



КУЗЬМІН

Віктор Євгенович — член-кореспондент НАН України, директор Фізико-хімічного інституту ім. О.В. Богатського НАН України

ХЕМОІНФОРМАТИКА ЯК ЕФЕКТИВНИЙ ІНСТРУМЕНТ ПРОГНОЗУВАННЯ І КОНСТРУЮВАННЯ НОВИХ РЕЧОВИН З КОМПЛЕКСОМ НЕОБХІДНИХ ВЛАСТИВОСТЕЙ

**Стенограма доповіді на засіданні
Президії НАН України 21 вересня 2022 року**

Доповідь присвячено результатам фундаментальних та прикладних наукових досліджень у галузі хемоінформатики, проведених у Фізико-хімічному інституті ім. О.В. Богатського НАН України і пов'язаних з аналізом зв'язків «структура — властивості» хімічних речовин з метою цілеспрямованого створення нових сполук і матеріалів з комплексом заданих властивостей.

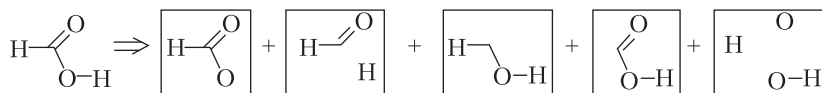
Вельмишановний Анатолію Глібовичу!

Шановні члени Президії!

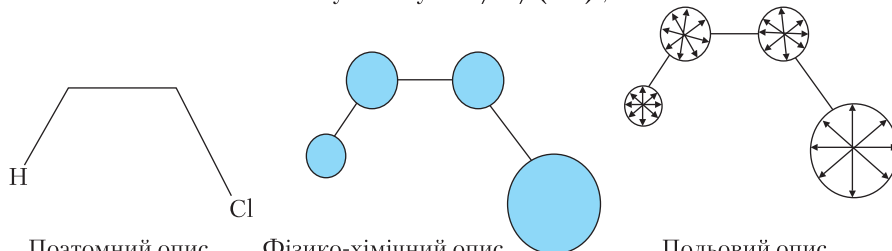
Пропоную до вашої уваги результати досліджень у галузі хемоінформатики, отримані останніми роками у Фізико-хімічному інституті ім. О.В. Богатського НАН України.

Приблизно 40 років тому виникла нова наукова дисципліна — хемоінформатика, в якій методами комп'ютерних технологій та інформатики вирішують найрізноманітніші хімічні проблеми. Ключові з них — це, зокрема, прогнозування властивостей сполук та конструювання нових речовин — потенційних реагентів, каталізаторів, лікарських засобів, нових матеріалів тощо, з комплексом необхідних властивостей.

Предметом хемоінформатики є побудова баз даних хімічних сполук та реакцій, прогнозування фізичних, хімічних та біологічних властивостей сполук і матеріалів, пошук нових лікарських препаратів, аналіз спектральної інформації, передбачення перебігу хімічних реакцій, планування органічного синтезу, розроблення підходів до зберігання, вилучення та оброблення хімічної інформації. Інакше кажучи, хемоінформатика шукає



Кількість симплексів у молекулі: $n!/4! / (n-4)!$, n — кількість атомів



Поатомний опис

- Індивідуальність атомів (мітки)

Фізико-хімічний опис

- Атомні заряди
- Вандерваальсовий об'єм
- Ліпофільність
- Поляризованість
- Донори/акцептори Н-зв'язку

Польовий опис

- Електростатика
- «Стерика»
- Інформація

Рис. 1. Симплексне представлення молекулярної структури (SiRMS)

відповіді на такі фундаментальні запитання: яка хімічна сполука має цільову властивість? як синтезувати цільову хімічну сполуку? яку хімічну сполуку буде отримано в результаті реакції?

