

УДК 621.384.668.8: 577.112.6

© Я. И. Лютвинский, С. С. Присяч, Н. В. Краснов,
С. В. Фионов, А. Н. Арсеньев

ИНТЕРАКТИВНОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ В БИОХИМИЧЕСКОЙ МАСС-СПЕКТРОМЕТРИИ

В работе рассмотрены проблемы интерфейсов пользователя программного обеспечения для биохимической масс-спектрометрии. Продемонстрированы подходы и способы преодоления данных проблем на примере программного обеспечения TOF Explorer для времяпролетного масс-спектрометра MX 5310/11. Приведены примеры уже реализованных интерфейсов пользователя в биохимической масс-спектрометрии.

Кл. сл.: программное обеспечение прибора, MX 5310, MX 5311, интерфейс пользователя

ВВЕДЕНИЕ

Новации высокомолекулярной масс-спектрометрии в области алгоритмов анализа данных играют не меньшую роль, чем совершенствование аналитической части самих масс-спектрометров. В то же время, какими бы совершенными и технологичными ни были алгоритмы, они останутся невостребованными конечным пользователем в случае неясной и неудобной реализации работы с ними.

Комплекс TOF Explorer, являющийся сопроводительным программным обеспечением для времяпролетных масс-спектрометров с источником ионизации "Электроспрей" серии MX 5310/11, разработанных в Лаборатории биомедицинской масс-спектрометрии Института аналитического приборостроения РАН в рамках целевой научно-технической программы "Разработка уникальных научно-исследовательских приборов и оборудования для учреждений РАН", интегрирует как стандартные, так и специально для него созданные алгоритмы обработки данных масс-спектрометрического эксперимента [1–7]. Основные цели, поставленные при проектировании его пользовательского интерфейса, — сделать понятными и востребованными реализованные алгоритмы. В данной работе мы представляем разработанные нами методы визуализации и приемы работы с данными масс-спектрометрических экспериментов.

ПРИНЦИП СОЗДАНИЯ TOF EXPLORER

Современные потребители программного обеспечения предпочитают использовать в работе не набор обособленных программ, выполняющих отдельные задачи, а единую multifunctionальную

рабочую среду, обеспечивающую доступ к различным алгоритмам, ясный и свободный переход от одного режима работы к другому. С этой целью TOF Explorer изначально проектировался в парадигме "predefined desktop", при которой единое главное окно является самодостаточной пользовательской средой.

По-видимому, такое решение проблемы становится общепринятым. Так, фирма-разработчик известного программного обеспечения MassLynx при смене версии с 3.5 на 4.0, потребовавшей более 3 лет, отказалась от концепции MDI (multi document interface), при которой каждая функция представляется в отдельном окне, в пользу обсуждаемого "predefined desktop".

Реализация "predefined desktop" в TOF Explorer основывается на специализированных управляющих структурах (рис. 1):

- "дерево рабочего пространства" (отображает иерархию полученных в процессе работы TOF Explorer данных);
- складывающиеся панели управления (содержат параметры алгоритмов, разделенные по направлениям работы);
- закладки визуализируемой информации (диаграммы и графики);
- сводная таблица данных (инструмент отображения всех данных в числовой форме, удобной для экспорта в другие программы).

ИНТЕГРАЦИЯ С ДРУГИМИ ПРОГРАММНЫМИ ПРОДУКТАМИ

Отдельной задачей проектирования программной системы является обеспечение ее взаимодействия с другими программными продуктами. Пользователь не должен испытывать трудностей

