АНОТАЦІЯ

Стельмах С. І. Аналіз і прогнозування властивостей молекулярних нанооб'єктів методами хемоінформатики. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 102-Хімія. Одеський національний університет імені І. І. Мечникова, МОН України, Одеса, 2021.

Дисертаційна робота присвячена дослідженню особливостей впливу природи та розмірних характеристик низки наночастинок оксидів на їх цитотоксичність та властивості в рамках симплексного представлення молекулярної будови методами nanoQSAR/QSPR.

В результаті ретельного пошуку та детального порівняльного аналізу існуючих баз даних нанооб'єктів, що містять інформацію необхідну для nanoQSAR/QSPR моделювання – виявлено потребу у створенні бази даних, придатної до застосування в рамках даного дослідження. Сконструйовано Комбіновану Базу Даних Нанооксидів (КБДН), яка містить інформацію щодо 188 наночастинок оксидів з огляду на такі їх характеристики як: будова, параметри розміру (радіус частинки, гідродинамічний радіус); та активності/властивості: дзета-потенціал, енергія E_g (заборонені зони – ширини проміжків значень енергії в яких не існує делокалізованих одноелектронних станів), дані щодо цитотоксичності до клітин *Escherichia coli* та клітин HaCaT.

Використовуючи дані з КБДН, проведено процедуру «курації» даних та сформовано чотири вибірки для дослідження впливу параметрів наночастинок на відповідну активність/властивість. Побудована система 1D дескрипторного опису наночастинок оксидів із застосуванням симплексних, інтегральних, «рідкої краплі», індивідуальних та крос-дескрипторів.

Оримані 1D моделі наночастинок оксидів були використані для побудови адекватних nanoQSAR/QSPR моделей для відповідних

активностей/властивостей за допомогою методу машинного навчання PLS (часткових найменших квадратів).

Одержані nanoQSAR моделі цитотоксичності до клітин *Escherichia coli* (R^2 = 0.93, R^2_{test} = 0.97) та клітин HaCaT (R^2 = 0.83, R^2_{test} = 0.91) були валідовані та інтерпретовані. Крос-валідацію (Q^2_{LOO} = 0.90, Q^2_{LOO} = 0.71 відповідно) здійснено за процедурою виключення по одному (Leave-One-Out). Проведений кластерний аналіз та інтерпретація за відносними дескрипторними внесками показали, що одним з основних факторів, що визначає цитотоксичність наночастинок оксидів, є величина заряду йону металу.

Побудовано консенсусну nanoQSPR модель для дзета-потенціалу (R^2 = 0.89, R^2_{test} = 0.81). Валідацію моделі (Q^2_{cv} = 0.81) проведено відповідно до процедури п'ятикратної перехресної перевірки (five-fold cross-validation). Здійснено додаткове оцінювання прогностичної здатності моделі за допомогою зовнішнього тестування ($R^2_{ext.test}$ = 0.83). Проведена інтерпретація показала значний вплив взаємодії структрних факторів – сумарний відносний внесок крос-дескрипторів склав ~81%.

Проведено nanoQSPR моделювання енергії E_g. Статистичні показники побудованої консенсусної моделі (R^2 = 0.83, Q^2_{cv} = 0.74, R^2_{test} = 0.73) свідчать про здатність до задовільного прогнозу досліджуваної властивості. Було проведено структурну інтерпретацію отриманої моделі, за результатами якої виявлено, що найбільш впливовими є фактори електростатичних та Ван дер Ваальсових взаємодій.

Отримані nanoQSAR/QSPR моделі для цитотоксичності до клітин *Escherichia coli* та HaCaT, дзета-потенціалу, енергії E_g було об'єднано в експертну систему «*nanoExpert*». Експертну систему було інтегровано у програмне забезпечення Methods of Data Analysis (© Artemenko A.G.) та інтерфейс користувача. Показано, що експертна система є програмним забезпеченням орієнтованим на широку аудиторію і придатним до проведення моделювання користувачами низки активностей/властивостей

наночастинок оксидів без наявності спеціальних навичок в галузі хемоінформатики та nanoQSAR/QSPR моделювання.

Ключові слова: наночастинки оксидів, комп'ютерне моделювання, взаємозв'язок структура-активність, взаємозв'язок структура-властивість, симплексне представлення молекулярної будови, цитотоксичність, дзета-потенціал, заборонена зона.

SUMMARY

Stelmakh S.I. Analysis and prediction of properties of molecular nanoobjects by chemoinformatics methods.

Thesis for scientific degree of Doctor of Philosophy in Chemistry (speciality 102-Chemistry). I.I. Mechnikov National University of Odesa, Ministry of Education and Science of Ukraine, Odesa, 2021.

The thesis is devoted to the study of influence of the peculiarities of nature and dimensional characteristics on cytotoxicity and other properties of a row of oxide nanoparticles by nanoQSAR/QSPR methods within simplex representation of molecular structure approach.

A thorough research and detailed comparative analysis of existing databases of nanoobjects containing information required for nanoQSAR/QSPR modeling revealed the need to create a database suitable for usage in scope of this study. The Combined Database of Nanooxides (CDN) has been constructed, which contains information on 188 oxide nanoparticles including such characteristics as structure and size parameters as well as some activities/properties: zeta potential, energy E_g (band gaps - an energy range in a solid where no electronic states can exist), data on cytotoxicity to *Escherichia coli* cells and HaCaT cells.

