

УДК 544.165:54.31:004.94

**С. І. Стельмах, В. Є. Кузьмін, Л. М. Огніченко**

Одеський національний університет ім. І. І. Мечникова, кафедра органічної та фармацевтичної хімії, вул. Преображенська, 24, Одеса, 65026, Україна.  
E-mail: s.i.stelmakh@gmail.com

## QSPR МОДЕЛІ ДЛЯ ПРОГНОЗУ ДЗЕТА-ПОТЕНЦІАЛІВ НАНОЧАСТИНОК ОКСИДІВ

Набір даних з експериментально визначеним дзета-потенціалом, що складався з 14 наночастинок оксидів різного розміру, було використано для розрахунку дескрипторів SiRMS та кросс-дескрипторів методу «рідкої краплі». Побудовано ефективну QSPR модель ( $R^2 = 0.88$ ,  $R^2_{\text{test}} = 0.81$ ). Прогностичну здатність моделі перевірено за допомогою зовнішнього тестування ( $R^2_{\text{test}} = 0.84$ ), для чого використано окрему вибірку даних з 5 оксидів. Моделі інтерпретовано та відображена можливість задовільного прогнозування дзета-потенціалу.

**Ключові слова:** QSPR моделювання, наночастинок, дзета-потенціал, SiRMS, liquid drop.

На протязі останніх років стає дедалі більше сфер використання наночастинок (далі – НЧ) оксидів. Як відомо, такі об'єкти можуть знайти особливо широке застосування в клінічній діагностиці і практиці. Наприклад, можливо їх застосування в якості транспортних агентів, клінічних маркерів, сигнальних систем, агентів адресної доставки лікарських засобів і т. д. [1].

Однак ефективне застосування НЧ оксидів на практиці ускладнене через цитотоксичність таких об'єктів. Ця цитотоксичність пов'язана не тільки з виділенням токсичних металів з НЧ в біологічне середовище, але також з ефектами, обумовленими фізико-хімічними параметрами: розміром і поверхневим зарядом частинок [2]. Відомо, що індукована зміна поверхневих зарядів є одним з факторів, які змінюють загальну токсичність НЧ [3].

Для характеристики фізико-хімічних параметрів НЧ використовується велика кількість методів і технік. Однією з найважливіших характеристик є поверхневий заряд. Величина поверхневого заряду НЧ на практиці характеризується значенням дзета-потенціалу ( $\zeta$ ) [4].

В роботі [5] була відображена користь цих даних для опису цілого ряду властивостей НЧ. Дзета-потенціал є визначальним параметром для агрегаційних властивостей, взаємодій частинок з середовищем і їхньої електростатичної стабільності [6].

Вимірювання дзета-потенціалу зазвичай супроводжується використанням утиліт і програмного забезпечення, які суттєво спрощують такі процеси, проте існує ряд труднощів, а саме: великі інтервали похибок в даних розмірів НЧ, велика залежність від умов експерименту (розчинник, рН, іонна сила розчину, температура), часта відсутність зазначення умов експерименту в публікаціях.

Комп'ютерне моделювання може бути використане для прогнозу різноманітних властивостей НЧ і розуміння зв'язків структура-властивість.

Завдання кожної комп'ютерної моделі (Quantitative Structure-Activity Relationship, Quantitative Structure-Property Relationship, Quantitative Nanostructure-Activity Relationship) – встановити кількісні відношення між молекулярними дескрип-

торами і конкретною властивістю (наприклад, цитотоксичність / біоактивність). Однак існують певні труднощі, що перешкоджають проведенню результативного QSPR-моделювання НЧ. Наприклад, пошук якісної бази даних із заданими властивостями виявився трудомістким: багато даних виявилися неточними, неповними або уривистими. Очевидно, що якість експериментальних даних безпосередньо залежить від багатьох факторів: лабораторного устаткування, організації внутрішнього розпорядку, навичок дослідників, тощо. Внаслідок порушень алгоритмів роботи з початковими потоками даних виникають помилки та спотворення даних, що часто потребує проведення додаткових процедур з «курації» даних. Відсутність інформації про фізичні умови проведення процедури виміру заданої властивості НЧ також стало причиною бракування таких баз даних.

