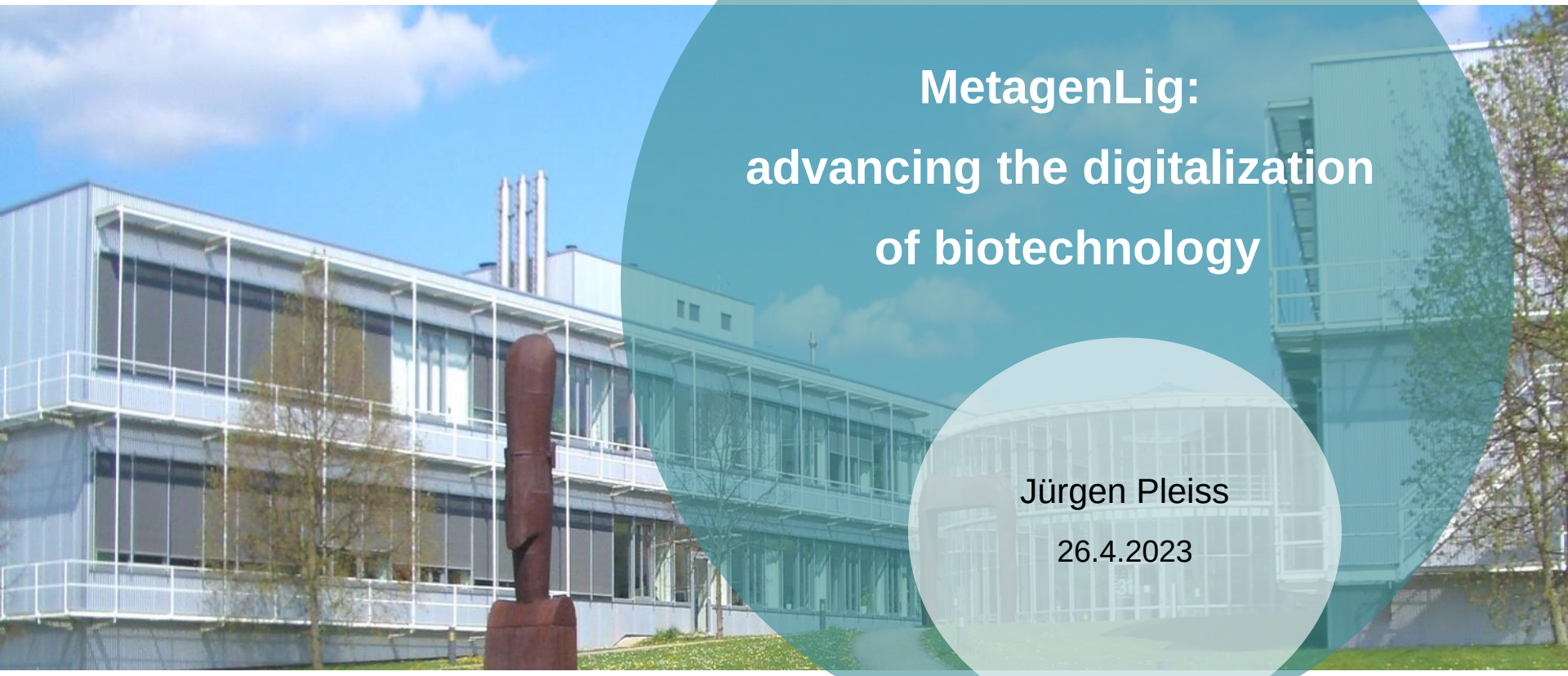
Universität Stuttgart
Institut für Biochemie und Technische Biochemie



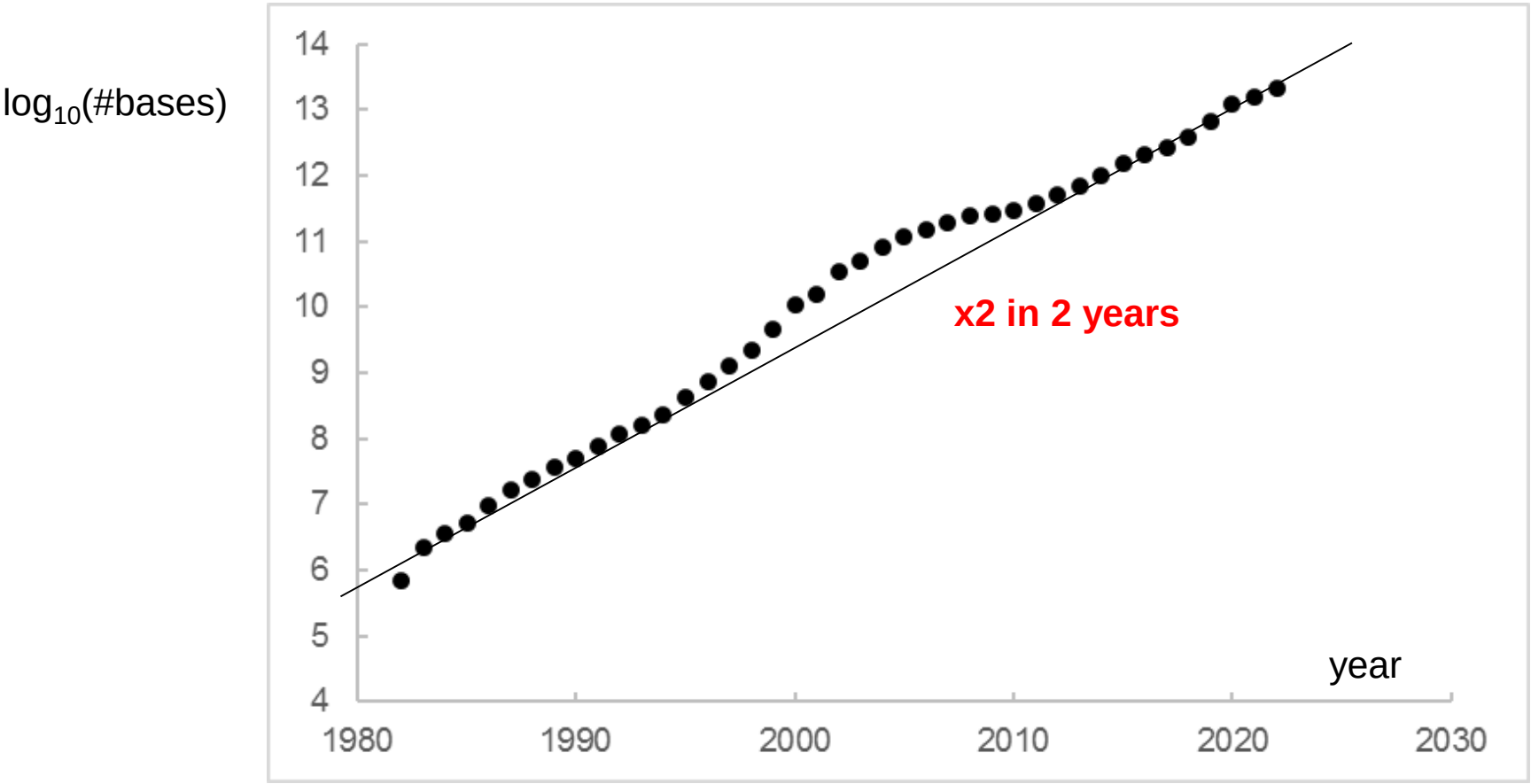
MetagenLig: advancing the digitalization of biotechnology

