

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
ІНСТИТУТ БІООРГАНІЧНОЇ ХІМІЇ ТА НАФТОХІМІЇ

На правах рукопису

ТАНЧУК Всеволод Юлійович

ПОШУК НАБОРІВ ДЕСКРИПТОРІВ, ЩО ПІДВИЩУЮТЬ ЕФЕКТИВНІСТЬ  
ПРОГНОЗУВАННЯ ФІЗІОЛОГІЧНОЇ АКТИВНОСТІ РЕЧОВИН НА ОСНОВІ  
ЇХ СТРУКТУРИ

02.00.10-біоорганічна хімія, хімія природних та фізіологічно активних речовин

АВТОРЕФЕРАТ

дисертації на здобуття наукового ступеня  
кандидата хімічних наук

Київ - 1996

#6

ЛНБ України ім.В.Стефаніка



00752966 (Z)

Дисертацією є рукопис.

Робота виконана в Інституті біоорганічної

Науковий керівник: доктор медичних наук,  
професор О.І. Луйк

Офіційні опоненти: доктор хімічних наук  
Гулько В.М.

доктор біологічних наук,  
професор Кібірев В.К.

Провідна організація - Інститут органічної хімії НАН України.

Захист відбудеться 27 вересня 1996 р. на засіданні спеціалізованої вченої ради  
Д 01.80.01 в Інституті біоорганічної хімії та нафтохімії НАН України (253660,  
Київ, вул. Мурманська, 1).

З дисертацією можна ознайомитись у бібліотеці Інституту біоорганічної хімії та  
нафтохімії НАН України (253660, Київ, вул. Мурманська, 1).

Автореферат розісланий "27" вересня 1996 р.

ЛНБ ім. В. Стефаніка  
АН України

Вчений секретар  
спеціалізованої вченої ради

Д.М. Федоряк

**Актуальність** У 1990 році було офіційно опубліковано статистику щодо створення нових лікарських препаратів провідними фармацевтичними фірмами світу. Згідно з цими даними, процес отримання лише одного комерційного препарату займає 8-12 років. За цей час синтезується та випробовується у медико-біологічних експериментах понад 10000 речовин. Вартість одної розробки складає в середньому 500-800 млн. доларів. На сьогоднішній день ситуація у цій галузі суттєво не змінилася.

Це, безперечно, свідчить про неефективність існуючих методів комп'ютерного прогнозування фізіологічної активності нових сполук на основі їх хімічної будови та розрахункового конструювання лікарських речовин (drug design). Вищезначене зумовлює необхідність вдосконалення таких підходів.

Поряд зі згаданим, переважно практичним, аспектом існує й більш фундаментальний. Він стосується подальшої розробки раціональної класифікації ліків, що була запропонована проф. Луйком у 1988 році. На вищому щаблі ієрархічної класифікації переважна більшість лікарських речовин розподіляється на дві великі групи, що якісно відрізняються за впливом на головні сигнальні системи клітин [О.І.Луйк зі співавторами, 1988, В.П. Кухар, О.І.Луйк, 1992].

На сьогоднішній день найбільш актуальною проблемою подальшого розвитку даної класифікаційної концепції є конкретизація будови загальних фармакофорних рис, що притаманні лікарським речовинам кожної з вищезгаданих груп. Треба зазначити, що розкриття таких найбільш загальних структурних ознак має не тільки теоретичну, а й практичну значимість, бо їх наявність треба враховувати при конструюванні лікарських речовин будь-якого конкретного типу дії.

**Ступінь дослідженості тематики** Вибір найбільш інформативних параметрів є одною з найважливіших і найменш досліджених проблем при побудові моделей “структура - активність”. В той же час дана робота є продовженням робіт по підтвердженню раціональної класифікації лікарських речовин. Попередні роботи довели тільки сам факт наявності спільних рис хімічної будови, що об’єднують речовини кожної з двох наведених класифікаційних груп та відрізняють ці групи одна від одної. Проте, в чому конкретно полягають ці риси, так і не було розкрито.

**Метою роботи** було вдосконалення існуючих методів комп’ютерного прогнозування фізіологічної дії речовин, виходячи з їх хімічної будови, та пошуку найбільш загальних фармакофорних ознак з подальшою перевіркою розроблених підходів на задачах різного ступеню складності.

#### **Основні завдання дослідження**

1) Розробка нових та вдосконалення існуючих методів кодування будови молекул за допомогою топологічних індексів.

2) Створення систем для вибору найбільш інформативних параметрів та розрахункового прогнозування фізіологічної дії речовин, виходячи з їх хімічної структури.

3) Перевірка ефективності розроблених методів на задачах прогнозування активності інгібіторів 5-ліпоксигенази, альдегіддегідрогенази, альдозоредуктази та наркотичних речовин.

4) Пошук найбільш загальних рис будови лікарських речовин, об’єднаних за принципами раціональної ієрархічної класифікації.

