

МОДЕЛЮВАННЯ МОЖЛИВОЇ ФАРМАКОФОРНОЇ ГРУПИ 4-ТІАЗОЛІДИНОНІВ З ПРОТИПУХЛИННОЮ АКТИВНІСТЮ

Б.С.Зіменковський, 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

* Державний вищий навчальний заклад «Ужгородський національний університет»

Ключові слова: 4-тіазолідинони; протипухлинна активність; кластерний аналіз; моделювання фармакофора

На основі кластерного аналізу бібліотеки 4-тіазолідинонів з протипухлинною активністю виділено стабільну та статистично значимо відмінну групу сполук зі спільним механізмом онкоцитотоксичної дії. Розроблено комплексну фармакофорну модель, що містить координати двох ароматичних чи інших циклів з π -зв'язками, координати гідрофобної групи та двох проєкцій донорів водневого зв'язку.

MODELING OF A POSSIBLE PHARMACOPHORE GROUP FOR 4-THIAZOLIDINONES WITH THE ANTICANCER ACTIVITY

B.S.Zimenzkovsky, O.T.Devinyak, R.B.Lesyk

Based on cluster analysis of in-home library of 4-thiazolidinones with the anticancer activity the stable and statistically significantly specific group of compounds possessing the cytotoxic action through common mechanism has been isolated. For this group a complex pharmacophore model that contains the coordinates of two aromatic or other cycles of π -bonds, the coordinates of a hydrophobic group and two projections of the hydrogen bond donors have been developed.

МОДЕЛИРОВАНИЕ ВОЗМОЖНОЙ ФАРМАКОФОРНОЙ ГРУППЫ 4-ТИАЗОЛИДИНОНОВ С ПРОТИВООПУХОЛЕВОЙ АКТИВНОСТЬЮ

Б.С.Зименковский, О.Т.Девиняк, Р.Б.Лесык

На основе кластерного анализа библиотеки 4-тиазолидинонов с противоопухолево́й активностью выделено стабильную и статистически значимо отличающуюся группу соединений с общим механизмом онкоцитотоксического действия. Разработана комплексная фармакофорная модель, содержащая координаты двух ароматических или других циклов с π -связями, координаты гидрофобной группы и двух проєкций доноров водородной связи.

Пошук антинеопластичних засобів серед похідних тіазолідину та споріднених гетероциклічних систем є ключовим напрямком досліджень кафедри фармацевтичної, органічної і біоорганічної хімії ЛНМУ ім. Данила Галицького [1-3]. Грунтовний аналіз взаємозв'язку «структура-активність» дозволив виявити певні закономірності та окреслити напрямки наступних *in silico* досліджень. Так, у попередній роботі [4] створено та верифіковано класифікаційну модель протипухлинної активності 4-тіазолідинонів. Аналіз прототипів класів показав, що для розділення активних та неактивних сполук необхідно враховувати чималу кількість дескрипторів молекулярної будови, серед яких найбільш важливими виявились топологічні. Зокрема, встановлено залежність протиракової дії 4-тіазолідинонів від кількості та взаємного розміщення п'ятичленних циклів, подвійних і ароматичних зв'язків, а також термінальних атомів/груп, що мають зв'язок лише з одним сусіднім атомом не враховуючи Гідрогенів. Таким

чином, дескриптори, що описують характеристики графа молекули, суттєво переважають значення дескрипторів, які визначають фізико-хімічні властивості в цілому. Наведений результат наводить на думку, що одержана нами модель Random Forest [4] може описати структуру фармакофорної групи. Тому метою даної роботи стало моделювання можливої фармакофорної групи протипухлинних 4-тіазолідинонів, яке проводилось за наступним алгоритмом (рис. 1).

