

Вибір дескрипторів для побудови кількісних залежностей «структура — властивість» із використанням генетичного алгоритму і модифікованого методу повного перебору

В.В. Марков^{1*}, О.П. Бойченко², Л.П. Логінова¹

¹ Харківський національний університет імені В.Н. Каразіна
пл. Свободи, 4, Харків, 61022, Україна

² Відділ аналітичної біохімії, Фармацевтичний центр університету,
9713, Гронінген, Нідерланди

Резюме. У роботі на основі обраних молекулярних дескрипторів побудовано кількісні залежності «структура — константа розподілу органічних речовин у двофазних системах «хлороформ — газ», «хлороформ — вода» та «1-октанол — вода». Для пошуку дескрипторів використано методи повного перебору і генетичний алгоритм. Проведено оптимізацію параметрів алгоритмів, що дало змогу скоротити час розрахунків. Побудовано моделі з використанням і без використання групи дескрипторів молекулярні властивості, що розраховуються на основі емпіричних залежностей. Побудовані моделі описують експериментальні дані з коефіцієнтом кореляції не нижче 0,92 та мають високу прогностичну здатність.

Ключові слова: залежності «структура — властивість», розподіл, дескриптор, генетичний алгоритм.

Вступ. Кількісні залежності «структура — властивість» (quantitative — structure property relationships, QSPR) базуються на припущенні, що структура молекули (геометричні, стеричні, електронні властивості) мають особливості, які пов'язані з фізичними, хімічними або біологічними властивостями і можуть бути представлені за допомогою одного або декількох числових дескрипторів. Спроби побудувати кількісні залежності «структура — біологічна активність» (quantitative — structure activity relationships, QSAR) почали з'являтися понад сто років тому [1]. Засновником сучасного QSAR, який уперше систематично дослідив можливість моделювання як біологічної активності, так і хімічних властивостей молекул із використанням дескрипторів, вважають Ганча [2–5].

Розрізняють два основні підходи до вибору

дескрипторів для побудови QSPR. Перший, механістичний, пов'язаний з апіорним вибором дескрипторів і базується на відомих даних про властивості речовини та найбільш важливі структурні особливості досліджуваних молекул. У більшості випадків вибір проводять з невеликого набору дескрипторів (наприклад, дескрипторами можуть виступати константа розподілу в системі 1-октанол — вода, полярність, молярна рефракція і т.д.). Другий підхід, статистичний, базується на ідеї, що вибір дескрипторів є результатом роботи певного алгоритму. Для побудови моделей і вибору найбільш адекватних дескрипторів (з декількох тисяч відомих) використовують хемометричні методи [6–8]. До переваг статистичного підходу можна віднести відсутність суб'єктивізму при відборі дескрипторів, а також можливість розрахувати дескриптори для ще не синтезованих молекул, а отже, і передбачити їх фізико-хімічні властивості або біологічну активність.

Використані методи дають змогу передбачити велику кількість фізико-хімічних влас-

* Corresponding author.

Tel.: +38057-7075248

E-mail address: markov.vadim@gmail.com

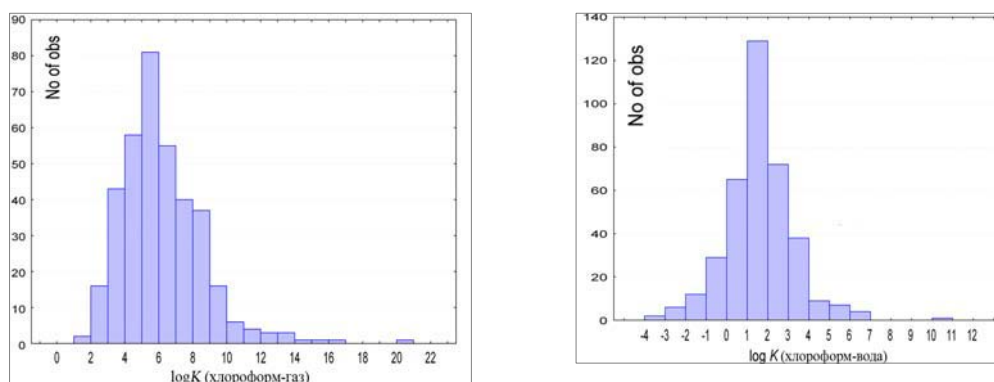


Рис. 1. Гістограма розподілу логарифмів констант розподілу тестових речовин у системі «хлороформ — газ» і «хлороформ — вода».

тивостей речовин, у тому числі ще не синтезованих, повністю абстрагувавшись від природи досліджуваного параметру. Застосованість методу не обмежується лише коефіцієнтами розподілу, на відміну від уже відомих адитивних схем розрахунку. Єдиною умовою є наявність набору експериментальних даних для побудови адекватної кореляції.

