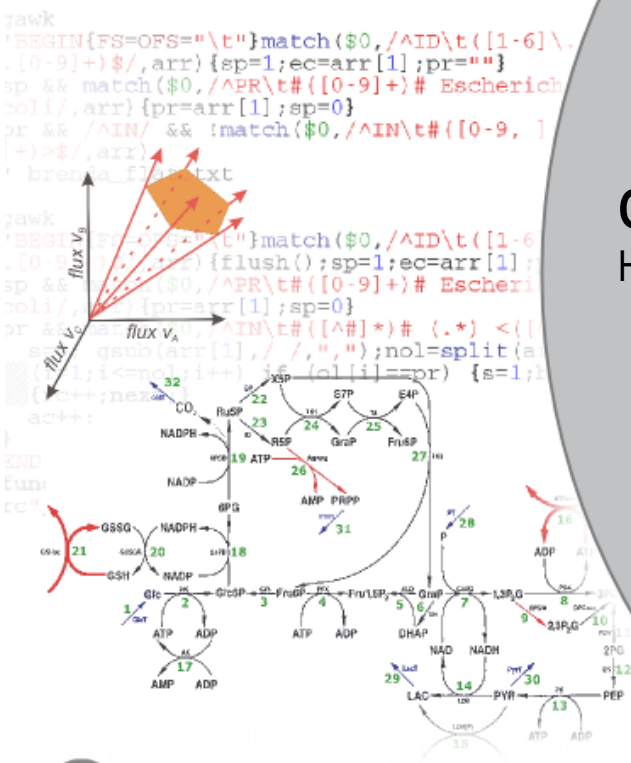
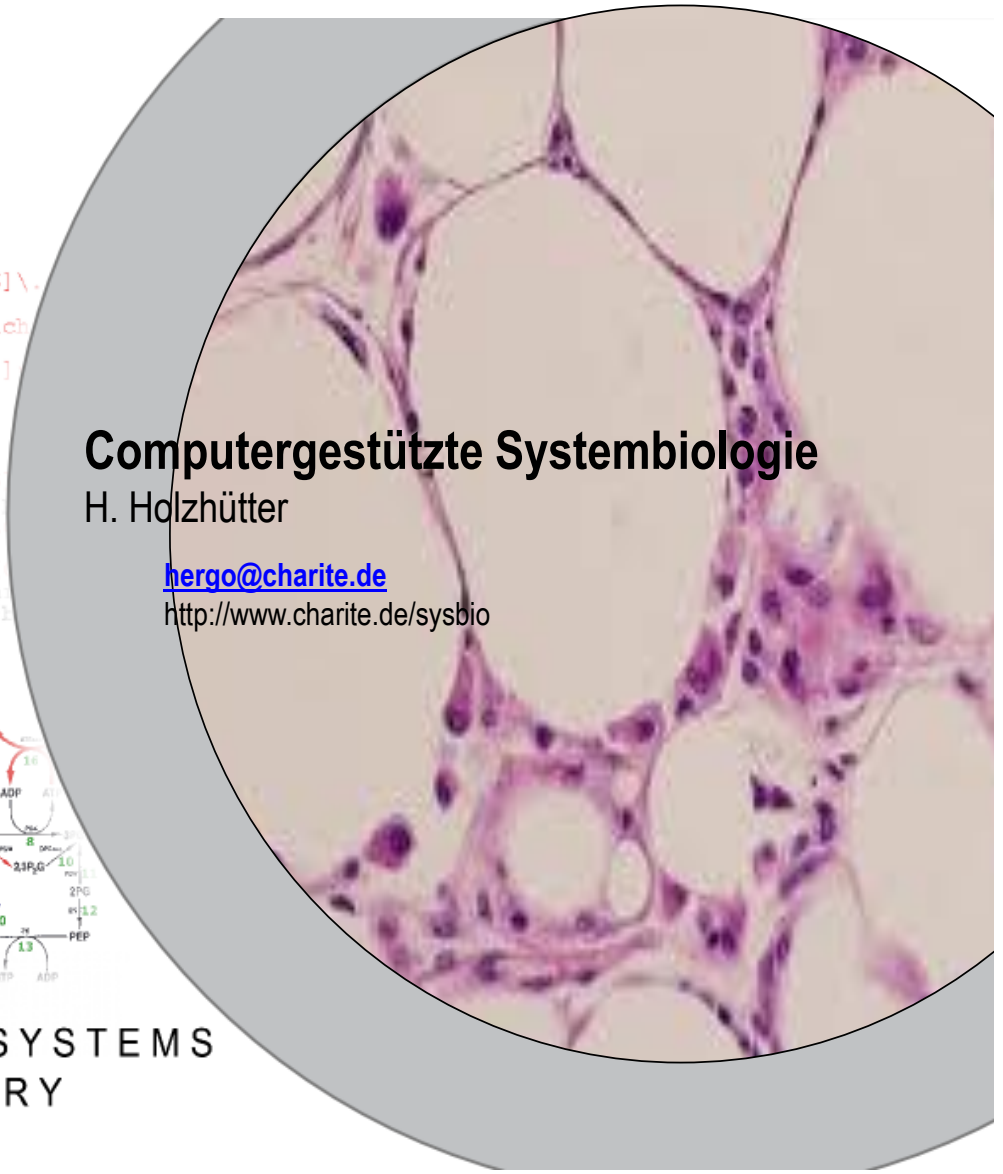
CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN



COMPUTATIONAL SYSTEMS BIOCHEMISTRY



Computergestützte Systembiologie

H. Holzhütter

hergo@charite.de
http://www.charite.de/sysbio

Gliederung

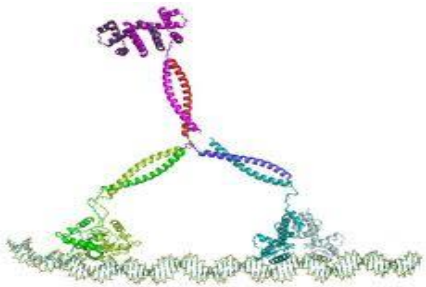
I Systembiologie als neue Wissenschaftsdisziplin

II Biologische Systeme als interagierende molekulare Netzwerke

III Netzwerkmodelle

Systembiologie – ein neues Gebiet Wissenschaftsgebiet ?

Ökologie

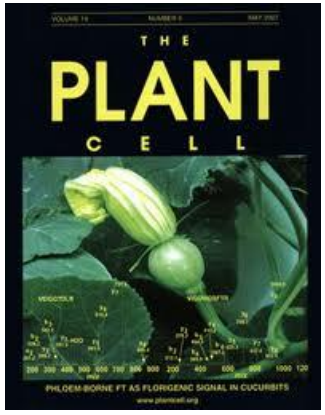


Molekularbiologie

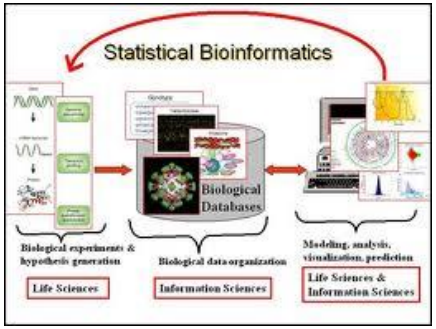
Neurobiologie

Entomologie

Systembiologie?



Pflanzenbiologie



Bioinformatik

Physiologie

Mikrobiologie

Entschlüsselung des Genotyps – The Human Genome Project

DNA (Genotyp)

↓

mRNA

↓

Protein

↓

Lipide
Kohlenhydrate
Nukleotide
Aminosäuren

↓

Organellen

↓

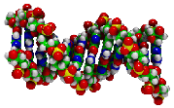
Zelle

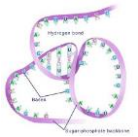
↓

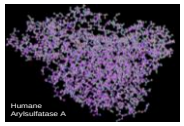
Gewebe / Organ

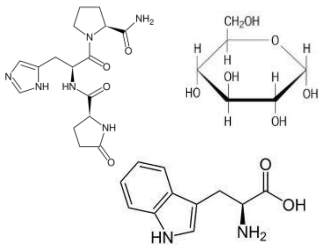
↓

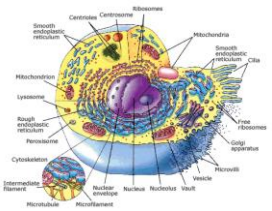
Organismus (Phänotyp)



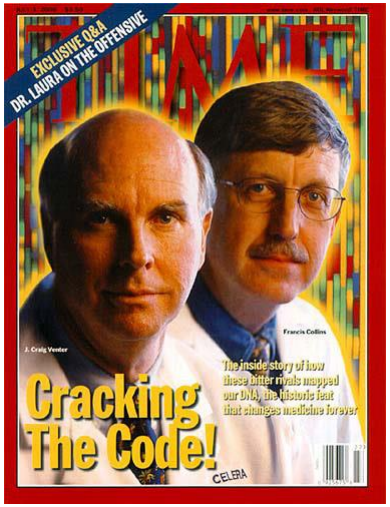








„knowledge of the human genome is as necessary to the continuing progress of medicine and other health sciences as knowledge of human anatomy has been for the present state of medicine



2 konkurrierende Projekte („race for the genome')

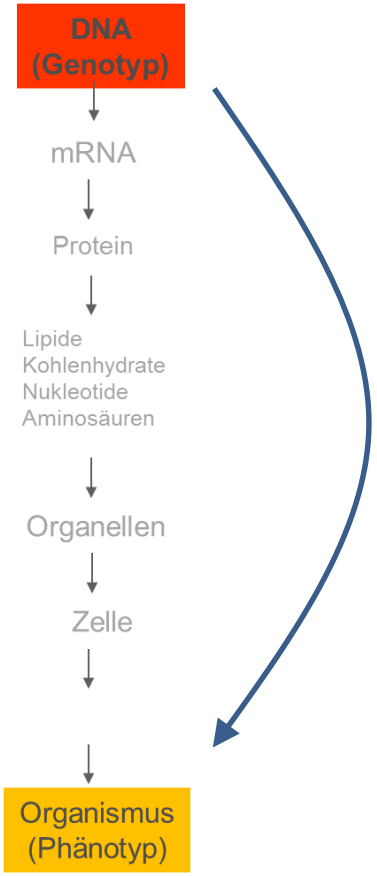
- öffentlich gefördert (USA, England, Frankreich, Japan, Australien Leitung: F. Colins)
- private Initiative von Celera Genomics (C. Venter)

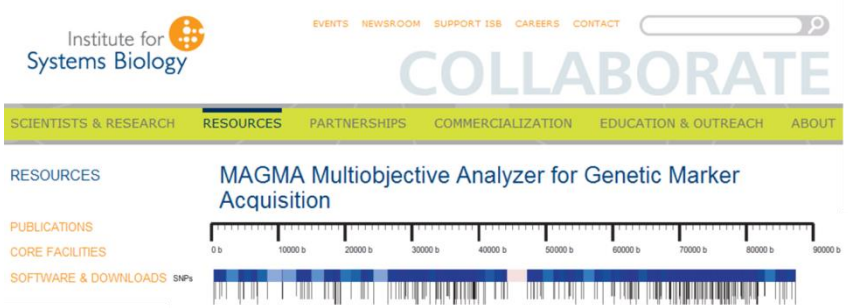
unvollständige DNA Squenz publiziert 2000
vollständige DNA Sequenz publiziert April 2003

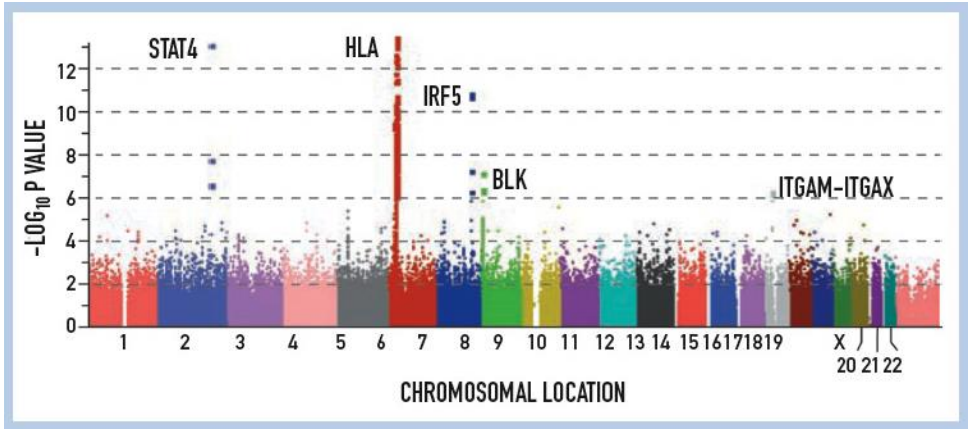
- es gibt etwa 23,000 Gene im humanen Genom (etwa genau so viel wie in der Maus und im Rundwurm)
- die humane DNA besitzt deutlich mehr repetitive Sequenzabschnitte als andere Säugergenome
- weniger als 7% Proteinfamilien sind spezifisch für Wirbeltiere

Genetische Marker für phänotypische Merkmalsausprägung

Heuristik
Der individuelle Phänotyp ist (letztlich) genetisch kodiert



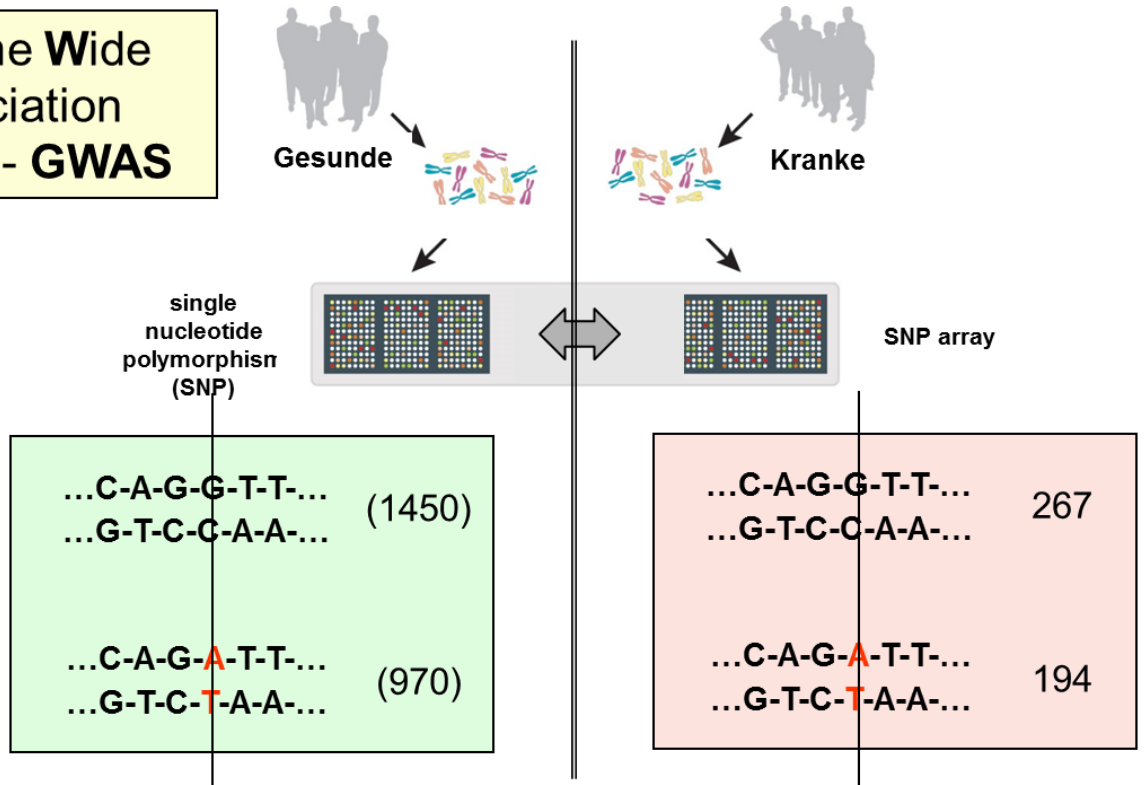




A **genetic marker** is any alteration in your DNA that may indicate an increased risk of developing a specific disease or disorder. Because SNP (Single Nucleotide Polymorphism)s are, by their very definition, variations in DNA, they can be used as flags or markers for nearby DNA that affects your health

