SABIO-RK

Integration and Curation of Reaction Kinetics Data

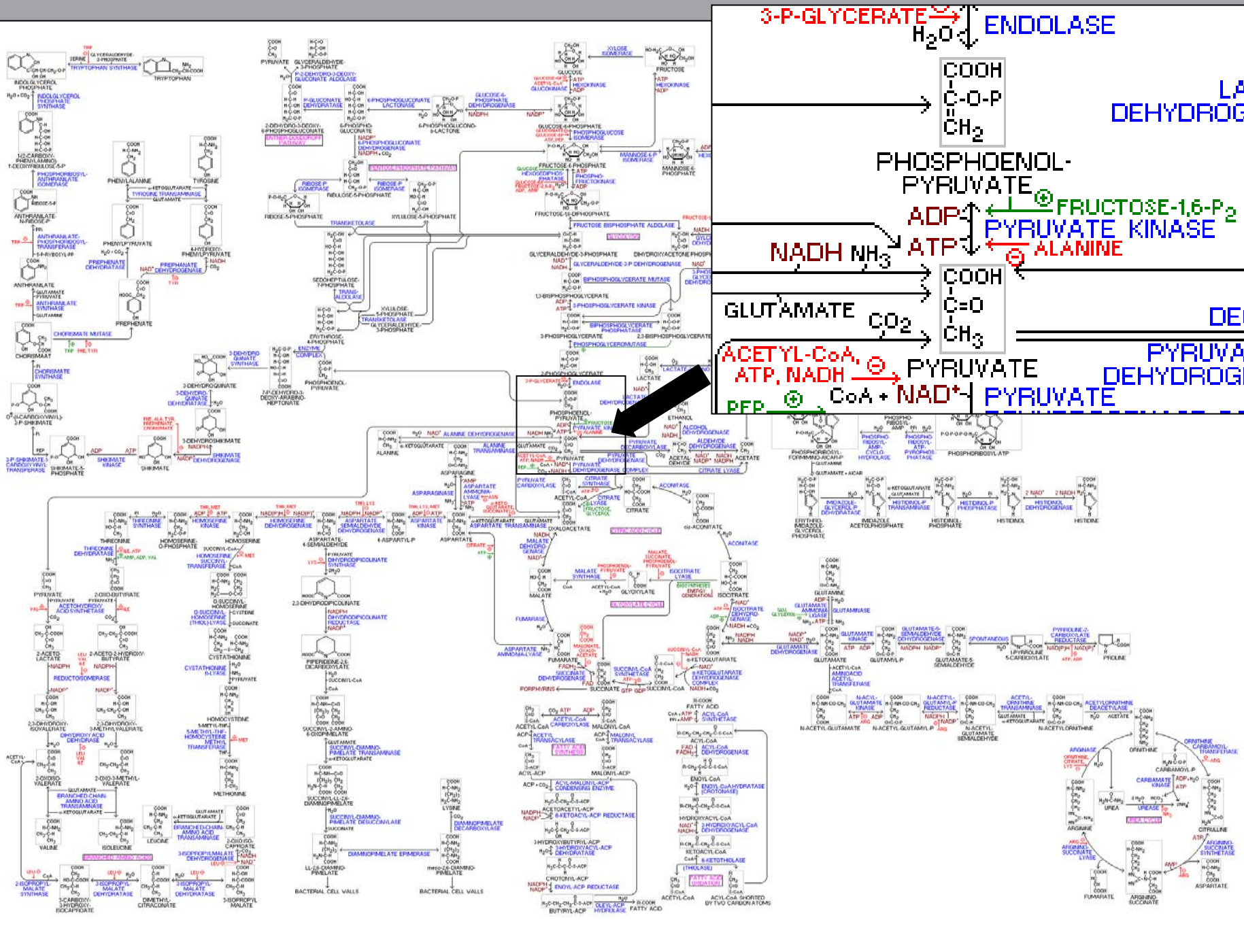
http://sabio.villa-bosch.de/SABIORK

Ulrike Wittig

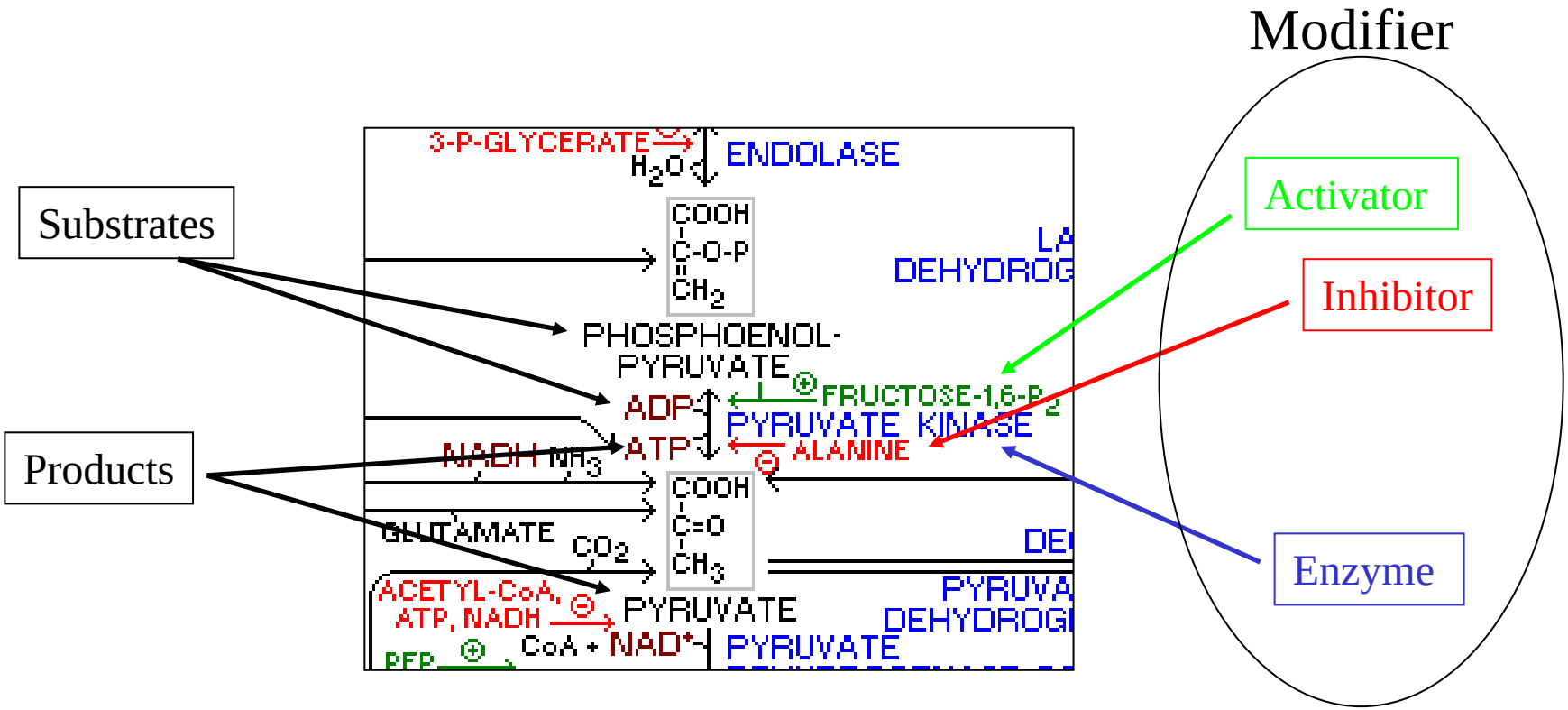
EML Research

Overview

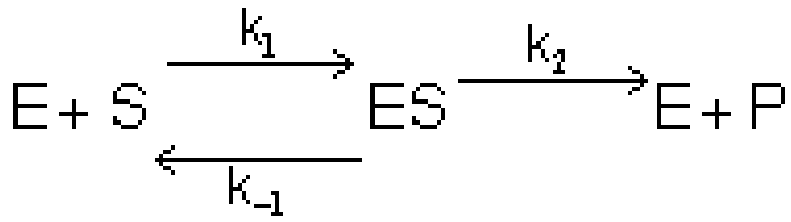
- Introduction /Motivation
- Database content /User interface
- Data integration
- Curation
- Conclusion /Future directions

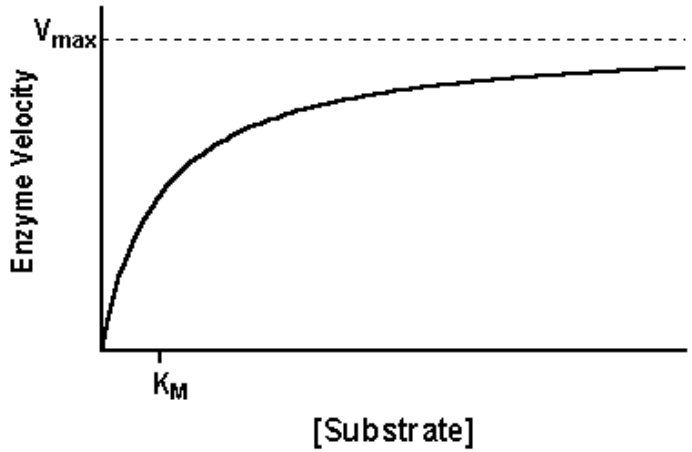

3-P-GLYCERATE
H2O
ENDOLASE
COOH
C-O-P
CH2
PHOSPHOENOL-
PYRUVATE
FRUCTOSE-1,6-P2
ADP
ATP
PYRUVATE KINASE
ALANINE
NADH NH3
COOH
C=O
CH3
GLUTAMATE
CO2
ACETYL-CoA,
ATP, NADH
PYRUVATE
CoA + NAD+
PFP
PYRUVATE

