

Standardized Data, Scalable Documentation, Sustainable Storage – EnzymeML As A Basis For FAIR Data Management In Biocatalysis

Jürgen Pleiss*^[a]

The often reported reproducibility crisis in the biomedical sciences also applies to enzymology and biocatalysis, and mainly results from incomplete reporting of reaction conditions. In this Concept article, an infrastructure based on EnzymeML is sketched, which enables reporting, exchange, and storage of enzymatic data according to the FAIR data principles. EnzymeML is a novel data exchange format for enzymology and biocatalysis, which facilitates the application of the STREND A Guidelines and thus makes data on enzyme-catalyzed reactions findable, accessible, interoperable, and reusable. EnzymeML enables the comprehensive documentation of metadata, thus

fostering reproducibility and replicability in enzymology and biocatalysis. An EnzymeML Application Programming Interface integrates electronic lab notebooks with modelling platforms and databases on enzymatic reactions, and thus enables the seamless flow of enzymatic data from measurement to modelling to publication, without the need for manual intervention such as reformatting or editing. EnzymeML serves as a valuable tool for the design of biocatalytic experiments and contributes to the vision of a unified research data infrastructure for catalysis research.

1. Challenges in biocatalysis


Systems biocatalysis and process intensification are considered as cornerstones of a sustainable bioeconomy.^[1] Bioinformatic tools for rapid identification of the required enzyme functions and bioprocess design tools have successfully contributed to rapid prototyping of the biocatalytic reactions. However, biocatalysis is not yet widely applied in the fine chemical industry,^[2] and the success of biocatalysis is still limited by the performance of the biocatalyst and the cost and timeline of development.^[3] Machine learning is a promising approach to handle the complexity of biosystems design^[4] and is based on big data from automated, high throughput, and reproducible experimentation. However, laboratory automation requires open standards and protocols,^[5] and standardization is key to promote interoperability, efficiency, and safety in bioprocess development.^[6] However, our current way to do biocatalytic research and development is still limited by low reproducibility of experimental results, limited scalability of experimentation, and limited access to data. This Concept article is a commentary on challenges in data management and a proposal of a roadmap towards a unified research data infrastructure for catalysis research.

The often reported reproducibility crisis in science^[7] also applies to biocatalytic research. It is mainly due to incomplete reporting of reaction conditions,^[8] which also hinders re-analysis of original data. Measured data and metadata such as reaction conditions is generally not accessible. In principle, all relevant information is stored locally as requested by all funding agencies, but in reality, original data and metadata is hardly findable by the corresponding author and certainly not accessible by the scientific community, especially if the PhD student who generated the data had already left the research group. Even if accessible, data and metadata cannot be easily exchanged between research groups because of missing interoperability, which needs additional efforts for reformatting and results in loss of information. Limited reproducibility severely hampers scientific development and causes additional costs and delays in academic research and industrial development.

A second limitation is the lack of scalability of biocatalytic research, due to our current practices of data management. Robotics and microfluidic systems enable high throughput experimentation and transform biocatalysis into a data-intensive science. Concepts such as retrobiosynthesis^[9] depend on the access to massive data from systematic experimentation and from public databases. Bioprocess design is based on kinetic modelling and on novel data-driven modelling such as machine learning, which require traceable data and reliable kinetic parameters from many thousands of previous experiments. Without a fundamental change of data management with a high level of standardization, an efficient design of scalable bioprocesses by concepts such as reconfigurable reactors^[10] or biofoundries^[6] is not feasible.

The major obstacle to transform biocatalysis into a data science is the inaccessibility of published and unpublished data.

