Beilstein Institut

STRENDA

Minutes of the

8th STRENDA Meeting

September 24 – 27, 2012

Hotel Jagdschloss Niederwald
Rüdesheim/Rhein, Germany

by Carsten Kettner

Agenda

(as approved by the participants)

Opening and Reports
General Discussion: Criteria of acceptance
Proposal of a STRENDA Check list
Use of different terminologies
List 2: organism-specific conditions
eForm: joint test, change requests, prioritization
General discussion: Potential ways to integrate the eForm in the publication process

Table of Contents

Participants

(in alphabetical order)

- Richard N. Armstrong, Vanderbilt University, Dept. of Biochemistry, Nashville, TN, U.S.A., (RA);
- Amos Bairoch, Swiss Institute of Bioinformatics, CALIPHO Group, Geneva, Switzerland, (AB);
- Hans Bisswanger, University of Tübingen, Interfaculty Institute of Biochemistry, Tübingen, Germany, (HB);
- Peter Halling, University of Strathclyde, Dept. of Pure and Applied Chemistry, Glasgow, United Kingdom, (PH);
- Carsten Kettner, Beilstein-Institut, Frankfurt am Main, Germany (co-ordination), (CK);
- Friedrich Lottspeich, Max Planck Institute of Biochemistry, Protein Analysis, Martinsried, Germany, (FL);
- Frank Raushel, Texas A & M University, Department of Chemistry, College Station, TX, U.S.A., (FR);
- Udo Reschel, Beilstein-Institut, Frankfurt am Main, Germany, (UR);
- Johann Rohwer, University of Stellenbosch, Dept. of Biochemistry, Stellenbosch, South Africa, (JR);
- Dietmar Schomburg, Technical University of Braunschweig, Dept. of Bioinformatics and Systems Biology, Braunschweig, Germany, (DS);
- Ulrike Wittig, H-ITS gGmbH, Scientific Databases and Visualization, Heidelberg, Germany, (UW);
- Roland Wohlgemuth, Sigma-Aldrich, Buchs, Switzerland, (RW)

Unfortunately absent with valid excuse:

- Athel Cornish-Bowden, CNRS-BIP, Marseille, France, (ACB);
- Thomas S. Leyh, The Albert Einstein College of Medicine, Dept. of Biochemistry, Bronx, NY, U.S.A., (TL);
- Christoph Steinbeck, EMBL Outstation – European Bioinformatics Institute, Cambridge, UK, (CS);
- Keith Tipton, Trinity College, Dept. of Biochemistry, Dublin, Ireland, (KT)

Introduction

The STRENDA Commission, founded 2004 under the auspices of the Beilstein Institut and funded and supported since then, has devised a number of check-lists, called STRENDA guidelines, that define the minimum information which should be provided when enzyme kinetic data are published. These guideline aim at the improvement of the quality of reporting functional enzyme data to enable researcher to reproduce the data of different laboratories. Additionally, the Commission is developing an electronic data acquisition software system for enzymology data (STRENDA eForm) that serves as a portal (i) to support authors and also journals as an assessment tool on the compliance with the STRENDA guidelines with an emphasis on providing information comprehensively rather than defining acceptance criteria and (ii) to store entered functional data along with the experimental conditions in a database that will be publicly accessible.

A further aim of STRENDA is to propose uniform assay standards for the standardization of data for single enzymes and groups of enzymes. Clearly, the conditions under which an enzyme operates will depend on the organism and organelle in which it occurs. However, this issue just has been tackled when standard assay conditions for the functional characterization of the yeast glycolysis enzymes were developed and tested (van Eunen et al. (2010) *FEBS J.* **227**(3)).

Today, 28 biochemistry journals recommend authors to consider the STRENDA guidelines when reporting kinetic data. Since BMC, PLoS and *OMICS* recommend the authors to refer to the MIBBI portal for prescriptive check-lists when reporting their research data and since STRENDA is registered at MIBBI that provides information about 30 odd check-lists from diverse bioscience communities, the number of journals is even higher.

Additionally, the major database producers in the enzyme field such as BRENDA and SABIO-RK are closely connected with STRENDA as there is a close exchange of information and ideas between STRENDA and ISO/DIN, YSBN, MIBBI/BioSharing.org, IUBMB/FEBS and further organizations that foster STRENDA to become widely accepted by the scientific community.

The primary objective of this year's meeting is to present and discuss the prototype of the STRENDA eForm, to subject the eForm a comprehensive test and to define further modifications and improvements.

Opening and Reports

The STRENDA Commission Meeting took place at the Jagdschloss Niederwald Hotel in Rüdesheim, Germany, over two days and started on Tuesday, the 25th and ended in the late afternoon of Wednesday, 26th of September 2012.

After a short self-introduction of the participants CK opened the meeting with a report on the past, current and future activities of STRENDA. He then led over to an opinion poll about the expectations and suggestions of the participants aiming at a consensus on the final agenda which then included a general discussion on the criteria of acceptance of standards and a proposal to set up a check-list of the STRENDA guidelines, the different use of terminologies, and the resumption of the discussion on list 2 (definition of organism-specific conditions). There was a strong agreement to carry out a joint test of the eForm software prototype with a subsequent discussion of modifications and extensions. In addition, a general discussion about the potential integration of the eForm in the publication work flow was appreciated.

The overall goal of this meeting was the generation of a prioritised change request list of the previously discussed modifications to enable the Beilstein development team to specify and plan the further implementations.

Reports

Resignations

The STRENDA Commission is active since more than 9 years but it is faced by resignations of old members and appointments of new members from time to time. This year, two members announced to leave the Commission: Christoph Steinbeck due to an overcommitment by his projects at EBI and Keith Tipton due to age which will not allow too much travelling any more. However, both former members declared their willingness to remain available for consultation. The participants of this meeting thanked Christoph and Keith for their efforts spent to advance the work of the STRENDA initiative.