Метою цієї роботи є демонстрація ефективності розроблених нами дескрипторних систем – фрагментного 1D-представлення неорганічних сполук [7] та методу «рідкої краплі» [7] для вирішення QSPR завдань щодо дзета-потенціалів нанооксидів.

## МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ

### Набір даних

Процеси відбору і так званої «курації» даних є одними з найважливіших етапів під час проведення будь-яких досліджень, що будуються на статистичному моделюванні. Основні та найчастіші помилки при QSPR-моделюванні виникають саме через застосування помилкових або неточних даних.

Об'єктами дослідження були оксиди, що мали щонайменше три експериментально виміряні властивості: дзета-потенціал, номінальний розмір (далі – НР) НЧ та гідродинамічний радіус (далі – ГР).

Для формування вибірки було проведено збір максимально повних даних з різних джерел, основою стала публікація [3] з найбільшою чисельністю записів, після чого проведена їх корекція, включаючи перевірку за декількома джерелами (оригінальними публікаціями), виправлення і видалення помилкових та неповних даних. Основну вибірку даних було доповнено додатковими записами [5].

При підготовці вибірки даних для побудови моделей QSPR, відбиралися лише ті оксиди, які мали три або більше експериментальних вимірів досліджуваної властивості.

Наступним етапом стала перевірка і корекція вхідних даних досліджуваних НЧ. Був проведений моніторинг значень дзета-потенціалів у вихідних публікаціях та базах даних, виявлені та виправлені вочевидь помилкові дані.

Явні експериментальні викиди, значення досліджуваної властивості яких суттєво різнилися з діапазоном значень властивостей елементів вибірки аналогічних їм по природі (напр., дзета-потенціал  $\zeta = -20.2$  мВ при діапазоні 30.3–43.0 мВ у групі з семи НЧ  $\text{Al}_2\text{O}_3$ ), були додатково проаналізовані, використовуючи геометричні параметри (такі як НР, НЧ та ГР) та виключені з вибірки даних через невідповідність основному тренду параметрів.

На цьому етапі, з початкових 99 записів [3, 5] для 27 НЧ, у вибірці внаслідок проведення процедур корекції даних, нарахувалось 77 записів для 16 НЧ ( $\text{Al}_2\text{O}_3$ ,  $\text{CeO}_2$ ,  $\text{CuO}$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{Fe}_3\text{O}_4$ ,  $\text{In}_2\text{O}_3$ ,  $\text{La}_2\text{O}_3$ ,  $\text{NiO}$ ,  $\text{Sb}_2\text{O}_3$ ,  $\text{SiO}_2$ ,  $\text{SnO}_2$ ,  $\text{TiO}_2$ ,  $\text{WO}_3$ ,  $\text{Y}_2\text{O}_3$ ,  $\text{ZnO}$ ,  $\text{ZrO}_2$ ).

Надалі, з вибірки були вилучені дані, до яких у оригінальних публікаціях не наведено або наведено не в повному обсязі фізико-хімічні параметри та умови вимірів досліджуваної величини та сформовано вибірку, що нараховує 14 НЧ та 37 відповідних записів досліджуваної величини (табл. 1) з вимірами при рН=7.

Таблиця 1  
Значення експериментальних та розрахованих параметрів НЧ оксидів  
для досліджуваної вибірки

