Minutes of the

The 20th STRENDA Meeting

November 12–13, 2024

Łódź University of Technology
Łódź, Poland

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

based on the notes of
Hans V. Westerhoff and Carsten Kettner

Table of Contents

The 20th STRENDA meeting took place on Tuesday and Wednesday, November 12 and 13, 9 am – 5 pm, at the Łódź University of Technology. As some members could not travel, the meeting was hosted as a hybrid event allowing the remote members to participate and contribute to the discussions.

List of Participants

- Andreas Bommarius, Georgia Institute of Technology, Atlanta, USA (online)
- Alan Bridge, Swiss Institute of Bioinformatics, Geneva, Switzerland
- Andrew McDonald, Trinity College Dublin, Ireland
- Peter Halling, Strathclyde University, Glasgow, UK
- Jürgen Pleiss, University of Stuttgart, Germany
- Frank M. Raushel, Texas A&M University, College Station, USA
- Johann M. Rohwer, University of Stellenbosch, South Africa
- 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) but eventually online attending

- Paul Fitzpatrick, University of Texas Health Science Center at San Antonio, TX, USA
- Ming-Daw Tsai, Academia Sinica, Taipei, Taiwan
- Robert Goldberg, Bethesda, ML, USA
- Thomas Leyh, The Albert Einstein College of Medicine, Bronx, NY, USA
- Barbara Bakker, University Medical Center Groningen, The Netherlands

Agenda

Tuesday, November 12

9:00	Welcome and Opening Prof. Małgorzata Szynkowska-Jóźwik (Dean of the Faculty of Chemistry) Carsten Kettner
	Overview of STRENDA Activities C. Kettner
10:30	Coffee break
11:00	Status of STRENDA DB Andrew McDonald, C. Kettner
11:30	Biocatalysis Reporting Guidelines / Metadata Catalogue Peter Halling, C. Kettner
12:00	EnzymeML - New developments and Applications Jürgen Pleiss
12:30	RDMkit Ulrike Wittig, C. Kettner
13:00	Lunch
14:00	Guidelines and Tables for Thermodynamics Data Reporting Status, next steps Hans Westerhoff
15:00	Coffee break
15:30	Organization of the STRENDA Commission C. Kettner
16:30	Introduction of Andreas Bommarius
17:00	Close
19:30	Dinner

Wednesday, November 13

9:00	Opening and Reconciliation C. Kettner
9:15	Discussion: What is minimum information, what is its purpose, which information is required by whom?
10:30	Coffee break
11:00	Inclusion of STRENDA into the EQUATOR Network? Discussion
11:30	STRENDA Outreach (1) - a general slide Break out groups, 3 groups à 3 persons task: create one (or two) slides that can be added to your slide desk to reach out on STRENDA
12:00	Reporting and discussion of the results
12:30	Lunch
13:30	RHEA and UniProt Alan Bridge Sorting out collaboration ways
14:00	STRENDA Outreach (2) - the Landscape Break out groups, 3 groups à 3 persons task: displaying the interactions of the project with other projects worldwide
15:00	Coffee break
15:30	Reporting and discussion of the results
16:00	Strategies for the outreach increasing awareness, increasing data input, what is missing, what needs to be done?
16:30	How to proceed? Tak list
17:00	Close
19:30	Dinner

STRENDA – news and activities

Withdrawals from the Commission

Meina Neumann-Schaal got another appointment and thus is not longer responsible for BRENDA. Therefore, she decided to step back but she emphasized that she will be stay in contact if the Commission asks for his advice. The Commission thanks Meina for her short engagement and for her contributions to the STRENDA project. Dr. Boyke Bunk who is now the project leader of BRENDA would be the first candidate to be contacted. Alternatively, Prof. Dieter Jahn whose department is associated with BRENDA could be contacted.