Один із напрямів хемоінформатики пов'язаний з прогнозуванням властивостей речовин на основі аналізу зв'язку «структура — властивості». Це побудова так званих моделей QSAR/QSPR (quantitative structure activity/property relationship): QSAR — якщо йдеться про біологічну активність, а QSPR — для будь-якої властивості. Спочатку QSAR був результатом спроби виразити відносну величину біологічної властивості в термінах набору фізичних параметрів, що характеризують гідрофобні, електронні та стеричні параметри молекул. Сьогодні під QSAR/QSPR фактично мається на увазі інтелектуальний аналіз (data mining) відомої експериментальної інформації щодо структури та властивостей конкретних сполук. За результатами цього аналізу методами машинного навчання будують певні закономірності, виражені у вигляді різних математичних співвідношень, точніше — в термінах дескрипторів хімічної структури.

Якщо говорити більш конкретно, то процес QSAR/QSPR-моделювання передбачає, що для побудови моделі необхідно мати набір

сполук відомої структури з відомою цільовою властивістю. Це так звана *навчальна вибірка*. Для всіх цих сполук розраховують структурні параметри, або дескриптори. Наприклад, такими параметрами можуть бути кількість певних фрагментів, молекулярний об'єм, дипольний момент тощо. Далі отриманий набір чисел, що характеризує структуру, зіставляють з відомим набором чисел, що характеризує властивість, і виводять відповідний математичний вираз, який відображає зв'язок між ними.

Для побудови таких співвідношень використовують різні методи машинного навчання, зокрема множинний регресійний аналіз, нейронні мережі, методи опорних векторів та ін.

Для оцінки якості побудованої моделі застосовують так звану *тестову вибірку*, тобто набір сполук з відомими цільовими властивостями, але які не були задіяні в побудові QSAR-моделі. Якщо побудована модель достатньо адекватно прогнозує відповідну властивість сполук з тестової вибірки, її можна використовувати для прогнозування цієї властивості інших сполук, які входять до так званої *області застосування моделі* (applicability domain). Якість моделі оцінюють за певним набором статистичних показників, зокрема за коефіцієнтами детермінації, які в разі ідеальної моделі наближаються до одиниці.

Слід зазначити, що в Україні перші роботи з використання математичних підходів для дослідження зв'язку «структура — біологічна активність» було започатковано ще у 70-х роках минулого століття з ініціативи та за безпосередньої участі мого вчителя, видатного хіміка-органіка академіка Олексія Всеволодовича Богатського (1929—1983). І сьогодні у Фізико-хімічному інституті, названому його ім'ям, ми продовжуємо розвивати цей перспективний напрям, застосовуючи нові сучасні підходи. Один з основних наукових напрямів Інституту сформульовано як «цілеспрямований пошук і молекулярний дизайн нових речовин та матеріалів з комплексом необхідних властивостей, який ґрунтується на структурно-функціональному аналізі сполук, що враховує їх топологію, геометрію і розподіл властивостей атомів у молекулі».

Основні завдання наших досліджень:

- створення методології розроблення адекватних моделей «структура — властивості» (QSAR/QSPR), що мають прогнозувальний характер;
- побудова та експлуатація QSAR/QSPR-моделей для віртуального скринінгу органічних сполук щодо різних видів їх біологічної активності, термодинамічних характеристик, оптичних властивостей тощо;
- виявлення закономірностей впливу структури різних органічних сполук на їхні властивості, структурна, фізико-хімічна та стереохімічна інтерпретація QSAR/QSPR-моделей для подальшого молекулярного дизайну нових перспективних сполук;
- моделювання сумішей органічних сполук та прогноз їхніх властивостей;
- моделювання різноманітних неорганічних речовин, наночастинок і молекулярних асоціацій та прогноз їхніх властивостей;
- створення комп'ютерних експертних систем для фахівців різних галузей науки — споживачів моделей QSAR/QSPR.

Далі я наведу кілька конкретних прикладів наших розробок у цій галузі.