Entdeckung von Marker-Genen: Genomweite Assoziationsstudien

Genome Wide Association Studies - GWAS



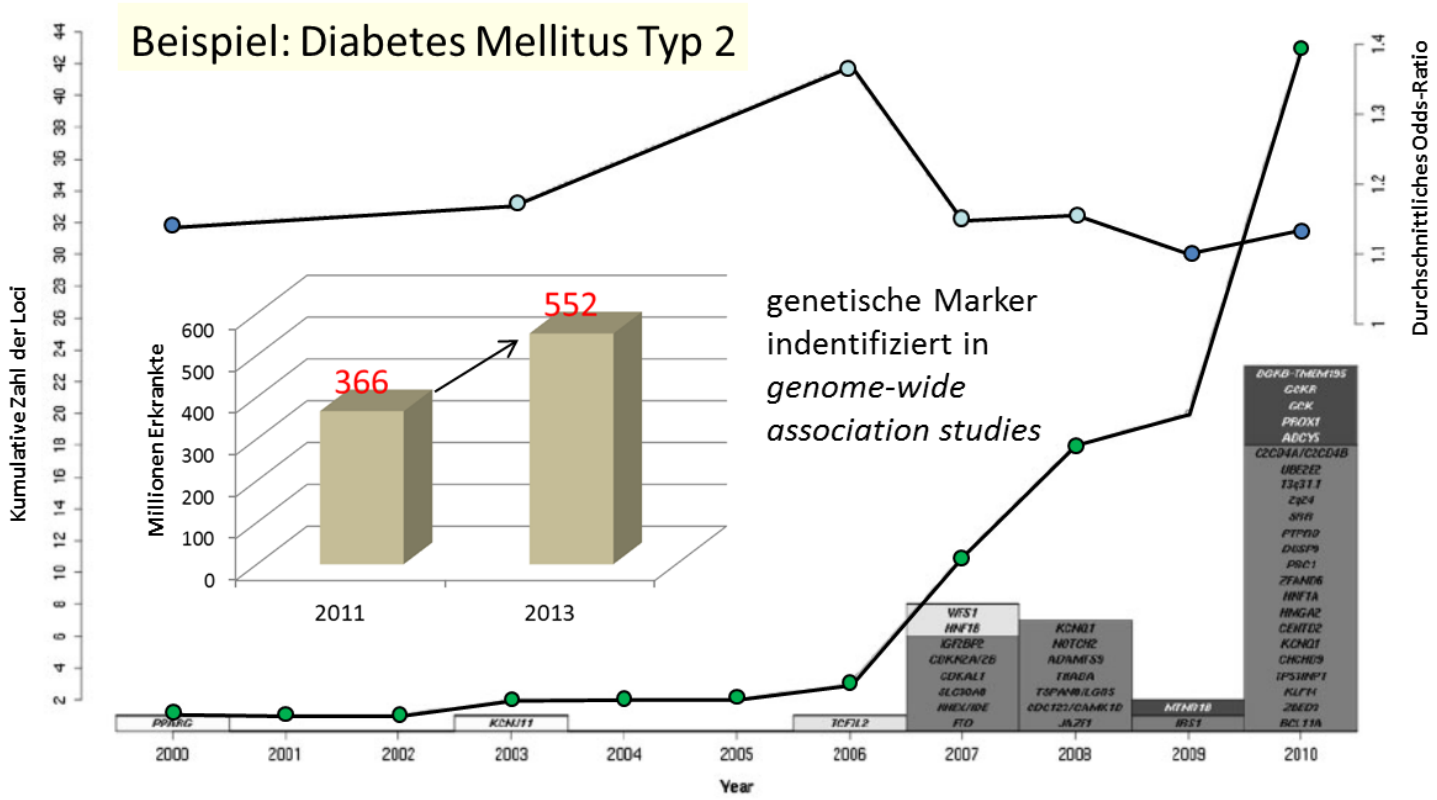
odds = Vorkommen/Nicht-Vorkommen

odds(G) = 970 : 1450 = 0.668966

odds(K) = 194 : 267 = 0.726592

odds ratio (OR) = odds(Kranke) : odds(Gesunde) = 0.726592 : 0.668966 = **1.0861**

Prädiktivität genetischer Marker für die Risikoabschätzung



Wheeler, E. & Barroso, I. (2011) Genome-wide association studies and type 2 diabetes, *Briefings in functional genomics.* ***10, 52-60.***

DNA - Sequenzdaten genügen nicht für das Verständnis von Krankheiten

*It is fair to say that **the Human Genome Project has not yet directly affected the health care of most individuals**."*

nature Vol 464|1 April 2010

OPINION

Has the revolution arrived?

Looking back over the past decade of human genomics, **Francis Collins** finds five key lessons for the future of personalized medicine — for technology, policy, partnerships and pharmacogenomics.

*However, when all this sequence data got into computers, it became obvious **that the genetic blueprints by themselves tell us very little about the functional behavior of cells and multicellular organisms**; that is, about what we really want to know about biological systems. In this way, the human genome project, which is perhaps the most spectacular success of molecular biology, also meant that a vast space of future research of a radically different kind became visible.*

GENOME RESEARCH

Systems Biology Is Taking Off

Måns Ehrenberg, Johan Elf, Erik Aurell, Rickard Sandberg and Jesper Tegnér

Genome Res. 2003 13: 2377-2380
Access the most recent version at doi:10.1101/gr.1763203

Systems Biology: Übersetzung des Genotyps in den Phänotyp

DNA

↓

mRNA

↓

protein

↓

lipids
carbohydrates
nucleotides
amino acids
organic molecules

↓

organelles

↓

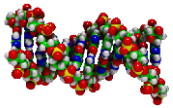
cell

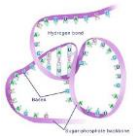
↓

tissue / organ

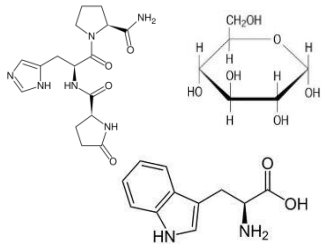
↓

organism

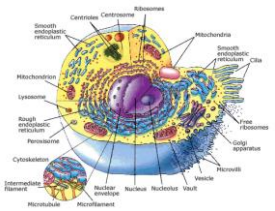




Molekularbiologie
Genetik
Biophysik



Biochemie



Zellbiologie

Physiologie
Anatomie
Medizine

Systembiologie

- ist ein **holistischer Forschungsansatz** – gleichzeitige Untersuchung von Lebensprozessen auf verschiedenen Stufen der hierarchischen Organisation
- erforscht die **molekularen Grundlagen** von komplexen Eigenschaften des Phänotyps (Altern, Wachstum, Krankheit…)
- benutzt moderne Verfahren der Bildgebung (“life imaging”), Sequenzierung und Quantifizierung von Molekülen (Chip-Technologien, Massenspektrometrie)
- produziert und archiviert Daten **genomweit** (**Datenbanken**)
- analysiert Daten mit Hilfe multivariater statistischer Methoden (**Bioinformatik**) und **mathematischer Modelle**

Systembiologie: Abkehr von reduktionistischer Forschung ?

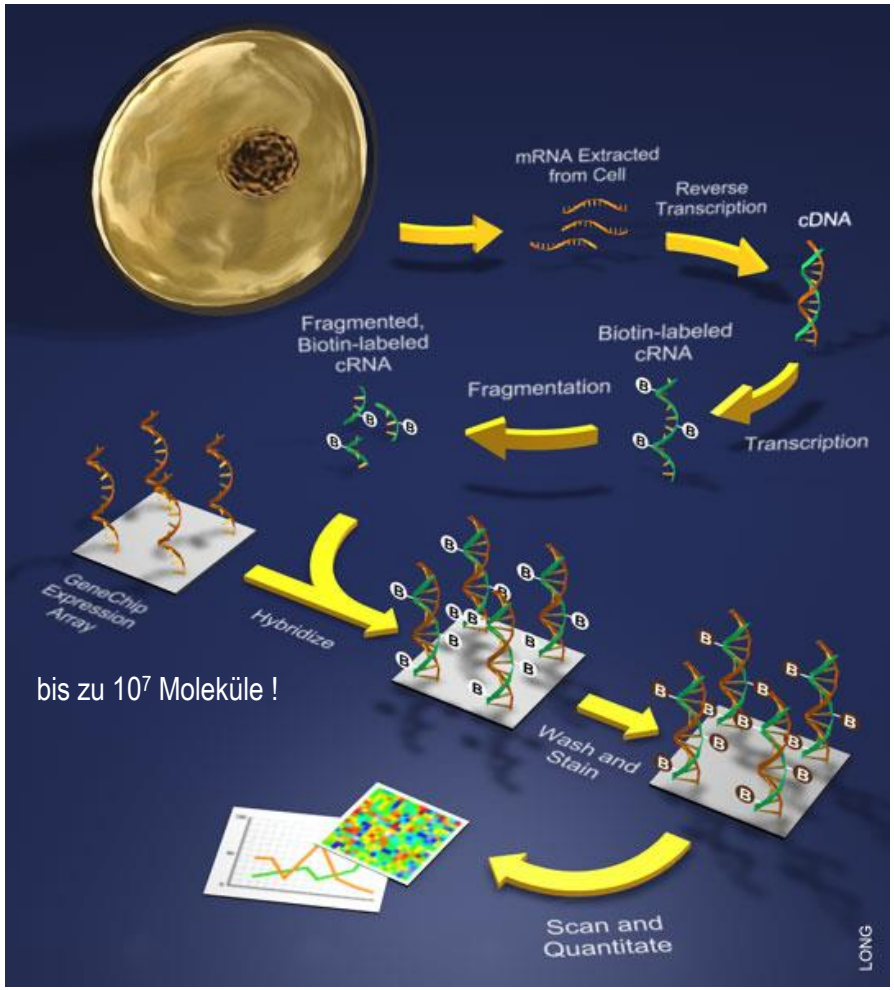
A New Approach to Decoding Life:
Systems Biology

Trey Ideker[1,2], Timothy Galitski[1], and Leroy Hood[1,2,3,4,5]
Institute for Systems Biology[1], Seattle, Washington 98105; Departments of Molecular Biotechnology[2], Immunology[3], Bioengineering[4], and Computer Science and Engineering[5], University of Washington, Seattle, Washington 98195; e-mail: tideker@systemsbiology.org, tgalitski@systemsbiology.org, lhood@systemsbiology.org

approach. Systems biology does not investigate individual genes or proteins one at a time, as has been the highly successful mode of biology for the past 30 years. Rather, it investigates the behavior and relationships of all of the elements in a particular biological system while it is functioning. These data can then be integrated, graphically displayed, and ultimately modeled computationally. How

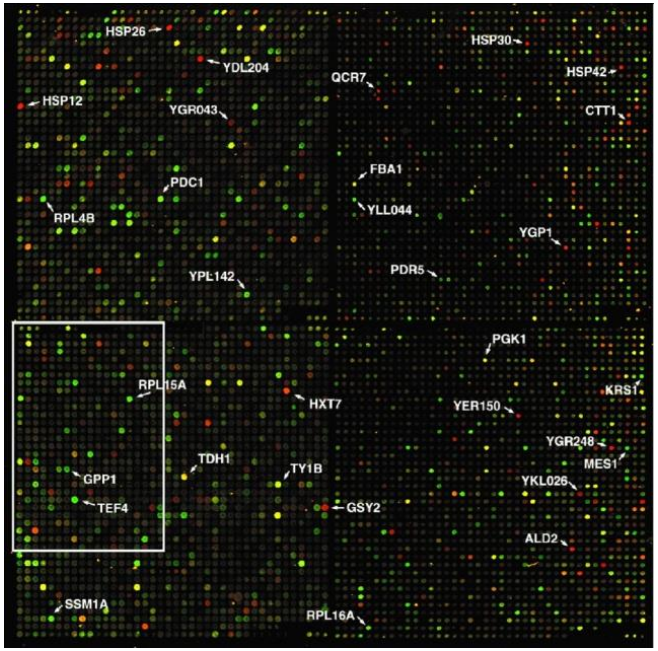
Ideker, T., et al. (2001). "A new approach to decoding life: systems biology." Annu Rev Genomics Hum Genet **2: 343-372.**

Entwicklung von Chip-Techologien (“microarrays”)



bis zu 10^7 Moleküle !

