

УДК 541.6:547.789:615.076

АНАЛІЗ ПРОТИПУХЛИННОЇ АКТИВНОСТІ ПОХІДНИХ 4-ТІАЗОЛІДИНОНІВ ТА ЇХ КЛАСИФІКАЦІЯ ЗА ІМОВІРНИМИ МЕХАНІЗМАМИ ДІЇ МЕТОДАМИ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ

Б.С.Зіменковський, О.Т.Девіняк*, Д.Я.Гаврилюк, Р.Б.Лесик

Львівський національний медичний університет ім. Данила Галицького
79010, м. Львів-10, вул. Пекарська 69. E-mail: dr_r_lesyk@org.lviv.net

* Державний вищий навчальний заклад “Ужгородський національний університет”

Ключові слова: 4-тіазолідинони; приховані дані; нейронні мережі; протипухлинна активність

Здійснено пошук внутрішніх закономірностей та “прихованих даних” (data mining) у базі похідних 4-тіазолідинонів з протипухлинною дією. На основі аналізу головних компонентів та нейромережевого моделювання показано, що активність сполук може реалізуватись шляхом двох самостійних механізмів дії та одного змішаного. Сформовані раціональні масиви для здійснення майбутнього QSAR моделювання.

ANALYSIS OF THE ANTITUMOR ACTIVITY OF 4-THIAZOLIDINONE DERIVATIVES AND THEIR CLASSIFICATION BY POSSIBLE MECHANISMS OF ACTIVITY USING MATHEMATIC MODELLING METHODS

B.S.Zimenkovsky, O.T.Devinyak, D.Ya.Havrylyuk, R.B.Lesyk

A search of intrinsic regularities and data mining in the results of the anticancer activity evaluation of 4-thiazolidinone derivatives has been carried out. Based on the Principal Components Analysis and neurone network approaches two different mechanisms of the antitumor activity and a mixed one have been determined. The rational data arrays for future QSAR investigations have been formed.

АНАЛИЗ ПРОТИВООПУХОЛЕВОЙ АКТИВНОСТИ ПРОИЗВОДНЫХ 4-ТИАЗОЛИДИНОНОВ И ИХ КЛАССИФИКАЦИЯ ПО ВЕРОЯТНЫМ МЕХАНИЗМАМ ДЕЙСТВИЯ МЕТОДАМИ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Б.С.Зименковский, О.Т.Девиняк, Д.Я.Гаврилюк, Р.Б.Лесык

Осуществлен поиск внутренних закономерностей и “скрытых данных” (data mining) в базе производных 4-тиазолидинонов с противоопухолевым эффектом. С помощью анализа главных компонент и нейронных сетей обнаружено два различных механизма противоопухолевого действия и один смешанный. Сформированы рациональные массивы для проведения будущего QSAR моделирования.

Одним з перспективних напрямків наукових досліджень похідних 4-тіазолідинонів є вивчення їх протипухлинної активності [1-6]. Ключовим фактором успіху у цьому напрямку є спрямований синтез нових сполук на підставі результатів QSAR-аналізу та молекулярного докінгу [7-9]. Значний обсяг одержаного нами експериментального матеріалу у галузі протипухлинних 4-тіазолідинонів та споріднених гетероциклічних систем став підґрунтям для проведення *in silico* досліджень, які полягають у виявленні досі невідомих закономірностей в базі однорідних даних та формуванні нових знань, і носить назву *data mining* (виявлення прихованих даних) [10]. Метою здійсненого аналізу було формування раціонального фундаменту для подальшого моделювання біологічної активності 4-тіазолідинонів, встановлення можливості прояву різних механізмів протипухлинної дії, кла-

сифікація активних молекул за виявленими механізмами та характеристика моделей чутливості утворених класів сполук. Всі розрахунки здійснено з використанням комп'ютерної програми SciLab 5.3.0.

Синтетичні дослідження, проведені нашою групою, дозволили запропонувати ряд нових напрямків молекулярного дизайну біологічно активних 4-тіазолідинонів та споріднених гетероциклічних систем, а також одержати понад 5000 нових сполук. Одержана реальна бібліотека гетероциклічних похідних стала об'єктом вивчення на предмет протиракової активності у рамках міжнародної наукової програми DTP (Developmental Therapeutic Program) Національного інституту раку США (NCI) [11-15]. На теперішній час протестовано 1076 сполук, серед яких 402 (37,4%) проявили активність різного рівня [2]. Згідно зі стандартним

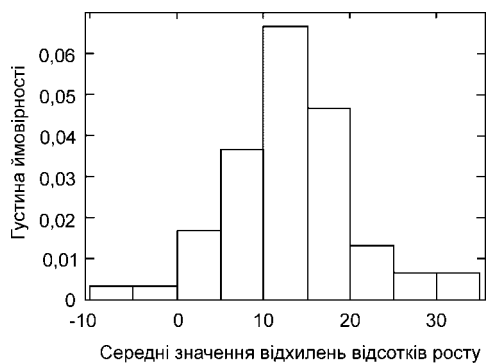


Рис. 1. Розподіл середніх значень відхилень відсотків росту за окремими лініями ракових клітин.

