1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.324

Published: 07 Sept. 2023

Ontology-Based Laboratory Data Acquisition with EnzymeML for Process Simulation of Biocatalytic Reactors

Alexander S. Behr[1][https://orcid.org/0000-0003-4620-8248], Elnaz Abbaspour[1][https://orcid.org/0009-0002-3335-6428], Katrin Rosenthal[2][https://orcid.org/0000-0002-6176-6224], Jürgen Pleiss[3][https://orcid.org/0000-0003-1045-8202], and Norbert Kockmann[1][https://orcid.org/0000-0002-8852-3812]

[1] Dept. of Biochemical and Chemical Engineering, TU Dortmund University, Germany

[2] School of Science, Constructor University, Bremen, Germany

[3] Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Germany

Abstract. The presented work explores the use of ontologies and standardized enzymatic data to set up enzymatic reactions in process simulators, such as DWSIM. Setting up an automated workflow to start a process simulation based on enzymatic data obtained from the laboratory can help save costs and time during the development phase. Standardized conditions are crucial for accurate comparison and analysis of enzymatic data, where ontologies provide a standardized vocabulary and semantic relations between relevant concepts. To ensure standardized data, an electronic lab notebook (ELN) is used based on EnzymeML, an open standard XML-based format for enzyme kinetics data. Furthermore, two ontologies are merged and the result is extended for the use in the Python-based workflow. The resulting data is stored in a knowledge graph for research data in a machine-accessible and human-readable format. Thus, the study demonstrates a workflow that allows for the direct translation of ELN data into a process simulation via ontologies.

Keywords: Electronic Lab Notebooks, Enzymatic Catalysis, Knowledge Graph, Process Simulation

1. Introduction

The industrial production of biocatalytic processes has significant potential [1]. However, the development of new bioprocesses is a challenging and highly specific task about reaction conditions. Therefore, there is a high demand for tools, such as process simulators, that can help save costs and time during the process development phase [2]. For instance, the open-source process simulator DWSIM [3] enables the calculation of process streams prior to the establishment of the process in a laboratory plant.

Standardized conditions are crucial for comparing and operating on enzymatic data in bioreactors set up with enzymatic reactions in process simulators. Important parameters include enzyme activities and reaction kinetics, which can vary depending on specific reaction conditions. Furthermore, laboratory technicians who record enzyme-specific data may not be the same individuals who execute process simulations. To address these challenges, ontologies can be used as they provide a standardized vocabulary and semantic relations between concepts relevant to the research domains, enabling accurate comparison and analysis of enzymatic data [4, 5].

Furthermore, electronic laboratory notebooks (ELNs) help laboratory experimenters to record laboratory data and generating a machine-readable data collection while mitigating data loss. Thus, they enable clean research data management in laboratories. As ELNs exist in multiple shapes and utilize different formats, the focus in this work lies on the use of data stored in EnzymeML-based ELNs in form of pre-structured Excel-sheets [6, 7]. EnzymeML is not only an open standard XML-based format for enzyme kinetics data, but also uses ontology classes, e.g., from the Systems Biology Ontology (SBO) [8].

This work shows an automated approach to translate data contained in ELNs into a process simulation by standardized concepts stated in ontologies. Reading EnzymeML-based Excel-sheets with Python, data is extracted and stored in an ontology-based knowledge graph. Furthermore, data regarding the process flow sheet and additional data needed to setup the process simulation are included. Then, DWSIM and its Python interface are used to import the needed data for the automated setup of a process simulation of a reactor scale-up. After the process simulation is conducted, resulting data is also stored in the knowledge graph, allowing for automated storage of research data in a machine and human readable way. This overall workflow, allowing for direct translation of ELN data into a process simulation via ontologies, is depicted schematically in Figure 1.

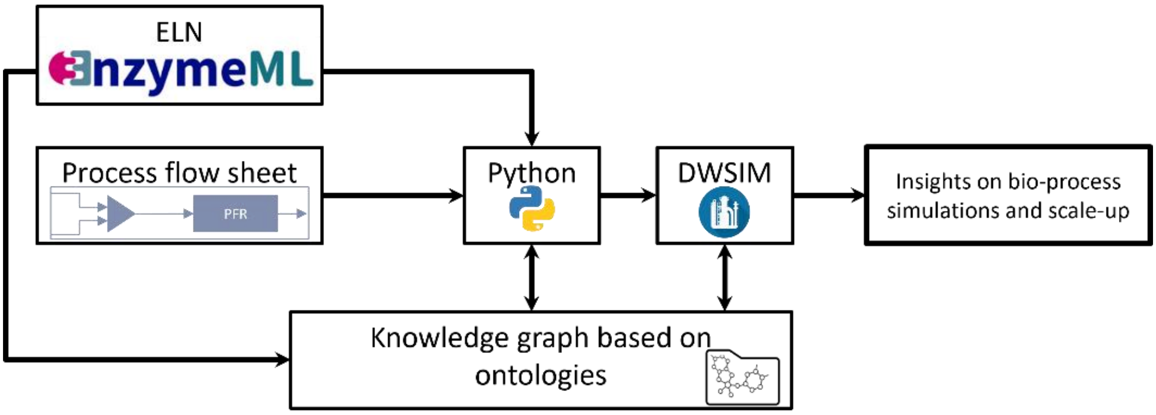

ELN
EnzymeML
Process flow sheet
PFR
Python
DWSIM
Insights on bio-process simulations and scale-up
Knowledge graph based on ontologies

Figure 1. Schematic overview of the workflow presented in this work for the automated execution of process simulations based on a knowledge graph and ELN-based data.

