

**Einführung in die Angewandte Bioinformatik:
Netzwerke
15.07.2010**

Prof. Dr. Sven Rahmann

Modellierung von Interaktionen

Ein lebendes System besteht aus Molekülen (verschiedener Art) und „funktioniert“ aufgrund der Interaktionen zwischen diesen.

Netzwerke dienen zur Beschreibung von **Interaktionen**.

Mathematische Grundlage: **Graphen**

Graphen kennen wir schon von phylogenetischen Bäumen.

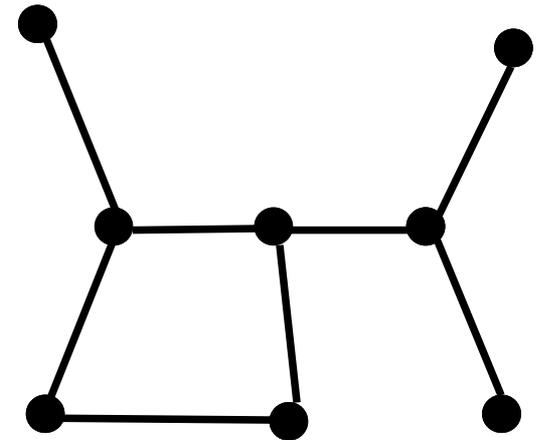
Graphen zur Modellierung von Interaktionen

Ein **Graph** besteht aus:

- **Knoten** (vertices V , nodes) und
- **Kanten** (edges E),
die zwischen den Knoten verlaufen.

Graphen können **gerichtet** oder **ungerichtet** sein
(Kanten haben eine Richtung oder nicht).

Natürlicherweise modellieren die Knoten die
Objekte des Systems, Kanten die Interaktionen.
Sowohl Knoten als auch Kanten erhalten weitere Annotationen.



Arten von Netzwerken (außerhalb der Biologie)

Soziale Netzwerke (z.B. studivz, facebook, xing):

Jeder kennt jeden über höchstens 6 Ecken ?!

Knoten: Personen

Kanten: „kennt“, im Prinzip ungerichtet (A kennt B \Leftrightarrow B kennt A)

Das Internet (WWW)

Knoten: Einzelne Seiten, oder aber Domains

Kanten: Links, gerichtet

(Problematische Modellierung für dynamische Inhalte, Web 2.0)

Projektpläne

Knoten: Tätigkeiten

Kanten: Abhängigkeiten zwischen Tätigkeiten, Voraussetzungen; gerichtet

Biologische Netzwerke

- Transkriptionelle Genregulationsnetzwerke
Transkriptionsfaktor-Protein bindet an DNA / Gen
Betrachte Gen und Genprodukt (Protein) als dasselbe Objekt (Knoten)
- Protein-Protein-Interaktionsnetzwerke
physikalischer Kontakt zwischen Proteinen
- Signaltransduktionsnetzwerke
- Metabolische Reaktionsnetzwerke

Übersicht: Pathguide

Einstiegspunkt

Pathguide

<http://www.pathguide.org/>

Suchkriterien in Pathguide

inhaltlich
(Art des Netzwerks)

nach Organismen
nach Verfügbarkeit
unterstützte Standards

Pathguide the pathway resource list

Home BioPAX cBio MSKCC

Navigation

- Protein-Protein Interactions
- Metabolic Pathways
- Signaling Pathways
- Pathway Diagrams
- Transcription Factors / Gene Regulatory Networks
- Protein-Compound Interactions
- Genetic Interaction Networks
- Protein Sequence Focused
- Other

Search

Organisms: All

Availability: All

Standards: All

Reset Search

Statistics

Analyze Pathguide

Contact

Comments, Questions, Suggestions are Always Welcome!

Complete Listing of All Pathguide Resources

Pathguide contains information about **240** biological pathway resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please [send us an e-mail](#).

News

Find it Easier
Many new search options are available

Get the Stats
Detailed Pathguide resource statistics now available

Pathguide Published

Protein-Protein Interactions

Database Name (Order: alphabetically by web popularity)	Full Record	Availability	Standards
3DID - 3D interacting domains	Details	Free	
ABCdb - Archaea and Bacteria ABC transporter database	Details	Free	
AfCS - Alliance for Cellular Signaling Molecule Pages Database	Details	Free	
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	Free	
ASEdb - Alanine Scanning Energetics Database	Details	Free	
ASPD - Artificial Selected Proteins/Peptides Database	Details	?	
BID - Binding Interface Database	Details	Free	
BIND - Biomolecular Interaction Network Database	Details	Free	PSI-MI
BioGRID - General Repository for Interaction Datasets	Details	Free	PSI-MI
BRITE - Biomolecular Relations in Information Transmission and Expression	Details	Free	
CA1Neuron - Pathways of the hippocampal CA1 neuron	Details	Free	
Cancer Cell Map - The Cancer Cell Map	Details	Free	BioPAX
CellCircuits - CellCircuits	Details	Free	
CSP - Cytokine Signaling Pathway Database	Details	X	
CTDB - Calmodulin Target Database	Details	Free	
DDIB - Database of Domain Interactions and Bindings	Details	Free	
DIP - Database of Interacting Proteins	Details	Free	PSI-MI