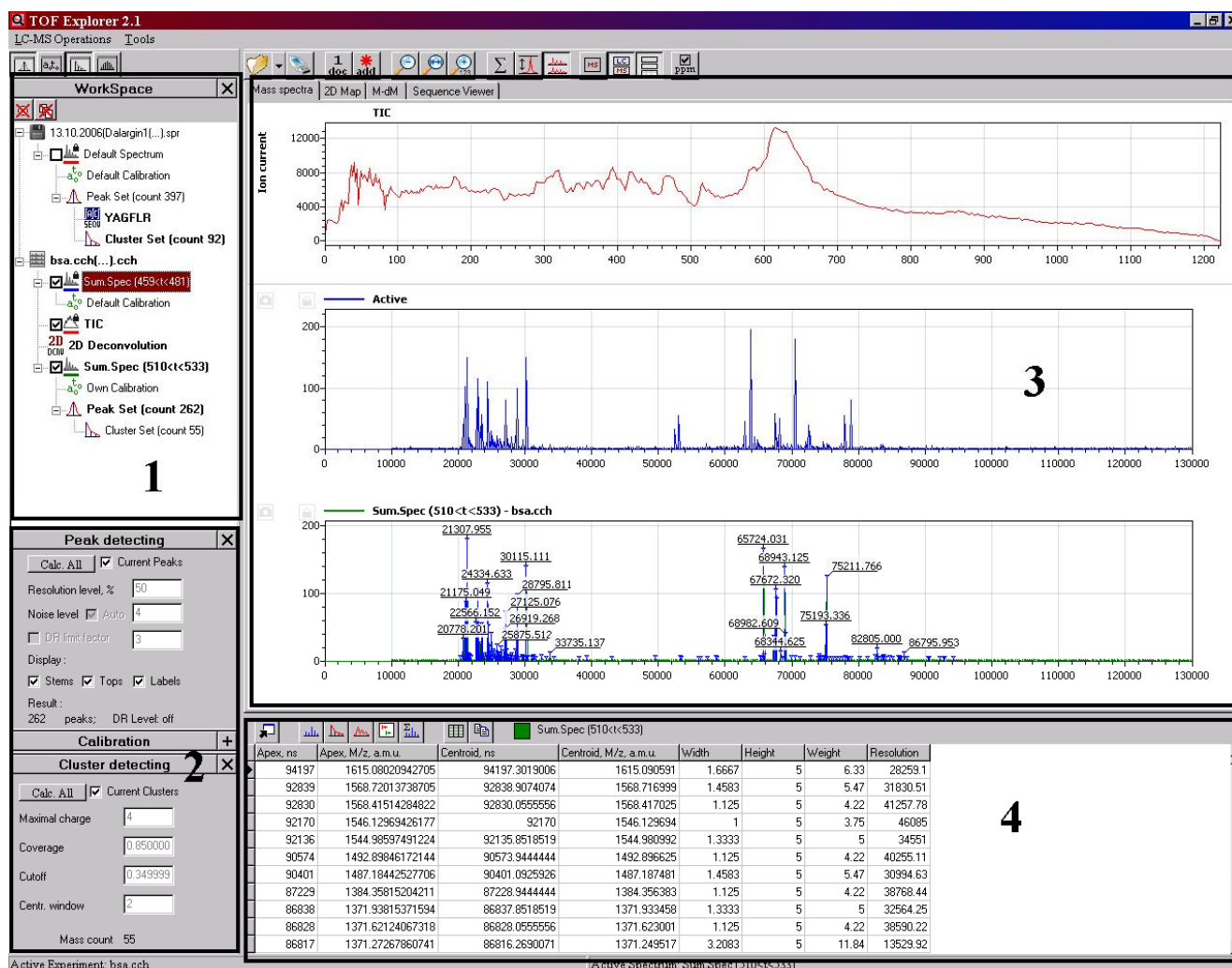


Рис. 1. Пример вида главного окна TOF Explorer.

Видны 4 основные графические структуры, на которых базируется "predefined desktop" интерфейс: 1 — "дерево рабочего пространства"; 2 — складывающиеся панели управления; 3 — закладки (активная "Mass spectra", а также "2D Map", "M-dM" и "Sequence Viewer"); 4 — сводная таблица данных

при переносе результатов работы алгоритмов в другие пакеты, однако реализовать конвертирование в множество закрытых коммерческих форматов хранения данных трудно осуществимо на практике.

Для восполнения этой потребности в TOF Explorer выработан собственный подход к экспорту данных. Он подразумевает использование в качестве единицы переноса информации не файл, а буфер обмена. Таким образом удастся избежать проблем с перекодировкой форматов файлов, поскольку этот механизм априори поддерживается большинством прикладных программ. Однако потребовалась дополнительная разработка инструмента для просмотра и экспорта сводной таблицы данных. В таблицу автоматически заносятся все

полученные и рассчитанные числовые и буквенные результаты. Здесь они могут быть отсортированы, отобраны и скопированы в буфер обмена. Далее пользователь имеет возможность перенести эти данные для обработки в необходимую программу. Пример работы со сводной таблицей данных представлен на рис. 2.

В состав TOF Explorer интегрированы следующие подпрограммы: визуализация "сырых" данных; визуализация изотопных распределений; визуализация данных (ВЭЖХ-МС)-экспериментов; визуализация определения зависимости "масса—дефект массы"; визуализация и поиск аминокислотных последовательностей на фрагментном масс-спектре.

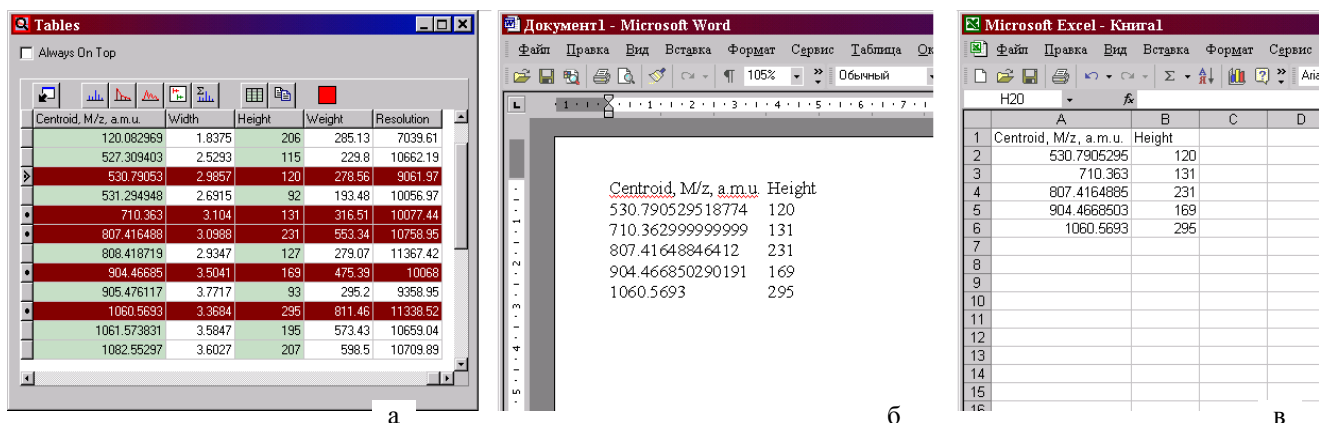


Рис. 2. Экспорт информации посредством сводной таблицы данных.

На примере: в сводную таблицу внесены данные о всех найденных на масс-спектре пиках (а); отображена информация (выделенные столбцы таблицы) о центрах масс (Centroid, M/z, a.m.u) и высотах пиков (Height); отображены данные о пяти пиках (выделенные строки таблицы) и перенесены в Word (б) и Excel (в)

