

IN SILICO: НОВИЙ НАПРЯМ У РОЗРОБЦІ ФАРМАКОЛОГІЧНИХ ТА ФАРМАЦЕВТИЧНИХ ВЛАСТИВОСТЕЙ ЛІКАРСЬКИХ ЗАСОБІВ

I.C.Чекман, Т.Ю.Небесна, П.В.Симонов

Національний медичний університет ім. О.О.Богомольця

Ключові слова: *in silico*; молекулярний докінг; віртуальний скринінг; фармакофор; QSAR

Стаття присвячена аналізу літератури та результатів досліджень авторів у сфері in silico технологій. В узагальненому вигляді представлена інформація щодо основних методів in silico. Розглянуті основні принципи молекулярного докінгу, методу побудови гомологічних моделей, віртуального скринінгу, моделювання фармакофорів та QSAR. Приділено увагу застосуванню підходів in silico у токсикології. Показано, що методи in silico відіграють ключову роль у створенні нових лікарських засобів та оптимізації існуючих медикаментів і сполук-лідерів.

Прогрес у сфері розвитку обчислювальної техніки триває. Потужність та складність комп'ютерних технологій безупинно зростає, поступово проникаючи у всі сфери нашого життя. Фармакологія та фармація не є виключенням. На теперішній час вже неможливо уявити розробку нових лікарських препаратів без застосування комп'ютерних технологій, які знайшли широке використання на всіх етапах створення і дослідження нових медикаментів, допомагають виявити фармакологічні властивості досліджуваних препаратів [9, 20].

У минулому для позначення різних етапів наукового дослідження у фармакології та фармації існувало три основні поняття: «*in cerebro*», «*in vitro*» та «*in vivo*». Перше характеризувало наукові ідеї щодо розробки препарату, друге – хімічні випробування поза організмом, третє – експеримент на живих тваринах. З розвитком комп'ютерних технологій з'явився четвертий термін – «*in silico*». Цей термін використовується для позначення експерименту, що проводиться за допомогою комп'ютера.

Наразі точно невідомо, хто перший застосував термін «*in*

silico». Декілька дослідників стверджують, що саме їм належить ідея створення цього поняття. Тим не менше, однією з найбільш ранніх робіт, в якій зустрічається цей термін, є «Фізіологічні дослідження *in silico*» [64].

Основні методи *in silico*. Методи *in silico* досліджень використовують підхід застосування комп'ютерних програм для виявлення, аналізу та узагальнення отриманої з різних джерел біологічної та медичної інформації. Отримана інформація використовується у створенні комп'ютерних моделей, що є підґрунтям для прогнозування, створення гіпотез та здійснення відкриттів у фармакології [21]. *In silico* дослідження широко застосовуються для відкриття та оптимізації молекул з високою афінністю до конкретних мішеней, виявлення параметрів їх абсорбції, розподілу, метаболізму, екскреції та токсичності (ADME-Tox), а також фізико-хімічних характеристик [20].

Комп'ютерні системи моделювання, задачею яких є пошук фармакофорів, сполук-лідерів та їх біологічних мішеней, об'єднуються під назвою «*in silico* фармакологія» [45].

Загалом існує два компоненти взаємодії біологічно активних речовин та систем організму (рис. 1), які можуть бути спрощено охарактеризовані «дією сполуки на біосистему» та «дією біосистеми на сполуку». Ліганд, що діє на мішень, може викликати фармакологічну чи токсичну відповідь. Це фармакодинамічний компонент взаємодії. У свою чергу, біосистема впливає на ксенобіотик, абсорбуючи, розподіляючи, метаболізуючи та екскретуючи його (ADME-параметри). Це фармакокінетичний компонент. Дуже важливо враховувати, що ці два компоненти взаємопов'язані. ADME-параметри впливають на тривалість і вираженість фармакодинамічних ефектів ксенобіотиків. І, навпаки, медикамент може впливати на гемодинаміку чи активність ферментів організму, змінюючи таким чином фармакокінетичні показники. Тільки системний підхід до врахування обох взаємопов'язаних компонентів взаємодії дає можливість виявити найбільш оптимальні сполуки для синтезу нових медикаментів. Методи *in silico* покликані враховувати як фармакодинамічні, так і фармакокінетичні аспекти взаємодії лігандів з мішенями [21].

Створення нового лікарського препарату – складний, тривалий у часі процес. На тепе-

I.C.Чекман – доктор мед. наук, професор, член-кореспондент НАН та НАМН України, завідувач кафедри фармакології та клінічної фармакології Національного медичного університету ім. О.О.Богомольця (м. Київ)

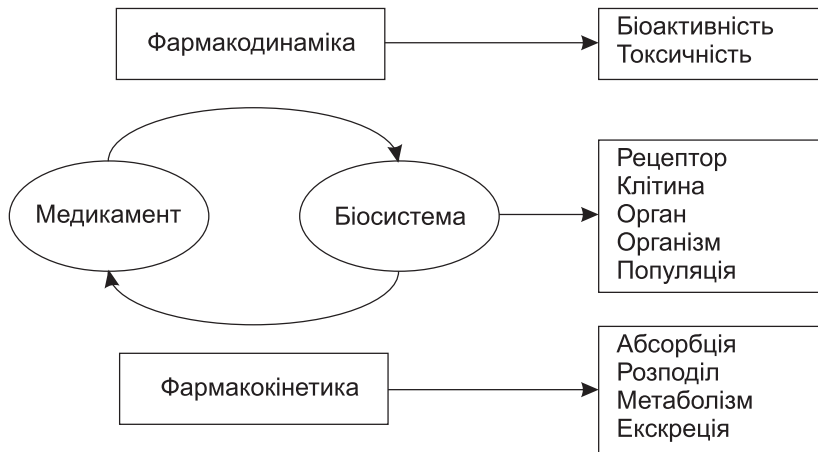


Рис. 1. Модель взаємодії «медикамент/біологічна система» [21], з доповненнями]

рішній час розробка лікарського засобу об'єднує експериментальні методи визначення структури речовин з теоретичними методиками CAMD (computer-aided molecular design – автоматизоване молекулярне проектування). Цей термін включає велику кількість розрахункових методів хімії. Вони поділяються на 2 групи в залежності від предмету досліджень. В основі досліджень першої лежить ліганд, другої – білкова структура клітини-мішені.

Взаємодія ліганду з мішенню досліджується за допомогою двох різних груп методів:

ліганд-орієнтованої та мішень-орієнтованої. Мішень-орієнтовані методи полягають у безпосередній побудові тривимірної структури білка-мішені. Ліганд-орієнтовані методи – це аналіз властивостей набору лігандів та розробка моделей взаємодії ліганду з мішенню без попередніх знань про структуру білка-мішені [9, 22, 75].

Вибір конкретного методу *in silico* досліджень залежить від обсягу вже відомої інформації про мішень чи ліганд. Процес вибору може бути зображений у вигляді алгоритму (рис. 2).

Мішень-орієнтовані методи в *in silico* дослідженнях. Мішень-орієнтовані методи CAMD засновані на використанні попередньо отриманої інформації щодо молекулярної структури білків. Ці методи, з одного боку, надають вичерпну інформацію щодо очікуваної конформації та орієнтації ліганду в активному центрі білка (докінг) та, з іншого боку, раціональну оцінку його афінності (оціночна функція) [21].

