

# ДОСЛІДЖЕННЯ ЗВ'ЯЗКУ МІЖ СТРУКТУРОЮ 4-ТІАЗОЛІДИНОНІВ ТА ЇХ ПРОТИПУХЛИННОЮ АКТИВНІСТЮ ЗА ДОПОМОГОЮ БАГАТОВИМІРНИХ АДАПТИВНИХ РЕГРЕСІЙНИХ СПЛАЙНІВ

О.Т.Девіняк\*, Р.Б.Лесик

Львівський національний медичний університет ім. Данила Галицького  
79010, м. Львів, вул. Пекарська, 69. E-mail: [dr\\_r\\_lesyk@org.lviv.net](mailto:dr_r_lesyk@org.lviv.net)

\* Державний вищий навчальний заклад «Ужгородський національний університет»

*Ключові слова:* 4-тіазолідинони; протипухлинна активність; QSAR; багатовимірні адаптивні регресійні сплайни; MARS; молекулярні дескриптори

*Використовуючи комбінування методу бэггінгу та багатовимірних адаптивних регресійних сплайнів отримано комплексну QSAR-модель протипухлинної активності 4-тіазолідинонів, яка містить 200 суб-моделей і перевищує прогностичну здатність існуючих моделей. Інтерпретація найбільш значимих дескрипторів QSAR-моделі виявила перспективні шаблони для дизайну сполук із протипухлинною активністю. Підтверджено раніше одержаний висновок, що невеликі розміри молекул та відсутність зв'язків, які характеризуються сильними дипольними моментами, підвищують шанси щодо прояву протипухлинної активності в ряду 4-тіазолідинонів. Встановлено вплив значень 3D-MoRSE дескрипторів на здатність сполук до інгібування росту клітин пухлини.*

## THE STUDY OF THE STRUCTURE – ANTICANCER ACTIVITY RELATIONSHIP OF 4-THIAZOLIDINONE DERIVATIVES USING MULTIVARIATE ADAPTIVE REGRESSION SPLINES

O. T. Devinyak, R. B. Lesyk

**Key words:** 4-thiazolidinones; anticancer activity; QSAR; multivariate adaptive regression splines; MARS; molecular descriptors

*Combining the bagging method with multivariate adaptive regression splines the complex QSAR-model that incorporates 200 submodels has been developed. The predictive ability of the new model has the superiority over already known 4-thiazolidinones anticancer activity models. The interpretation of the most significant descriptors used in the model reveals promising templates for the design of new anticancer agents. The earlier conclusion that small sizes of molecules together with the absence of bonds with high dipole moments increase the antitumor activity has been confirmed. The influence of 3D-MoRSE descriptors values on the capability of compounds to inhibit the cancer cells growth has been found.*

## ИССЛЕДОВАНИЕ СВЯЗИ МЕЖДУ СТРУКТУРОЙ 4-ТИАЗОЛИДИНОНОВ И ИХ ПРОТИВООПУХОЛЕВОЙ АКТИВНОСТЬЮ С ПОМОЩЬЮ МНОГОМЕРНЫХ АДАПТИВНЫХ РЕГРЕССИОННЫХ СПЛАЙНОВ

О.Т.Девиняк, Р.Б.Лесык

**Ключевые слова:** 4-тиазолидиноны; противоопухолевая активность; QSAR; багатомерные адаптивные регрессионные сплайны; MARS; молекулярные дескрипторы

*Используя комбинирование метода бэггинга и многомерных адаптивных регрессионных сплайнов получено комплексную QSAR-модель, которая содержит 200 субмоделей и превышает прогностическую способность известных моделей противоопухолевой активности 4-тиазолидинонов. Интерпретация наиболее значимых дескрипторов QSAR-модели обнаружила перспективные шаблоны для дизайна соединений с противораковым действием. Подтвержден ранее полученный вывод о том, что небольшие размеры молекул и отсутствие связей, что характеризуются сильными дипольными моментами, повышают шансы проявления противоопухолевой активности в ряду 4-тиазолидинонов. Установлено влияние значений 3D-MoRSE дескрипторов на способность соединений к ингибированию роста клеток опухоли.*