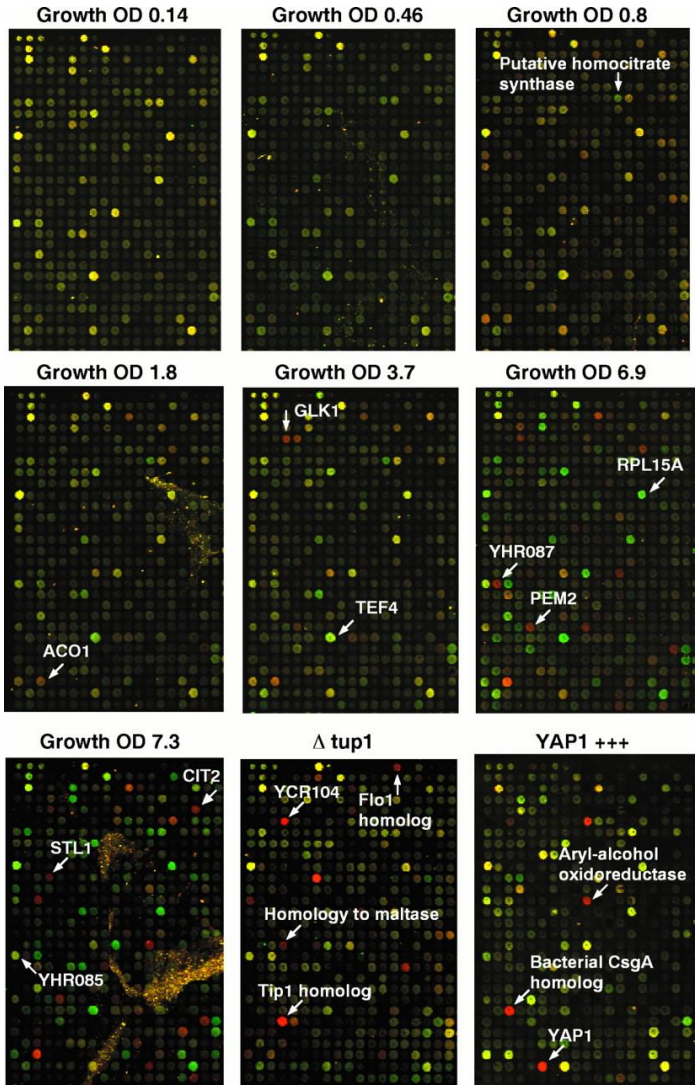
Allgemeines Schema eines *Affymetrix* GeneChips

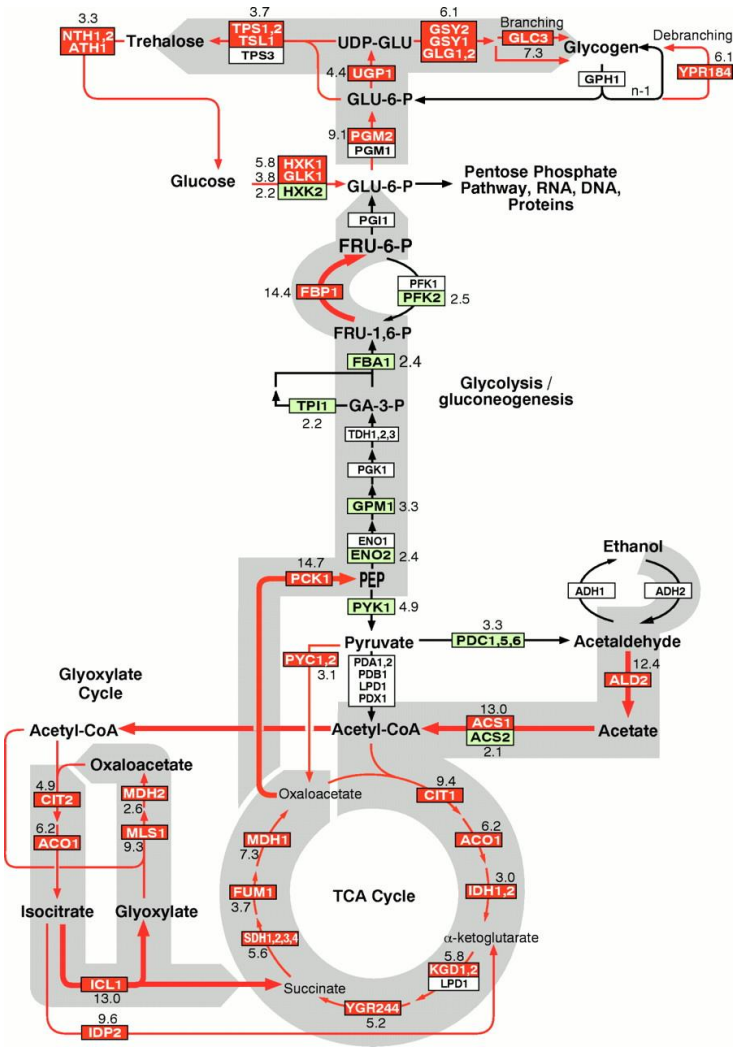


Yeast genome microarray. The size of the microarray is 18 x 18 mm. The image was obtained by a fluorescent scanning confocal microscope.

DeRisi, J.L., Iyer, V.R. & Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.

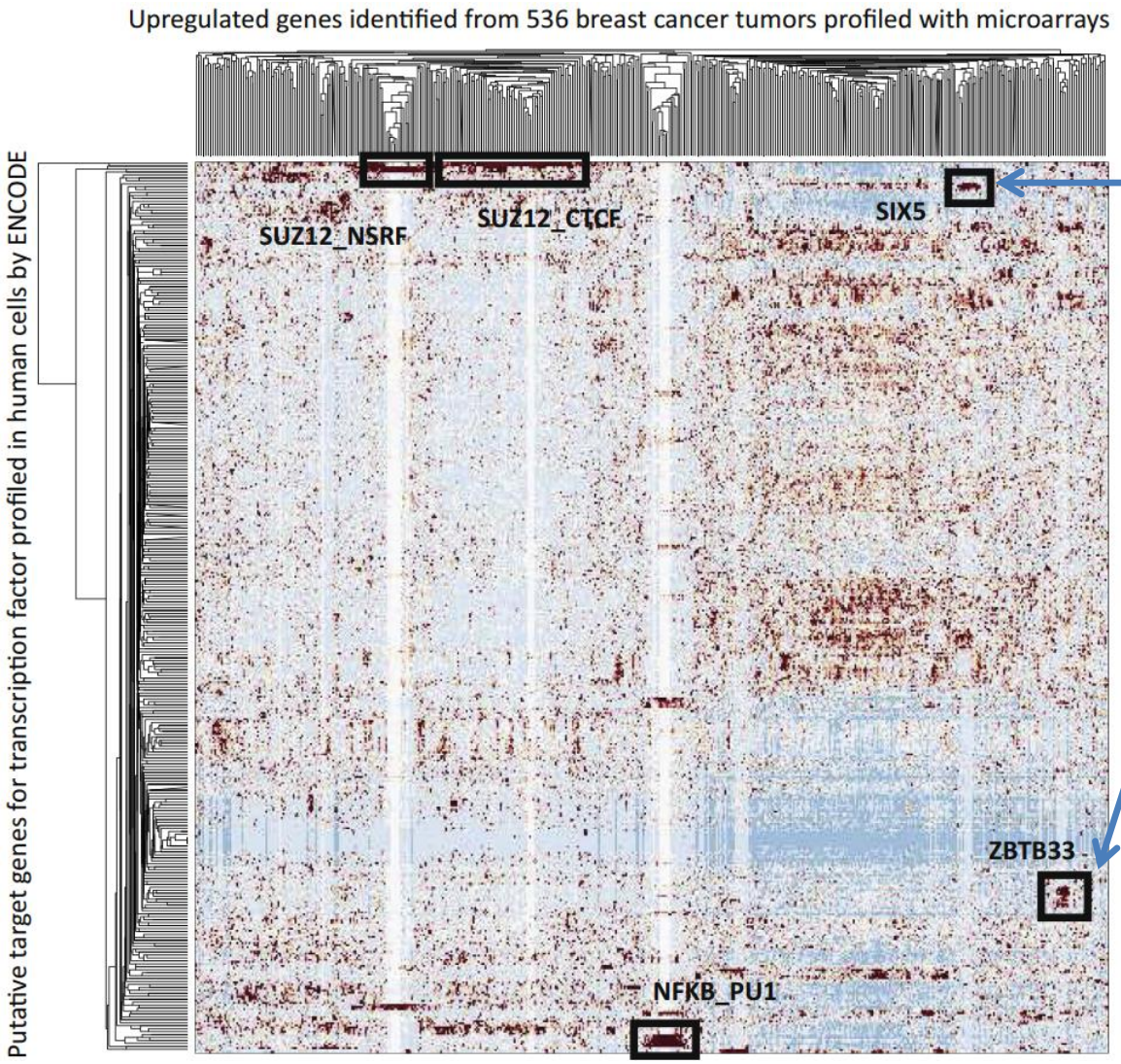
Zeitabhängige Genexpression metabolischer Enzyme in Hefe





DeRisi, J.L., Iyer, V.R. & Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.

Expressionsprofil individueller Brustkrebs-Tumore

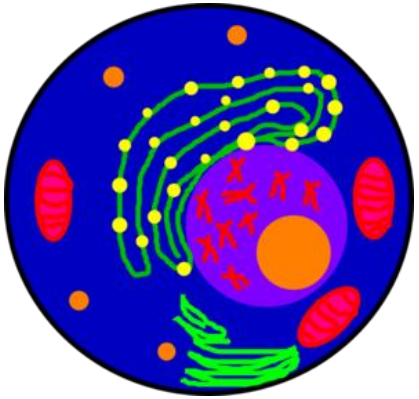


Cluster stark überlappender Gene

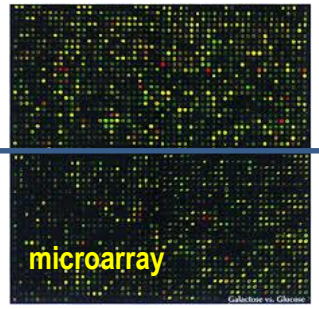
Ma'ayan, A., Rouillard, A. D., Clark, N. R., Wang, Z., Duan, Q. & Kou, Y. (2014) Lean Big Data integration in systems biology and systems pharmacology, *Trends Pharmacol Sci.* ***35, 450-60.***

Die Zelle als Punkt im hochdimensionalen molekularen Zustandsraum

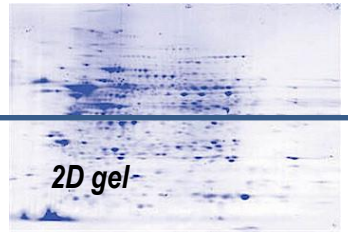
Zelle



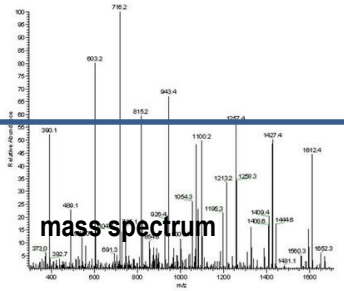
Transkriptom



Proteom



Metabolom



Zustandsvektor

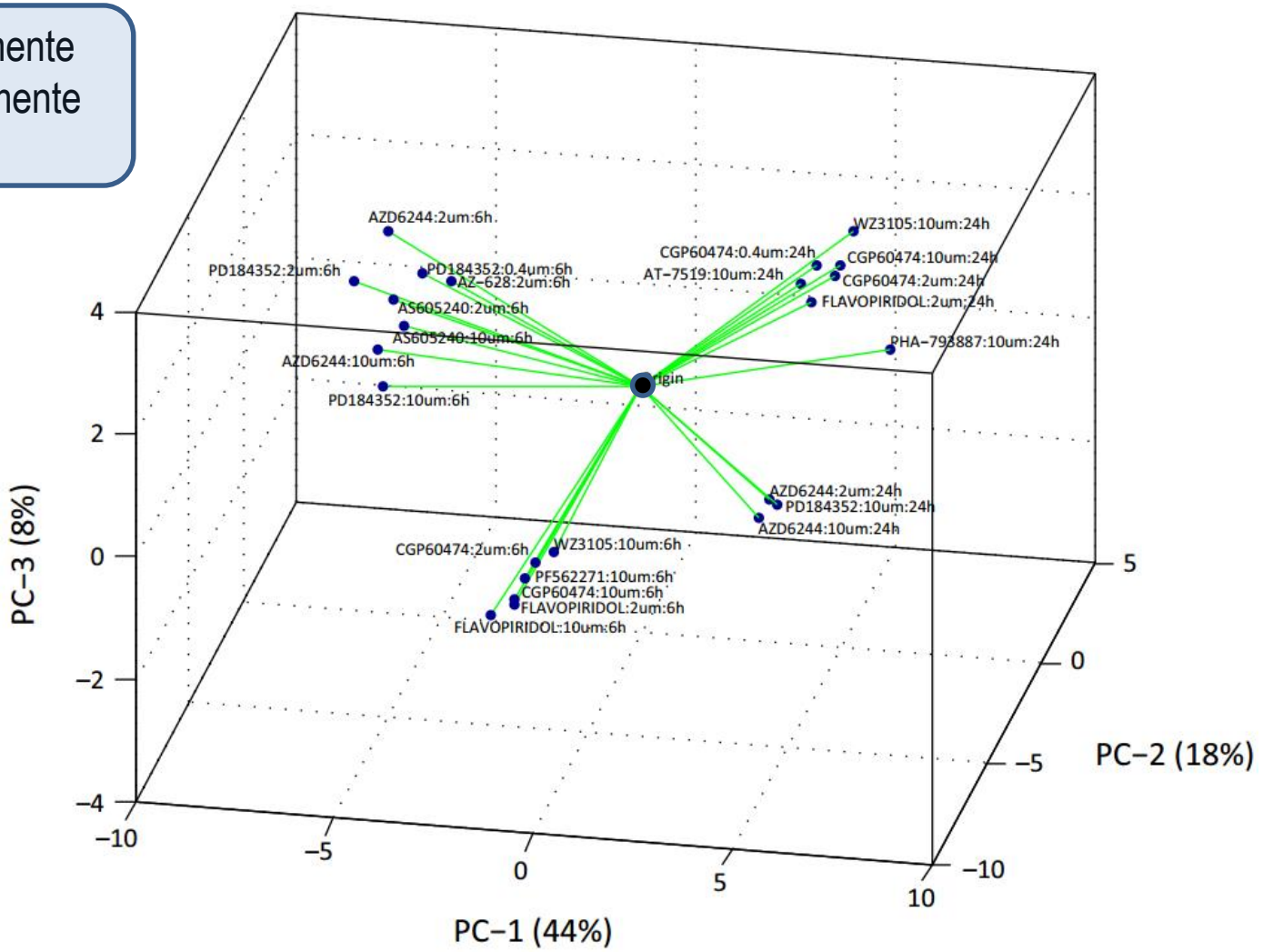
$mRNA_1$
$mRNA_2$
$mRNA_3$
•
•
$mRNA_{NR}$

$Protein_1$
$Protein_2$
$Protein_3$
•
•
$Protein_{NP}$

$Metabolit_1$
$Metabolit_2$
$Metabolit_3$
•
•
$Metabolit_{NM}$

Medikamentös induzierte Zustandsänderungen (breast cancer cell line)

88 Medikamente
352 Experimente
978 Gene



Ma'ayan, A., Rouillard, A. D., Clark, N. R., Wang, Z., Duan, Q. & Kou, Y. (2014) Lean Big Data integration in systems biology and systems pharmacology, *Trends Pharmacol Sci.* ***35, 450-60.***

Kontrolle der Genexpression

„DNA makes RNA makes Protein“

DNA

RNA-Polymerase

exon intron

promotor ATGCCTAGCGG TTTAAGCGC CCTAGGCCTTTAAGG AAAGGCGCGGGAAATT promotor

Transkriptionsfaktor

Transkriptionsfaktor

SIGNAL

Transkription
Splicing
Chemische
Modifikation

Transkriptionsfaktor

mRNA

cap TACGGATCGCC AAATTCGCG GGATCCGGAAATTCC TTTCCGCGCCCTTTAA AAAAAAAAAAAAAAA

Ribosom

Translation

Aminosäuren

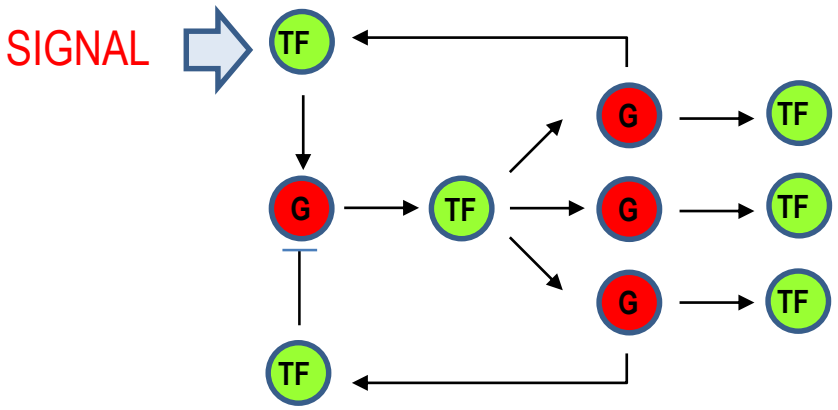
Protein

Strukturporotein
Enzym
Rezeptor
Membrankanal
Membrantransporter
Transkriptionsfaktor

Abbau
(Proteolyse)