Pathguide: Suchkriterien bezüglich Verfügbarkeit

- frei (für alle)
- frei (für akademische Nutzer)
- kommerziell

Pathguide: Suchkriterien bezüglich unterstützter Standards für den Datenaustausch

- **BioPAX** (<http://www.biopax.org/>):
Datenaustauschformat für biologische / biochemische Netzwerke,
aktuell Level 3
- **CellML** (<http://www.cellml.org/>):
XML-basiertes Datenaustauschformat für (mathematische) Modelle
von (Teilaspekten) von Zellen
- **PSI-MI** (<http://www.psidev.info/>):
PSI = Proteomics Standards Initiative
MI = Molecular Interactions
XML-basiertes Datenaustauschformat
für Interaktionsdaten, aktuell v2.5
- **SBML** (<http://www.sbml.org/>):
Systems Biology Markup Language
Format für biochemische Reaktionsnetze

Warum (so viele) Standards?

- Warum Standards?
 - Datenaustausch zwischen verschiedenen Anwendungen
 - offene Standards: Unabhängigkeit von Software-Anbietern
 - Standardisierungsprozess: schafft Klarheit über Modellierungsaspekte
- Warum viele?
 - Daten Technologie- und Anwendungs-spezifisch
 - Neue Entwicklungen schaffen neue Messmöglichkeiten
 - verschiedene Ziele (z.B. Daten menschen- oder maschinen-lesbar?)
 - Zu einfache Standards können nicht genug
 - Zu umfangreiche Standards sind nicht mehr benutzbar
 - Anwendungs-spezifischer Mittelweg zwischen Einfachheit und Komplexität

Datenbanken zu molekularen Interaktionen allgemein: BOND mit BIND

- **BIND** (Biomolecular Interaction Network Database),
mittlerweile integriert in das kommerzielle integrative System
- **BOND** (Biomolecular Object Network Databank,
<http://bond.unleashedinformatics.com>)
= BIND + Sequenzdatenbanken + Abfragemöglichkeiten

Inhalt von BIND

Interaktionen zwischen mindestens zwei Objekten, die experimentell in einem lebenden System nachgewiesen wurden (mit Literaturverweis)

Objekte: Proteine, DNA, RNA, Liganden, Komplexe, Gene, Photonen, ...

Molekulare Komplexe entstehen durch Verbindung von mindestens zwei Interaktionen mit gemeinsamen Objekten, zusätzliche Annotation der Struktur des Komplexes

Pathways entstehen durch Verbindung von mindestens zwei Interaktionen mit gemeinsamen Objekten, in einer festgelegten Reihenfolge; zusätzliche Annotation des Pathways (unter welchen Umständen benutzt?)

