

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ФІЗИКО-ХІМІЧНИЙ ІНСТИТУТ ІМ. О.В. БОГАТСЬКОГО

МОКШИНА ОЛЕНА ГЕОРГІЇВНА



УДК 547.898:544.16

**КРИТИЧНІ ВЛАСТИВОСТІ ТА ДРУГІ ВІРІАЛЬНІ КОЕФІЦІЄНТИ
ОРГАНІЧНИХ РЕЧОВИН ТА ЇХ СУМІШЕЙ. QSPR АНАЛІЗ З УРАХУВАННЯМ
МІЖМОЛЕКУЛЯРНИХ ВЗАЄМОДІЙ.**

02.00.03. – Органічна хімія

АВТОРЕФЕРАТ

дисертації на здобуття наукового ступеня кандидата хімічних наук

Одеса – 2016

Дисертацією є рукопис

Робота виконана у відділі молекулярної структури і хемоінформатики Фізико-хімічного інституту ім. О.В. Богатського НАН України.

Науковий керівник: доктор хімічних наук, професор
Кузьмін Віктор Євгенович,
Фізико-хімічний інститут ім. О.В. Богатського НАН України,
завідувач відділу молекулярної структури і хемоінформатики.

Офіційні опоненти: доктор хімічних наук, професор
Оковитий Сергій Іванович,
Дніпропетровський національний університет
ім. Олеся Гончара, завідувач кафедри органічної хімії;
кандидат хімічних наук, старший науковий співробітник
Танчук Всеволод Юлійович,
Інститут біоорганічної хімії та нафтохімії НАН України,
старший науковий співробітник.

Захист відбудеться «7» жовтня 2016 р. о 10:00 год. на засіданні спеціалізованої вченої ради Д 41.219.02 в Фізико-хімічному інституті ім. О.В. Богатського НАН України за адресою: 65080, м. Одеса, Люстдорфська дорога, 86.

З дисертацією можна ознайомитись в науковій бібліотеці Фізико-хімічного інституту ім. О.В. Богатського НАН України.

Автореферат розісланий «___» _____ р.

Вчений секретар спеціалізованої
вченої ради Д 41.219.02,
к.х.н., с.н.с.



С. М. Плужник-Гладир

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ¹

Актуальність теми. Термодинамічні властивості органічних сполук, зокрема параметри критичної точки, фактор Пітцера (ω) та другі віріальні коефіцієнти (ДВК), є фундаментальними властивостями, на основі яких проводять розрахунки при впровадженні та проектуванні енерго-хіміко-технологічних систем.

Експериментальні методи визначення критичних параметрів мають велику трудомісткість (жорсткі вимоги до проведення аналізу, висока чистота реагентів, складність фіксування термодинамічної рівноваги і т.ін.) та собівартість. Крім того, ці методи є незастосовними для великих молекул. Властивості бінарних сумішей є найбільш визначними при моделюванні фазових рівноваг, оскільки саме бінарні суміші є найбільш широко застосовними робочими середовищами.

Більшість з відомих термодинамічних методів прогнозування досліджуваних властивостей можуть бути застосовані до сполук одного класу або потребують додаткових експериментальних даних. Існуючі QSPR-моделі (Quantitative Structure-Property Relationships – кількісні співвідношення «структура-властивість»), що описують критичні властивості розроблені з використанням невеликих наборів споріднених сполук, та їх корисність обмежена. Невелика кількість досліджень, в яких QSPR моделювання проводилося для широкого діапазону органічних сполук, використовує 3D моделі молекул і характеризується вкрай трудомісткими чисельними методиками (конформаційний пошук з використанням квантово-хімічних методів). Майже усі з існуючих робіт для QSPR індивідуальних сполук використовують дескриптори, які інтерпретувати складно або неможливо. Щодо QSPR моделювання критичних властивостей сумішей, то попередніх досліджень немає, бо вибір сумішевих дескрипторів є вкрай обмеженим. Для других віріальних коефіцієнтів як індивідуальних сполук, так і сумішей QSPR моделей до сьогоднішнього дня не існувало.

Тому є актуальною розробка нових підходів до моделювання вищезгаданих термодинамічних властивостей. 2D QSPR моделі на основі симплексного представлення молекулярної структури має ряд переваг над аналогами. Наявність розробленої системи сумішевих дескрипторів дозволяє побудувати моделі для термодинамічних властивостей сумішей органічних сполук. Створення адекватних моделей для прогнозування досліджуваних властивостей органічних речовин різних класів та їх бінарних сумішей на базі структурної інформації є, таким чином, актуальною задачею. Досліджувані в роботі термодинамічні властивості і, в першу чергу, критичні властивості, безпосередньо зв'язані з енергією міжмолекулярної взаємодії. Це дозволяє через оцінку впливу структурних параметрів на термодинамічні характеристики виявити та проінтерпретувати структурні залежності енергії міжмолекулярної взаємодії. Результати моделювання також дозволять замінити тривалі, не завжди ефективні, неекономічні розрахунки на новий, більш ефективний підхід; виявити та проінтерпретувати структурні залежності для критичних властивостей та ДВК.

Зв'язок роботи з науковими програмами, планами, темами

Дана робота виконувалася відповідно до планів відомчої тематики відділу молекулярної структури і хемоінформатики ФХІ ім. О.В. Богатського НАН України «Експертна система прогнозування критичних параметрів органічних сполук та їх бінарних сумішей» (2015-2016, № державної реєстрації 015U001060). Частина роботи, присвячена

¹) У керівництві дисертацією брав участь професор, д.т.н. Недоступ Вадим Іоаннович

каузальному аналізу, виконувалася в рамках проекту NanoBridges (People Programme (Marie Curie Actions) of the European Union Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no 295128").

Мета і завдання дослідження. Аналіз впливу структури органічних сполук різноманітних класів і складу їх бінарних сумішей на критичні властивості, фактор Пітцера та другі віріальні коефіцієнти. Для досягнення вказаної мети необхідно було вирішити наступні завдання:

1. Збір, аналіз та верифікація експериментальної інформації про критичні властивості, ω та ДВК органічних сполук. Формування відповідних баз даних.
2. Створення системи дескрипторів для міжмолекулярних взаємодій в індивідуальних органічних речовинах та їх сумішах.
3. Розробка та валідація QSPR моделей на основі симплексного підходу та дескрипторів міжмолекулярних взаємодій для досліджуваних термодинамічних властивостей органічних речовин та їх бінарних сумішей.
4. Визначення структурних та фізико-хімічних чинників, що зумовлюють досліджувані властивості.
5. Створення експертної системи віртуального скрінінгу на основі побудованих QSPR моделей.

Об'єкт дослідження. Системи структурного опису органічних сполук та їх бінарних сумішей та моделі для прогнозування критичних властивостей, фактору Пітцера та других віріальних коефіцієнтів. Структурні та фізико-хімічні фактори, що впливають на досліджувані властивості.

Предмет дослідження. Органічні сполуки різних класів та їх бінарні суміші.

Методи дослідження. Генерація 2D структурних дескрипторів була виконана методом симплексного представлення молекулярної структури, «псевдосумішевим» методом та методом парних дескрипторів взаємодій. Для попередньої обробки даних використовувалися підходи т.з. «data curation». Залежності «структура – властивість» будувалися за допомогою методів машинного навчання – метода загального лінійного моделювання (GLM), метода часткових найменших квадратів (PLS), метода випадкового лісу (RF), метода опорних векторів (SVM) та метода градієнтного бустінгу (GBM). Валідація моделей проводилась за допомогою 5-кратної зовнішньої крос-валідації та (при наявності) зовнішньої тестової вибірки.

Наукова новизна отриманих результатів.

1. Розроблені нові QSPR моделі, які прогнозують, важливі термодинамічні властивості індивідуальних органічних речовин та їх бінарних сумішей – критичні параметри, ω та ДВК для органічних речовин різних класів.
2. Вперше на основі симплексного представлення молекулярної структури розроблені підходи, які більш глибоко відображають природу взаємодії досліджуваних речовин. Розроблено новітні дескриптори – «псевдосумішеві», які моделюють структуру індивідуальної речовини як суміш двох ідентичних сполук, та парні міжмолекулярні дескриптори, які базуються на параметрах атом-атомних потенціалів. Обидва типи дескрипторів дозволяють враховувати можливі міжмолекулярні взаємодії.
3. Вперше проведено каузальний аналіз залежності ω від характеристик форми та розміру молекули.
4. Використано новий «двошаровий» підхід для QSPR моделювання ДВК, в якому використовується теоретична залежність ДВК від температури.

5. Вперше побудовані QSPR моделі для критичних властивостей та ДВК бінарних сумішей.
6. Показано вплив фізико-хімічних та структурних факторів на досліджувані властивості.

Практичне значення отриманих результатів. Отримані в роботі результати визначають шляхи цілеспрямованого пошуку нових технічних агентів на основі індивідуальних органічних сполук та їх сумішей, зокрема теплоносіїв, холодоагентів і т.і. Отримані моделі можуть бути корисні для обчислення широкого кола хіміко-технологічних процесів. Розроблена на їх основі експертна система може використовуватися для оцінки критичних властивостей, ω та ДВК органічних речовин та їх сумішей. Експертна система є відкритою до використання та спрямованою на широке коло фахівців. Отримані результати можуть бути використані при викладанні спецкурсів для студентів спеціальності «Органічна хімія».

Особистий вклад здобувача. Самостійно розроблена система дескрипторів для більш детального опису міжмолекулярних взаємодій на 2D-рівні опису молекулярної структури, реалізована генерація дескрипторів у вигляді скриптів, побудовано QSPR моделі та проведено їх валідацію, визначено вплив структурних та фізико-хімічних факторів на досліджувані властивості. У формуванні завдань досліджень, інтерпретації та обговоренні результатів, а також у підготовці статей брали участь науковий керівник та професор, д.т.н. В.І. Недоступ. Частина дослідження була виконана у співавторстві з к.х.н. П.Г. Поліщуком. Академік НАН України Г.Л. Камалов висловив корисні поради щодо оформлення дисертації.

