

УДК 543.544.3:004.85

ОСОБЕННОСТИ ЛИНЕЙНОЙ КОРРЕЛЯЦИИ ГАЗОХРОМАТОГРАФИЧЕСКИХ ИНДЕКСОВ УДЕРЖИВАНИЯ СОЕДИНЕНИЙ РАЗЛИЧНЫХ ТАКСОНОМИЧЕСКИХ ГРУПП

© 2023 г. И. Г. Зенкевич*

Санкт-Петербургский государственный университет, Институт химии
Университетский просп., 26, Санкт-Петербург, 198504 Россия

*e-mail: izenkevich@yandex.ru

Поступила в редакцию 02.09.2022 г.

После доработки 10.09.2022 г.

Принята к публикации 10.09.2022 г.

Несмотря на постоянное совершенствование сложных компьютерных алгоритмов вычисления газохроматографических индексов удерживания, сохраняют свое значение простейшие способы их оценки на основании линейных корреляционных соотношений между индексами структурных аналогов разных таксономических групп, $RI^1 \approx aRI^2 + b$. Показано, что условием корректности подобных корреляций в простейших группах (замещенные метаны) являются симбатные вариации первых конечных разностей индексов удерживания, $\Delta_{RI}^1 = RI_{n+1} - RI_n$ (эквивалентны первым производным индексов удерживания по варьируемому в группе структурному параметру). Монотонные вариации Δ_{RI}^1 в одной из групп при наличии экстремумов в другой — однозначный признак отсутствия линейной корреляции индексов удерживания. Если в одной из групп значения Δ_{RI}^1 возрастают, а в другой — убывают, то порядок ранжирования соединений в какой-либо одной из них должен быть заменен обратным. Показано, что простейшее соотношение $RI^1 \approx aRI^2 + b$ также применимо в более сложных таксономических группах (замещенные этаны, бензолы и нафталины) и позволяет не только оценивать значения RI еще не охарактеризованных в настоящее время соединений, но и уточнять известные справочные данные.

Ключевые слова: газохроматографические индексы удерживания, таксономические группы структурных аналогов, линейная корреляция, контроль корректности индексов, оценка неизвестных значений.

DOI: 10.31857/S0044450223040175, EDN: LADISF

Возможности хромато-масс-спектрометрической идентификации следов органических соединений в сложных смесях могут быть эффективно реализованы только при совместном использовании масс-спектров с ионизацией электронами и газохроматографических индексов удерживания (RI). Однако если судить по содержанию современных баз данных (в том числе [1]), число соединений, охарактеризованных масс-спектрами, существенно уступает числу соединений, для которых известны значения RI на стандартных неподвижных фазах. Например, база данных NIST [1] (версия 2020 г.) содержит 350643 масс-спектров с ионизацией электронами 306869 соединений, но значения RI имеются только для 139498 соединений. Из них лишь для части (111768) объектов одновременно известны как масс-спектры, так и индексы удерживания.

Лимитирующей стадией формирования баз справочных значений RI является доступность объектов. Наибольших затрат времени требует, при необходимости, их целенаправленный синтез. По этой причине быстрое увеличение объема массивов справочных данных маловероятно, что придает особую актуальность совершенствованию способов расчета индексов удерживания на основании структур молекул. Известно огромное число методов расчета RI , рассмотрение которых является самостоятельной задачей. В последнее время “популярны” алгоритмы, предполагающие использование различных топологических дескрипторов (QSRR) [2–4], но все большее распространение приобретают самообучающиеся алгоритмы на принципах нейронных сетей [5–8] и более сложные (например, Gene Expression Programming [9]). Вместе с тем нельзя утверждать,

что исчерпаны возможности известных и широко применяемых аддитивных схем [10, 11].

Преимущества современных компьютерных методов расчета RI [5–9] проявляются при обработке максимально больших выборок данных, что не всегда рационально. Дело в том, что по мере накопления новых данных известные справочные значения RI могут быть уточнены или даже исправлены (в том числе это относится к базе [1]), что ведет к необходимости обработки массива информации для всех соединений заново. Кроме этого, в подобных алгоритмах точность оценок индексов удерживания фактически усреднена для всех рассматриваемых объектов. Это затрудняет выявление менее надежных данных, что часто приходится проводить вручную. Кроме того, в практике хроматографического анализа актуальна необходимость решения более “локальных” задач оценки RI единичных соединений более простыми способами. При этом характеристики точности таких оценок должны относиться к конкретным результатам, а не представлять собой средние значения для метода расчетов в целом. В таких случаях эффективны алгоритмы, применимые именно для малых таксономических групп, в том числе:

1. На основании корреляции RI с положением характеризуемого соединения в соответствующей группе гомологов, конгенеров или структурных аналогов, $RI = f(n_x)$;

2. В результате корреляции значений RI соединений выбранной таксономической группы с их физико-химическими характеристиками, $RI = f(A_x)$;

3. В результате корреляции RI соединений некоторой таксономической группы (1) с известными значениями RI соединений другой группы (2) (структурных аналогов), $RI^1 = f(RI^2)$.

Алгоритмы первой группы представляют собой расширенный вариант аддитивных схем, так что их возможности сопоставимы двух с другом. Наибольшую точность оценки газохроматографических индексов удерживания обеспечивают методы второй группы, особенно предполагающие использование нормальных температур кипения гомологов или конгенеров [12, 13]. Причина этого в том, что число соединений, для которых с высокой точностью определены значения $T_{кип}$, существенно превышает число соединений, охарактеризованных значениями RI. Однако некоторая сложность таких корреляций состоит в том, что используемые для вычислений соотношения нелинейны (линейно-логарифмические) [12, 13], что несколько затрудняет вычисления:

$$\lg RI = a \lg T_{кип} + bA + c, \quad (1)$$

где A – параметр, фиксирующий положение конкретного соединения в соответствующей таксономической группе (например, число атомов углерода в молекуле, число циклов, формальная не-предельность, и т.п.); a , b , c – коэффициенты, вычисляемые методом наименьших квадратов.

Вычисление индексов удерживания на основании корреляции их значений для структурных аналогов различных групп (третья группа; в общем случае такие зависимости нелинейны) предполагает использование аналогичных линейно-логарифмических соотношений [14]:

$$\lg RI^1 = a \lg RI^2 + bA + c, \quad (2)$$

при этом смысл параметров A , a , b , c – такой же, как и формуле (1).