Характеристики розподілу органічних речовин у різних двофазних системах є фундаментальними фізико-хімічними параметрами, необхідними для розробки методів хімічного аналізу, передбачення біологічної активності, проектування технологічних процесів тощо. Проте їх експериментальне визначення досить часто утруднюється через відсутність достатньої кількості реагенту або його недостатню чистоту, трудомісткість і тривалість стандартної «shake flask» процедури [9]. Крім того, важливими є дослідження розподілу речовин між водною фазою і нанорозмірними дисперсіями, наприклад, міцелярними розчинами поверхнево-активних речовин, в яких неможливо фізично розділити водну фазу і міцелярну псевдофазу [10-13].

У цій роботі за допомогою генетичного алгоритму й оптимізованого методу повного перебору ми обрали теоретичні дескриптори для побудови лінійних моделей, що описують характеристики розподілу органічних речовин у різних системах.

Тестові речовини і константи їх розподілу у двофазних системах. Як тестовий набір використано дані про константи розподілу в системах «хлороформ — газ» і «хлороформ — вода», наведені в роботі [14]. Для 287 із 368 речо-

вин із цього тестового набору в літературі було знайдено дані про розподіл у системі «1-октанол — вода» [4, с. 15-27]. На рис. 1 наведено гістограми розподілу логарифмів констант розподілу тестових речовин у системах «хлороформ — газ» і «хлороформ — вода».

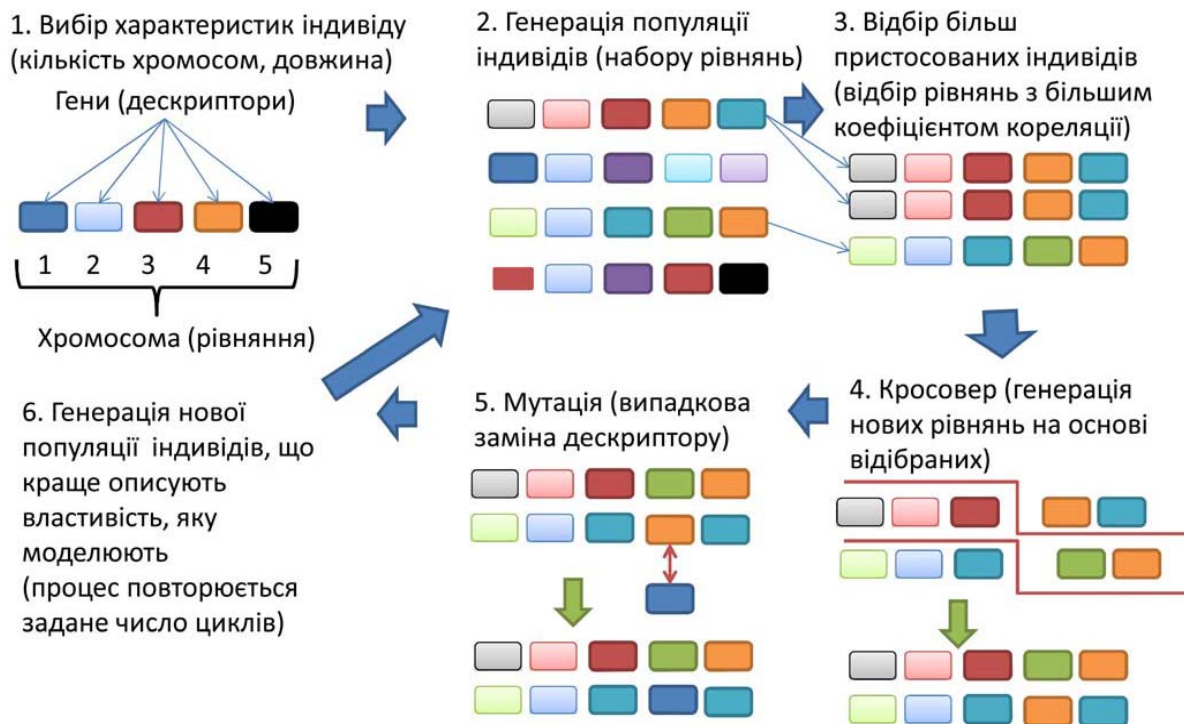
Діапазон значень констант розподілу дуже широкий і становить понад 10 порядків для обох систем, що говорить про репрезентативність обраного набору речовин для пошуку дескрипторів.

