

# РАЗРАБОТКА УНИВЕРСАЛЬНОГО ПРОГРАММНОГО ИНТЕРФЕЙСА ДЛЯ АНАЛИЗА МАСС-СПЕКТРОМЕТРИЧЕСКОЙ ИНФОРМАЦИИ

**Н. В. Махнач, И. В. Трофименков**

*Белорусский государственный университет, г. Минск;*

*m1kach3m@gmail.com, vanyu0759@gmail.com;*

*науч. рук. – А. Н. Коваленко, ст. науч. сотр.;*

*Е. И. Коваленко, доц., канд. биол. наук*

Разработан универсальный кросс-платформенный программный интерфейс для анализа масс-спектрометрической информации, позволяющий работать с различными форматами баз данных и обращаться к различным масс-спектрометрическим библиотекам. Программный интерфейс представляет собой языковую конструкцию типа интерфейс (контракт), позволяющий передавать неизвестный масс-спектр в модуль поиска и получать из него результат поиска. Разработаны два модуля поиска mssearch1 и mssearch2 – динамически загружаемые библиотеки DLL, написанные на ObjectPascal в среде Lazarus IDE, взаимодействующие с UniChrom. Данный интерфейс протестирован с UniChrome-97 под ОС Linux Debian GNU/Linux 10.9 и Windows 7/Windows 10 64bit. Модули поиска могут быть использованы с другими программными продуктами и другими базами данных, с Web-интерфейсом.

**Ключевые слова:** масс-спектрометрия; хроматография; программный интерфейс; базы данных; коэффициенты подобия; SQL; NIST2020.

## **ВВЕДЕНИЕ**

Масс-спектрометрия (МС) – аналитический метод, применяемый в биологии, химии, медицине, экологии, контроле технологических процессов, позволяющий определять молекулярную массу и элементный состав молекул и их фрагментов [1,2]. К приоритетным направлениям МС, в частности, относятся исследования биополимеров и лабильных веществ, многокомпонентных смесей, изучение метаболизма [1,2]. Применение МС реализуется в сочетании с хроматографией и другими методами разделения веществ. Масс-спектру соответствует массив пар значений отношения массы к заряду ( $m/z$ ) и интенсивности ионного тока. Важнейшей проблемой МС является расшифровка масс-спектров и интерпретация информации, для чего необходимо применять компьютерную обработку результатов. Процедура расшифровки масс-спектра требует обращения к базе данных (БД). Среди БД крупнейшими являются библиотеки Национального института стандартов и технологий США (NIST) платного доступа, содержащие до миллиона спектров, а также открытые библиотеки (MassBank и др.). Производители оборудования для МС сохраняют данные в проприетарных форматах, что ограничивает пользователя и принуждает применять программные пакеты

только этого производителя. Таким образом, для повышения продуктивности работы важным является создание универсального программного интерфейса для работы с различными форматами данных и библиотеками, разработка которого и явилась целью данной работы.

## РЕАЛИЗАЦИЯ ПРОГРАММНОГО ИНТЕРФЕЙСА

Работа выполнена с применением NIST-MS Library, набора библиотек разработчика NIST-MS SDK (system development kit), кросс-платформенного компилятора ObjectPascal Free Pascal Compiler, совместимого с Delphi Pascal, Lazarus IDE (интегрированная среда разработки для Free Pascal Compiler), SQL сервера Firebird, системы регистрации, обработки и хранения спектрометрической информации UniChrom-97. Программный интерфейс представляет собой языковую конструкцию типа интерфейс (контракт), позволяющий передавать неизвестный масс-спектр в модуль поиска и получать из него результат поиска. Разработаны два модуля поиска mssearch1 и mssearch2 – динамически загружаемые библиотеки DLL, написанные на ObjectPascal в среде Lazarus IDE, взаимодействующие с UniChrom.

Оба модуля реализуют общий алгоритм поиска в произвольной базе данных, включающий следующие шаги.

