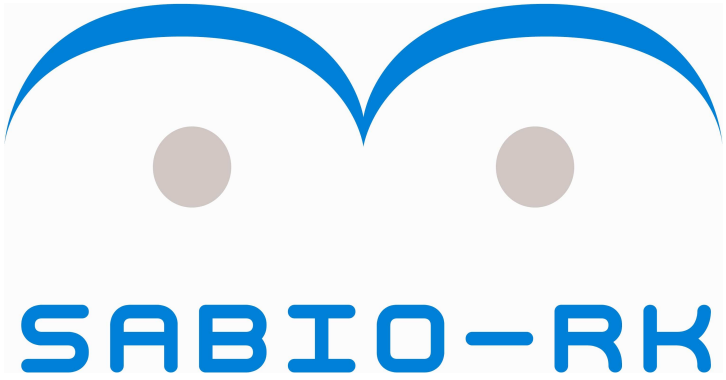


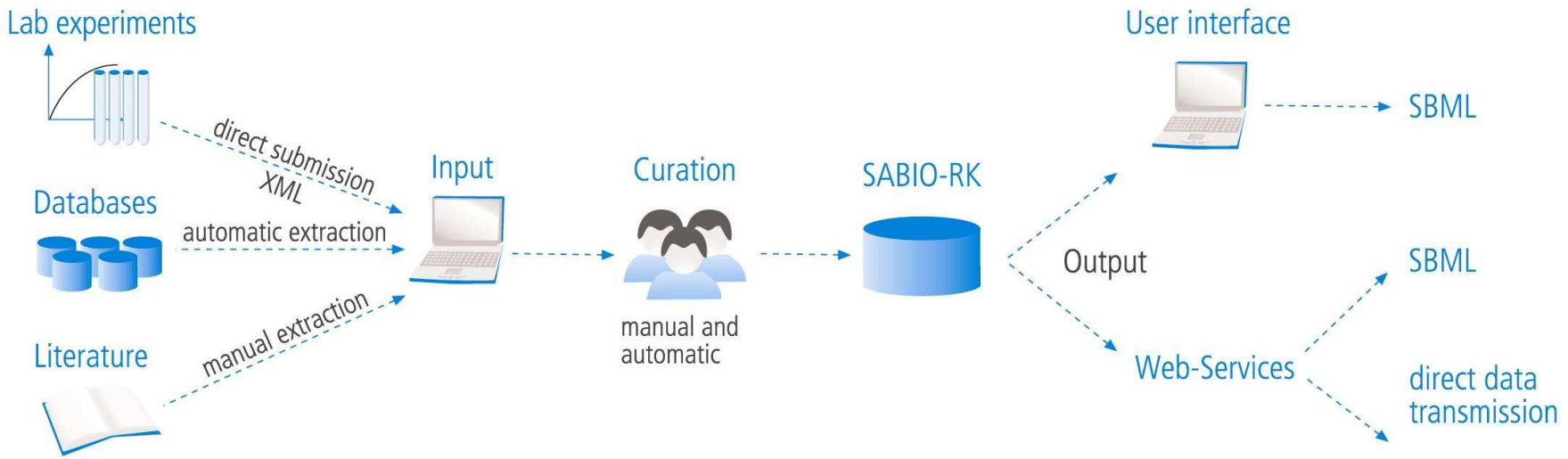
SABIO-RK: Curated Kinetic Data of Biochemical Reactions

Ulrike Wittig
Scientific Databases and Visualization Group
EML Research, Heidelberg, Germany

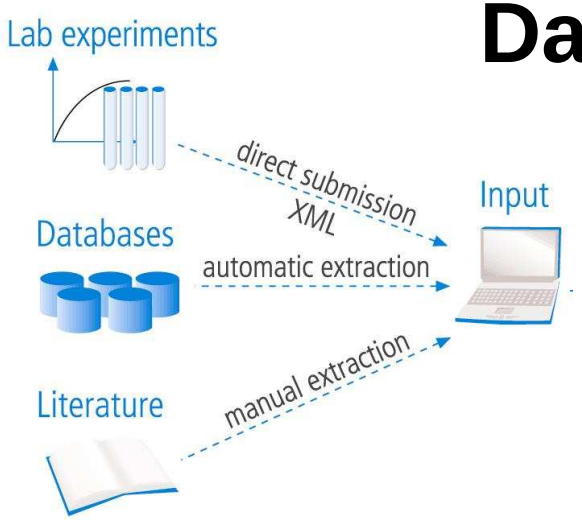
SABIO-RK Database

- general information ***(from other databases + literature)***
  - reaction (substrate, product, modifier), pathway
  - enzyme, protein information (wildtype, mutant etc.)
  - organism, tissue, cell location
  - information source
- kinetic information ***(from literature)***
  - kinetic law, formula
  - parameter (Km, Vmax, concentration etc.)
  - experimental condition (pH, temperature, buffer)

Overview



Data Input



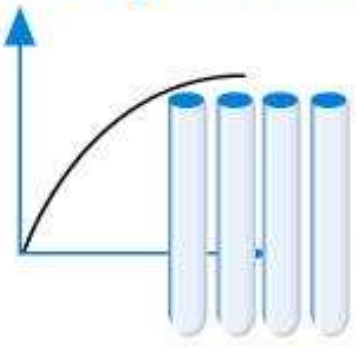
- Automatic extraction of information from other databases (e.g. KEGG)

  → *pathways, reactions, enzymes, compounds*
- Data manually extracted from literature

  → *reactions, protein details, kinetic parameter, environmental conditions, mechanism etc.*
- Data directly obtained from laboratory experiments

  → *kinetic parameter, environmental conditions etc.*

Automatic Input (XML based)



Generation of experimental data

Database of experimental raw data

Storage of kinetic data (e.g. Km, Vmax)

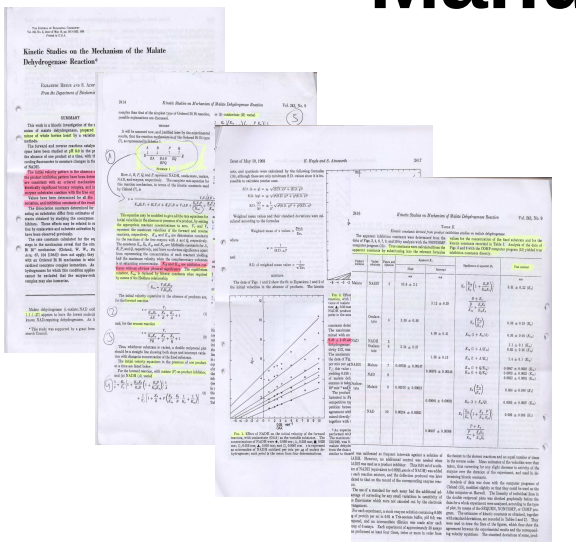
Link to original data

SABIO-RK

Data access

Collaboration with Douglas Kell, Neil Swainston (Manchester Center for Integrative Systems Biology)

Manual Data Extraction



- Data stored in publications (tables, formulas, graphs etc.)
- Some information is only noted as reference (e.g. environmental conditions)
- No controlled vocabulary in the literature (different names for compounds, kinetic laws etc.)
- Missing information

