

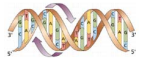
Kapitel 6: Pathway-Datenbanken

- n Motivation und biologische Grundlagen
 - Metabolische Pfade
 - Regulatorische Pfade
- n Pathway-Datenbanken und ihre Anwendung
 - Anforderungen
 - KEGG
 - EcoCyc/BioCyc



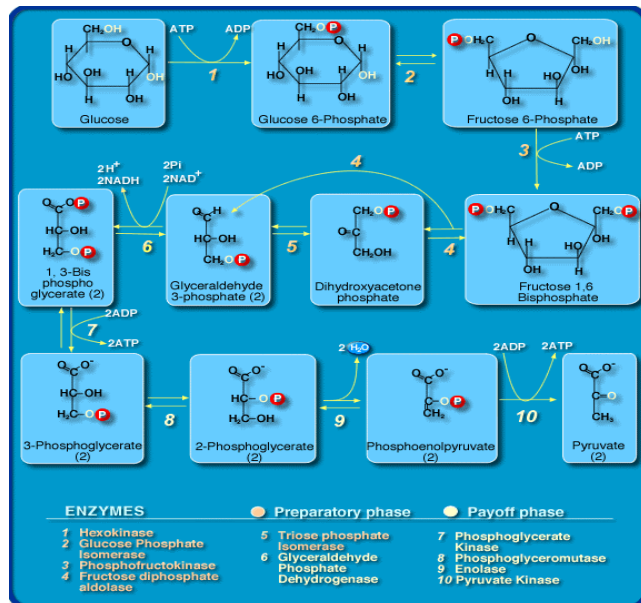
Pathway – biochemische Reaktionswege

- n Pathway: Folge von biochemischen Reaktionen, meist einer oder mehreren Funktion(en) im Organismus zugeordnet
- n Grobeinteilung der Pathways in
 - Stoffwechselwege (metabolic pathways)
 - Regulatorische Pfade (regulatory pathways)
- n Metabolische Pfade
 - Gesamtheit aller (lebens)notwendigen biochemischen Vorgänge beim Aufbau, Abbau und Umbau eines Organismus sowie dessen Stoffaustausch mit der Umwelt
 - 2 grundlegende Stoffwechselvorgänge: 1. Assimilation/Anabolismus (Photosynthese, Chemosynthese, Verdauungsprozesse); 2. Dissimilation/Katabolismus (Atmung, Gärung)
- n Regulatorische Pfade
 - Regulation der Genexpression (genetic networks, genetic-regulatory pathways)
 - Signalwege (signalling pathways, signal-transduction cascades)



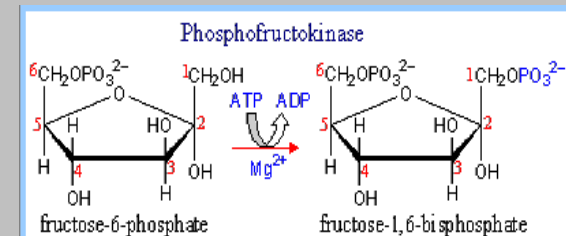
Metabolische Pfade: Bsp. Glykolyse

- n Zuckerabbau; wichtiger energieliefernder Prozess



Enzymatische Reaktion

Schritt 3 der Glykolyse – Phosphofruktokinase katalysiert Phosphorylierung von Fructose-6-phosphat:



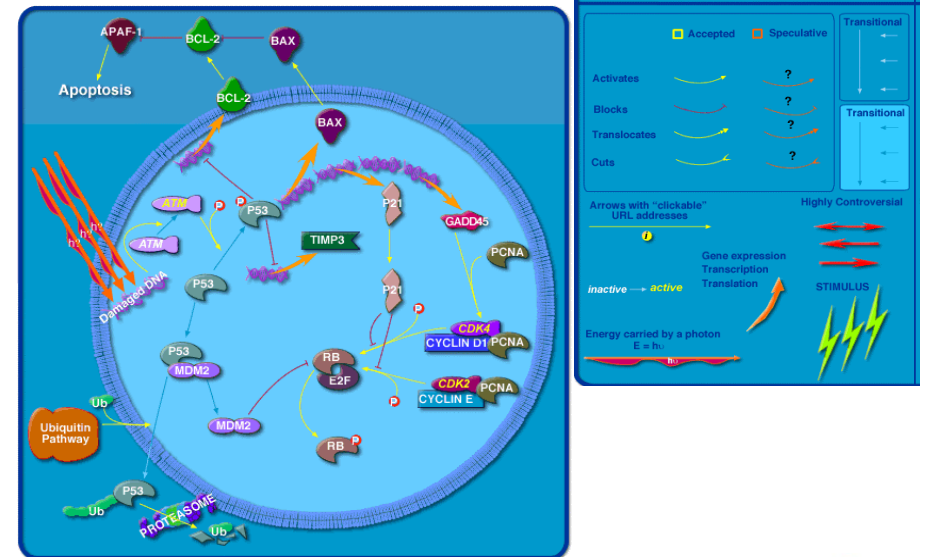
Hierarchie der Enzyme: E.C. - Code

- n Standardisierter Code zur Klassifizierung der Enzyme (IUBMB Enzyme Nomenclature Committee)
- n Codierung: Oxidoreduktasen (1), Transferasen (2), Hydrolasen (3), Lyasen (4), Isomerasen (5), Ligasen (6).
- n Bsp.: Phosphofruktokinase, E.C. 2.7.1.11, gehört zur Enzymklasse der Transferasen (2)

E.C. Code	Bezeichnung
2	Enzym der funktionellen Gruppe Übertragen
27	Enzym der funktionellen Gruppe Phosphotransferase
271	Enzym der funktionellen Gruppe Phosphotransferase, Acceptor
27111	Phosphofruktokinase



