

**МІНІСТЕРСТВО ОХОРОНИ ЗДОРОВ'Я
НАЦІОНАЛЬНИЙ МЕДИЧНИЙ УНІВЕРСИТЕТ
імені О.О. БОГОМОЛЬЦЯ**

**ФАРМАЦЕВТИЧНИЙ ФАКУЛЬТЕТ
КАФЕДРА АНАЛІТИЧНОЇ, ФІЗИЧНОЇ ТА КОЛОЇДНОЇ ХІМІЇ**

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА

**на тему «QSPR ПРОГНОЗУВАННЯ ВОДНОЇ РОЗЧИННОСТІ
ОРГАНІЧНИХ РЕЧОВИН: ФАРМАЦЕВТИЧНО АКТИВНИХ ТА
ДОПОМІЖНИХ»**

Виконав: здобувач вищої освіти 5 курсу
групи 10804
напряму підготовки 22 «Охорона здоров'я»
спеціальності 226 «Фармація, промислова
фармація»
освітньої програми «Фармація»
Ізмаїлов Раїм Муслімович

Керівник: кандидат хімічних наук, доцент
Пушкарьова Я.М.

Рецензент: кандидат медичних наук, доцент
Половинка В.О.

КИЇВ-2025

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	3
ВСТУП	4
ОСНОВНА ЧАСТИНА.....	7
1. ЛІТЕРАТУРНИЙ ОГЛЯД	7
1.1. Використання QSPR моделей у фармацевтичному прогнозуванні	7
1.2. Експериментальні методи визначення водної розчинності.....	9
Висновки до розділу 1	12
2. ЕКСПЕРИМЕНТАЛЬНА ЧАСТИНА	13
2.1. Масив даних для дослідження	13
2.2. Реалізація методології QSPR	23
2.2.1. Перевірка дескрипторів на мультиколінеарність	23
2.2.2. Побудова двопараметричної QSPR моделі	28
2.2.3. Побудова трипараметричної QSPR моделі	33
2.2.4. Побудова чотирипараметричної QSPR моделі	38
Висновки до розділу 2	39
ВИСНОВКИ.....	40
СПИСОК ЛІТЕРАТУРИ.....	41
ДОДАТОК.....	43
SUMMARY	44

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

QSPR	Quantitative Structure-Property Relationship (метод кількісного зв'язку між структурою сполуки та її фізико-хімічними властивостями)
$\log S_w$	логарифм експериментальних значень водної розчинності органічних речовин
R^2	коефіцієнт детермінації
R^2_{adj}	скоригований (adjusted) коефіцієнт детермінації
R	коефіцієнт кореляції
% RE	відносна похибка залишків у відсотках

ВСТУП

Актуальність теми. Водна розчинність органічних сполук є важливою молекулярною властивістю, яка відіграє значну роль у поведінці сполук у багатьох сферах. Враховуючи важливість розчинності, метод її прогнозування, заснований виключно на молекулярній структурі, може стати корисним інструментом, оскільки існує багато сполук, для яких значення розчинності просто невідоме.

Розчинність хімічних речовин і лікарських засобів у водному середовищі суттєво впливає на ступінь їх абсорбції та транспортування в організмі. Саме тому розчинність вважається дуже важливим параметром у сучасних дослідженнях ADMET (абсорбція, розподіл, метаболізм, виведення та токсичність).

Водна розчинність відіграє ключову роль у таких сферах фармації, як:

- розробка лікарських форм – впливає на вибір допоміжних речовин і методів формування дозованих форм;
- біодоступність лікарських засобів – визначає швидкість і ступінь всмоктування активної речовини в організмі;
- фармакокінетика – впливає на розподіл, метаболізм і виведення препаратів;
- контроль якості ліків – важливий параметр при стандартизації та сертифікації фармацевтичних продуктів;
- створення ін'єкційних та офтальмологічних лікарських форм – забезпечує стабільність і ефективність розчинів;
- формування стратегії доставки лікарських засобів – впливає на вибір носіїв для таргетного транспортування;
- оцінка екологічної безпеки фармацевтичних відходів – визначає їхній вплив на водні екосистеми.

Широка значущість даних про водну розчинність для різних галузей науки, медицини, технологій та інженерії сприяла розробці численних

моделей для її прогнозування. Тому доцільною є розробка моделі для прогнозування водної розчинності, використовуючи лише теоретично отримані дескриптори. У порівнянні з трудомісткими експериментальними процедурами для безпосереднього визначення водної розчинності, надійні комп'ютерні методи прогнозування стали більш популярними та необхідними в сучасних дослідженнях.

Мета і завдання дослідження. Мета дослідження – побудувати QSPR модель для прогнозування водної розчинності органічних сполук, що мають фармацевтичне значення, на основі теоретично обчислених молекулярних дескрипторів.

Завдання дослідження:

1) формування масиву даних для дослідження, що включає експериментальні значення водної розчинності органічних речовин та їх молекулярних дескрипторів;

2) аналіз значущості молекулярних дескрипторів та їхнього впливу на прогнозування водної розчинності органічних речовин з метою розробки оптимальної QSPR моделі;

3) оцінка якості прогностичної моделі за допомогою статистичних показників.

Методи дослідження: множинна лінійна регресія – для побудови QSPR моделі; кореляційний аналіз – для визначення сили взаємозв'язку між молекулярними дескрипторами і водною розчинністю; статистичний аналіз – для оцінки якості побудованої моделі за допомогою статистичних показників (коефіцієнт детермінації, скоригований коефіцієнт детермінації, залишки і відносна похибка залишків).

Програмне забезпечення – Matlab R2024b (trial license), Microsoft Excel 2010.

Новизна та значення одержаних результатів. Розроблено трипараметричну QSPR модель для прогнозування водної розчинності органічних речовин. Оцінка впливу різних типів молекулярних дескрипторів

на точність моделі дозволила обрати найбільш значущі параметри для прогнозування розчинності: логарифм коефіцієнта розподілу, топологічна площа полярної поверхні та молекулярна маса.

Розробка ефективної моделі для прогнозування водної розчинності може значно скоротити час та витрати на розробку нових органічних сполук у фармацевтичній та хімічній галузях, допомагаючи зменшити потребу в експериментальних тестах.

Модель може бути корисною для оптимізації властивостей молекул на ранніх етапах досліджень, що дозволяє спрогнозувати, які саме сполуки мають потенціал для подальших розробок та впровадження в різні галузі, зокрема у фармацевтичну практику.

Апробація результатів дослідження. Результати цієї роботи були представлені на IV Міжнародній науковій конференції «Наукові тренди постіндустріального суспільства», 21 березня 2025, м. Тернопіль, Україна.

Публікації. Опубліковано тези доповідей: Ізмаїлов Р.М., Пушкарьова Я.М. QSPR прогнозування водної розчинності лікоподібних органічних сполук. Матеріали міжнародної наукової конференції «Наукові тренди постіндустріального суспільства», м. Тернопіль, Україна, 21 березня 2025. С. 313-314.

Структура роботи: 45 сторінок, 2 розділи, 1 додаток, список використаних джерел містить 19 публікацій.

ОСНОВНА ЧАСТИНА

1. ЛІТЕРАТУРНИЙ ОГЛЯД

1.1. Використання QSPR моделей у фармацевтичному прогнозуванні

QSPR моделі є потужним інструментом для прогнозування фізико-хімічних властивостей молекул, включаючи водну розчинність. Ці моделі ґрунтуються на аналізі кореляцій між молекулярною структурою та її властивостями, що дозволяє передбачати характеристики нових сполук без потреби в проведенні дорогих і тривалих експериментальних досліджень. Особливо це важливо у фармацевтичному контексті, оскільки оптимальна водна розчинність органічних сполук є ключовим фактором для їхньої біодоступності та ефективності в організмі [1, 2].

Методи QSPR базуються на використанні молекулярних дескрипторів – кількісних характеристик молекул, що описують їхню структуру та фізико-хімічні властивості. Молекулярні дескриптори можуть включати в себе різноманітні параметри, такі як геометричні властивості молекул, їх електронну структуру, полярність, гідрофобність, тощо.

Моделі QSPR використовуються для аналізу та прогнозування численних молекулярних властивостей, серед яких водна розчинність є однією з найбільш досліджених. Вони дозволяють розробляти нові сполуки з передбачуваними властивостями ще на ранніх етапах досліджень, що значно скорочує час і витрати на дослідження [3].

Для фармацевтичних досліджень точне прогнозування водної розчинності є важливим етапом на шляху до створення лікарських засобів з оптимальними фармакокінетичними властивостями. Молекули з низькою водною розчинністю часто мають низьку біодоступність, що ускладнює їхнє ефективне використання в медицині. Тому QSPR моделі, що дозволяють на

ранніх етапах розробки лікарських засобів передбачати водну розчинність нових молекул, стають безцінним інструментом для фармацевтів та дослідників [4, 5].