1. Попереднє опрацювання даних

Вихідними даними для проведення фармакофорного пошуку стали структури 598 молекул та результати тестування протипухлинної активності цих сполук у рамках міжнародної наукової програми DTP (Developmental Therapeutic Program) Національного інституту раку США (NCI) [5]. Згідно зі стандартним протоколом програми DTP кожна сполука проходила тестування щодо можливості інгібування росту онкоклетин близько 60 різних

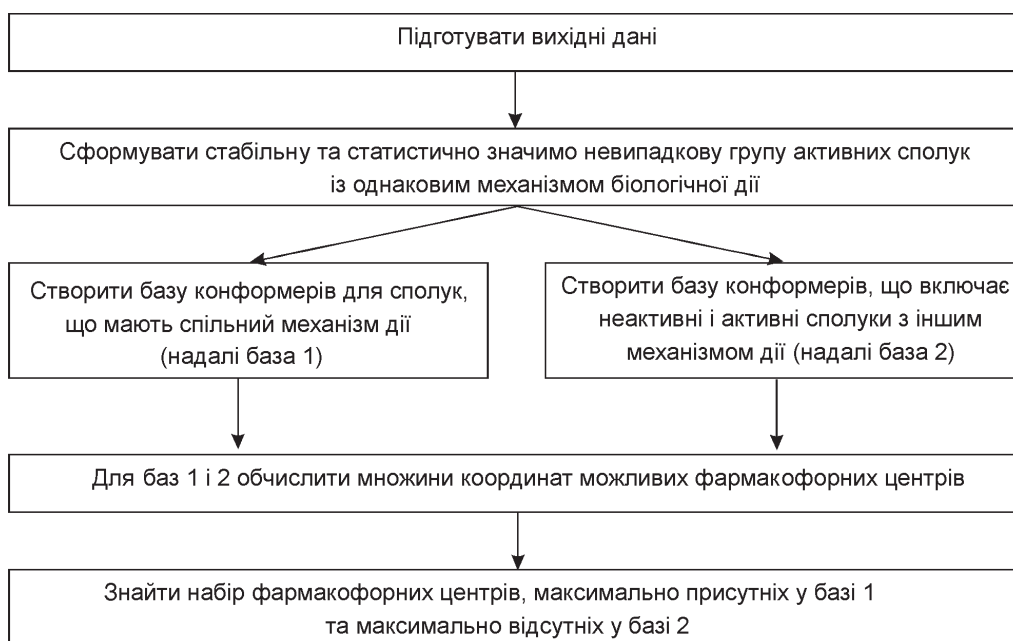


Рис. 1. Алгоритм проведення фармакофорного пошуку.

ліній, що характеризують різні анатомо-морфологічні класи неоплазій [6]. Результат вимірювався як відсоток росту клітин пухлини відносно контрольного інтактного досліджу. Рациональна межа між активними та неактивними сполуками була встановлена нами раніше і становила $\overline{GP}=86\%$ [7]. Подальший статистичний аналіз активності проводився за допомогою середовища R (версія 2.13.0) [8]. Близько 3,7% результатів для різних ліній були відсутні у протоколах досліджень. Щоб не втрачати корисної інформації внаслідок видалення сполук чи ліній із відсутніми відсотками росту, зазначені дані були відновлені на основі аналізу наявних результатів. Одним із ефективних методів для проведення такої операції є багатовимірне доповнення за допомогою ланцюгових рівнянь (Multivariate Imputation by Chained

Equations – MICE) [9]. Доповнені таким чином дані стали об'єктом подальших обчислювальних досліджень.

2. Пошук групи активних сполук із однаковим механізмом біологічної дії

На першому етапі здійснено розподіл активних сполук та підвибірки високоактивних сполук ($\overline{GP}<20\%$, ця межа виявилась оптимальною при побудові моделі Random Forest [4]) у самоорганізованій карті Кохонена (рис. 2), отриманій раніше [7].

