

УДК 543.061

НОВОЕ В МЕТОДОЛОГИИ ХИМИЧЕСКОЙ ИДЕНТИФИКАЦИИ

© 2024 г. Б. Л. Мильман^{а,*}, И. К. Журкович^а

^а Научно-клинический центр токсикологии имени академика С.Н. Голикова Федерального медико-биологического агентства России
ул. Бехтерева, 1, Санкт-Петербург, 192019, Россия

* E-mail: bormilman@yandex.ru

Поступила в редакцию 10.04.2023 г.

После доработки 28.04.2023 г.

Принята к публикации 03.05.2023 г.

Представлен обзор основных методов, способов, процедур и информационных продуктов, применяемых в последние годы при идентификации химических соединений. Методология, используемая в целевом анализе, во многом осталась без изменения; лишь критерии идентификации подверглись некоторой корректировке. Резко расширился фронт исследований в нецелевом анализе. В этом случае основные проблемы заключаются в выявлении кандидатов на идентификацию. Эти версии проверяются по типичным критериям целевого анализа. Эффективный поиск подходящих кандидатских соединений стал возможен при появлении современных хромато-масс-спектрометров высокого разрешения и прогрессе информатики. Последний включает разработку алгоритмов и программ обработки хроматографических и масс-спектрометрических данных, их сравнения со справочными характеристиками, прогнозирования масс-спектров и параметров удерживания и других величин. Химические базы данных позволяют оценить распространенность химических соединений и соответственно их перспективность как кандидатов на идентификацию.

Ключевые слова: идентификация, критерии идентификации, масс-спектрометрия, хроматография, нецелевой анализ, прогнозирование, базы данных.

DOI: 10.31857/S0044450224020029, EDN: vziyvk

Процедура идентификации определяемых соединений представляет собой обязательную часть химического анализа. Основные принципы, правила и критерии химической идентификации сформулированы в первом десятилетии XXI века ([1–3] и ссылки в этих публикациях). При этом методология целевого анализа (ЦА, определение заданных аналитов по соответствующим методам анализа) к настоящему времени изменилась мало, произошли лишь, например, некоторые корректировки критериев идентификации [4].

Гораздо более сложная задача – идентификация заранее неизвестных химических соединений в рамках нецелевого анализа (НЦА) [5]. Возможности решения этой задачи, которая связана с возросшими потребностями в глубоком анализе вещества, значительно выросли в последние 10–15 лет. Это обусловлено, во-первых, широким распространением новых аналитических

приборов (хроматографов и масс-спектрометров высокого и ультравысокого разрешения) и, во-вторых, очень быстрым развитием информатики, которое проявляется в появлении новых алгоритмов, программ, баз данных (БД), сетей. Эти обстоятельства во многих случаях способствуют сравнительно быстрому и надежному поиску кандидатов на идентификацию и сбору доказательств определенной идентичности аналитов.

Методология нецелевого анализа быстро развивается, что, в свою очередь, предполагает необходимость регулярного его рассмотрения и подготовки новых обзоров в этой области. В настоящей статье мы кратко суммировали результаты наиболее важных работ последних лет, связанных с методами и способами химической идентификации, ограничившись низкомолекулярными соединениями и в основном масс-спектрометрией (МС), хроматографией

и хромато-масс-спектрометрией (ХМС). Другую релевантную информацию, относящуюся, прежде всего, к анализу биологических проб, объектов окружающей среды, продуктов питания и природных соединений, можно найти в обзорах процедур современной идентификации [6–12].

ОБЩИЕ ВОПРОСЫ

Основные способы идентификации, основанные на применении методов (а) МС, в том числе тандемной масс-спектрометрии (МС²) и масс-спектрометрии высокого разрешения (МСВР, МС²ВР), (б) газовой (ГХ) и жидкостной хроматографии (ЖХ), в том числе высокоэффективной (ВЭЖХ) и ультраэффективной (УЭЖХ), а также (в) спектрометрии ионной подвижности (СИП) перечислены в табл. 1. Однозначно правильная идентификация достигается при использовании только первого способа идентификации (совместное определение аналита и стандарта) [1–3]. В качестве аналитических стандартов часто используют обычные реактивы и химикаты (хотя их чистота не всегда регламентирована). Большие потенциальные возможности рассматриваемого способа идентификации поддерживаются огромным рынком продаваемых химических соединений (см. далее). Использование собственно аналитических стандартов (стандартных образцов состава) было бы наиболее надежным вариантом идентификации, но номенклатура стандартов [14] меньше на несколько порядков величины.

В метаболомике [15] и анализе объектов окружающей среды [16] положительные результаты применения способа идентификации № 1 (табл. 1) относят к ее первому уровню. Остальные способы приводят лишь к кандидатам на идентификацию (уровни второй и ниже). При этом более надежные результаты, которые называют “предположительно аннотированными соединениями” (putatively annotated compounds) [15] или “вероятной структурой” (probable structure) [16], дает второй способ идентификации (табл. 1). Близкие условия регистрации экспериментальных и справочных данных предполагают тем не менее возможность правильной идентификации. Поэтому такой подход часто используется, он оправдан практически, но строя его доказательность требует валидации в случае многих аналитов и матриц.

В научной литературе последнего времени, особенно в области метаболомики, вместо термина “идентификация” и наряду с ним очень

часто употребляется термин “аннотация” [9, 15], означающий процедуру приписания аналиту (аналитическому сигналу) предполагаемой формулы/структуры или других характеристик химического соединения и имеющий смысл условной идентификации.

Можно ввести количественные показатели правильности (надежности) идентификации. В случае надежной идентификации (способ № 1) того или иного аналита логично считать, что соответствующая правильность равняется 100%. Остальные способы идентификации дают, как правило, меньшие значения, которые можно оценить, используя те или иные вероятностные модели обсуждаемой процедуры [1–3]. Если оценивать правильность того или иного метода/способа идентификации для серии анализов или группы аналитов, то этот показатель представляет собой долю полученных правильных положительных (ПП) результатов (традиционный показатель качественного анализа [2]).

Ранее предложено понятие точки идентификации (ТИ) — одной характеристики (одного значения измеряемой величины, одного свойства) аналита, используемой для идентификации при сравнении со свойствами известного соединения сравнения [2, 3]. Для надежной целевой идентификации требуется несколько ТИ (совпадение параметров хроматографического удерживания, масс ионов и интенсивностей массовых пиков), некоторые из них (точные массы ионов) обладают большим весом. Концепция ТИ недавно была распространена на НЦА, осуществляемый разными методами ЖХ-МСВР [6, 17]. Точки идентификации суммируются как целочисленные баллы [6]. Введена также дробная шкала оценок совпадения масс-спектров и хроматограмм, в частности, предусмотрено сравнение экспериментальных данных с соответствующими расчетными значениями масс ионов и времен удерживания (ВУ) [17].

ЦЕЛЕВОЙ АНАЛИЗ. НОРМАТИВНЫЕ КРИТЕРИИ ИДЕНТИФИКАЦИИ

Идентификация заранее заданных аналитов проводится по методикам анализа, многие из которых разработаны в соответствии с руководствами международных или национальных регулирующих организаций. Недавно опубликованы актуальные сводки нормативных хроматографических и масс-спектрометрических критериев идентификации [4]. После этого появились

Таблица 1. Способы идентификации

№	Способ	Принцип	Новые возможности
1	Совместный анализ проб и аналитических стандартов (веществ/ соединений сравнения)	Совпадение аналитических сигналов аналита и вещества сравнения* при использовании двух и более независимых методов анализа, прежде всего МС и хроматографии, и варьировании условий анализа	Полное использование возможностей синтеза химических соединений по требованию [13] расширяет рамки применения этого наиболее надежного метода идентификации
2	Сравнение результатов анализа со справочными данными экспериментального происхождения	Данные представляют собой масс-спектры и параметры удерживания, зарегистрированные в сходных, но неидентичных экспериментальных условиях (иное время, иные условия регистрации спектров и хроматограмм, другая лаборатория, другие модели приборов)	Идентификация обеспечивается расширением библиотек масс-спектров, сводок хроматографических параметров удерживания и значений поперечного сечения столкновений (ПСС) в СИП, появлением молекулярных сетей
3	Сравнение с естественными значениями физических постоянных	МСВР: относительные массы молекул и изотопная картина	Широкое распространение МСВР обеспечивает реализацию этого способа идентификации за счет уточнения молекулярных формул кандидатов на идентификацию
4	Сравнение результатов анализа с предсказанными справочными данными	Масс-спектры, параметры удерживания и другие характеристики рассчитывают, прежде всего, с использованием методов машинного обучения, и используют в качестве справочных данных	Методы предсказания совершенствуются, точность предсказания возрастает
5	Интерпретация масс-спектров	Выводы о структуре (субструктурах) аналитов делают, исходя из правил и закономерностей фрагментации и/или при использовании соответствующих компьютерных программ	Способы компьютерной интерпретации масс-спектров, широко развиваются - параллельно с разработкой способов их предсказания по предполагаемой структуре аналита (№ 4)

* Совместный, что более доказательно, или последовательный ввод в прибор аналита и стандарта.

новые руководства Управления по санитарному надзору за качеством пищевых продуктов и медикаментов США (FDA, 2020 г. [18]), Евросоюза (ЕС, пестициды, 2021 г. [19]) и Всемирного антидопингового агентства (2023 г. [20]) Последние два документа не содержат существенных изменений критериев в сравнении с предыдущей их редакцией [4].

В обновленных рекомендациях FDA [18], как и ранее в документе ЕС [4], введены универсальные

максимальные отклонения относительных интенсивностей пиков ($\pm 30\%$) от соответствующих значений для масс-спектров сравнения (например, спектров градуировочных стандартов) — независимо от вида масс-спектра (одинарный масс-анализ, MS^1 , минимум три иона; MS^2 , два и более фрагментных ионов) и типа скана (регистрация отдельных ионов/переходов или полные сканы). В случае МСВР критерии включают массовое окно (± 5 ppm) для двух ионов. Совпадения

интенсивностей, как и ранее, не требуется, поскольку при идентификации соответствие точных масс имеет гораздо большее значение, чем сходство интенсивностей пиков [21]. Интересно, что из обсуждаемого руководства исчез в явном виде такой распространенный хроматографический критерий, как максимальное отклонение от ВУ известного соединения; предусмотрена, однако, необходимость совпадения пиков аналита на всех масс-хроматограммах одного хроматографического пробега [18].

