УДК 577.322.23+616.379−008.64: 616.85

ЗАСТОСУВАННЯ МЕТОДІВ МОЛЕКУЛЯРНОГО МОДЕЛЮВАННЯ ДЛЯ ПОШУКУ НОВИХ БІОЛОГІЧНО АКТИВНИХ РЕЧОВИН

В. В. ГУРМАЧ[1], О. М. БАЛИНСЬКИЙ[1], М. О. ПЛАТОНОВ[2], О. М. БОЙКО[1], Ю. І. ПРИЛУЦЬКИЙ[1]

[1]Київський національний університет імені Тараса Шевченка, Україна;
e-mail: gyrmach@gmail.com;
[2]Інститут молекулярної біології та генетики НАН України, Київ

Пошук нових сполук зі специфічною біологічною дією потребує використання новітніх методів молекулярного моделювання. З метою пошуку потенційно активних речовин для всього класу SH2 доменів проведено порівняння відомих структур, їх кластерний аналіз, молекулярний докінг, виділено усі можливі фармакофорні моделі та застосовано GTM передбачення. Одержані дані свідчать про значну варіативність зв'язування SH2 доменів.

Ключові слова: біологічно активні речовини, молекулярне моделювання, SH2 домени, фармакофорні моделі, GTM передбачення.

Біологічні процеси у клітинах відбуваються за участю великої кількості протеїнових макромолекул, які функціонують, зокрема, у складі протеїнових і нуклеїнових комплексів. Різноманітність протеїнових взаємодій настільки значна, що їх графічне зображення має вигляд надзвичайно складної і заплутаної мережі [1, 2]. Відтак, знання просторової будови комплексів клітинних протеїнів та їхніх лігандів є важливим кроком на шляху до розуміння механізмів функціонування.

SH2 (Src Homology 2) – компактний глобулярний домен, що бере активну участь у внутрішньоклітинній сигналізації, відіграє важливу роль як посередник специфічних протеїново-протеїнових взаємодій [3, 4]. Він складається приблизно зі 100 амінокислот, які утворюють сім β-складок і дві α-спіралі. Його активний сайт характеризується строго визначеним розміщенням групи ArgβB5, яка утворює водневий (Н-) зв'язок із двома атомами оксигену, що входять до складу pTyr (фосфотирозину) [5, 6]. Кишеня зв'язування SH2 має велику гідрофобну частину; саме це створює передумови для пошуку різноманітних селективних лігандів проти цього класу доменів, що є найбільшим класом pTyr – розпізнавальних доменів [7, 8].

Здебільшого домен SH2 міститься в онкопротеїнах (Src oncoprotein) та у протеїнах, які входять до сигнальних каскадів клітини. Геном людини кодує близько 120 доменів SH2, що входять до складу 110 протеїнів, які присутні у найрізноманітніших класах: протеїнкінази (Src, Lck), фосфатази (SHP2, SHIP2), фосфоліпази (PLCγ1), фактори транскрипції (STAT), регуляторні протеїни (SOCS), адаптери протеїнів (Grb2), структурні протеїни (SHC) та ін. Вони широко представлені в організмі тварин і майже повністю відсутні в мікроорганізмах (наприклад, примітивний фрагмент SH2 у дріжджах). Це дозволяє припустити, що їх поява пов'язана з ускладненням механізмів передачі сигналів у багатоклітинних організмах [9].

Виявлено, що домени SH2 можна поділити залежно від специфічності розпізнавання pTyr-залишку із С-кінця: таке розпізнавання відбувається у позиціях +1, +2 та +3 [5]. Можна припустити, що кожен окремий домен SH2 зв'язується лише з конкретним pTyr-фрагментом. Наприклад, Src SH2 переважно розпізнають Glu-Glu-Ile (взаємодіючий фрагмент pYEEI), у той час як Grb2 SH2 домен зв'язується з іншим фрагментом – pYVNV. Однак повне розуміння цього ефекту потребує детального вивчення термодинамічних особливостей взаємодії фосфопептидів із доменами SH2.

Мета цієї роботи – комплексно дослідити домени SH2 методами молекулярного моделювання для створення бібліотеки лігандів на весь клас цих доменів, яку в подальшому можна було б застосувати для передбачення можливих

внутрішньоклітинних порушень, пов'язаних із дією SH2 доменів, та використовувати її у терапевтичних цілях.

Матеріали і методи

Усі розрахунки проводили на розрахунковому кластері Київського національного університету імені Тараса Шевченка [10], використовували базу хімічних речовин фірми Enamine, яка містить 1,2 млн. речовин [11]. Перед початком розрахунків усі речовини наблизили до свого найімовірнішого хімічного стану у водному середовищі: для цього провели протонування основ та депротонування кислот (наприклад, аміни визначалися, як кислоти, якщо у них міститься два сильних електрон–акцепторних замісники, але, якщо присутня комбінація гідрогену із sp3 замісником карбону, аміни вважаються основами). При цьому функціональні групи з характерними значеннями pKa 7,0 (не вдалося точно визначити, чи вони мають бути протоновані, чи депротоновані) нічого не змінювали.

На наступному етапі проводили побудову 3D-структур молекул та їх енергетичну мінімізацію, використовуючи опцію «four-dimensional energy minimization». Процедуру проводили в силовому полі MMFF94x [12] за RMSD градієнта 0,01 нм у чотири етапи: (1) ініціалізація – кожному атому молекули присвоювали координати у чотиривимірному просторі (xi,yi,zi,wi); (2) 4D-мінімізація енергії та побудова 4D-конформацій; (3) мінімізація функції U(x,y,z,w) – потенціальна енергія молекули; (4) 3D-мінімізація – видаляли w (четверта координата виміру), а після цього у тривимірному просторі мінімізували функцію U(x,y,z).