Introduction - Reaction



Introduction - Reaction kinetics



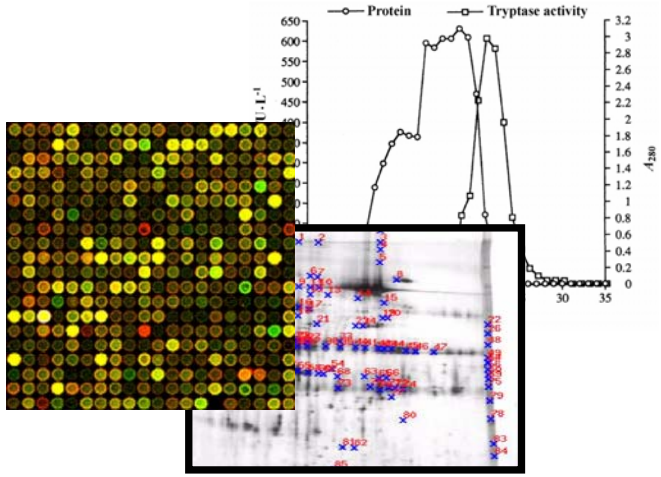


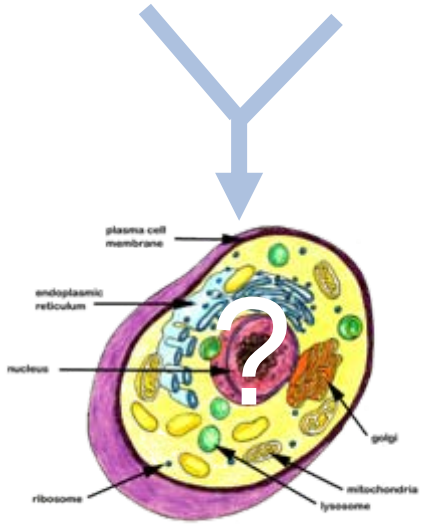
$$\text{Velocity} = V = \frac{V_{max}[S]}{[S] + K_M}$$

V_{max} → maximal enzyme velocity

K_M → Michaelis-Menten constant (k2+k-1)/k1

Systems Biology





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

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

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

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

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

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

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

Systems Biology

- Growing interest in simulation and analysis of complex biochemical networks requires:
  - Access to reaction kinetics data
  - Structuring and merging of information
  - Using and defining standard formats to facilitate the integration of data
  - Searching and re-use of data

Public sources for kinetic data

- **BRENDA** *http://www.brenda.uni-koeln.de/*
  - functional and molecular information about enzymes
  - parameters associated with enzymes but no kinetic laws

- **Biomodels database** *http://www.ebi.ac.uk/biomodels/*
  - information about complete published mathematical models of biochemical networks

- **KDBI** *http://xin.cz3.nus.edu.sg/group/kdbi/kdbi.asp*
  - kinetic data of binding or reaction events

- **UniProt/Swiss-Prot** *http://www.ebi.uniprot.org/*
  - comment line “biophysicochemical properties” contains data on kinetic parameters, pH and temperature dependence

- **JWS** *http://www.jjj.bio.vu.nl/database/*
  - information about complete published mathematical models of biochemical networks

Motivation for SABIO-RK

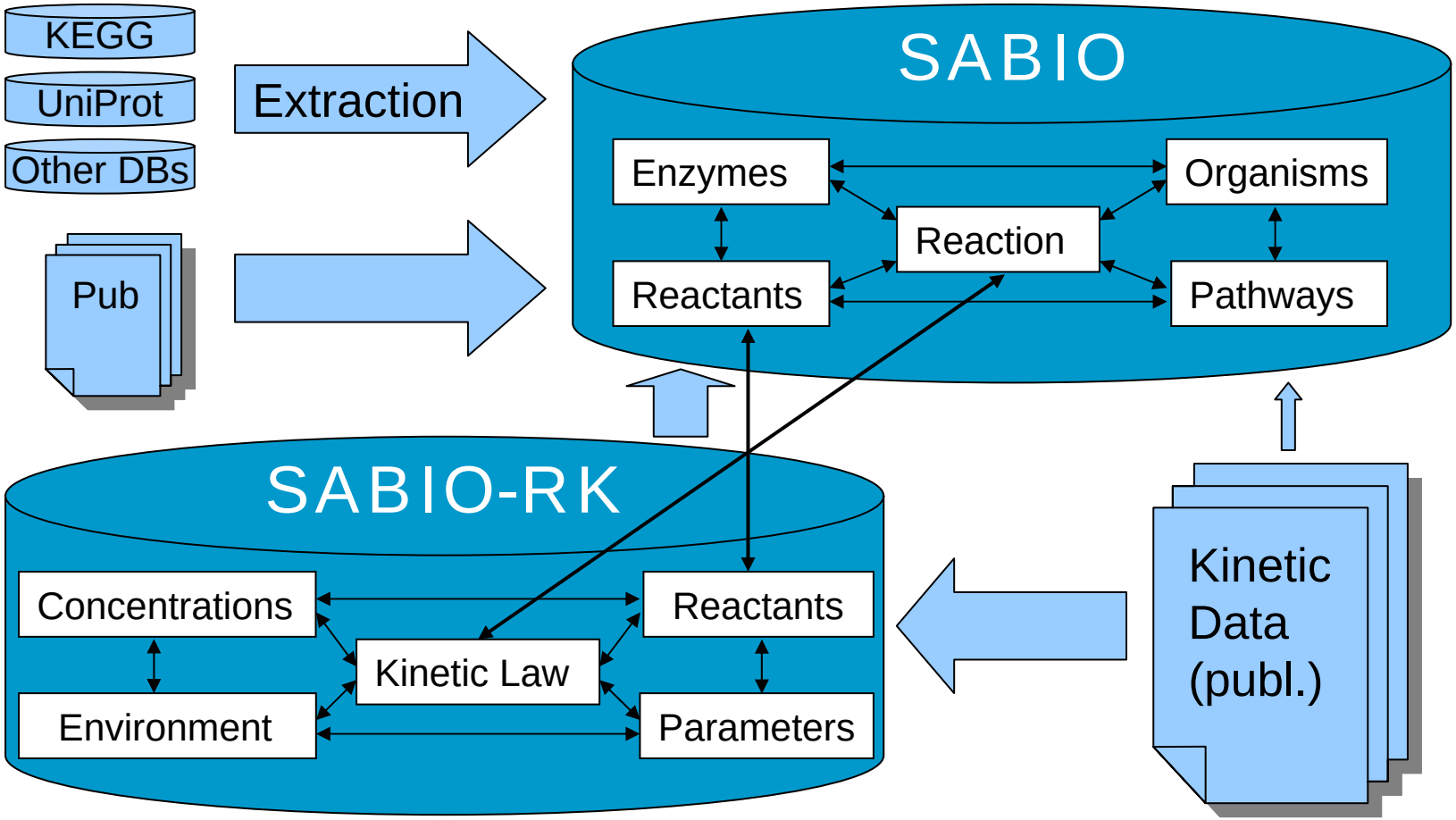
- Most information about reaction kinetics stored in literature

  → Structuring information from literature
- Information about biochemical reactions is rarely connected with information about their kinetics
- Need of kinetic data of biochemical reactions for Systems Biology groups → Data for computational analysis of biochemical reactions
- None of the existing databases links experimental kinetic data for single reactions to complete sets of information comprising:

  - Kinetic Law for the reaction rate
  - Environmental conditions
  - Concentrations of reactants and modifiers
  - Data source (original publication)
  - Organism, tissue and cellular location
- Kinetic data must be easily accessible and interchangeable
- SABIO (System for the Analysis of Biochemical Pathways) already developed at EML
- In house expertise in the area of systems biology