Основним мішень-орієнтованим методом є молекулярний докінг (рис. 3). В його основі лежить підхід співставлення малих молекул із активним центром рецептора з метою виявлення уявного ліганду з найбільшою афінністю – кандидата для створення нового лікарського засобу [12].

З цією метою використовуються багато програм, серед яких AutoDock, FlexX, GOLD, DOCK [10], Surflex-Dock, FRED [38], DynaDock [5]. За допомогою молекулярного докінгу були відкриті нові інгібітори циклін-залежних кіназ, кіназ рецепторів епідермального та судинного ендотеліального факторів росту [48], інгібітори ВІЛ протеаз (рис. 4)

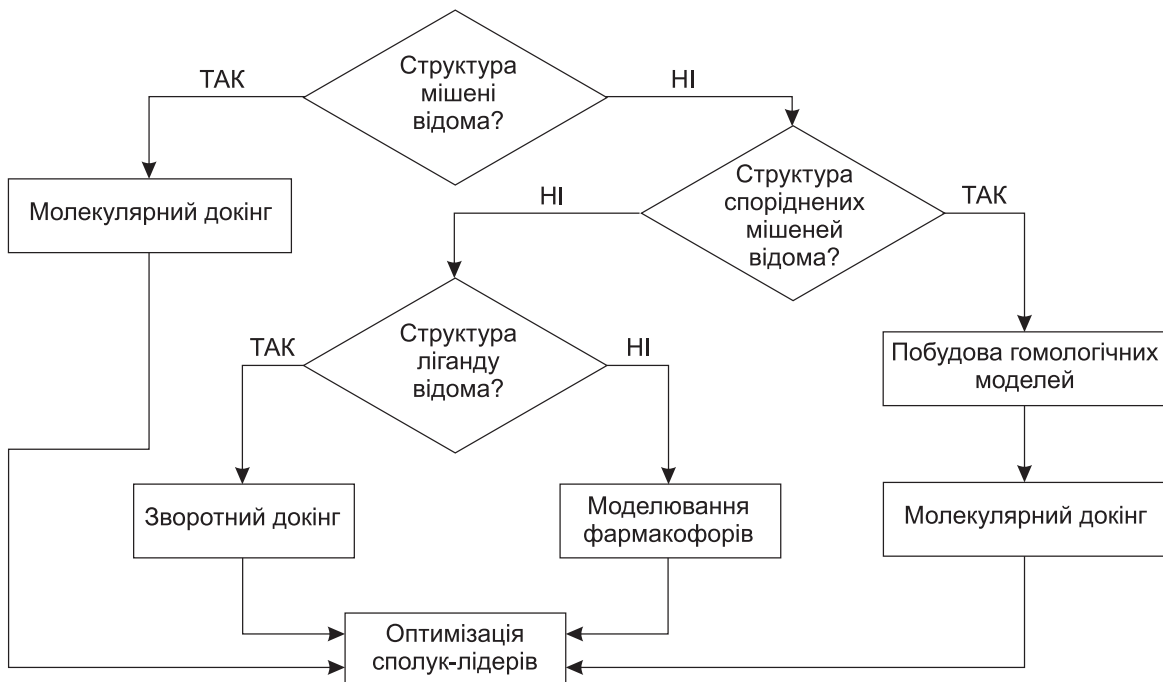


Рис. 2. Загальний алгоритм *in silico* досліджень

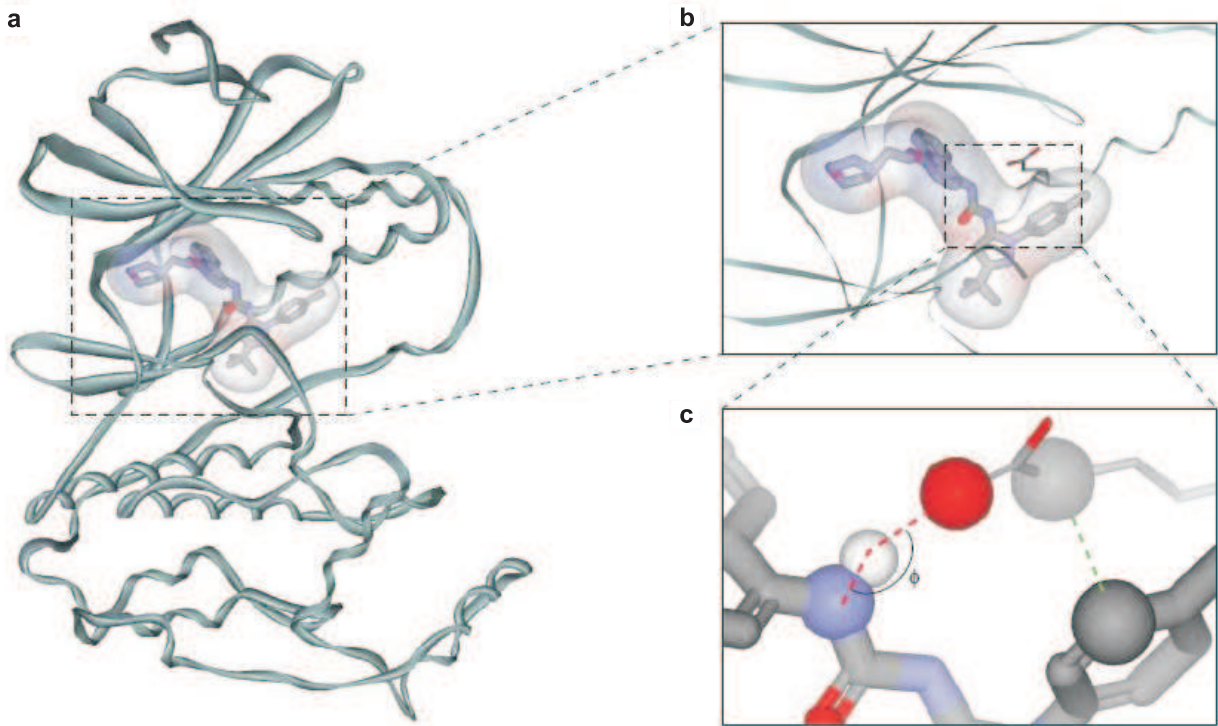


Рис. 3. Модель молекулярного докінгу: а – структура протеїнкінази р38 у сполученні з інгібітором BIRB796; б – збільшене зображення активного центру; с – взаємодія групи Glu71 з BIRB796 [33]

[33], моноацилгліцерол ліпази [74], стероїдної ароматази [16].

