МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

На правах рукопису

ГУРМАЧ ВАСИЛЬ ВАСИЛЬОВИЧ

УДК 577.322.23+616.379−008.64: 616.85

ПОШУК in Silico ЛІГАНДІВ ДЛЯ ДОМЕНІВ SH2

03.00.02-біофізика

Дисертація на здобуття наукового ступеня
кандидата біологічних наук

Науковий керівник:
Прилуцький Юрій Іванович
доктор фізико-математичних наук,
професор

Київ–2015

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ 5
ВСТУП 6
РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ 12
1.1. Протеїн-протеїнова взаємодія 12
1.2. Домени SH2 14
1.2.1. Будова та функції доменів SH2 14
1.2.2. Компоненти доменів SH2 людини 17
1.2.3. Неохарактеризовані/слабко охарактеризовані домени SH2 19
1.2.4. Дефекти, пов’язані з доменами SH2 20
1.2.5. Структурне та функціональне різноманіття доменів SH2 22
1.2.6. Фосфотирозин-незалежні взаємодії доменів SH2 26
1.2.7. Класифікація доменів SH2 27
1.3. *In silico* методи дослідження 32
1.3.1. Вирівнювання амінокислотних послідовностей протеїнів 32
1.3.2. Вирівнювання структур протеїнів 34
1.3.3. Розпізнавання функціональних сайтів 35
1.3.4. Молекулярний докінг 35
1.3.5. Алгоритми молекулярного докінгу 37
1.4. Узагальнення до огляду літератури 39
РОЗДІЛ 2. МАТЕРІАЛИ ТА МЕТОДИ 40
2.1. Апаратне та програмне забезпечення 40
2.2. Отримання 3D структур доменів SH2 та їх аналіз 40
2.2.1. Отримання сету 3D структур доменів SH2 40
2.2.2. Порівнювання отриманих структур доменів SH2 та їх поділ на групи за подібністю амінокислотних послідовностей 41
2.2.3. Конформаційний пошук доменів SH2 42
2.2.4. Кластеризація доменів SH2 методом k-середніх 43

2.2.5. Аналіз вирівнюваних амінокислотних позицій та одержаних груп 45
2.2.6. Визначення оточення сайту зв'язування доменів SH2 46
2.3. Підготовка бази хімічних речовин до проведення молекулярного докінгу 47
2.3.1. Відбір бібліотеки потенційних лігандів 47
2.3.2. Підготовка отриманих баз хімічних речовин до моделювання 49
2.4. Моделювання комплексів типу "домен SH2-ліганд" 51
2.4.1. Застосування пакету MOE 52
2.4.2. Застосування пакету flo+ 53
2.4.3. Побудова фармакофорних моделей 54
2.4.4. Генерація декоїв 55
2.4.5. ROC – аналіз 55
2.4.6. Віртуальний скринінг 56
2.4.7. Модель GTM 57
2.5. Біологічне тестування отриманих структур 57
2.5.1. Оцінка впливу відібраних речовин на фосфорилювання STAT3 57
2.5.2. Оцінка проліферативної активності різних клітинних ліній за дії досліджуваних речовин 58
2.5.3. Визначення кіназної активності Btk та Syk кіназ 59
РОЗДІЛ 3. РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ ТА ЇХ ОБГОВОРЕННЯ 60
3.1. Структурний аналіз доменів SH2 60
3.1.1. Відбір репрезентативних структур та їх розподіл на групи 60
3.1.2. Характеристика доменів SH2 в межах кожної групи 63
3.1.3. 3D оточення сайту зв'язування 69
3.1.4. Конформаційний пошук для доменів SH2 72
3.2. Моделювання комплексу "домен SH2-ліганд" 78
3.2.1. Молекулярний докінг доменів SH2 78
3.2.2. Фармакофорні моделі 87

3.2.3. Валідація фармакофорних моделей 91
3.2.4. Віртуальний скринінг та побудова моделей GTM 92
3.3. Експериментальна перевірка 94
3.3.1. Оцінка впливу відібраних речовин на фосфорилювання домену SH2 протеїну STAT3 та проліферацію пухлинних клітин 94
3.3.2. Дія відібраних речовин на кіназну активность Btk та Syk кіназ 101
РОЗДІЛ 4. УЗАГАЛЬНЕННЯ РЕЗУЛЬТАТІВ ДОСЛІДЖЕНЬ 103
ВИСНОВКИ 112
СПИСОК ЛІТЕРАТУРИ 114

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

PPI	протеїн-протеїнова взаємодія (protein-protein interaction)
SH2	родина доменів PPI (src homology 2)
ЯМР	ядерний магнітний резонанс
PDB	база даних протеїнів (Protein Data Bank)
Pfam	база родин протеїнів (Protein families database)
Btk	тирозинкіназа Брутона (Bruton's tyrosine kinase)
Syk	нерецепторна тирозинкіназа (Spleen tyrosine kinase)
QXP	швидкий пошук (Quick eXPlore)
GTM	генеративні топографічні карти (Generative topographic map)
STAT	сигнальний протеїн та активатор транскрипції (signal transducer and activator of transcription)
pTyr	фосфотирозин (phosphorylated tyrosine)
PTK	протеїн тирозин кінази (protein tyrosine kinase)
ADME	поглинання, розподіл, метаболізм, виділення (absorption, distribution, metabolism, excretion)
ID	ідентифікаційний номер (identification number)
Rmsd	середньоквадратичне відхилення від початкової координати атома (Root-mean-square deviation of atomic positions)
ASA	доступна площа поверхні (accessible surface area)
ROC	робоча характеристика сигналу (receiver operating characteristic)
AUC	площа під кривою (area under curve)

ВСТУП

Актуальність теми. Сукупність протеїнових структур формує складну мережу взаємодій, під час яких відбувається передача сигналів як всередині клітини, так і між ними [1]. Протеїн-протеїнова взаємодія (PPI) – це постійний або тимчасовий фізико-хімічний вплив протеїнових структур одна на одну (утворення/руйнування високомолекулярних комплексів). Ця взаємодія відбувається завдяки різноманіттю протеїнових доменів [2], які входять до складу протеїнових структур, характеризуються компактністю [3] та широким спектром функцій [4]. Так, наприклад, домени SH2 входять до складу протеїнкіназ (Src, Lck), фосфатаз (SHP2, SHIP2), фосфоліпаз (PLCγ1), факторів транскрипції (STAT), регуляторних (SOCS), адаптерних (Grb2) і структурних протеїнів (SHC). Враховуючи таку розповсюдженість протеїнових доменів, їх розглядають як зручні терапевтичні мішені [5].

Водночас, пошук *in silico* нових сполук зі специфічною біологічною дією є складним завданням, яке потребує використання сучасних методів молекулярного моделювання [6], достатньої кількості структурних і біологічних даних щодо мішені (підтверджені активні та неактивні речовини, відомі комплекси протеїну мішені та ліганду).

SH2 – компактний глобулярний домен, що містить близько 100 амінокислот, які формують дві α – спіралі (αA, αB) та сім β – складок (βG, βB, βC, βD, βD', βE і βF). Його основна функція - взаємодія зі структурами, що містять фосфорильований залишок тирозину (pTyr). Здебільшого домени SH2 містяться в онкопротеїнах (Src oncoprotein) та протеїнах, які входять до сигнальних каскадів клітини. Встановлено зв’язок доменів SH2 з процесами, пов’язаними з онкозахворюваннями [7]: Наприклад, порушення в їх функціонуванні викликають гострий лімфобластний [8], хронічний мієлоцитарний [9] та ювенальний мієломоноцитарний [10] лейкози тощо. Важливо, що для цієї родини доменів відомі зв’язані з лігандом або іншими

протеїнами структури (наприклад, PDB – 1O49, 2GE9). Саме тому поміж великої кількості доменів PPI, як мішень, обрали домени SH2.

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

Наразі відомо близько 10 низькомолекулярних структур, здатних інгібувати домени SH2. Враховуючи таку їх незначну кількість, пошук нових лігандів є важливим завданням, вирішення якого сприятиме створенню/виявленню нових хемотипів як елементів ефективних лікарських засобів для профілактики хвороб, пов'язаних з порушеннями функціонування доменів SH2.

Застосування комп'ютерних методів для пошуку низькомолекулярних лігандів на основі визначених тривимірних структур передбачає відбір сету потенційно активних лігандів щодо тієї чи іншої мішені. У роботі застосовано комплекс методів комп'ютерного моделювання для створення універсальної цільової бібліотеки потенційних інгібіторів доменів SH2. Тобто, отриманий сет речовин можна застосувати для пошуку активних речовин для будь-якого домену SH2 під час біологічного тестування. Активність і специфічність отриманих активних сполук з такого сету буде підвищуватись шляхом їх модифікування відносно функціональних особливостей конкретного домену SH2. Нарешті, одержані результати комп'ютерного моделювання потребують перевірки на різних біологічних моделях.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконувалась в межах наукової тематики ННЦ «Інститут біології» за номером держреєстрації 0111U004648 «Механізми реалізації адаптаційно-компенсаторних реакцій організму за умов розвитку різних патологій» (№ 11 БФ-036-01).

Мета і завдання дослідження. Мета роботи полягає у створенні цільової бібліотеки потенційних лігандів доменів SH2 та їх перевірці на біологічних моделях.

Для досягнення цієї мети були поставлені такі наукові завдання:

1. використовуючи наявні бази даних структур доменів SH2 та активних речовин щодо них, провести розподіл цих доменів на групи/кластери, виділити шаблонні структури та відібрати репрезентативний сет речовин з наявного пулу низькомолекулярних потенційних лігандів;

2. оцінити геометричні характеристики відібраних доменів SH2 та порівняти їх;

3. виділити ключові точки зв'язування доменів SH2 з нативними лігандами;

4. створити цільову бібліотеку потенційних лігандів доменів SH2;

5. провести експериментальну перевірку отриманих сполук щодо їх впливу на проліферацію пухлинних ліній клітин K562, U937, A431 і LNCap та кіназну активність Btk та Syk.

Об'єкт дослідження – домени SH2, низькомолекулярні ліганди.

Предмет дослідження – фізико-хімічні властивості кишені зв'язування доменів SH2, структурні особливості активних речовин щодо доменів SH2, нековалентна взаємодія комплексу "домен-ліганд".

Методи дослідження – попарне та множинне порівняння амінокислотних послідовностей; метод аналізу якості вирівнювання амінокислотних послідовностей (оцінка консервативності та ідентичності); метод k-середніх; метод геометричного аналізу (значення величини Rmsd

між структурами, які порівнюються, та доступність взаємодіючих поверхонь); чотиривимірна енергетична мінімізація структур; гнучкий докінг; побудова прямих/структурно-залежних та непрямих/ліганд-залежних фармакофорних моделей; метод GTM (генерування топографічних карт); оцінка впливу отриманих сполук на фосфорилювання протеїну STAT3, проліферацію пухлинних клітин шляхом детектуванням флуоресценції за використання резазурину та кіназну активность Btk та Syk шляхом детектування люмінесценції АДФ.

Наукова новизна отриманих результатів.

1. Усі наявні структури доменів SH2 поділено на групи за одновимірною та тривимірною подібністю.

2. Виділено основні структурні частини доменів SH2, які беруть участь у зв'язувані з лігандом; серед них виявлено найбільш та найменш гнучкі.

3. Запропоновано оригінальний алгоритм застосування методів ком'ютерного моделювання для створення цільової бібліотеки потенційних лігандів доменів SH2.

4. Побудовано фармакофорні моделі доменів SH2.

5. Одержано нові хемотипи, зокрема бензолсульфонамід та піримідин, які потенційно можуть виступати замінниками pTyr - частини пептидних лігандів при зв'язуванні з доменами SH2.

6. Встановлено речовини, які впливають на проліферацію пухлинних клітин ліній K562, U937, A431 та LNCap, проявляють інгібуючий ефект щодо домену SH2 протеїну STAT3 і здатні специфічно інгібувати кіназну активність Btk та Syk.

Практичне значення отриманих результатів.

1. Запропонований алгоритм відбору потенційних низькомолекулярних лігандів для доменів SH2 може бути ефективним для різних класів протеїнів.

2. Отримана бібліотека потенційно активних речовин може бути застосована для пошуку лігандів відносно будь-якого домену SH2 у біологічному тесті.

3. Знайдені біологічно активні сполуки можуть бути використані для розробки нових лікарських засобів.

Особистий внесок здобувача. Основна частина дисертації є самостійною науковою працею, у якій висвітлені власні ідеї/підходи щодо застосування різноманітних методів комп'ютерного моделювання для вирішення поставлених завдань. Автор самостійно опрацював літературні та експериментальні дані. Планування та обговорення усіх етапів дослідження проведені за участі наукового керівника, співробітників хімБіоЦентру Київського національного університету імені Тараса Шевченка та професора А. Варнека (Франція, Університет м. Страсбург). Аспірант виконав: розподіл доменів SH2 на групи і кластери; провів аналіз кристалографічних даних доменів; проаналізував відомі активні речовини стосовно доменів; провів молекулярний докінг; побудував фармакофорні моделі; виконав віртуальний і частково біологічний скринінги; побудовав GTM моделі; оцінив енергетичні характеристики зв'язування при утворенні комплексів "ліганд-домен"; самостійно встановлював та налаштовував програмне і апаратне забезпечення на усіх етапах розрахунків; активно обговорював отримані результати, займався написанням наукових статей і висвітленням результатів на конференціях.

Апробація результатів дисертації. Результати досліджень були представлені і доповідались на VII міжнародній науково-технічній конференції «БФФХ-2012» (23-27 квітня 2012 р., м. Севастополь, Україна); 2nd Ukrainian-French School-Seminar «Carbon Nanomaterials: Structure and Properties» (16-20 September 2013, Beregove, Ukraine); XI міжнародній науковій конференції студентів та аспірантів "Молодь і поступ в біології" (Львів, 20-23 квітня 2015р.); V Ukrainian–German Symposium on Physics and Chemistry of Nanostructures and on Nanobiotechnology (21–25 September 2015, Kyiv, Ukraine); International Conference for Young Scientists (21–25 September 2015, Kyiv, Ukraine), а також на наукових семінарах кафедри біофізики та

ННЦ “Інститут біології” Київського національного університету імені Тараса Шевченка.

Публікації. За матеріалами дисертації у наукових виданнях опубліковано 11 праць: 6 статей [6, 11, 12, 13, 14, 15] (одна з яких є оглядовою) у вітчизняних наукових фахових виданнях, з них 3 статті у журналах, включених до наукометричної бази даних Scopus; 5 тез наукових доповідей на міжнародних конференціях [16, 17, 18, 19, 20].

Структура та обсяг дисертації. Дисертаційна робота викладена на 135 сторінках тексту, складається зі вступу, огляду літератури, опису матеріалів та методів дослідження, результатів дослідження та їх обговорення, узагальнення результатів дослідження, висновків і списку літератури, який включає 209 джерела, з них 205 англомовні. Дисертаційна робота містить 19 таблиць, 22 рисунки.

РОЗДІЛ 1
ОГЛЯД ЛІТЕРАТУРИ

1.1. Протеїн-протеїнова взаємодія

Більшість клітинних процесів відбуваються завдяки дії протеїнів, які виконують специфічні функції, взаємодіючи між собою у складі стабільних або динамічних макромолекулярних комплексів [21]. Так, за допомогою PPI сигнали передаються з міжклітинного середовища всередину клітини та в межах клітини [22]. Внаслідок цього утворюється складна мережа PPI (інтерактом клітини), природа якої і є однією з головних причин складності біологічних систем. Широка розповсюдженість PPI в організмах тварин порівняно з іншими живими системами вказує на еволюційні ускладнення механізмів передачі сигналів. Тому при модуляції молекулярних біологічних процесів, окрім впливу на активний сайт протеїну, також можна регулювати зв’язування протеїну-мішені та його PPI структур, опосередковано впливаючи на мішень.

Під час дослідження PPI оцінюють типи взаємодіючих комплексів, геометричні характеристики (наприклад, розміри, комплементарність, зміну конформації контактуючих структур), гідрофобність, сегментацію [23]. Нажаль, більшість таких досліджень надають інформацію лише про поверхню протеїнів і практично нічого не кажуть про їхнє гідрофобне ядро [24]. Встановлено, що інтерфейси PPI є динамічними; плоскі і сферичні, зазвичай, містять багато випуклостей [25]. Найбільші інтерфейси протеїнів спостерігають у клітинних процесах зі значними конформаційними змінами [26]. Зазначимо, що PPI інтерфейси також характеризуються електростатичною комплементарністю [23]. Відомо, що PPI головним чином

відбуваються у напрямку зменшення епіструктурного напруження для індивідуальних патернів [27] завдяки структурним дефектам протеїнів (дегідрон) [28].

Враховуючи вище зазначене, можна зробити висновок, що протеїни, які беруть активну участь у PPI, є зручними мішенями для пошуку потенційних лікарськтих засобів [29, 30]. Такі протеїни використовують як ефективні мішені для розробки нових препаратів від різноманітних хвороб, зокрема онкозахворювання, захворювання кровоносної системи, тощо.

Наразі, під час клінічних досліджень, протестовано значну кількість хімічних структур та отримано декілька готових препаратів (наприклад, Titrobifan, Maraviroc) [31]. Але мала кількість підтверджених протеїнів-мішеней обмежує розробку нових терапевтичних засобів [29]. Крім того, наявні структурні дані свідчать, що лише 1% протеїнів протеому людини є прийнятними мішенями для інгібування низькомолекулярними сполуками [32]. Здебільшого це протеїни, які входять до складу одних родин або функціональної групи і характеризуються наявністю ідентичного/подібного нативного ліганду або субстрату (зазвичай це низькомолекулярні сполуки-метаболіти, нейромедіатори розміром 10 і більше амінокислотних залишків, що створює труднощі при моделюванні лігандів).

Важливо, що для отримання активних речовин в *in vitro* експерименті в діапазоні нмоль/л, відібрані в межах *in silico* досліджень потенційно активні сполуки повинні зв’язуватися з ключовими точками сайту зв’язування протеїну нековалентними зв’язками і відповідати правилам Ліпінскі [33]. Проте часто, відібрані речовини не відображають усі нековалентні взаємодії між протеїном і нативним лігандом. Це вказує на можливий різний характер зв’язування білка-патерна нативного ліганду і потенційного низькомолекулярного ліганду із сайтом зв’язування PPI. Однак розробка інгібіторів PPI, яка враховує частину пептидного ланцюга, не завжди приводить до позитивного результату, оскільки число обертальних зв’язків лікоподібної молекули, згідно правил Ліпінскі, не повинно перевищувати 10

(умова лікоподібності речовин пов’язана із селективністю і токсичністю), а один амінокислотний залишок містить як мінімум три таких зв’язки. Більше того, синтез сполук на основі поліпептидів є складним і не дешевим процесом.

Часто PPI відбувається через домени протеїнів, які є по суті структурними блоками макромолекулярних комплексів та мережі PPI [34]. Згідно вторинно- та третинно-структурного розподілів налічується близько 80 родин протеїнових доменів, кожна з яких містить близько 500 представників одного і того самого протеїнового домену та біля 260 протеїнів з доменами у своїй структурі [35]. Ідентичність амінокислотних послідовностей всередині однієї родини складає від 20% [36]. Крім того, часто домен/протеїн зв’язується лише з конкретною хімічною структурою.

Отже, ця робота спрямована на дослідження PPI родини доменів SH2, що вважається перспективною мішенню для розробки різноманітних лікарських засобів. Вона є добре охарактерезованою, зокрема відомий достатній масив тривимірних структурних даних (як вільних (apo), так і зв’язаних з лігандом (holo)), пептидні ліганди і механізми їх зв’язування з доменом. Також добре вивчено можливі зв’язки мішені з різноманітними процесами у клітині, а, отже, і терапевтичне значення доменів SH2.

1.2. Домени SH2

1.2.1. Будова та функції доменів SH2. SH2 – компактний глобулярний домен, який взаємодіє з протеїнами, що містять фосфорильований залишок тирозину (Tyr) [37, 38]. Здебільшого він знаходяться в онкопротеїнах (Src oncoprotein [39]) та протеїнах, які входять до сигнальних каскадів клітини [40, 41]. Домени SH2 мають високу спорідненість до фосфотирозину. Вони є найбільшим класом pTyr – розпізнавальних доменів [42].

Функціонально домени SH2, як правило, зв’язані з тирозинкіназами протеїнів (PTK). Вони беруть активну участь у внутрішньоклітинній передачі сигналів, що пов’язано з їхньою здатністю специфічно розпізнавати фосфотирозин. Цей процес має фундаментальне значення при передачі

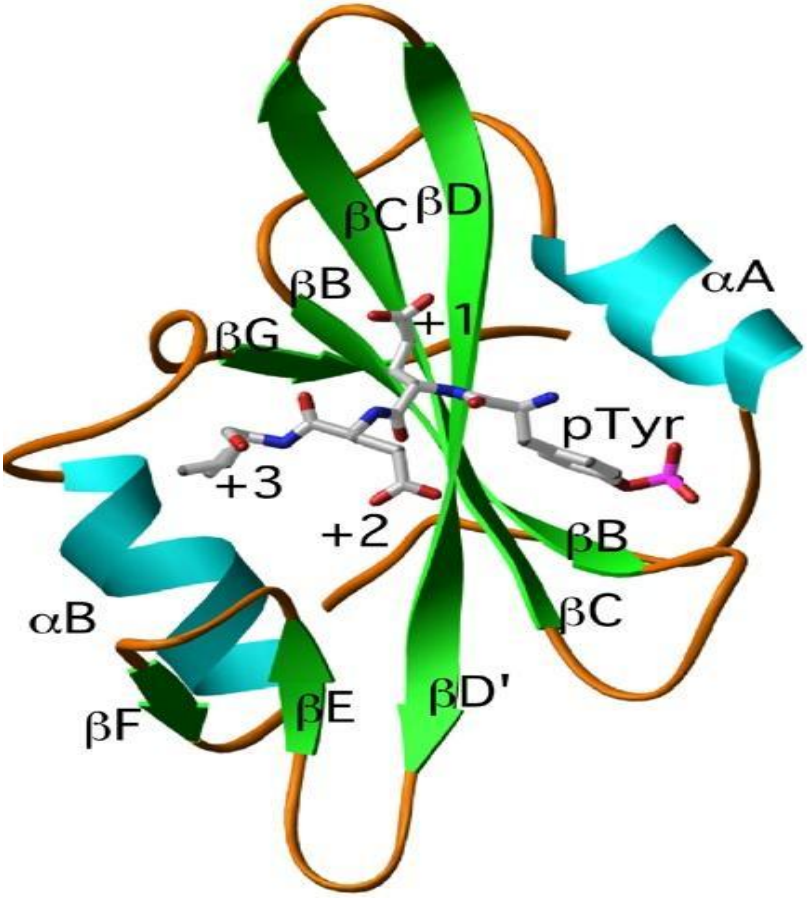


Рис. 1.1**.** Домен SH2, зв’язаний з фосфотирозин – вмісним лігандом.

сигналів через мембрану: коли сигнал (певна хімічна структура) надходить у позаклітинний простір, він контролюється за допомогою рецепторів [43]; потрапивши знов у клітину – контролюється структурами, які здатні розпізнати фосфотирозин. Фосфорилювання тирозину призводить до активації PPI, внаслідок чого SH2-вмісні протеїни зв’язуються з фосфотирозиновим скефолдом. Тому, мутації доменів SH2 викликають порушення PPI, що, в свою чергу, приводить до різноманітних патологій (Табл. 1.1, Табл. 1.2.).

Перші домени SH2 ідентифікували в онкопротеїнах Src та Fps. Встановлено, що домени SH2 містять близько 100 амінокислот, які формують дві α-спіралі (αA і αB) та сім β-складок (βG, βB, βC, βD, βD', βE і βF) (рис. 1.1.) [44].

Таблиця 1.1.

Захворювання людини пов'язані з мутаціями генів доменів SH2

№	Назва SH2-вмісного протеїну	Короткий опис мутацій	Вплив на організм
1	ABL1	t(9;22) транслокація призводить до злиття з геном BCR	Лейкемія, гостра лімфобластомна лейкемія, хронічний мієлоцитарний лейкоз (AML)
2	ABL2	t(1;12)(q25;p13) транслокація з ETV6/TEL	Гостра мієлогенна лейкемія
3	BLNK	Заміна пари основ у місці сплайсингу призводить до зменшення або повної втрати BLNK-транскриптів	Імунодефіцит людини Попередник гострого лімфатичного лейкозу
4	Btk	Мутації сайтів сплайсингу	Хвороба Брутона
5	CBL	Зміни в частинах MLL-гена: t(4;11)(q21;q23) t(11;14)(q23;q32) t(11;22)(q23;q12)	Гостра мієлогенна лейкемія Гостра лейкемія Лімфома В-клітин Саркома
6	CRK	Вилучення з локусу MDLS, MDS, MDCR, DEL17p13.3, C17DELp13.3 частини 17p13.3	Синдром лісенцефалії Міллера – Дікера
7	ITK	t(5;9)(q33;q22) транслокація в Syk	Периферійні Т-клітинні лімфоми
8	JAK2	t(9;12)(p24;p13) зливаються з TEL t(9;15;12)(p24;q15;p13) fusion with TEL and ETV6	Гостра лімфатична лейкемія Хронічна мієлогенна лейкемія,
9	JAK3	Унаслідок вставки або видалення нуклеотидів зсувається рамка зчитування або процес взагалі обривається	Імунодефіцит
10	LCK	Транслокація t(1;7)(p34;q34) націлена на LCK і TCRB	Гострий лейкоз Т-клітин

		Зниження експресії p56 (lck)	SCID
11	PTPN11	Унаслідок missense мутації амінокислота D61 замінюється на N-SH2, унаслідок чого посилюється функціональна активність.	Синдром Нунана
		Missense мутації у родині екзотів 7, 12 та 13 становлять 95%; мутації родини екзотів та дефекти у родині 13 впливають на протеїнові фосфатази тирозиновогодомена, унаслідок чого посилюється функціональність	Прогресивна кардіоміопатія (LS) Мієломоноцитарний лейкоз
12	SH2D1A	Мутації, пов'язані з неправильним функціонуванням протеїну	Х-хромосні лімфопроліферативні хвороби
13	STAT1	Унаслідок нуклеотидних замін передчасно генерується стоп-кодон	Зростає уразливість до вірусних і бактеріальних захворювань
14	STATT5B	Гомозиготні мутації A630P	Імунодефіцит
15	RASA1	Нонсенс-мутація мутації домену SH2, результат утворення стоп кодона	Базаліома, артеріовенозна мальформація
16	SH3BP2	Точкові мутації в екзоні 9 впливають на три амінокислоти в межах шести-амінокислтної посідовності (RSPPDG)	Херувизм
17	SHIP2	Мутації в SRC кодоні 531	Діабет, артеріальна гіпертензія
18	SRC	t(9;12)(q22;p12) транслокації в межах TEL	Колоректальний рак
19	Syk	t(5;9)(q33;q22) транслокація в межах ITK	Мієлодиспластичний синдром
20	TYK2	Гомологічні заміни амінокислот в JH ділянці JAK	Системний червоний вовчак
21	ZAP70	Мутація сайту сплайсингу	Дефекти Т-клітин

1.2.2. Компоненти доменів SH2 людини. Дослідження [43, 45] ідентифікували 120 доменів SH2, які знаходяться у 110 різних протеїнах, 10 з яких містять подвійні домени SH2 у своїй структурі. Для порівняння, більш ранні дослідження [46, 47] виявили лише 87 та 98 протеїнів доменів SH2. Ці

домени присутні у найрізноманітніших класах протеїнів, включаючи протеїнкінази (Src, Lck), фосфатази (SHP2, SHIP2), фосфоліпази (PLCγ1), фактори транскрипції (STAT), регуляторні протеїни (SOCS), адаптери протеїнів (Grb2), структурні протеїни (SHC) та ін.

Крім того, літературні дані свідчать про деяку гомологічність між доменами SH2 людини та інших організмів. Здебільшого цю гомологічність визначали за допомогою наявних кристалографічних даних щодо того чи іншого домену SH2 з використанням програмних пакетів BLSAT [48] та UniGene [49]. Так, наприклад: SH2D1C (ERT) – функціональний ген у мишей та псевдоген в організмі людини; SH2D1C є сигнальним регулятором, пов'язаним з SH2D1B/EAT2 та SH2D1A/SAP, які, в свою чергу, виступають регуляторами передачі сигналів [50]; з іншого боку, людський геном кодує SH2D3A, який відсутній у гризунів, але присутній у приматів [51].

Захворювання людини, пов'язані з генетичними пошкодженнями протеїнів, що містять домени SH2 [52], представлені в Табл. 1.1., а порушення генів миші – у Табл. 1.2. Наведено усі SH2-вмісні протеїни, закодовані у геномі людини. Однак важливо відзначити, що здебільшого результат робіт такого характеру залежить від набору тих методів, які використовують [53, 54].

Таблиця 1.2.

Захворювання мишей, пов'язані з порушеннями доменів SH2

№	Назва SH2-вмісного протеїну	Розлади, що винекають при порушення роботи відповідного протеїну	Частина організму, що піддається впливу
1	Abl1	Смерть через 1-2 тижні	Тимус, селезінка
2	Abl2	Порушення поведінки	Мозок
3	Aps	Гіперинсулинемия	B1 лімфоцити, адепоцити
4	Bks		Гепатоцити
5	Blk		Не впливає
6	Blnk	Лімфома	B-лімфоцити
7	Bmx		Хвороби серцевосудинної системи
8	Btk	XID фенотип	B-лімфоцити

9	Cbl		Тимус, Т-лімфоцити
10	Cblb	Аутоімунний енцефаломієліт	Т-лімфоцити
11	Cblc		Не впливає
12	Crk		Не впливає
13	Crkl	Ембріональна летільність	Тимус
14	Csk	Ембріональна летільність	Нервова трубка
15	Dapp1		В-лімфоцити
16	Fer		Не впливає
17	Fes	Вплив на імунну систему	Кістковий мозок, мієлоїдні клітини
18	Fgr	Дефекти легень	Ерозія дихальних шляхів
19	Frk		Не впливає
20	Fyn		Гіпокамп
21	Gads		Т-лімфоцити
22	Grap		Лімфоцити
23	Grb2	Ембріональна летальність	Ендодерм
24	Grb10	На ~30% збільшує органи	Мозок, печінка
25	Grb14	Поліпшення засвоєння глюкози	Печінка, склетні м’язи
26	Hck		Макрофаги
27	Itk		Т-лімфоцити
28	Jak1	Принатальна летальність	Т/В-лімфоцити
29	Jak2	Летальність ембріонів	Гемоцитобласти
30	Jak3	Лімфопенія	Тимус
31	Lck	Атрофація тимусу	Тимус
32	Lnk		Не впливає
33	Pik3r1		В-лімфоцити
34	Pik3r2	Гіпоглікемія	М’язи
35	Plcg2	Дефіцит В-клітин	В-лімфоцити
36	Ptpn6	Дефекти імунітету та кровотворення	Т-лімфоцити
37	Rasa1	Ембріональна летальність	Судинна система

1.2.3. Неохарактеризовані/слабко охарактеризовані домени SH2. Деякі домени SH2 раніше не були описаними взагалі або частково загадувалися у літературі. Наприклад, LOC284948 є представником родини SH2 скефолд протеїнів і міститься у SLP-76, BLNK та MIST/CLNK; LOC284948 продукт гена SLNK, пов’язаного з BLNK [43]; SH2D5 (LOC400745) є слабко охарактеризованим протеїном, що має подібну будову домену з Shc родини протеїнів, представлених у PTB [55]. Зазначимо, що у цьому випадку подібність обмежується лише будовою структур; SH2D5 має мінімальну подібність амінокислотних послідовностей зі структурами Shc родини. Крім того, родина Shc включає чотири члени (Rai-like протеїни (RaLP)) з невідомими досі функціями. Згідно сучасної номенклатури ця родина

називається Shc4 (або ShcD) [56]. Кожен член цієї родини містить один домен SH2: їх класифікували як поодинокі регуляторні протеїни, хоча мало що відомо про їхню клітинну роль. Ці структури можна також поділити на невеликі групи. Так, структури SH2D4A та SH2D4B і HSH2D та SH2D2A характеризуються високою подібністю попарно, але деякою відмінністю у наборах амінокислотних послідовностей між кожною окремою парою структур. Supt6h – архаїчний протеїн, який зберігся від дріжжів до людини [57, 58]. Він містить здебільшого примітивні SH2-подібні амінокислотні послідовності, які зустрічаються лише у дріжжів (хоча зазначимо, що деякі літературні дані свідчать про відсутність доменів SH2 у дріжжах [59]). На відміну від більшості інших SH2-вмісних протеїнів, Supt6h не набув значного поширення у різноманітних організмах та не містить значного різноманіття варіантів доменів SH2. Так, в організмі тварин міститься примітивний фрагмент, який навіть не зв'язується з pTyr-фрагментом (примітивна SH2-подібна послідовність) [60]. Однак, еволюційна консервативність і те, що він досі є представленим в організмі тварин, свідчать про його певну функціональну (невизначену досі) роль.

1.2.4. Дефекти, пов'язані з доменами SH2. Функціональна роль доменів SH2 детально досліджена для ссавців і безхребетних. Деякі SH2-вмісні протеїни є важливими для ембріогенезу, а саме: Grb2, RasGAP, Csk, Shp2, Shc1, PLC-g1, Jak2 та Stat3. Інші потрібні на більш пізніх стадіях розвитку організму. Здебільшого це ті, що функціонують лише у спеціалізованих типах клітин. У клітинах імунної системи функціонують такі SH2-вмісні протеїни, як Lck, ZAP-70, Slp-76 та Gads, що зв'язуються у сигнальні каскади, наприклад, Т-клітин [61, 62, 63, 64]. Для деяких SH2-вмісних протеїнів характерним є перекривання функцій. У Табл. 1.2. показано 14 випадків комбінованої інактивації однієї або декількох родин, внаслідок чого має місце комплексний вплив на фенотип мишей. Наприклад, у Vav12/2,

Vav22/2, або Vav32/2 мишей спостерігають мінімальні фенотипові аномалії [65, 66, 67], водночас потрійно накаутні миші Vav1/2/3 містять значні дефекти розвитку лімфоцитарних клітин і, як наслідок, лімфоцитної системи [65].