Говоря о возможностях корреляции газохроматографических индексов удерживания структурных аналогов различных таксономических групп, нельзя не отметить практически важный частный случай. Дело в том, что для некоторых сочетаний структурных аналогов различных классов зависимость $RI^1 = f(RI^2)$ с приемлемой точностью может быть аппроксимирована простейшим уравнением линейной регрессии:

$$RI^1 \approx aRI^2 + b. \quad (3)$$

В подобных случаях получение оценок RI существенно упрощается, так как не требует применения нестандартных методов вычислений. Однако до настоящего времени не сформулированы какие-либо априорные правила выявления групп конгенеров, для которых этот вариант обеспечивает приемлемую точность результатов.

Для иллюстрации в качестве примера на рис. 1а приведена зависимость индексов удерживания бромзамещенных метанов $CH_{4-x}Br_x$ с числом атомов брома от одного до четырех ($1 \leq x \leq 4$) от значений RI структурно аналогичных хлорзамещенных метанов $CH_{4-x}Cl_x$ ($1 \leq x \leq 4$). Казалось бы, не следует ожидать больших различий свойств таких галогенопроизводных; они характеризуются близкими значениями диэлектрических проницаемостей и дипольных моментов, а атомные радиусы брома и хлора близки (1.00 и 1.15 ангстрем). Однако зависимость $RI(CH_{4-x}Br_x) = f[RI(CH_{4-x}Cl_x)]$ нелинейна, так что вычисления в таких условиях возможны только с использованием уравнения (2). При этом аналогичная зависимость индексов удерживания йодметанов и бромметанов (рис. 1б) близка к линейной (коэффициент корреляции $R = 0.9989$), хотя различия атомных радиусов йода и брома больше (1.40 и 1.15 ангстрем). Таким образом, уже из этого примера следует, что линейность или нелинейность корреляций $RI^1 = f(RI^2)$ определяется не физико-химическими характеристиками сравни-

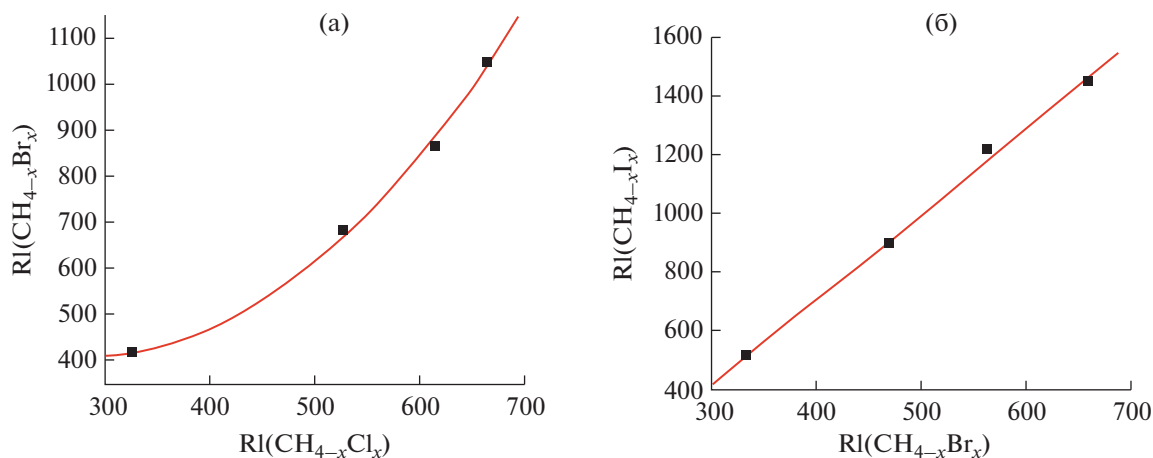


Рис. 1. Иллюстрация типичной нелинейной и практически линейной зависимостей индексов удерживания соединений разных таксономических групп: (а) – бромметаны и хлорметаны, $1 \leq x \leq 4$; (б) – йодметаны и бромметаны, $1 \leq x \leq 4$. При использовании уравнения (2) нелинейная зависимость (а) может быть приведена к линейному виду. Параметры линейной регрессии (б): $a = 1.49 \pm 0.05$, $b = -101 \pm 39$, $R = 0.9989$, $S_0 = 23$.

ваемых структурных аналогов, а какими-то иными причинами.

Комментируя рис. 1б, следует заметить, что значение RI тетраiodметана (1862) непропорционально завышено по сравнению с другими йодпроизводными: 516 ± 7 (CH_3I), 899 ± 9 (CH_2I_2) и 1209 ± 24 (CHI_3). В базу [1] оно включено на основании двух публикаций [15, 16], но, скорее всего, ошибочно. Дело в том, что CI_4 плавится при $168\text{--}171^\circ\text{C}$ с разложением, следовательно, его газохроматографический анализ невозможен из-за деградации в испарителе хроматографа. Поэтому значение RI , принадлежащее какому-то другому соединению (продукту разложения CI_4), было заменено экстраполированной теоретически ожидаемой оценкой 1446, вычисленной по линейно-логарифмическому соотношению (2) на основании данных для трех остальных йодметанов. Как отмечено в работе [17], проверка надежности исходных данных – необходимая стадия корреляций “свойство–свойство”.

Цель настоящей работы состоит в выявлении критериев применимости простейших линейных соотношений (3) для оценки газохроматографических индексов удерживания конгенов различных таксономических групп на основании значений индексов для конгенов других групп. Решаемые при этом задачи включают оценку неизвестных значений RI , проверку их правильности и, в отдельных случаях, идентификацию аналогов.

ВЫБОР ИСХОДНЫХ ДАННЫХ И ИХ ОБРАБОТКА (ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ)

Источником газохроматографических индексов удерживания рассматриваемых соединений на стандартных неполярных полидиметилсилоксановых неподвижных фазах служила база NIST [1]. Значения RI этоксисилоанолов $(C_2H_5O)_{4-x}Si(OH)_x$ ($1 \leq x \leq 3$) заимствовали из работы [18] (там же указаны условия газохроматографического анализа).

Для вычислений методом наименьших квадратов использовали программное обеспечение Excel (Microsoft 2007) и Origin (версии 8.1 и 4.1). В качестве критериев корректности линейных зависимостей $RI^1 \approx aRI^2 + b$ выбирали а) значения коэффициентов корреляции $R > 0.99$ и б) результаты сравнения минимальных разностей RI гомологов характеризуемой группы со значениями генеральных дисперсий S_0 . Величины S_0 в первом приближении характеризуют надежность значений, вычисляемых по уравнению линейной регрессии (для более точных оценок используют так называемые коридоры ошибок [19]). Результаты аппроксимации RI по соотношениям вида (3) значимы, если минимальные разности индексов удерживания соседних структурных аналогов превышают удвоенные значения S_0 :

$$\min |RI_i - RI_{i\pm 1}| > 2S_0. \quad (4)$$

Невыполнение неравенства (4) соответствует ненадежным оценкам RI , которые нецелесообразно использовать для решения аналитических задач.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Таксономические группы структурных аналогов.

