

УДК: 547.022:615.015.11

**К. А. Клименко¹, Л. Н. Огниченко², В. Е. Кузьмин^{1,2}, А. Г. Артеменко²,
Л. Г. Горб³**

¹ Одесский национальный университет им. И. И. Мечникова, Украина, Одесса, ул. Дворянская 2

² Физико-химический институт им. А. В. Богатского НАН Украины, Украина, Одесса, 65080, Люстдорфская дорога 86, e-mail: ogni@ukr.net

³ Институт молекулярной биологии и генетики НАН Украины, Украина, Киев, 03143, ул. Заболотного 150, e-mail: lgorb@icnanotox.org

QSPR АНАЛИЗ ТЕМПЕРАТУРНОЙ ЗАВИСИМОСТИ РАСТВОРИМОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ В ВОДЕ

Разработаны 2D-QSPR модели для прогноза растворимости органических соединений в воде при разных температурах. Проанализированы вклады в растворимость различных структурных факторов, в частности, определен вклад заместителей в бензольном кольце. Построена адекватная модель, позволяющие оценить влияние температуры на растворимость разнообразных соединений в воде.

Ключевые слова: водная растворимость, молекулярные симплексы, QSPR, температурная зависимость растворимости.

Введение

Растворимость в воде является одним из важнейших свойств веществ. Множество процессов, таких как, коррозия металлов, отложение минералов, а также загрязнение окружающей среды, основаны на способности веществ образовывать истинные растворы с водой. Такая характеристика вещества как растворимость в воде также очень важна при производстве биодобавок и лекарственных препаратов, т. к. растворимость связана с биодоступностью [1].

Информация о растворимости новых соединений – важный фактор для оценки экологической опасности, для моделирования транспорта и метаболизма химических соединений и для понимания фармакокинетического поведения загрязнителей в живых организмах. В большинстве случаев водная растворимость новых химических веществ – определяющее свойство для оценки влияния этих веществ на окружающую среду. Однако, экспериментальные данные растворимости соединений часто недоступны, поэтому внеэкспериментальный скрининг растворимости весьма актуален.

Целью настоящей работы явилось построение 2D-QSPR моделей, оценивающих растворимость различных органических соединений при определенных температурах.

Материалы и методы

В настоящей работе были задействованы три выборки соединений: (1) состоящая из 80 разнообразных органических молекул, использованная для построения моделей, предсказывающих значения растворимости в воде при 20 °С, 30 °С и 40 °С, (2) состоящая из 23 ароматических молекул, использованная для определения

влияния различных заместителей у бензойного кольца на растворимость в воде при 20 °С, 25 °С, 30 °С и 40 °С, (3) состоящая из 244 данных по растворимости азотсодержащих соединений при различных температурах в диапазоне от 0 °С до 100 °С, использованная при построении модели, включающей температуру как параметр.

Данные по растворимости в воде (Sw) в ммоль/л при различных температурах для исследуемых соединений были взяты из справочника Ялковского [2], при построении моделей использовали логарифм растворимости (LgSw).

Предварительно выборка (1) была разделена на две части – обучающую и тестовую, состоящие из 64 и 16 молекул соответственно. Для формирования тестовой выборки молекулы были упорядочены согласно значениям их растворимости в воде, и затем каждая пятая молекула в этом ряду была отобрана в тестовую выборку, т. е. 20 % от общего количества молекул было отобрано в тестовую выборку. Аналогичная процедура была проведена и для выборки (3). Все молекулы выборки (2) были включены в обучение. Но в этом случае для исключения возможности ложной корреляции была проведена процедура *Y Scrambling* – создание модели методом случайного подбора коэффициентов. Статистические характеристики моделей, полученных при использовании процедуры *Y Scrambling*, должны быть ниже по показателям, чем у рабочих моделей.

В данной работе для расчета структурных дескрипторов был использован подход SiRMS [3, 4], основанный на симплексном представлении молекулярной структуры. В рамках данного подхода молекула представляется как совокупность различных симплексов – четырехатомных фрагментов фиксированного состава, структуры. Для дифференциации атомов в симплексе могут быть использованы различные атомные характеристики, в частности: уникальность атома (природа атома или более детализированный тип), частичный заряд на атоме, липофильность атома, рефракция, способность атома быть донором/акцептором водорода в потенциальной Н-связи и т. д. Для атомных характеристик, имеющих действительные значения (заряд, липофильность и т. п.) на предварительной стадии проводится разделение диапазона значений на определенное количество дискретных групп. Количество групп (G) является настроечным параметром и может варьироваться (как правило, G=3-7). Атомы в симплексах метятся в зависимости от принадлежности к той или иной группе (рис.1).