Програмне забезпечення і бази даних. Отримані дані оброблялися в середовищі Excel. Оптимізація геометрії молекул тестових сполук проводилася в програмному пакеті ChemOffice 2005 (<http://www.cambridgesoft.com/software/ChemOffice/>) з використанням модуля GAMESS (<http://www.msg.ameslab.gov/gamess/>) напівемпіричним методом PM3. Реалізація генетичного алгоритму проводилася на мові програмування Python 2 (<http://www.python.org>) з використанням модулів SciPy (<http://www.scipy.org>) і Pyevolve (<http://pyevolve.sourceforge.net/>). Для розрахунку дескрипторів використовувалася програма TALETE Dragon ver. 5.5, яка дає змогу розрахувати 3224 дескриптори для кожної молекули.

Теоретична частина.

Генетичний алгоритм. Розвиток генетичних алгоритмів почався в середині 50-х років ХХ ст. з робіт біолога Баріцеллі [28, 29] і програміста Фрайзера [30], що застосували комп'ютерні моделі для вивчення процесів еволюції та селекції. З'ясувалося, що завдання еволюції — пошук пристосованих до певного середовища організмів серед величезної кіль-

Схема роботи генетичного алгоритму



кості можливих комбінацій — відповідає одному з найважливіших технологічних і наукових завдань — пошуку оптимального рішення.

У генетичних алгоритмах використовують біологічну термінологію. Так, хромосома — це ланцюжок, що складається із впорядкованої послідовності генів. Коефіцієнти рівняння визначаються на етапі перевірки якості рівняння. Гени можуть розташовуватися в різних позиціях рядків, які називають локусами хромосоми, і приймати різні значення — алелі. Генотип — це повний набір хромосом, що характеризує окремих індивід. Була використана найпростіша модель з використанням лише однієї хромосоми, яка відповідає набору дескрипторів QSPR. Пристосованість індивіда до середовища характеризують функцією пристосованості (у нашій роботі використано якість опису експериментальних даних згенерованим рівнянням). Популяція — це певна кількість одночасно існуючих індивідів незалежно від значення функції пристосованості в кожного з них, тобто набір згенерованих рівнянь QSPR.

Першим кроком роботи генетичного алгоритму є генерування випадковим чином початкової популяції індивідів, що являє собою набір з деякої кількості випадково згенерованих

рівнянь. Розмір популяції задається спочатку і не змінюється протягом роботи алгоритму, типовим значенням для цієї задачі є кілька сотень особин. На наступному кроці роботи алгоритму відбувається відбір більш пристосованих індивідів. Найпростіший пропорційний відбір — рулетка (roulette-wheel selection) — відбирає індивідів за допомогою n «запусків» рулетки. Колесо рулетки містить по одному сектору для кожного індивіда популяції. Розмір i -ого сектору пропорційний величині $P(i)$, що дорівнює відношенню пристосованості особи до сумарної пристосованості популяції. При такому відборі члени популяції з більш високою пристосованістю будуть вибиратися частіше, ніж особи з низькою пристосованістю. На кроці кросовера n індивідів випадковим чином розбиваються на $n/2$ пар, які набувають функцію батьків. Вибирається точка розриву, і кожна батьківська хромосома розривається на два не обов'язково рівні сегменти по цій точці. Потім відповідні сегменти хромосом від різних батьків склеюються і виходять два генотипи особин-нащадків. У нашій задачі два обрані набори дескрипторів обмінюються своїми частинами. Наприклад, рівняння з дескрипторами (A, B, C) і (D, E, F) можуть дати нащадків (A, E, F) і (D, B, C). Після завер-

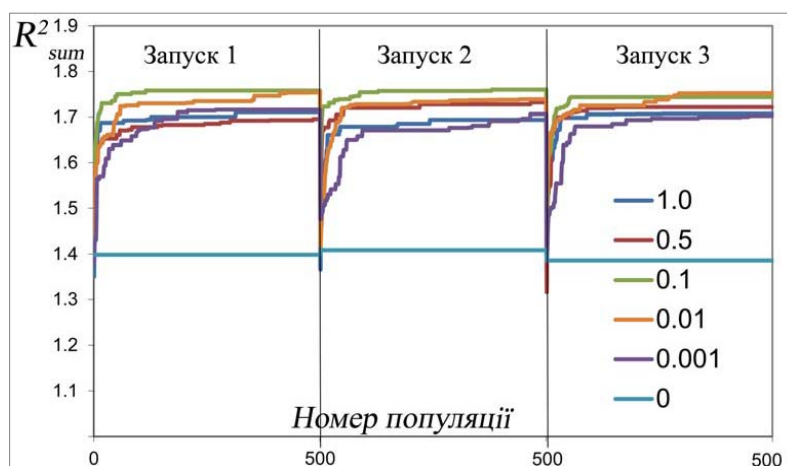


Рис. 2. Залежності значення функції пристосованості для найбільш життєздатного індивіду певної генерації від номера генерації для різних імовірностей мутацій і розміру популяції 500 індивідів.