As replacement following members of the scientific community were suggested:

Antoine Danchin (Evry, France), Karen Allen (Boston, MA, USA), Barbara Bakkers and/or Karen van Eunen (Groningen, The Netherlands) and Yasuhisa Asano (Toyama, Japan).

Action: CK will contact and invite these persons to join the STRENDA Commission.

STRENDA Book Project

According to the following presentation ACB concluded in his message to CK on September 24 "Seeing it set out like this it doesn't seem as depressing as I thought it would. We have at least made some progress."

Carsten Kettner (Beilstein) and Athel Cornish-Bowden
Introductory remarks : *DONE (but probably will need some revision when all the chapters are in)*

Robert Goldberg (National Institute of Standards and Technology)
Standards in Biothermodynamics *: DONE (long ago!)*

Keith Tipton (Trinity College Dublin)
Nomenclature for enzymes and proteins
Keith has never delivered anything or answered any letters, so ACB has incorporated this topic into his chapter on the 1983 recommendations on enzyme kinetics. KT could have done it better, but ACB's efforts will be better than nothing. So we can call this DONE

Dietmar Schomburg (Technical University Braunschweig)
Standards in Enzymology – Data Integration in the World's Enzyme Information System BRENDA : *DONE*

Minoru Kanehisa (Kyoto University)
Predictive Genomic and Metabolomic Analysis for the Standardization of Enzyme Data : *DONE*

Hans Bisswanger
Enzyme Assays : *DONE*

Douglas Auld (Novartis, Cambridge, MA, USA)
High-throughput assays : *by end of November 2012*

Athel Cornish-Bowden
Analysis and interpretation of enzyme kinetic data : *DONE*

Athel Cornish-Bowden
IUBMB recommendations (1983) : *DONE (with Enzyme Nomenclature material incorporated)*

Octavio Monasterio (University of Chile)
Magnetic resonance : *DONE*

Pedro Mendes (University of Manchester and Virginia Tech)
Applications in systems biology : *Nothing heard from him for a long time. Maybe we could get Hans Westerhoff to suggest one of his colleagues (it would be great if he would do it himself, but I fear he will not have time. Someone like Barbara Bakker would be good.)*

Peter Halling (University of Strathclyde)
Industrial Applications : *Due to overcommitmentsuggests to contact the previously agreed cococo-authorsLucia Gardossi and Munashwar Gupta. He also provided also their plans for the manuscript*

Action: CK or ACB will contact L. Gardossi and M. Gupta

Amnon Kohen and Kevin Francis (University of Iowa)
Isotope effects: : *as we invited him only this year he asked for a year to do it, and we agreed to submission by the end of February 2013.*

Thomas Leyh (The Albert Einstein College of Medicine):
Electronic data submission : *Nothing so far, but probably it needs to wait until the prototype is essentially finished. Questionable if Tom will want to be the main author as the final work has been done in Germany, but of course he's welcome to do it. Otherwise probably some who was involved in the software development.*

To be decided
Post-translational modifications

Members of STRENDA
STRENDA recommendations : *This probably needs to wait. When the time comes CK and ACB can draft something rapidly and include the other members of STRENDA as authors unless they refuse.*

Copyright Agreement for Guidelines and eForm

This was just a formal act. Since the STRENDA Guidelines are published on the website of the Beilstein-Institut and since the Guidelines are basis for the development of the eForm which also is available on the Beilstein website the foundation requires an official agreement in which the members of the STRENDA Commission grant the Beilstein-Institut a non-exclusive and royalty-free license to use (i.e. copy, reproduce and distribute) the Guidelines and all future versions thereof.

The agreement was signed by all members present on site.

Recommendations

Today, 28 biochemistry journals recommend authors to consider the STRENDA guidelines when reporting kinetic data.

Since BMC, PLoS and *OMICS* recommend the authors to refer to the MIBBI portal for prescriptive

check-lists for reporting their research data and since STRENDA is registered at MIBBI that provides information about 30 odd check-lists from diverse bioscience communities, the number of journals is even higher.

ACS Catalysis
Archives in Biochemistry and Biophysics
Antimicrobial Agents and Chemotherapy
BBA (all nine sections)
Biochemical and Biophysical Research Communications
Biochemical Journal
Biochemistry
Clinical and Vaccine Immunology
FEBS Journal
Free Radical Research
Infection and Immunity
Journal of the American Chemical Society
mBio
Molecular and Cellular Biology
Nature Chemical Biology
Proceedings of the National Academy of Sciences
The Journal of Bacteriology
The Journal of Biological Chemistry
The Journal of Virology
Trends in Biotechnology

STRENDA @ FEBS Congress 2012

Thanks to the efforts of ACB, Marìluz Cardenas and one of the organizers, Prof. Miguel de la Rosa, STRENDA had the opportunity to organize a so called Specific Activity. Specific activities are part of the official scientific program of the FEBS Congress and provide a variety a events that support discussions amongst delegates, corporate partner, exhibitors and scientists.

The aim of the STRENDA Specific Activity was to set up the stage for a workshop at which the issue of non-standardized and non-comparable functional enzyme data and its use in biochemical disciplines such as systems biology, enzyme characterizations and pathway studies was discussed. Under the sub-title "Making Biochemistry Work for Life" Hans Westerhoff (Amsterdam) chaired the small number of presentations and guided through the panel discussion thereafter. CK gave a brief introduction in the STRENDA initiative and the guidelines for reporting enzymology data, and presented the eForm

software prototype.