протоколом вивчення протипухлинної активності за програмою DTP [11, 12] сполуки, що проявили на етапі прескринінгу високий ступінь пригнічення росту ракових клітин у концентрації 10^{-5} М, проходять ґрунтове тестування на 60 ліній ракових клітин у градієнті концентрацій 10^{-4} - 10^{-8} М. Враховуючи повторюваність дослідження сполук у концентрації 10^{-5} М, ми висунули гіпотезу **1**: результати ґрунтового тестування протипухлинної активності сполуки в концентрації 10^{-5} М є однорідними та співставимими із результатами претестування. У випадку підтвердження гіпотези можливі наступні наслідки: 1) результати обох експериментів можна об'єднати при побудові QSAR-моделі, збільшивши загальний об'єм вхідних даних; 2) розходження результатів для однакових сполук дає можливість знайти похибку експерименту; 3) знайдена похибка експерименту є мінімально можливою похибкою для будь-якої достовірної QSAR-моделі (якщо середньоквадратична похибка моделі менша за середньоквадратичну похибку експерименту, отже більша відповідність (overfitting) є очевидною [16]).

Згідно з теоремою про центральну границю середнє значення відхилень результатів будь-яко-

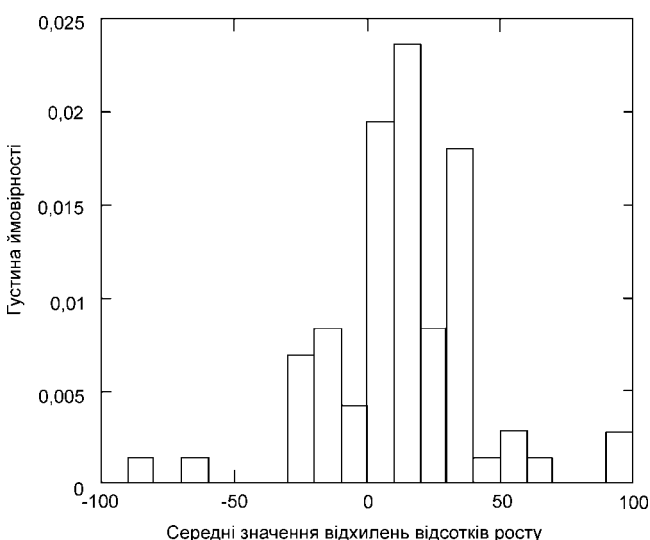


Рис. 2. Розподіл середніх значень відхилень відсотків росту за окремими сполуками.

го експерименту прямує до нуля, якщо кількість випробувань прямує до нескінченності. Таким чином, гіпотеза **1** вірна, якщо вірною є наступна нульова гіпотеза: відхилення в результатах двох тестувань є випадковою вибіркою із нормального розподілу із середнім значенням, рівним нулю, і невідомою дисперсією. Перевірка даної нульової гіпотези здійснювалась за допомогою *t*-тесту Стюдента для кожної з ліній ракових клітин. Загалом використано результати 73 пар випробувань (кожна пара випробувань містить відсотки росту 60 ліній клітин, отриманих на етапах претестування та ґрунтового скринінгу однієї і тієї ж сполуки). Результати показали, що із 5% рівнем імовірності нульова гіпотеза виявилась хибною для 41 із 60 культур. У той же час для інших 19 ліній ця гіпотеза є вірною або ж недостатньо даних, щоб її відхилити. Розподіл середніх значень відхилень за окремими лініями представлений на рис. 1.

Середнє з наведених середніх значень становить 13,18, тобто дані ґрунтового тестування показують у загальному менший відсоток росту ракових клітин. Наведені результати дають змогу відхилити гіпотезу **1** в цілому і стверджувати, що вхідні дані є неоднорідними і не можуть використовуватись разом при побудові QSAR-моделі. Найбільш імовірною причиною цього є певна відмінність у методиці та/або умовах проведення прескринінгу та ґрунтового тестування протипухлинної активності. Проте порівняння результатів двох експериментів дає можливість оцінити кількість грубих прорахунків. На рис. 2 зображено розподіл середніх значень відхилень відсотків росту за окремими сполуками. Крайні від'ємні та додатні значення відхилень вказують на грубі прорахунки в цілому для сполуки (на I чи II етапах досліджень). Вважаючи грубим прорахунком випадок, коли результати на окремій пухлинній лінії для однієї сполуки відрізняються більш як на 100%, нами виявлено 152 грубих прорахунки серед 3774 результатів (4,03%).

Наступним завданням наших досліджень була спроба класифікації похідних 4-тіазолідинонів за ймовірними механізмами протипухлинної активності. Побудова QSAR-моделі полягає у пошуку залежності біологічної активності сполуки від певних характеристик молекули, так званих дескрипторів. Існування такої залежності зумовлене тим, що біологічна активність є наслідком фізико-хімічної взаємодії ксенобіотика із біомішенями живого організму. Для того, щоб існувала афінність до конкретного протеїну, молекула-ліганд повинна володіти певним набором оптимальних значень молекулярних дескрипторів. Для різних біомішеней ці молекулярні характеристики відрізняються між собою, тому при вивченні кореляції "структура-активність" насамперед необхідно з'ясувати можливість прояву досліджуваної активності за участю не одного, а кількох механізмів дії.

Таблиця

Пояснювана розбіжність перших 10 головних компонентів (PC)

Головні компоненти	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Пояснювана розбіжність, %	67,48	3,40	2,28	2,08	1,92	1,86	1,72	1,64	1,42	1,24
Зміна пояснюваної розбіжності*, %	64.08	1.19	0.20	0.17	0.05	0.14	0.08	0.22	0.19	0.07

* Зміна пояснюваної розбіжності (i) = Пояснювана розбіжність (i) - Пояснювана розбіжність (i+1).