Наразі відомо 18 мутацій різних SH2-вмісних протеїнів, які здатні спричинити розлади організму людини (рак, лейкемія, порушення розвитку, діабет, імунодифіцит (Таблиця 1.1.)). Причиною цього може стати втрата функціональності доменом або, навпаки, посилення його функціональних можливостей. Так, можливі мутації, які впливають на властивості зв'язування SH2D1A/SAP або Btk з пептидом, внаслідок чого може виникнути імунодефіцит, зумовлений ензиматичною активністю тирозинфосфатази Shp2 (містить два N-кінцеві каталітичних домени SH2), яка контролюється взаємодією N-кінців доменів SH2 з активним сайтом фосфатазного домену [68]. Зв'язування pTyr-пептидів з доменами SH2 активує каталітичний домен та співставляє його з мішенню, завдяки чому відбувається стимуляція ERK/MAP кіназного шляху. Так, наприклад, синдром Нонада (характеризується вродженими хворобами серця, скелетними дефектами, когнітивними порушеннями) виникає внаслідок мутацій у гені людини Shp2 (PTPN11) [69]. Також важливо зазначити, що не всі мутації у цьому випадку викликають негативний ефект чи певні зміни в інтерфейсі доменів SH2, що безпосередньо зв'язується з pTyr, внаслідок чого можлива втрата автоінгібування без значного впливу на каталітичні чи pTyr властивості зв'язування домену. У цьому випадку консервативні амінокислотні заміни можуть викликати лейкемію (наприклад, ювенальний мієломоноцитарний лейкоз), пов'язану зі збільшенням активації Shp2 за таких мутацій [70].

Для доменів SH2 також характерні критичні функціональні компоненти химерних онкопротеїнів людини (наприклад, Bcr-Abl [71]), які сприяють специфічній дії кіназних інгібіторів (наприклад, imatinib/Gleevec [72]). Мутації сайту зв'язування доменів SH2 сприяють онкогенності людської

RTKs (наприклад, у випадку з Met RTK мутації викликають рак легень людини), яка активується внаслідок втрати Cbl SH2 сайту зв’язування [73].

1.2.5. Структурне та функціональне різноманіття доменів SH2. Перші структури доменів SH2 знайшли у 1992 р., а подальший аналіз виявив консервативні та варіативні особливості взаємодії доменів SH2 [74, 75, 76]. Загалом на цей момент PDB включає більше 340 структур доменів SH2, які містяться у 110 протеїнах людини [43, 45]. Нажаль, лише для половини з них існує певна кількість літературної інформації. Зазначимо, що багато з цих протеїнів перебуває у комплексі з пептидним або сентитичним лігандом.

Домени SH2 містять у середньому 100 амінокислот, які формують N-кінець – αA-спіраль та центральні антипаралельні β-складки (βA- βD). Крім того, формуються мала β-складка (βD0, βE, βF), друга αB-спіраль, C-кінець, βG – складка [77]. У таких структурах β-складка відокремлює консервативну pTyr кишеню зв’язування від менш консервативної частини домену (біля C-кінця pTyr – зв’язуючої частини кишені, амінокислоти якого частково втягнуті у pTyr сайт зв’язування). Вважається, що це і є одна з головних причин зв’язування доменів SH2 з фосфорильованими пептидними послідовностями тирозину (його внесок складає приблизно половину вільної енергії зв’язування) та амінокислот C-кінця, які, зазвичай, визначають специфічність зв’язування доменів SH2. Ці короткі фосфопептидні ділянки є компонентами гіганського поліпептиду, такого як RTKs [78, 79]. Крім того, зазначимо, що такі речовини можна використовувати для визначення потенційного сайту зв’язування доменів SH2 [80, 81, 82].

З урахуванням того, що домени SH2 є досить подібними за своїми властивостями, у багатьох випадках для ідентифікації можливого сайту

зв’язування використовують вже наявні дослідження. Так, вони вказують, що домени SH2 можуть зв’язуватись з фосфорильованим пептидом декількома різними шляхами (рис. 1.2.). Найчастіше у зв’язуванні з фосфорильованим пептидом беруть участь усі центральні β-складки та декідька амінокислот С-кінця. Наприклад, фосфоліпаза (PLC)-g1 зв’язується з коротким пептидом (простий варіант зв’язування) [83], тобто ліганд зв’язується у pTyr сайті та гідрофобній кишені.

На рис. 1.2. наведені варіанти зв’язування доменів SH2 з лігандами:

А – структура домену SH2, зв’язаного з pTyr-Glu-Glu-Ile пептидом (PDB, 1SPS). Поверхню домену SH2 показано як напівпрозору частину, а вторинні структурні елементи – голубим кольором; αA – спіраль справа, αB – спіраль зліва; Arg βB5, залишок для зв’язування pTyr, показано жовтим кольором; N-кінець pTyr пептиду (червоний колір) займає pTyr-зв’язувальну кишеню. Пептид рухається повз центральну β-частину домену SH2; +1 і + 2 глутамати зв’язуються з поверхнею домену, а бічна частина +3 (зліва) входить у гідрофобну кишеню.

Б – Grb2 SH2 у комплексі з pYVNV (червоний колір), (PBD, 1BMB) і Trp (зелений колір) стабілізують конформацію β-структури, необхідну для високоафінного зв’язування.

В – дві фосфотирозин – зв’язувальні кишені подано в одному домені SH2 APS. Одна APS SH2 молекула зв’язана з pYETDpY (червоний колір) пептидом активаційної петлі INSR; pTyr-1158 взаємодіє з Arg-438 (зелений колір), Lys-455 і Lys-457 (помаранчевий колір) утворюють іншу кишеню зв’язування для pTyr-1162 (PDB, 1RQQ).

Г – SH2D1A/SAP у комплексі з нефосфорильованим SLAM пептидом KSLTIYAQVQK (червоний колір) (PBD, 1D4T).

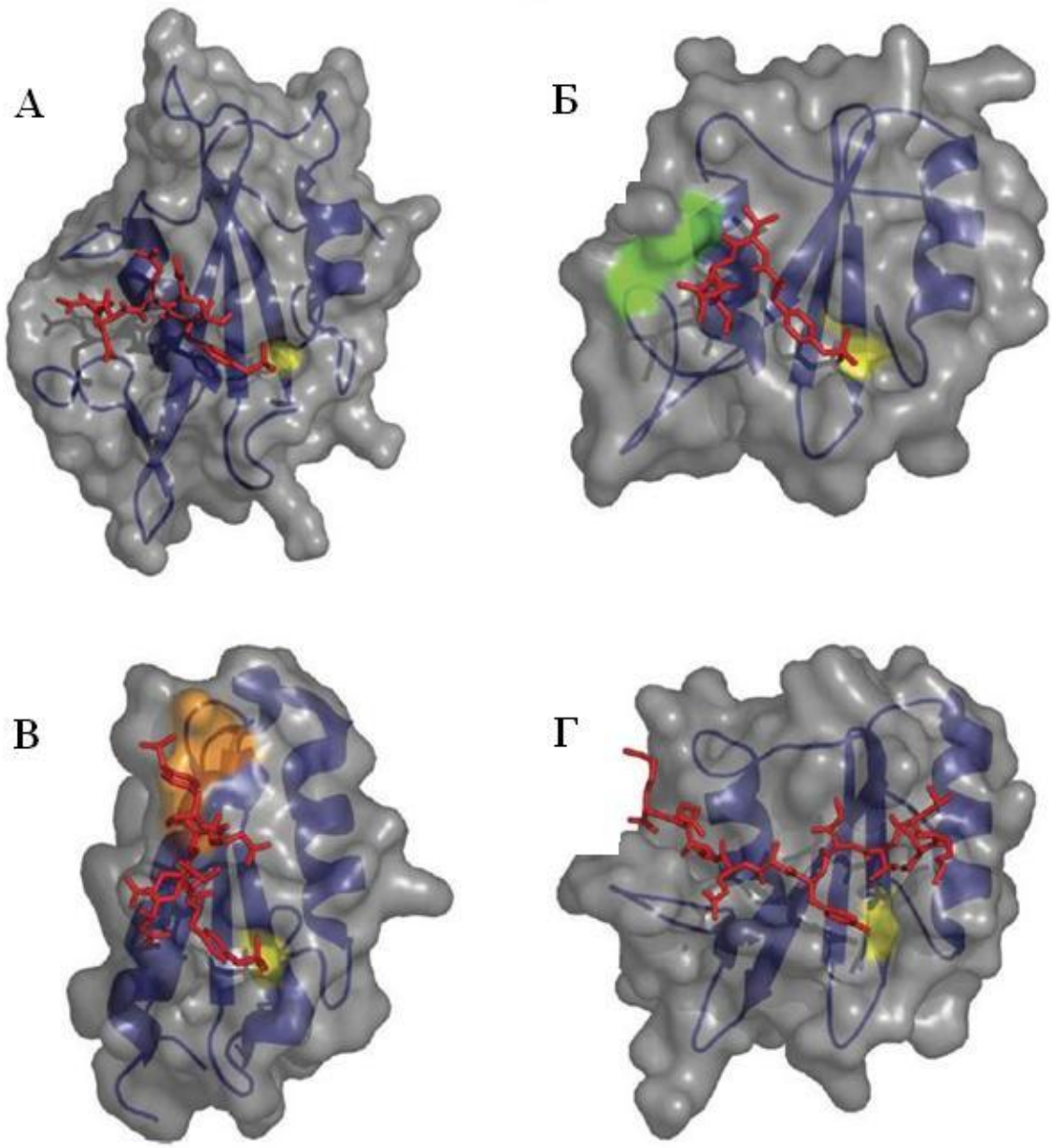


Рис.1.2. Варіанти зв’язування доменів SH2 (стрічкова та поверхнева діаграми) [43].

Значна частина доменів SH2 зв’язується з довгими фосфопептидами (складний варіант зв’язування) у розгорнутій конформації. Це відбувається завдяки розпізнаванню пептиду (як C, так і N кінцями), наприклад, SH2D1A доменом SH2. Крім того, цей домен може зв’язуватись з нефосфорильованим пептидом з досить високою афінністю [84, 85].

SOCS3 домен SH2 має розширений N-кінець, що формує гідрофобну кишеню, щільно зв’язується з pTyr пептидом рецептора IL-6 (Kd = 150 nM) [86]. Водночас, pTyr-X-Asn пептид при зв’язуванні з Grb2 доменом SH2 сфокусований на зв’язувані лише у pTyr-зв’язуючій кишені та не заходить у гідрофобну кишеню [87]. Ще один варіант зв’язування спостерігається з APS доменом SH2, де фосфорильована активаційна петля INSR є структурно обмеженою внаслідок взаємодії домену з кіназними доменами. Крім того, у деяких випадках можна спостерігати взаємодію pTyr із залишками Lys, що знаходиться у βD-складці [88].

Домени SH2 також можуть бути посередниками ідіосинкразійних PPI через поверхню зв’язування, яка відрізняється від звичайних фосфопептид-зв’язуючих ділянок [89]. Завдяки цьому деякі індивідуальні домени SH2 можуть зв’язуватись з декількома структурами. Так, домен SH2D1A має вторинну поверхню зв’язування, завдяки чому можливе комплексоутворення з доменами SH3 [90]. Домен SH2D1A Т-клітин входить у SLAM рецептор [91].

У Т-клітинах SH2D1A утворює структурний комплекс із SLAM рецептором шляхом зв’язування з лігандом. Однак, з комплексом можливе зв’язування ліганду домену SH3 Fyn тирозинкіназ. Це стимулює Fyn-кіназну активність, яка обмежує активність Т-клітин [90, 92].

Домен Crk SH2 може зв’язуватись з доменом Abl SH3 через пролін-збагачену петлю. За деяких взаємодій домени SH2 протеїнів Grb10, Grb14, APS та Grb7 формують нековалентні гомодимери через консервативний

інтерфейс αB-спіралі, одночасно фосфорилюючи активаційну петлю рецептора RTK [58, 88].

Також можливе утворення тандему доменів SH2 та структур PI3K p85, ZAP-70, Syk, Shp-2 і фосфоліпази C-g1, кожен з яких зв'язується з фосфорильованими тирозин-залежними ділянками [93].

1.2.6. Фосфотирозин-незалежні взаємодії доменів SH2. Як було зазначено вище, домен SH2 протеїну Supt6h може втрачати pTyr-зв'язуючу активність. Його можна віднести до низки інших доменів, що включають послідовності JAK тирозин кінази. Член родин JAK і TYK2 містять His замість критичного ArgbB5, що координує фосфатну групу pTyr. З іншого боку, не зважаючи на те, що домени SH2 людини протеїнів JAK1, JAK2 і JAK3 містять Arg у позиції βB5, вони відрізняються N-кінцями домену SH2, що охоплює βA, αA, βB і βC-складки [94]. Інтактний домен SH2 протеїну JAK подібним чином зв'язується з цитокіновими рецепторами. Звідси можна припустити, що домени SH2 мають незвичайні pTyr-зв'язуючі властивості, які є важливими для розпізнавання рецептора. Дійсно, при моделюванні зв'язування JAK2 виявили, що домени SH2 можуть контактувати з N-кінцем домену FERM [95]. Подібно до цього, домени SH2 протеїну Rin2 (HisbB5) та SH2D5 (TrpbB5) також не містять ArgbB5 і тому можуть не зв'язуватися з pTyr лігандами. Домени SH2 фосфатидилінозитолу (PI) 30-кінази зв'язуються з PI(3,4,5)P3, що перешкоджає розпізнаванню pTyr-вмісних пептидів, і цим самим забезпечують зворотне інгібування для взаємодій типу "домен SH2 – фосфопептид" [96].

Домен SH2 також може зв'язуватись з PIP3 (фосфатидилінозитол (3, 4, 5) трифосфат) [96], сульфогалактозою [97] та фосфопептидами. Це вказує на диференціацію доменів SH2 для різних варіантів зв'язування. Крім того, при аналізі 3D структур фосфотирозин-незалежних доменів виявили, що їх тривимірна упаковка є давнішою, ніж у pTyr-зв'язуючих доменів SH2.

BirA, E. coli містять структурну ділянку, аналогічну в доменах SH2 [98], яка зв’язується з фосфатом у формі АТФ.

Отже, за функціональної і структурної невизначеності амінокислотних послідовностей важко відслідкувати еволюційні взаємозв’язки, що, в свою чергу, підтверджує різне функціональне призначення доменів SH2 у живих організмах.

1.2.7. Класифікація доменів SH2. Домени SH2 знаходяться у протеїнових комплексах з різними біохімічними властивостями. Їх поділяють на 11 груп (рис. 1.3, 1.4.). Зазначимо, інформація щодо класифікації доменів визначає фізіологічні межі SH2-вмісних мішеней, доступних для pTyr сигналізації у клітинах людини. Функціональний вплив, наприклад, будь – якого рецептора тирозинкінази (РТК) залежить від тих SH2 – вмісних протеїнів (Табл. 1.1. і 1.2.), які вона мобілізує безпосередньо або опосередковано через фосфорильовані молекули скефолда. Ці дані визначають головні шляхи впливу РТК на внутрішньоклітинні процеси. pTyr – залежні взаємодії у SH2 – вмісних протеїнах спрямовують сигнал РТК у визначену точку. Цей процес охоплює фосфорилювання тирозину (через цитоплазматичні РТК і тирозин (фосфатазу)), контроль метаболізму фосфоліпідів (фосфатидилінозитол, інозитол фосфат), регуляцію маленьких ГТФаз (включаючи Ras, Rho і Rap родини) гуанін – вмісними обмінними факторами і генну експресію ГТФаз (в межах сигнальних каскадів).

SH2/SH3-вмісні протеїни – це здебільшого адаптори (наприклад, Grb2, Crk, та Nck [99]). Nck – цитоскелетні регулятори (N-Wasp і Pak серін/треонін кінази [100, 101]; Grb2 зв’язані із Sos та Gab1 [98, 102, 103]; беруть участь у MAPK/PI3K сигналізації [104]. Зазначимо, що значна кількість SH2 протеїнів регулює тривалість РТК сигналізації (SOCS протеїни [105]). Cbl протеїни спричиняють pTyr-залежне мільтиубіквитинування [106].

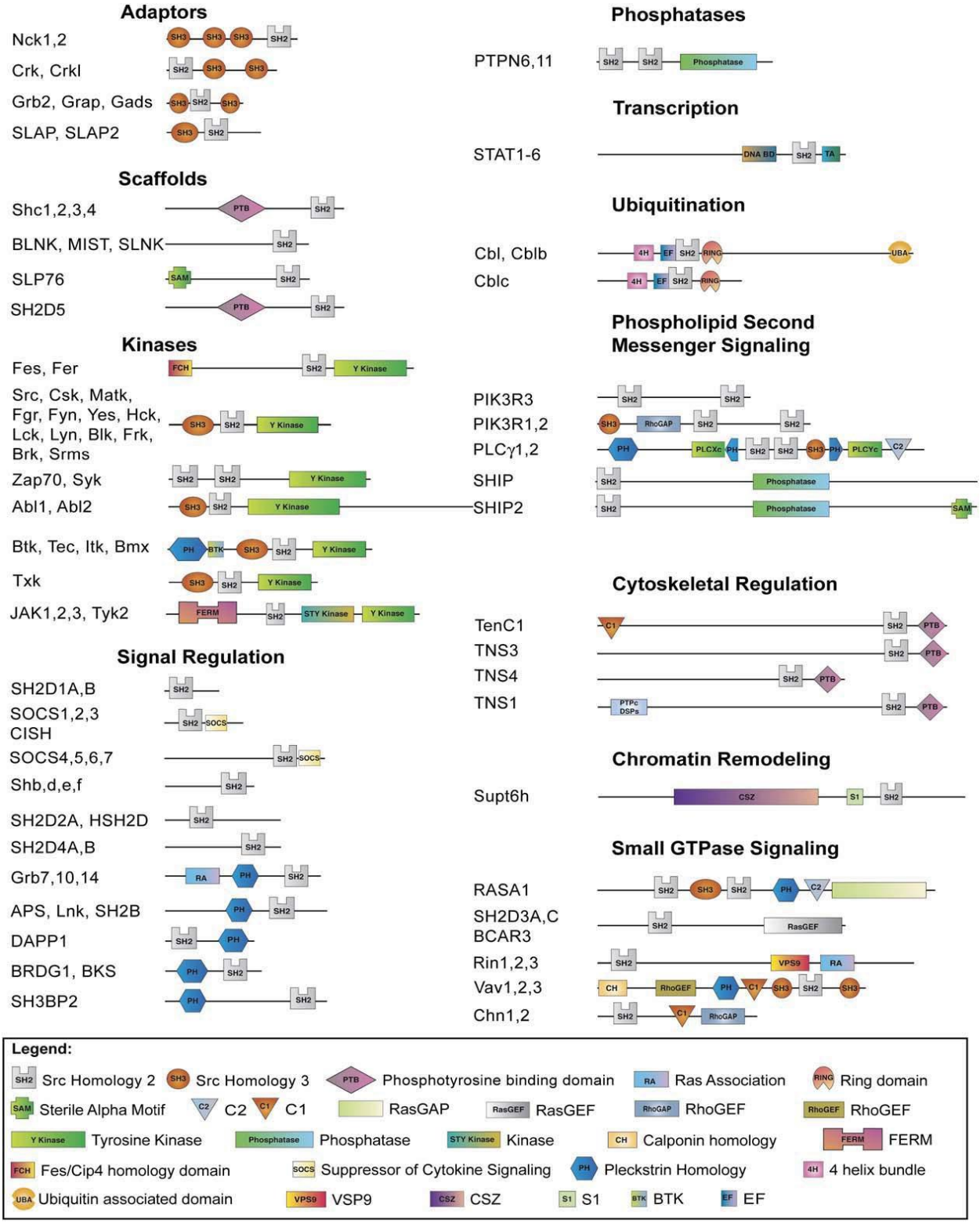


Рис.1.3. Протеїни, до складу яких входять домени SH2 [43]

Домени SH2 здатні взаємодіяти з активними сайтами RTKs. Наприклад, домен SH2 протеїну APS одночасно гомодимеризується та зв'язується з активаційною петлею інсулінового рецептора (INSR), стабілізуючи його активний сайт [88]. Навпаки, домени SH2 пртеїнів Grb14 можуть

зв’язуватись з фосфорильованою INSR активаційною петлею завдяки наявності BPS ділянки, розташованої між SH2 та PH доменами [58]. Таким чином, SH2-вмісні протеїни можуть виступати як регулятори внутрішньоклітинних біохімічних процесів, модифікуючи кінетичну активність, субстратну специфічність компонентів сигнальлних тирозин-кіназних каскадів. Беручи до уваги велику кількість різних взаємодій і каталітичних доменів у протеомі людини, очевидно, що домени SH2 є асоційованими з великою кількістю різноманітних біологічних структур, зокрема: SH3, PTB, PH, кінази, фосфатази, GAP та домени GEF (рис. 1.2, 1.4.). Причиною цього може бути те, що SH2-вмісні протеїни завжди знаходяться біля плазматичної мембрани, де вони беруть участь у специфічних сигнальних передачах, сприяють росту клітини, диференціюванню, морфологічним змінам та метаболізму. Хоча зазначимо, транскрипційні фактори протеїнів STAT знаходяться біля інших субклітинних компартментів.

Завдяки модульній конструкції, домени SH2 можна штучно віднести до доменів створення аномальних сигнальних шляхів [107]. Ймовірно, ці нефізіологічні комбінації виникли під час еволюції для захисту PTKs від зворотніх внутрішньоклітинних реакцій. Для кращого розуміння процесу еволюції доменів SH2 їх можна поділити на декілька дискретних родин.

Загалом у літературі застосовували кілька основних методів ідентифікації взаємозв’язків доменів SH2 у внутрішньоклітинному просторі, зокрема вирівнювання амінокислотних послідовностей та їх порівняння. Однак, лише цього методу не досить для поділу доменів на

різні родини [43]. Організація протеїнових доменів також залежить від будови самого домену, його композиції та порядку розміщення в межах того чи іншого протеїну. Усе це разом можна застосувати для еволюційного детектування гомології на противагу порівняння амінокислотних

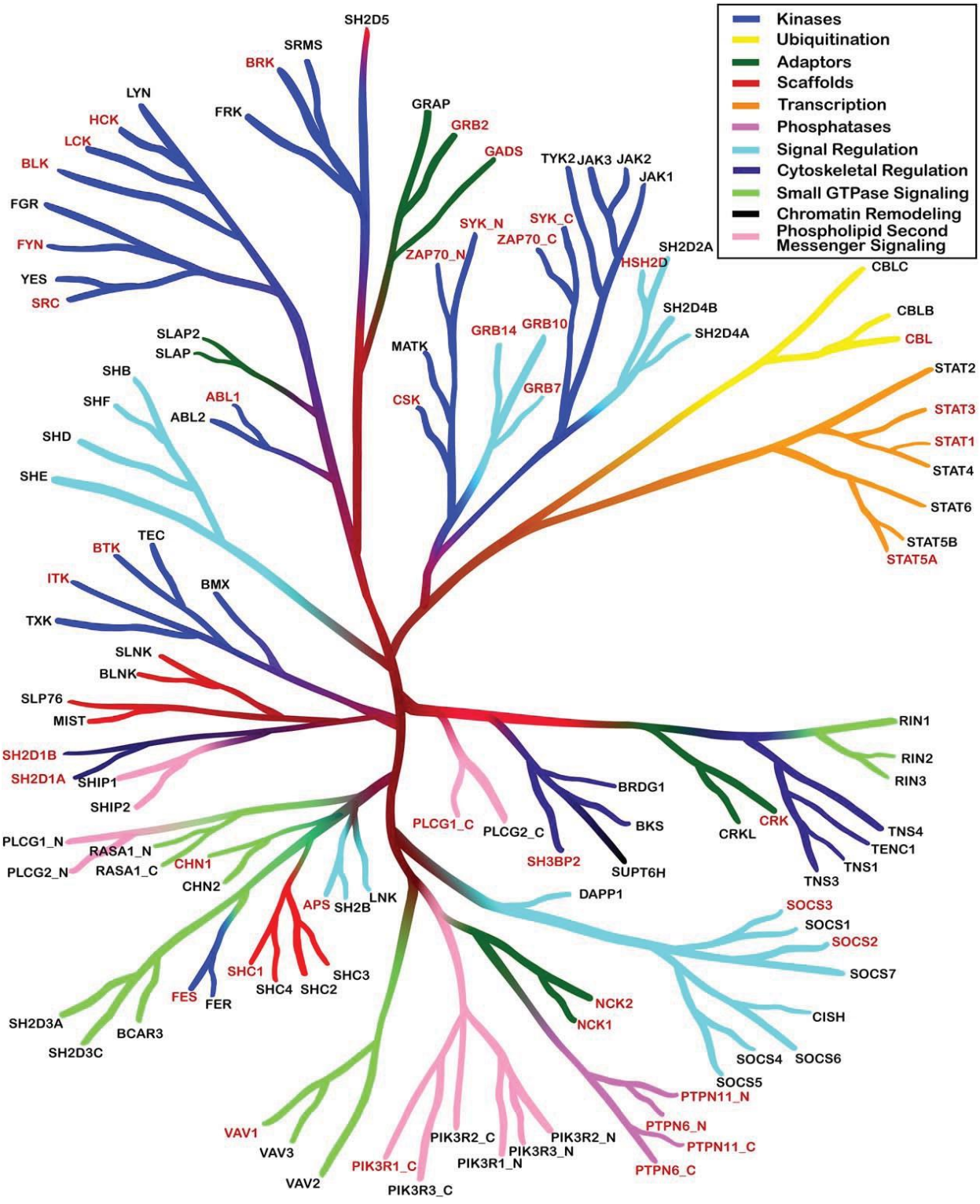


Рис.1.4. Дендрограма доменів SH2, яка демонструє подібність доменів SH2, наведених у табл. 1.1. та 1.2. [43]

послідовностей [108].

Попередні дослідження доменів SH2 людини свідчать про те, що багато з них мають ідентичну геометричну організацію, але не можуть бути віднесені до однієї родини, оскільки містять різні амінокислотні послідовності (гомології послідовностей) [43]. Також застосовували аналіз “інтрон/екзон коду” за відсутності достатньої кількості філогенетичної інформації [109]. На генетичному рівні сплайсинг та шифтінг екзонів демонструє еволюційний механізм збереження важливих консервативних амінокислотних послідовностей як результат клітинних функцій [110, 111]. На основі цього отримали моделі сплайсингу доменів SH2 людини [60]. Цей аналіз доповнили використанням програмного пакету SMART для визначення позицій інтрон/екзонного коду. На основі базових структур доменів SH2 визначили інтрон/екзонний код для усіх доменів SH2 за допомогою вирівнювання їх структур, які містять вторинні елементи (рис. 1.5.). Ієрархічна кластеризація інтрон/екзонного коду елементів вторинної структури доменів SH2 вказує на різні, але зі збереженням консервативності, моделі сплайсингу в межах подібних структур доменів. Цю класифікацію провели за такими параметрами: порівняння сіквенсів, організація доменів протеїну та моделі сплайсингу. Отримали 38 родин доменів SH2 в межах геному людини. Цей аналіз

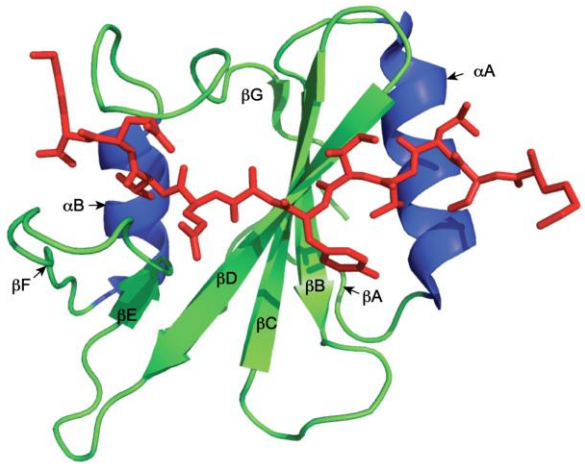


Рис. 1.5. Зображення інтронного та екзонного сайтів сплайсингу доменів SH2 людини. Наведені елементи вторинної структури (α-спіралі (синій) і β-складки та петлі (зелений)) домену SH2 SAP, зв’язаного з лігандом (червоний колір) (PDB: 1M27). Основні інтрон/екзонні сайти сплайсингу на елементах вторинної структури показані стрілочками [60].

підтвердив попередній аналіз функцій доменів SH2 та їх класифікацію. Наприклад, модель сплайсингу в родини FRK (складається з Brk, Srms, Frk) відрізняється від такої в родини SRC (складається із Blk, Fgr, Fyn, Hck, Lck, Lyn, Src і Yes). Члени родини FRK мають подібні амінокислотні послідовності до SFKs та частково подібну організацію білкових доменів, однак відрізняються функціонально [112].

Slap та Src є подібними за амінокислотними послідовностями та моделями сплайсингу, що вказує на близькі еволюційні зв'язки. Водночас, вони є членами різних родин, оскільки мають різну доменну організацію. Зазначимо, існують винятки у класифікації, якщо група містить Crk та CrkL, або Socs та Cish, що пов'язано з відсутністю чітких даних щодо їх структур. Хоча наявні дані свідчать, що Crk та CrkL слабко відрізняються моделями сплайсингу, однак, можливо, вони повинні знаходитися у різних групах. SOCS родина згрупована разом, не зважаючи на те, що у доменах SH2 протеїнів Socs1, Socs3, Socs4, Socs5 та Socs6 не виявили сайт сплайсингу. Крім того, сплайсинг сайт Socs2 пов'язаний із Cish, а сплайсинг сайт для Socs7 є унікальним [60].

1.3. *In silico* методи дослідження

1.3.1. Вирівнювання амінокислотних послідовностей протеїнів. Комп'ютерні методи вирівнювання послідовностей можна поділити на два типи: глобальні (шукає відповідності амінокислотних залишків по всій довжині досліджуваних послідовностей) та локальні (ідентифікує подібні ділянки у межах довгих послідовностей, які суттєво відрізняються на більшій частині своєї протяжності) вирівнювання. Для глобального вирівнювання застосовують алгоритм Нудельмана–Вунша [113], а для локального — Сміта–Вотермана [114].

Залежно від кількості послідовностей, що порівнюються, розрізняють попарне та множинне вирівнювання. Існує три основних методи для реалізації попарного вирівнювання: точково-матричний (вирівнювання достатньо подібних послідовностей) [115], динамічне програмування (як локальне, так і глобальне поетапне вирівнювання: (1) ініціалізація, (2) заповнення скоринг матриці, (3) вирівнювання) [116] і словесні методи [117]. Останні є найуживанішими. Вони визначають серію коротких амінокислотних послідовностей («слів»), які не перекриваються у структурі, що порівнюються з базою даних послідовностей. Їх найчастіше використовують у пошукових інструментах баз даних FASTA і BLAST [118].

Множинне вирівнювання застосовують для одночасного порівняння двох і більше послідовностей. Це дозволяє виявити консервативні ділянки, які можна використати для визначення каталітичних активних сайтів ензимів та еволюційної спорідненості протеїнів. До методів множинного вирівнювання належать: динамічне програмування (*див. вище*), прогресивні методи [119] (поступове порівняння послідовностей за спаданням їхньої подібності), ітегративні методи [120] (поетапне вирівнювання, коли спочатку вирівнюються і розбиваються на групи усі послідовності, а далі вирювнюється кожна група окремо) та профільний аналіз [121] (пошук коротких консервативних ділянок амінокислотних залишків).

У випадку множинного вирівнювання послідовностей застосовують алгоритм загальної оптимізації — приховану модель Маркова. Його покладено в основу програми FSA, яка спочатку здійснює попарне порівнювання послідовностей [122], а потім використовує технологію відбору послідовностей для реалізації множинного вирівнювання [123]. Цей підхід вирівнювання значно точніший порівняно з вищеописаними завдяки використанню алгоритму «крутого підйому» [117].

1.3.2. Вирівнювання структур протеїнів. Структурне вирівнювання дозволяє точніше визначити функціональну та еволюційну спорідненість протеїнів. Для цього потрібна інформація щодо їх вторинної і третинної структури. Тому цей підхід застосовують лише до послідовностей з експериментально встановленою просторовою структурою, зокрема для передбачення функцій протеїнів, гомологічного моделювання [124] та структурної класифікації [125], порівнювання двох і більше структур та локальних вирівнювань їх амінокислотних послідовностей. До методів структурного вирівнювання протеїнів відносять: DALI (попарне структурне вирівнювання протеїнів та конструювання баз даних структурного вирівнювання FSSP [126]); SSAP (дворазове застосування динамічого методу для визначення оптимальних локальних вимірювань, які потім використовують у сумарному вимірюванні структур [127]). Алгоритм використовують для конструювання ієрархічних баз даних протеїнових укладок CATH [128, 129]; метод комбінаторного подовження (попарне та множинне вирівнювання протеїнових структур, які базуються на визначенні відстані між амінокислотними залишками, характеристиках сусідніх амінокислотних залишків). Алгоритм використовують для сканування баз даних та детального аналізу великих родин [130].

Частина методів множинного вирівнювання дозволяє проводити глобальне попарне порівнювання структур та одночасно розглядати усі молекули структур, які беруть участь у порівнюванні: PrISM [131], MUSTA [132], MultiProt [133] і MASS [134]. Ці методи дозволяють визначати консервативні ділянки незалежно від послідовності.

Здебільшого при структурному вирівнюванню протеїни приймають за жорсткі тіла, хоча їхні конформаційні зміни мають важливе функціональне значення. У цьому випадку розглядають два типи таких змін: швидкі переміщення бічних ланцюгів та повільні великомасштабні згинання у шарнірній ділянці (важливо для розпізнавання сайтів зв’язування та

моделювання інгібіторів чи лікарських засобів). За допомогою алгоритму гнучкого вирівнювання структур розв’язують три завдання: (1) вирівнювання амінокислотних залишків; (2) співставлення (сурперпозиція) протеїнів; (3) ідентифікація шарнірних ділянок.

1.3.3. Розпізнавання функціональних сайтів. Подібність просторових структур протеїнів часто вказує на подібність функцій, хоча інколи протеїни з подібною укладкою, наприклад, ТІМ-бочка, відіграють різну біологічну роль [135]. Водночас, протеїни з різними укладками здатні виконувати однакові функції. Припускають, що саме ділянки зв’язування визначають подібність функції протеїнів. Дослідження у цьому напрямку були використані для пошуку потенційних лікарських засобів. Однак потрібно зазначити, що досить часто отримані сполуки приводять до побічних ефектів. Вважають, що простий шлях розв’язати цю проблему – це аналізувати масиви таких протеїнів (структурно різних), зв’язаних з тим самим лігандом. Порівнювання таких комплексів дозволяє вирівняти їхні ліганди і сайти зв’язування [136], що є важливим під час дослідження взаємодіючих функціональних груп. Проте недоліком цього підходу є можливість альтернативного зв’язування у кишені протеїну [137]. Тому для розуміння такого внутрішньоклітинного явища необхідне детальне дослідження фізико-хімчних властивостей амінокислотних послідовностей, їхніх взаємодій тощо.

1.3.4. Молекулярний докінг. Молекулярний докінг дозволяє передбачити можливі варіанти взаємодії молекул, оцінити просторову структуру їх комплексів та афінність взаємодії. Цей метод є ефективними для пошуку лікарських засобів. Найчастіше його використовують для

віртуального скринінгу бібліотек органічних речовин, оцінки токсичності та вивчення PPI.

Існує припущення, що надмолекулярні комплекси краще моделювати, використовуючи підходи квантової механіки (розв'язання рівняння Шредінгера для досліджуваних систем). Але це досить складний підхід. Тому здебільшого використовують класичну механічну модель, яка розглядає атоми як ньютонівські частинки і не враховує квантові взаємодії [138]. Потенціальна енергія взаємодії залежить від довжини зв'язків, валентних і торсійних кутів і нековалентних взаємодій.