У залежності від принципів, за якими відбувається підстановка ліганду в активний центр білка, розрізняють систематичні, стохастичні та симуляційні методи молекулярного докінгу. В основі систематичного методу лежить пошук лігандів у електронних базах даних з врахуванням всіх їх можливих конформацій, що вимагає значних витрат часу. У якості альтернативи використовується стратегія *de novo* дизайну, яка полягає у докінгу окремих фрагментів з подальшим їх об'єднанням у молекулу-ліганд. Такий підхід використовують програми Groupbuild, GenStar, Grow та Growmol. Також використовується метод визначення стійких (основний фрагмент) та гнучких (бокові ланцюги) частин молекули. Стійкі частини під час докінгу залишаються незмінними, а гнучкі компоненти змінюють кут обертання. Варіанти лігандів з найменшою афінністю поступово відкида-

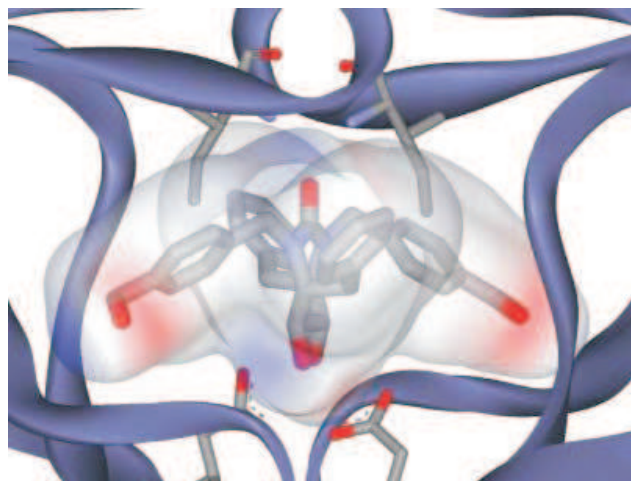


Рис. 4. Модель зв'язування інгібітора Dmp323 з BIA протеазою: зображений електростатичний потенціал інгібітора; групи Ile50 та Asp25 стабілізують зв'язок [33]

ються у процесі зміни бокових ланцюгів, що зменшує загальну складність обчислень.

Стохастичні методи молекулярного докінгу засновані на створенні випадкових змін у структурі ліганду та оцінці їх впливу на афінність. Один з найбільш популярних методів даної групи – Monte Carlo. Даний підхід реалізується через декілька етапів. Спочатку створю-

ється початкова конфігурація ліганду в активному центрі та проводиться його оцінка. Потім створюється новий ліганд, який також оцінюється. Для визначення того, чи підходить даний ліганд, використовується Метрополіс-критерій. Сутність критерію: якщо наступний ліганд має кращу оцінку, він одразу приймається, а якщо гіршу – застосовується функція імо-

вірності Больцмана. Етапи пошуку повторюються до отримання необхідної конфігурації [33].

До стохастичних належить також метод генетичних алгоритмів. Параметри моделі ліганду кодуються у «хромосомі». Випадковим чином створюється «популяція», тобто сукупність різних «хромосом», які оцінюються. «Хромосоми» лігандів, що отримали найкращу оцінку, піддаються «генетичним операціям» для формування нових «хромосом». Головними операціями є «кросингвер» – з'єднання частин «хромосом» попередньої генерації для створення нових «хромосом» та «мутація» – додавання у «хромосому» випадкової інформації. Процес триває до отримання необхідного набору параметрів ліганду [33, 72].

Великої популярності також набув стохастичний метод табу-пошуку. Процедура починається з побудови випадкового початкового рішення (структури ліганду) та його оцінки. Після цього створюються та оцінюються можливі «сусіди» – рішення, що можуть бути отримані з початкового шляхом впровадження елементарних змін. Нове рішення обирається з «сусідів» попереднього. Таким чином, завершується один крок алгоритму. Для запобігання повторів метод записує останні кроки у табу-список. Максимальна кількість кроків у табу-списку задається користувачем. Коли список повністю заповнений, кроки виключаються з нього за принципом «first-in, first-out» [69]. Табу-список зберігає крок з попереднім рішенням та забороняє пошуковій системі повторювати кроки, що приводили до цього рішення. Якщо кращий з «сусідів» відсутній у табу-списку, він обирається як нове рішення. Нове рішення буде прийнято навіть у тому випадку, коли краще рішення «сусід» гірше, ніж початкове рішення. Таким чином, метод запобігає зупинці алгоритму на рішенні, що

є локальним мінімумом. Як тільки «сусід» був обраний як нове рішення, він додається у табу-список. Якщо крок, що знаходиться у табу-списку, веде до нового оптимального рішення, застосовується критерій прагнення (aspiration criterion), який дозволяє знову виконати даний крок. Якщо отримане на певному етапі рішення задовольняє бажаним вимогам або кількість кроків досягає визначеного користувачем ліміту, процедура зупиняється [8, 62].

В основі симуляційних методів лежать закони молекулярної динаміки – комп'ютерних методик дослідження взаємодії атомів та молекул за певний період часу та створення картини руху частинок. Молекулярна динаміка є своєрідним «віртуальним мікроскопом» з надвисокою роздільною здатністю. Цей мультидисциплінарний метод дослідження взаємодії сполук поєднує закони математики, фізики і хімії з алгоритмами інформатики. Симуляція взаємодії атомів є надтривалим у часі процесом та вимагає великої потужності комп'ютерного обладнання. Для постановки експерименту необхідно вибрати кількість частинок, крок та час взаємодії таким чином, щоб процес створення симуляції закінчився якомога скоріше. Разом з тим симуляція повинна бути достатньо тривалою, щоб відповісти кінетиці досліджуваного процесу. Частковим вирішенням даної проблеми є створення паралельних алгоритмів, що дозволяють розподілити навантаження між комп'ютерами. Головною задачею процесора під час побудови симуляції є оцінка потенціалу (силового поля) як функції координат частинок. Інший фактор, що збільшує загальний час, необхідний для створення симуляції, – це розмір кроку, а саме проміжок часу між двома оцінками потенціалу. Крок повинен бути до-

статньо малим, щоб запобігти виникненню помилок дискретності, тобто меншим, ніж найбільш швидка частота вібрації у системі. Величина кроку зазвичай становить близько однієї фемтосекунди ($1 \cdot 10^{-15}$ с). Ця величина може бути збільшена завдяки алгоритму обмеження SHAKE, який стабілізує вібрацію атомів водню, та його модифікаціям. Найбільш популярними пакетами програмного забезпечення для моделювання динаміки біомолекул є AMBER, CHARMM та GROMOS [3, 31, 50].

Для оцінки афінності ліганду використовують 3 класи оціночних функцій: засновані на силовому полі, емпіричні та засновані на знаннях. Силове поле – функція, що виражає енергію системи як суму різних показників молекулярної механіки. Найчастіше це енергія взаємодії «протеїн-ліганд» та внутрішня енергія ліганду. Різні оціночні функції базуються на різних наборах параметрів силового поля. Наприклад, у програмі AutoDock використовується набір параметрів AMBER force field [33].

Емпіричні оціночні функції (empirical scoring functions) та засновані на знаннях оціночні функції (knowledge-based scoring functions) використовують експериментальні дані. Емпіричні оціночні функції базуються на моделі, де афінність виражена у якості суми усіх можливих складових взаємодії, таких як сили Ван-дер-Ваальса, водневі зв'язки, гідрофобний ефект тощо. Для побудови нової емпіричної оціночної функції необхідно використання тренувального набору кристалографічних структур бінарних комплексів «протеїн-ліганд», для яких експериментально обчислене значення афінності. Наступний крок – застосування регресійного аналізу для пошуку найкращого збігу між передбаченою та отриманою в експерименті афінністю та окрема оцінка кожної

