

EML
Research

Sabio-RK, SDBV, this meeting

An overview

Isabel Rojàs/Wolfgang Müller
Scientific Databases and Visualization Group
EML Research, Heidelberg, Germany

Sabio-RK user meeting, June 15th/16th, Heidelberg

- Scientific Databases (and Visualization) group

- Motivation of Sabio-RK

- Motivation of this meeting

- Sabio-RK and the meeting program

- A glimpse on the future of Sabio-RK

- ~10 people, focus on Sabio-RK

- Some people with lab experience, some computer scientists

- Funded by

- EML-R

- Klaus Tschira Stiftung (KTS)

- SysMO-LAB,

- SysMO-DB

- HepatoSys

In more detail...

➤ Curators (a.o.):

- Martin Golebiewski
- Renate Kania
- 1/2 Dr. Ulrike Wittig

➤ Related projects:

- Henriette Engelken
- Dr. Olga Krebs
- Saqib Mir

➤ Programmers:

- (Sylvestre Kegne-Kugne)
- (Saqib Mir)
- Dr. Wolfgang Müller
- 1/2 Dr. Isabel Rojàs
- 1/2 Dr. Andreas Weidemann

➤ DBA:

- Heidrun Sauer-Danzwith

Summarizing

$$[PLC]' = k_7[Ca] - k_8 \frac{[PLC]}{([PLC] + K_9)}$$

$$[Ca_{cyt}]' = (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{12} * PLC + k_{13} * [Ca] - k_{14} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} - k_{16} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)}$$

- Interdisciplinary group
- Differing experiences
- Specialists *both* in bio *and* in computer science

- Fun stuff 😊

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

Motivation of Sabio-RK

➤ Need for much concise data:

- E.g. Modelling

➤ Legacy data: Most information in publications

➤ Need for data in one place:

- kinetic laws for the reaction rate with their respective parameters
- environmental conditions
- concentrations of reactants and modifiers
- organism, tissue and cellular location

Previous work (Isabel's flashback slide)

- We had developed a database for the storage of the metabolism of Mycoplasma pn.
- This mainly concentrated on metabolic reactions, pathways, proteins (acting as enzymes), genes, etc.
- The core model from this database was then extended to include information about kinetics
- I.e. the origins of SABIO-RK date from aprox. 2000, with several and constant revisions to the model

Goals

➤ Kinetic data inside

➤ must be easily accessible and interchangeable

➤ should be available in a standard format

➤ should be related it to its environmental context

➤ Links to the outside

➤ Annotation to controlled vocabularies and ontologies should be incorporated

➤ Links to complementary databases should be made available

➤ Access it the way you like

➤ Webapp

➤ Web services

Solution

- **Hand curated data**

- **Need bio knowledge for succinctness and structuration**

- **Work on interchange:**

- **Use of SBML**

- **Miriam links**

- **Links out to other DBs**

- **Sabio-RK controlled vocabs and ontologies**

- **Use JAVA as strong basis for webapp**

Motivating this meeting, part I

What can you expect from this meeting?

- Find out our current goals
- Lobby for your goals: Tell us your
 - feature wishes
 - bugs
- Learn about how other people use Sabio-RK
- Learn about tools associated with Sabio-RK

Motivating this meeting, part II

What do we expect from this meeting?

➤ Getting better

- Learn how you use the system
- Prioritize new features
- Learn about related tools

➤ Reinforcing the community

- Us: Meeting the guys and gals who file these annoying bugs 😊
- You: Knowing who fixes those annoying bugs 😊
- You&Us: Building a network of people who know how to use Sabio-RK

$$[Ga]' = k_1 + (k_2[Ga]) - k_3 \frac{[Ga][PLC]}{([Ga] + K_4)} - k_5 \frac{[Ga][Ca_{cyt}]}{([Ga] + K_6)}$$

$$[PLC]' = k_7[Ga] - k_8 \frac{[PLC]}{([PLC] + K_9)}$$

$$[Ca_{cyt}]' = (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{12} * PLC + k_{13} * [Ga]$$