На рис. 2 окреслено та позначено літерами орієнтовні кластери сполук із однаковими імовірними механізмами активності. Варто зазначити, що кластер Б складається тільки із високоактивних сполук. Крім того, головним чином лише цей

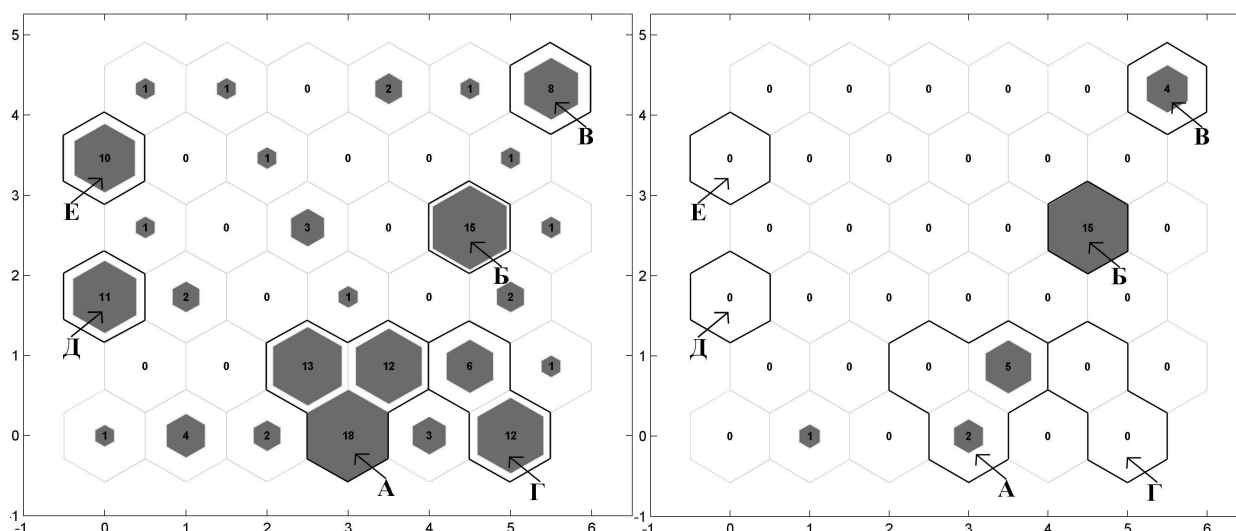


Рис. 2. Розподіл сполук із середніми відсотками росту онкоклітин $<86\%$ та $<20\%$, відповідно, у нейронній мережі (самоорганізованій карті Кохонена).

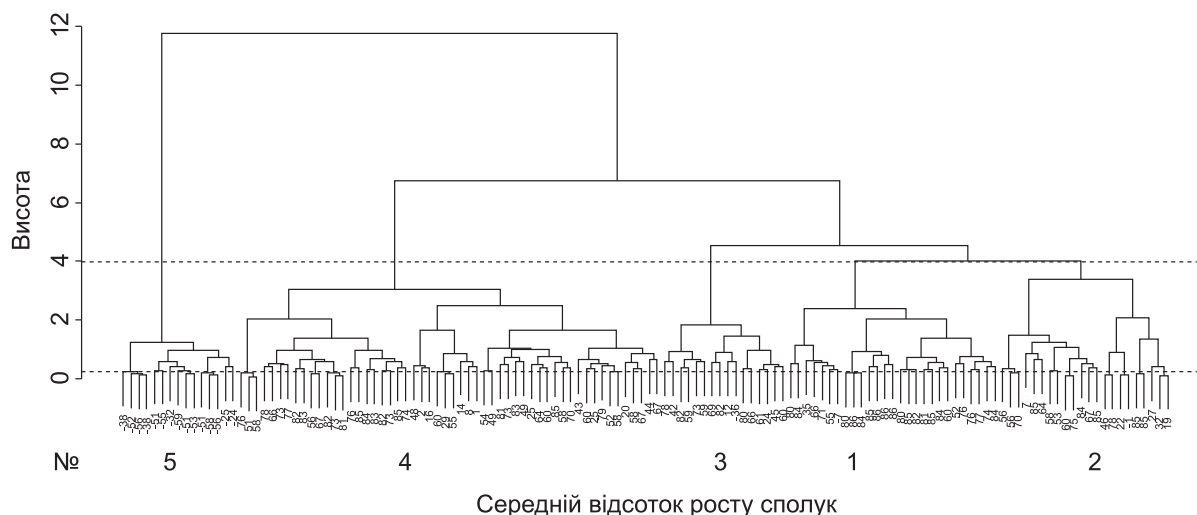


Рис. 3. Дендрограма протипухлинної активності досліджуваних сполук. Пунктиром позначено максимальну та мінімальну висоту розгалуження множини дендрограм симульованих об'єктів.