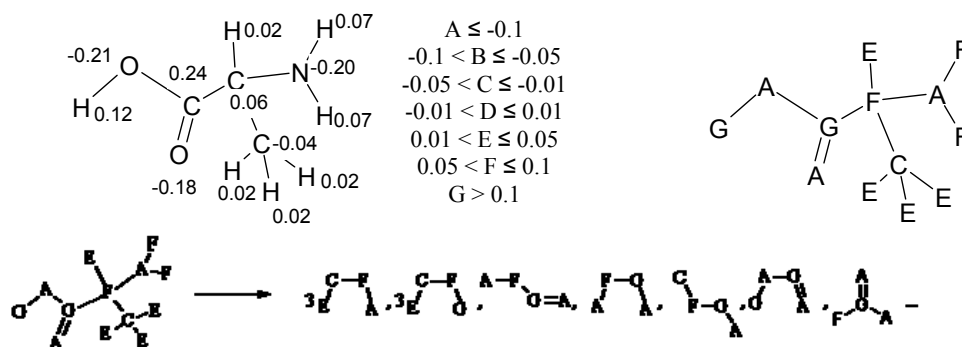


Рис. 1 Разбивка молекулы аланина на симплексы на 2D уровне (для дифференциации атомов использованы частичные атомные заряды).

Для построения моделей были использованы 2D¹ связанные симплексы, атомы в которых были взвешенные по типу атомов, частичным зарядам [5], липофильности [6], поляризуемости [7], способности выступать донором/акцептором водородной связи, ван-дер-ваальсовым взаимодействиям [7]. В качестве примера на рис. 1 представлена схема разбиения молекулы на симплексы для аланина.

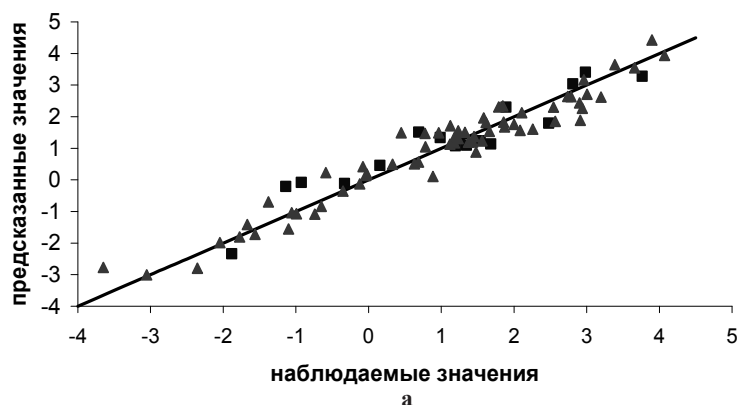
Для установления связи структура-свойство использовали популярный метод частичных наименьших квадратов (PLS) [8], хорошо зарекомендовавший себя при построении моделей с большим числом молекулярных дескрипторов. При построении PLS-соотношений предварительно были отсеяны взаимно-коррелирующие и постоянные параметры и использован генетический алгоритм [9] для формирования начальных наборов дескрипторов.