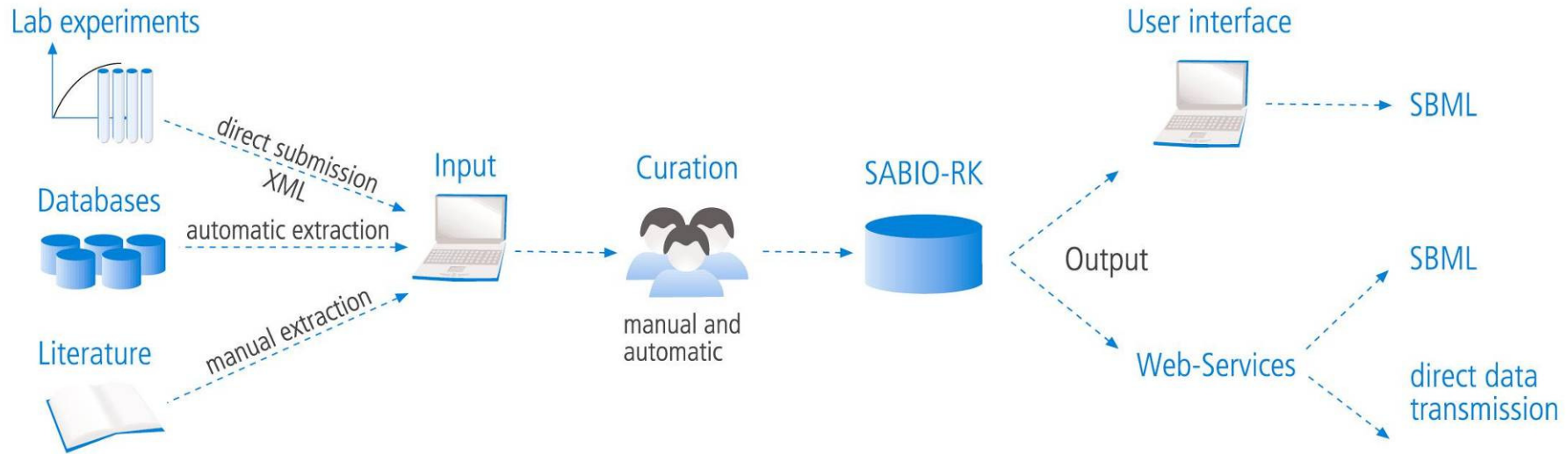
A quick Sabio-RK walkthrough

$$- k_{14} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{15})} - k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} - k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} + (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

