

... БІОТЕХНОЛОГІЯ ... BIOTECHNOLOGY ...

DOI: 10.26565/2075-5457-2025-44-1
УДК: 504:57.04

Система ідентифікації структурних маркерів мутагенності Еймса на основі подібності відбитків структури ксенобіотиків С.В. Кисляк, О.М. Дуган, М.О. Мороз, О.І. Яловенко

В статті приділено увагу питанням оцінки генотоксичного потенціалу хімічних сполук, що можуть потрапляти у навколишнє середовище. Доведена необхідність зміни базового вектору розвитку сучасної токсикології з урахуванням досягнень в області комп'ютерних наук та інформаційних технологій. В рамках проведення дослідження було акцентовано увагу на *in silico* підході, що дозволяє зробити висновки про генотоксичність хімічної сполуки відповідно до ідентифікованих функціональних груп, що можуть лежати в основі проявів мутагенності. Система визначення структурних маркерів мутагенності Еймса була реалізована відповідно до загальнодоступних баз даних хімічних сполук (EFSA, Kazius/Bursi та Hansen). Початкова кількість об'єднаного набору даних була збільшена за рахунок мікотоксинів, дублікати видалені. Для кожного ксенобіотику, який представлений в наборі даних був визначений мутагенний потенціал за допомогою *in vitro* тесту Еймса. З метою ефективної ідентифікації функціональних груп, що можуть бути сигналами мутагенності, було прийнято рішення щодо розподілу ксенобіотиків об'єднаного набору даних на п'ять структурних класів. Такий підхід відносно формування однорідних груп ксенобіотиків, що можуть проявляти потенційні генотоксичні властивості, дозволяє визначити структурні маркери мутагенності Еймса в рамках кожного класу мутагенів. В основі отримання достовірної інформації про наявність певної функціональної групи – сигнала мутагенності з урахуванням досліджуваного структурного класу ксенобіотиків, було запропоновано використовувати матриці відстаней, що були розраховані для кожної пари мутаген/не мутаген об'єднаного набору даних. При цьому схожість між сполуками оцінювалась за допомогою класичних метрик оцінки подібності (Танімото та Хемінга) відповідно до розрахованих трьох типів відбитків структури (molecular fingerprint) для кожного ксенобіотику. Останній етап реалізації системи виявлення структурних маркерів мутагенності Еймса був пов'язаний з пошуком та застосуванням ефективного алгоритму для візуалізації багатовимірних даних. Аналіз літератури дозволив нам обрати оптимальний алгоритм для вирішення цієї задачі. Обраний алгоритм (t-SNE) дозволяє багатовимірні дані (матриці відстаней для всіх мутагенів та не мутагенів) представити у двовимірному просторі. Така візуалізація дає змогу знайти всі пари (мутаген/не мутаген), що мають достатньо великий індекс подібності та зробити висновки про наявність певних функціональних груп, що можуть лежати в основі проявів мутагенності для кожного з п'яти структурних класів потенційних мутагенів. Досить цікавим в науковому відношенні є аналіз ефективності застосування різних типів відбитків структури для ідентифікації структурних попереджень мутагенності Еймса, що і було проведено в рамках даного дослідження. Результатом роботи є створене програмне забезпечення, що дозволяє визначати структурні маркери мутагенності Еймса на основі подібності відбитків структури хімічних сполук, що представлені у об'єднаному наборі даних. Показна можливість використання запропонованого підходу для вирішення задачі щодо пошуку причинно-наслідкових зв'язків між мутагенністю та наявністю певних функціональних груп в структурі досліджуваних ксенобіотиків.

Ключові слова: мутаген, мутація, генотоксичність, ксенобіотик, тест Еймса, індекс подібності, модель QSAR

Цитування: Кисляк С.В., Дуган О.М., Мороз М.О., Яловенко О.І. Система ідентифікації структурних маркерів мутагенності Еймса на основі подібності відбитків структури ксенобіотиків. Вісник Харківського національного університету імені В.Н. Каразіна. Серія «Біологія», 2025, 44, с. 6-14. <https://doi.org/10.26565/2075-5457-2025-44-1>

Про авторів:

Кисляк С.В. – Національний технічний університет України Київський політехнічний інститут імені Ігоря Сікорського, Берестейський проспект, 37, Київ, Україна, kisluak@ukr.net, <https://orcid.org/0000-0003-2097-3793>.
Дуган О.М. – Національний технічний університет України Київський політехнічний інститут імені Ігоря Сікорського, Берестейський проспект, 37, Київ, Україна, dugan51@gmail.com, <https://orcid.org/0000-0002-5646-917X>
Мороз О.І. – Національний технічний університет України Київський політехнічний інститут імені Ігоря Сікорського, Берестейський проспект, 37, Київ, Україна, moroz.maksym@ill.kpi.ua, <https://orcid.org/0009-0007-8573-0166>
Яловенко О.І. – Національний технічний університет України Київський політехнічний інститут імені Ігоря Сікорського, Берестейський проспект, 37, Київ, Україна, yalov89@i.ua, <https://orcid.org/0000-0002-5022-143X>

Подано до редакції: 09.03.2025 / Прорецензовано: 07.05.2025 / Прийнято до друку: 12.05.2025

