



By Mark Andrew CC BY 2.0 via Wikimedia Commons

Minutes of

STRENDA

the 14th **Meeting**



18 and 19 September, 2018

Manchester Institute of Biotechnology, The University of Manchester
Manchester, UK

By Dr. Carsten Kettner
Beilstein-Institut, Trakehner Str. 7 – 9, 60487 Frankfurt/Main, Germany
ckettner@beilstein-institut.de

Table of Contents

List of Participants	3
The Agenda	4
Results	6
Overview	6
Last Developments in STRENDA DB	8
Status of STRENDA DB	8
Data Transfer from STRENDA DB to SABIO-RK	8
Enforcing STRENDA Guidelines and STRENDA DB	8
Comparison of PDB with STRENDA DB	9
Manuscript Review	10
Incorporation of STRENDA DB in Instructions to Authors	11
Strategies and Tasks	13
Results of Data Rate Group	13
Results of Journals Group	14
Data Exchange Format	14
STRENDA DB – Change Requests	14
STRENDA DB – Extensions	15
Definition of Task List	15
Increasing the data rate in STRENDA DB	15
Increasing the number of journals actively supporting	17
Appendix	19

This minutes provide you with the topics and results discussed at the 14th STREND meeting held at The University of Manchester in September 2018.

Any further information on the STREND project is available on the project's website at www.beilstein-strend.org.

List of Participants

1. Barbara M. Bakker, University Medical Center Groningen, The Netherlands
2. Paul F. Fitzpatrick, The University of Texas Health Science Center, San Antonio, TX, USA
3. Peter Halling, University of Strathclyde, Glasgow, UK
4. Carsten Kettner, Beilstein-Institut, Frankfurt am Main, Germany
5. Thomas Leyh, The Albert-Einstein-College, Bronx, NY, USA
6. Jürgen Pleiss, University of Stuttgart, Germany
7. Frank M. Raushel, Texas A&M University, College Station, TX, USA
8. Johann Rohwer, University of Stellenbosch, South Africa
9. Santiago Schnell, University of Michigan, Ann Arbor, MI, USA
10. Neil Swainston, University of Manchester, UK
11. Ming-Daw Tsai, Academia Sinica, Taipei, Taiwan
12. Ulrike Wittig, Heidelberg Institute for Theoretical Studies, Germany
13. Roland Wohlgemuth, Lodz University of Technology, Poland
14. Hans Westerhoff, Vrije Universiteit Amsterdam, The Netherlands
15. Sameer Velankar, EMBL Outstation, European Bioinformatics Institute, Hinxton, UK

Unfortunately absent

- Toni Baici, University of Zurich, Switzerland
- Athel Cornish-Bowden, CNRS-BIP, France
- Dietmar Schomburg, Technical University Braunschweig, Germany

The Agenda

(as approved by the participants)

Tuesday, 18th

9.00 AM Welcome and Opening

- Introduction and information

Overview of past activities

- Contacts with journals
- DFG proposal
- Presentations on conferences

Carsten Kettner, Ulrike Wittig, Peter Halling, Jürgen Pleiss

Status report

- Publications
- Reviewers' comments - *Peter Halling*

- Last developments in STRENDA DB
- Status of STRENDA DB

- Issues with data transfer into SABIO-RK - *Ulrike Wittig*

1.30 PM Enforcing STRENDA Guidelines and STRENDA DB, requirements, obstacles and challenges

- The lead example for STRENDA DB – PDB - *Sameer Velankar*

Development of Strategies and definition of tasks

- a) Increasing data rate into STRENDA DB
- b) Increasing number of journals recommending STRENDA DB

Means and leverages, carrots and sticks to make use of STRENDA DB mandatory

Wednesday, 19th

9.00 AM Opening

Reconciliation of yesterday's results and agreements

STRENDA DB – data exchange format

- BioCatNet data for STRENDA DB - *Jürgen Pleiss*
- EnzymeML - *Santiago Schnell*

STRENDA DB – changes requested

Query section:

- organism/expression system
- protein, native/modified

STRENDA DB – Extensions

- Suggestions for detailed operation and wording – *Peter Halling*
- Equations, raw data and progress curve data sets – *Johann Rohwer*

Definition of Task list

Results

Overview

CK gave a brief overview of past activities and implementation of tasks agreed at the 13th STRENDA meeting in September 2017. He reported of visits at a number of journals and publishers which aimed at making the journals and publishers aware of both the STRENDA initiative and STRENDA DB and convincing them to recommend their authors to make use of the STRENDA Guidelines and to enter enzymology data into STRENDA DB. When in Boston, CK had intensive discussions with representatives of the following journals and publishers: Cell, Cell Chemical Biology, Cell Metabolism, Cell Biochemical Sciences, Cell Systems and Current Biology. He received much interest in the issue and positive responses as well as many interesting suggestions including the retrospective data input by the community. The major demand was that the journals seek for strong demands from the community before applying both the Guidelines and STRENDA DB and suggested CK to carry out a survey on the impact of Guidelines and whether STRENDA DB should be used. The Cell representatives agreed to discuss the recommendation of STRENDA and the decision will be made soon.

In addition, CK was in contact with Molecular Systems Biology but received the reply that the four EMBOPress journals only rarely publish papers that contain enzyme activity data. CK had the opportunity to make L. Kiessling, (ACS Chemical Biology) aware of STRNDA. She showed very interested, received additional material but no further action was taken. Last but not least, CK was in contact with F. Hollmann, member of the editorial board of Molecular Catalysis. Hollmann was very interested in agreed to bring this topic up in the editorial board meeting the following week.

CK reported about his proposal to apply for funding support from DFG. He's seeking for support for (1) a computer technician who is planned to develop STRENDA DB and (2) a scientist who is intended to become in charge for promoting, marketing and support of STRENDA DB. However, BI showed very hesitant to support this proposal as the directors noted that the data input rate is very low and there are still uncertainties regarding the adoption of STRENDA DB by the journals.

This issue has been intensively discussed by the STRENDA Commission and both the data input rate and the adoption by the journals have been identified as major tasks that need to be addressed in the future work of the Commission.

CK reported about a number of presentations he gave (PTB, Physikalisch-Technische Bundesanstalt; Open Science Days by May Planck Digital Library; Poster presentation at

Moosbach Kolloquium by German Society of Biochemistry and Molecular Biology). He received very much positive responses from the community and had very interesting discussions after the presentations.

UW reported from the results of discussions after she gave an overview of STRENDA DB at H-ITS and she provided the Commission with a list of issues identified of which some are based on misunderstandings in terminology or “marketing” of STRENDA DB and some based on real findings when comparing SABIO-RK and STRENDA DB. Much of these issues already are resolved or are easily to resolve (see appendix).

PH reported from his discussions at the CECAM Workshop on “Proteins in realistic environments: simulation meets experiments”. He received much interest and even experimentalists expressed a wish to enter data into it. However, they also indicated that most of them could not do so in the present version, at least not completely, as some have only progress curves, not kinetic constants; others have kinetic constants for a fitted model but this is more complex than accommodated (e.g. two substrates, four products, seven parameters). See for more in the appendix.

JP reported from his discussions at a biocatalysis meeting and has experienced similar responses from the community as the members above. The community appreciates the need for standardization but it is not sure how to handle this issue. There was also the question of the immediate use of depositing data in STRENDA DB and which were the ways to make it an immediate value (e.g. assigning DOIs to datasets → which is already done by STRENDA DB).

In summary, STRENDA DB has much impact but both the communication with the community needs to be improved and the database needs to be extended to meet the demands of the community more efficiently.

CK presented the latest publications from the STRENDA Commission, among them the paper on STRENDA DB (published in FEBS J), the correspondence with Nature on the impact of STRENDA DB and offering the SpringerNature group the service making data sets complete, and the analysis on omissions of information in published papers (published in Biophysical Chemistry). In their first year, the full papers have gained much attention with 253 and 66 reads on ResearchGate.

