

МОЛЕКУЛЯРНО-СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДЛЯ ИДЕНТИФИКАЦИИ НЕИЗВЕСТНЫХ СОЕДИНЕНИЙ

© 2020 г. Д. Д. Матюшин^{а,*}, А. Е. Карнаева^а, А. К. Буряк^а

^а Российская академия наук, Институт физической химии и электрохимии им. А.Н. Фрумкина, 119071 Москва, Россия

*e-mail: dm.matiushin@mail.ru

Поступила в редакцию 12.06.2019 г.

После доработки 12.06.2019 г.

Принята к публикации 03.09.2019 г.

Показано, что метод молекулярно-статистических расчетов, в разработке и совершенствовании которого принимали активное участие сотрудники Института физической химии и электрохимии РАН, позволяет оценивать хроматографическое удерживание аналитов с высокой точностью. Разработанный вариант метода молекулярно-статистических расчетов (с учетом конформационной нежесткости) применен для предсказания величин удерживания алкилбензолов в условиях газоадсорбционной хроматографии. Выполнено сравнение точности оценки индексов удерживания с результатами современных методов предсказания удерживания на основе машинного обучения (для газожидкостной хроматографии), а также сравнение эффективности разделения структурных изомеров углеводородов на колонках разных видов.

Ключевые слова: газовая хроматография, адсорбция, углеродные сорбенты, математическое моделирование

DOI: 10.31857/S004445372003022X

Графитированная термическая сажа (ГТС) находит применение в газоадсорбционной хроматографии в качестве сорбента. Одним из первых применять ГТС в газовой хроматографии начал сотрудник Института физической химии АН СССР А.В. Киселев с коллегами из Московского государственного университета [1]. Данный сорбент состоит из частиц углерода, имеющих графитоподобную структуру и сравнительно однородную поверхность, содержащую относительно немного функциональных групп. Это позволяет сравнительно легко и эффективно моделировать адсорбцию органических молекул на молекулярном уровне. Первые работы по теоретическому вычислению адсорбции на ГТС, использующие атом-атомное приближение для вычисления потенциала взаимодействия сорбат–сорбент, были выполнены в конце 50-х – начале 60-х годов сотрудниками ИФХ АН СССР [2–4].

В последующие годы эти работы легли в основу так называемого метода молекулярно-статистических расчетов, позволяющего, в приближении классической молекулярной динамики, рассчитывать константу Генри адсорбции и энергию Гиббса адсорбции на однородных поверхностях и предсказывать характеристики удерживания [5].

Данный метод отличается сравнительно небольшим набором эмпирически подбираемых параметров и высокой точностью (погрешность не более 1%) для некоторых классов органических соединений. Использование различных вариантов метода для идентификации химических соединений (в том числе для предсказания порядка элюирования для смесей изомеров) описано в обзоре [6]. Исходный вариант метода обладал очень ограниченной областью применения, ввиду возможности расчета только для конформационно-жестких молекул. Вычисления для молекул с несколькими углами внутреннего вращения оказывались очень трудоемкими.

Недавно был разработан новый вариант молекулярно-статистического метода с использованием алгоритма Метрополиса для перебора всех возможных конформаций молекулы и для оценки внутренней энергии конформации [7–9]. Метод был протестирован на ряде соединений (алканы, хлоралканы, простые эфиры), показана его достаточная точность для больших групп молекул.

На данный момент ГТС выпускается под торговыми марками Carbopack, Carboblack, Carbochrom, кроме того, коммерчески доступны и готовые микронабивные колонки на основе этого

сорбента [10]. ГТС используется в различных исследованиях, в том числе в химической кинетике для анализа реакционных смесей [11, 12]. Эффективное разделение смесей изомеров, соединений, отличающихся структурой и геометрическим строением, является характерной особенностью сорбентов данного класса [13].

Способы теоретического предсказания характеристик удерживания в газовой и жидкостной хроматографии активно разрабатываются в последние годы. Они могут быть применены при подготовке эксперимента и непосредственно для подтверждения или опровержения предположения о структуре неизвестного соединения, посредством сравнения рассчитанных и наблюдаемых характеристик, в частности, удерживания. Для этого используются самые разнообразные методы машинного обучения и подбора эмпирических формул и зависимостей [14, 15].