Тіазолідинове ядро є перспективною матрицею для створення нових сполук із високим потенціалом біологічної активності [1-4], в тому числі з протипухлинним ефектом [5-9]. Завдяки численним дослідженням науковців кафедри біоорганічної, органічної та фармацевтичної хімії ЛНМУ ім. Данила Галицького вдалося зібрати значну за обсягом базу даних «структура-дія» серед 4-тіазолідинонів та їх похідних. Ці результати стали підґрунтям для ряду обчислювальних наукових праць, що стосуються ідентифікації можливих ме-

ханізмів протипухлинної активності [10], моделювання ймовірної фармакофорної групи [11] та взаємодії з білками-мішенями [5, 12], а також QSAR-аналізу [13-17]. Розглядаючи питання опису зв'язків між структурою та протипухлинною активністю 4-тіазолідинонів, ми встановили, що апроксимація зв'язків лінійними функціями є адекватною для окремих класів сполук [13-15], однак не підходить у випадку використання всієї доступної бази даних протипухлинної активності [16]. Проведені QSAR-моделювання із використанням

гнучких алгоритмів, таких як Random Forest [16] та регресія Гаусівських процесів [17] показали свою спроможність та прогностичну здатність, однак отримані моделі містять деякі обмеження для застосування. Так, класифікаційна модель *Random Forest* здатна лише прогнозувати наявність чи відсутність протипухлинної активності без повідомлення навіть приблизного її значення, а модель Гаусівських процесів характеризується помірною точністю, що дає можливість для подальшого удосконалення. Тому метою даної роботи є створення та інтерпретація нелінійної *QSAR*-моделі, здатної прогнозувати чітке значення протипухлинної активності і в той же час характеризуватись кращими статистичними показниками, ніж існуюча модель Гаусівських процесів.

### 1. Створення моделі

Для досягнення поставленої мети було вирішено здійснити комбінування сильних сторін раніше використовуваних алгоритмів – *Random Forest* та регресії Гаусівських процесів. *Random Forest* – алгоритм машинного навчання, запропонований Лео Брейманом і Адель Катлер, що полягає у використанні комітету (ансамблю) дерев рішень [18]. Кожне з дерев рішень будується на випадковій субвибірці, отриманій з вихідної вибірки за допомогою сполучення з повтореннями (така процедура називається бутстрапом, *bootstrap*) на основі використання лише обмеженого підпростору факторів, причому побудова триває до повного вичерпання субвибірки і отримане дерево не підлягає процедурі підрізання (*pruning*). Класифікація об'єктів проводиться шляхом голосування: кожне дерево ансамблю відносить досліджуваній об'єкт до одного з класів, і перемагає клас, за який проголосувало найбільше число дерев. Таким чином, хоч одиничне дерево рішення (особливо те, що не зазнало підрізання) є слабкою моделлю, узагальнення багатьох слабких моделей, отриманих шляхом бутстрапу вихідної вибірки, дозволяє досягти високої точності. Такий підхід є сильною стороною алгоритму *Random Forest* і носить назву «беггінг», що є скороченням словосполучення англійською: *bootstrap aggregating* → *bagging*. Крім того, процедура беггінгу дозволяє встановити прогностичну здатність ансамблю на сполуках, що не були включені у ряд окремих бутстрапових моделей (*out-of-bag*). Такі статистичні показники зазвичай є досить подібними до результатів зовнішнього тестування і можуть використовуватись для вибору параметрів моделі. З іншого боку, замість дерева рішень як базової моделі ансамблю можна використати різні нелінійні регресійні моделі, що дасть змогу отримувати кількісний прогноз активності. У якості базового алгоритму в даному до-