Всім згаданим колегам автор щиро вдячна за допомогу та співпрацю.

Апробація результатів дисертації. Основні результати роботи доповідалися та обговорювалися на XIV, XVI та XVII конференціях молодих вчених та студентів-хіміків Південного регіону України (Одеса, 2012, 2014 та 2015) ; IV та VII Всеукраїнських наукових конференціях студентів і аспірантів «Хімічні Каразінські Читання» (Харків, 2012 та 2015); 3rd Strasbourg Summer School in Chemoinformatics (2012); 5th International Symposium: Methods and Applications of Computational Chemistry (Харків, 2013); I та II Kazan Summer School on Chemoinformatics (Казань, 2013 та 2015); XVIII Scientific Youth Conference «Problems and Achievements of Modern Chemistry» (Одеса, 2016).

Публікації. За матеріалами дисертації опубліковано 7 статей, 10 тез доповідей.

Структура та об'єм роботи. Дисертація складається зі вступу, п'яти розділів, висновків, бібліографії і 2 додатків. Робота викладена на 126 сторінках, містить 21 малюнки, 15 таблиць, список літератури з 135 найменувань.

Розділ 1 присвячено огляду літератури в розділах 2 і 3 викладено основний зміст роботи, розділ 4 містить опис експертної системи, у розділі 5 приведено опис комп'ютерних програм та алгоритмів, що були використані для рішення QSPR задач, в додатках наведено експериментальні та розрахункові дані для досліджуваних властивостей органічних сполук.

ОСНОВНИЙ ЗМІСТ РОБОТИ

1. Сучасні методи прогнозування термодинамічних властивостей органічних сполук та їх сумішей (огляд літератури)

У першому розділі проведено аналіз відомих моделей, що використовуються для вирішення задач QSPR для термодинамічних властивостей, таких як критичні

властивості, ω та ДВК. Розглянуто різні типи підходів, що використовуються для моделювання досліджуваних властивостей, такі як методи групових вкладів, методи, засновані на принципі відповідних станів, та деякі нечисленні QSPR моделі. У частині, присвяченій QSPR методології, стисло проаналізовані сучасні підходи, у тому числі до моделювання та прогнозування сумішей.

2. Розвиток методології симплексного представлення молекулярної структури для дослідження термодинамічних властивостей органічних речовин та їх сумішей

2.1. Розвиток симплексного представлення молекулярної структури

Для опису молекулярної структури використовувалось симплексне представлення молекулярної структури (СПМС) (Kuz'min V. E., 2010), в якому молекула відображується як граф, що розбивається на чотириатомні фрагменти зі сталими складом та структурою – симплекси. Для диференціації атомів при розрахунку симплексних дескрипторів використовуються різні підходи. Зокрема атому може бути присвоєна мітка (символ), що характеризує його унікальність (тип атома, донор/акцептор водневого зв'язку і т.д.), або мітка, що характеризує певні діапазони зміни деяких властивостей атомів: часткових зарядів на атомах, рефракції тощо. Дескрипторами є кількість симплексів конкретного вигляду з урахуванням різних способів диференціації атомів. Для сумішей СПМС було модифіковано (Orghiu I., 2012). Для суміші молекул А і В зв'язані симплекси генеруються для кожного компонента окремо, у той час як незв'язані симплекси можуть описувати обидва компоненти та суміш в цілому (рис.1). Для того, щоб відрізнити «сумішеві» симплекси від незв'язних

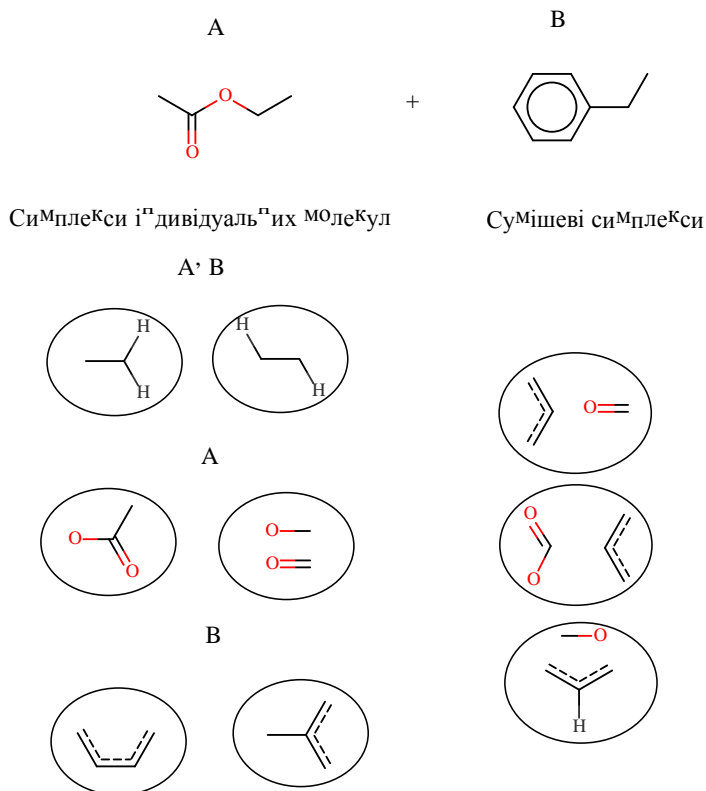
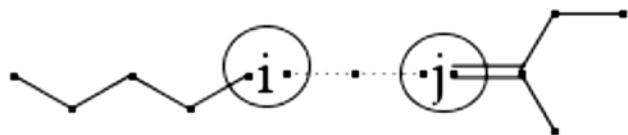


Рис. 1. Приклади симплексних дескрипторів для опису сумішей.

симплексів однієї молекули, використовуються мітки. Для опису властивостей індивідуальних органічних речовин було модифіковано сумішевий підхід. Цей підхід базується на природі досліджуваних властивостей та так званих «квазіхімічних» термодинамічних підходах. При розрахунку «псевдосумішевих» дескрипторів індивідуальні компоненти описуються як суміш двох ідентичних молекул. Завдяки специфіці сумішевих дескрипторів це не призводить до дублювання симплексів, а навпаки дозволяє отримати нові унікальні «псевдосумішеві» структурні параметри. Вони враховують можливі міжмолекулярні взаємодії в індивідуальних речовинах.

Стандартна схема сумішевих дескрипторів була модифікована і для других віріальних крос-коефіцієнтів. Оскільки другий віріальний крос-коефіцієнт описує тільки різномірні взаємодії в суміші, інформація щодо структури індивідуальних компонентів є надлишковою. Таким чином, для моделювання другого віріального крос-коефіцієнта були використані тільки сумішеві дескриптори.



Також було розроблено та застосовано новітні дескриптори парної взаємодії. На першому етапі розрахунку генеруються всі можливі незв'язані пари ij , де i – атом у першій молекулі, j – атом у

другій молекулі (порядок компонентів не має значення). Потім для кожної пари розраховуються параметри взаємодії згідно з обраним силовим полем. Для зважування незв'язаних пар використовуються параметри взаємодії (σ , ϵ) згідно обраного силового поля. Використовуються наступні правила зважування: $\sigma_{ij} = \frac{1}{2(\sigma_i + \sigma_j)}$; $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$, де ϵ – глибина мінімуму потенціалу міжмолекулярної взаємодії, σ – відстань, на якій енергія взаємодії дорівнює нулю.

В даному дослідженні було використано універсальне силове поле (UFF) (Rappe A. K., 1992), але можливо використання будь-якого іншого силового поля. На заключному етапі зважені атом-атомні пари розбиваються на групи [A; B; C] згідно з наступними інтервалами – ϵ ($A \leq 0.1 < B \leq 0.3 < C$) та σ ($A \leq 3.5 < B \leq 4 < C$). Подібно до СПМС число пар кожного типу і є дескриптором.

2.2. Технологія вирішення задач QSPR

У роботі використовувалася наступна технологія вирішення задач QSPR:

- Формування вибірок молекул.** Збереження та обробка даних. Перевірка та видалення помилкових та ненадійних значень (т.з. «data curation»). Стандартизація хімічних структур.
- Розрахунок симплексних та/або інших типів дескрипторів.** Попередня обробка отриманих дескрипторів, яка включала їх масштабування та нормалізацію.
- Побудування моделей** проводилося у середовищі програмування з відкритим кодом R. QSPR аналіз проводився з використанням методів машинного навчання: GLM, PLS, RF, GBM. Для оцінки якості моделей використовувалися наступні параметри: R^2 – коефіцієнт детермінації (середня похибка 0.03 розраховувалася в межах 95% довірчого інтервалу), RMSE – стандартна помилка прогнозу. Для критичних властивостей сумішей використовувалися також показники MAPE – середня абсолютна відсоткова помилка, AAD – середнє абсолютне відхилення, RMS – відносна стандартна помилка прогнозу ($RMS = 100 * \sqrt{(|\sum(y_{obs} - y_{pred})|) / y_{obs}}$), де y_{obs} – експериментальне значення, y_{pred} – спрогнозоване значення.
- Визначення області застосовності** отриманих моделей (domain applicability – DA). Область застосовності моделі визначає регіон хімічного простору, який адекватно описується сполуками навчальної вибірки. Статистичні моделі можуть надійно прогнозувати тільки ті сполуки, що належать цьому регіону. В даній роботі для визначення DA моделей був використаний метод фрагментного контролю (Solov'ev V. P., 2011). В цьому підході будь-яка молекула, що містить в собі фрагменти, які не

зустрічаються в навчальній вибірці, знаходиться за межами DA.

