

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ БІООРГАНІЧНОЇ ХІМІЇ ТА НАФТОХІМІЇ

Танін Володимир Олександрович

УДК: 577.151.4 : 577.152.3 : 519.6

**НОВІ ПІДХОДИ ДО *IN SILICO* ДОСЛІДЖЕННЯ ІНГІБІТОРІВ
ПРОТЕЇНТИРОЗИНФОСФАТАЗИ 1В**

02.00.10 – біоорганічна хімія

АВТОРЕФЕРАТ

дисертації на здобуття наукового ступеня
кандидата хімічних наук

Київ – 2016

Дисертацією є рукопис.

Робота виконана у відділі механізмів біоорганічних реакцій
Інституту біоорганічної хімії та нафтохімії НАН України

Науковий керівник: член-кореспондент НАН України,
доктор хімічних наук, професор
Вовк Андрій Іванович,
Інститут біоорганічної хімії та нафтохімії НАН України,
завідувач відділу механізмів біоорганічних реакцій

Офіційні опоненти: доктор хімічних наук, старший науковий співробітник
Роженко Олександр Борисович,
Інститут органічної хімії НАН України,
старший науковий співробітник
відділу фізико-хімічних досліджень

кандидат біологічних наук, старший науковий співробітник
Карпов Павло Андрійович,
Інститут харчової біотехнології та геноміки НАН України,
завідувач лабораторії біоінформатики
та структурної біології

Захист відбудеться 15 квітня 2016 р. о 10-й годині на засіданні спеціалізованої вченої ради Д 26.220.01 в Інституті біоорганічної хімії та нафтохімії НАН України, 02094, Київ-94, вул. Мурманська, 1.

З дисертацією можна ознайомитись у бібліотеці Інституту біоорганічної хімії та нафтохімії НАН України, 02094, Київ-94, вул. Мурманська, 1.

Автореферат розісланий 14 березня 2016 р.

Вчений секретар
спеціалізованої вченої ради

В.О. Євдокименко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Відомо, що наявність значної кількості лікарських препаратів з огляду на ряд причин не зупиняє пошук нових ефективних і безпечних речовин для застосування в медицині. В процесі створення нових ліків синтезується значна кількість органічних сполук, що досліджуються в різних біологічних тестах *in vitro* та *in vivo* з метою визначення сполук-лідерів. Альтернативний шлях пошуку потенційно біоактивних сполук та їх оптимізації до структурних прототипів можливих ліків включає застосування комп'ютерних методів. При цьому цілеспрямоване конструювання *in silico* включає оцінку біологічних властивостей ліганду на основі його хімічної структури або ж моделювання з урахуванням структури хімічної сполуки і відповідної білкової мішені. Завдяки накопиченню нових даних щодо просторової будови білків і їх комплексів з різними лігандами дослідження в цьому напрямі постійно розширюються.

Останнім часом значну увагу в процесі дизайну біоактивних сполук привертають протеїнтирозинфосфатази як ензими, що контролюють стан фосфорильованості білків і тим самим регулюють сигнальні клітинні та метаболічні шляхи. Внутрішньоклітинна протеїнтирозинфосфатаза 1В (PTP1B) каталізує дефосфорилування залишків фосфотирозину інсулінового і лептинового рецепторів і розглядається як негативний регулятор в процесах інсулінової і лептинової сигналізації. У цьому зв'язку PTP1B може бути однією із перспективних терапевтичних мішеней для органічних молекул в процесі лікування діабету другого типу та ожиріння. Крім того, інгібітори PTP1B розглядаються як можливі ліки від деяких ракових захворювань. Тому формування нових підходів до пошуку і комп'ютерного моделювання активності органічних сполук як інгібіторів PTP1B є актуальним завданням біоорганічної хімії. Застосування комп'ютерних розрахунків на основі узагальнення відомих експериментальних даних і вивчення зв'язку між структурою та активністю надає можливість прогнозування властивостей нових інгібіторів протеїнтирозинфосфатаз, а також з'ясування детальних механізмів інгібувальної здатності потенційно біоактивних сполук.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконувалась у відповідності з планами науково-дослідних робіт Інституту біоорганічної хімії та нафтохімії НАН України (тема 2.1.10.13-10 "Пошук і модельні дослідження потенційно біоактивних сполук", № держреєстрації 0110U000375; тема ЦНП 9.1-12 "Розвиток методів синтезу, дослідження властивостей та механізмів дії нових потенційно біоактивних сполук", № держреєстрації 012 U002657; тема 2/03-13 "Молекулярний дизайн, синтез і дослідження інгібіторів протеїнтирозинфосфатаз як потенційних лікарських засобів проти цукрового діабету та інших захворювань", № держреєстрації 0113U005097).

Метою дослідження було встановлення нових підходів до *in silico* дослідження малих органічних молекул як інгібіторів протеїнтирозинфосфатази 1В. Для досягнення цієї мети необхідно було виконати наступні завдання.

1. На основі аналізу структури комплексів РТР1В з відомими інгібіторами оцінити рухливість амінокислотних залишків активної поверхні ензиму і визначити основні репрезентативні конформації активного центру та основні конформації рухомої петлі 110-120 поза межами активного центру.

2. Вивчити можливість комп'ютерного моделювання взаємодії лігандів з РТР1В на основі класифікації конформацій та нової модифікації програми для докінгу AutoDock 4.2, що орієнтована на прискорений скринінг органічних сполук.

3. Побудувати нові QSAR моделі для прогнозування активності потенційних інгібіторів РТР1В.

4. На основі поєднання молекулярного докінгу і QSAR моделювання провести пошук інгібіторів РТР1В з використанням значної кількості органічних сполук.

5. Створити нові оціночні функції для молекулярного докінгу на основі функцій AutoDock, AutoDock Vina і RF-score, що забезпечуватимуть кращу відповідність розрахунків з експериментальними результатами.

Об'єкт дослідження – протеїнтирозинфосфатаза 1В та синтетичні інгібітори цього ензиму.

Предмет дослідження – конформації активного центру РТР1В, комплекси РТР1В з інгібіторами, методологія молекулярного докінгу, зв'язок структура-активність.

Методи дослідження – молекулярний докінг, QSAR, кінетичний і статистичний аналіз.

Наукова новизна одержаних результатів. Вперше було проведено порівняння конформацій РТР1В згідно даних PDB-файлів для характеристики ліганд-зв'язувальних центрів (поділ на кластери) та рухливості залишків амінокислот. Сформовано кластери, які відрізняються за конформацією WPD-петлі, а також Arg47, YRD-петлі, залишку Arg24 зі вторинного арилфосфатного зв'язувального центру та залишку Lys120 з рухливої петлі 110-121. Центроїди кластерів запропоновано як репрезентативні структури РТР1В для проведення розрахунків методом молекулярного докінгу та наступного аналізу одержаних результатів. Розроблено методологію молекулярного докінгу (модифікована версія AutoDock) для вивчення зв'язування лігандів у великих зв'язувальних порожнинах на поверхні ензиму та для швидшого і надійнішого віртуального скринінгу. Побудовано три моделі QSAR для оцінювання активності інгібіторів РТР1В із залученням зібраної навчальної бази із більше ніж 2000 сполук. Вперше було розроблено нові оціночні функції для молекулярного докінгу, що базуються на відомих оціночних функціях програм AutoDock і AutoDock Vina, а також RF Score, які показали вищу ефективність для більшості комплексів РТР1В у порівнянні з вихідними функціями. За результатами тестових

досліджень обґрунтовано застосування нових підходів для пошуку інгібіторів РТР1В.

