

EML
Research

Bridging Experiment and Model: The SABIO-RK Database

Martin Golebiewski

Scientific Databases and Visualization Group

EML Research, Heidelberg

Super-Hackathon “Standards and Ontologies for Systems Biology

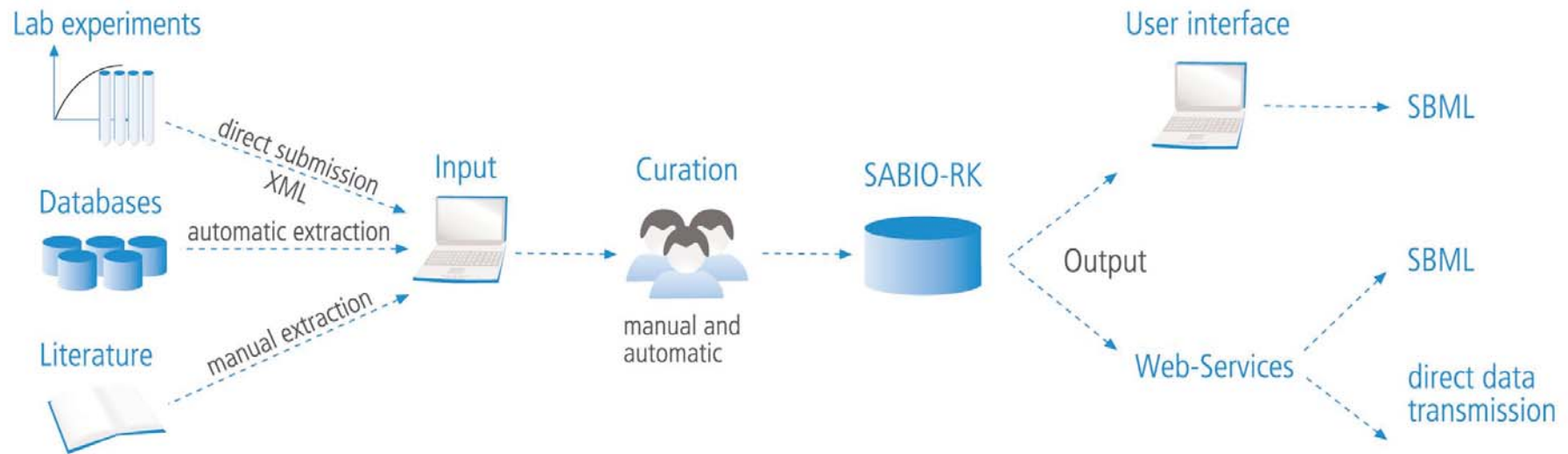
Okinawa, January 28 to February 2

Why we have developed SABIO-RK ?

- Biochemical model simulations need experimental reaction kinetics data
- Kinetic parameter values highly depend on environmental conditions
- Enzyme characteristics vary between organisms, tissues and cellular locations
- Kinetic parameters are only interpretable with their corresponding kinetic laws
- To link data for single reactions to complete sets of corresponding information
- Data must be easily accessible and interchangeable (data export for exchange)

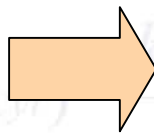
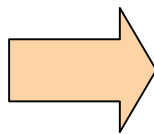
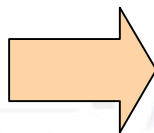
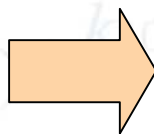
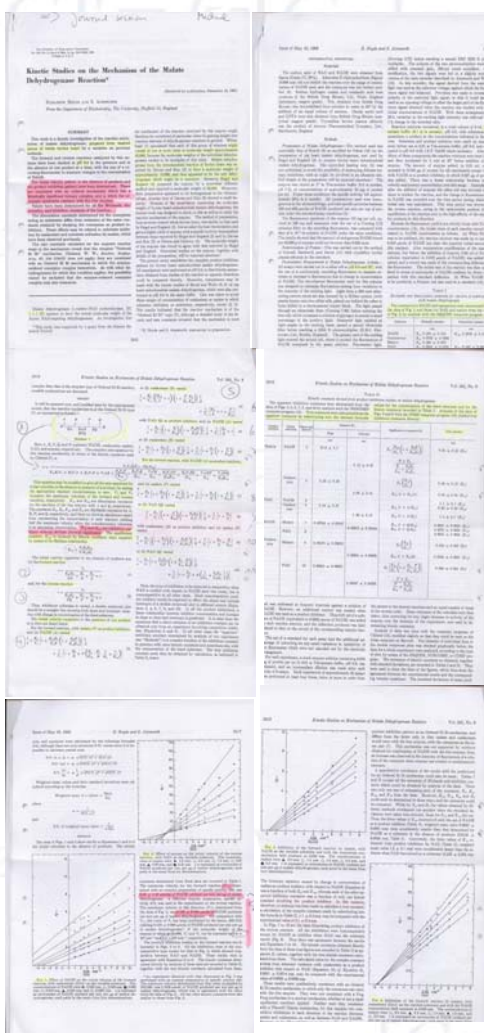
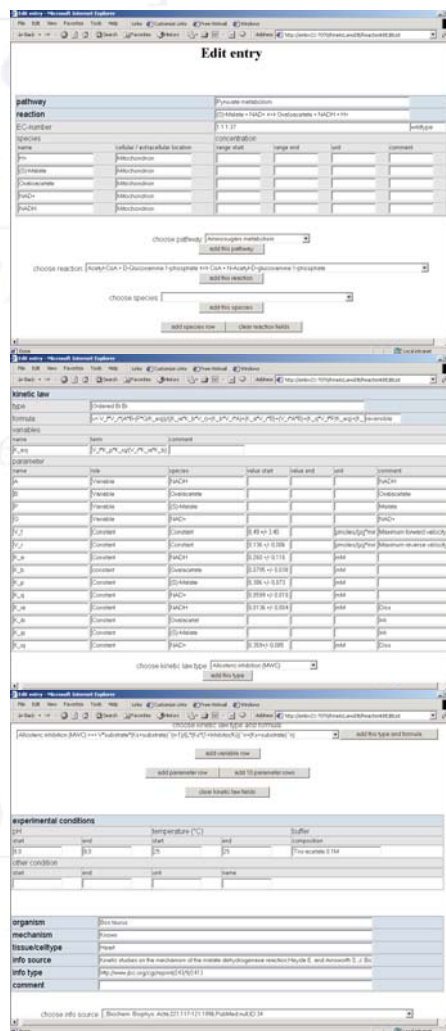
We aimed at creating a database that collects and standardizes kinetic data, relates the data to its biochemical, environmental and experimental context, cross-links corresponding data and associates it with external resources to make the data comparable and accessible in standard formats