Вступ

Продовж останніх кількох десятиліть завдяки науково-технічному прогресу спостерігається значне зростання кількості ксенобіотиків, що потрапляють у навколишнє середовище та здатні впливати на генетичний апарат людини. У серпні 2024 року кількість зареєстрованих ксенобіотиків, інформація про які зберігалась на серверах Американського хімічного товариства, склала більше ніж 280 млн. речовин (Kislyak, Dugan et al., 2025). Кількість хімічних сполук, що генерує людство зростає приблизно на 4000 за один день (Honma et al., 2018). При цьому кількість ксенобіотиків, що виробляються промисловістю, потрапляють у довкілля та можуть негативно впливати на спадковий апарат людини, складає більше ніж 100 000 (Honma, 2020). Наприкінці 2022 року Європейськими та Американськими агентствами хімічних речовин були оприлюднені відомості про більш ніж 800 тисяч ксенобіотиків, щодо яких немає повної інформації про прямий чи опосередкований вплив на генетичний апарат людини (Samanipour et al., 2023). На сьогоднішній день очевидно є проблема ефективного обліку та генетичної оцінки впливу всіх генотоксичних сполук, що можуть потрапляти у довкілля. Крім того, велика кількість хімічних сполук реєструється без врахування їх впливу на навколишнє середовище та здоров'я людини (Honma, 2020). При такій динаміці збільшення кількості ксенобіотиків, що можуть проявляти потенційні генотоксичні властивості та потрапляють у навколишнє середовище, наукова спільнота потребує максимальної активізації з метою підтримання генетичної та екологічної безпеки людської популяції. Необхідність у вирішенні даної проблеми особливо відчувається з урахуванням генетичних наслідків впливу хімічних мутагенів на спадковий апарат людини. Взаємодія ксенобіотиків-мутагенів з геномом людини може призводити до збільшення частоти народжених дітей зі спадковими захворюваннями (у випадку індукованих мутацій на рівні статевих клітин) або стати основою для розвитку онкологічних захворювань (індуковані мутації на рівні соматичних клітин). Молекула ДНК, що є носієм спадкової інформації, для забезпечення життєздатності організму на різних етапах онтогенезу та для збереження функції розмноження, повинна проявляти геномну стабільність, що може бути забезпечена за рахунок виключення впливу на спадкову інформацію генетично активних речовин (Tubbs, Nussenzweig, 2017; Valles et al., 2020).

Для отримання об'єктивної оцінки генетичної активності хімічних речовин, що представлені у навколишньому середовищі, необхідно врахувати три можливі кінцеві результати індукованого генотоксичними сполуками пошкодження ДНК, які пов'язані з виникненням генних мутацій, хромосомних аберацій та анеуплоїдії (Turkez et al., 2017). Оцінка потенційної генетичної небезпеки впливу факторів навколишнього середовища може бути отримана за допомогою класичної батареї короткострокових *in vitro* та *in vivo* тест-систем, що була прийнята науковою спільнотою та затверджена відповідними настановами міжнародних організацій (наприклад OECD, ECHA, UK-EMS, US-FDA, EFSA та ін.) (Mišík et al., 2022; Kislyak, Dugan et al., 2024). Серед різноманітних *in vitro* та *in vivo* експериментальних методів оцінки генотоксичності факторів навколишнього середовища *in vitro* тест Еймса заслуговує на особливу увагу. Тест на бактеріальну зворотну мутацію є швидким та простим у виконанні методом (Ren et al., 2017). Не зважаючи на те, що розроблена Брюсом Еймсом методика, яка була опублікована та апробована ще у 70-х роках 20 сторіччя, на сьогоднішній день тест Еймса все одно залишається одним з основних *in vitro* методів, що масово використовують для оцінки генетичної активності впливу факторів навколишнього середовища (Барил'як, 2002; Дуган, 2006).

Цікавим є той факт, що фундаментом розвитку та становлення сучасної комп'ютерної токсикології, як науки, стала зміна стандартної парадигми проведення тестування на генотоксичність з використанням батареї тест-систем. Розроблені у минулі десятиріччя класичні *in vitro* та *in vivo* методи оцінки генетичних ефектів факторів навколишнього середовища є складними з точки зору їх проведення, є дорогавартісними, тривалі в часі, мають проблему відтворюваності результатів експерименту в різних лабораторіях (Chu et al., 2021; Kislyak, Dugan et al., 2025). Крім того, при проведенні тестувань необхідно враховувати концепцію «3R» що керується принципами, які направлені на зменшення, вдосконалення та заміну моделей тварин при проведенні тестування на генотоксичність (Mellor et al., 2019; Yang et al., 2021). В таких умовах, з урахуванням розвитку інформаційних технологій та систем штучного інтелекту, спостерігається активізація наукової спільноти щодо розробки та впровадження сучасних *in silico* моделей QSAR (Quantitative Structure-Activity Relationship) оцінки генотоксичності, які дозволяють відмовитись від застосування в експериментах теплокровних тварин, є менш дорогавартісними, а також є ефективними з точки зору часових витрат. У наукових працях (Honma et al., 2018; Mao et al., 2021; Tolosa et al., 2023; Kislyak, Dugan et al., 2025) висвітлена методологія побудови кількісної залежності «структура-властивість», що лежить в основі *in silico* прогнозування мутагенності. Описані в науковій літературі сучасні моделі

Ames/QSAR, що використовуються для оцінки мутагенності Еймса базуються на двох підходах, що відповідає двом типам прогностичних моделей: на основі статистики та на основі правил (Honma et al., 2018). *Ames/QSAR* моделі на основі статистики побудовані відповідно до фізико-хімічних, просторових, електронних властивостей досліджуваних ксенобіотиків, що задаються різноманітними наборами молекулярних дескрипторів (наприклад PaDel, Mordred, RDkit та інші). Другий підхід, на основі правил, дозволяє спрогнозувати мутагенність Еймса відповідно до ідентифікованих підструктур або функціональних груп, що лежать в основі індукованих ксенобіотиками кінцевих генетичних ефектів. Серед сучасних QSAR моделей, які розроблені відповідно до даного підходу, виділяють методи: на основі фрагментів; на основі графів та на основі відбитків структури (Yang et al., 2018). При проведенні дослідження була приділена увага ідентифікації маркерів мутагенності, що відповідає другому типу прогностичних *Ames/QSAR* моделей, в основі яких були представлені відбитки молекулярної структури ксенобіотиків.