Відбір мішеней для проведення докінгу. Загалом відібрали 219 структур доменів SH2, які були взяті з бази PDB (Protein Data Bank) [13]: із них 66 структур одержані методом ЯМР у розчині і 153 – методом рентгеноструктурного аналізу. Всі ці структури містять інформацію про 67 доменів, які належать до 65 протеїнів 9 видів організмів. Для найякіснішого використання структурних даних враховували, що деякі структурні файли містили декілька різних копій одного і того ж домену. Тому всі файли розбили на 1129 окремих протеїнових структур, які розділили на 8 кластерів і повторно їх порівняли, використовуючи програмний пакет Chimera [14].

Молекулярний докінг. Перед проведенням докінгу з відібраних протеїнів видалили усі молекули води, а всі Arg та Lys запротонували. Молекулярний докінг проводили у два етапи: на першому етапі користувалися програмним пакетом MOE [15], а на другому – flo+ [16]. В обох випадках використали гнучкий ліганд і фіксований рецептор. Для того аби переконатися, що докінг є якісним, застосовували обидві програми для проведення процедури редокінгу (табл. 1 і 3).

Застосування пакета MOE. На першому етапі використали стандартні параметри проведення докінгу в рамках пакета MOE [17, 18]. Для кожної молекули генерували 30 можливих варіантів зв'язування із протеїном, з яких потім відбирали лише один.

Застосування пакета flo+. Застосували алгоритм систематичного докінгу (sdock+) [19]. Максимальна кількість кроків розрахунку складала 200; 10 найкращих комплексів (виходячи з внутрішніх скоринг-функцій QXP [16])

Таблиця 1. *Результати порівняння відібраних структур: їх гомологічність, % (верхня половина таблиці) та значення RMSD (Å) між цими структурами (нижня половина таблиці)*

PDB	1o49	2fci	2ge9	3in7	2jyq	2k7a	2kk6	1uus
1o49	----------	20,2	20	92	33	29	27,6	15
2fci	2,384	----------	22,3	19,1	21,3	14,9	13,8	13,8
2ge9	2,24	3,141	----------	20,6	21	40,6	27,6	13,1
3in7	1,395	2,665	2,287	----------	30,9	30,6	28,6	13,7
2jyq	1,022	2,619	2,285	1,149	----------	23,4	26,6	17
2k7a	1,491	2,733	2,16	1,287	1,287	----------	29,6	11,9
2kk6	1,444	2,708	2,352	1,347	1,363	1,486	----------	12,2
1uus	1,568	2,683	2,817	2,044	1,632	2,066	1,842	---------

залишали для подальшого аналізу. Важливо зазначити, що у процесі розрахунків було враховано, що діапазон рухливості взаємодіючих структур може бути різним, починаючи з невеликих бічних ланцюгів і закінчуючи масштабними доменними рухами. Одержані результати відфільтрували фільтрами multyRmsd та flo+ [20]. Спочатку оцінювали енергію комплексу «протеїн-ліганд», потім кількість Н-зв'язків між атомами протеїну та ліганду (під час відбору потенційних лігандів особливу увагу приділяли молекулам із великою кількістю акцепторів Н-зв'язку) і, нарешті, оцінювали площу ліганду, що контактує з протеїном.

Побудова фармакофорних моделей. Для побудови структурно-залежних фармакофорних моделей використали всі наявні PDB-структури. З них відібрали структури, що містили ліганди (1SKJ, 1BKM (PP60 V-SRC Tyrosine kinase transforming protein), 1O49, 1O48, 1O47, 1O46, 1O44, 1O43, 1O42 (proto-oncogene tyrosine-protein kinase SRC), 1IJR, 1FBZ (proto-oncogene tyrosine-protein kinase LCK), 1A1E (C-SRC tyrosine kinase)). Після цього за допомогою пакета UCFS Chimera провели порівняння відібраних структур. Далі їх використали для побудови фармакофорних моделей за допомогою програми LigandScout [21].

Для побудови лігандзалежних фармакофорних моделей відібрали всі відомі активні та неактивні сполуки щодо доменів SH2. Загалом одержали 78 активних та 38 неактивних речовин. Для них, використовуючи пакет LigandScout Omega за внутрішньої енергії 10 ккал/моль, згенерували усі можливі стереоізомери (для кожної речовин обмежували кількість можливих конформацій 500, при цьому межа значення RMSD становила 0,4 нм). Далі провели кластеризацію наявних структур: для кожної структури відібрали 25 енергетично найвигідніших стереоізомерів.

Валідація фармакофорних моделей (генерація декоїв – структур, які мають мінімальні шанси бути активними щодо будь-якої мішені). Процедуру виконали за допомогою інтернет-ресурсу «DUDE decoy generation». Спочатку порівняли ADME (absorption, distribution, metabolism, and excretion) декоїв та структур лігандів (молекулярна маса, LogP, обертальні зв'язки, акцептори Н-зв'язку, донори Н-зв'язків і заряди). Усі декої згенеровано в діапазоні pH від 6,0 до 8,0 з використанням програмного пакета «Schrödinger's Epik», застосовуючи параметри «pH 7,0-tp 0,20 (мінімальна вірогідність створення таутомерів)». Для кожної комбінації параметрів ADME згенерували 50 можливих декоїв. Наступні декої були відібрані за допомогою пакета ZINC45 [22]. Використовували протокол динаміки, застосований до локального хімічного простору, збільшуючись або зменшуючись навколо відібраних шести властивостей. У такий спосіб одержали декої, суміжні з початковим лігандом за основними структурними характеристикам (найчастіше позитивне передбачення спостерігали за pH 7,05). Потім декої сортували за спаданням подібності (коефіцієнт танімото (Tc)) відносно будь-якого ліганда. Найбільш неподібні декої, для яких $Tc \leq 25\%$ щодо активних речовин, відкинули. Загалом для 78 активних речовин згенерували 4200 декоїв, які використали для валідації фармакофорних моделей.

ROC – аналіз. ROC (receiver operating characteristic) – як інструмент оцінки результатів скринінгу, посідає особливе місце серед бінарних алгоритмів класифікації [23]. У нашому разі за допомогою ROC-кривих визначили здатність фармакофорних моделей класифікувати речовини на активні і неактивні, розраховуючи значення AUC (Area Under Curve) [24, 25]. ROC-криві вказують на специфічність (1 - Sp) і чутливість (Se), яка змінюється залежно від якості моделі [26, 27]. Зазначимо, що переваги такої валідації нещодавно небезпідставно були піддані критиці [28].

