




STRENDA Commission @ Work

**Minutes of the**

**16th STRENDA Meeting**



**online**

**October, 20 & 21, 2020**

by

Dr. Carsten Kettner

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

# Table of Contents

List of Participants.....	3
Agenda.....	4
Tuesday, 20 October.....	4
Wednesday, 21 October.....	5
1 General.....	6
1.1 Stackfield.....	6
1.2 Appointments / Memberships.....	6
1.3 Polls.....	6
2 Overview.....	7
2.1 Tasklist 2019/2020.....	7
2.2 STRENDA DB.....	8
2.2.1 Overview.....	8
2.2.2 Extension of STRENDA DB.....	9
3 Recommendations for reporting measurements of $K_{eq}$ of enzyme reactions.....	10
4 EnzymeML – data exchange format and proposal.....	10
5 Proposal: Extension of Guidelines by protocols.....	12
6 Outspread.....	14
6.1 Credits.....	14
6.2 Webinar.....	14

## List of Participants

- Barbara M. Bakker, University Medical Center Groningen, University of Groningen, The Netherlands
- Paul F. Fitzpatrick, University of Texas Health Science Center at San Antonio, TX, USA
- Robert Goldberg, NIST, Biosystems and Biomaterials Division, Gaithersburg, MA, USA
- Peter Halling, University of Strathclyde, Glasgow, UK
- Thomas S. Leyh, The Albert-Einstein-College, Bronx, NY, USA
- Jürgen Pleiss, University of Stuttgart, Germany
- Frank M. Raushel, Texas A&M University, College Station, TX, USA
- Johann M. Rohwer, University of Stellenbosch, South Africa
- Santiago Schnell, University of Michigan, Ann Arbor, MI, USA
- Neil Swainston, The University of Manchester, UK
- Ming-Daw Tsai, Academia Sinica, Taipei, Taiwan
- Hans V. Westerhoff, Universities of Amsterdam, The Netherlands
- Ulrike Wittig, Heidelberg Institute for Theoretical Studies, Germany
- Roland Wohlgemuth, Lodz University of Technology, Poland
- Carsten Kettner (co-ordination), Beilstein-Institut, Frankfurt am Main, Germany

## Agenda

### Tuesday, 20 October

Topics
<b>Opening and welcome</b> Proposal of the agenda Technical introductions (Carsten Kettner)
<b>STRENDA Overview</b> <ul style="list-style-type: none"><li>• Proposal: appointment A. McDonald, Dublin, Ireland</li><li>• Task list</li><li>• draft: Acceptance letter from journal</li><li>• Wikipedia Article</li><li>• Promotion (STRENDees, NFDI4Chem, DataCite...)</li><li>• GO FAIR Biocatalysis</li><li>• Progress with STRENDA DB – techniques and numbers</li><li>• Survey after data input</li><li>• EuropePMC – links from abstract to dataset</li></ul> (Carsten Kettner)
<b>STRENDA DB</b> <ul style="list-style-type: none"><li>• Technical implementations, status and perspectives (Andrew McDonald)</li></ul>
<b>Proposal - Thermodynamic data</b> <ul style="list-style-type: none"><li>• Extension STRENDA DB Guidelines (Peter Halling, Bob Goldberg)</li><li>• Extension STRENDA DB (Carsten Kettner)</li></ul>
General Discussion

## Wednesday, 21 October

Topics
<p><b>Come together</b>, opening and reconciliation of yesterday's discussions and decisions made (Carsten Kettner)</p>
<p><b>Data exchange format – EnzymeML</b></p> <ul style="list-style-type: none"><li>• progress, current status and future perspectives</li><li>• FAIRdom</li><li>• enzymeml.org</li></ul> <p>(Jürgen Pleiss, Santiago Schnell, Neil Swainston)</p>
<p><b>Proposal – extension of STRENDA Guidelines</b></p> <ul style="list-style-type: none"><li>• reporting of kinetic models, the modelling process and model's quality (Santiago Schnell, Jürgen Pleiss)</li></ul>
<p><b>Discussions</b> and decisions on the above topics (All)</p>
<p><b>Outspread</b></p> <ul style="list-style-type: none"><li>• stocktake of which journals currently recommend STRENDA and STRENDA DB,</li><li>• integration of reporting standards in education</li><li>• Strategy (again)</li><li>• increasing the data rate in STRENDA DB</li><li>• increasing the acceptance of journals to encourage the use of STRENDA DB</li><li>• what is missing?</li></ul> <p>(All)</p>
<p><b>Webinar</b></p> <ul style="list-style-type: none"><li>• infrastructure provided by Beilstein</li><li>• planning for the upcoming weeks and months</li></ul> <p>(All)</p>
Consensus on task list for 2020/2021

# 1 General

## 1.1 Stackfield

According to the Commission's consent, Stackfield ([www.stackfield.com](http://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.