Об'єкти та методи дослідження

При проведенні дослідження використовувалась база даних (Helma et al., 2021), що була отримана шляхом об'єднання трьох загальнодоступних наборів даних: Kazius-Bursi (Kazius et al., 2005), Hansen (Hansen et al., 2009) та EFSA (EFSA Journal 2016). Крім того, об'єднаний набір даних був розширений мікотоксинами, які в науковій праці (Tolosa et al., 2023) використовувались для побудови прогностичних *Ames/QSAR* моделей для оцінки мутагенних ефектів впливу даної групи генотоксичних сполук на навколишнє середовище. Загальна кількість ксенобіотиків у наборі даних склала 8454 хімічних сполук. Розширений датасет був збережений у csv. форматі, в якому для кожного ксенобіотика зберігалась наступна інформація: 1. Ідентифікатор (ID), що відповідає порядковому номеру ксенобіотика в базі даних; 2. SMILES (Simplified Molecular Input Line Entry System) лінійна нотація – це загальноприйнятий текстовий формат даних, який використовується для збереження інформації про структуру хімічних сполук (Кисляк, 2023); 3. Структурний клас, до якого відноситься ксенобіотик, що визначається відповідно до особливостей молекулярного каркасу ксенобіотика (табл 1); 4. Інформація про наявний або відсутній мутагенний потенціал, що була отримана експериментально за допомогою *in vitro* теста Еймса. З метою ефективного пошуку маркерів мутагенності було запропоновано розподілити хімічні сполуки розширеної бази даних на 5 однорідних структурних класів (табл.1). Для вирішення цієї задачі був обраний веб-сервіс ClassyFire (Djoumbou Feunang et al., 2016), який на вхід приймає SMILES нотацію певного ксенобіотика, на виході отримуємо повну класифікацію хімічної сполуки включно з інформацією про особливості її молекулярного каркасу (Molecular Framework).

Таблиця 1. Розподіл ксенобіотиків за структурними класами

Table 1. Distribution of xenobiotics by structural classes

Назва класу / Class name	№ класу / # of the class	Кількість мутагенів / Number of mutagens	Кількість не мутагенів / Number of non-mutagens	Загальна кількість / Overall
Аліфатичні ациклічні	1	548	774	1322
Аліфатичні гетеромоноциклічні	2	189	178	367
Аліфатичні гетерополіциклічні		79	141	220
Аліфатичні гомомоноциклічні	3	28	101	129
Аліфатичні гомополіциклічні		29	128	157
Ароматичні гетеромоноциклічні	4	355	675	1030
Ароматичні гетерополіциклічні		1248	881	2129
Ароматичні гомомоноциклічні	5	871	1176	2047
Ароматичні гомополіциклічні		780	273	1053
ЗАГАЛОМ / OVERALL	1-5	4127	4327	8454

На наступному етапі, з метою отримання оцінки ступеня подібності між всіма парами мутаген/не мутаген, в рамках кожного з п'яти структурних класів, були проведені розрахунки відбитків структур (molecular fingerprint), що є двовимірними дескрипторами. Відбитки структур дозволяють записати інформацію про структуру ксенобіотика за допомогою двійкового рядку. Відповідно до наукової праці (Willett, 2014) в сучасній класифікації дескрипторів виділяють три групи: 1D, 2D та 3D. Вирішення задачі отримання оцінки подібності між двома хімічними сполуками базується на використанні двовимірних 2D дескрипторів. При проведенні дослідження цікавим з наукової точки зору може бути зроблений

висновок відносно оцінки ефективності застосування трьох типів відбитків структури (MACCS, RDkit та FCFP), що відносяться до трьох класів: субструктурні ключі, топологічні та циркулярні відбитки структури (Willett et al., 2005), для вирішення задачі оцінки хімічної подібності та ідентифікації маркерів мутагенності. Після проведення розрахунків трьох типів відбитків структур для кожного ксенобіотика, оцінка подібності між парами – мутаген/не мутаген здійснювалась відповідно до метрик Танімото та Хемінга (Swamidass et al., 2005; Helma et al., 2021). Розглянутий в роботі підхід щодо оцінки подібності між хімічними сполуками найчастіше використовуються для вирішення подібних задач обчислювальної медичної хімії та хемоінформатики (Maggiore et al., 2014; Кисляк, 2023). Результати оцінки попарних відстаней між всіма мутагенами та не мутагенам зберігались у вигляді матриці. Останній етап реалізації системи виявлення структурних маркерів мутагенності Еймса був пов'язаний з пошуком та застосуванням ефективного алгоритму для візуалізації багатовимірних даних. Аналіз літературних джерел відповідно до даного напрямку досліджень дозволив обрати оптимальний алгоритм – t-SNE (Т-розподіленого вкладення стохастичної близькості), що дозволяє візуалізувати багатовимірні дані (хімічні сполуки в просторі молекулярних відбитків) на площині, при цьому він має найкращу точність збереження простору сусідів у порівнянні з іншими методами (Orlov et al., 2025). У науковій праці (Challa et al., 2020) дослідники також використовують t-SNE в якості основного алгоритму для візуалізації даних з метою прогнозування тератогенності хімічних сполук з малою молекулярною масою, вплив яких може порушувати розвиток плода на ранніх етапах онтогенезу. Подібна методологія щодо зменшення розмірності даних з наступною візуалізацією у двовимірному просторі дескрипторів MolPrint2D та CDK була запропонована авторами статті (Helma et al., 2021) при дослідженні мутагенності пірилізидинових алкалоїдів.