Regulatorische Pfade: Bsp. p53-Signalweg



Pathway-Datenbanken

- n Inhalt einer Pathway-Datenbank
 - Strukturierte Informationen über biochemische Reaktionswege, Einzelreaktionen, beteiligte Enzyme und Substrate
- n Motivation
 - "Lexikon der Biochemie" in Praxis, Lehre und Forschung
 - Identifizierung potentieller Angriffspunkte für Arzneimittel (drug targets)
 - Vorhersage von Stoffwechselwegen sequenzierter Organismen (Beispiel BioCyc)
- n Spezifische Anforderungen
 - Rechnerverwertbare Darstellung/Modellierung der Pfade und Netzwerke (erweiterte Datenmodelle; z.B. graphbasiert)
 - Algorithmen zur Pathway-Analyse
 - Methoden zur Visualisierung
 - Standards für den Datenaustausch



Pathway-Datenbanken: Übersicht

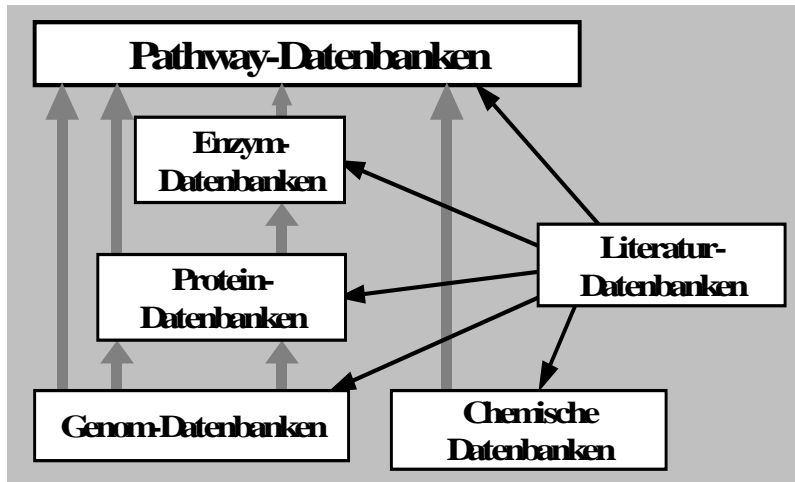
Metabolic Pathways and Cellular Regulation

EcoCyc	http://ecocyc.org/	<i>Escherichia coli</i> K-12 genome, metabolic pathways, transporters and gene regulation
ENZYME	http://www.expasy.ch/enzyme/	Enzyme nomenclature
EpoDB	http://www.cbil.upenn.edu/EpoDB/	Genes expressed during human erythropoiesis
Klotho	http://www.ibc.wustl.edu/klotho/	Collection and categorization of biological compounds
KEGG	http://www.genome.ad.jp/kegg	Metabolic and regulatory pathways
LIGAND	http://www.genome.ad.jp/ligand/	Chemical compounds and reactions in biological pathways
MetaCyc	http://ecocyc.org/	Metabolic pathways and enzymes from various organisms
UMBBD	http://umbdd.ahc.umn.edu/	Curated information on microbial catabolism and related biotransformations
PathDB	http://www.ncgr.org/pathdb	Biochemical pathways, compounds and metabolism
PRODORIC	http://prodoric.tu-bs.de	Prokaryotic database of gene regulation and regulatory networks
RegulonDB	http://www.cifn.unam.mx/Computational_Genomics/regulondb/	<i>Escherichia coli</i> transcriptional regulation and operon organization
UM-BBD	http://umbdd.ahc.umn.edu/	Microbial biocatalytic reactions and biodegradation pathways
WIT2	http://wit.mcs.anl.gov/WIT2/	Integrated system for metabolic models

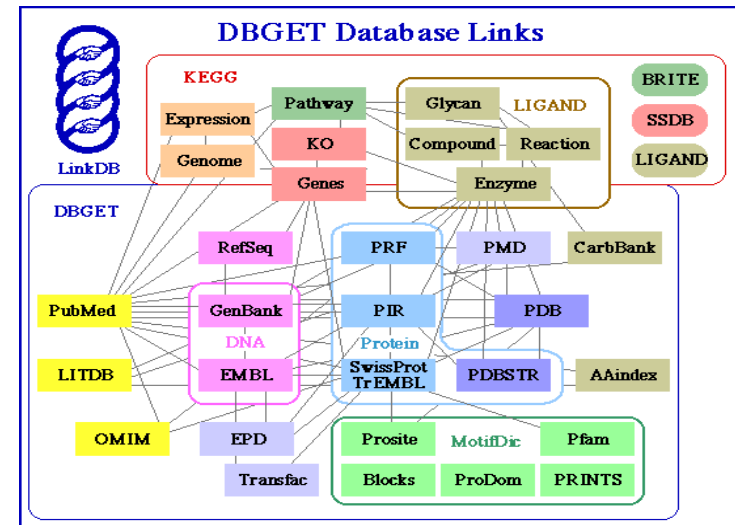
aus: The Molecular Biology Database Collection: 2003 update, Nucleic Acids Research, 2003, Vol. 31, No. 1 1-12



Zusammenhang mit anderen biologischen Datenbanken



Beispiel KEGG*



KEGG

- n Kyoto Encyclopedia of Genes and Genomes*
- n Umfangreiche Datenbanksammlung des japanischen GenomeNet-Service zu Pathway-Informationen
- n KEGG/PATHWAY-DB
 - enthält alle bekannten Stoffwechselwege
 - Auswahl regulatorischer Pathways
- n Methoden und Tools für Berechnungen und Analysen



KEGG: Pathway-Hierarchie

Metabolism

Carbohydrate Metabolism
 Energy Metabolism
 Lipid Metabolism
 Nucleotide Metabolism
 Amino Acid Metabolism
 Metabolism of Other Amino Acids
 Metabolism of Complex Carbohydrates
 Metabolism of Complex Lipids
 Metabolism of Cofactors and Vitamins
 Biosynthesis of Secondary Metabolites
 Biodegradation of Xenobiotics