Jürgen Pleiss

26.4.2023

Opportunities and challenges in biotechnology

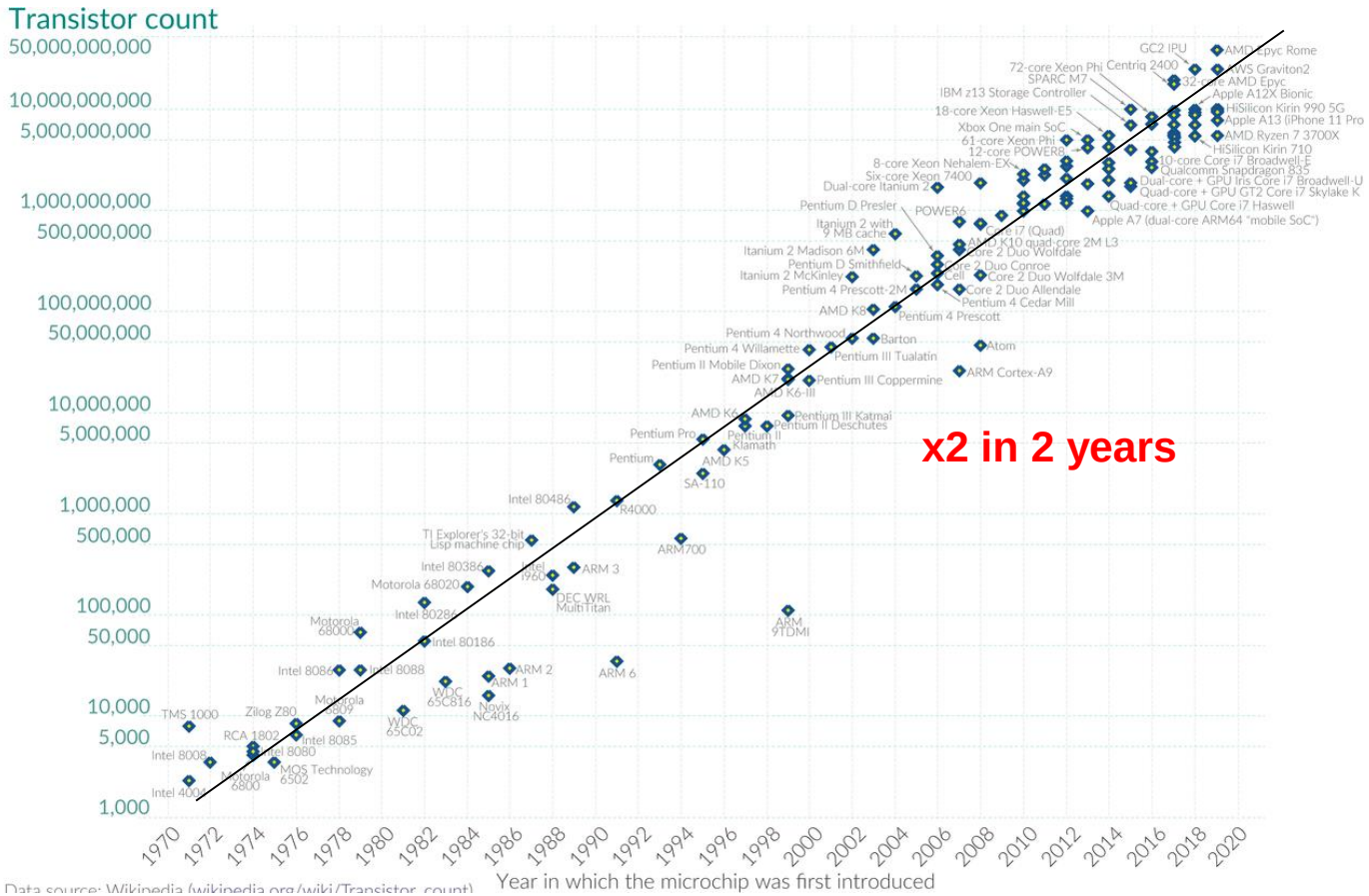
- Exponentially increasing volume of (meta)genomic data → novel enzyme candidates



GenBank statistics: *https://www.ncbi.nlm.nih.gov/genbank/statistics/*

Opportunities and challenges in biotechnology

- Exponentially increasing volume of (meta)genomic data → novel enzyme candidates
- Exponentially increasing computational resources → novel modeling paradigms



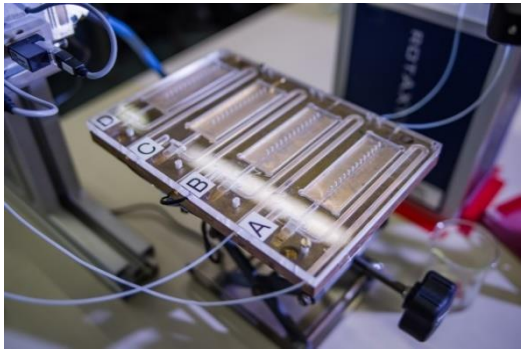
https://en.wikipedia.org/wiki/Moore%27s_law