Силові поля (довжина зв'язків, валентні кути, величина парціальних зарядів, ван-дер-вальсові параметри) використовують, щоб спрогнозувати формування комплексів. Приклади таких силових полів: AMBER (розроблено для протеїнів та нуклеїнових кислот; поле дозволяє враховувати кожен атом окремо або формувати кластери атомів з однаковими властивостями [139]); CHARMM (розроблено для протеїнів, ДНК, РНК та ліпідів; у додачу до попереднього це поле включає поляризовані силові поля, які базуються на моделі заряду або дисперсійній моделі осцилятора [140]); CVFF (використовують для розрахунку енергетичних параметрів різних низькомолекулярних органічних сполук [138]); CFF (використовують для оптимізації структур, враховуючи довжини зв'язків, валентні та торсійні кути, відстань між атомами, потенціальну енергію, дипольний момент та частоту коливань [141]) та COMPASS [142].

Існують узагальнені силові поля з невисокою точністю. Їх застосовують до систем, що не описуються більш точними моделями силового поля. Прикладами узагальнених полів є ESFF [143] і UFF [144] (їх створено для моделювання органічних, неорганічних та металоорганічних систем).

1.3.5. Алгоритми молекулярного докінгу. Головне завдання молекулярного докінгу – це пошук можливих варіантів зв'язування між структурами на молеклярному рівні організації. Алгоритм повинен знайти якнайбільше число зв'язувань між молекулами. Але це є занадто ресурсоємний процес. Для його оптимізації створили такі алгоритми молекулярного докінгу: молекулярна динаміка, метод Монте Карло, генетичний алгоритм, алгоритми, засновані на фрагментах, алгоритми комплементарності точок, пошук заборон, систематичний пошук.

Алгоритми молекулярної динаміки розв'язують рівняння руху Ньютона. Мінімальну енергію комплексу знаходять шляхом стандартної оптимізації.

Метод Монте Карло представляє групи чисельних методів, які базуються на отриманні масиву реалізацій стохастичного процесу. Цей метод генерує випадкові рухи системи, після чого фільтрує їх за ймовірністю перебування частинок у тому, чи іншому стані [145].

Генетичний алгоритм – еволюційний алгоритм пошуку, який застосовують для послідовного підбору та комбінування досліджуваних параметрів. Основна його ідея полягає в еволюції кількості можливих розв'язків через генетичні зміни (мутації) до отримання фінального результату. Спочатку кодують змінні (наприклад, ступені вільності у «генетичний код»). На наступному етапі відбирають випадкову кількість розв'язків, які оцінюють за енергетичними параметрами. Відзначимо, що генетичні алгоритми добре підходять для паралельного обчислення [146].

Методи, засновані на фрагментах, дозволяють кластеризувати ліганди. Тут головну роль відіграє підбір основного скефолду — функціональної групи ліганду (утворює ключові взаємодії з рецептором [147]).

Алгоритм комплементарності точок базується на оцінці стеричної та хімічної комплементарностей при взаємодії структур між собою. Крім того, існує можливість застосувати додаткові умови зв'язування, наприклад, спрямувати нормалі взаємодіючих структур у протилежні напрямки [148].

Метод пошуку заборон базується на дослідженні стохастичних процесів – генерації нових станів (зі стабільного/вихідного стану) системи при використанні набору заборон. Отримані результати оцінюють і сортують. Найкращий розв’язок вважають за новий вихідний стан системи і повторюють цикл. Набір заборон діє, як пам’ять, що містить інформацію стосовно попередніх вихідних станів [149].

Узагальнення до огляду літератури

Підсумовуючи дані літератури, можна зауважити наступне.

1. PPI має важливе фундаментальне і практичне значення. Зрозуміти природи такого процесу можна при застосуванні комплексу *in silico* та *in vitro* методів.

2. Домени SH2 є однією з понад 100 різноманітних родин доменів, які здатні до специфічних взаємодій та містяться у великій кількості протеїнів. Вони характеризуються сталою структурою (дві α-спіралі та сім β-складок). Зв'язування фосфотирозину, що відбувається у строго визначених місцях домену (рис. 1.2), робить їх зручними мішенями для біо/хемоінформатичних досліджень.

3. У сучасних дослідженнях широко застосовують комбінування різних методів біо/хемоінформатики. Описані підходи є як специфічними для конкретних мішеней (наприклад, протеїни), так і універсальними (для різних молекул, наприклад, ДНК, РНК, ліпіди та низькомолекулярні сполуки). Структурний аналіз є кращим за вирівнювання послідовностей і дає змогу більш точно оцінити функціональну та еволюційну подібність протеїнів. Молекулярний докінг широко використовують для пошуку нових лікарських засобів. Водночас, алгоритми докінгу потребують удосконалення, зокрема підвищення точності, швидкості, можливостей врахування гнучкості взаємодіючих молекул тощо.

Результати цього розділу викладені в оглядовій статті [11].

РОЗДІЛ 2
МАТЕРІАЛИ ТА МЕТОДИ

2.1. Апаратне та програмне забезпечення

Усі розрахунки проводили на обчислювальному кластері Київського національного університету імені Тараса Шевченка [150]. Використовували базу хімічних речовин фірми Enamine, яка містить 1,2 млн. [151].

Основні етапи розрахунків виконані за допомогою оригінальних програм та bash-скриптів [152], написаних на мові C++ та java [153].

2.2. Отримання 3D структур доменів SH2 та їх аналіз

2.2.1. Отримання сету 3D структур доменів SH2. Вихідні 3D структури доменів SH2 взяті з баз PDB [154] та Pfam [155]. Загалом відібрали 219 структур доменів SH2: з них 66 структур одержані методом ЯМР у розчині і 153 – методом рентгеноструктурного аналізу. Ці структури містять інформацію про 67 доменів, які належать до 65 протеїнів дев’ятьох видів організмів. Для найкращого використання структурних даних врахували, що деякі структурні файли містили декілька різних копій одного і того ж домену. Так, наприклад, структура 2K7A [156] містить 18 копій своєї структури в одному pdb файлі. Тому для більш повного використання отриманої інформації з даних ЯМР структури розбили на окремі моделі, а структури, отримані методом ренгеноструктурного аналізу, розбили на окремі протеїнові ланцюги. Загалом отримали 1129 протеїнових структур, які використали на наступних етапах дослідження.

2.2.2. Порівнювання отриманих структур доменів SH2 та їх поділ на групи за подібністю амінокислотних послідовностей. Усі отримані структури порівнювали. Враховуючи те, що кожен кристал містить декілька копій (які інколи відрізняються за структурою), спочатку порівнювання проводили в межах кожного окремого ID номера структур у програмному пакеті ClustalX (порівнювання амінокислотних послідовностей) [157]. Так, виконали множинне порівнювання аміноксилотних послідовностей (MSA) [158] за прогресивним методом [159]. Під час розрахунків використовували повільний [160], але більш точний алгоритм вирівнювання амінокислотних послідовностей. Порівнювання відбувалося у три етапи: (1) на першому етапі всі послідовності порівнювали одна з одною методом попарного порівнювання [161] (це відбувалося у матриці замін BLOSUM (застосовується на всіх етапах порівнювання)). У випадку, якщо порівнювані амінокислоти у певній точці матриці замін попадали на пропуск (мала розрахункова ймовірність такої заміни, або наявність вставок), тоді до оцінки якості вирівнювання застосовували такі штрафи: Gap Open (оцінка якості усіх попарно вирівнюваних послідовностей) – 10 та Gap Extension (оцінка попарного вирівнювання для кожного додаткового залишку при пропуску) – 0,1; (2) на другому етапі будували дендрограму, яка описує наближений поділ структур на групи за подібністю їх амінокислотних послідовностей. Відзначимо, що крім вище зазначених штрафів якості вирівнювання на цьому етапі додаються ще декілька, а саме: Gap Distances (якщо амінокислотні пропуски знаходяться на відстані, меншій ніж 5 позицій) – 5, No End Gaps (заборона штрафу, якщо пропуск у кінці послідовності) – false, KTUP (оцінка подібності між порівнюваними послідовностями) – 1, Iteration (тип ітерацій) - tree, Num Iter (максимальна кількість ітерацій, що проводиться на кожному кроці) – 1, Clustering (тип поділу на групи амінокислотних послідовностей) – NJ (метод приєднання сусідів [162]); (3) третій етап характеризується множинним вирівнюванням послідовностей згідно результатів їхнього поділу на групи, отриманих на попередньому етапі.

Найбільш подібні послідовності порівнювали в першу чергу, після чого додавали менш подібні послідовності (з урахуванням попарного вирівнювання), що у подальшому використовували при множинному вирівнюванні. Відзначимо, що така послідовність дій дозволяє покроково оцінювати амінокислотну варіативність у кожній окремій позиції.

Крім того, порівнювали геометричну подібність отриманих структур, використовуючи програмний пакет Chimera (вимірювання значення Rmsd між структурами) [163]. Зауважимо, що при такому порівнюванні на одну амінокислоту вирівнювали лише один атом (C^{α}), що дозволяло оцінити конформаційні відмінності між порівнюваними структурами.

Аналогічна процедура була застосована до структур з різними ID номерами і, таким чином, отримали 56 різних структур. З використанням програми ClustalX вони були повторно вирівняні та відредаговані, використовуючи Jalview [164], і поділені на 6 груп за подібністю амінокислотних послідовностей. У цих групах виділили ключові точки (кишені зв'зування), які використали як дескриптори під час поділу кишень зв'язування на кластери (локальне порівнювання).

2.2.3. Конформаційний пошук доменів SH2. Вісім доменів SH2 використали для конформаційного пошуку: 2KK6 (solution structure of sh2 domain of proto-oncogene tyrosine-protein kinase, Homo sapiens), 3IN7 (crystal structure of the Grb2 SH2 domain in complex with a Cyclopropyl-constrained Ac-pY-Q-N-NH2 Tripeptide Mimic, Homo sapiens), 2K7A (complex between the SH3 and SH2 domain of interleukin-2 tyrosine kinase, Mus musculus), 2JYQ (NMR structure of the apo v-Src SH2 domain, Rous sarcoma virus), 2GE9 (solution structures of the SH2 domain of Bruton's Tyrosine Kinase, Homo sapiens), 2FCI (Syk tyrosine kinase, Bos taurus), 1UUS (structure of an activated D. STAT in its DNA-unbound form, Dictyostelium discoideum) та 1O49 (crystal structure of SH2 proto-oncogene tyrosine protein kinase SRC, Homo sapiens).

Пошук потенційно можлививх конформацій виконали у програмному пакеті MOE [165] з використанням Low Mode Search (коротка молекулярна динаміка, 0.1 пс) методу. Так, можливі конформації доменів SH2 генерували згідно наступних опцій:

- Rejection Limit – 100 (число суміжних спроб, які потрібно виконати, щоб згенерувати нову конформацію перед зупинкою пошуку);
- Iteration Limit – 10 (максимальна кількість спроб для генерації нової конформації; не залежить від новизни кожної згенерованої конформації);
- RMS Gradient – 0.005 (якщо середньоквадратичний градієнт падає нижче зазначеного значення, тоді мінімізація енергії зупиняється);
- MM Iteration Limit – 50 (максимальне число мінімізацій енергії, яке застосовують до кожної згенерованої конформації);
- RMSD Limit - 0.25 Å, 0.5 Å і 0.75 Å (якщо між двома конформаціями Rmsd менше, ніж зазначене значення, тоді вони є ідентичними структурами).
- Strain Cutoff – 7 (енергетично мінімізована конформація відкидається, якщо її потенціальна енергія більша, ніж E_{min+s}, де E_{min} – найменша енергія серед мінімізованих конформацій, а s – значення, вказане для Strain Cutoff (Ккал/моль)).

Отримані конформації мінімізували у декартовій системі координат.

*2.2.4. Кластеризація доменів SH2 методом k-середніх.*На противагу поділу за подібністю амінокислотних послідовностей, провели кластеризацію дескрипторів сайтів (ключові амінокислоти виділені при їх порівнюванні (*див. 2.2.2*)) зв'язування на основі заданої метрики методом k-середніх [166]. Цей метод застосували для оцінки розподілу форми сайтів зв'язування доменів [167].

Порівняно з попереднім методом дослідження, він має такі відмінності:

- враховує лише геометричні характеристики досліджуваних структур, що дає змогу оцінити геометричне різноманіття сайту зв'язування доменів SH2;
- не використовує накладання структур згідно амінокислотних послідовностей, що дозволяє порівнювати домени зі значними відмінностями.

Уся поверхня зв'язування була поділена на триплети молекулярної поверхні (структури на молекулярній поверхні, що складаються з трьох точок). Випадковим чином з усіх вершин відібрали 107 груп по 3 вершини у кожній. Водночас, ці вершини поділили за довжиною їхніх сторін на різні групи (довжини сторін округлювали до найближчого меншого цілого числа в ангстремах). У випадку, якщо довжина сторони утвореного трикутника перевищувала 16 Å, він відкидався. Як результат, виділені вершини були поділені на 508 груп. 508-вимірний вектор, компоненти якого відповідають числу триплетів у кожному з кошиків, використовували як дескриптори сайту зв'язування.

Дескриптори порівнювали, використовуючи Евклівдову метрику, за якою різниця між векторами a та b обчислюється за формулою:

$$g(a,b) = \sqrt{\sum_{i=1}^{508}(a_i + b_i)^2}.$$

Передбачалося, що в межах розрахунків відстань між двома дескрипторами буде збільшуватись зі зростанням відмінностей між формою порівнюваних сайтів зв'язування.

Так, для усіх отриманих дескрипторів спочатку відібрали 8 випадкових центрів. На наступному етапі усі дескриптори поділили на 8 груп за їх подібністю до кожного з вибраних центрів. Потім для кожного кластеру визначали центр заново, як середнє з усіх його елементів. Після цього розподіл за кластерами повторювали доти, поки після визначення центру та перерозподілу дескрипторів за кластерами їхні центри не змінювалися.

2.2.5. Аналіз вирівнюваних амінокислотних позицій одержаних груп. Несуперечливість амінокислотних послідовностей в межах кожної окремої групи оцінювали з використанням PRALINE онлайн серверу [168]. У цьому випадку значення подаються як певна шкала при вирівнюванні (1-9; якщо зустрічається значення 10, тоді це означає 100% подібність амінокислот у певній позиції).

Консервативність розраховували з використанням Scorecons серверу [169]. У цьому випадку обчислювали значення можливої ентропійної складової вирівнюваних послідовностей. Оцінка ентропії, яку використовували, заснована на нормалізації ентропії Шенона [170]. Отже, якщо консервативність складає 1 (можливе максимальне значення у цьому випадку – 100%), це означає, що структури є досить подібними (низька ентропія порівнювання). У випадку, коли консервативність складає 0 (мінімальне можливе значення консервативності, абсолютно не подібні структури, у цьому випадку – 0%), це свідчить про значне різноманіття структур у цій позиції і, як наслідок, про значну ентропійну складову порівнювання. Заначення ентропії (у цьому випадку, фактично, консервативності) для кожної конкретної позиції на амінокислотній послідовносі визначали за формулою

$$C_{entropy} = \frac{-\sum_a^K P_a log_2 P_a}{log_2(\min(N,K))}.$$

де N – число амінокислотних залишків позиції, K – кількість типів амінокислотних залишків позиції, $P_a=n_a/N$ (n_a – число амінокислотних залишків типу a). Однак, зазначимо, що у цій частині роботи проводили два варіанти таких розрахунків. У першому - амінокислоти поділяли на K=21 тип (20 стандартних та 1, що є можливим пропуском послідовності за її вирівнювання). У другому - амінокислоти поділяли на K=7 типів (аліфатичні (AVLIMC), ароматичні (FWYH), полярні (STNQ), позитивно заряджені (KR), негативно заряджені (DE), зі спеціальною неоднозначною конформацією (GP) та пропуски).

Усі розрахунки виконані в матриці замін BLOSSUM62 [114]. Однак перед тим, як цю матрицю замін можна було б використовувати, її трансформували у зручний діапазон значень. Для цього застосували Karlinlike трансформацію [171].

Крім того, провели розрахунок подібності амінокислотних послідовностей, для чого використали параметр ідентичності [172]:

$$I = 100\left(\frac{identical\ positions}{Length\ of\ the\ aligment}\right).$$

2.2.6. Визначення оточення сайту зв'язування доменів SH2. Визначення оточення сайту зв'язування доменів виконали з використанням програмних пакетів Chimera [173] і helixweb [174, 175].

Амінокислотні залишки, що містять хоча б один атом, який бере участь у формуванні сайту зв'язування протеїну (навіть якщо він займає мінімальну площу сайту зв'язування), визначаються як такі, що беруть участь у формуванні сайту зв'язування. Амінокислоти, що беруть участь у

формуванні сайту зв'язування, знаходили шляхом оцінки значення їхньої ASA [176].

Так, якщо амінокислота має площу поверхні більшу, ніж 5 $Å^2$ та знаходиться на відстані в межах 5 Å, або менше до будь-якого атома зв'язаного ліганду (відносно міцна взаємодія), вона визначається як амінокислота зі значним внеском у формування сайту зв'язування. З іншого боку, амінокислоти, які мають меншу площу, або знаходяться на більшій відстані від будь-якого атома зв'язаного ліганду, визначаються як допоміжні структури без значного внеску у формування сайту зв'язування.

2.3. Підготовка бази хімічних речовин до проведення молекулярного докінгу

2.3.1. Відбір бібліотеки потенційних лігандів. Як пул з уже готовим набором потеційних лігандів зі значною кількістю структурних модифікацій, використали базу хімічних речовин підприємства Enamine [151]. Також при відборі речовин для молекулярного докінгу врахували те, що деякі варіанти моделювання є достатньо ресурсоємними, тому для їхнього проведення скоротили сет бази хімічних речовин. Крім того, поетапні розрахунки були застосовані до всієї бази. Але спочатку застосували фільтрування (за ADME параметрами) до бази хімічних речовин. Для цього використали програму JChem. Фільтрування проводили за такими параметрами: Mw (молекулярна вага), LogS (розчинність у воді), LogP (величина, що вказує на ліпофільність

Таблиця 2.1.

Молекулярні параметри відбору лігандів для проведення віртуального скринінгу

Параметри відбору	Min	Max
Mw	300	650
LogS	-6.5	-3
LogP	1	6
hb_a	4	11
RB	3	9

або гідрофільність речовини, тобто за цим параметром можна судити про її проникність крізь плазматичну мембрану), hb_a (акцептори водню), RB (обертальні зв'язки) (табл. 2.1). Після фільтрування відібрали близько 1 млн. речовин.

Для докінгу використали три варіанти відбору бази потенційних лігандів: (1) відбір для домену SH2 протеїну STAT3 (тестовий етап дослідження, наймеш ресурсоємний) - відібрали речовини, що мають максимальну кількість акцепторів та відповідають правилам Ліпінскі; (2) наявна база хімічних лікоподібних речовин була поетапно "стиснута" за різноманітністю до 50 тисяч структур (перший варіант основних розрахунків) (*див. нижче*); (3) застосували поетапний докінг, що дало змогу провести моделювання з усією наявною базою, - найбільш ресурсоємний варіант розрахунків (у цьому випадку не використовували фільтрування за ADME параметрами).

Вибірку для другого варіанту досліджень проводили за такими основними критеріями: об'ємна форма молекул, хімічна структура і фармакофорні властивості. Зменшення бази речовин проводили за такими етапами:

- застосування 169-бітних структурних відбитків MACCS [177] для скорочення попередньо профільтрованої бази до 500 тисяч структур;
- застосування програмного пакету ConfGen для генерації стереоізомерів [178]. Розрахунки проводили за таких параметрів: максимальна

відносна енергія складала 50 кДж/моль, мінімальне середньоквадратичне відхилення між двома атомами – 0.4 нм. Згенерували приблизно 3.5 млн. конформацій (близько 7 конформацій для кожної структури). Після цього застосували трьохточкові фармакофорні відбитки piDAPH3 [177] для опису отриманих конформацій. З урахуванням критеріїв тривимірного розподілу фармакофорних властивостей молекул (донор, заряд, акцептор водневих зв'язків, π-взаємодії) одержали вибірку у 200 тис. речовин;

- використання просторових відбитків EShape3D [177] і, таким чином отримали 48512 сполук.

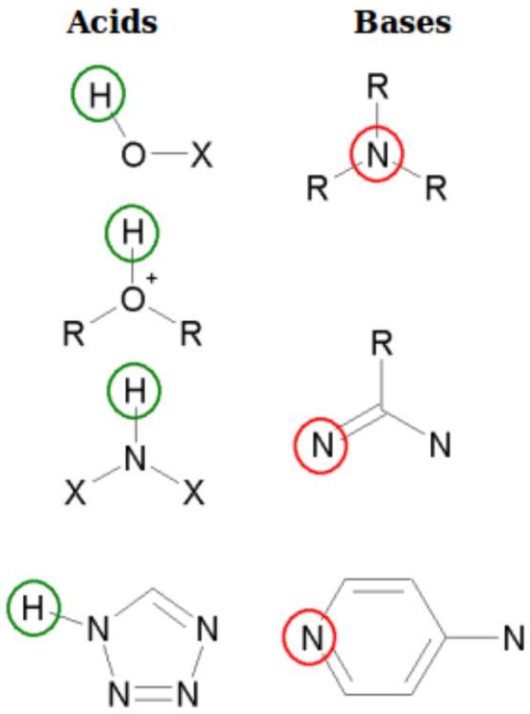
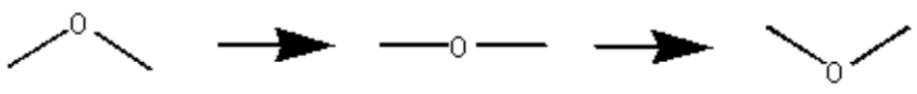


Рис 2.1. Хімічні фрагменти, що класифікуються як кислоти та основи.

2.3.2. Підготовка отриманих баз хімічних речовин до моделювання. Підготовку отриманих баз хімічних речовин до молекулярного докінгу проводили згідно такого алгоритму. Перед початком розрахунків усі речовини наблизили до найімовірнішого хімічного стану у водному середовищі: для цього провели протонування основ та депротонування

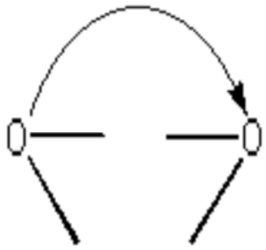
кислот (наприклад, аміни визначалися як кислоти, якщо вони містять два сильних електрон – акцепторних замісники, але, якщо присутня комбінація водню з sp3 замісником вуглецю, аміни вважалися основами). Для оксикислот приєднання X визначалось як конюгована електрон –акцепторна група (аналог замісника вуглецю). N=C-N скефолд вважався основою, якщо структура була аліфатичною і не містила електорон-акцепторних замісників. При цьому функціональні групи з характерними значеннями pKa=7 (не можна точно визначити, чи вони повинні бути протоновані, чи депротоновані) нічого не змінювали. Так, наприклад, хімічні фрагменти можна класифікувати як сильні кислоти, або основи (рис. 2.1). Фактично це дозволяє нейтралізувати кислоти і основи. По-перше, робиться спроба знайти стан протонування, при якому всі атоми є нейтральними. Якщо це не можливо, намагалися знайти стан протонування, при якому лише деякі частини молекули були б повністю нейтральними.

Потім побудували 3D структури молекул та провели їх енергетичну мінімізацію, використовуючи опцію «чотиривимірної мінімізації енергії». Процедуру проводили у силовому полі MMFF94x [179] за Rmsd градієнта 0.01 нм. Візуалізація чотиривимірного простору є складним завданням, тому спробуємо провести таку аналогію. Припустимо, що молекули існують лише у двовимірному просторі (наприклад, стандартний хімічний запис на папері). Розглянемо перетин енергетичного бар’єру між двома конформаціями молекули C-O-C:



Коли атоми перебувають на одній прямій, потенціальний бар’єр високий, завдяки чому молекулі важко його подолати для зміни конформації.

Водночас, при додавані третього виміру створюється додатковий шлях з низьким рівнем енергії між двома конформаціями:



Іншими словами, з’являється можливість руху між двома атомами без зміни кута між ними, зберігаючи сталу енергію комплексу. Ця ж ідея застосовується при переході від тривимірного до чотиривимірного простору (тобто, при такому переході існує можливість подолати енергетичний бар’єр в іншому просторі з меншими затратами енергії, завдяки чому можна мінімізувати молекулярну систему у чотирьох вимірах). Однак, у цьому випадку можливими є такі конформаційні зміни, за яких атоми, хімічні/фізичні зв’язки будуть перетинатися, що є більш складними сценаріями конформаційних змін молекул.

Отже, мінімізація у чотиривимірному просторі відбувається за таким алгоритмом: (1) ініціалізація: кожному атому молекули присвоювали координати у чотиривимірному просторі (x_i,y_i,z_i,w_i); (2) 4D-мінімізація енергії та побудова 4D-конформацій; (3) мінімізація функції U(x,y,z,w) (потенціальна енергія молекули); (4) 3D-мінімізація: видаляється w (четверта координата виміру) і після цього у тривимірному просторі мінімізується функція U (x,y,z).

2.4. Моделювання комплексів типу “домен SH2-ліганд”

Для першої (моделювання проводили з доменом SH2 протеїну STAT3) і другої (48512 різноманітних сполук) вибірок хімічних речовин моделювання

комплексу “домен-ліганд” проводили лише з використанням програмного пакету flo+ [180]. Розрахунки з використанням усієї бази проводили у декілька етапів (*опис див. нижче*):

- застовування пакету MOE [165, 181];
- застосування пакету flo+ [180];
- побудова фармакофорних моделей;
- віртуальний скринінг у програмному пакеті LigandScout [182];
- GTM-передбачення [183, 184].

Перед проведенням молекулярного докінгу з відібраних протеїнів видалили усі молекули води, а всі Arg і Lys запротонували. В обох випадках використали гнучкий ліганд і фіксований рецептор. Для того, аби переконатися, що докінг є якісним, за допомогою пакетів програм MOE і flo+ провели процедуру редокінгу.

2.4.1. Застосування пакету MOE. Усі розрахунки проводили у силовому полі MMFF94x [185]. На першому етапі використали стандартні параметри проведення докінгу в рамках пакету MOE. Спочатку всі ліганди докувались відповідно до примітивних фармакофорних моделей (містять 2-3 точки зв’язування). Зв’язування проводили з усією базою лікоподібних речовин підприємства Enamine. Також використовували наявні дані зв’язування доменів з нативними лігандами. Здебільшого, як основні точки простої моделі, брали pTyr-зв’язуючий сайт та одну з двох гідрофобних кишень. У процесі докування використовували протокол “flexible ligand docking” [186].

На наступному кроці оцінювали вільну енергію зв’язування комплексу “домен SH2- ліганд” у кожній з позицій ліганду, які задовольняли принятим обмеженням. Для цього використали London dG Scoring функцію [165]

$$\Delta G = c + E_{flex} + \sum_{h-bonds} c_{hb} f_{hb} + \sum_{m-lig} c_M f_M + \sum_{atoms\ i} \Delta D_i,$$

де c – середнє значення збільшення/зменшення обертальної і трансляційної ентропії, E_{flex} – енергія втрати гнучкості лігандів (розраховується лише з використанням топології лігандів, - фактично, топологічна оцінка ентропії ліганду), f_{hb} – характеризує геометричні недоліки водневих зв'язків, може набувати значень 0,1, c_{hb} – енергія ідеальних водневих зв'язків, f_M – характеризує недоліки зв'язування типу "ліганд-метал" і може набувати значень 0,1 (зазначимо, що у нашому випадку такого типу зв'язування не існує), c_M – енергія ідеального зв'язування металу з лігандом, D_i – десольватаційна енергія атома *i*. Після завершення розрахунків зберігали 30 енергетично найвигідніших комплексів.

Потім отримані комплекси фільтрували та провели енергетичну мінімізацію (за законами молекулярної механіки): якщо зсув підмножин фармакофорних моделей, а, відповідно, і молекул зв'язаної структури, перевищував 0.4 Å – вони відкидались. Після цього повторно оцінювали вільну енергію зв'язування і залишали лише один енергетично найвигідніший комплекс.

2.4.2. Застосування пакету flo+. Застосували алгоритм систематичного докінгу (sdock+) [187]. Усі розрахунки проводили в силовому полі AMBER. Перед початком розрахунків для кожної мішені виділяли ключові атоми для взаємодії з лігандом. Далі для кожної оптимізованої сполуки максимально генерували 200 комплексів, використовуючи метод Monte Carlo [188]. На наступному етапі, згідно внутрішніх скоринг-функцій QXP, відбирали 10 найбільш енергетично вигідних комплексів [180], які залишали для подальшого аналізу. Для кожного окремого варіанту докінгу оцінювали

енергію комплексу “протеїн-ліганд”, енергію зв’язування протеїну з лігандом, внутрішню енергію напруження ліганду, енергію ван-дер-вальсових та електростатичних взаємодій, кількість Н-зв’язків між атомами протеїну та ліганду (при відборі потенційних лігандів особливу увагу приділяли молекулам з великою кількістю акцепторів Н-зв’язку) і, нарешті, оцінювали площу ліганду, яка контактує з протеїном.

Важливо підкреслити, що під час розрахунків врахували діапазон рухливості взаємодіючих структур, який може бути різним, починаючи з невеликих бічних ланцюгів і закінчуючи масштабними доменними рухами. Одержані результати відфільтрували multyRmsd та внутрішніми фільтрами flo+ [189]. Фільтрування проводили за таким алгоритмом: (1) застосування фільтру за кількістю водневих зв’язків: (ліганд повинен мати не менше трьох таких зв’язків з протеїном); (2) зв’язування між конкретними атомами протеїну та ліганду (зазвичай у pTyr-зв’язуючій та гідрофобній кишенях); (3) результуюча вільна енергія комплексоутворення не повинна перевищувати -15 кДж/моль. Зазначимо, що інші фільтри застосовували залежно від особливостей кишені зв’язування тієї чи іншої PDB структури.

2.4.3. Побудова фармакофорних моделей. Для побудови структурно-залежних фармакофорних моделей використали усі наявні PDB-структури, зв’язані з лігандами: 1SKJ [190], 1BKM [191] (PP60 V-SRC Tyrosine kinase transforming protein), 1O49, 1O48, 1O47, 1O46, 1O44, 1O43, 1O42 [192] (proto-oncogene tyrosine-protein kinase SRC), 1IJR [193], 1FBZ [194] (proto-oncogene tyrosine-protein kinase LCK), 1A1E [195] (C-SRC tyrosine kinase). Після цього за допомогою пакету UCFS Chimera провели порівнювання відібраних структур. Далі їх використали для побудови фармакофорних моделей у програмі LigandScout [182], яка використовує алгоритм щодо інтерпретації ліганд-рецепторних взаємодій (водневі зв’язки, ван-дер-вальсові взаємодії, гідрофобні, ароматичні частини) і, зрештою, подає їх у вигляді зображення

Кекуле (Kekulé pattern) [196]. Згенерована модель включає основні частини зв'язування домену з лігандом [197]. Також до фармакофорних моделей додали опцію "excluded volumes" для покращення ефективності віртуального скринінгу [198, 199].

Для побудови ліганд-залежних фармакофорних моделей відібрали усі відомі активні та неактивні сполуки щодо доменів SH2. Загалом одержали 78 активних і 38 неактивних речовин. Для них, використовуючи пакет LigandScout Omega, для значення внутрішньої енергії 10 ккал/моль згенерували усі можливі стереоізомери (для кожної речовин обмежували кількість можливих конформацій 500, при цьому межа значення Rmsd складала 0.4 нм). Далі провели кластеризацію структур (для кожної структури відібрали 25 енергетично найбільш вигідних стереоізомерів).

2.4.4. Генерація декоїв. Процедуру виконали за допомогою інтернет-ресурсу «DUDE decoy generation». Спочатку порівняли параметри ADME декоїв та структур лігандів (молеклярна вага, LogP, обертальні зв'язки, акцептори Н-зв'язку, донори Н-зв'язків і заряди). Усі декої згенеровано в діапазоні pH 6-8 з використанням програмного пакету «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 відібрані за допомогою пакету ZINC45 [200], використовуючи протокол динаміки, застосований до локального хімічного простору, збільшуючись або змешуючись відносно відібраних вище шести параметрів. 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 (за коефіцієнтом Танімото (Tc)) відносно будь-якого ліганду. Найбільш неподібні декої, для яких Tc≤25% відносно

активних речовин, відкинули. Загалом для 78 активних речовин згенерували 4200 декоїв, які використали для валідації фармакофорних моделей.

2.4.5. ROC – аналіз. ROC – аналіз, як інструмент оцінки результатів віртуального скринінгу, посідає особливе місце серед бінарних алгоритмів класифікації [201]. За допомогою ROC - кривих визначали здатність фармакофорних моделей поділяти речовини на активні та неактивні, розраховуючи значення AUC [202, 203]. ROC - криві вказують на специфічність (Sp) і чутливість (Se), які змінюються залежно від якості моделі [204, 205]. Специфічність і чутливість визначали так

$$Sp = \frac{TN}{TN + FP},$$

$$Se = \frac{TP}{TP + FN},$$

де TP – істинно позитивне значення, TN – істинно негативне значення, FP – помилково позитивне значення, FN – помилково негативне значення. Величина AUC може бути розрахована, як [206]:

$$AUC = \sum_i [(Se_{i+1})(Sp_{i+1} - Sp_i)].$$

Так, якщо $0.9 \leq AUC \leq 1$ – ідеальний результат, $0.8 \leq AUC \leq 0.9$ – добрий, $0.7 \leq AUC \leq 0.8$ – посередній, $0.5 \leq AUC \leq 0.7$ – задовільний і $AUC<0.5$ – незадовільний, який свідчить, що відсоток відібраних декоїв (FHR) перевищує відсоток відібраних активних речовин (PHR) (модель не працює). Водночас зазначимо, що така валідація нещодавно була піддана критиці [207].

2.4.6. Віртуальний скринінг. Віртуальний скринінг – потужний метод для пошуку нових потенційно активних речовин і скефолдів [208]. Для його проведення використали програмний пакет LigandScout. Розрахунки виконали за таких функціональних параметрів програми: «Pharmacophore-Fit scoring function» (враховували лише властивості підмножин фармакофорних моделей та їхні Rmsd); «Match all query features» (опція для пошуку молекул, які 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).

2.4.7. Модель GTM. Результати, одержані на попередніх етапах дослідження (сет активних і неактивних речовин, результати віртуального скринінгу), були застосовані для побудови моделей GTM за алгоритмом, запропонованого у роботах [183, 184].

2.5. Біологічне тестування отриманих структур

2.5.1. Оцінка впливу відібраних речовин на фосфорилювання STAT3. Оцінка впливу відібраних для дослідження речовин на індуковане фосфорилювання STAT3 проводили методом імуноблотАналізу [209]. Дослідження проводили на клітинній лінії L540 (лімфогранулематоз), для

якої характерний високий рівень експресії фосфорильованого STAT3. Клітини культивували за температури 37^{o} у зволоженому середовищі у змішаній атмосфері: 95% O_2 і 5% CO_2. Концентрації речовин складали 30 μM, а час впливу – 24 год.