→ **Input Interface**

- Web interface
- Structuring the data from literature

Input Interface – Single Entry

experimental conditions

pH		temperature (°C)		buffer
start	end	start	end	composition
7.4		25		10 mM Tris/Cl, 150 mM NaCl, 5 mM

other condition				
start	end	unit	name	comment
				experimental conditions are taken

organism	Homo sapiens	strain	5
mechanism	unknown		unknown
tissue/celltype	muscle		1016
info source	Human cytosolic adenylate kinase allelozymes; purification and characterization; Luz CM, König I, Schirme		
info type	Journal		
comment	comment		

choose organism: Homo sapiens [add this organism]

choose tissue: unknown [add this tissue]

choose info source: Human monocyte carboxylesterase. Purification and kinetics; Saboori AM, Newcombe DS; J Biol Chem; 26 ID:1790

[add this infosource] [enter new infosource]

[set entry as completed] [update & set completed] [set as non-completed]

[insert as novel entry] [replace old entry with this entry (correct old entry)] [delete this entry]

Input Interface - Summary

Enzyme classification

Entry #	Complex information	EC number	Wildtype	Recombinant	Expressed in	edit
21630	P00568	2.7.4.3	wildtype allelozyme AK1*2	.	.	edit
21631	P00568	2.7.4.3	wildtype allelozyme AK1*2	.	.	edit
21632	P00568	2.7.4.3	wildtype allelozyme AK1*1	.	.	edit
21633	P00568	2.7.4.3	wildtype allelozyme AK1*1	.	.	edit

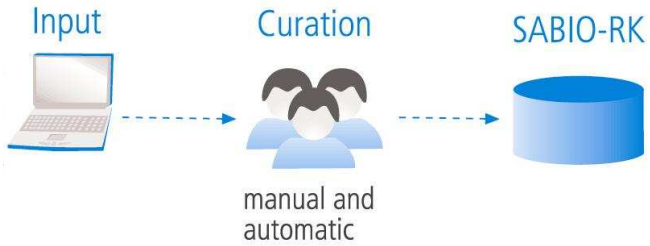
Enzyme description

Entry #	Stoe	Name	UniProt ID	MW (kDa)	Deviation (kDa)	edit
21630	1	native	complex	21.7	.	edit
21631	1	native	complex	21.7	.	edit
21632	1	native	complex	21.7	.	edit
21633	1	native	complex	21.7	.	edit

Environment conditions

Entry #	pH Start	pH End	Temp Start	Temp End	Other Start	Other End	Other Unit	Other Types	Buffer	Comment	edit
21630	7.4	.	25	.	.	.	.	.	10 mM Tris/Cl, 150 mM NaCl, 5 mM dithioerythritol, 0.1 mM EDTA, 4 mM phosphoenolpyruvate, 20 mM KCl, 0.5 mg/ml bovine serum albumin, 1 mM MgCl2, 0.2 mM NADH, 20 U/ml lactate dehydrogenase, 10 U/ml pyruvate kinase	experimental conditions are taken from: PubMed ID 214039	edit
21631	8.1	.	30	.	.	.	.	.	50 mM Tris/OAc, 2.1 mM Mg(OAc)2, 6.7 mM glucose, 0.67 mM NADP, 2.67 U/ml hexokinase, 2.33 U/ml glucose 6-phosphate dehydrogenase	experimental conditions are taken from: PubMed ID 6269633	edit
21632	7.4	.	25	.	.	.	.	.	10 mM Tris/Cl, 150 mM NaCl, 5 mM dithioerythritol, 0.1 mM EDTA, 4 mM phosphoenolpyruvate, 20 mM KCl, 0.5 mg/ml bovine serum albumin, 1 mM MgCl2, 0.2 mM NADH, 20 U/ml lactate dehydrogenase, 10 U/ml pyruvate kinase	experimental conditions are taken from: PubMed ID 214039	edit
21633	8.1	.	30	.	.	.	.	.	50 mM Tris/OAc, 2.1 mM Mg(OAc)2, 6.7 mM glucose, 0.67 mM NADP, 2.67 U/ml hexokinase, 2.33 U/ml glucose 6-phosphate dehydrogenase	experimental conditions are taken from: PubMed ID 6269633	edit

Curation



- Data first inserted in an intermediate database
- Curation process
  - *Manually by biological experts*
  - *Semi-automatically by consistency checks*
  - *Standardization*
  - *Unification*
  - *Annotation to controlled vocabularies and external databases*
- Transfer data from intermediate to public SABIO-RK database

Curation Problems

Missing or only partial information in the data source:

- *Incomplete reactions (e.g. products not mentioned)*
- *Organism specification is missing*
- *Assay conditions or methods description missing or reference to another paper (no temperature or room temperature etc.)*
- *Kinetic law equation (or fitting equation) not described or parameters incomplete*
- *Isoenzyme not specified (old papers)*

Identification of compounds, reactions and enzymes:

- *Ambiguous descriptions of chemical compounds or enzymes*

Automatic Consistency Checks

- → Check if all parameters in a kinetic law formula are in the parameters list
- → Check if all reaction participants are written and have a defined role in the reaction equation
- → Check if parameter type and its role (variable or constant) is defined
- → Check if compound-dependent parameter types have a defined species

Identification of Identical Chemical Compounds

Methods developed to support curation:

→Normalization of compound names

→Analysis of compound names and generation of SMILES string

→Generation of chemical structure based on SMILES string

→Classification of chemical compounds

Normalization of Compounds

Matching names with different databases (SABIO-RK, ChEBI, PubChem) to find synonyms of chemical compounds

Example rules:

-Lowercase letters only

-Normalise different types of brackets to one type

-Remove spaces

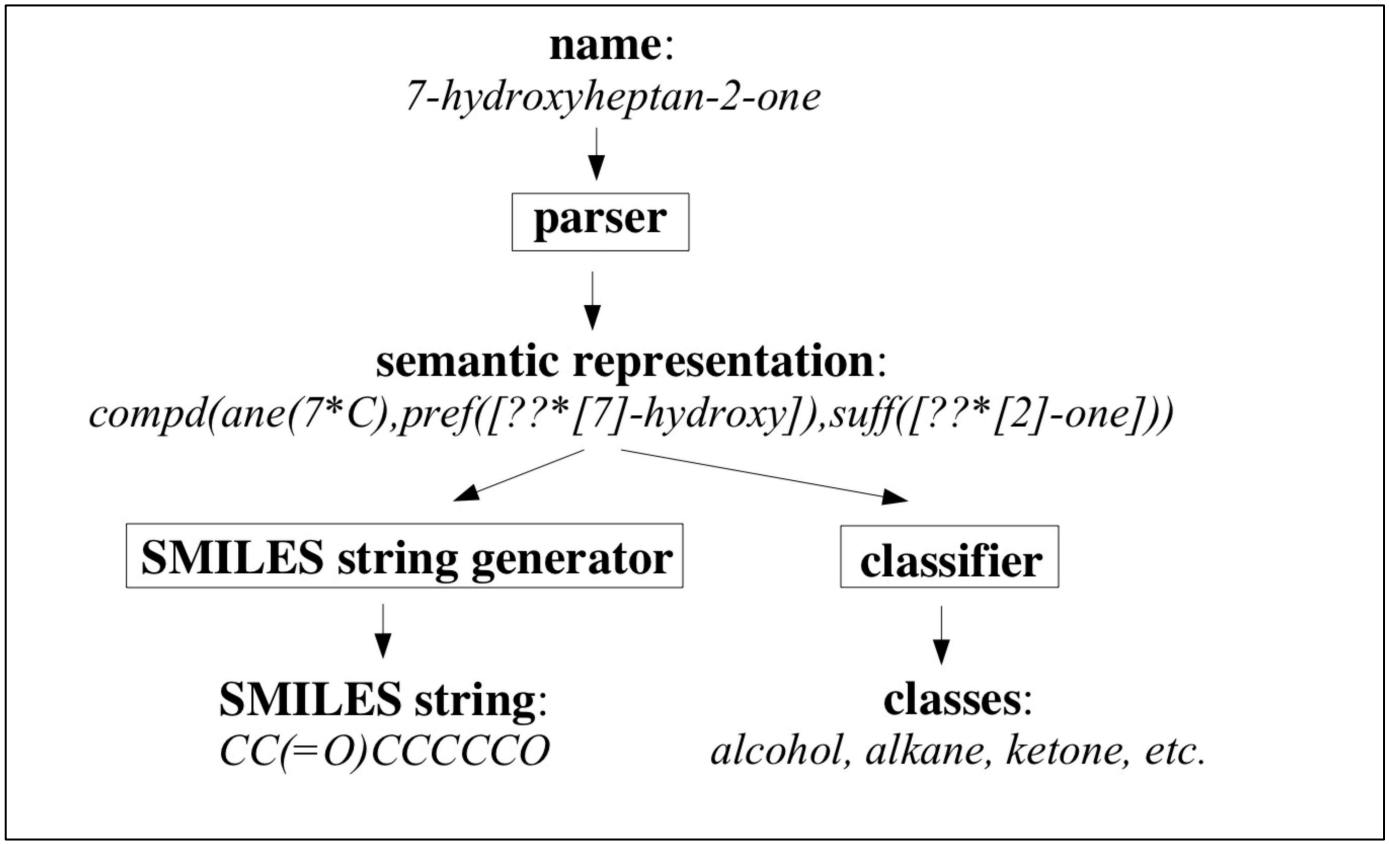
-Replace ‘-p’ at end of name by ‘-phosphate’

-Replace all suffixes ‘-phosphate’ by prefixes ‘phospho-’

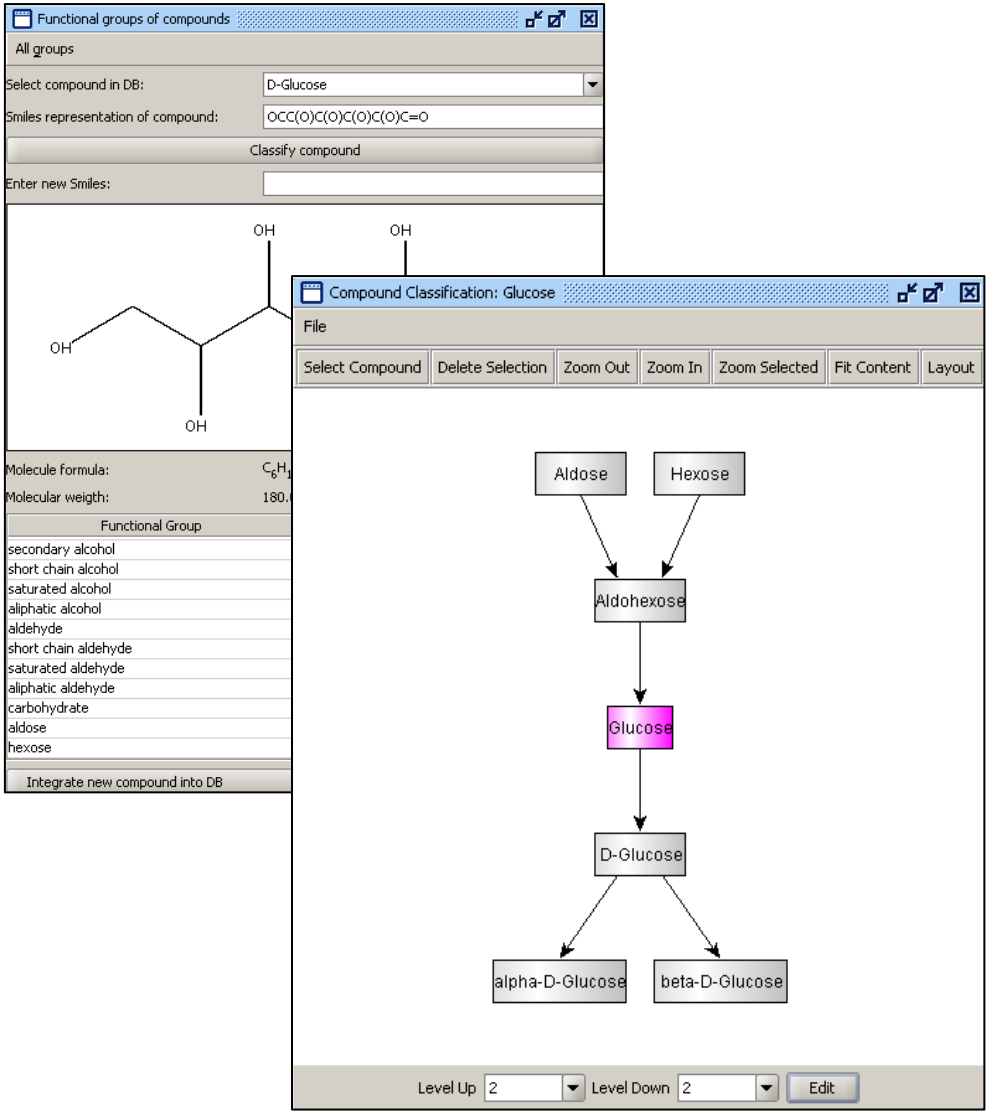
-Replace ‘ate’ followed by a delimiting character by ‘ic acid’

http://sabiork.villa-bosch.de/normaWeb

Analysis of Compound Names



Classification of Compounds



- Generation of structural formula, totals formula and molecular weight based on SMILES string
- Rules for compound classifications
- Graph and sub-graph search for functional groups
- Classification using different criteria, for example D-Glucose is a:
  -Aldose (functional group aldehyde)
  -Hexose (number of C-Atoms = 6)

→ **SABIO-RK search for general compounds and all sub-classes**

Ontologies used in SABIO-RK

→ **ChEBI:** database and ontology for chemical compounds

→ **NCBI taxonomy:** classification of organisms

→ **Gene Ontology:** ontology for molecular functions, biological processes and cellular components

→ **Systems Biology Ontology:** ontology for systems biology

Annotations/links to ontologies and external databases in SABIO-RK user interface and SBML export

