

# СИНТЕЗ ТА АНАЛІЗ БІОЛОГІЧНО АКТИВНИХ СПОЛУК

УДК 547, 615.1, 547.9; 577.1, 615.012:547.789:547.642

DOI: 10.32352/0367-3057.6.23.05

О. В. КЛЕНІНА (<https://orcid.org/0000-0002-8946-3698>), канд. фарм. наук, доцент,

Т. І. ЧАБАН (<https://orcid.org/0000-0003-0618-275X>), канд. фарм. наук, доцент

*Львівський національний медичний університет імені Данила Галицького*

## ВИКОРИСТАННЯ БАЗ ДАНИХ ХЕМОІНФОРМАТИКИ ТА БІОІНФОРМАТИКИ У ПРОЦЕСАХ КОМП'ЮТЕРНОГО КОНСТРУЮВАННЯ ЛІКІВ (огляд)

**Ключові слова:** комп'ютерне конструювання ліків, база даних, раціональне конструювання ліків, ідентифікація мішені, хемоінформатика, біоінформатика

### АНОТАЦІЯ

Впровадження методів комп'ютерного моделювання у процес розроблення лікарських засобів (computer-aided drug discovery – CADD) дає змогу мінімізувати час та вартість створення нових біологічно активних речовин порівнюючи з традиційними експериментальними підходами. Процес *in silico* розроблення ліків включає ідентифікацію біологічних мішеней для потенційних лікарських засобів та створення комбінаторних бібліотек низькомолекулярних сполук із подальшим проведенням їх віртуального скринінгу до обраних макромолекул. Такий підхід є можливим за наявності інформації про тривимірні хімічні структури лігандів та біомолекул. У цьому контексті бази даних хемоінформатики та біоінформатики є високоефективними ресурсами для використання на різних етапах процесу раціонального конструювання ліків.

Метою цього огляду став розгляд основних підходів у разі застосування методів молекулярного моделювання у процесі раціонального конструювання ліків, а також аналіз та узагальнення інформації про сучасні бази хемоінформатики та біоінформатики, які є у вільному доступі та широко використовуються на всіх етапах пошуку і розроблення потенційних лікарських засобів.

У роботі використано комплекс загальнонаукових методів пошуку та систематизації джерел літератури, аналіз та порівняння інформації з різних джерел, узагальнення зі зазначенням тенденцій розвитку, виділення нових та перспективних напрямів досліджень.

В огляді охарактеризовано основні сучасні хемоінформатичні та біоінформатичні бази даних, які є у вільному доступі. Ці інформаційні ресурси широко використовуються для проведення ліганд-орієнтованого та рецептор-орієнтованого віртуального скринінгу, що є основними підходами у процесі комп'ютерного конструювання ліків.

Успіхи у розвитку органічного синтезу та високопродуктивного скринінгу, експериментальної біології, хімії та медицини, а також розроблення і запровадження нових підходів та інструментів хемоінформатики та біоінформатики, технологій обробки великих даних (Big Data) та досягнень інформаційних технологій створили підґрунтя для розроблення та швидкого наповнення високоефективних і структурованих баз даних. Ресурси баз даних хемоінформатики та біоінформатики, які є у вільному доступі, широко використовуються у процесах *in silico* розроблення та удосконалення потенційних лікарських засобів.

О. В. KLENINA (<https://orcid.org/0000-0002-8946-3698>),

Т. І. CHABAN (<https://orcid.org/0000-0003-0618-275X>)

*Danylo Halytsky Lviv National Medical University, Lviv*

## USE OF CHEMOINFORMATICS AND BIOINFORMATICS DATABASES IN THE PROCESSES OF COMPUTER-AIDED DRUG DESIGN (Review)

**Key words:** computer-aided drug design, database, rational drug design, target identification, chemoinformatics, bioinformatics

### ABSTRACT

The introduction of molecular modeling methods into the computer-aided drug discovery (CADD) process allows to minimize the time and costs for construction and development of new biologically active

© О. В. Кленіна, Т. І. Чабан, 2023

substances in comparison with traditional experimental approaches. *In silico* drug development process includes the identification of potential biological targets for drug candidates and the combinatorial libraries creation of low molecular weight compounds followed by their virtual screening to selected macromolecules. This approach becomes possible due to the availability of information about the three-dimensional chemical structures of ligands and biomolecules. In this context, freely available chemoinformatics and bioinformatics databases are the highly effective resources for use at various stages of the rational drug design process.

The aim of this review was to consider the main approaches towards the application of molecular modeling methods in the process of rational drug discovery, as well as to analyse and summarize information concerning modern chemoinformatics and bioinformatics databases, which are freely available and widely used at all stages of the search and development of potential drug candidates.

In the process of the review preparation we used a complex of general scientific methods of searching and systematizing literary references, analysis and comparison of information from various sources, generalization with the indication of development trends, selection of new and promising research directions.

This review also describes the main modern freely available chemoinformatics and bioinformatics databases. These information resources are widely used for ligand-based and receptor-based virtual screening, which are the main approaches in the process of computer-aided drug design.

Successes in the development of organic synthesis, high-throughput screening, experimental biology, chemistry, and medicine, as well as the development and implementation of new approaches and tools in chemoinformatics and bioinformatics, Big Data and information technologies achievements have created the basis for the development and increasing updating of highly efficient and curated databases. Freely available chemoinformatics and bioinformatics database resources are widely used in the processes of *in silico* discovery and improvement of potential drug candidate.

---

## Вступ

В основу сучасних методів комп'ютерного конструювання ліків, що базуються на дослідженнях молекулярних механізмів біологічної дії існуючих лікарських засобів, покладено підхід раціональної розробки ліків, який передбачає наявність ідентифікованої біологічної мішені та полягає у поетапному відборі сполук-хітів за результатами віртуального у бібліотек сполук, їх подальшій оптимізації з метою поліпшення «лікоподібних» властивостей і селективності та зниження токсичності. Впровадження таких методів стало можливим завдяки їх широкій міждисциплінарності з можливістю залучення новітніх досягнень у галузях геноміки, молекулярної біології, протеоміки, фармакології, медицини, а також розробки обчислювальних алгоритмів та інструментів молекулярного моделювання. Такі міждисциплінарні дослідження є основою сучасного раціонального дизайну нових лікарських засобів, спрямованого на створення речовини з бажаними фізико-хімічними властивостями та заданою біологічною дією [1].

Традиційний підхід до процесу розроблення та удосконалення існуючих лікарських засобів є затратним і тривалим процесом, основними етапами якого є ідентифікація та валідація біологічних мішеней, створення комбінаторних бібліотек хімічних сполук, відбір із них сполук-хітів із поліпшеними «лікоподібними» властивостями, оптимізація їх структури, а також доклінічні *in vitro* та *in vivo* дослідження їхньої біологічної активності, які завершуються клінічними випробуваннями. За сучасними підрахунками, орієнтовна вартість виведення на ринок нового препарату становить приблизно 1,8 мільярда доларів США [2], а рівень відсіювання, особливо під час клінічних етапів дослідження потенційного лікарського засобу, досягає 96%. Причиною такого високого коефіцієнта відсіювання, що виражається низькою ефективністю лікоподібних речовин, є недостатній рівень їх адсорбції, розподілу, метаболізму та виведення, а також висока токсичність (ADME-Tox) [3]. Як правило, *in vivo* та *in vitro* методи використовують для дослідження безпечності ліків, включно зі встановленням їх побічних ефектів і токсичності. Для прискорен-

ня ідентифікації біологічно активних речовин серед великої кількості синтезованих сполук за допомогою автоматизованих підходів було розроблено методи висококоєфективного скринінгу (ВЕС) (High-throughput screening (HTS) methods). Хоча автоматизовані підходи ВЕС зменшують затрати часу і людської праці, масштаби застосування методів висококоєфективного скринінгу залишаються низькими порівняно з різноманітністю хімічних структур.

**Метою** цього огляду став розгляд основних підходів у разі застосування *in silico* методів у процесі розроблення лікарських засобів, а також аналіз та узагальнення інформації про сучасні хемоінформаційні та біоінформаційні бази даних, які є у вільному доступі і широко використовуються у процесі раціонального конструювання ліків.

## **Матеріали та методи дослідження**

У роботі використано комплекс загальнонаукових методів пошуку та систематизації літературних джерел, аналіз та порівняння інформації з різних джерел, узагальнення зі зазначенням тенденцій розвитку, виділення нових та перспективних напрямів досліджень.

## **Результати дослідження та обговорення**

### 1. Значення методів комп'ютерного моделювання у раціональному конструюванні ліків

Впровадження методів комп'ютерного моделювання у процес конструювання ліків (computer-aided drug design – CADD) дає змогу мінімізувати час та вартість створення нових лікарських засобів порівнюючи з традиційними експериментальними підходами. Процес раціонального конструювання ліків із використанням методів комп'ютерного моделювання ґрунтується на ідентифікації потенційних біологічних мішеней та передбаченні механізму взаємодії рецептора і ліганду. Наступні етапи процесу розроблення кандидатів у лікарські засоби, тобто речовин, які, як передбачається, будуть з високою афінністю та селективністю зв'язуватись з обраною мішенню, включають створення хімічних бібліотек низькомолекулярних сполук, проведення віртуального скринінгу та кількісне оцінювання афінності лігандів до біологічної мішені, а також моделювання зв'язування ліганд–мішень, при якому враховується специфічність взаємодії низькомолекулярної хімічної сполуки із відповідною біомолекулою [4]. Ряд наступних стадій складається з ідентифікації сполук-хітів за результатами віртуального скринінгу, їх подальшої оптимізації з метою підвищення афінності та селективності (оптимізація «від хіта до лідера»), *in silico* прогнозування основних показників біодоступності та потенційної токсичності, а також подальша оптимізація структур-лігандів із метою поліпшення цих показників і зниження токсичності при збереженні спорідненості до зв'язування з мішенню. Після проведення обчислень всіх зазначених характеристик для сполук, відібраних як кандидати у лікарські засоби, ці ж характеристики вивчають експериментально *in vitro/in vivo* для підтвердження теоретичних результатів. Таким чином, підходи комп'ютерного моделювання потенційних лікарських засобів можуть зменшити кількість хімічних сполук, для яких потрібно провести фармакологічний скринінг, одночасно збільшуючи кількість потенційних кандидатів у лікарські засоби за рахунок виключення з розгляду неефективних і токсичних хімічних сполук. Отже, використання *in silico* методів у процесах конструювання ліків дає змогу суттєво підвищити ефективність процесу створення потенційних лікарських засобів на стадії ідентифікації сполук-хітів та їх подальшої структурної оптимізації.

