

Плющакова Л.А. (ІнФОВ НАНУ), Косинова Ю.С., Недодай В.С. (ДонНУ),
Суйков С.Ю. (ІнФОВ НАНУ)

ЕМПІРИЧНИЙ АНАЛІЗ ВПЛИВУ ЖОРСТКОСТІ МОЛЕКУЛИ НА ЕФЕКТИВНІСТЬ ПРОГНОЗНИХ МОДЕЛЕЙ РОЗПОДІЛУ

Моделювання процесів у довкіллі потребує визначення значної кількості фізико-хімічних параметрів сполук, зокрема їх розподільчих властивостей. Витратність експериментального підходу призвела до поширення емпіричних прогнозних моделей. Ефективність та прийнятність їх для нових сполук значною мірою залежить від використаних для параметризації експериментальних наборів. У роботі проаналізовано ефективність прогнозу на літературних даних для ряду екологічно важливих сполук. Показано, що внутрішньомолекулярна рухливість значно погіршує ефективність наявних прогнозних моделей.

Існує техногенне навантаження на навколишнє природне середовище та необхідність подальшого розвитку техносфери зумовлює значну потребу у моделях довкілля, зокрема для проведення прогнозу впливу варіантів управлінських рішень [1]. Але для ефективного використання таких моделей потрібна значна кількість фізико-хімічних характеристик сполук [1].

Ключовими для багатьох моделей [2, 3] характеристиками є константи розподілу (зокрема у системі октанол/вода — P_{ow}). Експериментальне визначення розподільчих характеристик дуже витратне і такі дані відомі для обмеженої кількості сполук. Можливості для послідовного точного розрахунку коефіцієнтів розподілу *ab initio* досі ще не створено, не дивлячись на значний прогрес комп'ютерного моделювання [4]. Практично для прогнозу розподільчих характеристик органічних сполук знайшли широке використання емпіричні та напівемпіричні моделі. Але їх ефективність вочевидь залежна від вибору та характеристик тренувальних (та контрольних) наборів сполук (далі — наборів). Аналіз хімічних та статистичних характеристик наборів зазвичай не проводиться, зокрема через недостатню кількість надійних експериментальних даних (наприклад, у [3] для параметризації моделі використано лише 11 експериментальних значень).

Існує широке коло класифікацій прогнозних моделей розподілу (та розчинності), зокрема емпіричних. Одна з останніх (стосовно розчинності) наведена у [4]. Автори структурували моделі за методами їх реалізації. У практичних цілях емпіричні моделі параметрів розподілу зазвичай поділяють за структурою на [2]:

1. моделі атомних вкладів;
2. моделі групових вкладів;
3. комбіновані моделі;
4. моделі квантовохімічних дескрипторів;
5. так звані нейронні сітки.

Вочевидь, вимоги стосовно характеристик тренувальних наборів у цих класах відрізняються (наприклад, параметризація моделей першої групи

потребує меншої кількості сполук), але об'єм тренувального набору повинен бути пов'язаний з кількістю параметрів моделі.

Для параметризації моделей використовують малі та дуже малі (від сотень до, подеколи, десятків значень) тренувальні набори даних. Найбільш розповсюдженим джерелом є база даних фізичних властивостей PSYSPROP [5], база даних AQUASOL [6] та база даних Сангстера [7]. Перші дві бази доступні на комерційній основі, остання — вільно через Інтернет. Але слід зазначити, що вищезгадані бази даних значною мірою перекриваються та не є незалежними. Цікаво, що для ряду сполук за цими базами ми отримали абсолютно однакові значення, але з посиланням на різні експериментальні роботи. Це підтверджує невизначеність та складність верифікації та автентифікації експериментальних даних з розподілу.