Controlled Vocabularies in SABIO-RK

→ Chemical compounds names

→ Role of reaction participants (*substrate, product, modifier)*

→ Defined reactions and pathways

→ Organisms, tissues and cellular locations

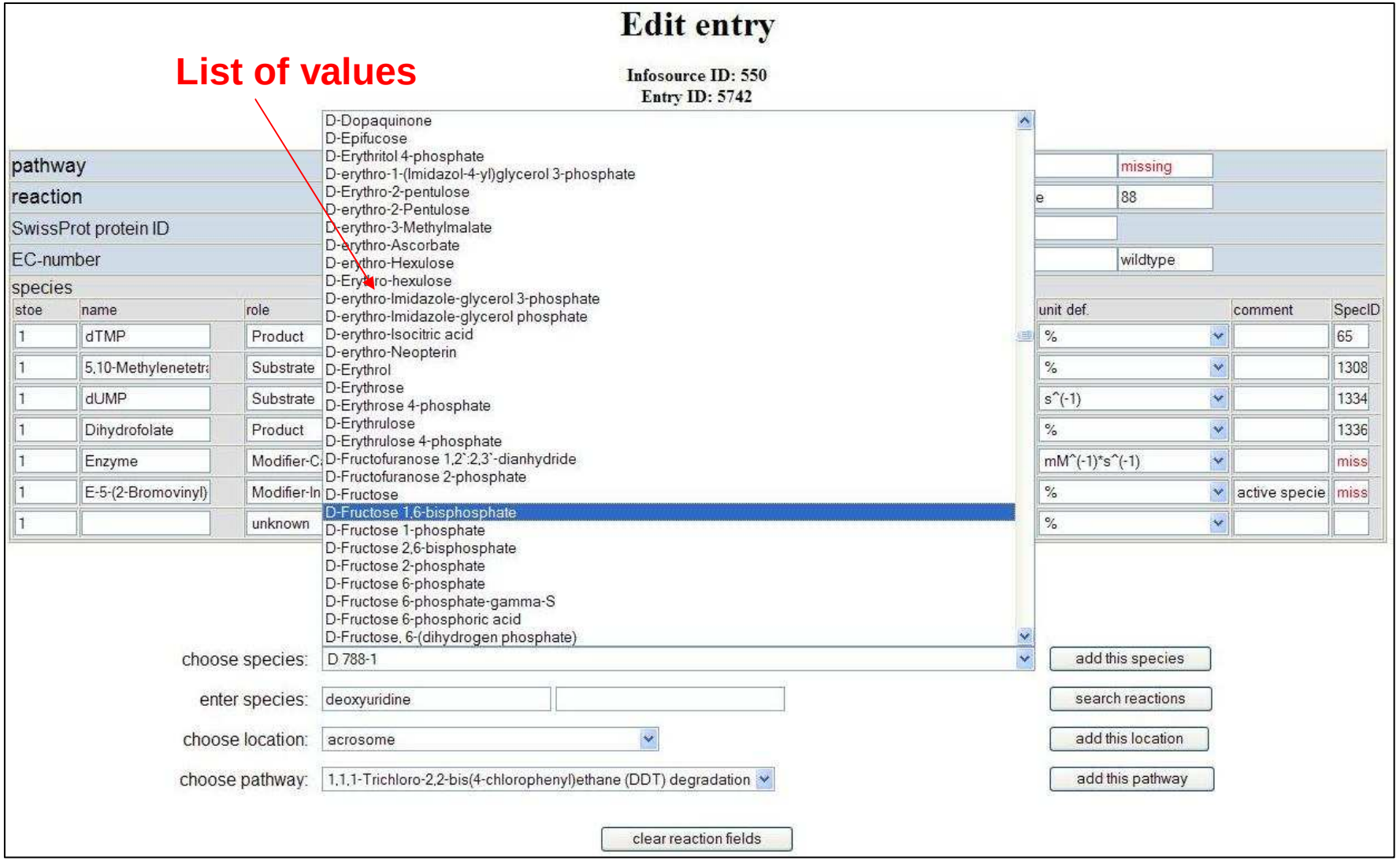
→ Kinetic law types (*Michaelis-Menten, Ordered Bi Bi*)

→ Parameter types (*Km, kcat*) and units (*mM, 1/sec*)

→ Protein details *(wildtype, mutant, isoenzyme; complex composition; information about recombinant enzymes)*

To avoid redundancies and typing errors by defined notations

Controlled Vocabularies in SABIO-RK



Protein Details

→ Structured information about proteins based on the literature

- *Complex composition (e.g. homotetramer)*
- *Information: recombinant; expressed in …*
- *wildtype, mutant, isozyme*
- *Molecular weights of complex and subunits*

Tissue:	**liver**
EC Class: 4.2.1.22	wildtype
Recombinant	Expressed in Escherichia coli

Substrates

name	location	comment
L-Homocysteine	-	-
L-Serine	-	-

Products

name	location	comment
H2O	-	-
L-Cystathionine	-	-

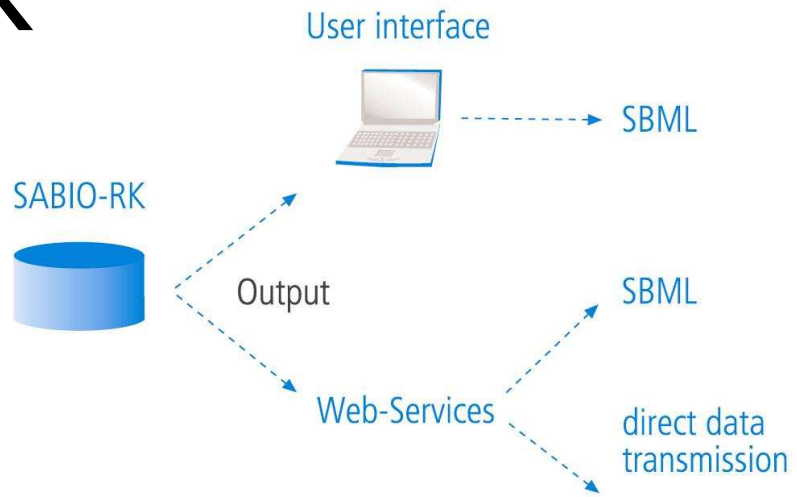
Modifiers

name	location	effect	comment	protein complex
Cystathionine beta-synthase(Enzyme)	-	Modifier-Catalyst	-	(P35520)*4;

Enzyme (protein data)

	UniProt-ID	name	mol. weight (kDa)	deviation (kDa)
subunit	-	-	-	-
complex	-	-	252.0	-

Access to SABIO-RK



- **Web-based user interface**

  for browsing and searching the data manually

- **Web Services (API access)**

  can be automatically called by external tools, e.g. by other databases or simulation programs for biochemical network models

Both interfaces support the export of the data in SBML

SABIO-RK Queries

Search for

→ reactions by defining participants (reactants, enzymes, proteins)

→ publications

→ pathways, organisms, tissues etc.