Завдяки використанню таких моделей можна [6, 7]:

1. Зменшити час та витрати на експериментальні дослідження, що дозволяє більш ефективно відбирати сполуки для подальших стадій розробки.

2. Оптимізувати склад лікарських форм, оскільки точне знання водної розчинності дозволяє створювати лікарські засоби з покращеними фармакокінетичними характеристиками, такими як швидке всмоктування та висока біодоступність.

3. Удосконалити процеси молекулярного дизайну – молекулярні дескриптори можуть бути використані для передбачення не тільки водної розчинності, але й інших властивостей, таких як токсичність, стійкість до метаболізму та взаємодія з біологічними мішенями.

Одним із підходів для підвищення точності прогнозування є використання мультидескрипторних моделей. Вони поєднують декілька молекулярних дескрипторів для створення більш складних і точних моделей. Оскільки водна розчинність молекул є результатом складної взаємодії різних фізико-хімічних факторів, такий підхід дозволяє більш точно передбачати поведінку молекул у водному середовищі, що є надзвичайно важливим для фармацевтичного прогнозування.

Завдяки сучасним досягненням в області обчислювальної хімії, молекулярного моделювання та машинного навчання, майбутнє QSPR методів виглядає надзвичайно перспективним. Ці методи дозволяють здійснювати швидко і точно прогнозування фізико-хімічних властивостей молекул, що є важливими для різних галузей, зокрема для фармацевтики. Сучасні підходи в обчислювальній хімії дозволяють удосконалювати QSPR моделі, які використовують більш складні дескриптори і алгоритми, здатні прогнозувати не лише водну розчинність, але й інші важливі параметри.

Інтеграція QSPR моделей з відкритими та комерційними базами даних, що містять інформацію про величезну кількість молекул і їх властивості, дозволяє значно підвищити якість та масштабованість моделей. Хмарні обчислення також дозволяють виконувати складні обчислення та моделювання без необхідності в потужному локальному апаратному забезпеченні. Це забезпечує швидкий доступ до актуальної інформації і дає можливість для розвитку інтероперабельних платформ, де науковці, фармацевти і розробники можуть спільно використовувати дані для прогнозування властивостей нових молекул [8, 9].

1.2. Експериментальні методи визначення водної розчинності

Експериментальні методи визначення водної розчинності органічних речовин включають кілька підходів, які дозволяють точно виміряти, як різні сполуки взаємодіють з водою. Ось кілька основних методів [10-13]:

1. Метод насичення

Метод насичення є одним з найпоширеніших для вимірювання водної розчинності. Він полягає в тому, що органічна сполука додається до води до тих пір, поки не буде досягнута насичена фаза (тобто, концентрація молекул в розчині більше не збільшується). Після цього розчин відфільтровується або центрифугується, щоб видалити нерозчинені частинки, і вимірюється концентрація розчиненої сполуки.

Переваги:

- Простота виконання.
- Використовуються стандартні лабораторні методи, такі як спектрофотометрія або хроматографія для вимірювання концентрації.

Недоліки:

- Часовитратність, оскільки розчинність може досягати рівноваги тільки після тривалого часу.

- Може бути складно для сполук, які розчиняються дуже повільно або мають низьку розчинність.

2. Метод витягування

Цей метод передбачає розчинення органічної речовини в певному розчиннику (наприклад, органічному розчиннику), а потім перенос її в воду. Використовується для оцінки здатності органічних молекул до розчинення в двофазних системах. Після цього змішуються фази і визначають концентрацію органічної сполуки у водній фазі.

Переваги:

- Швидкість отримання результатів, оскільки процес витягування часто займає менше часу, ніж досягнення насичення.
- Може бути корисним для дослідження систем з малою водною розчинністю.

Недоліки:

- Не підходить для молекул, що дуже швидко переходять в іншу фазу.
- Потрібна точна калібрування та підготовка для визначення рівня концентрації.

3. Спектрофотометрія

Цей метод використовується для визначення концентрації розчиненої сполуки шляхом вимірювання абсорбції світла на певній довжині хвилі. Якщо органічна сполука має специфічну оптичну активність, її концентрацію можна точно визначити через спектрофотометр.

Переваги:

- Висока чутливість, дозволяє вимірювати дуже низькі концентрації.
- Неконкурентний і ефективний метод для аналізу багатьох сполук одночасно.

Недоліки:

- Потрібна наявність молекул, що поглинають світло на певній довжині хвилі.
- Може бути складно для більш складних сполук.

4. Хроматографічні методи

Методи хроматографії, зокрема рідинна хроматографія високої ефективності, також застосовуються для визначення розчинності. Рідинна хроматографія дозволяє розділити молекули за їхніми властивостями взаємодії з розчинником і зібрати точні дані про розчинність сполуки.

Переваги:

- Дуже точні і можуть використовуватися для аналізу складних сумішей.

- Висока чутливість та точність.

Недоліки:

- Складний метод, який потребує спеціалізованого обладнання.

- Потребує довгих підготовчих етапів.

5. Метод рівноважних розчинів

Цей метод ґрунтується на створенні рівноважної концентрації органічної сполуки у воді шляхом змішування сполуки з водою до досягнення рівноваги. Визначення рівноважної концентрації може здійснюватися за допомогою таких методів, як спектрофотометрія або мас-спектрометрія.

Переваги:

- Визначення концентрації на основі рівноважних станів є надзвичайно корисним для дослідження стабільних молекул.

- Підходить для аналізу навіть складних сумішей.

Недоліки:

- Може бути тривалим процесом для досягнення рівноваги.

- Потрібно бути впевненим у стабільності молекул у тестованих умовах.

Висновки до розділу 1

З огляду на поточні досягнення в галузі обчислювальної хімії та машинного навчання, QSPR методи мають величезний потенціал для подальшого розвитку. Вони не тільки допомагають спрогнозувати основні фізико-хімічні властивості молекул, але й можуть бути використані для створення нових підходів у розробці лікарських засобів, персоналізованому лікуванні та вдосконаленні технологій, що використовуються в фармацевтичній галузі.

2. ЕКСПЕРИМЕНТАЛЬНА ЧАСТИНА

2.1. Масив даних для дослідження

Досліджуваний масив даних включав 54 органічні речовини, що є активними фармацевтичними інгредієнтами та допоміжними речовинами. Список досліджуваних речовин з коротким описом їх застосування:

1) Активні фармацевтичні інгредієнти:

- Флуфенамова кислота – нестероїдний протизапальний засіб, який зменшує біль і запалення.

- Іопанова кислота – використовується в терапії тиреоїдних захворювань, зокрема для лікування гіпертиреозу.

- Дифлунізал – використовується як протизапальний, знеболювальний і жарознижувальний засіб.

- Амітриптилін – застосовується як антидепресант.

- Галоперидол – застосовується як антипсихотичний препарат. Використовується для лікування психозів, таких як шизофренія, а також для контролю тривожності, агресії та симптомів манії в рамках біполярних розладів.

- Мефенамова кислота – має знеболювальні та протизапальні властивості. Використовується для лікування болю, зокрема при головному болю, менструальних болях, болях у суглобах, а також при запальних захворюваннях, таких як артрит.

- Флурбіпрофен – нестероїдний протизапальний засіб. Має знеболювальні, жарознижувальні та протизапальні властивості.

- Лоразепам – належить до класу бензодіазепінів. Має заспокійливі, снодійні, протитривожні та міорелаксанти властивості.

- Ібупрофен – має знеболювальні, жарознижувальні та протизапальні властивості. Використовується для полегшення болю різного походження (головний, зубний, менструальний, болі в суглобах), зменшення запалення

при захворюваннях опорно-рухового апарату (наприклад, артрит), а також для зниження температури при простудних захворюваннях.

– Налідиксова кислота – антибіотик групи хінолонів. Використовується для лікування інфекцій сечовивідних шляхів (цистит, пієлонефрит) та інших бактеріальних інфекцій.

– Карбамазепін – використовується як протисудомний засіб. Застосовується для лікування епілепсії, зокрема при парціальних та генералізованих судомах, а також для контролю біполярних розладів.

– 5-Етил-5-геptулбарбітурат – використовується як снодійний засіб. Застосовується для лікування безсоння та порушень сну.

– Бутамбен – застосовується як місцевий анестетик. Він використовується для знеболення в медичних процедурах, особливо для обробки шкіри та слизових оболонок.

– Алклофенак – належить до нестероїдних протизапальних препаратів. Використовується для зменшення болю, запалення та набряків при різних захворюваннях, таких як остеоартрит, ревматоїдний артрит і подагра.

– Гексетал – використовується як анестетик. Застосовується в медичній практиці для короткочасного знеболення при хірургічних або діагностичних процедурах.