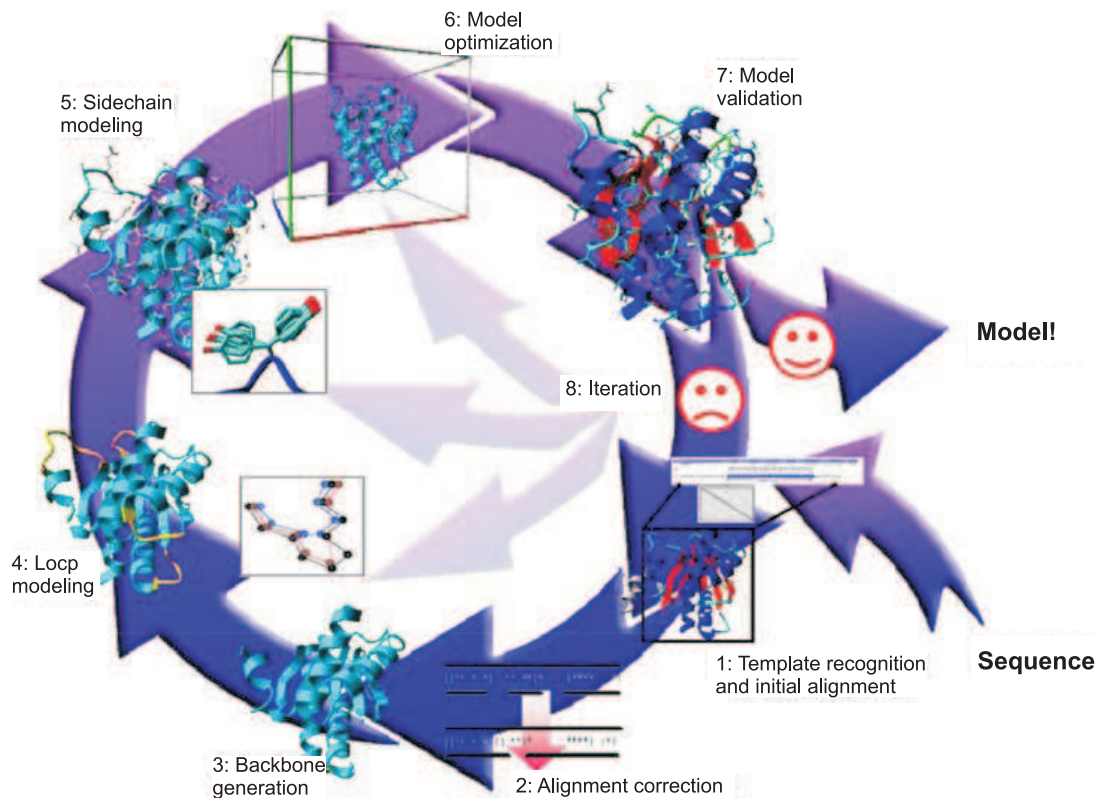


Рис. 5. Стадії побудови гомологічної моделі [71]

складової взаємодії [23, 66]. Емпіричні оціночні функції представлені у програмах SFCscore, ChemScore, Lig-Score, Validate [66] та X-SCORE [69].

Засновані на знаннях оціночні функції використовують статистичну обробку експериментально отриманих комплексів «протеїн-ліганд» для створення правил щодо геометрії взаємодії. Ці правила є «псевдопотенціалами», які використовуються для оцінки афінності лігандів у генерованих комп'ютером моделях. Такий статистичний підхід дає можливість врахувати фізичні ефекти, що на теперішній час ще не є цілком зрозумілими з теоретичної точки зору (сольватація, поляризація), а тому не можуть бути враховані при створенні оціночної функції силового поля. Відомою оціночною функцією є PMFScore, що базується на енергії Гельмгольца, отриманої з 697 кристалографічних комплексів «протеїн-ліганд» [26]. У 2000 р. німецькими вченими була розроблена оціночна функція DrugScore,

яка використовувала інформацію щодо взаємодії 1376 комплексів «протеїн-ліганд» з бази даних Relibase [25].

На теперішній час існують також «консенсус» оціночні функції, які об'єднують підходи всіх трьох перелічених вище класів [33, 63].

На початку досліджень у сфері докінгу молекулярна структура мішеней визначалася експериментально за допомогою методів рентгеноструктурного аналізу або ЯМР. Молекулярний докінг залишався досить складним процесом, незважаючи на більш ніж 25 років досліджень у даній сфері. Це пояснювалося складністю отримання максимально точної інформації щодо будови білка-мішені експериментальним шляхом [33, 39]. На жаль, через ці технічні складності навіть тепер відомі далеко не всі 3D структури терапевтично значущих родин протеїнів [21].

В останні роки велику увагу привернув докінг-метод визначення структури мішеней,

експериментальне отримання яких не є можливим – метод побудови гомологічних моделей (homology modelling techniques). Його сутність полягає у генеруванні структури мішені за допомогою комп'ютерного моделювання на основі шаблонної структури подібних мішеней із використанням інформації щодо лігандів як обмеження. Серед популярних засобів побудови гомологічних моделей – програма MODBASE [54], веб-сервери Robetta [14], CHOYCE [56] та SWISS-MODEL [6]. Прикладом застосування цього методу є відкриття структури декількох спряжених з G-білком рецепторів (GPCR) [7, 21].

Процес побудови гомологічної моделі включає 8 стадій (рис. 5):

1. Створення шаблону та початкове вирівнювання між шаблоном та модельною послідовністю (пошук у базі даних протеїнів Protein Data Bank за допомогою програми PSI-BLAST) [4].

2. Оцінка вирівнювання, заснована на використанні моле-

кулярних класових інформаційних систем – Molecular class-specific information systems (MCSISs), які містять багато додаткової інформації щодо конкретного класу білків; прикладом такої системи є GPCRDDB, що містить дані щодо послідовностей, структури, мутацій та взаємодії з лігандами родини спряжених з G-білком рецепторів [29].

3. Створення основи моделі.

4. Заповнення проміжків та створення петель; обирається петля з мінімальною вільною енергією [73].

5. Додавання бокових ланцюгів із використанням електронних бібліотек ротамерів, отриманих за допомогою методів рентгеноструктурного аналізу [13, 18, 44].

6. Молекулярно-динамічна симуляція моделі для виявлення помилок (ця стадія може бути вибірково включена до алгоритму).

7. Валідація моделі за допомогою спеціального програмного забезпечення WHAT_CHECK [28].

8. Отримання готової моделі чи повтор процесу моделювання із заміною шаблону, початкового вирівнювання чи повторним проходженням окремих стадій [71].

В останні роки увагу вчених привернули дослідження із зворотного докінгу, в якому здійснюється пошук у Protein Data Bank мішеней для вже відомих лігандів. Докінг заснований на процедурі, що включає множинно-конформаційне співставлення молекул з активними центрами білків. Оцінка збігу засновується на врахуванні енергії міжмолекулярного зв'язку та, якщо можливо, на визначенні інших лігандів, що зв'язуються з цим рецептором. За допомогою зворотного докінгу були визначені мішені 4Н-тамоксифену та токоферолу. Застосування цього методу дозволить визначити невідомі білки-мішені, зв'язок з якими є причиною виникнення фарма-

кологічних ефектів медикаментів [12].

Відома програма INVDOCK застосовує зворотній докінг для пошуку білків-мішеней та передбачення токсичності речовин. INVDOCK використовує базу даних з 147 протеїнів, які беруть участь у виникненні побічних ефектів (DART). За допомогою цієї програми були визначені 11 антиретровірусних лікарських засобів з високою активністю проти β -ДНК-полімерази та ДНК-топоізомерази I [30].

У 2006 р. створений веб-сервер зворотного докінгу TarFisDock, що дозволяє проводити скринінг бази даних Potential Drug Target Database, який містить 698 протеїнів [41]. Процедура пошуку білків-мішеней: введення інформації щодо біологічно активної сполуки чи існуючого медикаменту, зворотній докінг із використанням баз даних протеїнів, аналіз гомологів отриманих білків, подальша оцінка біологічної активності.

