

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА  
ІНСТИТУТ ВИСОКИХ ТЕХНОЛОГІЙ**

Завідувач кафедри супрамолекулярної хімії  
д.х.н., професор Ігор Володимирович  
Комаров  
Протокол №\_\_\_\_ засідання кафедри  
Від «\_\_\_\_»\_\_\_\_\_2021 року

**ДОСЛІДЖЕННЯ СИНТЕТИЧНОЇ ДОСТУПНОСТІ ДЕ  
НОВО ГЕНЕРОВАНИХ СПОЛУК НА ПРИКЛАДІ  
ІНГІБІТОРІВ MARK4**

Випускна кваліфікаційна робота магістра  
студента напряму підготовки  
хемоінформатика  
102 Хімія  
**Кощесенко Вікторія Михайлівна**

Науковий керівник:  
д.х.н., професор  
**Шиванюк Олександр Миколайович**

Оцінка захисту роботи

---

**Київ 2021 р.**

## Зміст

<b>Вступ</b> .....	3
<b>Огляд літератури</b> .....	4
<b>Практична частина</b> .....	20
<b>Висновок</b> .....	32
<b>Використана література:</b> .....	33

## Вступ

Хімічний простір надзвичайно великий - його розміри оцінюються в  $\sim 10^{33}$  сполуках [1]. Найближчим часом неможливо буде перелічити цей простір або виконати будь-який вичерпний пошук. Тому, методи та стратегії вивчення цього простору привертають науковий інтерес. Однією з популярних стратегій є *de novo* дизайн – моделювання нових хімічних структур з перспективними прогнозованими властивостями [2, 3]. Вже декілька десятиліть *de novo* дизайн невеликих молекул використовується для створення нових лікарських засобів. Останні публікації, що описують синтетичну доступність препаратів, розроблених на основі *de novo* дизайну, вказують на те, що все ще є потреба у знаходженні нових методів. Існують дві основні стратегії генерації структур: (i) ітераційна генерація структур для прогнозування моделей і (ii) генерація структур, що мають бажаний набір властивостей безпосередньо за допомогою моделей машинного навчання (ML) (наприклад, зворотний QSAR або генеративні нейронні мережі).

Загальний робочий час включає: (i) генерацію або вибір вихідних структур, (ii) оцінку генерованих структур за моделлю (моделями) (QSAR, фармакофори тощо), (iii) вибір найбільш перспективних кандидатів, генерація нових структур на основі вибраних та повернення до етапу (ii). Цю процедуру повторюють до утворення сполук з бажаними властивостями. У цьому випадку етапи генерації структури та оцінки властивостей розділяються. Отже, можна використовувати будь-яку комбінацію підходів до створення структури та у моделях *in-silico* для прогнозування властивостей сполук.

Ключові слова: QSAR, фармакофори, *in-silico*, хімічний простір.

## Огляд літератури

Ми можемо розділити традиційні підходи на три групи: генерація структур на основі атомів, на основі фрагментів та реакцій, кожен має свої переваги та проблеми (табл. 1).

	На основі атомів	На основі фраментів	На основі реакцій
Новизна	+++	++	+
Вичерпність дослідження хімічного простору	+++	++	+
Різноманітність	+++	++	+
Валідність	-	-/+	+++
Хімічна доступність	-	-/+	+++

Підходи на основі атомів представляють методи “ab initio” серед генераторів структур і використовують прості правила як "додати / видалити / замінити атом / зв'язок", щоб модифікувати структури та генерують нові [4]. Теоретично це повинна бути можливість генерувати всі можливі структури з використанням цих правил, що збільшує вірогідність новизни та різноманітності перелічених структур. Тому атомні підходи більше застосовуються для систематичного дослідження хімічного простору. Однак головне питання атомних підходів - це синтетична доступність, яка не контролюється протягом генерації та може призвести до менш синтетично доступних структур.

Підходи, засновані на реакціях створюють сполуки з більшою новизною та різноманітністю лише за декілька етапів генерації порівняно з підходами на основі атомів, для досягнення яких може знадобитися набагато більше кроків. Підходи, засновані на реакціях, вносять великі зміни у структуру під час утворення сполук і, отже, є більш придатними для грубої розвідки хімічного простору. Синтетична доступність отриманих сполук та наявний синтетичний шлях - головні переваги підходу, заснованого на реакціях. Застосовність цього виду підходів було продемонстровано в кількох дослідженнях [5-8]. Тим не менше, обмежена кількість правил (головним чином розглядаються лише реакції сполучення) та обмежені розміри бібліотек реагентів можуть обмежувати ці алгоритми у вивченні більшого хімічного простору (отже, втрачається новизна та різноманітність утворених сполук).

Підходи, засновані на фрагментах, є середнім між атомними та реакційними [9-11]. Набір початкових фрагментів безпосередньо визначає новизну та різноманітність згенерованих сполук. Тож ми очікуємо, що підхід на основі фрагментів перевершить підходи, засновані на реакції (з точки зору різноманітності та новизни), оскільки здається простішим зібрати різноманітну бібліотеку фрагментів, ніж різноманітну бібліотеку реагентів. Однак доступний хімічний простір менший ніж для атомних підходів. Здається, це простіше для контролю хімічної достовірності перелічених структур у випадку з фрагментами, але все ще існує проблема синтетичної доступності, оскільки зв'язування синтетично доступних фрагментів може призвести до синтетично недоступних молекул. Тому, при зв'язуванні фрагментів слід враховувати хімічний контекст зв'язаних фрагментів. Нещодавно Лю та ін. опублікували підхід, де вони враховували типи атомів, до яких можуть прикріпитися фрагменти [12]. Стратегія призвела до створення хімічно допустимих структур, але автори не вивчали синтетичну доступність згенерованих структур.

Інтерес до другої стратегії відновився нещодавно завдяки досягненню глибокого навчання та генеративних моделей [13–19]. Результуючі структури можуть створюватися без нагляду або контролю. При неконтрольованих підходах моделі ML (як правило, рекуррентні нейронні мережі) навчаються на структурах відомих сполук (зазвичай представлені SMILES) та стохастично вибирають вихідні структури. У контрольованих підходах модель оптимізована для пошуку комбінації дескрипторів, які відповідають активним сполукам, а потім структури сполук відновлюються з цього набору дескрипторів (так зване завдання "зворотного QSAR"). Останні досягнення глибокого навчання дозволили сформувавши репрезентацію вхідних сполук для знаходження оптимального набору змінних, пов'язаних з бажаними властивостями і вибірки структур із цього підпростору змінних [20]. Тим не менше, не існує жодних гарантій того, що знайдені оптимальні комбінації змінних відповідають дійсним хімічним структурам. Відсоток дійсних структур, створених моделями глибокого навчання, може сильно варіюватися залежно від моделі глибокого навчання: від майже 100 до 4% [21]. Більше того, розроблені в даний час підходи не контролюють синтетичну доцільність генерованих структур, яка зазвичай оцінюється після утворення сполук [20]. Незважаючи на недавні успіхи в генеративних моделях глибоких нейронних мереж, підходи, засновані на фрагментах, здаються привабливою альтернативою, оскільки вони забезпечують високу ефективність хімічного простору і можуть поєднуватися з будь-яким підходом до моделювання. Вони можуть також вирішити проблему синтетичної доступності, але це досі не досліджувалося.

CReM забезпечує усі основні функції (MUTATE, GROW та LINK) для маніпуляцій з фрагментами та легко інтегрується зі іншими програмними забезпеченнями. Заснований підхід ґрунтується на визначенні взаємозамінних фрагментів з бази даних відомих сполук для виконання правильної модифікації

початкових структур, що дозволяє генерувати хімічно правильні сполуки і контролювати їх синтетичну доступність.

Використовуючи невеликі фрагменти – дослідники з Astex характеризують їх як молекули, що відповідають правилу трьох ) [22,23] – можна подолати ці обмеження, адже малі фрагменти мають більш високі показники ефективності по відношенню до білкових мішеней, а отже є більш перспективними для драг дизайну. [24-27] До того ж фрагментні бібліотеки зазвичай дешевші за HTS. [28] De novo дизайн у поєднанні з технологіями Fragment based drug discovery пропонують кращу функціональність пошуку.

Правильний скринінг фрагментів та їх перетворення на сполуки-ліди, що специфічні для однієї мішені, можуть більш перспективними, ніж скринінг величезної кількості лігандів. Крім того, низькомолекулярні фрагменти показують кращу ефективність ліганду, а також їх легше оптимізувати. [30,33] Часто використовуваною стратегією є вирощування фрагментів. Вирощування фрагментів – це підхід de novo де якірний фрагмент служить відправною точкою для вирощування. [30,31,34]

Синтетична доступність є важливою вимогою в усіх de novo drug design працях починаючи з перших статей і до сьогодні. [34-39,29-32] Було розроблено багато інструментів, які не враховували синтетичну доступність згенерованих молекул, а зосереджувалися лише на різноманітності та комплементарності до сайту зв'язування. [31,35,40-44] Приклади інструментів de novo проектування, що займаються синтетичною доступністю, є Drug Guru [46], DOGS [45] та TOPAS [47]. Всі ці інструменти є ліганд-орієнтованими і не враховують обмежень первинного фрагменту. Їх можна поєднувати з інструментами для докінгу (наприклад, AutoDock [48], FlexX [49], Gold [50], GLIDE [51,52]),

інтегрований підхід дозволить ефективніше охопити відповідний простір пошуку.

Інструменти, що враховують синтетичну доступність можна розділити на дві групи. Інструменти першої групи виконують вирощування за допомогою заміни атомів водню на дрібні фрагменти. (39,52-55) Для прогнозування синтетичної доступності отриманих сполук використовуються різні стратегії. LigBuilder 2, наприклад, проводить ретро-синтетичний аналіз на основі комерційних білдинг блоків [54]. Інструменти другої групи забезпечують синтетичну доступність з самого початку, використовуючи правила хімічної реакції. [38,34,59,60] Використовування заздалегідь визначених правил, подібних тим, які зібрані Гартенфеллером [56] або клік-хімія [57] зазвичай призводять до утворення сполук з високою синтетичною доступністю. [58]

Хоча було запропоновано багато алгоритмів, важко знайти статистично обґрунтоване підтвердження їх використання, на прикладі пари таргетів як доказ концепції. У більшості випадків кілька вирощених молекул виявляють мілімолярну афінність або принаймні досягається поліпшення афінності до зв'язування. Хоча експериментальна афінність деяких запропонованих молекул демонструє користь цих програм у *de novo drug design*, надійні показники ефективності відсутні.