Головною методологічною розробкою в дослідженні зв'язку «структура — властивості»

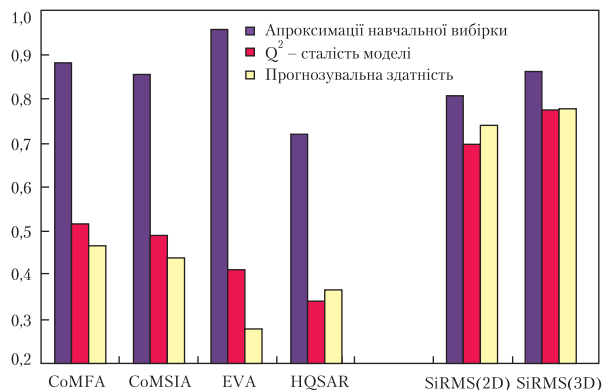


Рис. 2. Порівняльний аналіз різних QSAR-підходів, зокрема симплексних SiRMS(2D) і SiRMS(3D)

є симплексне представлення молекулярної структури [1]. Це різновид так званих фрагментних дескрипторів, коли молекулу розглядають як систему фрагментів, переважно 4-атомних, — так званих симплексів. Важливою особливістю цього підходу є аналіз не лише зв'язаних фрагментів, а й нез'язаних, а головне — симплекси можна охарактеризувати як стандартно, тобто атомами, які до них входять, так і фізико-хімічним описом (атомні заряди, вандерваальсовий об'єм, ліпофільність, поляризованість, донори/акцептори Н-зв'язку), а також описом різних полів, джерелом яких є вершини (атоми) симплексів (рис. 1). Симплексне представлення молекулярної структури забезпечує універсальність, різноманіття та гнучкість опису хімічних сполук різних типів; створює зрозумілі й добре інтерпретовані QSAR-моделі, які дозволяють легко здійснювати молекулярний дизайн речовин з потрібним рівнем активності. Симплексний підхід не має обмежень, властивих QSAR-підходам, які засновані на порівняльному аналізі молекулярних полів і які можна використовувати тільки для структурно однорідного набору молекул, тобто для кожної молекули розглядають лише один випадковий конформер.

На рис. 2 наведено результати порівняльного аналізу різних QSAR-підходів, який було проведено на вибірці інгібіторів ангіотензинперетворювального ензиму. Можна бачити,

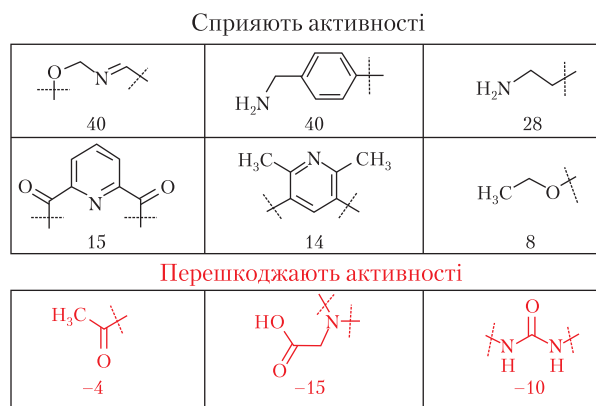


Рис. 3. Внески різних фрагментів у величину антигерпетичної активності, $\Delta\%$

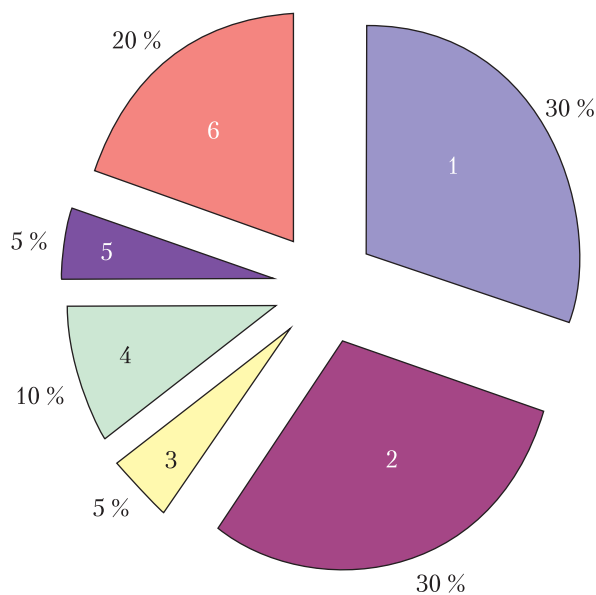


Рис. 4. Систематизація фізико-хімічних факторів, що визначають антигерпетичну активність: 1 – електростатичні; 2 – гідрофобні; 3 – Н-зв'язок; 4 – дисперсійні; 5 – інші; 6 – індивідуальні особливості атомів

що всі зазначені підходи з точки зору апроксимації навчальної вибірки приблизно однакові, але прогнозувальна здатність помітно вища у симплексного підходу.