[a] Prof. J. Pleiss
Institute of Biochemistry and Technical Biochemistry,
University of Stuttgart
Allmandring 31, 70569 Stuttgart (Germany)
E-mail: Juergen.Pleiss@itb.uni-stuttgart.de

 © 2021 The Authors. ChemCatChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

The bulk of data is hidden in scientific papers as text, figures, and tables; thus, data extraction is time consuming for a human and impossible for a machine. For enzyme-catalyzed reactions, a number of specialized public databases such as BRENDA,^[11] SABIO-RK,^[12] Enzyme Portal,^[13] and STRENDA DB^[14] provide structured data, but all databases use different data models and data formats, and none of them stores experimental data such as the time course of substrates or products, thus making re-analysis impossible. Storing experimental data on a local file server or on repositories such as Zenodo is requested by funding agencies. However, the use of *ad hoc* data models and formats for storing data and metadata prevents findability and interoperability. Local repositories have limited accessibility, and there is a high risk of data loss. But even publicly available repositories might shut down due to lack of funding, as it happened to the NCBI Peptidome and ProteomeCommons Tranche repositories in 2013, when rescuing data was a challenge.^[15] Therefore, establishing open science policies,^[16] FAIR (findable, accessible, interoperable, reusable) data principles,^[17] and a resilient storage of scientific data has become a major concern of funding agencies.

2. Solution for enzyme catalysis

The efficient design of improved or novel enzymes and the comprehensive characterization of enzyme-catalyzed reactions is one of the bottlenecks in the development of a bioprocess.^[2] For enzyme reactions, the three challenges (reproducibility, scalability, accessibility) are addressed by a novel infrastructure based on EnzymeML, a standardized exchange format for enzymatic data, which enables F.A.I.R data management in enzyme catalysis. The concept is not limited to enzyme catalysis but might be transferred to other research fields in catalytic sciences and beyond.

2.1. Building block 1: Standardized enzymatic data

Many 'omics communities have proposed or instituted best practices and reporting standards in their specific disciplines.^[18] Central to these efforts was the introduction of standardized exchange formats,^[19,20] allowing for the representation of experimental data according to the FAIR data principles.^[17] Previously, a group of enzymologists, enzyme engineers, and bioinformaticians developed EnzymeML to support data acquisition, data analysis, and sharing of data by providing a standardized exchange format for enzymatic data. EnzymeML follows the Standards for Reporting Enzymology Data (STRENDA)^[21] Guidelines, which define the minimum information to describe enzyme activity data – a description of the enzyme, of the reaction conditions, and the results. The STRENDA Guidelines are recommended by more than 50 international biochemistry journals. EnzymeML is written in eXtensible Markup Language (XML). It builds on the well-established Systems Biology Markup Language (SBML)^[22] and includes information about the enzyme, the substrate(s) and product(s), the reaction condi-

tions, the selected kinetic model, and estimated kinetic parameters. In addition, the measured time course of substrate or product concentrations are stored in a comma-separated values (CSV) formatted file. From the reaction conditions and the time course of the substrates and products, biochemical properties such as catalytic activity, yield, stereoselectivity, regioselectivity, and thermostability can be derived. The XML- and the CSV-file are combined into a single EnzymeML document using the widely-used OMEX format.^[23] Documentation and software of the EnzymeML project are freely available for non-commercial and commercial users at <https://EnzymeML.org>.

2.2. Building block 2: Exchange of enzymatic data

The typical user is not expected to read or write EnzymeML documents directly, but to use software to generate EnzymeML documents, which are then used as a standardized exchange format to transfer data between applications (Figure 1). Therefore, an Application Programming Interface (API) was developed to read, write, and edit EnzymeML documents, using the popular programming language Python. Because the API enables batch processing, management of biocatalytic data is scalable, and high throughput strategies of experimentation and data analysis become feasible. By data export in formats such as Pandas DataFrame, large datasets can be analyzed by data-driven modelling methods such as machine learning. Upon reading, writing, and editing of EnzymeML documents, the API controls data completeness and consistency, such as checking that scalar properties such as pH are within a given range. Additional validation tools control compatibility with SBML or with minimum requirements of applications such as STRENDA DB, SABIO-RK, or COPASI. As an alternative to a local installation of the API with each application, it is also accessible as a RESTful Web service, which makes use of standards such as HTTP, JSON, and XML. This Web service enables applications such as electronic laboratory notebooks (ELN), modelling platforms, or specialized database to read or write EnzymeML documents.

These two building blocks respond to the first challenge. They enable interoperability and reproducibility of data and metadata, and guarantee FAIR principles for data and metadata.