слідженні було вибрано багатомірні адаптивні регресійні сплайни (*multivariate adaptive regression splines* – *MARS*). Цей метод здатен описати складні нелінійні зв'язки між структурою молекул та їх активністю і довів свою ефективність у багатьох дослідженнях [19-23]. Алгоритм *MARS* створює адитивну нелінійну модель наступного вигляду:

$$\hat{y} = a_0 + \sum_{m=1}^M a_m B_m(x),$$

де:  $\hat{y}$  – прогнозоване значення активності;  $a_0$  – зсув моделі;  $M$  є кількістю базисних функцій, а  $B_m$  та  $a_m$  є  $m$ -ою базисною функцією та її коефіцієнтом [24]. Кожна базисна функція, в свою чергу, є двосторонньою ступеневою функцією, що описується рівняннями:

$$\begin{aligned} [+(x-t)]_+^q &= \begin{cases} (t-x)^q, & \text{якщо } x \geq t, \\ 0, & \text{у іншому випадку} \end{cases} \\ [-(x-t)]_+^q &= \begin{cases} (t-x)^q, & \text{якщо } x < t, \\ 0, & \text{у іншому випадку} \end{cases} \end{aligned}$$

де:  $t$  є координатою вузла (точки спряження),  $q(\geq 0)$  є ступенем сплайну, а запис  $[\ ]_+$  означає додатню частину аргументу. Створення моделі *MARS* здійснюється за допомогою двох проходів: прямого і зворотного. Під час прямого проходу дескриптори додаються в модель «жадібним» чином (підключись лише про оптимальність наступного кроку), що створює модель, занадто підіграну під дані. В подальшому ця модель спрощується при зворотному проході шляхом вилучення малозначимих дескрипторів та створених на їх основі базисних функцій. Деталі роботи алгоритму *MARS* можна почерпнути з основоположної роботи *J. Friedman* [24] або інших, більш сучасних джерел [19, 22, 23].

Для проведення *QSAR*-аналізу протипухлинної активності 4-тіазолідинонів за допомогою багатомірних адаптивних регресійних сплайнів використовувались 694 сполуки, протиракова активність яких була перевірена *in vitro* на 60 лініях ракових клітин у рамках програми *DTP* Національного Інституту Раку США (*NCI*). Визначення активності проводилось за стандартною процедурою *NCI* методом флуоресцентного зафарбовування (барвник – сульфородамін Б) [25, 26] і виражалось у формі відсотків росту *GP 60* різних ліній ракових клітин (*NCI-60 panel*) порівняно із контрольним дослідом. Виходячи із вимірних значень оптичної густини на початку експерименту ( $T_z$ ), оптичної густини у контрольному досліді ( $C$ ) та досліді в присутності речовини у концентрації  $10^{-5}$ М ( $T_i$ ), значення *GP* були обчислені за формулами:

$$\frac{T_i - T_z}{C - T_z} \times 100\%, \text{ якщо } T_i \geq T_z,$$

$$\frac{T_i - T_z}{T_z} \times 100\%, \text{ якщо } T_i < T_z$$

Таким чином, значення  $GP=100\%$  свідчить про відсутність будь-якого впливу на ріст зазначеної лінії, значення  $GP=0\%$  позначає повну зупинку росту клітин пухлини, а усі значення  $GP<0\%$  означають зменшення кількості ракових клітин порівняно з їхньою кількістю на початку експерименту. Як залежна змінна, що описує активність речовини, використовувалось середнє арифметичне відсотків росту,  $\overline{GP}$ . Біля 5,2% даних щодо відсотків росту були відсутні. Ці значення зазнали відновлення шляхом багатовимірної доповнення за допомогою ланцюгових рівнянь (*Multivariate Imputations by Chained Equations – MICE*) [27]. Більше 1600 молекулярних дескрипторів обчислювались для досліджуваних сполук з використанням *E-DRAGON* [28] з попередньою оптимізацією геометрії молекул програмою *CORINA*. Враховуючи досвід попередніх *QSAR*-аналізів [15, 17], нами було додано до масиву дескрипторів ще одну індикаторну змінну, яка дорівнює одиниці, якщо кількість самовертаючих обходів 9-го порядку дорівнює 3006, і нулю в іншому випадку. Дані розподілялись на навчальну та тестову вибірки з метою перевірки прогностичної здатності майбутньої *QSAR*-моделі. Розподіл залежної змінної є сильно асиметричним із відносно малою кількістю високоактивних молекул. Тому, щоб зберегти достатньо сполук-хітів у навчальній вибірці, пропорція розмірів навчальної до тестової вибірки була встановлена на рівні 4:1. Розподіл на вибірки проводився випадковим чином. У такий спосіб розмір навчальної вибірки склав 555 сполук, а тестової – 139. Всі подальші обчислення проводились у середовищі для статистичного аналізу *R 3.0.1* [29] із використанням додаткових пакетів «*caret*» [30] та «*earth*» [31]. Кількість бутстрапових моделей у ансамблі складала 200. З метою суттєвої економії обчислювального часу проводилась процедура попереднього відбору 50 найбільш впливових дескрипторів. Для цього було використано більш швидкий алгоритм *Random Forest*. Процедуру побудови ансамблю було виконано при різних значеннях ступеня сплайну та максимальної кількості змінних у спрощеній моделі як вхідних параметрів алгоритму *MARS*. Вибір кращої комбінації параметрів проводився на основі оцінки показника  $R^2_{out-of-bag}$ . Порівнюючи отримані комплексні моделі, ми встановили, що лінійний сплайн має перевагу над квадратичним. Найкраща *QSAR* модель (рис. 1), що є ансамблем із 200 багатовимірних адаптивних регресійних сплайнів, була отримана при ступені сплайну  $q=1$  та кількості змінних у спрощеній моделі  $nprune=5$  і характеризується наступними статистичними по-

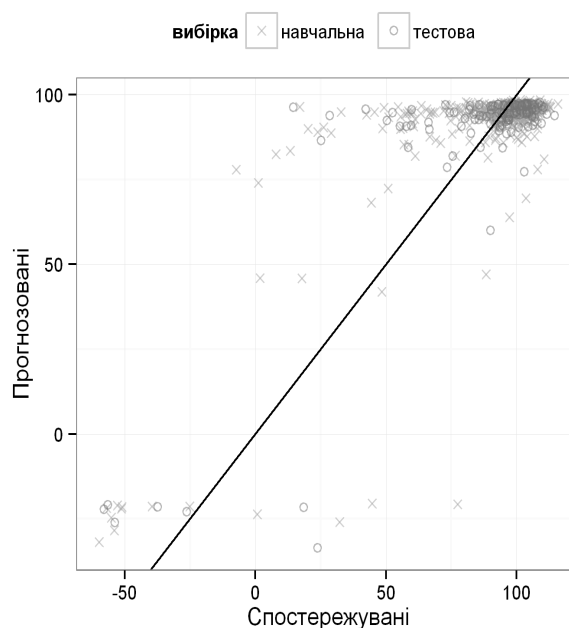


Рис. 1. Спостережувані та прогнозовані за допомогою *QSAR*-моделі значення середніх відсотків росту клітин пухлини при дії досліджуваних сполук.

казниками:  $R^2=0,687$ ,  $R^2_{out-of-bag}=0,572$ ,  $R^2_{ext}=0,675$ . Вказані значення переважають показники моделі Гаусівських процесів ( $R^2=0,656$  та  $R^2_{ext}=0,602$  після видалення декількох викидів) і підтверджують її прогностичну здатність.

## 2. Аналіз та інтерпретація моделі

При здійсненні зворотного проходу алгоритму *MARS* на кожному кроці видаляється найменш значимий дескриптор із тих, що залишились у моделі. Спрощуючи кожного учасника ансамблю – бутстрапову модель *MARS* із кількістю змінних, рівною 5, до вилучення останнього дескриптора, можна підрахувати кількість субмоделей, у яких зберігається кожен з дескрипторів –  $nsubsets$ . Так, для першого дескриптора, який буде виключено із бутстрапової моделі  $nsubsets=1$ , для другого – 2, а дескриптор, що залишиться останнім, отримає значення  $nsubsets=5$ . Сума показників  $nsubsets$  для окремого дескриптора при спрощуванні кожної бутстрапової моделі ансамблю може слугувати показником відносної важливості цього дескриптора [24]. Оскільки ансамбль містить 200 бутстрапових моделей, максимальне значення  $nsubsets$ , якого може досягти дескриптор, рівне  $5 \times 200 = 1000$ . А за допомогою частки цього дескриптора у загальній сумі  $nsubsets$  всіх дескрипторів, що були використані хоча б в одній бутстраповій моделі, можна оцінити відносний вклад кожного дескриптора (табл. 1).

