Photo by Rolf Süssbrich, CC-SA 3.0

Minutes of the

12th STRENDA Meeting

22 - 23 September, 2016

European Bioinformatics Institute

Hinxton, Cambridge, UK

by Carsten Kettner

Table of Contents

Agenda

(as approved by the participants)

- Introduction of Beilstein-Institut
- Introduction in STRENDA (for invited guests)
- Report on past and future STRENDA activities
- Experimental Standard Conditions – if and how to implement?
- Report from ESAB meeting
- Extension of the STRENDA Guidelines, Level 1A
- STRENDA DB
    - Report, progress and comments
    - Inhibition and activation parameters
    - Promotion and next steps

Participants

(in alphabetical order)

Baici, Antonio (AB)	University of Zurich, Switzerland	toni.baici@gmail.com
Fitzpatrick, Paul (PF)	University of Texas San Antonio, TX, USA	fitzpatrick@biochem.uthscsa.edu
Halling, Peter J. (PH)	University of Strathclyde, Glasgow, UK	p.j.halling@strath.ac.uk
Hicks, Martin G. (MH)	Beilstein-Institut, Frankfurt am Main, Germany	mhicks@beilstein-institut.de
Kettner, Carsten (CK)	Beilstein-Institut, Frankfurt am Main, Germany	ckettner@beilstein-institut.de

Leyh, Thomas S. (TL)	The Albert Einstein College of Medicine, Bronx, NY, USA	tsleyh@gmail.com
Mueller, Jonathan W. (JM)	University of Birmingham, UK	J.W.Mueller@bham.ac.uk
Raushel, Frank M. (FR)	Texas A&M University, College Station, TX, USA	raushel@chem.tamu.edu
Rohwer, Johann (JR)	University of Stellenbosch, South Africa	jr@sun.ac.za
Schnell, Santiago (SS)	University of Michigan, Ann Arbor, MI, USA	schnells@umich.edu
Swainston, Neil (NS)	University of Manchester, UK	Neil.Swainston@manchester.ac.uk
Westerhoff, Hans (HW)	University of Manchester, UK	hans.westerhoff@manchester.ac.uk
Wohlgemuth, Roland (RW)	Sigma-Aldrich, Buchs, Switzerland	roland.wohlgemuth@sial.com

Unfortunately absent

- Amos Bairoch, Swiss Institute of Bioinformatics, Geneva, Switzerland,
- Barbara Bakker, University Medical Center Groningen, The Netherlands,
- Athel Cornish-Bowden, CNRS-BIP, Marseille, France,
- Claire O'Donovan, EBI, Hinxton, Cambridge, UK
- Dietmar Schomburg, Technical University Braunschweig, Germany
- Ming-Daw Tsai, Academia Sinica, Taipeh, Taiwan

Introduction of Beilstein-Institut

The Open Access journals (*Beilstein Journal of Organic Chemistry* and *Beilstein Journal of Nanotechnology*) belong along with the standardization projects STRENDA and MIRAGE to the general mission of the Beilstein-Institut to disseminate chemical information and information from related disciplines – scientific and science-related – within the science community. The focus here is to propagate the free access to information (Open Access) and data (Open Data) which includes also means for sharing of science data. Since knowledge generation is driven by experimental data of high quality any action should be taken to ensure that data are both reported completely and collected comprehensively. It is expected that high quality data results in an increased reproducibility of experimental findings.
Thus, the Beilstein-funded standardization projects STRENDA and MIRAGE meet perfectly these objectives. The foundation is engaged in discussing all aspects of Open Science which includes the above mentioned aspects of fostering Open Access and supporting Open Data. This is reflected by the participation of MH and CK in in the Chemistry Research Data Interest Group at Research Data Alliance (www.rda.org) and in the BioSharing Working Group at FORCE11 (The Future of Research Communications and e-Scholarship, www.force11.org).

Please refer any details to the web site of the foundation, i.e. http://www.beilstein-institut.de

Introduction in STRENDA (for invited guests)

CK gave a short introduction in the STRENDA project, its roots and goals for invited guests who were not aware in detail of both the Commission and its missions, i.e.