**Наукова новизна** Запропоновано ряд нових топологічних індексів. Вперше застосовано еволюційну стратегію в комбінації з методом “к-найближчих сусідів” для пошуку найкращих наборів параметрів. Побудовано ефективні класифікатори речовин на активні та неактивні для інгібіторів 5-

ліпоксигенази, альдозоредуктази, альдегіддегідрогенази та наркотиків. Виявлено топологічні та геометричні фармакофорні фрагменти, характерні для двох основних груп речовин по типу дії на основні сигнальні системи клітини, що підтверджує справедливість використаних підходів до класифікації фізіологічно активних речовин та відкриває нові перспективи для їх конструювання.

**Теоретична цінність** Отримані нові розрахункові підтвердження доцільності побудови раціональної ієрархічної класифікації лікарських речовин, що базується на їх первинній дії на головні системи клітинної сигнальної трансдукції. Конкретизовані найзагальніші фармакофорні риси речовин, об'єднаних за вказаним принципом. Останнє свідчить про те, що кожен конкретний тип фармакологічної дії реалізується шляхом сполучення специфічних рис будови лікарської речовини з більш загальними базовими.

Вперше показано, що в будові речовин, які викликають наркотичну залежність, незважаючи на їх широку різноманітність існують спільні принципи будови.

**Практична цінність** Описані методи дозволяють швидко будувати класифікатори та отримувати прогнозуючі системи для любых видів біологічної та інших активностей хімічних речовин. Практичне значення мають також розроблені в рамках даної роботи системи пошуку 2-х та 3-х мірних фармакофорів.

**Рівень реалізації** Всі дослідження проведено за допомогою комплексу програм для персонального комп'ютера, який може і далі використовуватися для подібних досліджень.

**Апробація роботи** Основні результати роботи доповідалися на Всеросійській науковій конференції "Створення лікарських препаратів" (Москва, 1992); 12-й щорічній конференції товариства молекулярної графіки

(Швейцарія, 1993); на третій щорічній конференції по еволюційному програмуванню "EP - 94" (Сан - Дієго, 1994); на американсько - японському семінарі по альдозоредуктазі (Гаваї, 1994); 10-у Європейському симпозиумі по дослідженню зв'язку структура - активність (Барселона, 1994); 12-у міжнародному конгресі по фармакології (Монреаль, 1994); 1-й Європейській конференції по комп'ютерній хімії (Франція, 1994).

Публікації Основні результати роботи викладено в 5 статтях та 7 тезах доповідей.

Структура та обсяг роботи Дисертація складається із вступу, огляду літератури, 3-х розділів, висновків, списку літератури та додатку. Робота викладена на 170 аркушах машинописного тексту (з додатком) і ілюстрована 5 малюнками та 10 таблицями. Список літератури нараховує 148 джерел.

Конкретний особистий внесок дисертанта Всі основні результати отримані автором.

Методологія Комп'ютерне моделювання зв'язку між структурою та біологічною активністю хімічних сполук.

## *1. Класифікація молекул за допомогою методу k-найближчих сусудів.*

### *1.1 Параметри молекул*

В даному дослідженні будова молекул була описана за допомогою їхніх топологічних індексів. Топологічні індекси утворюються шляхом перетворення хімічного графа (математичний аналог хімічної формули) в число за певним алгоритмом.

Індекси, котрі можуть бути розраховані за допомогою нашої програми Ind, поділяються на такі групи:

- Індокси, що базуються на матриці відстаней та матриці суміжності. Недоліком таких індоксів є те, що вони не враховують різних типів атомів та зв'язків;

- Індокси, що базуються на матриці зв'язків, запропонованій Барішем для молекул, які містять гетероатоми. Елементи модифікованої матриці суміжності  $a'_{ij}$  обчислюються за таким правилом:

$$A'(i, j) = \frac{1}{n} \frac{36}{Z_i Z_j} A(i, j) \quad Z_i \text{ та } Z_j \text{ кількість}$$

електронів в атомах  $i$  та  $j$ , що сполучені даним зв'язком;  $n$  - являє собою порядок зв'язку (1- для простого зв'язку, 2-для подвійного, 3- для потрійного і 3/2 для ароматичного). Елементи матриці відстаней  $D'(G)$  визначаються як сума елементів матриці суміжності, що відповідають зв'язкам, які лежать на

відповідному найкоротшому шляху:  $D'(i, j) = \sum A'(k, l)$   $k$  та  $l$  у цій формулі позначають пару суміжних атомів, що лежать на найкоротшому шляху між вершинами  $i$  та  $j$ . Якщо існує кілька шляхів однакової довжини, то перевага віддається шляху з меншою сумою;

- Індокси молекулярних зв'язків, запропоновані Холлом та Кіром. Цей метод напряму вводить електронні характеристики в молекулярні графи. Атоми характеризуються їхніми простим та валентним  $\delta$  числами. Просте число  $\delta$  визначається як різниця між загальним числом електронів атома, що знаходяться на орбіталях сигма - типу та числом електронів, що утворюють сигма зв'язки з атомами водню. Це число є електронною інтерпретацією степеня вершини з теорії графів. Валентне число  $\delta_i^v$  обчислюється за такими правилами:  $\delta_i^v = Z_i^v - h_i$  для атомів другого періоду, та  $\delta_i^v = (Z_i^v - h_i) / (Z - Z^v - 1)$  для атомів інших періодів.  $Z_i^v$  - кількість валентних електронів скелетного атома  $i$ ,  $h_i$  - кількість зв'язаних з ним атомів водню,  $Z$  - атомний номер;