Human Diseases

Neurodegenerative Disorders

Genetic Information Processing

Transcription
 Translation
 Sorting and Degradation
 Replication and Repair

Environmental Information Processing

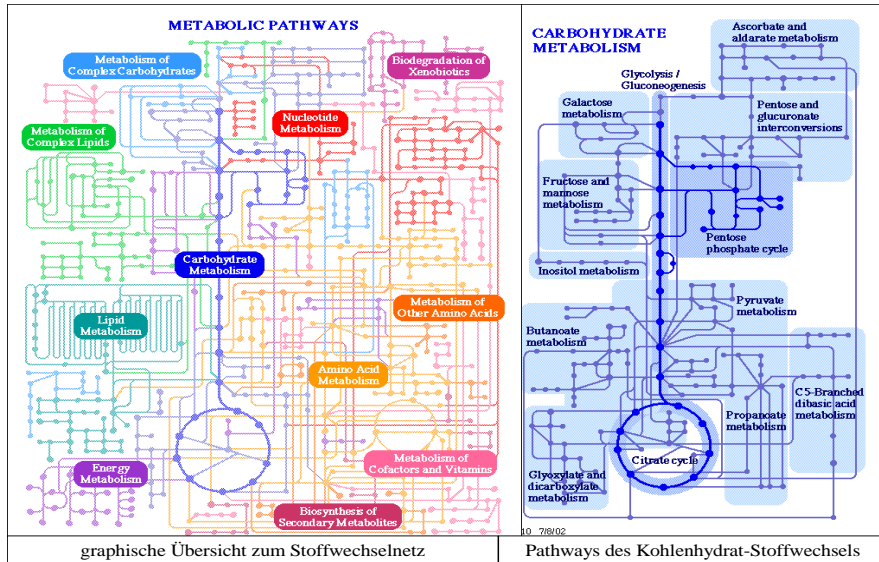
Membrane Transport
 Signal Transduction
 Ligand-Receptor Interaction

Cellular Processes

Cell Motility
 Cell Growth and Death
 Cell Communication
 Development
 Behavior



KEGG: Metabolische Übersichtskarten



KEGG: Einstiegsseite

Adresse <http://www.genome.ad.jp/kegg/kegg2.html>

KEGG - Table of Contents

[[PATHWAY](#) | [GENES](#) | [SSDB](#) | [LIGAND](#) | [EXPRESSION](#) | [BRITE](#) | [BLAST/FASTA](#) | [DBGET](#)]

Search for

1. Pathway Information

1-1. Pathway Maps and Ortholog Tables -- Update Status

Category	Pathway Map Ortholog Table	Search & Compute	DBGET Search
Pathway	Metabolic pathways	Search objects in pathway maps Color objects in pathway maps Search or color genes in ortholog tables Search similar sequences in pathway maps Search similar sequences in ortholog tables Generate possible reaction pathways	PATHWAY
	Regulatory pathways		

1-2. Hierarchical Classifications



KEGG: Suchbeispiel 1

Search against:

Enter a set of gene name(s) with the colors below:

Alternatively, enter the file name containing the data:

Color for the reference in map

Default color for the gene(s) in map

Display gene name(s) NOT found in the search

Adresse http://www.genome.ad.jp/kegg-bin/search_pathway_multi_www

Pathway Search Result

- [hsa00010 Glycolysis / Gluconeogenesis](#)
EC 2.7.1.11 6-phosphofruktokinase; phosphohexokinase; phosphofruktokinase I;
- [hsa00030 Pentose phosphate pathway](#)
EC 2.7.1.11 6-phosphofruktokinase; phosphohexokinase; phosphofruktokinase I;
- [hsa00051 Fructose and mannose metabolism](#)
EC 2.7.1.11 6-phosphofruktokinase; phosphohexokinase; phosphofruktokinase I;
- [hsa00052 Galactose metabolism](#)
EC 2.7.1.11 6-phosphofruktokinase; phosphohexokinase; phosphofruktokinase I;



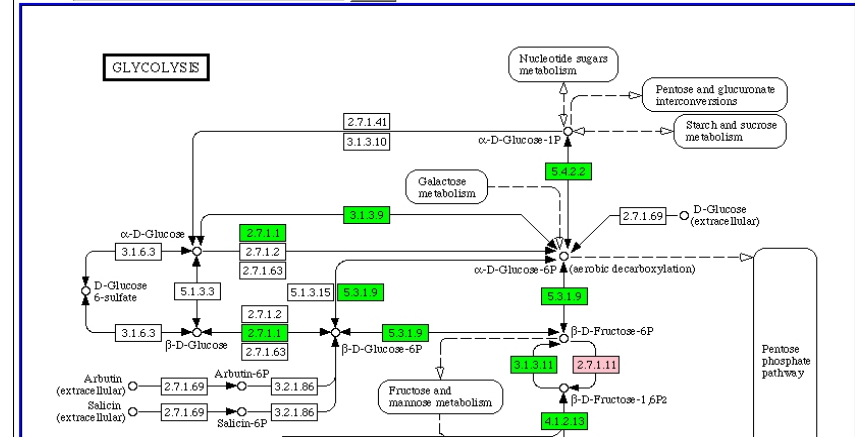
KEGG: Suchbeispiel 1 (2)

Adresse http://www.genome.ad.jp/kegg-bin/mark_pathway_www?opt/local/WWW/pub/tmp/search_pathway2586/hsa00010.args

Glycolysis / Gluconeogenesis - Homo sapiens


Go to: [LinkDB search](#) | [Ortholog Table](#) |

Go to:



KEGG: Suchbeispiel 2

Adresse http://www.genome.ad.jp/kegg-bin/check_cpd

 **Generate Possible Pathways between Two Compounds**

Exec Clear

Search against:

Select initial substrate: (searched for "glucose")

Select final product: (searched for "pyruvate")

Compound IDs may be found by [searching LIGAND database using DBGET](#)

Enter cut off length:

Select hierarchy for relaxation: with level

Select sort option: by path length by compound ID

Exec Clear

(C) Prof. R. Müller, Prof. E. Rahm

6 - 17



KEGG: Suchbeispiel 2 (2)

Adresse http://www.genome.ad.jp/kegg-bin/check_cpd

Result of Pathway Computation

Organism : all
 Initial substrate : C00293 Glucose
 Final product : C00022 Pyruvate
 Cutoff length : 5
 Relaxation : No relaxation
 Number of Results : 30

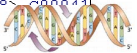
[[Show as Diagram](#)]

```

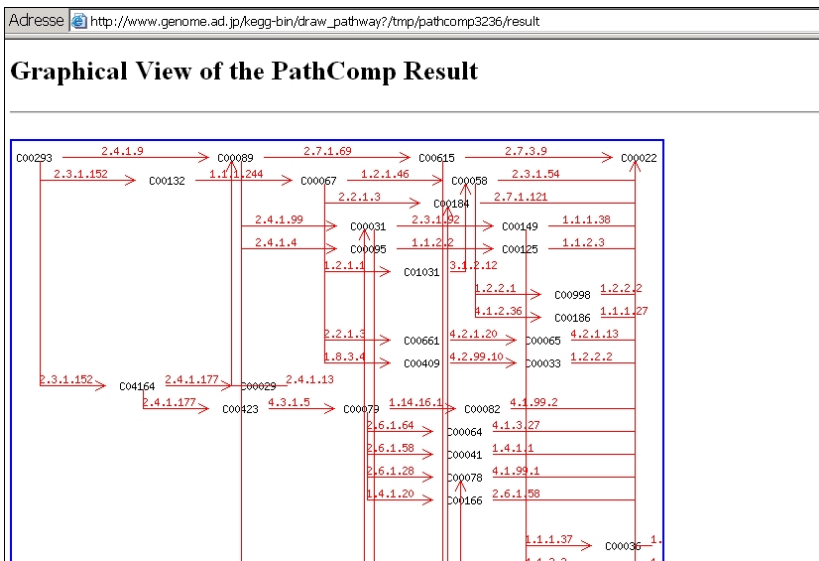
3 C00293 <2.4.1.9> C00089 <2.7.1.69> C00615 <2.7.3.9> C00022 [Known pathways]
4 C00293 <2.3.1.152> C00132 <1.1.1.244> C00067 <1.2.1.46> C00058 <2.3.1.54> C0002
4 C00293 <2.3.1.152> C00132 <1.1.1.244> C00067 <2.2.1.3> C00184 <2.7.1.121> C0002
4 C00293 <2.4.1.9> C00089 <2.4.1.99> C00031 <2.3.1.92> C00149 <1.1.1.38> C00022
4 C00293 <2.4.1.9> C00089 <2.4.1.4> C00095 <1.1.2.2> C00125 <1.1.2.3> C00022 [K
5 C00293 <2.3.1.152> C00132 <1.1.1.244> C00067 <1.2.1.1> C01031 <3.1.2.12> C00058
5 C00293 <2.3.1.152> C00132 <1.1.1.244> C00067 <1.2.1.46> C00058 <1.2.2.1> C00998
5 C00293 <2.3.1.152> C00132 <1.1.1.244> C00067 <1.2.1.46> C00058 <4.1.2.36> C0018
5 C00293 <2.3.1.152> C00132 <1.1.1.244> C00067 <2.2.1.3> C00661 <4.2.1.20> C00065
5 C00293 <2.3.1.152> C00132 <1.1.1.244> C00067 <1.8.3.4> C00409 <4.2.99.10> C0003
5 C00293 <2.3.1.152> C04164 <2.4.1.177> C00029 <2.4.1.13> C00089 <2.7.1.69> C0061
5 C00293 <2.3.1.152> C04164 <2.4.1.177> C00423 <4.3.1.5> C00079 <1.14.16.1> C0008
5 C00293 <2.3.1.152> C04164 <2.4.1.177> C00423 <4.3.1.5> C00079 <2.6.1.64> C00064
6 C00293 <2.3.1.152> C04164 <2.4.1.177> C00423 <4.3.1.5> C00079 <2.6.1.64> C00064
    
```

(C) Prof. R. Müller, Prof. E. Rahm

6 - 18

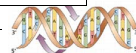


KEGG: Suchbeispiel 2 (3)