Database Population and Access



- Information about **biochemical reactions** and pathways mainly collected from other databases (e.g. KEGG) merged with corresponding **kinetic data**
- Manual curation, assisted by tools for standardization and integration
- Data is unified, systematically structured, normalized, interrelated and annotated
- Access through a web-based user interface and through web-services (API)
- Data export in Systems Biology Markup Language (SBML)

Database Population: Literature Extraction

Edit entry

pathway:

reaction:

EC number:

species	cellular / intracellular location	range start	range end	unit	comment
Glucose	intracellular				
Glucose-6-phosphate	intracellular				
H+	intracellular				
H2O	intracellular				

kinetic law:

parameter	unit	species	value start	value end	unit	comment
k_1	Variable	Protein				Protein
k_{-1}	Variable	Protein				Protein
k_2	Variable	Glucose				Glucose
k_{-2}	Variable	Glucose				Glucose
k_3	Variable	Protein				Protein
k_4	Variable	Protein				Protein
k_5	Variable	Protein				Protein
k_6	Variable	Protein				Protein
k_7	Variable	Protein				Protein
k_8	Variable	Protein				Protein
k_9	Variable	Protein				Protein
k_{10}	Variable	Protein				Protein
k_{11}	Variable	Protein				Protein
k_{12}	Variable	Protein				Protein
k_{13}	Variable	Protein				Protein
k_{14}	Variable	Protein				Protein
k_{15}	Variable	Protein				Protein
k_{16}	Variable	Protein				Protein
k_{17}	Variable	Protein				Protein
k_{18}	Variable	Protein				Protein
k_{19}	Variable	Protein				Protein
k_{20}	Variable	Protein				Protein
k_{21}	Variable	Protein				Protein
k_{22}	Variable	Protein				Protein
k_{23}	Variable	Protein				Protein
k_{24}	Variable	Protein				Protein
k_{25}	Variable	Protein				Protein
k_{26}	Variable	Protein				Protein
k_{27}	Variable	Protein				Protein
k_{28}	Variable	Protein				Protein
k_{29}	Variable	Protein				Protein
k_{30}	Variable	Protein				Protein
k_{31}	Variable	Protein				Protein
k_{32}	Variable	Protein				Protein
k_{33}	Variable	Protein				Protein
k_{34}	Variable	Protein				Protein
k_{35}	Variable	Protein				Protein
k_{36}	Variable	Protein				Protein
k_{37}	Variable	Protein				Protein
k_{38}	Variable	Protein				Protein
k_{39}	Variable	Protein				Protein
k_{40}	Variable	Protein				Protein
k_{41}	Variable	Protein				Protein
k_{42}	Variable	Protein				Protein
k_{43}	Variable	Protein				Protein
k_{44}	Variable	Protein				Protein
k_{45}	Variable	Protein				Protein
k_{46}	Variable	Protein				Protein
k_{47}	Variable	Protein				Protein
k_{48}	Variable	Protein				Protein
k_{49}	Variable	Protein				Protein
k_{50}	Variable	Protein				Protein
k_{51}	Variable	Protein				Protein
k_{52}	Variable	Protein				Protein
k_{53}	Variable	Protein				Protein
k_{54}	Variable	Protein				Protein
k_{55}	Variable	Protein				Protein
k_{56}	Variable	Protein				Protein
k_{57}	Variable	Protein				Protein
k_{58}	Variable	Protein				Protein
k_{59}	Variable	Protein				Protein
k_{60}	Variable	Protein				Protein
k_{61}	Variable	Protein				Protein
k_{62}	Variable	Protein				Protein
k_{63}	Variable	Protein				Protein
k_{64}	Variable	Protein				Protein
k_{65}	Variable	Protein				Protein
k_{66}	Variable	Protein				Protein
k_{67}	Variable	Protein				Protein
k_{68}	Variable	Protein				Protein
k_{69}	Variable	Protein				Protein
k_{70}	Variable	Protein				Protein
k_{71}	Variable	Protein				Protein
k_{72}	Variable	Protein				Protein
k_{73}	Variable	Protein				Protein
k_{74}	Variable	Protein				Protein
k_{75}	Variable	Protein				Protein
k_{76}	Variable	Protein				Protein
k_{77}	Variable	Protein				Protein
k_{78}	Variable	Protein				Protein
k_{79}	Variable	Protein				Protein
k_{80}	Variable	Protein				Protein
k_{81}	Variable	Protein				Protein
k_{82}	Variable	Protein				Protein
k_{83}	Variable	Protein				Protein
k_{84}	Variable	Protein				Protein
k_{85}	Variable	Protein				Protein
k_{86}	Variable	Protein				Protein
k_{87}	Variable	Protein				Protein
k_{88}	Variable	Protein				Protein
k_{89}	Variable	Protein				Protein
k_{90}	Variable	Protein				Protein
k_{91}	Variable	Protein				Protein
k_{92}	Variable	Protein				Protein
k_{93}	Variable	Protein				Protein
k_{94}	Variable	Protein				Protein
k_{95}	Variable	Protein				Protein
k_{96}	Variable	Protein				Protein
k_{97}	Variable	Protein				Protein
k_{98}	Variable	Protein				Protein
k_{99}	Variable	Protein				Protein
k_{100}	Variable	Protein				Protein