5. Для верифікації QSPR моделей використовувалася процедура 5-кратної зовнішньої крос-валідації (статистичні параметри позначаються нижнім індексом 5f) і/або зовнішня тестова вибірка (нижній індекс – ts). Сутність 5-кратної зовнішньої крос-валідації полягає в тому, що досліджувана вибірка розбивається на 5 частин. Кожна з цих частин використовується для верифікації, при цьому моделі будуються з тих частин вибірки, що залишилися. Процедура повторюється 5 разів. Таким чином усі досліджувані сполуки використовуються для прогнозу. Отримані результати потім об'єднуються у консенсусні моделі (передбачувані значення усереднюються по всіх відібраних моделях).

При валідації моделей для сумішей використовувалася модифікація раніше розробленого підходу «mixtures out» (Muratov E. N., 2012). В рамках даного підходу в тестову вибірку потрапляють всі точки суміші. У випадку побудованих моделей дані про індивідуальні компоненти не включалися до моделі.

6. Структурна та фізико-хімічна інтерпретація побудованих моделей. Фізико-хімічна інтерпретація моделей має на меті виявити, які саме змінні, зважені згідно з молекулярними властивостями (частковими зарядами, електронегативністю, наявністю водневих зв'язків та інші), мають найбільше значення для досліджуваних властивостей. Для фізико-хімічної інтерпретації була використана оцінка важливості змінних за Брімманом (Breiman, L., 2001). Для вимірювання важливості j-ї змінної після навчання моделі, значення j-ї змінної перемутьються (тобто переставляються) для всіх навчальних даних. Важливість змінної потім розраховується як різниця між помилками на оригінальній навчальній вибірці та на перемутьованих даних.

Для структурної інтерпретації моделей для індивідуальних сполук використовувався універсальний фрагментний підхід (Polishchuk, P. G., 2013), в якому вклад молекулярного фрагменту в досліджувану властивість розраховується як різниця між значенням для вибраної сполуки та прогнозованим значенням для вибраної сполуки за виключенням цього фрагменту. Для структурної інтерпретації моделей для сумішей оцінювалася (згідно зі схемою Бріммана) важливість окремих сумішевих дескрипторів.

7. Об'єднання побудованих QSPR моделей в єдину прогнозуючу експертну систему, що являє собою програмне середовище з графічною оболонкою.

3. Кількісна оцінка впливу структури на термодинамічні властивості індивідуальних органічних речовин та їх бінарних сумішей

3.1. Критичні властивості індивідуальних органічних речовин та їх бінарних сумішей

3.1.01. Критичні властивості індивідуальних органічних речовин

Для вирішення задач прогнозування термодинамічних властивостей маловивчених сполук на основі принципу термодинамічної подоби найчастіше використовуються дані про координати критичної точки (T_c – критична температура, P_c – критичний тиск, V_c – критичний об'єм) і, в якості критерію подоби, виступає т.з. фактор Пітцера (ω), що визначається за формулою:

$$\omega = -lg \frac{P_c}{P_{\tau=0.7}} - 1$$

,де τ – наведена температура, що дорівнює T/T_c . $P_{\tau=0.7}$ – тиск насиченої пари в точці, де $\tau = 0.7$. Для сферичних молекул благородних газів значення ω близьке до нуля, та

чим менш сферична молекула, тим далі значення ω від 0. Пітцер назвав цю величину ацентричним фактором, вважаючи, що вона характеризує несферичність молекули.

Тому нами було досліджено вплив структури на критичні параметри та фактор Пітцера для органічних молекул різних класів. При цьому була використана база даних (NIST WebBook: <http://webbook.nist.gov/chemistry/>), що включала в себе 407 сполук з відомими значеннями критичної температури, 382 сполуки – для критичного тиску, 309 сполук – для критичного об'єму, 331 сполуку – для фактора Пітцера. В навчальній вибірці були представлені карбон-, галоген-, киснев-, нітроген та сульфуровмісні сполуки.

Ряд моделей для критичних властивостей було побудовано з використанням «псевдосумішевого» підходу. Для порівняння обрані моделі, що базуються на звичайних симплексних дескрипторах (моделі «однієї молекули»). Найкращі результати були отримані з використанням методу RF. Усі фінальні моделі характеризуються високою якістю апроксимації ($R^2 = 0.97-0.99$) та високою прогнозуючою здатністю у рамках 5-кратної зовнішньої крос-валідації ($R^2_{sf} = 0.86-0.95$). Стандартна помилка прогнозу для моделей «однієї молекули» складає 32.7 K для критичних температур, 28 см³/моль – для критичних об'ємів, 4.1 бар для критичних тисків та 0.027 логарифмічних одиниць для фактора Пітцера. Стандартна помилка прогнозу для «псевдосумішевих» моделей складає 30.6 K для критичних температур, 24 см³/моль – для критичних об'ємів, 4 бар для критичних тисків.

Графіки «експериментальне-спрогнозоване» (рис. 2) демонструють, що найбільші відхилення спрогнозованих значень від експериментальних спостерігаються при граничних значеннях властивостей. Найбільшими є помилки прогнозу для невеликих значень критичних властивостей у зв'язку з важкістю опису малих молекул – наприклад, метану, оскільки для них генерується замало унікальних дескрипторів. З іншого боку, великі помилки прогнозу можуть бути обумовлені збільшенням експериментальної помилки для складних молекул. Наприклад, експериментальні помилки для критичних температур метану та етану складають тільки 0.3 K, у той час як для нонадекану та ейкозану помилки складають близько 8 K. Для кисневмісних сполук, наприклад, етанолу, ацетальдегіду та оцтової кислоти експериментальні помилку складають відповідно 7 K, 2 K та 30 K.

Порівняння моделей «однієї молекули» та «псевдосумішевих» моделей показує статистично значущу різницю у RMSE у 95% довірчому інтервалі. Результуючі моделі для фактора Пітцера виявилися ідентичними за своїми статистичними характеристиками. Такий результат був очікуваний, оскільки фактор Пітцера є індивідуальною характеристикою молекули. Відносний розрахований рівень входження «псевдосумішевих» дескрипторів для усіх трьох властивостей (T_c , P_c , V_c) складає біля 25%. Аналіз розподілу помилок прогнозу для тестової вибірки показує, що для усіх критичних властивостей максимум розподілу припадає на діапазон допустимої експериментальної помилки – менше 10%.

Аналіз впливу зважених дескрипторів показує домінуючий внесок електростатичних та вандерваальсових взаємодій. Це підтверджується тим фактом, що параметри критичної точки безпосередньо залежать від енергії міжмолекулярних взаємодій. Роль топології молекули (розгалуженість скелету без врахування природи атомів) порівняно мала для усіх критичних властивостей. Здатність атома бути донором

або акцептором водневого зв'язку є більш важливою для критичних тисків – 19%, порівняно з 6% та 9% для відповідно критичних температур та критичних об'ємів.

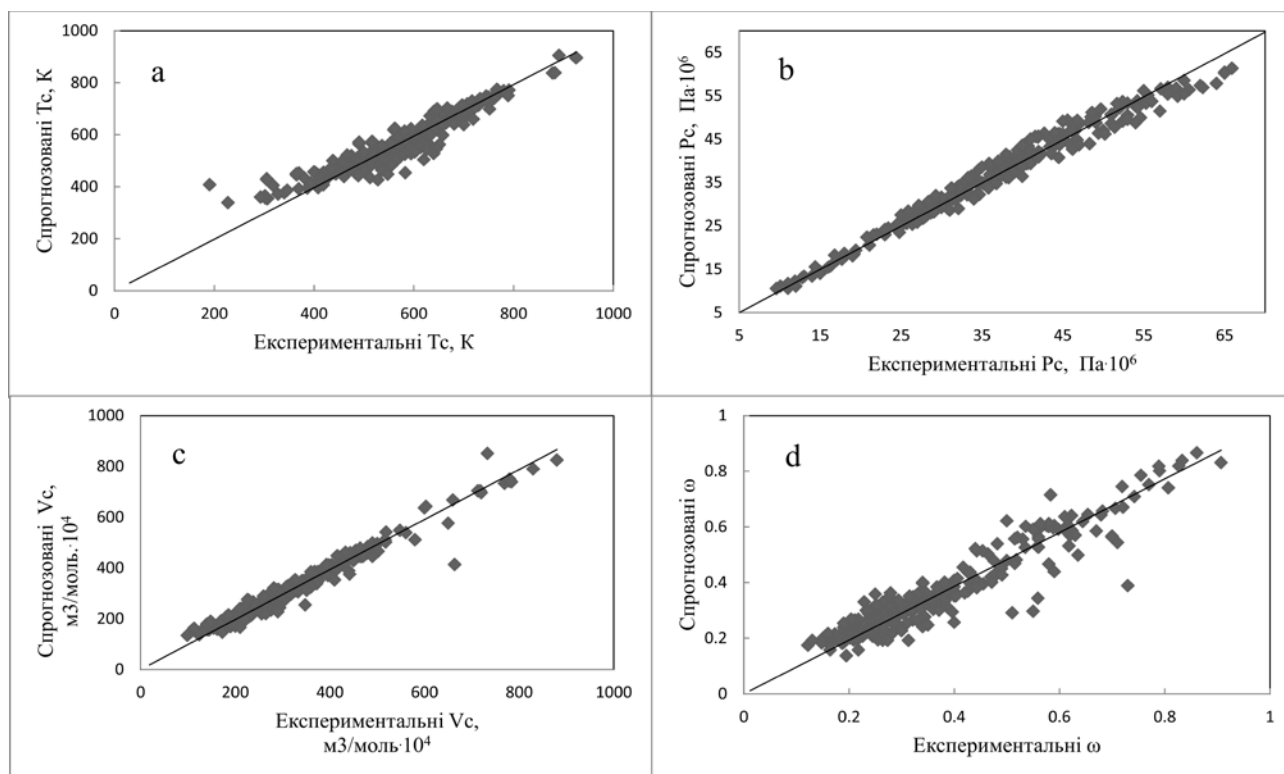


Рис. 2. Діаграма «експериментальне-спрогнозоване» для досліджуваних властивостей: а) критична температура T_c , б) критичний тиск P_c , в) критичний об'єм V_c , д) фактор Пітцера ω

Структурна інтерпретація виявила найбільш впливові фрагменти молекул. Для критичних температур максимальний позитивний вплив виявляють ароматичні та аліциклічні фрагменти та фрагменти, що містять карбонільну групу, а негативний вплив – фрагменти, що містять атоми фтору. Такі тенденції можна пояснити, виходячи з того, що критична температура зв'язана з енергією міжмолекулярної взаємодії і підвищується при зростанні здатності молекул до взаємопритягання. В свою чергу зростання здатності молекул до відштовхування негативно впливає на критичні температури. Для критичного тиску найбільший позитивний вплив мають дісульфідна група, аліфатичний фтор, первинна аміногрупа, потрійний $C\equiv C$ зв'язок та аліфатична гідроксогрупа. Закономірно, для критичних об'ємів найбільш об'ємні фрагменти мають позитивний вплив.

