МЕДИЦИНСКИ УНИВЕРСИТЕТ - СОФИЯ
ФАРМАЦЕВТИЧЕН ФАКУЛТЕТ
1000 СОФИЯ, ул. Дунав № 2
www.pharmfac.mu-sofia.bg

R E V I E W

by **Prof. Irini Doytchinova**

Faculty of Pharmacy, Medical University of Sofia

RE: PhD thesis of **Gergana Ilieva Tancheva**, PhD student in the Faculty of Chemistry, Plovdiv University, on theme: Application of Cheminformatics Methods for Multicomponent Substances and Nanomaterials, submitted for awarding a PhD degree in Higher education area: 4. Natural Sciences, Mathematics and Informatics, Professional direction: 4.2. Chemical Sciences, Doctoral program: Theoretical Chemistry, Supervisor: Assoc. Prof. Nikolay Kochev

The PhD thesis titled "Application of Cheminformatics Methods for Multicomponent Substances and Nanomaterials" explores a highly relevant and innovative field focused on utilizing cheminformatics methods for processing, analyzing, and modeling complex chemical entities. The topic is multifaceted, encompassing multicomponent substances, nanomaterials, and materials with enhanced functionalities (AdMa), highlighting the interdisciplinary nature of the research. The application of FAIR principles for data management (Findable, Accessible, Interoperable, Reusable) is particularly relevant in the context of international efforts to standardize and efficiently manage scientific information.

The PhD thesis is written on 196 pages and includes the following sections: Literature review, Aim and tasks, Original research, Conclusions, Scientific reports, resources and educational activities related to the PhD thesis, References, and Appendices. It is illustrated with 6 tables and 65 figures, and the References comprise 246 titles.

The Literature review covers definitions, classifications and regulations related to multicomponent substances and nanomaterials. It provides a detailed description of methods for representing, storing, and processing chemical data, as well as a comprehensive overview of quantitative structure-activity relationship (QSAR) modeling. The review is well-structured and gives conclusions leading to the aim of the PhD thesis.

The aim of the PhD study is to apply a semantic model for processing and storing information about multicomponent substances, nanomaterials, and advanced materials, based on FAIR principles. Sixteen tasks have been defined to achieve this aim, which could potentially be consolidated into four or five broader tasks.

In the "Original Research" section, Gergana applies a semantic model for representing chemical substances and nanomaterials, which imply the FAIR, CARE, and TRUST principles. The successful integration of this model into the eNanoMapper database demonstrates its applicability and significance. I consider this model a substantial contribution to the standardization of chemical data. Although I cannot assess Gergana's technical contribution, the applicability of the developed model, demonstrated through specific examples and collaborative research in international scientific projects, strengthens my belief in its practical value.

A method for data serialization of multicomponent substances is proposed, utilizing formats such as JSON, RDF, ISA, and NeXus, which facilitate integration and analysis. An open-source software library, Ambit-SLN, was developed for processing information on chemical substances using SYBYL Line Notation (SLN). This library provides a rich syntax for describing chemical structures, supporting macro and Markush atoms, search queries, and combinatorial libraries. Ambit-SLN is integrated with popular formats (SMILES, SMARTS), making it versatile and suitable for academia, industry, and regulatory agencies.

Electronic notebooks were developed to automate the processing, filtering, and visualization of data for nanomaterial risk assessment using the eNanoMapper database. The first notebook retrieves data through REST API queries, allowing customized filtering by substances, experimental protocols, and parameters (e.g., concentration, treatment time) and visualizing the results in interactive 3D graphics. The second notebook evaluates data completeness by comparing terminology from templates and the database, identifying missing or incorrect terms using a similarity algorithm based on Levenshtein distance. These tools facilitate the structuring and analysis of nanomaterial safety data.

A methodology for processing high-throughput screening (HTS) data on the toxicity of chemicals and nanomaterials is proposed, based on the principles of New Approach Methodologies (NAM) and the 3Rs (Replacement, Reduction, and Refinement). HTS uses *in vitro* analyses to assess toxicity through a combination of experiments involving various time points, concentrations, and cell lines. The traditional approach of measuring GI50 has been replaced with the Tox5-Score, which integrates multiple parameters into a single toxicological prioritization index, visualized using pie charts.

In the brief discussion, Gergana critically analyzes and summarizes the various stages of the research, highlighting the importance of applying FAIR principles to data processing for complex chemical substances and nanomaterials. She demonstrates the advantages of the developed software applications, electronic notebooks, and add-ons that facilitate data processing, serialization, and modeling of toxicological and other chemical data. Potential limitations of the proposed approaches and methods and suggestions for overcoming them should be commented.

Two groups of contributions are defined: scientific and applied scientific. The scientific contributions include the application of the Ambit/eNanoMapper semantic model for FAIR-ification of data, the development of a prototype nanomaterial identifier based on SLN notation and the integration of high-throughput screening data with metadata, including the Tox5-Score toxicological prioritization index. The applied scientific contributions consist

of the development of specialized software tools for processing, serialization, and modeling of toxicological and chemical data. I fully accept these contributions as defined.

The reports related to the PhD thesis include two publications in impact factor journals, one in Q1 (25 points) and one in Q2 (20 points), as well as one book chapter (15 points), totaling 60 points. This exceeds the required 30 points as per the regulations of Plovdiv University and the additional requirements for scientific and teaching activities of candidates for academic degrees and positions in the Faculty of Chemistry in the professional fields 4.2. Chemical Sciences and 1.3. Pedagogy of Chemistry Education (https://procedures.uni-plovdiv.bg/docs/ additional/hf_r.pdf). However, I am concerned that Gergana is neither first or second author in these publications, which would have emphasized her personal contribution to their development. One of the publications has 16 citations. The results of the PhD thesis have been presented at seven international and five national scientific forums and funded by five international projects.

Gergana’s active educational activity during her PhD studies is noteworthy. She successfully completed six programming courses. I believe that this education, combined with her ability to apply the acquired knowledge in solving specific scientific problems, establishes Gergana as a skilled professional in bioinformatics.

Finally, I give a positive evaluation of Gergana Ilieva Tancheva's PhD thesis and as a member of the scientific jury, I confidently vote for awarding her the educational and scientific degree of "Doctor" in the field of higher education 4. Natural Sciences, Mathematics, and Informatics, professional field 4.2. Chemical Sciences, doctoral program in Theoretical Chemistry.

4.12.2024 г. Reviewer:

Sofia

(Prof. Irini Doytchinova)