2.3. Building block 3: Automated data collection

In most enzyme catalysis projects, experimental procedures and the measured data are still recorded in paper notebooks and electronic spreadsheets, respectively. Since 30 years, the development of ELNs enables researchers to efficiently enter, store, and access experimental procedures and results with a long shelf life.^[24] Several commercial ELNs have become available, but they are still not widely adopted in academia. Recently, open source solutions became available such as openBIS,^[25] which addresses academic life science groups, Chemotion^[26] as a repository for chemistry research data, and BioCatHub as a specialized platform for the documentation of enzyme-cata-

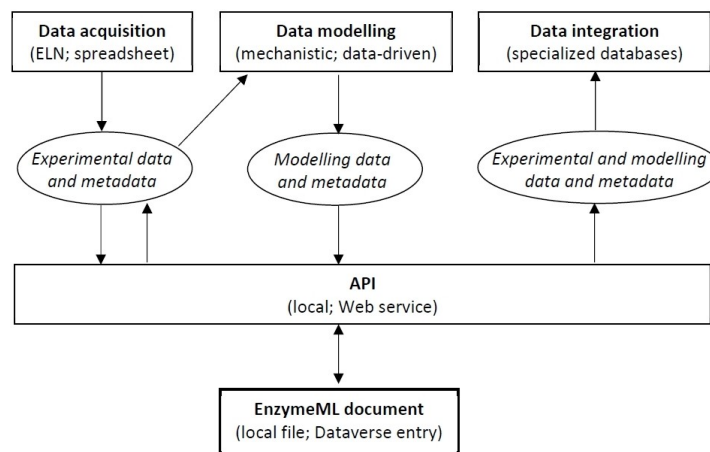


Figure 1. Seamless data flow between tools for data acquisition, data modelling, and data integration as described in Scenario 1. The Application Programming Interface (API) provides the functionality to read, write, and edit EnzymeML documents (as local files or as Dataverse entries).

lyzed reactions.^[27] Because an ELN stores all data at a single place and makes it easily accessible, it speeds up work, facilitates data retrieval, and enables automation of data collection from high throughput experimentation. BioCatHub is already using the API to read and write EnzymeML documents, and the implementation with Chemotion is an ongoing project. By enabling ELNs to read and write EnzymeML documents, the experimental data collected by an ELN can be exchanged with other ELNs, transferred to applications such as modelling tools, and uploaded to specialized databases such as SABIO-RK and STRENDA DB (Figure 1).

This building block responds to the second challenge. It enables upscaling, repeatability, and re-analysis, and guarantees FAIR principles for processes.

2.4. Building block 4: Data repository

Resilience and accessibility of scientific data is a major concern of the scientific community, funding agencies, and the public. A decentralized, distributed data repository reduces bandwidth requirements for data storage and retrieval and contributes to resilience, as compared to a centralized data repository. The Dataverse platform provides a generic open-source data repository system with a configurable metadata schema, with more than 60 installations worldwide (<https://dataverse.org/>).^[28] Upon conversion of the EnzymeML data model into a Dataverse metadata schema, a local implementation of a standardized repository on experimental and modelling results of enzyme-catalyzed reactions is feasible. Individual datasets can be private, shared with project partners, or public. Because the EnzymeML metadata schema is identical on all participating Dataverse installations, all EnzymeML Dataverses form a decentralized, distributed repository of enzymatic data, which uses the same data model and is searchable by the Dataverse API (<https://guides.dataverse.org/>). The EnzymeML Dataverse system is robust, redundant, and resilient, because transferring and

copying of EnzymeML datasets between Dataverse installations is straightforward, and data can be easily rescued in case of shutdown of individual repositories. Each EnzymeML Dataverse entry forms a micropublication with a unique digital object identifier, thus specialized databases for enzymatic data such as STRENDA DB, SABIO-RK, or Enzyme Portal can use EnzymeML Dataverse entries as a literature reference. Because storing enzymatic data as an EnzymeML Dataverse entry is an alternative to creating an EnzymeML file, writing, reading, and editing of an EnzymeML Dataverse entry is an additional functionality of the RESTful EnzymeML API.

This building block responds to the third challenge. It guarantees accessibility, sustainability, resilience, and long-term data security of experimental data.

