

Валідація *in silico* методів визначення ліпофільності/розчинності для раціонального дизайну лікарських засобів

М.В. Протопопов^{1*}, О.Ю. Нипорко¹, В.Г. Бджола²,
С.А. Старосила¹, С.М. Ярмолюк²

¹ Київський національний університет імені Тараса Шевченка
вул. Володимирська, 64, Київ, 01033, Україна

² Інститут молекулярної біології і генетики НАН України
вул. Академіка Заболотного, 150, Київ, 03680, Україна

Резюме. Точність розрахунку значень LogP і LogS є надзвичайно важливою під час розробки лікарських препаратів. Нами проведено валідацію найбільш розповсюджених методів оцінки цих параметрів з метою визначення найбільш точного. Проаналізовано літературу, що висвітлює алгоритми роботи цих методів, розраховано коефіцієнти кореляції між експериментально визначеними і розрахованими значеннями LogP/LogS та похибки обчислень. Встановлено, що найбільш достовірними є методи XLOG2, метод програми JCHEM і OCHEM із коефіцієнтами кореляції 0,93952, 0,91122 та 0,90215 відповідно. Окрім найбільшого коефіцієнта кореляції, метод XLOGP2 має найменші похибки (MAE=0,44695, RMSE=0,8382). За нашими розрахунками, найбільш оптимальними для оцінки значень LogP є методи XLOGP2 та JCHEM.

Ключові слова: LogP, LogS, ліпофільність, правила Ліпінського, ADME/Tox.

Вступ. Неодмінним складником раціонального дизайну лікарських засобів є дослідження фармакокінетики потенційних ліків, що проводиться на основі параметрів ADME/Tox. Ці фізико-хімічні та фізіологічні показники характеризують абсорбцію (absorption), розподіл (distribution), метаболізм (metabolism), виділення (excretion) і токсичність (toxicity) [1-3]. Одними з головних параметрів ADME/Tox, які обов'язково враховують під час розробки лікарських препаратів (крім власне токсичності), є ефективна проникність у тонкій кишці людини, проникність крізь шкіру, ліпофільність, розчинність у воді, проникність через ге-

матоенцифалічний бар'єр, зв'язування із білками плазми крові та деякі інші [4, 8].

Серед фізико-хімічних показників особлива увага приділяється таким параметрам, як логарифмічний коефіцієнт розподілу між октанолом і водою (LogP), що є показником ліпофільності речовини, та логарифмічний коефіцієнт розчинності речовини у воді (LogS). Саме ліпофільність і розчинність визначають здатність речовини проходити через мембрану і, відповідно, проникати всередину клітини, а отже, є визначальними для передбачення біодоступності препарату [5-7, 9, 10].

Статистично встановлено, що ключові фізико-хімічні показники лікарських препаратів, включаючи показники ADME, знаходяться в цілком певних діапазонах величин [14]. Допустимі значення таких фізико-хімічних параметрів, як LogP, молекулярна маса, кількість атомів, а також стабільність речовини у

* Corresponding author.

Tel.: +38044-5222458

E-mail address: mykola.protopopov@gmail.com

фізіологічному розчині та відсутність реакційноздатних хімічних груп і фрагментів, формують поняття лікоподібності [15, 16].

Найбільш поширеним принципом визначення лікоподібності речовини є застосування правил Ліпінського (або правил п'яти) [11]. Відповідно до цих правил значення ліпофільності LogP для (потенційних) лікарських має бути в межах від -2 до 5. Однак, оскільки розробка лікарських препаратів — процес багатостадійний, вимоги до проміжних сполук є жорсткішими, тому верхня гранична межа ліпофільності становить 3 [13, 14].

У процесі пошуку лікоподібних сполук (і наступної розробки лікарських препаратів) виникає нагальна потреба перевірки властивостей великої кількості (сотень тисяч) сполук, тому експериментальна оцінка фізико-хімічних параметрів ADME для такої вибірки є надзвичайно ресурсо- і часозатратною. З огляду на це дослідники все більшу увагу приділяють розробці комп'ютерних методів розрахунку (оцінки) фізико-хімічних властивостей сполук, за якими можна передбачувати ADME/Tox властивості кандидатів на лікарські засоби [12]. Це значно пришвидшує і з дешевлює процес розробки лікарських препаратів.

При великому різноманітті методів і підходів, що застосовуються до обчислення параметрів ліпофільності та розчинності, постає проблема у виборі оптимального методу розрахунку, оскільки існуючі методи оцінюють величини LogP і LogS з певною розбіжністю, хоча при цьому є статистично достовірними на великих вибірках речовин. Крім того, існує проблема меж застосування тих чи інших методів обчислення LogP/S — практично кожен із розроблених підходів не спрацьовує на певних речовинах/класах речовин. Тому визначення спектра методів розрахунку LogP та LogS, результати яких найкращим чином узгоджуються з даними інструментальних експериментів щодо існуючих *лікарських препаратів*, і встановлення меж застосування цих методів, стали метою нашої роботи.