Результати та обговорення

Результатом проведеного дослідження є розроблене програмне забезпечення, що дозволяє через візуалізацію у двовимірному просторі відбитків структури MACCS, RDkit та FCFP з урахуванням розрахованих індексів подібності для ксенобіотиків – потенційних мутагенів, що належать до п'яти структурних класів, визначати функціональні групи або підструктури, які можуть лежати в основі проявів їх мутагенності. Після завантаження програмного застосунку користувач має можливість обрати не тільки клас ксенобіотиків відповідно до особливостей будови їх молекулярного каркасу, але й тип відбитка структури (MACCS, RDkit та FCFP) та метрику оцінки подібності (Танімото та Хемінга). Відповідно до отриманих відбитків структури FCFP та розрахованих матриць відстаней на основі індекса Танімото, на рисунку 1 показані всі ксенобіотики, що належать до класу ароматичних гетерополіциклічних та ароматичних гетеромоноциклічних хімічних сполук (табл.1, 4 клас). Зеленим кольором (рис.1) виділені хімічні сполуки, які відповідно до *in vitro* тесту Еймса не вважаються мутагенами. Червоний колір використовується для позначення ксенобіотиків, мутагенність яких була підтверджена експериментально на штаммах *Salmonella typhimurium*.

Досить важливим з наукової точки зору та для досягнення основної мети проведеного дослідження, є порівняння структурних формул ксенобіотиків з метою виявлення таких розбіжностей (або ідентифікації подібних функціональних груп), які можуть розглядатись в якості визначальних, з точки зору проявів їх мутагенності. Запропонована в роботі методологія ідентифікації маркерів мутагенності може бути реалізована відповідно до двох підходів. Перший – базується на порівнянні двох сполук - мутаген/не мутаген. При цьому функціональна група, яка зустрічається на рівні мутагена та відсутня у не мутагена може бути цікавою з точки зору формулювання гіпотези щодо проявів мутагенності певного ксенобіотика. Другий підхід дозволяє отримати інформацію про певні маркери мутагенності відповідно до порівняння структурних формул подібних між собою мутагенів. В такому випадку необхідно звертати увагу на функціональні групи або підструктури, що наявні на рівні кожної досліджуваної хімічної сполуки – мутагена. Для проведення зручного відбору ксенобіотиків за допомогою системи ідентифікації структурних маркерів мутагенності Еймса реалізована можливість масштабування зображень результатів візуалізації. При цьому користувач має можливість сконцентрувати увагу на ксенобіотиках, між якими буде мінімальна відстань (рис. 2.)

На рис. 2, в якості демонстрації ефективності запропонованого підходу, показані результати порівнянь пари (мутаген/не мутаген), що дають змогу сформулювати гіпотезу щодо мутагенності хімічної сполуки з ID2011. Заміна ОН групи на NO₂ призвела до того, що ксенобіотик з ID6212 – не мутаген, перетворився на хімічну сполуку з ID 2011, що відповідно до *in vitro* тесту Еймса, є мутагеном.

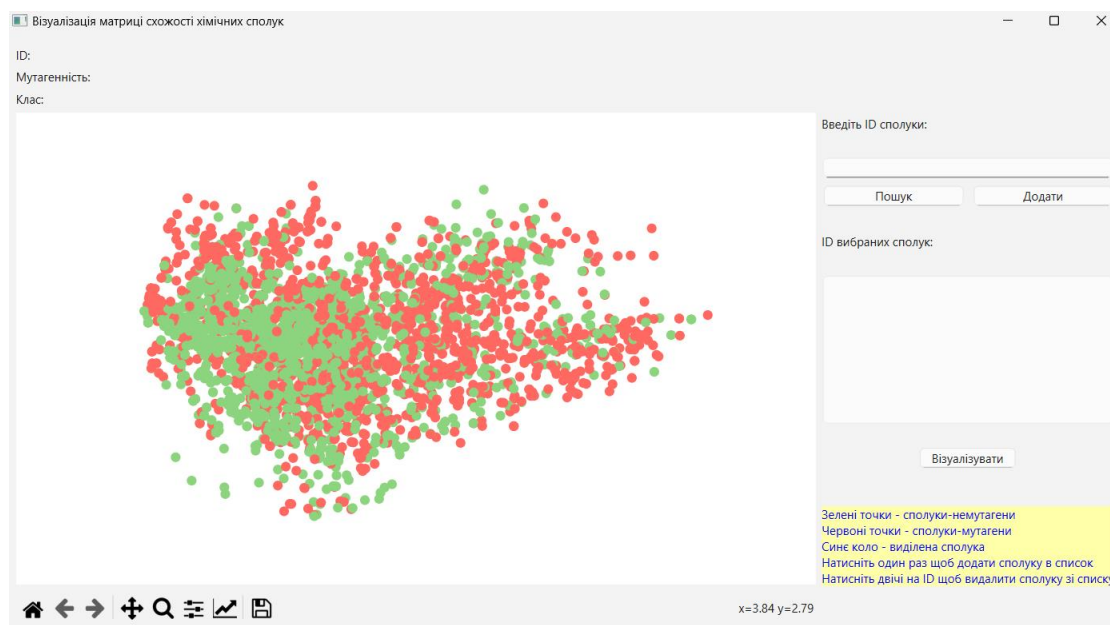


Рис. 1. Візуалізація t-SNE даних ароматичних гетерополіциклічних та ароматичних гетеромоноциклічних хімічних сполук у просторі FCFP на основі розрахованих відстаней Танімото
Fig. 1. Visualization of t-SNE data of aromatic heteropolycyclic and aromatic heteromonocyclic chemical compounds in FCFP space based on calculated Tanimoto distances

