NAVIGATING COMPLEX BIOLOGICAL SYSTEMS WITH PHYSIOLOGY-BASED PHARMACOKINETICS, TEXT MINING AND ARTIFICIAL INTELLIGENCE: IN SILICO NEW APPROACH METHODOLOGIES FOR THE DEVELOPMENT OF RELIABLE AND ROBUST QUANTITATIVE ADVERSE OUTCOME PATHWAYS

A. Karakoltizids[1,2], E. Renieri[1,2], N. Papaioannou[1,2], I. S. Frydas[1,2], T. Papageorgiou[1,2], D. R. Schultz[1,2], C. Gabriel[1,2], N. Georgiou[1,2], S. Karakitsios[1,2], D. A. Sarigiannis[1,2,3,4]*

[1]HERACLES Research Center on the Exposome and Health, Center for Interdisciplinary Research and Innovation, Aristotle University of Thessaloniki, Thessaloniki, Greece

[2]Environmental Engineering Laboratory, Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

[3]University School of Advanced Study IUSS, Pavia, Italy

[4]National Hellenic Research Foundation, Athens, Greece

*(*sarigiannis@auth.gr)*

ABSTRACT

In the context of this study, a computational methodology is introduced which starts from environmental exposure and concludes by quantitatively linking it to a disease by constructing qAOPs. Numerous tools and methodologies are developed encompassing NLP, AI, exposure, PBTK, and systems biology models. Commencing with exposure scenarios, the exposure is quantified across the three routes. Pharmacokinetic models are employed to estimate internal exposure. To delve into the molecular level, in vitro and vivo experiments to generate omics data were conducted. The data were integrated to generate multi-omics, aiding in the identification of metabolic pathways perturbed following exposure. This data was input into MINER, a text mining toolbox developed to convert metabolic pathways into mechanistic models. The incorporation of as many biological pathways as possible resulted in the construction of a big systems biology model, consisting of 1348 differential equations. The kinetics of the model were parameterized with data originating from the BRENDA and SABIORK databases as well as with the use of Deep Learning models to estimate enzyme properties of interest. Data derived from the HMDB database were utilized to initialize the model. In addition, a methodology was developed using ML and GANs to initialize the model. An ML model was generated for each endogenous metabolite, irrespective of whether its concentration was known. This was conducted for validation purposes. Subsequently, the model was executed twice. The second run also incorporated the fold change results obtained from the omics. The simulation output unveiled metabolites whose concentrations changed by an order of magnitude. Owing to this circumstance, publicly available data were collected to scrutinize the variations in concentrations when an individual is impacted by a disease. The mathematical equation describing the endogenous metabolite provides a comprehensive network of interactions involving metabolites, genes, and proteins. For the AOP development, a bottom-up approach was employed, leveraging our knowledge of the AO and the available information about the MIE. By utilizing transcriptomics, conducting network analysis, and performing a literature review, the precise MIE was determined. Utilizing the developed NLP tools, the AOP was established. Additionally, experimental studies were identified for the quantitative description of KERs.

KEYWORDS: NLP, Topic Modelling, Liver Fibrosis, Liver Diseases, Exposure

REFERENCES