механізм активності описувала модель Random Forest. Тому саме цей кластер став об'єктом нашого дослідження. Щоб підтвердити не випадковість вказаної групи сполук, застосовувався ієрархічний агломеративний кластерний аналіз вибірки активних ($GP < 86\%$) молекул із використанням кореляційної відстані як індикатора віддаленості сполук за механізмом дії та методу Уарда для визначення відстані між кластерами. Вибір кореляційної відстані ($1-r(a,b)$, де $r(a,b)$ – коефіцієнт кореляції Пірсона між векторами відсотків інгібування росту онкоклетин сполуками a та b відповідно) зумовлений характерними її особливостями, а саме: нечутливістю до очікуваної наявності різних середніх значень та дисперсій відсотків росту у сполук одного механізму дії і в той же час урахуванням метричності залишків (тобто врахування, наскільки сильно відхиляються значення середніх відсотків росту обох сполук від спільної лінійної моделі на відміну від рангових кореляційних коефіцієнтів). Використаний нами метод Уарда визначає відстань між двома кластерами, як приріст суми квадратів відстані від кожного об'єкта до центру об'єднаного кластера відносно суми квадратів відстані від кожного об'єкта до центру відповідного йому меншого кластера.

$$\Delta(A, B) = \sum_{i \in A \cup B} \|\bar{x}_i - \bar{m}_{A \cup B}\|^2 - \sum_{i \in A} \|\bar{x}_i - \bar{m}_A\|^2 - \sum_{i \in B} \|\bar{x}_i - \bar{m}_B\|^2 = \frac{n_A n_B}{n_A + n_B} \|\bar{m}_A - \bar{m}_B\|^2,$$

де: \bar{m}_j – центром кластера j ; n_j – кількість об'єктів у ньому; $\Delta(A, B)$ також ще називають «вартістю злиття» кластерів A і B .

Таким чином, метод Уарда є оптимальним для формування компактних сферичних кластерів [10]. З іншого боку, якщо вважати, що існує група сполук із однаковим механізмом дії та певним паттерном чутливості онкоклетин, то відхилення цих

сполук від спільного паттерну будуть випадковими, а отже очікується саме сферична форма кластерів.

Паралельно, щоб встановити коректну кількість кластерів (визначити коректну висоту їх поділу), необхідно дослідити нульову гіпотезу про відсутність будь-яких закономірних паттернів чутливості серед досліджуваних сполук. Для цього проведено генерацію сполук із випадковими паттернами чутливості, однак із тим самим розподілом відсотків росту (перестановочний тест Монте-Карло). З цією метою для кожної сполуки здійснено випадкову перестановку відсотків росту ракових клітин різних ліній. Загалом створення множини об'єктів, що виходять із нульової гіпотези, та наступне їх порівняння із досліджуваним об'єктом є методом перевірки статистичних гіпотез, альтернативним до статистичних критеріїв. У нашому випадку кількість таких об'єктів становила 1000. Консолідовані результати ієрархічного агломеративного кластерного аналізу реальних та симульованих видів активності представлено на рис. 3. Усі сполучення вище інтервалу випадкових сполучень позначають не випадково різні кластери молекул, а усі сполучення нижче – є не випадково однаковими за механізмом дії.

П'ять кластерів сполучаються вище лінії максимуму, що і є необхідною коректною кількістю. Присутність вузлів вище цієї лінії також дозволяє відкинути нульову гіпотезу про те, що усі сполуки діють за одним механізмом. З огляду на дендрограму (рис. 3) існує 5 не випадково розмежованих кластерів, однак для твердження щодо існування саме п'яти механізмів цей аргумент є недостатнім. Нижче від лінії мінімальної висоти випадкового розгалуження знаходяться групи сполук, що мають не випадково схожий паттерн чутливості. Серед цих груп цікавими є сполуки, що відрізняються середньою активністю на величини

