

Minutes of the





Meeting

October 5 & 6, 2022

Hotel Jagdschloss Niederwald Rüdesheim, Germany

by
Dr. Carsten Kettner
Beilstein-Institut, Trakehner Str. 7 – 9, 60487 Frankfurt, Germany

based on the notes of Stephan Malzacher, Research Center Jülich, Germany



Table of Contents

List of Participants	3
Agenda	4
Wednesday, October 5	4
Thursday, October 6	
STRENDA DB: Overview	5
Technical Development of STRENDA DB	
STRENDA Guidelines V.1.8	7
M-CSA and EBI enzyme resources	7
Extending the STRENDA Guidelines – Reporting Guidelines for Biocatalysis Experiments.	8
EnzymeML – Status, progresses and perspectives	9
Recommendations for Reporting Measurements of Equilibrium Constants of Enzyme	
Reactions	10
Proposal: Paper on the underpinning thermodynamics	11
New thermodynamics tables	12
Perspectives on and for STRENDA	13
Results	
Group 1	14
Group 2	15
Group 3	16
Announcements	



The 18th STRENDA meeting took place over two days on Wednesday, October 5, 9 am – 5 pm, and Thursday, October 6, 9 am - 5 pm at the Hotel Jagdschloss Niederwald in Rüdesheim, Germany. As some members could not travel, the meeting was hosted as an hybrid event allowing the remote members to participate and contribute to the discussions.

The material presented at this meeting is deposited on Google Drive at:

https://drive.google.com/drive/folders/1Is3O2ArFrEQp8_lqpKvw2hfJ8OiTvn7Q?usp=sharing

List of Participants

- Paul F. Fitzpatrick, University of Texas Health Science Center at San Antonio, TX, USA (online),
- Robert Goldberg, NIST, Gaithersburg, MA, USA (online)
- Peter Halling, University of Strathclyde, Glasgow, UK
- Thomas S. Leyh, The Albert-Einstein-College of Medicine, Bronx, NY, USA
- Stephan Malzacher, Research Center Jülich, Germany
- Andrew Mcdonald, Trinity College Dublin, Ireland (online)
- Meina Neumann-Schaal, Leibniz Institute DSMZ, Braunschweig, Germany
- Jürgen Pleiss, University of Stuttgart, Germany
- Frank M. Raushel, Texas A&M University, College Station, USA
- Antonio Ribeiro, EMBL Outstation, European Bioinformatics Institute, Hinxton, UK
- Johann M. Rohwer, University of Stellenbosch, South Africa
- Katrin Rosenthal, Jakobs-University Bremen, Germany
- Ming-Daw Tsai, Academia Sinica, Taipei, Taiwan (online)
- 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, Germany



Unfortunately absent (with excuse)

- Barbara M. Bakker, University Medical Center Groningen, University of Groningen, The Netherlands
- Athel Cornish-Bowden, CNRS-BIP, Marseille, France
- Claire O'Donovan, EMBL Outstation, European Bioinformatics Institute, Hinxton, UK
- Santiago Schnell, University of Michigan, Ann Arbor, MI, USA

Withdrawals from the Commission

Neil Swainston, formerly University of Liverpool. As Neil has left academia and his new position does not allow his further engagement he decided to step back but he emphasized that he will be stay in contact if the Commission asks for his advice. The Commission thanks Neil for his many contributions and hands-on work for the STRENDA project.

Agenda

Wednesday, October 5

Topics	
Opening and Welcome	
Overview	

Proposal of the agenda Technical instructions (Carsten Kettner)

STRENDA Guidelines V. 1.8

Overview of the latest news

(Carsten Kettner)

STRENDA perspectives – stock taking in break-out groups

discussions with leading questions

preparation of the results

M-CSA and EBI enzyme resources

(Antonio Ribeiro)

Extending the STRENDA Guidelines – Reporting Guidelines for Biocatalysis Experiments

(Stephan Malzacher, Katrin Rosenthal and Peter Halling)

Proposal and discussion



Thursday, October 6

Topics

EnzymeML

status, progress and perspectives

(Jürgen Pleiss)

Recommendations on the reporting of equilibrium constants

Proposal for the extension of the STRENDA Guidelines

(Bob Goldberg, Hans Westerhoff, Peter Halling)

Paper on the underpinning thermodynamics (update and upgrade of Alberty et al., 1994), and

New thermodynamics tables

(Hans Westerhoff)

Perspectives on and for STRENDA

Reports from the groups

General topics

STRENDA DB: Overview

The current technical version is 1.3 (from September 26, 2022). The database provides 87 datasets to the public, 18 additional datasets are going to be finalized (assignment to the publication is missing). The current STRENDA DB survey has been answered by 12 participants. They explain their data deposition in STRENDA DB with the pressure by journals that demand the data deposition in a repository or database by the authors.

