

УДК 54.022.1:54.021:543.544

**КОДИРОВАНИЕ ОСОБЕННОСТЕЙ СТРУКТУРЫ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ
ДЛЯ ОЦЕНКИ ХРОМАТОГРАФИЧЕСКИХ ИНДЕКСОВ УДЕРЖИВАНИЯ
С ИСПОЛЬЗОВАНИЕМ АДДИТИВНЫХ СХЕМ**

© 2010 И.Г. Зенкевич*, А.И. Уколов

*Санкт-Петербургский государственный университет, Химический факультет**Статья поступила 27 февраля 2009 г.**С доработки — 28 июля 2009 г.*

Один из современных вариантов аддитивных схем для оценки газохроматографических индексов удерживания органических соединений, не охарактеризованных такими константами, предполагает "сборку" их структур из более простых предшественников и последующие арифметические операции с известными значениями индексов в соответствии с выбранными схемами. Для оптимизации применения такого алгоритма (подтверждения правильности выбора структурных аналогов, исключения неопределенностей их суперпозиции и т.д.) предлагается использование системы линейного кодирования не полных структур молекул, а только тех их фрагментов, которые в наибольшей степени влияют на хроматографические параметры удерживания.

Ключевые слова: органические соединения, особенности структуры, кодирование, хроматографические индексы удерживания, аддитивные схемы расчета.

Различные системы номенклатуры органических соединений (прежде всего, IUPAC) обеспечивают взаимно-однозначное соответствие структур молекул их названиям, которые можно рассматривать как соответствующие им линейные коды [1]. Такое программное обеспечение, как ChemDraw и IsisDraw, включает опции для генерирования названий по заданным структурам и наоборот. При этом часто применяют и несистемные варианты номенклатуры, в частности, для природных соединений широко распространена практика использования тривиальных и гибридных названий [2].

Помимо систематической номенклатуры известно большое число способов линейной записи структурных формул органических соединений, исторически первым из которых стал так называемый код Висвессера (WLN — Wiswesser Line Notation), предложенный в 1947 г. [3]. В нем используют числовые символы и буквы латинского алфавита; например структуре $\text{CH}_3\text{CO—S—P(S)(OC}_2\text{H}_5)_2$ соответствует следующая запись — 2COS PZS/2(Z2). Однако такая система кодирования оказалась неудобной как для визуальной, так и для компьютерной интерпретации, в результате чего ее "популярность" в последнее время заметно снизилась. Более широкое распространение получил способ, в котором возможно визуальное восприятие информации — система SMILES (Simplified Molecular Input Line Entry Specification) [4]; в качестве простейшего примера можно привести соответствующий код для бицикло[3.3.0]октана: (C—C—C—C1—C—C—C—C2), C1—C2. Такой вариант кодирования структур предусмотрен, например, в современной базе масс-спектров и хроматографических индексов удерживания Национального института стандартов и технологии США [5].

Почти все известные варианты линейной записи органических структур ориентированы на конкретные соединения, т.е. фактически являются альтернативными систематической номенклатуре. Между тем в практике современной хроматографической и хромато-масс-спектромет-

* E-mail: izenkevich@mail15.com

рической идентификации органических соединений достаточно часто встречаются варианты, предусматривающие на первой стадии отнесение неизвестных аналитов к соответствующим гомологическим рядам или иным крупным таксономическим категориям (групповая идентификация). Уточнение же структуры углеродного скелета молекул (индивидуальная идентификация) становится следующим (достаточно независимым) этапом интерпретации данных. Подобный подход реализуется, например, при использовании таких инвариантов гомологических рядов как масс-спектры ионных серий [6], разности газохроматографических индексов удерживания (ИУ) на стандартных полярных и неполярных фазах [7] и коэффициенты распределения аналитов в гетерофазных системах органических растворителей (K_p) совместно с их ИУ в форме дифференциальных параметров $j = kRI - \lg K_p$ [8], где k — коэффициент, приводящий вариации значений RI и $\lg K_p$ гомологов к сопоставимым шкалам.

Общей проблемой всех перечисленных алгоритмов групповой идентификации [6—8] является не сама концепция двухстадийной интерпретации данных, а неожиданные терминологические сложности при представлении названий рядов. Это, правда, неудивительно, так как существующие системы номенклатуры ориентированы исключительно на конкретные структуры и не всегда могут быть адекватно распространены на более крупные таксономические категории. Например, в ряду ароматических карбонильных соединений (арилалканоны и арилалканы) как по закономерностям масс-спектрометрической фрагментации, так и по хроматографическим параметрам удерживания отчетливо различаются не альдегиды и кетоны, а изомеры с сопряженными и несопряженными ароматическими системами и карбонильными группами. В подобном относительно простом случае еще можно использовать названия "сопряженные (или несопряженные) ароматические карбонильные соединения", как это сделано в монографии [6]. Однако если рассмотреть такой класс как ароматические непредельные карбонильные соединения (арилалкеноны и арилалкены), то число различимых на стадии групповой идентификации подгрупп гомологов в зависимости от типа сопряжения фрагментов Ar , $C=C$ и $C=O$ возрастает до семи:

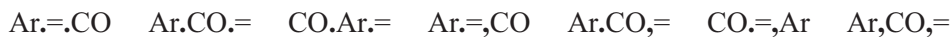
Сопряженные фрагменты структуры			Несопряженные фрагменты структуры		
Ar	$C=C$	$C=O^*$			
Ar	$C=O$	$C=C$			
$C=O$	Ar	$C=C$			
Ar	$C=C$		$C=O$		
Ar	$C=O$		$C=C$		
$C=O$	$C=C$		Ar		
			Ar	$C=O$	$C=C$

* Разделительные символы при перечислении указанных фрагментов структуры здесь опущены для того, чтобы избежать путаницы при обсуждении этого примера далее.