ACTION POINTS

- Carsten will contact Dr. Boyke Bunk.
- Roland will contact Prof. Jahn about how to continue contacts between BRENDA and STRENDA
- Andrew and Ulrike offer to talk to BRENDA database people to see how the STRENDA-BRENDA interaction can be deepened.

New appointments

Professor Andreas Bommarius, Georgia Institute of Technology, Atlanta, GA, USA, accepted the invitation to the join the STRENDA Commission. He introduced himself as being much interested in comprehensive and proper reporting of enzyme and biocatalysis data and making these data publicly available. He has contributed a number of datasets to STRENDA DB. Andreas' background includes industry at Degussa from 1990 to 2000. He then finally moved to the Georgia Institute of Technology after some years in Munich and at the MIT. His research interest is in the development of new redox enzymes and reactor engineering. The STRENDA Commission voted unanimously for Andreas Bommarius being appointed to the STRENDA Commission.

Data links to Europe PMC

EuropePMC provides users with links from the abstract of the published paper to data stored in repositories and databases. The generation of this link requires the use of a specific metadata template to upload the connections of datasets and EuropePMC identifiers. This needs to be done manually as there is no API or any other automatic process available. An example, how this linking looks like, is a paper of Taylor et al. JACS, 2017, DOI:10.1021/jacs.7b04824, PMID: 28650156, PMC5629633. On EuropePMC scroll down the hit to “Data behind the article”.

Almost all data sets in STRENDA DB are now linked with EuropePMC. The update of datasets is carried out twice per year.

Data links to PubMed

Similar to the process of linking datasets from STRENDA DB to EuropePMC, STRENDA DB datasets are now also linked with PubMed using a specific metadata template. The process is the same as for EuropePMC as due to the lack of a API the linking process has to be done manually.

Currently, 85 datasets from STRENDA DB are linked with PubMed. The update process is carried out twice per year.

RMDkit

Suggested by Ulrike at the previous STRENDA Meeting, we have created a new domain “Enzymology” on RDMkit pages. RDM stands for Research Data Management and provides information on guidelines, initiative, tools, software, data exchange formats as well as databases and repositories for all those who are interested in data deposition and sharing.

RDMKit helps to make your data FAIR. It is a community effort with 14 editorial board members. Presently it consists of approximately 2000 web pages. It uses Github or Googledocs and has 7719 users (since 2012) and 53 thousand views. RDMKit can help researchers to make a data management plan to put into their grant applications. Such a facility is not available yet.

The idea now proposed to the STRENDA commission is to have add to the RDMKit:

a domain page for enzymology and a page about using EnzymeML

The same idea has already been used by Chemistry in the mean time and by te research community for rare diseases. The aim of our activity is to increase the visibility of STRENDA among RDM maintainers.

A first draft of this document is provided on Google Drive at https://tinyurl.com/3spny8sk.

Discussion

Roland**:** This is a great and ambitious project. It would be good to extend to Inborn Errors of Metabolism database system. Is there a connection with the omics (?) database. Is there metabolomics community? Not yet.

Ulrike responds that the target is not so much researchers as data managers and it only gives guidelines; it is no repository of research data. One should use this before one starts with a project, when writing a data management plan. An editorial board only deals with the guidelines and the text (checked for consistency) but not with the data management structure.

Hans raises that FAIR is limited; assurance of reproducibility is an aspect that must be added to data management plans. How can this be added the RDMKit? Ulrike**:** Users can make suggestions for changes using the pen.

STRENDA DB

Overview

The current technical version is 1.3.3 (from December 23, 2024). The unit handling was bugfixed and the special characters font was replaced. Greek symbols are represented by an empty square in the PDF but they are properly displayed in the HTML. The symbols are all done UniCode characters. However, as font Helvetica has been used which does not include Greek symbols. From our Beilstein journals we already are using useful fonts that include all the special characters and as the generation of the PDF is affected only, the exchange won't take too much efforts. Helvetica seems to be the default font with iText but makes a number of trouble as it is not a font that is ubiquitous available on all OS but a font that needs to be licensed.