Незважаючи на деякі успіхи у застосуванні, методи *de novo drug design* широко не застосовуються медичними хіміками. Однією з головних проблем є відсутність широкого набору даних про ретроспективну перевірку, що призводить до складності оцінки області застосування кожного алгоритму. Ще один важливий факт - вже згадана відсутність синтетичної доцільності передбачень.

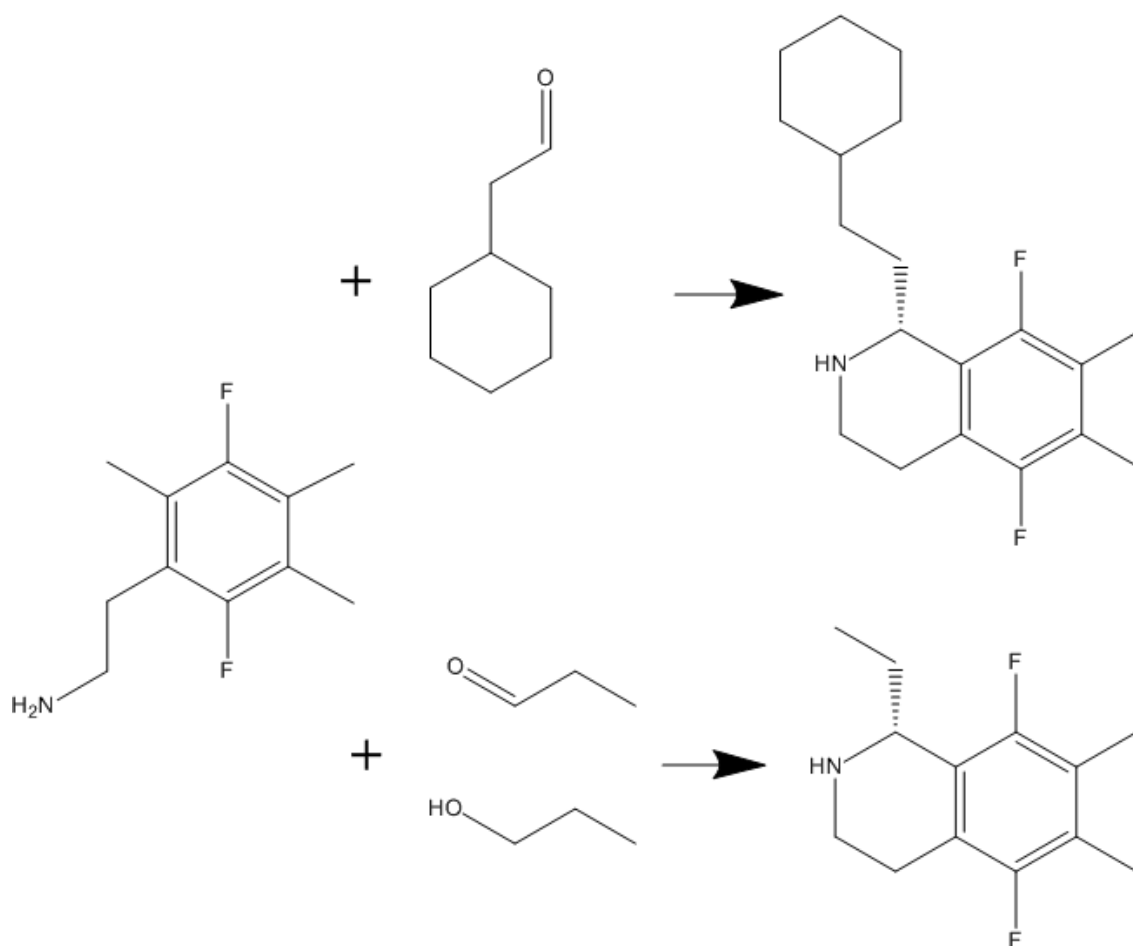


NAOMInext – це новий, ефективний метод для вирощування фрагментів у сайті зв'язування білку. Починаючи з якірного фрагмента, вмикається автоматичний режим вирощування сполук *de novo* в межах даного місця зв'язування за допомогою набору опублікованих реакції органічного синтезу [56], що забезпечують синтетичну доцільність. NAOMInext конструює сполуки-кандидати в межах місця зв'язування цільового білка.

Якірний фрагмент (кристалізований або докований) служить відправною точкою для дослідження доступного хімічного простору. Якірний фрагмент у своїй зв'язаній конформації вже належним чином прилягає до місця зв'язування, конформаційний пошук обмежується прикріпленим блоком.

Отже, у NAOMInext є: моделювання реакцій, конформаційна вибірка та оцінка. NAOMInext обробляє схеми реакцій у Daylight' SMIRKS та Reaction SMARTS форматі [62,63]. Молекулярні структури, описані в SMARTS, використовуються для пошуку певного підграфу у молекулі (графі). Більше того, SMARTS дозволяють детально описувати атоми, що оточують, полегшуючи включення груп, що є донорами або акцепторами електронів. Реакції SMARTS потребували деяких модифікацій, щоб правильно виконувати задані реакції та мати можливість перевірки результатів на основі реакцій продукту SMARTS. Необхідні типи зв'язків для нових сформованих зв'язків між атомами, що беруть участь, походять від продуктів SMARTS з використанням наданих ярликів і ковалентні зв'язки утворюються між відповідні реагентами. Довжина та кути зв'язку адаптуються відповідно з використанням вбудованих значень бібліотеки програмного забезпечення NAOMI [64,65], забезпечуючи валідні 3D координати. Stereo дескриптори перераховуються враховуючи геометричні зміни модифікованих атомів. Нові утворені зв'язки можуть утворити нові стереоцентри, також можуть виникати регіоізмери через симетричне визначення підструктури SMARTS. NAOMInext

обробляє всі стерео- та регіоізомери, що зустрічаються та перелічує можливі результати(РІС.1). Навпаки, проста інверсія потенційних стереоцентрів під час реакції, наприклад, за допомогою реакції Міцунобу, неможлива. У цьому випадку, NAOMInext однаково перераховує можливі стереоцентри, що, на жаль, призводить до одного неіснуючого продукту реакції.



РІС.1

NAOMInext використовує комбінацію ітеративного Breadth-First-Search (BFS) та алгоритми рекурсивного Depth-First-Search (DFS) і включає декілька обмежень для зменшення комбінаторної складності. Вже було показано, що підхід поєднання алгоритмів DFS та BFS є сприятливим для ефективного конформаційного пошуку. [59,66]

Для найкращих  $n$  набраних стартових розміщень (за замовчуванням: 50 та настроювані користувачем) дійсні розміщення для білдинг блоку перераховуються на сайті зв'язування. Збільшення використовуваної кількості стартових розміщень дозволяє отримати кращу вибірку конформаційного простору і в основному кращі результати за рахунок трохи збільшеного часу обчислень. Процедура вирощування фрагментів обробляє молекулу в цілому, таким чином уникаючи спотвореної геометрії молекули. Ефективність виконання під час відбору проб досягається за допомогою динамічних адаптацій під час процедури відбору проб для кожної молекули та конкретної мішені. Розміщення будуються відповідно до стандартної стратегії поступової побудови. [49,67]

Алгоритм DFS є основною рушійною силою в реалізованій конформаційній стратегії вибірки і призводить до швидких, але переважно неадекватних результатів. Це точка, де основні обмеження цілі, а саме стеричне вписування в активний сайт, відіграють вирішальну роль. Ці обмеження використовуються для керування, прискорення та спрощення вибірки. Розміщення, що стикаються, відкидаються на ранній стадії. Одночасно алгоритм адаптується до цієї умови і переходить у режим BFS, що дозволяє отримати більш точну вибірку в певних точках. Це дозволяє NAOMInext обробляти навіть вузькі сайти зв'язування. Для підтримки швидкості дискретизації алгоритм негайно переходить у режим DFS. Розміщення спочатку сортуються за зменшенням скору. Починаючи з розміщення з найбільшим скором як початкового центру кластера, всі, що залишились додаються до кластера, якщо їх RMSD нижчий за заданий поріг (0,25 Å). Якщо RMSD більший, формується новий центр кластера. Якщо кількість кластерів перевищує заданий поріг (за замовчуванням: 128), поріг кластера збільшується з кроком 0,05 Å. Сортування розміщень у збільшеному порядку їх придатності

гарантує, що найкраще набране розміщення завжди використовується як представник кластера.

## **Скоринг**

Для оцінки кожного кроку вибірки під час нарощування фрагментів використовується NAOMInext емпірична функція підрахунку, що відповідає ChemScore [68,69]. Через велику кількість функцій оцінки, більш точні, але повільніші схеми оцінювання, такі як HYDE [70,71] не можна застосувати. Функція ChemScore будується з парних внесків взаємодій, таких як показник ліпофільності, здатності до зв'язування металу, і показнику водневого зв'язку. Крім того, використовуються дескриптори зв'язку, такі як дескриптор утруднення зв'язку, що обертається та внутрішньої деформації. Дескриптори зв'язку оцінюються лише під час остаточного оцінювання сполук. Всі інші (на основі атомів) використовуються під час додаткового процесу побудови молекули. Дескриптор зіткнення обмежений атомами, які мають принаймні три ковалентні зв'язки. Зіткнення атом-атом (між атомами, що не є воднем) відбувається, якщо міжатомна відстань менше суми обох радіусів Ван-дер-Ваальса (vdW), помноженої на постійний коефіцієнт пом'якшення  $k$  (0,7, що означає, що дозволено перекриття радіусів 30% vdW). Внутрішнє зіткнення підсумовується для кожної дійсної пари зіткнень  $ij$ . Дескриптор обмеження заснований на моделі гармонічного осцилятора, яка часто використовується в силових полях для моделювання зв'язку в молекулі як пружини, що з'єднує два атоми.

## **Підготовка набору даних**

Щоб перевірити ефективність підходу до вирощування фрагментів, був використаний вже опублікований набір даних від Malhotra і Karanicolas [72].

Цей набір даних містить пари молекул, де одна молекула є передбачуваним прекурсором (принаймні фрагментом) більшої lead-like молекули.