Результаты и их обсуждение

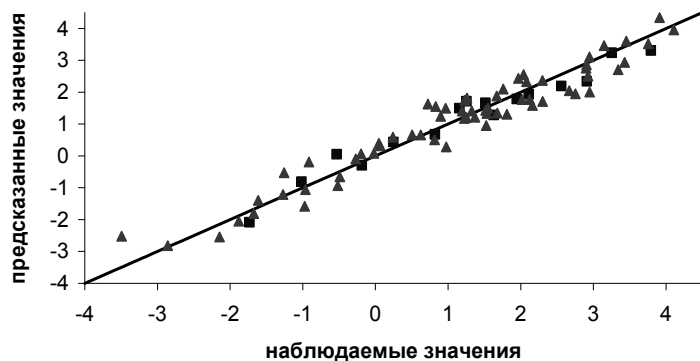
При построении моделей для предсказания растворимости органических соединений при температурах 20 °С, 30 °С и 40 °С для выборки (1) из 80 соединений был найден оптимальный набор дескрипторов, описывающих связь между растворимостью и структурой веществ. При определении набора дескрипторов кроме стандартной процедуры отбора мы использовали дескрипторы, для других температур. Т. е. дескрипторы, отобранные для построения модели при температуре 20 °С (модель T20a), были использованы при построении моделей T30a, T40a для температур 30°С и 40°С соответственно. Аналогично, на основе дескрипторов, отобранных для моделей при температурах 30 °С (модель T30б) и 40 °С (модель T40в) были созданы модели T20б, T40б и T20в, T30в соответственно.

Далее для групп моделей, полученных для определенных температур, результаты прогноза усреднялись.

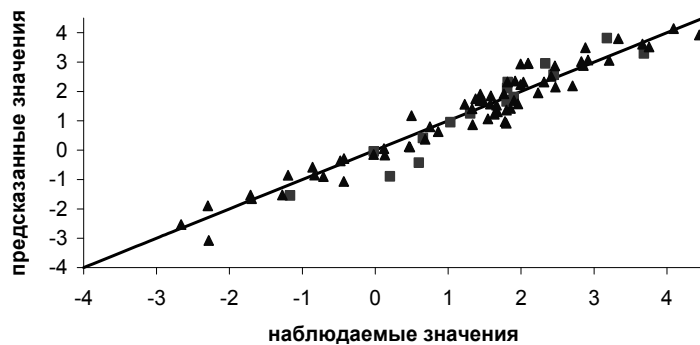
В результате такой процедуры были получены вполне адекватные консенсусные модели K20, K30, K40, описывающие связь структуры исследуемых молекул с их растворимостью в воде, для температур 20 °С, 30°С и 40 °С соответственно. На рис. 2 представлены зависимости наблюдаемых и предсказанных значений LgSw по моделям K20, K30, K40. Статистические характеристики полученных моделей свидетельствуют об их высокой устойчивости и предсказательной способности (см. табл. 1).



¹ 2D модель – учитывается только информация из структурных формул.



б



в

Рис. 2 Зависимости предсказанных и наблюдаемых значений $LgSw$ при температурах а) 20 °С, б) 30 °С и в) 40 °С, где треугольниками отмечены данные из обучающей выборки, а квадратами отмечены данные из тестовой выборки.

Низкие статистические показатели, полученные для процедуры *Y Scrambling* позволяют говорить о неслучайности установленной связи между структурой веществ и их растворимостью для всех рассмотренных моделей (см. табл.1).

Таблица 1

Сравнение статистических характеристик моделей

модель	N	A	k	R ²	Q ²	R ² (ts)	S (ts)	S (ws)	R ² (scr)	Q ² (scr)
T20a	80	2	35	0.95	0.90	0.84	0.64	0.41	0.35	0.19
T20б	80	2	45	0.93	0.85	0.90	0.50	0.47	0.39	0.22
T20в	80	2	36	0.92	0.84	0.87	0.58	0.49	0.30	0.18
K20	-	-	-	0.94	-	0.89	0.53	0.42	-	-
T30a	80	2	34	0.94	0.86	0.92	0.44	0.43	0.34	0.18
T30б	80	2	51	0.93	0.85	0.89	0.52	0.46	0.41	0.22

Продолжение таблицы 1

модель	N	A	k	R ²	Q ²	R ² (ts)	S (ts)	S (ws)	R ² (scr)	Q ² (scr)
T30в	80	2	37	0.91	0.84	0.93	0.42	0.51	0.33	0.19
K30	-	-	-	0.93	-	0.90	0.41	0.43	-	-
T40a	80	2	32	0.94	0.88	0.87	0.58	0.42	0.31	0.17
T40б	80	2	30	0.93	0.89	0.84	0.63	0.44	0.30	0.15
T40в	80	2	38	0.93	0.86	0.87	0.58	0.45	0.31	0.17
K40	-	-	-	0.92	-	0.87	0.50	0.47	-	-
M-20	23	3	24	0.95	0.90	-	-	0.24	0.56	0.35
M-25	23	3	24	0.96	0.91	-	-	0.23	0.63	0.38
M-30	23	3	24	0.95	0.91	-	-	0.24	0.60	0.35
M-40	23	3	24	0.96	0.91	-	-	0.22	0.57	0.33
B	244	5	37	0.91	0.84	0.88	0.38	0.30	-	-