We thus changed the PDF font towards a font which is well proven to display all the “foreign” characters we are experienced with and thus using that font which is also used for the Beilstein Journals.

In addition, the UnitConverter has been extended by micro l (µl). Finally, Java classes using Cayenne Modeler Update UniProt Searching have been updated in a way that the new RESTful API can used to report true or false as “yes” or “no”.

The database provides >108 datasets to the public, 24 additional datasets are going to be finalized (assignment to the publication is missing). There are some entries of doubt (no reference found, incomplete and “sloppy” data) which will be removed after another double-check. Many more new data sets just have been started but no further progress could have been observed.

When contacting the authors, these messages often remain without any replies which is a strong indication that the data sets generated are “death” ones which have been created for testing reasons. These datasets have been removed as well.

The user manual has been updated from version 0.92 to version 1.3.3 which reflects now the current version of STRENDA DB.

Currently, the developer’s framework is updated by replacing the Apache Cayenne Modeler by Hibernate. Both frameworks do similar things: object-relational mapping, DB schema is mapped to Java classes and SQL statements are created during runtime. However, Hibernate does this more efficiently by creating DataTransferObjects which allows a faster and easier data handling.

Andrew now signs off developing for STRENDA DB as he has no experiences with Hibernate which is now handed back to Udo.

Carsten reports that he had underestimated the Java complexity. He has learned that infrastructure is a difficult issue. STRENDA DB is now state of the art however and ready to have additional development done by a new person, i.e., Udo. Carsten believes that to keep the database system running is safe now, as there is money for maintenance. Keeping the database running will be manageable because it is Just 100 lines of code. However, adding new feature into STRENDA DB may be harder though.

Roland noticed a much increased interest in Brussels last week in modern enzymology; models, and thermodynamics data. Therefore we will need to implement further progressions; more complex things like kinetic models.

We are now observing that more and more journals require the authors to provide a data availability statement. Carsten showed an example how this looks like. Frank: there is reluctance with editors to require statement that data have been put into data repositories. Carsten: but this has been changing as e.g. funders now require this. Johann sees that suggesting to deposit their data somewhere is having effects. Carsten: yes, the DOI of the datasets has some impact. However, a uniform way of how to reference datasets in STRENDA should be provided for the journals.

ACTION POINTS

- Carsten will provide a bit of text on how authors can cite the data entry in STRENDA DB in their submitted paper.

Recommendations for STRENDA DB

Wiley-VCH updated its data deposition policy for all journals of European Chemical Societies that publish the following journals: *European Journal of Inorganic Chemistry, European Journal of Organic Chemistry, ChemCatChem, ChemSusChem, ChemBioChem, Chemistry Europe, ChemSystemsChem*. With this addition, more than 25 journals recommend their authors to deposit their data in STRENDA DB. And, of course, the list of journal that recommend the STRENDA Guidelines has increased by these seven journals as well.

Biocatalysis Reporting Guidelines

Peter reported about the progress of the STRENDA Biocatalysis Guidelines for Cataloging Metadata. The corresponding paper has been accepted by Nature Catalysis.

This contains a catalogue for what might be needed in an experiment in Biocatalysis. Much of this is already in the original STRENDA guidelines. Other guidelines are new and part of a new catalogue:

- more than one phase may be present (liquid plus solid, for instance);
- then the mixing,

- dimensions of the reaction vessel (need to be specified).
- all the components of the ration mixture, operation mode, vessel type, catalyst components, reaction conditions, sampling need to be reported.

Decision trees are being discussed.

The catalogue is available on the Github system.