Last developments in STRENDA DB

From the discussion at the previous STRENDA meeting, there was the demand to slightly change the wording in STRENDA DB. The members felt unhappy with the terms experiment and experimental subset. This has been changed in the meantime, with keeping “Experiment” due to documented reasons and the replacement of Experimental Subset to ‘Dataset’.

Status of STRENDA DB

So far, ~30 datasets are published in STRENDA DB. Each dataset published is advertised on Twitter with the author names, the journal of the original work and the DOIs of the datasets in STRENDA given. There are ~10 datasets finalized and are awaiting publication in STRENDA DB, and there is still a reasonable number of datasets created but without any progress. These datasets will be deleted in due course.

Users who have entered data in STRENDA DB were asked how long they took for the entire input process. The answer was ~30 min once you are familiar with the system.

Data transfer from STRENDA DB to SABIO-RK

UW reported about the attempt to transfer data from STRENDA DB to SABIO-RK and found some inconsistencies between both databases as well as between the STRENDA DB entries and publications. For more, see appendix.

Enforcing STRENDA Guidelines and STRENDA DB

CK reported that currently more than 55 biochemistry journals recommend the STRENDA Guidelines, and more than 10 to 15 journals recommend their authors to also share their data using STRENDA DB. The odd estimate derives from the fact that some publishers/journals share the same policy. For example, there is a data policy for all Nature journals that include (as journals relevant for enzyme activity data) Nat. Biotech., Nat. Chem., Nat. Microbiol., Nat. Systems Biol., etc. The same is true for the PLoS journals (with e.g. PLoS One, PLoS Biology, PLoS Computational Biology etc.).

CK proposed that the next step(s) could be to transfer STRENDA from the recommending level into a mandatory level (via adoption on voluntary basis) and in this context Sameer Velankar (EBI, Cambridge) gave an overview of the history of PDB. The development of PDB serves an example for the potential route for STRENDA DB.

Comparison of PDB with STRENDA DB

In the subsequent discussion, the Commission compared STRENDA DB and PDB in order to find a way to accelerate the adoption of STRENDA DB by the community.

Table 1. Comparison of PDB and STRENDA DB

Common	Differences
Data used by consumers, not necessarily producers.	Raw data not currently supported by STRENDA-DB, while PDB does.
Both dependent on raw experimental and derived data (e.g. kinetic assay progress curves and kinetic parameters).	Specialized enzymology community is too small, but general user-base of STRENDA is large.
Data provides confidence in quality.	Structural biology data producers generally more expert than kinetics data producers.
Derived data consumers primarily less interested in raw experimental data. (?)	PDB data deposition is required for publication. STRENDA-DB entry is not required.
Reliant on community engagement.	PDB more engaged with external engagement than STRENDA currently.
Need to support / engage with broad user-base of data producers.	PDB data submission system more mature, integrate with instrumentation output, provide pre-processing tools, support multiple experimental data formats.
Raw data can be reanalysed to re-derive different derived data.	Kinetics assay raw experimental data less standardised, more heterologous.
Both PDB and STRENDA are free and open-access.	PDB initialised from a smaller, more focussed community.

Data producers rewarded by data submission, supports additional citation?	Enzymology community traditionally less computationally focussed than structural biology community.
Both systems support reproducible science.	Unlikely to get new insight from reanalysing raw experimental kinetics assay data, unlike re-analysis raw structural data. (This was disputed by Barbara.)
	STREND-DB can be considered a pre-publication validation tool (primarily of metadata).
	PDB validates interpretation derived from experimental raw data.
	PDB perhaps more diplomatic in communicating with journals.
	Publication <i>requires</i> PDB id before publication. At best, STREND submission is a recommendation before submission.
	PDB integrated with Elixir.

The cells labelled green characterise the only impactful differences between PDB and STREND DB that need to be addressed by the Commission, i.e. outreach with the community and rapid and easy input procedures supported by data exchange formats.

