

REVIEW

by **Prof. Ivanka Milosheva Tsakovska, PhD**
Institute of Biophysics and Biomedical Engineering – BAS

On the PhD for awarding the educational and scientific degree "Doctor"

Area of Higher Education: 4. Natural Sciences, Mathematics, and Informatics

Professional Field: 4.2. Chemical Sciences

Doctoral Program: "Theoretical Chemistry"

Author: Gergana Ilieva Tancheva

Title: APPLICATION OF CHEMOINFORMATICS METHODS TO MULTICOMPONENT SUBSTANCES AND NANOMATERIALS

Scientific Supervisor: Assoc. Prof. Dr. Nikolay Todorov Kochev, Plovdiv University "Paisii Hilendarski"

1. General description of the submitted materials

By Order No. RD-21-1840 dated 23.10.2024 of the Rector of Plovdiv University "Paisii Hilendarski" (PU), I was appointed as a member of the Scientific Jury for the defense procedure of the dissertation titled "*Application of Chemoinformatics Methods to Multicomponent Substances and Nanomaterials*" for the acquisition of the educational and scientific degree "Doctor" in the area of higher education 4. Natural Sciences, Mathematics, and Informatics, professional field 4.2. Chemical Sciences, doctoral program "Theoretical Chemistry." The author is Gergana Tancheva, a part-time doctoral student in the Department of Analytical Chemistry and Computer Chemistry, Faculty of Chemistry, with scientific supervision by Assoc. Prof. Dr. Nikolay Kochev from Plovdiv University.

The set of materials submitted by Gergana Tancheva complies with Article 36 (1) of the Regulations for the Development of the Academic Staff at PU and includes the following documents: 1. Application form to the Rector of PU for opening the dissertation defense procedure; European-format CV; 2. Protocol from the Department council confirming readiness for the defense procedure and a preliminary discussion of the dissertation; 3. PhD thesis manuscript; 4. Abstract; 5. List of scientific publications on the dissertation topic; 6. Copies of three scientific publications; 7. Declaration of originality and authenticity of the submitted documents; 8. Statement of meeting the minimum academic requirements; 9. Opinion by the scientific supervisor.

The review of the submitted materials confirms compliance with all legal requirements for the defense procedure.

2. Brief Biographical data of the doctoral candidate

Gergana Tancheva holds a Bachelor's degree in Medical Chemistry and a Master's degree in Spectrochemical Analysis from the Faculty of Chemistry, Plovdiv University "Paisii

Hilendarski," Department of Analytical Chemistry and Computer Chemistry. In 2019, she was admitted as a part-time doctoral student in the same department and was granted the right to defend her dissertation in 2024. Her dissertation was reviewed and approved for defense by the expanded Department Council at a meeting held on 14.10.2024.

During her doctoral studies, Gergana Tancheva completed numerous training courses, carefully selected to support her high-level dissertation research. Noteworthy is her active participation in international research projects, in frame of which the research for the dissertation was conducted. Since late 2022, she has been employed at IDEAKONSULT Ltd., a key partner in these projects.

3. Relevance of the topic and appropriateness of the objectives and tasks

Chemoinformatics plays a pivotal role in drug design and computational toxicology for modeling the well-defined chemical structures. However, when it comes to substances composed of multiple chemical structures or nanomaterials, chemoinformatics methods are still adapting to properly represent this different structural space and to enable the structured storage of essential metadata for modeling and decision-making, especially in the field of toxicology. In this regard, the dissertation research is highly relevant and forward-looking. Its results have already been applied to several international research projects aimed at supporting regulatory decision-making regarding the safety of nanomaterials.

The goal is clearly formulated, and the research tasks necessary to achieve it are identified. However, I have a critical remark regarding the inclusion of tasks related to studying existing software systems, semantic models, ontologies, etc. In my opinion, these should be part of the literature review, while the research tasks of the dissertation should reflect the specific studies undertaken to achieve the dissertation's goal.

4. Familiarity with the problem

The doctoral candidate demonstrates excellent knowledge of the field of chemoinformatics and a broad range of expertise — from the representation and modeling of chemical compound effects to approaches for solving complex problems in nanoinformatics. This is evident in the literature review section of the PhD thesis.