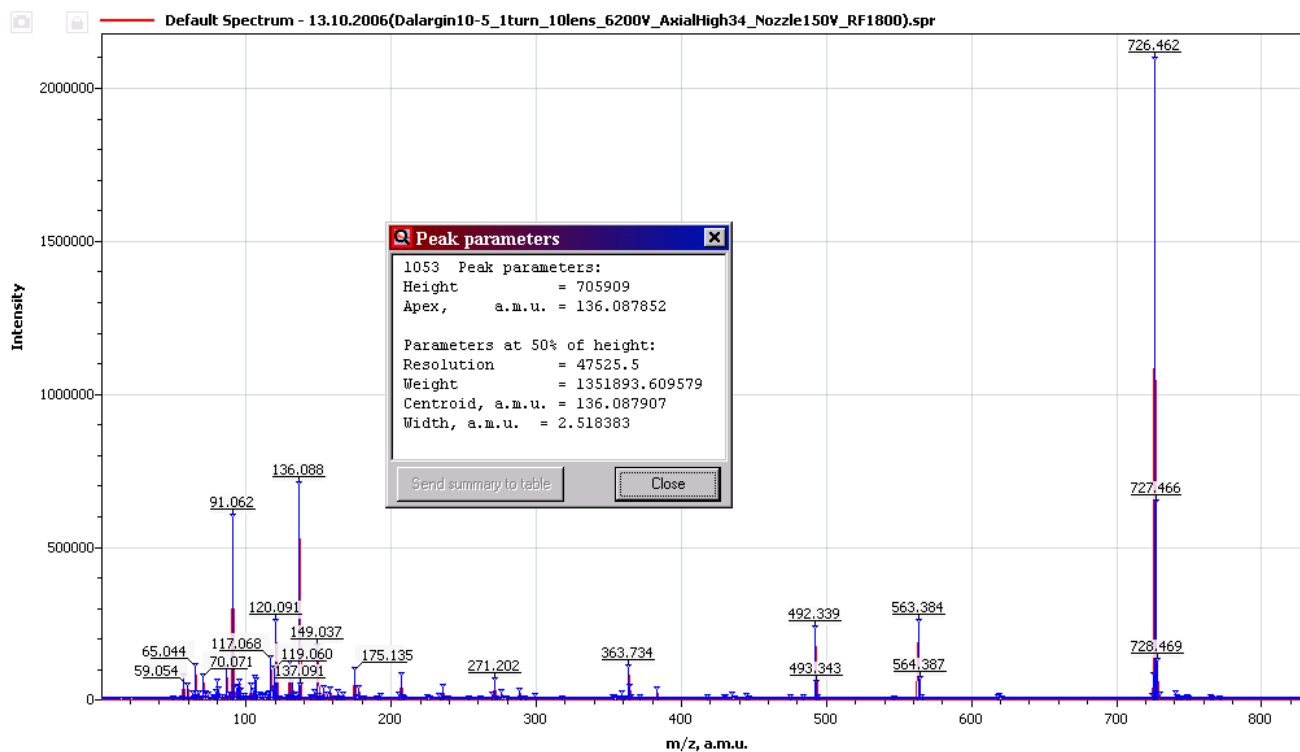


Рис. 3. Результат работы алгоритма TOF Explorer по неперекрывающемуся расположению меток пиков. Метки располагаются преимущественно над самими пиками. Но в случаях, когда такое расположение влечет наложения, метки смещаются

ВИЗУАЛИЗАЦИЯ НЕОБРАБОТАННЫХ МАСС-СПЕКТРОМЕТРИЧЕСКИХ ДАННЫХ

Типичный масс-спектр содержит порядка нескольких сотен пиков. В каждый момент времени пользователю необходимо видеть точные значения

масс только для наиболее важных пиков спектра, что требует наличия механизма скрытия остальных меток пиков. Наиболее очевидный подход, при котором "неважными" считаются пики с интенсивностью ниже определенного уровня, не является достаточным. Этот подход слишком часто

исключает информативные, но малые пики, особенно в области высоких масс, и не гарантирует того, что метки с указанием масс пиков не будут перекрываться. В программном обеспечении TOF Explorer применен алгоритм визуализации пиков, учитывающий не только их величину, но и плотность распределения на спектре (рис. 3). Во избежание избыточности отображаемой информации только двадцать локально наиболее высоких пиков снабжаются метками с указанием величины апекса. Место для меток находится через решение геометрической задачи о неперекрывающемся расположении прямоугольников на площади. Допускается смещение метки относительно центра пика в виде выноски. Сохранение возможности наблюдать все выявленные пики одновременно обеспечивается построением на экране так называемого "стержня" — отрезка, соответствующего каждому пику. Также в любой момент времени для любого представленного на экране пика могут быть показаны более подробные характеристики в отдельном окне.

ВИЗУАЛИЗАЦИЯ ИЗОТОПНЫХ РАСПРЕДЕЛЕНИЙ

Современный математический анализ часто является многоступенчатым. При этом последующие данные наслаиваются на полученные ранее. Так, при проведении зарядной и изотопной деконволюции (преобразование спектра в список пар "моноизотопная молекулярная масса—интегральная интенсивность сигнала", необходимое для преодоления зарядового и изотопного разнообразия ионов вещества) возникает потребность зрительно и функционально выделить *изотопные кластеры* (группы пиков на спектре, соответствующие изотопному распределению молекулярных масс заданного химического соединения).

Масс-спектрометры с источником ионов "электроспрей" наиболее востребованы в белковом анализе, поэтому в состав TOF Explorer включен алгоритм распознавания изотопных кластеров с изотопным распределением, типичным для пептидов. При распознавании изотопных кластеров используются модельные изотопные распределения для пептидов заданной массы. Этот алгоритм может быть использован для других классов веществ (сахара, нуклеотиды), если рассчитать изотопные распределения, типичные для веществ этих классов.

При визуализации изотопного распределения вещества встают две антагонистичные проблемы. С одной стороны, размеченные ЭВМ изотопные кластеры должны быть представлены специалисту для верификации, что требует сохранения возможности видеть исходный спектр. С другой стороны, важно иметь возможность легко зрительно

идентифицировать предложенный изотопный кластер и иметь быстрый доступ к его параметрам.

В TOF Explorer данную проблему решает особый механизм визуализации изотопных кластеров. Их разметка производится поверх графика спектра, с которого убрана вся дополнительная информация помимо "стержней" пиков, отображающих их апексы. Это сохраняет для пользователя возможность судить о правильности результатов.

Пики одного химического соединения отмечаются на экране монитора квадратной меткой. Уровень расположения метки означает интенсивность, которая была бы у данного пика в случае полного соответствия имеющегося изотопного распределения его модели. Все пики изотопного распределения объединены общей ломаной линией, пересекающей пики на уровне их меток. Это позволяет видеть изотопное распределение таким, каким оно должно бы было быть для класса соединений. Примеры изотопных распределений изображены в Приложении на рис. П1. Дополнительно метка первого пика в распределении выделяется цветом.