3. Application

Applications of the EnzymeML-based infrastructure of standardized data, scalable documentation, and sustainable storage are demonstrated for three selected scenarios: the comprehensive description of enzymatic reactions, the analysis of large-scale experiments, and the data exchange between databases.

3.1. Scenario 1: Characterization of an enzymatic reaction

The detailed characterization and kinetic modelling of enzyme-catalyzed reactions is the basis of a successful enzyme engineering and bioprocess design to overcome the limitation of natural enzymes and to tune catalytic activity and selectivity toward the non-natural substrate and improve stability under non-natural conditions.^[2] A comprehensive documentation of the measured time courses of the substrate or product concentrations (usually as replicates) and of the reaction conditions according to the STRENDA Guidelines (identification of substrate, product, and enzyme; concentrations of enzyme, sub-

strate, product, inhibitors, or activators; temperature, solvent, pH) enables advanced modelling approaches and deepens our understanding of the bottlenecks of an enzyme-catalyzed reaction. Reading and writing of an EnzymeML document by a modelling tool such as COPASI^[29] is enabled by the API (Figure 1). The modelling tool reads in an EnzymeML document with experimental data and metadata. After model selection and parameter estimation, the kinetic law, the estimated kinetic parameters, parameter uncertainty, and estimators of model quality are added to the EnzymeML document. Thus, an EnzymeML document is a human- and machine-readable micropublication, which contains the complete information about the experiment and the kinetic modelling, which is either stored locally as a file or as an EnzymeML Dataverse entry with a unique DOI. This micropublication serves as machine-readable literature source for a new entry in a specialized database such as STREDA DB or SABIO-RK. Instead of a manual upload of enzyme activity data, a new entry is created upon upload of an EnzymeML document and the automated extraction of the relevant information.

The analysis of time course data differentiates between kinetic models and avoids misinterpretation of kinetic parameters,^[30] and using thermodynamic activity rather than concentration separates enzyme-substrate from enzyme-solvent interactions.^[31,32] However, not only the experimental reaction conditions mediate the kinetic parameters, but also the modelling method^[33] and even the computer program used for modelling.^[34] Therefore, in order to enable reproducibility of the obtained kinetic parameters, the EnzymeML format will be extended by metadata describing the process of kinetic modelling.

The discussion about an unequivocal definition of an “enzyme unit” is a showcase of the need for stringent standardization. Although the units of enzyme activity have been defined by the International Union of Biochemistry,^[35] there is still ambiguity in published work. The value of “1 U” depends not only on the substrate and the direction of reaction,^[36] but also on the assay conditions and the kinetic modelling.^[37] Therefore, a comprehensive, standardized reporting of the experiment by EnzymeML is pivotal to an unambiguous definition of the enzyme unit. Because EnzymeML provides tools for comprehensive and interoperable documentation of experiments and modelling, it supports the implementation and communication of best practices in enzymology and biocatalysis.

3.2. Scenario 2: Screening experiment

A systematic and extensive investigation of enzymes is needed to understand their biochemical properties and to guide protein engineering. Microfluidic systems developed in the Hollfelder group enabled the analysis of hundreds of combinations of enzymes, substrates, and inhibitors in less than 5 min, resulting in conclusive Michaelis–Menten kinetics and inhibition curves.^[38] Using their HT-MEK microfluidic platform, the Fordyce group analyzed thousands of kinetic experiments for thousands of

enzyme variants and modelled the kinetic and thermodynamic constants.^[39] Thus, they were able to remove effects from misfolding and to quantify mutational effects on intrinsic catalytic activity and other kinetic parameters. Continuous measurement by a flow reactor enabled automatic collecting of large amounts of kinetic data by the Woodley group.^[40] For these experiments, the complete original data and the reaction conditions are not available in a standardized, machine-readable format. However, large data obtained by identical experimental procedures under controlled reaction conditions would be a valuable training set for machine learning approaches or for mechanistic modelling. Storing the complete datasets as a standardized EnzymeML entry on Dataverse would enable the re-analysis by novel data-driven modelling methods.