2.5.2. Оцінка проліферативної активності різних клітинних ліній за дії досліджуваних речовин. Оцінку проліферативної активності проводили методом флуоресцентного детектування проліферації клітин з використанням резазурину. В такому експерименті кількість клітин після інкубації (з додаванням або без додавання до розчину потенційного ліганду) визначали за інтенсивностю флуоресценції резофурину, що є продуктом відновлення резазурину в клітинах. Як контроль використовували доксорубіцин, що є відомим інгібітором проліферації клітин, механізм дії якого полягає в інтеркаляції ДНК та інгібуванні біосинтезу. Експеримент проводили у 384-коміркових (40 μл/комірка) платах у дублікатах. Вихідний сигнал (інтенсивність флуоресценції) з плат конвертували у значення інгібування, після чого проводили корекцію крайових ефектів з використанням програми HTS Corrector. Результуюче значення інгібування визначали як середнє з трьох повторів. Дослідження проводили на чотирьох лініях пухлинних клітин: A431 (плоскоклітинна карцинома), K562 (хронічний мієлоцитарний лейкоз), U937 (гістоцитна лімфома) та LNCap (аденокарцинома простати). Їх висівали у таких середовищах:

K562 - 5×10^5 клітин/мл у середовищі RPMI-1640 з додаванням пеніцеліну/стрептоміцину та L-глутаміну;

U937 - 7×10^5 клітин/мл у середовищі RPMI-1640 з додаванням пеніцеліну/стрептоміцину та L-глутаміну;

A431 та LNCap - 5×10^5 клітин/мл у середовищі DMEM пеніцеліну/стрептоміцину та L-глутаміну.

Речовини доводили до кінцевої концентрації 10 мкМ. Після цього клітини інкубували 17 год у зволоженій атмосфері за 37°С та 5% CO_2.

Для визначення життєздатності вихідних тест – клітин кількість резазурину доводили до концентрації 50 мкМ та інкубували 2 год у зволоженій атмосфері за 37°С та 5% CO_2. Наявність резазурину (пропорційна числу життєздатних клітин) оцінювали шляхом вимірювання флуоресценції при E_x = 490 нм і E_m = 540 нм. Дослідження виконували з використанням FLIPR® (Fluorometric Imaging Plate Reader).

2.5.3. Визначення кіназної активності Btk та Syk кіназ. Для 80 сполук з бібліотеки SH2, обраних за результатами клітинного скринінгу, визначали їхній вплив на кіназні активності Syk та Btk (Homo sapiens). Ці кінази містять 2 та 1 домени SH2, задіяні у молекулярному механізмі кіназної активності, відповідно.

Активність Syk та Btk кіназ визначали з використанням наборів Btk (cat. V2941) та Syk (cat. V3801) kinases, отриманих від Promega corporation (Madison, WI, USA), згідно стандартного протоколу.

Дослідження починали з додавання 2 μL АТФ маточного розчину до кінцевої концентрації 50 μM. Після 1.5 год плавного збовтування (300 rpm) за температури 27°С додавали АДФ-Glo реагент (7 μL). Через 40 хв додавали люциферазу (14 μL) для оцінки люмінесценції. Після 30 хв інкубації вимірювали люмінесценцію.

Активність кіназ є пропорційною інтенсивності люмінесценції. Як позитивний контроль використали сигнал з комірок, у яких досліджувані реагенти інкубувалися без наявності субстрату (нульова кіназна активність). Як негативний контроль використали сигнал з комірок, куди не додавали

потенційний ліганд. Люмінесценцію детектували за допомогою BMG Polarstar Omega Reader упродовж 0.5 с.

Результати виражали у % інгібування кіназної активності.

РОЗДІЛ 3
РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ ТА ЇХ ОБГОВОРЕННЯ

3.1. Структурний аналіз доменів SH2

3.1.1. Відбір репрезентативних структур та їх розподіл на групи. Зазначимо, що відібрані репрезентативні структури у кожній з груп володіють 2D та 3D властивостями значної кількості подібних до них субструктур (рис. 3.1). Вони здебільшого є 100% ідентичними (з незначними геометричними відмінностями, середнє значення Rmsd між ними становить 1.6 Å). Так, наприклад, 2K7A (група 5) відображає інформацію про сім інших субструктур: 1LUI (residues 238-344), 1LUK.A (residues 238-344), 1LUM.A (residues 238-344), 1LUN.A (residues 238-344), 2K79.B, 2ETZ.A і 2EU0.A (усі вони є Tyrosine-protein kinase ITK/TSK та містяться у м’язах).

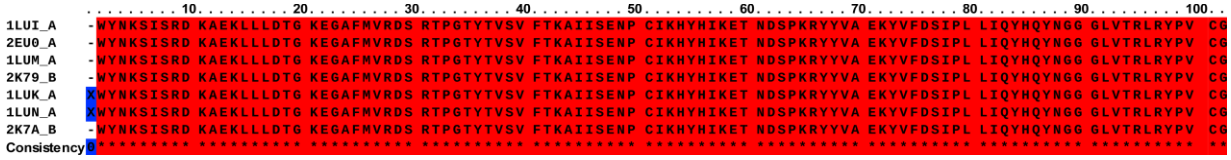


Рис. 3.1. Порівнювання послідовностей, які представлені репрезентативною структурою 2K7A.

Зі 1129 структур, після порівнювання амінокислотних послідовностей, отримали лише 56 (*більш детально див. Розділ 3.1.2.*). Враховуючи подібність амінокислотних послідовностей (по всій їх довжині), ці структури поділили на шість груп (рис. 3.2.).

Кластеризація дескрипторів сайтів зв’язування заснована на ключових консервативних позиціях кишені зв’язування, виділених після поділу на групи *(див. нижче)*. З них відібрали 8 найбільш репрезентативних сайтів зв’язування для проведення молекулярного докінгу (максимальна кількість,

яку можна було розрахувати, виходячи з апаратного та програмного забезпечення) методом k-середніх. Кожен такий сайт зв'язування визначали як структуру-центроїд, тобто такий, дескриптор якої є найближчим до центру відповідного кластера. Таким чином, кожний кластер представляє сукупність структур з геометрично подібними сайтами зв'язування, а структура-центроїд є певним усередненням, тобто «типовим представником» сукупності. Припускається, що центроїди представляють усю геометричну різноманітність і, обираючи їх для докінгу, відкидаємо надто подібні структури.

Кількість структур у кластері склала від 48 до 223. Повний перелік кластерів із зазначенням відповідних центроїдів та кількості структур наведений у табл. 3.1.

Таблиця 3.1.

Результат кластеризації доменів SH2

№ кластеру	Центроїд: структура (код PDB, ланцюг, модель)	Центроїд: організм, білок та домен	Кількість структур
1	1o49, ланцюг A	*Homo Sapiens* SRC SH2	223
2	2fci, ланцюг A, модель 6	*Bos Taurus* PLCG1 SH2-2	48
3	2ge9, ланцюг A, модель 15	*Homo Sapiens*Btk SH2	72
4	3in7, ланцюг A	*Homo Sapiens*GRB2 SH2	183
5	2jyq, ланцюг A, модель 9	*Homo Sapiens* GRB2 SH2	114
6	2k7a, ланцюг B, модель 5	*Mus Musculus*ITK SH2	218
7	2kk6, ланцюг A, модель 14	*Homo Sapiens* FER SH2	149
8	1uus, ланцюг A	*Dictyostelium Discoideum* DSTA SH2	129

Отримані 8 кластерів здебільшого подібні до 6 груп, виділених під час порівнювання амінокислотних пословностей. Як виняток, не отримали кластера №4 (табл. 3.1; 3IN7). Він потрапив до групи 1 (у середньому його подібність до цієї групи становить понад 75%, а до деяких членів понад 90%; табл. 3.2). Крім того, до груп 2 та 4 не увійшов жоден з центроїдів, одержаних при кластеризації. З аналізу їх сайтів зв'язування визначили, що групи 1 та 2 практично ідентичні, група 4 є подібною до групи 3 (у цьому випадку подібність є меншою, ніж у попередньому). Беручи до уваги те, що при відборі мішеней для докінгу насамперед цікавлять сайти зв'язування, за допомогою розподілу на групи визначали загальне різноманіття та характеристики подібності усієї структури доменів SH2. Детальний аналіз груп дав змогу оцінити не лише параметри кишені зв'язування доменів, але і виявити нові варіанти зв'язування (зв'язування в іншій гідрофобній кишені, точка 4, рис. 3.3), які врахували у подальших розрахунках.

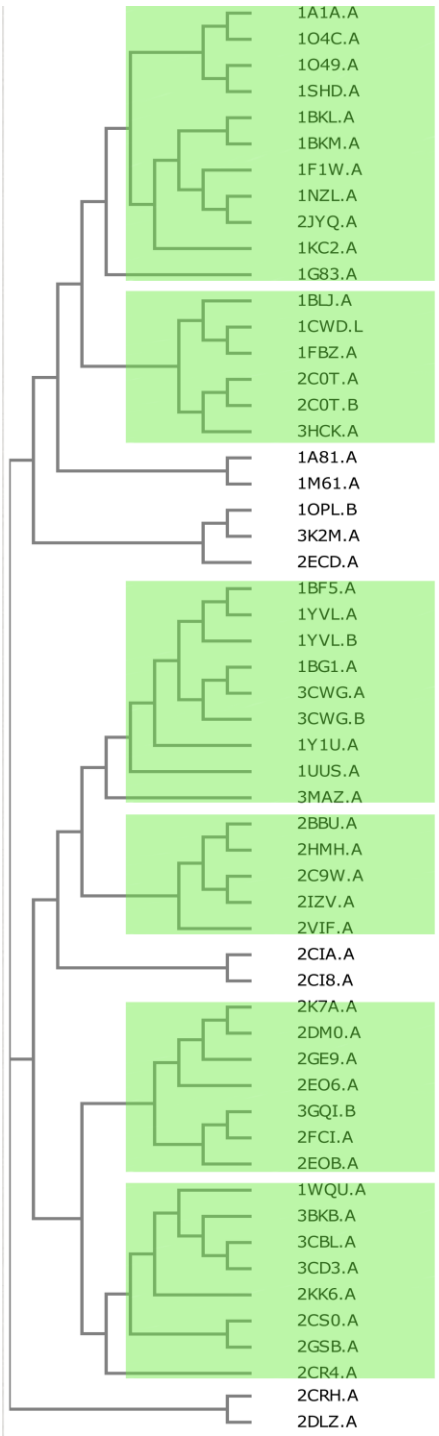


Рис. 3.2. Дендрограма розподілу доменів SH2 на групи. Зеленим кольором виділені репрезентативні структури з кожної групи.

3.1.2. Характеристика доменів SH2 в межах кожної групи. Отримані групи характеризуються значним зростанням параметрів подібності порівняно з

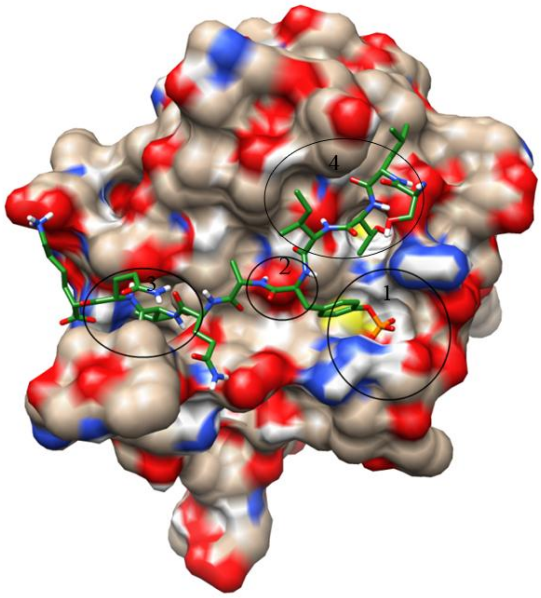


Рис. 3.3. Ключові точки взаємодії для доменів SH2 (PDB 1d4w). Показана молекулярна поверхня, де зліва направо: точки 1-3; кишеня зв'язування фосфотирозину; центральний атом O; основна гідрофобна кишеня 4; не основна гідрофобна кишеня.

усіма наявними доменами: наприклад, ідентичність усіх доменів становить 31%, тоді як у групі 1 – 91%, а у кишені зв'язування групи 1 – 92% (табл. 3.3). Також важливо, що типи протеїнів у кожній групі відрізняються. Це свідчить про структурні і, як наслідок, функціональні особливості, притаманні лише тому чи іншому типу протеїну. Так, група 1 містить – c-src tyrosine kinase (Homo sapiens - 1A1A, 1SHD, Rous sarcoma virus - 1BKL, 1KC2), Proto-oncogene tyrosine-protein kinase src (Homo - 1O4C, 1O49, Gallus gallus - 1F1W), Tyrosine-protein kinase transforming protein src (Rous sarcoma virus - 1NZL, 2JYQ), Proto-oncogene tyrosine-protein kinase fyn (Homo sapiens - 1G83), v-src tyrosine kinase transforming protein (Rous sarcoma virus - 1BKM); група 2 –p55 blk protein tyrosine kinase (Mus musculus - 1BLJ), p56lck tyrosine kinase (Homo sapiens -

1CWD), Proto-oncogene tyrosine-protein kinase lck (Homo sapiens - 1FBZ), Tyrosine-protein kinase hck (Homo sapiens - 2C0T), hck SH2

(Homo sapiens - 3HCK); група 3 –DNA(5'D(AP CP AP GP TP TP TP CP CP CP GP TP AP AP AP TP G P C)-3') (Homo sapiens - 1BF5), Signal transducer and activator of transcription 1-alpha/beta (Homo sapiens - 1YVL), DNA(5'D(TP GP CP AP TP TP TP CP CP CP GP TP AP AP AP TP CPT)-3') (Mus musculus - 1BG1), Signal transducer and activator of transcription 3 (Mus musculus - 3CWG), Signal transducer and activator of transcription 5A (Mus musculus - 1Y1U), Stat protein (Dictyostelium discoideum - 1UUS), Signal-transducing adaptor protein 1 (Homo sapiens – 3MAZ);

Таблиця 3.2.

Результати порівнювання відібраних структур : гомологічність (%;

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	---------

верхня половина таблиці) та значення Rmsd (Å) між цими структурами (нижня половина таблиці)

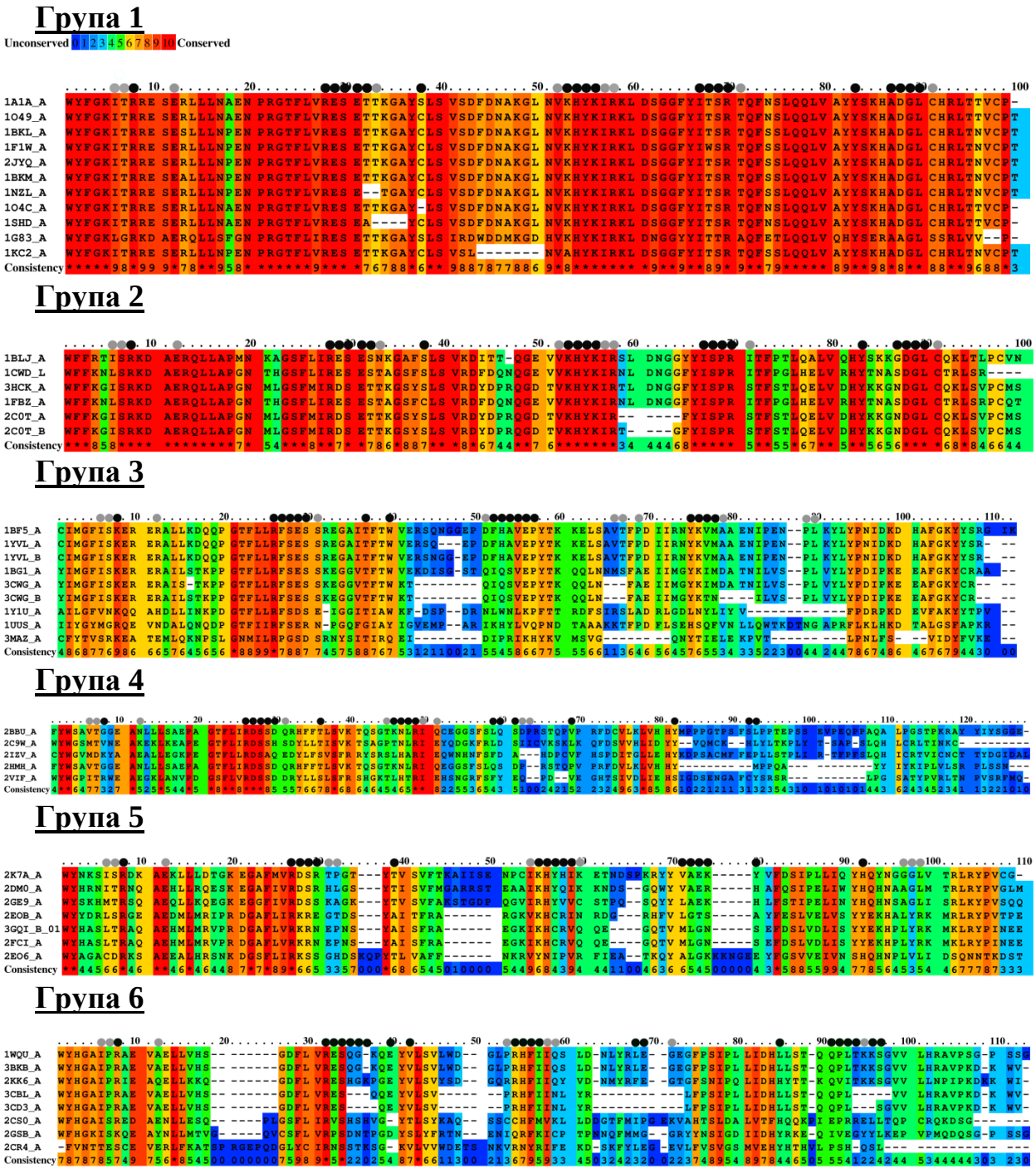


Рис. 3.4. Множинне вирівнювання послідовностей доменів SH2. Ключові амінокислотні залишки сайту зв'язування позначені чорними крапками (утворюють основну частину сайту зв'язування), сірими позначено амінокислотні залишки, внесок яких в утворення сайту з'язування є незначним.

група 4 – Suppressor of cytokine signaling (Mus musculus - 2BBU, 2HMH, Homo sapiens - 2IZV, 2VIF); група 5 – SH3 domain of Tyrosine-protein kinase ITK/TSK (Mus musculus - 2K7A), Tyrosine-protein kinase TXK (Homo sapiens - 2DM0), Tyrosine protein kinase Btk (Homo sapiens - 2GE9), B-cell linker protein (Mus musculus - 2EO6), Basic fibroblast growth factor receptor 1 (Homo sapiens - 3GQI), Doubly phosphorylated peptide derived from Syk kinase (Bos taurus - 2FCI), 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma 2 (Rattus norvegicus - 2EOB); група 6 – Proto-oncogene tyrosine-protein kinase FES/FPS (Homo sapiens - 1WQU, 3BKB, 3CBL, 3CD3), Proto-oncogene tyrosine-protein kinase FER (Homo sapiens - 2KK6), Hematopoietic SH2 domain containing (Homo sapiens - 2CS0), Ras GTPase-activating protein 1 (Homo sapiens - 2GSB).

Оцінювання доступної площі поверхні доменів SH2 показало, що сайт зв'язування займає в середньому ¼ від загальної поверхні домену. У всіх випадках він плоский і формується двома основними частинами кишені – pTyr-зв'язуючою та гідрофобною. Крім того, порівнювання структур виявило зростання параметрів подібності у сайті зв'язування в межах кожної групи та під час порівнювання з усіма доменами. Однак, якщо брати до уваги лише групи 1 і 2, тоді відмінність в параметрах подібності сайтів зв'язування та груп є незначними (зростання параметрів становить (1-2)%, табл. 3.3).

Групи 1 і 2 є висококонсервативними: зростання ідентичності структур (якщо порівнювати з усіма доменами) складає 61% і 54% та подібності 55.1% і 35.2%, відповідно. Вони містять найбільші консервативні ділянки, які формують поверхню домену. Важливо підкреслити, що цей факт не є очікуваним, оскільки для цих структур характерний високий рівень подібності. Водночас, це впливає на загальну площу поверхні цих частин домену. Так, наприклад, група 1 має найменшу площу сайту зв'язування, тоді як менш

консервативний сайт зв'язування групи 2 займає більшу площу поверхні (табл. 3.3, рис. 3.4). Це досягається завдяки варіативності амінокислотного складу у різних позиціях послідовностей.

Таблиця 3.3.

Розташування амінокислот сайту зв'язування доменів SH2. Параметри окремих груп доменів SH2 та їх сайту зв'язування показані як Рос. – сайт зв'язування і Pro.-весь протеїн. Точки на рис. 3.4 показані в (...), номер ділянки показаний в [...]. Значення консервативності наведені в одній колонці (K = 21/K = 7)

Група та шаблонний PDB	Амінокислотне оточення сайту зв'язування	Ідентичність сіквенсів, %		Середнє значення Rmsd, Å		Середня проща поверхні, Å^2		Значення консерватив-ності	
		Рос.	Pro.	Рос.	Pro.	Рос.	Pro.	Рос.	Pro.
Група 1 1A1A	[1]. Ile 156 (6), Thr 157 (7), Arg 158 (8). [2]. Glu 162 (12). [3]. Arg 178 (28), Glu 179 (29), Ser 180 (30), Glu 181 (31), Thr 182 (32), Thr 182 (33). [4]. Ser 188 (38). [5]. Val 202 (52), Lys 203 (53), His 204 (54), Tyr 205 (55), Lys 206 (56). Ile 207 (57), Arg 208 (58). [6]. Ile 217 (67), Thr 218 (68), Ser 219 (69), Arg 220 (70), Thr 221 (71). [7]. Tyr 233 (83). [8]. Ala 237 (87), Asp 238 (88), Gly 239 (89), Leu 240 (90), Cys 241 (91).	0.92	0.91	0.72	0.79	1268	5876	0.94/ 0.93	0.92/ 0.91
Група 2 1BLJ	[1]. Ile 19 (6), Ser 20 (7), Arg 21 (8). [2]. Ser 25 (12). [3]. Arg 41 (28), Glu 42 (29), Ser 43 (30), Glu 44 (31), Ser 45 (32), Asn 46 (33). [4]. Ser 51 (38). [5]. Val 64 (52), Lys 65 (53), His 66 (54), Tyr 67 (55), Lys 68 (56), Ile 69 (57), Arg 70 (58). [6]. Ile 79 (67), Ser 80 (68), Pro 81 (69), Arg 82 (70), Ile 83 (71). [7]. Tyr 95 (83). [8]. Gly 99 (87), Asp 100 (88), Gly 101 (89), Leu 102 (90), Cys 103 (91).	0.85	0.78	1.14	1.32	1491	5856	0.89/ 0.96	0.81/ 0.87
Група 3 1BF5	[1]. Ile 582 (6), Ser 583 (7), Lys 584 (8). [2]. Arg 588 (12). [3]. Arg 602 (26), Phe 603 (27), Ser 604 (28), Glu 605 (29), Ser 606 (30), Ser	0.5	0.47	1.53	1.65	1552	6945	0.59/ 0.67	0.56/ 0.62

	607 (31). [4]. Trp 616 (40). [5]. Asp 627 (51), Phe 628 (52), His 629 (53), Ala 630 (54), Val 631 (55), Glu 632 (56), Pro 633 (57). [6]. Val 642 (66), Thr 643 (67), Pro 645 (69). [7]. Tyr 651 (75), Lys 652 (76), Val 653 (77), Met 654 (78), Ala 655 (79). [8]. Pro 633 (89), Leu 634 (90). [9]. Thr 613 (37).								
Група 4 2C9W	[1]. Met 53 (6) Thr 54 (7), Val 55 (8). [2]. Lys 59 (12). [3]. Arg 73 (26), Asp 74 (27), Ser 75 (28), Ser 76 (29), His 77 (30), Ser 78 (31). [4]. Thr 83 (36). [5]. Pro 92 (45), Thr 93 (46), Asn 94 (47), Leu 95 (48), Arg 96 (49), Ile 97 (50), Glu 98 (51). [6]. Leu 106 (59), Asp 107 (60), Ser 108 (61), Ile 109 (62), Ile 110 (63). [7]. Tyr 129 (82). [8]. His 149 (92), Leu 150 (93). [9]. Leu 116 (69).	0.45	0.32	1.82	1.99	1355	7161	0.49/ 0.60	0.37/ 0.51
Група 5 2DMO	[1]. Ile 23 (6), Thr 24 (7), Arg 25 (8). [2]. Glu 29 (12). [3]. Arg 44 (27), Asp 45 (28), Ser 46 (29), Arg 47 (30), His 48 (31), Leu 49 (32). [4]. Thr 53 (39). [5]. Ile 68 (54), Lys 69 (55), His 70 (56), Tyr 71 (57), Gln 72 (58), Lys 73 (59), Ile 74 (60). [6]. Val 83 (71), Ala 84 (72), Glu 85 (73), Arg 86 (74), His 87 (80). [7]. His 79 (92). [8]. Ala 104 (97), Glu 105 (98), Lleu 106 (99).	0.39	0.39	1.81	2.00	1645	6499	0.53/ 0.70	0.50/ 0.68
Група 6 1WQU	[1]. Ile 23 (6). Pro 24 (7), Arg 25 (8). [2]. Ala 29 (12). [3]. ARG 41 (32), GLU 42 (33), SER 43 (34), GLN 44 (35), GLY 45 (36), -//- (37). [4]. Val 50 (42). [5]. Pro 59 (53), Arg 60 (54), His 61 (55), Phe 62 (56), Ile 63 (57), Ile 64 (58), Gln 65 (59). [6]. Leu 73 (68), Glu 74 (69), Gly 75 (72). [7]. Leu 88 (85). [8]. Gln 93 (91), Pro 94 (92), Leu 95 (93), Thr 96 (94), Lys 98 (96), Ser 99 (97). [9]. Glu 48 (40).	0.45	0.45	1.27	1.45	1639	5429	0.53/ 0.59	0.51/ 0.61

Групи 3-6 є набагато менш консервативними порівняно з першими двома: зростання ідентичності (при порівнянні з усіма доменами) становить 19%, 14%, 8% і 14%, а подібності 14,5%, 12,2%, 15,5 і 11,63%, відповідно, що, ймовірно, і призводить до значного зростання середньої площі поверхні доменів. Так, наприклад, позицію 55 (ділянка 5) у групі 3 займає переважно Val (S=22.06 $Å^2$), але у випадку структури 1Y1U він замінюється на Ile (48.6 $Å^2$) (має місце зростання величини ASA на 26 $Å^2$).

Зростання ідентичності та подібності у групах 3-6 не настільки значне, як у групах 1 і 2. Але зростання цих параметрів є достатнім для ділянок, які формують сайт зв'язування та близькі до нього (рис. 3.4; табл. 3.3). Зростання подібності у групі 3 складає 21%. У всіх випадках консервативність та ідентичність в межах кожної групи перевищували 40%.

Виявлено декілька консервативних ділянок, що знаходяться не в межах сайту зв'язування (не взаємодіють з короткими фосфопептидами, але беруть участь у PPI), але просторово наближені до нього (позиції 1-5 та 10-11 на усіх множинно вирівнюваних амінокислотних послідовностях) (рис. 3.4). Отже, наявність таких консервативних ділянок, які не є частинами сайту зв'язування, ставить під сумнів можливість зв'язування доменів SH2 лише з лінійними фосфопептидними послідовностями.

3.1.3. 3D оточення сайту зв'язування. Вісім основних ділянок та поодиноких амінокислот представлені у сайті зв'язування доменів SH2. Однак групи 3, 4 і 6 містять по одній додатковій позиції зв'язування (номер 9, табл. 3.3). Загалом сайт зв'язування доменів SH2 містить 20-24 амінокислоти, що було визначено за допомогою оцінки доступної площі поверхні доменів.

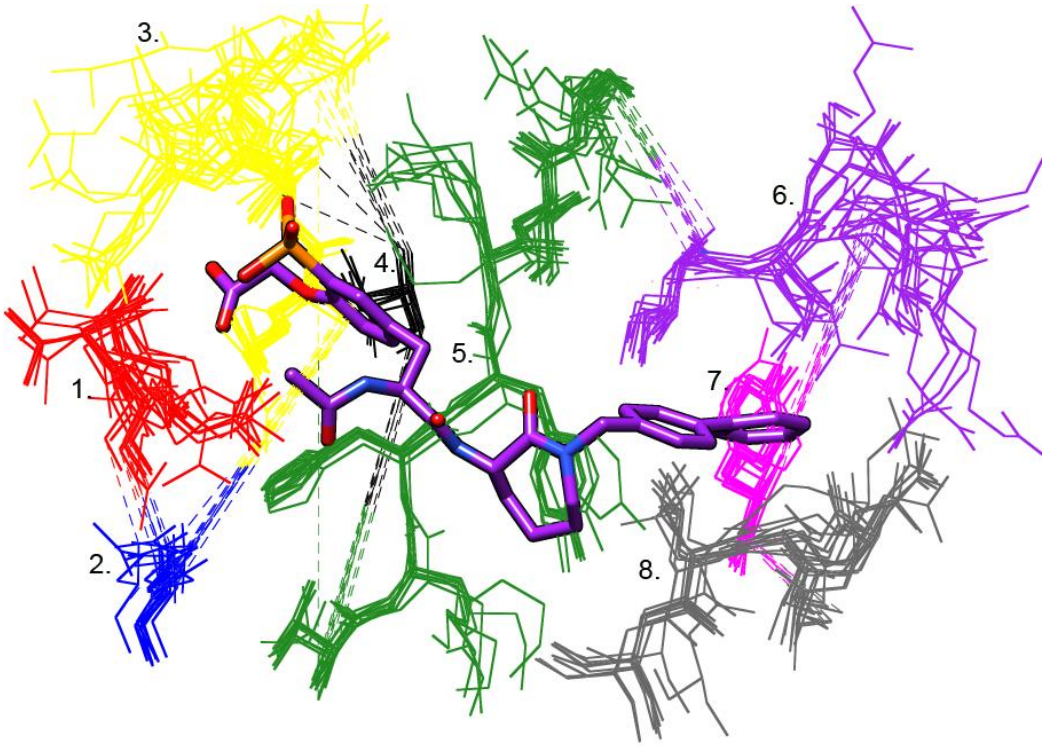


Рис. 3.5. Сайт зв’язування групи 1 з нативним лігандом і розподіл ділянок сайту зв’язування: ділянка 1 (ITR), ділянка 2 (E), ділянка 3 (RESETT), ділянка 4 (C), ділянка 5 (VKHYKIR), ділянка 6 (ITSRT), ділянка 7 (Y) і ділянка 8 (ADGLC).

Сайт зв’язування доменів SH2 формує значна кількість вторинних структур (βB-sheet, βG-sheet, βG-sheet та αB-helix) (рис. 3.4, 3.5; табл. 3.3). Так, pTyr-зв’язуючу частину формують ITR (ділянка 1), FLVRESETT (ділянка 3) та C (ділянка 4). Лише друга частина ділянки 3 (RESETT) бере участь у формуванні можливих взаємодій з лігандом. Ділянки 2 (E) і 5 (VKHYKIR) (βD-sheet і αA-helix) формують центральну частину кишені зв’язування. Третя частина включає ITSRT (ділянка 6), Y (ділянки 7), ADGLC (ділянка 8) та формує гідрофобну частину кишені зв’язування (βG-sheet, B-helix та αA-helix), позиції 67-70 і 87-91. Крім того, існує чотири висококонсервативних позиції (>70%), якщо брати до уваги всі домени SH2: позиція 3 ділянки 1, позиції 1 і 3 ділянки 3 та позиція 3 ділянки 5. Усі інші частини сайту зв’язування є менш консервативними (<70%).

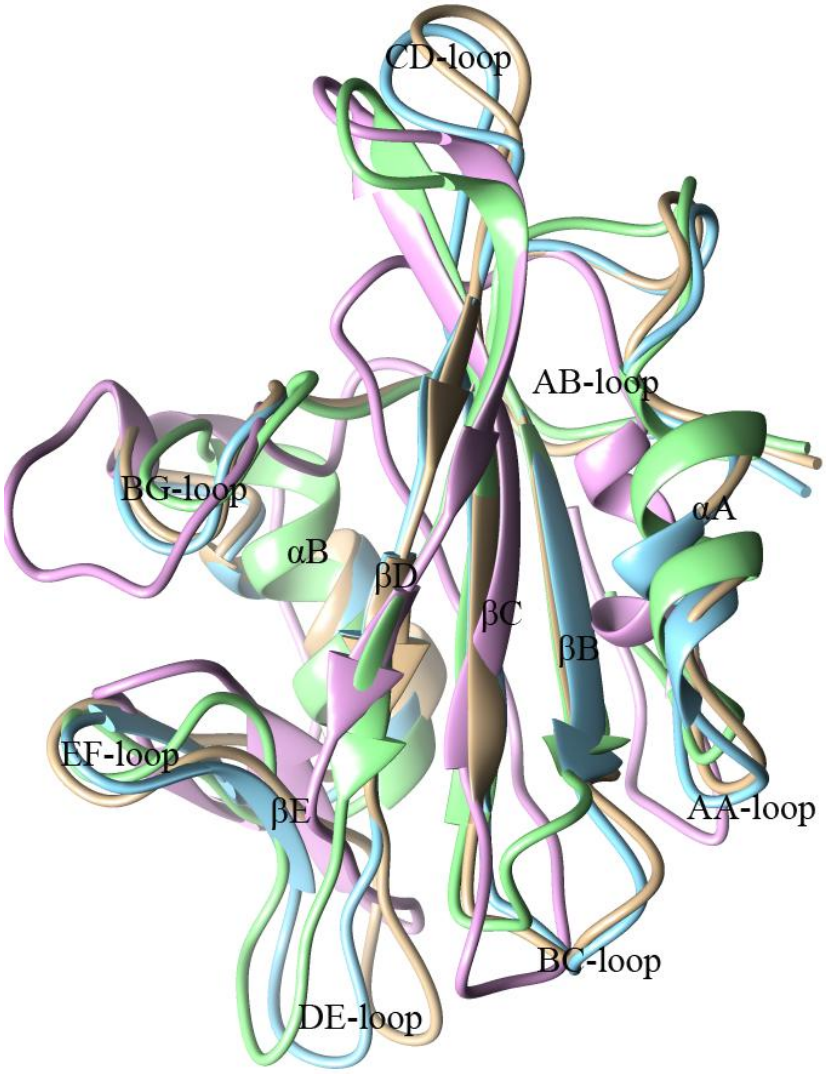


Рис. 3.6. Приклад накладання отриманих конформацій – PDB ID 1O49: сірий – початкова форма домену, зелений – перша конформація (Rmsd - 3.2 Å); синій – друга конформація (2.6 Å); рожевий – третя конформація (5.17 Å).