Особенности линейной корреляции газохроматографических индексов удерживания структурных аналогов выявляли на примерах соединений нескольких таксономических типов, в том числе следующих:

1. Замещенные метаны: $\text{CH}_4 \rightarrow \text{CH}_3\text{X} \rightarrow \text{CH}_2\text{X}_2 \rightarrow \text{CHX}_3 \rightarrow \text{CX}_4$. Поскольку в таких группах первый член (CH_4 , $\text{RI} = 100$) одинаков, то его исключали из рассмотрения. В результате в них остается всего по четыре объекта, что снижает практическую ценность таких корреляций RI , но упрощает выявление их закономерностей и особенностей.

2. Замещенные этаны: $\text{C}_2\text{H}_6 \rightarrow \text{C}_2\text{H}_5\text{X} \rightarrow \text{CH}_3\text{CHX}_2 \rightarrow \text{XCH}_2\text{CH}_2\text{X} \rightarrow \text{XCH}_2\text{CHX}_2 \rightarrow \text{XCH}_2\text{CX}_3 \rightarrow \text{X}_2\text{CHCHX}_2 \rightarrow \text{X}_2\text{CHCX}_3 \rightarrow \text{X}_3\text{C-CX}_3$. После исключения незамещенного этана ($\text{RI} = 200$) в таких группах остается по восемь соединений, часть из которых (в зависимости от природы X) не охарактеризована значениями RI , что делает решение подобных задач более актуальным, чем соединений первого типа.

3. Замещенные бензолы: $\text{C}_6\text{H}_6 \rightarrow \text{C}_6\text{H}_5\text{X} \rightarrow \text{C}_6\text{H}_4\text{X}_2$ (три изомера) $\rightarrow \text{C}_6\text{H}_3\text{X}_3$ (три изомера) $\rightarrow \text{C}_6\text{H}_2\text{X}_4$ (три изомера) $\rightarrow \text{C}_6\text{HX}_5 \rightarrow \text{C}_6\text{X}_6$. Без незамещенного бензола ($\text{RI} = 654 \pm 7$ [1]) число объектов в таких группах равно 13. Для многих из них получение оценок RI является актуальной задачей из-за отсутствия экспериментальных данных.

4. Замещенные нафталины: $\text{C}_{10}\text{H}_8 \rightarrow \text{C}_{10}\text{H}_7\text{X}$ (два изомера) $\rightarrow \text{C}_{10}\text{H}_6\text{X}_2$ (10 изомеров). Эти группы чаще всего приходится ограничивать дизамещенными соединениями, так как для многих изомеров с тремя даже простейшими заместителями (метильные группы) надежные значения RI неизвестны [1].

Особенности корреляции газохроматографических индексов удерживания в группе замещенных метанов и родственных соединений¹. Проверка взаимной корреляции значений RI замещенных метанов при разных X ($\text{X} = \text{Cl}, \text{Br}, \text{I}, \text{OCH}_3, \text{OC}_2\text{H}_5, \text{CH}_3, \text{C}_6\text{H}_5$ и др.) позволяет выявить примеры как нелинейности зависимостей $\text{RI}^1 = f(\text{RI}^2)$ (рис. 1а), так и (реже) их линейности (как на рис. 1б). Для характеристики рассматриваемого подхода необходимо выяснить причины подобных различий, которые, как отмечено выше, не связаны с химической природой фрагментов X .

Необходимым, но не достаточным условием взаимной линейной корреляции индексов удерживания структурных аналогов являются их симбатные изменения в обеих сравниваемых груп-

пах. Если в одной из групп значения RI монотонно возрастают или убывают, а во второй имеется экстремум, то применение соотношений вида (3) не может обеспечить необходимую точность оценок. Экстремумы индексов удерживания (максимумы) наблюдаются, например, в группах метоксиметанов $\text{CH}_3\text{O} \rightarrow \dots \rightarrow \text{C}(\text{OCH}_3)_4$, фторметанов $\text{CH}_3\text{F} \rightarrow \dots \rightarrow \text{CF}_4$, метилсиланов $\text{CH}_3\text{SiH}_3 \rightarrow \dots \rightarrow \text{Si}(\text{CH}_3)_4$ и других.

Для получения более определенных заключений об особенностях вариации значений RI в разных группах необходимо сопоставить не просто значения индексов удерживания, а их первых конечных разностей, $\Delta_{\text{RI}}^1 = \text{RI}_{i+1} - \text{RI}_i$ (табл. 1). Так, для рассмотренного выше примера бром- и хлорметанов (рис. 1) индексы удерживания монотонно возрастают в каждой из групп. Однако разности Δ_{RI}^1 в ряду $\text{CH}_3\text{Cl} \rightarrow \text{CCl}_4$ монотонно убывают от 201 до 48, тогда как в ряду $\text{CH}_3\text{Br} \rightarrow \text{CBr}_4$ последнее значение Δ_{RI}^1 (186) несколько больше предыдущего (181), т.е. набор значений Δ_{RI}^1 характеризуется незначительным экстремумом (минимумом). В математическом смысле первые конечные разности Δ_{RI}^1 эквивалентны первой производной функции $\text{RI} = f(n_{\text{X}})$ по числу варьируемых фрагментов структуры, однако выявление причин вариаций ее знака (появления экстремумов) даже на феноменологическом уровне представляет собой сложную задачу. Наиболее вероятным представляются стерические взаимодействия фрагментов X в молекулах, аналогичные эффектам, выявленным в работе [20] на примере метилбензолов.

Таким образом, если в одной из сопоставляемых таксономических групп первые разности индексов удерживания Δ_{RI}^1 изменяются монотонно, а во второй характеризуются экстремумами, то корреляция таких значений RI не может обеспечить приемлемой точности их оценок. К таким примерам относятся сочетания набора данных $\text{CH}_3\text{Cl} \rightarrow \text{CCl}_4$ (значения Δ_{RI}^1 монотонно убывают) с наборами данных $\text{CH}_3\text{OCH}_3 \rightarrow \text{C}(\text{OCH}_3)_4$, $\text{CCl}_3\text{Br} \rightarrow \text{CBr}_4$, $(\text{CH}_3)_3\text{SiCl} \rightarrow \text{SiCl}_4$ (во всех случаях существуют максимумы значений Δ_{RI}^1) и другими.