На сьогодні методи комп'ютерного моделювання успішно використовують для виведення нових лікарських сполук на ринок для лікування різних захворювань [5–7], включаючи препарати, що інгібують вірус імунодефіциту людини (ВІЛ) (атазанавір), протипухлинних препаратів (ралтитрексед) та антибіотиків (норфлоксацин).

## 2. Основні підходи при застосуванні *in silico* методів для розроблення лікарських засобів

Основною метою раціонального конструювання ліків є пошук невеликих молекул-лігандів, які можуть модулювати функцію ідентифікованої біологічної мішені і, тим самим, модулювати фенотип захворювання. Крім того, необхідно ідентифікувати невеликі молекули, які виявляють ефективні фармакокінетичні властивості та низьку токсичність. Застосування *in silico* методів у процесі розроблення потенційних лікарських засобів можливе також на всіх етапах скринінгу включно із доклінічним та клінічним етапами, що помітно знижує ризик невдач.

На початковій стадії розробки лікарських засобів широко використовують два основних типи конструювання ліків: структура-орієнтоване (мішень-орієнтоване) [8] та ліганд-орієнтоване конструювання ліків [9], які дають змогу проводити скринінг бібліотек потенційних лікарських засобів. Вибір відповідного типу конструювання залежить від наявності інформації про структуру цільової біологічної мішені. У випадку структура-орієнтованого типу конструювання необхідна наявність інформації про структуру цільової біомолекули, яку зазвичай отримують експериментально за допомогою ядерного магнітного резонансу або рентгенівської кристалографії. Якщо експериментальна інформація про 3D-структуру мішені є недоступною, для її прогнозування можна використовувати методи моделювання гомології або *ab initio*. Після встановлення структури можливе проведення віртуального скринінгу, зокрема молекулярного докінгу. Ліганд-орієнтований тип раціонального конструювання ліків зазвичай використовують як альтернативний метод у тих випадках, коли неможливо отримати структуру цільової біологічної мішені. Хоча цей підхід потребує попередньої наявності інформації про вже відомі біологічно активні речовини, що мають високу афінність до цільової біомолекули, багато сполук із різними видами активності було виявлено і зібрано у загальнодоступні бази даних, якщо мішень для них вже відома [10].

### *2.1. Ліганд-орієнтований тип раціонального конструювання ліків*

Ліганд-орієнтований підхід до раціонального конструювання ліків зазвичай використовується, коли тривимірна структура цільової біологічної мішені невідома. Такі методи, як пошук подібності, фармакофорне моделювання та QSAR, надають корисну інформацію про взаємодію між мішенню та лігандом за відсутності знань про структуру біомолекули [11]. При застосуванні цього типу розробки ліків використовують попередні знання про ефективність та властивості лікоподібних сполук, зокрема їх структурні особливості, фізичні та хімічні властивості, з метою конструювання нових кандидатів у лікарські засоби з подібними біологічними ефектами (рис. 1).

Відомі біологічно активні сполуки використовують для прогнозування активності нових потенційних лікарських засобів з-поміж великої кількості синтезованих або віртуальних структур. Прогнозування біологічної активності сполук при ліганд-орієнтованому типі раціонального конструювання ліків ґрунтується на подібності ознак (наприклад, ароматичних властивостях, наявності акцепторів та донорів водневих зв'язків, гідрофобності) між хімічними сполуками, виходячи з припущення, що сполуки, які мають високу структурну та фізико-хімічну подібність, найбільш імовірно, проявлятимуть подібну біологічну активність. Іншим інструментом при ліганд-орієнтованому типі створення ліків є розробка фармако-

форних моделей із використанням баз даних відомих біологічно активних сполук, які визначають та об'єднують мінімально необхідні структурні характеристики ліганду, необхідні для його ефективного та селективного зв'язування з обраною біологічною мішенню [12]. Крім того, у ліганд-орієнтованому пошуку ліків широко використовуються кількісні співвідношення структура–активність (QSAR) [13], за допомогою яких встановлюється кореляція між розрахованими властивостями молекул та їхньою експериментально визначеною біологічною активністю; ці кореляції в подальшому використовують для віртуального скринінгу нових потенційно активних сполук серед похідних цього класу.



Рис. 1. Процес ліганд-орієнтованого комп'ютерного конструювання ліків

У сучасній медичній хімії завдяки суттєвому розширенню хімічних баз даних та розвитку комп'ютерних технологій віртуальний скринінг на основі пошуку подібності та за допомогою розроблених фармакофорних і QSAR моделей став одним із важливих засобів пошуку сполук-хітів, для яких здійснюється подальша оптимізація та синтез потенційних біологічно активних сполук із прогнозованою активністю.

## 2.2. Структура-орієнтований тип конструювання ліків

При застосуванні структура-орієнтованого типу раціонального конструювання ліків визначають комплементарність ліганду до цільової біологічної мішені та розраховують афінність (енергію зв'язку) між лігандом і біомолекулою, використовуючи відому інформацію про їх структуру (рис. 2). Отже, початковими етапами структура-орієнтованого конструювання ліків є ідентифікація та валідація біологічної мішені, а також пошук ефективних лігандів для неї [15].

На першому етапі визначають потенційну терапевтичну мішень із встановленою 3D структурою. Зазвичай, тривимірні структури всіх терапевтично важливих біомолекул визначаються експериментально методами інтегративної структурної біології, наприклад ЯМР, рентгенівська кристалографія або криоелектронна мікроскопія. Якщо повна тривимірна структура біомолекули недоступна, використовують методи молекулярного моделювання, зокрема гомологічне, порівняльне або *ab initio* моделювання. Валідація структури біологічної мішені передбачає пере-

вірку її стереохімії та можливих конформацій для всіх амінокислотних залишків, присутніх у її структурі. Для ідентифікації ділянки зв'язування виконують ретельне дослідження тривимірної структури цільової біомолекули шляхом аналізу фізико-хімічних характеристик, електростатичних властивостей і характеристик форми активних та алостеричних центрів макромолекули, а також здійснюють перевірку на наявність порожнин, щілин і додаткових алостеричних кишень.

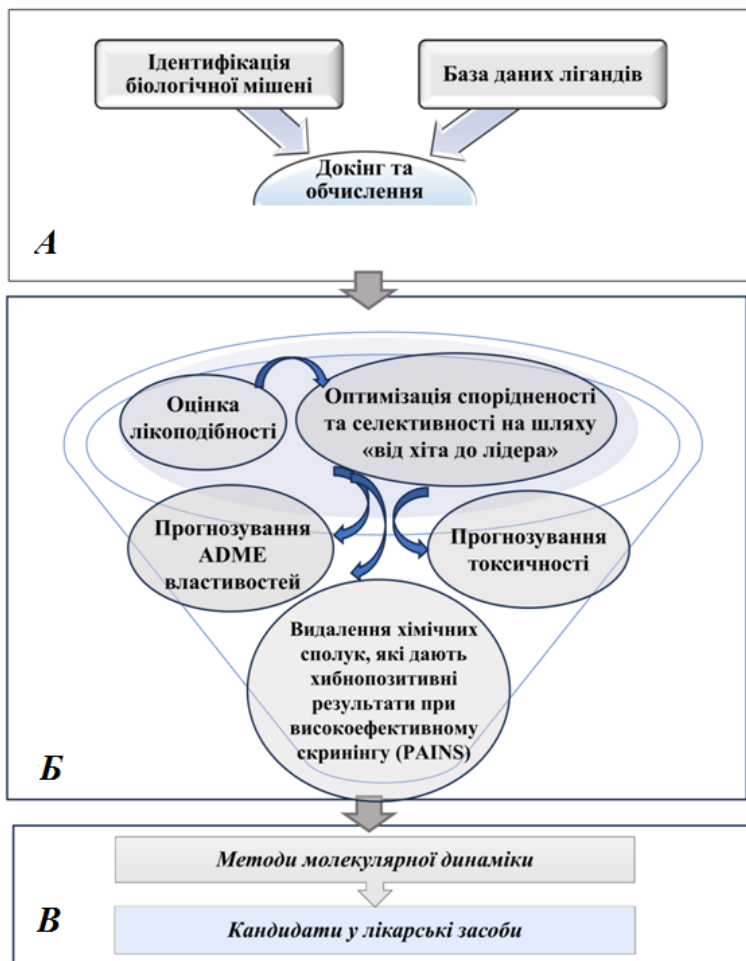


Рис. 2. Процес структура-орієнтованого комп'ютерного конструювання ліків

*А* – Здійснюється ідентифікація біологічної мішені та її ділянки зв'язування, а також створюється бібліотека потенційних лігандів. Проводиться молекулярний докінг лігандів до цільової біологічної мішені, отримані комплекси мішень–ліганд розташовуються у послідовності посилення афінності ліганду до мішені та поліпшення їх комплементарності, необхідної для досягнення оптимального типу зв'язування.

*Б* – Здійснюється структурна оптимізація сполук-хітів, виконується їх фільтрація за фізико-хімічними властивостями, що визначають ADME-Tox та інші фармакологічні характеристики.

*В* – Додаткове моделювання молекулярної динаміки виконується для вдосконалення структур хімічних речовин, і, зрештою, відбираються кандидати у лікарські засоби. В подальшому для них здійснюють фармакологічний скринінг.

Основними методами пошуку потенційних терапевтично ефективних лігандів, здатних модулювати функції обраної біологічної мішені, є віртуальний скринінг великих хімічних баз даних та *de novo* дизайн, спрямований на конструювання молекул лігандів з окремих фрагментів, у якому окремі «будівельні блоки» (атоми, цикли, функціональні групи тощо) з'єднують між собою безпосередньо у ділянці зв'язування біологічної мішені таким чином, щоб утворена молекула була максимально комплементарною до активного центру і могла проявляти високу афінність до обраної біомолекули.