(C) Prof. R. Müller, Prof. E. Rahm

6 - 19



KEGG: XML-Darstellung

n KGML (KEGG Markup Language)

```

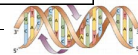
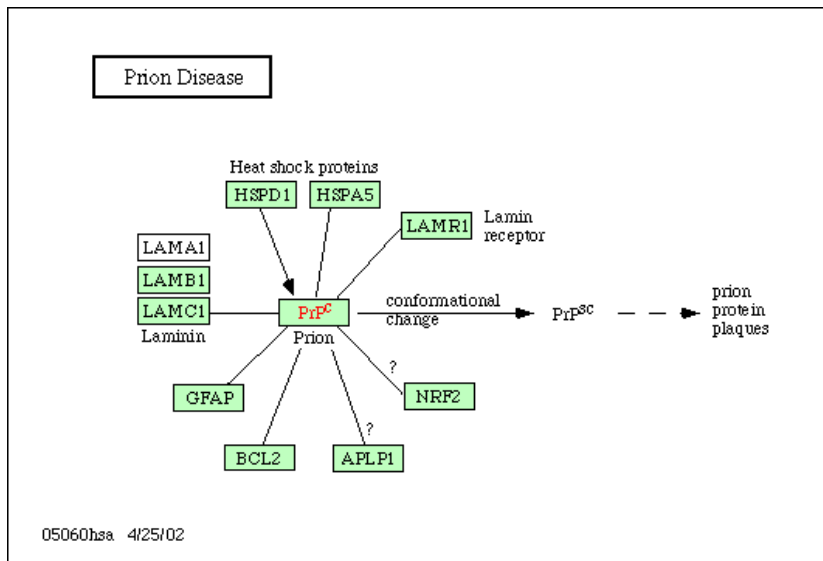
<?xml version="1.0" ?>
<!DOCTYPE pathway (View Source for full doctype...)>
- <pathway name="path:map00010" number="00010" org="map" title="Glycolysis /
  Gluconeogenesis"
  image="http://www.genome.ad.jp/kegg/pathway/map/map00010.gif"
  link="http://www.genome.ad.jp/dbget-bin/show_pathway?map00010">
- <entry name="ec:1.2.1.3" type="enzyme" id="1" reaction="rn:R00710">
  link="/dbget-bin/www_bget?enzyme+1.2.1.3">
  <graphics name="1.2.1.3" x="171" y="1019" type="rectangle" width="45"
  height="17" fgcolor="#000000" bgcolor="#FFFFFF" />
</entry>
- <entry name="ec:6.2.1.1" type="enzyme" id="2" reaction="rn:R00235">
  link="/dbget-bin/www_bget?enzyme+6.2.1.1">
  <graphics name="6.2.1.1" x="102" y="917" type="rectangle" width="46"
  height="17" fgcolor="#000000" bgcolor="#FFFFFF" />
</entry>
- <entry name="ec:1.2.1.5" type="enzyme" id="3" reaction="rn:R00711">
  link="/dbget-bin/www_bget?enzyme+1.2.1.5">
  <graphics name="1.2.1.5" x="171" y="1040" type="rectangle" width="45"
  height="17" fgcolor="#000000" bgcolor="#FFFFFF" />
</entry>
- <entry name="cpd:C00033" type="compound" id="4">
  link="http://www.genome.ad.jp/dbget-bin/www_bget?
  compound+C00033">
  <graphics name="C00033" x="102" y="971" type="circle" width="8" height="8">
  fgcolor="#000000" bgcolor="#FFFFFF" />
</entry>
    
```

(C) Prof. R. Müller, Prof. E. Rahm

6 - 20



KEGG: Beispiel für pathogenen Pfad



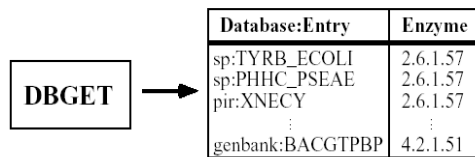
KEGG: Interne Darstellung

n Graphorientiertes Modell

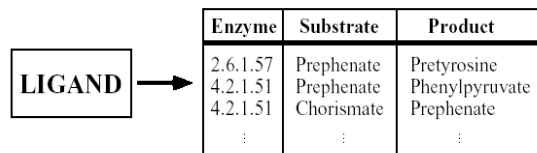
Graph	Node	Edge	Database
Gene universe	Gene		GENES
	Protein-coding gene	Sequence similarity (orthology, paralogy, etc.)	SSDB
	Gene	Adjacency	GENES, GENOME
	Gene	Expression similarity	EXPRESSION
	Gene or gene product	Interaction or relation	BRITE
Protein network	Gene product or subnetwork	Direct protein-protein interaction Gene expression relation Enzyme-enzyme relation	PATHWAY
Chemical universe	Chemical compound		COMPOUND
	Chemical compound	Chemical reaction	REACTION



KEGG: Interne Darstellung (2)



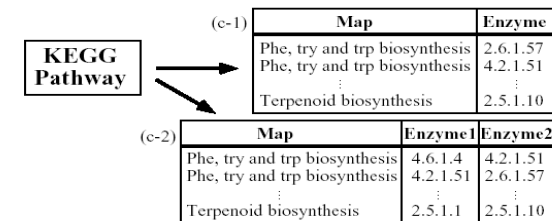
(a) Relation between a database entry and an enzyme. This relation is extracted from LinkDB⁶ in the DBGET system.⁴



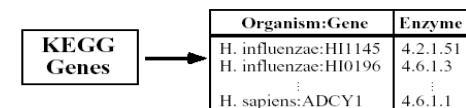
(b) Binary relation between chemical compounds (substrate and product) that participate in a chemical reaction. This relation is extracted from LIGAND Chemical Database for Enzyme Reaction.⁵



KEGG: Interne Darstellung (3)



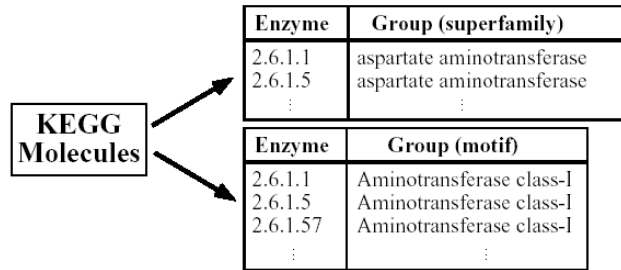
(c) Relationship between an enzyme and its location (map) in the metabolic pathways (c-1). Binary relation between two enzymes that appear consecutively in the pathway (c-2). These relations are extracted from the pathway diagrams of KEGG.



(d) Relation between a gene and an enzyme. It is extracted from the KEGG hierarchical text data in the genes section.



KEGG: Interne Darstellung (4)



(e) Relation between an enzyme and the group to which it belongs relation is derived from the KEGG hierarchical text data in the m section. Currently, there are four classifications based on EC number superfamilies,⁸ SCOP 3D-folds⁹ and PROSITE motifs.³



KEGG: Pfad-Deduktion

n Verwendung der deduktiven Datenbank CORAL*

$$reaction(E, X, Y).$$

When the conversion of compound X to compound Y is a multistep process consisting of a number of enzymes, the enzymatic pathway is represented by:

$$path(X, Y, [E]) \leftarrow reaction(E, X, Y).$$

$$path(X, Y, [E | EL]) \leftarrow reaction(E, X, Z), path(Z, Y, EL).$$

n Probleme

- Aufsplitterung der Abhängigkeitsgraphen in Relationen (vermehrte Notwendigkeit von Joins)
- Zyklenerkennung