Указанные нормативные/стандартизованные критерии, проверенные в специальных исследованиях, в основном оказались действенными, но содержали возможности улучшения [22]. Так, определено 169 ветеринарных лекарств, добавленных в яичные продукты, методом УЭЖХ-МС² (с тройным квадрупольным масс-анализатором, QqQ) и прослежено влияние числа регистрируемых пиков фрагментных ионов на показатели ложных (положительных, ЛП, и отрицательных, ЛО) результатов идентификации. При учете двух рекомендованных фрагментов невысокий уровень ЛП (наблюдаемый в случае <10% анализов) обнаружен для 95% анализов, а регистрация трех фрагментов, доступная для современных приборов, обеспечила указанный низкий уровень ЛП уже в случае 99% анализов. Сравнительно высокий уровень ЛО (более 10% анализов) также существенно уменьшался (с 25–45 до 10–12% анализов) при регистрации трех вместо двух переходов [22].

НЕЦЕЛЕВОЙ АНАЛИЗ

Смещение химико-аналитических исследований в сторону НЦА — ярко выраженная тенденция наших дней [5]. Идентификация неизвестных компонентов вещества в рамках его НЦА, даже если эти компоненты обнаружены, представляет собой гораздо более сложную задачу, чем подтверждение или неподтверждение идентичности заранее заданных аналитов в ЦА (см. ранее). В отличие от ЦА, где предусмотрены, как правило, первый или, реже, второй способы идентификации (табл. 1), в рамках НЦА применимы все рассмотренные способы, но их реализация весьма сложна и часто приводит к неоднозначным результатам.

Совместный анализ с аналитическими стандартами (способ № 1, табл. 1) возможен, но малопродуктивен, если о природе аналита известно мало и сложно выбрать вещества сравнения. Справочные данные (№ 2) могут отсутствовать

или относиться к непохожим экспериментальным условиям. Решению задач НЦА способствует широкое применение МСВР (способ № 3), но при увеличении молекулярной массы аналита выбор его единственной молекулярной формулы, и тем более структуры, становится непростым. Еще одно направление прогресса НЦА связано с быстрым развитием алгоритмов предсказания масс-спектров и других характеристик (способ № 4) и предсказания структур по таким данным (способ № 5). Результаты работы соответствующих программ восполняют отсутствие справочных данных экспериментального происхождения и устраняют трудности “ручной” интерпретации спектров.

Способы идентификации при НЦА показаны на рис. 1 и более подробно рассмотрены ниже, начиная с понятия химического пространства, которое представляет собой множество (подмножества) химических соединений как потенциальных результатов НЦА. Отсутствующие в статье детали и трактовки можно найти в публикациях [6, 9, 11, 24–27].

Химическое пространство. В общем случае неизвестным компонентом анализируемых проб с переменной вероятностью может оказаться практически любое известное (неизвестное известное, unknown known) или даже неизвестное (unknown unknown) соединение. Допускается существование огромного количества (10^{11} и много больше, табл. 2) индивидуальных соединений, пути синтеза которых в принципе известны. Гораздо более определённой является информация о соединениях, которые могут быть синтезированы по требованию. Если их объединить с реактивами и химикатами, поставляемыми со склада, то общее количество доступных соединений как будто бы приближается к 1 млрд (табл. 2), что, однако, трудно проверить.

Реальное химическое пространство, по-видимому, ограничено ~200 млн индивидуальных соединений и их несложных смесей (веществ, substances) (табл. 2). Среди них имеется очень много редких соединений, вряд ли присутствующих в распространенных объектах анализа: биопробах человека и животных, продуктах питания, растительных материалах, объектах окружающей среды. Различные подмножества/подпространства наиболее распространенных соединений (их до ~1–2 млн) включены в специализированные БД, которые также указаны в табл. 2. При прочих равных условиях следует ориентироваться на присутствие в анализируемых объектах

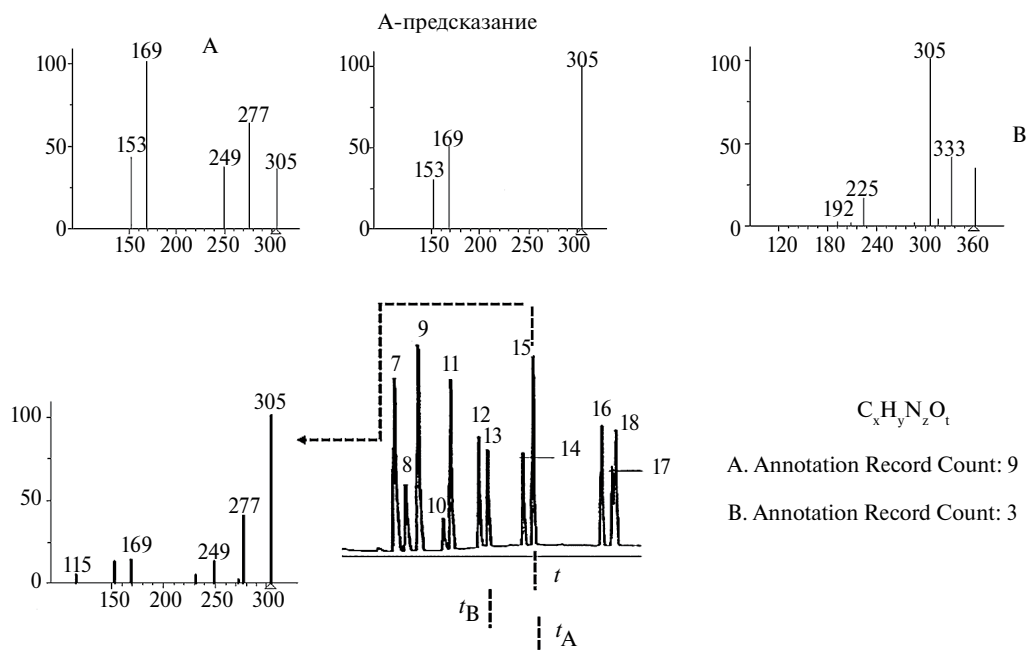


Рис. 1. Идентификация условного соединения А в рамках нецелевого анализа. Наблюдается удовлетворительное совпадение тандемного масс-спектра (левый нижний угол), соответствующего пику 15 на хроматограмме, со справочными спектрами соединения А (верхний ряд), в том числе с *in silico* спектром. Отличие от второго кандидатского соединения (В, правый верхний угол) гораздо более существенное. Время удерживания t также гораздо ближе к справочному значению для соединения А (t_A). С использованием масс-спектрометрии высокого разрешения установлена брутто-формула неизвестного аналита $C_xH_yN_zO_t$, причем для аналита А обнаружено гораздо больше информации, чем для соединения В; приведен показатель Annotation Record Count – размер информационной справки – в базе данных PubChem [23] (см. далее).

распространенных/популярных и достаточно хорошо изученных соединений, таких как основные компоненты биологических образцов (метаболиты), химическая продукция и часто используемые реактивы, продукты их превращений в окружающей среде и т.п. Таким образом, среди аналитов преобладают или могут преобладать те соединения, которые ранее уже были обнаружены в рассматриваемых объектах анализа. Математический эквивалент этого утверждения – статистика Байеса, которая используется в простых тестах [2, 3], но применима также к современным методам идентификации [38].

Показатели популярности/распространенности представляют собой, по существу, относительный объем априорной информации (метаинформации) о данном соединении и частоту ссылок на него и поэтому весьма эффективно оцениваются по существующим химическим БД (табл. 3). Итак, если две версии аналита, имеющие похожие масс-спектры и другие характеристики, значительно различаются по информационным показателям, то более популярное

соединение следует считать основным кандидатом на идентификацию и соответствующую гипотезу необходимо проверять в первую очередь. То или иное ненулевое значение такого показателя логично рассматривать как пороговое при рассмотрении возможности присутствия соединения в анализируемой пробе; такие же выводы можно делать на основании простого присутствия или отсутствия в специализированной БД.

Показатели популярности химических соединений рассматриваются в дополнение к результатам химического анализа. Факт распространенности того или иного подходящего аналита сам по себе нельзя рассматривать как строгое доказательство его присутствия в пробе, но он очень важен для поддержки или отклонения идентификационной гипотезы и первоочередного отбора аналитических стандартов при получении исчерпывающих доказательств определенной идентичности.

Табл. 3 показывает, что для рассматриваемых оценок популярности чаще других используют большие БД ChemSpider и PubChem, к которым

Таблица 2. Химическое пространство и подпространства*

Соединения	Комментарии, источники информации	Количество соединений
Виртуальное пространство		
Потенциальные, молекулярная масса ≤ 500 Да	Возможен синтез, оценка см. [28]	$10^{11} - 10^{200}$
Виртуальное пространство и известные соединения		
Продаваемые и синтезируемые по требованию	Каталоги и реклама, собранные в БД ZINC 20 [13]:	
	— соединения, готовые к продаже; — синтез по требованию	230 000 000 750 000 000
Известные соединения		
Отраженные в научной литературе	Зарегистрированы в системе Chemical Abstracts Service (CAS) [29]	$\leq 204\,000\,000$
Важные, входящие в другие БД**	БД ChemSpider [30]	115 000 000
Биологически важные	БД PubChem [23]	114 000 000
Наиболее распространенные соединения		
Регулируемые, загрязнители окружающей среды, токсиканты	БД CompTox Chemistry Dashboard [31].	1 200 059
	БД CHEMLIST [29].	>417 000
	БД NORMAN-SLE [32]	115 248
Метаболиты	БД METLIN (данные 2017 г. [28]). БД HMDB [33]	961 829 220 945
Природные соединения	Разные БД [34]	>320 000
В продуктах питания	БД FooDB [35]	70 926

* Информацию о других химических/биохимических БД см. [27, 36, 37].