Цель данной работы – сравнение метода машинного обучения (на основе нейронной сети) и молекулярно-статистического метода на примере анализа смесей изомеров алкилбензолов.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Были использованы экспериментальные данные по удерживанию на ГТС из литературных источников: для алкилбензолов – из работы [13] (приведены индексы удерживания, измеренные при различных температурах). Взаимодействие сорбата и однородной поверхности графита описывалось атом-атомными потенциалами в форме Леннард-Джонса, с параметрами из [16]. Для атомов углерода с поверхности ГТС использовались параметры: $\epsilon = 0.295$ кДж/моль, $r = 1.908$ Å [7].

Внутренняя энергия конформаций, необходимая для применения алгоритма Метрополиса, рассчитывалась с использованием параметров взаимодействий MMFF94 [17] и соответствующих правил. Выполнялся перебор конформаций посредством алгоритма Метрополиса, использовался вариант молекулярно-статистического метода с учетом внутреннего вращения [7–9]. Произвольные шаги включали в себя поворот молекулы вокруг произвольной связи, перемещение произвольного атома (до 0.1 Å) и изменение длины произвольной связи (до 0.1 Å) с вероятностями 0.25, 0.5, 0.25 соответственно. Для каждой молекулы выполнялось 10^7 шагов алгоритма Метрополиса, изменение результата при дальнейшем увеличении числа шагов составляли не более 0.05 в единицах логарифма константы Генри адсорбции. Все расчеты выполнялись при помощи собственного программного обеспечения [7].

Вычисление индексов удерживания выполнялось с помощью глубокой сверточной нейронной сети по методам [18, 19]. Нейронная сеть состояла

из двух сверточных слоев со 150 выходными каналами, объединяющего слоя (усреднение) и двух полносвязанных слоев (200 скрытых узлов, 1 выходной узел). Использовалась функция активации ReLU (на всех слоях кроме последнего, на последнем – линейная функция активации):

$$f = \begin{cases} 0 & \text{для } x < 0, \\ x & \text{для } x \geq 0. \end{cases}$$

В качестве входных данных использовались записанные в форме унитарного кода строки SMILES [20], соответствующие молекулам аналитов. Обучение производилось с использованием базы данных масс-спектров и индексов удерживания NIST 2008, из которой были предварительно исключены алкилбензолы, для которых выполнялось тестирование. Подробнее метод расчета описан в работе [19]. В случае наличия в базе NIST 2008 нескольких индексов удерживания для одной молекулы – они усреднялись. Индексы рассматривались для неполярных (полиметилсилоксан, 5%-фенил-полидиметилсилоксан, сквалан) жидких неподвижных фаз.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В табл. 1 приведены результаты молекулярно-статистического расчета логарифмов констант Генри адсорбции для всех рассмотренных алкилбензолов и индексы удерживания: экспериментальные на ГТС и неполярных жидких фазах и рассчитанные для неполярных жидких фаз. Экспериментальное [13] определение индексов удерживания выполнялось при разных температурах, все расчеты выполнялись для 150°C. Выполнение расчетов для одной температуры необходимо, чтобы получить данные для разных соединений, сравнимые между собой [7].

Корреляционная зависимость между индексом удерживания и рассчитанными константами Генри приведена на рис. 1а. Коэффициент корреляции составляет 0.994. Среднеквадратичное отклонение между индексами, рассчитанными по уравнению корреляционной зависимости, и экспериментальными составляет 19.1. Среднее абсолютное отклонение составляет 14.1, среднее относительное – 1.3%, медианное абсолютное отклонение – 10.2. Корреляционная зависимость для индексов удерживания на неполярной жидкой неподвижной фазе приведена на рис. 1б. Среднеквадратичное отклонение между индексами, рассчитанными по уравнению корреляционной зависимости, и экспериментальными составляет 20.2. Среднее абсолютное отклонение составляет 15.4, среднее относительное – 1.4%, медианное абсолютное отклонение – 11.5.

