Database generation workflow for supporting SSbD risk assessments

José Luis Vallés-Pardo[1], Salvador Moncho[1], Eva Serrano-Candelas[1], Rafael Gozalbes[2]

1. Introduction

The new paradigm of green toxicology aims to make chemicals safer by design, focussing on hazard issues earlier during development stages and keeping track on them during their whole life cycle. This evolves in the development of Safety and Sustainable by Design (SSbD) frameworks [1], defining how to perform the risk assessment, from the raw materials until the final disposal and degradation in the environment. The risk assessment on the SSbD framework requires a big amount of information about physical and (eco)toxicological properties of substances used or released along the production, processing and application of the product. However, the green toxicology paradigm includes the reduction of experimental tests [2]. Under these premises we have started different European projects, such as BIO-SUSHY[3], SiToLub[4] and CheMatSustain[5], including data-based approaches for SSbD assessment. For this reason, an automatic workflow for the creation of specific databases for each project, combining experimental data together with QSAR predictions, will become an interesting tool to help in the SSbD hazard statement.

2. Database structure

For the generation of these databases, existing experimental data and QSAR predictions will be used. The automatised workflow to build the databases extracts information from three different kinds of sources (vendor-provided data, public databases and QSAR predictions), as shown in Figure 1.

Vendor's information will be gathered using their safety data sheets (SDSs). Automatic text scrapping will be used to gather the information from the PDF documents and organise it into tabulated data.

Data from a series of reliable public databases will be also extracted, including databases coming from regulatory entities (such as ECHA and EPA). Our databases will include not only the parameters/endpoints values, but also other information about the experimental records -as the sources and experimental protocols- to facilitate the selection among multiple records and to evaluate the reliability and pertinence of the different values. A special treatment will be applied for the mixtures, including data about both, the materials of interest and their constituents. In addition to experimental records for specific properties, our databases will also gather previous safety

[1] ProtoQSAR S.L., CEEI Valencia, Avda. Benjamin Franklin 12, 46980 Paterna (Spain); jlvalles@protoqsar.com
[2] MolDrug AI Systems S.L., c/Olimpia Arozena Torres, 45, 46108 Valencia (Spain)

assessments, such as CLP harmonized categories and data from the ECHA list of substances of very high concern.

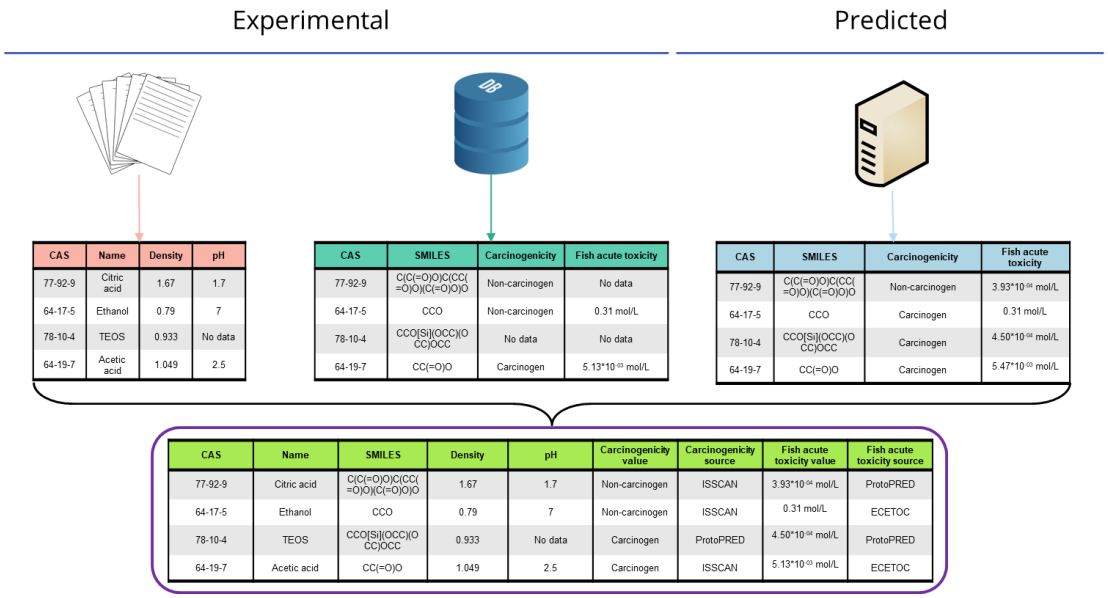

Figure 1. Databases sources.

Predicted data will be obtained from QSAR models. Several prediction tools will be explored to use the most appropriate one, –including our own in-house models available at the ProtoPRED® platform [6]–, keeping track of the selected tool, together with the metrics of the model and other relevant information.

Once the tables with experimental and predicted data will be compiled, they will be combined into a single one for each project, using tailored selection rules to select the information when more than one record exists.

3. Conclusions and future steps

Following the aforementioned protocol, comprehensive databases will be obtained without performing any new experimental test, and therefore without the costs and the ethical considerations when *in vivo* tests are involved.

To broaden the range of this approach, several steps will be taken in the future:

- The number of public databases and QSAR models included will be expanded.
- An automatic process to transform the data obtained into regulatory-based categories used in SSbD risk assessment will be developed.
- Additionally, the remaining gaps will be filled up with *ad hoc* tests.

4. Acknowledgements

The BIO-SUSHY project is funded by the European Union under the Grant Agreement number 101091464. The SiToLub project is funded by the European Union under the Grant Agreement number 101138807. The CheMatSustain project is funded by the European Union under the Grant Agreement number 101137990. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of

the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

5. References

[1] European Commission, Joint Research Centre, Caldeira, C., Farcal, L., Garmendia Aguirre, I. et al., *Safe and sustainable by design chemicals and materials – Framework for the definition of criteria and evaluation procedure for chemicals and materials*, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2760/487955

[2] Maertens, A., Luechtefeld, T., Knight, J. and Hartung, T. (2024) "Alternative methods go green! Green toxicology as a sustainable approach for assessing chemical safety and designing safer chemicals", *ALTEX - Alternatives to animal experimentation*, 41(1), pp. 3–19. doi: 10.14573/altex.2312291.

[3] https://www.bio-sushy.eu/

[4] https://sitolub.eu/

[5] https://www.chematsustain.eu/

[6] https://protopred.protoqsar.com/