//github.com/Strenda-biocatalysis

//w3id.org/strenda/biocatalysis-guideins

Discussion

Hans finds what an experimenter would need to report massive and suggests that there could be simpler standard routes through the tree and demanded for simplification. Herewith it could be recommended that experimenters uses certain standard reaction media, rather then to develop and then report their own (example: the use of tissue culture media on cell biology). Some commission members expressed that such standardization is not desired because it limits freedom of experimenting.

Roland mentioned that there is an added disadvantage of all using the same standard media; certain pitfalls may not be noted such as for instance a limitation by amino acids.

Johann mentioned that the catalogue helps to reflect that the biocatalysis environment is wider than enzymology and it demonstrates this. The problem of the complexity of reporting is alleviated by the fact that not all fields need to be populated by the user. Jürgen added that most of this will be done by the computer rather than the experimenter. Then, the datamodel of the catalogue is needed but complexity will not be there. Peter mentioned that biocatalysis people will only work towards maximum yield. Being prescriptive will not work therefore.

Carsten: at present subgroups are being made.

Jürgen: we now still need to read methods sections of papers to see if our datamodel is complete.

Hans asked if there is a way now to put in kinetic models and Km's (he missed this in the presentation) and if the connection with other STRENDA guidelines is made (e.g. The Golberg recommendation on equilibrium constants)?

Jürgen: Often the people miss the tools and motivation. Reaction engineering is more than screening. It involved accurately defining enzymes for a network reaction system. The progress here includes that it is no longer text but already a mark-down document (in Python). These guidelines already include the datamodel (STRENDA guidelines were still written text).

EnzymeML – Status, progresses and perspectives

Jürgen introduced that at the 14th STRENDA meeting it was agreed that a markup language was needed for enzymology data. In 2022 the concept was presented and since then annual EnzymeML workshops were hosted.

The project has developed tools, and convince people to use these. It has made tools usable and accessible. In order to enable integration of different types of experiments. it has produced an extensible data model and developed a tool to enable making a reproducible workflow.

The EnzymeMLplatform provides:

- an extensible data model (Python library)
- tools for extension of the data model
- standard exchange formats
- tools for data acquisition
- a modular workflow platform (Jupyter)

Discussion

Formally STRENDA DB and EnzymeML are two different projects. Carsten explains the various STRENDA-linked initiatives and finds it useful to do two different things; different working groups on different aspects.

We should perhaps change names to make visible that the two projects are connected and that they use the same data model.

Frank: does the Enzyme ML help to deal with a three enzyme pathway and to optimize this?

Jürgen: First step is collecting experimental information (on the individual enzymes). This is in Enzyme ML. The next step is modelling of the cascade. Choose the best model of the three. Next step is to use a neural ODE model.
Best model is Ken Johnson's model. The modelling tools are the choice of the user and therefore made independent of enzyme ML.
Hans: Does Enzyme ML provide help for the users to then approach the three enzyme pathway? Jürgen: Such best practices will become part of Enzyme ML.
Johann: Enzyme ML has the flexibility and possibilities of SBML. Jürgen: and Enzyme ML has extended SBML by now also getting the experimental data in. We do not yet have metadata o the process of modeling. What is included is the rate laws, but not the process of modeling itself. This will be feasible. For this there is already SBML (includes simulation: to capture what you do with the model; the modeling steps).
Peter: It is good that you know to recognize AniML and do not develop things yourself.
Roland: how can enzyme ML facilitate association of datasets with STRENDA DB.
Jürgen: it is automatic (Ulrike: with a quality control).
To provide enzymologists with alerts early on about what is missing (or what is incorrect) for SABIO or STRENDA UPLOAD will be useful.
Uploading lots of data is not a problem for STRENDA or SABIO; It is the human quality control (read the literature) that is still important.
Ulrike: The user will want to know whether the data are directly empirical or the results of modeling (like alpha-fold). A database differs from an exchange format (the latter is not curated).