3.1.02. Каузальний аналіз фактора Пітцера: вплив параметрів форми та розміру молекул

Наведене вище визначення фактора Пітцера має емпіричний характер. Важливо знати, як впливають структура та сферичні характеристики на цю величину. Каузальний аналіз дозволяє *in silico* відтворити рандомізований контрольований експеримент, який є стандартним способом виявити причинно-наслідкові взаємозв'язки. Алгоритми каузального аналізу знаходяться у відкритому доступі, вихідний код знаходиться на <http://webdav.tuebingen.mpg.de/causality/>. В даній роботі ми використали декілька алгоритмів, найбільш адекватним з яких виявився алгоритм LTr (Linear Trace method).

Цей метод був розроблений для виявлення каузальних відношень у багатовимірних даних. Головна ідея цього методу – використання асиметрії між розподілами ймовірних причини та ефекту. Така асиметрія спостерігається у тому випадку, коли коваріаційна матриця причини та матриця, що зв'язує причину з ефектом, ймовірно незалежні. Цей метод не залежить від типу розподілу даних.

Таблиця 1

Дескриптори, що впливають на фактор Пітцера

Дескриптор	Опис
ECC – eccentricity, 2D	Сума атомних ексцентриситетів (максимальних відстаней від одного атома до будь-якого іншого) за всіма неводневими атомами $ECC = \sum_i \eta_i$, де η_i – ексцентриситет <i>i</i> -го атома.
AECC – average eccentricity, 2D	$AECC = \frac{ECC}{n_{noHAt}}$, де n_{noHAt} – кількість неводневих атомів.
DECC – eccentric, 2D	Міра середнього зміщення атомних ексцентриситетів від AECC. $DECC = \frac{\sum_i \eta_i - AECC}{n_{At}}$, де η_i – ексцентриситет <i>i</i> -го атома, n_{At} – кількість атомів.
ASP – asphericity, 3D	Анізометричний дескриптор, який вимірює відхилення конформації від сферичної форми: $ASP = \frac{1}{2} \cdot \frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}$, де λ – власні значення молекулярної матриці інерції. Асферичність змінюється від 0 для сферичних молекул до 1 для лінійних молекул.

Проведений каузальний аналіз дозволив оцінити причинно-наслідкові взаємозв'язки між фактором Пітцера та молекулярними параметрами, що зв'язані з формою та розміром (табл. 1). Метод LTr показав, що дескриптори ECC, AECC, DECC значно причинно впливають на фактор Пітцера. Несподівані результати були отримані для дескриптора ASP (асферичність), лінійний коефіцієнт кореляції з ω для якого дорівнює 0.39. Але метод LTr показав наявність значимого каузального впливу ASP на фактор Пітцера. При цьому інтерпретація ASP відповідає традиційній інтерпретації фактора Пітцера. Для вуглеводнів дескриптори форми не є впливовими, що може бути зв'язано з тим, що оптимальна конформаційна форма близька до сферичної для вуглеводнів в неідеальних умовах. В порядку зменшення впливу дескриптори утворюють наступний ряд: ECC > AECC > DECC ~ ASP. Таким чином, найбільш впливовими виявляються саме 2D дескриптори, що відображають розмір молекули.

3.1.03. Критичні властивості бінарних сумішей органічних речовин

Дані щодо критичних властивостей бінарних сумішей є значно більш обмеженими, ніж для індивідуальних сполук, що пояснюється передусім експериментальною складністю дослідження сумішей. У виборці представлені 94 сполуки, близько 300 їх сумішей. З урахуванням різного складу сумішей кількість експериментальних точок для кожної властивості – близько 1000 (Hicks C., 1975). За

типами суміші було поділено на суміші неполярна сполука – неполярна сполука (НП-НП), неполярна сполука – полярна сполука (НП-ПЛ), полярна сполука – полярна сполука (ПЛ-ПЛ). Найбільша кількість точок належала сумішам типу НП-НП.

При формуванні QSPR моделей для сумішей були використані парні дескриптори. Було побудовано ряд моделей за допомогою методів GBM, PLS, SVM и RF. Найкращі результати показав метод RF для критичних температур ($R^2_{5f} = 0.93$, $RMSE_{5f} = 23.2$ К) та критичних тисків ($R^2_{5f} = 0.81$, $RMSE_{5f} = 24.6$ бар). Для критичних об'ємів найкращі результати показав метод SVM ($R^2_{5f} = 0.93$, $RMSE_{5f} = 33$ см³/моль). Помилка прогнозу MAPE є порівнянною с експериментальною MAPE. Для критичних температур $MAPE_{5f} = 6.8\%$, $MAPE_{exp} = \sim 5\%$; для критичних тисків $MAPE_{5f} = 11.5\%$, $MAPE_{exp} = \sim 10\%$; для критичних об'ємів $MAPE_{5f} = 14.6\%$, $MAPE_{exp} = \sim 10\%$.

В даному випадку критичні об'єми та критичні температури є найбільш точно змодельованими властивостями. Для критичних температур це може бути пояснено високою точністю експериментальних даних, щодо критичних об'ємів, то вони є адитивною властивістю. Для критичних тисків експериментальна точність вимірювання є найнижчою.

Аналіз важливості змінних за Бріманом для сумішей (рис. 3) висвітлює наступні тенденції. Для критичних об'ємів найбільше значення параметра важливості спостерігається для електронної поляризованості. Як електронна поляризованість, так і критичні об'єми є величинами, що мають розмірність об'єму. Найбільш важливими для критичних температур та тисків є вандерваальсові взаємодії (що включає парні міжмолекулярні дескриптори), потім електростатичні взаємодії. Незначний внесок дескрипторів, що описують водневий зв'язок, може бути пояснений невеликою кількістю сполук, що здатні до його утворення – від 11 до 22% всієї вибірки. Порівняння з важливістю змінних для моделей індивідуальних речовин показує схожість тенденцій.

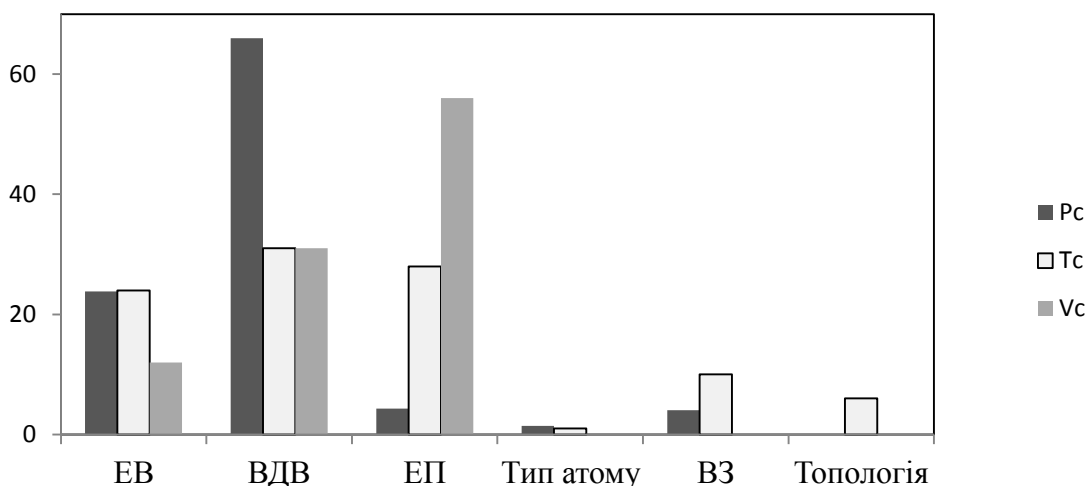


Рис. 3. Відносна важливість змінних (%) в QSPR моделі для критичних властивостей сумішей. (ЕВ – електростатична взаємодія, ВДВ – вандерваальсові взаємодії, ЕП – електронна поляризованість, ВЗ – водневий зв'язок)

Внесок парних міжмолекулярних дескрипторів для критичних температур складає 7%, для критичних об'ємів – 1%, для критичних тисків – 34%. Значний вклад парних

дескрипторів для критичних тисків може бути зумовлений тим, що величина критичного тиску найбільшою мірою залежить від здатності сполук до асоціації.

