) DirectMS1: MS/MS-Free identification of 1000 proteins of cellular proteomes in 5 minutes, *Anal. Chem.*, **92**, 4326-4333, <https://doi.org/10.1021/acs.analchem.9b05095>.
79. Ivanov, M. V., Bubis, J. A., Gorshkov, V., Abdrakhimov, D. A., Kjeldsen, F., and Gorshkov, M. V. (2021) Boosting MS1-only proteomics with machine learning allows 2000 protein identifications in single-shot human proteome analysis using 5 min HPLC gradient, *J. Proteome Res.*, **20**, 1864-1873, <https://doi.org/10.1021/acs.jproteome.0c00863>.

80. Abdrakhimov, D. A., Bubis, J. A., Gorshkov, V., Kjeldsen, F., Gorshkov, M. V., and Ivanov, M. V. (2021) Biosaur: An open-source Python software for liquid chromatography-mass spectrometry peptide feature detection with ion mobility support, *Rapid Commun. Mass Spectrom.*, e9045, <https://doi.org/10.1002/rcm.9045>.
81. Ivanov, M. V., Tarasova, I. A., Levitsky, L. I., Solovyeva, E. M., Pridatchenko, M. L., Lobas, A. A., Bubis, J. A., and Gorshkov, M. V. (2017) MS/MS-free protein identification in complex mixtures using multiple enzymes with complementary specificity, *J. Proteome Res.*, **16**, 3989-3999, <https://doi.org/10.1021/acs.jproteome.7b00365>.
82. Solovyeva, E. M., Bubis, J. A., Tarasova, I. A., Lobas, A. A., Ivanov, M. V., Nazarov, A. A., Shutkov, I. A., and Gorshkov, M. V. (2022) On the feasibility of using an ultra-fast DirectMS1 method of proteome-wide analysis for searching drug targets in chemical proteomics, *Biochemistry (Moscow)*, **87**, 1342-1353, <https://doi.org/10.1134/S000629792211013X>.

ULTRAFast PROTEOMICS

Mini-Review

I. I. Fedorov^{1,2}, S. A. Protasov^{1,2}, I. A. Tarasova², and M. V. Gorshkov^{2*}

¹ Moscow Institute of Physics and Technology (National University),
141700 Dolgoprudny, Moscow Region, Russia

² Talrose Institute for Energy Problems of Chemical Physics,
N. N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences,
119334 Moscow, Russia; e-mail: mike.gorshkov@gmail.com

The current stage of development of proteomic research in the field of biology, medicine, development of new drugs, population screening, or personalized approaches to therapy dictates the need to analyze large sets of samples within the reasonable experimental time. Until recently, mass spectrometry measurements in proteomics were characterized as unique in identifying and quantifying cellular protein composition, but low throughput, requiring many hours to analyze a single sample. This was in conflict with the dynamics of changes in biological systems at the whole cellular proteome level upon the influence of external and internal factors. Thus, the low speed of whole proteome analysis has become the main factor limiting the developments in functional proteomics, where it is necessary to annotate the intracellular processes not only in a wide range of conditions, but also over a long period of time. The enormous level of heterogeneity of tissue cells or tumors, even of the same type, dictates the need to analyze the biological system at the level of individual cells. These studies involve obtaining molecular characteristics for tens, if not hundreds of thousands of individual cells, including their whole proteome profiles. The development of mass spectrometry technologies providing high resolution and mass measurement accuracy, predictive chromatography, new methods for peptide separation by ion mobility and processing proteomic data based on artificial intelligence algorithms have opened a way for significant, if not radical, increase in the throughput of whole proteome analysis and led to the implementation of the novel concept of ultrafast proteomics. Work done just in the last few years has demonstrated proteome-wide analysis throughput of several hundred samples per day at a depth of several thousand proteins, levels unimaginable three or four years ago. The review examines the background of these developments, as well as modern methods and approaches that implement ultrafast analysis of the entire proteome.

Keywords: proteomics, mass spectrometry, peptides, proteins, ultrafast analysis, quantitative proteomics