Так как с практической точки зрения наибольший интерес представляет не абсолютная вели-

Таблица 1. Экспериментально измеренные индексы [13] удерживания Ковача алкилбензолов на ГТС (I_{gtcb}), на неполярных жидких неподвижных фазах (I_n) и рассчитанные индексы удерживания на неполярных жидких фазах (I_{cn}) и ГТС (I_{cg}), константы Генри адсорбции ($\ln K_1$ [cm^3/m^2]) на ГТС

Вещество	I_{gtcb}	I_n	I_{cn}	I_{cg}	$\ln K_1$
Бензол	561	655	680	567	-0.17
Толуол	699	756	775	703	1.50
Этилбензол	762	852	876	765	2.26
1,4-Диметилбензол	840	866	891	836	3.12
1,3-Диметилбензол	824	863	885	832	3.09
1,2-Диметилбензол	836	883	880	848	3.28
Изопропилбензол	798	919	927	786	2.51
<i>n</i> -Пропилбензол	852	945	956	862	3.45
<i>m</i> -Метилэтилбензол	894	953	963	891	3.81
<i>n</i> -Метилэтилбензол	882	952	984	893	3.83
<i>o</i> -Метилэтилбензол	888	969	977	908	4.01
1,3,5-Триметилбензол	945	970	982	952	4.56
1,2,4-Триметилбензол	970	986	1003	974	4.83
1,2,3-Триметилбензол	979	1010	1021	995	5.08
<i>трет</i> -Бутилбензол	845	993	996	845	3.24
<i>втор</i> -Бутилбензол	862	1002	1014	839	3.17
Изобутилбензол	910	996	996	916	4.11
1-Метил-4-изопропилбензол	918	1020	1032	892	3.81
1,3-Диэтилбензол	941	1040	1044	952	4.55
<i>n</i> -Бутилбензол	945	1049	1051	956	4.61
1-Метил-4- <i>n</i> -пропилбензол	988	1044	1058	986	4.98
1,4-Диэтилбензол	954	1058	1045	948	4.51
1,2-Диэтилбензол	941	1049	1051	959	4.64
1-Метил-2- <i>n</i> -пропилбензол	980	1056	1055	1002	5.17
1,4-Диметил-2-этилбензол	1019	1069	1040	1036	5.59
1,3-Диметил-4-этилбензол	1015	1072	1080	1028	5.48
1,2-Диметил-4-этилбензол	1021	1078	1093	1027	5.48
1,2,4,5-Тетраметилбензол	1111	1112	1120	1113	6.53
1,2,3,5-Тетраметилбензол	1107	1117	1151	1112	6.52
1,2,3,4-Тетраметилбензол	1131	1144	1138	1138	6.84
<i>втор</i> -Пентилбензол	933	1105	1110	909	4.02
1-Этил-4- <i>n</i> -пропилбензол	1047	1135	1123	1045	5.70
<i>n</i> -Пентилбензол	1045	1150	1158	1075	6.06
1-Этил-4- <i>n</i> -бутилбензол	1078	1147	1167	1084	6.18
Пентаметилбензол	1271	1266	1275	1271	8.48
1,3-Диизопропилбензол	998	1138	1149	993	5.07
1,3-Диметил-5- <i>трет</i> -бутилбензол	1021	1160	1158	1039	5.62
1,4-Диизопропилбензол	1015	1163	1169	979	4.89
1-Изопропил-4- <i>n</i> -пропилбензол	1079	1187	1208	1063	5.92
1,3-Ди- <i>n</i> -пропилбензол	1121	1192	1189	1142	6.89
1,4-Ди- <i>n</i> -пропилбензол	1149	1224	1215	1147	6.95
1-Этил-4- <i>n</i> -бутилбензол	1142	1230	1238	1141	6.88
<i>n</i> -Гексилбензол	1140	1247	1257	1176	7.31