Manuscript Review

FR suggested to enforce the use of both the Guidelines and STREND DB when a member of the Commission is reviewing a paper: He will make a statement to the reviewing editor (not to the author) when he believes the new kinetic data reported in the manuscript should be submitted to STREND DB.

He asked whether it would be appropriate to have a standard sentence (from the STREND hierarchy) or two that could be used to convey this request to journal editors when reviewing manuscripts. Even though there was broad consensus on this

procedure, the Commission decided not to create an universal statement as the contexts may vary. Thus, it will be up to each member to find the right words.

FM added that if this reviewer's request is subsequently ignored by the journal editors, he's going to decide to put such a request directly in the review to the authors. If this request will be ignored, too, it is very doubtful that future reviews will be made for that editor.

This procedure is supported by the Commission and PH provided a suggestion for a three-step approach that might be considered:

1. The journal already recommends STRENDA-DB.
Ask the authors to deposit in accordance with the recommendation and quote the reference number or DOI.
2. I can identify at least one missing item that STRENDA-DB would catch.
Point out missing item, say STRENDA-DB will catch omissions like this, and ask authors to use it to check for any others.
3. Otherwise, perhaps at this stage a mild recommendation to use STRENDA-DB would be reasonable

The Commission agreed to this approach as well but let it open to each member to act individually.

SS added that he found himself in the situation of explaining to both the authors and editors the importance of depositing the enzyme assays in STRENDA DB (or at least using the STRENDA Guidelines to report the assay conditions) but he admitted that he has not been very successful up to this point even with those journals that already recommend the STRENDA Guidelines and DB.

Incorporation of STRENDA DB into instructions to authors

PH proposed some suggestions for wording and mechanisms that might be used from journals to incorporate STRENDA DB into their instructions to authors. He pointed out that the key wording is important. It should be preferred "require" or something similar but also appreciate that journals may not be willing to go this far yet. The hope at least will be for "strongly recommend".

In detail, he suggested a first statement like:

“Details of experiments on enzyme function should be submitted to STRENDA DB (<https://www.beilstein-strenda-db.org/strenda/>) in order to ensure that the description is complete, with all essential meta-data included.”

Followed by:

“Authors should include the STRENDA reference number(s) (SRNs) of the data included in the manuscript.”

“Authors should include in Supplementary Materials the pdf file(s) generated by STRENDA DB which summarise the experimental conditions and results in their manuscript.”

If the journal uses headings to identify requirements on particular topics, the heading “Enzyme Function Measurements” could be suggested.

If the journal uses checklists or similar for authors, reviewers and/or editors, it should be asked that these include a question in relation to STRENDA DB, e.g.

“If the manuscript includes enzyme function data, has this been submitted to STRENDA DB?”

The journal may already require information that is also captured by STRENDA DB: UniProtKB identifier(s) for the protein(s) studied; full sequences of these proteins; details of post-translational modifications; EC numbers; PubChem identifiers; InChi strings. If so, it could be indicated that this requirement is satisfied by STRENDA DB entry. For example, items could be asterisked with a footnote like:

“* If these relate to enzymes and compounds used in enzyme function experiments, the requirement is satisfied if there is a reference to a STRENDA DB entry that will include this information.”

The Commission agreed with this proposal.

Strategies and Tasks

- a) Increasing data rate into STRENDA DB
- b) Increasing number of journals recommending STRENDA DB

As the Commission has identified the two above issues the most important for the future development of STRENDA DB (which is in accordance to the findings from the comparison with PDB) two out-break groups were formed to address either of these issues. The results were presented to the entire group and discussed.

Results of the data rate group:

What can the strenda commission do?

Check papers with enzymology recently and convince people to submit! Sabio-RK could offer support. E-mail people who recently published enzymology data. Ulrike has overview of scientists with e-mail addresses since 2-3 years. Carsten plus commission can draft a Newsletter. It should be personalised, with the name of the recent publication. They may upload already published data, but can also be made aware for uploading future datasets.

Where is external help required?

Neil can help Ulrike to get the mailing semi-automated. We may need help from a communication expert to avoid the letter lost in spam. (Suggestions to avoid this: feature Beilstein logo prominently, include Strenda commission members at the end of the letter with photographs, affiliations.)