→ reactions containing defined kinetic parameters or laws

→ reactions under defined experimental conditions

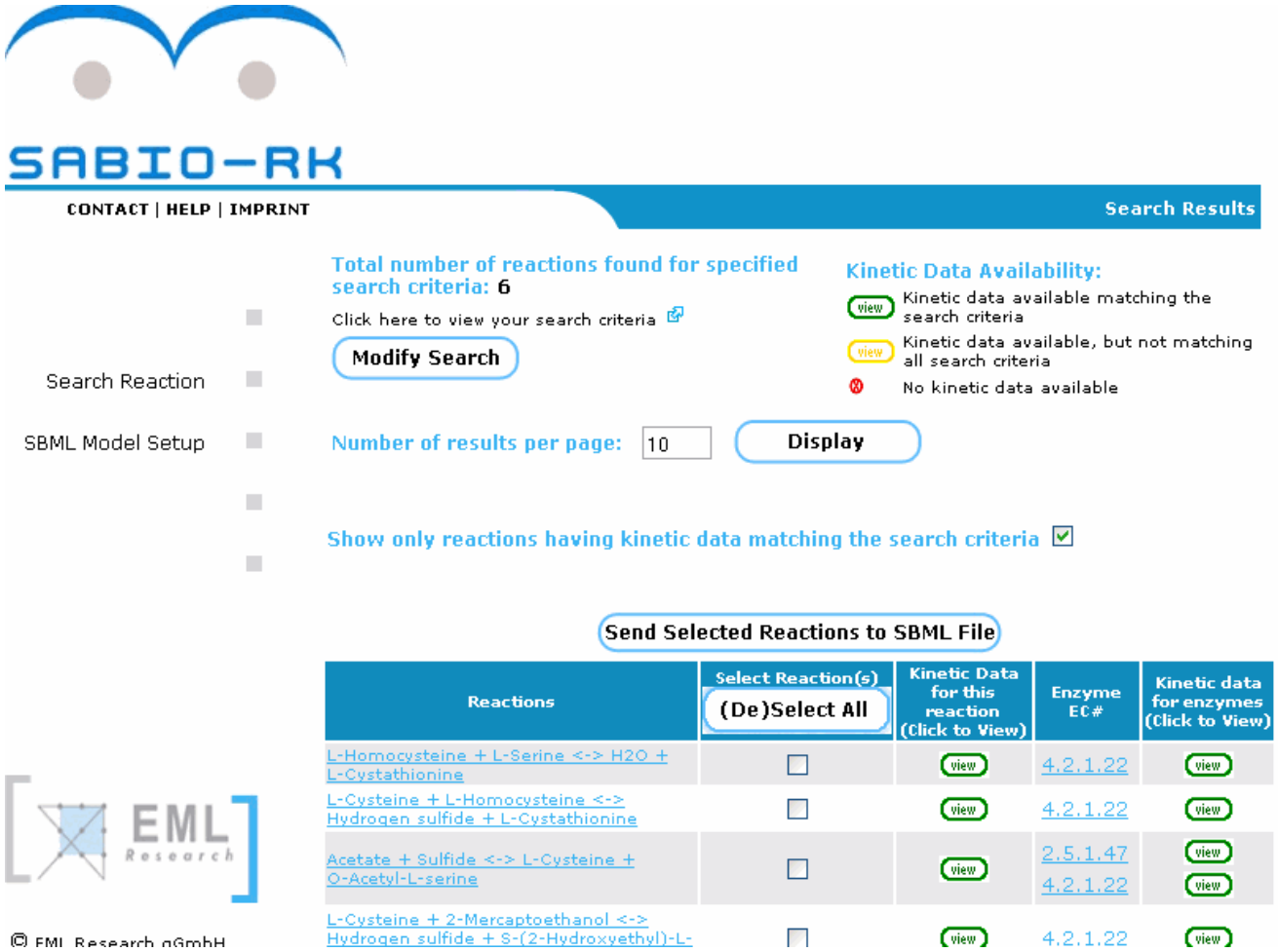
Definition of queries by the combination of all the search criteria

SABIO-RK Database

http://sabio.villa-bosch.de/

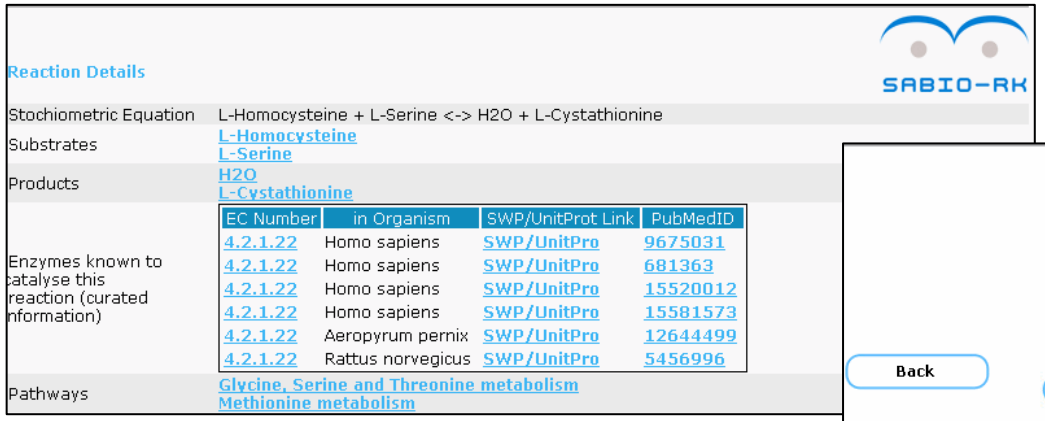
SABIO-RK Database

http://sabio.villa-bosch.de/



SABIO-RK Database

http://sabio.villa-bosch.de/



Reaction Details

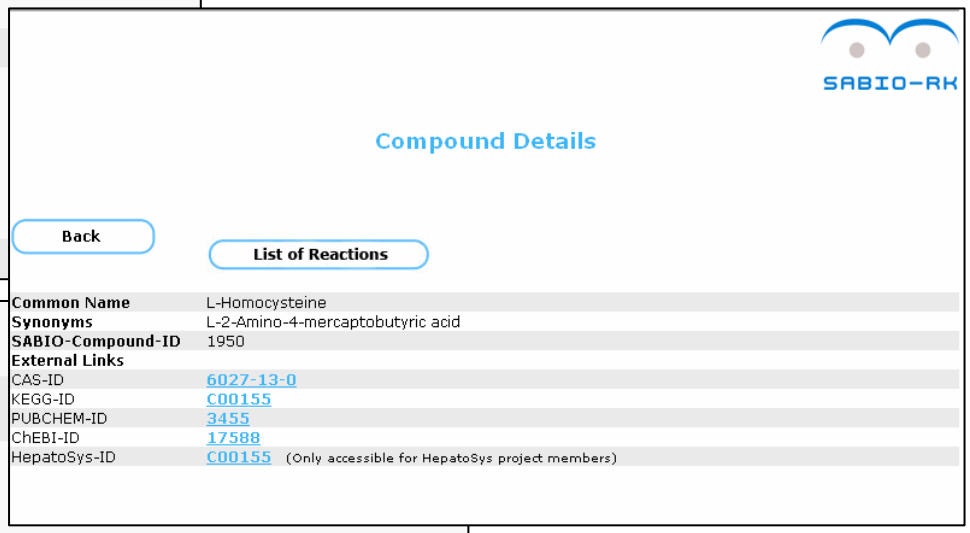
SABIO-RK

Stochiometric Equation	L-Homocysteine + L-Serine <-> H2O + L-Cystathionine
Substrates	L-Homocysteine L-Serine
Products	H2O L-Cystathionine