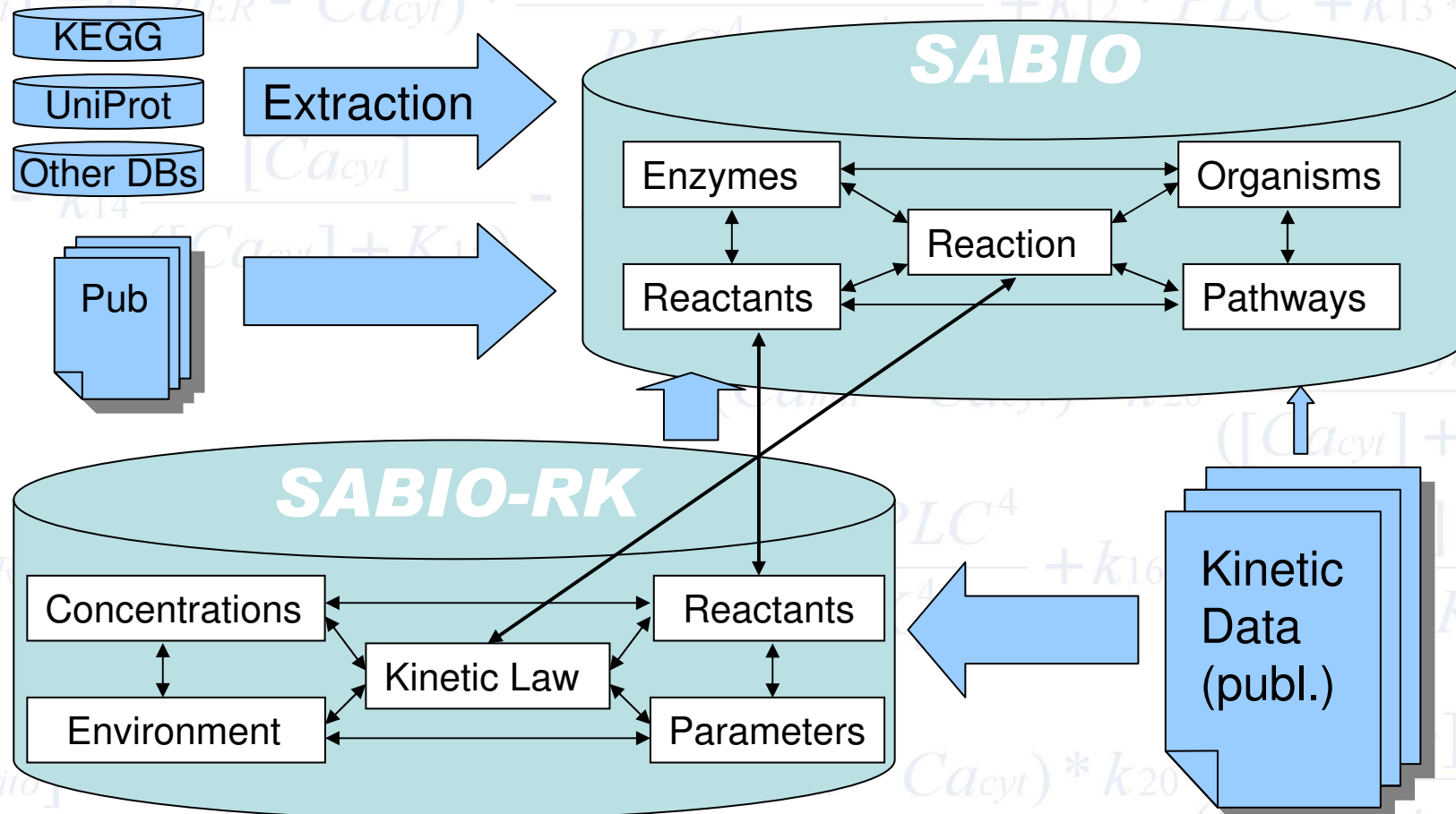
The overview



- Data input
- Curation and annotation
- Storage
- Access

History: why this name?

SABIO-RK describes **R**eaction **K**inetics and is an extension of **SABIO** (**S**ystem for the **A**nalysis of **B**iochemical Pathways)



Problems in curation process

- Missing or only partial information

- incomplete reactions (products not mentioned)