Вагомою характеристикою сполуки є її конформаційна лабільність (або кількість конформерів). Спосіб врахування її у моделях не є добре дослідженим. Зокрема у [4] проводиться точка зору щодо достатності двомірних статичних дескрипторів для задовільного прогнозу розподільчих характеристик органічних сполук. Очевидно, що динамічна характеристика не може бути безпосередньо отримана ані з двомірних [8], ані з тримірних статичних дескрипторів молекули, а потребує для оцінки окремих процедур.

Ми спробували оцінити вплив лабільності на результати прогнозу. Було використано дві вибірки з літературних даних — конформаційно жорсткі сполуки рядів безодіоксинів та бензофуранів [9–13] та конформаційно лабільні похідні 1,4-бензодіазепіну [7]. Слід зазначити, що наявні дані для класу діоксиноподібних сполук демонструють значну варіативність отриманих експериментальних значень, різниця між ними складає до двох порядків [14] (аналогічні висновки для набагато більш дослідженого DDT наведено у [15]). Хоч для діоксиноподібних сполук елементний склад сполук не є випадковим (загальна формула $C_{12}H_nO_2Cl_{(8-n)}$ або $C_{12}H_nOCl_{(8-n)}$), але вони і не є гомологами.

Нажаль, на сьогоднішній день немає загальноприйнятого однозначного набору значень. Нами використано ряд з [9], як найбільш, на наш погляд, внутрішньо узгоджений набір (таблиця 1).

Оцінку прогнозних можливостей для вільно доступних через Інтернет (<http://vcclab.org>) моделей проведено для ABlogP, ALOGPs, COSMOFrag, IALogP, KOWWIN, miLogP, XLOGP. Слід зазначити, що у ряду присутні як найпростіші за

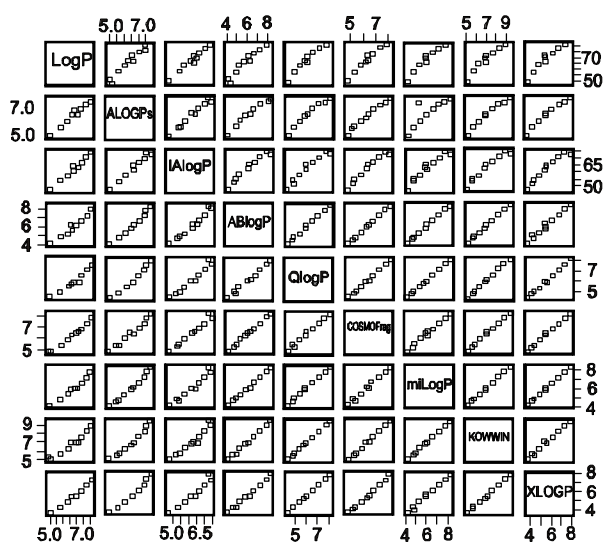


Рис. 1. Діоксини, $IgPow_{calc.}$ vs $IgPow_{exper.}$

реалізацією моделі (XLOGP — адитивна схема, 80 вкладів та 5 поправок), так і набагато складніші (ALOGPs — адаптивна нейрона сітка та створена на основі квантовохімічної континуальної моделі розчину COSMOFrag). Результати розрахунку наведені на рис. 1 у вигляді матриці графіків Експеримент vs Розрахунок [16].

Перший рядок на графіку відповідає залежності прогнозу від експерименту для кожної моделі. Легко бачити, що вона практично лінійна, тобто прогнозна ефективність є дуже високою, залежність близька до теоретичної прямої з нахилом у 45° .

Таблиця 1. Експериментальне значення коефіцієнтів розподілу етанол/вода [9]