Відома система прогнозування механізму протипухлинної активності за алгоритмом *COMPARE* [11, 12] здійснює віднесення досліджуваної молекули до механізму дії, властивого відомим лікарським засобам. В основі цієї системи знаходиться твердження, що в залежності від механізму одні лінії ракових клітин є більш чутливими, а інші — менш чутливими до дії протипухлинного агента. Проведений нами раніше *COMPARE*-аналіз [4, 5] показав, що для активних 4-тіазолідинонів неможливо з високою ймовірністю прогнозувати жоден відомий механізм дії. Тому для виявлення можливості прояву декількох, у тому числі невідомих механізмів протипухлинної активності 4-тіазолідинонів, використано метод аналізу головних компонентів (*principal components analysis, PCA*) [17]. Даний метод дає можливість знайти такі лінійні комбінації змінних (головні компоненти, *principal components, PC*), що проекція початкових даних на отримані вектори буде мати максимальну дисперсію. PCA передбачає істинність наступних тверджень: 1) в залежності від механізму одні лінії ракових клітин є більш чутливими, а інші — менш чутливими до дії протипухлинного агента. Тобто різні механізми дії мають різні моделі (зразки) чутливості; 2) похибка визначення відсотків росту ракових клітин є меншою за різницю у моделях чутливості для різних механізмів.

Очевидно, що перший PC (PC1) буде охоплювати інформацію про середній відсоток росту (надалі середню активність) сполуки, оскільки варіабельність між активними та неактивними 4-тіазолідинонами значно більша за варіабельність між різними механізмами дії сполук. Наступні PC

будуть охоплювати інформацію про різні механізми (якщо їх $\epsilon > 1$) та відхилення експерименту. Критерієм розділення варіабельності між механізмами та відхиленням експерименту є зміна пояснюваної розбіжності при порівнянні двох суміжних PC. Перед проведенням PCA загальний масив даних було очищено від відсутніх значень шляхом видалення рядка (дані про тестування певної сполуки на різних лініях) або стовпця (дані про тестування різних сполук за певною лінією ракових клітин), керуючись принципом мінімізації втраченої інформації. При такому підході для проведення аналізу залишилися результати тестування 487 сполук на 38 лініях. Ці дані були стандартизовані за лініями, тобто приведені до одиначної дисперсії та нульового середнього значення з метою забезпечення однакового впливу результатів кожної з досліджуваних ліній на процес аналізу. Одержані PC характеризуються відсотками пояснюваної розбіжності (табл.).

Як зазначалось вище, PC1 охоплює інформацію щодо середньої активності сполуки. Зміна пояснюваної розбіжності PC2 більш ніж у 5 раз перевищує аналогічні показники інших PC. Даний факт є індикатором наявності двох різних механізмів протипухлинної активності. Зміни пояснюваної розбіжності усіх інших PC перебувають на проміжку [0;0.22]. Таким чином, інші PC інкапсулюють похибки експерименту. Проекція досліджуваних сполук на площину PC1 та PC2 (70.88% від сукупної розбіжності) наведена на рис. 3, причому еліпсами окреслено орієнтовні кластери сполук, активність яких реалізується завдяки двом різним механізмам. Проекція точки з відсотком

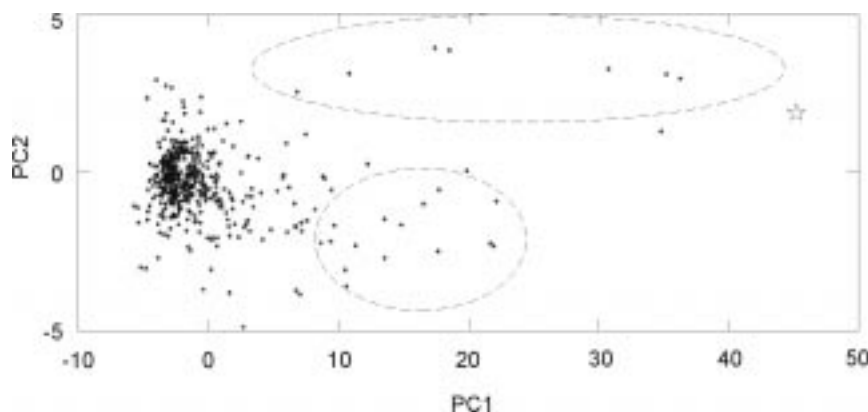


Рис. 3. Проекція досліджуваних сполук на площину PC1 та PC2. Зірочками позначено проекції точок із відсотками росту 100% або -100%.