С другой стороны, если оба сравниваемых набора значений RI характеризуются экстремумами Δ_{RI}^1 , то можно полагать, что их корреляция окажется вполне приемлемой по критерию точности оценок индексов удерживания. В табл. 2 сопоставлены индексы удерживания серии метоксиметанов $\text{CH}_3\text{OCH}_3 \rightarrow \text{C}(\text{OCH}_3)_4$ (данные для которых недостаточно надежны [1] и заслуживают уточнения), рассматриваемые как функция индексов удерживания хлорбромметанов

¹ Топологическими аналогами замещенных метанов являются, например, замещенные силаны.

Таблица 1. Индексы удерживания бром- и хлорметанов [1] и их разности

Бромметан	RI	Δ_{RI}^1	Хлорметан	RI	Δ_{RI}^1
CH ₃ Br	415 ± 5	268 181* 186	CH ₃ Cl	327 ± 3	201
CH ₂ Br ₂	683 ± 6		CH ₂ Cl ₂	528 ± 10	87
CHBr ₃	864 ± 11		CHCl ₃	615 ± 6	48
CBr ₄	1050 ± 5		CCl ₄	663 ± 16	

* Минимум значений Δ_{RI}^1 .**Таблица 2.** Индексы удерживания хлорбром- и метоксиметанов [1] и их разности

Метоксиметан	RI	Δ_{RI}^1	Аппроксимированное значение RI	Хлорбромметан	RI	Δ_{RI}^1
CH ₃ OCH ₃	427 ± 19	81 194* 128	408 531 689 824	CCl ₃ Br	757 ± 5	87
CH ₂ (OCH ₃) ₂	508 ± 2			CCl ₂ Br ₂	844 ± 1	111*
CH(OCH ₃) ₃	702			CClBr ₃	955	95
C(OCH ₃) ₄	830 ± 4			CBr ₄	1050 ± 5	

* Максимумы значений Δ_{RI}^1 .

CCl₃Br → CBr₄. В обеих группах регистрируются максимумы Δ_{RI}^1 , однако корреляция значений RI этих совокупностей структурных аналогов, иллюстрируемая рис. 2, вполне пригодна для получения практически полезных оценок индексов. Все значения Δ_{RI}^1 превышают удвоенную генеральную дисперсию $S_0 = 23$. Параметры уравнения линейной регрессии и расчетные значения RI указаны в табл. 2; среднее значение абсолютного отклонения расчетных и справочных величин составляет всего 15 ± 5 ед. индекса.

Значения Δ_{RI}^1 в двух сравниваемых таксономических группах должны либо одновременно возрастать, либо убывать. Поэтому еще одной причиной отсутствия линейной корреляции $RI^1 = f(RI^2)$ являются антибатные изменения Δ_{RI}^1 , что эквивалентно разным знакам первой производной dRI/dn_X . В качестве примера можно привести корреляцию индексов удерживания фторхлорметанов CF_{4-x}Cl_x и хлорметанов CH_{4-x}Cl_x ($1 \leq x \leq 4$), результаты которой иллюстрирует рис. 3а, б. Если обе группы структурных аналогов сформированы по увеличению числа атомов хлора в молекуле, т.е. CF₃Cl → ... → CCl₄ и CH₃Cl → ... → CCl₄, то за-

висимость $RI^1 = f(RI^2)$ нелинейна (рис. 3а). Причина этого в том, что в первой из этих серий значения Δ_{RI}^1 возрастают, а во второй — убывают. Чтобы преобразовать данные к линейной зависи-

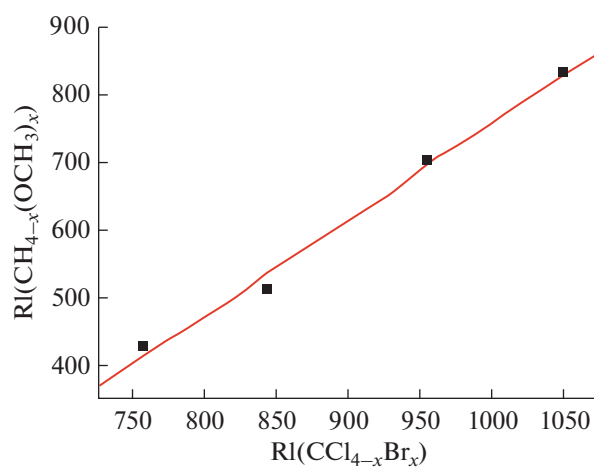


Рис. 2. Корреляция индексов удерживания метоксиметанов и хлорбромметанов ($1 \leq x \leq 4$). Параметры линейной регрессии: $a = 1.42 \pm 0.10$, $b = -667 \pm 95$, $R = 0.995$, $S_0 = 23$.

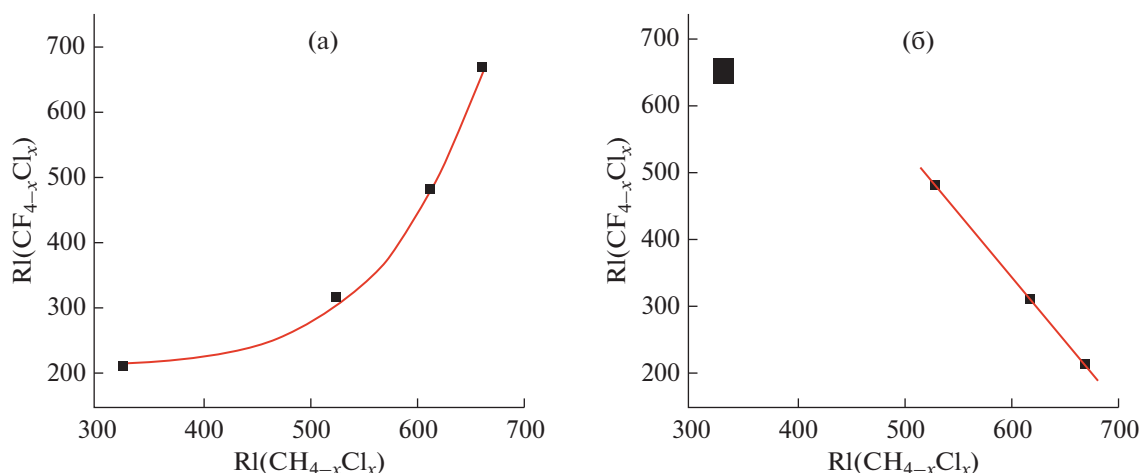


Рис. 3. Корреляция индексов удерживания фторхлорметанов и хлорметанов ($1 \leq x \leq 4$). (а): Нелинейная зависимость: обе группы структурных аналогов ранжированы по увеличению числа атомов хлора в молекуле: $CF_3Cl \rightarrow CCl_4$ и $CH_3Cl \rightarrow CCl_4$; (б): группы аналогов ранжированы по увеличению значений Δ_{RI}^1 : $CF_3Cl \rightarrow CCl_4$ и $CCl_4 \rightarrow CF_3Cl$ и $CH_3Cl \rightarrow CCl_4$. После исключения точки $CH_3Cl - CCl_4$ (слева сверху) параметры линейной регрессии для трех оставшихся пар: $a = -2.00 \pm 0.06$, $b = 1536 \pm 38$, $R = 0.9995$, $S_0 = 6$.