- Development of experimental standard conditions;
- Definition of minimum information for reporting functional enzyme data (STRENDA Guidelines);
- Generation of a comprehensive data acquisition and storage system (STRENDA DB)

For further details, please refer to the web site of STRENDA, i.e. http://www.beilstein-strenda.org.

Report on past and future STRENDA activities

CK looked back on the last twelve months since the last STRENDA meeting, held in Rüdesheim, September 2015. Some activities in relationship with STRENDA are to be reported:

- presentation of STRENDA and STRENDA DB in two talks at the ACS Spring meeting, March 2016 in San Diego. The CINF division (Chemical Information) provided an ideal platform to present and discuss the current issues of scientific data reporting, standardization, sharing and reproduction (the slides are available here: http://acscinf.org/content/cinf-2016-data-summit-slides-251st-acs-meeting-march-2016). The feedback on the presentations was very positive and helped to make STRENDA more visible in a community that is mainly active in the area of data collection and processing. However, here, CK learned that not only publications can be registered with a DOI (Digital Object Identifier) but also datasets and thus the idea was born to register datasets in STRENDA DB with DataCite (see below).

- FEBS Congress: It is planned to organize another Special Session on Data Quality together with Israel Pecht at the upcoming FEBS Congress in Jerusalem in 2017. CK's proposal has been approved by the organization committee. Further organizational actions will take place in due course.

- Biosharing: STRENDA is registered with biosharing (www.biosharing.org), a portal that collects standards, databases and procedures and links related activities to each other. The STRENDA Guidelines still are the only ones that support authors to report functional enzyme data completely and comprehensively.

- DIN Standards: CK was appointed to a working group within the DIN Standards Committee Food and Agricultural Products that is concerned with standards for Biotechnology. The aim of this working group was to develop standards for traceable, searchable and interoperable data together with integrated data process for biotechnology and life sciences. Unfortunately, the German standardization body, DIN, decided to charge a fee from the members of this working group which caused the termination of Beilstein's activity in this group.

- Registration of Guidelines: Before registering datasets, first the STRENDA Guidelines have been registered with CrossRef after this registrant also offered the registration of standards. The STRENDA Guidelines are now registered with the following DOIs:

Level 1A (Experimental Conditions): doi:10.3762/strenda.1

Level 1B (Experimental Results): doi:10.3762/strenda.2

Generic Text: doi:10.3762/strenda.3

It is now possible to refer to the STRENDA Guidelines by also citing the DOI which guide the user directly to the Guidelines at the corresponding web sites (e.g., type http://dx.doi.org/10.3762/strenda.1 in your browser's address line).

- Registration of datasets with DataCite: inspired by the ACS meeting (see above) Beilstein decided to also register datasets stored in STRENDA DB with DataCite which is a registrant for data sets under the leadership of the German National Library of Science and Technology (TIB Hannover). A specification for both the process and the technical implementation has been prepared, the implementation has been started whilst these minutes are written.
- Developments, tests and bug fixing of STRENDA DB (s. below)

Experimental Standard Conditions – if and how to implement?

The STRENDA Commission initially started with the idea to standardize the experimental conditions for assaying enzymes. However, it became quickly obvious that for this task much more expertise would be required than was then available. Thus, the Commission decided to propose guidelines for reporting functional enzymology data – the STRENDA Guidelines.

However, the task of setting experimental standard conditions was and is still on the agenda and in order to come to a clear agreement for addressing this issue in the future it has been extensively discussed. The following questions intended to guide through this discussion:

- Experimental standard conditions - what would this mean in practice?
- reports from the practice – how are experiments carried out?
- what is possible to standardize, and what isn't?
- Ii it useful – is it doable?
    - If yes, how to implement this?
- what could be potential ways:
    - guidelines**:** Organism-related definitions of experimental conditions.