Опублікований набір даних складається з 297 молекулярних пар, де бібліотека програмного забезпечення NAOMI ідентифікувала 14 пар набору даних як ковалентно-пов'язані, засновані на тісних контактах атомів принаймні одного ліганда та атома білка. Три додаткові пари були виключені через те, що вони є ковалентно зв'язаним, оскільки SIENA [73] не змогла виділити дійсний ліганд. Крім того, потрібно було вилучити 16 додаткових пар з оцінки через неоднозначність даних: для вирощування фрагментів нам потрібно визначити правильний якірний фрагмент серед двох лігандів.

### **Робочий процес перевірки**

ASCONA [74], інтегрована у SIENA [73], використовувалася для накладання відповідних білкових структур споріднених пар молекул, якірного фрагменту та молекули ліду. Максимальний RMSD місця зв'язування (усі важкі атоми) окремих накладених білків був вищим за 3 Å, як це також згадували Malhotra та Karanicolas[72]. Суперпозиція сайтів зв'язування білка була необхідною для розрахунку RMSD генерованих розміщень на еталонну структуру. Щоб уникнути будь-якого зміщення вхідних координат, попередньо перераховуються 3D-координати еталонного ліганду за допомогою UNICON [64].

Для оцінки було використано GLIDE - сучасний інструмент для докінгу. Робочий процес докінгу GLIDE виконується за допомогою Python API Schroedinger MAESTRO Suite версії 2017.01, що дозволяє автоматизувати великомасштабний аналіз. [51,52] Усі відповідні команди python додаються до розділу Supporting Information в Glide Docking Workflow, а також реалізований Python скрипт. Спочатку майстер підготовки MAESTRO використовується для підготовки цільової структури білка включаючи якірний фрагмент. Потім ко-

кристалізований ліганд екстрагують з мішені білка. Отже, попередній вилучений ліганд використовується для визначення центру сітки. Розмір активного сайту подовжується на 20Å, оскільки ліганди більші, ніж кристалізований якірний фрагмент. Нарешті, GLIDE використовується для докінгу еталонної молекули з перерахованими 3D-координатами в цільовий білок. Отримані розміщення [75] потім використовуються для обчислення RMSD для еталонної структури за допомогою калькулятора RMSD на основі бібліотеки програмного забезпечення NAOMI. RMSD обчислюється лише між важкими атомами.

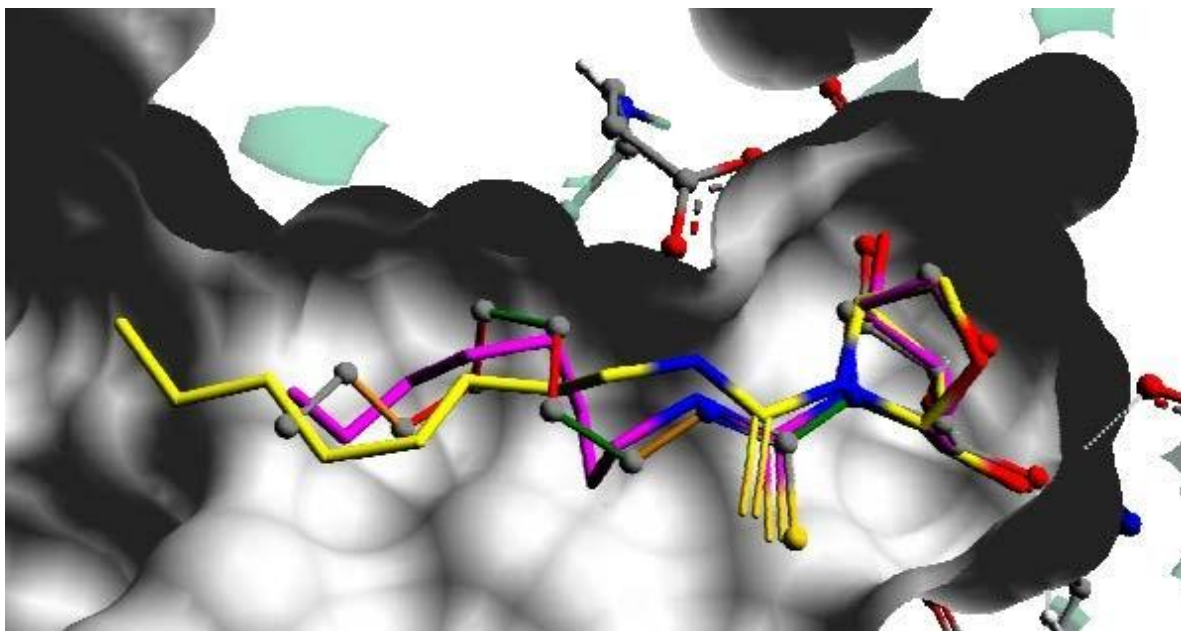
Оскільки NAOMInext призначений для широкої застосовності, програмне забезпечення було оцінено на великому наборі даних. Перша і друга оцінки перевіряють здатність програми відтворювати більшу молекулу на основі якірного фрагмента вже кристалізованої малої молекули в межах даного місця зв'язування. Тести є аналогами селф-докінгу та перехресного докінгу. Отже, для тесту проводиться вирощування та відбір проб у місці зв'язування більшої еталонної молекули. 3D координати якірного фрагмента переносяться у відповідні координати еталонного ліганду. Тест два проводиться в місці зв'язування кристалізованого якірного фрагмента, що сприяє реалістичному сценарію тесту. Успішність тесту визначається як RMSD нижче 2Å до кристалічної структури генерованої пози. У третьому оцінюванні показується кілька результатів реакції на важливі мішені у відкритті ліків та обговорюється ретроспективне дослідження.

Спочатку було використано набір даних для оцінки ефективності вибірки обох підходів, вирощування фрагментів та докінгу відповідно як базовий експеримент. Тому білкову структуру більшого ліганду було використано як мішень для селф-докінгу та вирощування фрагментів, щоб скласти перше враження про їхні особливості. Неможливо виконати докінг GLIDE для мішеней

4HV7, 2W8Y, 1YHS та 1O5F через збої під час підготовки білка за допомогою майстра підготовки SCHROEDINGER. Усі таргети в наборі даних можна обробляти за допомогою NAOMInext. GLIDE успішно добував 75% тестових сполук нижче 2 Å RMSD до еталонної структури. Підхід до вирощування фрагментів NAOMInext досяг 94% тестових даних нижче 2 Å RMSD з використанням однакових критеріїв оцінки. Це не дивно, оскільки NAOMInext використовує більше доступної інформації та прив'язує вибірку ліганду до початкових 3D-координат кристалізованого якірного фрагмента. Через відсутність повної автоматизації, провести тест на 264 випадках не вдалося.

### **Бета-глюкозидазний комплекс**

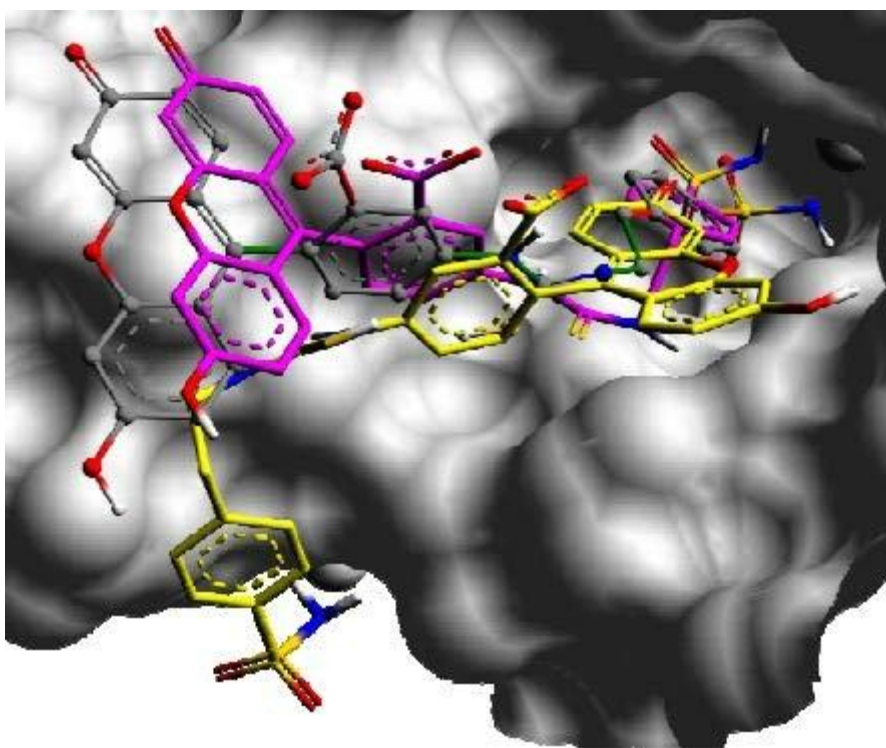
Розробка нових інгібіторів білка глюкозидази є важливою для лікування раку, ВІЛ та діабету. [77] Результати для структурної пари з ідентифікаторами PDB 2J77 і 2VRJ можна побачити на малюнку 6, найкраще розміщення NAOMInext має RMSD (порівняно з кристалічною структурою) 0,9 Å (пурпуровим), тоді як найкраще розміщення GLIDE має RMSD 1,5 Å (жовтим).



## СА II

Іншим прикладом, коли використання додаткової інформації про координати якірного фрагмента є корисним, є пара структури СА II з ідентифікаторами PDB 2F14 та 4E3G (див. Малюнок 7). Виявлено понад 14 різних ізоформ СА, і всі відіграють важливу фізіологічну роль. Неправильне регулювання може призвести до таких захворювань, як глаукома та рак. [78,79] Аніон цинку має важливе значення для каталітичної функціональності і знаходиться у заглибленні сайту зв'язування (15 Å), координований трьома залишками гістидину. [78] GLIDE не зміг створити правильне розміщення і змістив сульфаніламідний фрагмент не в кишеню, що зв'язує, а на поверхню білка (див. Малюнок 7), де найкраща поза має RMSD 10,9 Å порівняно з кристалічною структурою. Основним класом інгібіторів СА є сульфаніламідни, які зв'язуються з Zn (II) - іоном через свій атом азоту. Використання підходу до вирощування фрагментів із уже кристалізованим фрагментом може бути гарною відправною точкою для утворення нових сполук лідів. Хоча у інструментів для докінгу іноді виникають проблеми з докінгом з ферментами в яких наявний метал [94], вирощування фрагментів може подолати цю проблему, використовуючи наявні експериментальні дані (кристалізований якірний фрагмент). У цьому прикладі NAOMInext знайшов позу з RMSD 2,4 Å на 16-ій ітерації.

