

УДК 621.384.668.8 + 537.534.3]: 663.1 + 579.2

© Н. В. Краснов, Я. И. Лютвинский, Е. П. Подольская

МАСС-СПЕКТРОМЕТРИЯ С МЯГКИМИ МЕТОДАМИ ИОНИЗАЦИИ В ПРОТЕОМНОМ АНАЛИЗЕ (ОБЗОР)

В обзоре рассмотрено использование масс-спектрометрии с мягкими методами ионизации в протеомике. Описаны основы масс-спектрометрического анализа белковых соединений, показаны возможности масс-спектрометрии при идентификации белков и пептидов, поиске и идентификации посттрансляционных модификаций, при сравнительном количественном анализе.

Кл. сл.: масс-спектрометрия с мягкими методами ионизации, идентификация белков, тандемная масс-спектрометрия, посттрансляционные модификации, базы данных белков

1. МАСС-СПЕКТРОМЕТРИЯ С МЯГКИМИ МЕТОДАМИ ИОНИЗАЦИИ В ИССЛЕДОВАНИИ БЕЛКОВ

Исследованием белкового комплекса всего организма как единого целого занимается научная дисциплина — протеомика. Различают множество направлений в протеомных исследованиях. Вот некоторые из них.

– Субклеточная протеомика [1, 2]: локализация местонахождения белков в живой клетке.

– Функциональная протеомика [3, 4]: определение функций белков и исследование межбелковых взаимодействий [5].

– Количественная дифференциация белков [6, 7]: определение разницы в экспрессии белков как реакции на изменение внешних условий, патологию или блокаду выбранного гена.

– Определение маркеров болезней [8]: выявление белков, измененный уровень экспрессии которых может послужить средством ранней, доклинической диагностики заболеваний.

– Определение "белков-мишеней" [9]: определение белков, целевое воздействие на которые фармацевтическими средствами может скорректировать течение болезней.

Каждое из этих направлений чрезвычайно важно как для получения фундаментального знания о природе жизни, так и для практических задач медицины и фармакологии. При проведении исследований в этих направлениях масс-спектрометрия стала одним из наиболее универсальных инструментов [10] при анализе сложных биологических проб, позволяющих успешно решать следующие типовые задачи биологической химии:

- идентификация белков [11];
- восстановление первичной аминокислотной последовательности белка [12, 13];
- выявление посттрансляционных модификаций [14];
- сравнительный количественный анализ белков [15, 16].

Объемы данных, получаемых в современных масс-спектрометрических экспериментах, достаточно велики — несколько тысяч спектров за час эксперимента, и их обработка невозможна или чрезвычайно затруднена без средств автоматизации. Поэтому на практике исследователи имеют дело в основном с результатами автоматической обработки массива данных и только для некоторых наиболее важных спектров прибегают к ручной верификации данных для уточнения результата.

В этой статье описаны общепринятые подходы к обработке и интерпретации данных, полученных в ходе масс-спектрометрического эксперимента для решения каждой из перечисленных выше задач.

1.1. Подготовка пробы к анализу

В протеомике исследуемыми объектами являются ткани и культуры клеток, содержащие чрезвычайно сложные смеси белков в большом диапазоне концентраций (6–9 порядков) [17]. Анализ таких сложных смесей стал возможен в результате развития высокоразрешающих методов разделения, таких как аффинное обогащение, SDS и 2D гель-электрофорез и др. [18]. Далее полученный белковый материал готовится для масс-спектрометрического анализа.

При масс-спектрометрическом анализе белков различают два подхода:

– "top-down": получение информации на основе масс-спектрометрического анализа целых, неповрежденных белковых молекул;

– "bottom-up": восстановление информации о белках за счет анализа отдельных пептидов этих белков.

Подход top-down используется редко и для специфических задач. Причиной этого служит недостаточная селективность подхода, поскольку для каждой массы белка можно подобрать множество белков-кандидатов с такой массой.

Далее в этом обзоре обсуждается подход bottom-up. Подробнее о подходе top-down можно узнать из статей, посвященных этому методу [19, 20].

Для получения смеси пептидов при подготовке масс-спектрометрического анализа проводится предварительный избирательный ферментативный гидролиз. Чаще всего для пробоподготовки в масс-спектрометрии используется трипсин. Этот фермент расщепляет белки по связи Lys – X и Arg – X (где X — любая аминокислота, отличная от Pro и Nup). В результате триптического гидролиза (трипсинолиза) получается набор пептидов массой 500–4000 дальтон (триптических пептидов). Подробные описания этой методики подготовки даны в [21]. В некоторых случаях используются и другие агенты избирательного гидролиза.

Таким образом, объектом масс-спектрометрического анализа становится смесь пептидов, являющихся фрагментами исходных белков.

1.2. Методы ионизации

Известно, что пептиды и белки относятся к классу термолабильных, нелетучих соединений. Масс-спектрометрический анализ данных соединений требует применения мягких методов ионизации молекул, позволяющих перевести молекулы в газовую фазу без разрушения ковалентных связей. В данное время применяются два основных метода ионизации, создание которых было отмечено в 2002 г. Нобелевской премией по химии.

Источник с ионизацией электрораспылением, электроспрей (ElectroSpray Ionization source, ESI) [22–26]: образование ионов идет за счет распыления раствора веществ в присутствии сильных электрических полей при атмосферном давлении. Поток анализируемого раствора, поступающего в распылитель, варьируется в пределах от 5 нл (*наноспрей* [27]) до 100 мкл в минуту. Прибор, снабженный таким источником, может быть легко состыкован с жидкостным хроматографом, что позволяет выполнять дополнительное разделение смеси во время эксперимента. В результате стыковки появляется возможность обеспечить анализ смесей более высокой сложности.

Источник с матрично-ассоциированной лазерной десорбцией/ионизацией (Matrix-Assisted Laser Desorption/Ionization, MALDI) [28, 29]: при реализации данного метода ионизации анализируемое вещество сокристаллизуется с матрицей, представляющей собой легколетучую слабую органическую кислоту, одновременно являющуюся источником протонов. При облучении полученных кристаллов короткими лазерными импульсами происходит совместный переход образца и матрицы в газообразное ионное состояние.

Каждому из перечисленных методов ионизации присущи свои недостатки. Например, при обработке данных существенно то, что для источника ионизации *электроспрей* характерно образование многозарядных ионов, т. е. ионов, образованных за счет присоединения более одного катиона. Это делает совокупность сигналов масс-спектра более сложной для интерпретации.

С другой стороны, в масс-спектрах MALDI нижний диапазон масс заселен сигналами, связанными с наличием матрицы, на фоне которых информативные сигналы пептидов могут быть потеряны.

1.3. Первичная обработка сигналов масс-спектра

На рис. 1 представлен фрагмент масс-спектра. Здесь показан исходный сигнал масс-спектрометра (а именно ESI-TOF-MS, MX 5303 разработки Института аналитического приборостроения РАН) в том виде, в каком он получен из системы регистрации масс-спектрометра [30]. Пики в сигнале соответствуют детектированным молекулам с одним молекулярным весом. Поскольку молекулы с одной молекулярной массой поступают на детектор с некоторой разницей по времени, пик имеет некоторую ширину.

На основании ширины пика рассчитывается такая характеристика масс-спектрометра, как разрешающая способность (Resolving Power, R). Эта характеристика рассчитывается как отношение значения m/z к ширине пика на его полувысоте. Для спектра на рис. 1 разрешающая способность прибора равна примерно 4000. Это означает, что молекулы с массой 4000 и 4001 дальтон будут представлены в масс-спектре двумя пиками, разрешенными (разделенными) до половины высоты этих пиков.

Существует множество специальных алгоритмов, которые на основании исходного сигнала распознают в нем пики и определяют значение соответствующей этому пику массы иона. Результаты работы одного из таких алгоритмов представлены на рис. 1 вертикальными линиями и метками пиков. Именно эти алгоритмы дают вторую