мости последовательность структурных аналогов в первой из этих групп (целевой) нужно изменить на противоположную, т.е. ранжировать по уменьшению значений Δ_{RI}^1 , $CCl_4 \rightarrow \dots \rightarrow CF_3Cl$. Однако в результате этого преобразования данных на графике появляется точка (показана на рис. 3б в увеличенном масштабе слева вверху), отвечающая паре соединений $CH_3Cl - CCl_4$, которые не содержат фтора и, следовательно, по химическим признакам не соответствуют остальным фторпроизводным. Поэтому вполне закономерно, что она выпадает из общей линейной регрессии; коэффициент корреляции для трех оставшихся пар соединений составляет $R = 0.9995$.

В качестве примера применения обсуждаемых особенностей корреляции значений RI в группе замещенных метанов и их аналогов (в том числе кремнийсодержащих) небезынтересно рассмотреть газохроматографическую идентификацию необычных продуктов частичного гидролиза тетраэтоксисилана $(C_2H_5O)_4Si$ ($RI\ 978 \pm 10$). В их числе обнаружены триэтоксисиланол $(C_2H_5O)_3SiOH$ ($RI\ 727 \pm 3$), диэтоксисиландиол $(C_2H_5O)_2Si(OH)_2$ ($RI\ 608 \pm 1$) и этоксисилантриол $(C_2H_5O)Si(OH)_3$ ($RI\ 537 \pm 2$) [18]. Если сравнивать значения RI таких этоксисиланолов, ранжированные “естественным образом” по увеличению числа этоксигрупп в молекулах (соответствует возрастанию значений Δ_{RI}^1), с данными для хлорметанов $CH_3Cl \rightarrow CCl_4$ (соответствует уменьшению Δ_{RI}^1), то зависимость $RI^1 = f(RI^2)$ оказывается нелинейной (рис. 4а). Для ее линеаризации одну

из этих последовательностей необходимо заменить обратной, например, рассматривать совокупность этоксисиланолов по увеличению числа OH -групп в молекуле, $(C_2H_5O)_4Si \rightarrow \dots \rightarrow (C_2H_5O)Si(OH)_3$. В результате получаем практически “идеальную” линейную регрессию (коэффициент корреляции R достигает 0.9995), график которой изображен на рис. 4б. Примечательно, что разметка осей рис. 4а и 4б, естественно, одинакова, и для того чтобы различить два рассматриваемых варианта, необходимы специальные комментарии.

Для дополнительного подтверждения того, что продукты частичного гидролиза тетраэтоксисилана образуют совокупность конгенов, можно рассмотреть еще одну корреляцию их значений RI с данными для какой-либо иной совокупности структурных аналогов, например йодалканов Cl_4 ($RI\ 1446$) $\rightarrow CHI_3$ (1209 ± 24) $\rightarrow CH_2I_2$ (899 ± 9) $\rightarrow CH_3I$ (516 ± 7). Бромалканы для этого использовать нельзя из-за присущего им небольшого максимума значений Δ_{RI}^1 (см. табл. 1). График линейной зависимости $RI^1 \approx aRI^2 + b$ приведен на рис. 5. После исключения точки, соответствующей паре соединений $CH_3I - (C_2H_5O)_4Si$, значение R для трех оставшихся пар структурных аналогов равно -0.998 при $S_0 = 9$.

Рассматриваемые корреляции $RI^1 \approx aRI^2 + b$ рекомендуется использовать не только для оценки, но и для проверки надежности справочных значений индексов удерживания наименее подробно охарактеризованных “экзотических” соединений. При этом в качестве реперных объектов

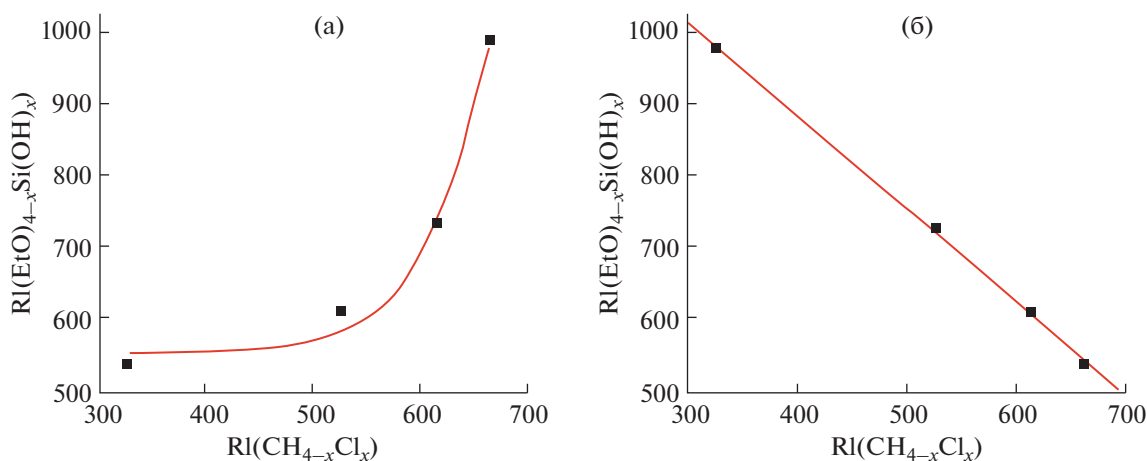


Рис. 4. Корреляция индексов удерживания этоксисилоанов $(C_2H_5O)_{4-x}Si(OH)_x$ и хлорметанов ($1 \leq x \leq 4$). (а): Нелинейная зависимость: значения RI этоксисилоанов ранжированы по увеличению молекулярных масс (или, иначе, числа этоксигрупп в молекуле), что соответствует возрастанию Δ_{RI}^1 ; (б): обе группы ранжированы по убыванию значений Δ_{RI}^1 , что приводит к обращению последовательности алкоксисилоанов: $(C_2H_5O)_4Si \rightarrow (C_2H_5O)Si(OH)_3$. Параметры регрессии (б): $a = 1.30 \pm 0.03$, $b = 1408 \pm 16$, $R = -0.9995$, $S_0 = 8$.