Are there any external stakeholdes with a strong voice?

We need enzymologists with a strong voice (Ken Johnson and ..). They can advocate Strenda by:

- Writing a letter to many journals
- Writing editorials
- Uploading a dataset and very explicitly state this in their paper

Other stakeholders: biochemical engineers or systems biologists with a strong voice

What can Beilstein (CK) do?

Hire a student to support people in the submission process. (Ming Daw offers that he can also hire a person). Teaching about Strenda DB in enzymology courses. Beilstein could take the lead in developing a tutorial, but any of us could present this at a conference/course. Beilstein could even offer a summer school for enzyme kinetics.

What can Strenda offer?

... on a short/long term perspective

draft task list available

Results of the journals group:

- Focus initially on 3-4 key journals, perhaps JBC, FEBS J, Biochem
- Invite key editors to our meeting
- A STRENDEE presents to Editorial Board meeting – we have members on many
- For trial period, offer help to enter to STRENDA DB from submitted papers
- Keep publishing “white papers”, letters etc on STRENDA and DB
- Count number of enzyme function papers not using STRENDA and/or DB
- Tick box for authors saying where you have deposited data
- Details of presentation in instructions to authors that make it easy to check, e.g. submit STRENDA DB pdf, accession code
- Reward junior authors (cash? recognition?)

Data exchange format

JP briefly introduced in the BioCatNet, the modelling platform for biocatalytic activity kinetics. He proposed the development of an exchange format for this data for the data transfer from the bench to the publication, including modelling platforms such as BioCatNet and COPASI and databases such as STRENDA DB and SABIO-RK.

He together with SS proposed the development of EnzymeML for which first workshop is planned to be held in November 2018 in Stuttgart.

The Commission agreed with this proposal.

STRENDA DB – Changes requested

It has been noted that the query section provides some uncertain entries due to historic reasons in which the input section was modified but these modifications have not been implemented in the query section. This especially concerns the term (host) organism when an expression system has been used to express a protein. The second issue is that the query system does not reveal whether the report was about a native or a modified

protein. It is just shown the UniProtKB AC and the name of the protein but there is no information about the nature of the protein.

CK is asked to arrange the modification.

STRENDA DB – Extensions

JR proposed to extend STRENDA DB by the following functionalities:

1. Allowing the user to specify the kinetic equation used to fit the data, including inhibition experiments.
2. Allowing the inclusion of raw data (e.g. initial rate data).
3. Dealing with progress curve data sets.

In particular, for the first bullet point he identified the most important one, JR provided a very detailed solution for the implementation in STRENDA DB which is very close to that included in COPASI.

The Commission was hesitant to agree and thus, this issue need further examination and discussion.

Definition of Task list

The data input rate in STRENDA DB and the number of journals actively supporting STRENDA have been identified the major issues for which tasks have been defined. In the following, there are two task lists each addressing one of the issues mentioned above.

Table 2. Increasing the data rate in STRENDA DB

Task	Definition	Who? By when / when to start?
Upload of published data by community	<ul style="list-style-type: none"> • Check papers with enzymology recently and convince people to submit! • Potential source: SABIO-RK could offer support (can provide contact details, 	CK, UW, TL, BB, NS