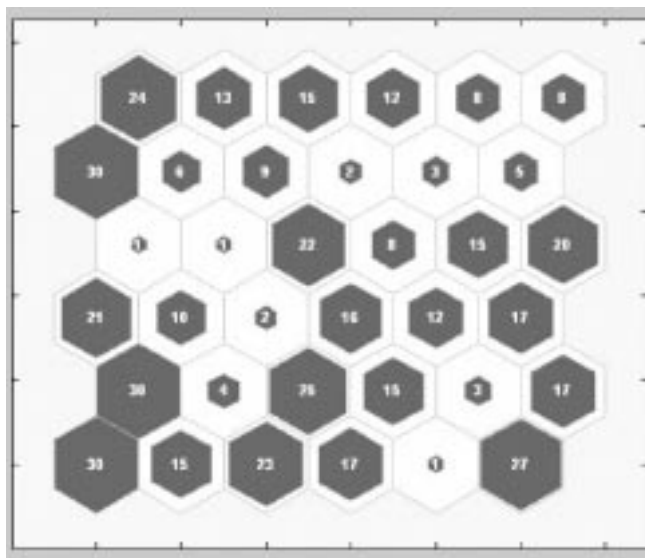


Рис. 4. Розподіл досліджуваних сполук у нейронній мережі. Діаметр забарвленої зони пропорційний кількості сполук у кожному кластері.

росту -100% (позначена зірочкою) є зміщеною від осі PC1 в бік одного з механізмів. Це зумовлюють дві взаємодоповнюючі обставини: 1) ближній кластер охоплює механізм дії з меншою варіабельністю чутливостей клітинних ліній відносно другого механізму, тому притягує до себе дану точку; 2) дальній кластер містить значно більше активних сполук, тому притягує до себе вісь PC1.

Однак із наведеного (рис. 3) неможливо визначити приналежність сполук до одного з кластерів та встановити можливу реалізацію їх активності за двома механізмами дії. Крім того, питання чіткого окреслення меж кластерів також залишилося невирішеним. Для вирішення даної проблеми застосовано моделювання за допомогою нейронних мереж. Зокрема, проводилась кластеризація шляхом нескерованого навчання (*unsupervised learning*) методом "самоорганізованих карт Кохонена" (*Cohonen's Self-Organizing Maps, SOM*) [18].

Кожен нейрон такої мережі охоплює певну кількість сполук, споріднених за зразком чутливості. Перед здійсненням навчання показники біологічної активності були стандартизовані за сполуками. Наведена процедура відкидає дані про середню активність, причому на перший план виходить інформація про відносну чутливість, а сполуки з подібними зразками складатимуть один кластер. Однак через відсутність інформації про середню активність в одному кластері поряд з активними сполуками буде присутня певна кількість неактивних, що потрапили внаслідок ймовірного характеру похибки експерименту (їх модель чутливості випадково схожа на модель активних речовин). Розподіл усіх 487 досліджуваних сполук у даній нейронній мережі зображено на рис. 4.

Проаналізувавши дані щодо середніх значень активності, ми виявили, що максимум досягає

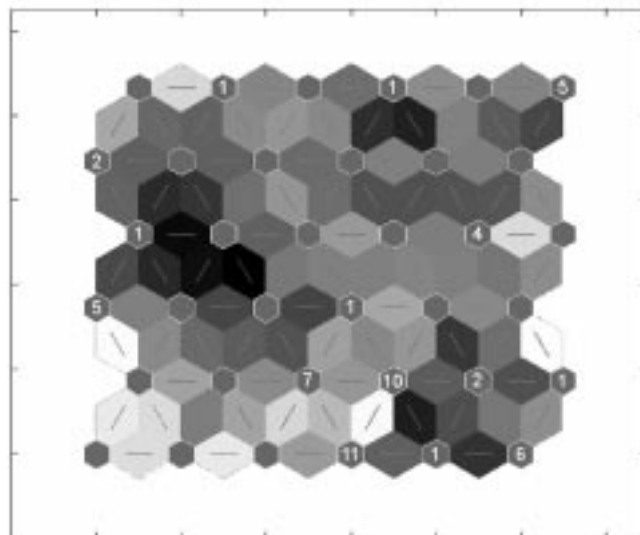


Рис. 5. Карта відстаней між нейронами мережі. Найближчі сусіди з'єднані білим кольором, найдальші - чорним. Числами позначено кількість сполук у кластері, що проявляють протипухлинну активність.

близько 120% росту. Таким чином, можна очікувати, що максимальне відхилення середніх значень активності для неактивних сполук сягає $\approx 20\%$, тому всі сполуки із середнім відсотком росту ракових клітин $< 80\%$ певною мірою проявляють протипухлинну активність. Таке проведення меж між активними та неактивними сполуками дає змогу виявити розташування молекул, що гальмують ріст пухлин, у створеній нейронній мережі (рис. 5).

На рис. 5 чітко виділяється одна велика сукупність активних сполук (3 нейрони) та декілька менших. Хоча слід зауважити, що важливість кожного нейрона (з точки зору пошуку найбільш активних сполук) слабко пов'язана із кількістю інкапсульованих ним активних сполук, адже інгибування росту пухлинних клітин характеризується дуже широким діапазоном. Тому введено поняття інтегральної активності (IA) сполук нейрона, яка є сумою вкладів охоплених нейроном молекул і визначається за формулою:

$$IA_i = -\sum (GI_i - 100),$$

де: IA — інтегральна активність сполук i -го нейрона; GI_i — середня активність сполуки, яка належить i -му нейрону.

Вклад неактивної сполуки із середнім відсотком росту 100% в інтегральну активність сполук нейрона рівний 0. Вклад гіпотетичної сполуки із середнім відсотком росту -100% (максимально можлива активність) в інтегральну активність сполук нейрона становить 200.

Поверхня інтегральної активності сполук над площиною нейронної мережі, згладжена за допомогою кубічної інтерполяції (рис. 6).