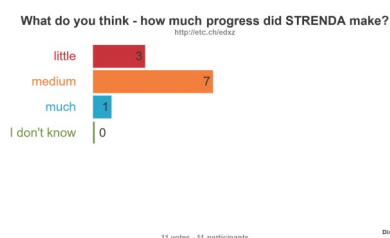
## 1.2 Appointments / Memberships

Dr. Andrew McDonald, former Trinity College Dublin Ireland, was appointed to the STRENDA Commission. He accepted the invitation.

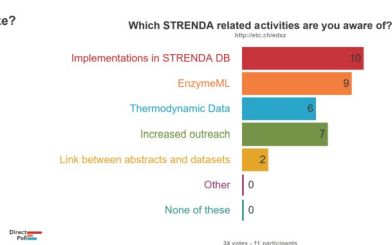
Prof. Antonio Baici (University of Zurich) and Prof. Dietmar Schomburg (TU Braunschweig) resigned from the STRENDA Commission due to their retirements. Both expressed their will to advice in the future whenever necessary. Due to its role with the BRENDA database, a replacement for Dietmar Schomburg was discussed but a decision has been postponed.

## 1.3 Polls

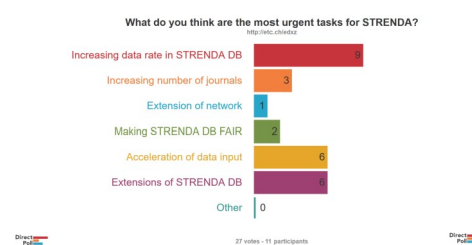
As a kind of a warm up, a series of polls has been carried out with the Commission. Here, the most relevant results will be shown:



Poll #1



Poll #2



Poll #3

The majority of the Commission is convinced that the project has made progress since the previous meeting at a medium degree (poll #1). The members of the Commission are pretty well informed about the diverse activities within the project (poll #2) but the fact that there is no 100% awareness

reached may mean that the members need to be better informed by CK over the year. The majority has identified the data rate in STRENDA DB, the acceleration of the data input and the extension of STRENDA DB has questions of concern (poll #3).

## 2 Overview

### 2.1 Tasklist 2019/2020

CK presented the task list as agreed to on the previous meeting and went through the items:

- due to the pandemic, CK did manage to attend only a very few conferences to promote STRENDA DB
- there is not yet any teaching or educational material available on the STRENDA webpages
- Acceptance letter from journal which recommends authors to contact STRENDA DB to submit manuscript data to database. Proposal from PF which is also implemented for ABB:

“Submission of enzyme kinetic data. ABB follows the STRENDA guidelines for publication of enzyme kinetic data. Since your manuscript contains such data, we invite you to submit your data to the STRENDA DB at <https://www.beilstein-strenda-db.org/strenda/index.xhtml>. Doing so will generate a pdf document summarizing the data. This document can be included with your published manuscript as supplementary information if desired. It will also be assigned a DOI number describing the data once the manuscript is published, providing a citable document for future manuscripts.”

- Upload published data: 30 datasets have been entered by a student of SS, however, only 50% out of these are considered to qualify for further efforts because only these contain STRENDA relevant data. More data will be delivered from FR.
- Journals: RW brought STRENDA up at the last Frontiers global board meeting. Frontiers is interested and invited the Commission to prepare a manuscript for one suitable journal and to include into author instructions. As the publisher has a lot of diverse journals in its portfolio, the Commission decided first to go with *F. in Bioechnol. Bioeng.*, *F. in Cell and Dev. Biol.*, and *F. in Molecular Biosciences*. In the meanwhile, RW and CK sent Frontiers a selection of Frontiers journals for which the STRENDA guidelines may be relevant.

*Biotechnol. Journal*: board meeting was supportive, CK has sent out additional material but did not receive any response yet.

*Biotechnol. & Bioeng.* Recommends STRENDA guidelines after BG has persuaded them when reviewing a paper and found many missing pieces of information.

- Advertisement video: this video hasn't been produced but tutorial videos are already available.
- Google Search: SEO check has been carried out, the STRENDA webpage has been modified accordingly to allow this site to be better findable by search machines.
- 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](https://en.wikipedia.org/wiki/Standards_for_Reporting_Enzymology_Data).
- General promotion: CK looked at the Commission's members' webpages and found that the papers on STRENDA are well cited, and some members have some info on STRENDA included on their webpages. However, the members are asked to do more.
- Activities with relevance for STRENDA DB: partner of the German **NFDI4Chem** project funded by DFG, CK is representative in the **TIB Consortium of the DataCite** initiative, 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.