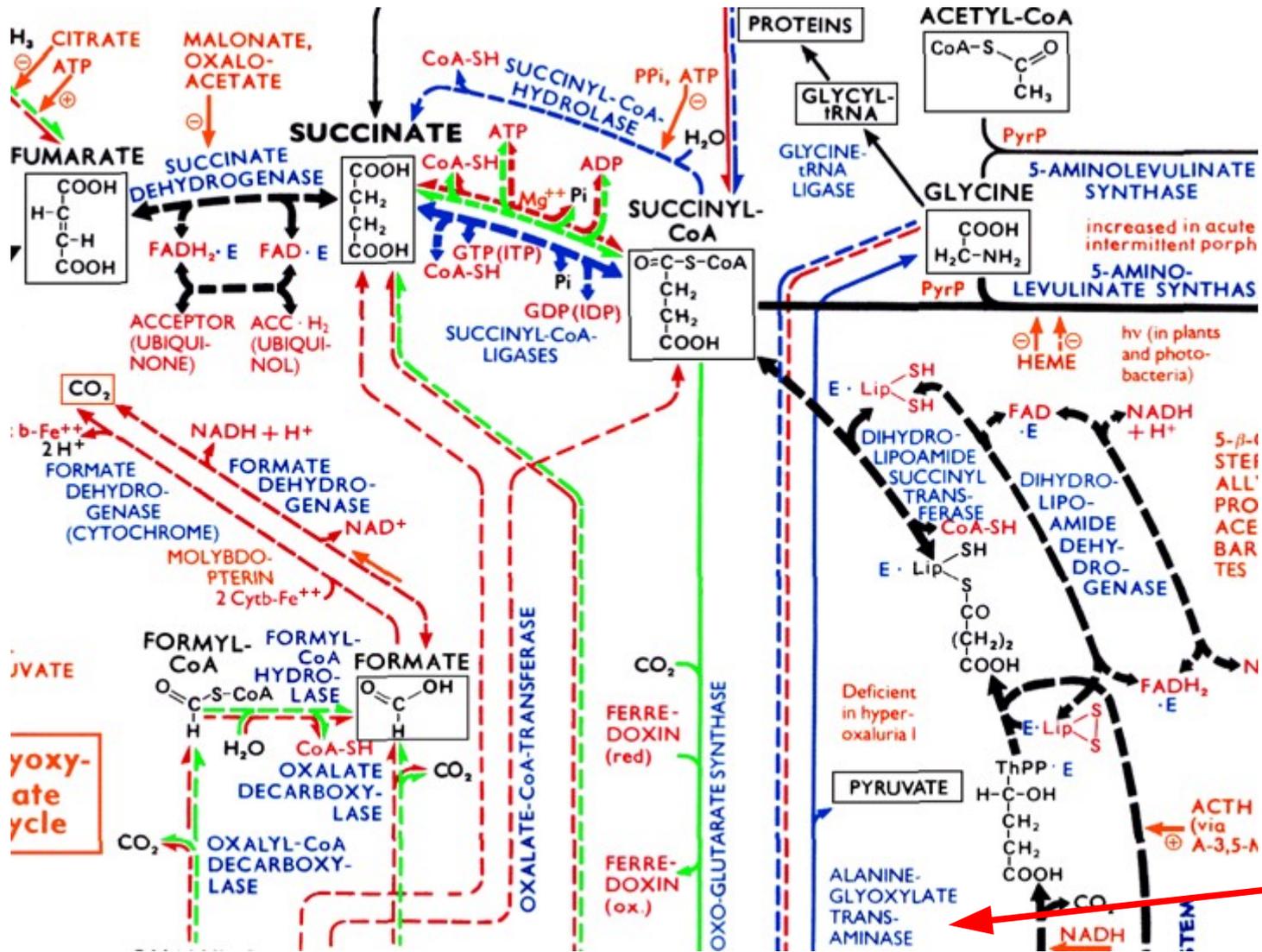
„Biochemical Pathways“

Definition [absichtlich schwammig]

Unter einem biochemischen Pfad („pathway“) versteht man einen zusammenhängenden Teil eines biochemischen Reaktionsnetzwerkes (der in einem inhaltlichen Zusammenhang steht).

Datenbanken mit Zugriff über ExPASy

- Digitalisierte Version der Wandtafel „Biochemical Pathways“ von Roche Applied Science (<http://www.expasy.ch/tools/pathways/>), Enzyme-DB-Einträge im Bild verlinkt
- Enzyme-Datenbank (<http://www.expasy.ch/enzyme/>)
- UniPathway (<http://www.grenoble.prabi.fr/obiwarehouse/unipathway>), eine metabolische Sichtweise auf UniProt (insbes. Swiss-Prot)



Link zur ENZYME-Datenbank

NiceZyme View of ENZYME: EC 2.6.1.44

Official Name

Alanine--glyoxylate transaminase.

Alternative Name(s)

AGT.

Alanine--glyoxylate aminotransferase.

Reaction catalysed

L-alanine + glyoxylate <=> pyruvate + glycine

Cofactor(s)

Pyridoxal-phosphate.

Comment(s)

- With one component of the animal enzyme, 2-oxobutanoate can replace glyoxylate.
- A second component also catalyzes the reaction of EC 2.6.1.51.

Cross-references

Biochemical Pathways; map number(s)	B7 ; G5
PROSITE	PDOC00514
BRENDA	2.6.1.44
PUMA2	2.6.1.44
PRIAM enzyme-specific profiles	2.6.1.44
KEGG Ligand Database for Enzyme Nomenclature	2.6.1.44
IUBMB Enzyme Nomenclature	2.6.1.44
IntEnz	2.6.1.44
MEDLINE	Find literature relating to 2.6.1.44
MetaCyc	2.6.1.44

UniProtKB/Swiss-Prot

Q940M2, AGT21_ARATH;	Q94AL9, AGT22_ARATH;	Q9SR86, AGT23_ARATH;
Q17QF0, AGT2_BOVIN;	Q9BYV1, AGT2_HUMAN;	Q3UEG6, AGT2_MOUSE;
Q5RFA3, AGT2_PONAB;	Q64565, AGT2_RAT;	P43567, AGX1_YEAST;
Q56YA5, SGAT_ARATH;	P84187, SGAT_MAIZE;	P84188, SGAT_WHEAT;
P31029, SPYA_CALJA;	Q54GT6, SPYA_DICDI;	P41689, SPYA_FELCA;
P21549, SPYA_HUMAN;	Q35423, SPYA_MOUSE;	Q5RDPO, SPYA_PONAB;
P31030, SPYA_RABIT;	P09139, SPYA_RAT;	

Zugriff über ExPASy,
Informationen über Enzyme
und die von ihnen katalysierten
Reaktionen
Literaturverweise
Verweise auf Sequenz-DBen

KEGG (Kyoto Encyclopedia of Genes and Genomes) besteht aus mehreren Datenbanken.



KEGG Home
Introduction
Overview
Release notes
Current statistics

KEGG Identifiers

[KGML](#)

KEGG API

KEGG FTP

KegTools

Feedback

GenomeNet

KEGG: Kyoto Encyclopedia of Genes and Genomes

A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. Towards this end we have been developing a bioinformatics resource named KEGG as part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.