Також виявили декілька позицій у сайті зв’язування, які не є частиною будь-якого із сайтів зв’язування з вище описаних, наприклад, T (позиція 37 - група 3), L (позиція 69 - група 4) та Q (позиція 40 - група 6). Така різноманітність є важливою для розпізнавання лігандів.

3.1.4. Конформаційний пошук для доменів SH2. Оцінимо рухливость 8 доменів SH2 з попередньо відібраними найбільш репрезентативними сайтами зв'язування. Приклад накладання різних конформацій доменів SH2 наведений на рис. 3.6, загальні характеристики – в табл. 3.4, а рухливість кожної структури показана на рис. 3.7. Загалом для наведених зразків згенерували 39 (від 4 до 6 на одну структуру) можлививх конформацій. Значення Rmsd для усіх конформацій змінювалось в межах (1.0 – 3.5) Å за винятком однієї конформації, де воно становить 5.17 Å (конформація № 3 PDB ID 1O49). Також врахували те, що "DE-петля, βE – складка, EF-петля та FB-петля" знаходяться в одній частині домену на незначній відстані, тому їх описували як один структурний комплекс (рис. 3.6).

Так, має місце такий середній внесок у рухливість домену SH2 кожної окремої вторинної структури αA – спіралі (середне значення Rmsd для отриманих конформацій знаходиться в діапазоні (1.5–2.7) Å), BC – петлі (1.3 – 3.1) Å, CD – петлі (1.3 – 6.7) Å, αB – петлі (1.0 – 2.5) Å, BG – петлі (1.8 – 2.5) Å, DE-петля, βE – складка, EF-петля і FB-петля (1.6–2.8) Å.

Однак підкреслимо, що рухливість різних частин вище зазначених структур не є однаковою. Так ***αA – спіраль*** (1O49 – RRESERLL, 3IN7 – RAKAEEML, 1UUS – RQEVNDA, 2FCI – RAQAEHMLMR, 2GE9 – RSQAEQLLKQE, 2JYQ – RRESERLL, 2K7A – RDKAEKLLLDT, 2KK6 – RIEAQE) містить дві високорухливі амінокислоти (номер один (0.4 – 7.5) Å та номер п'ять (0.7 – 5.1) Å). На противагу цьому в конформаціях отриманих щодо 2KK6 найбільш рухливою є амінокислота номер 2 (2.5 –5.7) Å. Усі інші амінокислоти спіралі є менш гнучкими (1.0 – 2.5) Å. Подібна ситуація спостерігається для ***BG – петлі*** (1O49 – HADGLCHRLTTVCPT, 1UUS – HSQFV, 2FCI – HPLYRKMK, 2GE9 – HNSAGLISRLKYPVSQQNKNAPST, 2K7A – YNGGGLVTRLRYPVCG, 2JYQ – SKHADGLCHRLTNVC, 2KK6 – TKQVITKKSGVVLLNPIP, 3IN7 – STSVSRNQQIFLRDIEQ). Середня конформаційна рухливість у випадку 1O49, 2FCI, 2GE9, 2K7A, 2JYQ, 2JYQ і

2KK6 знаходиться в діапазоні (2.4 – 2.5) Å, для 1UUS та 3IN7 – (1.3 – 1.8) Å. Одна частина BG – петлі (наприклад, 2K7A – RLRYPVCG) є жорсткою (1.0 – 1.8) Å, інша є більш гнучкою (наприклад, 2K7A – YNGGGLVT) – (1.0 – 4.0) Å. Для ***BC – петлі*** (1O49 – SETTKGA, 1UUS – SERNPGQ, 2FCI – KRNEPNSY, 2K7A – DSRTPGTY, 2GE9 – SSKAGK, 2K7A – DSRTPGTY, 2JYQ – SETTKG, 2KK6 – SHGKPGE, 3IN7 – SESAPGD) аналогічна ситуація: центральна частина петлі (наприклад, 2FCI – EPN) є більш рухливою, ніж бічні частини (наприклад, 2FCI – KRN і SY); Rmsd діапазон становить (2.0 – 4.1) Å і (1.5 – 1.9) Å, відповідно. Таку мобільність можна пояснити тим, що ця петля контактує із βB– та βC – складками, загальна рухливість яких є мінімально. Рухливість ***CD – петлі*** (1O49 – NAK, 1UUS – GEVMPAR, 2FCI – KRNEPNSY, 2GE9 – KSTGDPQG, 2K7A – TKAIISENPC, 2JYQ – NAK, 2KK6 – DG, 3IN7 – GN), ***αB-спіралі*** (1O49 – LQQLVAYYSK, 1UUS – FPDFLSE, 2FCI – SLVDLISYYEK, 2GE9 – IPELINYHQ, 2K7A – IPLLIQYHQ, 2JYQ – LQQLVAYY, 2KK6 – IPQLIDHHYT, 3IN7 – LNELVDYHR) та ***DE-петлі, βE – складки, EF-петлі і FB-петлі*** (1O49 –RKLDSGGFYITSRTQFNS, 1UUS – VQPNDTAAKKT, 2FCI – QQEGQTVMLGNSEFD, 2GE9 – VCSTPQSQYYLAEKHLFST, 2K7A –, 2JYQ – KLDSGGFYITSRTQFSS, 2KK6 – IKETNDSPKRYYVAEKYVFDS, 3IN7 – RDGAGKYFLWVVKFNS) є порівняно однаковою. Однак спостерігається значна відмінність у рухливості між різними структурами. Так, для 1UUS, 2FCI, 2K7A, 2KK6 і 3IN7 середня гнучкість для CD – петлі становить 1.5 Å, 2.8 Å, 2.7 Å, 1.3 Å і 2.8 Å, відповідно. На противагу цьому CD – петля для структур 1O49, 2GE9 і 2JYQ є набагато більш гнучкою – 6.1 Å, 3.1 Å і 4.6 Å. Рухливість αB – спіралі структур 1UUS та 3IN7 становить 1.0 Å та 1.5 Å. Усі інші структури є більш гнучкими (2.1 – 2.5) Å. З іншого боку, рухливість AA – петлі (0.8 – 2.0) Å, AB – петлі (1.4 – 2.0) Å, βB – складки (0.9 – 1.9) Å, βC – складки (1.2 – 1.9) Å та βD – складки (0.9 – 2.3) Å вторинних структур є значно меншою. Рухливість ***AA – петлі*** (1O49 – WYFGKIT, 3IN7 – WFFGKIP, 2FCI – WHASLT, 2GE9 – WYSKHMT, 2JYQ – WYFGKIT, 2K7A – WYNKSIS, 2KK6 – WYNKSIS,

1UUS – IIYGYMG) становить від 0.3 Å до 8.0 Å. Практично у всіх випадках перші чотири амінокислоти АА – петлі є найбільш рухливими (наприклад, для 1O49 (WYFG) – (2.0 – 8.0) Å, всі інші є менш рухливими (0.3 – 2.0) Å.

Таблиця 3.4.

Результати конформаційного пошуку репрезентативних структур доменів SH2

PDB	Rmsd обмеження, Å	Загальна кількість отриманих конформацій	Rmsd, Å (порівняно з нативним доменом)
1O49	0.25	3	3.2
			2.6
			5.17
	0.5	1	1.34
	0.75	1	1.44
1UUS	0.25	1	1.09
	0.5	2	1.15
			3.37
	0.75	1	2.8
3IN7	0.25	2	0.94
			3.74
	0.5	1	0.95
	0.75	1	2.39
2FCI	0.25	2	2.78
			2.12
	0.5	2	3.14
			2.07
	0.75	1	2.61
2GE9	0.25	1	2.67
	0.5	2	2.49
			2.26
	0.75	2	2.17
			3.10
2JYQ	0.25	1	2.3
	0.5	2	2.54
			2.33
	0.75	3	2.79

			2.6
			2.62
2K7A	0.25	1	2.75
	0.5	2	2.11
			3.21
	0.75	3	2.87
			2.95
			2.03
2KK6	0.25	1	2.31
	0.5	1	2.21
	0.75	2	2.32
			3.14

Рухливість ***AB – петлі*** (1O49 – LNAENPRGT, 1UUS – LQNQDPGT, 2FCI – VPRD, 2GE9 – GKEG, 2K7A – GKEGA, 2JYQ – LNPENPRGT, 2KK6 – GKEGA, 3IN7 – SKQRHDGA), ***βB – складки*** (1O49 – FLVRE, 1UUS – FIIRF, 2FCI –GAFLVR, 2GE9 – GFIVRD, 2K7A – FMVR, 2JYQ – FLVRE, 2KK6 – FLVRE, 3IN7 – FLIRE), ***βC – складки*** **(**1O49 – YCLSVSDFD, 1UUS – FGIAYI, 2FCI – ATFSR, 2GE9 – YTVSVFA, 2K7A – TVSVF, 2JYQ – AYCLSVSDFD, 2KK6 – YVLSVYS, 3IN7 – FSLSVKF) та ***βD – складки*** (1O49 – GLNVKHYKI, 1UUS – IKHYL, 2FCI – AISFRAEGKIKHCRV, 2GE9 – VIRHYV. 2K7A – IKHYH, 2JYQ – GLNVKHYKIR, 2KK6 – IKHYH, 3IN7 – DVQHFKV) є практично однаковою на всіх частинах структури (за декількома винятками). Так, гнучкість AB – петлі знаходиться в діапазоні (1.6 – 2.0) Å. А рухливість LNA (1O49) і LNPENP (2JYQ) – (2.0 – 3.0) Å і (1.5 – 3.0) Å, відповідно. З іншого боку ділянки ENPRGT(1O49) і RGT (2JYQ) менш гнучкі (0.5 – 1.8) Å і (1.0 – 1.7) Å. Ділянки GL (1O49) і DVQ (3IN7) звязані з CD – петлею, що може слугувати причиною значної рухливості цих структур (2.1 – 3.3) Å.

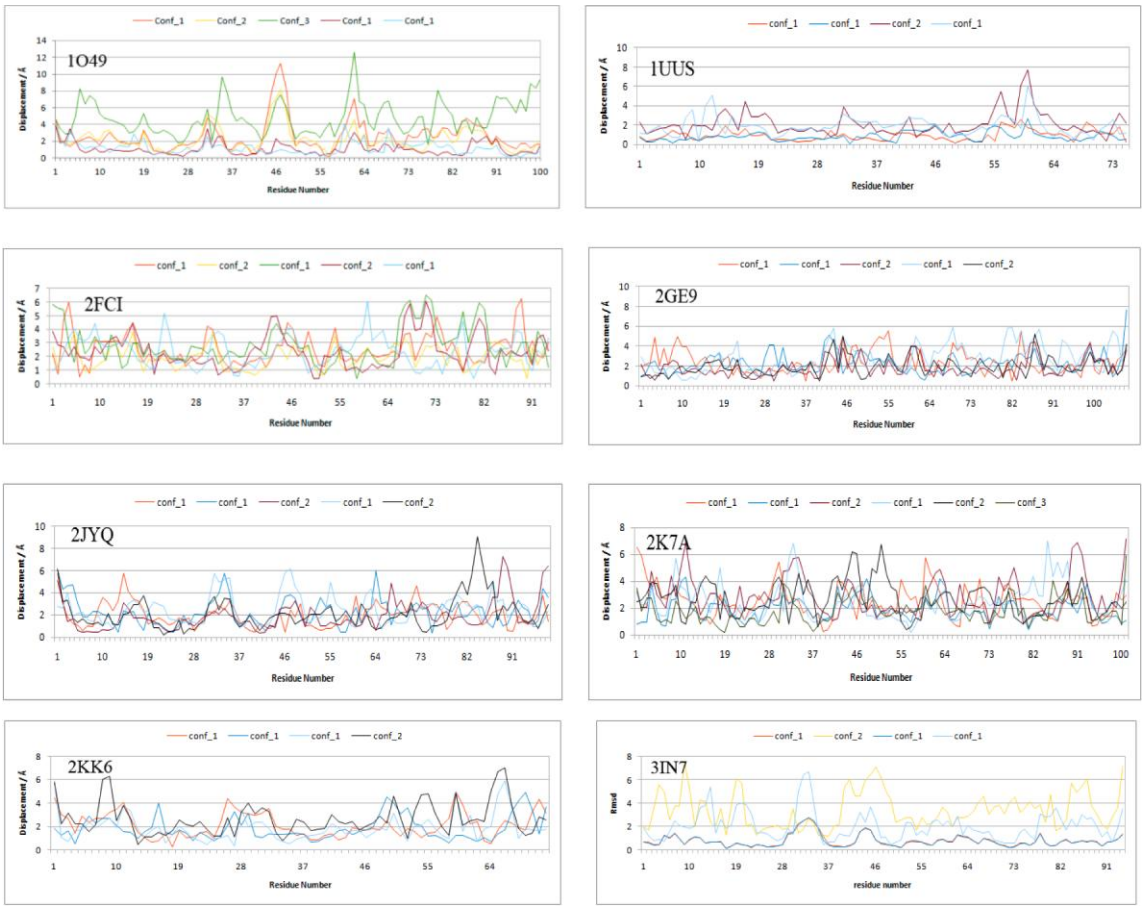


Рис. 3.7. Загальна рухливість усіх конформацій, наведених в табл. 3.4: (1) криві за Rmsd обмеження 0.25 Å позначені: червоний (конформація №1); жовтий (конформація №2) і зелений (конформація №3); (2) Rmsd обмеження 0.5 Å: синій (конформація №1) і коричневий (конформація №2); (3) Rmsd обмеження 0.75 Å: блакитний (конформація № 1); чорний (конформація №2) і темно-зелений (конформація №3).

Також оцінили рухливість усіх ділянок, що формують сайт зв'язування доменів SH2. Виявили, що загальна рухливість сайту зв'язування є меншою ніж усього домену. Загалом, це очікувано оскільки сайт зв'язування є більш консервативним, ніж усі інші структури домену (*див. вище*). Отже, його середня рухливість для різних структур становить: 1O49 - 2.26 Å, 1UUS - 1.27 Å, 2FCI - 2.27 Å, 2GE9 - 2.29 Å, 2JYQ - 2.12 Å, 2K7A - 2.38 Å, 2KK6 - 2.28 Å і

3IN7 - 1.742 Å. Звичайно, кожна частина сайту зв'язування дає свій внесок у гнучкість домену: так, pTyr – зв'язуюча частина (утворюється AA-петлею, αA-спіраллю, βB-складкою, BC-петлею і βC-складкою) і гідрофобна кишеня (βB-складка, BC-петля, βC-складка і "DE-петля, βE – складка, EF-петля і FB-петля") є порівняно високорухливими. Отже, pTyr – зв'язуюча частина характеризується рухливістю в межах (1.1 – 2.5) Å. Водночас, рухливість pTyr –зв'язуючої частини 1UUS є незначною (1.1 Å) а рухливість pTyr – зв'язуючої частини 2GE9 (1.7 Å), 2JYQ (1.8 Å) і 3IN7 (1.8 Å) посередня; для усіх інших рухливість pTyr – зв'язуючого сайту перевищує 2 Å. Відмітимо, що практично всі вторинні структури дають значний внесок у рухливість pTyr – зв'язуючого сайту: AA-петля (наприклад, PDB 1O49 – 2.8 Å), αA-спіраль (2.7 Å), βC-складка (2.7 Å), BC-петля (2.8 Å), (за винятком βB-складки (1.6 Å)). Згідно з отриманими результатами, гідрофобна частина кишені є більш гнучкою, ніж інші частини. Середня рухливість становить: 1O49 – 2.46 Å, 1UUS – 1.63 Å, 2FCI – 2.3 Å, 2GE9 – 2.6 Å, 2JYQ – 2.3 Å, 2JYQ – 2.4 Å, 2K7A – 2.5 Å, 2KK6 – 2.5 Å і 3IN7 – 1.7 Å. Як і у попередньому випадку, для різних центроїдів внески у рухливість залежно від вторинної структури є різними, наприклад; 1O49, 2FCI і 2GE9, BG-петля є високорухливими (2.7 Å, 2.3 Å і 3 Å, відповідно), а для 1UUS і 3IN7 рухливість становить 1.3 Å і 1.7 Å. Найбільш стабільною виявилась центральна частина кишені зв'язування. Тільки у двох випадках (2FCI – 2.3 Å і 2GE9 – 2.5 Å) середня рухливість перевищує 2 Å. У всіх інших випадках рухливість менша (1O49 – 1.3 Å, 1UUS – 1 Å, 2JYQ – 1.8 Å, 2K7A – 2 Å, 2KK6 – 1.8 Å і 3IN7 – 1.6 Å). Практично у всіх випадках βD-складка і αA-спіраль формують цю частину. Але і тут спостерігаються винятки: 1UUS центральну частину також формують BC-складка (TYR 625) і "DE-петля, βE – складка, EF-петля і FB-петля" частина (VAL 639 і GLN 640); 2FCI - CD-петля та 2K7A - BC-складка і "DE-петля, βE – складка, EF-петля і FB-петля".

3.2. Моделювання комплексу “домен SH2-ліганд”

3.2.1. Молекулярний докінг доменів SH2. Спочатку провели редокінг на прикладі зв’язаного домену 1O49. Нативний ліганд видалили з протеїну і повторно докували у сайт зв’язування. Результат вважався задовільним, якщо значення Rmsd не перевищувало 2 Å (табл. 3.5 і 3.6). Використовуючи пакет MOE, одержали менший діапазон значень Rmsd – (1-1.7) Å (табл. 3.5). Однак, з метою одержання найкращого варіанту зв’язування застосували програму flo+ (табл. 3.6). Крім того, порівнювали енергетичні параметри отриманих комплексів (табл. 3.6), зокрема 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 ліганда). 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; нарешті, до третьої групи можна віднести положення 5, яке порівняно з усіма іншими має найгірші характеристики. Як видно з табл. 3.6, у будь-якому

Таблиця 3.5.

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

Таблиця 3.6.

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

положенні є декілька практично ідентичних параметрів щодо нативної форми. Винятком можна назвати лише положення 1, але у цьому випадку має місце найкраще значення Rmsd, що є одним з основних параметрів оцінки якості редокінгу.

У всіх випадках докінг в рамках програми MOE проводили за трьома ключовими точками: pTyr-зв'язуюча частина, взаємодія з центральним His або Tyr та утворення зв'язків у гідрофобній кишені (рис. 3.3, точки 1-3, 1, 2, 4).

Таблиця 3.7.

Результати докінгу в програмному пакеті MOE

Структура	1o49	2fci	2ge9	2jyq	2k7a	2kk6	1uus
Кількість речовин	86008	717214	372977	751379	711509	407885	266225

Загалом для кожного репрезентативного домену SH2 до 86 тис. речовин (табл. 3.7). Їх використали на наступному етапі розрахунків у рамках програмного пакету flo+.

Перед проведенням докінгу молекули води прибрали зі структур доменів. Усі Arg та Lys амінокилотні залишки запротонували (збільшили потенційну кількість водневих зв'язків). Для усіх хімічних структур згенерували стереоізомери (їх максимальна кількість для однієї речовини складала шість). Незважаючи на деякі відмінності, у всіх випадках ключовими точками зв'язування були pTyr-зв'язуюча та гідрофобна кишені. Докування програмним пакетом flo+ відбувалось за таким алгоритмом.

Спочатку домени SH2 дослідили на прикладі простих взаємодій з метою відбору принципово важливих амінокислот для зв'язування лігандів. Для цього було розроблено п'ять моделей, які відрізнялися механізмом зв'язування (моделі 1-4 показані на рис. 3.8, 3.9). Проведені розрахунки

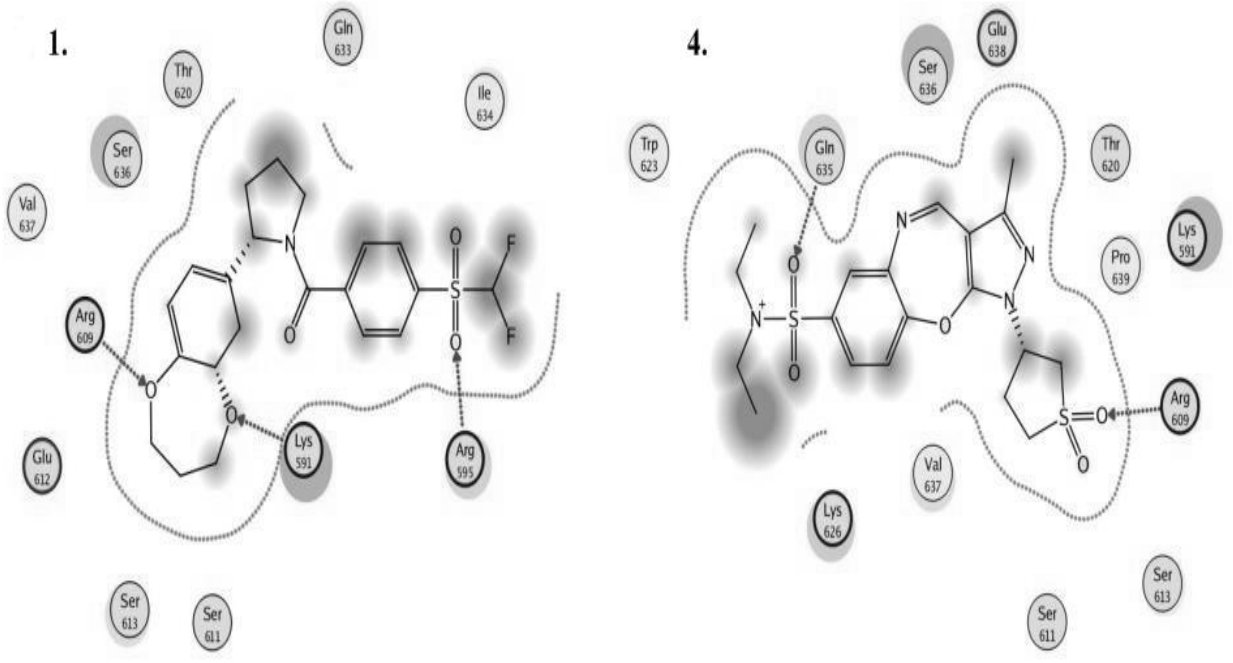


Рис. 3.8. 1 і 4 - моделі STAT3 білку (схематичне зображення кишені зв'язування з лігандом).

виявили, що усі ліганди, які зв'язуються з доменом SH2, є короткими - їх довжина не перевищувала довжину 2-3-х пептидних зв'язків, тоді як довжина значної частини нативних лігандів становить 4-5 пептидних зв'язків (не кажучи вже про велику кількість масивних пептидів, зв'язування з якими нагадує взаємодію із субстратом). У цьому полягає принципова відмінність взаємодії субстрату і низькомолекулярного ліганду, тобто ліганд повинен на менший об'єм субстрату надати більшу кількість зв'язків. Тому попередньо був проведений детальний аналіз поверхні та ключових точок зв'язування домену і на його основі сформульовано оригінальні правила зв'язування у кожній з п'яти (чотири основних та комбінації їх варіантів, які не зустрічались у нашому випадку (не було таких речовин)) запропонованих моделей. Важливо зазначити, що на рис. 3.8 і 3.9 наведені результати лише для перших чотирьох моделей, оскільки п'ята модель відрізнялася досить широкими енергетичними межами щодо утворення можливих структурних конформацій, завдяки чому в ній відтворювалися усі взаємодії, характерні для попередніх моделей.

Так, спочатку в тестовому форматі провели моделювання на доменах SH2 протеїну STAT3. У цьому випадку найпоширенішими виявились чотири моделі зв'язування (рис 3.8, 3.9). Як результат, отримали понад 1 тисячу потенційно активних речовин.

Незважаючи на відмінності у механізмах зв'язування з амінокислотним оточенням, першій та четвертій моделям притаманні спільні риси. В обох випадках ліганд щільно заповнює сайт зв'язування фосфотирозину, утворюючи міцні водневі зв'язки: у першому випадку з Lys 591 та Arg 609, а в другому - з Arg 609. Ліганди в обох моделях стерично взаємодіють з багатьма амінокислотами за принципом розпізнавання ароматичного залишку, до яких входять Glu612, Val 637, Pro 639, Ser 613 та 611.

Перша модель передбачає утворення другого кластеру водневих зв'язків, утворених з Arg 595, який просторово наближений до фосфотирозинової кишені. Ліганди четвертої моделі заповнюють основний простір кишені при

зв’язуванні з 1 або 2 фосфотирозиновим залишком Gln 635 водневим зв’язком. Також, зважаючи на гнучкість лізинового залишку (Lys 626), можливе утворення водневого зв’язку між ним і лігандом. Варто зазначити, що для першої моделі характерні просторово і стерично жорсткі молекули з невеликим об’ємом. На противагу їй, у четвертій моделі присутні сполуки з ароматичним каркасом, які можуть бути достатньо гнучкими.

Моделі 2 і 3 мають однаковий механізм зв’язування з фосфотирозиновим сайтом і демонструють гнучкість у підборі низькомолекулярних структур (рис. 3.9). Саме ця хімічна послідовність є ключовою щодо здатності речовин інгібувати мішень. Проте, існують інші функціональні групи, які за складом та просторовою структурою відрізняються від неї. Так, у моделі 2 - це циклосульфолановий фрагмент, а також аліфатичний і конформаційно-гнучкий цикл. У моделі 3 - арилсульфамідна група, яка стерично імітує фосфотирозиновий фрагмент. У рамках моделі 2 сполуки заповнюють кишеню, утворену Phe 716, Trp 623, Thr 714 та Lys 626. Ці амінокислотні залишки відносяться до різних структурних фрагментів, але просторово наближені одна до одної. Для стабільної взаємодії необхідна конформаційна жорсткість, яка досягається завдяки спряженю складної системи гетероциклів і домінуваню в кишені стекінг – та катіон-π-взаємодій. У моделі 3 відбуваються взаємодії, подібні до зв’язування субстрату. Це єдина схема інгібування, за якої ліганд досягає Pro 715, тобто заповнює об’єм і реалізує контакти з оточенням подібно субстрату. Також важливим є утворення водневого зв’язку із Ser 636, який імітує субстрат.

У подальших розрахунках використали базу даних, яка включала 50 тис. речовин. Враховуючи, що у цьому випадку початкова кількість речовин є незначною, ми використали всі центроїди, які були відібрані за допомогою

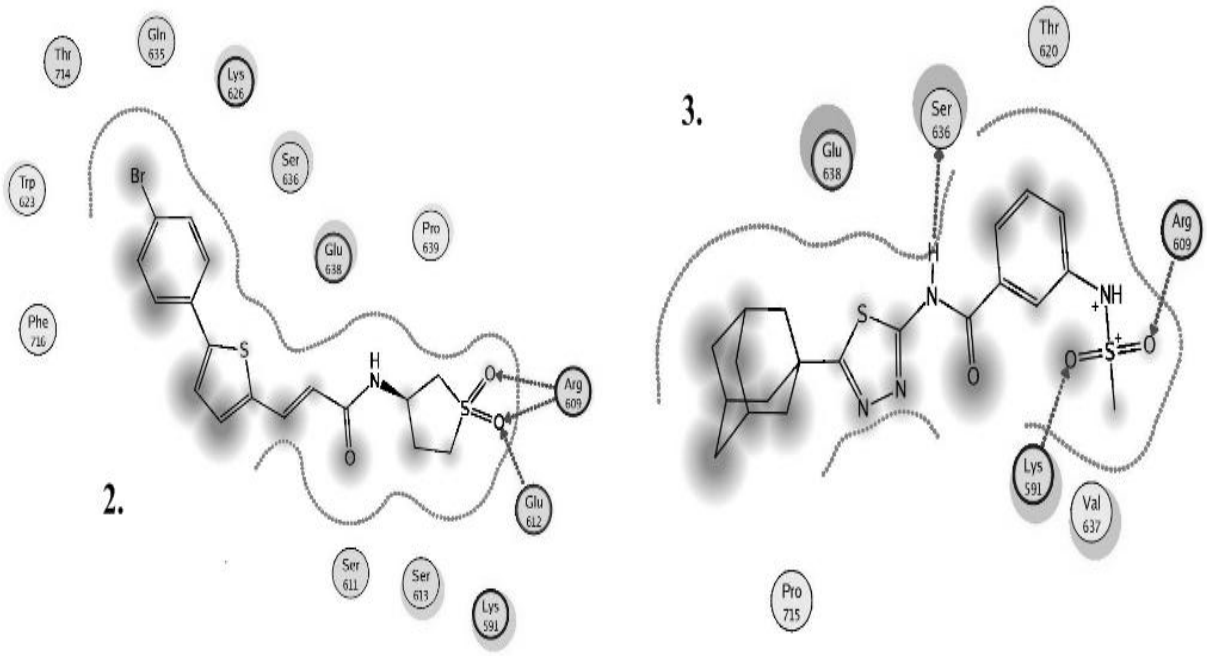


Рис. 3.9. 2 і 3 - моделі STAT3 білку (схематичне зображення кишені зв'язування з лігандом).

методу *k-середніх*. Після докінгу отримали близько 10 млн. комплексів, які були відфільтровані за значеннями Rmsd та геометричними фільтрами. Хімічні групи, які імітують фосфотирозин, та чотири основні моделі зв'язування (A, B, C та D) представлені на рис. 3.10, 3.11.

A. Модель характеризується наявністю додаткових водневих зв'язків за участю полярних радикалів фосфотирозинімітуючого фрагменту: утворюються водневі зв'язки з Arg 26 та Arg 46. Крім того, має місце міцне зв'язування з Arg 26 завдяки катіон-π взаємодії з піримідиновим гетероциклом. Також ліганд частково заповнює іншу частину кишені та стерично взаємодіє з Gln 30, Val 52 та His 63. Поліцикл є просторово наближеним до Arg 61, існує можливість утворення катіон-π зв'язку.

B. Модель характеризується зв’язуванням масивного гетероциклічного ядра, об’єм якого перевищує об’єм фосфотирозинового фрагменту. Стабілізація відбувається завдяки катіон-π зв’язкам з Arg 26 та утворенням водневого зв’язку з Arg 42. У цьому процесі задіяна майже уся поверхня трициклічного гетероциклу. Важливим є утворення водневого зв’язку між His та арилсульфомідною групою (вона також може виступати як імітатор фосфотирозину). Отже, цей водневий зв’язок надає значної жорсткості зв’язуванню ліганду з білком, а фурановий фрагмент додає стабілізації комплексу ліганд-протеїн завдяки Ван-Дер-Ваальсовій взаємодії.

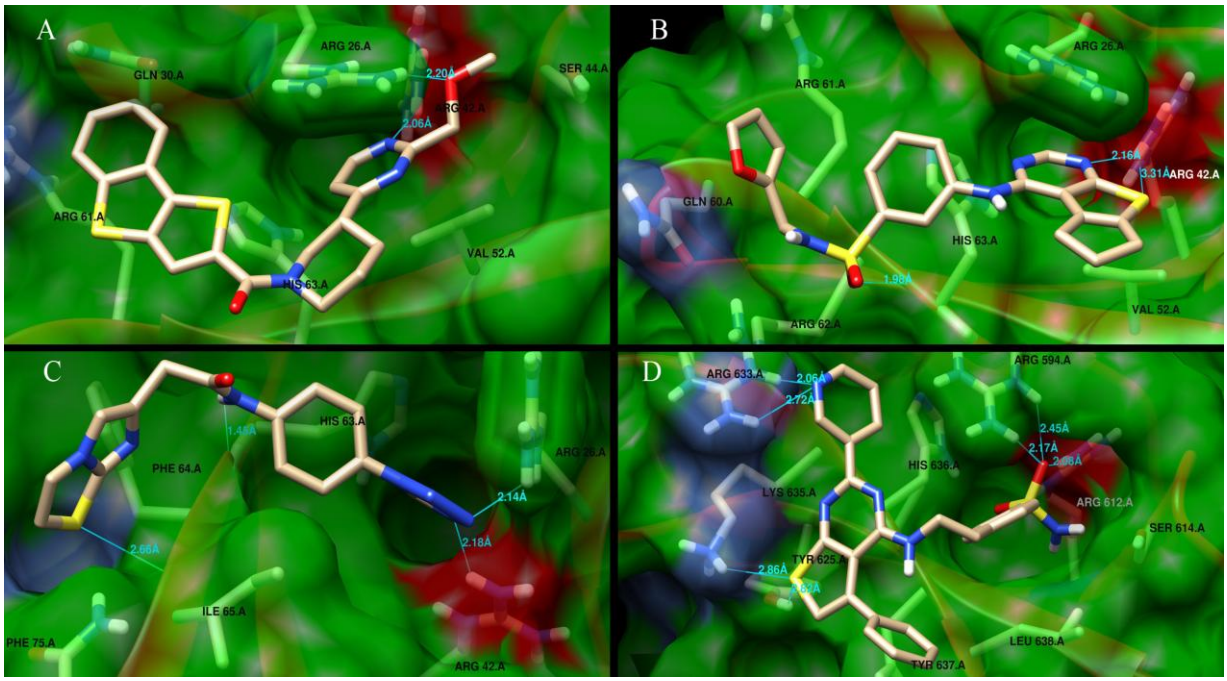


Рис. 3.10. Варіанти зв’язування доменів SH2 на прикладі центроїдів 2KK6 (A, B, C) та 1UUS (D) (стрічкова та поверхнева діаграми). Поверхня домену показана як напівпрозора частина, вторинні елементи – червоного кольору; αA- і αB-спіралі - справа. Ключові точки, що визначають модель зв’язування, розфарбовані у синій та червоний кольори.