– Гептабарбітал – належить до групи барбітуратів. Використовується як снодійний засіб для короткочасного лікування безсоння.

– Метаквалон – належить до групи снодійних препаратів. Використовувався для лікування безсоння, а також як заспокійливий засіб.

– Празиквантел – застосовується для лікування паразитарних захворювань, зокрема гельмінтозів.

– Хлорзоказон – використовується як міорелаксант. Застосовується для зняття м'язових спазмів, які можуть виникати при різних захворюваннях.

– 5,5-Ді-і-пропілбарбітурова кислота – барбітурат. Використовується як снодійний препарат.

– Примідон – протисудомний засіб. Застосовується для лікування епілепсії та судомних розладів.

– Етил-4-амінобензоат (Бензокаїн) – анестезуючий засіб. Використовується для місцевого знеболення при лікуванні болю в горлі, зубного болю, а також для знеболення шкіри при незначних травмах або опіках.

– Ломефлосацин – застосовується як антибіотик для лікування бактеріальних інфекцій, зокрема інфекцій сечовивідних шляхів, а також респіраторних та шлунково-кишкових інфекцій.

– Амобарбітал – використовується як снодійний засіб.

– Ізокарбоксазид – антидепресант з групи інгібіторів моноаміноксидази. Застосовується для лікування депресії, зокрема депресій, що не піддаються лікуванню іншими засобами.

– Гептобарбітал – барбітурат. Застосовується як снодійний засіб для лікування безсоння.

– Фенобарбітал – барбітурат. Використовується як протисудомний засіб для лікування епілепсії та інших судомних розладів.

– Глутетимід – снодійний засіб. Використовується для лікування безсоння та інших порушень сну.

– Пропілтіоурацил – застосовується для лікування гіпертиреозу (надмірної активності щитовидної залози). Пропілтіоурацил знижує рівень гормонів щитовидної залози, блокуючи їх синтез, і таким чином нормалізує функцію щитовидної залози.

– Ідобутал – антиаритмічний препарат. Застосовується для лікування порушень серцевого ритму, зокрема при шлуночкових та суправентрикулярних аритміях.

– Пробарбітал – належить до групи барбітуратів. Використовується як снодійний засіб для лікування безсоння та інших порушень сну.

– Атропін – належить до групи антихолінергічних засобів. Використовується в медицині для лікування різних станів, таких як серцеві

розлади, брадикардія (повільний пульс), а також як міорелаксант при підготовці до хірургічних операцій.

- Буталбітал – належить до групи барбітуратів. Має заспокійливу та анксіолітичну дію і використовується переважно для лікування психоемоційних розладів, таких як тривожність та безсоння.

- Міноксидил – застосовується для лікування облісіння та стимулювання росту волосся.

- Саліціламід – знеболювальний та протизапальний засіб. Зазвичай використовують для лікування більших болів та запальних процесів, таких як артрит та міозит.

- 7-Бутилтеофілін – застосовується при дихальних розладах. Використовується для лікування таких захворювань, як астма та хронічна обструктивна хвороба легень.

- Саліцилова кислота – широко застосовується в медицині та фармацевтиці. Основне застосування полягає у лікуванні шкірних захворювань, таких як псоріаз, акне, грибкові інфекції тощо.

- Аллобарбітал – барбітурат із седативною та гіпнотичною дією. Використовується як ліки для лікування безсоння, а також в деяких випадках для контролю тривожних розладів.

- Аспірин (або ацетилсаліцилова кислота) – анальгезуючий, протизапальний та жарознижувальний препарат. Широко застосовується для лікування болю, запалення та жарів, а також для профілактики тромбозів.

- Апробарбітал – барбітурова кислота з групи барбітуратів, що має заспокійливу та снодійну дію.

- Баклофен – належить до групи міорелаксантів. Використовується для лікування м'язових спазмів та м'язової слабкості, зокрема при таких захворюваннях, як розсіяний склероз, спінальні травми, церебральний параліч, а також для зменшення симптомів дифузного спастичного паралічу.

– Бутобарбітон (Бутетал) – належить до класу барбітуратів. Використовується як седативний, анхіолітичний та гіпнотичний засіб, що допомагає при лікуванні безсоння, тривожних розладів та депресії.

– Циметидин – належить до класу антигістамінних засобів. Застосовується для лікування виразкової хвороби шлунка, гастроезофагеального рефлюксного захворювання і для зменшення кислотності шлунка.

– Секбутабарбітал – похідне барбітурової кислоти, що має седативну, гіпнотичну та протисудомну дію.

– Диклофенак – нестероїдний протизапальний препарат, що широко застосовується в медицині.

– Напроксен – нестероїдний протизапальний препарат, що використовується для зменшення болю, запалення та жару.

– Буметанід – сечогінний препарат.

– Сальбутамол – бета-2-адреноміметичний препарат, що використовується як бронхолітичний засіб.

– Сульфамеразін – сульфаніламідний антибіотик, що використовується для лікування бактеріальних інфекцій.

– 5-метил-5-аллілбарбітурат – барбітурова кислота, що має седативний та анксіолітичний ефект.

2) Допоміжні інгредієнти

– Бутилпарабен – консервант, що широко використовується в фармацевтичній і косметичній промисловості. Він запобігає росту мікроорганізмів, що дозволяє продовжити термін придатності продуктів. Зазвичай бутилпарабен додається до мазей, кремів, шампунів та інших косметичних засобів для їх збереження. У фармацевтичних препаратах може використовуватися для стабілізації складових, запобігаючи їх псуванню.

– Етилпарабен – консервант, що використовується у фармацевтиці та косметиці для протимікробної дії. Він допомагає продовжити термін зберігання продуктів, запобігаючи розвитку бактерій та грибків.

– Сахарин – штучний підсолоджувач, що широко використовується як заміник цукру. Сахарин часто додають до лікарських засобів, зокрема до сиропів, щоб надати їм приємний смак і зробити їх приємнішими для вживання, зокрема для пацієнтів, які мають труднощі зі смаковими відчуттями при прийомі ліків.

– Метил-п-гідроксибензоат – органічна сполука, що належить до ефірів пара-гідроксибензоєвої кислоти. Широко використовується як консервант у косметичних та фармацевтичних препаратах.

Для опису досліджених органічних речовин використали набір із 10 фізико-хімічних та топологічних дескрипторів:

- молекулярна маса (molecular weight),
- логарифм коефіцієнта розподілу (logP) – логарифм коефіцієнта розподілу молекули між двома розчинниками,
- кількість донорів водневих зв'язків (hydrogen bond donor count),
- кількість акцепторів водневих зв'язків (hydrogen bond acceptor count),
- кількість обертових зв'язків (rotatable bond count) – кількість зв'язків, навколо яких молекула може обертатися,
- топологічна площа полярної поверхні (topological polar surface area),
- кількість важких атомів (heavy atom count) – кількість атомів, що важчі за Гідроген,
- молекулярна складність (complexity) – міра складності молекули на основі її структурних характеристик, таких як кількість різних типів зв'язків, атомів, циклів та інших факторів,
- кількість визначених стереоцентрів атомів (defined atom stereocenter count) – кількість атомів, які є стереоцентрами та мають чітко визначену стереохімію,
- кількість невизначених стереоцентрів атомів (undefined atom stereocenter count) – кількість атомів, для яких стереохімія не визначена або відсутня.

Значення цих 10 параметрів були отримані з бази даних PubChem [14]. Експериментальні значення водної розчинності взяті із наукової літератури [4, 15]. Експериментальні значення водної розчинності були виміряні за температури 25°C, моль/л. Ми використали логарифм розчинності ($\log S_w$).

Досліджені органічні речовини випадковим чином були поділені на навчальну (85%) та тестову вибірки (15%). Навчальну вибірку використовують для побудови та налагодження моделі, тестову вибірку – для оцінки якості створеної моделі на абсолютно нових даних, для зовнішньої валідації.

Значення 10 дескрипторів, експериментальні значення водної розчинності та приналежність речовин до навчальної / тестової вибірок представлено у таблиці 2.1.