3.3. Scenario 3: Exchange of data between specialized databases

Currently, the content of specialized databases such as STREDA DB, SABIO-RK, or Enzyme Portal on enzymatic data is still limited, because information is retrieved semi-automatically from scientific literature, followed by a manual and therefore labor-intensive and error-prone curation step. If each publication would provide the relevant experimental data and metadata in a standardized format (as a supplementary EnzymeML document or as an EnzymeML Dataverse entry with a DOI), the upload of new datasets could be fully automated, and a complete coverage of scientific literature is guaranteed. As is, the contents of specialized databases can hardly be compared, because they use different data models and output formats. EnzymeML might serve as a universal, standardized exchange format for specialized databases to exchange enzymatic data.

4. Conclusion

Combining EnzymeML as a standardized data exchange format, an EnzymeML API for interoperability, an EnzymeML-compatible ELN for scalable data acquisition, and a distributed, resilient, and accessible EnzymeML Dataverse repository enables the seamless flow of enzymatic data from measurement to modelling to publication, without the need for manual intervention such as reformatting or editing. It is scalable from one to thousands of experiments and guarantees a complete description of metadata. Data and metadata of the experiment and the modelling process are combined into a single file, or in an EnzymeML Dataverse entry that is addressable by a DOI and serves as a machine-readable micropublication. The ensemble of EnzymeML documents guarantees reproducibility of enzymatic experiments and enables re-analysis of enzymatic data. Though the current version of EnzymeML is still limited to free enzymes in a batch experiment, it is currently extended to comprise immobilized enzymes, whole cell biocatalysts, enzyme-catalyzed cascade reactions, and flow reactions. It serves as a valuable tool for the design of biocatalytic experiments

and contributes to the vision of a unified research data infrastructure for catalysis research.^[41]

Acknowledgements

J. P. acknowledges Mercè Crosas, Carsten Kettner, Maria J. Martin, Marilize Le Roess-Hill, Santiago Schnell, and Neil Swainston for inspiring discussions, Frank T. Bergmann, Colin Halupczok, Dorothea Iglezakis, Jens Lohmann, Jan Range, Santiago Schnell, and Ulrike Wittig for implementing the EnzymeML infrastructure, and Deutsche Forschungsgemeinschaft DFG (EXC 2075) for funding. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Application Programming Interface · enzymatic reactions · FAIR data principles · reproducibility crisis · STREND A Guidelines

