UNIVERSITÀ DEGLI STUDI
DELL'INSUBRIA

DIPARTIMENTO DI SCIENZE
TEORICHE E APPLICATE - DiSTA

Review report on the doctoral dissertation by Viacheslav Muratov

"A properties-to-omics predictive framework for hazard assessment of metal- and carbon-based nanomaterials"

The dissertation by M.Sc. Viacheslav Muratov, supervised by Prof. Dr. Tomasz Puzyn and Dr. Karolina Jagiello, presents a computational framework integrating transcriptomics, QSAR modelling, and machine learning for nanomaterial hazard assessment. The work focuses on the development of novel in silico tools to establish relationships between large-scale transcriptomic responses induced by multi-walled carbon nanotubes (MWCNTs) and titanium dioxide (TiO_2) nanomaterials and their structural characteristics, represented through physicochemical properties and theoretical descriptors. The thesis is positioned at the intersection of two rapidly evolving research fields, namely New Approach Methodologies (NAMs) and high-throughput omics technologies, both of which are highly relevant for advancing current knowledge in nanotoxicology and nanomaterial risk assessment. The proposed predictive framework addresses the growing availability of transcriptomic data, whose biological interpretation remains a major challenge, by employing computational approaches capable of linking molecular-level responses with the physicochemical characteristics of nanomaterials.

The dissertation demonstrates a high level of scientific productivity and impact, resulting in six articles published in peer-reviewed international journals three of which are the principal publications presented and discussed in the thesis, and three are listed as additional publications. Furthermore, the work has been communicated through oral and poster presentations at five international conferences.

The thesis consists of an introduction, two chapters describing the research objectives and results, and a concluding discussion, for a total of 79 pages which include 211 references. Two additional sections are dedicated to statements of authorship and the collection of published articles. In particular, Chapter 1 reviews the state of the art regarding the toxicology of MWCNTs and metal oxide nanoparticles, motivating the need for novel computational approaches and transcriptomic endpoints to support hazard assessment. Chapter 2 briefly introduces the main research hypotheses and objectives of the study. Chapter 3 summarizes the three case studies and their principal findings. Finally, Chapter 4 shortly presents the overall conclusions of the work.
The three articles presented in Chapter 3 address, respectively: (i) the development of the Transcriptomic Response Index (TRI) using PCA to condense transcriptomic responses, and the creation of nano QSARs to link MWCNTs structural properties to the TRI (Article 1 in Chapter 7); (ii) the extension and optimization of the TRI framework defined for MWCNTs to TiO_2 nanoparticles as a compact predictive omics endpoint for nanosafety assessment (Article 2 in Chapter 7); and (iii) a machine-learning benchmarking framework comparing QSARs for the

prediction of gene expression due to TiO_2-exposure in order to identify transcriptomic signatures possibly associated with inflammation and fibrosis, thereby complementing TRI-based nanosafety assessment (Article 3 in Chapter 7).

A major strength of the dissertation is its highly interdisciplinary nature, requiring expertise across multiple methodological domains, including transcriptomics, multivariate analysis, nano-QSAR modelling, and machine learning, thereby reflecting a broad and solid scientific background. The principal contribution of the thesis is the development of the Transcriptomic Response Index (TRI), an innovative approach designed to condense complex transcriptomic responses into a single endpoint suitable for predictive modelling. The proposed methodology addresses the challenge of incorporating thousands of gene expression measurements into computational models through the derivation, using principal component analysis (PCA), of a compact and informative indicator of the biological perturbation induced by nanomaterial exposure.
An additional strength of the work lies in its effort to establish a bridge between transcriptomic signatures and nano-QSAR modelling. The proposed TRI framework suggests a possible conceptual route for integrating transcriptomic information into nano-QSAR workflows and thus represents a potentially important contribution to the future development of predictive models based on omics data. Moreover, findings reported in Article 3 demonstrate that Machine Learning can be applied to model gene expression changes using simple experimental properties of TiO_2 nanoparticles, while model's quality was used to benchmark genes highly informative for pulmonary inflammation.