- **Main entry point to the KEGG web service**
[KEGG2](#) [KEGG Table of Contents](#) [Update notes](#) [Help](#)
- **Data-oriented entry points**
 - [KEGG Atlas](#) Global maps of cell/organism functions
 - [KEGG PATHWAY](#) Pathway maps and pathway modules
 - [KEGG BRITE](#) Functional hierarchies and ontologies
 - [KEGG GENES](#) Genomes, genes, proteins, and orthologs
 - [KEGG LIGAND](#) Chemical compounds, drugs, glycans, and reactions
- **Organism-specific entry points**
[KEGG Organisms](#) Select (example) [hsa](#)
- **Subject-specific entry points**
 - [KEGG DISEASE](#) Gene/molecule based disease information resource
 - [KEGG DRUG](#) Chemical structure based drug information resource
 - [KEGG GLYCAN](#) Glycome informatics resource
 - [KEGG COMPOUND](#) Knowledge base for biochemical compounds
 - [KEGG REACTION](#) Knowledge base for biochemical reactions
 - [KAAS](#) KEGG automatic annotation server

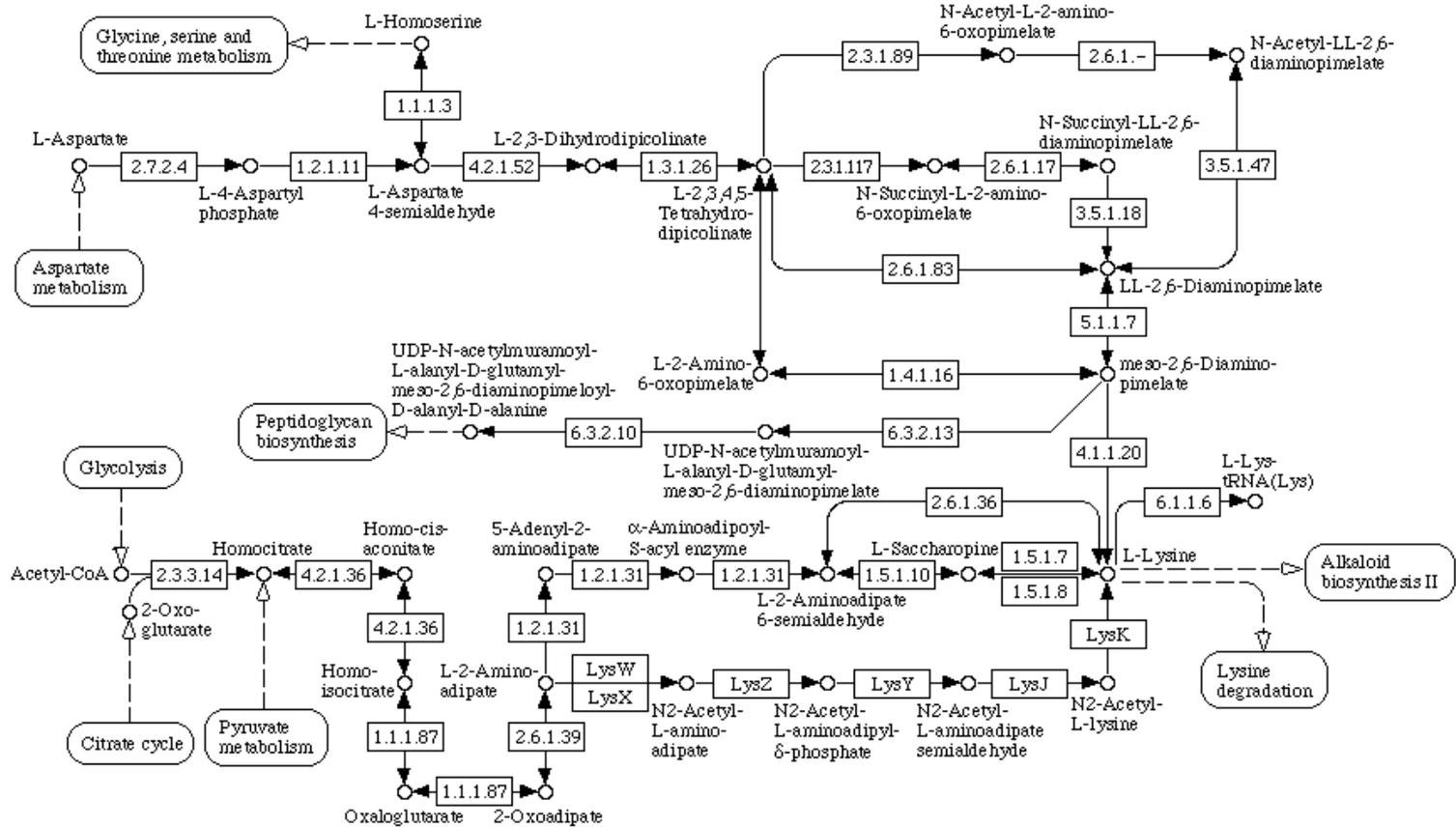
Copyright 1995-2008 Kanehisa Laboratories