* <http://www.cs.wisc.edu/coral/>



EcoCyc

n Encyclopedia of *Escherichia coli* Genes and Metabolism

n Pathway/Genom-Datenbank (PGDB)

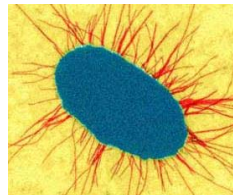
- Enthält das vollständig sequenzierte Genom und die metabolischen sowie regulatorischen Pathways des *E. coli*-Bakteriums

n Rechnergestützte Analysemethoden

n Teil der BioCyc* Knowledge Library (Datenbanksammlung zu Mikroorganismen) mit ca. 15 Datenbanken

n Datenpflege

- 5 hauptamtliche Curators
- Hauptinformationsquellen: Bio-Literatur, Generierung von Pfaden auf der Basis von Gensequenzen
- Korrektur ggf. durch Forschungsgruppen
- Vierteljährliche Releases



* <http://BioCyc.org/>



BioCyc: Datenbanken

Literature-derived Pathway/Genome Databases

- [EcoCyc](#) -- *Escherichia coli* K12
- [MetaCyc](#) -- Metabolic pathways and enzymes from 150 species

Computationally-derived Pathway/Genome Databases

- [AgroCyc](#) -- *Agrobacterium tumefaciens*
- [BsubCyc](#) -- *Bacillus subtilis*
- [CtraCyc](#) -- *Chlamydia trachomatis*
- [CauloCyc](#) -- *Caulobacter crescentus*
- [EcoO157Cyc](#) -- *Escherichia coli* O157:H7 **NEW**
- [HinCyc](#) -- *Haemophilus influenzae*
- [HpyCyc](#) -- *Helicobacter pylori*
- [HumanCyc](#) -- *Homo sapiens*
- [MtbCdcCyc](#) -- *Mycobacterium tuberculosis* CDC1551
- [MtbRvCyc](#) -- *Mycobacterium tuberculosis* H37Rv
- [MpneuCyc](#) -- *Mycoplasma pneumoniae*
- [PseudoCyc](#) -- *Pseudomonas aeruginosa*
- [ShigellaCyc](#) -- *Shigella flexneri* **NEW**
- [TpalCyc](#) -- *Treponema pallidum*
- [VchoCyc](#) -- *Vibrio cholerae*



EcoCyc: Datenmodell

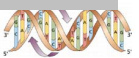
- n Ca. 1000 Klassen, die Schlüsselkonzepte der Biochemie und Molekularbiologie codieren, z.B. 'Pathways', 'Compounds', 'Genes'
- n Ca. 200 Slots (Attribute), die Eigenschaften und Relationen zwischen den Klassen beschreiben, z.B. Slot 'REACTION-LIST' eines 'Pathway'-Objektes
- n Frame-Wissensrepräsentationssystem (via Lisp)
 - ähnlich einem OODBS
 - Web of frames': Frame = ein einzelnes biologisches Objekt (z.B. Gen, Protein), beschrieben durch seine Eigenschaften (Slots)
 - Vernetzung zwischen den Frames durch Slots, die semantische Relationen beschreiben, z.B. Protein x = Genprodukt von Gen y



EcoCyc: Pathway-Taxonomie

- n Vererbungshierarchie:

- [Pathways](#)
 - [Energy metabolism](#)
 - [Super-Pathways](#)
 - [Amino acid families](#)
 - [Biosynthesis](#)
 - [Carbohydrates](#)
 - [Cell-structures](#)
 - [Murein sacculus](#)
 - [Surface structures](#)
 - [Fatty acids and lipids](#)
 - [Cofactors, prosthetic groups, electron carriers](#)
 - [Nucleotides](#)
 - [Ribonucleotides](#)
 - [purines and pyrimidines](#)
 - [2'-deoxyribonucleotides](#)
 - [Polyamines](#)
 - [Amino acid biosynthesis](#)
 - [Amino acid families](#)
 - [Individual amino acids](#)



EcoCyc: Lisp-Frame

```

(define-instance FUMHYDR_RXN
  :template <REACTION>
  :slots
  (COMMON-NAME "fumarate hydration"
    :max-cardinality 1 :value-class STRING)
  (SYNONYMS "malate dehydration"
    :value-class STRING)
  (EC-NUMBER "4.2.1.2" :max-cardinality 1
    :value-class STRING)
  (CATALYZED-BY FUMARA_ENZRKN
    :value-class ENZYMATI-C-REACTION)
  (LEFT (fumarate H2O) :value-class CHEMICAL)
  (LEFT-COEFFICIENTS (1 1) :value-class integer)
  (RIGHT (malate) :value-class CHEMICAL)
  (RIGHT-COEFFICIENTS (1) :value-class integer)
  (DELTA G0 :max-cardinality 1
    :value-class NUMBER)
  (COMMENT)
  (CITATIONS)
  )
    
```

Figure 2: An instance frame of types **reaction**. This definition gives the name of the frame, its template (defining) classes, and the slots it contains. The definitions of individual slots list the slot names, their values (if present), and constraints on the allowable values of the slots, such as the maximum number of values, and the type (value-class) of individual values.



EcoCyc: Datenmengen

EcoCyc KB Statistics by Year

	2003	2002	2001	2000	1999	Description
Pathways	176	164	165	165	159	Number of metabolic plus signaling pathways. Excludes super-pathways.
Reactions	3177	2862	2604	2115	946	Number of reactions -- includes metabolic reactions, transport reactions, reactions involving binding of transcription factors to their binding sites.
Enzymes	992	918	905	884	629	Number of enzymes that catalyze biochemical reactions.
Transporters	169	168	162	158	13	Number of transporters.
Protein comments	1929	1030	921	846	486	Number of proteins that contain comments.
Genes	4477	4393	4393	4393	4390	Number of genes, including some that have not been pinned to the DNA sequence.
Transcription Units	828	724	629	NA	NA	Number of transcription units -- includes operons and single-gene transcription-units.
Citations	6223	3701	3508	3208	1944	Number of distinct references cited in EcoCyc.