(с которыми проводят сравнение) целесообразно выбирать соединения с хорошо известными и воспроизводимыми значениями RI , например подробно охарактеризованные углеводороды. Следующий пример относится именно к такому типу: необходимо проверить известные значения RI метилсиланов $CH_3SiH_3 \rightarrow Si(CH_3)_4$. Данные для простейшего монотетилсилана в базе [1] отсутствуют, а следующие два гомолога охарактеризованы в единственной работе [21], поэтому их приходится дополнять расчетными величинами на основании температур кипения [12–14]. Значения RI метилсиланов можно сопоставлять с данными для изоструктурных алканов C_2-C_5 , что иллюстрирует табл. 3.

График зависимости $RI^1 = f(RI^2)$ в данном случае настолько близок к линейному, что его вполне можно не приводить и ограничиться только параметрами линейной регрессии: $a = 1.01 \pm 0.03$, $b = 39 \pm 9$, $R = 0.9992$, $S_0 = 4.5$. Таким образом, подтверждено, что известные значения RI метилсиланов достаточно надежны.

При соответствующем контроле вариаций Δ_{RI}^1 в сравниваемых группах структурных аналогов рассмотренный подход не имеет принципиальных ограничений и отличается исключительной простотой.

Корреляция газохроматографических индексов удерживания в других таксономических группах (большого объема) существенно расширяет возможности рассматриваемого подхода. С другой стороны, можно ожидать непредсказуемого усложнения зависимостей $RI^1 = f(RI^2)$, например,

за счет проявления трудно формализуемых стерических эффектов, заметных лишь для отдельных соединений в пределах таких групп. В качестве примера целесообразно рассмотреть группу замещенных этанов (9 конгенов), значения RI которых сопоставлены с данными для изоструктурных алканов (табл. 4). Из их графического представления (рис. 6) следует, что весь набор точек “распадается” на две подгруппы, соответствующие двум линейным участкам: (I) для соединений

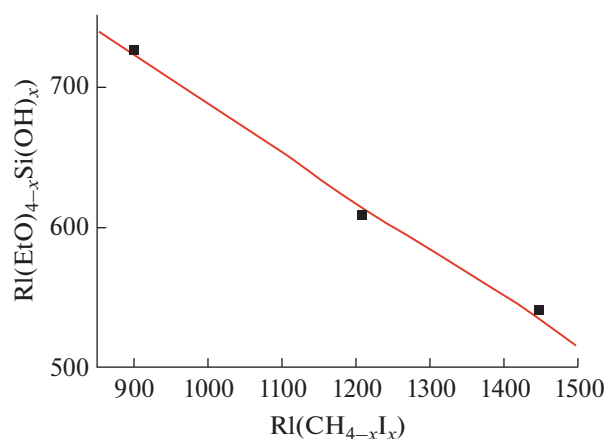


Рис. 5. Корреляция индексов удерживания этоксисилоанов $(C_2H_5O)_{4-x}Si(OH)_x$ и йодметанов ($2 \leq x \leq 4$). Обе группы структурных аналогов ранжированы по убыванию значений Δ_{RI}^1 . Параметры линейной регрессии: $a = -0.35 \pm 0.02$, $b = 1038 \pm 29$, $R = -0.998$, $S_0 = 9$.

Таблица 3. Индексы удерживания метилсиланов и алканов C₂–C₅ [1]

Метилсилан	RI	Δ_{RI}^1	Аппроксимированное значение RI	Алкан	RI	Δ_{RI}^1
CH ₃ SiH ₃	243 ± 8	94	240	C ₂ H ₆	200	100
(CH ₃) ₂ SiH ₂	337 ± 8	71	342	C ₃ H ₈	300	62
(CH ₃) ₃ SiH	408	48	402	<i>изо</i> -C ₄ H ₁₀	362 ± 7	51
(CH ₃) ₄ Si	456 ± 16		413	<i>нео</i> -C ₅ H ₁₂	413 ± 2	

Таблица 4. Индексы удерживания хлорэтанов C₂H_{6-x}Cl_x и алканов C₃–C₈ [1]

Номер соединения	Хлорэтан	RI	Аппроксимированное значение RI (номер подгруппы)	Алкан (структурный аналог)	RI
1	C ₂ H ₅ Cl	422 ± 8	436 (I)	C ₃ H ₈	300
2	ClCH ₂ -CH ₂ Cl	642 ± 7	624 (I)	C ₄ H ₁₀	400
3	CH ₃ -CHCl ₂	558 ± 19	553 (I)	Изобутан	362 ± 7
4	ClCH ₂ -CHCl ₂	768 ± 4	761 (I)	Изопентан	473 ± 4
5	CH ₃ -CCl ₃	641 ± 7	648 (I)	Неопентан	413 ± 2
6	ClCH ₂ -CCl ₃	869	882 (I), 876 (II)	2,2-Диметилбутан	537 ± 2
7	Cl ₂ CH-CHCl ₂	916	905 (II)	2,3-Диметилбутан	566 ± 3
8	C ₂ HCl ₅	969	974 (II)	2,2,3-Триметилбутан	635 ± 4
9	C ₂ Cl ₆	1064	1063 (II)	2,2,3,3-Тетраметилбутан	724 ± 9

№№ 1–6 (содержат от одного до четырех атомов хлора) и (II) для соединений №№ 6–9 (от четырех до шести атомов хлора). Параметры двух уравне-

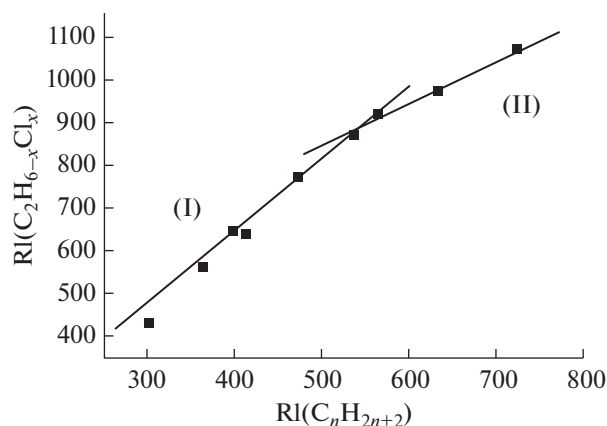


Рис. 6. Корреляция индексов удерживания хлорэтано- и алканов C₂H_{6-x}Cl_x (1 ≤ x ≤ 6) и алканов C₃–C₈. Параметры линейной регрессии участка (I): a = 1.88 ± 0.08, b = 128 ± 32, R = 0.997, S₀ = 14; участка (II): a = 1.00 ± 0.07, b = 339 ± 43, R = 0.995, S₀ = 10.