где N – количество молекул,
 A – количество латентных переменных,
 k – количество дескрипторов,
 R² – коэффициент детерминации для обучающей выборки,
 Q² – коэффициент детерминации в условиях скользящего контроля,
 R²(ts) – коэффициент детерминации для тестовой выборки,
 S (ts) – стандартное отклонение для тестовой выборки,
 S (ws) – стандартное отклонение для обучающей выборки,
 R²(scr) – коэффициент детерминации для процедуры *Y Scrambling*,
 Q²(scr) – коэффициент кросс-валидации для процедуры *Y Scrambling*,

На рис. 3 представлена диаграмма, отражающая относительное влияние некоторых физико-химических свойств на растворимость веществ в воде при определенной температуре. Так как, влияние физико-химических параметров в указанном диапазоне температур практически не изменяется, мы приводим диаграмму только для 20°C.

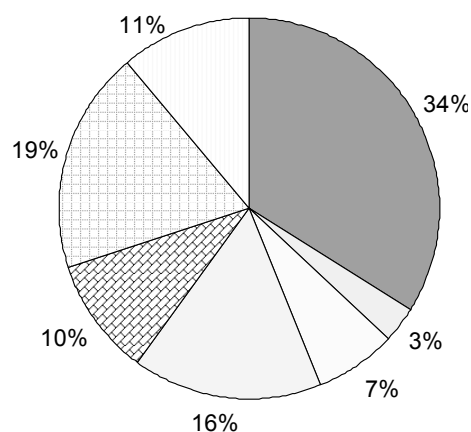


Рис.3 Относительное влияние некоторых физико-химических факторов на LgSw при 20°C.
 А – электростатика, В – ван-дер-ваальсовое отталкивание, С – ван-дер-ваальсовое притяжение, D – типы атомов, E – поляризуемость, F – липофильность, G – способность атома выступать в роли донора или акцептора водородной связи.

Из диаграмм видно, что, как и следовало ожидать, наибольшее значение для растворимости соединений в воде имеют электростатические факторы и липофильность фрагментов молекулы. Также можно отметить, что процентное соотношение вкладов параметров в растворимость с увеличением температуры практически не изменяется.

На втором этапе настоящей работы была сделана попытка проанализировать влияние заместителей в ароматическом кольце с изменением температуры для различных производных бензола (выборка 2). Статистические характеристики полученных моделей М-20, М-25, М-30, М-40 вполне приемлемы (табл. 1).

Из анализа QSPR моделей были определены вклады $-Cl$, $-OH$, $-COOH$, $-NO_2$ групп (рис. 4), связанных с бензойным кольцом. Как видно из рис.4 влияние анализируемых заместителей незначительно изменяется в данном диапазоне температур. Для более детального анализа в качестве примера мы приводим «лепестковые» диаграммы вкладов заместителей только при одной из температур $-40^{\circ}C$ (рис. 5). Такие диаграммы позволяют оценить не только вклад заместителя в изучаемое свойство, но и взаимное влияние заместителей.