The further program of this workshop was as following:

Luis Serrano Pubull (CRG, Barcelona)
The quality of experimental data in the literature and the complicated use of this data in Systems Biology applications

Chris Whitman (University of Texas, Austin)
The tautomerase enzyme super-family and the case for STRENDA

Richard Perham (University of Cambridge, UK and Editor-in-Chief of FEBS Journal)
A view from a journal

Aleksander Benjak (University of Heidelberg and Managing Editor of FEBS Letters)
Why guidelines for reporting data are important for scientific journals

General discussion: Getting the most out of STRENDA.
Hans Westerhoff: Chair; Athel Cornish-Bowden: rapporteur

The discussion covered the following questions?

(1) How far should the requirement of data standardization reach? All publications? All funded grants? All university work? What should be exemptions?
(2) Should there be a STRENDA data repository (STRENDA) exclusive for the STRENDA compliant data?
(3) Who (which journal) would be willing to alpha-test the STRENDA input form?
(4) Should primary data be stored in STRENDA?

Question #1 was answered in a different way. In particular, the question was discussed towards the impact of standards for science.

yes,

- standards are important
- standards are appreciated by the journals
- provided that guidelines are understandable and short but sufficient for the improvement of data quality
- provided that they can be handled by journals (see high number of guidelines)
- but STRENDA should also be concerned with the naming of enzymes (this is in fact to the EC nomenclature commission!)

no,

- if a big collection of guidelines is overwhelming for authors, editors and reviewers (see. *Nat. Chem. Biol.* references)

Question #2 was discussed in a way if databases that provide a user-interface would be preferred to repositories without user-friendly querying tools.

- repository preferred if data from STRENDA database is flagged as STRENDA compliant data in BRENDA or SABIO-RK
- one-stop-shopping preferred by community when searching for enzymology data
- however, at the beginning of the STRENDA database it also should provide a user interface for querying and retrieving entered data
- xml as data exchange format is fine

The questions #3 and 4 were skipped due to some reasons and replaced by the discussion on potential copyright conflicts and the future perspectives of the STRENDA eForm.

Regarding the issue whether there might be legal conflicts when data provided by authors will be freely available in a STRENDA database, neither the representatives from the journals nor the representative from a publisher discovered any copyright problems since (according to their arguments) only a subset of the entire data from the publication will appear in the database which is not considered a threat for the business model of the publishers ("you need the full paper to understand the experiments" - "the papers present interpreted data, not just data"). Additionally, the journals also noted that the database provides some benefits for the journals since the users are referred from the data in the STRENDA database to the original references. A very good example for the realization of the co-existence of data-housing databases and data-publishing journals is PDB which is widely accepted by the publishers.

Surprisingly enough, the capabilities of the eForm were considered very enthusiastic by the attendees:

- eForm can be such a repository for enzymology data as PDB is for structural data (this is indeed the master plan of STRENDA!)
- editorial board meeting of *FEBS J.* end of October; Prof. Perham will discuss the eForm at this meeting!
- Suggestion: seek for strong support from the systems biology community rather from the enzymology community since the latter one is fragmented and small
- *FEBS Lett.*: Special Issue on STRENDA, enzymology in general, systems biology, modelling etc.. This issue could be co-edited by the members of STRENDA.
- publication: September 2013 at the 6th ESCEC Symposium?

- The use of both the eForm and the STRENDA ID is suggested as a proof-of-principle for research papers

6th ESCEC Symposium

The 6th ESCEC Symposium will take place at the Jagdschloss Niederwald Hotel in Rüdesheim, Germany, September 16 – 20, 2013. The subtitle "*...celebrating the 100th anniversary of Michaelis-Menten kinetics*" indicates the aspects covered by the conference:

- biochemistry and molecular biology coming of age: 100 years of Michaelis-Menten kinetics, (60 years of the structure of DNA, and 40 years of Metabolic Control Analysis)
- systems biology and systems medicine: the new kids on the block
- sequence, structure, kinetics, control and regulation
- physiological meaning of enzyme kinetics
- network kinetics versus enzyme kinetics
- enzymes in metabolic collaboration
- pitfalls in data reproduction
- The end to the tower of Babel: making biochemist speak a lingua franca

Additionally, the ESCEC conference series will become 10 years old....

The speakers list already consists of 11 speakers and will be increased within the next weeks and months to up to 25 speakers.

- Karen Allen (Boston, MA, USA)
- Patricia Babbitt (San Francisco, CA, USA)
- Roger Goody (Dortmund, Germany)
- P. Fred Guengerich (Nashville, TN, USA)
- Manfred Konrad (Göttingen, Germany)
- Dietmar Schomburg (Braunschweig, Germany)
- Reiner Sterner (Regensburg, Germany)
- Keith Tipton (Dublin, Ireland)
- Hans Westerhoff (Amsterdam, The Netherlands)
- Chris Whitman (Austin, TX, USA)
- Jan-Olof Winberg (Tromsö, Norway)

The following ideas will be realized at this symposium:

- One session on the impact of MM kinetics for today's research;
- Involvement of CSMB, GBM and Humboldt University (CSMB has been contacted without any response so far, GBM is on board with Dr. Goody, the contact with Humboldt University is still pending);
- Poster session and oral presentation (5 – 7 min).

General Discussions

Fundamental Criteria of Acceptance

CK presented his ideas on the fundamental criteria of the acceptance of guidelines such as the STRENDA guidelines and he was seeking for agreement of these suggestions.. On the one hand, when presenting the STRENDA initiative he usually comments that the guidelines are not mandatory but should kept in mind when authors submit papers which contain functional enzyme data. The phrase "not mandatory" was disliked by the participants and there was strong preference to change it to "will become mandatory" or "should become mandatory" to foster the establishment of the guidelines within the community.

On the other hand there was strong support regarding the second part of the slide:

> STRENDA and other minimum information initiatives such as MIRAGE or MIAPE will also share two fundamental criteria required for its broad acceptance by the enzymology community:
>
> [1] sufficiency: The guidelines should require sufficient information about a data set and its experimental context to allow a reader to understand and critically evaluate the interpretation and conclusions, and to support their experimental corroboration.
>
> This means that the guidelines should support authors to provide information about materials and methods applied and results obtained in a comprehensive way.
>
> [2] practicability: Achieving compliance should not be that burdensome as to prohibit its widespread use.
>
> However, even guidelines should be useful for users; on the one hand they should be understandable but not too long, on the other hand.