Task	Definition	Who? By when / when to start?
	<p>Ulrike has overview of scientists with e-mail addresses since 2-3 years)</p> <ul style="list-style-type: none"> • E-mail people who recently published enzymology data. • Neil can help Ulrike to get the mailing semi-automated. We may need help from a communication expert to avoid the letter lost in spam. (Suggestions to avoid this: feature Beilstein logo prominently, include STREND A commission members at the end of the letter with photographs, affiliations.) • Carsten plus commission can draft a Newsletter. It should be personalised, with the name of the recent publication. They may upload already published data, but can also be made aware for uploading future datasets. 	
Advocacy for STREND A DB	<p>We need enzymologists with a strong voice (Ken Johnson and ..). They can advocate STREND A by:</p> <ul style="list-style-type: none"> • Writing a letter to many journals • Writing editorials • Uploading a dataset and very explicitly state this in their paper <p>Other stakeholders: biochemical engineers or systems biologists with a strong voice</p> <ul style="list-style-type: none"> • Identification required • Encourage to engage • Eventually coordination of wording in letters and editorials required. 	<p>Requirement: more data in DB</p> <p>CK, RW (identification persons, contact)</p> <p>Next step: encourage persons to vote for STREND A DB</p>
Upload (new) data	<p>Hire a student to support people in the submission process. (Ming Daw offers that he can also hire a person). Aligned with “Package Deal” (s. task list ‘Increasing number of journals’)</p>	<p>MDT: entering data from ABB CK</p>
Education	<p>Teaching about STREND A DB in enzymology/DMP courses.</p>	<p>All (teaching), CK: STREND A</p>

Task	Definition	Who? By when / when to start?
		“play ground” online?
Education (ctd)	Development of tutorial, → compilation material (but any of us could present this at a conference/course) Deposition and using enzyme data	BB, UW, JR
	Beilstein could even offer a summer school for enzyme kinetics. → Organization (place, date, invitations, etc) by Beilstein <ul style="list-style-type: none"> • Teaching by STRENDees • Participation based on application, • Travel grants • Low fee + boarding and lodging 	CK: Check for potential clashes (FEBS Course)

Table 3. Increasing the number of journals actively supporting

Task	Definition	Who? By when / when to start?
“Package Deal”	Prerequisite: increased number of datasets in STRENDA DB	
Focus initially on 3-4 key journals, perhaps JBC, FEBS J, Biochemistry, ABB, Biochem J.	Select journals, based on which criteria	HW
Count number of enzyme function papers not using STRENDA and/or DB	<ul style="list-style-type: none"> • Access journal(s) • Retrieve/read papers • count 	PH (2), JR
	Invite key editors to our meeting	Alanna Shepartz for Enzymology Symp.

Task	Definition	Who? By when / when to start?
	A STRENDee presents to Editorial Board meeting – we have members on many <ul style="list-style-type: none"> • Contact with journal • invitation, • presentation and discussion • follow-up discussion until decision 	After Paul's trial phase, data number has increased
For trial period, offer help to enter data to STRENDA DB from submitted papers	<ul style="list-style-type: none"> • select journal(s) to support, • create a workflow, • define what to do • contact the journal 	PF
Keep publishing “white papers”, letters etc. on STRENDA and DB	<ul style="list-style-type: none"> • What to write? • Where to submit to? • Does it include blogging and social media? 	NS, SS, UW, JP: e.g. Why we need a data exchange format?
Tick box for authors saying where (whether) you have deposited data (and if yes, where).	Comment: not clear where this tick box will appear.	PH: preparation paragraph to be sent to journals
Details of presentation in instructions to authors that make it easy to check, e.g. submit STRENDA DB pdf, accession code		PH: preparation of paragraph
Reward junior authors (cash? recognition?)	Define the threshold, Specify award (T-shirt, talk at conference, travel grant,...) Define “junior”	CK

Appendix

1. UW' report: STRENDA DB data in SABIO-RK

Questions from discussion after HITS internal talk in Heidelberg, Germany and talk at de.NBI Summer School (*Riding the Data Life Cycle*) in Braunschweig, Germany

- change data later (DOI is fixed)
 - while publication/reviewing process
 - revise errors
 - versioning in STRENDA DB?
 - who is allowed to change the data? (author, reviewer, curator?)
- what is the difference between STRENDA (DB) and BRENDA (DB)
- „DB“ is confusing → validation tool
- how long Beilstein will be able to host the server/database and provide the data
- what about if *Nature* (or any other publisher) is interested in the tool plus the data and takes over (no longer open access?)