In addition, some constraints were identified:

- searching for CIDs is not possible not clear what is meant here,
- issues when using Google Chrome needs to be tested, again,
- need to automate the data input process planned with the implementation of an upload function using EnzymeML.

The community is indeed willing to share their data but the deposition process must not take too much time. The average time to enter data in STRENDA DB is between 20 and 45 minutes (depending of the complexity of the dataset).



Technical Development of STRENDA DB

(presented by Andrew McDonald)

Just before the STRENDA meeting, UniProt changed its legacy API for the query of protein data. This failure could be fixed but it is obvious that UniProt is going to get rid of the legacy API and that we need to be aware of this.

AMD presented his first steps towards the granulation of the methodology field which is currently a free-text field. This software task raised a number of questions:

Q: Change of a database schema is challenging: Are the current schemas fixed? How to adapt the old 87 entries to the schema which is now present in version 1.3

A: This could be quite a labour intensive task and depends on the depth of information provided in the datasets before. As the initial free-text field 'methodology' is kept, the information of older datasets is still kept here. For new entries this field will be still available but for data only that could not captured by the fields before.

The flexibility of the free text entry for the methodology is a powerful and useful method for users to enter data in a bunch. However, free text fields are inefficient since the content is very error-prone and disturb the consistency of all data sets. Therefore, the granulation of this section contributes to a standardized description of the methodology.

Q: When UniProt changes, then STRENDA DB would also stop proper functioning – would it be possible to user older versions of UniProt?

A: It is not a matter of versioning of UniProt but a matter of technical access. If the legacy API is stopped by UniProt, any version of UniProt will become inaccessible.

Q: How much time is needed to maintain the changes made by other public databases (e.g. PubChem)?

A: If these databases provide a well-documented API (and PubChem does), it is only a matter of a few hours to reconnect STRENDA DB with PubChem and others.



STRENDA Guidelines V.1.8

Based on the discussions on the 2021 meeting the STRENDA Guidelines were extended by a few recommendations on the measurement of equilibrium constants of enzyme-catalyzed reactions. For this, in List Level 1A, the paragraph "Equilibrium measurements" with four statements was introduced. In List Level 1B (Experimental Results) the section on "Data necessary for reporting kinetic parameters" was extended by equilibrium-measurement related parameters. Finally, a new paragraph was included that refers to data which are required for the reporting of equilibrium measurements.

The Commission approved the new version of the STRENDA Guidelines which is published online at https://www.beilstein-institut.de/en/projects/strenda/guidelines/.

M-CSA and EBI enzyme resources

(presentation by António Ribeiro)

- there is no "one enzmye" mechanism, since enzymes show a mixture of conformations
- due to conformational changes, the rate limiting factor can be found outside of the active site
- many energetic transition states are calculated based on the chemical reaction itself, not regarding conformational changes
- there should be more awareness about the way in which enzyme mechanisms are reported. Standards should be applied when data are published in papers.
- many EC numbers are present, without the respective sequence available.
- the guidelines are about reporting a mechanism and reporting data required to describe a mechanism.



Extending the STRENDA Guidelines – Reporting Guidelines for Biocatalysis Experiments

(presentation by Stephan Malzacher, Katrin Rosenthal, Peter Halling, et al.)

A working group is currently working on the extension of the STRENDA Guidelines towards the inclusion of reporting of data from biocatalysis experiments (as proposed in the paper by Gardossi et al. (2010), DOI:10.1016/j.tibtech.2010.01.001).

This working group includes S. Malzacher, D. Rother, Z. Findrik-Blažević, R. Giessmann, P. Halling, C. Kettner, J. Pleiss, J. Range, K. Rosenthal, P. Wied, R. Wohlgemuth, and J. Woodly.

With regards to the STRENDA Guidelines, in particular those parameters need to be introduced and specified that apply for biocatalysis experiments, i.e. the description of the enyzme in its environment (immobilised or soluble), the reaction system, and the phases in which the enzyme reaction takes place. In contrast to the procedure when the STRENDA Guidelines were created, this extension is generated in a modularized way:

- use cases were determined by the experts in the field for the description of the modules,
- the modularization was directly implemented in a software-based decision tree (see https://github.com/StephanM87/Strenda-biocatalysis),
- an experimental workflow with regards to the data reporting was generated,
- meta-data are defined in parallel to the definition and description of the experimental procedures,

The presentation was followed by an intense discussion on the role of these guidelines and whether they will be incorporated in STRENDA DB. There were some critical voices that demanded for focusing on the current STRENDA Guidelines. However, as these have reached a stable version, revisions may be necessary from time to time. The focus is the complete implementation of the guidelines in STRENDA DB. There was the note that the guidelines are not 100% covered by STRENDA DB and vice versa. The first is a question of time until all parameters from the guidelines will be implemented in STRENDA DB. However, additional efforts are required to specify those parameters in the guidelines that appear very generic. First ideas for this are already available. In some sections, STRENDA DB provides more information than asked in the guidelines, e.g. metadata on compound identifiers (InChI), database identifiers, etc. as these data can be automatically captured from other repositories.