On the one hand, it is commendable how many approaches and methodologies the candidate has explored. On the other hand, the literature review is somewhat difficult to follow due to the large amount of information beyond the specific objectives and tasks and lacks precision in structuring the content. For example, in section 1.2 *Classification of Chemical Substances*, topics such as *Properties of Nanomaterials* and *Approaches for Nanomaterial Synthesis* are included; the development of Adverse Outcome Pathways (AOPs) is discussed under *Model Validation* topic, further it is presented very briefly and lacks clarity.

The doctoral candidate also extensively discusses the OECD's principles for validation of QSAR models, referring to the widely accepted regulatory document from the past 15 years (note: correction of the cited literary source 180 is needed). However, I recommend including the latest, updated framework document developed by OECD in late 2023, titled “(Q)SAR Assessment

Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure-Activity Relationship Models and Predictions” [<https://doi.org/10.1787/d96118f6-en>].

5. Research methodology

The research methodology is appropriate and enables the achievement of the PhD thesis's objective as well as addressing the specific research tasks.

6. Characterization and evaluation of the PhD thesis

The PhD thesis is structured into 196 pages, including 70 pages of literature review, objectives, and tasks; 78 pages of original research and summarized discussion; conclusions and contributions. The bibliography consists of 246 references, providing an excellent overview of the research field. Additionally, four appendices are included, contributing to the full transparency and openness of the data.

In the doctoral student's original research, the Ambit/eNanoMapper data model was selected for describing data of multicomponent substances and nanomaterials. FAIR descriptions of multicomponent substances and nanomaterials are presented through three main metadata layers: (i) substance identification, (ii) composition description, and (iii) measurement records with rich metadata. Based on this foundation, improvements were made to the computational infrastructure for managing toxicological data on nanomaterials within eNanoMapper. For instance, the NMDataParser tool was developed to facilitate data preparation and uploading to the eNanoMapper database. Noteworthy is the work on processing a large number of Excel files containing experimental data collected within various international scientific projects for data FAIRification. The converted data enriched the eNanoMapper database with a wide variety of physicochemical and biological analyses data.

A web-based software tool, the eNanoMapper Template Wizard, was also developed to harmonize data descriptions with the necessary metadata, which is highly beneficial for specialists when harmonization and standardization of their experimental data is needed. Additionally, the eNanoMapper ontology was enriched by the introduction of 15 new scientific terms as part of the dissertation research.

A significant contribution of the research is the development of identifiers for the structural representation of multicomponent substances, extending in this way the approaches used for chemical compound identifiers. In this regard, the Ambit-SLN open-source software library was developed to represent chemical substances using the SYBYL Line Notation. This library has been integrated into the Ambit open-source software platform.

To address the need for processing experimental data related to nanomaterials' characterization and safety, electronic notebooks were developed for data extraction, processing, and visualization from the eNanoMapper database, as well as for verifying data completeness according to the harmonized templates. Notable results were achieved in automating data processing for high-throughput screening (HTS). Specifically, the ToxFairy software library was developed for reading and annotating HTS data, performing preliminary processing,

FAIRification, and toxicity assessment. Additionally, a user interface for the Orange platform was developed for this library. The library was successfully used for automated data processing and assessment of data within two international scientific projects.

7. Contributions and Significance of the Research for Science and Practice

The dissertation results have already been applied to structuring and processing information on nanomaterials within several European projects, demonstrating a clear contribution to extending cheminformatics approaches for multicomponent chemical systems, particularly nanomaterials. Numerous software components were implemented, primarily within the eNanoMapper project.

I accept the doctoral student's stated contributions as significant, with both scientific and applied value. The scientific contributions include:

- (i) developing the NMDDataParser tool for preparing and uploading data into the eNanoMapper database in line with the FAIRification concept for chemical substances and nanomaterials;
- (ii) developing a prototype identifier for nanomaterials based on SLN linear notation that encodes both structural and experimental data for specific nanomaterials;
- (iii) creating a concept for annotating high-throughput screening data, including metadata, preprocessing, and calculating a toxicological prioritization index.