experimental conditions

pH	temp	temperature [°C]	buffer
7.0	37	37	Tris

organization:

Data source:

- Scientific Publications
 - Data found in Text, Tables, Formula, Graphs, Pictures...
- Manual data collection is often difficult and tedious

Full-text publication

SABIO-RK input interface

Database Population: Literature Extraction

Lists of values (LOV)

SABIO-RK
Input Interface

kinetic law		
type	Sequential Bi Bi	
formula	$(E \cdot k_{cat} \cdot A \cdot B) / (K_{ia} \cdot K_b + K_a \cdot B + K_b \cdot A + A \cdot B)$	
variables		
name	term	
parameter		
name	role	type
Kb	Constant	Km
Ka	Constant	Km
kcat	Constant	kcat
Kia	Constant	Kd
kcat_Kb	Constant	kcat/Km
kcat_Ka	Constant	kcat/Km
A	Variable	concentration
B	Variable	concentration
E	Variable	concentration

kinetic law type	unit	unit def.
Product inhibition		
Pseudo first order		
Pseudo first order transition of enzyme into active form		
Random Bi Bi		
Random Bi Bi (modified)		
Random sequential		
Random sequential Bi		
Random Ter Ter		
Random Uni Bi		
Rapid equilibrium ordered		
Rapid equilibrium random		
Rapid equilibrium random Bi Bi		
Rapid equilibrium random Ter Ter		
Reversible Michaelis-Menten	mM	%
Second-order mass action		
Sequential	mM	%
Sequential Bi Bi	1/s	1/s
Sequential Bi Bi with competitive substrate inhibition		
Sequential Bi Bi with substrate inhibition		
Sequential ordered		
Sequential ordered (modified)	1/(M*sec)	1/(M*sec)
Sequential ordered Bi Bi		
Sequential ordered Bi Ter	1/(M*sec)	1/(M*sec)
Sequential ordered Bi Uni		
Sequential ordered forming a dead end complex	mM	%
Sequential ordered multistep		
Sequential ordered Uni Bi	mM	%
Sequential with competitive substrate inhibition		
Steady-state ordered Ping Pong Uni Uni Bi Bi		
Steady-state random		%
Second-order mass action		