Об'єкти дослідження. Визначення оптимального методу оцінки коефіцієнтів LogP і LogS для сполук — потенційних лікарських препаратів — було проведено на вибірці речо-

вин з бази даних Approved Drug. Ця база даних є складником біохемоінформатичного банку даних DrugBank (www.DrugBank.ca) [17] і містить інформацію про низькомолекулярні сполуки, що вже запроваджені як лікарські засоби. У результаті попередньої обробки Approved Drug було виявлено ряд сполук, для яких відомі експериментальні значення параметрів LogP і LogS, та було сформовано дві вибірки. Виборка для розрахунків параметра LogP міститила 693 сполуки. Виборка для розрахунків параметра LogS, відповідно, містила 117 сполук.

Інструменти дослідження. Розрахунок параметрів LogP і LogS для відповідних виборок лікарських препаратів проводився за допомогою методів, інтегрованих у веб-портالي ALOGPS 2.1 (<http://www.vcclab.org/lab/alogps>) та Ochem (<https://ochem.eu>). Серед цих методів у роботі використано XLOGP3, XLOGP2, KOWWIN, ALOGPS LogP, ALOGPS LogS, AC LogP, AC LogS, OchemLogP, OchemLogS. Крім того, для порівняння з іншими методами розрахунків були використані значення LogP, розраховані за методом, реалізованим у програмному забезпеченні JCHEM (www.chemaxon.com), які безпосередньо містяться в базі даних Approved Drug.

Методи розрахунку LogP/S. Усі методи розрахунку LogP/S розділяють на дві великі групи. Перша — методи, що ґрунтуються на аналізі підструктур, які входять до складу молекули (substructure-based methods), друга група методів заснована на певних властивостях досліджуваних сполук (property-based methods).

Методи аналізу підструктур у свою чергу діляться на фрагментні й атомні в залежності від того, підструктури якого рівня (хімічні групи або окремі атоми) розглядаються в досліджуваній молекулі. Остаточне значення LogP вираховується шляхом підсумовування внесків окремих підструктур [18]. Використовуються також комбіновані атом-фрагментні підходи.

Методи, що ґрунтуються на властивостях молекул, також поділяються на три великі групи: методи, засновані на емпіричних даних, методи, що використовують особливості просторової структури молекули, та методи, що

грунтуються на використанні топологічних дескрипторів [18].

XLOGP3 і XLOGP2. Ці підходи до розрахунку коефіцієнта розподілу октанол/вода, розроблені Вангом і співавторами, належать до атомних методів [19]. Значення LogP отримуються як сума двох компонентів. Перший — внесок у LogP різних груп та атомів, другий — сума корекційних факторів:

$$\text{Log}P = \sum_i a_i A_i + \sum_j b_j B_j \quad (1),$$

де A_i — внесок певного атома; a_i — кількість i -х атомів; B_j — вклад певного корегувального коефіцієнта; b_j — кількість j -х атомів. Коефіцієнти отримують шляхом побудови багатовимірної регресивної моделі для великої кількості органічних сполук.

XLOGP3 є наступним поколінням методу XLOGP2 і має принципову відмінність від попередника. У XLOGP3 як відправну точку для розрахунку використовують схожу молекулу з уже відомим значенням LogP. Відповідно значення відомої молекули $\text{Log}P^0$ дорівнює:

$$\text{Log}P^0 = \sum_i a_i A_i^0 + \sum_j b_j B_j^0 \quad (2).$$

Щоб розрахувати LogP невідомої молекули за методом XLOGP3, необхідно відняти формулу 2 від формули 1 [19, 20]:

$$\text{Log}P = \text{Log}P^0 + \sum_i a_i (A_i - A_i^0) + \sum_j b_j (B_j - B_j^0) \quad (3).$$

KOWWIN. Мейлан і Говард розробили комбінований атом-фрагментний (atom/fragment contribution, AFC) метод оцінки LogP, заснований на використанні множинної лінійної регресії та відомих достовірних значень LogP. Тренувальний сет складався із 2473 переважно простих речовин, тоді як валідаційна вибірка становила 10589 сполук простої, складної та комплексної будови. Модель містить 150 типів атомів і 250 типів корегувальних факторів, які і визначають сумарне значення LogP [21-23].

Регресії будували у два послідовні етапи. У першій регресії розраховували кореляцію між значеннями атом-фрагментних внесків і відповідними значеннями LogP, у другій — між корегувальними факторами та LogP. Розрахунок LogP проводили простим підсумовуванням атом-фрагментного внеску та внеску корегувальних факторів, що містяться в структурі молекули.