В таких случаях названия подгрупп гомологов становятся громоздкими и неудобными, например, "ароматические непредельные карбонильные соединения с арильными фрагментами, сопряженными с двойной связью $C=C$ и изолированными карбонильными группами".

Изложенные терминологические сложности послужили объективной посылкой для разработки системы линейного кодирования не формул конкретных соединений, а общих элементов структуры гомологов [9]. Примеры ее использования при групповой идентификации по разностям ИУ на фазах неодинаковой полярности приведены в статье [7], а в качестве классификаторов аналитов различной химической природы — в публикации [10]. Она включает набор визуально легко воспринимаемых дескрипторов разных фрагментов состава и структуры молекул (большинство из них соответствует обычным химическим символам, например " CO " — карбонильная группа, "=" — двойная связь $C=C$, "≡" — тройная связь $C≡C$, и т.д.), указателей их взаимного расположения (например, "." — сопряженные, ";" — гомосопряженные, "," — изолированные) и различных вспомогательных обозначений. Главное же, что такие коды допускают арифметические операции с входящими в них символами: вычеты молекулярных мас-

совых чисел по модулю 14 (или, иначе, номера гомологических групп соответствующих рядов), $y \equiv M(\text{mod}14)$ аддитивно складываются из значений y_i кодируемых фрагментов структуры, $y \equiv \sum y_i(\text{mod}14)$. Преимущества такой записи легко проиллюстрировать, вернувшись к упомянутому выше примеру непредельных ароматических карбонильных соединений; семь вариантов сопряжения структурных фрагментов в молекулах могут быть компактно записаны следующим образом, исключая необходимость их дополнительной словесной характеристики:



Таким образом, не подвергая сомнению ценность линейной записи полных структур органических соединений, можно отметить не менее важную проблему, а именно кодирование не всех, а только ключевых фрагментов структуры молекул, детерминируемых характером аналитических задач с целью оптимизации их решения. Одна из них — групповая хроматографическая или хроматоспектральная идентификация предполагает кодирование только структурных фрагментов, определяющих принадлежность аналитов к соответствующим гомологическим рядам. Другим же примером целесообразности линейной записи фрагментов структур органических соединений можно считать оценку их хроматографических ИУ с использованием аддитивных схем.

Настоящая статья посвящена не столько исчерпывающе подробному изложению системы кодирования особенностей структуры органических соединений, в наибольшей степени влияющих на их газохроматографические параметры удерживания (это требует существенного расширения числа рассматриваемых примеров ее применения), сколько обоснованию самой необходимости ее введения и использования.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Применение модифицированной аддитивной схемы расчета газохроматографических ИУ иллюстрировано примерами, приведенными в работах [11—13]. Свободнорадикальное хлорирование циклогексана молекулярным хлором в соответствии с рекомендациями работы [14] проведено к.х.н. Е.В. Елисеенковым и к.х.н. А.Н. Касаточкиным (Химический факультет СПбГУ); авторам предоставлены образцы реакционных смесей.

Анализ продуктов хлорирования циклогексана проведен на газовом хроматографе Цвет-100 с пламенно-ионизационным детектором и кварцевой WCOT колонкой со стандартной неполярной фазой DB-1, длиной 30 м, внутренним диаметром 0,32 мм, толщина пленки неподвижной фазы 0,25 мкм. Газ-носитель — азот, объемная скорость 1,0 мл/мин, деление потока при дозировании проб (0,5—1,0 мкл) 1:30. Использован режим линейного программирования температуры от 100 до 200 °C со скоростью 4 °C/мин, температура испарителя 250 °C. В качестве реперных компонентов для определения ИУ выбраны *n*-алканы C₆—C₉ (все гомологи) и C₁₁—C₁₇ (гомологи только с нечетным числом атомов углерода в молекуле). Регистрацию и первичную обработку данных проводили с использованием пакета программ МультиХром (версия 15). Для расчета линейно-логарифмических ИУ (алгоритм и программа приведены в руководстве [15]) использовано программное обеспечение Microsoft® Office Excel 2003.

Часть простейших продуктов хлорирования циклогексана (монохлор-, 1,1-дихлор, *цис*-1,2-дихлор-) была однозначна идентифицирована в составе реакционных смесей по известным или предварительно определенным (для синтезированных образцов) значениям ИУ. Значения ИУ семи изомерных дихлорциклогексанов были получены пересчетом приведенных в работе [14] их абсолютных времен удерживания, а для 12 изомерных три- (с учетом диастереомеров) и 32 изомерных тетрахлорциклогексанов оценены с использованием обсуждаемой ниже модифицированной аддитивной схемы [13].