ВИЗУАЛИЗАЦИЯ ДАННЫХ (ВЭЖХ-МС)-ЭКСПЕРИМЕНТОВ

Одним из режимов работы МС с источником ионов "электроспрей" является режим "on-line" с ВЭЖХ. Получаемые в этом случае данные содержат серию последовательно записанных спектров. Стандартный способ визуализации — это представление таких данных с помощью так называемых трейсов (диаграмм интеграла ионного тока в пределах заданного диапазона масс за время всего эксперимента) и интегрированных спектров (интегрированных масс-спектров в полном диапазоне масс в заданном диапазоне времени (ВЭЖХ-МС)-эксперимента). Такой подход реализован в TOF Explorer (рис. П2, Приложение), однако он не дает представления о результате ВЭЖХ-МС в целом и при анализе сложных смесей приводит к большому объему рутинной работы.

Чтобы преодолеть эти недостатки, в TOF Explorer реализована возможность визуализации всего эксперимента в целом как двумерной диаграммы (рис. П3). Ось ординат сопоставляется время регистрации спектров, ось абсцисс — отношение массы к заряду. Интенсивность ионного тока отображается интенсивностью цвета.

На рис. П2–П3 отображен один и тот же фрагмент данных (ВЭЖХ-МС)-эксперимента. Как можно видеть, рис. П3 гораздо более информативен. Простое наблюдение одного интегрированного трейса и спектра расширяется возможностью видеть сразу точное время и отношение масса/заряд иона одновременно для каждого пика и кластера. Метками на двумерной диаграмме

обозначены изотопные кластеры, найденные при помощи алгоритма двумерной деконволюции [8, 9]. Он работает со всей совокупностью данных (ВЭЖХ-МС)-эксперимента в целом и определяет не только моноизотопную молекулярную массу и совокупную интенсивность сигнала вещества, но и время элюции вещества из хроматографа. Также этот алгоритм обладает большей устойчивостью к шумам спектра, т. к. опирается в расчетах на серию последовательных спектров, а не на единственный спектр.

ВИЗУАЛИЗАЦИЯ ОПРЕДЕЛЕНИЯ ЗАВИСИМОСТИ "МАССА—ДЕФЕКТ МАССЫ"

Отклонения массы пиков от ближайшего целого значения может быть использовано для выявле-

ния соединений с участием тяжелых химических элементов (металлов), поскольку дефект масс элементов со значительным атомным весом отрицательный. Определение зависимости "масса—дефект массы" вручную по апексам пиков на масс-спектре очень трудоемко. Для TOF Explorer разработан специальный способ визуализации спектра, удобный при наблюдении распределения дефектов масс пиков (рис. 4). Диаграмма анализируемого спектра с размеченными пиками отображается согласованно с диаграммой распределения дефектов масс пиков, по вертикальной оси которой отложен дефект масс, а по горизонтальной — отношение "масса иона / заряд иона".