С. Модель характеризується частковим заповненням фосфотирозинового сайту. Нітроген – вмісний гетероцикл утворює водневі зв’язки з Arg 42 та Arg 26. У цьому випадку імітатором фосфотирозину виступає тетразольний фрагмент, що є класичним електростатичним імітатором кислотної групи. Також модель передбачає збільшення жорсткості зв’язування пептиду

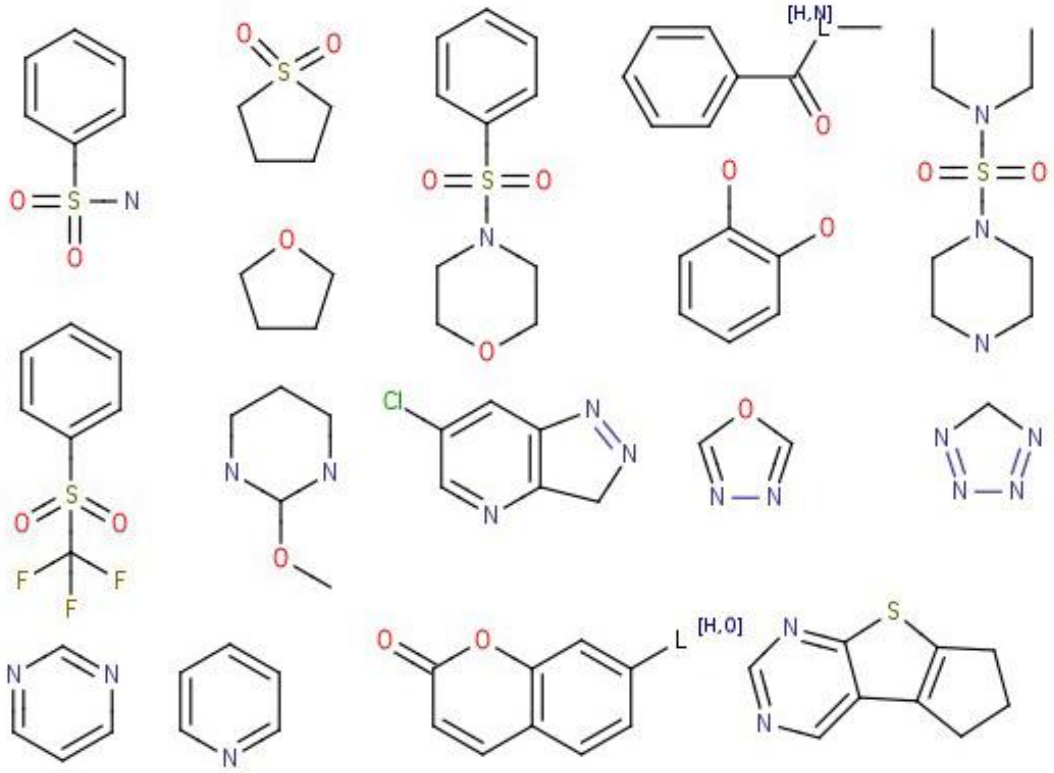


Рис. 3.11. Фрагменти структур, які виступали як імітатори фосфотирозину.

завдяки взаємодії NH групи ліганду з His 63, що входить до іншої частини кишені зв’язування, але відіграє важливу роль щодо зв’язування. Крім того, мають місце стеричні взаємодії з Gln 67, Ile 65, His 63, Phe 75 та Phe 64. Завдяки тому, що ліганд має доволі плоску просторову конформацію, він практично повністю розкиданий по кишені домену і в обох частинах центроїду зв’язування є неповним. Однак, не зважаючи на це, структури, які були відібрані в рамках цієї моделі, виявили значний інгібуючий ефект.

D. Модель представлена центроїдом, в основі якого лежить PDB 1UUS. Принципово відрізняється за механізмом зв'язування з амінокислотним оточенням. Так, в рамках цієї моделі ліганд щільно заповнює сайт зв'язування фосфотирозину, утворюючи міцні водневі зв'язки між Arg 594 і Arg 612 та арилсульфатною групою, що виступає як імітатор фосфотирозину. Модель передбачає утворення додаткового кластеру водневих взаємодій між піридином та Arg633, який просторово наближений до центру кишені. Також масивна частина ліганду утворює стекінг та катіон-π взаємодії з Lys635 (зазначимо, що завдяки гнучкості Lys635 можливе утворення водневого зв'язку між ним та лігандом), His636 і Tyr637 залишками. Отже, ліганд міцно закріплюється в активному сайті і виявляє значний інгібуючий ефект.

Одержана бібліотека включала 3018 структур, кожна з яких містила 20-40 атомів. Результати комп'ютерного моделювання показують, що відібрані структури утворюють складну мережу хімічних взаємодій у фосфотирозин – зв'язуючій кишені та відображають усі можливі варіанти блокування доменів SH2. Це свідчить про значне хімічне різноманіття цієї бібліотеки та велику гнучкість фосфотирозинової кишені. Крім того, виявлені специфічні групи, які здатні імітувати pTyr (рис. 3.11).

На останньому етапі докінг проводили з відібраними структурами при застуванні MOE. Зазначимо, що використали лише сім центроїдів, враховуючи велику ресурсоємність розрахунків такого характеру. Так, у кожен центроїд докувалось 50 – 800 тис. речовин. Кінцевий сет речовин склав 10463. Відзначимо, що структури, одержані для різних мішеней, значно відрізняються одна від одної: з майже десяти тисяч перекриваються лише 705 речовин. Отримані моделі зв'язування повністю перекрилися з моделями, які були описані вище.

3.2.2. Фармакофорні моделі. При порівнюванні сайтів зв'язування відібраних структур та їхніх лігандів виявили значну подібність як перших (табл. 3.2 і 3.8), так і других. Для порівнювання використали програмний пакет UCFS Chimera. Розглянули два варіанти порівняння структур: у першому випадку застосували алгоритм Нідлмана-Вулша та матриці порівнювань BLOSUM-62 (усі структури порівнювали з однією шаблонною структурою (1O49)). Значення Rmsd не перевищувало 1 Å (табл. 3.2 і 3.8); у другом випадку усі структури порівнювали між собою. Одержали: Q-score 0.525 і середне значення Rmsd 0.860 Å. Беручи до уваги вищезазначені дані, порівняли механізм зв'язування відібраних PDB структур (рис. 3.12, верхній). Так, у всіх випадках зв'язування включає такі ключові позиції (рис. 3.12): (1) активна частина кишені, яка зв'язується з pTyr/карбоновою кислотою і характеризується великою кількістю донорів H-зв'язку (Arg, Lys); (2) безпосередньо біля сайту зв'язування pTyr знаходиться ароматична/гідрофобна частина ліганду, яка зумовлена 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 з групою NH пептидного ліганду; (4) наявність однієї або двох гідрофобних кишень, які знаходяться за центром активного сайту і, зазвичай, заповнені гідрофобною частиною ліганду (рис. 3.3.). Виходячи з цього, побудували одну структурно-залежну фармакофорну модель, представлену на рис. 3.12, нижній.

Для побудови ліганд-залежних 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.

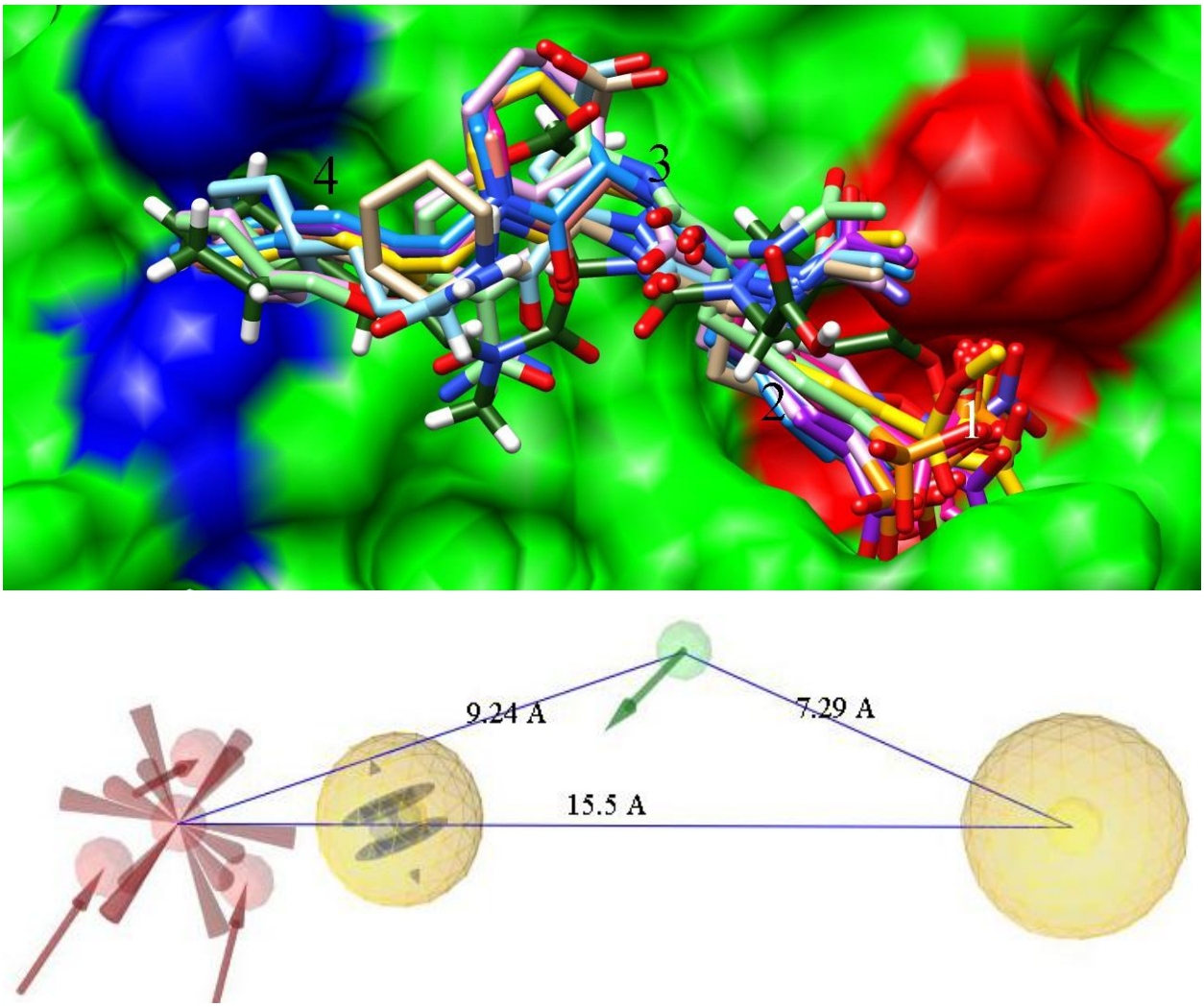


Рис. 3.12. Накладання усіх наявних кристалів доменів SH2 з лігандами (верхня частина рисунку: 1A1E - сірий, 1BKM - блакитний, 1FBZ - рожевий, 1IJR - світлозелений, 1SKJ - темнозелений, 1O42 - помаранчевий, 1O43 – білий, 1O44 - темнорожевий, 1O46 - жовтий, 1O47 - синій, 1O48 - пурпуровий, 1O49 - фіолетовий). Також наведено чотири основні місцеположення зв’язування (1 2, 3 і 4), на основі яких була побудована загальна фармакофорна модель (нижня частина рисунку).

При порівнянні одержаних моделей (табл. 3.9) і моделі, представленої на рис. 3.12, виявили, що практично усі ці моделі характеризуються такими властивостями: наявність акцептора Н-зв'язку (місцеположення 1, рис. 3.12); поряд знаходиться гідрофобна або ароматична частина ліганду (місцеположення 2, рис. 3.12) та гідрофобна частина ліганду (місцеположення 4, рис. 3.12).

Таблиця 3.8.

Порівнювання відібраних структур зі структурою 1O49, яку використовували у докінгу

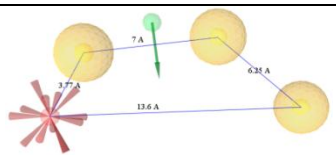
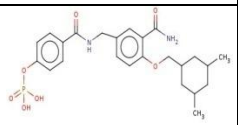
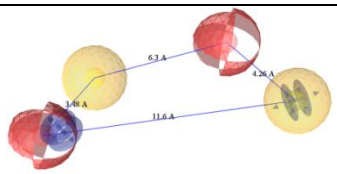
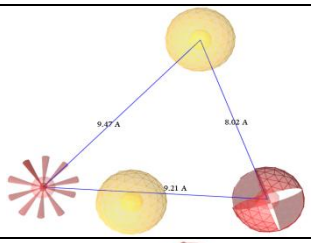
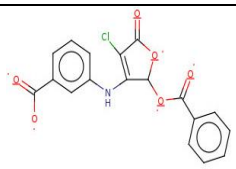
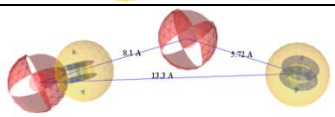
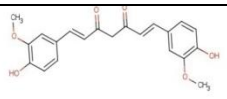
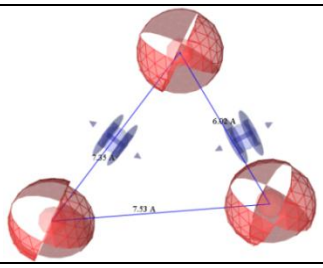
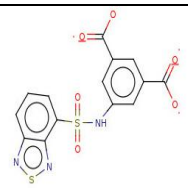
PDB	1A1E	1BKM	1FBZ	1IJR	1SKJ	1O42	1O43	1O44	1O46	1O47	1O48
Rmsd, Å	0.76	0.65	0.80	0.84	0.69	0.23	0.21	0.19	0.21	0.21	0.16
sequence alignment score	528	518	334	336	515	551	547	551	551	551	547

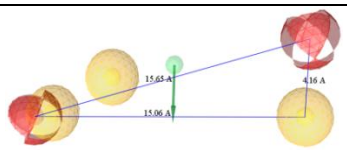
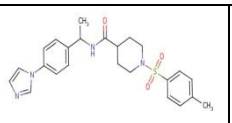
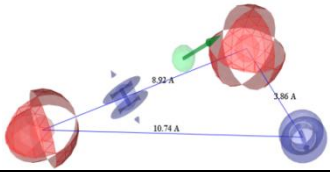
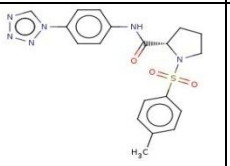
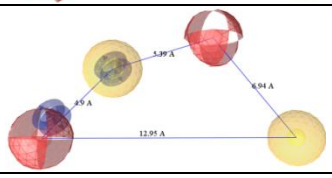
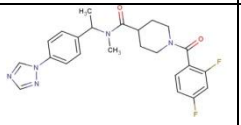
Після завершення побудови фармакофорних моделей звернули увагу на те, що, незважаючи на значні конформаційні обмеження при генерації усіх можливих стереоізомерів, речовини здатні значно вигинатися. Для вирішення цієї проблеми логічно було б провести вирівнювання усіх структур кластера за максимально витягнутою структурою (у нашому випадку це складає 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, такого вирівнювання не проводили. До того ж, як правило, такі вигини не є випадковим явищем. Так, запропонована 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 4 основних підмножин фармакофорної моделі

(описаних вище, як механізм зв'язування ліганду з протеїном) за винятком того, що у моделях № 2, 4 і 8 донорна група у положенні 3 замінена на акцепторну (це може бути пов'язано з тим, що, зазвичай, поряд з O His знаходиться группа NH Tyr і, можливо, у цьому випадку утворення H-зв'язку відбувається саме з цією амінокислотою).

Таблиця 3.9.

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

№	Фармакофорна модель	Репрезентативна речовина	AUC (area under ROC curve)	EF (enrich ment factor)	THR,% true hit rate	FHR,% false hit rate	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
8			1.00	209	100	0.41	522	17

3.2.3. Валідація фармакофорних моделей. В рамках ROC-аналізу визначали якість розподілу речовин на активні та неактивні, використовуючи значення AUC, загальну специфічність, істинно – позитивні значення (TPR), помилково – позитивні значення (FPR). З табл. 3.9 видно, 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; абсолютно протилежна ситуація має місце з моделлю №5. З одержаних результатів 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 у загальних рисах відповідають основній фармакофорній моделі з незначними змінами.

3.2.4. Віртуальний скринінг та побудова моделей GTM. Усі фармакофорні моделі використали для проведення віртуального скринінгу, базуючись на результатах докінгу (понад 10 тис. речовин). Загалом після цього одержали 1816 активних речовин (результати по кожнаму кластеру наведені в табл. 3.9), які потім аналізували за допомогою методу GTM.

Моделі GTM будували на основі 78 активних і 43 неактивних речовин. Використовуючи програму ISDA Fragmentor, відібрали дескриптори (у нашому випадку фрагменти речовин) в діапазоні від 2 до 8 атомів та хімічних

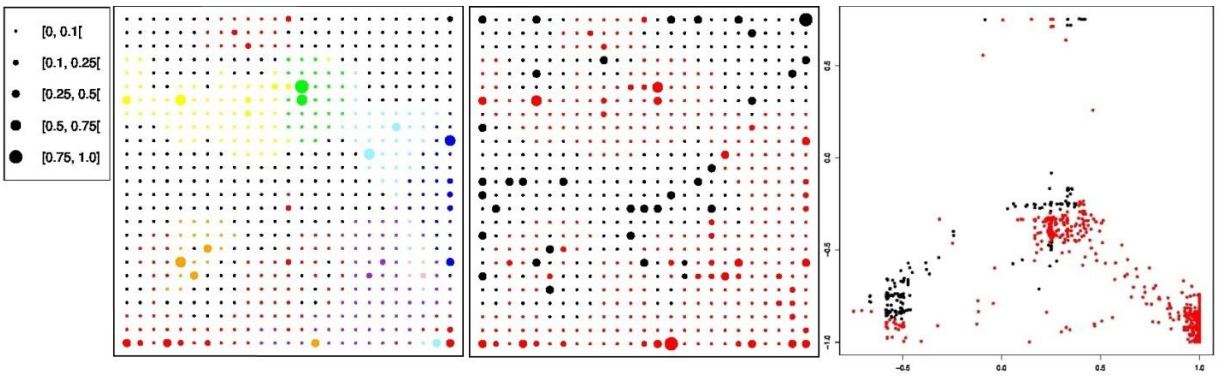


Рис. 3.13. GTM візуалізація одержаної моделі. Карта зліва - GTM модель з урахуванням розбиття усіх активних речовин на різні кластери: чорний колір – неактивні речовини (не враховували при наведені вище досліджуваних кластерів); червоний колір – кластер 1 (сінгелтоно; речовини, які не піддались кластеризації); синій колір – кластер 2 (дигідрокси/метокси-фенілаланіни); зелений колір – кластер 3 (тріазолпірадозіни-6-піридіни); помаранчевий колір – кластер 4 (арілсульфаміди та арілсульфони); пурпурнйи колір – кластер 5 (карбонові кислоти); блакитний колір – кластер 6 (фосфати і фосфанати); зелений колір – кластер 7 (паразаміщені арілазоли). Карта в центрі - GTM модель без урахування розбиття на кластери (червоний – активні, чорний – неактивні речовини). Карта справа - результат передбачення в рамках GTM моделі.

зв’язків. Одержали 4155 дескрипторів і далі побудували моделі GTM, з яких відібрали декілька найбільш вдалих. Розрахунки проводили за усіма 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 (при побудові моделі за кожною комбінацією параметрів трейнінг-сет поділяли на три частини, дві з яких використовували для побудови моделі, а одну – для її валідації). Найкращі параметри розрахунків наведено на рис. 3.13 і табл. 3.10.

Зазначимо, що оскільки не всі активні речовини були використані для побудови фармакофорних моделей у програмі LigandScout, вони були повторно перекластеризовані для оцінки взаємного розміщення одна відносно одної. Як результат, на GTM-картах отримали 7 (не враховуючи неактивні речовини) кластерів, які суттєво відрізнялися один від одного (середнє значення Tc=0.22) і характеризувалися значною подібністю в межах кожного окремого кластеру (середнє значення Tc=0.92). Це такі сполуки, як сульфіди, карбонові кислоти, фосфати і фосфанати, паразаміщені арілазоли та інші. Беручи до уваги фармакофорні моделі, наведені вище, таке різноманіття активних речовин свідчить про значну структурну варіативність сайту зв’язування доменів SH2. Доказом цього є також і те, що загалом вдалося виділити 7 доменів кластерів, у яких гомологічність не перевищує 30%. Крім того, важливо зазначити, що ці фармакофорні моделі будували на незначній кількості структур, що, в свою чергу, вказує на можливість існування інших варіантів зв’язування доменів SH2.

Незважаючи на незначну подібність одержаних кластерів між собою, спостерігається їхнє часткове перекриття. Так, усі кластери мають перекриття з неактивними речовинами (середнє значення Tc=0.26, максимальне - Tc=0.56, мінімальне - Tc=0.12) (рис. 3.13). В межах GTM моделі передбачення мало місце лише за кластерами 1, 4, 5 і частково 6. Отже, кінцева бібліотека склала 1607 речовин (273 речовини було відкинуто).

Таблиця 3.10.

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

3.3. Експериментальна перевірка

3.3.1. Оцінка впливу відібраних речовин на фосфорилювання домену SH2 протеїну STAT3 та проліферацію пухлинних клітин. Спочатку протестували відібрані структури на доменах SH2 протеїну STAT3. Перевірку проводили на лінії клітин L540 за концентрації 30 мкмоль/л. Так, теоретичні розрахунки відрізняються від експериментальних результатів, де

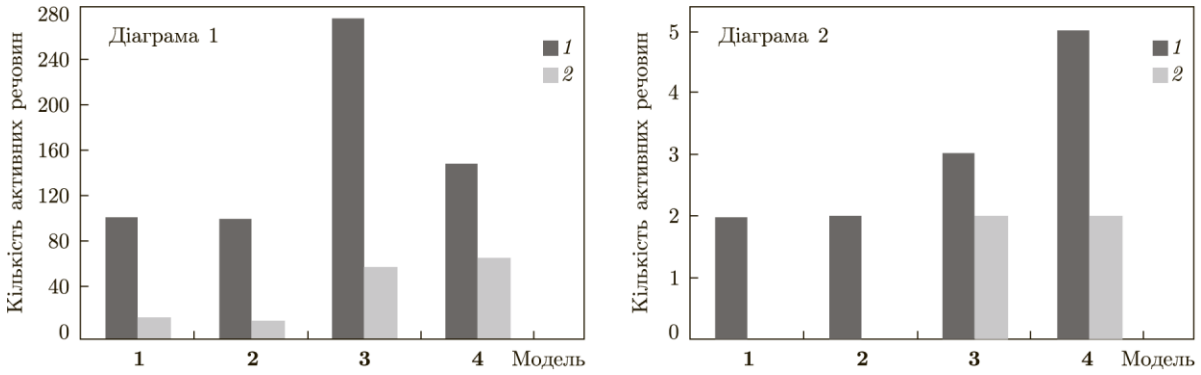


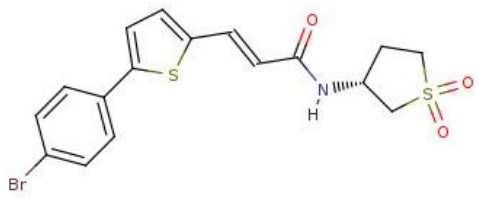
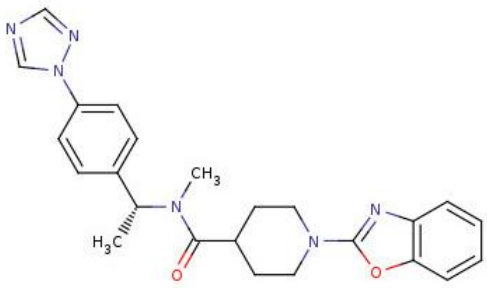
Рис. 3.14. Порівняння теоретичних даних інгібування (Діаграма 1) з експериментальними даними (Діаграма 2). Тут O (1) та N (2) - хімічні елементи, за якими відбувалося фільтрування потенційних лігандів у кишені зв'язування. По вісі *модель* позначені групи зв'язування (рис. 3.8, 3.9).

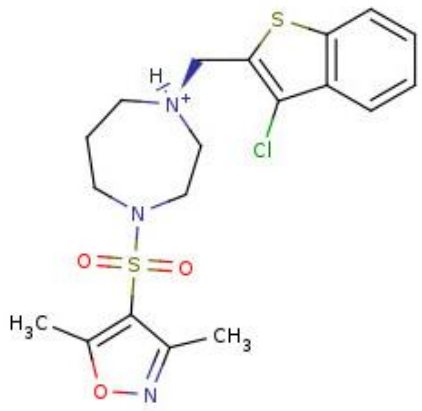
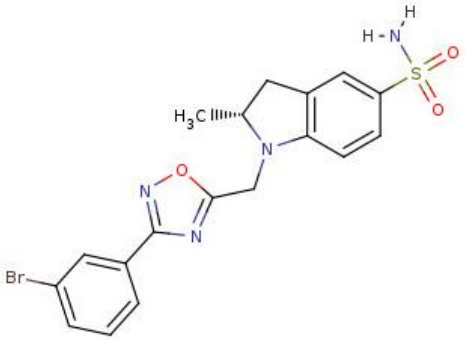
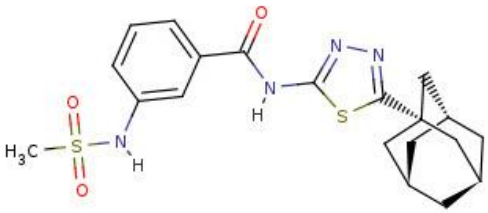
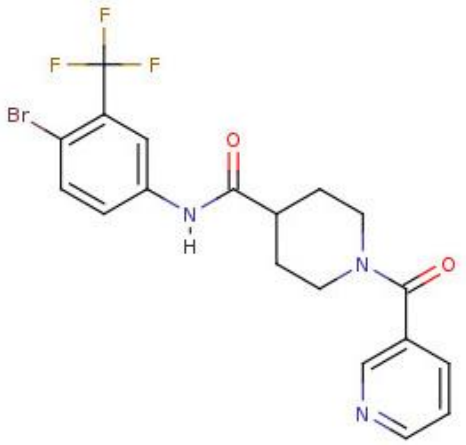
найбільш активними виявились ліганди, отримані в рамках четвертої, а не третьої моделі, як передбачила теорія. Як видно з діаграми 1 (див. рис. 3.14), після фільтрування залишилося трохи більше тисячі потенційних лігандів домену SH2. При цьому найбільш активними були ті ліганди, які зв'язувались з атомами кисню. Найбільш вдалою виявилася саме третя модель, де спостерігається найбільша кількість потенційно активних лігандів.

Це може бути пов'язано з тим, що 715 Pro істотно виступає над кишенею зв'язування, завдяки чому ліганд практично у всіх випадках змушений зв'язуватись всередині кишені і, відповідно, закривати її поверхню. Діаграма 2 (див. рис. 3.14) демонструє результати впливу відібраних речовин на фосфорилювання STAT3.

Таблиця 3.11.

Репрезентативні активні сполуки щодо STAT3

Речовина	Інгібування, %
	68.7
	67.5

	62.7
	65.0
	69.9
	72.3

Однак варто зазначити, що усі чотири теоретичні моделі враховують активні сполуки, які суттєво відрізняються за механізмом зв’язування з амінокислотним оточенням. Тому наразі першочерговим завданням є застосування розробленої методики для цілеспрямованого відбору

низькомолекулярних лігандів на усю платформу доменів SH2. Також підкреслимо, що активні речовини, отримані шляхом цього тестування, застосували для побудови не прямих/ліганд-залежних фармокофорних моделей (табл. 3.9). Репрезентативні активні сполуки щодо STAT3 наведені у табл.3.11.

На наступному етапі для оцінки біологічної активності структур використали дві бібліотеки: одна, одержана внаслідок застосування методів молекулярного моделювання (результати GTM та репрезентативна вибірка з нересурсоємного варіанту молекулярного докінгу; загалом 3018 структур), та випадковий сет бази лікоподібних речовин підприємства Enamine (2905 речовин). Відзначимо, що випадкова бібліотека містить достатньо різні речовини, які навіть можуть бути токсичними (наприклад, акцептори Міхаеліса) щодо тієї чи іншої лінії клітин. Відібрана ж нами бібліотека потенційних лігандів не містить таких речовин.

Чотири клітинних ліній K562, U937, A431 і LNCap використали для біологічного тесту. Біологічну активність визначали шляхом оцінки проліферації клітин. У всіх відібраних клітин функції залежать від гіперекспресії конкретного протеїну, що містить домени SH2, або взаємодії з іншими протеїнами через ці домени. Наприклад, проліферація клітинної лінії A431 пов’язана з гіперекспресією EGFR (epidermal growth factor receptor), який взаємодіє з GRB2, NCK1 і NCK2 через домен SH2. Клітинна лінія K562 містить гіперекспресовайний SH2-зв’язуючий протеїн, наприклад CD7, що взаємодіє із субодиницею p85 кінази Pi3. Характерною особливістю клітинної лінії K562 є залежність від комплексу BCR-Abl, що також містить домен SH2. Очікувалось, що зв’язування з доменами SH2 викличе негативний ефект на клітинну проліферацію. Інші дві лінії клітин застосували для диференціації сполук, що мають загальну токсичність за відсутності впливу на домени SH2. Наприклад, у клітинній лінії U937 присутня незначна експресія EGFR. Існують докази того, що проліферaційні ефекти цієї клітинної лінії за впливу на кіназні шляхи

передачі сигналів (у тому числі EGFR) є повільними. Тоді як для LNCap клітин, експресія EGFR та стимуляція потребують додаткової активації. Однак, важливо зазначити, що одержані результати можна інтерпретувати як непряме зв'язування доменів SH2 з відповідними лігандами, оскільки клітинні лінії містять надзвичайно велику кількість протеїнових комплексів, які здатні або не здатні впливати на домени SH2 та їхнє зв'язування.

Саме тому використали випадковий сет для перевірки ефективності якості розрахованої бібліотеки потенційних лігандів. Основною характеристикою сету, які використовували для порівнювання результатів, було число речовин, що селективно інгібували проліферацію клітинних ліній. Крім того, активні сполуки впливають на проліферацію лише однієї лінії клітин і практично або зовсім не впливають на проліферацію іншої лінії клітин. Активність речовин визначалася у % інгібування щодо позитивного контролю. Відомий протипухлинний агент доксорубіцин використали як позитивний контроль за концентрації, що призводить до 100% проліферацийної інгібіючої дії. Клітинні лінії з DMSO використали як негативний контроль. Відібрані речовини визначали як активні, коли вони впливали лише на одну лінію клітин серед чотирьох, вважаючи, що контрольні лінії клітин зазнають значно меншого впливу на проліферацію при зв'язуванні з доменами SH2. Також селективність між клітинними лініями може виникати завдяки високій вібірковості досліджуваних речовин до певного домену SH2. Результати порівнювання кількості активних речовин для кожної лінії клітин наведені в табл. 3.12.

Таблиця 3.12.

Кількість селективних хітів у таргетній бібліотеці доменів SH2 та випадковій вибірці

Лінія клітин	K562	U937	A431	LNCap	Всього
Випадкова вибірка	5	7	26	19	57
Таргетна бібліотека доменів SH2	19	3	55	6	83

Як видно з табл. 3.12, загальна кількість селективних хітів у бібліотеці в 1.5 рази більша за кількість хітів у випадковій вибірці, причому для ліній A431 та K562 це значення складає відповідно 2.1 та 3.8.

Чисельно оцінимо результат експериментальної перевірки для бібліотеки SH2. Нехай N_R = 2905 та N = 3018 – кількість сполук у випадковій вибірці і таргетній бібліотеці, відповідно. Припустимо, що частка селективних хітів певного сорту у початковій колекції сполук складає p. Тоді очікувана кількість хітів у випадковій вибірці складає pN_R і, якщо в експерименті їх кількість виявилася рівною n_R, тоді їх частку в колекції можна оцінити як $p = n_R/N_R$. У нашому випадку, коли N =3018, а p <<1, можна скористатися наближенням нормального розподілу з математичним сподіванням Np та дисперсією Np(1-p). Розглядаючи довірчий інтервал у три середньоквадратичних відхилення, можна стверджувати, що з ймовірністю у 99.7% кількість хітів у вибірці розміром N лежатиме в межах:

$$n = pN \pm 3\sqrt{Np(1-p)} = n_R \frac{N}{N_R} \pm 3\sqrt{n_R \frac{N}{N_R} \frac{N_R - n_R}{N_R}}$$

Результати обчислення за цією формулою очікуваної кількості хітів у випадковій вибірці розміром 2905 сполук наведені у табл. 3.13.

Таблиця 3.13.

Межі очікуваної кількості селективних хітів для випадкової вибірки розміром 2905 сполук для довірчого інтервалу у три середньоквадратичних відхилення та реальна кількість хітів у бібліотеці SH2

Лінія клітин	K562	U937	A431	LNCap	Всього
Очікування для випадкової вибірки	5.2 ± 6.8	7.3 ± 8.1	27.0 ± 15.5	19.7 ± 13.3	59.2 ± 22.9
Таргетна бібліотека доменів SH2	19	3	55	6	83

Наприклад, максимальна очікувана кількість «випадкових» селективних хітів для лінії A431 у бібліотеці складає 42 (табл. 3.13). Кількість виявлених селективних хітів цієї лінії у бібліотеці (55) є вдвічі більшою за їх середню кількість у випадковій вибірці (26) (табл. 3.12).

З отриманих результатів можна зробити два висновки щодо практичної цінності бібліотеки SH2:

1. застосування бібліотеки для пошуку селективних інгібіторів для певних ліній клітин є щонайменше у 2 – 3 рази ефективнішим з практичної точки зору, ніж застосування випадкової вибірки;

2. з ймовірністю 99.7% можна стверджувати, що серед 55 селективних хітів для лінії A431 є щонайбільше 42 «випадкових», а решта 13 (щонайменше) є наслідком успішного відбору за допомогою застосованих методів.

3.3.2. Дія відібраних речовин на кіназну активность Btk та Syk кіназ. 80 селективних інгібіторів, які виявили у клітинному скринінгу, відібрали для перевірки у біохімічному експерименті. Їх тестували на кіназах людини Btk і Syk (містять відповідно 1 та 2 домени SH2). Використали метод люмінесцентної детекції АДФ. Ензиматичну активність кінази вимірювали за швидкістю перетворення АТФ на АДФ.

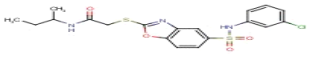
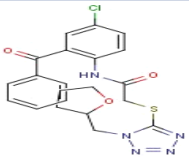
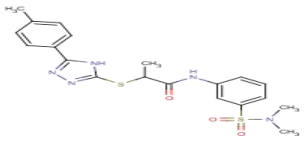
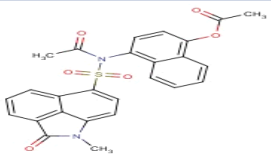
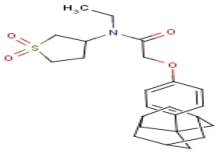
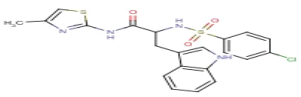
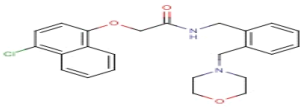
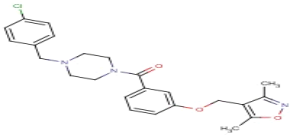
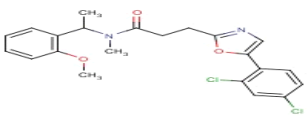
Передбачалося, що зв'язування сполук з доменами SH2, які входять до складу кіназ, впливатиме на кіназну активність. Як позитивний контроль використовували сигнал у комірках, до яких не додавали відповідний пептидний субстрат. Експеримент проводили у квадруплікатах, тобто кожна сполука була додана до чотирьох комірок плати, сигнал по яких усереднювався.

Кількість хітів, тобто сполук, що виявили інгібування на рівні не менше 50% для кіназ Btk та Syk склала відповідно 1 та 31. Інформацію про селективні хіти, які виявили найбільшу активність як у біохімічному, так і у клітинному експериментах, наведено у табл. 3.14.