choose kinetic law type: add this type

Data Integration Problems: Parameter Units

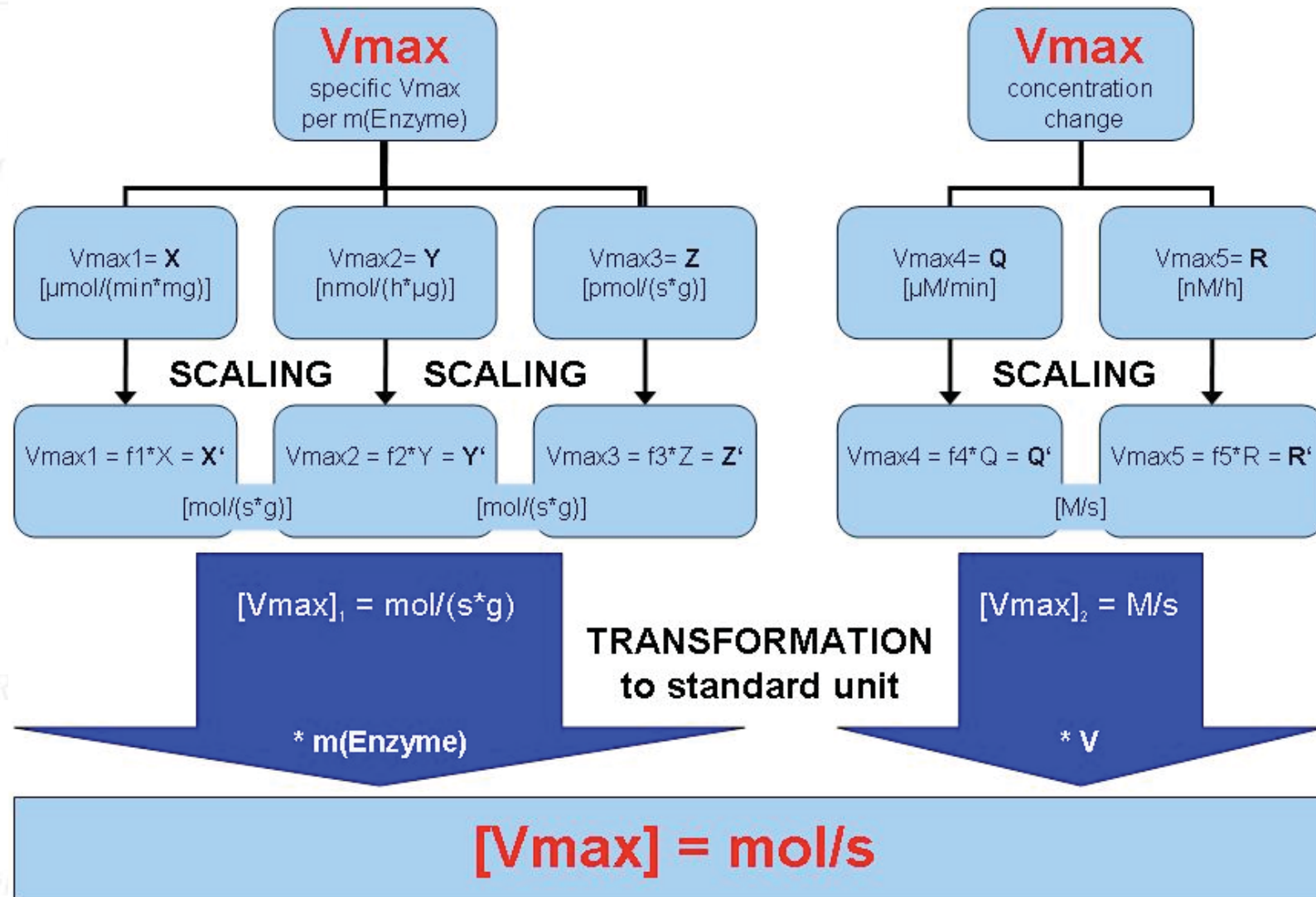
Units from SABIO-RK for kinetic parameter Vmax:

U/g	mM/min	mmol/(min*mg)
mol/(min*µg)	nM/(min*mg)	µmol/(h*mg)
mol/(h*mg)	nmol/min	U/mg
µmol/(min*mol)	U/µmol	pmol/(mg*sec)
mM/s	U/ml	nM/min
mmol/(min*µg)	nmol/(s*mg)	µM/(min*mg)
mmol/(h*mg)	mol/(mol*min)	µmol/(sec*mg)
µM/(sec*mg)	fmol/(pmol*min)	pmol/(min*mg)
µmol/(min*mg)	µM/s	µM/(min*µM)
µmol/(min*g)	katal/mg	pmol/(min*pmol)
U/µg	µmol/min	nmol/(h*mg)
µmol/(min*µmol)	nmol/(min*mg)	pmol/min
µM/min	µM/(h*mg)	fmol/min
mU/ml	nM/sec

Problems:

- Multiple definitions:
e.g. *Katal* or *Unit*
- Different scaling:
 $mM/s = 10^{-3} M / s$
 $\mu M/s = 10^{-6} M / s$
 $nM/min = 10^{-9} M / min \dots$
- Different compositions:
e.g. $\mu mol/s$ or $\mu mol/(s*mg)$
- Wrong parameter unit:
e.g. $1/s \rightarrow$ describes kcat

Normalization of Parameter Units





Controlled Vocabularies and Ontologies

- To unambiguously identify entities or terms in database entries
- Facilitate the search, interpretation and comparison of the data
- Permit a matching with other databases based on a shared vocabulary
- Facilitate the integration of different database entries into kinetic models
- Allows to trace back the provenience of exported data

Some Biomedical Ontologies:

- **ChEBI (Chemical Entities of Biological Interest)**
- **Gene Ontology (GO)**
- **Systems Biology Ontology (SBO)**
- **NCBI taxonomy**



Annotations

Element:
Entity, Process...

Annotation

Qualifier



represents



represents



Biological entity
A

Biological entity
B

Relationship



e.g. Reaction:



isVersionOf

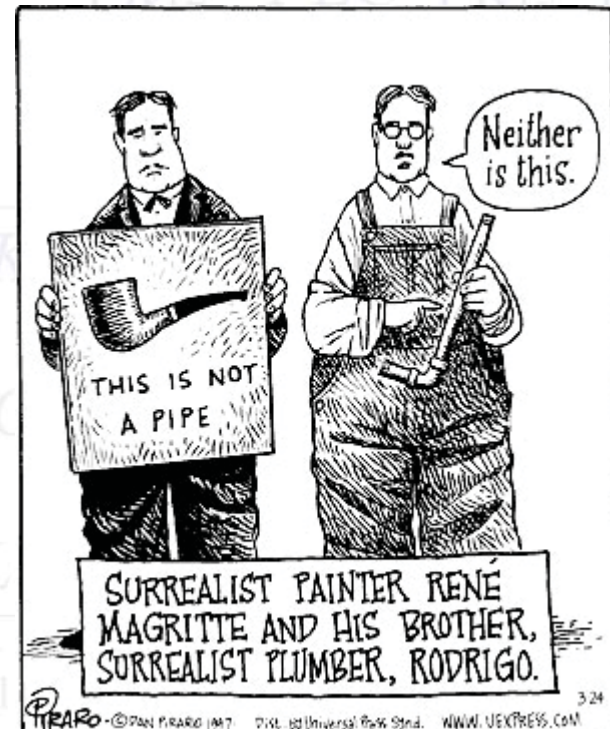


e.g. Enzymatic activity:
EC 1.1.1.1

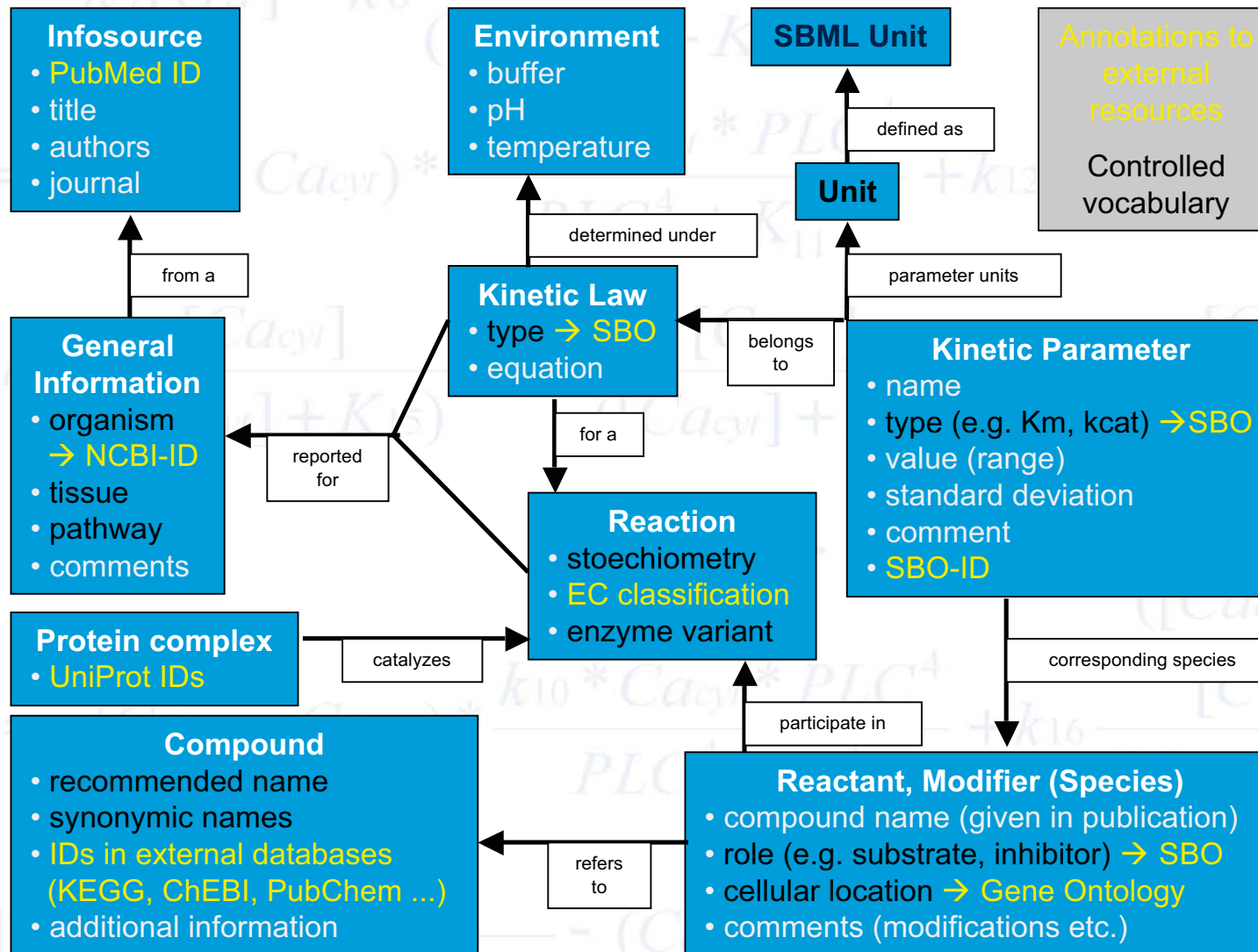


Annotations

It is all about **REPRESENTATION** and **MAPPING**



Annotations and Controlled Vocabulary



Cross-Linking External Databases

Reaction Details

Stoichiometric Equation [dUMP + 5,10-Methylenetetrahydrofolate <-> dTMP + Dihydrofolate](#)

Substrates [dUMP](#)
[5,10-Methylenetetrahydrofolate](#)

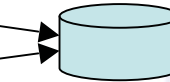
Products [dTMP](#)
[Dihydrofolate](#)

Enzymes known to catalyse this reaction (curated information)

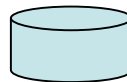
EC Number	in Organism	SWP/UnitProt Link	PubMedID
2.1.1.45	Homo sapiens	SWP/UnitPro	12859954
2.1.1.45	Escherichia coli	SWP/UnitPro	14967040

Pathways [One carbon pool by folate](#)
[Pyrimidine metabolism](#)

