

УДК 541.6

DOI: 10.15587/2313-8416.2015.51463

ОГЛЯД НАЙБІЛЬШ ПОШИРЕНИХ ОБЧИСЛЮВАЛЬНИХ МЕТОДІВ ВИВЧЕННЯ ЗВ'ЯЗКІВ МІЖ СТРУКТУРОЮ МОЛЕКУЛ ТА ЇХ БІОЛОГІЧНОЮ ДІЄЮ

© О. Т. Девіняк

Ціль. Систематизувати найбільш поширені методи дослідження зв'язків «молекулярна структура-біологічна активність» (QSAR) та розкрити їх принципи реалізації, сильні та слабкі сторони.

Методи. Проводився огляд сучасної наукової літератури, присвяченої QSAR-моделюванню. Найбільш часто використовувані методи розробки моделей «структура-активність» вибирались для подальшого опису.

Результати. Визначено місце аналізу зв'язків «молекулярна структура-активність» серед комп'ютерних методів розробки нових лікарських засобів та описано найбільш поширені алгоритми побудови QSAR моделей із акцентуванням на механізмі їх роботи. У останній час все більшої і більшої популярності набувають підходи, що базуються на використанні ансамблів моделей, прикладом яких є Random Forest.

Висновки. Прогрес у розвитку методів машинного навчання для побудови моделей «структура-активність» є запорукою подальшого розвитку напрямку QSAR та знаходження нових біологічно активних речовин із його допомогою

Ключові слова: QSAR, розробка ліків, молекулярне моделювання, методи QSAR, математичні моделі, методи вибору змінних, машинне навчання

Aim. To systematize the most used methods of “molecular structure-biological activity” relationship studies and to disclose their principles of application, strong and weak sides.

Methods. The review of modern scientific literature devoted to QSAR modeling was carried out. The most frequently used methods for “structure-activity” models development were selected for further description.

Results. The place of “molecular structure-activity” relationships analysis among computer assisted drug design methods is discussed in the current review and the most used algorithms of QSAR model development with emphasis on the mechanisms of their work are described. The approaches based on model ensembles become more and more popular, one of which is Random Forest.

Conclusions. The progress in machine learning methods development is the key to the further evolution of QSAR direction and to the discovering of new biologically active substances

Keywords: QSAR, drug development, molecular modeling, QSAR methods, mathematical models, variable selection methods, machine learning

1. Вступ

Комп'ютерна розробка ліків (*Computer-Assisted Drug Design, CADD*) є сучасним напрямком фармацевтичної науки і характеризується швидким та інтенсивним розвитком. Прикладами успішного застосування моделювання зв'язку між структурою молекул та їх дією є відкриття інгібіторів зворотнього захоплення серотоніну та норадреналіну, селективних інгібіторів ЦОГ-2, інгібіторів ВІЛ-протеази та інтегрази, антагоністів 5-HT₃ рецепторів, антагоністів CCR5, та агоністів 5-HT_{1B/1D}, Н-холіноміметиків та багато інших.

2. Постановка проблеми у загальному вигляді, актуальність теми та її зв'язок із важливими науковими чи практичними питаннями

Існують два основні напрямки CADD: *target-based*, що здійснює пошук активних сполук, котрі зв'язуються з певним білком; та *ligand-based*, що здійснює пошук активних сполук, схожих на інші, вже досліджені молекули. Недоліком *target-based* підходу є необхідність для проведення моделювання рентгеноструктурних даних досліджуваного білка, кристалізованого із приєднаним до нього активним лігандом, що не завжди є доступно. Пе-

ревагою *ligand-based CADD* є відсутність потреби у будь-яких знаннях щодо механізму дії досліджуваних молекул, а необхідні лише результати біологічних випробувань ряду сполук. Обмеженням наведеного напрямку є втрата прогностичної здатності моделей для молекул, які сильно відрізняються від структур, що були застосовані для розробки та навчання. У рамках *ligand-based CADD* існують два чітко виражені підходи: моделювання зв'язків структура-активність (*quantitative structure-activity analysis, QSAR*), та фармакофорне моделювання. Методи, що використовуються для створення таких моделей, базуються на останніх досягненнях інформатики та статистичного аналізу, є надзвичайно різноманітними та потребують систематизації.