Таблиця 2.1. Значення 10 дескрипторів, експериментальні значення водної розчинності та приналежність досліджуваних органічних речовин до навчальної / тестової вибірок

№	Органічна речовина	Вибірка	log S _w	Номер дескриптора*									
				1	2	3	4	5	6	7	8	9	10
1	Флуфенамова кислота	навчальна	-4,623	281,23	5,2	2	6	3	49,3	20	346	0	0
2	Іопанова кислота	навчальна	-4,580	570,93	3,9	2	3	4	63,3	17	277	0	1
3	Дифлунісал	навчальна	-4,479	250,20	4,4	2	5	2	57,5	18	311	0	0
4	Амітриптилін	навчальна	-4,456	277,40	5,0	0	1	3	3,2	21	331	0	0
5	Галоперидол	навчальна	-4,429	375,90	3,2	1	4	6	40,5	26	451	0	0
6	Мефенамова кислота	навчальна	-3,770	241,28	5,1	2	3	3	19,3	18	292	0	0
7	Флурабіпрофен	навчальна	-3,740	244,26	4,2	1	3	3	37,3	18	286	0	1
8	Лоразепам	навчальна	-3,604	321,20	2,4	2	3	1	61,7	21	443	0	1
9	Ібупрофен	навчальна	-3,420	206,28	3,5	1	2	4	37,3	15	203	0	1
10	Налідиксова кислота	навчальна	-3,366	232,23	1,4	1	5	2	70,5	17	378	0	0
11	Карбамазепін	навчальна	-3,294	236,27	2,5	1	1	0	46,3	18	326	0	0
12	5-Етил-5-гептилбарбітурова кислота	навчальна	-3,218	254,33	3,0	2	3	7	75,3	18	319	0	0
13	Бутамбен	навчальна	-3,131	193,24	2,9	1	3	5	52,3	14	174	0	0
14	Алклофенак	навчальна	-3,125	226,65	2,5	1	3	5	46,5	15	230	0	0
15	Бутилпарабен	навчальна	-3,101	194,23	3,6	1	3	5	46,5	14	171	0	0
16	Гексетал	навчальна	-3,049	240,30	2,5	2	3	6	75,3	17	306	0	0
17	Гептабарбітал	навчальна	-3,000	250,29	2,0	2	3	2	75,3	18	406	0	0
18	Метаквалон	навчальна	-2,921	250,29	2,5	0	2	1	32,7	19	390	0	0
19	Празиквантел	навчальна	-2,893	312,40	2,7	0	2	1	40,6	23	472	0	1
20	Хлорзоксазон	навчальна	-2,831	169,56	1,8	1	2	0	38,3	11	185	0	0
21	5,5-Ді-ізопропілбарбітурова кислота	навчальна	-2,766	212,25	1,6	2	3	2	75,3	15	291	0	0

22	Примідон	навчальна	-2,640	218,25	0,9	2	2	2	58,2	16	279	0	0
23	Етил-4-амінобензоат (Бензокаїн)	навчальна	-2,616	165,19	1,9	1	3	3	52,3	12	151	0	0
24	Ломефлораксацин	навчальна	-2,533	351,35	-0,8	2	8	3	72,9	25	585	0	1
25	Амобарбітал	навчальна	-2,470	226,27	2,1	2	3	4	75,3	16	303	0	0
26	Ізокарбоксамід	навчальна	-2,461	231,25	1,5	2	4	4	67,2	17	254	0	0
27	Гептобарбітал	навчальна	-2,380	218,21	1,1	2	3	1	75,3	16	325	0	0
28	Фенобарбітал	навчальна	-2,366	232,23	1,5	2	3	2	75,3	17	339	0	0
29	Етилпарабен	навчальна	-2,346	166,17	2,5	1	3	3	46,5	12	148	0	0
30	Глутетимід	навчальна	-2,337	217,26	1,9	1	2	2	46,2	16	294	0	1
31	Пропілтіоурацил	навчальна	-2,185	170,23	0,8	2	2	2	73,2	11	222	0	0
32	Пробарбітал	навчальна	-2,153	198,22	1,0	2	3	2	75,3	14	276	0	0
33	Атропін	навчальна	-2,124	289,40	1,8	1	4	5	49,8	21	353	2	1
34	Буталбітал	навчальна	-2,119	224,26	1,7	2	3	4	75,3	16	327	0	0
35	Саліциламід	навчальна	-1,836	137,14	1,3	2	2	1	63,3	10	136	0	0
36	7-Бутилтеофілін	навчальна	-1,805	236,27	0,7	0	3	3	58,4	17	333	0	0
37	Саліцилова кислота	навчальна	-1,804	138,12	2,3	2	3	1	57,5	10	133	0	0
38	Аллобарбітал	навчальна	-1,796	208,21	1,0	2	3	4	75,3	15	315	0	0
39	Аспірин	навчальна	-1,733	180,16	1,2	1	4	3	63,6	13	212	0	0
40	Сульфамеразин	навчальна	-1,218	264,31	0,1	2	6	3	106	18	360	0	0
41	5-Метил-5-алкілбарбітурова кислота	навчальна	-1,160	182,18	0,5	2	3	2	75,3	13	276	0	0
42	Бутобарбітагон (Бутетал)	навчальна	-1,686	212,25	1,7	2	3	4	75,3	15	279	0	0
43	Циметидин	навчальна	-1,613	252,34	0,4	3	4	7	114,0	17	296	0	0
44	Метил-п-гідроксибензоат	навчальна	-1,705	152,15	2,0	1	3	2	46,5	11	136	0	0
45	Сальбутамол	навчальна	-1,224	239,31	0,3	4	4	5	72,7	17	227	0	1
46	Баклофен	навчальна	-1,696	213,66	-1,0	2	3	4	63,3	14	191	0	1
47	Секбутабарбітал	тестова	-2,333	212,25	1,6	2	3	3	75,3	15	292	0	1

48	Диклофенак	тестова	-5,097	296,10	4,4	2	3	4	49,3	19	304	0	0
49	Напроксен	тестова	-4,155	230,26	3,3	1	3	3	46,5	17	277	1	0
50	Буметанід	тестова	-3,562	364,40	2,8	3	7	8	127,0	25	528	0	0
51	Ідобутал	тестова	-2,172	224,26	2,0	2	3	5	75,3	16	316	0	0
52	Сахарин	тестова	-1,725	183,19	0,9	1	3	0	71,6	12	303	0	0
53	Апробарбітал	тестова	-1,710	210,23	1,1	2	3	3	75,3	15	314	0	0
54	Міноксидил	тестова	-1,978	209,25	1,2	3	3	1	88,9	15	329	0	0

*1 – молекулярна маса, 2 – логарифм коефіцієнта розподілу, 3 – кількість донорів водневих зв'язків, 4 – кількість акцепторів водневих зв'язків, 5 – кількість обертових зв'язків, 6 – топологічна площа полярної поверхні, 7 – кількість важких атомів, 8 – молекулярна складність, 9 – кількість визначених стереоцентрів атомів, 10 – кількість невизначених стереоцентрів атомів.

2.2. Реалізація методології QSPR

Для характеристики якості QSPR моделей використали такі статистичні показники [16-18]:

- коефіцієнт детермінації R^2 – показує, яка частка варіації залежної змінної пояснюється незалежними змінними в моделі; чим вище коефіцієнт детермінації, тим точніше модель;

- скоригований (adjusted) коефіцієнт детермінації R^2_{adj} – коригована версія коефіцієнта детермінації, враховує кількість предикторів (незалежних змінних) у моделі та є більш об'єктивним при порівнянні моделей з різною кількістю змінних:

$$R^2_{adj} = 1 - \frac{(1-R^2)(n-1)}{n-k-1}, \quad (2.1)$$

де n – загальна кількість органічних речовин, k – кількість незалежних змінних у моделі (дескрипторів);

- залишки (residuals) – різниця між експериментальним та прогнозованим значенням водної розчинності, та відносна похибка залишків у відсотках (% RE):

$$\%RE = \frac{\text{Residual}}{\text{експериментальне значення}} 100\%. \quad (2.2)$$

Усі розрахунки здійснені із застосуванням програмного забезпечення Matlab R2024b (trial individual license) [19].

2.2.1. Перевірка дескрипторів на мультиколінеарність

Першим кроком розробки QSPR моделі є перевірка дескрипторів на мультиколінеарність. Перевірка дескрипторів на мультиколінеарність потрібна для того, щоб:

- забезпечити інформативність моделі – якщо два дескриптори несуть однакову інформацію, один із них можна виключити без втрати прогностичної здатності моделі,

– покращити інтерпретацію моделі – мультиколінеарність ускладнює визначення впливу кожного окремого дескриптора на залежну змінну (водну розчинність), оскільки їхні ефекти можуть «змішуватися».

Для цього розраховали коефіцієнти кореляції між кожною парою параметрів. Розраховані значення коефіцієнтів кореляції представлено у таблиці 2.2. Результати свідчать, що дескриптор № 7 «кількість важких атомів» та дескриптор № 8 «молекулярна складність» характеризуються високим значенням R 0,8870. Це вказує на наявність мультиколінеарності між ними. Таке високе значення коефіцієнта детермінації свідчить про те, що ці два дескриптори несуть подібну інформацію.

Для видалення одного з цих дескрипторів визначили, який із них є більш значущим для прогностичної моделі шляхом розрахунку коефіцієнта детермінації між значеннями дескрипторів та експериментальними значеннями водної розчинності (рисунки 2.1. та 2.2).