- repository of well described/performing procedures

In addition, further aspects came up during the discussion:

- do we talk about *in vivo* standards? *In vivo* conditions can change, they are very complex and complicated.
- enzymologists ask different questions as data users, there are different user groups, how to suffice the groups' requirements?
- there is a need for standardized assays, in particular from the systems biology community;
- what is possible to standardize? Just pH and T?
- what can be specified in general? What are the most important molecules in tissues/cells/organisms? Can SOPs be created for enzymology?
- how to systematically organize these standardization efforts? Is STRENDA the right body doing this? Will STRENDA be accepted by the community proposing such standards? Or is this aim out of the scope of STRENDA?
- it is not the aim of STRENDA to require certain experimental procedures, the group's aim is just to monitor information rather than policing.

Conclusion

Although there is an agreement that standardized assay conditions would be very desirable the Commission was hesitant to really take up this big challenge “may be too big for STRENDA”

Decision

Two options have been worked out, either to remove this task from the agenda or to consider it an item which is put in a basket for future activities. The Commission decided to keep the experimental standard conditions in principle on the agenda but to address this issue in the future. The major task now is to finalize and promote STRENDA DB.

Report from ESAB meeting

The European Section of Applied Biocatalysis (ESAB) within the European Federation of Biotechnology (EFB) has held its meeting as a satellite to the EFB conference in Krakow, Poland, in July 2016. RW presented both the mission of STRENDA and STRENDA DB and he received positive feedback from the members of the ESAB Commission on these efforts. However, it also turned out that STRENDA seems to be under-represented in Asia since the Asian members were not aware of the activities of the Commission. This is an issue which should be addressed in the near future.

In addition, a number of comments and questions were raised at the ESAB meeting and these were discussed and answered at the STRENDA meeting as following:

- how can the existing data in the STRENDA DB be searched?

  There is an easy-to-use query page which provides just one input field (comparable to Google), hits are displayed in a table. Details are accessible either as a PDF file or via the 'Experiment Overview' page. However, there is currently no browsing through the data base implemented since the focus of the implementation was on the data input rather the data representation.

- Is the statistics available on the number of enzyme functions entered ?

  Not yet but this issue will be addressed when more data will come in the database.

- Is there some recognition for the researcher who enters data, e.g. by a citable location

  Yes – each data set is registered with DataCite (www.datacite.org), is annotated with metadata and has a DOI (Digital Object Identifier). DOIs along with the metadata make data more visible/searchable and trackable. For example, accesses of individual data sets can be referred to the authors of the corresponding paper, and vice versa.

- What is the best practice for novel enzymes in the sequence application for new EC number - adding data to STRENDA DB – publication of research article?

  STRENDA DB not only collects data from already published proteins which are registered with UniProtKB but also collects data from novel proteins. In this case, the unambiguous identifier is the protein sequence which needs to be entered together with a reasonable name by the user.

In addition, any further information on e.g. sequence modifications, post-translational modifications, source, organism etc. need to be entered as well.

- How does STRENDA relate to enzyme databases for specialized enzyme classes?

  Currently, there are no connections between STRENDA DB and databases that are specialized on the collection of certain enzyme classes. This issue can be addressed when STRENDA DB will have gained reasonable impact for the community.

- How will long-term access and support be secured?
  As long as the science community contributes to STRENDA DB with the input of data and the journals embedding STRENDA DB in the publication process more consequently, the Beilstein-Institut is willing to secure long-term access and support.

- Additional Interest from EiC *Journal of Molecular Catalysis B*: *Enzymatic*.

  This comment refers to the list of journals that recommend their authors to refer to the STRENDA Guidelines when reporting functional enzyme data. The STRENDA Commission appreciates much this news and will follow up this issue.

- What is the interest of *Nature* and *Science*?

  Not sure – *Nature Chemical Biology* is recommending the STRENDA Guidelines and actually it is expected that other relevant journals from the *Nature* family are following but it is not clear if there are common instructions for authors for all *Nature* journals or if there the instructions are made individually for each *Nature* journal. For *Science* it is not clear whether there is interest in the Guidelines but this will be sorted out.

- How can the authors be motivated to enter the enzyme data?