Ліганд-орієнтовані методи в *in silico* дослідженнях. На теперішній час далеко не всі 3D структури терапевтично значущих родин протеїнів відомі, тому широко використовуються ліганд-орієнтовані методи – аналіз властивостей набору лігандів та розробка моделей взаємодії ліганду з мішенню без попередніх знань про структуру білка-мішені.

До ліганд-орієнтованих методів належить віртуальний скринінг, побудова фармакофорів та метод кількісного співвідношення структура-активність (QSAR). Останній знайшов широке використання також у оптимізації сполук-лідерів [2]. Віртуальний скринінг полягає у пошуку необхідного ліганду серед великої кількості відомих сполук у молекулярних базах даних [22, 49, 75]. Інструментом скринінгу баз даних є параметри, що описують власти-

вості необхідного ліганду – фармакофорні дескриптори. Мета пошуку – відбір молекул з потенційно високою афінністю до мішені. Сполуки, що підходять найбільше (сполуки-лідери), замовляються у підприємства, яке надає базу даних, та проходять подальші випробування та оптимізацію *in vitro*. Таким чином, дорогого та тривалого процесу хімічного синтезу можна уникнути, обравши декілька комерційно доступних сполук. Як результат, віртуальний скринінг порівняно з методом хімічного синтезу є більш ефективним та знайшов широке використання [9, 21].

Серед найбільш відомих баз даних віртуального скринінгу – WOMBAT (Sunset Molecular Discovery LLC), що містить 120 400 молекул, про які згадувалося у журналах з медичної хімії протягом останніх 30 років. База даних MDL Drug Data Report чи MDDR (MDL Information Systems) містить інформацію, зібрану з наукової літератури, журналів та конгресів, про терапевтичну дію та біологічну активність більш ніж 132 000 сполук. База даних AurSCOPE (Aureus Pharma) надає колекцію хімічних бібліотек з 320 000 молекул. І нарешті, бази даних MedChem та Target Inhibitor (GVK Biosciences) зібрали більше 2 000 000 молекул та інформацію з 20 000 наукових публікацій щодо їх активності та токсичності по відношенню до відповідних родин протеїнів [21].

На велику увагу заслуговує метод моделювання фармакофорів. Згідно з визначенням ІЮПАК (1997) фармакофор – це сукупність просторових та електронних властивостей, необхідних для забезпечення оптимальних молекулярних взаємодій зі структурою специфічної біологічної мішені, що викликають її біологічний відгук. Фармакофор не є реальною молекулою або реальною сукупністю функціональних груп, а вказує на загальні

молекулярні властивості груп сполук стосовно структури мішені. Фармакофор також може розглядатися як велика сукупність наборів активних молекул [27].

Фармакофор містить інформацію щодо структури, фізико-хімічних та квантово-фармакологічних властивостей, які, в свою чергу, визначають фармакологічну активність речовини, особливості впливу на органи і системи організму, обмін речовин тощо. Одного разу знайдений фармакофор може виступати потужною моделлю у різноманітних методах раціонального дизайну ліків, таких як віртуальний скринінг та *de novo* дизайн [59, 60]. Розроблені фармакофори перспективних лікарських засобів для лікування лейшманіозів [55], хвороби Альцгеймера [11], малярії [24, 51], діабету другого типу [42, 67], онкозахворювань [43] та багатьох інших патологій.

У 2002 р. американськими вченими розроблена модель фармакофору для лігандів, що зв'язуються з прегнан-Х-рецептором (PXR) – протеїном множинної лікарської стійкості (MDR), який регулює метаболізм та виведення солей жовчних кислот, стероїдів та ксенобіотиків [35, 47, 57, 65]. Ця модель може використовуватися як фільтр для ідентифікації сполук, що зв'язуються з PXR, до їх випробування *in vitro*. Це допоможе у виявленні молекул з меншою афінністю до PXR, а отже – у запобіганні індукції великої кількості метаболізуючих ліки ферментів [19].

Модель фармакофору може бути отримана за допомогою різних програм, найбільш відомими серед яких є ALADDIN, DISCO, GERM, COMPASS, GASP, Catalyst HIPHOP, SCAMPI, Catalyst Hypogen, CLEW GAMMA, PARM та DANTE [9]. Багато вчених світу використовує веб-сервер для створення фармакофорів PharmaGist. Розглянемо алгоритм по-

шуку фармакофорів, представлений у цій системі. Вихідними даними є тривимірні моделі лігандів, факт взаємодії яких з мішенню науково підтверджений, та фізико-хімічні властивості цих лігандів. Процес пошуку складається з декількох етапів.

1. Кожен з лігандів розглядається системою окремо. Визначаються зв'язки ліганду, що обертаються. За цими зв'язками ліганди об'єднуються у групи. Один з лігандів набуває статусу «штифт» (pivot). Це може бути попередньо задана користувачем чи випадково обрана системою сполука. Якщо «штифт» обраний алгоритмом випадково, то цей статус послідовно надається кожному ліганду. У якості попередньо заданого користувачем «штифту» може виступати сполука з найвищою афінністю до рецептора або ліганд з найменшою кількістю зв'язків, що обертаються у досліджуваній структурі.

2. «Штифт» (стала величина) попарно порівнюється з кожним лігандом (перемінна величина). Кожній парі сполук надається оцінка, яка є сумою властивостей, що співпали. Результат етапу – велика кількість пар сполук, що мають високий ступінь збігу.

3. Парні групи об'єднуються у множинні групи. Завдання етапу – знайти значимі властивості «штифту», що співпадають з властивостями більшості парних груп різних лігандів.

4. Фармакофори-кандидати, отримані з різних «штифтів», об'єднуються. Ті, що отримали найвищу оцінку, виводяться програмою як результат. Отримані фармакофори можуть бути використані у процесі дизайну ліків для подальшого створення нових сполук-лідерів [59, 60].

Оптимізація сполук-лідерів. Сполуки-лідери підлягають подальшій оптимізації, що має на меті максимізацію бажаних та мінімізацію небажаних по-

бічних ефектів. Найбільш відомим методом оптимізації є QSAR.

Виходячи з монументального правила хемогеноміки, згідно з яким сполуки, що проявляють подібну специфічну дію на рецептор, мають бути подібні за структурою [17, 39], можна дійти висновку, що існує залежність між структурою речовин та їх активністю. Ця залежність може бути виражена рівнянням. Пошук кількісних співвідношень структура-активність та процес побудови відповідних моделей-рівнянь отримав назву QSAR (Quantitative Structure-Activity Relationship).

QSAR використовується у якості інструменту передбачення та оцінки фармакокінетичних та фармакодинамічних властивостей хімічних речовин та заснований на моделі тренувального набору сполук з відомою активністю та визначеною структурою. Традиційно QSAR може бути представлений і без комп'ютера. Проблема відношення структура-активність була висвітлена ще у 1899 р. німецьким ученим Гансом Мейєром, який запропонував «ліпідну теорію інгібування клітин», у якій встановлювався зв'язок між величиною коефіцієнту розподілу розчинності речовин та їх дією. Більш стрімкий розвиток QSAR відбувся у 60-ті роки ХХ ст., коли почали впроваджуватися комп'ютерні технології для побудови математичних моделей структура-активність [45].