Для того, щоб оцінити напрям та силу впливу дескрипторів на активність, була проведена побудова додаткової однієї моделі *MARS* із тими ж дескрипторами та параметрами алгоритму і вико-

Таблиця 1

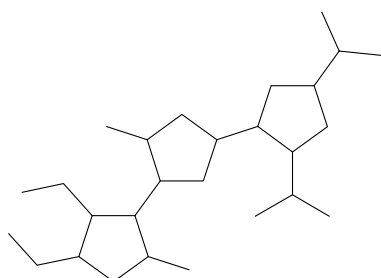
Важливості та вклади топ-10 дескрипторів у отриману комплексну QSAR-модель

Дескриптор	nsubsets	Частка, %
<i>SRW09=3006</i> (індикаторна змінна)	919	37.14
<i>nArCO</i> (кількість ароматичних кето-груп)	479	19.36
<i>EEig02d</i> (друге власне значення матриці суміжності ребер, зваженої дипольними моментами)	158	6.39
<i>Mor12v</i> (3D-MoRSE дескриптор із параметром розсіювання $11 \text{ \AA}^{-1}$ , зважений ван-дер-Ваальсовими об'ємами атомів)	153	6.18
<i>Mor11v</i> (3D-MoRSE дескриптор із параметром розсіювання $10 \text{ \AA}^{-1}$ , зважений ван-дер-Ваальсовими об'ємами атомів)	140	5.66
<i>MATS4m</i> (автокореляційний коефіцієнт Морана із зсувом 4, зважений атомними масами)	101	4.08
<i>R8v</i> (автокореляційний коефіцієнт <i>R</i> із зсувом 8, зважений ван-дер-Ваальсовими об'ємами атомів)	82	3.31
<i>Mor12u</i> (3D-MoRSE дескриптор із параметром розсіювання $11 \text{ \AA}^{-1}$ , не зважений)	65	2.63
<i>Mor16e</i> (3D-MoRSE дескриптор із параметром розсіювання $15 \text{ \AA}^{-1}$ , зважений атомними електронегативностями Сандерсона)	63	2.54
<i>EEig03d</i> (третє власне значення матриці суміжності ребер, зваженої дипольними моментами)	58	2.34

ристанням всієї навчальної вибірки (табл. 2). Ця модель містить 4 дескриптори і має статистичні показники  $R^2=0,546$ ,  $R^2_{\text{ext}}=0,638$ , які є нижчими за показники ансамблю, що підтверджує успішність вибраної стратегії щодо беггінгу багатьох моделей MARS.

Найбільш важливим дескриптором виявилась індикаторна змінна «*SRW09=3006*» із *nsubsets*=919 (табл. 1), що узгоджується із попередніми дослідженнями [15, 17]. Кількість самоповертаючих обходів 9-го порядку є топологічним дескриптором і залежить від взаємного розташування та локального оточення циклів з непарною кількістю атомів (у нашому випадку – п'ятичленних). Специфічна кількість самоповертаючих обходів 9-го порядку – 3006 відображає структурний шаблон, наведений на рис. 2. Цей структурний шаблон зустрічається у ряду кон'югатів 4-тіазолідинону, ізатину та діарилзаміщеного піразоліну (рис. 3), які у більшості випадків проявляють високу протипухлинну активність.