До відібраних на попередньому етапі сполук застосовують процедуру молекулярного докінгу. При цьому спорідненість ліганду до зв'язування з біомолекулою визначають як за допомогою оцінюючих функцій, так і за стеричним розташуванням ліганду у центрі зв'язування мішені та його здатністю приймати конформацію, що забезпечує утворення саме тих типів зв'язків та міжмолекулярних взаємодій у комплексі «ліганд–мішень», які необхідні для існування стабільного комплексу. За результатами проведеного докінгу здійснюють ранжування лігандів за енергіями зв'язку у комплексах «ліганд–мішень» та за величинами різноманітних оцінюючих функцій.

Для сполук-хітів, виявлених за результатами молекулярного докінгу, на наступному етапі проводиться оцінка їх фармакокінетичних властивостей, показників біодоступності і токсичності. Завдяки прогресу в алгоритмах машинного навчання та накопичених базах даних, ADME-Tox властивості також можна передбачити за допомогою *in silico* методів. Для прогнозування фармакокінетичних властивостей хімічних сполук розроблено низку моделей із використанням алгоритмів машинного навчання, зокрема PreADMET [15] та SwissADME [16].

Проведення симуляції комплексу «ліганд–мішень» методами молекулярної динаміки надає більш детальну інформацію про динамічну поведінку молекулярної системи та стабільність утвореного комплексу у часі, зокрема дає змогу проаналізувати конформаційні зміни ліганду, динаміку взаємодії між лігандом і біомолекулою та оцінити можливі конформаційні зміни у структурі макромолекули, спричинені її взаємодією з лігандом. Ця інформація може бути корисною під час подальшого розроблення лікарських засобів і прогнозуванні потенційної взаємодії ліганду з мішенню.

Отже, методи молекулярного моделювання та їх впровадження у процес раціонального конструювання ліків продемонстрували величезний потенціал для прискорення розроблення нових ефективних та безпечних лікарських засобів.

### 3. Бази даних хемоінформатики та біоінформатики, що використовують у процесах комп'ютерного конструювання ліків

Необхідною умовою застосування ліганд-орієнтованого або структура-орієнтованого типів комп'ютерного конструювання ліків є наявність і доступність інформації про структури біологічних мішеней або потенційних лігандів із відомою біоактивністю. Протягом останніх десятиліть за допомогою експериментальних методів структурної біології, фармакологічного скринінгу та органічного синтезу було отримано масштабні дані про сотні тисяч біомолекул та низькомолекулярних сполук. Тому необхідною передумовою досягнення максимальної ефективності у використанні загальнодоступних баз даних в галузях хемо- та біоінформатики, та з метою використання повного потенціалу наявної в них інформації, є знання того, яка саме інформація міститься у кожній базі даних, основні принципи її структуризації, індексування та пошуку.

#### 3.1. Хемоінформатика та біоінформатика як сучасні мультидисциплінарні галузі знань

Хемоінформатика – це новий мультидисциплінарний напрям сучасної хімічної науки, у якому запроваджено використання інформаційних технологій та інструментів для зберігання, аналізу, оброблення та візуалізації великої кількості

хімічної інформації. Також хемоінформатика є сучасним обчислювальним інструментом для індексування та пошуку хімічної інформації, необхідної для виконання завдань, пов'язаних із конструюванням ліків, кращим розумінням механізмів хімічної взаємодії, стандартизації протоколів виробництва ліків тощо. [17]. Таким чином, трьома основними аспектами хемоінформатики є: 1) збір інформації, тобто процес генерування та збору хімічної інформації про сполуки, отриманої як емпіричним шляхом (за експериментальними результатами синтезу та аналізу), так і теоретичної інформації, одержаної в результаті застосування методів молекулярного моделювання; 2) адміністрування інформації, що уможлиблює її зберігання та ефективний пошук; 3) використання інформації, що передбачає аналіз даних, пошук кореляцій та побудова моделей, а також її застосування у хімічних та біохімічних науках, зокрема для прогнозування властивостей потенційних лікарських засобів, віртуальний скринінг хімічних бібліотек та інші [18].

*Біоінформатика* – це міждисциплінарна галузь знань, яка використовує обчислювальні інструменти та алгоритми для вивчення структури і функцій біомолекул та аналізу біологічних даних. Шляхом дослідження та інтерпретації величезної кількості біологічної інформації, зокрема нуклеотидних послідовностей РНК і ДНК, структури протеїнів та профілів експресії генів, біоінформатика отримує цінну інформацію про молекулярні механізми, які лежать в основі захворювань, що дає змогу ідентифікувати потенційні терапевтичні мішені.

Класичними об'єктами дослідження у біоінформатиці є вивчення послідовностей нуклеотидів РНК та ДНК у генах або повних геномів, встановлення амінокислотних послідовностей протеїнів, визначення тривимірних структур протеїнів, нуклеїнових кислот і білково-нуклеїнових комплексів. Інші типи даних, якими оперує біоінформатика, отримуються при застосуванні оміксних технологій до біологічних об'єктів та включають у себе зокрема [19]:

- геноміку (вивчає всі ДНК організму, тобто його геном, дозволяє ідентифікувати та охарактеризувати всі гени та функціональні елементи у геномі, а також способи їх взаємодії);

- епігенетику (вивчає спадкові зміни у фенотипі або експресії генів, що зумовлені іншими механізмами, ніж зміна нуклеотидної послідовності в ДНК, наприклад, метилювання ДНК або модифікацію гістонів);

- транскриптоміку (вивчає транскриптому, тобто сукупності всіх РНК-транскриптів організму, а також рівні регуляції експресії генів);

- протеоміку (ідентифікація та кількісне визначення білків у складних біологічних зразках, їх субклітинної локалізації та функцій, прогнозування тривимірних структур, вивчення протеїн-протеїнових взаємодій);

- метаболоміку (комплексне вивчення низькомолекулярних сполук (метаболітів) у біологічних системах, у тому числі закономірності їх перетворень біохімічними шляхами);

- інтерактоміку (дослідження закономірностей взаємодії між молекулами, зокрема взаємодій протеїн-протеїн та протеїн-нуклеїнова кислота).

Основними завданнями, які покликана вирішувати біоінформатика, є [20]:

- Розробка ефективних алгоритмів для вимірювання подібності послідовностей. Сучасними інструментами для вирівнювання послідовностей та пошуку гомологічних послідовностей є BLAST, FASTA, SSearch, GL Search та інші. Послідовності можна задати у вигляді файла або у форматі fasta. Наприклад, програма BLAST (Basic Local Alignment Search Tool) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), розроблена Національним центром біотехнологічної інформації (National Center for Biotechnology Information, NCBI), використовує алгоритм Сміта-Вотермена для по-



рівняння амінокислотної послідовності у досліджуваному протеїні або послідовності нуклеотидів з нуклеотидними послідовностями у цільовій базі даних з метою ідентифікації областей локального вирівнювання та створення профілю із вирівнювання множинних послідовностей, подібність яких вище заданого порогового значення.

- Розширення експериментальних даних шляхом генерування прогнозованих біологічних даних. Фундаментальною метою обчислювальної біології є передбачення структури білка за амінокислотною послідовністю.

- Прогнозування взаємодії між протеїнами, враховуючи індивідуальні структури окремих білків. Застосування докінгу для дослідження протеїн-протеїнових комплексів дозволяє виявити, що протеїни у таких комплексах демонструють високу комплементарність за формою поверхні та полярністю і стабілізуються переважно за рахунок слабких взаємодій, зокрема шляхом утворення водневих зв'язків та сил Ван-дер-Ваальса. Окремим завданням, яка може мати важливе терапевтичне застосування, є розробка антитіл, які здатні з високою спорідненістю зв'язуватись із цільовим білком.

У процесах конструювання ліків дані біоінформатики використовують перш за все для порівняння певного набору даних між носіями досліджуваної симптоматики (пацієнти, тестування захворювань на тваринах, культивування клітин *in vitro* тощо) і контрольними групами без ознак цієї симптоматики. Такі порівняння здатні успішно вирішувати декілька завдань: 1) пов'язати симптоми захворювання з генетичними мутаціями, епігенетичними модифікаціями та іншими факторами навколишнього середовища, що модулюють експресію генів; 2) ідентифікувати біологічні мішені ліків, які можуть або відновлювати порушені клітинні функції, або знищувати «несправні» клітини; 3) прогнозувати потенційні лікарські засоби, які можуть впливати на цільову біомішень для досягнення запланованого терапевтичного результату та мінімізації побічних ефектів; 4) оцінити вплив лікарських засобів на навколишнє середовище та потенційну появу резистентності до них [21].

Проте на даний час спостерігається тенденція до зближення біоінформатики та хімоінформатики у формі запровадження перехресних посилань. Компоненти візуалізації низки веб-ресурсів спрямовано на візуалізацію не лише окремих молекул, але й молекулярних механізмів перебігу процесів у біологічних системах, які охоплюють шкалу від окремих клітин до цілих організмів. Зрештою, такі зусилля можуть уможливити створення персоналізованих профілів функцій клітин і тканин, що може призвести до більш спрямованих підходів до спрямованого конструювання ліків, діагностики та лікування захворювань.

### 3.2. Бази даних хемоінформатики

Хімічна інформація, яку акумулюють і адмініструють сучасні хемоінформатичні бази даних, складається з хімічних формул сполук, їх хімічних структур та властивостей, спектральних характеристик тощо. Сучасні бази даних хемоінформатики містять інформацію здебільшого про хімічні сполуки з невеликою молекулярною масою (до 500). Це пов'язане з тим, що саме невеликі молекули, які здатні проявляти «лікоподібні» властивості, є важливими при вивченні молекулярних та клітинних функцій організмів, оскільки вони здатні модулювати певні біологічні мішені, що проявляється у їх терапевтичній ефективності. Більшість існуючих на даний час ліків належить до цього класу речовин [22].

Важливою характеристикою сучасних хемоінформатичних баз даних є те, що вони анують хімічні дані інформацією з інших наукових областей, наприклад, генетики, геноміки, фармакології, токсикології, медицини, біоінформатики та системної біології, що є надзвичайно корисним при їх використанні у процесах конструювання ліків.

Завдяки бібліотекам хімічної інформації медична хімія отримала потужний інформаційний ресурс для ідентифікації потенційних біологічно активних речовин при проведенні широкомасштабного віртуального скринінгу. На даний час суттєво збільшилась кількість хімічних баз даних та засобів програмного забезпечення, що є у вільному доступі – із відкритим вихідним кодом або з веб-підтримкою.

Існує три типи хемоінформатичних баз даних: 1) архівовані, або «глобальні»; 2) спеціалізовані, з високим рівнем адміністрування; 3) структурні бази даних [17].