Завдяки симплексному підходу для кожної досліджуваної молекули можна виділити фрагменти, які або сприяють, або перешко-

джають прояву активності, або є відносно ін-диферентними. Аналізуючи в такий спосіб усю вибірку сполук, що досліджуються, можна виділити фрагменти, які стало впливають на активність – як позитивно, так і негативно (рис. 3). Ця інформація є корисною для молекулярного дизайну нових більш активних молекул.

Додатково аналізуючи симплекси молекул навчальної вибірки, можна оцінити відносний вплив різних фізико-хімічних чинників, що визначають певну властивість (рис. 4).

Один із прикладів наших QSAR/QSPR-досліджень, проведених спільно з американськими колегами, пов'язаний з вирішенням досить нестандартного завдання – прогнозування водної розчинності органічних сполук за різних температур. Причому йдеться не лише про оцінку розчинності речовин у воді, а й про обов'язковий аналіз впливу на неї температури. З використанням бази даних, яка містить різноманітні відомості про органічні сполуки, було побудовано нетрадиційну високоякісну «двошарову» систему QSPR-моделей, за допомогою якої можна досить легко оцінити для кожної конкретної сполуки показники розчинності у воді за певної температури [2]. Інтерпретація цієї моделі показала, що наявність у молекулах органічних сполук груп, які потенційно можуть утворювати водневий зв'язок, сприяє розчинності, що й слід було очікувати.

З іншого боку, QSPR-дослідження водної розчинності деяких гетариламонієвих солей гексафторкремнієвої кислоти виявили несподіваний ефект – групи, що є донорами водневого зв'язку, не сприяють, а навпаки, перешкоджають розчинності. Ми показали, що причиною такого незвичайного явища є водневі зв'язки груп OH-, COOH-, NH-, NH₂- органічної частини солі з атомами фтору аніона (SiF₆)⁻, які зміцнюють зв'язок між катіоном і аніоном і, відповідно, перешкоджають дисоціації солі та її розчинності у воді [3].

Інший наш спільний проєкт з колегами зі Страсбурзького університету був присвячений віртуальному скринінгу сконструйованих RGD-пептидоміметиків – ефективних анти-

агрегаційних засобів. Ми «просіяли» понад 20 тис. сполук (з урахуванням стереоізомерів) через систему моделей, що містить фармакофорні моделі, QSAR-моделі та молекулярний докінг. Потім ми теоретично оцінили відібрані сполуки-кандидати з точки зору їх можливої токсичності, певних фармакокінетичних характеристик, розчинності у воді, синтетичної доступності. Відібрані з них 4 перспективні сполуки було синтезовано та протестовано експериментально. Результати цих досліджень довели їх високу ефективність [4]. Зрештою сполуку, яка виявилася найперспективнішою для створення ефективних антиагрегаційних засобів, було відправлено для проведення додаткових біологічних досліджень.

Разом з американськими дослідниками ми проводили оцінку екологічної небезпеки різних високоенергетичних нітроароматичних сполук військового призначення. Результатом цієї роботи стало створення адекватних QSAR-моделей для визначення різних видів їх токсичності, розчинності у воді, ліпофільності та біодоступності. У комплексі всі ці характеристики визначають екологічну небезпеку цих сполук. Зважаючи на актуальність оборонної тематики, ми плануємо розробити комп'ютерну експертну систему для прогнозування (позаекспериментального скринінгу) високоенергетичних сполук військового призначення (зокрема, вибухівки) щодо їх шкідливого впливу на навколишнє середовище та організм людини і тварин [5].

Інша робота, яка добре продемонструвала можливості нашого підходу щодо прогнозування властивостей сумішей речовин, стосувалася оцінки кривих рівноваги «рідина — пара» для сумішей різних розчинників [6]. Це дослідження ми проводили разом з французькими колегами. Виявилось, що наші моделі точно прогнозують наявність азеотропів та їхній склад. Загалом збіг отриманих нами результатів розрахунку з експериментальними даними можна вважати задовільним, хоча ми вже маємо ідеї, як поліпшити модель.