Крім того, прості фрагменти комбінували між собою, щоб уникнути ситуації, коли немає значення елементарного фрагмента для роз-

рахунку LogP молекули. Формула для розрахунку методом KOWWIN:

$$\text{Log}P = \sum(f_i n_i) + \sum(c_j n_j) + 0.229 \quad (4),$$

де f_i — внесок відповідного атома-фрагмента, c_j — внесок відповідного корегувального фактора, n_i, n_j — кількість відповідних значень атом-фрагментного внеску і корегувальних факторів, 0,229 — константа, що виводиться регресією.

Метод KOWWIN розроблено для розрахунку LogP органічних, кремнійорганічних сполук і деяких солей органічних молекул із використанням відповідних інструментальних значень. Експериментальні значення для неорганічних сполук не використовували [21-23].

JCHEM. Метод розрахунку LogP, запропонований фахівцями компанії «ChemAxon» (www.chemaxon.com), є комбінованим атом-фрагментним підходом. Кожному атому і фрагменту досліджуваної молекули присвоюється певний тип. Атомні внески відповідають модифікованим типам атомів, запропонованим Вішвандханом та співавторами [24]. Модифікації в основному стосуються атомів із делокалізованою електронною густиною, а також певних іонних форм. Внески окремих фрагментів і функціональних груп ураховуються за Хлопманом та співавторами [25]. Остаточне значення LogP розраховується як сума внесків окремих атомів і фрагментів, представлених у молекулі.

У новіших версіях JCHEM використовують «trainable logP», де користувач може сам обрати вибірку, за якою будуть розраховуватись значення LogP, що може значно покращити точність методу. Користувач може як вибирати з готових вибірок, так і створювати власні вибірки.

ALOGPS та OCHEM. Метод ALOGPS, розроблений Тетко і співавторами з використанням асоціативної нейронної мережі (associative neural network), належить до так званих електротопологічних і поєднує в собі підходи для розрахунку коефіцієнта розподілу октанол/вода LogP [27, 28] та коефіцієнта розчинності у воді LogS [29]. Авторами показано, що для коректного моделювання LogP і LogS важливо враховувати молекулярну масу. Додатково до молекулярної маси введено два простих параметри — кількість атомів водню і кількість важких атомів. Важливими вхідними даними для програми є електротопологічні індекси атомів та зв'язків [26]. Індекси розра-

Статистичні показники різних методів розрахунку LogP

Метод	R	R ²	Adj R ²	RMSE (s)	MAE
XLOGP2	0,93952	0,88271	0,88254	0,8382	0,44695
JCHEM	0,91122	0,83033	0,83008	1,08872	0,70696
OCHEM	0,90215	0,81388	0,81361	0,92426	0,58738
ALOGPS	0,89906	0,8083	0,80802	0,92596	0,65732
ACLOGP	0,89886	0,80795	0,80767	1,1457	0,74549
XLOGP3	0,89722	0,805	0,80472	1,10548	0,55326
KOWWIN	0,88092	0,77603	0,7757	1,36001	0,76427

Примітка: R — коефіцієнт кореляції; R² — коефіцієнт детермінації; Adj R² — скорегований коефіцієнт детермінації; RMSE — середньоквадратичне відхилення; MAE — середня абсолютна похибка. Методи подано в порядку зменшення коефіцієнта кореляції.

ховуються із вхідного формату SMILES. Для розрахунку LogP використовували 75 електротопологічних індексів і брали 64 нейронні мережі.

Наступна генерація методів ALOGPS, що реалізована на веб-порталі Occhem (далі — методи OCHEM) [30], також дає змогу оцінювати як показники LogP, так і показники LogS для досліджуваних сполук.

Статистична обробка даних. Для сформованих вибірок було розраховано коефіцієнт кореляції, коефіцієнт детермінації, скорегований коефіцієнт детермінації, стандартне квадратичне відхилення та середня абсолютна похибка між експериментальними і розрахованими значеннями LogP/S.

Середню абсолютну похибку розраховували за формулою

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|,$$

де f_i — експериментальне значення, y_i — розраховане значення, n — кількість значень.

Результати й обговорення. Попередній аналіз вмісту бази даних Approved Drugs дав змогу виявити, що серед вибірки лікарських засобів експериментальні значення LogP відомі для 693 речовин. Ситуація із значеннями LogS є суттєво відмінною: лише для 117 речовин із лікарського списку значення цього параметра встановлені експериментально. Імовірно, така різниця в наявності експериментальних значень обумовлена тим, що у визначеннях лікоподібності фігурує саме ліпофільність, тому інструментальне визначення коефіцієнта розчинності не актуальне для розробників лікарських препаратів [15].