❖ Глобальні хімічні бази даних зберігають інформацію про велику кількість хімічних сполук, їх властивості та біологічну активність. Прикладом архівованої бази даних хімічних сполук є PubChem (<https://pubchem.ncbi.nlm.nih.gov/>).

❖ Спеціалізовані бази даних є меншими ресурсами порівняно з глобальними. Вони зазвичай містять інформацію про тисячі чи десятки тисяч біоактивних сполук, проте, на відміну від глобальних баз, вони також містять детальну інформацію про цикли біологічних перетворень низькомолекулярних речовин в організмі, біологічні мішені, механізми дії, асоційовані з патогенезом відповідних захворювань, токсикологічні дані та фармакогеномні наслідки. Більшість із цих баз даних також мають широкий інструментарій пошуку і можливостей перегляду, включаючи пошук тексту, послідовності та структурної подібності. ChEBI (<https://www.ebi.ac.uk/chebi/>) та DrugBank (<https://go.drugbank.com/>) є спеціалізованими базами даних з високим рівнем адміністрування.

❖ Третьою категорією хемоінформатичних баз даних є структурні бази, що містять 3D координати сполук. ZINC20 (<https://zinc.docking.org/>) можна віднести до структурної бази даних. Деякі бази даних, наприклад, Cambridge Structure Database (<http://www.ccdc.cam.ac.uk>), містять експериментально визначені 3D координати хімічних сполук. Більш детальну характеристику деяких із зазначених баз даних буде наведено нижче.

Також очевидно, що для здійснення ефективного пошуку у хемоінформатичних базах даних критичною необхідністю є розробка і доступність інструментів пошуку, тобто програмного забезпечення для введення та зчитування хімічної інформації. Такими інструментами у теперішній час слугують, наприклад, комп'ютерний застосунок OpenBabel (<http://openbabel.org/>), який використовують як програмне забезпечення для конвертації форматів файлів, MolConvert від Chemaxon (<https://docs.chemaxon.com/display/docs/molconvert.md>) та Corina від Molecular Networks GmbH and Altamira, LLC (<https://mn-am.com/products/corina/>), які використовують як програмне забезпечення для прогнозування 2D і 3D структур хімічних сполук, відповідно.

Короткий огляд деяких важливих хемоінформатичних баз даних, що є у вільному доступі, представлено на рис. 3. Для них наведено уніфіковані локатори ресурсів (URL), які також вказані нижче, у більш детальному описі кожної бази даних.

Ці бази даних суттєво різняться за вмістом, розміром (рис. 4) та якістю. Деякі з них надають доступ до широкого спектру інформації про мільйони молекул, агрегованих з сотень зовнішніх джерел. Інші містять спеціалізовану інформацію про кількості тисяч молекул, яка підтримується та адмініструється.

Оскільки поширеною практикою є обмін даними між цими загальнодоступними базами даних, у них трапляються деякі збіги та дублювання вмісту, проте кожна база даних має унікальне наповнення, цінне для отримання певного виду необхідної інформації. Також кожна з цих баз даних забезпечує користувачів унікальними інструментами для пошуку, перегляду, візуалізації та завантаження інформації.

• ZINC20 (<https://zinc20.docking.org/>) – це оновлена версія ZINC, відкритої бази даних комерційно доступних сполук для проведення віртуального скринінгу [23]. На даний час ZINC містить інформацію про понад 230 мільйонів доступних для при-

дбання хімічних речовин, хімічні структури яких доступні для завантаження у 2D і 3D форматах, що є зручним для проведення їх подальшого докінгу. Веб-ресурс забезпечує швидкий пошук серед хімічних структур та їх аналогів. ZINC також містить інформацію про понад 750 мільйонів комерційно доступних речовин, пошук яких можна здійснювати на сайті як за базовими структурами, так і за структурами аналогів. Такий веб-дизайн сайту повністю відповідає стрімкому зростанню хімічного простору: від менше 1 мільйону сполук у 2005 році до майже 2 мільярдів на сьогодні.

<ul style="list-style-type: none"> <li>• <b>ZINC20</b> (<a href="https://zinc20.docking.org/">https://zinc20.docking.org/</a>) Містить інформацію про &gt; 230 млн. хімічних сполук, доступних для завантаження у 2D і 3D форматах, а також містить інформацію про &gt; 750 млн. комерційно доступних хімічних речовин.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>ChemSpider</b> (<a href="http://www.chemspider.com/">http://www.chemspider.com/</a>) Містить інформацію про &gt; 128 млн. сполук, зібрану з ~ 500 джерел.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>PubChem</b> (<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>) Надає доступ до комплексної інформації про &gt; 128 млн. унікальних сполук, зібраної з &gt; 7500 джерел, а також експериментальні дані щодо їх біологічної активності.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>ChEMBL</b> (<a href="https://www.ebi.ac.uk/chembl/">https://www.ebi.ac.uk/chembl/</a>) Містить інформацію про ~ 2,4 млн. біологічно активних низькомолекулярних сполук та дані щодо їх біологічної активності, реферовані з першоджерел</li> </ul>
<ul style="list-style-type: none"> <li>• <b>DrugBank</b> (<a href="https://go.drugbank.com/">https://go.drugbank.com/</a>) Забезпечує доступ до комплексної інформації про &gt; 500 тис. лікарських засобів, затверджені FDA, а також потенційні лікарські засоби на різних стадіях досліджень.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>Відкрита база даних Національного Інституту раку США (Open National Cancer Institute (NCI) database)</b> (<a href="https://cactus.nci.nih.gov/">https://cactus.nci.nih.gov/</a>) Надає доступ до 250250 хімічних сполук, включено можливість прогнозування для них різних видів біологічної активності.</li> </ul>

Рис. 3. Загальна характеристика хімічних баз даних із відкритим доступом

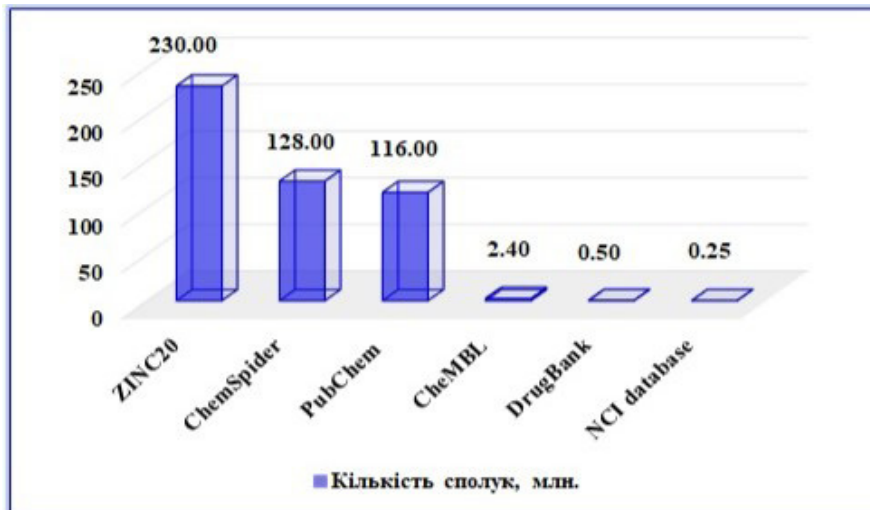


Рис. 4. Кількість хімічних сполук, що містяться у хемоінформатичних базах даних, що розглядаються у цьому огляді (станом на листопад 2023 р.)

Окрім даних про хімічні сполуки, ZINC також містить набір інструментів для їх ефективного пошуку. Важливим новим інструментом ZINC є пошук аналогів біологічно активної молекули з використанням так званого підходу «аналоги-за-каталогом», заснованого на принципі молекулярної подібності для ідентифікації сполук, подібних до вихідної, для яких можна створювати і досліджувати зв'язки

«структура–активність». Нові безкоштовні інструменти ZINC дають змогу створювати бібліотеки хімічних структур «на вимогу» шляхом здійснення пошуку за подібністю як для структури цілої молекули, так і для структур окремих її фрагментів (підструктур, або будівельних блоків). Пошук у базі даних за подібністю для всієї молекули проводиться з використанням SmallWorld – методу «відстані редагування графів» (graph-edit-distance) і пошуку максимального подібного підграфа, анованого традиційними методами пошуку подібності за «відбитками пальців» або обчислених коефіцієнтів подібності Tanimoto. Пошук подібності для підструктур здійснюють зі застосуванням підходу Arthor, що використовує налаштований збіг шаблонів хемотипу, який дозволяє вказувати типи атомів та варіанти їх заміни, бажану наявність певних функціональних груп і скафолдів, комбінації функціональних груп, тому цей підхід є зручним у використанні з метою пошуку реагентів для побудови віртуальних бібліотек сполук.

- *ChemSpider* [24] (<http://www.chemspider.com/>) – це безкоштовна онлайн-база даних хімічних речовин, що надає доступ до їх фізичних та хімічних властивостей, молекулярної структури, спектральних даних, методів синтезу, інформації про токсичність та номенклатуру для приблизно 128 мільйонів унікальних хімічних сполук, отриманих і пов'язаних із майже 400 окремими джерелами даних в Інтернеті.

- *PubChem* (<https://pubchem.ncbi.nlm.nih.gov/>) – це база даних хімічних сполук та їх активностей, визначених на різних біологічних моделях і за різними протоколами [25]. Система підтримується Національним центром біотехнологічної інформації (NCBI), підрозділом Національної медичної бібліотеки, яка є частиною Національного інституту здоров'я США (NIH). Доступ до PubChem із можливістю завантаження структурної інформації можна отримати безкоштовно через веб-інтерфейс користувача. Дані у PubChem систематизовані у вигляді трьох окремих менших баз: Речовини (Substance), Сполуки (Compound) і Біопротоколи (BioAssay), у яких можна здійснювати пошук хімічних речовин за назвою, молекулярною формулою, структурою та іншими ідентифікаторами, визначати їх хімічні та фізичні властивості, біологічну активність, знаходити інформацію про безпеку і токсичність, патенти, зовнішні посилання тощо.