** Наша характеристика.

обеспечен свободный доступ, что особенно значимо практически. Тем не менее их сложно отнести к курируемым БД, они создаются путем компьютерной компиляции данных из других электронных массивов информации и могут быть неполными, содержать неточности [50]. Вызывает сожаление, что наиболее полная курируемая БД Chemical Abstracts, содержащая выверенную химическую информацию и не имевшая конкурентов в эпоху бумажной информатики, в настоящее время доступна только по коммерческим онлайн-каналам.

Присутствие того или иного соединения в специализированной БД (нижняя часть табл. 2) означает возможность его обнаружения в образцах/матрицах соответствующего вида. В случае ненулевой информации о возможном составе матрицы НЦА приобретает статус

“определения ожидаемых/подозреваемых соединений” (ООС) [5]. Поэтому составление сводок подозреваемых аналитов [32, 56], имеющих те или иные шансы на обнаружение и идентификацию, существенно облегчает работу аналитиков.

Библиотеки масс-спектров. Использование компьютерных библиотек масс-спектров — один из первых подходов к автоматической идентификации и до сих пор наиболее распространенный ее способ [2, 3, 57, 58]. Идентификация особенно эффективна в случае летучих аналитов и масс-спектров ионизации электронами (ИЭ-МС¹), достигая правильности ~80%. Две самые большие библиотеки: NIST 20 (350 643 спектра 306 869 соединений) и Wiley Registry 12th (соответственно 817 290 и 668 435) [5]. Эти массивы информации также включают справочные индексы удерживания (ИУ, ГХ-ИУ). Многие вопросы

создания и использования таких библиотек решены, но отдельные проблемы остаются [59–61].

История библиотек tandemных масс-спектров (спектров MS^2) гораздо короче, и они далеко не полны [57, 58]. Считается, что с их использованием возможна идентификация не более нескольких процентов метаболитов и других низкомолекулярных соединений [62, 63]. Далее, вид спектров MS^2 варьируется в широких пределах; наблюдается зависимость от типа приборов (QqQ; квадрупольно-времяпролетный масс-анализатор, Q-TOF; орбитальная ловушка, Orbitrap), энергии столкновений и некоторых других факторов [3, 62, 64]. Это существенно влияет на показатель ПП при поисках в рассматриваемых библиотеках. Таким образом, наряду с расширением этих библиотек необходимы стандартизация условий их создания и повышение качества спектральных данных [58, 62]. Это тем более важно, поскольку спектральные библиотеки служат основой для отбора тренировочных массивов спектров в процедурах их предсказания (спектры *in silico*, см. далее).

Наиболее крупные библиотеки спектров MS^2 перечислены в табл. 4; сведения о других их коллекциях можно найти в работах [57, 58]. Библиотеки удобно разделить на коммерческие и свободного доступа, причем последние могут быть открыты для пополнения пользователями и аннотирования их спектров (депозитарии, такие как GNPS, см. табл. 4). При специализации аналитических лабораторий удобно пользоваться библиотеками (домашними или межлабораторными), ориентированными на определенные классы и группы соединений [58, 68–70]. Некоторые библиотеки tandemных масс-спектров содержат справочную информацию о ВУ и сечениях столкновений (см. далее) [58, 71].

В литературе отмечено несколько путей видоизменения библиотек, которые приводят к улучшению идентификации. Так, предложено расширение библиотек за счет спектров, соответствующих новым энергиям столкновений; такие данные получают интерполяцией существующих спектров, приписанных другим значениям энергии [72]. Перспективно также использование фрагментов всех ионов — предшественников данного анализата [71].

Модернизация библиотечных поисков. Сравнение экспериментальных масс-спектров со справочными спектрами приводит к ранжированию последних в зависимости от значений

показателя их сходства. Наиболее похожие справочные спектры (компьютерные ответы 1-го ранга) соответствуют главным кандидатам на идентификацию. Именно кандидаты 1-го ранга чаще всего учитываются в статистике правильных результатов (как % ПП).

В качестве меры сходства традиционно используют функцию косинуса (точечную функцию) [2, 3], не исключающую, однако, появление ложных кандидатов на идентификацию. Один из способов повышения эффективности библиотечных поисков и расширения их возможности — использование усовершенствованных мер сходства экспериментальных и справочных масс-спектров. Это особенно важно для tandemных масс-спектров с их недостаточной воспроизводимостью, присутствием фоновых пиков, множественной природой ионов-предшественников, пропуском некоторых фрагментных ионов, а также их ограниченным количеством. Многие меры сходства, удачно дополняющие/заменяющие функцию косинуса, рассмотрены в публикациях [12, 73].

В новых работах обращают внимание не только на сходство масс ионов, но и на соответствие разностей масс (нейтральных потерь), что позволяет, например, осуществлять поиск структурно сходных соединений [12]. Соответствующий показатель сходства — модифицированная функция косинуса, где дополнительно учитываются разности масс предшественников и фрагментов в сопоставляемых спектрах [74]. Еще один новый перспективный подход состоит в использовании спектральной энтропии в мерах сходства tandemных масс-спектров [75]. Предложено также учитывать вариативность масс-спектров высокого разрешения при их сравнении [76].

Перечисляя различные инновации, отметим также, что использование глубоких нейронных сетей позволяет улучшить результаты традиционных поисков в библиотеках спектров ИЭ- MS^1 [77] и полезно для сравнения tandemных масс-спектров [78].

Молекулярные сети [79, 80]. Ими называют графы подмножества химических соединений, связанных между собой, чаще всего по степени сходства спектров MS^2 (рис. 2). В узлах графа — массы ионов (ионов-предшественников) различных соединений, причем каждому из них может соответствовать один или несколько похожих спектров. Ребра графа соединяют в итоге пары структурно сходных соединений, масс-спектры

Таблица 3. Показатели популярности/распространенности химических соединений и их возможного присутствия в анализируемых пробах

Год	Показатель, комментарии	Литература
2000	Количество рефератов (справок, ссылок) в журнале Chemical Abstracts*	[39, 40]
2011	Количество рефератов (справок, ссылок) в БД Chemical Abstracts	[43]
2012	Количество ссылок на соединение в БД ChemSpider	[44]
2014	БД PubChem, пороговое значение показателя refscore (>5) — числа веществ (substances), в состав которых входит данное химическое соединение	[45]
2016	Присутствие в объектах анализа по данным предшествующих отчетов	[46]
2016	Количество ссылок, в том числе патентов, и источников информации в БД ChemSpider и PubChem	[47]
2017	Присутствие соединения в БД по метаболомике и загрязнителям окружающей среды	[48]
2020	Количество источников данных в БД CompTox Chemistry Dashboard (см.табл. 2)	[49]
2020 — 2021	Количество источников данных и фирм — производителей химикатов в БД ChemSpider и PubChem, размер справок о соединении и число патентов в БД PubChem	[50–52]
2021	Различные показатели БД PubChemLite — версии PubChem, предназначенной для решения задач экспосомики [54] и объединения с предсказателем масс-спектров MetFrag	[53]
2023	Количество определенных лексических единиц в описании соединения в БД PubChem	[55]

* Значительное количество совместных справок (цитирований) в журнале/БД Chemical Abstracts, относящихся к (а) паре соединений или (б) соединению и матрице (веществу), может указывать соответственно на вероятное совместное присутствие двух и более соединений в данной пробе или присутствие соединения в данной матрице (веществе) [40–42].

которых весьма близки. Молекулярная сеть существует совместно с библиотекой масс-спектров, по которой можно идентифицировать “неизвестные известные” аналиты. Если узел, встраиваемый в сеть, относится к такому новому для информационной системы соединению, то оно может быть идентифицировано также по разности масс ионов этого аналита и связанного с ним известного соединения с учетом структуры последнего.

Наиболее известна молекулярная сеть в области природных соединений (GNPS, табл. 4), которая представляет собой часть глобальной информационной системы, включающей библиотеки масс-спектров, депозитарии для них, различные программные инструменты для обработки данных и взаимодействия с пользователями системы и др. Методология сетей сочетается с возможностями групповой идентификации и использованием предсказанных масс-спектров [81, 82], применима к ИЭ-МС¹ [83]. Способы

построения сетей и варианты их наполнения совершенствуются [12, 84–87].

Предсказание масс-спектров [10, 12, 63, 88]. Компьютерные методы предсказания масс-спектров (спектров *in silico*) позволяют получать данные, заменяющие справочные масс-спектры экспериментального происхождения, что особенно важно для МС². Различные алгоритмы предсказания, как правило, основаны на машинном обучении с использованием тренировочного набора известных масс-спектров и/или на правилах фрагментации (эвристическая информация). Наиболее известны программы CFM-ID, MetFrag и MS-FINDER. Легче прогнозировать тандемные масс-спектры соединений регулярного строения, например липидов, чьи ионы подвергаются распаду по одним и тем же правилам. Новые способы предсказания и соответствующие компьютерные программы появляются постоянно (см., например, [36, 89–91]), некоторые из них позволяют

Таблица 4. Основные библиотеки tandemных масс-спектров низкомолекулярных соединений [5]

Название	Количество		Комментарии	Литература
	спектры	соединения		
NIST 20, MC ² *	1 320 389	31 000		
METLIN		~600 000	Метаболиты	[65]
MassBank of North America (MONA)	691548	226303	Преобладают биологически активные соединения, присутствуют <i>in silico</i> масс-спектры	
mzCloud	10 326 515	20 820	MC ⁿ , лекарственные соединения, метаболиты	[66]
The Global Natural Product Social Molecular Networking (GNPS)	586 647 [58]	> 20 000 (оценка)	Природные соединения.	
MassBank	90 471	16 881	MC ¹ , MC ⁿ	[67]
Human Metabolome Database (HMDB):			Метаболиты	[33]
— эксперимент;	64 923	4 064		
— предсказание (<i>in silico</i>)	1 787 163	206 809		

* Включая масс-спектры высокого разрешения 27 840 соединений.

оценивать степень правильности условной идентификации (аннотации) [92]

Масс-спектры *in silico* в роли справочных спектров сравнения — сами по себе далеко не панацея, и результат соответствующих библиотечных поисков зависит от многих факторов [93]. Доля правильных ответов 1-го ранга, полученных в большинстве значимых примеров, может достигать ~50–70% и иногда выше этого уровня [48, 63, 94, 95]. Более высокие показатели наблюдаются при сочетании с другими способами идентификации и включении априорной информации, например показателей распространенности соединений (см. ранее).