- assay conditions missing or reference to another paper

- kinetic law (or fitting equation) not described

- Complexity in the description of buffers

- e.g. coupled enzyme assay

- Identification of compounds, reactions and enzymes

- fuzziness of definitions, conflicting names

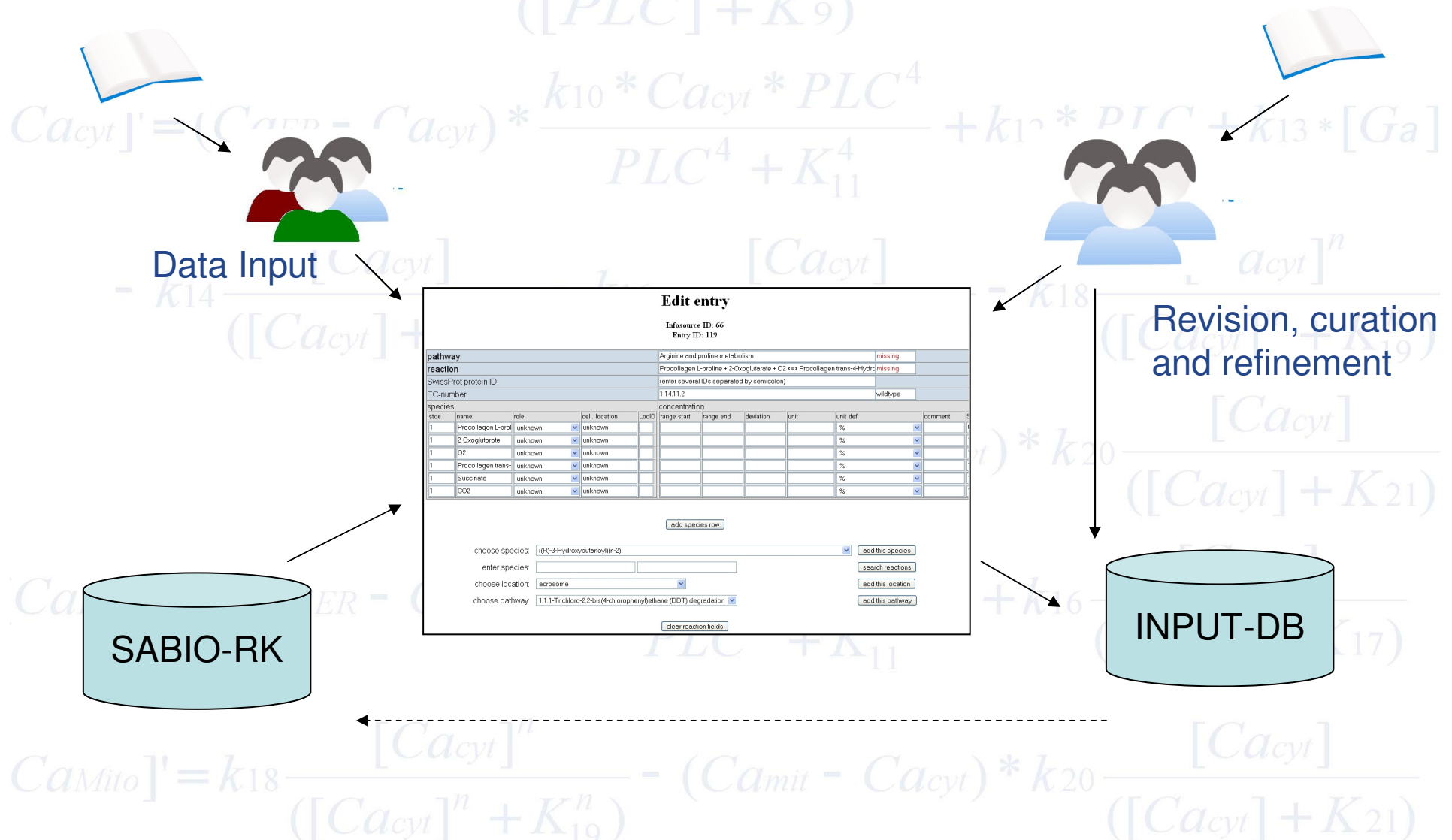
- isoenzyme not specified

- no clear unit conventions for the expression of parameters and result values

- Kinetic law types

- no controlled vocabulary available (changing SBO)

Curation workflow



Input interface

At the center of Sabio-RK entry

- Facilitates use of closed vocabularies via drop-down lists
- Facilitates structuring of data by form fill-in
- Facilitates collaboration between curators and data entry people

Compound names

p-Formylbenzoic acid
p-Carboxybenzaldehyde
4-Carboxybenzaldehyde
Terephthalaldehydic acid
Terephthalaldehydic acid
4-Formylbenzoic acid
4-Formyl-benzoic acid

Multiple names
one compound



Advanced compound name matching

New and still under development: Automatic data submission via web service

- Works: submit fresh experimental data for inclusion into Sabio-RK

- Web service

- Authentication/authorization

- Work in progress: Display what's arrived at our place

- Todo:

- Support workflow between curator and submitter

- Share rights

Last but not least

$$[PLC]' = k_7[Ca_{ER}] - k_8 \frac{[PLC]}{([PLC] + K_9)}$$

Unit normalisation feature

$$[Ca_{cyt}]' = (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{12} * PLC + k_{13} * [Ga] - k_{14} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{15})} - k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} - k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} + (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

➤ Normalize many units

(Make values comparable)

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

Hear more about it...