Практичне значення одержаних результатів. На основі запропонованих підходів було створено нове програмне забезпечення для встановлення репрезентативних конформацій ензиму, прийнятних для моделювання комплексів з потенційними інгібіторами. Це надає можливість використання ряду типових структур (у вигляді PDB-файлів, що є стандартним форматом для білків) для віртуального скринінгу. Проведення комп'ютерних досліджень із використанням модифікованої версії AutoDock може забезпечувати результати з прогнозованими значеннями констант інгібування, що відповідають експериментальним даним. Запропоновано нові гібридні оціночні функції на основі AutoDock та AutoDock Vina, а також RF Score можуть мати практичне застосування для молекулярного докінгу у випадку дослідження різних ферментів. Рейтинговий підхід для проведення оцінки великого масиву сполук як потенційних інгібіторів з урахуванням різноманіття конформацій білкової мішені та одночасним використанням QSAR і функцій AutoDock може бути використаний для розробки і подальшої структурної оптимізації інгібіторів РТР1В та інших протеїнтирозинфосфатаз.

Особистий внесок здобувача. Дисертантом самостійно виконано дослідження, пов'язані з молекулярним моделюванням. Проведено аналіз даних літератури, виконано теоретичні розрахунки, здійснено порівняння результатів розрахунків ензим-інгібіторних комплексів та кореляцій між структурою і активністю. Постановка наукових завдань, обговорення та інтерпретація одержаних результатів і формулювання висновків були здійснені спільно з науковим керівником – чл. кор. НАН України А.І. Вовком та канд. хім. наук В.Ю. Танчуком. Дослідження активності інгібіторів *in vitro* виконано канд. біол. наук Л.А. Кононець.

Апробація результатів дисертації. Результати роботи були представлені на наукових конференціях ІБОНХ НАН України (Київ, 2009 р., 2011 р., 2015 р.), IV Міжнародному симпозиумі "Methods and applications of computational chemistry" (Львів, 2011), III Міжнародному симпозиумі "Intracellular signalling and bioactive molecules design" (Львів, 2012), міжнародній конференції "Биологические активные вещества и материалы: фундаментальные и прикладные вопросы получения и применения" (Новий Світ, 2013), III Міжнародній конференції "High performance computing" (Київ, 2013).

Публікації. За матеріалами дисертації опубліковано 12 робіт, з них 6 статей у фахових наукових журналах і 6 тез доповідей в матеріалах наукових конференцій.

Структура та об'єм роботи. Дисертація складається зі вступу, п'яти розділів, висновків і списку літератури (165 найменувань). Дисертаційна робота налічує 152 сторінки друкованого тексту, проілюстрована 15 таблицями та 33 рисунками. Розділ 1 є оглядом літератури, а в наступних розділах дисертації викладено основний зміст роботи.

ОСНОВНИЙ ЗМІСТ РОБОТИ

Огляд літератури (розділ 1) присвячено результатам дослідження структури і функцій РТР1В, аналізу активності ряду відомих інгібіторів РТР1В, сучасним *in silico* методам, які використовуються для пошуку і створення потенційних інгібіторів ензимів.

Класифікація конформацій активного центру РТР1В (розділ 2). Рентгеноструктурний аналіз комплексів РТР1В з інгібіторами свідчить про відмінності між конформаціями зв'язувального центру. Очевидно, що конформаційні зміни в активному центрі РТР1В, які представлені значною кількістю структур в білковій базі даних, повинні братися до уваги при проведенні комп'ютерного моделювання взаємодії між ензимом та інгібітором. На сьогодні успішно використовується ряд програм для докінгу, в тому числі молекулярний докінг з використанням ансамблю різних конформацій, що враховує рухливість амінокислотних залишків ліганд-зв'язувальних центрів. Однак, для віртуального скринінгу, який включає обробку баз даних, неможливо використовувати дані близько сотні кристалічних структур РТР1В. Це привело нас до рішення, що всі наявні структури РТР1В можуть бути розподілені на групи (кластери), базуючись на факторі подібності конформацій ліганд-зв'язувальних центрів. Центроїди (репрезентативні конформації), що представляють кластери, можуть бути використані для подальшого молекулярного моделювання. Для цього було необхідно мати спеціальний інструмент, який би дозволив порівнювати лише обрані частини (активні центри) білкових структур, а також надавав інформацію щодо рухливості окремих залишків.

Алгоритм порівняння і кластеризації конформацій активної поверхні ензиму. Аналіз кристалічних структур РТР1В було проведено з використанням програмного забезпечення, створеного нами. Ідея нового підходу полягала в наступному. Алгоритм АСТРДВСМР (Active Part of PDB Comparison) складався з валідації структур, їх попарного порівняння, кластерного аналізу та визначення рухливості кожного залишку. Алгоритм включав: 1) зчитування та упорядкування ланцюгів амінокислот ензиму; 2) пошук відповідного заданого фрагменту в усіх ланцюгах ензиму; 3) попарне порівняння знайдених фрагментів ланцюга; 4) створення матриці середніх квадратичних відхилень між залишками амінокислот; 5) кластерний аналіз, за допомогою якого обирається попередньо обумовлена кількість кластерів; 6) визначення рухливості кожного залишку амінокислот; 7) за необхідності генерування PDB-файлів для кожної пари накладених один на одного фрагментів (для центроїдів кластерів). Для такого аналізу було обрано залишки амінокислот як у каталітичному центрі, так і в сусідньому додатковому зв'язувальному центрі РТР1В. При цьому включено амінокислотні залишки WPD-петлі, яка відіграє важливу роль в каталітичному механізмі. Залишки, які знаходились на відстані до 5Å від молекули інгібітора та від залишків амінокислот WPD-петлі, створювали результуючий фрагмент, що складався з 32 залишків амінокислот.

Рухливість залишків амінокислот РТР1В. 102 фрагменти активних центрів, які пройшли відбір, було знайдено серед 91 структури РТР1В в Protein Data Bank. Використовуючи створену програму АСТРDВСМР, було проведено аналіз відмінностей між кристалічними структурами РТР1В і визначено рухливість залишків амінокислот, які можуть бути важливими для зв'язування інгібітора. Найрухливішими виявилися залишки амінокислот Phe182 та Asp181, які розташовані приблизно всередині WPD-петлі, яка в процесі конформаційних змін забезпечує перебіг каталітичних перетворень. Значною мірою зміщеним було також положення залишку Arg47, що належить до так званої YRD-петлі (залишки амінокислот Tyr46-Asp48), які залучена до зв'язування фосфотирозинового субстрату. Несподівано досить рухливим виявився залишок Lys120, розташований на певній відстані від активного центру. Помітні зміни спостерігалися для залишків амінокислот, передусім, Arg24, які утворюють вторинний центр зв'язування.

Класифікація конформацій РТР1В. Після проведення ієрархічного кластерного аналізу було вирішено обмежити кількість кластерів до п'яти. Ця кількість видавалася прийнятною для подальшого моделювання. Табл. 1 характеризує кластери та центроїди, які є структурами із мінімальною сумою відстаней по відношенню до всіх структур у кластері. Фактично було отримано сім кластерів, два з яких представлені одним або двома PDB-файлами.

Таблиця 1.

Кластери PDB-структур

Кластер	Центроїд (PDB код)	Кількість структур	RMSD від 2HNP (Å)	WPD-петля
1	2HNP ^a	1	0	Відкрита
2	1OEM ^b	2	2,84	Відкрита
3	1NL9	7	1,37	Відкрита
4	1PH0	15	1,02	Відкрита
5	2CNF	22	2,39	Закрита
6	1Q6M	11	2,51	Закрита
7	2CM8	44	2,23	Закрита

^a вільна від ліганду форма ферменту; ^b окиснена форма ферменту.