Aminosäuren

Genregulatorische Netzwerke

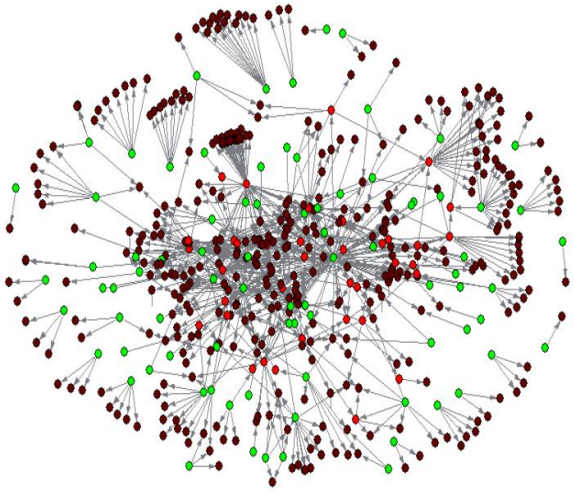


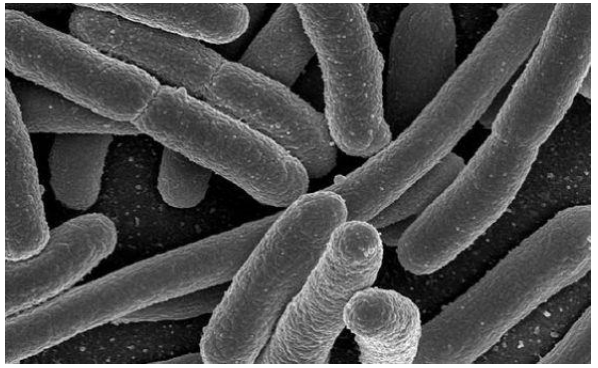
bipartiter Graph

Knoten = Gene, Transkriptionsfaktoren
Kanten = Expressionskontrolle (aktivierend, hemmend)

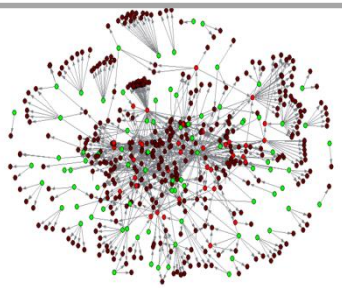
Hemmung
Aktivierung

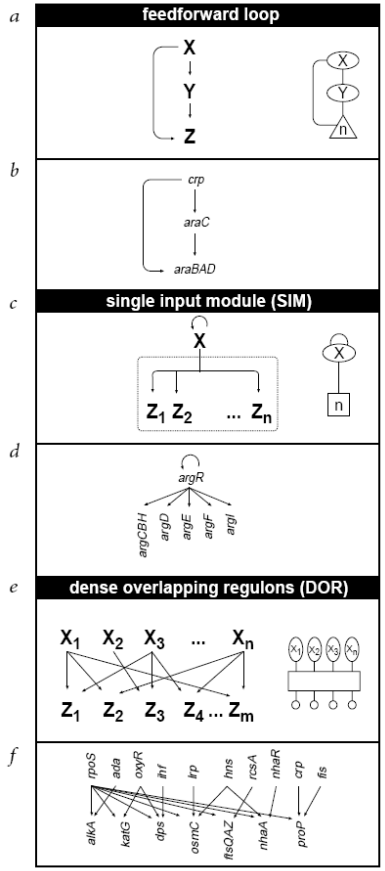
Beispiel: Genregulatorisches Netzwerk (E.coli)

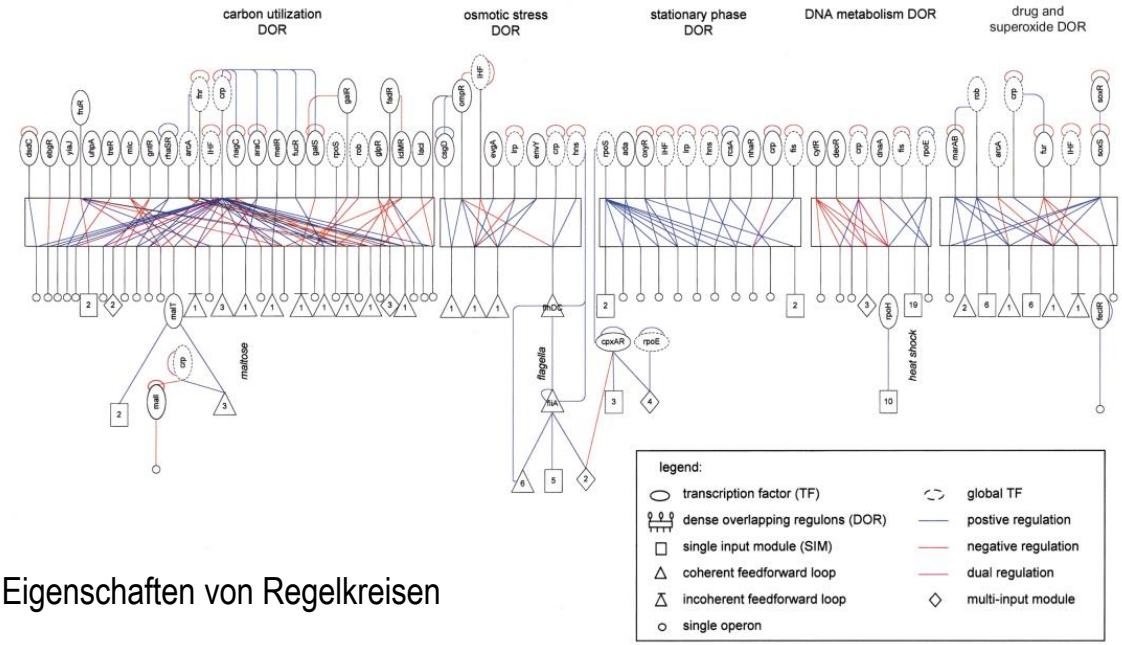




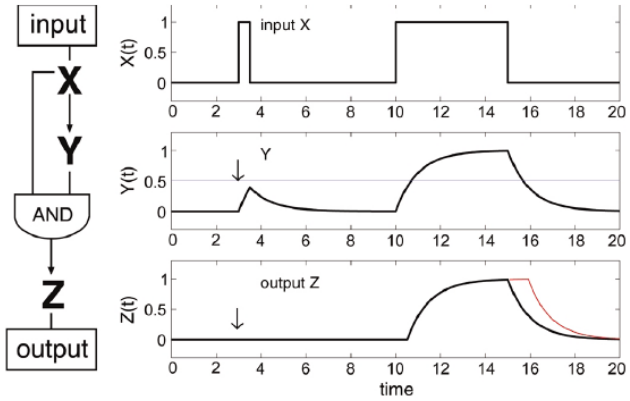
Allgemeine Netzwerkeigenschaften: Netzwerkmotive





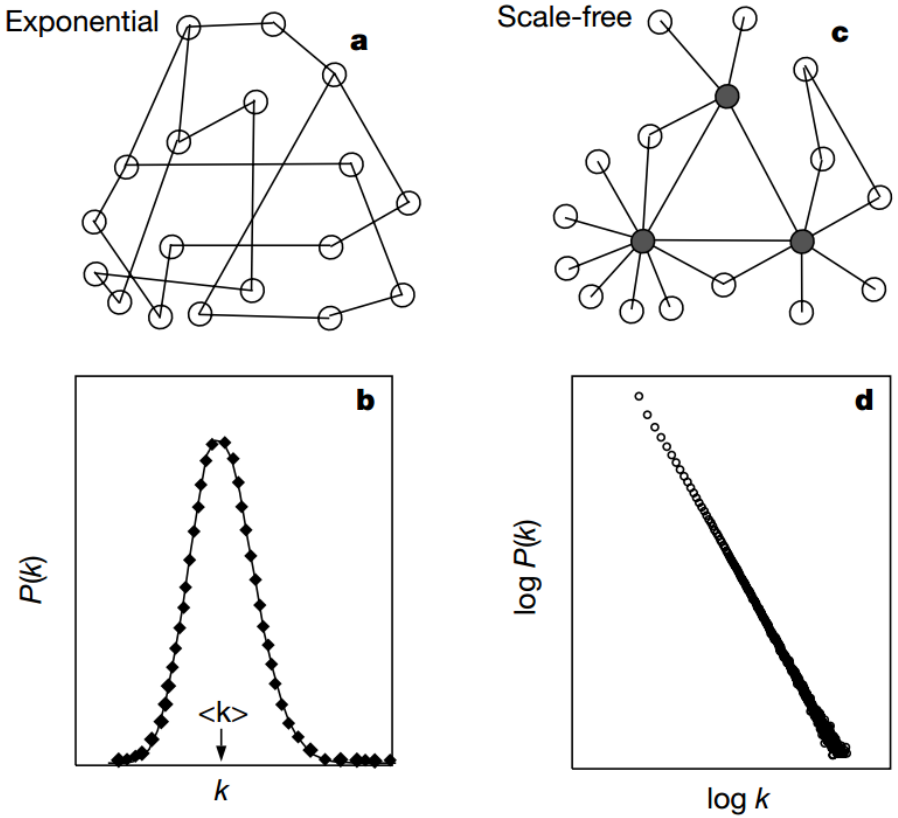


dynamische Eigenschaften von Regelkreisen

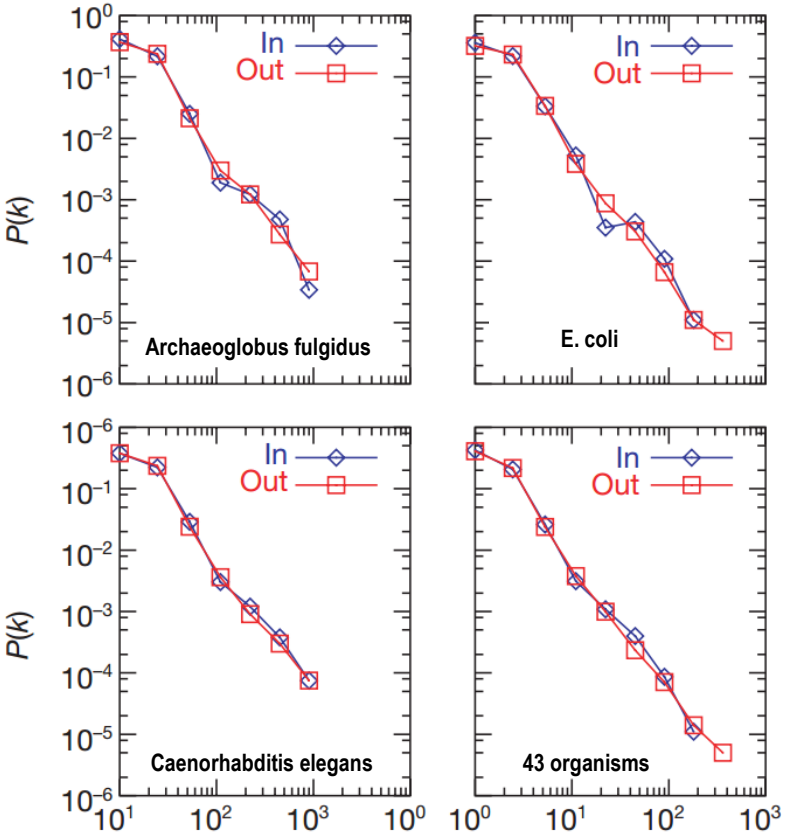


Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002) Network motifs in the transcriptional regulation network of Escherichia coli, *Nat Genet.* ***31, 64-8.***

Allgemeine Netzwerkeigenschaften: Stabilität (Vernetzungsgrad)



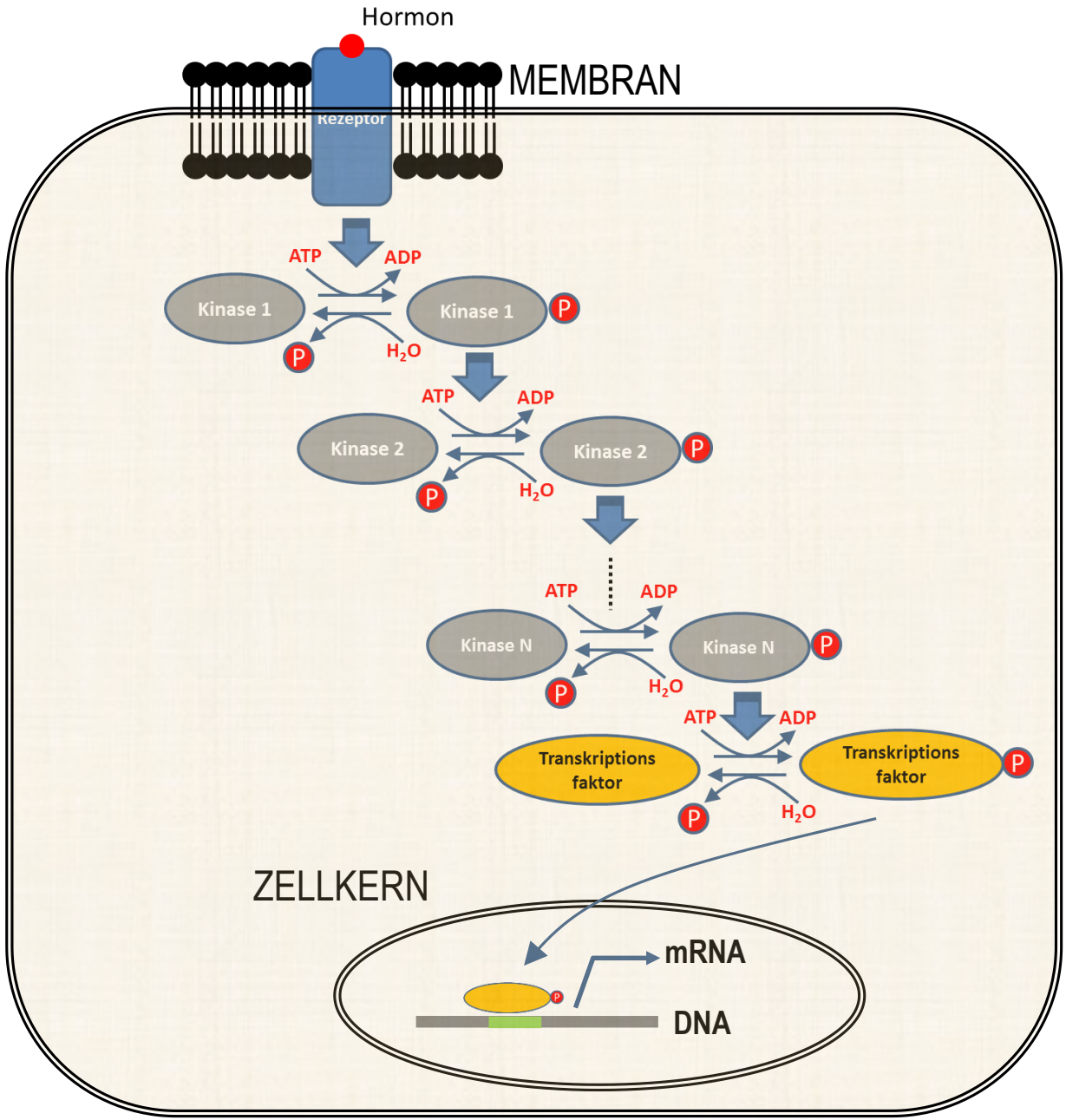
Wahrscheinlichkeit (Häufigkeit)
dass ein Knoten k Nachbarn hat:
$P(k) \sim k^{-\gamma}$