На рис. 6 вирізняються три вершини (**A**, **B**, **B**), розташовані на одній лінії. З огляду на результати

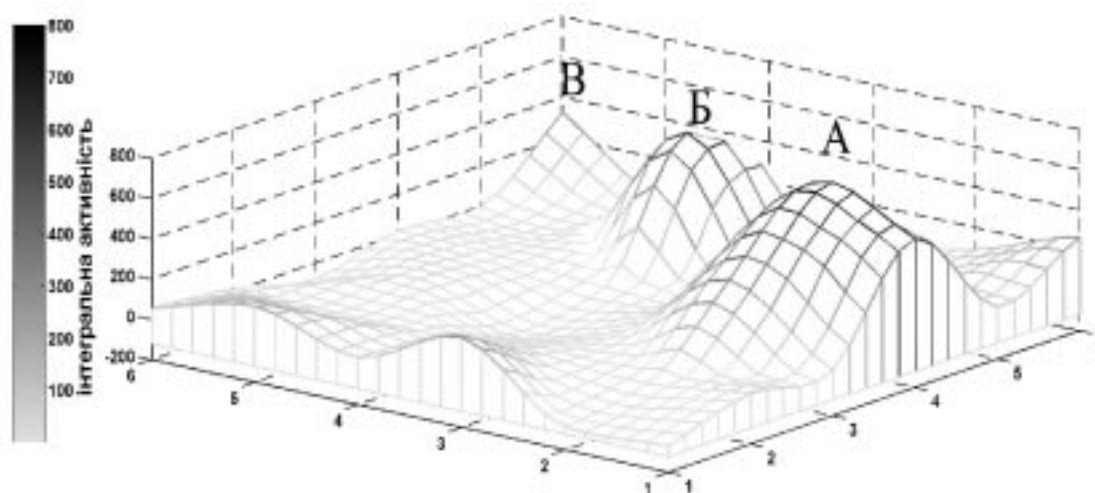


Рис. 6. Поверхня інтегральної активності сполук над площиною нейронної мережі.

аналізу головних компонентів можна стверджувати, що для масиву протипухлинних 4-тіазолідонів реалізуються два різні механізми інгібування росту (А і В) ракових клітин та один змішаний (Б). Застосування нейронної логіки, на відміну від

РСА, дозволяє провести між ними чітку межу. Крім того, по краях площини нейронної мережі знаходяться три класи із невисокою інтегральною інтенсивністю, тобто існують й інші механізми протипухлинної активності досліджуваних спо-

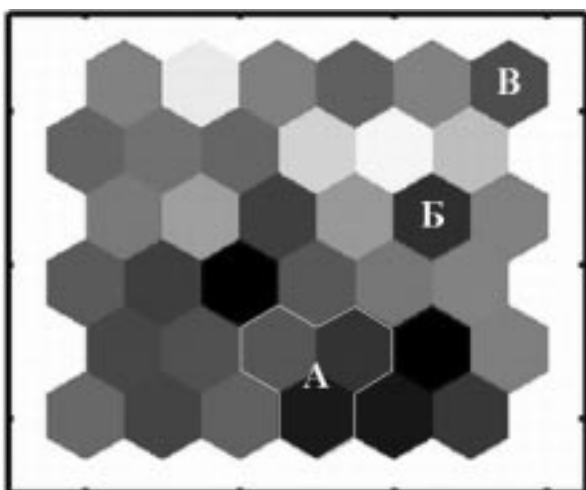


Рис. 7а. Площина ваги лінії NCI-H460*.

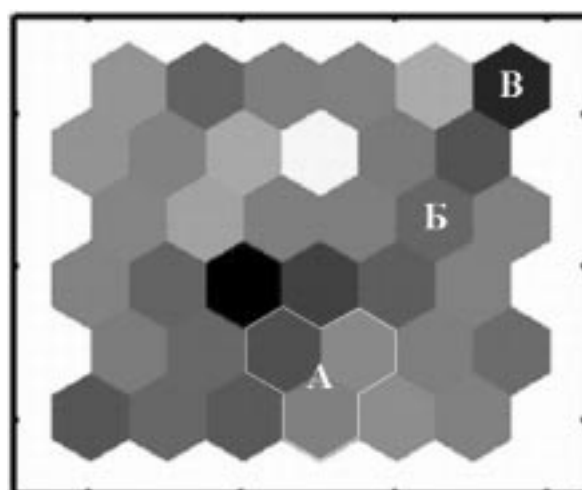


Рис. 7б. Площина ваги лінії MDA-MB-231/ATCC*.

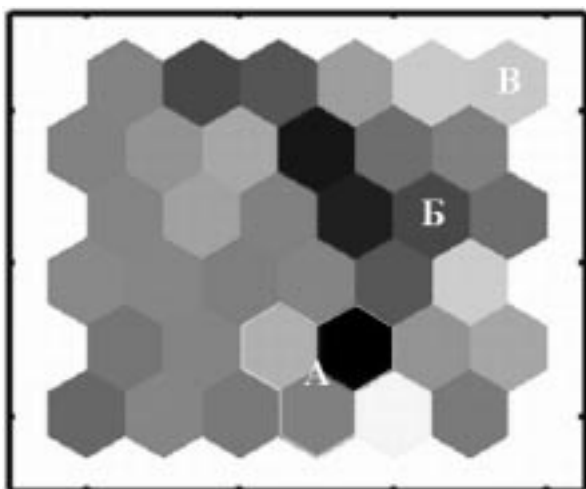


Рис. 7в. Площина ваги лінії M14*.