В качестве источников справочных значений ИУ других соединений использована база данных Национального института стандартов и технологии (NIST, США) [5] и коллекция одного из авторов (И.З.). Ее принципиальное отличие от базы данных [5] состоит в том, что для оптимизации практического применения все литературные и экспериментальные значения ИУ представлены в статистически обработанном виде (средние значения и соответствующие стан-

дартные отклонения, $RI \pm s_{RI}$). База NIST в настоящее время включает литературные значения ИУ со ссылками на конкретные источники информации, но без какой-либо дополнительной обработки исходных данных.

ОБСУЖДЕНИЕ

Одним из важнейших преимуществ представления хроматографических параметров удерживания в интерполяционной форме ИУ, помимо достигаемой при этом межлабораторной воспроизводимости и, следовательно, возможности использования справочных данных для идентификации [5], является аддитивность ИУ, отражающая аддитивность энергий взаимодействия сорбат/сорбент или сорбат/неподвижная фаза [16]. Из этого следует возможность получения достаточно точных оценок ИУ для ранее неизвестных или не охарактеризованных соединений с использованием простейших аддитивных схем.

Традиционный "дизайн" аддитивных схем в органической химии предполагает применение совокупности параметров, называемых "инкрементами". Если некоторая базовая структура ХН характеризуется свойством Ω , то значение этого свойства для структурного аналога XR в первом приближении можно оценить суммированием величины $\Omega(XH)$ и инкремента $\Delta\Omega$, соответствующего преобразованию структуры $H \rightarrow R$:

$$\Omega(XR) \approx \Omega(XH) + \Delta\Omega(H \rightarrow R). \quad (1)$$

В отдельных случаях оценка значений свойств Ω для сложных структур $[AB...YZ]$ возможна суммированием вкладов всех структурных составляющих $\Omega(A) + \Omega(B) + \dots + \Omega(Y) + \Omega(Z)$. Одним из самых известных и наиболее разработанных примеров такого подхода являются системы атомных, групповых и связевых молярных рефракций [17], сохраняющие свое значение до настоящего времени.

Необходимые значения инкрементов оценивают по обратному к уравнению (1) соотношению, используя данные для известных веществ. Применительно к хроматографическим ИУ, это соответствует вычитанию индексов соединений с некоторой базовой структурой из значений ИУ аналогов, содержащих характеризуемые заместители. Например, оценка инкремента ΔRI (здесь и далее все значения ИУ и инкрементов указаны для стандартных неполярных полидиметилсилоксановых неподвижных фаз) замещения атома водорода в циклогексане атомом хлора равна:

$$(870 \pm 3)(\text{цикло-}C_6H_{13}Cl) - (666 \pm 8)(\text{цикло-}C_6H_{12}) = (204 \pm 9).$$

Полученная величина закономерно отличается от инкремента замещения $H \rightarrow Cl$ в *цис*-2-положении относительно уже присутствующего в молекуле атома хлора:

$$(1092 \pm 2)(\text{цис-1,2-цикло-}C_6H_{10}Cl_2) - (870 \pm 3)(\text{цикло-}C_6H_{11}Cl) = (222 \pm 4).$$

Многообразие подобных инкрементов представляет собой один из главных недостатков аддитивных схем. Другой проблемой эффективного использования величин ΔRI представляется неизбежная "потеря" информации о структурных особенностях молекул, которые были выбраны для их оценки, что затрудняет контроль правомерности их последующего применения. Указанные неопределенности, приводящие к неизбежному снижению точности оценок ИУ, послужили причиной модификации аддитивных схем, впервые предложенной в 2004 г. [11], с отказом от использования самого понятия "инкремент". Операции предварительного вычисления значений ΔRI и их последующего применения можно совместить, что эквивалентно "сборке" целевых структур из более простых молекул путем их суммирования и вычитания. Такой способ предусматривает выбор некоторого (желательно максимально близкого) структурного аналога целевого соединения, к которому следует добавить молекулу, содержащую недостающие структурные элементы и вычесть структуру, включающую дублирующие друг друга фрагменты.

Суть способа можно представить следующей общей схемой. Если для формирования целевой структуры ABCD выбраны предшественники ABC и BCD, так что



то

$$RI(ABCD) \approx RI(ABC) + RI(BCD) - RI(BC). \quad (2)$$

Иными словами, дублирующие друг друга фрагменты структуры молекул выбранных предшественников в итоге должны "сократиться", что обеспечивает требуемую стехиометрию такой операции. При этом возможна оценка стандартных отклонений результатов вычислений (s_{RI}) с использованием известного соотношения:

$$s_{RI(ABCD)} \approx [s_{RI(ABC)}^2 + s_{RI(BCD)}^2 + s_{RI(BC)}^2]^{1/2}. \quad (3)$$

Главными преимуществами такого способа оценок ИУ, помимо исключения необходимости предварительного вычисления инкрементов и связанных с этим неопределенностей, оказывается его структурная наглядность, а также возможности варьирования выбираемых структурных аналогов (способов "сборки" целевых структур), непосредственного использования справочных значений их ИУ и применения даже для диастереомеров [13].