Таблиця 2.2. Розраховані значення коефіцієнтів кореляції між кожною парою дескрипторів

Параметр [*]	1	2	3	4	5	6	7	8	9	10
1	1,0000	0,2438	0,0214	0,2712	0,2305	-0,0212	0,7124	0,6080	0,1128	0,4526
2	0,2438	1,0000	-0,3355	-0,2108	0,0519	-0,6587	0,2156	-0,0300	-0,0265	-0,0600
3	0,0214	-0,3355	1,0000	0,3381	0,2394	0,6700	-0,1197	-0,0598	-0,1094	0,0238
4	0,2712	-0,2108	0,3381	1,0000	0,2529	0,4124	0,3639	0,3890	0,0964	0,0865
5	0,2305	0,0519	0,2394	0,2529	1,0000	0,2274	0,2034	-0,0298	0,1710	0,0421
6	-0,0212	-0,6587	0,6700	0,4124	0,2274	1,0000	-0,1323	0,1064	-0,0768	-0,1457
7	0,7124	0,2156	-0,1197	0,3639	0,2034	-0,1323	1,0000	0,8870	0,1956	0,3496
8	0,6080	-0,0300	-0,0598	0,3890	-0,0298	0,1064	0,8870	1,0000	0,0992	0,2400
9	0,1128	-0,0265	-0,1094	0,0964	0,1710	-0,0768	0,1956	0,0992	1,0000	0,2828
10	0,4526	-0,0600	0,0238	0,0865	0,0421	-0,1457	0,3496	0,2400	0,2828	1,0000

^{*} 1 – молекулярна маса, 2 – логарифм коефіцієнта розподілу, 3 – кількість донорів водневих зв’язків, 4 – кількість акцепторів водневих зв’язків, 5 – кількість обертових зв’язків, 6 – топологічна площа полярної поверхні, 7 – кількість важких атомів, 8 – молекулярна складність, 9 – кількість визначених стереоцентрів атомів, 10 – кількість невизначених стереоцентрів атомів.

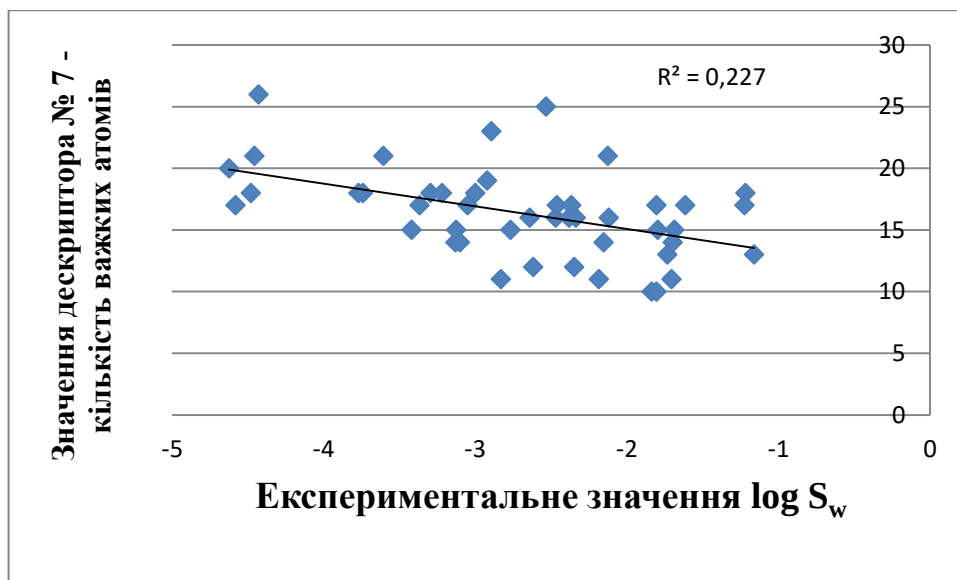


Рисунок 2.1. Залежність дескриптора «кількість важких атомів» від логарифма експериментальних значень водної розчинності органічних речовин

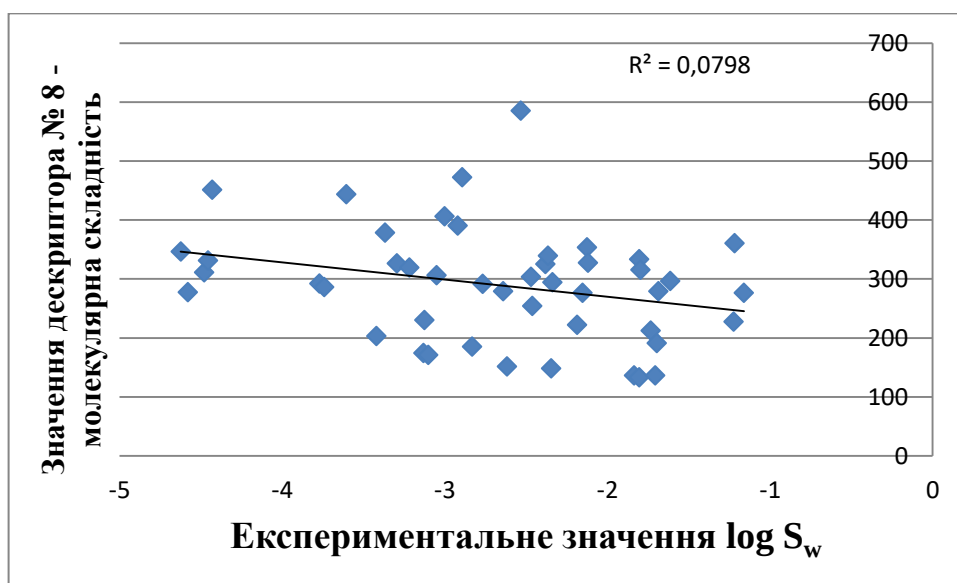


Рисунок 2.2. Залежність дескриптора «молекулярна складність» від логарифма експериментальних значень водної розчинності органічних речовин

Як видно з рисунків 2.1 та 2.2, дескриптор «кількість важких атомів» має суттєво більше значення коефіцієнта детермінації із експериментальними значеннями водної розчинності органічних речовин, що вказує на його вищу значущість для прогнозування. Тому з набору дескрипторів доцільно

видалити саме параметр «молекулярна складність». Отже, для побудови QSPR моделі застосували 9 дескрипторів.

Для визначення базисного набору дескрипторів для початку побудови QSPR моделі також розраховували коефіцієнти детермінації між дескрипторами № 1, 2, 3, 4, 5, 6, 9, 10 та експериментальними значеннями водної розчинності органічних речовин. Результати цих розрахунків наведено у таблиці 2.3.

Таблиця 2.3. Розраховані значення коефіцієнтів детермінації між дескрипторами та експериментальними значеннями водної розчинності органічних речовин

Параметр	R^2
молекулярна маса	0,2782
логарифм коефіцієнта розподілу	0,6729
кількість донорів водневих зв'язків	0,0862
кількість акцепторів водневих зв'язків	0,0006
кількість обертових зв'язків	0,0007
топологічна площа полярної поверхні	0,3039
кількість важких атомів	0,2270
кількість визначених стереоцентрів атомів	0,0085
кількість невизначених стереоцентрів атомів	0,0050

Представлені у таблиці 2.3 результати свідчать про надзвичайно низьку значущість для прогнозування водної розчинності органічних речовин таких параметрів як: кількість невизначених стереоцентрів атомів, кількість обертових зв'язків, кількість визначених стереоцентрів атомів, кількість акцепторів водневих зв'язків та кількість донорів водневих зв'язків.

Найбільшою силою кореляції із експериментальними значеннями водної розчинності характеризуються дескриптори «логарифм коефіцієнта розподілу» та «топологічна площа полярної поверхні». Тому саме ці 2

параметра є основою для побудови QSPR моделі. До цієї двопараметричної моделі поступово додавалися дескриптори «молекулярна маса» та «кількість важких атомів» для оцінки їх значущості для прогнозування водної розчинності органічних речовин.

2.2.2. Побудова двопараметричної QSPR моделі

Рівняння двопараметричної регресії, що описує залежність логарифму експериментальних значень водної розчинності органічних речовин від логарифма коефіцієнта розподілу та топологічної площі полярної поверхні має вигляд:

$$\log S_w = -1,66161 - 0,52849 \cdot x_1 + 0,00895 \cdot x_2, \quad (2.3)$$

де x_1 – логарифм коефіцієнта розподілу, x_2 – топологічна площа полярної поверхні.

Статистичні показники для моделі, що описується рівнянням 2.3, мають такі значення: $R = 0,8204$, $R^2 = 0,6731$, $R_{adj}^2 = 0,6579$.