Enzymes known to catalyse this reaction (curated information)

EC Number	in Organism	SWP/UnitProt Link	PubMedID
4.2.1.22	Homo sapiens	SWP/UnitPro	9675031
4.2.1.22	Homo sapiens	SWP/UnitPro	681363
4.2.1.22	Homo sapiens	SWP/UnitPro	15520012
4.2.1.22	Homo sapiens	SWP/UnitPro	15581573
4.2.1.22	Aeropyrum pernix	SWP/UnitPro	12644499
4.2.1.22	Rattus norvegicus	SWP/UnitPro	5456996

Pathways	Glycine, Serine and Threonine metabolism Methionine metabolism

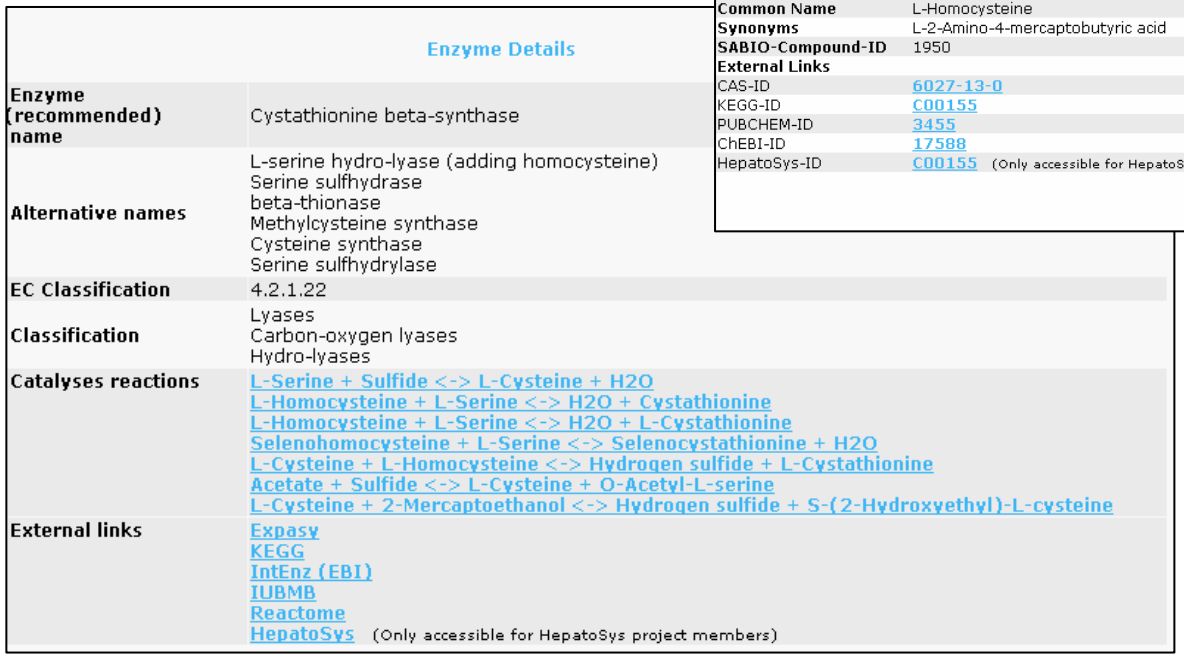


SABIO-RK

Compound Details

Back | List of Reactions

Common Name	L-Homocysteine
Synonyms	L-2-Amino-4-mercaptobutyric acid
SABIO-Compound-ID	1950
External Links	
CAS-ID	6027-13-0
KEGG-ID	C00155
PUBCHEM-ID	3455
ChEBI-ID	17588
HepatoSys-ID	C00155 (Only accessible for HepatoSys project members)



Enzyme Details

Enzyme (recommended) name	Cystathionine beta-synthase
Alternative names	L-serine hydro-lyase (adding homocysteine) Serine sulfhydrase beta-thionase Methylcysteine synthase Cysteine synthase Serine sulfhydrylase
EC Classification	4.2.1.22
Classification	Lyases Carbon-oxygen lyases Hydro-lyases
Catalyses reactions	L-Serine + Sulfide <-> L-Cysteine + H2O L-Homocysteine + L-Serine <-> H2O + Cystathionine L-Homocysteine + L-Serine <-> H2O + L-Cystathionine Selenohomocysteine + L-Serine <-> Selenocystathionine + H2O L-Cysteine + L-Homocysteine <-> Hydrogen sulfide + L-Cystathionine Acetate + Sulfide <-> L-Cysteine + O-Acetyl-L-serine L-Cysteine + 2-Mercaptoethanol <-> Hydrogen sulfide + S-(2-Hydroxyethyl)-L-cysteine
External links	Expasy KEGG IntEnz (EBI) IUBMB Reactome HepatoSys (Only accessible for HepatoSys project members)

SABIO-RK Database

http://sabio.villa-bosch.de/

Kinetic
L-Homocysteine +
Expar

Entry Nr. 3101

Organism:	**Homo sapiens**
Tissue:	**unknown**
EC Class: 4.2.1.22	wildtype

Entry Nr. 3102

Organism:	**Homo sapiens**
Tissue:	**unknown**
EC Class: 4.2.1.22	mutant delta411-551

Entry Nr. 3103

Organism:	**Homo sapiens**
Tissue:	**unknown**
EC Class: 4.2.1.22	wildtype

Entry Nr. 5535

Organism:	**Homo sapiens**
Tissue:	**unknown**
EC Class: 4.2.1.22	wildtype

Entry Nr. 8224

Tissue:	**liver**
EC Class: 4.2.1.22	wildtype
Recombinant	Expressed in Escherichia coli

Substrates

name	location	comment
L-Homocysteine	-	-
L-Serine	-	-

Products

name	location	comment
H2O	-	-
L-Cystathionine	-	-

Modifiers

name	location	effect	comment	protein complex
Cystathionine beta-synthase(Enzyme)	-	Modifier-Catalyst	-	(P35520)*4;

Enzyme (protein data)

	UniProt-ID	name	mol. weight (kDa)	deviation (kDa)
subunit	-	-	-	-
complex	-	-	252.0	-

Kinetic Law

type	formula
Michaelis-Menten	E*kcat*S/(Km+S)

Parameters

name	species	type	start value	end value	deviation	unit	comment
E	Enzyme	concentration	-	-	-	-	-
S	L-Serine	concentration	0.005	6.4	-	mM	-
A	L-Homocysteine	concentration	5	-	-	mM	-
Km	L-Serine	Km	3	-	-	mM	-
kcat_Km	L-Serine	kcat/Km	2	-	-	mM^(-1)*s^(-1)	-
kcat	-	kcat	6	-	-	s^(-1)	-