  The authors need to be convinced by the benefits when entering data in STRENDA DB, i.e. automatic check on completeness (contributes to high quality of the manuscript), visibility (by registration of the data sets with DataCite and receiving DOIs), contribution to a fast growing functional enzyme database. However, on the other side, it is the role of the journals to leverage the impact of STRENDA DB by making the use of the assessment (and input of data) on compliance with the STRENDA Guidelines a requirement for the submission of enzyme data to a journal.

- What recognition can be given to each author for the enzyme function data entered?

  See above with regards to the registration of data sets.

- Is there in addition to the website a user forum?

  No, not yet, but this issue can be addressed when the need for such a forum becomes obvious. The advantage of the user forum is that user can help each other with problems without contacting the institution that runs the service. However, it is also important for Beilstein and STRENDA to learn from the community what needs to be improved or even changed.

- How are reviewers made aware of STRENDA DB?

  Ideally by the journals that refer their reviewers to the datasets stored in the non-public part of STRENDA DB. In addition, since the reviewers are part of the science community and thus are "writing" researchers as well they will be made aware of STRENDA DB and its use as an assessment tool by the journals, again.

- What is the process for enabling the access of reviewers?

  This depends on the way how STRENDA DB will be embedded in the publication process and needs to be sorted out either individually with any journal or via a common interface. Currently, there is no general exclusive access for reviewers.

- How are other DBs willing to adopt the STRENDA guidelines be harmonized with STRENDA DB and provide massive input to the STRENDA DB?

  Currently, there is only loose contact to other databases such as BRENDA and SABIO-RK with regards to the exchange of data. STRENDA DB is open in principle and the willingness to share data is indicated by the provision of XML files for each data set. Since STRENDA DB is just beginning to collect data details on the ways of exchanging and sharing data need to be discussed with other databases in the future.

- Is a ranking of biochemistry journals according to the number of STRENDA DB PMIDs already available?

  No, it is not since the number of datasets in STRENDA DB is currently to low in order make statistics on rankings, access numbers etc. However, it is obvious that statistical analysis gains more and more impact with the growth of the database.

Extension of the STRENDA Guidelines, Level 1A

PH suggested to extend the STRENDA Guidelines by parameters that describe the storage procedures of isolated and prepared enzymes. The Commission agreed to this extension and made the following suggestion:

STRENDA Guidelines Level 1A

[…]

3. Preparation

[…]

4. Storage conditions

Data	Comments
Storage temperature	If frozen, freezing method, e.g. -20 °C flash
Atmosphere if not air	
pH	e.g. pH 7.0
At which temperature was the pH measured?	e.g. 25 °C
Buffer & concentrations (including counter-ion)	e.g., 200 mM potassium phosphate, 100 mM HEPES-KOH
Metal salt(s) & concentrations	e.g., 10 mM KCl, 1.0 mM $MgSO_4$
Other components	e.g., 1.0 mM EDTA, 1.0 mM dithiothreitol, 10% glycerol, 20% DMSO, 1 mg/ml PEG2000, 2 mg/ml BSA, peptidase inhibitors
Enzyme/protein concentration	Molar concentration if known, otherwise mass concentration e.g., mg ml^{-1} or better µM
Optional: Statement about observed loss of activity under the above conditions	e.g., less than 10% lost after 1 month
Statement about the thawing procedure	e.g., on ice

5. Assay Conditions

[…]

STRENDA DB

Report, progress and comments

CK gave an overview of the general idea of STRENDA DB regarding both the data being captured and the workflow of interactions between authors, journals and STRENDA DB. In addition, he summarized the latest developments, i.e. mainly change requests, testing and bug fixing since the last STRENDA meeting in September 2015.

The following 19 changes requests have been implemented and tested/bug fixed:

- extensions of help text,
- extension of user guidance (tool tips),
- input of synthetic proteins, or proteins without UniProtKB AC,
- validation with protein modifications less strict,
- re-ordering fields (protein source),
- correction of units (including such with negative powers),
- LogIn procedure: password forgotten,
- Pre-PDF for proof-reading,

In agreement with the Commission (decision made at STRENDA meeting, Sept. 2015) the following change requests were re-classified as feature requests and will be addressed in subsequent versions of STRENDA DB:

- inclusion of equations
- time course data, etc.
- databases in mind: ChEBI for lipids, UniProtKB for proteins
- copy experiment #1 to experiment #2