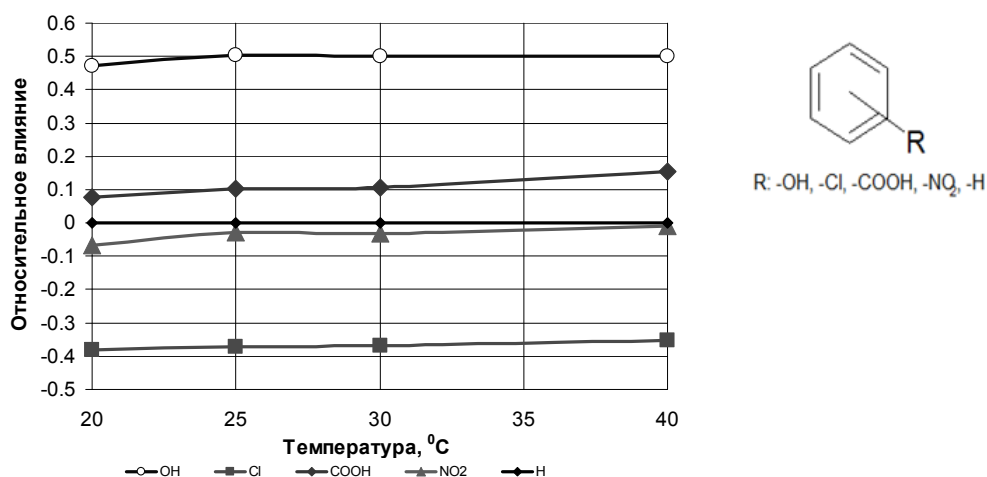


Рис. 4. Изменение вкладов заместителей с ростом температуры.

Из диаграмм видно, что группы, которые являются донорами водородных связей ($-OH$, $-COOH$), способствуют увеличению растворимости. Что касается электронных эффектов и полярности заместителей, то они не оказывают однозначного влияния на исследуемое свойство. Среди анализируемых заместителей максимальным положительным вкладом в растворимость обладает $-OH$ группа, а максимальным отрицательным – $-Cl$. Из рис. 5 однозначно следует, что влияние заместителей на растворимость неаддитивно. Вклады заместителей существенно зависят не только от их природы, но и от их взаимного положения. Отметим также, что в зависимости от характера замещения значительно изменяется вклад $C-H$ групп.

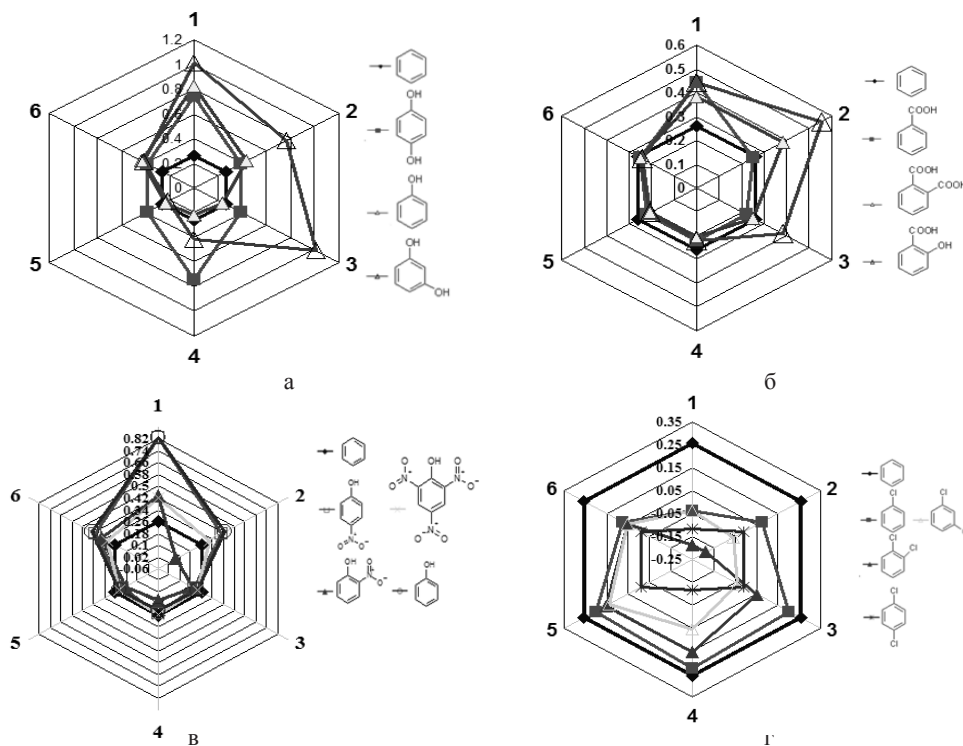


Рис. 5. Диаграммы влияния вкладов заместителей:
 а) -ОН, б) -СООН, в) -ОН, -NO₂, г) -Сl на растворимость исследуемых соединений.

На следующем этапе работы была построена модель, в которой температура была задействована непосредственно как параметр (выборка 3). Построение модели В (см. табл. 1), в которой температура была включена как параметр в виде $e^{-\frac{1}{T}}$, в соответствии с уравнением Шредера [10], дало нам возможность вычислить вклад температуры в растворимость.