Віртуальний скринінг. Віртуальний скринінг – один із найпотужніших методів для пошуку нових потенційно активних речовин і скефолдів [29]. Для його проведення використали програмний пакет LigandScout. Розрахунки виконали за таких функціональних параметрів програми: «Pharmacophore-Fit scoring function» (враховували лише властивості фічів фармакофорних моделей та їхні RMSD), «Match all query features» (опція для пошуку молекул, що повністю відповідають фармакофорній моделі), «Stop after first matching conformation» (для кожної молекули скринінг зупиняли після того, як була знайдена перша конформація молекули, що відповідає фармакофорній моделі).

Модель GTM (Generative topographic map). Результати, одержані на попередніх етапах

дослідження (сет активних і неактивних речовин, результати віртуального скринінгу), були застосовані для побудови моделей GTM згідно з алгоритмом, запропонованим у роботах [30, 31].

Результати та обговорення

Відбір мішеней для проведення докінгу. Базуючись на аналізі усіх наявних структур, відібрали 8 PDB структур для проведення молекулярного докінгу (табл. 1). Але, враховуючи ресурсоємність розрахунків програмним пакетом Chimera, повторно порівняли відібрані структури. Так, виявили значну подібність між структурами 3in7 і 1O49, внаслідок чого структуру 3in7 відкинули. Отже, загалом для пошуку лігандів використали 7 мішеней, а саме: 1O49 (ланцюг A, *HomoSapien*s SRC SH2 (центроїд: організм, протеїн і домен, надалі все записано в аналогічному порядку)), 2fCI (ланцюг A, *HomoSapiens*, BTK, SH2), 2GE9 (ланцюг A, *HomoSapiens*, BTK, SH2), 2JYQ (ланцюг A, *HomoSapiens*, GRB2, SH2), 2K7A (ланцюг B, *MusMusculus*, ITK, SH2), 2KK6 (ланцюг A, *HomoSapiens*, FER, SH2), 1UUS (ланцюг A, *DictyosteliumDiscoideum*, DSTA, SH2).

Молекулярний докінг. Спочатку провели редокінг на прикладі мішені 1O49. Нативний ліганд видалили з протеїну і повторно докували в сайт зв'язування. Результат вважався задовільним, якщо значення RMSD не перевищувало 2 Å [32] (табл. 2 і 3). Використовуючи пакет MOE, одержали менший діапазон значень RMSD – 1–1,5 Å. Однак найкращий варіант зв'язування (табл. 3) отримали у програмі flo+. Тому ми спочатку докували з використанням програми MOE, а потім одержані результати використали в роботі з пакетом flo+. Крім того, порівнювали енергетичні параметри

Таблиця 2. Результати редокінгу в програмному пакеті MOE

Положення	RMSD, Å	E_conf, ккал/моль
1	1,11	4,51
2	1,71	1,84
3	1,69	3,01
4	1,65	1,44
5	1,55	1,24
Initial	0,00	3,05

конформацій ліганду (табл. 2). Далі провели редокінг за допомогою програми flo+ (табл. 3). Оцінювали такі параметри: FreE (загальна вільна енергія комплексоутворення), Cntc (контактна енергія зв'язування між усіма молекулами ліганду і протеїну), Hbnd (енергія Н-зв'язків між протеїном і лігандом), Bump (енергія стеричних зіткнень), Intl (енергія напруженості ліганду). Виявилося, що одержані результати можна розділити на три групи: до першої групи відносяться положення (1, 2), які характеризуються практично у всіх випадках мінімальними енергетичними параметрами; до другої групи – положення (3, 4), які мають характеристики, подібні до нативної форми зв'язування; до третьої групи можна віднести положення 5, яке порівняно з усіма іншими має найгірші характеристики. Як видно (табл. 2) у будь-якому положенні є декілька практично ідентичних параметрів щодо нативної форми. Винятком можна назвати лише положення 1, але у цьому разі має місце найкраще значення RMSD, що є одним із основних параметрів оцінки якості редокінгу.

Таблиця 3. Результати редокінгу в програмному пакеті flo+

Положення	RMSD, Å	FreE, кДж/моль	Cntc, кДж/моль	Hbnd, кДж/моль	Bump, кДж/моль	Intl, кДж/моль
1	0,64	-29,6	-81,8	-17,7	3,1	15,7
2	2,71	-25	-77,5	-18,2	4,4	15,5
3	2,86	-17,9	-76,6	-17,8	4,5	17,1
4	3,83	-18,1	-75,4	-18,3	4,8	16,5
5	1,04	-23,4	-73,8	-11,4	3,5	12,6
Initial	0,00	-17,5	-77,0	-14,9	4,3	19,9

Після редокінгу провели докінг. Так, за застосування пакета MOE відібрали 150 тис. речовин для кожного кластера. Ці речовини використали на наступному етапі розрахунків у рамках програмного пакета flo+. З усіх PDB структур повністю видалили воду і провели протонування Arg та Lys (ці залишки у протонізованій катіонній формі здатні утворювати велику кількість водневих зв'язків). Завдяки відмінності орієнтації pTyr-зв'язуючого сайта і гідрофобної кишені для кожного кластера застосували окремі правила відбору лігандів. Так, загалом докували майже 500 тис. структур у всі мішені і одержали понад 10 млн. комплексів протеїн-ліганд. Завдяки фільтрації скоротили кількість комплексів до 50 тис., які аналізували візуально. Кінцевий сет речовин становив 10463. Зазначимо, що структури, одержані щодо різних мішеней, значно відрізняються одна від одної. Так, майже з десяти тисяч перекриваються лише 705 речовин. Ці моделі зв'язування повністю перекрилися з моделями, які були одержані на тестовому варіанті розрахунків (наведені в роботі [33]).