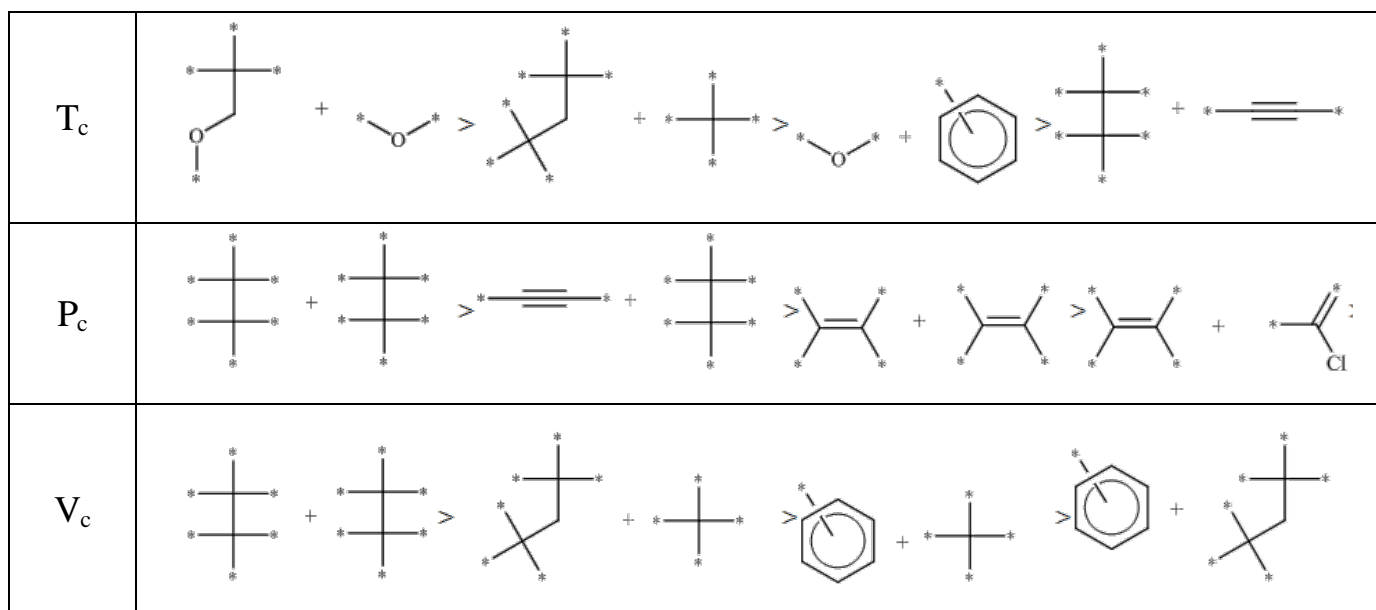


Рис. 4. Вплив деяких фрагментів сумішей органічних сполук на їх критичні властивості (T_c , P_c , V_c).

Структурна інтерпретація виявила найбільш впливові сумішеві фрагменти (рис. 4). Для всіх властивостей найбільший внесок роблять розгалужені вуглеводневі фрагменти, що відповідає відомим експериментальним даним, адже зі зростанням довжини та розгалуженості вуглеводневого ланцюга компонентів суміші значення її критичних властивостей зростають. Для критичних температур найбільший позитивний внесок мають оксигенвмісні фрагменти (а також наявність в суміші ароматичних та оксигенвмісних сполук), при наявності яких в суміші можуть протікати процеси асоціації. Для критичних тисків сумішей, як і для критичних тисків індивідуальних сполук важливими є фрагменти з ненасиченими зв'язками карбон-карбон. Для критичних об'ємів, як і в випадку індивідуальних сполук, найімовірніше великий вплив фрагментів зумовлений розмірами фрагментів.

3.2. Другі віріальні коефіцієнти індивідуальних органічних речовин та їх бінарних сумішей

3.2.01. Другі віріальні коефіцієнти (B) індивідуальних органічних речовин

Інтерес до ДВК визначається по-перше тим, що це важлива частина рівняння стану, і її значення доволі точно визначаються на основі експериментальних PVT -даних, по-друге, ДВК безпосередньо зв'язаний з енергією взаємодії з рівнянням:

$$B = 2\pi N \int \left(e^{-\frac{U}{kT}} - 1 \right) r^2 dr.$$

З цього рівняння витікає, що ті структурні залежності, що впливають на ДВК, таким же чином впливають на енергію взаємодії.

Вибірка по ДВК містить 250 сполук (Dymond J., 2002) різних класів. З урахуванням даних при різних температурах загальна кількість точок складає понад 4500 точок. Попередній аналіз даних показує їх значну кластеризацію – галогенвмісні

сполуки та вуглеводні мають вищі значення B ; нітроген-, оксиген- та сульфуровмісні сполуки у середньому мають нижчі значення B .

Фазові рівноваги описуються великою кількістю рівнянь стану, одним з найвідоміших є рівняння Ван дер Ваальса:

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT,$$

де V – мольний об'єм, a та b – константи, характерні для кожного реального газу, P – тиск, T – температура, R – універсальна газова константа.

Нами запропоновано підхід до моделювання температурно-залежних властивостей, що базується на рівнянні Клаузіуса, що використовує константи з рівняння стану Ван дер ваальса для реальних газів: $B = b - \exp(a/RT)$, де a , b – константи з рівняння Ван дер Ваальса; T – температура. Якість апроксимації $B=f(a, b)$ рівняння була високою – R^2 більше 0.95. Після розрахунку коефіцієнтів a та b шляхом апроксимації для них будувалися окремі QSPR моделі, з яких згодом розраховувався B : $B = f(a, b)$; $a = f(D_1, D_2, \dots, D_n)$; $b = f(D_1, D_2, \dots, D_m)$, де D – дескриптори. При побудованні моделей використовувалися як стандартні симплексні дескриптори, так і «псевдосумішеві». Для моделі «однієї молекули» $R^2_{5f} - 0.78$, $RMSE_{5f} - 255 \pm 28$, для «псевдосумішевих» моделей $R^2_{5f} - 0.81$, $RMSE_{5f} - 185 \pm 25$. Таким чином, використання «псевдосумішевих» дескрипторів дозволяє статистично значимо покращити якість прогнозу. Таке покращення пояснюється природою досліджуваної властивості – другі віріальні коефіцієнти описують парні міжмолекулярні взаємодії, а «псевдосумішеві» дескриптори спрямовані на знаходження усіх можливо взаємодіючих незв'язаних фрагментів. Огляд залежності між помилками даних та помилками моделі показує, що дані з меншою помилкою також мають меншу помилку прогнозу, що є закономірним. Для нітроген- та оксигенвмісних сполук подвійний ріст помилки даних призводить майже до 4-кратного зростання помилки прогнозу. У свою чергу більші експериментальні помилки можуть бути пояснені нижчою термальною стабільністю та складнішими міжмолекулярними взаємодіями цих сполук. Отримані прогнози були порівнянні з розрахунками різними термодинамічними методами (табл. 2). Для більшості сполук прогнози, отримані з QSPR моделей, значно точніші.

Фінальна «псевдосумішева» модель була обрана для аналізу важливості змінних. Коефіцієнти a та b мають чіткий фізичний сенс, a є мірою притягання між частинками, b – виключений об'єм (який є результатом відштовхування). Обидві ці величини насамперед залежать від електронної поляризованості, потім від електростатичних та вандерваальсових взаємодій. Отже, аналіз важливості факторів показує, що отримання фізично значимих моделей можливо з використанням 2D дескрипторів. Структурна інтерпретація проводилася окремо для коефіцієнтів a та b . Найбільший позитивний вплив на коефіцієнт a справляють фрагменти з ненасиченими зв'язками карбон-карбон та галогенвмісні фрагменти; найбільший негативний внесок роблять ароматичні та неароматичні циклічні фрагменти та оксигенвмісні фрагменти. За своїм фізичним змістом коефіцієнт a відображає міжмолекулярне притягання. У свою чергу для коефіцієнта b тенденції прямо протилежні. Та фрагменти, здатні до асоціації (за рахунок утворення водневих зв'язків, стекінгових взаємодій, т.ін.), знижують енергію відштовхування. Оскільки

коефіцієнт b описує молекулярний об'єм, об'ємні та/або здатні до асоціації фрагменти мають позитивний вплив.

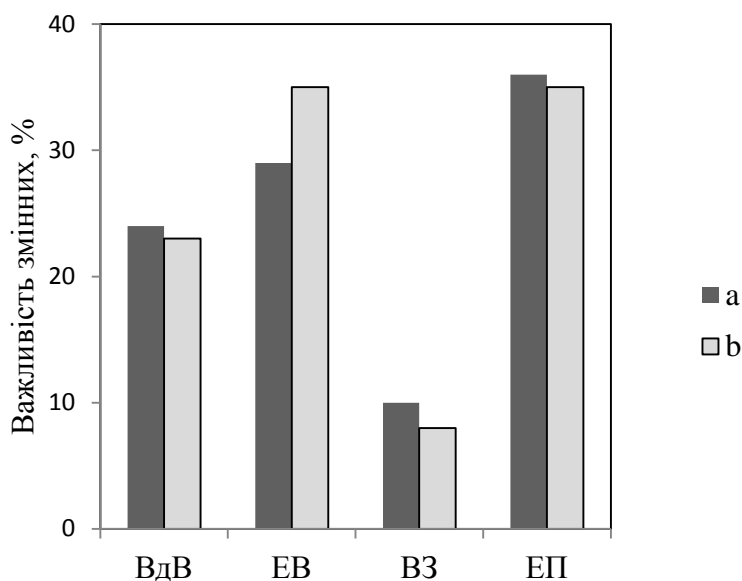


Рис. 5. Відносна важливість змінних (%) для других віріальних коефіцієнтів індивідуальних органічних речовин. (ЕВ – електростатична взаємодія, ВдВ – вандерваальсові взаємодії, ЕП – електронна поляризованість, ВЗ – водневий зв'язок)

Таблиця 2

Порівняння помилок (AAD та RMS) побудованих QSPR моделей для B з іншими підходами

Сполука	AAD PC	RMS PC	AAD TD	RMS TD	AAD ECS	RMS ECS	AAD СПМС	RMS СПМС
Етин	2.4	2.6	2.1	2.5	2.1	2.3	0.04	0.2
1,3-Диметилбензен	6.4	8	7	8.1	4.8	5.7	3.6	8
1,4-Диметилбензен	7.8	8.6	4.9	6.6	5.2	5.8	0.2	0.8
Фенол	10	10	6.2	6.3	4.9	5.5	0.2	0.8
Фторометан	11	11	7.2	7.4	9.4	9.5	0.1	0.3
1,2-Дихлороетан	6	9.8	7.2	9.4	7.7	9	1.6	5.7
Бромометан	7.1	9.1	14	17	6.9	9.1	1.7	7.3
Диметиловий етер	14	15	6.4	9.5	15	16	0.6	2.8
Метанол	24	29	15	17	22	25	0.2	1.5
Етанол	23	28	15	18	18	20	0.1	0.4
1-Пропанол	10	13	10	13	14	16	0.9	4.2
2-Пропанол	9.9	12	10	12	16	18	0.8	3.7
2-Метил-1-пропанол	4.9	7.3	23	23	19	20	3.2	6.5
2-Метил-2-	8	10	25	25	22	25	3.3	6.7

пропанол								
1-Бутанол	8.9	11	7.1	9.7	19	23	0.8	2.9
(RS)-2-Бутанол	15	16	17	17	9.7	11	3.6	7.2
Ацетонітрил	54	55	7.6	9.9	13	14	0.6	3.8
Нітрометан	49	49	3.7	4.5	14	14	0.02	0.1

PC – Pitzer, Curl, 1975; TD – Tarakad, Danner, 1977; ECS – розширений принцип відповідних станів (extended corresponding state principle) – Xiang, 2002.