SABIO-RK

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

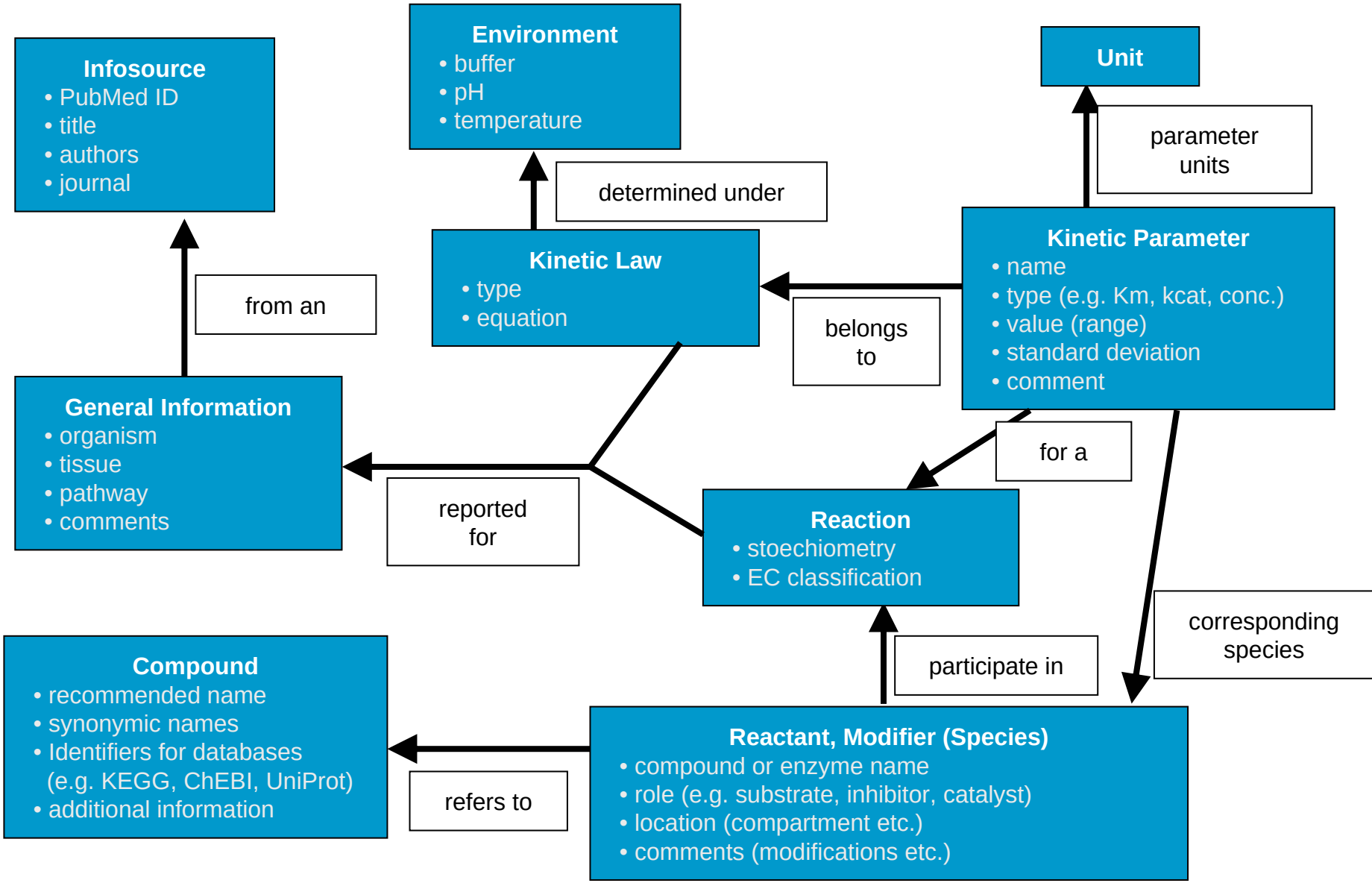


SABIO-RK - Database content

- general information related to SABIO
  - reaction (substrate, product, modifier), pathway
  - enzyme, protein information (wildtype, mutant etc.)
  - organism, tissue, cell location
  - information source

- kinetic information
  - kinetic law, formula
  - parameter (Km, Vmax, concentration etc.)
  - experimental condition (pH, temperature, buffer)
  - information source

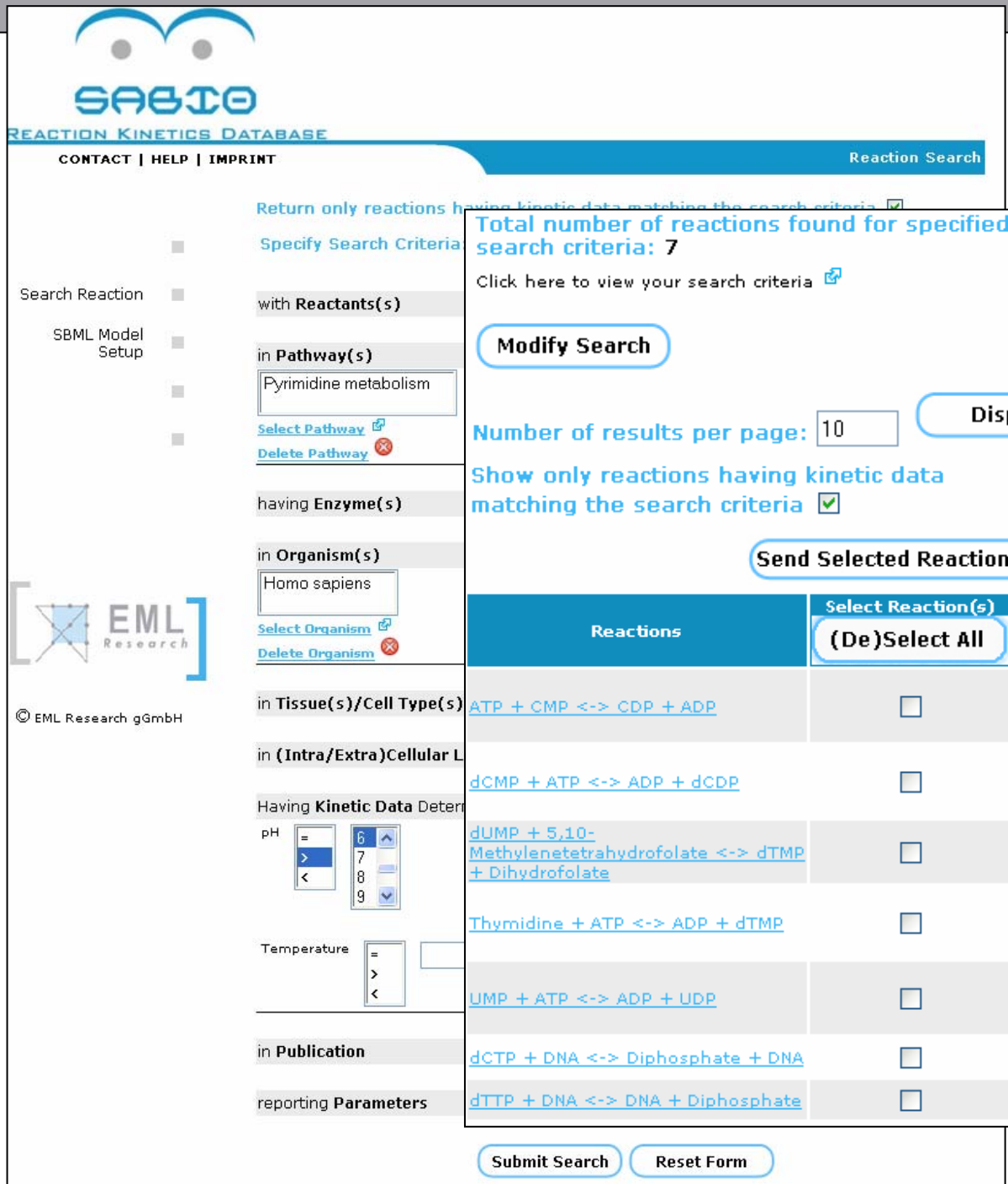
SABIO-RK - Data model (schematic)



SABIO-RK web interface

- Web accessible database to provide information about the kinetics of biochemical reactions
- Search for general reaction information, kinetic laws, kinetic parameters, experimental conditions etc.
- Complex queries (combining different search criteria)
  - Give me all reactions in human liver for pathway Glycolysis measured at pH 7.5!

- Colour-coded representation of results
  - Kinetic data available matching search criteria
  - Kinetic data available but not matching search criteria
  - No kinetic data available
- Export of kinetic data in SBML (Systems Biology Mark-up Language)


SABIO
REACTION KINETICS DATABASE
CONTACT | HELP | IMPRINT
Reaction Search
Return only reactions having kinetic data matching the search criteria
Specify Search Criteria
Search Reaction
SBML Model Setup
with Reactants(s)
in Pathway(s)
Pyrimidine metabolism
Select Pathway
Delete Pathway
having Enzyme(s)
in Organism(s)
Homo sapiens
Select Organism
Delete Organism
in Tissue(s)/Cell Type(s)
in (Intra/Extra)Cellular L
Having Kinetic Data Deter
pH
Temperature
in Publication
reporting Parameters
EML Research
© EML Research gGmbH
Total number of reactions found for specified search criteria: 7
Click here to view your search criteria
Modify Search
Kinetic Data Availability:
Kinetic data available matching the search criteria
Kinetic data available, but not matching all search criteria
No kinetic data available
Number of results per page: 10
Display
Show only reactions having kinetic data matching the search criteria
Send Selected Reactions to SBML File
Reactions
Select Reaction(s)
(De)Select All
Kinetic Data for this reaction (Click to View)
Enzyme EC#
Kinetic data for enzymes (Click to View)
ATP + CMP <-> CDP + ADP
2.7.4.14
2.7.4.4
dCMP + ATP <-> ADP + dCDP
2.7.4.14
2.7.4.4
dUMP + 5,10-Methylenetetrahydrofolate <-> dTMP + Dihydrofolate
2.1.1.45
Thymidine + ATP <-> ADP + dTMP
2.7.1.21
2.7.1.145
UMP + ATP <-> ADP + UDP
2.7.4.14
2.7.4.4
dCTP + DNA <-> Diphosphate + DNA
2.7.7.7
dTTP + DNA <-> DNA + Diphosphate
2.7.7.7
Submit Search
Reset Form

Entry Nr. 5046 dUMP + 5,10-Methylenetetrahydrofolate <-> dTMP + Dihydrofolate

Organism:	Homo sapiens
Tissue:	unknown
EC Class: 2.1.1.45	**Variant:** wildtype

Reversability: reversible

Substrates

name	location	comment	External References
5,10-Methylenetetrahydrofolate	unknown	-	[KEGG: C00143; CHEBI: 15636;]
dUMP	unknown	-	[KEGG: C00365; CHEBI: 17622;]

Products

name	location	comment	External References
Dihydrofolate	unknown	-	[KEGG: C00415; CHEBI: 15633;]
dTMP	unknown	-	[KEGG: C00364; CHEBI: 17013;]

Modifiers

name	location	effect	comment	External References
E-5-(2-Bromovinyl)-2'-deoxyuridine monophosphate	unknown	Modifier-Inhibitor	active species of anticancer drug NB1011	
Thymidylate synthase(Enzyme)	unknown	Modifier-Catalyst	-	

Kinetic Law

```
(kcat*E*A)/(A+Km_A*(1+I/(Ki)))
```

Kinetic Law Type: Competitive inhibition

Parameters

name	species	type	St_value	Deviation	End_value	unit	comment
I	E-5-(2-Bromovinyl)-2'-deoxyuridine monophosphate	concentration	0	-	40	µM	-
Ki	E-5-(2-Bromovinyl)-2'-deoxyuridine monophosphate	Ki	4.17	0.29		µM	-
E	Enzyme	concentration	0.1	-		µM	-
kcatKm_A	dUMP	kcat/Km	79000	-		1/(M*sec)	-
Km_A	dUMP	Km	2.7	0.33		µM	-
kcat		kcat	0.213	0.020		1/sec	-
B	5,10-Methylenetetrahydrofolate	concentration	50	-		µM	saturated
A	dUMP	concentration	2	-	32	µM	-