Важливо відзначити, що певні структури (1OEM, 1OES та 2HNP) виявили себе як такі, що є відмінними від усіх інших PDB-структур. Дві з них (1OEM та 1OES) є окисненою формою ензиму і були виокремлені до спеціального кластеру. Ензим без ліганду (2HNP) також утворює спеціальний окремий кластер. Це означає, що ензим, який не взаємодіє із жодним лігандом, також значною мірою відрізняється від будь-якої іншої конформації комплексу ензим-ліганд.

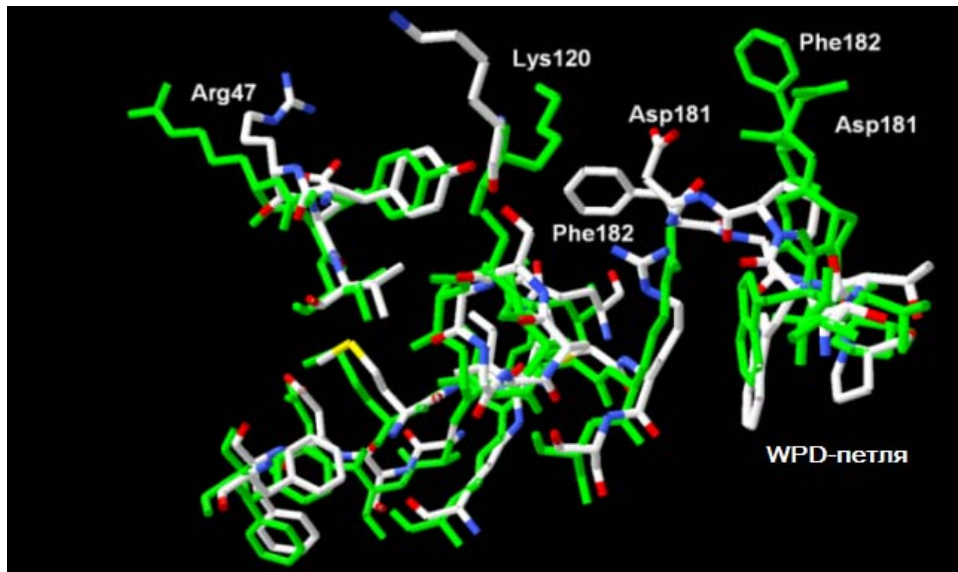


Рис. 1. Суперпозиція конформацій центрів 1Q6M та 1NL9. У цьому випадку представлено найбільш відмінні центроїди (середнє квадратичне відхилення (RMSD) складає 2,63 Å).

Як видно з рис. 1, конформації ензиму характеризуються різним положенням бокових ланцюгів Arg47 та Lys120, а також положенням залишків WPD-петлі, для яких спостерігається найбільше зміщення.

Загалом існує велика відмінність між конформаціями всіх центрів (табл. 1) із відкритою та закритою WPD-петлею, яка може бути доповнена відмінностями в бокових ланцюгах Arg47 і Lys120. Відмінність поміж кластерами, які належать до однієї і тієї ж групи, не є такою значною, але вона є помітною. У випадку із закритими WPD-петлями відмінність поміж конформаціями залишків Lys120 є найбільшою, тоді як у випадках із відкритими петлями (1NL9 та 1PH0), або в іншому випадку із парою центрів із закритою петлею це стосується положення залишку Arg47. У всіх випадках конформації комплексів досить відрізнялися від конформації вільного ензиму (2HNP), який утворює окремий кластер (середнє значення RMSD від усіх інших структур складає 2,06 Å). Це робить комп'ютерне конструювання з використанням конформацій вільного ензиму дещо проблематичним з огляду на індуковане вбудовування, тобто суттєву зміну конформації протеїну при утворенні комплексу з інгібітором.

Аналіз конформаційної рухливості петлі 110-120 РТР1В. Вище було описано спеціальний підхід для аналізу відомих конформацій активного центру РТР1В за даними PDB-файлів. Фрагмент, що його було піддано аналізу, включав ділянку ензиму з каталітичним Cys215, WPD-петлею та вторинним зв'язувальним центром (Arg24, His25, Ala27, Phe52, Arg254, Met258 і Gly259). При цьому майже всі найрухливіші залишки амінокислот належали до WPD-петлі. Однак досить рухливим виявився також залишок Lys120, який відносять до петлі 110-120. WPD-петля та петля 110-120 виглядають як такі, що контролюють вхід до каталітичної кишені РТР1В.

У відповідності до результатів виконаних розрахунків кристалічних структур, петля 110-120 має стандартну конформацію незалежно від того, відкрита чи закрита WPD-петля. Однак, виявилось, що петля 110-120 змінює свою конформацію у разі утворення комплексів із певними інгібіторами з об'ємною структурою. Цікаво відзначити той факт, що WPD-петля у таких випадках є закритою (рис. 2.) Три нестандартні конформації петлі 110-120 спостерігаються виключно для обмеженої кількості структур ензиму. Крім того, більшість нестандартних конформацій петлі 110-120 знаходяться у PDB-файлах із кількома ланцюгами РТР1В.

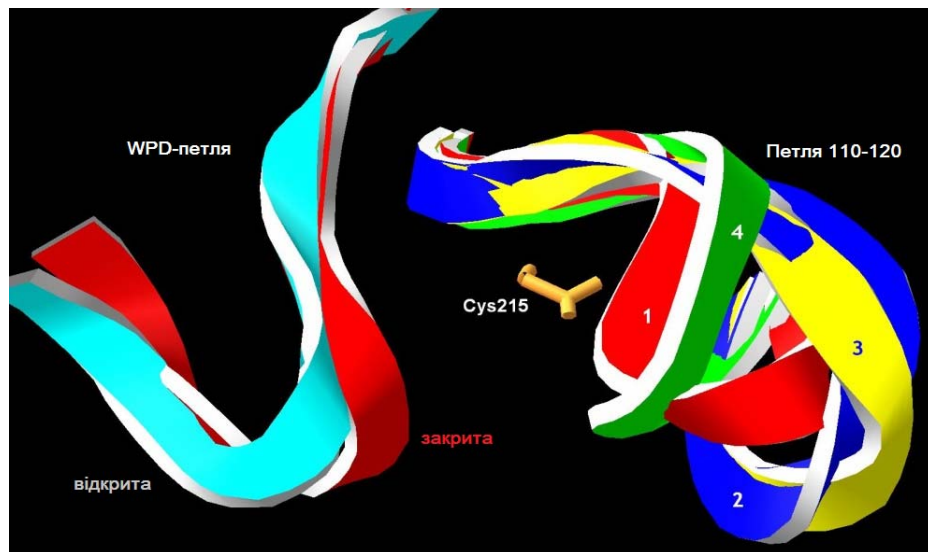


Рис. 2. Чотири конформації петлі 110-120 в структурах 2СМ8 (1), 1Q6P-В (2) 1Q6N-А (3) і 1Q6Т-В (4). WPD-петля представлена відкритою конформацією (1РНО) та закритою конформацією (2СМ8).

Докінгові дослідження РТР1В (розділ 3).