Другим дескриптором за важливістю є кількість ароматичних кето-груп у структурі молеку-

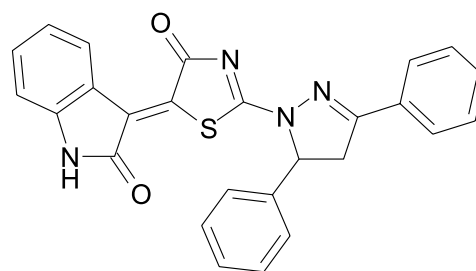
Рис. 2. Мінімальний структурний шаблон, що забезпечує *SRW09* рівний 3006.

Таблиця 2

Структура моделі MARS для визначення впливу дескрипторів

Базисні функції	Коефіцієнт
Зсув	93.61571
« <i>SRW09=3006</i> »	-117.621
$[+(nArCO) - 1]_+^1$	-54.9801
$[-(EEig02d - 3,641)]_+^1$	-38.3857
$[+(Mor12u - 0,085)]_+^1$	-43.5887
$[+(EEig02d - 4,347)]_+^1$	18.55004

ли. Так, згідно з табл. 1 відсутність цих груп або наявність однієї групи не чинить істотного впливу на активність, тоді як сполуки з двома кето-групами мають значний приріст активності (прогнозований середній відсоток росту клітин пухлини зменшується на 54,98). Серед досліджуваної бібліотеки сполук двома ароматичними кето-групами володіють декілька похідних тіопірано

Рис. 3. Спільна основа кон'югатів 4-тіазолідинону, ізатину та діарилзаміщеного піразоліну, яка володіє *SRW09=3006*.

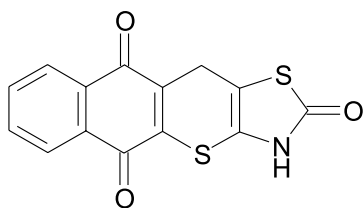


Рис. 4. Перспективний структурний шаблон похідних тіопірано[2,3-*d*][1,3]тіазол-2-ону.

[2,3-*d*][1,3]тіазол-2-онів, що є іншим перспективним шаблоном для пошуку високоактивних протипухлинних засобів (рис. 4).

Наступним за важливістю дескриптором є *EEig02d* – друге власне значення матриці суміжності ребер, зваженої дипольними моментами. Цей дескриптор відігравав істотну роль у моделі Гаусівських процесів, згідно з якою молекули з меншими значеннями *EEig02d* проявляють кращу протипухлинну активність [17]. У табл. 1 *EEig02d* зустрічається у двох базисних функціях. Згідно з функцією  $[-(EEig02d - 3,641)]_+^1$  та її коефіцієнтом, що дорівнює -38,3857, якщо значення *EEig02d* менше 3,64, то кожна одиниця різниці між поточним значенням *EEig02d* та числом 3,64 зменшує прогнозований середній відсоток росту клітин пухлини на 38,39. Якщо ж *EEig02d* спостерігається більшим, ніж 3,641, то зміни активності не відбувається. Інша базисна функція  $[(EEig02d - 4,347)]_+^1$  має додатній коефіцієнт 18,55. Це означає, що якщо значення *EEig02d* перевищує 4,347, то за кожен одиницю перевищення прогнозований відсоток росту клітин пухлини збільшується на 18,55. Узагальнюючи обидві базисні функції, можна зазначити, що при низьких значеннях *EEig02d* активність сполуки підвищується, при середніх значеннях – не змінюється, а при високих значеннях *EEig02d* протипухлинна активність зменшується. Таким чином, цей висновок співпадає із результатом регресії Гаусівських процесів, що засвідчує узгодженість обох моделей. В свою чергу, низькі значення *EEig02d* пов'язані з невеликими розмірами молекул та відсутністю зв'язків, які характеризуються сильними дипольними моментами.

Четвертий за важливістю фактор – *Mor12v* не потрапив до моделі *MARS* для визначення впливу дескрипторів, однак він тісно пов'язаний (і корельований:  $r = 0,860$ ) із *Mor12u*. Обидва фактори належать до класу *3D-MoRSE* дескрипторів і мають спільний параметр розсіювання, лише доданки дескриптора *Mor12v* зважені ван-дер-ваальсівським об'ємом атомів, а дескриптора *Mor12u* мають однакову вагу, рівні одиниці для всіх атомів.