ний линейной регрессии, соответствующие каждому из этих участков, приведены в подписи к рисунку. Объяснение причин появления двух линейных участков в настоящий момент представляется затруднительным, и к этому факту следует относиться как к феноменологически выявленной закономерности. Аналогично два линейных участка зависимости RI¹ = f(RI²) наблюдаются при корреляции значений RI хлор- и бромэтано-

Средняя точность оценок индексов удерживания хлорэтано- и алканов для изоструктурных углеводородов составляет 9 ± 5 ед. индекса, что вполне достаточно для практических целей и не уступает возможностям гораздо более сложных методов. Поскольку атомы хлора могут находиться при двух атомах углерода, это усложняет использование первых конечных разностей индексов удерживания Δ_{RI}^1 (массив Δ_{RI}^1 становится двумерным).

Проверку возможностей рассматриваемого подхода в группах замещенных бензолов целесообразно начать с сопоставления индексов удерживания бром- и хлорбензолов. Справочные данные приведены в табл. 5; для некоторых бромбен-

Таблица 5. Индексы удерживания бромбензолов $C_6H_{6-x}Br_x$ и хлорбензолов $C_6H_{6-x}Cl_x$ [1]

Хлорбензол	RI	Бромбензол	RI	Расчетное значение RI
Монохлор-	839 ± 7	Монобром-	839 ± 7	
1,2-Дихлор-	1016 ± 11	1,2-Дибром-	1216 ± 5	
1,3-Дихлор-	989 ± 10	1,3-Дибром-	1190 ± 7	
1,4-Дихлор-	995 ± 10	1,4-Дибром-	1193	
1,2,3-Трихлор-	1199 ± 12	1,2,3-Трибром-	—*	1530
1,2,4-Трихлор-	1160 ± 13	1,2,4-Трибром-	—	1464**
1,3,5-Трихлор-	1117 ± 14	1,3,5-Трибром-	1390	
1,2,3,4-Тетрахлор-	1364 ± 11	1,2,3,4-Тетрабром-	—	1808
1,2,3,5-Тетрахлор-	1322 ± 11	1,2,3,5-Тетрабром-	—	1737
1,2,4,5-Тетрахлор-	1326 ± 18	1,2,4,5-Тетрабром-	1697	
Пентахлор-	1501 ± 5	Пентабром-	—	2039
Гексахлор-	1691 ± 5	Гексабром-	2379	

* Прочерк означает отсутствие справочных данных в базе [1]; ** оценка RI по аддитивной схеме [10] равна 1460.

золов значения RI неизвестны, что объясняет актуальность их оценки. Поскольку значение коэффициента корреляции R достигает 0.9989, график зависимости $RI^1 = f(RI^2)$ (рис. 7) подтверждает ее линейность.

Объективным ограничением такого способа оценки индексов удерживания является невозможность различить некоторые изомеры положения, характеризующиеся относительно небольшими различиями RI, например *орто*-, *мета*- и *пара*-дибромбензолы с индексами 1216, 1190 и 1193 соответственно. Их попарные разности меньше удвоенной величины $S_0 = 24$ (см. уравнение (4)). Примечательно, что для изомерных трибромбензолов таких совпадений уже нет: 1530 (1,2,3-изомер), 1464 (1,2,4-изомер) и 1390 (1,3,5-изомер).

Аналогичные оценки RI на основании их взаимных корреляций для соединений различных таксономических групп получены для Br- и Cl-бензолов $C_6H_{6-x}Hal_x$ ($1 \leq x \leq 6$) в результате их сравнения с данными для метилбензолов. Для индексов удерживания йодбензолов $C_6H_{6-x}I_x$ ($1 \leq x \leq 3$) эффективным оказалось их сравнение с данными для соответствующих хлорбензолов, метиловых эфиров бензолкарбоновых кислот $C_6H_{6-x}(CO_2CH_3)_x$ ($1 \leq x \leq 4$) — с хлорбензолами, хлорнафталинов $C_{10}H_{8-x}Cl_x$ ($1 \leq x \leq 4$) — с метилнафталинами) и т.д.

Таким образом, анализ возможностей и особенностей простейшей линейной корреляции газохроматографических индексов удерживания соединений разных таксономических групп

структурных аналогов, $RI^1 \approx aRI^2 + b$ позволил установить следующее:

— Если каждое соединение в соответствующей группе однозначно задано молекулярной формулой, т.е. не имеет изомеров (это возможно для простейших замещенных метанов, $CH_3X \rightarrow \dots \rightarrow CX_4$), то условием линейной корреляции индексов удерживания в таких группах являются симбатные вариации первых конечных разностей индексов удерживания, $\Delta_{RI}^1 = RI_{n+1} - RI_n$. Наличие экстремумов Δ_{RI}^1 в одной из групп при их монотонном изменении в другой — однозначный признак отсутствия линейной корреляции $RI^1 \approx aRI^2 + b$. Если в одной из групп значения Δ_{RI}^1

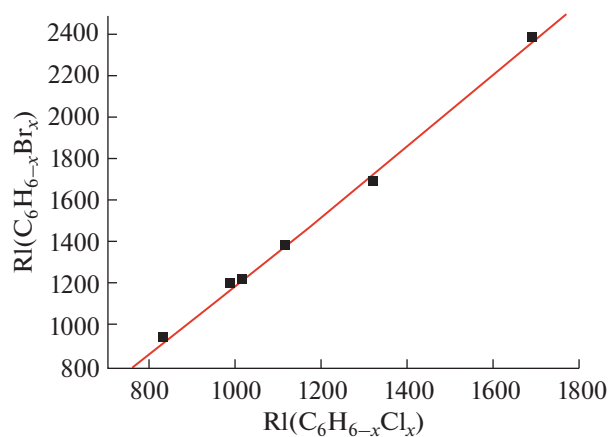


Рис. 7. Корреляция индексов удерживания бромбензолов $C_6H_{6-x}Br_x$ ($1 \leq x \leq 6$) и соответствующих хлорбензолов. Параметры линейной регрессии: $a = 1.68 \pm 0.03$, $b = -489 \pm 40$, $R = 0.9989$, $S_0 = 24$.