Фармакофорні моделі. Під час порівняння сайтів зв'язування відібраних структур та їхніх лігандів виявили значну подібність як перших (табл. 2), так і других. Для порівняння використовували програмний пакет UCFS Chimera. Розглянули два варіанти порівняння структур: у першому із застосуванням алгоритму Нідлмана-Вулша та матриці порівнянь BLOSUM-62 [34, 35] (усі структури порівнювали з однією шаблонною структурою (1O49)) – значення RMSD не перевищувало 1 Å, що є підтвердженням просторової подібності цих структур (табл. 4); у другому випадку, всі структури порівнювали між собою. Одержали: Q-score 0,525 і середнє значення RMSD 0,860 Å. Враховуючи ці результати, на основі відібраних PDB-структур побудували одну загальну фармакофорну модель (рис. 1).

Беручи до уваги вищезазначені дані, порівняли механізм зв'язування відібраних PDB структур (рис. 1, верхній). Так, у всіх випадках зв'язування включає такі ключові моменти (рис. 1): (1) активна частина кишені, яка зв'язується з pTyr/карбоновою кислотою і характеризується великою кількістю донорів H-зв'язку (Arg, Lys); (2) безпосередньо біля сайта зв'язування pTyr знаходиться ароматична/гідрофобна частина ліганду, яка обумовлена проявом ван-дер-вальсових взаємодій, здебільшого між Arg і Lys; (3) центр кишені зв'язування визначається позицією O амінокислотного залишку другої амінокислоти (His), який під час зв'язування утворює водневий зв'язок із групою NH пептидного ліганду; (4) наявність однієї або двох гідрофобних кишень, які знаходяться за центром активного сайта і, зазвичай, заповнені гідрофобною частиною ліганду. Виходячи з цього, побудували одну залежну від структури фармакофорну модель, представлену на рис. 1.

Для побудови лігандзалежних фармакофорних моделей використали усі наявні активні речовини. Для оцінки і категоризації хімічних речовин застосували кластерний аналіз. Як результат, одержали 27 кластерів і 9 сінгелтонів, з яких у подальшому побудували 8 фармакофорних моделей. За порівняння одержаних моделей (табл. 5) і моделі, представленої на рис. 1, виявили, що практично всі ці моделі характеризуються такими властивостями: наявність акцептора H-зв'язку (місцеположення 1, рис. 1); поряд знаходяться гідрофобна або ароматична частина ліганду (місцеположення 2, рис. 1) та гідрофобна частина ліганду (місцеположення 4, рис. 1).

Після побудови фармакофорних моделей звернули увагу на те, що, не зважаючи на значні конформаційні обмеження у разі генерації всіх можливих стереоізомерів, речовини могли знач-

Таблиця 4. Результати порівняння відібраних структур зі структурою 1O49, яку використовували в докінгу

PDB	1A1E	1BKM	1FBZ	1IJR	1SKJ	1O42	1O43	1O44	1O46	1O47	1O48
RMSD, Å	0,763	0,651	0,808	0,84	0,696	0,234	0,219	0,192	0,212	0,210	0,161
Sequence alignment score	528,2	518	334,1	336,2	515	551,2	547,6	551,2	551,2	551,2	547,6