Experimental conditions

	St_value	end_value	unit
pH	7.5	-	
Temperature	30	-	°C

Buffer: 40 mM Tris-HCl, 25 mM MgCl2, 1mM EDTA, 100 mM beta-mercaptoethanol

Comment: expressed in E.coli

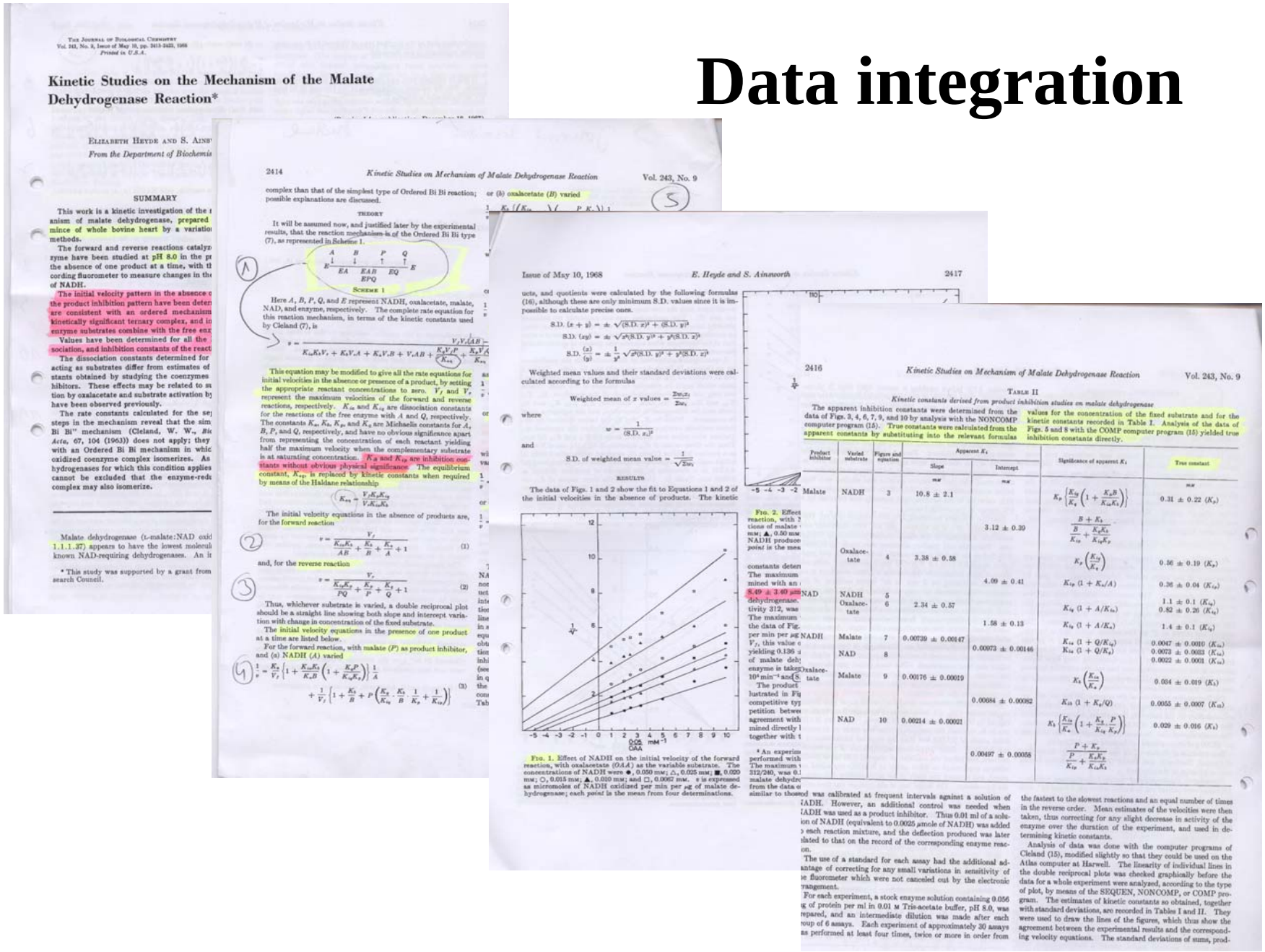
PUBMEDID: 12859954

SBML export

```
<listOfReactions>
- <reaction id="reac_0">
  - <listOfReactants>
      <speciesReference species="spc_1" />
      <speciesReference species="spc_2" />
    </listOfReactants>
  - <listOfProducts>
      <speciesReference species="spc_3" />
      <speciesReference species="spc_4" />
    </listOfProducts>
  - <listOfModifiers>
      <modifierSpeciesReference species="spc_5" />
      <modifierSpeciesReference species="enz_1" />
    </listOfModifiers>
  - <kineticLaw>
    - <math xmlns="http://www.w3.org
      - <apply>
          <ci>kl_0</ci>
          <ci>spc_5</ci>
          <ci>Ki</ci>
          <ci>enz_1</ci>
          <ci>kcatKm_A</ci>
          <ci>Km_A</ci>
          <ci>kcat</ci>
          <ci>spc_1</ci>
          <ci>spc_2</ci>
        </apply>
      </math>
    - <listOfParameters>
        <parameter id="Ki" value="4.17" units="microM" />
        <parameter id="kcatKm_A" value="79000" units="onedivMsec" />
        <parameter id="Km_A" value="2.7" units="microM" />
        <parameter id="kcat" value="0.213" units="onedivsec" />
```

```
<species id="spc_3" name="Dihydrofolate" compartment="compart_1">
- <annotation>
  - <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns" xmlns:dc="http://purl.org/dc/elements/1.1/">
    - <rdf:Description rdf:about="#Dihydrofolate">
      - <dc:relation>
        - <rdf:Bag>
            <rdf:li rdf:resource="http://www.genome.jp/dbget-bin/www_bget?cpd:C00415" />
            <rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/searchId.do?chebiId=15633" />
          </rdf:Bag>
        </dc:relation>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
<species id="spc_4" name="dTMP" compartment="compart_1">
- <annotation>
  - <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns" xmlns:dc="http://purl.org/dc/elements/1.1/">
    - <rdf:Description rdf:about="#dTMP">
      - <dc:relation>
        - <rdf:Bag>
            <rdf:li rdf:resource="http://www.genome.jp/dbget-bin/www_bget?cpd:C00364" />
            <rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/searchId.do?chebiId=17013" />
```

Data integration



THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 243, No. 9, Issue of May 10, pp. 2413-2423, 1968
Printed in U.S.A.

Kinetic Studies on the Mechanism of the Malate Dehydrogenase Reaction*

ELIZABETH HEYDE AND S. AINS
From the Department of Biochemis

SUMMARY

This work is a kinetic investigation of the anism of malate dehydrogenase, prepared mince of whole bovine heart by a variatio methods.

The forward and reverse reactions catalyz zyme have been studied at pH 8.0 in the pr the absence of one product at a time, with th cording fluorometer to measure changes in the of NADH.

The initial velocity pattern in the absence the product inhibition pattern have been deter are consistent with an ordered mechanism kinetically significant ternary complex, and in enzyme substrates combine with the free enz

Values have been determined for all the sociation, and inhibition constants of the react

The dissociation constants determined for acting as substrates differ from estimates of stants obtained by studying the coenzymes hibitors. These effects may be related to st tion by oxalacetate and substrate activation by have been observed previously.

The rate constants calculated for the se steps in the mechanism reveal that the sim Bi Bi" mechanism (Cleland, W. W., *Bio Acta*, 67, 104 (1963)) does not apply; they with an Ordered Bi Bi mechanism in whic oxidized coenzyme complex isomerizes. As hydrogenases for which this condition applies cannot be excluded that the enzyme-redu complex may also isomerize.

Malate dehydrogenase (L-malate:NAD oxid 1.1.1.37) appears to have the lowest molecul known NAD-requiring dehydrogenases. An i

* This study was supported by a grant from search Council.

2414 *Kinetic Studies on Mechanism of Malate Dehydrogenase Reaction* Vol. 243, No. 9

complex than that of the simplest type of Ordered Bi Bi reaction; or (*b*) oxalacetate (*B*) varied possible explanations are discussed.

THEORY

It will be assumed now, and justified later by the experimental results, that the reaction mechanism is of the Ordered Bi Bi type (7), as represented in Scheme 1.

SCHEME 1

Here *A*, *B*, *P*, *Q*, and *E* represent NADH, oxalacetate, malate, NAD, and enzyme, respectively. The complete rate equation for this reaction mechanism, in terms of the kinetic constants used by Cleland (7), is

This equation may be modified to give all the rate equations for initial velocities in the absence or presence of a product, by setting the appropriate reactant concentrations to zero. V_f and V_r represent the maximum velocities of the forward and reverse reactions, respectively. K_{ia} and K_{ib} are dissociation constants for the reactions of the free enzyme with *A* and *Q*, respectively. The constants K_a, K_b, K_p, and K_q are Michaelis constants for *A*, *B*, *P*, and *Q*, respectively, and have no obvious significance apart from representing the concentration of each reactant yielding half the maximum velocity when the complementary substrate is at saturating concentration. K_{ia} and K_{ib} are inhibition constants without obvious physical significance. The equilibrium constant, K_{eq}, is replaced by kinetic constants when required by means of the Haldane relationship

The initial velocity equations in the absence of products are, for the forward reaction

(1)

and, for the reverse reaction

(2)

Thus, whichever substrate is varied, a double reciprocal plot should be a straight line showing both slope and intercept variation with change in concentration of the fixed substrate.

The initial velocity equations in the presence of one product at a time are listed below.

For the forward reaction, with malate (*P*) as product inhibitor, and (*a*) NADH (*A*) varied

(3)

Issue of May 10, 1968 *E. Heyde and S. Ainsworth* 2417

ucts, and quotients were calculated by the following formulas (16), although these are only minimum S.D. values since it is impossible to calculate precise ones.