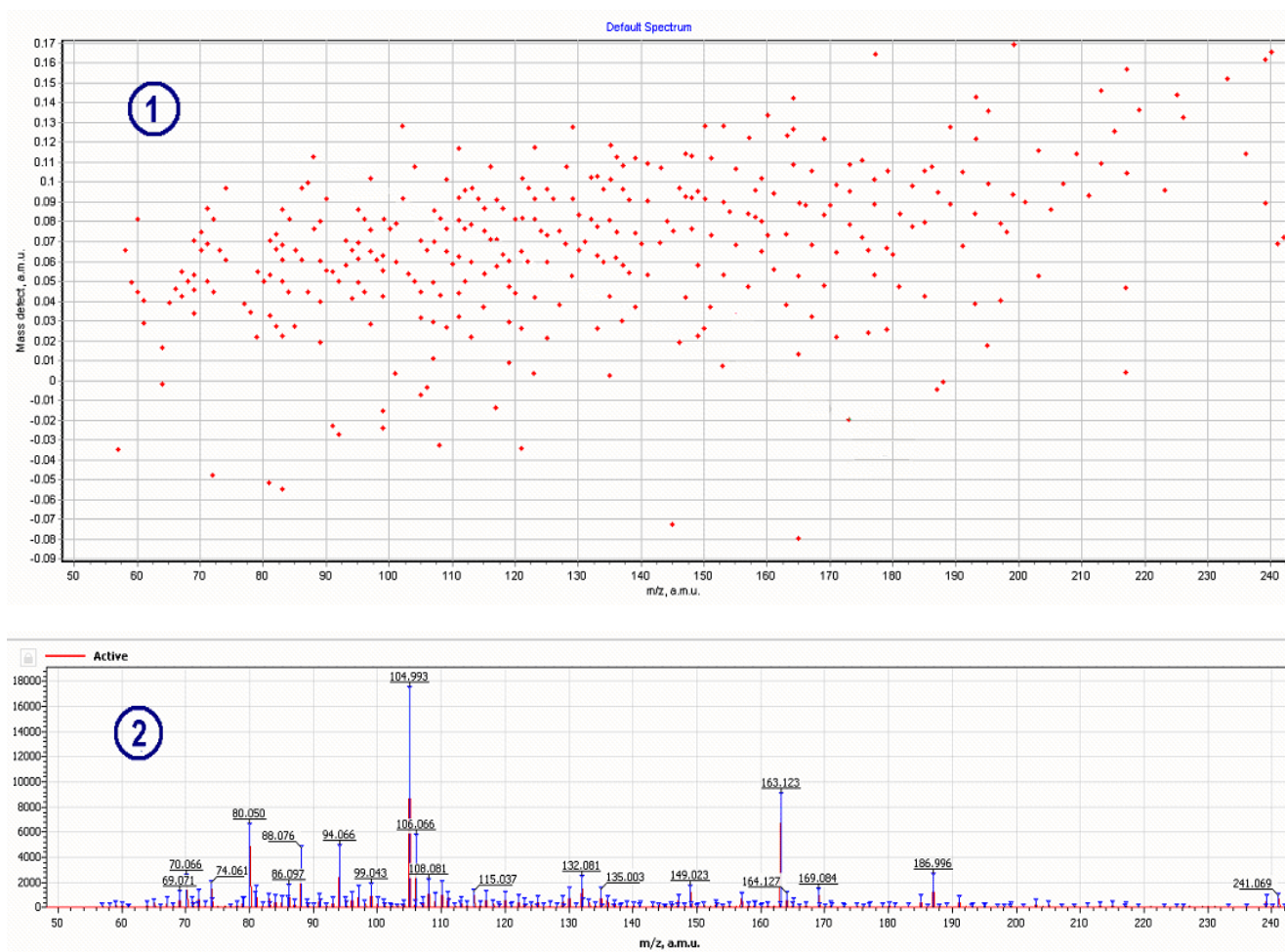


Рис. 4. Наблюдение распределения дефектов масс пиков средствами TOF Explorer.
1 — диаграмма распределения дефектов масс пиков; 2 — график масс-спектра

ВИЗУАЛИЗАЦИЯ И ПОИСК АМИНОКИСЛОТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ФРАГМЕНТНОМ МАСС-СПЕКТРЕ

Приборы серии MX 5310/11 успешно используются для определения аминокислотных последовательностей относительно коротких (до 15 аминокислотных последовательностей) пептидов [10–14] при условии чистоты препарата (одно вещество в образце). Спектры пептидов, полученные в режиме фрагментации в источнике, аналогичны спектрам, полученным на тандемных (МС-МС)-приборах и содержат сигналы, коррелирующие с аминокислотной последовательностью пептида. Обозначение информативных ионов и распознанных аминокислотных последовательностей в TOF Explorer в целом базируется на общепринятой нотации аминокислотных последовательностей [15, 16], однако для работы с неполностью распознанными последовательностями, и особенно спектрами модифицированных пептидов, была разработана и введена расширенная нотация записи аминокислотных последовательностей. Ее описание в виде синтаксической диаграммы представлено в таблице. Эта нотация позволяет выражать как известные, хранящиеся в словаре TOF Explorer химические модификации аминокислот, так и неизвестные модификации, представленные химическими формулами или изменением молекулярной массы. Например, эту нотацию можно использовать для записи такого пептида:

[214.15][S+ 47.985]YpAMP[K-H+COCH₃],

где

214.15 — невосстановленный участок аминокислотной последовательности массой 214.15 а.е.м.;

S+ 47.985 — трижды окисленный метионин, записанный как изменение массы аминокислотного остатка;

Yp — фосфорилированный тирозин;

K-H+COCH₃ — ацетилованный лизин, где модификация представлена, как химическая формула.

Решение проблемы предоставления пользователю максимально удобного и знакомого метода работы с аминокислотными последовательностями на фрагментных масс-спектрах в TOF Explorer реализовано полным набором возможных подходов:

- верификация известной аминокислотной последовательности по ее нотационному выражению;
- разметка аминокислотной последовательности, выполняемая экспертом;
- автоматическое определение и поиск последовательности в протеомных базах данных [4, 5].

Нотация TOF Explorer для записи аминокислотной последовательности

<Capital letter>	::=	"A" ... "Z"
<lower case letter>	::=	"a" ... "z"
<number>	::=	{ "1" ... "9" "0" }
<float number>	::=	<number> "." <number>
<Valid symbol>	::=	{ <lower case letter> "#" "\$" "&" "*" "." "/" "\" ":" "_" "~" "!" "@" }
<Chemical element>	::=	<Capital letter> { <lower case letter> } { <number> }
<Chemical addition>	::=	"+" "-" (<float number> <number> { <Chemical element> })
<Elements>	::=	<Chemical element> [<number>]
<Residue modification>	::=	{ <valid symbol> }
<End modification>	::=	{ <valid symbol> }
<Pure amino acid>	::=	"G" "A" "R" "L" "Y" "F" "E" "C" "H" "K" "V" "N" "Q" "M" "S" "T" "W" "P" "I"
<Amino acid>	::=	<Pure amino acid> [<Residue modification>]
<Token>	::=	"[" (<Amino acid> <float number>) [{ <Chemical addition> }] "]"
<end modified element>	::=	(<Pure amino acid> <Token>) [<End modification>]
<peptide>	::=	<end modified element> [{ <Token> <Amino acid> }] <end modified element>

Алгоритмы TOF Explorer построены таким образом, что ввод символьных значений последовательности и графическая разметка дополняют друг друга. В любой момент указанная символами последовательность может быть дополнена графической разметкой и наоборот. Дополнительные управляющие и информационные элементы в алгоритмах разметки последовательности общие (см. рис. П4, Приложение).

ЗАКЛЮЧЕНИЕ

Уровень сопроводительного программного обеспечения отечественного масс-спектрометра MX 5310/11, разработанного в ИАП РАН, по основным направлениям соответствует аналогичным разработкам зарубежным фирм и по некоторым направлениям опережает их. Программный комплекс TOF Explorer является самостоятельным продуктом, который постоянно развивается.