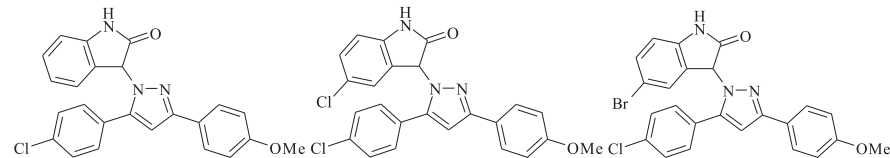
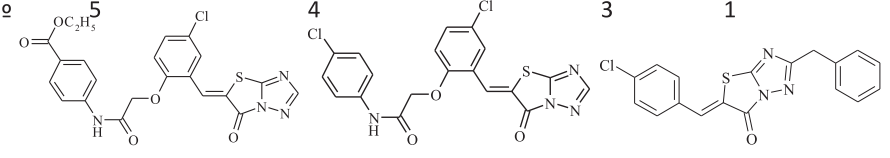
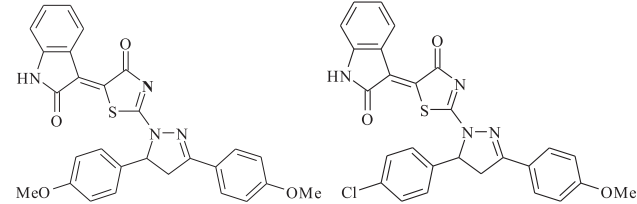
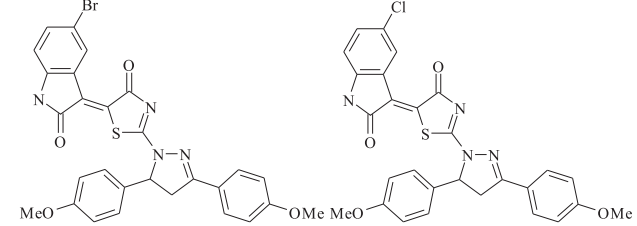
 <p>Les-4025, \overline{GP} = 76% Les-4026, \overline{GP} = 51% Les-4027, \overline{GP} = 58%</p>	група 1
<p>№ 5</p>  <p>Les-2798, \overline{GP} = 60% Les-2785, \overline{GP} = 29% Les-2829, \overline{GP} = 55%</p>	група 2
 <p>Les-3970, \overline{GP} = 46% Les-3974, \overline{GP} = 78%</p>	група 3
 <p>Les-3971, \overline{GP} = 22% Les-3988, \overline{GP} = -1%</p>	група 4

Рис. 4. Групи сполук із статистично значимо однаковим механізмом дії та різною величиною протипухлинної активності.

ну, більшу від похибки експерименту. Адже у такому випадку є достатні статистичні підстави стверджувати, що виявлена різниця у структурі цих сполук дійсно має вплив на активність за одним і тим же механізмом. Для цього, взявши сукупність неактивних сполук ($\overline{GP} > 86\%$), обчислимо для них стандартне відхилення, рівне 5,63%. Це відхилення можна прийняти тотожним стандартній похибці експерименту. З теорії ймовірностей відомо, що нормально розподілена випадкова величина у 99,73% випадків буде набувати значення, що знаходиться в межах трьох стандартних відхилень. Цього достатньо, щоб прийняти, що у даному контексті поняття «відрізняються не випадково» означає «відрізняються більш ніж на $3 \cdot 5,63\% \approx 17\%$ ». Така різниця значень середнього відсотка росту виявлена для сполук, які віднесено до чотирьох споріднених груп (рис. 4).

Так, у групі 1 введення атома галогену до фрагменту ізатину призводить до підвищення протипухлинної активності. Аналогічна картина спостерігається у випадку присутності атома Хлору в 4 положенні фенілацетамідного фрагменту (гру-

па 2), однак протилежна – у випадку *n*-хлорфенільного (положення 5 піразолінового залишку) замісника у молекулі Les-3974 (група 3). У групі 4 тривіальна, на перший погляд, заміна атома Бром на Хлор привела до суттєвого покращення здатності до інгібування росту онкоклетин. Слід також зазначити, що механізм дії сполук першої групи істотно відрізняється від механізму третьої та четвертої, а отже наявність 4-тіазолідинового ядра відіграє ключову роль у реалізації онкостатичного впливу сполуками третьої та четвертої груп.