Weighted mean values and their standard deviations were calculated according to the formulas

Weighted mean of x values $= \frac{\Sigma w_i x_i}{\Sigma w_i}$

where

$$w = \frac{1}{(\text{S.D. } x_i)^2}$$

and

S.D. of weighted mean value $= \frac{1}{\sqrt{\Sigma w_i}}$

RESULTS

The data of Figs. 1 and 2 show the fit to Equations 1 and 2 of the initial velocities in the absence of products. The kinetic

FIG. 1. Effect of NADH on the initial velocity of the forward reaction, with oxalacetate (*OAA*) as the variable substrate. The concentrations of NADH were ●, 0.050 mM; △, 0.025 mM; ■, 0.020 mM; ○, 0.015 mM; ▲, 0.010 mM; and □, 0.0067 mM. *v* is expressed as micromoles of NADH oxidized per min per μg of malate dehydrogenase; each point is the mean from four determinations.

FIG. 2. Effect
reaction, with 7
tions of malate
mM; ▲, 0.50 mM
NADH produce
point is the mea

constants deter
The maximum
mined with an
8.49 ± 2.03 μm
dehydrogenase.
tivity 312, was
The maximum
the data of Fig.
per min per μg
V_r, this value
yielding 0.126
of malate deh
enzyme is take
10^5 min^{-1} and
The product
lustrated in Fi
competitive typ
petition betwe
agreement with
mined directly
together with

[a] An experim
performed with
The maximum
312/240, was 0.
malate dehydr
from the data
similar to those

2416 *Kinetic Studies on Mechanism of Malate Dehydrogenase Reaction* Vol. 243, No. 9

TABLE II

Kinetic constants derived from product inhibition studies on malate dehydrogenase

The apparent inhibition constants were determined from the data of Figs. 3, 4, 6, 7, 9, and 10 by analysis with the NONCOMP computer program (15). True constants were calculated from the apparent constants by substituting into the relevant formulas values for the concentration of the fixed substrate and for the kinetic constants recorded in Table I. Analysis of the data of Figs. 5 and 8 with the COMP computer program (15) yielded true inhibition constants directly.

Product inhibitor	Varied substrate	Figure and equation	Apparent K_i		Significance of apparent K_i	True constant
			Slope	Intercept		
			mM	mM		mM
Malate	NADH	3	10.8 ± 2.1		$K_p\left(\frac{K_{ia}}{K_a}\left(1+\frac{K_bB}{K_{ia}K_b}\right)\right)$	0.31 ± 0.22 (K_p)
				3.12 ± 0.39	$\frac{B+K_b}{\frac{B}{K_{ia}}+\frac{K_bK_b}{K_{ia}K_p}}$	
	Oxalacetate	4	3.38 ± 0.58		$K_p\left(\frac{K_{ia}}{K_a}\right)$	0.56 ± 0.19 (K_p)
				4.09 ± 0.41	$K_{ip}(1+K_a/A)$	0.36 ± 0.04 (K_{ia})
NAD	NADH	5				1.1 ± 0.1 (K_{iq})
	Oxalacetate	6	2.34 ± 0.57		$K_{iq}(1+A/K_{ia})$	0.82 ± 0.26 (K_{ia})
				1.58 ± 0.13	$K_{iq}(1+A/K_a)$	1.4 ± 0.1 (K_q)
NADH	Malate	7	0.00739 ± 0.00147		$K_{ia}(1+Q/K_{iq})$	0.0047 ± 0.0010 (K_{ia})
				0.00973 ± 0.00146	$K_{ia}(1+Q/K_a)$	0.0072 ± 0.0033 (K_{ia})
	NAD	8				0.0022 ± 0.0001 (K_{ia})
Oxalacetate	Malate	9	0.00176 ± 0.00019		$K_b\left(\frac{K_{ia}}{K_a}\right)$	0.004 ± 0.019 (K_b)
				0.00084 ± 0.00082	$K_{ib}(1+K_a/Q)$	0.0055 ± 0.0007 (K_{ia})
	NAD	10	0.00214 ± 0.00021		$K_b\left(\frac{K_{ia}}{K_a}\left(1+\frac{K_a}{K_{iq}}\frac{P}{K_p}\right)\right)$	0.029 ± 0.016 (K_b)
				0.00497 ± 0.00058	$\frac{P+K_p}{\frac{P}{K_{ip}}+\frac{K_pK_p}{K_{ia}K_b}}$	

was calibrated at frequent intervals against a solution of ADH. However, an additional control was needed when ADH was used as a product inhibitor. Thus 0.01 ml of a solution of NADH (equivalent to 0.0025 μmole of NADH) was added to each reaction mixture, and the deflection produced was later related to that on the record of the corresponding enzyme reaction.

The use of a standard for each assay had the additional advantage of correcting for any small variations in sensitivity of the fluorometer which were not canceled out by the electronic arrangement.

For each experiment, a stock enzyme solution containing 0.056 mg of protein per ml in 0.01 M Tris-acetate buffer, pH 8.0, was prepared, and an intermediate dilution was made after each group of 6 assays. Each experiment of approximately 30 assays was performed at least four times, twice or more in order from the fastest to the slowest reactions and an equal number of times in the reverse order. Mean estimates of the velocities were then taken, thus correcting for any slight decrease in activity of the enzyme over the duration of the experiment, and used in determining kinetic constants.

Analysis of data was done with the computer programs of Cleland (15), modified slightly so that they could be used on the Atlas computer at Harwell. The linearity of individual lines in the double reciprocal plots was checked graphically before the data for a whole experiment were analyzed, according to the type of plot, by means of the SEQUEN, NONCOMP, or COMP program. The estimates of kinetic constants so obtained, together with standard deviations, are recorded in Tables I and II. They were used to draw the lines of the figures, which thus show the agreement between the experimental results and the corresponding velocity equations. The standard deviations of sums, prod-

Information source

- Publications
  - Manual extraction

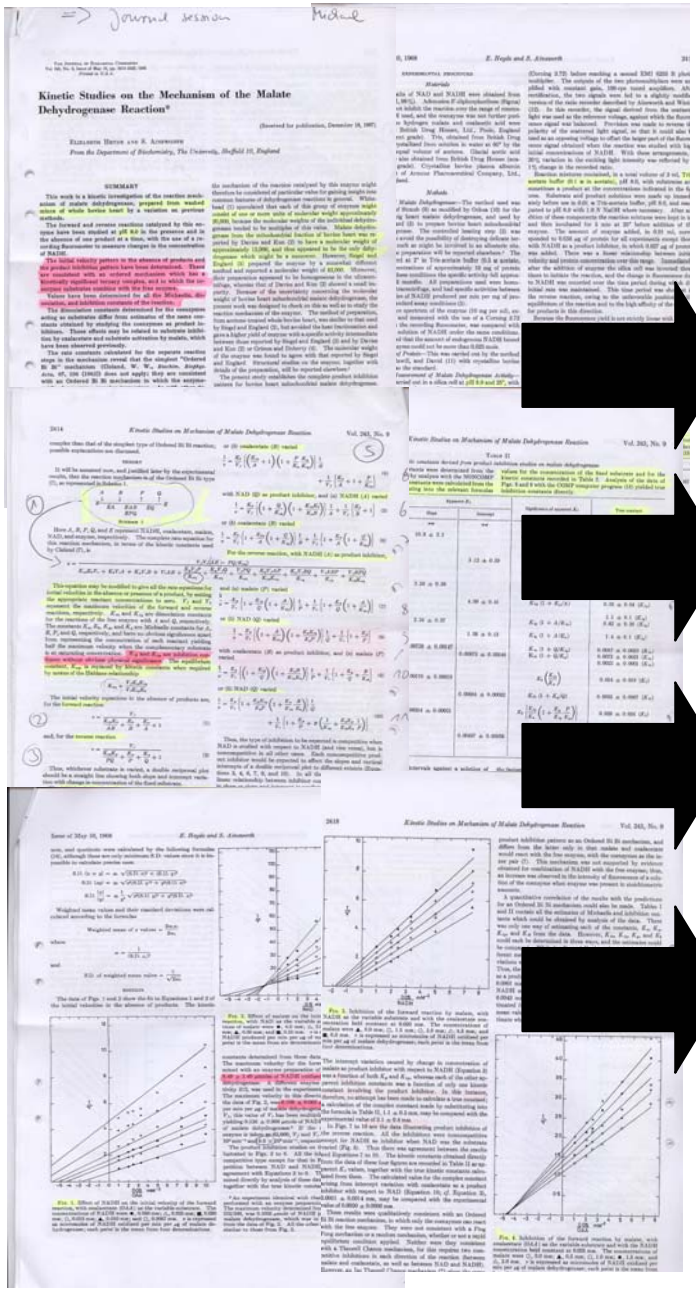
    → no automatic information extraction at the moment

    → data stored in tables, formulas, graphs

→ **Input interface**

- web interface
- structuring of data from literature

Input interface


Kinetic Studies on the Mechanism of the Malate Dehydrogenase Reaction*

ELIZABETH HAYON AND S. AINSWORTH

From the Department of Biochemistry, The University, Sheffield 10, England

SUMMARY

This work is a kinetic investigation of the reaction mechanism of malate dehydrogenase, prepared from washed mince of whole bovine heart by a variation on previous methods.