ПРИЛОЖЕНИЕ

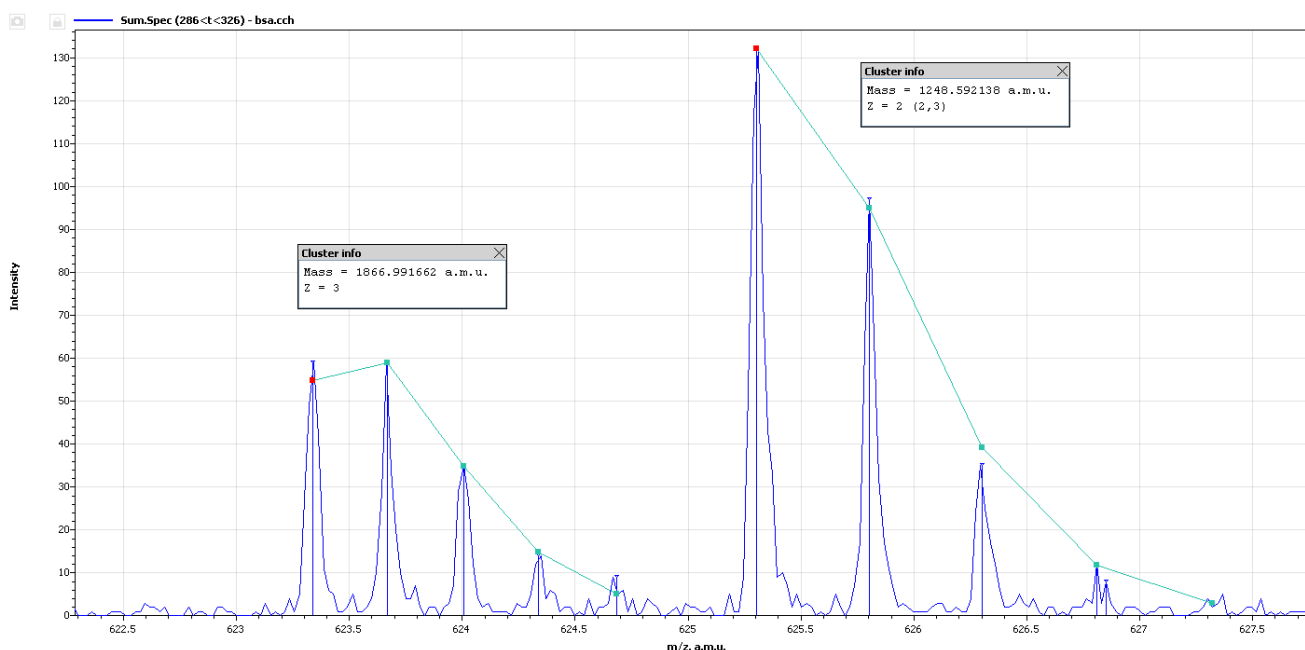


Рис. П1. Выделение на графике масс-спектра групп пиков, соответствующих отдельным химическим веществам, средствами TOF Explorer. Пики изотопного распределения отмечены квадратными значками и объединены отрезками

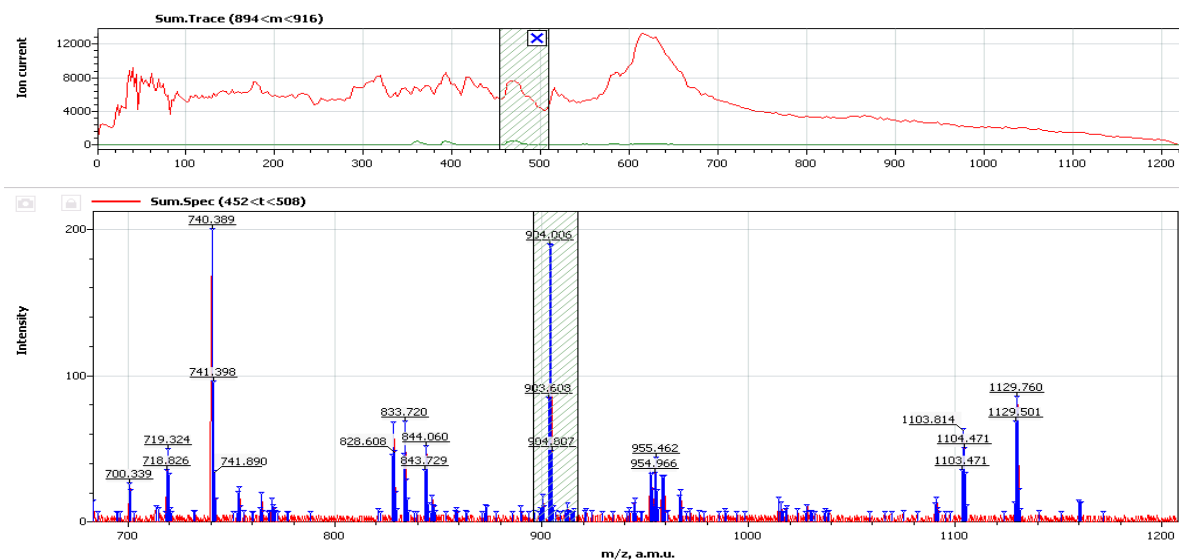


Рис. П2. Визуализация данных (ВЭЖХ-МС)-экспериментов, реализованная в программе TOF Explorer