Концепція QSAR побудована на переведенні хімічної структури у числову форму за допомогою молекулярних дескрипторів. На теперішній час існує безліч дескрипторів, що описують різні властивості речовин: електронні, стеричні, квантово-механічні, конформаційні та термодинамічні. Необхідність дослідження нових властивостей сполук приводить до створення нових молекулярних дескрипторів, які об'єднуються у бази да-

них. Прикладом такої бази даних є MOLE db, що містить інформацію стосовно 234 773 молекул та 1124 дескрипторів.

Задача QSAR полягає у порівнянні великої кількості дескрипторів та визначенні, які з них корелюють з досліджуваними параметрами фармакодинаміки чи фармакокінетики. Встановлення такого зв'язку здійснюється за допомогою побудови регресійного рівняння. Чим більша кореляція, тим більше дескриптор впливає на параметр. Для QSAR розрахунків використовують спеціальні програми, наприклад, E-DRAGON. Ця програма містить інформацію стосовно 1600 дескрипторів, які застосовуються для оцінки відношень структура-активність.

Моделі, побудовані на основі одночасно лінійних та топографічних дескрипторів, називаються 2D-QSAR та широко використовуються для передбачення фізико-хімічних властивостей речовин та надання кількісних оцінок біологічних ефектів хімічних сполук [17].

На перевагу попереднім, 3D-QSAR моделі враховують просторову тривимірну структуру молекул. Характерною рисою 3D дескрипторів є їх залежність від молекулярної конформації. Тому багато 3D-QSAR методів вимагає «вирівнювати» молекули перед побудовою моделей. Існує декілька способів «вирівнювання» молекул, серед яких найбільш широко застосовується метод проєкції атомів на уявну площину (одна зі складових методу CoMFA) [15, 40] та докінг сполук у активний центр мішені, що застосовується у випадках, коли відома структура цього центру [52].

Конформаційні обмеження розглянутих методів мотивували появу альтернативного 3D-QSAR підходу, що базується на статистичному аналізі просторового розташування молекул GRIND без проєкції дескрипто-

рів на площину-сітку (GRid-INdependent descriptors). Цей метод представлений у програмі ALMOND [53].

Існує декілька процедур підрахунку оптимальних результатів QSAR. Найбільш часто використовуються методи внутрішньої перехресної оцінки *leave-one-out* (q2 LOO) та *leave-10%-out* (q2_10). У методі q2_10 застосовується процедура 10% x 10%, згідно з якою сполуки, що входять до масиву даних, у випадковому порядку розділяються на 10 рівних за кількістю частин (кожна по 10%). При кожному повторі процесу оцінки одна з цих частин видаляється з масиву даних, а нова модель будується з інших 90%. Видалені 10% є масивом порівняння для цієї нової моделі. Потім видаляються інші 10%, і процес пошуку повторюється 10 разів, поки кожен 10 сполук не будуть використані у якості масиву порівняння. Метод q2 LOO подібний до q2_10 за процедурою порівняння, але відрізняється тим, що за один цикл замість десяти сполук з масиву даних видаляється тільки одна. Цей метод дає більш точні результати, але є більш тривалим та вимагає застосування комп'ютерного обладнання більшої потужності.

Існує декілька проблем, пов'язаних із застосуванням QSAR моделей. Першою є обов'язковість забезпечення її достовірними вихідними даними та вибір доцільного набору дескрипторів. Якщо модель буде містити невірні дані, то результати також будуть носити помилковий характер. Передбачення QSAR є більш точними, якщо для побудови моделей використовуються більш широкий обсяг інформації стосовно біологічних та хімічних властивостей досліджуваних сполук. Другою проблемою є складність застосування методів QSAR для органометалічних комплексів, рослинних екстрактів та високомолекулярних сполук-полімерів через від-

сутність достатньо повного уявлення щодо структури та властивостей цих речовин. Інша проблема полягає у поліфармакології. Медикаменти можуть одночасно діяти за різними механізмами та впливати на різні мішені. Враховуючи дані обмеження, можна зробити висновок, що QSAR дослідження повинні проводитися не ізольовано, а у поєднанні з іншими джерелами інформації та методами дослідження [45].

Методи QSAR зазвичай не враховують шляхи подальшого метаболізму сполук в організмі та їх метаболічну активацію. Тому перед побудовою моделі необхідно дослідити накопичені наукові дані щодо метаболізму сполук чи спрогнозувати можливу структуру метаболітів, а потім досліджувати ці метаболіти за допомогою QSAR [45]. Однак існують програми, що використовують бази даних *in vitro* та *in vivo* досліджень для передбачення метаболізму речовин в організмі людини: MetabolExpert, META та METEOR [34, 37, 68].

До відомих та вдалих програм оптимізації сполук-лідерів слід віднести програму PASS, яка прогнозує 2468 видів біологічної активності, включаючи фармакологічні ефекти, мутагенність, канцерогенність, тератогенність та ембріотоксичність. Точність прогнозу в 85% достатня для практичного використання системи PASS з метою прогнозу спектра біологічної активності нових речовин. Результати прогнозу виводяться програмою у вигляді списку назв можливих видів активності з розрахованими оцінками вірогідності наявності (Pa) та відсутності (Pi) кожного виду активності, які мають значення від 0 до 1. Чим більше для конкретної активності величина Pa і чим менше величина Pi, тим більше шанс підтвердити дану активність в експерименті [1, 36].

Важливим методом у виборі найкращих сполук-лідерів, зокрема при розробці медикаментів, що приймаються *per os*, є фільтрація отриманих у результаті *in silico* досліджень сполук за молекулярними та фізико-хімічними властивостями, що забезпечують належне ADME (абсорбція, розподілення, метаболізм та екскреція) [22]. Цей метод реалізований через «правило п'яти» американського хіміка Ліпінського – набір критеріїв для передбачення оральної біодоступності сполуки на основі простих молекулярних ознак. Згідно з цим правилом ліки, що приймаються орально, мають відповідати щонайменше трьом з наведених нижче критеріїв:

- не більше 5 донорів водневих зв'язків;
- не більше 10 акцепторів водневих зв'язків;
- молекулярна маса менше 500 дальтон;
- логарифм константи розподілу октанол/вода $\log P$ менше 5 ($\log P$ є мірою розчинності ліків у воді та гідрофобних середовищах, прикладом яких є клітинні мембрани) [9].

Використання методів *in silico* у токсикології. Останнім часом великий інтерес викликає використання технологій *in silico* для доклінічного передбачення токсикологічних ефектів медикаментів. Науковий світ зацікавлений у використанні таких технологій як інструмента, що визначає безпечність препаратів та інших ксенобіотиків. Активні хімічні речовини можуть бути досліджені за допомогою спеціальних програм та баз даних. *In silico* токсикологія використовує скринінг токсикологічних баз даних, QSAR моделювання, ліганд- та мішень-орієнтовані методи.

Головна концепція створення токсикологічних баз даних – використання отриманої з токсикологічних досліджень науко-

вої інформації для побудови електронного ресурсу, за допомогою якого можна передбачувати та порівнювати токсичні ефекти сполук [45].

Зараз створені бази даних для вивчення канцерогенних властивостей сполук (CCRIS – Chemical carcinogenesis research information system), гепатотоксичності (Acutoxbase [32]), кардіотоксичності (AERS – Adverse Event Reporting System [46]) та алергічних реакцій на ксенобіотики [58]. Розробки спеціалізованих баз даних ведуться у напрямку створення електронних ресурсів для дослідження токсичного впливу медикаментів на дихальну, імунну та центральну нервову системи.