При детальному аналізі дендрограми (рис. 3) серед виявлених класів спостерігається компактний та добре відділений від інших кластер №5, що містить 15 онкоцитотоксичних (усі $\overline{GP} < -20\%$) сполук. Він є тотожним кластеру Б на самоорганізованій карті Кохонена (рис. 2) і є найбільш цікавим для пошуку фармакофорної групи. Однак відомо, що ієрархічний агломеративний кластерний аналіз належить до «жадібних» алгоритмів, які на кожному кроці піклуються лише про оптимальність безпосередньо наступного стану, а не

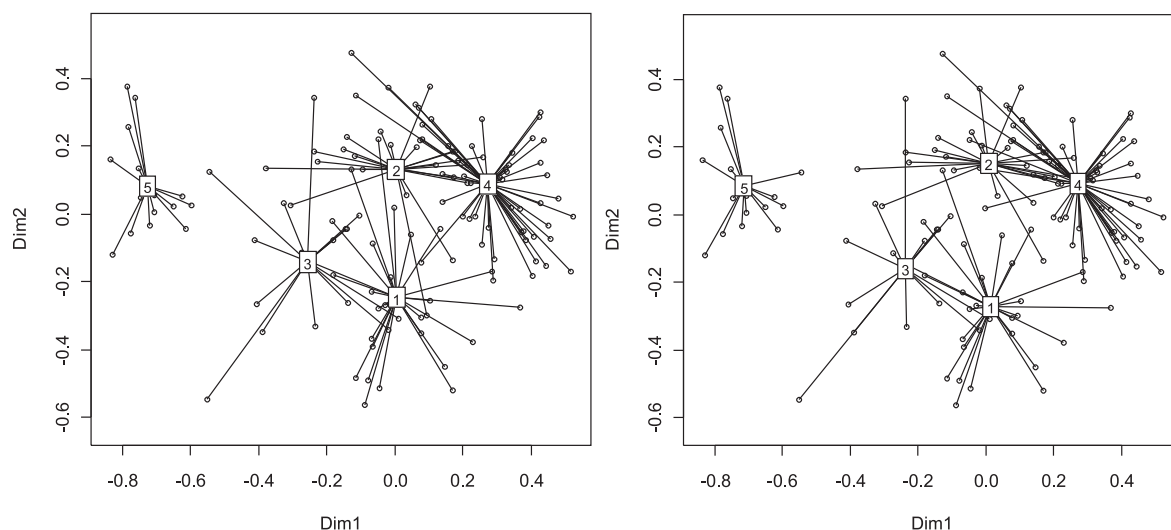


Рис. 5. Розміщення та структура кластерів до та після оптимізації на площині перших двох головних координат.

Таблиця 1

Стабільності кластерів

№ кластера	1	2	3	4	5
Стабільність, %	0,8531111	0,7197391	0,9344000	0,8696038	0,9785000

кінцевого результату, тому здійснене віднесення сполук до різних кластерів часто відхиляється від оптимального. Поширеним є підхід, який полягає у встановленні числа кластерів за допомогою ієрархічного кластерного аналізу з наступною оптимізацією отриманого поділу за допомогою методу *k*-середніх. Взаємне розміщення кластерів до та після оптимізації, апроксимоване за допомогою багатовимірного масштабування [11], наведено на рис. 5.

Найбільш суттєвою відмінністю після оптимізації є внесення ще однієї сполуки (шістнадцятої, яка раніше була у третьому кластері) до п'ятого кластера. Також помітним є достатнє відділення цього кластера від усіх інших. Таким чином, ми сформували статистично значимо не випадкову оптимальну групу високоактивних сполук. Стабільність усіх кластерів (і, головним чином, п'ятого) можна оцінити, додавши до кожного значення відсотка росту для кожної сполуки так званий «шум» – випадкову величину, одержану із нормального розподілу з нульовим центром та стандартним відхиленням, рівним стандартному відхиленню відповідної лінії для сукупності неактивних сполук. Отримавши 1000 вибірок, оптимізуємо кожну з них методом *k*-середніх із заданими кластерами. Середня частка сполук, які залишались у тих самих кластерах, слугувала оцінкою стабільності (табл. 1).