Прогнозовані значення водної розчинності для органічних речовин навчальної та тестової вибірок, а також значення залишків та відносної похибки залишків у відсотках наведено у таблиці 2.4.

Рисунок 2.3 свідчить про високу можливість застосування побудованої двопараметричної QSPR моделі (2.3) до абсолютно нових даних ($R^2 = 0,9695$).

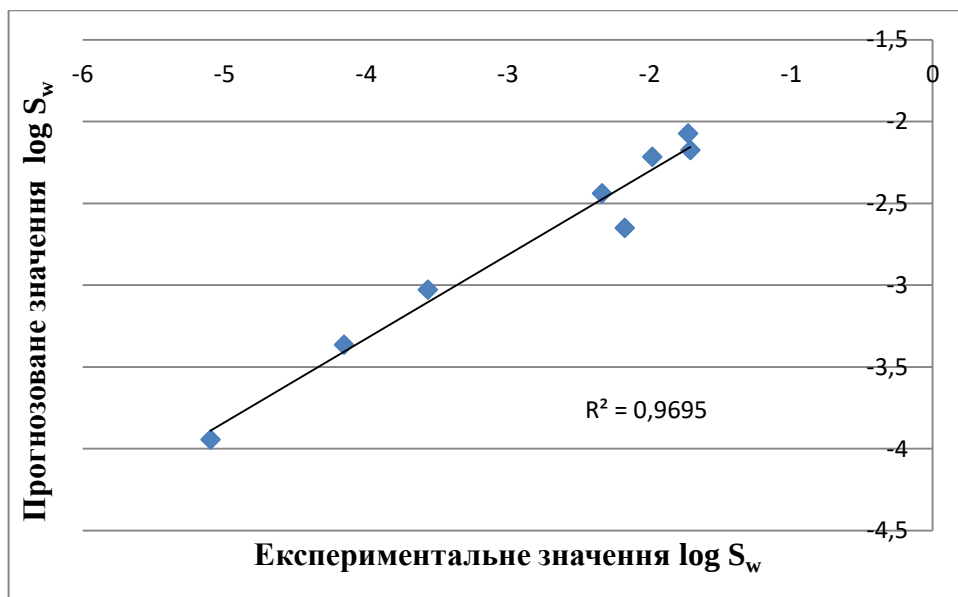


Рисунок 2.3. Залежність прогнозованих значень водної розчинності від експериментальних значень водної розчинності органічних речовин тестової вибірки у випадку двопараметричної QSPR моделі

Таблиця 2.4. Значення залишків та відносної похибки залишків у відсотках для навчальної та тестової вибірок для випадку двопараметричної QSPR моделі

№	Органічна речовина	Вибірка	$\log S_w$	Прогнозоване значення $\log S_w$	Residual	%Re
1	Флуфенамова кислота	навчальна	-4,623	-4,366	-0,257	5,567
2	Іопанова кислота	навчальна	-4,580	-3,666	-0,914	19,955
3	Дифлунісал	навчальна	-4,479	-3,936	-0,543	12,134
4	Амітриптилін	навчальна	-4,456	-4,301	-0,155	3,474
5	Галоперидол	навчальна	-4,429	-3,317	-1,112	25,118
6	Мефенамова кислота	навчальна	-3,770	-4,340	0,570	-15,110
7	Флурабіпрофен	навчальна	-3,740	-3,848	0,108	-2,885
8	Лоразепам	навчальна	-3,604	-2,875	-0,729	20,234
9	Ібупрофен	навчальна	-3,420	-3,478	0,058	-1,694
10	Налідиксова кислота	навчальна	-3,366	-2,338	-1,028	30,529
11	Карбамазепін	навчальна	-3,294	-2,941	-0,353	10,704
12	5-Етил-5-гептилбарбітурова кислота	навчальна	-3,218	-3,180	-0,038	1,191
13	Бутамбен	навчальна	-3,131	-3,147	0,016	-0,525
14	Алклофенак	навчальна	-3,125	-2,941	-0,184	5,881
15	Бутилпарабен	навчальна	-3,101	-3,523	0,422	-13,594
16	Гексетал	навчальна	-3,049	-2,915	-0,134	4,380
17	Гептабарбітал	навчальна	-3,000	-2,651	-0,349	11,627
18	Метаквалон	навчальна	-2,921	-2,954	0,033	-1,115
19	Празиквантел	навчальна	-2,893	-3,052	0,159	-5,503
20	Хлорзоксазон	навчальна	-2,831	-2,579	-0,252	8,915
21	5,5-Ді-ізопропілбарбітурова кислота	навчальна	-2,766	-2,440	-0,326	11,793

22	Примідон	навчальна	-2,640	-2,085	-0,555	21,017
23	Етил-4-амінобензоат (Бензокаїн)	навчальна	-2,616	-2,619	0,003	-0,112
24	Ломефлораксацин	навчальна	-2,533	-1,174	-1,359	53,669
25	Амобарбітал	навчальна	-2,470	-2,704	0,234	-9,476
26	Ізокарбоксамід	навчальна	-2,461	-2,394	-0,067	2,714
27	Гептобарбітал	навчальна	-2,380	-2,176	-0,204	8,590
28	Фенобарбітал	навчальна	-2,366	-2,387	0,021	-0,885
29	Етилпарабен	навчальна	-2,346	-2,941	0,595	-25,372
30	Глутетімід	навчальна	-2,337	-2,624	0,287	-12,297
31	Пропілтіоурацил	навчальна	-2,185	-2,019	-0,166	7,603
32	Пробарбітал	навчальна	-2,153	-2,123	-0,030	1,407
33	Атропін	навчальна	-2,124	-2,568	0,444	-20,919
34	Буталбітал	навчальна	-2,119	-2,493	0,374	-17,633
35	Саліциламід	навчальна	-1,836	-2,292	0,456	-24,836
36	7-Бутилтеофілін	навчальна	-1,805	-1,979	0,174	-9,655
37	Саліцилова кислота	навчальна	-1,804	-2,826	1,022	-56,634
38	Аллобарбітал	навчальна	-1,796	-2,123	0,327	-18,191
39	Аспірин	навчальна	-1,733	-2,239	0,506	-29,191
40	Сульфамеразин	навчальна	-1,218	-1,620	0,402	-32,971
41	5-Метил-5-алкілбарбітурова кислота	навчальна	-1,160	-1,858	0,698	-60,212
42	Бутобарбітагон (Бутетал)	навчальна	-1,686	-2,493	0,807	-47,844
43	Циметидин	навчальна	-1,613	-1,771	0,158	-9,794
44	Метил-п-гідроксибензоат	навчальна	-1,705	-2,677	0,972	-57,007
45	Сальбутамол	навчальна	-1,224	-1,755	0,531	-43,389
46	Баклофен	навчальна	-1,696	-1,076	-0,620	36,530
47	Секбутабарбітал	тестова	-2,333	-2,440	0,107	-4,578

48	Диклофенак	тестова	-5,097	-3,943	-1,154	22,644
49	Напроксен	тестова	-4,155	-3,364	-0,791	19,037
50	Буметанід	тестова	-3,562	-3,028	-0,534	15,000
51	Ідобутал	тестова	-2,172	-2,651	0,479	-22,062
52	Сахарин	тестова	-1,725	-2,073	0,348	-20,184
53	Апробарбітал	тестова	-1,710	-2,176	0,466	-27,225
54	Міноксидил	тестова	-1,978	-2,216	0,238	-12,044

2.2.3. Побудова трипараметричної QSPR моделі

Рівняння трипараметричної регресії, що описує залежність логарифму експериментальних значень водної розчинності органічних речовин від логарифма коефіцієнта розподілу, топологічної площі полярної поверхні та молекулярної маси має вигляд:

$$\log S_w = -0,98629 - 0,43173 \cdot x_1 + 0,005072 \cdot x_2 - 0,00476 \cdot x_3, \quad (2.4)$$

де x_1 – логарифм коефіцієнта розподілу, x_2 – топологічна площа полярної поверхні, x_3 – молекулярна маса.

Статистичні показники для моделі, що описується рівнянням 2.4, мають такі значення: $R = 0,8907$, $R^2 = 0,7934$, $R_{adj}^2 = 0,7786$. Спостерігається значне збільшення значень статистичних показників у порівнянні із двопараметричною моделлю, що свідчить про важливість та необхідність застосування молекулярної маси як предиктора для прогнозування водної розчинності органічних речовин.

Прогнозовані значення водної розчинності для органічних речовин навчальної та тестової вибірок, а також значення залишків та відносної похибки залишків у відсотках наведено у таблиці 2.5.

Рисунок 2.4 свідчить про високу можливість застосування побудованої трипараметричної QSPR моделі (2.4) до абсолютно нових даних ($R^2 = 0,9579$).