- Індeksi молекулярної форми  $k$  (капша) 1-го, 2-го та 3-го порядків а також модифіковані індeksi  $k_\alpha$ , до яких внесено поправку на тип атома  $\alpha$ :

$$\alpha = \sum (r_x / r_{\text{ср}}) - 1, \text{ де } r - \text{ковалентні радіуси};$$

- Статистика по різних атомах та групах( кількість певних атомів, або груп в молекулі);
- Параметри, розраховані на основі тримірної структури молекул - 3-х мірний індeкс Вінера, функції дисиметрії та ін.

### *1.2 Методи вибору найінформативніших параметрів.*

Загальна кількість топологічних та інших параметрів, які розраховувалися за допомогою програми Ind, складала 277, тому велика увага в даній роботі приділялася вибору найбільш інформативних параметрів. Використання великої кількості малоінформативних величин значно погіршує прогноз біологічної активності. З самого початку, як правило, не відомо, які з параметрів найкраще описують залежність структура - активність, хоч велика кількість і різноманітність індeксів, за допомогою яких представлена молекула, дозволяє надіятися, що деякі з них відображають властивості, відповідальні за проявлення даної активності. В нашій роботі для вибору параметрів застосовувалося кілька евристичних стратегій:

- Поступове вдосконалення набору параметрів. На кожному кроці розглядається кожен із параметрів і або відкидається від існуючого набору (якщо цей параметр до нього входить), або додається до нього. Якщо така зміна призводить до покращення прогнозу, вона залишається в силі, якщо ні - то відкидається. Робота алгоритму припиняється, якщо на певному кроці, після розгляду всіх параметрів, не було досягнуто покращення;
- Еволюційна стратегія (ЕС) - найсучасніший метод, що моделює біологічний відбір, основні закони якого були сформульовані Дарвіном. Цей

метод моделює еволюцію систем через процеси відбору та відтворення, які залежать від пристосованості індивідів до середовища. В загальних рисах алгоритм виглядає так:

а) Створити випадковим чином початкову популяцію (групу наборів параметрів) заданого розміру  $m$ ;

б) Оцінити пристосованість кожного індивіда (якість набору параметрів);

в) З застосуванням до кожного із індивідів початкової популяції операторів пошуку (генетичних операторів), таких як рекомбінація і мутація, отримати нове покоління. Отримується приблизно  $m * 7$  потомків попередньої популяції, причому більш пристосовані батьки дають більше потомків;

г) Оцінити пристосованість кожного індивіда нового покоління, і з них, а також старих індивідів вибрати  $m$  найкращих, які вижили. Повернутися до кроку в.

Породження нових поколінь продовжується поки не буде досягнуто заданого результату, або не буде перевищено задане число поколінь. Таким чином здійснюється адаптивний пошук найкращих індивідів. У нашому випадку індивід - це набір параметрів, а генетичний оператор (мутація) - це додавання до набору або відкидання одного із параметрів. Оскільки така робота вимагає величезного обсягу обчислень, оцінка якості наборів проводилася за допомогою швидкого класифікатора, відомого як "к-найближчих сусідів".

Пристосованість вираховувалася за формулою:

$$f = \frac{n_c}{N} - \beta Q, \text{ де } n_c - \text{кількість вірно класифікованих молекул, } N -$$

загальна їх кількість,  $Q$  - кількість параметрів в наборі,  $\beta$  - вага параметра (звичайно  $\beta = 0.001$ , але при перевищенні обмеження на кількість параметрів в наборі  $\beta = \infty$ ). Другий член введено, щоб забезпечити перевагу меншим наборам;

Нова двоступенева процедура, що є комбінацією методу часткових найменших квадратів (ЧНК, PLS) та еволюційної стратегії. На першій стадії за допомогою методу часткових найменших квадратів із початкового набору параметрів утворюється менший набір ортогональних латентних змінних, що в максимальній мірі корельовані з молекулярною активністю. На другій стадії за допомогою ЕС вибирається найкращий набір латентних змінних.

В нашій роботі для класифікації молекул на активні та неактивні використовувався метод “к-найближчих сусідів” в комбінації з переліченими вище методами пошуку наборів параметрів. Цей метод включав оригінальну методику перехресної оцінки (cross-validation), що полягала у відкиданні від двох до половини молекул навчального набору і оцінці набору параметрів за результатами класифікації відкинутих молекул. Кількість таких перевірок підбиралася такою, щоб забезпечити задану точність проценту правильно класифікованих молекул на основі статистичного критерію:

$$\hat{P}_e = \frac{n_c}{n_{all}}$$

$$\Delta \hat{P}_e = \pm 1.96 \cdot \sqrt{\frac{\hat{P}_e \cdot (1 - \hat{P}_e)}{n_{all}}}$$

$n_c$  - кількість вірно зпрогнозованих молекул в результаті  $n_{all}$  відкидань.