3.2.02. Другі віріальні крос-коефіцієнти (B_{12}) бінарних сумішей органічних сполук

Другий віріальний крос-коефіцієнт (B_{12}) – це розрахована властивість, що визначається виключно різнорідними взаємодіями компонентів в суміші. На відміну від інших властивостей сумішей другий віріальний крос-коефіцієнт характеризує тільки взаємодію молекул типу 1 з молекулами типу 2. Вибірка по другому віріальному крос-коефіцієнту включала 126 сумішей сполук різних класів та 1211 значень B_{12} при різних температурах (Dymond J., 2002). Суміші, для яких було менше 4 значень, були включені в зовнішню тестову вибірку, яка включала 102 суміші з 188 значеннями при різних температурах. Біля 798 точок доступні для сумішей НП-НП; 302 – для сумішей НП-ПЛ та 108 точок – для сумішей ПЛ-ПЛ. Як і індивідуальні полярні речовини, полярні суміші характеризуються нижчими значеннями B_{12} .

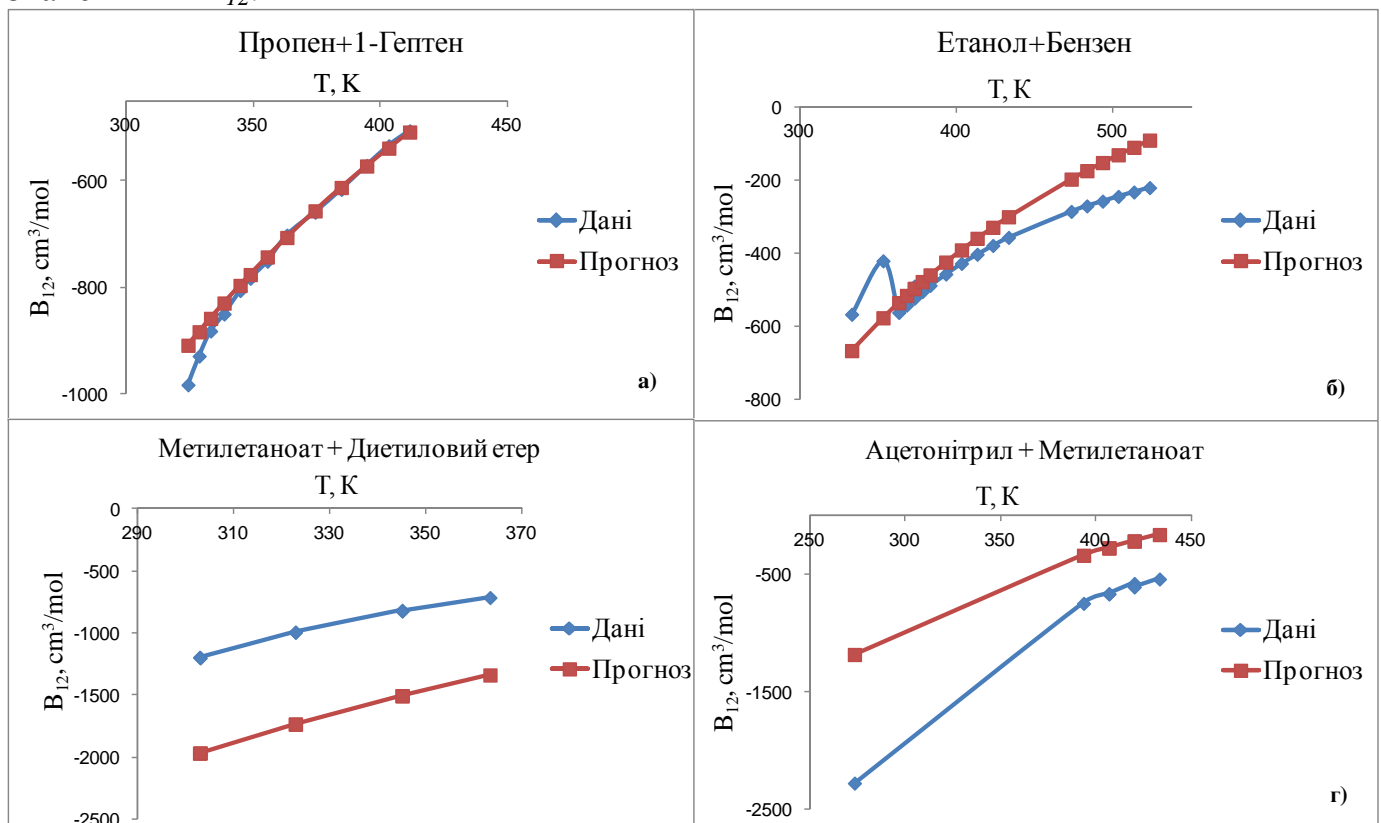


Рис. 6. Приклади відновлених температурних кривих B_{12} для сумішей різного типу (а – суміш НП-НП; б – суміш НП-ПЛ; в, г – ПЛ-ПЛ)

Згідно з вищеописаною процедури було розраховано коефіцієнти a та b , використовуючи апроксимацію для B_{12} . Якість апроксимації була високою з

коефіцієнтом детермінації близько 0.90. Для коефіцієнтів a та b були отримані моделі методом GBM з R^2_{5f} відповідно 0.52 та 0.56. Для найкращих моделей для B_{12} , побудованих методом GBM, R^2_{5f} дорівнює 0.75, $RMSE_{5f} = 253.1 \text{ см}^3/\text{моль}$. Для зовнішньої тестової вибірки $R^2_{ts} = 0.52$, $RMSE_{ts} = 341.5 \text{ см}^3/\text{моль}$; в межах DA $R^2_{ts} = 0.65$, $RMSE_{ts} = 224.5 \text{ см}^3/\text{моль}$ (при цьому покриття тестової вибірки DA моделі дорівнює 92%).

Помилка прогнозу 5-кратної зовнішньої крос-валідації збільшується від неполярних сумішей до полярних симбатно з помилками даних. Для неполярних сумішей помилки даних приблизно $25 \text{ см}^3/\text{моль}$, у той час як помилка прогнозу $80 \text{ см}^3/\text{моль}$. Середня помилка для сумішей неполярна-полярна вища – близько $55 \text{ см}^3/\text{моль}$, середня помилка прогнозу – $150 \text{ см}^3/\text{моль}$. Отже, найвища помилка даних для полярних сумішей – близько $90 \text{ см}^3/\text{моль}$, помилка прогнозу також найвища – близько $250 \text{ см}^3/\text{моль}$. З отриманих моделей були відновлені температурні криві для другого віріального крос-коефіцієнта (рис. 6). Найкраще відтворюються температурні криві для неполярних сумішей та для сумішей полярна-неполярна речовина. Результат розрахунку є гладкою кривою, що свідчить про коректне відображення теоретичної залежності.

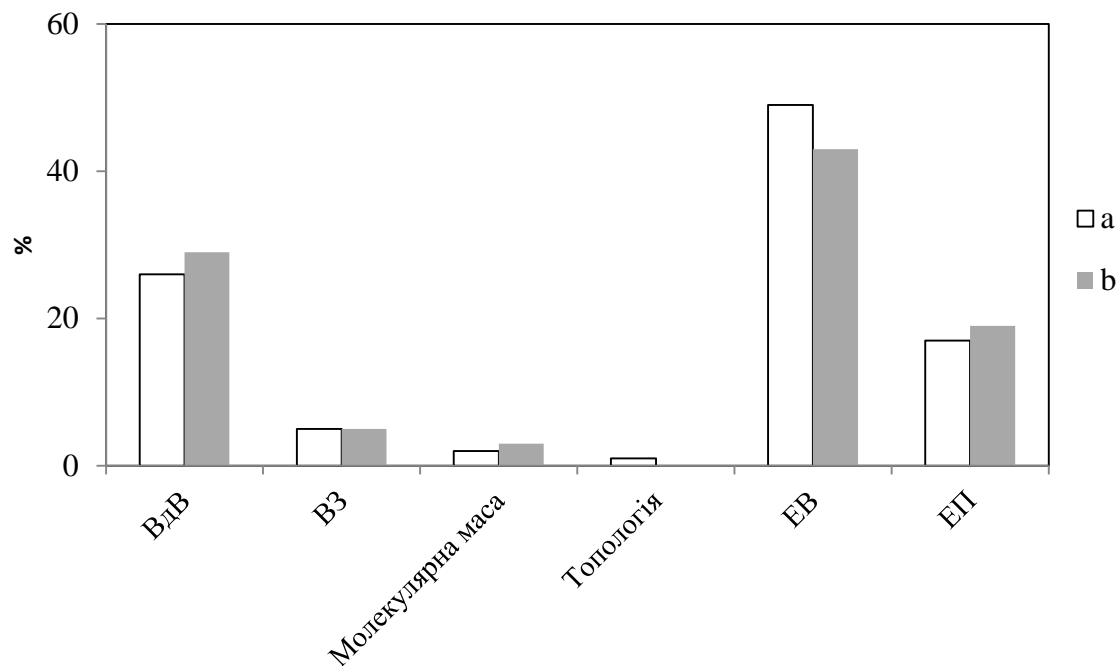


Рис. 7. Відносна важливість змінних (%) для B_{12} . (ЕВ – електростатична взаємодія, ВДВ – вандерваальсові взаємодії, ЕП – електронна поляризованість, ВЗ – водневий зв'язок)