Note: The statistics for each year pertain to the last EcoCyc version released in that year.



EcoCyc: Beispiel-Suche

Adresse <http://biocyc.org:1555/server.html>

BioCyc Query Page

Upcoming Pathway Tools User Group Meetings at PSB and ASM/TIGR Micro
For more information, contact ptools-support@ai.sri.com

This form provides several different mechanisms for querying Pathway/Genome Databases.

Select a dataset:

Links to summary information about the selected organism:

- [Summary page for dataset](#)
- **Metabolic Overview** [Diagram](#)/[Expression Viewer](#) (not available for MetaCyc)
- [History of updates to this dataset](#)
- [PathoLogic Pathway Analysis](#) (not available for *E. coli* or MetaCyc)

• [Choose from a list of pathways](#)

• **Query by name or EC number:**

To retrieve objects by name, first select the type of object you wish to retrieve, then enter the name of

(C) Prof. R. Müller, Prof. E. Rahm

6 - 33



EcoCyc: Beispiel-Suche (2)

Adresse <http://biocyc.org:1555/ECOLI/substring-search?type=ENZYME&object=2.7.1.11>

Query Results

The query "2.7.1.11" matched 2 enzymes:

[6-phosphofructokinase-1](#)
[tagatose-6-phosphate kinase / 6-phosphofructokinase-2](#)

This page is Copyright © SRI International 1996,1999-2002, Marine Biological Laboratory 1996-2001, DoubleTwist Inc. 1996-2002. Please cite EcoCyc as [Nucleic Acids Res. 30\(1\):59 2002](#) in publications resulting from its use. SRI International Pathway Tools software, page generated on Tue Jan 14, 2003.


(C) Prof. R. Müller, Prof. E. Rahm

6 - 34



EcoCyc: Beispiel-Suche (3)

Adresse <http://biocyc.org:1555/ECOLI/NEW-IMAGE?type=ENZYME&object=6PFK-1-CPX>



E. coli Enzyme: 6-phosphofructokinase-1

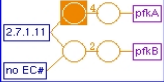
Superclasses: [Protein-Complexes](#)

Component composition: [6-phosphofructokinase-1 monomer x 4](#)

Citations: [[1](#), [2](#), [3](#), [4](#)]

Comment: This enzyme is an isozyme with phosphofructokinase-2. The nucleotide sequences of the tetrameric species is the only one which can bind both substrates and effectors, and thus has unique properties. The C terminal end of the peptide is required for allosteric properties. [[3](#)] Crystals with and without activators and inhibitors. [[1](#), [2](#)]

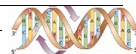
Gene-Reaction Schematic:



Enzymatic reaction of: 6-phosphofructokinase-1


(C) Prof. R. Müller, Prof. E. Rahm

6 - 35

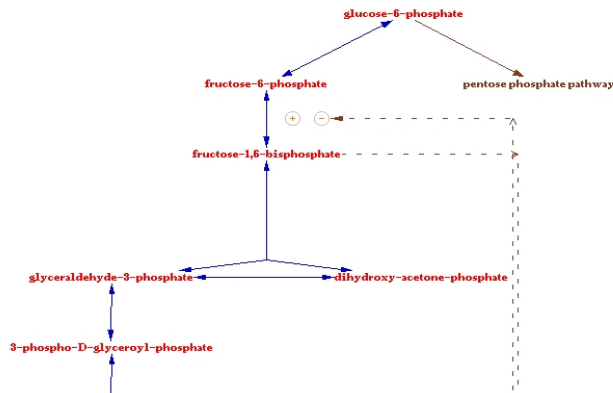


EcoCyc: Beispiel-Suche (4)

Adresse <http://biocyc.org:1555/ECOLI/new-image?type=PATHWAY&object=GLYCOLYSIS>



E. coli Pathway: glycolysis



The diagram illustrates the glycolysis pathway in *E. coli*. It starts with glucose-6-phosphate, which is converted to fructose-6-phosphate. Fructose-6-phosphate is then converted to fructose-1,6-bisphosphate. Fructose-1,6-bisphosphate is cleaved into glyceraldehyde-3-phosphate and dihydroxy-acetone-phosphate. Glyceraldehyde-3-phosphate is further converted to 3-phospho-D-glyceroyl-phosphate. The diagram also shows the pentose phosphate pathway branching off from fructose-6-phosphate.

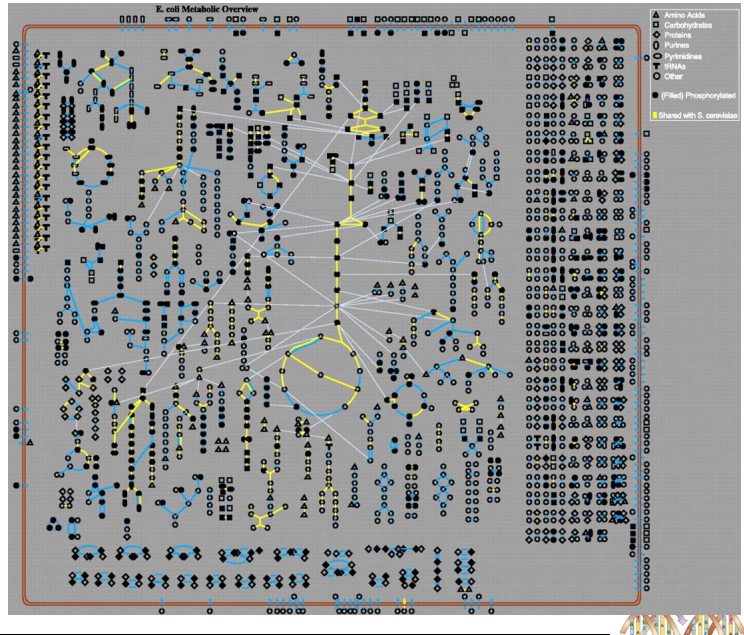
(C) Prof. R. Müller, Prof. E. Rahm

6 - 36



EcoCyc: Übersicht

- "Karten" für Gesamtstoffwechsel
- Überlagerung mit anderen Spezies



EcoCyc: Tools

- PathoLogic**
 - Prediction of metabolic network from genome
 - Computational creation of new Pathway/Genome Databases
- Pathway/Genome Editors**
 - Distributed curation of PGDBs
 - Distributed object database system, interactive editing tools
- Pathway/Genome Navigator**
 - WWW publishing of PGDBs
 - Querying, visualization of pathways, chromosomes, operons
 - Analysis operations
 - Pathway visualization of gene-expression data
 - Global comparisons of metabolic networks
- Bioinformatics 18:S225 2002**