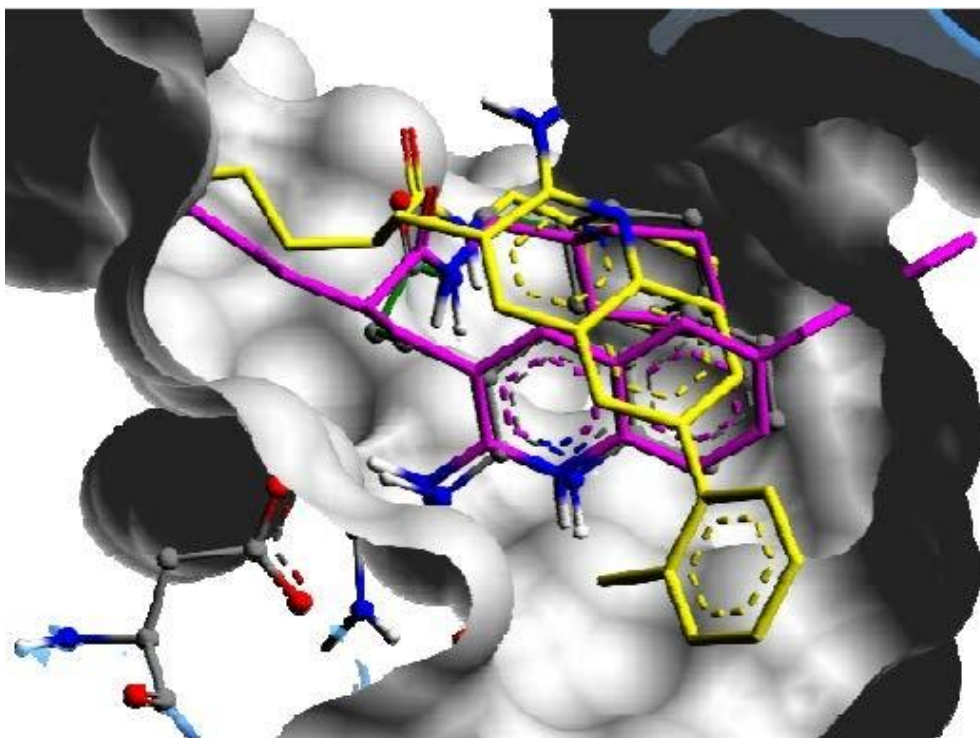




Тривалість роботи обох інструментів порівнянна з урахуванням необхідного часу для кожного ліганду. Майстер попередньої обробки GLIDE виконує процедуру мінімізації цілого білка, що може зайняти значні обчислювальні ресурси. Час докінгу на ліганд порівнянний із часом роботи NAOMInext. Вибірка фрагментів, що мають до восьми зв'язків, що обертаються виконується переважно протягом однієї секунди. Час роботи для більш гнучких фрагментів (до 13 зв'язків, що обертаються) взагалі не перевищує десяти секунд. Хоча NAOMInext використовує кілька стартових позицій (верхня межа за замовчуванням: 50), щоб врахувати навантаження на зростаючий фрагмент, медіана часу виконання на фрагмент збільшується лише лінійно до приблизно 24 секунд. Просте лінійне збільшення часу виконання досягається завдяки здатності алгоритму дискретизації динамічно адаптуватися до існуючих станів та обмежень.

## Оцінка результатів перехресних вибірок

Метою другого оцінювання було підібрати більш реалістичний варіант використання. Традиційно підходи до вирощування фрагментів у FBDD використовуються під час проектування ліду для еволюції фрагмента в невелику молекулу. Отже, вирощування слід проводити з використанням білкової структури кристалізованого фрагмента. У цій оцінці NAOMInext не зміг дати жодного результату для 16 із 264 записів, а GLIDE провалився у 17 випадках. Невдача в NAOMInext - це, наприклад, пара, де сайт зв'язування містить гнучкі залишки, зазначені через різні конформації бічного ланцюга в двох білкових структурах. Отже, вирощування не може відбуватися через зіткнення, оскільки гнучкість цілі ще не враховується (наприклад, інгібітор Vase-1 Рисунок 11). Для понад 70% з 264 тестових випадків NAOMInext досягла мінімального значення RMSD нижче 2 Å у верхніх 32 позиціях (див. Малюнок 9a). Використовуючи GLIDE у вищеписаному робочому процесі, продуктивність докінгу становить близько 52% (див. Малюнок 9b).



## **Лактатдегідрогеназа А (LDHA)**

Його функція каталізувати перетворення з пірувату на лактат і той факт, що він в основному регулюється в клінічних пухлинах, робить LDHA потенційною терапевтичною мішенню в області метаболізму раку. [81] Показано, що його інгібування призводить до окисного стресу та загибелі клітин [82,83] Пара структури з ідентифікаторами PDB 4AJK та 4AL4 базується на опублікованому дослідженні фрагментів [81]. У цьому дослідженні кристалізований фрагмент використовується як відправна точка для розвитку ліду (PDBid 4AJ1). Ліганд з PDBid 4AJK вже був розроблений за допомогою вирощування фрагментів, щоб мати можливість здійснювати зв'язування фрагментів з фрагментом, націленим на кишеню нікотинаміду (PDBid 4AJH, ліганд 2B4).

## Практична частина

База in-Stock речовин Life Chemicals була використана як джерело стартових фрагментів. Бази даних були отримані з усіх сполук Life Chemicals, що мають обмежені значення Sascore. Кількість сполук і результуючих пар фрагментів істотно зменшилась із зниженням порогу Sascore.

Вихідна база сполук Life Chemicals була перевірена на правильність за допомогою інструментів Chemaxon та RDKit.

Structure Checker - це інструмент для виявлення помилок в зображенні або структурних елементах у хімічних структурах. Окрім цього, небажані подання також можуть бути виправлені за допомогою перевірки структури, якщо це необхідно.

Structure Checker працює із заздалегідь визначеними параметрами перевірки, і ви можете вибрати, які з них слід враховувати під час запуску.

Після того, як всі структури пройшли перевірку, базу було фрагментовано.

Ідея взаємозамінних фрагментів – основа розробленого підходу - безпосередньо пов'язана із відповідним підходом до молекулярних пар з урахуванням їх локального контексту [33]. Взаємозамінні фрагменти - це фрагменти, які відповідають тому ж локальному хімічному контексту в структурах сполук. Атоми в межах конкретного радіусу навколо точок кріплення фрагмента представляють цей місцевий хімічний контекст. Замінюємо один фрагмент на інший, що має такий самий хімічний контекст, і результатом є хімічно обгрунтована і доступна структура. Хімічну «правильність» генерованих структур гарантовано. Можна також очікувати, що згенеровані сполуки є синтетично доступними.

Створення бази даних взаємозамінних фрагментів є двоетапною процедурою. На першому кроці структури відомих сполук фрагментуються шляхом розрізання до 4 нециклічних одинарних зв'язків між двома важкими атомами, що використовують алгоритм RDKit для зіставлення молекулярних пар, запропонований Хусейном та Реєю[34]. Водні відділяються окремо. На другому кроці, контекст заданого радіуса визначається для точок прикріплення кожного фрагмента і кодується в SMILES. Рядок SMILES, був канонізований, щоб отримати канонічну нумерацію точок прикріплення та канонічне подання SMILES контексту. Точки прикріплення у відповідному фрагменті перенумеровані відповідно. SMILES контексту заданого радіуса та пов'язаного фрагмента зберігається в таблиці бази даних як пара ключ-значення для наступного пошуку взаємозамінних фрагментів (значень), що мають ідентичний контекст (ключ). Якщо контекст двох або більше точок приєднання ідентичний, усі можливі перестановки цих точок кріплення у відповідному фрагменті виконуються. Нумери точок кріплення в контексті не змінюється, оскільки це призведе до однакового подання SMILES. Фрагменти з альтернативною нумерацією точок прикріплення зберігаються окремо як пари ключ-значення (контекст, у даному випадку, однаковий). Для заміни фрагмента в молекулі визначається контекст заданого радіуса і канонічно кодується. Даний рядок контексту SMILES шукається у базі даних фрагментів, а фрагменти з однаковим контекстом отримуються та використовуються для заміни фрагментів.

Є три режими формування структури: MUTATE, GROW, LINK.

MUTATE - це заміна довільно вибраного фрагмента іншим. GROW - це окремий випадок операції MUTATE - заміна водню іншим фрагментом.

LINK - являє собою заміщення атомів водню у двох молекулах на фрагмент, що їх зв'язує(РІС.1).