шення кросовера особини популяції піддаються мутації. З певною заданою ймовірністю кожен алель хромосоми (дескриптор у цій задачі) змінюється на інший, випадково обраний. Після кросовера та мутації нова популяція замінює стару і завершується цикл одного покоління. Робота генетичного алгоритму продовжується до виконання заданого числа поколінь чи іншого критерію зупинки.

Необхідно відзначити, що, крім описаних вище і використаних нами операторів відбору, кросовера і мутації, в літературі використовується також цілий ряд інших, наприклад, турнірний відбір, елітні методи відбору і т.д.

Оптимізований метод перебору. Перебір усіх можливих комбінацій дескрипторів для пошуку мультипараметричних моделей є найбільш загальним підходом, проте обчислювальна складність повного перебору всіх можливих поєднань зростає експоненційно зі збільшенням кількості дескрипторів у наборі, що в більшості випадків приводить до необхідності застосовувати занадто дорогі обчислювальні потужності. Суть оптимізації алгоритму повного перебору полягала в покроковому збільшенні кількості дескрипторів у наборі на підставі відсіювання напевне неоптимальних сполучень дескрипторів. Для переходу від набору з n дескрипторів до $(n+1)$ алгоритм виконував такі кроки: (1) формування всіх можливих сполучень з n дескрипторів; (2) сортування отриманого списку сполучень відповідно до якості опису експериментальних даних [як такий критерій використовували коефіцієнт ко-

реляції між експериментальними й передбаченими значеннями]; (3) відбір m сполучень, що описують експериментальні дані з найбільшою точністю [нами обрано m рівне 200, подальше збільшення не приводило до зміни результатів розрахунків]; (4) формування всіх можливих сполучень набору, отриманого в п. 3, із дескрипторами з повного списку. Отриманий у результаті список сполучень можливо знову ввести до першого пункту алгоритму.

Складність такого алгоритму зростає лінійно зі збільшенням числа дескрипторів.

Результати й обговорення.

Оптимізація параметрів генетично алгоритму. Критерій пристосованості та зупинки роботи генетичного алгоритму. Як функції пристосованості кожного індивіда, що представляє лінійне мультипараметричне рівняння з наперед заданим числом параметрів, використовували суму квадратів коефіцієнтів кореляції між передбаченими й експериментальними значеннями констант розподілу в системах «хлороформ — газ» і «хлороформ — вода» або «хлороформ — вода» та «1-октанол — вода». Критерієм зупинки роботи генетичного алгоритму був номер популяції.

Вибір оптимальної імовірності мутацій. На рис. 2 наведено результати трьох паралельних запусків генетичного алгоритму для різних значень імовірності мутацій (0, 0,001, 0,01, 0,1, 0,5, 1) та однакового значення числа індивідів (500). Кількість популяцій задавали рівною 500, кожен індивід мав одну хромосому з 10 локусами. Для всіх значень імовірності

мутацій, крім нуля, залежності виглядають як відносно гладкі лінії, що наближаються до максимуму. При цьому спостерігається немотонна залежність максимального значення функції пристосованості від імовірності мутацій з максимумом при значенні імовірності мутацій 0,1.

Оптимізація розміру популяції. Для визначення достатнього розміру популяції досліджували еволюції з імовірністю мутацій 0,01 та розміром 2000, 500, 200, 100, 50, 20 і 10 індивідів протягом 500 поколінь. Геном кожного індивіда складався з однієї хромосоми з 10 локусами. На залежностях функції пристосованості індивіда від номера генерації спостерігається зростання, яке стає більш різким при збільшенні розміру популяції. Беручи до уваги обчислювальні можливості та отримані залежності як оптимальні параметри роботи генетичного алгоритму, обрано розмір популяції 2000 та ймовірність мутацій 0,1.

Вибір дескрипторів. Для побудови кількісних залежностей «структура — властивість» часто використовують не тільки теоретичні дескриптори, які розраховують лише на підставі структури молекули, а й так звані молекулярні властивості, наприклад, логарифм розподілу в системі «1-октанол — вода, розчинність і т.д. У більшості випадків молекулярні властивості розраховують із використанням багатопараметричних емпіричних рівнянь, коефіцієнти яких отримані на основі експериментальних даних. А отже, такі дескриптори вже не можуть вважатися теоретичними. Тому в нашій роботі за допомогою генетичного алгоритму й оптимізованого методу повного перебору ми спробували побудувати моделі, що містять емпіричні молекулярні дескриптори, а також моделі, у яких вони відсутні.