3. Аналіз останніх досліджень і публікацій

Обговорення методів QSAR-моделювання та створених за їх допомогою моделей відображено у роботах А. Dudek [1], М. Arakawa [2] та ін, які були опубліковані у 2006-2007 роках. Класична схема QSAR аналізу була описана А. Tropsha та А. Colbraikh [3]. Питаннями майбутнього розвитку QSAR переймалися А. Cherkasov та співавтори [4].

4. Виділення невирішених раніше частин загальної проблеми

Інтенсивний розвиток галузі молекулярного моделювання призвів до появи нових методів побудови QSAR моделей. Крім того, щороку публікується значна кількість наукових статей із описом нових розроблених моделей зв'язків структура-активність. Тому є нагальною є потреба у оновленні та повторній систематизації знань у галузі QSAR-моделювання. Крім того, статті, що описують методи QSAR, акцентують більше увагу на розробці моделей, часто не розкриваючи механізму роботи того чи іншого алгоритму.

Тому метою даного огляду є систематизація сучасних методів QSAR із розкриттям їх принципів реалізації, сильних та слабких сторін, підкріплених прикладами їх використання для створення моделей з високою прогностичною здатністю.

1. Етапи моделювання функціональної залежності активності від структури молекул.

Загалом процедуру моделювання біологічного ефекту можна поділити на наступні стадії:

1. Збір, первинна обробка та попередній аналіз результатів експериментального тестування активності.

2. Оптимізація геометрії та обчислення молекулярних дескрипторів досліджуваних сполук.

3. Відсіювання малоінформативних та видалення лінійно залежних дескрипторів.

4. Вибір та відображення (mapping) дескрипторів в активність. Це є ключовим етапом у QSAR-дослідженнях, під час якого власне і відбувається розробка математичної моделі. Методи, що використовуються для цього, детально розглянуто нижче.

5. Статистична оцінка, визначення домена застосовності, випробування та оптимізація моделі.

6. Прогнозування за допомогою отриманої моделі.

2. Найбільш поширені методи розробки QSAR-моделей

2.1. Множинна лінійна регресія

Множинна лінійна регресія (Multiple Linear Regression, MLR) представляє активність речовини як лінійну функцію від всіх наданих дескрипторів. Коефіцієнти цієї функції встановлюються на основі сполук із навчальної вибірки таким чином, щоб мінімізувати квадрати розходжень між прогнозованою та істинною активністю (метод найменших квадратів). Для застосування методу необхідними є переважання числа молекул навчальної вибірки над кількістю дескрипторів та відсутність мультиколінеарних дескрипторів. Множинна лінійна регресія є найбільш застосовуваним засобом встановлення відображення з дескрипторів у активність. Так, за допомогою даного методу було здійснено моделювання інгібування тирозин кінази похідними хіназоліну, пошук зв'язку між структурою похідних тієно[2,3-*d*]піримідиніонів та афінністю до монокарбоксилат транспортеру 1, інгібування JNK1 сполуками ізохінолінового ряду, вивчення протипухлинної дії піразолінзамішених 4-тіазолідинонів тощо [5, 6].

2.2. Метод часткових найменших квадратів

За допомогою цього методу лінійна регресійна модель знаходиться шляхом проєкції і активності, і дескрипторів у новий простір (Partial Least Squares, PLS). Таким чином ліквідується проблема мультиколінеарності та великої кількості дескрипторів. Алгоритм базується на судженні, що, незважаючи на велику кількість дескрипторів, модельований процес керується відносно малою кількістю незалежних латентних змінних. У процесі знаходження цих латентних змінних здійснюється декомпозиція матриці дескрипторів на матриці оцінок (scores) та коефіцієнтів (loadings), причому матриця оцінок є ортогональною, здатна охопити значну частину дескрипторної інформації і дозволяє створити регресійну модель активності із хорошою прогностичною здатністю. Метод часткових найменших квадратів використовувався для моделювання інгібування зворотнього захоплення серотоніну та антагонізму гістаміну H(3), антиоксидативної властивості низькомолекулярних пептидів, інгібування циклін-залежної кінази [7, 8].