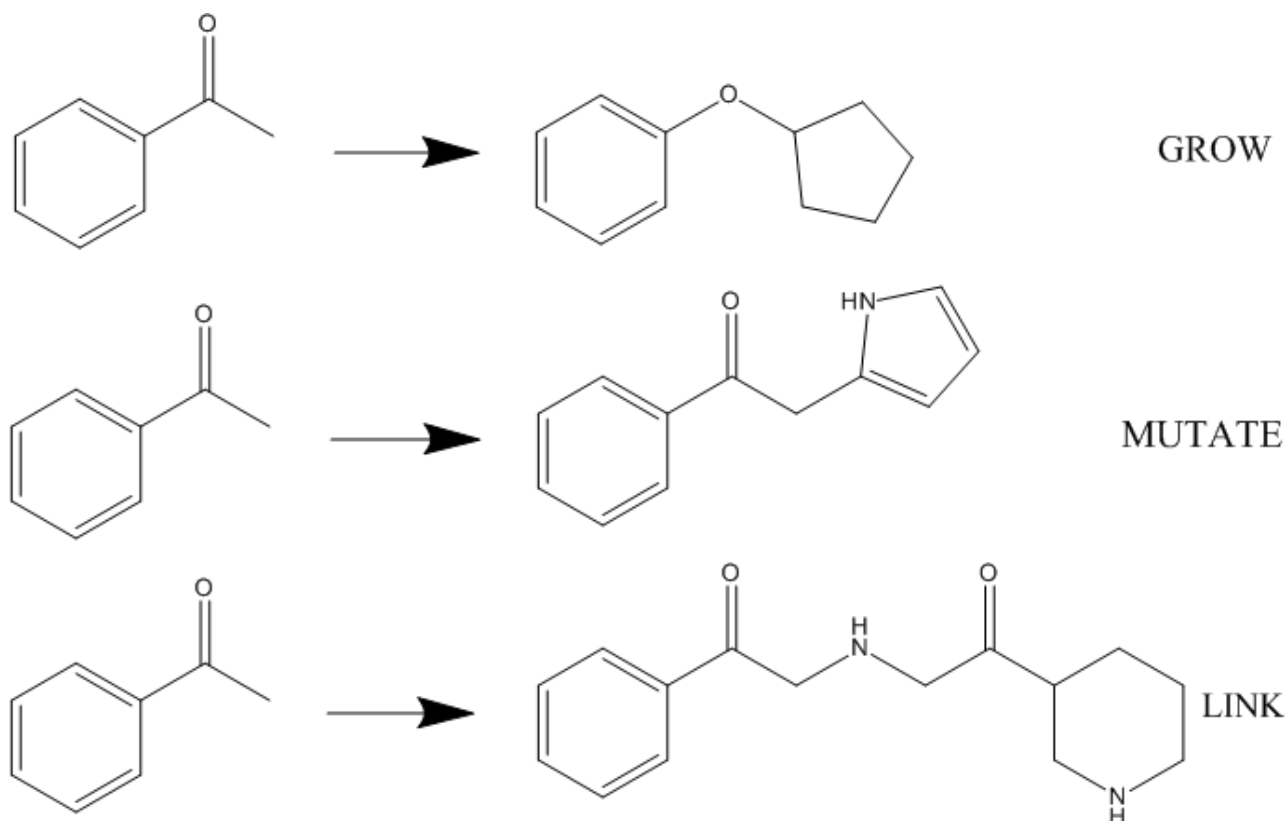


РИС.1 Нарощування фрагментів

Структури вхідних сполук використовуються для створення бази даних взаємозамінних фрагментів. Управління вмістом бази вхідних даних, що використовується для фрагментації, дає опосередкований контроль над переліченими структурами та. Вибір синтетично здійснених вихідних сполук може поліпшити синтетичну доступність отриманих сполук. У той же час попередній вибір сполук для переліку бібліотеки фрагментів може зменшити різноманітність та новизну створених структур. Збільшення радіуса розглянутого молекулярного контексту зменшить появу нових хемотипів у перелічених сполуках та зробить заміни більш консервативними. Подібно до оцінки синтетичної доступності, запропонованої Ertl & Schufenhauer [35], можна

припустити, що заміна фрагментами, що частіше трапляються, призведе до більш синтетично здійснених сполук. Це також зменшить кількість заміни і збільшить швидкість пошуку. Розміри змінних фрагментів можуть контролювати вичерпність досліджень хімічного простору, збільшуючи або зменшуючи кроки пошуку, і залежить від мети конкретного дослідження. Дослідження оптимізації ліду можуть вимагати невеликих кроків для вивчення місцевого хімічного простору навколо вихідної сполуки, тоді як генерація ліду може вимагати грубого дослідження більшого хімічного простору і менші кроки в кінці, щоб точно вивчити утворені структури. Обмеження максимальної кількості заміни може пришвидшити дослідження хімічного простору, оскільки генеровані бази даних фрагментів можуть бути дуже великими, і здійснення всіх можливих заміни може коштувати дорого. Захист виділених атомів від модифікації або модифікації лише вибраних атомів може бути корисним для досліджень оптимізації властивостей / активності для захисту масштабних чи фармакофорних ознак від змін або для модифікації молекул лише у визначених положеннях.

### **Принципи вибору стартових фрагментів:**

- 1) щоб молекули були досить складними (число атомів);
- 2) могли зв'язуватися досить специфічно (число Н-донорів і акцепторів),
- 3) не дуже ліпофільними ( $\log P$ , TPSA);
- 4) не мали б великих циклів (більше 7), щоб не виникало проблем з семпліруванням конформерів;
- 5) обмежена загальна кількість циклів, щоб виключити поліциклічні структури, вони не мають найчастіше сенсу при дизайні;
- 6) не були б занадто гнучкими (RTB), тому що це відразу дасть багато ступенів свободи, а це ускладнить пошук правильних поз в докінгу і обмежить пошук, тому що RTB використовується як один із критеріїв зупинки пошуку;
- 7) виключити полігалогенні структури (крім фтору), тому що якщо буде потрібно алгоритм сам зможе додати галогени;
- 8) щоб фрагменти не містили небажаних груп ні по одному зі структурних фільтрів, щоб виключити реакційно здатні і не драг-лайк молекули і зменшити число стартових фрагментів до розумного значення (кілька тисяч)



*Фізико-хімічні параметри, що застосовувались для вибору стартових фрагментів:*

H-donors/H-acceptors 2-5, загальне число атомів, які можуть бути донорами або акцепторами Н-зв'язку, якщо атом і донор і акцептор він враховується лише 1 раз, наприклад ОН;

NumRings 1-3, число циклів;

NumRingsFused  $\leq 2$ , число циклів з урахуванням конденсованих і спіро;

RTB  $\leq 2$ , число зв'язків, що обертаються;

logP  $\leq 2$ ,

TPSA  $> 25$ ,

НАС 8-15, число важких атомів;

nHal  $\leq 1$ , число галогенів крім фтору;

max\_ring\_size  $\leq 7$ , розмір найбільшого циклу;

*Структурні фільтри, що застосовувались для вибору стартових фрагментів:*

Rule\_BMS

Rule\_Dundee

Rule\_Glaxo

Rule\_Inpharmatica

Rule\_LINT

Rule\_MLSMR

Rule\_PAINS

Rule\_SureChEMBL

## **Про стратегії:**

### **1 - жадібний алгоритм**

Серед усіх сполук здокованих на поточній інтеракції вибирають ті, які задовільняють MW, RTB та rmsd, потім серед них вибирається фіксоване число молекул (зазначене у значущих аргументах) з максимальним скорингом. Вибрані молекули нарощуються і передаються на наступну інтеракцію.

### **4 - алгоритм Парето**

серед усіх здокованих сполук поточної інтеракції вибирають ті, які задовільняють MW, RTB та rmsd. Середні показники вибирають з'єднання, що лежать на парето-фронті в координатах MW - docking\_score (см. Малюнок нижче, лінія, що з'єднує точки - парето фронт), тобто вибираються сполуки з максимальним скорингом і мінімальною масою. Це повинно дати шанс менш крупним сполукам з меншими скорингами дорости до чогось більшого (у жадібній стратегії фокус був лише на скорингу). Очікувалось, що це підвищило структурну різноманітність серед лігандів з високим скорингом.

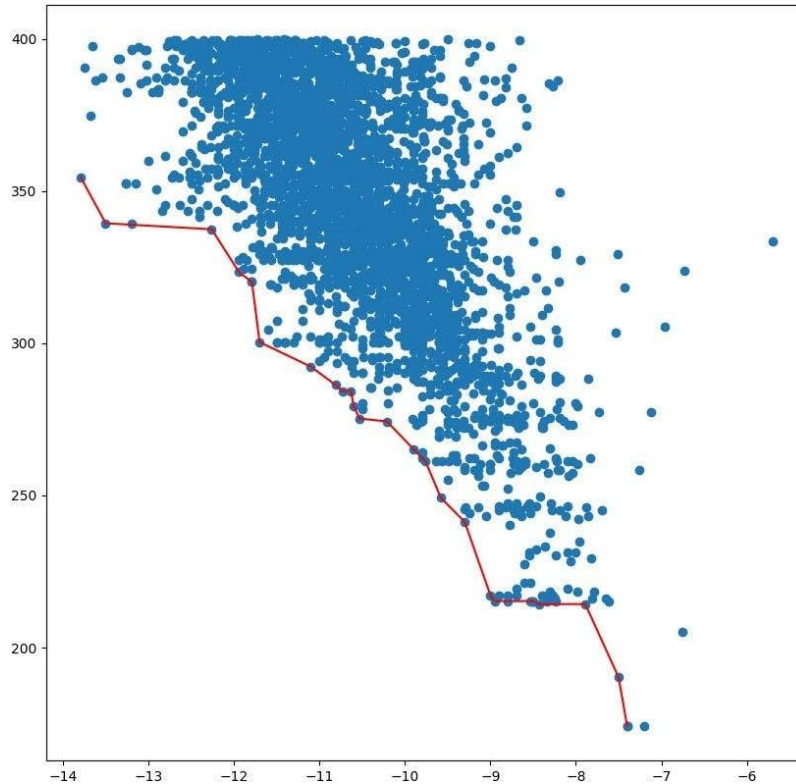


РИС.2 Парето фронт

**Параметри скрипта для алгоритмів:**

"db": "replacements02\_sa2.db", - база даних з фрагментами, які будуть додаватися до докованих молекул, ми використовували 3 бази (усі, sa2, sa2.5);

"input\_fragments": "selected.smi", - файл зі смайлами стартових фрагментів;

"max\_atoms": 10, - максимальне число атомів у доданому фрагменті з баз;

"max\_replacements": 2000, - максимальне число нарощування для кожної молекули, якщо якусь молекулу можна збільшити більше ніж цим числовим способом, то програма вибереть для неї випадкові 2000 замен;

"mgl\_install\_dir": "bin / mgltools\_x86\_64Linux2\_1.5.6", - путь до папки з mgltools;

"min\_atoms": 1, - мінімальне число атомів у доданому фрагменті з баз;

"min\_freq": 0, - мінімальна зустрічаємость фрагментів у базі (0, означає зустрічаємость ігнорується);

"mol\_weight": 450, - максимальна молекулярна маса, якщо молекули стоять більшими, вони не вибираються при наступній ітерації (якщо молекул з меншою масою не залишається, ітерація завершується)

"ncpu": 32, - число ядер

"ntop": 20, - число сполук, які вибираються на наступну ітерацію;