- [1] R. Wohlgemuth, *Nat. Biotechnol.* **2021**, *60*, 113–123.
- [2] J. M. Woodley, *Appl. Microbiol. Biotechnol.* **2019**, *103*, 4733–4739.
- [3] B. Hauer, *ACS Catal.* **2020**, *10*, 8418–8427.
- [4] M. J. Volk, I. Lourentzou, S. Mishra, L. T. Vo, C. Zhai, H. Zhao, *ACS Synth. Biol.* **2020**, *9*, 1514–1533.
- [5] H. Bär, R. Hochstrasser, B. Papenfuß, *J. Lab. Autom.* **2012**, *17*, 86–95.
- [6] N. Hillson, M. Caddick, Y. Cai, J. A. Carrasco, M. W. Chang, N. C. Curach, D. J. Bell, R. L. Feuvre, D. C. Friedman, X. Fu, N. D. Gold, M. J. Herrgård, M. B. Holowko, J. R. Johnson, R. A. Johnson, J. D. Keasling, R. I. Kitney, A. Kondo, C. Liu, V. J. J. Martin, F. Menolascina, C. Ogino, N. J. Patron, M. Pavan, C. L. Poh, I. S. Pretorius, S. J. Rosser, N. S. Scrutton, M. Storch, H. Tekotte, E. Travník, C. E. Vickers, W. S. Yew, Y. Yuan, H. Zhao, P. S. Freemont, *Nat. Commun.* **2019**, *10*, 1038–1041.
- [7] M. Baker, D. Penny, *Nature* **2016**, *533*, 452–454.
- [8] P. Halling, P. F. Fitzpatrick, F. M. Rauschel, J. Rohwer, S. Schnell, U. Wittig, R. Wohlgemuth, C. Kettner, *Biophys. Chem.* **2018**, *242*, 22–27.
- [9] N. J. Turner, E. O'Reilly, *Nat. Chem. Biol.* **2013**, *9*, 285–288.
- [10] A. C. Bédard, A. Adamo, K. C. Aroh, M. G. Russell, A. A. Bedermann, J. Torosian, B. Yue, K. F. Jensen, T. F. Jamison, *Science* **2018**, *361*, 1220–1225.
- [11] I. Schomburg, A. Chang, D. Schomburg, *Nucleic Acids Res.* **2002**, *30*, 47–49.
- [12] U. Wittig, R. Kania, M. Golebiewski, M. Rey, L. Shi, L. Jong, E. Algaa, A. Weidemann, H. Sauer-Danzwith, S. Mir, O. Krebs, M. Bittkowski, E. Wetsch, I. Rojas, W. Müller, *Nucleic Acids Res.* **2011**, *40*, D790–D796.
- [13] R. Alcántara, J. Onwubiko, H. Cao, P. De Matos, J. A. Cham, J. Jacobsen, G. L. Holliday, J. D. Fischer, S. A. Rahman, B. Jassal, M. Goujon, F. Rowland, S. Velankar, R. López, J. P. Overington, G. J. Kleywegt, H. Hermjakob, C. O'Donovan, M. J. Martín, J. M. Thornton, C. Steinbeck, *Nucleic Acids Res.* **2013**, *41*, 773–780.
- [14] N. Swainston, A. Baici, B. M. Bakker, A. Cornish-Bowden, P. F. Fitzpatrick, P. Halling, T. S. Leyh, C. O'Donovan, F. M. Rauschel, U. Reschel, J. M. Rohwer, S. Schnell, D. Schomburg, K. F. Tipton, M. D. Tsai, H. V. Westerhoff, U. Wittig, R. Wohlgemuth, C. Kettner, *FEBS J.* **2018**, *285*, 2193–2204.
- [15] L. Martens, *Proteomics* **2013**, *13*, 1548–1550.
- [16] E. A. Zerhouni, *Science* **2004**, *306*, 1895.
- [17] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, *Sci. Data* **2016**, *3*, 160018.
- [18] C. E. Mason, E. Afshinnekoo, S. Tighe, S. Wu, S. Levy, *J. Biomol. Tech.* **2017**, *28*, 8–18.
- [19] C. F. Taylor, D. Field, S. A. Sansone, J. Aerts, R. Apweiler, M. Ashburner, C. A. Ball, P. A. Binz, M. Bogue, T. Booth, A. Brazma, R. R. Brinkman, A. Michael Clark, E. W. Deutsch, O. Fiehn, J. Fostel, P. Ghazal, F. Gibson, T. Gray, G. Grimes, J. M. Hancock, N. W. Hardy, H. Hermjakob, R. K. Julian, M. Kane, C. Kettner, C. Kinsinger, E. Kolker, M. Kuiper, N. Le Novère, J. Leebens-Mack, S. E. Lewis, P. Lord, A. M. Mallon, N. Marthandan, H. Masuya, R. McNally, A. Mehrle, N. Morrison, S. Orchard, J. Quackenbush, J. M. Reecy, D. G. Robertson, P. Rocca-Serra, H. Rodriguez, H. Rosenfelder, J. Santoyo-Lopez, R. H. Scheuermann, D. Schober, B. Smith, J. Snape, C. J. Stoeckert, K. Tipton, P. Sterk, A. Untergasser, J. Vandesompele, S. Wiemann, *Nat. Biotechnol.* **2008**, *26*, 889–896.