The data curation procedure was performed using data from the CDN and four datasets were formed to study the influence of nanoparticle parameters on the corresponding activity/property. A system of 1D descriptors of oxide nanoparticles is constructed with the application of simplex, integral, "liquid drop", individual and cross-descriptors.

The obtained 1D models of oxide nanoparticles were used to construct adequate nanoQSAR / QSPR models for the corresponding activities/properties using the PLS (partial least squares) machine learning method.

The obtained nanoQSAR models of cytotoxicity to *Escherichia coli* cells ($R^2 = 0.93$, $R^2_{test} = 0.97$) and HaCaT cells ($R^2 = 0.83$, $R^2_{test} = 0.91$) were validated and interpreted. Cross-validation ($Q^2_{LOO} = 0.90$, $Q^2_{LOO} = 0.71$ respectively) was performed by the Leave-One-Out procedure. The cluster analysis and

interpretation by descriptors relative influence showed that one of the main factors determining the cytotoxicity of oxide nanoparticles is the magnitude of the metal ion charge.

A consensus nanoQSPR model for the zeta potential was constructed (R^2 = 0.89, R^2_{test} = 0.81). Model validation (Q^2_{cv} = 0.81) was performed according to the five-fold cross-validation procedure. An additional assessment of the predictive ability of the model was performed using external testing ($R^2_{ext.test}$ = 0.83). The performed interpretation showed a significant influence of the interaction of structural factors - total relative contribution of cross-descriptors was ~ 81%.

NanoQSPR modeling of E_g energy was performed. Statistical indicators of the constructed consensus model (R^2 = 0.83, Q^2_{cv} = 0.74, R^2t_{est} = 0.73) indicate the ability to adequately predict the studied property. A structural interpretation of the obtained model was performed, it was found that the most influential factors are electrostatic and Van der Waals interactions.

The obtained nanoQSAR/QSPR models for cytotoxicity to *Escherichia coli* and HaCaT cells, zeta potential, E_g energy were combined into the "*nanoExpert*" expert system. The expert system was integrated into the Methods of Data Analysis software (© Artemenko A.G.) and the corresponding user interface. It is shown that the expert system is software aimed at a wide audience and suitable for modeling of a row of activities/properties by users without special skills in chemoinformatics and nanoQSAR/QSPR modeling.

Keywords: oxide nanoparticles, computer modeling, structure-activity relationship, structure-property relationship, simplex representation of molecular structure, cytotoxicity, zeta potential, band gap.

Список публікацій здобувача за темою дисертації:

1. Kuz'min V., Ognichenko L., Sizochenko N., Chapkin V., **Stelmakh S**., Shyrykalova A., Leszczynski J. Combining Features of Metal Oxide Nanoparticles: Nano-QSAR for Cytotoxicity. *International Journal of Quantitative Structure-Property Relationships*. 2019. Vol. 4, № 1. P. 28-40.

   Особистий внесок здобувача: проведено розрахунок дескрипторів, проведено побудову, оптимізацію та валідацію моделей.
2. Kuz'min V., Artemenko A., Ognichenko L., Hromov A., Kosinskaya A., **Stelmakh S.**, Sessions Z., Muratov E. Simplex representation of molecular structure as universal QSAR/QSPR tool. *Structural Chemistry*. 2021. Vol. 32, № 5. P. 1365–1392.

   Особистий внесок здобувача: брав участь у написанні розділу огляду, присвяченого моделюванню наночастинок.
3. **Стельмах С. І.**, Кузьмін В. Є., Огніченко Л. М. QSPR моделі для прогнозу дзета-потенціалів наночастинок оксидів. *Вісник ОНУ*. 2021. Том 26, № 2(78), С. 89–100.

   Особистий внесок здобувача: проведено розрахунок дескрипторів, проведено побудову, оптимізацію та валідацію моделей. Сумісно зі співавторами проведено інтерпретацію та обговорення результатів.
4. Ognichenko L., Shyrykalova A., **Stelmakh S.**, Tinkov O., Kuz'min V. The importance of Effects of Structural Factors Interaction for Metal Oxides Nanoparticles in QSAR Models for Cytotoxicity. *Nanoscience & Nanotechnologies* : book of abstracts of 15th International Conference, Thessaloniki, 3-6 July 2018. Thessaloniki, Greece, 2018. P. 203.
5. **Stelmakh S.**, Ognichenko L., Kuz'min V. Nano-QSPR for zeta potential of metal oxides. *Chemistry, Physics and Technology of Surface* : proceedings of Ukrainian Conference with International participation dedicated to 90th birthday of Aleksey Chuiko, Academician of NAS of Ukraine, Kyiv 21-22 October 2020. Kyiv, Ukraine, 2020. P. 171.

6. **Stelmakh S.**, Kuz'min V. QSPR vs Molecular Docking. Adducts of [60] fullerene as potential HIV-1 PR inhibitors. *Science, Innovation, Quality* : book of Papers of 1st International Scientific-Practical Conference, Berdyansk 17-18 December 2020. Berdyansk, Ukraine, 2020. P. 122.