Моделі з використанням емпіричних дескрипторів (молекулярних властивостей). Для пошуку дескрипторів залежності «константа розподілу — структура речовини» використовували результати, отримані після паралельних запусків генетичного алгоритму з оптимальними параметрами ймовірності мутацій і розміру популяції (геном кожного індивіду складався з однієї хромосоми із 7 локусами), а також набори моделей, отримані після запуску оптимізованого алгоритму повного перебо-

ру. На підставі отриманих результатів відібрано 6 дескрипторів, які дали змогу з високою точністю описати розподіл речовин у трьох системах. Нижче наведено отримані рівняння зі значеннями коефіцієнтів кореляції між передбаченими й експериментальними значеннями констант розподілу (у дужках подано значення стандартного відхилення для коефіцієнтів):

система «хлороформ — газ»

$$\log K = -0,66(0,04)A\text{LOGP} + 0,87(0,04)X1\text{sol} + 0,25(0,08)(O-057) + 0,38(0,05)(H-050) - 0,41(0,17)n\text{HBonds} + 0,078(0,006)AMR + 0,32(0,10); R = 0,96;$$

система «хлороформ — вода»

$$\log K = 1,06(0,04)A\text{LOGP} + 0,08(0,04)X1\text{sol} - 1,03(0,08)(O-057) - 0,74(0,05)(H-050) + 0,96(0,17)n\text{HBonds} + 0,010(0,006)AMR + 0,20(0,08); R = 0,92;$$

система «1-октанол — вода»

$$\log K = 1,04(0,04)A\text{LOGP} - 0,16(0,08)X1\text{sol} + 0,12(0,06)(O-057) - 0,12(0,03)(H-050) + 0,27(0,11)n\text{HBonds} + 0,028(0,01)AMR - 0,22(0,08); R = 0,95$$

Для опису розподілу речовин у двофазних системах найбільш значущим виявився дескриптор ALOGP (константа розподілу в системі «1-октанол — вода»), що розраховується на підставі простих вкладів атомів в коефіцієнт розподілу з використанням емпіричного рівняння, коефіцієнти якого розраховані на основі експериментальних даних про розподіл 893 речовин. Також у модель були включені дескриптор X1sol (solvation connectivity index chi-1), що описує дисперсійні взаємодії в розчині; O-057, що представляє кількість спиртових груп у молекулі (ураховуються спиртові, фенольні та енольні атоми кисню); H-050 — кількість атомів водню, пов'язаних з гетероатомами; nHBonds — кількість внутрішньомолекулярних водневих зв'язків і AMR — молярна рефракція Ghose-Crippen, яка розраховується аналогічно ALOGP на основі емпіричного рівняння.

Для валідації моделі розраховані логарифми констант розподілу в системі «1-октанол — вода» для 13 речовин, які досліджувались нами в попередній роботі [31] та не ввійшли до набору речовин, що використовували для побудови моделі: 1,3,5-триметилбензол, 1,4-диметилбензол, 1-етил-4-нітробензол, 2,3-дихлор-

Кореляційна матриця 14 дескрипторів, розрахованих для 368 речовин

	H-046	MAXDP	RBF	nX	nROH	X1sol	H-050	F01[C-C]	Jhetm	nCar	H-052	N-075	O-057	nHDon
H-046	1,00													
MAXDP	0,05	1,00												
RBF	0,35	0,33	1,00											
nX	-0,18	-0,08	-0,25	1,00										
nROH	0,34	0,27	0,25	-0,09	1,00									
X1sol	0,19	0,45	0,10	-0,08	0,24	1,00								
H-050	0,13	0,30	0,01	-0,16	0,40	0,19	1,00							
F01[C-C]	0,36	0,33	0,05	-0,27	0,29	0,81	0,24	1,00						
Jhetm	-0,06	-0,09	-0,08	0,40	-0,07	0,08	-0,09	-0,16	1,00					
nCar	-0,34	0,09	-0,25	-0,15	-0,26	0,39	0,03	0,56	-0,11	1,00				
H-052	0,18	0,12	0,29	-0,11	0,36	0,12	0,08	0,11	-0,04	-0,35	1,00			
N-075	-0,16	-0,21	-0,16	-0,05	-0,21	-0,06	-0,25	-0,23	-0,03	0,07	-0,11	1,00		
O-057	-0,04	0,28	0,07	-0,02	0,27	0,03	0,28	0,11	-0,04	0,20	-0,19	-0,21	1,00	
nHDon	0,13	0,30	0,01	-0,16	0,40	0,19	1,00	0,24	-0,09	0,03	0,08	-0,25	0,28	1,00

фенол, 2,5-дихлорфенол, 2,6-дихлорфенол, 3,4-дихлорфенол, 3,5-дихлорфенол, 2-нітроанізол, етилбензол, пентилбензол, гексилбензол, флуорен. Отримане значення коефіцієнта кореляції між передбаченими й експериментальними значеннями логарифмів констант розподілу склало 0,96, тангенс кута нахилу прямої значимо не відрізнявся від одиниці, а вільний член від нуля, що говорить про високу передбачувальну спроможність побудованої моделі.