In total, the intention of the guidelines is not to create a substitute for the review process.

Proposal of a Check list

There is a difference between guidelines and check-list. Even within the STRENDA Commission this has not been defined properly yet.

By definition (*Wikipedia, The free Encyclopedia*. Retrieved: 10 October 2012 08:57 UTC , permanent link: http://en.wikipedia.org/w/index.php?title=Guideline&oldid=493640427) a "*guideline* is a statement by which to determine a course of action. A guideline aims to streamline particular processes according to a set routine or sound practice". In this term, the STRENDA generic introduction (http://www.beilstein-institut.de/STRENDA/STRENDA_Generic_Introduction.pdf) can be considered a guideline.

In contrary, a check-list is defined as a type of informational job aid used to reduce failure by compensating for potential limits of human memory and attention. It helps to ensure consistency and completeness in carrying out a task (Source: *Wikipedia, The Free Encyclopedia*. Retrieved: 10 October 2012 09:00 UTC, Permanent link: http://en.wikipedia.org/w/index.php?title=Checklist&oldid=514560211).

In this sense, the STRENDA Guideline lists 1A and 1B can be considered check-lists, however without providing check-lists for "done", "still pending" etc.

Check list for Authors / Reviewers

List Level 1A

defines data that are recommended for the methods section for publishing enzyme data. This information should allow the reproducibility of the results.

Manuscript Title .....

Corresponding Author(s) .....

STRENDA ID (if available)

Data	Definition	yes	If yes,on which page of your ms.	no	NA
Identity of the enzyme					
Name of reaction catalyst	name, preferably the accepted name from the IUBMB Enzyme List	☐	☐	☐	☐
EC number		☐	☐	☐	☐
Sequence accession number		☐	☐	☐	☐
Organism/species & strain	NCBI Taxonomy ID	☐	☐	☐	☐
Additional information on the enzyme					
Isoenzyme	naturally occurring variant				
Tissue					
Organelle					

Figure 1. Proposed check-list on basis of the STRENDA Guidelines.

CK brought up the idea to provide authors and reviewers the guidelines as a check-list which contains text boxes to include the manuscript title, the name(s) of the corresponding author(s) and the STRENDA ID (if available) (see figure above). Additionally, tick boxes are suggested to be included that ask the authors/reviewers for each line if the corresponding information is available, if yes, on which page.

After a short discussion this proposal was refused by the participants due to the following major reasons:

- too burdensome for authors
- redundant to the eForm
- Why should authors (and/or reviewers) complete both the check-list and the eForm?
- What would be the benefit for authors to go through the check-list? Who at the editorial office should do this?
- Page number (information available on page xy) can change during the publication process

The Consistent Use of Terminologies

HB reported about his observation that the STRENDA Guidelines obviously do not comply totally with the IUBMB Recommendations which he applied when writing text books. In particular, he noted K_m which often is written as K_M and V (maximum velocity) which is commonly expressed as V_{max}. Regarding the appearance of V_{max} in the STRENDA Guidelines HB is right to complain about this since IUBMB discouraged the use of V_{max} since in the mathematical sense a limit is defined rather than a maximum.

> Action: The aspects Guidelines as described above need to be modified in accordance to the IUBMB recommendations (*European Journal of Biochemistry*, 1982, **128**:281-291).
>
> All: approval: change V_{max} to V.

Additionally, HB suggests to replace the designators for products (P1, P2,..) and substrate (S1, S2) by those used by the Cleland Nomenclature (A,B,C...P,Q,R). Even though these designators are not used explicitly in the STRENDA Guidelines, S1, S2,...P1, P2 are used in both versions of the eForm.

> Action: all, compliance with IUBMB recommendations in this regard needs to be discussed.

General agreement: any differences between IUBMB recommendations and STRENDA guidelines should be avoided to avoid confusion of the authors by the requirement contradictory rules.

The review of the Guidelines (Level 1A) showed a misleading comment in the line of the description of the enzyme activity. The phrase

> "Ideally k_{cat} otherwise expressed as amount product formed per amount enzyme protein present - sometimes referred to as enzyme unit or international unit (1 IU = 1 μmol min^{-1}). The katal (mol/s) may alternatively be used as a unit of activity (conversion factor 1 unit = 16.67 nkat)."

should be changed to the suggested one:

> "k_{cat}, if the enzyme concentrations are known. Otherwise IU (International Unit, 1 IU = 1 μmol min^{-1}) is acceptable."

Reason: the use of katal is not accepted any more.

Action: all, approval of suggestion above or suggestion of an alternative (better) comment.

List 2: Organism-specific Experimental Conditions

The so-called List 2 has been created in 2004 to separate the reporting guidelines from the proposition of experimental conditions which need to be defined in dependence of the organism used. However, already in 2004 there was a general consensus that this list will need the expertise from scientists who would be willing to provide their knowledge about the requirements for the establishment of appropriate rules and thus this issue was put on hold.

DS reminded that a reopening of the List 2 should be done as soon as possible. The participants agreed with him but also noted that the situation did not change since 2004: external expertise will still be required. CK suggested to start with by discussing organizational issues (creation of a sub-group within the STRENDA Commission?, identification of experts who would be willing to contribute, etc.) first before starting with the creation of the List itself. Finally, further discussions have been deferred.

Action: none so far.