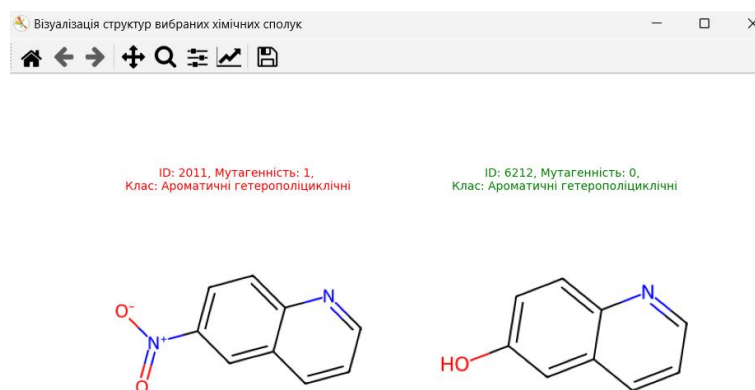


Рис. 2. Структурні формули пари мутаген/не мутаген, що відносяться до класу гетеромоноциклічних та гетерополіциклічних сполук
Fig. 2. Structural formulas of mutagen/non-mutagen pairs belonging to the class of heteromonocyclic and heteropolycyclic compounds

Функціональна група NO_2 може виступати в ролі такої, що визначає мутагенність хімічних сполук, що відносяться до класу гетеромоноциклічних та гетерополіциклічних сполук. Додатковим підтвердженням цієї гіпотези можуть бути результати порівняння подібних ксенобіотиків – мутагенів з ID2607,5706,5831,3871 (рис.3.). Можемо побачити, що кожний з чотирьох мутагенів містить функціональну групу NO_2 . Відповідно до результатів дослідження, що були опубліковані у науковій праці (Müller et al.,2006), має підтвердження сформульована нами гіпотеза щодо причинно-наслідкових зв'язків між проявами мутагенності та наявністю групи NO_2 .

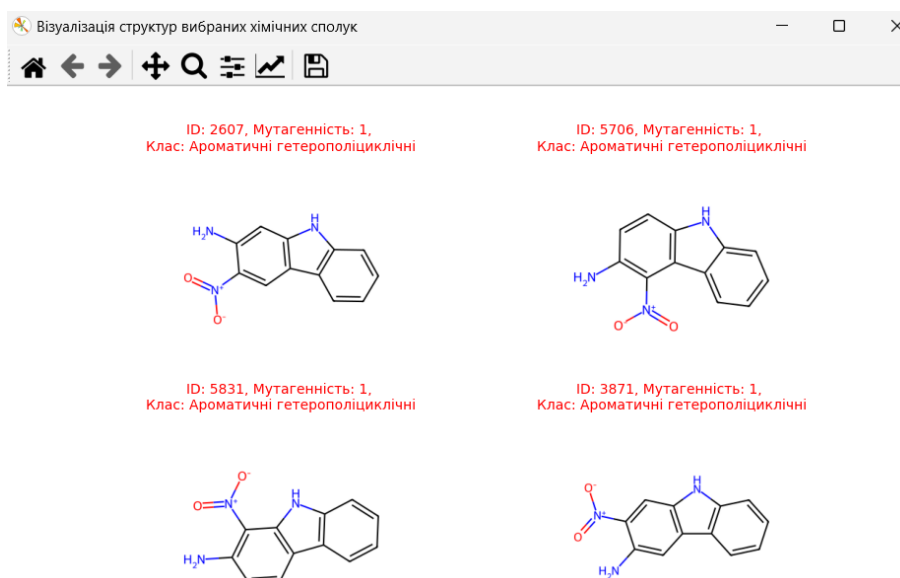


Рис. 3. Структурні формули ксенобіотиків – мутагенів, що відносяться до класу гетеромоноциклічних та гетерополіциклічних сполук

Fig. 3. Structural formulas of xenobiotics – mutagens belonging to the class of heteromonocyclic and heteropolycyclic compounds

Необхідно відмітити, що в рамках проведеного дослідження, нами не була поставлена задача щодо визначення всіх можливих функціональних груп та підструктур для кожного класу ксенобіотиків, що можуть бути пов'язані з мутагенністю. В нашій роботі була показана лише можливість застосування запропонованого в роботі підходу для отримання генетичної оцінки впливу факторів навколишнього середовища.

Досить важливим результатом проведеного нами дослідження є сформульовані висновки щодо ефективності застосування різних типів молекулярних відбитків структур ксенобіотиків, відповідно до розробленої в роботі методики. Було показано, що використання різних відбитків структури, що відносяться до трьох класів, можуть мати суттєвий вплив на кінцевий результат моделювання. В цьому контексті відбитки структури MACCS мають певні недоліки, що пов'язані з особливостями збереження інформації про наявність функціональних груп, що можуть повторюватись. MACCS не враховує кількість однакових підструктур, що представлені на рівні молекули. В такій ситуації, при візуалізації матриць подібності для ксенобіотиків, що можуть бути близькими сусідами, з урахуванням розрахованих відстаней (Танімото, Хемінга), вони не будуть схожими з точки зору порівняння їх структурних формул. Така проблема не може сприяти ефективному пошуку маркерів проявів мутагенності. У науковій праці (Cortes-Ciriano, 2016) автор також акцентує увагу на недоліках застосування іншого відбитка структури – Моргана, що відноситься до класу циркулярних та використовується у дослідженні для вирішення задачі оцінки токсичності хімічних сполук, що можуть бути кандидатами на роль лікарських препаратів. Для дескрипторів FCFP та RDKit, що належать до двох інших класів відбитків структури, відсутні недоліки, що характерні для MACCS. Для них була отримана однакова оцінка ефективності для вирішення задачі пошуку маркерів мутагенності.