*3D-MoRSE* дескриптори базуються на ідеї отримання інформації із *3D*-координат атомів у молекулі, використовуючи трансформацію, подібну до тієї, що застосовується при дослідженнях електронної дифракції [32-34]. Ці дескриптори обчислюються за формулою:

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin sr_{ij}}{sr_{ij}},$$

де:  $s$  – параметром розсіювання, що може набувати цілочисельних значень від 0 до  $31 \text{ \AA}^{-1}$  (тобто цифра 12 у записі *Mor12v* відображає параметр розсіювання  $11 \text{ \AA}^{-1}$ );  $r_{ij}$  – Евклідовою відстанню між  $i$ -им та  $j$ -им атомами;  $N$  – загальна кількість атомів;  $A_i$  та  $A_j$  – різними характеристиками атомів, які використовуються як ваги доданків. Таким чином, *3D-MoRSE* дескриптори відображають взаємне просторове розміщення атомів у тривимірній молекулярній структурі. Від'ємний коефіцієнт базисної функції  $[-(Mor12u - 0,085)]_+^1$ , що дорівнює -43,59, свідчить, що менші значення вказаного *3D-MoRSE* дескриптора (і, так само, дескриптора *Mor12v*, оскільки між ними існує сильна позитивна кореляція) підвищують здатність сполуки інгібувати ріст клітин пухлини.

## Висновки

1. За допомогою комбінування методу беггінгу та багатовимірних адаптивних регресійних сплайнів отримано комплексну *QSAR*-модель з високою прогностичною здатністю, яка дозволяє удосконалити підходи до раціонального дизайну потенційних протипухлинних агентів серед 4-тіазолідинонів та їх похідних.

2. Інтерпретація найбільш значимих дескрипторів моделі виявила, крім відомого перспективного структурного шаблону, представленого 3-[2-(3,5-діарил-4,5-дигідропіразол-1-іл)-4-оксо-4,5-дигідро-1,3-тіазол-5-іліден]-2,3-дигідро-1*H*-індол-2-оном, ще одну матрицю для дизайну сполук із протипухлинною активністю – тіопірано[2,3-*d*][1,3]тіазол-2-оновий фрагмент.

3. Підтверджено, що невеликі розміри молекул та відсутність зв'язків, які характеризуються сильними дипольними моментами, підвищують шанси щодо прояву протипухлинної активності в ряду 4-тіазолідинонів.

4. Встановлено вплив *3D-MoRSE* дескрипторів із параметром розсіювання  $11 \text{ \AA}^{-1}$  на здатність 4-тіазолідинонів та їх похідних до інгібування росту клітин пухлини.