Исследуется возможность предсказания масс-спектров, в том числе tandemных, с использованием квантово-химических расчетов [96, 97].

Структурная интерпретация масс-спектров [10, 12, 63, 88]. Соответствующие программы на основе машинного обучения решают задачу, обратную предсказанию масс-спектров: выводят варианты структурных фрагментов и кандидатских структур в целом, исходя из масс-спектра неизвестного аналита. Так, широко известен программный инструмент CSI:FingerID, который, в частности, был интегрирован в программу

SIRIUS. Последняя, исходя из масс ионов-предшественников, изотопной картины и спектров MC², выбирает наиболее вероятные кандидатские соединения из тех, которые включены в химические БД [98]. Позднее для этих же целей была создана программа MAD HATTER, обеспечивающая более 70% правильных идентификаций в модельной выборке tandemных спектров [55]. Разрабатываются и другие новые программы автоматического извлечения структурных данных из масс-спектров [36, 99, 100]. В целом отмечается высокая значимость алгоритмов глубокого машинного обучения, применяемого для решения рассматриваемой задачи [11, 101]. Возможности “ручной” интерпретации tandemных масс-спектров, однако, далеко не исчерпаны [102, 103].

Параметры удерживания в хроматографии. Совместное применение масс-спектрометрии и хроматографии — требование нормативных документов ЦА и условие надежной идентификации в условиях НЦА.

В ГХ давно и успешно используют справочные ИУ экспериментального происхождения, которые сейчас в значительном объеме представлены в БД NIST (139693 соединений [5]) и других сводках данных [2, 3, 104]. Большинство

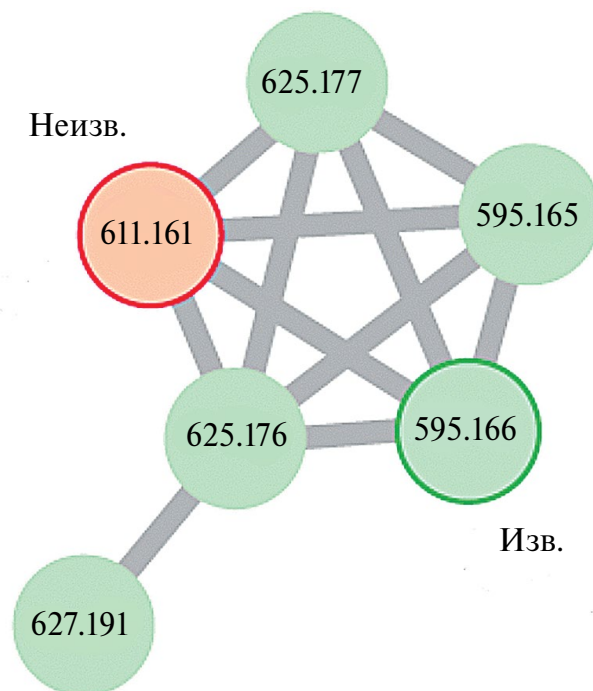


Рис. 2. Участок молекулярной сети (адаптирован из работы [80]). Числами указаны массы (Да) ионов — предшественников каждого из соединений. Разность масс 611.161 и 595.166 соответствует атому кислорода, что, с учетом сходства масс-спектров, позволяет считать неизвестный анализ (Неизв.) продуктом окисления известного соединения (Изн.).

практически значимых летучих соединений, по-видимому, охвачены этими данными. Тем не менее появляются новые соединения, подходящие для определения методом ГХ, например летучие производные малолетучих соединений, поэтому требуется информация об их справочных ИУ. Соответствующие оценки выполняются различными вариантами машинного обучения и некоторыми другими методами [104–106].

Концепция ИУ для В/УЭЖХ менее значима, здесь значительно чаще оперируют величиной ВУ. Начинают появляться достаточно большие коллекции экспериментальных ВУ, например для 80 тыс. соединений в информационной системе METLIN [107]; эти характеристики установлены в условиях обращенно-фазовой ВЭЖХ.

Массивы экспериментальных данных служат тренировочными выборками для предсказания ВУ методами машинного обучения (обзоры [6, 10, 108]). В качестве структурных переменных, коррелируемых с ВУ (количественная связь структура—удерживание), используют молекулярные дескрипторы. Обращенно-фазовые и колонки HILIC рассматривают отдельно [108, 109]. Различные ошибки современного предсказания ВУ в основном находятся в диапазоне от 10 с до 1 мин [10, 108–112]. Подобная точность

предсказаний, по-видимому, недостаточна, чтобы отразить тонкие различия в структуре аналита, но вполне подходит для отклонения многих кандидатов на идентификацию. Так, 68% кандидатских соединений, предсказанных по масс-спектрам компонентов плазмы крови, не соответствовали оценкам характеристик удерживания: оказались вне интервала ($V_U \pm 1$ мин) [109]. Отметим, что вследствие сравнительно невысокой воспроизводимости ВУ и их сильной зависимости от экспериментальных условий, важен пересчет справочных наборов ВУ при их адаптации к текущему эксперименту [10, 113].

Спектрометрия ионной подвижности [114]. Этот метод представляет собой разделение ионов по их подвижности (ИП) в газовой фазе и сочетается с МС и ВЭЖХ-МС. Приборы СИП-МС и ВЭЖХ-СИП-МС с различными техническими (физическими) вариантами реализации ИП выпускаются основными производителями масс-спектрометрических приборов.

Стадия ИП повышает степень разделения аналитов, что особенно наглядно проявляется для смесей изомеров или изобарных соединений. Разделение зависит от ПСС, которое представляет собой индивидуальную характеристику химических соединений и подходит для их

идентификации. Значения этой величины, как экспериментального происхождения, так и рассчитанные, табулированы в нескольких БД [12, 112, 114, 115] и могут быть использованы как справочные данные совместно с масс-спектрометрической и хроматографической информацией [6, 12].

Совмещение методов идентификации необходимо для повышения ее надежности. Совместное применение MS^1BP , MS^2 , У/ВЭЖХ и СИП названо “четырёхмерным” подходом к идентификации [12]. При высоком качестве (точности) всех экспериментальных и справочных данных критерию распознавания аналита — незначительной разности соответствующих масс ионов-предшественников, значений ВУ и ПСС и больших показателей сходства спектров MS^2 — должно удовлетворять только одно кандидатское соединение. В реальности различные справочные данные могут приводить к противоречивым и неоднозначным результатам. Для выбора наилучшей версии и дальнейшей ее проверки можно оценивать степень совпадения данных по подсчету точек идентификации в их численном выражении [6, 17].

Пакеты программ (fusion tools). Аналитику удобно, если различные программы, необходимые для идентификации нецелевых аналитов, интегрированы в единое программное обеспечение (на компьютере прибора или доступном сайте). Это позволяет совмещать поиски в библиотеках масс-спектров, электронных коллекциях ИУ, ВУ и ПСС и химических БД с оценками указанных характеристик в экономном режиме. В обзоре [10] упомянуто восемь таких пакетов, компоненты которых частично рассмотрены ранее. В случае анализа сложных образцов критически важное значение имеют программы обработки первичных данных ХМС и выделения соответствующих аналитических сигналов [116]. В целом в литературе описаны сотни программных продуктов, предназначенных для различного преобразования данных ХМС и идентификации аналитов по масс-спектрам и хроматографическим параметрам; создают специальные программы для ориентации в этой области [117].

Сличения. Традиционная практика аналитической химии, в том числе химической идентификации [2, 3], включает сравнительные/сличительные эксперименты, оценивающие профессиональный уровень и качество работы лабораторий, эффективность различных методов и подходов к анализу, пригодность аналитических стандартов и др. Важные меж- и внутрилабораторные

сличения последних лет, касающиеся идентификации компонентов предоставленных образцов как проб неизвестного состава или только анализа данных, отражены в табл. 5. Основным вывод из результатов этих работ заключается в далеко не полной воспроизводимости итогов работы различных лабораторий (№ 1, 8) и умозаключений, сделанных с использованием разных вариантов ЖХ-МС и разных режимов сбора данных в одной лаборатории (№ 7, а также [123]), итогов применения разных компьютерных программ (№ 2, 4) и разных мер сходства масс-спектров (№ 10). При этом возможности идентификации оказались весьма удовлетворительными (№ 6, 9), они возрастают в случае дополнительной информации об ожидаемом составе проб и распространенности аналитов (№ 1, 3, 5).

ЗАКЛЮЧЕНИЕ: СМЕНА ПАРАДИГМЫ И СМЕНА ПОКОЛЕНИЙ

Суммируя изложенное ранее, можно утверждать, что основная тенденция исследований последних лет представляет собой развитие методологии поисков кандидатов на идентификацию (условно идентифицированных или аннотированных соединений в итоге) в рамках НЦА. Если требуется окончательная идентификация по типичным критериям ЦА, то, располагая перечнем кандидатских соединений, целесообразно (а) ввести в оборот аналитические стандарты или, что менее надежно и требует валидации, (б) регистрировать масс-спектры и хроматограммы (а также спектры ИП) в условиях, максимально близких к тем, в которых определены наиболее качественные/точные справочные данные. Тематический переход от разработки принципов, правил, критериев идентификации к методологии поиска кандидатских соединений представляет, на наш взгляд, смену *парадигмы* в целях и содержании химико-аналитических исследований.