Зрозуміло, що наведені результати біохімічного, а тим більше клітинного скринінгу бібліотеки SH2, не є прямим доказом зв'язування виявлених хітів саме з доменами SH2. Дійсно, у клітині присутні безліч протеїнів та інших макромолекулярних мішеней, зв'язування з якими може призводити до отриманого ефекту. Аналогічно, причиною інгібування кіназної активності у біохімічному експерименті не обов'язково є зв'язування саме з доменом SH2. Проте такі застереження не зменшують практичної цінності одержаних результатів, найважливішим з яких є виявлення десятків селективних хімічно різноманітних біологічно активних сполук. Їхня активність та селективність може бути предметом подальшої оптимізації, а валідація зв'язування саме з доменами SH2 може бути проведена у подальшому біофізичними методами з використанням ізольованих доменів SH2.

Таблиця 3.14.

Сполуки, які виявили селективну активність у біохімічному та клітинному експериментах

	Інгібування (%)					
	Кінази		Лінії клітин			
	SYK	BTK	K562	A431	U937	LNCap
T6056336 	-7	83	43	49	80	33
T5826298 	20	80	28	74	47	7
T5821946 	11	74	6	62	52	-6
T5409531 	17	61	27	32	129	40
T5415855 	14	60	39	78	55	33
T6534000 	22	58	43	27	48	101
T5572569 	16	54	73	53	94	17
T6247952 	23	52	7	101	66	101
T6452563 	-1	50	108	95	39	104

РОЗДІЛ 4
УЗАГАЛЬНЕННЯ РЕЗУЛЬТАТІВ ДОСЛІДЖЕНЬ

Розуміння механізмів взаємодії доменів SH2 з різноманітними субстанціями базується на сучасних уявленнях про їх структуру та участь у клітинних PPI, в межах чого можливе зменшення епіструктурного напруження [27], що часто викликає структурні дефекти протеїнів [28]. Результатом порушень роботи доменів SH2 також можуть бути різні захворювання (табл. 1.1, табл. 1.2.). Більше того, враховуючи значний масив структурних даних, цей клас доменів є зручною мішенню для пошуку нових лікарських засобів (здебільшого щодо профілактики онкозахворювань). Наразі виявлено 18 мутацій, здатних спричиняти розлади організму людини (лейкемія, порушення розвитку, діабет, імунодефіцит (табл. 1.1.)). Вони пов’язані з втратою/посиленням функцій доменів SH2. Наприклад, деякі мутації впливають на зв’язування SH2D1A/SAP або Btk з пептидом, що викликає імунодефіцит, зумовлений ензиматичною активністю тирозинфосфатази Shp2 [68].

Під час виконання роботи застосували комплекс теоретичних і експерементальних методів для пошуку і тестування потенційних лігандів щодо доменів SH2. Застосований алгоритм цих методів дав змогу охарактеризувати 219 наявних структур доменів SH2, відібрати три різні варіанти баз потенційних лігандів щодо них, протестувати їх на п’яти лініях клітин та двох SH2-вмісних кіназах.

Для охарактеризування та відбору репрезентативних структур доменів SH2 поетапно використали методи структурного аналізу (геометричний аналіз, множинне вирівнювання амінокислотних послідовностей тощо). Врахували, що кожна структура з PDB може містити декілька копій.

Загалом усі структури поділили на 1129 окремих копій. Отримали 56 унікальних структур, які потім знову поділили на шість груп (враховуючи всю довжину амінокислотних послідовностей доменів). Використовуючи програмні пакети Chimera, PRALINE та Jalview порівняли одержані структури та їхні групи. Оцінка якості цих груп виявила високий та середній рівні консервативності (група 1 – 92/91%, група 2 – 81/87%, група 3 – 56/62%, група 4 – 37/51%, група 5 – 50/68%, група 6 – 51/61%) та незначні коливання Rmsd (група 1 – 0.8 Å, група 2 – 1.16 Å, група 3 – 1.6 Å, група 4 – 1.9 Å, група 5 – 2 Å, група 6 – 1.4 Å). Важливо, що середнє значення консервативності є значно вищим у сайті зв'язування порівняно з усім доменом. Це підтверджує загально відомий факт, що функціональні ділянки протеїнових структур завжди є більш консервативними, що і забезпечує виконання їх функцій. Так, консервативність сайту зв'язування для різних груп становить: група 1 – 94/93%, група 2 – 89/96%, група 3 – 59/67%, група 4 – 49/60%, група 5 – 53/70%, група 6 – 53/59% (*більш детально див. табл. 3.3*). Групи 1 і 2 є найбільш консервативними, зростання ідентичності структур (якщо порівнювати з усіма доменами) складає 61% і 54% та подібності 55.1% і 35.2%. Групи 3-6 є набагато менш консервативними порівняно з першими двома: зростання ідентичності (порівняно з усіма доменами) становить 19%, 14%, 8% і 14%, подібності 14,5%, 12,2%, 15,5 і 11,63%, відповідно, що, ймовірно, і призводить до значного зростання середньої площі зв'язуючої поверхні доменів.

Кишеня зв'язування займає значну частину поверхні протеїну – в середньому ¼ від загальної площі домену. У всіх випадках вона є плоскою і складається з 20-24 амінокислот, які формують вісім основних послідовностей кишені зв'язування. Водночас, групи 3, 4 і 6 містять одну додаткову точку (табл. 3.3).

Кишеню зв'язування доменів SH2 формують такі структури: βB-, βG-складки і αB-спіраль (рис. 3.4, 3.5; табл. 3.3). Загалом її поділили на такі ділянки: pTyr-зв'язуюча, яка формується послідовністю ITR (ділянка 1),

FLVRESETT (ділянка 3) та C (ділянка 4) (лише послідовність RESETT взаємодіє з лігандом); центральна частина - ділянки 2 (E) і 5 (VKHYKIR) (βD-складка та αA-спіраль); гідрофобна частина включає ITSRT (ділянка 6), Y (ділянка 7), ADGLC (ділянка 8) (βG-складка, αB- та αA-спіралі), позиції 67-70 і 87-91. Виявили чотири висококонсервативні позиції (>70%), якщо брати до уваги усі домени SH2: позиція 3 ділянки 1, позиції 1 і 3 ділянки 3 та позиція 3 ділянки 5.

Знайшли висококонсервативні позиції кишені зв'язування, які не є частиною будь-якого із сайтів зв'язування з вище описаних, наприклад, T (позиція 37 - група 3), L (позиція 69 - група 4) та Q (позиція 40 - група 6). Враховуючи ці дані, для кожного сайту зв'язування були відібрані дескриптори, на основі яких одержано вісім кластерів.

Конформаційний пошук виявив рухливість доменів SH2, яка знаходиться в діапазоні (1.5 – 5.5) Å (рис. 3.6). Так, pTyr – зв'язуюча частина є доволі гнучкою, а центральна – жорсткою. Виявили 39 можливих конформацій за мінімального значення Rmsd.

Редокінг виявив задовільні як геометричні, так і енергетичні можливості відтворення застосованих методів. Rmsd знаходиться в межах 2 Å (табл. 3.5 і 3.6). Отримані енергетичні параметри поділили на три групи (*більш детально див. Розділ 3.2.1*).

У результаті молекулярного докінгу виявили чотири основні моделі зв'язування типу "домен-ліганд" (рис. 3.10) та декілька хімічних груп, які здатні імітувати фосфотирозин (рис. 3.11):

A. водневі зв'язки утворюються з Arg 26 (крім того, спостерігається катіон-π взаємодія) і Arg 46. Відбувається стерична взаємодія з Gln 30, Val 52 не основної гідрофобної кишені (рис. 3.3, точка 4) та центральним His 63;

B. зв'язування з масивними гетероциклами; стабілізація відбувається завдяки взаємодіям з Arg 26 (катіон-π) і Arg 42 (воднева). Важливим є утворення водневого зв'язку між His 63 та арилсульфомідною групою, що надає значної міцності зв'язуванню ліганда з доменом;

C. утворення міцних водневих зв’язків з Arg 42 та Arg 26 і NH групи потенційного ліганду з His 63 (центр кишені зв’язування). Присутність стеричних взаємодій з Gln 67, Ile 65, His 63, Phe 75 та Phe 64. Структури, відібрані в рамках цієї моделі, виявили значний інгібуючий ефект;

D. модель принципово відрізняється за механізмом зв’язування. Фосфотирозин-зв’язуюча кишеня щільно заповнюється, утворюються міцні водневі зв’язки з Arg 594 і Arg 612. Arg633 утворює другий кластер водневих зв’язків (потенційно інший фосфотирозин-зв’язуючий сайт). Спостерігається значна стекінг взаємодія з центральною частиною кишені. Речовини цієї моделі також виявили значний інгібуючий ефект.

Таблиця 4.1.

Енергетичні параметри зв’язування “домен-ліганд” (кДж / моль { Pi = -log10 (ки) })

модель	Cntc	Hbnd	Intl	Bump	FreE
A	-52.8	-1.9	3.8	2.5	-16.0
B	-58.5	-2.7	6.6	4.6	-16.6
C	-47.9	-2.9	3.3	2.9	-19.5
D	-71.6	-4.7	8.2	4.1	-23.5

Відібрані моделі відображають основні варіанти зв’язування, які виявили після фільтрування. Як бачимо з табл. 4.1, загальна енергія зв’язування (FreE) є мінімальною для моделі D, але у цьому випадку спостерігається значне внутрішнє напруження ліганду (Intl), незначна контактна енергія (Cntc) і відносно вище значення стеричних зіткнень (Bump). Незважаючи на те, що у моделях A і B спостерігають практично однакові значення Cntc, Bump, Intl, Hbond і Cntc. Такі відмінності пов’язані зі складностями зв’язування компактного фосфотирозинового сайту з великим поліциклом.

Згідно з наведених результатів модель C повинна бути найбільш відтворюваною. У цьому випадку спостерігається найбільша Cntc, практично мінімальні Bump, FreE та Intl.

У клітинному біологічному тесті виявили 56 лігандів з моделі C, 14 з D та 13 з A. Модель B не підтвердилась у жодному з випадків (можливо це пов'язано зі складностями зв'язування масивного гетероциклічного ядра з малим фосфотирозиновим сайтом зв'язування домену).

Загалом отримали три бібліотеки потенційних лігандів щодо доменів SH2. Основна з них склала 10463 (результат молеклярного докінгу), (перекривання між різними центроїдами становило 705 речовин (рис. 4.1)).

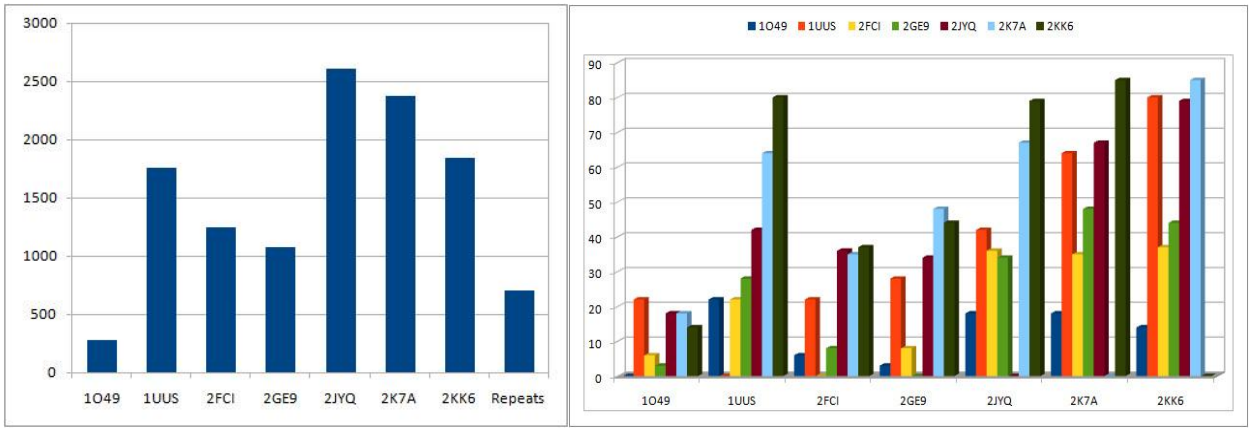


Рис. 4.1. Кількість відібраних речовин для кожного кластеру та їх перекривання.

Побудували дев'ять фармакофорних моделей: одна структурно-залежна і вісім ліганд-залежних. У всіх випадках зв'язування включає такі ключові моменти (рис. 3.12, табл. 3.10): (1) зв'язуння з pTyr/карбоновою кислотою, велика кількість донорів H-зв'язку (Arg, Lys); (2) ароматична/гідрофобна частина ліганду, Ван-Дер-Вальсова взаємодія здебільшого між Arg і Lys; (3) His утворює водневий зв'язок з групою NH пептидного ліганду; (4) взаємодія у гідрофобній кишені, яка знаходиться за центром.

Ліганд-залежні моделі № 1, 3, 6, 8 є найкращими (AUC 100%), № 4, 5 і 7 трохи гірші (AUC 75%). Моделі № 1, 2, 6 і 8 найкраще відбирають активні речовини і відсіюють декої; модель №3 відбирає речовини, надані тренінг-сетом, але відкидає декої; протилежна ситуація спостерігається у моделі №5. Так, всі отримані моделі доповнюють одна одну, частина відбирає усі активні речовини, інші відкидають більшість декоїв. Моделі № 3 і 5 не відповідають геометрії сайту зв'язуваання (причиною може бути здатність набувати інших конформацій молекулами, з яких сформовані моделі).

Після віртуального скринігу одержали 1816 активних речовин (результати по кожнаму кластеру наведені в табл. 3.10).

Таблиця 4.2.

Діапазон ADME параметрів для початкового сету, відібраної бібліотеки та речовин з відомою активністю

ADME	Вихідна база (10463 речовин)		Результуючий сет (1543 речовин)		Активні речовини (78)	
	Min	Max	Min	Max	Min	Max
Донори водневих зв'язків	0	6	0	5	0	5
Акцептори водневих зв'язків	2	14	2	11	3	10
LogP	-0.002	8.62	-0.035	6.326	0.123	5.571
LogS	-0.004	2.0	-0.88	2.0	-7.373	-2.235
Обертальні зв'язки	0	11	2	11	3	9
Молекулярна маса	241.2	644.6	254.2	545.6	236.6	545.5

Ці речовини у подальшому аналізували за допомогою методу GTM, в рамках якого розглянуто 78 активних і 43 неактивних речовин, які поділили на 4155 дескрипторів. Застосували такі параметри моделі GTM: 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). Одержали чотири найкращі варіанти розрахунків (табл. 3.11), які сформовані сімома кластерами речовин (сульфіди, карбонові кислоти, фосфати і фосфанати, паразаміщені арілазоли та інші): середнє значення Tc=0.22 між кластерами і Tc=0.92 – всередені кожного кластеру. Також спостерігали часткову подібність з неактивними речовинами (середнє значення Tc=0.26, максимальне - Tc=0.56 і мінімальне - Tc=0.12). Використовуючи одержані моделі, мало місце передбачення за кластерами 1, 4, 5 і частково 6. Результуюча бібліотека склала 1607 речовин.

Для перевірки того, наскільки репрезентативною є ця вибірка порівняно з використаною базою сполук, для неї побудували розподіл ADME-пов’язаних властивостей. На рис. 4.2 наведено розподіл у результуючому сеті порівняно з розподілом у вихідній бібліотеці. В табл. 4.2 наведено діапазони значень ADME параметрів для початкового сету, відібраної бібліотеки та речовин з відомою активністю.

Як видно з рис. 4.2, відмінності у розподілах є незначними. Це означає, що застосовані методи вибірки є репрезентативними відносно основних фізико-хімічних властивостей і дають змогу зменшити кількість сполук, не зменшуючи загальної якості вибірки. При цьому слід зазначити, що порівняно з початковим сетом у результуючого ADME параметри наблизились до параметрів активних речовин. Це дає підстави вважати, що результуюча вибірка є потенційно більш активною, ніж вибірка, отримана на попередніх етапах. Водночас, має місце помітне зміщення розподілу кількості обертальних зв’язків у бік їх збільшення порівняно з активними сполуками. Але це цілком логічно, оскільки молекули з більшою кількістю обертальних зв’язків здатні до утворення більш різноманітних конформацій.

Потенційні ліганди для домену SH2 протеїну STAT3 перевірили на L540 лінії клітин. Речовини тестували за концентрації 30 мкмоль/л упродовж

24 год. Найбільш вдалою виявилася модель № 4 (хоча за теоретичними розрахунками вона гірша за модель № 3). Найбільш активними виявилися

ліганди, в яких pTyr частину замінював скефорд з атомом кисню. Припускаємо, що це пов’язано з координацією Pro 715.

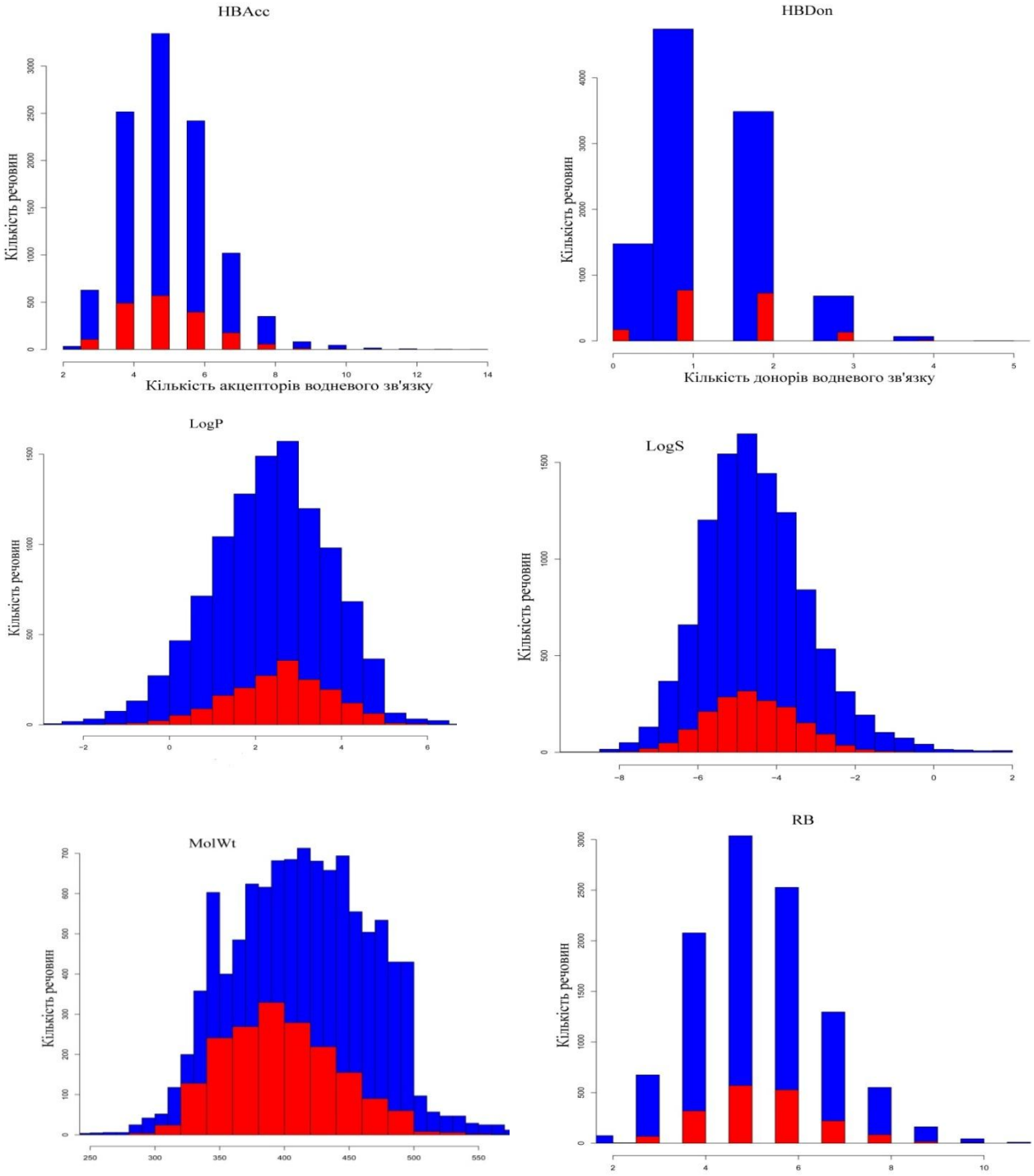


Рис.4.2. Розподіл ADME-пов’язаних фізико-хімічних параметрів у вибірці сполук (червоний) порівняно з вихідною бібліотекою (синій). HBAcc – акцептори водневих зв’язків, HBDon – донори водневих зв’язків, LogP –

коефіцієнт розподілу між октанолом і водою, LogS – розчинність у воді, Molwt – молекулярна маса, RB – обертальні зв'язки.

Активні речовини, отримані у цьому тесті, застосували для побудови не прямих/ліганд-залежних фармокофорних моделей (табл. 3.9).

Репрезентативні активні сполуки щодо STAT3 наведені у табл.3.11. Репрезентативний сет результуючих бібліотек потенційних лігандів та випадково відібраної бібліотеки протестували на K562, U937, A431 і LNCap (у кожній гіперекспресується протеїн, який містить домен SH2) лініях клітин та оцінили їх проліферацію. Загалом виявили 83 активних речовини у відібраній бібліотеці та 57 у випадковій (табл. 3.12). З активних речовин за моделлю A – 13, B – не містить жодної, C – 56 і D – 14. Кількість хітів у бібліотеці виявилась у 1.5 рази більшою за кількість таких у випадковій вибірці, зокрема, для A431 та K562 – у 2.1 та 3.8 рази, відповідно.

Інгібітори, виявлені у клітинному скринінгу, застосували у біохімічному експерименті на Btk і Syk кіназах людини. Використали метод люмінесцентної детекції АДФ. Для Btk та Syk знайшли 1 та 31 активних селективних сполуки, відповідно. Інформацію про селективні хіти, які виявили найбільшу активність як у біохімічному, так і у клітинному експериментах, наведено у табл. 3.14.

ВИСНОВКИ

1. Домени SH2 є висококонсервативними структурами, особливо в межах сайту зв’язування. Найбільш консервативним є pTyr-зв’язуюча частина домену, а найменш – гідрофобна частина кишені звя’зування.

2. За геометричними параметрами домени SH2 є практично ідентичними (значення Rmsd змінюється в межах 2 Å). Вони характеризуються незначною рухливістю, яка змінюється в межах 1.5–5.5 Å. Найбільш рухливою виявилась αA-спіраль, а найменш – βD-складка.

3. Отримані групи доменів SH2 характеризуються значною подібністю: в межах кожної окремої групи спостерігаються домени лише одного або подібних типів протеїнів.

4. Окрім консервативних ділянок в межах кожної окремої групи виявили чотири висококонсервативних позиції (>70%): позиція 3 ділянки 1, позиції 1 і 3 ділянки 3 та позиція 3 ділянки 5. Усі інші частини сайту зв’язування є менш консервативними (<70%). Також встановили декілька позицій, які не є унікальними для тієї чи іншої групи: T (позиція 37 - група 3), L (позиція 69 - група 4) та Q (позиція 40 - група 6).

5. Виділено ключові точки зв’язування типу “домен SH2-ліганд”: (1) зв’язування з pTyr/карбоновою кислотою (характеризується великою кількістю донорів H-зв’язку (Arg, Lys)); (2) біля pTyr-зв’язуючої частини знаходиться ароматична/гідрофобна частина ліганду; (3) центр кишені зв’язування визначається позицією O амінокислотного залишку другої амінокислоти (His); (4) наявність однієї або двох гідрофобних кишень, які знаходяться за центром активного сайту.

6. Отримали чотири основні моделі молекулярного докінгу зв’язування типу “домен SH2-ліганд”. Практично у всіх випадках ліганд щільно заповнює фосфотирозинову кишеню зв’язування та бере участь у водневих взаємодіях з ключовими амінокислотами Arg або Lys. Крім того,

присутні стеричні взаємодії між лігандом та амінокислотами Glu, Val, Pro, His і Ser. Під час біологічного тестування, модель, яка передбачає зв'язування з масивними гетероциклами не підтвердилась. Отримали три бібліотеки потенційних лігандів доменів SH2, що складали: 1015, 3018 і 10463 структур (останні дві використані у біологічному тестуванні).

7. Побудовано дев'ять фармакофорних моделей, які відображають декілька варіантів зв'язування доменів SH2 з лігандом. Валідація показала, що усі одержані фармакофори можуть бути застосовані для пошуку потенційних лігандів доменів SH2. Значення величини AUC в отриманих моделях знаходяться в діапазоні 75-100%. 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. Як наслідок, відібрано 1816 речовин, які застосували у GTM передбаченні і одержали 1607 речовин, що відповідають ключовим точкам взаємодії, застосованих під час докінгу та фармакофорного пошуку.

8. Перевірка отриманих потенційних лігандів доменів SH2 щодо їх впливу на проліферативну активність пухлинних клітин та кіназну активність Btk та Syk кіназ людини виявила значний інгібуючий ефект (65-95%).

СПИСОК ЛІТЕРАТУРИ

1. Visualization and targeted disruption of protein interactions in living cells / D. H. Henry, D. Wen; H. Jonas [et al.] // Nature Communications. – 2013. – Vol. 4. – P. 2660 –2670.

2. Sara C. Structural Classification of Small, Disulfide-rich Protein Domains / C. Sara, S. S. Krishna, V. G. Nick // J. Mol. Biol. – 2006. – V. 359. P. 215–237.

3. Richardson J. S. "The anatomy and taxonomy of protein structure" / J. S. Richardson. // Adv Protein Chem. - 1981. - Vol. 34. - P 167–339.

4. Bork P. "Shuffled domains in extracellular proteins" / P. Bork. // FEBS Lett. – 1991. – Vol. 286. – P. 47–54.

5. Smithgall T. E. SH2 and SH3 domains: potential targets for anti-cancer drug design / T. E. Smithgall. // J Pharmacol Toxicol Methods. – 1995. – Vol. 34. – P. 125 – 132.

6. Застосування методів молекулярного моделювання для пошуку нових біологічно активних речовин / В. В. Гурмач, О. М. Балинський, М. О. Платонов та інші. // Укр. біохім. журн. – 2015. – Т. 87. – С. 109 – 120.

7. Genome wide analysis of pathogenic SH2 domain mutations. / I. Lappalainen, J. Thusberg, B. Shen [et al.] // Proteins. – 2008. – Vol. 72. – P. 779 – 792.

8. TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. / V. Lacronique, A. Boureux, V. D. Valle [et al.] // Science. – 1997. – Vol. 278. – P. 1309 – 1312.

9. Fusion of TEL, the ETS-variant gene 6 (ETV6), to the receptor-associated kinase JAK2 as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia. / P. Peeters, S. D. Raynaud, J. Cools [et al.] // Blood. – 1997. – Vol. 90. – P. 2535–2540.

10. Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. / M. Tartaglia, C. M. Niemeyer, A. Fragale [et al.] // Nat. Genet. – 2003. – Vol. 34. – P. 148–150.

11. Метод молекулярного докінгу за участю SH2-доменів / В. В. Гурмач, О. М. Балинський, М. О Платонов та інші. // "Біотехнологія". – 2012. – Т. 5. – С. 31 – 40.

12. Пошук низькомолекулярних лігандів для SH2-доменів методом докінгу міжмолекулярних взаємодій / В. В. Гурмач, О. М. Балинський, П. О. Бориско та інші. // Доповіді НАН України – 2012. – Т. 8. – С. 141 – 146.

13. Comparative analysis of SH2 domain structures / V. V. Hurmach, M. O. Platonov, O. M. Boyko [et al.]. // Studia Biologica – 2015. – V. 9. – P. 5 – 14.

14. Design of potentially active ligands for SH2 domains by molecular modeling methods / V. V. Hurmach, O. M. Balinskyi, M. O. Platonov [et al.]. // Biopolymers and Cell – 2014. – V. 30. – P. 321-325.

15. New conformational properties of SH2 domain binding pocket / V. V. Hurmach, M. O. Platonov, Yu. I. Prylutskyy. // Ukr. Biochem. J. – 2015. – Vol. 87. – P. 133 – 140.

16. Гурмач В. В. Пошук низькомолеку-лярних лігандів для SH2 доменів методом молекулярного докінгу / В. В. Гурмач, О. М. Балинський, М. О. Платонов, П. О. Бориско, Ю. І. Прилуцький // VII міжнародній науково-технічній конференції «БФФХ-2012». – 2012. – С – 100.

17. Hurmach V. V. Desing and test of potentially active lihands for SH2-domains by molecular modeling methods / V. V. Hurmach, A. M. Balynskyy, M. O. Platonov, A. N. Boyko, P. O. Borysko, Y. I. Prylutskyy, A. Varnek// 2nd Ukrainian-French School-Seminar «Carbon Nanomaterials: Structure and Properties». – 2013. – P. 79.

18. Hurmach. V. SH2 doamins structure analisy / V. Hurmach // XI міжнародній науковій конференції студентів та аспірантів “Молодь і поступ в біології”. – 2015. – С. 12.

19. Hurmach V. V. Computational study of SH2 domains / V. V. Hurmach, A. A. Varnek, Yu. I. Prylutskyy // Ukrainian–German Symposium on Physics and Chemistry of Nanostructures and on Nanobiotechnology. – 2015. – P. – 262.

20. Hurmach V. Conformational analisys of Sh2 domain / V. Hurmach, A. A. Varnek, O. M. Platonov, Yu. I. Prylutskyy // International Conference for Young Scientists (CYS). – 2015. – P. – 93.

21. Phizicky E. M. "Protein-protein interactions: Methods for detection and analysis" / E. M. Phizicky, S. Fields //. Microbiological reviews. – 1995. – Vol. 59. – P. 94–123.

22. Rivas D. L. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. / D. L. Rivas, C. Fontanillo // PLoS Comput Biol. - 2010. - Vol. 6. - P. 10 - 18.

23. Jones S. «Principles of protein-protein interactions.» / S. Jones, J. M. Thornton // Proceedings of the National Academy of Sciences of the United States of America. – 1996. – Vol. 93. – P. 13–20.

24. «Characterization of protein-protein interfaces.» / C. Yan, F. Wu; L Jernigan R. [et al.] . // The protein journal. – 2008. – Vol. 27 – P. 59–70.

25. Jones S. «Analysis of protein-protein interaction sites using surface patches.». / S. Jones, J. M. Thornton // Journal of molecular biology. – 1997. – Vol. 272 – P. 121–132.

26. Janin J. The structure of protein-protein recognition sites. / J. Janin, C. Chothia // The Journal of biological chemistry. — 1990. — Vol. 265. — P. 16027–16030.

27. Fernandez A. «Epistructural tension promotes protein associations.». / A. Fernandez // Physical Review Letters – 2012. – Vol. 108. – P. 18 – 25.

28. Ariel F. A Structurally Encoded Signal for Protein Interaction Ariel / F. Ariel, S Ridgway // Biophys J. – 2003. – Vol. 85. – P. 1914–1928.

29. Arkin M. R. «Small-molecule inhibitors of protein–protein interactions: progressing towards the dream». / M. R. Arkin, J. A. Wells // Nature Reviews Drug Discovery. – 2004. – Vol. 3 – P. 301–317.

30. Chen J. «Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area». / J. Chen, N. Sawyer; L. Regan // Protein Science – 2013. – Vol. 22. – P. 510-515.

31. Ivanov A. A. "Targeting protein–protein interactions as an anticancer strategy". A. A. Ivanov, F. R. Khuri; H. Fu // Trends in Pharmacological Sciences. – 2013. – Vol. 34. – P. 393–400.

32. Whitty A. Between a rock and a hard place? / A. Whitty, G. Kumaravel // Nature Chemical Biology. – 2006. – Vol. 2. – P. 112–118.

33. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. / C. A. Lipinski, F. Lombardo, B. W. Dominy [et al.]. // Advanced drug delivery reviews. – 2001. – Vol. 46. – P. 3–26.

34. The role of protein interaction domains in the human cancer network. / S. S. Ibrahim, M. A. R. Eldeeb, M. A. H. Rady [et al.]. // Network Biology. – 2011. – Vol. – 1. – P. 59 – 71.

35. Pawson T. Assembly of cell regulatory systems through protein interaction domains. / T. Pawson, P. Nash // Science – 2003. – Vol. 300. – P. 445–452.

36. Chothia C. The relation between the divergence of sequence and structure in proteins. / C. Chothia, A. M. Lesk // The EMBO Journal. – 1986. – Vol. 5. – P. 823 – 826.

37. Pawson T. "SH2 domains, interaction modules and cellular wiring". / T. Pawson, G. D. Gish, P. Nash // Trends in Cell Biology. – 2001. – Vol. 11. – P. 504 – 511.

38. "Defining the specificity space of the human SRC homology 2 domain". / H. Huang, L. Li, C. Wu [et al.] . // Molecular & Cellular Proteomics. – 2008. – Vol. 7. – P. 768–784.

39. Sadowski I. "A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps". / I. Sadowski, J. C. Stone, T. Pawson // Mol. Cell. Biol. – 1986. – Vol. 6. – P. 4396–4408.

40. Russell R. B. "Conservation analysis and structure prediction of the SH2 family of phosphotyrosine binding domains". / R. B. Russell, J. Breed, G. J. Barton // FEBS Lett. – 1992. – Vol. 304. – P. 15 – 20.

41. "Phosphotyrosine signaling proteins that drive oncogenesis tend to be highly interconnected". / G. Koytiger, A. Kaushansky, A. Gordus [et al.] . // Mol. Cell Proteomics. – 2013. – Vol. 12. – P. 1204 – 1213.

42. Gu H. The Shp'ing news: SH2 domain-containing tyrosine phosphatases in cell signaling / H. Gu, B. G. Neel, L. Pao // Trends Biochem. Sci. – 2003. – V. 28, N 6. – P. 284–293.

43. The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling / A. L. Bernard, J. Karl, R. Monica [et al.] . // Molecular Cell. – 2006. – Vol. 22. – P. 851 – 868.

44. Brian S. SH2 domain structure and function / S. Brian // Biochim Biophys Acta. – 1995. – Vol. 1242. – P. 61 – 75.

45. "The SH2 Domain-Containing Proteins in 21 Species Establish the Provenance and Scope of Phosphotyrosine Signaling in Eukaryotes". / B. A. Liu, E. Shah, K. Jablonowski [et al.] . // Science Signaling – 2011. – Vol. 4. – P. 1 – 39.

46. The sequence of the human genome. / J. C. Venter, M. D. Adams, E. W. Myers [et al.]. // Science. – 2001. Vol. 291. – P. 1304–1351.

47. A quantitative protein interaction network for the ErbB receptors using protein microarrays. / R. B. Jones, A. Gordus, J. A. Krall [et al.] . // Nature – 2006. – Vol. 439. – P. 168–174.

48. Oehmen C. "ScalaBLAST: A Scalable Implementation of BLAST for High-Performance Data-Intensive Bioinformatics Analysis". / C. Oehmen, J. Nieplocha // IEEE Transactions on Parallel and Distributed Systems. – 2006. Vol. 17. – P. 740 – 749.