- *ChEMBL* [26] (<https://www.ebi.ac.uk/chembl/>) – це відкрита база даних, яка містить інформацію про типи зв'язування, біологічну активність та ADME-Tox властивості великої кількості «лікоподібних» сполук. Ці дані регулярно реферуються вручну за первинними науковими публікаціями, а потім додатково систематизуються та стандартизуються, щоб максимально збільшити ефективність їх застосування у широкому діапазоні дослідницьких проблем хімічної біології та розробки ліків. На даний момент база даних містить 5,4 мільйона записів про біологічну активність для близько 2,4 мільйона низькомолекулярних сполук та 5 200 білків-мішеней. Доступ до бази здійснюється через веб-інтерфейс, який надає доступ до завантаження даних і ряду веб-сервісів.

- *DrugBank* [27] (<https://go.dragbank.com/>) – це унікальний веб-ресурс для біоінформатики / хемоінформатики, який поєднує детальні дані про ліки (тобто хімічні речовини) з вичерпною інформацією про їх біологічні мішені (цільові білки). База даних містить близько 13 850 записів про біологічно активні речовини, включаючи > 2 660 схвалених FDA низькомолекулярних і біотехнологічних препаратів, а також > 3 200 кандидатів у лікарські засоби на різних стадіях розробки. Крім того, у бані наявна інформація про > 14 000 цільових послідовностей білків або інших біологічних мішеней, пов'язаних із певним низькомолекулярним лігандом. До багатьох записів створено гіперпосилання на інші бази даних (KEGG, PubChem, ChEBI, PDB, Swiss-Prot і GenBank) і різноманітні аплети для перегляду структури. База да-

них повністю доступна та підтримує розширений пошук по тексту, послідовності, хімічній структурі та за іншими пов'язаними запитамі. Потенційні застосування DrugBank включають *in silico* виявлення біологічних мішеней, докінг або інші види віртуального скринінгу, прогнозування метаболізму ліків та ін.

• *Відкрита база даних Національного Інституту раку США (Open National Cancer Institute (NCI) database)* (<https://cactus.nci.nih.gov/>) створена групою NCI/CADD, що є дослідницьким підрозділом лабораторії хімічної біології Національного інституту раку. База хімічних структур NCI – це колекція з приблизно півмільйона структур, накопичена протягом останніх 45 років під час скринінгу сполук на протиракову активність, а пізніше також на дію проти ВІЛ-інфекції. Приблизно 50% сполук бази даних захищена угодами про конфіденційність із постачальниками зразків, тоді як інша половина є відкритою і доступною для завантаження, при цьому користувач може сортувати і завантажувати з використанням веб-браузера близько 250 000 структур за більш ніж 600 критеріями. До веб-сервісу також включено можливість прогнозування порядку 550 різних видів біологічної активності, розрахованих програмою PASS.

### 3.3. Бази даних біоінформатики

Як правило, бази даних біоінформатики складаються з великих колекцій послідовностей та/або структур білків, РНК або ДНК. Біоінформатичні бази даних також поділяються на (1) архівовані, або «глобальні», прикладом яких є GenBank, що є базою послідовностей нуклеотидів (<https://www.ncbi.nlm.nih.gov/nucleotide/>); (2) спеціалізовані, з високим рівнем адміністрування, до яких належать бази даних SwissProt (<https://www.expasy.org/resources/uniprotkb-swiss-prot>) та UniProt (<https://www.uniprot.org/>); (3) структурні бази даних, наприклад PDB (<https://www.rcsb.org/>). Окремою групою баз даних у біоінформатиці є бази даних біологічних механізмів, що містять інформацію про біологічні процеси у серіях молекулярних подій. Прикладами таких баз даних є Reactome (<https://reactome.org/>) і KEGG (<https://www.genome.jp/kegg/>).

Короткий огляд деяких важливих біоінформатичних баз даних, які є у вільному доступі, представлено на рис. 5. Для них наведено уніфіковані локатори ресурсів (URL), які також будуть вказані нижче, при більш детальному описі кожної бази даних.

• *GenBank* ([www.ncbi.nlm.nih.gov/Genbank](http://www.ncbi.nlm.nih.gov/Genbank)) – це повна база даних нуклеотидних послідовностей і допоміжних бібліографічних і біологічних анотацій із відкритим доступом [28]. Базу було створено та підтримується Національним центром біотехнологічної інформації (National Center for Biotechnology Information, NCBI), підрозділом Національної медичної бібліотеки (National Library of Medicine, NLM) Національного інституту здоров'я США (US National Institutes of Health, NIH) у м. Бетесда, штат Меріленд, США. Станом на жовтень 2023 р. містить загалом 19,6 трильйонів пар основ із понад 2,9 мільярдів нуклеотидних послідовностей РНК та ДНК для 504 000 офіційно описаних видів організмів. GenBank доповнюється даними як з окремих лабораторій, так і з великомасштабних проєктів повного секвенування геномів. Щоденний обмін даними з Європейським нуклеотидним архівом (ENA) і Банком даних ДНК Японії (DDBJ) забезпечує надходження до бази інформації зі всього світу. GenBank категоризує послідовності за кількома розділами на основі таксономії джерела або стратегії секвенування, використаної для отримання даних. Отже, GenBank складається з 12 таксономічних підрозділів і 5 високопродуктивних підрозділів. Окрім того, підрозділ PAT містить дані, надані патентними відомствами, а підрозділ TSA містить послідовності, надані проєктами зі складання транскриптомів. Кожному запису GenBank (нуклеотидна послідовність та анотація до неї) присвоюється унікальний ідентифікатор – номер доступу, що є спільним для

трьох баз даних (GenBank, DDBJ, ENA). Записи послідовностей у GenBank доступні через пошукову систему NCBI Entrez. Записи GenBank зберігаються у базі даних Nucleotide. Послідовності GenBank, які є частиною популяційних або філогенетичних досліджень, також акумулюються у базі даних PopSet, а концептуальні трансляції послідовностей CDS доступні також у базі даних Protein. Кожна з цих баз даних має відповідні посилання на оригінальне опубліковане дослідження у PubMed.

<ul style="list-style-type: none"> <li>• <b>GenBank</b> (<a href="http://www.ncbi.nlm.nih.gov/Genbank">www.ncbi.nlm.nih.gov/Genbank</a>). Містить дані геноміки, транскриптоміки та інформацію про протеїни. Станом на жовтень 2023 р. містить загалом 19,6 трильйонів пар основ із понад 2,9 мільярдів нуклеотидних послідовностей РНК та ДНК для 504 000 офіційно описаних видів організмів.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>PDB</b> (<a href="https://www.rcsb.org/">https://www.rcsb.org/</a>). Архівує структурні дані у 3D форматі для біологічних макромолекул (білків, ДНК і РНК), що є пріоритетно важливими для наукових досліджень і освіти у галузях фундаментальної біології, охорони здоров'я, енергетики та біотехнології. Станом на 04.10.2023 р. містить дані про 212061 експериментально встановлених структур та 1068577 теоретично передбачених структур.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>UniProt</b> (<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>). Випуск 2023_05 від 08.11.2023 р. бази даних UniProtKB/Swiss-Prot містить дані про 570420 амінокислотних послідовностей, реферованих з 295467 оригінальних посилань, які сумарно містять 206321560 амінокислот.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>Therapeutic Target Database (TTD)</b> (<a href="http://db.idrblab.net/ttd/">http://db.idrblab.net/ttd/</a>). Містить інформацію про відомі та досліджені терапевтичні біологічні мішені - білки та нуклеїнові кислоти, а також інформацію про роль у молекулярних механізмах цільового захворювання та відповідні лікарські засоби, які проявляють спрямовану дію на кожну з цих мішеней. Загалом містить дані про 426 клінічно підтверджених біомішеней, 1014 мішеней на стадії клінічних випробувань, 212 запатентованих мішеней, що перебувають на стадії доклінічних випробувань, та 1479 біомішеней, реферованих за оригінальними посиланнями.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>SuperTarget</b> (<a href="http://iisilico.clarite.de/supertarget/">http://iisilico.clarite.de/supertarget/</a>). На даний момент оновлена база даних містить дані про &gt; 6000 білків-мішеней, які анотовані &gt; 330 000 гіперпосиланнями до 196 000 сполук (включно із затвердженими препаратами); для переважної більшості комплексів ліганд-мішень наведено величини афінності.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>BRENDA</b> (<a href="https://www.brenda-enzymes.org/">https://www.brenda-enzymes.org/</a>). Містить понад 5000000 даних, реферованих вручну із ~157000 первинних літературних джерел, які анотують інформацію про ~90000 протеїнів з ~13000 організмів, доповнену інформацією, отриманою в результаті аналізу текстових описів та інших даних, а також при інтеграції даних із зовнішніх джерел та згенеровану з використанням алгоритмів прогнозування.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>SCOP</b> (<a href="https://scop.mrc-lmb.cam.ac.uk/">https://scop.mrc-lmb.cam.ac.uk/</a>). База даних надає детальний опис структурних та еволюційних зв'язків між усіма протеїнами з відомою структурою, який включає широкий огляд усіх відомих «складок» певного білка, докладну інформацію про споріднені до нього протеїни, що становить основу класифікації білків. Останнє оновлення від 29.06.2022 р. включає 72 544 модульні домени, що представляють 861631 структур протеїнів.</li> </ul>
<ul style="list-style-type: none"> <li>• <b>Reactome</b> (<a href="https://reactome.org/">https://reactome.org/</a>). Це адміністрована база даних про біологічні взаємодії між молекулами в клітині, яка також надає інструменти для візуалізації, інтерпретації та аналізу даних про молекулярні механізми процесів, аналізу генома та моделювання. Включає дані про 14803 біохімічні реакції, які відбуваються в організмі людини, організовані у 2647 ланцюги перетворень, у яких беруть участь 30338 білків і модифікованих форм білків, кодованих 11154 різними генами людини, дані про 14441 комплексів, 2025 низькомолекулярних сполук та 1119 лікарських засобів. Інформація анотована посиланнями з 37156 літературних джерел.</li> </ul>

Рис. 5. Загальна характеристика баз даних біоінформатики з відкритим доступом

- *Банк даних білків (Protein Data Bank, PDB)* (<https://www.rcsb.org/>) є найбільшою на даний час базою тривимірних структурних даних біологічних мішеней, що знаходиться у публічному домені та може використовуватись безкоштовно. Архів PDB містить >212 000 тривимірних структур біомолекул на атомному рівні, експериментально визначених за допомогою кристалографії, спектроскопії ядерного магнітного резонансу та електронної мікроскопії. RCSB PDB обслуговує мільйони користувачів у всьому світі, його дані є інтегрованими з ~40 зовнішніми ресурсами біологічних даних. Було виявлено, що архівні фонди PDB сприяли відкриттю

~90% із 210 нових ліків, схвалених Управлінням з контролю за продуктами й ліками США у 2010–2016 роках [29]. Майже 3/4 (~73%) структур PDB також містять дані про структуру одного або більше лігандів (наприклад, ферментних кофакторів та інгібіторів, лікарських засобів, схвалених FDA, і металів), а ~10% структур PDB включають один або більше вуглеводних компонентів. Архів PDB зараз вважається міжнародним суспільним надбанням, оскільки практично всі результати структурних досліджень було одержано за фінансової підтримки урядів різних країн або приватних благодійних організацій, відновна вартість поточних архівних фондів PDB консервативно оцінюється у >15 мільярдів доларів США.