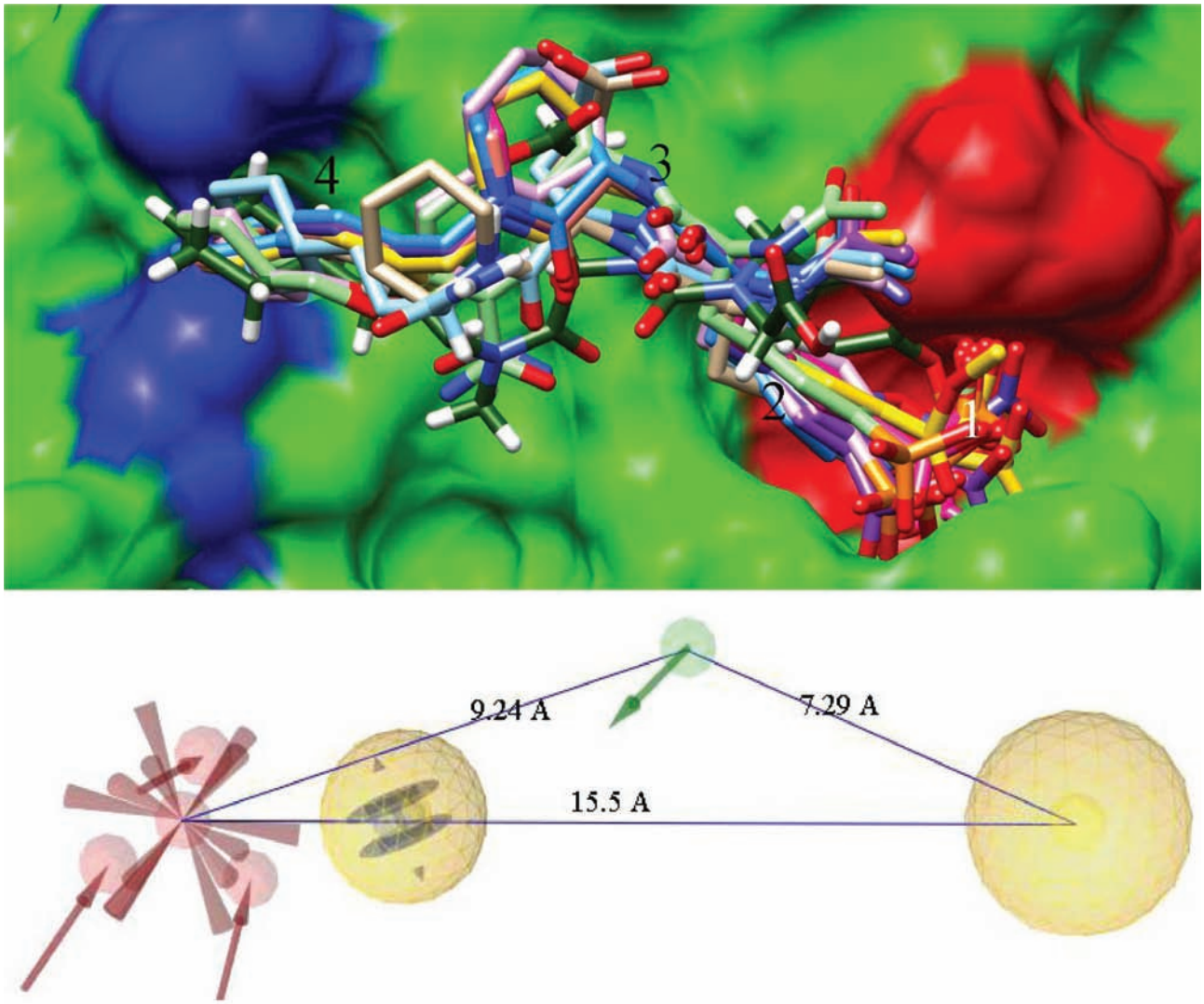
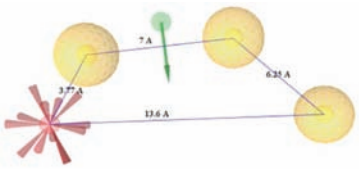
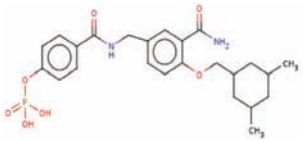
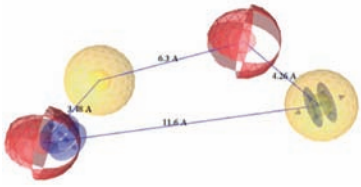
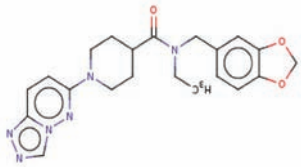
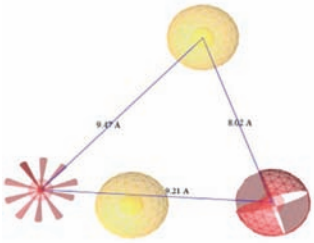
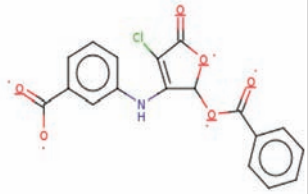
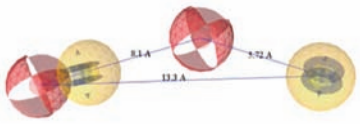
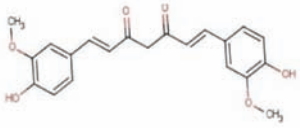
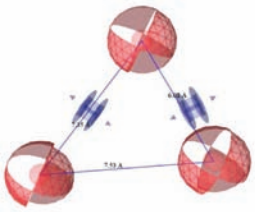
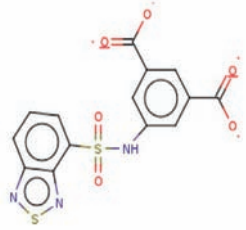
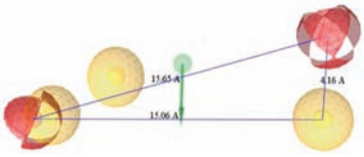
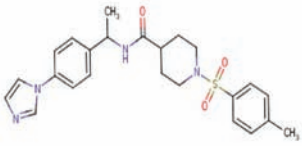
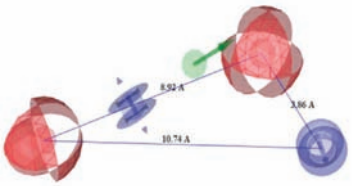
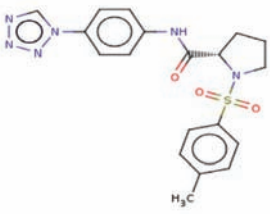


Рис. 1. Накладання усіх наявних кристалів SH2 доменів з лігандами (верхня частина рисунку: 1A1E – сірий, 1BKM – блакитний, 1FBZ – рожевий, 1IJR – світло-зелений, 1SKJ – темно-зелений, 1O42 – помаранчевий, 1O43 – білий, 1O44 – темно-рожевий, 1O46 – жовтий, 1O47 – синій, 1O48 – пурпуровий, 1O49 – фіолетовий). Також наведено чотири основні місцеположення зв'язування (1 2, 3 і 4), на основі яких була побудована загальна фармакофорна модель (нижня частина рисунка)

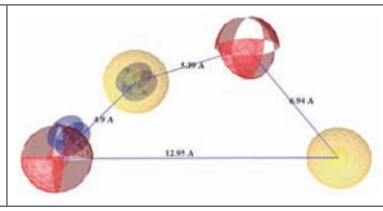
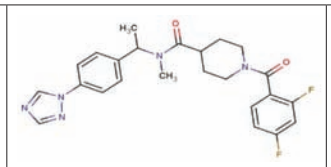
но вигинатися. Для вирішення цієї проблеми логічно було б провести вирівнювання усіх структур кластера за максимально витягнутою структурою (у нашому разі це було б від 10,74 до 15,06 Å). Але, враховуючи те, що одержали лише 2 (№ 3, 5) моделі з 8, які були побудовані саме на таких речовинах, ми такого вирівнювання не проводили. До того ж, як правило, такі вигини не є випадковим явищем. Отже, ця модель може відображати інший механізм блокування доменів SH2 (на жаль, не існує протеїнових структур у комплексі зі структурами, з яких побудовані моделі № 3, 5). Усі інші фармакофорні моделі (№ 1, 2, 4, 6-8) незначно відрізняються одна від одної. Ці моделі характеризуються наявністю чотирьох основних фічів (описаних вище як механізм зв'язування ліганду з протеїном) за винятком того, що в моделях № 2, 4 і 8 донорна група в положенні 3 (рис. 1) замінена на акцепторну (це може бути пов'язано з тим, що, зазвичай, поряд із O His знаходиться група NH Tyr і, можливо, у цьому разі утворення H-зв'язку відбувається саме з цією амінокислотою).