Thermodynamics Project: Standard Conditions and Standard States for physiological Chemistry

Hans presents the history of the project within STRENDA DB. He shows that the aims have largely been attained:

- a bulleted list with recommendations for new biothermodynamics use (finished; many contributors)
- a Thermodynamic Table with corresponding standard metabolic energies (finished in excel form (Hans); now final checking and Johann is making a simpler user interface, which he will presented later) (Hans, Johann). These is also text on how to use the Table and some results obtained when analyzing the Table. Paper plus Table to go into the *Beilstein Journal of Organic Chemistry*.
- paper with the underpinning thermodynamics (90% complete in LaTeX) including the rationalization of the new transformations.
- paper for TIBS for generating interest (Peter, Hans) and for having the bulleted list as part thereof.

He also showed a number of scientific results that have come form inspecting the Table. One example was showing cases where the group contribution was not quite right compared to the experimental method represented by the new Table. Hans considers that it is better to have the paper with the Thermotable out before the recommendations are published.

The working group for this work is Hans, Peter, Johann, Carsten

Johann then presented a tool based on Python and running in Google that enables getting the data out of the excel sheet.

Discussion

A suggestion was made to produce output in the form as could be used for publication: output.

Carsten: Let us try to get this on the STRENDA-website. He will ask his 'hoster' how to do this. A paper plus the Table as paper in *Beilstein Journal of Organic Chemistry* containing the Thermotable, should indeed be feasible. Carsten: one may be interested to have a license.

Peter: Equilibrator may have wrong values for compounds that only have a single metabolic reaction connecting and this may explain some of the things Hans showed.

ACTION POINTS

- the working group will finalize the manuscript and the table.
- Johann will further work on the development of the software tool together with Hans
- Carsten will contact the hosting service of the Beilstein-Institut webpage regarding hosting and running Johann's tool.

Onsite Working Results

Organization of STRENDA Commission

The question has been raised if there is anything to do for the STRENDA Commission after 20 years of activity in the field. What can we do? What are we willing to do and what could be STRENDA's future goals?

This topic led to the following discussion:

Allen: What is a data set in Brenda?

Carsten: Reaction conditions, kinetic parameters, description of enzyme, assay components (linked to chem databases). Only dataset when it has been published in a journal. STRENDA DB is still weak in kinetic parameters.

Recently developed by Andrew the field with reaction conditions was diversified in more detailed subfields. The large fields of the old entries were left as such. Perhaps use large

language model to extract the new category information from the large text field in the original large text fields.

Enzyme ML: SBML but now for the enzyme data, not (just) for the modelling. Python library as data model. Mark-down file format description of Enzyme ML. Biocatalyis guidelines are not in the data model. A next challenge will be to also integrate modeling tools. Now add quality control for data and modeling and enabling AI. Original data and analysis methods need to be associated with the enzyme.

STRENDA DB needs to be ready for calculated K_M from alpha fold. Currently only datasets that have been curated by humans and that have an experimental basis. Screening data are presently not included. BUT there will be a preference for keeping experimental and computed 'data' separate in separate databases. (The Nielsen paper predicting K_M's, states that more experimental work is now needed to train the network.). But we should keep links between experimental community and computation community.

Frank does not understand the utility of K_M and k_{cat} data. Hans explains that this would be nice to have for modeling entire pathways/ networks. Frank does not believe that this could work. Ulrike: Gotenzymes (Metabolic enzymes) is database with predicted Km and kcat values.

Back to "is there still anything to do?"

Currently: one group focusing on single task like STRENDA DB, but more projects have emerged; these are done by subgroups (like Enzyme ML, thermodynamics)

Carsten proposes to change the organization of STRENDA now formally into something that is de facto already there:

- One group: commission
- Plus working groups (per subproject)' each of these subgroups has a lead, who drives the progress and reports back to the commission
- Other members from the commission provide fresh ideas to the subgroups
- external subgroup members of subgroups (being not members of the Commission) will be mentioned/acknowledged as members of the STRENDA subgroup.