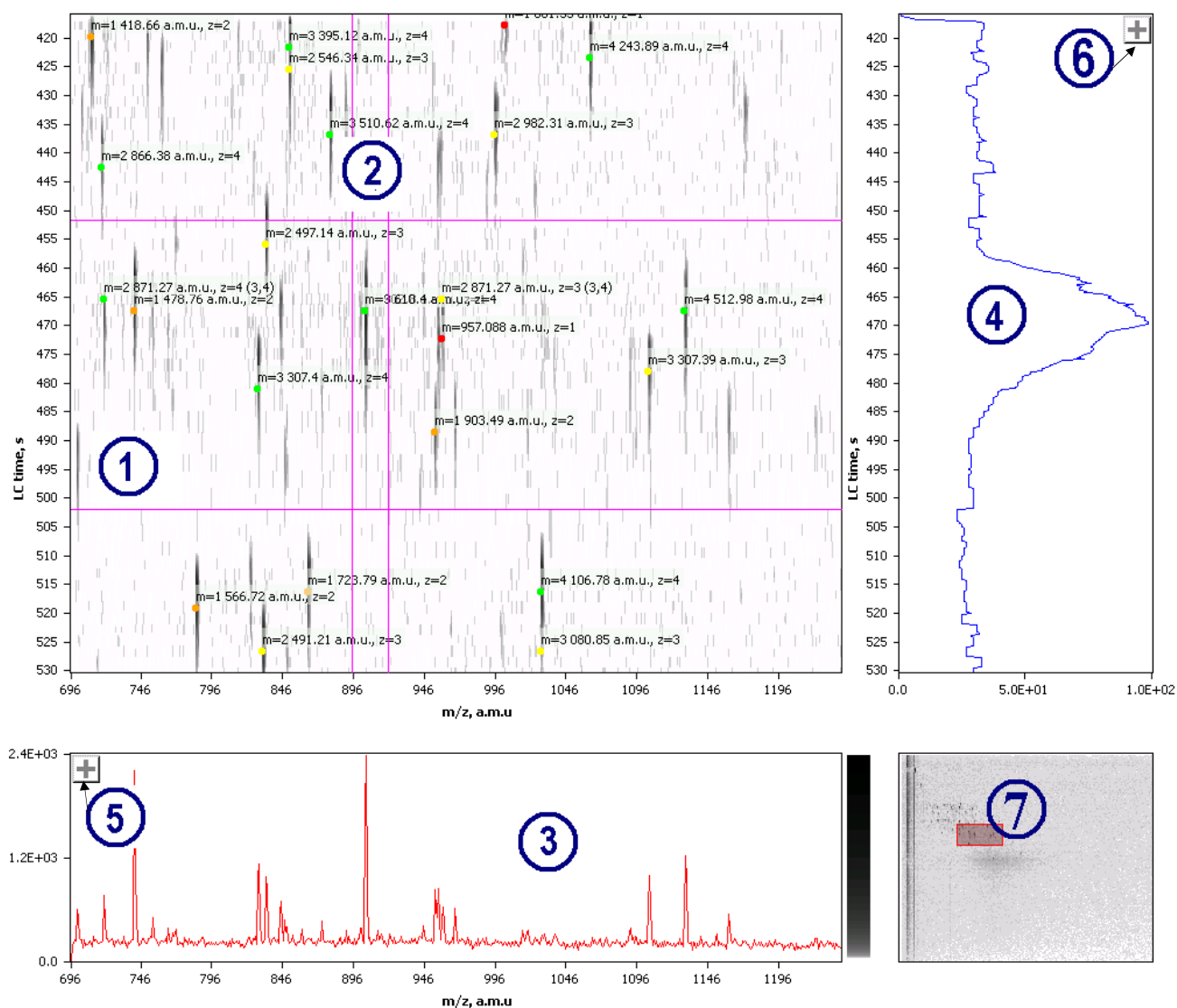


Рис. ПЗ. Отображение временной динамики спектров всего хромато-масс-спектрометрического эксперимента.

- 1 — указан интервал спектра;
- 2 — указан интервал трейса;
- 3 и 4 — соответствующие указанным интервалам текущие спектр и трейс;
- 5 и 6 — кнопки использования указанных интервалов спектра и трейса;
- 7 — макет диаграммы с обозначенным увеличенным фрагментом