Валідація фармакофорних моделей. В рамках ROC-аналізу визначали якість розподілу речовин на активні та неактивні, використо-

Таблиця 5. Моделі, побудовані на основі інгібіторів SH2 доменів. Результати їх валідації і віртуального скринінгу

№	Фармакофорна модель	Репризентативна речовина	AUC	EF	THR, %	FHR, %	Хіти	
							Тест сет	Декої
1			1,00	559	100	0,12	90	5
2			0,92	318	83	0,14	159	6
3			1,00	54,5	100	1,79	400	75
4			0,75	39,6	50	1,23	604	52
5			0,75	91,2	50	0,59	216	25
6			1,00	233	100	0,33	129	14
7			0,75	110	100	0,35	334	15

Таблиця 5. Продовження

8			1,00	209	100	0,41	522	17

вуючи значення AUC, загальну специфічність, positive rate (TPR), false positive rate (FPR) [24]. Видно (табл. 5), що моделі 1, 3, 6, 8 мають кращі параметри, якщо судити за загальною площею під ROC-кривою, а саме 100%; трохи гіршою є модель 2 та найгіршими є моделі 4, 5 і 7 – 75%. Моделі 1, 2, 6 і 8 з максимальною якістю відбирають наявні активні речовини і відсіюють декої; модель 3 відбирає речовини, надані тренінг-сетом, але при цьому погано відкидає декої; нарешті, абсолютно протилежна ситуація має місце з моделлю 5. З одержаних даних можна зробити висновок, що практично усі моделі доповнюють одна одну, частина з них відбирає усі активні речовини, а інша відкидає більшість декоїв. Якщо порівняти ці фармакофорні моделі з активним сайтом протеїну, тоді майже усі вони можуть добре вписатися в його геометрію. За винятком моделей 3 і 5, усі інші в загальних рисах відповідають основній фармакофорній моделі з незначними змінами.

Віртуальний скринінг. Усі фармакофорні моделі використали для проведення віртуального скринінгу на результатах докінгу (понад 10 тис. речовин). Загалом, після віртуального скринінгу одержали 1816 активних речовин (результати для кожного кластера наведено в табл. 5). Ці речовини у подальшому аналізували за допомогою методу GTM.

Побудова моделі GTM. Моделі GTM будували на основі 78 активних і 43 неактивних речовин. Використовуючи програму ISDA Fragmentor, відібрали дискриптори (у нашому разі фрагменти речовин) у діапазоні від 2 до 8 атомів та хімічних зв'язків. Одержали 4155 дискрипторів і на їх основі побудували моделі GTM, з яких відібрали декілька найвдаліших. Розрахунки проводили за всіма можливими комбінаціями параметрів k (number R of RBF) (25), m (the grid resolution) (5), w (the RBF width) (0,25, 0,5, 0,75, 1, 1,25, 1,5, 1,75, 2), l (the weight regularization coefficient) (0,01, 0,1, 1, 10, 100) з використанням крос-валідації (для побудови моделі за кожною комбінацією параметрів трейнінг-сет розділяли на три частини, дві з яких використовували для побудови моделі, а одну – для її валідації). Найкращі параметри розрахунків наведено в табл. 5, 6.

Зазначимо, що оскільки не всі наявні активні речовини були використані для побудови фармакофорних моделей у програмі LigandScout, вони були повторно перекластеризовані для оцінки їх взаємного розміщення різних активних речовин одна відносно одної. Як результат, на GTM-картах отримали 7 (не враховуючи неактивні речовини) кластерів, які істотно відрізнялися один від одного (середній Tc 0,22) і характеризувалися значною подібністю в межах кожного окремого кластера (середній Tc 0,92). Це такі сполуки, як сульфіди, карбонові кислоти, фосфати і фосфанати, паразаміщені арілазоли та інші. Беручи до уваги фармакофорну модель, наведену вище, таке різноманіття наявних активних речовин свідчить про значну структурну варіативність сайта зв'язування SH2 доменів. Доказом цього є також і те, що загалом нам вдалося виділити

Таблиця 6. Розрахункові параметри для GTM моделі

№	k	m	w	l	Likelihood	Delta	Bac (kNNode)
1	25	5	0,75	10	143,4980	65,16	0,7046
18	25	5	0,25	0.01	170,8029	88,54	0,6577
32	25	5	1	10	131,1180	58,44	0,6307
37	25	5	0,25	10	162,1306	79,30	0,678

