

ОТЗЫВ

на автореферат диссертационной работы **ХРИСАНФОВА Михаила Дмитриевича** на тему «Поиск ошибок в хроматографических базах данных и моделирование свойств идентифицируемых молекул методами машинного обучения», представленной на соискание ученой степени кандидата химических наук по специальности 1.4.2 – Аналитическая химия

Совершенствование хроматографической, масс-спектрометрической и, следовательно, хромато-масс-спектрометрической идентификации, бесспорно, относится к актуальным задачам современной аналитической химии. В диссертационной работе М.Д. Хрисанфова рассмотрены несколько аспектов такого совершенствования, а именно, во-первых, выявление потенциальных ошибок в существующих базах газохроматографических индексов удерживания (на примере базы NIST 17). Кроме того, затронуты вопросы моделирования свойств молекул (см. комментарии ниже) с использованием весьма популярных в последнее время методов машинного обучения. Таким образом, *актуальность* работы не вызывает сомнений, что оправдывает ее *цели и задачи*, а именно разработку оригинального подхода к поиску ошибок в базах данных, создание алгоритма округления нецелочисленных значений m/z , а также совершенствование подходов к предсказанию масс-спектров электронной ионизации. Пункты **научной новизны и практической значимости** полностью согласованы с целями и задачами, равно как и с *выносимыми на защиту положениями* и **выводами**. Диссертационная работа соответствует паспорту научной специальности 1.4.2 – Аналитическая химия – по перечисленным в автореферате пунктам. Апробацию работы следует считать более чем достаточной.

Судя по содержанию автореферата, особенностью работы оказывается преимущественно «хеометрический» уровень рассмотрения результатов (без подробного обсуждения деталей для конкретных соединений). Иные аспекты, по-видимому, рассмотрены в самой диссертации (глава 3). Поэтому, учитывая ограниченный объем отзыва, целесообразно сразу перейти к дискуссионным моментам работы. Сразу же следует подчеркнуть, что их не стоит классифицировать как «ошибки». В журнальных публикациях большая их часть подлежала бы уточнениям без каких-либо «карательных» мер в отношении автора.

Прежде всего, следовало бы дать определение такому термину как «молекулярные отпечатки пальцев». Из фразы на стр. 4 следует, что он не эквивалентен масс-спектрам. Из текста на стр. 17 трудно составить адекватное представление о возможностях их предсказания. То же относится к «кросс-платформенной совместимости», упомянутой на стр. 23. Не совсем понятен смысл фразы «Подтверждено, что предсказательные модели способны правильно предсказывать даже ошибочные записи в химических наборах данных благодаря способности к обобщению» (стр. 21).

Термин «ошибки» (начиная с названия работы) применительно к хроматографическим базам данных выглядит не совсем корректно и требует уточнения. Действительно, к ошибкам можно отнести экспериментальные погрешности определения индексов удерживания (RI), ошибки при вводе их значений в компьютер(ы) при формировании баз данных, и даже ошибки в химических названиях соединений. Кроме того, база NIST 17 содержит значения RI двух типов; во-первых, это единичные значения RI, определенные в единичных лабораториях, но, во-вторых, заметная часть данных представляет собой усредненные значения RI. Совершенно очевидно, что если индексы удерживания из разных источников незначительно отличаются друг от друга, то вероятность считать их ошибочными несоизмеримо меньше, чем в первом случае. Было бы весьма полезно, если бы автору удалось сравнить количество ошибок в каждом из этих случаев.

О термине «свойства молекул» (в том числе, в названии работы). К свойствам молекул можно отнести их молекулярную массу, электронную поляризуемость (расчетную), дипольный момент (в вакууме, тоже расчетная величина), средние валентные углы между связями и, пожалуй, это почти все. Индексы удерживания, времена удерживания, равно как температуры кипения, плотности и проч., не являются свойствами молекул, так как это характеристики веществ.

Автор отмечает, что в базе NIST 17 найдено 2093 потенциально ошибочных записи, но, к сожалению, не приводит ни одного конкретного примера. Возможно, что они есть в тексте диссертации. В связи с этим возникает еще один любопытный вопрос: сообщал ли автор в NIST об обнаруженных подозрительных значениях индексов удерживания?