1. Получить масс-спектр в указанной временной точке.
2. Выполнить вычитание шума.
3. Создать контекст поиска (для каждого хромато-масс-спектра создается свой контекст).
4. Передать спектр в выбранный контекст в модуле поиска.
5. Вызвать метод поиска кандидатов (масс-спектров веществ, подобных неизвестному).
6. Выполнить подготовку БД: открыть соединение с БД.
7. Выполнить предварительный поиск.
8. Выполнить полный поиск кандидатов масс-спектров в БД.
9. Сопоставить масс-спектры полученных кандидатов с информацией по веществам (наименование, молекулярная масса (mw), формула и другие атрибуты).
10. Сформировать результат (список масс-спектров, подобных искомому) и снабдить необходимыми атрибутами.
11. Возвратить успешный или неуспешный код выхода (успешный при обнаружении подобных спектров, иначе – неуспешный).
12. Вывести список результатов со всеми атрибутами в виде таблицы.

Пункты 1, 2, 4, 5, 11, 12 выполняются в UniChrom. Специфичными для определённого типа БД являются пункты 3, 6-10, реализуемые с помощью различных функций в модулях поиска mssearch1 и mssearch2.

Модуль `mssearch1` разработан с использованием SQL (от англ. structured query language – «язык структурированных запросов») [3]. Данный модуль обращается к серверу БД, в которой реализована структура, содержащая следующие таблицы. `COMP` – информация о компоненте (веществе); `MSLIB` – информация о спектре, содержит ссылку на вещество `COMP_ID` в таблице `COMP`; `SPECTRA` – пики всех спектров, содержит ссылку `SP_ID` на спектр в таблице `MSLIB`; `SPECTRA_TMP` – временная таблица, содержащая пики неизвестного спектра; `SYNON` – все наименования веществ, содержит ссылку `COMP_ID` на таблицу `MSLIB`. Предварительный поиск и полный поиск объединены в одну хранимую процедуру. В процессе предварительного поиска выполняется выбор первых 5 наиболее интенсивных пиков неизвестного спектра, определяется область пересечения всей таблицы библиотечных пиков (`inner join`) с отобранными пиками при совпадении  $m/z$  и отличии значений интенсивности не более чем на 10 отн. ед. Результат группируется по идентификатору спектра, что позволяет рассчитать агрегатные значения для дальнейшего определения коэффициентов подобия, а именно, прямой коэффициента подобия по S. Stein [4] (`St`), косинуса угла спектральной контрастности (`SC`), коэффициент корреляции спектров Пирсона (`Pearson`). Далее, результат выводится в `UniChrom`.

Модуль `mssearch2` реализован с применением БД NIST. Для разработки использовался NIST SDK. Контекст поиска представляет собой объект, инкапсулирующий в себе структуры для управления поиском и структуры для получения результата. Основная функция, управляющая контекстом, называется `nistms_search`. В контексте поиска передаются пути к БД, типы используемых БД (`mainlib`, `replib`) и выбираются активные библиотеки. После выполнения поиска составляется список кандидатов с масс-спектрами, подобными исследуемому веществу. Коэффициенты подобия по S. Stein [4], возвращаются в массивы `sim_num` (прямой коэффициент подобия), `rev_sim_num` (обратный коэффициент подобия), `hit_prob` (вероятность совпадения). По списку кандидатов производится вычитывание масс-спектров и передается в `UniChrom`.

## ТЕСТИРОВАНИЕ

Тестирование модулей поиска производилось в 64-битных системах Windows 7/10 и Linux Debian GNU/Linux 10.9 с использованием 64-битной версии `UniChrom 5.1`, результат проиллюстрирован на рис. 1 и 2. На рис. 1 представлена хроматограмма полного ионного тока. Маркер установлен внутри пика выбранного вещества (указано стрелкой), выходящего из хроматографа при  $t = 4,75$  мин, для которого анализируется масс-спектр. На рис. 2 показаны слева масс-спектры и справа таблица кандидатов с атрибутами (название, формула,  $m_w$ , `St`, `SC`, `Pearson`). Масс-спектр

неизвестного вещества размещен сверху (область 1), а выбранного кандидата из таблицы – снизу (область 2).