Перевірки припиняються коли  $\Delta \hat{P}_e$  стане меншим заданої величини.

Навчальними наборами молекул фізіологічно активних речовин, що застосовувалися для вибору найбільш інформативних параметрів та класифікації молекул, були відповідні банки даних, складені для кожної конкретної задачі (див. 1.3 - 1.6) на основі літературних джерел.

### 1.3 Прогнозування активності гідроксаматів як інгібіторів

#### 5-ліпоксигенази.

Надмірний біосинтез таких надзвичайно активних речовин, як простагландини та інші ейкозаноїди, відіграє важливу роль у патогенезі запалень, бронхіальної астми, інших алергій, інфаркту міокарда та інших хвороб. Через це велике значення має розрахункове прогнозування активності інгібіторів одного із ключових ферментів у біосинтезі лейкотриєнів - 5-ліпоксигенази(5-LO) як потенційних засобів лікування перелічених захворювань.

Ми проаналізували 44 хімічні речовини, вибрані на основі літературних даних. Вони були розділені на три групи - активні речовини ( $IC_{50} < 10$   $\mu$ моль), неактивні речовини ( $IC_{50} > 10$   $\mu$ моль) і контрольна група речовин для перевірки прогнозуючої здатності методу. Групи нараховували 14, 22 та 8 речовин. В результаті застосування методу "к-найближчих сусідів" з вибором параметрів за методом поступового вдосконалення їх набору було знайдено, що з застосуванням лише 3-х параметрів вдається вірно класифікувати всі 36 речовин навчального набору, а також вірно передбачити активність 7 із 8 речовин контрольного набору. Два із цих параметрів є логарифмами суми модулів коефіцієнтів характеристичного полінома матриці суміжності поділеними на кількість вершин у молекулярному графі. Один із цих індексів розраховується для матриці, що враховує гетероатоми. Третій параметр - це індекс відстаней між вершинами в графі, поділений на кількість вершин у молекулярному графі:

$$GDI'(G) = \frac{1}{N} \sum (g_i)^2, \text{ де } g_i - \text{кількість повторень}$$

найкоротшої довжини  $i$  в графі,  $N$  - кількість вершин у молекулярному графі.

Застосувавши для пошуку параметрів еволюційну стратегію вдалося отримати такий же результат з використанням лише двох параметрів:

$$h\chi'(G) = \sum (v'_1 v'_2 \dots v'_{h+1})^{-1/2}, \text{ де сума по найкоротших}$$

шляхах молекулярного графа,  $v'_i$ - модифіковані степені вершин, що лежать на даному шляху.

Індекс молекулярних звязків 0-го порядку:

$${}^0\chi = \sum_{i=1}^n \delta_i^{-1/2} / n$$

$\delta_i$  - просте число дельта  $i$  - го атома,  $n$ -кількість атомів у молекулі (кількість вершин у молекулярному графі).

#### 1.4 Інгібітори альдегіддегідрогенази.

В основі дії антиалкогольних препаратів сенсibiliзуючого типу лежить пригнічення активності одного з ключових ферментів основного шляху метаболізму етанолу - альдегіддегідрогенази (альдегід : NAD-оксидоредуктаза, АльДГ). Хоча поряд з відомими лікарськими препаратами (тетраетилтіурамдисульфід (дисульфірам); карбімід кальцію) існує ще ряд речовин, що пригнічують активність цього ферменту, проблеми токсичності та підвищення селективності у відношенні до АльДГ роблять актуальним цілеспрямований пошук нових інгібіторів цього ферменту. Серед потенційних інгібіторів розглядаються речовини різної природи - тіурамдисульфіди, тіокарбонілдисульфіди, дигіопіримідини, амід бензойних кислот, аміни та амід аліфатичних та ароматичних кислот, що містять циклопропілетильний радикал. Застосувавши метод "к-найближчих сусідів" з оптимізацією ознакового простору за допомогою еволюційної стратегії, вдалося знайти оптимальний набір з 5-ти параметрів для передбачення можливості зниження

активності АльДГ матриксу мітохондрій новими сполуками вищезначених типів:

- Узагальнений індекс молекулярних зв'язків Кіра, поділений на кількість вершин у молекулярному графі:

$$h\chi_n(G) = \frac{1}{N} \sum (v_1 v_2 \dots v_{h+1})^{-1/2}$$

сума по всіх шляхах даного молекулярного графа,  $v_i$ - степені вершин, що лежать на даному шляху;

- Аналог першого індексу, розрахований з використанням модифікованої матриці суміжності, котрий враховує тільки найкоротші шляхи;
- Сольватаційний топологічний індекс 1-го порядку, який описує дисперсійні взаємодії у розчині:

$$1_{\chi s} = (1 / 4) \sum Z_i Z_j / (\delta_i \delta_j)^{1/2}$$

$Z_i$  - коефіцієнт, що характеризує розмір атома  $i$ . Цей коефіцієнт рівний номеру періода, в якому знаходиться атом,  $\delta_i$  - ступень  $i$ -ї вершини. Сума по всіх парах суміжних вершин;

- Функція диссиметрії, що не враховує тип атома. Для розрахунку даного параметра необхідно знати реальну геометрію молекули, яку розраховували з використанням методу MNDO пакету MOPAC;
- Молекулярний парахор, розрахований по адитивній схемі.