## 2.2 STRENDA DB

### 2.2.1 Overview

- Status: as of end of September: 61 datasets published, 11 more datasets to be finalized in the pipeline, ~30 datasets started but not completed
- Matomo analytics implemented
- STRENDA DB survey (created by JR) – highly appreciated by Commission. Users whose datasets have been published in STRENDA DB are asked by CK to participate in this survey.

Suggestion (I): to put a link on the STRENDA DB home page because it is considered valuable to also learn about the reasons for failing rather than from successful entries.

Suggestion (ii): send survey to all 134 registered users to get more info.

**Decision:** Commission looks at survey and makes suggestions for improvement. JR will collate and implement.



- EuropePMC – abstracts in EuropePMC link to datasets in STRENDA DB. Implementing these links is very time consuming on both sides as on our side special xml files need to be generated and on the EuropePMC side this information need to be added to the current datasets. An additional issue arises from the fact that EuropePMC can only accept papers indexed by PMID. UW suggested to do the same with PubMed as the general structure is the same.

### 2.2.2 Extension of STRENDA DB

CK presented an extension project which is carried out together with AMD who will make the software development. The project is divided into three phases with increasing degree of complexity. At the time of this meeting, more than the half of the tasks have been finalized, tested and approved.

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

The specification of phase 1 is available on request, those for the phases 2 and 3 are currently in work. JP asked for the data model and would like to see the plans because he sees some implications for the interaction between STRENDA DB and the development of EnzymeML.

NS wondered about the flexibility of the system with regards to extensions such as thermodynamic data.

HW expressed worry about the openness of the code in terms of understandability to third party programmers. As Java and Oracle are standard tools at Beilstein-Institut and as the code is well commented, AMD did not see any reasons of concern.

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

- PH and BG report on the requirement of including additional parameters in the STRENDA Guidelines by showing examples.
- thermodynamic data; this is  $K_{eq}$ s not calorimetric data. Follow IUMB guidelines, ref. to Alberty et al. 2011
- consensus that this is an issue for STRENDA to take over,
- separate set guidelines for reporting thermodynamic data (as defined above) → integration in STRENDA Guidelines → implementation in STRENDA DB,

**Decision:** PH takes lead, BG, HW forming task group to work on this. A draft will be presented and discussed at the next meeting in 2021.

### 4 EnzymeML – data exchange format and proposal

As this was a longer discussion and for a better understanding of the future development, here is a summary of the discussion:

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.

Challenge is to get consistency between STRENDA guidelines and EnzymeML. EnzymeML is based on SBML and extended by STRENDA parameters. There is overlap between contents and focus. STRENDA DB takes only MM as a kinetic model. Enzyme ML rather provides time course data, is limited in metadata and can use any enzyme kinetic model. Storage conditions, purity etc are not included yet. SBML description of kinetic laws is used by EnzymeML. Products, kinetic model etc. are included.

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.

Example: Emanuele & Fitzpatrick (1995) STRENDA entertains only MM Model. If there are two substrates one has difficulty reporting. This may cause difficulty in people uploading their data. FR comments on the example (which he coauthored): we discussed about this years ago, but it was thought that 90% of the data in the literature only vary one substrate and have the other substrate saturating. FR agrees that this should work for 95% of the users. JP has observed examples where people failed to upload their data due to this problem.

### **Discussion:**

HW asks how to enter product inhibition? It is crucial for systems biology to study the effects of downstream metabolites.

JP: STRENDA should also suggest people what to do better in terms of data collection, e.g. substrate inhibition, two substrates.

PF: we wanted to just capture the data; has been the goal all along.

SS: e-Labnotebooks are being used more and more; we need to prepare for the future and therefore have guidelines also for the future.

PH: has argued for a long time that in STRENDA DB you should state the equation you have used to fit your data.

NS: we have been discussing these things before; what is keeping us from doing this?

JP: part of the problem is that we only ask our students to upload their data at the end of their PhD thesis.

HW: is there any technical problem preventing STRENDA DB to entertain multiple substrates, products.

CK: No, not anymore; now there is a developer.

HW: Should we form a subcommittee also here (supported by JP).

CK: many things are already included in STRENDA DB; but not (yet) in STRENDA Guidelines. There are 2 major points not yet included: (i) go beyond  $K_M$  and integrate time course data and (ii) include the concept of the modifiers as suggested by TB. Advantage is also the broadening the audience, e.g. by systems biologists, clinical enzymologists, applied bioengineers. TB is willing to collaborate and help insert data. He has a database on enzyme modifiers.

JP and HW: 'Modifiers' should then include products and co-substrates.

CK: Guidelines indeed need to be reviewed.

JR: equation issue was one of the questions asked in the survey. Most people reported that in papers they report the equation used to infer the  $K_M$  etc.. Those data cannot be inserted into the database at the moment.