2. Methods

To test the workflow, a set of ELN files describing experiments of oxidation of ABTS with the enzyme Laccase is used. The data of the experiments conducted in the lab in a continuous milliliter-reactor contain among others information of the reaction kinetics. Once, the EnzymeML-based ELN is filled in with the information of the laboratory experiments, the Python-package PyEnzyme [8] allows for automated data extraction from the ELN files.

In order to setup a knowledge graph for the experiments and the simulations in DWSIM, an ontology is needed. As the ELN is setup with classes from the SBO, it is used as base ontology. In addition, the metadata4ing ontology [9] is used to describe process-related concepts. Thus, classes from metadata4ing are included into the SBO and own classes and relations added where necessary to obtain an extended ontology tailored to the needs of this workflow.

Utilizing the owlready2 [10] module in Python, the ontology can then be loaded and extended automatically with the data from the ELN files. Figure 2 depicts an excerpt of the resulting knowledge graph, where individuals are assigned automatically to the ontology based on the input ELN data. This allows for structured access on, e.g., reaction kinetic parameters.

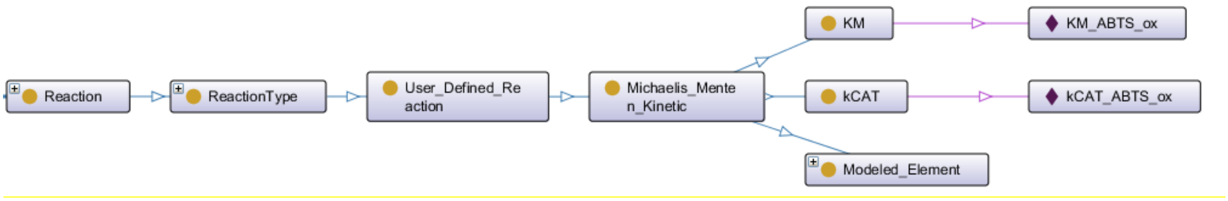
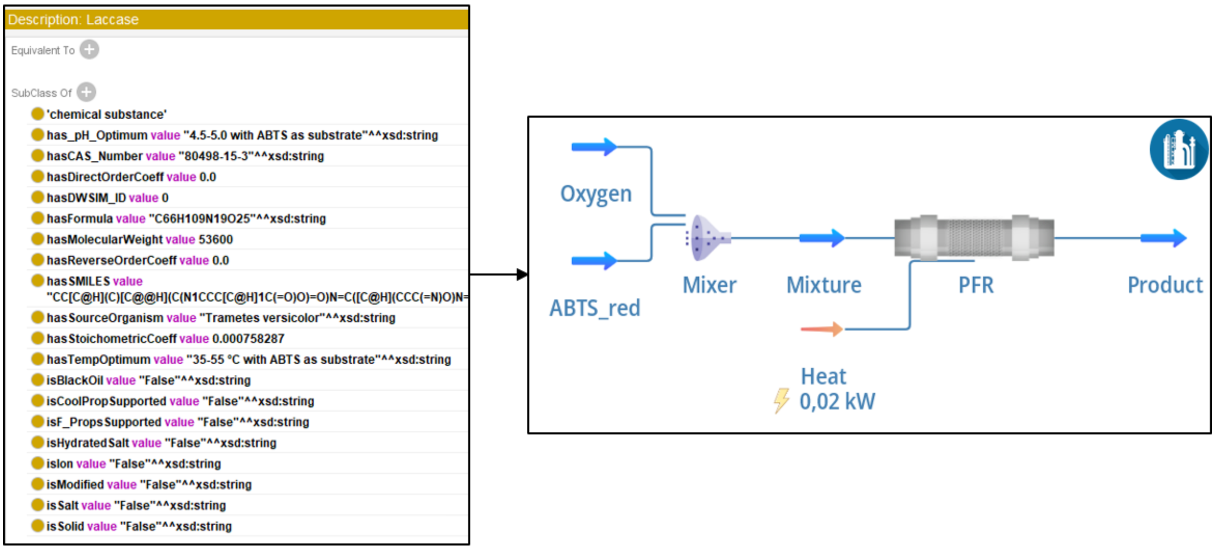


Figure 2. Excerpt of the ontology visualized in Protégé describing the class hierarchy (yellow circles) leading to the individuals for the kinetic parameters K_M and k_{Cat} of the Michaelis-Menten kinetics with regards to the substance $ABTS_{ox}$

Furthermore, the process simulator DWSIM is equipped with a Python-API allowing for automated setup of flow-sheets and execution of process simulations. Thus, after the knowledge graph is generated and stored, the information contained is transferred via the API, creating and executing a new process simulation.