Für uns interessant:
KEGG PATHWAY,
enthält handgezeichnete
Reaktionsdiagramme,
geordnet z.B. nach
metabolischer Funktion

KEGG PATHWAY

LYSINE BIOSYNTHESIS

Diagramme

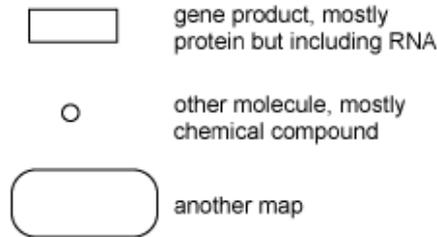


00300 8/24/07

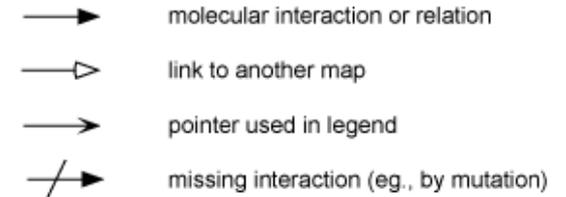
Legende für KEGG PATHWAY

Objekte (Moleküle)
= verschiedene Knotentypen

Objects

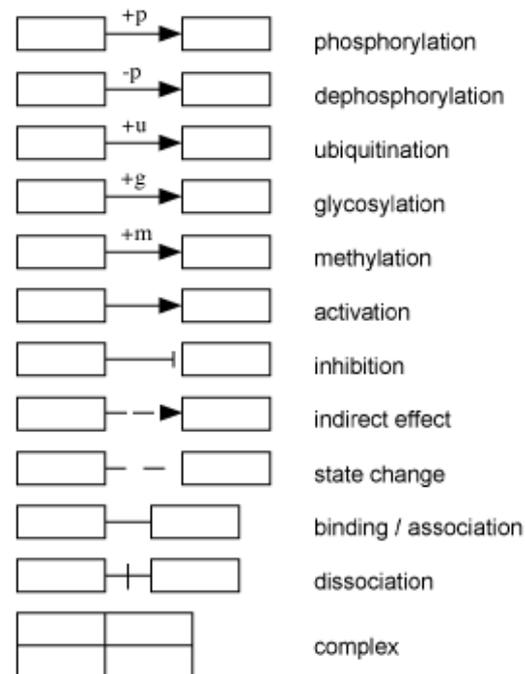


Arrows

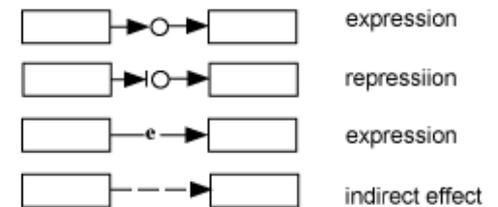


Beziehungen (Reaktionen)
= verschiedene Kantentypen

Protein-protein interactions



Gene expression relations



Enzyme-enzyme relations



BioCyc, EcoCyc, MetaCyc

BioCyc (v14.1, Juli 2010, <http://biocyc.org/>)

enthält Daten zu metabolischen Netzen und Genomen von 673 verschiedenen Organismen, größtenteils automatisch generiert. Von Hand gepflegt werden nur EcoCyc und MetaCyc.

EcoCyc (<http://ecocyc.org>) ist eine umfassende Datenbank zu verschiedenen Netzwerken des Modellorganismus E.coli, u.a. Metabolismus, transkriptionelle Regulation, Transport.

MetaCyc (<http://metacyc.org/>) enthält nicht-redundante Repräsentanten von >1100 Metabolischen Pathways aus >1500 verschiedenen Organismen

Größte Abdeckung des „metabolischen Universums“

- durchstöberbar (browsable)
- durchsuchbar
- aus eigenen Programmen abfragbar
- unterstützt BioPAX und SBML-Standards

Protein-Interaktionen: DIP

Database of Interacting Proteins, UCLA; <http://dip.doe-mbi.ucla.edu/>
Registrierung erforderlich (kostenlos für akademische Nutzer)

Number of proteins	19935
Number of organisms	204
Number of interactions	56638
Number of distinct experiments describing an interaction	64241
Number of data sources (articles)	3516
Number of data sources (other)	34