The STRENDA eForm

Comments from participants at the ESCEC Symposium 2011 on the eForm developed by Tom Leyh *et al.* in 2011

drafted by Peter Halling, 15th September 2011

- **General feeling**. Need to be very strict on number of required fields - more tolerance for optional fields that user is allowed to leave blank.
- Beware of having items that are considered more **conclusions** than data - but not easy to draw line. e.g. kinetic constants at different pH values might be better than pK's deduced from them. What about inhibition type - could be regarded as conclusion?
- Mixed views on whether the form should include a **free text box**. Valuable for entering details of conditions or results not covered by form - but don't want to end up with whole paper pasted in as text! Probably do need some box that is free text as far as system goes, but where user is warned to make strictly limited use of it.
- **Wording** on form should always make it absolutely clear what the user is asked to provide - problems here are key aspect of "unfriendly".
- Must be able to return entered data in easily **human readable form** - either for user who has entered data to see that is properly recorded, or to convince others that form is working.
- How do we handle H+ (and H2O) in reaction equations?
- Perhaps boxes to tick saying more data can be found in paper (e.g. for full pH profile)
- **Tabular data**: Form operates more on modular basis. That means, if I have entered full details describing one or more enzymes I have studied, these are recorded in system, and then when I'm entering more data, I am asked something like "Please select which of your enzymes this is for, or enter details of a new enzyme". The same could go for a reaction module, perhaps a conditions module. Then on entering results, I might select with 3 clicks a combination of previously described enzyme, reaction, conditions - and then enter results for this combination.
- **LIMS:** Could the eform write part of the experimental section of my paper for me?
- Default values can aid user friendliness, but beware of user accepting wrong defaults without thinking. So e.g. if I have chosen an enzyme, the system might say, "Have you studied the reaction: XXXXX?". Only if I say yes is this automatically entered.

Conclusion: Requirements for a new eForm prototype

- **rapid and easy entry of data**, i.e. less than five minutes per data set. The data set consists of the description of the protein along with the results at one given experimental conditions set.
- entry of experimental data sets which represent e.g. pH dependent kinetic profiles. In general, the ability to enter data from tables without the repetitive entry of already entered data is essential.

Development: The Deliveries

Due to the shortness of time (3 months) of the development of the prototype at the Beilstein-Institut the developers together with CK defined a number of items which are required to be implemented or realized:

- background: development of a prototype that is based on TL's form
- simple design of the interface comparable to the existing e-Form;
- auto-fill function of defined fields using connections to web-based freely accessible scientific databases which provide the data as public-domain data. Extraction and inclusion of the relevant data in the eForm;
- help text where necessary;
- entry of data from tables;
- elementary plausibility control as proof-of-concept;
- display of all entered data for proofreading (but without editing wrong entries);
- generation and output of the STRENDA ID;
- storage of this data in a database;
- output of the data entered as XML file (readable by other data repository systems);

Development: The Non-deliveries

Due to the limited amount of time, a number of decisions have been made to create a simple software prototype which remains capable of further developments. The hope was to cover at least 80% of the requirements defined by common experimental designs and results:

- no tab-separated masks as presented in version 2 (developed by the BRENDA team) as agreed at the 6th STRENDA meeting 2010.
- complex database interface to query data

- special cases of data entries won't be considered. Such cases are e.g. unknown protein or protein name, protein without UniProtDB entry, only the reaction is known (e.g. crude extracts)
- no user support regarding the input of protein modifications; this section needs to be completed manually by the author, basic help will be provided,
- only the assay conditions are variable to map data ranges in tables. Assay conditions include substrates, products, buffers, salts, other compounds, pH, temperature as well as inhibitors and activators along with their concentrations.
- The effect of sequence changes on initial rate parameters will not be considered. Alternatively, such scenarios can be represented in different experiments

Modifications after Video Conference in March 2012

In March 2012 the first version of the prototype was presented during a video conference even though some requirements were still open at this time. The following items have been finished since then:

- Matching EC Numbers complete / partial DONE
- Automatic completion of TaxonID DONE
- Remove compound: last compound in the prototype DONE
- Temperature and pH: plausibility check DONE
- Initial rate parameters: true / apparent DONE
- Finalize Experiment: results for inhibition and activation parameters DONE
- Help texts where appropriate and necessary DONE
- Tool tips (mouse-over info, e.g. for compounds) still open, if still necessary
- Confirmation page: create STRENDA ID DONE
- Error handling DONE (relying on error messages from the server)
- Browser back (and back button) not yet implemented
- „wait“ sign (e.g. hourglass) to indicate search process still open, if still necessary
- Browser tests (IE, Safari, Firefox) partly DONE

The following items were discussed and the participants agreed with their implementation during subsequent developments:

- •Back buttons, where appropriate
- •Editing/correction before storage to database

- •Remove compound: last compound in the prototype, possible for each other compounds in later versions
- •User handling: registration / log in → yes, required when pausing or for later corrections/additions
- •Data base query interface → simple interface providing simple query functions will be sufficient

The Joint eForm Test

Based on a "real" research article the participants carried out a joint software test of the eForm. The eForm can be accessed via the following URL: http://195.227.158.199/eform/ or via the web site of the Beilstein-Institut: http://www.beilstein-institut.de/en/projects/strenda/e-form/ (select "View Prototype Form").
The test case was to enter the relevant data from this research article (Marchand, M. et al. (1998) Glucosephosphate isomerase from *Trypanosoma brucei*. *Eur. J. Biochem.* **184**:455-464) and to complete a questionnaire (see appendix). The results and comments were compiled as follows:

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
Start page	Introductory text answering the question why the user is here	X	
Manuscript title:	Would be useful to relate different proteins to one publication; otherwise many StrendaIDs would be related to one manuscript		
Authors	Format definition is needed (full name or initials of first name) □ use PubMed format	X	
Values	Format definition is needed (7.5 or 7,5?)	X	
Values	Should be possible to have empty fields (not to have fill in "0") exceptions: pH, Temp., fixed concentrations,	X	