Skalenfreie Netzwerke gelten als robust gegen zufällige Knotenausfälle

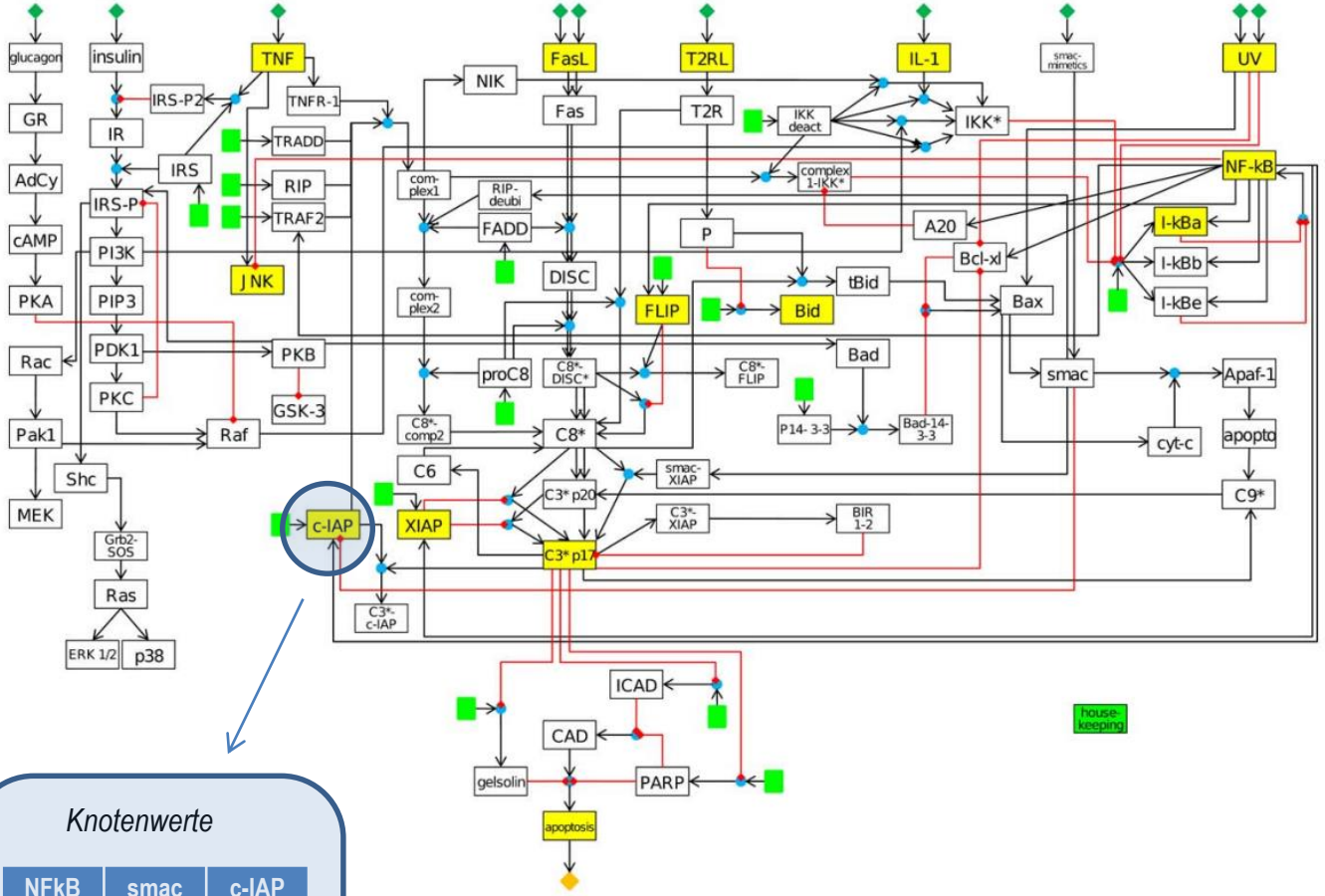
Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. (2000) The large-scale organization of metabolic networks, *Nature*. **407**, 651-4.

Regulation der Genexpression: Signalnetzwerke



Signalnetzwerk = Gesamtheit der molekularen Prozesse, die ein **Signal** (z.B. Hormon) in einen aktivierten Transkriptionsfaktor überführen

Boolesche Netzwerkmodelle (Input-Output-Verhalten)



Knotenwerte

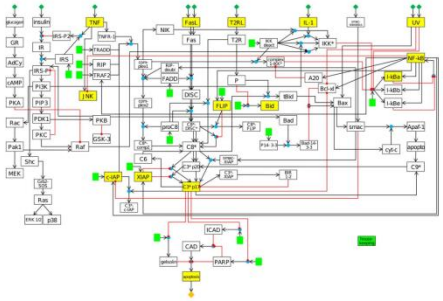
NFkB	smac	c-IAP
0	0	0
0.5	0	0.5
1	0	1
0	0.5	0
0.5	0.5	0
1	0.5	0
…	…	…

Graph eines Booleschen Booleschen Netzwerkes (Apoptose)

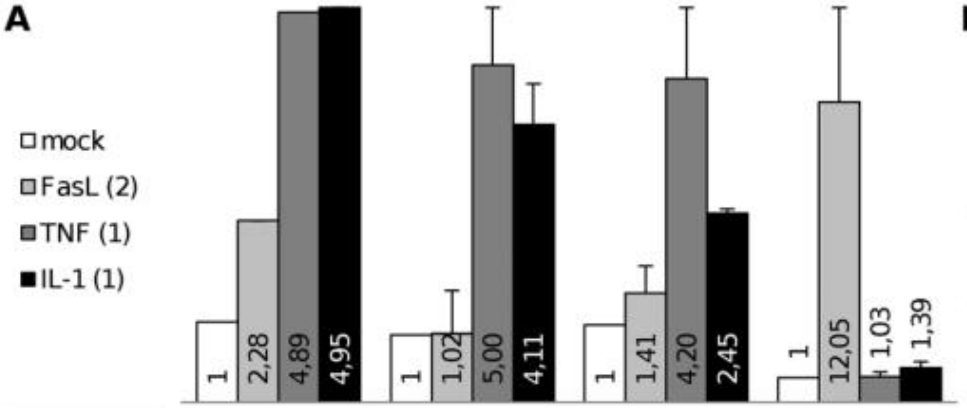
- ❑ 8 Input-Variable
- ❑ 86 Knoten
- ❑ 127 Kanten (Übergänge)
- ❑ Mehrfachwerte

Schlatter R, Schmich K, Avalos Vizcarra I, Scheurich P, Sauter T, Borner C, Ederer M, et al. ON/OFF and beyond--a boolean model of apoptosis. PLoS Comput Biol 2009;5:e1000595.

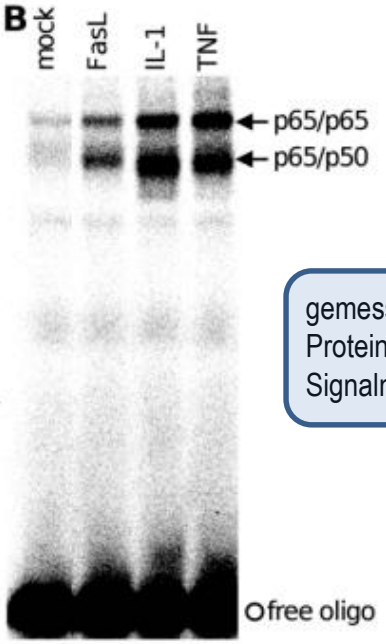
Kalibrierung von Booleschen Signalnetzwerken (“reverse engineering”)



Kombinationen von Zytokinen



node	NF-κB	c-IAP	FLIP	C3*p17
node value	0 0 1 1	1 1 2 2	1 1 2 2	0 2 0 0
measure	NF-κB	c-IAP2	FLIP	C3*p17
assay	EMSA	qRT-PCR	qRT-PCR	DEVDase
repeats	1	3	3	3



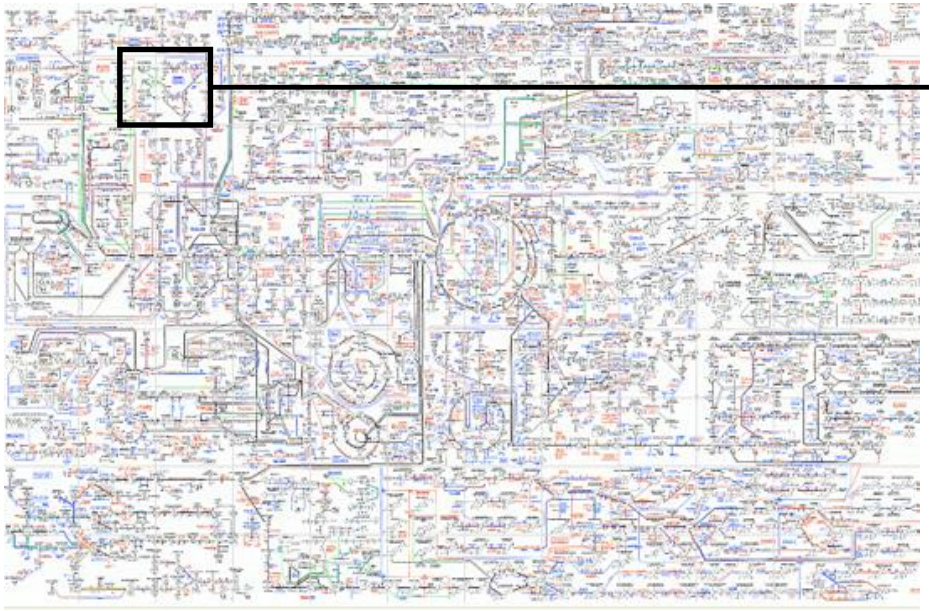
gemessene Proteinmengen des Signalnetzwerkes

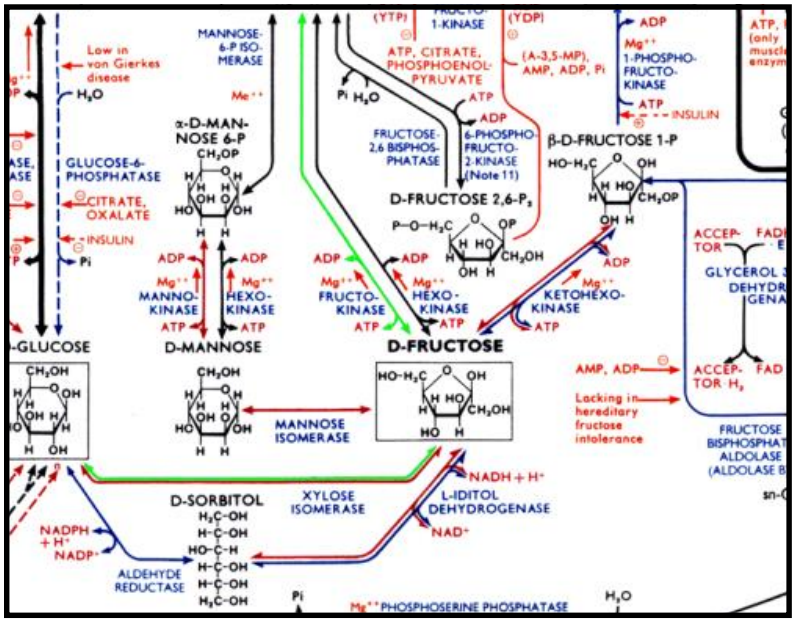
Vergleich Modell - Experiment

Schlatter R, Schmich K, Avalos Vizcarra I, Scheurich P, Sauter T, Borner C, Ederer M, et al. ON/OFF and beyond--a boolean model of apoptosis. PLoS Comput Biol 2009;5:e1000595.

Stoffwechselnetzwerke

Stoffwechselnetzwerk

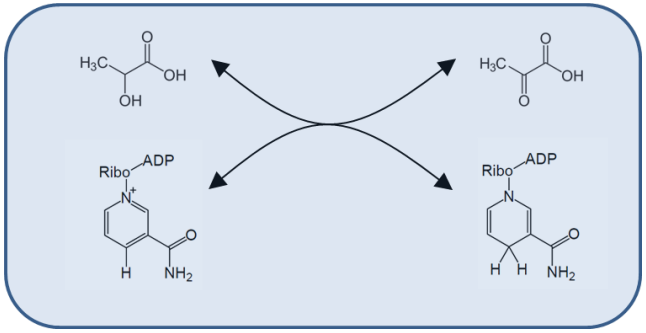




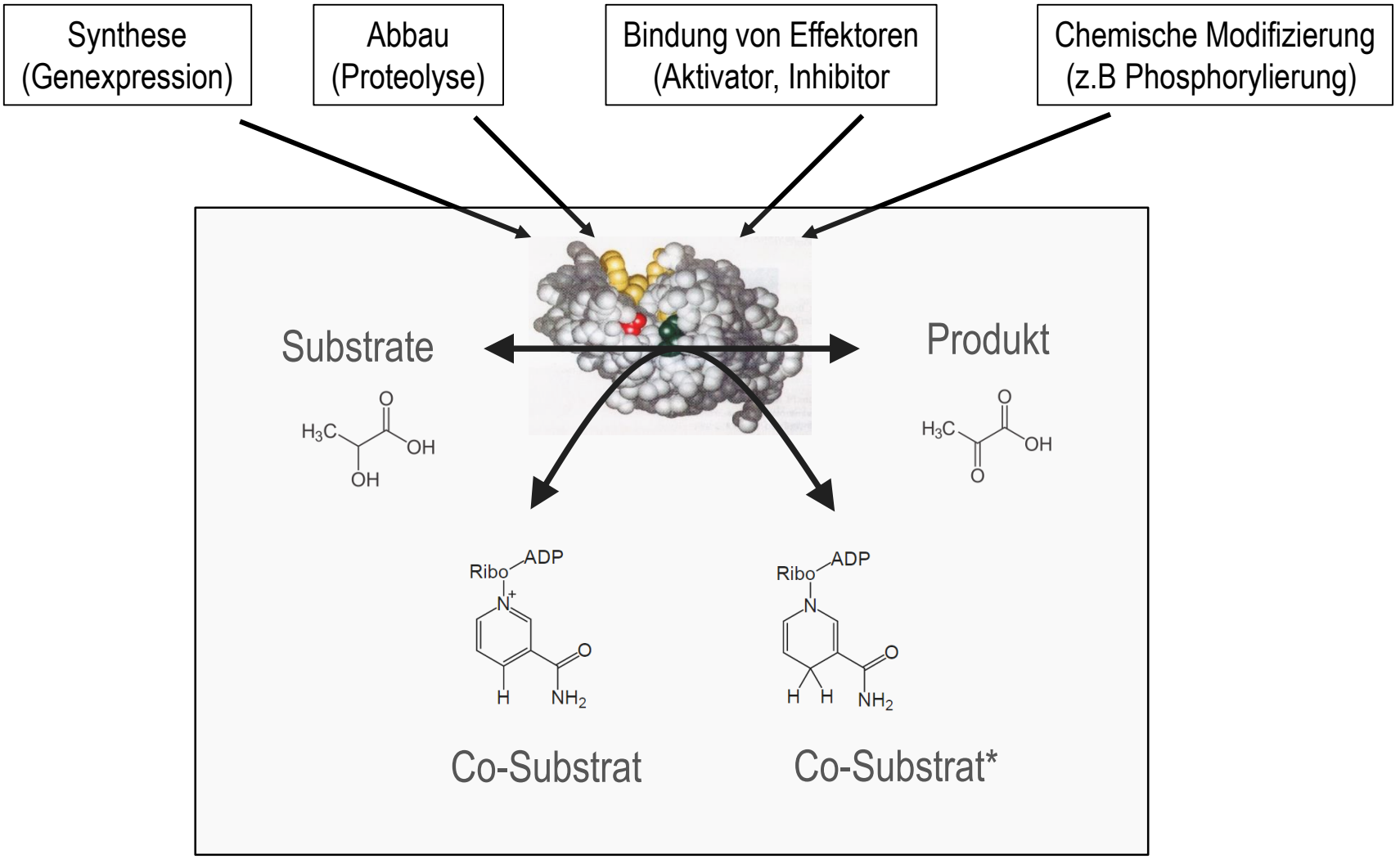
Reaktionsschema des zellulären Stoffwechselnetzwerks
(Aus: Biochemical Pathways, Roche Applied Science, http://www.expasy.org/tools/pathways/)