Carsten then presented a table with six working groups:

- STRENDA guidelines (CK; FR)

- STRENDA DB (CK; FR, JH)
- Biocatalysis (DR(?), PH, KR Kathryn Rosenthal))
- EnzymeML (JP, CK, KR, DR, AB, JR)
- Thermodynamics (HW, PH; CK, JR)
- Outreach: produce a slide for us to use during presentations; social media strategies; interfacing with scientific communities and industrial activities (UW companies).

With respect to subgroup 1 Carsten notes an outstanding problem with multisubunit enzymes (like the vacuolar ATPases). Solutions are suggested such as the 'complex portal'. Peter mentions that for biocatalysis there is also a mixture of pathway or whole cell of various proteins catalyzing.

In this context Alan asks what makes people submit their data to STRENDA. +20 journals recommend. Frank's lab has much experience with input of data into STRENDA DB.

What is the potential demand for STRENDA DB? Potential submitters: 5000 papers of enzymology per year, dropped to 3000. Potential consumers: 700 000 papers with proteins; SWISSProt deals with 10% of all proteins. A 1000 papers report K_M.

Frank: Could you send the propaganda to all 1000 authors? Carsten has tried for one publication, but this did not work.

Frank: The journal *Biochemistry* required enzymes to have Uniprot IDs. Can you send STRENDA DB advertisements to these authors? Ulrike: Are there many initial entries that have stopped? How to motivate people to enter data into STRENDA DB? There should be a reward. Make contributors author to a paper.

It is important that PIs use electronic lab books. Then the data can be sent to STRENDA DB automatically. Ulrike: This needs to be taught.

Hans proposed to add a seventh project: Enzyme networks. Johann: This is relevant for both systems biology and biocatalysis. Roland: And for cascades; how can we connect say 10 enzymes to produce something. Also connect to communities where organic catalysts are used. Roland further points at connecting the different types of enzymes (like metabolic, protein kinases, DNA modifying enzymes).

Roland: Biocatalysis and Thermodynamics

Ulrike: Outreach (lead, CK member).

Enzyme networks (HW, RW, JR) (Barbara?) but start at slow pace

Peter prefers to serve as member of working groups. Because at the moment only people around the table are listing themselves.

Preliminary organization:

Working groups	Aims	Lead	Members
STRENDA Guidelines	Revision of guidelines	CK	
STRENDA DB	Feature requests, specifications, test	CK	FR, JR
Biocatalysis Guidelines	Revision of guidelines	DR, KR	CK, PH, RW, JP
EnzymeML	Use case definitions, extensions	JP	CK, KR, DR, AB, JR, SS
Thermodynamics	Finalization concepts and publications	HW	CK, JR, RW, PH
Enzyme networks	Cascades, synthetic pathways	HW	RW, JR
Outreach	Creation of materials, collaborations, networking e.g. RDMkit, flyers, slide, Online course, online training materials,	UW	CK

ACTION POINTS

- Carsten: inform other STRENDA members about the new organization and ask them to which project they'd be willing to contribute.

EQUATOR Network

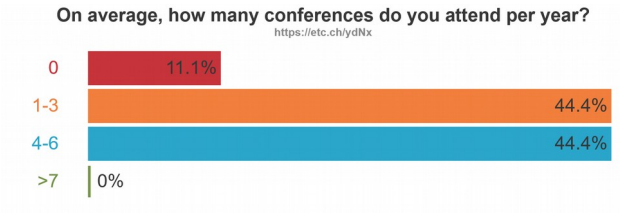
EQUATOR network: Enhancing the quality and transparency of health research. This collects online resources; promotion of guidelines, assistance icontact n development of guidelines, support. Carsten would find this our worth the while. Roland: interfacing biocatalysis with medicine is hard, but worth a try. Mutant enzymes are accommodated in STRENDA DB. Also completely synthetic sequences are possible.