Кроком уперед у квантовій фармакології є впровадження заснованих на людських знаннях методів *in silico* (knowledge-based methods), які часто є більш достовірними та точними, ніж машинні алгоритми, засновані лише на статистичних даних. У таких методах створюються експертні правила, засновані на даних літератури та на власному досвіді вчених. Це варіант симбіозу та поєднання методів *in cerebri* з комп'ютерними QSAR методами. Найбільш відомі програми, що застосовують даний підхід – EPA OncoLogic™ та Lhasa Ltd.'s Derek for Windows (DFW).

У *in silico* ADME-Tox-підходах застосовуються правило Ліпінського та багато інших важливих загальних положень відносно властивостей, які треба враховувати протягом доклінічної розробки ліків, наприклад, проникність гематоенцефалічного бар'єру для нейротропних лікарських засобів. Таким чином, використання ADME-Tox програм для аналізу індивідуальних сполук та класів сполук, знайдених за допомогою віртуального скринінгу токсикологічних баз даних, може бути надзвичайно інформативним. ADME-Tox підходи вирішують

фармакокінетичні та токсикологічні питання стосовно участі ліків у організмі [45].

Практичний приклад застосування *in silico* у токсикології представлений англійським вченим M. Segall та колегами. Вони провели скринінг 3000 сполук з віртуальної бібліотеки, що нараховувала більше 13000 речовин. Ці сполуки перевірялися на водорозчинність, всмоктуваність у кишечнику, $\log P$, зв'язування з білками плазми крові, афінність до цитохрому P₄₅₀2D6 та 2C9, IC₅₀ для калієвих hERG каналів кардіоміоцитів, проникність гематоенцефалічного бар'єру. В результаті молекули, відповідно до ADME показників, були класифіковані на 13 серій хемотипів. Надалі було синтезовано 185 сполук із передбаченими ADME властивостями. Результати випробувань *in vitro* підтвердили вірність передбачень *in silico* [61].

ВИСНОВКИ

Створення та удосконалення технологій *in silico* – новий крок у розробці та оптимізації лікарських засобів, дослідженні фармакокінетичних і фармакодинамічних параметрів сполук-лідерів та медикаментів. Методи *in silico* включають молекулярний докінг, метод побудови гомологічних моделей, зворотній докінг, віртуальний скринінг, моделювання фармакофоров, QSAR та використання електронних баз даних хімічних сполук. Ці методи знайшли широке застосування у відкритті та оптимізації нових сполук з високою афінністю щодо біологічних мішеней, визначенні параметрів їх абсорбції, розподілу, метаболізму та екскреції, а також інших фізико-хімічних характеристик.

Дослідження *in silico*, на відміну від експериментальних досліджень, потребують менше часу та коштів, мають високий рівень відтворюваності, не потребують хімічного синтезу великої кількості сполук, змен-

шують необхідність проведення дослідів на тваринах та можуть проходити постійну оптимізацію.

In silico фармакологія безупинно продовжує свій розвиток та створює багато інструментів, які допомагають

ученим у відкритті нових мішеней та сполук-лідерів з передбаченою біологічною дією на них.

ЛІТЕРАТУРА

1. Бобкова Л.С., Чекман І.С., Яворовський О.П. та ін. // *Сучасні проблеми токсикол.* – 2008. – №2. – С. 78-86.
2. Чекман І.С., Небесна Т.Ю., Барабаш О.В. // *Вісник фармакол. та фармацевції.* – 2009. – №11-12. – С. 7-14.
3. Alonso H., Bliznyuk A.A., Gready J.E. // *Med. Res. Rev.* – 2006. – Vol. 26, №5. – P. 531-568.
4. Altschul S.F., Madden T.L., Schaffer A.A. et al. // *Nucl. Acids Res.* – 1997. – Vol. 25, №17. – P. 3389-3402.
5. Antes I. // *Proteins.* – 2010. – Vol. 78, №5. – P. 1084-1104.
6. Arnold K., Bordoli L., Kopp J., Schwede T. // *Bioinformatics.* – 2006. – Vol. 22, №2. – P. 195-201.
7. Bissantz C., Bernard P., Hibert M., Rognan D. // *Proteins.* – 2003. – Vol. 50, №1. – P. 5-25.
8. Blazewicz J., Lukasiak P., Milostan M. // *Artif. Intell. Med.* – 2005. – Vol. 35, №1-2. – P. 135-145.
9. Chang C., Ekins S., Bahadduri P., Swaan P.W. // *Adv. Drug Deliv. Rev.* – 2006. – Vol. 58, №12-13. – P. 1431-1450.
10. Chang M.W., Ayeni C., Breuer S., Torbett B.E. // *PLoS ONE.* – 2010. – Vol. 5, №8. – P. e11955.
11. Chen D., Martin Z.S., Soto C., Schein C.H. // *Bioorg. Med. Chem.* – 2009. – Vol. 17, №14. – P. 5189-5197.
12. Chen Y.Z., Zhi D.G. // *Proteins.* – 2001. – Vol. 43, №2. – P. 217-226.
13. Chinea G., Padron G., Hooft R.W. et al. // *Proteins.* – 1995. – Vol. 23, №3. – P. 415-421.
14. Chivian D., Kim D.E., Malmstrom L. et al. // *Proteins.* – 2003. – Vol. 53, suppl. 6. – P. 524-533.
15. Cramer R.D., Patterson D.E., Bunce J.D. // *J. Am. Chem. Soc.* – 1988. – Vol. 110, №18. – P. 5959-5967.
16. Dai Y., Wang Q., Zhang X. et al. // *Eur. J. Med. Chem.* – 2010. – Vol. 45, №12. – P. 5612-5620.
17. Dudek A.Z., Arodz T., Galvez J. // *Comb. Chem. High Throughput Screen.* – 2006. – Vol. 9, №3. – P. 213-228.
18. Dunbrack R.L. Jr., Karplus M. // *J. Mol. Biol.* – 1993. – Vol. 230, №2. – P. 543-574.
19. Ekins S., Erickson J.A. // *Drug Metab. Dispos.* – 2002. – Vol. 30, №1. – P. 96-99.
20. Ekins S., Mestres J., Testa B. // *Br. J. Pharmacol.* – 2007. – Vol. 152, №1. – P. 21-37.
21. Ekins S., Mestres J., Testa B. // *Br. J. Pharmacol.* – 2007. – Vol. 152, №1. – P. 9-20.
22. Ekins S., Shimada J., Chang C. // *Adv. Drug Deliv. Rev.* – 2006. – Vol. 58, №12-13. – P. 1409-1430.
23. Filgueira de Azevedo W., Dias R. // *Bioorg. Med. Chem.* – 2008. – Vol. 16, №20. – P. 9378-9382.
24. Gemma S., Campiani G., Butini S. et al. // *J. Med. Chem.* – 2009. – Vol. 52, №2. – P. 502-513.
25. Gohlke H., Hendlich M., Klebe G. // *J. Mol. Biol.* – 2000. – Vol. 295, №2. – P. 337-356.
26. Gohlke H., Klebe G. // *Curr. Opin. Struct. Biol.* – 2001. – Vol. 11, №2. – P. 231-235.
27. Guerin G.A., Pratuangdejkul J., Alemany M. et al. // *Drug Discov. Today.* – 2006. – Vol. 11, №21-22. – P. 991-998.
28. Hooft R.W., Vriend G., Sander C., Abola E.E. // *Nature.* – 1996. – Vol. 381, №6580. – P. 272.
29. Horn F., Bettler E., Oliveira L. et al. // *Nucleic Acids Res.* – 2003. – Vol. 31, №1. – P. 294-297.
30. Ji Z.L., Wang Y., Yu L. et al. // *Toxicol. Lett.* – 2006. – Vol. 164, №2. – P. 104-112.
31. Karplus M., McCammon J.A. // *Nat. Struct. Biol.* – 2002. – Vol. 9, №9. – P. 646-652.
32. Kinsner-Ovaskainen A., Rzepka R., Rudowski R. et al. // *Toxicol. In Vitro.* – 2009. – Vol. 23, №3. – P. 476-485.
33. Kitchen D.B., Decornez H., Furr J.R., Bajorath J. // *Nat. Rev. Drug Discov.* – 2004. – Vol. 3, №11. – P. 935-949.
34. Klopman G., Dimayuga M., Talafous J. // *J. Chem. Inf. Comput. Sci.* – 1994. – Vol. 34, №6. – P. 1320-1325.
35. Krasowski M.D., Yasuda K., Hagey L.R., Schuetz E.G. // *Mol. Endocrinol.* – 2005. – Vol. 19, №7. – P. 1720-1739.
36. Lagunin A., Stepanchikova A., Filimonov D. et al. // *Bioinformatics.* – 2000. – Vol. 16, №8. – P. 747-748.
37. Langowski J., Long A. // *Adv. Drug Deliv. Rev.* – 2002. – Vol. 54, №3. – P. 407-415.
38. Lape M., Elam C., Paula S. // *Biophys. Chem.* – 2010. – Vol. 150, №1-3. – P. 88-97.
39. Leach A.R., Shoichet B.K., Peishoff C.E. // *J. Med. Chem.* – 2006. – Vol. 49, №20. – P. 5851-5855.