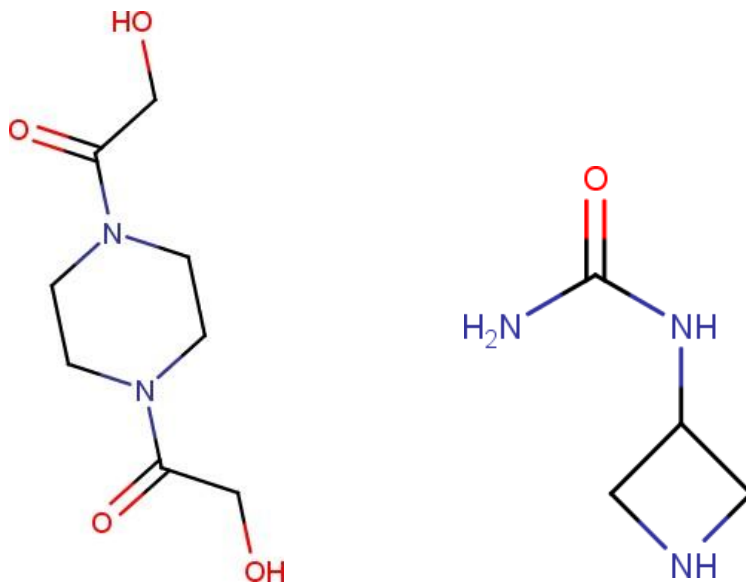
"output": "5es1\_A\_t1\_sa2 / 5es1\_A\_t1\_sa2.db", - база даних будеуть зберігатися результати генерації;

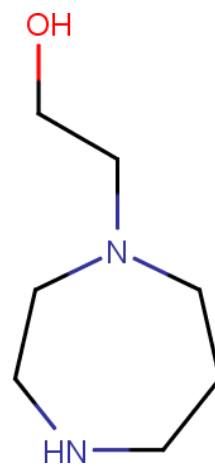
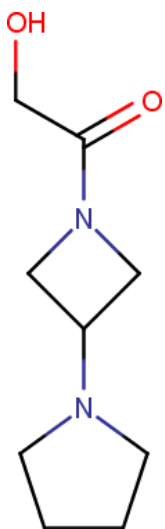
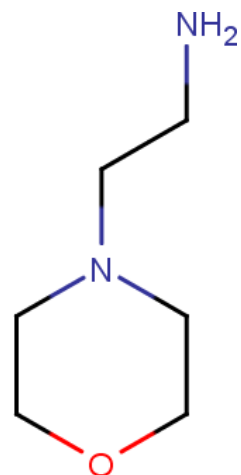
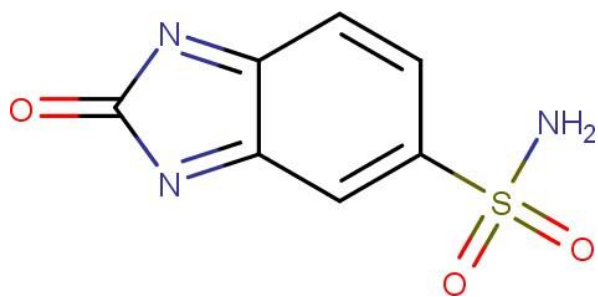
"protein": "5es1\_A.pdbqt", - файл з підготовленим белком;

"protein\_setup": "5es1.log", - файл із параметрами положень, куди докувати;

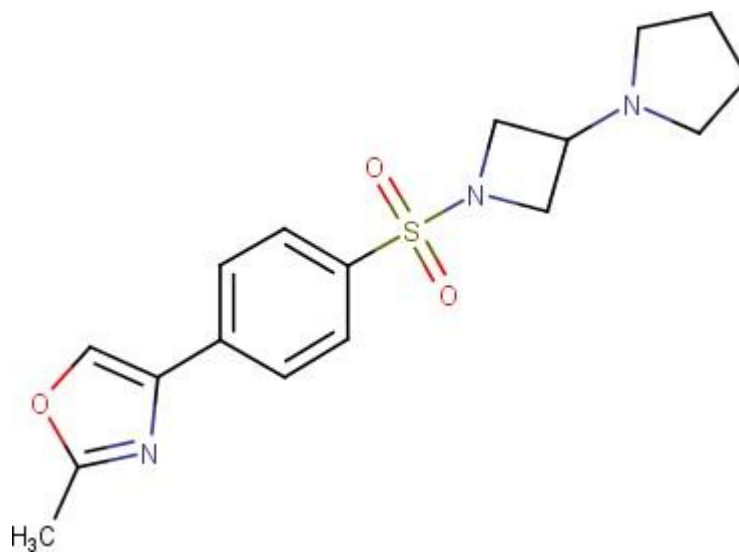
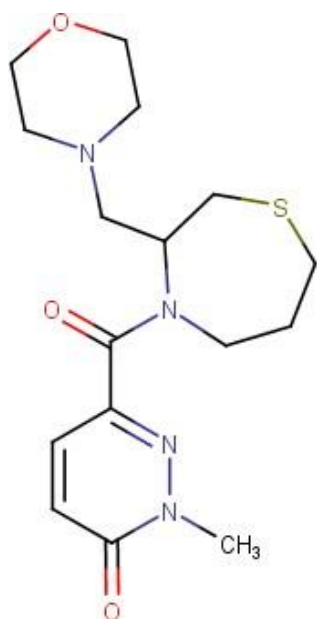
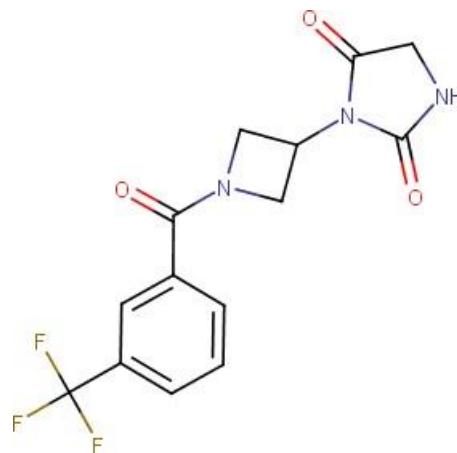
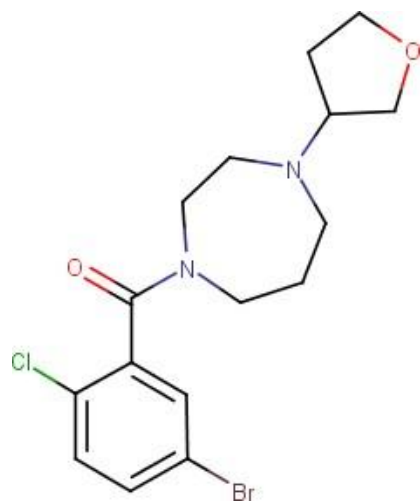
"radius": 3, - радіус контексту для пошуку фрагментів у базі, щоб додати їх до молекул;

**Приклади якірних фрагментів, що використовувались для нарощування:**





Приклади сполук, на основі яких формувалась бібліотека фрагментів:



Нарощування фрагментів відбувалося в активному центрі MARK4 кінази. Протеїнкінази відіграють важливу роль у нейродегенеративних, ракових та інших обмінних захворюваннях, оскільки вони беруть участь у регуляції різних сигнальних шляхів [1]. MARK4 - це кіназа Ser -Thr, яка відіграє ключову роль у динаміці мікротрубочок, оскільки вона фосфорилує сериновий мотив у домені зв'язування мікротрубочок тау. Надмірна експресія MARK4 призводить до розвитку метастатичних захворювань молочної залози та інших карцином [4 - 6]. Крім того, MARK4 бере участь в адипогенезі та апоптозі, індукованому шляхом JNK1, демонструючи важливість MARK4 у життєво важливих шляхах [7]. MARK4 безпосередньо пов'язаний з нейродегенеративними захворюваннями, а його надмірна експресія індукує дефекти синапсів та дендритних шипів [8]. MARK4 пропонується як одна з можливих мішеней для лікування хвороби Альцгеймера[9]. Всі ці дослідження вказують на зв'язок MARK4 з діабетом, раком та іншими нейродегенеративними захворюваннями [10 -13].

## **Висновок**

Розроблений підхід генерує хімічно допустимі структури. Цей підхід також дозволяє побічно впливати на результат генерації шляхом налаштування бази даних вхідних даних, яка потім використовується для формування бази даних взаємозамінних фрагментів. Вибір більш синтетично доступних вихідних сполук для фрагментації може призвести до отримання синтетично доступніших сполук. Проведені експерименти показали, що навіть маючи невелику бібліотеку синтетично здійсненних сполук, можна створити досить різноманітні набори структур з більш передбачуваною синтетичною доступністю.



## Використана література:

1. Polishchuk PG, Madzhidov TI, Varnek A (2013) Estimation of the size of drug-like chemical space based on GDB-17 data. *J Comput Aided Mol Des.* 27:675–679. <https://doi.org/10.1007/s10822-013-9672-4>
2. Schneider P, Schneider G (2016) De novo design at the edge of chaos. *J Med Chem* 59:4077–4086. <https://doi.org/10.1021/acs.jmedchem.5b01849>
3. Schneider G (2017) Automating drug discovery. *Nat Rev Drug Discovery* 17:97. <https://doi.org/10.1038/nrd.2017.232>
4. Hoksza D, Škoda P, Voršilák M, Svozil D (2014) Molpher: a software framework for systematic chemical space exploration. *J Cheminform* 6:7. <https://doi.org/10.1186/1758-2946-6-7>
5. Hartenfeller M, Zettl H, Walter M, Rupp M, Reisen F, Proschak E, Weggen S, Stark H, Schneider G (2012) DOGS: reaction-driven de novo design of bioactive compounds. *PLoS Comput Biol* 8:e1002380
6. Szymkuć S, Gajewska EP, Klucznik T, Molga K, Dittwald P, Startek M, Bajczyk M, Grzybowski BA (2016) Computer-assisted synthetic planning: the end of the beginning. *Angew Chem Int Ed* 55:5904–5937. <https://doi.org/10.1002/anie.201506101>
7. Batiste L, Unzue A, Dolbois A, Hassler F, Wang X, Deerrain N, Zhu J, Spiliotopoulos D, Nevado C, Cafisch A (2018) Chemical space expansion of bromodomain ligands guided by in silico virtual couplings (AutoCouple). *ACS Cent Sci* 4:180–188. <https://doi.org/10.1021/acscentsci.7b00401>
8. Merk D, Grisoni F, Friedrich L, Gelzinyte E, Schneider G (2018) Computerassisted discovery of retinoid X receptor modulating natural products and isofunctional

mimetics. *J Med Chem* 61:5442–5447. <https://doi.org/10.1021/acs.jmedchem.8b00494>