Інтерпретація згідно з найбільш важливими факторами (рис. 7) для двошарової моделі показує домінуючий вплив електростатичних факторів, менший вплив вандерваальсових взаємодій (що описуються парними дескрипторами). На відміну від індивідуальних речовин вклад електронної поляризованості найменший. Другий віріальний крос-коефіцієнт також напряму зв'язаний з енергією міжмолекулярних взаємодій $U(R_{12})$, яка в даному випадку описує взаємодію молекули 1 з молекулою 2 в термінах міжмолекулярної відстані R_{12} . З теорії міжмолекулярних взаємодій відомо, що на великих відстанях молекули притягуються у зв'язку з дисперсійними взаємодіями; а на малих відстанях відштовхування превалює у зв'язку з обмінними взаємодіями. Постійні електричні

моменти мають найбільший вплив. Таким чином, розраховані значення важливостей змінних узгоджуються з загальнотеоретичним описом другого віріального крос-коефіцієнта.

Результати структурної інтерпретації наведені на рис. 8. Для коефіцієнта a для сумішей, як і для індивідуальних сполук, найбільш впливовими є фрагменти з ненасиченими подвійними/потрійними карбон-карбовоними зв'язками, а також наявність суміші одного ароматичного компонента та одного галогенвмісного компонента, що може призводити до утворення відносно стабільних π -комплексів. Ряд для коефіцієнта b показує схожі тенденції, окрім ароматичних та галогенвмісних фрагментів важливий внесок також вносять розгалужені вуглеводневі фрагменти, що може бути пов'язано з їх впливом на сумарний об'єм суміші, що загалом повторює тенденції впливу для індивідуальних сполук.

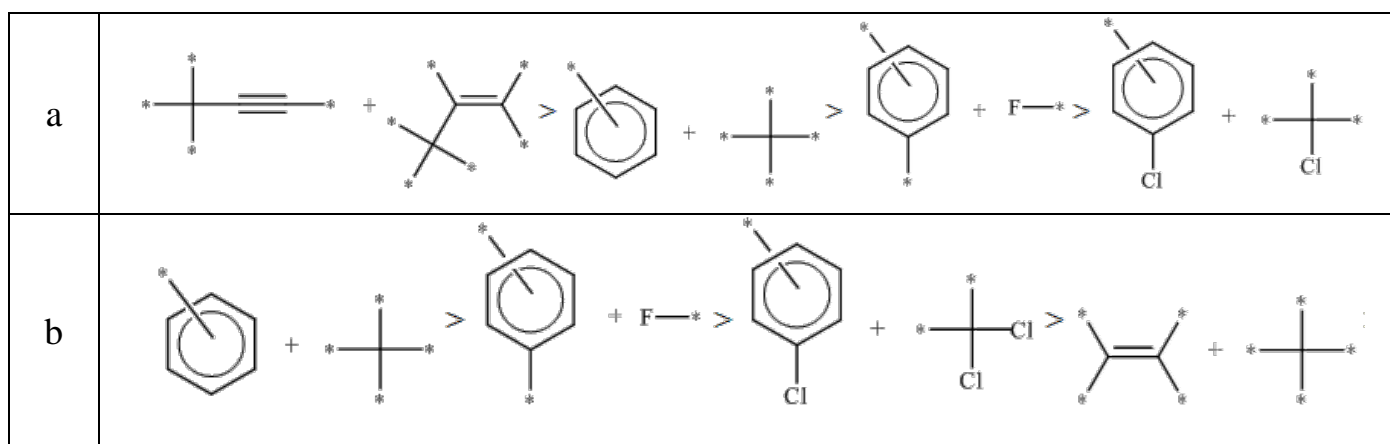


Рис. 8. Вплив деяких фрагментів сумішей органічних сполук на їх другий віріальний крос-коефіцієнт (B_{12})

ВИСНОВКИ

Розроблена система моделей, що адекватно прогнозує критичні властивості органічних речовин та їх бінарних сумішей. Показана ефективність застосування новітніх дескрипторів міжмолекулярних взаємодій для побудування моделей властивостей, що безпосередньо залежать від енергії взаємодії між молекулами. Показано, що 2D-дескриптори здатні описувати складні фундаментальні властивості, прогнозувати їх з помилкою близькою до експериментальної.

1. Використання «псевдосумішевих» дескрипторів у QSPR моделях дозволяє зменшити середньоквадратичну помилку прогнозування критичних властивостей та других віріальних коефіцієнтів індивідуальних органічних сполук різних класів. Оцінка важливості дескрипторів парних взаємодій у QSPR моделях щодо критичних властивостей сумішей показала, що найбільший вклад ці дескриптори вносять в опис критичних тисків – біля 30%. Для критичних об'ємів навпаки їх вклад не є значимим – біля 1%.
2. На критичні властивості речовин та їх сумішей найбільше впливають електростатичні та вандерваальсові взаємодії. Для критичних температур превалюючу роль грають електростатичні фактори, у той час як для критичних об'ємів істотно зростає роль вандерваальсових взаємодій. Аналіз важливості змінних для критичних властивостей сумішей демонструє провідну роль електронної поляризованості для критичних об'ємів (56%) та незначний вплив

електронної поляризованості на критичні тиски (4%). Показано, що фактор Пітцера залежить не тільки від асферичності молекулярної форми (як вважалося раніше), а передусім від розміру молекули.

3. Для всіх критичних властивостей найбільший внесок роблять розгалужені вуглеводневі фрагменти. Для критичних температур максимальний позитивний вплив виявляють циклічні фрагменти та фрагменти, що містять карбонільну групу, негативний – фрагменти, що містять атоми фтору. Це можна пояснити, виходячи з того, що критична температура зв'язана з енергією міжмолекулярної взаємодії і підвищується при зростанні здатності молекул до взаємопритягання. Для критичного тиску найбільший позитивний вплив має дісульфідна група, аліфатичний фтор, первинна аміногрупа, потрійний карбон-карбонів зв'язок та аліфатична гідроксогрупа. Для критичних об'ємів вплив фрагментів зумовлений розмірами фрагментів. Для сумішей спостерігаються аналогічні тенденції.
4. «Псевдосумішеві» QSPR моделі для других віріальних коефіцієнтів індивідуальних речовин проявили високу прогнозуючу здатність, помилка прогнозу в середньому на порядок нижча, ніж для інших термодинамічних методів. Завдяки використанню незв'язаних симплексів та дескрипторів парної взаємодій розраховані другі віріальні крос-коефіцієнти (B_{12}), які безпосередньо характеризують різні міжмолекулярні взаємодії в сумішах. Розраховані криві залежності $B_{12} = f(T)$ у більшості випадків адекватно відтворюють дані експерименту.
5. Структурна інтерпретація параметрів рівняння Ван-дер-Ваальса a та b , на основі яких розраховуються другі віріальні коефіцієнти, показала наступні тенденції. Найбільший позитивний вплив на коефіцієнт a справляють фрагменти з ненасиченими зв'язками карбон-карбон та галогенвмісні фрагменти; найбільший негативний вклад вносять ароматичні та неароматичні циклічні фрагменти та оксигенвмісні фрагменти, які сприяють асоціації молекул (за рахунок утворення водневих зв'язків, стекінгових взаємодій, т.ін.) та знижують енергію відштовхування. Об'ємні та/або здатні до асоціації фрагменти мають позитивний вплив на коефіцієнт b , який залежить від молекулярного об'єму. Для сумішей вплив структурних фрагментів аналогічний, але для коефіцієнта a наявність в суміші одного ароматичного компонента та одного галогенвмісного компонента, що може призводити до утворення відносно стабільних комплексів.

ПЕРЕЛІК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Мокшина Е.Г. QSPR моделирование критических параметров органических веществ различных классов в рамках симплексного представления молекулярной структуры. / Е.Г. Мокшина, В.Е. Кузьмин, В.И. Недоступ // Журнал Органической Химии. – 2014. – Т. 50, №3. – С. 326-333 (*Особистий внесок автора – побудування моделей для критичних параметрів індивідуальних сполук, обробка та участь у інтерпретації отриманих результатів*).
2. Мокшина Е.Г. Метод определения критических параметров и ацентрического фактора многоатомных экспериментально неизученных веществ / Е.Г. Мокшина, В.Е. Кузьмин, В.И. Недоступ // Технические газы. – 2014. – № 1. – С. 52-58. (*Особистий внесок автора – побудування моделей для критичних параметрів індивідуальних сполук, обробка та участь в інтерпретації отриманих результатів*).
3. Mokshyna E. 'Quasi-Mixture' Descriptors for QSPR Analysis of Molecular Macroscopic Properties. The Critical Properties of Organic Compounds / E. Mokshyna, V.I. Nedostup,

P.G. Polishchuk, V.E. Kuz'min // Mol. Inf. – 2014. – V.33, I.10. – P. 647-654. (*Особистий внесок автора – розробка «псевдосумішевих» дескрипторів, побудування моделей для критичних параметрів індивідуальних сполук, обробка та участь в інтерпретації отриманих результатів*).