Модифікована версія AutoDock 4.2 для роботи з ензимами з об'ємним зв'язувальним сайтом. Ми припустили, що способи зв'язування біоізостерних фрагментів інгібіторів в активному центрі РТР1В можуть бути уніфіковані для порівняльного прискореного скринінгу органічних сполук. Виконана з цією метою модифікація AutoDock забезпечувала: 1) здатність застосовувати одне або кілька просторових обмежень по відношенню до положень ліганда в сайті зв'язування, що є корисним у випадку, коли відомі групи, відповідальні за фіксацію інгібітора; 2) здатність використовувати стандартні багатовимірні техніки оптимізації замість стохастичного пошуку. Пошук із обмеженнями, накладеними на оптимізацію зв'язування, потребує приблизно в 100 разів менше часу у порівнянні із оригінальною версією. Крім того, модифікований AutoDock знаходить нижчий енергетичний мінімум; 3) здатність виконувати послідовний пошук щодо кращої конформації, яка проходить до наступного циклу. Це приводить до глобальної оптимізації; 4) здатність використовувати заряди, розраховані методом МОРАС 2009 РМ6, який відомий тим, що покращує положення зв'язування; 5) багатопотоковість.

Особливості нового пошукового алгоритму включають наступне.

Обмеження положень ліганду. Ключовою ідеєю вбачалося додавання певних обмежень у такий спосіб, щоб певні атоми інгібітора розміщувались тільки у певних місцях простору. Цей підхід було реалізовано шляхом додавання нової характеристики - обмеження положень атома. У тому випадку, коли відстань від атома до центру області перевищує максимальне значення, додається штрафне значення до енергії, розрахованої AutoDock.

Новий режим оптимізації. Було застосовано новий режим пошуку, який базується на стандартній багатовимірній оптимізації.

Поетапна (загальна) оптимізація. При використанні нової альтернативи, названої поетапною оптимізацією, найкращий загальний результат не втрачається. Після завершення часу проведення операції залишається найкращий окремих варіант, який і переходить до наступної операції. Інші окремі варіанти вихідної множини генеруються в довільний спосіб та змагаються із найкращим результатом попередніх операцій. Ідея є схожою до «елітної селекції» у генетичному алгоритмі. У такому випадку паралельні запуски операцій генетичного алгоритму робляться послідовними.

На основі викладеного вище було створено нову версію AutoDock, яка виявилась результативною для дослідження фіксації лігандів у великих зв'язувальних порожнинах та для надійнішого віртуального скринінгу. Нова процедура пошуку є від 10 до 30 разів швидшою, ніж оригінальний варіант генетичного алгоритму. Багатопотоковість приводить до мінімізації потреб щодо пам'яті, дозволяючи використовувати повну потужність наявних процесорів. Завдяки цьому віртуальний скринінг великої кількості органічних молекул може бути доступнішим.

Докінг відомих інгібіторів з використанням типових конформацій активного центру РТР1В. Кінцевою метою цієї частини роботи було вивчення відмінностей між кластерами конформацій РТР1В шляхом докінгу в активний центр ензиму відомих інгібіторів різної структури. Для цього було використано 205 описаних в літературі інгібіторів РТР1В. Для оптимізації розрахунків докінг було проведено за допомогою модифікованої версії AutoDock. Розподіл найкращих результатів за кластерами (табл. 2) демонструє, що вони характерні для всіх кластерів, в тому числі з конформацією ензиму з закритою (останні три) та відкритою WPD-петлею (перші два). Виявилось, що немає єдиного кластеру, який би мав достатньо добру кореляцію з експериментальними результатами. Також не було доброї кореляції між найвищими або найнижчими розрахованими і експериментальними значеннями. З іншого боку, кожна сполука мала варіант зв'язування із передбаченим значенням rK_i , що є досить близьким до експериментального. Ці значення корелюють з експериментальними значеннями досить добре ($R^2 = 0,69$, RMSE (середня квадратична похибка) = 0,79). Виявлено також випадки, коли AutoDock не зміг розрахувати належно значення rK_i з причини певних обмежень методу (три сполуки). Кореляція після вилучення цих трьох сполук була значно кращою ($R^2=0,76$, RMSE = 0,86).

Кластери PDB-структур та розподіл між ними результатів докінгу, які найбільше відповідали експериментальним даним

Кластер	Центроїд (PDB код)	Кількість найкращих результатів докінгу для кластеру
1	1NL9	24
2	1PH0	39
3	2CNF	46
4	1Q6M	48
5	2CM8	48

Комплексний підхід до прогнозування інгібування PTP1B інгібіторами різних класів (розділ 4). Для прогнозування біологічних властивостей нових сполук поряд з молекулярним докінгом використовується метод QSAR (Quantitative Structure-Activity Relationship). Враховуючи описані у попередньому розділі модифікації AutoDock, спрямовані на збільшення швидкості розрахунків і якості можливих комплексів PTP1B з інгібіторами, з залученням QSAR з'являється можливість прогнозування активності лігандів обома методами і побудови консенсусної моделі.

QSAR моделі для прогнозування інгібування PTP1B інгібіторами різних класів. QSAR моделі можуть слугувати як важливі інструменти для автоматизованого перед-віртуального скринінгу з використанням баз даних. Однак, опубліковані QSAR-моделі для PTP1B переважно створені для певних класів хімічних речовин і не можуть використовуватись для віртуального скринінгу великих баз даних. Метою цієї частини роботи була побудова QSAR моделей для оцінки інгібувальної здатності органічних сполук різноманітної структури стосовно PTP1B.

Генерування набору даних. Для генерування набору даних було зібрано 2237 сполук із більше ніж двадцяти джерел. Біологічні дані були представлені у вигляді pK_i (або у вигляді pIC_{50}). Було використано молекулярні дескриптори, які були розраховані за допомогою пакетів Dragon, ADRIANA.Code, OEState та AlogPS. Загальне число молекулярних дескрипторів, розрахованих всіма модулями, перевищувало 6000. Константні та близькі до константних дескриптори, так само як і дескриптори із високим ступенем кореляції (кореляція $> 0,95$ з іншими дескрипторами) були виключені.

Побудова QSAR моделей. Ми спробували застосувати різні інструменти моделювання та різні набори дескрипторів, які є доступними на сайті OCHEM для побудови QSAR моделей. Найкращі результати були отримані при використанні асоціативних нейронних мереж - методу, який поєднує можливості традиційного методу штучних нейронних мереж із методом k-

найближчих сусідів. Завжди присутня дилема, чи варто використовувати всі наявні дані як тренувальний набір, чи використовувати їх частину як тестовий набір. У нашому випадку було вирішено не втрачати жодної частини інформації та покластися на досить потужні методи перехресної оцінки, що використовуються OCHEM.

Валідація моделей QSAR. Тестовий набір складався зі сполук із нового доповнення бази даних NIH - 507 нових речовин (набір 1). Деякі з них представляли нові типи сполук, тоді як інші мали структурні аналоги у тренувальному наборі (113 нових молекул мають аналоги у тренувальному наборі). Ці сполуки було використано як другий валідаційний набір. Обидва набори було використано для тестування здатності передбачення моделей QSAR.

Оцінка отриманих моделей. Модель 1 покладається тільки на дескриптори електронно-топологічного стану, які не потребують 3D-структури. Параметри цієї моделі такі: $R^2 = 0,73$, коефіцієнт перехресної оцінки q^2 (R^2 при перехресній оцінці) = 0,73; RMSE = 0,71; MAE (усереднена абсолютна похибка) = 0,5. Необхідно зазначити, що модель є дуже швидкою та може бути застосована для попереднього оцінювання потенційних інгібіторів PTP1B.