49. Database resources of the National Center for Biotechnology Information. / D. L. Wheeler, T. Barrett, D. A Benson [et al.]. // Nucleic Acids Res. – 2005. Vol. 33. P. 39 – 45.

50. Negative regulation of natural killer cell function by EAT-2, a SAP-related adaptor. / R. Roncagalli, J. E. Taylor, S Zhang [et al.]. // Nat. Immunol. – 2005. Vol. 6. – P. 1002–1010.

51. Lu Y. "NSP1 defines a novel family of adaptor proteins linking integrin and tyrosine kinase receptors to the c-Jun N-terminal kinase/stress-activated protein kinase signaling pathway". / Y. Lu, J. Brush, T. A. Stewart // The Journal of Biological Chemistry. – 1999. – Vol. 274. – P. 10047 – 10047.

52. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. / A. Hamosh, A. F. Scott, J. S. Amberger [et al.]. // Nucleic Acids Res. – 2005. – Vol. 33. – P. 514–517.

53. Construction of a nonredundant human SH2 domain database. / H. Huang, Y. Jiao, R. Xu [et al.]. // Genomics Proteomics Bioinformatics. – 2004. Vol. 2. – P. 119 – 122.

54. Initial sequencing and analysis of the human genome. / E. S. Lander, L. M. Linton, B. Birren [et al.]. // Nature. – 2001. – Vol. 409. – P. 860–921.

55. Src Homology 2 Domain Containing Protein 5 (SH2D5) Binds the Breakpoint Cluster Region Protein, BCR, and Regulates Levels of Rac1-GTP / J G. Elizabeth, P. Evangelia, J. Andrew [et al.]. // J Biol Chem. – 2014. – Vol. 19. – P. 35397 – 35408.

56. Analysis of a Shc family adaptor protein, ShcD/Shc4, that associates with muscle-specific kinase." / N. Jones, W. R. Hardy, M. B. Friese [et al.]. // Mol. Cell. Biol. – 2007. – Vol. 27. – P. 4759 – 4773.

57. Identification and analysis of the human and murine putative chromatin structure regulator SUPT6H and Supt6h. / P. W. Chiang, S. Wang, P. Smithivas [et al.]. // Genomics. – 1996. – Vol. 34. – P. 328 – 333.

58. Depetris R. S. Structural basis for inhibition of the insulin receptor by the adaptor protein grb14 / R. S. Depetris, J. Hu, I. Gimpelevich // Mol. Cell. – 2005. – V. 20. – P. 325–333.

59. Eichinger L. "The genome of the social amoeba Dictyostelium discoideum". / L. Eichinger, J. A. Pachebat, G. Glockner // Nature. – 2005. – Vol. 435. – P. 43 – 57.

60. The SH2 Domain–Containing Proteins in 21 Species Establish the Provenance and Scope of Phosphotyrosine Signaling in Eukaryotes / A. L. Bernard, S. Eshana, J. Karl [et al.]. // Sci Signal. – 2011. – Vol. 4. – P. 1 – 39.

61. Defective T cell receptor signaling and CD8+ thymic selection in humans lacking zap-70 kinase. / E. Arpaia, M. Shahar, H. Dadi [et al.]. // Cell. – 1994. – Vol. 76. – P. 947 – 958.

62. Profound block in thymocyte development in mice lacking p56lck. / T. J. Molina, K. Kishihara, D. P. Siderovski [et al.]. // Nature. – 1992. – Vol. 357. – P. 161 – 164.

63. Impaired viability and profound block in thymocyte development in mice lacking the adaptor protein SLP-76. / V. Pivniouk, E. Tsitsikov, P. Swinton [et al.]. // Cell. – 1998. – Vol. 94. – P. 229 – 238.

64. Requirement for the SLP-76 adaptor GADS in T cell development. / J. Yoder, C. Pham, Y. M. Iizuka [et al.]. // Science. – 2001. Vol. 291. – P. 1987–1991.

65. Vav1/2/3-null mice define an essential role for Vav family proteins in lymphocyte development and activation but a differential requirement in MAPK signaling in T and B cells. / K. Fujikawa, A. V. Miletic, F. W. Alt [et al.]. // J. Exp. Med. – 2033. Vol. 198. – P. 1595–1608.

66. Stein P. L. Combined deficiencies of Src, Fyn, and Yes tyrosine kinases in mutant mice. / P. L. Stein, H. Vogel, P. Soriano // Genes Dev. – 1994. Vol. 8. – P. 1999–2007.

67. Compensationbetween Vav-1 and Vav-2 in B cell development and antigen receptor signaling. / K. Tedford, L. Nitschke, I. Girkontaite [et al.]. // Nat. Immunol. – 2001. – Vol. 2. – P. 548–555.

68. Crystal structure of the tyrosine phosphatase SHP-2. / P. Hof, S. Pluskey, S. Dhe-Paganon [et al.]. // Cell. – 1998. – Vol. 92. – P. 441–450.

69. Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. / M. Tartaglia, E. L. Mehler, R. Goldberg [et al.]. // Nat. Genet. – 2001. Vol. 29. – P. 465–468.

70. Activating mutations of the noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia. / M. Bentires-Alj, J. G. Paez, F. S. David [et al.]. // Cancer Res. – 2004. Vol. 64. – P. 8816–8820.

71. The SH2 domain of bcr-Abl is not required to induce a murine myeloproliferative disease; however, SH2 signaling influences disease latency and phenotype. / X. Zhang, R. Wong, S. X. Hao [et al.]. // Blood. – 2001. Vol. 97. – P. 277–287.

72. Structural basis for the autoinhibition of c-Abl tyrosine kinase. / B. Nagar, O. Hantschel, M. A. Young [et al.]. // Cell. – 2003. Vol. 112. – P. 859–871.

73. Somatic mutations lead to an oncogenic deletion of met in lung cancer. / M. Kong-Beltran, S. Seshagiri, J. Zha [et al.]. // Cancer Res. – 2006. Vol. 66. – P. 283–289.

74. Structure of an SH2 domain of the p85 alpha subunit of phosphatidylinositol-3-OH kinase. / G. W. Booker, A. L. Breeze, A. K. Downing [et al.]. // Nature. – 1992. Vol. 358. – P. 684–687.

75. Three-dimensional solution structure of the src homology 2 domain of c-abl. / M. Overduin, C. B. Rios, B. J. Mayer [et al.]. // Cell. – 1992. – Vol. 70. – P. 697–704.

76. Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. / G. Waksman, D. Kominos, S. C. Robertson [et al.]. // Nature. – 1992. Vol. 358. – P. 646–653.

77. Bradshaw J. M. Molecular recognition by SH2 domains. / J. M. Bradshaw, G. Waksman // Adv. Protein Chem. – 2002. – Vol. 61. – P. 161–210.

78. Kuriyan J. Modular peptide recognition domains in eukaryotic signaling. / J. Kuriyan, D. Cowburn // Annu. Rev. Biophys. Biomol. Struct. – 1997. – Vol. 26. – P. 259–288.

79. Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, SYK, and Vav. / Z. Songyang, S. E. Shoelson, J. McGlade [et al.]. // Mol. Cell. Biol. – 1994. – Vol. 14. – P. 2777–2785.

80. Yaffe M. B. Phosphotyrosine-binding domains in signal transduction. / M. B. Yaffe // Nat. Rev. Mol. Cell Biol. – 2002. – Vol. 3. – P. 177–186.

81. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. / J. C. Obenauer, L. C. Cantley, M. B. Yaffe // Nucleic Acids Res. – 2003. Vol. 31. – P. 3635–3641.

82. Phospho. ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. / F. Diella, S. Cameron, C. Gemund [et al.]. // BMC Bioinformatics – 2004. Vol. 5. – P. 1 – 5.

83. Nuclear magnetic resonance structure of an SH2 domain of phospholipase Cgamma 1 complexed with a high affinity binding peptide. / S. M. Pascal, A. U. Singer, G. Gish [et al.]. // Cell. – 1994. Vol. 77. – P. 461–472.

84. Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/ SH2D1A. / S. C. Li, G. Gish, D. Yang [et al.]. // Curr. Biol. – 1999. – Vol. 9. – P. 1355–1362.

85. Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition. / F. Poy, M. B. Yaffe, J. Sayos [et al.]. // Mol. Cell. – 1999. – Vol.4. – P. 555–561.

86. The structure of SOCS3 reveals the basis of the extended SH2 domain function and identifies an unstructured insertion that regulates stability. / J. J. Babon, E. J. McManus, S. Yao [et al.]. // Mol. Cell. – 2006. Vol. 22. – P. 205–216.

87. Structural basis for specificity of Grb2-SH2 revealed by a novel ligand binding mode. / J. Rahuel, B. Gay, D. Erdmann [et al.]. // Nat. Struct. Biol. – 1996. Vol. 3. – P. 586–589.

88. Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. / J. Hu, J. Liu, R. Ghirlando [et al.]. // Mol. Cell. – 2003. – Vol. 12. – P. 1379–1389.

89. John E. L. Energetics of Src Homology Domain Interactions in Receptor Tyrosine Kinase-Mediated Signaling / E. L. John, T. Stefan // Arold Methods in Enzymology. – 2011. – Vol. 488. – P. 147 – 183.

90. Binding of SAP SH2 domain to FynT SH3 domain reveals a novel mechanism of receptor signalling in immune regulation. / S. Latour, R. Roncagalli, R. Chen [et al.]. // Nat Cell Biol. – 2003. – Vol. 5. – P. 149 – 154.

91. SH2D1A and SLAM protein expression in human lymphocytes and derived cell lines. / N. Nagy, C. Cerboni, K. Mattsson [et al.]. // Int J Cancer. – 2000. – Vol. 88. – P. 439 – 447.

92. Regulation of SLAM-mediated signal transduction by SAP, the X-linked lymphoproliferative gene product. / S. Latour, G. Gish, C. D. Helgason [et al.]. // Nat. Immunol. – 2001. – Vol. 2. – P. 681–690.

93. Ottinger E. A. Tandem SH2 domains confer high specificity in tyrosine kinase signaling. / E. A. Ottinger, M. C. Botfield, S. E. Shoelson // J. Biol. Chem. – 1998. Vol. 273. – P. 729–735.

94. The Jak1 SH2 domain does not fulfill a classical SH2 function in Jak/STAT signaling but plays a structural role for receptor interaction and up-regulation of receptor surface expression. / S. Radtke, S. Haan, A. Jorissen [et al.]. // J. Biol. Chem. – 2005. Vol. 280. – P. 25760–25768.

95. Giordanetto F. Prediction of the structure of human Janus kinase 2 (JAK2) comprising JAK homology domains 1 through 7. / F. Giordanetto, R. T. Kroemer // Protein Eng. – 2002. – Vol. 15. – P. 727–737.

96. Rameh L. E. Phosphatidylinositol (3,4,5)P3 interacts with SH2 domains and modulates PI 3-kinase association with tyrosine-phosphorylated

poteins. / L. E. Rameh, C. S. Chen, L. C. Cantley // Cell. – 1995. – Vol. 83. – P. 821–830.

97. The sulfogalactose moiety of sulfoglycosphingolipids serves as a mimic of tyrosine phosphate in many recognition processes. Prediction and demonstration of Src homology 2 domain/sulfogalactose binding. / C. Lingwood, M. Mylvaganam, F. Minhas [et al.]. // J. Biol. Chem. – 2005. – Vol. 280. – P. 12542–12547.

98. The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1. / M. Rozakis-Adcock, R. Fernley, J. Wade [et al.]. // Nature. -1993. – Vol. 363. – P. 83–85.

99. Pawson T. The SH2 domain: a prototype for protein interaction domains. / T. Pawson, G. D. Gish, P. Nash // In Modular Protein Domains. – 2004. Vol. 7. – P. 5–36.

100. Buday L. The Nck family of adapter proteins: regulators of actin cytoskeleton. / L. Buday, L. Wunderlich, P. Tamas // (2002). Cell. Signal. – 2002. – Vol. 14. – P. 723–731.

101. Inducible clustering of membrane-targeted SH3 domains of the adaptor protein Nck triggers localized actin polymerization. / G. M. Rivera, C. A. Briceno, F. Takeshima [et al.]. // Curr. Biol. – 2004. – Vol. 14. – P. 11–22.

102. Identification of an atypical Grb2 carboxyl-terminal SH3 domain binding site in Gab docking proteins reveals Grb2-dependent and –independent recruitment of Gab1 to receptor tyrosine kinases. / L. S. Lock, I. Royal, M. A. Naujokas [et al.]. // J. Biol. Chem. – 2000. Vol. 275. – P. 31536–31545.

103. Takenawa T. Signaling through Grb2/Ash-control of the Ras pathway and cytoskeleton. / T. Takenawa, H. Miki, K. Matuoka // Curr. Top. Microbiol. Immunol. – 1998. Vol. 228. – P. 325–342.

104. Feller S. M. Crk family adaptors-signalling complex formation and biological roles. / S. M. Feller // Oncogene. – 2001. Vol. 20. P. - 6348–6371.

105. Ilangumaran S. Regulation of the immune system by SOCS family adaptor proteins. / S. Ilangumaran, S. Ramanathan, R. Rottapel // Semin. Immunol. – 2004. – Vol. 16. – P. 351–365.

106. Haglund K. Ubiquitylation and cell signaling. / K. Haglund, I. Dikic // EMBO J. – 2005. – Vol. 24. – P. 3353–3359.

107. Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. / P. L. Howard, M. C. Chia, S. Del Rizzo [et al.]. // Proc. Natl. Acad. Sci. USA. – 2003. – Vol. 100. – P. 11267–11272.

108. Modeling the evolution of protein domain architectures using maximum parsimony. / J. H. Fong, L. Y. Geer, A. R. Panchenko [et al.]. // J. Mol. Biol. – 2007. – Vol. 366. – P. 307–315.

109. Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. / S. D'Aniello, M. Irimia, I. Maeso [et al.]. // Mol. Biol. Evol. – 2008. – Vol. 25. – P. 1841–1854.

110. Arrangements in the modular evolution of proteins. / A. D. Moore, A. K. Bjorklund, D. Ekman [et al.]. // Trends Biochem. Sci. – 2008. – Vol. 33. – P. 444–451.

111. Liu M. Protein domains correlate strongly with exons in multiple eukaryotic genomes—Evidence of exon shuffling? / M. Liu, A. Grigoriev // Trends Genet. – 2004. – Vol. 20. – P. 399–403.

112. Serfas M. S. Brk, Srm, Frk, and Src42A form a distinct family of intracellular Src-like tyrosine kinases. / M. S. Serfas, A. L. Tyner // Oncol. Res. – 2003. – Vol. 13. – P. 409–419.

113. Needleman S. B. A general method applicable to the search for similarities in the amino acid sequence of two proteins / S. B. Needleman, C. D. Wunsch // J. Mol.Biol. — 1970. — V. 4. — P. 443–453.

114. Smith T. F. Identification of Common Molecular Subsequences / T. F. Smith, M. S. Waterman // Ibid. —1981. — V. 147. — P. 195–197.

115. Mount D. M. Dot Matrix Pairwise Sequence Comparison / D. M. Mount // Cold Spring Harb. Protoc. —2007.

116. IAN H. Dynamic Programming Alignment Accuracy / H. IAN, D. RICHARD / Journal of Computational Biology. – 2009. – Vol. 5. – P. 493-504.

117. Volynets. G. P. Methods of structural bioinformatics / G. P. Volynets, V. G. Bdzhola, S. M. Yarmoluk // Institute of Molecular Biology and Genetics of the National Academy of Sciences of Ukraine. – 2010. – Vol. 3. – P. 9 – 19.

118. Mount D. M. Sequence and Genome Analysis / D. M. Mount // Bioinformatics. — 2004. – Vol. 11. – P. 495 – 508.

119. Julie D. T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice / D. T. Julie, G. H. Desmond, J. G. Toby // Oxford JournalsScience & Mathematics Nucleic Acids. – 1994. – Vol. 22. – P. 4673-4680.

120. Comprehensive study on iterative algorithms of multiple sequence alignment / M. Hirosawa, Y. Totoki, M. Hoshida [et al.]. // Comput. Appl. Biosci. — 1995. — V. 11. — P.13–18.

121. Marshall T. Efficient exact motif discovery / T. Marshall, S. Rahmann // Bioinformatics. — 2009. — V. 25. — P. 356–364.

122. Bradley R. K. Fast statistical alignment / R. K. Bradley, A. Roberts, M. Smoot // PloS Comput. Biol. — 2009. — V. 5. — P. 45 – 50.

123. Schwartz A. S. Multiple alignment by sequence annealing / A. S. Schwartz, L. Pachter // Bioinformatics. — 2007. — V. 23. — P. 24–29.

124. Goldsmith-Fischman S. Structural genomics: Computational methods for structure analysis / S. Goldsmith-Fischman, B. Honig // Protein Sci. — 2003. — V. 12. — P. 1813–1821.

125. fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3 / S. Dietmann, J. Park, C. Notredame [et al.]. // Nucl. Acids Res. — 2001. — V. 29. — P. 55–57.

126. Holm L. "Mapping the protein universe" / L. Holm, C. Sander // Science. – 1996. – Vol. 273. – P. 595–603.

127. Taylor W. R. Multiple protein structure alignment / W. R. Taylor, T. P. Flores, C. A. Orengo // Protein Sci. — 1994. — V. 3. — P. 1858–1870.

128. Orengo C. A. CATH — a hierarchic classification of protein domain structures / C. A. Orengo, A. D. Michie, Jones S. // Structure. — 1997. — V. 5. — P. 1093–1108.

129. Orengo C. A. "SSAP: Sequential structure alignment program for protein structure comparison". / C. A. Orengo, W. R. Taylor // Methods in enzymology. – 1996. – Vol. 266. – P. 617–635.

130. Shindyalov I. N. "Protein structure alignment by incremental combinatorial extension (CE) of the optimal path" / I. N. Shindyalov, P. E. Bourne // Protein Eng. – 1998. – Vol. 11. – P. 739–747.

131. Yang A. An integrated approach to the analysis and modeling of proteinsequences and structures. I. Protein strutural alignment and a quantitative measure for protein structural distance / A. Yang, B. Honig // J. Mol. Biol. — 2000. — V. 301. — P. 691–711.

132. Leibowitz N. MUSTA — a general, efficient, automated method for multiple structure alignment and detection of common motifs: application to proteins / N. Leibowitz, R. Nussinov, H. J. Wolfson // J. Comp. Biol. — 2001. — V. 8. — P. 93–121.

133. Shatsky M. MultiProt — a multiple protein structural alignment algorithm / M. Shatsky, R. Nussinov, H. J. Wolfson // Work. Algor. Bioinform. — 2002. — V. 2452. — P. 235–250.

134. Multiple structural alignment by secondary structures: algorithm and applications / O. Dror, H. Benyamini, R. Nussinov [et al.]. // Protein Sci. — 2003. — V. 12, N. 11. — P. 2492–2507.

135. Nagano N. One fold with many functions — the evolutionary relationships between TIM barrel families based on their sequences, structures and functions / N. Nagano, C. A. Orengo, J. M. Thornton // J. Mol. Biol. — 2002. — V. 321. — P. 741–765.

136. A consensus-binding structure for adenine at the atomic level permits searching for the ligand site in a wide spectrum of adenine-containing complexes / Y. Y. Kuttner, V. Sobolev, A. Raskind [et al.]. // Proteins. — 2003. — V. 52. — P. 400–411.

137. Denessiouk K. A. Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins / K. A. Denessiouk, V. Rantanen, M. S. Johnson // Ibid. — 2001. — V. 44. — P. 282–291.

138. Burger S. K. Quantum mechanics/molecular mechanics strategies for docking pose refinement: distinguishing between binders and decoys in cytochrome C peroxidase. / S. K. Burger, D. C. Thompson, P. W. Ayers // J Chem Inf Model. – 2011. – Vol. 24. – P. 93 – 101.

139. Cornell W. D. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules / W. D. Cornell, P. Cieplak, C. I. Bayly // J. Am. Chem. Soc. — 1996. — V. 117. — P. 5179–5197.

140. Lamoureux G. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm / G. Lamoureux, B. Roux // J. Chem. Phys. — 2003. — V. 119. — P. 3025–3039.

141. Jynsdyttir S. У. The consistent force field. Part 6: an optimized set of potential energy functions for primary amines / S. У. Jynsdyttir, K. Rasmussen // New J. Chem. — 2000. — V. 24. — P. 243–247.

142. Sun H. COMPASS: an ab initio force-field optimized for condensed-phase applications — overview with details on alkane and benzene compounds / H. Sun // J. Phys. Chem. B. — 1998. — V. 102, N 38. — P. 7338–7364.

143. Shi S. An extensible and systematic force field, ESFF, for molecular modeling of organic, inorganic, and organometallic systems / S. Shi, L. Yan, Y. Yang // J. Comput. Chem. — 2003. — V. 24. — P. 1059–1076.

144. Rappe A. K. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations / A. K. Rappe, C. J. Casewit, K. S. Colwell // J. Am. Chem. Soc. — 1992. — V. 114. — P. 10024–10035.

145. Metropolis N. The Monte Carlo Method / N. Metropolis, S. Ulam // J. Am. Stat. Assoc. — 1949. — V. 44. — P. 335–341.

146. Fogel D. B. Evolutionary Computation: Towards a New Philosophy of Machine Intelligence. — New York / D. B. Fogel // IEEE Press – 2000.

147. Taylor R. D. A review of protein small molecule docking methods / R. D. Taylor, P. J. Jewsbury, J. W. Essex // J. Comput. Aid. Mol. Des. — 2004. — V. 16. — P. 151–166.

148. Gabb H. A. Modeling Protein Docking using shape complementarity, electrostatics and biochemical information / H. A. Gabb, R. M. Jackson, M. J. E. Sternberg // J. Mol. Biol. — 1997. — V. 272. — P. 106–120.

149. Baxter C. A. Flexible docking using Tabu search and an empirical estimate of binding affinity / C. A. Baxter, C. W. Murray, D. E. Clark // Proteins. — 1998. — V. 33. — P. 367–382.

150. Kyiv National Taras Shevchenko University High Performance Computing Cluster. / Y. V. Boyko, O. O. Vystoropsky, T. V. Nychyporuk [et al.]. // Third International Young Scientists Conference on Applied Physics. – 2003. – P. 189-181.

151. Drug- and leadlikeness, target class, and molecular diversity analysis of 7.9 million commercially available organic compounds provided by 29 suppliers. / A. Chuprina, O. Lukin, R. Demoiseaux [et al.]. // J. Chem. Inf. Model. – 2010. – Vol 50. – P. 470-479.

152. Daniel R. Python Scripts as a Replacement for Bash Utility Scripts / R. Daniel // Linux Journal. – 2013.

153. Naugler D. "C# 2.0 for C++ and Java programmer: conference workshop" / D. Naugler // Journal of Computing Sciences in Colleges. - 2007. - Vol. 22.

154. The protein databank. / H. M. Berman, J. Westbrook, Z. Feng [et al.]. // Nucleic Acids Res. – 2000. – Vol. 28. – P. 235-242.

155. The pfam protein families database. / M. Punta, P. C. Coggill, R. Y. Eberhardt [et al.]. // Nucleic acids research. – 2012. – Vol. 40. – P. 290–301.

156. Proline Isomerization Preorganizes the Itk SH2 Domain for Binding to the Itk SH3 Domain / A. J. Severin, R. E. Joseph, S. Boyken [et al.]. // J.Mol.Biol. – 2009. – Vol. 387. – P. 726 – 743.

157. Thompson J. D. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. / J. D. Thompson, T. J. Gibson, F. Plewniak // Nucleic Acids Res. – 1997. – Vol. 25. – P. 4876-4882.

158. Multiple sequence alignment with the Clustal series of programs / C. Ramu, S. Hideaki, K. Tadashi [et al.]. // Nucleic Acids Research. – 2003. – Vol. 31. – P. 3497–3500.

159. Hogeweg P. "The alignment of sets of sequences and the construction of phyletic trees: an integrated method" / P. Hogeweg, B. Hesper // J Mol Evol. – 1984. – Vol. 20. – P. 175 – 186.

160. Chang J. M. "TCS: A New Multiple Sequence Alignment Reliability Measure to Estimate Alignment Accuracy and Improve Phylogenetic Tree Reconstruction." / J. M. Chang, T. P. Di, C. Notredame // Molecular Biology and Evolution. – 2014. – Vol. 31. – P. 1625–1637.

161. Pairwise Alignment of Protein Interaction Networks / K. Mehmet, K. Yohan, T. Umut [et al.]. // Journal of Computational Biology. – 2006. – Vol. 13. – P. 182-199.

162. Saitou N. "The neighbor-joining method: a new method for reconstructing phylogenetic trees." / N. Saitou, M. Nei, // Molecular Biology and Evolution. – 1987. – Vol. 4 – P. 406-425.

163. Erick F. P. UCSF Chimera A visualization system for exploratory research and analysis. / F. P. Erick, D. G. Thomas, C. H. Conrad // J. Comput. Chem. – 2004. – Vol. 25. – P. 1605-1612.

164. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins / C. A. Kocsh, D. Anderson, M. F. Moran [et al.]. // Science. – 1991. – V. 252. – P. 668-674.

165. Corbeil C. R. Variability in docking success rates due to dataset preparation / C. R. Corbeil, C. I. Williams, P. Labute // J. Comp.-Aided Mol. – 2012. – V. 26. – P. 775-786.

166. Vattani. A. "k-means requires exponentially many iterations even in the plane" / A. Vattani. // Discrete and Computational Geometry. - 2011. - Vol. 45. - P. 596–616.

167. Osada R. Shape distributions. / R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin // ACM Trans. Graph. – 2002. – Vol. 21. – P. 807–832.

168. Pirovano W. Homology-extended alignment strategy. / W. Pirovano, K. A. Feenstra, J. Heringa // Bioinformatics. – 2009. – Vol. 24. – P. 492-497.

169. Asieh S. Comparison and Evaluation of Multiple Sequence Alignment Tools in Bininformatics. / S. Asieh, B. A. Rodziah, T. A. Khairina // Int. J. Computer Sci. Network Security. – 2009. – Vol. 9. – P. 51-56.

170. Shannon C. E. "A Mathematical Theory of Communication". / C. E. Shannon // Bell System Technical Journal. – 1948. – Vol. 27. – P. 379–423.

171. Arlin S. Evolutionary conservation of RecA genes in relation to protein structure and function. / S. Arlin, L. Brocchieri // J. Bacteriol. – 1996, - Vol. 178. – P. 1881-1894.

172. Livingstone D. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation / D. Livingstone, J. B. Geoffrey // CABIOS. – 1993. – Vol. 9. – P. 745-756.

173. Goddard T. D. Software Extensions to UCSF Chimera for Interactive Visualization of Large Molecular Assemblies. / T. D. Goddard, C. C. Huang, T. E. Ferrin // Structure. – 2005. – Vol. 13. – P. 473-482.

174. Guntur F. A potent anti-dengue human antibody preferentially recognizes the conformation of E protein monomers assembled on the virus surface. / F. Guntur, L. Joanne, A. Scott // EMBO Mol. Med. – 2014. – Vol. 6. – P. 358-351.

175. Lee B. The Interpretation of Protein Structures: Estimation of Static Accessibility. / B. Lee, M. Richards // J. Mol. Biol. – 1971. – Vol. 55. – P. 379-400.

176. Adamczak R. "Combining prediction of secondary structure and solvent accessibility in proteins". / R. Adamczak, A. Porollo, J. Meller // Proteins. - 2005. - Vol. 59. - P. 467–75.

177. How similar are similarity searching methods? a principal component analysis of molecular descriptor space. / A. Bender, J. L. Jenkins, J. Scheiber [et al.]. // Journal of Chemical Information and Modeling. – 2009. – Vol. 49. – P. 108 – 119.

178. A conformational search method for efficient generation of bioactive conformers. / K. S. Watts, P. Dalal, R. B. Murphy [et al.]. // Journal of Chemical Information and Modeling. – 2010. – Vol. 50. – P. 534-546.

179. Thomas A. H. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94 / A. H. Thomas // J. Comp. Chem. – 1996. – V. 17. – P. 490-519.

180. Mcmartin C. QXP: Powerful, rapid computer algorithms for structure-based drug design / C. Mcmartin, R. Bohacek // J. Comput.-Aid. Mol. Design. – 1997. – V. 11. – P. 333–344.

181. Chang D. T. a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm / D. T. Chang, Y. J. Oyang, J. H. Lin // Nucleic acids. – 2005. – V. 33. – P. 233-238.

182. Gerhard W., LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters / W. Gerhard, L. Thierry // J. Chem. Inf. Model. – 2005. – Vol. 45. – P. 160–169.

183. Generative Topographic Mapping-Based Classification Models and Their Applicability Domain: Application to the Biopharmaceutics Drug Disposition Classification System (BDDCS) / A. G. Helena, M. Gilles, H. Dragos [et al.]. // chemical information and modelling. – 2013. – V. 53. – P. 3318-3325.

184. Generative Topographic Mapping (GTM): Universal Tool for Data Visualization, Structure-Activity Modeling and Dataset Comparison / N. Kireeva, I. I. Baskin, H. A. Gaspar [et al.]. // Molecular Informatics. – 2012. – V. 31. – P. 301-312.

185. Thomas A. H. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94 / A. H. Thomas // J. Comp. Chem. – 1996. – V. 17. – P. 490-519.

186. Lorber D. M. Flexible ligand docking using conformational ensembles. / D. M. Lorber, B. K. Shoichet // Protein Sci. - 1998. - V. 7. - P. 938 - 950.

187. A critical assessment of docking programs and scoring functions / G. L. Warren, C. W. Andrews, A. M. Capelli [et al.]. // J. Med. Chem. – 2006. – V. 49. – P. 5912-5931.

188. Scheraga A. Flexible docking simulations: Scaled collective variable Monte Carlo minimization approach using Bezier splines, and comparison with a standard Monte Carlo algorithm / A. Scheraga // Journal of Computational Chemistry. – 1999. – Vol. 20. – P. 244–252.

189. Geometric filters for protein–ligand complexes based on phenomenological molecular models / O. O. Sudakov, O. M. Balinskyi, M. O. Platonov [et al.]. // Biopolym. Cell. – 2013. – V. 29. – P. 395–400.

190. Design, synthesis, and cocrystal structure of a nonpeptide Src SH2 domain ligand. / M. S. Plummer, D. R. Holland, A. Shahripour [et al.]. // J.Med.Chem. – 1997. – Vol. 40. – P. 3719-3725.

191. Structure-Based Design of a Novel Series of Nonpeptide Ligands that Bind to the Pp60Src Sh2 Domain / D. R. Holland, E. A. Lunney, M. S. Plummer [et al.]. // J.Am.Chem.Soc. – 1997. – Vol. 119. – P. 12471–12476.

192. Primary Citation Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of (pp60)Src are identical to those for high affinity binding of full length inhibitors. / G. Lange, D. Lesuisse, P. Deprez [et al.] // J.Med.Chem. – 2003. – Vol.46. – P. 5184-5195.

193. A novel phosphotyrosine mimetic 4'-carboxymethyloxy-3'-phosphonophenylalanine (Cpp): exploitation in the design of nonpeptide inhibitors of pp60(Src) SH2 domain. // N. Kawahata, M. G. Yang, G. P. Luke [et al.]. // Bioorg.Med.Chem.Lett. – 2001. – Vol. 11. – P. 2319-2323.

194. Structure-based design of an osteoclast-selective, nonpeptide src homology 2 inhibitor with in vivo antiresorptive activity. / W. Shakespeare, M. Yang, R. Bohacek [et al.]. // Proc.Natl.Acad.Sci.Usa. – 2000. – Vol. 97. – P. 9373-9378.

195. Peptide ligands of pp60(c-src) SH2 domains: a thermodynamic and structural study. / P. S. Charifson, L. M. Shewchuk, W. Rocque [et al.]. // Biochemistry. – 1997. – Vol. 36. – P. 6283-6293.

196. Krovat E. M. Pharmacophore identification, in silico screening, and virtual library design for inhibitors of the human factor Xa. / E. M. Krovat, K. H. Fruhwirth, T. Langer // J Chem Inf Model. – 2005. – Vol. 45. – P. 146–59.

197. Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. / M. Rella, C. A. Rushworth, J. Guy [et al.]. // J Chem Inf Model. – 2000. – Vol. 46. – P. 708–716.

198. Identification of novel cyclooxygenase-2 selective inhibitors using pharmacophore models. / A. Palomer, F. Cabre, J. Pascual [et al.]. // J Med Chem. – 2002. – Vol. 45. – P. 1402–1411.

199. Pharmacophores incorporating numerous excluded volumes defined by X-ray crystallographic structure in three-dimensional database searching: application to the thyroid hormone receptor. / P. A. Greenidge, B. Carlsson, L. G. Bladh [et al.]. // J Med Chem. – 1998. – Vol. 41. – P. 2503–2512.

200. Irwin J. J. ZINC a free database of commercially available compounds for virtual screening / J. J. Irwin, B. K. Shoichet // J. Chem. Inf. Model. – 2005. – V. 45 – P. 177 – 182.

201. Braga C. Assessing the Performance of 3D Pharmacophore Models in Virtual Screening: How Good are They? Rodolpho / C. Braga, H. A. Carolina // Curr Top Med Chem. – 2013. – V. 13. – P. 1127-1238.

202. Virtual screening using protein-ligand docking: avoiding artificial enrichment / M. L. Verdonk, V. Berdini, M. J. Hartshorn [et al.]. // Chem. Inf. Model. – 2004. – V. 44. – P. 793-806.

203. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? / J. Kirchmair, P. Markt, S. Distinto [et al.]. // Comput.-Aided Mol. – 2008. – V.22. – P. 213-228.

204. Sonego P. ROC analysis: applications to the classification of biological sequences and 3D structures / P. Sonego, A. Kocsor, S. Pongor // Brief. Bioinformatics – 2008. – V. 9. – P. 198-209.

205. pROC: an open-source package for R and S+ to analyze and compare ROC curves / X. Robin, N. Turck, A. Hainard [et al.]. // BMC Bioinformatics – 2011. – V. 12. – P. 1471-2105.

206. Virtual screening workflow development guided by the “receiveroperating characteristic” curve approach. Application to highthroughput docking on metabotropic glutamate receptor subtype 4. / N. Triballeau, F. Acher, I. Brabet [et al.]. // J. Med. Chem. – 2005. – Vol. 48. – P. 2534-2347.

207. An effective docking strategy for virtual screening based on multi-objective optimization algorithm / L. Honglin, Z. Hailei, Z. Mingyue [et al.]. // BMC Bioinformatics – 2009. – Vol. 10. – P. - 2105-2110.

208. Ligand and structure based pharmacophore modeling to facilitate novel histone deacetylase 8 inhibitor design / S. Sakkiah, S. Thangapandian, Y. J. John [et al.]. // Eur. J. Med. Chem. – 2010. – V. 45. – P. 4409-4417.

209. Imaging Systems for Westerns: Chemiluminescence vs. Infrared Detection / T. M. Suresh, T. M. Suresh, P. P. Eric [et al.]. // Protein Blotting and Detection. - 2009. - Vol. 536, - P. 499-513.