- [20] S. A. Sansone, P. McQuilton, P. Rocca-Serra, A. Gonzalez-Beltran, M. Izzo, A. L. Lister, M. Thurston, *Nat. Biotechnol.* **2019**, *37*, 358–367.
- [21] K. F. Tipton, R. N. Armstrong, B. M. Bakker, A. Bairoch, A. Cornish-Bowden, P. J. Halling, J.-H. Hofmeyr, T. S. Leyh, C. Kettner, F. M. Rauschel, J. Rohwer, D. Schomburg, *Perspect. Sci.* **2014**, *1*, 131–137.
- [22] M. Hucka, F. T. Bergmann, A. Dräger, S. Hoops, S. M. Keating, N. Le Novère, C. J. Myers, B. G. Olivier, S. Sahle, J. C. Schaff, L. P. Smith, D. Waltemath, D. J. Wilkinson, *J. Integr. Bioinform.* **2018**, *15*, 1–181.
- [23] F. T. Bergmann, R. Adams, S. Moodie, J. Cooper, M. Glont, M. Golebiewski, M. Hucka, C. Laibe, A. K. Miller, D. P. Nickerson, B. G. Olivier, N. Rodriguez, H. M. Sauro, M. Scharm, S. Soiland-Reyes, D. Waltemath, F. Yvon, N. Le Novère, *BMC Bioinf.* **2014**, *15*, 369.
- [24] R. E. Dessy, *Chemom. Intell. Lab. Syst.* **1990**, *8*, 2–3.
- [25] C. Barillari, D. S. M. Ottoz, J. M. Fuentes-Serna, C. Ramakrishnan, B. Rinn, F. Rudolf, *Bioinformatics* **2016**, *32*, 638–40.
- [26] P. Tremouilhac, C. L. Lin, P. C. Huang, Y. C. Huang, A. Nguyen, N. Jung, F. Bach, R. Ulrich, B. Neumair, A. Streit, S. Bräse, *Angew. Chem. Int. Ed.* **2020**, *59*, 22771–22778; *Angew. Chem.* **2020**, *132*, 22960–22968.
- [27] S. Malzacher, J. Range, C. Halupczok, J. Pleiss, D. Rother, *Chem. Ing. Tech.* **2020**, *92*, 1251–1251.
- [28] B. McKinney, P. A. Meyer, M. Crosas, P. Sliz, *Ann. N. Y. Acad. Sci.* **2017**, *1387*, 95–104.
- [29] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, U. Kummer, *Bioinformatics* **2006**, *22*, 3067–3074.
- [30] P. C. F. Buchholz, R. Ohs, A. C. Spiess, J. Pleiss, *Biotechnol. J.* **2019**, *14*, 1800183.
- [31] J.-H. Grosch, D. Wagner, V. Nistelkas, A. C. Spieß, *Biotechnol. Prog.* **2017**, *33*, DOI 10.1002/btpr.2401.
- [32] J. Pleiss, *Trends Biotechnol.* **2017**, *35*, 379–382.
- [33] W. Stroberg, S. Schnell, *Biophys. Chem.* **2016**, *219*, 17–27.
- [34] M. Zavrel, K. Kochanowski, A. C. Spiess, *Eng. Life Sci.* **2010**, *10*, 191–200.
- [35] *Eur. J. Biochem.* **1979**, *97*, 319–320.
- [36] C. Bauer, G. Cercignani, M. Paolini, *Lancet* **1996**, *348*, 271.
- [37] H. Bisswanger, *Perspect. Sci.* **2014**, *1*, 41–55.
- [38] F. Gielen, T. Buryška, L. Van Vliet, M. Butz, J. Damborsky, Z. Prokop, F. Hoffelder, *Anal. Chem.* **2015**, *87*, 624–632.
- [39] C. J. Markin, D. A. Mokhtari, F. Sunden, M. J. Appel, E. Akiva, S. A. Longwell, C. Sabatti, D. Herschlag, P. M. Fordyce, *bioRxiv* **2020**.
- [40] R. H. Ringborg, A. Toftgaard Pedersen, J. M. Woodley, *ChemCatChem* **2017**, *9*, 3285–3288.
- [41] C. Wulf, M. Beller, T. Boenisch, O. Deutschmann, S. Hanf, N. Kockmann, R. Kraehnert, M. Oezaslan, S. Palkovits, S. Schimmler, S. A. Schunk, K. Wagemann, D. Linke, *ChemCatChem* **2021**, DOI 10.1002/cctc.202001974.

Manuscript received: June 8, 2021

Revised manuscript received: June 30, 2021

Accepted manuscript online: July 17, 2021

Version of record online: July 22, 2021