Before releasing STRENDA DB in Version 1.0 a final development needs to be implemented, i.e. the registration of data sets with DataCite and combining each data set with a DOI and metadata automatically. It is believed that the implementation of this feature attracts authors to enter their manuscript data into STRENDA due to the following benefits:

- data permanently accessible regardless of physical location,
- easy citation of data for reuse and verification,
- each dataset with its unambiguous DOI,
- impact of data is tracked à rewards to authors

The implementation of the DOI registration, however, requires a small change in the generation of the STRENDA Registry Number (SRN). This number is currently a time stamp and indicates at which time in seconds after 1 January 1970 the data set has been saved in STRENDA DB. This results in a 13 digits SRN which can even become longer in the near future. The creation of both long SRNs and the DOI which is planned to include the SRN makes the use of the SRN rather cumbersome. Thus, an alternative solution for the creation of SRNs is required and going to be investigated.

The requirements are:

- unambiguous
- short (< 6 digits)
- letters and numbers only
- not counting in sequential series
- non-speaking

Potential approaches are:

- encryption using hash algorithm (e.g. MD5, UOWHF, etc.)
- collision-resistant
- decryption not relevant, no security issues

It is planned to finalize the implementation of the data set registration by mid of October.

The Commission agreed to the benefits of registering data sets with DataCite and discussed various aspects of the registration procedure and consequences that can arise by changes in the manuscript and/or the data themselves.

Presentation of STRENDA DB

NS introduced in STRENDA DB and gave a short overview of the system by a data input on-the-fly.
The presentation showed that the current implementation of STRENDA DB in Version 0.9.9 has still some weaknesses, i.e.

- input of K_m is only possible if a concentration range is applied; if the concentration is set to fixed there should be no way to enter K_m. This feature belongs to the group of validation processes and will be addressed in subsequent versions of the system;

- repeat the name of the ESS in the results sections “Edit Kinetic Parameter”, meaning an additional line for the name previously given above the line “Kinetic Parameter for <substrate>. It should be possible to include this feature in Version 1.0;
- include journal name and DOI of the paper in the query hit list. This feature request will be reserved for subsequent versions.
- The significant digits of the pH entered by the authors should not be modified. E.g. pH = 7.0 should not be cut down to '7' since 7.00 means 7.00 ± 0.01 but '7' means 7 ± 1.
- add units (that have impact in particular for “other compounds” for which the exact molecular mass is uncertain): % v/v and % w/v.
- Methodology: it seems not to be clear what to enter here and why this field appears here. After there has not been found an obvious alternative place to move this field to (and since this dilemma has been expressed in STRENDA DB as well “At present we ar enot sure whether the positioning and availability to this box is best, and would appreciate comments on this.”) there was a general agreement to at least change the text describing the anticipated content.

On 29 September, JR made the following suggestions that modify the already existing texts slightly (which can be implemented in Version 1.0):

Experiment Details

Any information about the methodology used and techniques applied in your experiment should be given. A literature reference may suffice for an established procedure but any modification should be detailed.

For further details, please refer to the quick help.

Quick help text:

Experiment Details

In this field any information about the methodology used and techniques applied should be given. A literature reference may suffice for an established procedure but any modification should be detailed.

The intention here is to provide the user with a free text box to give details that are not currently captured in the more structured fields that follow.

There will be input forms for details on the components of the assay such as protein assayed, reactant compounds, concentrations, pH, etc., on subsequent pages. **These should not be entered here.**

Following information should be provided in this text box:

Type of Assay	e.g., continuous or discontinuous, direct or coupled
Stopping procedure	In the case of discontinuous assays, e.g., quenched in 15% perchloric acid
Direction of the assay	With respect to the reaction equation provided, e.g., NAD reduction by alcohol dehydrogenase; alcohol + NAD+ --> aldehyde or ketone + NADH + H+
Reactant determined	e.g., NADH formation, O_2 utilization
Quantification method	e.g., Absorbance at 340nm, mass spectrometry

Inhibition and activation parameters

AB gave an overview of inhibition and activation of enzymes and presented mathematical and graphical representations of 17 different inhibition and activation types. He proposes an extension of STRENDA DB which currently just can cover inhibition data for five linear inhibition mechanisms but leaving out any allosteric interactions. A detailed introduction along with many examples from the literature is provided on www.enzyme-modifier.ch.