Кореляцію між експериментальними (EXPLogP) і розрахованими значеннями LogP (за методами ALOGPS, AC LogP, KOWWIN, XLOGP3, XLOGP2) наведено на рис. 1. Відповідні чисельні характеристики представлено в табл. 1.

За даними кореляційного аналізу, значення коефіцієнтів розподілу октанол/вода, що були оцінені всіма зазначеними вище методами, характеризуються високим ступенем кореляції з даними інструментальних експериментів. Найбільшими значеннями коефіцієнтів кореляції характеризуються набори даних, отримані за допомогою методів XLOGP2 (R=0,93952), JCHEM (R=0,91122) і OCHEM (R=0,90215). Величини коефіцієнтів детермінації для обчислених вибірок відповідають значенням коефіцієнтів кореляції. У той же час величини абсолютної похибки демонструють дещо іншу картину. Найменшим значенням цієї величини також характеризуються дані, отримані методом XLOGP2 (MAE=0,44695), проте друге місце за цим показником посідає метод XLOGP3 (MAE=0,55326), третє (подібно до даних кореляційного аналізу) — OCHEM — 0,58738. Подібні відмінності можуть бути пов'язані з особливостями роботи алгоритму, реалізованому в XLOGP3. Використання попередньо верифікованих значень LogP⁰ [20], з одного боку, приводить у ряді випадків до абсолютно точного визначення LogP, з іншого — до певного розмиття обрахованих значень на великих вибірках сполук.

Слід зазначити, що для абсолютно всіх методів характерні крайові випадки, тобто точки, які знаходяться на великій відстані від середнього значення. Величини LogP, розраховані

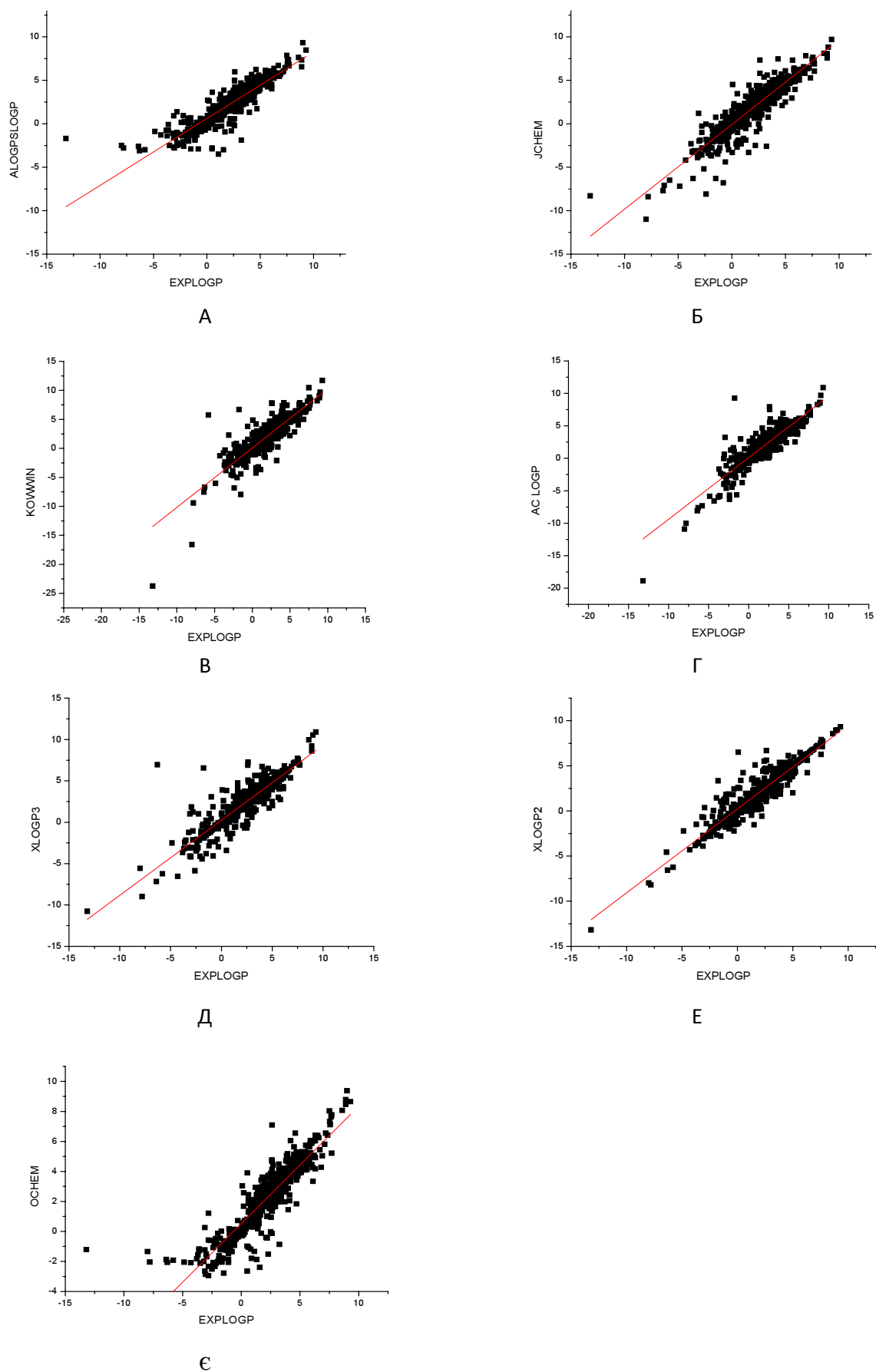
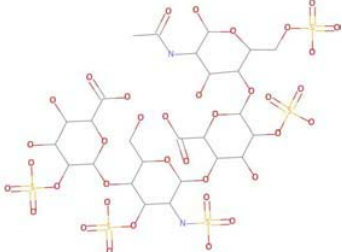
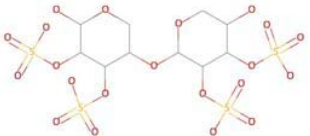