➤ Tutorials

➤ Data input

➤ Closed vocabs

➤ Compound name matching

➤ Automatic data submission

➤ Talk

➤ Hannan Messiha („the other side of automatic data submission“)

$$[Ga]' = k_1 + (k_2[Ga]) - k_3 \frac{[Ga][PLC]}{([Ga] + K_4)} - k_5 \frac{[Ga][Ca_{cyt}]}{([Ga] + K_6)}$$

$$[PLC]' = k_7[Ga] - k_8 \frac{[PLC]}{([PLC] + K_9)}$$

$$[Ca_{cyt}]' = (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{12} * PLC + k_{13} * [Ga]$$

Data access

$$- k_{14} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{15})} - k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} - k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} + (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

Possibilities of data access

➤ Web interface

➤ Search reactions

➤ Get details

➤ Add data to shopping cart

➤ Export to SBML

➤ Web services

➤ Discover services via WSDL

➤ Simple datatypes only

SABIO-RK User Interface

Querying the Database

CONTACT | HELP | IMPRINT Reaction Search

Return only reactions having kinetic data matching all criteria (blue and grey)

Search criteria in blue are used to define the search conditions for reactions, independently if there is or not kinetic data for these reactions.

Specify Search Criteria: Submit Search Reset Form

with **Reactants(s)** [+][-]

alpha-D-Glucose Join entries with AND or OR

- alpha-D-Glucose [+][-]
- alpha-D-Glucose 1-phosphate [+][-]
- alpha-D-Glucose 1,6-biphosphate [+][-]
- alpha-D-Glucose 1,6-bisphosphate [+][-]
- alpha-D-Glucose 3-phosphate [+][-]
- alpha-D-Glucose 6-phosphate [+][-]
- alpha-D-Glucose-1-phosphate [+][-]

related to **Protein** (UniProtID) [+][-]

in **Organism(s)** [+][-]

in **Tissue(s)/Cell Type(s)** [+][-]

in **(Intra/Extra)Cellular Location(s)** [+][-]

having **Kinetic Data** Determined for Specific Experimental Conditions [+][-]

having **Kinetic data** [+][-]

<input type="checkbox"/> Vmax	<input type="checkbox"/> Km	<input type="checkbox"/> Vmax/Km	<input type="checkbox"/> kcat
<input type="checkbox"/> kcat/Km	<input type="checkbox"/> Kd	<input type="checkbox"/> rate const.	<input type="checkbox"/> enz. activity
<input type="checkbox"/> EC50	<input type="checkbox"/> Hill coefficient	<input type="checkbox"/> Hill constant	<input type="checkbox"/> S half



Data statistics

Coarse grained:

Publications 2175

Structured entries 24962

Rate equations 12342

Distinct reactions 3455

Enzymes 1944

Fine grained:

Reaction rates (rate constant, kcat or Vmax) 30123

Affinity constants (Km, Kd, or Hill constants) 22467

Inhibition constants (KI or IC50) 30123

Hear more about it...

➤ Tutorials

- Data access tutorial
- Web service tutorial

➤ Talks:

➤ WS:

- Akira Funahashi
- Matthias König
- Peter Li
- Ismael Navas-Delgado

➤ Web interface + Modelling:

- Ursula Kummer
- Ulf Liebal

➤ Searching reactions:

- Paula de Matos

$$[Ga]' = k_1 + (k_2[Ga]) - k_3 \frac{[Ga][PLC]}{([Ga] + K_4)} - k_5 \frac{[Ga][Ca_{cyt}]}{([Ga] + K_6)}$$

$$[PLC]' = k_7[Ga] - k_8 \frac{[PLC]}{([PLC] + K_9)}$$

$$[Ca_{cyt}]' = (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{12} * PLC + k_{13} * [Ga]$$

Collaborations

$$- k_{14} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{15})} - k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} - k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} + (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

Types of collaborations

➤ Ongoing

➤ Collaborations via Webservice

- You help me, I build a nice tool/include your data in my tool
CellDesigner, BioUML, SBMM...

➤ Common tool building

- Sycamore
- MeMo/Sabio-RK
- SBML2LaTeX

➤ ...