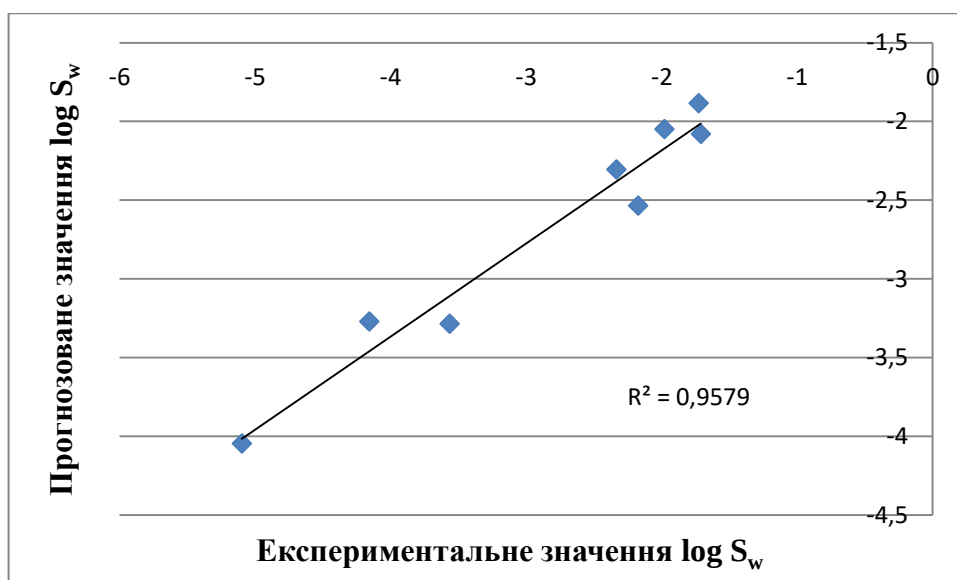


Рисунок 2.4. Залежність прогнозованих значень водної розчинності від експериментальних значень водної розчинності органічних речовин тестової вибірки у випадку трипараметричної QSPR моделі

Таблиця 2.5. Значення залишків та відносної похибки залишків у відсотках для навчальної та тестової вибірок для випадку трипараметричної QSPR моделі

№	Органічна речовина	Вибірка	$\log S_w$	Прогнозоване значення $\log S_w$	Residual	%Re
1	Флуфенамова кислота	навчальна	-4,623	-4,366	-0,303	6,561
2	Іопанова кислота	навчальна	-4,580	-3,666	0,486	-10,617
3	Дифлунісал	навчальна	-4,479	-3,936	-0,694	15,494
4	Амітриптилін	навчальна	-4,456	-4,301	-0,007	0,158
5	Галоперидол	навчальна	-4,429	-3,317	-0,478	10,782
6	Мефенамова кислота	навчальна	-3,770	-4,340	0,469	-12,429
7	Флурабіпрофен	навчальна	-3,740	-3,848	0,033	-0,879
8	Лоразепам	навчальна	-3,604	-2,875	-0,366	10,150
9	Ібупрофен	навчальна	-3,420	-3,478	-0,130	3,804
10	Налідиксова кислота	навчальна	-3,366	-2,338	-1,028	30,529
11	Карбамазепін	навчальна	-3,294	-2,941	-0,339	10,283
12	5-Етил-5-гептилбарбітурова кислота	навчальна	-3,218	-3,180	-0,108	3,356
13	Бутамбен	навчальна	-3,131	-3,147	-0,238	7,610
14	Алклофенак	навчальна	-3,125	-2,941	-0,217	6,929
15	Бутилпарабен	навчальна	-3,101	-3,523	0,128	-4,130
16	Гексетал	навчальна	-3,049	-2,915	-0,222	7,269
17	Гептабарбітал	навчальна	-3,000	-2,651	-0,341	11,365
18	Метаквалон	навчальна	-2,921	-2,954	0,170	-5,819
19	Празиквантел	навчальна	-2,893	-3,052	0,540	-18,661
20	Хлорзоксазон	навчальна	-2,831	-2,579	-0,455	16,067
21	5,5-Ді-ізопропілбарбітурова кислота	навчальна	-2,766	-2,440	-0,461	16,656

22	Примідон	навчальна	-2,640	-2,085	-0,522	19,758
23	Етил-4-амінобензоат (Бензокаїн)	навчальна	-2,616	-2,619	-0,288	11,028
24	Ломефлораксацин	навчальна	-2,533	-1,174	-0,590	23,278
25	Амобарбітал	навчальна	-2,470	-2,704	0,118	-4,773
26	Ізокарбоксамид	навчальна	-2,461	-2,394	-0,067	2,737
27	Гептобарбітал	навчальна	-2,380	-2,176	-0,262	11,017
28	Фенобарбітал	навчальна	-2,366	-2,387	-0,009	0,371
29	Етилпарабен	навчальна	-2,346	-2,941	0,275	-11,706
30	Глутетімід	навчальна	-2,337	-2,624	0,269	-11,522
31	Пропілтіоурацил	навчальна	-2,185	-2,019	-0,414	18,967
32	Пробарбітал	навчальна	-2,153	-2,123	-0,174	8,059
33	Атропін	навчальна	-2,124	-2,568	0,764	-35,978
34	Буталбітал	навчальна	-2,119	-2,493	0,287	-13,527
35	Саліциламід	навчальна	-1,836	-2,292	0,043	-2,351
36	7-Бутилтеофілін	навчальна	-1,805	-1,979	0,312	-17,273
37	Саліцилова кислота	навчальна	-1,804	-2,826	0,541	-29,988
38	Аллобарбітал	навчальна	-1,796	-2,123	0,231	-12,864
39	Аспірин	навчальна	-1,733	-2,239	0,306	-17,670
40	Сульфамеразин	навчальна	-1,218	-1,620	0,532	-43,658
41	5-Метил-5-алкілбарбітурова кислота	навчальна	-1,160	-1,858	0,527	-45,455
42	Бутобарбітагон (Бутетал)	навчальна	-1,686	-2,493	0,662	-39,292
43	Циметидин	навчальна	-1,613	-1,771	0,169	-10,460
44	Метил-п-гідроксибензоат	навчальна	-1,705	-2,677	0,633	-37,128
45	Сальбутамол	навчальна	-1,224	-1,755	0,662	-54,087
46	Баклофен	навчальна	-1,696	-1,076	-0,446	26,275
47	Секбутабарбітал	тестова	-2,333	-2,440	-0,028	1,181

48	Диклофенак	тестова	-5,097	-3,943	-1,052	20,634
49	Напроксен	тестова	-4,155	-3,364	-0,884	21,271
50	Буметанід	тестова	-3,562	-3,028	-0,276	7,762
51	Ідобутал	тестова	-2,172	-2,651	0,363	-16,727
52	Сахарин	тестова	-1,725	-2,073	0,159	-9,199
53	Апробарбітал	тестова	-1,710	-2,176	0,370	-21,635
54	Міноксидил	тестова	-1,978	-2,216	0,071	-3,615

2.2.4. Побудова чотирипараметричної QSPR моделі

Рівняння чотирипараметричної регресії, що описує залежність логарифму експериментальних значень водної розчинності органічних речовин від логарифма коефіцієнта розподілу, топологічної площі полярної поверхні, молекулярної маси та кількості важких атомів має вигляд:

$$\log S_w = -0,6649 - 0,43732 \cdot x_1 + 0,004121 \cdot x_2 - 0,00356 \cdot x_3 - 0,0328 \cdot x_4, \quad (2.5)$$

де x_1 – логарифм коефіцієнта розподілу, x_2 – топологічна площа полярної поверхні, x_3 – молекулярна маса, x_4 – кількість важких атомів.

Статистичні показники для моделі, що описується рівнянням 2.5, мають такі значення: $R = 0,8949$, $R^2 = 0,8010$, $R_{adj}^2 = 0,7816$. У порівнянні із трипараметричною моделлю можемо зробити такі спостереження та висновки:

- коефіцієнт детермінації змінився з 0,7934 до 0,8010, що є незначним підвищенням (0,0076). Це означає, що загальна частка варіації, яку модель пояснює, трохи збільшилась після додавання дескриптора, але зміна не є великою. Отже, дескриптор «кількість важких атомів» вносить деякий вклад у пояснення варіації, але цей вклад досить малий.

- скорегований коефіцієнт детермінації змінився з 0,7786 до 0,7816, що також є невеликою зміною (0,0030). Це вказує на те, що новий дескриптор не суттєво покращив модель з точки зору пояснення варіації, при цьому не знижуючи точність.

Отже, додавання до QSPR моделі дескриптора «кількість важких атомів» не змінює суттєво точність моделі, і можна сказати, що він не є критично важливим для прогнозування.