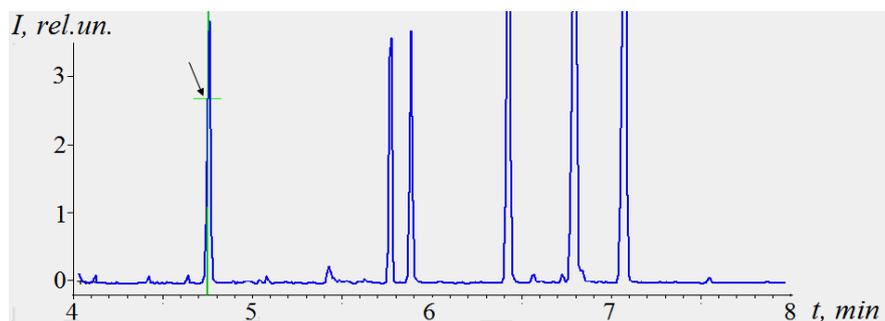


Рис. 1. Хроматограмма полного ионного тока для смеси вещества

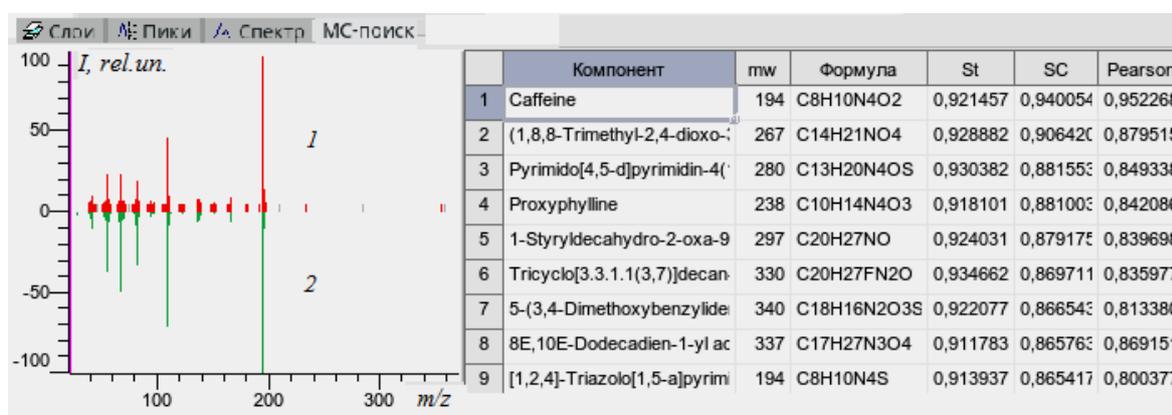


Рис. 2. Масс-спектры (слева) и таблица кандидатов с атрибутами (справа)

Таким образом, разработан универсальный кросс-платформенный программный интерфейс к БД для идентификации масс-спектров. Модули поиска могут быть использованы с различными программными продуктами и БД, в том числе с Web-интерфейсом.

#### Библиографические ссылки

1. Лебедев А. Т. Масс-спектрометрия в органической химии. М.: БИНОМ. Лабор. знаний, 2003, 493 с.
2. Беризовская Е. И. Методы обработки масс-спектрометрических данных при идентификации пептидов и белков // Вестн. Моск. Ун-та. Сер. 2. Химия. 2015. Т. 56. № 5. С. 266-278.
3. Грофф Д. Р., Вайнберг П. Н., Оппель Э. Дж. SQL: полное руководство, 3-е изд. SQL: The Complete Reference, Third Edition. - М.: «Вильямс», 2014. 960 с.
4. Stephen E. Estimating probabilities of correct identification from results of mass spectral library searches. // J. Am. Soc. Mass. Spectrom. 1994, V. 5, P. 316-323.