Overall, the scientific ambition of the dissertation is considerable, and the work contains several innovative elements. At the same time, the proposed methodology raises some conceptual and methodological questions that merit further discussion.

General observations:
Despite providing a comprehensive overview of the toxicity of MWCNTs and TiO_2 nanoparticles, as well as the growing use of omics-based approaches as alternative methods for hazard assessment, the Introduction offers only a limited discussion of the existing regulatory frameworks. While concepts such as New Approach Methodologies (NAMs), Integrated Approaches to Testing and Assessment (IATAs), and Safe and Sustainable by Design (SSbD) are briefly introduced (pages 21-23), a clear description of the regulatory context in which these approaches are expected to operate in relation to nanomaterials (such as the Safe and Sustainable by Design framework, the REACH Regulation, or OECD documentation) is not provided.
Similarly, the regulatory principles and guidelines governing QSAR model development and validation are only briefly mentioned (page 30), despite representing key requirements for the regulatory acceptance of *in silico* models. Moreover, Reference 112, cited on page 30, does not specifically address the aforementioned OECD regulatory principles, and Reference 131, which would be more appropriate, is not correctly reported in the reference list (i.e., it should be cited

as: "OECD (2024), (Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure–Activity Relationship Models and Predictions - Second Edition, OECD Series on Testing and Assessment No. 405, OECD Publishing, Paris").

A more explicit contextualization of current regulatory requirements and hazard and risk assessment workflows related to the use of NAMs for nanomaterials, as well as their strengths and limitations, would have helped readers better appreciate the practical significance and applicability of the proposed methodologies.

At present, computational methodologies are discussed only briefly in the final section of the Introduction (pages 29-30), through a general overview of methods applicable to data analysis and modelling, with no clear reference to the methods applied in Chapter 3 and in the articles listed in Chapter 7. This section could have been expanded to include a more detailed description of nanoparticle characterization procedures, descriptor calculation strategies, and the theoretical foundations of the multivariate statistical and machine-learning methods employed throughout the work. In addition, the inclusion of schematic workflows summarizing the main stages of the analyses would have greatly facilitated comprehension of the overall research strategy, from data preprocessing and curation, through the unsupervised analyses used for TRI generation, to the subsequent modelling steps, including variable selection, model development, and validation procedures.

Another aspect that is missing is a more extensive discussion of the limitations of the proposed approaches and results. In particular, the thesis would have benefited from a dedicated critical discussion addressing the boundaries, limitations, and practical implementation of the proposed methodologies.

A final remark concerns the lack of cross-referencing between the content discussed in Chapter 3 and the original articles presented in Chapter 7. Not all the data and figures supporting the results discussed in Chapter 3 (e.g., descriptors, scores, loadings, and figures) are included in the text. Readers need to consult the original publications to access relevant supporting information which is not indicated in Chapter3. This affects the readability of the work.

Points of discussion

1. A first aspect that merits further discussion concerns the biological interpretability of the TRI presented in Chapter 3. The proposed index represents an innovative strategy for condensing the overall magnitude of transcriptomic perturbation into a single predictive endpoint, thereby facilitating downstream modelling and nano-QSAR applications. At the same time, although the original PCA loadings remain available and preserve a formal link with the transcriptomic space, the final TRI is derived from an aggregation of variance-weighted principal component scores (equation 1 page 38), making the contribution of individual genes and biological pathways less straightforward to interpret. In addition, components contributing with opposite signs may partially compensate each other, potentially leading to similar TRI values for transcriptomic profiles characterized by different biological features. This issue may be particularly relevant given the large number of genes involved in the analyses, ranging from approximately 600 in the TiO_2 dataset to more than 5,000 in the MWCNT dataset, which makes biological interpretation of the final index challenging in

practice. These considerations do not detract from the value of the TRI as a global measure of transcriptomic perturbation, but they raise questions regarding its use as a mechanistic index.