In addition to his presentation, AB provides the Commission with detailed suggestions on how to modify STRENDA DB in order to enable the system to capture enzyme-ligand interaction data accurately and comprehensively (see appendix).

It is recommended to discuss this issue at the next STRENDA meeting, again.

Promotion and next steps

The Commission agreed on focusing the finalization of the development of STRENDA DB and then promoting the system within the science community. In his introduction, TL divided the latter task into bottom-up and top-down approaches. After collecting of a number of potential tasks the next step was to sort out both the responsibilities for each task and the starting time. A few tasks were left unassigned and will be later moved on the agenda, again.

Bottom-up approach

What?	Who?	Starting? / Status
STRENDA Poster	CK provides updated “old” poster	
STRENDA logo on top slide	CK sends logo to Commission	DONE
2 – 3 slides on STRENDA	Tom sends his slides to Commission	
Contact to scientific press	Santiago	In progress
Disseminate STRENDA fact sheet (PDF)	ALL	ASAP
Providing DOI of datasets in slides, in the paper	ALL	ASAP
“data reported in compliance with STRENDA Guidelines” - at end of paper, last slide	ALL	ASAP
Reference to STRENDA web site (i.e. www.beilstein-strenda.org)	ALL	ASAP
Nucleic Acids Research paper: publication on STRENDA DB (or Preprint Server to publish rapidly)	Neil drafts for first call, on Google Docs	In progress
Citing STRENDA DB on individual's web pages	ALL	ASAP
Short paragraph in EFB newsletter	CK draft (by mid of October)	10 October
Contact to Asian community	Roland	
Flyer on STRENDA for conferences	CK	
STRENDA T-shirt		“item for the basket”

What?	Who?	Starting? / Status
STRENDA on social media	CK: news on new datasets (along with information about orginal paper), video advertising STRENDA DB	DONE (with Beilstein): Twitter: @BeilsteinInst RG LinkedIn: https://www.linkedin.com/company/beilstein-institut?trk=top_nav_home
Movie on mission of STRENDA (who are we, what do we?)	CK	
Announcement at conference in Groningen, Oct. 2016	Roland	ASAP
BioTrans, Budapest, 2017	Roland	
Enzyme Engineering, Toulouse, Sept., 2017	Roland	
Announcement at conference in Como	Hans	
FEBS Conference, Jerusalem, July 2017 – special session	CK	In progress
Links from STRENDA to conferences	CK	
1 – 4 July, 2018, ECB conference, Geneva	Roland	
1st week of October, 2016, Systems biology meeting, Jena	Johann	DONE
On STRENDA DB webpage, list of people who helped us getting better	CK	ASAP
Input of already published data, input of new data	ALL (Deal: any new data in STRENDA DB will be posted on Twitter)	ASAP

Top down approach

What?	Who?	Starting? / Status
Submit data to STRENDA DB and submit PDF to journal	ALL	ASAP
Request (in letter to EiC): Suggestion from journals in letter to authors recommending STRENDA DB Letter to Editors-in-Chief	Paul -draft	ASAP
Getting journals interested Aim: Journlas recommend authors to use STRENDA DB Including the recommended use of STRENDA DB in Instructions for Authors		
Biochemistry	Tom	After release
JBC	Fred (CK will contact him)	After release
Biophys. Chem.	Santiago	In progress
FEBS J.	Nigel Scrutton (who contacts him?)	After release
FEBS Lett.	CK	After release
Japanese journals, sort out the most appropriate ones	Roland	
Journal of Biochemistry (Tokyo)		
Funding agencie		
BBSRC		After release
EU – one page proposal: what do we intent to achieve in 10 years?		After release ASAP
EBI	Claire?	

Appendix

in a separate document: Suggestions for the STRENDA-DB Version 1 by Antonio Baici.