Модель 2 використовує 1750 відфільтрованих дескрипторів, більшість яких потребують 3D-структури. Ця модель є схожою за своїми параметрами з попередньою з $R^2 = 0,73$, $q^2 = 0,73$, RMSE = 0,72, MAE = 0,5, але вимагає набагато більше ресурсів для проведення розрахунків. Тим не менше, вона відрізняється від першої та може надати альтернативний погляд на протестовані сполуки (може підтвердити високу активність, передбачену першою моделлю, або зробити її сумнівною). Ми припускаємо використання першої моделі для всіх сполук, а використання другої моделі – для підтвердження прогнозів із високою активністю.

Комбінована модель (модель 3) може бути побудована, використовуючи просту формулу

$$\text{Модель3} = \text{Модель1} \cdot 0,497276 + \text{Модель2} \cdot 0,479049 + 0,120085.$$

У цьому разі спостерігається певне покращення з $R^2 = 0,76$, RMSE = 0,71, MAE = 0,48. Це є ще одним доказом того, що модель 1 та модель 2 засновані на різному підґрунті щодо активності.

Два валідаційних набори були використані для тестування обох QSAR моделей, які надали порівнювані результати. Набір 1 надав досить слабкі кореляції для обох моделей, але MAE була досить низькою в обох випадках (0,83 для моделі 1 та 0,80 для моделі 2). Набір 2 надав набагато кращі результати із MAE 0,60 для моделі 1 та MAE 0,52 для моделі 2. Модель 3 перевершила модель 1 та модель 2 на обох тестових наборах (MAE = 0,73, 128 значень, що випадають, для набору 1; MAE = 0,52, 13 значень, що випадають, та $R^2 = 0,76$ для набору 2). Це означає, що модель 3 не втратила прогнозуючої здатності та може використовуватись для комбінування результатів за двома моделями (рис. 3).

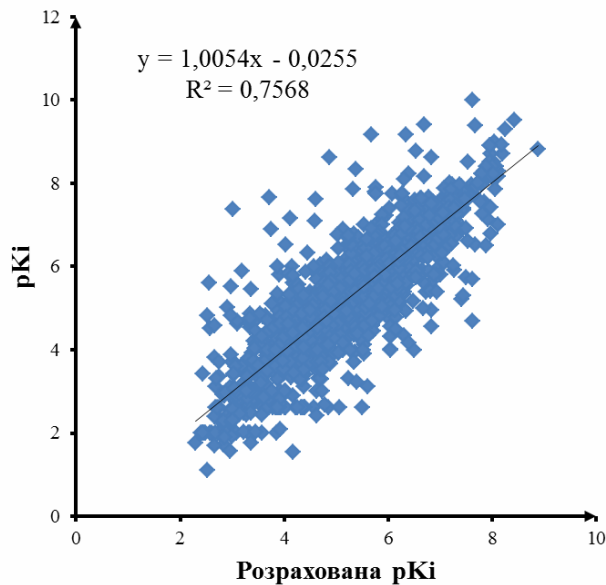
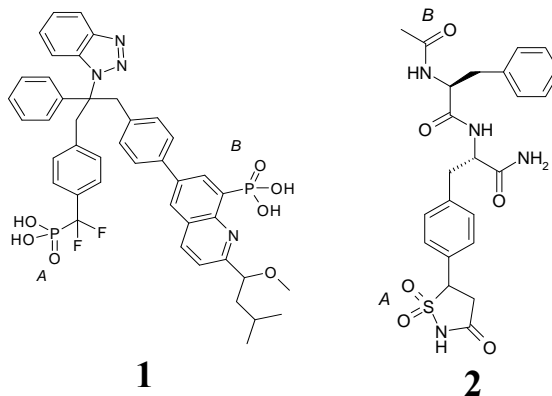


Рис. 3. Комбінована модель 3.

Пошук інгібіторів РТР1В серед набору з понад 64000 органічних сполук. З метою демонстрації можливостей запропонованих підходів було поставлено завдання провести віртуальний скринінг бібліотеки гетероциклічних сполук, яка налічувала понад 64000 речовин, синтезованих у відділі хімії біоактивних азотовмісних гетероциклічних основ ІБОНХ НАН України під керівництвом проф. В.С. Броварця. При виборі моделі фармакофора було використано структуру комплексу РТР1В (1Q6T) з відомим інгібітором **1** - бісфосфоновою кислотою на об'ємній молекулярній платформі з фрагментом бензотриазолу ($IC_{50} = 5$ нМ). Аналіз структури цього комплексу показує, що обидві фосфонатні групи беруть участь у закріпленні інгібітора в активному центрі. При цьому дифторметиленфосфонатний фрагмент (А) знаходиться поряд з залишком каталітичного Cys215, а інша фосфонатна група (В) взаємодіє з залишком Arg24. Відстань між двома фосфонатними групами у комплексі з РТР1В складає 12,76 Å.



При аналізі іншого комплексу РТР1В з інгібітором **2** (2CM7, $K_i = 210$ нМ) два фрагменти, що забезпечують формування водневих зв'язків, розташовані на відстані 12,29 Å. При цьому атом сірки ізотіазолідинового фрагменту А наближений до Cys215, а амідна група інгібітора **2** (фрагмент В) взаємодіє з

Arg47, тобто спрямована в іншому напрямку, якщо порівнювати з інгібітором 1. Це вказує на те, що потенційний інгібітор РТР1В може мати як мінімум дві групи, важливі для зв'язування, які розташовані на відстані 12-13 Å. У випадку комплексу 1Q6T між групами 16 зв'язків, а у випадку структури 2CM7 таких зв'язків 12. З огляду на це на першому етапі пошуку були відібрані сполуки, що мають від 10 до 15 зв'язків між такими групами. Серед наданої бази даних було знайдено 740 таких сполук.

Молекулярний докінг з використанням модифікованої версії AutoDock 4.2. РТР1В характеризується наявністю відносно великого заглиблення в активному центрі і має кілька додаткових ділянок, де можуть зв'язуватись функціональні групи інгібітора. Поверхня ензиму має щонайменше п'ять положень для зв'язування інгібітора за допомогою водневих зв'язків. Можна було припустити, що як мінімум один з акцепторів при утворенні водневих зв'язків розташовується поблизу Cys215, тобто може бути «прив'язаним» неподалік від нього (наприклад, на відстані не більше 5 Å від атома сірки Cys215). Решті атомів ліганду було надано можливість вільно розміститися під час оптимізації.

Паралельне використання молекулярного докінгу і QSAR. Загалом, для 740 відібраних сполук, що можуть бути доковані в активний центр РТР1В з конформаціями, які представляють п'ять кластерів, було знайдено 8001 можливий спосіб зв'язування. Для всіх способів зв'язування проведено докінг. Речовини були відсортовані за кількістю варіантів зв'язування, які давали високі прогнозовані значення pK_i (≥ 7) (так званий докінг-рейтинг). Найкращим виявився результат сполуки RAM3004098 (рис. 4). В докінговому рейтингу ця похідна роданіну, 4-{5-[3-(4-нітрофеніл)-1-фенілпіразол-4-ілметилен]-4-оксо-2-тіотіазолідин-3-іл}бутанова кислота, поділяє третю позицію, а в рейтингу QSAR знаходиться на 63-й, хоча її константа інгібування, що передбачалася, була гіршою у порівнянні з найкращою лише на порядок. Значення IC_{50} , визначене для цієї сполуки експериментально, складає 4 мкМ.