В работе [11] подобная схема была предложена и использована для оценки ИУ 839 конгенов* полихлорированных гидроксифенилов. В отличие от полихлорированных дибензо-*n*-диоксинов (все 75 конгенов синтезированы и доступны в качестве образцов сравнения) или -бифенилов (то же для 209 конгенов), столь большое число структурно-родственных соединений не может быть синтезировано за разумное время, что полностью исключает возможность экспериментального определения ИУ каждого из них. Наиболее рациональный подход к решению задачи включал синтез лишь нескольких (не более десяти) соединений этого класса, определение их ИУ без выделения из реакционных смесей и последующий расчет ИУ всех остальных конгенов по рассматриваемой аддитивной схеме.

Этот же вариант аддитивной схемы столь же эффективно был использован для оценки ИУ 211 структурных изомеров 4-нонилфенолов [12], а также продуктов свободнорадикального хлорирования алкиларенов [18] и циклогексана [13], в число которых при $n_{Cl} \geq 2$ входят не только структурные изомеры, но и диастереомеры. Во всех случаях подтверждена его простота, наглядность и достаточно высокая точность получаемых оценок, что достигается оптимальным выбором схем формирования целевых структур. Однако расширение числа примеров применения этого подхода показывает, что однозначная "сборка" некоторых структур, например диастереомеров, из более простых аналогов возможна не всегда. Максимальная точность оценок ИУ рассматриваемым способом достигается только тогда, когда в молекулах предшественников представлены все особенности структуры целевых соединений, влияющие на их газохроматографические параметры удерживания. Общее соотношение (3), формально соответствующее выбранной схеме "сборки", не может отразить всех таких особенностей. Следовательно, для предотвращения возможных ошибок необходима некоторая минимальная формализация правил выбора структур предшественников и корректности операций с ними.

Один из способов решения такой задачи предполагает введение системы линейного кодирования ключевых фрагментов структуры, предназначенной для использования именно при получении оценок ИУ с применением модифицированной аддитивной схемы. Основные структурные факторы, в наибольшей степени влияющие на хроматографические ИУ органических соединений, представлены вместе с предлагаемыми для них кодами в табл. 1.

К числу таких факторов (различные элементы указывают, используя разделитель "+") следует отнести присутствие в молекулах вторичных и третичных атомов углерода (разветвления sp^3 -углеродного скелета молекул), положение заместителей при первичных, вторичных или третичных атомах углерода, и, главное, расположение различных групп в стерически затрудненных положениях. Известно, что именно стерические препятствия внутримолекулярным колебательным и вращательным процессам проявляются в увеличении как температур кипения таких изомеров по сравнению с другими, не имеющими подобных особенностей, так и изме-

* Конгенерами (англ. Congeners) принято называть соединения родственной химической природы, отличающиеся числом и расположением гетероатомов или более сложных функциональных групп (более общая таксономическая группа по сравнению с совокупностями изомеров или гомологов).

Т а б л и ц а 1

Некоторые особенности структуры молекул органических соединений, в наибольшей степени влияющие на их газохроматографические параметры удерживания, и их линейное кодирование

Фрагмент структуры	Символ	Примеры записи	
		Соединение	Код
Алкильные фрагменты, содержащие n атомов углерода	nC	Октилбензол	Ph + 8C
Кратные связи C=C, C≡C	=	Трибутиламин	N + 3(4C)
	≡	2-Октин	8C + ≡ или 8C + ≡ + 1C. + 5C.
Цикл, содержащий m атомов углерода	(m)	Бутилциклогексан	(6) + 4C
Сопряжение фрагментов X и Y	X.Y	Фенол	Ph.OH
Разветвление sp^3 -углеродного скелета (вторичный атом углерода)	*	Изопентан	5C + *
	* k	Изобутилбензол 3-Этилпентан 2-Этилгексилакрилат	Ph + 4C + * 5C + *2 =.CO.O + 8C + *2
Третичный атом углерода (два разветвления углеродного скелета при одном атоме углерода)	**	2,2,4-Триметилпентан	8C + ** + *
Два разветвления углеродного скелета в α -положении	* α	2,3-Диметилгексан Кумол	8C + 2* α Ph + 3C + * α
	[H]	Диэаноламин	NH + 2[2C + OH] + [H]
Заместители (символ X) при первичных (p), вторичных (s) или третичных (t) атомах углерода	p -X	2-Метил-2-хлорпропан	4C + * + t -Cl
	s -X	α -Метилбензиламин	Ar + 2C + s -NH ₂ α
	t -X	1,1-Диметилциклогексан	(6) + 2C(0)
Две функциональных группы (символы X,Y) в геминальных положениях	X,Y(0), если X=Y, то 2X(0)	2-Бутил-1,1-дихлорциклопропан	(3) + 4C + 2Cl. + 2Cl(0)
	X,Y(1)	1-Фенил-1,2-этандиол	Ph + 2C + + s -OH α . p -OH(1) + [H]
Частный случай: <i>орто</i> -положение заместителей (X,Y) в арильных фрагментах	X,Y(o)	2-Нитротолуол	Ph + C + .NO ₂ + + C.NO ₂ (o)
	<i>цис</i> -(<i>транс</i> -)Заместители в циклах в положениях n и $n+k$	<i>цис</i> - k Диметилциклопентан	(5) + 2C.() + <i>цис</i> -1
Любой фрагмент структуры (W или Ar), не затрагиваемый при трансформации структур	W	1-(Хлорметил)-4-	Ar + C + p -Cl
	Ar	трифторметилбензол	