• *Універсальна база даних протеїнів UniProt (Universal Protein Knowledgebase)* (<https://www.uniprot.org/>) – це повний, високоякісний і вільнодоступний ресурс амінокислотних послідовностей та функціональної інформації про протеїни [30]. UniProt складається з чотирьох компонентів, кожен з яких оптимізований для різних цілей: 1) База знань UniProt (UniProtKB) є центральною інформаційною частиною, що забезпечує доступ до великої систематизованої інформації про білки, включно з їх функціями, класифікацією та перехресними посиланнями. UniProtKB складається з двох розділів: 1.1) UniProtKB/Swiss-Prot, який анотують вручну та періодично переглядається; 1.2) UniProtKB/TrEMBL, який автоматично анотується та не переглядається. 2) Бази даних UniProt Reference Clusters (UniRef) надають кластеризовані набори амінокислотних послідовностей із UniProtKB та відповідні їм анотації, з метою забезпечення максимально повного покриття хімічного простору, приховуючи зайві послідовності. 3) Архів UniProt (UniParc) – це комплексне сховище, яке використовується для відстеження послідовностей та їхніх ідентифікаторів. Випуск UniProt 2022\_03 містить понад 227 мільйонів амінокислотних послідовностей в UniProtKB. Портал UniProtKB Proteomes (<https://www.uniprot.org/proteomes/>) надає доступ до більш ніж 451 000 протеомів, які є наборами білкових послідовностей, що походять від повністю секвованих вірусних, бактеріальних, архейних та еукаріотичних геномів. Крім того, платформа UniProt забезпечує можливість використання п'яти інструментів для аналізу структурних даних протеїнів: 1) BLAST є базовим інструментом пошуку локального вирівнювання для пошуку подібності амінокислотних послідовностей; 2) інструмент 'Align' використовують для вирівнювання кількох послідовностей; 3) інструмент 'Peptide search' застосовують для створення запитів з коротких пептидних послідовностей, щонайменше з трьох амінокислотних залишків, та пошуку всіх послідовностей UniProtKB, які мають точну відповідність послідовності запиту; 4) інструмент 'ID mapping' служить для використання списку ідентифікаторів з метою отримання пакетів записів UniProtKB і перетворення ідентифікаторів бази даних із UniProt у зовнішні бази даних або навпаки; 5) інструмент «SPARQL» є загальним пошуковим інструментом, що створює запити до всієї бази даних.

• *Therapeutic Target Database (TTD)* [31] (<http://db.idrblab.net/ttd>) – онлайн-ресурс, що містить дані експериментальної валідації 37 316 біологічно активних речовин і близько 3 419 біологічних мішеней. У цій базі даних біологічні мішені систематизовано за дев'ятьма категоріями, серед яких здатність протеїна бути мішенню для низькомолекулярних інгібіторів, агоністів, модуляторів, тощо, підтверджена закономірностями молекулярних взаємодій; функціональні властивості біомолекули у протеїн-протеїнових взаємодіях; профіль подібності мішені до протеїнів людини за межами свого сімейства; участь мішені у встановлених циклах метаболізму; розподіл по тканинах і органах людини. Базу було створено в результаті систематичного огляду реферованих посилань на 426 клінічно підтверджених біомішеней, 1 014 мішеней на стадії клінічних випробувань, 212 запатентованих

мішеней, що перебувають на стадії доклінічних випробувань, та 1 479 біомішеней, реферованих за оригінальними посиланнями.

- *SuperTarget* (<http://iisilico.cliarite.de/supertarget/>) [32] – ресурс з інформацією про 195 770 низькомолекулярних лікарських сполук і 6 219 біологічних мішеней.

- *BRENDA* (BRaunschweig ENzyme DAtabase) (<https://www.brenda-enzymes.org/>) була створена у 1987 р. у Німецькому національному дослідницькому центрі біотехнології в Брауншвейзі (GBF) [33], у теперішній час база підтримується Інститутом біохімії Кельнського університету. Цю інформаційну систему було розроблено для збору та зберігання даних щодо функціональних властивостей ферментів та їх метаболізму, зібраних з первинних наукових джерел. База містить дані щонайменше про 90 000 різних ферментів із 13 000 організмів, біохімічну та молекулярну інформацію про їх класифікацію та номенклатуру, реакційну здатність та специфічність, функціональні параметри, структуру, застосування, стабільність, пов'язані захворювання, способи виділення, літературні посилання. Класи ферментів визначаються відповідно до типу каталізованих реакцій, на теперішній час база класифікує ензими за близько 7 600 типів різних реакцій. Реферовані дані, які складають базу, отримано з приблизно 157 000 посилань на оригінальні дослідження, які цитуються в PubMed. Інформація про ензими анотована реферованими даними про пов'язані захворювання, даними амінокислотних послідовностей протеїнів, їх 3D-структурами, також наведено прогнозовані активні сайти ферментів та анотації геномів.

- *SCOP* (Structure classification of Proteins) (<https://scop.mrc-lmb.cam.ac.uk/>) – це база даних, яка класифікує білкові домени відповідно до областей структурної подібності, використовуючи як ручне, так і автоматичне адміністрування [34]. Ця база даних використовує відомі структури з PDB для ієрархічної класифікації конформацій білків на основі структурних та еволюційних зв'язків. Користувачі можуть ідентифікувати відомі конформації білка і використовувати цю інформацію, щоб визначити потенційні функції досліджуваного протеїна.

- *Reactome* (<https://reactome.org/>) – це адміністрована база даних, яка містить інформацію про молекулярні взаємодії та механізми у широкому діапазоні фізіологічних і патологічних біологічних процесів людини, включаючи також спадкові та набуті захворювання. Біохімічні процеси у базі анотуються як упорядкована мережа молекулярних перетворень в єдиній узгодженій моделі даних. Таким чином, Reactome функціонує як цифровий адміністрований архів біологічних процесів людини і, у той же час, як інструмент для виявлення функціональних зв'язків серед таких даних, як профілі експресії генів або каталоги соматичних мутацій пухлинних клітин. Також ручне адміністрування бази дозволило розширити анотації нормальних і пов'язаних із захворюваннями сигнальних процесів і препаратів, що спричиняють дію на дані патологічні ланки, включно з інфекційними та вірусними процесами. Нові інструменти підтримують кращий одночасний аналіз високопродуктивних даних із кількох джерел і розміщення недостатньо вивчених («темних») білків із проаналізованих наборів даних у контексті введених вручну механізмів.

Крім окремих баз даних, на сьогодні створено великі портали хімічної та біологічної інформації, що об'єднують інформацію з декількох баз та забезпечують її використання за допомогою спеціально розроблених інструментів пошуку і навігації. Прикладами таких величезних конгломератів веб-даних, що є у вільному доступі, можуть слугувати EMBL-EBI та ExPASy.

- ❖ Веб-портал *EMBL-EBI* створено Європейським інститутом біоінформатики (European Bioinformatic Institute, EBI) ([www.ebi.ac.uk](http://www.ebi.ac.uk)), який є підрозділом Європейської лабораторії молекулярної біології (EMBL) та спеціалізується на дослідженнях та послугах у галузі біоінформатики. Веб-портал надає доступ до широкого спектра



основних баз даних і інструментів аналізу, що мають ключове значення у біоінформатиці. Окрім надання веб-інтерфейсів до цих ресурсів, веб-сервіси порталу забезпечуються програмним доступом за допомогою протоколів REST і SOAP та дозволяють інтегрувати їх в інші програми та аналітичні робочі процеси. Ресурси даних EMBL-EBI включають: бази даних, які архівують експериментальну інформацію; доповнювані бази даних, які забезпечують анотування, адміністрування, повторний аналіз та інтеграцію депонованих даних; інструменти програмного забезпечення з відкритим вихідним кодом, які дозволяють повторно використовувати ці ресурси. Двома найбільшими архівними ресурсами є Європейський архів нуклеотидів (ENA) та Європейський архів геному-феному (EGA), на які припадає понад 90% усіх даних порталу на сьогодні. Помітно швидке зростання даних за останні роки спостерігалось для ресурсу даних зображень – BioImage Archive (BIA), який зберігає та надає доступ до біологічних зображень, пов'язаних з рецензованою публікацією. Іншими веб-ресурсами є: ArrayExpress (база експериментальних даних функціональної геноміки), BioModels (репозиторій рецензованих опублікованих обчислювальних моделей), BioSamples (база даних, що описує біологічні зразки та надає посилання на відповідні експериментальні дані), BioStudies (база даних, що містить описи біологічних досліджень, посилання на них у інших базах даних у складі EMBL-EBI або за її межами), ChEBI (база еталонних хімічних структур, їх номенклатура та онтологічна класифікація), ChEMBL (база даних, пов'язаних з біологічно активними сполуками – функціональні дані прояву їх біоактивності та ADME-Tox властивості), а також багато інших ресурсів та інструментів. Новим ресурсом EMBL-EBI є AlphaFold – система штучного інтелекту, розроблена компанією DeepMind, яка дає змогу прогнозувати тривимірну структуру протеїна за його амінокислотною послідовністю з точністю, яка може конкурувати з експериментальною.

❖ *ExPASy* – це веб-портал ресурсів біоінформатики Швейцарського інституту біоінформатики (Bioinformatics Resource Portal, SIB) (<http://expasy.org/>), який надає доступ до наукових баз даних і програм для їх аналізу в різних галузях біомедичних наук. Цей інтегрований портал надає доступ до понад 160 баз даних і програмних інструментів, розроблених SIB Groups у галузі ряду наук про життя та клінічних досліджень, від геноміки, протеоміки, транскриптоміки та структурної біології до еволюції та філогенії, системної біології та медичної хімії.