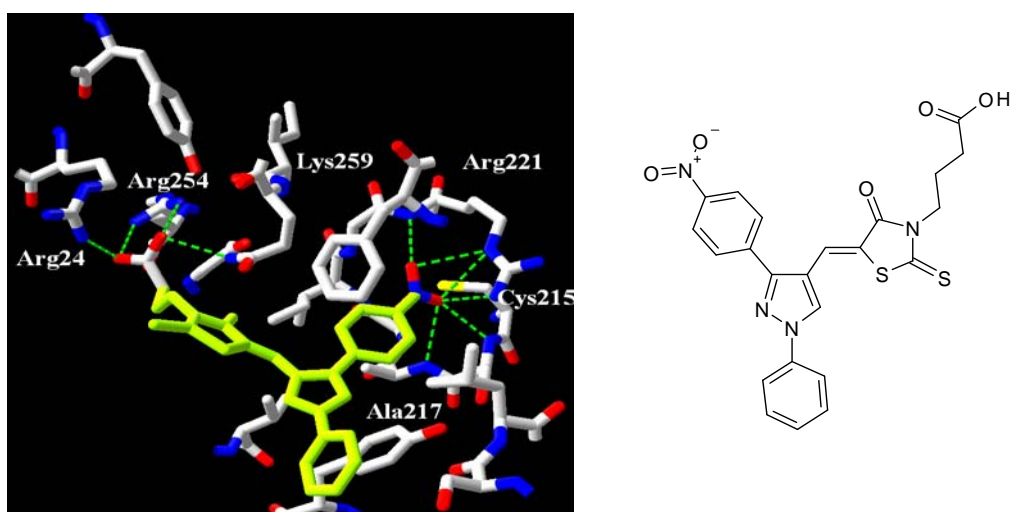


Рис 4. Структура інгібітора RAM3004098 та спосіб його зв'язування в активному центрі РТР1В (репрезентативна структура кластера 1Q6M).

Нові оціночні функції для молекулярного докінгу та оцінки констант інгібування (розділ 5).

Необхідність вдосконалення оціночної функції для молекулярного докінгу. Відомо, що молекулярний докінг складається з двох основних складових: генерування положень ліганду за допомогою просторового конформаційного пошуку (для визначення способів зв'язування ліганду з ензимом) та оцінки згенерованих положень за допомогою оціночної функції. Генерування положень ліганда залежить від алгоритмів оптимізації. У нашому випадку оптимізація досить ефективно знаходить мінімуми, що в багатьох випадках є значно нижчими за експериментальні значення. Це означає, що покращення вимагає сама оціночна функція, яка є важливою для ранжування положень ліганду у відповідності до їх потенційних спорідненостей щодо зв'язування ензимом. Метою цієї частини дослідження була перевірка двох нових оціночних функцій, створених на основі відомих раніше.

1. Об'єднання оціночних функцій AutoDock та AutoDock Vina. AutoDock і AutoDock Vina мають значно відмінні оціночні функції. Оціночна функція AutoDock походить від силового поля AMBER та включає п'ять компонентів для взаємодій між атомами та оцінку втрати конформаційної ентропії після зв'язування ліганда (ΔS_{conf}). Усі складові мають свої вагові коефіцієнти. AutoDock Vina має абсолютно відмінну оціночну функцію, побудовану на основі формалізму «машинного навчання», без жодної фізичної моделі. Не дивлячись на це, вона також базується на попарних взаємодіях між атомами, які визначаються п'ятьма складовими, кожна зі своїм ваговим коефіцієнтом. Кінцева формула включає також штраф за кількість зв'язків, які обертаються, що додає деяку нелінійність до оціночної функції. Обидва підходи мають той самий формат введення даних, завдяки чому функції можуть бути легко об'єднані.

Нова покращена оціночна функція H1 була побудована з використанням покращеного набору бази даних PDBbind версії 2012 року як тренувального набору (2412 комплексів езим-інгібітор) та 313 комплексів, що з'явилися у редакції 2013, як тестового набору. Нові оціночні функції є лінійними комбінаціями усіх складових обох оціночних функцій. Окрім того, значення функції AutoDock Vina включено в нові оціночні функції як окрему складову. Усі 11 складових елементів енергії були розраховані для кожного комплексу з тренувального та тестового наборів інгібіторів з використанням спеціально модифікованої версії AutoGrid. Ця версія AutoGrid була також здатною розрахувати значення для оціночних функцій AutoDock та AutoDock Vina. Коефіцієнти були отримані методом множинної лінійної регресії.

Нова оціночна функція H1 тестувалася під час CSAR Benchmark Exercise 2014 (<http://csardock.org>). Оціночна функція (команда T) правильно визначила 3 найкращих положення з 200 для всіх 22 лігандів, що тестувалися. З 22 випадків була лише одна помилка з визначенням найкращого положення ліганду. При цьому справжня структура комплексу була розташована на другому місці в ряду, що сформувався в результаті виконаних розрахунків.

Результуюча оціночна функція показала себе набагато краще, ніж оригінальні оціночні функції AutoDock та AutoDock Vina (таблиця 3). Результати, отримані для тестового набору, є не набагато гіршими від результатів для тренувального набору сполук. Це означає, що функція може використовуватися для прогнозування констант інгібування РТР1В.

2. *Лінійна функція на основі RF Score (RFL)*. Оціночна функція RF Score використовує новітній алгоритм машинного навчання, відомий як Random Forest, що базується на дуже простому наборі дескрипторів, які описують міжмолекулярні взаємодії. Беруться типи атомів в ензимі (C, N, O, S) та типи атомів в ліганді (ті ж C, N, O, S, а також атоми Cl, F, Br, I, P), а інші типи атомів ігноруються. Для кожної пари типів атомів знаходиться кількість відстаней, що є меншими від 12 Å, і в результаті виходить 36 цілочисельних дескрипторів. Отримана функція дозволяє прогнозувати rK_i для комплексів ензим-інгібітор з високою точністю (можливо найвищою на сьогодні). Успіх подібного підходу пояснювався високою якістю моделей, побудованих за методом Random Forest.

В результаті дисертаційного дослідження на основі зазначених вище дескрипторів побудовано лінійну регресійну модель, яка є значно простішою у порівнянні з моделлю, побудованою Random Forest. Як видно з таблиці 3, лінійна функція RFL лише незначною мірою поступається RF Score на тестовому наборі (високі результати RF Score на тренувальному наборі свідчать швидше про можливість методу Random Forest в «підгонці» даних, в той час як здатність до прогнозу перевіряється саме на тестовому наборі). Досить цікавим є те, що такий досить примітивний набір дескрипторів непогано описує міжмолекулярні взаємодії навіть у комбінації з множинною лінійною регресією.

Таблиця 3.

Ефективність нових оціночних функцій H1 та RFL у порівнянні із функціями RF Score, AutoDock та AutoDock Vina.

Оціночна функція	Тренувальний набір, 2412 комплексів			Тестовий набір, 313 комплексів		
	Коефіцієнт кореляції R	RMSE	MAE	Коефіцієнт кореляції R	RMSE	MAE
H1	0,62	1,54	1,22	0,49	1,74	1,45
RFL ^a	0,58	1,59	1,26	0,56	1,64	1,36
RF Score	0,93	0,64	0,5	0,63	1,54	1,25
AutoDock Vina	0,53	2,10	1,65	0,36	2,59	2,08
AutoDock	0,43	2,58	1,96	0,28	2,81	2,13

^a Тренувальний набір у цьому випадку складався з 2897 комплексів.

Використання нових оціночних функцій на комплексах РТР1В. При перевірці нових оціночних функцій для прогнозування структури комплексів РТР1В з потенційними інгібіторами 26 комплексів РТР1В з даних PDBbind було віднесено до тренувального набору. Як тестовий набір розглядалися інші 28 комплексів.