Достатньою стабільністю для коректного пошуку фармакофорної групи (вважаємо мінімально прийнятним значенням 95%) володіє тільки п'ятий кластер. Саме він є шуканою стабільною

та не випадковою групою сполук з однаковим механізмом активності.

3. Створення баз конформерів та обчислення для них координат можливих фармакофорних центрів

Бази конформерів 1 та 2 були сформовані за допомогою стохастичного конформаційного пошуку із використанням МОЕ [12]. Зазначений метод створення конформацій полягає у виконанні випадкових поворотів кожного доступного для обертання зв'язку з кроком у 30° і подальшої градієнтної мінімізації енергії отриманої структури, яка зупиняється при досягненні коренем середньоквадратичного відхилення градієнта (RMS gradient) значення, меншого від 0,001. Конформери, енергія яких перевищувала мінімальне знайдене значення для даної молекули більше, ніж на 7 ккал/моль, визначались як енергетично невідповідні. Ймовірністю існування таких конформерів в умовах біологічної системи можна знехтувати, тому вони автоматично вилучались із баз.

Для кожного з конформерів встановлювалися координати усіх центрів ароматичних циклів чи циклів з π-зв'язками, усіх гідрофобних груп, а також акцепторів та донорів водневого зв'язку та їх відповідних проєкцій.

4. Встановлення структури фармакофора

Подальше виявлення фармакофорної групи дозволило встановити ряд наборів фармакофорних центрів, що здійснюють коректне відділення сполук п'ятого кластера (одного механізму) від 582

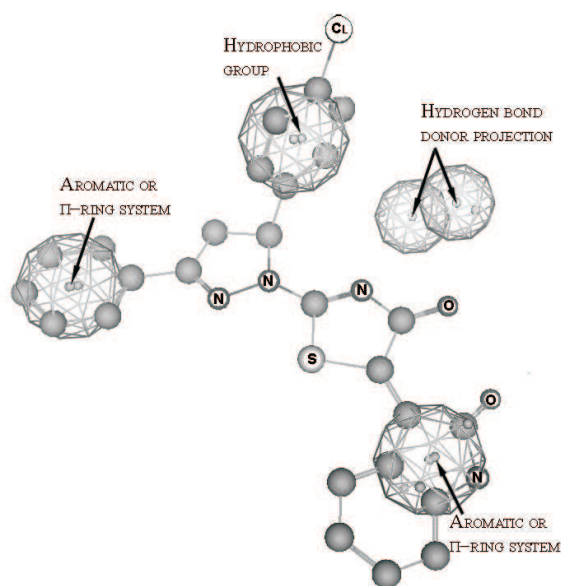


Рис. 6. Сполука Les-3989 узгоджена з розташуванням фармакофора.

інших. Найбільшою точністю характеризується фармакофор, що містить координати двох ароматичних чи інших циклів з π-зв'язками, координати однієї гідрофобної групи та двох проєкцій донора водневого зв'язку (акцептора електронної пари) (рис. 6).

Матриця помилок наведеної фармакофорної моделі (табл. 2) свідчить про її високу точність та адекватність.

Слід зазначити, що наявність двох донорів водневого зв'язку у структурі біомішені, відстань між якими (рис. 7) близька до 1.5Å (така відстань співмірна із довжиною σ-зв'язку), є малоюмовірною. Вважаємо, що отримана модель вказує на присутність області зі значною електронною густиною, що утворюється за рахунок неспарених p-електронів атомів Нітрогену та Оксигену (обидва входять до структурного фрагменту 4-тіазолідину). Електронодефіцитний атом Гідрогену біомішені, потрапивши в таку область, утворює достатньо сильний водневий зв'язок, стабілізований резонансом між двома можливими структурами, в одній з яких донором електронної пари виступає Нітроген, а в іншій – Оксиген. Ароматичний цикл, π-зв'язок та гідрофобний фрагмент можуть відігравати певну роль у приєднанні молекули до мішені за рахунок слабкої Ван-дер-Ваальсової взаємодії.