Table 1

Observed and predicted values for parameters of oxide NP set

НЧ	$\zeta_{\text{експ}}$ (мВ)	$\zeta_{\text{розрах}}$ (мВ)	НР (нм)	ГР (нм)	НЧ	$\zeta_{\text{експ}}$ (мВ)	$\zeta_{\text{розрах}}$ (мВ)	НР (нм)	ГР (нм)
Al <sub>2</sub> O <sub>3</sub>	39.2	39.6	11.4	94.7	Sb <sub>2</sub> O <sub>3</sub>	-35.3	-27.2	11.8	147.6
Al <sub>2</sub> O <sub>3</sub>	38	39.2	13.56	312.6	SiO <sub>2</sub>	-8.1	-23.4	15	809.7
Al <sub>2</sub> O <sub>3</sub>	43	38.0	30	210	SiO <sub>2</sub>	-31.8	-21.4	13.5	113.4
CeO <sub>2</sub>	26.5	27.8	13.0	200.7	SiO <sub>2</sub>	-29.8	-25.7	20	1230
CeO <sub>2</sub>	21.4	27.3	18.3	197.6	SnO <sub>2</sub>	-21.1	-23.1	15	264.9
CuO	37.4	13.9	23.1	171.7	TiO <sub>2</sub>	-13.5	-14.7	12.2	171
CuO	17	20.5	12.5	130	TiO <sub>2</sub>	-19.4	3.9	12.6	166
CuO	7.6	20.2	12.8	263.3	TiO <sub>2</sub>	15	3.7	10	116
CuO	24.4	18.9	28	285	TiO <sub>2</sub>	7.09	-0.1	16	1500
Fe <sub>2</sub> O <sub>3</sub>	-22.8	-15.9	30	942	WO <sub>3</sub>	-45.2	-54.2	10.6	62.8
Fe <sub>2</sub> O <sub>3</sub>	-11.2	-16.5	20	1565	WO <sub>3</sub>	-61.3	-53.4	16.6	176.6
Fe <sub>2</sub> O <sub>3</sub>	-2.1	-10.6	12.3	144.7	WO <sub>3</sub>	-54.4	-55.2	15.9	545.5
Fe <sub>3</sub> O <sub>4</sub>	-17.7	-10.1	20	685	ZnO	-10.8	2.3	20	161.3
Fe <sub>3</sub> O <sub>4</sub>	-2.1	-8.2	12	198.4	ZnO	16.4	9.7	30	501
La <sub>2</sub> O <sub>3</sub>	54.3	54.1	24.6	672.9	ZnO	0.02	7.8	20	759
NiO	13.3	16.3	30	1634	ZnO	28.8	9.4	22.6	204.5
NiO	27.6	25.2	13.1	228	ZnO	-15.0	9.3	13	413
NiO	26	24.6	14	399	ZrO <sub>2</sub>	-6.9	-4.2	27	2337
Sb <sub>2</sub> O <sub>3</sub>	-24.2	-25.2	20.8	124.7					

З метою оцінки прогнозуючої здатності QSPR-моделей, з бази даних «Cancer Nanotechnology Laboratory» [8], була сформована вибірка для проведення зовнішнього тестування, що нараховувала 8 записів по 5 наноксидам (табл. 2).

Таблиця 2

## Параметри НЧ оксидів для зовнішнього тестування

Table 2

Parameters of oxide NP set for external testing			
НЧ	$\zeta_{\text{експ}}$ (мВ)	НР (нм)	ГР (нм)
CuO	31	42	220
Fe <sub>2</sub> O <sub>3</sub>	-17.3	29	1580
Fe <sub>3</sub> O <sub>4</sub>	1.8	30	200
SiO <sub>2</sub>	47.9	60	218
SiO <sub>2</sub>	-30.94	60	1000
SiO <sub>2</sub>	-27.34	60	99
SiO <sub>2</sub>	-25.68	60	114
TiO <sub>2</sub>	5.9	63	300

**Дескриптори**

Для виконання задачі з QSPR-моделювання нами використовувались наступні типи дескрипторів:

1. Дескриптори 1D рівня розраховувались із використанням підходу симплексного представлення неорганічних сполук [7]. Цей підхід базується на врахуванні кількості різних комбінацій атомів (від одиниць до чотирьох), які включені в «brutto» або емпіричну формулу (формулу Хілла). У той же час атоми характеризуються мітками, що відображають їх положення в періодичній системі елементів, електро-негативність та інші легко інтерпретовані характеристики.

2. Дескриптори, що характеризують атоми елементу у оксиді: заряд ядра атома елементу (Z), потенціал іонізації (Ip) та дескриптори, розраховані на основі геометричних характеристик НЧ (радіус Вігнера-Зейтца, кількість молекул у нанокластері, площа поверхні, тощо). Ці геометричні дескриптори базуються на моделі «рідкої краплі», яка дозволяє припустити що найбільш ймовірною формою НЧ є сферична крапля [7].

Крім індивідуальних дескрипторів використовувались парні крос-дескриптори (добуток величин різних пар дескрипторів), які характеризують вплив на властивість, що вивчається, ефекту взаємодії різних структурних факторів сполук.

Початковий набір структурних параметрів становив 696 дескрипторів, з них 276 дескрипторів першої групи, 20 – другої групи. Набір парних крос-дескрипторів було побудовано на основі дескрипторів другої групи і нараховував 400 параметрів.