При обсуждении округления значений m/z до эквивалентных им целочисленных величин в тексте автореферата рассмотрен пример с ионами с m/z 83.5362 и 82.5284 в масс-спектре карбазола (стр. 16), причем для этих ионов указаны составы $[^{13}\text{C}_{12}\text{H}_9\text{N}]^{++}$ и $[^{13}\text{C}_{12}\text{H}_7\text{N}]^{++}$. С этим категорически нельзя со-

гласиться. Во-первых, формулы двухзарядных ионов $[\text{C}_{12}\text{H}_9\text{N}]^{++}$ (m/z 83.5) и $[\text{C}_{12}\text{H}_7\text{N}]^{++}$ (m/z 82.5) указаны неверно, поскольку они не имеют к изотопу ^{13}C никакого отношения. Во-вторых (и это главное), величины m/z 83.5362 и 82.5284 в масс-спектре карбазола *округлять до целых значений не нужно*, так как они характеризуют двухзарядные ионы с регистрируемыми массовыми числами $83.5 = 167/2$ и $82.5 = 165/2$. В связи с этим, автору, прежде чем округлять значения m/z , можно порекомендовать решить «вспомогательную» задачу: отличить подлежащие округлению значения m/z от тех, для которых такое округление не является необходимым.

В тексте несколько раз упоминается выявление ошибок в базах индексов удерживания (газовая хроматография) и времен удерживания (ВЭЖХ). Следует заметить, что эти задачи совершенно не эквивалентны одна другой, поскольку в условиях обращенно-фазовой ВЭЖХ на времена удерживания влияет значительно большее количество различных параметров, чем в ГХ. Например, если в массив времен удерживания бензойной кислоты при рН элюента менее четырех попадет одно значение времени удерживания при рН > 6.0, то оно неизбежно будет классифицировано как ошибочное, хотя таким не является.

В тексте встречаются неудачные выражения: «масс-спектрометрическая характеристика (стр. 3), «собрать данные с прибора» (там же), «некоторые алгоритмы ... могут ... оставаться неизвестными широкой публике» (там же). «Несмотря на то, что ошибки линейно возрастают, ..., угол наклона небольшой» (стр. 9), часто выбирают один или другой подход исходя из личных предпочтений (стр. 13, чьих предпочтений?), «большая часть ошибок лежит слева, остальные записи – справа» (стр. 14), «что обусловлено явлением дефекта массы» (стр. 15), «лучше медианы предсказаний нескольких» (стр. 21). Выражение «требуется большого опыта» (стр. 17) всегда нуждается в уточнении «чей именно опыт здесь нужен?». Однако количество подобных погрешностей невелико и, как отмечено выше, их исправление относится к рутинной редакторской работе. Они не влияют на общее положительное впечатление о диссертационной работе.

В результате рассмотрения автореферата можно утверждать, что по критериям актуальности, научной новизны, уровню экспериментальных решений, а также научной и практической значимости полученных результатов диссертационная работа М.Д. Хрисанфова вполне соответствует требованиям, установленным Московским государственным университетом имени

М.В. Ломоносова к работам подобного рода. Содержание диссертации можно полагать соответствующим паспорту специальности 1.4.2 – «Аналитическая химия», а также критериям, определенным п.п. 2.1–2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В.Ломоносова. Оформление работы соответствует приложениям №№ 5, 6 Положения о диссертационном совете Московского государственного университета имени М.В.Ломоносова.

Следовательно, на основании автореферата можно заключить, что соискатель – **Михаил Дмитриевич Хрисанфов** – заслуживает присуждения **ученой степени кандидата химических наук по специальности 1.4.2 – Аналитическая химия.**

ЗЕНКЕВИЧ Игорь Георгиевич

докт. хим. наук, профессор

Профессор Института химии Федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет»

Адрес: Университетский проспект 26, С-Петербург 198504

Специальность докторской диссертации: 02.00.03 – Органическая химия

29
25 30.06.2026

ОПЕЧАТКА