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. Havrylyuk D., Kovach N., Zimenkovsky B. et al. // *Archiv der Pharmazie*. – 2011. – Vol. 344, №8. – P. 514-522.
4. Зіменковський Б.С., Девіняк О.Т., Лесик Р.Б. // *ЖОФХ*. – 2012. – Т. 10, вип. 2 (38). – С. 43-49.

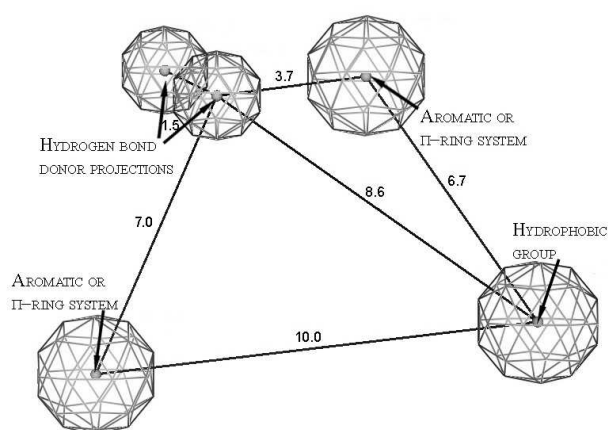


Рис. 7. Відстані між фармакофорними центрами протитуберкульозних 4-тіазолідинів (Å).

Таблиця 2

Матриця помилок фармакофорної моделі 1

Висновки	Кластер №5	Інші
Узгоджуються з фармакофором	14	3
Не узгоджуються з фармакофором	2	579

Висновки

1. За допомогою ієрархічного кластерного аналізу у поєднанні з методом *k*-середніх сформовано стабільну та статистично значимо не випадкову групу 4-тіазолідинів, що можуть проявляти однаковий механізм протипухлинної активності.

2. Запропоновано фармакофорну модель протитуберкульозних 4-тіазолідинів, що містить координати двох ароматичних чи інших циклів з π-зв'язками, координати однієї гідрофобної групи та двох проєкцій донора водневого зв'язку (акцептора електронної пари). Зазначена фармакофорна модель може розглядатись як теоретична платформа для раціонального дизайну молекул із потенційною протипухлинною активністю.

3. Вперше висунуто припущення про ключову роль атомів Нітрогену та Оксигену 4-тіазолідинового фрагменту, як донорів електронної пари при взаємодії з біомішенню, що може розглядатись як аргумент на користь віднесення «тіазолідинової» матриці до групи базових молекулярних фрагментів для спрямованого пошуку протитуберкульозних лікарських засобів.

5. <http://dtp.nci.nih.gov>
6. Shoemaker R. // *Nature Rev. Cancer*. – 2006. – №6. – P. 813-823.
7. Зіменковський Б.С., Девіняк О.Т., Гаврилюк Д.Я., Лесик Р.Б. // *ЖОФХ*. – 2011. – Т. 9, вип. 3 (35). – С. 64-71.
8. *Development Core Team. R: A language and environment for statistical computing*. – Vienna, Austria: R Foundation for Statistical Computing, 2011. ISBN 3-900051-07-0, URL <http://www.R-project.org>
9. Van Buuren S., Groothuis-Oudshoorn K. // *J. of Statistical Software*. – 2011. – Vol. 45 (3). – P. 1-67.
10. Ward J.H. // *J. of the Am. Statistical Association*. – 1963. – Vol. 58 (301). – P. 236-244.
11. Cox T.F., Cox M.A.A. *Multidimensional Scaling*. 2-nd ed. – Chapman and Hall, 2000. – 328 p.
12. *Chemical Computing Group Inc. MOE 2007.09 (Molecular Operating Environment software)*, <http://www.chemcomp.com>.

Надійшла до редакції 06.09.2012 р.