EnzymeML – a data exchange format for biocatalysis and enzymology

Jan Range[1,2], Jürgen Pleiss[2]

[1] Cluster of Excellence "Data-Integrated Simulation Science (SimTech)", University of Stuttgart, Pfaffenwaldring 5a, 70569 Stuttgart, Germany

[2] Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

The design of biocatalytic reaction systems can be quite intricate, as numerous factors can impact the estimated kinetic parameters, including the enzyme itself, reaction conditions, and the chosen modelling method. This complexity can make reproducing enzymatic experiments and reusing enzymatic data challenging. EnzymeML[1] was created as an XML-based markup language to address this issue, enabling the storage and exchange of enzymatic data, including reaction conditions, substrate and product time course, kinetic parameters, and kinetic models. This approach makes enzymatic data accessible, findable, interoperable, and reusable (FAIR). Furthermore, the EnzymeML team and community have developed a toolbox that helps researchers report on experiments, perform modelling tasks, and upload EnzymeML documents to data repositories. The usefulness and feasibility of the EnzymeML toolbox have been demonstrated in various scenarios by collecting and analysing the data and metadata of different enzymatic reactions[2]. EnzymeML is a seamless communication channel that connects experimental platforms, electronic lab notebooks, tools for modelling enzyme kinetics, publication platforms, and enzymatic reaction databases. EnzymeML is an open standard and available via *https://github.com/EnzymeML* and *http://enzymeml.org*.

References

1. Range, J. *et al.* (2021). EnzymeML—a data exchange format for biocatalysis and enzymology. *The FEBS Journal*, 289(19), pp.5864–5874. doi: 10.1111/febs.16318.

2. Lauterbach, S. *et al.* (2023). EnzymeML: seamless data flow and modeling of enzymatic data. *Nature Methods*, 20(3), pp.400–402. doi: 10.1038/s41592-022-01763-1.