Узагальнені результати цього дослідження, які демонструють коефіцієнти кореляції залежності розрахованих прогнозних значень констант інгібування РТР1В від експериментальних значень цих констант для всіх комплексів, наведено в табл. 4.

Нові функції Н1 та RFL показали себе набагато краще як для тестового набору, так і для загального набору з 54 комплексів.

Таблиця 4.

Порівняння ефективності функцій на комплексах РТР1В.

Функція	Тренувальний набір, 26 комплексів			Тестовий набір, 28 комплексів			Загалом, 54 комплекси		
	R	RMSE	MAE	R	RMSE	MAE	R	RMSE	MAE
Н1	0,45	1,15	1,0	0,47	1,09	0,84	0,50	1,12	0,91
RFL	0,42	1,09	0,88	0,63	1,09	0,84	0,60	1,09	0,85
Auto-Dock Vina	0,39	1,16	0,92	0,15	2,58	2,25	0,05	2,03	1,61
Auto-Dock	0,23	2,89	2,58	0,39	1,83	1,38	0,21	2,41	1,96

Таким чином, базуючись на перевагах точнішого прогнозування зв'язування інгібіторів, комбінована гібридна оціночна функція Н1 та RFL як лінійний варіант функції RF Score можуть бути використані для виявлення перспективних хімічних сполук в процесі віртуального скринінгу можливих інгібіторів РТР1В. Очевидно, наявність двох оціночних функцій з близькими статистичними характеристиками є швидше перевагою, ніж недоліком, оскільки на кожному конкретному класі речовин перевагу може мати будь-яка з них.

ВИСНОВКИ

В дисертаційній роботі запропоновано нові підходи до вивчення протеїнтирозинфосфатази 1В як мішені для потенційних інгібіторів *in silico*, що включають класифікацію ліганд-зв'язувальних центрів, застосування нових докінгових процедур та використання даних QSAR аналізу.

1. Теоретичний аналіз відомих кристалічних структур РТР1В у комплексах з інгібіторами демонструє можливість п'яти репрезентативних конформацій активного центру та чотирьох конформацій петлі 110-120. Використання репрезентативних конформацій для молекулярного докінгу інгібіторів забезпечує врахування конформаційних змін активного центру РТР1В.

2. Фіксація біоізостерних фрагментів інгібіторів в активному центрі РТР1В може бути обмежена для порівняльного прискореного *in silico* скринінгу органічних сполук. Запропоновано нову модифікацію програми AutoDock 4.2, що орієнтована на дослідження великої кількості сполук для пошуку інгібіторів РТР1В з об'ємним центром зв'язування.

3. Для пошуку нових інгібіторів РТР1В і прогнозування їх інгібувальної здатності побудовано нові QSAR моделі на базі 2D-дескрипторів, 3D-дескрипторів та 2D+3D-дескрипторів

4. На основі класифікації ліганд-зв'язувальних центрів РТР1В, поєднання молекулярного докінгу і QSAR моделювання серед набору з понад 64000 сполук ідентифіковано нові потенційні інгібітори, що виявляють активність *in vitro* в мікромолярному діапазоні концентрацій.

5. Для молекулярного докінгу потенційних інгібіторів РТР1В запропоновано нові гібридні оціночні функції H1 та RFL на основі функцій AutoDock 4.2, AutoDock Vina і функції RF Score. У випадку комплексів РТР1В з відомими лігандами нові функції забезпечують кращі результати у порівнянні з базовими розрахунками.

СПИСОК ПУБЛІКАЦІЙ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Tanchuk V.Y. Classification of binding site conformations of protein tyrosine phosphatase 1B / Tanchuk V.Y., Tanin V.O., Vovk A.I. // Chem. Biol. Drug Des. – 2012. – Vol. 80, №1. – P. 121-128.
2. Tanchuk V.Yu. QSAR models for predicting protein tyrosine phosphatase 1B inhibition by structurally diverse inhibitors / Tanchuk V.Yu., Tanin V.O., Vovk A.I. // J. Org. Pharm. Chem. – 2013. – Vol. 11, №42. – P. 51-56.
3. Tanchuk V.Yu. Analysis of conformational flexibility of loop 110-120 of protein tyrosine phosphatase 1B / Tanchuk V.Yu., Tanin V.O., Vovk A.I. // Ukr. Biochem. J. – 2013. – Vol. 85, №5. – P. 73-80.

4. Tanchuk V. A new scoring function for molecular docking based on AutoDock and AutoDock Vina / Tanchuk V., Tanin V., Vovk A., Poda G. // Curr. Drug Discov. Technol. – 2015. – Vol. 12, №3. – P. 170-178.
5. Tanchuk V.Yu. A new scoring function for estimating protein-ligand binding based on simple RF-score like approach // Tanchuk V.Yu., Tanin V.O. // Modern Science. – 2015. – Vol. 11, № 5. – P. 51-56.
6. Tanchuk V.Yu. Testing of a new docking scoring function on the example of inhibitors of protein tyrosine phosphatase 1B / Tanchuk V.Yu., Tanin V.O. // J. Org. Pharm. Chem. – 2015. – Vol. 13 , № 4 (52). – P. 16-19.
7. Tanchuk V. Multithreaded version of AutoDock 4.2 suitable for massive virtual screening of potential biologically active compounds (enzyme inhibitors) / Tanchuk V., Tanin V., Vovk A. // Third International Conference "High Performance Computing" HPC-UA 2013 – Ukraine – Kyiv – 2013 – P. 399-401 - <http://hpc-ua.org/hpc-ua-13/files/proceedings/76.pdf>.
8. Tanchuk V.Yu. Faster, easier and better docking by modified version of AutoDock 4.2. Introduction of spacial constraints, use of stepwise optimization and automatic preparations for massive virtual screening. / Tanchuk V.Yu., Prokopenko V.V., Tanin V.O. // Biologically active substances and materials: Fundamental and applied problems. Proceedings of the conference – 2013 – Vol. 2 – P. 312.
9. Tanchuk V.Yu. QSAR models for prediction of PTP1B inhibition by structurally diverse compounds / Tanchuk V.Yu. Tanin V.O. Vovk A.I. // Abstracts 3rd International Symposium “Intracellular Signalling and Bioactive Molecules Design” – 2012. – P. 99.
10. Tanchuk V.Yu. Comparative analysis of binding site conformations of PTP1B / Tanchuk V.Yu. Tanin V.O. Vovk A.I. // Abstracts 3rd International Symposium “Intracellular Signalling and Bioactive Molecules Design” – 2012 – P. 97.
11. Tanchuk V.Yu. Classification of binding site conformations on the example of protein tyrosine phosphatase 1B (PTP1B) / Tanchuk V.Yu., Tanin V.O., Vovk A.I. // Methods and applications of computational Chemistry. 4th International symposium, Book of abstracts. – 2011. – P. 67.
12. Tanchuk V.Yu. QSAR studies of protein tyrosine phosphatase 1B (PTP1B) inhibitors / Tanchuk V.Yu. Tanin V.O. Vovk A.I. // Methods and applications of computational Chemistry. 4th International symposium, Book of abstracts. – 2011. – P. 148.

АНОТАЦІЯ

Танін В.О. – Нові підходи до *in silico* дослідження інгібіторів протеїнтирозинфосфатази 1B – Рукопис.

Дисертація на здобуття наукового ступеня кандидата хімічних наук за спеціальністю 02.00.10 - біоорганічна хімія. – Інститут біоорганічної хімії та нафтохімії НАН України, Київ, 2015.