Проблему щодо ефективної ідентифікації структурних маркерів мутагенності намагалися вирішити Ешбі і Теннат ще у далекому 1988 році (Ashby, Tennant, 1988). Аналіз сучасних наукових праць, відповідно до даного напрямку досліджень, дозволяють зробити висновки про те, що незважаючи на велику кількість розроблених *in silico* методів та підходів (Kazius al., 2006; Shen al., 2010; Lepailleur et al., 2013; Floris et al., 2017), проблема ефективного пошуку структурних попереджень мутагенності, у повній мірі, залишається ще не вирішеною. Необхідним є створення нових сучасних *in silico* методів та вдосконалення існуючих підходів для генетичної оцінки впливу факторів навколишнього середовища. В цьому контексті, вирішення нагальних питань токсикології, може бути здійснено з урахуванням досягнень в області сучасних інформаційних технологій, систем штучного інтелекту та хемоінформатики.

Список літератури / References

- Ashby, J., Tennant, R.W. (1988). Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. *Mutation Research*, 204, 17–115. [https://doi.org/10.1016/0165-1218\(88\)90114-0](https://doi.org/10.1016/0165-1218(88)90114-0)
- Challa, A.P., Beam, A.L., Shen, M., Peryea, T., Lavieri, R.R., Lippmann, E.S., Aronoff, D.M. (2020). Machine learning on drug-specific data to predict small molecule teratogenicity. *Reproductive toxicology (Elmsford, N.Y.)*, 95, 148–158. <https://doi.org/10.1016/j.reprotox.2020.05.004>
- Chu, C.S.M., Simpson, J.D., O'Neill, P.M., Berry, N.G. (2021). Machine learning - Predicting Ames mutagenicity of small molecules. *Journal of molecular graphics & modelling*, 109, 108011. <https://doi.org/10.1016/j.jmgm.2021.108011>
- Cortes-Ciriano, I. (2016). Bioalerts: a python library for the derivation of structural alerts from bioactivity and toxicity data sets. *Journal of cheminformatics*, 8, 13. doi: 10.1186/s13321-016-0125-7
- Djombou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., Wishart, D.S. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of cheminformatics*, 8, 61. <https://doi.org/10.1186/s13321-016-0174-y>
- EFSA Journal (2016). EFSA (European Food Safety Authority), 2016. Dietary exposure assessment to pyrrolizidine alkaloids in the European population. European Food Safety Authority. *the EFSA Journal*, 14(8), 4572. <https://doi.org/10.2903/j.efsa.2016.4572>
- Floris, M., Raitano, G., Medda, R., Benfenati, E. (2017). Fragment prioritization on a large mutagenicity dataset. *Molecular Informatics*, 36, 1600133. <https://doi.org/10.1002/minf.201600133>
- Hansen, K., Mika, S., Schroeter, T., Sutter, A., ter Laak, A., Steger-Hartmann, T., Heinrich, N., Müller, K.R. (2009). Benchmark data set for *in silico* prediction of Ames mutagenicity. *Journal of chemical information and modeling*, 49(9), 2077–2081. <https://doi.org/10.1021/ci900161g>
- Helma, C., Schöning, V., Drewe, J., Boss, P. (2021). A Comparison of Nine Machine Learning Mutagenicity Models and Their Application for Predicting Pyrrolizidine Alkaloids. *Frontiers in pharmacology*, 12, 708050. <https://doi.org/10.3389/fphar.2021.708050>
- Honma, M. (2020). An assessment of mutagenicity of chemical substances by (quantitative) structure-activity relationship. *Genes and environment : the official journal of the Japanese Environmental Mutagen Society*, 42, 23. <https://doi.org/10.1186/s41021-020-00163-1>
- Honma, M., Kitazawa, A., Cayley, A., Williams, R.V., Barber, C., Hanser, T., Saiakhov, R., Chakravarti, S., Myatt, G.J., Cross, K.P., Benfenati, E., Raitano, G., Mekenyan, O., Petkov, P., Bossa, C., Benigni, R., Battistelli, C.L., Giuliani, A., Tcheremenskaia, O., DeMeo, C., Norinder, U., Koga, H., Jose, C., Jeliaskova, N., Kochev, N., Paskaleva, V., Yang, Ch., Daga, P.R., Clark, R.D., Rathman, J. (2019). Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project. *Mutagenesis*, 34(1), 3–16. <https://doi.org/10.1093/mutage/gey031>
- Kazius, J., McGuire, R., Bursi, R. (2005). Derivation and validation of toxicophores for mutagenicity prediction. *Journal of medicinal chemistry*, 48(1), 312–320. <https://doi.org/10.1021/jm040835a>
- Kazius, J., Nijssen, S., Kok, J., Bäck, T., Ijzerman, A.P. (2006). Substructure mining using elaborate chemical representation. *J. Chem. Inf. Model.* 46, 597–605. <https://doi.org/10.1021/ci0503715>
- Kislyak, S., Dugan, O., Yalovenko, O. (2024). Systems for Genetic Assessment of the Impact of Environmental Factors. *Innovative Biosystems and Bioengineering*, 8(2), 3–27. <https://doi.org/10.20535/ibb.2024.8.2.288127>
- Kislyak, S., Dugan, O., Yesypenko, R., Starosyla, D., Yalovenko, O. (2025). *In silico* the Ames Mutagenicity Predictive Model of Environment. *Innovative Biosystems and Bioengineering*, 9(2), 42–52. <https://doi.org/10.20535/ibb.2025.9.2.316239>
- Lepailleur, A., Poezevara, G., Bureau, R. (2013). Automated detection of structural alerts (chemical fragments) in (eco)toxicology. *Comput. Struct. Biotechnol. J.* 5, e201302013. <https://doi.org/10.5936/csbj.201302013>
- Maggiore, G., Vogt, M., Stumpfe, D., Bajorath, J. (2014). Molecular similarity in medicinal chemistry. *Journal of medicinal chemistry*, 57(8), 3186–3204. <https://doi.org/10.1021/jm401411z>
- Mao, J., Akhtar, J., Zhang, X., Sun, L., Guan, S., Li X., Chen, G., Liu, J., Jeon, H.N., Kim, M.S., No, K.T., Wang, G. (2021). Comprehensive strategies of machine-learning-based quantitative structure-activity relationship models. *iScience*, 24(9), 103052. <https://doi.org/10.1016/j.isci.2021.103052>
- Mellor, C.L., Marchese Robinson, R.L., Benigni, R., Ebbrell, D., Enoch, S.J., Firman, J.W., Madden, J.C., Pawar, G., Yang, C., Cronin, M.T.D. (2019). Molecular fingerprint-derived similarity measures for