2.3. Байєсівський класифікатор

Наївний Байєсівський класифікатор (Naive Bayes Classifier, NBC) є простим імовірнісним класифікатором, заснованим на теоремі Байєса зі строгими (наївними) припущеннями щодо незалежності дескрипторів. Ці припущення на практиці часто не справджуються, однак перевагою класифікатора є невелика кількість даних необхідних для навчання. Байєсівський класифікатор використовувався при моделюванні процесів кишкового всмоктування [9], при дослідженні відворотності мультилікарської стійкості ракових клітин [10], при розпізнаванні протитуберкульозної активності [11] та ін.

2.4. Метод k-найближчих сусідів

Метод k-найближчих сусідів (k-Nearest Neighbors, k-NN) є простою системою прийняття рішень, яка для кожної сполуки аналізує її k найближчих (найбільш подібних) сполук і здійснює прогноз належності до того класу, який найчастіше зустрічається серед сусідів. Кількість сусідів є змінним параметром, який варто оптимізувати в процесі створення моделі для отримання кращих результатів. Метод k-найближчих сусідів використовувався для розпізнавання інгібіторів ацетилхолінестерази [12], антибактеріальної активності [13], засобів для лікування трипаносомозу [14] та ін.

2.5. Штучні нейронні мережі

Штучні нейронні мережі (Artificial Neural Networks, ANN) є прогностичним методом, що сформувався шляхом наслідування природних процесів та базується на архітектурі мережі нейронів. Існує багато різновидів та алгоритмів, заснованих на даній парадигмі. Найширшого застосування при моделюванні зв'язку структура – активність набули моделі багатопшарового перцептрона та радіально-базисні нейронні мережі. Модель багатопшарового перцептрона (Multi-Layer Perceptron, MLP) складається із декількох шарів зв'язаних між собою нейронів, кожен з яких здатен виконувати лінійну комбінацію своїх входів і за допомогою передаточної функції виводити бінарний чи дійсний результат. Кожен вхід нейрона

має адаптовану вагу, яка характеризує важливість цього входу. При навчанні входи перцептрона формуються молекулярними дескрипторами, а вихід є прогнозом активності сполуки. Радіально-базисні нейронні мережі (*Radial-Basis Function Neural Networks, RBF-NN*) складаються із вхідного шару, одного прихованого, та вихідного шару. На відміну від багатшарового перцептрона, нейрони прихованого шару не обчислюють свого виходу як добуток ваг і вхідних значень. Кожен нейрон прихованого шару визначається своїм центром, тобто точкою у просторі дескрипторів. Вихід нейрона обчислюється як функція відстані між цією точкою та аналізованою сполукою у даному просторі. Нейрони вихідного шару діють за принципом перцептрону, обчислюючи вихід як передаточну функцію добутку власних ваг та виходів радіально-базисних нейронів. Прикладами використання радіально-базисних нейронних мереж для моделювання зв'язку структура-активність може бути дослідження інгібіторів циклін-залежної кінрази 4 [15], моделювання інгібіторів зворотньої транскриптази ВІЛ-1 [16] та інгібіторів нейрамінідази [17].

2. 6. Деревя рішень

Класифікаційна модель дерева рішень (*Decision tree, DT*) складається із деревоподібної структури, що містить вузли та зв'язки між ними. В загальному, у кожному вузлі здійснюється тест із використанням одного дескриптору. На основі результатів цього тесту алгоритм спрямовується до одного із нащадків даного батьківського вузла. У вузлі-нащадку здійснюється наступний тест, і так далі поки алгоритм не зупиниться на кінцевому вузлі, який називають ще на листком. Остаточне рішення визначається класом активності, асоційованим із цим листком. Таким чином, загалом процес прийняття рішення базується на послідовності простих тестів, що складають собою шлях від кореня дерева до одного з листків. Деревя рішень використовувались при дослідженні кишкового всмоктування [9], при моделюванні інгібування агрегації β -амілоїду [18], тощо.