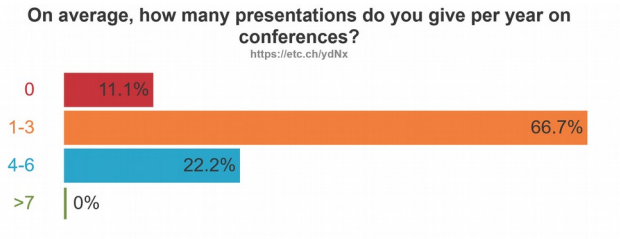
ACTION POINTS

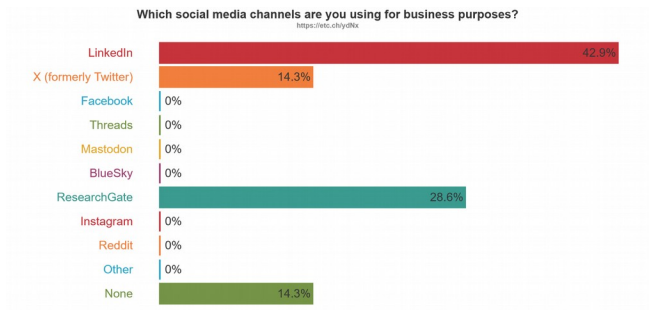
- Carsten: contact the network for possibilities for collaborations.

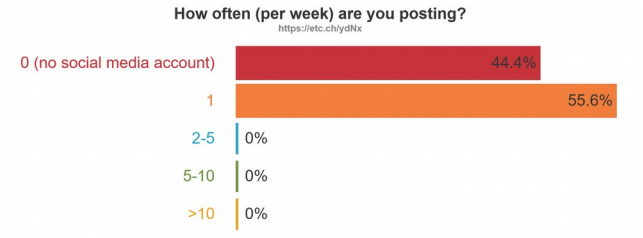
STRENDA Outreach

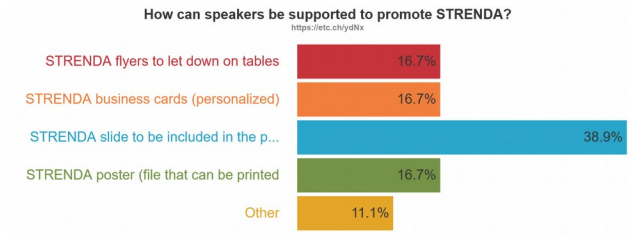
As an introduction into the topic, the Commission performed an online poll:











As a result, the Commission agreed on the development of a promotional slide that can be used in the end of a talk given by each Commission's member. Business cards, flyers and posters were not very convincing but intensively discussed.

Slide

Three outbreak groups had the task to invent – at least – a concept of such a slide that may include elements of the slide currently available (but little used) on the STRENDA websites at https://www.beilstein-institut.de/en/projects/strenda/meetings/ (scroll down to "Materials to be used by the STRENDA Commission").

The drafts are deposited on Google Drive at https://tinyurl.com/3rcjd3sk.

ACTION POINT:

- Carsten: all three slides will be open for comments to both the members of the Commission not present and to people outside of the Commission.
- Peter: will add QR code and the STRENDA logo to their slide draft.

STRENDA landscape

In the subsequent session the commission members tried to get to sketching the landscape around STRENDA. However, the group decided to first create a list of established collaborations:

- SABIO RK
- IUBMB not that perfect
- STRENDA Biocatalysis and ESAB
- BRENDA needs to be reopened
- FAIRDOM through EnzymeML (STRENDA is mentioned in the FAIRDOM standard connection) and on the FAIRDOM website. There is a link back from SABIO RK. Ulrike: One can now define metadata in FAIRdom: Metadata for STRENDA. Dataverse – Enzyme ML – STRENDA
- STRENDA DB is core repository for NFDI4Chem
- 60 journals that recommend STRENDA guidelines and their societies.
- [UniProt] also via EnzymeML
- PubMed PMC connection via its linkout.