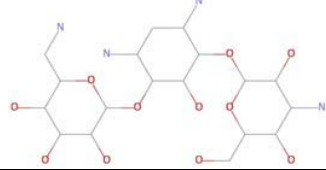
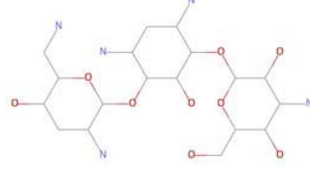

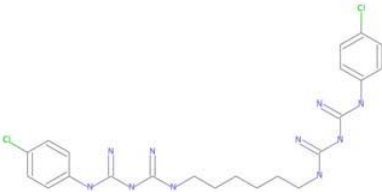
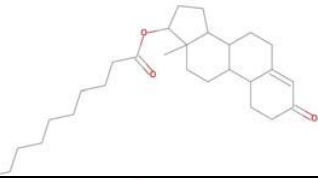
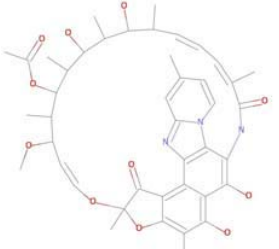


Рис. 1. Кореляції методів оцінки значень LogP ALOGPS (а), JCHEM (б), KOWWIN (в), AC LogP (г), XLOGP3 (д), XLOGP2 (е) та OCHEM (є) з експериментально визначеними даними.

Таблиця 2

Сполуки, розраховані значення LogP яких лежать за межами допустимих похибок

	DrugBank ID	ALOGPS	JCHEMLOGP	KOWWIN	EXPLOGP	ACLOGP	XLOGP3	XLOGOP2	Структура
1	DB01225	-1,7	-8,3	-23,75	-13,2	-18,88	-10,77	-13,18	
2	DB00686	-2,5	-11	-16,6	-8	-10,9	-5,58	-7,99	
3	DB00452	-2,8	-8,4	-9,41	-7,8	-10,03	-8,99		
4	DB01172	-3,1	-7,1	-6,7	-6,3	-7,61	6,95	-6,56	
5	DB00684	-3	-6,5	5,76	-5,8	-7,32	-6,23	-6,24	
6	DB00944	0,65	-1,4	6,69	-1,75	9,27	6,57	3,34	
7	DB00878	2,71	4,51	4,85	0,08	2,56	3,89	6,51	
8	DB08804	5,96	7,32	7,75	2,62	7,5	7,27	6,69	
9	DB01220	4,94	4,37	7,81	2,6	7,94	6,89	4,32	

для деяких сполук, не узгоджуються з даними інструментального експерименту. Наявність цих точок свідчить про певні обмеження для застосування досліджуваних методів обчислення.

Структурні формули сполук-виключень і відповідні дані LogP наведено в табл. 2. Як видно з представлених даних, ці речовини мають певні структурні особливості, зокрема, всі вони містять дуже велику кількість донорів водневих зв'язків. Експериментально визначені значення LogP для цих сполук менше значення -5, тобто крайнього значення із правила Ліпінського. У свою чергу, розраховані величини LogP для цих сполук також знаходяться у від'ємному діапазоні, проте є вищими за своїми абсолютними значеннями. Можна припустити, що правила, за якими сумуються внески донорів водневого зв'язку в ліпофільність молекули, систематично занижують їх загальний внесок.

Результати кореляційного аналізу величин LogS, розрахованих методами ALOGPS, ACLOGS та OCHEM, з відповідними експериментальними даними наведено на рис. 2. Числові значення основних статистичних показників вказані в табл. 3.