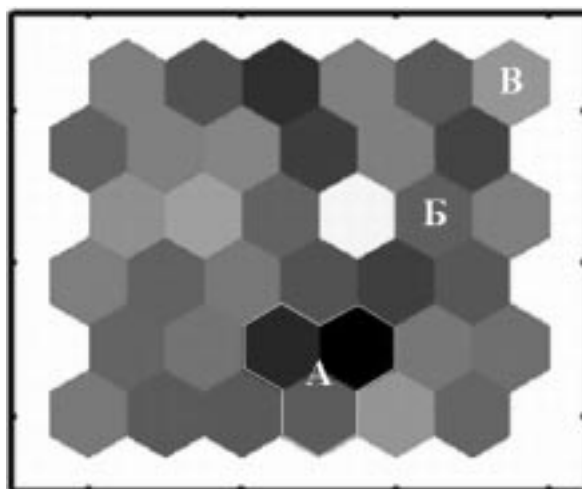


Рис. 7г. Площина ваги лінії SK-MEL-5*.

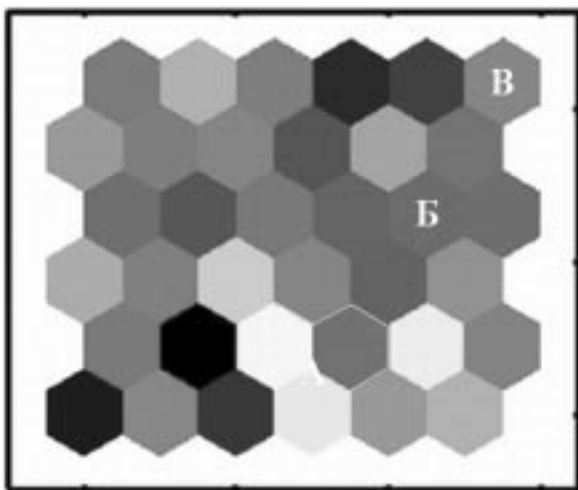


Рис. 7д. Площина ваги лінії NCP-62*.

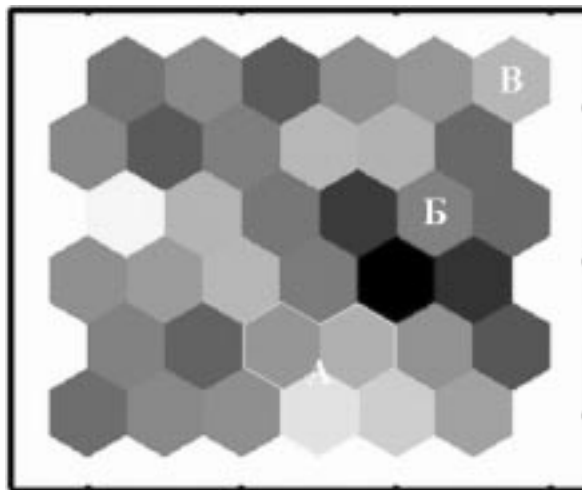


Рис. 7е. Площина ваги лінії OVCAR-5*.

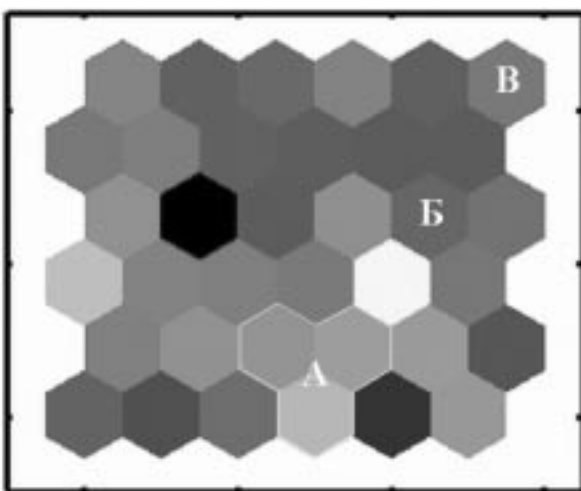


Рис. 7з. Площина ваги лінії SK-OV-3*.

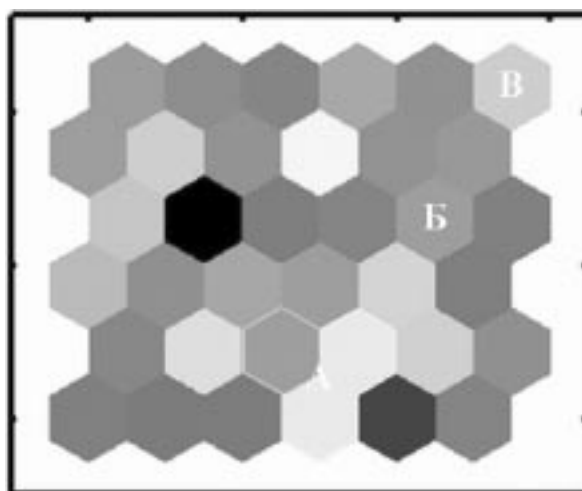


Рис. 7ж. Площина ваги лінії SNB-19*.