- toxicological read-across: Recommendations for optimal use. *Regulatory toxicology and pharmacology: RTP*, 101, 121–134. <https://doi.org/10.1016/j.yrtph.2018.11.002>
- Mišík, M., Nersesyan, A., Ferk, F., Holzmann, K., Krupitza, G., Herrera Morales, D., Staudinger, M., Wultsch, G., Knasmueller, S. (2022). Search for the optimal genotoxicity assay for routine testing of chemicals: Sensitivity and specificity of conventional and new test systems. *Mutation research. Genetic toxicology and environmental mutagenesis*, 881, 503524. <https://doi.org/10.1016/j.mrgentox.2022.503524>
- Müller, L., Mauthe, R.J., Riley, C.M., Andino, M.M., Antonis, D.D., Beels, C., DeGeorge, J., De Knaep, A.G., Ellison, D., Fagerland, J.A., Frank, R., Fritschel, B., Galloway, S., Harpur, E., Humfrey, C.D., Jacks, A.S., Jagota, N., Mackinnon, J., Mohan, G., Ness, D.K., O'Donovan, M.R., Smith, M.D., Vudathala, G., Yotti, L. (2006). A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity. *Regulatory toxicology and pharmacology: RTP*, 44(3), 198–211. <https://doi.org/10.1016/j.yrtph.2005.12.001>
- Orlov, A.A., Akhmetshin, T.N., Horvath, D., Marcou, G., Varnek, A. (2025). From High Dimensions to Human Insight: Exploring Dimensionality Reduction for Chemical Space Visualization. *Molecular informatics*, 44(1), e202400265. <https://doi.org/10.1002/minf.202400265>
- Ren, N., Atyah, M., Chen, W.Y., Zhou, C.H. (2017). The various aspects of genetic and epigenetic toxicology: testing methods and clinical applications. *Journal of translational medicine*, 15(1), 110. <https://doi.org/10.1186/s12967-017-1218-4>
- Samanipour, S., O'Brien, J.W., Reid, M.J., Thomas, K.V., Praetorius, A. (2023). From Molecular Descriptors to Intrinsic Fish Toxicity of Chemicals: An Alternative Approach to Chemical Prioritization. *Environmental science & technology*, 57(46), 17950–17958. <https://doi.org/10.1021/acs.est.2c07353>
- Shen, J., Cheng, F., Xu, Y., Li, W., Tang, Y. (2010). Estimation of ADME properties with substructure pattern recognition. *Journal of Chemical Information and Modeling*, 50, 1034–1041. <https://doi.org/10.1021/ci100104j>
- Swamidass, S.J., Chen, J., Bruand, J., Phung, P., Ralaivola, L., Baldi, P. (2005). Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity. *Bioinformatics (Oxford, England)*, 21 Suppl 1, i359–i368. <https://doi.org/10.1093/bioinformatics/bti1055>
- Tolosa, J., Serrano Candelas, E., Vallés Pardo, J.L., Goya, A., Moncho, S., Gozalbes, R., Palomino Schätzlein, M. (2023). MicotoXilico: An Interactive Database to Predict Mutagenicity, Genotoxicity, and Carcinogenicity of Mycotoxins. *Toxins*, 15(6), 355. <https://doi.org/10.3390/toxins15060355>
- Tubbs, A., Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell*, 168(4), 644–656. <https://doi.org/10.1016/j.cell.2017.01.002>
- Turkez, H., Arslan, M.E., Ozdemir, O. (2017). Genotoxicity testing: progress and prospects for the next decade. *Expert opinion on drug metabolism & toxicology*, 13(10), 1089–1098. <https://doi.org/10.1080/17425255.2017.1375097>
- Valles, G.J., Bezsonova, I., Woodgate, R., Ashton, N.W. (2020). USP7 Is a Master Regulator of Genome Stability. *Frontiers in cell and developmental biology*, 8, 717. <https://doi.org/10.3389/fcell.2020.00717>
- Willett, P. (2014). The Calculation of Molecular Structural Similarity: Principles and Practice. *Molecular informatics*, 33(6-7), 403–413. <https://doi.org/10.1002/minf.201400024>
- Yang, H., Sun, L., Li, W., Liu, G., Tang, Y. (2018). In Silico Prediction of Chemical Toxicity for Drug Design Using Machine Learning Methods and Structural Alerts. *Frontiers in chemistry*, 6, 30. <https://doi.org/10.3389/fchem.2018.00030>
- Yang, X., Zhang, Z., Li, Q., Cai, Y. (2021). Quantitative structure-activity relationship models for genotoxicity prediction based on combination evaluation strategies for toxicological alternative experiments. *Scientific reports*, 11(1), 8030. <https://doi.org/10.1038/s41598-021-87035-y>
- Бариляк, І.Р., Дуган, О.М. (2002). Еколого-генетичні дослідження в Україні. *Цитологія і генетика*, 5, с. 3-10. [Baryliak, I.R., Duhan, O.M. (2002). Ecological and genetic research in Ukraine. *Cytology and Genetics*, 5, 3–10 (in Ukrainian)]
- Дуган, О.М., Яловенко, О.І. (2006). Потенційна мутагенна дія інгредієнтів фарб для волосся в альтернативних тест-системах. *Проблеми екол. та мед. генетики і клін. імунології: Зб. наук. пр. К.; Льв.; Х.*, 1(64). [Duhan, O.M., Yalovenko, O.I. (2006). Potential mutagenic effect of hair dye ingredients in alternative test systems. *Problems of Ecological and Medical Genetics and Clinical Immunology: Collection of Scientific Papers. Kyiv–Luhansk–Kharkiv*, 1(64) (in Ukrainian)]
- Кисляк, С.В., Голуб, Н.Б., Дуган, О.М., Аверьянова, О.А. (2023). Моделювання молекулярної взаємодії [Електронний ресурс]: Підручник для здобувачів ступеня магістра за спеціальністю 162 «Біотехнології та біоінженерія» (Електронні текстові дані, 1 файл, 26 Мбайт). Київ: КПІ ім. Ігоря Сікорського. [Kysliak, S.V., Holub, N.B., Duhan, O.M., Averyanova, O.A. (2023). Molecular Interaction