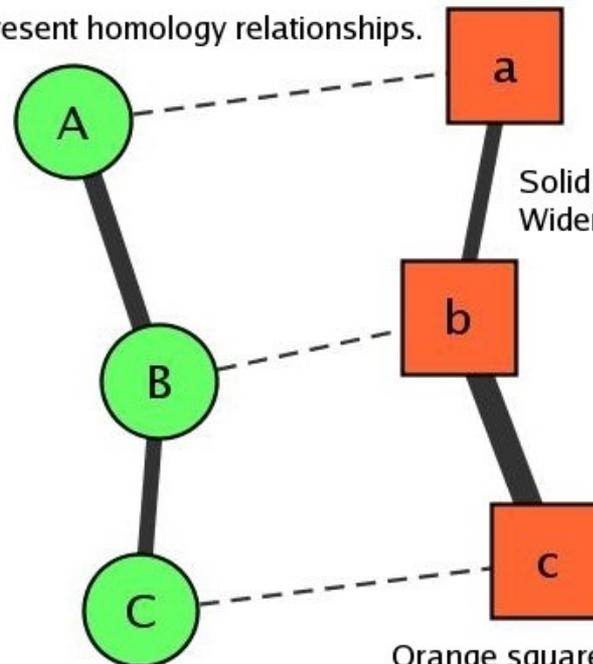
ORGANISM	PROTEINS	INTERACTIONS	EXPERIMENTS	Details
<i>Drosophila melanogaster</i> (fruit fly)	7088	21022	21141	●
<i>Saccharomyces cerevisiae</i> (baker's yeast)	4923	18343	22755	●
<i>Escherichia coli</i>	1858	7440	8842	●
<i>Caenorhabditis elegans</i>	2644	4038	4084	●
<i>Homo sapiens</i> (Human)	1298	1923	2819	●
<i>Helicobacter pylori</i>	711	1426	1426	[---]
<i>Mus musculus</i> (house mouse)	346	453	654	●
<i>Rattus norvegicus</i> (Norway rat)	128	166	243	●
Others (196)	939			

PathBLAST: Transfer von Protein-Interaktionen

- Eingabe von interagierenden Proteinen in einem Organismus
- Eingabe des Zielorganismus
- Ausgabe: Homologe Proteine im Zielorganismus mit konserviertem Interaktionsmuster

- benutzt die DIP-Daten

Dashed lines represent homology relationships.



Solid lines represent protein interactions.
Wider lines imply more certain interactions.

Green circles represent query proteins.

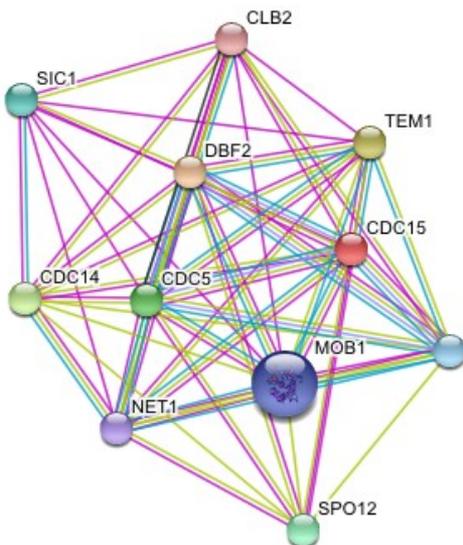
Orange squares represent target species proteins.

Protein-Assoziationen: STRING

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) am EMBL (<http://string.embl.de>):

sagt Protein-Assoziationen aus verschiedenen Arten von Evidenz vorher: genomische Nachbarschaft, Genfusion, gemeinsames Auftreten über viele Spezies, gemeinsame Genexpression, experimentelle physikalische Interaktions-Evidenz, Assoziation in hand-gepflegten Datenbanken, automatisches Text-Mining

This is the **evidence view**. Different line colors represent the types of evidence for the association.



Your Input:

- CDC15 Protein kinase of the Mitotic Exit Network that is localized to the spindle pole bodies at late anaphase; promotes mitotic exit by directly switching on the kinase activity of Dbf2p (974 aa) (*Saccharomyces cerevisiae*)

Predicted Functional Partners:

- DBF2 Ser/Thr kinase involved in transcription and stress response; functions as part of a network of genes in exit from mitosis; localiz
- TEM1 GTP-binding protein of the ras superfamily involved in termination of M-phase; controls actomyosin and septin dynamics during c
- CDC14 Protein phosphatase required for mitotic exit; located in the nucleolus until liberated by the FEAR and Mitotic Exit Network in ana
- CDC5 Polo-like kinase with similarity to Xenopus Plx1 and S. pombe Plo1p; found at bud neck, nucleus and SPBs; has multiple function
- SPO12 Nucleolar protein of unknown function, positive regulator of exit from mitosis; involved in regulating the release of Cdc14p from
- SIC1 Inhibitor of Cdc28-Clb kinase complexes that controls the G1/S phase transition, preventing premature S phase entry and ensur
- DBF20 Ser/Thr kinase involved in late nuclear division, one of the mitotic exit network (MEN) proteins; necessary for the execution of c
- MOB1 Component of the mitotic exit network; associates with and is required for the activation and Cdc15p-dependent phosphorylation
- NET1 Core subunit of the RENT complex, which is a complex involved in nucleolar silencing and telophase exit; stimulates transcription
- CLB2 B-type cyclin involved in cell cycle progression; activates Cdc28p to promote the transition from G2 to M phase; accumulates dur

Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
●	●	●	●	●	●	●	●	0.999
●	●	●	●	●	●	●	●	0.999
●	●	●	●	●	●	●	●	0.999
●	●	●	●	●	●	●	●	0.998
●	●	●	●	●	●	●	●	0.999
●	●	●	●	●	●	●	●	0.996
●	●	●	●	●	●	●	●	0.965
●	●	●	●	●	●	●	●	0.991
●	●	●	●	●	●	●	●	0.988
●	●	●	●	●	●	●	●	0.984

Views:



Transkriptionelle Genregulation

Struktur der Graphen bei Transkriptioneller Genregulation

$X \rightarrow Y \rightarrow Z$ heißt:

Das Produkt (Protein) von Gen X ist ein Transkriptionsfaktor (TF), der die Transkription von Y reguliert (aktiviert oder reprimiert).