EcoCyc: Tools (2)

Query, visualization and editing tools for these datatypes:

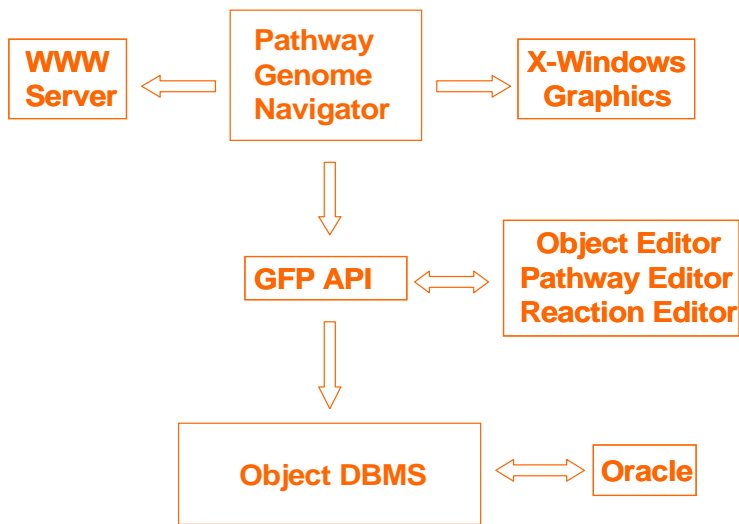
- Full Metabolic Map**
 - Paint gene expression data on metabolic network; compare metabolic networks
- Pathways**
 - Pathway prediction
- Reactions**
 - Balance checker
- Compounds**
 - Chemical substructure comparison
- Enzymes, Transporters, Transcription Factors**
- Genes:**
- Chromosomes**
- Operons**
 - Operon prediction



EcoCyc: Architektur

- Allegro Common Lisp**
- Sun and PC platforms**
- Ocelot object database**
- 250,000 lines of code**
- Lisp-based WWW server at BioCyc.org**
 - Manages 15 PGDBs

EcoCyc: Architektur (2)

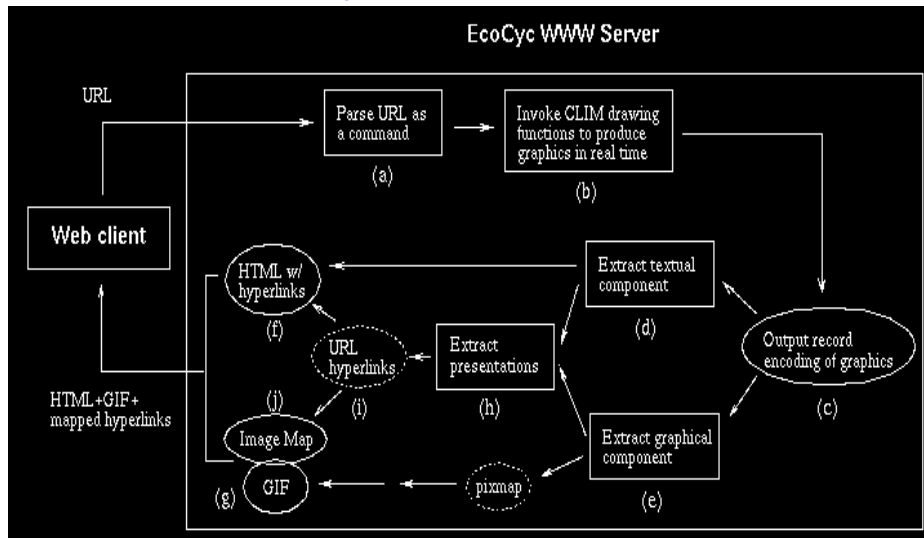


EcoCyc: Architektur (3)

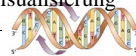
- **Persistent storage via disk files, Oracle DBMS**
 - Concurrent development: Oracle
 - Single-user development: disk files
 - Read-only delivery: bundle data into binary program
- **Oracle storage**
 - DBMS is submerged within Ocelot, invisible to users
 - Relational schema is domain independent, supports multiple KBs simultaneously
 - Frames transferred from DBMS to Ocelot
 - ◆ On demand
 - ◆ By background prefetcher
 - ◆ Memory cache
 - ◆ Persistent disk cache to speed performance via Internet



EcoCyc: WWW-Server



– GLIM (Generalized Interactive Modelling): Software für Datenauswertung und -visualisierung



Weitere Quellen zu Pathway-Datenbanken

- n The BioPathways Consortium,
– <http://www.biopathways.org/>
- n Pacific Symposium on Biocomputing
– <http://psb.stanford.edu/>
- n ASM/TIGR Conference on Microbial Genomes
– <http://www.tigr.org/conf/mg/index.htm>
- n International Conference on Systems Biology 2004
– <http://www.icsb2004.org/>



Zusammenfassung

n Motivation und biologische Grundlagen

- Metabolische Pfade
- Regulatorische Pfade

n Pathway-Datenbanken und ihre Anwendung

- Anforderungen
- KEGG
- EcoCyc

n Spezifische Problematik

- Abbildung der Netzwerke (Aufsplitterung der Abhängigkeitsgraphen in Relationen versus OO-Konzepte)
- Visualisierung
- Zyklenerkennung