няющихся симбатно с ними газохроматографических ИУ [19, 20]. К таким особенностям структуры относятся группы в геминальных, вицинальных или *орто*-положениях (в арильных фрагментах). Наиболее сложным и трудно предсказуемым образом на значения ИУ влияет присутствие нескольких *цис*- или *транс*-заместителей в циклических фрагментах структуры, поскольку это связано с их аксиальными или экваториальными положениями, проявлениями стерических факторов (в случае *цис*-1,2), а также возможностью изменения положения конформационных равновесий самих циклов. К еще более непредсказуемому уменьшению ИУ по сравнению с аддитивными оценками может приводить образование внутримолекулярных водородных связей.

Таким образом, предлагаемая символика включает перечисление всех фрагментов состава молекул (что необходимо для проверки стехиометрической корректности схем с их использованием) в сочетании с дополнительным указанием важнейших особенностей их относительного расположения. Обусловленное этим дублирование некоторых символов не представляется чрезмерным, так как при алгебраических операциях с такими линейными кодами при "сборке" целевых структур большая часть из них сокращается. Перечень символов, представленных в табл. 1, ограничен основными из них и при развитии рассматриваемого подхода может быть уточнен и дополнен.

Некоторые примеры "сборки" структур различных соединений для оценки их ИУ и проверки корректности таких схем с применением предлагаемого способа линейного кодирования структурных особенностей приведены в табл. 2 и обсуждаются ниже.

Первый пример (№ 1) иллюстрирует оценку ИУ (1-метил-1-хлорэтил)бензола с использованием данных для кумола, 2-хлорпропана и пропана ("вычитаемое"). Такой, по суммарной стехиометрии формально непротиворечивый вариант, приводит к совершенно неприемлемой (сильно завышенной) оценке ИУ. Причиной этого является выбор исходной структуры, содержащей атом хлора при вторичном атоме углерода, тогда как в целевой он должен быть локализован при третичном атоме углерода. Эта особенность однозначно проявляется при арифметических операциях с линейными кодами, так как символ "*s*-Cl" в итоге остается не сокращенным. Более того, полученная совокупность символов невероятна (вторичный атом хлора в разветвленном заместителе с тремя атомами углерода присутствовать не может), следовательно, соответствующая выбранной схеме аддитивная оценка ИУ заведомо неверна и должна быть отклонена. Получение подобных "отрицательных" заключений и является одной из важнейших целей рассматриваемого подхода, поскольку иные приемы не позволяют придти к аналогичным выводам.

Корректный вариант "сборки" той же структуры (№ 2) требует замены 2-метилпропана на 2-метил-2-хлорпропан, молекула которого содержит третичный атом хлора. Вычитаемым фрагментом в этом случае оказывается 2-метилпропан, а арифметические операции с линейными кодами приводят к последовательности, однозначно соответствующей требуемой структуре. Значение ИУ, вычисленное с применением такого варианта, в пределах стандартных отклонений закономерно совпадает с экспериментальным.

Пример № 3 иллюстрирует правомерность оценки ИУ 3,6-дихлор-1,1'-бифенил-2-ола с использованием данных для более простых 3-хлор-1,1'-бифенил-2-ола, 2,5-дихлор-1,1'-бифенила и 3-хлор-1,1'-бифенила [11].

Более сложным случаем, когда применение системы кодирования особенностей структур молекул приводит к необходимости проверки дополнительных гипотез, оказывается трансформация углеродного скелета 4-(1,4-диметил-1-этилпентил)фенола в структуру 4-(1-метил-1-пропилпентил)фенола [12]. Результат алгебраических действий с кодами фрагментов структуры содержит несокращенное слагаемое [$*\alpha_1, *\alpha_2 + *, *3 - *, *2$](0), которое можно привести к требуемому виду [$*\alpha_1, *\alpha_3$](0) только в том случае, если допустить, что $*, *3 \approx *, *2$, т.е. если хроматографические эффекты разветвления основной цепи атомов углерода слабо зависят от числа атомов углерода во вторичных цепях. Такое предположение нельзя принять без дополнительной проверки, которую проще всего провести сравнением справочных значений ИУ изомерных алканов. Например, для такой пары изомеров, как 2,3-диметилгексан [код $8C + 2*\alpha$; ИУ 761 ± 3] и 2-метил-3-этилпентан [$8C + *1*2\alpha$; 762 ± 2], значения ИУ практически одинаковы, однако для пары 2,3,5-триметилгептан [$10C + 2*\alpha + *$; 914 ± 2] и 2,5-диметил-3-этилгексан [$10C + *1, *2\alpha + *$; 894 ± 5] различия достигают 20 единиц индекса. Следовательно, вариант трансформации углеродного скелета в примере № 4 принципиально приемлем, но может приводить к некоторому возрастанию погрешностей получаемых оценок.