Moleküle/Molekülkomplexe/Organellen etc. = **Knoten**
Prozesse (chemische Reaktionen, Transporte) = **Kanten**

kleinste funktionelle Netzwerkeinheit = chemische Reaktion

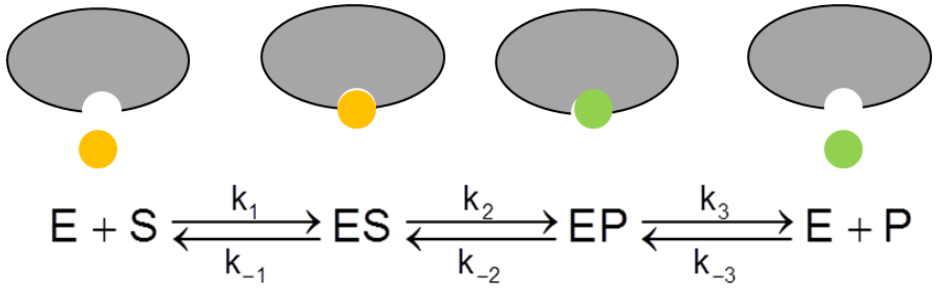


Stoffwechselregulation = Regulation der Enzymaktivität



Das Konzept der enzymatischen Ratengleichung

Reaktionsscheme einer einfachen monomolekularen Reaktion (Haldane 1930)



Approximation
(quasi steady-state)

$$v = \frac{\frac{V^+}{K_m^+}S - \frac{V^-}{K_m^-}P}{1 + \frac{S}{K_m^+} + \frac{P}{K_m^-}}$$

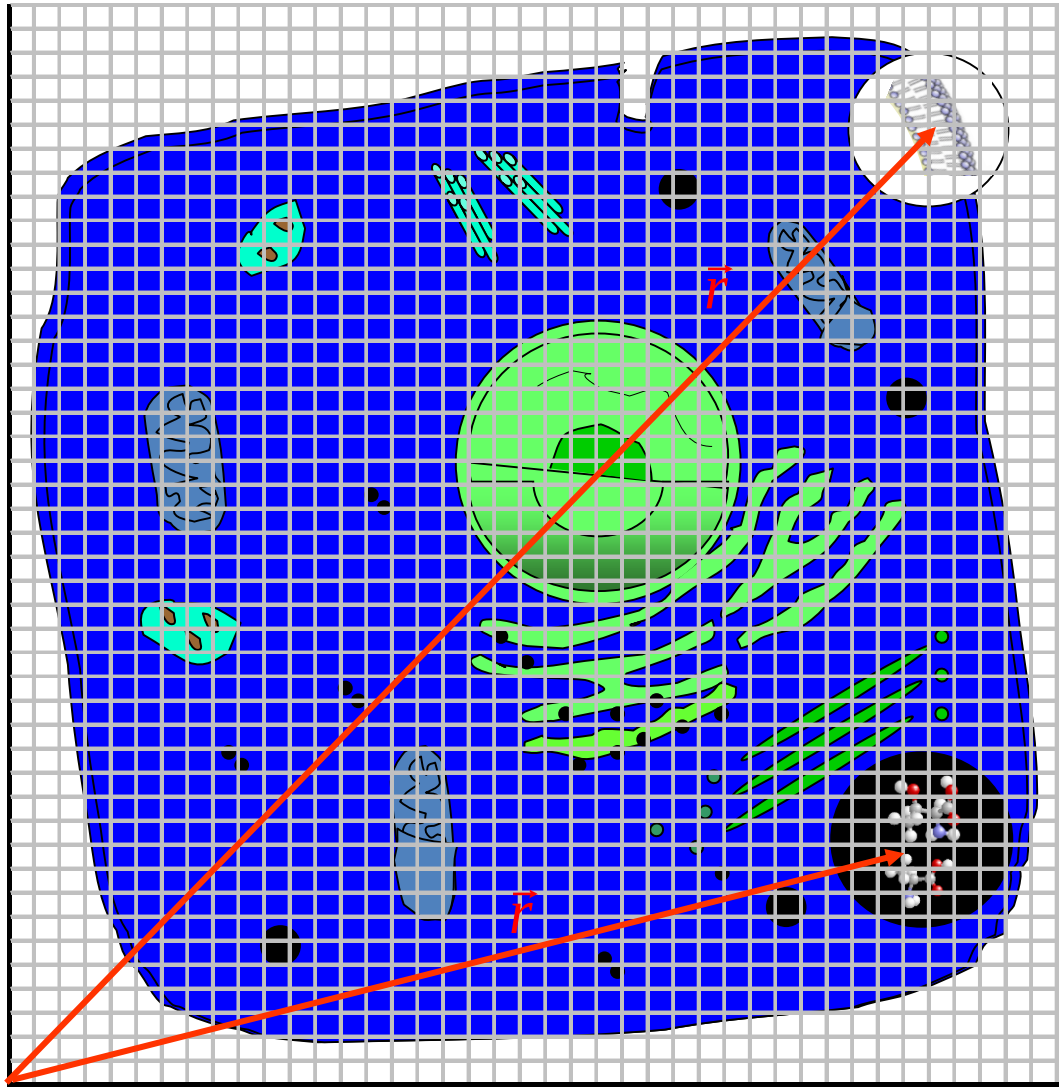
Ratengleichung

$$V^+ = \frac{k_2 k_3}{k_2 + k_3 + k_{-2}},\quad V^- = \frac{k_{-1}k_{-2}}{k_2 + k_{-1} + k_{-2}}$$

$$K_m^+ = \frac{k_2 k_3 + k_{-1}k_3 + k_{-1}k_{-2}}{k_1(k_2 + k_3 + k_{-2})}$$

$$K_m^- = \frac{k_2 k_3 + k_{-1}k_3 + k_{-1}k_{-2}}{k_{-3}(k_2 + k_{-1} + k_{-2})}.$$

Dynamische Netzwerkmodelle von Zellen



Ultimatives Ziel: Berechne

$$N_i(\vec{r},t)$$

= Zahl der Moleküle zum Zeitpunkt t an der Position $\vec{r}$

$$\frac{\partial N_i}{\partial t} = \sum_k D_k \frac{\partial^2 N_i}{\partial x_k^2} + \sum_j S_{ij}\, v_j(N_1, N_2, \ldots)$$

Reaktions-Diffusionsgleichungen

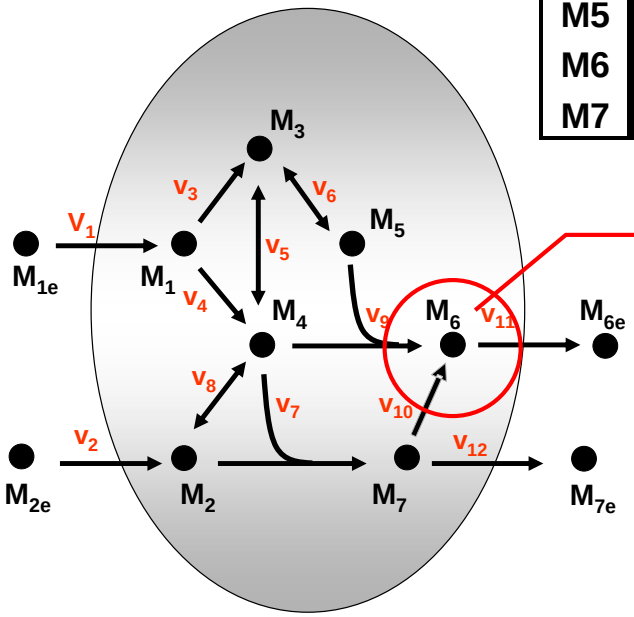
für komplette Zelle nicht explizit formulierbar:

- ❑ fehlende Parameterwerte
- ❑ keine räumlich aufgelösten Messungen
- ❑ Modellierung der Biogenese von Organellen unterentwickelt

Matrix-Repräsentation von Stoffwechselnetzwerken

Stöchiometrische Matrix

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12
M1e	-1	0	0	0	0	0	0	0	0	0	0	0
M2e	0	-1	0	0	0	0	0	0	0	0	0	0
M6e	0	0	0	0	0	0	0	0	0	0	1	0
M7e	0	0	0	0	0	0	0	0	0	0	0	1
M1	1	0	-1	-1	0	0	0	0	0	0	0	0
M2	0	1	0	0	0	0	-1	1	0	0	0	0
M3	0	0	1	0	-1	-1	0	0	0	0	0	0
M4	0	0	0	1	1	0	-1	-1	-1	0	0	0
M5	0	0	0	0	0	1	0	0	-1	0	0	0
M6	0	0	0	0	0	0	0	0	1	1	-1	0
M7	0	0	0	0	0	0	1	0	0	-1	0	-1



$$\frac{dM}{dt} = Sv$$

Traditionell: Kinetische Modellierung einzelner Stoffwechselwege

Biochem. J. (1976) **154**, 449–469
Printed in Great Britain

The Regulatory Principles of Glycolysis in Erythrocytes *in vivo* and *in vitro*

A MINIMAL COMPREHENSIVE MODEL DESCRIBING STEADY STATES, QUASI-STEADY STATES AND TIME-DEPENDENT PROCESSES

By TOM A. RAPOPORT
Akademie der Wissenschaften der DDR, Zentralinstitut für Molekularbiologie, Berlin-Buch, German Democratic Republic

and REINHART HEINRICH and SAMUEL M. RAPOPORT
Humboldt-Universität, Institut für Physiologische und Biologische Chemie, Berlin, German Democratic Republic

(*Received* 18 *April* 1975)

Eur. J. Biochem. *42*, 89–95 (1974)

A Linear Steady-State Treatment of Enzymatic Chains

General Properties, Control and Effector Strength

Reinhart HEINRICH and Tom A. RAPOPORT
Institut für Physiologische und Biologische Chemie der Humboldt-Universität Berlin

(Received March 13/August 7, 1973)

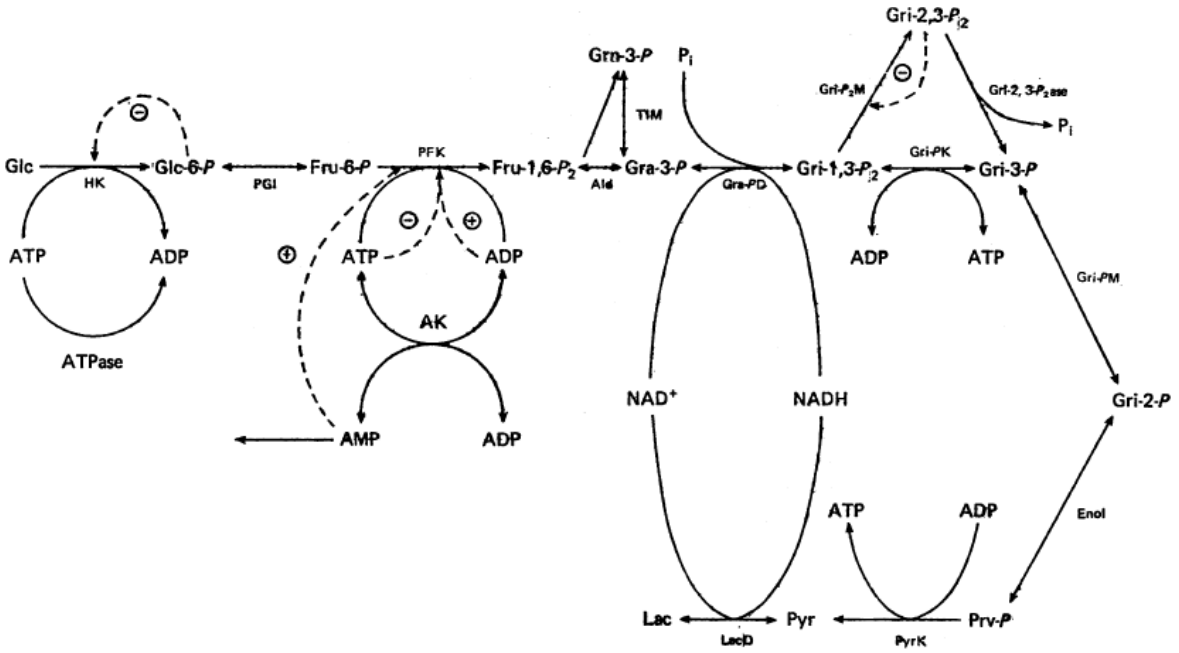
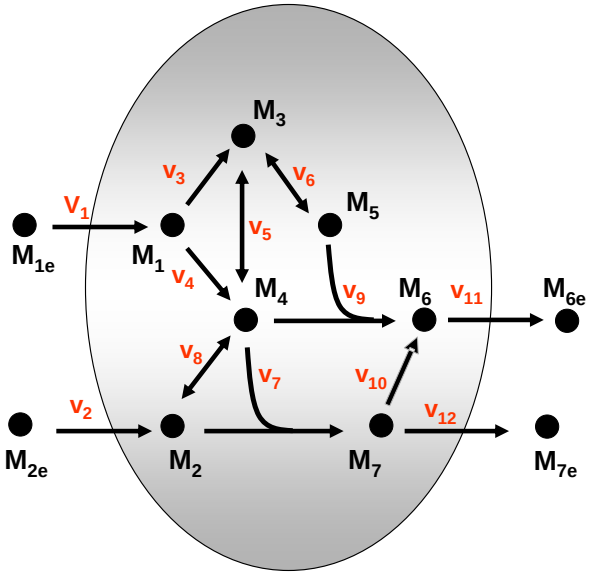


Fig. 1. *Scheme for glycolysis in erythrocytes considered for the model*

All reactions with arrows in both directions were considered to be near to equilibrium. Reactions indicated by an arrow in one direction were considered to be practically irreversible. The broken lines indicate activations (+) or inhibitions (–) of enzymes by metabolites, which were taken into account for the model.