It would be interesting to discuss the extent to which specific TRI values can be related to biologically meaningful pathways or mechanisms of toxicity considering its mathematical formulation, and how cases should be interpreted in which nanomaterials with distinct transcriptomic profiles converge to similar TRI values.

2. The combination of Genetic Algorithms and Ridge regression represents an interesting methodological choice. Nevertheless, the rationale for combining these approaches should be discussed in greater depth. Ridge regression is generally used to address multicollinearity among predictors, whereas the Genetic Algorithm performs variable selection from a very large pool of candidate descriptors. Given the relatively small number of nanomaterials available compared with the thousands of descriptors initially considered for instance in the MWCNT dataset (about 3500 SiRMS descriptors are listed in Table S4 of Article 1, Chapter 7. Data not reported in Chapter 3), it would be valuable to examine the potential impact of the selection process on model generalizability and the risk of selection bias inherent to high-dimensional, low-sample-size modelling problems. Finally, given the combination of large descriptor spaces, feature-selection procedures, and relatively small numbers of nanomaterials, a discussion of the potential risk of chance correlations would strengthen the methodological assessment.

The candidate could further elaborate on the rationale for preferring a GA-Ridge regression over alternative strategies, such as PLS or GA-MLR, and comment on the impact of the selection process on the risk of selection bias inherent to high-dimensional, low-sample-size modelling problems. It would also be interesting to know whether procedures such as Y-randomization, permutation testing, or related validation approaches were considered to assess the possibility of obtaining comparable model performances by chance.

3. The thesis would also benefit from a broader discussion of the limitations and applicability of the proposed framework. While the TRI is in general presented as a powerful means of integrating transcriptomic information into nano-QSAR modelling, the candidate recognizes (page 49) that TRI "is dependent on both the particular material under study and on the characteristics of the available data" and that "TRI should be re-optimized whenever a new class of nanomaterials, or new datasets, are examined". Therefore, it remains unclear to what extent the approach can be generalized beyond the specific nanomaterials and datasets considered in this work. This issue appears particularly relevant also in light of the different best TRI formulations adopted in Articles 1 and 2 (specifically referring to 99% and 44% of the total variance, respectively). Whereas the MWCNT study employs a TRI capturing almost the entire transcriptomic variance, the TiO_2 study identifies a more compact index explaining a substantially smaller proportion of the variance while yielding improved predictive performance. This result may appear contradictory and raises broader questions regarding

the balance between transcriptomic representation and predictive capability, as well as the criteria used to define an optimal TRI.

It would be interesting to discuss how the TRI might evolve as additional nanomaterials and data are incorporated into the dataset and whether the primary objective of TRI optimization should be the maximization of the retained transcriptomic variance or the predictive performance of the subsequent nano-QSAR model.

Concluding remarks

The presented work addresses an important scientific problem, proposes innovative computational methodologies, and demonstrates the candidate's ability to integrate concepts from transcriptomics, machine learning, statistics, and computational toxicology. The candidate clearly possesses the skills required for independent scientific research and has produced work that contributes to the advancement of nanoinformatics and nanotoxicology.

Despite the methodological questions outlined above, I consider the dissertation to constitute a substantial and original scientific contribution. The candidate has demonstrated significant expertise, scientific maturity, and research independence.

In summary, I conclude that the doctoral dissertation submitted by Mr. Viacheslav Muratov entitled "A properties-to-omics predictive framework for hazard assessment of metal- and carbon-based nanomaterials" meets the requirements set forth in the Act – Law on Higher Education and Science (Article 187, paragraphs 1–3) of 2018, as amended.

Therefore, I submit to the Council of the Discipline of Chemical Sciences of the University of Gdansk a recommendation to allow Mr. Viacheslav Muratov to proceed to the subsequent stages of the procedure for awarding the academic degree of Doctor.

Varese, 17/11/2026

Sincerely,

Prof. Ester Papa