#### 3.4. *Посаднання досягнень хемоінформатики та біоінформатики у процесах конструювання ліків*

Хемоінформатика та біоінформатика доповнюють одна одну у галузі дослідження біомолекулярних процесів, зокрема при вивченні структури та функцій білків, зв'язування ліганду із біологічною мішенню та видів взаємодій у ділянках зв'язування, перетворення субстрату в складі фермент-субстратного комплексу та ферментативного каталізу біохімічних реакцій. Розробка нових методів та інструментів хемо- і біоінформатики дозволяє збільшувати ефективність адміністрування накопичених біологічних та хімічних даних та їх використання.

Інструменти хемоінформатики та біоінформатики є взаємодоповнюючими на різних етапах процесу комп'ютерного конструювання ліків (рис. 7, адаптований за *Raslan M. A. et al.* [36]). Ідентифікувати біологічну мішень на початковому етапі раціонального конструювання ліків можна за допомогою різних біоінформатичних інструментів, наприклад, методами геноміки та протеоміки. Пошук і оптимізацію сполук-хітів та сполук-лідерів зазвичай здійснюють з використанням ряду інструментів хемоінформатики, наприклад, аналізу та скринінгу хімічних баз даних, QSAR аналізу та *in silico* прогнозування ADME-Tox властивостей з метою пошуку потенційних лігандів з мінімальними побічними ефектами.

Сучасні наукові тенденції в галузі конструювання ліків спрямовані на застосування існуючих алгоритмів пошуку та аналізу для обробки величезної кількості накопичених цифрових даних [37]. Такий підхід сприяє розширеному прогнозуванню біологічних особливостей поведінки біомолекул при їх модуляції потенційними лігандами. У процесах конструювання більш ефективних та безпечних ліків широко запроваджується інтеграція структурної, генетичної інформації та фармакологічних даних від рівня окремих молекул до рівня всього організму.



Рис. 6. Застосування інструментів біоінформатики та хемоінформатики на різних етапах процесу конструювання ліків

Деякі приклади результативного застосування низки інструментів хемоінформатики та біоінформатики у галузі конструювання ліків та класифікації захворювань наведено в таблиці, зокрема у разі розроблення нового антибіотика, діагностиці COVID-19 на основі даних рентгенографії і комп'ютерної томографії, при ідентифікації мікрорезистентних організмів та з метою пошуку нових протипухлинних агентів – інгібіторів гену ECT2.

Велика кількість існуючих баз даних хімічної та біологічної інформації зумовила необхідність розробки великомасштабного веб-інструмента, який би дозволив встановлювати зв'язки та здійснювати пошук між всіма збереженими і описаними хімічними структурами та їхніми хімічними ідентифікаторами з усіх окремих хімічних ресурсів. Таким веб-інструментом може слугувати UniChem.

UniChem (<https://www.ebi.ac.uk/unichem/>) – це масштабна служба відображення ідентифікаторів та гіперпосилань між хімічними структурами та інтернет-ресурсами, що містять первинну хімічну інформацію про дану структуру, а також іншими базами даних, що стосуються макромолекул, проте містять інформацію, певним чином пов'язану із зазначеною хімічною сполукою (наприклад, UniProt, PDB тощо) [42]. UniChem є безкоштовним вільнодоступним інтернет-сервісом, у якому обмін даними між веб-інструментом і браузером надається за допомогою безпечного протоколу HTTPS.

**Узагальнена інформація про успішне застосування інструментів  
хемоінформатики та біоінформатики у конструюванні ліків**

№	Галузь знань, інструмент	Медична галузь	Результат
1	Хемоінформатика (поліпшення фармакологічного профілю, зокрема показників біодоступності та стійкості, шляхом обчислення показника відносної біодоступності)	Розробка антибіотиків	Тетрацикліновий аналог В (йодоциклін). Більш активний, ніж тетрациклін, і менш бактеріорезистентний [38]
2	Біоінформатика (застосування методів глибокого навчання, зокрема методу створення конволюційної нейронної мережі, Convolutional Neural Networks (CNNs), та розробка моделі логістичної регресії, що складається з ансамблю з п'яти класифікаторів: SGD, SVM, Naive Bayes, Random Forest і KNN	Діагностика захворювання	Ґрунтуючись на даних рентгенографії і комп'ютерної томографії, було проведено дослідження можливості діагностики COVID-19. Було показано, що таку діагностику можливо забезпечити шляхом використання декількох класифікаторів, нашарованих у вигляді ансамблю. Кількісна оцінка результатів дослідження вказує на ефективність запропонованої техніки на рівні 99% точності діагностики [39]
3	Хемоінформатика (біоізостерне заміщення скафолда, SAR аналіз, докінг); Біоінформатика (метод 3D-SAR аналізу з метою біологічно-релевантного порівняння молекул, зокрема з метою пошуку кращих конформерів шляхом порівняння бази конформерів з еталонною молекулою)	Визначення антибактеріальної активності відносно резистентних мікроорганізмів	Три досліджені сполуки на основі піразолу та бензімідазолу показали помірну бактерицидну ефективність проти MSSA, MRSA та ванкомицин-резистентного <i>Staphylococcus aureus</i> (VRSA) [40]
4	Хемоінформатика (генетична зміна ECT2 та її кореляція з виживанням пацієнтів, молекулярний докінг із наступним моделюванням методом молекулярної динаміки)	Дія на ракові клітини	Результати дослідження демонструють, що ген ECT2 здатний підвищувати концентрацію як мРНК, так і білка при різних типах раку людини, тим самим забезпечуючи більше елімінування супресорних клітин мієлоїдного походження (MDSC – (myeloid-derived suppressor cells) і зменшуючи популяцію природних Т-лімфоцитів, що безпосередньо впливає на прогноз виживання хворого. У дослідженні здійснено пошук протипухлинних агентів – інгібіторів ECT2 [41]

На сьогодні UniChem містить інформацію про більш ніж 145 мільйонів хімічних структур та акумулює дані з 40 джерел, включаючи такі ресурси:

– бази даних первинної хімічної інформації, що відноситься до низькомолекулярних сполук, зокрема ChEMBL, DrugBank, PubChem, ZINC, NIH Clinical Collection та інші;

– бази даних, які мають інший основний фокус, наприклад PDBe та Gene Expression Atlas, основними цільовими об'єктами яких є структури протеїнів та

дані про експресію генів відповідно, проте ці дані часто анотуються інформацією про низькомолекулярні речовини, наприклад, про положення та взаємодію хімічної сполуки-інгібітора у центрі зв'язування білкової макромолекули, або про регуляцію транскрипції генів низькомолекулярними сполуками.

Таким чином, веб-сервіси в UniChem надають користувачам єдиний простий автоматизований механізм для підтримки усіх посилань з їхнього ресурсу на всі інші джерела, представлені в UniChem, тобто для кожного надісланого запиту усі посилання на та з усіх інших джерел автоматично ідентифікуються та стають доступними для використання. Тому веб-сервіс UniChem можна вважати гнучким інтеграційним рішенням для створення, підтримки та адміністрування зв'язків та гіперпосилань між іншими інтернет-ресурсами хімічної та біологічної інформації, яке потребує мінімальних зусиль з боку користувача, важливою властивістю якого є також можливість залучати нові додаткові джерела в майбутньому.

Завдяки стрімкому зростанню кількості загальнодоступних онлайн-ресурсів було розроблено багато методів прогнозування, заснованих на машинному навчанні, для встановлення взаємодії лікарських засобів та біологічних рецепторів, кількісної оцінки проникності гематоенцефалічного бар'єру для різних речовин [43] та прогнозування ADME-Tox властивостей потенційних лікарських засобів [44]. Напрямок комп'ютерного моделювання лікарських засобів стрімко розвивається, зокрема активно розробляються нові методики та інструменти. Таким чином, протягом останніх кількох років інтеграція великих баз біологічних даних і методів машинного навчання відкрила нові можливості для підвищення ефективності комп'ютерного конструювання ліків із заданою активністю та прогнозованою низькою точністю.

## Висновки

1. Технології комп'ютерного конструювання ліків є важливою складовою сучасної медичної хімії, оскільки дають змогу суттєво скоротити час і затрати для створення нових лікарських засобів.

2. Завдяки прогресу в обчислювальних алгоритмах і накопичених базах даних, інструменти комп'ютерного моделювання інтегровані у кожен етап процесу конструювання ліків.

3. Із розвитком органічного синтезу та високопродуктивного скринінгу, біоінформатики, експериментальної біології, хімії та медицини, великі обсяги накопичених результатів дали змогу розробити численні бази даних хімічних сполук, лікарських засобів та біологічних мішеней, що містять інформацію про структуру речовин, фармакологічну активність сполук, роль протеїнів у патогенезі захворювань та механізми інгібіторної дії лігандів. Використання таких баз даних у процесі *in silico* раціонального конструювання ліків є надзвичайно ефективним з огляду на їх високу структурованість, що дає можливість здійснювати пошук і фільтрування даних за ключовими словами, а також наявність великої кількості перехресних посилань.