Пример № 5 иллюстрирует выявление иных неопределенностей, возможных в ходе "сборки" целевых структур из более простых предшественников. Для оценки ИУ 1-цис-2-транс-3-транс-4-тетрахлорциклогексана формально можно исходить из структуры 1-цис-2-транс-3-

Примеры контроля корректности "сборки" молекулярных структур для оценки газохроматографических ИУ с использованием их линейного кодирования

№	Соединение	Схема "сборки"
1	(1-Метил-1-хлорэтил)бензол	
2	(1-Метил-1-хлорэтил)бензол	
3	3,6-Дихлор-1,1'-бифенил-2-ол	
4	4-(1-Метил-1-пропилпентил)фенол	
5	1-цис-2-транс-3-транс-4-Тетрахлорциклогексан	
6	1-цис-2-цис-3-транс-4-Тетрахлорциклогексан	
7	1-цис-2-транс-3-транс-4-Тетрахлорциклогексан	
8	1-цис-2-цис-3-транс-4-Тетрахлорциклогексан	
9	цис-(транс)-1,1,3,5-Тетрахлорциклогексаны	
10	транс-1,1,3,5-Тетрахлорциклогексан	
11	То же (продолжение № 9)	
№	Оценка ИУ	Операции с линейными кодами (заключение)
1	2	3
1	$(914 \pm 6) + (496 \pm 5) - (300) \approx (1110 \pm 8)$ (эксп. значение ИУ 1070 ± 2)	$\text{Ph} + 3\text{C} + * \alpha + 3\text{C} + s\text{-Cl} - 3\text{C} = \text{Ph} + 3\text{C} + * \alpha + s\text{-Cl}$ (вариант невозможен)
2	$(914 \pm 6) + (526 \pm 5) - (366 \pm 6) \approx (1074 \pm 10)$ (эксп. значение ИУ 1070 ± 2)	$\text{Ph} + 3\text{C} + * \alpha + 4\text{C} + * + t\text{-Cl} - 4\text{C} + * = \text{Ph} + 3\text{C} + * \alpha + t\text{-Cl}$ (однозначный вариант)
3	$(1672 \pm 2) + (1691 \pm 1) - (1560 \pm 4) \approx (1803 \pm 5)$ (эксп. значение ИУ 1802 ± 2)	$\text{Ph} + .[\text{Ph} + \text{OH} + \text{Cl} + \text{OH}, \text{Ph}(\text{o}) + \text{OH}, \text{Cl}(\text{o})] +$ $+ \text{Ph} + .[\text{Ph} + 2\text{Cl} + \text{Cl}, \text{Ph}(\text{o})] - \text{Ph} + .[\text{Ph} + \text{Cl}.] =$ $= \text{Ph} + .[\text{Ph} + \text{OH} + 2\text{Cl} + \text{OH}, \text{Ph}(\text{o}) + \text{OH}, \text{Cl}(\text{o}) + \text{Cl}, \text{Ph}(\text{o})]$ (однозначный вариант)

О к о н ч а н и е т а б л . 2

1	2	3
4	$(1709 \pm 1) + (981 \pm 4) - (894 \pm 3) \approx$ $\approx (1736 \pm 5)$ (эксп. значение ИУ 1732 ± 1)	НО.Ph + 9C + [$*\alpha_1, * \alpha_2$](0) + * + 9C + [$*, *3$](0) - 9C + + [$*, *2$](0) + * = НО.Ph + 9C + [$*\alpha_1, * \alpha_2 + *, *3 - *, *2$](0) (требуется дополнительной проверки)
5	$(1247 \pm 5) + (1092 \pm 2) - (870 \pm 3) \approx$ $\approx (1469 \pm 6)$ (эксп. значение ИУ неизвестно)	$(6) + 3Cl.(.) + \text{цис-1} + \text{транс-1} + (6) + 2Cl.(.) + \text{цис-1} -$ $- (6) + Cl.(.) = (6) + 4Cl.(.) + 2\text{цис-1} + \text{транс-1}$ (неоднозначный вариант; см. следующую строку)
6	То же	То же
7	$2 \times (1247 \pm 5) - (1037 \pm 3) \approx (1457 \pm 8)$ (эксп. значение ИУ неизвестно)	$(6) + 3Cl.(.) + \text{цис-1} + \text{транс-1} + (6) + 3Cl.(.) + \text{транс-1} +$ $+ \text{цис-1} - (6) + 2Cl.(.) + \text{транс-1} = (6) + 4Cl.(.) + \text{цис-1} +$ $+ \text{транс-1} + \text{цис-1}$ (однозначный вариант)
8	$(1247 \pm 5) + (1302 \pm 5) - (1092 \pm 2) \approx$ $\approx (1457 \pm 7)$ (эксп. значение ИУ неизвестно)	$(6) + 3Cl.(.) + \text{цис-1} + \text{транс-1} + (6) + 3Cl.(.) + \text{цис-1} + \text{цис-1} -$ $- (6) + 2Cl.(.) + \text{цис-1} = (6) + 4Cl.(.) + \text{цис-1} + \text{цис-1} + \text{транс-1}$ (однозначный вариант)
9	$2 \times (1142 \pm 3) - (978 \pm 3) \approx (1306 \pm 5)$ (эксп. значение ИУ неизвестно)	$(6) + 3Cl.(.) + 2Cl(0) + (6) + 3Cl.(.) + 2Cl(0) - (6) + 2Cl.(.) +$ $+ 2Cl(0) = (6) + 4Cl.(.) + 2Cl(0)$ (неоднозначный вариант)
10	$(978 \pm 3) + (1223 \pm 5) - (870 \pm 3) \approx$ $\approx (1331 \pm 7)$ (эксп. значение ИУ неизвестно)	$(6) + 2Cl.(.) + 2Cl(0) + (6) + 3Cl.(.) + \text{цис-2} + \text{транс-2} -$ $- (6) + Cl.(.) = (6) + 4Cl.(.) + 2Cl(0) + \text{цис-2} + \text{транс-2}$ (переопределенный вариант)
11	$2 \times (1142 \pm 3) - (978 \pm 3) + (1048 \pm 2) -$ $- 2 \times (870 \pm 3) + (666 \pm 8) \approx (1280 \pm 11)$ (эксп. значение ИУ неизвестно)	$(6) + 3Cl.(.) + 2Cl(0) + (6) + 3Cl.(.) + 2Cl(0) - (6) + 2Cl.(.) -$ $- 2Cl(0) + (6) + 2Cl.(.) + \text{транс-2} - (6) + Cl.(.) -$ $- (6) + Cl.(.) + (6) = (6) + 4Cl.(.) + 2Cl(0) + \text{транс-2}$ (однозначный, но чрезмерно сложный вариант)