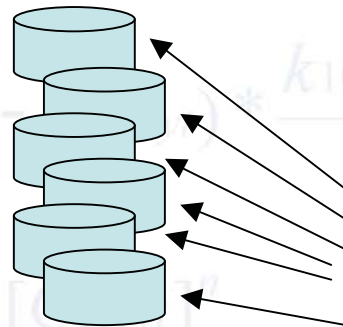
Literature Database



Protein Database



Chemical Databases



Compound Details

Back
List of Reactions

Common Name dUMP
2'-Deoxyuridine 5'-phosphate
Deoxyuridine 5'-phosphate
Deoxyuridine monophosphate
Deoxyuridylic acid

Synonyms

SABIO-Compound-ID 1334

External Links

CAS-ID [964-26-1](#)

Kegg-ID [C00365](#)

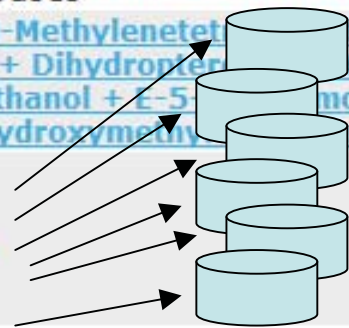
PUBCHEM Substance ID [3656](#)

CHEBIIDs [17622](#)

Cross-Linking Databases: Enzyme

Enzyme Details

Enzyme (recommended) name	Thymidylate synthase
Alternative names	5,10-Methylenetetrahydrofolate:dUMP C-methyltransferase Methylenetetrahydrofolate:dUMP C-methyltransferase Synthase, thymidylate TMP synthetase TS TSASE Thymidylate synthetase dTMP synthase
EC Classification	2.1.1.45
Classification	Transferases Transferring one-carbon groups Methyltransferases
Catalyses reactions	dUMP + 5,10-Methylenetetrahydrofolate <-> dTMP + Dihydrofolate L-Glutamate + Dihydrofolate + ATP <-> ADP + Phosphate + Dihydrofolate 2-Mercaptoethanol + L-5-(methylthio)-2-thiouridine-5'-phosphate <-> 1-(3-hydroxy-4-hydroxymethyl-5-methylthio-2-thiouridyl)-5-[1,4]oxathian-2-yl-1H-pyrimidine-2,4-dione
External links	Expasy KEGG IntEnz (EBI) IUBMB Reactome



**Biochemical
Databases**

$$([Ca_{cyt}]^n + K_{10}^n)$$

$$([Ca_{cyt}] + K_{21})$$

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
5,10-Methylenetetrahydrofolate	unknown	-
dUMP	unknown	-

Products

name	location	comment
Dihydrofolate	unknown	-
dTMP	unknown	-

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	-	Protein: [P04816]?

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](#)



User Interface: Database Entry with Kinetic Data

Compound Details

Back

List of Reactions

Common Name	dUMP 2'-Deoxyuridine 5'-phosphate Deoxyuridine 5'-phosphate
Synonyms	Deoxyuridine monophosphate Deoxyuridylic acid
SABIO-Compound-ID	1334
External Links	
CAS-ID	964-26-1
Kegg-ID	C00365
PUBCHEM Substance ID	3656
CHEBIIDs	17622