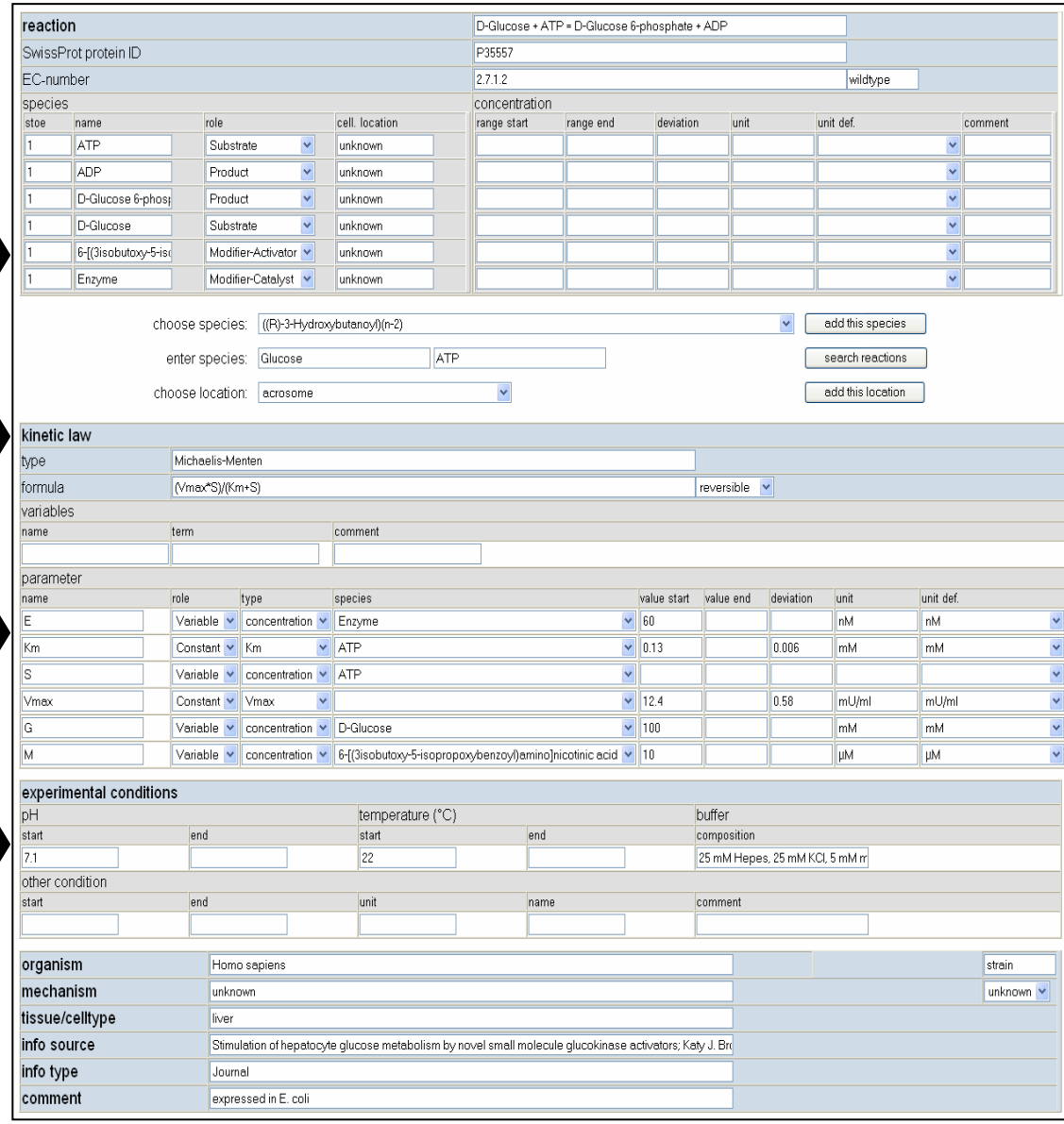
The forward and reverse reactions catalyzed by this enzyme have been studied at pH 8.6 in the presence and in the absence of one product at a time, with the use of a recording fluorometer to measure changes in the concentration of NADH.

The initial velocity pattern in the absence of products and no product inhibition pattern have been determined. These are consistent with an ordered mechanism which has a kinetically significant ternary complex, and in which the latter reaction substrate combines with the free enzyme.

Values have been determined for all the Michaelis...

The dissociation constants determined for the enzyme-acting as substrates differ from estimates of the same constants obtained by studying the reactions as product inhibitors. These effects may be related to substrate inhibition by oxaloacetate and substrate activation by malate, which have been observed previously.

The rate constants calculated for the separate reaction steps in the mechanism reveal that the simplest "Ordered Bi Bi" mechanism (Dalziel, W. W., Biochim. Biophys. Acta, 05, 106 [1962]) does not apply; they are consistent with an Ordered Bi Bi mechanism in which the enzyme...



Insert procedure

- Input interface
- Data first inserted in an intermediate database
- Curation process (search for errors and inconsistencies)
  - Manually by biological experts
  - Semi-automatically (supported by NLP tools)
- Automatic search for already existing compounds, reactions, organisms, etc. in SABIO-RK
- Insert new compounds, reactions, etc. if not already in SABIO-RK
- Transfer data from intermediate to relational SABIO-RK database (Oracle)

- User interface (output, export)

Database population and annotation

- Most of the reactions, their associations with biochemical pathways as well as enzyme classifications are downloaded from KEGG Ligand database *(http://www.genome.ad.jp/kegg/ligand.html)*

- Use of controlled vocabularies
  - for systematic names of organism → NCBI taxonomy *(http://www.ncbi.nlm.nih.gov/Taxonomy/)*
  - for enzymes → IUBMB recommendations *(http://www.chem.qmul.ac.uk/iubmb/enzyme/)*
  - for compound names → IUPAC recommendations *(http://www.chem.qmul.ac.uk/iupac/)*
  - for parameter units → SI system for unit notation

  etc.

- Links to other databases (KEGG, ChEBI, Swiss-Prot, PubMed etc.) and in future annotations (Systems Biology Ontology *http://www.ebi.ac.uk/compneur-srv/sbo/*)

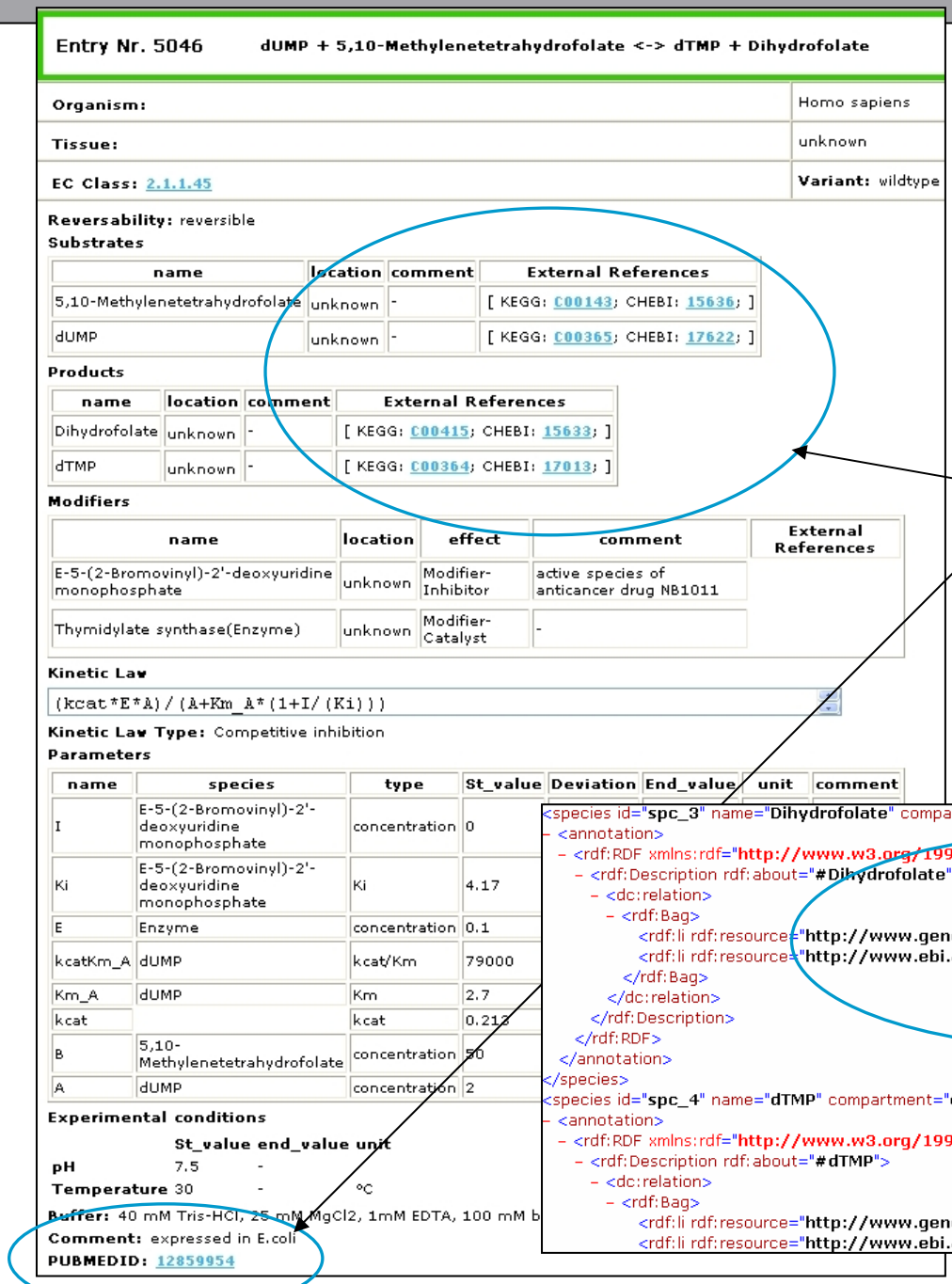
Multiplicity of units

	Extracted from paper	Internal identified/grouped as
1	unknown	-
2	%	%
3	µg/µl	µg/µl
4	µM	µM
4	µmol/l	µM
5	(µM)^(2)	µM*µM
6	µM/min	µM/min
7	µmole	µmol
7	µmoles	µmol
8	µmol*min^(-1)*µl^(-1)	µmol/(min*µl)
9	µmol*s^(-1)*mg^(-1)	µmol/(sec*mg)
10	µmol/ml	µmol/ml
11	µM^(-1)* min^(-1)	1/(µM*min)
11	µM^(-1)*min^(-1)	1/(µM*min)
11	min^(-1)*µM^(-1)	1/(µM*min)
12	µM^(-1)*s^(-1)	1/(µM*sec)
12	s^(-1)*µM^(-1)	1/(µM*sec)
13	1/h*1/mg	1/(h*mg)
14	h^(-1)	1/h
14	h^-1	1/h
14	1/h	1/h
15	1/min	1/min