трихлорциклогексана, комбинируя ее со структурой *цис-1,2*-дихлорциклогексана. Однако можно заметить, что полученный код не содержит указаний на относительное расположение четырех атомов хлора в молекуле, так как полученную последовательность четырех вицинальных заместителей можно интерпретировать и как *цис + цис + транс*, и как *цис + транс + цис*. Следовательно, ему соответствует не только структура *1-цис-2-транс-3-транс-4*-тетрахлорциклогексана, но и альтернативная структура *1-цис-2-цис-3-транс-4*-изомера (пример № 6), которая получается в результате иной взаимной ориентации молекул предшественников.

Таким образом подтверждается эмпирически найденный в работе [13] факт, что схему сборки молекул изомерных тетрахлорциклогексанов исходя из данных для три- и дихлорзамещенных (т.е. вида $3 + 2 - 1 = 4$ по числу атомов хлора в молекуле) нельзя использовать при оценке их ИУ, так как для отдельных диастереомеров она не обеспечивает получения однозначных результатов. Для достижения требуемой однозначности оценок необходимо изменить выбор предшественников, т.е. перейти к схеме "сборки" вида $3 + 3 - 2 = 4$. В этом случае "перекрытие" структур двух выбранных трихлорпроизводных увеличивается до фрагмента, содержащего два вицинальных атома хлора, что приводит к более жесткому фиксированию их взаимной ориентации. Например, для формирования структуры *1-цис-2-транс-3-транс-4*-тетрахлорциклогексана необходима комбинация двух молекул *1-цис-2-транс-3*-трихлорциклогексана с вычитанием *транс-1,2*-дихлорциклогексана (пример № 7). При этом в последовательности линейных кодов можно учесть взаимное расположение структурных фрагментов, что отражено смещением их символов. В результате такого изменения схемы структуру требуемого *1-цис-2-транс-3-транс-4*-тетрахлорциклогексана удается задать однозначно. Получение альтер-

нативной структуры 1-*цис*-2-*цис*-3-*транс*-4-тетрахлорциклогексана из тех же предшественников становится принципиально невозможным, так как для этого необходим выбор иных структур (пример № 8).

Однако даже при использовании схемы "сборки" вида $3 + 3 - 2 = 4$ для некоторых диастереомеров не удастся избежать неопределенностей. Например, комбинируя две молекулы 1,1,3-трихлорциклогексана, нельзя однозначно придти к структурам ни *цис*-, ни *транс*-1,1,3,5-тетрахлорциклогексанов (пример № 9). Операции с соответствующими кодами закономерно приводят к неопределенной последовательности символов, в которой не указано расположение двух из четырех атомов хлора. Альтернативный способ "сборки" (пример № 10) также не приводит к требуемой однозначности. Здесь в результате арифметических операций с кодами часть символов (любой из дескрипторов *цис*- или *транс*-) остается не сокращенной, так что соответствующая им структура переопределена (указано положение пяти атомов хлора в молекуле вместо четырех) и, следовательно, лишена смысла.