Проблема QSPR-моделювання неорганічних сполук полягає в тому, що поняття «струк-

турна формула» для багатьох з них є умовним, а отже, найпопулярніші 2D-моделі для органічних сполук у цих випадках малоприменні. Тому ми запропонували 1D-підхід, у якому неорганічну сполуку представляють у вигляді бруто-формули. Далі аналізують різні комбінації (двійки, трійки, четвірки і т.д.) елементів, що входять до досліджуваних сполук. Кількість таких конкретних комбінацій та фундаментальні властивості відповідних елементів періодичної системи становлять основу для розрахунку нових 1D-дескрипторів. Показано, що такої інформації в QSPR-моделях цілком достатньо для прогнозування температур плавлення та показників заломлення різноманітних неорганічних сполук (оксидів, галогенідів, халькогенідів, карбонатів, сульфатів тощо). Середня відносна помилка прогнозу в обох випадках становить 10 % [7].

Особливості моделювання наночастинок потребують урахування не лише структури молекул, що їх утворюють, а й характеристик самих наночастинок. У проведеному нами дослідженні було використано систему 1D-дескрипторів для опису неорганічних речовин, а для характеристики безпосередньо наночастинок — запозичений у фізиків метод «рідкої краплі» (з огляду на розмір наночастинки оцінюють її об'єм, площу поверхні, відносну кількість молекул в об'ємі та на поверхні та ін.). Ефективність запропонованого підходу було продемонстровано на прикладі дослідження цитотоксичності (на клітинах *E. coli* і *HaCaT*) нанорозмірних оксидів металів [8]. Отримані QSAR-моделі цілком адекватні і мають достатню прогнозувальну здатність. Показано, що зазначена активність на 60 % визначається природою оксиду і на 40 % — розмірними характеристиками наночастинки.

Крім того, завдяки результатам фундаментальних досліджень було отримано низку прикладних розробок. Створено 12 комп'ютерних експертних систем для прогнозування різноманітних властивостей органічних сполук та їхніх сумішей, зокрема різних видів токсичності й біодоступності, водної розчинності, ліпофільності, деяких термодинамічних власти-

востей (теплоти утворення та згоряння) тощо. Зазначу, що експертна система — це програма для комп'ютера, яка оперує знаннями в певній предметній галузі з метою вироблення рекомендацій або вирішення певної проблеми. Наші системи мають дружній інтерфейс, що спрощує їх використання фахівцями різних галузей науки — споживачами моделей QSAR/QSPR.

Комплекс комп'ютерних програм HIT QSAR, в якому реалізовано наш підхід до аналізу зв'язку «структура — властивості», та деякі експертні системи захищено свідоцтвами на авторське право України.

Насамкінець хотів би презентувати новий інноваційний анальгетичний засіб «Пропоксазепам», який гальмує гострий і, що дуже важливо, хронічний біль та має полімодальний механізм дії. Зараз триває друга стадія клінічних випробувань цього препарату, спонсором яких є ТДВ «Інтерхім». У процесі його розроблення нам вдалося спрогнозувати деякі властивості

діючої речовини цього препарату, зокрема ліпофільність та водну розчинність. У попередніх дослідженнях було показано, що цей препарат не має мутагенної дії, є біодоступним і може проникати з крові в мозок.

Дослідницький колектив хемоінформатиків Фізико-хімічного інституту ім. О.В. Богатського НАН України має високий міжнародний авторитет, а географія співпраці та кооперації охоплює провідні наукові центри 14 країн світу.

Отже, хемоінформатика — це прорив у майбутнє — від емпіричної хімії до передбачальної. Сьогодні успіх у створенні нової корисної речовини, зокрема ліків, багато в чому залежить від взаєморозуміння та взаємодії принаймні трьох груп фахівців — хіміків-синтетиків, хемоінформатиків та біологів, фармакологів.

Дякую за увагу!