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
UniProtID The ID is the entry name which can change.	The field called "UniProtID" should be called "UniProtKB AC" Help text: define UniProtKB And you should not show the entry name but the AC!	X X	
The search in UniProt should be done with a number of possibilities: Name Author EC number		(X)	X
UniProt entry with multiple EC numbers. In some cases you enter only one of them (the first one). Example: P08907 In some case you enter all of them. Example: P14060 In the case that more than one is entered, no reaction are entered.	Statement in the help text: at the moment we cannot handle the following		X
How to deal with multi-component enzymes, e.g. heterodi/tri/x-ers?			X
UniProt search	Back button at UniProt search more visible → different colour → on the bottom, too		
Subcellular compartment	You could load the data available from the UniProt entry.	(X)	X

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
Protein modification	Should be changed to "Protein sequence modifications"	X	
Protein modification	Should be yes, no or unknown	X	
Protein modification	Left of dots sometimes represented as. | → agreed alternative: see line „Sequence modifications: yes“		
Protein modification	Spelling check is ON → needs some investigation to check out how to overcome with it, depends on browser add-ins (Safari by default? Firefox add-on that needs to be installed additionally	(X)	X
And the help text to	"Does the protein contain any sequence modification(s) in comparison to that of the UniProtKB entry?"	X	
if you load the sequence from UniProtKB [...] you should not allow the sequence field to be edited.		X	
Sequence modification: yes	New field(s) expand(s), auto-filled with the native sequence, allowing modifications	X	
Sequence alignments			X
There should be an additional part of the form on "Post-translational modifications"	Example: see Fig. 2	X	
Cross-check of modifications			X

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
Cell lines (different from cell types) along with the definition of the organ		(X)	X
Compounds	In substrates: "CID" -> "PubChem CID"	X	
ChemDraw structure	Additional information!	(X)	X
EC Number	Plausibility check at ExplorEnz (does this EC number exist?)	(X)	X
Organism	Plausibility check at NCBI Taxonomy (not complete) or http://www.organismnames.com/ e.g. typos	(X)	X
Tissue	Plausibility check at BTO (does it exist)	(X)	X
Compartments	Should be replaced by „Localisation“	X	
Strains	New field	X	
Relationship tissue – organism?	Plausibility check: does the tissue exist in the organism		X
Varies	Replace by variable „initial value varied“	X	
Substrate. Variable and fixed. When input is variable it was not clear how to I put the range of concentrations	Range of concentrations: fields for minimum and maximum concentration. Help text	X	
Temperature	No SD	X	
pH/pD	No SD	X	

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
Vmax vs. Vm	Consistency check throughout the form	X	
Macro molecules and other substrates	New field(s): initially free text field (can be modified depending on the entries during the test) --> tick box: protein, peptide with sequence (One-letter-code) --> polysaccharides (ID from database – which to recommend?) --> DNA, RNA by sequence or database identifier (if known) → Peptides with UniProtKB AC	X	
Recommendation: set „variable“ for all compounds to change later variable as default but can be changed to constant	Or statement on the meaning/consequence of setting constant/variable	Open (tbd)	
STRENDA ID	STRENDA specific to avoid mixing up with Ids from other databases, e.g. SIDxyz	X	
Required fields	Consistency check for the * label throughout the form not relevant anymore		
Kinetic Parameters	Parameters fitted: kcat, Vm, kcat/Km, Vm/Km, Km, Ki (substrate inhibition). Only the first two should be compulsory fields – in principle should also make at least one of the final six compulsory, but very unlikely that all will be left blank.	X	

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
Kinetic Parameters	Ki: „Substrate Inhibition“ must be made clearer	X	
When type of inhibition already defined remember for additional values	Keep the initial definitions for inhibition types when conditions are changed	X	
Affinity constant: Affinity not essential for activating substances	Ka not required	X	
Saturation: not always reachable for activation substances	defined concentration if not saturating → entry of defined concentration (additional field) and set saturation as default	X	
Kinetic constants: Variable substrate/product concentration(s) nonsense for just one Km/Vmax	Help text (?) declaring that only the exchange of the compounds themselves are useful when entering Km/Vmax additional fields: minimum / maximum concentrations used (connection of both fields with the kinetic params)	X	
Buffers	Add on main form „Please see help text for advice on how to enter buffer details“	X	
Buffers – Help text	Please select from PubChem an appropriate salt form of the buffer compound, with the correct counter-ion (e.g. Tris hydrochloride, potassium acetate). Then give the total concentration of the buffer compound in both protonation states, as in the usual convention (i.e. not the actual concentration of the individual PubChem	X	

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
	compound). Together with the pH value entered elsewhere, this will fully characterise the buffer system used. In the case of buffers like phosphate, either the mono- or di-salt can be entered. If a buffer is made by, for example, adjusting Tris hydrochloride with NaOH, the additional NaCl generated may be declared as an added salt. Note that Tris base is named TROMETHAMINE in PubChem.		
EC reaction vs. entry of substrates and products: no link between those fields and therefore no automatic data transfer from reaction to substrates/products	The reaction according to the EC number is often vague and contains just compound classes rather than explicitly the compound names. Therefore the EC reaction has been called „default reaction“ but this reaction can be assayed different in the experiment	clarified	
“Default reaction” doesn’t encourage me to change it.	Should read “Reaction studied (edit default shown if necessary)”.	clarified	
Default reaction	Addition of declaration: “Reaction as in data base – you will enter actual substrates and products used later” field non-editable and not required	X	
Reaction	check reaction and substrates + products automatically for consistency		X