Згідно з отриманими даними, моделі розрахунку величини LogS ALOGPS та OCHEM характеризуються високими коефіцієнтами кореляції між обрахованими та інструментально визначеними величинами LogS, а також допустимими значеннями стандартного відхилення.

Варто зауважити, що дані, отримані з використанням метода OCHEM, мають дуже високий коефіцієнт кореляції, проте з протилежним знаком. Такий феномен може бути пов'язаний з особливостями технічної реалізації алгоритму розрахунку. Статистично метод є правильним, але щоб отримувати реальні дані з його допомогою, необхідно змінювати знак розрахованої величини LogS на протилежний.

Узагальнення. За результатами кореляційного аналізу, всі протестовані методи оцінки LogP є придатними для подальшого використання в дослідженнях з оптимізації структури відомих/розроблених рецептор-специфічних інгібіторів до лікоподібного стану, про що свідчать високі значення коефіцієнтів кореляції між розрахованими й експериментальними величинами LogP для всіх без винятку вибірок

Таблиця 3
Статистичні параметри різних методів розрахунку LogS

Метод	R	R ²	Adj R ²	RMSE(s)	MAE
OCHEM	0,97271	0,94617	0,9457	0,33505	0,21932*
ALOGPS	0,91006	0,82821	0,82672	0,58786	0,4906
ACLOGS	0,67372	0,4539	0,44915	1,13776	0,71427

Примітка: R – коефіцієнт кореляції; R² – коефіцієнт детермінації; AdjR² – скорегований коефіцієнт детермінації; RMSE – середньоквадратичне відхилення; MAE – середня абсолютна похибка. Методи подані за порядком зменшення коефіцієнта кореляції.

сполук (найнижче значення коефіцієнта кореляції становить 0,88 для метода KOWWIN). Найбільшими коефіцієнтами кореляції характеризуються атомарний метод XLOGP2 і фрагментарний метод JCHEM (коефіцієнти кореляції дорівнюють 0,94 та 0,91 відповідно).

Серед досліджених методів оцінки LogS придатними для подальшого використання виявилися методи OCHEM та ALOGPS. Протестований метод ACLOGS, на жаль, показав недостатню кореляцію з відомими експериментальними даними. Найкращим для обрахунків LogS виявився метод OCHEM, що базується на топографічних дескрипторах (коефіцієнт кореляції між розрахованими та експериментальними значеннями дорівнює 0,97).

Для подальшого використання в роботі з розробки лікарських препаратів, зокрема оптимізації структури відомих/розроблених рецептор-специфічних інгібіторів до лікоподібного стану є доцільним використання комбінації підходів XLOGP2 для розрахунку LogP та OCHEM для розрахунку LogS.

Подяки. Висловлюємо щирі подяку м.н.с. відділу біомедичної хімії Інституту молекулярної біології і генетики НАН України О.В. Остринській за сприяння у написанні статті.

Надійшла в редакцію 25.06.2013 р.