Вид зависимости предсказанных и наблюдаемых значений растворимости по модели В представлен на рис. 6, а относительное влияние некоторых физико-химических факторов на изучаемое свойство в рамках этой модели – на рис. 7.

Как и следовало ожидать (см. рис. 7), температура значительно влияет на растворимость веществ в воде (вклад 30 %). Также, как и для моделей К20, К30, К40, для модели В наибольшее влияние оказывают электростатические факторы и липофильность (31 % и 12 % соответственно). Кроме того, заметное влияние оказывают ван-дер-ваальсовы взаимодействия (15 %).

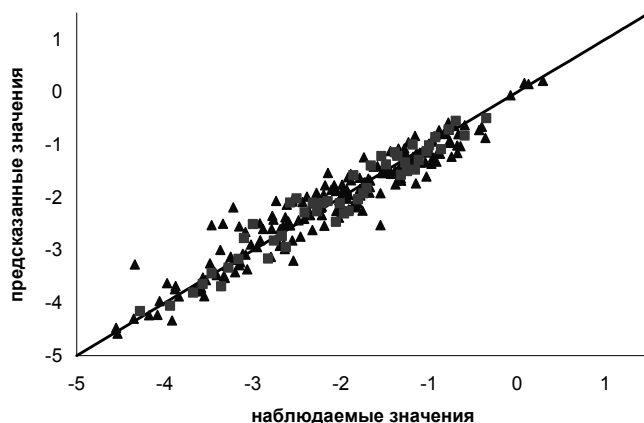


Рис. 6. Зависимость предсказанных и наблюдаемых значений $LgSw$ для модели В, где треугольниками отмечены данные из обучающей выборки, а квадратами отмечены данные из тестовой выборки.

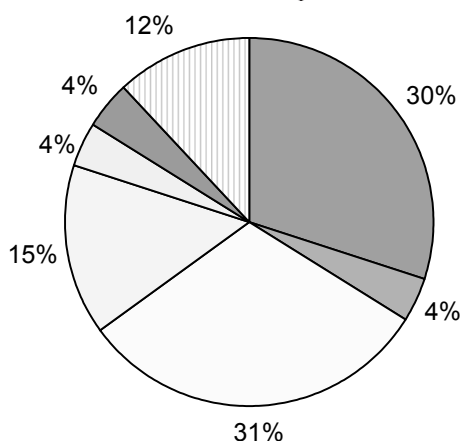


Рис.7 Относительное влияние некоторых физико-химических факторов на $LgSw$ для модели В, где А – температура, Р – поляризуемость, С – электростатика, D – ван-дер-ваальсовое притяжение, Е – способность атома выступать в роли донора или акцептора водородной связи, F – тип атома, G – липофильность.

Заключение

Таким образом, были получены адекватные 2D-QSPR модели, описывающие влияние структурных факторов на растворимость в воде разнообразных органических соединений при различных температурах. Основное влияние на растворимость в воде оказывают частичные заряды на атомах и липофильность фрагментов молекулы. Показано, что вклады структурных факторов мало меняются в диапазоне температур 20 – 40 °С .

Для производных бензола вклад заместителей в водную растворимость существенно неаддитивен. Направление влияния заместителя на растворимость соединения, главным образом, определяется способностью заместителя образовывать водородные связи. Нет однозначного влияния электронных эффектов заместителей в бензольном ядре на их вклад в растворимость.

Для QSPR модели, что непосредственно включает температуру как параметр, обнаружено, что ее относительный вклад в растворимость весьма существенен (примерно 30 %).

Данная работа является первым этапом широкомасштабных исследований. Авторы планируют существенно увеличить объем и структурное разнообразие обучающих выборок и, по возможности, расширить температурный диапазон. Конечным результатом этого цикла исследований, будет экспертная система для внеэкспериментального скрининга растворимости в воде различных органических соединений.