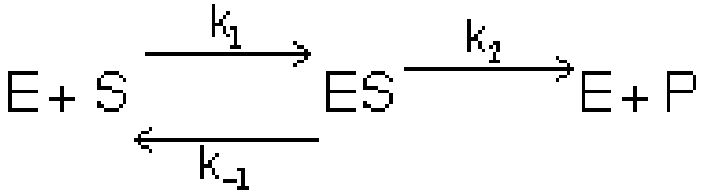
Experimental conditions

	start value	end value	unit
pH	8.6	-	-
temperature	37	-	°C
buffer: 10 mM Tris/HCl			

General comment: -

PUBMEDID: 9675031

Reaction Mechanism



$$\text{Velocity} = V = \frac{V_{max}[S]}{[S] + K_M}$$

V_{max} = maximal enzyme velocity
K_M = Michaelis-Menten constant (k2+k-1)/k1

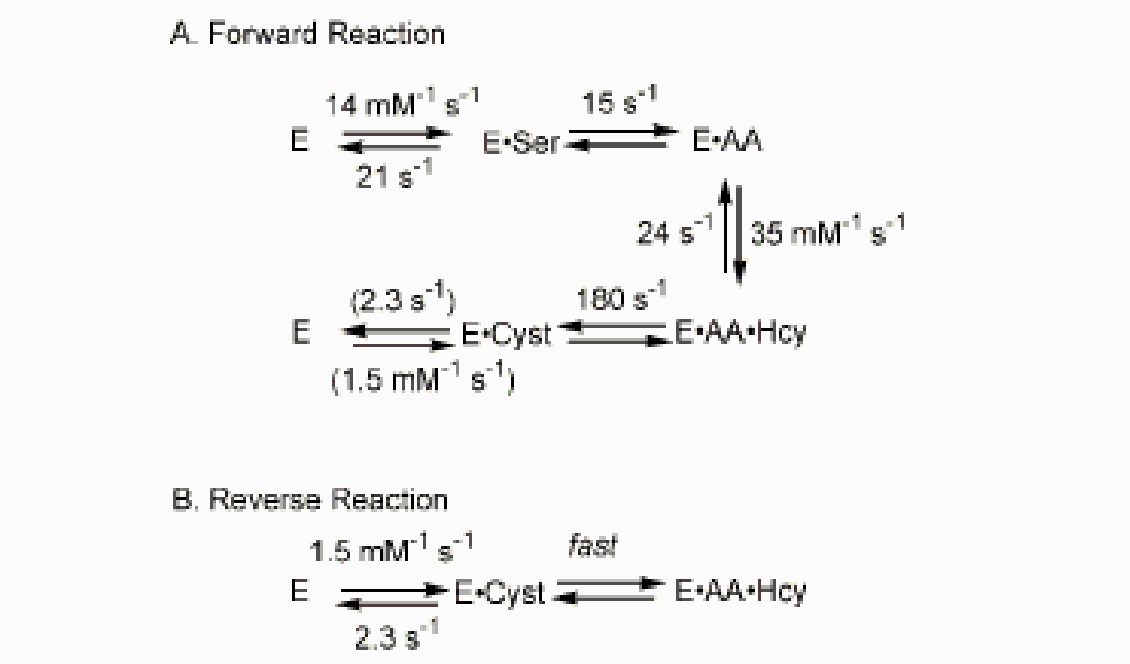
→ Simulation experiments need detailed information about the reaction mechanism and kinetic parameters for the steps

→ No quantitative data for mechanism steps available in structured format *(no database contains such data at the moment)*

Reaction Mechanism

L-Serine + L-Homocysteine = Cystathionine + H2O

A **B** **P** **Q**



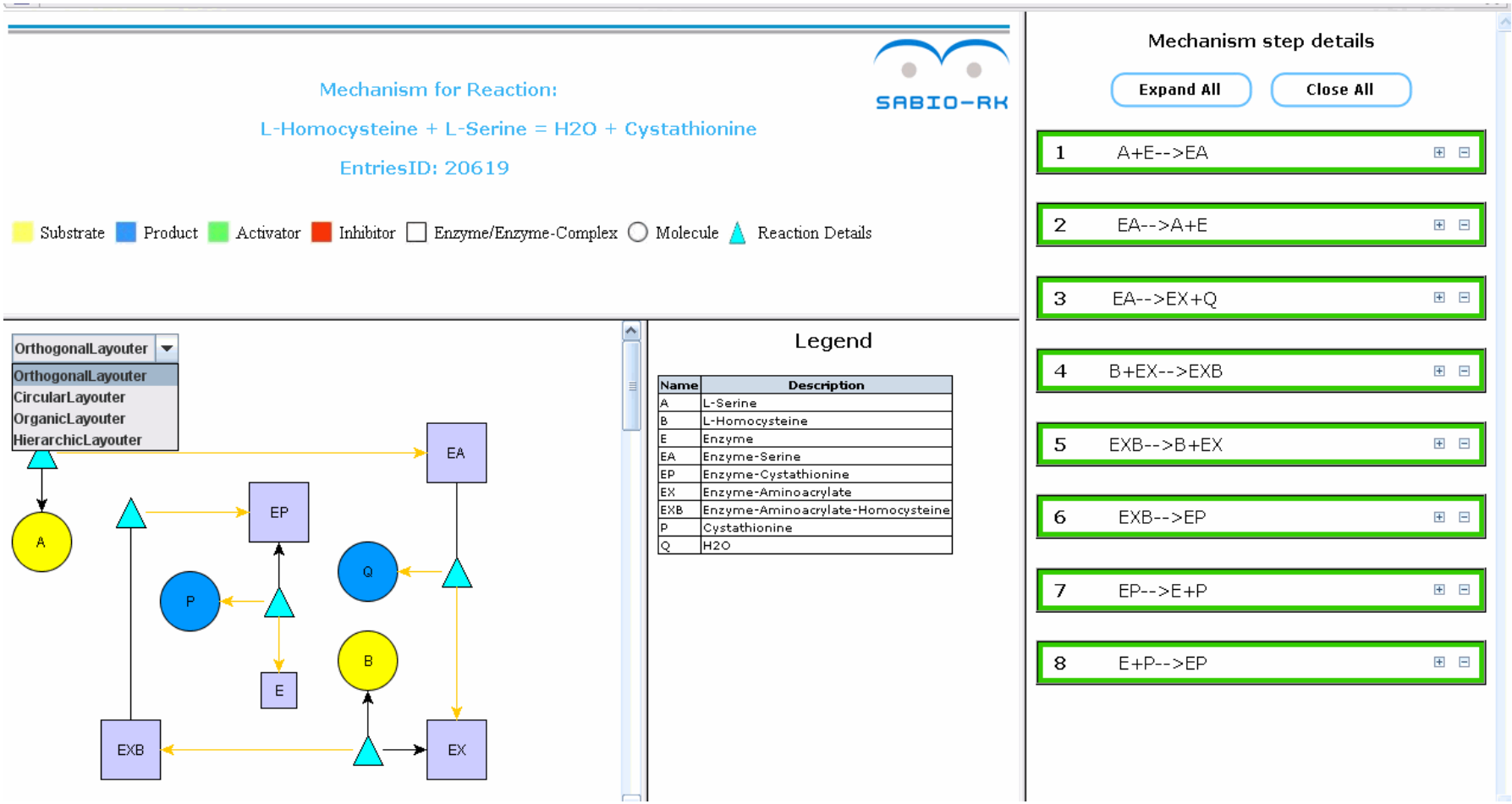
THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2002 by The American Society for Biochemistry and Molecular Biology, Inc.
Vol. 277, No. 25, Issue of June 21, pp. 22421–22425, 2002
Printed in U.S.A.