3. Results

Executing the workflow described in the previous chapters results in a knowledge graph containing not only the ontology classes of the extended ontology, but also data obtained from the ELN. Figure 3 shows an excerpt of the class Laccase visualized in Protégé with the data annotations used to setup the corresponding substance in DWSIM. Furthermore, the resulting process simulation is visualized. Thus, this workflow allows for a quick and automated setup of process simulations based on laboratory data previously recorded in an EnzymeML-based ELN.


Description: Laccase

Equivalent To

SubClass Of

- 'chemical substance'
- has_pH_Optimum value "4.5-5.0 with ABTS as substrate"^^xsd:string
- hasCAS_Number value "80498-15-3"^^xsd:string
- hasDirectOrderCoeff value 0.0
- hasDWSIM_ID value 0
- hasFormula value "C66H109N19O25"^^xsd:string
- hasMolecularWeight value 53600
- hasReverseOrderCoeff value 0.0
- hasSMILES value "CC[C@H](C)[C@@H](C(N1CCC[C@H]1C(=O)O)=O)N=C([C@H](CCC(=N)O)N=
- hasSourceOrganism value "Trametes versicolor"^^xsd:string
- hasStoichometricCoeff value 0.000758287
- hasTempOptimum value "35-55 °C with ABTS as substrate"^^xsd:string
- isBlackOil value "False"^^xsd:string
- isCoolPropSupported value "False"^^xsd:string
- isF_PropsSupported value "False"^^xsd:string
- isHydratedSalt value "False"^^xsd:string
- isIon value "False"^^xsd:string
- isModified value "False"^^xsd:string
- isSalt value "False"^^xsd:string
- isSolid value "False"^^xsd:string

Figure 3. Excerpt of the annotation of the Enzyme Laccase within the knowledge graph visualized with Protégé (left) and resulting process simulation of a plug flow reactor (PFR) in DWSIM (right).

Data availability statement

The data, code and markdown files presented in this abstract will be available at GitHub here: https://github.com/TUDoAD/EnzymeMLandDWSIM

Author contributions

Conceptualization: A.S.B.; Methodology: A.S.B., E.A.; Software: E.A., A.S.B.; Validation: E.A., J.P., K.R.; Data Curation: K.R.; Writing – Original Draft: A.S.B., Writing – Review & Editing: N.K., J.P., K.R.; Visualization: A.S.B., E.A.; Supervision: A.S.B., N.K.

Competing interests

The authors declare that they have no competing interests.

Funding

The Deutsche Forschungsgemeinschaft (DFG) is acknowledged for funding this research as part of the Nationale Forschungsdateninfrastruktur (NFDI) initiative (grant No.: NFDI/2-1 - 2021).

Acknowledgement

The authors thank Julia Suhrkamp for the supervision of the laboratory experiments enabling the data recording with EnzymeML used in this work.

A.S.B. thanks the networking program ‘Sustainable Chemical Synthesis 2.0’ (SusChemSys 2.0) for the support and fruitful discussions across disciplines.

References

1. R. Siedentop et al., “Getting the Most Out of Enzyme Cascades: Strategies to Optimize In Vitro Multi-Enzymatic Reactions,” Catalysts 2021, 11, 1183., doi: https://doi.org/10.3390/catal11101183
2. P. De Santis et al., “The rise of continuous flow biocatalysis – fundamentals, very recent developments and future perspectives,” In React. Chem. Eng. 5 (12), pp. 2155–2184., doi: https://doi.org/10.1039/D0RE00335B
3. D. Medeiros, “DWSIM - Open Source Process Simulator,” URL: https://dwsim.org/
4. M.J. Menke et al., “Development of an Ontology for Biocatalysis,” Chemie Ingenieur Technik, 2022, 94: 1827-1835, doi: https://doi.org/10.1002/cite.202200066
5. J. Grühn et al., ”From Coiled Flow Inverter to Stirred Tank Reactor – Bioprocess Development and Ontology Design,” Chemie Ingenieur Technik, 2022, 94: 852-863., doi: https://doi.org/10.1002/cite.202100177
6. J. Range et al., “EnzymeML—a data exchange format for biocatalysis and enzymology,” FEBS J, 2022, 289: 5864-5874., doi: https://doi.org/10.1111/febs.16318
7. S. Lauterbach et al., “EnzymeML: seamless data flow and modeling of enzymatic data,” Nat. Methods, 2023, 20, 400–402., doi: https://doi.org/10.1038/s41592-022-01763-1
8. Jan Range, Frank Bergmann, Johann Rohwer, AnnaReisch, Hannah Dienhart, & SL-2204. (2022). EnzymeML/PyEnzyme: PyEnzyme 1.1.3 (v1.1.3). Zenodo. https://doi.org/10.5281/zenodo.6457299
9. S. Arndt et al., “Metadata4Ing: An ontology for describing the generation of research data within a scientific activity”. (1.1.0). Zenodo. DOI: https://doi.org/10.5281/zenodo.770601
10. J. B. Lamy, “Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies,” Artificial Intelligence in Medicine. 80., doi: https://doi.org/10.1016/j.artmed.2017.07.002