In the end, there was a general agreement to proceed with the development and establishment of the STRENDA Biocatalysis Guidelines as they are considered very important for the data reporting in a sustainable way with regards to the FAIR principles.

EnzymeML - Status, progresses and perspectives

(presented by J. Pleiss)

Philosophy of EnzymeML:

- easy-to-use and interesting tools which increase the efficiency of their research work
- an extensible data model
- adaptable workflows which implement best practices on data reporting and modelling
- scalable workflows to analyse high-throughput experimental data
- structured data for mechanistic and data-driven modelling
- a feedback loop for the design of experiments
- a flexible platform which is not limiting, but which enables and encourages novel approaches
- publication of original and derived data according to the FAIR principles

Data workflow:

- experimental data is captured through ELNs (e.g. Chemotion, Sciformation or openBIS), data platforms such as BioCatHub or simple Excel spreadsheets,
- additionally, experimental data can be modelled using PySCeS or COPASI,
- data from both sources is stored in an EnzymeML document,
- this file can be uploaded into databases and repositories, such as SABIO-RK and STRENDA DB,
- additionally, experimental data can be deposited in EnzymeML Dataverse and is assigned a DOI,
- databases can refer to the data via the DOI



General discussion:

- EnzymeML provides the possibility to validate, if datasets are compliant with standards, for example the STRENDA Guidelines
- Upload of EnzymeML documents is not possible to STRENDA DB at the moment.
- How to guarantee the interoperability with other platforms like Scifinder?!
- Searching for individual experiments stored as EnzymeML documents in Google is at the moment not possible

Plans

- extension of the group of EnzymeML partners in order to create and test use cases for the extension of EnzymeML,
- test of data transfer from the "bench" to modelling platforms,
- connection with electronic lab notebooks developers in order to implement EnzymeML there.
- development and test of data transfers into databases and repositories.

Recommendations for Reporting Measurements of Equilibrium Constants of Enzyme Reactions

(presented by B. Goldberg, group: R. Giessmann, P. Halling, C. Kettner)

Decision from the 2020 meeting:

- thermodynamic data; this is K_{eqs} not calorimetric data. Follow IUMB guidelines and ref. to Alberty et al. 2011,
- consensus that this is an issue for STRENDA to take over,
- separate set of guidelines for reporting thermodynamic data (as defined above), which will be integrated in STRENDA Guidelines → and implemented in STRENDA DB,
- draft will be presented and discussed in the Commission



BG presented the final version of these recommendations. Details on the recommendations were intensively discussed and the results were introduced in the final draft manuscript.

The manuscript has been published recently in the *Beilstein Journal of Organic Chemistry* (https://www.beilstein-journals.org/bjoc/articles/19/26).

Action Point: publish a checklist on the STRENDA website

Proposal: Paper on the underpinning thermodynamics

(update and upgrade of Alberty et al., 1994) (presented by H. Westerhoff)

Decision made in 2020 to produce a paper that presents the description of the underpinning thermodynamics in addition to the Recommendations paper.

Aims of STRENDA is to recommend

- all the relevant parameters to be measured,
- the important enzymes to be measured,
- the conditions to be used.

Motivational questions raised in the presentation:

- Transfer from *in vivo* to *in vitro*? The medium composition is important to compare in vitro to *in vitro*.
- The equilibrium coefficients can be more informational than taking a view on the activity coefficients.
- Is the definition of recommendations in flux analysis coupled highly to the research question?
- See the application not only *in in vivo* application but also in applications like synthesis.
- Non equilibrium thermodynamics in compartmentalized environments.



- the application of thermodynamics in metabolic diseases might also bring benefits in understanding mechanisms in e.g. cancer.
- seeing the thermodynamics and activities in environments where water activity plays a major role. What are the activity coefficients like in these environments?
- how to measure the substrate concentrations *in vivo*? •• Metabolomics
- additionally recommended:
 - ∘ report K_{eq},
 - o report on differences between the actual medium and *in vivo* conditions.

Requirements:

- Improve the definition of *in vivo* medium,
- use the same *in vivo* medium for thermodynamic and kinetic standardization experiments
- measurement of the dependences on T, pH, pMg, I, P, S, buffer type in order to enable extrapolations
- \circ what are the best pMg conditions, is T=310 applicable as second standard temperature?

The Commission agreed with HW that this underpinning paper would be a useful addition to the Recommendations paper and encouraged HW to proceed.