**No decision** taken as the development and implementation of EnzymeML has been previously approved and will be continued.

## 5 Proposal: Extension of Guidelines by protocols

SS proposes the extension of the STRENDA Guidelines to cover protocols. In the US 44% of labs use an electronic lab book of some kind. STRENDA still lacks standardization of the measurement protocols. Various published studies (his and Peter's) on relevance of steady state type assumptions and on how time course experiments need to be used, e.g. where and how to measure 'initial rates'. Should one use slope, or two or three time points; nonlinear regression. Report on replicability (same team)/reproducibility (different teams) as often missing and this may cause people not to trust each other's results.

Conclusions: We need to:

- develop STRENDA guidelines for reporting measurement protocols and for modelling protocols.
- prepare for our STRENDA data moving seamlessly into STRENDADB from labbooks.

### Discussion:

CK: what should be reported for measurement protocol.

SS: e.g. at which time point the initial rate is evaluated.

RW: minimum requirements needed for reporting replicability, e.g. vis-à-vis bad quality of substrate such as glycerol batches containing aldehyde inactivating enzyme. STRENDA DB should have some modular way to cover >95% of global applications; make software such that it starts with simple but then has ability to open an extra window for more complete reporting.

SS: We need to become more physicist, less biochemist.

FR:  $V_{max}$  or  $K_M$  is not equivalent to Planck's constant.

NS tries to define a roadmap. Existing STRENDA DB, then kinetics, then more understanding of the quality of the data.

---

AMD: new extensions should not break the existing structure; if possible increment. Much is done already EnzymeML (e.g. complex equations).

JP sees 3 different areas: language, database and model, but we need guidelines also to teach our students how to do enzyme kinetic studies. Go back to the guidelines and set up a working group to address how to modify the guidelines.

CK agrees to go back and start finding out new recommendations and see what is acceptable for the community. STRENDA DB is modular so this should be doable. It is more challenging to guide the users carefully through the system.

TL (who just joined) agrees on extending the guidelines. What remains an issue is extending the usage the adoption of STRENDA.

JP recognizes a role of these guidelines in terms of improving science.

MDT: we need to face the issue of uniformity versus diverse reality; we have the problem of the paradoxon of unifying a system that should be diverse. On the other hand, STRENDA could indeed serve to improve the science here.

CK sees that there could be more progress if we widen the community and is therefore agreeing with extending the guidelines. Also, it will be good to serve making the datasets more complete, so that they can be more efficiently used by others. Existing guidelines are a good blueprint. In addition, users of STRENDA DB are often different from the people entering data into it. He supports the idea of creating a task group.

SS also sees STRENDA to make a goldbook of enzymology.

TL proposed the idea to bring in experts in kinetic isotope effects; he thinks that there will be lots of areas to include on such a goldbook.

RW: keep simple so that activation barrier for entering data remains low. So differentiate the essential from the nice.

PF thinks that it is worth revising the guidelines, but STRENDA DB should be considered separately from the STRENDA guidelines. It will be doable to update the guidelines.

UW comments that the STRENDA DBb would be considered as a validation tool of adhering to the guidelines.

NS: Guidelines could drive adjustments in STRENDA DB.

CK: We could keep the guidelines as they are and add a separate document for thermodynamics, which could then mature and later be integrated into STRENDA guidelines and later into STRENDA DB.

---

RW: Is thermodynamic database at NIST maintained? CK: No. He has been trying to contact them. HW is worried that the NIST data gets lost. UW has checked this database, but there is no way to export those data.

**Decision:** create a second task group with JP, JR, TL and CK chaired by SS.

## 6 Outspread

Today, 55 journals recommend their authors to consult the STRENDA Guidelines. More than 10 journals recommend their authors to enter their data into STRENDA DB. In addition, STRENDA DB is listed as recommended repository by other gathering databases. Now it needs to go deeper into the research community. The STRENDA DB website documents which data can be uploaded, but still people try and then fail.

How to increase the data rate? EnzymeML may be a good means.

### 6.1 Credits

UW: idea to give credit to the people who have uploaded; there is a possibility in ORCID to have contributions not just papers mentioned. However, STRENDA DB has to register with ORCID so that this information will be included into the ORCID record.

UW suggests to hold courses to practise the data upload into STRENDA DB. This is done with SABIO-RK, but these people are often modellers.

RW: these would have to be advanced courses. Perhaps in a Beilstein webinar?

### 6.2 Webinar

CK proposes webinars as 20 minutes lectures, followed by QA session, on a regular basis in the late afternoons, up to 500 participants possible.

UW comments that it will be crucial to get the right audience (experimentalists).

First pipeline will be made available on Stackfield.

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