* Темні кольори відповідають більшій вазі, а світлі - меншій, тобто клітини аналізованої лінії більш чутливі до дії сполук з темних кластерів та більш резистентні до дії сполук із світлих кластерів у порівнянні з іншими онкоклітинами.

лук, які однак представлені малою кількістю молекул-хітів. Площини ваги кожної лінії клітин у нейронній мережі (рис. 7а-ж) дозволяють проаналізувати моделі чутливості активних сполук.

Представники усіх трьох аналізованих кластерів проявляють певну селективність дії щодо лінії ракових клітин NCI-H460 (рис. 7а), проте лише активні сполуки кластера **В** характеризуються селективністю щодо MDA-MB-231/ATCC (рис. 7б). Лінії M14 та SK-MEL-5 вибірково інгібуються активними сполуками із класу **А** (рис. 7в,г). З іншого боку, до цитостатичної дії представників класу **А** найбільш резистентними є клітини NCP-62, OVCAR-5, SK-OV-3 та SNB-19 (рис. 7д-ж), а до дії речовин із групи **В** — клітини M14, OVCAR-5 та SNB-19 (рис. 7в,е,ж). Ваги чутливості молекул кластера **Б** завжди знаходяться в інтервалі, утвореному вагами сполук з механізмами дії **А** та **В**, що підтверджує гіпотезу про змішаний механізм інгібування росту пухлин представниками **Б**. Представники класів сполук **А**, **Б**, **В** наведені на рис. 8.

Таким чином, щоб здійснити спрямований синтез на основі результатів QSAR аналізу, QSAR-моделі повинна з достатньою точністю відрізнити активні сполуки від неактивних. Така модель повинна будуватись окремо для кожного з трьох кластерів сполук з виділеними механізмами активності, тому наступним завданням наших досліджень було формування трьох масивів даних для QSAR-аналізу. У кожен з них повинні входити активні сполуки лише з відповідного класу. У той же час неактивні сполуки у всіх масивах будуть однаковими.

Поняття неактивності в контексті даної роботи являє собою середній відсоток росту клітин пухлини $GP = 100\% \pm SE$ (стандартна похибка експерименту). Причому сукупність значень GP для різних неактивних сполук повинна мати нормальний розподіл Гауса. Вибірка, визначена нами вище як неактивна (масив середніх результатів із значенням $>80\%$), не відповідає вказаним вимогам (нормальний розподіл та середнє значення рівне 100%) на основі t -тесту. Середнє значення