Das Gen Y codiert wiederum einen TF, der Z reguliert.

Darstellung von Transkriptionsfaktorbindestellen

Transkriptionsfaktoren binden an bestimmten Stellen an die DNA, (Transkriptionsfaktorbindestellen, TFBS),

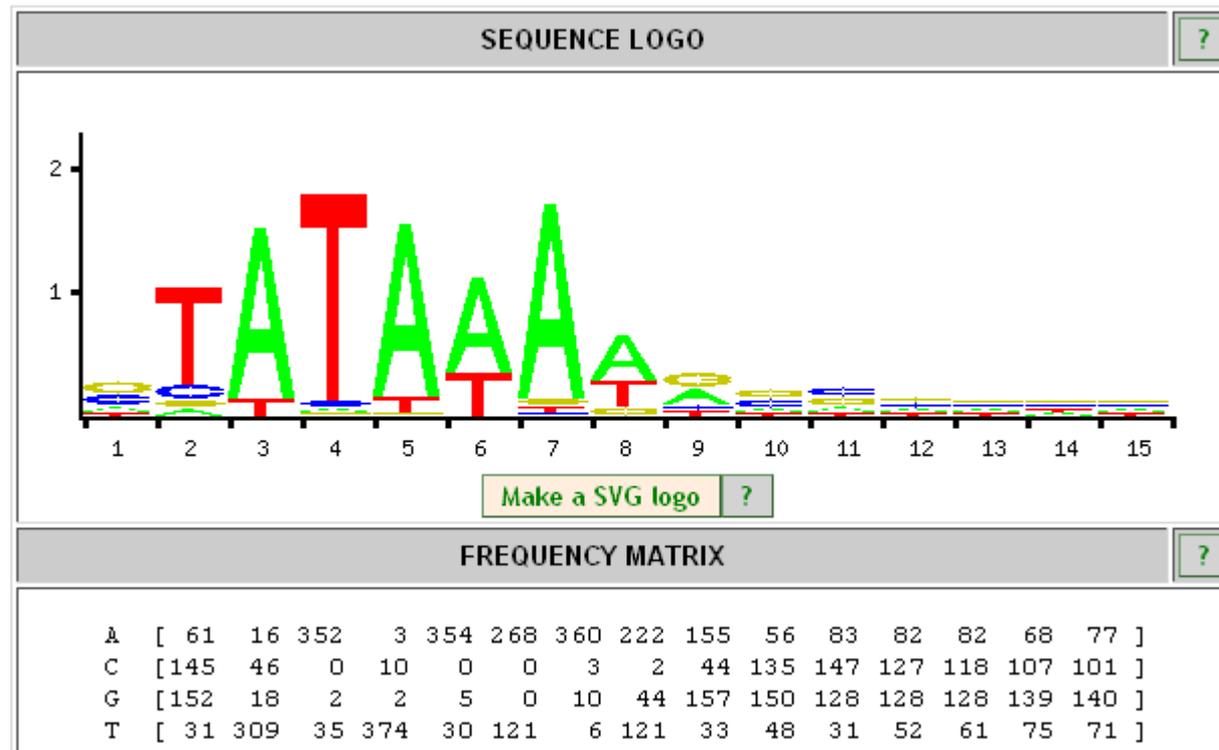
die durch DNA-Sequenz-Motive beschrieben werden können.

Diese wiederum werden oft durch Sequenzlogos visualisiert.

Transkriptionelle Genregulation – Datenbanken

JASPAR (<http://jaspar.genereg.net/>, JASPAR-core):

von Hand annotierte kleine Menge an TFBS-Modellen in mehreren Organismen,
sorgfältig ausgewählt, experimentell stark abgesichert,
frei verfügbar,
stellt TFBSen durch Position-Frequency-Matrices dar.



Transkriptionelle Genregulation - Datenbanken

TRANSFAC (<http://www.gene-regulation.com/>, kommerziell):
umfangreiche Informationen zu experimentell validierten und vorhergesagten
TFBSen in vielen Organismen,
erfordert Registrierung

CoryneRegNet und CoryneCenter (<http://www.CoryneCenter.de>):
Integration von Genregulations-Wissen mit weiteren Genom-, Transkriptomdaten,
speziell auf Corynebakterien.
Erlaubt die Übertragung von bekannten Teilnetzwerken auf verwandte Spezies
mittels homologiebasierter Vorhersage,
sowie die interaktive Visualisierung der Netzwerke