9. Wang R, Gao Y, Lai L (2000) LigBuilder: a multi-purpose program for structure-based drug design. *Mol Model Annu* 6:498–516. <https://doi.org/10.1007/s0089400060498>

10. Chéron N, Jasty N, Shakhnovich EI (2016) OpenGrowth: an automated and rational algorithm for finding new protein ligands. *J Med Chem* 59:4171–4188. <https://doi.org/10.1021/acs.jmedchem.5b00886>

11. Kutchukian PS, Lou D, Shakhnovich EI (2009) FOG: fragment optimized growth algorithm for the de novo generation of molecules occupying druglike chemical space. *J Chem Inf Model* 49:1630–1642. <https://doi.org/10.1021/ci9000458>

12. Liu T, Naderi M, Alvin C, Mukhopadhyay S, Brylinski M (2017) Break down in order to build up: decomposing small molecules for fragment-based drug design with eMolFrag. *J Chem Inf Model* 57:627–631. <https://doi.org/10.1021/acs.jcim.6b00596>

13. Olivecrona M, Blaschke T, Engkvist O, Chen H (2017) Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics* 9:48. <https://doi.org/10.1186/s13321-017-0235-x>

14. Popova M, Isayev O, Tropsha A (2018) Deep reinforcement learning for de-novo drug design. *Scie Adv* 4(7):eaap7885. <https://doi.org/10.1126/sciad.v.aap7885>

15. Yuan W, Jiang D, Nambiar DK, Liew LP, Hay MP, Bloomstein J, Lu P, Turner B, Le Q-T, Tibshirani R, Khatri P, Moloney MG, Koong AC (2017) Chemical space mimicry for drug discovery. *J Chem Inf Model* 57:875–882. <https://doi.org/10.1021/acs.jcim.6b00754>

16. Li Y, Zhang L, Liu Z (2018) Multi-objective de novo drug design with conditional graph generative model. *J Cheminform.* 10(1):33. <https://doi.org/10.1186/s13321-018-0287-6>
17. Polykovskiy D, Zhebrak A, Vetrov D, Ivanenkov Y, Aladinskiy V, Mamoshina P, Bozdaganyan M, Aliper A, Zhavoronkov A, Kadurin A (2018) Entangled conditional adversarial autoencoder for de novo drug discovery. *Mol Pharm* 15:4398–4405. <https://doi.org/10.1021/acs.molpharmaceut.8b00839>
18. Putin E, Asadulaev A, Ivanenkov Y, Aladinskiy V, Sanchez-Lengeling B, Aspuru-Guzik A, Zhavoronkov A (2018) Reinforced adversarial neural computer for de novo molecular design. *J Chem Inf Model.* <https://doi.org/10.1021/acs.jcim.7b00690>
19. Segler MHS, Kogej T, Tyrchan C, Waller MP (2018) Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent Sci* 4:120–131. <https://doi.org/10.1021/acscentsci.7b00512>
20. Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, Aguilera-Iparraguirre J, Hirzel TD, Adams RP, Aspuru-Guzik A (2018) Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci* 4:268–276. <https://doi.org/10.1021/acscentsci.7b00572>
21. Elton DC, Boukouvalas Z, Fuge MD, Chung PW (2019) Deep learning for molecular design—a review of the state of the art. *Mol Syst Des Eng* 4:828–849. <https://doi.org/10.1039/C9ME00039A>
22. M. Congreve, R. Carr, C.W. Murray, H. Jhoti, A “Rule of Three” for fragment-based lead discovery?, *Drug Discov. Today.* 8 (2003) 876–877. doi:10.1016/S1359-6446(03)02831-9.

23. V. Velvadapu, B.T. Farmer, A.B. Reitz, Fragment-Based Drug Discovery, in: C. Wermuth, D. Aldous, P. Raboisson, D. Rognan (Eds.), *Pract. Med. Chem.*, 4th ed., Academic Press, 2015: pp. 161–180. doi:10.1016/B978-0-12-417205-0.00007-9.
24. P.J. Hajduk, J. Greer, A decade of fragment-based drug design: Strategic advances and lessons learned, *Nat. Rev. Drug Discov.* 6 (2007) 211–219. doi:10.1038/nrd2220.
25. M.M. Hann, A.R. Leach, G. Harper, Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery, *J. Chem. Inf. Comput. Sci.* 41 (2001) 856–864. doi:10.1021/ci000403i.
26. D.A. Erlanson, Introduction to Fragment-Based Drug Discovery, in: T.G. Davies, M. Hyvönen (Eds.), *Fragm. Drug Discov. X-Ray Crystallogr.*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012: pp. 1–32. doi:10.1007/128\_2011\_180.
27. A.L. Hopkins, C.R. Groom, A.A. Alex, Ligand efficiency: a useful metric for lead selection, *Drug Discov. Today.* 9 (2004) 430–431. doi:10.1016/S1359-6446(04)03069-7.
- 28.[10] D.E. Scott, A.G. Coyne, S.A. Hudson, C. Abell, Fragment-based approaches in drug discovery and chemical biology, *Biochemistry.* 51 (2012) 4990–5003. doi:10.1021/bi3005126.
- 29.[24] J.B. Moon, W.J. Howe, Computer Design of Bioactive Molecules - a Method for Receptor-Based de Novo Ligand Design, *Proteins-Structure Funct. Genet.* 11 (1991) 314–328. //a1991gv03200008.
- 30.[25] R.S. DeWitte, E.I. Shakhnovich, SMOG: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence, *J. Am. Chem. Soc.* 118 (1996) 11733–11744. doi:10.1021/ja960751u.
- 31.[26] S.H. Rotstein, M.A. Murcko, GenStar: A method for de novo drug design, *J. Comput. Aided. Mol. Des.* 7 (1993) 23–43. doi:10.1007/BF00141573.
- 32.[27] N.-O. Friedrich, C. de Bruyn Kops, F. Flachsenberg, K. Sommer, M. Rarey, J. Kirchmair, Benchmarking Commercial Conformer Ensemble Generators, *J. Chem. Inf. Model.* 57 (2017) 2719–2728. doi:10.1021/acs.jcim.7b00505.
33. [11] P.G. Wyatt, A.J. Woodhead, V. Berdini, J.A. Boulstridge, M.G. Carr, D.M. Cross, D.J. Davis, L.A. Devine, T.R. Early, R.E. Feltell, E.J. Lewis, R.L. McMenamain, E.F. Navarro, M.A. O'Brien, M. O'Reilly, M. Reule, G. Saxty, L.C.A.A. Seavers, D.-M.M. Smith, M.S. Squires, G. Trewartha, M.T. Walker, A.J.A. Woolford, M.A. O'Brien, M. O'Reilly, Identification of N-(4-piperidinyl)-4-(2,6-

- dichlorobenzoylamino)-1Hpyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design, *J. Med. Chem.* 51 (2008) 4986–4999. doi:10.1021/jm800382h.
34. S.H. Rotstein, M.A. Murcko, GroupBuild: a fragment-based method for de novo drug design., *J. Med. Chem.* 36 (1993) 1700–1710. doi:10.1016/0263-7855(94)80069-3.
35. V.J. Gillet, A.P. Johnson, P. Mata, S. Sike, P. Williams, SPROUT: A program for structure generation, *J. Comput. Aided. Mol. Des.* 7 (1993) 127–153. doi:10.1007/BF00126441.
36. P.C.D. Hawkins, Conformation Generation: The State of the Art, *J. Chem. Inf. Model.* 57 (2017) 1747–1756. doi:10.1021/acs.jcim.7b00221.
37. J.C. Baber, M. Feher, Predicting Synthetic Accessibility: Application in Drug Discovery and Development, *Mini-Reviews Med. Chem.* 4 (2004) 681–692. doi:10.2174/1389557043403765.
38. F. Chevillard, H. Rimmer, C. Betti, E. Pardon, S. Ballet, N. van Hilten, J. Steyaert, W.E. Diederich, P. Kolb, Binding-Site Compatible Fragment Growing Applied to the Design of  $\beta$  2 -Adrenergic Receptor Ligands, *J. Med. Chem.* 61 (2018) 1118–1129. doi:10.1021/acs.jmedchem.7b01558.
39. N. Chéron, N. Jasty, E.I. Shakhnovich, OpenGrowth: An Automated and Rational Algorithm for Finding New Protein Ligands, *J. Med. Chem.* 59 (2016) 4171–4188. doi:10.1021/acs.jmedchem.5b00886.
40. D.A. Pearlman, M.A. Murcko, CONCEPTS: New dynamic algorithm for de novo drug suggestion, *J. Comput. Chem.* 14 (1993) 1184–1193. doi:10.1002/jcc.540141008.
41. V.J. Gillet, W. Newell, P. Mata, G. Myatt, S. Sike, Z. Zsoldos, A.P. Johnson, SPROUT: Recent developments in the de novo design of molecules, *J. Chem. Inf. Model.* 34 (1994) 207–217. doi:10.1021/ci00017a027.
42. P. Mata, V.J. Gillet, A.P. Johnson, J. Lampreia, G.J. Myatt, S. Sike, A.L. Stebbings, SPROUT: 3D Structure Generation Using Templates, *J. Chem. Inf. Model.* 35 (1995) 479–493. doi:10.1021/ci00025a016. 43.[42] V.J. Gillet, G. Myatt, Z. Zsoldos, A.P. Johnson, SPROUT, HIPPO and CAESA: Tools for de novo structure generation and estimation of synthetic accessibility, *Perspect. Drug Discov. Des.* 3 (1995) 34–50. doi:10.1007/BF02174466.