Таблица 1. Окончание

Вещество	I_{gtcb}	I_n	I_{cn}	I_{cg}	$\ln K_1$
1-Метил-4- <i>n</i> -пентилбензол	1178	1240	1266	1204	7.65
Гексаметилбензол	1414	1433	1403	1417	10.3
1- <i>n</i> -Пропил-4- <i>n</i> -бутилбензол	1240	1311	1331	1240	8.10
1-Этил-4- <i>n</i> -пентилбензол	1244	1322	1341	1255	8.28
1-Метил-4- <i>n</i> -гексилбензол	1273	1333	1364	1309	8.95
1,4-Ди- <i>трет</i> -бутилбензол	1091	1295	1305	1045	5.70
1- <i>трет</i> -Бутил-4-изобутилбензол	1205	1291	1311	1180	7.36
1- <i>трет</i> -Бутил-4- <i>втор</i> -бутилбензол	1138	1291	1313	1096	6.33
1,4-Ди- <i>втор</i> -бутилбензол	1182	1304	1356	1131	6.75
1- <i>втор</i> -Бутил-4-изобутилбензол	1235	1306	1305	1196	7.56
1,4-Диизобутилбензол	1281	1308	1304	1272	8.49
1- <i>трет</i> -Бутил-4- <i>n</i> -бутилбензол	1216	1345	1389	1208	7.70
1- <i>втор</i> -Бутил-4- <i>n</i> -бутилбензол	1245	1358	1429	1212	7.75
1-Изобутил-4- <i>n</i> -бутилбензол	1309	1360	1376	1309	8.84
1,4-Ди- <i>n</i> -бутилбензол	1334	1411	1420	1337	9.29
1- <i>n</i> -Пропил-4- <i>n</i> -пентилбензол	1347	1419	1438	1352	9.47

чина адсорбции, а порядок элюирования изомеров, то был также рассчитан коэффициент Спирмена (K_s), который составил 0.993 для ГТС и

молекулярно-статистического метода и 0.994 для нейронной сети и жидкой неподвижной фазы. Коэффициент Спирмена является ранговым,

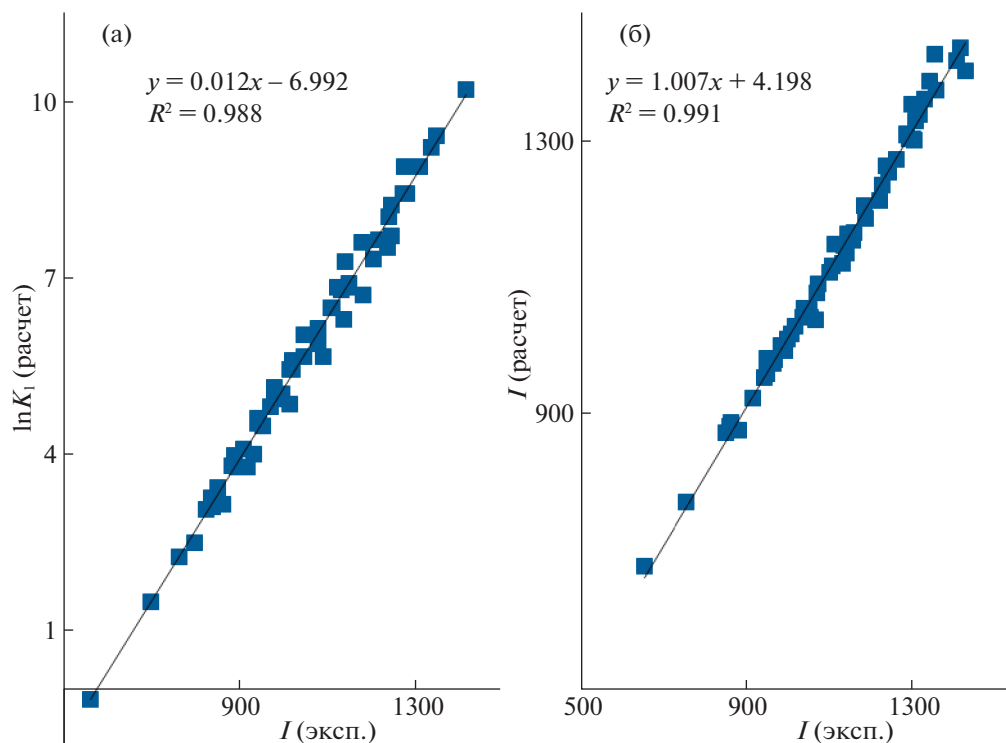


Рис. 1. Корреляция экспериментально измеренных индексов удерживания Ковача алкилбензолов на ГТС и рассчитанных констант Генри адсорбции ($\ln K_1$ [$\text{см}^3/\text{м}^2$]) (а) и корреляция рассчитанных с помощью нейронной сети и экспериментальных индексов удерживания для неполярной жидкой неподвижной фазы (б) для всех рассмотренных алкилбензолов.