Opportunities and challenges in biotechnology

- Exponentially increasing volume of (meta)genomic data → novel enzyme candidates
- Exponentially increasing computational resources → novel modeling paradigms
- Laboratory automation → high-throughput experimentation

Pipetting robot
https://www.kiwi.tu-berlin.de/

Nanopore sequencing
https://nanoporetech.com



Microfluidics
https://www.niemeyer-lab.de/applications/

Portable, modular mini factory
https://www.invite-research.com/

Opportunities and challenges in biotechnology

- Exponentially increasing volume of (meta)genomic data → novel enzyme candidates
- Exponentially increasing computational resources → novel modeling paradigms
- Laboratory automation → high-throughput experimentation

Digital transformation:

From a **qualitative** description of properties of biochemical systems …

… to complex **quantitative** models of enzymes, organisms, and bioprocesses

From **trial-and-error**-based development strategies…

… to the **predictive** design of enzymes, solvent systems, and bioprocesses

Opportunities and challenges in biotechnology

- Exponentially increasing volume of (meta)genomic data → novel enzyme candidates
- Exponentially increasing computational resources → novel modeling paradigms
- Laboratory automation → high-throughput experimentation

Digital transformation:

From a **qualitative** description of properties of biochemical systems …

… to complex **quantitative** models of enzymes, organisms, and bioprocesses

From **trial-and-error**-based development strategies…

… to the **predictive** design of enzymes, solvent systems, and bioprocesses

MS Excel or vendor-specific software are not sufficient

→ need for **digitalization** in biotechnology (**Industry 4.0**)

The key to success: manage and analyze big data

1. Databases (sequence, structure, function)

2. Extensible workflow platforms

3. Standardized data formats

The key to success: manage and analyze big data

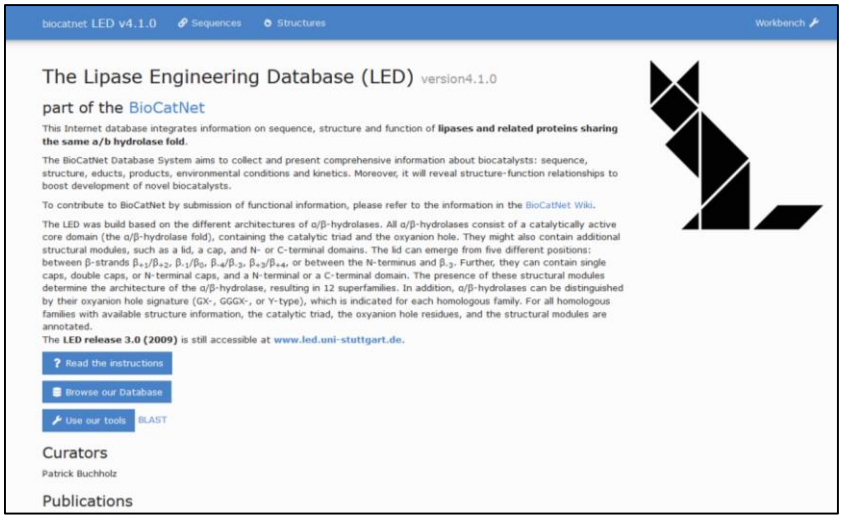
1. Databases (sequence, structure, function) $\rightarrow$ plastics active enzymes, expansins, α/β hydrolases, glycoside hydrolases

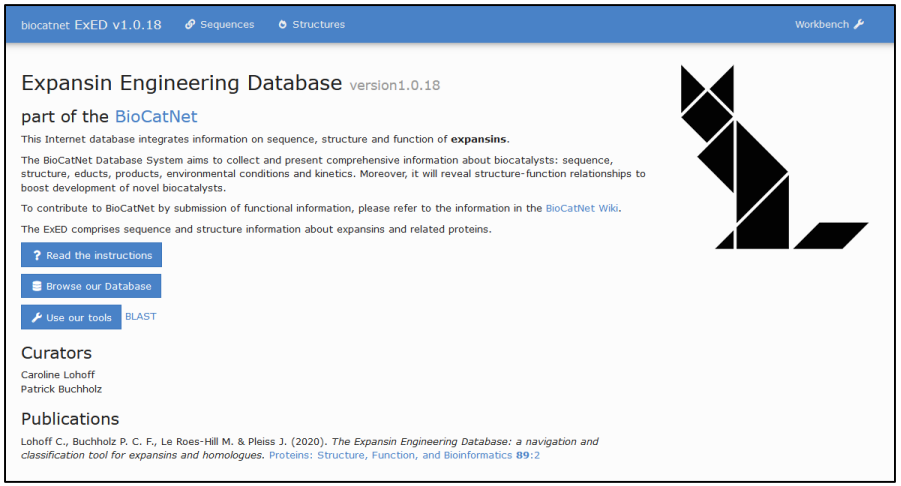
2. Extensible workflow platforms $\rightarrow$ PyEED

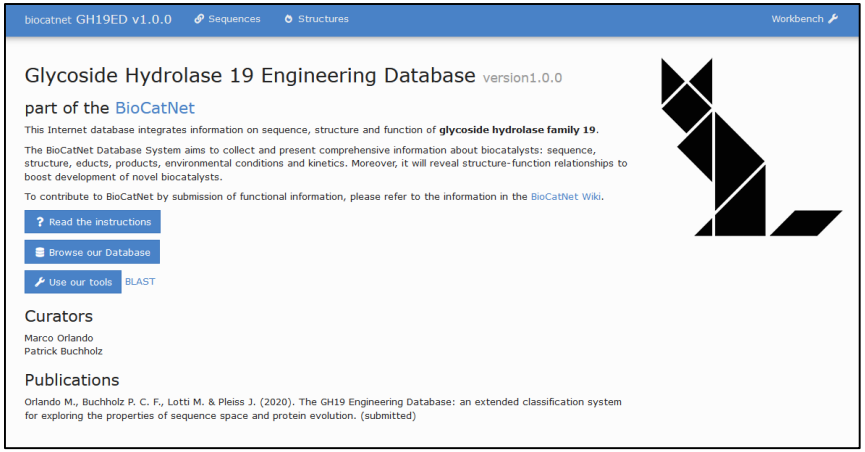
3. Standardized data formats $\rightarrow$ EnzymeML

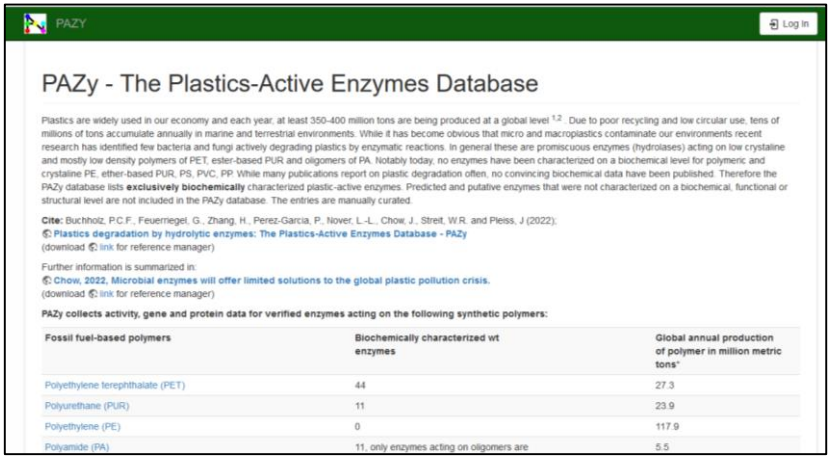
1. Databases

1. Lipase Engineering Database: *https://led.biocatnet.de/*
2. Expansin Engineering Database: *https://exed.biocatnet.de/*
3. Glycoside Hydrolase 19 Engineering Database: *https://gh19ed.biocatnet.de/*
4. PAZy - The Plastics-Active Enzymes Database: *https://pazy.eu/*





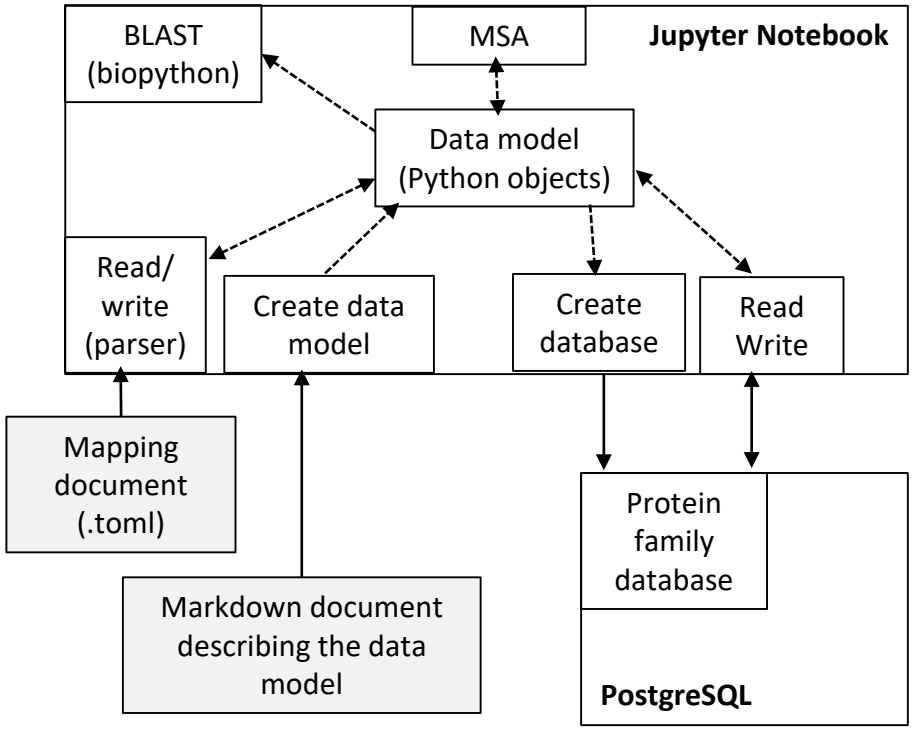




2. PyEED: Python Enzyme Engineering Database

Modular and reusable **bioinformatics workflow platform** based on:

- Python / Biopython
- Jupyter Notebook
- PostgreSQL
- GitHub

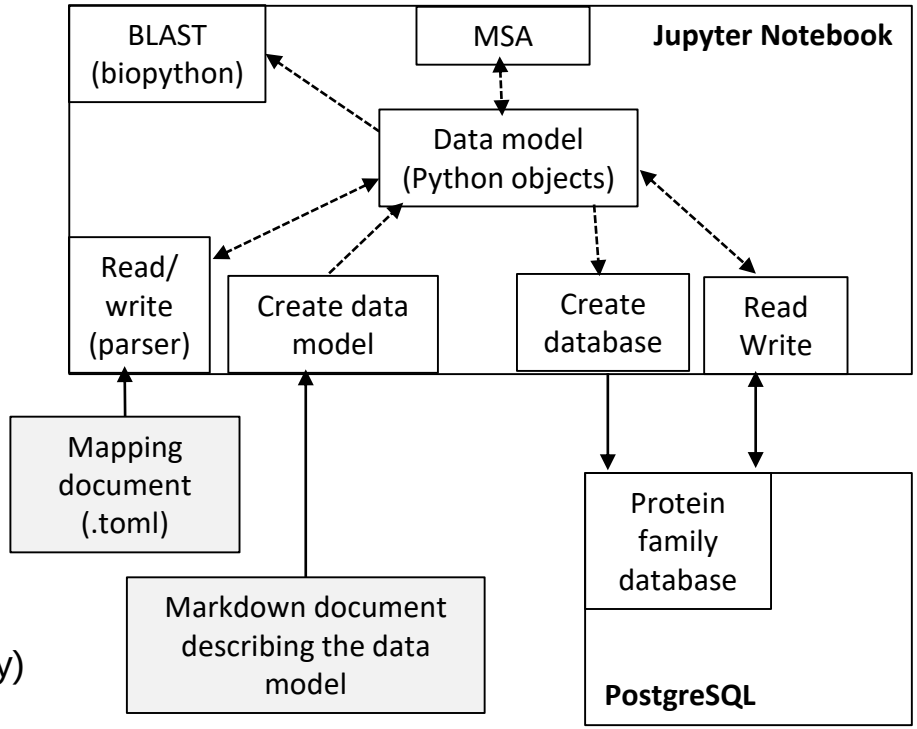


https://github.com/PyEED

2. PyEED: Python Enzyme Engineering Database

Modular and reusable **bioinformatics workflow platform** based on:

- Python / Biopython
- Jupyter Notebook
- PostgreSQL
- GitHub



International collaborations:

Jennifer Andexer (University of Freiburg)

Tom Desmet (University of Gent)

Ho Huynh Thuy Duong (KTEST, Ho Chi Minh City)

Marilize Le Roes-Hill (CPUT, Cape Town)

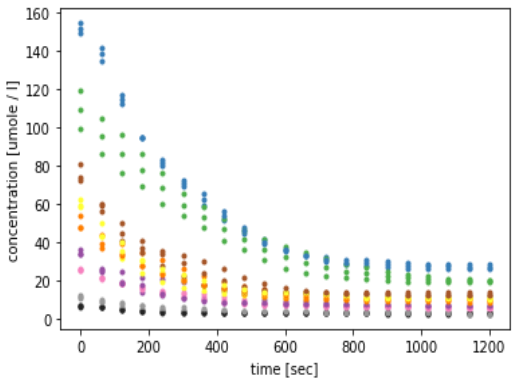
Peter Oelschlaeger (Western University of Health Sciences, Pomona)

Wolfgang Streit (University of Hamburg)

Truong Nam Hai (Vietnam Academy of Science and Technology, Ha Noi)

https://github.com/PyEED

3. EnzymeML: a standardized data exchange format for biocatalysis



Experimental platforms

↓

EnzymeML

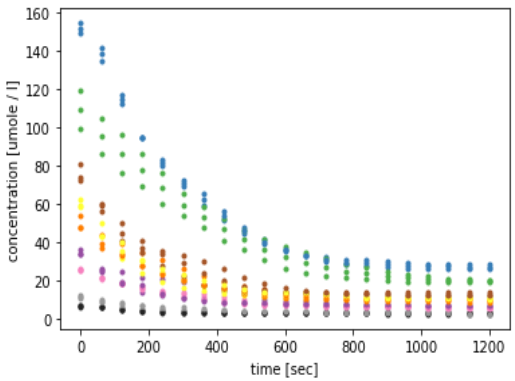
complete reporting of experimental and modelling (meta-)data

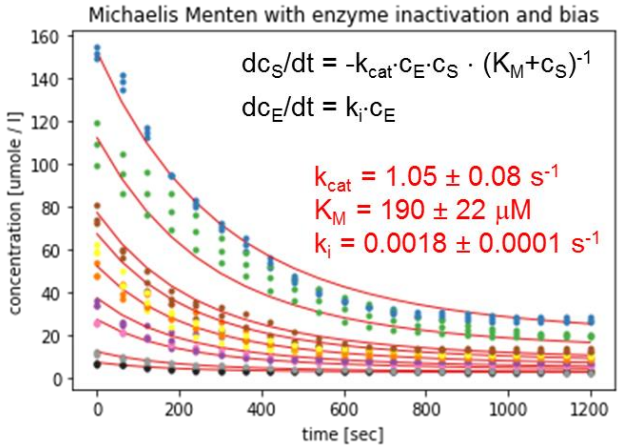
Lauterbach S, Dienhart H, Range J, Malzacher S, Spöring JD, Rother D, Pinto MF, Martins P, Lagerman CE, Bommarius AS, Vang Høst A, Woodley JM, Ngubane S, Kudanga T, Bergmann FT, Rohwer JM, Iglezakis D, Weidemann A, Wittig U, Kettner C, Swainston N, Schnell S, Pleiss J (2023) *Nat Methods* **20**: 400

Range J, Halupczok C, Lohmann J, Swainston N, Kettner C, Bergmann FT, Weidemann A, Wittig U, Schnell S, Pleiss J. (2022) *FEBS J* **289:** 5864

Pleiss J (2021) *ChemCatChem* **13:** 3909

3. EnzymeML: a standardized data exchange format for biocatalysis





Experimental platforms

Modelling platforms

EnzymeML

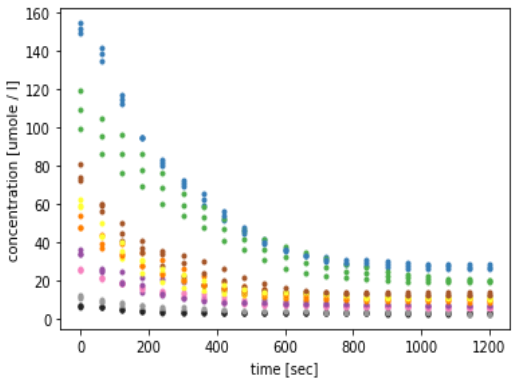
complete reporting of experimental and modelling (meta-)data

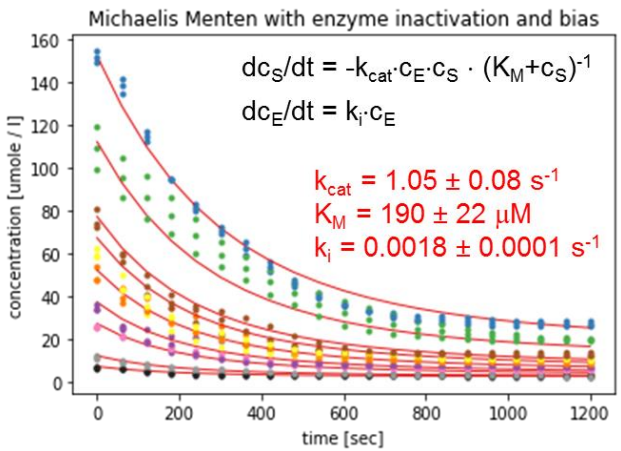
Lauterbach S, Dienhart H, Range J, Malzacher S, Spöring JD, Rother D, Pinto MF, Martins P, Lagerman CE, Bommarius AS, Vang Høst A, Woodley JM, Ngubane S, Kudanga T, Bergmann FT, Rohwer JM, Iglezakis D, Weidemann A, Wittig U, Kettner C, Swainston N, Schnell S, Pleiss J (2023) *Nat Methods* **20**: 400

Range J, Halupczok C, Lohmann J, Swainston N, Kettner C, Bergmann FT, Weidemann A, Wittig U, Schnell S, Pleiss J. (2022) *FEBS J* **289:** 5864

Pleiss J (2021) *ChemCatChem* **13:** 3909

3. EnzymeML: a standardized data exchange format for biocatalysis





Assay Conditions

Small Assay Components

Name	Role	Concentration
Ethylenediaminetetraacetic acid disodium...	Other Compound	1.0 mM
HEPES	Buffer	20.0 mM
CHAPS	Other Compound	2.0 mM
DL-1,4-Dithiothreitol	Other Compound	10.0 mM
sucrose	Other Compound	0.292 M
sodium chloride	Salt	100.0 mM
Ac-DEVD-AMC	Substrate	3.125 - 300.0 µM
7-Amino-4-methylcoumarin	Product	3.125 - 300.0 µM

Physical Properties

pH	pD	Temperature
7.2		37.0 °C

Experimental platforms

Modelling platforms

Publication platforms

EnzymeML

complete reporting of experimental and modelling (meta-)data

Lauterbach S, Dienhart H, Range J, Malzacher S, Spöring JD, Rother D, Pinto MF, Martins P, Lagerman CE, Bommarius AS, Vang Høst 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) *Nat Methods* **20**: 400

Range J, Halupczok C, Lohmann J, Swainston N, Kettner C, Bergmann FT, Weidemann A, Wittig U, Schnell S, Pleiss J. (2022) *FEBS J* **289:** 5864

Pleiss J (2021) *ChemCatChem* **13:** 3909

EnzymeML: an international collaborative project

Experiments

Jennifer Andexer (University of Freiburg)
Andreas Bommarius (Georgia Tech, Atlanta)
Tomasz Borowski (Jerzy Haber Institute of Catalysis and Surface Chemistry, Krakow)
Rebecca Buller (ZHAW, Zürich)
Mehdi Davari (Leibniz Institute of Plant Biochemistry, Halle)
Tom Desmet (University of Gent)
Zvjezdana Findrik Blazevic (University of Zagreb)
Marco Fraaije (University of Groningen)
Harald Gröger (University of Bielefeld)
Egon Heuson (University of Lille)
Frank Hollmann (TU Delft)
Dirk Holtmann (THM Giessen)
Roland Kontermann (University of Stuttgart)
Tukayi Kudanga (DTU Durban)
Marilize Le Roes-Hill (CPUT, Cape Town)
Marina Lotti (University of Milano-Bicocca)
Ligia Martins (NOVA University, Lisboa)
Daniel Ohde (TU Hamburg)
Caroline Paul (TU Delft)
Brett Pletschke (Rhodes University, Makhanda)
Katrin Rosenthal (TU Dortmund)
Dörte Rother (Forschungszentrum Jülich)
Wolfgang Streit (University of Hamburg)
Dirk Tischler (University of Bochum)
John Woodley (Technical University of Denmark)
Anne Zaparucha (Genoscope, Évry)

Modeling

Frank Bergmann (Heidelberg): COPASI
William Finnigan (University of Manchester): RetroBioCat
Pedro Martins (University of Porto): interferENZY
Nicole Radde (University of Stuttgart)
Johann Rohwer (University of Stellenbosch): PySCeS

Database

Dorothea Iglezakis (University of Stuttgart) : Dataverse
Carsten Kettner (Beilstein-Institut, Frankfurt): STRENDA DB
Ulrike Wittig (Heidelberg Institute for Theoretical Studies): SABIO-RK

Electronic Lab Notebook

Caterina Barillari (ETH Zürich): openBIS
Nicole Jung (KIT Karlsruhe): Chemotion
Stephan Malzacher (Forschungszentrum Jülich): BioCatHub
Felix Rudolphi : sciformation

Digitalization of biotechnology: what is needed?

Scientific methods

- automation of experimentation: model-based design of experiments
- data analysis: mechanistic and data-driven modelling

Digitalization of biotechnology: what is needed?

Scientific methods

- automation of experimentation: model-based design of experiments
- data analysis: mechanistic and data-driven modelling

Interdisciplinary research projects

- tight integration of experiment with data analysis, modelling, and design

Digitalization of biotechnology: what is needed?

Scientific methods

- automation of experimentation: model-based design of experiments
- data analysis: mechanistic and data-driven modelling

Interdisciplinary research projects

- tight integration of experiment with data analysis, modelling, and design

Infrastructure

- personnel: research software engineers, data scientists
- computational ressources
- data infrastructure: standards, repositories

Digitalization of biotechnology: what is needed?

Scientific methods

- automation of experimentation: model-based design of experiments
- data analysis: mechanistic and data-driven modelling

Interdisciplinary research projects

- tight integration of experiment with data analysis, modelling, and design

Infrastructure

- personnel: research software engineers, data scientists
- computational ressources
- data infrastructure: standards, repositories

International training

- concepts and scientific methods
- computational tools
- programming skills

Acknowledgements

MetagenLig partners

Truong Nam Hai (Institute of Biotechnology, VAST, Ha Noi)

Wolfgang Streit (University of Hamburg)

Acknowledgements

MetagenLig partners

Truong Nam Hai (Institute of Biotechnology, VAST, Ha Noi)

Wolfgang Streit (University of Hamburg)

EnzymeML Development Team

Jan Range, Torsten Giess, Max Häußler (University of Stuttgart)

Stephan Malzacher (Forschungszentrum Jülich)

Frank Bergmann (University of Heidelberg)

Carsten Kettner (Beilstein-Institut, Frankfurt)

Santiago Schnell (University of Notre Dame)

PyEED Fellows

Maxim Schwarzberg (University of Stuttgart)

Golo Feuerriegel (University of Hamburg)

Alex Windels (University of Gent)

Lars-Hendrik Köppl (University of Freiburg)

Alaric Prins (CPUT, Cape Town)

Acknowledgements

MetagenLig partners

Truong Nam Hai (Institute of Biotechnology, VAST, Ha Noi)

Wolfgang Streit (University of Hamburg)

EnzymeML Development Team

Jan Range, Torsten Giess, Max Häußler (University of Stuttgart)

Stephan Malzacher (Forschungszentrum Jülich)

Frank Bergmann (University of Heidelberg)

Carsten Kettner (Beilstein-Institut, Frankfurt)

Santiago Schnell (University of Notre Dame)

PyEED Fellows

Maxim Schwarzberg (University of Stuttgart)

Golo Feuerriegel (University of Hamburg)

Alex Windels (University of Gent)

Lars-Hendrik Köppl (University of Freiburg)

Alaric Prins (CPUT, Cape Town)

Funding

DFG: Excellence Cluster EXC 2075 "Data-Integrated Simulation Science"

DFG: Collaborative Research Center 1333 "Molecular heterogeneous catalysis in confined geometries"

BMBF: German – South African cooperation in science and research

BMBF: Bioeconomy International "MetagenLig"

EU: COST Action "COZYME"

Results from MetagenLig

Research stays

1.4. - 30.9.2019	Dao Trong Khoa (IBT,Hanoi) in Stuttgart
1.10.19 -31.3.2020	Dao Trong Khoa (IBT,Hanoi) in Hamburg
1.5.-5.7.2019	Marco Orlando (University of Milano-Bicocca) in Stuttgart

Project meetings, workshops, conferences

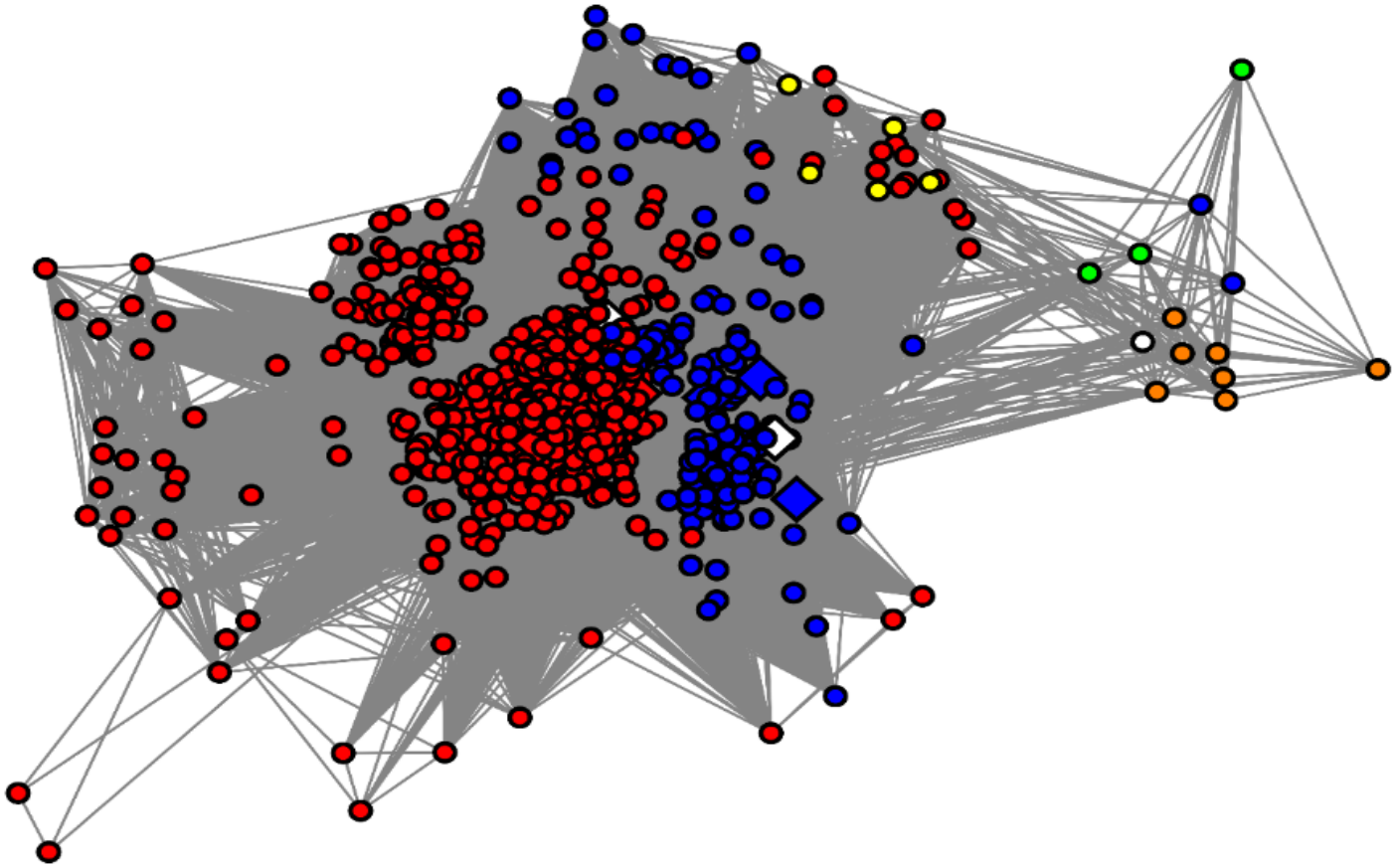
18.-20.6.2018	Kickoff meeting (Hanoi)
8.-10.10.2018	Project meeting (Stuttgart)
2.-6.7.2019	Project meeting (Stuttgart)
3.-5. 7.2019	International Workshop "Nature as a teacher: big data, complex systems, new chemistry" (Stuttgart)
30.10.2019	Poster at German Science Day in Hanoi
5.-7.10.2020	International Workshop "Systematic profiling of multicopper oxidases" (online)
9.11.2022	Final MetagenLig Workshop (online)

Results from MetagenLig

Scientific papers:

1. Le TTH, Nguyen TB, Nguyen HD, Nguyen HD, Le NG, Dao TK, Nguyen TQ, Do TH, Truong NH (2022) De novo metagenomic analysis of microbial community contributing in lignocellulose degradation in humus samples harvested from Cuc Phuong tropical forest in Vietnam. *Diversity* **14**: 220.
2. Zhang H, Pérez-García P, Dierkes RF, Danso D, Pleiss J, Almeida A, Höcker B, Schmitz-Streit RA, Chow J, Streit WR (2022) The Bacteroidetes Aequorivita sp. and Kaistella jeonii produce promiscuous esterases with PET-hydrolyzing activity. *Front Microbiol* **12**: 803896.
3. Buchholz PCF, Feuerriegel G, Zhang H, Pérez-García P, Nover LL, Chow J, Streit WR, Pleiss J (2022) Plastics degradation by hydrolytic enzymes: the Plastics-Active Enzymes Database - PAZy. *Proteins* **90:** 1443.
4. Chow J, Pleiss J, Streit WR (2022) PAZy – mikrobielle Enzyme für den Abbau künstlicher Polymere. *BIOspektrum* **4:** 451.
5. Antranikian G, Streit WR (2022) Microorganisms harbor keys to a circular bioeconomy making them useful tools in fighting plastic pollution and rising CO_2 levels. *Extremophiles* **26**:10.
6. Lohoff C, Buchholz PCF, Le Roes-Hill M, Pleiss J (2021) The Expansin Engineering Database: a navigation and classification tool for expansins and homologues. *Proteins* **89**: 149.
7. Orlando M, Buchholz PCF, Lotti M, Pleiss J (2021) The GH19 Engineering Database: sequence diversity, substrate scope, and evolution in glycoside hydrolase family 19. *PLoS One* **16**: e0256817.
8. Pérez-García P, Danso D, Zhang H, Chow J, Streit WR (2021) Exploring the global metagenome for plastic-degrading enzymes. *Methods in Enzymology* **648**:137
9. Dao TK, Do TH, Le NG, Nguyen HD, Nguyen TQ, Le TTH, Truong NH (2021) Understanding the role of Prevotella genus in the digestion of lignocellulose and other substrates in Vietnamese native goats' rumen by metagenomic deep sequencing. *Animals* **11**: 3257.
10. Bauer TL, Buchholz PCF, Pleiss J (2020) The modular structure of α/β-hydrolases. *FEBS J* **287**: 1035.
11. Danso D, Chow J, Streit WR (2019) Plastics: environmental and biotechnological perspectives on microbial degradation. *Appl Environ Microbiol* **85**:e01095-19

PAZy: The Plastics-Active Enzymes Database



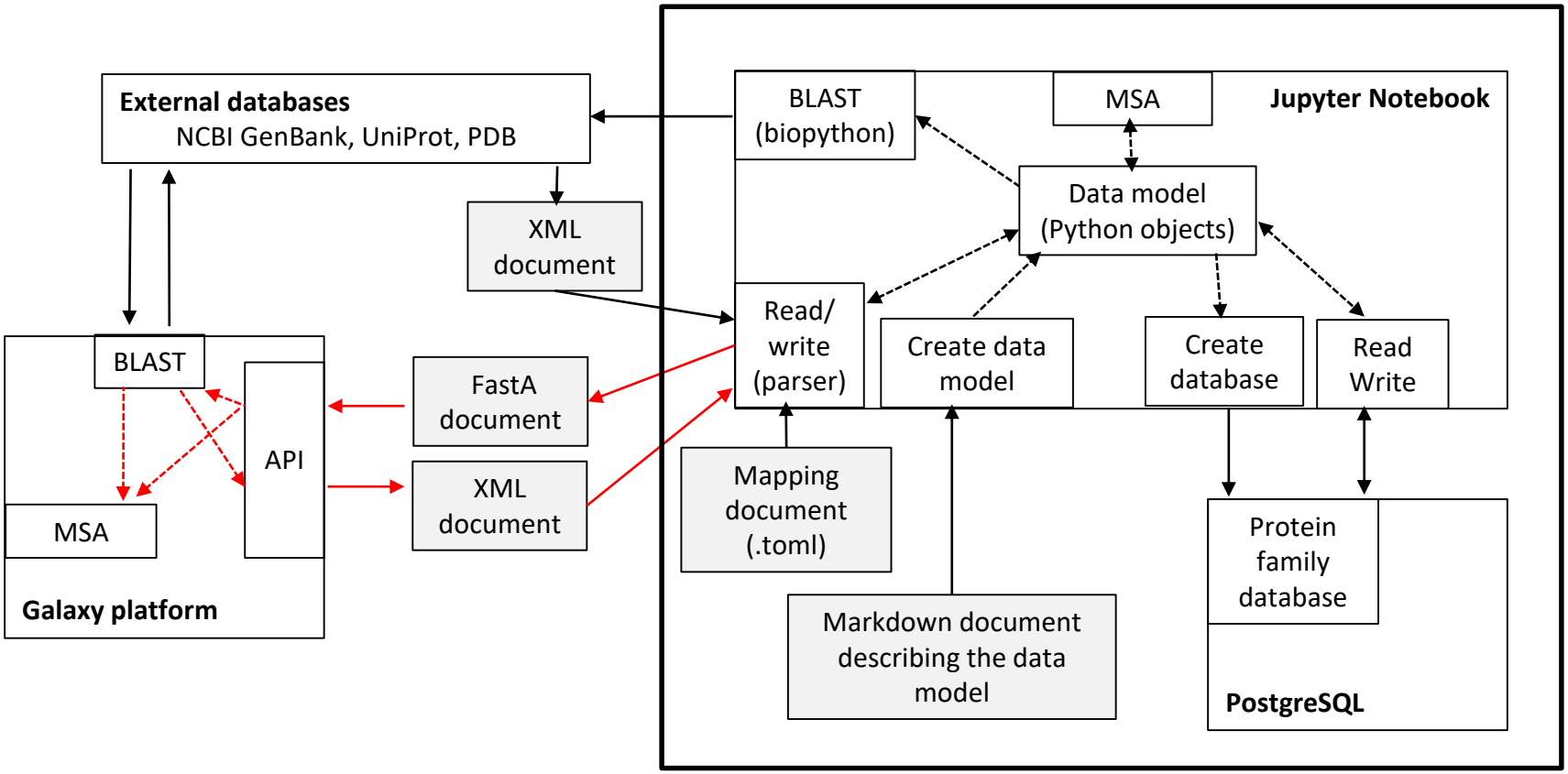
Buchholz PCF, Feuerriegel G, Zhang H, Perez-Garcia P, Nover L-L, Chow J, *et al.*(2022) *Proteins* **90**:1443

Zhang H, Pérez-García P, Dierkes R, Danso D, Pleiss J, Almeida A, *et al.*(2022) *Front Microbiol* **12:** 803896

Chow J, Pleiss J, Streit WR (2022) *BIOspektrum* **28**: 451

https://pazy.eu

2. PyEED: Python Enzyme Engineering Database



https://github.com/PyEED