44. H.J. Böhm, The computer program LUDI: A new method for the de novo design of enzyme inhibitors, *J. Comput. Aided. Mol. Des.* 6 (1992) 61–78. doi:10.1007/BF00124387.
45. M. Hartenfeller, H. Zettl, M. Walter, M. Rupp, F. Reisen, E. Proschak, S. Weggen, H. Stark, G. Schneider, Dogs: Reaction-driven de novo design of bioactive compounds, *PLoS Comput. Biol.* 8 (2012). doi:10.1371/journal.pcbi.1002380.
46. K.D. Stewart, M. Shiroda, C.A. James, Drug Guru: A computer software program for drug design using medicinal chemistry rules, *Bioorganic Med. Chem.* 14 (2006) 7011–7022. doi:10.1016/j.bmc.2006.06.024.
47. G. Schneider, M.L. Lee, M. Stahl, P. Schneider, De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks, *J. Comput. Aided. Mol. Des.* 14 (2000) 487–494. doi:10.1023/A:1008184403558.
48. G. Morris, R. Huey, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J. ....* 30 (2009) 2785–2791. doi:10.1002/jcc.21256.AutoDock4.
49. M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm., *J. Mol. Biol.* 261 (1996) 470–89. doi:10.1006/jmbi.1996.0477.
50. G. Jones, P. Willett, R. Glen, A. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol.* 267 (1997) 727–748. doi:10.1006/jmbi.1996.0897.
51. R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, *J. Med. Chem.* 47 (2004) 1739–1749. doi:10.1021/jm0306430.
52. M.P. Repasky, R.B. Murphy, J.L. Banks, J.R. Greenwood, I. Tubert-Brohman, S. Bhat, R.A. Friesner, Docking performance of the glide program as evaluated on the Astex and DUD datasets: A complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide, *J. Comput. Aided. Mol. Des.* 26 (2012) 787–799. doi:10.1007/s10822-012-9575-9.
53. R. Wang, Y. Gao, L. Lai, LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design, *J. Mol. Model.* 6 (2000) 498–516. doi:10.1007/s0089400060498.

54. Y. Yuan, J. Pei, L. Lai, LigBuilder 2: A practical de novo drug design approach, *J. Chem. Inf. Model.* 51 (2011) 1083–1091. doi:10.1021/ci100350u.
55. E. Shang, Y. Yuan, X. Chen, Y. Liu, J. Pei, L. Lai, De novo design of multitarget ligands with an iterative fragment-growing strategy, *J. Chem. Inf. Model.* 54 (2014) 1235–1241. doi:10.1021/ci500021v.
56. M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K.H. Altmann, G. Schneider, E. Jacoby, S. Renner, A collection of robust organic synthesis reactions for in silico molecule design, *J. Chem. Inf. Model.* 51 (2011) 3093–3098. doi:10.1021/ci200379p.
57. H.C. Kolb, M.G. Finn, K.B. Sharpless, Click Chemistry: Diverse Chemical Function from a Few Good Reactions, *Angew. Chemie - Int. Ed.* 40 (2001) 2004–2021. doi:10.1002/1521-3773(20010601)40:113.0.CO;2-5.
58. F. Chevillard, Improved approaches to ligand growing through fragment docking and fragment-based library design Florent Chevillard, Philipps-Universität Marburg, 2016.
59. S. Makino, T.J. Ewing, I.D. Kuntz, DREAM++: flexible docking program for virtual combinatorial libraries., *J. Comput. Aided. Mol. Des.* 13 (1999) 513–32. <http://www.ncbi.nlm.nih.gov/pubmed/10483532> (accessed August 4, 2017).
60. J.D. Durrant, S. Lindert, J.A. Mccammon, AutoGrow 3 . 0 : An improved algorithm for chemically tractable , semi-automated protein inhibitor design, *J. Mol. Graph. Model.* 44 (2013) 104–112. doi:10.1016/j.jm gm.2013.05.006.
61. H.M. Vinkers, M.R. de Jonge, F.F.D. Daeyaert, J. Heeres, L.M.H. Koymans, J.H. van Lenthe, P.J. Lewi, H. Timmerman, K. Van Aken, P.A.J. Janssen, SYNOPSIS: SYNthesize and OPTimize System in Silico, (2003). doi:10.1021/JM030809X.
62. Daylight Theory: SMARTS - A Language for Describing Molecular Patterns, (n.d.). <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html> (accessed October 5, 2018).
63. Daylight Theory: SMIRKS - A Reaction Transform Language, (n.d.). <http://www.daylight.com/dayhtml/doc/theory/theory.smirks.html> (accessed October 5, 2018).
64. K. Sommer, N.-O. Friedrich, S. Bietz, M. Hilbig, T. Inhester, M. Rarey, UNICON: A Powerful and Easy-to-Use Compound Library Converter, *J. Chem. Inf. Model.* 56 (2016) 1105–1111. doi:10.1021/acs.jcim.6b00069.

65. S. Urbaczek, A. Kolodzik, J.R. Fischer, T. Lippert, S. Heuser, I. Groth, T. SchulzGasch, M. Rarey, NAOMI: On the almost trivial task of reading molecules from different file formats, *J. Chem. Inf. Model.* 51 (2011) 3199–3207. doi:10.1021/ci200324e.
66. A. Griewel, O. Kayser, J. Schlosser, M. Rarey, Conformational sampling for largescale virtual screening: accuracy versus ensemble size, *J. Chem. Inf. Model.* 49 (2009) 2303–2311. doi:10.1021/ci9002415.
67. W. Welch, J. Ruppert, A.N. Jain, Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites, *Chem. Biol.* 3 (1996) 449–462. doi:10.1016/S1074-5521(96)90093-9.
68. M.D. Eldridge, C.W. Murray, T.R. Auton, G. V Paolini, R.P. Mee, Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, *J. Comput. Aided. Mol. Des.* 11 (1997) 425–445. doi:10.1023/a:1007996124545.
69. C.A. Baxter, C.W. Murray, D.E. Clark, D.R. Westhead, M.D. Eldridge, Flexible docking using Tabu search and an empirical estimate of binding affinity, *Proteins Struct. Funct. Genet.* 33 (1998) 367–382. doi:10.1002/(SICI)1097-0134(19981115)33:33.0.CO;2-W.
70. I. Reulecke, G. Lange, J. Albrecht, R. Klein, M. Rarey, Towards an integrated description of hydrogen bonding and dehydration: Decreasing false positives in virtual screening with the HYDE scoring function, *ChemMedChem.* 3 (2008) 885–897. doi:10.1002/cmdc.200700319.
71. N. Schneider, G. Lange, S.A. Hindle, R. Klein, M. Rarey, A consistent description of HYdrogen bond and DEhydration energies in protein-ligand complexes: Methods behind the HYDE scoring function, *J. Comput. Aided. Mol. Des.* 27 (2013) 15–29. doi:10.1007/s10822-012-9626-2.
72. S. Malhotra, J. Karanicolas, When Does Chemical Elaboration Induce a Ligand To Change Its Binding Mode?, *J. Med. Chem.* 60 (2017) 128–145. doi:10.1021/acs.jmedchem.6b00725.
73. S. Bietz, M. Rarey, SIENA: Efficient Compilation of Selective Protein Binding Site Ensembles, *J. Chem. Inf. Model.* 56 (2016) 248–259. doi:10.1021/acs.jcim.5b00588.



74. S. Bietz, M. Rarey, ASCONA: Rapid Detection and Alignment of Protein Binding Site Conformations, *J. Chem. Inf. Model.* 55 (2015) 1747–1756. doi:10.1021/acs.jcim.5b00210.
75. F. Chevillard, P. Kolb, SCUBIDOO: A Large yet Screenable and Easily Searchable Database of Computationally Created Chemical Compounds Optimized toward High Likelihood of Synthetic Tractability, *J. Chem. Inf. Model.* 55 (2015) 1824–1835. doi:10.1021/acs.jcim.5b00203.
76. P.S. Kutchukian, D. Lou, E.I. Shakhnovich, FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules Occupying Druglike Chemical Space, *J. Chem. Inf. Model.* 49 (2009) 1630–1642. doi:10.1021/ci9000458.
77. T.M. Gloster, P. Meloncelli, R. V. Stick, D. Zechel, A. Vasella, G.J. Davies, Glycosidase inhibition: An assessment of the binding of 18 putative transition-state mimics, *J. Am. Chem. Soc.* 129 (2007) 2345–2354. doi:10.1021/ja066961g.
78. C.T. Supuran, A. Scozzafava, A. Casini, Carbonic anhydrase inhibitors, *Med. Res. Rev.* 23 (2003) 146–189. doi:10.1002/med.10025.
79. Vijay M. Krishnamurthy, George K. Kaufman, Adam R. Urbach, Irina Gitlin, Katherine L. Gudiksen, and Douglas B. Weibel, G.M. Whitesides\*, Carbonic Anhydrase as a Model for Biophysical and Physical-Organic Studies of Proteins and Protein–Ligand Binding, (2008). doi:10.1021/CR050262P.
80. D.T. Moustakas, P.T. Lang, S. Pegg, E. Pettersen, I.D. Kuntz, N. Brooijmans, R.C. Rizzo, Development and validation of a modular, extensible docking program: DOCK 5, *J. Comput. Aided. Mol. Des.* 20 (2006) 601–619. doi:10.1007/s10822-006-9060-4.
81. R.A. Ward, C. Brassington, A.L. Breeze, A. Caputo, S. Critchlow, G. Davies, L. Goodwin, G. Hassall, R. Greenwood, G.A. Holdgate, M. Mrosek, R.A. Norman, S. Pearson, J. Tart, J.A. Tucker, M. Vogtherr, D. Whittaker, J. Wingfield, J. Winter, K. Hudson, Design and synthesis of novel lactate dehydrogenase a inhibitors by fragment-based lead generation, *J. Med. Chem.* 55 (2012) 3285–3306. doi:10.1021/jm201734r.
82. A. Le, C.R. Cooper, A.M. Gouw, R. Dinavahi, A. Maitra, L.M. Deck, R.E. Royer, D.L. Vander Jagt, G.L. Semenza, C. V. Dang, Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression, *Proc. Natl. Acad. Sci.* 107 (2010) 2037–2042. doi:10.1073/pnas.0914433107.

83. C. Granchi, S. Roy, C. Giacomelli, M. Macchia, T. Tuccinardi, A. Martinelli, M. Lanza, L. Betti, G. Giannaccini, A. Lucacchini, N. Funel, L.G. León, E. Giovannetti, G.J. Peters, R. Palchauthuri, E.C. Calvaresi, P.J. Hergenrother, F. Minutolo, Discovery of N -Hydroxyindole-Based Inhibitors of Human Lactate Dehydrogenase Isoform A (LDH-A) as Starvation Agents against Cancer Cells, *J. Med. Chem.* 54 (2011) 1599–1612. doi:10.1021/jm101007q.