Такого рода изменения проявляются, во-первых, в появлении и широкой эксплуатации новых аналитических приборов (хромато-масс-спектрометров высокого разрешения и др.) и, во-вторых, в резком усилении роли информатики. Последнее представляет собой появление многочисленных программ автоматического поиска кандидатов на идентификацию на основе машинного обучения, расширения библиотек масс-спектров, электронных коллекций других справочных характеристик и химических баз данных. Этот тренд отражает привлечение к работе нового поколения

Таблица 5. Примеры сравнительных исследований в области идентификации

№	Год	Объекты анализа или данные, метод	Содержание проекта	Число лабораторий	Результаты, комментарии	Литература
1	2015	Экстракты речной воды	Сравнение различных методов ХМС и разных подходов к анализу	19	Акцент на НЦА, ООС и МСВР. Выявлена необходимость интегральных пакетов программ, полных библиотек масс-спектров, многих аналитических стандартов и актуальность обмена данными	[118]
2	2012-2017, CASMI*	Масс-спектры набора соединений как неизвестных аналитов	Использование различных программ предсказания масс-спектров, их интерпретации, библиотек масс-спектров и других данных	≤8 (наша оценка)	36–88% ПП – результаты ежегодных победителей [49]	[119]
3	2017	Масс-спектры, представленные в CASMI*	Предсказание масс-спектров и других данных	1	22–25% ПП для наилучших предсказанных соединений, 73–78% при их наличии в химической БД, 87–93% при добавлении поисков в библиотеках масс-спектров	[48]
4	2019	Данные ХМС	Сравнение заданных сигналов (значений m/z при определенных ВУ) в четырех программах обработки данных	≤5 (наша оценка)	Лишь 10% сигналов определены для всех программ. Идентификация 22 тест-соединений: 64–88% ПП	[116]
5	2020	Масс-спектры, представленные в CASMI*	Использование программы предсказания спектров CFM-ID с учетом дополнительной информации	1	Правильный результат в числе пяти наилучших предсказаний: 36–81% соединений (CFM-ID) и 53–96% (CFM-ID + оценка числа источников информации, см. табл. 3)	[49]
6	2020	Множественные спектры MS ² BP 15 соединений	Сравнение спектров каждой лаборатории со справочными данными и другие сравнения	7	78–96% ПП	[62]
7	2020	Образцы плазмы крови человека, УЭЖХ-СИП-MS ² BP	Сравнение идентификации метаболитов в различных вариантах метода анализа и сбора данных	1	Наилучшие результаты: режим фрагментации всех ионов-предшественников (all ion fragmentation) при росте энергии столкновения (ramped) в ходе распада ионов	[120]

Таблица 5. Окончание

					вместо фиксированных ее значений	
8	2021	Экстракты чая, УЭЖХ-МС ² BP	Воспроизводимость сигналов ХМС (МС ¹) в интервалах $\Delta m/z$: ± 5 ppm и ΔBU : ± 3 с или ± 6 с; сравнение Orbitrap и Q-TOF	2	Совпадение <30% сигналов, причины—различия ионов-предшественников, степени фрагментации и др.	[121]
9	2021	51 пестицид в овощах и фруктах, массовая доля 5–50 мкг/кг, УЭЖХ-МС ² BP	Возможности методологии НЦА в целевом определении пестицидов	25	ЛП и ЛО < 5%. 88% лабораторий идентифицировали все из ≤ 7 пестицидов, добавленных в каждый из образцов	[122]
10	2022	10 млн пар спектров МС ²	Сравнение трех показателей сходства: функции косинуса, модифицированной функции косинуса и нейтральных потерь	1	Функция модифицированного косинуса превосходит два других показателя в корреляциях со структурой соединений	[74]

* Проект CASMI (Critical Assessment of Small Molecule Identification — критическая оценка идентификации малых молекул). Исходные файлы использовались затем в других исследованиях возможности компьютерных программ (№ 3, 5).

специалистов в области информатики. Итак, смена поколений приборов и специалистов и смена парадигмы взаимозависимы и являются ответом на широкие потребности в подробном анализе вещества.

Смена эпох не обходится без проблем. В частности, отдельные издания, прежде всего реферативный журнал Chemical Abstracts, содержащий выверенную информацию о большинстве известных химических соединений, при переходе в электронную форму не обрели открытого доступа, что можно рассматривать как потерю в сфере информатики. Далее, многие опытные химики последних поколений непосредственно работали с веществом и могут многое сказать о содержании пробы по цвету, запаху, форме кристаллов, растворимости, прозрачности, простейшим тестам и т.п. Полученные результаты весьма важны при анализе неизвестных объектов сложного состава. Идентификация индивидуальных соединений методами ХМС — еще одна составляющая таких работ. Как показано ранее, важная роль в них принадлежит специалистам

с хорошей подготовкой в области информатики. Наилучшие результаты достигаются, по нашему мнению, когда их деятельность сочетается с работой классических химиков.

Работа выполнена в соответствии с тематическим планом прикладных и научных исследований и разработок в сфере здравоохранения по государственному заданию ФМБА России (код: 64.001.23.800).

СПИСОК ЛИТЕРАТУРЫ

1. *Milman B.L.* Identification of chemical compounds // Trends Anal. Chem. 2005. V. 24. № 6. P. 493. <https://doi.org/10.1016/j.trac.2005.03.013>
2. *Мильман Б.Л.* Введение в химическую идентификацию. СПб: ВВМ, 2008. 180 с. (*Milman B.L.* Introduction to chemical identification. Saint Petersburg: VVM, 2008. 180 p.)
3. *Milman B.L.* Chemical identification and its quality assurance. Berlin: Springer, 2011. 281 p.
4. *Мильман Б.Л., Журкович И.К.* Обобщенные критерии идентификации химических соединений

- методами хроматографии—масс-спектрометрии // Аналитика и контроль. 2020. Т. 24. № 3. С. 164. <https://doi.org/10.15826/analitika.2020.24.3.003>
5. Мильман Б.Л., Журкович И.К. Современная практика нецелевого химического анализа // Журн. аналит. химии. 2022. Т. 77. № 5. С. 412. <https://doi.org/10.31857/S0044450222050085> (Milman B.L., Zhurkovich I.K. Present-day practice of non-target chemical analysis // J. Anal. Chem. 2022. V. 77. № 5. P. 537. <https://doi.org/10.1134/S1061934822050070>)
 6. Monge M.E., Dodds J.N., Baker E.S., Edison A.S., Fernández F.M. Challenges in identifying the dark molecules of life // Annu. Rev. Anal. Chem. 2019. V. 12. P. 177. <https://doi.org/10.1146/annurev-anchem-061318-114959>
 7. Molyneux R.J., Beck J.J., Colegate S.M., Edgar J.A., Gaffield W., Gilbert J., Hofmann T., McConnell L.L., Schieberle P. Guidelines for unequivocal structural identification of compounds with biological activity of significance in food chemistry (IUPAC Technical Report) // Pure Appl. Chem. 2019. V. 91. № 8. P. 1417. <https://doi.org/10.1515/pac-2017-1204>
 8. Nash W.J., Dunn W.B. From mass to metabolite in human untargeted metabolomics: Recent advances in annotation of metabolites applying liquid chromatography-mass spectrometry data // Trends Anal. Chem. 2019. V. 120. Article 115324. <https://doi.org/10.1016/j.trac.2018.11.022>
 9. Place B.J., Ulrich E.M., Challis J.K., Chao A., Du B., Favela K., Feng Y.L., Fisher C.M., Gardinali P., Hood A., Knolhoff A.M., McEachran A.D., Nason S.L., Newton S.R., Ng B., Nuñez J., Peter K.T., Phillips A.L., Quinete N., Renslow R., Sobus J.R., Sussman E.M., Warth B., Wickramasekara S., Williams A.J. An introduction to the benchmarking and publications for non-targeted analysis working group // Anal. Chem. 2021. V. 93. № 49. P. 16289. <https://doi.org/10.1021/acs.analchem.1c02660>
 10. Tian Z., Liu F., Li D., Fernie A.R., Chen W. Strategies for structure elucidation of small molecules based on LC-MS/MS data from complex biological samples // Comput. Struct. Biotechnol. J. 2022. V. 20. P. 5085. <https://doi.org/10.1016/j.csbj.2022.09.004>
 11. De Jonge N.F., Mildau K., Meijer D., Louwen J.J., Bueschl C., Huber F., Van der Hoof J.J. Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools // Metabolomics. 2022. V. 18. № 12. P. 103. <https://doi.org/10.1007/s11306-022-01963-y>
 12. Cai Y., Zhou Z., Zhu Z.J. Advanced analytical and informatic strategies for metabolite annotation in untargeted metabolomics // Trends Anal. Chem. 2023. V. 158. Article 116903. <https://doi.org/10.1016/j.trac.2022.116903>
 13. ZINC20. <https://zinc20.docking.org> (18.03.2023).
 14. Reference Materials. <https://www.sigmaaldrich.com/RU/en/products/analytical-chemistry/reference-materials> (18.03.2023).
 15. Sumner L.W., Amberg A., Barrett D., Beale M.H., Berger R., Daykin C.A., Fan T.W.M., Fiehn O., Goodacre R., Griffin J.L., Hankemeier T., Hardy N., Harnly J., Higashi R., Kopka J., Lane A.N., Lindon J.C., Marriott P., Nicholls A.W., Reilly M.D., Thaden J.J., Viant M.R. Proposed minimum reporting standards for chemical analysis: chemical analysis working group (CAWG) metabolomics standards initiative (MSI) // Metabolomics. 2007. V. 3. P. 211. <https://doi.org/10.1007/s11306-007-0082-2>
 16. Schymanski E.L., Jeon J., Gulde R., Fenner K., Ruf M., Singer H.P., Hollender J. Identifying small molecules via high resolution mass spectrometry: Communicating confidence // Environ. Sci. Technol. 2014. V. 48. № 4. P. 2097. <https://doi.org/10.1021/es5002105>
 17. Alygizakis N., Lestremieu F., Gago-Ferrero P., Gil-Solsona R., Arturi K., Hollender J., Schymanski E.L., Dutilleul V., Slobodnik J., Thomaidis N.S. Towards a harmonized identification scoring system in LC-HRMS/MS based non-target screening (NTS) of emerging contaminants // Trends Anal. Chem. 2023. V. 159. Article 116944. <https://doi.org/10.1016/j.trac.2023.116944>
 18. Methods, Method Verification and Validation. <https://www.fda.gov/media/73920/download> (19.03.2023).
 19. Analytical quality control and method validation procedures for pesticide residues analysis in food and feed. https://www.eurl-pesticides.eu/userfiles/file/EURLALL/SANTE_11312_2021.pdf (19.03.2023).
 20. Minimum criteria for chromatographic-mass spectrometric confirmation of the identity of analytes for doping control purposes. https://www.wada-ama.org/sites/default/files/2023-02/td2023idcrv1.1_eng_final.pdf (19.03.2023).
 21. Milman B.L., Zhurkovich I.K. Identification performance of low-molecular compounds by searching tandem mass spectral libraries with simple peak matching // Mass Spectrom. Lett. 2018. V. 9. № 3. P. 73. <https://doi.org/10.5478/MSL.2018.9.3.73>
 22. Lehotay S.J. Comparison of analyte identification criteria and other aspects in triple quadrupole tandem mass spectrometry: Case study using UHPLC-MS/MS for regulatory analysis of veterinary drug residues in liquid and powdered eggs // Anal. Bioanal. Chem. 2022. V. 414. № 1. P. 287. <https://doi.org/10.1007/s00216-021-03380-x>
 23. PubChem. <https://pubchem.ncbi.nlm.nih.gov> (20.03.2023).
 24. Dührkop K., Nothias L.F., Fleischauer M., Reher R., Ludwig M., Hoffmann M.A., Petras D., Gerwick W.H., Rousu J., Dorrestein P.C., Böcker S. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra // Nat. Biotechnol. 2021. V. 39. № 4. P. 462. <https://doi.org/10.1038/s41587-020-0740-8>