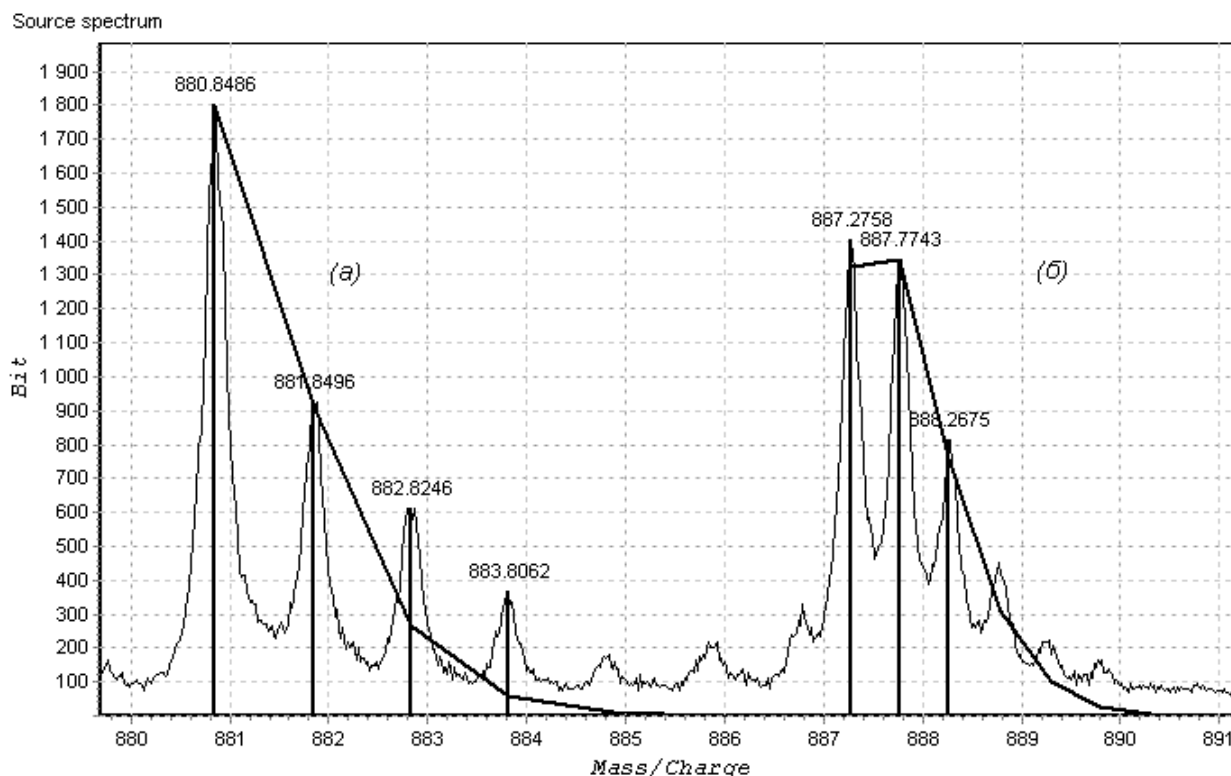


Рис. 1. Пример обработки исходных данных системы регистрации масс-спектрометра.
 (а) — изотопное распределение молекулярных масс однозарядного иона $[M+H^+] = 880.8486$ Да;
 (б) — изотопное распределение молекулярных масс двузарядного иона $[M+2H^+]/2 = 887.2758$ Да

важнейшую характеристику масс-спектрометра — точность определения масс. Точность измерения масс определяется как относительная средняя ошибка для массы, соответствующей пику. Поскольку для современных приборов ошибка невелика, обычно точность измеряют в ppm (point per million). Точность в 1 ppm означает, что молекулярные массы в районе 1000 Да будут определены с точностью в 0.001 Да.

На рис. 1 представлены сигналы двух видов молекулярных ионов с их изотопными распределениями. Наличие изотопного распределения объясняется тем, что в природе углерод существует в виде смеси изотопов: 99 % C^{12} и 1 % C^{13} . В молекулах пептидов около 60 % молекулярной массы вещества приходится на углерод. В случае пептида с молекулярной массой до 1000 Да вероятность того, что в молекуле будет 0 атомов C^{13} — около 40 %, 1 атом C^{13} — около 25 %, 2 атома C^{13} — около 14 %. Соответственно примерно такое же соотношение интенсивности сигналов изотопного распределения, полученного от молекулярного иона с массой 880.8486 Да, мы наблюдаем на (а) рис. 1. Первый пик соответствует моноизотопной молекулярной массе, т. е. массе пепти-

да, не содержащего ни одного атома C^{13} . Расстояние между пиками составляет примерно 1 Да, что соответствует разнице масс между изотопами C^{12} и C^{13} . Следует отметить, что сигнал (а) принадлежит однозарядному молекулярному иону $[M+H^+]$, т. е. в процессе ионизации незаряженная молекула присоединяет к себе один катион (в нашем случае протон H^+).

Для источника ионов электроспрей чем больше молекулярная масса пептидов, тем больше вероятность присоединения более одного катиона. На практике пептиды с массой более 1000 Да, как правило, представлены в спектре многозарядными ионами.

Существенно, что масс-спектрометры регистрируют не молекулярную массу, а отношение молекулярной массы иона к его зарядовому состоянию (масс-зарядовое отношение). Для ионов с одним присоединенным катионом масс-зарядовое отношение соответствует сумме молекулярных масс соединения и катиона $[M+H^+]$. Для ионов, образованных с присоединением произвольного количества протонов справедлива формула: $m = (M + H \cdot z) / z$, где M — масса пептида, m — зарегистрированное масс-зарядовое отношение, H — масса

протона, z — число присоединенных протонов.

Для многозарядных ионов расстояние между изотопными пиками будет обратно пропорционально зарядовому состоянию. Разница между пиками в сигнале (δ) на рис. 1 составляет 0.5 Да, что соответствует изотопному распределению, полученному от двухзарядного иона $[M+2H]^+$. Этому сигналу с масс-зарядовым отношением 887.2758 Да соответствует масса пептида, составляющая 1772.5516 Да.

С увеличением массы пептида растет среднее количество атомов углерода C^{13} на одну молекулу. Соответственно изменяется вид изотопного распределения, как демонстрирует (δ) на рис. 1.

На практике исследователям крайне редко приходится работать с исходными масс-спектрометрическими данными. Как уже говорилось, обработка исходных данных масс-спектрометра — это задача программного обеспечения первичной обработки спектров. Кроме алгоритмов распознавания пиков также существуют алгоритмы, предназначенные для декомпозиции изотопных и зарядных распределений ионов [31–33]. В частности, в данном случае для вычисления молекулярных масс использовалась статистическая информация о типичных изотопных распределениях пептидов. Ломаные линии на рис. 1 демонстрируют теоретическое среднее распределение изотопов в пептидах заданной массы, рассчитанное на основе баз данных белков.

Обработка первичных данных масс-спектров с учетом зарядового состояния и изотопного распределения молекулярных ионов позволяет отчетливо разделить сигналы ионов даже в случае интерференции (перекрывания) их изотопных распределений и представить эти сигналы как набор пар значений "моноизотопная масса—интенсивность сигнала" [33].

Таким образом, вместо исходных данных масс-спектрометра после первичной обработки масс-спектр исходной смеси соединений будет представлен как набор пар значений "масса—интенсивность сигнала". Масс-спектрометрические данные, представленные в такой форме, готовы для дальнейшей обработки с целью получения биологически значимого ответа.

1.4. Идентификация белков, метод PMF

Для идентификации белков средствами MS-анализа стандартом является методика, получившая в научной литературе название *peptide mass fingerprinting*, PMF [34]. Данная методика основана на том, что набор продуктов ферментативного гидролиза — пептидов — уникален для каждого белка подобно отпечаткам пальцев человека. Суть метода состоит в том, что для каждого белка с известной последовательностью рассчитываются

массы пептидов — продуктов теоретического ферментативного гидролиза с учетом указанного пользователем числа пропущенных сайтов гидролиза и возможных модификаций. Далее фактический масс-спектр продуктов гидролиза определяемого белка сравнивается со всеми теоретическими спектрами белков и выявляется белок, у которого степень соответствия наивысшая. Процедура расчета степени соответствия должна учитывать такие априорные данные, как число обнаруженных в масс-спектре пептидов, точность определения масс пептидов, массу белка и т. д. Пользователю выдается список белков по мере убывания степени соответствия.

Для решения этой задачи разработан ряд систем программного обеспечения, таких как Mascot [35], Sequest [36], X!Tandem [37] и многие другие. На рис. 2 приведен пример отчета, полученного от системы идентификации белков Mascot. Бычий бета-лактоглобулин (~19 кДа) был подвергнут триптическому гидролизу, и затем из полученного масс-спектра продуктов гидролиза был выделен набор, состоящий из 73 обнаруженных значений моноизотопных масс, который был введен в поисковую систему.

Отчет содержит данные о 20 наиболее вероятных белках для полученного масс-спектра. Только один белок (бета-лактаглобулин) показан на гистограмме, располагающейся в начале отчета, за пределами заштрихованной области, что показывает, что только этот белок был идентифицирован достаточно достоверно.

Сами по себе поисковые системы не пытаются делать какие-либо выводы о природе пробы. Их задача только соотнести полученные масс-спектры с белковыми последовательностями по некоторому алгоритму и оценить степень соответствия экспериментальных и рассчитанных спектров. Эта оценка всегда носит вероятностный характер и поэтому никогда не достигает 100%.

Главный недостаток метода PMF — это невозможность идентификации белка в смеси. Попытки адаптировать метод PMF для распознавания смеси белков предпринимались неоднократно, однако ни один метод не стал общепринятым. Применение метода PMF для смесей из 2–5 белков требует тщательного обоснования, обязательного учета априорной информации и большого количества ручной работы [38, 39].