Список использованной литературы

1. Аксельруд Г. А., Молчанов А. Д. Растворение твердых веществ – М.: Химия, 1977. – 272 с.
2. Yalkovsky S. H., Yan H., Jain P. Handbook of aqueous solubility data – second ed. – Boca Raton – 2010. – P. 1608.
3. Kuz'min V. E., Artemenko A. G., Muratov E. N. Hierarchical QSAR Technology on the Basis of Simplex Representation of Molecular Structure // J. Comput.-Aided Mol. Des. – 2008. – Vol. 22 – P. 403–421
4. Kuz'min V. E., Muratov E. N., Artemenko E. N., Gorb L. G., Qasim M., Leszczynski J. The Effect of Nitroaromatics' Composition on Their Toxicity in Vivo: Novel, Efficient Non-Additive 1D QSAR Analysis // Chemosphere – 2008 – Vol. 72 – P. 1373–1380.
5. Jolly W. L., Perry W. B. Estimation of atomic charges by an electronegativity equalization procedure calibration with core binding energies // J. Am. Chem. Soc. – 1973. – Vol. 95. – P. 5442-5450.
6. Wang R., Fu Y., Lai L. A new atom-additive method for calculating partition coefficients // J. Chem. Inf. Comput. Sci. -1997.-Vol.37.-P.615-621.
7. Иоффе Б. В. Рефрактометрические методы химии – Ленинград: Химия, 1983. – 350 с.
8. Rännar S., Lindgren F., Geladi P., Wold S. A PLS Kernel Algorithm for Data Sets with Many Variables and Fewer objects. Part 1: Theory and Algorithm // J. Chemometrics. – 1994. – Vol. 8. – P. 111-125.
9. Rogers D., Hopfinger A. J. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships // J. Chem. Inf. Comp. Sci. – 1994. – Vol. 34 – P. 854-866.
10. Уэйлес С. Фазовые равновесия в химической технологии : Т. 2. Пер. с англ. – М.: Мир, 1989. – 360 с.

**К. О. Кліменко¹, Л. М. Огніченко², В. Є. Кузьмін^{1,2}, А. Г. Артеменко²,
Л. Г. Горб³**

¹ Одеський національний університет імені І. І. Мечникова, Україна, Одеса, 65052, вул. Дворянська 2

² Фізико-хімічний інститут ім. О. В. Богатського НАН України, Україна, Люстдорфська дорога 86, Одеса, 65080, e-mail: ogni@ukr.net

³ Інститут молекулярної біології та генетики НАН України, Україна, Київ, 03143, вул. Заболотного 150, e-mail: lgorb@icnanotox.org

QSPR АНАЛІЗ ТЕМПЕРАТУРНОЇ ЗАЛЕЖНОСТІ РОЗЧИННОСТІ ОРГАНІЧНИХ СПЛУК У ВОДІ

Резюме

Створено 2D-QSPR моделі для прогнозування розчинності органічних сполук у воді при різних температурах. Проаналізовано вклади у розчинність різних структурних факторів, зокрема, визначено вклад замісників у бензеновому кільці. Створена адекватна модель, яка дає можливість оцінити вплив температури на розчинності різноманітних сполук у воді.

Ключові слова: водна розчинність, молекулярні симплекси, QSPR, температурна залежність розчинності.

**К. А. Klimenko¹, L. N. Ognichenko², V. E. Kuz'min^{1,2}, A. G. Artemenko²,
L. Gorb²**

QSPR ANALYSIS OF TEMPERATURE RELATED AQUEOUS SOLUBILITY OF ORGANIC COMPOUNDS

¹ I. I. Mechnikov National University, Dvorianskaya 2, Odessa 65026, Ukraine.

² Laboratory of Theoretical Chemistry, Department of Molecular Structure, A. V. Bogatsky Physical-Chemical Institute National Academy of Sciences of Ukraine, Lustdorfskaya Doroga 86, Odessa 65080, Ukraine, e-mail: ognichenko@mail.ru

³ Institute of Molecular Biology and Genetics National Academy of Sciences of Ukraine, 150, Zabolotnogo Str, Kyiv – 143, 03680, Ukraine, e-mail: lgorb@icnanotox.org

Summary

2D-QSPR models were developed in order to predict aqueous solubility of organic compounds at different temperatures. Impact of various structure features was analyzed and benzene residues contribution to solubility was defined. A reliable model which allows to assess temperature influence on forming compounds solubility was developed.

Key words: aqueous solubility, molecular simplex, QSPR, temperature dependence of solubility

Стаття надійшла до редакції 12.03.12