General problems:

in STRENDA DB

- expressed in *E. coli* = organism *E. coli*
- all inserted values (temperature, pH, concentrations, parameters) automatically with one decimal place (e.g. 100.0) → this implied precision is maybe not correct
- PubmedID is not included in the exported data set → no reference to paper!

in SABIO-RK

- currently linkage to more than one data source (STRENDA DB + publication) not possible
- currently not allowed to have the same publication more than once

Specific problems:

while transfer of current data from STRENDA DB to SABIO-RK:

(already published papers; data were not checked by editors/reviewers!)

- incomplete reactions (missing H₂O)
- no products given for the reaction or sometimes only in „Comment for Protein Reaction“ as free text
- no buffer details
- missing organism (protein from rat expressed in *E. coli* → STRENDA DB overview shows *E. coli* as organism)
- cofactor specified as „Salt“ as part of the buffer
- same parameter values but different units in STRENDA DB and publication
- missing EC number
- missing cofactor
- wrong UniprotID (different organism)
- wrong substrate compared to publication
- not all kinetic data from publication inserted in STRENDA DB

Improvements for STRENDA DB:

- add PubmedID to data set
- interlink entries of the same publication? (e.g. 3 entries and same PubmedID)
- delete one decimal place (e.g. 100.0) which is automatically inserted for temperature, pH, concentrations, and parameters
- distinguish between original organism/strain and host organism/strain
- distinguish between compound role and chemical property (inhibitor and salt)
 - add cofactor as compound role
- allow changes during publication process

2. PH's report from CECAM Workshop

I found considerable interest in STRENDA-DB, and a wish to see it succeed from everyone I spoke to. The experimentalists all expressed a wish to enter data into it. Unfortunately, most of them could not do so in the present version, at least not completely. Some have only progress curves, not kinetic constants. Others have kinetic constants for a fitted model, but this is more complex than accommodated (e.g. two substrates, 4 products, 7 parameters).

Several attendees (notably Florian Hollfelder) are generating large amounts of data using automated methods. This is of course an important general trend. STRENDA-DB needs to be able to accept such data by electronic transfer from instruments, custom software etc – with perhaps manual addition of essential meta-data when necessary. After further discussion I came to the view that the way to do this is perhaps to define a standard format for data transfer (exactly analogous to PDB file format). STRENDA-DB should be able to input or output in this format. If/when it becomes accepted other software (from instrument manufacturers, or general experiment management) would want to be able to output in this format. It could be XML, but there is something to be said for the more human-readable type of system as used by PDB. Another possibility is CSV, which can be read or written from a spreadsheet if wanted. This all links also to the view, already raised in discussions with Carsten and Juergen Pleiss, that STRENDA-DB should link to a Laboratory Information Management System (LIMS) that is appropriate for enzyme function studies. Indeed, STRENDA-DB could be the basis for (part of) a very effective LIMS.

I noted that a valuable extension to STRENDA-DB will be plausibility checks on entered data, to flag up possible mistakes in entry for authors to check. A good suggestion made, I think, was that if data identified as unlikely is indeed confirmed as correct by the author, they be offered a text box (optionally) to comment on or explain the issue. Data that is unexpected but correct is often the sign of an important discovery.

People liked my STRENDA-DB T-shirt. Can I get one if I enter my data into STRENDA-DB? Not a bad idea – a reward for anyone who enters a certain amount of data during the early phase?

When enzymes are used for organic synthesis, the data record should be linked to databases concerned with available synthetic reactions and retro-synthetic opportunities (e.g.

REAXYS). Perhaps we should think about a general field that shows links to related data in a wide range of other external databases.

A couple of people raised the issue of whether STREND-DB should be able to make public only part of the data entered in a Manuscript. For example it might have been used to enter data during drafting, and only some of this data is eventually included in a submitted paper. Or referees/editors might object to the validity of part of the data, so it doesn't all make it to the paper. This could be handled if there was a mechanism to split a Manuscript into two – probably initially just two identical copies, with a simple means to delete data from each one and then re-finalise.

A couple of people (from the Engineering side) wanted to enter quite complex models that had been fitted to enzyme function data, with details of equations, model selection between alternatives, reasoning etc.