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

Types of collaborations

➤ Ongoing (continued)

➤ Data need collaboration

➤ Curation services from Sabio-RK

➤ Specification of data need

➤ HepatoSys, SysMO-LAB

➤ Extension

➤ Signalling

➤ Theses

➤ Saqib Mir: Prof. Dr. Steffen Staab, ISWeb, Uni Koblenz

➤ Henriette Engelken: PD Dr. Uwe Reyle, INLP, Uni Stuttgart

Types of collaboration

➤ „Working on it“

➤ Compound name matching

➤ Information Extraction + Sabio-RK

➤ More experimental data entry

$$\begin{aligned}
 [Ga]' &= k_1 - k_2[Ga] - k_3 \frac{[Ga][PLC]}{([Ga] + K_4)} - k_5 \frac{[Ga][Ca_{cyt}]}{([Ga] + K_5)} \\
 [PLC]' &= k_7[Ga] - k_8 \frac{[PLC]}{([PLC] + K_9)} \\
 [Ca_{cyt}]' &= (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{12} * PLC + k_{13} * [Ga] \\
 &\quad - k_{14} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{15})} - k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} - k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} \\
 &\quad + (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})} \\
 [Ca_{ER}]' &= - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} \\
 [Ca_{Mito}]' &= k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}
 \end{aligned}$$

The meeting program

- Invited speakers

- Power users

- Tool builders

- Other members of information landscape

- Tutorials: From Data to User interface

- Data input

- Curation & annotation

- Using the data

- Web interface

- Web services

- SBML Export

Future of Sabio-RK

➤ My favourite term: „Easier“

➤ Make data access easier

➤ Explorative search

➤ Faster

➤ More previews

➤ Make curation easier

➤ Information extraction

➤ Improved input interface

➤ Merge input interface into Sabio-RK

Future of Sabio-RK

➤ Another favourite term: participation

➤ Online support for detailed feedback (ongoing grant proposal)

➤ Rights:

➤ Right to participate in data entry

➤ Right to participate in curation

The egotist in me... ...gets a slide with 2 favourite terms

➤ Make software development **easier**⁴

➤ Refactor Sabio-RK

➤ Simpler development

➤ Simpler **participation** in development (if wanted)

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

Not that catchy, but as important

Data mining

- Less tables
- More graphs
- More aggregated information
- ➔ More uses for Sabio-RK

$$[Ga]' = k_1 + (k_2[Ga]) - k_3 \frac{[Ga][PLC]}{([Ga] + K_4)} - k_5 \frac{[Ga][Ca_{cyt}]}{([Ga] + K_6)}$$

$$[PLC]' = k_7[Ga] - k_8 \frac{[PLC]}{([PLC] + K_9)}$$

$$[Ca_{cyt}]' = (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{12} * PLC + k_{13} * [Ga]$$

Not yet the end...

$$- k_{14} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{15})} - k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} - k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)}$$

...two more things

$$+ (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

Tell us how you feel

- Survey in your folder
- Please do fill it
- You **are** allowed to tell us about things you'd like to stay as they are
- Please do complain

Saqib needs some work this evening

➤ Saqib will do WS tutorial

➤ Part of it will be JAVA

➤ Wants to get from you information about your language of choice chosen

➔ Maybe able to do a tiny example in Python/Perl/Ruby/Groovy...

$$[Ga]' = k_1 + (k_2[Ga]) - k_3 \frac{[Ga][PLC]}{([Ga] + K_4)} - k_5 \frac{[Ga][Ca_{cyt}]}{([Ga] + K_6)}$$

$$[PLC]' = k_7[Ga] - k_8 \frac{[PLC]}{([PLC] + K_9)}$$

$$[Ca_{cyt}]' = (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{12} * PLC + k_{13} * [Ga]$$

The end!

$$- k_{14} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{15})} - k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})} - k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} + (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$

$$[Ca_{ER}]' = - (Ca_{ER} - Ca_{cyt}) * \frac{k_{10} * Ca_{cyt} * PLC^4}{PLC^4 + K_{11}^4} + k_{16} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{17})}$$

$$[Ca_{Mito}]' = k_{18} \frac{[Ca_{cyt}]^n}{([Ca_{cyt}]^n + K_{19}^n)} - (Ca_{mit} - Ca_{cyt}) * k_{20} \frac{[Ca_{cyt}]}{([Ca_{cyt}] + K_{21})}$$