При застосуванні даного набору параметрів вдається вірно зпрогнозувати активність 51 з 55 (92.7%) молекул навчального набору. При перехресній оцінці з відкиданням 22-х молекул(40%) процент вірних прогнозів зменшується лише до 82.5%, що вказує на високу стабільність прогнозів.

### 1.5 Залежність наркотичної дії речовин від їхньої структури.

Вибірка складалася з 28 наркотиків та 21 речовини, що не викликає наркотичної залежності, але близька за будовою до наркотиків. З 210 розрахованих параметрів на попередньому етапі було вибрано 37, попарні коефіцієнти кореляції яких не перевищували 0.8. В результаті оптимізації простору за допомогою еволюційної стратегії знайдено набір параметрів, що забезпечував вірне передбачення активності для 46 молекул із 49. При перехресній оцінці з відкиданням 10 молекул процент вірних прогнозів зменшився лише до 90,7%. Цей набір складався з таких параметрів:

кількість атомів в ароматичних циклах; кількість гідроксильних груп; кількість сульфонових груп S=O; індекс молекулярної форми 3-го порядку; валентний індекс молекулярних зв'язків 1-го порядку:

$${}^1\chi^v = \sum (\delta_i^v \delta_j^v)^{-1/2}, \text{ де } \delta_i^v \text{ та } \delta_j^v - \text{валентні дельта}$$

атомів, що зв'язані хімічним зв'язком. Сума по всіх парах зв'язаних атомів.

Таким чином показано, що незважаючи на широку різноманітність структур наркотиків, їм притаманні певні спільні риси, які, можливо, і зумовлюють розвиток наркотичної залежності. Це сприятиме новому розумінню механізмів розвитку звикання та пошуку антинаркотичних засобів.

### 1.6 Вивчення залежності структура - активність інгібіторів

#### альдозоредуктази.

Фермент альдозоредуктаза, що відіграє важливу роль в метаболізмі глюкози, є одним із важливих факторів етіології та патогенезу діабету. Тому великі зусилля направлені на пошук, у тому числі і комп'ютерний, інгібіторів цього ферменту. При дослідженні залежності структура - активність інгібіторів альдозоредуктази нами вперше застосовано метод ЧНК у

комбінації з еволюційною стратегією. Застосування ЧНК зменшує кількість параметрів і дозволяє отримувати взаємно ортогональні і більш корельовані з активністю параметри. Цей метод дає результати подібно методу головних компонент з тією різницею, що отримані латентні змінні не просто моделюють незалежні змінні  $X$  (у нашому випадку набори із 152 топологічних індексів для кожної молекули), а й корелюють з  $Y$  (молекулярною активністю). Для дослідження залежності структура - активність було використано 224 речовини у навчальному наборі і 30 речовин у контрольному. Для активних молекул ми приписали  $y = 1$ , а для неактивних  $y = 0$  і застосували ЧНК для отримання 30 латентних змінних. Латентні змінні використовувалися як вхідні параметри для вибору параметрів за допомогою еволюційної стратегії. Найкращий набір параметрів складався з 6 латентних змінних: 2, 3, 5, 9, 18 та 27. Активність 191(85,2%) із 224 речовин навчального набору була передбачена вірно. На контрольному наборі результат був навіть кращим - 27(90%) із 30.

## *2. Методи пошуку фармакофорів.*

Хоч описані вище методи дозволяють будувати класифікатори, що забезпечують досить високий процент правильних прогнозів, вони не дають саме хімічної інформації про молекули різного типу дії (не дають можливості знаходити фармакофор). Для пошуку 2-х і 3-х мірних фармакофорів застосовувалися спеціально розроблені методи. Для пошуку 2-х мірних фармакофорів розроблено близьку до відомих методів ФКСП(фрагментарний код суперпозиції підструктур) та МСДЦ(матриця суміжності дескрипторних центрів) систему представлення молекул, яка базується на типах атомів у силовому полі ММ2. Молекула є матрицею, діагональними елементами якої є дескрипторні центри(ДЦ) - атоми типів передбачених ММ2, або прості цикли розміром від 3-х до 9-ти членів (можуть бути ароматичними), а

недіагональними - набори (до 5) найкоротших відстаней між відповідними ДЦ. Пошук фармакофорів розділяється на пошук лінійних та 2-х мірних фрагментів. У першому випадку просто розглядається статистика по всіх типах трійок (2 ДЦ + відстань) присутніх в молекулах обох типів. У другому застосовується досить складний метод пошуку 2-х мірних фрагментів. Головним елементом цього методу є пошук клік графа сумісності двох структур. Вершинами такого графа є пари еквівалентних ДЦ. Ребра відповідають парам центрів, відстань між якими однакова. Клікою графа називається максимальний повний підграф, тобто підграф, усі вершини якого зв'язані ребрами. Таким чином, кожна кліка графа сумісності відповідає максимальній спільній частині структур, що порівнюються. Загальна схема пошуку 2-х мірних фармакофорів передбачає виконання для кожної молекули обох класів наступних дій:

- Порівняння з усіма іншими молекулами обох типів та знаходження максимальних спільних елементів;
- Для кожного виділеного фрагмента перевіряється, чи входить він до складу всіх інших структур, крім тих, у яких він був виділений. Така перевірка виконується також шляхом побудови графа сумісності і знаходження його клік. У цьому випадку розмір кліки повинен дорівнювати розміру фрагмента. Здійснюється оцінка фрагмента за одним із статистичних критеріїв, наведених нижче, і, якщо він задовольняє певним вимогам, то включається до складу фармакофорів.

Для оцінки статистичної значущості фармакофорів застосовувалися два критерії:

- Байєсівська оцінка імовірності того, що сполука, яка має даний фрагмент, буде відноситись до класу сполук  $A_k$ :

$$P(A_k / f_r) = \frac{m_{rk} + 1}{m_r + 2},$$

де  $m_{rk}$  - кількість сполук, що мають даний фрагмент і належать до класу  $A_k$

$m_r$  - загальна кількість сполук, що мають даний фрагмент;

- коефіцієнт кореляції якісних ознак Юла:

$$\alpha = \frac{n_1 n_4 - n_2 n_3}{\sqrt{N_1 N_2 N_3 N_4}},$$

де  $n_1$  - кількість сполук, що мають даний фрагмент і належать до класу  $A_k$ ,  $n_2$  -

кількість сполук, що не мають даного фрагмента і належать до класу  $A_k$ ,  $n_3$  та

$n_4$  - відповідні величини для всіх сполук, що не належать до класу  $A_k$ ,  $N_1$ ,  $N_2$  -

кількості сполук, що належать і не належать класу  $A_k$ ,  $N_3 = n_1 + n_3$ ,  $N_4 = n_2 +$

$n_4$ .

Ці ж самі критерії застосовувалися для оцінки лінійних фрагментів, а також фрагментів, отриманих електронно-топологічним методом.

Описані методи було застосовано для класифікації лікарських препаратів по типу їх дії на сигнальні системи клітини. В інституті під керівництвом проф. О.І.Луїка інтенсивно проводяться дослідження з розробки раціональних підходів до класифікації лікарських препаратів. На першому шаблі ієрархічної класифікації пропонується розподіл речовин на два основні класи - "+1/-2" - ксенобіотики, які підсилюють сигнали, що проходять через аденілатциклазну систему та пригнічують передачу сигналів через фосфоліпідну сигнальну систему, та "-1/+2" - речовини, які мають протилежний тип дії. В результаті аналізу 123 речовин першої групи та 69 речовин другої виділено 101 фармакофорний фрагмент (45 2-х мірних та 56 лінійних). За допомогою поступової оптимізації цей набір було скорочено до 22-х фрагментів, за допомогою яких можна правильно визначити активність 147 із 192 сполук ( 105 із 123 у класі "+1/-2" та 42 із 69 у класі "-1/+2", при цьому кількість невірно класифікованих сполук була відповідно 3 і 11, а

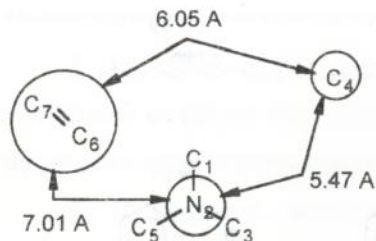
кількість сполук, активність яких визначити не вдалося, 15 і 16). Для прогнозування активності застосовувався найпростіший метод "голосування ознак": молекула відносилася до певного класу, якщо вона мала більше ознак, характерних для цього класу, ніж ознак характерних для іншого класу. Виділені фрагменти наведені на Мал. 2.

Наведений вище метод знаходження спільних частин молекул застосовано у новій модифікації електронно-топологічного методу, запропонованій автором. Цей метод вперше розроблений Дімогло, у найпростішому варіанті розглядає молекулу як матрицю, діагональними елементами якої є заряди атомів, а недіагональними - відстані між атомами. Застосування способу знаходження спільних фрагментів як клік графа сумісності і використання для цього одного з найкращих алгоритмів такого типу [Рейнгольд, 1979] дозволило значно прискорити розрахунки і розглядати значно більший обсяг даних. Загалом алгоритм роботи ЕТМ дещо відмінний, що зумовлено більшою складністю роботи з 3-х мірними структурами молекул:

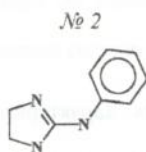
- квантово-хімічні розрахунки структур і вибір конформацій;
- вибір контрольної молекули, з якою будуть порівнюватися всі інші молекули. Це або найактивніша молекула, або молекула з найбільш дослідженими конформаціями;
- парне порівняння усіх молекул з контрольною і виділення спільних фрагментів з використанням пошуку клік графа сумісності. Фрагменти заносяться в базу даних і по них набирається попередня статистика;
- перевірка менших фрагментів на входження в більші, яка проводиться також за допомогою графа сумісності. В результаті набирається повна статистика по фрагментах. Тепер їх можна оцінювати за наведеними вище статистичними критеріями;

- статистична оцінка параметрів та вибір серед них статистично достовірних.