2. БЕЛКОВЫЕ БАЗЫ ДАННЫХ

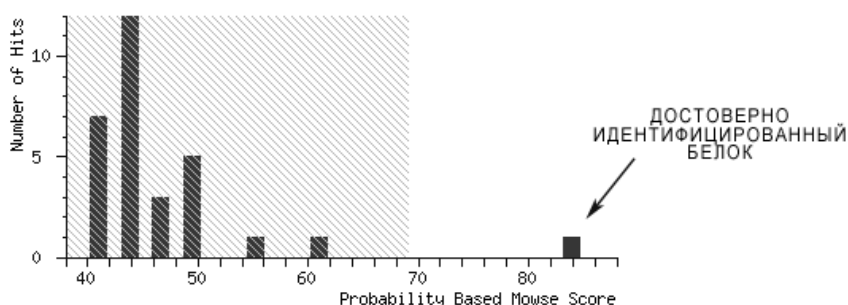
Исходя из задач, решаемых современной масс-спектрометрией, идентифицировать белок — значит соотнести данные масс-спектра с аминокислотной последовательностью белка.

{MATRIX}
{SCIENCE} **Mascot Search Results**

Database : NCBI nr 20051210 (3093834 sequences; 1063507484 residues)
Taxonomy : Mammalia (mammals) (445334 sequences)
Timestamp : 15 Dec 2005 at 00:49:39 GMT
Top Score : 84 for gi|49259423, Chain X, The Cys121ser Mutant Of Beta-Lactoglobulin

Probability Based Mowse Score

Ions score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Protein scores greater than 69 are significant ($p < 0.05$).



Concise Protein Summary Report

1.	gi 49259423	Mass: 18339	Score: 84	Expect: 0.0018	Queries matched: 10
	Chain X, The Cys121ser Mutant Of Beta-Lactoglobulin				
	gi 229460	Mass: 18355	Score: 82	Expect: 0.0027	Queries matched: 10
	lactoglobulin beta				
	gi 7245834	Mass: 18363	Score: 82	Expect: 0.0027	Queries matched: 10
	Chain A, Structural Changes Accompanying Ph-Induced Dissociation Of The B-Lactoglobulin Dimer				
				
	gi 125910	Mass: 19870	Score: 64	Expect: 0.16	Queries matched: 8
	Beta-lactoglobulin precursor (Beta-LG) (Allergen Bos d 5)				
	gi 72079	Mass: 18255	Score: 57	Expect: 0.89	Queries matched: 7
	beta-lactoglobulin - water buffalo				
	gi 3808097	Mass: 20010	Score: 53	Expect: 2.4	Queries matched: 7
	beta-lactoglobulin [Bubalus bubalis]				
2.	gi 73988428	Mass: 8016	Score: 61	Expect: 0.33	Queries matched: 9
	PREDICTED: similar to nuclear receptor interacting protein 3 isoform 2 [Canis familiaris]				
3.	gi 51829208	Mass: 32826	Score: 56	Expect: 1	Queries matched: 11
	PREDICTED: similar to PYRIN-containing APAF1-like protein 7 isoform 2 [Mus musculus]				
				

Рис. 2. Фрагмент отчета Mascot для метода PMF

Информация об аминокислотных последовательностях белков, пептидов накапливается мировым научным сообществом еще со времен открытия первых методов расшифровки аминокислотных последовательностей белков и пептидов, таких как деградация по Эдману [40], равно как и информация о нуклеотидных генетических последовательностях.

последовательностях.

В последние 10–15 лет рост объемов этой информации носит лавинообразный характер. За это время были установлены полные генетические последовательности нескольких десятков биологических видов, получены аминокислотные последовательности миллионов белков. Централь-

ный репозиторий генетической информации National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) недавно сообщил о достижении объема публично доступных генетических баз данных в 1011 базовых пар.

Развитие методов обработки этих данных позволяет говорить о появлении новой научной дисциплины, предметом которой является алгоритмическая обработка информации, содержащейся в биологических объектах, — вычислительной молекулярной биологии, или биоинформатики.

Сопоставление и интерпретация масс-спектрометрических данных и баз данных, составленных из известных биологических последовательностей, — одно из направлений исследований в области биоинформатики.

Для анализа белков в первую очередь интересны базы данных, содержащие информацию об аминокислотных последовательностях белков. На сегодняшний момент одной из самых надежных считается база данных SWISS-PROT [41] (www.expasy.org/sprot/), курируемая группой Швейцарского института биоинформатики. Данные о каждом белке, внесенном в эту базу, тщательно верифицированы и включают в себя такую информацию, как вариации структуры, функции белка в организме, ссылки на публикации и многое другое. На данный момент база данных SWISS-PROT сравнительно невелика и содержит около 190 000 аминокислотных последовательностей белков.

Другие базы, например TREMBL и PIR-PSD [42], содержат информацию о миллионах последовательностей. Такие базы данных составляются автоматически, как правило из данных, непосредственно предоставленных исследователями. При этом значительную часть базы данных могут составлять гипотетические белки — белки, существование которых предсказано при анализе генетических последовательностей.

Поскольку в настоящее время исследования белков ведутся сразу во многих научных центрах, часто бывает так, что несколько групп исследователей независимо друг от друга размещают в базах данных информацию об одних и тех же белках. Для того чтобы избежать дублирования информации при передаче в базу данных, последовательность проходит автоматическую проверку и аннотацию. Базы данных, проверенные таким способом, называются неизбыточными (non-redundant) и гарантируют, что в базе данных для каждого белка есть только одна последовательность [43]. Хорошим примером такой базы данных, собранной из различных источников, является NCBIInr [44], доступная по адресу (<ftp://ncbi.nlm.nih.gov/blast/db/>).

Кроме белковых баз данных при обработке данных масс-спектрометрии могут быть использованы и генетические базы данных. Например,

весьма популярны базы данных EST, Expressed Sequence Tags [45], содержащие последовательности, полученные на основе матричных РНК [46].

Также можно использовать информацию, полученную в ходе расшифровки генома человека, быка, крысы и т. д. Теоретически в полной генетической последовательности зашифрованы все возможные для организма белки, однако идентификация белков по этой информации осложняется неопределенностью рамки считывания, сложным механизмом трансляции белковой последовательности и другими факторами. Поэтому исследователи стараются избегать использования данного вида информации для белковых анализов.

Вопрос выбора базы данных для сопоставления данных масс-спектрометрического эксперимента — это всегда компромисс между надежностью и полнотой информации. Для актуальных исследований чаще всего используются базы данных класса NCBIInr. Данные SWISS-Prot надежны, но неполны — многие данные актуальных исследований в этой базе не представлены. При решении специфической задачи могут использоваться и некоторые подмножества генетических баз данных. Полные генетические базы данных используются редко из-за сложности интерпретации результатов и сильной избыточности данных.

Наиболее полный список биологических баз данных разного назначения можно найти в электронном каталоге DBCat [47], расположенном на (www.infobiogen.fr/services/dbcat/).

3. ТАНДЕМНАЯ МАСС-СПЕКТРОМЕТРИЯ

Для анализа смесей белков, по сложности превышающих аналитические возможности метода PMF, применяется тандемная масс-спектрометрия. Хорошее определение метода тандемного анализа приведено в русскоязычном обзоре, посвященном тандемной масс-спектрометрии [48]. Тандемная масс-спектрометрия (МС-МС) используется для структурного анализа и идентификации веществ в составе смесей.

Методика МС-МС состоит из следующих операций.

- Разделение в первой МС-ступени первичных, или "родительских", ионов и селекция ионов с единственным значением отношения массы к заряду (m/z).
- Фрагментация этих ионов с образованием разнообразных структурно значимых ионных фрагментов, называемых вторичными, или "дочерними", ионами.
- Масс-анализ дочерних ионов.

Существенно, что если при использовании метода PMF масс-спектр показывает набор триптических пептидов исследуемого белка, то в случае

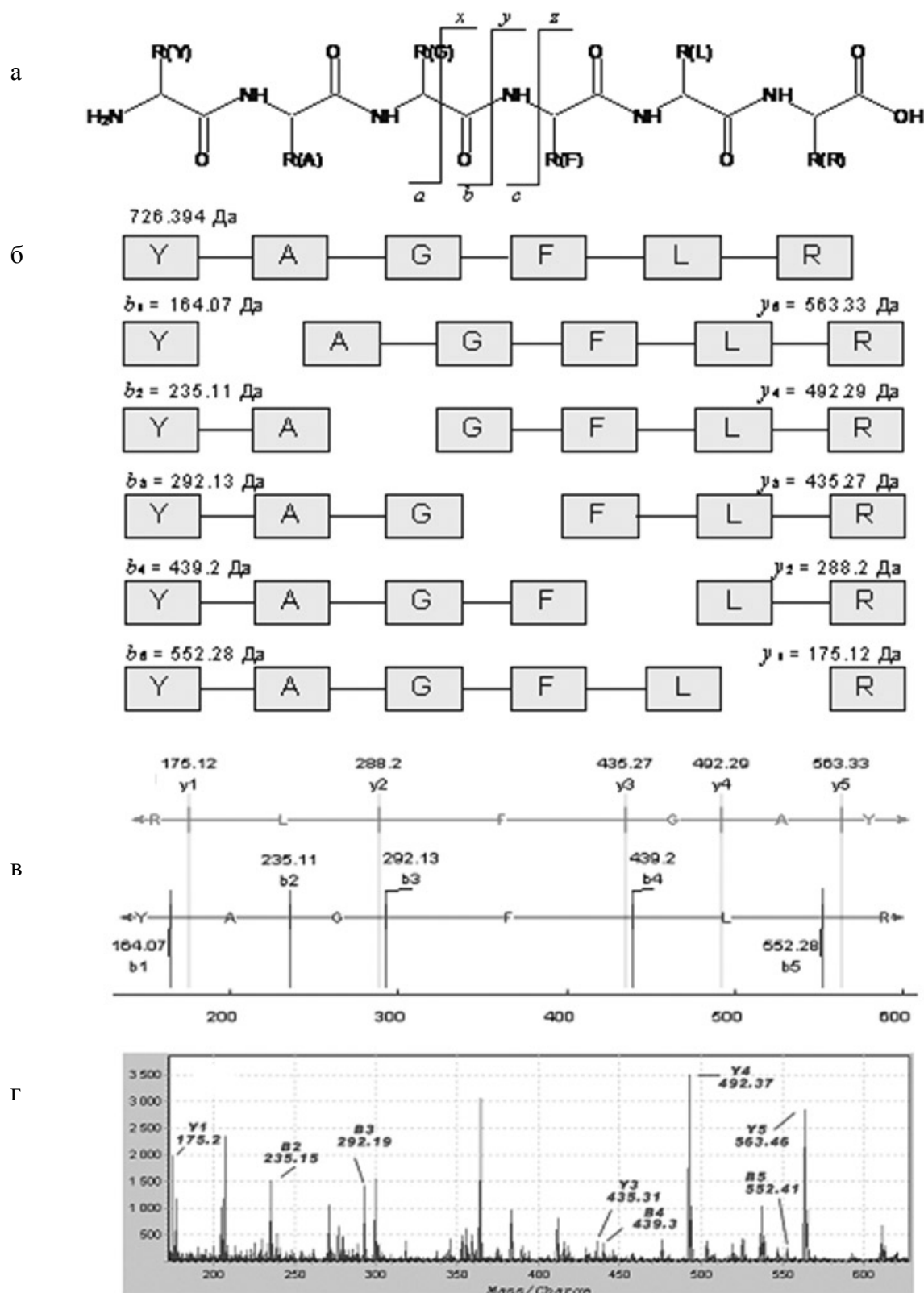


Рис. 3. Фрагментация пептидов на примере даларгина.
 а — нотация пептидных фрагментов; б — образование основных фрагментов; в — рассчитанный фрагментный масс-спектр; г — зарегистрированный фрагментный масс-спектр

танDEMной масс-спектрометрии каждый отдельный масс-спектр отображает набор фрагментов **одного пептида**.

В каждый момент времени из всех ионов, образующихся в источнике, для последующей фрагментации отбираются ионы только с одного соединения с заданным масс-зарядовым отношением. Такая селекция происходит при помощи специально разработанных алгоритмов DDA (Data Dependent Acquisition) на основе обработки данных масс-спектра родительских ионов. Данные одного MS-MS анализа, как правило, содержат множество фрагментных масс-спектров, каждый из которых характеризует фрагменты только одного соединения.

При анализе очень сложных смесей используют предварительное разделение смеси — чаще всего за счет применения методов хроматографии [49], для того чтобы исключить попадание в ячейку фрагментации ионов с близкими массами.

Состав фрагментов отражает структуру анализируемого вещества и напрямую зависит от способа фрагментации. В данное время наиболее часто используется столкновительная ячейка [50], в

которой ионы сталкиваются с нейтральными молекулами газа, заполняющего ячейку, что приводит к разрыву ковалентных связей в полипептидной цепи (рис. 3).

Для фрагментных ионов пептидов принята классификация, предложенная в 1984 году в работе [51], в соответствии с которой фрагменты, содержащие N-концевую аминокислоту, в зависимости от разорванной связи обозначают латинскими буквами *a*, *b* и *c*. Подобно фрагменты C-конца обозначены *x*, *y* и *z* (рис. 3, а).

Кроме разрыва полипептидной цепи происходят и такие процессы, как фрагментация по боковым радикалам, многократная фрагментация и ряд других процессов. Однако, несмотря на то что масс-спектр содержит множество сигналов помимо основных предполагаемых, в спектре, как правило, доминируют *b* и *y* ионы. После фрагментации достаточного количества родительских ионов накопленные сигналы *b* и *y* ионов могут образовывать полные наборы — серии. Пример приведен на рис. 3, в: для пептида, состоящего из шести аминокислот мы можем наблюдать по пять сигналов ионов, принадлежащих *b* и *y* сериям.

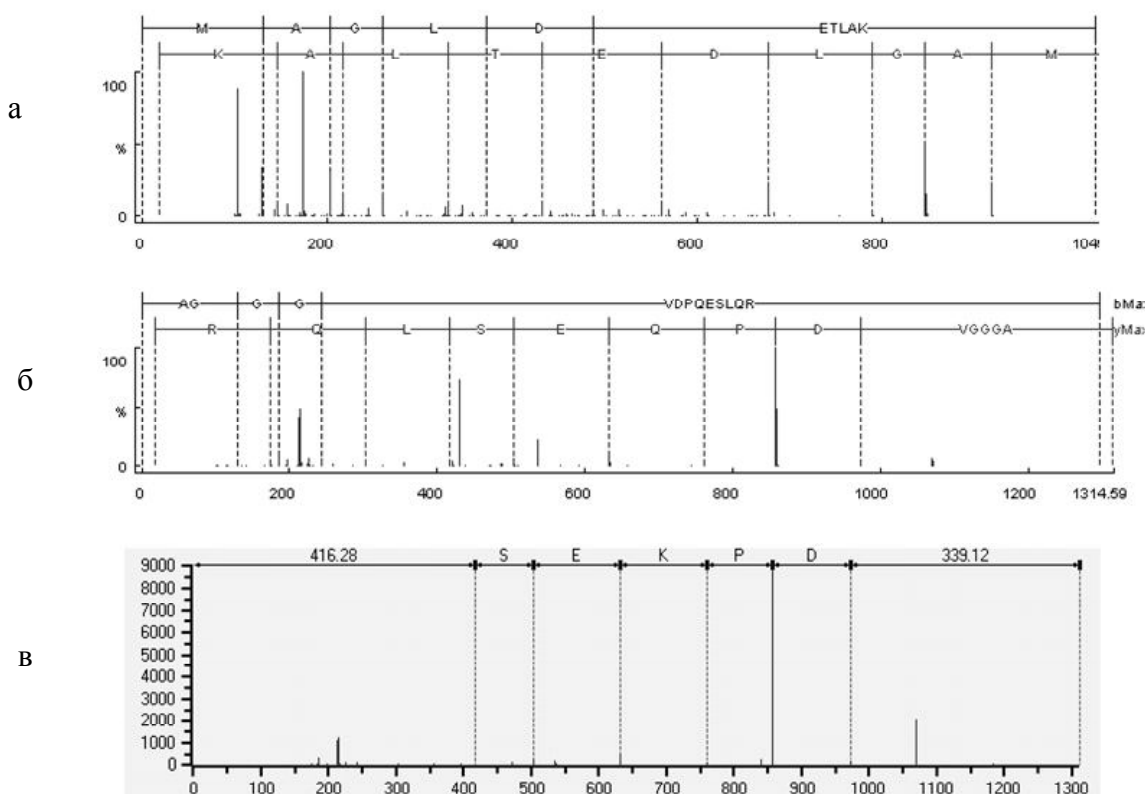


Рис. 4. Восстановление аминокислотной последовательности.

а — пептид MAGLDETIAK тяжелой цепи миозина, кролик; б — некорректная попытка распознавания de Novo пептида ELPDPQESIQR бычьего трансферина; в — масс-тег [416.28]SEKPD[339.12] пептида ELPDPQESIQR бычьего трансферина

При ручной обработке фрагментного масс-спектра зачастую удается обнаружить фрагменты серий *y* и *b* ионов и на основании этих серий восстановить последовательность фрагментированного пептида. Полученные таким образом данные пользуются наибольшим доверием. Однако число фрагментных масс-спектров от одного МС-МС эксперимента измеряется сотнями и тысячами, что вызывает необходимость автоматизации обработки данных МС-МС. Существует ряд программ, которые практически воспроизводят действия при ручной интерпретации фрагментных масс-спектров. В настоящее время используются два основных подхода.

3.1. Секвенирование de Novo

Секвенирование de Novo — определение полной последовательности пептида по его масс-спектру. По этому принципу работают такие программные продукты, как Peaks Studio [52], Lutefisk [53] и др. По массе родительского иона и наиболее интенсивным сигналам от дочерних ионов строится ряд предположений о том, какие пептиды могли бы дать подобный фрагментный масс-спектр. Для каждого предположенного пептида рассчитывается теоретический фрагментный масс-спектр и сопоставляется с фактическим. Затем выстраивается рейтинг, ранжирующий предположенные пептиды по степени соответствия (рис. 4).

Базовый недостаток этого метода состоит в том, что зачастую в спектре представлены не все фрагменты. Серии ионов могут быть представлены не полностью, а для восстановления пропущенных участков последовательности используются слабо обоснованные предположения. Например, спектр пептида на рис. 4, б, не содержит *y*-ионов в области масс более 1000 Да, из-за этого программам de Novo не удается восстановить часть аминокислотной последовательности пептида и предложить правильную версию распознавания. Дополнительные сложности возникают при выделении сигналов фрагментных ионов в сильно зашумленных областях спектра.

3.2. Поиск Sequence Tag

При использовании методов поиска Sequence Tag [54] в масс-спектре дочерних ионов выявляются гипотетические последовательности пиков (масс-теги), предположительно принадлежащие к одной из серий фрагментных ионов пептида. Таким образом, восстанавливается часть аминокислотной последовательности пептида. В качестве примера приведем тег, извлеченный из спектра, представленного на рис. 4, б, с точностью до замены аминокислот с близкими массами — [416.28] SEKPD [339.21].

Этот метод представляется более обоснованным, чем секвенирование de Novo, т. к. расшифровывается только часть последовательности пептида, но на основании надежной информации, и она будет корректна даже для модифицированных пептидов [55, 56]. Кроме того, реализация данного метода отличается малым временем вычисления. Но в то же время задача осложняется тем, что для насыщенного масс-спектра можно построить множество гипотетических тэгов и вместо полной последовательности восстанавливается только ее часть.

Качественная реализация поиска Sequence Tag, как и распознавания de Novo, является сложной задачей. Удачные алгоритмы для распознавания Sequence Tag стали появляться только в последнее время [57–59]; хотя в коммерческих пакетах программного обеспечения реализация алгоритма, как правило, присутствует, но ее качество оставляет желать лучшего.

Фрагменты аминокислотных последовательностей, полученные с помощью перечисленных методов и записанные в виде буквенного кода, могут быть использованы для текстовых методов поиска в базах данных, хорошо развитых в классической биоинформатике. Например, по набору пептидов можно найти ряд гомологов данного белка в базах данных и тем самым сделать обоснованное предположение о функции белка, даже если этот белок не был ранее описан [60].

3.3. Идентификация белков методами тандемной масс-спектрометрии

Основным преимуществом метода тандемной масс-спектрометрии является возможность идентификации белков без их предварительного разделения; число белков, идентифицированных из одной неразделенной смеси, может достигать 4000 [61]. В ходе типичного МС-МС-анализа триптического гидролизата смеси белков регистрируется до нескольких тысяч фрагментных масс-спектров отдельных триптических пептидов.

Полученные спектры сопоставляются с теоретическими масс-спектрами, рассчитанными на основании последовательностей всех возможных триптических пептидов для всех белков базы данных. Если найден хотя бы один пептид, принадлежащий некоторому белку, то этот белок присутствует в исходной смеси с некоторой долей вероятности. Если идентифицировано несколько пептидов одного белка, то вероятность того, что белок присутствует в исходной смеси, значительно возрастает.

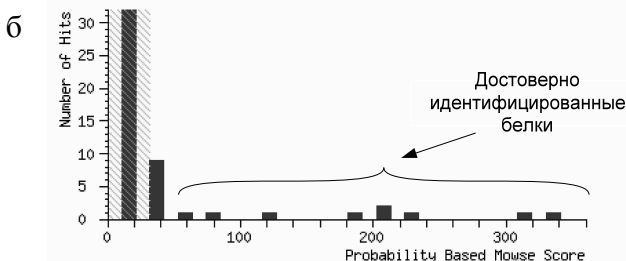
Информации, полученной из нескольких тысяч МС-МС-спектров, может быть достаточно для идентификации нескольких десятков и даже сотен белков, входящих в состав исходной смеси.

Significant hits:

gi 29135265	transferrin [<i>Bos taurus</i>]
gi 1664	unnamed protein product [<i>Oryctolagus cuniculus</i>]
gi 23306688	glutamate dehydrogenase 1 [<i>Bos taurus</i>]
gi 229552	albumin
gi 115698	Catalase
gi 27806851	lactoperoxydase [<i>Bos taurus</i>]
gi 68184	fructose-bisphosphate aldolase (EC 4.1.2.13) A - rabbit
gi 203480	creatine kinase
gi 34860177	similar to amylase 1; amylase 1, salivary [<i>Rattus norvegicus</i>]
gi 37546089	similar to RIKEN cDNA 4921529018 [<i>Homo sapiens</i>]
gi 28972560	mKIAA0979 protein [<i>Mus musculus</i>]
gi 130353	Phosphoglycerate mutase 2 (Phosphoglycerate mutase isozyme M)
gi 38083658	similar to pcy1 [<i>Mus musculus</i>]
gi 34098395	Kinectin

Probability Based Mowse Score

Score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 32 indicate identity or extensive homology ($p < 0.05$).



7. [gi|68184](#) **Mass:** 39643 **Total score:** 130 **Peptides matched:** 5
fructose-bisphosphate aldolase (EC 4.1.2.13) A - rabbit
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss Score	Rank	Peptide
<input checked="" type="checkbox"/> 487	401.2450	800.4755	800.4756	-0.0000	0	17	1 ALQASALK
<input checked="" type="checkbox"/> 626	470.7455	939.4765	939.4774	-0.0009	0	22	1 ELSDIAHR
<input checked="" type="checkbox"/> 716	547.2857	1092.5568	1092.5563	0.0005	1	33	1 AAQEEYVKR
<input checked="" type="checkbox"/> 733	567.2841	1132.5536	1132.5546	-0.0010	0	31	1 ALANSLACQEK + Deamidation (NQ)
<input checked="" type="checkbox"/> 825	447.9018	1340.6837	1340.6837	0.0000	0	28	1 PHSHPALTPEQK

Proteins matching the same set of peptides:

[gi|113608](#) **Mass:** 39774 **Total score:** 130 **Peptides matched:** 5
Fructose-bisphosphate aldolase A (Muscle-type aldolase)

[gi|229506](#) **Mass:** 39449 **Total score:** 130 **Peptides matched:** 5
aldolase C

[gi|6730618](#) **Mass:** 39586 **Total score:** 130 **Peptides matched:** 5
Chain A, Rabbit Muscle Aldolase AFRUCTOSE-1,6-Bisphosphate Complex

[gi|13096347](#) **Mass:** 39636 **Total score:** 130 **Peptides matched:** 5
Chain A, Fructose 1,6-Bisphosphate Aldolase From Rabbit Muscle

[gi|13096351](#) **Mass:** 39646 **Total score:** 130 **Peptides matched:** 5
Chain A, Fructose 1,6-Bisphosphate Aldolase From Rabbit Muscle

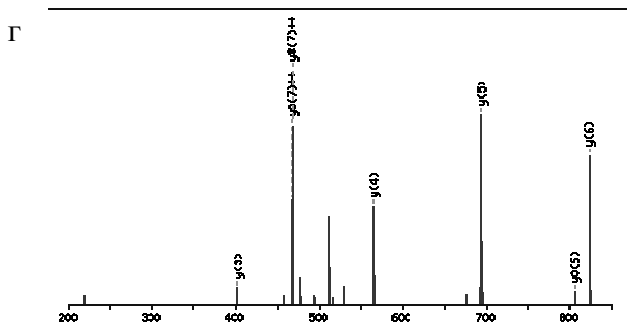


Рис. 5. Отчет Mascot для МС-МС-эксперимента. а — список наиболее вероятных белков, входящих в исходную смесь; б — гистограмма потенциально идентифицированных белков; в — идентификация адолазы, кролик; г — карта фрагментов пептида AAQEEYVKR

В качестве примера приведен результат обработки системой Mascot набора примерно из четырех тысяч MS-MS-спектров, полученных при анализе смеси, состоящей из восьми модельных белков (рис. 5). Эксперимент проводился в университете Южной Дании (г. Оденс) на масс-спектрометре класса IT-FTICR.

В заголовке отчета в списке наиболее вероятных белков на первых восьми позициях стоят белки, составляющие анализируемую смесь (рис. 5, а), они же показаны на гистограмме за пределами заштрихованной зоны низкой достоверности идентификации (рис. 5, б).

В оставшейся части списка представлены варианты интерпретации, полученные за счет случайных ложных совпадений данных масс-спектра с фрагментами пептидов.

Присутствие каждого белка подтверждено несколькими MS-MS-спектрами триптических пептидов этого белка (рис. 5, в). Для спектров, сопоставленных пептидам белков, построены карты фрагментов (рис. 5, г).

Следует отметить, что результаты идентификации, полученные в ходе MS-MS-эксперимента, обладают большей достоверностью, нежели результаты применения метода PMF.

3.4. Идентификация посттрансляционных модификаций

Часть пептидов, содержащихся в подготовленной для MS-анализа пробе, так или иначе модифицированы. Соответственно модификации можно разделить на три группы.

– Посттрансляционные модификации, являющиеся следствием присоединения каких-либо

функциональных групп. Например, фосфорилирование, ацетилирование и т. п.

– Посттрансляционные модификации, связанные с процессингом белка до зрелой формы.

– Модификации, возникшие в процессе пробоподготовки.

Наибольший интерес представляют модификации первых двух типов, поскольку именно они связаны с жизненным циклом клетки. Методами масс-спектрометрии легко детектируют модификации первого типа [62, 63].

Исследования модификации второго типа требуют специальных методов пробоподготовки и обработки данных, отличных от стандартных.

Модификации третьего типа, включающие в себя восстановление дисульфидных связей, окисление метионина и т. п., учитываются при обработке данных MS-анализа, для того чтобы не потерять актуальную информацию. Частично этих модификаций можно избежать при тщательной пробоподготовке.

Каждая из модификаций проявляется в масс-спектрах за счет характерного изменения массы аминокислотного остатка. Во фрагментных спектрах это выглядит как смещение серий пиков на определенную величину, соответствующую молекулярной массе присоединенной функциональной группы. Именно по этой разнице и, в некоторых случаях, по нейтральным потерям определяется точная локализация модификации. Например, фосфорилированный треонин в пятой позиции пептида на рис. 6 четко определяется за счет изменения массы +80 Да ($+HPO_3^+$) и потери слабо связанной группы H_3PO_4 с массой 98 Да при фрагментации.

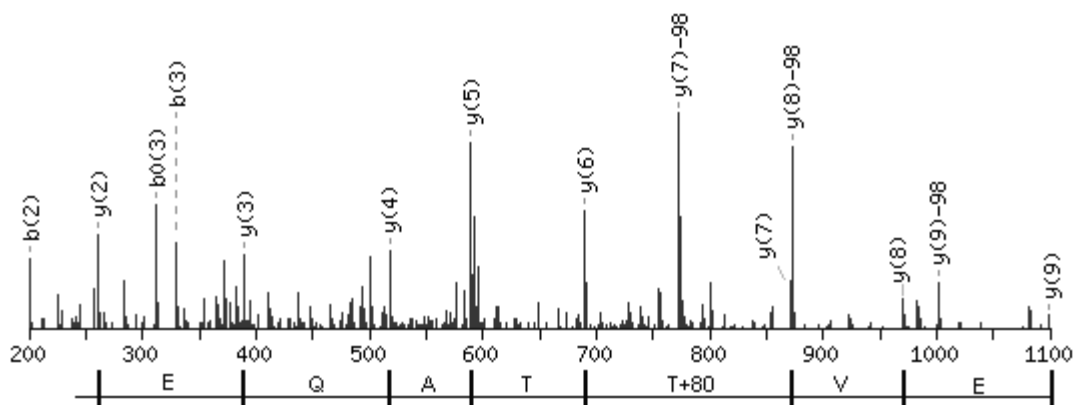


Рис. 6. Фрагментный спектр пептида EAEVTTAQENK, фосфорилированного по треонину

Стоит отметить, что при идентификации за счет сопоставления масс-спектров количество рассчитанных теоретических масс-спектров растет в геометрической прогрессии при учете каждой принимаемой во внимание модификации. Для каждого пептида приходится строить множество вариантов теоретического масс-спектра для всех возможных комбинаций модификаций анализируемого пептида. Это приводит к тому, что время работы алгоритмов при учете модификаций значительно возрастает, а достоверность результатов понижается.

Особый класс модификаций, выявление которых масс-спектрометрическими методами осложнено, — это модификации, образованные высокомолекулярными соединениями, например остатками сахаров и липидов. В этом случае присоединенные к пептиду группы могут иметь молекулярную массу, сопоставимую с самим пептидом, сложную структуру и соответственно обладать собственным набором фрагментов [64]. При детектировании и характеристике такой модификации требуется не только зарегистрировать ее наличие и определить место в последовательности, несущее модификацию, но и понять структурную формулу присоединенной группы. На сегодняшний момент расшифровка спектров пептидов, несущих подобные модификации, чаще всего происходит вручную, а получение спектров требует специальной пробоподготовки.

4. СРАВНИТЕЛЬНЫЙ КОЛИЧЕСТВЕННЫЙ АНАЛИЗ БЕЛКОВ

Проведение сравнительного количественного анализа методом масс-спектрометрии становится возможным при введении внутренних стандартов в исследуемую смесь. При исследовании белков и пептидов в качестве внутреннего стандарта чаще всего используются изотопно-меченные аналоги анализируемых соединений. В качестве метки могут быть использованы такие стабильные изотопы, как C^{13} , D , O^{18} . Интенсивности сигналов в масс-спектре двух веществ, отличающихся только изотопной меткой, будут находиться в таком же соотношении, как концентрации этих веществ.

На данный момент наработано множество методик использования изотопных меток для сравнительного анализа концентраций каждого из белков, присутствующих в двух и более пробах. Эти подходы вызывают острый интерес, поскольку дают возможность непосредственно оценивать отклик биосистемы в целом. Например, выделить из нескольких сотен белков те заранее неизвестные белки, уровень экспрессии которых изменился под воздействием каких-либо факторов [65].

Одним из самых популярных методов является ICAT (Isotope Coded Affinity Tag) [66]. Метка, используемая в этом методе, содержит тиолреагирующую группу для ковалентного присоединения по цистеину, изотопно-меченную связку (линкер) и остаток биотина для аффинного связывания и последующего выделения пептидов, модифицированных по остатку цистеина. Меткой обрабатываются обе переназначенные для сравнения пробы, но в одном случае используется метка, содержащая только $1H$ (легкая метка), в другом — восемь атомов протия замещены на дейтерий (тяжелая метка), что дает разницу масс, равную 8 Да. Перед масс-спектрометрическим анализом (ЖХ-МС-МС) цистеин-содержащие пептиды, связанные с меткой ICAT, в одинаковых условиях выделяются на аффинной колонке и смешиваются.

Благодаря одинаковой химической природе время удерживания в хроматографической колонке для пептидов с тяжелой и легкой метками одинаково. Таким образом, два вещества анализируются при одинаковых условиях. При проведении масс-спектрометрического анализа в спектрах родительских ионов наблюдаются характерные пары сигналов с разницей масс в 8 Да, и по отношению интенсивностей сигналов в таких парах мы можем судить об их концентрационном соотношении. Для идентификации белков, содержащих анализируемые пептиды, а также для того, чтобы убедиться, что пары сигналов были определены правильно, следующим шагом проводят МС-МС-анализ обоих родительских ионов из каждой пары.

После того как пептиды были отнесены к соответствующим белкам, по соотношению интенсивностей сигналов мы можем судить о соотношении концентраций интересующих нас белков в сравниваемых пробах. Информация об относительной концентрации белков получает подтверждение за счет регистрации сигналов нескольких пептидов каждого белка.

Методика SILAC (Stable Isotope Labeling by Amino acids in cell Culture) [67] предлагает вводить изотопную метку за счет естественного метаболизма живых клеток. Для этого одна из двух сравниваемых клеточных линий выращивается на питательной среде, обогащенной, например, аргинином с 6 атомами изотопа ^{13}C . По мере роста клеток весь аргинин в них замещается на тяжелую форму. После выращивания двух клеточных линий — на легком и на тяжелом аргинине — пробы смешиваются и проводится масс-спектрометрический анализ, такой же как описан выше. Преимущество этой методики заключается в том, что пробы с тяжелой и легкой метками смешиваются перед процессом пробоподготовки, что позволяет избежать дополнительных ошибок анализа.

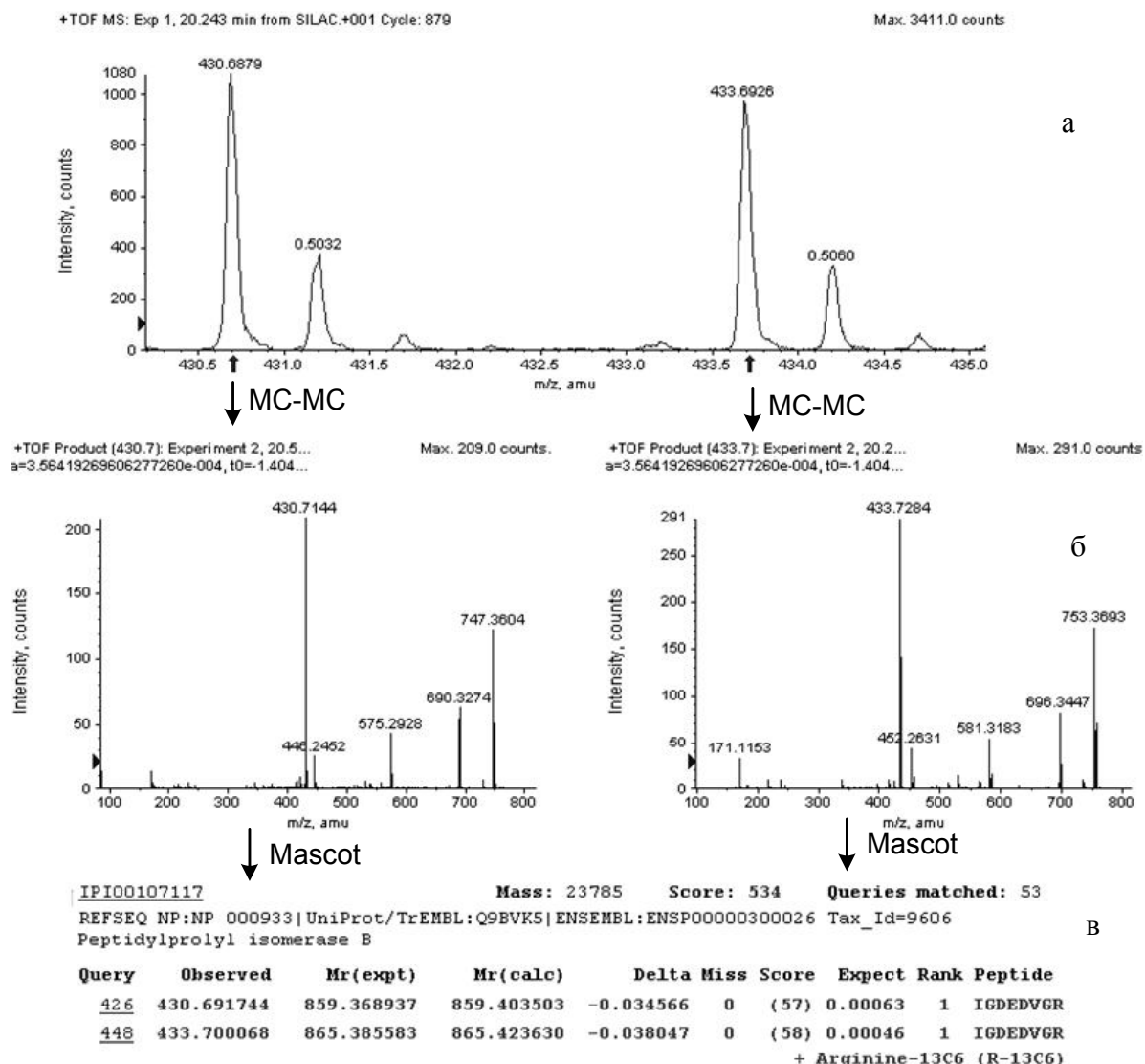


Рис. 7. Пример обработки данных сравнительного количественного анализа методом SILAC. а — спектр родительских ионов; б — спектр дочерних ионов для показанных родительских ионов; в — результат идентификации Mascot для приведенных спектров дочерних ионов

Пример применения метода SILAC приведен на рис. 7, где представлены результаты масс-спектрометрического анализа. На рис. 7, а, показана пара сигналов родительских ионов с разницей масс 3 Да (с учетом зарядности — 6 Да), что соответствует разнице масс между меченым и немеченым аргинином, примерно одинаковой интенсивности, что говорит о близком уровне концентраций пептидов, а соответственно и белков в сопоставляемых пробах.

Данные MC-MC-анализа (рис. 7, б) для каждого пептида, обработанные при помощи ПО Mascot (рис. 7, в), демонстрируют идентичность этих

пептидов.

В последнее время предпринимаются попытки увеличить количество проб, сопоставляемых в результате одного эксперимента. Например, методика iTRAQ [68] предполагает анализ до четырех проб одновременно за счет введения четырех различных изотопных меток.

ЗАКЛЮЧЕНИЕ

Масс-спектрометрия с мягкими методами ионизации становится неотъемлемой частью многих исследований. Этот метод отличается высокой

чувствительностью, что позволяет его использование во многих областях, касающихся молекулярной биологии, биохимии и протеомики человека. Для ряда задач масс-спектрометрия может являться основным или даже единственным средством их решения.

СПИСОК ЛИТЕРАТУРЫ

1. Blackstock W.P., Weir M.P. Proteomics: quantitative and physical mapping of cellular proteins // Trends Biotechnol. 1999. N 17. P. 121–127.
2. Jung E., Heller M., Sanchez J.C., Hochstrasser D.F. Proteomics meets cell biology: the establishment of subcellular proteomes // Electrophoresis. 2000. N 21. P. 3369–3377.
3. Pandey A., Mann M. Proteomics to study genes and genomes // Nature. 2000. V. 405. P. 837–846.
4. Eisenstein E., Gilliland G.L., Herzberg O., et al. Biological function made crystal clear — annotation of hypothetical proteins via structural genomics // Curr. Opin. Biotechnol. 2000. N 11. P. 25–30.
5. Shevchenko A., Schaft D., Roguev A., et al. Deciphering protein complexes and protein interaction networks by tandem affinity purification and mass spectrometry: analytical perspective // Mol. Cell. Proteomics. 2002. V. 1, N 3. P. 204–212.
6. Gygi S.P., Rist B., Gerber S.A., et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags // Nat. Biotechnol. 1999. N 17. P. 994–999.
7. Ong S.E., Blagoev B., Kratchmarova I., et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics // Mol Cell Proteomics. 2002. V. 1, May, N 5. P. 376–386.
8. Heinke M.Y., Wheeler C.H., Yan J.X., et al. Changes in myocardial protein expression in pacing-induced canine heart failure // Electrophoresis. 1999. N 20. P. 2086–2093.
9. Page M.J., Amess B., Rohlf C., et al. Proteomics: a major new technology for the drug discovery process // Drug Discov. Today. 1999. N 4. P. 55–62.
10. Aebersold R., Mann M. Mass spectrometry-based proteomics // Nature. 2003. Mar. 13, v. 422, N 6928. P. 198–207.
11. Washburn M.P., Wolters D., Yates III J.R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology // Nat. Biotechnol. 2001. Mar., v. 19, N 3. P. 242–247.
12. Schmelzer C.E., Getie M., Neubert R.H. Mass spectrometric characterization of human skin elastin peptides produced by proteolytic digestion with pepsin and thermitase // J. Chromatogr. A. 2005. Aug. 12, v. 1083, N 1-2. P. 120–126.
13. Wu S.L., Kim J., Hancock W.S., Karger B. 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) // J. Proteome Res. 2005. Jul-Aug, v. 4, N 4. P. 1155–1170.
14. Mann M., Jensen O.N. Proteomic analysis of post-translational modifications // Nat. Biotechnol. 2003. V. 21, N 3. P. 255–261.
15. Lill J. Proteomic tools for quantitation by mass spectrometry // Mass Spectrom. Rev. 2003. V. 22, N 3. P. 182–194.
16. Ong S.E., Blagoev B., Kratchmarova I., et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics // Mol. Cell. Proteomics. 2002. V. 1, N 5. P. 376–386.
17. Corthals G.L., Wasinger V.C., Hochstrasser D.F., Sanchez J.C. The dynamic range of protein expression: a challenge for proteomic research // Electrophoresis. 2000. V. 21, N 6. P. 1104–1115.
18. Nilsson C.L., Davidsson P. New separation tools for comprehensive studies of protein expression by mass spectrometry // Mass Spectrom. Rev. 2000. V. 19, N 6. P. 390–397.
19. Thevis M., Loo R.R., Loo J.A. Mass spectrometric characterization of transferrins and their fragments derived by reduction of disulfide bonds // J. Am. Soc. Mass Spectrom. 2003. V. 14, N 6. P. 635–647.
20. Jensen P.K., Pasa-Tolic L., Anderson G.A., et al. Probing proteomes using capillary isoelectric focusing-electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry // Anal. Chem. 1999. V. 71, N 11. P. 2076–2084.
21. Shevchenko A., Wilm M., Vorm O., Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels // Anal. Chem. 1996. V. 68, N 5. P. 850–858.
22. Александров М.Л., Галь Л.Н., Краснов Н.В. и др. Экстракция ионов из растворов при атмосферном давлении — новый способ масс-спектрометрического анализа биоорганических веществ // ДАН. 1984. Т. 277, № 2. С. 379–383.
23. Александров М.Л., Галь Л.Н., Краснов Н.В. и др. Прямая стыковка микроколоночного жидкостного хроматографа с масс-спектрометром // Биоорганическая химия. 1984. Т. 10, № 5. С. 710–712.
24. Александров М.Л., Галь Л.Н., Краснов Н.В. и др. Метод масс-спектрометрического анализа труднолетучих термически нестабильных веществ, основанный на экстракции ионов из раствора при атмосферном давлении // ЖАХ. 1985. Т. 40, № 6. С. 160–172.
25. Whitehouse C.M., Dreyer R.N., Yamashita M., Fenn J.B. Electrospray Interface for liquid chromatographs and mass spectrometers // Anal. Chem. 1985. V. 57. P. 675–679.
26. Fenn J.B., Mann M., Meng C.K., Wong S.F., Whitehouse C.M. Electrospray ionization for mass spectrometry of large biomolecules // Science. 1989. V. 246. P. 64–71.
27. Wilm M., Mann M. Analytical properties of the nano-electrospray ion source // Anal. Chem. 1996. V. 68, N 1. P. 1–8.
28. Tanaka K., Waki H., Ido Y., et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry // Rapid Commun. Mass Spectrom. 1988. V. 2, N 8. P. 151–153.
29. Karas M., Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10 000