**Побудова QSPR моделей**

Зважаючи на велику кількість отриманих структурних параметрів, для побудови моделей застосовувався метод машинного навчання PLS [9] (часткових найменших квадратів), що дозволяє зменшувати оброблювану структурну інформацію до невеликої кількості латентних змінних.

PLS- рівняння може бути представлено як

$$Y = b_0 + \sum_{i=1}^N b_i x_i,$$

де  $Y$  – досліджувана властивість,  $b_i$  – регресійні коефіцієнти PLS,  $x_i$  – значення  $i$ -го дескриптора,  $N$  – загальне число дескрипторів.

Для оцінки надійності побудованих моделей та здатності до задовільного прогнозу було проведено процедуру п'ятикратного перехресного затвердження. Для цього основну вибірку було поділено на п'ять репрезентативних тестових підвбірок однакового розміру (20% сполук). Для кожної тестової підвбірки решта сполук (80%) використовується як навчальна підвбірка, у результаті чого натреновуються та випробовуються п'ять моделей.

## РЕЗУЛЬТАТИ ТА ЇХ ОБГОВОРЕННЯ

У результаті побудови консенсусної PLS QSPR моделі було задіяно 106 дескрипторів, які складали 2 латентні змінні. Співвідношення між спрогнозованими на спостережуваними значеннями досліджуваної властивості можна проаналізувати за допомогою табл. 1 та рис. 1.

У табл. 3 наведені параметри з максимальним відносним внеском в розрахункове значення досліджуваної властивості.

Жирним шрифтом виділено групи, в які входять атоми в залежності від того, в який діапазон зміни значень певної властивості вони потрапляють.

При розрахунку 1D дескрипторів для диференціації атомів використовувалися наступні атомні характеристики:

- номер групи – максимальне значення головного квантового числа: від 1 до 18  
 $1 = A < 2 = B < 3 = C < 4 = D < 5 = E < 6 = F < 7 = G < 8 = H < 9 = I < 10 = J < 11 = K < 12 = L < 13 = M < 14 = N < 15 = O < 16 = P < 17 = Q < 18 = R$ ;
- ступінь окиснення – інтервали значень:  $A < +2 = B < +3 = C < +4 = D < +6 = E$ ;
- заряд ядра – діапазон зміни значень на інтервали:  $A < 25 \leq B < 45 \leq C < 65 \leq D$ ;
- внесок атомів у йонність парних взаємодій атомів:  $A < 0.5 \leq B < 1.0 \leq C < 2.0 \leq D < 3.0 \leq E$ ;
- Ван дер Ваальсові взаємодії: притягання  $A < 50 \leq B < 100 \leq C < 250 \leq D < 400 \leq E < 650 \leq F < 2000 \leq G$ , відштовхування  $A < 20000 \leq B < 32000 \leq C < 50000 \leq D < 100000$ .

Відповідно до концепції Л. Полінгу, йонність парних взаємодій (BI) для атомів A і B характеризує ступінь полярності, яка залежить від різниці електронегативностей пар атомів, що беруть участь у взаємодії:

$$BI(A-B) = 1 - \exp[-0.25 (X_A - X_B)^2],$$

де  $X_A$  та  $X_B$  – електронегативність A та B, відповідно. Далі, значення BI розподіляється між атомами, що утворюють цю пару –  $BI(A) = BI(B) = 0,5 (BI(A-B))$ .