Таблица 2. Коэффициенты корреляции и коэффициенты Спирмена (K_S) между экспериментальными и рассчитанными индексами удерживания алкилбензолов на ГТС (I) и на неполярной жидкой неподвижной фазе (II) для групп изомеров в зависимости от молекулярной массы

M , а.е.м.	I		II	
	R^2	K_S	R^2	K_S
106	0.981	0.800	0.123	0.400
120	0.991	0.929	0.960	0.857
134	0.990	0.960	0.954	0.882
148	0.990	0.900	0.985	0.900
162	0.987	0.927	0.986	0.976
176	0.996	1	0.977	1
190	0.994	0.991	0.904	0.715

т.е. характеризует не абсолютные значения, а именно ранг (порядок), в котором располагаются переменные [21].

Большинство (57 из 59) рассмотренных в табл. 1 соединений составляют шесть групп изомеров с молекулярными массами 106, 120, 134, 148, 162, 176, 190. В табл. 2 для каждой группы изомеров приведены коэффициент корреляции и коэффициент Спирмена между рассчитанными и наблюдаемыми параметрами удерживания. Рассмотренные вычислительные методы не позволяют всегда предсказать, какой из изомеров имеет больший индекс удерживания для всех пар изомеров, однако в большинстве случаев, когда индексы удерживания отличаются достаточно сильно, чтобы можно было наблюдать два полностью разделенных пика на хроматограмме, расчет позволяет правильно предсказать порядок элюирования изомеров.

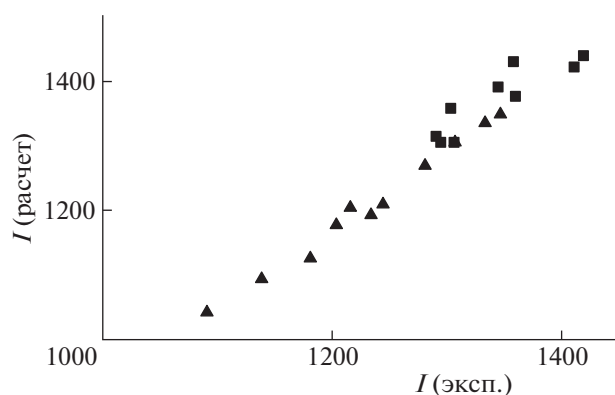


Рис. 2. Корреляционная зависимость между рассчитанными и экспериментальными индексами удерживания для неполярной жидкой неподвижной фазы (■) и ГТС (▲) соответственно.

Молекулярно-статистический метод лишь немногим лучше предсказывает индексы удерживания для всех алкилбензолов вместе, но для каждой из серий изомеров этот метод лучше предсказывает порядок элюирования изомеров.

На рис. 2 показано сравнение результатов расчета и эксперимента для изомеров октилбензола (10 изомерных *para*-дибутилбензолов и *para*-пропилбутилбензол) для обеих неподвижных фаз. На ГТС наблюдается значительно больший диапазон индексов удерживания среди изомеров, а также лучшая корреляция между результатами расчета и эксперимента. Аналогичная картина наблюдается и для других серий изомеров. Таким образом, несмотря на то, что точность этих методов для предсказания индексов удерживания почти одинакова, для серий изомеров молекулярно-статистический метод расчета и неподвижная фаза ГТС подходят несколько лучше.

Однако работа с ГТС зачастую связана с рядом сложностей (уширение пиков, невысокая эффективность разделения, низкая воспроизводимость времен удерживания, необратимая адсорбция ряда аналитов) и молекулярно-статистический расчет более требователен к ресурсам по сравнению с расчетом с помощью нейронной сети. Выбор неподвижной фазы и способа расчета должен зависеть от конкретной задачи, однако, рассмотренный пример алкилбензолов показывает, что применение молекулярно-статистического метода перспективно при анализе смесей структурных изомеров.

Работа выполнена при поддержке программы фундаментальных исследований президиума РАН № 14-П “Физическая химия адсорбционных явлений и актинидных наночастиц” за 2018–2020 годы.