Додавання до трипараметричної моделі (2.4) таких параметрів як кількість невизначених стереоцентрів атомів, кількість обертових зв'язків, кількість визначених стереоцентрів атомів, кількість акцепторів водневих зв'язків та кількість донорів водневих зв'язків підтвердило їх незначущість для прогнозування водної розчинності органічних лікарських речовин.

Висновки до розділу 2

У результаті детального аналізу значущості 10 фізико-хімічних та топологічних дескрипторів для прогнозування водної розчинності органічних речовин встановлено оптимальний вигляд QSPR моделі, що характеризується високими значеннями статистичних параметрів, та включає лише три параметра:

$$\log S_w = -0,98629 - 0,43173 \cdot x_1 + 0,005072 \cdot x_2 - 0,00476 \cdot x_3,$$

де x_1 – логарифм коефіцієнта розподілу, x_2 – топологічна площа полярної поверхні, x_3 – молекулярна маса.

ВИСНОВКИ

1. Визначено, що для прогнозування водної розчинності органічних речовин, зокрема фармацевтично активних та допоміжних, найбільш значущими дескрипторами є логарифм коефіцієнта розподілу, площа полярної поверхні та молекулярна маса.

2. Побудована трипараметрична QSPR модель для прогнозування водної розчинності органічних речовин демонструє високу точність та надійність, підтверджуючи ефективність обраних дескрипторів.

3. Результати дослідження спрощують процес прогнозування водної розчинності органічних речовин та можуть бути корисними на етапі розробки лікарських засобів з оптимальними фармакокінетичними характеристиками.

СПИСОК ЛІТЕРАТУРИ

1. Ghaslani, D., Gorji, Z. E., Gorji, A. E., & Riahi, S. (2017). Descriptive and predictive models for Henry's law constant of CO₂ in ionic liquids: a QSPR study. *Chemical Engineering Research and Design*, 120, 15-25.
2. Das, R. N., & Roy, K. (2013). Advances in QSPR/QSTR models of ionic liquids for the design of greener solvents of the future. *Molecular diversity*, 17, 151-196.
3. Khajeh, A., Modarress, H., & Zeinoddini-Meymand, H. (2013). Modified particle swarm optimization method for variable selection in QSAR/QSPR studies. *Structural Chemistry*, 24, 1401-1409.
4. Duchowicz, P. R., Talevi, A., Bruno-Blanch, L. E., & Castro, E. A. (2008). New QSPR study for the prediction of aqueous solubility of drug-like compounds. *Bioorganic & medicinal chemistry*, 16(17), 7944-7955.
5. Duchowicz, P. R., Talevi, A., Bellera, C., Bruno-Blanch, L. E., & Castro, E. A. (2007). Application of descriptors based on Lipinski's rules in the QSPR study of aqueous solubilities. *Bioorganic & medicinal chemistry*, 15(11), 3711-3719.
6. Akbar, J., Iqbal, M. S., Chaudhary, M. T., Yasin, T., & Massey, S. (2012). A QSPR study of drug release from an arabinoxylan using ab initio optimization and neural networks. *Carbohydrate polymers*, 88(4), 1348-1357.
7. Begam, B. F., & Kumar, J. S. (2016). Computer assisted QSAR/QSPR approaches – a review. *Indian Journal of Science and Technology*, 9(8), 1-8.
8. Kiralj, R., & Ferreira, M. (2009). Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *Journal of the Brazilian Chemical Society*, 20, 770-787.
9. Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, 22(1), 69-77.

10. Etzweiler, F., Senn, E., & Schmidt, H. W. (1995). Method for measuring aqueous solubilities of organic compounds. *Analytical Chemistry*, 67(3), 655-658.
11. Birch, H., Redman, A. D., Letinski, D. J., Lyon, D. Y., & Mayer, P. (2019). Determining the water solubility of difficult-to-test substances: a tutorial review. *Analytica chimica acta*, 1086, 16-28.
12. Mota, F. L., Queimada, A. J., Pinho, S. P., & Macedo, E. A. (2008). Aqueous solubility of some natural phenolic compounds. *Industrial & Engineering Chemistry Research*, 47(15), 5182-5189.
13. Bodor, N., & Huang, M. J. (1992). A new method for the estimation of the aqueous solubility of organic compounds. *Journal of pharmaceutical sciences*, 81(9), 954-960.
14. <https://pubchem.ncbi.nlm.nih.gov/>
15. Rytting, E., Lentz, K. A., Chen, X. Q., Qian, F., & Venkatesh, S. (2005). Aqueous and cosolvent solubility data for drug-like organic compounds. *The AAPS Journal*, 7, E78-E105.
16. Patel, H. M., Noolvi, M. N., Sharma, P., Jaiswal, V., Bansal, S., Lohan, S., ... & Bhardwaj, V. (2014). Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Medicinal chemistry research*, 23, 4991-5007.
17. Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., & Prachayasittikul, V. (2009). A practical overview of quantitative structure-activity relationship. *EXCLI J*, 8(7), 74-88.
18. Roy, K., Kar, S., Das, R. N., Roy, K., Kar, S., & Das, R. N. (2015). Statistical methods in QSAR/QSPR. *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*, 37-59.
19. <https://www.mathworks.com/products.html>

ДОДАТОК



СЕРТИФІКАТ

ПРО УЧАСТЬ У КОНФЕРЕНЦІЇ (З ПУБЛІКАЦІЄЮ)

ICSR № 25/2103-050

✓ 0,4 ECTS

РЕКОМЕНДОВАНО

Вченою Радою

Науковий установи
Інститут науково-
технічної інженерної
співпраці

Протокол № 11 від 20.03.2025

✓ Конференцію
зарєєстровано

в Державній науковій
установі у сфері
управління Міністерства
освіти і науки «Інститут
Посвідчення № 92 від 04.01.2025.

✓ Офіційний
видавець

Свідчення суб'єкта
видавничої справи:
ДК № 7860 від 22.04.2025.

www.mcnd.org.ua



Ізмаїлов Раїм Муслімович

взяв(ла) участь у IX Міжнародній науковій конференції

«НАУКОВІ ТRENДИ
ПОСТІНДУСТРІАЛЬНОГО СУСПІЛЬСТВА»

21 березня 2025 року у м. Тернопіль, Україна

та опублікував(ла) наукову роботу в збірці конференції

з ISBN 978-617-8440-60-2

DOI 10.62731/mcnd-21.03.2025



ВІЦЕ-ПРЕЗИДЕНТ МЦНА
ГОЛОВА ОРГКОМІТЕТУ
СОТНИК СОЛОМІЯ

SUMMARY

Raim Izmailov

QSPR PREDICTION OF AQUEOUS SOLUBILITY OF ORGANIC COMPOUNDS: ACTIVE PHARMACEUTICAL INGREDIENTS AND EXCIPIENTS

Department of Analytical, Physical and Colloid Chemistry

Scientific supervisor: associate professor Yaroslava Pushkarova

Keywords: drug design, pharmacy, Quantitative Structure-Property Relationship, aqueous solubility

Introduction. The aqueous solubility of organic compounds is an important molecular property that plays a significant role in the behavior of compounds in various fields. Compared to labor-intensive experimental procedures for direct determination of aqueous solubility, reliable computational prediction methods have become more popular and essential in modern research.

Materials and methods. Data set – 54 organic compounds, which are characterized by 10 physical-chemical and topological descriptors. Methods of investigation: multiple linear regression – for constructing the QSPR model; correlation analysis – for determining the strength of the relationship between molecular descriptors and aqueous solubility; statistical analysis – for evaluating the quality of the constructed model using statistical indicators (coefficient of determination, adjusted coefficient of determination, residuals, and relative error of residuals).

Results. Given the current advancements in computational chemistry and machine learning, QSPR methods hold enormous potential for further development. They not only assist in predicting the key physicochemical properties of molecules but can also be used to create new approaches in drug development, personalized medicine, and the enhancement of technologies used in the pharmaceutical industry.

As a result of a detailed analysis of the significance of 10 physicochemical and topological descriptors for predicting the aqueous solubility of organic

compounds, the optimal form of the QSPR model was established. This model is characterized by high values of statistical parameters and includes only three parameters:

$$\log S_w = -0,98629 - 0,43173 \cdot x_1 + 0,005072 \cdot x_2 - 0,00476 \cdot x_3,$$

where x_1 – logarithm of the partition coefficient, x_2 – topological polar surface area, x_3 – molecular weight.

Conclusions. It has been determined that for predicting the aqueous solubility of organic compounds, including pharmaceutically active and excipient substances, the most significant descriptors are the logarithm of the partition coefficient, topological polar surface area, and molecular weight.

A three-parameter QSPR model for predicting the aqueous solubility of organic compounds demonstrates high accuracy and reliability.

The results of the study simplify the process of predicting the aqueous solubility of organic compounds and may be useful in the development of drug products with optimal pharmacokinetic characteristics.