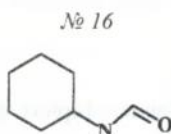
За допомогою ЕТМ було виявлено фармакофор, характерний для молекул класу "+1/-2".



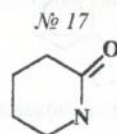
Мал. 1. 3-мірний фармакофор, характерний для молекул класу "+1/-2" (знайдено в 20 молекулах). Атоми вуглецю 4, 6, 7, як правило, належать ароматичним циклам.



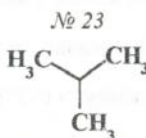
0 - 4



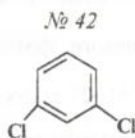
0 - 4



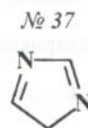
18 - 2



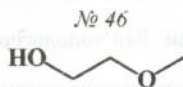
0 - 7



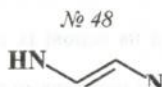
3 - 11



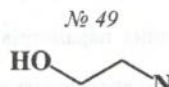
4 - 12



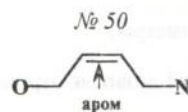
0 - 6



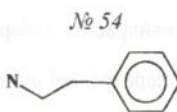
1 - 7



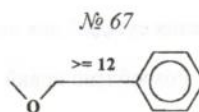
10 - 0



0 - 6



1 - 7

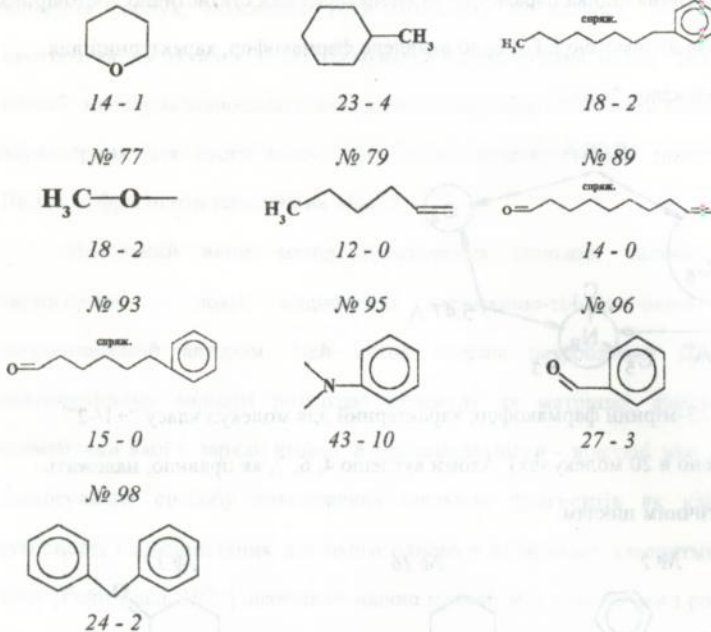


&gt;= 12

№ 70

№ 71

ЛНБ ім. В. Стефаника  
АН України  
№ 76



Мал. 2. Оптимізований набір фармакофорних фрагментів, характерних для молекул класів "+1/-2" та "-1/+2". Для кожного фрагменту наведено його номер у початковому наборі, а також кількість молекул обох класів, у яких він присутній (наприклад, для останнього фрагменту 24-2 означає, що він присутній в 24-х молекулах класу "+1/-2" та 2-х молекулах класу "-1/+2").

## ВИСНОВКИ

1. Створено програму для обчислення великої кількості топологічних та геометричних параметрів молекул на основі їх структури. Ряд топологічних індексів, що враховують відмінності між типами атомів використано вперше.
2. Застосовано еволюційну стратегію в комбінації з методом "к-найближчих сусідів" для пошуку найкращих наборів параметрів.
3. Розроблено новий метод перехресної оцінки, який дозволяє отримати статистично достовірну оцінку прогностичної здатності наборів параметрів.

4. Наведені вище методи використано для вирішення реальних задач пошуку зв'язку структура - активність. Побудовано ефективні класифікатори речовин на активні та неактивні для інгібіторів 5-ліпоксигенази, альдозоредуктази, альдегіддегідрогенази та наркотиків.

5. На основі пошуку клік графа сумісності розроблено власну модифікацію електронно-топологічного методу (ETM) та систему пошуку 2-х мірних (топологічних фармакофорів).

6. В результаті застосування ETM та пошуку топологічних фармакофорів до групи речовин, що по різному діють на основні сигнальні системи клітини, виявлено один 3-х мірний та 22 топологічних фармакофорних фрагмента, які цілком задовільно описують належність молекул до одної з двох основних груп речовин по типу дії на основні сигнальні системи клітини.