Export of Annotated Data in SBML

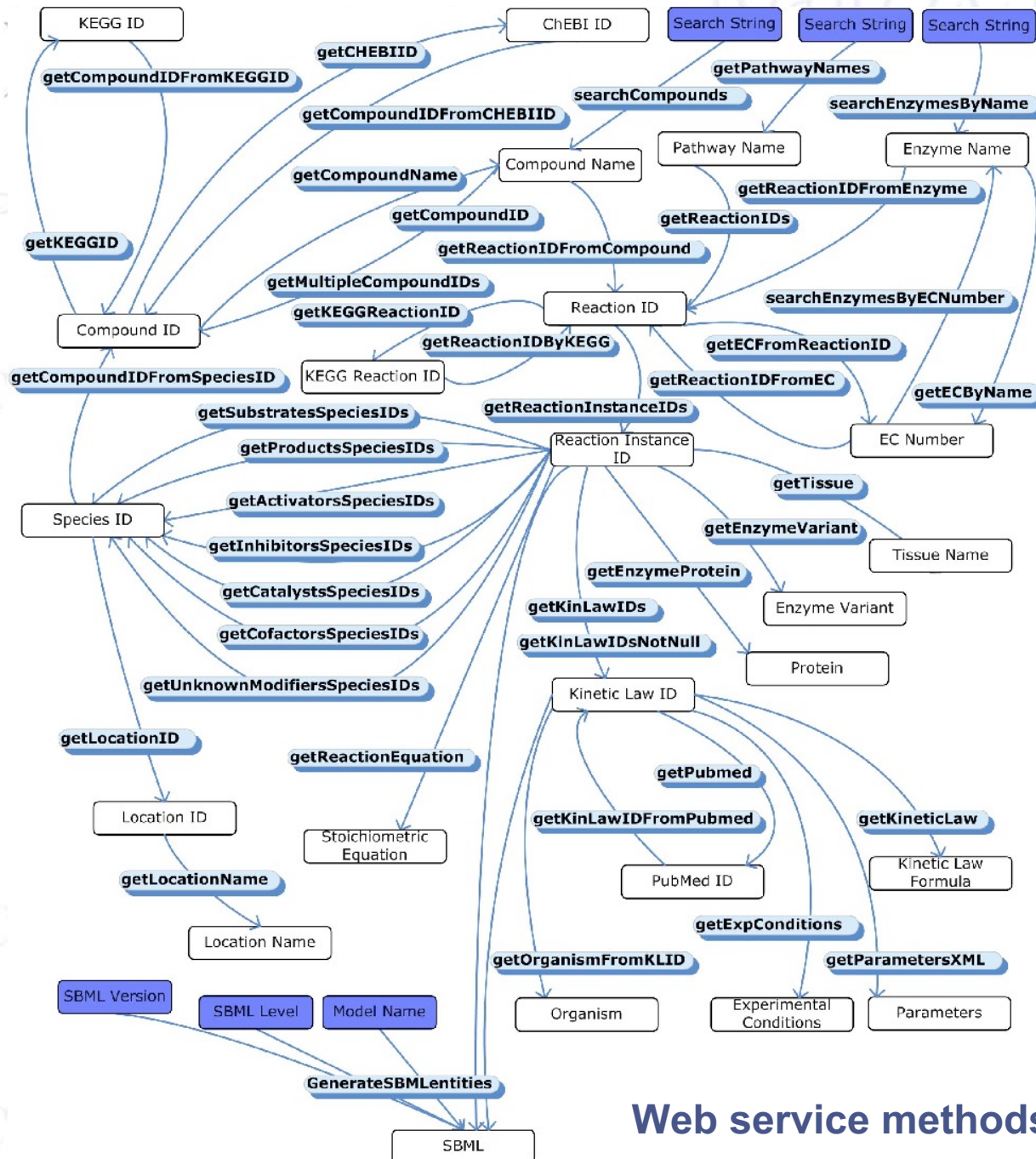
```

    <rdf:li rdf:resource="http://www.genome.jp/kegg/reaction/#R02101" />
  </rdf:Bag>
  <bqbiol:is>
  </rdf:Description>
</rdf:RDF>
</annotation>
<listOfReactants>
  <speciesReference species="spc_1" />
  <speciesReference species="spc_2" />
</listOfReactants>
<listOfProducts>
  <speciesReference species="spc_3" />
  <speciesReference species="spc_4" />
</listOfProducts>
<listOfModifiers>
  <modifierSpeciesReference species="enz_1" />
  <modifierSpeciesReference species="spc_5" />
</listOfModifiers>
<kineticLaw metaid="meta_kl_0">
- <annotation>
- <rdf:RDF xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
- <rdf:Description rdf:about="#meta_kl_0">
- <bqbiol:isDescribedBy>
- <rdf:Bag>
- <rdf:li rdf:resource="http://www.pubmed.gov/#12859954" />
- </rdf:Bag>
- </bqbiol:isDescribedBy>
</rdf:Description>
</rdf:RDF>
</annotation>
- <math xmlns="http://www.w3.org/1998/Math/MathML">
- <apply>
  <ci>k0</ci>
  <ci>kcat</ci>
  <ci>enz1</ci>
  <ci>Km,A_spc_1</ci>
  <ci>spc2</ci>
  <ci>Ki_spc_5</ci>
  <ci>kcatKm,A_spc_1</ci>
  <ci>spc5</ci>
  <ci>spc1</ci>
</apply>
</math>
- <listOfParameters>
  <parameter id="kcat" value="0.213" units="onedivsec" />
  <parameter id="Km,A_spc_1" value="2.7" units="microM" />
  <parameter id="Ki_spc_5" value="4.17" units="microM" />
</listOfParameters>
</species>
<species metaid="meta_spc_2" id="spc_2" name="5,10-Methylenetetrahydrofolate" compartment="compart_1">
- <annotation>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
- <rdf:Description rdf:about="#meta_spc_2">
- <bqbiol:is>
- <rdf:Bag>
- <rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/#CHEBI:15636" />
- <rdf:li rdf:resource="http://www.genome.jp/kegg/compound/#C00143" />
- </rdf:Bag>
- </bqbiol:is>
</rdf:Description>
</rdf:RDF>
</annotation>
</species>
<species metaid="meta_spc_3" id="spc_3" name="Dihydrofolate" compartment="compart_1">
- <annotation>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
- <rdf:Description rdf:about="#meta_spc_3">
- <bqbiol:is>
- <rdf:Bag>
- <rdf:li rdf:resource="http://www.ebi.ac.uk/chebi/#CHEBI:15633" />
- <rdf:li rdf:resource="http://www.genome.jp/kegg/compound/#C00415" />
- </rdf:Bag>
- </bqbiol:is>
</rdf:Description>
</rdf:RDF>
</annotation>
</species>
<listOfUnitDefinitions>
- <unitDefinition id="onedivsec" name="onedivsec">
- <listOfUnits>
  <unit kind="second" exponent="-1" multiplier="1" offset="0" />
</listOfUnits>
</unitDefinition>
- <unitDefinition id="onedivMsec" name="onedivMsec">
- <listOfUnits>
  <unit kind="second" exponent="-1" multiplier="1" offset="0" />
  <unit kind="litre" multiplier="1" offset="0" />
  <unit kind="mole" exponent="-1" multiplier="1" offset="0" />
</listOfUnits>
</unitDefinition>
- <unitDefinition id="microM" name="microM">
- <listOfUnits>
  <unit kind="mole" scale="-6" multiplier="1" offset="0" />
  <unit kind="litre" exponent="-1" multiplier="1" offset="0" />
</listOfUnits>
</unitDefinition>
</listOfUnitDefinitions>