Таблиця 3

## Параметри з максимальними внесками

Table 3

## High contribution parameters

Дескриптор	Внесок (%)
(йонний радіус за Бокієм) <sup>2</sup>	8.97
кількість двійок <b>J, P</b> , які диференційовані за номером групи періодичної системи (максимальне значення головного квантового числа: від 1 до 18)	5.09
кількість четвірок <b>A, C, C, C</b> , зважених за електронною поляризованістю	4.83
кількість одиниць <b>D</b> , зважених за ступенем окиснення	4.54
кількість одиниць <b>B</b> , зважених за Ван дер Ваальсовими взаємодіями (притягання)	4.38
потенціал іонізації металу	3.77
кількість четвірок <b>C, C, C, C</b> , зважених за Ван дер Ваальсовими взаємодіями (притягання)	3.49
кількість трійок <b>C, P, P</b> , які диференційовані за номером групи періодичної системи	3.18
(йонний радіус за Бокієм)·(радіус Вігнера-Зейца)	3.12
кількість одиниць <b>A</b> , зважених за внеском атомів у йонність парних взаємодій атомів	3.12
кількість одиниць <b>C</b> , зважених Ван дер Ваальсовими взаємодіями (відштовхування)	2.82
(потенціал іонізації металу)·(атомний радіус)	2.09
густина оксиду	2.07
кількість двійок <b>K, P</b> , які диференційовані за номером групи періодичної системи	2.05
(електронегативність металу) <sup>2</sup>	1.91
(F-поверхня молекул)·(атомний радіус металу)	1.88
(густина оксиду)·(йонний радіус за Бокієм)	1.84
кількість двійок <b>B, D</b> , зважених за Ван дер Ваальсовими взаємодіями (відштовхування)	1.78
кількість трійок <b>O, O, P</b> , які диференційовані за номером групи періодичної системи	1.77
(атомний радіус металу)·(z –заряд йону металу)	1.41
кількість одиниць <b>A</b> , зважених за електронною поляризованістю	1.36
(ковалентний індекс йону металу)·(F-поверхня молекул)	1.35
кількість одиниць <b>H</b> , які диференційовані за номером групи періодичної системи	1.30
кількість одиниць <b>B</b> , зважених за зарядами ядер атомів	1.27
кількість одиниць <b>C</b> , зважених за зарядами ядер атомів	1.23
(номінальний розмір)·(F-поверхня молекул)	1.18
(електронегативність металу)·(потенціал іонізації металу)	1.10
кількість одиниць <b>D</b> , зважених за Ван дер Ваальсовими взаємодіями (відштовхування)	1.03
кількість одиниць <b>E</b> , зважених за ступенем окиснення	1.02
кількість двійок <b>A, C</b> , зважених за електронною поляризованістю	1.02

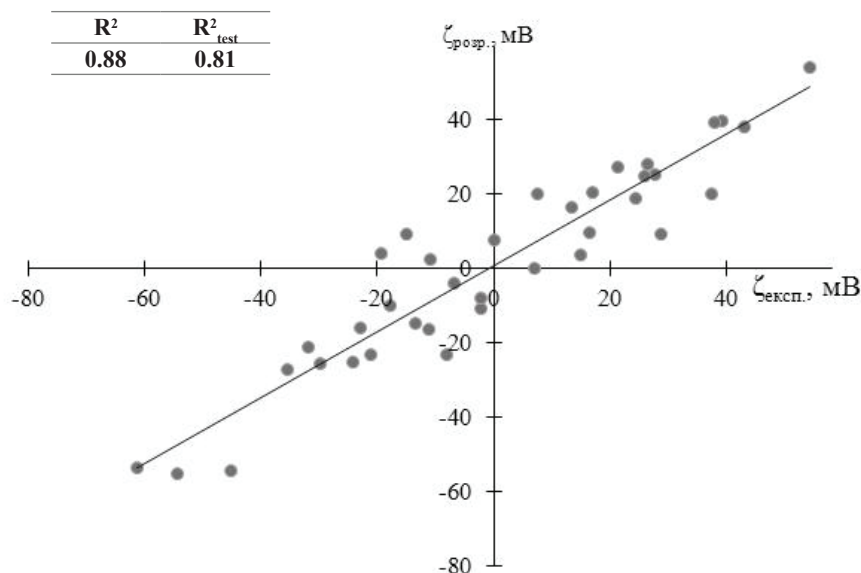


Рис. 1. Спостережувані/спрогнозовані значення дзета-потенціалів

Fig. 1. Observed/predicted zeta-potential values

Таким чином, для всіх атомів парних взаємодій отримуємо усереднені значення. Потім для кожного атому молекули підсумовуються усі усереднені значення його внесків у йонність парних взаємодій.

Загальний внесок групи дескрипторів з максимальними відносними внесками склав ~80%. Інтерпретація моделі проводилась застосовуючи цю групу дескрипторів та було виявлено, що внесок дескрипторів розподілився між індивідуальними дескрипторами та крос-дескрипторами і становить 46% та 54% відповідно. Це свідчить про важливість урахування ефектів взаємодії структурних факторів. Внесок комбінаторних дескрипторів першої групи склав 59%, а другої групи дескрипторів – 41% (дескрипторів рідкої краплі – 29%, дескрипторів, що характеризують атом металу – 12%).