25. *Rey-Stolle F., Dudzik D., Gonzalez-Riano C., Fernández-García M., Alonso-Herranz V., Rojo D., Barbas C., García A.* Low and high resolution gas chromatography-mass spectrometry for untargeted metabolomics: A tutorial // *Anal. Chim. Acta.* 2022. V. 1210. Article 339043. <https://doi.org/10.1016/j.aca.2021.339043>
26. *Caballero-Casero N., Belova L., Vervliet P., Antignac J.P., Castaño A., Debrauwer L., López M.E., Huber C., Klanova J., Krauss M., Lommen A., Mol H.G.J., Oberacher H., Pardo O., Price E.J., Reinstadler V., Vitale C.M., Van Nuijs A.L.N., Covaci A.* Towards harmonized criteria in quality assurance and quality control of suspect and non-target LC-HRMS analytical workflows for screening of emerging contaminants in human biomonitoring // *Trends Anal. Chem.* 2021. V. 136. Article 116201. <https://doi.org/10.1016/j.trac.2021.116201>
27. *Misra B.B.* New software tools, databases, and resources in metabolomics: Updates from 2020 // *Metabolomics.* 2021. V. 17. № 5. P. 49. <https://doi.org/10.1007/s11306-021-01796-1>
28. *Milman B.L., Zhurkovich I.K.* The chemical space for non-target analysis // *Trends Anal. Chem.* 2017. V. 97. P. 179. <https://doi.org/10.1016/j.trac.2017.09.013>
29. CAS. <https://www.cas.org/about/cas-content> (20.03.2023).
30. ChemSpider. <http://www.chemspider.com> (20.03.2023).
31. CompTox Chemistry Dashboard. <https://comptox.epa.gov/dashboard> (20.03.2023).
32. NORMAN-SLE. <http://www.norman-network.com/?q=node/236> (20.03.2023).
33. The human metabolome database (HMDB). <https://hmdb.ca> (20.03.2023).
34. *Sorokina M., Steinbeck C.* Review on natural products databases: where to find data in 2020 // *J. Cheminf.* 2020. V. 12. № 1. P. 20. <https://doi.org/10.1186/s13321-020-00424-9>
35. FooDB. <https://foodb.ca/compounds> (20.03.2023).
36. *O'Shea K., Misra B.B.* Software tools, databases and resources in metabolomics: Updates from 2018 to 2019 // *Metabolomics.* 2020. V. 16. № 3. P. 35. <https://doi.org/10.1007/s11306-020-01657-3>
37. *Banimfreg B.H., Shamayleh A., Alshraideh H.* Survey for computer-aided tools and databases in metabolomics // *Metabolites.* 2022. V. 12. № 10. P. 1002. <https://doi.org/10.3390/metabo12101002>
38. *Ludwig M.* Bayesian methods for small molecule identification. Doctoral dissertation. Friedrich-Schiller-Universität, Jena. 2020. https://www.db-thueringen.de/servlets/MCRFileNodeServlet/dbt_derivate_00050369/dissludwig.pdf (20.03.2023).
39. *Milman B.L., Konopelko L.A.* Identification of chemical substances by testing and screening of hypotheses I. General // *Fresenius J. Anal. Chem.* 2000. V. 367. P. 621. <https://doi.org/10.1007/s002160000426>
40. *Milman B.L., Kovrizhnykh M.A.* Identification of chemical substances by testing and screening of hypotheses II. Determination of impurities in n-hexane and naphthalene: II. Determination of impurities in n-hexane and naphthalene // *Fresenius J. Anal. Chem.* 2000. V. 367. P. 629. <https://doi.org/10.1007/s002160000427>
41. *Milman B.L.* A procedure for decreasing uncertainty in the identification of chemical compounds based on their literature citation and cocitation. Two case studies // *Anal. Chem.* 2002. V. 74. № 7. P. 1484. <https://doi.org/10.1021/ac010611p>
42. *Milman B.L.* Literature-based generation of hypotheses on chemical composition using database co-occurrence of chemical compounds // *J. Chem. Inf. Model.* 2005. V. 45. № 5. P. 1153. <https://doi.org/10.1021/ci049716u>
43. *Little J.L., Cleven C.D., Brown S.D.* Identification of “known unknowns” utilizing accurate mass data and chemical abstracts service databases // *J. Am. Soc. Mass Spectrom.* 2011. V. 22. № 2. P. 348. <https://doi.org/doi:10.1007/s13361-010-0034-3>
44. *Little J.L., Williams A.J., Pshenichnov A., Tkachenko V.* Identification of “known unknowns” utilizing accurate mass data and ChemSpider // *J. Am. Soc. Mass Spectrom.* 2012. V. 23. № 1. P. 179. <https://doi.org/10.1007/s13361-011-0265-y>
45. *Ridder L., Van der Hooft J.J.J., Verhoeven S.* Automatic compound annotation from mass spectrometry data using MAGMa // *Mass Spectrom.* 2014. V. 3. № 2. P. S0033. <https://doi.org/10.5702/massspectrometry.S0033>
46. *Woldegebriel M., Vivó-Truyols G.* A new Bayesian approach for estimating the presence of a suspected compound in routine screening analysis // *Anal. Chem.* 2016. V. 88. № 19. P. 9843. <https://doi.org/10.1021/acs.analchem.6b03026>
47. *Ruttkies C., Schymanski E.L., Wolf S., Hollender J., Neumann S.* MetFrag relaunched: incorporating strategies beyond in silico fragmentation // *J. Cheminf.* 2016. V. 8. № 1. P. 1. <https://doi.org/10.1186/s13321-016-0115-9>
48. *Blaženović I., Kind T., Torbašinović H., Obrenović S., Mehta S.S., Tsugawa H., Wermuth T., Schauer N., Jahn M., Biedendieck R., Jahn D., Fiehn O.* Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: Database boosting is needed to achieve 93% accuracy // *J. Cheminf.* 2017. V. 9. P. 32. <https://doi.org/10.1186/s13321-017-0219-x>
49. *McEachran A.D., Chao A., Al-Ghoul H., Lowe C., Grulke C., Sobus J.R., Williams A.J.* Revisiting five years of CASMI contests with EPA identification