```

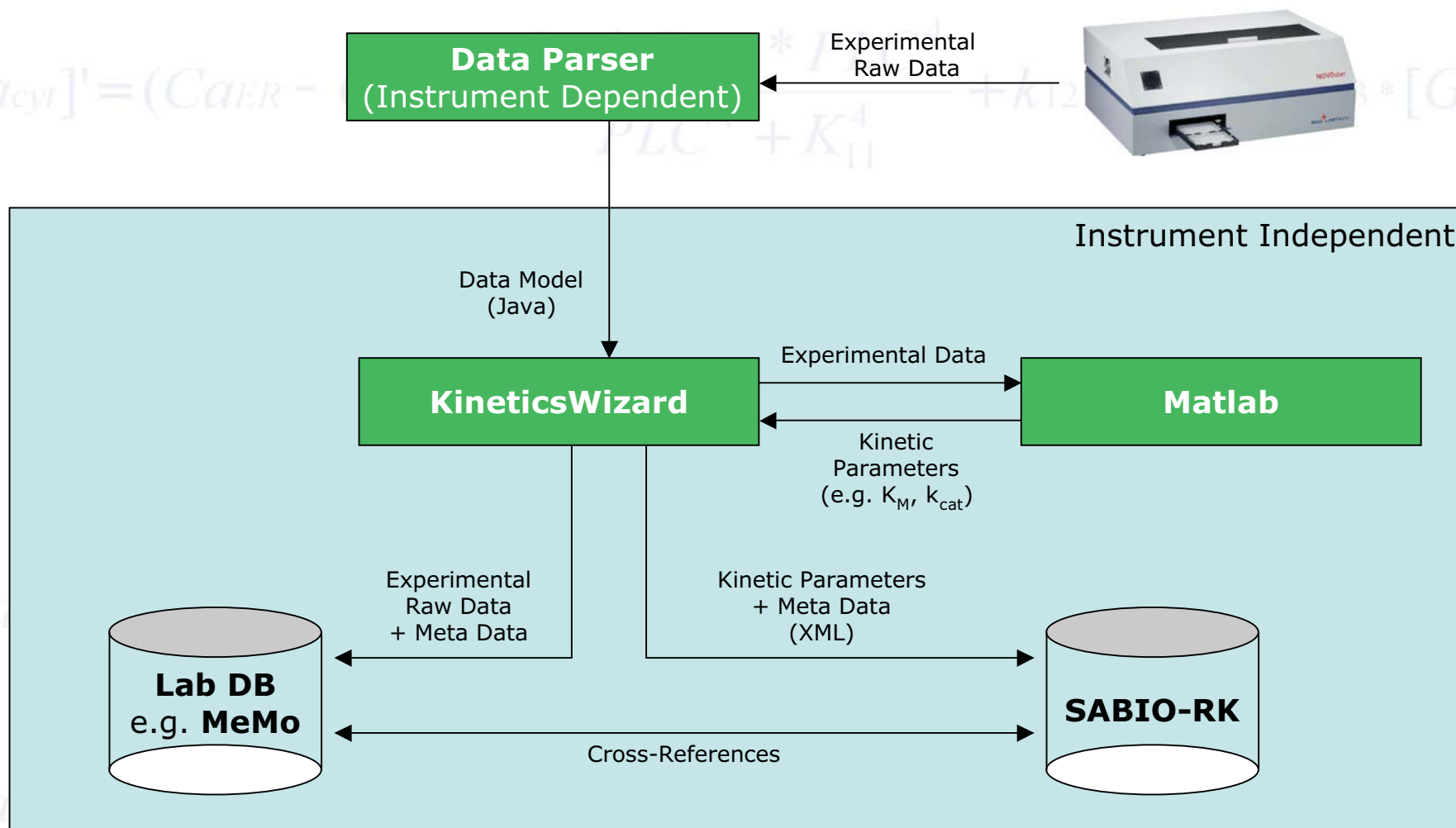
SABIO-RK API Access (Web Services)

- Integration possible in simulation tools
- Cross-linking with other databases
- Several entry points
- Data export in **SBML**
- **ChEBI** already uses API for cross-reference
- **CellDesigner** currently is implementing access



Web service methods

Data Storage System for Kinetic Data



Conclusions

- SABIO-RK is a web-accessible database that contains interrelated data describing biochemical reaction kinetics embedded in the environmental conditions' context
- Access through a web-based user interface or through web services (API)
- Manual curation of the data with some semi-automatic support
- Controlled vocabularies are applied and developed
- Content is annotated to ontologies and external resources (also exported in SBML)
- Tools for normalization of the data help to integrate heterogeneous information
- Direct import of experimental data from the instrumentation to SABIO-RK will speed-up database population (cooperation with the University of Manchester)
- We are working on the graphical display of pathway and reaction data in SABIO-RK



Acknowledgements

SDBV group:

Henriette Engelken

Renate Kania

Olga Krebs

Saqib Mir

Andreas Weidemann

Ulrike Wittig

Isabel Rojas

Sylvestre Kegne (student)

and many more: students, colleagues at EML Research and other collaborators...

Financial support:



External Collaborators:

SBO/MIRIAM: Nicolas Le Novère & Camille Laibe

CellDesigner: Hiroaki Kitano & Akira Funahashi
& Akiya Jouraku

ChEBI: Kirill Degtyarenko & Paula de Matos &
Janna Hastings & Christoph Steinbeck

MCISB (University of Manchester):

Douglas Kell & Neil Swainston

SABIO-RK is member of:

