STRENDA Commission @ Work

Minutes of the

17th STRENDA Meeting

online

September 20 & November 9, 2021

by

Dr. Carsten Kettner

Beilstein-Institut, Trakehner Str. 7 – 9, 60487 Frankfurt am Main, Germany

Table of Contents

List of Participants

- Athel Cornish-Bowden, Marseille, France (ACB)
- Paul F. Fitzpatrick, University of Texas Health Science Center at San Antonio, TX, USA (PF)
- Robert Giessmann, Bayer AG, Berlin, Germany (RG)
- Robert Goldberg, NIST, Biosystsems and Biomaterials Division, Gaithersburg, MA, USA (BG)
- Peter Halling, University of Strathclyde, Glasgow, UK (PH)
- Thomas S. Leyh, The Albert-Einstein-College, Bronx, NY, USA (TL)
- Jürgen Pleiss, University of Stuttgart, Germany (JP)
- Frank M. Raushel, Texas A&M University, College Station, TX, USA (FM)
- Johann M. Rohwer, University of Stellenbosch, South Africa (JR)
- Dörte Rother, Research Center Jülich, Germany (DR)
- Santiago Schnell, University of Notre Dame, IN, USA (SS)
- Neil Swainston, The University of Liverpool, UK (NS)
- Ming-Daw Tsai, Academia Sinica, Taipei, Taiwan (MDT)
- Hans V. Westerhoff, Universities of Amsterdam, The Netherlands (HW)
- Ulrike Wittig, Heidelberg Institute for Theoretical Studies, Germany (UW)
- Roland Wohlgemuth, Lodz University of Technology, Poland (RW)
- Carsten Kettner (co-ordination), Beilstein-Institut, Frankfurt am Main, Germany (CK)

Unfortunately absent (with excuse)

- Barbara M. Bakker, University Medical Center Groningen, University of Groningen, The Netherlands
- Claire O'Donovan, EBI, Cambridge, UK

Agenda

Monday, 20 September

Topics
Opening and welcome Proposal of the agenda Technical introductions (Carsten Kettner)
STRENDA Overview • Welcome new Commission member • past activities (Carsten Kettner)
STRENDA DB • Technical implementations, status and perspectives (Andrew McDonald)
Proposal - Thermodynamic data • Extension STRENDA DB Guidelines (Peter Halling, Bob Goldberg)
General Discussion

Tuesday, 9 November

Topics
Opening and welcome Proposal of the updated agenda (Carsten Kettner)

Topics
Follow up: Thermodynamic data – reporting equilibrium constants • Extension of the STRENDA Guidelines • additional documents (Peter Halling)
Data exchange format – EnzymeML • progress, current status and future perspectives • consequences for STRENDA DB (Jürgen Pleiss)
Reporting of kinetic models, the modeling process and model's quality • overview and introduction • proposal for the extension of the STRENDA Guidelines (Santiago Schnell)
Discussions and decisions on the above topics (All)

1 General

1.1 Stackfield

According to the Commission's consent, Stackfield (www.stackfield.com) is still open to be used as a platform for the STRENDA Commission to share files, discuss results obtained in the sub-groups and to recall past information. You may log on the platform using your username and password chosen for the first log in. If you login attempt fails due to any reason, please let me know.

1.2 Appointments / Memberships

Dr. Meina Neumann-Schaal, Leibniz Institute DSMZ, Braunschweig, Germany, was appointed to the STRENDA Commission. She accepted the invitation. @Meina: Welcome to the Commission!

2 Overview

2.1 Activities 2020/2021

CK gave a brief overview of the past activities run by and related with the STRENDA Commission. He also referred to some activities started in the previous year:

- due to the pandemic, CK did manage to attend only a very few events to promote STRENDA DB, among these were a webinar talk given with the European Section of Applied Biocatalysis, short presentation at the eScience conference, and a lecture at the Open Science Days.
- Activities with relevance for STRENDA:
  - partner of the German **NFDI4Chem** project funded by DFG, STRENDA DB plays an exemplary role for the infrastructure required in chemistry and biochemistry, In addition, CK is involved the development of standards and metadata for chemistry,
  - CK is representative in the **TIB Consortium of the DataCite** initiative, which is helpful to bring in additional views on e.g. FAIR metadata to be delivered to DataCite,
  - CK has initiated an interest group of **GoFAIR Biocatalysis**. The idea is to create an implementation platform for biocatalysis FAIRdata. As STRENDA DB is starting to become FAIR for computers and humans but more efforts need to be spent.
  - Advisory role of CK terminated at BioRoboost, a standardization initiative for synthetic biology. This EU funded project has ended in September 2021. CK took part in only a few of the meetings but was able to bring in some of his expertise and to make this community aware of the much work already done by STRENDA.
  - Invitation to the DataFAIRy project (standards for bioassay reporting), a project of the Pistoia Alliance. CK agreed to take part as a guest.
- Beilstein Webinars (Beilstein Talks) on enzymology have been started with talks by Robert Goldberg, Robert Giessmann and Ming-Daw Tsai. The audience size was between 40 and 80 participants per lecture. CK repeated his invitation to contribute.
- Wikipedia: after a longer editorial process, there is now an article on STRENDA published. This is also part of the Minimum information standard section. STRENDA can be found at: https://en.wikipedia.org/wiki/Standards_for_Reporting_Enzymology_Data. This page will be permanently revised and modified.

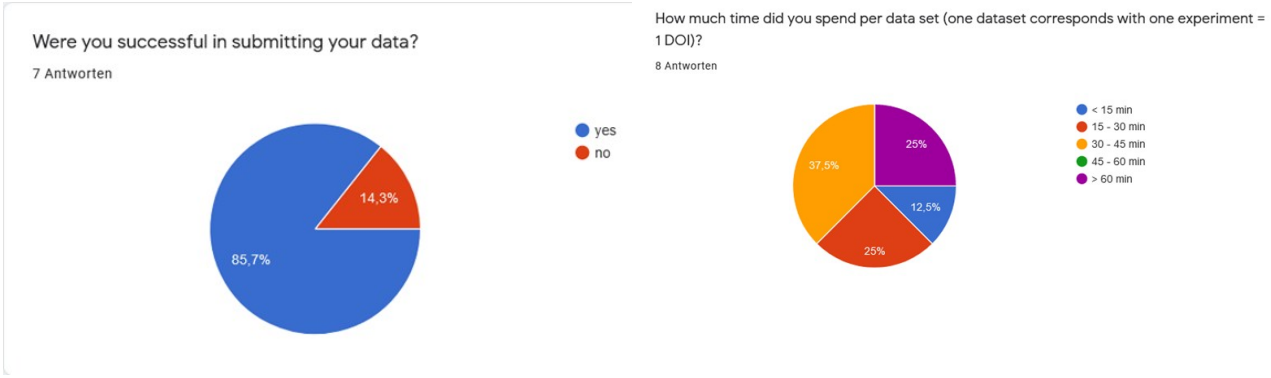
2.2 Proposal for increasing the visibility of STRENDA DB

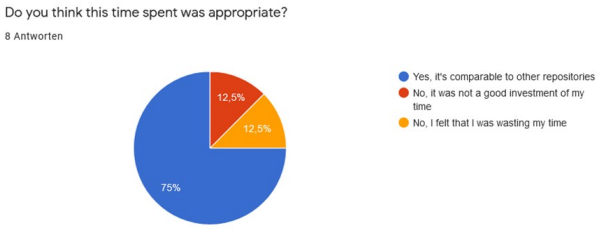
R. Giessmann suggested adding STRENDA Reference as a mechanism of identifying who is using our Guidelines. Guidelines should contain text requesting that author place reference in their paper. Suggestion was well received. Santiago… F1000 Research may be a useful place to position the Guidelines for references. Suggestion to request that authors mention STRENDA compliance in Acknowledgements. Goldberg: Mention STRENDA in "archival" literature which would include links to STRENDA Guidelines URL – Keith Tipton and Gerry Moss were mentioned as persons that could do this.

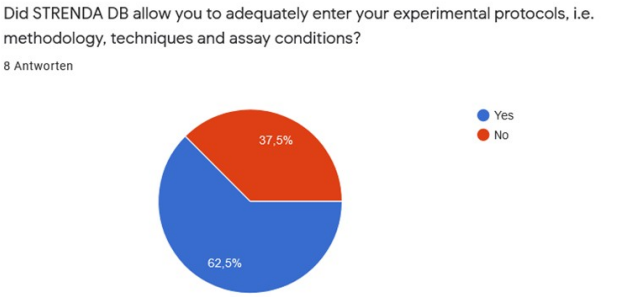
2.3 STRENDA DB

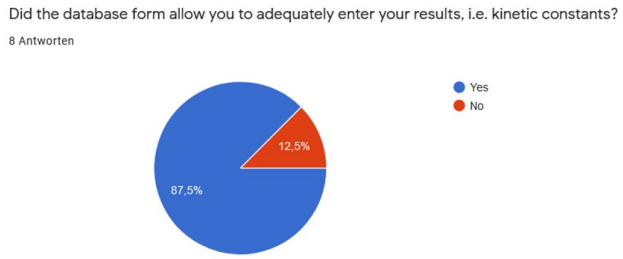
2.3.1 Overview

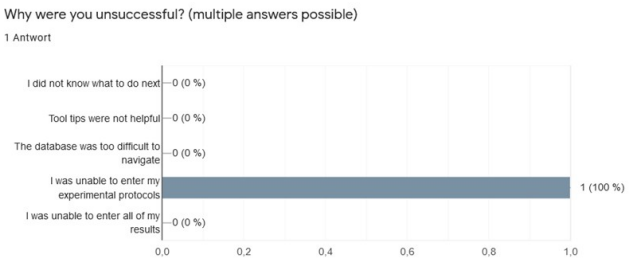
- Status: as of end of September: 71 datasets published, 18 more datasets to be finalized in the pipeline, ~30 datasets started but not completed
- Matomo analytics implemented
- STRENDA DB survey (https://forms.gle/DMYosCgweaQPX54N7): only a few answers available so far:

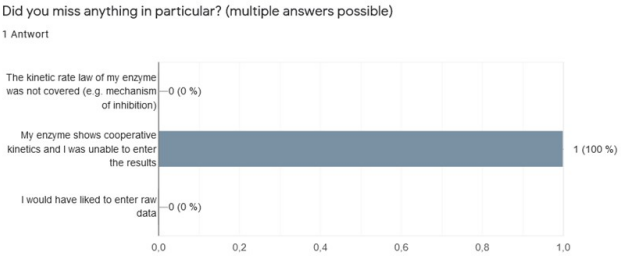












2.3.2 Extension of STRENDA DB

AMD presented his work on the extensions of STRENDA DB and reported about the current difficulties to upgrade the Cayenne Modeller, a developer's framework for the integration and the mapping of the GUI with the database management system. It supports reverse-engineering of RDBMS schemas, editing object-relational mapping projects, generation of Java source code for the persistent objects and other functions. As the upgrade has to be carried out step by step from version 3.0 (last version in use) via 3.1.3 and 4.0.2 to version 4.1 and as each upgrade resulted in a series of (often cryptic) error messages, the upgrade process is time consuming. In addition, a small number of bugs have been identified and fixed. The deployment in the productive system will be carried out very soon.

Lots of details were reported and effort expended. NS codifies problem: the original STRENDA DB system assumed that substrate and products are small molecules, but the upgrade could incorporate that substrates could be proteins.

However, this is part of the next steps of the extension when the macromolecule section (part of the definition of the assay conditions) are made more user friendly.

Tasks	
Set up of infrastructure (GitHub, Andrew's system, etc.)	✓
Administration	
• number of chars in manuscript title field	✓
• Order of buttons	✓
• ORCID in user profile	✓
• linked DOI and PMID	✓
Experiment Data	
• Correction of XML file and query table	✓
• additional units in pull down menu,	✓
• integer digits (e.g. pH)	✓
• Granulation of methodology section	
• Macromolecules assigned to determine kinetic parameters	

The last two points (granulation of methodology section, macromolecules assigned to determine kinetic parameters) are moved into phase 2 of the development.

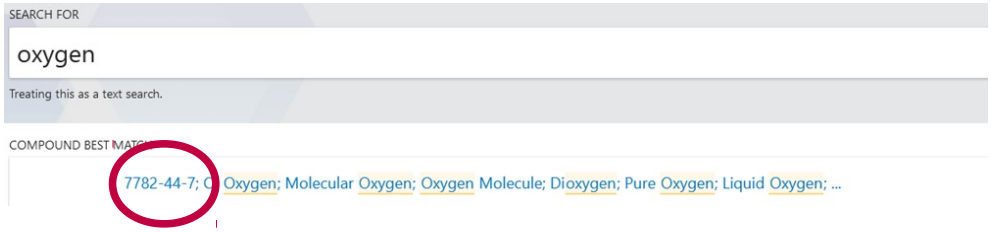
2.3.3 Syntax problem for compounds identified

PH identified a problem with chemical compound names in STRENDA DB:

Query in PubChem	STRENDA DB entry
nitrocefin	UNII-EWP54G0J8F
Oxygen	7782-44-7
hydrogen	1333-74-0
Tris (Tromethamine)	Trometamol
D(+) Glucose	D-Glc

Obviously, PubChem has reordered the names of synonyms of chemical compounds in a way that CAS numbers (see for oxygen) or odd abbreviations (see for glucose) appear at first and then will be retrieved by STRENDA DB using the request syntax.

Example:



This bug has been resolved by replacing the SynonymList (previously used) by the MeSHHeadingList of PubChem which is the equivalent of the title. This bugfix will be implemented in the next minor release.

The problem of compound identifiers was discussed at length – NS made the comment that we distinguish between what the computational expert and bench scientist needs to know.

3 Recommendations for reporting measurements of K_{eq} of enzyme reactions

On the previous STRENDA meeting, PH and BG reported on the requirement of including additional parameters in the STRENDA Guidelines by showing examples.

Difficulty in accessing NIST database, which BG estimates contains ~100 constants. Suggestion to emphasize the relationship between the Haldane Relationship and the overall equilibrium constant for an enzyme catalysed reaction. There was a general consensus that this is an issue for STRENDA to take over.
TL raised the question of whether the current “thermodynamics text” is intended for an expansion of the STRENDA Guidelines, an independent paper, or as a web-site recommendation from the Commission. This question has been addressed by the preparation of the following four documents by the working group (PH, BG, HW, RG and CK):

1. revised version of the STRENDA Guidelines with general edits,
2. an updated version of the STRENDA Guidelines that includes the reporting of equilibrium data,

3. on the basis of the Alberty et al., 2011 paper RG created a new version of recommendations on the measurement of equilibrium data, a so called short version which has been discussed within the task group intensively,
4. HW drafted recommendations for best practice experiments, in particular for systems biology.

For the latter two, both welcome any support and input from the Commission.

The updated versions of the STRENDA Guidelines were discussed. JR raised the question of whether and how to include equilibrium constants predicted by rate data (e.g., the Haldane Equation). It was agreed that a section addressing this issue would be pursued.

BG raised the issue of which symbols should be used to indicate an equilibrium constant – K' or K_{eq}. The use of the word "canonical" was discussed. Is it ambiguous? Perhaps "balanced," along with an appropriate reference to what that means, is preferable. It was agreed that concerns regarding systematic errors should be mentioned, but perhaps not explored in detail. The question of whether the experimentalist should have to justify why at a particular model was chosen to fit their data.

Most of the suggestions made in the Guidelines have been approved. There were only a few parameters for which an improved and more precise wording was demanded. PH took over this task.

ACTION ITEMS (all):

- Review of the updated STRENDA Guidelines, comment Peter's suggestions for alternative wording.
- Contact BG and/or HW with regards to the recommendation documents.