- tools // *Metabolites*. 2020. V. 10. № 6. P. 260.
<https://doi.org/10.3390/metabo10060260>
50. Мильман Б.Л., Островидова Е.В., Журкович И.К. Большие химические базы данных свободного доступа в нецелевом масс-спектрометрическом анализе // *Масс-спектрометрия*. 2020. Т. 17. № 2. С. 87. <https://doi.org/10.25703/MS.2020.17.23> (Milman B.L., Ostrovidova E.V., Zhurkovich I.K. Big free-access chemical databases in non-target mass spectrometry analysis // *J. Anal. Chem.* 2021. V. 76. P. 1477.) <https://doi.org/10.1134/S1061934821130086>
51. Мильман Б.Л., Журкович И.К. Популярность химических соединений. О чем это? // *Аналитика*. 2020. Т. 10. № 6. С. 464.
<https://doi.org/10.22184/2227-572X.2020.10.6.464.469>
52. Milman B.L., Zhurkovich I.K. Statistics of the popularity of chemical compounds in relation to the non-target analysis // *Molecules*. 2021. V. 26. № 8. P. 2394.
<https://doi.org/10.3390/molecules26082394>
53. Schymanski E.L., Kondić T., Neumann S., Thiessen P.A., Zhang J., Bolton E.E. Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag // *J. Cheminf.* 2021. V. 13. № 1. P. 19.
<https://doi.org/10.1186/s13321-021-00489-0>
54. Мильман Б.Л., Журкович И.К. Экспосом и химии // *Аналитика*. 2023. Т. 13. № 1. С. 56.
<https://doi.org/10.22184/2227-572X.2023.13.1.56.59>
55. Hoffmann M.A., Kretschmer F., Ludwig M., Böcker S. MAD HATTER correctly annotates 98% of small molecule tandem mass spectra searching in PubChem // *Metabolites*. 2023. V. 13. № 3. P. 314.
<https://doi.org/10.3390/metabo13030314>
56. Cave J.R., Parker E., Lebrilla C., Waterhouse A.L. Omics forecasting: Predictive calculations permit the rapid interpretation of high-resolution mass spectral data from complex mixtures // *J. Agric. Food Chem.* 2019. V. 67. № 48. С. 13318.
<https://doi.org/10.1021/acs.jafc.9b04384>
57. Milman B.L., Zhurkovich I.K. Mass spectral libraries: A statistical review of the visible use // *Trends Anal. Chem.* 2016. V. 80. P. 636.
<https://doi.org/10.1016/j.trac.2016.04.024>
58. Bittremieux W., Wang M., Dorrestein P.C. The critical role that spectral libraries play in capturing the metabolomics community knowledge // *Metabolomics*. 2022. V. 18. № 12. P. 94.
<https://doi.org/10.1007/s11306-022-01947-y>
59. Samokhin A., Sotnezova K., Revelsky I. Predicting the absence of an unknown compound in a mass spectral database // *Eur. J. Mass Spectrom.* 2019. V. 25. № 6. P. 439.
<https://doi.org/10.1177/1469066719855503>
60. Chua C.K., Lv Y., Zhao W., Ren Y., Zhang H.J. Improving annotation of known-unknowns with accurately reconstructed mass spectra // *Int. J. Mass Spectrom.* 2020. V. 451. Article 116321.
<https://doi.org/10.1016/j.ijms.2020.116321>
61. Samokhin A.S., Matyushin D.D. How searching against multiple libraries can lead to biased results in GC/MS-based metabolomics // *Rapid Commun. Mass Spectrom.* 2023. V. 37. № 3. P. e9437.
<https://doi.org/10.1002/rcm.9437>
62. Oberacher H., Sasse M., Antignac J.P., Guitton Y., Debrauwer L., Jamin E.L., Schulze T., Krauss M., Covacci A., Caballero-Casero N., Rousseau K. A European proposal for quality control and quality assurance of tandem mass spectral libraries // *Environ. Sci. Eur.* 2020. V. 32. P. 1.
<https://doi.org/10.1186/s12302-020-0031>
63. Kretzler C.A., Thallinger G.G. A map of mass spectrometry-based in silico fragmentation prediction and compound identification in metabolomics // *Briefings Bioinf.* 2021. V. 22. № 6. P. bbab073.
<https://doi.org/10.1093/bib/bbab073>
64. Milman B.L. General principles of identification by mass spectrometry // *Trends Anal. Chem.* 2015. V. 69. P. 24.
<https://doi.org/10.1016/j.trac.2014.12.009>
65. Montenegro-Burke J.R., Guigas C., Siuzdak G. METLIN: A tandem mass spectral library of standards / *Computational Methods and Data Analysis for Metabolomics* / Ed. Li S. USA: Clifton, 2020. P. 149.
https://doi.org/10.1007/978-1-0716-0239-3_9
66. m/zCloud. <https://www.mzcloud.org> (22.03.2023).
67. MassBank. <https://massbank.eu/MassBank/Contents> (22.03.2023).
68. Lee S., Hwang S., Seo M., Shin K.B., Kim K.H., Park G.W., Kim J.Y., Yoo J.S., No K.T. BMDMS-NP: A comprehensive ESI-MS/MS spectral library of natural compounds // *Phytochemistry*. 2020. V. 177. Article 112427.
<https://doi.org/10.1016/j.phytochem.2020.112427>
69. Davidsen A., Mardal M., Linnet K., Dalsgaard P.W. How to perform spectrum-based LC-HR-MS screening for more than 1,000 NPS with HighResNPS consensus fragment ions // *PloS One*. 2020. V. 15. № 11. Article e0242224.
<https://doi.org/10.1371/journal.pone.0242224>
70. Li Y., Zhu W., Xiang Q., Kim J., Dufresne C., Liu Y., Li T., Chen S. Creation of a plant metabolite spectral library for untargeted and targeted metabolomics // *Int. J. Mol. Sci.* 2023. V. 24. № 3. P. 2249.
<https://doi.org/10.3390/ijms24032249>
71. Tada I., Tsugawa H., Meister I., Zhang P., Shu R., Katsumi R., Wheelock C.E., Arita M., Chaleckis R. Creating a reliable mass spectral-retention time library for all ion fragmentation-based metabolomics // *Metabolites*. 2019. V. 9. № 11. P. 251.
<https://doi.org/10.3390/metabo9110251>
72. King E., Overstreet R., Nguyen J., Ciesielski D. Augmentation of MS/MS Libraries with spectral interpolation for improved identification // *J. Chem. Inf. Model.* 2022. V. 62. № 16. P. 3724.
<https://doi.org/10.1021/acs.jcim.2c00620>

73. Kim S., Kato I., Zhang X. Comparative analysis of binary similarity measures for compound identification in mass spectrometry-based metabolomics // *Metabolites*. 2022. V. 12. № 8. P. 694. <https://doi.org/10.3390/metabo12080694>
74. Bittremieux W., Schmid R., Huber F., Van der Hooft J. J., Wang M., Dorrestein P.C. Comparison of cosine, modified cosine, and neutral loss based spectrum alignment for discovery of structurally related molecules // *J. Am. Soc. Mass Spectrom.* 2022. V. 33. № 9. P. 1733. <https://doi.org/10.1021/jasms.2c00153>
75. Li Y., Kind T., Folz J., Vaniya A., Mehta S.S., Fiehn O. Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification // *Nat. Methods*. 2021. V. 18. № 12. P. 1524. <https://doi.org/10.1038/s41592-021-01331-z>
76. Roberts M.J., Moorthy A.S., Sisco E., Kearsley A.J. Incorporating measurement variability when comparing sets of high-resolution mass spectra // *Anal. Chim. Acta*. 2022. V. 1230. Article 340247. <https://doi.org/10.1016/j.aca.2022.340247>
77. Matyushin D.D., Sholokhova A.Y., Buryak A.K. Deep learning driven GC-MS library search and its application for metabolomics // *Anal. Chem.* 2020. V. 92. № 17. P. 11818. <https://doi.org/10.1021/acs.analchem.0c02082>
78. Huber F., Van der Burg S., Van der Hooft J.J., Ridder L. MS2DeepScore: A novel deep learning similarity measure to compare tandem mass spectra // *J. Cheminform.* 2021. V. 13. № 1. P. 84. <https://doi.org/10.1186/s13321-021-00558-4>
79. GNPS. <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp> (23.03.2023).
80. Аксенов А.А. Гармония химического пространства // *Лаборатория и производство*. 2019. № 6. P. 8. <https://doi.org/10.32757/2619-0923.2019.6.10.8.15>
81. Quinlan Z.A., Koester I., Aron A.T., Petras D., Aluwihare L.I., Dorrestein P.C., Nelson C.E., Kelly L.W. ConCISE: Consensus annotation propagation of ion features in untargeted tandem mass spectrometry combining molecular networking and in silico metabolite structure prediction // *Metabolites*. 2022. V. 12. № 12. P. 1275. <https://doi.org/10.3390/metabo12121275>
82. Neto F.C., Raftery D. Expanding urinary metabolite annotation through integrated mass spectral similarity networking // *Anal. Chem.* 2021. V. 93. № 35. P. 12001. <https://doi.org/10.1021/acs.analchem.1c02041>
83. Elie N., Santerre C., Touboul D. Generation of a molecular network from electron ionization mass spectrometry data by combining MZmine2 and MetGem software // *Anal. Chem.* 2019. V. 91. № 18. P. 11489. <https://doi.org/10.1021/acs.analchem.9b02802>
84. Olivon F., Elie N., Grelier G., Roussi F., Litaudon M., Touboul D. MetGem software for the generation of molecular networks based on the t-SNE algorithm // *Anal. Chem.* 2018. V. 90. № 23. P. 13900. <https://doi.org/10.1021/acs.analchem.8b03099>
85. Chen L., Lu W., Wang L., Xing X., Chen Z., Teng X., Zeng X., Muscarella A.D., Shen Y., Cowan A., McReynolds M.R., Kennedy B.J., Lato A.M., Campagna S.R., Singh M., Rabinowitz J.D. Metabolite discovery through global annotation of untargeted metabolomics data // *Nat. Met.* 2021. V. 18. № 11. P. 1377. <https://doi.org/10.1038/s41592-021-01303-3>
86. Zhou Z., Luo M., Zhang H., Yin Y., Cai Y., Zhu Z.J. Metabolite annotation from knowns to unknowns through knowledge-guided multi-layer metabolic networking // *Nat. Commun.* 2022. V. 13. № 1. P. 6656. <https://doi.org/10.1038/s41467-022-34537-6>
87. Treen D.G., Wan, M., Xing S., Louie K.B., Huan T., Dorrestein P.C., Northen T.R., Bowen B.P. SIMILE enables alignment of tandem mass spectra with statistical significance // *Nat. Commun.* 2022. V. 13. № 1. P. 2510. <https://doi.org/10.1038/s41467-022-30118-9>
88. Ljoncheva M., Stepišnik T., Džeroski S., Kosjek T. Cheminformatics in MS-based environmental exposomics: Current achievements and future directions // *Trends Environ. Anal. Chem.* 2020. V. 28. P. e00099. <https://doi.org/10.1016/j.teac.2020.e00099>
89. Fan Z., Alley A., Ghaffari K., Ransom H.W. MetFID: Artificial neural network-based compound fingerprint prediction for metabolite annotation // *Metabolomics*. 2020. V. 16. P. 1. <https://doi.org/10.1007/s11306-020-01726-7>
90. Young A., Wang B., Röst H. MassFormer: Tandem mass spectrum prediction with graph transformers // *arXiv Preprint*. 2021. arXiv:2111.04824. <https://doi.org/10.48550/arXiv.2111.04824>
91. Murphy M., Jegelka S., Fraenkel E., Kind T., Healey D., Butler T. Efficiently predicting high resolution mass spectra with graph neural networks // *arXiv Preprint*. 2023. arXiv:2301.11419. <https://doi.org/10.48550/arXiv.2301.11419>
92. Hoffmann M.A., Nothias L.F., Ludwig M., Fleischauer M., Gentry E.C., Witting M., Dorrestein P.C., Dührkop K., Böcker S. High-confidence structural annotation of metabolites absent from spectral libraries // *Nat. Biotechnol.* 2022. V. 40. № 3. P. 411. <https://doi.org/10.1038/s41587-021-01045-9>
93. Bremer P.L., Vaniya A., Kind T., Wang S., Fiehn O. How well can we predict mass spectra from structures? Benchmarking competitive fragmentation modeling for metabolite identification on untrained tandem mass spectra // *J. Chem. Inf. Model.* 2022. V. 62. № 17. P. 4049. <https://doi.org/10.1021/acs.jcim.2c00936>
94. Milman B.L., Ostrovidova E.V., Zhurkovich I.K. Isomer differentiation using in silico MS² spectra. A case study for the CFM-ID mass spectrum predictor // *Mass*