*В.В.В.*

## СПИСОК РОБІТ, ОПУБЛІКОВАНИХ ПО ТЕМІ ДИСЕРТАЦІЇ.

1. Пода Г.И., Танчук В.Ю., Кошель М.И., Луйк А.И. Применение топологических индексов для прогнозирования 5-липноксигеназной ингибирующей активности в ряду гидроксаматов. // Теор. и эксп. химия. - 1993. - т. 29, № 2. - С. 122-125.
2. Tetko I. V., Tanchuk V. Yu., Chentsova N. P., Antonenko S. V., Poda G. I., Kukhar V. P., Luik A. I. HIV-1 Reverse Transcriptase Inhibitor Design Using Artificial Neural Networks. // J. Med. Chem. - 1994. т. 37, № 16 С. 2520-2526.
3. Tetko I. V., Tanchuk V. Yu., Luik A. I. Application of an Evolutionary Algorithm to the Structure-Activity Relationship. // Proceedings of the Third Annual Conference on Evolutionary Programming. A.V. Sebald and L.J. Fogel (eds.), World Scientific, River Edge, NJ, 1994, pp. 109-119.
4. Tanchuk V. Yu., Mogilevich S. E., Tetko I. V., Poda G.I., Luik A. I. Application of the evolutionary strategy for the investigation of structure-activity relationships in the aldose reductase inhibitors. // Доповіді академії наук України. - 1994, № 12, С. 140-143.
5. Tetko И.В., Танчук В.Ю., Васильев С.А., Хиля В.П., Пода Г.И., Луйк А.И. Эволюционное программирование для выявления закономерностей "Структура - активность" в ряду производных 3-феноксихромена и 3-фенокси-4-гидроксикумарина. // Биоорганическая химия. - 1995. т. 21, № 10 С.809 - 815.
6. Tetko I. V., Tanchuk V. Yu., Vasiliev S. A., Luik A. I. Application of an evolutionary algorithm to the structure-activity relationship studies in flavonoids. // Abstracts of 10th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modelling. September 4-9, 1994, Barcelona (Spain). A188.
7. Tanchuk V. Yu., Tetko I. V., Vinogradova K.G., Luik A. I. Application of the evolution strategy (ES) to the investigation of aldose reductase and alcohol dehydrogenase inhibitors. *ibid.*, B208.
8. Tanchuk, V.Yu., Kislova, O.V., Koshel, M.I, Luik A.I. On correlation between IC<sub>50</sub> of aldehyde dehydrogenase inhibitors and topological indexes K<sub>50</sub> and t. *ibid.*, B209.
9. Luik A.I., Tanchuk V.Yu., Poda G.I. Determination of structural moieties relevant to the interaction of molecules with cellular signalling systems. *ibid.*, B234.
10. Tetko, I. V.; Tanchuk, V. Yu. Anti AIDS Drug Design with the Help of Neural Networks. // Canad. J. Physiol. Pharmac. (Abstracts of XIII<sup>th</sup> International Congress of Pharmacology, Montreal, Canada, 24-29 July, 1994) 1994, т. 72, С.214.
11. Poda G.I., Dimoglo A.S., Tanchuk V.Yu. Luik A.I. Structure - activity investigations of the two sets of molecules similarly influencing the main cellular signalling systems. // AIP Conference proceedings 330. E.C.C.C. 1 Computational Chemistry, F.E.C.S. Conference, Nancy, France 1994. Bernardi F., Rivail J-L. (eds.) Woodbery, New York. p. 791.
12. Poda G.I., Tanchuk V.Yu. Luik A.I. Statistical view to the pattern recognition. *ibid.*, p. 632.

## АННОТАЦИЯ

Танчук В.Ю. Поиск наборов дескрипторов, повышающих эффективность прогнозирования физиологической активности веществ, исходя из их структуры. Рукопись.

Диссертация на соискание ученой степени кандидата химических наук по специальности - 02.00.10 -биоорганическая химия, химия природных и физиологически активных веществ. Институт биоорганической химии и нефтехимии НАН Украины, Киев, 1996.

Защищается 12 работ, в которых описаны новые методы выбора параметров в задачах "структура - активность", приведены примеры их использования, а также определены фармакофорные фрагменты, характерные для двух групп физиологически активных веществ, одинаково влияющих на клеточные сигнальные системы.

## SUMMARY

Tanchuk V. Yu. The search for the sets of descriptors that make predictions of physiological activity from chemical structure more efficient. Manuscript.

Ph.D. thesis in speciality 02.00.10 - Bioorganic chemistry, chemistry of natural and physiologicaly active compounds.

12 scientific publications are defended. They include description of new methods for parameter selection in SAR(Structure-activity relationship) problems, the examples of their application and determination of pharmacophoric patterns of the two groups of physiologicaly active compounds, similarly influencing the main cellular signalling systems.

**Ключові слова:** структура - активність, QSAR, клітинні сигнальні системи, пошук фармакофорів, еволюційна стратегія.

*В.Ю. Танчук*









AB 35.356