40. Lemmen C., Lengauer T. // *J. Comput. Aided Mol. Des.* – 2000. – Vol. 14, №3. – P. 215-232.
41. Li H., Gao Z., Kang L. et al. // *Nucleic Acids Res.* – 2006. – Vol. 34. – P. W219-W224.
42. Lin C.H., Peng Y.H., Coumar M.S. et al. // *J. Med. Chem.* – 2009. – Vol. 52, №8. – P. 2618-2622.
43. Ling Q., Huang Y., Zhou Y. et al. // *Bioorg. Med. Chem.* – 2008. – Vol. 16, №15. – P. 7399-7409.
44. Lovell S.C., Word J.M., Richardson J.S., Richardson D.C. // *Proteins.* – 2000. – Vol. 40, №3. – P. 389-408.
45. Luis G., Valerio J. // *Toxicol. Appl. Pharmacol.* – 2009. – Vol. 241, №3. – P. 356-370.
46. Matthews E.J., Frid A.A. // *Regul. Toxicol. Pharmacol.* – 2010. – Vol. 56, №3. – P. 247-275.
47. Moore L.B., Maglich J.M., McKee D.D. et al. // *Mol. Endocrinol.* – 2002. – Vol. 16, №5. – P. 977-986.
48. Muegge I., Enyedy I.J. // *Curr. Med. Chem.* – 2004. – Vol. 11, №6. – P. 693-707.
49. Nicolle E., Boumendjel A., Macalou S. et al. // *Adv. Drug Deliv. Rev.* – 2009. – Vol. 61, №1. – P. 34-46.
50. Okimoto N., Futatsugi N., Fuji H. et al. // *PLoS Comput. Biol.* – 2009. – Vol. 5, №10. – P. e1000528.
51. Opsenica I., Opsenica D., Lanteri C.A. et al. // *J. Med. Chem.* – 2008. – Vol. 51, №19. – P. 6216-6219.
52. Ortiz A.R., Pisabarro M.T., Gago F., Wade R.C. // *J. Med. Chem.* – 1995. – Vol. 38, №14. – P. 2681-2691.
53. Pastor M., Cruciani G., McLay I. et al. // *J. Med. Chem.* – 2000. – Vol. 43, №17. – P. 3233-3243.
54. Pieper U., Esvar N., Webb B.M. et al. // *Nucleic Acids Res.* – 2009. – Vol. 37. – P. D347-D354.
55. Ponte-Sucre A., Gulder T., Wegehaupt A. et al. // *J. Med. Chem.* – 2009. – Vol. 52, №3. – P. 626-636.
56. Rawi R., Whitmore L., Topf M. // *Bioinformatics.* – 2010. – Vol. 26, №13. – P. 1673-1674.
57. Reschly E.J., Krasowski M.D. // *Curr. Drug Metab.* – 2006. – Vol. 7, №4. – P. 349-365.
58. Roberts D.W., Patlewicz G., Kern P.S. et al. // *Chem. Res. Toxicol.* – 2007. – Vol. 20, №7. – P. 1019-1030.
59. Schneidman-Dukhovny D., Dror O., Inbar Y. et al. // *J. Comput. Biol.* – 2008. – Vol. 15, №7. – P. 737-754.
60. Schneidman-Dukhovny D., Dror O., Inbar Y. et al. // *Nucleic Acids Res.* – 2008. – Vol. 36. – P. W223-W228.
61. Segall M.D., Beresford A.P., Gola J.M. et al. // *Expert Opin. Drug Metab. Toxicol.* – 2006. – Vol. 2, №2. – P. 325-337.
62. Shen Q., Shi W.M., Kong W. // *Artif. Intell. Med.* – 2010. – Vol. 49, №1. – P. 61-66.
63. Shoichet B.K., McGovern S.L., Wei B., Irwin J.J. // *Curr. Opin. Chem. Biol.* – 2002. – Vol. 6, №4. – P. 439-446.
64. Sieburg H. // *Stud. Sci. Compl.* – 1990. – №12. – P. 321-342.
65. Song X., Li Y., Liu J. et al. // *J. Pharmacol. Exp. Ther.* – 2005. – Vol. 312, №2. – P. 571-582.
66. Sotriffer C.A., Sanschagrín P., Matter H., Klebe G. // *Proteins.* – 2008. – Vol. 73, №2. – P. 395-419.
67. Taha M.O., Dahabiyeh L.A., Bustanji Y. et al. // *J. Med. Chem.* – 2008. – Vol. 51, №20. – P. 6478-6494.
68. Testa B., Balmat A., Long A. // *Pure Appl. Chem.* – 2004. – №76. – P. 907-914.
69. Timmers L.F., Caceres R.A., Vivian A.L. et al. // *Arch. Biochem. Biophys.* – 2008. – Vol. 479, №1. – P. 28-38.
70. Tingjun H., Junmei W., Lirong C., Xu X. // *Protein Eng.* – 1999. – Vol. 12, №8. – P. 639-648.
71. Venselaar H., Joosten R.P., Vroliing B. et al. // *Eur. Biophys. J.* – 2010. – Vol. 39, №4. – P. 551-563.
72. Willett P. // *Trends Biotechnol.* – 1995. – Vol. 13, №12. – P. 516-521.
73. Xiang Z., Soto C.S., Honig B. // *Proc. Natl. Acad. Sci. USA.* – 2002. – Vol. 99, №11. – P. 7432-7437.
74. Yuriev E., Agostino M., Ramsland P.A. // *J. Mol. Recognit.* – 2011. – Vol. 24, №2. – P. 149-164.
75. Zou J., Xie H.Z., Yang S.Y. et al. // *J. Mol. Graph. Model.* – 2008. – Vol. 27, №4. – P. 430-438.