Annotations

Links to other Databases

Annotation in SBML



Entry Nr. 5046 dUMP + 5,10-Methylenetetrahydrofolate <-> dTMP + Dihydrofolate

Organism:	Homo sapiens
Tissue:	unknown
EC Class: 2.1.1.45	**Variant:** wildtype

Reversability: reversible

Substrates

name	location	comment	External References
5,10-Methylenetetrahydrofolate	unknown	-	[KEGG: C00143; CHEBI: 15636;]
dUMP	unknown	-	[KEGG: C00365; CHEBI: 17622;]

Products

name	location	comment	External References
Dihydrofolate	unknown	-	[KEGG: C00415; CHEBI: 15633;]
dTMP	unknown	-	[KEGG: C00364; CHEBI: 17013;]

Modifiers

name	location	effect	comment	External References
E-5-(2-Bromovinyl)-2'-deoxyuridine monophosphate	unknown	Modifier-Inhibitor	active species of anticancer drug NB1011	
Thymidylate synthase(Enzyme)	unknown	Modifier-Catalyst	-	

Kinetic Law

(kcat*E*A)/(A+Km_A*(1+I/(Ki)))

Kinetic Law Type: Competitive inhibition

Parameters

name	species	type	St_value	Deviation	End_value	unit	comment
I	E-5-(2-Bromovinyl)-2'-deoxyuridine monophosphate	concentration	0				
Ki	E-5-(2-Bromovinyl)-2'-deoxyuridine monophosphate	Ki	4.17				
E	Enzyme	concentration	0.1				
kcatKm_A	dUMP	kcat/Km	79000				
Km_A	dUMP	Km	2.7				
kcat		kcat	0.213				
B	5,10-Methylenetetrahydrofolate	concentration	50				
A	dUMP	concentration	2				

Experimental conditions

	St_value	end_value	unit
pH	7.5	-	
Temperature	30	-	°C

Buffer: 40 mM Tris-HCl, 25 mM MgCl2, 1mM EDTA, 100 mM b

Comment: expressed in E.coli

PUBMEDID: 12859954

```
<species id="spc_3" name="Dihydrofolate" compartment="compart_1">
 <annotation>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns" xmlns:dc="http://purl.org/dc/elements/1.1/">
   <rdf:Description rdf:about="#Dihydrofolate">
    <dc:relation>
     <rdf:Bag>
      <rdf:li rdf:resource="http://www.genome.jp/dbget-bin/www_bget?cpd:C00415" />
      <rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/searchId.do?chebiId=15633" />
     </rdf:Bag>
    </dc:relation>
   </rdf:Description>
  </rdf:RDF>
 </annotation>
</species>
<species id="spc_4" name="dTMP" compartment="compart_1">
 <annotation>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns" xmlns:dc="http://purl.org/dc/elements/1.1/">
   <rdf:Description rdf:about="#dTMP">
    <dc:relation>
     <rdf:Bag>
      <rdf:li rdf:resource="http://www.genome.jp/dbget-bin/www_bget?cpd:C00364" />
      <rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/searchId.do?chebiId=17013" />
```

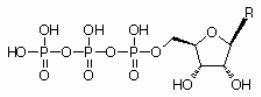
Problems in curation process

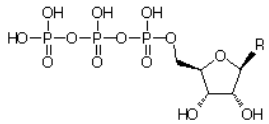
- Missing or only partial information
  - incomplete reactions (products not mentioned)
  - assay conditions missing or reference to another paper
  - kinetic law (or fitting equation) not described
- Complexity in the description of buffers
  - e.g. coupled enzyme assay
- Identification of compounds, reactions and enzymes
  - usage of unusual synonymic names
  - isoenzyme not specified
- Multiplicity of parameter units
  - e.g. katal, U, µmol/(s*mg), mM/min for enzymatic activity
- Kinetic law types
  - no controlled vocabulary available

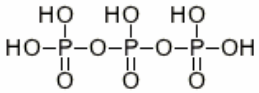
Curation

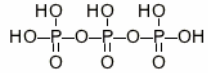
- Search for multiple entries for identical compounds

examples from KEGG database

Entry	C00201 Compound
Name	Nucleoside triphosphate; NTP
Formula	C5H12O13P3R
Mass	372.9489
Structure	 C00201 Mol file KCF file DB search
Reaction	R00331 R00333 R00444 R01532 R02319 R02320 R02321 R02432 R03149 R05512
Enzyme	2.7.1.40 2.7.1.74 2.7.4.6 2.7.4.10 2.7.7.6 2.7.7.28 2.7.7.46 2.7.7.48 3.6.1.15 3.6.1.19
Other DBs	PubChem: 3501
LinkDB	All DBs
KCF data	Show

Entry	C03802 Compound
Name	Ribonucleoside triphosphate
Formula	C5H12O13P3R
Mass	372.9489
Structure	 C03802 Mol file SIMCOMP
Reaction	R04315
RPair	A03950
Enzyme	1.17.4.2
Other DBs	PubChem: 6551
LinkDB	All DBs
KCF	Show

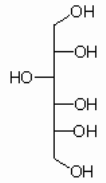
Entry	C03279 Compound
Name	Inorganic triphosphate
Formula	H5O10P3
Mass	257.9096
Structure	 C03279 Mol file KCF file DB search
Other DBs	PubChem: 6138
LinkDB	All DBs
KCF data	Show

Entry	C00536 Compound
Name	Triphosphate; Tripolyphosphate
Formula	H5O10P3
Mass	257.9096
Structure	 C00536 Mol file KCF file DB search
Reaction	R00136 R00138 R01492 R01856 R02504 R04286 R05220 R07268
Pathway	PATH: map00190 Oxidative phosphorylation
Enzyme	2.5.1.17 2.7.4.1 3.1.5.1 3.6.1.2 3.6.1.25 4.2.3.12
Other DBs	PubChem: 3818 ChEBI: 29203 3DMET: B00124
LinkDB	All DBs

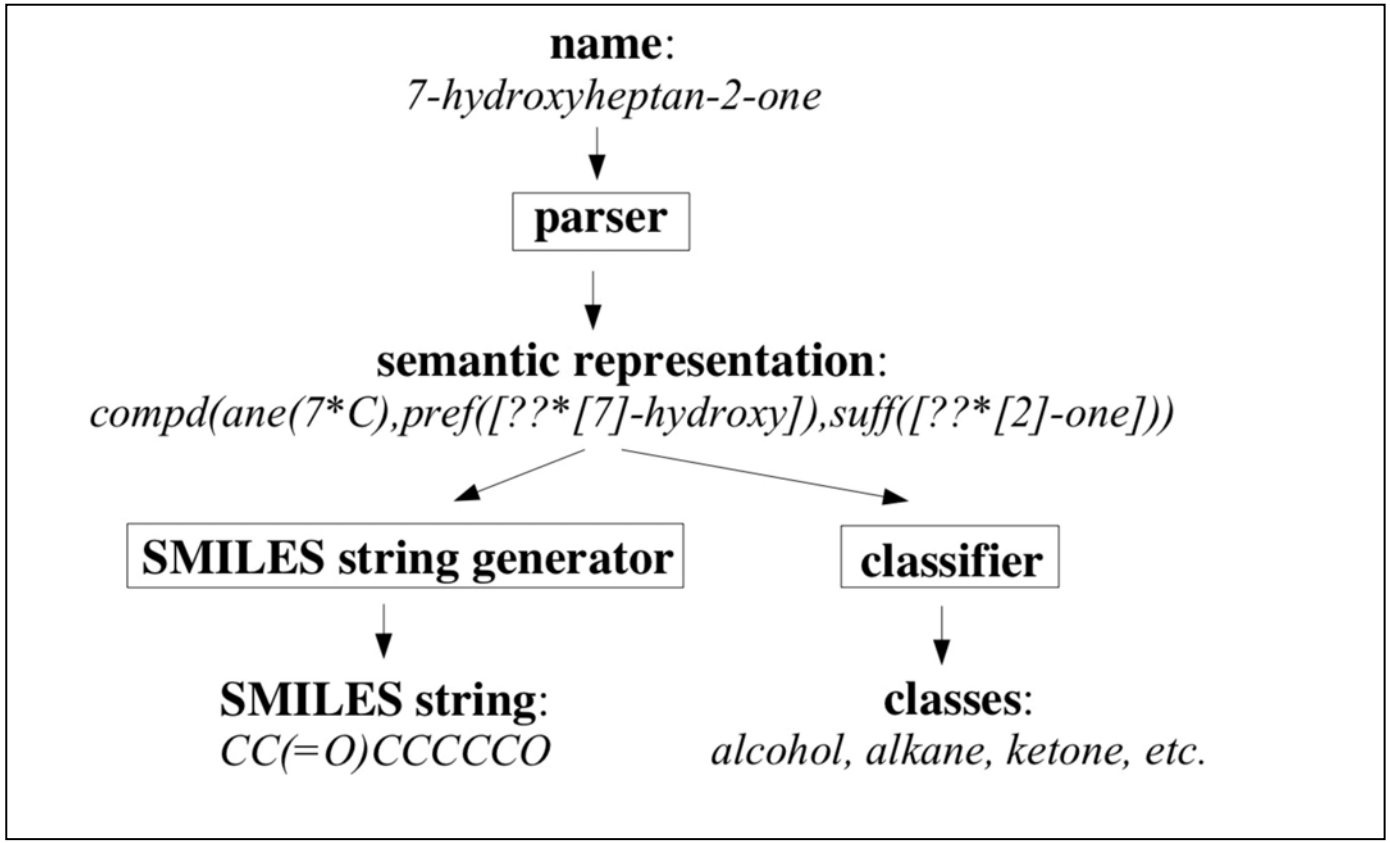
Curation

- Search for multiple entries for identical compounds
  - ID 1371 D-Sorbitol 6-phosphate
  - ID 21224 D-Glucitol 6-phosphate

example from SABIO-RK database

Entry	C00794 Compound
Name	D-Sorbitol; D-Glucitol; L-Gulitol; Sorbitol
Formula	C6H14O6
Mass	182.0791
Structure	 C00794 Mol file KCF file DB search
Reaction	R00874 R00875 R01697 R01787 R02865 R02866 R02867 R02868 R02925 R02926 R05820
Pathway	PATH: map00051 Fructose and mannose metabolism PATH: map00052 Galactose metabolism PATH: map02060 Phosphotransferase system (PTS)
Enzyme	1.1.1.14 1.1.1.15 1.1.1.21 1.1.99.21 1.1.99.28 2.7.1.1 2.7.1.69 3.1.3.50 3.2.1.22
Other DBs	CAS: 50-70-4 PubChem: 4052