2. 7. Алгоритм Random Forest

За допомогою даного алгоритму здійснюється побудова ансамблю дерев рішень, поєднуючи в собі ідею *bagging*, згідно якої моделі будуються не на одній навчальній вибірці, а на декількох (репліках), отриманих із вихідної шляхом техніки бустрепінгу (*bootstrapping*), та ідею випадкових підпросторів (*Random Subspaces*), що означає використання випадково сформованих наборів дескрипторів. Класифікація сполук проводиться шляхом голосування: перемагає той клас, за який проголосувало найбільше дерев. Оптимальне число дерев та параметр кількості дескрипторів m підбирається таким чином, щоб мінімізувати похибку *out-of-bag* (*OOB*): частку сполук навчальної вибірки, що класифікується неправильно, якщо не враховувати голоси дерев, побудованих на цих сполуках як навчальних. Цей досить новий алгоритм швидко завоював популярність у *QSAR*-дослідників, і використовувався для моделювання інгібіторів *JNK3* [19], *QSAR*-аналізу протипухлинної дії 4-тіазолідинонів [20] та ін. У порівняльному аналізі *QSAR*-класифікаторів [21] *Random Forest* показав

кращий результат відносно машини опорних векторів та інших ансамблів дерев рішень.

5. Висновки

У даному огляді визначено місце аналізу зв'язків «молекулярна структура-активність» серед комп'ютерних методів розробки нових лікарських засобів та описано найбільш поширені алгоритми побудови *QSAR* моделей із акцентуванням на механізмі їх роботи. У останній час все більшої і більшої популярності набувають підходи, що базуються на використанні ансамблів моделей, прикладом яких є *Random Forest*. Прогрес у розвитку методів машинного навчання для побудови моделей «структура-активність» є запорукою подальшого розвитку напрямку *QSAR* та знаходження нових біологічно активних речовин із його допомогою.

Література

1. Dudek, A. Z. Computational methods in developing quantitative structure-activity relationships (QSAR): a review [Text] / A. Z. Dudek, T. Arodz, J. Galvez // *Combinatorial Chemistry & High Throughput Screening*. – 2006. – Vol. 9, Issue 3. – P. 213–228. doi: 10.2174/138620706776055539
2. Arakawa, M. The recent trend in QSAR modeling-variable selection and 3D-QSAR methods [Text] / M. Arakawa, K. Hasegawa, K. Funatsu // *Current Computer-Aided Drug Design*. – 2007. – Vol. 3, Issue 4. – P. 254–262. doi: 10.2174/157340907782799417
3. Tropsha, A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening [Text] / A. Tropsha, A. Golbraikh // *Current Pharmaceutical Design*. – 2007. – Vol. 13, Issue 34. – P. 3494–3504. doi: 10.2174/138161207782794257
4. Cherkasov, A. QSAR modeling: where have you been? Where are you going to? [Text] / A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin et. al. // *Journal of Medicinal Chemistry*. – 2014. – Vol. 57, Issue 12. – P. 4977–5010. doi: 10.1021/jm4004285
5. Devinyak, O. T. Quantitative structure-activity relationship study and directed synthesis of thieno[2,3-d]pyrimidine-2,4-diones as monocarboxylate transporter 1 inhibitors [Text] / O. T. Devinyak, M. V. Slivka, M. V. Slivka, V. M. Vais, V. G. Lendel // *Medicinal Chemistry Research*. – 2011. – Vol. 21, Issue 9. – P. 2263–2272. doi: 10.1007/s00044-011-9748-4
6. Девіняк, О. Т. QSAR аналіз 2(4)-тіазолідинонів із піразоліновим фрагментом в молекулах, що проявляють протипухлинну активність щодо клітин недрібноклітинного раку легень *in vitro* [Текст] / О. Т. Девіняк, Д. Я. Гаврилюк, Б. С. Зіменковський, Р. Б. Лесик // *Клінічна фармація, фармакотерапія та медична стандартизація*. – 2011. – № 3-4. – С. 163–168.
7. Li, Y. W. Structure-activity relationship study of antioxidative peptides by QSAR modeling: the amino acid next to C-terminus affects the activity [Text] / Y. W. Li, B. Li, J. He, P. Qian // *Journal of Peptide Science*. – 2011. – Vol. 17, Issue 6. – P. 454–462. doi: 10.1002/psc.1345
8. Mazanetz, M. P. Prediction of cyclin-dependent kinase 2 inhibitor potency using the fragment molecular orbital method [Text] / M. P. Mazanetz, O. Ichihara, R. J. Law, M. Whittaker // *Journal of Cheminformatics*. – 2011. – Vol. 3, Issue 1. – P. 2. doi: 10.1186/1758-2946-3-2
9. Suenderhauf, C. Combinatorial QSAR modeling of human intestinal absorption [Text] / C. Suenderhauf, F. Hamann, A. Maunz, C. Helma, J. Huwyler // *Molecular Pharmaceutics*. – 2010. – Vol. 8, Issue 1. – P. 213–224. doi: 10.1021/mp100279d