- Spectrom. Lett. 2019. V. 10. № 3. P. 93.
<https://doi.org/10.5478/MSL.2019.10.3.93>
95. Wang F., Liigand J., Tian S., Arndt D., Greiner R., Wishart D.S. CFM-ID 4.0: More accurate ESI-MS/MS spectral prediction and compound identification // *Anal. Chem.* 2021. V. 93. № 34. P. 11692.
<https://doi.org/10.1021/acs.analchem.1c01465>
96. Koopman J., Grimme S. From QCEIMS to QCxMS: A tool to routinely calculate CID mass spectra using molecular dynamics // *J. Am. Soc. Mass Spectrom.* 2021. V. 32. № 7. P. 1735.
<https://doi.org/10.1021/jasms.1c00098>
97. Schnegotzki R., Koopman J., Grimme S., Süßmuth R.D. Quantum chemistry-based molecular dynamics simulations as a tool for the assignment of ESI-MS/MS spectra of drug molecules // *Chem. – Eur. J.* 2022. V. 28. № 27. Article e202200318.
<https://doi.org/10.1002/chem.202200318>
98. Dührkop K., Fleischauer M., Ludwig M., Aksenov A.A., Melnik A.V., Meusel M., Dorrestein P.C., Rousu J., Böcker S. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information // *Nat. Methods.* 2019. V. 16. № 4. P. 299.
<https://doi.org/10.1038/s41592-019-0344-8>
99. Stravs M.A., Dührkop K., Böcker S., Zambo-ni N. MSNovelist: De novo structure generation from mass spectra // *Nat. Methods.* 2022. V. 19. № 7. P. 865.
<https://doi.org/10.1038/s41592-022-01486-3>
100. Zulfiqar M., Gadelha L., Steinbeck C., Sorokina M., Peters K. MAW: The reproducible metabolome annotation workflow for untargeted tandem mass spectrometry // *J. Cheminform.* 2023. V. 15. № 1. P. 32.
<https://doi.org/10.1186/s13321-023-00695-y>
101. Liu Y., De Vijlder T., Bittremieux W., Laukens K., Heyndrickx W. Current and future deep learning algorithms for tandem mass spectrometry (MS/MS)-based small molecule structure elucidation // *Rapid Commun. Mass Spectrom.* 2021. P. e9120.
<https://doi.org/10.1002/rcm.9120>
102. Niessen W.M.A., Correa C.R.A. Interpretation of MS-MS Mass Spectra of Drugs and Pesticides. Hoboken: Wiley, 2017. https://toc.library.ethz.ch/objects/pdf03/e01_978-1-118-50018-7_01.pdf (24.03.2023)
103. Steckel A., Schlosser G. An organic chemist's guide to electrospray mass spectrometric structure elucidation // *Molecules.* 2019. V. 24. № 3. P. 611. <https://doi.org/10.3390/molecules24030611>
104. Matyushin D.D., Buryak A.K. Gas chromatographic retention index prediction using multimodal machine learning // *IEEE Access.* 2020. V. 8. P. 223140.
<https://doi.org/10.1109/ACCESS.2020.3045047>
105. Matyushin, D.D., Sholokhova A.Y., Karnaeva A.E., Buryak A.K. Various aspects of retention index usage for GC-MS library search: A statistical investigation using a diverse data set // *Chemom. Intell. Lab. Syst.* 2020. V. 202. P. 104042.
<https://doi.org/10.1016/j.chemolab.2020.104042>
106. Kireev A., Osipenko S., Mallard G., Nikolaev E., Kostyukevich Y. Comparative prediction of gas chromatographic retention indices for GC/MS identification of chemicals related to chemical weapons convention by incremental and machine learning methods // *Separations.* 2022. V. 9. № 10. P. 265.
<https://doi.org/10.3390/separations9100265>
107. Domingo-Almenara X., Guijas C., Billings E., Montenegro-Burke J.R., Uritboonthai W., Aisporna A.E., Chen E., Benton H.P., Siuzdak G. The METLIN small molecule dataset for machine learning-based retention time prediction // *Nat. Commun.* 2019. V. 10. № 1. P. 5811.
<https://doi.org/10.1038/s41467-019-13680-7>
108. Witting M., Böcker S. Current status of retention time prediction in metabolite identification // *J. Sep. Sci.* 2020. V. 43. № 9-10. P. 1746.
<https://doi.org/10.1002/jssc.202000060>
109. Bonini P., Kind T., Tsugawa H., Barupal D. K., Fiehn O. Retip: Retention time prediction for compound annotation in untargeted metabolomics // *Anal. Chem.* 2020. V. 92. № 11. P. 7515.
<https://doi.org/10.1021/acs.analchem.9b05765>
110. Fedorova E.S., Matyushin D.D., Plyushchenko I.V., Stavrianidi A.N., Buryak A.K. Deep learning for retention time prediction in reversed-phase liquid chromatography // *J. Chromatogr. A.* 2022. V. 1664. Article 462792.
<https://doi.org/10.1016/j.chroma.2021.462792>
111. Osipenko S., Nikolaev E., Kostyukevich Y. Retention time prediction with message-passing neural networks // *Separations.* 2022. V. 9. № 10. P. 291.
<https://doi.org/10.3390/separations9100291>
112. Lenski M., Maallem S., Zarcone G., Garçon G., Lo-Guidice J. M., Anthérieu S., Allorge D. Prediction of a large-scale database of collision cross-section and retention time using machine learning to reduce false positive annotations in untargeted metabolomics // *Metabolites.* 2023. V. 13. № 2. P. 282.
<https://doi.org/10.3390/metabo13020282>
113. Bouwmeester R., Martens L., Degroeve S. Generalized calibration across liquid chromatography setups for generic prediction of small-molecule retention times // *Anal. Chem.* 2020. V. 92. № 9. P. 6571.
<https://doi.org/10.1021/acs.analchem.0c00233>
114. Paglia G., Smith A.J., Astarita G. Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics // *Mass Spectrom. Rev.* 2022. V. 41. № 5. P. 722.
<https://doi.org/10.1002/mas.21686>
115. Belova L., Caballero-Casero N., Van Nuijs A.L., Covaci A. Ion mobility-high-resolution mass spectrometry (IM-HRMS) for the analysis of contaminants of emerging concern (CECs): Database compilation and application to urine samples // *Anal.*

- Chem. 2021. V. 93. № 16. P. 6428.
<https://doi.org/10.1021/acs.analchem.1c00142>
116. Hohrenk L., Itzel F., Baetz N., Tuerk J., Vosough M., Schmidt T.C. Comparison of software tools for LC-HRMS data processing in non-target screening of environmental samples // *Anal. Chem.* 2019. V. 92. № 2. P. 1898.
<https://doi.org/10.1021/acs.analchem.9b04095>
117. Dekermanjian J., Labeikovsky W., Ghosh D., Kechris K. MSCAT: A machine learning assisted catalog of metabolomics software tools // *Metabolites*. 2021. V. 11. № 10. P. 678.
<https://doi.org/10.3390/metabo11100678>
118. Schymanski E.L., Singer H.P., Slobodnik J., Ipolyi I.M., Oswald P., Krauss M., Schulze T., Haglund P., Letzel T., Grosse S., Thomaidis N.S., Bletsou A., Zwiener C., Ibáñez M., Portolés T., De Boer R., Reid M.J., Onghena M., Kunkel U., Schulz W., Guillon A., Noyon N., Leroy G., Bados P., Bogialli S., Stipaničev D., Rostkowski P., Hollender J. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis // *Anal. Bioanal. Chem.* 2015. V. 407. P. 6237.
<https://doi.org/10.1007/s00216-015-8681-7>
119. CASMI. <http://www.casmi-contest.org/2022/index.shtml> (25.03.2023).
120. Pezzatti J., González-Ruiz V., Boccard J., Guillarme D., Rudaz S. Evaluation of different tandem MS acquisition modes to support metabolite annotation in human plasma using ultra high-performance liquid chromatography high-resolution mass spectrometry for untargeted metabolomics // *Metabolites*. 2020. V. 10. № 11. P. 464.
<https://doi.org/10.3390/metabo10110464>
121. Clark T.N., Houriet J., Vidar W.S., Kellogg J.J., Todd D.A., Cech N.B., Linington R.G. Interlaboratory comparison of untargeted mass spectrometry data uncovers underlying causes for variability // *J. Nat. Prod.* 2021. T. 84. №. 3. C. 824.
<https://doi.org/10.1021/acs.jnatprod.0c01376>
122. Wong J.W., Wang J., Chang J.S., Chow W., Carlson R., Rajski Ł., Fernández-Alba A.R., Self R., Cooke W.K., Lock C.M., Mercer G.E., Mastovska K., Schmitz J., Vavclavik L., Li L., Panawennage D., Pang G.F., Zhou H., Miao S., Ho C., Lam T.C.H., To Y.B.S., Zomer P., Hung Y.C., Lin S.W., Liao C.D., Culberston D., Taylor T., Wu Y., Yu D., Lim P.L., Wu Q., Schirlé-Keller J.P.X., Williams S.M., Johnson Y.S., Nason S.L., Ammirata M., Eitzer B.D., Willis M., Wyatt S., Kwon S.Y., Udawatte N., Priyasantha K., Wan P., Filigenzi M.S., Bakota E.L., Sumarah M.W., Renaud J.B., Parinet J., Biré R., Hort V., Prakash S., Conway M., Pyke J.S., Yang D.H.D., Jia W., Zhang K., Hayward D.G. Collaborative study of a nontarget data acquisition for target analysis (nDATA) workflow using liquid chromatography-high-resolution accurate mass spectrometry for pesticide screening in fruits and vegetables // *J. Agric. Food Chem.* 2021. V. 69. № 44. P. 13200.
<https://doi.org/10.1021/acs.jafc.1c04437>
123. Anderson B.G., Raskind A., Habra H., Kennedy R.T., Evans C.R. Modifying chromatography conditions for improved unknown feature identification in untargeted metabolomics // *Anal. Chem.* 2021. V. 93. № 48. P. 15840.
<https://doi.org/10.1021/acs.analchem.1c02149>