4 EnzymeML – data exchange format and proposal

JP presented the progress of the development of EnzymeML:

- the "philosophy" - a major objective is to unburden the experimentalist, and enhance, not restrict, creativity, to enable researchers to store data in a standardized format early in the data life cycle, to provide adaptable workflows that implement best practices on data reporting and modelling, scalable, including feedback loop for the design of experiments, promoting novel approaches by providing an adaptable platform,

- publicizing EnzymeML in a variety of ways: publications, posters, presentations,
- presentation of some example of data acquisition and analysis.

JP presents the needs for a data exchange format which can be implemented as markup language. Various workshops, hackathons and biweekly meetings resulted in further development of EnzymeML, SABIO-RK became part of the project, and the project has been registered with FAIRDOMHub. There is also an API based on Python and Java available which allows reading and writing of EnzymeML files.

EnzymeML's next steps: more use cases, more platforms. Demonstrate extension to enzyme cascades. Whole cell biocatalysts? Metadata about assays system such as flow versus batch reactor.

Discussion:

HW asked what the time frame will be for the implementation of the EnzymeML file upload functionality in STRENDA DB – first steps will be part of Phase 2 of the extension project.

NS raised the question of whether STRENDA DB should be positioned as the back-end data repository of EnzymeML – it will be one of the applications for import and export of enzyme data.

UW raised the question of which STRENDA Recommendations are required for data entry into STRENDA DB – this initiated a discussion on the differences between the Guidelines and the data fields in the database on the one hand and on EnzymeML on the other other hand.

Why extending STRENDA DB, aren't the rate parameters already included in the database sufficient? - no, the community asks for more and additional data fields such as for data on progress curves, time course data etc.

The less parameters are requested in the Guidelines the larger is the variability of kinetic parameter data in the database; if one follow guidelines you'll diminish the variability of database.

RW: contact with IBM in Switzerland, they seems to be interested in STRENDA. Could IBM become partner of STRENDA? - Maybe. Definitely yes for the data exchange format EnzymeML but rather no for STRENDA DB. It is suspected that IBM is working on the development of AI (as many of the technology companies are) and thus they'd be interested in the data rather than in the project. NS agreed but commented that AI research relies on huge amounts of data and thus STRENDA DB would not be an interesting data source. TL mentioned that he's also in contact with an individual at IBM's research centre in New York and he is willing to ask his contact person on their interest in STRENDA.

ACTION ITEM(s):

- none as the development and implementation of EnzymeML will be continued with STRENDA and along with other biocatalysis groups performing experiments and developing tools.

5 Proposal: Extension of Guidelines by protocols

SS gives a brief presentation on his analysis of kinetic rate parameters to be found in the literature. The result of his analysis is an undesirable heterogeneity of kinetic data in terms of kinetics to be found over several orders of magnitude even though those kinetics have been obtained from same experimental set ups. Obviously, there seems to be biases in the design of experiments (details not fully stated) and data reporting (not any detail considered worth to be reported). In addition, in the reports the information on replicability (same team) vs. reproducibility (different teams) often is missing and this may cause people not to trust each other's results.

SS wonders whether the precision with experiments could be improved if they'd follow the STRENDA Guidelines. He proposes (again) the extension of the STRENDA Guidelines with statements on the availability of raw data, the inclusion of detailed protocols, and the description of software or algorithms for enzyme kinetic experiments. In conclusion, the time is ripe for research practices in order to improve transparency and reproducibility.

A longer discussion followed this presentation, in particular on the findings and its conclusions. In addition, the role of the Guidelines were discussed, i.e. best practices recipe vs. reporting guidelines. There was general consensus that the STRENDA Guidelines are reporting guidelines whose goal should be reproducibility - "not all details provided but the most essential information should be given" (MDT).

However, in this vein the Guidelines need another extension with regards to the reporting of measurement protocols and for models used and equations applied to obtain the kinetic data.

ACTION ITEM(s):

- SS will draft an extension of the Guidelines. Any assisting hand is highly appreciated on this. CK would be happy to organize subgroup meetings to discuss this task.

CK thanks TL for taking the notes that are basis for these minutes.