В дисертаційній роботі виконано аналіз відомих кристалічних структур РТР1В у комплексах з інгібіторами, який було проведено за допомогою спеціально розробленої програми. Результати цього аналізу свідчать про можливість п'яти репрезентативних конформацій активного центру та чотирьох конформацій петлі 110-120. Найрухливішими виявилися залишки амінокислот Phe182 та Asp181 WPD-петлі і залишок Lys120. Розроблено методологічний підхід, за яким врахування конформаційних змін активної поверхні РТР1В втілюється шляхом використання центроїдів кластерів, і це може бути покладено в основу системного підходу до проведення молекулярного докінгу інгібіторів в активний центр РТР1В. З'ясовано, що способи зв'язування біоізостерних фрагментів інгібіторів в активному центрі РТР1В можуть бути уніфіковані для порівняльного прискореного скринінгу органічних сполук. Запропоновано нову модифікацію програми для докінгу AutoDock 4.2, що орієнтована на масовий скринінг сполук у випадку ферментів з об'ємним центром зв'язування. Побудовано нові QSAR моделі для прогнозування інгібувальної здатності синтетичних сполук. Згідно запропонованих підходів, на основі класифікації ліганд-зв'язувальних центрів РТР1В, поєднання молекулярного докінгу і QSAR серед набору з понад 64000 сполук ідентифіковано нові потенційні інгібітори РТР1В, що виявляють активність *in vitro* зі значеннями IC₅₀ в мікромольному діапазоні. Для розробки моделей взаємодії потенційних інгібіторів з активним центром РТР1В запропоновано нові оціночні функції для молекулярного докінгу на основі функцій AutoDock 4.2, AutoDock Vina і функції RF Score. У випадку комплексів РТР1В нові функції дають значно кращі результати у порівнянні з базовими розрахунками.

Ключові слова: протеїнтирозинфосфатаза 1В, інгібітори, AutoDock, QSAR, *in silico* дослідження, кластерифікація, докінг, структурна подібність.

АННОТАЦІЯ

Танин В.А. - Новые подходы к *in silico* исследованию ингибиторов протеинтирозинфосфатазы 1В – Рукопись.

Диссертация на соискание ученой степени кандидата химических наук по специальности 02.00.10 – биоорганическая химия. - Институт биоорганической химии и нефтехимии НАН Украины, Киев, 2015.

В диссертационной работе выполнен анализ известных кристаллических структур РТР1В в комплексах с ингибиторами, проведенный с помощью специально разработанной программы. Результаты этого анализа свидетельствуют о возможности пяти репрезентативных конформаций активного центра и четырех конформаций петли 110-120. Подвижными оказались остатки аминокислот Phe182 и Asp181 WPD-петли и остаток Lys120. Разработан методологический подход, согласно которому учет конформационных изменений активной поверхности РТР1В воплощается путем использования центроидов кластеров, и это может быть положено в основу системного подхода к проведению молекулярного докинга ингибиторов в активный центр РТР1В. Установлено, что способы связывания

биоизостерических фрагментов ингибиторов в активном центре PTP1B могут быть унифицированы для сравнительного ускоренного скрининга органических соединений. Предложена новая модификация программы для докинга AutoDock 4.2, ориентированная на массовый скрининг соединений в случае ферментов с объемным центром связывания. Построены новые QSAR модели для прогнозирования ингибирующих способностей новых синтетических соединений на основе известных данных о низкомолекулярных ингибиторах PTP1B. На основе классификации лиганд-связывающих центров PTP1B, использования молекулярного докинга и QSAR в наборе из более 64000 соединений идентифицированы новые потенциальные ингибиторы PTP1B, проявляющие активность *in vitro* со значениями IC₅₀ в микромолярном диапазоне. Для разработки моделей взаимодействия потенциальных ингибиторов с активным центром PTP1B предложены новые оценочные функции для молекулярного докинга на основе функций AutoDock 4.2, AutoDock Vina и функции RF Score. В случае комплексов PTP1B новые функции дают значительно лучшие результаты по сравнению с базовыми расчетами.

Ключевые слова: протеинтирозинфосфатаза 1B, ингибиторы, AutoDock, QSAR, *in silico* исследования, связывающие центры, кластерификация, докинг, структурное сходство.

SUMMARY

Tanin V.O. - New approaches for *in silico* study of inhibitors of protein tyrosine phosphatase 1B – Manuscript.

Dissertation for the candidate degree in chemical sciences, speciality 02.00.10 - bioorganic chemistry. - Institute of bioorganic chemistry and petrochemistry, NAS Ukraine, Kyiv, 2015.

Dissertation provides development of theoretical analysis of the known crystal structures of complexes of PTP1B inhibitors and it was first time, when conformations of PTP1B were classified in accordance with datum, provided in PDB-files to characterize ligand-binding centers and mobility of amino acid residues. This was conducted using specially developed program. The results of this analysis indicate the possibility to select five representative conformations of active site and four conformations of 110-120 loop. Amino acid residues of Phe182 and Asp181 of WPD-loop and the residue of Lys120 appeared to be the most flexible. In the same time it was found, that the loop 110-120 of the PTP1B is not as mobile as it was expected. In the most of cases it adopts the same conformation. Most non-standard conformations of 110-120-loop were found in PDB files with multiple chains. The result may help in further understanding of PTP1B functioning and provide a starting point for the computer simulations of inhibitor binding. It means, that offered

program provides solid instruments to study mechanisms of enzyme functioning on the base of experimental data, carrying multiple analyzing. The developed methodological approach takes into consideration changes of the active site of PTP1B by using centroids of clusters. It can be used as the basis for a systematic approach for molecular docking of potential inhibitors into the active site of PTP1B. It was found that the methods of binding of bioisosteric fragments of inhibitors in the active site of PTP1B can be unified for comparatively high-speed screening of organic compounds. The new modification of AutoDock 4.2 was created for this purpose. It focuses on mass screening of enzymes' ligands with spacy binding center. It was proved that the results of docking are heavily dependent on starting conditions. The results of docking into different cluster centroids were very different. It was not possible to find a cluster that gave the best docking results for the whole set of compounds, but each compound had a simulated docking with predicted K_i not very different from the experimental result. New QSAR models, based on known data on small-molecule-inhibitors of PTP1B, were built to predict inhibitory activity of new synthetic compounds. The models have been built using software for molecular descriptors calculations. Two models are the ASNN (Associative Neural Network) models of equal quality, but the first one, built using 2D descriptors, is much easier and faster to calculate. The third model, being a linear combination of the first two, gives better results. The models can be used separately or can be combined for virtual screening aimed at the design of PTP1B inhibitors. The first two models can be used directly from the free OCHEM site, while the third one can be calculated by the user. The dataset of molecules is also available from the OCHEM site. More than 64,000 compounds were virtually screened using the novel method based on a combination of molecular docking and QSAR, which is presented in the form of rating approach for potential inhibitor, but different functions of these two methods, resulting estimations by providing potential inhibitor with predicted activity and rating position. New potential inhibitors of PTP1B, which showed *in vitro* activity with IC_{50} values in the micro-molar range, were identified. New scoring functions for molecular docking, based on the functions of AutoDock 4.2, AutoDock Vina and RF Score were developed, showing much higher efficacy (up to several times), comparing with basic functions of AutoDock 4.2 and AutoDock Vina. In addition, the novel functions give much better results for the complexes of PTP1B, than the basic ones.

Keywords: protein tyrosine phosphatase 1B, inhibitors, AutoDock, QSAR, *in silico* research, binding centers clustering, docking, structural similarities.