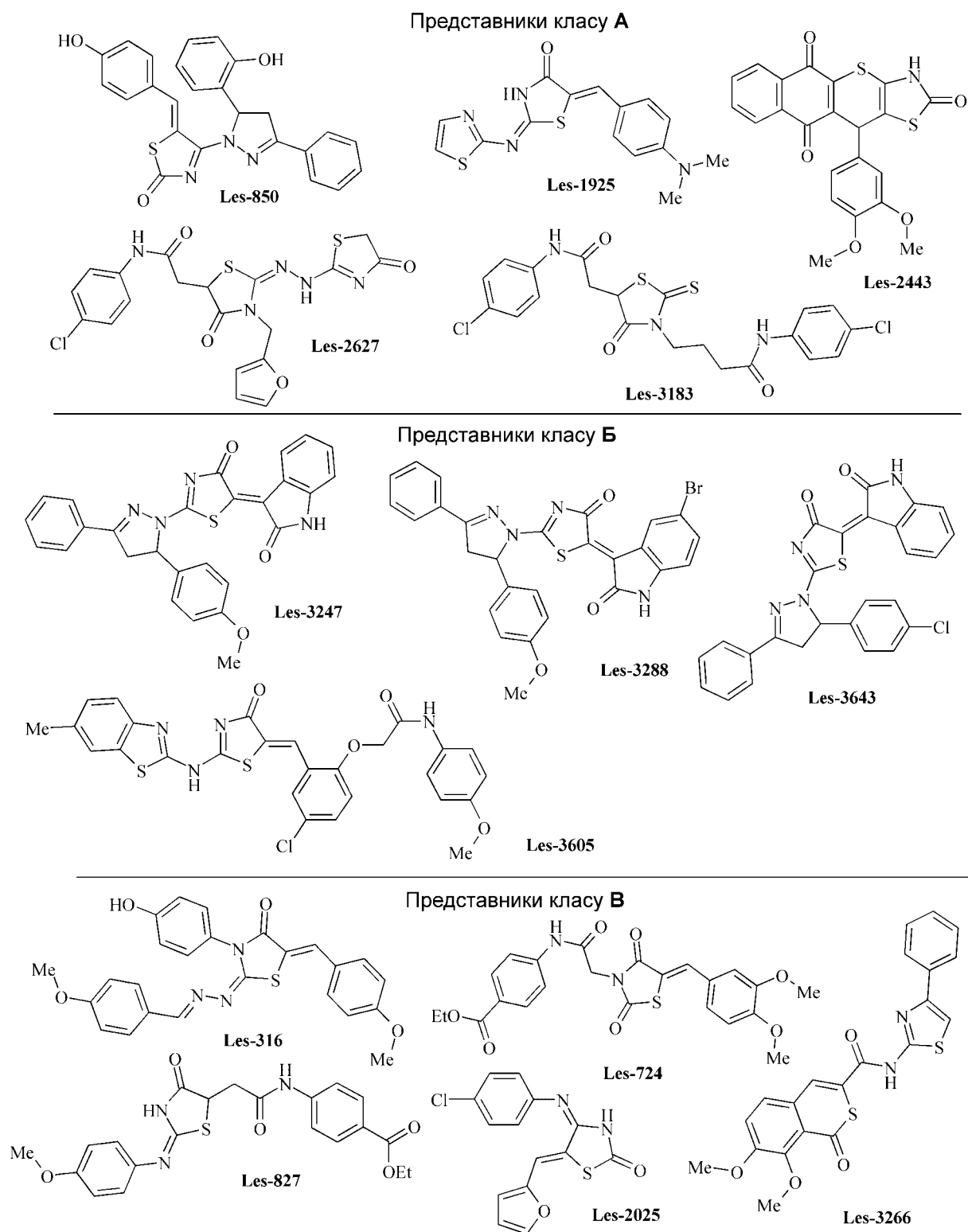


Рис. 8. Представники класів сполук А, Б, В.

цієї вибірки становить 98.48. Тому, використавши обчислення *t*-критерію та поступово підвищуючи поріг відсотка інгібування, здійснено пошук масиву, для якого нульова гіпотеза про нормальний розподіл із середнім значенням, рівним 100%, не може бути відкинута з рівнем достовірності $\alpha = 0.05$. Таким чином, сформовано вибірку із 426

неактивних сполук, GP яких перевищує встановлений мінімальний поріг, рівний 86% (середнє значення GP становить 99.60). Для класу А масив даних становить 454 сполуки (426 неактивних та 28 активних), для класу Б — 430 сполук (426 неактивних та 4 активних), а для механізму В — 431 сполука (426 неактивних та 5 активних).

Висновки

1. Проведено порівняльний аналіз результатів претестування та ґрунтового *in vitro* скринінгу протипухлинної активності сполук на 60 лініях ракових клітин у концентрації 10^{-5} М. За допомогою статистичних критеріїв доведено неоднорідність досліджуваних вибірок, що означає неприпустимість їх поєднання при здійсненні QSAR-досліджень. Показано, що близько 4% результатів

тестування похідних 4-тіазолідинонів як потенційних протиракових агентів (за програмою DTP) є грубими прорахунками.

2. На основі аналізу головних компонентів та нейромережевого моделювання встановлено, що протипухлинна активність похідних 4-тіазолідинонів реалізується шляхом двох самостійних механізмів дії та одного змішаного, що можна розглядати як основу для раціонального пошуку біомішеней та створення точних і адекватних QSAR-моделей.

Література

1. Зіменковський Б.С., Лесик Р.Б. 4-Тіазолідони. Хімія, фізіологічна дія, перспективи. — Вінниця: Нова книга, 2004. — 106 с.
2. Lesyk R.B., Zimenkovsky B.S., Kaminsky D.V. et al. // *Biopolymers and Cell*. — 2011. — Vol. 27, №2. — P. 107-117.
3. Гаврилюк Д.Я., Лесик Р.Б., Матійчук В.С., Обушак М.Д. // *ЖОФХ*. — 2006. — Т. 4, вип. 2 (14). — С. 42-47.
4. Havrylyuk D., Mosula L., Zimenkovsky B. et al. // *Eur. J. Med. Chem.* — 2010. — Vol. 45. — P. 5012-21.
5. Havrylyuk D., Zimenkovsky B., Vasylenko O. et al. // *Eur. J. Med. Chem.* — 2009. — Vol. 44, №4. — P.1396-1404.
6. Kaminsky D., Zimenkovsky B., Lesyk R. // *Eur. J. Med. Chem.* — 2009. — Vol. 44. — P. 3627-3636.
7. Камінський Д.В., Роман О.М., Атаманюк Д.В., Лесик Р.Б. // *ЖОФХ*. — 2006. — Т. 4, вип. 1 (13). — С. 41-48.
8. Гаврилюк Д.Я., Зіменковський Б.С., Драпак І.В. та ін. // *Фармац. журн.* — 2009. — №6. — С. 69-75.
9. Мосула Л.М., Зіменковський Б.С., Огурцов В.В. та ін. // *Фармац. журн.* — 2010. — №2. — С. 77-83.
10. Kantardzic M. *Data Mining: Concepts, Models, Methods, and Algorithms*. — John Wiley & Sons, 2003. — 345 p.
11. Shoemaker R. // *Nature Rev. Cancer*. — 2006. — №6. — P. 813-823.
12. <http://dtp.nci.nih.gov>.
13. Boyd M.R., Paull K.D. // *Drug Develop. Res.* — 1995. — Vol. 34. — P. 91-109.
14. Alley M.C., Scudiero D.A., Monks P.A. et al. // *Cancer Res.* — 1988. — Vol. 48. — P. 589-601.
15. Grever M.R., Schepartz S.A., Chabner B.A. // *Seminars in Oncol.* — 1992. — Vol. 19, №6. — P. 622-638.
16. Everitt B. *Cambridge Dictionary of Statistics*. — CUP, 2002. — 410 p.
17. Jolliffe I. *Principal Component Analysis, 2-nd ed.* — Springer, NY, 2002. — 487 p.
18. Kohonen T. *Self-Organizing Maps (Third Extended Ed.)*. — New York, 2001. — 501 p.

Надійшла до редакції 24.06.2010 р.