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
Product: adding any products in initial conditions?		clarified	
Assay components: "Varies" and "Concentration"	"Initial Concentration Varied?", with answer Yes/No. Then "Fixed initial concentration used" – only required if not varied.	X	
Assay Components: PubChem hits	Free text field to allow to enter chirality info	X	
Limiting Rate Parameters: Neither of us had ever heard this term. Later expansion shows they are Km and Vmax etc	Change to Kinetic Parameters	X	
Protein Identification: Commercial product	"If a commercial product, state vendor's name, product code and/or lot no.	(X)	X
Search for Proteins: Protein Identification	Should be renamed to "Search string"	X	
Assay method / Description of the Methodology	New field, short description, giving an example	X	
Reactions in crude extracts			X
Bisubstrate reactions Vmax and Km values of a specific depends on the conc.. how to handle grid datasets with the concentration of both/all substrates changed?	Comment: needs more input from JR		X
Progress curves, time-course data	Comment: needs more input from JR		X
Registration page	Name, affiliation, email address, approval by activation link	X	

Comment on field (description)	Change request	Priority high (by mid of 2013)	Mid term
Overview	Print/save to PDF. This file could be used as supporting info for papers	X	

Several possibilities of how to realize post-translational modifications in the eform have been discussed and there was a general agreement that the fields of the sequence modifications could be used as a template for PTMs which is illustrated in Fig. 2.

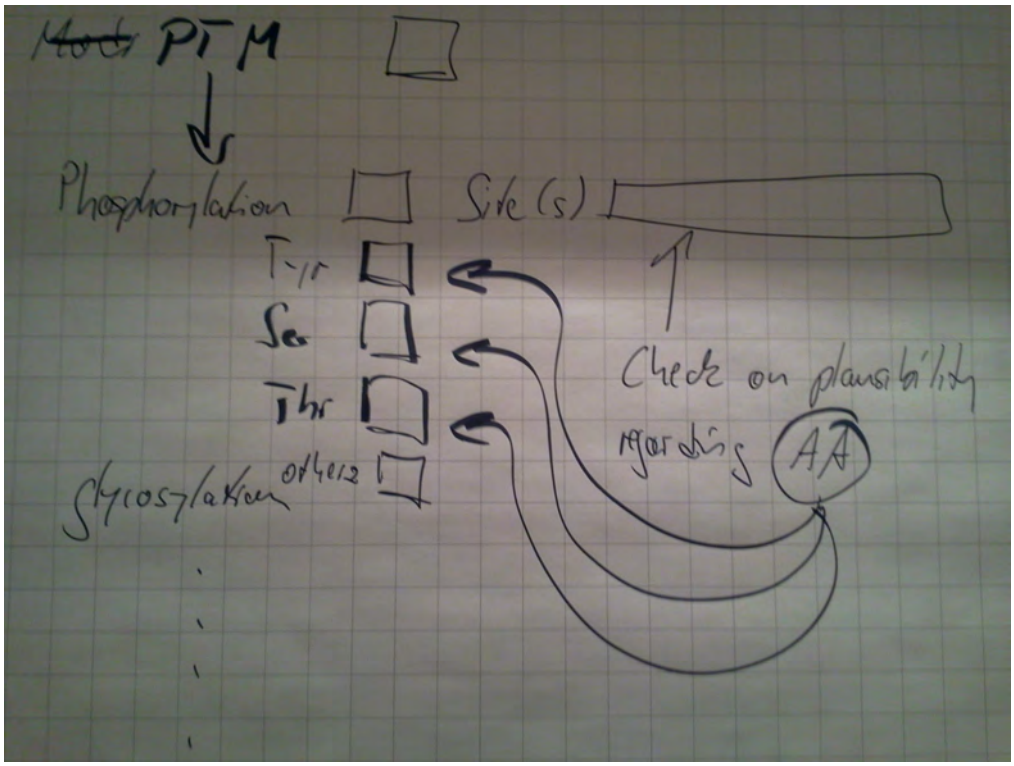

Figure 2. Example for the implementation of fields for PTM.

For each kind of modification (i.e. phosphorylation, glycosylation, methylation etc.) corresponding fields are provided. Fields for phosphorylation were worked out in detail: if the protein was phosphorylated, the user clicks in the tick box and further fields open providing tick boxes for amino acids that are phosphorylation sites (Tyr, Ser, Thr) accompanied by free text fields in which the amino acid numbers are entered. These entries could be checked on the plausibility whether the amino acid identified is the appropriate one for phosphorylation.

Similar solution could be implemented for glycosylation, methylation and other modifications.

Change Requests: Next steps

This list of change requests will be the basis for the following steps:

- sorting of the items according the complexity for the implementation;
- definition of working packages;
- prioritisation packages regarding the release date (latest mid of 2013);
- specification of the changes;
- presentation of the specification to the members of the STRENDA Commission and seeking for comments, suggestions for improvements and approval;
- implementation and test;
- presentation of the software product to the STRENDA Commission;
- release as alpha version;
- information of the community (and journals) about this release;
- gathering and subsequent analysis of comments for subsequent software versions.

The Integration of the eForm in the Publication Process

The Commission was concerned with the integration of the eForm in the work flow of the daily publication process. There was a general agreement that data should be entered by the authors rather than by the members of the editorial office. Additionally, it was consensus that only published kinetic data should be available from a public data base. Those data that are still under review will remain in a non-public („closed“) database. This data can be reviewed, edited and modified if necessary by the author and the author (or somebody else) will be in charge to add the bibliographical data to the experimental data after publication of the manuscript.