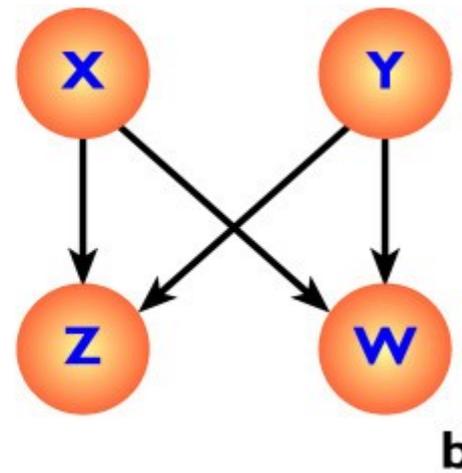
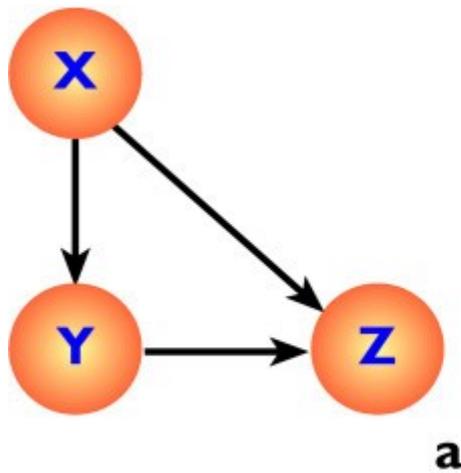
Netzwerk-Motive

Motive sind wiederkehrende Teile von Netzwerken (formal: Subgraphen), z.B.

- a) feed-forward-Motiv
- b) bi-fan-Motiv

Hoffnung, komplexe Netzwerke modular in Motive zu zerlegen,
aus der Funktion einzelner Motive die Funktionen des Netzwerks zu bestimmen.

Problematisch, wenn Beziehungen zwischen den Motiv-Instanzen
außer Acht gelassen werden.

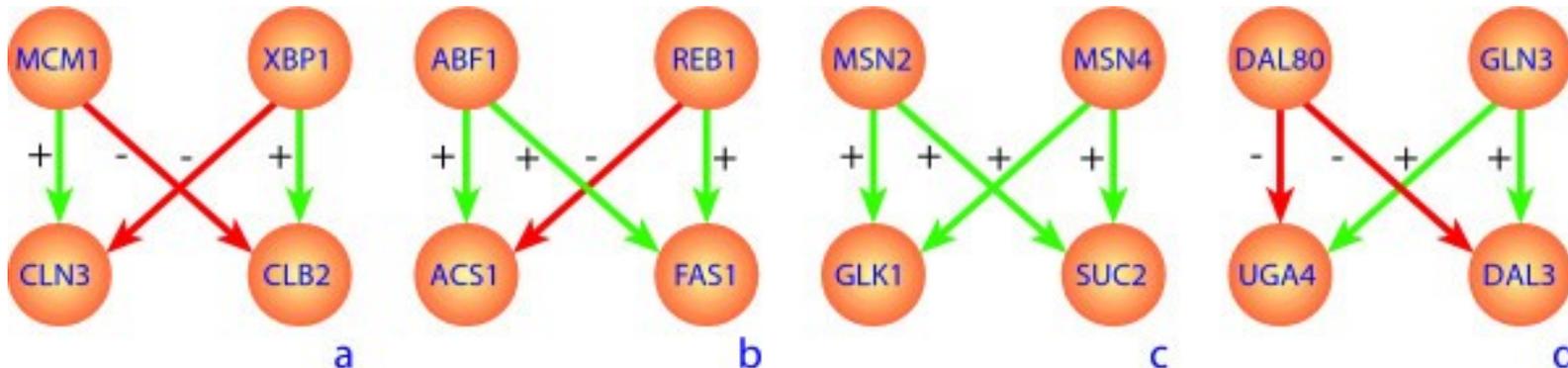


Netzwerk-Motive

Verschiedene Ausprägungen des bi-fan-Motivs
im transkriptionellen Regulationsnetzwerk der Bierhefe

Grün: Aktivierung
Rot: Inhibition

Quelle: Ingram et al., BMC Genomics. 2006; 7: 108.



Trotz gleicher Motiv-Topologie ist die Funktion der vier Motive unterschiedlich.

Analyse und Visualisierung von Netzwerken

Cytoscape (<http://www.cytoscape.org/>)

durch Plug-ins erweiterbare Software-Plattform zur Visualisierung und Analyse von Netzwerken jeder Art, bietet dabei Methoden zur Datenintegration aus verschiedenen Quellen. Man muss Cytoscape herunterladen, installieren und mit Daten füttern.

MapMan (<http://gabi.rzpd.de/projects/MapMan/>)

bildet genomweite Daten (z.B. Genexpressionsdaten) auf funktionelle Kategorien (Module) des Metabolismus ab und verschafft so eine visuelle Übersicht von im Metabolismus betroffenen Bereichen unter verschiedenen experimentellen Bedingungen

MapMan Beispiel

Experiment:
Bestrahlung der Wurzeln von Arabidopsis-Keimlingen mit UVB-Licht.

Veränderungen in der Genexpression werden abgebildet auf Module des Metabolismus

Die Gene wurden per Hand zugeordnet, die Module per Hand visualisiert.