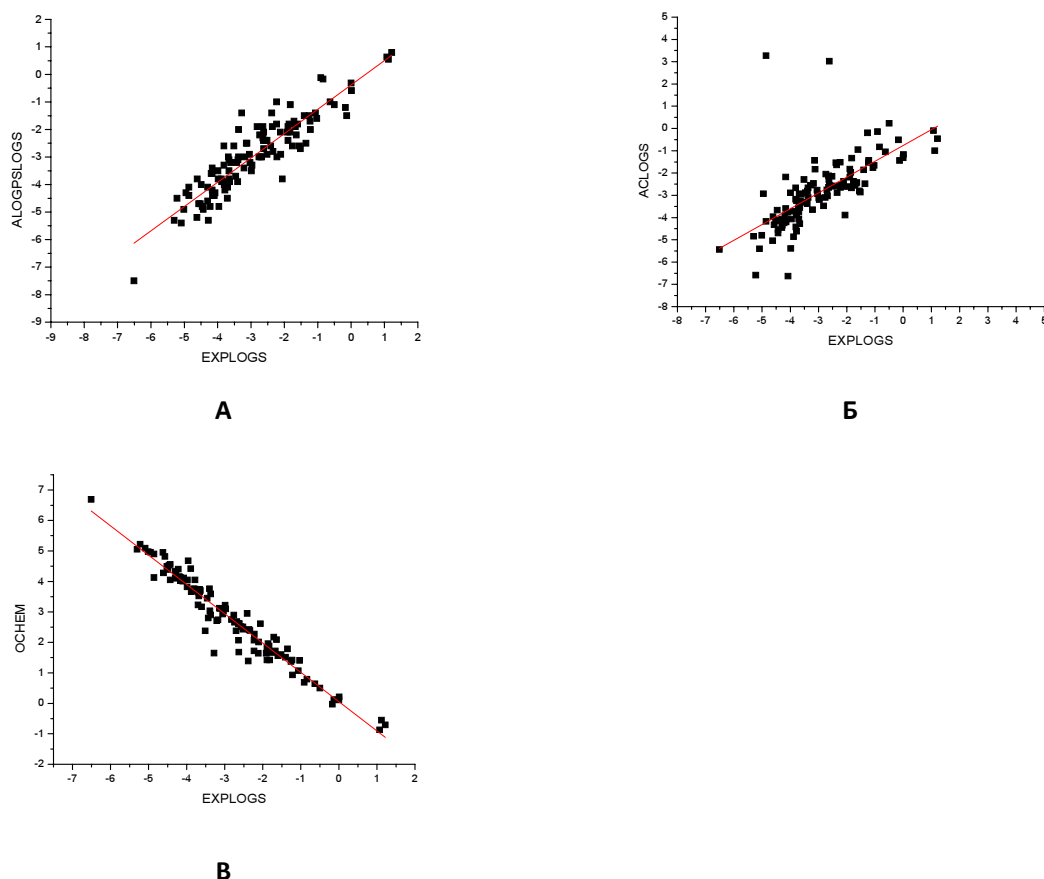


Рис. 2. Кореляції методів оцінки значень *LogS* (а), *ACLogS* (б), *OCHEM* (в) з експериментально визначеними даними за цим параметром.

In silico validation of methods for estimation of drugs lipophilicity/solubility

M.V. Protopopov¹, A.Yu. Nyporko¹, V.G. Bdzhola², S.A. Starosila¹, S.M. Yarmoluk²

¹ Taras Shevchenko Kyiv National University
60, Volodymirska Str., Kyiv, 01601, Ukraine

² Institute of Molecular Biology and Genetics, NAS of Ukraine
150, Zabolotnoho Str., Kyiv, 03680, Ukraine

Summary. Accuracy evaluation of *LogP* and *LogS* is crucial for drugs development. With the aim to determine the most exact techniques of these parameters estimation general methods have been validated. We had reviewed literature data about principles of operation of these methods, calculated the correlation coefficient between experimental and predicted values of *LogP/LogS* and respective calculation errors. The most valid methods were XLOG2, JCEM and OCHEM with its correlation coefficients 0.93952, 0.91122 and 0.90215 respectively. Despite the largest one XLOGP2 has the lowest error (MAE=0.44695, RMSE=0.8382). According to our calculation, the most appropriate methods for *LogP* values estimation were XLOGP2 and JCHEM.

Keywords: *LogP*, *LogS*, lipophilicity, Lipinski's rule of five, ADME/Tox.

Перелік літератури

1. Yu H., Adedoyin A. ADME-Tox in drug discovery: integration of experimental and computational technologies // *Drug Discov. Today*. — 2003. — Vol. 8, No. 18. — P. 852-861.
2. Prueksaritanont T., Tang C. ADME of biologics—what have we learned from small molecules? // *AAPS J*. — 2012. — Vol. 14, No. 3. — P. 410-419.
3. Hilmer S.N. ADME-tox issues for the elderly // *Expert Opin. Drug Metab. Toxicol*. — 2008. — Vol. 4, No. 10. — P. 1321-1331.
4. Boobis A., Gundert-Remy U., Kremers P., Macheras P., Pelkonen O. *In silico* prediction of ADME and pharmacokinetics // *European Journal of Pharmaceutical Sciences*. — 2002. — 17. — P. 183-193.
5. Rutkowska E., Pajlk K., Józwiak K. Lipophilicity—methods of determination and its role in medicinal chemistry // *Acta Pol Pharm*. — 2013. — Vol. 70, No. 1. — P. 3-18.
6. Fagerholm U. Prediction of human pharmacokinetics — gastrointestinal absorption // *JPP*. — 2007. — 59. — P. 905-916.