Stopped-flow Kinetic Analysis of the Reaction Catalyzed by the Full-length Yeast Cystathionine β-Synthase*

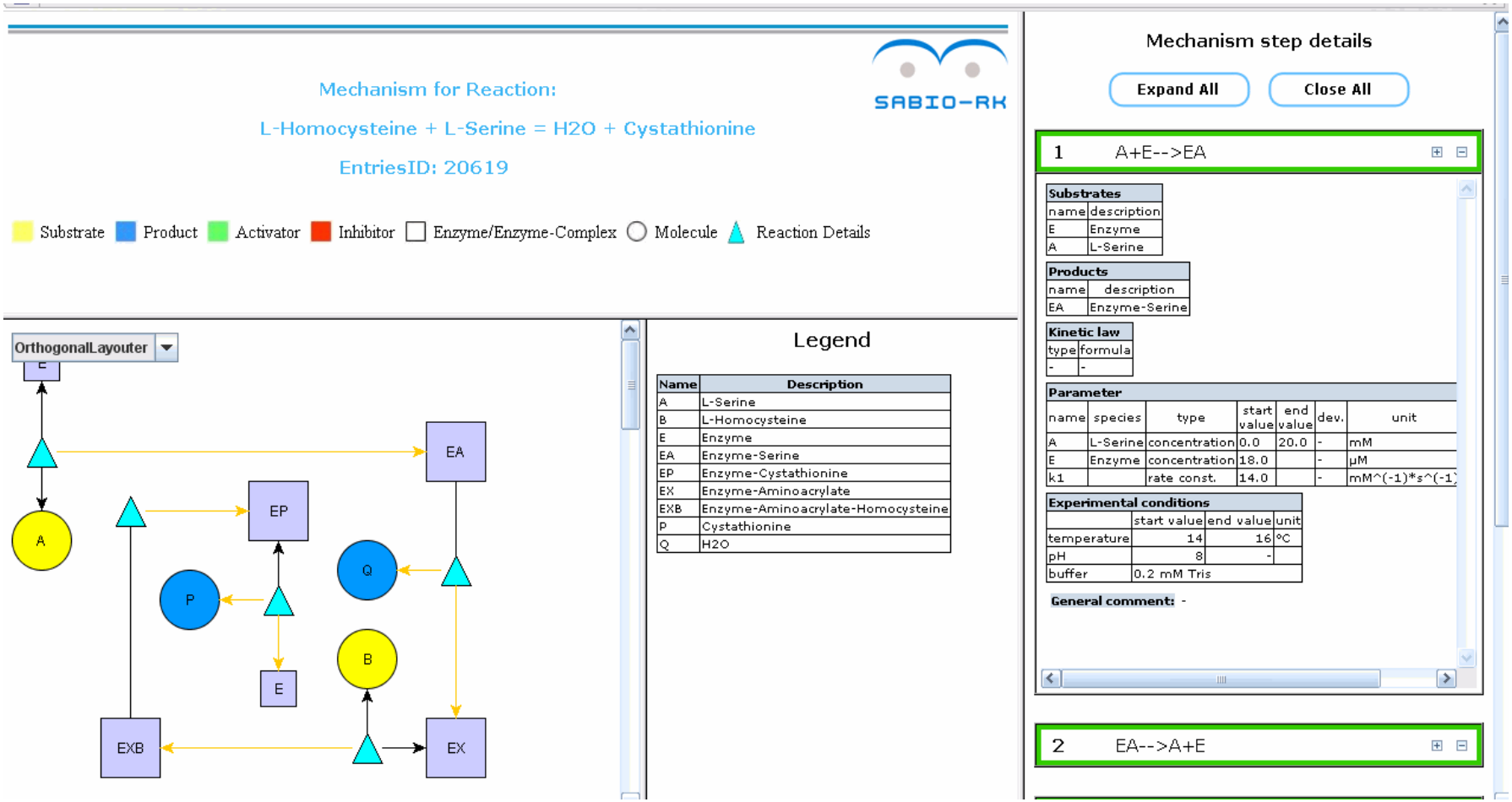
Mechanism Steps

E + A → EA
EA → E + A
EA → EX + Q
EX + B → EXB
EXB → EX + B
EXB → EP
EP → E + P
E + P → EP

User Interface – Mechanism



User Interface – Mechanism



SABIO-RK Summary

→ curated database for biochemical reactions and their kinetics

→ platform for kinetic data storage and exchange

→ free available for academic use

→ use of standard data formats, controlled vocabularies and ontologies

→ links to original data sources (literature, databases etc.)

SABIO-RK Statistics

Publications (curated)	2208
Entries (curated)	30710
Reactions	9596
with kinetic data:	
Reactions	4924
Enzymes	779
Organisms	504
Tissues	175

as of April 2009

Future Perspectives

- SABIO-RK as platform for experimental kinetic data
  - *Scientists producing the data can directly enter it into SABIO-RK*
- Input interface as storage and exchange tool for authors of publications and journal editors
- Kinetic data for signaling reactions
- Information about in vivo metabolite concentrations
- Development and implementation of tools for information extraction and supporting data curation

SABIO-RK team:

Renate Kania
Olga Krebs
Andreas Weidemann
Saqib Mir
Henriette Engelken
Martin Golebiewski
Ulrike Wittig
Isabel Rojas
Wolfgang Müller
Sylvestre Kengne
Heidrun Sauer-Danzwith

Financial support:

KTS

Bundesministerium für Bildung und Forschung

SABIO-RK is member of:

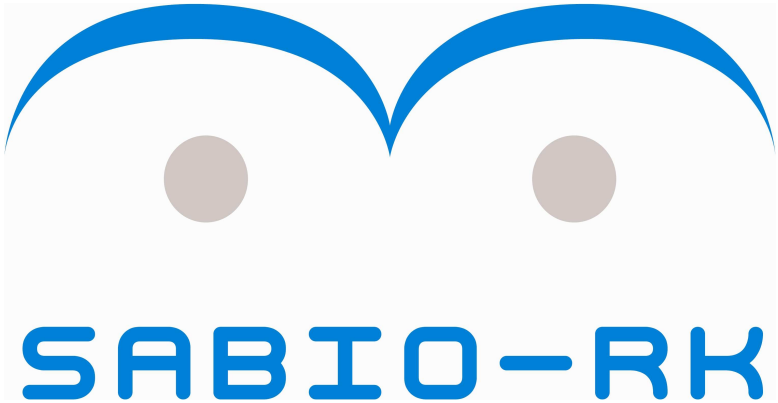
Network Systems Biology HepatoSys

SysMO Systems Biology of Microorganisms

SABIO-RK User Meeting

June 15-16, 2009

Heidelberg, Germany



http://sabio.villa-bosch.de/