10. Sun, H. A naive Bayes classifier for prediction of multidrug resistance reversal activity on the basis of atom typing [Text] / H. Sun // *Journal of Medicinal Chemistry*. – 2005. – Vol. 48, Issue 12. – P. 4031–4039. doi: 10.1021/jm050180t
11. Prathipati, P. Global Bayesian models for the prioritization of antitubercular agents [Text] / P. Prathipati, N. L. Ma, T. H. Keller // *Journal of Chemical Information and Modeling*. – 2009. – Vol. 48, Issue 12. – P. 2362–2370. doi: 10.1021/ci800143n
12. Lv, W. Prediction of acetylcholinesterase inhibitors and characterization of correlative molecular descriptors by machine learning methods [Text] / W. Lv, Y. Xue // *European Journal of Medicinal Chemistry*. – 2010. – Vol. 45, Issue 3. – P. 1167–1172. doi: 10.1016/j.ejmech.2009.12.038
13. Yang, X.-G. Prediction of antibacterial compounds by machine learning approaches [Text] / X.-G. Yang, D. Chen, M. Wang, Y. Xue, Y.-Z. Chen // *Journal of Computational Chemistry*. – 2009. – Vol. 30, Issue 8. – P. 1202–1211. doi: 10.1002/jcc.21148
14. Boiani, M. Modeling anti-Trypanosoma cruzi activity of N-oxide containing heterocycles [Text] / M. Boiani, H. Cerecetto, M. González, J. Gasteiger // *Journal of Chemical Information and Modeling*. – 2008. – Vol. 48, Issue 1. – P. 213–219. doi: 10.1021/ci7002768
15. Saghiaie, L. Application of partial least squares and radial basis function neural networks in multivariate imaging analysis-quantitative structure activity relationship: study of cyclin dependent kinase 4 inhibitors [Text] / L. Saghiaie, M. Shahlaei, A. Madadkar-Sobhani, A. Fassihi // *Journal of Molecular Graphics and Modelling*. – 2010. – Vol. 29, Issue 4. – P. 518–528. doi: 10.1016/j.jmgm.2010.10.001
16. Chen, H. F. In silico log P prediction for a large data set with support vector machines, radial basis neural networks and multiple linear regression [Text] / H. F. Chen // *Chemical Biology & Drug Design*. – 2009. – Vol. 74, Issue 2. – P. 142–147. doi: 10.1111/j.1747-0285.2009.00840.x
17. Lü, W. J. QSAR study of neuraminidase inhibitors based on heuristic method and radial basis function network [Text] / W. J. Lü, Y. L. Chen, W.P. Ma, X. Y. Zhang, F. Luan, M. C. Liu, X. G. Chen, Z. D. Hu // *European Journal of Medicinal Chemistry*. – 2008. – Vol. 43, Issue 3. – P. 569–576. doi: 10.1016/j.ejmech.2007.04.011
18. Stempler, S. Quantitative structure-activity relationship analysis of β -amyloid aggregation inhibitors [Text] / S. Stempler, M. Levy-Sakin, A. Frydman-Marom, Y. Amir, R. Scherzer-Attali, L. Buzhansky, E. Gazit, H. Senderowitz // *Journal of Computer-Aided Molecular Design*. – 2010. – Vol. 25, Issue 2. – P. 135–144. doi: 10.1007/s10822-010-9405-x
19. Schattel, V. V. Modeling and benchmark data set for the inhibition of c-Jun N-terminal kinase-3 [Text] / V. V. Schattel, G. Hinselmann, A. Jahn, A. Zell, S. Laufer // *Journal of Chemical Information and Modeling*. – 2011. – Vol. 51, Issue 3. – P. 670–679. doi: 10.1021/ci100410h
20. Зіменковський, Б. С. Вивчення зв'язку «структура – протипухлинна активність» похідних 4-тіазолідинонів методами регресійного аналізу та класифікаційного моделювання [Текст] / Б. С. Зіменковський, О. Т. Девіняк, Р. Б. Лесик // *Журнал органічної та фармацевтичної хімії*. – 2012. – Т. 10, № 2 (38). – С. 43–49.
21. Bruce, C. L. Contemporary QSAR classifiers compared [Text] / C. L. Bruce, J. L. Melville, S. D. Pickett, J. D. Hirst // *Journal of Chemical Information and Modeling*. – 2007. – Vol. 47, Issue 1. – P. 219–227. doi: 10.1021/ci600332j
- High Throughput Screening, 9 (3), 213–228. doi: 10.2174/138620706776055539