За допомогою Williams plot [10] було виявлено, що одна з  $\text{HfWO}_3$  та  $\text{La}_2\text{O}_3$  не входять у межі області застосовності побудованої QSPR моделі, що свідчить про неможливість прогнозу досліджуваної властивості для даних сполук.

Додатково для оцінки прогнозуючої здатності QSPR-моделі, була використана зовнішня вибірка даних, де представлені 8 записів по 5 оксидам, що не використовувалася при побудові моделей. Співвідношення між спрогнозованими на спостережуваними значеннями досліджуваної властивості зовнішньої тестової вибірки можна проаналізувати за допомогою рис. 2. Підтвердження прогнозування в межах області застосовності побудованої QSPR моделі було здійснене за допомогою Williams plot.

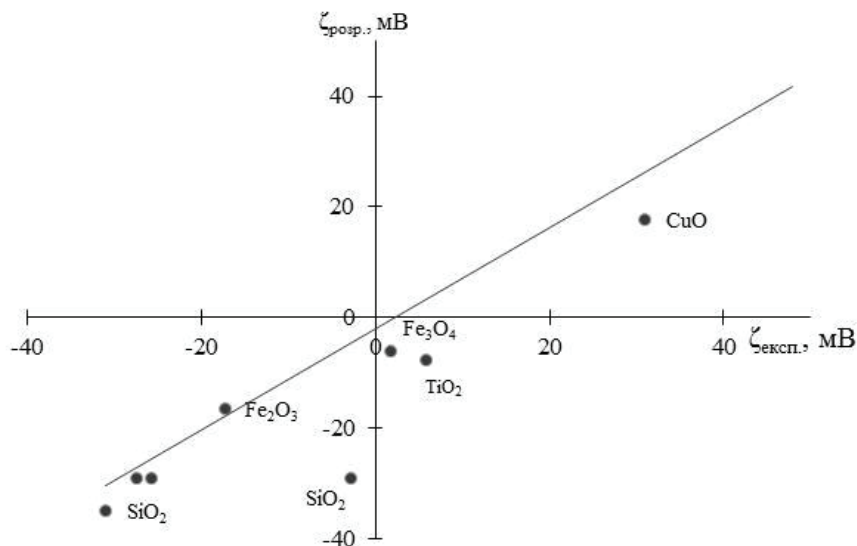


Рис. 2. Спостережувані/спрогнозовані значення дзета-потенціалів для зовнішньої тестової вибірки

Fig. 2. Observed/predicted zeta-potential values for external test set

## ВИСНОВКИ

У результаті проведеної роботи вдалося побудувати QSPR модель на 1D рівні представленні молекулярної структури неорганічних сполук за допомогою комбінаторних дескрипторів із задовільною здатністю прогнозування дзета-потенціалів наночастинок оксидів. При цьому вклад парних крос-дескрипторів, що відображають ефекти сумісного впливу, склав більше 50%. Виявлено, що найбільш впливовими параметрами є характеристики, що відображають природу оксидів, серед яких максимальний вклад мають параметри електростатичних взаємодій.

Автори висловлюють щире подяку член-кореспонденту НАН України М.О. Мchedlov-Петросяну за проведені консультації та допомогу з проведенням процедур «курації» даних.

## СПИСОК ЛІТЕРАТУРИ

1. Salata O. Applications of nanoparticles in biology and medicine. // J. Nanobiotechnol. – 2004. – Vol. 2, N3. <https://doi.org/10.1186/1477-3155-2-3>
2. Sukhanova A., Poly S., Bozrova S., Lambert É., Ewald M., Karaulov A., Molinari M., Nabiev I. Nanoparticles With a Specific Size and Surface Charge Promote Disruption of the Secondary Structure and Amyloid-Like Fibrillation of Human Insulin Under Physiological Conditions. // Front. Chem. – 2019. – Vol. 7. – An. 480. <https://doi.org/10.3389/fchem.2019.00480>
3. Toropov A.A., Sizochenko N., Toropova A. P., Leszczynski J. Towards the Development of Global Nano-Quantitative Structure-Property Relationship Models: Zeta Potentials of Metal Oxide Nanoparticles. // Nanomater. (Basel). – 2018. – Vol. 8, N4. – An 243. <https://doi.org/10.3390/nano8040243>
4. Chen L., Fu X., Li J. Ultrasensitive surface-enhanced Raman scattering detection of trypsin based on anti-aggregation of 4-mercaptopyridine-functionalized silver nanoparticles: An optical sensing platform toward proteases. // Nanoscale. – 2013. – Vol. 5, N13. – P. 5905–5911. <https://doi.org/10.1039/C3NR00637A>