СПИСОК ЛИТЕРАТУРЫ

1. Васильев В.С., Киселев А.В., Никитин Ю.С. и др. // Журн. физ. химии. 1961. Т. 35. № 8. С. 1889.
2. Avgul N.N., Berezin G.I., Kiselev A.V. et al. // Bulletin of the Academy of Sciences of the USSR, Division of Chemical Science. 1956. V. 5. № 11. P. 1339. <https://doi.org/10.1007/BF01177699>
3. Kiselev A.V., Lygina I.A. // Ibid. 1962. V.11. № 8. P. 1271. <https://doi.org/10.1007/BF00907967>
4. Kiselev A.V. et al. // Ibid. 1962. V. 11. № 8. P. 1264. <https://doi.org/10.1007/BF00907966>
5. Kiselev A.V., Poshkus D.P., Grumadas A.J. // J. Chem. Soc., Faraday Trans. 1: Physical Chemistry in Condensed Phases. 1979. V. 75. P. 1288. <https://doi.org/10.1039/F19797501288>
6. Буряк А.К. // Успехи химии. 2002. Т. 71. № 8. С. 788. Buryak A.K. // Russ. Chem. Revs. 2002. V. 71. № 8. P. 695. <https://doi.org/10.1070/RC2002v071n08ABEH000711>

7. *Матюшин Д.Д., Буряк А.К.* // Сорбционные и хроматографические процессы. 2017. Т. 17. № 2. С. 204.
<https://doi.org/10.17308/sorpchrom.2017.17/372>
8. *Полунин К.Е., Матюшин Д.Д., Ульянов А.В. и др.* // Коллоидн. журн. 2019. Т. 81. № 4. С.493. *Polunin K.E., Matyushin D.D., Ulyanov A.V. et al.* // Colloid Journal. 2019. V. 81. № 4. P. 431
<https://doi.org/10.1134/S1061933X19030104>
9. *Матюшин Д.Д., Буряк А.К.* // Журн. аналит. химии. 2019. Т. 74. № 7 (приложение). *Matyushin D.D., Buryak A.K.* // J. Analyt. Chem. 2019. V.74. Suppl. 7. P. 775.
<https://doi.org/10.1134/S1061934819070165>
10. CarboBlack Columns [Электронный ресурс]. Режим доступа <http://www.restek.com/catalog/view/7019> Дата обращения 23.05.2019.
11. *Subramoney S.C., Nelson W.M., Naidoo P. et al.* // Fluid Phase Equilibria. 2015. V. 406. P. 156.
<https://doi.org/10.1016/j.fluid.2015.07.039>
12. *Bengesai P.N., Nelson W.M., Naidoo P., Ramjugernath D.* // J. Chem. & Eng. Data. 2016. V. 61. № 9. P. 3363.
<https://doi.org/10.1021/acs.jced.6b00409>
13. *Engewald W., Wennrich L., Pörschmann J.* // Chromatographia. V. 11. № 8. P. 434.
14. *Mihaleva V.V., Verhoeven H.A.* //Bioinformatics. 2009. V. 25. № 6. P. 787.
<https://doi.org/10.1093/bioinformatics/btp056>
15. *Qiu F., Lei Z., Sumner L.W.* //Anal. Chim. Acta. 2018. V. 1037. P. 316.
<https://doi.org/10.1016/j.aca.2018.03.052>
16. *Wang J., Wolf R.M., Caldwell J.W. et al.* // J. Comput. Chem. 2004. V. 25. P. 1157.
<https://doi.org/10.1002/jcc.20035>
17. *Halgren T.A.* // Ibid. 1996. V. 17. № 5–6. P. 490.
[https://doi.org/10.1002/\(SICI\)1096-987X\(199604\)17:5/6<490::AID-JCC1>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P)
18. *Gomez-Bombarelli R., Wei J.N., Duvenaud D. et al.* // ACS Central Sci. 2018. V. 4. № 2. P. 268.
<https://doi.org/10.1021/acscentsci.7b00572>
19. *Matyushin D.D., Sholokhova A.Y., Buryak A.K.* // J. Chromatography A. 2019. P. 460395.
<https://doi.org/10.1016/j.chroma.2019.460395>
20. *Weininger D.* // J. Chem. information and computer sciences. 1988. V. 28. № 1. P. 31.
<https://doi.org/10.1021/ci00057a005>
21. *Kobayashi K., Pillai K.S.* A Handbook of Applied Statistics in Pharmacology. Boca Raton: CRC Press, 2012. P. 230.