возрастают, а в другой — убывают, то в любой из групп порядок ранжирования соединений должен быть заменен обратным.

— Показано, что корреляции индексов удерживания возможны в таких группах структурных аналогов, как замещенные этаны, бензолы, нафталины и других. Во всех случаях использование простейшего соотношения $RI^1 \approx aRI^2 + b$ позволяет не только оценивать значения RI еще не охарактеризованных в настоящее время соединений, но и уточнять ненадежные справочные данные.

СПИСОК ЛИТЕРАТУРЫ

1. The NIST Mass Spectral Library (NIST/EPA/NIH EI MS Library, 2020 Release). Software/Data Version; NIST Standard Reference Database, Number 69, May 2020. National Institute of Standards and Technology, Gaithersburg, MD 20899: <http://webbook.nist.gov> (дата обращения: сентябрь 2022 г.).
2. *Pompe M., Novic M.* Prediction of gas-chromatographic retention indices using topological descriptors // *J. Chem. Inf. Comput. Sci.* 1999. V. 39. № 1. P. 59. <https://doi.org/10.1021/ci980036z>
3. *Farkas O., Heberger K., Zenkevich I.G.* Quantitative structure-retention relationships XIV. Prediction of GC retention indices for saturated O-, N-, and S-heterocyclic compounds // *Chemometr. Intel. Lab. Systems.* 2004. V. 72. P. 173.
4. *Heberger K.* Quantitative structure – (chromatographic) retention relationships (QSRR) // *J. Chromatogr. A.* 2007. V. 1158. P. 273. <https://doi.org/10.1016/chroma.2007.03.108>
5. *Matyushin D., Buryak A.* Gas chromatographic retention index prediction using multimodal machine learning // *IEEE Access.* 2020. V. 8. P. 223140. <https://doi.org/10.1109/ACCESS.2020.3045047>
6. *Matyushin D.D., Sholokhova A.Yu., Buryak A.K.* Deep learning based prediction of gas chromatographic retention indices for a wide variety of polar and mid-polar liquid stationary phases // *Int. J. Mol. Sci.* 2021. V. 22. 13 p. <https://doi.org/10.3390/ijms22179194>
7. *Idroes R., Noviandy T.R., Maulana A., Syhendra R., Sasmita N.R., Muslem M., Idroes G.M., Kemala P., Irvanizam I.* Application of genetic algorithm – multiple linear regression and artificial neural network determination for prediction of Kovats retention index // *Int. Rev. Model. Simulat.* 2021. V. 14. № 2. 137 p. <https://doi.org/10.15866/iremos.v14i2.20460>
8. *Qu C., Schneider B.I., Kearsley A.J., Keyrouz W., Allison T.C.* Predicting Kovats retention indices using graph neural networks // *J. Chromatogr. A.* 2021. V. 1646. Art. 462100. <https://doi.org/10.1016/j.chroma.2021.462100>
9. *Zhang X., Shi L., Ding L., Sun Z., Song L., Qu H., Sun T.* Study on quantitative structure – retention relationships (QSRR) for oxygen-containing organic compounds based on gene expression programming (GEP) // *J. Chromatogr. Sep. Tech.* 2015. V. 6. № 7. 7 p. <https://doi.org/10.4172/2157-7064.1000306>
10. *Stein S.E., Babushok V.I., Brown R.L., Linstrom P.J.* Estimation of Kovats retention indices using group contribution // *J. Chem. Inf. Model.* 2007. V. 47. P. 975. <https://doi.org/10.1021/ci600548y>
11. *Зенкевич И.Г., Елисеенков Е.В., Касаточкин А.Н.* Идентификация продуктов органических реакций при отсутствии аддитивности хроматографических индексов удерживания. Хлорпроизводные метил-трет-бутилкетона // *Журн. структ. химии.* 2013. Т. 54. № 3. С. 453. <https://doi.org/10.1134/S0022476613030050>
12. *Зенкевич И.Г.* Формирование таксономических групп органических соединений для расчета газохроматографических индексов удерживания по физико-химическим константам // *Журн. структ. химии.* 1994. Т. 35. № 6. С. 176.
13. *Zenkevich I.G.* Reciprocally unambiguous conformity between GC retention indices and boiling points within two- and multidimensional taxonomic groups of organic compounds // *J. High Resolut. Chromatogr. Chromatogr. Commun.* 1998. V. 21. № 10. P. 565.
14. *Зенкевич И.Г.* Принцип структурной аналогии при оценке газохроматографических индексов удерживания // *Журн. структ. химии.* 1996. Т. 37. № 4. С. 793.
15. *Waggott A., Davies I.W.* Identification of organic pollutants using linear temperature programmed retention indices (LTPRIs) – Part II // 1984. <http://dwi.defra.gov.uk/research/completed/research/reports/dwi0383.pdf> (дата обращения: август 2022 г.).
16. *Yasuhara A., Morita M., Fuwa K.* Temperature-programmed retention indices of 221 halogenated organic compounds with 1-bromoalkanes as references // *J. Chromatogr.* 1985. V. 328. P. 35. [https://doi.org/10.1016/S0021-9673\(01\)87374-X](https://doi.org/10.1016/S0021-9673(01)87374-X)
17. *Зенкевич И.Г.* Контроль, коррекция и восстановление значений физико-химических свойств органических соединений с использованием рекуррентных зависимостей // *Журн. физ. химии.* 2021. Т. 95. № 5. С. 700. <https://doi.org/10.1134/S0036024421040294>
18. *Зенкевич И.Г., Баранов Д.А.* Газохроматографическая идентификация необычных нестабильных продуктов частичного гидролиза тетраэтоксисилана // *Журн. аналит. химии.* 2023. Т. 78. № 1. С. .
19. *Линник Ю.В.* Метод наименьших квадратов и основы теории обработки наблюдений. М.: Физматгиз, 1958. 334 с.
20. *Зенкевич И.Г.* Выявление аномалий газохроматографических индексов удерживания гомологов на основании их гомологических инкрементов // *Журн. физ. химии.* 2021. Т. 95. № 8. С. 1269. <https://doi.org/10.1134/S003602442107030X>
21. *Szekely T., Nefedov O.M., Gazzo G., Shinyayav V.I., Fritz D.* Gas chromatographic investigation of the thermal decomposition of polydimethylsilylene // *Acta Chem. Acad. Sci. Hung.* 1967. V. 54. № 3–4. P. 241.