- Nomenclature committee of IUBMB through Andrew. Not a reversible connection. Gerry Moss maintains the membership list of both committees. Andrew will put that in place.
- FAIRSHARING.Org and RE3DATA.org (repository for repositories) are linked with STRENDA

Johann: we need to get STRENDA DB submission compulsory.

Ulrike: to get STRENDA submission compulsory a community initiative is needed.

Frank: citations and H-index are motivating; so therefore people will be motivated for STRENDA. Plus, there will possibly be a link to the Journal *Biochemistry* (through Tobias Erb). Tobias may recommend the use of STRENDA DB.

Could STRENDA DB users be activated to push STRENDA DB?

ACTION POINT

- Carsten will draw the landscape and put this onto google.docs.

Presentation: RHEA and UniProt

by Alan Bridge

Enzyme annotation with Rhea and UniProt'

SIB about 900 people. Research groups and service groups. Alan's is a service group; 10 developers etc with in total about 50 members. There are 20 curators on UniProt.

Rhea is a vocabulary for talking about enzyme functions in UniProt; describes also reaction chemistry.

They use CheBI for small molecules. ENZYME uses a copy of the EC number database. They are also linked with MetaboLights and with Reactome (a pathway database).

260k publications have been curated for UniProt (?).

ECnumbers are used; they remain popuflar.

Interactome. For structures pdb and alpha fold.

Rhea is more recent and also has a standalone website. It works through chemical reactions rather than enzymes or genes; Reaction are defined through chemical structures interconversion. It has 17000 reactions.

TrEMBL 17% annotated of all enzymes (Swiss Prot half of all proteins).

The SPARQL software does the data integration. It can do tissue specific metabolic networks.

InChiKey is used for the representation of compounds

Recone3 now with much more explicit connection with the underlying literature references.

SwissLipids: based on predicted possible structure; *in silico* lipidome. Number was reduced by focusing on only the variants known to occur in human.

LitSuggest used in curation.

One seventh of articles about enzymes have been curated.

Discussion

Frank**:** Can one ask UniProt to add a certain enzyme that has not been added yet? Indeed they work towards community curation in their Open Reaction Database. AI is used but as assistant only; there is always a human curator involved. Protonation of metabolites is about to be dealt with.

Alan's advice is to have separate databases for experiment based and computed data. Otherwise people will ignore the column that annotates whether it is experimental or computational. This is happening in the use of other databases.

In the UniProt cross ref section, reference to STRENDA DB can be added.

ACTION POINT

- Carsten and Alan's production team will be in contact to establish reference in pdb to STRENDA.

STRENDA-ESAB Symposium

The Need for Reproducibility of Enzyme Reaction Data

November 14

Lecture hall at Lodz University of Technology, Lodz, Alchemium, Faculty of Chemistry

The STRENDA Project of the Beilstein Institut and the European Society of Applied Biocatalysis (ESAB) share common interests in standardization and have joined forces to organize this first STRENDA/ESAB Symposium.

The Beilstein Symposia addresses contemporary issues in chemistry and neighboring sciences by emphasizing interdisciplinarity. Scientists from a wide range of areas – often outside chemistry – are invited to present (provoking) aspects of their work for "out of the box" discussion with the aim not only to advance science, but also to enhance interdisciplinary communication.

The European Society of Applied Biocatalysis (ESAB) takes initiatives in areas of growing scientific and industrial interest and importance in the field of applied biocatalysis and identifies key topics which may be rate-limiting in the development of scientific progress and technological prospects in applied biocatalysis and takes steps to stimulate these areas.

The STRENDA/ESAB Symposium provides a forum for discussing, highlighting and learning from different perspectives the need for reproducibility of enzyme reaction data.

More information is available at

https://www.beilstein-institut.de/en/projects/strenda/meetings/strenda-esab-symposium/