- daltons // *Anal. Chem.* 1988. V. 60. P. 2299–2301.
30. Лютвинский Я.И., Петров Д.М., Веренчиков А.Н., Хасин Ю.И., Гаврик М.А. Система регистрации для параллельного ВПМС-анализа // *Научное приборостроение.* 2004. Т. 14, № 2. С. 80–91.
 31. Zhang Z., Marshall A.G. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra // *J. Am. Soc. Mass Spec.* 1998. V. 9, N 3. P. 225–233.
 32. Reinhold B.R., Reinhold V.N. Electrospray ionisation mass spectrometry: deconvolution by an entropy-based algorithm // *J. Am. Soc. Mass Spec.* 1992. V. 3, N 3. P. 207–215.
 33. Макаров В.В., Самокиш А.В., Лютвинский Я.И. Метод извлечения значимой информации из масс-спектров пептидов // *Научное приборостроение.* 2004. Т. 14, № 2. С. 96–104.
 34. Pappin D.J., Hojrup P., Blesby A.J. Rapid identification of proteins by peptide-mass fingerprinting // *Curr. Biol.* 1993. V. 3, N 6. P. 327–332.
 35. Perkins D.N., Pappin D.J., Creasy D.M., Cottrell J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data // *Electrophoresis.* 1999. V. 20, N 18. P. 3551–3567.
 36. Eng J.K., McCormack A.L., Yates III J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database // *JASMS.* 1994. V. 5, N 11. P. 976–989.
 37. Craig R., Beavis R.C. TANDEM: matching proteins with tandem mass spectra // *Bioinformatics.* 2004. V. 20, N 9. P. 1466–1467.
 38. Тюряева И.И., Миргородская О.А., Черепанова О.А. и др. Выявление и идентификация ламинина в составе плазматических мембран клеток асцитной гепатомы Зайдела крысы // *Цитология.* 2005. Т. 47, № 2. С. 150–163.
 39. Тюряева И.И., Миргородская О.А., Черепанова О.А. и др. Взаимодействие ламинина с компонентами плазматических мембран клеток асцитной гепатомы Зайдела // *Цитология.* 2005. Т. 47, № 12. С. 1039–1048.
 40. Edman P. Method for determination of the amino acid sequence in peptides // *Acta. Chem. Scan.* 1950. N 4. P. 283–293.
 41. Bairoch A., Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000 // *Nucleic Acids Res.* 2000. V. 28, N 1. P. 45–48.
 42. Bairoch A., Apweiler R., et al. The Universal Protein Resource (UniProt) // *Nucleic Acids Res.* 2005. V. 33, suppl. 1, Jan. (Database issue). P. D154–D159.
 43. Li W., Jaroszewski L., Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases // *Bioinformatics.* 2001. V. 17, N 3. P. 282–283.
 44. Holm L., Sander C. Removing near-neighbour redundancy from large protein sequence collections // *Bioinformatics.* 1998. V. 14, N 5. P. 423–429.
 45. Neubauer G., King A., Rappsilber J., et al. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex // *Nat. Genet.* 1998. V. 20, N 1. P. 46–50.
 46. Parkinson J., Blaxter M. Expressed sequence tags: analysis and annotation // *Methods Mol. Biol.* 2004. V. 270. P. 93–126.
 47. Discala C., Benigni X., Barillot E., Vaysseix G. DBcat: a catalog of 500 biological databases // *Nucleic Acids Res.* 2000. V. 28, N 1. P. 8–9.
 48. Веренчиков А.Н., Краснов Н.В., Галь Л.Н. Tandemные масс-спектрометры в биохимии // *Научное приборостроение.* 2004. Т. 14, № 2. С. 4–23.
 49. Nilsson C.L., Davidsson P. New separation tools for comprehensive studies of protein expression by mass spectrometry // *Mass Spectrom. Rev.* 2000. V. 19, N 6. P. 390–397.
 50. Cody Jr. R.B., Amster I.J., McLafferty F.W. Peptide mixture sequencing by tandem Fourier-transform mass spectrometry // *Proc. Natl. Acad. Sci. USA.* 1985. V. 82, N 19. P. 6367–6370.
 51. Roepstorff P., Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides // *Biomed Mass Spectrom.* 1984. V. 11, N 11. P. 601.
 52. Ma B., Zhang K., Hendrie C., Liang C., Li M., Doherty-Kirby A., Lajoie G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry // *Rapid Commun. Mass Spectrom.* 2003. V. 17, N 20. P. 2337–2342.
 53. Taylor J.A., Johnson R.S. Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry // *Anal. Chem.* 2001. V. 73, N 11. P. 2594–2604.
 54. Mann M., Wilm M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags // *Anal. Chem.* 1994. V. 66, N 24. P. 4390–4399.
 55. Sunyaev S., Liska A.J., Golod A., Shevchenko A. MultiTag: multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry // *Anal. Chem.* 2003. V. 75, N 6. P. 1307–1315.
 56. Tsur D., Tanner S., Zandi E., Bafna V., Pevzner P.A. Identification of post-translational modifications by blind search of mass spectra // *Nat. Biotechnol.* 2005. V. 23, N 12. P. 1562–1567.
 57. Frank A., Tanner S., Bafna V., Pevzner P. Peptide sequence tags for fast database search in mass spectrometry // *J. Proteome Res.* 2005. V. 4, N 4. P. 1287–1295.
 58. Savitski M.M., Nielsen M.L., Zubarev R.A. New database-independent, sequence tag-based scoring of peptide MS/MS data validates Mowse scores, recovers below threshold data, singles out modified peptides, and assesses the quality of MS/MS techniques // *Mol. Cell. Proteomics.* 2005. V. 4, N 8. P. 1180–1188.
 59. Лютвинский Я.И., Краснов Н.В. Разработка и первичная характеристика алгоритма частичного распознавания фрагментных масс-спектров пептидов — CrystalTag // *Научное приборостроение.* 2005. Т. 14, № 2. С. 108–113.
 60. Shevchenko A., Sunyaev S., Loboda A., et al. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching // *Anal. Chem.* 2001. V. 73, N 9. P. 1917–1926.
 61. Wei J., Sun J., Yu W., Jones A., et al. Global proteome discovery using an online three-dimensional LC-MS/MS // *J. Proteome Res.* 2005. V. 4, N 3. P. 801–808.
 62. Bennett K.L., Stensballe A., Podtelejnikov A.V., Mo-

- niatte M., Jensen O.N.* Phosphopeptide detection and sequencing by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry // *J. Mass Spectrom.* 2002. V. 37, N 2. P. 179–190.
63. *MacCoss M.J., McDonald W.H., Saraf A., et al.* Shotgun identification of protein modifications from protein complexes and lens tissue // *Proc. Natl. Acad. Sci. USA.* 2002. V. 99, N 12. P. 7900–7905.
64. *Mirgorodskaya E., Krogh T.N., Roepstorff P.* Characterization of protein glycosylation by MALDI-TOFMS // *Methods Mol. Biol.* 2000. V. 146. P. 273–292.
65. *Stewart J.J., White J.T., Yan X., et al.* Proteins associated with cisplatin resistance in ovarian cancer cells identified by quantitative proteomics technology and integrated with mRNA expression levels // *Mol. Cell. Proteomics.* 2006. V. 5, N 3. P. 433–443.
66. *Gygi S.P., Rist B., Gerber S.A., Turecek F., Gelb M.H., Aebersold R.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags // *Nat. Biotechnol.* 1999. V. 17, N 10. P. 994–999.
67. *Ong S.E., Blagoev B., Kratchmarova I., et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics // *Mol. Cell. Proteomics.* 2002. V. 1, N 5. P. 376–386.
68. *Shadforth I.P., Dunkley T.P., Lilley K.S., Bessant C.* i-Tracker: for quantitative proteomics using iTRAQ™ // *BMC Genomics.* 2005. V. 6. P. 145.

*Институт аналитического приборостроения РАН,
Санкт-Петербург*

Контакты: Подольская Екатерина Петровна,
Kpodolskaya@yandex.ru

Материал поступил в редакцию 17.09.2010.

MASS SPECTROMETRY WITH SOFT IONIZATION METHODS IN PROTEOMIC ANALYSIS (REVIEW)

N. V. Krasnov, Ya. I. Lutvinsky, E. P. Podolskaya

Institute for Analytical Instrumentation RAS, Saint-Petersburg

The review covers the use of mass spectrometry with soft ionization methods in proteomics. The principles of mass spectrometric analysis of proteins, the possibilities of mass spectrometry in the identification of proteins and peptides, in the search and identification of post-translational modifications, and in comparative quantitative analysis are described.

Keywords: mass spectrometry with soft ionization methods, peptide identification, tandem mass spectrometry, posttranslational modifications, peptide databases