Моделі, які не містять емпіричних дескрипторів. При побудові моделей з використанням тільки теоретичних дескрипторів, для того, щоб оцінити вплив наявності водної фази на набір отриманих дескрипторів, ми розділили дані про розподіл тестових речовин на дві групи: (i) хлороформ — газ, хлороформ — вода; (ii) хлороформ — вода, 1-октанол — вода.

Після застосування генетичного алгоритму й методу повного перебору отримано два набори з 9 дескрипторів, що описують розподілу речовин в системах, які належать до групи (i) та (ii). При цьому виявилось, що тільки 14 із них є різними. Застосування кожного з рівнянь для опису розподілу речовин з іншої групи («хлороформ — газ» або «1-октанол — вода») призводить до зниження коефіцієнтів кореляції, а отже, необхідне об'єднання набору

дескрипторів. У табл. 1 наведено кореляційну матрицю для 14 відібраних дескрипторів.

Аналіз отриманої кореляційної матриці вказує на присутність двох дескрипторів з коефіцієнтом кореляції, рівним одиниці, — H-050 (кількість атомів водню, пов'язаних з гетероатомами) і nHDon (кількість внутрішньомолекулярних водневих зв'язків). Для подальшого використання обрано nHDon. Високе значення коефіцієнта кореляції також спостерігалось для дескрипторів F01[C-C] (частка зв'язків вуглець-вуглець) і X1sol. Для подальшого використання обрано X1sol. У побудованих моделях, що містять 12 дескрипторів, коефіцієнт при дескрипторі Jhetm (індекс Балабана для зваженої за масою матриці відстаней) виявився незначущим. Тому остаточна модель містить 11 дескрипторів. Нижче наведено рівняння, отримані для трьох досліджених систем:

система «хлороформ — газ»

$$\log K = -0,078(0,015)nCar - 0,064(0,01)(H-046) - 0,083(0,019)(H-052) + 0,23(0,05)(N-075) - 0,28(0,09)(O-057) + 0,49(0,05)nHDon + 0,28(0,03)MAXDP - 2,54(0,63)RBF - 0,64(0,05)nX - 0,01(0,09)nROH + 1,27(0,03)X1Sol + 0,56(0,11); R = 0,97$$

система «хлороформ — вода»

$$\log K = 0,31(0,015)nCar + 0,23(0,01)(H-046) + 0,13(0,02)(H-052) - 0,51(0,05)(N-075) -$$

0,60(0,09)(O-057)-0,89(0,05)nHDon-
0,18(0,03)MAXDP+3,69(0,63)RBF+
0,52(0,05)nX-0,36(0,09)nROH+
0,10(0,02)X1Sol+0,44(0,10); R=0,92
система «1-октанол — вода»
logK=0,28(0,017)nCar+0,19(0,015)(H-046)+
0,12(0,018)(H-052)-0,50(0,04)(N-075)+
0,36(0,07)(O-057)-0,29(0,04)nHDon-
0,24(0,03)MAXDP+3,52(0,54)RBF+
0,70(0,06)(nX)-0,18(0,07)nROH+
0,044(0,03)X1Sol+0,064(0,09); R=0,95

Більшість із дескрипторів у моделях досить прості для розуміння та є цілочисловими. Вони представляють кількість гідроксильних груп (nROH), кількість ароматичних атомів вуглецю (nCar), кількість атомів водню без віцинальних галогензаміщених атомів (H-046), кількість атомів водню з однозаміщеним віцинальним атомом (H-052), число фрагментів R-N-R і R-N-X (N-075), загальну кількість фенольних, енольних і карбоксильних гідроксигруп (O-057), кількість донорів водневих зв'язків (nHDon), кількість атомів галогенів у молекулі (nX). Фізичний зміст інших дескрипторів можна інтерпретувати як інтенсивність дисперсійних взаємодій у розчині для дескриптора X1sol, який відноситься до групи індексів зв'язності. Конституційний дескриптор RBF дорівнює відсотку зв'язків, що допускають обертання, і вказує на можливість утворення конформацій.

MAXDP відноситься до групи електротопологічних дескрипторів і пов'язаний з максимальною по молекулі щільністю π -електронів.

Як і у випадку валідації моделі, що містила емпіричні дескриптори, ми передбачили логарифми константи розподілу для 13 речовин, які не використовували для побудови моделі. При цьому значення коефіцієнта кореляції між передбаченими й експериментальними значеннями логарифмів констант розподілу склав 0,96, тангенс кута нахилу прямої не відрізнявся від одиниці, а вільний член від нуля.