Интересно заметить, что именно использование линейного кодирования фрагментов структуры в последнем случае все-таки позволяет выявить способ однозначного решения задачи, который, однако, связан с существенным усложнением процедуры "сборки". Если схема № 9 приводит к неопределенной геометрии молекулы, то ее можно дополнительно скомбинировать со структурой требуемого изомера (например, *транс*-1,2-дихлорциклогексана), вычестить две молекулы монохлорпроизводного и скомпенсировать недостающее число циклов молекулой незамещенного циклогексана. Однако использование такой последовательности менее удобно для практических целей, так как при усложнении схемы "сборки" (увеличении числа исходных структур от трех до семи) неизбежно возрастает неопределенность получаемых оценок ИУ, что, в соответствии с формулой 3, отражается в увеличении их стандартных отклонений. Такой вариант (№ 11 в табл. 2) для 1,1,3-*транс*-5-тетрахлорциклогексана дает значение 1280 ± 11 . Экспериментальное значение ИУ этого конгенера неизвестно, но компонент с близкой величиной ИУ обнаружен среди продуктов свободнорадикального хлорирования циклогексана [13].

С целью дополнительной характеристики возможностей применяемой для хлорциклогексанов аддитивной схемы она была проверена на примере метилциклогексанов с использованием доступных справочных данных [5]. Оказалось, что аддитивные оценки ИУ всех 12 изомерных триметилциклогексанов по известным значениям ИУ семи диметилпроизводных характеризуются средним отклонением ± 8 ед. индекса, а для 15 из 32 изомерных тетраметилзамещенных оно закономерно увеличивается до ± 12 ед. индекса. Такие погрешности можно считать вполне допустимыми для аддитивных схем.

Конечно же, рассмотренные варианты линейного кодирования особенностей структуры органических соединений для оценки их хроматографических ИУ с использованием аддитивных схем не охватывают всего многообразия возможностей такого подхода. Его распространение на иные классы соединений может привести к дополнительной детализации предлагаемых символов и уточнению их минимально необходимого перечня. Тем не менее нельзя сомневаться в целесообразности такого кодирования для контроля правильности применения аддитивных схем при оценке газохроматографических ИУ и, следовательно, для повышения надежности результатов хроматографической и хроматоспектральной идентификации органических соединений.

Работа выполнена в соответствии с темой гранта Российского фонда фундаментальных исследований № 08-03-00980.

СПИСОК ЛИТЕРАТУРЫ

1. Хлебников А.Ф., Новиков М.С. Современная номенклатура органических соединений. – СПб.: "Профессионал", 2004.
2. Растительные ресурсы России. Т. 1. / Под ред. А.Л. Буданцева. – СПб.-М.: КМК, 2008.
3. Smith E.G. The Wiswesser Line Formula Notation. – N.Y.: McGraw Hill Co., 1968.

4. http://en.wikipedia.org/wiki/Simplified_molecular_input_line_entry_specification (февраль 2009)
5. NIST Standard Reference Database, Number 69, June 2005. National Institute of Standards & Technology, Gaithersburg, MD, 20899. (<http://webbook.nist.gov>) (апрель 2009 г.)
6. *Зенкевич И.Г., Иоффе Б.В.* Интерпретация масс-спектров органических соединений. – Л.: Химия, 1986.
7. *Зенкевич И.Г.* // Журн. аналит. химии. – 2003. – **58**, № 2. – С. 110 – 129.
8. *Зенкевич И.Г., Васильев А.В.* // Там же. – 1993. – **48**, № 3. – С. 473 – 486.
9. *Зенкевич И.Г., Чупалов А.А.* // Вестн. СПб. ун-та. Сер. физ.-хим. – 1994. – № 4. – С. 45 – 51.
10. *Zenkevich I.G.* // Process Control Quality. – 1997. – **9**. – P. 67 – 78.
11. *Zenkevich I.G., Moeder M., Koeller G., Schrader S.* // J. Chromatogr. A. – 2004. – **1025**. – P. 227 – 236.
12. *Zenkevich I.G., Makarov A.A., Schrader S., Moeder M.* // Ibid. – 2009. – **1216**. – P. 4097 – 4106.
13. *Зенкевич И.Г., Елисеенков Е.В., Касаточкин А.Н.* // Масс-спектрометрия. – 2009. – **6**, № 2. – С. 137 – 148.
14. *Rainer K.D., Luszyk J., Ingold K.U.* // J. Amer. Chem. Soc. – 1988. – **85**. – P. 2161 – 2171.
15. *Столяров Б.В., Савинов И.М., Витенберг А.Г. и др.* Практическая газовая и жидкостная хроматография. – СПб.: Изд. СПб. ун-та, 2002.
16. *Zenkevich I.G.* // Encyclopedia of Chromatography / Ed. J. Cazes. – N.Y.: Taylor & Francis, 2005. – P. 901 – 907.
17. *Eisenlohr F.* Spektrochemie Organischer Verbindungen. Molekularrefraktion und-dispersion. – Stuttgart: Verlag von F. Enke, 1912.
18. *Зенкевич И.Г.* // Журн. орган. химии. – 2001. – **37**, № 2. – С. 283 – 293.
19. *Зенкевич И.Г.* // Там же. – 1998. – **34**, № 10. – С. 1463 – 1470.
20. *Зенкевич И.Г.* // Там же. – 2003. – **39**, № 8. – С. 1127 – 1133.