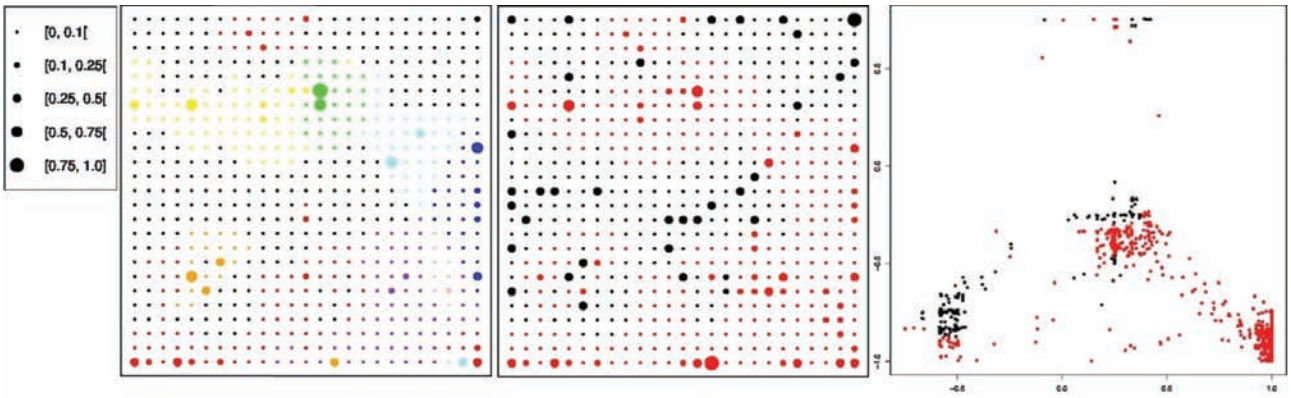


Рис. 2. GTM візуалізація одержаної моделі. Карта зліва - GTM модель з урахуванням розбиття усіх активних речовин на різні кластери: чорний колір – кластер 1 (неактивні речовини (не враховували за наведення вищедосліджуваних кластерів)); червоний колір – кластер 2 (сінгелтоно, речовини, які не піддались кластеризації); синій колір – кластер 3 (дигідрокси/метокси-фенілаланіни); зелений колір – кластер 4 (тріазолпірадозіни-6-піридіни); помаранчевий колір – кластер 5 (арілсульфаміди та арілсульфони); пурпурнйи колір – кластер 6 (карбонові кислоти); блакитний колір – кластер 7 (фосфати і фосфанати); зелений колір – кластер 8 (паразаміщені арілазоли). Карта в центрі – GTM модель без врахування розбиття на кластери (червоний – активні, чорний – неактивні речовини). Карта справа – результат передбачення в рамках GTM моделі

7 доменів кластерів, в яких гомологічність не перевищує 30%. Крім того, важливо зазначити, що ця фармакофорна модель будувалася на незначній кількості структур, що, в свою чергу, вказує на можливість існування інших варіантів зв'язування доменів SH2.

Не зважаючи на незначну подібність одержаних кластерів між собою, спостерігається їхнє часткове перекриття. Так, усі наявні кластери мають перекриття з неактивними речовинами (максимальне значення – Tc = 0,5572275714, середнє – Tc = 0,264242635 і мінімальне – Tc = 0,1166415714) (рис. 2). Внаслідок використання одержаної моделі передбачення мало місце лише за кластерами 2, 5, 6 і частково 7. Отже, кінцева бібліотека складалася з 1607 речовин, бо 273 речовини було відкинуто.

Таким чином, базуючись на сучасних методах молекулярного моделювання (молекулярний докінг, фармакофорне моделювання, GTM передбачення), знайдено нові речовини, здатні зв'язуватись з доменами SH2. Зокрема, для ідентифікації основних фічів взаємодії ліганд-домен SH2 після проведення докінгу побудували фармакофорні моделі, що залежать від ліганду та структури. Загалом знайдено 9 фармакофорних моделей, що виражають три варіанти зв'язування доменів SH2 з лігандом. Валідація показала, що всі одержані фармакофори можуть бути застосовані для пошуку потенційних лігандів доменів SH2. Значення величини AUC в отриманих моделях знаходиться в діапазоні від 75 до 100%. Одна частина моделей краще відбирає активні речовини і відсіює декої (моделі 1, 2, 3, 6, 8 визначені як найкращі), а інша (моделі 4, 5, 7) – краще відсіює декої або відбирає активні речовини. Тому припускаємо, що ці моделі доповнюють одна одну. Підтвердженням цього є те, що частина лігандзалежних моделей (1, 2, 4, 6, 7, 8) комплементарна активному сайту протеїнів із наявним нативним лігандом, інша (3, 5) – має значні відмінності, а, отже, ці моделі характеризують іншу геометрію сайта зв'язування домену SH2. Як наслідок, відібрано 1816 речовин, які застосували у GTM передбаченні і одержано 1607 речовин, що відповідають основним точкам взаємодії, застосованих під час докінгу та фармакофорного пошуку.

ПРИМЕНЕНИЕ МЕТОДОВ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ ДЛЯ ПОИСКА НОВЫХ БИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ

В. В. Гурмач[1], А. М. Балинский[1], М. О. Платонов[2], А. Н. Бойко[1], Ю. И. Прилуцкий[1]

[1]Киевский национальный университет имени Тараса Шевченко, Украина;
e-mail: gyrmach@gmail.com;
[2]Институт молекулярной биологии и генетики НАН Украины, Киев

Поиск новых соединений со специфическим биологическим действием требует использования современных методов молекулярного моделирования. С целью поиска потенциально активных веществ для всего класса SH2 доменов проведено сравнение всех известных структур, их кластерный анализ, молекулярный докинг, выделены все возможные фармакофорные модели и применено GTM предсказание. Полученные данные свидетельствуют о значительной вариативности связывания SH2 доменов.

Ключевые слова: биологически активные вещества, молекулярное моделирование, SH2 домены, фармакофорные модели, GTM предсказание.

APPLICATION OF THE METHODS OF MOLECULAR MODELING TO THE SEARCH FOR NEW BIOLOGICALLY ACTIVE SUBSTANCES

V. V. Hurmach[1], O. M. Balinskyi[1], M. O. Platonov[2], O. M. Boyko[1], Yu. I. Prylutskyy[1]

[1]Taras Shevchenko National University of Kyiv, Ukraine;
e-mail: gyrmach@gmail.com;
[2]Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kyiv

The searching for new chemical compounds possessing specific biological activity is a complex problem that needs the usage of modern methods of molecular modeling. In particular, for the prupose of searching for potentially active compounds for whole class of SH2 domains, a comparison of all available structures, their cluster analysis, molecular docking, selection of all possible pharmacophore models and GTM prediction were done. Obtained results testify to the considerable variability of binding of SH2 domains.

Key words: biologically active compounds, molecular modeling, SH2 domains, pharmacophore models, GTM prediction.