4. Mokshyna E. Predictive QSPR modelling for the second virial coefficient of the pure organic compounds / E. Mokshyna, V.I. Nedostup, P.G. Polishchuk, V.E. Kuz'min // Mol. Inf. – 2015. – V.34, I.1. – P. 53-59. (*Особистий внесок автора – побудування моделей для других віриальних коефіцієнтів індивідуальних сполук, обробка та участь в інтерпретації отриманих результатів*).
5. Мокшина Е.Г. QSPR моделирование критических свойств бинарных смесей органических соединений/ Е.Г. Мокшина, П.Г. Полищук, В.И. Недоступ, В.Е. Кузьмин // Журнал Органической Химии – 2016 – Т.52, №1. – С. 13-18. (*Особистий внесок автора – побудування моделей для критичних параметрів бінарних сумішей, обробка та участь в інтерпретації отриманих результатів*)
6. Mokshyna E.G. QSPR modelling for the Second Virial Cross-coefficients of the Binary Organic Mixtures / Mokshyna E.G., Polishchuk P.G., Nedostup V.I., Kuz'min V.E. // International Journal of Quantitative Structure-Property Relations – V.I, I.1. – P. 72-84. (*Особистий внесок автора – побудування моделей для других віриальних коефіцієнтів бінарних сумішей, обробка та участь в інтерпретації отриманих результатів*).
7. Sizochenko N. Causal inference methods to assist in mechanistic interpretation of classification nano-SAR models / N. Sizochenko, B. Rasulev, A. Gajewicz, E. Mokshyna, V. Kuz'min, J. Leszczynski, T. Puzyn// Royal Soc. Chem. Adv. – 2015. – No.5. – P. 77739-77745. (*Особистий внесок автора – адаптація методів каузального аналізу для задач хемоінформатики, розрахунки методами каузального аналізу, участь в інтерпретації результатів*)
8. Мокшина Е.Г. QSPR моделирование критических параметров органических веществ различных классов // XIV Конференция молодых ученых и студентов-химиков Южного региона Украины: Сборник тезисов. – Одесса, 2012. – С. 23.
9. Мокшина Е.Г. Модели прогнозирования критической температуры органических веществ, основанные на симплексном представлении молекулярной структуры // “Хімічні Каразінські Читання”: IV Всеукраїнська наукова конференція студентів і аспірантів: Тези доповідей. – Харків, 2012. – С. 195-196.
10. Mokshina E. “Quasi-mixture”-based QSPR study of critical parameters of pure organic compounds /E. Mokshina, A. G. Artemenko, P.G. Polishchuk, V. I. Nedostup, V. E. Kuz'min // 3rd Strasbourg Summer School in Chemoinformatics. – Strasbourg, 2012.
11. Mokshyna E.G. Study of second virial coefficient: new challenge for QSPR / Mokshyna E.G., Kuz'min V.E., Nedostup V.I. // Methods and Applications of Computational Chemistry. 5th International Symposium. – Kharkiv, 2013. – P. 48.
12. Mokshina E.G. Critical properties: QSPR of binary organic mixtures / E.G. Mokshina, P.G. Polishchuk, V.I. Nedostup, V.E. Kuz'min// Kazan Summer School on Chemoinformatics. – Kazan, 2013. – P. 24
13. Мокшина Е.Г. QSPR анализ вторых вириальных коэффициентов бинарных органических смесей / Е.Г. Мокшина, П.Г. Полищук//XVI конференция молодых ученых и студентов-химиков Южного региона Украины с международным участием. – 2014. – С. 20.
14. Мокшина О.Г. Фактор Пітцера та його зв'язок з структурою органічних сполук: дослідницький SPR аналіз // О.Г. Мокшина, В.І. Недоступ, В.Є. Кузьмін // “Хімічні Каразінські Читання”: VII Всеукраїнська наукова конференція студентів і аспірантів: Тези доповідей. – Харків, 2015 – С. 225.

15. Мокшина Е.Г. QSPR моделирование термодинамических свойств бинарных смесей органических соединений на примере критических свойств и вторых вириальных кросс-коэффициентов / Е.Г. Мокшина // XVII Конференция молодых ученых и студентов химиков Южного региона Украины: Сборник тезисов. – Одесса, 2015. – С. 21.
16. Mokshyna E. Atom-pair descriptors, temperature-dependent property and binary mixtures: QSPR for second virial cross-coefficients / E. Mokshyna, P. Polishchuk, V. Nedostup, V. Kuz'min // II Kazan Summer School on Chemoinformatics. – Kazan, 2015. – P. 40.
17. Moshyna E. Influence of Molecular Shape and Size on Acentric Factor: Causal Analysis / E. Mokshyna // XVIII Scientific Youth Conference "Problems and Achievements of Modern Chemistry". – Odessa, 2016. – P. 13.

АНОТАЦІЯ

Мокшина О.Г. Критичні властивості та другі віріальні коефіцієнти органічних речовин та їх сумішей. QSPR аналіз з урахуванням міжмолекулярних взаємодій. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата хімічних наук за спеціальністю 02.00.03 – органічна хімія. Фізико-хімічний інститут ім. О.В. Богатського НАН України, Одеса, 2015.

Дисертаційна робота присвячена розробці системи прогнозування термодинамічних властивостей органічних речовин та їх бінарних сумішей на основі QSPR аналізу у рамках симплексного підходу до опису молекулярної структури на прикладі критичних властивостей, фактора Пітцера та других віріальних коефіцієнтів.

Розроблено дескрипторний опис індивідуальних органічних сполук, що неявно відображує міжмолекулярні взаємодії – «псевдосумішевий» підхід та дескриптори парних взаємодій. Розроблено та валідовано QSPR моделі, що описують та прогнозують досліджувані термодинамічні властивості органічних сполук та їх бінарних сумішей. Проведено каузальний аналіз молекулярних параметрів, що впливають на фактор Пітцера. Показано, що використання «псевдосумішевих» дескрипторів у QSPR моделях дозволяє зменшити середньоквадратичну помилку (RMSE) прогнозування критичних властивостей та других віріальних коефіцієнтів індивідуальних органічних сполук різних класів. Визначені структурні та фізико-хімічні фактори, що впливають на досліджувані властивості. Вперше проведено розрахункову оцінку других віріальних крос-коефіцієнтів (B_{12}), які безпосередньо характеризують різномірні міжмолекулярні взаємодії у сумішах. Розраховані криві залежності $B_{12} = f(T)$. За допомогою зовнішньої валідації показано, що отримані моделі придатні для прогнозування властивостей нових сполук та раніше недосліджених сумішей.

Ключові слова: QSPR, симплексне представлення молекулярної структури, дескриптори сумішей, термодинамічні властивості.

АННОТАЦИЯ

Мокшина Е.Г. Критические свойства и вторые вириальные коэффициенты органических веществ и их смесей. QSPR анализ с учетом межмолекулярных взаимодействий. – Рукопись.

Диссертация на соискание ученой степени кандидата химических наук по специальности 02.00.03 – органическая химия. Физико-химический институт им. А.В. Богатского НАН Украины, Одесса, 2015.

Диссертационная работа посвящена разработке системы прогнозирования термодинамических свойств органических веществ и их бинарных смесей на основе QSPR анализа в рамках симплексного подхода к описанию молекулярной структуры на примере критических свойств, фактора Питцера и вторых вириальных коэффициентов.

Разработано дескрипторное описание индивидуальных органических соединений, которое неявно отображает межмолекулярные взаимодействия – «псевдосмесевой» подход и дескрипторы парных взаимодействий. Построены и валидированы QSPR модели, которые описывают и прогнозируют исследуемые термодинамические свойства органических соединений и их бинарных смесей. Проведен каузальный анализ молекулярных параметров, которые влияют на фактор Питцера. Показано, что использование «псевдосмесевых» дескрипторов в QSPR моделях позволяет уменьшить среднеквадратичную ошибку (RMSE) прогноза критических свойств и вторых вириальных коэффициентов индивидуальных органических соединений разных классов. Определены структурные и физико-химические факторы, которые влияют на исследуемые свойства. Впервые проведена расчетная оценка вторых вириальных кросс-коэффициентов (B_{12}), которые непосредственно характеризуют разнородные межмолекулярные взаимодействия в смесях. Рассчитаны кривые зависимости $B_{12} = f(T)$. С помощью внешней валидации показано, что полученные модели пригодны для прогнозирования свойств новых соединений и ранее неисследованных смесей.

Ключевые слова: QSPR, симплексное представление молекулярной структуры, дескрипторы смесей, термодинамические свойства.

SUMMARY

Mokshyna O.G. Critical properties and second virial coefficients of organic compounds and their mixtures. QSPR analysis with the account for the intermolecular interactions. – Manuscript.

Thesis of Candidate of Sciences (PhD) in specialty 02.00.03 – organic chemistry. O.V. Bogatsky Physico-Chemical Institute NAS of Ukraine, Odessa, 2016.

The main purpose of the thesis was to develop a system to predict thermodynamic properties (critical properties, acentric factor, second virial coefficient) of organic compounds and their binary mixtures using QSPR analysis based on the simplex representation of molecular structure.

Novel descriptor representations were developed able to account for intermolecular interactions – “quasi-mixture” approach and pair interaction descriptors. The developed QSPR models were validated and proved able to describe and predict investigated thermodynamic properties of individual compounds and their binary mixtures. The causal analysis was performed for molecular parameters that influence the acentric factor. Usage of “quasi-mixture” descriptors proved to decrease a root-mean-square prediction error for critical properties and second virial coefficients of individual compounds. Influential structural and physic-chemical factors were detected. For the first time the second virial cross-coefficients (B_{12}) were successfully modeled and predicted. Temperature dependence curves for B_{12} were restored from predictions and compared to the initial data. The external validation showed that the obtained models are applicable towards a prediction of new compounds and previously unstudied mixtures.

Key words: QSPR, simplex representation of molecular structure, descriptors of mixtures, thermodynamic properties.