Висновки. На підставі результатів роботи генетичного алгоритму й оптимізованого методу повного перебору нами із 3224 відомих дескрипторів обрано набори, які задовільно описують і передбачають розподіл різних класів органічних речовин у різних двофазних системах. Обрані набори дескрипторів можуть стати універсальними при моделюванні різних фізико-хімічних процесів, пов'язаних з розподілом у двофазних системах. Подальша робота буде пов'язана з описом розподілу речовин у системах «міцелярна псевдофаза — вода», а також розробкою більш інформативних неемпіричних дескрипторів, які дадуть змогу зменшити число параметрів у моделях.

Надійшла в редакцію 17.01.2012 р.

The selection of descriptors for QSAR (quantitative structure-activity relationships) building using genetic algorithm and modified multiple-searching method

V.V. Markov¹, A.P. Boichenko², L.P. Loginova¹

¹ V.N. Karazin Kharkiv National University
4, Svobody Sq., 61022, Kharkiv, Ukraine

² Analytical Biochemistry, University Centre for Pharmacy
1, A. Deusinglaan, 9713, AV Groningen, The Netherlands

Summary. In this work on the basis of selected descriptors the quantitative relationships structure-partition constant in two phase systems chloroform — gas, chloroform — water, 1-octanol — water has been constructed. The search of descriptors has been done by using the exhaustive search and genetic algorithm. The optimization of algorithm parameters was done that allow to reduce the time of calculations. The relationships were constructed with and without using the group of molecular properties descriptors which are calculated on the basis of empirical models. The constructed relationships have high goodness-of-fit (the correlation coefficient is not lower than 0.92) and high predictive capacity.

Keywords: structure-property relationships, partition, descriptor, genetic algorithm.

Перелік літератури

1. Selassie C., Verma R.P., Abraham D.J. History of Quantitative Structure-Activity Relationships // Burger's Medicinal Chemistry and Drug Discovery: John Wiley & Sons, Inc., 2003.

2. Hansch C., Maloney P.P., Fujita T., Muir R.M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients // Nature. — 1962. — Vol. 194, No. 4824. — P. 178-180.