Modeling [Electronic resource]: Textbook for Master's Degree Students, Specialty 162 "Biotechnology and Bioengineering" (Electronic text data, 1 file, 26 MB). Kyiv: Igor Sikorsky Kyiv Polytechnic Institute (in Ukrainian)]

A system for identification of structural markers of Ames mutagenicity based on similarity of xenobiotic structure fingerprints **S.V. Kislyak, O.M. Dugan, M.O. Moroz, O.I. Yalovenko**

The article focuses on the assessment of the genotoxic potential of chemical compounds that may be released into the environment. The necessity of changing the basic vector of development of modern toxicology in view of the achievements in the field of computer science and information technology is proved. In the framework of the study, attention was focused on the *in silico* approach, which allows to draw conclusions about the genotoxicity of a chemical compound in accordance with the identified functional groups that may underlie the manifestations of mutagenicity. The Ames system for determining structural markers of mutagenicity was implemented in accordance with publicly available databases of chemical compounds (EFSA, Kazius/Bursi and Hansen). The initial number of the merged dataset was increased by mycotoxins, and duplicates were removed. For each xenobiotic presented in the dataset, the mutagenic potential was determined using the *in vitro* Ames test. In order to effectively identify functional groups that may be signals of mutagenicity, it was decided to divide the xenobiotics of the combined data set into five structural classes. Such an approach to the formation of homogeneous groups of xenobiotics that may exhibit potential genotoxic properties allows us to identify structural markers of Ames mutagenicity within each class of mutagens. To obtain reliable information on the presence of a certain functional group - mutagenicity signal, taking into account the studied structural class of xenobiotics, it was proposed to use distance matrices calculated for each mutagen/non-mutagen pair of the combined data set. The similarity between the compounds was evaluated using classical similarity evaluation metrics (Tanimoto and Heming) according to the calculated three types of molecular fingerprints for each xenobiotic. The last stage of the implementation of the Ames system for detecting structural markers of mutagenicity was associated with the search for and application of an effective algorithm for visualizing multidimensional data. The literature analysis allowed us to choose the optimal algorithm for solving this problem. The chosen algorithm (t-SNE) allows multidimensional data (distance matrices for all mutagens and non-mutagens) to be represented in two-dimensional space. This visualization allows us to find all pairs (mutagen/non-mutagen) that have a sufficiently high similarity index and draw conclusions about the presence of certain functional groups that may underlie the manifestations of mutagenicity for each of the five structural classes of potential mutagens. It is quite interesting from the scientific point of view to analyze the effectiveness of using different types of structure fingerprints to identify structural warnings of Ames mutagenicity, which was carried out in the framework of this study. The result of the work is the developed software that allows determining structural markers of Ames mutagenicity based on the similarity of the structure fingerprints of chemical compounds represented in the combined data set. The possibility of using the proposed approach to solve the problem of finding cause-and-effect relationships between mutagenicity and the presence of certain functional groups in the structure of the studied xenobiotics is demonstrated.

Key words: *mutagen, mutation, genotoxicity, xenobiotic, Ames test, similarity index, QSAR model*

Cite this article: Kislyak S.V., Dugan O.M., Moroz M.O., Yalovenko O.I. (2025). A system for identification of structural markers of Ames mutagenicity based on similarity of xenobiotic structure fingerprints. *The Journal of V. N. Karazin Kharkiv National University. Series Biology*, 44, p. 6-14. <https://doi.org/10.26565/2075-5457-2025-44-1> (in Ukrainian)

About the authors:

Kislyak S.V. - National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Beresteysky Avenue, Kyiv, Ukraine, kisluak@ukr.net, <https://orcid.org/0000-0003-2097-3793>

Dugan O.M. - National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Beresteysky avenue, Kyiv, Ukraine, dugan51@gmail.com, <https://orcid.org/0000-0002-5646-917X>

Moroz O.I. - National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Beresteysky avenue, Kyiv, Ukraine, moroz.maksym@i.kpi.ua, <https://orcid.org/0009-0007-8573-0166>

Yalovenko O.I. - National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Beresteysky avenue, Kyiv, Ukraine, yalov89@i.ua, <https://orcid.org/0000-0002-5022-143X>

Внесок авторів: Усі автори зробили рівнозначний внесок у підготовку цієї роботи. / **Authors Contribution:** All authors have contributed equally to this work.

Конфлікт інтересів: Автори заявляють про відсутність конфлікту інтересів. / **Conflict of Interest:** The authors declare no conflict of interest.

Received: 09.03.2025 / Revised: 07.05.2025 / Accepted: 12.05.2025