Fluss-Bilanz-Analyse von metabolischen Netzwerken (FBA)



1378 FBA-Publikationen
im Zeitraum 2010-2016

für stationäre Zustände

$$\frac{dM}{dt} = Sv = 0 \qquad (1)$$

liefert Abhängigkeiten der Flüsse untereinander (Knotenbedingungen)

↓

Vorgabe einer Zielfunktion Z(v)

Beispiel: $Z = 2\ v_{11} + 3\ v_{12}$

↓

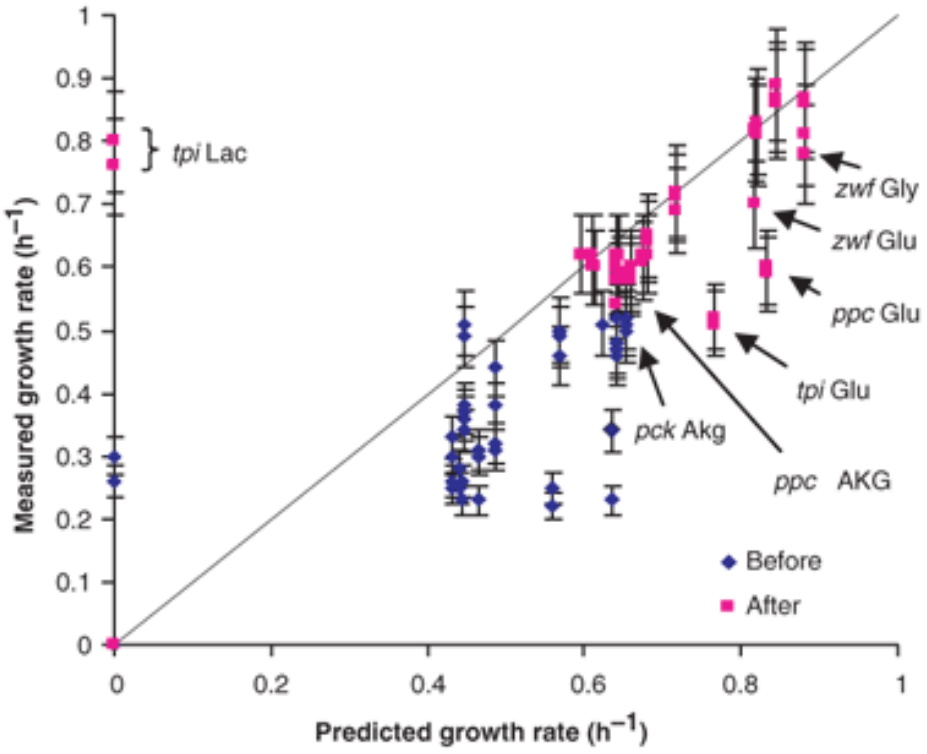
weitere Nebenbedingungen für v (2)

Beispiel: $v_1 <= 4$; $v_2 <= 3$

↓

Lösung des Optimierungsproblems
Z(v) → MAXIMUM
unter den Nebenbedingungen (1) und (2)

FBA-Modell des Stoffwechselnetzwerkes von E.coli



FBA model
762 Metabolite
932 Reaktionen
145 Membrantransporter

50 E. coli –Stämme mit Deletion eines der folgenden Enzyme

ackA (acetate kinase A)
frd (fumarate reductase)
pck (phosphoenolpyruvate carboxykinase)
ppc (phosphoenolpyruvate carboxylase)
tpi (triosephosphate isomerase)
zwf (glucose 6-phosphate-1-dehydrogenase)

Zielfunktion: Wachstumsrate

Für 39 der 50 Stämme (= 78%) weicht die berechneten Wachstumsrate < 10% von der gemessenen Wachstumsrate ab

Stephen S Fong & Bernhard Ø Palsson
Metabolic gene–deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes
Nature Genetics 36, 1056 - 1058 (2004)

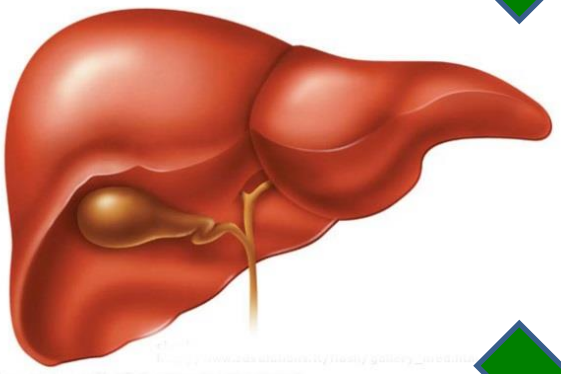
Die Leber: Das zentrale Stoffwechselorgan

SYNTHESE VON PLASMAPROTEINEN
Albumin
a1-Antitrypsin
Fibrinogen
Lipoproteine (VLDL, HDL A-I)

NAHRUNG

SYNTHESE VON PLASMAMETABOLITEN
Aminosäuren
Glukose
Ketonkörper
Kreatin

CIRCADIANE RHYTHMEN



KRANKHEIT
Diabetes
Fettleber
Tumore

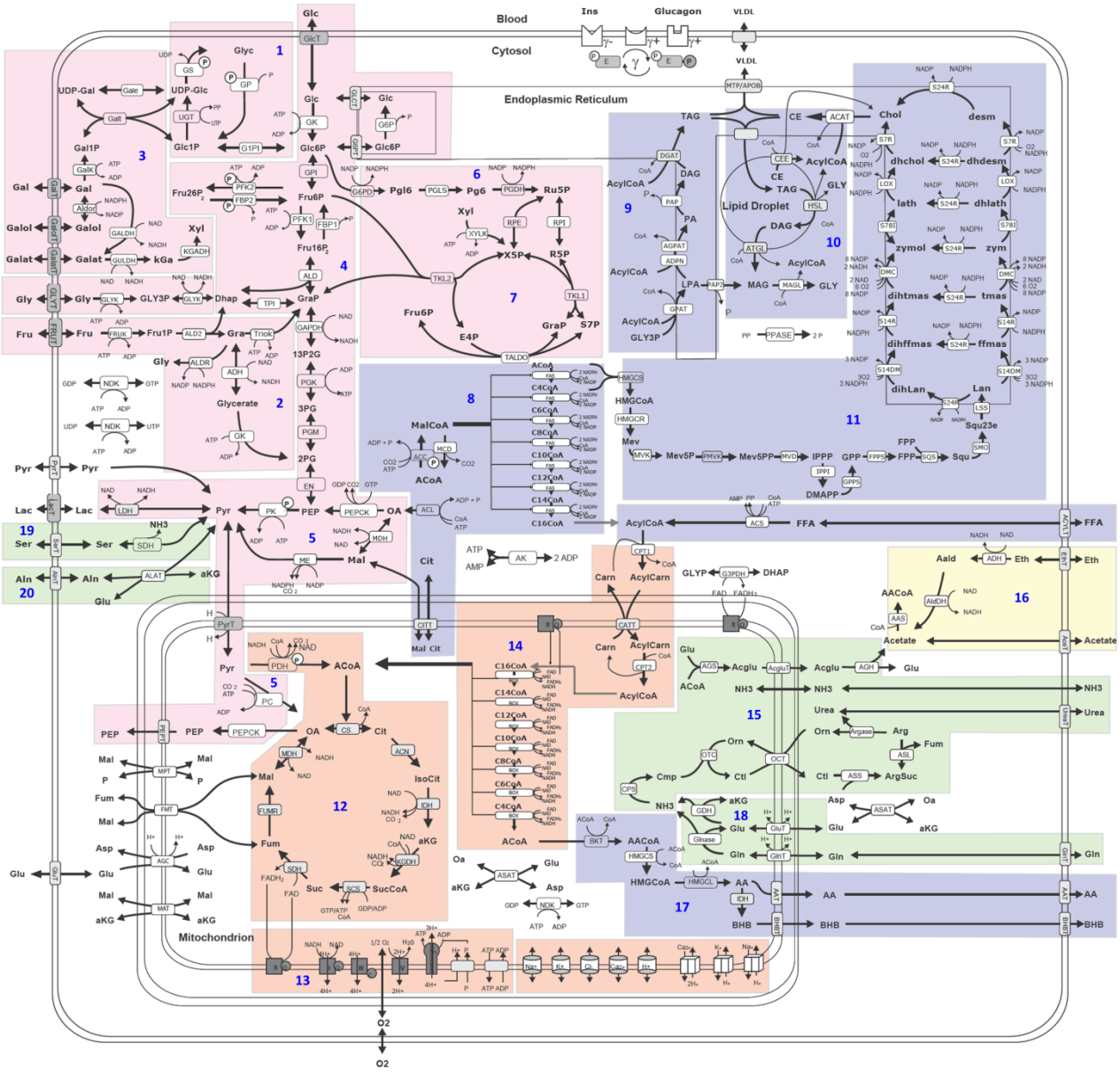
ALTERN

GALLEBILDUNG
Gallensäuren
Cholesterol
Phospholipide
Bilirubin
Medikamente

AKTIVITÄT

DETOXIFIZIERUNG
Medikamente
Hormone
Harnstoff

Beispiel: Kinetisches Modell des Zentralstoffwechsel von Leberzellen



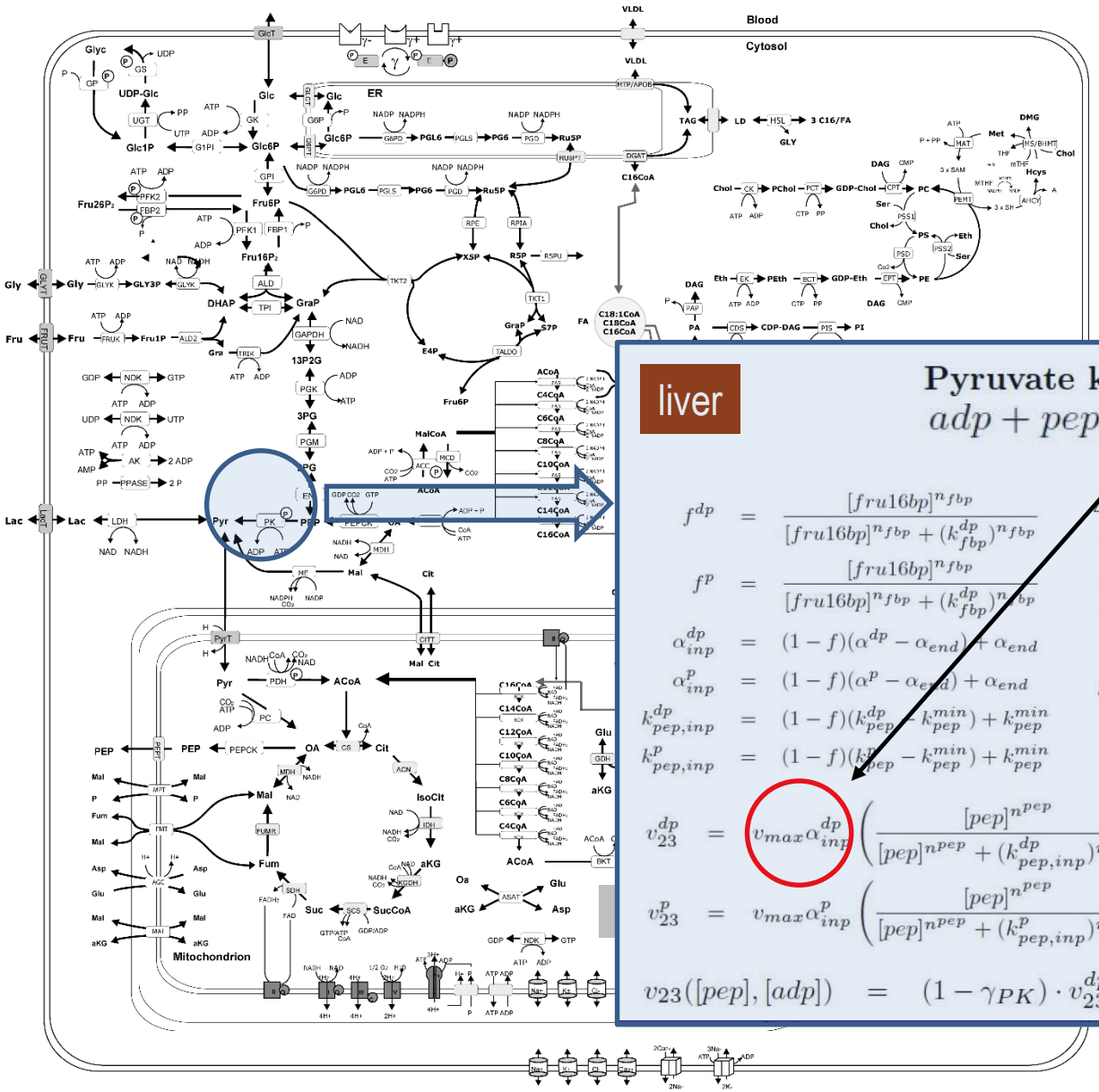
Stoffwechselwege

- Glykolyse
- Glukoneogenese
- Glykogenstoffwechsel
- Zitronensäurezyklus
- Pentosephosphat-Zyklus
- Fettsäuresynthese
- Fettsäureabbau (β-Oxidation)
- Triglyzeridsynthese
- Triglyzeridabbau
- Synthese von ketonkörpern
- Phospholipidsyntyhese
- Harnstoffzyklus

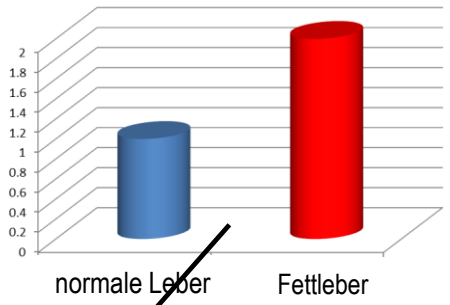
- Hormonelle Regulation von Enzymen (Insulin, Glukagon)

≈ 350 Variable
≈ 400 Reaktionen

Skalierung von Enzymaktivitäten mit gemessenen Proteinmengen



relative Proteinmengen



liver

Pyruvate kinase

$$adp + pep \rightarrow atp + pyr$$

$$f^{dp} = \frac{[fru16bp]^{n_{fbp}}}{[fru16bp]^{n_{fbp}} + (k^{dp}_{fbp})^{n_{fbp}}}$$

$$f^{p} = \frac{[fru16bp]^{n_{fbp}}}{[fru16bp]^{n_{fbp}} + (k^{dp}_{fbp})^{n_{fbp}}}$$

$$\alpha^{dp}_{inp} = (1-f)(\alpha^{dp} - \alpha_{end}) + \alpha_{end}$$

$$\alpha^{p}_{inp} = (1-f)(\alpha^{p} - \alpha_{end}) + \alpha_{end}$$

$$k^{dp}_{pep,inp} = (1-f)(k^{dp}_{pep} - k^{min}_{pep}) + k^{min}_{pep}$$

$$k^{p}_{pep,inp} = (1-f)(k^{p}_{pep} - k^{min}_{pep}) + k^{min}_{pep}$$

$$v^{dp}_{23} = v_{max}\alpha^{dp}_{inp}\left(\frac{[pep]^{n^{pep}}}{[pep]^{n^{pep}} + (k^{dp}_{pep,inp})^{n^{pep}}}\right)\left(\frac{[adp]}{[adp] + k_{adp}}\right)\left(a^{dp}_{base} + (1 - a^{dp}_{base})f^{dp}\right)$$

$$v^{p}_{23} = v_{max}\alpha^{p}_{inp}\left(\frac{[pep]^{n^{pep}}}{[pep]^{n^{pep}} + (k^{p}_{pep,inp})^{n^{pep}}}\right)\left(\frac{[adp]}{[adp] + k_{adp}}\right)\left(a^{p}_{base} + (1 - a^{p}_{base})f^{p}\right)$$

$$v_{23}([pep],[adp]) = (1-\gamma_{PK})\cdot v^{dp}_{23} + \gamma_{PK}\cdot v^{p}_{23}$$