References

1. Claverie J. M. Gene number. What if there are only 30,000 human genes? *Science.* 2001;291:1255-1257.
2. Alm E., Arkin A. P. Biological networks. *Curr. Opin. Struct. Biol.* 2003;13:193-202.
3. Koch C. A., Anderson D., Moran M. F., Ellis C., Pawson T. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science.* 1991;252:668-674.
4. Liu B. A., Jablonowski K., Raina M., Arcé M., Pawson T., Nash P. D. The human and mouse complement resource of SH2 domain proteins–establishing the boundaries of phosphotyrosine signaling. *Mol. Cell.* 2006;22:851-868.
5. Dominguez C., Boelens R., Bonvin A., M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Society.* 2003;125:1731-1737.
6. Brovarets O. O., Yurenkoc Y. P., Hovorun D. M. The significant role of the intermolecular CH···O/N hydrogen bonds in governing the biologically important pairs of the DNA and RNA modified bases: a comprehensive theoretical investigation. *J. Biomol. Struct. Dyn.* 2014:1-29.
7. Pawson T., Gish G. D., Nash P. SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol.* 2001;11:504-511.
8. Huang H., Li L., Wu C. Defining the specificity space of the human SRC homology 2 domain. *Mol. Cell. Proteomics.* 2008;7:768-784.
9. Silva C. M. Role of STATs as downstream signal transducers in Src family kinase-mediated tumorigenesis. *Oncogene.* 2004;23:8017-8023.
10. Boyko Y. V., Vystoropsky O. O., Nychyporuk T. V., Sudakov O. O. Kyiv National Taras Shevchenko University High Performance Computing Cluster. Third International Young Scientists Conference on Applied Physics. 2003. P. 189-181.
11. Chuprina A., Lukin O., Demoiseaux R., Buzko A., Shivanyuk A. Drug- and lead-likeness, target class, and molecular diversity

analysis of 7.9 million commercially available organic compounds provided by 29 suppliers. *J. Chem. Inf. Model.* 2010;50:470-479.
12. Thomas A. H. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comp. Chem.* 1996;17:490-519.
13. Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne, P. E. The protein databank. *Nucleic Acids Res.* 2000;28:235-242.
14. Balinskyi O. M., Sudakov O. O., Platonov M. O. Small-molecule ligand search for protein interaction domains involving shape-based molecular modeling methods. *Bulletin of Taras Shevchenko National University of Kyiv. Series Physics & Mathematics.* 2012;3:287-291.
15. Benini S., Gessa C., Ciurli S. *Bacillus pasteurii* urease: a heteropolymeric enzyme with a binuclear nickel active site. *Soil Biol. Biochem.* 1996;28:819-821.
16. Mcmartin C., Bohacek R. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aid. Mol. Design.* 1997;11:333-344.
17. Corbeil C. R., Williams C. I., Labute P. Variability in docking success rates due to dataset preparation. *J. Comp.-Aided Mol. Design.* 2012;26:775-786.
18. Chang D. T., Oyang Y. J., Lin J. H. MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. *Nucleic Acids Res.* 2005;33:233-238.
19. Warren G. L., Andrews C. W., Capelli A. M., Clarke B., LaLonde J., Lambert M. H., Lindvall M., Nevins N., Semus S. F., Senger S., Tedesco G., Wall I. D., Woolven J. M., Peishoff C. E., Head M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 2006;49:5912-5931.
20. Sudakov O. O., Balinskyi O. M., Platonov M. O., Kovalskyy D. B. Geometric filters for protein–ligand complexes based on phenomenological molecular models. *Biopolym. Cell.* 2013;29:395-400.
21. Wolber G., Langer T., J. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Chem. Info. Comput. Sci.* 2005;45:160-169.
22. Irwin J. J., Shoichet B. K. ZINC a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 2005;45:177-182.
23. Braga1 C., Carolina H. Andradel Assessing the Performance of 3D Pharmacophore Models in Virtual Screening: How Good are They? Rodolpho. *Curr. Top. Med. Chem.* 2013;13:1127-1238.
24. Verdonk M. L., Berdini V., Hartshorn M. J., Murray C. W., Taylor R. D., Watson J. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Model.* 2004:44:793-806.
25. Kirchmair J., Markt P., Distinto S., Wolber G., Langer T. J. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *Comput.-Aid. Mol. Design.* 2008;22:213-228.
26. Sonego P.; Kocsor A.; Pongor S. ROC analysis: applications to the classification of biological sequences and 3D structures. *Brief. Bioinformatics.* 2008;9:198-209.
27. Robin X., Turck N., Hainard A., Tiberti N., Lisacek F., Sanchez J. C., Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:1471-2105.
28. Truchon, J. F.; Bayly, C. I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* 2007;47:488-508.
29. Sakkiah S., Thangapandian S., John Y. J., Kwon K. W. Ligand and structure based pharmacophore modeling to facilitate novel histone deacetylase 8 inhibitor design. *Eur. J. Med. Chem.* 2010;45:4409-4417.
30. Héléna A. G., Gilles M., Dragos H., Alban A., Sylvain L., Philippe V., Alexandre V. Generative Topographic Mapping-Based Classification Models and Their Applicability Domain: Application to the Biopharmaceutics Drug Disposition Classification System (BDDCS). *J. Chem. Inf. Model.* 2013;53:3318-3325.
31. Kireeva N., Baskin I. I., Gaspar H. A., Horvath D., Marcou G., Varnek A. Generative Topographic Mapping (GTM): Universal Tool for Data Visualization, Structure-Activity Modeling and Dataset Comparison. *Mol. Inf.* 2012;31:301-312.

32. Abdul W., Shandana A., Muhammad R., Syed B. J., Taj Ur R., Sahib G., Mukhtiar H. Pakhtunkhwa Molecular Docking Study of 5-substituted-8-methyl-2H-pyrido [1, 2-a] pyrimidine-2, 4 (3H) – diones As Inhibitors of Basilluspasteurii urease. *J. Life Sci.* 2013;1:145-153.
33. Hurmach V. V., Balinskyi O. M., Platonov M. O., Boyko O. M., Borysko P. O., Prylutskyy Yu. I. Design of potentially active ligands for SH2 domains by molecular modeling methods. *Biopolymers and Cell.* 2014;30:321-325.
34. Needleman S. B.; Christian D. W. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 1970;48:443-453.
35. Angana C., Sanghamitra B. FOGSAA: Fast Optimal Global Sequence Alignment Algorithm. *Scie. Rep.* 2013;1746:1-9.

Отримано 01.10.2014