New thermodynamics tables

(presented by H. Westerhoff)

- NIST database, database is almost inaccessible, maybe it would be possible to automatically scrape the data.
- as this database is not that accessible as one would which it does not appear FAIR (findable, accessible, interoperable, reusable),
- in addition, it is not clear which database has been used by Alberty apart from the fact that the information used is incomplete and those numbers provided by Alberty are not easily accessible,



HW's second project/proposal:

- make all relevant data available electronically
- ensure the consistency with pre-existing thermodynamic Tables.
- recompute (check) results obtained by Alberty
- enable corrections for water activity (osmotic strength)
- feal with limitations of ionic strength approach by circumventing the extended Debye-Hűckel theory used by Alberty
- enlarge the number of substances in the Table
- using Alberty's formulas (which checked for their thermodynamic basis) HW then calculated $\Delta_f G_o$ ' for all reactants, for three different T's, 5 different pH's and the ionic strength of 0.25 M used by Alberty,
- comparison of these with the corresponding tabulated values in his book
- comparison of $\Delta_f G_0$'s computed using $\Delta_f G^0$ s from different sources

Proposal:

- thermodynamic data base under the wings of STRENDA DB
- linked to STRENDA DB
- needed: facility for entry of new data

After intensive discussion, the Commission agreed with this proposal and encouraged HW to proceed.

Perspectives on and for STRENDA

(reports from the break-out groups)

Three break-out groups were created to think frankly and unlimited about STRENDA, its impact and steps to increase the visibility and adoption of STRENDA DB.

Guiding questions (intended to create a kind of framework but not intended to think out-of-the-box) were:



- what is STRENDA delivierung to the community?
- how do you feel about the progess/status of the STRENDA project?
- without having heard about the status of STRENDA DB yet, what do you think about this service, its potential issues, challenges and benefits?
- which issues do you see regarding visibility and integration?
- suggestions on how to resolve these issues?
- what are you able/willing to contribute?

Results

Group 1

- STRENDA not widely known among enzymologists
- students often overwhelmed starting PhD but STRENDA should come in at the beginning
- students generally keen & motivated to try new things, but don't find STRENDA on their own
- some universities teach BRENDA, students know about this, seeds need to be planted early
- where? enzyme kinetics courses

Requirement for successful uptake

- raise awareness that standardisation problem exists, broader problem of data management
- provide online data and teaching materials, online data management course (e.g. team up with Chemotion, electronic lab notebooks)
- get students interested to use it

Aim: Teaching materials

- offer built courses at early stage, attractive also to biology students
- team up with NFDI for course development (STRENDA on its own insufficient to fill up whole course)



- hire 1st-year bio. students to help with testing
- <u>Plan B</u>: provide teaching material for STRENDA

Linkages

- Elixir (German bioinformatics) what is done in terms of teaching?
- EnzymeML/PyEnzyme
- grants more likely to be funded if containing a demonstrable teaching component
- involve STRENDees to obtain support

Group 2

Strategies to increase adoption

- Required by Journals: invite editors to meetings, STRENDees attend editorial board meetings,
- Required by Funding Agencies: speak with Program Directors interested in data integrity. NIH has previously expressed a willingness to recommend STRENDA Guidelines in the Data Sharing section of grand applications.
- Participant Adoption: STRENDEe presentations at meetings, brochures at meetings, introduce concepts to students (teaching)

additionally:

- Publications
- Advertise in select journals ... Mission Statement that point to web information
- STRENDA DB entries enhance publication profile
- Database entries = lab notebook

Other topics

- mechanisms in the EU (funding agencies?) to require database utilization?
- data-entry interface is narrow and difficult to use (prerequisite to approaching journals),
- should STRENDA expand to include catalysis more generally?



• are we requiring too much information (metadata)? Let's not forget that a peerreviewed paper is published.

Additional suggestions:

- STRENDA DB could expand to include inhibition of enzyme catalyzed reactions. Competitive, uncompetitive, and no competitive etc. (This requires altering the DB interface).
- suggestion that the "biocatalysis experts" decide how they would like their type of data "inputted" into STRENDA DB.
- some journal editors or associative editors perhaps could be considered for talks at the next enzyme conference (October 2023) and be then asked to participate in STRENDA discussions. They could discuss how best to get their authors to upload their data to STRENDA.

Group 3

- Add a "enzymology" domain page to RDMkit (https://rdmkit.elixir-europe.org
 - enzymology standards + repositories
 - reference to STRENDA and STRENDA DB
 - relations to other domains
- STRENDA could develop recommendations for experimental design (in addition to reporting)
- too many enzyme function experiments still cannot be reported in STRENDA DB
- mandatory data submission to STRENDA DB
- better interlinking between STRENDA DB, BRENDA and SABIO-RK

Announcements

- 1. Beilstein Enzymology Symposium 2023: planned for September 2023 in Rudesheim.
- 2. STRENDA Meeting 2023, planned to be the day before the Enzymology Symposium