## References

1. Shaker B., Ahmad S., Lee J. et al. *In silico* methods and tools for drug discovery // Computers in Biology and Medicine. – 2021. – V. 137 – P. 104851. <https://doi.org/10.1016/j.compbiomed.2021.104851>
2. Paul S. M., Mytelka D. S., Dunwiddie C. T. et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge // Nature Reviews Drug Discovery. – 2009. – V. 9. – P. 203–214. <https://doi.org/10.1038/nrd3078>
3. Giri S., Bader A. A low-cost, high-quality new drug discovery process using patient-derived induced pluripotent stem cells // Drug Discov. Today. – 2015. – V. 20. – P. 37–49. <https://doi.org/10.1016/j.drudis.2014.10.011>

4. Rognan D. The impact of in silico screening in the discovery of novel and safer drug candidates // *Pharmacology & Therapeutics*. – 2017. – V. 175 – P. 47–66. <https://doi.org/10.1016/j.phar>
5. Robinson B. S., Riccardi K. A., Gong Y.-f. et al. BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents // *Antimicrobial Agents and Chemotherapy*. – 2000. – V. 44. – P. 2093–2099. <https://doi.org/10.1128/AAC.44.8.2093-2099.2000>
6. Anderson A. C. The process of structure-based drug design // *Chem. Biol.* – 2003. – V. 10. – P. 787–797. <https://doi.org/10.1016/j.chembiol.2003.09.002>
7. Rutenber E. E., Stroud R. M. Binding of the anticancer drug ZD1694 to *E. coli* thymidylate synthase: assessing specificity and affinity // *Structure*. – 1996. – V. 4. – P. 1317–1324. [https://doi.org/10.1016/s0969-2126\(96\)00139-6](https://doi.org/10.1016/s0969-2126(96)00139-6)
8. Jhoti H., Leach A. R. *Structure-based drug discovery*. – Dordrecht (Netherland): Springer, 2007. – 250 p.
9. Vidal D., Garcia-Serna R., Mestres J. Ligand-based Approaches to *in silico* pharmacology // *Cheminformatics and Computational Chemical Biology*. – 2011. – V. 672. – P. 489–502. [https://doi.org/10.1007/978-1-60761-839-3\\_19](https://doi.org/10.1007/978-1-60761-839-3_19)
10. Talele T. T., Khedkar S. A., Rigby A. C. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic // *Current Topics in Medicinal Chemistry*. – 2010. – V. 10. – P. 127–141. <https://doi.org/10.2174/156802610790232251>
11. Klenina O. V., Tsepelevski H. O., Bannyi I. P. *ta in*. Vyvchennia zalezhnosti diuretychnoi, hemostatychnoi, depyrmuiuchoi, analhetychnoi ta protyzapalnoi aktyvnosti pokhidnykh  $\varepsilon$ -karboksypentylamidiv *R*-benzolsulfoniloksamnoovykh kyslot, a takozh ikh toksychnosti vid kvantovokhimichnykh parametriv budovy ikh molekul // *Farmats. zhurn.* – 2011. – № 1 – S. 54–63.
12. Yang S.-Y. Pharmacophore modeling and applications in drug discovery: Challenges and recent advances // *Drug Discovery Today*. – 2010. – V. 15. – P. 444–450. <https://doi.org/10.1016/j.drudis.2010.03.013>
13. Vucicevic J., Nikolic K., Mitchell J. B. Rational drug design of antineoplastic agents using 3D-QSAR, cheminformatic, and virtual screening approaches // *Current Medicinal Chemistry*. – 2019. – V. 26. – P. 3874–3889. <https://doi.org/10.2174/0929867324666170712115411>
14. Batool M., Ahmad B., Choi S. A Structure-based drug discovery paradigm // *Inter. J. Mol. Sci.* – 2019. – V. 20, N 11. – P. 2783. <https://doi.org/10.3390/ijms20112783>
15. Nunes A. M. V., de Andrade F. D. C. P., Filgueiras L. A. et al. preADMET analysis and clinical aspects of dogs treated with the Organotellurium compound RF07: A possible control for canine visceral leishmaniasis? // *Environ. Toxicol. Pharmacol.* – 2020. – V. 80. – P. 103470. <https://doi.org/10.1016/j.etap.2020.103470>
16. Daina A., Michielin O., Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules // *Sci. Reports*. – 2017. – V. 7. – P. 42717. <https://doi.org/10.1038/srep42717>
17. Wishart D. S. Introduction to Cheminformatics // *Current Protocols in Bioinformatics*. – 2007. <https://doi.org/10.1002/0471250953.bi1401s18>
18. Begam B. F., Kumar J. S. A study on cheminformatics and its applications on modern drug discovery // *Procedia engineering*. – 2012. – V. 38. – P. 1264–1275. <https://doi.org/10.1016/j.proeng.2012.06.156>
19. Manzoni C., Kia D. A., Vandrovicova J. et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences // *Briefings in bioinformatics*. – 2018. – V. 19, N 2 – P. 286–302. <https://doi.org/10.1093/bib/bbw114>
20. Lesk A. M. «Bioinformatics». *Encyclopedia Britannica*, 12 Sep. 2023. – URL: <https://www.britannica.com/science/bioinformatics>.
21. Xia X. Bioinformatics and drug discovery // *Current Topics in Medicinal Chemistry*. – 2017. – V. 17, N 15. – P. 1709–1726. <https://doi.org/10.2174/156802661766616116143440>
22. Wishart D. S., Knox C., Guo A. C. et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets // *Nucleic Acids Research*. – 2008. – V. 36. – P. D901–D906. <https://doi.org/10.1093/nar/gkm958>
23. Irwin J. J., Tang K. G., Young J. et al. ZINC20 – a free ultralarge-scale chemical database for ligand discovery // *J. Chemical Information and Modeling*. – 2020. – V. 60, N 12. – P. 6065–6073. <https://doi.org/10.1021/acs.jcim.0c00675>
24. Pence H. E., Williams A. ChemSpider: an Online Chemical Information Resource // *J. Chem. Education*. – 2010. – V. 87, N 11. – P. 1123–1124. <https://doi.org/10.1021/ed100697w>
25. Kim S., Chen J., Cheng T. et al. PubChem in 2021: new data content and improved web interfaces // *Nucleic Acids Research*. – 2021. – V. 49, N D1. – P. D1388–D1395. <https://doi.org/10.1093/nar/gkaa971>

26. Zdrzil B., Felix E., Hunter F. et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods // *Nucleic Acids Research*. – 2023. – gkad1004. <https://doi.org/10.1093/nar/gkad1004>
27. Wishart D. S., Feunang Y. D., Guo A. C. et al. DrugBank 5.0: a major update to the DrugBank database for 2018 // *Nucleic Acids Research*. – 2018. – V. 46, N D1. – P. D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
28. Sayers E. W., Cavanaugh M., Clark K. et al. GenBank 2023 update // *Nucleic Acids Research*. – 2023. – V. 51, N D1. – P. D141–D144. <https://doi.org/10.1093/nar/gkac1012>
29. Goodsell D. S., Zardecki C., Di Costanzo L. et al. RCSB Protein Data Bank: Enabling biomedical research and drug discovery // *Protein Sci*. – 2020. – V. 29, N 1. – P. 52–65. <https://doi.org/10.1002/pro.3730>
30. UniProt: the universal protein knowledgebase in 2023 // *Nucleic Acids Research*. – 2023. – V. 51, N (D1). – P. D523–D531. <https://doi.org/10.1093/nar/gkac1052>
31. Wang Y., Zhang S., Li F. et al. Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics // *Nucleic Acids Research*. – 2020. – V. 48. – P. DI031–DI041. <https://doi.org/10.1093/nar/gkz981>
32. Hecker N., Ahmed J., von Eichborn J. et al. SuperTarget goes quantitative: update on drug-target interactions // *Nucleic Acids Research*. – 2012. – V. 40. – P. D1113–D1117. <https://doi.org/10.1093/nar/gkr912>
33. Chang A., Jeske L., Ulbrich S. et al. BRENDA, the ELIXIR core data resource in 2021: new developments and updates // *Nucleic Acids Research*. – 2021. – V. 49, N D1. – P. D498–D508. <https://doi.org/10.1093/nar/gkaa1025>
34. Andreeva A., Howorth D., Brenner S. E. et al. SCOP database in 2004: refinements integrate structure and sequence family data // *Nucleic Acids Research*. – 2004. – V. 1, N 32 (Database issue). – P. D226–D229. <https://doi.org/10.1093/nar/gkh039>
35. Gillespie M., Jassal B., Stephan R. et al. The reactome pathway knowledgebase 2022 // *Nucleic Acids Research*. – 2022. – V. 50, N D1. – P. D687–D692. <https://doi.org/10.1093/nar/gkab1028>
36. Raslan M. A., Raslan S. A., Shehata E. M. et al. Advances in the Applications of Bioinformatics and Chemoinformatics // *Pharmaceuticals*. – 2023. – V. 16, N 7. – P. 1050. <https://doi.org/10.3390/ph16071050>
37. Lin X., Li X., Lin X. A review on applications of computational methods in drug screening and design // *Molecules*. – 2020. – V. 25, N 6. – P. 1375. <https://doi.org/10.3390/molecules25061375>
38. Kassab M. M. Development of novel antimicrobial tetracycline analog b (iodocycline) by chemoinformatics // *Ain Shams Med. J.* – 2022. – V. 73. – P. 969–981. <https://doi.org/10.21608/ASMJ.2022.285476>
39. Dawod E. F., Mahmoud N., Elsisi A. Hybrid approach for COVID-19 detection from chest radiography/IJCI // *Inter. J. Computers and Information*. – 2021. – V. 8. – P. 71–76. <https://doi.org/10.21608/IJCI.2021.207754>
40. Shalaby M.-A. W., Dokla E. M., Serya R. A. T. et al. Identification of novel pyrazole and benzimidazole based derivatives as PBP2a inhibitors: Design, synthesis, and biological evaluation // *Archives of Pharmaceutical Sciences Ain Shams University*. – 2019. – V. 3. – P. 228–245. <https://doi.org/10.21608/APS.2019.16625.1010>
41. Soltan M. A., Eldeen M. A., Sajer B. H. et al. Integration of Chemoinformatics and Multi-Omics Analysis Defines ECT2 as a Potential Target for Cancer Drug Therapy // *Biology*. – 2023. – V. 12. – P. 613. <https://doi.org/10.3390/biology12040613>
42. Chambers J., Davies M., Gaulton A. et al. UniChem: a unified chemical structure cross-referencing and identifier tracking system // *J. Cheminformatics*. – 2013. – V. 5, N 1. – P. 3. <https://doi.org/10.1186/1758-2946-5-3>
43. Wang W., Kim M. T., Sedykh A. et al. Developing enhanced blood-brain barrier permeability models: integrating external bio-assay data in QSAR modeling // *Pharmac. Res.* – 2015. – V. 32, N 9. – P. 3055–3065. <https://doi.org/10.1007/s11095-015-1687-1>
44. Schyman P., Liu R., Desai V. et al. vNN web server for ADMET predictions // *Frontiers in Pharmacology*. – 2017. – V. 8 – P. 889. <https://doi.org/10.3389/fphar.2017.00889>

Надійшла до редакції 20 жовтня 2023 р.  
Прийнято до друку 6 грудня 2023 р.

Електронна адреса для листування з авторами: [olena\\_klenina@yahoo.com](mailto:olena_klenina@yahoo.com);  
[chabantaras@ukr.net](mailto:chabantaras@ukr.net)  
(Кленіна О. В., Чабан Т. І.)