BIOTRANS
LA ROCHELLE 2023

N°1180 / PC
TOPIC(s) : Artifical intelligence / computational methods / Biocatalytic cascade reactions

Digitalization of biocatalysis: the data exchange format EnzymeML

AUTHORS
Jürgen PLEISS / UNIVERSITY OF STUTTGART, INSTITUTE OF BIOCHEMISTRY AND TECHNICAL BIOCHEMISTRY, ALLMANDRING 31, STUTTGART

PURPOSE OF THE ABSTRACT
The design of biocatalytic reaction systems is highly complex due to the dependency of the estimated kinetic parameters on the enzyme, the reaction conditions, and the modelling method. Consequently, reproducing enzymatic experiments and reusing enzymatic data are challenging.
To enable storage, retrieval, and exchange of enzymatic data such as the reaction conditions, the measured time courses of substrate and product, the selected kinetic model, and the estimated kinetic parameters, the XML–based markup language EnzymeML has been developed [1]. EnzymeML is based on SBML and the community Standards for Reporting Enzyme Data (STRENDA). The EnzymeML toolbox supports biocatalysis research by making enzymatic data findable, accessible, interoperable, and reusable [2].
An EnzymeML document contains information about reaction conditions and the measured time course of substrate or product concentrations (Fig. 1). It is generated from an EnzymeML spreadsheet or by the webtool BioCatHub. Kinetic modelling is performed by uploading EnzymeML documents to the modelling platforms COPASI or PySCeS. The estimated kinetic parameters are then added to the EnzymeML document. The EnzymeML document containing the experimental and modelling results is then uploaded to a Dataverse installation or to the reaction kinetics database SABIO-RK. The workflow of a project is encoded as Jupyter Notebook, which can be re-used, modified, or extended.
EnzymeML serves as a seamless communication channel between experimental platforms, electronic lab notebooks, tools for modelling of enzyme kinetics, publication platforms, and enzymatic reaction databases (Fig. 2). The feasibility and usefulness of the EnzymeML toolbox was demonstrated in six scenarios, where data and metadata of different enzymatic reactions are collected, analysed, and uploaded to public data repositories for future re-use [3].
Because EnzymeML documents are structured and standardized, the experimental results encoded in an EnzymeML document are interoperable and reusable by other groups. Because an EnzymeML document is machine-readable, it can be used in an automated workflow for storage, visualization, data analysis, and re-analysis of previously published data, without limitations of the size of each dataset or the number of experiments. Thus, EnzymeML contributes to the digitalization of (bio)catalytic sciences.

FIGURES

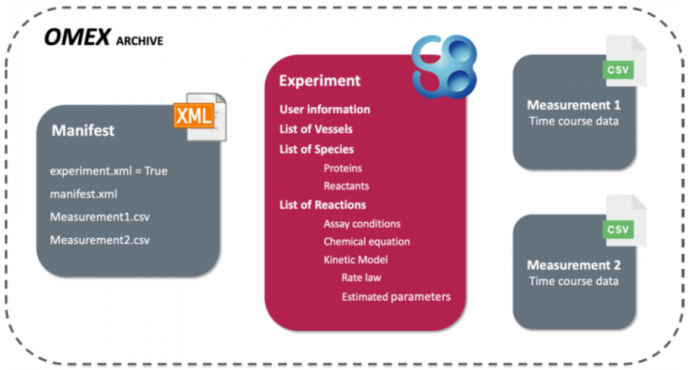

FIGURE 1

Structure of an EnzymeML document.

An EnzymeML document is a ZIP container in OMEX format and contains the experiment file (SBML) and the measurement files (CSV).

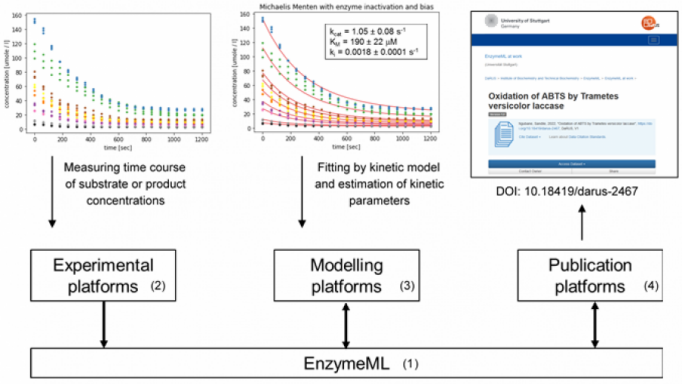

FIGURE 2

Seamless data transfer

An EnzymeML document integrates (meta-)data from experiment and modelling), and serves as a communication channel between experimental, modelling, and publication platforms.

KEYWORDS

research data management | FAIR data principles | enzyme kinetics | reproducibility

BIBLIOGRAPHY

1. Pleiss J, 2021. Standardized data, scalable documentation, sustainable storage - EnzymeML as a basis for FAIR data management in biocatalysis. ChemCatChem 13: 3909-3913; doi: 10.1002/cctc.202100822

2. Range J, Halupczok C, Lohmann J, Swainston N, Kettner C, Bergmann FT, Weidemann A, Wittig U, Schnell S, Pleiss J, 2022. EnzymeML - a data exchange format for biocatalysis and enzymology. FEBS J 289: 5864-5874; doi: 10.1111/febs.16318

3. Lauterbach S, Dienhart H, Range J, Malzacher S, Sp□ring JD, Rother D, Pinto MF, Martins P, Lagerman CE, Bommarius AS, Vang H□st A, Woodley JM, Ngubane S, Kudanga T, Bergmann FT, Rohwer JM, Iglezakis D, Weidemann A, Wittig U, Kettner C, Swainston N, Schnell S, Pleiss J, 2023. EnzymeML: seamless data flow and modeling of enzymatic data. Nat Methods; doi: 10.1038/s41592-022-01763-1