№	Назва	IgPow
1	2,8-діхлордібензофуран	5,04
2	2,3,7,8-тетрахлордібензофуран	6,46
3	2,3,4,7,8-пентахлордібензофуран	7,11
4	1,2,3,4,7,8-гексахлордібензофуран	7,53
5	1,2,3,6,7,8- гексахлордібензофуран	7,57
6	1,2,3,4,6,7,8- гептахлордібензофуран	8,01
7	1,2,3,4,7,8,9- гептахлордібензофуран	8,23
8	Октахлордібензофуран	8,60
9	1-хлордибензодіоксин	5,17
10	2,8-діхлордибензодіоксин	5,68
11	1,2,4- трихлордибензодіоксин	6,29
12	1,2,3,4- тетрахлордибензодіоксин	6,92
13	2,3,7,8- тетрахлордибензодіоксин	6,96
14	1,2,3,4,7- пентахлордибензодіоксин	7,39
15	1,2,3,4,7,8- гексахлордибензодіоксин	7,94
16	1,2,3,4,6,7,8- гептахлордибензодіоксин	8,40
17	Октахлордибензодіоксин	8,75

Надзвичайно низьким є рівень випадкових відхилень, який не перевищує 0.1–0.2 логарифмічні одиниці. Зазвичай, емпіричні моделі мають гірші показники ефективності. Ми вважаємо, що у даному випадку ефективність пов'язана з відомою жорсткою структурою і відсутністю внутрішньомолекулярної рухливості у досліджених сполук. Цей факт для діоксиноподібних сполук є добре відомим та встановленим як експериментальними методами, так і методами квантової хімії [17].

В таблиці 2 наведені експериментальні дані для ряду бензодіазепінів з [7].

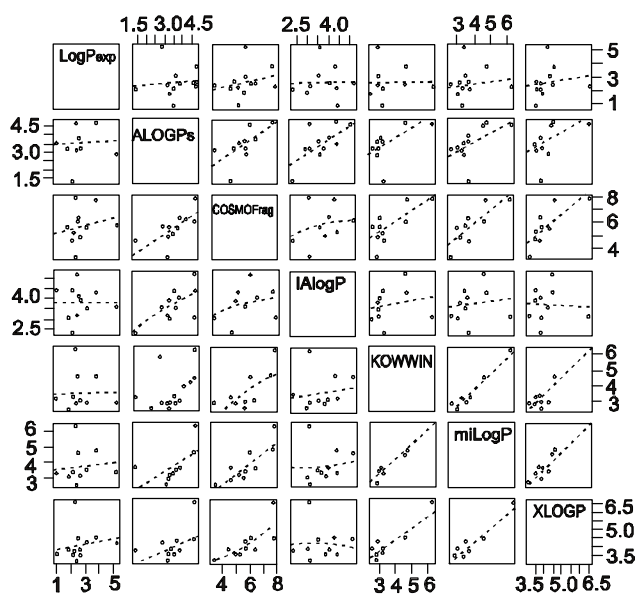


Рис. 2. Бензодіазепіни, $IgPow_{calc.}$ vs $IgPow_{exper.}$.

Результат розрахунку наведений на рис. 2. У даному випадку на жодному з графіків не спостерігається прийнятна лінійна залежність і жодна з моделей не забезпечує задовільний рівень прогнозу. На всіх графіках лінія регресії близька до горизонтальної, замість очікуваного проходження під кутом 45°, відповідно коефіцієнт кореляції має дуже мале значення. Суттєво те, що візуально не фіксується наявність функціональної залежності, неадекватність моделей майже ідеальна.

Таблиця 2. Експериментальні значення коефіцієнту розподілу октанол/вода похідних 1.4-бензодіазепіну [7]