3. *Hansch C., Muir R.M., Fujita T., Maloney P.P., Geiger F., Streich M.* The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients // *Journal of the American Chemical Society.* — 1963. — Vol. 85, No. 18. — P. 2817-2824.
4. *Fujita T., Iwasa J., Hansch C.* A new substituent constant, π , derived from partition coefficients // *Journal of the American Chemical Society.* — 1964. — Vol. 86, No. 23. — P. 5175-5180.
5. *Hansch C., Fujita T.* ρ - σ - π analysis. A method for the correlation of biological activity and chemical structure // *Journal of the American Chemical Society.* — 1964. — Vol. 86, No. 8. — P. 1616-1626.
6. *Myint K.Z., Xie X.Q.* Recent advances in fragment-based QSAR and multi-dimensional QSAR methods // *International Journal of Molecular Sciences.* — 2010. — Vol. 11, No. 10. — P. 3846-3866.
7. *Warr W.A.* Some Trends in Chem(o)informatics // *Methods in molecular biology (Clifton, N.J.).* — 2011. — Vol. 672, No. — P. 1-37.
8. *Taboureaux O.* Methods for building quantitative structure — activity relationship (QSAR) descriptors and predictive models for computer-aided design of antimicrobial peptides // *Methods in Molecular Biology (Clifton, N.J.).* — 2010. — Vol. 618, No. — P. 77-86.
9. *Boichenko A. P., Loginova L.P., Kulikov A.U., Iwashchenko A.L., Galat M.N.* Retention-hydrophobicity relationships for polyaromatic hydrocarbons and 4-hydroxybenzoic acid esters by micellar liquid chromatography data // *Ukr. Bioorg. Acta.* — 2007. — Vol. 2, No. — P. 3-16.
10. *Boichenko A., Dung L., Loginova L.* Solubilization of aliphatic carboxylic acids (C3-C6) by sodium dodecyl sulfate and brij 35 micellar pseudophases // *Journal of Solution Chemistry.* — 2011. — Vol. 40, No. 6. — P. 968-979.
11. *Zhao B., Zhu L., Li W., Chen B.* Solubilization and biodegradation of phenanthrene in mixed anionic-nonionic surfactant solutions // *Chemosphere.* — 2005. — Vol. 58, No. 1. — P. 33-40.
12. *Mohamed A., Mahfoodh A.-S.M.* Solubilization of naphthalene and pyrene by sodium dodecyl sulfate (SDS) and polyoxyethylene sorbitan monooleate (Tween 80) mixed micelles // *Colloids and Surfaces A: Physicochemical and Engineering Aspects.* — 2006. — Vol. 287, No. 1-3. — P. 44-50.
13. *Atanackovi^a M., Po a M., Heinle H., Gojkovi^a-Bukarica L., Cveji^a J.* Solubilization of resveratrol in micellar solutions of different bile acids // *Colloids and Surfaces B: Biointerfaces.* — 2009. — Vol. 72, No. 1. — P. 148-154.
14. *Sprunger L.M., Achi S.S., Acree Jr W.E., Abraham M.H., Leo A.J., Hoekman D.* Correlation and prediction of solute transfer to chloroalkanes from both water and the gas phase // *Fluid Phase Equilibria.* — 2009. — Vol. 281, No. 2. — P. 144-162.
15. *Leo A., Hansch C.* Substituent Constants for Correlation Analysis in Chemistry and Biology. — New York: Wiley, 1979.
16. *Hansch C., Anderson S.M.* The effect of intramolecular hydrophobic bonding on partition coefficients // *Journal of Organic Chemistry.* — 1967. — Vol. 32, No. 8. — P. 2583-2586.
17. *Sangster J.* Octanol-water partition coefficients of simple organic compounds // *Journal of Physical and Chemical Reference Data.* — 1989. — Vol. 8, No. 3. — P. 1111-1230.
18. *Leo A.J.* Calculating log P oct from structures // *Chemical Reviews (Washington, DC, United States).* — 1993. — Vol. 93, No. 4. — P. X-1306.
19. *Leo A.* The octanol-water partition coefficient of aromatic solutes: The effect of electronic interactions, alkyl chains, hydrogen bonds, and ortho-substitution // *Journal of the Chemical Society, Perkin Transactions 2.* — 1983. — No. 6. — P. 825-838.
20. *Dillingham E.O., Mast R.W., Bass G.E., Autian J.* Toxicity of methyl- and halogen-substituted alcohols in tissue culture relative to structure-activity models and acute toxicity in mice // *Journal of Pharmaceutical Sciences.* — 1973. — Vol. 62, No. 1. — P. 22-30.
21. *Dunn W.J.* Binding of certain nonsteroid antiinflammatory agents and uricosuric agents to human serum albumin // *Journal of Medicinal Chemistry.* — 1973. — Vol. 16, No. 5. — P. 484-486.
22. *Dunn W.J., Koehler M.G., Grigoras S.* The role of solvent-accessible surface area in determining partition coefficients // *Journal of Medicinal Chemistry.* — 1987. — Vol. 30, No. 7. — P. 1121-1126.
23. *Hansch C., Anderson S.M.* The effect of intramolecular hydrophobic bonding on partition coefficients // *The Journal of Organic Chemistry.* — 1967. — Vol. 32, No. 8. — P. 2583-2586.
24. *Herzog K.A., Swarbrick J.* Drug permeation through thin-model membranes III: Correlations between *in vitro* transfer, *in vivo* absorption, and physicochemical parameters of substituted benzoic acids // *Journal of Pharmaceutical Sciences.* — 1971. — Vol. 60, No. 11. — P. 1666-1668.
25. The pH dependent «lipid solubility» of drugs // *Pharmazie.* — 1977. — Vol. 32, No. 5. — P. 289-291.
26. *Yaguzhinsky L.S., Smirnova E.G., Ratnikova L.A., Kolesova G.M., Krasinskaya I.P.* Hydrophobic sites of the mitochondrial electron transfer system // *Journal of Bioenergetics and Biomembranes.* — 1973. — Vol. 5, No. 3. — P. 163-174.
27. *Yamagami C., Takao N., Fujita T.* Hydrophobicity Parameter of Diazines (1) Analysis and Prediction of Partition Coefficients of Monosubstituted Diazines // *Quantitative Structure-Activity Relationships.* — 1990. — Vol. 9, No. 4. — P. 313-320.
28. *Barricelli N.* Numerical testing of evolution theories // *Acta Biotheoretica.* — 1962. — Vol. 16, No. 1. — P. 69-98.
29. *Barricelli N.A.* Numerical testing of evolution theories // *Acta Biotheoretica.* — 1963. — Vol. 16, No. 3. — P. 99-126.
30. *Fraser A.S.* Simulation of genetic systems by automatic digital computers. II. Effects of linkage on rates of advance under selection // *Australian Journal of Biological Science.* — 1957. — Vol. 10. — P. 492-499.
31. *Марков В.В., Бойченко А.П., Логінова Л.П.* Количественные зависимости структура — параметры распределения органических веществ в истинных двухфазных, хроматографических и псевдофазных системах. I. Линейные зависимости энергий сольватации // *Вісник Харківського національного університету.* — 2011. — прийнята до друку.