$\Delta G^{0'} = -31.4\,\mathrm{kJ/mol}$

$n^{pep} = 3.5$

$n^{fbp} = 1.8$

$k^{dp}_{pep} = 0.58\,\mathrm{mM}$

$k^{p}_{pep} = 1.10\,\mathrm{mM}$

$k^{min}_{pep} = 0.08\,\mathrm{mM}$

$\alpha^{dp} = 1.0$

$\alpha^{p} = 1.1$

$\alpha_{end} = 1.0$

$k^{dp}_{fbp} = 0.16 \cdot 10^{-3}\,\mathrm{mM}$

$k^{p}_{fbp} = 0.35 \cdot 10^{-3}\,\mathrm{mM}$

$a^{dp}_{base} = 0.08$

$a^{p}_{base} = 0.04$

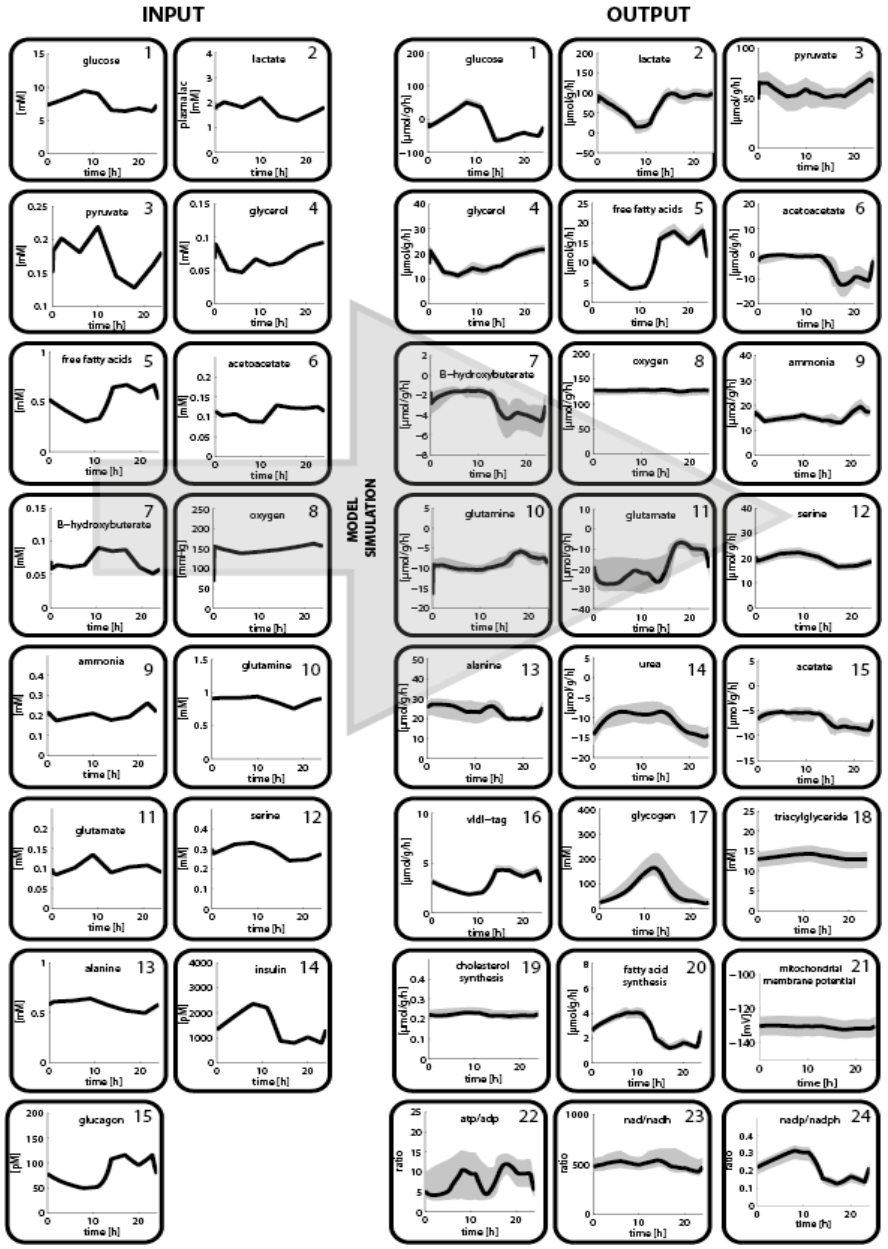
$k_{adp} = 2.3\,\mathrm{mM}$

Modellsimulation: Tagesrhythmus von Leberfunktionen

Eingangsgrößen

24h Plasmaprofil von Metaboliten und Hormonen

Plasma Metabolite & Hormone
glucose [mM]
glycerol [mM]
pyruvate [mM]
lactate [mM]
acetoacetate [mM]
β-hydroxybuterate [mM]
fatty acids [mM]
oxygen [Torr]
ammonia [mM]
glutamine [mM]
glutamate [mM]
alanine [mM]
serine [mM]
Insulin [pM]
Glucagon [pM]



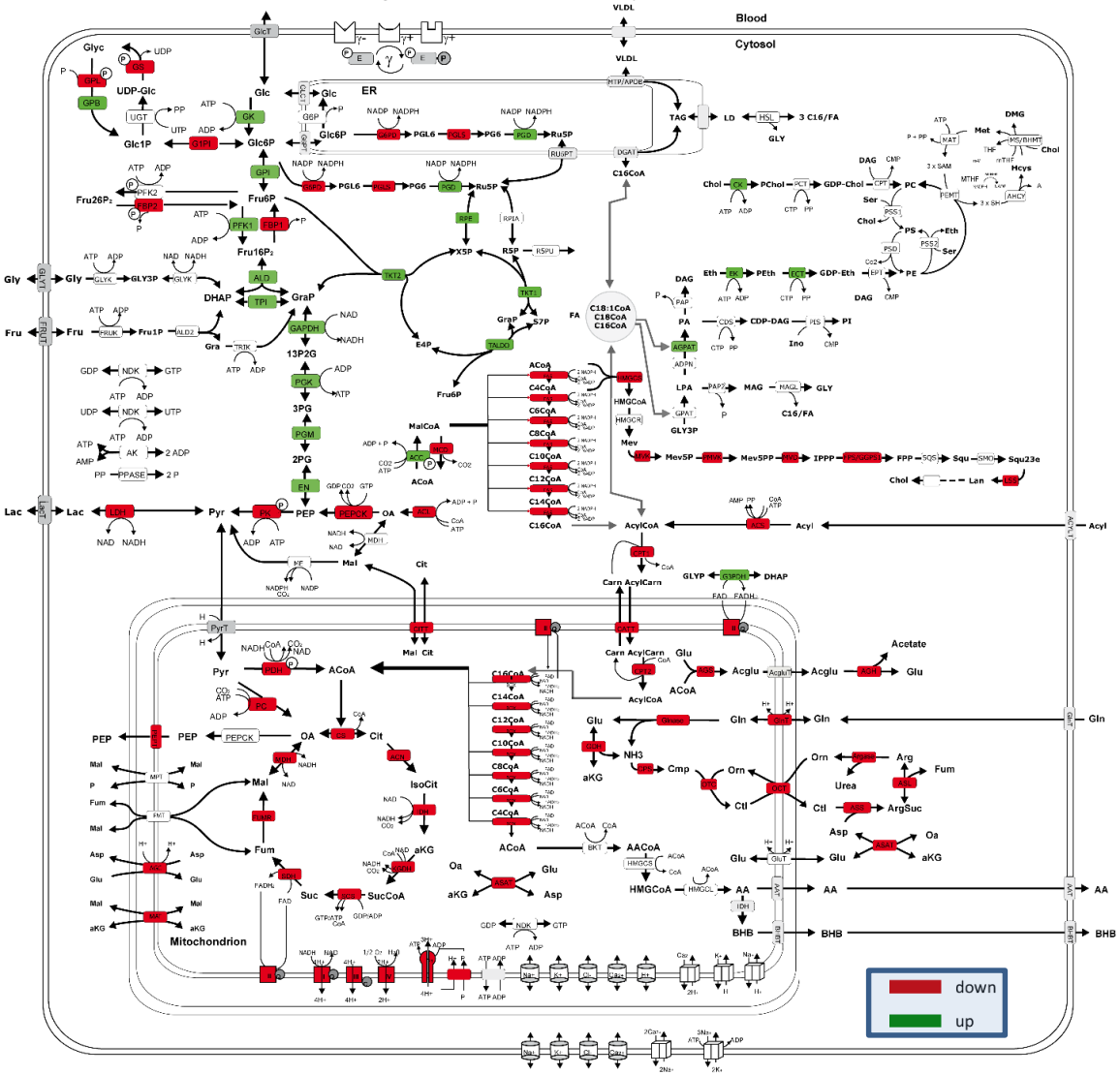
Ausgangsgrößen

24h Plasmaprofil des inneren Zustands und der Austauschflüsse

Interne Metabolie & Austausch-flüsse
glucose [µmol/g/h]
glycerol [µmol/g/h]
pyruvate [µmol/g/h]
lactate [µmol/g/h]
acetoacetate [µmol/g/h]
β-hydroxybuterate [µmol/g/h]
fatty acids [µmol/g/h]
oxygen [µmol/g/h]
ammonia [µmol/g/h]
glutamine [µmol/g/h]
glutamate [µmol/g/h]
alanine [µmol/g/h]
serine [µmol/g/h]
urea [µmol/g/h]
acetate [µmol/g/h]
vldl-tag [µmol/g/h]
glycogen [mM]
Triglycerides [mM]
cholesterolsynthesis [µmol/g/h]
fatty acid synthesis [µmol/g/h]

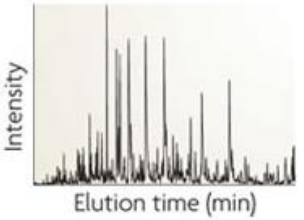
Veränderte Genexpression im Hepatozellulärem Karzinom

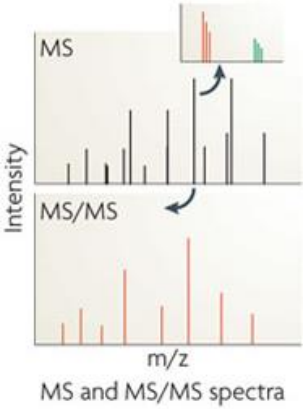
Proteinmengen HCC : hepatocyte



Proteinmengen bestimmt mit Massenspektrometrie

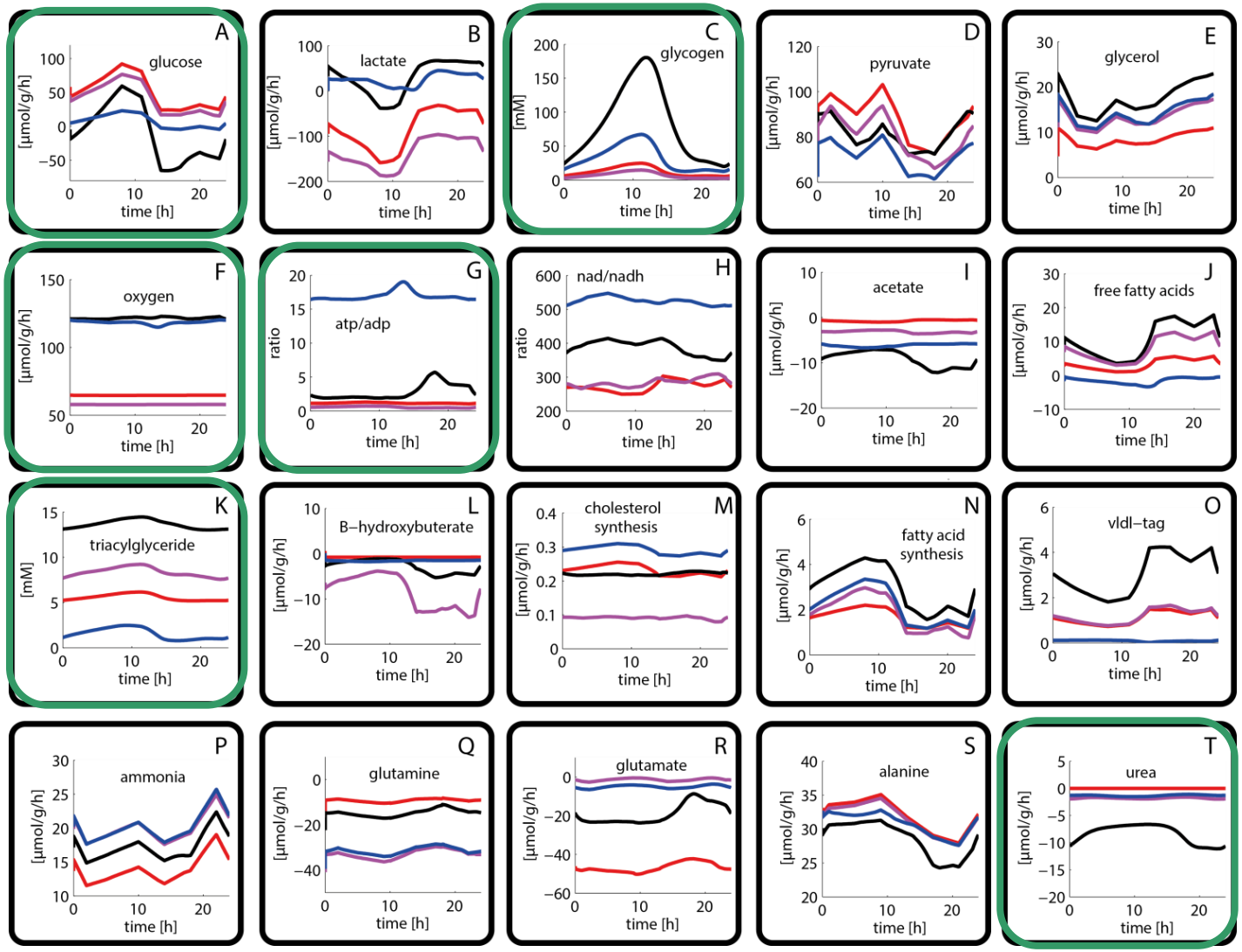
nLC-LTQ-OrbitrapMS/MS





MS and MS/MS spectra

Metabolische Charakterisierung von Lebertumoren



hepatocyte

T1 HCC

T3a HCC

adenoma

experimentell validiert

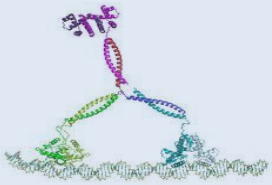
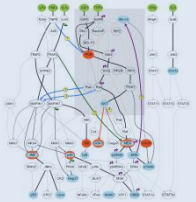
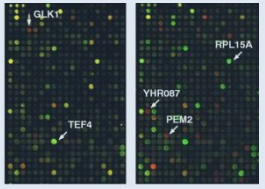
SYSTEMBIOLOGIE

ist ein stark *interdisziplinäres* Forschungsgebiet

erforscht die *molekularen Grundlagen* komplexer Lebensvorgänge

rekonstruiert, analysiert und modelliert *molekulare Netzwerke* auf verschiedenen Stufen der hierarchischen Organisation von Zellen, Geweben und Organen

benutzt moderne *Hochdurchsatz-Methoden* für die genomweite Quantifizierung von Nukleinsäuren, Proteinen, organischen Molekülen

braucht *Bioinformatik* und *Biostatistik* für die effektive Speicherung, Verwaltung und explorative Analyse der ständig anwachsenden Masse molekularer Daten („*BigData*“)

Biochemie
Molekularbiologie
Zellbiologie
Biophysik
Physiologie
Bioinformatik
Mathematik