5. Mikolajczyk A, Gajewicz A, Rasulev B., Schaeublin N., Maurer-Gardner E., Hussain S., Leszczynski J., Puzyn T. Zeta Potential for Metal Oxide Nanoparticles: A Predictive Model Developed by a Nano-Quantitative Structure–Property Relationship Approach // *Chem. Mater.*– 2015.– Vol. 27, N7.– P. 2400–2407. <https://doi.org/10.1021/cm504406a>
6. Wyrzykowska E., Mikolajczyk A., Sikorska C., Puzyn T. Development of a novel in silico model of zeta potential for metal oxide nanoparticles: a nano-QSPR approach. // *Nanotechnol.*– 2016.– Vol. 27, N44. <https://doi.org/10.1088/0957-4484/27/44/445702>
7. Kuz'min V., Artemenko A., Ognichenko L., Hromov A., Kosinskaya A., Stelmakh S., Sessions Z.L., Muratov E. N. Simple representation of molecular structure as universal QSAR/QSPR tool. // *Struct. Chem.*– 2021.– Vol. 32.– P. 1365–1392. <https://dx.doi.org/10.1007/s11224-021-01793-z>
8. <https://cananolab.nci.nih.gov/caNanoLab/>
9. Rännar S., Lindgren F., Geladi P., Wold S. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. // *J. Chemom.*– 1994.– Vol. 8, N2.– P. 111–125. <https://doi.org/10.1002/cem.1180080204>
10. Meloun M., Militku J., Hill M., Brereton R. G. Crucial problems in regression modelling and their solutions. // *Analyst.*– 2002.– Vol. 127, N4.– P. 433–450. <https://doi.org/10.1039/B110779H>

Стаття надійшла до редакції 04.05.2021

**С. И. Стельмах, В. Е. Кузьмин, Л. Н. Огниченко**

Одесский национальный университет им. И. И. Мечникова, кафедра органической и фармацевтической химии, ул. Преображенская, 24, Одесса, 65026, Украина.  
E-mail: s.i.stelmakh@gmail.com

## QSPR МОДЕЛИ ДЛЯ ПРОГНОЗА ДЗЕТА-ПОТЕНЦИАЛОВ НАНОЧАСТИЦ ОКСИДОВ

Набор данных с экспериментально определенными значениями дзета-потенциала, состоящий из 37 записей по 14 наночастицам оксидов разного размера, был использован для расчета дескрипторов SiRMS и кросс-дескрипторов метода «жидкой капли». Построена эффективная консенсусная QSPR модель ( $R^2 = 0.88$ ,  $R^2_{\text{test}} = 0.81$ ). Прогностическая способность модели проверена с помощью внешнего тестирования ( $R^2_{\text{test}} = 0.84$ ), для чего использовалась отдельная выборка данных из 5 оксидов. Модели интерпретированы и отражена возможность удовлетворительного прогнозирования дзета-потенциала.

**Ключевые слова:** QSPR моделирование, наночастицы, дзета-потенциал, SiRMS, liquid drop.

**S. I. Stelmakh, V. E. Kuz'min, L. M. Ognichenko**

Odesa Mechnikov National University, Department of Organic and Pharmaceutical Chemistry, Preobrazhenska str., 24, Odesa, 65026, Ukraine.  
E-mail: s.i.stelmakh@gmail.com

## QSPR MODELS FOR ZETA-POTENTIAL OF NANO-OXIDES PREDICTION

Nano-QSPR modeling often requires considering variety of factors, if neglected, may lead to erroneous result of the study. Frequently, the data turned out to be inaccurate, incomplete, or fragmentary. Obviously, the quality of experimental data directly depends on many factors: laboratory equipment, organization of internal regulations, skills of researchers, and so on. As a result of violations of algorithms and protocols of initial data streams processing – there are

errors and distortions of data, that is why performing a solid multistep data-curation process is crucial for such procedures. Data curation procedure was performed and approximately 60% was rejected (due to various errors, incomplete or absent records for physicochemical parameters or conditions of performed experiment), followed up by using zeta-potential value dataset for 37 various sizes nanoparticles of 14 metal oxides for calculation of 1D SiRMS descriptors as well as «liquid drop» model cross-descriptors. An efficient consensus model was built ( $R^2 = 0.88$ ,  $R^2_{\text{test}} = 0.81$ ). Predictive power ( $R^2 = 0.84$ ) of the model was tested using an external set of 5 nano-oxides and the possibility of satisfactory zeta-potential prediction was shown. Prediction of zeta-potential value within domain applicability of obtained QSPR model confirmed using a Williams plot. The interpretation of the final model was carried out and it was found that the contribution of descriptors was distributed between individual descriptors and cross-descriptors by 46% and 54% respectively. The contribution 1D SiRMS descriptors was 59%, the second group – 41% (liquid drop model descriptors – 29%, descriptors characterizing the metal atom – 12%). It was found that the most influential parameters are the characteristics that reflect the nature of the oxides. The parameters of electrostatic interactions have the highest contribution.

**Key words:** QSPR modeling, nanoparticles, zeta-potential, SiRMS, liquid drop.

## REFERENCES

1. Salata O. *Applications of nanoparticles in biology and medicine*. J. Nanobiotechnol., 2004, vol. 10, no 3. <https://doi.org/10.1186/1477-3155-2-3>
2. Sukhanova A., Poly S., Bozrova S., Lambert É., Ewald M., Karaulov A., Molinari M., Nabiev I. *Nanoparticles With a Specific Size and Surface Charge Promote Disruption of the Secondary Structure and Amyloid-Like Fibrillation of Human Insulin Under Physiological Conditions*. Front. Chem., 2019, vol. 7, an 480. <https://doi.org/10.3389/fchem.2019.00480>
3. Toropov A.A., Sizochenko N., Toropova A.P., Leszczynski J. *Towards the Development of Global Nano-Quantitative Structure-Property Relationship Models: Zeta Potentials of Metal Oxide Nanoparticles*. Nanomater. (Basel), 2018, vol. 8, no 4, an 243. <https://doi.org/10.3390/nano8040243>
4. Chen L., Fu X., Li J. *Ultrasensitive surface-enhanced Raman scattering detection of trypsin based on anti-aggregation of 4-mercaptopyridine-functionalized silver nanoparticles: An optical sensing platform toward proteases*. Nanoscale, 2013, vol. 5, no 13, pp. 5905–5911. <https://doi.org/10.1039/C3NR00637A>
5. Mikolajczyk A., Gajewicz A., Rasulev B., Schaeublin N., Maurer-Gardner E., Hussain S., Leszczynski J., Puzyn T. *Zeta Potential for Metal Oxide Nanoparticles: A Predictive Model Developed by a Nano-Quantitative Structure-Property Relationship Approach*. Chem. Mater., 2015, vol. 27, no 7, pp. 2400–2407. <https://doi.org/10.1021/cm504406a>
6. Wyrzykowska E., Mikolajczyk A., Sikorska C., Puzyn T. *Development of a novel in silico model of zeta potential for metal oxide nanoparticles: a nano-QSPR approach*. Nanotechnol., 2016, vol. 27, no 44. <https://doi.org/10.1088/0957-4484/27/44/445702>
7. Kuz'min V., Artemenko A., Ognichenko L., Hromov A., Kosinskaya A., Stelmakh S., Sessions Z.L., Muratov E.N. *Simplex representation of molecular structure as universal QSAR/QSPR tool*. Struct. Chem., 2021, vol. 32, pp. 1365–1392. <https://doi.org/10.1007/s11224-021-01793-z>
8. <https://cananolab.nci.nih.gov/caNanoLab/>
9. Rännar S., Lindgren F., Geladi P., Wold S. *A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm*. J. Chemom., 1994, vol. 8, no 2, pp. 111–125. <https://doi.org/10.1002/cem.1180080204>
10. Meloun M., Militku J., Hill M., Brereton R.G. *Crucial problems in regression modelling and their solutions*. Analyst, 2002, vol. 127, no 4, pp. 433–450. <https://doi.org/10.1039/B110779H>