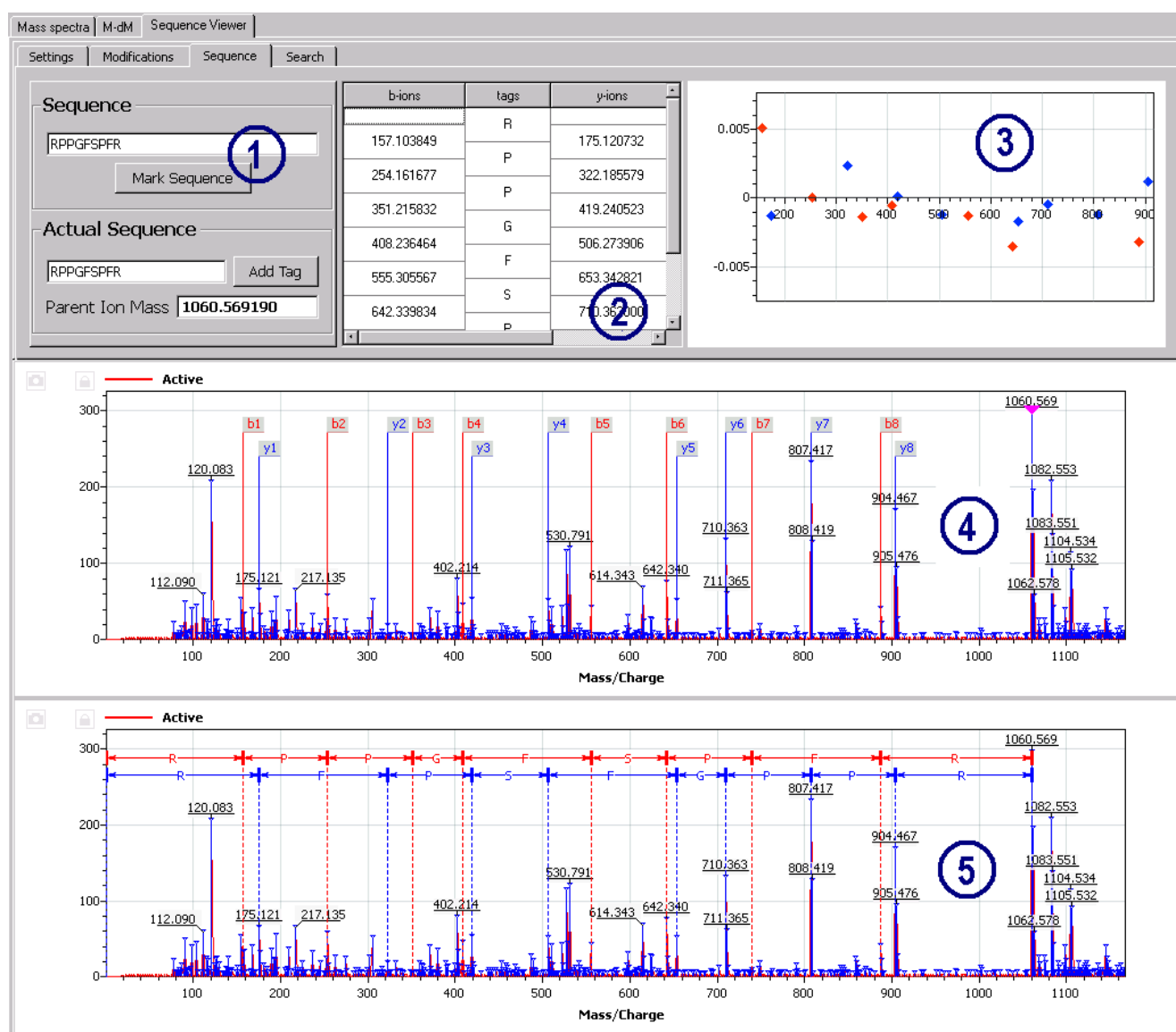


Рис. П4. Элементы управления, используемые при поиске аминокислотной последовательности на спектре.
 1 — область ввода известной последовательности;
 2 — таблица обнаруженных ионов (отображает информацию о массах найденных ионов);
 3 — диаграмма ошибок (показывает отклонение размеченных на спектре масс элементов от хранящихся в словаре их значений);
 4 — диаграмма ионов (служит для разметки и редактирования ионов b- и y-серий поверх пиков или изотопных кластеров спектра);
 5 — диаграмма последовательности (отображает информацию о найденных на спектре элементах последовательности, таких как аминокислоты, модифицированные аминокислоты, химические соединения, массы)

СПИСОК ЛИТЕРАТУРЫ

1. 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.
2. Макаров В.В., Савельев С.К., Лютвинский Я.И., Веренчиков А.Н., Краснов Н.В. Алгоритм извлечения аналитически значимой информации из масс-спектрометрических данных экспериментов протеомики // Научное приборостроение. 2006. Т. 16, № 2. С. 92–100.
3. Makarov V.V., Saveliev S.K. Data mining method for electrospray mass spectra of peptide mixtures // Proceedings of SPIE. 2005. V. 6251. P. 355–362.
4. Лютвинский Я.И., Краснов Н.В. Разработка CrystalTag — алгоритма частичного распознавания фрагментных масс-спектров пептидов // Научное

- приборостроение. 2005. Т. 15, № 3. С. 108–113.
5. *Фиронов С.В., Лютвинский Я.И., Краснов Н.В.* Использование недетерминированных конечных автоматов и индекса масс пептидов для быстрого сопоставления фрагментных масс-спектров пептидов базам данных аминокислотных последовательностей // *Научное приборостроение*. 2008. Т. 18, № 4. С. 114–119.
 6. *Веренчиков А.Н., Макаров В.В., Подольская Е.П. и др.* Интерпретация масс-спектров ионизации воздуха при атмосферном давлении с высокой точностью измерения масс // *Научное приборостроение*. 2006. Т. 16, № 3. С. 88–102.
 7. *Лютвинский Я.И., Макаров В.В., Краснов Н.В., Подольская Е.П., Веренчиков А.Н.* Частичная расшифровка аминокислотной последовательности пептида по его фрагментному масс-спектру: алгоритм и результаты применения // *Научное приборостроение*. 2006. Т. 16, № 3. С. 122–131.
 8. *Макаров В.В., Лютвинский Я.И., Веренчиков А.Н.* Алгоритм IPEX-2D для извлечения информации о компонентах пробы из массива данных (ВЭЖХ-МС)-экспериментов протеомики // *Научное приборостроение*. 2006. Т. 16, № 3. С. 107–112.
 9. *Лютвинский Я.И., Макаров В.В., Козлов Б.Н. и др.* Оценка емкости масс-спектров высокого разрешения при анализе сложных смесей // *Научное приборостроение*. 2006. Т. 16, № 3. С. 113–121.
 10. *Манойлов А.В., Торопыгин И.Ю., Козьмин Ю.П. и др.* Клиническая протеомика: новые диагностические возможности масс-спектрометрии // *Научное приборостроение*. 2008. Т. 18, № 4. С. 16–24.
 11. *Новиков А.В., Кошелева Г.Н., Бубляев Р.А. и др.* Исследование неферментативного взаимодействия селена и серебра с цистеином и глутатионом с помощью приборного комплекса MX5310 // *Научное приборостроение*. 2008. Т. 18, № 4. С. 79–88.
 12. *Лютвинский Я.И., Федорова Г.А., Новиков А.В.* Фрагментация пептидов в источнике электроспрей как способ извлечения информации о первичной структуре пептида на масс-спектре MX5305 // *Третий съезд Общества биотехнологов России им. Ю.А. Овчинникова*, 2005 г.
 13. *Безуглов В.В., Грецакая Н.М., Блаженкова А.В. и др.* Арахидоноиламинокислоты и арахидоноилпептиды: синтез и свойства // *Биоорганическая химия*. 2006. Т. 32, № 3. С. 258–267.
 14. *Новиков А.В., Назимов И.В., Русанов В.А. и др.* Контроль производства генно-инженерного инсулина человека с помощью тандема "микроколоночный жидкостный хроматограф—масс-спектрометр" в режиме прямого ввода образца // *Научное приборостроение*. 2004. Т. 14, № 2. С. 109–116.
 15. *Roepstorff P., Fohlman J.* Proposal for a common nomenclature for sequence ions in mass spectra of peptides // *Biomed. Mass Spectrom.* 1984. Nov., V. 11, N 11. P. 601.
 16. *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.

*Институт аналитического приборостроения РАН,
г. Санкт-Петербург*

Контакты: *Присяч Сергей Сергеевич,*
Prisyatch@yandex.ru

Материал поступил в редакцию 27.10.2010.

INTERACTIVE SOFTWARE FOR BIOCHEMICAL MASS SPECTROMETRY

Y. I. Lyutvinsky, S. S. Prisyatch, N. V. Krasnov, S. V. Fironov, A. N. Arseniev

Institute for Analytical Instrumentation RAS, Saint-Petersburg

The paper discusses the problem of designing interfaces to present data in biochemical mass spectrometry. The ways of overcoming it are represented. The examples of data interfaces that are implemented in software TOF Explorer for MX 5310/11 are shown.

Keywords: TOF Explorer, MX 5310, MX 5311, user interface