**Література**

1. Lesyk R.B., Zimenkovsky B.S. // *Current Org. Chem.* – 2004. – Vol. 16. – P. 1547-1577.
2. Verma A., Saraf S.K. // *Eur. J. of Medicinal Chem.* – 2008. – Vol. 5. – P. 897-905.
3. Lesyk R., Zimenkovsky B., Kaminsky D. et al. // *Biopolymers and Cell.* – 2011. – Vol. 2. – P. 107-117.
4. Jain A., Vaidya A., Ravichandran V. et al. // *Bioorg. & Med. Chem.* – 2012. – Vol. 11. – P. 3378.
5. Lesyk R., Zimenkovsky B., Atamanyuk D. et al. // *Bioorg. & Med. Chem.* – 2006. – Vol. 15. – P. 5230-40.
6. Subtel'na I., Atamanyuk D., Szymanska E. et al. // *Bioorg. & Med. Chem.* – 2010. – Vol. 14. – P. 5090-102.
7. Kaminsky D., Khylyuk D., Vasylenko O. et al. // *Sci. Pharm.* – 2011. – Vol. 4. – P. 763-77.
8. Havrylyuk D., Zimenkovsky B., Vasylenko O. et al. // *J. of Med. Chem.* – 2012. – Vol. 20. – P. 8630-41.
9. Kryshchyshyn A., Atamanyuk D., Lesyk R. // *Sci. Pharm.* – 2012. – Vol. 3. – P. 509-29.
10. Зіменковський Б.С., Девіняк О.Т., Гаврилюк Д.Я., Лесик Р.Б. // *ЖОФХ.* – 2011. – №3. – P. 64-71.
11. Зіменковський Б.С., Девіняк О.Т., Лесик Р.Б. // *ЖОФХ.* – 2012. – №4. – P. 76-82.
12. Крищишин А.П., Драпак І.В., Зіменковський Б.С. та ін. // *Клінічна фармація, фармакотерапія та медична стандартизація.* – 2011. – №1-2. – P. 190-198.
13. Мирко І.І., Огурцов В.В., Камінський Д.В. // *Фармац. журн.* – 2010. – №4. – P. 50-54.
14. Огурцов В.В., Зіменковський Б.С., Олійник І.І. та ін. // *Фармац. журн.* – 2010. – №4. – P. 50-54.
15. Девіняк О.Т., Гаврилюк Д.Я., Зіменковський Б.С., Лесик Р.Б. // *Клінічна фармація, фармакотерапія та медична стандартизація.* – 2011. – №3-4. – P. 163-168.
16. Зіменковський Б.С., Девіняк О.Т., Лесик Р.Б. // *ЖОФХ.* – 2012. – №2. – P. 43-49.
17. Devinyak O., Zimekovsky B., Lesyk R. // *Current Topics in Med. Chem.* – 2012. – Vol. 24. – P. 2763-2784.
18. Breiman L. // *Machine Learning.* – 2001. – Vol. 1. – P. 5-32.
19. Nguyen-Cong V., Van Dang G., Rode B.M. // *Eur. J. of Med. Chem.* – 1996. – Vol. 10. – P. 797-803.
20. Jalali-Heravi M., Mani-Varnosfaderani A. // *QSAR & Combinatorial Sci.* – 2009. – Vol. 9. – P. 946-958.
21. Alamdari R.F., Mani-Varnosfaderani A., Asadollahi-Baboli M., Khalafi-Nezhad A. // *SAR and QSAR in Environmental Res.* – 2012. – Vol. 7-8. – P. 665-682.
22. Deconinck E., Xu Q.S., Put R. et al. // *J. of Pharmac. and Biomed. Analysis.* – 2005. – Vol. 5. – P. 1021-1030.
23. Xu Q.S., Daszykowski M., Walczak B. et al. // *Chemometrics and Intelligent Laboratory Systems.* – 2004. – Vol. 1. – P. 27-34.
24. Friedman J.H. // *The Annals of Statistics.* – 1991. – P. 1-67.
25. Boyd M.R., Paull K.D. // *Drug Development Res.* – 1995. – Vol. 2. – P. 91-109.
26. Shoemaker R.H. // *Nature Reviews Cancer.* – 2006. – Vol. 10. – P. 813-823.
27. Buuren van S., Groothuis-Oudshoorn K. // *J. of Statistical Software.* – 2011. – Vol. 3. – P. 98-101.
28. Tetko I., Gasteiger J., Todeschini R. et al. // *J. of Computer-Aided Molecular Design.* – 2005. – Vol. 6. – P. 453-463.
29. *R: a language and environment for statistical computing / Core Team R. – Vienna, Austria: R Foundation for Statistical Computing, 2012.*
30. Kuhn M. // *J. of Statistical Software.* – 2008. – Vol. 5. – P. 1-26.
31. Milborrow S. // *R Software Package.* – 2009. – P. 38-45.
32. Gasteiger J., Sadowski J., Schuur J. et al. // *J. of Chemical Information and Computer Sci.* – 1996. – Vol. 5. – P. 1030-1037.
33. Schuur J.H., Selzer P., Gasteiger J. // *J. of Chemical Information and Computer Sci.* – 1996. – Vol. 2. – P. 334-344.
34. Todeschini R., Consonni V. *Handbook of Molecular Descriptors.* – Wiley-VCH, 2008. – 688 p.

Надійшла до редакції 12.06.2013 р.