2. Arakawa, M., Hasegawa, K., Funatsu, K. (2007). The Recent Trend in QSAR Modeling – Variable Selection and 3D-QSAR Methods. *Current Computer Aided-Drug Design*, 3 (4), 254–262. doi: 10.2174/157340907782799417
3. Tropsha, A., Golbraikh, A. (2007). Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Current Pharmaceutical Design*, 13 (34), 3494–3504. doi: 10.2174/138161207782794257
4. Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M. et. al. (2014). QSAR Modeling: Where Have You Been? Where Are You Going To? *Journal of Medicinal Chemistry*, 57 (12), 4977–5010. doi: 10.1021/jm4004285
5. Devinyak, O. T., Slivka, M. V., Slivka, M. V., Vais, V. M., Lendel, V. G. (2011). Quantitative structure-activity relationship study and directed synthesis of Thieno[2,3-d]pyrimidine-2,4-diones as monocarboxylate transporter 1 inhibitors. *Medicinal Chemistry Research*, 21 (9), 2263–2272. doi: 10.1007/s00044-011-9748-4
6. Devinyak, O. T., Havrylyuk, D.Y., Zimenkovsky, B. S., Lesyk, R. B. (2011). QSAR analiz 2(4)-тиазолідинонів із піразоліновим фрагментом в молекулах, щодо проіаєвляють протипухлинну активність швидко клітин недрибноклітинного рaku леhen in vitro [QSAR Study of 2(4)-Thiazolidinones with Pyrazoline Scaffold Possesing Antitumor Activity in vitro on Nonsmall Cell Lung Cancer Cells]. *Clinical Pharmacy, Pharmacotherapy and Medical Standardization*, 3-4, 163–168.
7. Li, Y.-W., Li, B., He, J., Qian, P. (2011). Structure-activity relationship study of antioxidative peptides by QSAR modeling: the amino acid next to C-terminus affects the activity. *Journal of Peptide Science*, 17 (6), 454–462. doi: 10.1002/psc.1345
8. Mazanetz, M. P., Ichihara, O., Law, R. J., Whittaker, M. (2011). Prediction of cyclin-dependent kinase 2 inhibitor potency using the fragment molecular orbital method. *Journal of Cheminformatics*, 3 (1), 2. doi: 10.1186/1758-2946-3-2
9. Suenderhauf, C., Hammann, F., Maunz, A., Helma, C., Huwyler, J. (2011). Combinatorial QSAR Modeling of Human Intestinal Absorption. *Molecular Pharmaceutics*, 8 (1), 213–224. doi: 10.1021/mp100279d
10. Sun, H. (2005). A Naive Bayes Classifier for Prediction of Multidrug Resistance Reversal Activity on the Basis of Atom Typing. *Journal of Medicinal Chemistry*, 48 (12), 4031–4039. doi: 10.1021/jm050180t
11. Prathipati, P., Ma, N. L., Keller, T. H. (2008). Global Bayesian Models for the Prioritization of Antitubercular Agents. *Journal of Chemical Information and Modeling*, 48 (12), 2362–2370. doi: 10.1021/ci800143n
12. Lv, W., Xue, Y. (2010). Prediction of acetylcholinesterase inhibitors and characterization of correlative molecular descriptors by machine learning methods. *European Journal of Medicinal Chemistry*, 45 (3), 1167–1172. doi: 10.1016/j.ejmech.2009.12.038
13. Yang, X.-G., Chen, D., Wang, M., Xue, Y., Chen, Y.-Z. (2009). Prediction of antibacterial compounds by machine learning approaches. *Journal of Computational Chemistry*, 30 (8), 1202–1211. doi: 10.1002/jcc.21148
14. Boiani, M., Cerecetto, H., González, M., Gasteiger, J. (2008). Modeling anti-Trypanosoma cruzi Activity of N - Oxide Containing Heterocycles. *Journal of Chemical Information and Modeling*, 48 (1), 213–219. doi: 10.1021/ci7002768
15. Saghiaie, L., Shahlaei, M., Madadkar-Sobhani, A., Fassihi, A. (2010). Application of partial least squares and radial basis function neural networks in multivariate imaging analysis-quantitative structure activity relationship: Study of cyclin dependent kinase 4 inhibitors. *Journal of Molecular Graphics and Modelling*, 29 (4), 518–528. doi: 10.1016/j.jmgm.2010.10.001