The applied contributions include:

- (i) extending the eNanoMapper database with FAIR data on physicochemical and biological parameters of nanomaterials studied within European projects with the active participation of the research group to which the doctoral candidate belongs;
- (ii) enriching the eNanoMapper ontology in the fields of ecotoxicology and environmental protection;
- (iii) developing the ToxFAIRy software library for annotating, processing, toxicological prioritization, and FAIRification of high-throughput screening data;
- (iv) implementation of the Orange3-ToxFAIR module for the Orange platform, providing a user interface for the ToxFAIRy library;
- (v) developing an automated workflow for processing, clustering, and FAIRification of high-throughput screening data, enabling efficient simultaneous handling of large data volumes.

The doctoral student has proposed ambitious future directions logically based on the achievements in representing structural and experimental information for multicomponent substances that are presented in the PhD thesis and further aiming at modeling the properties/effects of chemical substances/nanomaterials.

8. Assessment of publications related to the PhD thesis

Three scientific papers have been published in relation to the PhD thesis — two in impact factor-journals within Q1 and Q2 categories, and one book chapter. The publications have been cited 17 times, that is a demonstration of good recognition within the scientific community. The publications' content is closely related to the dissertation research. The results have been actively disseminated through numerous scientific forums and applied in several international scientific projects under the Horizon 2020 program. Based on this, I conclude that the minimum national

requirements for obtaining the educational and scientific degree of "Doctor" have been convincingly met.

9. Personal Contribution of the Doctoral Student

Although the research is entirely team-based, in line with modern scientific trends, I have no doubt that the PhD thesis is the doctoral student's personal work and that she has actively participated in the reported research. This is supported by the entirely positive opinion of her academic advisor.

10. Abstract

The synopsis successfully conveys the essential achievements of the comprehensive PhD thesis research and accurately reflects the main results achieved in the PhD thesis.

11. Critical Remarks, Recommendations, and Questions

Few critical remarks and recommendations are presented in the earlier sections of the review.

Questions to the doctoral candidate:

1. How are the concepts of multicomponent substance and chemical mixture distinguished? For instance, why is Composition 2 in Figure 17, consisting of two main components (benzene and toluene), classified as a substance rather than a chemical mixture?
2. Could the doctoral candidate discuss possibilities for modeling the toxicity of multicomponent chemical substances, considering the dissertation's statement that their induced toxicity (referring to materials with new or improved functionalities) "may not be accurately assessed based on the measured toxicity induced by the chemistry of their individual components, but on the unique combination of all components"?
3. In the "Chemical Databases" section of the literature review, examples are primarily given for chemical compound databases, with further discussion on the eNanoMapper database. Are there other such structured databases for chemical substances/nanomaterials?

12. Personal Impressions

I have no personal impressions of the doctoral candidate. However, I am well-acquainted with the work of her research team, which is among the leading scientific groups in applying cheminformatics for regulatory toxicology, pioneering the field of nanoinformatics in several European projects for nanomaterial safety assessment. Undoubtedly, this environment has provided an excellent foundation for her further development as an independent, well-established researcher.

CONCLUSION

The PhD thesis contains scientific and applied results representing an original contribution to science and meets all requirements of the Law on the Development of Academic Staff in the Republic of Bulgaria, its Regulations, and the corresponding regulations of Plovdiv University "Paisii Hilendarski".

The PhD thesis demonstrates that the doctoral candidate possesses in-depth theoretical knowledge and professional skills in cheminformatics and exhibits the qualities and abilities necessary for conducting independent scientific research.

Based on the above, I confidently provide a positive evaluation of the research presented in the reviewed PhD thesis, abstract, publications, results, and contributions and propose to the esteemed Academic Jury to positively vote for awarding the educational and scientific degree of "Doctor" to Gergana Ilieva Tancheva in the Area of Higher Education: 4. Natural Sciences, Mathematics, and Informatics, Professional Field 4.2. Chemical Sciences, Doctoral Program "Theoretical Chemistry."

18.12.2024

Reviewer:.....
/Prof. Ivanka Tsakovska/