In detail, CK presented the data flow as follows:

- Data Input: prior to publication (or latest during the manuscript processing) the author(s) enter their data into the form;
- Data Control: automatic assessment according to the STRENDA Guidelines since the form covers the most important aspects; if the form is complete the data for publication are guideline compliant. Additionally, data should be automatically controlled on plausibility;
- Storage: Data are temporarily stored in a non-public (closed) database;

- STRENDA ID: automatically generated by the system.
- (Re-)Submission: The author submits his manuscript together with the STRENDA ID. This ID indicates to the Editorial Office that the data within the manuscript are STRENDA compliant.
- Editor's Decision: editor decides whether to accept the manuscript or not
  - if not (rejected), the author can resubmit the manuscript to another journal using the same ID.
- if yes (accepted), the manuscript (ms) is passed to the referees.
- Review of the manuscript and the referees recommend whether to accept or to reject the ms.
  - if not (rejected), the author can resubmit the ms. to another journal.
  - if yes (accepted), the ms. can be published.
- Publication: the ms. is processed and made available to the public
- Bibliography: bibliography data (journal name, issue, page numbers etc.) is added to the dataset entered previously by the authors
- Storage: the complete dataset is stored in a database and made available to the public. The temporarily buffered data in the closed DB are deleted.

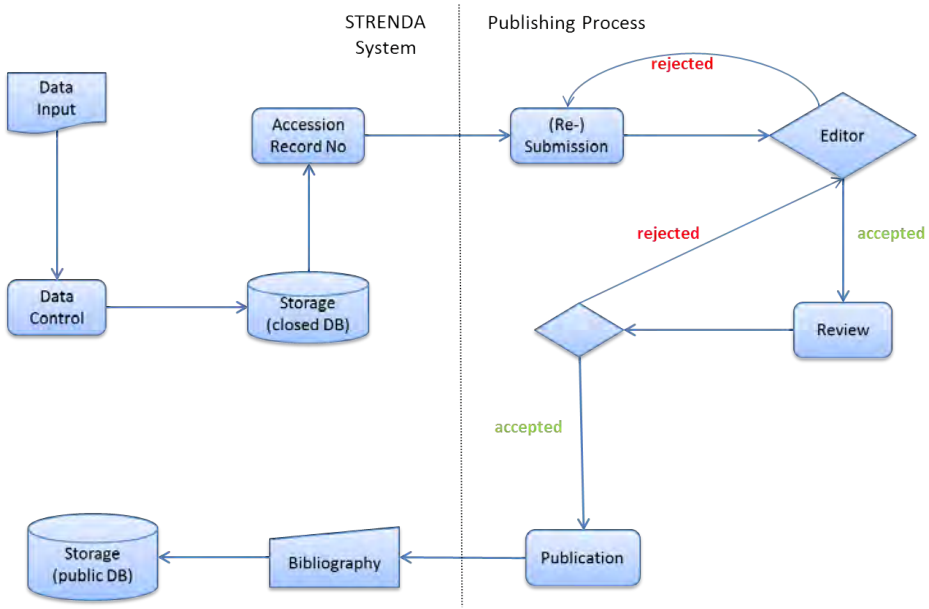

Figure 3. Possible integration of the eForm in the publication work flow.

There were no objections to this model.

RA announced his intention to introduce the alpha-release of the eForm on the web site of *Biochemistry* provided that the webmaster of this site would be willing to integrate it both appropriately link to

eForm on the Beilstein pages and an introductory text. The use of the eForm will be declared as non-manadatory but will be promoted by an editorial. However, the tool will just be offered as a quality assessment tool to enable authors to check if their data are in compliance with the STRENDA Guidelines. The tool is not intended to store entered data in a database. A PDF file presenting the entered data will be accepted as supplementary information and could be used for reviewers as an overall overview on the kinetic data.

Action: CK will contact P.F. Guengerich from JBC asking for participation in the alpha-release test of the eForm.

Incentives of the eForm

Regarding the question on the value of the eForm with and without a database it was discussed how to deal with the situation if any journals on the one hand would recommend the authors to use the eForm but on the other hand would not be willing to support the storage of this data in databases. This lead to the question which incentives to the users are provided by the eForm (including the database model).

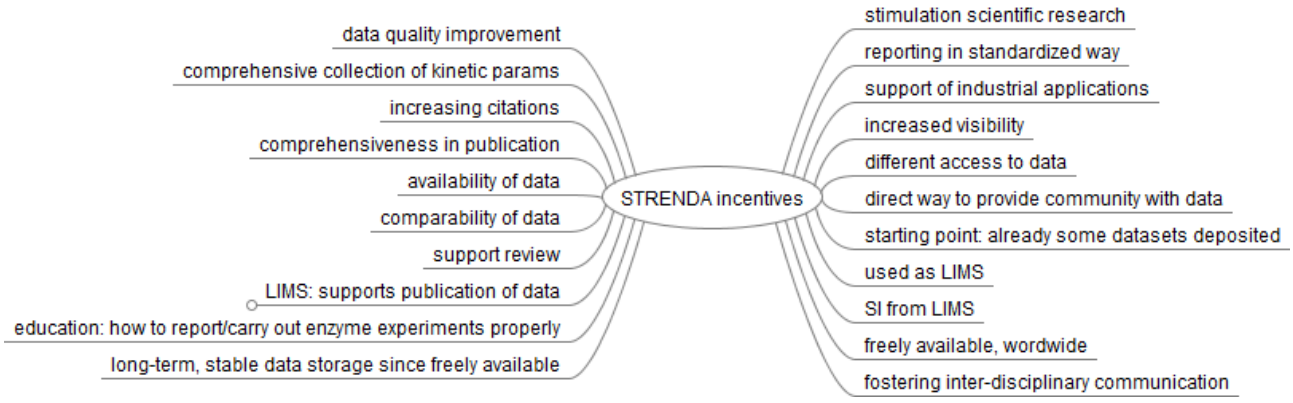

Figure 4. Mind map on the incentives of the eFom.

A mind map (Figure 4) was created to collect all ideas that support the discussion on the impact of the eForm for the scientific community, in particular for authors and editors. This mind map has not been discussed concludingly and therefore includes a number of redundant entries. However, this discussion should be resumed at the next STRENDA meeting since the clear presentation of incentives for the community is inherently important for the process of promotion and conviction for acceptance of the eForm.