References

1. Dudek, A., Arodz, T., Galvez, J. (2006). Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. *Combinatorial Chemistry &*

16. Chen, H.-F. (2009). In Silico Log P Prediction for a Large Data Set with Support Vector Machines, Radial Basis Neural Networks and Multiple Linear Regression. *Chemical Biology & Drug Design*, 74 (2), 142–147. doi: 10.1111/j.1747-0285.2009.00840.x

17. Lü, W. J., Chen, Y. L., Ma, W. P., Zhang, X. Y., Luan, F., Liu, M. C., Chen, X. G., Hu, Z. D. (2008). QSAR study of neuraminidase inhibitors based on heuristic method and radial basis function network. *European Journal of Medicinal Chemistry*, 43 (3), 569–576. doi: 10.1016/j.ejmech.2007.04.011

18. Stempler, S., Levy-Sakin, M., Frydman-Marom, A., Amir, Y., Scherzer-Attali, R., Buzhansky, L., Gazit, E., Senderowitz, H. (2010). Quantitative structure–activity relationship analysis of β -amyloid aggregation inhibitors. *Journal of Computer-Aided Molecular Design*, 25 (2), 135–144. doi: 10.1007/s10822-010-9405-x

19. Schattel, V., Hinselmann, G., Jahn, A., Zell, A., Laufer, S. (2011). Modeling and Benchmark Data Set for the Inhibition of c-Jun N-terminal Kinase-3. *Journal of Chemical Information and Modeling*, 51 (3), 670–679. doi: 10.1021/ci100410h

20. Zimenkovsky, B. S., Devinyak, O. T., Lesyk, R. B. (2012). Vychennia vzaïmozv'язku «struktura-protypukhlyna aktyvni» pokhidnykh 4-tiazolidinoniv metodamy rehresiinoho analizu ta klasyfikatsiinoho modeliuвання [QSAR study of 4-thiazolidinones as anticancer agents using regression analysis and classification modeling]. *Journal of organic and pharmaceutical chemistry*, 10/2 (38), 43–49.