7. *Dabrowska M., Stareka M., Skucinski J.* Lipophilicity study of some non-steroidal anti-inflammatory agents and cephalosporin antibiotics: A review // *Talanta*. — 2011. — 86. — P. 35-51.
8. *Abbott N.J.* Blood-brain barrier structure and function and the challenges for CNS drug delivery // *J. Inher. Metab. Dis.* — 2013. — Vol. 36, No. 3. — P. 437-449.
9. *Lyman W.J., Reehl W.F., Rosenblatt D.H.* Handbook of chemical property estimation methods: environmental behavior of organic compounds. — Washington, DC, American Chemical Society, 1990. — 960 p.
10. *Sangster J.* Octanol-water partition coefficients: fundamentals and physical chemistry, Vol. 2 of Wiley Series in Solution Chemistry. — Chichester: John Wiley & Sons Ltd., 1997. — 178 p.
11. *Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.* Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings // *Advanced Drug Delivery Reviews*. — 2001. — 46. — P. 3-26.
12. *Yamashita F., Hashida M.* *In silico* approaches for predicting ADME properties of drugs // *Drug Metab. Pharmacokinet.* — 2004. — Vol. 19, No. 5. — P. 327-338.
13. *Petit J., Meuricea N., Kaiser C., Maggiora G.* Softening the Rule of Five — where to draw the line? // *Bioorg. Med. Chem.* — 2012. — 20. — P. 5343-5351.
14. *Muegge I.* Selection criteria for drug-like compounds // *Med Res Rev.* — 2003. — Vol. 23, No. 3. — P. 302-303.1.
15. *Ghose A., Viswanadhan V., Wendoloski J.* A Knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases // *J. Comb. Chem.* — 1999. — 1. — P. 55-68.
16. *Bleicher K., Böhm H.J., Müller K., Alanine A.* A guide to drug discovery: Hit and lead generation: beyond high-throughput screening // *Nature Reviews Drug Discovery*. — 2003. — Vol. 2. — P. 369-378.
17. *Knox C., Law V., Jewison T., Liu P., Ly S., Frolkis A., Pon A., Banco K., Mak C., Neveu V., Djoumbou Y., Eisner R., Guo A.C., Wishart D.S.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs // *Nucleic Acids Res.* — 2011. — 39, Database issue. — P. D1035-D1041.
18. *Mannhold R., Poda G.I., Ostermann C., Tetko I.V.* Calculation of molecular lipophilicity: state-of-the-art and comparison of log p methods on more than 96,000 compounds // *J. Pharm. Sci.* — 2009. — Vol. 98, No. 3. — P. 861-893.
19. *Wang R., Gao Y., Lai L.* Calculating partition coefficient by atom-additive method // *Perspectives in Drug Discovery and Design*. — 2000. — 19. — P. 47-66.
20. *Cheng T., Zhao Yuan, Li X., Lin F., Xu Y., Zhang X., Li Y., Wang R.* Computation of octanol-water partition coefficients by guiding an additive model with knowledge // *J. Chem. Inf. Model.* — 2007. — 47. — P. 2140-2148.
21. *Meylan W.M., Howard P.H.* Estimating log P with atom/fragments and water solubility with log P // *Perspect Drug Discov Des.* — 2000. — 19. — P. 67-84.
22. *Meylan W.M., Howard P.H., Boethling R.S.* Improved method for estimating water solubility from octanol water partition coefficient // *Environ Toxicol Chem.* — 1996. — 15. — P. 100-106.
23. *Meylan W.M., Howard P.H.* Atom/fragment contribution method for estimating octanol-water partition coefficients // *J. Pharm. Sci.* — 1995. — Vol. 84, No. 1. — P. 83-92.
24. *Viswanadhan V.N., Ghose A.K., Revankar G.R., Robins R.K.* Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics // *J. Chem. Inf. Comput. Sci.* — 1989. — Vol. 29, No. 3. — P. 163-172.
25. *Klopman G., Li J-Y., Wang S., Dimayuga M.* Computer automated log P calculations based on an extended group contribution approach // *J. Chem. Inf. Comput. Sci.* — 1994. — Vol. 34, No. 4. — P. 752-781.
26. *Kier L.B., Hall L.H.* *Molecular Structure Description: The Electrotopological State*. — London: Academic Press, 1999. — 286 p.
27. *Tetko I.V., Tanchuk V.Y., Villa A.E.* Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices // *J. Chem. Inf. Comput. Sci.* — 2001. — 41. — P. 1407-1421.
28. *Tetko I.V., Tanchuk V.Y.* Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program // *J. Chem. Inf. Comput. Sci.* — 2002. — 42. — P. 1136-1145.
29. *Tetko I.V., Tanchuk V.Y., Kasheva T.N., Villa A.E.* Estimation of aqueous solubility of chemical compounds using E-state indices // *J. Chem. Inf. Comput. Sci.* — 2001. — 41. — P. 1488-1493.
30. *Sushko I., Novotarskyi S., Körner R., Pandey A.K., Rupp M., Teetz W., Brandmaier S., Abdelaziz A., Prokopenko V.V., Tanchuk V.Y., Todeschini R., Varnek A., Marcou G., Ertl P., Potemkin V., Grishina M., Gasteiger J., Schwab C., Baskin I.I., Palyulin V.A., Radchenko E.V., Welsh W.J., Kholodovych V., Chekmarev D., Cherkasov A., Aires-de-Sousa J., Zhang Q.Y., Bender A., Nigsch F., Patiny L., Williams A., Tkachenko V., Tetko I.V.* Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information // *J. Comput. Aided Mol. Des.* — 2011. — Vol. 25, No. 6. — P. 533-554.