Curation support
NLP



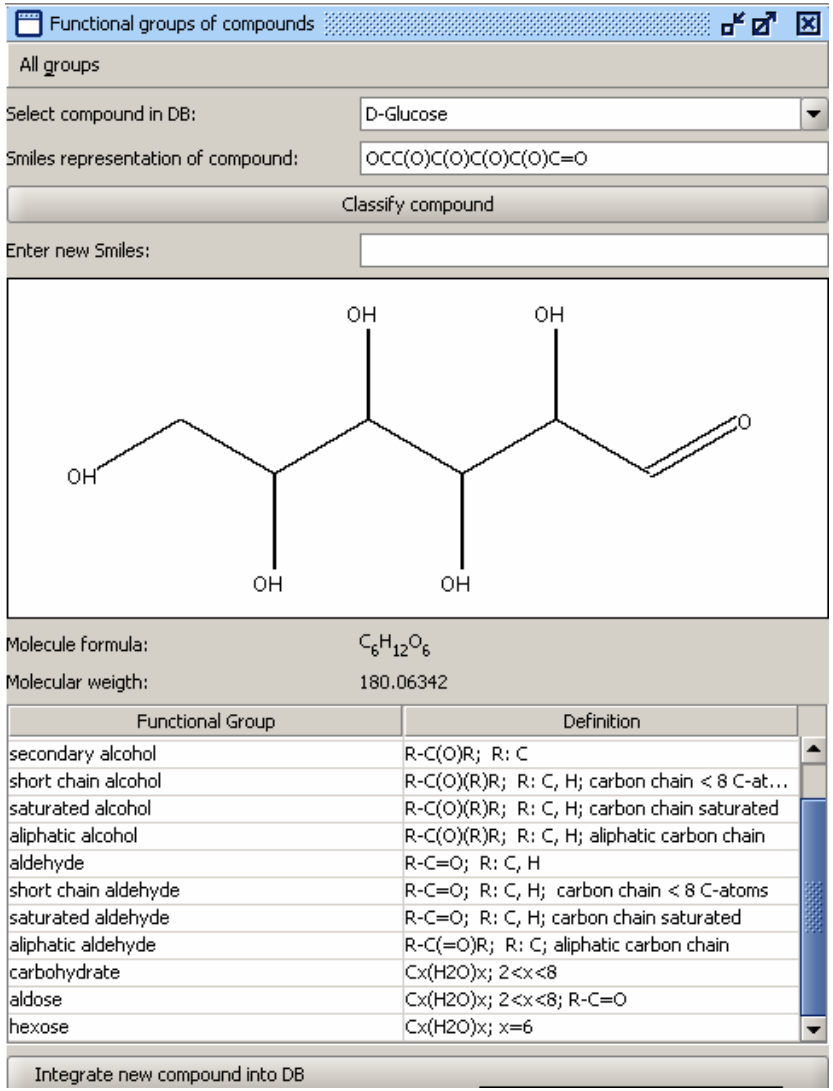
Classification of Compounds

- List of definitions for compound classes and functional groups

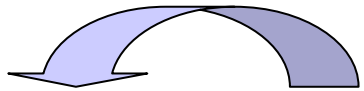
- Automatic generation of structural formula, totals formula and molecular weight

- Classification using different criteria

Thus D-Glucose is a:
- Aldose
  (functional group aldehyde)
- Hexose
  (number of C-Atoms = 6)



Classification of Compounds: The overall architecture

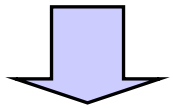


Structured Input Data

Import of structured data: SMILES, Mol-File....

Unstructured Input Data

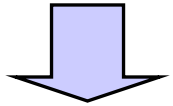
Import of chemical compound names



Conversion into graphs

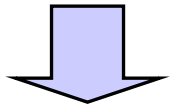
Atoms are represented as nodes
Bonds are represented as edges

Based on Chemical Development Kit API
(http://cdk.sourceforge.net/api.html)



Classification

- Analysis of graph structure, i.e. detection of simple functional groups (e.g. aldehyde, amines, ketones, etc.).
- Use of combinations of simple functional groups to detect higher order structures (e.g. nucleotides, carbohydrates, aldoses, hexoses...)



Output and Visualisation

- Group definitions (at present: about 200 definitions)
- Graphical representation of the molecule
- Storage of graph object as file for structure comparisons

Querying for chemical compounds

Querying PubMed or a database:
Find all biochemical reactions with D-Glucose as participant!

Output with means of string matching:

EC 5.1.3.3	alpha-D-Glucose 1-epimerase
	alpha-D-Glucose ⇔ beta-D-Glucose

Missing reactions for general molecules:

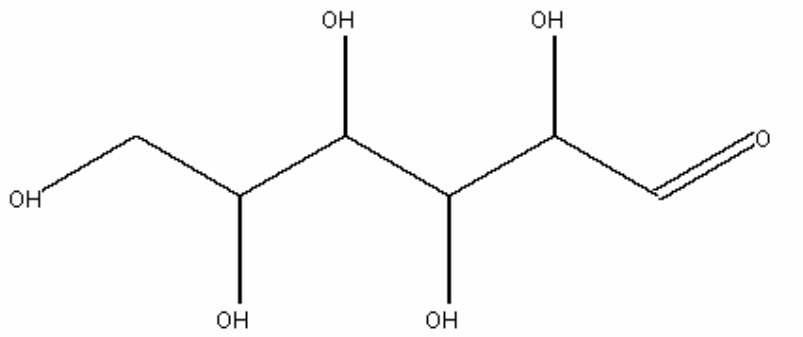
EC 1.1.1.21	aldehyde reductase
	alditol + NAD(P) → aldose + NAD(P)H
EC 2.7.1.1.	hexokinase
	ATP + D-hexose → ADP + D-hexose-6-phosphate

Reason:
D-Glucose is a specific aldose or a specific hexose

Deriving a semantic annotation

Glucose can be classified as

- alcohol C-OH
- aldehyde C=O
- carbohydrate $(C H_2O)_x$
- hexose $C_6(H_2O)_6$ 6 C-atoms
- aldose $(C H_2O)_x$ + -CH=O
- aldohexose $C_6(H_2O)_6$ + -CH=O
- monosaccaride $[(C H_2O)_x]_1$



Glucose $C_6H_{12}O_6$

Therefore the correct output should be

EC 5.1.3.3 a-D-Glucose 1-epimerase
a-D-Glucose ⇔ ß-D-Glucose

EC 1.1.1.21 aldehyde reductase
alditol + NAD(P) → aldose + NAD(P)H

EC 2.7.1.1. hexokinase
ATP + D-hexose → ADP + D-hexose-6-phosphate

Curation support
NLP

- Search for multiple entries for identical compounds
  - Linguistic analysis of chemical compound names
  - Representation of compound structure (SMILES) based on compound name
  - Building graphs based on the SMILES
  - Search for identical graphs or subgraphs

  - Classification of compounds based on structural similarities

  → basis for automatic information extraction of compound information from publication

Standardization

SABIO-RK implements the recommendations of the STRENDA commission

STRENDA = **St**andards for **R**eporting **E**nzymology **Da**ta

(http://www.strenda.org/)

→ Authors of publications insert their data into a database

- structuring of data
- full documentation of experimental conditions is needed
- online access to the data

SABIO-RK statistics

(as of July 2006)

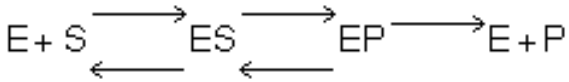
- SABIO-RK data extacted from literature
  - 623 curated publications
  - 5550 database entries (40% with rate equation)
  - kinetic data for 210 organisms
  - 1160 biochemical reactions (340 enzymes) related to kinetic data
  - 19838 chemical compound names
  - 13470 different entries (IDs) for chemical compounds
  - Numbers of synonyms per entry:
    - Maximum 28
    - Average 1,5

Conclusion

- **SABIO-RK**
  - Is a web-accessible database containing biochemical reaction kinetics
  - Merges general reaction information retrieved from other databases and kinetic data manually extracted from literature
  - Is structuring literature information
  - Is curated by biological experts
  - Has a high degree of interrelation (all necessary information is linked)
  - Offers data export in SBML format

Future directions

- Information about reaction mechanism
  - separate reactions for intermediate steps



  - no database contains such data at the moment
- More information about signalling reactions/pathways
- Information about protein complexes
  - information from literature and/or from other databases (UniProt etc.)
- Use of the database as a standard source for reaction kinetics data
  - Scientist could use the database to store data in a structured format (Input interface)

Acknowledgement

SDBV group at EML Research:
Isabel Rojas
Renate Kania
Saqib Mir
Andreas Weideman
Olga Krebs
Martin Golebiewski
Stefanie Anstein
Jasmin Saric

Systems biology people at EML Research:
Ursula Kummer
Ralph Gauges
Sven Sahle
Rebecca Wade
Mathias Stein

Financial support:

KTS

Bundesministerium
für Bildung
und Forschung

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