За матеріалами засідання підготувала О.О. Мележик

REFERENCES

1. Kuz'min V., Artemenko A., Ognichenko L., Hromov A., Kosinskaya A., Stelmakh S., Sessions Z., Muratov E. Simplex representation of molecular structure as universal QSAR/QSPR tool. *Structural Chemistry*. 2021. **32**: 1365–1392. <https://doi.org/10.1007/s11224-021-01793-z>
2. Klimenko K., Kuz'min V., Ognichenko L., Gorb L., Shukla M., Vinas N., Perkins E., Polishchuk P., Artemenko A., Leszczynski J. Novel Enhanced Applications of QSPR Models: Temperature Dependence of Aqueous Solubility. *J. Comput. Chemistry*. 2016. **37**: 2045–2051. <https://doi.org/10.1002/jcc.24424>
3. Gelboldt V., Ognichenko L., Shyshkin I., Kuz'min V. QSPR models for water solubility of ammonium hexafluoro-silicates: analysis of the effects of hydrogen bonds. *Structural Chemistry*. 2021. **32**: 309–319. <https://doi.org/10.1007/s11224-020-01652-3>
4. Polishchuk P.G., Samoylenko G.V., Khristova T.M., Krysko O.L., Kabanova T.A., Kabanov V.M., Kornilov A.Yu., Klimchuk O., Langer T., Andronati S.A., Kuz'min V.E., Krysko A.A., Varnek A. Design, Virtual Screening, and Synthesis of Antagonists of $\alpha\text{IIb}\beta\text{3}$ as Antiplatelet Agents. *J. Med. Chem.* 2015. **58**(19): 7681–7694. <https://doi.org/10.1021/acs.jmedchem.5b00865>
5. Ognichenko L.N., Kuz'min V.E., Gorb L., Muratov E.N., Artemenko A.G., Kovdienko N.A., Polishchuk P.G., Hill F.C., Leszczynski J. New advances in QSPR/QSAR analysis of nitrocompounds: solubility, lipophilicity and toxicity. In: Leszczynski J., Shukla M. (Eds) *Practical aspects of computational chemistry II*. Springer, London, 2012. P. 279–334. https://doi.org/10.1007/978-94-007-0923-2_8
6. Oprisiu I., Varlamova E., Muratov E., Artemenko A., Marcou G., Polishchuk P., Kuz'min V., Varnek A. QSPR Approach to Predict Nonadditive Properties of Mixtures. Application to Bubble Point Temperatures of Binary Mixtures of Liquids. *Molecular Informatics*. 2012. **31**(6-7): 491–502. <https://doi.org/10.1002/minf.201200006>
7. Kuz'min V.E., Ognichenko L.N., Zinchenko V.F., Artemenko A.G., Shyrykalova A.O., Kozhukhar A.V. QSPR Models for Predicting of the Melting Points and Refractive Indices for Inorganic Substances: Components of the Optical Film-Forming Materials. *International Journal of Quantitative Structure Property Relationships*. 2020. **5**(1): 1–21. <https://doi.org/10.4018/IJQSPR.2020010101>

8. Kuz'min V., Ognichenko L., Sizochenko N., Chapkin V., Stelmakh S., Shyrykalova A., Leszczynski J. Combining Features of Metal Oxide Nanoparticles: Nano-QSAR for Cytotoxicity. *International Journal of Quantitative Structure-Property Relationships*. 2019. 4(1): 28–40. <https://doi.org/10.4018/978-1-7998-8591-7.ch014>

Victor E. Kuz'min

O.V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Odesa, Ukraine

ORCID: <https://orcid.org/0000-0002-2753-0453>

CHEMOINFORMATICS AS AN EFFECTIVE TOOL FOR FORECASTING
AND DESIGNING NEW SUBSTANCES WITH A COMPLEX OF NECESSARY PROPERTIES

Transcript of scientific report at the meeting of the Presidium of NAS of Ukraine, September 21, 2022

The report is devoted to the results of fundamental and applied scientific research in the field of chemoinformatics conducted at the O.V. Bogatsky Physico-Chemical Institute of the NAS of Ukraine and related to the analysis of structure-property relationships in chemical substances for the purposeful design of new compounds and materials with a set of prescribed properties.