21. Bruce, C. L., Melville, J. L., Pickett, S. D., Hirst, J. D. (2007). Contemporary QSAR Classifiers Compared. *Journal of Chemical Information and Modeling*, 47 (1), 219–227. doi: 10.1021/ci600332j

Рекомендовано до публікації д-р мед. наук, професор Болдіжар О. О.
Дата надходження рукопису 17.09.2015

Девіняк Олег Теодозійович, кандидат фармацевтичних наук, завідувач кафедри фармацевтичних дисциплін, ДВНЗ «Ужгородський національний університет», пл. Народна, 3, м. Ужгород, Україна, 88000
E-mail: o.devinyak@gmail.com

УДК: 615.281:582.949.27:581.45
DOI: 10.15587/2313-8416.2015.51699

МОРФОЛОГО-ТАКСОНОМІЧНЕ ДОСЛІДЖЕННЯ РОДУ *SALVIA* ЗА ВЕГЕТАТИВНИМИ ОЗНАКАМИ

© О. М. Кошовий

Вступ. Рід *Salvia L.* нараховує понад 700 видів у світовій флорі, на території країн СНД зростає 78 видів, у флорі України рід представлено 21 видом. Усі види цього роду є ефіроолійними. Широко застосовуються та детально вивчені тільки шавлія мускатна і шавлія лікарська.

Мета. Виявити за допомогою морфолого-таксономічного аналізу за вегетативними ознаками перспективні види роду *Salvia L.* у флорі України для використання у фармацевтичній та медичній галузі.

Методи дослідження. Для виявлення перспективних видів цього роду, представлених у флорі України, для фармацевтичної та медичної галузі було проведено морфолого-таксономічне дослідження роду за вегетативними ознаками, в основі якого лежить математичний підхід (метод граф-аналізу). Для систематичної обробки даних морфологічної будови шавлій було використано морфологічні характеристики 78 видів шавлій, які представлені у флорі колишнього СРСР. У вихідну таблицю-матрицю було поміщено 615 морфологічних таксономічних ознак для 78 видів роду *Salvia L.* Проаналізовано 48585 позитивних і негативних станів таксономічних ознак.

Результати дослідження. Були розраховані коефіцієнти парної та групової спорідненості. Враховуючи найбільшу спорідненість між видами побудовано дендрограму, яка характеризує таксономічні відстані між видами роду за вегетативними ознаками і вперше побудовано ієрархічний ряд шавлій за загальними морфологічними ознаками, в основі якого знаходяться *S. adenostachya*, *S. Demetrii*, *S. Forskaehlei* та *S. scabiosifolia*, що характеризує еволюційний розвиток роду.

На основі морфолого-таксономічного дослідження виявлено основну групу таксонів (*S. scabiosifolia*, *S. adenostachya* Juz., *S. Demetrii*, *S. horminum*, *S. viridis L.L.*, *S. intercedens Pobed.*, *S. ceratophylla L.*, *S. semilanata*, *S. pratensis L.*, *S. dumetorum Andrz.*, *S. pachystachya*, *S. glutinosa*, *S. Forskaehlei*), характеристика якої узагальнює морфологічну будову роду *Salvia L.*

Висновки. Морфолого-таксономічний аналіз роду *Salvia L.* за вегетативними ознаками, показує, що рід поділяється на п'ять основних гілок, серед яких 1,8-цинеол розповсюджений у гілках, в основі яких знаходяться *S. pratensis*, *S. scabiosifolia*, *S. karabachensis* та *S. virgata*, що створює передумови для подальшого цілеспрямованого пошуку цинеоловмісної сировини серед цих груп таксонів

Ключові слова: шавлія, рід, вегетативні ознаки, морфологія, таксономія, граф-аналіз, дендрограма, ієрархічний ряд