№	Назва	lgPow
1	2Н-1,4-бензодіазепін-2-он, 7-хлоро-5-(2-хлорфеніл)-3-[(2,2-діметил-1,3-діоксалан-4-іл) метоксі]-1,3-дігідро-	2,48
2	2Н-1,4- бензодіазепін -2-он, 1-[2-(ацетілокі)етил]-7-хлоро-3-етоксі-1,3-дігідро-5-феніл-	2,61
3	2Н-1,4- бензодіазепін -2-он, 7-хлоро-3-[(2,2-діметил-1,3-діоксалан-4-іл)метоксі]-1,3-дігідро-5-феніл	2,14
4	2Н-1,4- бензодіазепін -2-он, 7-хлоро-1-[2-(діметиламіно)етил]-5-(2-фторфеніл)-1,3-дігідро-	2,35
5	2Н-1,4- бензодіазепін -2-он, 7-хлоро-1-[2-(діметиламіно)етил]-5-(2-фторфеніл)-1,3-дігідро-	2,54
6	2Н-1,4- бензодіазепін -2-он, 7-хлоро-1-[2-(діметиламіно)етил]-5-(2-фторфеніл)-1,3-дігідро-3-гідроксі-	1,89
7	2Н-1,4- бензодіазепін -2-он, 7-хлоро-5-(2-хлорфеніл)-1-[2-(діетиламіно)етил]-1,3-дігідро-3-гідроксі-	1,76
8	2Н-1,4- бензодіазепін -2-он, 3-(ацетілокі)-1-[3-(ацетілокі)пропіл]-7-хлоро-1,3-дігідро-5-феніл-	3,08
9	2Н-1,4- бензодіазепін -2-он, 7-хлоро-3-[(2,2-діетил-1,3-діоксолан-4-іл)метоксі]-1,3-дігідро-1-метил-5-феніл-	2,37
10	1Н- бензодіазепін -5-ацетоксі, -[(4,6-діметил-2-піримідиніл)оксі]-2,3,4,5-тетрагідро-1-метил-2-оксо-5-феніл-, (5S*)-	-2,00
11	1Н-1,4- бензодіазепін -5- ацетоксі , 5-феніл-[(3,5-діметилфеніл)оксі]-2,3,4,5-тетрагідро-2-оксо-1-[(2,4,6-трифторфеніл)метил]-, (5S*)-	2,10
12	1Н-1,4- бензодіазепін -5- ацетоксі , 5-(3-бутілфеніл)-[(4,6-діметил-2-піримідиніл)оксі]-2,3,4,5-тетрагідро-2-оксо-1-[(2,4,6- трифторфеніл)метил]-, (5S*)-	2,30
13	1Н-1,4- бензодіазепін -5- ацетоксі , 5-[1,1'-бифеніл]-3-іл-[(4,6-діметил-2-піримідиніл)оксі]-2,3,4,5-тетрагідро-2-оксо-1-[(2,4,6- трифторфеніл)метил]-, (5S*)-	1,80

На рис. 3, приведеному у вигляді VoxPlot з відомої методики EDA [18], проаналізувано нев'язки (тобто значення $\text{LogPow}_{\text{calc.}} - \text{LogPow}_{\text{exper}}$) моделей. Видно, що для останнього набору моделі неефективні і як засіб індивідуального прогнозу: вони дають зміщену оцінку зі значною долею викидів та найбільш вірогідною варіацією біля 2 логарифмічних одиниць.

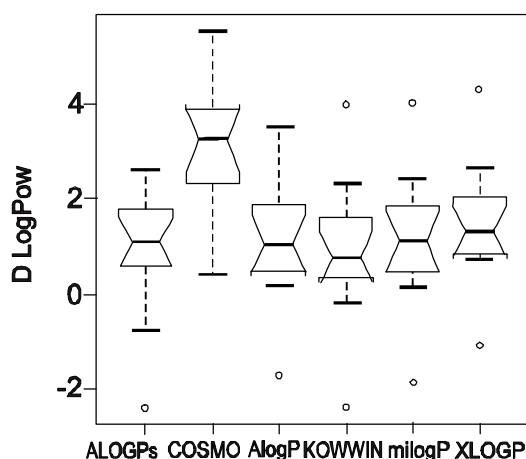


Рис. 3. Нев'язки прогнозних значень для похідних 1.4 бензодіазепіну.

Таким чином, показано наявність значного впливу внутрішньомолекулярної рухливості на можливості прогнозу коефіцієнтів розподілу октанол–вода для органічних сполук.

Тому, на наш погляд:

- уява про достатню ефективність двомірних та статичних дескрипторів [4] значною мірою пов'язана з відповідною структурою контрольних наборів сполук, але адекватна прогнозна модель повинна враховувати внутрішньомолекулярну рухливість [19];
- прогнозна модель повинна дозволяти корегувати кількість рухливих зв'язків (або інший відповідний параметр), оскільки міру рухливості не завжди можливо визначити формальними методами;
- необхідно враховувати наявність внутрішньомолекулярної рухливості при складанні тренувальних та контрольних наборів для параметризації і перевірки якості моделей;
- у рамках подальшого розвитку прогнозних моделей потрібна розробка простих формальних методів оцінки внутрішньомолекулярної рухливості та методів її відображення.

Література

1. **Comber, Mike H.I., Walker, John D., Watts, Chris, Hermens** // Environmental Toxicology and Chemistry, 2003. — V. 22. — P. 1822–1828.
2. **Yaffe D., Cohen Y., Espinosa G., Arenas A., Giralto Fuzzy** // Chem. Inf. Comput. Sci., 2002. — V. 42. — P. 162–183.
3. **Aizpuru A., Malhautier L., Fanlo J. L.** // Envir. Engrg., 2002. — V. 128. —N 10. — P. 953–959.
4. **Konstantin V. Balakin¹, Nikolay P. Savchuk¹ and Igor V.** // Tetko Problems and Solutions Current Medicinal Chemistry, 2006. — V. 13. — P. 223–241.
5. **PHYSPROP** database is available from Syracuse Research Corporation, <http://www.syrres.com>.
6. **Yalkowsky, S. H.; He, Y.** Handbook of Aqueous Solubility Data; CRC Press: Boca Raton, 2003.
7. <http://logkow.cisti.nrc.ca> 1986-2006 Sangster Research Laboratories
8. **Estrada, E., Molina, E. and Perdomo-LCpez, I.** // Chem. Inf. Comput. Sci., 2001. — V. 41. — P. 1015–1021.
9. **Govers, H. and Krop, H.** // Chemosphere, 1998. — V. 37. — P. 2139–2152.
10. <http://vcclab.org> Virtual Computational Chemistry Laboratory.
11. <http://ineris.fr> INERIS Parametres physico-chimiques et coefficients de transfert des dioxins pour levaluation des risques. INERIS DRC-03-45959 / ERSA – n° 272 - RBn /param-coef-dioxinestri.
12. www.hc-sc.gc.ca Liste des substances d'interet prioritaire rapport d'evaluation n° 1 :dibenzodioxines polychlorees et dibenzofurannes polychlores. ISBN 0-662-96003-3 N° de cat. MAS En40-2I5/1F.
13. <http://www.atsdr.cdc.gov>. The Agency for Toxic Substances and Disease Registry (ATSDR).
14. **Плющакова Л.А., Косинова Ю.С., Недодай В.С., Суйков С.Ю.** Збірка тез доповідей Першої всеукраїнської наукової конференції студентів, аспірантів та молодих учених „Хімічні проблеми сьогодення” Донецьк: ДонНУ, 2007. — С. 72.
15. **Pontolillo J., Eganhouse R.P.** The Search for Reliable Aqueous Solubility (Sw) and Octanol-Water Partition Coefficient (Kow) Data for Hydrophobic Organic Compounds: DDT and DDE as a Case Study U.S. Geological Survey Water-Resources Investigations Report 01-4201 Reston, Virginia 2001. — P. 51.
16. **Gnu R.** <http://www.R-project.org/>.
17. **J.E.Lee, W.Choi, B.J.Mhin** // Korean Chem. Soc., 2003. — Vol. 24. — No. 6. — P. 792–796.
18. **Tukey J.** Exploratory Data Analysis, 1977, Addison-Wesley Massachusetts. — 688 p.
19. **M. Randic** // J. Math. Chem., 1996. — V. 19. — P. 375–392.

Ó Плющакова Л.А., Косинова Ю.С., Недодай В.С., Суйков С.Ю., 2007